

Worldwide Environmental Health and Mental Health

Part 1

Data Wrangling & Exploratory Data Analysis

Data Sources

Health Data Sources: 1990-2017

This notebook uses the following works:

Hannah Ritchie and Max Roser (2018) - "Opioids, cocaine, cannabis and illicit drugs". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/illicit-drug-use>' (<https://ourworldindata.org/illicit-drug-use>) [Online Resource]

Saloni Dattani, Hannah Ritchie and Max Roser (2021) - "Mental Health". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/mental-health>' (<https://ourworldindata.org/mental-health>) [Online Resource]

Hannah Ritchie and Max Roser (2018) - "Alcohol Consumption". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/alcohol-consumption>' (<https://ourworldindata.org/alcohol-consumption>) [Online Resource]

Hannah Ritchie and Max Roser (2019) - "Drug Use". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/drug-use>' (<https://ourworldindata.org/drug-use>) [Online Resource]

Which are based on the following data sources:

Institute of Health Metrics and Evaluation (IHME), Global Burden of Disease (GBD)

Data: Deaths, DALYs and prevalence of mental health and substance use disorders, by age and sex

Geographical coverage: Global by country and region

Time span: from 1990 onwards

Available at: <http://ghdx.healthdata.org/gbd-results-tool> (<http://ghdx.healthdata.org/gbd-results-tool>)

World Health Organization (WHO) International Classification of Diseases (ICD) World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines (Vol. 1). World Health Organization.

Data: Definitions and classifications of mental and substance use disorders

Available at: ICD-10 Classification of Mental and Behavioural Disorders

World Health Organization (WHO) Global Health Observatory (GHO)

Data: Prevalence of substance use disorders, trends in alcohol consumption

Geographical coverage: Global by country

Time span: Variable depending on datasets. Most inconsistent years post-2000
Available at: <http://www.who.int/gho/en/> (<http://www.who.int/gho/en/>)

Institute of Health Metrics & Evaluation (IHME), Global Burden of Disease (GBD)
Data: Deaths, DALYs and prevalence of substance use disorders, by age and sex
Geographical coverage: Global by country and region
Time span: since 1990
Available at: <http://ghdx.healthdata.org/gbd-results-tool> (<http://ghdx.healthdata.org/gbd-results-tool>)

World Health Organization (WHO) International Classification of Diseases (ICD) World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines (Vol. 1). World Health Organization.
Data: Definitions and classifications of mental and substance use disorders
Available at: ICD-10 Classification of Mental and Behavioural Disorders

World Health Organization (WHO) Global Health Observatory (GHO)
Data: Prevalence of substance use disorders, trends in alcohol consumption
Geographical coverage: Global by country
Time span: Variable depending on datasets. Most inconsistent years post-2000
Available at: <http://www.who.int/gho/en/> (<http://www.who.int/gho/en/>)

Micronutrient Status Data Sources: 1990-2019

World Health Organization (WHO) Vitamin and Mineral Nutrition Information System (VMNIS)
Data: Micronutrient/Indicators - Anaemia, Calcium, Copper, Fluoride, Folate, Iodine, Iron, Magnesium, Riboflavin, Selenium, Thiamine, Vitamin A, Vitamin B12, Vitamin B6, Vitamin C, Vitamin D, Vitamin E, Zinc.
Time Period: 1990-2019 (available for earlier years as well)
Available at: <https://www.who.int/teams/nutrition-and-food-safety/databases/vitamin-and-mineral-nutrition-information-system/data> (<https://www.who.int/teams/nutrition-and-food-safety/databases/vitamin-and-mineral-nutrition-information-system/data>)

Environmental Health Data Sources:

As a measure of outdoor air pollution:
World Health Organization (WHO) Global Health Observatory (GHO)
Data: Concentrations of fine particulate matter (PM2.5)
Geographical coverage: Global by country
Time span: 2010-2016
Available at: [\(https://www.who.int/data/gho/data/indicators/indicator-details/GHO/concentrations-of-fine-particulate-matter-\(pm2-5\)\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/concentrations-of-fine-particulate-matter-(pm2-5)) ([\(https://www.who.int/data/gho/data/indicators/indicator-details/GHO/concentrations-of-fine-particulate-matter-\(pm2-5\)\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/concentrations-of-fine-particulate-matter-(pm2-5)))

As a measure of indoor air pollution:
World Health Organization (WHO) Global Health Observatory (GHO)
Data: Proportion of population with primary reliance on clean fuels and technology for cooking (%)
Geographical coverage: Global by country

Time span: 2000-2019

Available at: <https://www.who.int/data/gho/data/themes/air-pollution/household-air-pollution>
<https://www.who.int/data/gho/data/themes/air-pollution/household-air-pollution>)

As a measure of water quality/clean water supply: World Health Organization (WHO) Global Health Observatory (GHO)

Data: Population using at least basic drinking-water services (%)

Geographical coverage: Global by country

Time span: 2000-2020

Available at: [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-at-least-basic-drinking-water-services-\(.\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-at-least-basic-drinking-water-services-(.))
[https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-at-least-basic-drinking-water-services-\(.\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-at-least-basic-drinking-water-services-(.)))

World Health Organization (WHO) Global Health Observatory (GHO)

Data: Proportion of population using safely managed drinking water services (%)

Geographical coverage: Global by country

Time span: 2000-2020

Available at: [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-safely-managed-drinking-water-services-\(.\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-safely-managed-drinking-water-services-(.))
[https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-safely-managed-drinking-water-services-\(.\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-using-safely-managed-drinking-water-services-(.)))

Temperature:

Data Hub

Data: Global Temperature Time Series

Time span: 1880-2016

Available at: <https://datahub.io/core/global-temp#python> (<https://datahub.io/core/global-temp#python>)

Abnormal weather events:

EM-DAT: The International Disaster Database

Time span: 1999-2025

Available at: <https://www.emdat.be/> (<https://www.emdat.be/>)

Health Expenditures Data Source

World Health Organization (WHO) Global Health Expenditure Database Data: Global Health

Expenditure Database Time span: 2000-2018 Available at:

<https://apps.who.int/nha/database>Select/Indicators/en>
<https://apps.who.int/nha/database>Select/Indicators/en>)

Imports

```
In [1]: import pandas as pd  
import matplotlib.pyplot as plt
```

Mental Health Data

In [2]: #mental health

```
df3 = pd.read_csv('WorldInData/alcohol-drug-use-disorders-share-total-disea
df4 = pd.read_csv('WorldInData/alcohol-or-drug-use-disorders-male-vs-female
df6 = pd.read_csv('WorldInData/annual-number-of-deaths-by-cause.csv')
df12 = pd.read_csv('WorldInData/daly-rates-from-anxiety-disorders-by-age.cs
df13 = pd.read_csv('WorldInData/daly-rates-from-bipolar-disorder-by-age.csv'
df14 = pd.read_csv('WorldInData/daly-rates-from-schizophrenia-by-age.csv')
df16 = pd.read_csv('WorldInData/dalys-from-eating-disorders-by-age.csv')
df17 = pd.read_csv('WorldInData/dalys-from-mental-and-substance-disorder-by
df18 = pd.read_csv('WorldInData/dalys-from-mental-health-and-substance-use-
df19 = pd.read_csv('WorldInData/dalys-rate-mental-and-neurodevelopmental-di
df20 = pd.read_csv('WorldInData/death-rate-amphetamine.csv')
df21 = pd.read_csv('WorldInData/death-rate-from-opioid-use.csv')
df22 = pd.read_csv('WorldInData/death-rates-cocaine.csv')
df23 = pd.read_csv('WorldInData/death-rates-from-alcohol-use-disorders.csv'
df24 = pd.read_csv('WorldInData/death-rates-from-drug-use-disorders.csv')
df26 = pd.read_csv('WorldInData/death-rates-from-mental-and-substance-disor
df28 = pd.read_csv('WorldInData/death-rates-substance-disorders.csv')
df30 = pd.read_csv('WorldInData/deaths-due-to-illicit-drug-use-by-age.csv')
df31 = pd.read_csv('WorldInData/deaths-from-alcohol-and-drug-use-disorders.
df33 = pd.read_csv('WorldInData/deaths-from-drug-use-disorders-by-age.csv')
df34 = pd.read_csv('WorldInData/deaths-from-eating-disorders.csv')
df35 = pd.read_csv('WorldInData/deaths-illicit-drugs.csv')
df36 = pd.read_csv('WorldInData/deaths-substance-disorders-age.csv')
df37 = pd.read_csv('WorldInData/deaths-substance-disorders.csv')
df39 = pd.read_csv('WorldInData/depression-daly-rates-by-age.csv')
df43 = pd.read_csv('WorldInData/drug-disorder-dalys-rate.csv')
df48 = pd.read_csv('WorldInData/mental-health-share-of-total-disease-burden
df49 = pd.read_csv('WorldInData/number-of-deaths-by-risk-factor.csv')
df50 = pd.read_csv('WorldInData/number-of-people-with-bipolar-disorder.csv'
df51 = pd.read_csv('WorldInData/number-of-people-with-depression.csv')
df53 = pd.read_csv('WorldInData/number-with-adhd.csv')
df55 = pd.read_csv('WorldInData/number-with-an-eating-disorder.csv')
df56 = pd.read_csv('WorldInData/number-with-anorexia-and-bulimia-nervosa.cs
df58 = pd.read_csv('WorldInData/number-with-anxiety-disorders.csv')
df64 = pd.read_csv('WorldInData/number-with-mental-and-substance-use-disord
df65 = pd.read_csv('WorldInData/number-with-mental-health-disorders-by-sex.
df66 = pd.read_csv('WorldInData/number-with-schizophrenia.csv')
df69 = pd.read_csv('WorldInData/prevalence-by-mental-and-substance-use-diso
df71 = pd.read_csv('WorldInData/prevalence-of-alcohol-use-disorders-by-age.
df73 = pd.read_csv('WorldInData/prevalence-of-anxiety-disorders-by-age.csv'
df75 = pd.read_csv('WorldInData/prevalence-of-bipolar-disorder-by-age.csv')
df78 = pd.read_csv('WorldInData/prevalence-of-depression-by-age.csv')
df80 = pd.read_csv('WorldInData/prevalence-of-drug-use-disorders-by-age.csv
df82 = pd.read_csv('WorldInData/prevalence-of-eating-disorders-by-age.csv')
df84 = pd.read_csv('WorldInData/prevalence-of-mental-and-substance-disorder
df85 = pd.read_csv('WorldInData/prevalence-of-schizophrenia-by-age.csv')
df92 = pd.read_csv('WorldInData/share-of-total-disease-burden-by-cause.csv'
df94 = pd.read_csv('WorldInData/share-with-alcohol-or-drug-use-disorders.cs
df107 = pd.read_csv('WorldInData/share-with-mental-health-or-development-di
df110 = pd.read_csv('WorldInData/substances-risk-factor-vs-direct-deaths.cs
```

Alcohol and substance use disorders

In [3]:

```
df3.info()
df3.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6156 entries, 0 to 6155
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype  
---  --  
0   Entity    object 
6156 non-null   object 
1   Code      object 
5292 non-null   object 
2   Year      int64  
6156 non-null   int64  
3   DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent) 6156 non-null   float64 
dtypes: float64(1), int64(1), object(2)
memory usage: 192.5+ KB
```

Out[3]:

	Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)
0	Afghanistan	AFG	1990	0.334482
1	Afghanistan	AFG	1991	0.338543
2	Afghanistan	AFG	1992	0.348761
3	Afghanistan	AFG	1993	0.353914
4	Afghanistan	AFG	1994	0.351862

Checking for missing values:

In [4]:

```
missing = pd.concat([df3.isnull().sum(), 100 * df3.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[4]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)		0	0.000000
Code	864	14.035088	

In [5]: `df3.describe()`

Out[5]:

Year	DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)
count	6156.000000
mean	2003.000000
std	7.789514
min	1990.000000
25%	1996.000000
50%	2003.000000
75%	2010.000000
max	2016.000000

In [6]: `v1='DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)'`

In [7]: `df3.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()`

Out[7]:

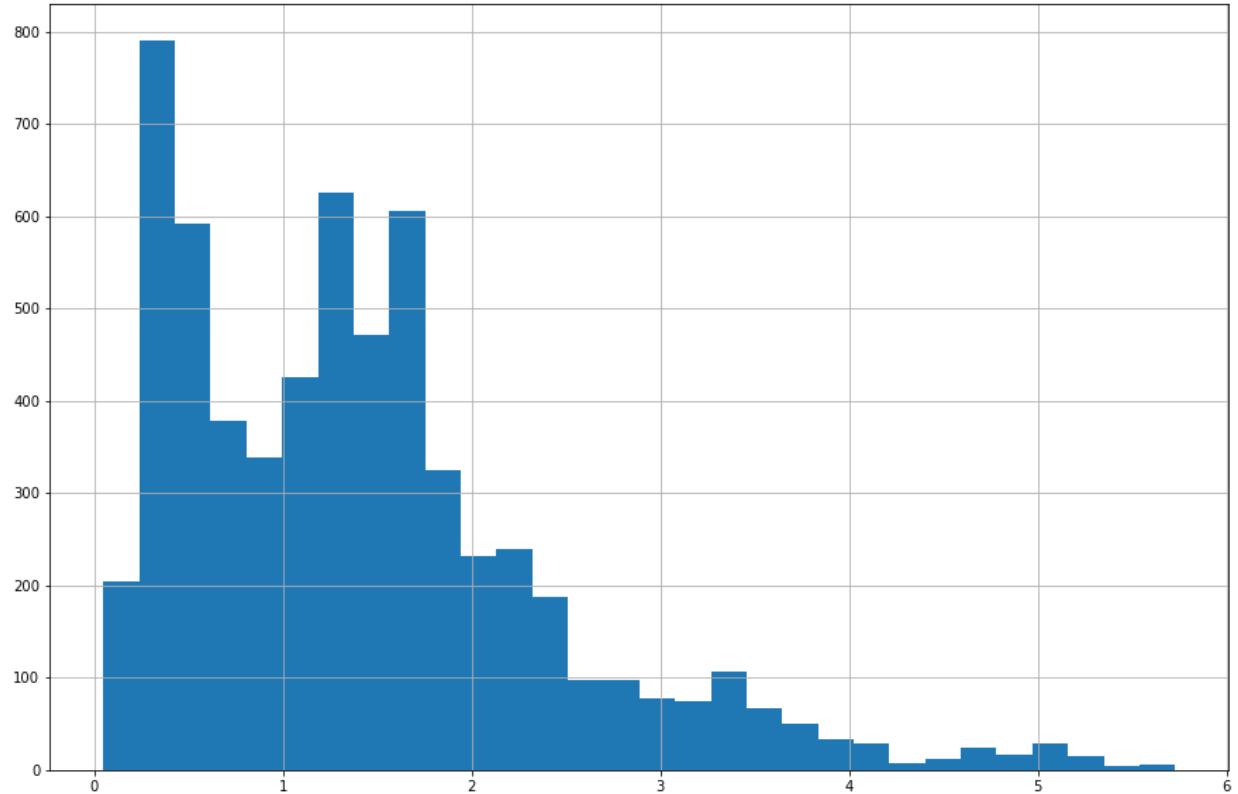
Entity	Year	DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)	
4495	Russia	2003	5.727319
4496	Russia	2004	5.725393
4494	Russia	2002	5.692551
4493	Russia	2001	5.667678
4497	Russia	2005	5.659647

In [8]: `df3.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()`

Out[8]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)	
4513	Rwanda	1994	0.048979
3861	Niger	1990	0.131985
3862	Niger	1991	0.132554
3863	Niger	1992	0.134088
3864	Niger	1993	0.135072

```
In [9]: df3[v1].hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [10]: df3.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[10]: Entity
Niger          0.203441
Mali           0.219912
Burkina Faso  0.230399
South Sudan    0.232517
Chad           0.238980
Name: DALYs (Disability-Adjusted Life Years) - Alcohol and substance use
disorders (Age-standardized percent), dtype: float64
```

```
In [11]: df3.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[11]: Entity
North America   4.449736
United States   4.575588
Estonia         4.639869
Eastern Europe   4.718923
Russia          5.240242
Name: DALYs (Disability-Adjusted Life Years) - Alcohol and substance use
disorders (Age-standardized percent), dtype: float64
```

```
In [12]: df3_mean = df3.groupby('Year').mean()  
df3_mean.head()
```

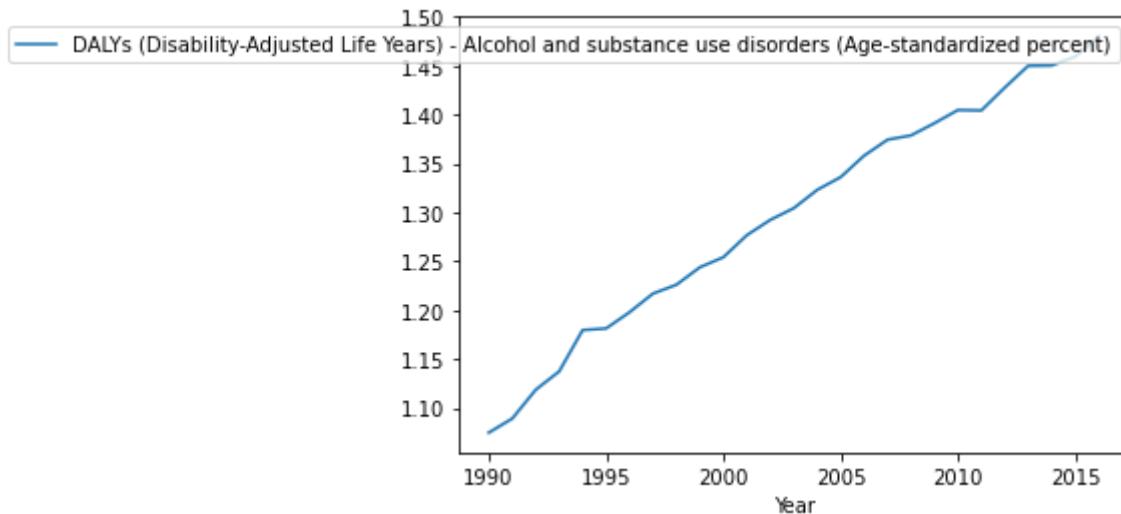
Out[12]:

DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)

Year	
1990	1.195516
1991	1.216156
1992	1.238201
1993	1.256913
1994	1.278350

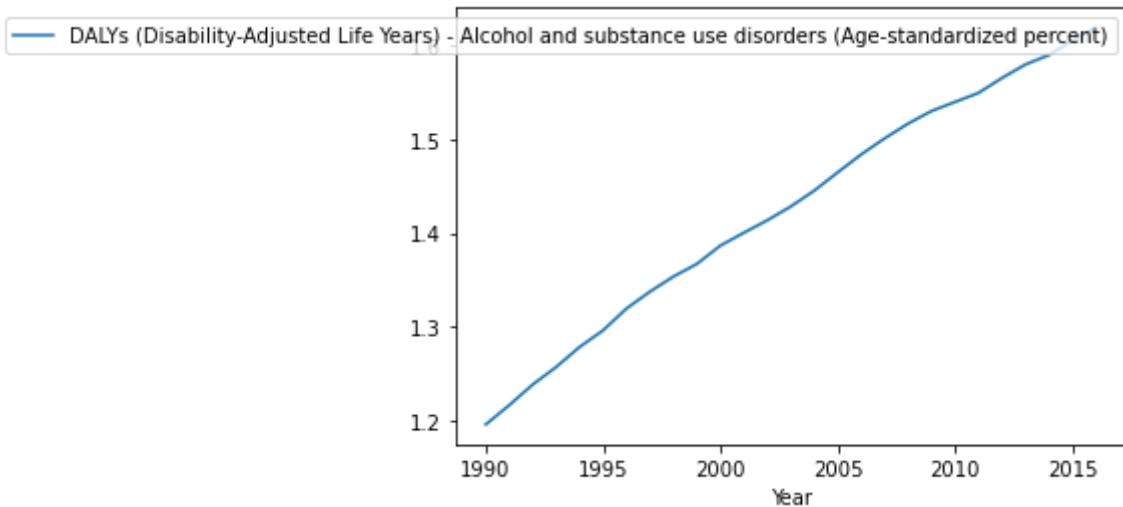
```
In [13]: df3.groupby('Year')[v1].median().plot(legend=True)
```

Out[13]: <AxesSubplot:xlabel='Year'>



```
In [14]: df3.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[14]: <AxesSubplot:xlabel='Year'>
```



```
In [15]: df3_country=df3.set_index('Year')
df3_country.groupby('Entity')[v1].plot(legend=True)
```

```
Out[15]: Entity
Afghanistan          AxesSubplot(0.125,0.125;0.775x0.755)
Albania              AxesSubplot(0.125,0.125;0.775x0.755)
Algeria              AxesSubplot(0.125,0.125;0.775x0.755)
American Samoa      AxesSubplot(0.125,0.125;0.775x0.755)
Andean Latin America AxesSubplot(0.125,0.125;0.775x0.755)
...
Western Sub-Saharan Africa AxesSubplot(0.125,0.125;0.775x0.755)
World                AxesSubplot(0.125,0.125;0.775x0.755)
Yemen                AxesSubplot(0.125,0.125;0.775x0.755)
Zambia               AxesSubplot(0.125,0.125;0.775x0.755)
Zimbabwe             AxesSubplot(0.125,0.125;0.775x0.755)
Name: DALYs (Disability-Adjusted Life Years) - Alcohol and substance use
disorders (Age-standardized percent), Length: 228, dtype: object
```

In [16]: df4.info()
df4.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6246 entries, 0 to 6245
Data columns (total 6 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6246 non-null   object
1   Code      object
5382 non-null   object
2   Year      int64
6246 non-null   int64
3   Prevalence - Alcohol and substance use disorders: Male (age-standardized percent) 6156 non-null   float64
4   Prevalence - Alcohol and substance use disorders: Female (age-standardized percent) 6156 non-null   float64
5   Continent object
dtypes: float64(2), int64(1), object(3)
memory usage: 292.9+ KB
```

Out[16]:

	Entity	Code	Year	Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)	Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	Asia
1	Afghanistan	AFG	1990	1.740499	1.314530	NaN
2	Afghanistan	AFG	1991	1.742372	1.310338	NaN
3	Afghanistan	AFG	1992	1.741842	1.307518	NaN
4	Afghanistan	AFG	1993	1.745838	1.305118	NaN

```
In [17]: df4.Entity.unique()
```

```
Out[17]: array(['Abkhazia', 'Afghanistan', 'Akrotiri and Dhekelia', 'Albania',  
   'Algeria', 'American Samoa', 'Andean Latin America', 'Andorra',  
   'Angola', 'Anguilla', 'Antarctica', 'Antigua and Barbuda',  
   'Argentina', 'Armenia', 'Aruba', 'Australasia', 'Australia',  
   'Austria', 'Austria-Hungary', 'Azerbaijan', 'Baden', 'Bahamas',  
   'Bahrain', 'Bangladesh', 'Barbados', 'Bavaria', 'Belarus',  
   'Belgium', 'Belize', 'Benin', 'Bermuda', 'Bhutan', 'Bolivia',  
   'Bonaire Sint Eustatius and Saba', 'Bosnia and Herzegovina',  
   'Botswana', 'Bouvet Island', 'Brazil',  
   'British Indian Ocean Territory', 'British Virgin Islands',  
   'Brunei', 'Bulgaria', 'Burkina Faso', 'Burundi', 'Cambodia',  
   'Cameroon', 'Canada', 'Cape Verde', 'Caribbean', 'Cayman Islands',  
   'Central African Republic', 'Central Asia', 'Central Europe',  
   'Central Latin America', 'Central Sub-Saharan Africa', 'Chad',  
   'Channel Islands', 'Chile', 'China', 'Christmas Island',  
   'Cocos Islands', 'Colombia', 'Comoros', 'Congo', 'Cook Islands',  
   'Costa Rica', "Cote d'Ivoire", 'Croatia', 'Cuba', 'Curacao',  
   'Cyprus', 'Czechia', 'Czechoslovakia',  
   'Democratic Republic of Congo', 'Denmark', 'Djibouti', 'Dominica',  
   'Dominican Republic', 'East Asia', 'East Germany',  
   'Eastern Europe', 'Eastern Sub-Saharan Africa', 'Ecuador', 'Egypt',  
   'El Salvador', 'England', 'Equatorial Guinea', 'Eritrea',  
   'Eritrea and Ethiopia', 'Estonia', 'Eswatini', 'Ethiopia',  
   'Faeroe Islands', 'Falkland Islands', 'Fiji', 'Finland', 'France',  
   'French Guiana', 'French Polynesia', 'French Southern Territories',  
   'Gabon', 'Gambia', 'Georgia', 'Germany', 'Ghana', 'Gibraltar',  
   'Greece', 'Greenland', 'Grenada', 'Guadeloupe', 'Guam',  
   'Guatemala', 'Guernsey', 'Guinea', 'Guinea-Bissau', 'Guyana',  
   'Haiti', 'Hanover', 'Heard Island and McDonald Islands',  
   'Hesse Electoral', 'Hesse Grand Ducal', 'High SDI',  
   'High-income Asia Pacific', 'High-middle SDI', 'Honduras',  
   'Hong Kong', 'Hungary', 'Iceland', 'India', 'Indonesia', 'Iran',  
   'Iraq', 'Ireland', 'Isle of Man', 'Israel', 'Italy', 'Jamaica',  
   'Japan', 'Jersey', 'Jordan', 'Kazakhstan', 'Kenya', 'Kiribati',  
   'Kosovo', 'Kuwait', 'Kyrgyzstan', 'Laos',  
   'Latin America and Caribbean', 'Latvia', 'Lebanon', 'Lesotho',  
   'Liberia', 'Libya', 'Liechtenstein', 'Lithuania', 'Low SDI',  
   'Low-middle SDI', 'Luxembourg', 'Macao', 'Madagascar', 'Malawi',  
   'Malaysia', 'Maldives', 'Mali', 'Malta', 'Marshall Islands',  
   'Martinique', 'Mauritania', 'Mauritius', 'Mayotte',  
   'Mecklenburg Schwerin', 'Mexico', 'Micronesia (country)',  
   'Middle SDI', 'Modena', 'Moldova', 'Monaco', 'Mongolia',  
   'Montenegro', 'Montserrat', 'Morocco', 'Mozambique', 'Myanmar',  
   'Nagorno-Karabakh', 'Namibia', 'Nauru', 'Nepal', 'Netherlands',  
   'Netherlands Antilles', 'New Caledonia', 'New Zealand',  
   'Nicaragua', 'Niger', 'Nigeria', 'Niue', 'Norfolk Island',  
   'North Africa and Middle East', 'North America', 'North Korea',  
   'North Macedonia', 'Northern Cyprus', 'Northern Ireland',  
   'Northern Mariana Islands', 'Norway', 'Oceania', 'Oman',  
   'Pakistan', 'Palau', 'Palestine', 'Panama', 'Papua New Guinea',  
   'Paraguay', 'Parma', 'Peru', 'Philippines', 'Pitcairn', 'Poland',  
   'Portugal', 'Puerto Rico', 'Qatar', 'Republic of Vietnam',  
   'Reunion', 'Romania', 'Russia', 'Rwanda', 'Saint Barthélemy',
```

```
'Saint Helena', 'Saint Kitts and Nevis', 'Saint Lucia',
'Saint Martin (French part)', 'Saint Pierre and Miquelon',
'Saint Vincent and the Grenadines', 'Samoa', 'San Marino',
'Sao Tome and Principe', 'Saudi Arabia', 'Saxony', 'Scotland',
'Senegal', 'Serbia', 'Serbia and Montenegro',
'Serbia excluding Kosovo', 'Seychelles', 'Sierra Leone',
'Singapore', 'Sint Maarten (Dutch part)', 'Slovakia', 'Slovenia',
'Solomon Islands', 'Somalia', 'Somaliland', 'South Africa',
'South Asia', 'South Georgia and the South Sandwich Islands',
'South Korea', 'South Ossetia', 'South Sudan', 'Southeast Asia',
'Southern Latin America', 'Southern Sub-Saharan Africa', 'Spain',
'Sri Lanka', 'Sub-Saharan Africa', 'Sudan', 'Suriname',
'Svalbard and Jan Mayen', 'Sweden', 'Switzerland', 'Syria',
'Taiwan', 'Tajikistan', 'Tanzania', 'Thailand', 'Timor', 'Togo',
'Tokelau', 'Tonga', 'Transnistria', 'Trinidad and Tobago',
'Tropical Latin America', 'Tunisia', 'Turkey', 'Turkmenistan',
'Turks and Caicos Islands', 'Tuscany', 'Tuvalu', 'Two Sicilies',
'USSR', 'Uganda', 'Ukraine', 'United Arab Emirates',
'United Kingdom', 'United Korea', 'United States',
'United States Minor Outlying Islands',
'United States Virgin Islands', 'Uruguay', 'Uzbekistan', 'Vanuatu',
'Vatican', 'Venezuela', 'Vietnam', 'Wales', 'Wallis and Futuna',
'West Germany', 'Western Europe', 'Western Sahara',
'Western Sub-Saharan Africa', 'World', 'Wuerttemburg', 'Yemen',
'Yemen Arab Republic', "Yemen People's Republic", 'Yugoslavia',
'Zambia', 'Zanzibar', 'Zimbabwe', 'Åland Islands'], dtype=object)
```

Checking for missing values:

```
In [18]: missing = pd.concat([df4.isnull().sum(), 100 * df4.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[18]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)		90	1.440922
Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)		90	1.440922
	Code	864	13.832853
	Continent	5961	95.437080

In [19]: df4.describe()

Out[19]:

	Year	Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)	Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)
count	6246.000000	6156.000000	6156.000000
mean	2003.172911	3.121449	1.532282
std	7.864314	1.260856	0.709345
min	1990.000000	1.480319	0.806518
25%	1996.000000	2.102214	1.144971
50%	2003.000000	2.843404	1.375345
75%	2010.000000	3.721584	1.559393
max	2016.000000	9.402884	5.071643

In [20]: v1='Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)'

In [21]: df4.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[21]:

	Entity	Year	Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)
4549	Russia	2000	9.402884
4550	Russia	2001	9.384413
4551	Russia	2002	9.367462
4548	Russia	1999	9.356063
4552	Russia	2003	9.351228

In [22]: df4.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[22]:

	Entity	Year	Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)
2604	Indonesia	1995	1.480319
2603	Indonesia	1994	1.481526
2605	Indonesia	1996	1.482580
2602	Indonesia	1993	1.483252
2600	Indonesia	1991	1.483527

In [23]: v2='Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)'

```
In [24]: df4.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()
```

Out[24]:

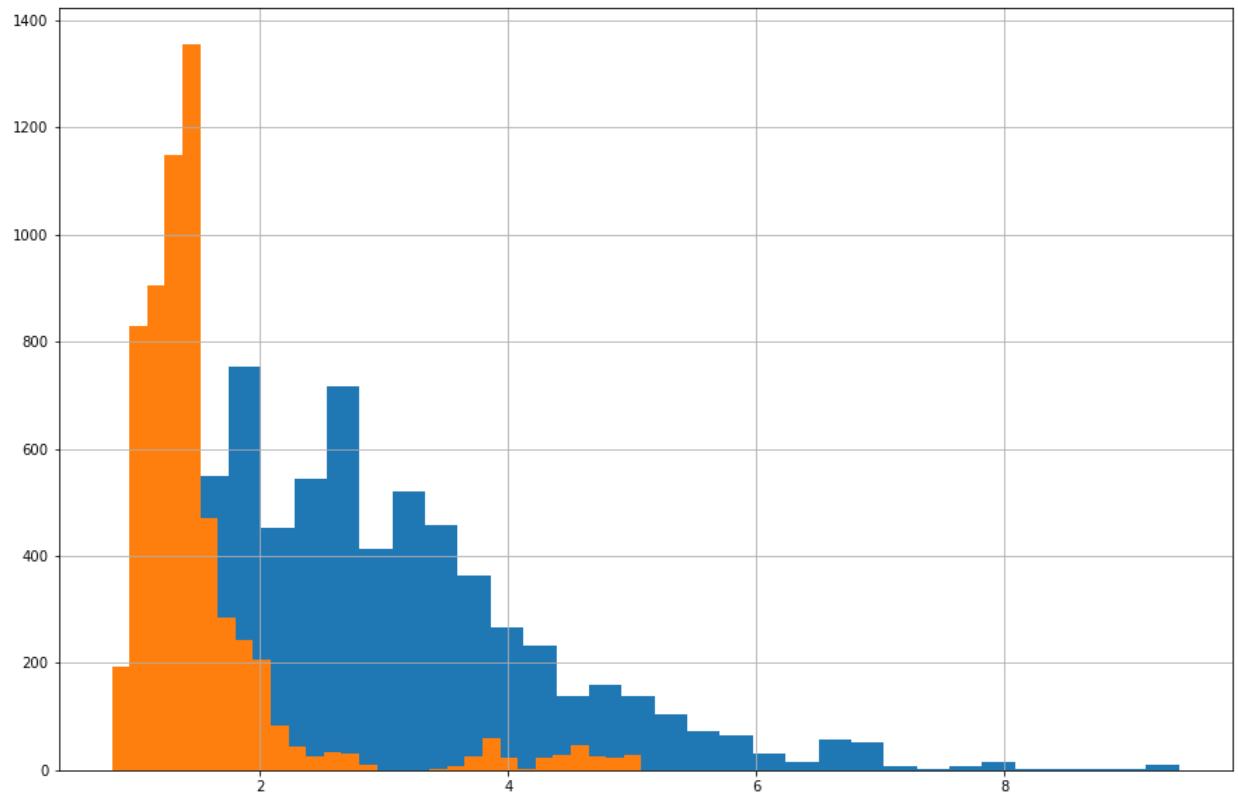
	Entity	Year	Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)
5801	Ukraine	2016	5.071643
3170	Lithuania	1990	5.057323
5800	Ukraine	2015	5.056950
5799	Ukraine	2014	5.041881
1857	Estonia	1990	5.036433

```
In [25]: df4.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()
```

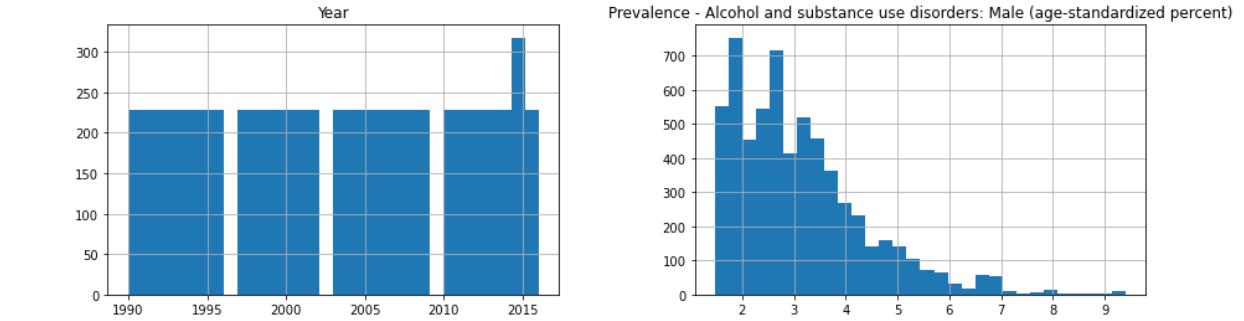
Out[25]:

	Entity	Year	Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)
2590	India	2008	0.806518
2591	India	2009	0.806539
2592	India	2010	0.806577
2589	India	2007	0.806785
2588	India	2006	0.807771

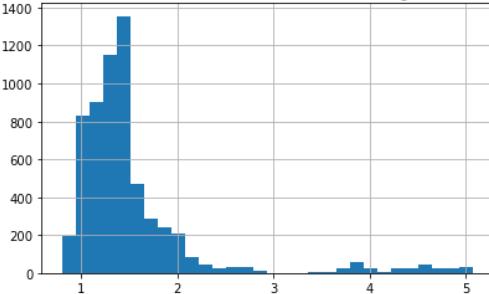
```
In [26]: df4[v1].hist(bins=30, figsize=(15,10))
df4[v2].hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [27]: df4.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)



```
In [28]: df4.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[28]: Entity
Indonesia      1.520101
Morocco        1.570231
Saudi Arabia   1.575613
Syria          1.610855
Yemen          1.620286
Name: Prevalence - Alcohol and substance use disorders: Male (age-standardized percent), dtype: float64
```

```
In [29]: df4.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[29]: Entity
Yemen Arab Republic      NaN
Yemen People's Republic  NaN
Yugoslavia              NaN
Zanzibar                 NaN
Åland Islands            NaN
Name: Prevalence - Alcohol and substance use disorders: Male (age-standardized percent), dtype: float64
```

```
In [30]: df4.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[30]: Entity
India          0.850959
Malaysia       0.855355
Israel          0.883177
Iran            0.886708
Indonesia       0.938945
Name: Prevalence - Alcohol and substance use disorders: Female (age-standardized percent), dtype: float64
```

```
In [31]: df4.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[31]: Entity
Yemen Arab Republic      NaN
Yemen People's Republic  NaN
Yugoslavia              NaN
Zanzibar                 NaN
Åland Islands            NaN
Name: Prevalence - Alcohol and substance use disorders: Female (age-standardized percent), dtype: float64
```

```
In [32]: df4_mean = df4.groupby('Year').mean()  
df4_mean.head()
```

Out[32]:

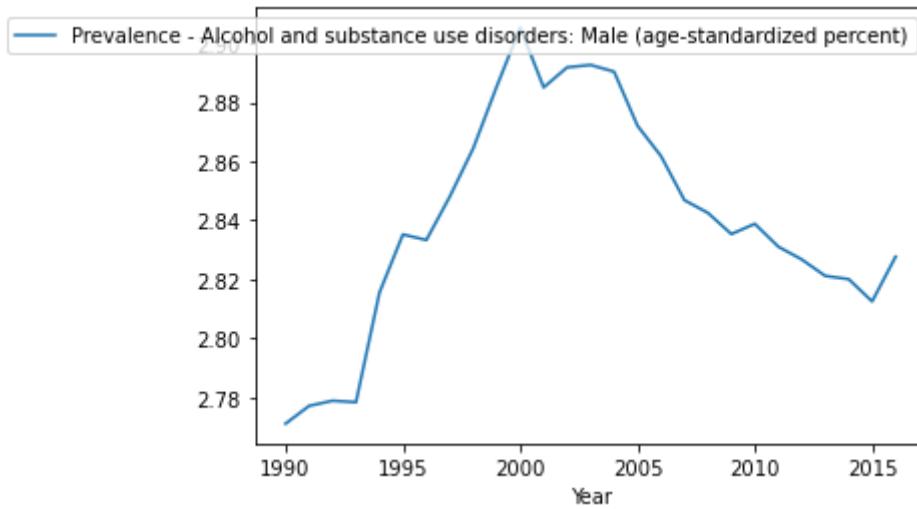
Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)

Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)

Year	Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)	Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)
1990	3.072778	1.526307
1991	3.081035	1.525942
1992	3.089954	1.526076
1993	3.099274	1.526165
1994	3.108663	1.526574

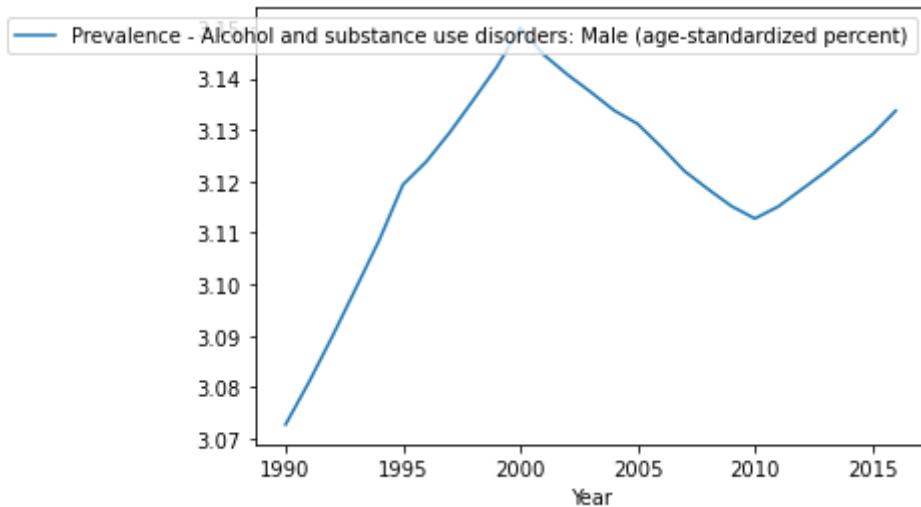
```
In [33]: df4.groupby('Year')[v1].median().plot(legend=True)
```

Out[33]: <AxesSubplot:xlabel='Year'>



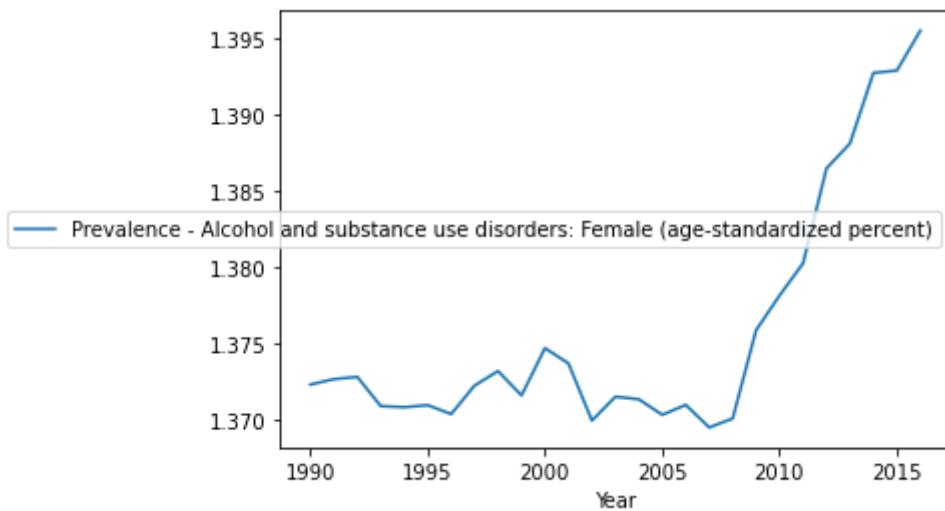
```
In [34]: df4.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[34]: <AxesSubplot:xlabel='Year'>
```



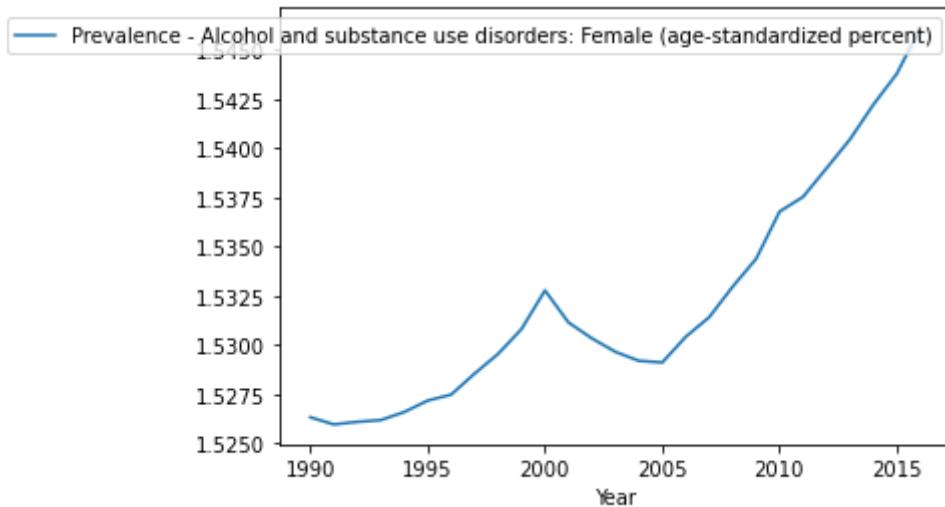
```
In [35]: df4.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[35]: <AxesSubplot:xlabel='Year'>
```



```
In [36]: df4.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[36]: <AxesSubplot:xlabel='Year'>
```



```
In [37]: df4_country=df4.set_index('Year')
df4_country.groupby('Entity')[v1].plot(legend=True)
```

```
Out[37]: Entity
Abkhazia           AxesSubplot(0.125,0.125;0.775x0.755)
Afghanistan        AxesSubplot(0.125,0.125;0.775x0.755)
Akrotiri and Dhekelia AxesSubplot(0.125,0.125;0.775x0.755)
Albania            AxesSubplot(0.125,0.125;0.775x0.755)
Algeria             AxesSubplot(0.125,0.125;0.775x0.755)
...
Yugoslavia         AxesSubplot(0.125,0.125;0.775x0.755)
Zambia              AxesSubplot(0.125,0.125;0.775x0.755)
Zanzibar            AxesSubplot(0.125,0.125;0.775x0.755)
Zimbabwe            AxesSubplot(0.125,0.125;0.775x0.755)
Åland Islands       AxesSubplot(0.125,0.125;0.775x0.755)
Name: Prevalence - Alcohol and substance use disorders: Male (age-standardized percent), Length: 318, dtype: object
```

```
In [38]: df4_country=df4.set_index('Year')
df4_country.groupby('Entity')[v2].plot(legend=True)
```

```
Out[38]: Entity
Abkhazia           AxesSubplot(0.125,0.125;0.775x0.755)
Afghanistan        AxesSubplot(0.125,0.125;0.775x0.755)
Akrotiri and Dhekelia AxesSubplot(0.125,0.125;0.775x0.755)
Albania            AxesSubplot(0.125,0.125;0.775x0.755)
Algeria             AxesSubplot(0.125,0.125;0.775x0.755)
...
Yugoslavia         AxesSubplot(0.125,0.125;0.775x0.755)
Zambia              AxesSubplot(0.125,0.125;0.775x0.755)
Zanzibar            AxesSubplot(0.125,0.125;0.775x0.755)
Zimbabwe            AxesSubplot(0.125,0.125;0.775x0.755)
Åland Islands       AxesSubplot(0.125,0.125;0.775x0.755)
Name: Prevalence - Alcohol and substance use disorders: Female (age-standardized percent), Length: 318, dtype: object
```

In [39]: df71.info()
df71.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 14 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year        int64
6468 non-null   int64
3   Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)
6468 non-null   float64
4   Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)
6468 non-null   float64
5   Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
6   Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)
6468 non-null   float64
7   Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)
6468 non-null   float64
8   Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)
6468 non-null   float64
9   Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)
6468 non-null   float64
10  Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)
6468 non-null   float64
11  Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)
6468 non-null   float64
12  Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)
6468 non-null   float64
13  Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)
6468 non-null   float64
dtypes: float64(11), int64(1), object(2)
memory usage: 707.6+ KB
```

Out[39]:

	Entity	Code	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)
				Entity	Code	Year	Entity	Code	Year	Entity	Code	Year	Entity	Code
0	Afghanistan	AFG	1990	0.053520	0.498033	0.501266	1.358526	1.354713	1.201890					
1	Afghanistan	AFG	1991	0.053620	0.499952	0.507430	1.361619	1.350155	1.202776					
2	Afghanistan	AFG	1992	0.053734	0.503599	0.528792	1.368903	1.344693	1.202468					
3	Afghanistan	AFG	1993	0.053822	0.506347	0.550169	1.372687	1.340029	1.203614					

Entity	Code	Year	Prevalence						
			- Alcohol use						
4	Afghanistan	AFG	1994	0.053874	0.506726	0.553151	1.370159	1.336986	1.206901

Checking for missing values:

```
In [40]: missing = pd.concat([df71.isnull().sum(), 100 * df71.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[40]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Code	980	15.151515	

```
In [41]: df71.columns
```

```
Out[41]: Index(['Entity', 'Code', 'Year',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)',
       'Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)'],
      dtype='object')
```

```
In [42]: v1='Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent'
v2='Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent'
v3='Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent'
v4='Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent'
v5='Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent'
v6='Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent'
v7='Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent'
v8='Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent'
v9='Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent'
v10='Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent'
v11='Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized'
```

In [43]: df71.describe()

Out[43]:

	Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.105648	1.063202	1.565993	3.005899	2.884627
std	8.078372	0.040766	0.473978	0.979419	1.486558	1.548540
min	1990.000000	0.043191	0.378933	0.434616	0.924971	0.756776
25%	1996.750000	0.070354	0.632190	0.916527	1.907675	1.905823
50%	2003.500000	0.104129	1.118708	1.356565	2.949014	2.742245
75%	2010.250000	0.123966	1.308383	1.909182	3.452614	3.310244
						3.191792

In [44]: df71.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[44]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)
504	Belarus	1990	0.256646
505	Belarus	1991	0.256030
506	Belarus	1992	0.255471
507	Belarus	1993	0.254996
508	Belarus	1994	0.254585

In [45]: df71.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[45]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)
5992	Ukraine	1990	2.966194
5993	Ukraine	1991	2.962748
5994	Ukraine	1992	2.955355
5995	Ukraine	1993	2.945152
4925	Scotland	2015	2.937054

In [46]: `df71.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()`

Out[46]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)
527	Belarus	2013	6.476958
526	Belarus	2012	6.471538
528	Belarus	2014	6.471127
525	Belarus	2011	6.458989
529	Belarus	2015	6.452227

In [47]: `df71.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()`

Out[47]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)
4714	Russia	2000	9.623866
4715	Russia	2001	9.607836
4716	Russia	2002	9.556062
4713	Russia	1999	9.551820
4717	Russia	2003	9.482203

In [48]: `df71.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()`

Out[48]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)
4715	Russia	2001	10.915594
4714	Russia	2000	10.912681
4716	Russia	2002	10.864503
4713	Russia	1999	10.803305
4717	Russia	2003	10.789822

In [49]: `df71.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()`

Out[49]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)
4927	Scotland	2017	8.203841
4926	Scotland	2016	8.196813
4925	Scotland	2015	8.191610
4924	Scotland	2014	8.185437
4923	Scotland	2013	8.179755

In [50]: df71.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[50]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)
531	Belarus	2017	4.183879
530	Belarus	2016	4.050404
529	Belarus	2015	3.941700
528	Belarus	2014	3.843992
527	Belarus	2013	3.760376

In [51]: df71.sort_values(by=v8, ascending=False)[['Entity', 'Year', v8]].head()

Out[51]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)
4714	Russia	2000	8.207603
4713	Russia	1999	8.197247
4715	Russia	2001	8.122494
4712	Russia	1998	8.109251
4725	Russia	2011	8.039942

In [52]: df71.sort_values(by=v9, ascending=False)[['Entity', 'Year', v9]].head()

Out[52]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)
515	Belarus	2001	0.152865
514	Belarus	2000	0.151972
516	Belarus	2002	0.151812
513	Belarus	1999	0.149663
517	Belarus	2003	0.149401

In [53]: df71.sort_values(by=v10, ascending=False)[['Entity', 'Year', v10]].head()

Out[53]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)
527	Belarus	2013	8.141410
526	Belarus	2012	8.135358
528	Belarus	2014	8.130052
529	Belarus	2015	8.112764
525	Belarus	2011	8.105998

In [54]: df71.sort_values(by=v11, ascending=False)[['Entity', 'Year', v11]].head()

Out[54]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)
524	Belarus	2010	5.474668
525	Belarus	2011	5.473536
526	Belarus	2012	5.467508
523	Belarus	2009	5.461357
527	Belarus	2013	5.455420

In [55]: df71.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[55]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)
2718	Indonesia	1992	0.043191
2717	Indonesia	1991	0.043198
2719	Indonesia	1993	0.043219
2716	Indonesia	1990	0.043250
2720	Indonesia	1994	0.043279

In [56]: df71.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[56]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)
2721	Indonesia	1995	0.378933
2722	Indonesia	1996	0.379347
2720	Indonesia	1994	0.379409
2723	Indonesia	1997	0.380256
2719	Indonesia	1993	0.380874

In [57]: df71.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[57]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)
2883	Italy	2017	0.434616
2882	Italy	2016	0.438455
2881	Italy	2015	0.441809
2880	Italy	2014	0.445086
2879	Italy	2013	0.448108

In [58]: df71.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[58]:

Entity Year Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)			
2861	Italy	1995	0.924971
2862	Italy	1996	0.925115
2863	Italy	1997	0.926129
2860	Italy	1994	0.926974
2864	Italy	1998	0.927679

In [59]: df71.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[59]:

Entity Year Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)			
2862	Italy	1996	0.756776
2870	Italy	2004	0.757052
2861	Italy	1995	0.757087
2863	Italy	1997	0.757173
2869	Italy	2003	0.757260

In [60]: df71.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[60]:

Entity Year Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)			
2721	Indonesia	1995	0.919192
2722	Indonesia	1996	0.919496
2723	Indonesia	1997	0.920945
2720	Indonesia	1994	0.921646
2724	Indonesia	1998	0.922991

In [61]: df71.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[61]:

Entity Year Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)			
5040	Singapore	1990	0.135407
5056	Singapore	2006	0.135409
5055	Singapore	2005	0.135419
5057	Singapore	2007	0.135436
5058	Singapore	2008	0.135597

In [62]: `df71.sort_values(by=v8, ascending=True)[['Entity', 'Year', v8]].head()`

Out[62]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)
2883	Italy	2017	0.699540
2871	Italy	2005	0.700668
2872	Italy	2006	0.700683
2882	Italy	2016	0.702411
2870	Italy	2004	0.702996

In [63]: `df71.sort_values(by=v9, ascending=True)[['Entity', 'Year', v9]].head()`

Out[63]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)
2744	Iran	1990	0.020697
10	Afghanistan	2000	0.021082
11	Afghanistan	2001	0.021101
2745	Iran	1991	0.021295
9	Afghanistan	1999	0.021366

In [64]: `df71.sort_values(by=v10, ascending=True)[['Entity', 'Year', v10]].head()`

Out[64]:

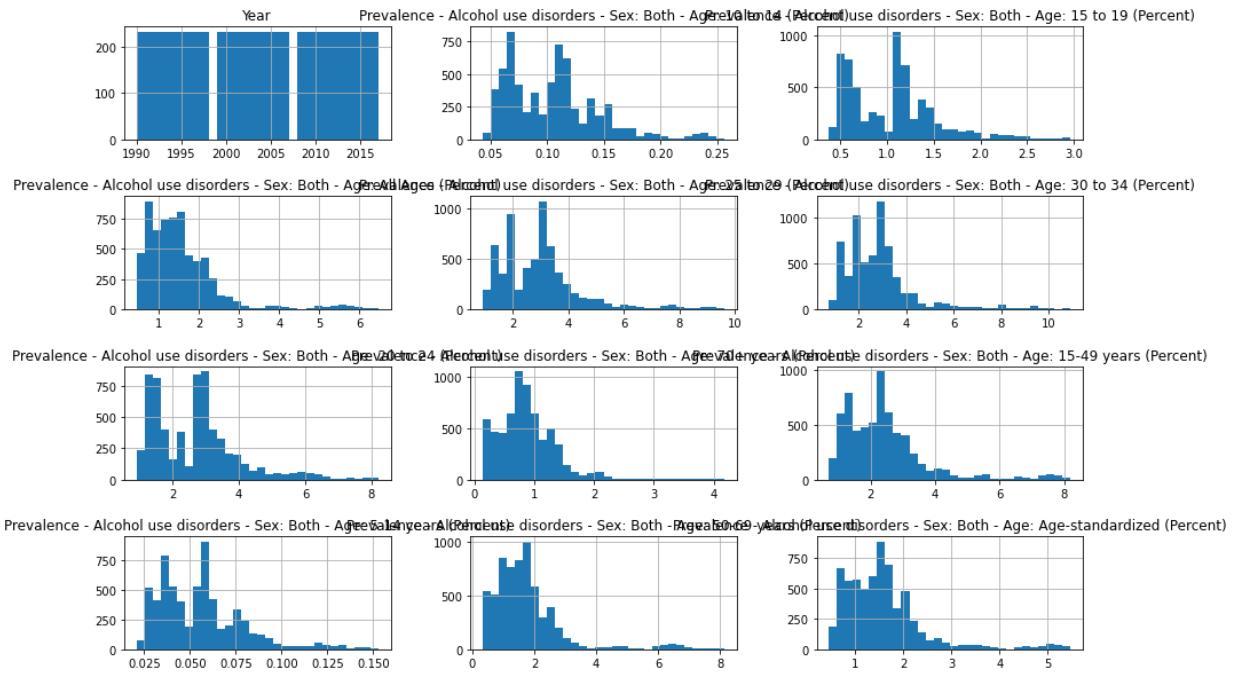
	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)
2870	Italy	2004	0.332686
2871	Italy	2005	0.332726
2862	Italy	1996	0.332973
2861	Italy	1995	0.333403
2869	Italy	2003	0.333704

In [65]: `df71.sort_values(by=v11, ascending=True)[['Entity', 'Year', v11]].head()`

Out[65]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)
2861	Italy	1995	0.446940
2862	Italy	1996	0.447098
2860	Italy	1994	0.447662
2863	Italy	1997	0.447723
2864	Italy	1998	0.448558

```
In [66]: df71.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [67]: df71.groupby('Entity')[v11].mean().sort_values().head()
```

```
Out[67]: Entity
Italy          0.456138
Singapore      0.479829
Israel         0.533354
Morocco        0.563468
Japan          0.565542
Name: Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [68]: df71.groupby('Entity')[v11].mean().sort_values().tail()
```

```
Out[68]: Entity
Estonia        4.917451
Ukraine        5.059726
Eastern Europe  5.106033
Russia          5.158491
Belarus         5.286263
Name: Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

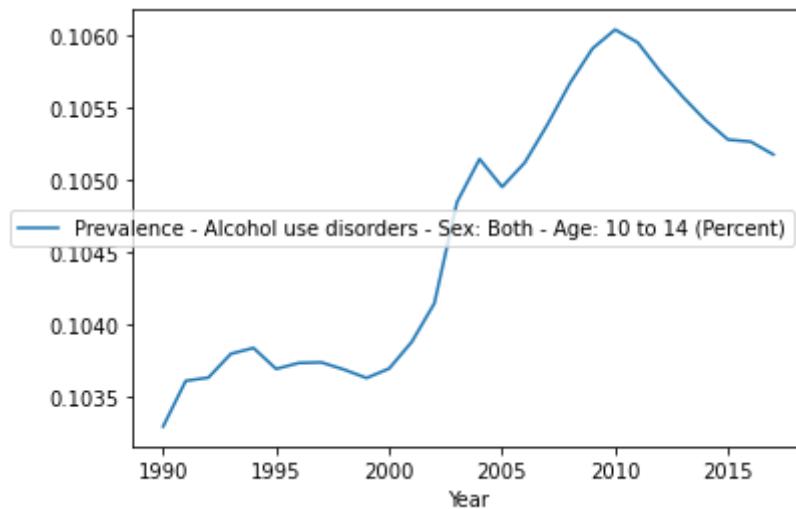
```
In [69]: df71_mean = df71.groupby('Year').mean()
df71_mean.head()
```

Out[69]:

	Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 15 (Percent)
Year								
1990	0.103682	1.046755	1.456609	2.955716	2.837985	2.642899	0.831115	2.3494
1991	0.103985	1.049101	1.467102	2.959101	2.843368	2.649025	0.837009	2.3581
1992	0.104270	1.051481	1.476878	2.962774	2.848450	2.654952	0.843125	2.3651
1993	0.104516	1.053740	1.486171	2.967044	2.852976	2.659651	0.848701	2.3721
1994	0.104709	1.055523	1.494500	2.971078	2.856356	2.662337	0.853121	2.3774

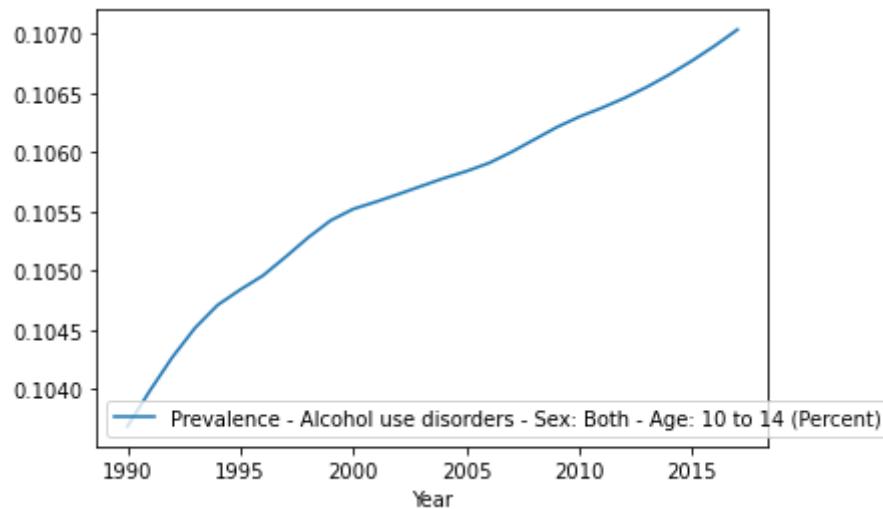
```
In [70]: df71.groupby('Year')[v1].median().plot(legend=True)
```

Out[70]: <AxesSubplot:xlabel='Year'>



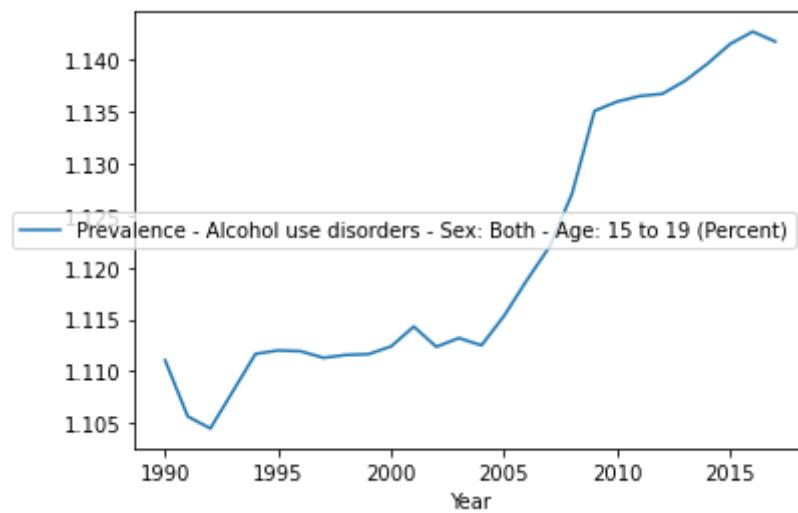
```
In [71]: df71.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[71]: <AxesSubplot:xlabel='Year'>
```



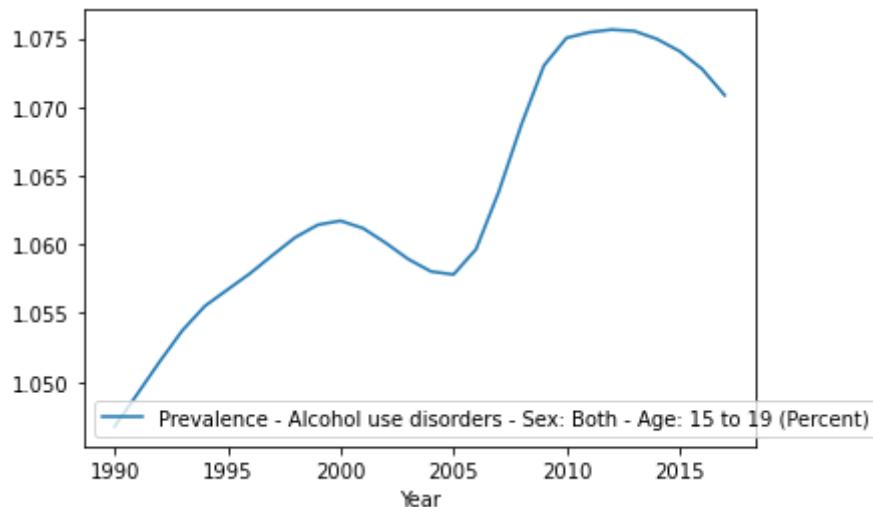
```
In [72]: df71.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[72]: <AxesSubplot:xlabel='Year'>
```



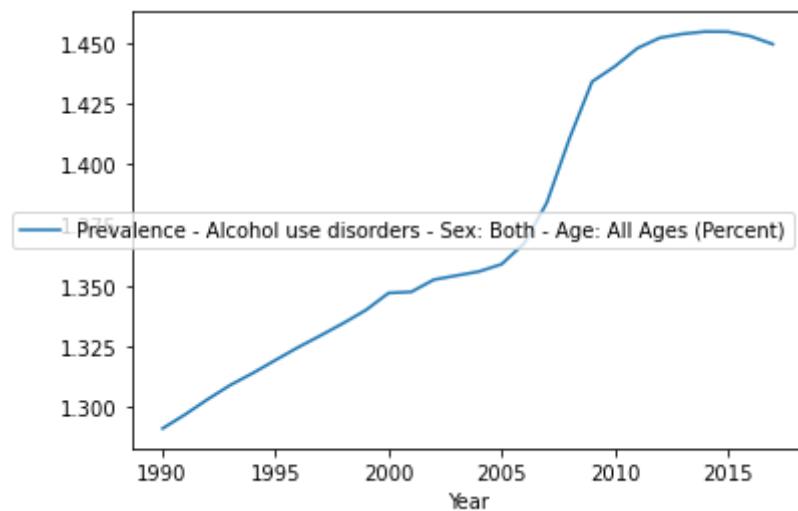
```
In [73]: df71.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[73]: <AxesSubplot:xlabel='Year'>
```



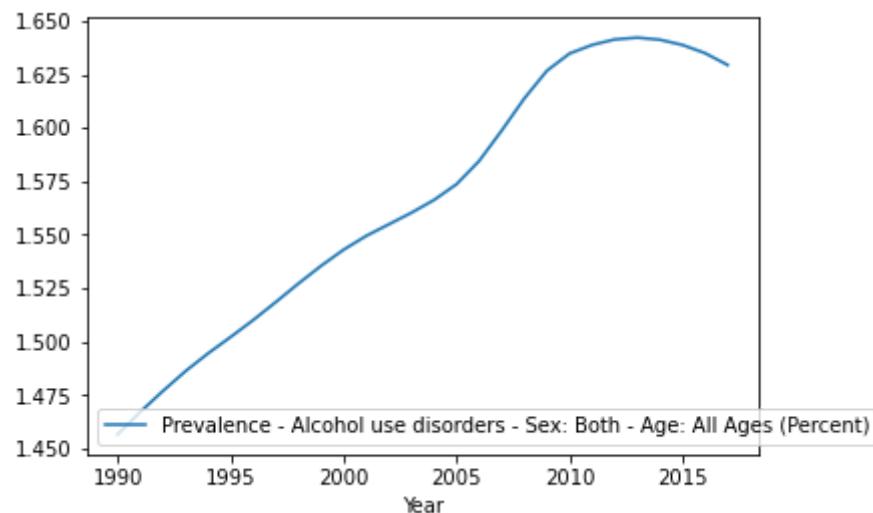
```
In [74]: df71.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[74]: <AxesSubplot:xlabel='Year'>
```



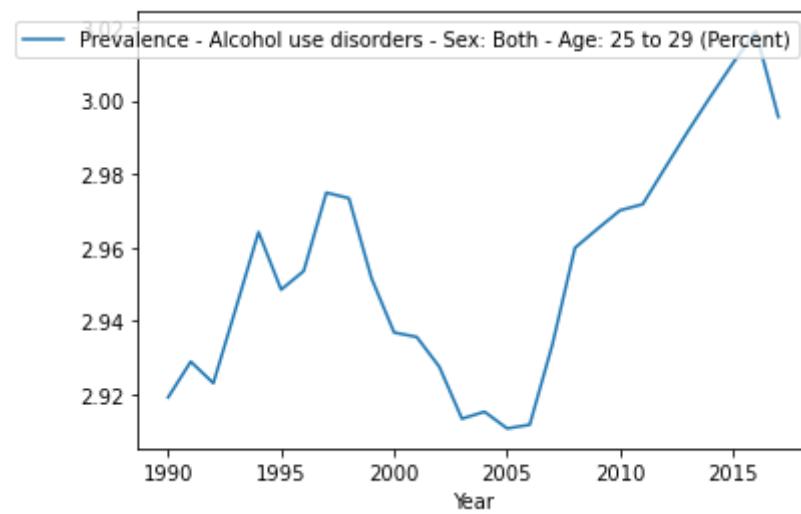
```
In [75]: df71.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[75]: <AxesSubplot:xlabel='Year'>
```



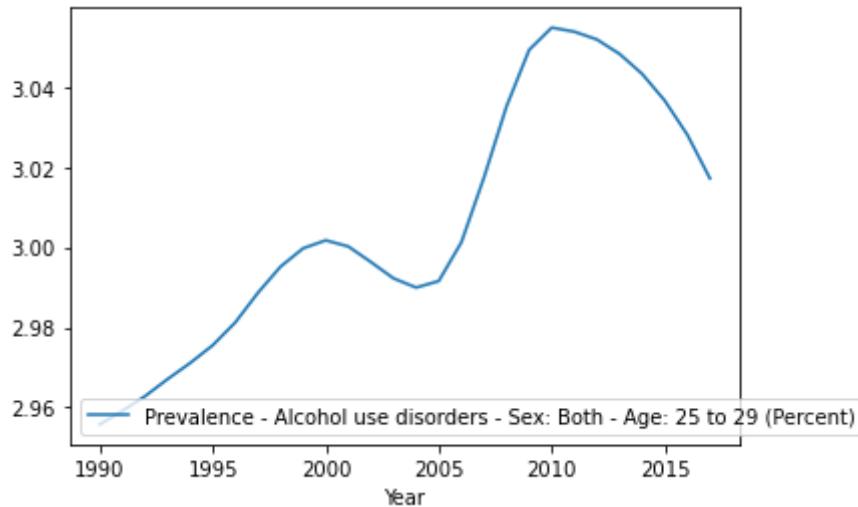
```
In [76]: df71.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[76]: <AxesSubplot:xlabel='Year'>
```



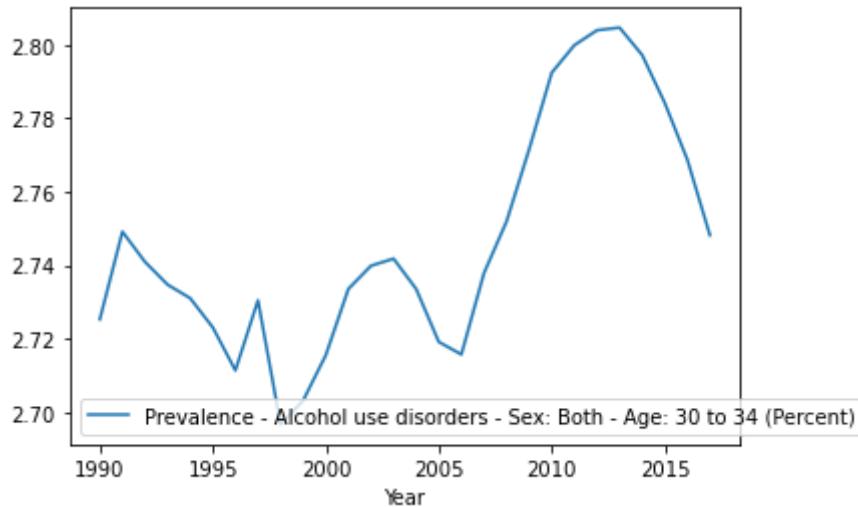
```
In [77]: df71.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[77]: <AxesSubplot:xlabel='Year'>
```



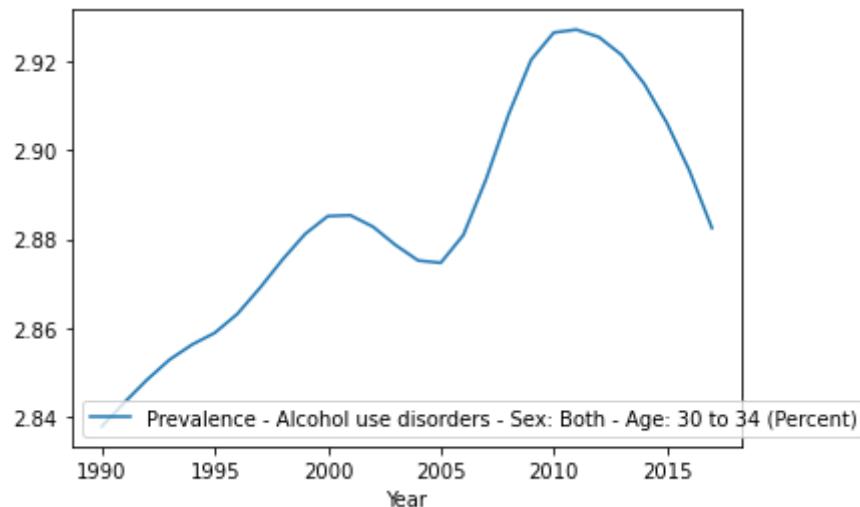
```
In [78]: df71.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[78]: <AxesSubplot:xlabel='Year'>
```



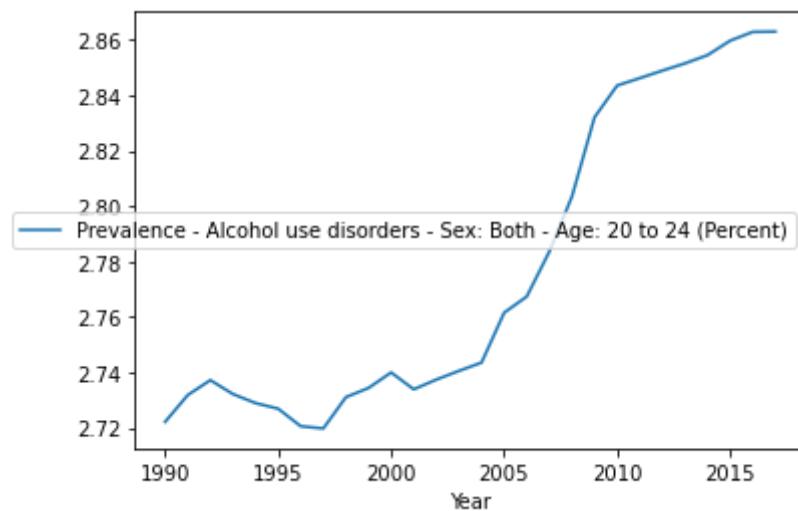
```
In [79]: df71.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[79]: <AxesSubplot:xlabel='Year'>
```



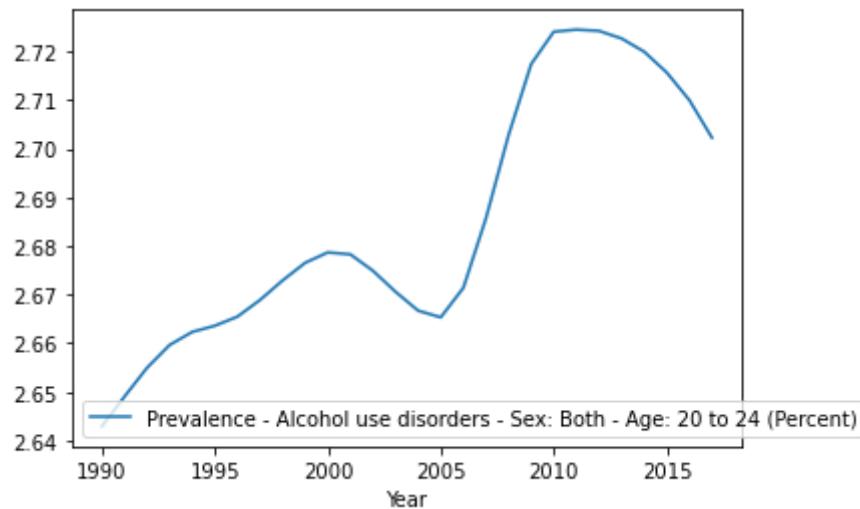
```
In [80]: df71.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[80]: <AxesSubplot:xlabel='Year'>
```



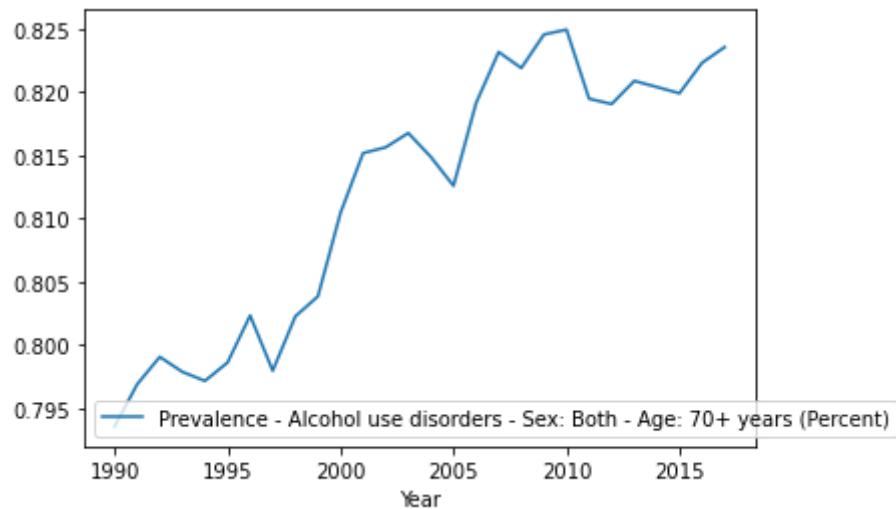
```
In [81]: df71.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[81]: <AxesSubplot:xlabel='Year'>
```



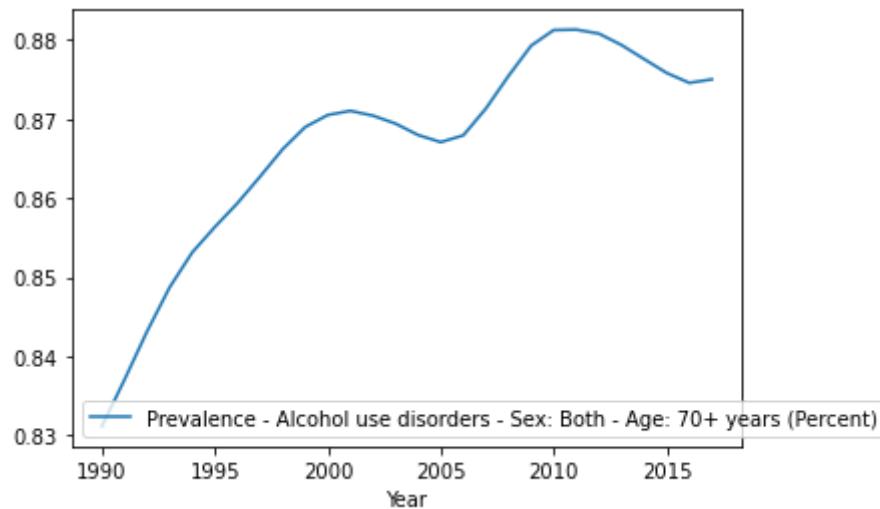
```
In [82]: df71.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[82]: <AxesSubplot:xlabel='Year'>
```



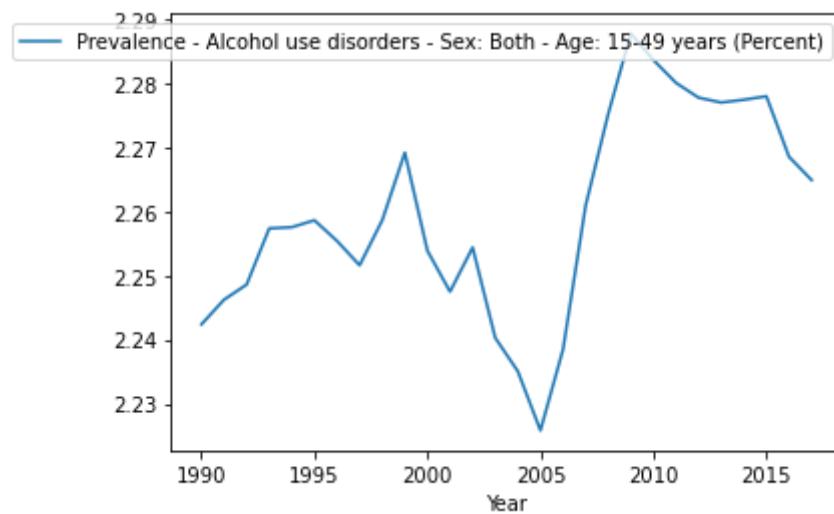
```
In [83]: df71.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[83]: <AxesSubplot:xlabel='Year'>
```



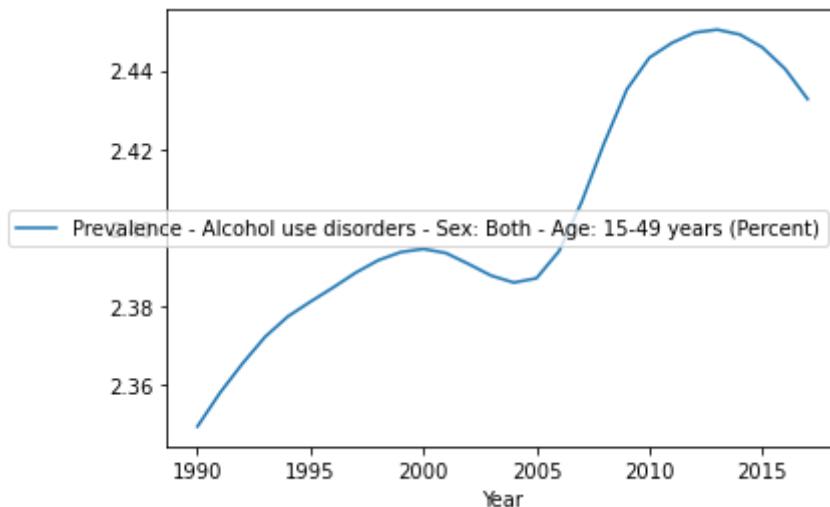
```
In [84]: df71.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[84]: <AxesSubplot:xlabel='Year'>
```



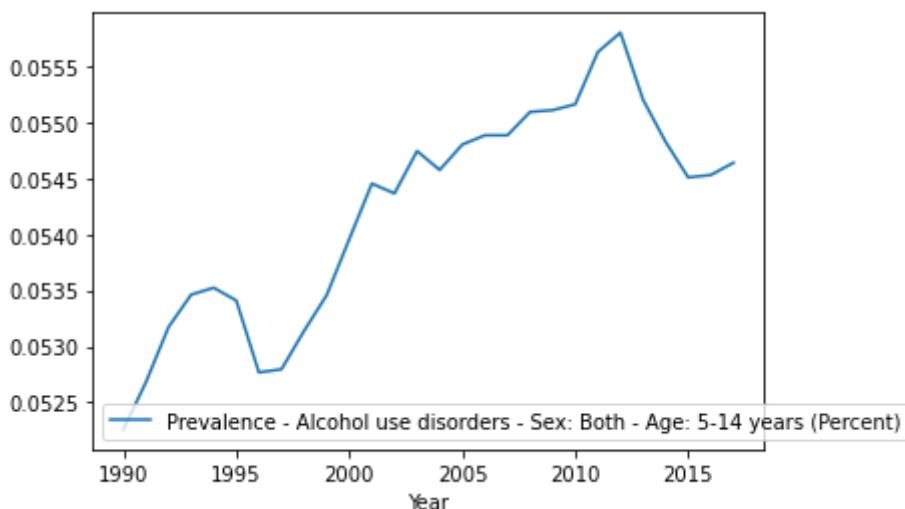
```
In [85]: df71.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[85]: <AxesSubplot:xlabel='Year'>
```



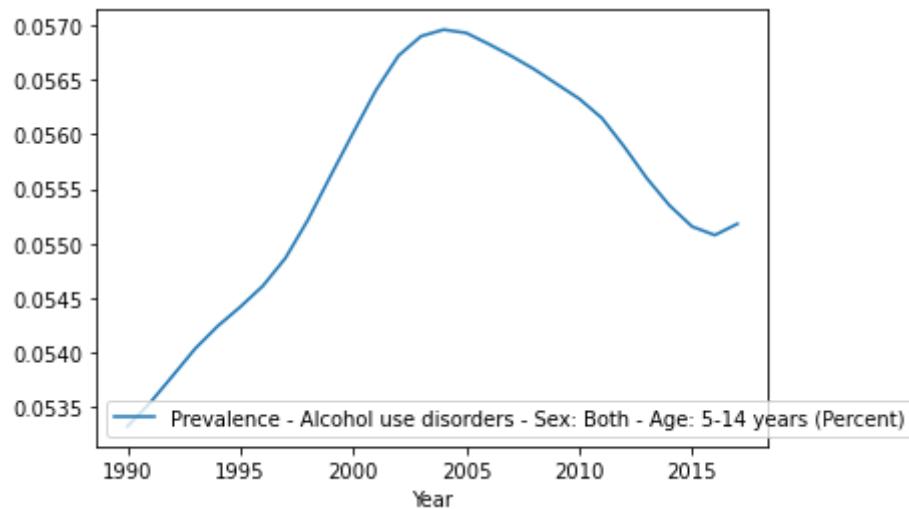
```
In [86]: df71.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[86]: <AxesSubplot:xlabel='Year'>
```



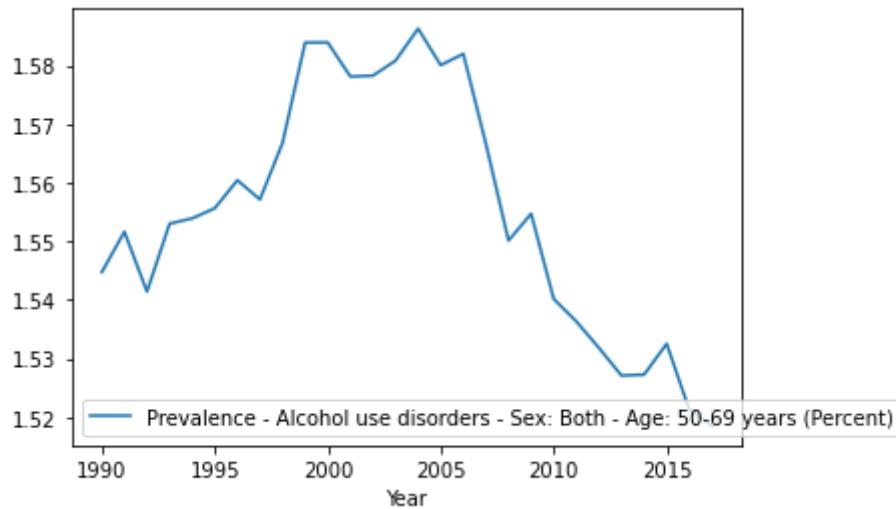
```
In [87]: df71.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[87]: <AxesSubplot:xlabel='Year'>
```



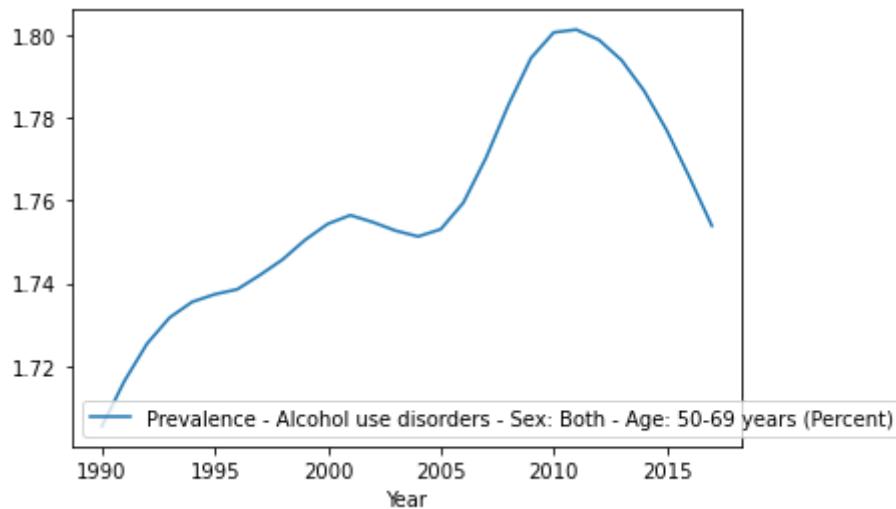
```
In [88]: df71.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[88]: <AxesSubplot:xlabel='Year'>
```



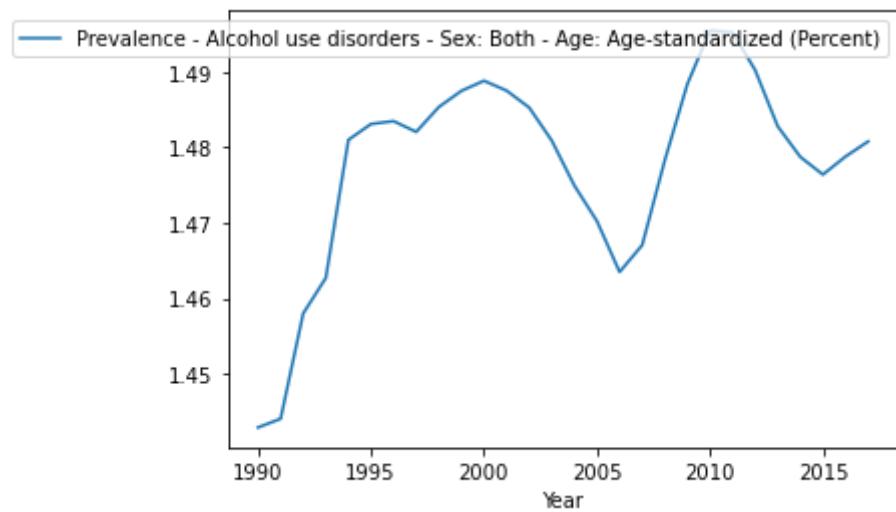
```
In [89]: df71.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[89]: <AxesSubplot:xlabel='Year'>
```



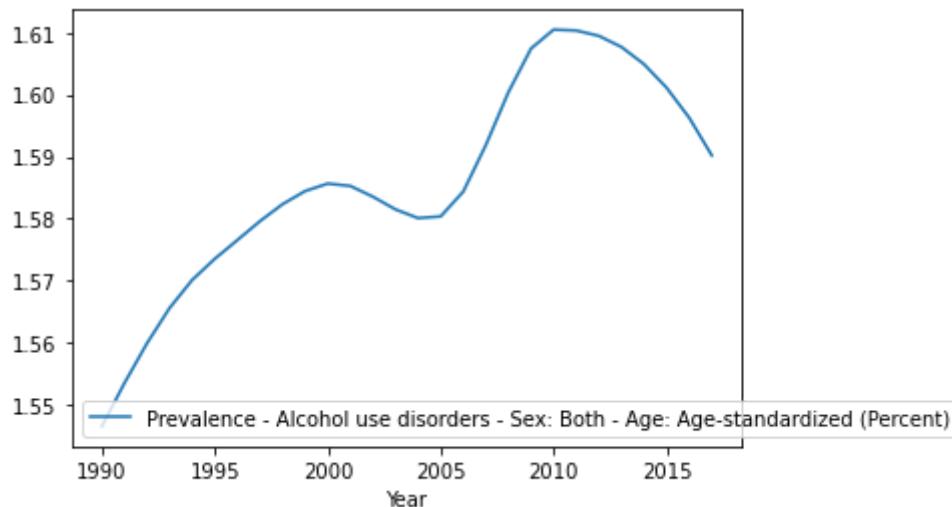
```
In [90]: df71.groupby('Year')[v11].median().plot(legend=True)
```

```
Out[90]: <AxesSubplot:xlabel='Year'>
```



```
In [91]: df71.groupby('Year')[v11].mean().plot(legend=True)
```

```
Out[91]: <AxesSubplot:xlabel='Year'>
```



```
In [92]: df71_country=df71.set_index('Year')
df71_country.groupby('Entity')[v11].plot(legend=True)
```

```
Out[92]: Entity
Afghanistan           AxesSubplot(0.125,0.125;0.775x0.755)
Albania               AxesSubplot(0.125,0.125;0.775x0.755)
Algeria                AxesSubplot(0.125,0.125;0.775x0.755)
American Samoa        AxesSubplot(0.125,0.125;0.775x0.755)
Andean Latin America   AxesSubplot(0.125,0.125;0.775x0.755)
                           ...
Western Sub-Saharan Africa AxesSubplot(0.125,0.125;0.775x0.755)
World                  AxesSubplot(0.125,0.125;0.775x0.755)
Yemen                   AxesSubplot(0.125,0.125;0.775x0.755)
Zambia                 AxesSubplot(0.125,0.125;0.775x0.755)
Zimbabwe                AxesSubplot(0.125,0.125;0.775x0.755)
Name: Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent), Length: 231, dtype: object
```

In [93]: df94.info()
df94.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6156 entries, 0 to 6155
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6156 non-null   object
1   Code      object
5292 non-null   object
2   Year      int64
6156 non-null   int64
3   Prevalence - Alcohol and substance use disorders: Both (age-standardized percent) 6156 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 192.5+ KB
```

Out[93]:

	Entity	Code	Year	Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)
0	Afghanistan	AFG	1990	1.530549
1	Afghanistan	AFG	1991	1.530105
2	Afghanistan	AFG	1992	1.529283
3	Afghanistan	AFG	1993	1.530944
4	Afghanistan	AFG	1994	1.532076

Checking for missing values:

In [94]: missing = pd.concat([df94.isnull().sum(), 100 * df94.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[94]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)		0	0.000000
Code	864	14.035088	

In [95]: df94.columns

Out[95]: Index(['Entity', 'Code', 'Year',
 'Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)',
 'dtype='object')]

In [96]: v1='Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)'

In [97]: df94.describe()

Out[97]:

	Year	Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)
count	6156.000000	6156.000000
mean	2003.000000	2.311523
std	7.789514	0.910629
min	1990.000000	1.196984
25%	1996.000000	1.651095
50%	2003.000000	2.054947
75%	2010.000000	2.618732
max	2016.000000	6.933015

In [98]: df94.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[98]:

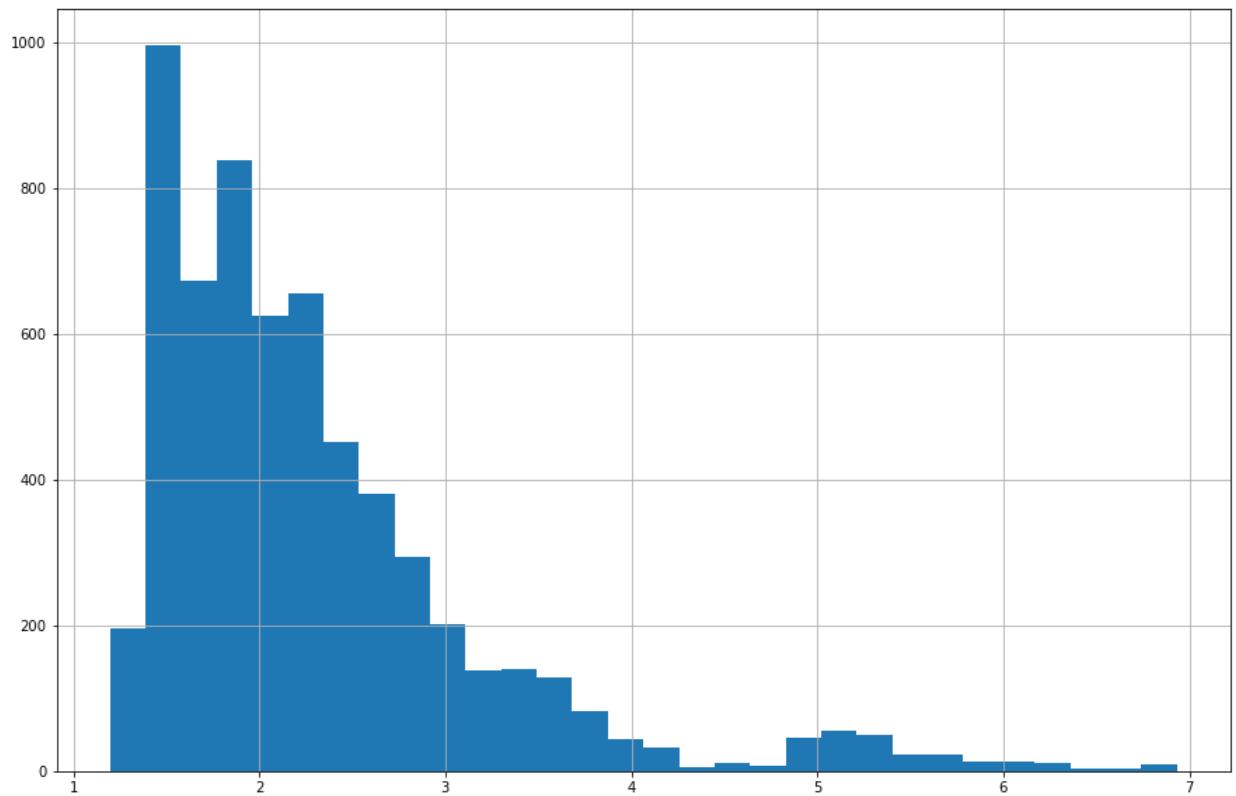
	Entity	Year	Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)
4492	Russia	2000	6.933015
4493	Russia	2001	6.909119
4494	Russia	2002	6.885858
4491	Russia	1999	6.875835
4495	Russia	2003	6.862802

In [99]: df94.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[99]:

	Entity	Year	Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)
2570	Indonesia	1995	1.196984
2569	Indonesia	1994	1.197815
2571	Indonesia	1996	1.198498
2568	Indonesia	1993	1.199902
2572	Indonesia	1997	1.200334

```
In [100]: df94[v1].hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [101]: df94.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[101]: Entity  
Indonesia      1.227345  
Morocco        1.248185  
Israel          1.326891  
Yemen           1.352224  
Sudan            1.355575  
Name: Prevalence - Alcohol and substance use disorders: Both (age-standardized percent), dtype: float64
```

```
In [102]: df94.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[102]: Entity  
United States    5.339861  
Greenland         5.363822  
Estonia           5.702670  
Eastern Europe     5.998806  
Russia             6.503332  
Name: Prevalence - Alcohol and substance use disorders: Both (age-standardized percent), dtype: float64
```

```
In [103]: df94_mean = df94.groupby('Year').mean()  
df94_mean.head()
```

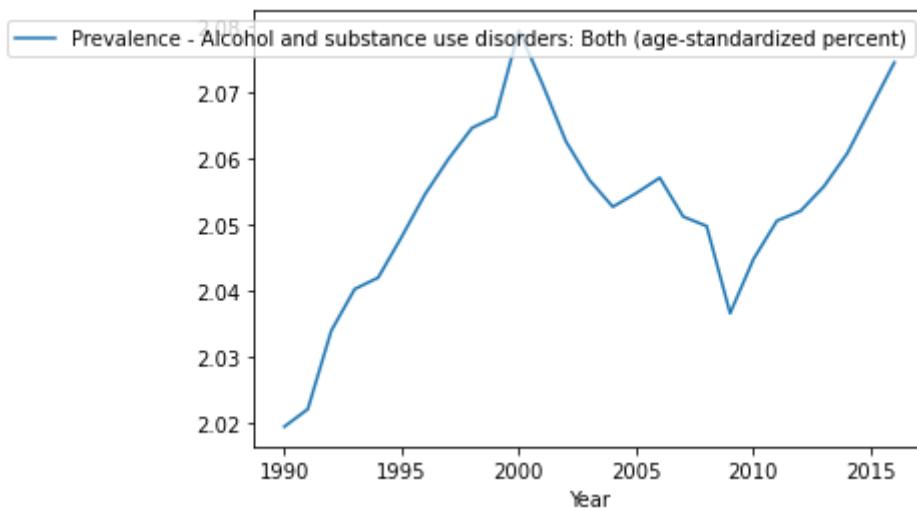
Out[103]:

Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)

Year	
1990	2.285541
1991	2.289356
1992	2.293757
1993	2.298235
1994	2.302819

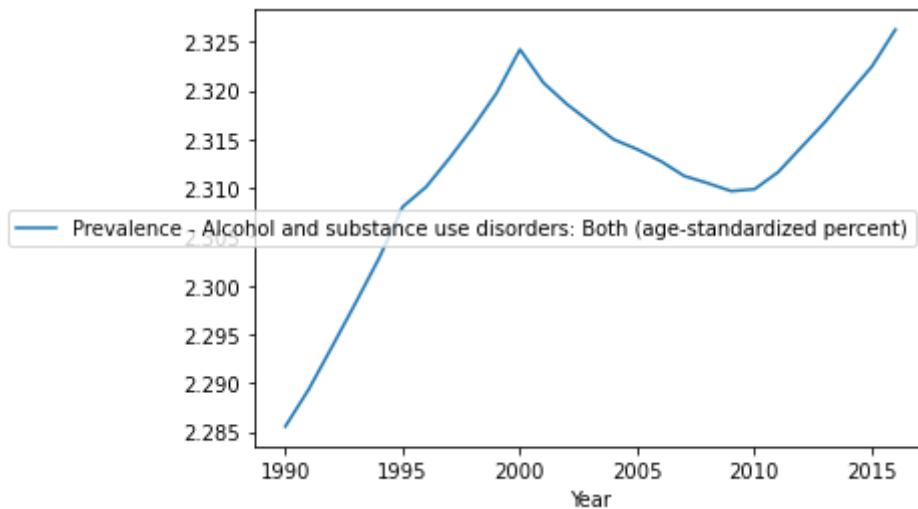
```
In [104]: df94.groupby('Year')[v1].median().plot(legend=True)
```

Out[104]: <AxesSubplot:xlabel='Year'>



```
In [105]: df94.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[105]: <AxesSubplot:xlabel='Year'>
```



In [106]: df20.info()
df20.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   float64
3   Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate) 6468 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

Out[106]:

	Entity	Code	Year	Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate)
0	Afghanistan	AFG	1990	0.046590
1	Afghanistan	AFG	1991	0.047434
2	Afghanistan	AFG	1992	0.048088
3	Afghanistan	AFG	1993	0.049568
4	Afghanistan	AFG	1994	0.051632

Checking for missing values:

In [107]: missing = pd.concat([df20.isnull().sum(), 100 * df20.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[107]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

In [108]: v1='Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized'

In [109]: df20.describe()

Out[109]:

	Year	Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate)
count	6468.000000	6468.000000
mean	2003.500000	0.040934
std	8.078372	0.052208
min	1990.000000	0.000459
25%	1996.750000	0.014080
50%	2003.500000	0.024147
75%	2010.250000	0.041248
max	2017.000000	0.542966

In [110]: df20.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[110]:

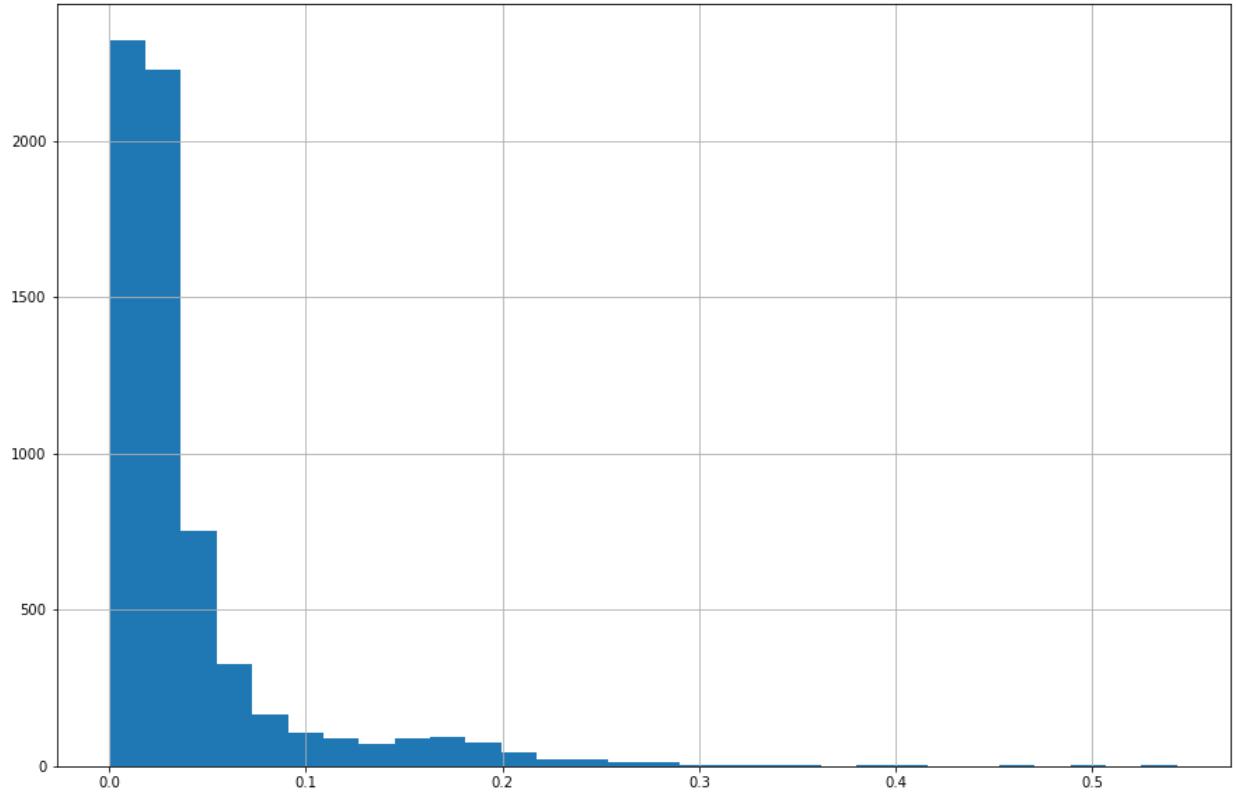
	Entity	Year	Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate)
6103	United States	2017	0.542966
6102	United States	2016	0.542785
6101	United States	2015	0.528149
4170	North America	2016	0.503404
4171	North America	2017	0.503207

In [111]: df20.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[111]:

	Entity	Year	Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate)
700	Bosnia and Herzegovina	1990	0.000459
708	Bosnia and Herzegovina	1998	0.000475
701	Bosnia and Herzegovina	1991	0.000486
707	Bosnia and Herzegovina	1997	0.000494
702	Bosnia and Herzegovina	1992	0.000505

```
In [112]: df20[v1].hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [113]: df20.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[113]: Entity
Egypt           0.001059
Armenia         0.002071
Bosnia and Herzegovina  0.002514
Singapore       0.003074
Zimbabwe        0.003207
Name: Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [114]: df20.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[114]: Entity
Denmark          0.178963
Finland          0.209884
Russia           0.213595
North America    0.243204
United States    0.259743
Name: Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [115]: df20_mean = df20.groupby('Year').mean()  
df20_mean.head()
```

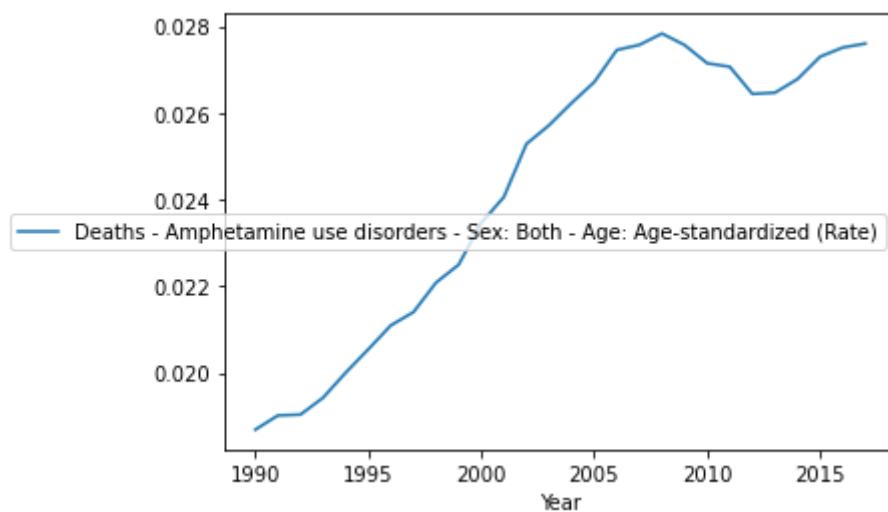
Out[115]:

Deaths - Amphetamine use disorders - Sex: Both - Age: Age-standardized (Rate)

Year	
1990	0.026969
1991	0.027594
1992	0.028419
1993	0.029440
1994	0.030814

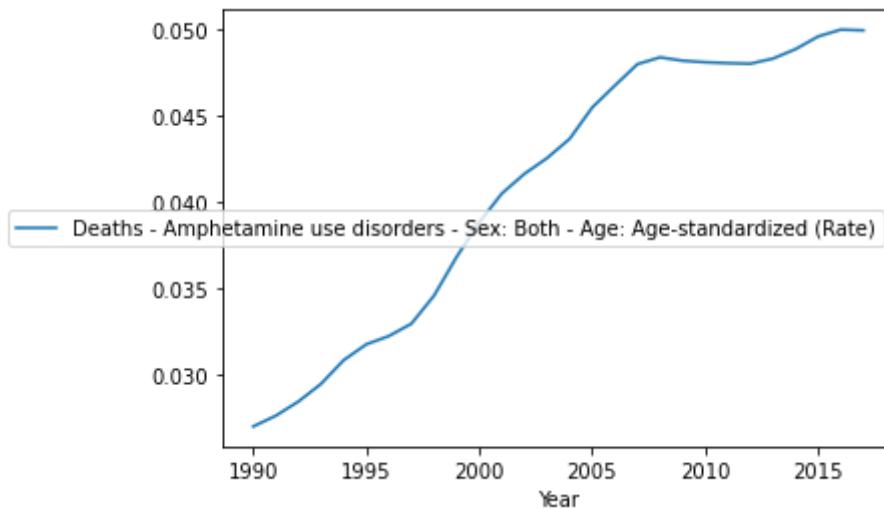
```
In [116]: df20.groupby('Year')[v1].median().plot(legend=True)
```

Out[116]: <AxesSubplot:xlabel='Year'>



```
In [117]: df20.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[117]: <AxesSubplot:xlabel='Year'>
```



```
In [118]: df21.info()
df21.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  --
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate) float64
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

```
Out[118]:
```

	Entity	Code	Year	Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate)
0	Afghanistan	AFG	1990	0.610114
1	Afghanistan	AFG	1991	0.622036
2	Afghanistan	AFG	1992	0.634234
3	Afghanistan	AFG	1993	0.656677
4	Afghanistan	AFG	1994	0.684709

Checking for missing values:

```
In [119]: missing = pd.concat([df21.isnull().sum(), 100 * df21.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[119]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

```
In [120]: v1='Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate')
```

```
In [121]: df21.describe()
```

Out[121]:

	Year	Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate)
count	6468.000000	6468.000000
mean	2003.500000	0.896465
std	8.078372	1.002894
min	1990.000000	0.041125
25%	1996.750000	0.321669
50%	2003.500000	0.577485
75%	2010.250000	1.051537
max	2017.000000	13.337479

```
In [122]: df21.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[122]:

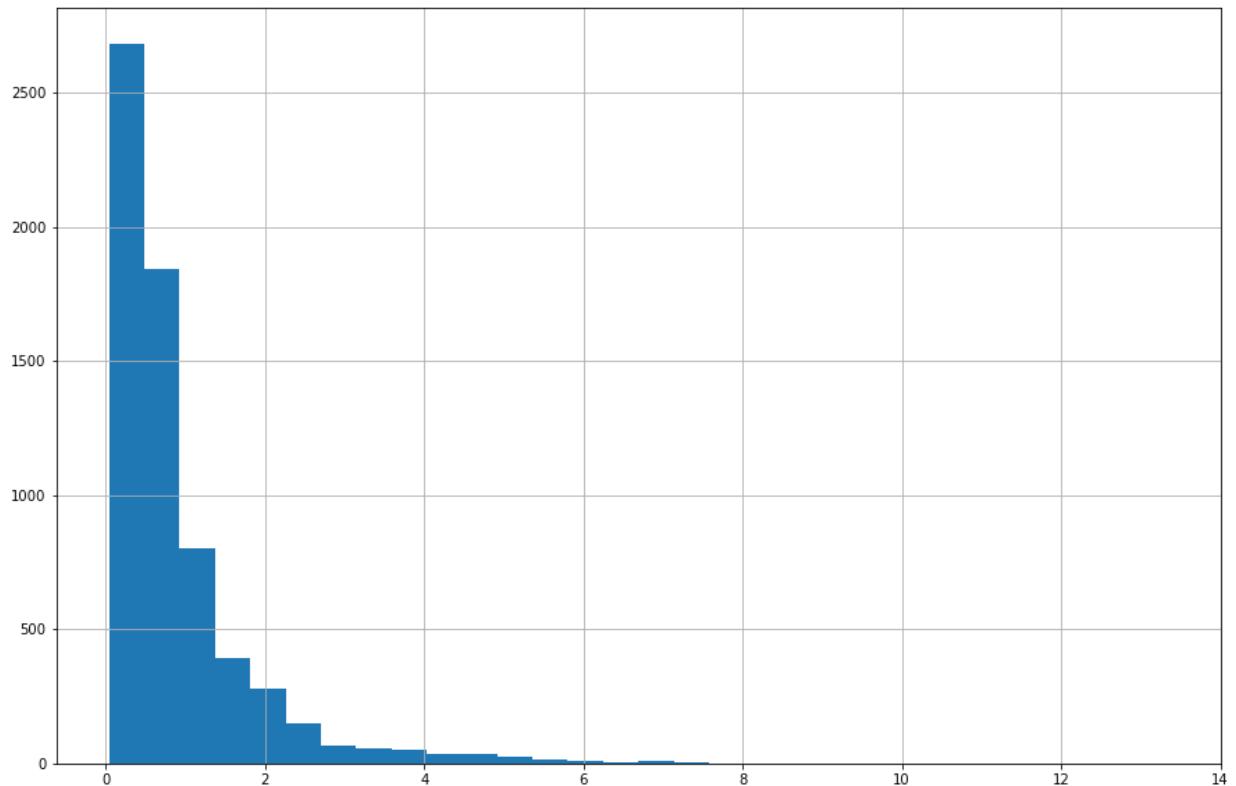
	Entity	Year	Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate)
6103	United States	2017	13.337479
6102	United States	2016	12.654711
4171	North America	2017	12.310067
4170	North America	2016	11.692211
6101	United States	2015	10.617001

```
In [123]: df21.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[123]:

	Entity	Year	Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate)
1792	Egypt	1990	0.041125
1793	Egypt	1991	0.041865
1794	Egypt	1992	0.042113
1795	Egypt	1993	0.045627
1796	Egypt	1994	0.048628

```
In [124]: df21[v1].hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [125]: df21.groupby('Entity')[v1].mean().sort_values().head()
```

Out[125]: Entity

Egypt	0.069787
Singapore	0.111500
Nepal	0.112146
Papua New Guinea	0.117811
Solomon Islands	0.120968

Name: Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64

```
In [126]: df21.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[126]: Entity
United Arab Emirates    3.971407
North America           4.240958
Russia                  4.262821
United States            4.511293
Libya                   4.944454
Name: Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized
(Rate), dtype: float64
```

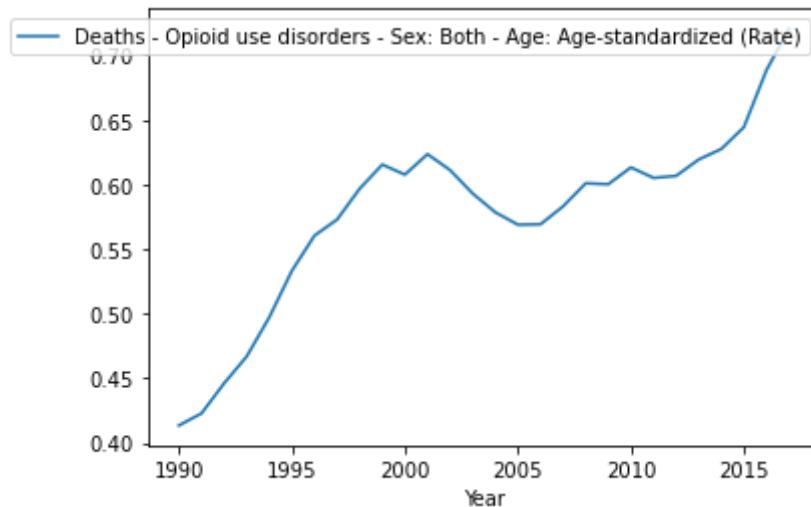
```
In [127]: df21_mean = df21.groupby('Year').mean()
df21_mean.head()
```

```
Out[127]: Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate)
```

Year	
1990	0.610466
1991	0.644234
1992	0.686798
1993	0.731864
1994	0.788686

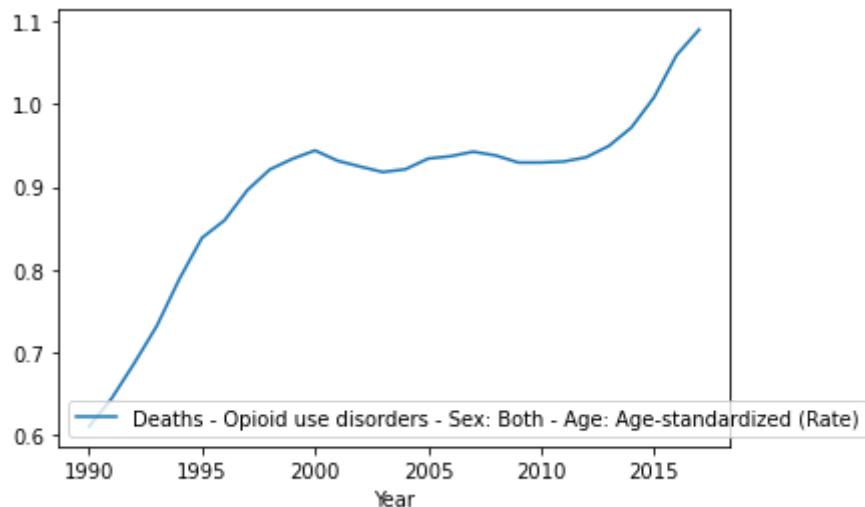
```
In [128]: df21.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[128]: <AxesSubplot:xlabel='Year'>
```



```
In [129]: df21.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[129]: <AxesSubplot:xlabel='Year'>
```



In [130]: df22.info()
df22.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized
(Rate) 6468 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

Out[130]:

	Entity	Code	Year	Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)
0	Afghanistan	AFG	1990	0.037949
1	Afghanistan	AFG	1991	0.038762
2	Afghanistan	AFG	1992	0.039532
3	Afghanistan	AFG	1993	0.040926
4	Afghanistan	AFG	1994	0.042738

Checking for missing values:

In [131]: missing = pd.concat([df22.isnull().sum(), 100 * df22.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[131]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

In [132]: v1='Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)'
df22[v1].value_counts()

In [133]: df22.describe()

Out[133]:

Year Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)		
	Entity	Year
count	6468.000000	6468.000000
mean	2003.500000	0.054284
std	8.078372	0.086960
min	1990.000000	0.000414
25%	1996.750000	0.011781
50%	2003.500000	0.031793
75%	2010.250000	0.060071
max	2017.000000	1.123634

In [134]: df22.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[134]:

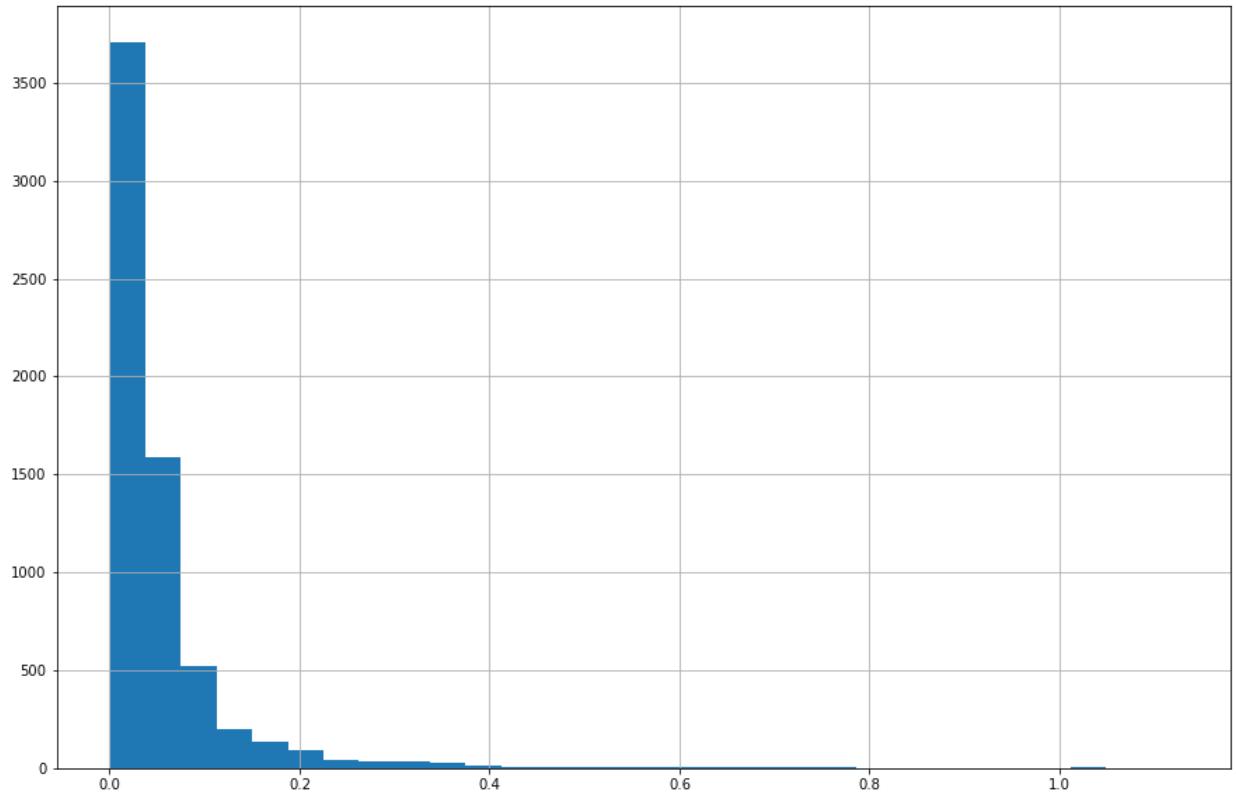
	Entity	Year	Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)
6103	United States	2017	1.123634
6102	United States	2016	1.106564
4171	North America	2017	1.047396
6101	United States	2015	1.040852
4170	North America	2016	1.032272

In [135]: df22.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[135]:

	Entity	Year	Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)
2241	Greece	1991	0.000414
2242	Greece	1992	0.000440
700	Bosnia and Herzegovina	1990	0.000458
2240	Greece	1990	0.000471
701	Bosnia and Herzegovina	1991	0.000502

```
In [136]: df22[v1].hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [137]: df22.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[137]: Entity  
Indonesia      0.001446  
Thailand       0.001703  
Egypt          0.001800  
Malaysia        0.002134  
Timor           0.002153  
Name: Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized  
(Rate), dtype: float64
```

```
In [138]: df22.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[138]: Entity  
Estonia         0.338552  
Russia          0.389446  
Puerto Rico     0.409847  
North America   0.550722  
United States   0.585157  
Name: Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized  
(Rate), dtype: float64
```

```
In [139]: df22_mean = df22.groupby('Year').mean()
df22_mean.head()
```

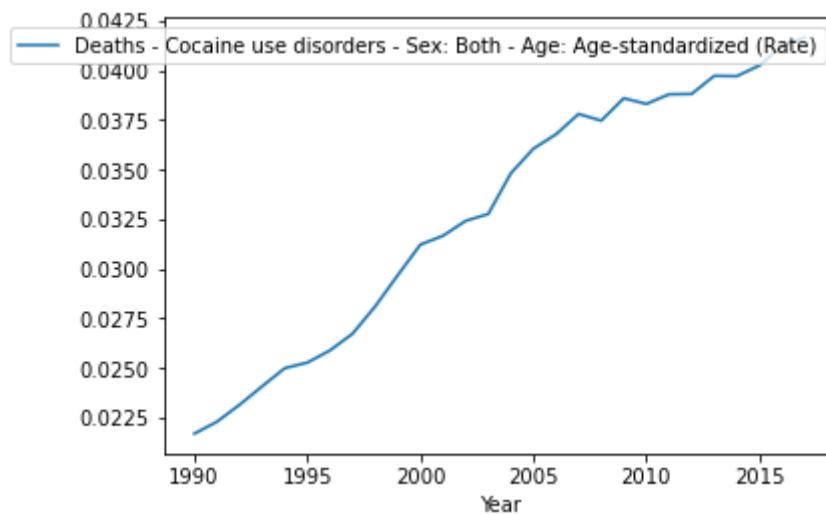
Out[139]:

Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)

Year	
1990	0.031518
1991	0.032841
1992	0.034707
1993	0.037382
1994	0.040709

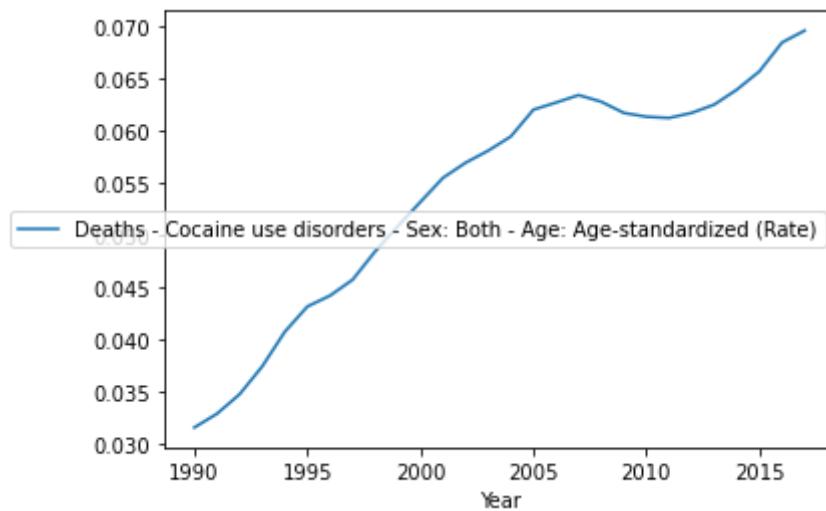
```
In [140]: df22.groupby('Year')[v1].median().plot(legend=True)
```

Out[140]: <AxesSubplot:xlabel='Year'>



```
In [141]: df22.groupby('Year')[v1].mean().plot(legend=True)
```

Out[141]: <AxesSubplot:xlabel='Year'>



In [142]: df23.info()
df23.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype  
---  --  
0   Entity    object 
6468 non-null   object 
1   Code      object 
5488 non-null   object 
2   Year      int64  
6468 non-null   int64  
3   Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate) 6468 non-null   float64 
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

Out[142]:

	Entity	Code	Year	Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)
0	Afghanistan	AFG	1990	1.053208
1	Afghanistan	AFG	1991	1.051713
2	Afghanistan	AFG	1992	1.044974
3	Afghanistan	AFG	1993	1.050948
4	Afghanistan	AFG	1994	1.066433

Checking for missing values:

In [143]: missing = pd.concat([df23.isnull().sum(), 100 * df23.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[143]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

In [144]: v1='Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)'
df23[v1].value_counts()

In [145]: df23.describe()

Out[145]:

	Year	Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)
count	6468.000000	6468.000000
mean	2003.500000	3.117192
std	8.078372	3.914501
min	1990.000000	0.174744
25%	1996.750000	1.107823
50%	2003.500000	2.103025
75%	2010.250000	3.383763
max	2017.000000	32.423774

In [146]: df23.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[146]:

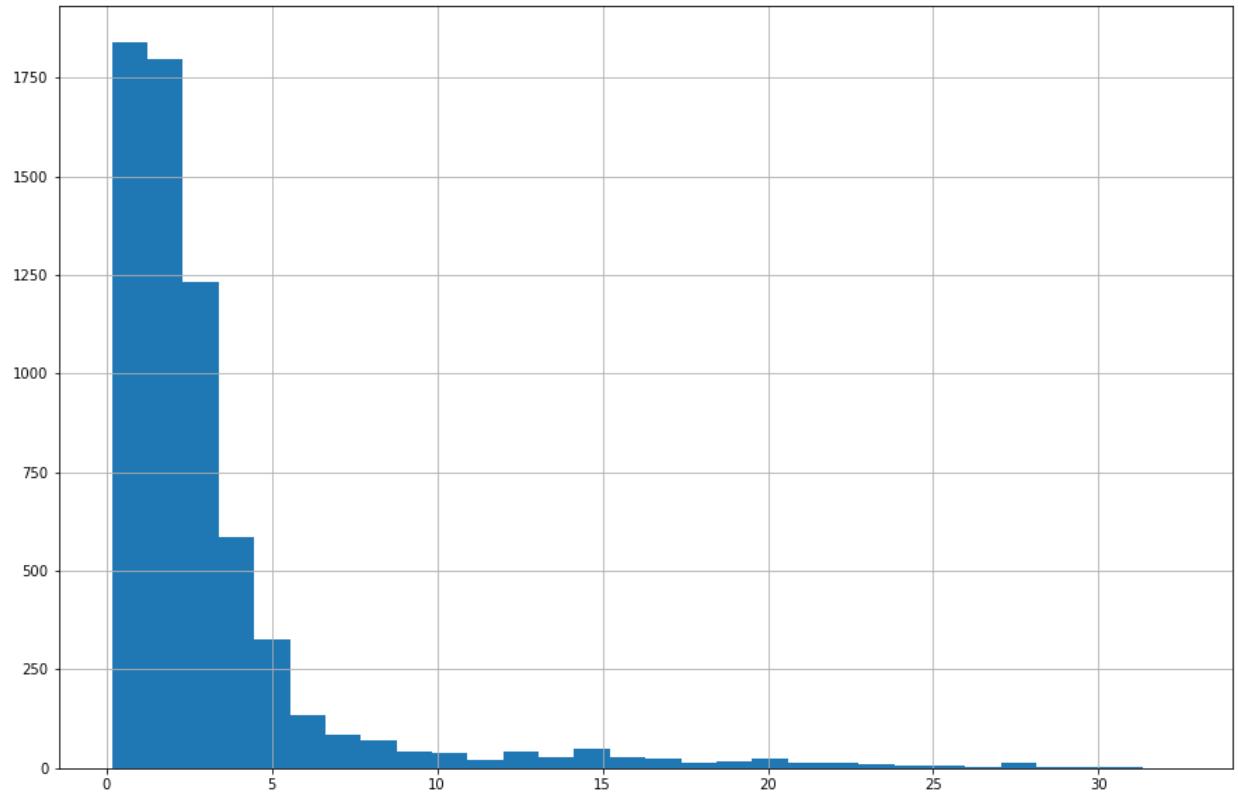
	Entity	Year	Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)
4708	Russia	1994	32.423774
4709	Russia	1995	30.823213
4717	Russia	2003	30.714187
4719	Russia	2005	30.710435
4718	Russia	2004	30.034615

In [147]: df23.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[147]:

	Entity	Year	Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)
5067	Singapore	2017	0.174744
5066	Singapore	2016	0.175719
5065	Singapore	2015	0.178954
5064	Singapore	2014	0.179878
5062	Singapore	2012	0.180693

```
In [148]: df23[v1].hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [149]: df23.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[149]: Entity  
Singapore      0.186827  
Bangladesh    0.221765  
Greece         0.226014  
Fiji           0.228064  
Egypt          0.228689  
Name: Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized  
(Rate), dtype: float64
```

```
In [150]: df23.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[150]: Entity
Ukraine           17.457360
Belarus            20.052555
Eastern Europe     20.801953
El Salvador        21.157093
Russia             22.584395
Name: Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized
(Rate), dtype: float64
```

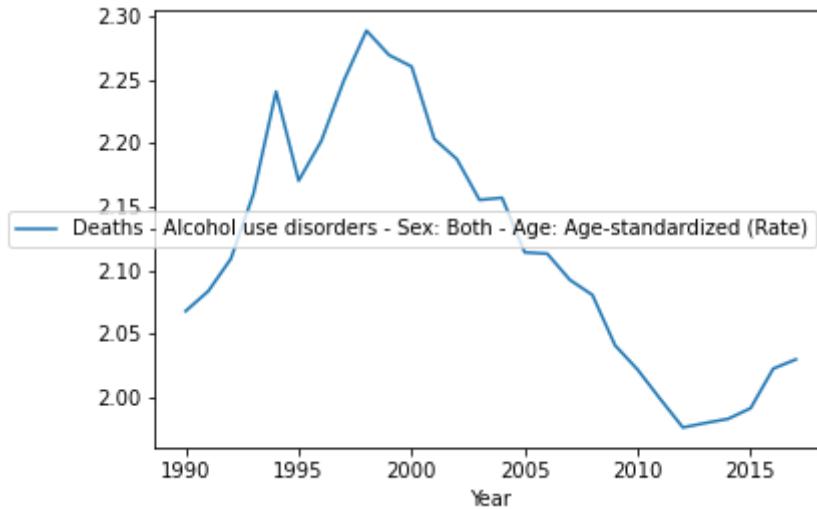
```
In [151]: df23_mean = df23.groupby('Year').mean()
df23_mean.head()
```

```
Out[151]: Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)
```

Year	
1990	2.891912
1991	2.981238
1992	3.105594
1993	3.297586
1994	3.452817

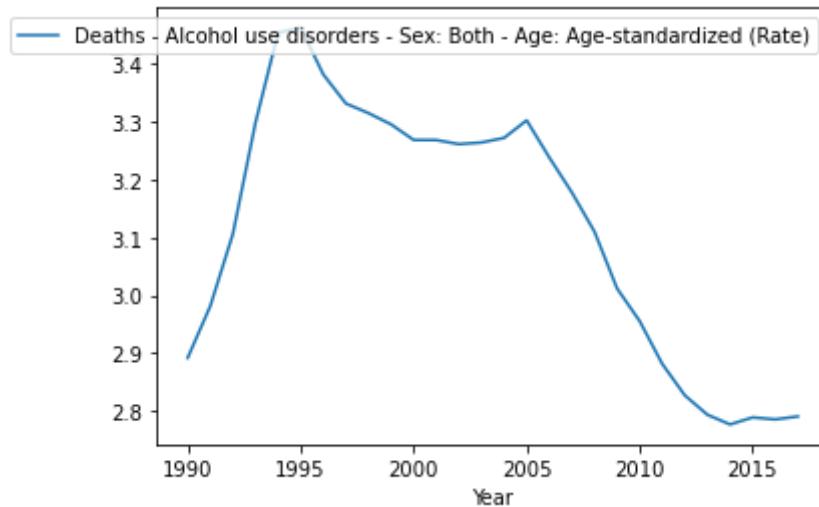
```
In [152]: df23.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[152]: <AxesSubplot:xlabel='Year'>
```



```
In [153]: df23.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[153]: <AxesSubplot:xlabel='Year'>
```



In [154]: df24.info()
df24.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   float64
3   Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)
e) 6468 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

Out[154]:

	Entity	Code	Year	Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)
0	Afghanistan	AFG	1990	1.137523
1	Afghanistan	AFG	1991	1.160556
2	Afghanistan	AFG	1992	1.183699
3	Afghanistan	AFG	1993	1.225011
4	Afghanistan	AFG	1994	1.276909

Checking for missing values:

In [155]: missing = pd.concat([df24.isnull().sum(), 100 * df24.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[155]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

In [156]: v1='Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)'

In [157]: df24.describe()

Out[157]:

Year Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)		
count	6468.000000	6468.000000
mean	2003.500000	1.468126
std	8.078372	1.419180
min	1990.000000	0.057751
25%	1996.750000	0.618545
50%	2003.500000	1.040799
75%	2010.250000	1.735679
max	2017.000000	18.749939

In [158]: df24.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[158]:

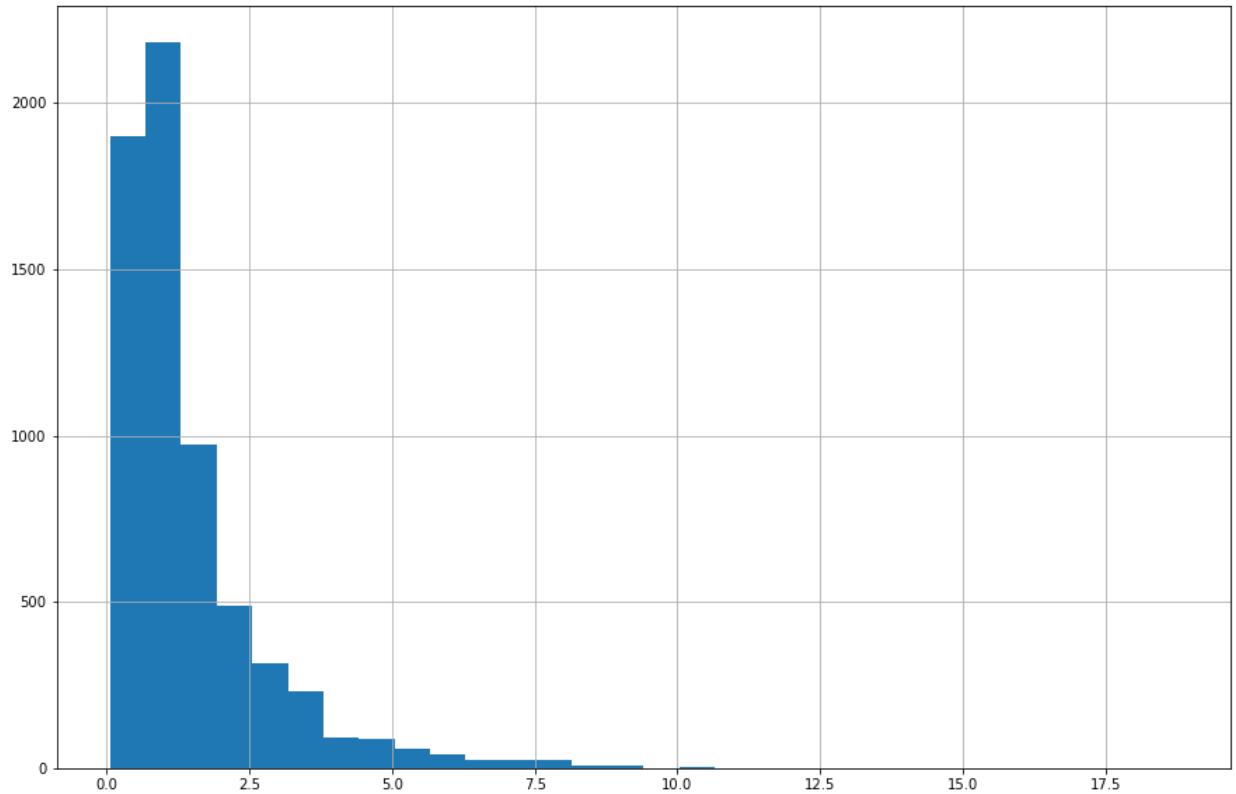
Entity Year Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)			
6103	United States	2017	18.749939
6102	United States	2016	17.999939
4171	North America	2017	17.353520
4170	North America	2016	16.677038
6101	United States	2015	15.688455

In [159]: df24.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[159]:

Entity Year Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)			
1792	Egypt	1990	0.057751
1794	Egypt	1992	0.058208
1793	Egypt	1991	0.058305
1795	Egypt	1993	0.062304
1796	Egypt	1994	0.065738

```
In [160]: df24[v1].hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [161]: df24.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[161]: Entity  
Egypt      0.088644  
Singapore  0.171825  
Indonesia 0.252608  
Cuba       0.264370  
Japan      0.271383  
Name: Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [162]: df24.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[162]: Entity
Scotland      5.380489
Libya        5.698478
Russia       6.254108
North America 6.851453
United States 7.278659
Name: Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

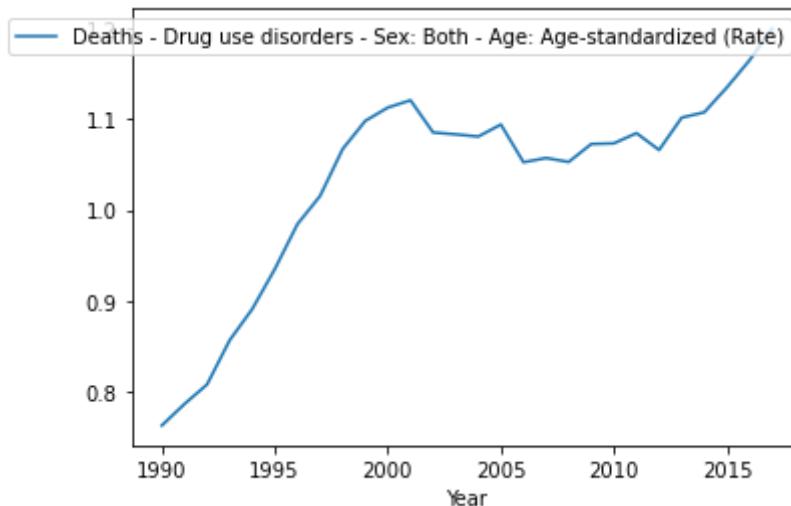
```
In [163]: df24_mean = df24.groupby('Year').mean()
df24_mean.head()
```

```
Out[163]: Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)
```

Year	
1990	1.019495
1991	1.064609
1992	1.120933
1993	1.181064
1994	1.258087

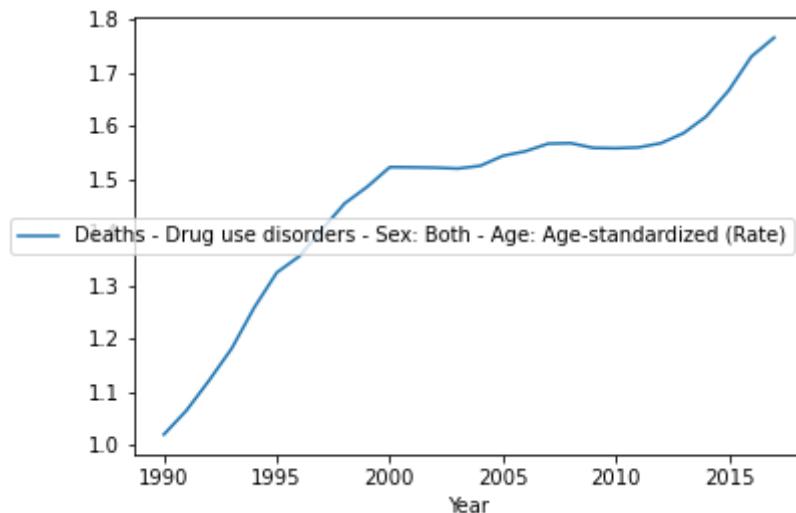
```
In [164]: df24.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[164]: <AxesSubplot:xlabel='Year'>
```



```
In [165]: df24.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[165]: <AxesSubplot:xlabel='Year'>
```



In [166]: df28.info()
df28.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Deaths - Substance use disorders - Sex: Both - Age: Age-standardized
(Rate) 6468 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

Out[166]:

	Entity	Code	Year	Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate)
0	Afghanistan	AFG	1990	2.190731
1	Afghanistan	AFG	1991	2.212270
2	Afghanistan	AFG	1992	2.228673
3	Afghanistan	AFG	1993	2.275958
4	Afghanistan	AFG	1994	2.343342

Checking for missing values:

In [167]: missing = pd.concat([df28.isnull().sum(), 100 * df28.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[167]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

In [168]: v1='Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (R

In [169]: df28.describe()

Out[169]:

Year Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate)		
count	6468.000000	6468.000000
mean	2003.500000	4.585319
std	8.078372	4.593365
min	1990.000000	0.259457
25%	1996.750000	2.191153
50%	2003.500000	3.315159
75%	2010.250000	5.131153
max	2017.000000	39.697337

In [170]: df28.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[170]:

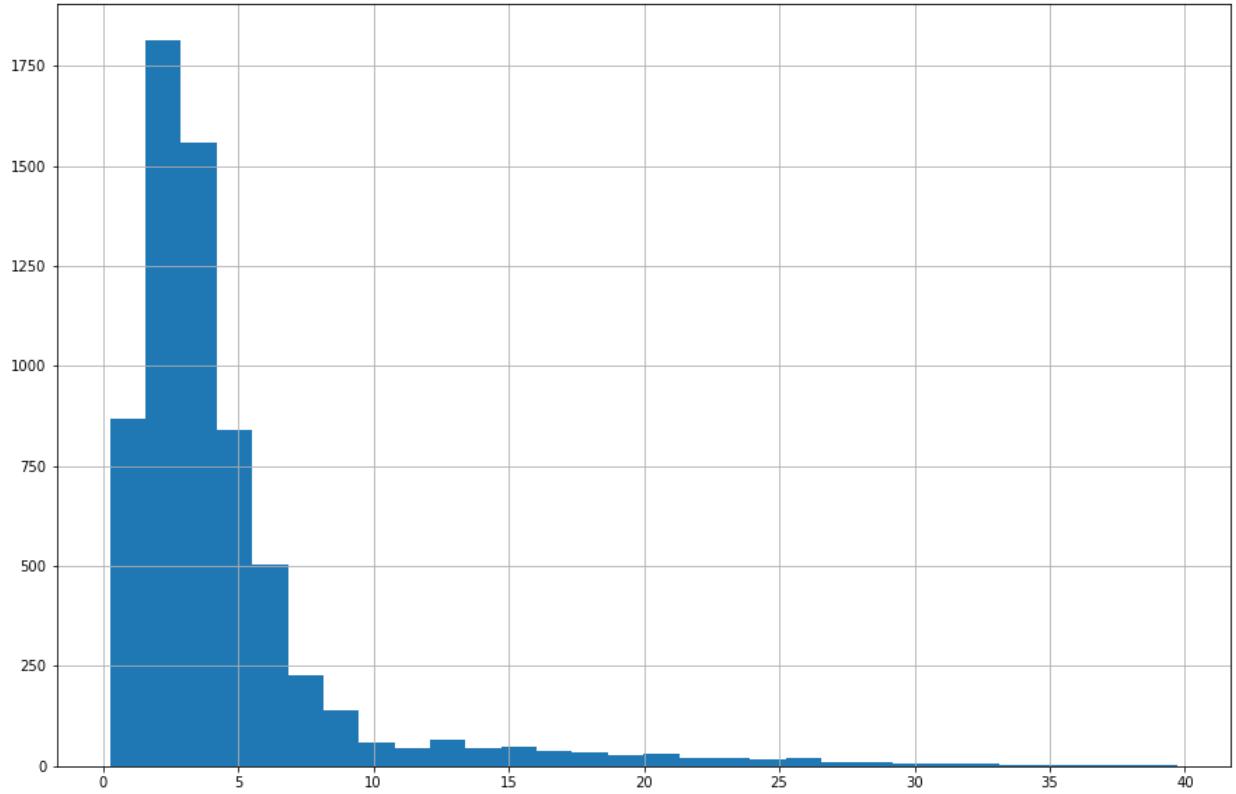
Entity Year Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate)			
4719	Russia	2005	39.697337
4708	Russia	1994	38.727243
4717	Russia	2003	38.007441
4718	Russia	2004	37.820639
4709	Russia	1995	36.953941

In [171]: df28.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[171]:

Entity Year Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate)			
1794	Egypt	1992	0.259457
1792	Egypt	1990	0.261869
1793	Egypt	1991	0.261898
1795	Egypt	1993	0.270092
1797	Egypt	1995	0.277304

```
In [172]: df28[v1].hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [173]: df28.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[173]: Entity  
Egypt      0.317333  
Singapore  0.358652  
Indonesia 0.549195  
Oman       0.627769  
Japan      0.650525  
Name: Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [174]: df28.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[174]: Entity
Estonia           21.393220
El Salvador       22.047560
Belarus           23.465071
Eastern Europe    25.985578
Russia            28.838503
Name: Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

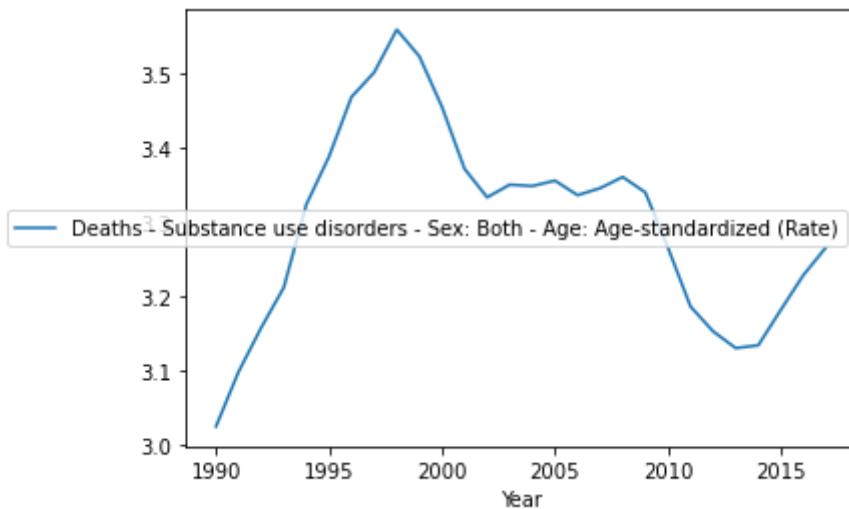
```
In [175]: df28_mean = df28.groupby('Year').mean()
df28_mean.head()
```

```
Out[175]: Deaths - Substance use disorders - Sex: Both - Age: Age-standardized (Rate)
```

Year	
1990	3.911408
1991	4.045847
1992	4.226527
1993	4.478650
1994	4.710904

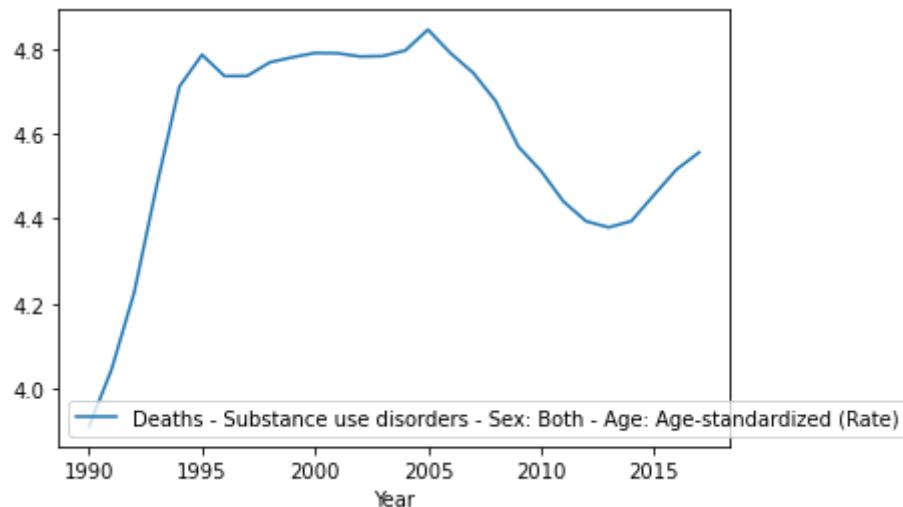
```
In [176]: df28.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[176]: <AxesSubplot:xlabel='Year'>
```



```
In [177]: df28.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[177]: <AxesSubplot:xlabel='Year'>
```



In [178]: df30.info()
df30.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 8 columns):
 #   Column           Non-Null
Count  Dtype
---  --  
0    Entity          6468 non-
null  object
1    Code            5488 non-
null  object
2    Year            6468 non-
null  int64
3    Deaths - Drug use - Sex: Both - Age: Under 5 (Number) 6468 non-
null  float64
4    Deaths - Drug use - Sex: Both - Age: 70+ years (Number) 6468 non-
null  float64
5    Deaths - Drug use - Sex: Both - Age: 50-69 years (Number) 6468 non-
null  float64
6    Deaths - Drug use - Sex: Both - Age: 15-49 years (Number) 6468 non-
null  float64
7    Deaths - Drug use - Sex: Both - Age: 5-14 years (Number) 6468 non-
null  float64
dtypes: float64(5), int64(1), object(2)
memory usage: 404.4+ KB
```

Out[178]:

	Entity	Code	Year	Deaths - Drug use - Sex: Both - Age: Under 5 (Number)	Deaths - Drug use - Sex: Both - Age: 70+ years (Number)	Deaths - Drug use - Sex: Both - Age: 50-69 years (Number)	Deaths - Drug use - Sex: Both - Age: 15-49 years (Number)	Deaths - Drug use - Sex: Both - Age: 5-14 years (Number)
0	Afghanistan	AFG	1990	0.0	24.017008	79.098359	105.207819	0.002243
1	Afghanistan	AFG	1991	0.0	24.901094	81.839095	111.027345	0.002157
2	Afghanistan	AFG	1992	0.0	25.901977	86.028561	135.900417	0.002296
3	Afghanistan	AFG	1993	0.0	27.008016	90.446762	167.578873	0.002531
4	Afghanistan	AFG	1994	0.0	28.146958	94.415585	184.081646	0.002608

Checking for missing values:

```
In [179]: missing = pd.concat([df30.isnull().sum(), 100 * df30.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[179]:

	count	%
Entity	0	0.000000
Year	0	0.000000
Deaths - Drug use - Sex: Both - Age: Under 5 (Number)	0	0.000000
Deaths - Drug use - Sex: Both - Age: 70+ years (Number)	0	0.000000
Deaths - Drug use - Sex: Both - Age: 50-69 years (Number)	0	0.000000
Deaths - Drug use - Sex: Both - Age: 15-49 years (Number)	0	0.000000
Deaths - Drug use - Sex: Both - Age: 5-14 years (Number)	0	0.000000
Code	980	15.151515

```
In [180]: v1='Deaths - Drug use - Sex: Both - Age: Under 5 (Number)'
v2='Deaths - Drug use - Sex: Both - Age: 70+ years (Number)'
v3='Deaths - Drug use - Sex: Both - Age: 50-69 years (Number)'
v4='Deaths - Drug use - Sex: Both - Age: 15-49 years (Number)'
v5='Deaths - Drug use - Sex: Both - Age: 5-14 years (Number)'
```

In [181]: df30[v1].isnull()

```
Out[181]: 0      False
1      False
2      False
3      False
4      False
...
6463    False
6464    False
6465    False
6466    False
6467    False
Name: Deaths - Drug use - Sex: Both - Age: Under 5 (Number), Length: 646
8, dtype: bool
```

In [182]: df30.describe()

Out[182]:

	Year	Deaths - Drug use - Sex: Both - Age: Under 5 (Number)	Deaths - Drug use - Sex: Both - Age: 70+ years (Number)	Deaths - Drug use - Sex: Both - Age: 50-69 years (Number)	Deaths - Drug use - Sex: Both - Age: 15-49 years (Number)	Deaths - Drug use - Sex: Both - Age: 5-14 years (Number)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.182077	1467.756992	2946.107240	4475.243668	0.952172
std	8.078372	1.183375	6466.685475	12133.824891	17411.533915	4.975542
min	1990.000000	0.000000	0.089418	0.387010	0.592723	0.000002
25%	1996.750000	0.000000	10.338780	30.160590	44.001977	0.000457
50%	2003.500000	0.000000	57.710932	130.331084	195.489700	0.002968
75%	2010.250000	0.000000	316.563570	698.710449	1067.528897	0.024194
max	2017.000000	28.358228	116684.846334	222893.905271	245763.599372	101.592724

In [183]: df30.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[183]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: Under 5 (Number)
6378	World	2012	28.358228
6379	World	2013	27.948511
6377	World	2011	22.404175
6380	World	2014	21.394403
6381	World	2015	18.649640

In [184]: df30.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[184]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: 70+ years (Number)
6383	World	2017	116684.846334
6382	World	2016	110554.907463
6381	World	2015	104883.602720
6380	World	2014	99627.789164
6379	World	2013	95311.875529

In [185]: df30.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[185]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: 50-69 years (Number)
6383	World	2017	222893.905271
6382	World	2016	216914.136760
6381	World	2015	206121.336954
6380	World	2014	193680.946274
6379	World	2013	182715.317870

In [186]: df30.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[186]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: 15-49 years (Number)
6383	World	2017	245763.599372
6382	World	2016	245443.575662
6381	World	2015	238041.173028
6374	World	2008	234813.376578
6371	World	2005	234247.001430

In [187]: df30.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[187]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: 5-14 years (Number)
6379	World	2013	101.592724
6380	World	2014	93.877996
6381	World	2015	92.081569
6378	World	2012	90.117507
6377	World	2011	85.663062

In [188]: df30.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[188]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: Under 5 (Number)
0	Afghanistan	1990	0.0
4245	Northern Ireland	2007	0.0
4244	Northern Ireland	2006	0.0
4243	Northern Ireland	2005	0.0
4242	Northern Ireland	2004	0.0

In [189]: df30.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[189]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: 70+ years (Number)
4256	Northern Mariana Islands	1990	0.089418
4257	Northern Mariana Islands	1991	0.099736
4258	Northern Mariana Islands	1992	0.111398
4259	Northern Mariana Islands	1993	0.123710
4260	Northern Mariana Islands	1994	0.135823

In [190]: df30.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[190]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: 50-69 years (Number)
4261	Northern Mariana Islands	1995	0.387010
4262	Northern Mariana Islands	1996	0.388606
4260	Northern Mariana Islands	1994	0.403443
4263	Northern Mariana Islands	1997	0.403578
4256	Northern Mariana Islands	1990	0.420918

In [191]: df30.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[191]:

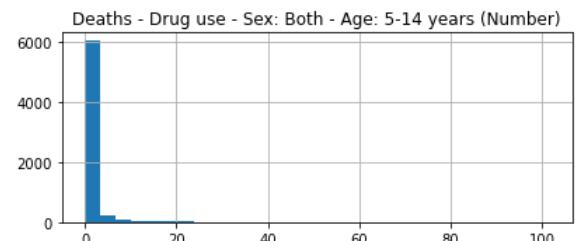
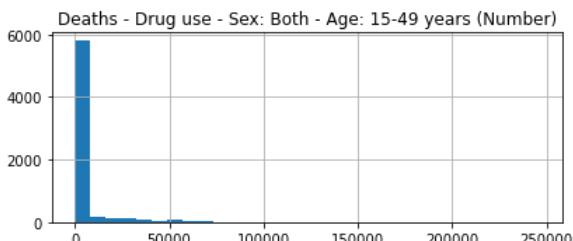
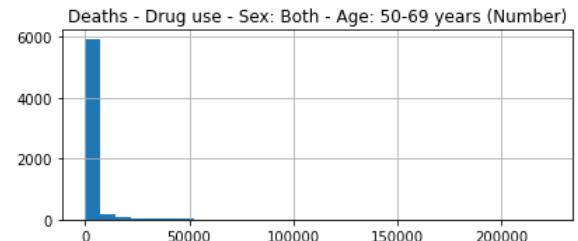
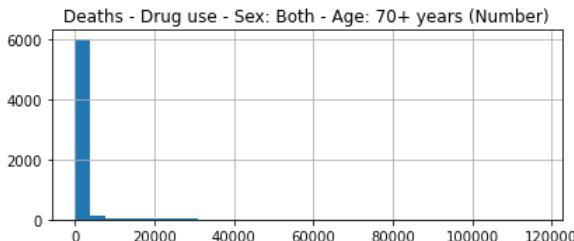
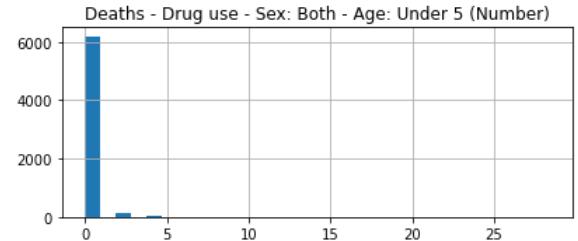
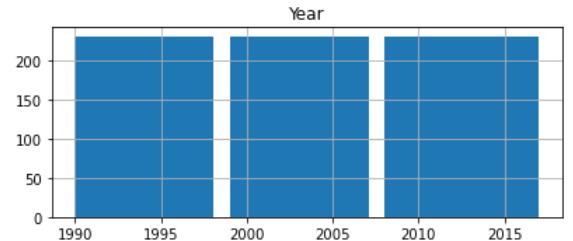
	Entity	Year	Deaths - Drug use - Sex: Both - Age: 15-49 years (Number)
4844	Sao Tome and Principe	1990	0.592723
84	American Samoa	1990	0.605287
196	Antigua and Barbuda	1990	0.617663
85	American Samoa	1991	0.629795
197	Antigua and Barbuda	1991	0.630554

In [192]: df30.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[192]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: 5-14 years (Number)
643	Bermuda	2017	0.000002
642	Bermuda	2016	0.000002
638	Bermuda	2012	0.000002
641	Bermuda	2015	0.000002
639	Bermuda	2013	0.000002

```
In [193]: df30.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [194]: df30.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[194]: Entity
Afghanistan      0.0
Namibia          0.0
Nepal            0.0
Netherlands     0.0
New Zealand      0.0
Name: Deaths - Drug use - Sex: Both - Age: Under 5 (Number), dtype: float
64
```

```
In [195]: df30.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[195]: Entity
Marshall Islands      0.204784
American Samoa        0.265908
Northern Mariana Islands 0.270602
Greenland             0.332583
Kiribati              0.353759
Name: Deaths - Drug use - Sex: Both - Age: 70+ years (Number), dtype: flo
at64
```

```
In [196]: df30.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[196]: Entity
American Samoa          0.671383
Northern Mariana Islands 0.719104
Marshall Islands         0.816320
Dominica                  1.126345
Kiribati                  1.130142
Name: Deaths - Drug use - Sex: Both - Age: 50-69 years (Number), dtype: float64
```

```
In [197]: df30.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[197]: Entity
American Samoa          0.746852
Antigua and Barbuda      1.023348
Northern Mariana Islands 1.032857
Dominica                  1.036928
Bermuda                   1.125690
Name: Deaths - Drug use - Sex: Both - Age: 15-49 years (Number), dtype: float64
```

```
In [198]: df30.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[198]: Entity
Antigua and Barbuda      0.000004
Bermuda                   0.000004
Sao Tome and Principe     0.000005
Barbados                  0.000009
United States Virgin Islands 0.000012
Name: Deaths - Drug use - Sex: Both - Age: 5-14 years (Number), dtype: float64
```

```
In [199]: df30.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[199]: Entity
Middle SDI                3.175032
Western Europe              3.594348
High SDI                   4.490094
High-income                 5.091682
World                      8.414478
Name: Deaths - Drug use - Sex: Both - Age: Under 5 (Number), dtype: float64
```

```
In [200]: df30.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[200]: Entity
Middle SDI                21556.096485
China                      27037.739587
East Asia                  28219.973003
Southeast Asia, East Asia, and Oceania 30819.570124
World                      69300.331228
Name: Deaths - Drug use - Sex: Both - Age: 70+ years (Number), dtype: float64
```

```
In [201]: df30.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[201]: Entity
Middle SDI           39902.136548
China                40033.927145
East Asia             42546.330984
Southeast Asia, East Asia, and Oceania 51378.636654
World                139727.961407
Name: Deaths - Drug use - Sex: Both - Age: 50-69 years (Number), dtype: float64
```

```
In [202]: df30.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[202]: Entity
High SDI              44574.805540
High-middle SDI       50327.033952
Middle SDI             58592.887801
Southeast Asia, East Asia, and Oceania 61061.145875
World                 213727.740508
Name: Deaths - Drug use - Sex: Both - Age: 15-49 years (Number), dtype: float64
```

```
In [203]: df30.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[203]: Entity
Western Europe        12.602979
Middle SDI             16.616184
High SDI               21.030255
High-income            24.488393
World                  44.063062
Name: Deaths - Drug use - Sex: Both - Age: 5-14 years (Number), dtype: float64
```

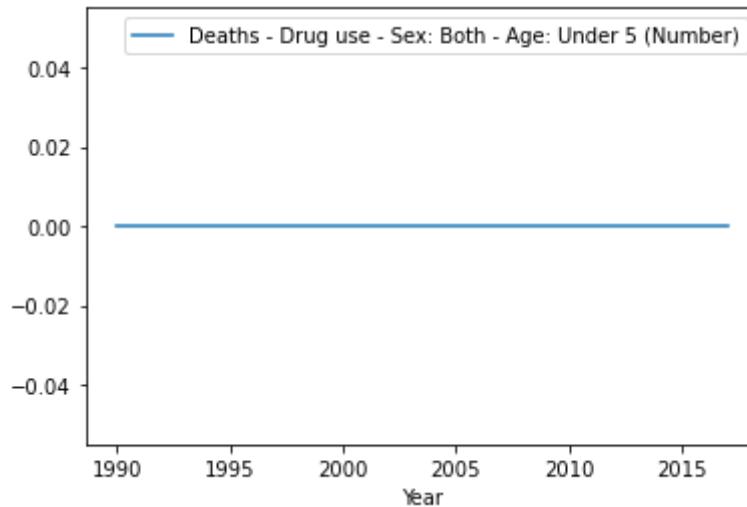
```
In [204]: df30_mean = df30.groupby('Year').mean()
df30_mean.head()
```

```
Out[204]:
```

Year	Deaths - Drug use - Sex: Both - Age: Under 5 (Number)	Deaths - Drug use - Sex: Both - Age: 70+ years (Number)	Deaths - Drug use - Sex: Both - Age: 50-69 years (Number)	Deaths - Drug use - Sex: Both - Age: 15-49 years (Number)	Deaths - Drug use - Sex: Both - Age: 5-14 years (Number)
1990	0.108632	700.600464	1781.207930	2814.434814	0.196254
1991	0.093136	750.374253	1867.597477	3036.186949	0.200130
1992	0.000000	797.017559	1948.329691	3320.840678	0.208308
1993	0.000000	845.732087	2037.536745	3600.720445	0.212771
1994	0.044126	896.401042	2115.964416	3872.615788	0.307455

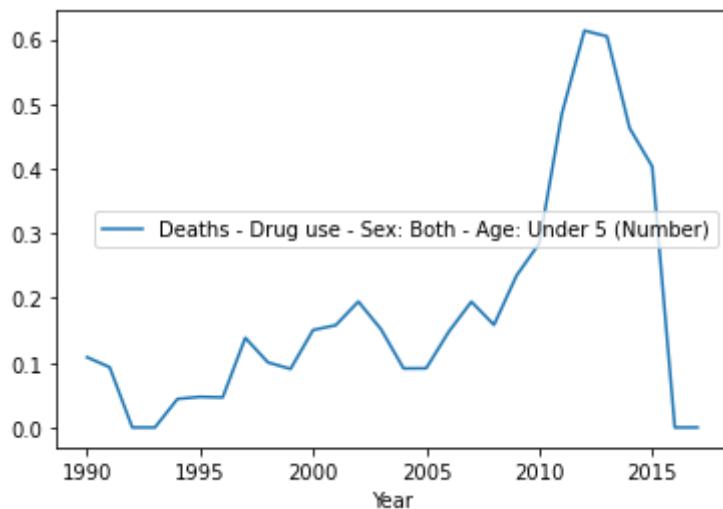
```
In [205]: df30.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[205]: <AxesSubplot:xlabel='Year'>
```



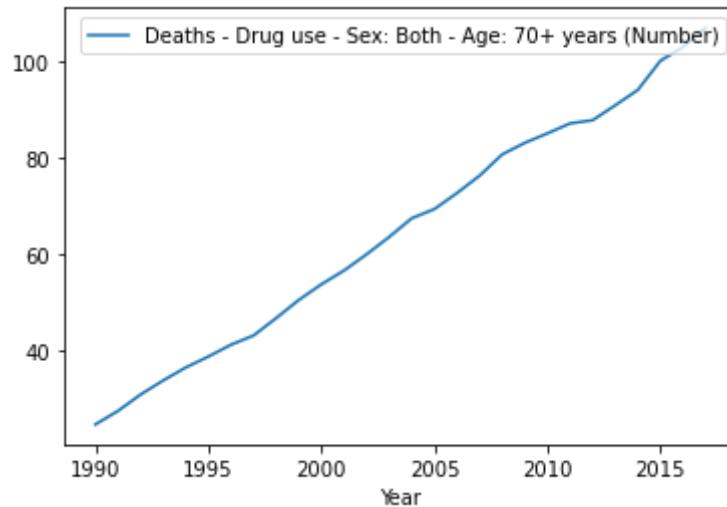
```
In [206]: df30.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[206]: <AxesSubplot:xlabel='Year'>
```



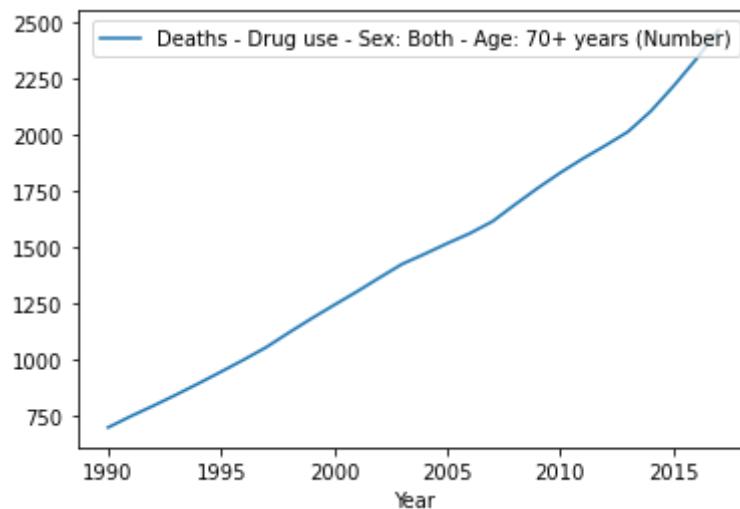
```
In [207]: df30.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[207]: <AxesSubplot:xlabel='Year'>
```



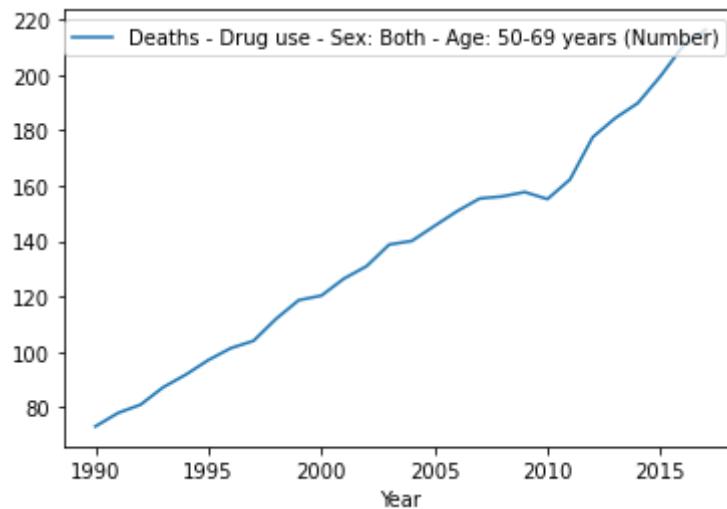
```
In [208]: df30.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[208]: <AxesSubplot:xlabel='Year'>
```



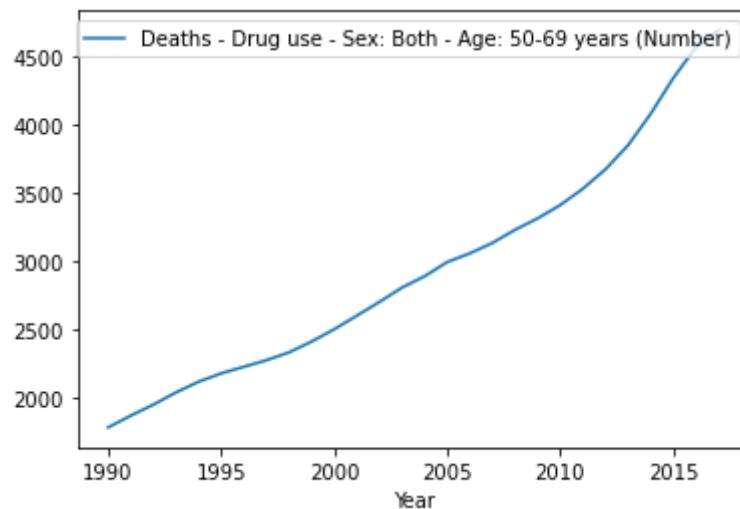
```
In [209]: df30.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[209]: <AxesSubplot:xlabel='Year'>
```



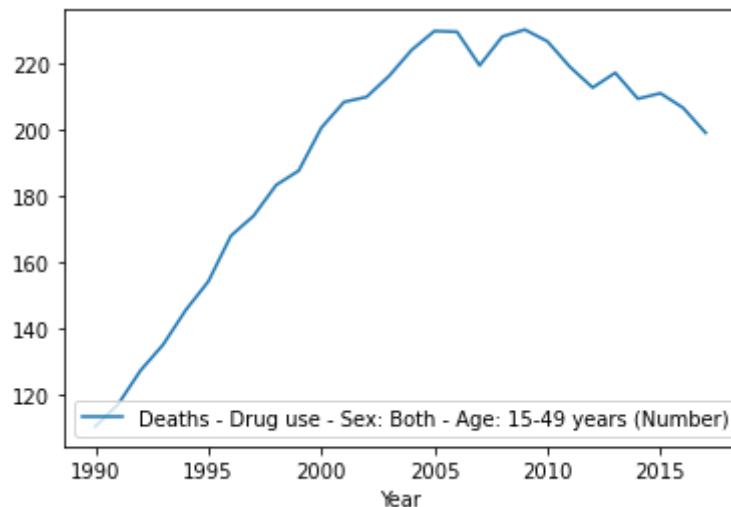
```
In [210]: df30.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[210]: <AxesSubplot:xlabel='Year'>
```



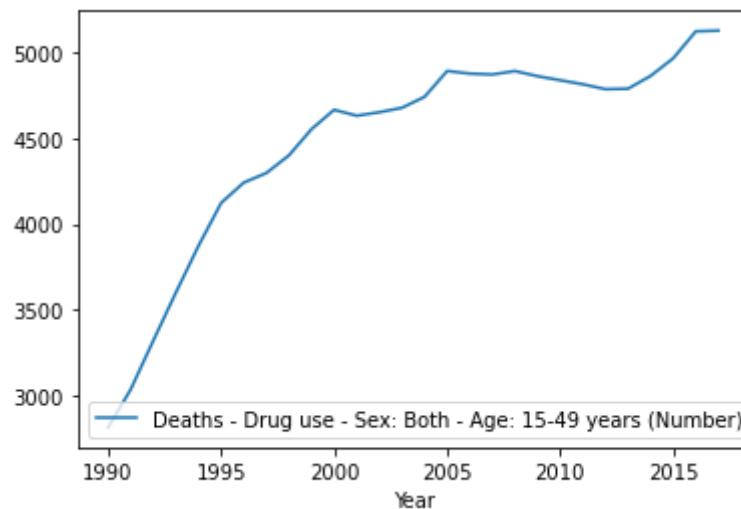
```
In [211]: df30.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[211]: <AxesSubplot:xlabel='Year'>
```



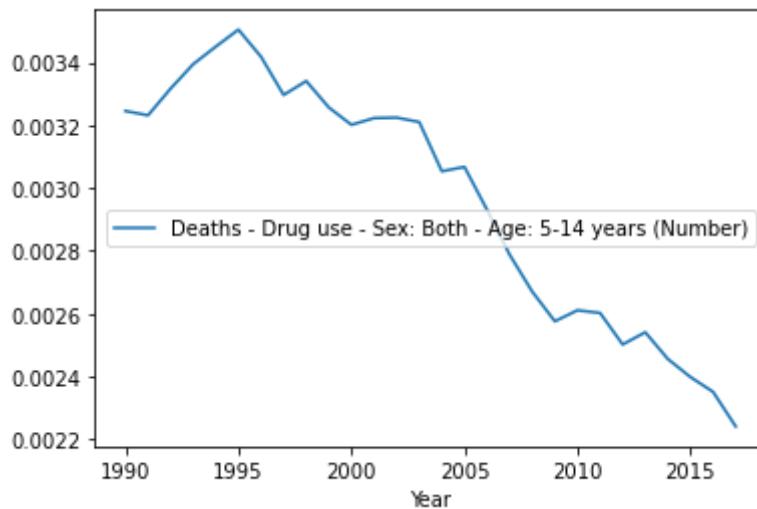
```
In [212]: df30.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[212]: <AxesSubplot:xlabel='Year'>
```



In [213]: df30.groupby('Year')[v5].median().plot(legend=True)

Out[213]: <AxesSubplot:xlabel='Year'>



In [214]: df31.info()
df31.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6156 entries, 0 to 6155
Data columns (total 4 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Entity            6156 non-null    object  
 1   Code              5292 non-null    object  
 2   Year              6156 non-null    int64  
 3   Deaths - Alcohol and substance use disorders  6156 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 192.5+ KB
```

Out[214]:

	Entity	Code	Year	Deaths - Alcohol and substance use disorders
0	Afghanistan	AFG	1990	67.324664
1	Afghanistan	AFG	1991	72.698996
2	Afghanistan	AFG	1992	79.808062
3	Afghanistan	AFG	1993	88.426374
4	Afghanistan	AFG	1994	97.675907

Checking for missing values:

```
In [215]: missing = pd.concat([df31.isnull().sum(), 100 * df31.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[215]:

	count	%
Entity	0	0.000000
Year	0	0.000000
Deaths - Alcohol and substance use disorders	0	0.000000
Code	864	14.035088

```
In [216]: v1='Deaths - Alcohol and substance use disorders'
```

```
In [217]: df31.describe()
```

Out[217]:

	Year	Deaths - Alcohol and substance use disorders
count	6156.000000	6156.000000
mean	2003.000000	2151.939444
std	7.789514	9900.863174
min	1990.000000	0.571512
25%	1996.000000	15.982657
50%	2003.000000	71.894964
75%	2010.000000	446.653002
max	2016.000000	143776.581600

```
In [218]: df31.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[218]:

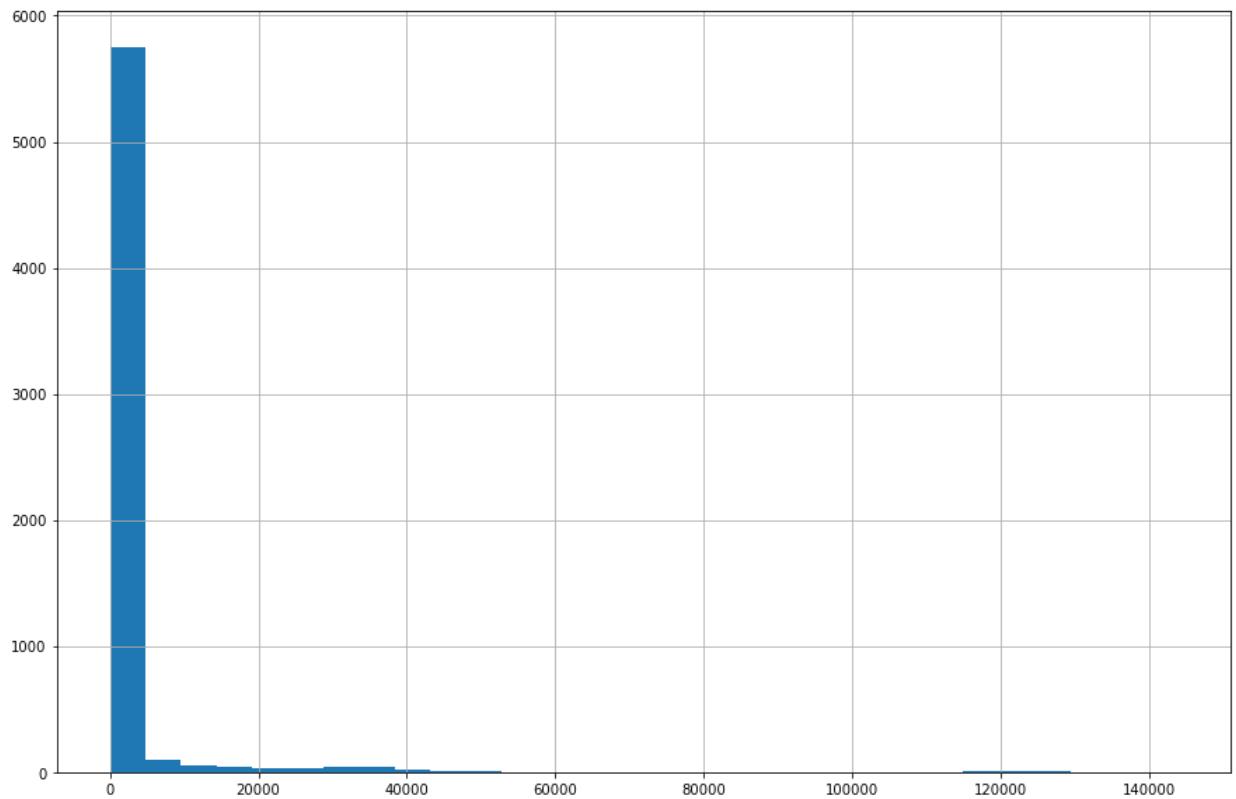
	Entity	Year	Deaths - Alcohol and substance use disorders
6074	World	2016	143776.5816
6073	World	2015	141007.4511
6072	World	2014	138556.4601
6071	World	2013	135736.3682
6070	World	2012	132941.2518

```
In [219]: df31.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[219]:

	Entity	Year	Deaths - Alcohol and substance use disorders
756	Brunei	1990	0.571512
4050	Northern Mariana Islands	1990	0.586343
4051	Northern Mariana Islands	1991	0.588056
757	Brunei	1991	0.593194
4052	Northern Mariana Islands	1992	0.594016

```
In [220]: df31[v1].hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [221]: df31.groupby('Entity')[v1].mean().sort_values().head()
```

Out[221]: Entity

Northern Mariana Islands	0.788138
Brunei	0.822249
Sao Tome and Principe	1.001119
Marshall Islands	1.155896
Barbados	1.246030

Name: Deaths - Alcohol and substance use disorders, dtype: float64

```
In [222]: df31.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[222]: Entity
High-middle SDI      34281.190394
China                 34318.965667
East Asia              35146.284570
Middle SDI             37415.651049
World                 120130.287554
Name: Deaths - Alcohol and substance use disorders, dtype: float64
```

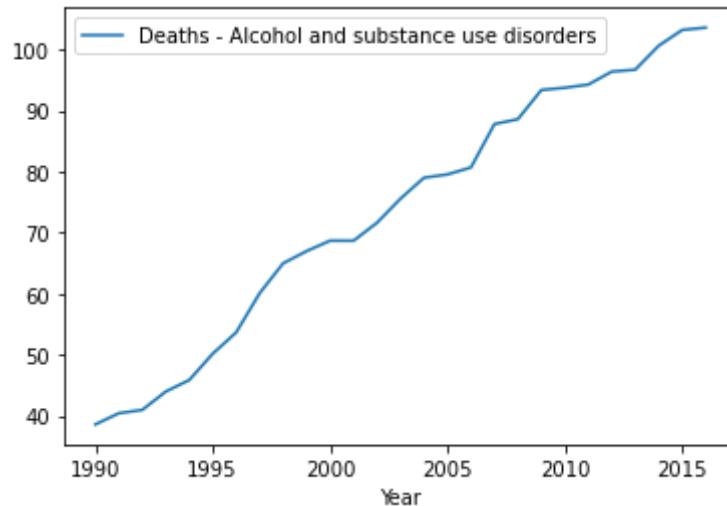
```
In [223]: df31_mean = df31.groupby('Year').mean()
df31_mean.head()
```

```
Out[223]: Deaths - Alcohol and substance use disorders
```

Year	
1990	1595.173681
1991	1660.925959
1992	1740.037997
1993	1825.546909
1994	1911.743534

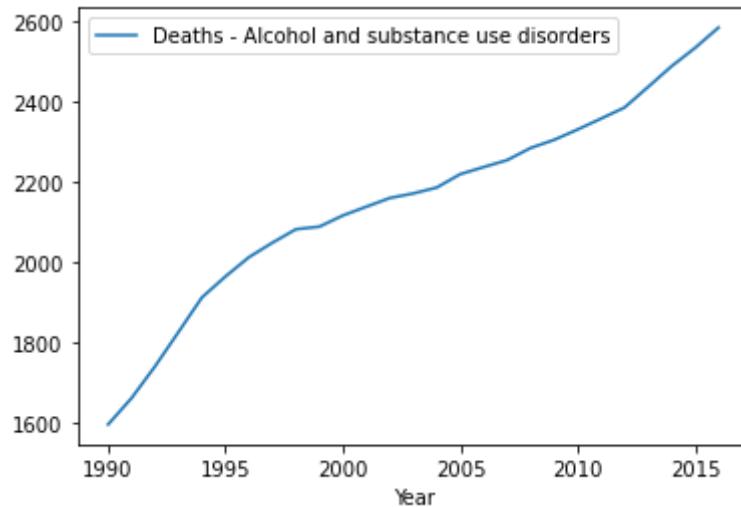
```
In [224]: df31.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[224]: <AxesSubplot:xlabel='Year'>
```



```
In [225]: df31.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[225]: <AxesSubplot:xlabel='Year'>
```



In [226]: df33.info()
df33.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 8 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)
6468 non-null   float64
4   Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)
6468 non-null   float64
5   Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)
6468 non-null   float64
6   Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)
6468 non-null   float64
7   Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)
6468 non-null   float64
dtypes: float64(5), int64(1), object(2)
memory usage: 404.4+ KB
```

Out[226]:

	Entity	Code	Year	Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)	Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)
0	Afghanistan	AFG	1990	0.0	11.379343	38.238407	28.636442	0.0
1	Afghanistan	AFG	1991	0.0	11.694379	40.001265	29.588942	0.0
2	Afghanistan	AFG	1992	0.0	12.077258	47.691407	31.241533	0.0
3	Afghanistan	AFG	1993	0.0	12.528090	57.800352	33.153549	0.0
4	Afghanistan	AFG	1994	0.0	13.013363	63.415100	34.867935	0.0

Checking for missing values:

```
In [447]: missing = pd.concat([df33.isnull().sum(), 100 * df33.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[447]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)		0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)		0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)		0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)		0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)		0	0.000000
Code	980	15.151515	

```
In [448]: v1='Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)'
v2='Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)'
v3='Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)'
v4='Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)'
v5='Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)'
```

In [449]: df33.describe()

Out[449]:

	Year	Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)	Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.182077	286.867573	1506.297338	482.373542	0.784290
std	8.078372	1.183375	1252.061052	6368.429348	2216.080184	4.546291
min	1990.000000	0.000000	0.009793	0.137206	0.019799	0.000000
25%	1996.750000	0.000000	1.531245	11.024519	2.605122	0.000000
50%	2003.500000	0.000000	9.876037	48.756793	14.379320	0.000000
75%	2010.250000	0.000000	50.517224	289.773766	82.571464	0.000000
max	2017.000000	28.358228	24164.577947	97649.177492	44798.793009	95.749609

In [450]: df33.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[450]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)
6378	World	2012	28.358228
6379	World	2013	27.948511
6377	World	2011	22.404175
6380	World	2014	21.394403
6381	World	2015	18.649640

In [451]: df33.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[451]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)
6383	World	2017	24164.577947
6382	World	2016	22656.053024
6381	World	2015	21125.819262
6380	World	2014	19823.720758
6379	World	2013	18741.458990

In [452]: df33.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[452]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)
6383	World	2017	97649.177492
6382	World	2016	94010.951906
6381	World	2015	86993.932562
6380	World	2014	82040.845694
6379	World	2013	77819.257944

In [453]: df33.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[453]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)
6383	World	2017	44798.793009
6382	World	2016	43019.141906
6381	World	2015	39389.699879
6380	World	2014	35915.054227
6379	World	2013	32675.466589

In [454]: df33.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[454]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)
6379	World	2013	95.749609
6380	World	2014	87.897372
6381	World	2015	86.015763
6378	World	2012	84.375237
6377	World	2011	79.935766

In [455]: df33.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[455]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)
0	Afghanistan	1990	0.0
4245	Northern Ireland	2007	0.0
4244	Northern Ireland	2006	0.0
4243	Northern Ireland	2005	0.0
4242	Northern Ireland	2004	0.0

In [456]: df33.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[456]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)
2268	Greenland	1990	0.009793
2269	Greenland	1991	0.010144
2271	Greenland	1993	0.010385
2270	Greenland	1992	0.010420
2272	Greenland	1994	0.010578

In [457]: df33.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[457]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)
4283	Northern Mariana Islands	2017	0.137206
4282	Northern Mariana Islands	2016	0.140526
4281	Northern Mariana Islands	2015	0.145856
4280	Northern Mariana Islands	2014	0.145867
4256	Northern Mariana Islands	1990	0.146933

In [458]: df33.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[458]:

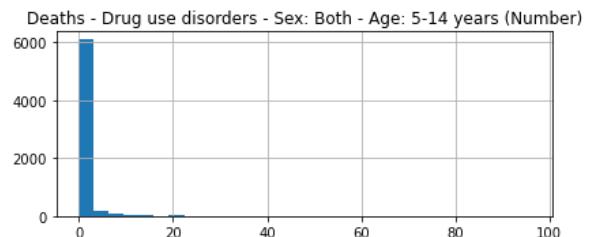
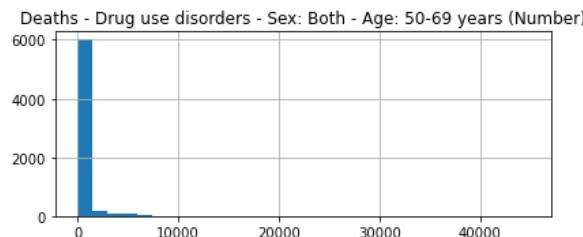
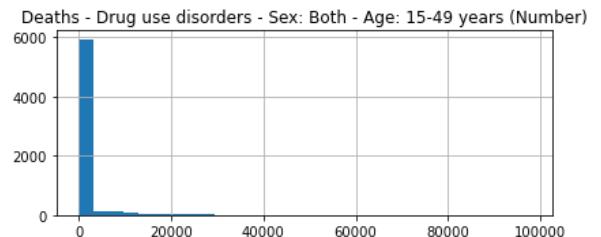
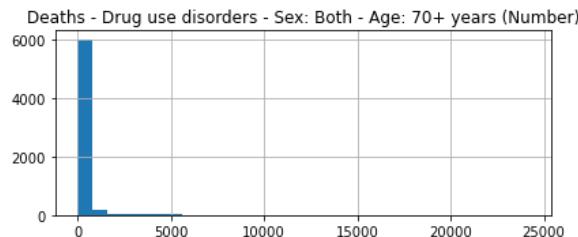
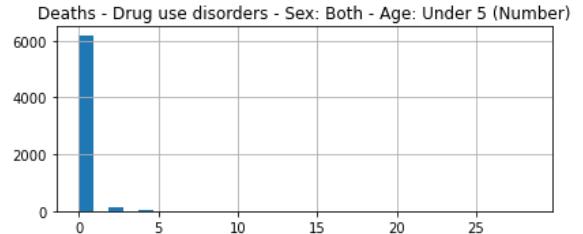
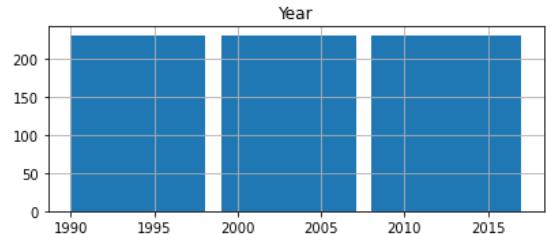
	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)
4256	Northern Mariana Islands	1990	0.019799
4262	Northern Mariana Islands	1996	0.020232
4261	Northern Mariana Islands	1995	0.020350
4263	Northern Mariana Islands	1997	0.021046
4260	Northern Mariana Islands	1994	0.021101

In [459]: df33.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[459]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)
0	Afghanistan	1990	0.0
4228	Northern Ireland	1990	0.0
4227	North Macedonia	2017	0.0
4226	North Macedonia	2016	0.0
4225	North Macedonia	2015	0.0

```
In [460]: df33.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [461]: df33.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[461]: Entity
Afghanistan      0.0
Namibia          0.0
Nepal            0.0
Netherlands      0.0
New Zealand      0.0
Name: Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number), dt
      type: float64
```

```
In [462]: df33[v2].sort_values(ascending=False).head()
```

```
Out[462]: 6383    24164.577947
6382    22656.053024
6381    21125.819262
6380    19823.720758
6379    18741.458990
Name: Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number),
      dtype: float64
```

```
In [463]: df33[v3].sort_values(ascending=False).head()
```

```
Out[463]: 6383    97649.177492
6382    94010.951906
6381    86993.932562
6380    82040.845694
6379    77819.257944
Name: Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number), dtype: float64
```

```
In [464]: df33[v4].sort_values(ascending=False).head()
```

```
Out[464]: 6383    44798.793009
6382    43019.141906
6381    39389.699879
6380    35915.054227
6379    32675.466589
Name: Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number), dtype: float64
```

```
In [465]: df33[v5].sort_values(ascending=False).head()
```

```
Out[465]: 6379    95.749609
6380    87.897372
6381    86.015763
6378    84.375237
6377    79.935766
Name: Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number), dtype: float64
```

```
In [466]: df33.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[466]: Entity
Middle SDI      3.175032
Western Europe  3.594348
High SDI        4.490094
High-income     5.091682
World           8.414478
Name: Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number), dtype: float64
```

```
In [467]: df33.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[467]: Entity
Middle SDI      4268.612875
China            5290.685618
East Asia        5467.397909
Southeast Asia, East Asia, and Oceania 5752.121170
World            13695.637664
Name: Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number), dtype: float64
```

```
In [468]: df33.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[468]: Entity
North America           18349.142591
Southeast Asia, East Asia, and Oceania   18382.033234
High-income              25808.051847
High SDI                 26343.420240
World                    71308.089117
Name: Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number), dtype: float64
```

```
In [469]: df33.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[469]: Entity
United States          6183.697949
North America           6461.894028
High-income              8202.666128
High SDI                 8249.822770
World                    23091.609099
Name: Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number), dtype: float64
```

```
In [470]: df33.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[470]: Entity
Western Europe          12.413323
Middle SDI               13.801849
High SDI                  18.435608
High-income                21.803158
World                     36.243902
Name: Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number), dtype: float64
```

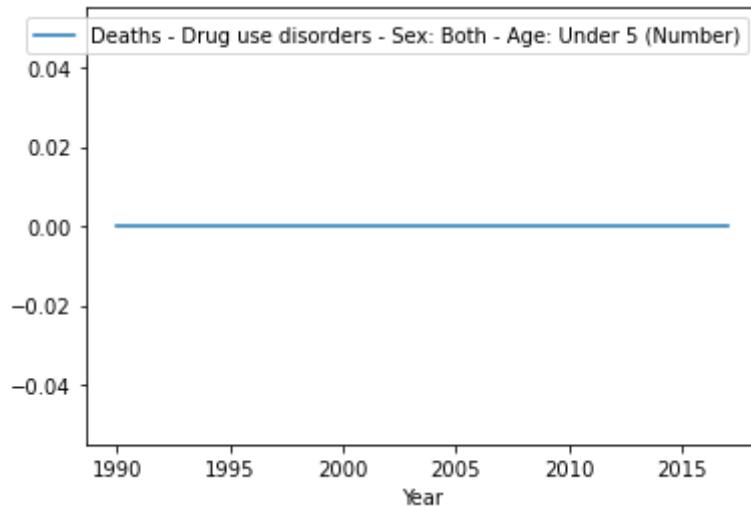
```
In [471]: df33_mean = df33.groupby('Year').mean()
df33_mean.head()
```

```
Out[471]:
```

	Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)	Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)	Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)
Year					
1990	0.108632	167.265386	1048.630085	279.469630	0.000000
1991	0.093136	179.658989	1142.091066	298.707310	0.000000
1992	0.000000	187.962337	1257.926373	315.193663	0.000000
1993	0.000000	194.471331	1352.237656	332.361588	0.000000
1994	0.044126	202.702055	1440.932595	344.661816	0.091031

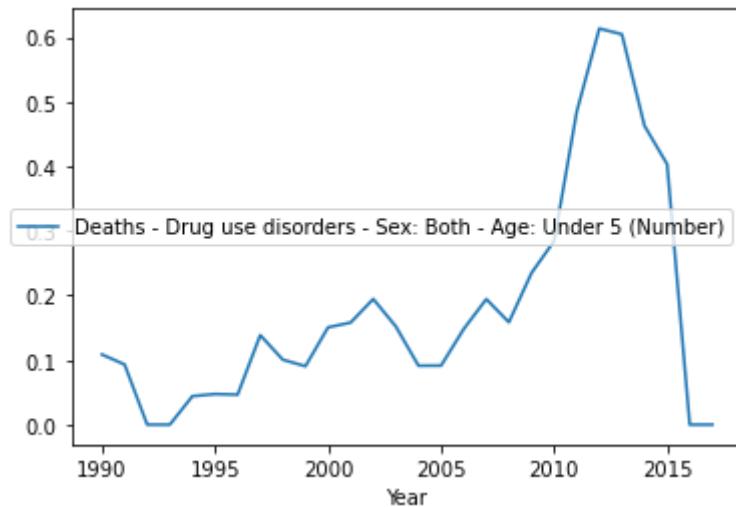
```
In [472]: df33.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[472]: <AxesSubplot:xlabel='Year'>
```



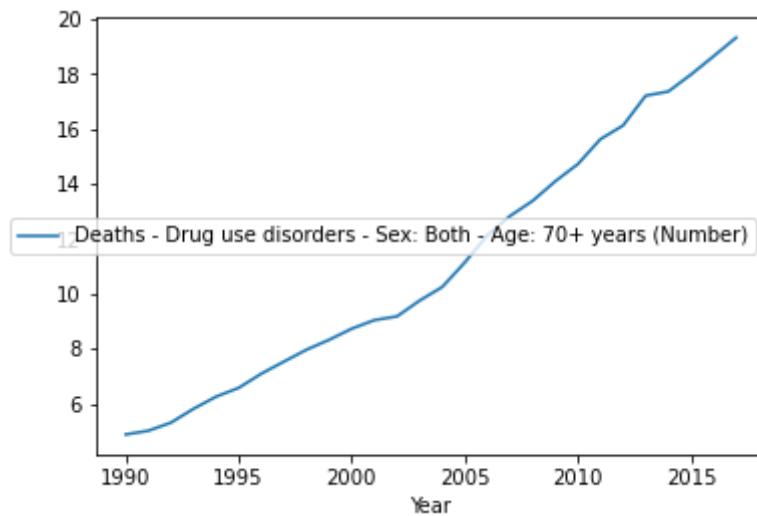
```
In [473]: df33.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[473]: <AxesSubplot:xlabel='Year'>
```



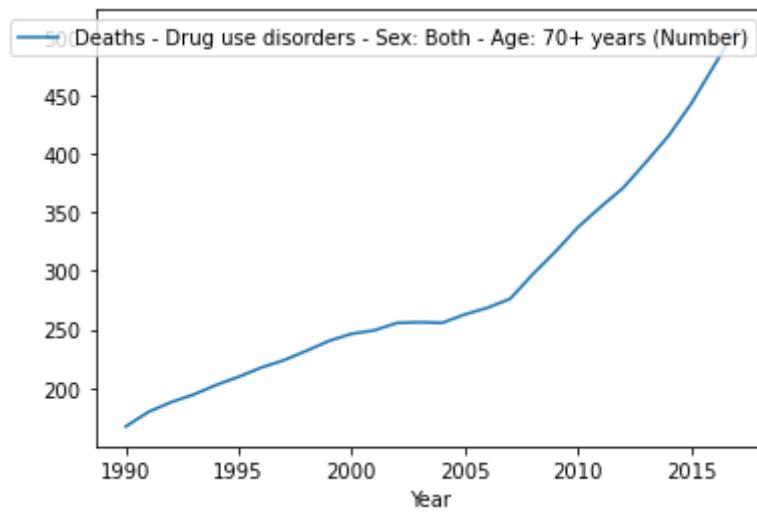
```
In [474]: df33.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[474]: <AxesSubplot:xlabel='Year'>
```



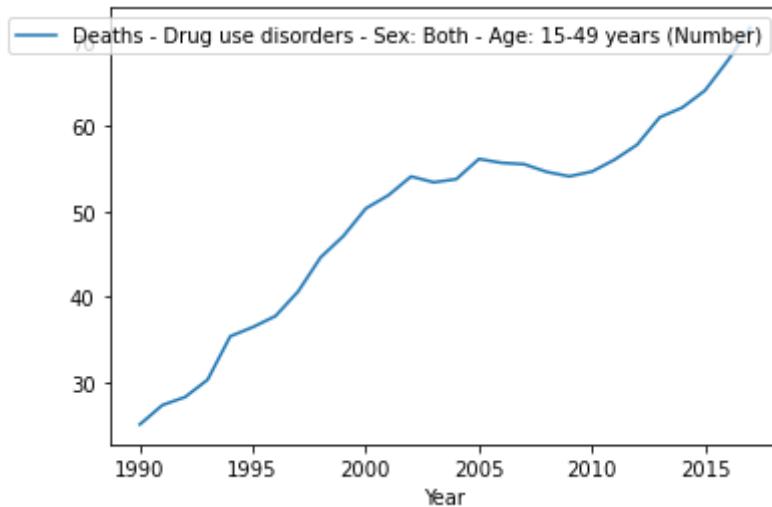
```
In [475]: df33.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[475]: <AxesSubplot:xlabel='Year'>
```



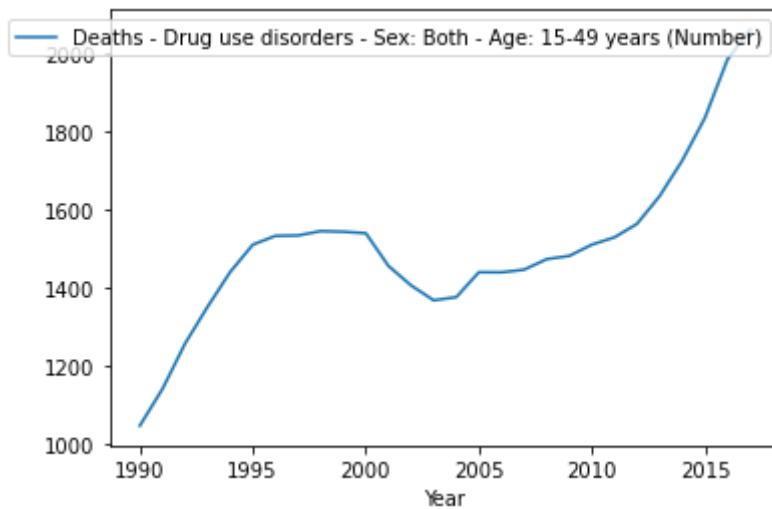
```
In [476]: df33.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[476]: <AxesSubplot:xlabel='Year'>
```



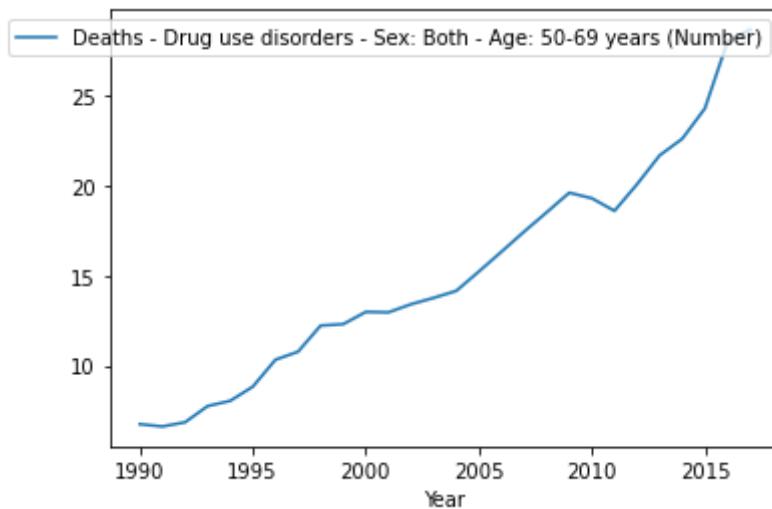
```
In [477]: df33.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[477]: <AxesSubplot:xlabel='Year'>
```



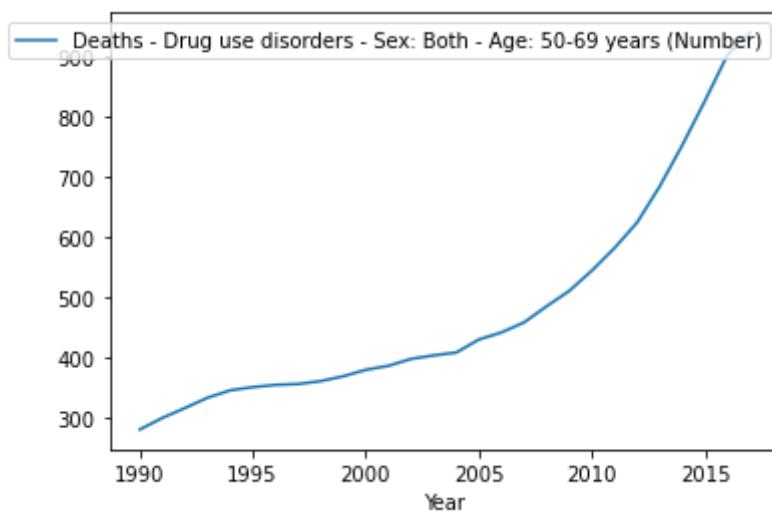
```
In [478]: df33.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[478]: <AxesSubplot:xlabel='Year'>
```



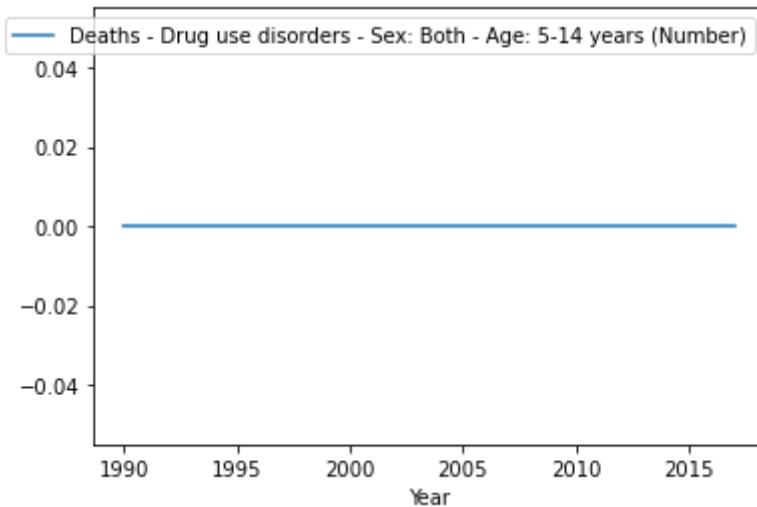
```
In [479]: df33.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[479]: <AxesSubplot:xlabel='Year'>
```



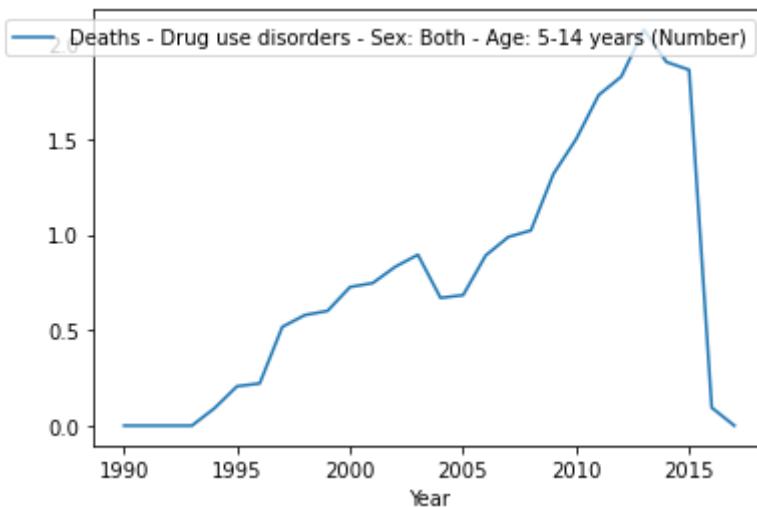
```
In [480]: df33.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[480]: <AxesSubplot:xlabel='Year'>
```



```
In [481]: df33.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[481]: <AxesSubplot:xlabel='Year'>
```



In [482]: df35.info()
df35.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column           Dtype      No
n-Null Count Dtype
---  --  -----  --
0   Entity        object      64
68  non-null     object      54
1   Code          object      54
88  non-null     object      64
2   Year          int64      64
68  non-null     float64    64
3   Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number) 64
68  non-null     float64    64
4   Deaths - Drug use - Sex: Both - Age: All Ages (Number)           64
68  non-null     float64    64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[482]:

	Entity	Code	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	1990	78.254192	208.325430
1	Afghanistan	AFG	1991	81.284586	217.769691
2	Afghanistan	AFG	1992	91.010197	247.833251
3	Afghanistan	AFG	1993	103.481991	285.036181
4	Afghanistan	AFG	1994	111.296397	306.646796

Checking for missing values:

In [483]: missing = pd.concat([df35.isnull().sum(), 100 * df35.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[483]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Drug use - Sex: Both - Age: All Ages (Number)		0	0.000000
Code	980	15.151515	

In [484]: v1='Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)'
v2='Deaths - Drug use - Sex: Both - Age: All Ages (Number)'

In [485]: df35.describe()

Out[485]:

	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
count	6468.000000	6468.000000	6468.000000
mean	2003.500000	2276.504819	8890.242150
std	8.078372	9663.540155	35415.115589
min	1990.000000	0.179388	1.240062
25%	1996.750000	17.214436	92.909932
50%	2003.500000	75.975217	408.586291
75%	2010.250000	460.087489	2170.843581
max	2017.000000	166612.548448	585348.180194

In [486]: df35.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[486]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	166612.548448
6382	World	2016	159690.500590
6381	World	2015	147614.117105
6380	World	2014	137888.912454
6379	World	2013	129359.881643

In [487]: df35.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[487]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
6383	World	2017	585348.180194
6382	World	2016	572922.966987
6381	World	2015	549156.843911
6380	World	2014	526924.912151
6379	World	2013	508271.643532

In [488]: `df35.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()`

Out[488]:

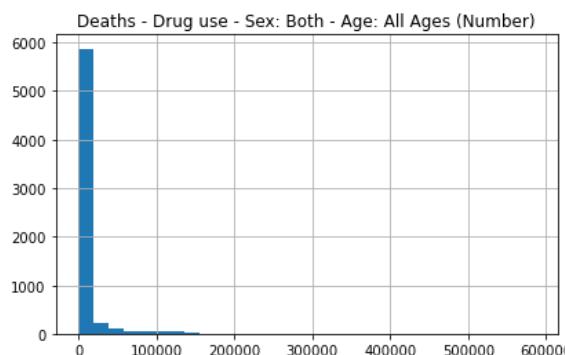
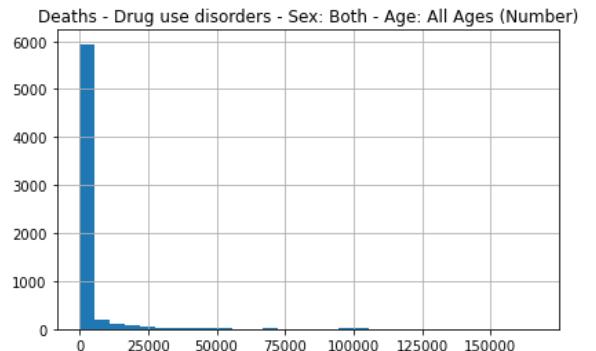
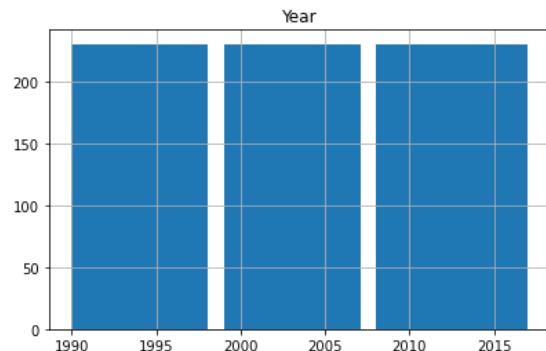
	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
4256	Northern Mariana Islands	1990	0.179388
4260	Northern Mariana Islands	1994	0.186184
4259	Northern Mariana Islands	1993	0.187125
4257	Northern Mariana Islands	1991	0.188218
4258	Northern Mariana Islands	1992	0.189438

In [489]: `df35.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()`

Out[489]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
84	American Samoa	1990	1.240062
85	American Samoa	1991	1.289510
86	American Samoa	1992	1.330492
87	American Samoa	1993	1.366904
88	American Samoa	1994	1.408779

In [490]: `df35.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [491]: df35.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[491]: Entity
Northern Mariana Islands      0.261072
American Samoa                0.343522
Kiribati                      0.434098
Tonga                          0.457850
Marshall Islands               0.464715
Name: Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [492]: df35.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[492]: Entity
American Samoa                 1.684194
Northern Mariana Islands        2.022616
Marshall Islands                2.563348
Antigua and Barbuda            3.102877
Dominica                        3.162885
Name: Deaths - Drug use - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [493]: df35.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[493]: Entity
East Asia                       26753.759737
Southeast Asia, East Asia, and Oceania 29992.877713
High-income                      37310.363509
High SDI                         37696.290110
World                            108139.994261
Name: Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [494]: df35.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[494]: Entity
High-middle SDI                  108734.845836
East Asia                        112786.815006
Middle SDI                       120070.912049
Southeast Asia, East Asia, and Oceania 143273.160750
World                            422808.510683
Name: Deaths - Drug use - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [495]: df35_mean = df35.groupby('Year').mean()  
df35_mean.head()
```

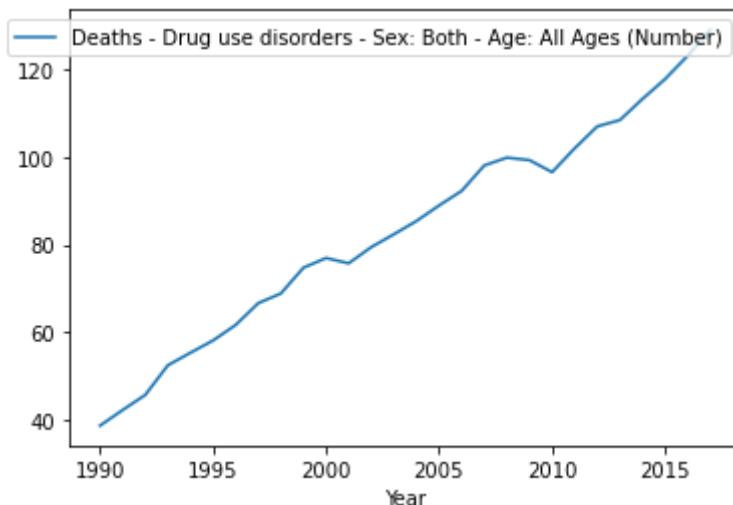
Out[495]:

	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
--	--	--

Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
1990	1495.473734	5296.548094
1991	1620.550501	5654.451944
1992	1761.082372	6066.396237
1993	1879.070576	6484.202048
1994	1988.431624	6885.332827

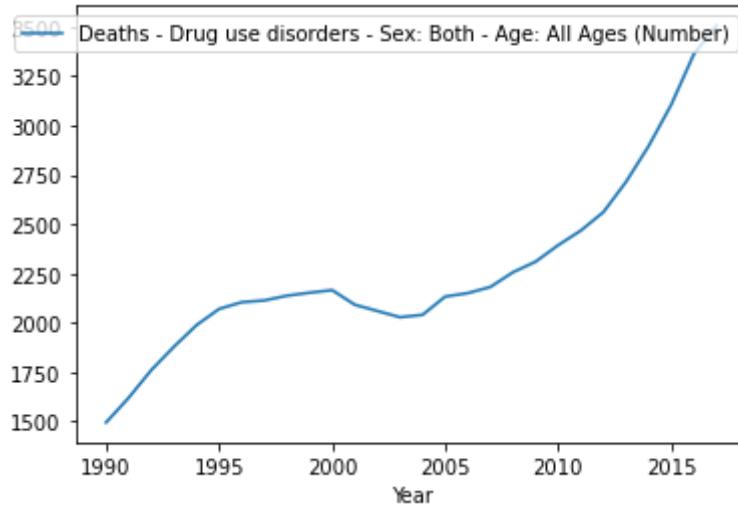
```
In [496]: df35.groupby('Year')[v1].median().plot(legend=True)
```

Out[496]: <AxesSubplot:xlabel='Year'>



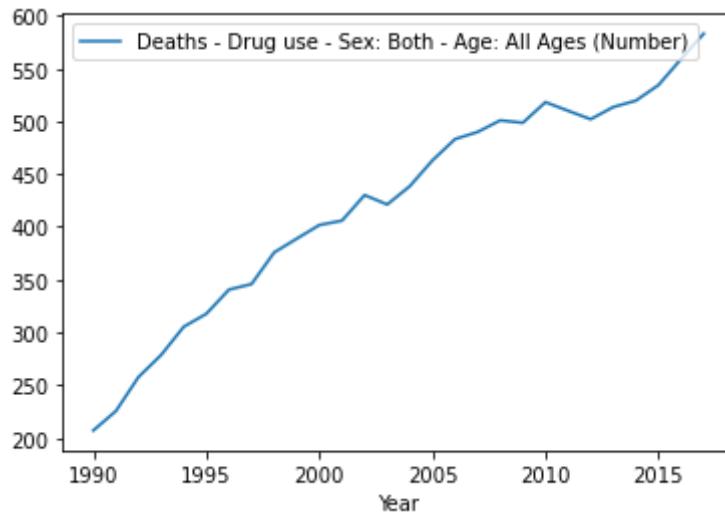
```
In [497]: df35.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[497]: <AxesSubplot:xlabel='Year'>
```



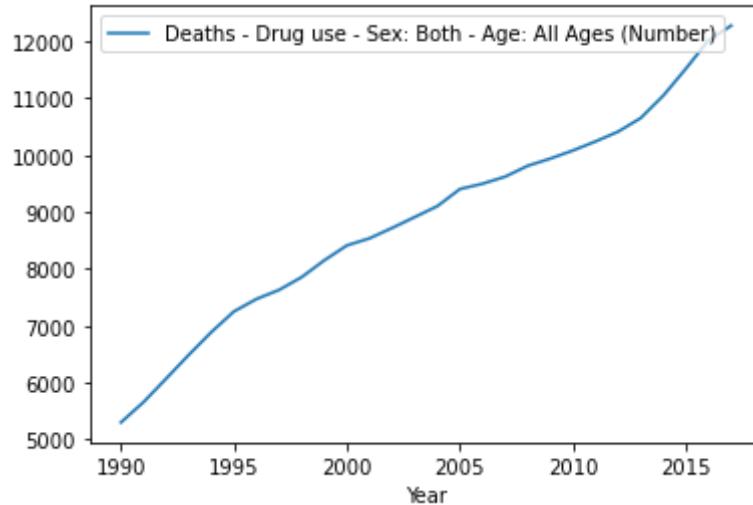
```
In [498]: df35.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[498]: <AxesSubplot:xlabel='Year'>
```



```
In [499]: df35.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[499]: <AxesSubplot:xlabel='Year'>
```



In [500]: df36.info()
df36.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 8 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number) 6468 non-null   float64
4   Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number) 6468 non-null   float64
5   Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number) 6468 non-null   float64
6   Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number) 6468 non-null   float64
7   Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number) 6468 non-null   float64
dtypes: float64(5), int64(1), object(2)
memory usage: 404.4+ KB
```

Out[500]:

	Entity	Code	Year	Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)	Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)	Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)	Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)	Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)
0	Afghanistan	AFG	1990	16.925449	58.650803	78.273161	0.0	0.0
1	Afghanistan	AFG	1991	17.339286	59.889574	80.165367	0.0	0.0
2	Afghanistan	AFG	1992	17.844703	62.307064	91.996741	0.0	0.0
3	Afghanistan	AFG	1993	18.432649	65.135370	107.428717	0.0	0.0
4	Afghanistan	AFG	1994	19.054541	67.597559	115.108447	0.0	0.0

Checking for missing values:

```
In [501]: missing = pd.concat([df36.isnull().sum(), 100 * df36.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[501]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)		0	0.000000
Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)		0	0.000000
Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)		0	0.000000
Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)		0	0.000000
Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)		0	0.000000
	Code	980	15.151515

```
In [502]: v1='Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)'
v2='Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)'
v3='Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)'
v4='Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)'
v5='Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)'
```

In [503]: df36.describe()

Out[503]:

	Year	Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)	Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)	Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)	Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)	Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	597.306265	1921.972944	3282.135316	0.182077	0.784290
std	8.078372	2404.890517	7819.519192	13068.563106	1.183375	4.546291
min	1990.000000	0.045373	0.105460	0.365676	0.000000	0.000000
25%	1996.750000	4.634170	14.582862	27.631791	0.000000	0.000000
50%	2003.500000	31.614097	80.516523	150.417466	0.000000	0.000000
75%	2010.250000	137.748142	379.426353	673.120478	0.000000	0.000000
max	2017.000000	45260.628019	126402.541203	179883.620987	28.358228	95.749609

In [504]: df36.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[504]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)
6383	World	2017	45260.628019
6382	World	2016	43128.098120
6381	World	2015	41090.284259
6380	World	2014	39278.908488
6379	World	2013	37733.826830

In [505]: df36.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[505]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)
6383	World	2017	126402.541203
6382	World	2016	123626.993378
6381	World	2015	118272.836504
6380	World	2014	111823.152772
6379	World	2013	106526.460182

In [506]: df36.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[506]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)
6383	World	2017	179883.620987
6382	World	2016	174610.149860
6371	World	2005	167234.863412
6381	World	2015	166518.816917
6366	World	2000	164387.517435

In [507]: df36.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[507]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)
6378	World	2012	28.358228
6379	World	2013	27.948511
6377	World	2011	22.404175
6380	World	2014	21.394403
6381	World	2015	18.649640

In [508]: df36.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[508]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)
6379	World	2013	95.749609
6380	World	2014	87.897372
6381	World	2015	86.015763
6378	World	2012	84.375237
6377	World	2011	79.935766

In [509]: df36.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[509]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)
3024	Kiribati	1990	0.045373
3025	Kiribati	1991	0.046642
3027	Kiribati	1993	0.046801
3028	Kiribati	1994	0.048055
3026	Kiribati	1992	0.048322

In [510]: df36.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[510]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)
3029	Kiribati	1995	0.105460
3030	Kiribati	1996	0.106399
3031	Kiribati	1997	0.107785
3028	Kiribati	1994	0.108320
3037	Kiribati	2003	0.108498

In [511]: df36.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[511]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)
84	American Samoa	1990	0.365676
85	American Samoa	1991	0.382705
86	American Samoa	1992	0.398108
5796	Tonga	1990	0.401382
105	American Samoa	2011	0.402750

In [512]: df36.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[512]:

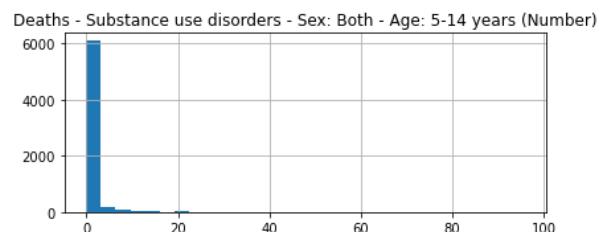
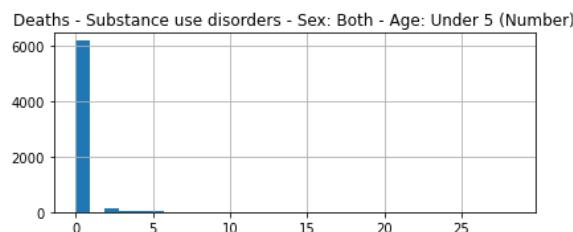
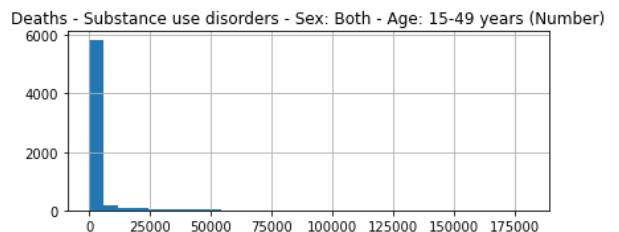
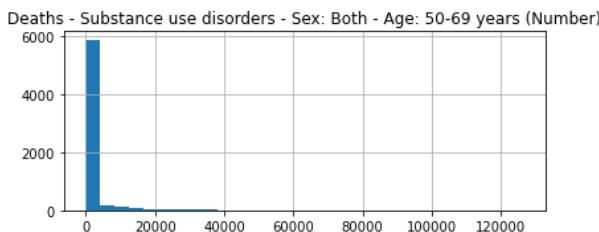
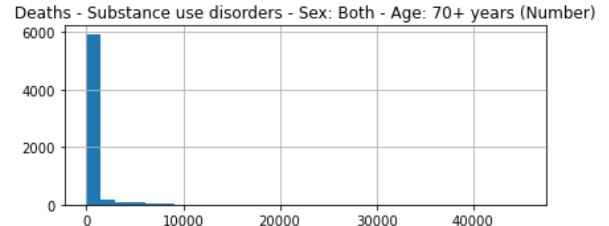
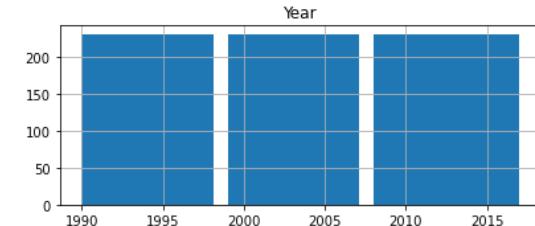
	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)
0	Afghanistan	1990	0.0
4245	Northern Ireland	2007	0.0
4244	Northern Ireland	2006	0.0
4243	Northern Ireland	2005	0.0
4242	Northern Ireland	2004	0.0

In [513]: df36.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[513]:

	Entity	Year	Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)
0	Afghanistan	1990	0.0
4228	Northern Ireland	1990	0.0
4227	North Macedonia	2017	0.0
4226	North Macedonia	2016	0.0
4225	North Macedonia	2015	0.0

```
In [514]: df36.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [515]: df36.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[515]: Entity
Kiribati          0.055865
Marshall Islands  0.063441
American Samoa   0.076141
Northern Mariana Islands  0.122557
Micronesia (country) 0.129831
Name: Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number), dtype: float64
```

```
In [516]: df36.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[516]: Entity
Kiribati          0.123493
American Samoa   0.157166
Marshall Islands 0.201897
Tonga            0.231243
Micronesia (country) 0.317522
Name: Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number), dtype: float64
```

```
In [517]: df36.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[517]: Entity
American Samoa   0.435521
Tonga            0.589623
Kiribati          0.594683
Marshall Islands 0.705991
Seychelles         0.902687
Name: Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number), dtype: float64
```

```
In [518]: df36.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[518]: Entity
Afghanistan      0.0
Namibia           0.0
Nepal             0.0
Netherlands       0.0
New Zealand       0.0
Name: Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number), dtype: float64
```

```
In [519]: df36.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[519]: Entity
Afghanistan      0.0
Namibia           0.0
Nepal             0.0
Netherlands       0.0
New Zealand       0.0
Name: Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number), dtype: float64
```

```
In [520]: df36.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[520]: Entity
High-middle SDI          7436.333600
High SDI                  7546.718095
High-income                7712.322167
Southeast Asia, East Asia, and Oceania 7876.922898
World                      28465.338983
Name: Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number), dtype: float64
```

```
In [521]: df36.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[521]: Entity
High SDI           24988.809273
Eastern Europe     27134.671380
Central Europe, Eastern Europe, and Central Asia 32421.042928
High-middle SDI    34789.012571
World              90987.684580
Name: Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number), dtype: float64
```

```
In [522]: df36.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[522]: Entity
Eastern Europe      39358.911158
High SDI            39923.540638
Central Europe, Eastern Europe, and Central Asia 45980.865045
High-middle SDI     54415.352014
World               155659.052940
Name: Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number), dtype: float64
```

```
In [523]: df36.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[523]: Entity
Middle SDI          3.175032
Western Europe       3.594348
High SDI             4.490094
High-income          5.091682
World                8.414478
Name: Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number), dtype: float64
```

```
In [524]: df36.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[524]: Entity
Western Europe       12.413323
Middle SDI           13.801849
High SDI             18.435608
High-income          21.803158
World                36.243902
Name: Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number), dtype: float64
```

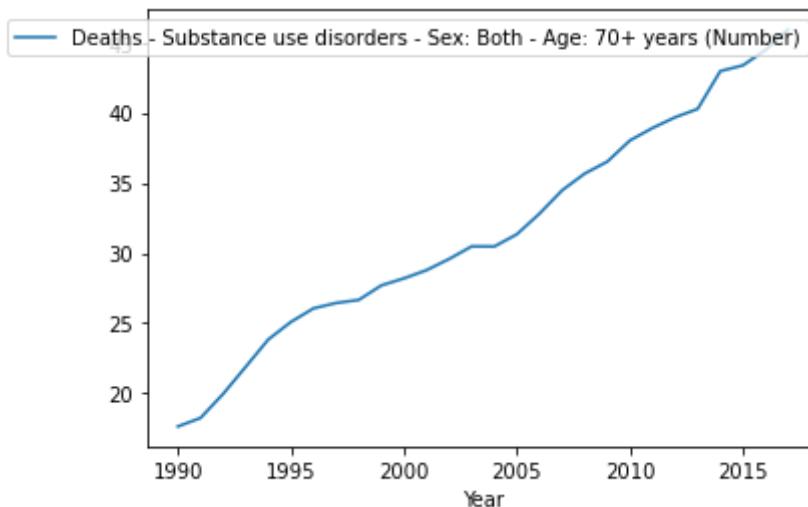
```
In [525]: df36_mean = df36.groupby('Year').mean()
df36_mean.head()
```

Out[525]:

Year	Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)	Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)	Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)	Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)	Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)
1990	351.354432	1250.185112	2269.897939	0.108632	0.000000
1991	371.542548	1326.240950	2440.641754	0.093136	0.000000
1992	389.422185	1439.762205	2706.706872	0.000000	0.000000
1993	407.415483	1598.723642	3035.804508	0.000000	0.000000
1994	425.011537	1699.033540	3317.573954	0.044126	0.091031

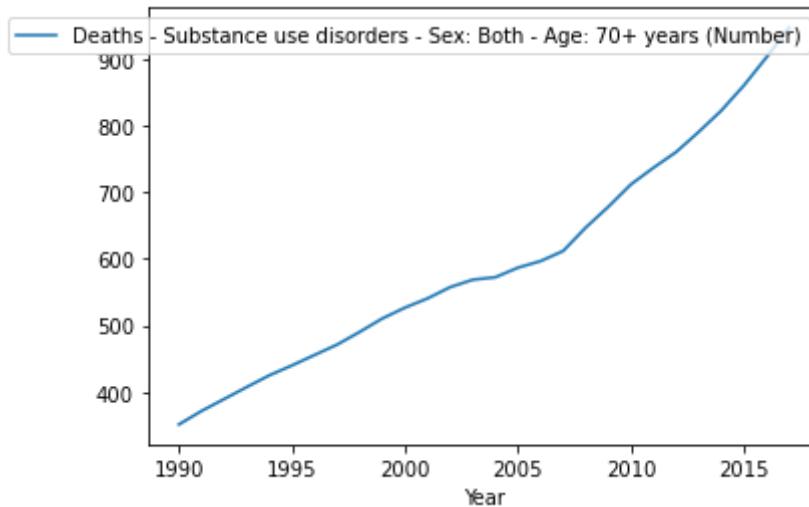
```
In [526]: df36.groupby('Year')[v1].median().plot(legend=True)
```

Out[526]: <AxesSubplot:xlabel='Year'>



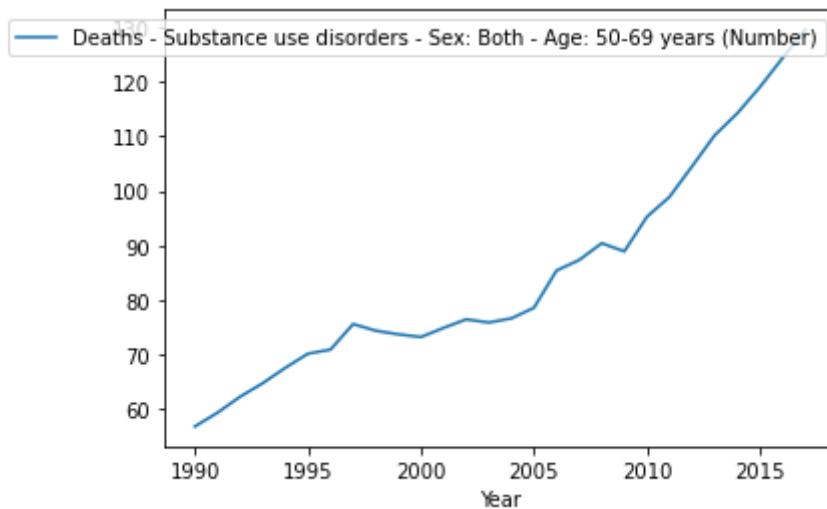
```
In [527]: df36.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[527]: <AxesSubplot:xlabel='Year'>
```



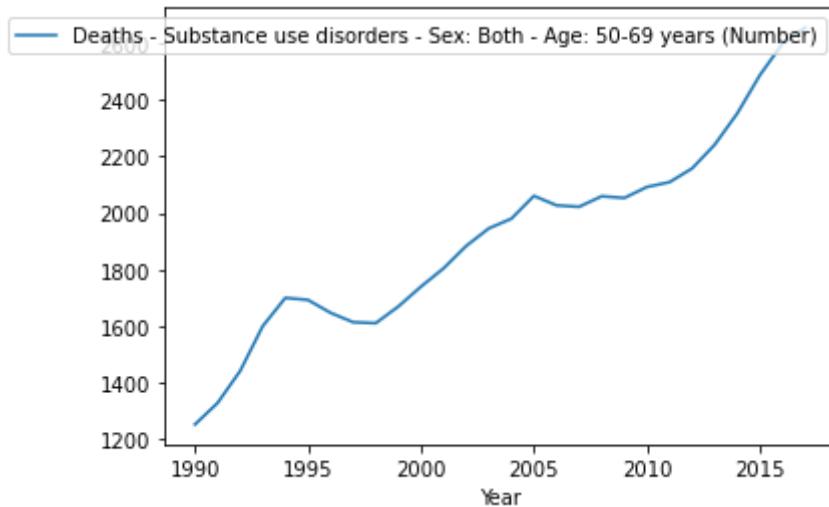
```
In [528]: df36.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[528]: <AxesSubplot:xlabel='Year'>
```



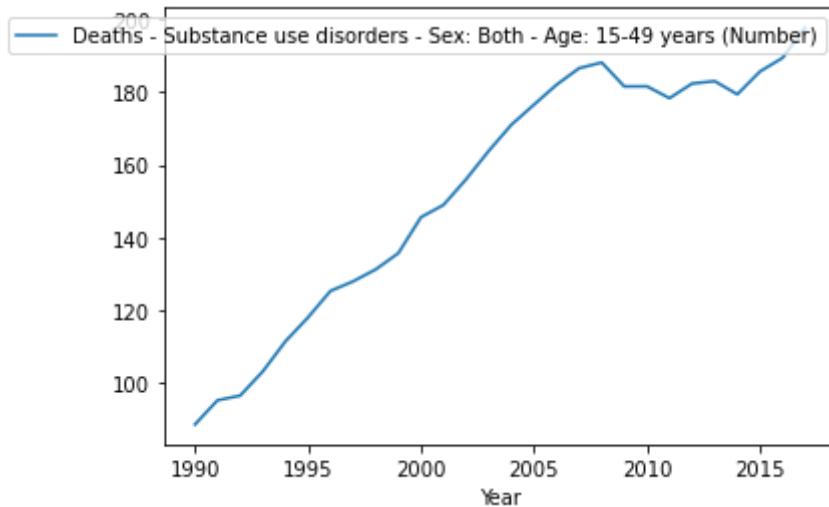
```
In [529]: df36.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[529]: <AxesSubplot:xlabel='Year'>
```



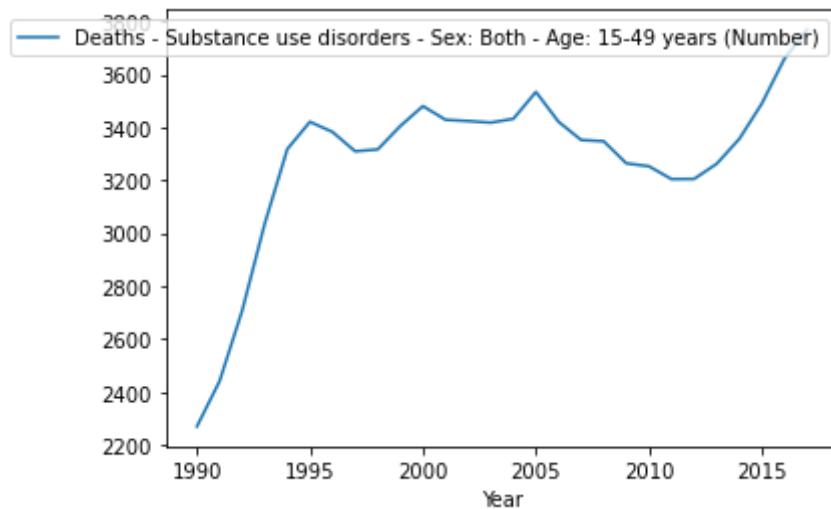
```
In [530]: df36.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[530]: <AxesSubplot:xlabel='Year'>
```



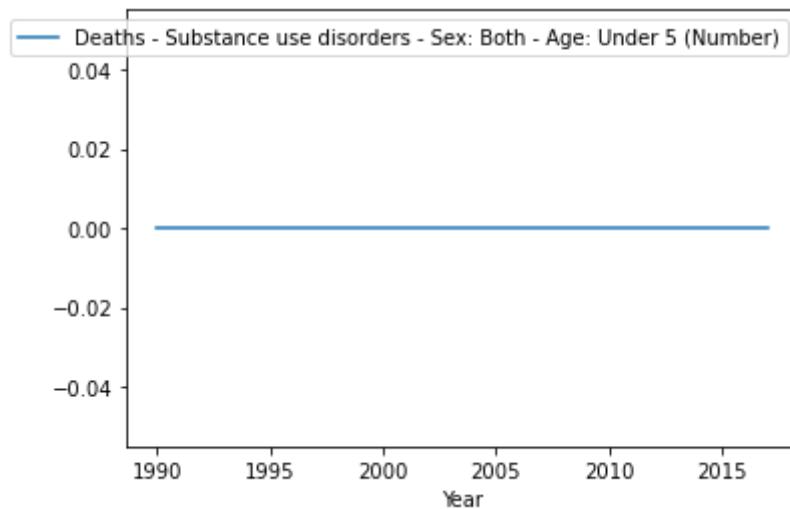
```
In [531]: df36.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[531]: <AxesSubplot:xlabel='Year'>
```



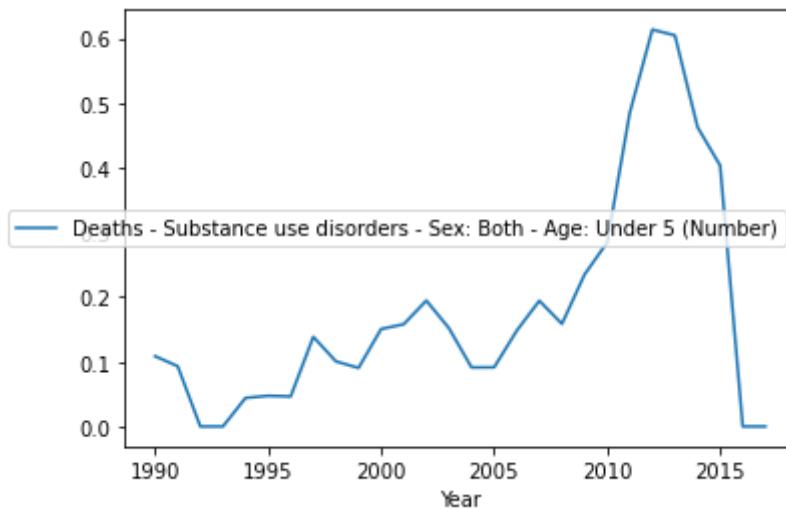
```
In [532]: df36.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[532]: <AxesSubplot:xlabel='Year'>
```



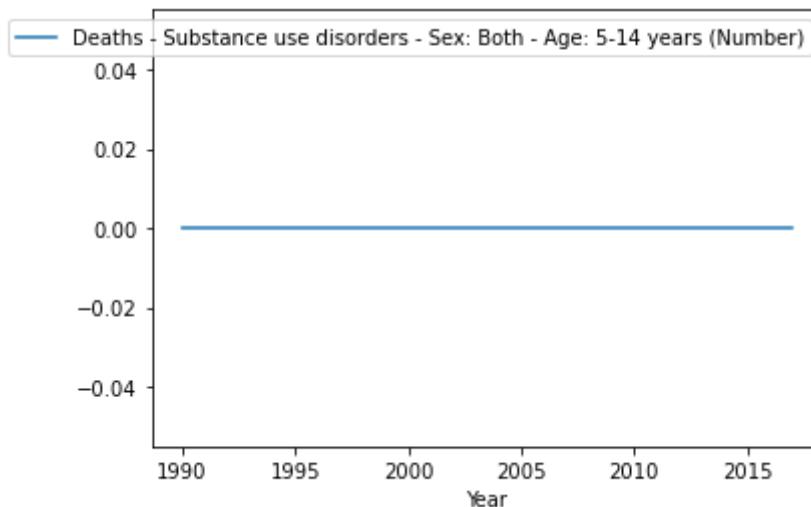
```
In [533]: df36.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[533]: <AxesSubplot:xlabel='Year'>
```



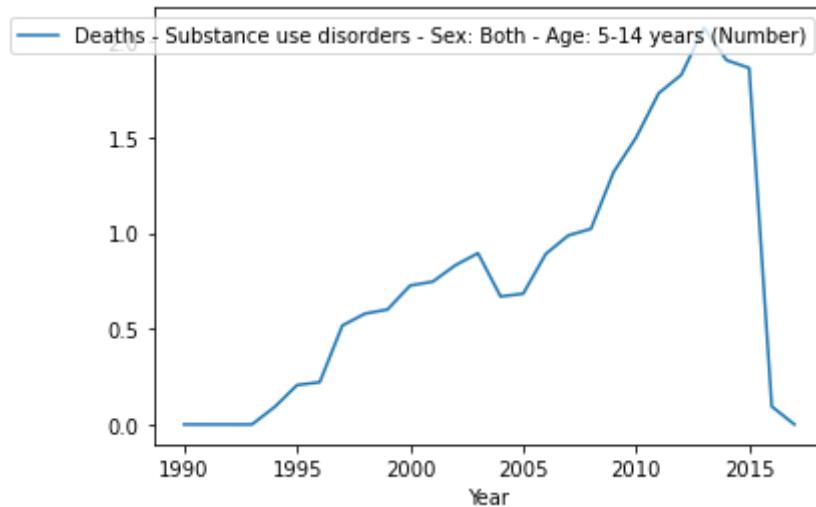
```
In [534]: df36.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[534]: <AxesSubplot:xlabel='Year'>
```



```
In [535]: df36.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[535]: <AxesSubplot:xlabel='Year'>
```



In [536]: df37.info()
df37.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 9 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year        int64
6468 non-null   int64
3   Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
4   Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
5   Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
6   Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
7   Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
8   Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
dtypes: float64(6), int64(1), object(2)
memory usage: 454.9+ KB
```

Out[536]:

	Entity	Code	Year	Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	1990	2.742366	78.254192	42.846584	75.595221	29.295752	3.369489
1	Afghanistan	AFG	1991	2.840492	81.284586	44.434209	76.109641	30.499450	3.510436
2	Afghanistan	AFG	1992	3.152032	91.010197	49.637945	81.138310	34.211945	4.008275
3	Afghanistan	AFG	1993	3.550867	103.481991	56.357910	87.514744	38.935044	4.638171
4	Afghanistan	AFG	1994	3.806900	111.296397	60.581453	90.464149	41.911427	4.996617

Checking for missing values:

```
In [537]: missing = pd.concat([df37.isnull().sum(), 100 * df37.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[537]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Code		980	15.151515

```
In [667]: v1='Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)'
v2='Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)'
v3='Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)'
v4='Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)'
v5='Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)'
v6='Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)'
```

In [668]: df37.describe()

Out[668]:

	Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)	Ampl
Year						Age:
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	96.482678	2276.504819	1357.032901	3525.876073	743.563462
std	8.078372	468.173836	9663.540155	5807.267010	14336.535999	3154.034909
min	1990.000000	0.001076	0.179388	0.076997	0.293634	0.072901
25%	1996.750000	0.341601	17.214436	9.749777	23.700635	4.299914
50%	2003.500000	2.099679	75.975217	48.095586	158.916099	22.448007
75%	2010.250000	11.581614	460.087489	291.475864	696.238005	134.544821
max	2017.000000	7286.568083	166612.548448	109519.748113	190843.575458	45269.630443

In [669]: df37.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[669]:

	Entity	Year	Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	7286.568083
6382	World	2016	7105.450902
6381	World	2015	6770.799542
6380	World	2014	6413.471099
6379	World	2013	6084.169645

In [670]: df37.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[670]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	166612.548448
6382	World	2016	159690.500590
6381	World	2015	147614.117105
6380	World	2014	137888.912454
6379	World	2013	129359.881643

In [671]: df37.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[671]:

	Entity	Year	Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	109519.748113
6382	World	2016	103961.489042
6381	World	2015	93950.981636
6380	World	2014	86622.112342
6379	World	2013	80222.164127

In [672]: df37.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[672]:

	Entity	Year	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
6371	World	2005	190843.575458
6370	World	2004	185851.759450
6383	World	2017	184934.241760
6369	World	2003	184025.573933
6372	World	2006	183932.669260

In [673]: df37.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[673]:

	Entity	Year	Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	45269.630443
6382	World	2016	44136.610854
6381	World	2015	42495.506609
6380	World	2014	40622.185093
6379	World	2013	38988.053252

In [674]: df37.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[674]:

	Entity	Year	Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	4536.601809
6382	World	2016	4486.949791
6381	World	2015	4396.829318
6380	World	2014	4231.143920
6379	World	2013	4065.494618

In [675]: df37.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[675]:

	Entity	Year	Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)
84	American Samoa	1990	0.001076
85	American Samoa	1991	0.001127
86	American Samoa	1992	0.001174
93	American Samoa	1999	0.001207
87	American Samoa	1993	0.001214

In [676]: df37.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[676]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
4256	Northern Mariana Islands	1990	0.179388
4260	Northern Mariana Islands	1994	0.186184
4259	Northern Mariana Islands	1993	0.187125
4257	Northern Mariana Islands	1991	0.188218
4258	Northern Mariana Islands	1992	0.189438

In [677]: df37.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[677]:

	Entity	Year	Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)
4256	Northern Mariana Islands	1990	0.076997
4260	Northern Mariana Islands	1994	0.080183
4259	Northern Mariana Islands	1993	0.080207
4257	Northern Mariana Islands	1991	0.080541
4258	Northern Mariana Islands	1992	0.081023

In [678]: df37.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[678]:

	Entity	Year	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
99	American Samoa	2005	0.293634
3027	Kiribati	1993	0.294713
84	American Samoa	1990	0.295547
98	American Samoa	2004	0.295837
3028	Kiribati	1994	0.301382

In [679]: df37.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[679]:

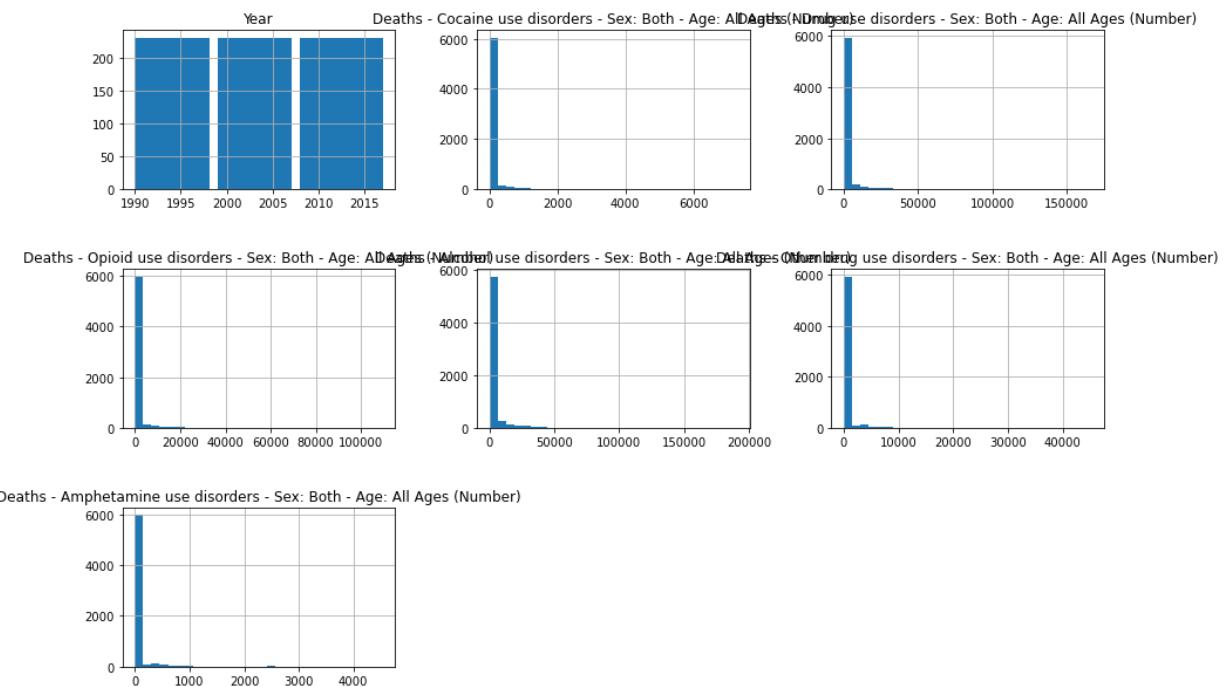
	Entity	Year	Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)
196	Antigua and Barbuda	1990	0.072901
4844	Sao Tome and Principe	1990	0.077110
197	Antigua and Barbuda	1991	0.079583
4845	Sao Tome and Principe	1991	0.079663
4788	Saint Vincent and the Grenadines	1990	0.081003

```
In [680]: df37.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()
```

Out[680]:

	Entity	Year	Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)
4788	Saint Vincent and the Grenadines	1990	0.001342
4260	Northern Mariana Islands	1994	0.001704
4259	Northern Mariana Islands	1993	0.001713
4258	Northern Mariana Islands	1992	0.001788
4261	Northern Mariana Islands	1995	0.001819

```
In [681]: df37.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [682]: df37.groupby('Entity')[v1].mean().sort_values().head()
```

Out[682]: Entity

American Samoa	0.001865
Northern Mariana Islands	0.002571
Tonga	0.002833
Marshall Islands	0.002931
Seychelles	0.003732

Name: Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number), dtype: float64

```
In [683]: df37.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[683]: Entity
Northern Mariana Islands      0.261072
American Samoa                 0.343522
Kiribati                        0.434098
Tonga                           0.457850
Marshall Islands                0.464715
Name: Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number), d
type: float64
```

```
In [684]: df37.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[684]: Entity
Northern Mariana Islands      0.119878
Kiribati                        0.135278
American Samoa                  0.145114
Tonga                            0.181176
Marshall Islands                 0.193154
Name: Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number),
dtype: float64
```

```
In [685]: df37.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[685]: Entity
American Samoa                  0.325306
Kiribati                         0.339944
Marshall Islands                  0.506614
Tonga                            0.554691
Maldives                          0.685093
Name: Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Numbe
r), dtype: float64
```

```
In [686]: df37.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[686]: Entity
Northern Mariana Islands      0.135722
American Samoa                  0.191850
Sao Tome and Principe          0.196049
Seychelles                       0.215136
Antigua and Barbuda            0.226404
Name: Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Numb
er), dtype: float64
```

```
In [687]: df37.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[687]: Entity
Northern Mariana Islands      0.002901
American Samoa                  0.004693
Marshall Islands                 0.009825
Tonga                            0.009913
Seychelles                        0.010399
Name: Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Num
ber), dtype: float64
```

```
In [688]: df37.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[688]: Entity
United States      2050.199750
North America     2146.621236
High SDI          2326.015470
High-income        2336.636464
World             4535.046951
Name: Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [689]: df37.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[689]: Entity
East Asia           26753.759737
Southeast Asia, East Asia, and Oceania 29992.877713
High-income         37310.363509
High SDI            37696.290110
World              108139.994261
Name: Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [690]: df37.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[690]: Entity
Southeast Asia, East Asia, and Oceania 16324.525623
High-middle SDI                  17062.604141
High-income                      21870.522787
High SDI                         22068.112637
World                           64471.735663
Name: Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [691]: df37.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[691]: Entity
Russia                41905.007733
Eastern Europe         56897.859931
Central Europe, Eastern Europe, and Central Asia 67698.268996
High-middle SDI       69974.087894
World                 167016.740623
Name: Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [692]: df37.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[692]: Entity
East Asia             10836.998504
Southeast Asia, East Asia, and Oceania 11862.461503
High-income           11870.932088
High SDI              12037.958481
World                 35375.641779
Name: Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [693]: df37.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[693]: Entity
High-income                                1232.272170
East Asia                                    1253.849568
High SDI                                     1264.203523
Southeast Asia, East Asia, and Oceania     1336.254760
World                                         3757.569868
Name: Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

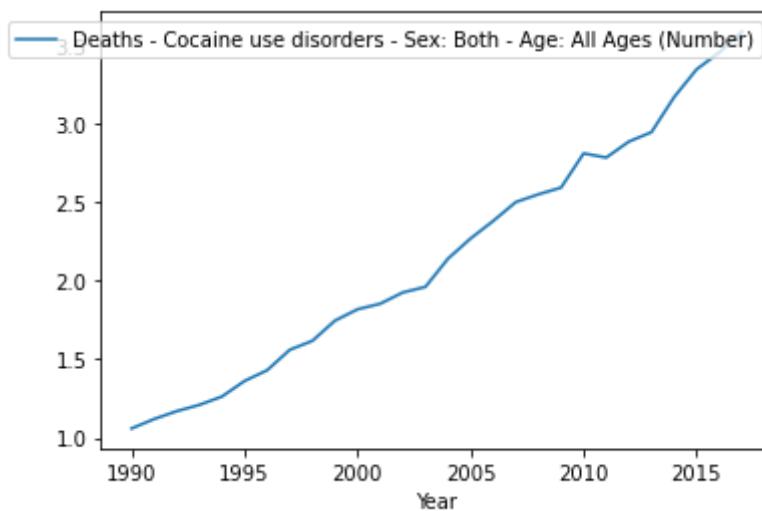
```
In [694]: df37_mean = df37.groupby('Year').mean()
df37_mean.head()
```

Out[694]:

	Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)
Year						
1990	45.947778	1495.473734	784.220940	2376.072381	596.704983	68.600032
1991	49.604421	1620.550501	861.587727	2517.967886	636.295478	73.062876
1992	54.461839	1761.082372	953.474242	2774.808891	675.100756	78.045536
1993	60.667033	1879.070576	1037.201292	3162.873057	700.458890	80.743360
1994	65.523250	1988.431624	1114.801117	3453.322565	725.267905	82.839352

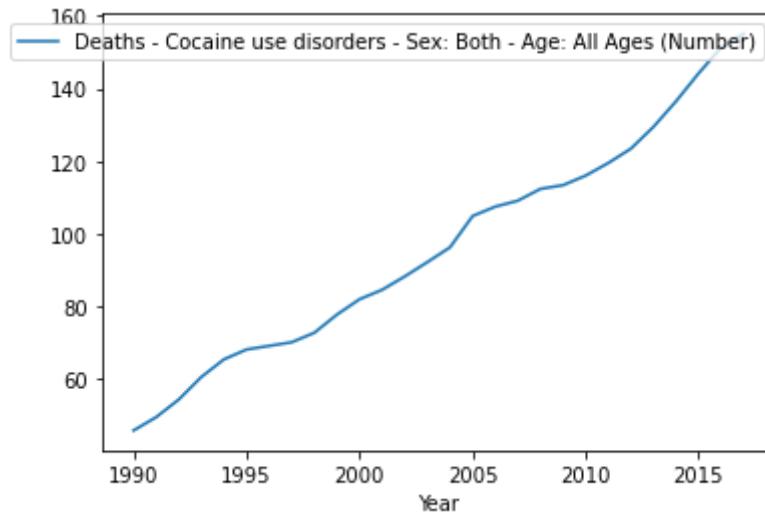
```
In [695]: df37.groupby('Year')[v1].median().plot(legend=True)
```

Out[695]: <AxesSubplot:xlabel='Year'>



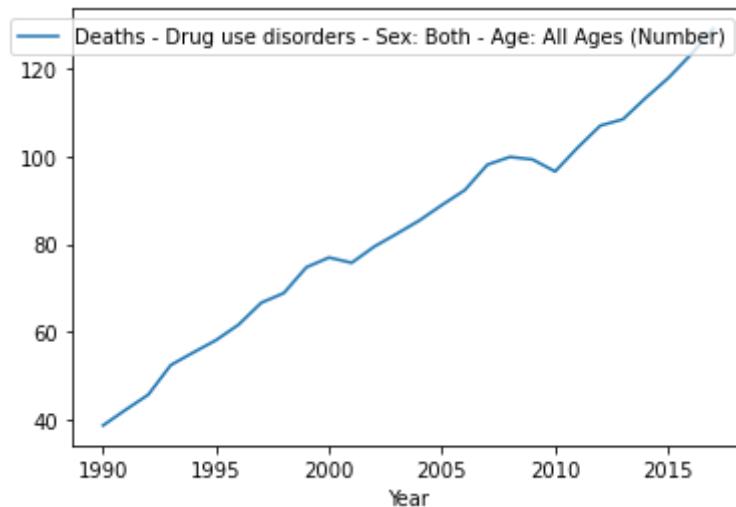
```
In [696]: df37.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[696]: <AxesSubplot:xlabel='Year'>
```



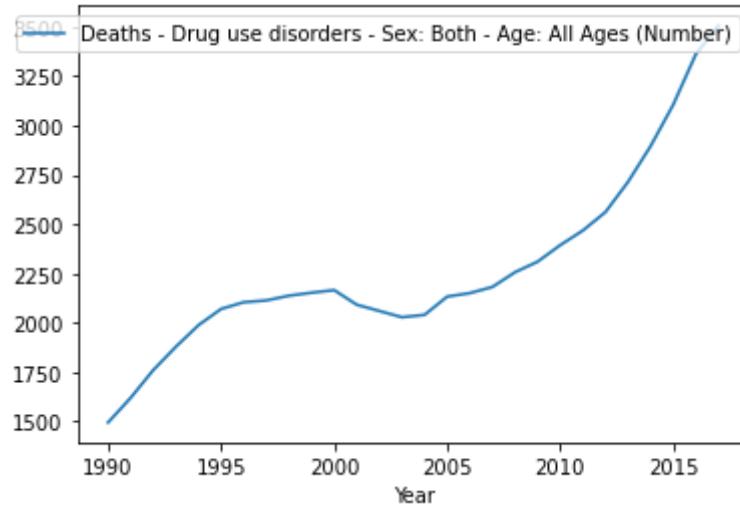
```
In [697]: df37.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[697]: <AxesSubplot:xlabel='Year'>
```



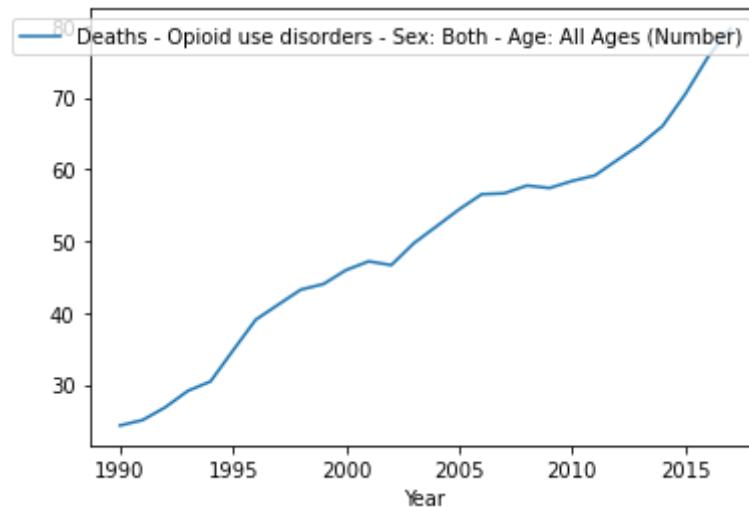
```
In [698]: df37.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[698]: <AxesSubplot:xlabel='Year'>
```



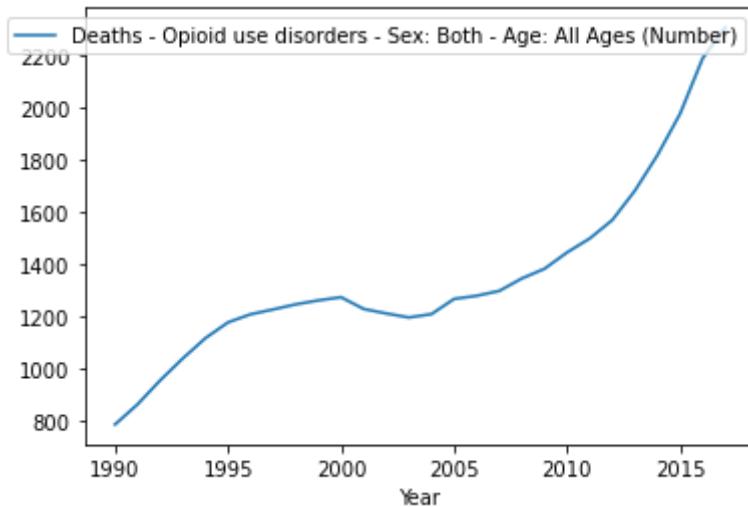
```
In [699]: df37.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[699]: <AxesSubplot:xlabel='Year'>
```



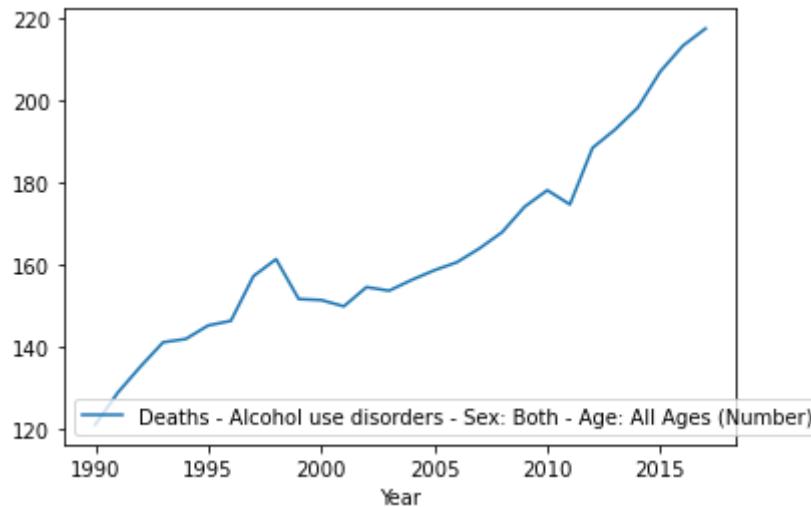
```
In [700]: df37.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[700]: <AxesSubplot:xlabel='Year'>
```



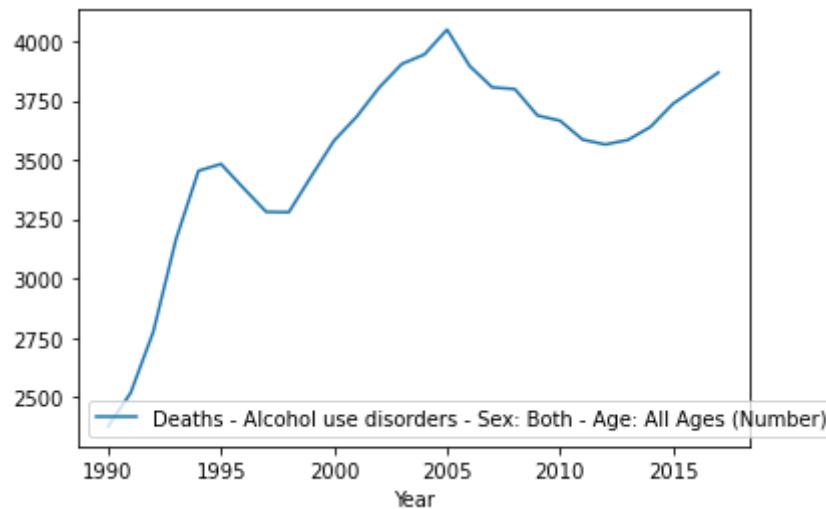
```
In [701]: df37.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[701]: <AxesSubplot:xlabel='Year'>
```



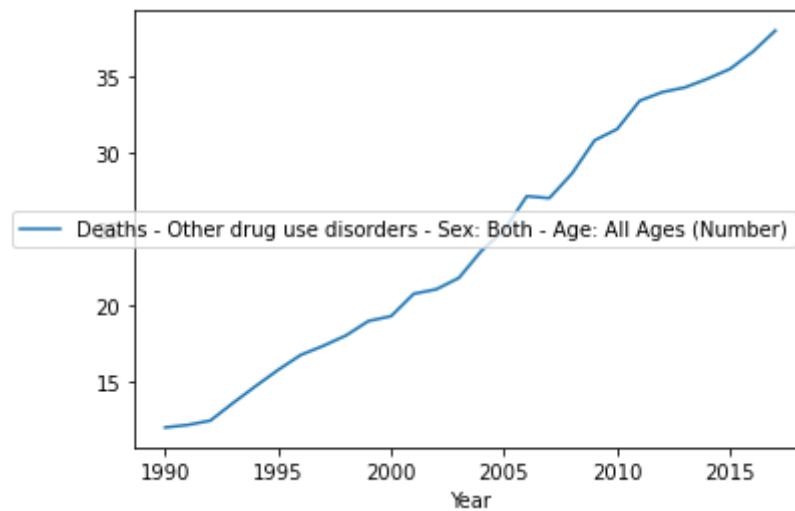
```
In [702]: df37.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[702]: <AxesSubplot:xlabel='Year'>
```



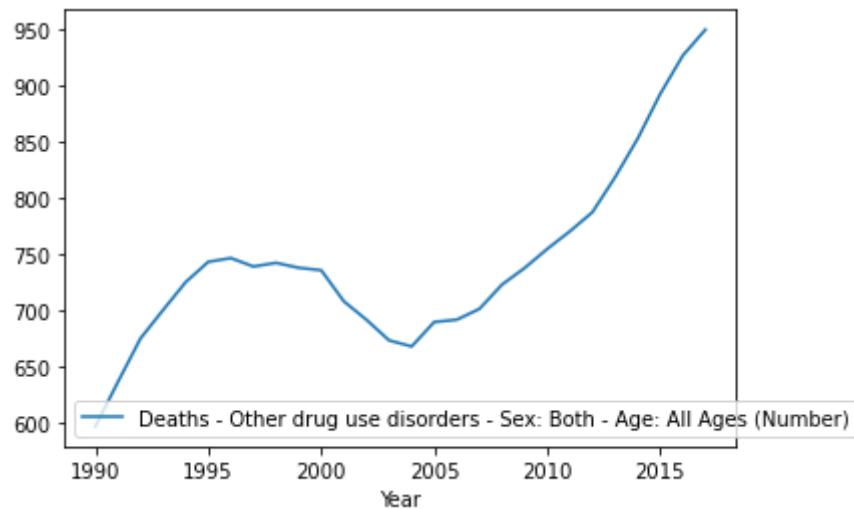
```
In [703]: df37.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[703]: <AxesSubplot:xlabel='Year'>
```



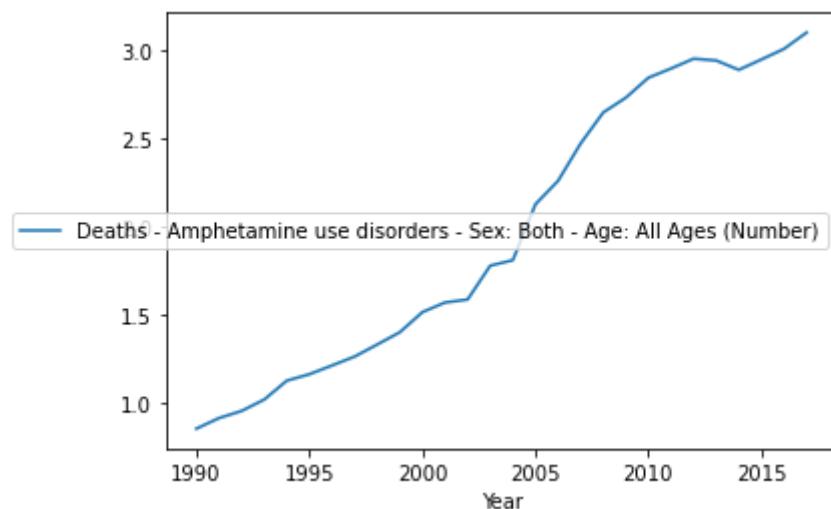
```
In [704]: df37.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[704]: <AxesSubplot:xlabel='Year'>
```



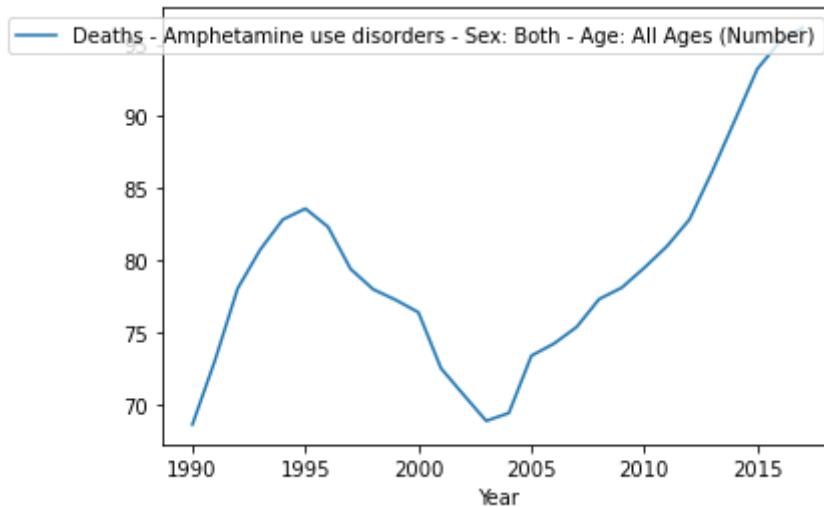
```
In [705]: df37.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[705]: <AxesSubplot:xlabel='Year'>
```



```
In [706]: df37.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[706]: <AxesSubplot:xlabel='Year'>
```



In [707]: df43.info()
df43.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate) 6468 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

Out[707]:

	Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)
0	Afghanistan	AFG	1990	648.307638
1	Afghanistan	AFG	1991	652.190016
2	Afghanistan	AFG	1992	657.044222
3	Afghanistan	AFG	1993	663.458236
4	Afghanistan	AFG	1994	670.067282

Checking for missing values:

In [708]: missing = pd.concat([df43.isnull().sum(), 100 * df43.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[708]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

In [709]: v1='DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both'

In [710]: df43.describe()

Out[710]:

	Year	DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)
count	6468.000000	6468.000000
mean	2003.500000	277.953956
std	8.078372	186.725298
min	1990.000000	79.640005
25%	1996.750000	164.182889
50%	2003.500000	207.336095
75%	2010.250000	306.419986
max	2017.000000	1695.546545

In [711]: df43.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[711]:

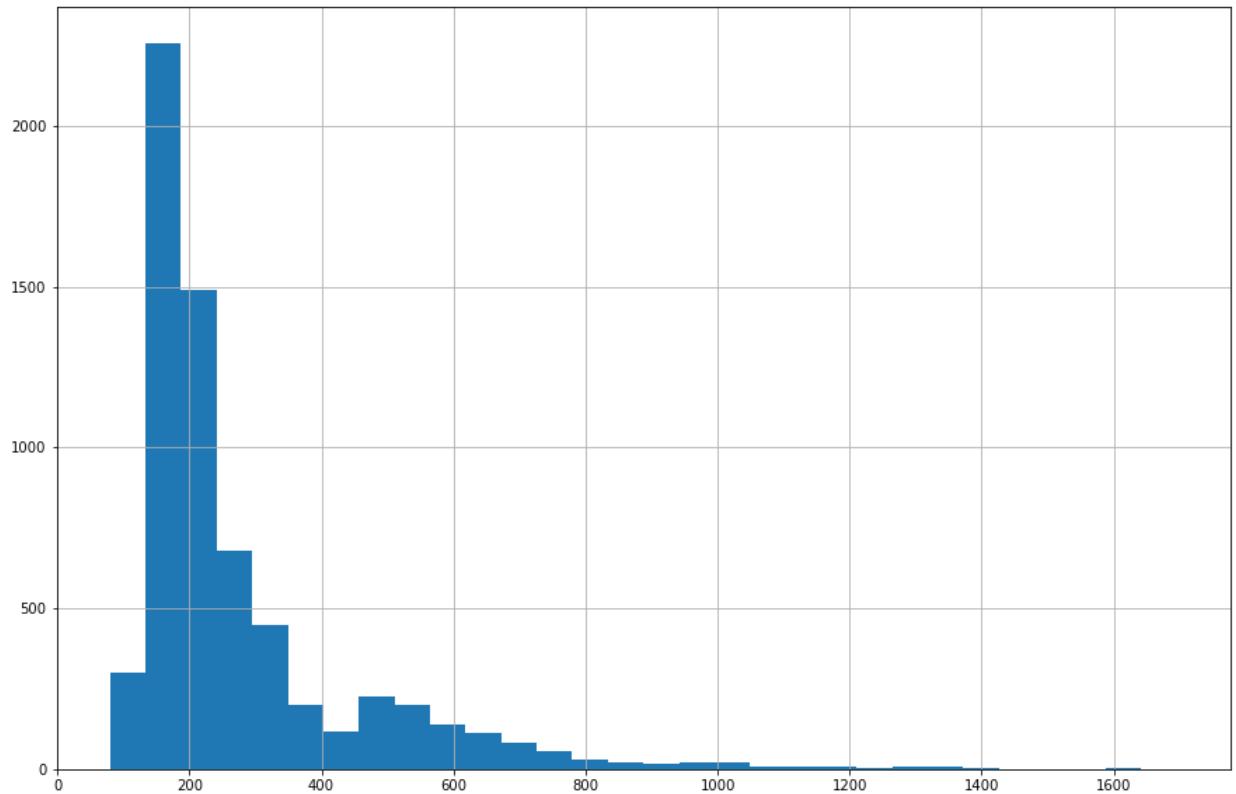
	Entity	Year	DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)
6103	United States	2017	1695.546545
6102	United States	2016	1628.835006
4171	North America	2017	1602.145320
4170	North America	2016	1540.513156
6101	United States	2015	1488.321404

In [712]: df43.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[712]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)
705	Bosnia and Herzegovina	1995	79.640005
704	Bosnia and Herzegovina	1994	79.862127
706	Bosnia and Herzegovina	1996	79.896646
700	Bosnia and Herzegovina	1990	79.963525
703	Bosnia and Herzegovina	1993	80.026679

```
In [713]: df43[v1].hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [714]: df43.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[714]: Entity  
Bosnia and Herzegovina    107.323850  
Tajikistan                 111.170866  
North Macedonia            116.450724  
Azerbaijan                118.874351  
Albania                    120.136349  
Name: DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex:  
Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [715]: df43.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[715]: Entity  
Afghanistan                848.281767  
North America               919.931894  
United States               958.021918  
Libya                      1044.978681  
United Arab Emirates        1169.042857  
Name: DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex:  
Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [716]: df43_mean = df43.groupby('Year').mean()  
df43_mean.head()
```

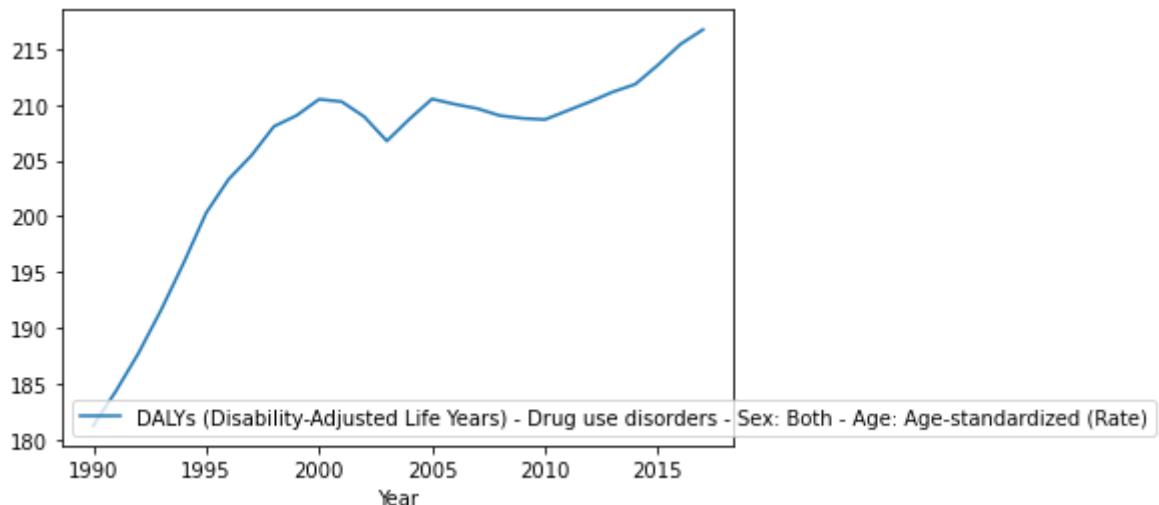
Out[716]:

DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)

Year	
1990	240.934276
1991	244.961484
1992	249.437576
1993	253.926167
1994	259.122120

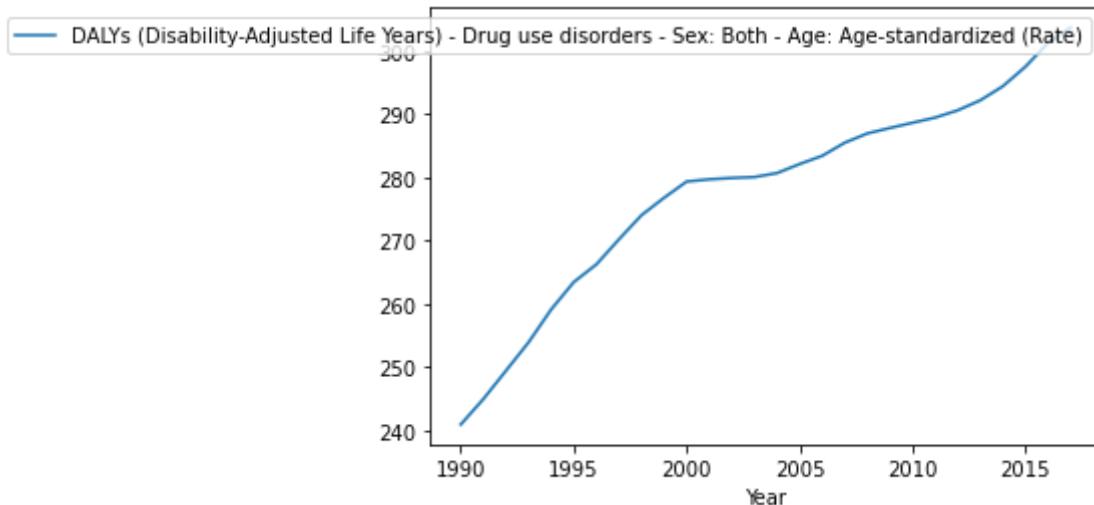
```
In [717]: df43.groupby('Year')[v1].median().plot(legend=True)
```

Out[717]: <AxesSubplot:xlabel='Year'>



```
In [718]: df43.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[718]: <AxesSubplot:xlabel='Year'>
```



In [719]: df80.info()
df80.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 13 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)
6468 non-null   float64
4   Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)
6468 non-null   float64
5   Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)
6468 non-null   float64
6   Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)
6468 non-null   float64
7   Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
8   Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)
6468 non-null   float64
9   Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)
6468 non-null   float64
10  Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent)
6468 non-null   float64
11  Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent)
6468 non-null   float64
12  Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)
6468 non-null   float64
dtypes: float64(10), int64(1), object(2)
memory usage: 657.0+ KB
```

Out[719]:

	Entity	Code	Year	Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)
0	Afghanistan	AFG	1990	3.784742	0.028011	1.033674	4.359556	1.213738	3.428878				
1	Afghanistan	AFG	1991	3.797962	0.028178	1.038626	4.391779	1.248699	3.438238				
2	Afghanistan	AFG	1992	3.820028	0.027875	1.055584	4.455005	1.347775	3.441508				
3	Afghanistan	AFG	1993	3.832659	0.028033	1.067455	4.508078	1.448685	3.462818				
4	Afghanistan	AFG	1994	3.864735	0.028460	1.070848	4.526207	1.482176	3.485138				

Checking for missing values:

```
In [720]: missing = pd.concat([df80.isnull().sum(), 100 * df80.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[720]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Code	980	15.151515	

```
In [721]: v1='Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)'
v2='Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)'
v3='Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)'
v4='Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)'
v5='Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)'
v6='Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)'
v7='Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)'
v8='Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent)'
v9='Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent)'
v10='Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)'
```

In [722]: df80.describe()

Out[722]:

	Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	1.587379	0.080175	1.454589	2.042085	0.863315
std	8.078372	0.922457	0.069691	1.054978	1.082283	0.507226
min	1990.000000	0.699521	0.000000	0.292252	0.881301	0.302919
25%	1996.750000	1.032626	0.031090	0.789712	1.349699	0.523956
50%	2003.500000	1.261302	0.059132	1.295034	1.697789	0.736383
75%	2010.250000	1.633889	0.104953	1.699037	2.125954	0.971229
max	2017.000000	6.928637	0.427862	7.926805	8.067019	5.338159

In [723]: df80.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[723]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)
6039	United Arab Emirates	2009	6.928637
6040	United Arab Emirates	2010	6.910006
6041	United Arab Emirates	2011	6.822064
6038	United Arab Emirates	2008	6.748892
6042	United Arab Emirates	2012	6.716159

In [724]: df80.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[724]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)
962	Canada	2000	0.427862
963	Canada	2001	0.426063
964	Canada	2002	0.420079
4007	New Zealand	1993	0.419368
4008	New Zealand	1994	0.418564

In [725]: df80.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[725]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)
6088	United States	2002	7.926805
6087	United States	2001	7.924069
6089	United States	2003	7.906434
6086	United States	2000	7.879789
6090	United States	2004	7.879623

In [726]: df80.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[726]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)
6045	United Arab Emirates	2015	8.067019
6044	United Arab Emirates	2014	8.028973
6046	United Arab Emirates	2016	8.000135
6043	United Arab Emirates	2013	7.930709
6047	United Arab Emirates	2017	7.816739

In [727]: df80.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[727]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)
6040	United Arab Emirates	2010	5.338159
6039	United Arab Emirates	2009	5.313234
6041	United Arab Emirates	2011	5.281151
6042	United Arab Emirates	2012	5.191838
6043	United Arab Emirates	2013	5.084227

In [728]: df80.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[728]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)
6103	United States	2017	8.569555
6102	United States	2016	8.378083
4171	North America	2017	8.326586
6101	United States	2015	8.211388
4170	North America	2016	8.148014

In [729]: df80.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[729]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)
6047	United Arab Emirates	2017	0.794887
3301	Libya	2015	0.780872
3302	Libya	2016	0.779940
3300	Libya	2014	0.779289
3299	Libya	2013	0.776962

In [730]: df80.sort_values(by=v8, ascending=False)[['Entity', 'Year', v8]].head()

Out[730]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent)
4005	New Zealand	1991	0.279673
4004	New Zealand	1990	0.277661
4006	New Zealand	1992	0.276968
4007	New Zealand	1993	0.276243
4008	New Zealand	1994	0.272873

In [731]: `df80.sort_values(by=v9, ascending=False)[['Entity', 'Year', v9]].head()`

Out[731]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent)
6040	United Arab Emirates	2010	6.141347
6041	United Arab Emirates	2011	6.102868
6039	United Arab Emirates	2009	6.093978
6042	United Arab Emirates	2012	6.027535
6043	United Arab Emirates	2013	5.931516

In [732]: `df80.sort_values(by=v10, ascending=False)[['Entity', 'Year', v10]].head()`

Out[732]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)
6103	United States	2017	3.452476
6102	United States	2016	3.353142
4171	North America	2017	3.338079
6101	United States	2015	3.262674
4170	North America	2016	3.245661

In [733]: `df80.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()`

Out[733]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)
705	Bosnia and Herzegovina	1995	0.699521
704	Bosnia and Herzegovina	1994	0.700788
706	Bosnia and Herzegovina	1996	0.701970
703	Bosnia and Herzegovina	1993	0.707128
702	Bosnia and Herzegovina	1992	0.707658

In [734]: df80.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[734]:

Entity Year Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)			
857	Burkina Faso	2007	0.0
5788	Togo	2010	0.0
5787	Togo	2009	0.0
5786	Togo	2008	0.0
5784	Togo	2006	0.0

In [735]: df80.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[735]:

Entity Year Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)			
5783	Togo	2005	0.292252
5785	Togo	2007	0.295484
5782	Togo	2004	0.295711
5784	Togo	2006	0.295724
5786	Togo	2008	0.297020

In [736]: df80.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[736]:

Entity Year Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)			
705	Bosnia and Herzegovina	1995	0.881301
706	Bosnia and Herzegovina	1996	0.883196
704	Bosnia and Herzegovina	1994	0.886102
703	Bosnia and Herzegovina	1993	0.888016
702	Bosnia and Herzegovina	1992	0.895539

In [737]: df80.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[737]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)
840	Burkina Faso	1990	0.302919
841	Burkina Faso	1991	0.305380
842	Burkina Faso	1992	0.308274
843	Burkina Faso	1993	0.310690
844	Burkina Faso	1994	0.312770

In [738]: df80.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[738]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)
706	Bosnia and Herzegovina	1996	0.950768
704	Bosnia and Herzegovina	1994	0.951189
705	Bosnia and Herzegovina	1995	0.953087
840	Burkina Faso	1990	0.954195
703	Bosnia and Herzegovina	1993	0.954413

In [739]: df80.sort_values(by=v6, ascending=True)[['Entity', 'Year', v7]].head()

Out[739]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)
706	Bosnia and Herzegovina	1996	0.113944
704	Bosnia and Herzegovina	1994	0.113811
705	Bosnia and Herzegovina	1995	0.113717
840	Burkina Faso	1990	0.151741
703	Bosnia and Herzegovina	1993	0.113585

In [740]: `df80.sort_values(by=v6, ascending=True)[['Entity', 'Year', v8]].head()`

Out[740]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent)
706	Bosnia and Herzegovina	1996	0.035053
704	Bosnia and Herzegovina	1994	0.033592
705	Bosnia and Herzegovina	1995	0.034728
840	Burkina Faso	1990	0.004509
703	Bosnia and Herzegovina	1993	0.033382

In [741]: `df80.sort_values(by=v6, ascending=True)[['Entity', 'Year', v9]].head()`

Out[741]:

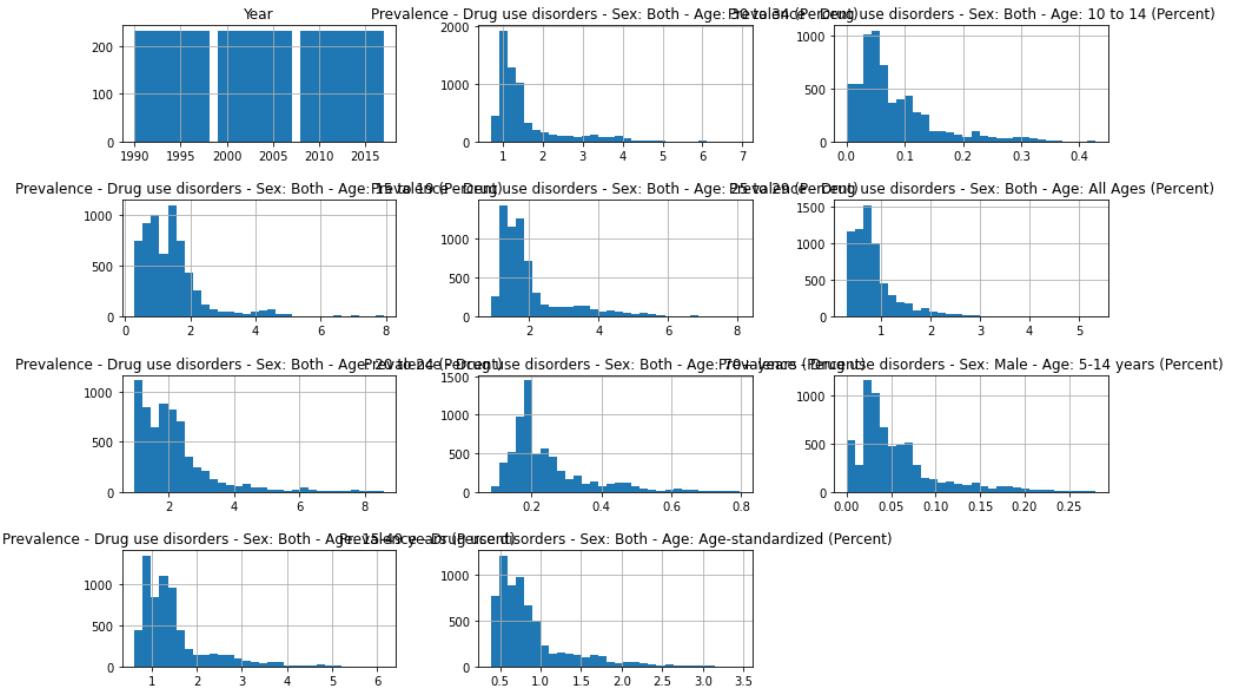
	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent)
706	Bosnia and Herzegovina	1996	0.621801
704	Bosnia and Herzegovina	1994	0.645085
705	Bosnia and Herzegovina	1995	0.630895
840	Burkina Faso	1990	0.683421
703	Bosnia and Herzegovina	1993	0.655700

In [742]: `df80.sort_values(by=v6, ascending=True)[['Entity', 'Year', v10]].head()`

Out[742]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)
706	Bosnia and Herzegovina	1996	0.384598
704	Bosnia and Herzegovina	1994	0.384059
705	Bosnia and Herzegovina	1995	0.383650
840	Burkina Faso	1990	0.396884
703	Bosnia and Herzegovina	1993	0.384720

```
In [743]: df80.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [744]: df80.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[744]: Entity
Bosnia and Herzegovina      0.776101
Tajikistan                   0.823659
Romania                      0.826098
Uzbekistan                   0.828751
Azerbaijan                   0.831396
Name: Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent),
      dtype: float64
```

```
In [745]: df80.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[745]: Entity
Togo                         0.000000
Nigeria                      0.001622
Burkina Faso                 0.002226
Western Sub-Saharan Africa    0.004536
Chad                          0.006330
Name: Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent),
      dtype: float64
```

```
In [746]: df80.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[746]: Entity
Burkina Faso    0.316065
Togo            0.316528
Nigeria         0.339107
Mali             0.356710
Guinea          0.358029
Name: Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [747]: df80.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[747]: Entity
Bosnia and Herzegovina  0.979473
Romania                 1.015709
Serbia                  1.065739
Tajikistan               1.066498
Uzbekistan              1.075598
Name: Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [748]: df80.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[748]: Entity
Burkina Faso    0.335646
Chad            0.340930
Mali             0.345423
Niger           0.345886
Guinea          0.355164
Name: Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [749]: df80.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[749]: Entity
Bangladesh      0.992401
Burkina Faso    0.993519
Bosnia and Herzegovina 1.022297
Mali             1.027031
Guinea          1.034778
Name: Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [750]: df80.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[750]: Entity
Sweden          0.091747
Greece          0.106946
Cambodia        0.107837
Finland         0.112863
Indonesia       0.117286
Name: Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [751]: df80.groupby('Entity')[v8].mean().sort_values().head()
```

```
Out[751]: Entity
Togo           0.000000
Nigeria        0.001482
Burkina Faso   0.001980
Western Sub-Saharan Africa  0.003987
Niger          0.005408
Name: Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent), dtype: float64
```

```
In [752]: df80.groupby('Entity')[v9].mean().sort_values().head()
```

```
Out[752]: Entity
Bosnia and Herzegovina  0.689244
Burkina Faso            0.721092
Mali                    0.738038
Romania                 0.739358
Guinea                  0.746699
Name: Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [753]: df80.groupby('Entity')[v10].mean().sort_values().head()
```

```
Out[753]: Entity
Bosnia and Herzegovina  0.412833
Burkina Faso            0.416361
Mali                    0.426686
Guinea                  0.429939
Chad                    0.437192
Name: Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [754]: df80.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[754]: Entity
Lebanon          4.269439
Iran             4.421749
Afghanistan     4.923217
Libya            4.931532
United Arab Emirates 5.935201
Name: Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [755]: df80.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[755]: Entity
England          0.295334
United States    0.303209
North America    0.306025
Canada           0.333189
New Zealand      0.339683
Name: Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [756]: df80.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[756]: Entity
New Zealand      4.445701
Greenland        4.818326
Canada           5.920938
North America    6.920693
United States    7.022303
Name: Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [757]: df80.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[757]: Entity
Libya            5.437177
North America    5.525339
United States    5.664498
Afghanistan      5.691282
United Arab Emirates 6.739712
Name: Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [758]: df80.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[758]: Entity
Libya            2.326631
North America    2.660094
Qatar            2.680388
United States    2.729106
United Arab Emirates 3.940635
Name: Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [759]: df80.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[759]: Entity
Australia        5.892120
Australasia      5.908292
New Zealand       5.988269
North America     7.199276
United States     7.348218
Name: Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [760]: df80.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[760]: Entity
North America     0.623778
United States     0.640730
Afghanistan       0.661034
United Arab Emirates 0.697102
Libya             0.713026
Name: Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [761]: df80.groupby('Entity')[v8].mean().sort_values().tail()
```

```
Out[761]: Entity
United Kingdom      0.189884
North America       0.189890
England             0.193582
Canada              0.207720
New Zealand         0.231954
Name: Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent), dtype: float64
```

```
In [762]: df80.groupby('Entity')[v9].mean().sort_values().tail()
```

```
Out[762]: Entity
Australia           3.626713
Libya                3.715900
North America        4.460450
United States        4.577161
United Arab Emirates 4.984197
Name: Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [763]: df80.groupby('Entity')[v10].mean().sort_values().tail()
```

```
Out[763]: Entity
Australia           2.194212
Libya                2.197783
United Arab Emirates 2.631126
North America        2.777057
United States        2.844713
Name: Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

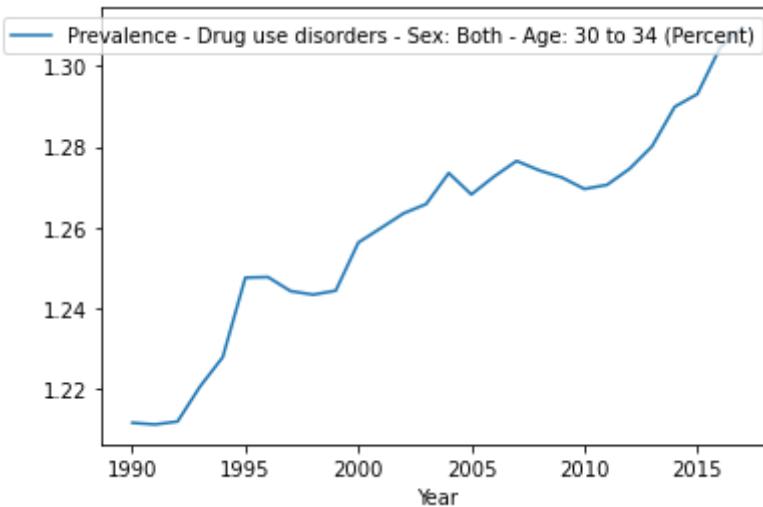
```
In [764]: df80_mean = df80.groupby('Year').mean()
df80_mean.head()
```

```
Out[764]:
```

	Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Drug use disorders - Sex: Ma Age: 5 ye (Percent)
Year								
1990	1.484877	0.077483	1.386676	1.918698	0.797671	2.003030	0.231364	0.0511
1991	1.493782	0.078057	1.392662	1.930814	0.802892	2.015536	0.232966	0.0511
1992	1.502803	0.078342	1.399945	1.943115	0.808145	2.027647	0.234460	0.0521
1993	1.512082	0.078749	1.408162	1.955019	0.813257	2.039161	0.235773	0.0521
1994	1.520564	0.079179	1.416090	1.965379	0.817822	2.048523	0.236760	0.0521

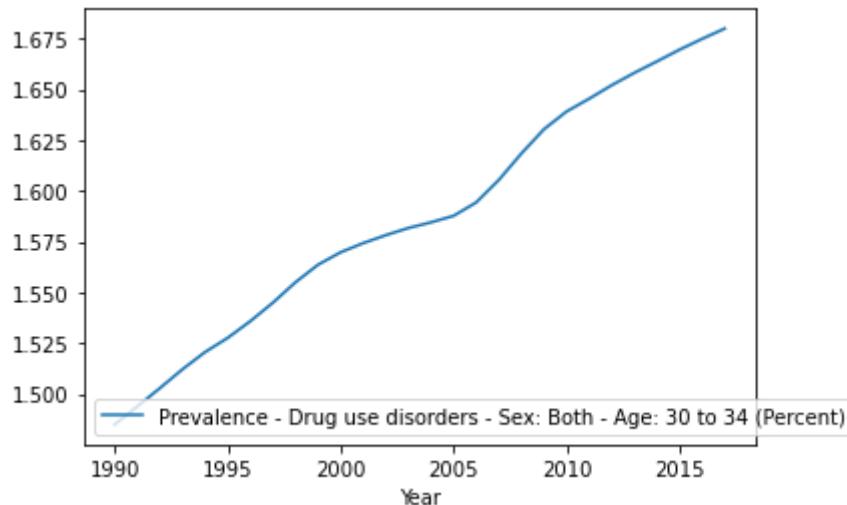
```
In [765]: df80.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[765]: <AxesSubplot:xlabel='Year'>
```



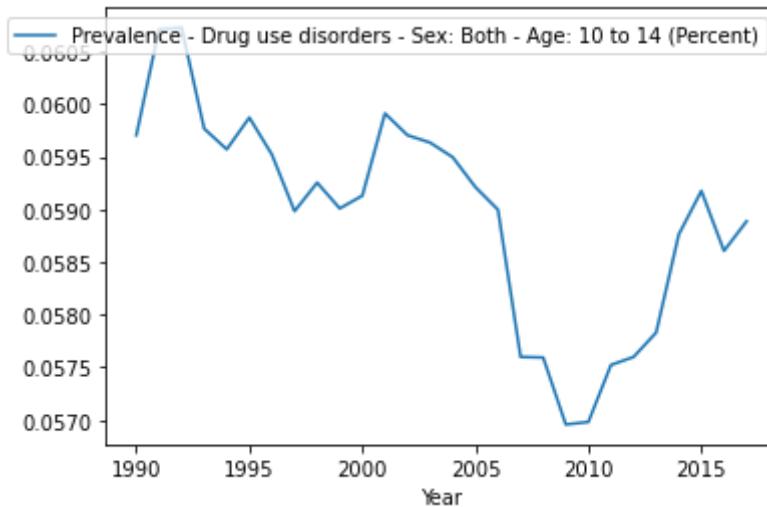
```
In [766]: df80.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[766]: <AxesSubplot:xlabel='Year'>
```



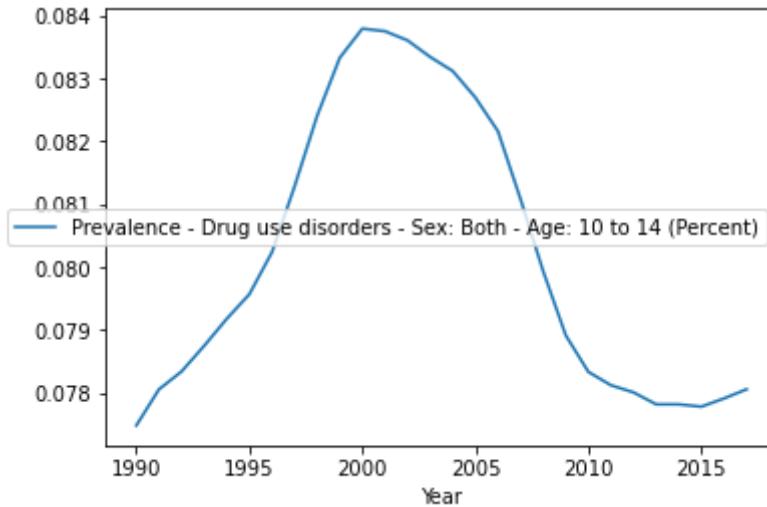
```
In [767]: df80.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[767]: <AxesSubplot:xlabel='Year'>
```



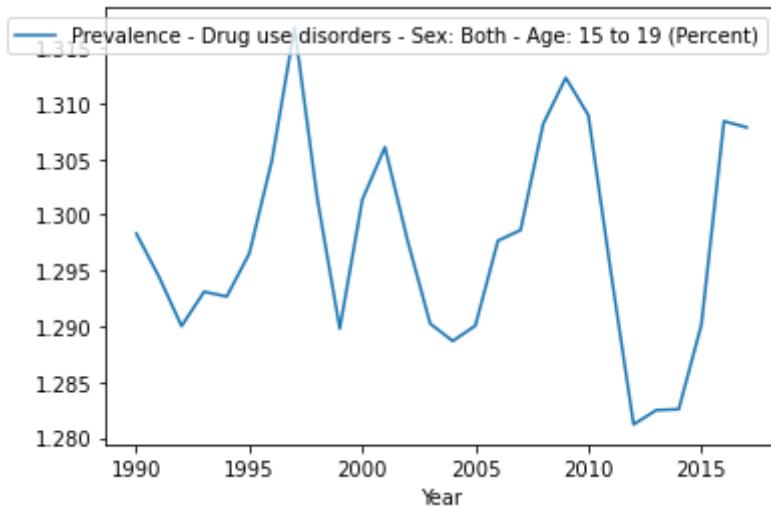
```
In [768]: df80.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[768]: <AxesSubplot:xlabel='Year'>
```



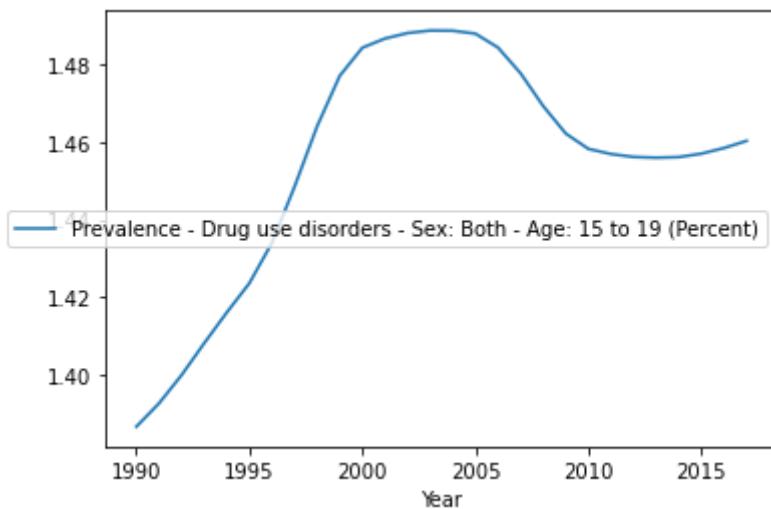
```
In [769]: df80.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[769]: <AxesSubplot:xlabel='Year'>
```



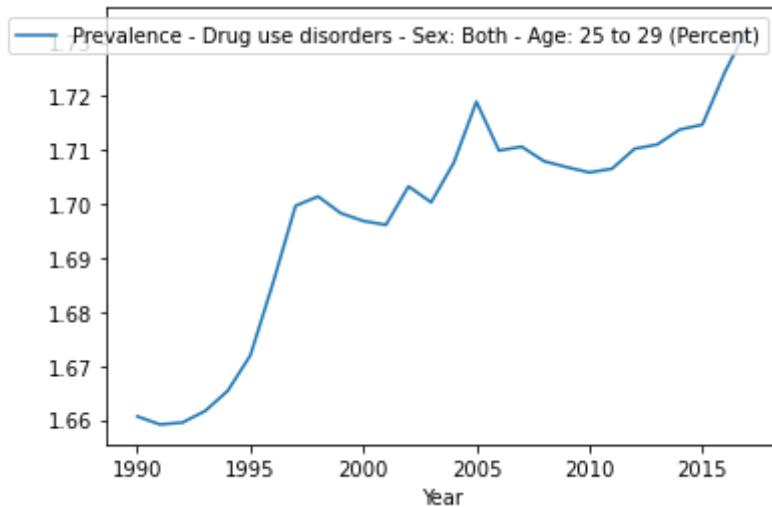
```
In [770]: df80.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[770]: <AxesSubplot:xlabel='Year'>
```



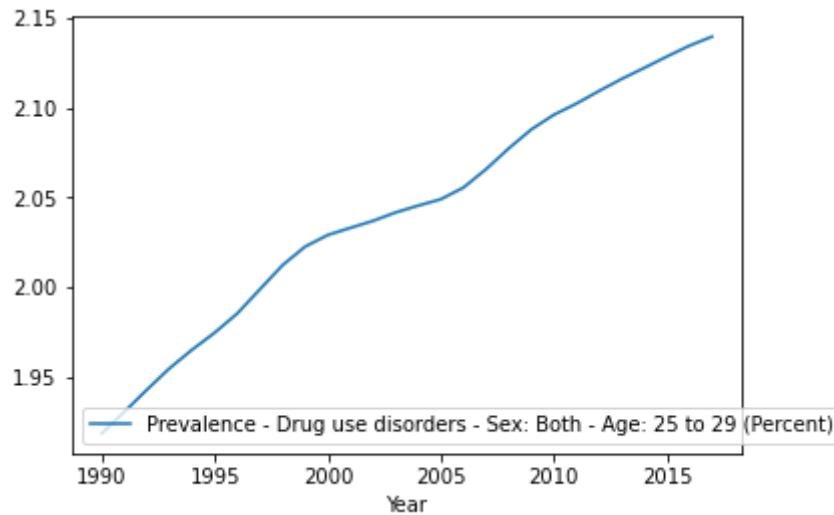
```
In [771]: df80.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[771]: <AxesSubplot:xlabel='Year'>
```



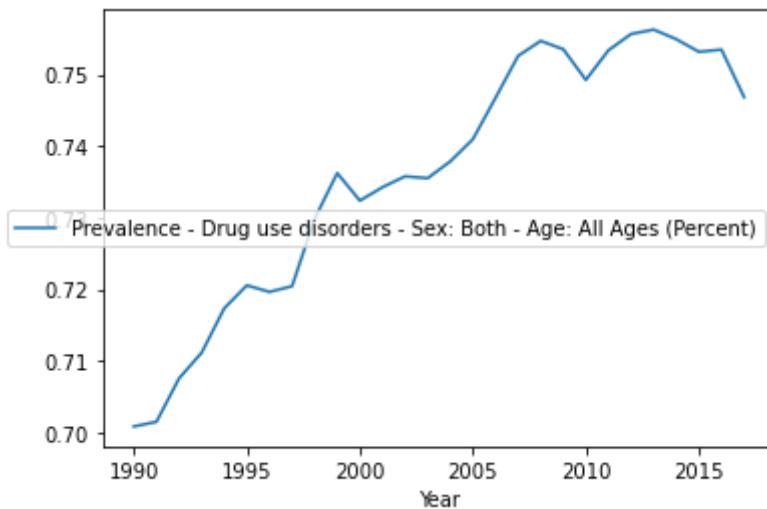
```
In [772]: df80.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[772]: <AxesSubplot:xlabel='Year'>
```



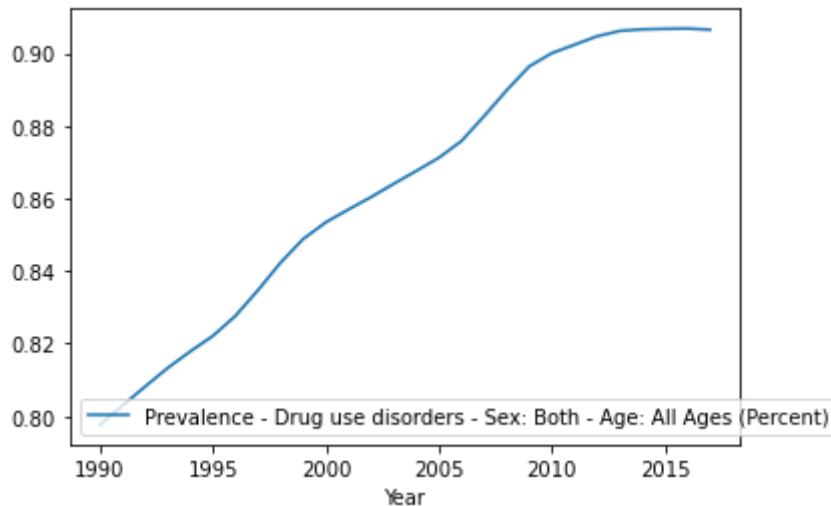
```
In [773]: df80.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[773]: <AxesSubplot:xlabel='Year'>
```



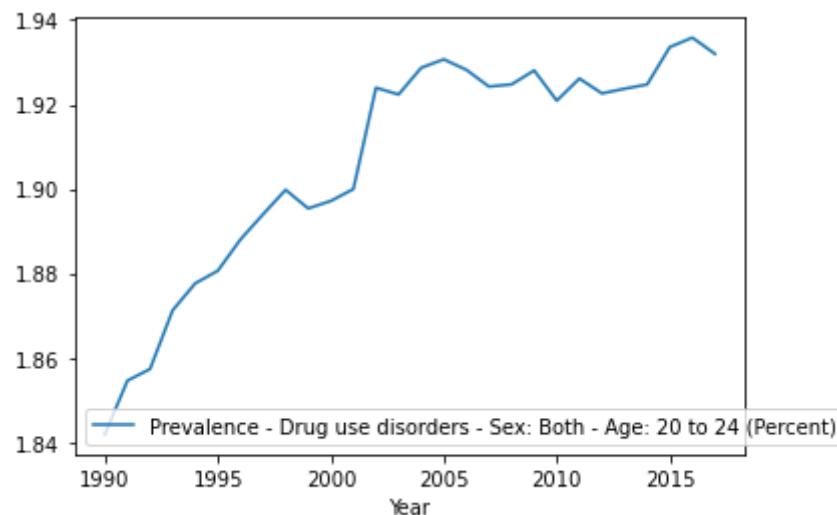
```
In [774]: df80.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[774]: <AxesSubplot:xlabel='Year'>
```



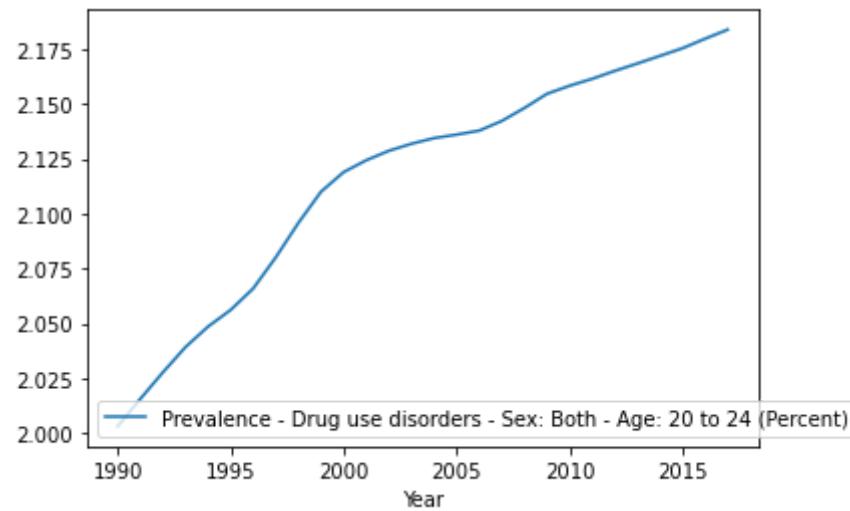
```
In [775]: df80.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[775]: <AxesSubplot:xlabel='Year'>
```



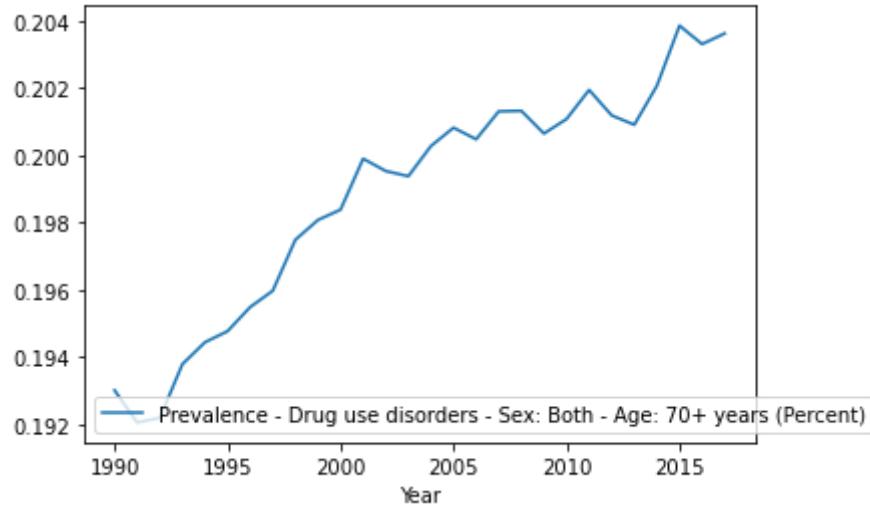
```
In [776]: df80.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[776]: <AxesSubplot:xlabel='Year'>
```



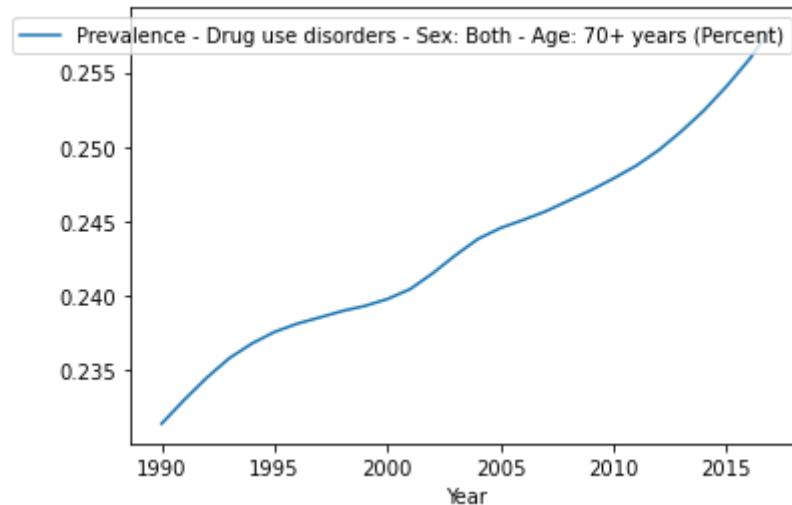
```
In [777]: df80.groupby('Year')[v7].median().plot(legend=True)
```

Out[777]: <AxesSubplot:xlabel='Year'>



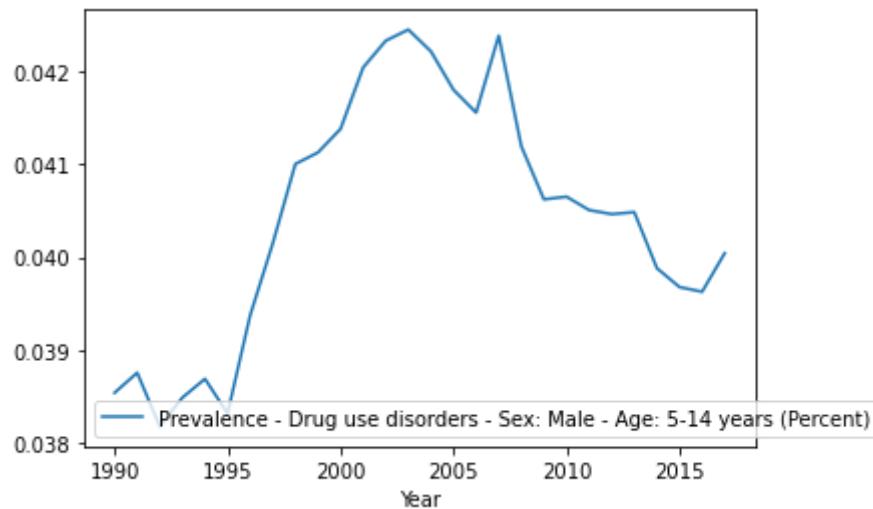
```
In [778]: df80.groupby('Year')[v7].mean().plot(legend=True)
```

Out[778]: <AxesSubplot:xlabel='Year'>



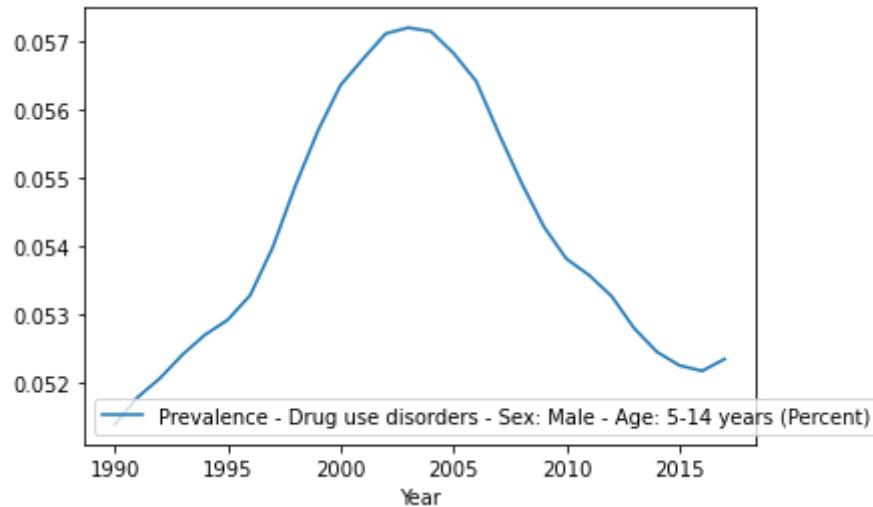
```
In [779]: df80.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[779]: <AxesSubplot:xlabel='Year'>
```



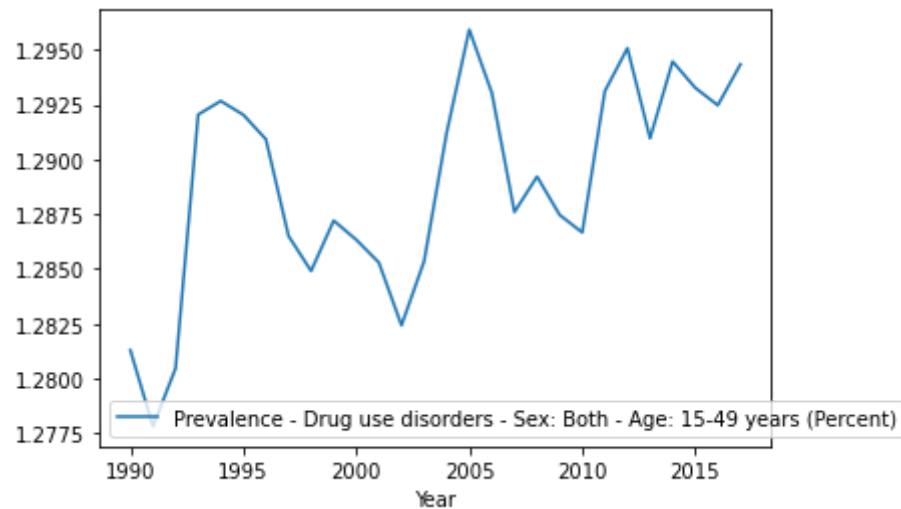
```
In [780]: df80.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[780]: <AxesSubplot:xlabel='Year'>
```



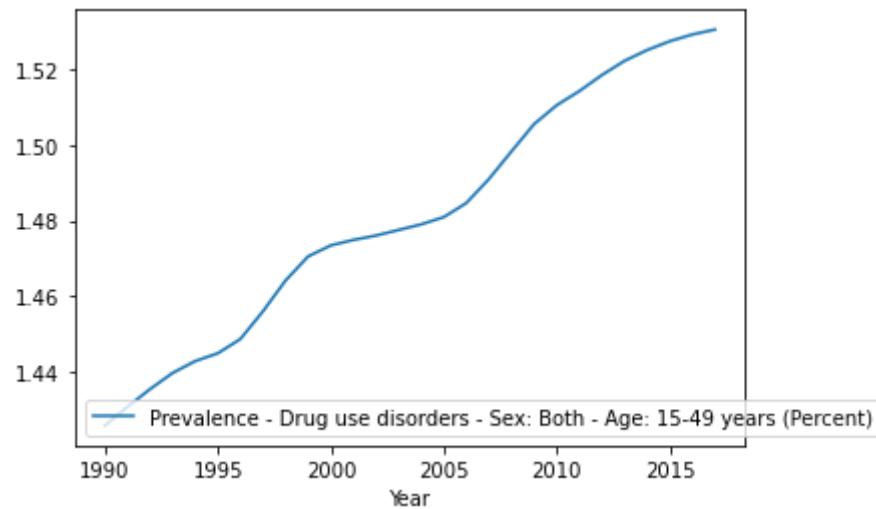
```
In [781]: df80.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[781]: <AxesSubplot:xlabel='Year'>
```



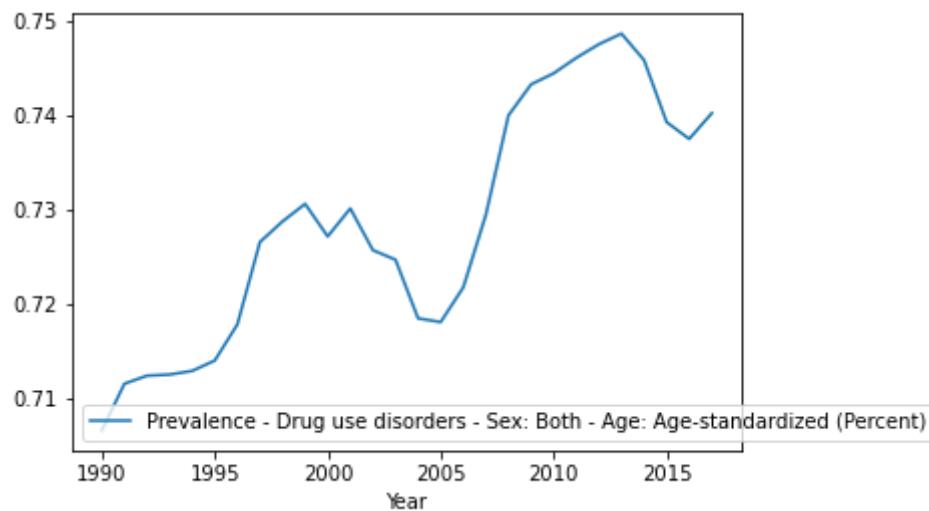
```
In [782]: df80.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[782]: <AxesSubplot:xlabel='Year'>
```



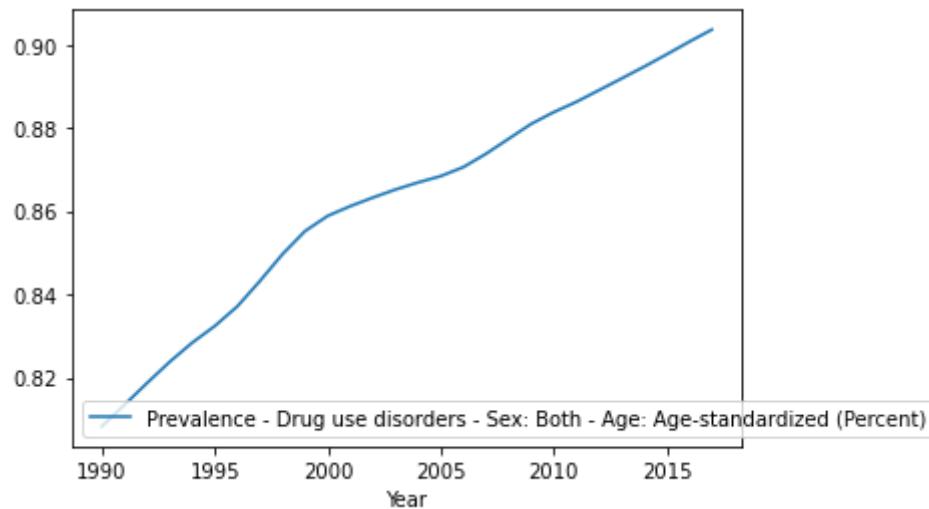
```
In [783]: df80.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[783]: <AxesSubplot:xlabel='Year'>
```



```
In [784]: df80.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[784]: <AxesSubplot:xlabel='Year'>
```



In [785]: df110.info()
df110.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype  
---  --  
 0   Entity           6468 non-null    object  
 1   Code              5488 non-null    object  
 2   Year              6468 non-null    int64  
 3   Deaths - Tobacco - Sex: Both - Age: All Ages (Number) 6468 non-null    float64 
 4   Deaths - Alcohol use - Sex: Both - Age: All Ages (Number) 6468 non-null    float64 
 5   Deaths - Drug use - Sex: Both - Age: All Ages (Number) 6468 non-null    float64 
 6   Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number) 6468 non-null    float64 
 7   Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number) 6468 non-null    float64 
dtypes: float64(5), int64(1), object(2)
memory usage: 404.4+ KB
```

Out[785]:

	Entity	Code	Year	Deaths - Tobacco - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)	Deaths - Drug use - Sex: Both - Age: All Ages (Number)	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	1990	10440.113942	356.529307	208.325430	78.254192	75.595221
1	Afghanistan	AFG	1991	10461.208026	320.598461	217.769691	81.284586	76.109641
2	Afghanistan	AFG	1992	10741.910669	293.257016	247.833251	91.010197	81.138310
3	Afghanistan	AFG	1993	11398.448085	278.129758	285.036181	103.481991	87.514744
4	Afghanistan	AFG	1994	11978.701505	250.691569	306.646796	111.296397	90.464149

Checking for missing values:

```
In [786]: missing = pd.concat([df110.isnull().sum(), 100 * df110.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[786]:

	count	%
Entity	0	0.000000
Year	0	0.000000
Deaths - Tobacco - Sex: Both - Age: All Ages (Number)	0	0.000000
Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)	0	0.000000
Deaths - Drug use - Sex: Both - Age: All Ages (Number)	0	0.000000
Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	0	0.000000
Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	0	0.000000
Code	980	15.151515

```
In [787]: v1='Deaths - Tobacco - Sex: Both - Age: All Ages (Number)'
v2='Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)'
v3='Deaths - Drug use - Sex: Both - Age: All Ages (Number)'
v4='Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)'
v5='Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)'
```

In [788]: df110.describe()

Out[788]:

	Year	Deaths - Tobacco - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)	Deaths - Drug use - Sex: Both - Age: All Ages (Number)	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
count	6468.000000	6.468000e+03	6.468000e+03	6468.000000	6468.000000	6468.000000
mean	2003.500000	1.530933e+05	5.020334e+04	8890.242150	2276.504819	3525.876073
std	8.078372	6.081494e+05	1.958226e+05	35415.115589	9663.540155	14336.535999
min	1990.000000	1.437328e+01	-2.315345e+03	1.240062	0.179388	0.293634
25%	1996.750000	1.623422e+03	3.639522e+02	92.909932	17.214436	23.700635
50%	2003.500000	7.242169e+03	2.803322e+03	408.586291	75.975217	158.916099
75%	2010.250000	3.594890e+04	1.289127e+04	2170.843581	460.087489	696.238005
max	2017.000000	8.101891e+06	2.842854e+06	585348.180194	166612.548448	190843.575458

In [789]: df110.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[789]:

	Entity	Year	Deaths - Tobacco - Sex: Both - Age: All Ages (Number)
6383	World	2017	8.101891e+06
6382	World	2016	8.039344e+06
6381	World	2015	7.912180e+06
6380	World	2014	7.717034e+06
6379	World	2013	7.607322e+06

In [790]: df110.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[790]:

	Entity	Year	Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)
6383	World	2017	2.842854e+06
6382	World	2016	2.817610e+06
6381	World	2015	2.773307e+06
6380	World	2014	2.706467e+06
6376	World	2010	2.677118e+06

In [791]: df110.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[791]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
6383	World	2017	585348.180194
6382	World	2016	572922.966987
6381	World	2015	549156.843911
6380	World	2014	526924.912151
6379	World	2013	508271.643532

In [792]: df110.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[792]:

	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	166612.548448
6382	World	2016	159690.500590
6381	World	2015	147614.117105
6380	World	2014	137888.912454
6379	World	2013	129359.881643

In [793]: df110.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[793]:

	Entity	Year	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
6371	World	2005	190843.575458
6370	World	2004	185851.759450
6383	World	2017	184934.241760
6369	World	2003	184025.573933
6372	World	2006	183932.669260

In [795]: df110.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[795]:

	Entity	Year	Deaths - Tobacco - Sex: Both - Age: All Ages (Number)
4260	Northern Mariana Islands	1994	14.373283
4256	Northern Mariana Islands	1990	14.433493
4259	Northern Mariana Islands	1993	14.571902
4261	Northern Mariana Islands	1995	14.581797
4258	Northern Mariana Islands	1992	14.870746

In [796]: df110.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[796]:

	Entity	Year	Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)
1848	England	1990	-2315.344758
1849	England	1991	-1529.132771
1850	England	1992	-789.545092
4284	Norway	1990	-720.204848
4285	Norway	1991	-718.888470

In [797]: df110.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[797]:

	Entity	Year	Deaths - Drug use - Sex: Both - Age: All Ages (Number)
84	American Samoa	1990	1.240062
85	American Samoa	1991	1.289510
86	American Samoa	1992	1.330492
87	American Samoa	1993	1.366904
88	American Samoa	1994	1.408779

In [798]: `df110.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()`

Out[798]:

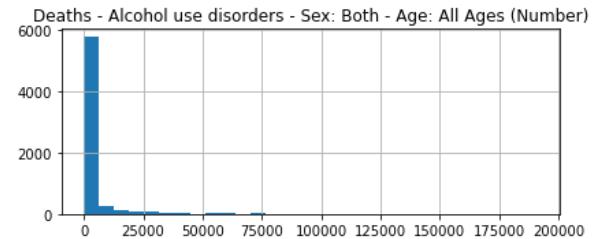
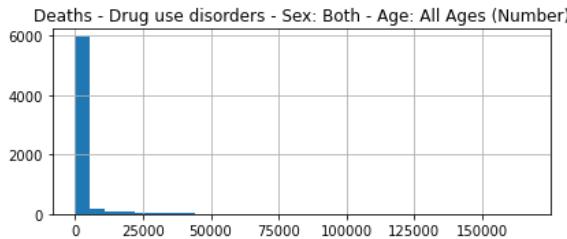
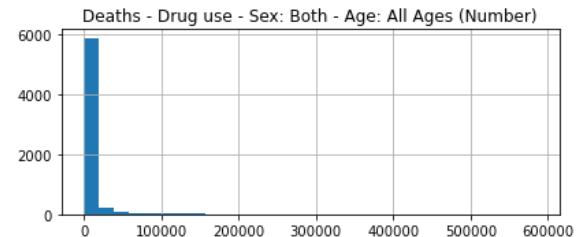
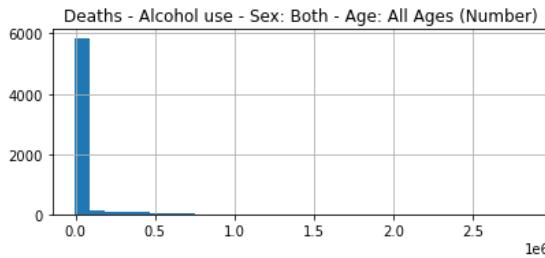
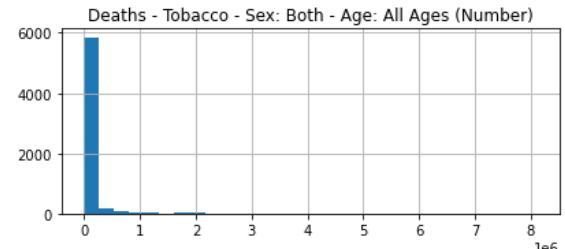
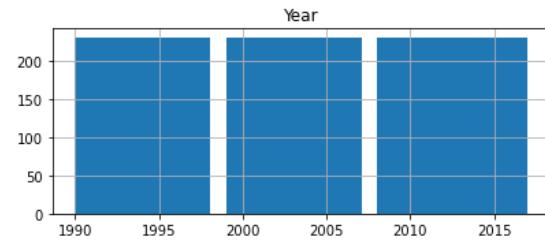
	Entity	Year	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
4256	Northern Mariana Islands	1990	0.179388
4260	Northern Mariana Islands	1994	0.186184
4259	Northern Mariana Islands	1993	0.187125
4257	Northern Mariana Islands	1991	0.188218
4258	Northern Mariana Islands	1992	0.189438

In [799]: `df110.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()`

Out[799]:

	Entity	Year	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
99	American Samoa	2005	0.293634
3027	Kiribati	1993	0.294713
84	American Samoa	1990	0.295547
98	American Samoa	2004	0.295837
3028	Kiribati	1994	0.301382

In [801]: `df110.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [802]: df110.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[802]: Entity
Northern Mariana Islands    21.383920
American Samoa              33.034350
Antigua and Barbuda        38.734569
Sao Tome and Principe      39.928382
Dominica                     41.810463
Name: Deaths - Tobacco - Sex: Both - Age: All Ages (Number), dtype: float
64
```

```
In [803]: df110.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[803]: Entity
Malta                      -36.022957
Iceland                    -12.328735
Norway                     -12.051495
American Samoa             1.752214
Marshall Islands           3.258901
Name: Deaths - Alcohol use - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [804]: df110.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[804]: Entity
American Samoa              1.684194
Northern Mariana Islands    2.022616
Marshall Islands            2.563348
Antigua and Barbuda        3.102877
Dominica                     3.162885
Name: Deaths - Drug use - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [805]: df110.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[805]: Entity
Northern Mariana Islands    0.261072
American Samoa               0.343522
Kiribati                     0.434098
Tonga                        0.457850
Marshall Islands             0.464715
Name: Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [806]: df110.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[806]: Entity
American Samoa              0.325306
Kiribati                     0.339944
Marshall Islands             0.506614
Tonga                        0.554691
Maldives                     0.685093
Name: Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [807]: df110.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[807]: Entity
High-middle SDI           1.964241e+06
China                      1.998778e+06
East Asia                  2.091979e+06
Southeast Asia, East Asia, and Oceania 2.661503e+06
World                      7.363845e+06
Name: Deaths - Tobacco - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [808]: df110.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[808]: Entity
East Asia                  5.298323e+05
Middle SDI                 6.352024e+05
High-middle SDI            6.780146e+05
Southeast Asia, East Asia, and Oceania 6.825221e+05
World                      2.417500e+06
Name: Deaths - Alcohol use - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [809]: df110.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[809]: Entity
High-middle SDI            108734.845836
East Asia                  112786.815006
Middle SDI                 120070.912049
Southeast Asia, East Asia, and Oceania 143273.160750
World                      422808.510683
Name: Deaths - Drug use - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [810]: df110.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[810]: Entity
East Asia                  26753.759737
Southeast Asia, East Asia, and Oceania 29992.877713
High-income                37310.363509
High SDI                   37696.290110
World                      108139.994261
Name: Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [811]: df110.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[811]: Entity
Russia                     41905.007733
Eastern Europe              56897.859931
Central Europe, Eastern Europe, and Central Asia 67698.268996
High-middle SDI             69974.087894
World                      167016.740623
Name: Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

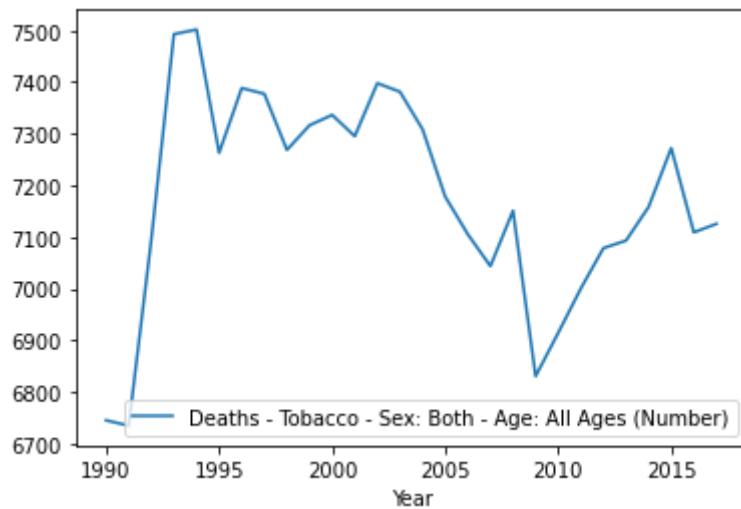
```
In [812]: df110_mean = df110.groupby('Year').mean()
df110_mean.head()
```

Out[812]:

	Deaths - Tobacco - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)	Deaths - Drug use - Sex: Both - Age: All Ages (Number)	Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)	Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
Year					
1990	139976.366533	37493.585640	5296.548094	1495.473734	2376.072381
1991	141774.124744	38811.034472	5654.451944	1620.550501	2517.967886
1992	143824.662228	40399.737904	6066.396237	1761.082372	2774.808891
1993	147099.247666	42198.338194	6484.202048	1879.070576	3162.873057
1994	148947.112020	43712.474709	6885.332827	1988.431624	3453.322565

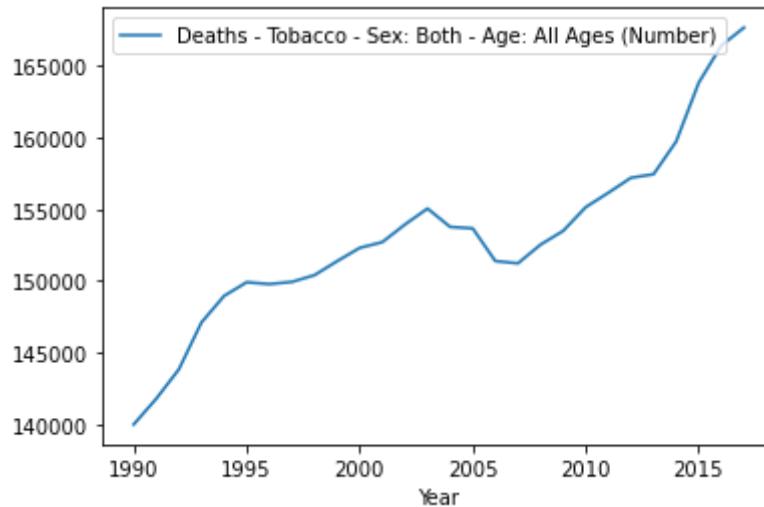
```
In [813]: df110.groupby('Year')[v1].median().plot(legend=True)
```

Out[813]: <AxesSubplot:xlabel='Year'>



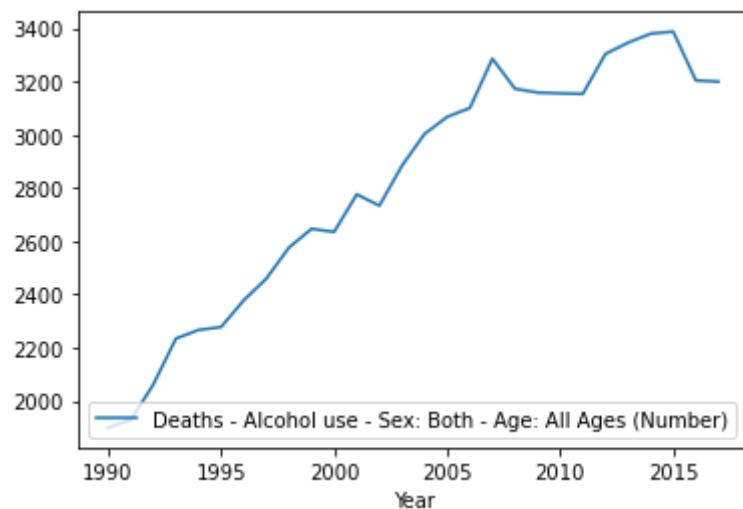
```
In [814]: df110.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[814]: <AxesSubplot:xlabel='Year'>
```



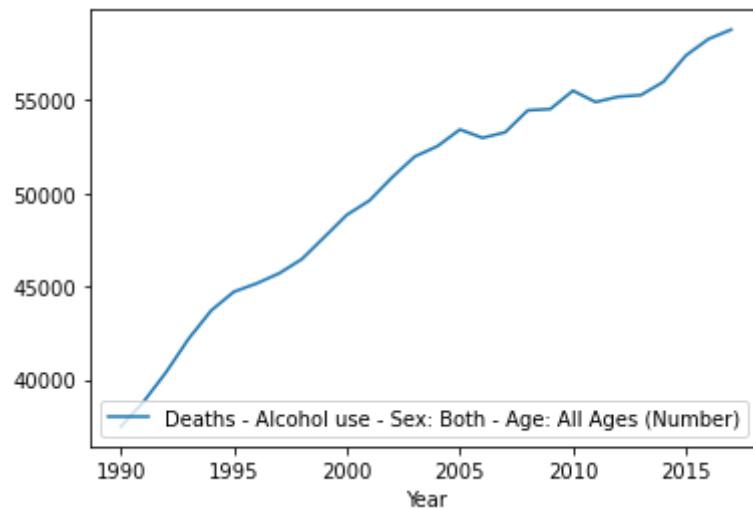
```
In [815]: df110.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[815]: <AxesSubplot:xlabel='Year'>
```



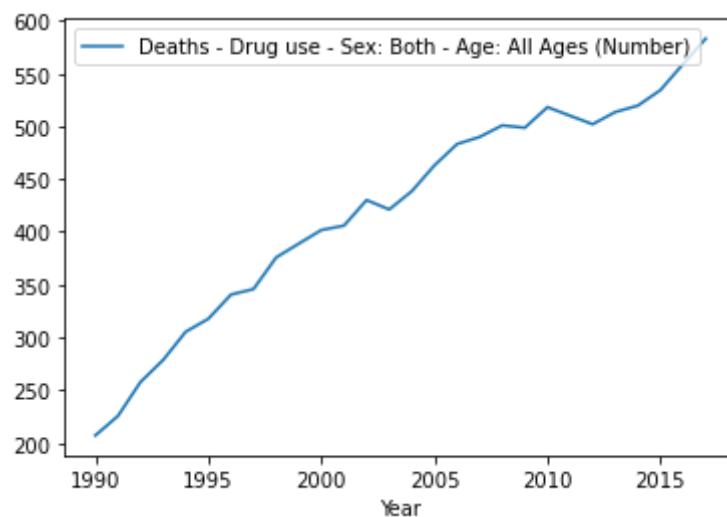
```
In [816]: df110.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[816]: <AxesSubplot:xlabel='Year'>
```



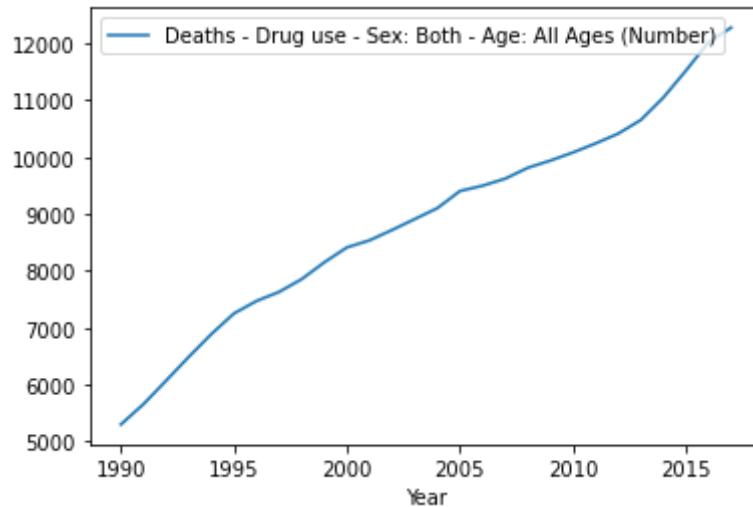
```
In [817]: df110.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[817]: <AxesSubplot:xlabel='Year'>
```



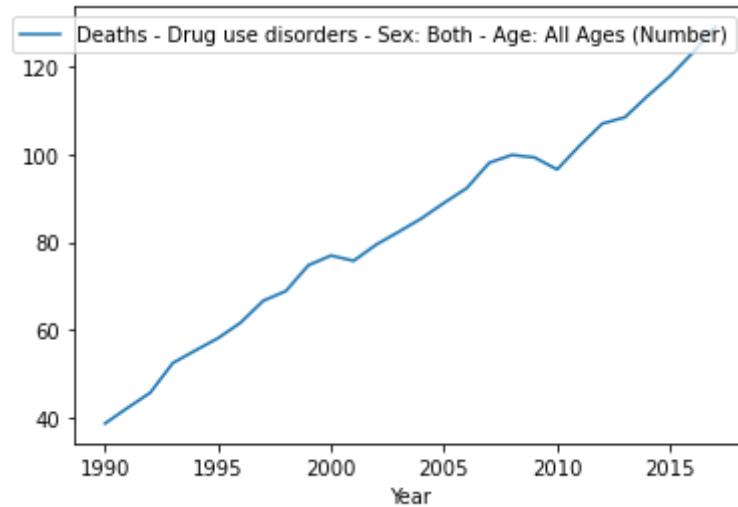
```
In [818]: df110.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[818]: <AxesSubplot:xlabel='Year'>
```



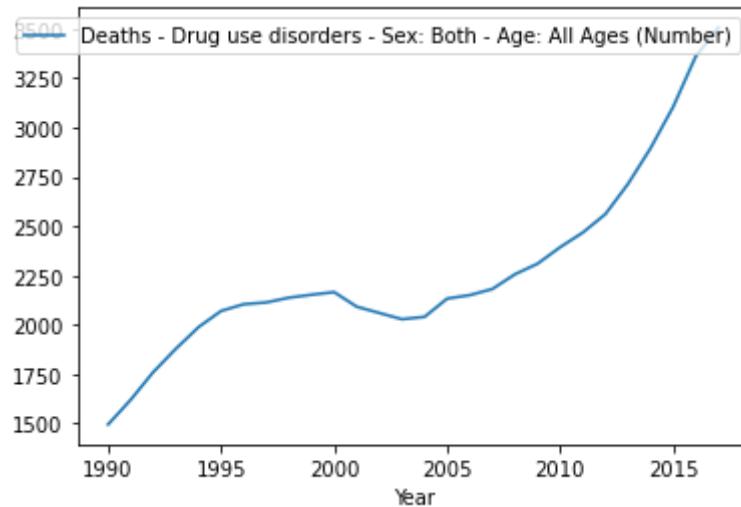
```
In [819]: df110.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[819]: <AxesSubplot:xlabel='Year'>
```



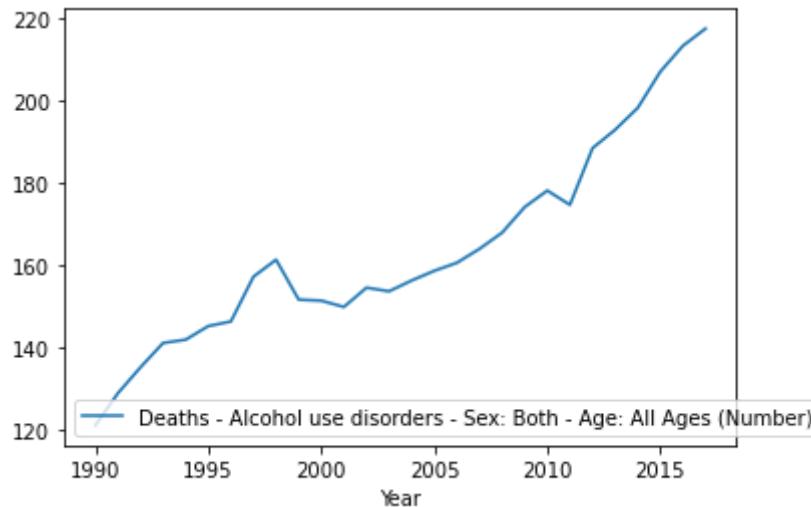
```
In [820]: df110.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[820]: <AxesSubplot:xlabel='Year'>
```



```
In [821]: df110.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[821]: <AxesSubplot:xlabel='Year'>
```



```
In [822]: df110.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[822]: <AxesSubplot:xlabel='Year'>
```



Mental Health and Substance Use Disorders

In [823]: df17.info()
df17.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 10 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year        int64
6468 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)          6468 non-null   float
64
4   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)           6468 non-null   float
64
5   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)           6468 non-null   float
64
6   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)           6468 non-null   float
64
7   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)           6468 non-null   float
64
8   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)            6468 non-null   float
64
9   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)  6468 non-null   float
64
dtypes: float64(7), int64(1), object(2)
memory usage: 505.4+ KB
```

Out[823]:

	Entity	Code	Year	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)
0	Afghanistan	AFG	1990	324.554760	1588.173650	1170.718747	2111.946976	2356.342730	1651..
1	Afghanistan	AFG	1991	342.900038	1592.705059	1175.847513	2120.935143	2364.547269	1661..

Entity	Code	Year	DALYs (Disability- Adjusted Life Years) - Mental and substance use	DALYs (Disability- Adjusted Life Years) - Mental and substance use						
			- disorders - Sex: Both - Age: Under 5 (Rate)	- disorders - Sex: Both - Age: 70+ years (Rate)	- disorders - Sex: Both - Age: 5-14 years (Rate)	- disorders - Sex: Both - Age: 50-69 years (Rate)	- disorders - Sex: Both - Age: 15-49 years (Rate)	- disorders - Sex: Both - Ages	- disorders - Sex: Both - Ages	
2	Afghanistan	AFG	1992	358.895522	1595.618716	1180.299534	2134.267133	2367.520133	1670.	
3	Afghanistan	AFG	1993	372.279506	1599.046250	1183.375231	2146.064953	2372.125291	1673.	
4	Afghanistan	AFG	1994	381.085610	1601.235717	1180.753541	2150.676280	2382.439831	1662.	

Checking for missing values:

```
In [824]: missing = pd.concat([df17.isnull().sum(), 100 * df17.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[824]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

```
In [825]: v1='DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)'
v2='DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)'
v3='DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)'
v4='DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)'
v5='DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)'
v6='DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)'
v7='DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age-standardized (Rate)'
```

In [826]: df17.describe()

Out[826]:

	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	129.801718	1716.288883	907.919476	1995.849972	1988.806482
std	8.078372	35.382507	237.674209	134.866270	209.469595	366.091705
min	1990.000000	76.921883	1175.263989	679.036719	1516.483890	1410.228401
25%	1996.750000	107.611038	1536.705787	811.919783	1839.744332	1714.926243
50%	2003.500000	122.063068	1665.654320	868.362727	2000.664290	1888.685286
75%	2010.250000	141.657080	1878.130667	969.097977	2119.697050	2230.457726
max	2017.000000	384.620835	2701.384674	1356.886685	2751.407118	3444.000026
						2530.077812

In [827]: df17.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[827]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	
5	Afghanistan	1995	384.620835
6	Afghanistan	1996	384.321595
7	Afghanistan	1997	383.405286
8	Afghanistan	1998	381.845656
4	Afghanistan	1994	381.085610

In [828]: df17.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[828]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	
5980	Uganda	2006	2701.384674
5978	Uganda	2004	2700.808597
5979	Uganda	2005	2700.746762
5977	Uganda	2003	2700.255860
5976	Uganda	2002	2699.261325

In [829]: df17.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[829]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)
4247	Northern Ireland	2009	1356.886685
4246	Northern Ireland	2008	1354.936717
4248	Northern Ireland	2010	1353.033396
4245	Northern Ireland	2007	1351.343106
4239	Northern Ireland	2001	1350.431943

In [830]: df17.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[830]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)
3852	Morocco	2006	2751.407118
3854	Morocco	2008	2750.214511
3851	Morocco	2005	2750.101814
3855	Morocco	2009	2748.817766
3853	Morocco	2007	2748.609753

In [831]: df17.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[831]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)
2273	Greenland	1995	3444.000026
2274	Greenland	1996	3442.522899
2272	Greenland	1994	3439.758030
2275	Greenland	1997	3438.351992
2276	Greenland	1998	3431.140689

In [832]: df17.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[832]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)
2288	Greenland	2010	2530.077812
2287	Greenland	2009	2528.345624
2289	Greenland	2011	2527.253578
2286	Greenland	2008	2524.428428
2270	Greenland	1992	2524.107702

In [833]: df17.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[833]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)
2274	Greenland	1996	2388.486688
2275	Greenland	1997	2388.220586
2273	Greenland	1995	2387.516967
2276	Greenland	1998	2386.226584
2272	Greenland	1994	2383.648307

In [834]: df17.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[834]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)
5653	Taiwan	2015	76.921883
5651	Taiwan	2013	77.115897
5655	Taiwan	2017	77.120770
5652	Taiwan	2014	77.140940
5654	Taiwan	2016	77.277811

In [835]: df17.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[835]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	
789	Brunei	1995	1175.263989
788	Brunei	1994	1175.770190
790	Brunei	1996	1176.116693
791	Brunei	1997	1176.234209
787	Brunei	1993	1177.555357

In [836]: df17.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[836]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	
1263	China	1993	679.036719
1264	China	1994	680.023505
1683	East Asia	1993	681.487516
1684	East Asia	1994	682.408156
6269	Vietnam	2015	683.119869

In [837]: df17.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[837]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	
1315	Colombia	2017	1516.483890
1314	Colombia	2016	1516.981611
1313	Colombia	2015	1518.281844
1312	Colombia	2014	1519.341088
1311	Colombia	2013	1521.561445

In [838]: df17.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[838]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)
3914	Myanmar	2012	1410.228401
3913	Myanmar	2011	1410.476831
3912	Myanmar	2010	1410.733801
3915	Myanmar	2013	1411.552201
3916	Myanmar	2014	1412.502842

In [839]: df17.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[839]:

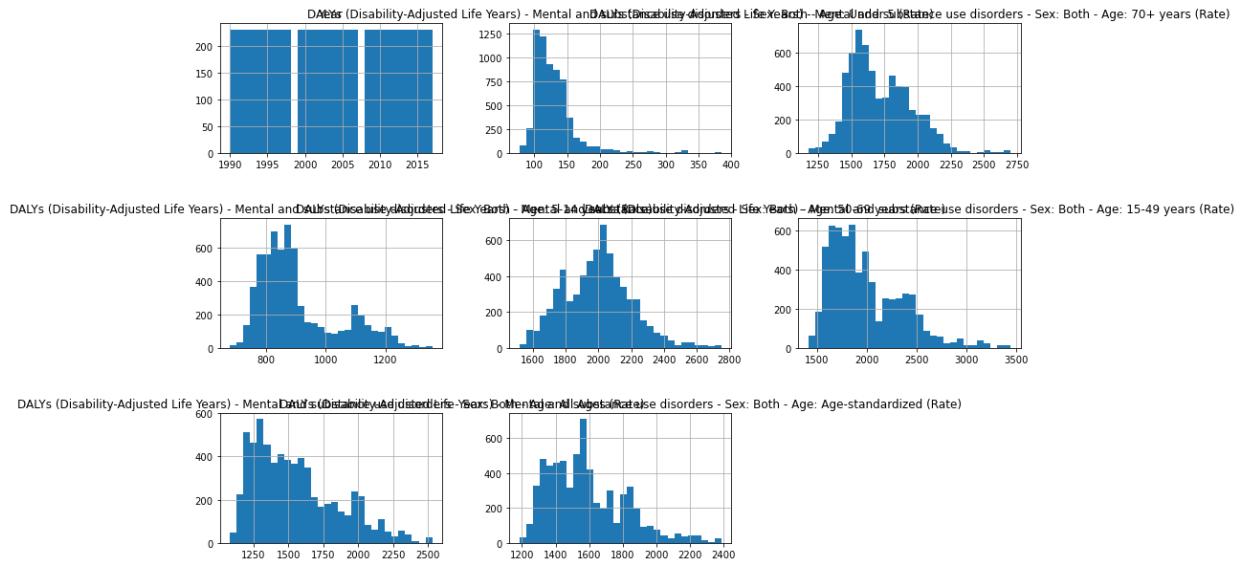
	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)
3545	Mali	2007	1080.423028
3544	Mali	2006	1080.698897
3546	Mali	2008	1080.705350
3543	Mali	2005	1081.443440
3547	Mali	2009	1082.048040

In [840]: df17.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[840]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)
3917	Myanmar	2015	1183.145977
3916	Myanmar	2014	1183.320100
3915	Myanmar	2013	1183.498466
3914	Myanmar	2012	1183.769238
3918	Myanmar	2016	1184.019663

```
In [841]: df17.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [842]: df17.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[842]: Entity
Taiwan           80.603399
Guam            85.025160
Northern Mariana Islands  86.187983
American Samoa  91.140988
Venezuela        93.160913
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate), dtype: float64
```

```
In [843]: df17.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[843]: Entity
Brunei          1189.913275
Colombia         1268.871798
Canada           1292.947391
Japan            1303.823606
High-income Asia Pacific 1321.570724
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [844]: df17.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[844]: Entity
Vietnam      722.016016
Colombia    726.623107
China        727.586807
East Asia   728.887167
Honduras    747.499945
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [845]: df17.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[845]: Entity
Brunei       1558.993520
Colombia    1560.458720
Albania      1586.989699
Azerbaijan  1593.934444
Tajikistan   1608.947698
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [846]: df17.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[846]: Entity
Myanmar     1424.375366
Vietnam     1469.429308
Tajikistan  1498.620137
Azerbaijan  1499.193394
Taiwan      1503.184990
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [847]: df17.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[847]: Entity
Mali         1095.934525
Tajikistan  1132.609979
Niger        1145.302073
Mauritania  1168.904944
Zimbabwe    1174.765360
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [848]: df17.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[848]: Entity
Myanmar     1206.838712
Vietnam     1232.954082
Azerbaijan  1240.017227
Tajikistan  1251.928040
Armenia     1267.223421
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [849]: df17.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[849]: Entity
Yemen           248.723703
South Asia     278.554692
Sudan           284.924365
India            327.091717
Afghanistan    338.933725
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate), dtype: float64
```

```
In [850]: df17.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[850]: Entity
Rwanda          2222.416837
Ukraine         2257.495481
Estonia          2297.009569
Lesotho          2517.751476
Uganda           2652.073981
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [851]: df17.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[851]: Entity
Australia        1221.827842
Australasia      1231.063676
Greece            1245.627614
New Zealand       1274.643729
Northern Ireland  1328.155545
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [852]: df17.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[852]: Entity
Lesotho           2488.434163
Northern Ireland  2489.383050
Iran              2578.152500
Uganda            2643.107445
Morocco           2704.692822
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [853]: df17.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[853]: Entity
United States     2936.914308
New Zealand       3099.859332
Australasia       3151.992604
Australia          3161.941943
Greenland          3385.631112
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [854]: df17.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[854]: Entity
Northern Ireland    2284.603765
New Zealand        2289.750280
Australasia         2341.658809
Australia           2351.842024
Greenland            2510.623735
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [855]: df17.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[855]: Entity
Northern Ireland    2178.061339
New Zealand        2225.266763
Australasia         2248.281186
Australia           2252.375446
Greenland            2368.473530
Name: DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

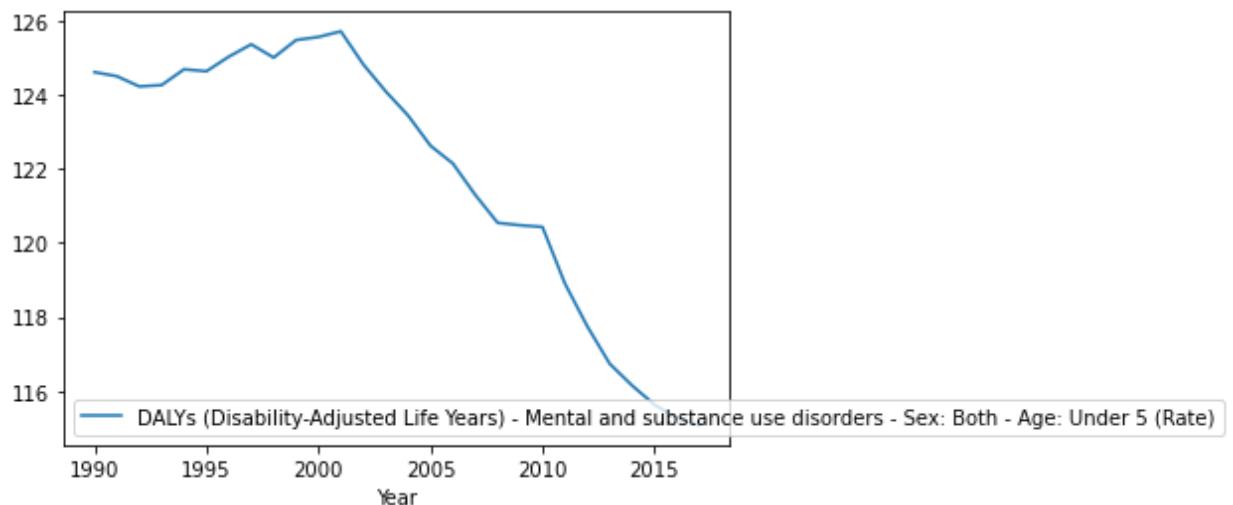
```
In [856]: df17_mean = df17.groupby('Year').mean()
df17_mean.head()
```

```
Out[856]:
```

DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)	DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Age- standardized (Rate)
Year						
1990	134.554569	1722.006434	900.824544	1996.809264	1984.846303	1498.618487
1991	134.510842	1723.127648	901.088955	1997.877326	1986.763734	1502.713927
1992	134.497099	1724.319400	901.697465	1998.932598	1988.343357	1506.812642
1993	134.420806	1725.425821	902.383128	1999.970212	1989.883123	1511.160797
1994	134.252482	1726.411166	902.761834	2000.846257	1991.200275	1515.535756

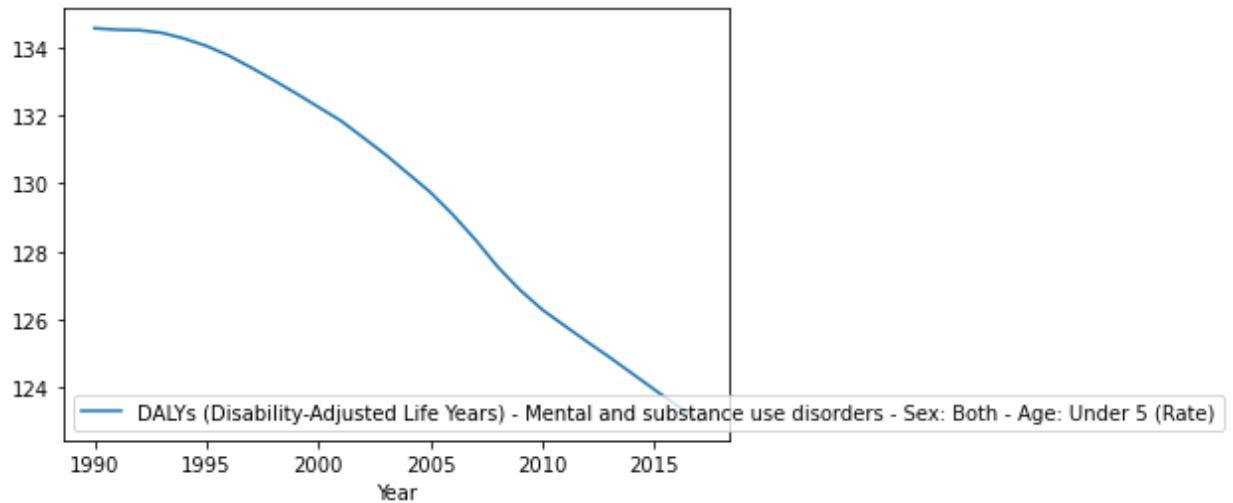
```
In [857]: df17.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[857]: <AxesSubplot:xlabel='Year'>
```



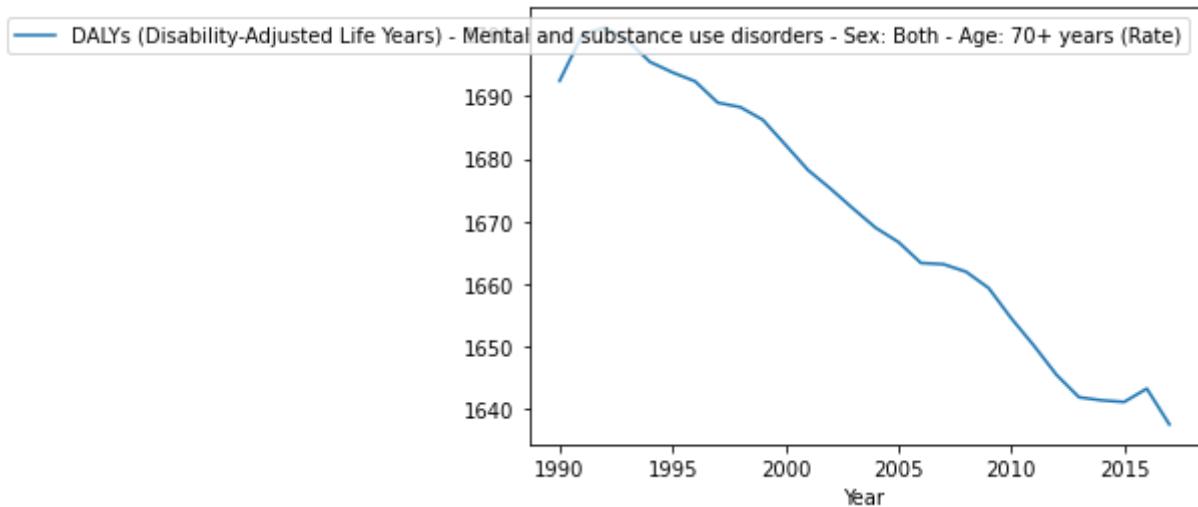
```
In [858]: df17.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[858]: <AxesSubplot:xlabel='Year'>
```



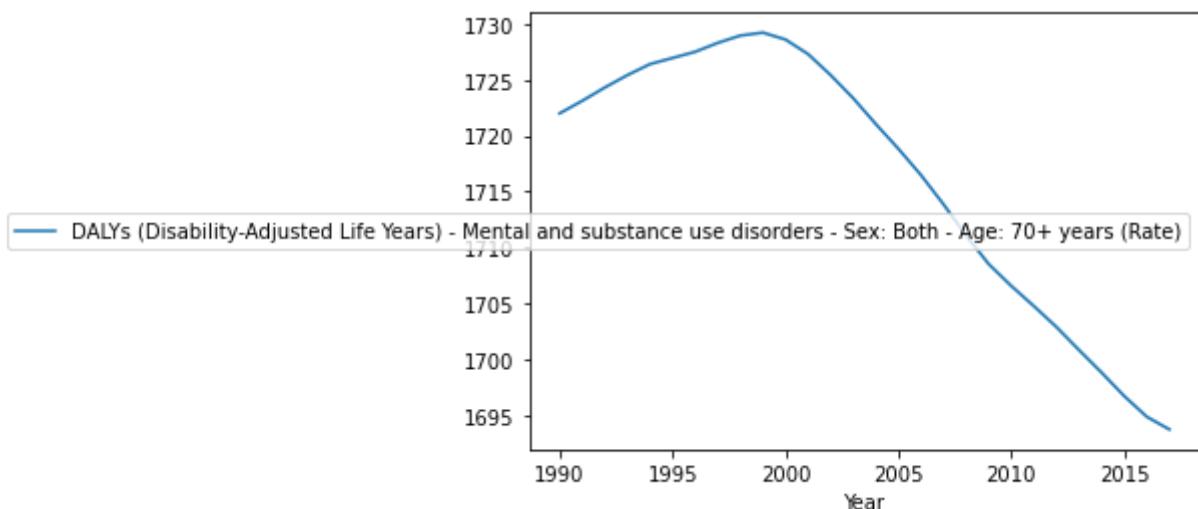
```
In [859]: df17.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[859]: <AxesSubplot:xlabel='Year'>
```



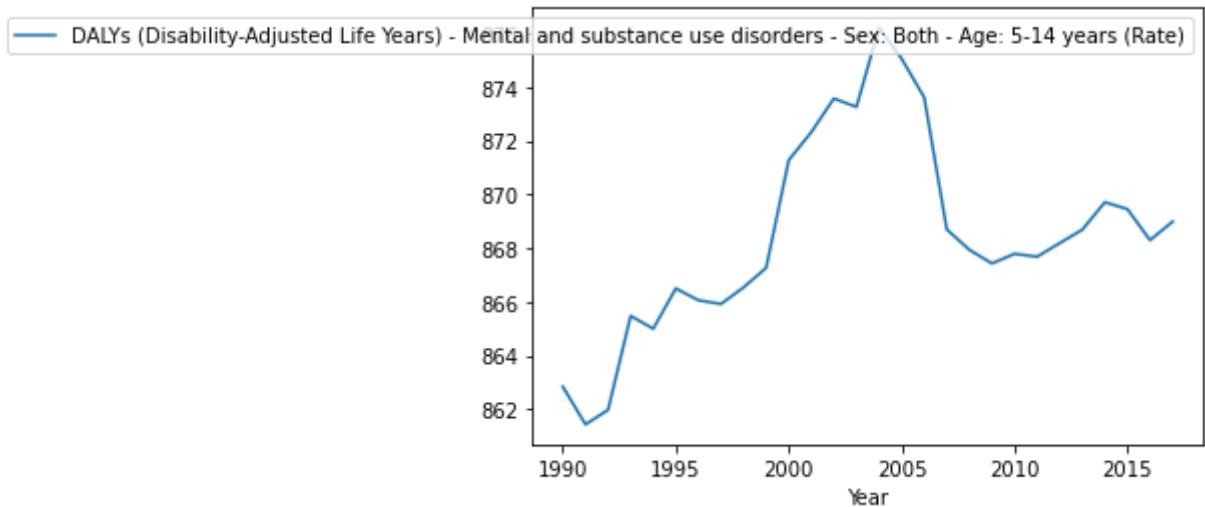
```
In [860]: df17.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[860]: <AxesSubplot:xlabel='Year'>
```



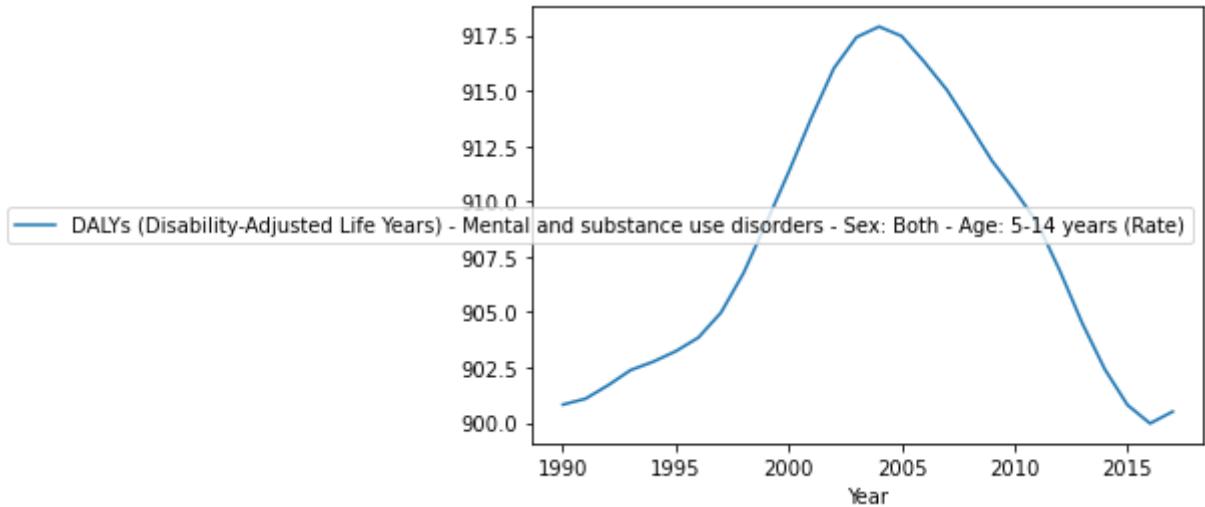
```
In [861]: df17.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[861]: <AxesSubplot:xlabel='Year'>
```



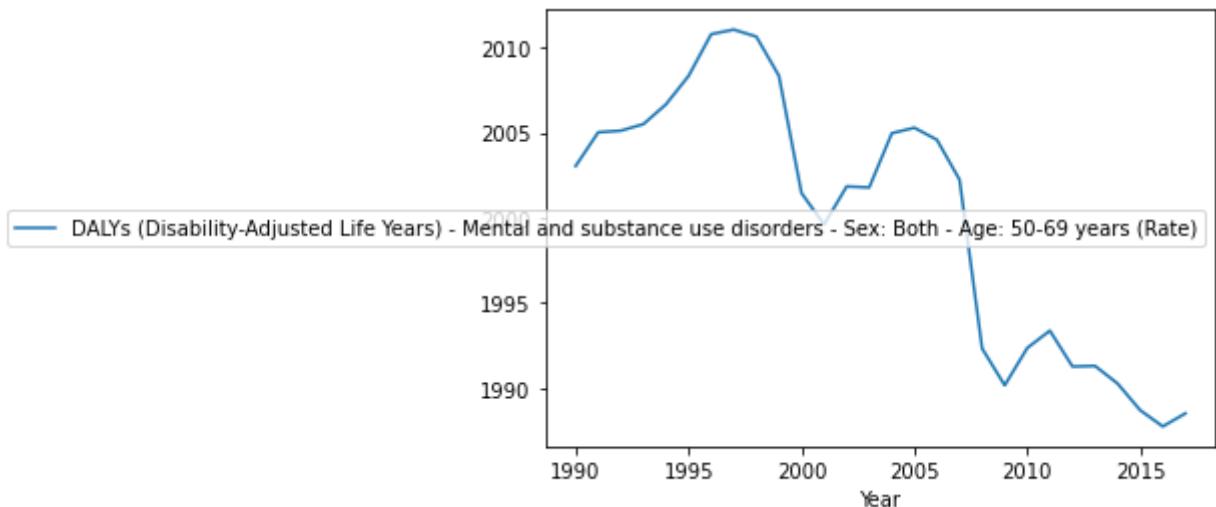
```
In [862]: df17.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[862]: <AxesSubplot:xlabel='Year'>
```



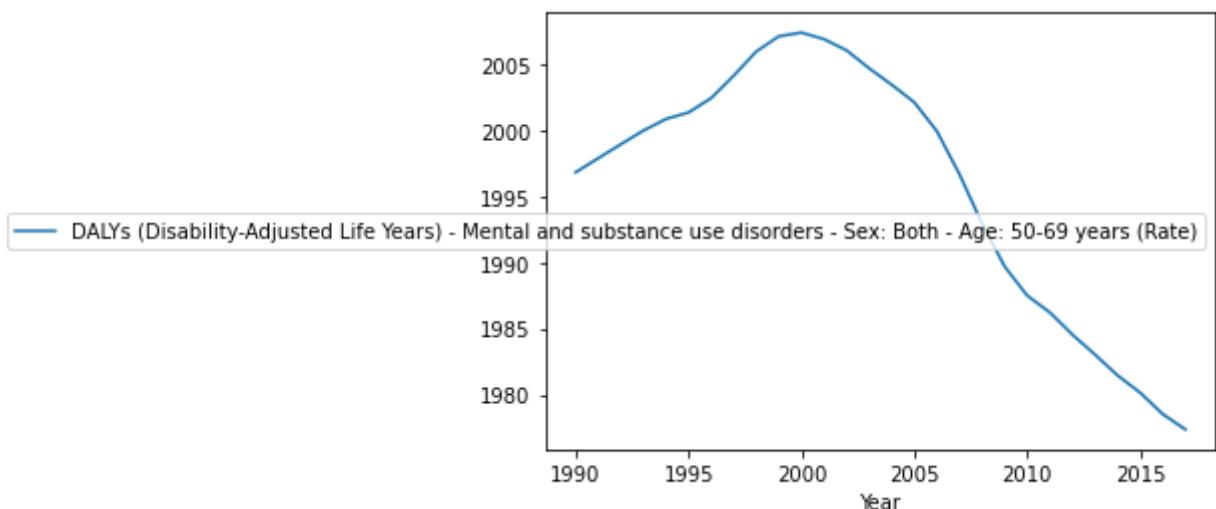
```
In [863]: df17.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[863]: <AxesSubplot:xlabel='Year'>
```



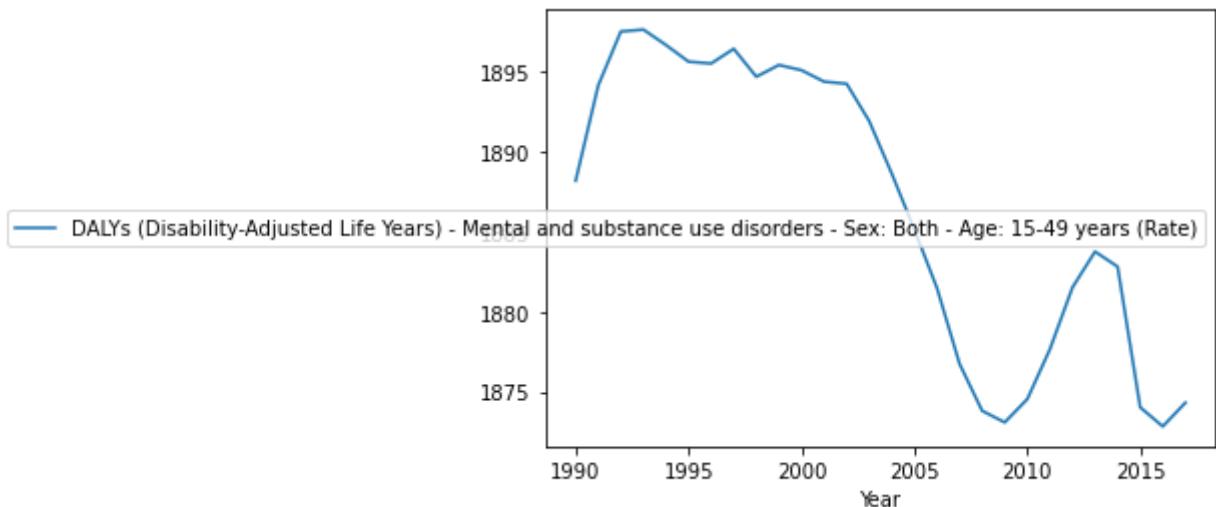
```
In [864]: df17.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[864]: <AxesSubplot:xlabel='Year'>
```



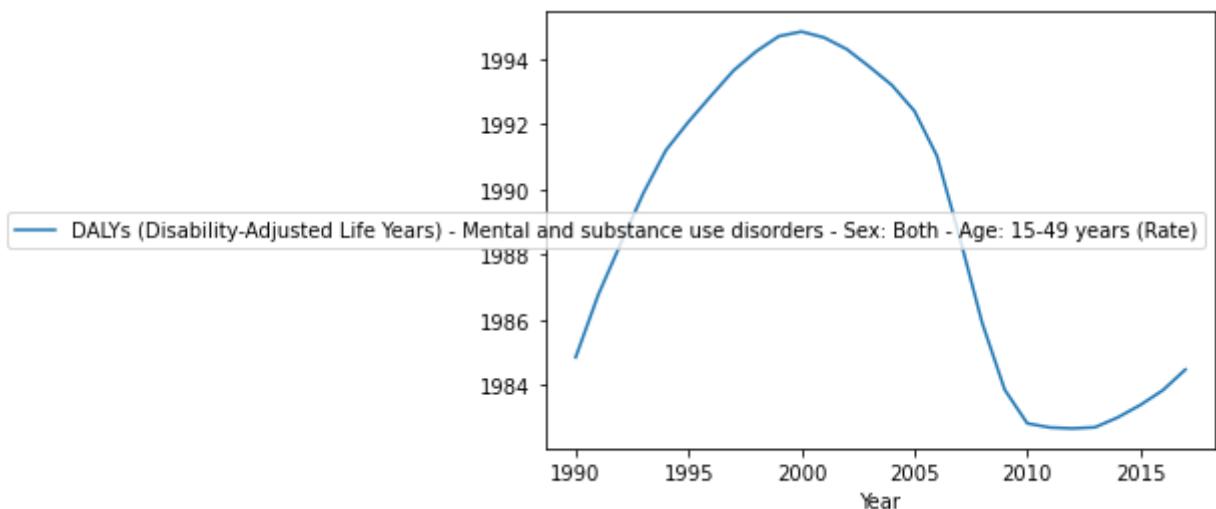
```
In [865]: df17.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[865]: <AxesSubplot:xlabel='Year'>
```



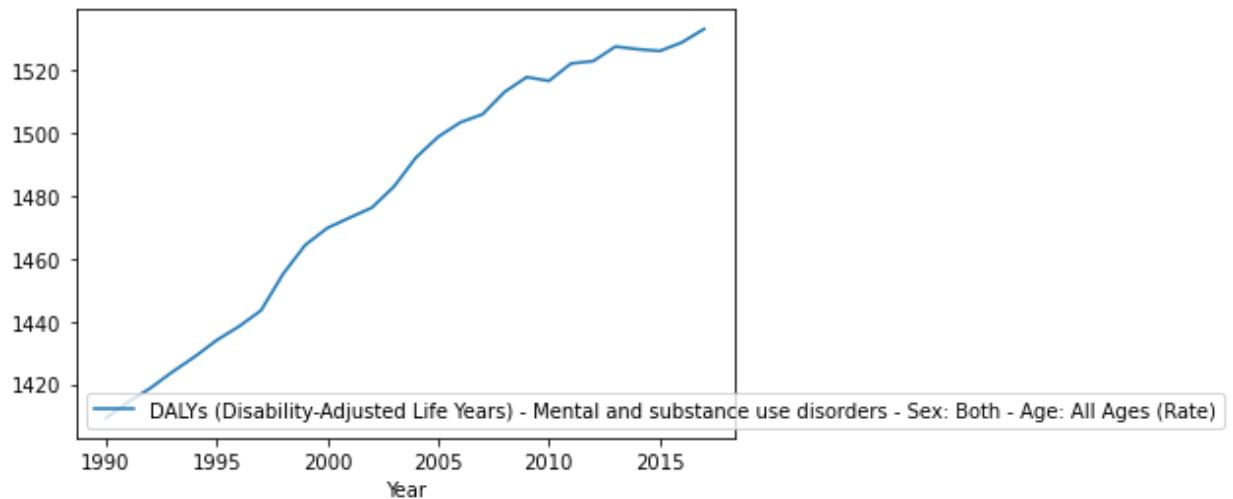
```
In [866]: df17.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[866]: <AxesSubplot:xlabel='Year'>
```



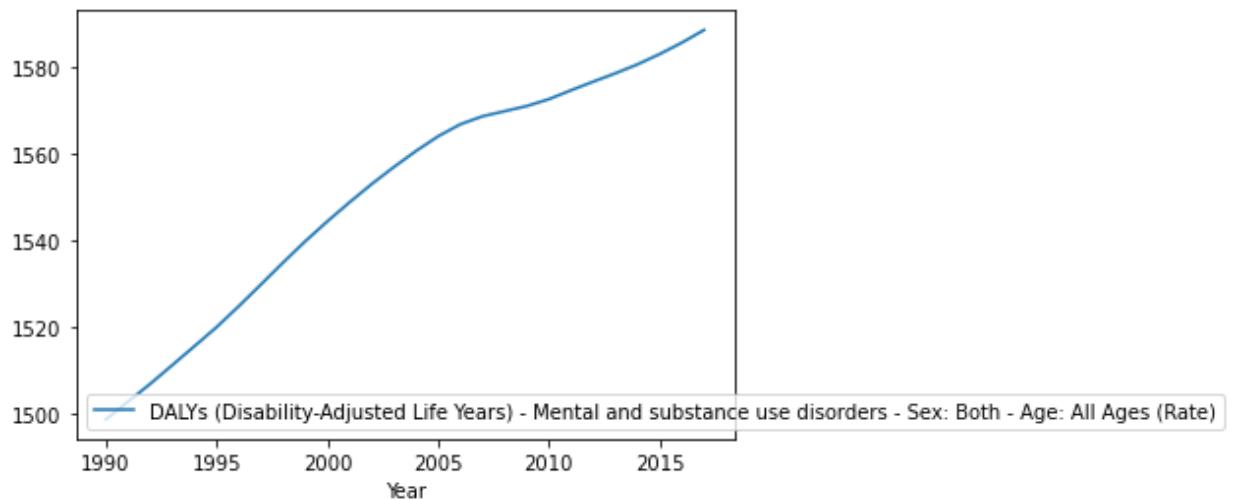
```
In [867]: df17.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[867]: <AxesSubplot:xlabel='Year'>
```



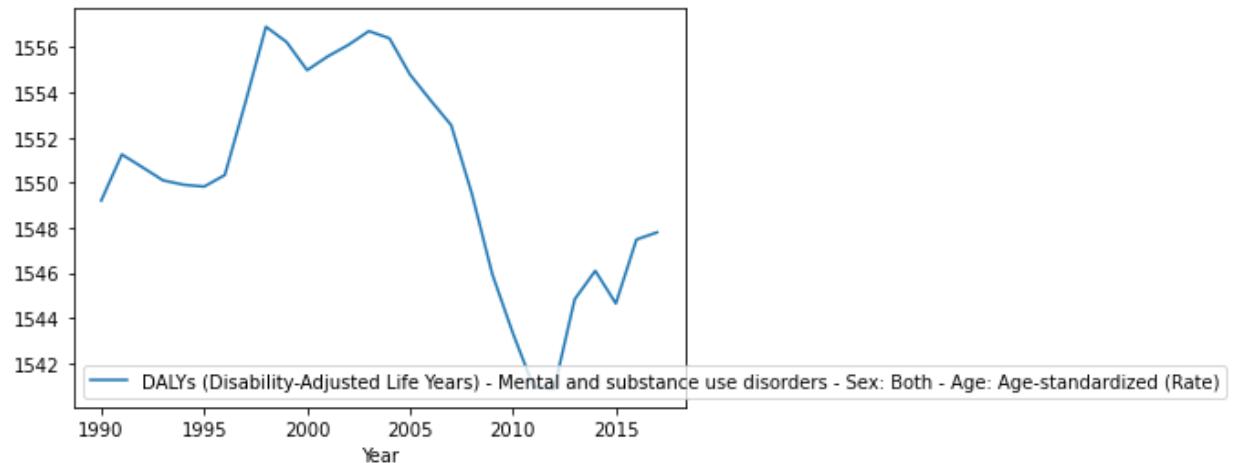
```
In [868]: df17.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[868]: <AxesSubplot:xlabel='Year'>
```



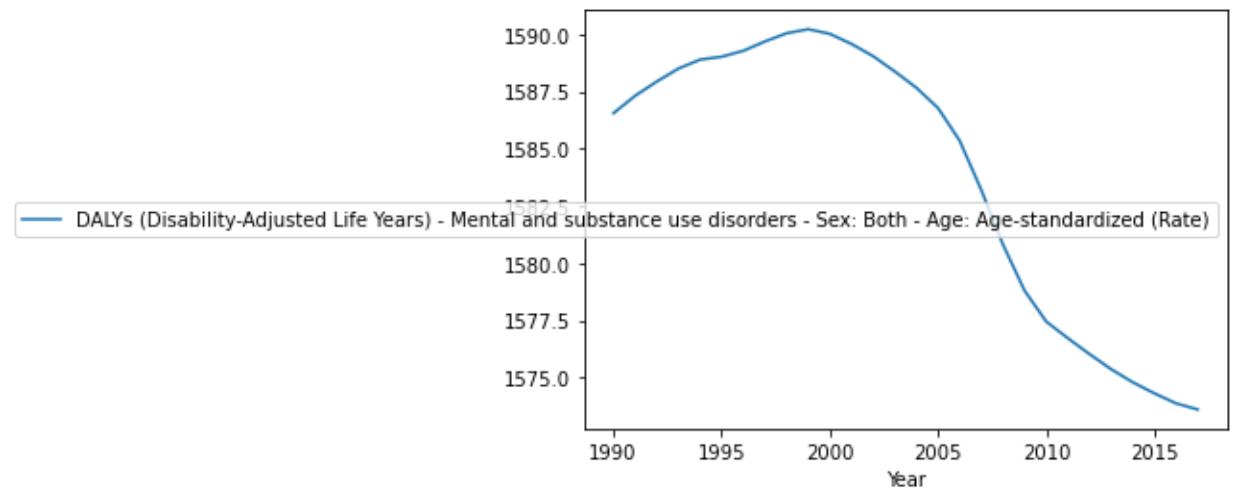
```
In [869]: df17.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[869]: <AxesSubplot:xlabel='Year'>
```



```
In [870]: df17.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[870]: <AxesSubplot:xlabel='Year'>
```



In [871]: df18.info()
df18.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 10 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)          6468 non-null   float64
4
4   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)           6468 non-null   float64
5
5   DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)          6468 non-null   float64
6
6   DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)  6468 non-null   float64
7
7   DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)          6468 non-null   float64
8
8   DALYs (Disability-Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number)  6468 non-null   float64
9
9   DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: All Ages (Number)          6468 non-null   float64
dtypes: float64(7), int64(1), object(2)
memory usage: 505.4+ KB
```

Out[871]:

	Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	1990	55175.002114	38693.841080	1749.842958	7745.147587	6808.953361
1	Afghanistan	AFG	1991	57088.309599	40134.004346	1808.943022	8021.598366	6991.706727

Entity	Code	Year	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)	
			(Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)	(Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)	(Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)	(Disability- Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	(Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)	
2	Afghanistan	AFG	1992	66721.220078	47591.190108	2222.046368	9481.777550	8037.327148
3	Afghanistan	AFG	1993	77128.443901	55362.651814	2647.076559	11143.050198	9246.501248
4	Afghanistan	AFG	1994	80106.045449	57473.350648	2727.391280	11740.701766	9690.785362

Checking for missing values:

```
In [872]: missing = pd.concat([df18.isnull().sum(), 100 * df18.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[872]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Code	980	15.151515	

```
In [882]: v1='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)'
v2='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)'
v3='DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)'
v4='DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)'
v5='DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)'
v6='DALYs (Disability-Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number)'
v7='DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: All Ages (Number)'
```

In [883]: df18.describe()

Out[883]:

	Year	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)	(Dis- Adjus Other sul use di - Sex Age: / (N
count	6468.000000	6.468000e+03	6.468000e+03	6.468000e+03	6.468000e+03	6.468000e+03	6.468000e+03
mean	2003.500000	7.257287e+05	4.654754e+05	5.544459e+04	3.249369e+05	2.094016e+05	1.8330
std	8.078372	2.843799e+06	1.803373e+06	2.153720e+05	1.260508e+06	8.608181e+05	7.2297
min	1990.000000	1.499374e+02	1.163656e+02	9.485065e+00	5.183047e+01	4.131976e+01	4.3273
25%	1996.750000	1.271145e+04	7.616426e+03	8.936216e+02	4.049180e+03	2.740689e+03	3.2375
50%	2003.500000	4.618646e+04	3.129093e+04	3.594727e+03	1.966998e+04	9.714427e+03	1.1830
75%	2010.250000	2.232327e+05	1.579110e+05	2.000231e+04	8.305441e+04	5.229994e+04	5.9538
max	2017.000000	4.309991e+07	2.712144e+07	3.369479e+06	1.746289e+07	1.265791e+07	1.1109

In [884]: df18.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[884]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	4.309991e+07
6382	World	2016	4.247073e+07
6381	World	2015	4.185286e+07
6380	World	2014	4.124378e+07
6379	World	2013	4.063114e+07

In [885]: df18.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[885]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)
6383	World	2017	2.712144e+07
6382	World	2016	2.682837e+07
6381	World	2015	2.653322e+07
6380	World	2014	2.624057e+07
6379	World	2013	2.594008e+07

```
In [886]: df18.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()
```

Out[886]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)	
6383	World	2017	3.369479e+06
6382	World	2016	3.318853e+06
6381	World	2015	3.266776e+06
6380	World	2014	3.216189e+06
6379	World	2013	3.164958e+06

```
In [887]: df18.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()
```

Out[887]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	
6383	World	2017	1.746289e+07
6382	World	2016	1.729257e+07
6381	World	2015	1.712330e+07
6380	World	2014	1.690357e+07
6371	World	2005	1.683854e+07

```
In [888]: df18.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()
```

Out[888]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)	
6383	World	2017	1.265791e+07
6382	World	2016	1.249532e+07
6381	World	2015	1.232534e+07
6380	World	2014	1.215137e+07
6379	World	2013	1.196965e+07

In [889]: df18.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[889]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number)	
6383	World	2017	1.110961e+07
6382	World	2016	1.094998e+07
6381	World	2015	1.078850e+07
6380	World	2014	1.062637e+07
6379	World	2013	1.046219e+07

In [890]: df18.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[890]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: All Ages (Number)	
6383	World	2017	2.718745e+07
6382	World	2016	2.654647e+07
6381	World	2015	2.567941e+07
6380	World	2014	2.493910e+07
6379	World	2013	2.425148e+07

In [891]: df18.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[891]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)	
3584	Marshall Islands	1990	149.937435
3585	Marshall Islands	1991	155.028373
3586	Marshall Islands	1992	160.254664
3587	Marshall Islands	1993	165.448489
84	American Samoa	1990	167.928200

In [892]: `df18.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()`

Out[892]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	116.365628
3585	Marshall Islands	1991	119.801116
3586	Marshall Islands	1992	123.136895
3587	Marshall Islands	1993	126.082007
3588	Marshall Islands	1994	129.225565

In [893]: `df18.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()`

Out[893]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	9.485065
3585	Marshall Islands	1991	9.898804
3586	Marshall Islands	1992	10.277163
3587	Marshall Islands	1993	10.637372
3588	Marshall Islands	1994	10.984500

In [894]: `df18.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()`

Out[894]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	51.830469
3585	Marshall Islands	1991	53.669184
3586	Marshall Islands	1992	55.431110
3587	Marshall Islands	1993	57.116491
3588	Marshall Islands	1994	58.873586

In [895]: df18.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[895]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	41.319756
3585	Marshall Islands	1991	42.592843
3586	Marshall Islands	1992	43.766536
3587	Marshall Islands	1993	44.906423
3588	Marshall Islands	1994	46.189171

In [896]: df18.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[896]:

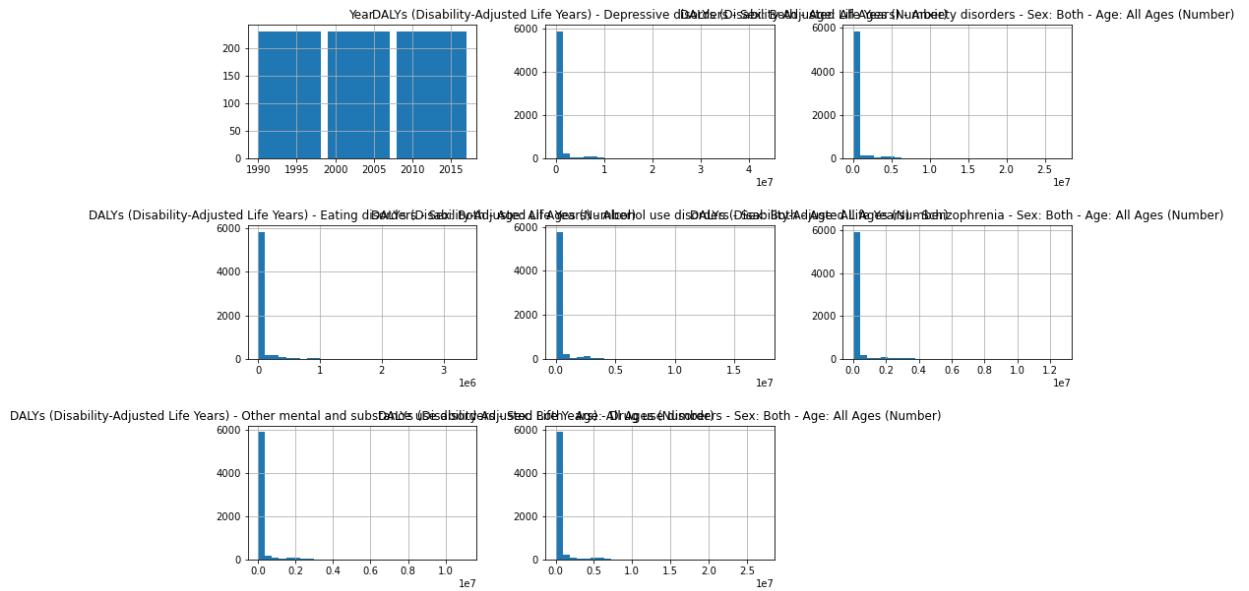
	Entity	Year	DALYs (Disability-Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	43.273521
3585	Marshall Islands	1991	44.559678
3586	Marshall Islands	1992	45.757935
3587	Marshall Islands	1993	47.023285
3588	Marshall Islands	1994	48.302732

In [897]: df18.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[897]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	72.389205
3585	Marshall Islands	1991	75.121045
3586	Marshall Islands	1992	77.649099
4283	Northern Mariana Islands	2017	78.671539
4282	Northern Mariana Islands	2016	78.809439

```
In [898]: df18.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [899]: df18.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[899]: Entity
Marshall Islands      199.193011
American Samoa        202.499336
Northern Mariana Islands  239.589560
Dominica              297.761282
Antigua and Barbuda   325.691358
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [900]: df18.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[900]: Entity
Marshall Islands      144.095189
American Samoa        158.716073
Northern Mariana Islands  169.852200
Kiribati               264.129014
Seychelles              268.091126
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [901]: df18.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[901]: Entity
Marshall Islands      12.595089
Kiribati              19.304984
American Samoa        21.118357
Micronesia (country)  24.558656
Tonga                 25.129982
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [902]: df18.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[902]: Entity
American Samoa        68.893183
Marshall Islands       74.811194
Kiribati               102.123853
Seychelles             107.824470
Tonga                  115.412492
Name: DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [903]: df18.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[903]: Entity
Marshall Islands       57.384768
American Samoa         75.228713
Dominica               85.264405
Northern Mariana Islands 100.962255
Kiribati                104.249502
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [904]: df18.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[904]: Entity
Marshall Islands       59.046246
American Samoa          67.759496
Northern Mariana Islands 84.312840
Dominica                 100.104201
Greenland                100.123395
Name: DALYs (Disability-Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [905]: df18.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[905]: Entity
American Samoa          93.159140
Marshall Islands         101.110512
Northern Mariana Islands 103.571122
Dominica                  146.424424
Kiribati                   149.990468
Name: DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [906]: df18.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[906]: Entity
East Asia           7.710532e+06
South Asia          7.754565e+06
Middle SDI          8.996749e+06
Southeast Asia, East Asia, and Oceania 9.867576e+06
World               3.562007e+07
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [907]: df18.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[907]: Entity
High-income         5.185098e+06
High SDI            5.214654e+06
Middle SDI          5.773954e+06
Southeast Asia, East Asia, and Oceania 5.935017e+06
World               2.277481e+07
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [908]: df18.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[908]: Entity
Southeast Asia, East Asia, and Oceania 5.320793e+05
Middle SDI                  6.266318e+05
High SDI                   9.172624e+05
High-income                 9.187661e+05
World                      2.673903e+06
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [909]: df18.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[909]: Entity
Middle SDI           3.413256e+06
Eastern Europe        3.459662e+06
Central Europe, Eastern Europe, and Central Asia 4.311872e+06
High-middle SDI      5.376926e+06
World                1.558242e+07
Name: DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [910]: df18.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[910]: Entity
Middle SDI           3.039195e+06
China                3.067832e+06
East Asia             3.198591e+06
Southeast Asia, East Asia, and Oceania 3.987688e+06
World                1.019252e+07
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [911]: df18.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[911]: Entity
China                               2.027189e+06
East Asia                           2.134908e+06
Middle SDI                          2.525292e+06
Southeast Asia, East Asia, and Oceania 2.897661e+06
World                                8.921175e+06
Name: DALYs (Disability-Adjusted Life Years) - Other mental and substance
use disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [912]: df18.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[912]: Entity
Middle SDI                         5.665345e+06
China                                5.737376e+06
East Asia                            5.964775e+06
Southeast Asia, East Asia, and Oceania 6.939355e+06
World                                2.100900e+07
Name: DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex:
Both - Age: All Ages (Number), dtype: float64
```

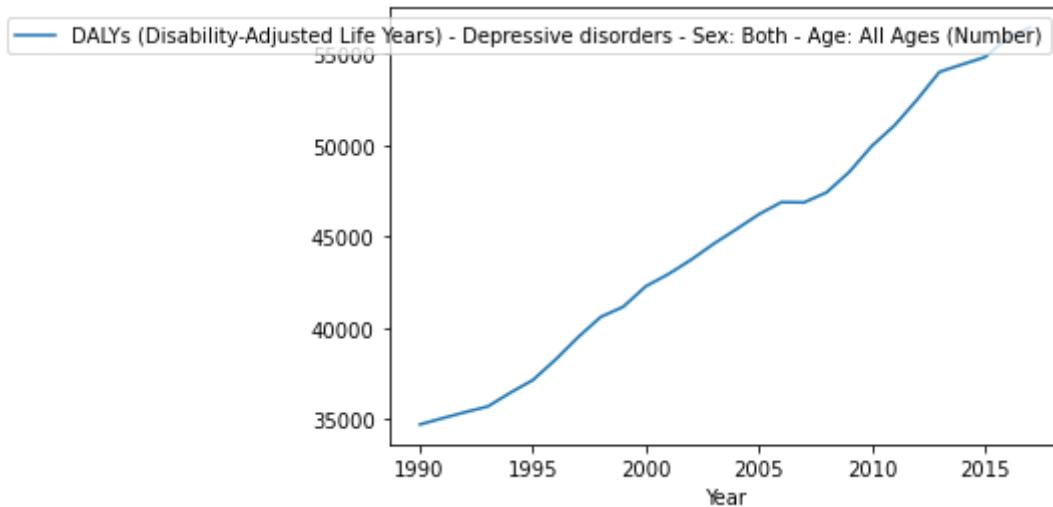
```
In [913]: df18_mean = df18.groupby('Year').mean()
df18_mean.head()
```

```
Out[913]:
```

DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Alcohol use disorders - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)	DALYs (Disability- Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number)
Year					
1990	579860.586727	373623.045716	43089.202737	242096.439055	160910.179278
1991	590696.338993	379694.369656	43855.723861	251147.629100	164425.863171
1992	601428.250402	385911.721961	44619.912002	265003.991459	167926.828374
1993	611830.957962	392154.408513	45390.023357	284054.114991	171460.011399
1994	622083.673166	398501.718223	46160.579683	299086.460524	174937.504893

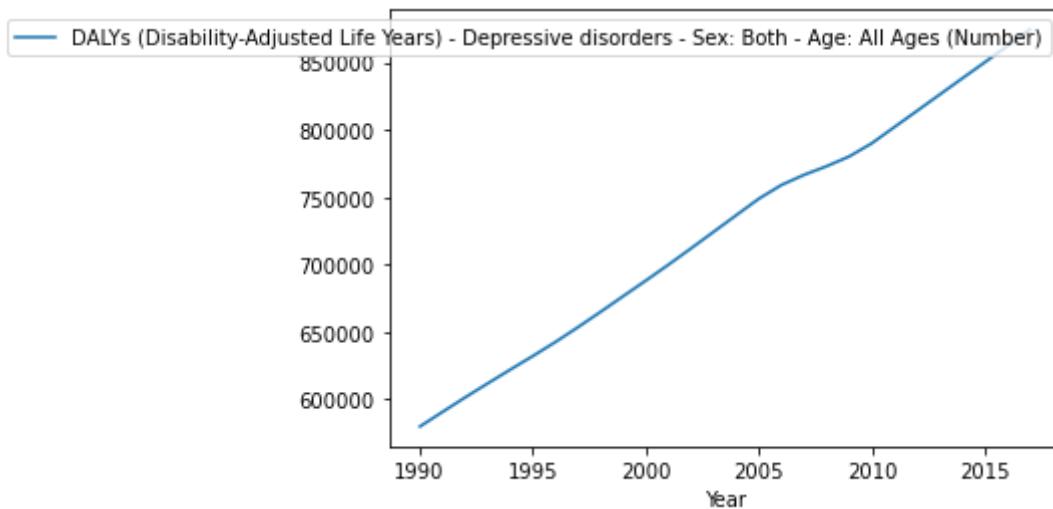
```
In [914]: df18.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[914]: <AxesSubplot:xlabel='Year'>
```



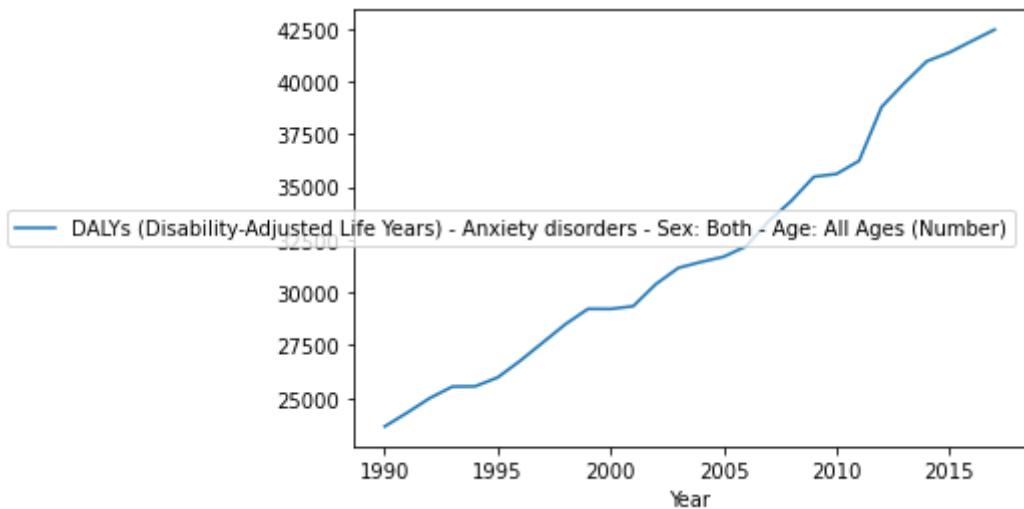
```
In [915]: df18.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[915]: <AxesSubplot:xlabel='Year'>
```



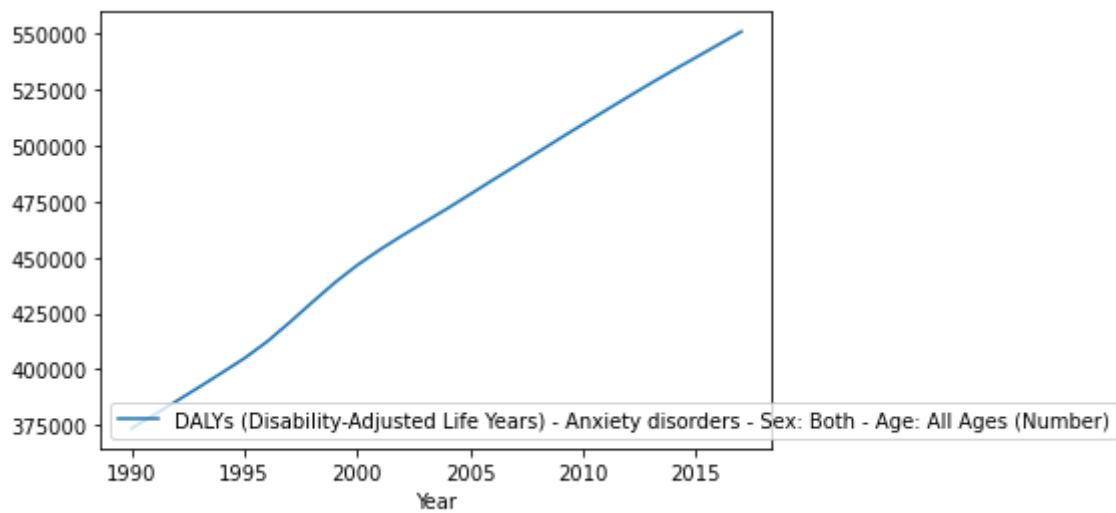
```
In [916]: df18.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[916]: <AxesSubplot:xlabel='Year'>
```



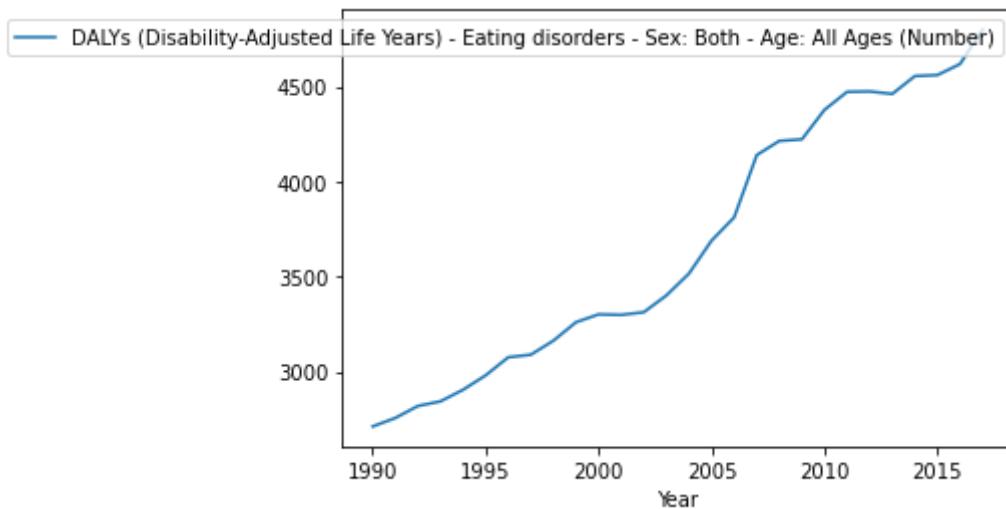
```
In [917]: df18.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[917]: <AxesSubplot:xlabel='Year'>
```



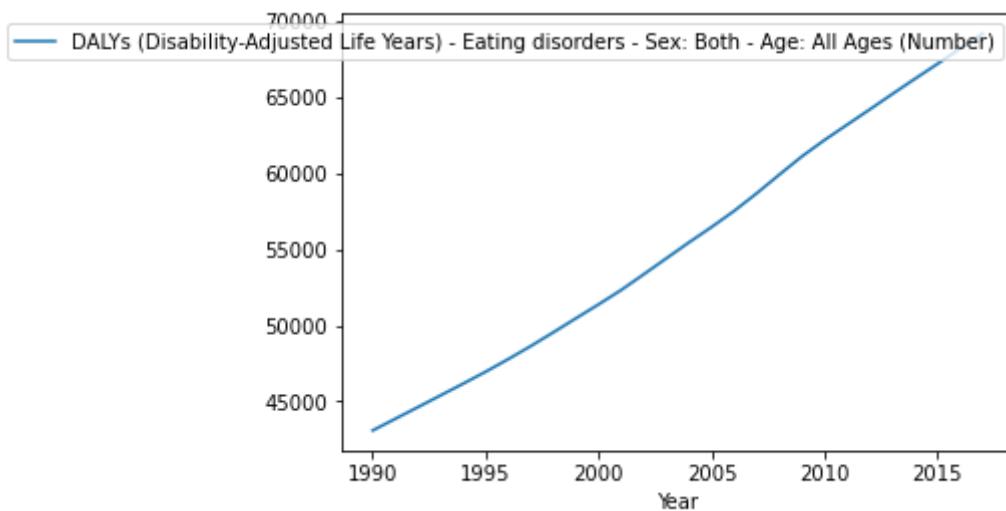
```
In [918]: df18.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[918]: <AxesSubplot:xlabel='Year'>
```



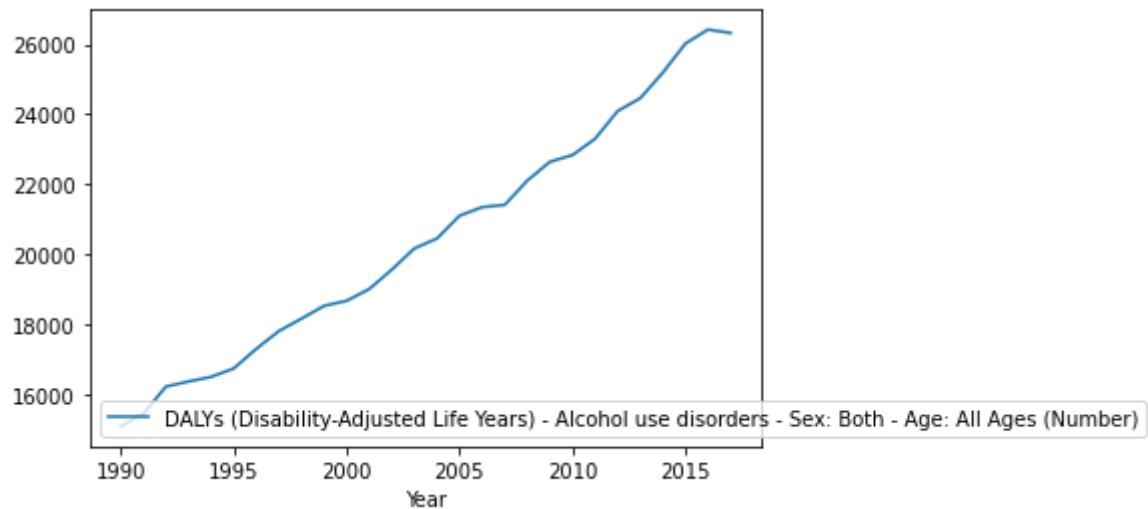
```
In [919]: df18.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[919]: <AxesSubplot:xlabel='Year'>
```



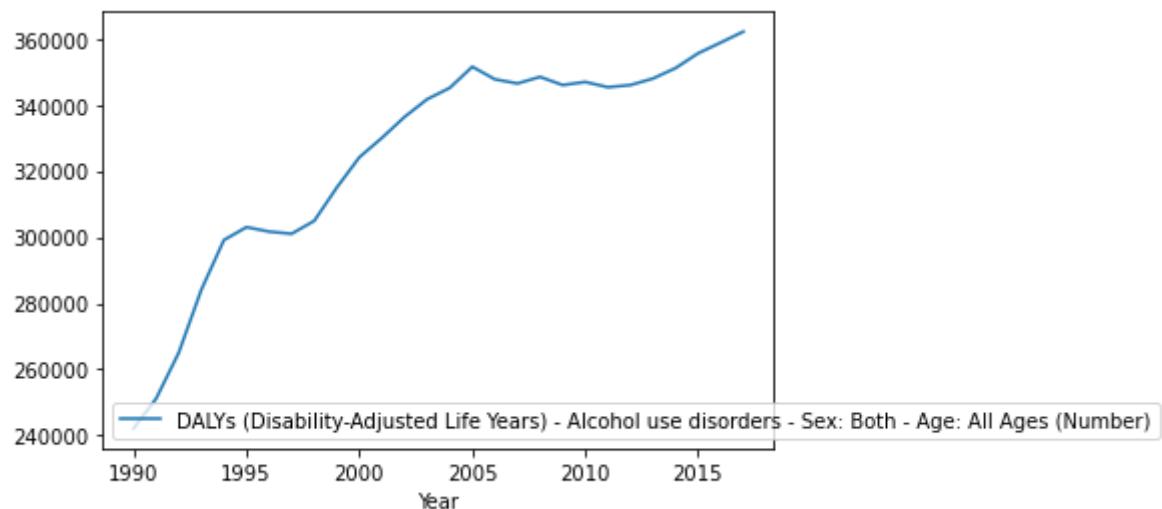
```
In [920]: df18.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[920]: <AxesSubplot:xlabel='Year'>
```



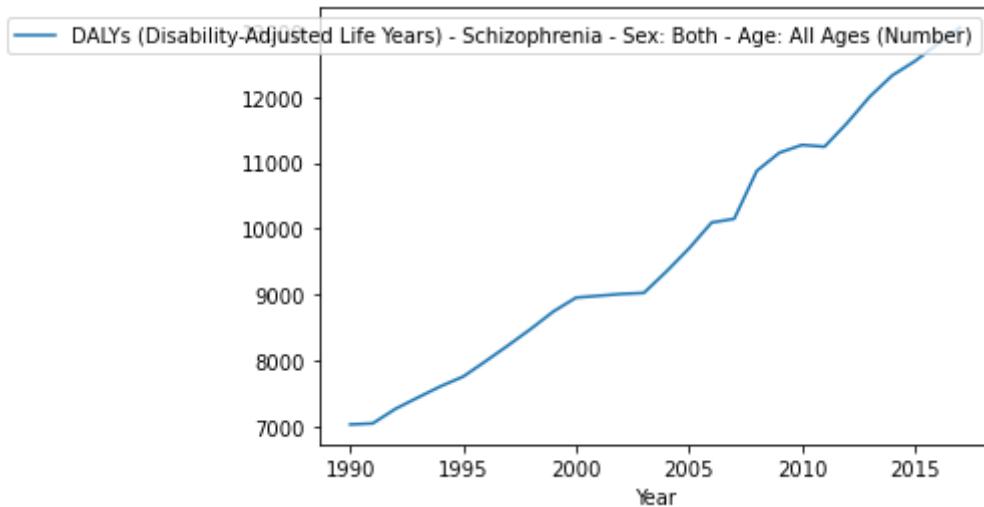
```
In [921]: df18.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[921]: <AxesSubplot:xlabel='Year'>
```



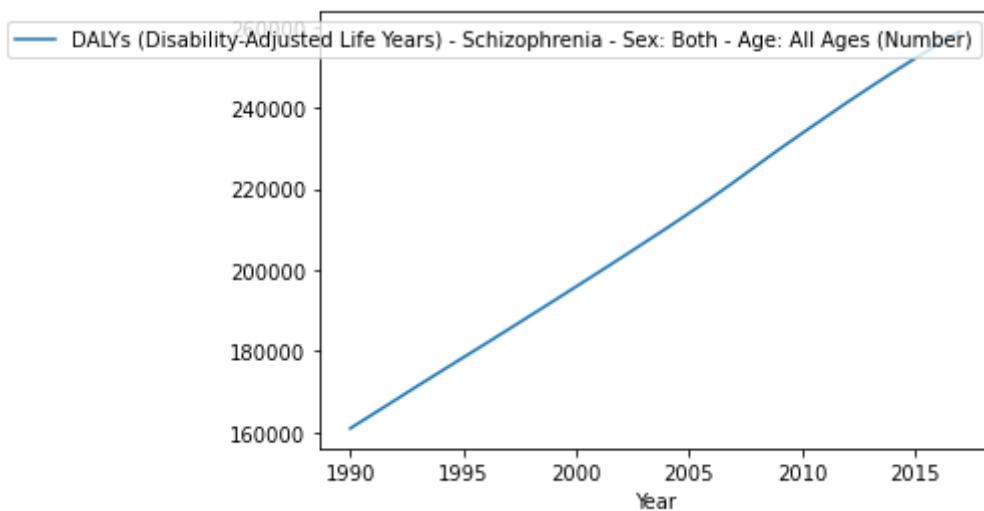
```
In [922]: df18.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[922]: <AxesSubplot:xlabel='Year'>
```



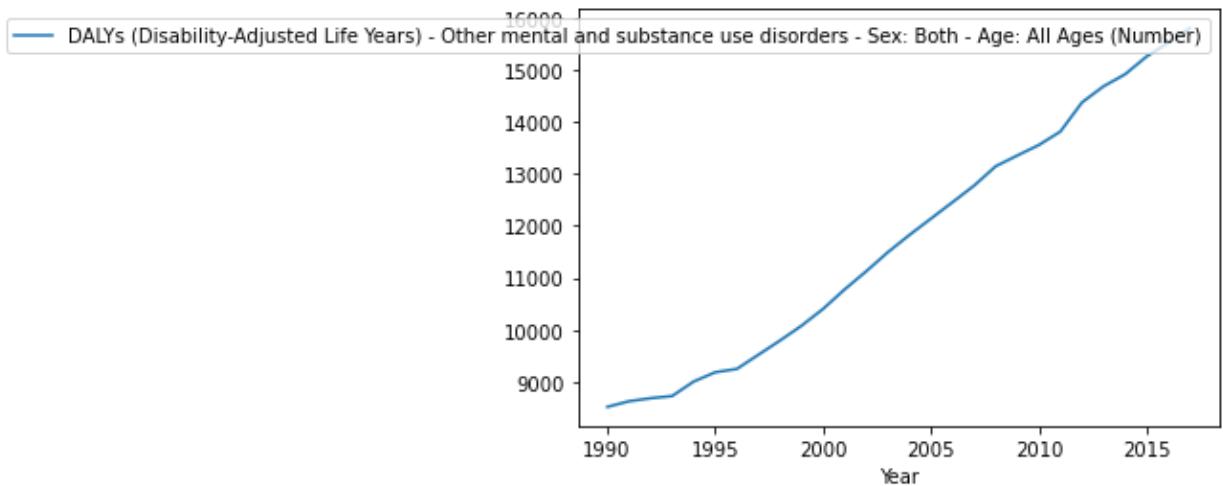
```
In [923]: df18.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[923]: <AxesSubplot:xlabel='Year'>
```



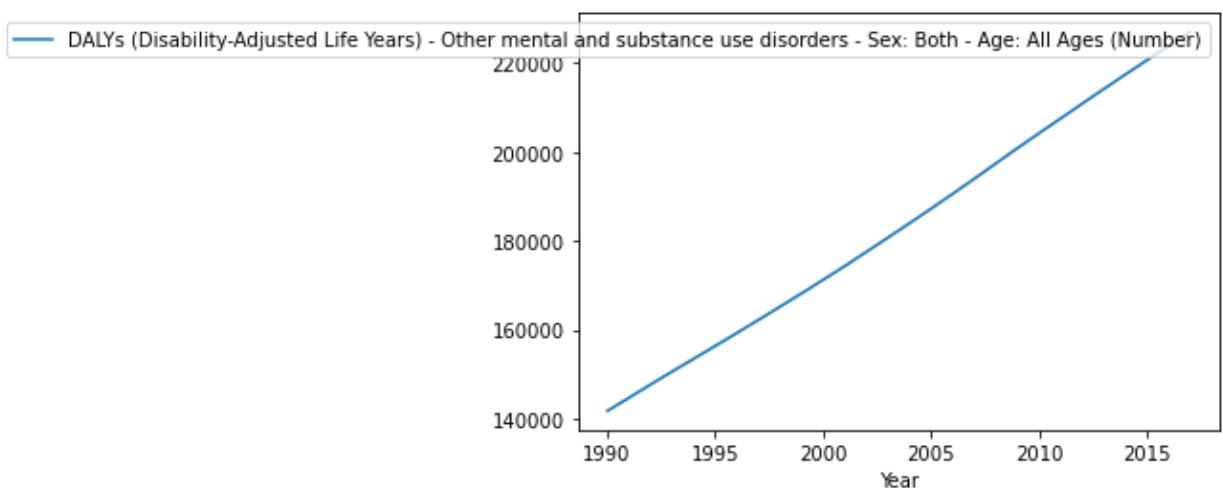
```
In [924]: df18.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[924]: <AxesSubplot:xlabel='Year'>
```



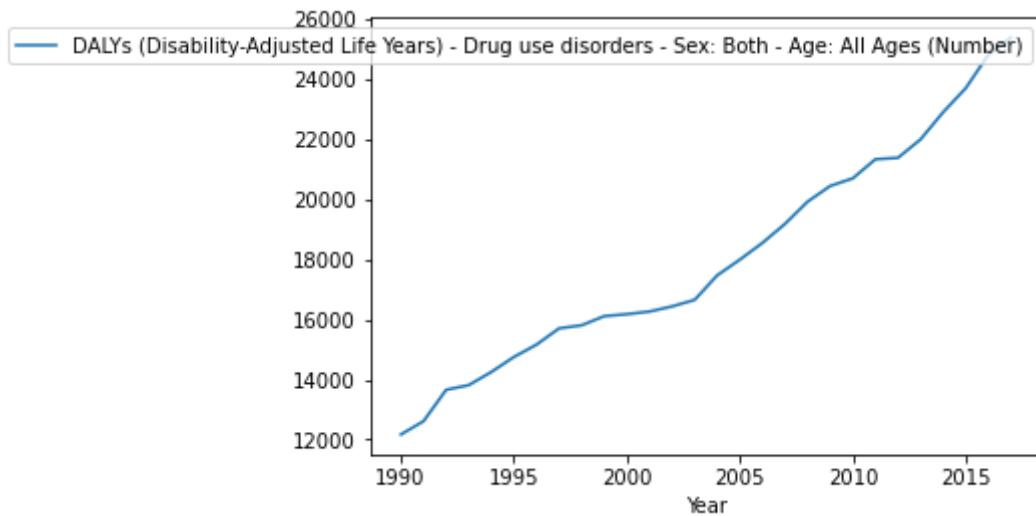
```
In [925]: df18.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[925]: <AxesSubplot:xlabel='Year'>
```



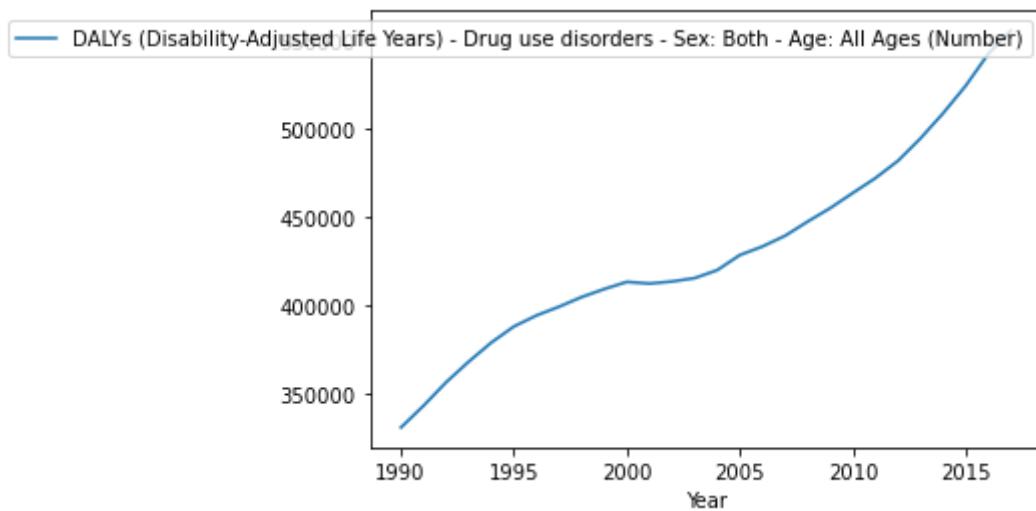
```
In [926]: df18.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[926]: <AxesSubplot:xlabel='Year'>
```



```
In [927]: df18.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[927]: <AxesSubplot:xlabel='Year'>
```



In [928]: df19.info()
df19.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6156 entries, 0 to 6155
Data columns (total 4 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6156 non-null   object
1   Code      object
5292 non-null   object
2   Year      int64
6156 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate) 6156 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 192.5+ KB
```

Out[928]:

	Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)
0	Afghanistan	AFG	1990	1946.055854
1	Afghanistan	AFG	1991	1949.616897
2	Afghanistan	AFG	1992	1953.343213
3	Afghanistan	AFG	1993	1956.545238
4	Afghanistan	AFG	1994	1957.157365

Checking for missing values:

In [929]: missing = pd.concat([df19.isnull().sum(), 100 * df19.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[929]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)		0	0.000000
Code	864	14.035088	

In [930]: v1='DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-s

In [931]: df19.describe()

Out[931]:

	Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)
count	6156.000000	6156.000000
mean	2003.000000	1727.433801
std	7.789514	215.329711
min	1990.000000	1332.128150
25%	1996.000000	1552.941226
50%	2003.000000	1701.216819
75%	2010.000000	1863.561830
max	2016.000000	2445.800370

In [932]: df19.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[932]:

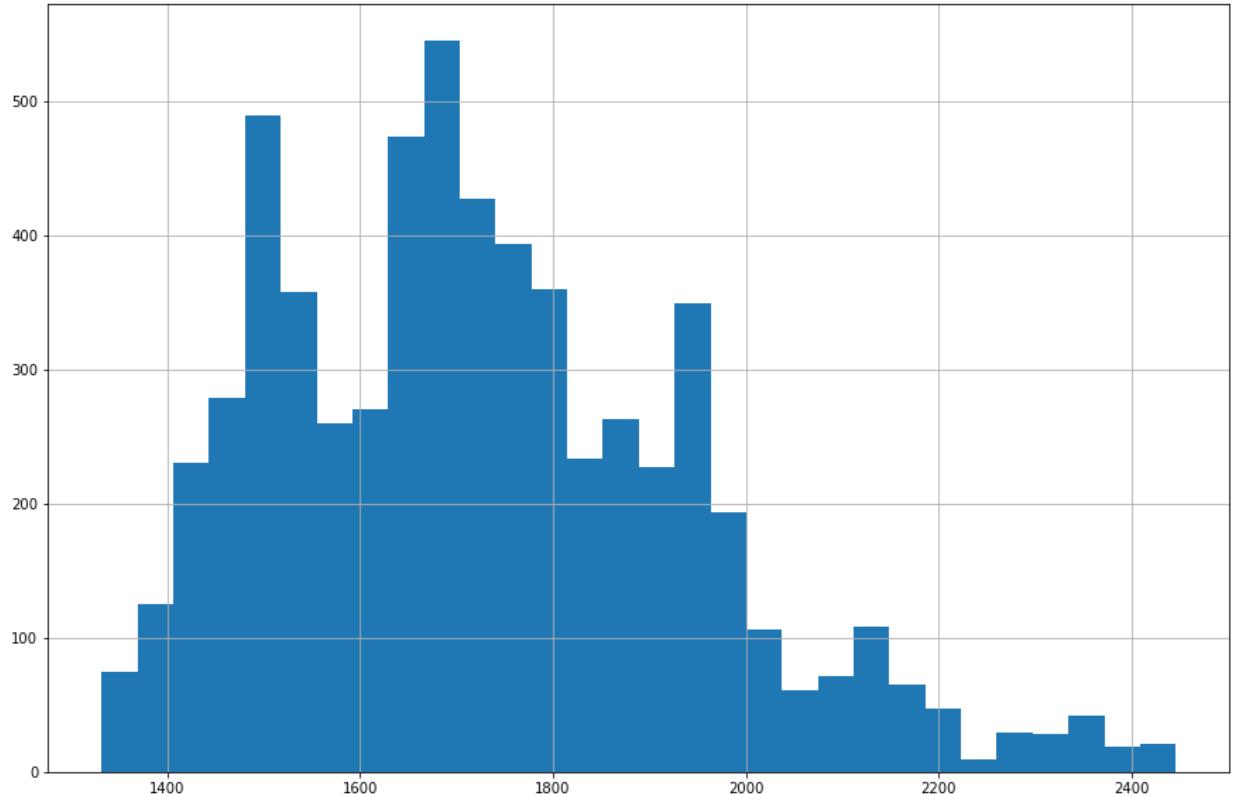
	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)
2165	Greenland	1995	2445.800370
2166	Greenland	1996	2442.580175
2167	Greenland	1997	2441.315404
2168	Greenland	1998	2439.042613
2169	Greenland	1999	2438.442377

In [933]: df19.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[933]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)
3719	Myanmar	2010	1332.128150
3718	Myanmar	2009	1332.181361
3720	Myanmar	2011	1333.359376
3717	Myanmar	2008	1334.460912
3716	Myanmar	2007	1335.691952

```
In [934]: df19[v1].hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [935]: df19.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[935]: Entity  
Myanmar      1340.528995  
Azerbaijan   1354.190913  
Tajikistan    1373.283680  
Armenia       1383.889453  
Albania       1388.646948  
Name: DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate), dtype: float64
```

```
In [936]: df19.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[936]: Entity
New Zealand      2295.714010
Morocco          2308.677455
Australasia      2351.319482
Australia        2362.072300
Greenland         2418.474687
Name: DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate), dtype: float64
```

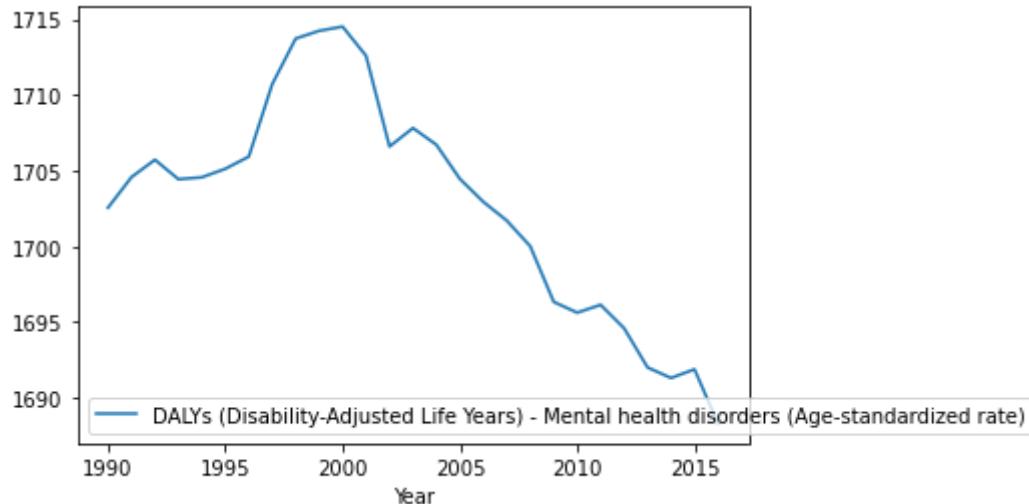
```
In [937]: df19_mean = df19.groupby('Year').mean()
df19_mean.head()
```

```
Out[937]:
DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)
```

Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)
1990	1730.147829
1991	1731.931148
1992	1732.198350
1993	1732.408747
1994	1732.781996

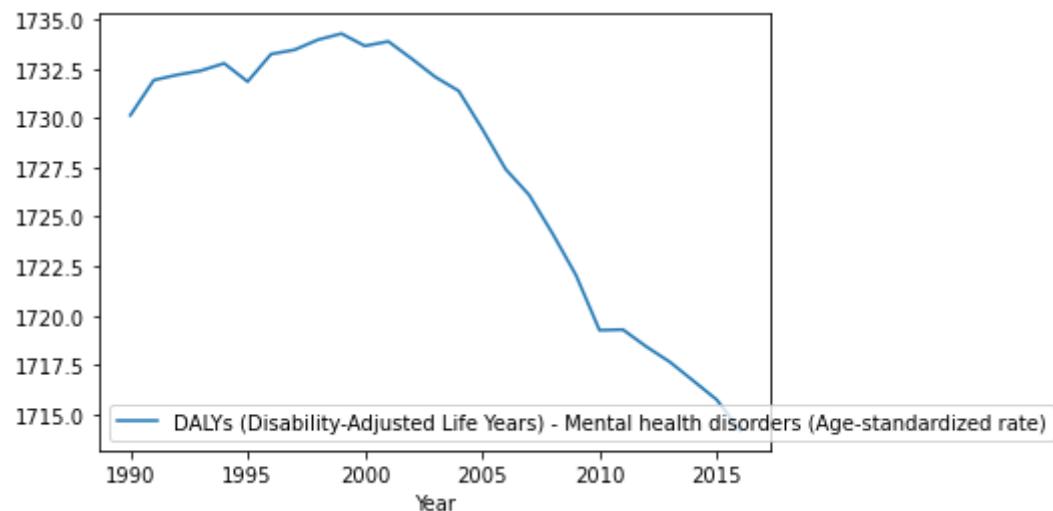
```
In [938]: df19.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[938]: <AxesSubplot:xlabel='Year'>
```



```
In [939]: df19.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[939]: <AxesSubplot:xlabel='Year'>
```



In [940]: df26.info()
df26.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 10 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   Deaths - Mental and substance use disorders - Sex: Both - Age: Under
5 (Rate)           6156 non-null   float64
4   Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ y
ears (Rate)          6156 non-null   float64
5   Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14
years (Rate)          6468 non-null   float64
6   Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49
years (Rate)          6468 non-null   float64
7   Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69
years (Rate)          6156 non-null   float64
8   Deaths - Mental and substance use disorders - Sex: Both - Age: Age-s
tandardized (Rate)  6468 non-null   float64
9   Deaths - Mental and substance use disorders - Sex: Both - Age: All A
ges (Rate)            6468 non-null   float64
dtypes: float64(7), int64(1), object(2)
memory usage: 505.4+ KB
```

Out[940]:

				Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: Age- standardized (Rate)
Entity	Code	Year							
0	Afghanistan	AFG	1990	0.021469	4.115177	0.013152	0.236228	2.773907	0.144067
1	Afghanistan	AFG	1991	0.021909	4.183183	0.012804	0.233502	2.832835	0.140653
2	Afghanistan	AFG	1992	0.022349	4.251692	0.012525	0.231883	2.898354	0.135279
3	Afghanistan	AFG	1993	0.022804	4.350942	0.012949	0.237781	2.973570	0.133047
4	Afghanistan	AFG	1994	0.023260	4.459888	0.013219	0.239914	3.052016	0.130073

Checking for missing values:

```
In [941]: missing = pd.concat([df26.isnull().sum(), 100 * df26.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[941]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)		0	0.000000
Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)		0	0.000000
Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Deaths - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)		0	0.000000
Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)		312	4.823748
Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)		312	4.823748
Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)		312	4.823748
Code		980	15.151515

```
In [942]: v1='Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years'
v2='Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years'
v3='Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized'
v4='Deaths - Mental and substance use disorders - Sex: Both - Age: All Ages'
v5='Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5'
v6='Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years'
v7='Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years'
```

In [943]: df26.describe()

Out[943]:

	Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)	
Year	Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)	
count	6468.000000	6156.000000	6156.000000	6468.000000	6468.000000	6156.000000	6468.000000
mean	2003.500000	0.005314	9.789904	0.497955	10.033683	9.555160	5.014424
std	8.078372	0.011795	7.361838	1.627249	19.715789	11.518568	9.795962
min	1990.000000	0.001263	0.677456	0.007797	0.059130	0.647502	0.034565
25%	1996.750000	0.002807	5.274254	0.023295	0.680799	3.314448	0.356321
50%	2003.500000	0.003492	8.428751	0.044972	2.429893	5.838692	1.254213
75%	2010.250000	0.004615	12.200324	0.341105	8.205105	10.696765	4.235866
max	2017.000000	0.246067	78.482540	24.978386	172.672348	90.302454	89.797874

In [944]: df26.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[944]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)
1945	Estonia	2003	24.978386
1938	Estonia	1996	23.341860
1934	Estonia	1992	22.889428
1944	Estonia	2002	22.599776
1935	Estonia	1993	21.807803

In [945]: df26.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[945]:

Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)				
Entity	Year			
1936	Estonia	1994		172.672348
1937	Estonia	1995		169.660411
1940	Estonia	1998		167.816903
1938	Estonia	1996		162.463238
1939	Estonia	1997		162.105626

In [946]: df26.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[946]:

Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)				
Entity	Year			
1936	Estonia	1994		89.797874
1937	Estonia	1995		87.384741
1940	Estonia	1998		85.146098
1938	Estonia	1996		83.704111
1941	Estonia	1999		82.708846

In [947]: df26.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[947]:

Deaths - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)				
Entity	Year			
1936	Estonia	1994		85.909921
1937	Estonia	1995		84.829013
1940	Estonia	1998		84.728868
1941	Estonia	1999		82.545207
1938	Estonia	1996		82.227912

In [948]: df26.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[948]:

Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)				
Entity	Year			
2755	Iran	2001		0.246067
2754	Iran	2000		0.240016
2753	Iran	1999		0.236581
2756	Iran	2002		0.232140
2752	Iran	1998		0.231983

In [949]: df26.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[949]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)
2283	Greenland	2005	78.482540
2282	Greenland	2004	77.556472
2281	Greenland	2003	75.583326
2284	Greenland	2006	75.581044
2280	Greenland	2002	74.463597

In [950]: df26.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[950]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)
4717	Russia	2003	90.302454
4716	Russia	2002	89.978610
4718	Russia	2004	87.333464
4715	Russia	2001	86.062385
4719	Russia	2005	84.216935

In [951]: df26.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[951]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)
896	Cambodia	1990	0.007797
897	Cambodia	1991	0.007876
898	Cambodia	1992	0.007982
899	Cambodia	1993	0.008118
900	Cambodia	1994	0.008264

In [952]: df26.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[952]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)
896	Cambodia	1990	0.059130
897	Cambodia	1991	0.060462
898	Cambodia	1992	0.062107
5740	Timor	1990	0.063753
899	Cambodia	1993	0.063992

In [953]: df26.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[953]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)
896	Cambodia	1990	0.034565
5740	Timor	1990	0.035158
897	Cambodia	1991	0.035242
898	Cambodia	1992	0.036097
5741	Timor	1991	0.036197

In [954]: df26.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[954]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)
896	Cambodia	1990	0.028276
897	Cambodia	1991	0.028918
898	Cambodia	1992	0.029746
899	Cambodia	1993	0.030757
900	Cambodia	1994	0.031744

In [955]: df26.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[955]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)
5124	Solomon Islands	1990	0.001263
5125	Solomon Islands	1991	0.001273
5126	Solomon Islands	1992	0.001281
5127	Solomon Islands	1993	0.001301
5128	Solomon Islands	1994	0.001315

In [956]: df26.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[956]:

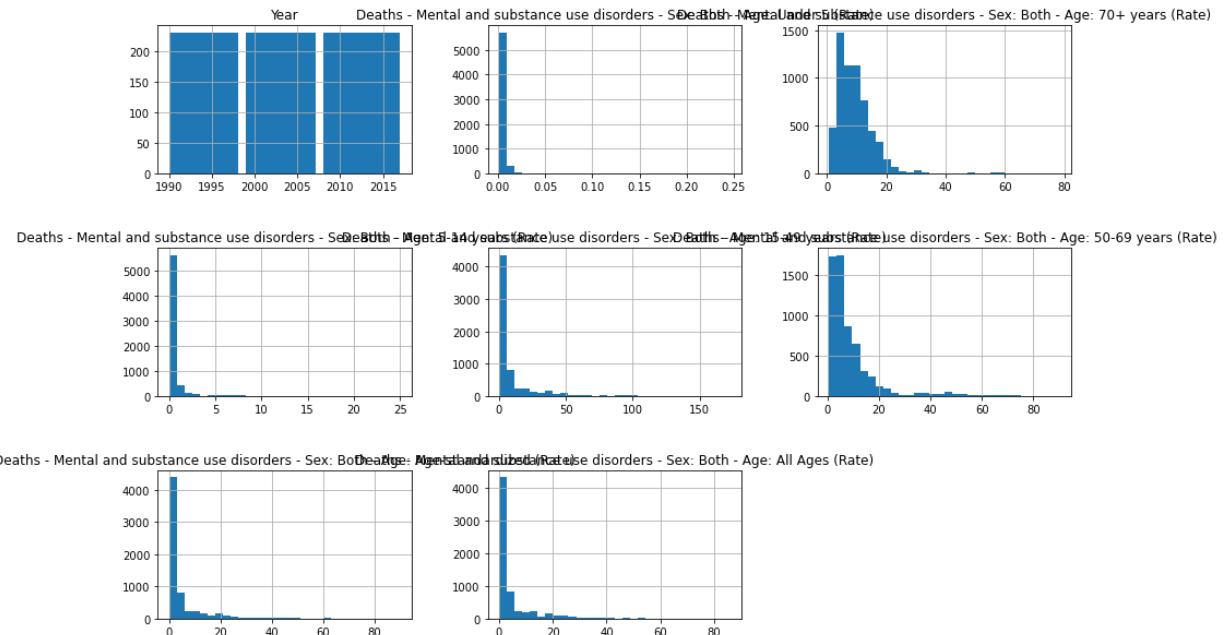
Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)				
	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	
1792	Egypt	1990	0.677456	
1794	Egypt	1992	0.681267	
1793	Egypt	1991	0.684990	
1795	Egypt	1993	0.693891	
1802	Egypt	2000	0.698001	

In [957]: df26.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[957]:

	Entity	Year	Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)
5051	Singapore	2001	0.647502
5050	Singapore	2000	0.647876
5052	Singapore	2002	0.651808
1794	Egypt	1992	0.656633
5053	Singapore	2003	0.658803

In [958]: df26.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);



```
In [959]: df26.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[959]: Entity
Cambodia      0.010250
Burkina Faso  0.012544
Nepal          0.013434
Niger          0.014911
Belize         0.015583
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14
years (Rate), dtype: float64
```

```
In [960]: df26.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[960]: Entity
Cambodia      0.097918
Timor          0.133953
Maldives       0.141827
Laos           0.157656
Yemen          0.178308
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 15-4
9 years (Rate), dtype: float64
```

```
In [961]: df26.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[961]: Entity
Cambodia      0.054266
Timor          0.073120
Maldives       0.079104
Laos           0.087595
Yemen          0.097346
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: Age-
standardized (Rate), dtype: float64
```

```
In [962]: df26.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[962]: Entity
Cambodia      0.051882
Timor          0.066966
Maldives       0.083150
Yemen          0.083336
Laos           0.084009
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: All
Ages (Rate), dtype: float64
```

```
In [963]: df26.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[963]: Entity
Solomon Islands 0.001656
Andorra         0.001666
Nepal           0.001667
India            0.001731
South Sudan     0.002001
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: Unde
r 5 (Rate), dtype: float64
```

```
In [964]: df26.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[964]: Entity
Egypt      0.721018
Singapore  0.993769
Greece     1.095144
Brunei     1.229526
Japan      1.952071
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 70+
years (Rate), dtype: float64
```

```
In [965]: df26.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[965]: Entity
Egypt      0.712494
Singapore  0.720052
Brunei     0.781039
Colombia   0.986059
Fiji       1.038783
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 50-6
9 years (Rate), dtype: float64
```

```
In [966]: df26.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[966]: Entity
Moldova        5.982458
Brunei         6.312190
High-income Asia Pacific  6.885188
Japan          8.012651
Estonia        18.841388
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14
years (Rate), dtype: float64
```

```
In [967]: df26.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[967]: Entity
High-income Asia Pacific  71.196652
Japan                 84.035594
Luxembourg            102.406647
Switzerland            119.986914
Estonia                137.139554
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 15-4
9 years (Rate), dtype: float64
```

```
In [968]: df26.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[968]: Entity
High-income Asia Pacific  35.694578
Japan                  41.947824
Luxembourg             50.558816
Switzerland            57.501724
Estonia                 70.641484
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: Age-
standardized (Rate), dtype: float64
```

```
In [969]: df26.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[969]: Entity
High-income Asia Pacific      35.511929
Japan                         39.147305
Luxembourg                     52.668084
Switzerland                    59.960059
Estonia                        68.601637
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: All
Ages (Rate), dtype: float64
```

```
In [970]: df26.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[970]: Entity
North Africa and Middle East          0.029543
Iran                                    0.162064
Central Europe, Eastern Europe, and Central Asia    NaN
High-income                           NaN
Southeast Asia, East Asia, and Oceania        NaN
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: Unde
r 5 (Rate), dtype: float64
```

```
In [971]: df26.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[971]: Entity
Azerbaijan                          50.488958
Greenland                            65.610658
Central Europe, Eastern Europe, and Central Asia    NaN
High-income                           NaN
Southeast Asia, East Asia, and Oceania        NaN
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 70+
years (Rate), dtype: float64
```

```
In [972]: df26.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[972]: Entity
Greenland                           60.972850
Russia                             67.895792
Central Europe, Eastern Europe, and Central Asia    NaN
High-income                           NaN
Southeast Asia, East Asia, and Oceania        NaN
Name: Deaths - Mental and substance use disorders - Sex: Both - Age: 50-6
9 years (Rate), dtype: float64
```

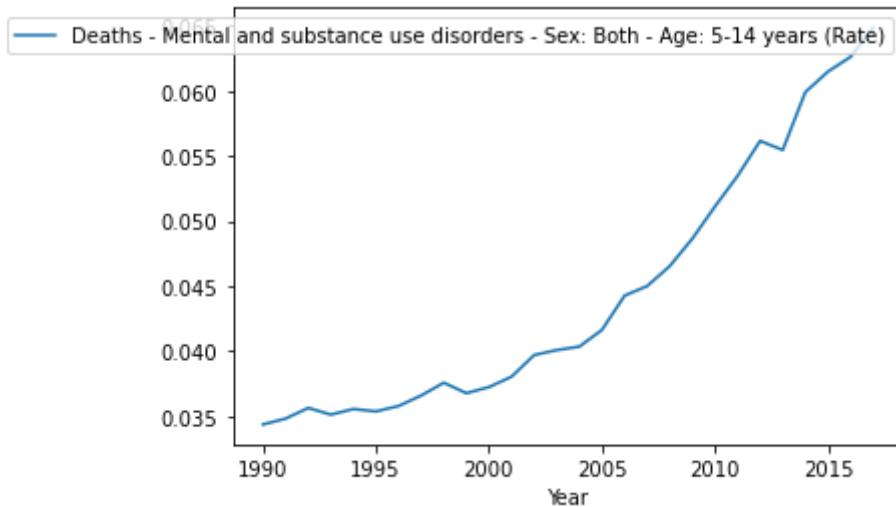
```
In [973]: df26_mean = df26.groupby('Year').mean()
df26_mean.head()
```

Out[973]:

	Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)	Deaths - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)
Year							
1990	0.005538	9.210626	0.418176	8.083244	9.406961	4.103350	4.212572
1991	0.005406	9.245414	0.431564	8.212001	9.514974	4.169240	4.278296
1992	0.005319	9.333661	0.450859	8.347589	9.655258	4.233742	4.346789
1993	0.005255	9.424439	0.466389	8.790168	9.898809	4.448892	4.566114
1994	0.005233	9.546360	0.473220	9.143256	10.133364	4.618771	4.744153

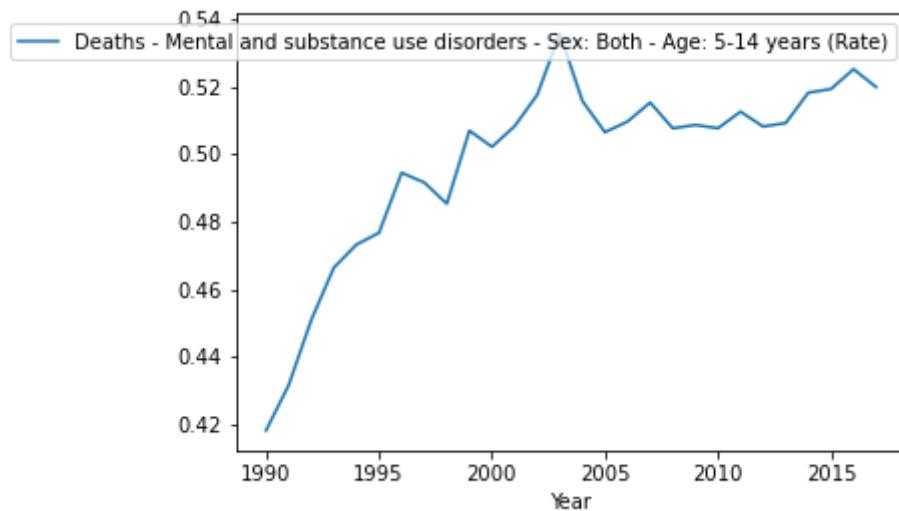
```
In [974]: df26.groupby('Year')[v1].median().plot(legend=True)
```

Out[974]: <AxesSubplot:xlabel='Year'>



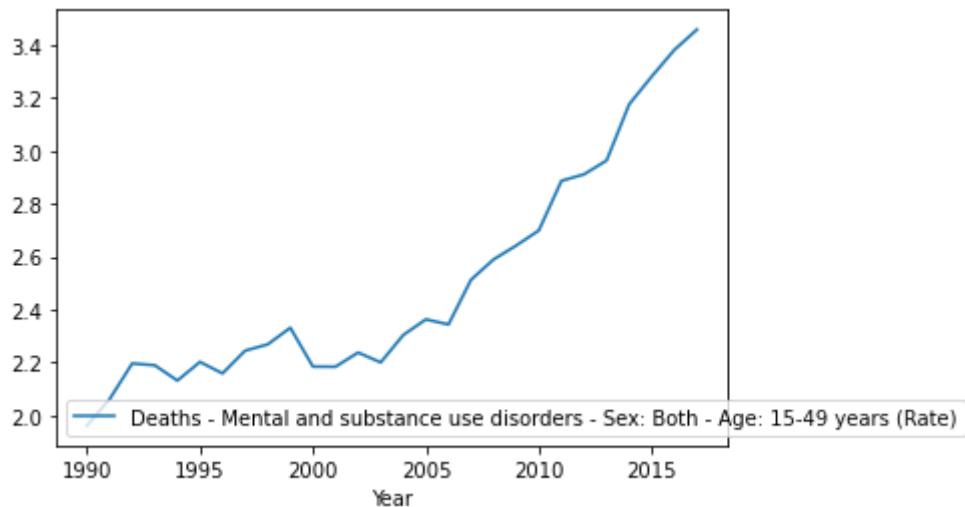
```
In [975]: df26.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[975]: <AxesSubplot:xlabel='Year'>
```



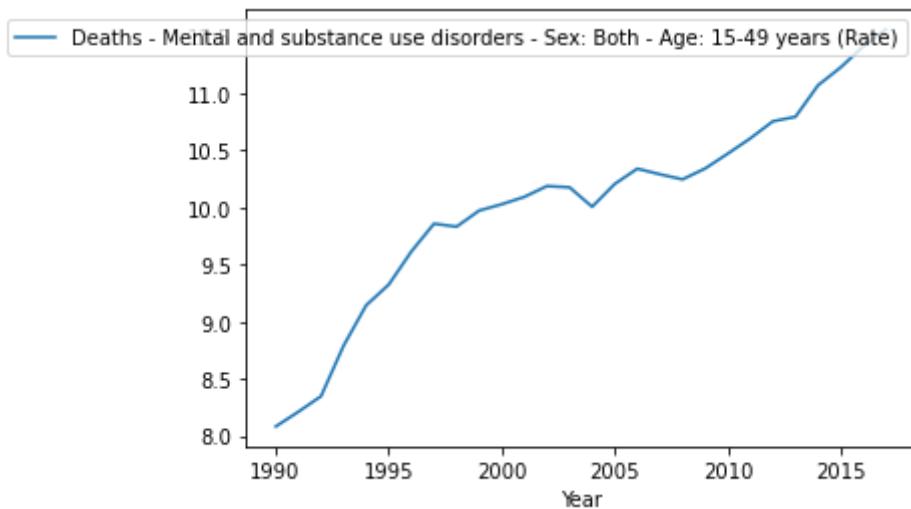
```
In [976]: df26.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[976]: <AxesSubplot:xlabel='Year'>
```



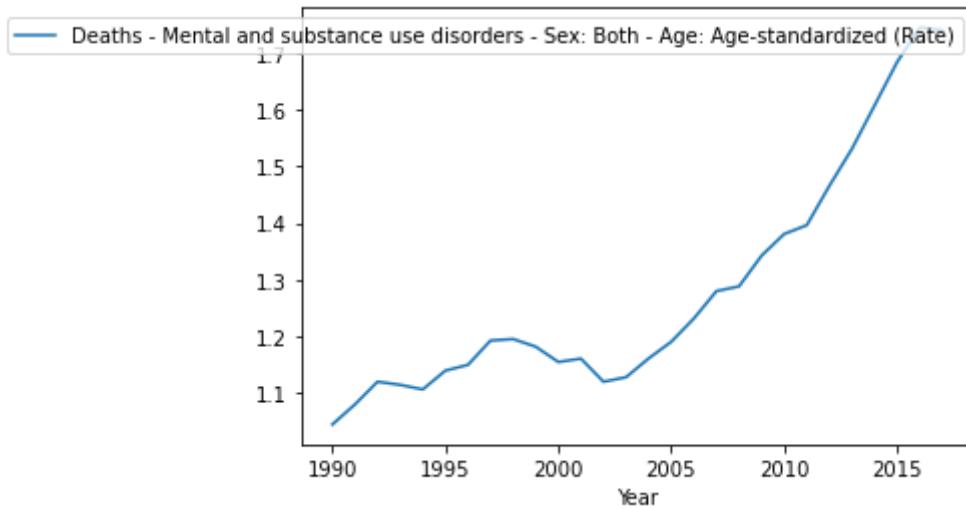
```
In [977]: df26.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[977]: <AxesSubplot:xlabel='Year'>
```



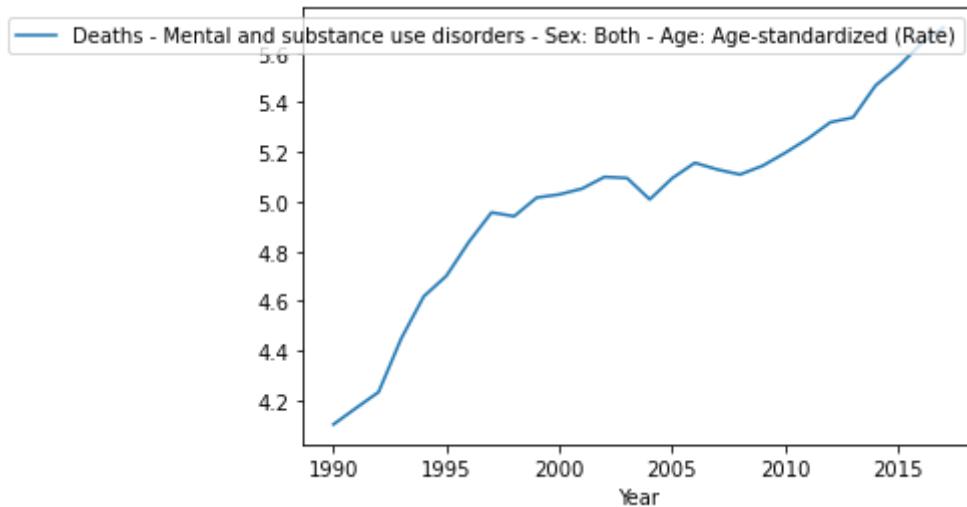
```
In [978]: df26.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[978]: <AxesSubplot:xlabel='Year'>
```



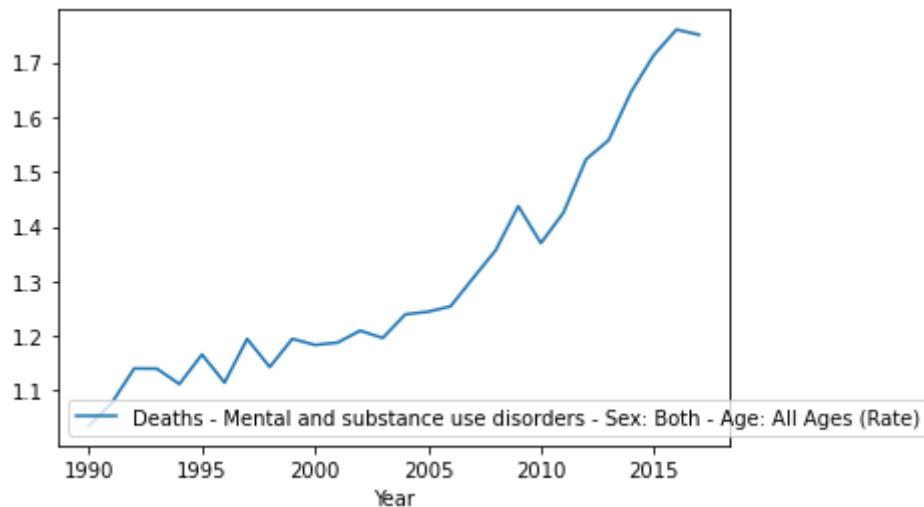
```
In [979]: df26.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[979]: <AxesSubplot:xlabel='Year'>
```



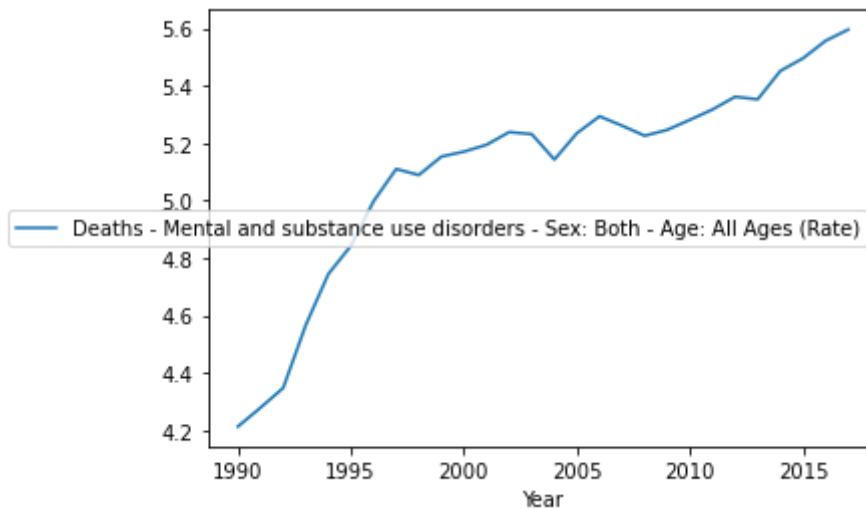
```
In [980]: df26.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[980]: <AxesSubplot:xlabel='Year'>
```



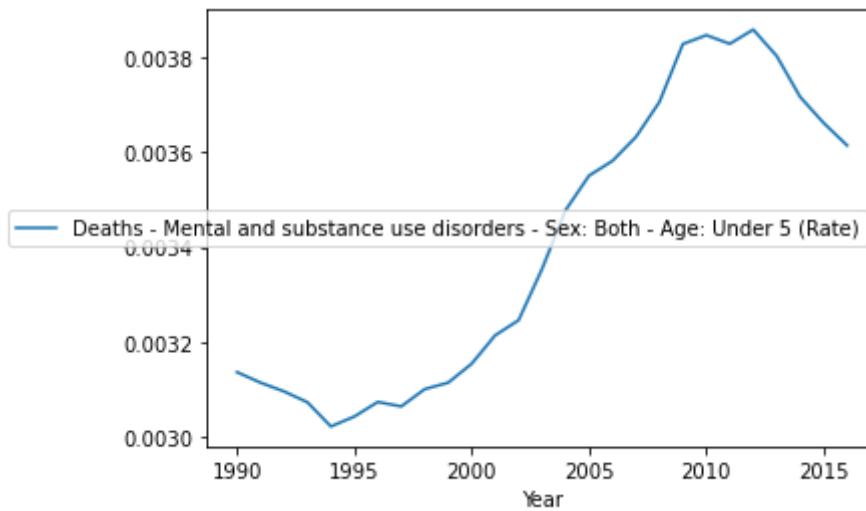
```
In [981]: df26.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[981]: <AxesSubplot:xlabel='Year'>
```



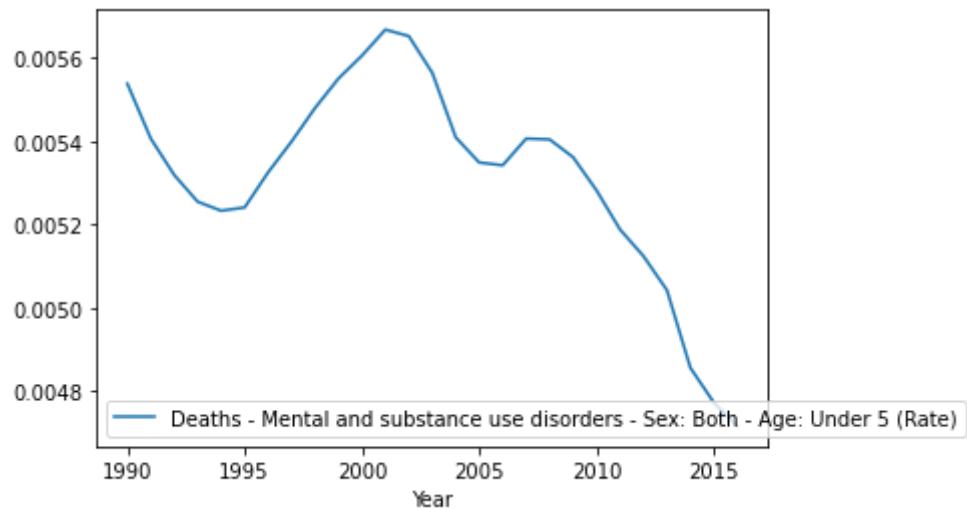
```
In [982]: df26.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[982]: <AxesSubplot:xlabel='Year'>
```



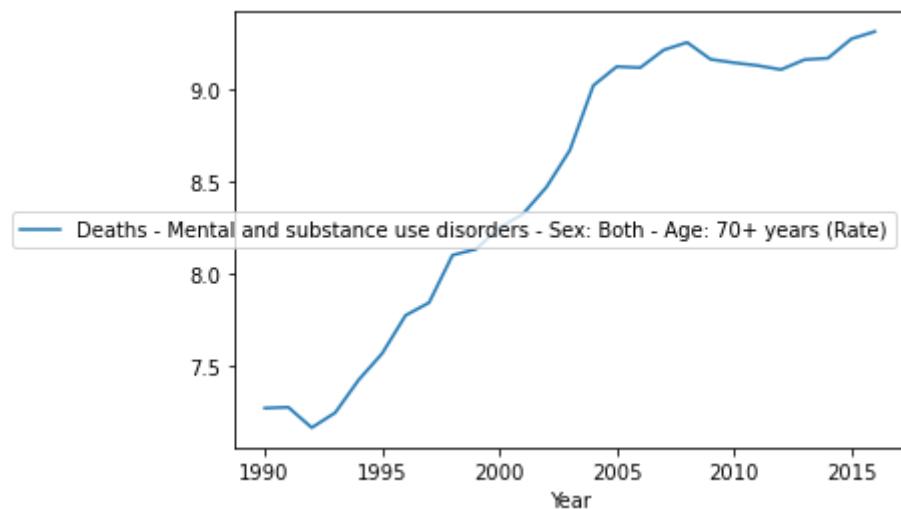
```
In [983]: df26.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[983]: <AxesSubplot:xlabel='Year'>
```



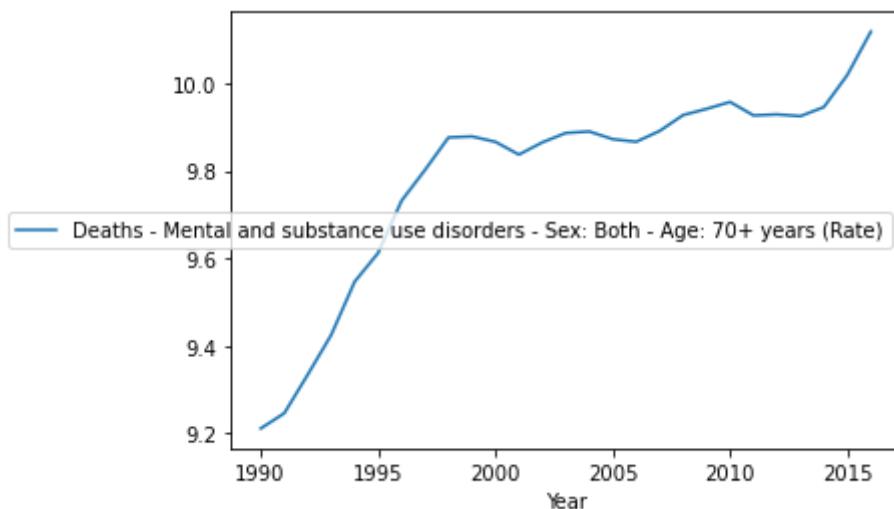
```
In [984]: df26.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[984]: <AxesSubplot:xlabel='Year'>
```



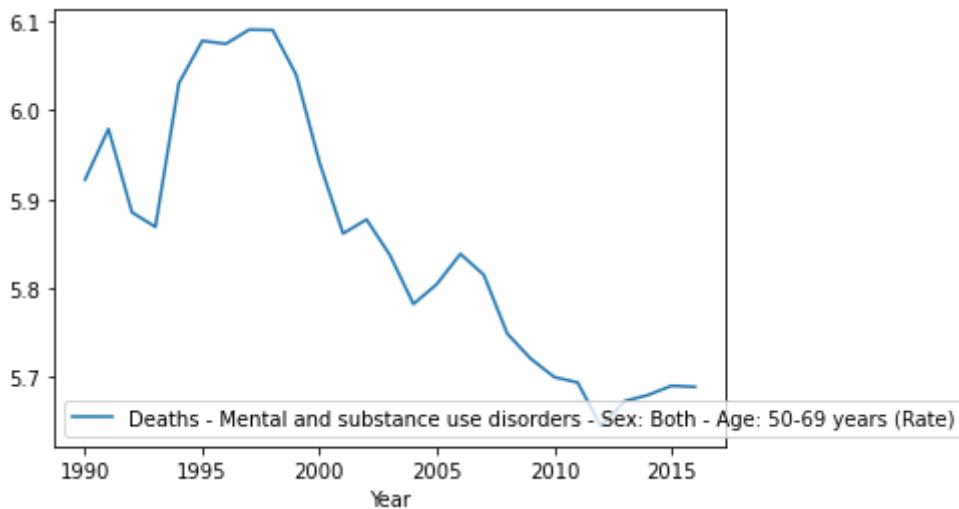
```
In [985]: df26.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[985]: <AxesSubplot:xlabel='Year'>
```



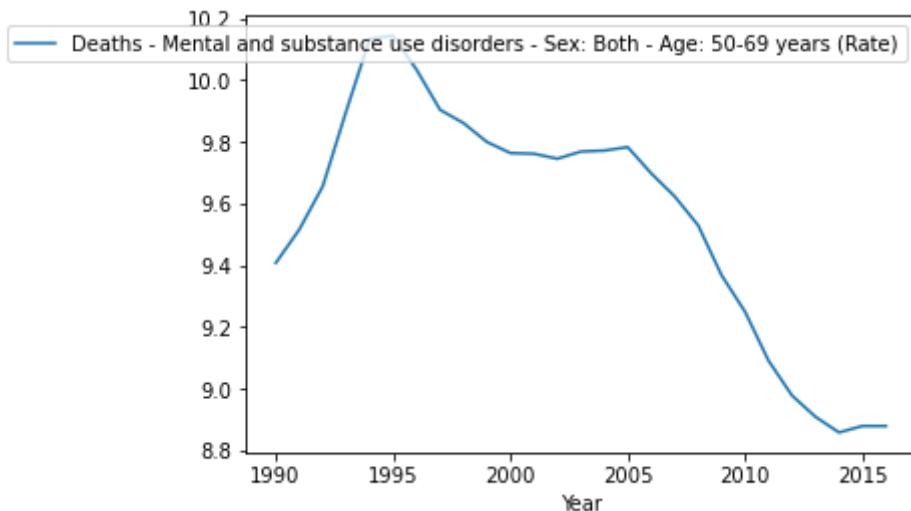
```
In [986]: df26.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[986]: <AxesSubplot:xlabel='Year'>
```



In [987]: `df26.groupby('Year')[v7].mean().plot(legend=True)`

Out[987]: <AxesSubplot:xlabel='Year'>



In [988]: `df48.info()`
`df48.head()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6156 entries, 0 to 6155
Data columns (total 4 columns):
 #   Column
 Non-Null Count  Dtype  
 ---  --  
 0   Entity    object 
 1   Code      object 
 2   Year      int64  
 3   DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent) float64 
dtypes: float64(1), int64(1), object(2)
memory usage: 192.5+ KB
```

Out[988]:

	Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)
0	Afghanistan	AFG	1990	1.702507
1	Afghanistan	AFG	1991	1.724710
2	Afghanistan	AFG	1992	1.776862
3	Afghanistan	AFG	1993	1.805277
4	Afghanistan	AFG	1994	1.783730

Checking for missing values:

```
In [989]: missing = pd.concat([df48.isnull().sum(), 100 * df48.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[989]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)		0	0.000000
Code		864	14.035088

```
In [990]: v1='DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)'
```

```
In [991]: df48.describe()
```

Out[991]:

Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)
count	6156.000000
mean	2003.000000
std	7.789514
min	1990.000000
25%	1996.000000
50%	2003.000000
75%	2010.000000
max	2016.000000

```
In [992]: df48.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[992]:

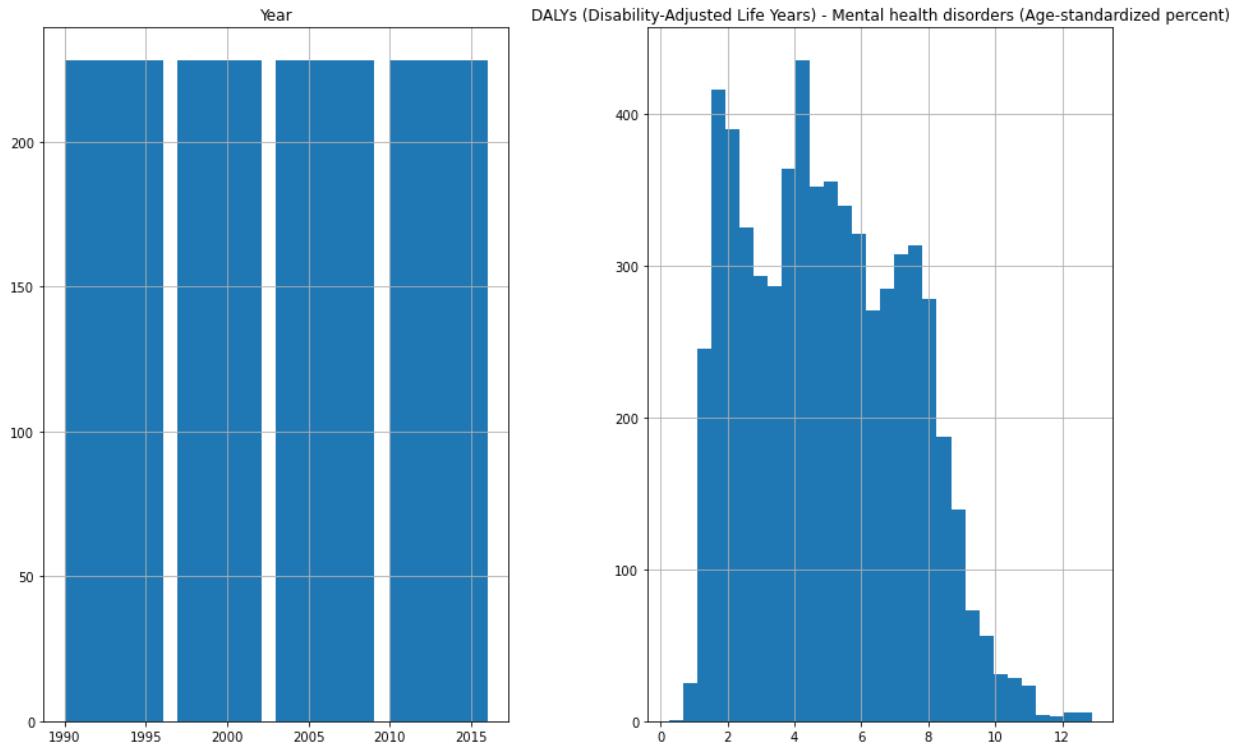
Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)
4453	Qatar	2015
4451	Qatar	2013
4452	Qatar	2014
4454	Qatar	2016
4450	Qatar	2012

In [993]: `df48.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()`

Out[993]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)	
4513	Rwanda	1994	0.258215
3861	Niger	1990	0.886586
3862	Niger	1991	0.901711
2369	Haiti	2010	0.916272
3863	Niger	1992	0.919022

In [994]: `df48.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



In [995]: `df48.groupby('Entity')[v1].mean().sort_values().head()`

Out[995]: Entity

Mali	1.357239
Niger	1.362908
Malawi	1.451206
Central African Republic	1.505243
South Sudan	1.507401

Name: DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent), dtype: float64

```
In [996]: df48.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[996]: Entity
Bahrain      9.616825
Australasia   10.236805
Australia     10.374070
Qatar         10.480428
Kuwait        10.865734
Name: DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent), dtype: float64
```

```
In [997]: df48_mean = df48.groupby('Year').mean()
df48_mean.head()
```

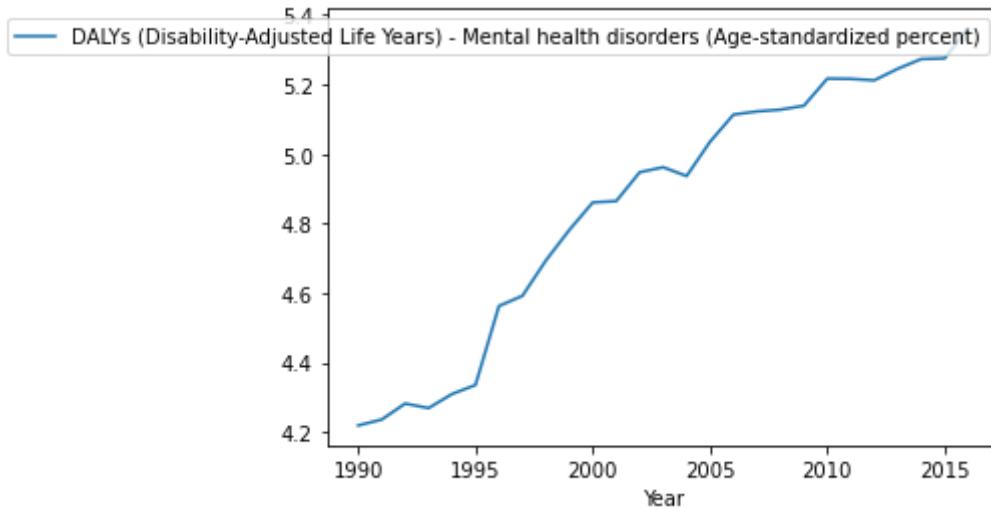
```
Out[997]:
```

DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)

Year	
1990	4.274255
1991	4.329026
1992	4.381049
1993	4.418688
1994	4.457298

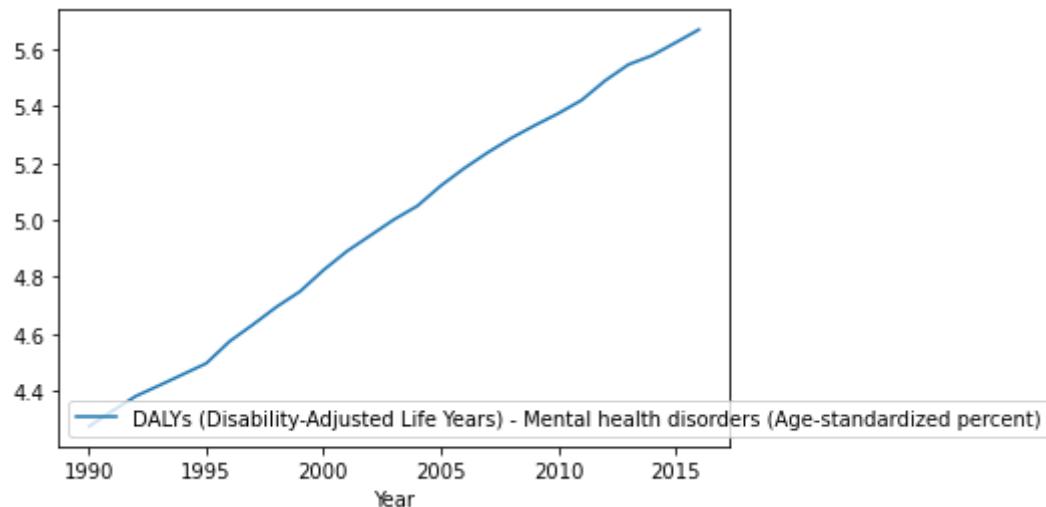
```
In [998]: df48.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[998]: <AxesSubplot:xlabel='Year'>
```



```
In [999]: df48.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[999]: <AxesSubplot:xlabel='Year'>
```



In [1000]: df64.info()
df64.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number) float64
4   Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number) float64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1000]:

	Entity	Code	Year	Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)	Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)
0	Afghanistan	AFG	1990	8.353371e+05	7.941608e+05
1	Afghanistan	AFG	1991	8.831918e+05	8.373352e+05
2	Afghanistan	AFG	1992	1.059854e+06	1.003135e+06
3	Afghanistan	AFG	1993	1.249472e+06	1.178946e+06
4	Afghanistan	AFG	1994	1.314736e+06	1.239610e+06

Checking for missing values:

```
In [1001]: missing = pd.concat([df64.isnull().sum(), 100 * df64.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1001]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)		0	0.000000
Code	980	15.151515	

```
In [1002]: v1='Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)'
v2='Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)'
```

Out[1003]:

Year	Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)	Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)
count	6468.000000	6.468000e+03
mean	2003.500000	8.046755e+06
std	8.078372	3.176805e+07
min	1990.000000	2.292667e+03
25%	1996.750000	1.438431e+05
50%	2003.500000	5.311530e+05
75%	2010.250000	2.514502e+06
max	2017.000000	4.672594e+08

```
In [1004]: df64.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[1004]:

Entity	Year	Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)
6383	World	2017
6382	World	2016
6381	World	2015
6380	World	2014
6379	World	2013

In [1005]: df64.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1005]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)
6383	World	2017	5.035530e+08
6382	World	2016	4.976906e+08
6381	World	2015	4.918710e+08
6380	World	2014	4.861023e+08
6379	World	2013	4.801523e+08

In [1006]: df64.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1006]:

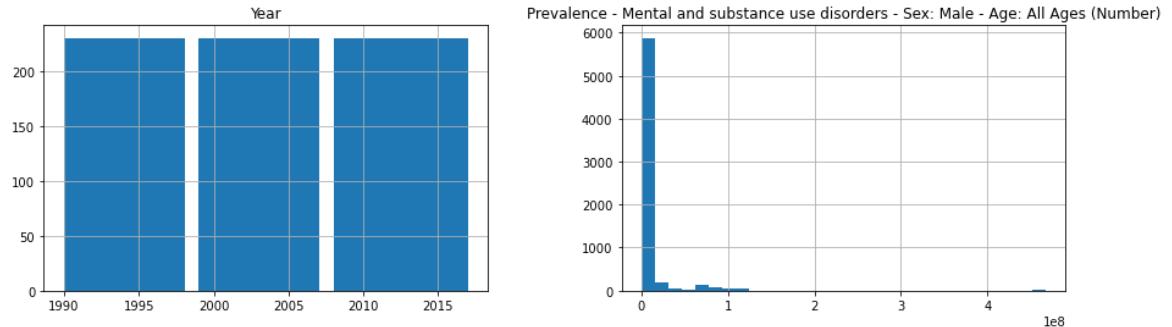
	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)
3584	Marshall Islands	1990	2292.667281
3585	Marshall Islands	1991	2349.940018
84	American Samoa	1990	2355.249302
3586	Marshall Islands	1992	2406.925085
85	American Samoa	1991	2435.044253

In [1007]: df64.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

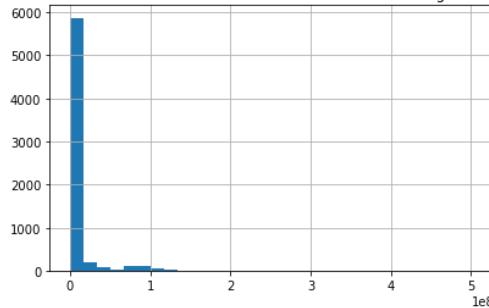
Out[1007]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)
3584	Marshall Islands	1990	2091.609955
3585	Marshall Islands	1991	2146.340355
4256	Northern Mariana Islands	1990	2167.980950
3586	Marshall Islands	1992	2201.154946
84	American Samoa	1990	2223.216328

```
In [1008]: df64.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)



```
In [1009]: df64.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1009]: Entity
Marshall Islands      2760.462926
American Samoa        2776.604583
Northern Mariana Islands 2924.655036
Bermuda                4060.131010
Seychelles              4508.553865
Name: Prevalence - Mental and substance use disorders - Sex: Male - Age:
All Ages (Number), dtype: float64
```

```
In [1010]: df64.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1010]: Entity
Marshall Islands      2565.169602
American Samoa        2719.505010
Northern Mariana Islands 2937.847718
Dominica                4545.352256
Seychelles              4644.982848
Name: Prevalence - Mental and substance use disorders - Sex: Female - Ag
e: All Ages (Number), dtype: float64
```

```
In [1011]: df64.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1011]: Entity
Low-middle SDI           8.630404e+07
South Asia                1.004193e+08
Middle SDI                1.051778e+08
Southeast Asia, East Asia, and Oceania 1.094321e+08
World                      3.984913e+08
Name: Prevalence - Mental and substance use disorders - Sex: Male - Age:
All Ages (Number), dtype: float64
```

```
In [1012]: df64.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1012]: Entity
Low-middle SDI           8.702137e+07
South Asia                9.661847e+07
Middle SDI                1.089256e+08
Southeast Asia, East Asia, and Oceania 1.137032e+08
World                      4.239598e+08
Name: Prevalence - Mental and substance use disorders - Sex: Female - Ag
e: All Ages (Number), dtype: float64
```

```
In [1013]: df64_mean = df64.groupby('Year').mean()
df64_mean.head()
```

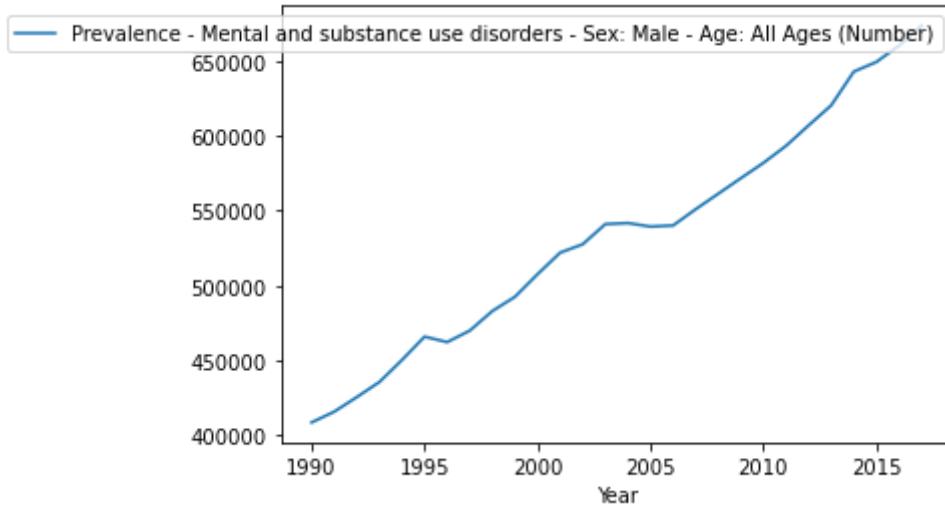
```
Out[1013]:
```

	Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)	Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)
--	---	---

Year		
1990	6.628189e+06	7.021926e+06
1991	6.737517e+06	7.132340e+06
1992	6.847914e+06	7.244494e+06
1993	6.956911e+06	7.355580e+06
1994	7.064276e+06	7.466933e+06

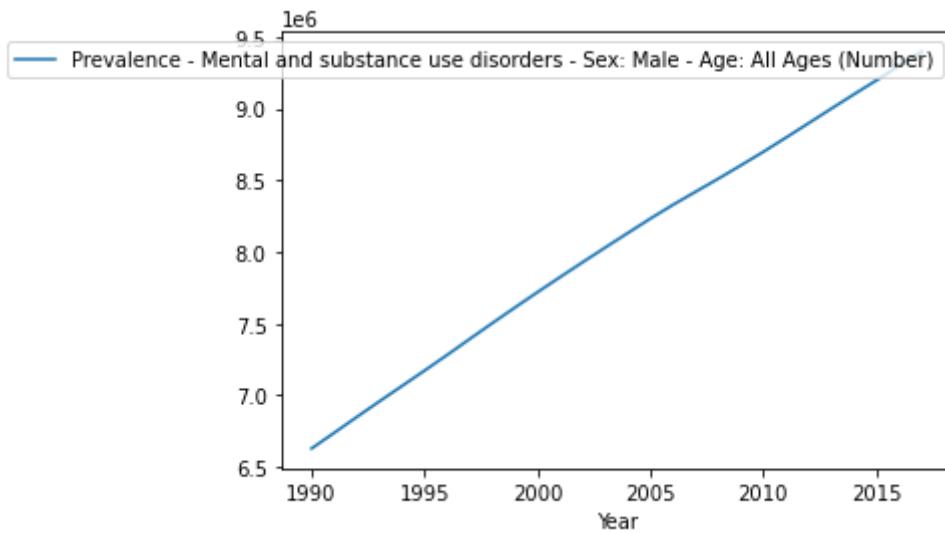
```
In [1014]: df64.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1014]: <AxesSubplot:xlabel='Year'>
```



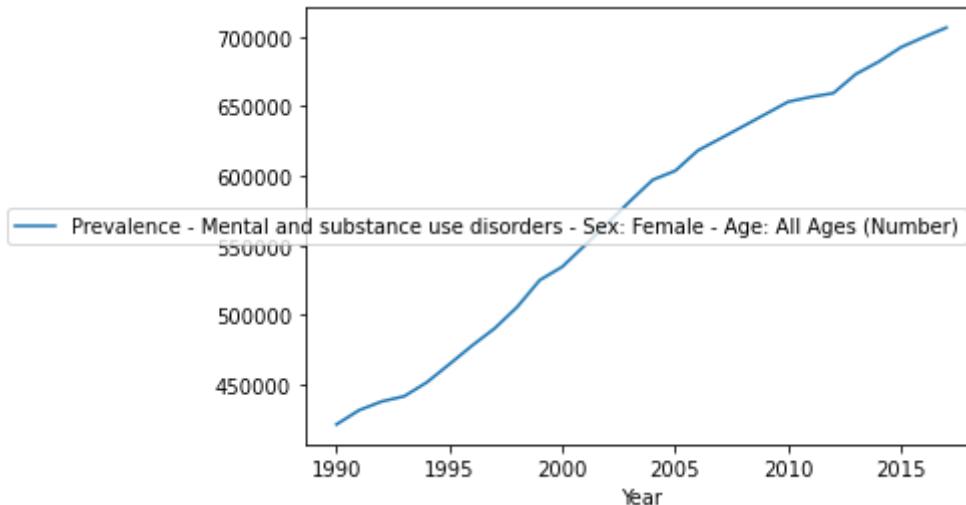
```
In [1015]: df64.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1015]: <AxesSubplot:xlabel='Year'>
```



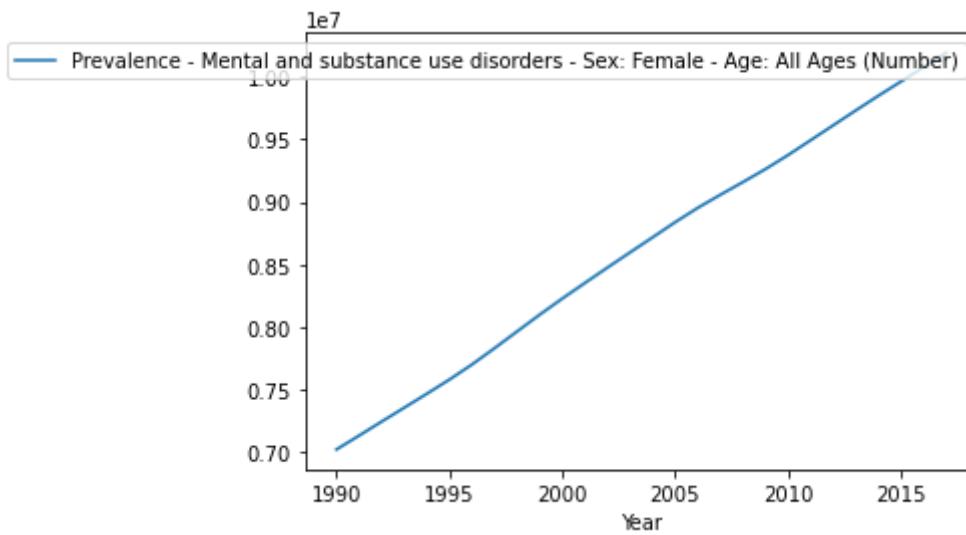
```
In [1016]: df64.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1016]: <AxesSubplot:xlabel='Year'>
```



```
In [1017]: df64.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1017]: <AxesSubplot:xlabel='Year'>
```



Annual Deaths by Cause

In [1018]:

```
df6.info()  
df6.head()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 6686 entries, 0 to 6685  
Data columns (total 37 columns):  
 #   Column  
Non-Null Count Dtype  
---  --  
-----  
 0   Entity      object  
6686 non-null   object  
 1   Code        object  
5545 non-null   object  
 2   Year        int64  
6686 non-null   int64  
 3   Number of executions (Amnesty International) object  
267 non-null   object  
 4   Deaths - Meningitis - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 5   Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 6   Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number) float64  
6156 non-null   float64  
 7   Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 8   Terrorism (deaths) float64  
2891 non-null   float64  
 9   Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 10  Deaths - Alzheimer disease and other dementias - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 11  Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 12  Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 13  Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 14  Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 15  Deaths - Hepatitis - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 16  Deaths - Neoplasms - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 17  Deaths - Parkinson disease - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 18  Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 19  Deaths - Malaria - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 20  Deaths - Drowning - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 21  Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 22  Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Number) float64  
6468 non-null   float64  
 23  Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number) float64
```

```

6468 non-null    float64
24 Deaths - Tuberculosis - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
25 Deaths - Road injuries - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
26 Deaths - Maternal disorders - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
27 Deaths - Neonatal disorders - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
28 Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
29 Deaths - Exposure to forces of nature - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
30 Deaths - Diarrheal diseases - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
31 Deaths - Environmental heat and cold exposure - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
32 Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
33 Deaths - Self-harm - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
34 Deaths - Conflict and terrorism - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
35 Deaths - Diabetes mellitus - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
36 Deaths - Poisonings - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
dtypes: float64(33), int64(1), object(3)
memory usage: 1.9+ MB

```

Out[1018]:

	Entity	Code	Year	Number of executions (Amnesty International)	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)	Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)	Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)	Terr (d)
0	Afghanistan	AFG	2007	15	9121.085992	29066.442137	461.195202	1846.996686	-
1	Afghanistan	AFG	2008	17	8387.057275	26623.480551	437.718960	1681.270324	-
2	Afghanistan	AFG	2009	0	7318.273004	24792.335792	415.776634	1568.095029	-
3	Afghanistan	AFG	2011	2	6919.757958	23115.144836	299.758258	1468.203744	-
4	Afghanistan	AFG	2012	14	6631.942601	22155.754481	302.255411	1406.214788	-

5 rows × 37 columns

Checking for missing values:

```
In [1019]: missing = pd.concat([df6.isnull().sum(), 100 * df6.isnull().mean()], axis=1)
missing.columns=[ 'count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1019]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Tuberculosis - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Road injuries - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Maternal disorders - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Neonatal disorders - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Exposure to forces of nature - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Diarrheal diseases - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Environmental heat and cold exposure - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Self-harm - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Conflict and terrorism - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Drowning - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Malaria - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Parkinson disease - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Meningitis - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Diabetes mellitus - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Alzheimer disease and other dementias - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Hepatitis - Sex: Both - Age: All Ages (Number)		218	3.260544
Deaths - Neoplasms - Sex: Both - Age: All Ages (Number)		218	3.260544

		count	%
Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number)	218	3.260544	
Deaths - Poisonings - Sex: Both - Age: All Ages (Number)	218	3.260544	
Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)	530	7.927012	
Code	1141	17.065510	
Terrorism (deaths)	3795	56.760395	
Number of executions (Amnesty International)	6419	96.006581	

In [1020]: `#dropping features with >50% data missing`
`df6.drop(labels=['Terrorism (deaths)', 'Number of executions (Amnesty Inter`

In [1021]: `df6.head(10)`

Out[1021]:

	Entity	Code	Year	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)	Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)	Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)	Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)	Deaths - d
0	Afghanistan	AFG	2007	9121.085992	29066.442137	461.195202	1846.996686	53532.680495	24
1	Afghanistan	AFG	2008	8387.057275	26623.480551	437.718960	1681.270324	53402.322328	24
2	Afghanistan	AFG	2009	7318.273004	24792.335792	415.776634	1568.095029	53024.450772	24
3	Afghanistan	AFG	2011	6919.757958	23115.144836	299.758258	1468.203744	52815.737495	26
4	Afghanistan	AFG	2012	6631.942601	22155.754481	302.255411	1406.214788	52961.704530	26
5	Afghanistan	AFG	2013	6774.892059	22417.229523	402.174478	1423.283631	53387.554213	27
6	Afghanistan	AFG	2014	6795.163086	22167.851944	427.419194	1421.239332	53858.559775	27
7	Afghanistan	AFG	2015	6667.310323	21627.195092	432.539812	1384.973801	54221.895746	27
8	Afghanistan	AFG	2016	6672.896174	21359.253797	435.834889	1363.976372	54963.454084	28
9	Afghanistan	AFG	1990	6469.977091	22836.912346	295.382207	1607.703797	46498.085024	19

10 rows × 35 columns

In [1022]: df6.describe()

Out[1022]:

	Year	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)	Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)	Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)	Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)	A
count	6686.000000	6468.000000	6.468000e+03	6156.000000	6468.000000	6.468000e+03	6.
mean	2003.469488	7780.768953	5.852767e+04	3430.235605	7968.171555	3.002555e+05	3.
std	8.076979	34193.574520	2.326967e+05	19322.536951	35566.546169	1.173251e+06	1.
min	1990.000000	0.214225	6.462032e+00	0.002202	0.035425	3.767702e+01	2.
25%	1996.000000	32.844678	6.106540e+02	0.476189	8.939738	4.225455e+03	3.
50%	2003.000000	263.326804	4.313073e+03	6.272934	139.132563	1.885626e+04	1.
75%	2010.000000	2093.966414	1.862988e+04	261.856157	2337.484551	8.015074e+04	9.
max	2017.000000	462419.725356	3.415941e+06	215726.866669	548711.952423	1.779095e+07	2.

8 rows × 33 columns

```
In [1023]: df49.info()  
df49.head()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 6468 entries, 0 to 6467  
Data columns (total 32 columns):  
 #   Column  
Non-Null Count Dtype  
---  --  
-----  
 0   Entity      object  
6468 non-null   object  
 1   Code        object  
5488 non-null   object  
 2   Year        int64  
6468 non-null   int64  
 3   Deaths - Unsafe water source - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 4   Deaths - Unsafe sanitation - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 5   Deaths - No access to handwashing facility - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 6   Deaths - Household air pollution from solid fuels - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 7   Deaths - Non-exclusive breastfeeding - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 8   Deaths - Discontinued breastfeeding - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 9   Deaths - Child wasting - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 10  Deaths - Child stunting - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 11  Deaths - Low birth weight for gestation - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 12  Deaths - Secondhand smoke - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 13  Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 14  Deaths - Drug use - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 15  Deaths - Diet low in fruits - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 16  Deaths - Diet low in vegetables - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 17  Deaths - Unsafe sex - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 18  Deaths - Low physical activity - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 19  Deaths - High fasting plasma glucose - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 20  Deaths - High total cholesterol - Sex: Both - Age: All Ages (Number)  
1561 non-null   float64  
 21  Deaths - High body-mass index - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 22  Deaths - High systolic blood pressure - Sex: Both - Age: All Ages (Number)  
6468 non-null   float64  
 23  Deaths - Smoking - Sex: Both - Age: All Ages (Number)
```

```

6468 non-null    float64
24 Deaths - Iron deficiency - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
25 Deaths - Vitamin A deficiency - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
26 Deaths - Low bone mineral density - Sex: Both - Age: All Ages (Number)
r)                               6468 non-null    float64
27 Deaths - Air pollution - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
28 Deaths - Outdoor air pollution (all ages) (IHME)
6467 non-null    float64
29 Deaths - Diet high in sodium - Sex: Both - Age: All Ages (Number)
6468 non-null    float64
30 Deaths - Diet low in whole grains - Sex: Both - Age: All Ages (Number)
r)                               6468 non-null    float64
31 Deaths - Diet low in nuts and seeds - Sex: Both - Age: All Ages (Number)
r)                               6468 non-null    float64
dtypes: float64(29), int64(1), object(2)
memory usage: 1.6+ MB

```

Out[1023]:

	Entity	Code	Year	Deaths - Unsafe water source - Sex: Both - Age: All Ages (Number)	Deaths - Unsafe sanitation - Sex: Both - Age: All Ages (Number)	Deaths - No access to handwashing facility - Sex: Both - Age: All Ages (Number)	Deaths - Household air pollution from solid fuels - Sex: Both - Age: All Ages (Number)	Deaths - Non-exclusive breastfeeding - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	1990	7554.049543	5887.747628	5412.314513	22388.497233	3221.138842
1	Afghanistan	AFG	1991	7359.676749	5732.770160	5287.891103	22128.758206	3150.559597
2	Afghanistan	AFG	1992	7650.437822	5954.804987	5506.657363	22873.768789	3331.349048
3	Afghanistan	AFG	1993	10270.731383	7986.736613	7104.620351	25599.756284	4477.006100
4	Afghanistan	AFG	1994	11409.177112	8863.010065	8051.515953	28013.167200	5102.622054

5 rows × 32 columns

Checking for missing values:

```
In [1024]: missing = pd.concat([df49.isnull().sum(), 100 * df49.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1024]:

		count	%
	Entity	0	0.000000
Deaths - Diet high in sodium - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Air pollution - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Low bone mineral density - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Vitamin A deficiency - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Iron deficiency - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Smoking - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - High systolic blood pressure - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - High body-mass index - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - High fasting plasma glucose - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Low physical activity - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Unsafe sex - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Diet low in vegetables - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Diet low in whole grains - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Diet low in fruits - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)		0	0.000000
	Year	0	0.000000
Deaths - Unsafe water source - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Unsafe sanitation - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - No access to handwashing facility - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Drug use - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Non-exclusive breastfeeding - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Household air pollution from solid fuels - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Child wasting - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Child stunting - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Low birth weight for gestation - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Secondhand smoke - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Discontinued breastfeeding - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Diet low in nuts and seeds - Sex: Both - Age: All Ages (Number)		0	0.000000
Deaths - Outdoor air pollution (all ages) (IHME)		1	0.015461
	Code	980	15.151515
Deaths - High total cholesterol - Sex: Both - Age: All Ages (Number)		4907	75.865801

```
In [1025]: #drop features with >50% data missing
df49.drop(labels=['Deaths - High total cholesterol - Sex: Both - Age: All A
                  #drop infant and child specific features
                  df49.drop(labels=['Deaths - Non-exclusive breastfeeding - Sex: Both - Age: All A

```

```
In [1026]: df49.describe()
```

Out[1026]:

Year	Deaths - Unsafe water source - Sex: Both - Age: All Ages (Number)	Deaths - Unsafe sanitation - Sex: Both - Age: All Ages (Number)	Deaths - No access to handwashing facility - Sex: Both - Age: All Ages (Number)	Deaths - Household air pollution from solid fuels - Sex: Both - Age: All Ages (Number)	Deaths - Secondhand smoke - Sex: Both - Age: All Ages (Number)	Alcoh Sex Age: / (N)
count	6468.000000	6.468000e+03	6.468000e+03	6.468000e+03	6.468000e+03	6.468000e+03
mean	2003.500000	3.156632e+04	2.337436e+04	1.893305e+04	4.308421e+04	2.428225e+04
std	8.078372	1.527731e+05	1.144930e+05	8.981037e+04	1.877345e+05	1.002562e+05
min	1990.000000	8.650193e-03	6.495981e-03	7.791357e-02	2.058533e-02	2.890665e+00
25%	1996.750000	1.019665e+01	4.603845e+00	1.688487e+01	8.759783e+01	2.780677e+02
50%	2003.500000	2.790317e+02	1.601965e+02	2.524991e+02	1.091671e+03	1.196228e+03
75%	2010.250000	5.301718e+03	3.832344e+03	3.811442e+03	9.161964e+03	5.963666e+03
max	2017.000000	2.111659e+06	1.638021e+06	1.239519e+06	2.708905e+06	1.260994e+06

8 rows × 24 columns

Anxiety Disorders

In [1027]: df12.info()
df12.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 10 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year        int64
6468 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate)    6468 non-null   float64
4   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Rate)     6468 non-null   float64
5   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate)    6468 non-null   float64
6   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate)  6468 non-null   float64
7   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate)   6468 non-null   float64
8   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate)  6468 non-null   float64
9   DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age-standardized (Rate) 6468 non-null   float64
dtypes: float64(7), int64(1), object(2)
memory usage: 505.4+ KB
```

Out[1027]:

	Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate)	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Rate)	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate)	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate)
0	Afghanistan	AFG	1990	10.851451	386.691683	336.764668	453.736877	290.799110	583.050750
1	Afghanistan	AFG	1991	10.758383	386.069958	336.346214	453.919631	287.732096	582.755568
2	Afghanistan	AFG	1992	10.387921	387.606438	336.558722	456.690054	284.531098	585.121129
3	Afghanistan	AFG	1993	10.289752	386.885092	336.870769	459.518593	281.185300	585.995450
4	Afghanistan	AFG	1994	10.464435	382.079120	337.476586	459.959994	278.113533	587.736708

Checking for missing values:

```
In [1028]: missing = pd.concat([df12.isnull().sum(), 100 * df12.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1028]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

```
In [1063]: v1='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both'
v2='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both'
v3='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both'
v4='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both'
v5='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both'
v6='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both'
v7='DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both'
```

In [1064]: df12.describe()

Out[1064]:

Year	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate)	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14	DALYs (Disability- Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	10.667841	360.300699	335.051311	420.883333	265.224424
std	8.078372	3.573766	110.775677	72.916462	105.986344	82.099643
min	1990.000000	5.421487	174.599768	197.059073	257.617569	129.669977
25%	1996.750000	8.646651	276.916613	277.439992	336.264466	218.039847
50%	2003.500000	9.276162	322.569279	344.333795	400.139300	234.733150
75%	2010.250000	11.271450	435.065320	371.161221	489.641882	284.897799
max	2017.000000	24.913957	786.105689	637.587946	827.969223	602.488920

In [1065]: df12.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1065]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate)
4240	Northern Ireland	2002	24.913957
4241	Northern Ireland	2003	24.870098
4239	Northern Ireland	2001	24.827560
4233	Northern Ireland	1995	24.716760
4234	Northern Ireland	1996	24.682782

In [1066]: df12.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1066]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Rate)
4016	New Zealand	2002	786.105689
4015	New Zealand	2001	785.954791
4017	New Zealand	2003	785.451777
4014	New Zealand	2000	785.169351
4018	New Zealand	2004	784.994491

In [1069]: df12.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1069]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate)
4238	Northern Ireland	2000	637.587946
4239	Northern Ireland	2001	636.579968
4240	Northern Ireland	2002	634.525120
4237	Northern Ireland	1999	633.775392
4241	Northern Ireland	2003	633.702056

In [1070]: df12.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1070]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate)
4238	Northern Ireland	2000	827.969223
4239	Northern Ireland	2001	827.356239
4240	Northern Ireland	2002	827.186923
4241	Northern Ireland	2003	826.397831
4242	Northern Ireland	2004	825.224446

In [1071]: df12.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1071]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate)
4247	Northern Ireland	2009	602.488920
4248	Northern Ireland	2010	601.518624
4246	Northern Ireland	2008	601.044825
4245	Northern Ireland	2007	600.065302
4239	Northern Ireland	2001	599.254152

In [1072]: df12.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1072]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate)
4015	New Zealand	2001	1062.642792
4016	New Zealand	2002	1061.863757
4014	New Zealand	2000	1061.604931
4017	New Zealand	2003	1059.848242
4018	New Zealand	2004	1058.001762

In [1073]: df12.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1073]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Age-standardized (Rate)
4015	New Zealand	2001	765.267392
4016	New Zealand	2002	765.148801
4014	New Zealand	2000	764.405461
4017	New Zealand	2003	764.165641
4018	New Zealand	2004	763.331296

```
In [1074]: df12.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[1074]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate)
6260	Vietnam	2006	5.421487
6259	Vietnam	2005	5.440761
6262	Vietnam	2008	5.474217
6261	Vietnam	2007	5.478233
6263	Vietnam	2009	5.515967

```
In [1075]: df12.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()
```

Out[1075]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Rate)
6244	Vietnam	1990	174.599768
6245	Vietnam	1991	175.333211
6246	Vietnam	1992	176.204569
6247	Vietnam	1993	177.078827
6248	Vietnam	1994	178.072080

```
In [1076]: df12.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()
```

Out[1076]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate)
1314	Colombia	2016	197.059073
1313	Colombia	2015	197.120006
1315	Colombia	2017	197.298866
1311	Colombia	2013	197.659255
1312	Colombia	2014	197.797046

In [1077]: df12.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1077]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate)
6244	Vietnam	1990	257.617569
6245	Vietnam	1991	258.000626
6248	Vietnam	1994	258.027297
6259	Vietnam	2005	258.029127
6258	Vietnam	2004	258.083478

In [1078]: df12.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1078]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate)
6267	Vietnam	2013	129.669977
6268	Vietnam	2014	129.711356
6266	Vietnam	2012	129.802780
6269	Vietnam	2015	129.883591
6244	Vietnam	1990	129.962082

In [1079]: df12.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[1079]:

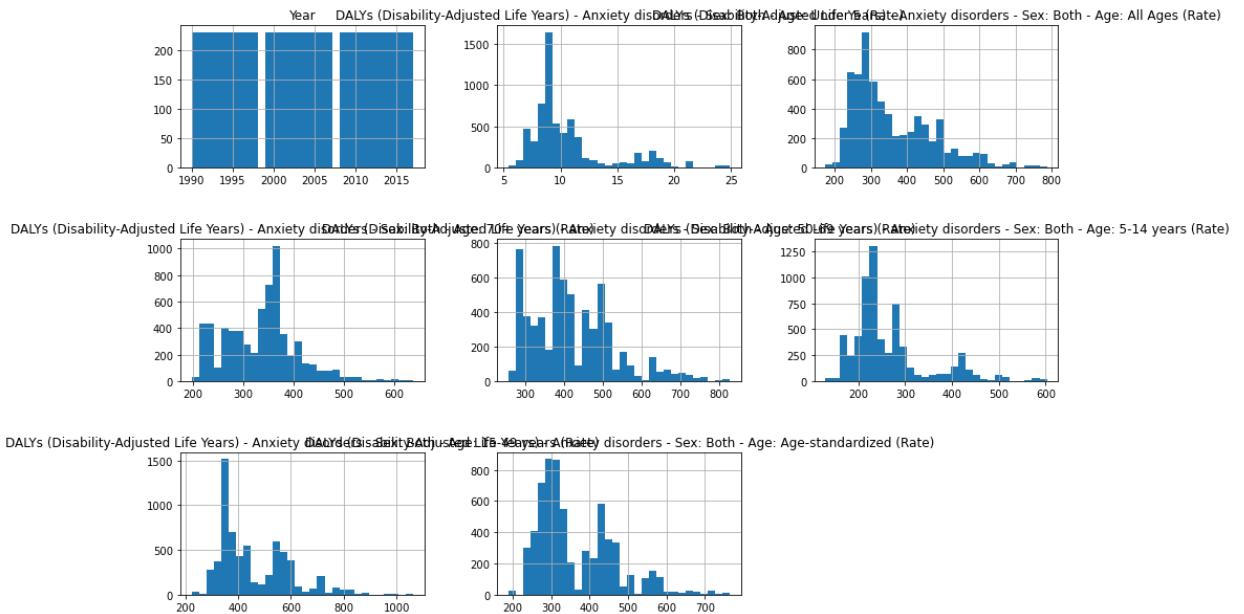
	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate)
6261	Vietnam	2007	223.959514
6244	Vietnam	1990	223.991333
6260	Vietnam	2006	223.999505
6264	Vietnam	2010	224.022350
6259	Vietnam	2005	224.059129

In [1080]: df12.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[1080]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age-standardized (Rate)
6261	Vietnam	2007	189.545964
6262	Vietnam	2008	189.625634
6264	Vietnam	2010	189.709623
6260	Vietnam	2006	189.730831
6263	Vietnam	2009	189.753134

```
In [1081]: df12.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1082]: df12.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1082]: Entity
Vietnam      5.628131
Colombia     6.551298
Tajikistan    6.740182
Mongolia      6.742935
Uzbekistan    6.813208
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate), dtype: float64
```

```
In [1083]: df12.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1083]: Entity
Vietnam      188.574428
Tajikistan    214.257177
Uzbekistan    224.372684
Mongolia      225.547126
Kyrgyzstan    226.652374
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1084]: df12.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1084]: Entity
Colombia     200.899583
Honduras      216.095048
Panama        217.198181
Nicaragua     217.335244
Liberia        218.556124
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [1085]: df12.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1085]: Entity
Vietnam      258.477351
Colombia    273.894706
Liberia     277.282593
Nigeria    281.572337
Gambia      282.415678
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [1086]: df12.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1086]: Entity
Vietnam      132.382698
Colombia    162.461393
Tajikistan   165.027791
Uzbekistan  166.220429
Kyrgyzstan  166.524899
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [1087]: df12.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1087]: Entity
Vietnam      225.144026
Mongolia    281.818894
Turkmenistan 284.353240
Tajikistan   284.554034
Uzbekistan  284.697411
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1088]: df12.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1088]: Entity
Vietnam      191.253854
Mongolia    232.397408
Tajikistan   233.292795
Uzbekistan  234.212024
Turkmenistan 234.761525
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [1089]: df12.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1089]: Entity
Netherlands  21.276881
Germany     21.306577
France       21.325949
Norway       24.077579
Northern Ireland 24.456373
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate), dtype: float64
```

```
In [1090]: df12.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1090]: Entity
Netherlands      627.687265
Iran             637.608821
Norway           696.282642
Northern Ireland 721.037315
New Zealand      756.392096
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1091]: df12.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1091]: Entity
Germany          518.322281
France            518.833152
Netherlands       522.080374
Norway            582.692673
Northern Ireland  612.979070
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [1092]: df12.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1092]: Entity
United States     681.571687
Iran              698.152459
New Zealand       719.245748
Norway            755.160813
Northern Ireland  791.133481
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [1093]: df12.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1093]: Entity
France            506.297895
Germany           506.380769
Netherlands        509.607854
Norway             579.546630
Northern Ireland   586.966931
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [1094]: df12.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1094]: Entity
Australasia        801.134163
Norway              816.550626
Iran                839.393222
Northern Ireland    851.909571
New Zealand         1008.299026
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1095]: df12.groupby('Entity')[v7].mean().sort_values().tail()
```

```
/opt/anaconda3/lib/python3.8/site-packages/IPython/core/displayhook.py:27
5: UserWarning: Output cache limit (currently 1000 entries) hit.
  Flushing oldest 200 entries.
  warn('Output cache limit (currently {sz} entries) hit.\n'
```

```
Out[1095]: Entity
Netherlands      596.332915
Iran             630.064996
Norway           669.859414
Northern Ireland 693.375842
New Zealand      736.019172
Name: DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [1096]: df21_mean = df12.groupby('Year').mean()
df12_mean.head()
```

```
-----
-->
NameError                                 Traceback (most recent call last)
t)
<ipython-input-1096-b6a36ba7995f> in <module>
      1 df21_mean = df12.groupby('Year').mean()
----> 2 df12_mean.head()

NameError: name 'df12_mean' is not defined
```

```
In [ ]: df12.groupby('Year')[v1].median().plot(legend=True)
```

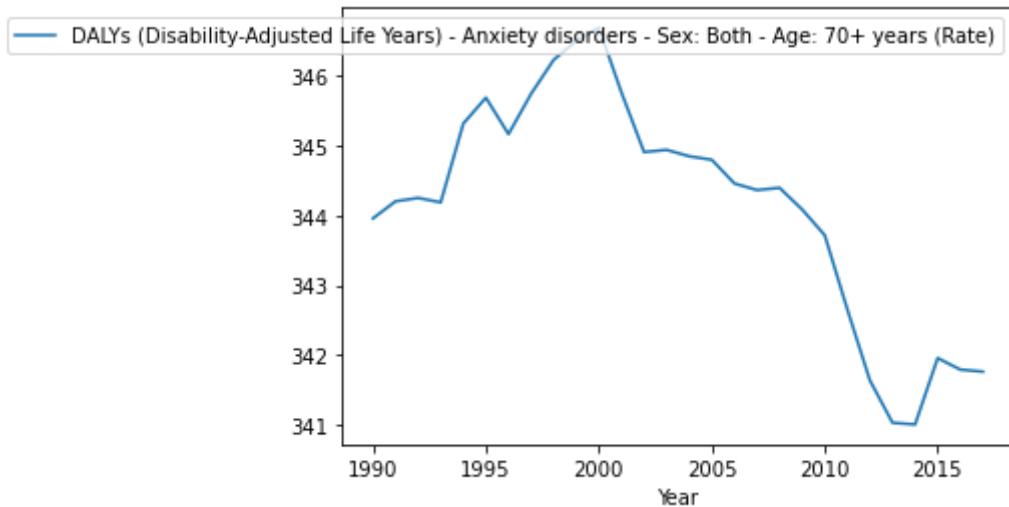
```
In [ ]: df12.groupby('Year')[v1].mean().plot(legend=True)
```

```
In [ ]: df12.groupby('Year')[v2].median().plot(legend=True)
```

```
In [ ]: df12.groupby('Year')[v2].mean().plot(legend=True)
```

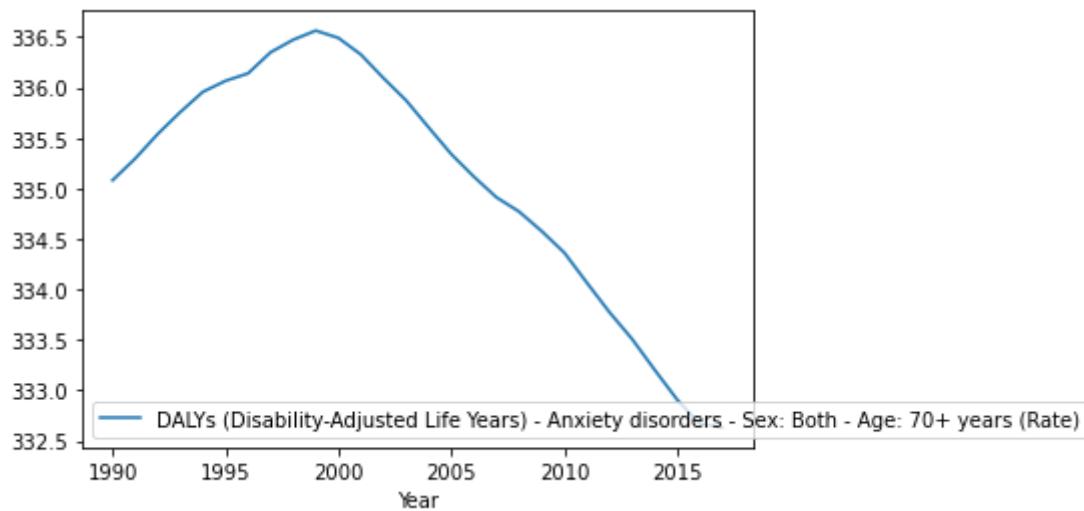
```
In [1097]: df12.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1097]: <AxesSubplot:xlabel='Year'>
```



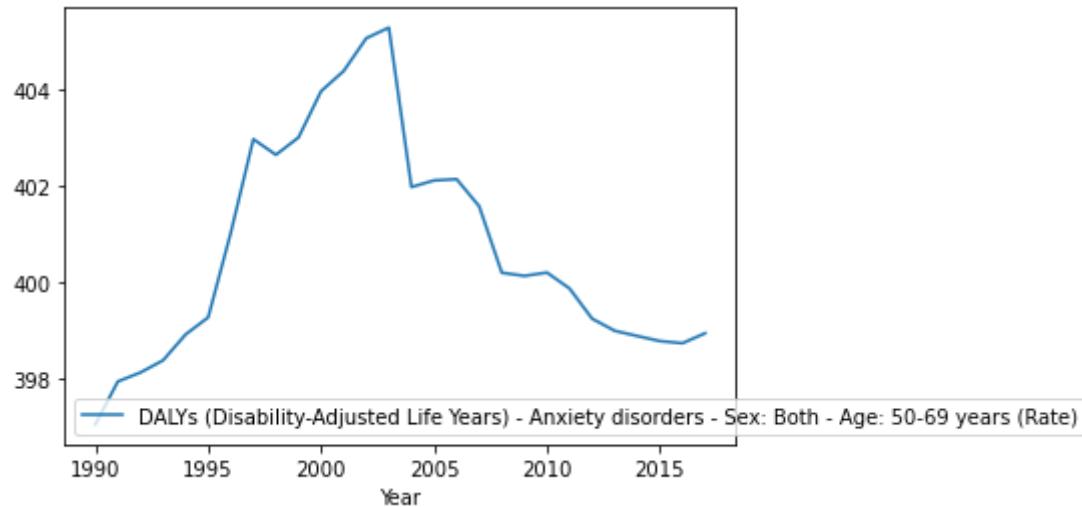
```
In [1098]: df12.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1098]: <AxesSubplot:xlabel='Year'>
```



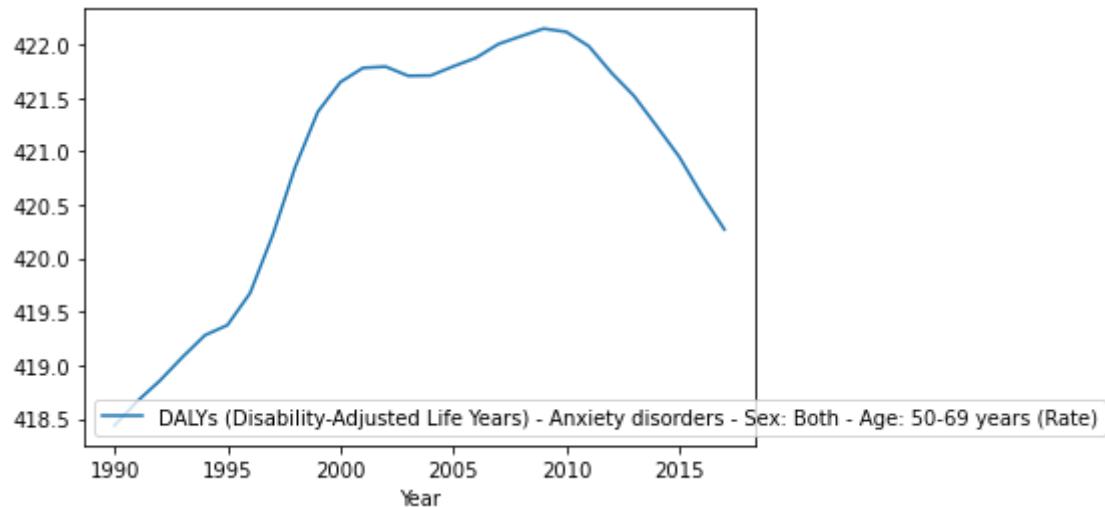
```
In [1099]: df12.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1099]: <AxesSubplot:xlabel='Year'>
```



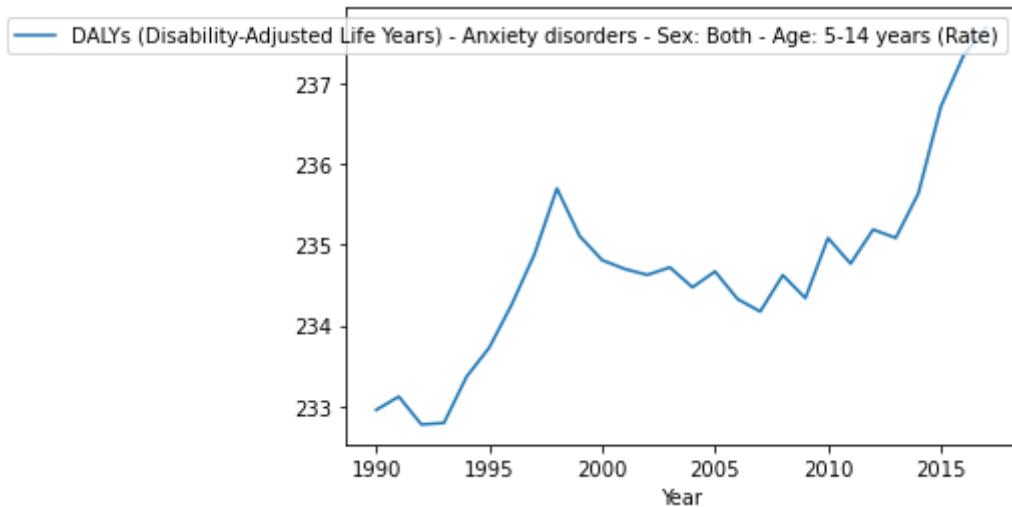
```
In [1100]: df12.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1100]: <AxesSubplot:xlabel='Year'>
```



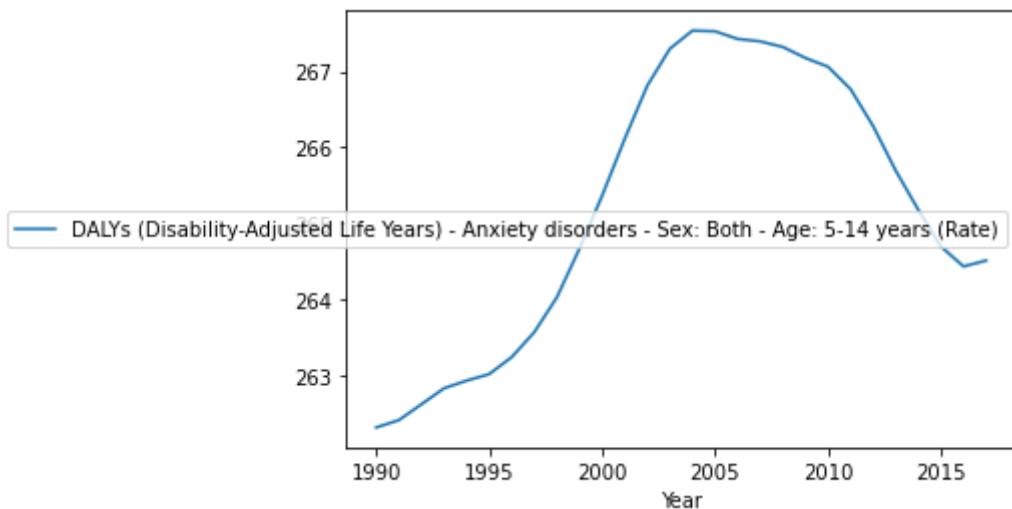
```
In [1101]: df12.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1101]: <AxesSubplot:xlabel='Year'>
```



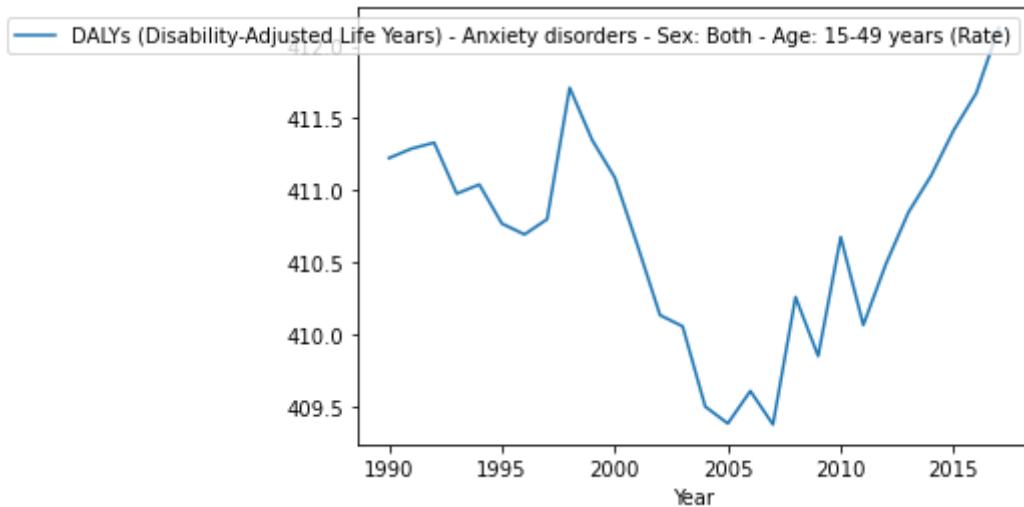
```
In [1102]: df12.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1102]: <AxesSubplot:xlabel='Year'>
```



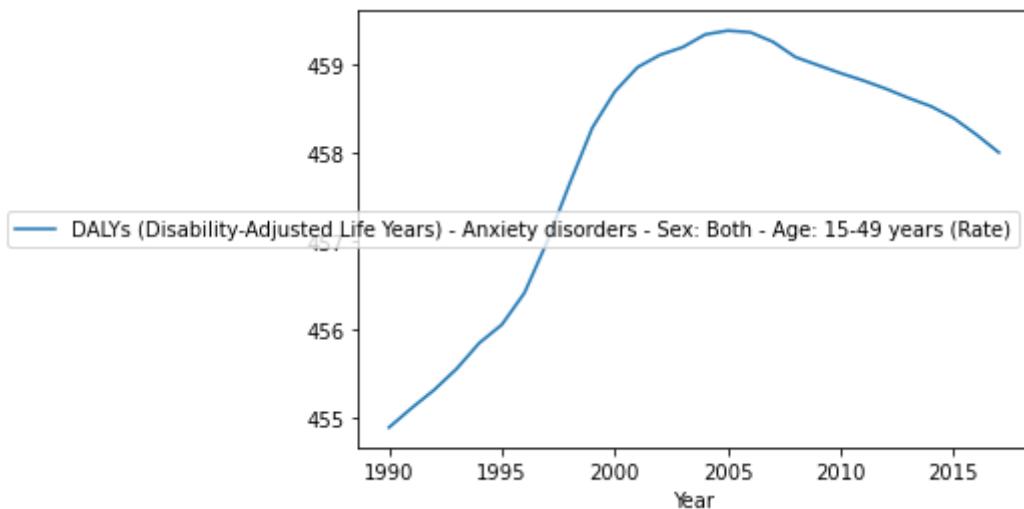
```
In [1103]: df12.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1103]: <AxesSubplot:xlabel='Year'>
```



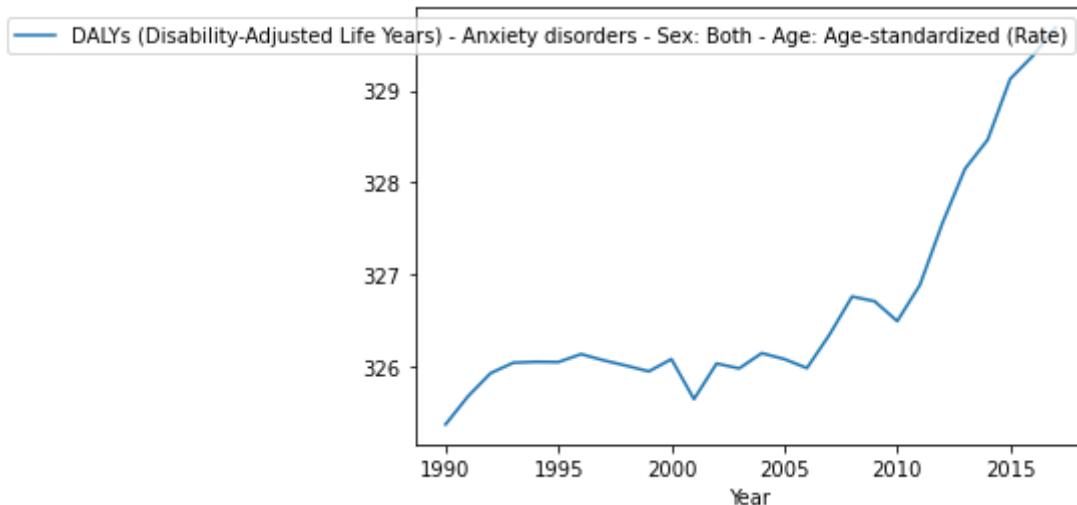
```
In [1104]: df12.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1104]: <AxesSubplot:xlabel='Year'>
```



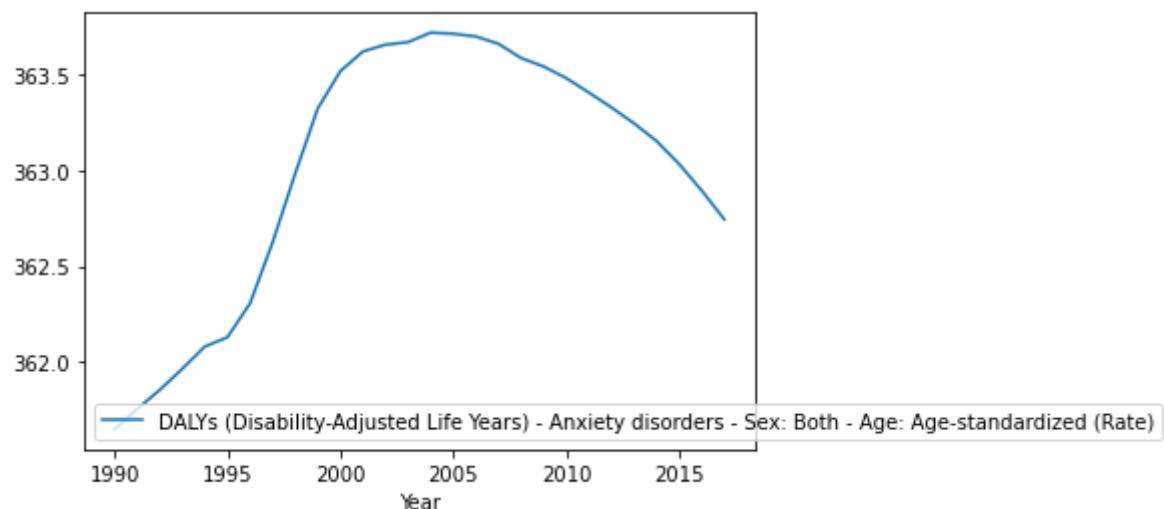
```
In [1105]: df12.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1105]: <AxesSubplot:xlabel='Year'>
```



```
In [1106]: df12.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1106]: <AxesSubplot:xlabel='Year'>
```



In [1107]: df58.info()
df58.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)
r) 6468 non-null   float64
4   Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)
6468 non-null   float64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1107]:

	Entity	Code	Year	Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)
0	Afghanistan	AFG	1990	252379.267264	164475.899445
1	Afghanistan	AFG	1991	262036.445129	170087.377635
2	Afghanistan	AFG	1992	310572.487808	199894.632147
3	Afghanistan	AFG	1993	361936.608315	231159.684414
4	Afghanistan	AFG	1994	376216.938911	238886.225430

Checking for missing values:

In [1108]: missing = pd.concat([df58.isnull().sum(), 100 * df58.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1108]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)		0	0.000000
Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)		0	0.000000
Code	980	15.151515	

```
In [1109]: v1='Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)'
v2='Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)'
```

```
In [1110]: df58.describe()
```

Out[1110]:

	Year	Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)
count	6468.000000	6.468000e+03	6.468000e+03
mean	2003.500000	3.077669e+06	1.790311e+06
std	8.078372	1.188462e+07	6.984347e+06
min	1990.000000	7.345057e+02	4.673009e+02
25%	1996.750000	4.807765e+04	3.047222e+04
50%	2003.500000	2.060209e+05	1.198175e+05
75%	2010.250000	1.055032e+06	5.703277e+05
max	2017.000000	1.787190e+08	1.056411e+08

```
In [1111]: df58.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[1111]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)
6383	World	2017	1.787190e+08
6382	World	2016	1.768644e+08
6381	World	2015	1.749929e+08
6380	World	2014	1.731254e+08
6379	World	2013	1.711806e+08

```
In [1112]: df58.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()
```

Out[1112]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)
6383	World	2017	1.056411e+08
6382	World	2016	1.043643e+08
6381	World	2015	1.030714e+08
6380	World	2014	1.018099e+08
6379	World	2013	1.005302e+08

```
In [1113]: df58.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[1113]:

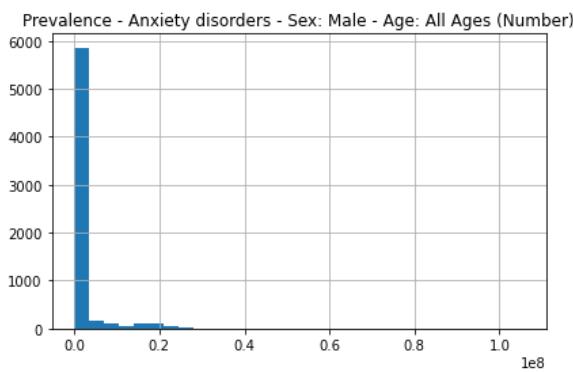
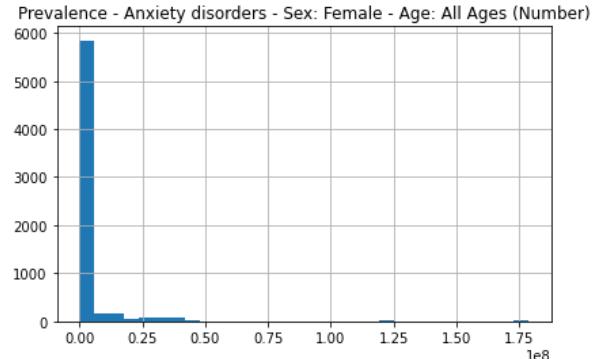
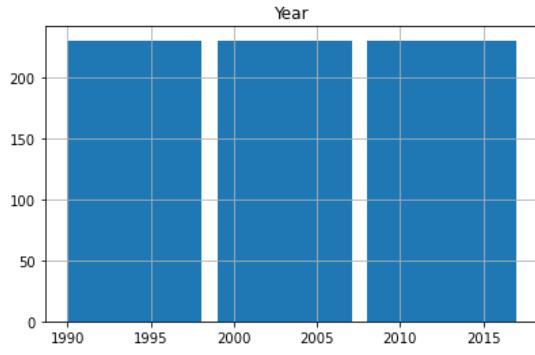
	Entity	Year	Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)
3584	Marshall Islands	1990	734.505655
3585	Marshall Islands	1991	756.288952
3586	Marshall Islands	1992	777.516125
3587	Marshall Islands	1993	798.027482
4256	Northern Mariana Islands	1990	814.426290

```
In [1114]: df58.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()
```

Out[1114]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)
3584	Marshall Islands	1990	467.300861
3585	Marshall Islands	1991	480.149371
3586	Marshall Islands	1992	492.563363
3587	Marshall Islands	1993	504.542618
3588	Marshall Islands	1994	516.089794

```
In [1115]: df58.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1116]: df58.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1116]: Entity
Marshall Islands      915.888072
American Samoa        1024.470180
Northern Mariana Islands 1101.461351
Seychelles             1743.804826
Kiribati               1765.397340
Name: Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number), dtype: float64
```

```
In [1117]: df58.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1117]: Entity
Marshall Islands      578.348053
American Samoa         628.844474
Northern Mariana Islands 659.478276
Kiribati                1000.929032
Seychelles              1038.037143
Name: Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1118]: df58.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1118]: Entity
High SDI           3.633891e+07
High-income        3.646265e+07
Middle SDI         3.737777e+07
Southeast Asia, East Asia, and Oceania 3.885392e+07
World              1.499981e+08
Name: Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number), dtype: float64
```

```
In [1119]: df58.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1119]: Entity
High SDI           1.865648e+07
South Asia          1.927474e+07
Middle SDI          2.261340e+07
Southeast Asia, East Asia, and Oceania 2.272819e+07
World              8.819313e+07
Name: Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1120]: df58_mean = df58.groupby('Year').mean()
df58_mean.head()
```

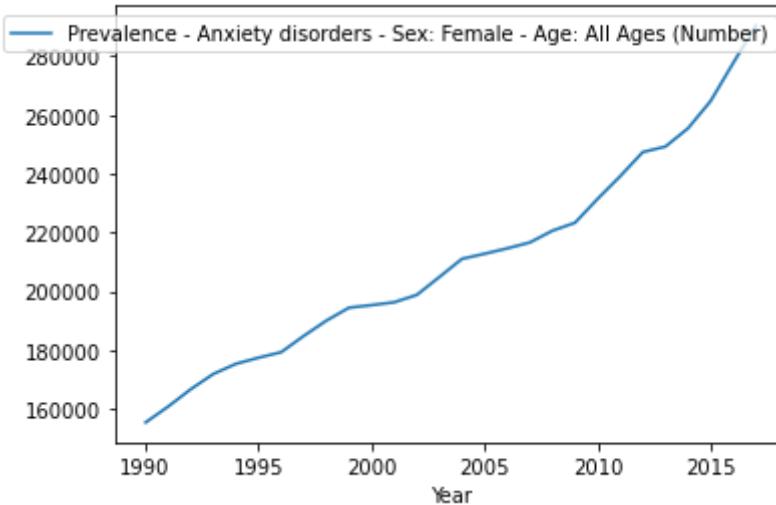
```
Out[1120]:
```

	Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)
--	---	---

Year	Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Anxiety disorders - Sex: Male - Age: All Ages (Number)
1990	2.460723e+06	1.442139e+06
1991	2.499291e+06	1.466819e+06
1992	2.538984e+06	1.491782e+06
1993	2.579116e+06	1.516749e+06
1994	2.620277e+06	1.541843e+06

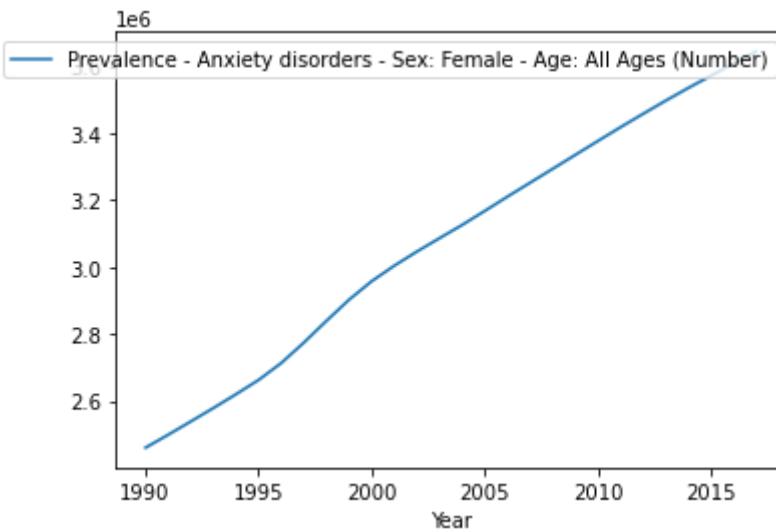
```
In [1121]: df58.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1121]: <AxesSubplot:xlabel='Year'>
```



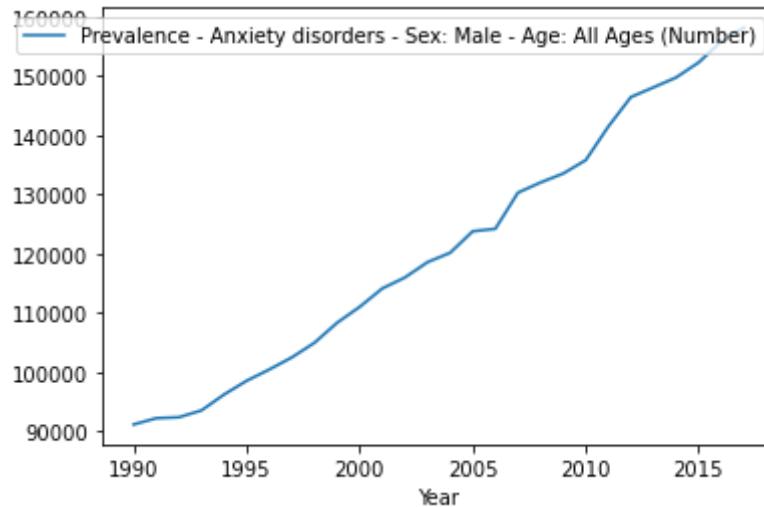
```
In [1122]: df58.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1122]: <AxesSubplot:xlabel='Year'>
```



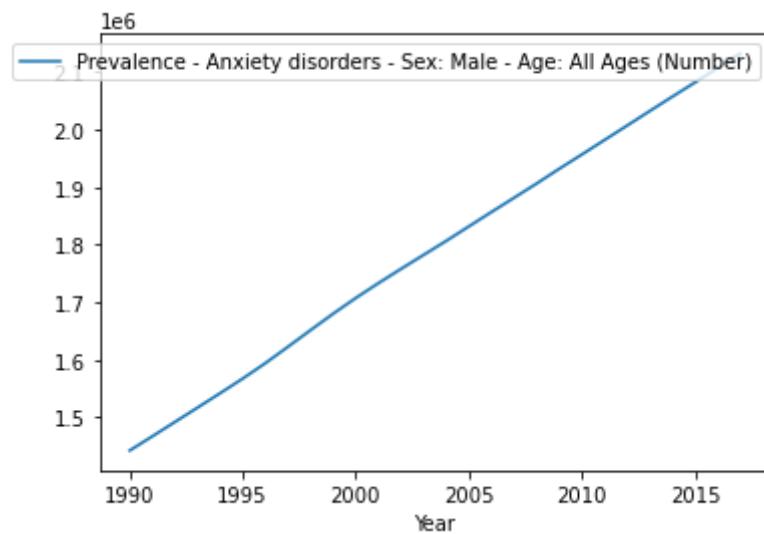
```
In [1123]: df58.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1123]: <AxesSubplot:xlabel='Year'>
```



```
In [1124]: df58.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1124]: <AxesSubplot:xlabel='Year'>
```



In [1125]: df73.info()
df73.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 15 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   float64
3   Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)
6468 non-null   float64
4   Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)
6468 non-null   float64
5   Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)
6468 non-null   float64
6   Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)
6468 non-null   float64
7   Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)
6468 non-null   float64
8   Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)
6468 non-null   float64
9   Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
6468 non-null   float64
10  Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)
6468 non-null   float64
11  Prevalence - Anxiety disorders - Sex: Both - Age: 15-49 years (Percent)
6468 non-null   float64
12  Prevalence - Anxiety disorders - Sex: Both - Age: 70+ years (Percent)
6468 non-null   float64
13  Prevalence - Anxiety disorders - Sex: Both - Age: 20 to 24 (Percent)
6468 non-null   float64
14  Prevalence - Anxiety disorders - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
dtypes: float64(12), int64(1), object(2)
memory usage: 758.1+ KB
```

Out[1125]:

	Entity	Code	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)
0	Afghanistan	AFG	1990	6.802932	0.113877	5.794202	4.354406	6.826513	3.035495
1	Afghanistan	AFG	1991	6.808945	0.112789	5.788069	4.354839	6.769529	2.993475
2	Afghanistan	AFG	1992	6.823713	0.109435	5.764448	4.355385	6.664152	2.958275

Entity	Code	Year	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence
			- Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)	- Anxiety disorders - Sex: Both - Age: Under 5 (Percent)	- Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)	- Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)	- Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)	- Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)
3	Afghanistan	AFG	1993	6.828456	0.108736	5.747657	4.356357	6.591816
4	Afghanistan	AFG	1994	6.814015	0.110961	5.751007	4.357629	6.586557

Checking for missing values:

```
In [1126]: missing = pd.concat([df73.isnull().sum(), 100 * df73.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1126]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 15-49 years (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
Code	980	15.151515	

```
In [1127]: v1='Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)'
v2='Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)'
v3='Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)'
v4='Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)'
v5='Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)'
v6='Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)'
v7='Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)'
v8='Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Per'
v9='Prevalence - Anxiety disorders - Sex: Both - Age: 15-49 years (Percent)'
v10='Prevalence - Anxiety disorders - Sex: Both - Age: 70+ years (Percent)'
v11='Prevalence - Anxiety disorders - Sex: Both - Age: 20 to 24 (Percent)'
v12='Prevalence - Anxiety disorders - Sex: Both - Age: All Ages (Percent)'
```

```
In [1128]: df73.describe()
```

Out[1128]:

	Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	4.780737	0.133574	4.852750	4.176475	4.689114
std	8.078372	1.528279	0.064727	1.475895	1.410520	1.505836
min	1990.000000	2.259521	0.059571	2.176986	1.904346	2.115601
25%	1996.750000	3.499611	0.094541	3.872644	3.336487	3.465122
50%	2003.500000	4.239070	0.107777	4.386963	3.627355	4.161835
75%	2010.250000	5.778770	0.135586	5.776919	4.613487	5.650790
max	2017.000000	12.325191	0.386139	10.829729	9.595047	12.465135
						7.085504

```
In [1129]: df73.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[1129]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)
4016	New Zealand	2002	12.325191
4017	New Zealand	2003	12.323527
4018	New Zealand	2004	12.319391
4015	New Zealand	2001	12.315547
4019	New Zealand	2005	12.311847

In [1130]: df73.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1130]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)
2087	France	2005	0.386139
2086	France	2004	0.385927
2088	France	2006	0.385415
2085	France	2003	0.384910
2089	France	2007	0.384717

In [1131]: df73.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1131]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)
4015	New Zealand	2001	10.829729
4016	New Zealand	2002	10.826727
4014	New Zealand	2000	10.826551
4017	New Zealand	2003	10.816047
4018	New Zealand	2004	10.801801

In [1132]: df73.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1132]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)
4299	Norway	2005	9.595047
4298	Norway	2004	9.590943
4300	Norway	2006	9.576803
4245	Northern Ireland	2007	9.575263
4242	Northern Ireland	2004	9.573940

In [1133]: df73.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1133]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)
4015	New Zealand	2001	12.465135
4016	New Zealand	2002	12.459262
4014	New Zealand	2000	12.459133
4017	New Zealand	2003	12.449190
4018	New Zealand	2004	12.433978

In [1134]: df73.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1134]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)
4301	Norway	2007	7.085504
4302	Norway	2008	7.084494
4299	Norway	2005	7.084474
4300	Norway	2006	7.082112
4298	Norway	2004	7.072774

In [1135]: df73.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1135]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
4238	Northern Ireland	2000	8.974926
4239	Northern Ireland	2001	8.974867
4240	Northern Ireland	2002	8.970964
4241	Northern Ireland	2003	8.963484
4242	Northern Ireland	2004	8.954033

In [1136]: df73.sort_values(by=v8, ascending=False)[['Entity', 'Year', v7]].head()

Out[1136]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
4016	New Zealand	2002	7.928377
4017	New Zealand	2003	7.923068
4015	New Zealand	2001	7.935640
4018	New Zealand	2004	7.916047
4014	New Zealand	2000	7.925490

In [1137]: df73.sort_values(by=v9, ascending=False)[['Entity', 'Year', v7]].head()

Out[1137]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
4014	New Zealand	2000	7.925490
4015	New Zealand	2001	7.935640
4016	New Zealand	2002	7.928377
4017	New Zealand	2003	7.923068
4018	New Zealand	2004	7.916047

```
In [1138]: df73.sort_values(by=v10, ascending=False)[['Entity', 'Year', v7]].head()
```

Out[1138]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
4238	Northern Ireland	2000	8.974926
4239	Northern Ireland	2001	8.974867
4240	Northern Ireland	2002	8.970964
4237	Northern Ireland	1999	8.872940
4241	Northern Ireland	2003	8.963484

```
In [1139]: df73.sort_values(by=v11, ascending=False)[['Entity', 'Year', v7]].head()
```

Out[1139]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
4014	New Zealand	2000	7.925490
4015	New Zealand	2001	7.935640
4016	New Zealand	2002	7.928377
4017	New Zealand	2003	7.923068
4018	New Zealand	2004	7.916047

```
In [1140]: df73.sort_values(by=v12, ascending=False)[['Entity', 'Year', v7]].head()
```

Out[1140]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
4014	New Zealand	2000	7.925490
4015	New Zealand	2001	7.935640
4016	New Zealand	2002	7.928377
4017	New Zealand	2003	7.923068
4018	New Zealand	2004	7.916047

```
In [1141]: df73.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[1141]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)
6259	Vietnam	2005	2.259521
6260	Vietnam	2006	2.259598
6261	Vietnam	2007	2.260905
6258	Vietnam	2004	2.262545
6262	Vietnam	2008	2.263134

In [1142]: df73.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1142]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)
6244	Vietnam	1990	0.059571
6245	Vietnam	1991	0.059825
6246	Vietnam	1992	0.060100
6247	Vietnam	1993	0.060357
6248	Vietnam	1994	0.060576

In [1143]: df73.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1143]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)
6259	Vietnam	2005	2.176986
6260	Vietnam	2006	2.177064
6261	Vietnam	2007	2.177348
6262	Vietnam	2008	2.177726
6263	Vietnam	2009	2.178651

In [1144]: df73.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1144]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)
6244	Vietnam	1990	1.904346
6245	Vietnam	1991	1.906147
6246	Vietnam	1992	1.907942
6247	Vietnam	1993	1.909950
6259	Vietnam	2005	1.911107

In [1145]: df73.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1145]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)
6259	Vietnam	2005	2.115601
6260	Vietnam	2006	2.116801
6258	Vietnam	2004	2.118164
6261	Vietnam	2007	2.118688
6262	Vietnam	2008	2.121011

In [1146]: df73.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[1146]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)
6244	Vietnam	1990	1.323106
6245	Vietnam	1991	1.325896
6246	Vietnam	1992	1.329689
6247	Vietnam	1993	1.334280
6248	Vietnam	1994	1.339204

In [1147]: df73.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[1147]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
6261	Vietnam	2007	2.773031
6262	Vietnam	2008	2.773125
6263	Vietnam	2009	2.773348
6260	Vietnam	2006	2.773381
6265	Vietnam	2011	2.773412

In [1148]: df73.sort_values(by=v8, ascending=True)[['Entity', 'Year', v7]].head()

Out[1148]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
6259	Vietnam	2005	2.774619
6260	Vietnam	2006	2.773381
6258	Vietnam	2004	2.777370
6261	Vietnam	2007	2.773031
6262	Vietnam	2008	2.773125

In [1149]: df73.sort_values(by=v9, ascending=True)[['Entity', 'Year', v7]].head()

Out[1149]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)
6244	Vietnam	1990	2.787883
6245	Vietnam	1991	2.787575
6246	Vietnam	1992	2.787480
6247	Vietnam	1993	2.787571
6248	Vietnam	1994	2.787838

```
In [1150]: df73.sort_values(by=v10, ascending=True)[['Entity', 'Year', v7]].head()
```

Out[1150]:

Entity Year Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)			
1314	Colombia	2016	2.788966
1315	Colombia	2017	2.785428
1313	Colombia	2015	2.793697
1312	Colombia	2014	2.800010
1311	Colombia	2013	2.806848

```
In [1151]: df73.sort_values(by=v11, ascending=True)[['Entity', 'Year', v7]].head()
```

Out[1151]:

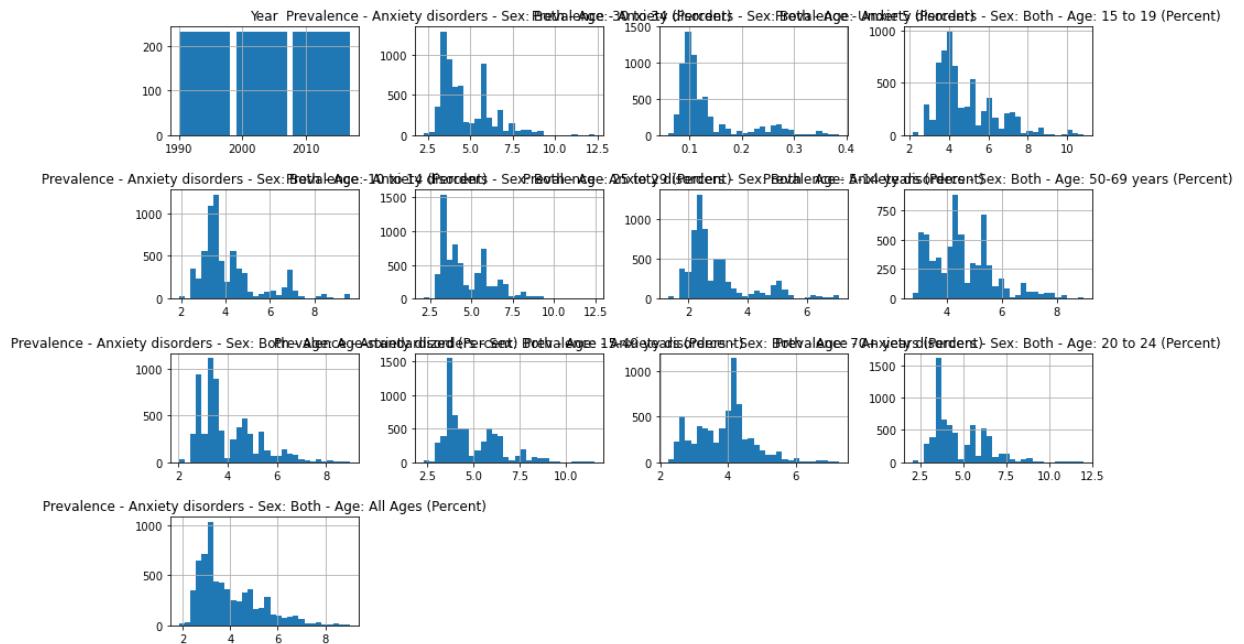
Entity Year Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)			
6259	Vietnam	2005	2.774619
6260	Vietnam	2006	2.773381
6261	Vietnam	2007	2.773031
6262	Vietnam	2008	2.773125
6263	Vietnam	2009	2.773348

```
In [1152]: df73.sort_values(by=v12, ascending=True)[['Entity', 'Year', v7]].head()
```

Out[1152]:

Entity Year Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)			
6244	Vietnam	1990	2.787883
6245	Vietnam	1991	2.787575
6246	Vietnam	1992	2.787480
6247	Vietnam	1993	2.787571
6248	Vietnam	1994	2.787838

```
In [1153]: df73.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1154]: df73.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1154]: Entity
Vietnam      2.284951
Mongolia    2.908695
Turkmenistan 2.928727
Tajikistan   2.933818
Kyrgyzstan   2.935627
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [1155]: df73.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1155]: Entity
Vietnam      0.061608
Tajikistan   0.074833
Colombia    0.077201
Uzbekistan   0.077564
Honduras     0.077985
Name: Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent), dtype: float64
```

```
In [1156]: df73.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1156]: Entity
Vietnam      2.202259
Mongolia     2.954250
Tajikistan   2.966982
Turkmenistan 2.969097
Uzbekistan   2.976720
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [1157]: df73.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1157]: Entity
Vietnam      1.927015
Colombia    2.506093
Tajikistan   2.520336
Mongolia     2.537035
Uzbekistan   2.541247
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1158]: df73.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1158]: Entity
Vietnam      2.141780
Mongolia     2.842670
Turkmenistan 2.863184
Tajikistan   2.875251
Kyrgyzstan   2.875912
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1159]: df73.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1159]: Entity
Vietnam      1.374382
Tajikistan   1.747413
Colombia    1.759027
Uzbekistan   1.782667
Mongolia     1.785691
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent), dtype: float64
```

```
In [1160]: df73.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1160]: Entity
Vietnam      2.782177
Colombia    2.923553
Israel       3.084673
Nigeria     3.089714
Gambia      3.103109
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent), dtype: float64
```

```
In [1161]: df73.groupby('Entity')[v8].mean().sort_values().head()
```

```
Out[1161]: Entity
Vietnam      2.035264
Tajikistan   2.512893
Mongolia     2.526100
Uzbekistan   2.541930
Kyrgyzstan   2.553457
Name: Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1162]: df73.groupby('Entity')[v9].mean().sort_values().head()
```

```
Out[1162]: Entity
Vietnam      2.327489
Mongolia     2.951706
Tajikistan   2.970691
Turkmenistan 2.976486
Uzbekistan   2.983257
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [1163]: df73.groupby('Entity')[v10].mean().sort_values().head()
```

```
Out[1163]: Entity
Colombia    2.282888
Honduras    2.476952
Panama      2.478360
Costa Rica  2.498171
Nicaragua   2.499954
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [1164]: df73.groupby('Entity')[v11].mean().sort_values().head()
```

```
Out[1164]: Entity
Vietnam      2.106643
Mongolia     2.862981
Turkmenistan 2.885037
Tajikistan   2.898242
Kyrgyzstan   2.901618
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1165]: df73.groupby('Entity')[v12].mean().sort_values().head()
```

```
Out[1165]: Entity
Vietnam      1.993180
Tajikistan   2.298930
Uzbekistan   2.423248
Niger        2.434416
Mongolia     2.434849
Name: Prevalence - Anxiety disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [1166]: df73.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1166]: Entity
North America 8.664276
United States 8.854439
Australasia   8.977026
Iran          9.353655
New Zealand   11.533277
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [1167]: df73.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1167]: Entity
New Zealand   0.341489
Norway        0.351041
Northern Ireland 0.351092
Germany       0.360590
France         0.379119
Name: Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent), dtype: float64
```

```
In [1168]: df73.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1168]: Entity
Germany       8.707728
Netherlands   8.929120
Norway        10.118733
Northern Ireland 10.131394
New Zealand   10.384905
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [1169]: df73.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1169]: Entity
France        8.284871
Netherlands   8.479243
New Zealand   8.722836
Northern Ireland 9.450220
Norway        9.511781
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1170]: df73.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1170]: Entity
Australia           8.457670
Northern Ireland    8.574515
Australasia         8.949906
Iran                9.266796
New Zealand          11.580087
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1171]: df73.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1171]: Entity
Germany            6.163624
France              6.201208
New Zealand         6.450876
Northern Ireland   6.874145
Norway              6.990689
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent), dtype: float64
```

```
In [1172]: df73.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1172]: Entity
United States       7.468828
Iran                7.636515
New Zealand         7.863288
Norway              8.235737
Northern Ireland   8.583744
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent), dtype: float64
```

```
In [1173]: df73.groupby('Entity')[v8].mean().sort_values().tail()
```

```
Out[1173]: Entity
Australasia         6.895850
Iran                6.932546
Norway              7.637363
Northern Ireland   7.826927
New Zealand          8.651309
Name: Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1174]: df73.groupby('Entity')[v9].mean().sort_values().tail()
```

```
Out[1174]: Entity
Australasia         8.678005
Norway              8.767304
Iran                8.935980
Northern Ireland   9.107339
New Zealand          10.944710
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [1175]: df73.groupby('Entity')[v10].mean().sort_values().tail()
```

```
Out[1175]: Entity
Germany           5.953351
Iran              5.985369
Netherlands       6.000585
Norway            6.738138
Northern Ireland  7.003374
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [1176]: df73.groupby('Entity')[v11].mean().sort_values().tail()
```

```
Out[1176]: Entity
Australasia        8.801687
Iran               8.896348
Norway             8.932567
Northern Ireland   9.094446
New Zealand        11.234584
Name: Prevalence - Anxiety disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1177]: df73.groupby('Entity')[v12].mean().sort_values().tail()
```

```
Out[1177]: Entity
Netherlands         7.035372
United States       7.045792
Norway              7.819327
Northern Ireland    8.046305
New Zealand         8.718694
Name: Prevalence - Anxiety disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

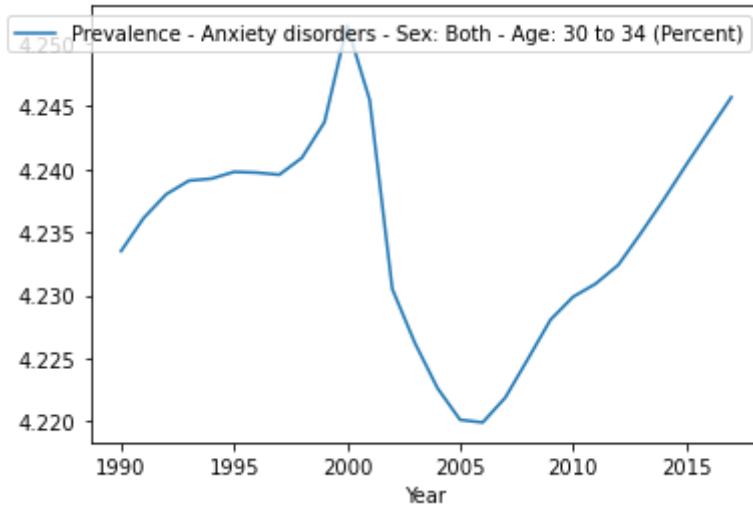
```
In [1178]: df73_mean = df73.groupby('Year').mean()
df73_mean.head()
```

```
Out[1178]:
```

	Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: standard (Percent)
Year								
1990	4.754114	0.129941	4.820296	4.124690	4.665480	2.882426	4.568713	3.95
1991	4.756808	0.130249	4.822124	4.126757	4.667315	2.886765	4.570396	3.96
1992	4.759444	0.130655	4.823608	4.129559	4.668646	2.892242	4.572467	3.96
1993	4.761710	0.131092	4.824780	4.132718	4.669719	2.897643	4.574573	3.96
1994	4.764358	0.131470	4.826022	4.136045	4.671066	2.901648	4.576437	3.96

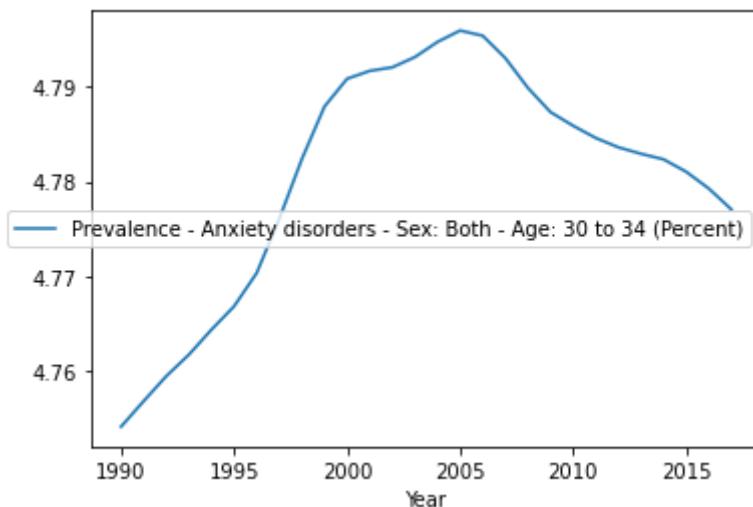
```
In [1179]: df73.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1179]: <AxesSubplot:xlabel='Year'>
```



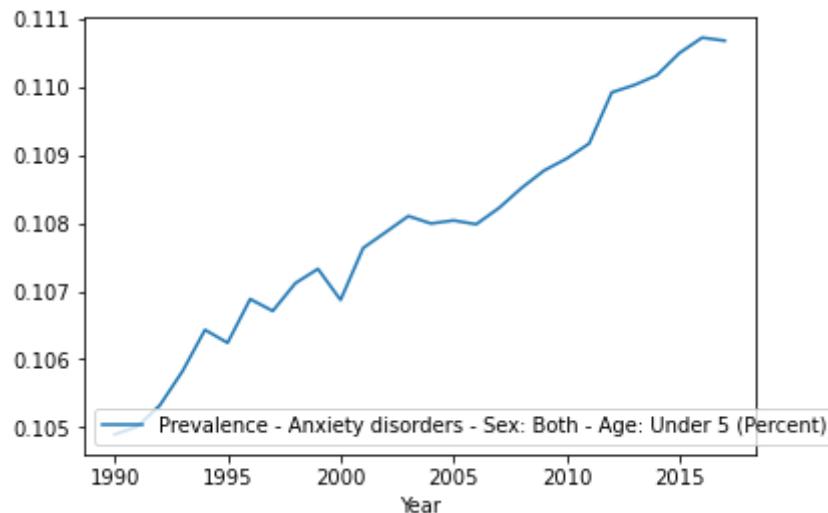
```
In [1180]: df73.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1180]: <AxesSubplot:xlabel='Year'>
```



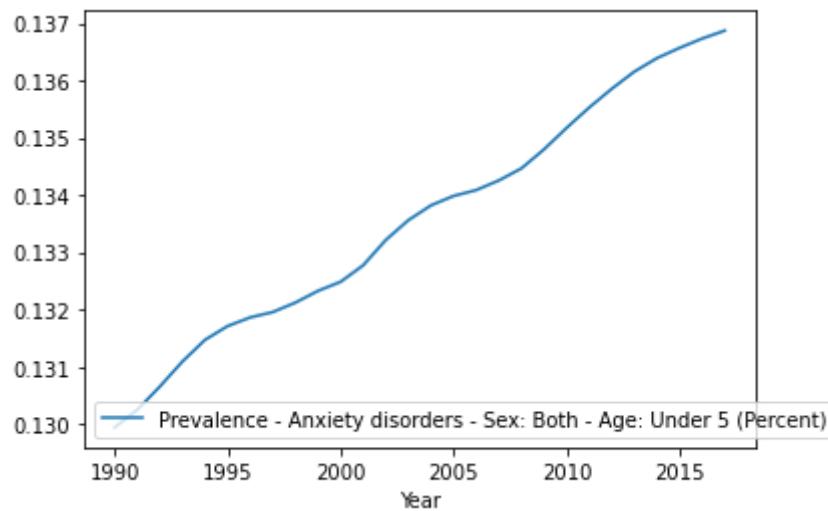
```
In [1181]: df73.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1181]: <AxesSubplot:xlabel='Year'>
```



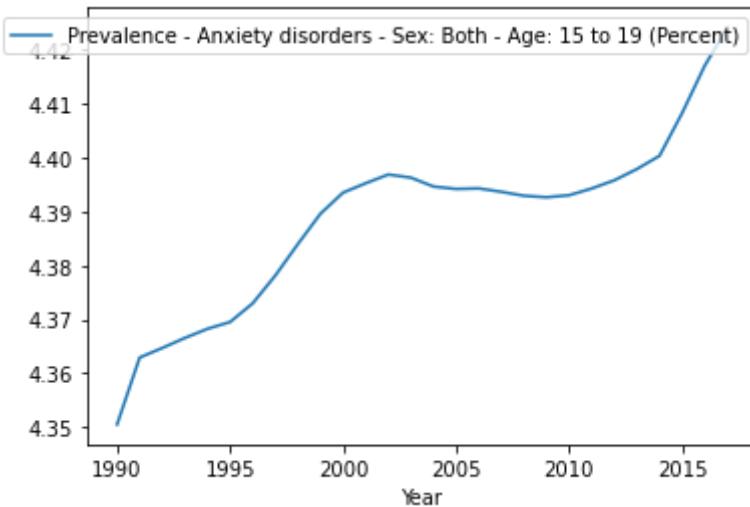
```
In [1182]: df73.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1182]: <AxesSubplot:xlabel='Year'>
```



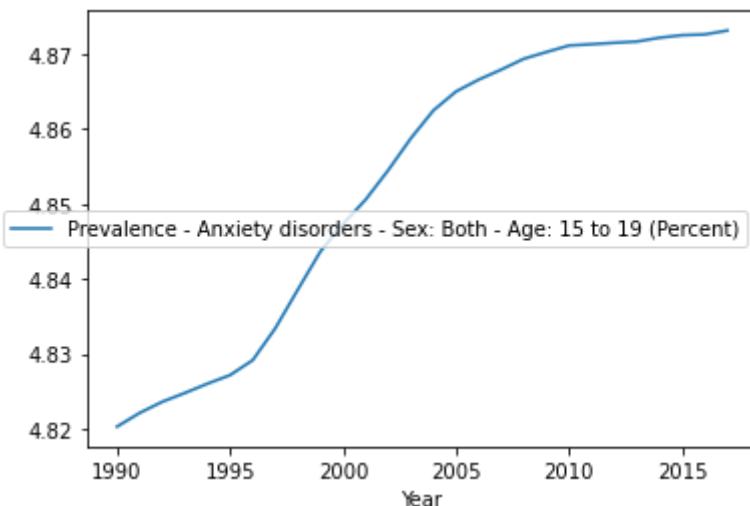
```
In [1183]: df73.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1183]: <AxesSubplot:xlabel='Year'>
```



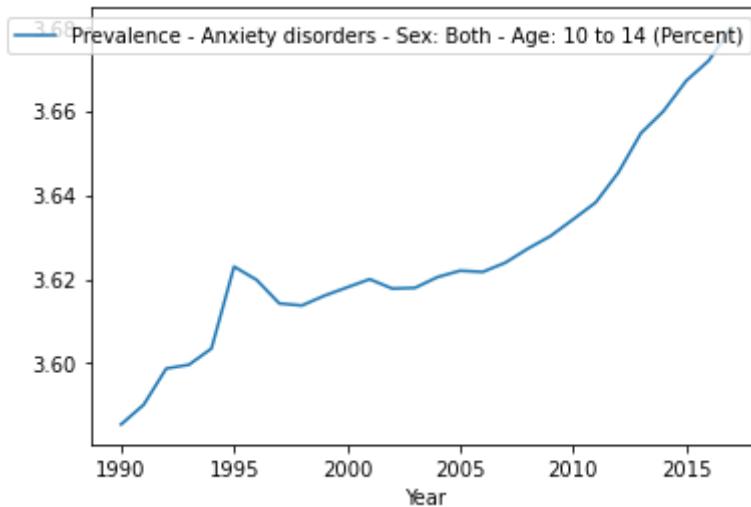
```
In [1184]: df73.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1184]: <AxesSubplot:xlabel='Year'>
```



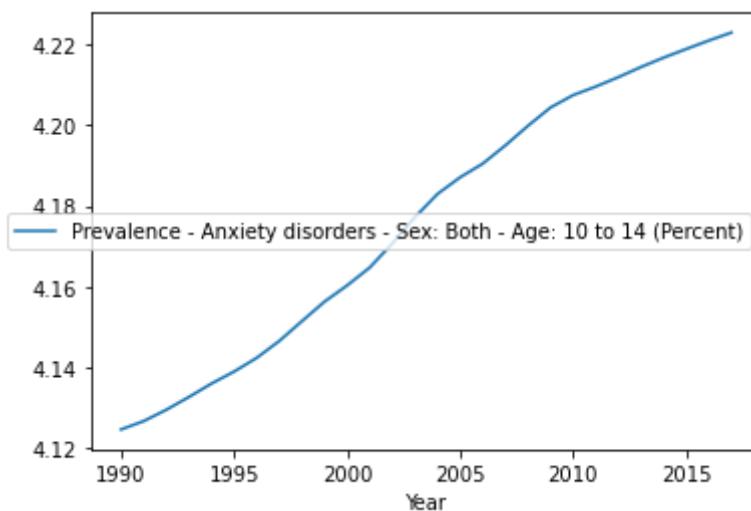
```
In [1185]: df73.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1185]: <AxesSubplot:xlabel='Year'>
```



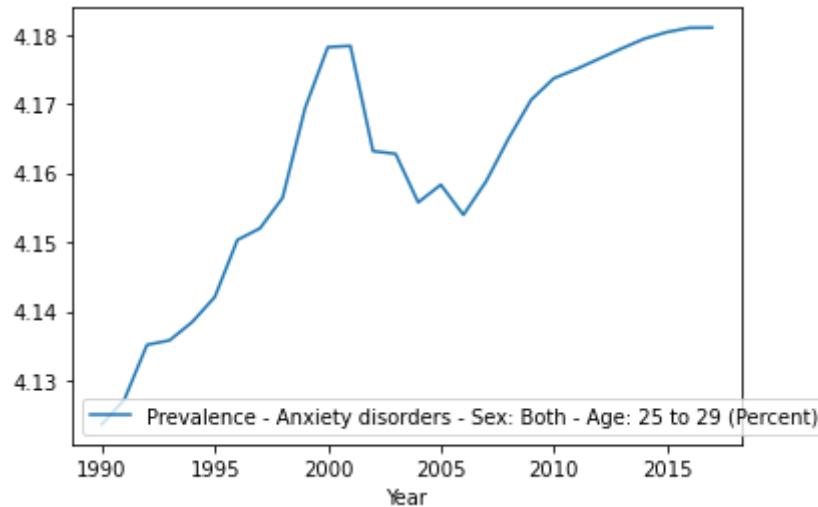
```
In [1186]: df73.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1186]: <AxesSubplot:xlabel='Year'>
```



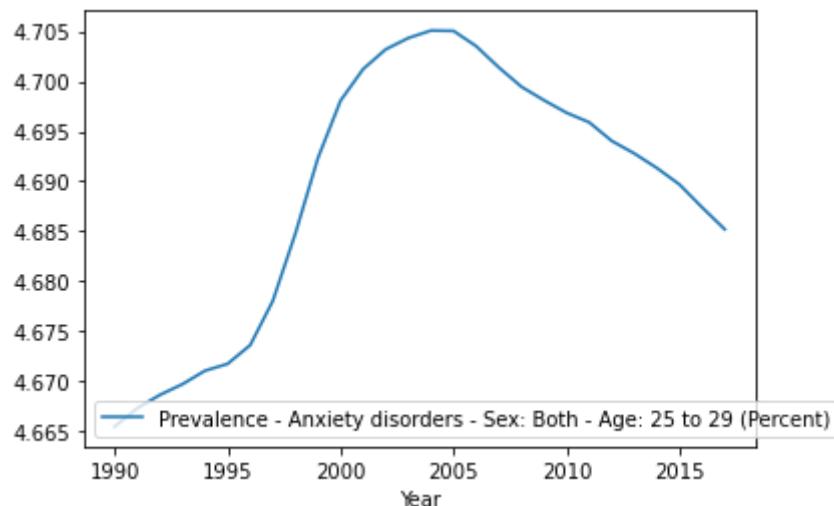
```
In [1187]: df73.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1187]: <AxesSubplot:xlabel='Year'>
```



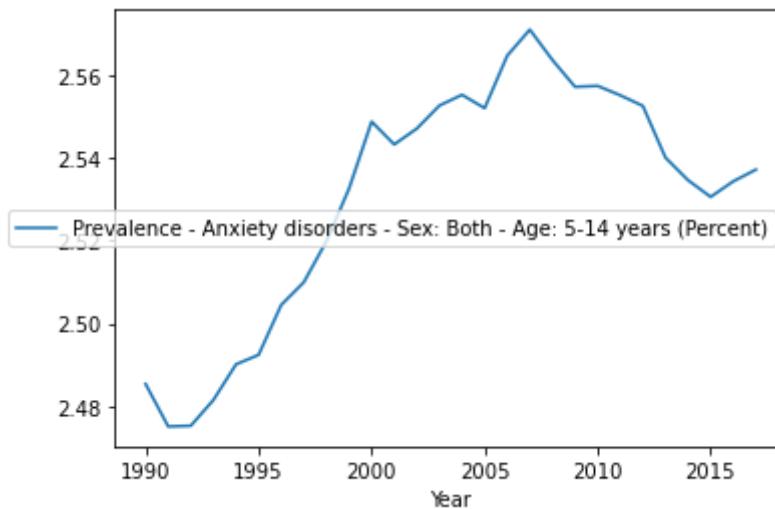
```
In [1188]: df73.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1188]: <AxesSubplot:xlabel='Year'>
```



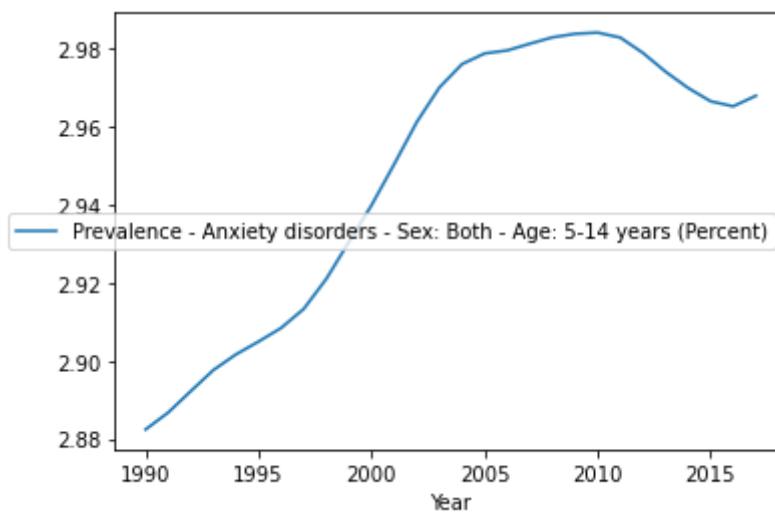
```
In [1189]: df73.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1189]: <AxesSubplot:xlabel='Year'>
```



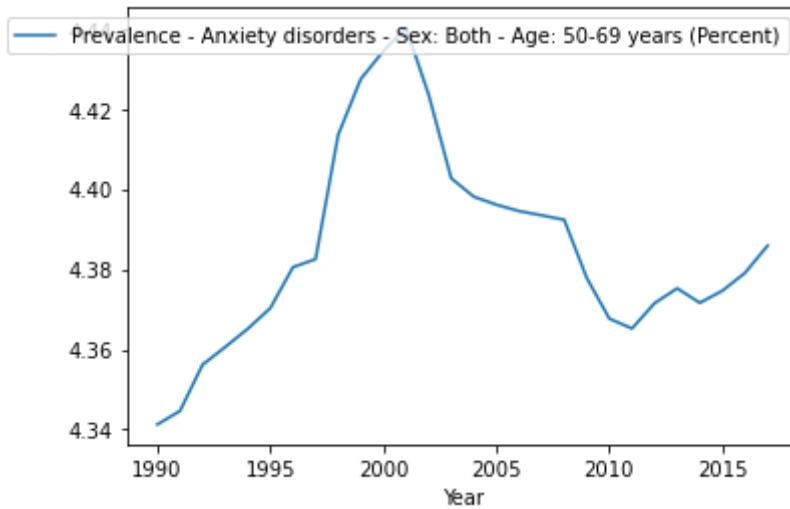
```
In [1190]: df73.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1190]: <AxesSubplot:xlabel='Year'>
```



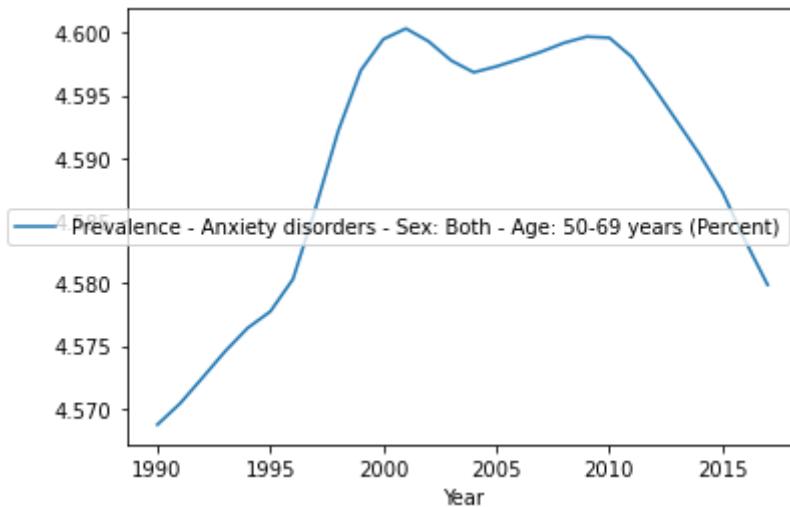
```
In [1191]: df73.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1191]: <AxesSubplot:xlabel='Year'>
```



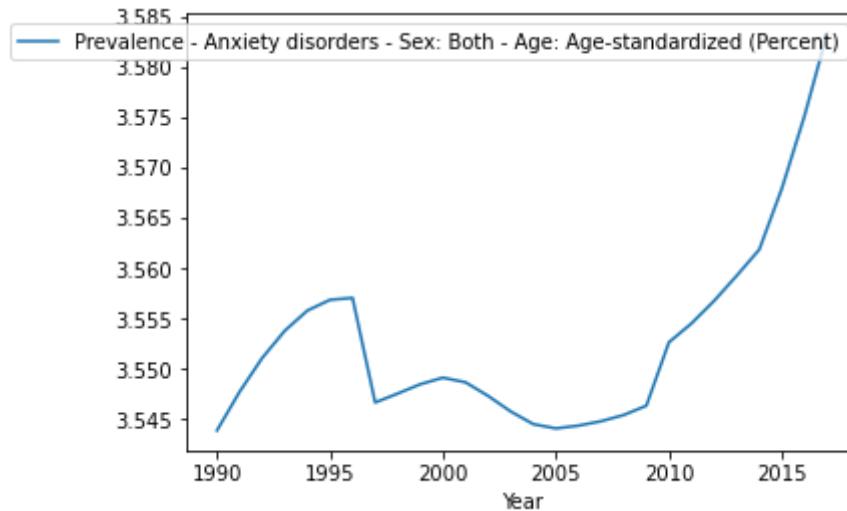
```
In [1192]: df73.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1192]: <AxesSubplot:xlabel='Year'>
```



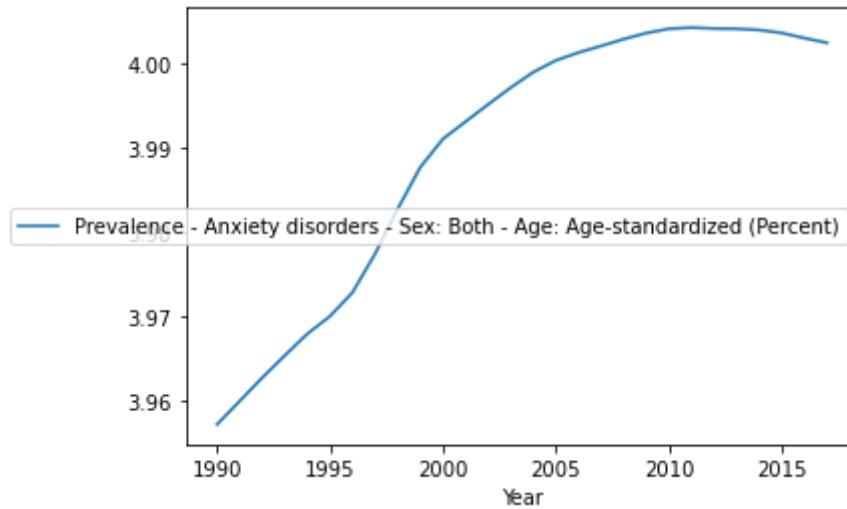
```
In [1193]: df73.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[1193]: <AxesSubplot:xlabel='Year'>
```



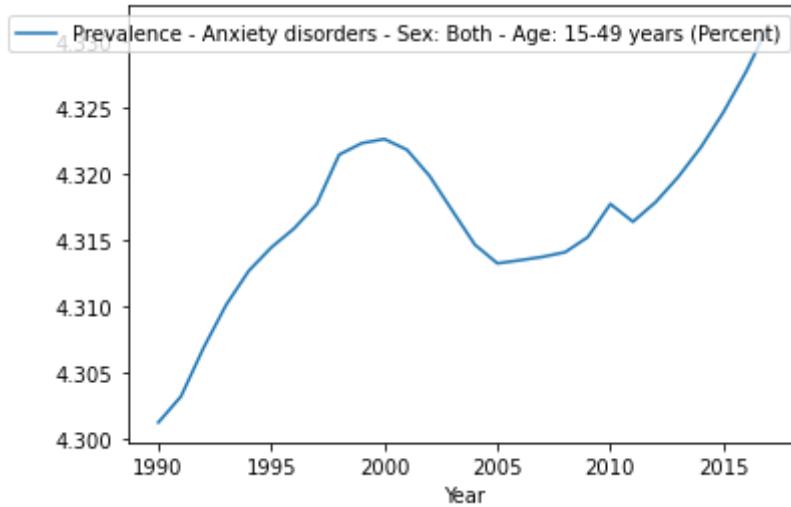
```
In [1194]: df73.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[1194]: <AxesSubplot:xlabel='Year'>
```



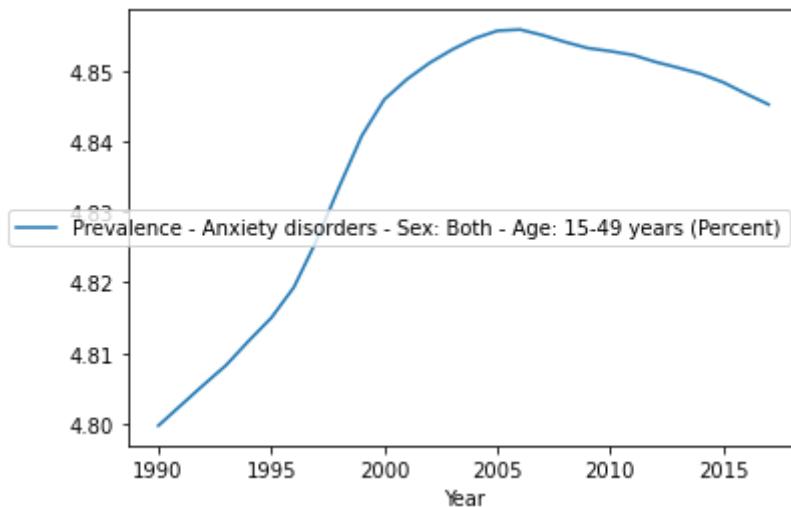
```
In [1195]: df73.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[1195]: <AxesSubplot:xlabel='Year'>
```



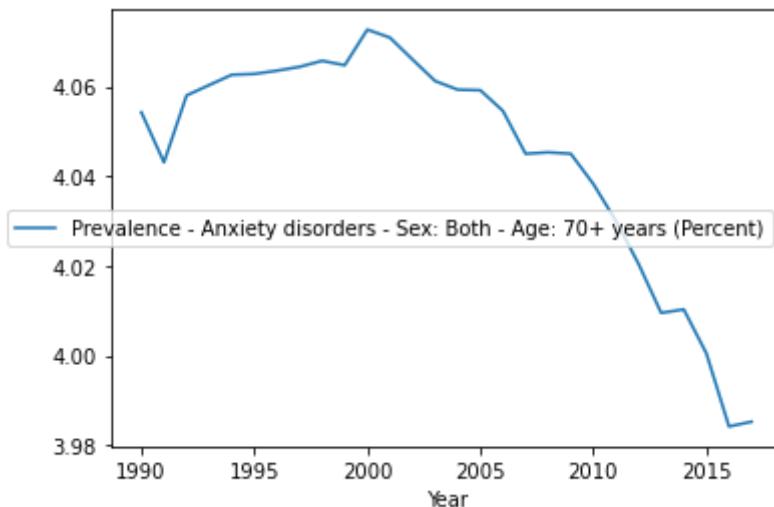
```
In [1196]: df73.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[1196]: <AxesSubplot:xlabel='Year'>
```



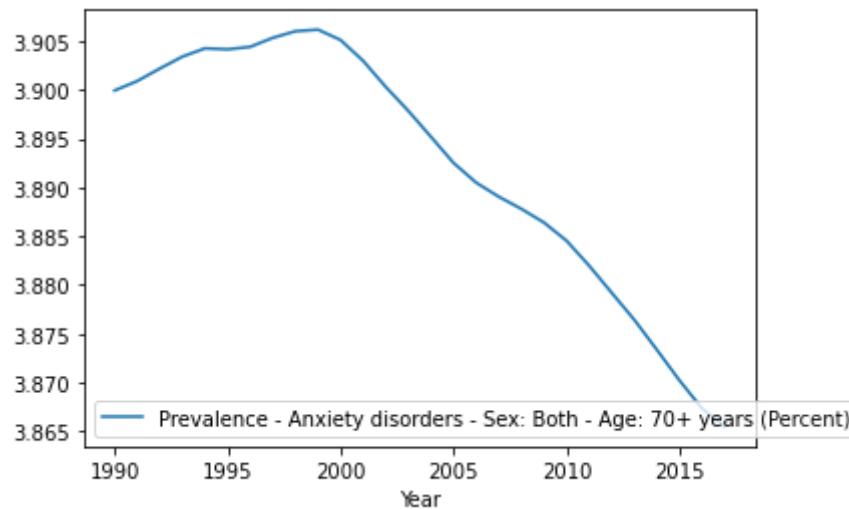
```
In [1197]: df73.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[1197]: <AxesSubplot:xlabel='Year'>
```



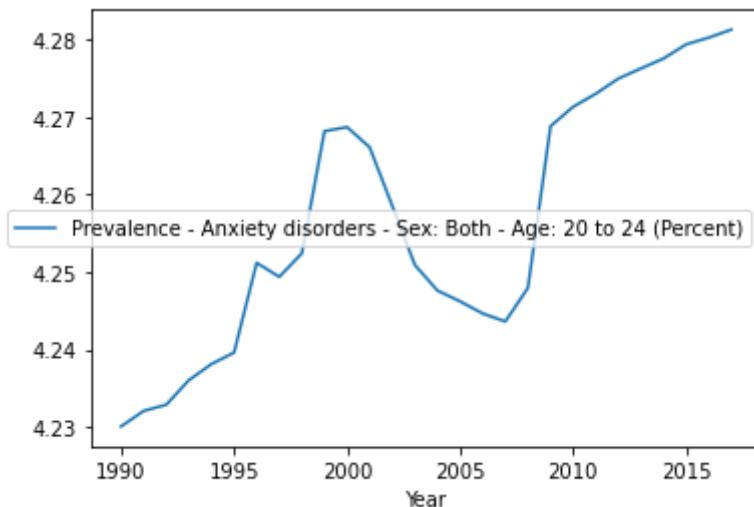
```
In [1198]: df73.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[1198]: <AxesSubplot:xlabel='Year'>
```



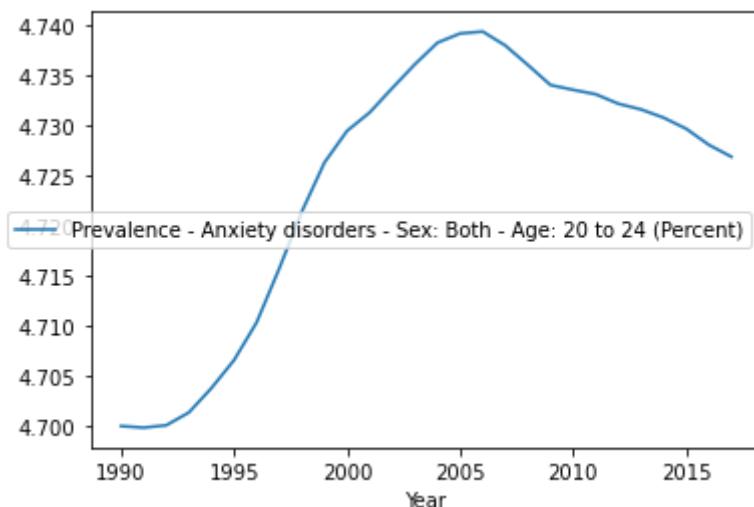
```
In [1199]: df73.groupby('Year')[v11].median().plot(legend=True)
```

```
Out[1199]: <AxesSubplot:xlabel='Year'>
```



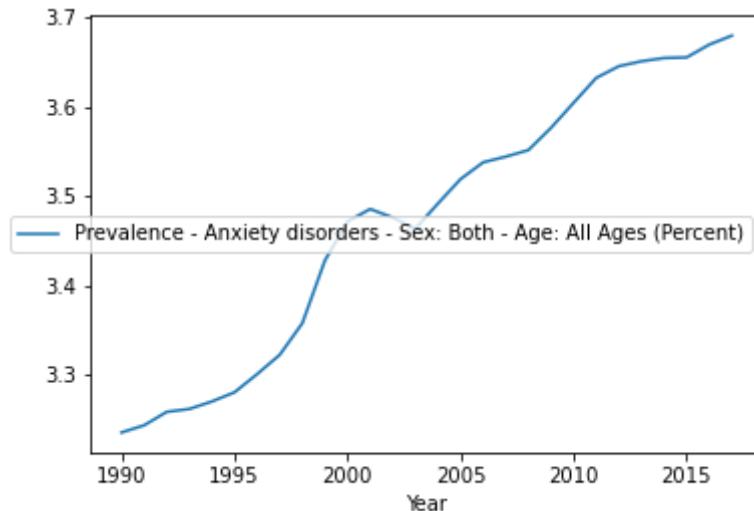
```
In [1200]: df73.groupby('Year')[v11].mean().plot(legend=True)
```

```
Out[1200]: <AxesSubplot:xlabel='Year'>
```



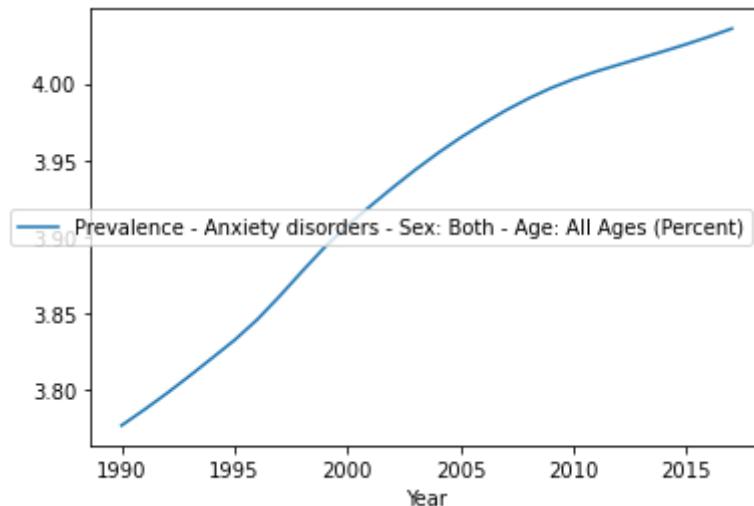
```
In [1201]: df73.groupby('Year')[v12].median().plot(legend=True)
```

```
Out[1201]: <AxesSubplot:xlabel='Year'>
```



```
In [1202]: df73.groupby('Year')[v12].mean().plot(legend=True)
```

```
Out[1202]: <AxesSubplot:xlabel='Year'>
```



Polar Disorder

BIPOLAR DISORDERS

In [1203]: df13.info()
df13.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 9 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
3   DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both
h - Age: 70+ years (Rate)       6468 non-null   float64
4   DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both
h - Age: All Ages (Rate)       6468 non-null   float64
5   DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both
h - Age: 5-14 years (Rate)     6468 non-null   float64
6   DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both
h - Age: 50-69 years (Rate)    6468 non-null   float64
7   DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both
h - Age: 15-49 years (Rate)    6468 non-null   float64
8   DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both
h - Age: Age-standardized (Rate) 6468 non-null   float64
dtypes: float64(6), int64(1), object(2)
memory usage: 454.9+ KB
```

Out[1203]:

	Entity	Code	Year	DALYs (Disability- Adjusted Life Years)					
0	Afghanistan	AFG	1990	100.461534	117.002539	35.611379	165.088225	203.210665	135.1209
1	Afghanistan	AFG	1991	100.330219	117.444972	34.711181	164.692554	203.527586	135.1421
2	Afghanistan	AFG	1992	100.040891	118.723259	33.931767	164.973584	203.884052	135.3138
3	Afghanistan	AFG	1993	100.286266	119.092281	33.095571	165.030504	203.768239	135.4345
4	Afghanistan	AFG	1994	100.295590	116.990653	32.225286	165.171913	202.959894	135.5125

Checking for missing values:

```
In [1204]: missing = pd.concat([df13.isnull().sum(), 100 * df13.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1204]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All Ages (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

```
In [1205]: v1='DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate)'
v2='DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All Ages (Rate)'
v3='DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14 years (Rate)'
v4='DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69 years (Rate)'
v5='DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49 years (Rate)'
v6='DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age-standardized (Rate)'
```

In [1206]: df13.describe()

Out[1206]:

Year	DALYs (Disability- Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All	DALYs (Disability- Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14	DALYs (Disability- Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69	DALYs (Disability- Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49	DALYs (Disability- Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: Age- standardized (Rate)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	101.744861	138.572339	33.827866	172.759367	203.578714
std	8.078372	12.226197	39.128467	9.892530	28.146030	49.807081
min	1990.000000	74.291836	62.166600	11.011586	96.735152	80.452406
25%	1996.750000	94.040086	105.351183	28.008345	156.788925	176.391726
50%	2003.500000	97.621699	137.814577	31.382137	166.595818	192.308649
75%	2010.250000	108.736912	167.370707	39.722929	196.412457	239.394521
max	2017.000000	148.058252	233.199891	65.584478	249.574553	334.730653
						218.149477

In [1207]: df13.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1207]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate)
4231	Northern Ireland	1993	148.058252
4229	Northern Ireland	1991	148.036587
4232	Northern Ireland	1994	147.866877
4230	Northern Ireland	1992	147.825271
4228	Northern Ireland	1990	147.727679

In [1208]: df13.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1208]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All Ages (Rate)
783	Brazil	2017	233.199891
782	Brazil	2016	233.157239
781	Brazil	2015	233.008440
780	Brazil	2014	232.867562
779	Brazil	2013	232.687318

In [1209]: df13.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1209]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14 years (Rate)
776	Brazil	2010	65.584478
775	Brazil	2009	65.580032
5872	Tropical Latin America	2010	65.432423
5871	Tropical Latin America	2009	65.419556
777	Brazil	2011	65.354992

In [1210]: df13.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1210]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69 years (Rate)
4254	Northern Ireland	2016	249.574553
4255	Northern Ireland	2017	249.556874
4251	Northern Ireland	2013	249.517202
4253	Northern Ireland	2015	249.291745
4245	Northern Ireland	2007	249.201157

In [1211]: df13.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1211]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49 years (Rate)
756	Brazil	1990	334.730653
5852	Tropical Latin America	1990	334.527077
757	Brazil	1991	334.421911
5853	Tropical Latin America	1991	334.221999
758	Brazil	1992	334.189022

In [1212]: df13.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1212]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: Age-standardized (Rate)
4025	New Zealand	2011	218.149477
4024	New Zealand	2010	218.127361
4023	New Zealand	2009	218.079320
4026	New Zealand	2012	218.036234
4028	New Zealand	2014	217.994816

In [1213]: df13.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1213]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate)
5654	Taiwan	2016	74.291836
5655	Taiwan	2017	74.302728
5653	Taiwan	2015	74.373286
5652	Taiwan	2014	74.782785
4199	North Korea	2017	74.807526

In [1214]: df13.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1214]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All Ages (Rate)
1260	China	1990	62.166600
1680	East Asia	1990	62.327560
1261	China	1991	62.454512
1681	East Asia	1991	62.611114
1262	China	1992	62.729126

In [1215]: df13.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1215]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14 years (Rate)
1263	China	1993	11.011586
1264	China	1994	11.043957
1683	East Asia	1993	11.088234
1684	East Asia	1994	11.113804
1262	China	1992	11.170043

In [1216]: df13.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1216]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69 years (Rate)
1275	China	2005	96.735152
1276	China	2006	96.782183
1695	East Asia	2005	96.849137
1696	East Asia	2006	96.887783
1274	China	2004	96.889031

In [1217]: `df13.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()`

Out[1217]:

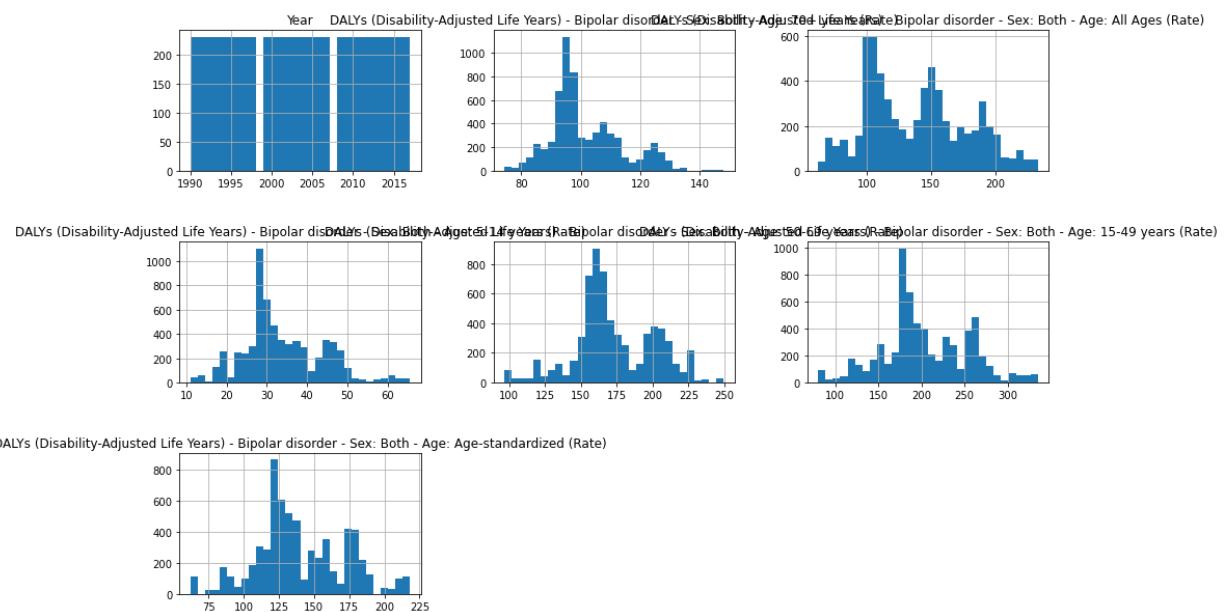
Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49 years (Rate)
1260	China 1990	80.452406
1261	China 1991	80.520243
1262	China 1992	80.535632
1264	China 1994	80.614319
1263	China 1993	80.621560

In [1218]: `df13.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()`

Out[1218]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: Age-standardized (Rate)
1260	China 1990	62.078692
1261	China 1991	62.127037
1262	China 1992	62.148625
1264	China 1994	62.187529
1263	China 1993	62.196767

In [1219]: `df13.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [1220]: df13.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1220]: Entity
North Korea      76.207309
Taiwan          77.962684
United States    80.115514
North America    80.820980
Greenland        81.713696
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [1221]: df13.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1221]: Entity
North Korea      68.239436
China            68.781626
East Asia         68.879696
Papua New Guinea 71.509959
Solomon Islands   72.130867
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1222]: df13.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1222]: Entity
China            12.593942
East Asia         12.624335
North Korea       12.946801
Taiwan           13.976110
Southeast Asia, East Asia, and Oceania 16.483511
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [1223]: df13.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1223]: Entity
China            97.272471
East Asia         97.371193
North Korea       97.573015
Taiwan           102.492463
Southeast Asia, East Asia, and Oceania 110.293696
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [1224]: df13.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1224]: Entity
China            81.225814
East Asia         81.418982
North Korea       85.343785
Taiwan           89.094144
Southeast Asia, East Asia, and Oceania 99.762744
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1225]: df13.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1225]: Entity
China                      62.360476
East Asia                  62.466473
North Korea                64.145872
Taiwan                     66.999443
Southeast Asia, East Asia, and Oceania    74.722884
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [1226]: df13.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1226]: Entity
Sweden                    129.726162
United Kingdom            130.488291
Finland                   130.999104
England                   131.404127
Northern Ireland          143.964705
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [1227]: df13.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1227]: Entity
Australasia               219.958459
Tropical Latin America    226.159310
New Zealand                226.178961
Northern Ireland           226.329645
Brazil                     226.768847
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1228]: df13.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1228]: Entity
New Zealand                60.109576
Australasia                61.009575
Australia                  61.200926
Tropical Latin America     63.684579
Brazil                     63.859130
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [1229]: df13.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1229]: Entity
England           227.785745
Sweden            227.825820
Paraguay          228.249218
New Zealand       236.523962
Northern Ireland  248.356401
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [1230]: df13.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1230]: Entity
Northern Ireland   316.123480
New Zealand        322.452188
Paraguay           325.857166
Tropical Latin America 330.359811
Brazil             330.480020
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1231]: df13.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1231]: Entity
Paraguay           212.211054
Northern Ireland   215.965908
Tropical Latin America 216.648784
Brazil             216.784845
New Zealand        217.416623
Name: DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

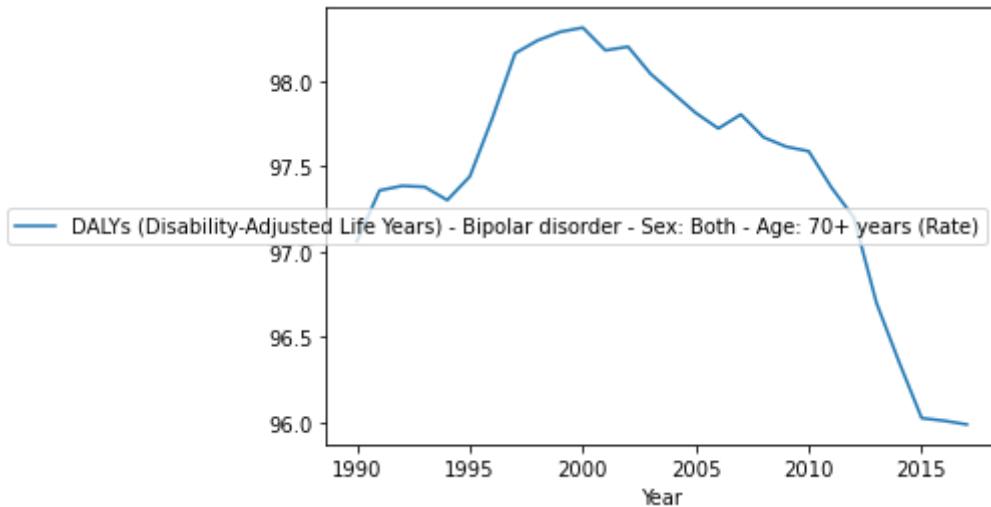
```
In [1232]: df13_mean = df69.groupby('Year').mean()
df13_mean.head()
```

```
Out[1232]:
```

	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Alcohol disorders - Sex: Both - Age: Age-standardized (Percent)
Year							
1990	0.209548	0.715392	0.221774	3.957269	0.808283	3.506288	1.540
1991	0.209586	0.715740	0.222076	3.960009	0.813466	3.510948	1.550
1992	0.209634	0.716091	0.222481	3.962778	0.818692	3.515033	1.550
1993	0.209690	0.716430	0.223033	3.965405	0.823780	3.518531	1.560
1994	0.209751	0.716755	0.223710	3.967976	0.828389	3.521437	1.570

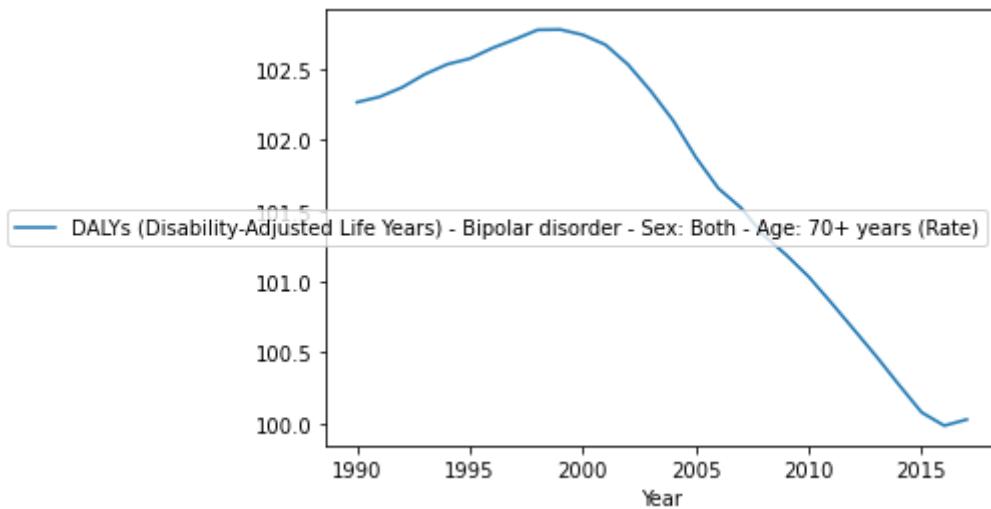
```
In [1233]: df13.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1233]: <AxesSubplot:xlabel='Year'>
```



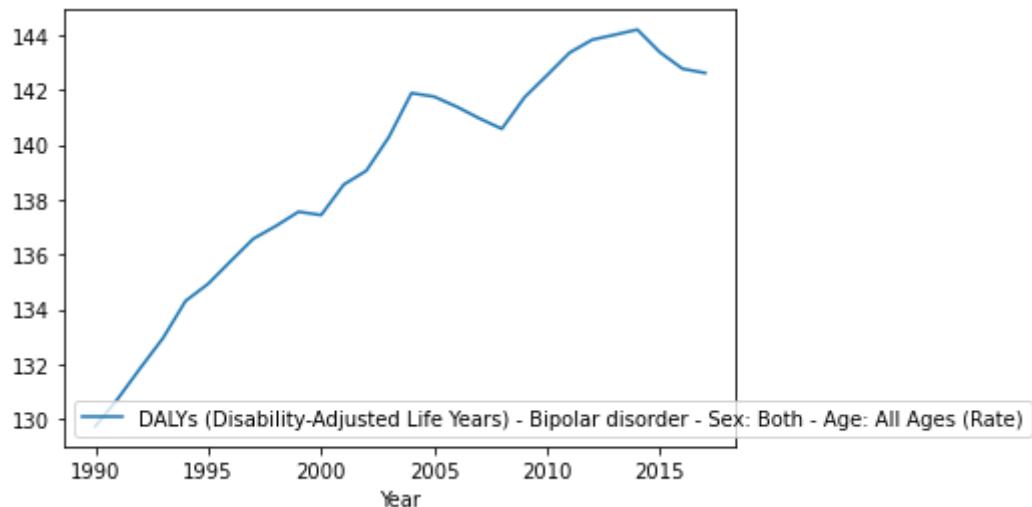
```
In [1234]: df13.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1234]: <AxesSubplot:xlabel='Year'>
```



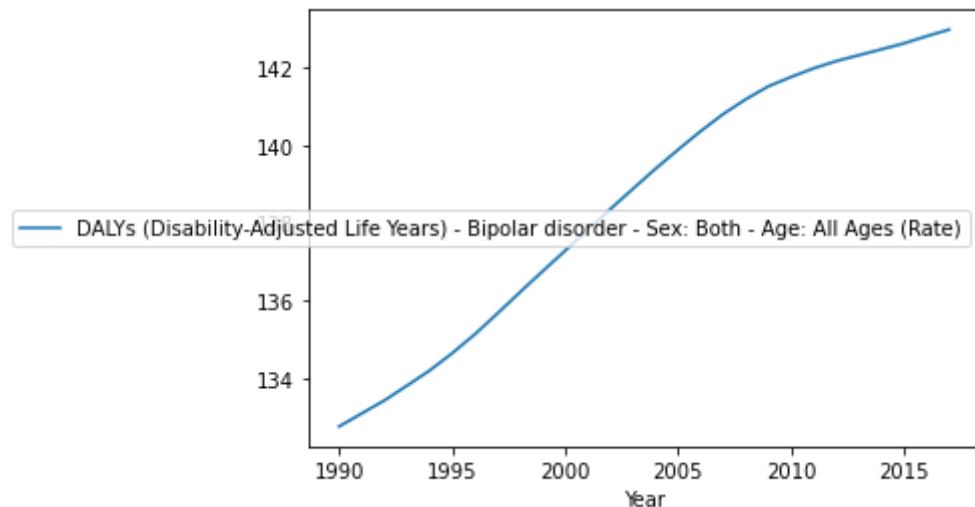
```
In [1235]: df13.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1235]: <AxesSubplot:xlabel='Year'>
```



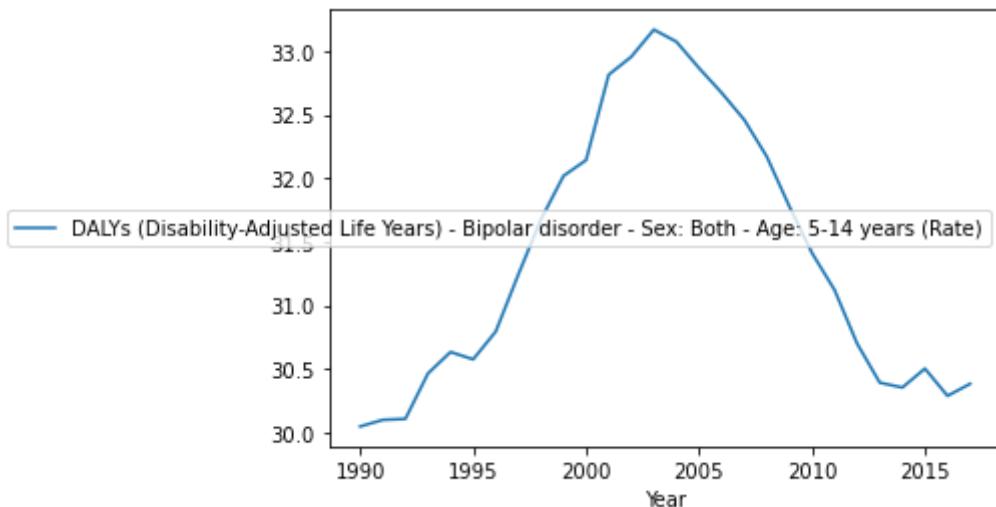
```
In [1236]: df13.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1236]: <AxesSubplot:xlabel='Year'>
```



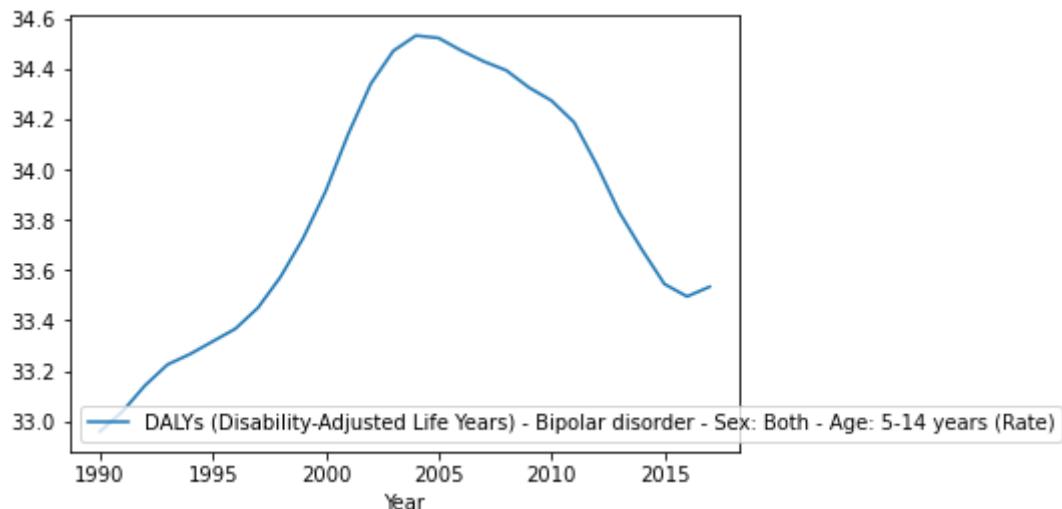
```
In [1237]: df13.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1237]: <AxesSubplot:xlabel='Year'>
```



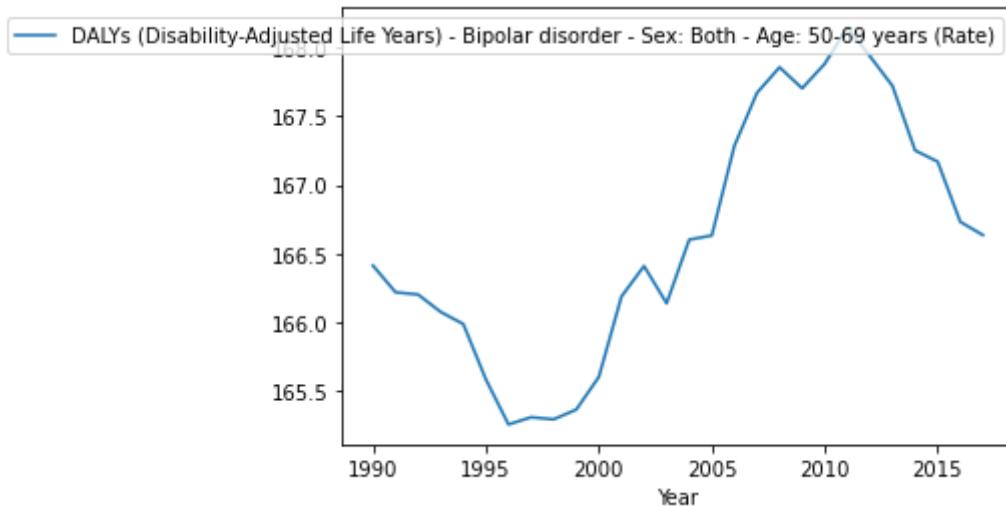
```
In [1238]: df13.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1238]: <AxesSubplot:xlabel='Year'>
```



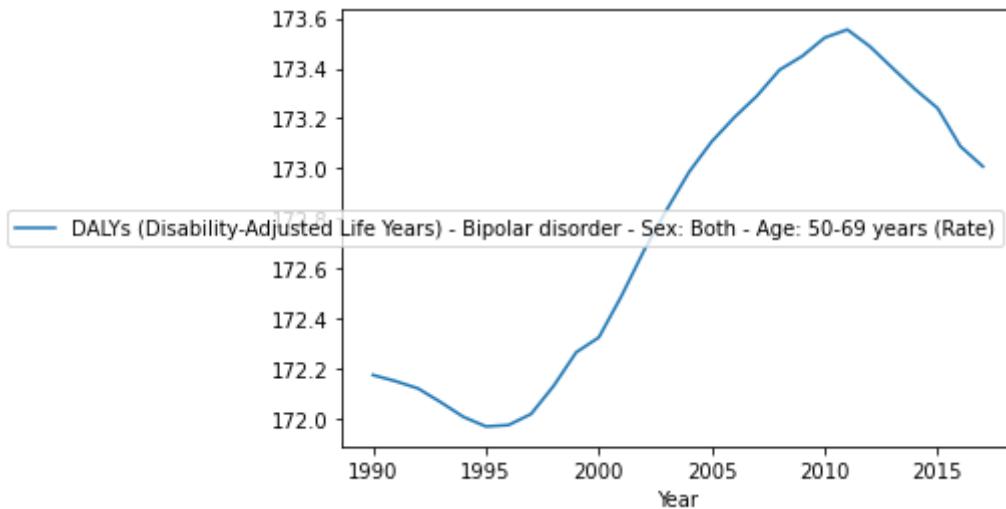
```
In [1239]: df13.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1239]: <AxesSubplot:xlabel='Year'>
```



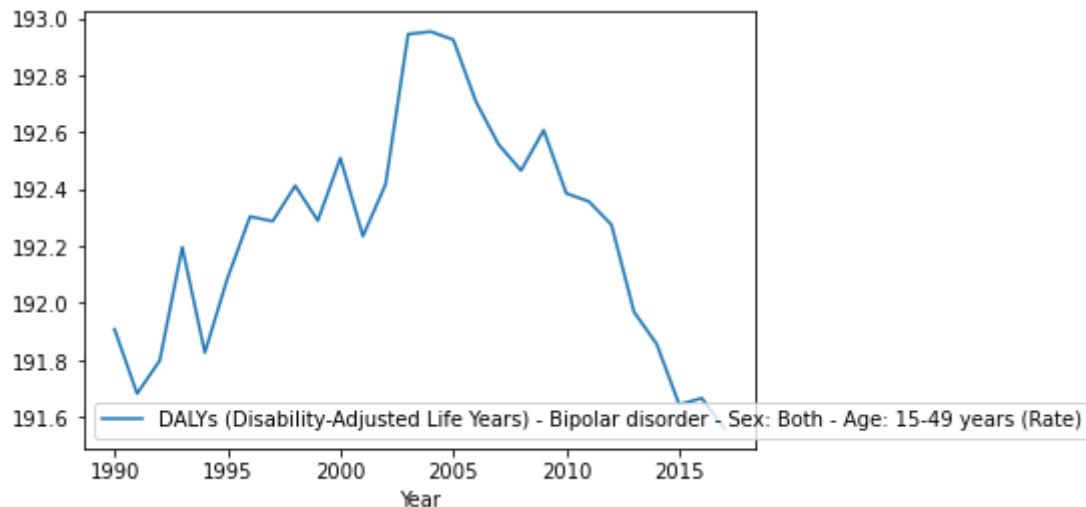
```
In [1240]: df13.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1240]: <AxesSubplot:xlabel='Year'>
```



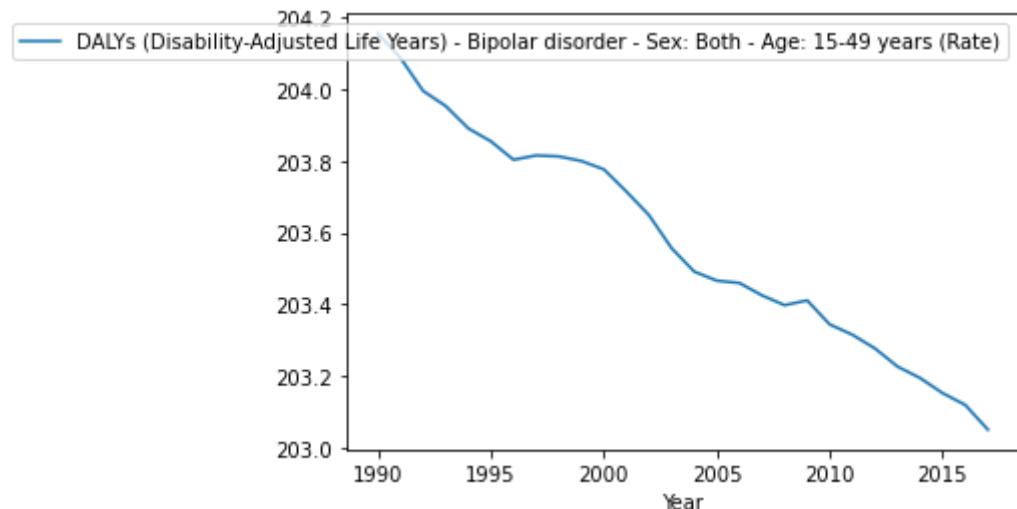
```
In [1241]: df13.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1241]: <AxesSubplot:xlabel='Year'>
```



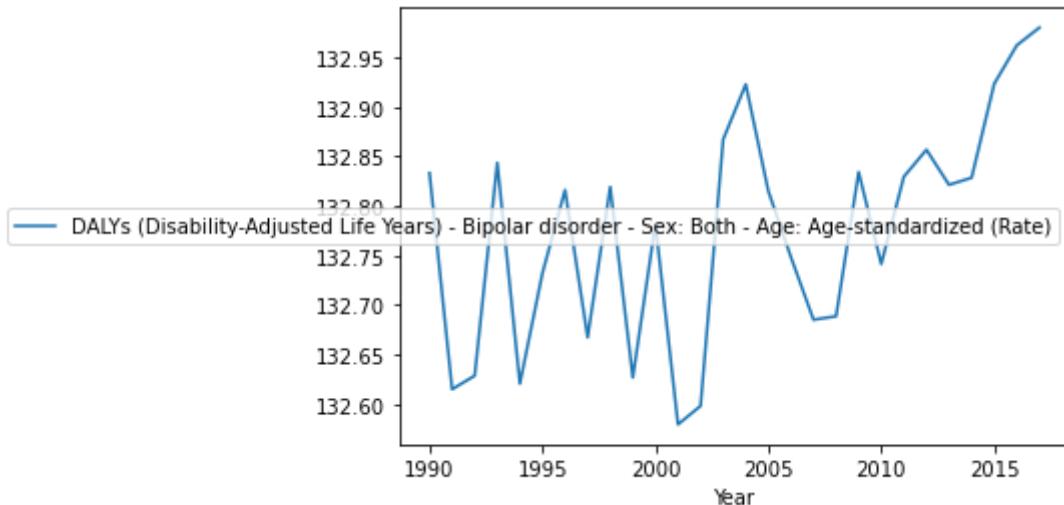
```
In [1242]: df13.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1242]: <AxesSubplot:xlabel='Year'>
```



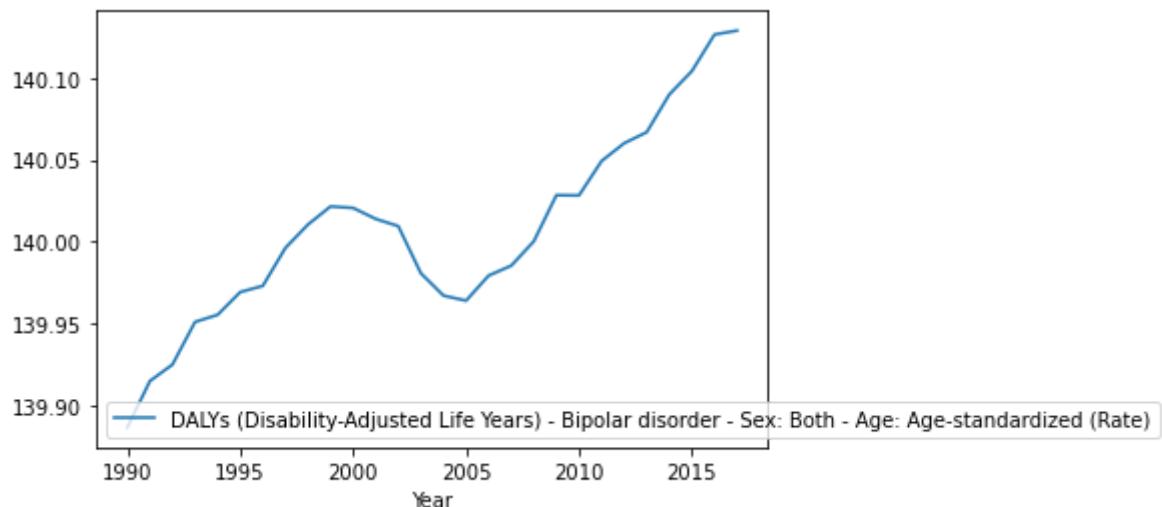
```
In [1243]: df13.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1243]: <AxesSubplot:xlabel='Year'>
```



```
In [1244]: df13.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1244]: <AxesSubplot:xlabel='Year'>
```



In [1245]: df50.info()
df50.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)
6468 non-null   float64
4   Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)
6468 non-null   float64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1245]:

	Entity	Code	Year	Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)	Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)
0	Afghanistan	AFG	1990	28930.352533	30481.738451
1	Afghanistan	AFG	1991	30104.899474	31779.833154
2	Afghanistan	AFG	1992	35784.508576	37793.317571
3	Afghanistan	AFG	1993	41773.459665	43951.501684
4	Afghanistan	AFG	1994	43145.724983	45350.008752

Checking for missing values:

In [1246]: missing = pd.concat([df50.isnull().sum(), 100 * df50.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1246]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)		0	0.000000
Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)		0	0.000000
Code	980	15.151515	

```
In [1247]: v1='Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)'
v2='Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)'
```

```
In [1248]: df50.describe()
```

```
Out[1248]:
```

	Year	Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)	Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)
count	6468.000000	6.468000e+03	6.468000e+03
mean	2003.500000	3.634099e+05	3.993715e+05
std	8.078372	1.393764e+06	1.512047e+06
min	1990.000000	7.431531e+01	7.087182e+01
25%	1996.750000	7.960718e+03	8.425412e+03
50%	2003.500000	2.906671e+04	3.270289e+04
75%	2010.250000	1.339652e+05	1.571981e+05
max	2017.000000	2.183191e+07	2.371746e+07

```
In [1249]: df50.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

```
Out[1249]:
```

	Entity	Year	Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)
6383	World	2017	2.183191e+07
6382	World	2016	2.154060e+07
6381	World	2015	2.124709e+07
6380	World	2014	2.095999e+07
6379	World	2013	2.067144e+07

```
In [1250]: df50.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()
```

```
Out[1250]:
```

	Entity	Year	Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)
6383	World	2017	2.371746e+07
6382	World	2016	2.340858e+07
6381	World	2015	2.309943e+07
6380	World	2014	2.279331e+07
6379	World	2013	2.248299e+07

```
In [1251]: df50.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[1251]:

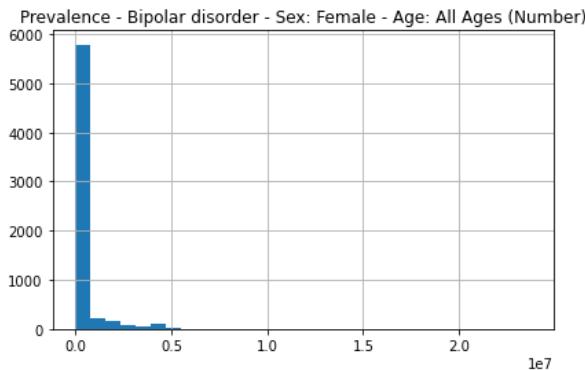
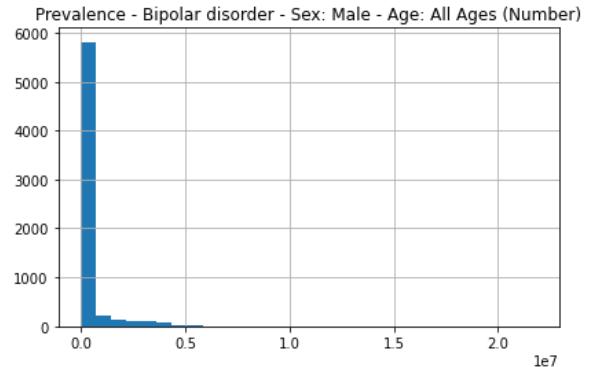
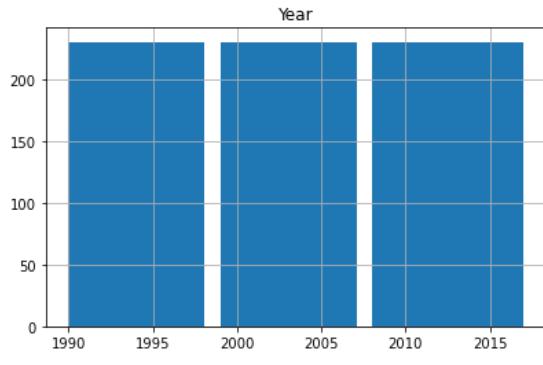
	Entity	Year	Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)
3584	Marshall Islands	1990	74.315310
3585	Marshall Islands	1991	76.583455
3586	Marshall Islands	1992	78.863475
3587	Marshall Islands	1993	81.162074
3588	Marshall Islands	1994	83.478514

```
In [1252]: df50.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()
```

Out[1252]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)
3584	Marshall Islands	1990	70.871824
3585	Marshall Islands	1991	73.019413
3586	Marshall Islands	1992	75.206093
3587	Marshall Islands	1993	77.433679
3588	Marshall Islands	1994	79.685878

```
In [1253]: df50.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1254]: df50.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1254]: Entity
Marshall Islands      98.127856
American Samoa       114.772093
Northern Mariana Islands 132.888531
Greenland             166.184854
Kiribati              168.004801
Name: Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number),
dtype: float64
```

```
In [1255]: df50.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1255]: Entity
Marshall Islands      93.259317
American Samoa        114.527169
Northern Mariana Islands 137.060511
Greenland              159.738276
Kiribati               183.943763
Name: Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number),
r, dtype: float64
```

```
In [1256]: df50.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1256]: Entity
Low-middle SDI        3.690940e+06
South Asia             3.826127e+06
Southeast Asia, East Asia, and Oceania 3.844390e+06
Middle SDI              4.675873e+06
World                  1.785197e+07
Name: Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number),
dtype: float64
```

```
In [1257]: df50.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1257]: Entity
High-middle SDI        3.822399e+06
High-income             4.350488e+06
High SDI                4.428088e+06
Middle SDI              4.931024e+06
World                  1.945979e+07
Name: Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number),
r, dtype: float64
```

```
In [1258]: df50_mean = df50.groupby('Year').mean()
df50_mean.head()
```

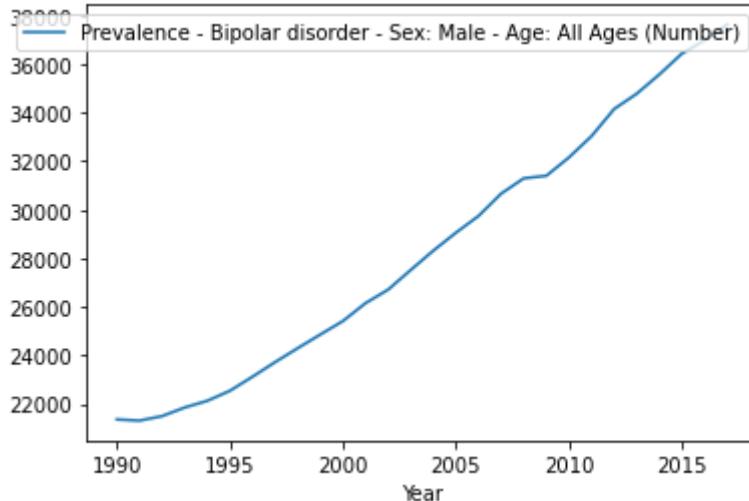
Out[1258]:

Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)	Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)
--	--

Year	Prevalence - Bipolar disorder - Sex: Male - Age: All Ages (Number)	Prevalence - Bipolar disorder - Sex: Female - Age: All Ages (Number)
1990	287957.381603	318235.668171
1991	292987.189096	323600.564137
1992	298073.379034	329037.003962
1993	303124.692244	334448.834527
1994	308206.825258	339941.043628

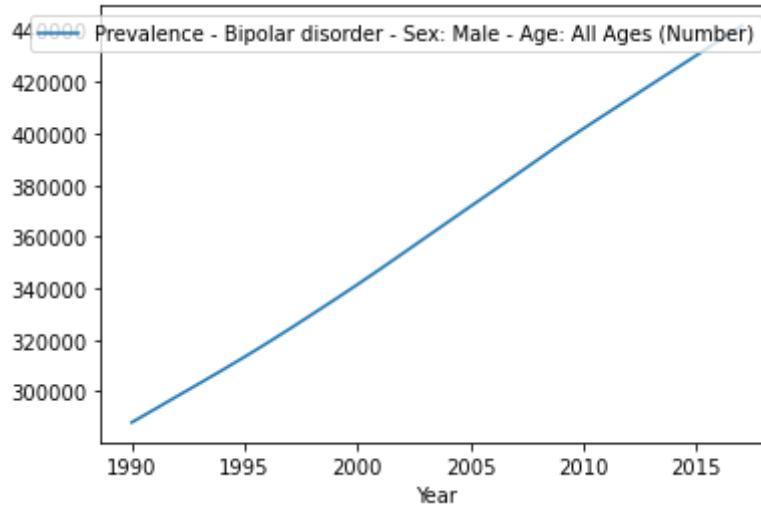
```
In [1259]: df50.groupby('Year')[v1].median().plot(legend=True)
```

Out[1259]: <AxesSubplot:xlabel='Year'>



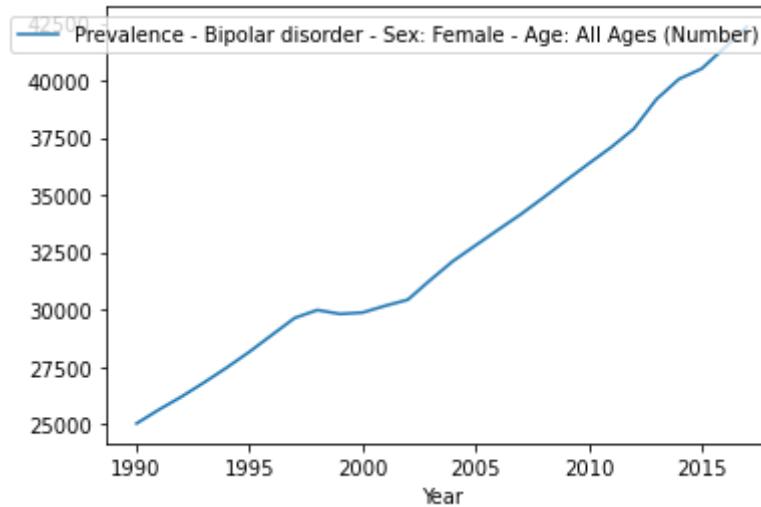
```
In [1260]: df50.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1260]: <AxesSubplot:xlabel='Year'>
```



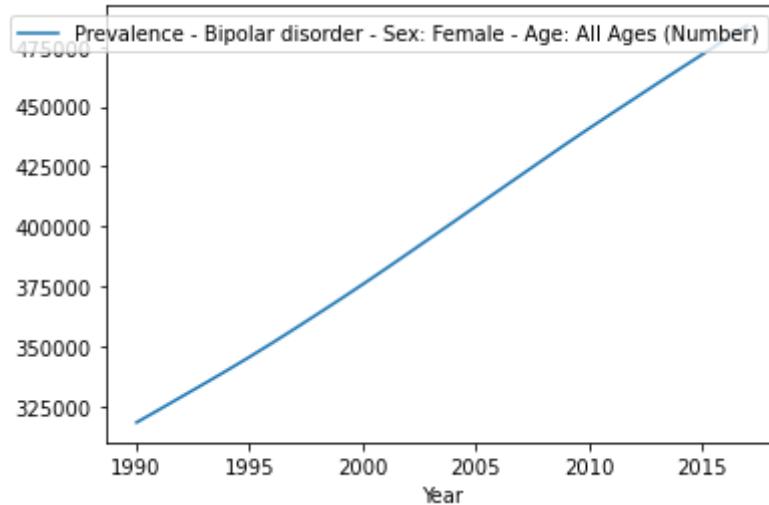
```
In [1261]: df50.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1261]: <AxesSubplot:xlabel='Year'>
```



```
In [1262]: df50.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1262]: <AxesSubplot:xlabel='Year'>
```



In [1263]: df75.info()
df75.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 13 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   float64
3   Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)
6468 non-null   float64
4   Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)
6468 non-null   float64
5   Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)
6468 non-null   float64
6   Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)
6468 non-null   float64
7   Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)
6468 non-null   float64
8   Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)
6468 non-null   float64
9   Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
6468 non-null   float64
10  Prevalence - Bipolar disorder - Sex: Both - Age: 70+ years (Percent)
6468 non-null   float64
11  Prevalence - Bipolar disorder - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
12  Prevalence - Bipolar disorder - Sex: Both - Age: 30 to 34 (Percent)
6468 non-null   float64
dtypes: float64(10), int64(1), object(2)
memory usage: 657.0+ KB
```

Out[1263]:

	Entity	Code	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: All Ages (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 30 to 34 (Percent)	Bipo
											disorde
											Sex: Bot
											standardiz
											(Perce
0	Afghanistan	AFG	1990	0.935195	0.347879	1.047479	0.175163	1.025670	0.6977		
1	Afghanistan	AFG	1991	0.934762	0.348015	1.047705	0.169763	1.026357	0.6979		
2	Afghanistan	AFG	1992	0.933837	0.348121	1.048098	0.165198	1.025695	0.6981		
3	Afghanistan	AFG	1993	0.933224	0.348278	1.048190	0.160896	1.022701	0.6982		
4	Afghanistan	AFG	1994	0.933310	0.348497	1.047928	0.156515	1.018788	0.6984		

Checking for missing values:

```
In [1264]: missing = pd.concat([df75.isnull().sum(), 100 * df75.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1264]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: All Ages (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Code	980	15.151515	

```
In [1265]: v1='Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)'
v2='Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)'
v3='Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)'
v4='Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)'
v5='Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)'
v6='Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)'
v7='Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)'
v8='Prevalence - Bipolar disorder - Sex: Both - Age: 70+ years (Percent)'
v9='Prevalence - Bipolar disorder - Sex: Both - Age: All Ages (Percent)'
v10='Prevalence - Bipolar disorder - Sex: Both - Age: 30 to 34 (Percent)'
```

In [1266]: df75.describe()

Out[1266]:

	Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age- standardized (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.958576	0.352264	1.073972	0.174889	1.002640
std	8.078372	0.262120	0.105831	0.306089	0.057302	0.249537
min	1990.000000	0.371677	0.121939	0.398073	0.056045	0.392653
25%	1996.750000	0.790133	0.298344	0.899652	0.136956	0.867002
50%	2003.500000	0.915396	0.314743	0.995804	0.162048	0.953705
75%	2010.250000	1.156216	0.426939	1.286817	0.215524	1.182740
max	2017.000000	1.600653	0.707979	1.868140	0.380115	1.649386

In [1267]: df75.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1267]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)
4026	New Zealand	2012	1.600653
4027	New Zealand	2013	1.600513
4025	New Zealand	2011	1.600467
4028	New Zealand	2014	1.600005
4024	New Zealand	2010	1.599629

In [1268]: df75.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1268]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)
327	Australia	2009	0.707979
328	Australia	2010	0.707896
326	Australia	2008	0.707868
329	Australia	2011	0.707829
325	Australia	2007	0.707787

In [1269]: df75.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1269]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)
783	Brazil	2017	1.868140
782	Brazil	2016	1.867705
761	Brazil	1995	1.867694
760	Brazil	1994	1.867641
759	Brazil	1993	1.867568

In [1270]: df75.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1270]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)
323	Australia	2005	0.380115
324	Australia	2006	0.380050
325	Australia	2007	0.379995
295	Australasia	2005	0.379892
296	Australasia	2006	0.379779

In [1271]: df75.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1271]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)
4004	New Zealand	1990	1.649386
4005	New Zealand	1991	1.646403
4006	New Zealand	1992	1.643103
4029	New Zealand	2015	1.640746
4028	New Zealand	2014	1.640583

In [1272]: df75.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1272]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)
4027	New Zealand	2013	1.206597
4026	New Zealand	2012	1.206549
4028	New Zealand	2014	1.206502
4025	New Zealand	2011	1.206429
4029	New Zealand	2015	1.206423

In [1273]: df75.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1273]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
4255	Northern Ireland	2017	1.268557
4254	Northern Ireland	2016	1.267900
4253	Northern Ireland	2015	1.267724
4244	Northern Ireland	2006	1.267434
4245	Northern Ireland	2007	1.267350

In [1274]: df75.sort_values(by=v8, ascending=False)[['Entity', 'Year', v7]].head()

Out[1274]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
4228	Northern Ireland	1990	1.253463
4229	Northern Ireland	1991	1.253602
4230	Northern Ireland	1992	1.254173
4231	Northern Ireland	1993	1.255559
4232	Northern Ireland	1994	1.256938

In [1275]: df75.sort_values(by=v9, ascending=False)[['Entity', 'Year', v7]].head()

Out[1275]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
4004	New Zealand	1990	1.193503
4005	New Zealand	1991	1.194842
4006	New Zealand	1992	1.196424
4022	New Zealand	2008	1.222675
4021	New Zealand	2007	1.223667

In [1276]: df75.sort_values(by=v10, ascending=False)[['Entity', 'Year', v7]].head()

Out[1276]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
4031	New Zealand	2017	1.213181
4030	New Zealand	2016	1.214081
4029	New Zealand	2015	1.215885
4028	New Zealand	2014	1.217593
4027	New Zealand	2013	1.219163

In [1277]: df75.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1277]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)
1271	China	2001	0.371677
1270	China	2000	0.371770
1272	China	2002	0.371993
1269	China	1999	0.372192
1691	East Asia	2001	0.372425

In [1278]: df75.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1278]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)
1260	China	1990	0.121939
1680	East Asia	1990	0.122399
1261	China	1991	0.123055
1681	East Asia	1991	0.123499
1262	China	1992	0.124208

In [1279]: df75.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1279]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)
1260	China	1990	0.398073
1261	China	1991	0.398590
1680	East Asia	1990	0.398752
1267	China	1997	0.398869
1266	China	1996	0.398911

In [1280]: df75.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1280]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)
1263	China	1993	0.056045
1683	East Asia	1993	0.056452
1264	China	1994	0.056548
1262	China	1992	0.056623
1684	East Asia	1994	0.056928

In [1281]: df75.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1281]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)
1260	China	1990	0.392653
1261	China	1991	0.392991
1262	China	1992	0.393331
1680	East Asia	1990	0.393630
1263	China	1993	0.393664

In [1282]: df75.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[1282]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)
1260	China	1990	0.314535
1261	China	1991	0.314987
1680	East Asia	1990	0.315198
1262	China	1992	0.315466
1681	East Asia	1991	0.315645

In [1283]: df75.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[1283]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
1275	China	2005	0.484986
1276	China	2006	0.485309
1274	China	2004	0.485432
1695	East Asia	2005	0.485673
1273	China	2003	0.485823

```
In [1284]: df75.sort_values(by=v8, ascending=True)[['Entity', 'Year', v7]].head()
```

Out[1284]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
5655	Taiwan	2017	0.519068
5654	Taiwan	2016	0.518670
5653	Taiwan	2015	0.518270
5652	Taiwan	2014	0.517760
4172	North Korea	1990	0.490772

```
In [1285]: df75.sort_values(by=v9, ascending=True)[['Entity', 'Year', v7]].head()
```

Out[1285]:

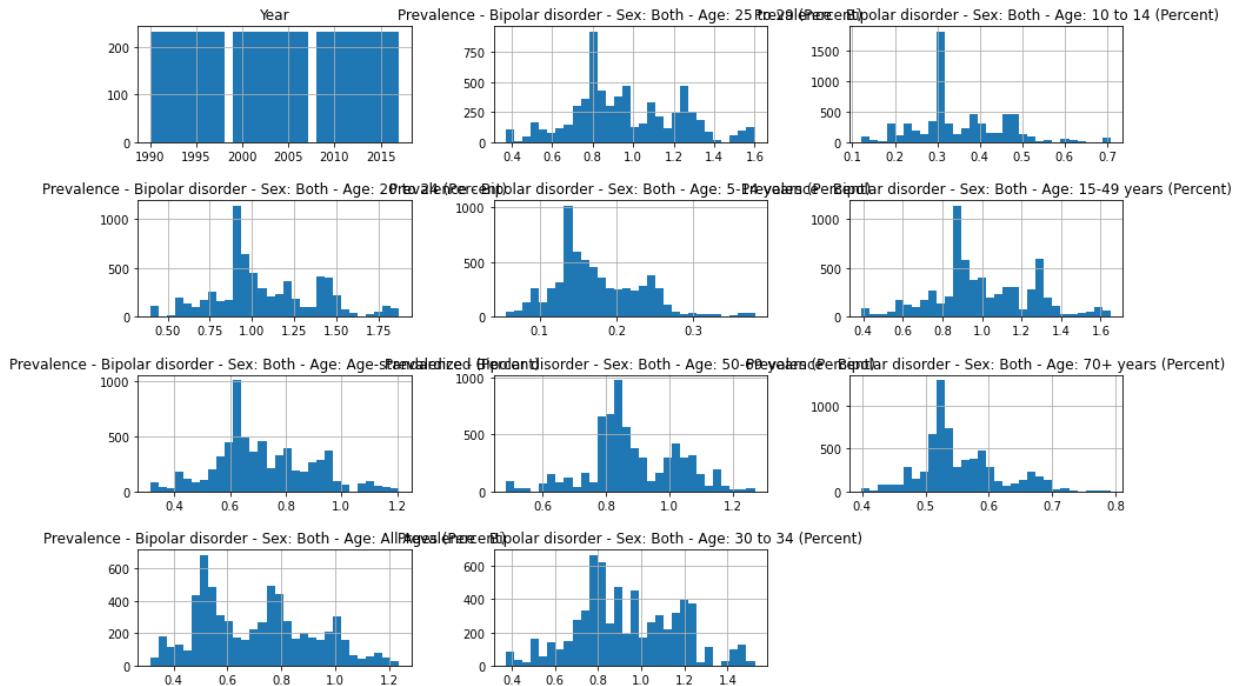
	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
1260	China	1990	0.488357
1680	East Asia	1990	0.488936
1261	China	1991	0.488650
1681	East Asia	1991	0.489226
1262	China	1992	0.488883

```
In [1286]: df75.sort_values(by=v10, ascending=True)[['Entity', 'Year', v7]].head()
```

Out[1286]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)
1275	China	2005	0.484986
1276	China	2006	0.485309
1274	China	2004	0.485432
1273	China	2003	0.485823
1277	China	2007	0.486146

```
In [1287]: df75.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1288]: df75.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1288]: Entity
China                               0.376960
East Asia                            0.377573
North Korea                           0.383630
Taiwan                                0.410322
Southeast Asia, East Asia, and Oceania 0.461096
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1289]: df75.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1289]: Entity
China                               0.127904
East Asia                            0.128266
North Korea                           0.132474
Taiwan                                0.145340
Southeast Asia, East Asia, and Oceania 0.165798
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1290]: df75.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1290]: Entity
China                               0.405846
East Asia                            0.406436
North Korea                           0.409997
Taiwan                                0.439894
Southeast Asia, East Asia, and Oceania 0.505603
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1291]: df75.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1291]: Entity
China                               0.065726
East Asia                            0.065892
North Korea                           0.066111
Taiwan                                0.075935
Southeast Asia, East Asia, and Oceania 0.083973
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent), dtype: float64
```

```
In [1292]: df75.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1292]: Entity
China                               0.397942
East Asia                            0.398900
North Korea                           0.415794
Taiwan                                0.439805
Southeast Asia, East Asia, and Oceania 0.487414
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [1293]: df75.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1293]: Entity
China                               0.319829
East Asia                            0.320392
North Korea                           0.325294
Taiwan                                0.350378
Southeast Asia, East Asia, and Oceania 0.379516
Name: Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1294]: df75.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1294]: Entity
China                               0.488404
East Asia                            0.488994
North Korea                           0.492759
Taiwan                                0.517126
Southeast Asia, East Asia, and Oceania 0.555784
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent), dtype: float64
```

```
In [1295]: df75.groupby('Entity')[v8].mean().sort_values().head()
```

```
Out[1295]: Entity
North Korea                          0.409309
Taiwan                                0.416829
East Asia                             0.434473
China                                 0.435284
United States                         0.437843
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [1296]: df75.groupby('Entity')[v9].mean().sort_values().head()
```

```
Out[1296]: Entity
North Korea                          0.344253
China                                 0.348472
East Asia                            0.349027
Papua New Guinea                     0.356175
Solomon Islands                      0.357843
Name: Prevalence - Bipolar disorder - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [1297]: df75.groupby('Entity')[v10].mean().sort_values().head()
```

```
Out[1297]: Entity
China                               0.376064
East Asia                            0.376989
North Korea                           0.392753
Taiwan                                0.415537
Southeast Asia, East Asia, and Oceania 0.458876
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [1298]: df75.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1298]: Entity
Northern Ireland                     1.552621
England                               1.575448
Tropical Latin America                1.589180
Brazil                                1.590765
New Zealand                           1.597062
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1299]: df75.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1299]: Entity
Tropical Latin America    0.624599
Brazil                      0.625685
New Zealand                 0.691016
Australasia                  0.701304
Australia                     0.703505
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1300]: df75.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1300]: Entity
Australia                   1.800134
Australasia                  1.808872
New Zealand                  1.852994
Tropical Latin America     1.863623
Brazil                        1.865992
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1301]: df75.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1301]: Entity
Tropical Latin America    0.324759
Brazil                      0.325859
New Zealand                 0.365324
Australasia                  0.367911
Australia                     0.368492
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent), dtype: float64
```

```
In [1302]: df75.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1302]: Entity
Northern Ireland            1.577845
Paraguay                     1.585359
Tropical Latin America     1.614241
Brazil                        1.615017
New Zealand                  1.635555
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [1303]: df75.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1303]: Entity
England           1.105841
Northern Ireland  1.142926
Australia         1.144660
Australasia       1.153425
New Zealand        1.199014
Name: Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized
(Percent), dtype: float64
```

```
In [1304]: df75.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1304]: Entity
Sweden            1.158809
England           1.160662
Lebanon            1.170637
New Zealand        1.212849
Northern Ireland   1.263120
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Perce
nt), dtype: float64
```

```
In [1305]: df75.groupby('Entity')[v8].mean().sort_values().tail()
```

```
Out[1305]: Entity
Sweden            0.697515
United Kingdom    0.701065
England           0.706231
Finland            0.709367
Northern Ireland   0.770897
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 70+ years (Percen
t), dtype: float64
```

```
In [1306]: df75.groupby('Entity')[v9].mean().sort_values().tail()
```

```
/opt/anaconda3/lib/python3.8/site-packages/IPython/core/displayhook.py:27
5: UserWarning: Output cache limit (currently 1000 entries) hit.
Flushing oldest 200 entries.
warn('Output cache limit (currently {sz} entries) hit.\n'
```

```
Out[1306]: Entity
Brazil             1.147365
Australia          1.170686
Australasia        1.179228
Northern Ireland   1.182487
New Zealand         1.222516
Name: Prevalence - Bipolar disorder - Sex: Both - Age: All Ages (Percen
t), dtype: float64
```

```
In [1307]: df75.groupby('Entity')[v10].mean().sort_values().tail()
```

```
Out[1307]: Entity
Tropical Latin America    1.455406
Brazil                      1.456287
England                     1.460669
Northern Ireland            1.480028
New Zealand                 1.517098
Name: Prevalence - Bipolar disorder - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

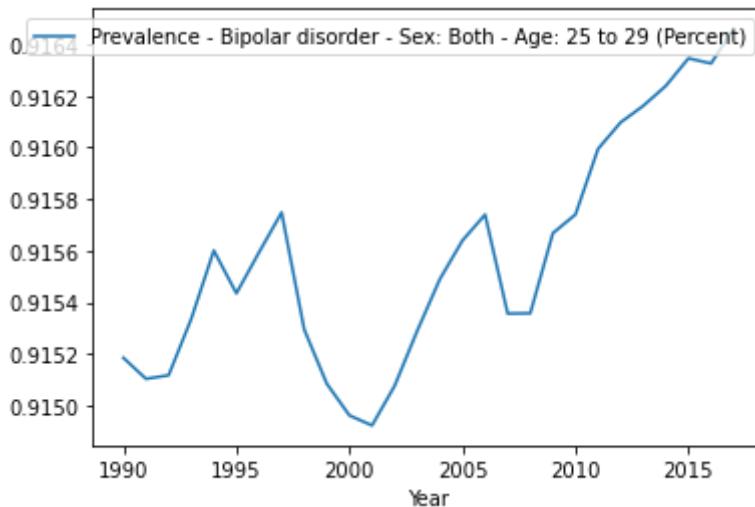
```
In [1308]: df75_mean = df75.groupby('Year').mean()
df75_mean.head()
```

```
Out[1308]:
```

	Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age- standardized (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)	Prevalence - Bi disor Sex: B Age !
Year								
1990	0.957667	0.347933	1.071368	0.168649	1.003945	0.715392	0.880793	0.55
1991	0.957732	0.348338	1.071622	0.169184	1.003671	0.715740	0.880459	0.55
1992	0.957824	0.348723	1.071906	0.169823	1.003421	0.716091	0.880149	0.55
1993	0.957971	0.349087	1.072230	0.170443	1.003208	0.716430	0.879822	0.55
1994	0.958144	0.349405	1.072594	0.170869	1.003100	0.716755	0.879539	0.55

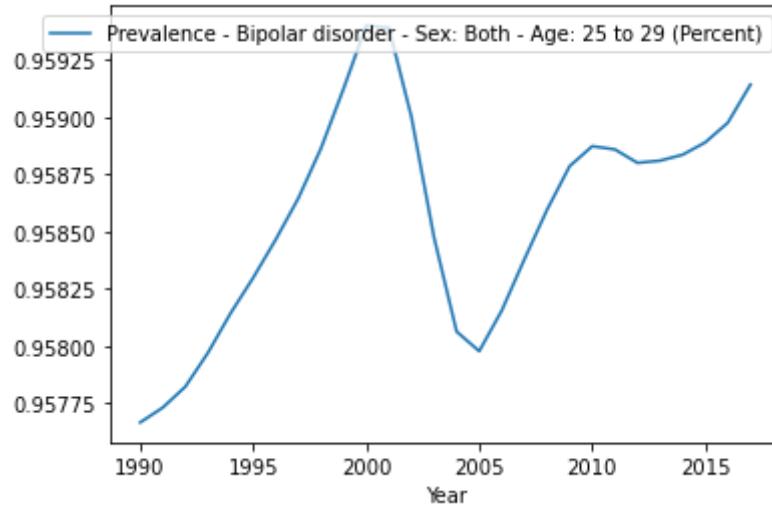
```
In [1309]: df75.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1309]: <AxesSubplot:xlabel='Year'>
```



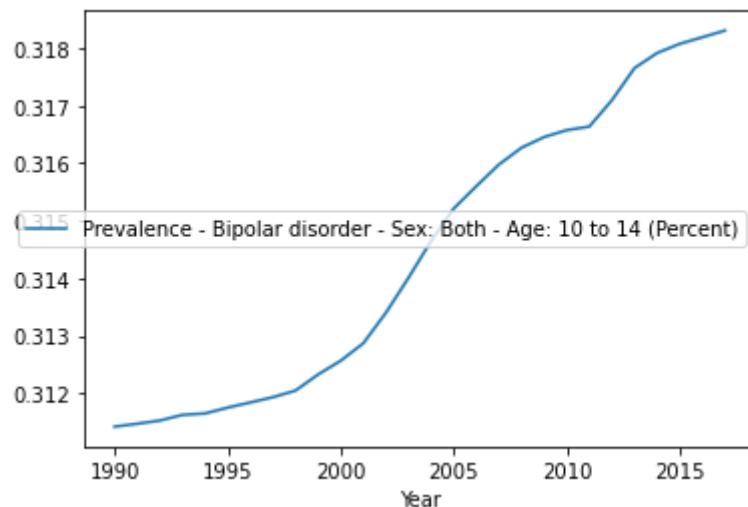
```
In [1310]: df75.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1310]: <AxesSubplot:xlabel='Year'>
```



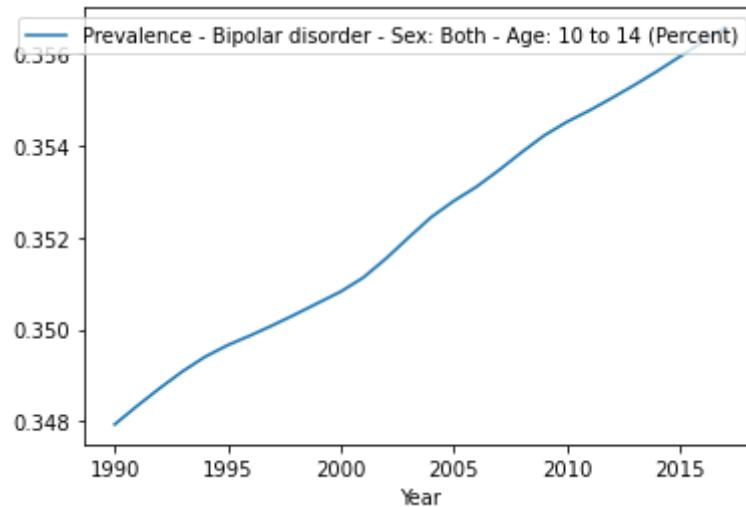
```
In [1311]: df75.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1311]: <AxesSubplot:xlabel='Year'>
```



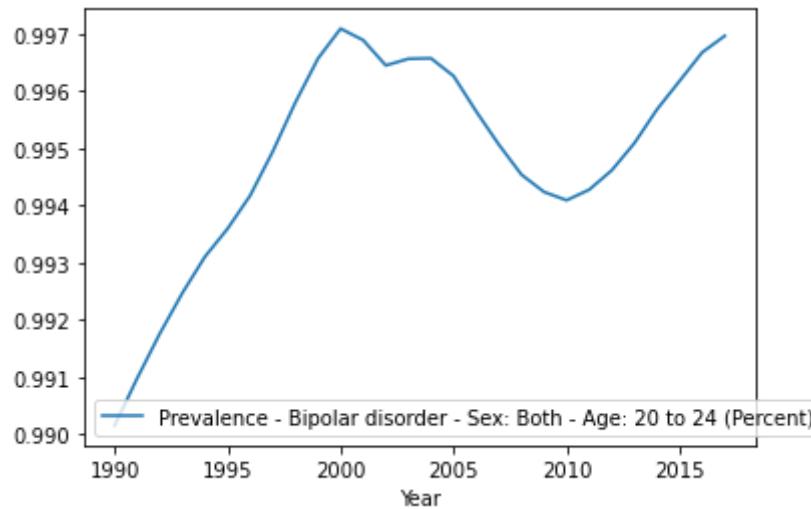
```
In [1312]: df75.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1312]: <AxesSubplot:xlabel='Year'>
```



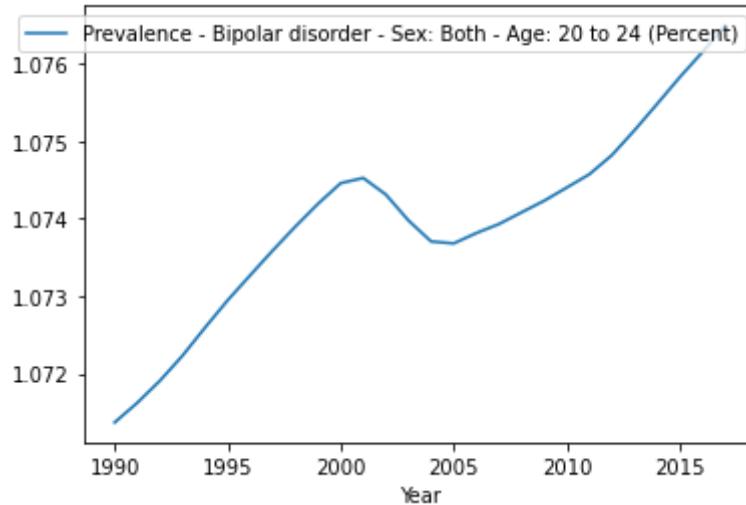
```
In [1313]: df75.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1313]: <AxesSubplot:xlabel='Year'>
```



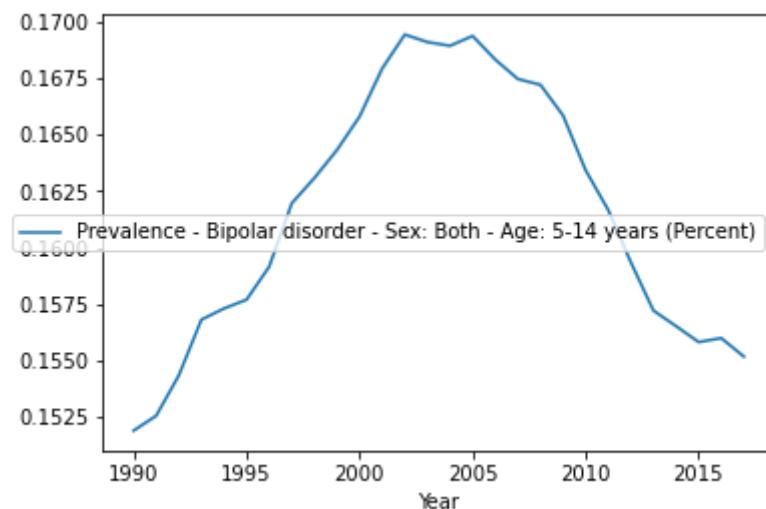
```
In [1314]: df75.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1314]: <AxesSubplot:xlabel='Year'>
```



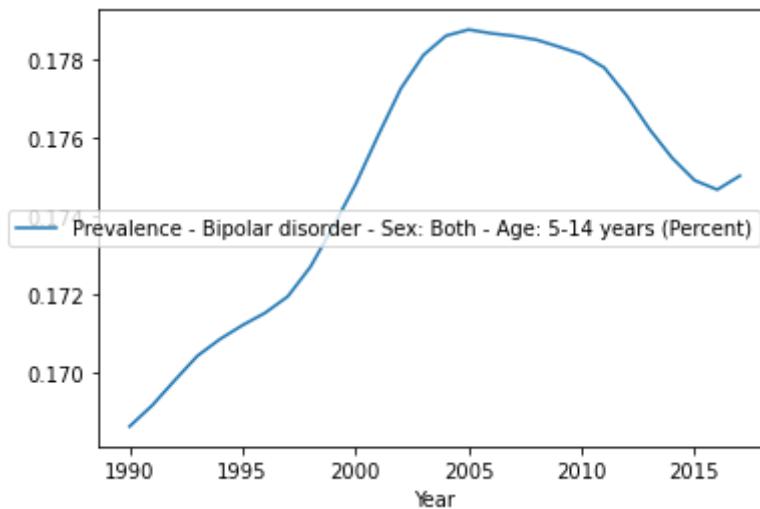
```
In [1315]: df75.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1315]: <AxesSubplot:xlabel='Year'>
```



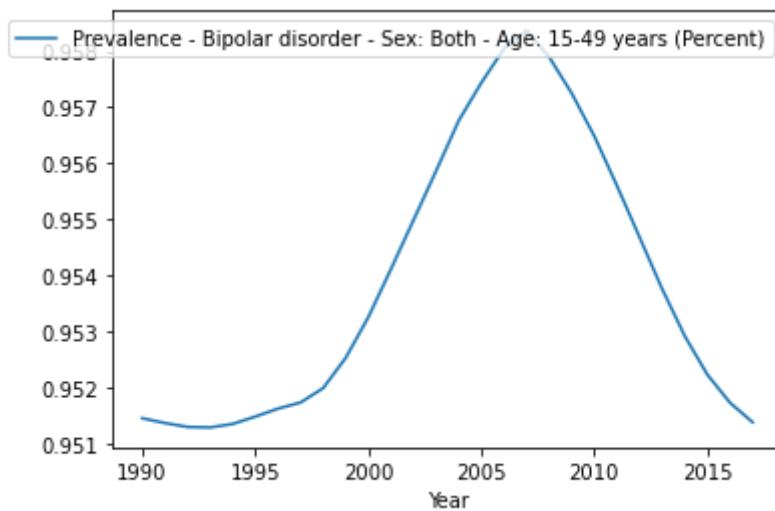
```
In [1316]: df75.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1316]: <AxesSubplot:xlabel='Year'>
```



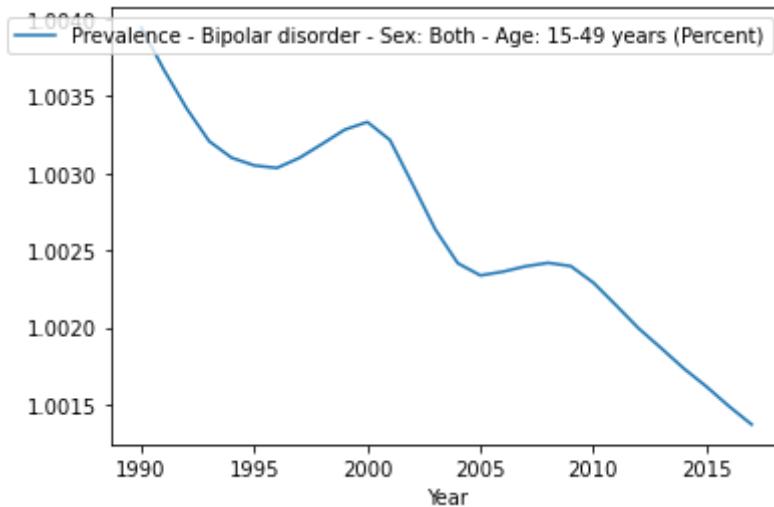
```
In [1317]: df75.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1317]: <AxesSubplot:xlabel='Year'>
```



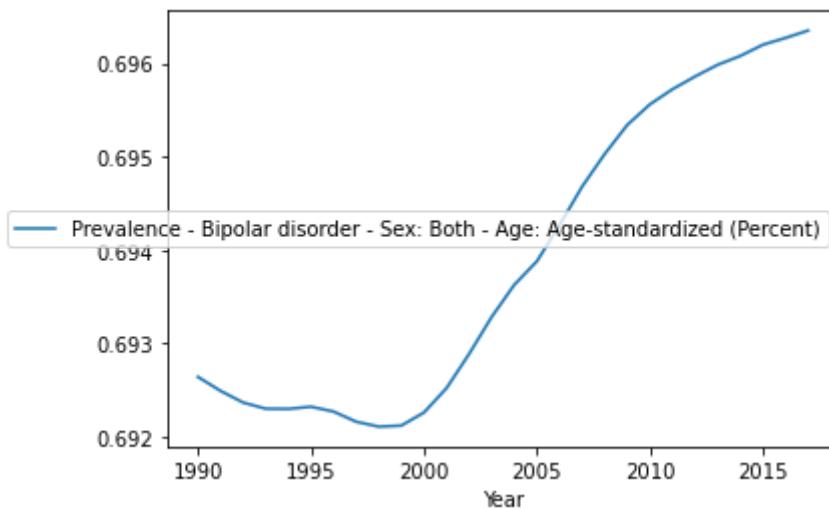
```
In [1318]: df75.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1318]: <AxesSubplot:xlabel='Year'>
```



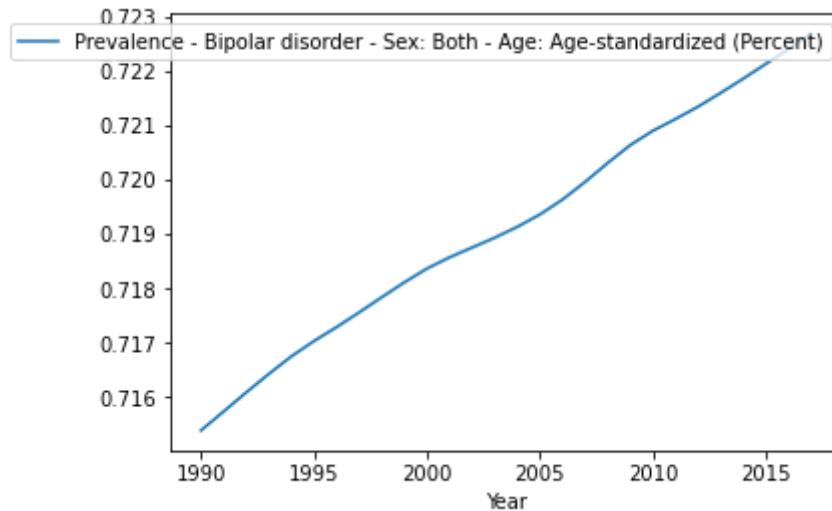
```
In [1319]: df75.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1319]: <AxesSubplot:xlabel='Year'>
```



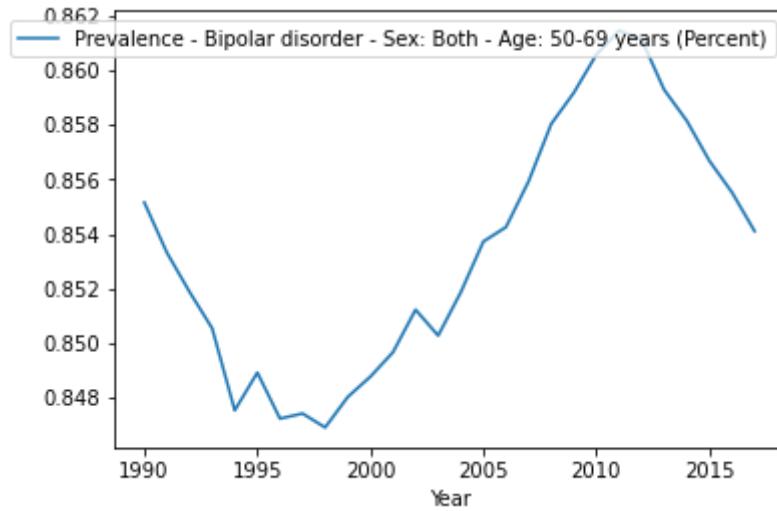
```
In [1320]: df75.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1320]: <AxesSubplot:xlabel='Year'>
```



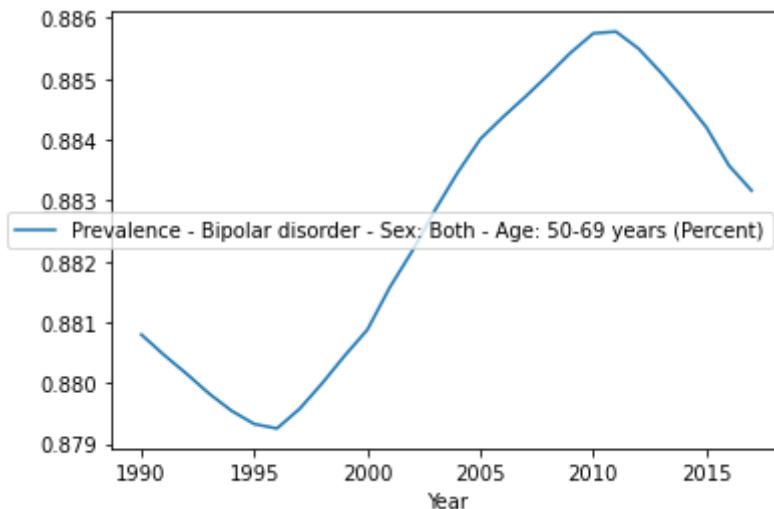
```
In [1321]: df75.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1321]: <AxesSubplot:xlabel='Year'>
```



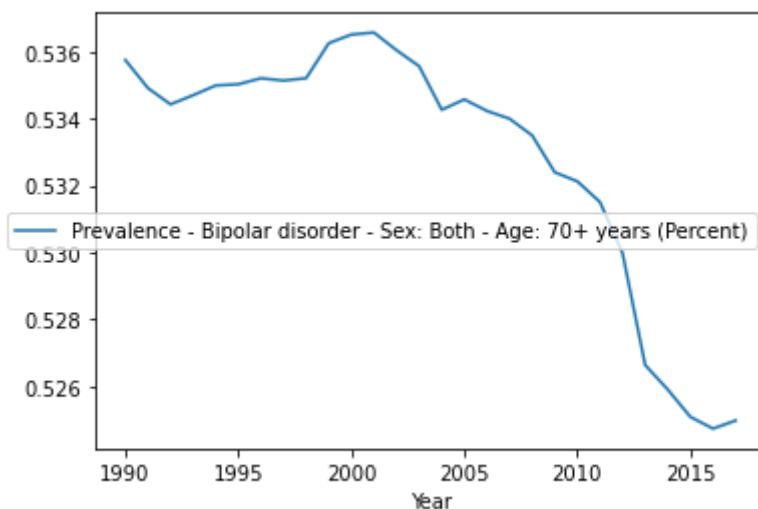
```
In [1322]: df75.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1322]: <AxesSubplot:xlabel='Year'>
```



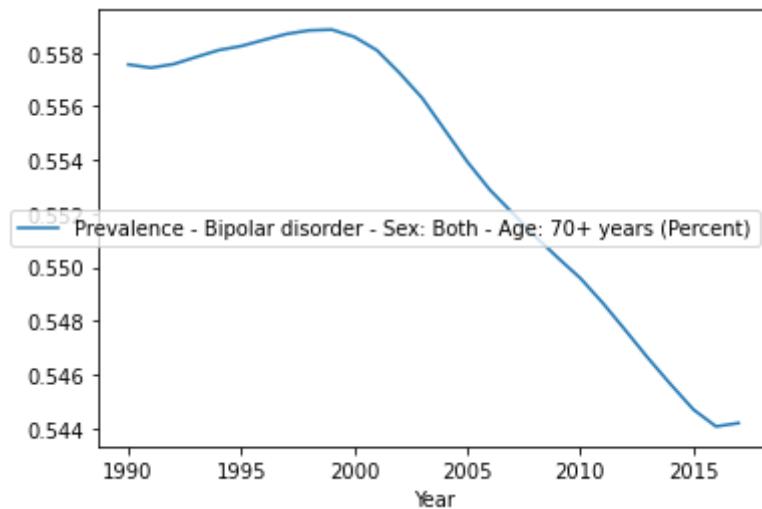
```
In [1323]: df75.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[1323]: <AxesSubplot:xlabel='Year'>
```



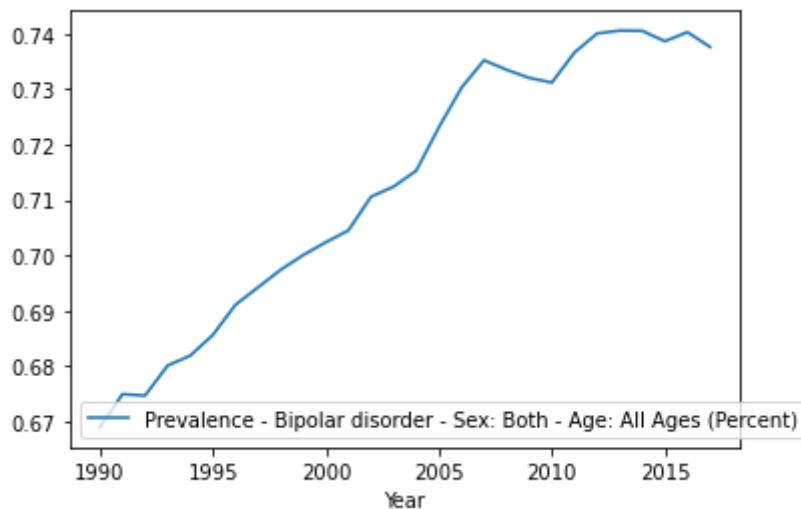
```
In [1324]: df75.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[1324]: <AxesSubplot:xlabel='Year'>
```



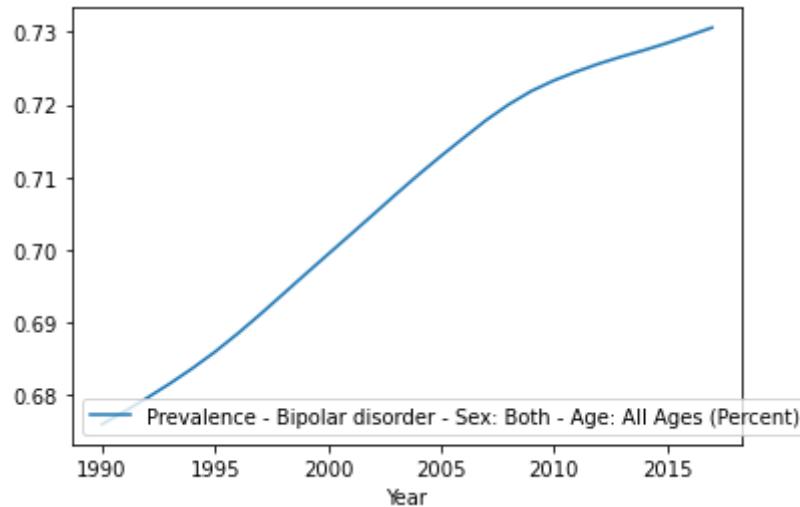
```
In [1325]: df75.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[1325]: <AxesSubplot:xlabel='Year'>
```



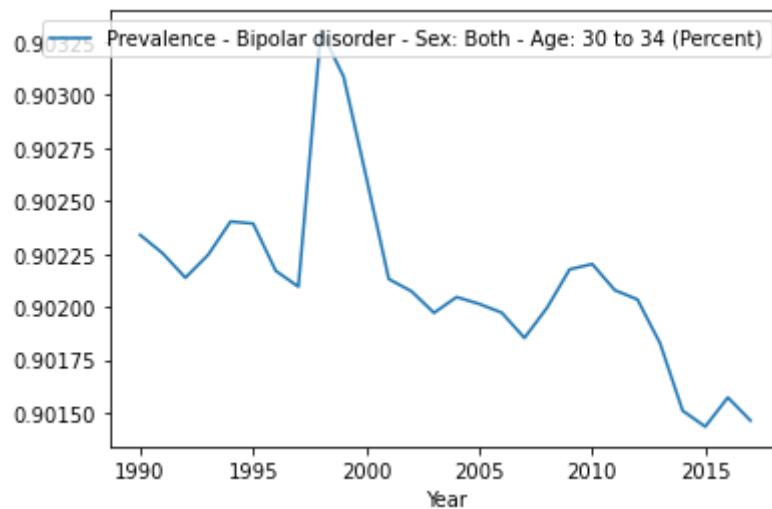
```
In [1326]: df75.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[1326]: <AxesSubplot:xlabel='Year'>
```



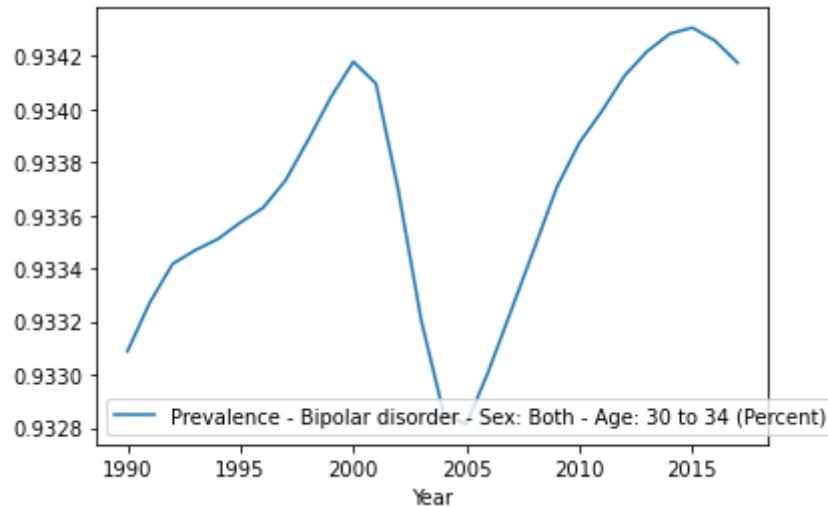
```
In [1327]: df75.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[1327]: <AxesSubplot:xlabel='Year'>
```



```
In [1328]: df75.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[1328]: <AxesSubplot:xlabel='Year'>
```



Schizophrenia

In [1329]: df14.info()
df14.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 9 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both -
Age: 70+ years (Rate)          6468 non-null   float64
4   DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both -
Age: 50-69 years (Rate)        6468 non-null   float64
5   DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both -
Age: Age-standardized (Rate)  6468 non-null   float64
6   DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both -
Age: 15-49 years (Rate)       6468 non-null   float64
7   DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both -
Age: 5-14 years (Rate)        6468 non-null   float64
8   DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both -
Age: All Ages (Rate)          6468 non-null   float64
dtypes: float64(6), int64(1), object(2)
memory usage: 454.9+ KB
```

Out[1329]:

	Entity	Code	Year	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age- standardized (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)	DAI (Disabil ity- Adjusted I Year Schizophre - Sex: Bot Age: 5 years (Ra
0	Afghanistan	AFG	1990	71.174164	156.973110	95.808405	109.809896	1.519%
1	Afghanistan	AFG	1991	71.740384	156.997946	95.798815	107.444655	1.469%
2	Afghanistan	AFG	1992	71.764364	156.807269	95.701359	104.219959	1.428%
3	Afghanistan	AFG	1993	71.960985	156.425287	95.580215	103.675981	1.389%
4	Afghanistan	AFG	1994	72.130356	155.894898	95.430753	105.670436	1.349%

Checking for missing values:

```
In [1330]: missing = pd.concat([df14.isnull().sum(), 100 * df14.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1330]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 5-14 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Rate)		0	0.000000
Code	980	15.151515	

```
In [1331]: v1='DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)'
v2='DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)'
v3='DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)'
v4='DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 5-14 years (Rate)'
v5='DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Rate)'
v6='DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age-standardized (Rate)'
```

In [1332]: df14.describe()

Out[1332]:

Year	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age- standardized (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 5-14 years (Rate)	Ac
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	97.381431	211.995128	128.484643	177.732247	1.660489
std	8.078372	19.499990	39.782122	24.675580	44.573811	0.372483
min	1990.000000	66.141911	152.877148	91.305279	103.675981	1.117175
25%	1996.750000	82.465462	182.809924	111.493224	142.679432	1.455660
50%	2003.500000	92.944338	201.652925	122.472151	170.022954	1.532795
75%	2010.250000	110.838721	233.158362	145.055396	207.777686	1.669258
max	2017.000000	171.830250	378.850070	217.406148	332.300012	4.081111

In [1333]: df14.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1333]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)	
3980	Netherlands	1994	171.830250
3986	Netherlands	2000	171.195298
3983	Netherlands	1997	171.114208
3982	Netherlands	1996	170.977181
3981	Netherlands	1995	170.967530

In [1334]: df14.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1334]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)
3988	Netherlands 2002	378.850070
3986	Netherlands 2000	378.583223
3989	Netherlands 2003	378.536035
3987	Netherlands 2001	378.340652
3990	Netherlands 2004	378.132674

In [1335]: df14.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1335]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age-standardized (Rate)
3992	Netherlands 2006	217.406148
3990	Netherlands 2004	217.387034
3994	Netherlands 2008	217.300478
3991	Netherlands 2005	217.285495
3993	Netherlands 2007	217.221392

In [1336]: df14.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1336]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)
3990	Netherlands 2004	332.300012
3988	Netherlands 2002	331.768927
3991	Netherlands 2005	331.687532
3989	Netherlands 2003	331.663616
3992	Netherlands 2006	331.499385

In [1337]: df14.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1337]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 5-14 years (Rate)
1271	China	2001	4.081111
1270	China	2000	4.049060
1272	China	2002	4.048921
1691	East Asia	2001	4.027326
1273	China	2003	3.999342

In [1338]: df14.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1338]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Rate)
3992	Netherlands	2006	265.799512
3993	Netherlands	2007	265.565418
3994	Netherlands	2008	265.517908
3991	Netherlands	2005	265.504280
3990	Netherlands	2004	265.500818

In [1339]: df14.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1339]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)
3265	Liberia	2007	66.141911
3268	Liberia	2010	66.204755
3269	Liberia	2011	66.280339
3267	Liberia	2009	66.323741
3259	Liberia	2001	66.374316

In [1340]: df14.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1340]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)
18	Afghanistan	2008	152.877148
21	Afghanistan	2011	153.248286
19	Afghanistan	2009	153.502390
20	Afghanistan	2010	153.585957
17	Afghanistan	2007	153.689181

In [1341]: df14.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1341]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age-standardized (Rate)
1047	Central African Republic	2001	91.305279
1054	Central African Republic	2008	91.355799
1050	Central African Republic	2004	91.358121
1052	Central African Republic	2006	91.360210
1048	Central African Republic	2002	91.386012

In [1342]: df14.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1342]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)
3	Afghanistan	1993	103.675981
2	Afghanistan	1992	104.219959
4	Afghanistan	1994	105.670436
1	Afghanistan	1991	107.444655
5	Afghanistan	1995	108.310107

```
In [1343]: df14.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()
```

Out[1343]:

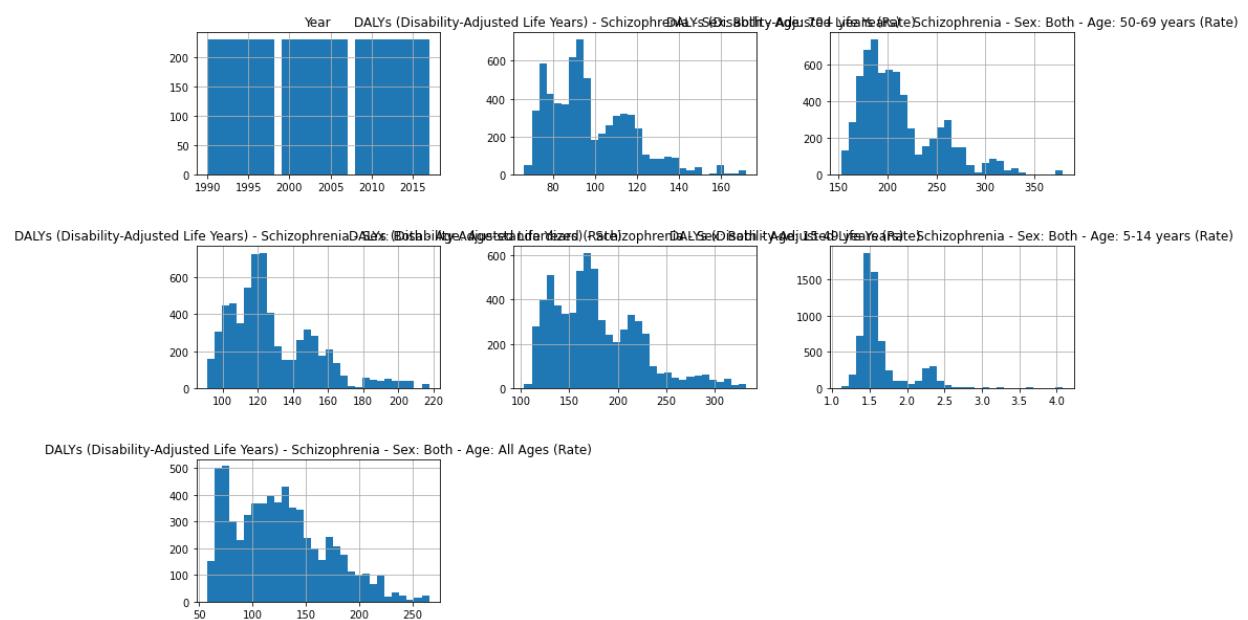
	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 5-14 years (Rate)
10	Afghanistan	2000	1.117175
11	Afghanistan	2001	1.119371
9	Afghanistan	1999	1.133332
5824	Trinidad and Tobago	1990	1.159959
12	Afghanistan	2002	1.161936

```
In [1344]: df14.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()
```

Out[1344]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Rate)
5974	Uganda	2000	57.603409
5973	Uganda	1999	57.748002
5975	Uganda	2001	57.829305
5970	Uganda	1996	57.863162
5971	Uganda	1997	57.878834

```
In [1345]: df14.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1346]: df14.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1346]: Entity
Liberia      67.709531
Mozambique   70.274928
Malawi       70.347577
Sierra Leone 71.714851
Lesotho       71.756491
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 70+ years (Rate), dtype: float64
```

```
In [1347]: df14.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1347]: Entity
Afghanistan    156.041407
Central African Republic 156.170048
Lesotho        160.283865
Mozambique     160.347857
Malawi         160.727924
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 50-69 years (Rate), dtype: float64
```

```
In [1348]: df14.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1348]: Entity
Central African Republic 91.752290
Mozambique      93.568644
Malawi          94.289783
Lesotho          95.159981
Burundi          95.591971
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: Age-standardized (Rate), dtype: float64
```

```
In [1349]: df14.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1349]: Entity
Uganda          113.129665
Malawi          114.666158
Central African Republic 115.212544
Mozambique      115.341636
Zimbabwe        116.517121
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 15-49 years (Rate), dtype: float64
```

```
In [1350]: df14.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1350]: Entity
Trinidad and Tobago 1.269129
Tajikistan        1.317022
Suriname          1.326598
Afghanistan       1.343379
Kyrgyzstan        1.344756
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 5-14 years (Rate), dtype: float64
```

```
In [1351]: df14.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1351]: Entity
Uganda      59.929518
Afghanistan 64.118045
Chad        64.684791
Mozambique   64.703387
Malawi       64.977025
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: All Ages (Rate), dtype: float64
```

```
In [1352]: df14.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1352]: Entity
Greenland          147.769784
Southeast Asia, East Asia, and Oceania 148.186774
East Asia           159.143835
China               160.397657
Netherlands         169.025478
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 70+ years (Rate), dtype: float64
```

```
In [1353]: df14.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1353]: Entity
United States     314.105907
Greenland          314.809277
Australasia        328.585084
Australia          331.837659
Netherlands         373.267130
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 50-69 years (Rate), dtype: float64
```

```
In [1354]: df14.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1354]: Entity
East Asia          199.689006
China              201.415764
Australasia        204.668846
Australia          207.130001
Netherlands         215.900248
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: Age-standardized (Rate), dtype: float64
```

```
In [1355]: df14.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1355]: Entity
East Asia          298.017944
China              300.606480
Australasia        307.741742
Australia          312.136188
Netherlands         323.749932
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 15-49 years (Rate), dtype: float64
```

```
In [1356]: df14.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1356]: Entity
Taiwan                               2.548216
North Korea                          2.652255
Southeast Asia, East Asia, and Oceania 3.138226
East Asia                            3.636607
China                                3.674941
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: 5-14 years (Rate), dtype: float64
```

```
In [1357]: df14.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1357]: Entity
East Asia      229.829584
China          231.934044
Australasia   232.699331
Australia     236.317813
Netherlands   261.419344
Name: DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both
- Age: All Ages (Rate), dtype: float64
```

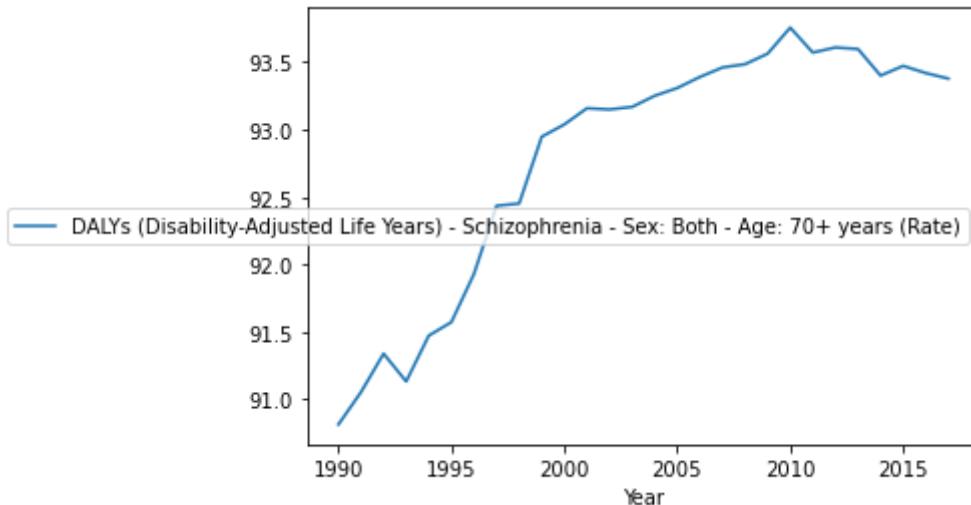
```
In [1358]: df14_mean = df14.groupby('Year').mean()
df14_mean.head()
```

```
Out[1358]:
```

DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age- standardized (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 5-14 years (Rate)	DALYs (Disability- Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Rate)
Year					
1990	97.121388	210.155718	127.808286	170.717037	1.612222
1991	97.071727	210.033312	127.797549	171.403852	1.613013
1992	97.073410	209.896203	127.765908	172.062563	1.615144
1993	97.126627	209.770861	127.770828	172.785553	1.618313
1994	97.209684	209.693755	127.752762	173.380104	1.621362

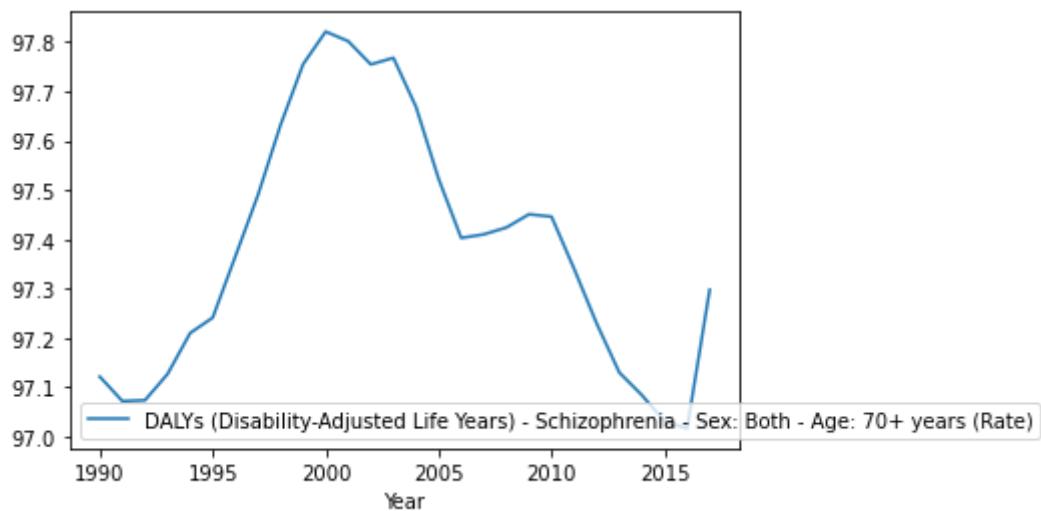
```
In [1359]: df14.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1359]: <AxesSubplot:xlabel='Year'>
```



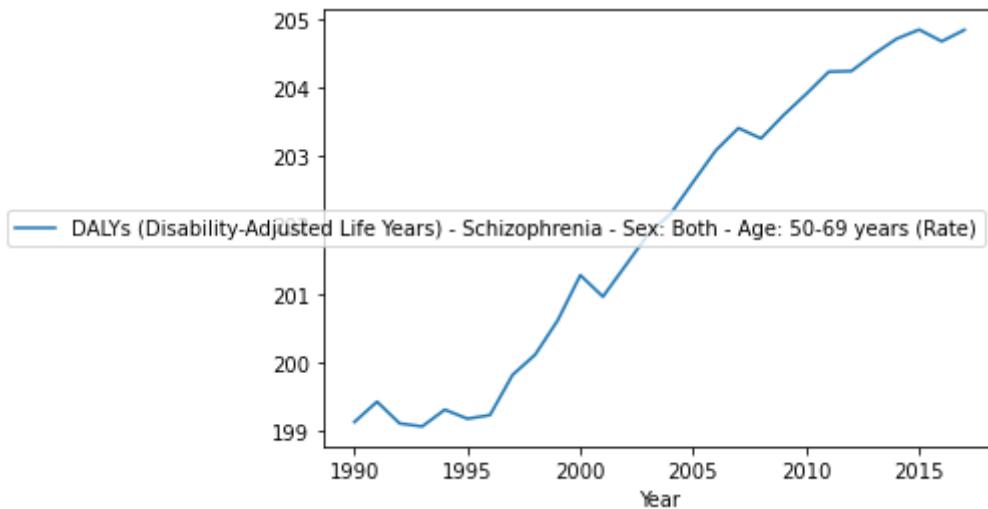
```
In [1360]: df14.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1360]: <AxesSubplot:xlabel='Year'>
```



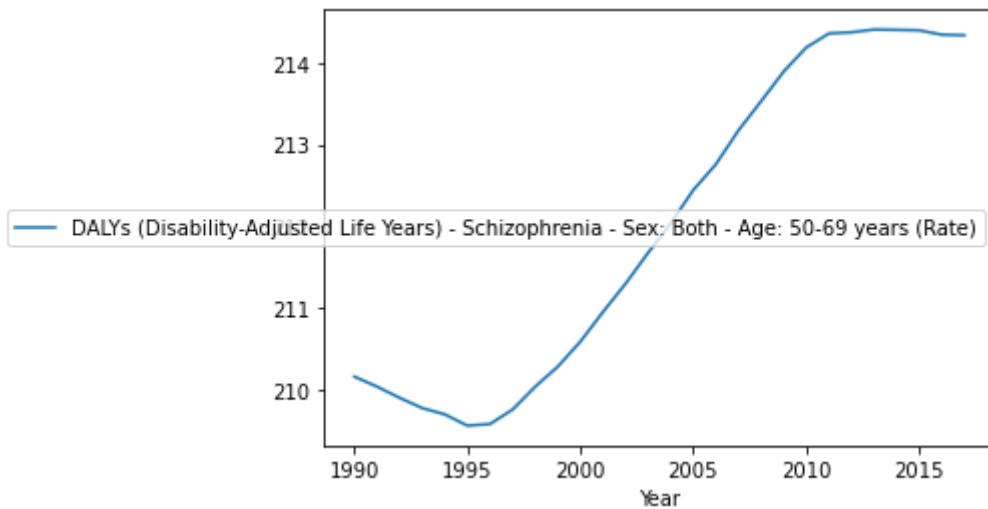
```
In [1361]: df14.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1361]: <AxesSubplot:xlabel='Year'>
```



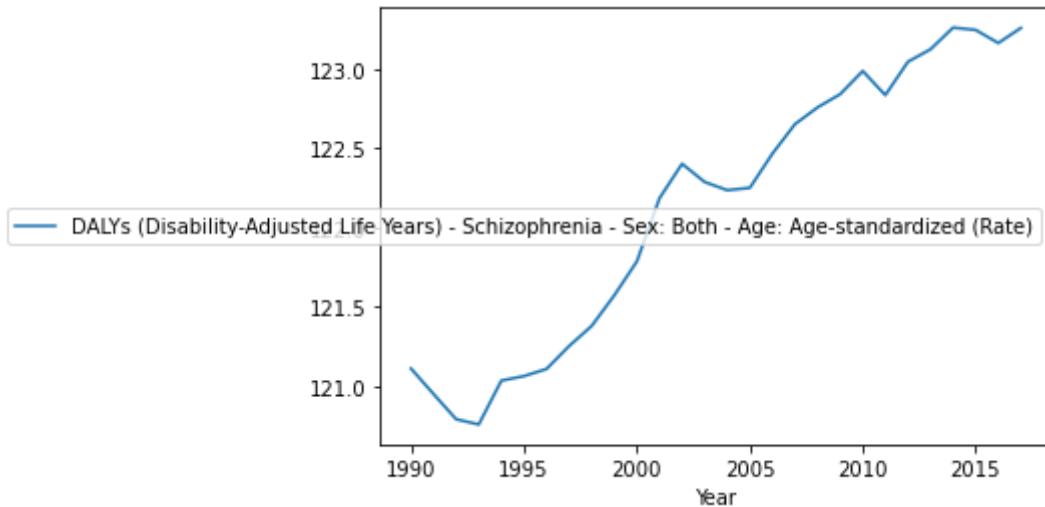
```
In [1362]: df14.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1362]: <AxesSubplot:xlabel='Year'>
```



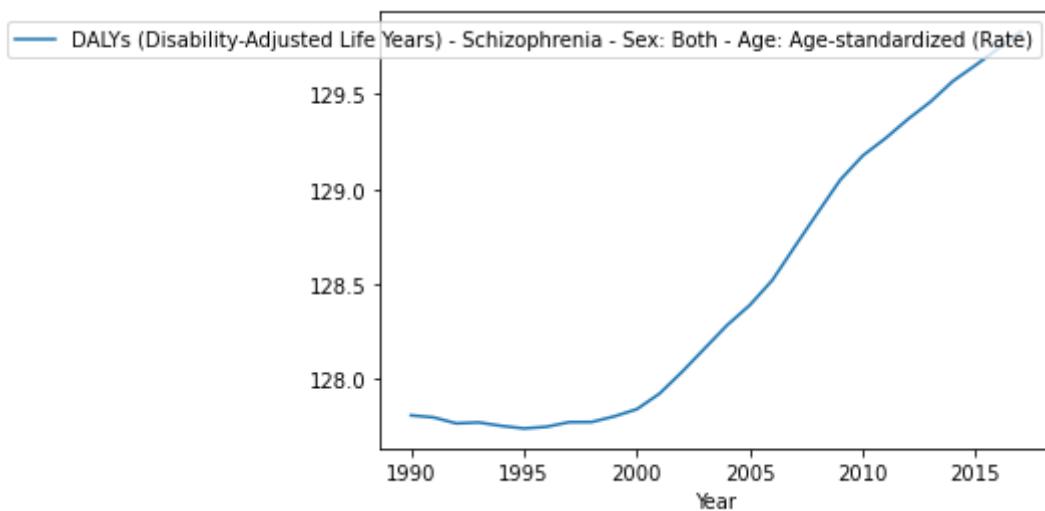
```
In [1363]: df14.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1363]: <AxesSubplot:xlabel='Year'>
```



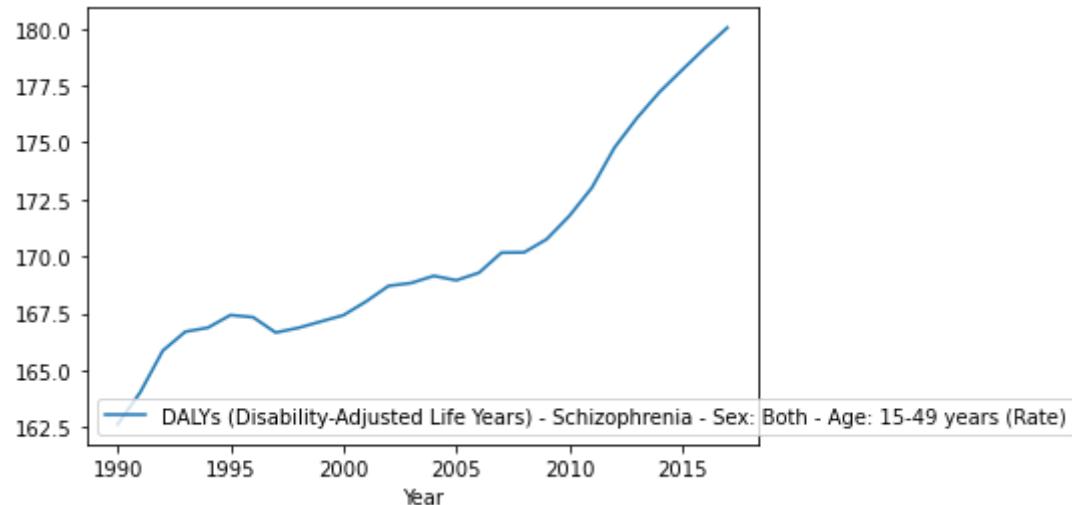
```
In [1364]: df14.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1364]: <AxesSubplot:xlabel='Year'>
```



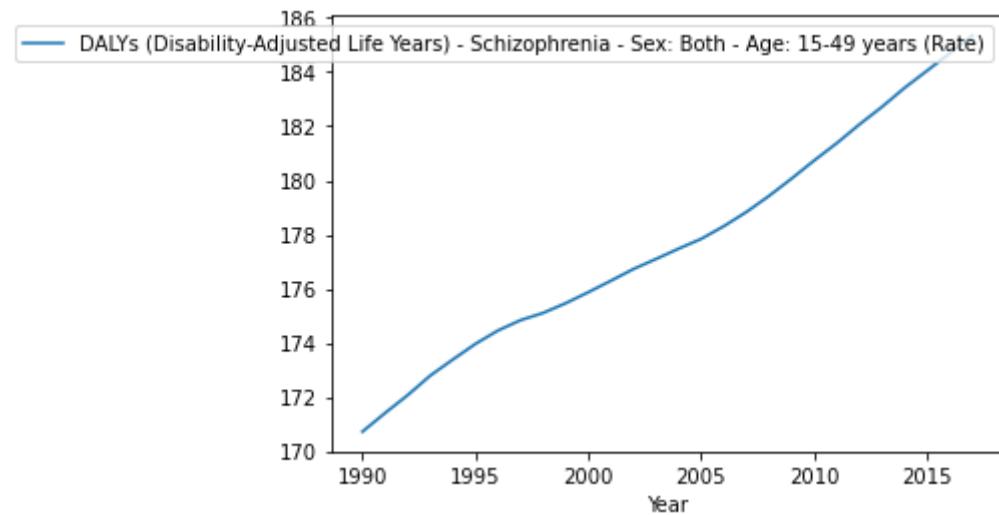
```
In [1365]: df14.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1365]: <AxesSubplot:xlabel='Year'>
```



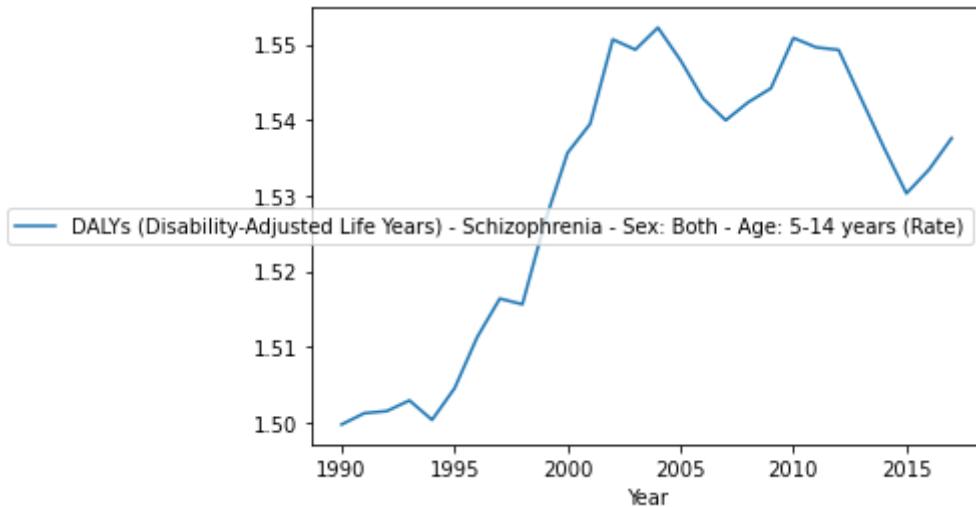
```
In [1366]: df14.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1366]: <AxesSubplot:xlabel='Year'>
```



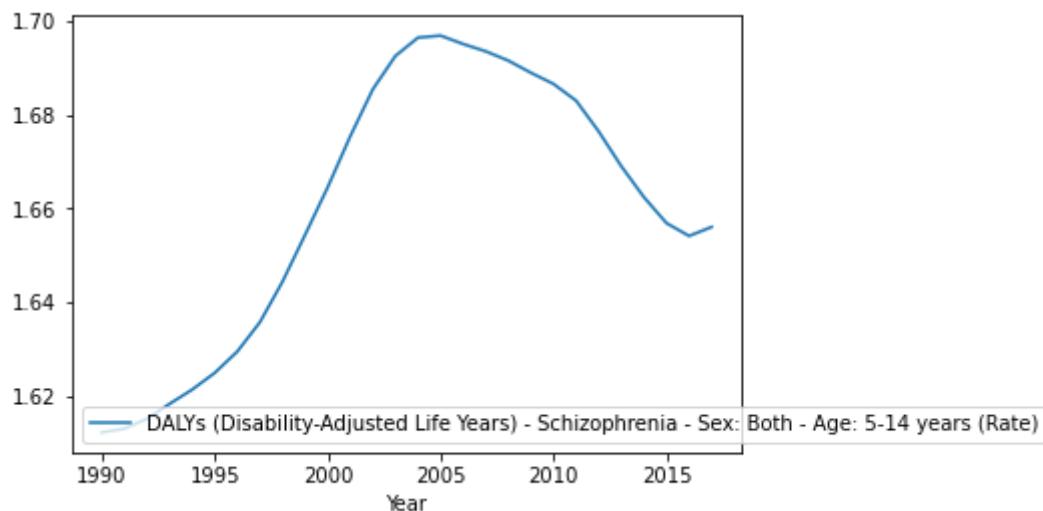
```
In [1367]: df14.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1367]: <AxesSubplot:xlabel='Year'>
```



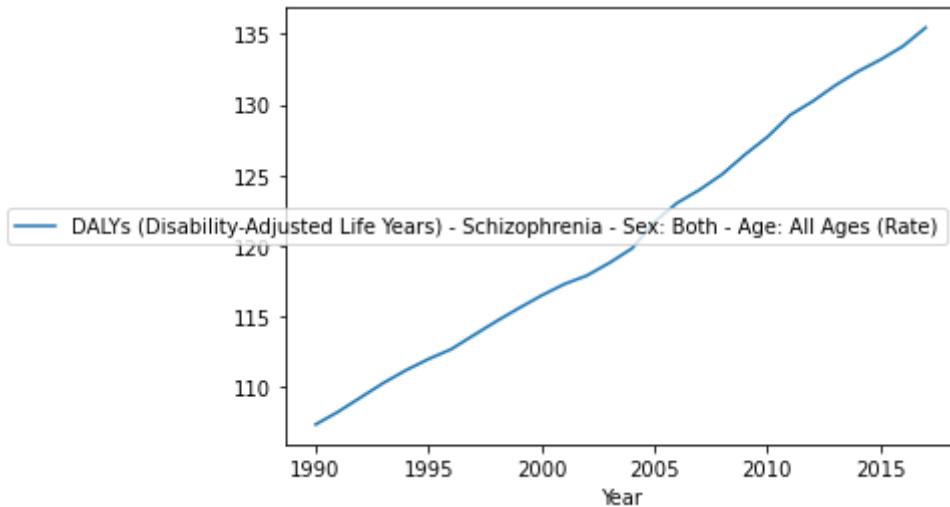
```
In [1368]: df14.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1368]: <AxesSubplot:xlabel='Year'>
```



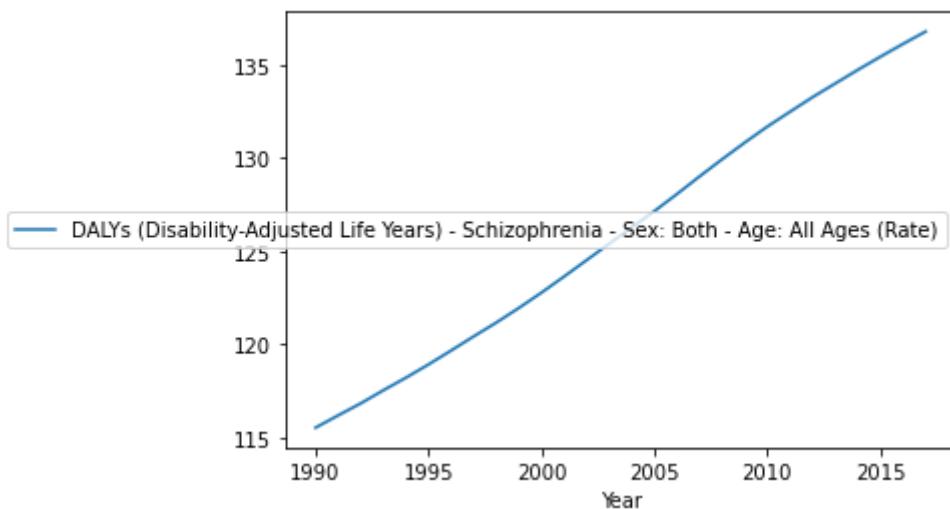
```
In [1369]: df14.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1369]: <AxesSubplot:xlabel='Year'>
```



```
In [1370]: df14.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1370]: <AxesSubplot:xlabel='Year'>
```



In [1371]: df66.info()
df66.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column           Dtype    
on-Null Count  Dtype  
---  --  
0   Entity          object    
468 non-null    object    
1   Code            object    
488 non-null    object    
2   Year            int64    
3   Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number) float64 
468 non-null    float64  
4   Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number) float64 
468 non-null    float64  
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1371]:

	Entity	Code	Year	Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number)	Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number)
0	Afghanistan	AFG	1990	5808.078611	5495.205799
1	Afghanistan	AFG	1991	5932.383469	5643.976535
2	Afghanistan	AFG	1992	6737.397285	6512.851828
3	Afghanistan	AFG	1993	7705.839889	7488.625513
4	Afghanistan	AFG	1994	8076.375017	7838.703381

Checking for missing values:

In [1372]: missing = pd.concat([df66.isnull().sum(), 100 * df66.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1372]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
	Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number)	0	0.000000
	Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number)	0	0.000000
	Code	980	15.151515

In [1373]: v1='Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number)'
v2='Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number)'

In [1374]: df66.describe()

Out[1374]:

	Year	Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number)	Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number)
count	6468.000000	6.468000e+03	6.468000e+03
mean	2003.500000	1.667691e+05	1.596019e+05
std	8.078372	6.861867e+05	6.533298e+05
min	1990.000000	3.385062e+01	3.009595e+01
25%	1996.750000	2.196877e+03	2.052597e+03
50%	2003.500000	7.546623e+03	7.656414e+03
75%	2010.250000	4.113871e+04	4.123744e+04
max	2017.000000	1.009944e+07	9.677443e+06

In [1375]: df66.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1375]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number)
6383	World	2017	1.009944e+07
6382	World	2016	9.978765e+06
6381	World	2015	9.850307e+06
6380	World	2014	9.716350e+06
6379	World	2013	9.577464e+06

In [1376]: df66.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1376]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number)
6383	World	2017	9.677443e+06
6382	World	2016	9.536969e+06
6381	World	2015	9.393973e+06
6380	World	2014	9.248785e+06
6379	World	2013	9.101670e+06

In [1377]: `df66.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()`

Out[1377]:

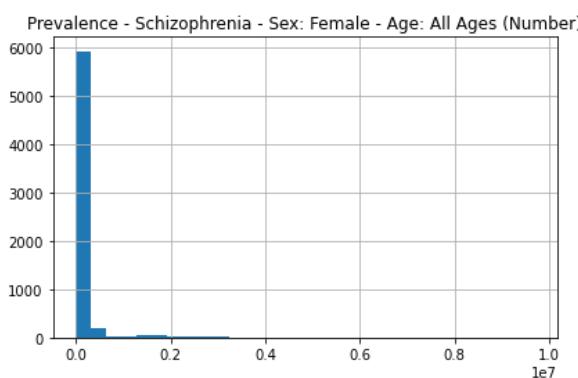
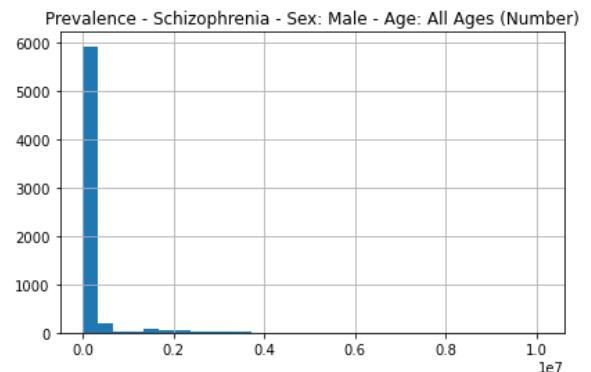
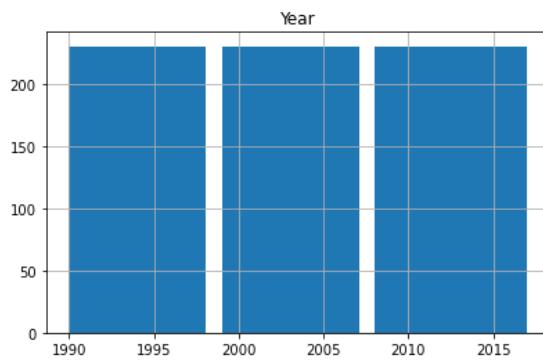
	Entity	Year	Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number)
3584	Marshall Islands	1990	33.850621
3585	Marshall Islands	1991	34.860134
3586	Marshall Islands	1992	35.876365
3587	Marshall Islands	1993	36.901361
3588	Marshall Islands	1994	37.938460

In [1378]: `df66.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()`

Out[1378]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number)
3584	Marshall Islands	1990	30.095951
3585	Marshall Islands	1991	30.949404
3586	Marshall Islands	1992	31.825111
3587	Marshall Islands	1993	32.724015
3588	Marshall Islands	1994	33.645466

In [1379]: `df66.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [1380]: df66.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1380]: Entity
Marshall Islands      47.338837
American Samoa        60.961729
Dominica              71.650578
Kiribati              79.665811
Northern Mariana Islands 81.135625
Name: Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1381]: df66.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1381]: Entity
Marshall Islands      41.967870
American Samoa        56.797384
Dominica              60.941092
Northern Mariana Islands 74.712980
Bermuda               80.912659
Name: Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number), dtype: float64
```

```
In [1382]: df66.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1382]: Entity
China                  2.357423e+06
Middle SDI             2.420358e+06
East Asia              2.458493e+06
Southeast Asia, East Asia, and Oceania 3.086098e+06
World                  8.158210e+06
Name: Prevalence - Schizophrenia - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1383]: df66.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1383]: Entity
Middle SDI             2.274715e+06
China                  2.356506e+06
East Asia              2.457062e+06
Southeast Asia, East Asia, and Oceania 3.049312e+06
World                  7.731674e+06
Name: Prevalence - Schizophrenia - Sex: Female - Age: All Ages (Number), dtype: float64
```

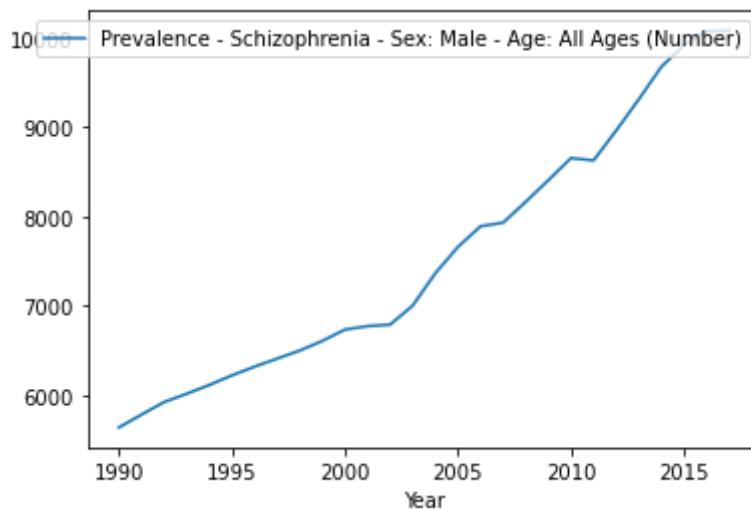
```
In [1384]: df66_mean = df69.groupby('Year').mean()
df66_mean.head()
```

Out[1384]:

Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Alcohol disorders - Sex: Both - Age: Age-standardized (Percent)
Year						
1990	0.209548	0.715392	0.221774	3.957269	0.808283	3.506288
1991	0.209586	0.715740	0.222076	3.960009	0.813466	3.510948
1992	0.209634	0.716091	0.222481	3.962778	0.818692	3.515033
1993	0.209690	0.716430	0.223033	3.965405	0.823780	3.518531
1994	0.209751	0.716755	0.223710	3.967976	0.828389	3.521437

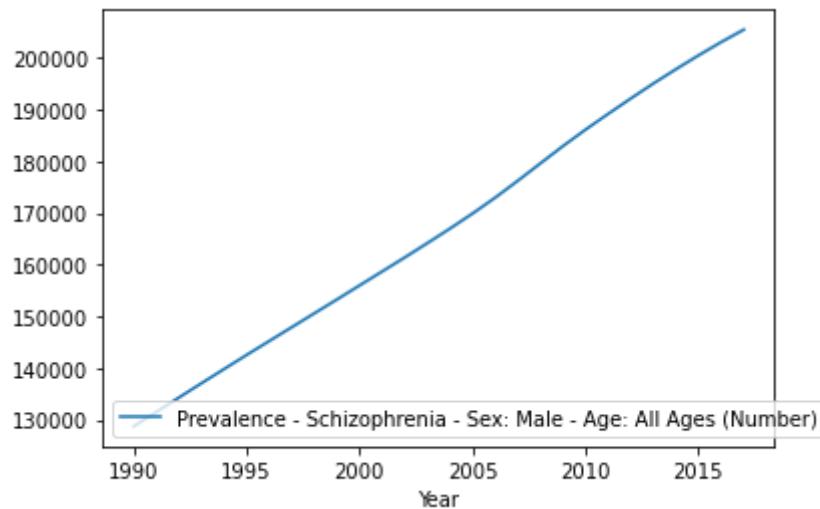
```
In [1385]: df66.groupby('Year')[v1].median().plot(legend=True)
```

Out[1385]: <AxesSubplot: xlabel='Year'>



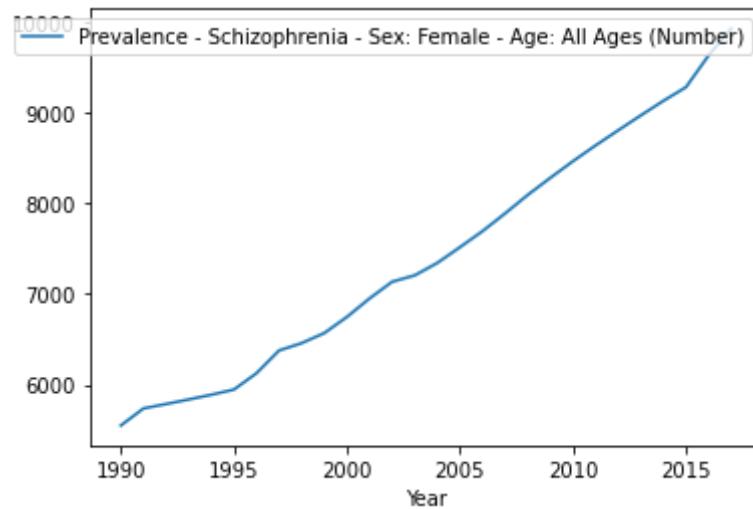
```
In [1386]: df66.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1386]: <AxesSubplot:xlabel='Year'>
```



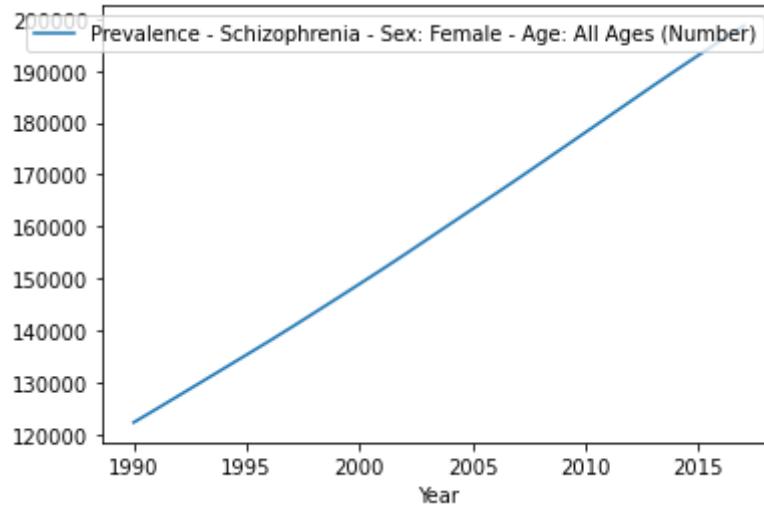
```
In [1387]: df66.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1387]: <AxesSubplot:xlabel='Year'>
```



```
In [1388]: df66.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1388]: <AxesSubplot:xlabel='Year'>
```



In [1389]: df85.info()
df85.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 14 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   float64
3   Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)
6468 non-null   float64
4   Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)
6468 non-null   float64
5   Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)
6468 non-null   float64
6   Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)
6468 non-null   float64
7   Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)
6468 non-null   float64
8   Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)
6468 non-null   float64
9   Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent)
6468 non-null   float64
10  Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent)
6468 non-null   float64
11  Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent)
6468 non-null   float64
12  Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent)
6468 non-null   float64
13  Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
dtypes: float64(11), int64(1), object(2)
memory usage: 707.6+ KB
```

Out[1389]:

	Entity	Code	Year	Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 50+ years (Percent)
0	Afghanistan	AFG	1990	0.036130	0.004323	0.225878	0.002177	0.2672
1	Afghanistan	AFG	1991	0.036175	0.004319	0.225491	0.002107	0.2666
2	Afghanistan	AFG	1992	0.036245	0.004316	0.225336	0.002048	0.2662
3	Afghanistan	AFG	1993	0.036297	0.004315	0.225255	0.001993	0.2662
4	Afghanistan	AFG	1994	0.036312	0.004314	0.225156	0.001937	0.2660

Checking for missing values:

```
In [1390]: missing = pd.concat([df85.isnull().sum(), 100 * df85.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1390]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent)		0	0.000000
Code	980	15.151515	

```
In [1391]: v1='Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)'
v2='Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)'
v3='Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)'
v4='Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)'
v5='Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)'
v6='Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent'
v7='Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent)'
v8='Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent)'
v9='Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent)'
v10='Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent)'
v11='Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent)'
```

In [1392]: df85.describe()

Out[1392]:

	Year	Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)	P
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6
mean	2003.500000	0.042666	0.005188	0.288960	0.002565	0.345096	
std	8.078372	0.009310	0.001130	0.061503	0.000601	0.062829	
min	1990.000000	0.030337	0.003779	0.202374	0.001607	0.256615	
25%	1996.750000	0.036697	0.004569	0.245407	0.002188	0.299534	
50%	2003.500000	0.038052	0.004791	0.271532	0.002351	0.325965	
75%	2010.250000	0.046085	0.005223	0.316101	0.002694	0.375770	
max	2017.000000	0.090072	0.011866	0.524593	0.006480	0.614857	

In [1393]: df85.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1393]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)
1287	China	2017	0.090072
1286	China	2016	0.089634
1285	China	2015	0.089203
1707	East Asia	2017	0.088952
1284	China	2014	0.088774

In [1394]: df85.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1394]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)
1287	China	2017	0.011866
1286	China	2016	0.011792
1707	East Asia	2017	0.011741
1285	China	2015	0.011728
1284	China	2014	0.011676

In [1395]: df85.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1395]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)
1287	China	2017	0.524593
1286	China	2016	0.523339
1285	China	2015	0.522192
1284	China	2014	0.521194
1283	China	2013	0.520373

In [1396]: df85.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1396]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)
1271	China	2001	0.006480
1272	China	2002	0.006463
1273	China	2003	0.006421
1270	China	2000	0.006404
1691	East Asia	2001	0.006394

In [1397]: df85.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1397]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)
3989	Netherlands	2003	0.614857
3990	Netherlands	2004	0.614674
3987	Netherlands	2001	0.614577
3988	Netherlands	2002	0.614566
3991	Netherlands	2005	0.614276

In [1398]: df85.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1398]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)
3992	Netherlands	2006	0.375110
3991	Netherlands	2005	0.375092
3993	Netherlands	2007	0.375087
3994	Netherlands	2008	0.374991
3990	Netherlands	2004	0.374961

```
In [1399]: df85.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()
```

Out[1399]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent)
3990	Netherlands	2004	0.532016
3991	Netherlands	2005	0.531912
3989	Netherlands	2003	0.531774
3988	Netherlands	2002	0.531276
3992	Netherlands	2006	0.531156

```
In [1400]: df85.sort_values(by=v8, ascending=False)[['Entity', 'Year', v8]].head()
```

Out[1400]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent)
3981	Netherlands	1995	0.296528
3980	Netherlands	1994	0.296308
3982	Netherlands	1996	0.295984
3983	Netherlands	1997	0.295751
3979	Netherlands	1993	0.295606

```
In [1401]: df85.sort_values(by=v9, ascending=False)[['Entity', 'Year', v9]].head()
```

Out[1401]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent)
323	Australia	2005	0.308804
324	Australia	2006	0.308612
322	Australia	2004	0.308510
325	Australia	2007	0.308005
321	Australia	2003	0.307756

```
In [1402]: df85.sort_values(by=v10, ascending=False)[['Entity', 'Year', v10]].head()
```

Out[1402]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent)
1287	China	2017	0.614880
1286	China	2016	0.613682
1285	China	2015	0.612315
1284	China	2014	0.610656
1707	East Asia	2017	0.609698

In [1403]: df85.sort_values(by=v11, ascending=False)[['Entity', 'Year', v11]].head()

Out[1403]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent)
3992	Netherlands	2006	0.447585
3993	Netherlands	2007	0.447426
3991	Netherlands	2005	0.447384
3990	Netherlands	2004	0.447102
3994	Netherlands	2008	0.447036

In [1404]: df85.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1404]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)
5829	Trinidad and Tobago	1995	0.030337
5828	Trinidad and Tobago	1994	0.030354
5830	Trinidad and Tobago	1996	0.030355
5827	Trinidad and Tobago	1993	0.030399
5831	Trinidad and Tobago	1997	0.030401

In [1405]: df85.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1405]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)
5829	Trinidad and Tobago	1995	0.003779
5828	Trinidad and Tobago	1994	0.003783
5830	Trinidad and Tobago	1996	0.003784
5827	Trinidad and Tobago	1993	0.003793
5831	Trinidad and Tobago	1997	0.003796

In [1406]: df85.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1406]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)
3868	Mozambique	1994	0.202374
3869	Mozambique	1995	0.202392
3867	Mozambique	1993	0.202414
3870	Mozambique	1996	0.202486
3866	Mozambique	1992	0.202508

In [1407]: df85.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1407]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)
10	Afghanistan	2000	0.001607
11	Afghanistan	2001	0.001612
9	Afghanistan	1999	0.001630
12	Afghanistan	2002	0.001677
8	Afghanistan	1998	0.001681

In [1408]: df85.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1408]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)
3865	Mozambique	1991	0.256615
3866	Mozambique	1992	0.256653
3864	Mozambique	1990	0.256671
3867	Mozambique	1993	0.256779
3868	Mozambique	1994	0.256969

In [1409]: df85.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[1409]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)
3866	Mozambique	1992	0.146902
3867	Mozambique	1993	0.146907
3865	Mozambique	1991	0.146938
3868	Mozambique	1994	0.146949
3869	Mozambique	1995	0.147017

In [1410]: df85.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[1410]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent)
3	Afghanistan	1993	0.169030
2	Afghanistan	1992	0.170344
4	Afghanistan	1994	0.172134
5163	Somalia	2001	0.173466
5162	Somalia	2000	0.173644

In [1411]: df85.sort_values(by=v8, ascending=True)[['Entity', 'Year', v8]].head()

Out[1411]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent)
3268	Liberia	2010	0.120007
3267	Liberia	2009	0.120077
3269	Liberia	2011	0.120197
3271	Liberia	2013	0.120281
3266	Liberia	2008	0.120506

In [1412]: df85.sort_values(by=v9, ascending=True)[['Entity', 'Year', v9]].head()

Out[1412]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent)
5824	Trinidad and Tobago	1990	0.113679
5825	Trinidad and Tobago	1991	0.113721
5826	Trinidad and Tobago	1992	0.113768
5827	Trinidad and Tobago	1993	0.113814
5828	Trinidad and Tobago	1994	0.113868

In [1413]: df85.sort_values(by=v10, ascending=True)[['Entity', 'Year', v10]].head()

Out[1413]:

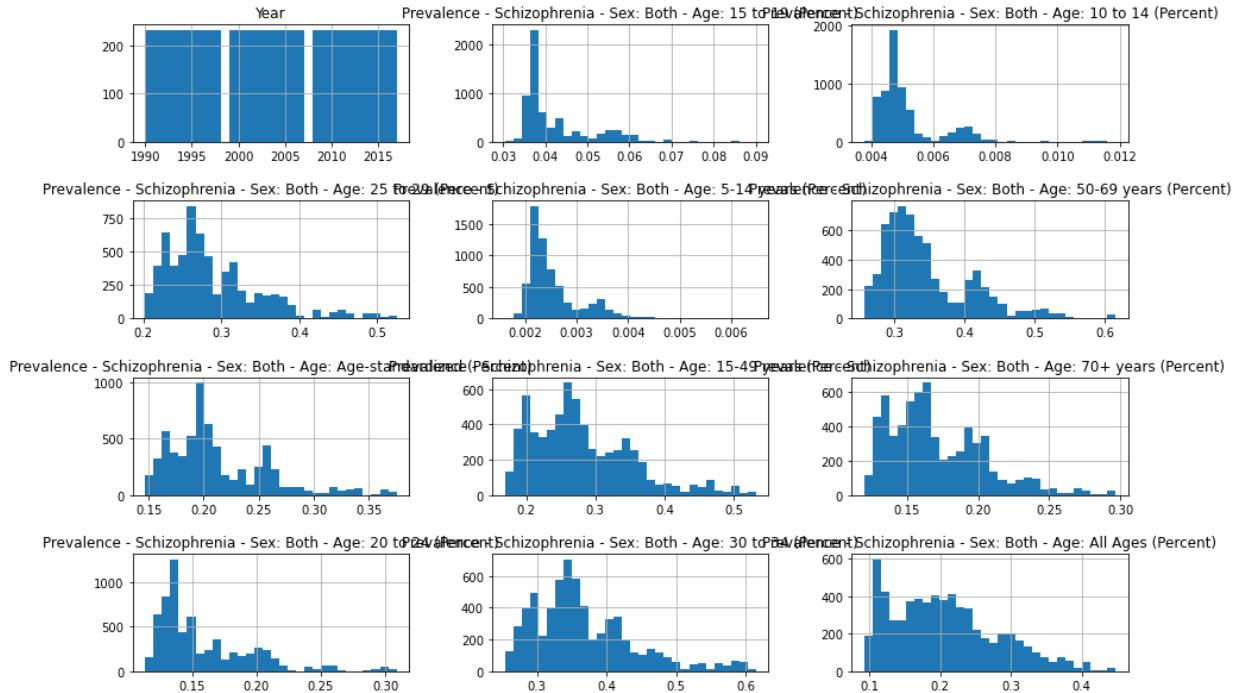
	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent)
3869	Mozambique	1995	0.253653
3868	Mozambique	1994	0.253676
3870	Mozambique	1996	0.253797
3867	Mozambique	1993	0.253804
3866	Mozambique	1992	0.254042

In [1414]: df85.sort_values(by=v11, ascending=True)[['Entity', 'Year', v11]].head()

Out[1414]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent)
5974	Uganda	2000	0.092744
5975	Uganda	2001	0.092778
5973	Uganda	1999	0.092814
5972	Uganda	1998	0.092966
5976	Uganda	2002	0.092969

```
In [1415]: df85.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1416]: df85.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1416]: Entity
Trinidad and Tobago      0.030659
Suriname                  0.033282
Haiti                     0.033537
Guyana                    0.034582
Tajikistan                 0.034624
Name: Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent), d
type: float64
```

```
In [1417]: df85.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1417]: Entity
Trinidad and Tobago      0.003843
Tajikistan                 0.004088
Suriname                   0.004104
Uzbekistan                  0.004137
Haiti                      0.004146
Name: Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent), d
type: float64
```

```
In [1418]: df85.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1418]: Entity
Central African Republic    0.204767
Mozambique                  0.207000
Malawi                      0.207724
Burundi                     0.210949
Somalia                     0.211965
Name: Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent), d
dtype: float64
```

```
In [1419]: df85.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1419]: Entity
Afghanistan                 0.001949
Tajikistan                  0.001961
Suriname                    0.002001
Haiti                       0.002007
Trinidad and Tobago        0.002015
Name: Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent),
dtype: float64
```

```
In [1420]: df85.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1420]: Entity
Central African Republic    0.261923
Mozambique                  0.263719
Malawi                      0.264461
Somalia                     0.265359
Burundi                     0.266460
Name: Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percen
t), dtype: float64
```

```
In [1421]: df85.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1421]: Entity
Central African Republic    0.148887
Mozambique                  0.150924
Malawi                      0.152340
Somalia                     0.153407
Burundi                     0.153603
Name: Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Per
cent), dtype: float64
```

```
In [1422]: df85.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1422]: Entity
Uganda          0.177339
Malawi          0.179678
Mozambique      0.181110
Central African Republic  0.181511
Somalia          0.183178
Name: Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent),
      dtype: float64
```

```
In [1423]: df85.groupby('Entity')[v8].mean().sort_values().head()
```

```
Out[1423]: Entity
Liberia          0.122798
Mozambique       0.123914
Malawi           0.124006
Somalia          0.126408
Burundi          0.127479
Name: Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent),
      dtype: float64
```

```
In [1424]: df85.groupby('Entity')[v9].mean().sort_values().head()
```

```
Out[1424]: Entity
Trinidad and Tobago  0.114454
Central African Republic  0.117076
Mozambique          0.118121
Malawi              0.119013
Burundi              0.120347
Name: Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent), d
type: float64
```

```
In [1425]: df85.groupby('Entity')[v10].mean().sort_values().head()
```

```
Out[1425]: Entity
Central African Republic  0.256519
Mozambique          0.259754
Malawi              0.260072
Burundi              0.264966
Lesotho              0.266694
Name: Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent), d
type: float64
```

```
In [1426]: df85.groupby('Entity')[v11].mean().sort_values().head()
```

```
Out[1426]: Entity
Uganda          0.096065
Niger           0.103366
Mozambique      0.103865
Chad            0.104133
Ethiopia         0.104354
Name: Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent), d
type: float64
```

```
In [1427]: df85.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1427]: Entity
Australasia          0.068133
Australia            0.069387
Southeast Asia, East Asia, and Oceania 0.075331
East Asia             0.085542
China                0.086397
Name: Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [1428]: df85.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1428]: Entity
Taiwan               0.008088
North Korea          0.008210
Southeast Asia, East Asia, and Oceania 0.009566
East Asia             0.011242
China                0.011358
Name: Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1429]: df85.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1429]: Entity
United States        0.461824
Australasia          0.484100
Australia            0.490892
East Asia             0.504667
China                0.509231
Name: Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1430]: df85.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1430]: Entity
North Korea          0.004100
Taiwan               0.004227
Southeast Asia, East Asia, and Oceania 0.004847
East Asia             0.005773
China                0.005834
Name: Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent), dtype: float64
```

```
In [1431]: df85.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1431]: Entity
Greenland            0.515008
United States        0.515606
Australasia          0.534153
Australia            0.538887
Netherlands          0.606893
Name: Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent), dtype: float64
```

```
In [1432]: df85.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1432]: Entity
United States      0.338117
New Zealand        0.340113
Australasia        0.360540
Australia          0.364498
Netherlands        0.372330
Name: Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1433]: df85.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1433]: Entity
China              0.467363
United States      0.468521
Australasia        0.496422
Australia          0.503125
Netherlands        0.519111
Name: Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [1434]: df85.groupby('Entity')[v8].mean().sort_values().tail()
```

```
Out[1434]: Entity
Southeast Asia, East Asia, and Oceania  0.253756
Greenland                         0.259705
East Asia                          0.271381
China                            0.273436
Netherlands                       0.292468
Name: Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [1435]: df85.groupby('Entity')[v9].mean().sort_values().tail()
```

```
Out[1435]: Entity
Southeast Asia, East Asia, and Oceania  0.262544
East Asia                           0.292715
China                             0.295510
Australasia                        0.297176
Australia                          0.305392
Name: Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1436]: df85.groupby('Entity')[v10].mean().sort_values().tail()
```

```
Out[1436]: Entity
Australasia          0.587695
East Asia            0.588361
Netherlands         0.592636
Australia           0.593558
China               0.593668
Name: Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

In [1437]: `df85.groupby('Entity')[v1].mean().sort_values().tail()`

Out[1437]: Entity

North America	0.375982
United States	0.377144
Australasia	0.399857
Australia	0.405343
Netherlands	0.439929

Name: Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent), d
type: float64

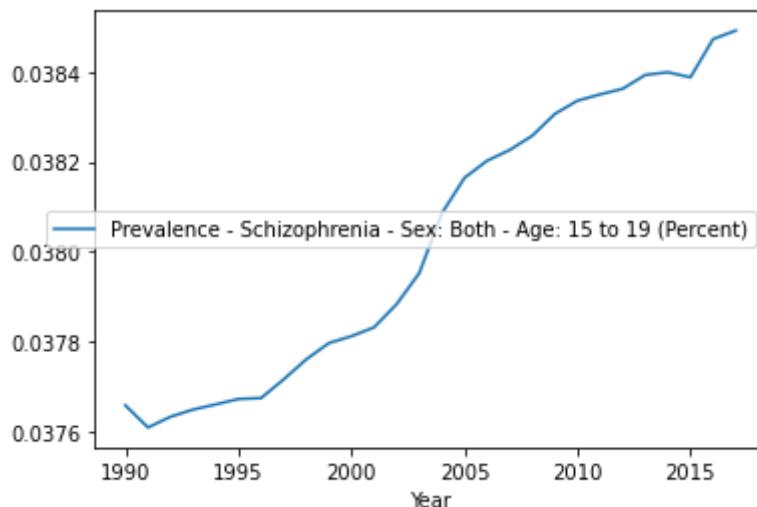
In [1438]: `df85_mean = df85.groupby('Year').mean()
df85_mean.head()`

Out[1438]:

Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)	Sc
Year						
1990	0.042259	0.005097	0.287310	0.002456	0.342562	0.209548
1991	0.042255	0.005101	0.287285	0.002460	0.342292	0.209586
1992	0.042254	0.005105	0.287268	0.002467	0.342042	0.209634
1993	0.042259	0.005109	0.287258	0.002474	0.341811	0.209690
1994	0.042268	0.005114	0.287255	0.002481	0.341626	0.209751

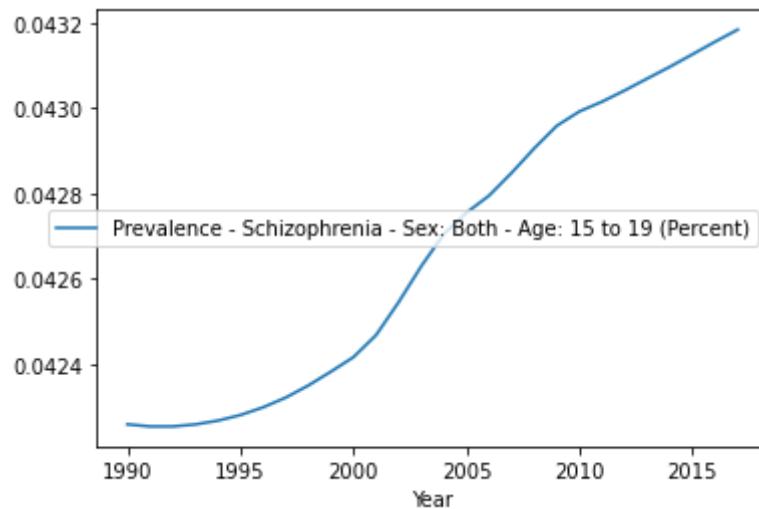
In [1439]: `df85.groupby('Year')[v1].median().plot(legend=True)`

Out[1439]: <AxesSubplot:xlabel='Year'>



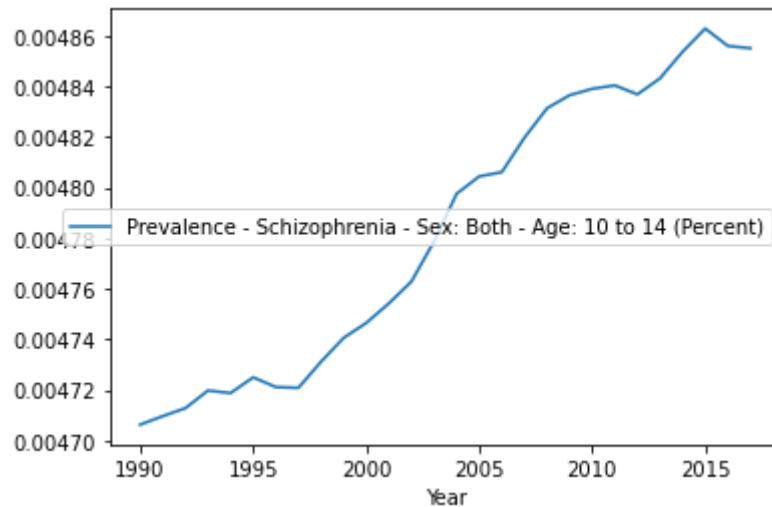
```
In [1440]: df85.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1440]: <AxesSubplot:xlabel='Year'>
```



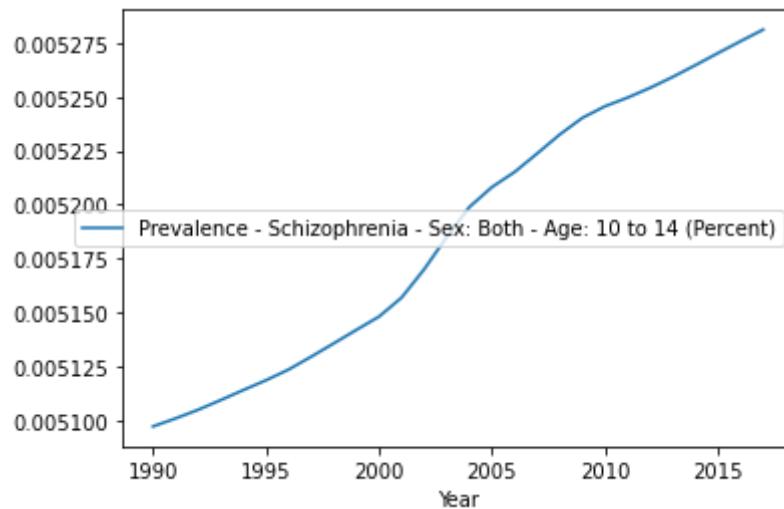
```
In [1441]: df85.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1441]: <AxesSubplot:xlabel='Year'>
```



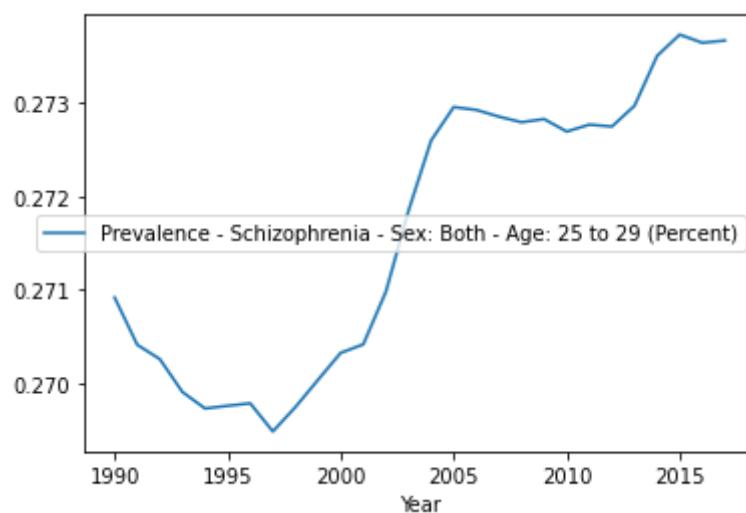
```
In [1442]: df85.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1442]: <AxesSubplot:xlabel='Year'>
```



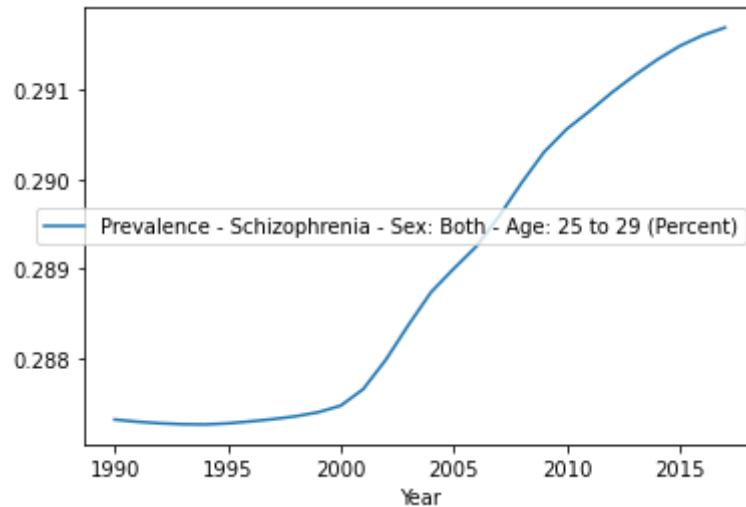
```
In [1443]: df85.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1443]: <AxesSubplot:xlabel='Year'>
```



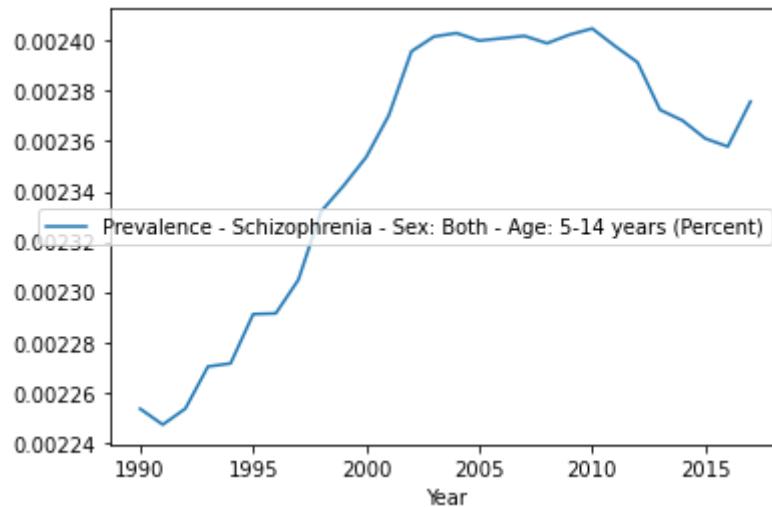
```
In [1444]: df85.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1444]: <AxesSubplot:xlabel='Year'>
```



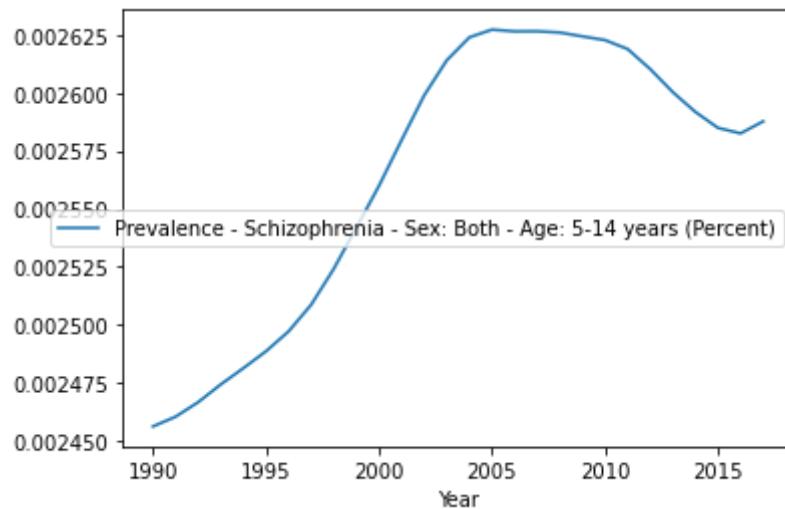
```
In [1445]: df85.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1445]: <AxesSubplot:xlabel='Year'>
```



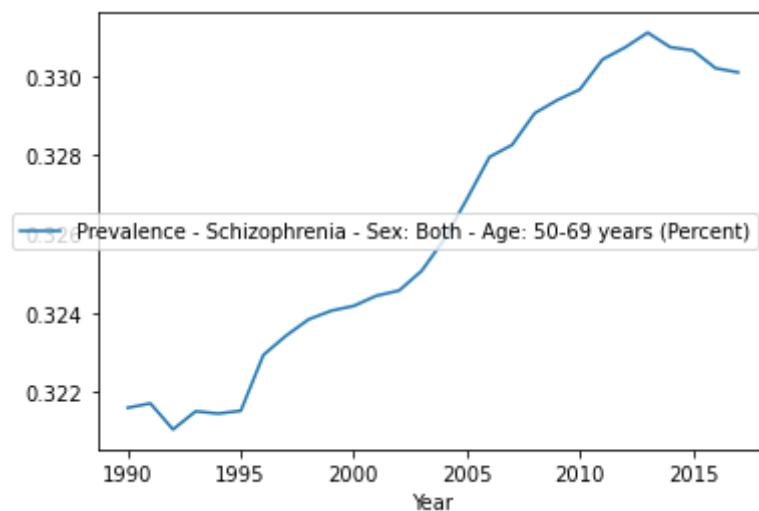
```
In [1446]: df85.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1446]: <AxesSubplot:xlabel='Year'>
```



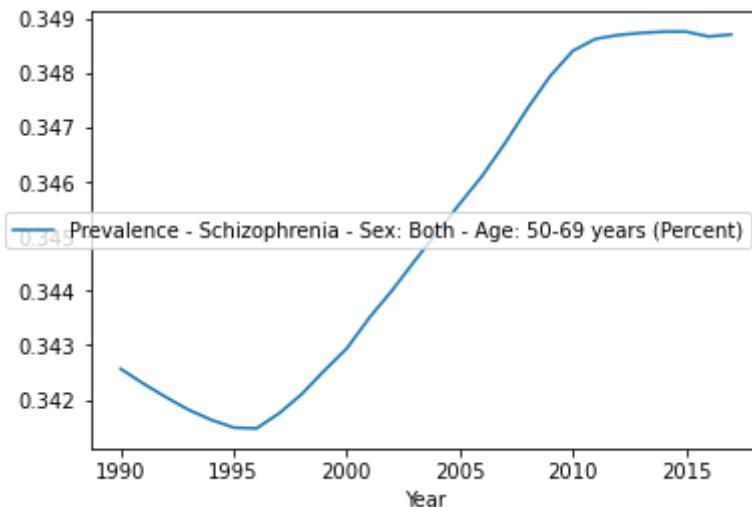
```
In [1447]: df85.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1447]: <AxesSubplot:xlabel='Year'>
```



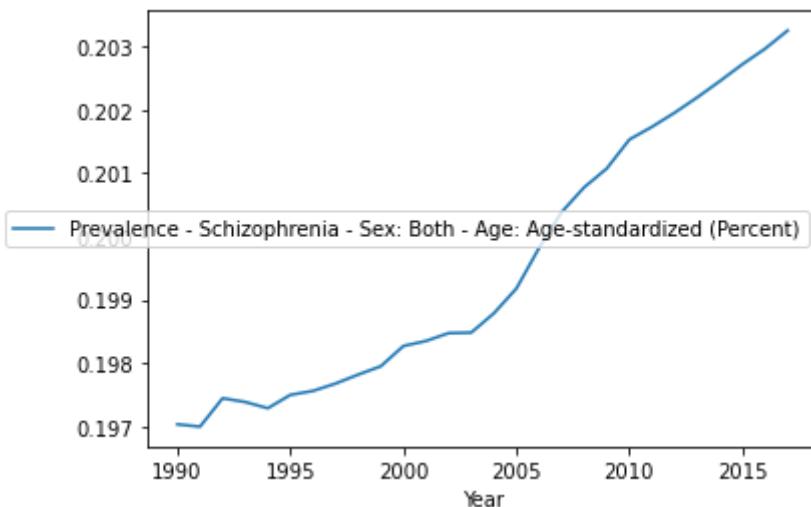
```
In [1448]: df85.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1448]: <AxesSubplot:xlabel='Year'>
```



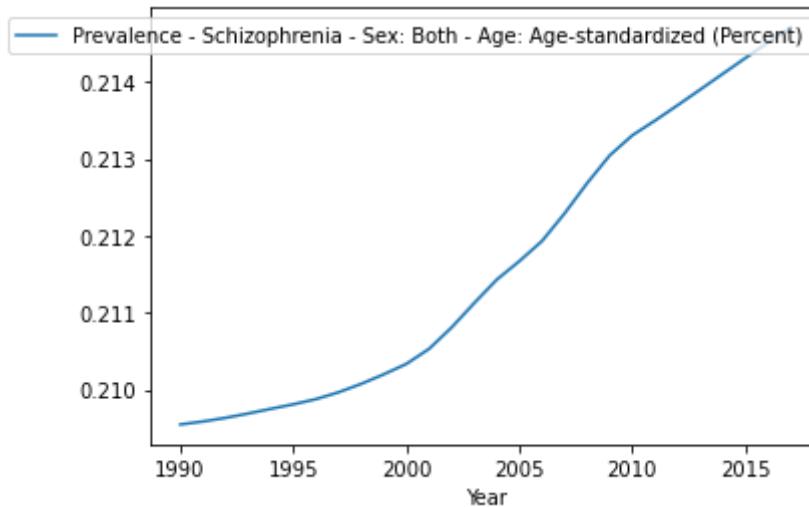
```
In [1449]: df85.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1449]: <AxesSubplot:xlabel='Year'>
```



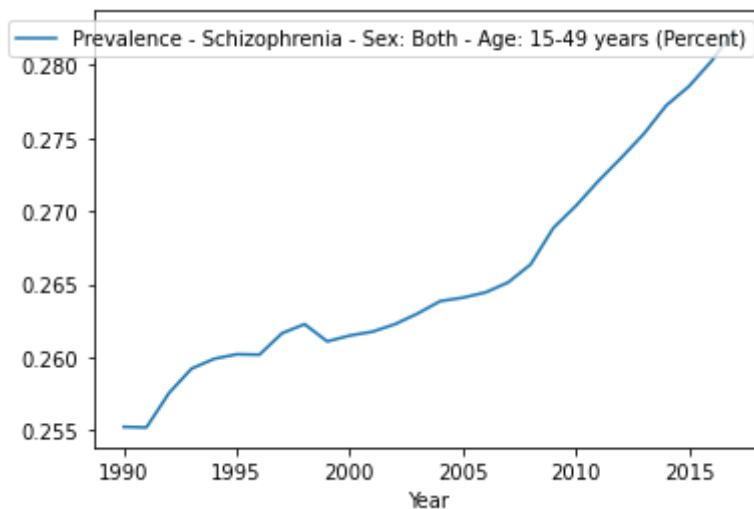
```
In [1450]: df85.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1450]: <AxesSubplot:xlabel='Year'>
```



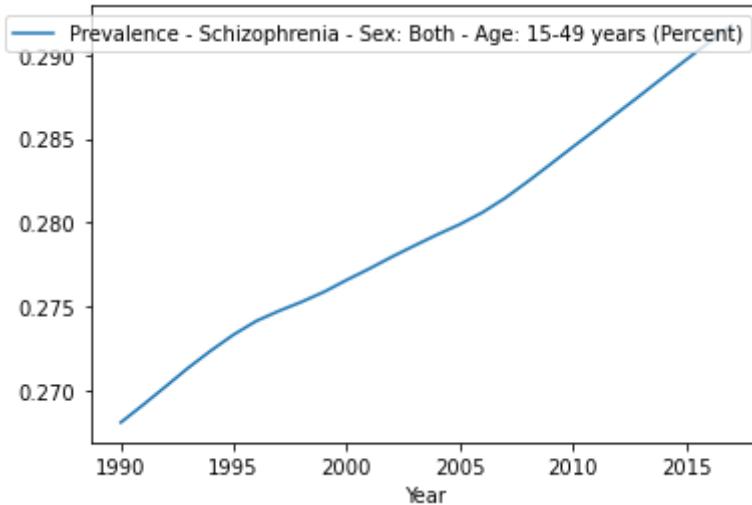
```
In [1451]: df85.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1451]: <AxesSubplot:xlabel='Year'>
```



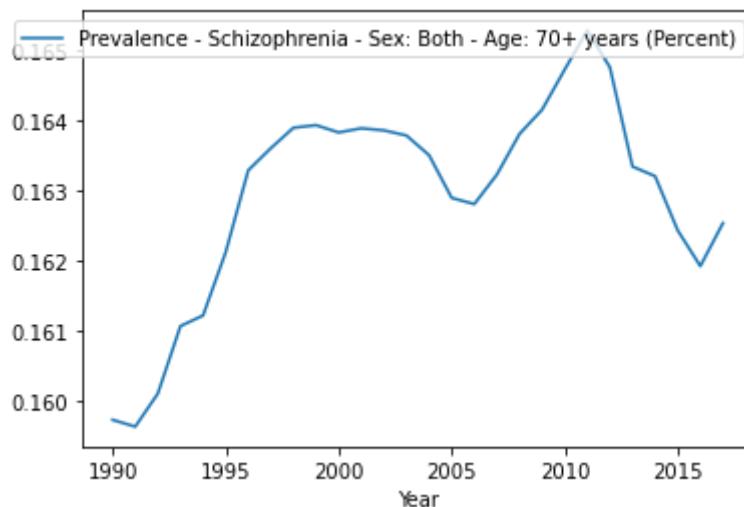
```
In [1452]: df85.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1452]: <AxesSubplot:xlabel='Year'>
```



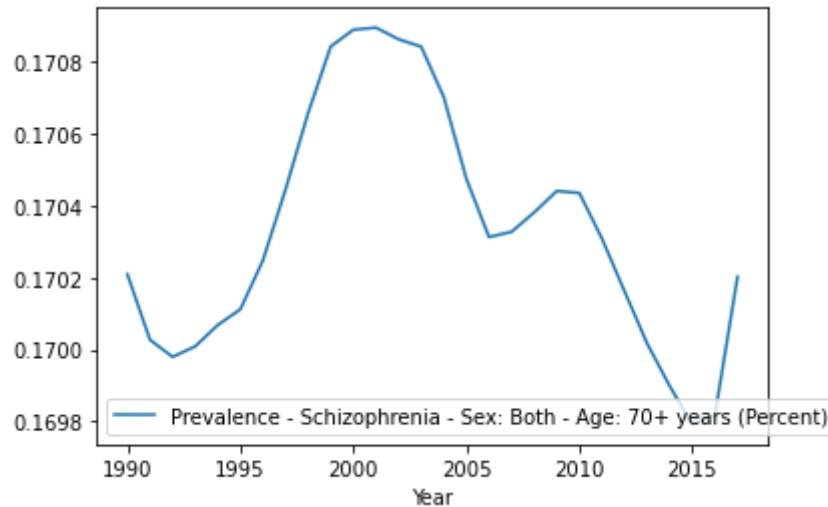
```
In [1453]: df85.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[1453]: <AxesSubplot:xlabel='Year'>
```



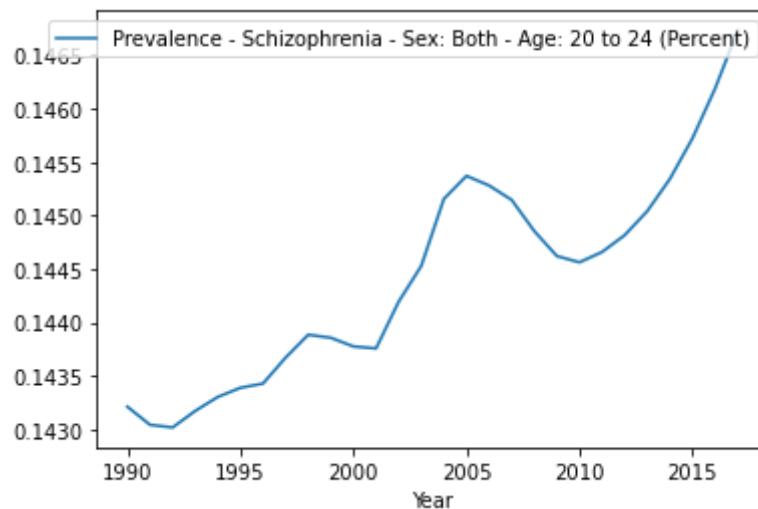
```
In [1454]: df85.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[1454]: <AxesSubplot:xlabel='Year'>
```



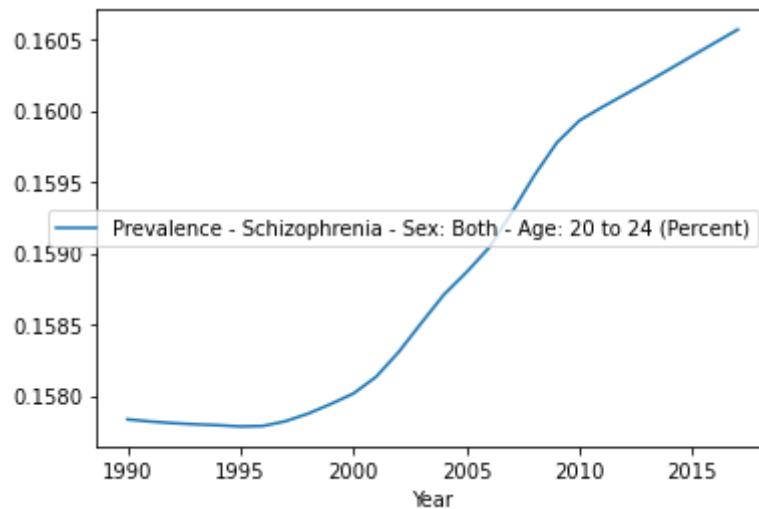
```
In [1455]: df85.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[1455]: <AxesSubplot:xlabel='Year'>
```



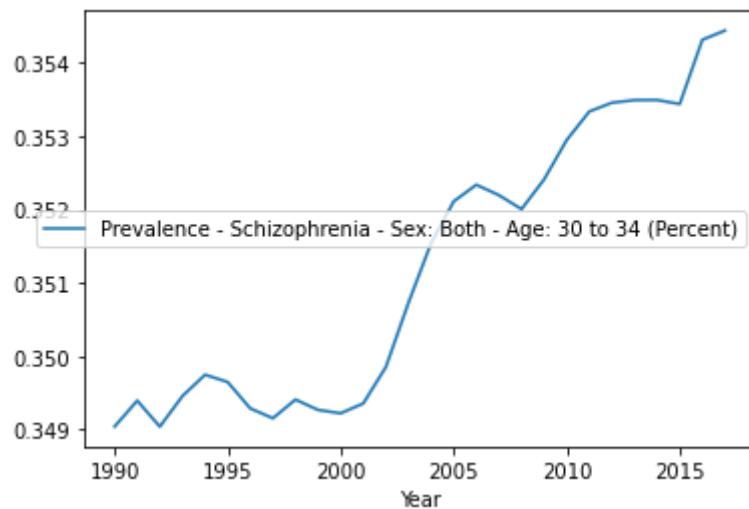
```
In [1456]: df85.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[1456]: <AxesSubplot:xlabel='Year'>
```



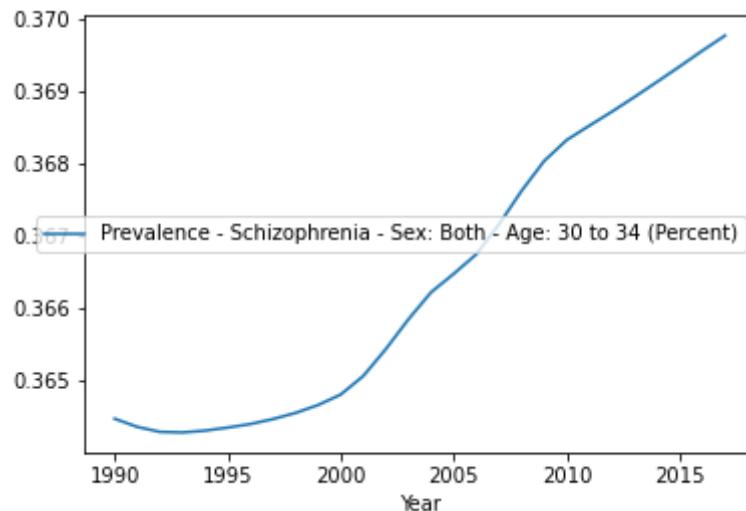
```
In [1457]: df85.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[1457]: <AxesSubplot:xlabel='Year'>
```



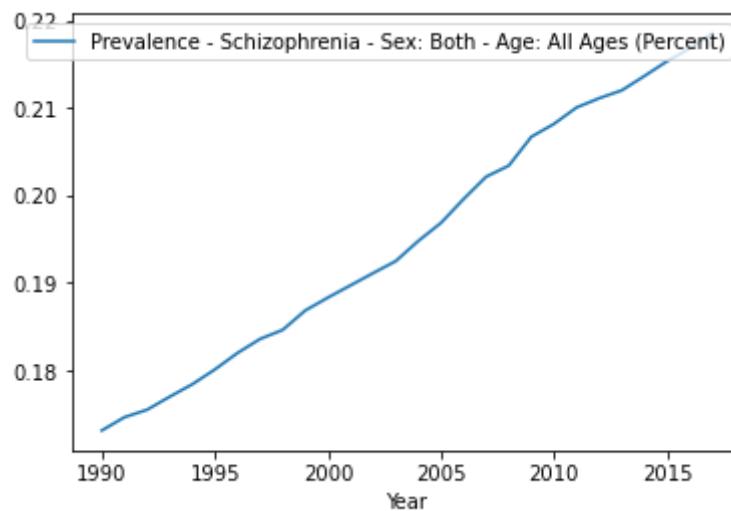
```
In [1458]: df85.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[1458]: <AxesSubplot:xlabel='Year'>
```



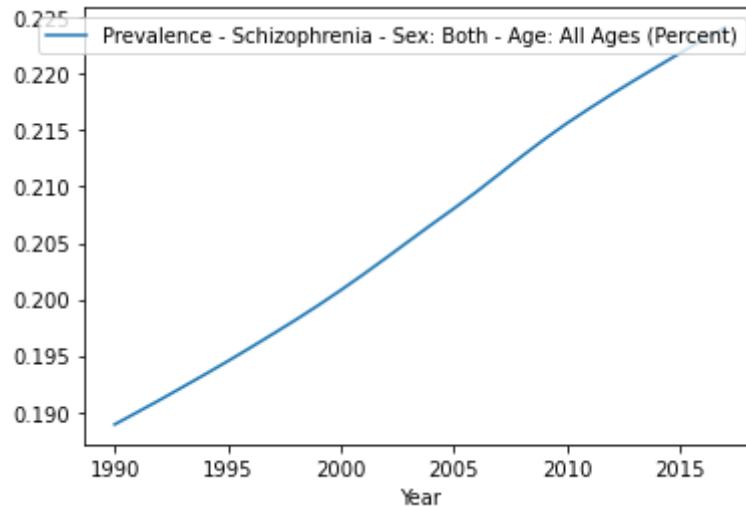
```
In [1459]: df85.groupby('Year')[v11].median().plot(legend=True)
```

```
Out[1459]: <AxesSubplot:xlabel='Year'>
```



```
In [1460]: df85.groupby('Year')[v11].mean().plot(legend=True)
```

```
Out[1460]: <AxesSubplot:xlabel='Year'>
```



Depressive Disorders

In [1461]: df39.info()
df39.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 10 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate)       6468 non-null   float64
4   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14 years (Rate)      6468 non-null   float64
5   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49 years (Rate)     6468 non-null   float64
6   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 50-69 years (Rate)    6468 non-null   float64
7   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 70+ years (Rate)      6468 non-null   float64
8   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate)      6468 non-null   float64
9   DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age-standardized (Rate) 6468 non-null   float64
dtypes: float64(7), int64(1), object(2)
memory usage: 505.4+ KB
```

Out[1461]:

				DALYs (Disability- Adjusted Life Years)	DALYs (Disability- Adjusted Life Years)	DALYs (Disability- Adjusted Life Years)	DALYs (Disability- Adjusted Life Years)	DALYs (Disability- Adjusted Life Years)	DALYs (Disability- Adjusted Life Years)
	Entity	Code	Year	Depressive disorders - Sex: Both - Age: Under 5 (Rate)	Depressive disorders - Sex: Both - Age: 5-14 years (Rate)	Depressive disorders - Sex: Both - Age: 15-49 years (Rate)	Depressive disorders - Sex: Both - Age: 50-69 years (Rate)	Depressive disorders - Sex: Both - Age: 70+ years (Rate)	Depressive disorders - Sex: Both - Age: All Ages (Rate)
0	Afghanistan	AFG	1990	0.271187	171.879841	875.280470	950.293471	766.893212	551.398203
1	Afghanistan	AFG	1991	0.268624	166.950366	870.018856	950.792194	765.393867	549.162279
2	Afghanistan	AFG	1992	0.260249	162.987003	859.514720	954.455355	762.681493	543.410963
3	Afghanistan	AFG	1993	0.258343	159.584534	856.438776	957.746985	760.617836	538.988725
4	Afghanistan	AFG	1994	0.263607	155.962684	861.397337	958.766015	759.161396	532.539812

Checking for missing values:

```
In [1462]: missing = pd.concat([df39.isnull().sum(), 100 * df39.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1462]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 50-69 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 70+ years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
Code	980	15.151515	

```
In [1463]: v1='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both'
v2='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both'
v3='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both'
v4='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both'
v5='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both'
v6='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both'
v7='DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both'
```

In [1464]: df39.describe()

Out[1464]:

	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 50-69	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 70+	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.232838	132.324244	679.344977	894.799816	933.990147
std	8.078372	0.040015	38.473408	178.660612	193.140777	273.545942
min	1990.000000	0.176387	65.557408	303.852234	468.441748	414.618681
25%	1996.750000	0.204791	102.491070	552.578225	750.043786	724.862304
50%	2003.500000	0.222612	123.581776	652.543869	855.322047	837.181963
75%	2010.250000	0.244009	155.095995	784.221981	1039.410953	1157.724780
max	2017.000000	0.389520	303.308292	1724.945246	1741.285902	2000.445977
						1180.048882

In [1465]: df39.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1465]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate)
5404	Spain 1990	0.389520
5414	Spain 2000	0.381926
5415	Spain 2001	0.380361
5416	Spain 2002	0.379535
5417	Spain 2003	0.377634

In [1466]: df39.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1466]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14 years (Rate)
2286	Greenland	2008	303.308292
2287	Greenland	2009	303.043981
2288	Greenland	2010	302.073443
2285	Greenland	2007	300.424481
2284	Greenland	2006	300.043383

In [1467]: df39.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1467]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49 years (Rate)
2273	Greenland	1995	1724.945246
2272	Greenland	1994	1723.782798
2274	Greenland	1996	1718.402413
2271	Greenland	1993	1715.902657
2275	Greenland	1997	1711.513772

In [1468]: df39.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1468]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 50-69 years (Rate)
5975	Uganda	2001	1741.285902
5976	Uganda	2002	1741.251298
5974	Uganda	2000	1740.836864
5977	Uganda	2003	1738.534355
5978	Uganda	2004	1734.726577

In [1469]: df39.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1469]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 70+ years (Rate)	
5978	Uganda	2004	2000.445977
5977	Uganda	2003	1999.738440
5976	Uganda	2002	1999.203123
5979	Uganda	2005	1999.143115
5980	Uganda	2006	1997.685754

In [1470]: df39.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1470]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate)	
2271	Greenland	1993	1180.048882
2272	Greenland	1994	1177.747401
2270	Greenland	1992	1176.737668
2273	Greenland	1995	1172.970775
2269	Greenland	1991	1171.050707

In [1471]: df39.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1471]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Age-standardized (Rate)	
2273	Greenland	1995	1105.100519
2274	Greenland	1996	1104.889557
2275	Greenland	1997	1103.961709
2276	Greenland	1998	1101.225224
2272	Greenland	1994	1101.018393

In [1472]: df39.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1472]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate)	
1287	China	2017	0.176387
4723	Russia	2009	0.177052
1707	East Asia	2017	0.177120
2727	Indonesia	2001	0.177215
2726	Indonesia	2000	0.177306

In [1473]: df39.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1473]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14 years (Rate)	
4590	Poland	2016	65.557408
4591	Poland	2017	65.696987
4589	Poland	2015	65.965244
4588	Poland	2014	67.232442
3902	Myanmar	2000	67.367820

In [1474]: df39.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1474]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49 years (Rate)	
3892	Myanmar	1990	303.852234
3901	Myanmar	1999	305.001979
3893	Myanmar	1991	305.096655
3902	Myanmar	2000	305.360037
3903	Myanmar	2001	305.677902

In [1475]: df39.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1475]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 50-69 years (Rate)	
804	Brunei	2010	468.441748
805	Brunei	2011	468.574534
803	Brunei	2009	468.942650
806	Brunei	2012	469.426053
802	Brunei	2008	469.604405

In [1476]: df39.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1476]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 70+ years (Rate)	
971	Canada	2009	414.618681
970	Canada	2008	415.017899
972	Canada	2010	415.305212
969	Canada	2007	415.635424
973	Canada	2011	415.908770

In [1477]: `df39.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()`

Out[1477]:

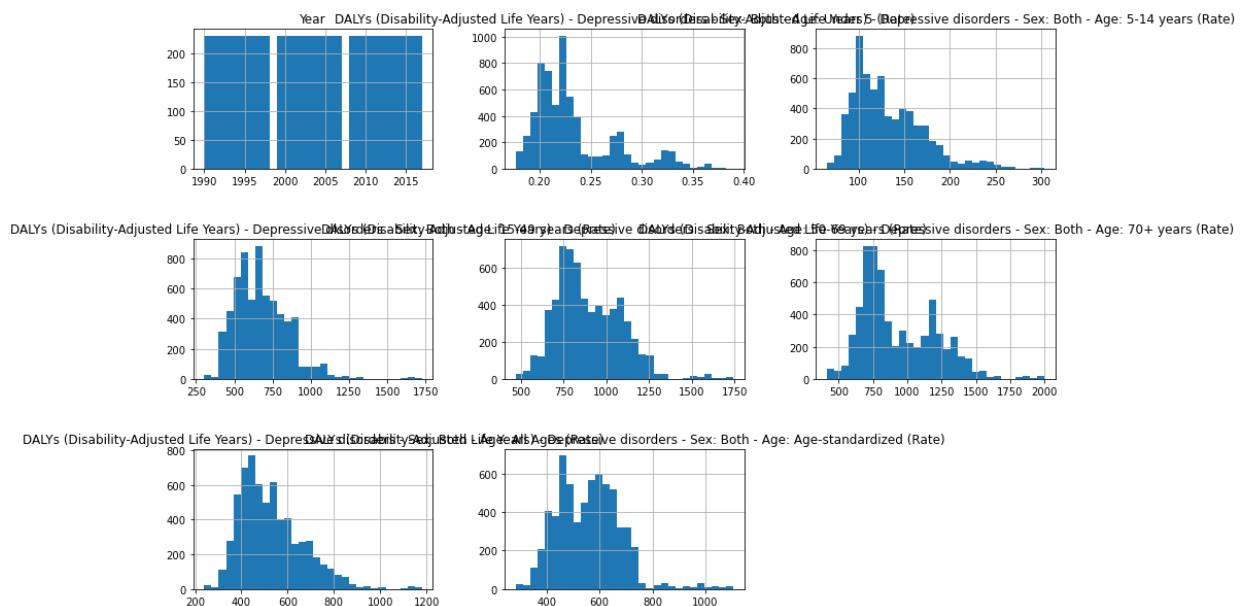
Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate)	
3892	Myanmar	1990	239.566756
3893	Myanmar	1991	241.201574
3894	Myanmar	1992	242.608265
3895	Myanmar	1993	244.079110
3896	Myanmar	1994	245.311016

In [1478]: `df39.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()`

Out[1478]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Age-standardized (Rate)	
3912	Myanmar	2010	282.190619
3911	Myanmar	2009	282.253931
3913	Myanmar	2011	282.567662
3910	Myanmar	2008	282.761695
3909	Myanmar	2007	283.058267

In [1479]: `df39.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [1480]: df39.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1480]: Entity
Indonesia      0.179061
Myanmar        0.181682
India          0.181751
Mexico         0.182003
Russia         0.182526
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: Under 5 (Rate), dtype: float64
```

```
In [1481]: df39.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1481]: Entity
Myanmar        68.482659
Poland         72.257298
India          79.073984
Taiwan         81.757709
Mexico         84.142384
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [1482]: df39.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1482]: Entity
Myanmar        307.912132
Albania         397.437054
Indonesia       410.737674
North Macedonia 417.451300
Peru            420.883934
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1483]: df39.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1483]: Entity
Brunei         473.005882
Albania         535.425960
Myanmar        550.910491
Poland         579.861105
Japan           584.851799
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [1484]: df39.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1484]: Entity
Canada          442.545056
New Zealand     459.692560
Brunei          460.212533
Myanmar         530.524176
North America   535.934687
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [1485]: df39.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1485]: Entity
Myanmar      259.088047
Tajikistan   318.842343
Mali         328.666857
Peru          331.769285
Philippines   335.932651
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1486]: df39.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1486]: Entity
Myanmar      284.939023
Albania       335.595769
Poland        345.686952
Indonesia    362.320168
Peru          363.338689
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [1487]: df39.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1487]: Entity
Northern Ireland  0.339812
Finland           0.343630
Portugal          0.355507
Spain              0.356614
Greece             0.363844
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate), dtype: float64
```

```
In [1488]: df39.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1488]: Entity
Uganda            239.722250
Morocco           244.684171
Greece             245.860838
Portugal           248.951169
Greenland          282.631322
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [1489]: df39.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1489]: Entity
Lesotho           1107.081642
Finland           1128.748422
Iran               1130.221342
Morocco           1280.424353
Greenland          1658.496138
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1490]: df39.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1490]: Entity
Ethiopia    1260.553674
Ukraine    1262.672080
Morocco    1546.835751
Lesotho     1552.834170
Uganda      1671.624884
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: 50-69 years (Rate), dtype: float64
```

```
In [1491]: df39.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1491]: Entity
Gambia      1548.685093
Ukraine    1555.625947
Estonia     1587.808872
Lesotho     1848.314194
Uganda      1951.732890
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: 70+ years (Rate), dtype: float64
```

```
In [1492]: df39.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1492]: Entity
Portugal    836.432453
Ukraine    848.739360
Finland     891.840965
Morocco     945.415412
Greenland   1145.369112
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1493]: df39.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1493]: Entity
Iran        840.474697
Uganda     938.452363
Lesotho     958.533462
Morocco    1005.736803
Greenland   1075.341406
Name: DALYs (Disability-Adjusted Life Years) - Depressive disorders - Se
x: Both - Age: Age-standardized (Rate), dtype: float64
```

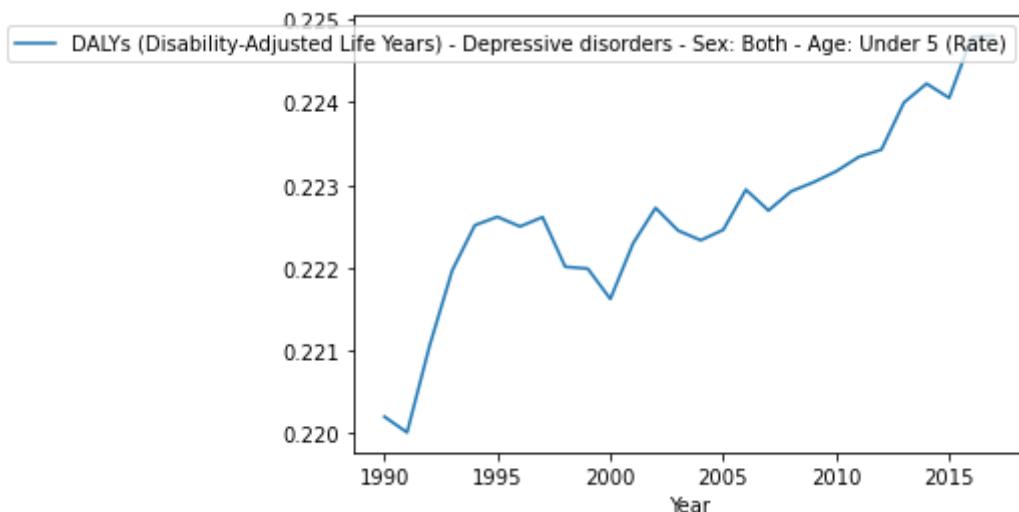
```
In [1494]: df39_mean = df39.groupby('Year').mean()
df39_mean.head()
```

Out[1494]:

DALYs (Disability- Adjusted Life Years)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14 years (Rate)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate)	DALYs (Disability- Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Age- standardized (Rate)
Year							
1990	0.231689	129.775105	682.280112	898.421760	938.443861	504.288010	564.208884
1991	0.231725	129.645250	683.724204	899.490405	939.324371	506.394772	564.845759
1992	0.231918	129.654267	684.932890	900.600454	940.178348	508.425414	565.386328
1993	0.232232	129.747168	685.973018	901.737357	940.887583	510.478334	565.822230
1994	0.232472	129.816598	686.864298	902.701204	941.549008	512.528678	566.156732

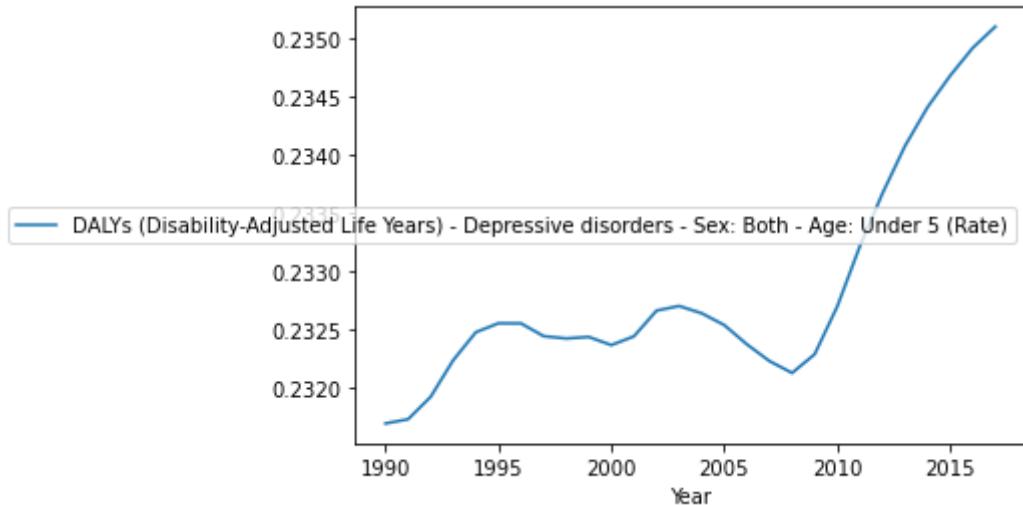
```
In [1495]: df39.groupby('Year')[v1].median().plot(legend=True)
```

Out[1495]: <AxesSubplot:xlabel='Year'>



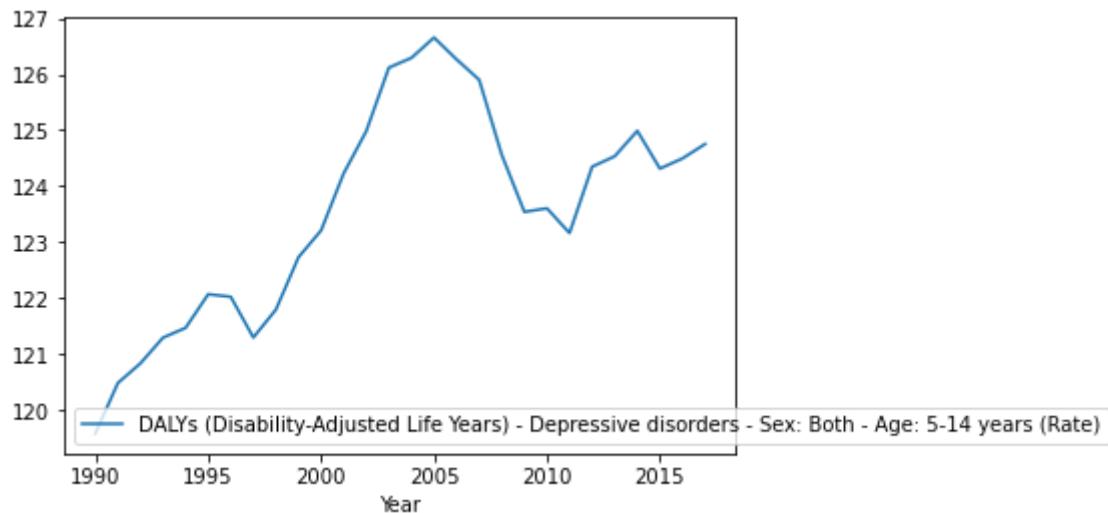
```
In [1496]: df39.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1496]: <AxesSubplot:xlabel='Year'>
```



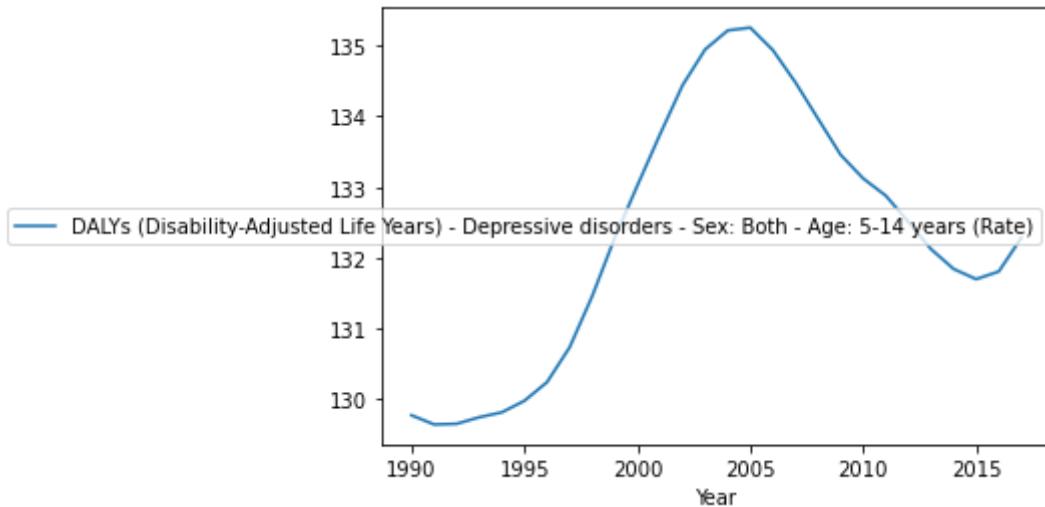
```
In [1497]: df39.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1497]: <AxesSubplot:xlabel='Year'>
```



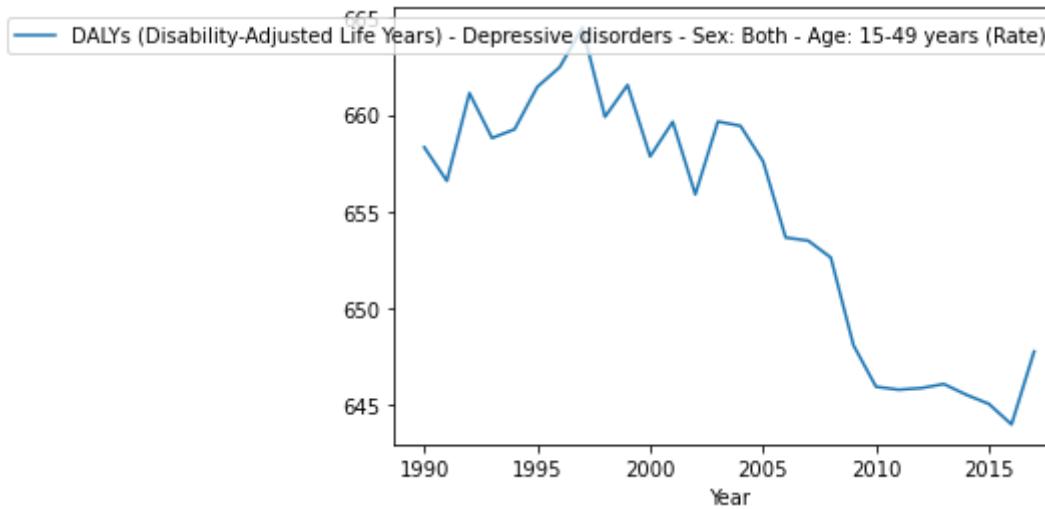
```
In [1498]: df39.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1498]: <AxesSubplot:xlabel='Year'>
```



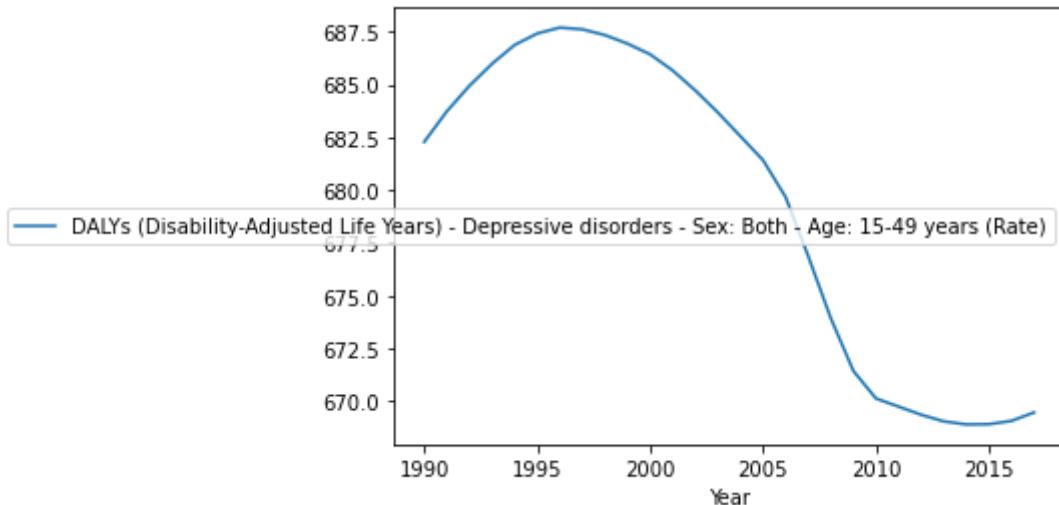
```
In [1499]: df39.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1499]: <AxesSubplot:xlabel='Year'>
```



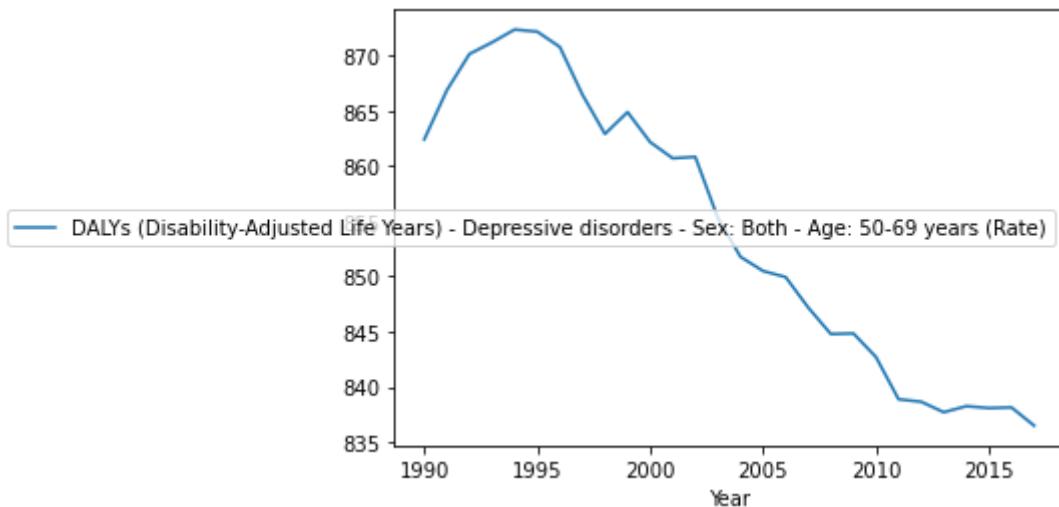
```
In [1500]: df39.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1500]: <AxesSubplot:xlabel='Year'>
```



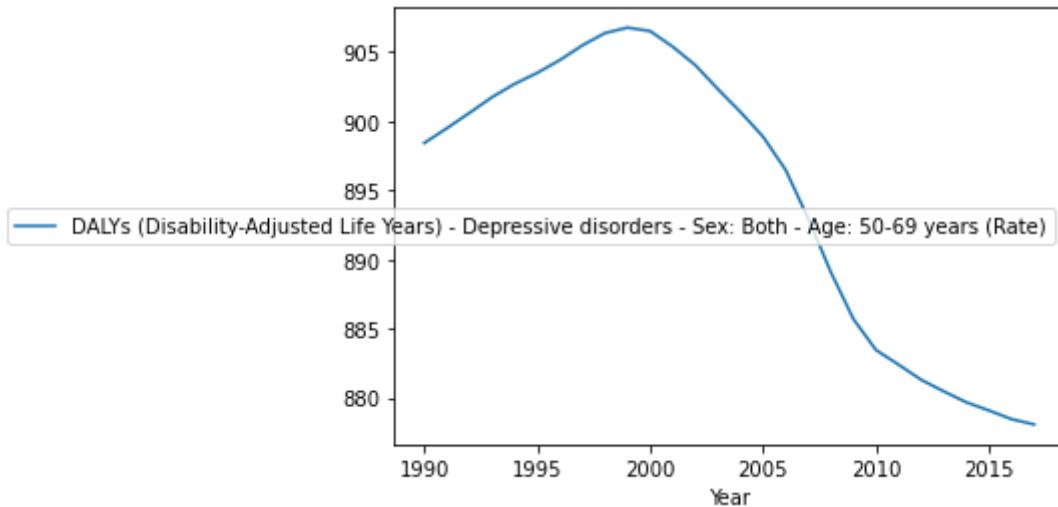
```
In [1501]: df39.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1501]: <AxesSubplot:xlabel='Year'>
```



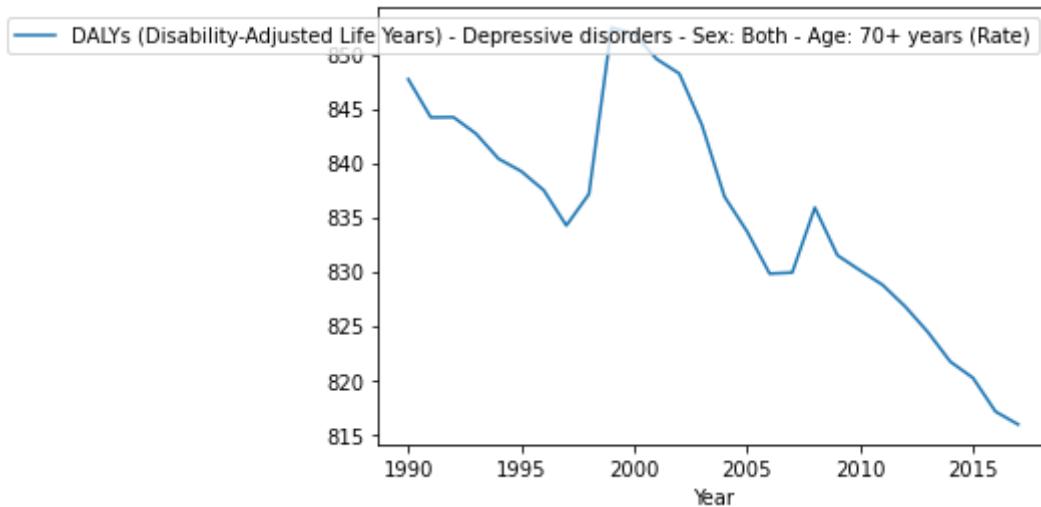
```
In [1502]: df39.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1502]: <AxesSubplot:xlabel='Year'>
```



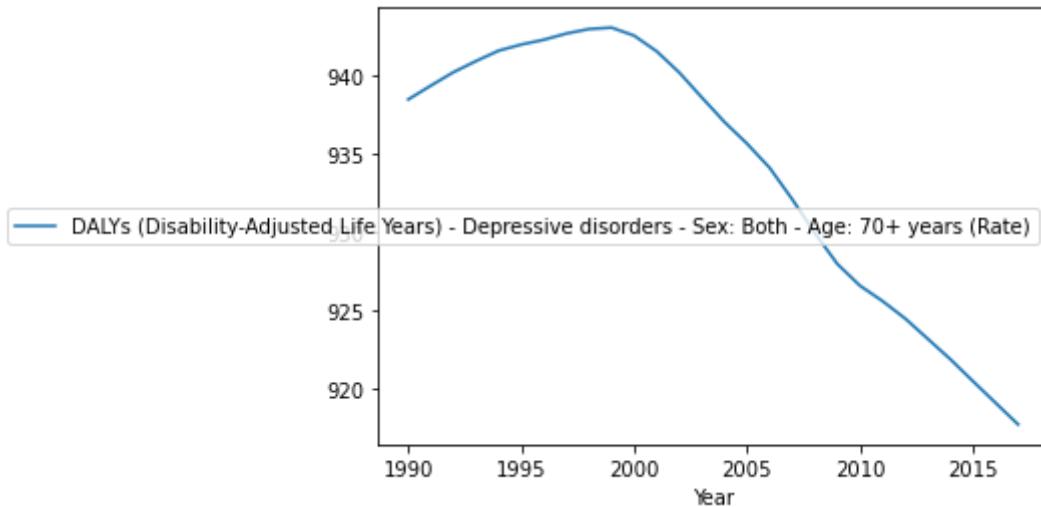
```
In [1503]: df39.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1503]: <AxesSubplot:xlabel='Year'>
```



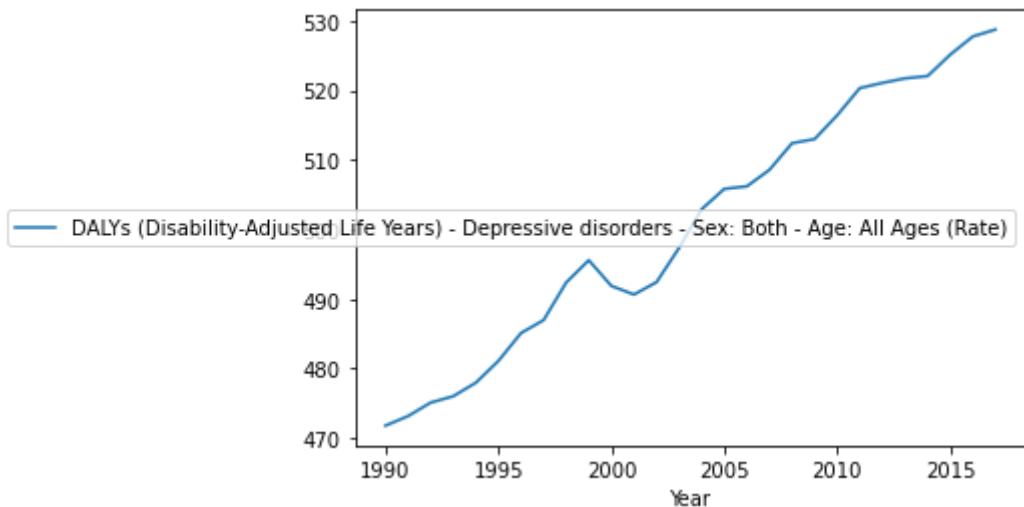
```
In [1504]: df39.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1504]: <AxesSubplot:xlabel='Year'>
```



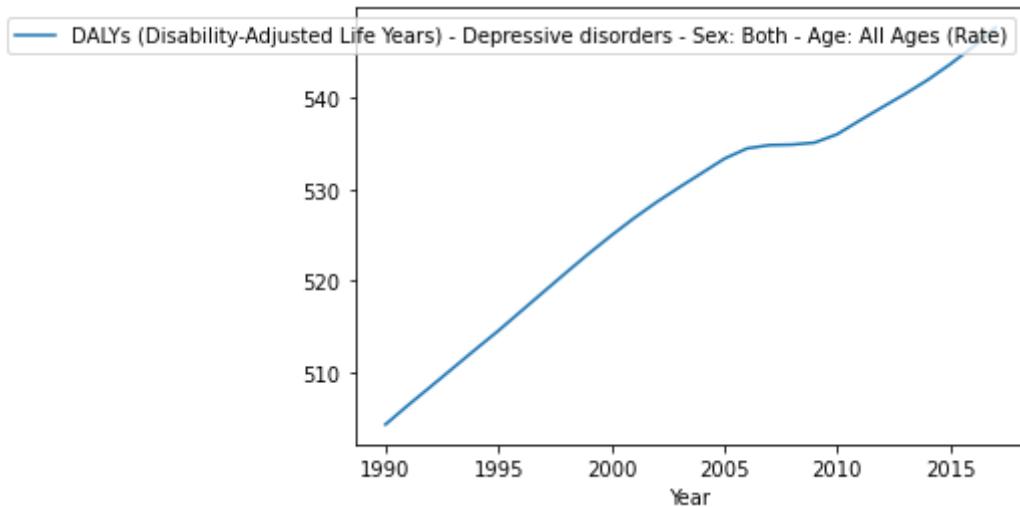
```
In [1505]: df39.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1505]: <AxesSubplot:xlabel='Year'>
```



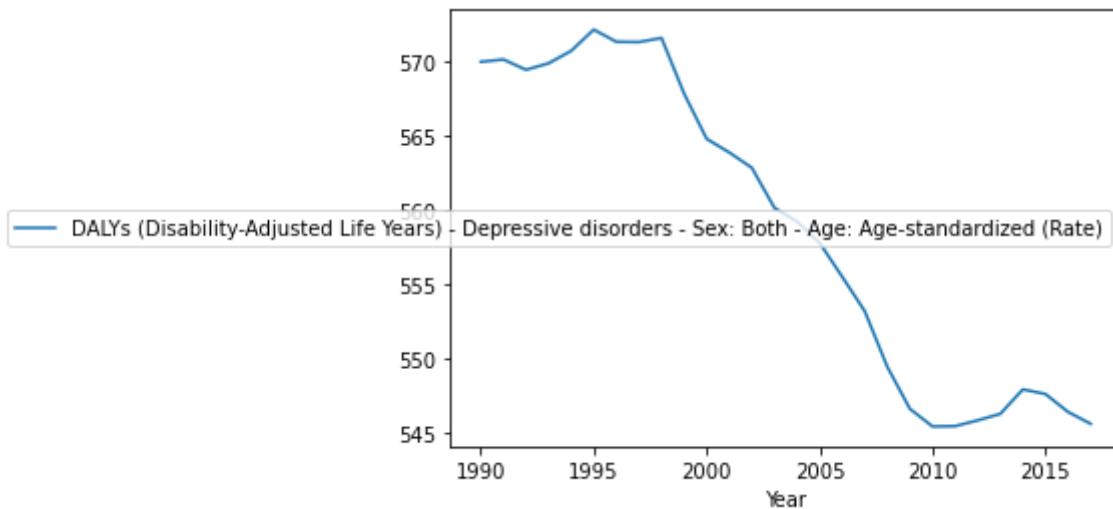
```
In [1506]: df39.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1506]: <AxesSubplot:xlabel='Year'>
```



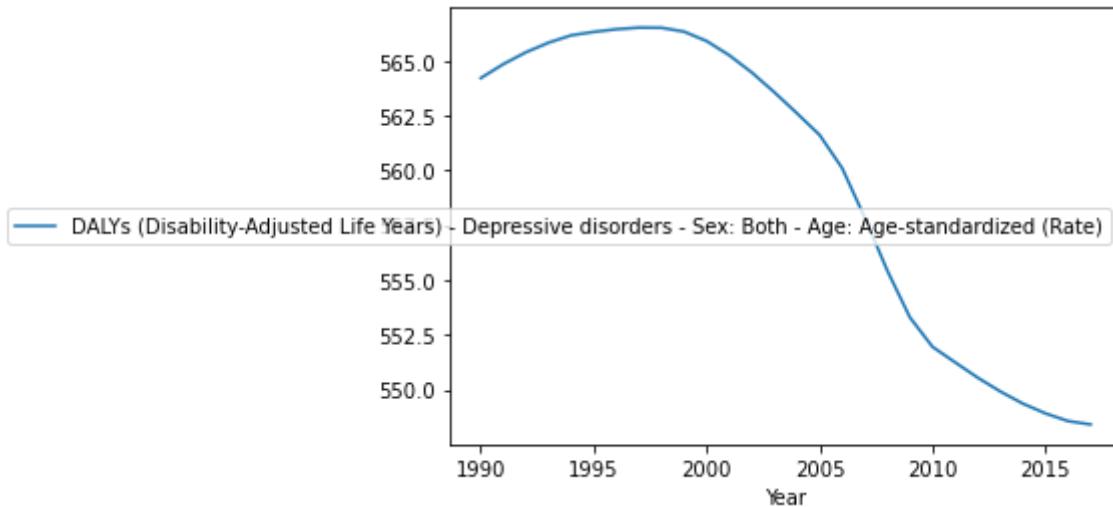
```
In [1507]: df39.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1507]: <AxesSubplot:xlabel='Year'>
```



```
In [1508]: df39.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1508]: <AxesSubplot:xlabel='Year'>
```



In [1509]: df51.info()
df51.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)
r) 6468 non-null   float64
4   Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)
6468 non-null   float64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1509]:

	Entity	Code	Year	Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)	Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)
0	Afghanistan	AFG	1990	138546.239872	179889.573799
1	Afghanistan	AFG	1991	142794.361829	186250.412127
2	Afghanistan	AFG	1992	165378.586277	217165.986617
3	Afghanistan	AFG	1993	190374.620067	250006.887326
4	Afghanistan	AFG	1994	197485.698675	259430.946814

Checking for missing values:

In [1510]: missing = pd.concat([df51.isnull().sum(), 100 * df51.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1510]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
	Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)	0	0.000000
	Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)	0	0.000000
	Code	980	15.151515

In [1511]: v1='Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)'
v2='Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)

In [1512]: df51.describe()

Out[1512]:

	Year	Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)	Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)
count	6468.000000	6.468000e+03	6.468000e+03
mean	2003.500000	1.700116e+06	2.707247e+06
std	8.078372	6.724542e+06	1.062258e+07
min	1990.000000	4.129049e+02	5.178167e+02
25%	1996.750000	2.967284e+04	4.429027e+04
50%	2003.500000	1.038870e+05	1.688615e+05
75%	2010.250000	5.065292e+05	8.137806e+05
max	2017.000000	1.028776e+08	1.615780e+08

In [1513]: df51.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1513]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)
6383	World	2017	1.028776e+08
6382	World	2016	1.012653e+08
6381	World	2015	9.967023e+07
6380	World	2014	9.811626e+07
6379	World	2013	9.656165e+07

In [1514]: df51.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1514]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)
6383	World	2017	1.615780e+08
6382	World	2016	1.591196e+08
6381	World	2015	1.567037e+08
6380	World	2014	1.543061e+08
6379	World	2013	1.518796e+08

```
In [1515]: df51.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

```
/opt/anaconda3/lib/python3.8/site-packages/IPython/core/displayhook.py:27
5: UserWarning: Output cache limit (currently 1000 entries) hit.
  Flushing oldest 200 entries.
  warn('Output cache limit (currently {sz} entries) hit.\n'
```

```
Out[1515]:
```

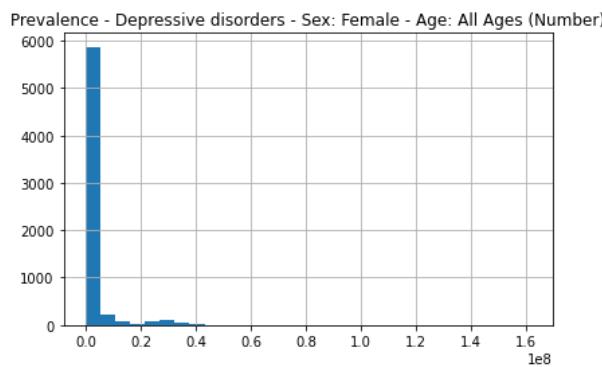
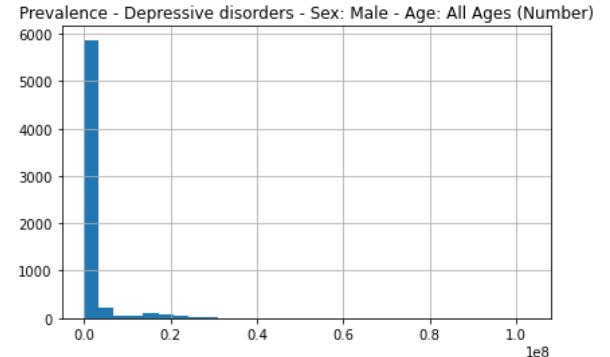
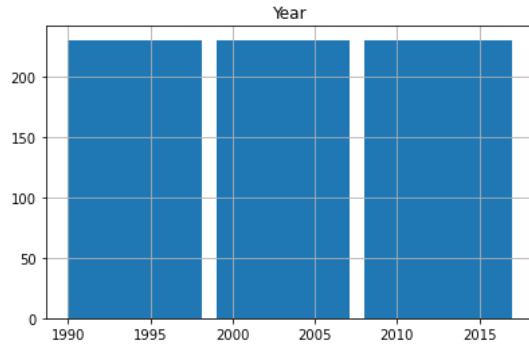
	Entity	Year	Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)
3584	Marshall Islands	1990	412.904915
3585	Marshall Islands	1991	427.264614
3586	Marshall Islands	1992	441.664009
3587	Marshall Islands	1993	455.837882
3588	Marshall Islands	1994	469.305070

```
In [1516]: df51.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()
```

```
Out[1516]:
```

	Entity	Year	Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)
3584	Marshall Islands	1990	517.816721
3585	Marshall Islands	1991	533.710544
3586	Marshall Islands	1992	550.102916
3587	Marshall Islands	1993	566.371792
4256	Northern Mariana Islands	1990	579.670562

```
In [1517]: df51.hist(bins=30, figsize=(15,10))  
plt.subplots_adjust(hspace=0.5);
```



```
In [1518]: df51.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1518]: Entity
Marshall Islands      559.158513
American Samoa        568.172664
Antigua and Barbuda   715.636284
Northern Mariana Islands 718.429514
Dominica              726.526117
Name: Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1519]: df51.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1519]: Entity
Marshall Islands      689.875514
American Samoa        771.524111
Northern Mariana Islands 851.091927
Dominica              1065.685414
Antigua and Barbuda   1256.343050
Name: Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number), dtype: float64
```

```
In [1520]: df51.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1520]: Entity
South Asia            1.866860e+07
East Asia              1.906103e+07
Middle SDI             2.212859e+07
Southeast Asia, East Asia, and Oceania 2.484054e+07
World                  8.354253e+07
Name: Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1521]: df51.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1521]: Entity
High-middle SDI        2.906128e+07
East Asia               3.029062e+07
Middle SDI              3.385294e+07
Southeast Asia, East Asia, and Oceania 3.903563e+07
World                   1.322342e+08
Name: Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number), dtype: float64
```

```
In [1522]: df51_mean = df51.groupby('Year').mean()
df51_mean.head()
```

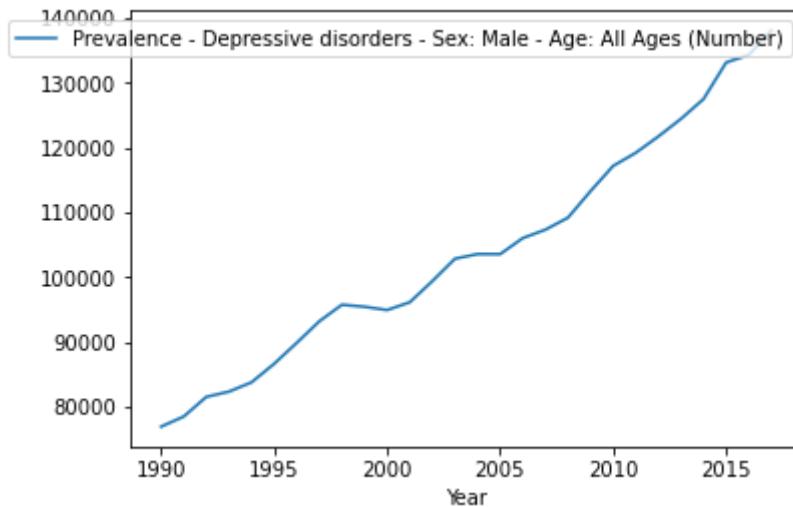
Out[1522]:

	Prevalence - Depressive disorders - Sex: Male - Age: All Ages (Number)	Prevalence - Depressive disorders - Sex: Female - Age: All Ages (Number)
--	--	--

Year		
1990	1.329713e+06	2.160287e+06
1991	1.356468e+06	2.199700e+06
1992	1.383213e+06	2.238812e+06
1993	1.409514e+06	2.276532e+06
1994	1.435698e+06	2.313926e+06

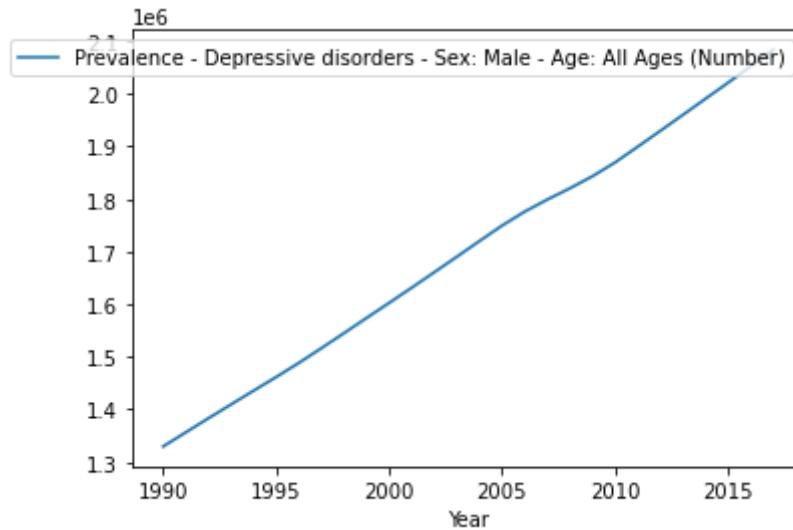
```
In [1523]: df51.groupby('Year')[v1].median().plot(legend=True)
```

Out[1523]: <AxesSubplot:xlabel='Year'>



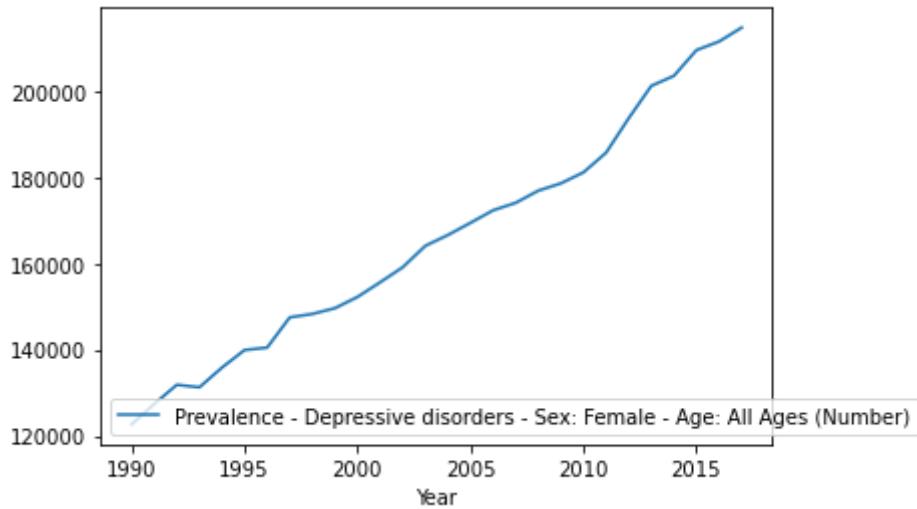
```
In [1524]: df51.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1524]: <AxesSubplot:xlabel='Year'>
```



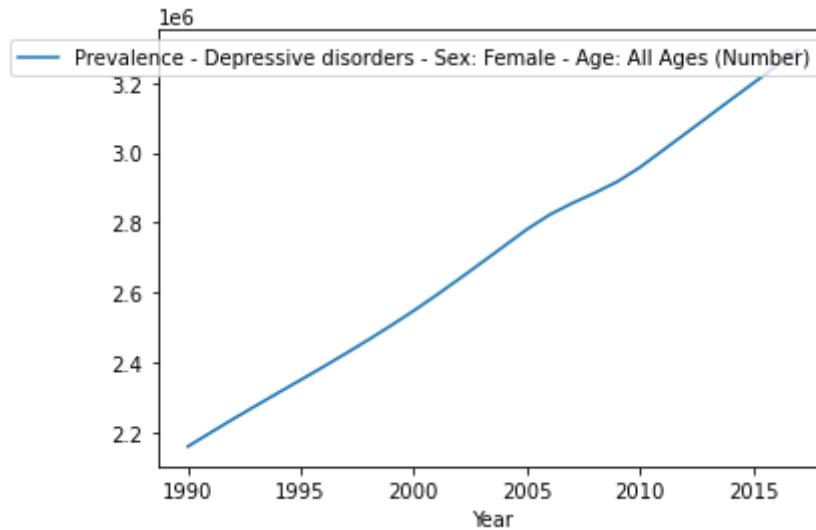
```
In [1525]: df51.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1525]: <AxesSubplot:xlabel='Year'>
```



```
In [1526]: df51.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1526]: <AxesSubplot:xlabel='Year'>
```



In [1527]: df78.info()
df78.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 13 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year        int64
6468 non-null   int64
3   Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)
6468 non-null   float64
4   Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)
6468 non-null   float64
5   Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
6   Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent)
6468 non-null   float64
7   Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)
6468 non-null   float64
8   Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)
6468 non-null   float64
9   Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent)
6468 non-null   float64
10  Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent)
6468 non-null   float64
11  Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)
6468 non-null   float64
12  Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent)
6468 non-null   float64
dtypes: float64(10), int64(1), object(2)
memory usage: 657.0+ KB
```

Out[1527]:

	Entity	Code	Year	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence
				Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)	Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)	Depressive disorders - Sex: Both - Age: All Ages (Percent)	Depressive disorders - Sex: Both - Age: 70+ years (Percent)	Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)	Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)
0	Afghanistan	AFG	1990	4.417802	1.594676	3.218871	5.202803	5.799034	3.455708
1	Afghanistan	AFG	1991	4.433524	1.588356	3.203468	5.192849	5.814828	3.451880
2	Afghanistan	AFG	1992	4.453689	1.577980	3.156559	5.176872	5.829745	3.434982
3	Afghanistan	AFG	1993	4.464517	1.577201	3.120655	5.167355	5.853060	3.420210
4	Afghanistan	AFG	1994	4.462960	1.570846	3.082179	5.157549	5.852851	3.425222

Checking for missing values:

```
In [1528]: missing = pd.concat([df78.isnull().sum(), 100 * df78.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1528]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent)		0	0.000000
Code	980	15.151515	

```
In [1529]: v1='Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)'
v2='Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)'
v3='Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent)'
v4='Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent)'
v5='Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)'
v6='Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)'
v7='Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent)'
v8='Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent)'
v9='Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)'
v10='Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent)'
```

In [1530]: df78.describe()

Out[1530]:

	Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	3.788878	1.376053	3.279958	6.137538	4.076232
std	8.078372	1.075230	0.404854	0.861814	1.501730	1.030536
min	1990.000000	1.718949	0.710318	1.805876	3.249315	2.231433
25%	1996.750000	3.071424	1.066155	2.609811	4.874158	3.248436
50%	2003.500000	3.528323	1.259560	3.031299	5.928127	3.969812
75%	2010.250000	4.309666	1.589738	3.793446	7.361696	4.725525
max	2017.000000	11.275973	3.303273	6.990122	11.531880	9.684816

In [1531]: df78.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1531]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)
2275	Greenland	1997	11.275973
2274	Greenland	1996	11.264110
2276	Greenland	1998	11.248950
2278	Greenland	2000	11.245934
2272	Greenland	1994	11.218985

In [1532]: df78.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1532]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)
2288	Greenland	2010	3.303273
2289	Greenland	2011	3.299501
2274	Greenland	1996	3.298973
2277	Greenland	1999	3.296261
2278	Greenland	2000	3.294942

In [1533]: df78.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1533]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent)
2271	Greenland	1993	6.990122
2272	Greenland	1994	6.980573
2270	Greenland	1992	6.973786
2273	Greenland	1995	6.957096
2269	Greenland	1991	6.944094

In [1534]: df78.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1534]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent)
5976	Uganda	2002	11.531880
5977	Uganda	2003	11.529887
5978	Uganda	2004	11.528105
5975	Uganda	2001	11.524187
5979	Uganda	2005	11.516791

In [1535]: df78.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1535]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)
2273	Greenland	1995	9.684816
2274	Greenland	1996	9.662793
2275	Greenland	1997	9.638313
2272	Greenland	1994	9.606278
2276	Greenland	1998	9.595571

In [1536]: df78.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1536]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)
2273	Greenland	1995	9.053459
2276	Greenland	1998	9.017674
2275	Greenland	1997	9.015748
2277	Greenland	1999	8.998987
2272	Greenland	1994	8.985850

In [1537]: df78.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1537]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent)
2274	Greenland	1996	10.377815
2273	Greenland	1995	10.366052
2272	Greenland	1994	10.346647
2275	Greenland	1997	10.321711
2276	Greenland	1998	10.307589

In [1538]: df78.sort_values(by=v8, ascending=False)[['Entity', 'Year', v8]].head()

Out[1538]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent)
5974	Uganda	2000	9.777747
5975	Uganda	2001	9.772732
5976	Uganda	2002	9.764527
5977	Uganda	2003	9.747140
5978	Uganda	2004	9.720096

In [1539]: df78.sort_values(by=v9, ascending=False)[['Entity', 'Year', v9]].head()

Out[1539]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)
2273	Greenland	1995	6.602754
2274	Greenland	1996	6.601250
2275	Greenland	1997	6.598258
2276	Greenland	1998	6.586734
2272	Greenland	1994	6.582469

In [1540]: df78.sort_values(by=v10, ascending=False)[['Entity', 'Year', v10]].head()

Out[1540]:

Entity Year Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent)				
2273	Greenland	1995		9.626314
2272	Greenland	1994		9.617917
2274	Greenland	1996		9.596353
2271	Greenland	1993		9.575569
2275	Greenland	1997		9.567272

In [1541]: df78.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1541]:

Entity Year Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)				
3907	Myanmar	2005		1.718949
3906	Myanmar	2004		1.719932
3904	Myanmar	2002		1.723394
3911	Myanmar	2009		1.723564
3905	Myanmar	2003		1.724097

In [1542]: df78.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1542]:

Entity Year Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)				
3902	Myanmar	2000		0.710318
3904	Myanmar	2002		0.711563
3901	Myanmar	1999		0.711596
3900	Myanmar	1998		0.712356
3903	Myanmar	2001		0.712798

In [1543]: df78.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1543]:

Entity Year Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent)				
3892	Myanmar	1990		1.805876
3893	Myanmar	1991		1.818176
3894	Myanmar	1992		1.829397
3895	Myanmar	1993		1.842453
3896	Myanmar	1994		1.853443

In [1544]: df78.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1544]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent)
789	Brunei	1995	3.249315
790	Brunei	1996	3.252276
788	Brunei	1994	3.253235
791	Brunei	1997	3.254807
787	Brunei	1993	3.259755

In [1545]: df78.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1545]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)
3911	Myanmar	2009	2.231433
3913	Myanmar	2011	2.233211
3912	Myanmar	2010	2.233687
3915	Myanmar	2013	2.235943
3914	Myanmar	2012	2.237961

In [1546]: df78.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[1546]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)
3904	Myanmar	2002	1.497964
3906	Myanmar	2004	1.498129
3909	Myanmar	2007	1.498861
3905	Myanmar	2003	1.499583
3902	Myanmar	2000	1.500277

In [1547]: df78.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[1547]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent)
3912	Myanmar	2010	1.946369
3913	Myanmar	2011	1.946504
3911	Myanmar	2009	1.947617
3910	Myanmar	2008	1.948577
3905	Myanmar	2003	1.950189

In [1548]: df78.sort_values(by=v8, ascending=True)[['Entity', 'Year', v8]].head()

Out[1548]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent)
805	Brunei	2011	3.248216
804	Brunei	2010	3.248593
803	Brunei	2009	3.248751
802	Brunei	2008	3.253359
806	Brunei	2012	3.253996

In [1549]: df78.sort_values(by=v9, ascending=True)[['Entity', 'Year', v9]].head()

Out[1549]:

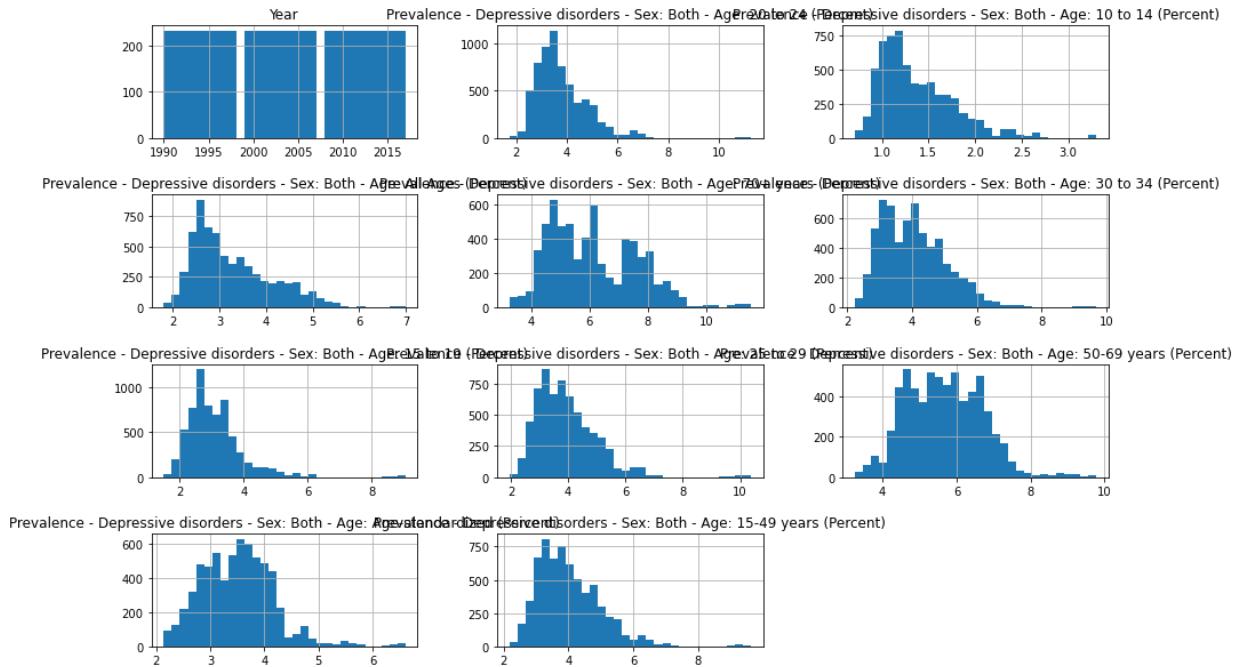
	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)
28	Albania	1990	2.139903
29	Albania	1991	2.141201
30	Albania	1992	2.143395
31	Albania	1993	2.145263
32	Albania	1994	2.150070

In [1550]: df78.sort_values(by=v10, ascending=True)[['Entity', 'Year', v10]].head()

Out[1550]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent)
3892	Myanmar	1990	2.177050
3893	Myanmar	1991	2.187027
3894	Myanmar	1992	2.192575
3895	Myanmar	1993	2.202094
3896	Myanmar	1994	2.208822

```
In [1551]: df78.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1552]: df78.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1552]: Entity
Myanmar      1.744484
Taiwan       2.133104
Indonesia    2.309471
Philippines   2.385798
Romania      2.430961
Name: Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1553]: df78.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1553]: Entity
Myanmar      0.722022
Poland       0.752656
India        0.847983
Central Europe 0.865900
Slovakia     0.867735
Name: Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1554]: df78.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1554]: Entity
Myanmar      1.996869
Mali         2.032776
Tajikistan   2.058413
Honduras     2.104790
Peru         2.107406
Name: Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [1555]: df78.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1555]: Entity
Brunei       3.306077
Canada        3.432255
New Zealand   3.569477
Japan          3.781599
High-income Asia Pacific 3.870936
Name: Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [1556]: df78.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1556]: Entity
Myanmar      2.250132
Albania       2.437914
North Macedonia 2.535215
Poland        2.540740
Romania       2.543456
Name: Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [1557]: df78.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1557]: Entity
Myanmar      1.516286
Taiwan        1.784620
Poland        1.852485
Romania       1.967061
Slovakia     1.973584
Name: Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [1558]: df78.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1558]: Entity
Myanmar          1.964888
Albania          2.450240
Indonesia        2.459383
Romania          2.475404
North Macedonia  2.492207
Name: Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1559]: df78.groupby('Entity')[v8].mean().sort_values().head()
```

```
Out[1559]: Entity
Brunei           3.270322
Albania          3.574977
Japan            3.728055
High-income Asia Pacific 3.752097
Colombia         3.766970
Name: Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent), dtype: float64
```

```
In [1560]: df78.groupby('Entity')[v9].mean().sort_values().head()
```

```
Out[1560]: Entity
Albania          2.190442
Myanmar          2.252113
Poland           2.284375
Colombia         2.311194
Peru              2.317500
Name: Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1561]: df78.groupby('Entity')[v10].mean().sort_values().head()
```

```
Out[1561]: Entity
Myanmar          2.244731
Albania          2.467109
North Macedonia 2.579064
Peru              2.594309
Romania          2.619816
Name: Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

```
In [1562]: df78.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1562]: Entity
United States    6.610652
Australasia      6.884311
Finland          6.885881
Australia        6.957891
Greenland         10.958194
Name: Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1563]: df78.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1563]: Entity
North America    2.423633
United States    2.462045
Portugal         2.482565
Australia        2.512964
Greenland        3.257075
Name: Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1564]: df78.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1564]: Entity
Morocco          5.242925
Ukraine          5.308271
Portugal          5.415400
Finland           5.765739
Greenland         6.830883
Name: Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [1565]: df78.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1565]: Entity
Gambia            9.240600
Ukraine          9.714017
Estonia           9.806835
Lesotho           11.033590
Uganda            11.279064
Name: Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent), dtype: float64
```

```
In [1566]: df78.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1566]: Entity
Australia         6.259872
Iran              6.748484
Finland           6.823468
Morocco           7.335285
Greenland         9.285031
Name: Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [1567]: df78.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1567]: Entity
Australasia       5.551051
Australia         5.652849
North America     5.783778
United States     5.874488
Greenland         8.865492
Name: Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [1568]: df78.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1568]: Entity
Australasia    6.434570
Australia      6.531959
Morocco        6.596658
Finland         6.792700
Greenland       9.995732
Name: Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1569]: df78.groupby('Entity')[v8].mean().sort_values().tail()
```

```
Out[1569]: Entity
Estonia        7.551284
Ukraine        7.827788
Morocco         8.710710
Lesotho         8.953203
Uganda          9.403496
Name: Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent), dtype: float64
```

```
In [1570]: df78.groupby('Entity')[v9].mean().sort_values().tail()
```

```
Out[1570]: Entity
Finland         5.183080
Uganda          5.227817
Lesotho          5.423811
Morocco          5.610298
Greenland        6.459109
Name: Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1571]: df78.groupby('Entity')[v10].mean().sort_values().tail()
```

```
Out[1571]: Entity
North America   6.349330
United States   6.440084
Finland         6.779927
Morocco          6.842995
Greenland        9.320678
Name: Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent), dtype: float64
```

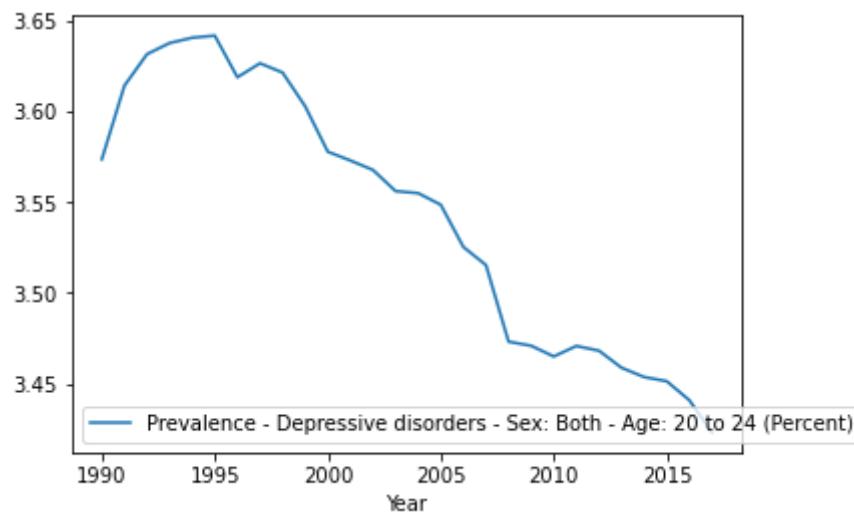
```
In [1580]: df78_mean = df78.groupby('Year').mean()
df78_mean.head()
```

Out[1580]:

	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence
	Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)	Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)	Depressive disorders - Sex: Both - Age: All Ages (Percent)	Depressive disorders - Sex: Both - Age: 70+ years (Percent)	Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)	Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)	Depressive disorders - Sex: Both - Age: 25 to 29 (Percent)	Depressive disorders - Sex: Both - Age: 50+ years (Percent)
Year								
1990	3.818881	1.359294	3.110991	6.175283	4.137331	3.069637	3.968893	5.6888
1991	3.833584	1.357751	3.124947	6.177954	4.140805	3.075707	3.976802	5.6948
1992	3.844902	1.357220	3.138378	6.180622	4.142956	3.080654	3.982677	5.7008
1993	3.854340	1.357125	3.151890	6.182858	4.144735	3.084604	3.986352	5.7068
1994	3.861149	1.357328	3.165363	6.184569	4.145596	3.088210	3.989088	5.7118

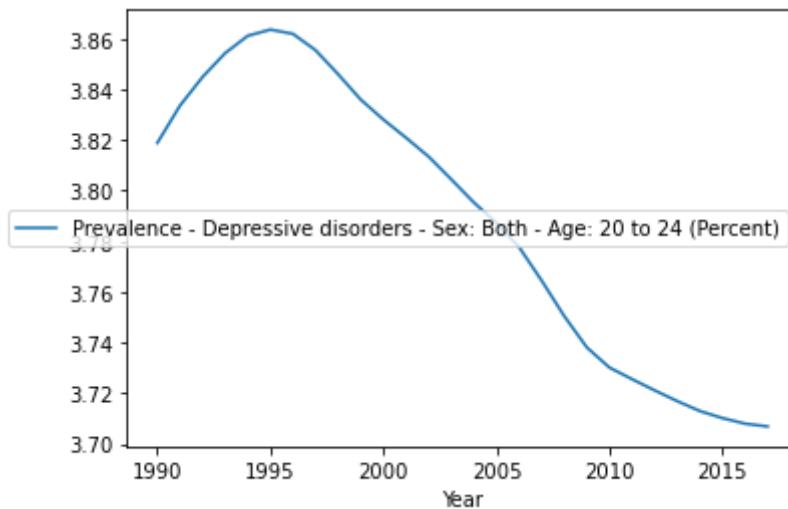
```
In [1573]: df78.groupby('Year')[v1].median().plot(legend=True)
```

Out[1573]: <AxesSubplot:xlabel='Year'>



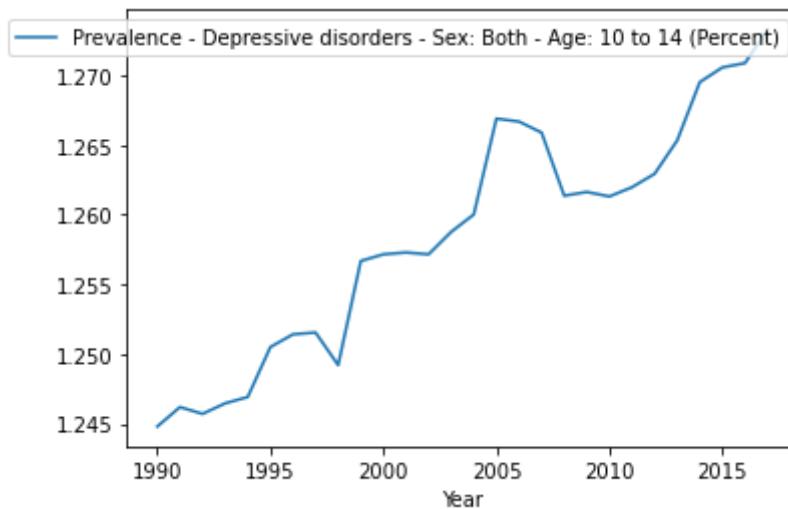
```
In [1574]: df78.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1574]: <AxesSubplot:xlabel='Year'>
```



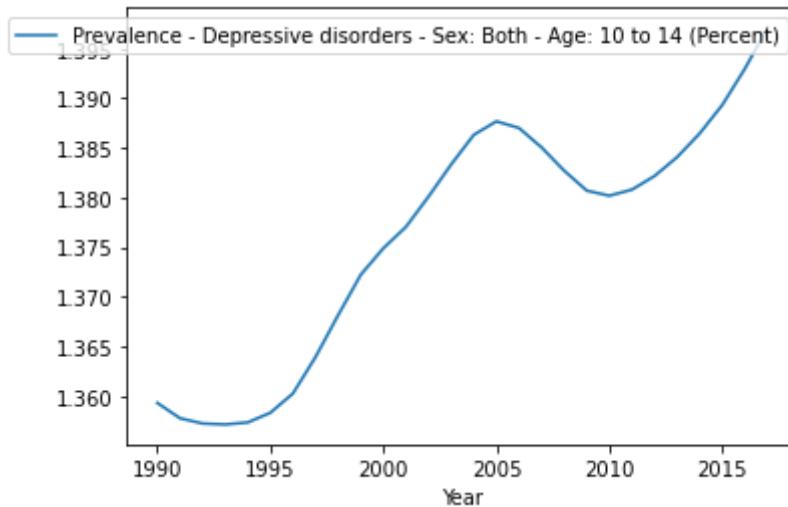
```
In [1575]: df78.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1575]: <AxesSubplot:xlabel='Year'>
```



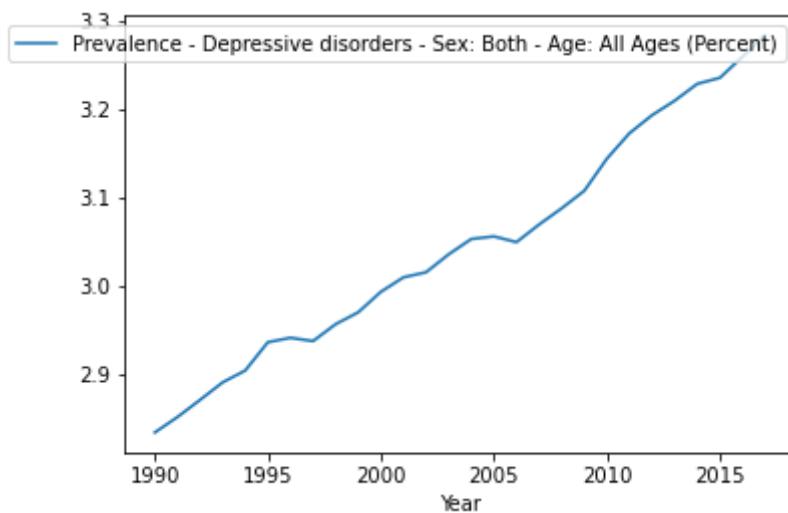
```
In [1576]: df78.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1576]: <AxesSubplot:xlabel='Year'>
```



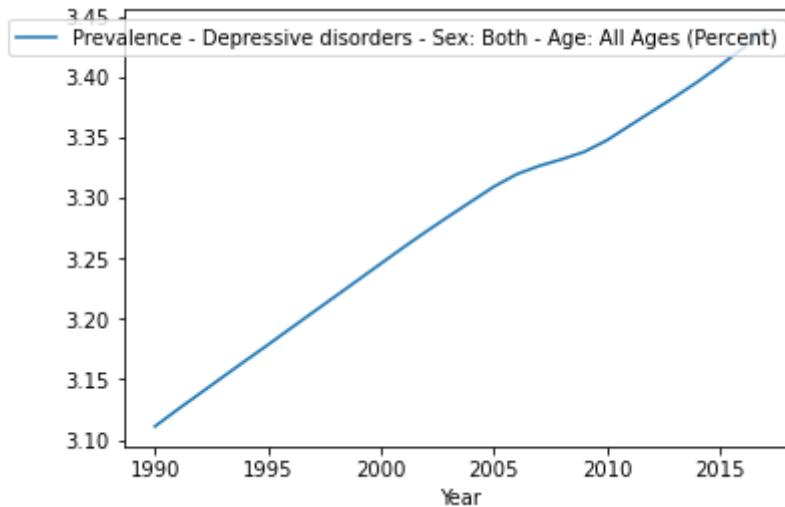
```
In [1577]: df78.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1577]: <AxesSubplot:xlabel='Year'>
```



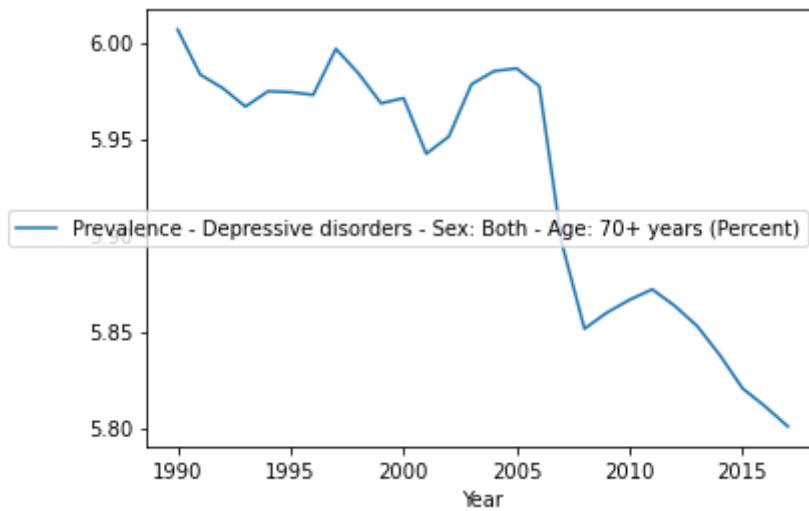
```
In [1578]: df78.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1578]: <AxesSubplot:xlabel='Year'>
```



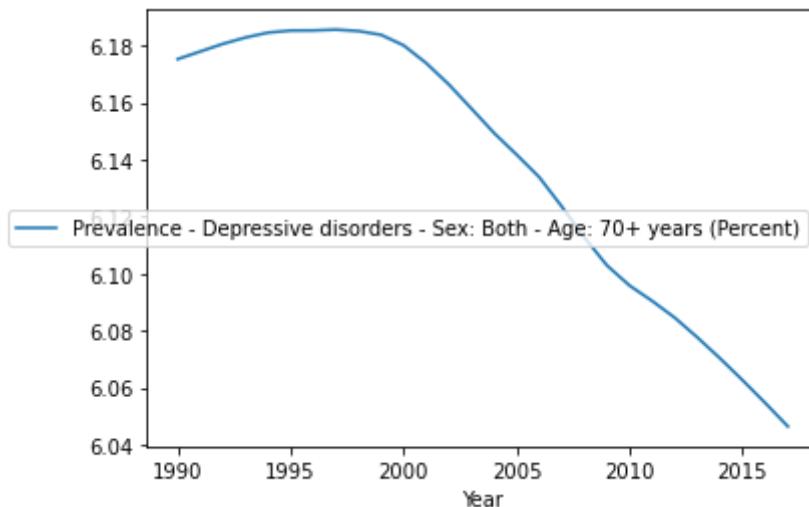
```
In [1579]: df78.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1579]: <AxesSubplot:xlabel='Year'>
```



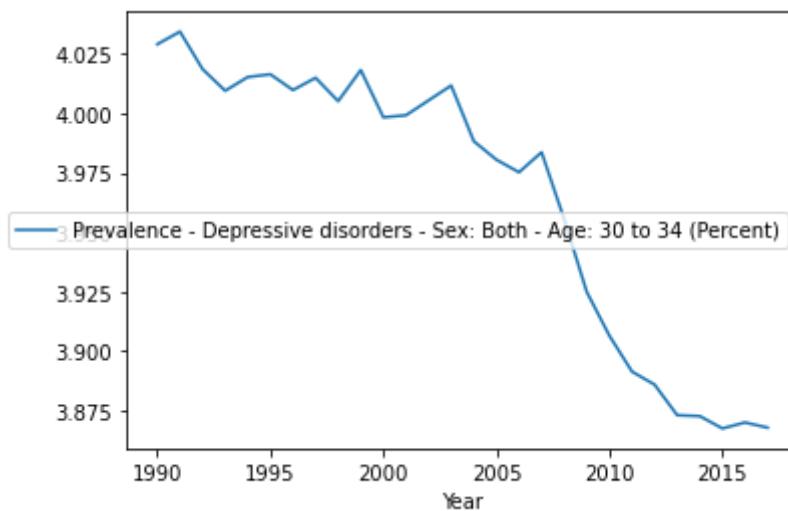
```
In [1581]: df78.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1581]: <AxesSubplot:xlabel='Year'>
```



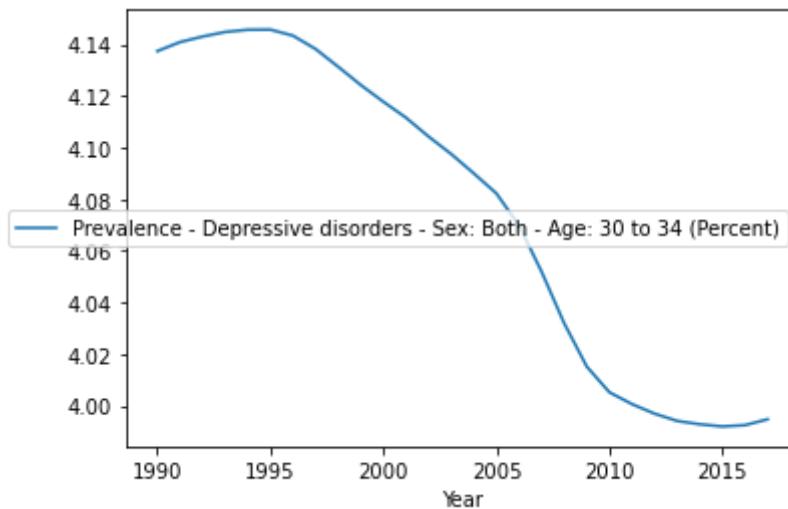
```
In [1582]: df78.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1582]: <AxesSubplot:xlabel='Year'>
```



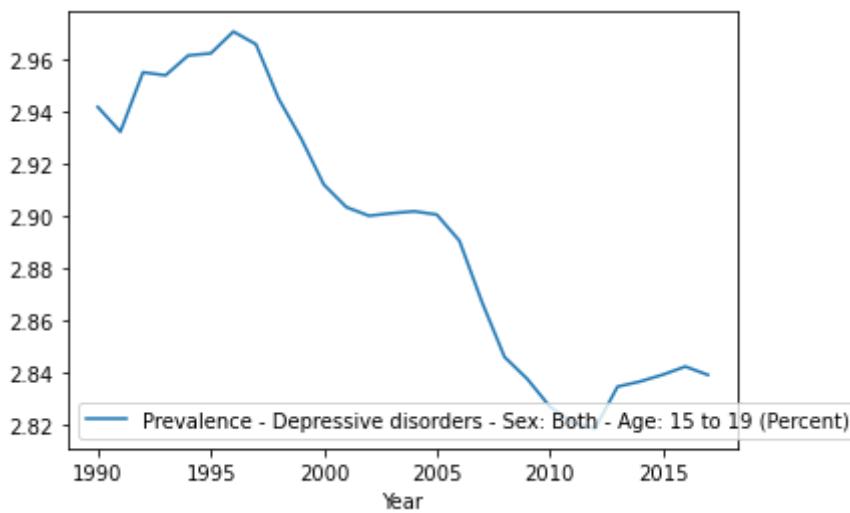
```
In [1583]: df78.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1583]: <AxesSubplot:xlabel='Year'>
```



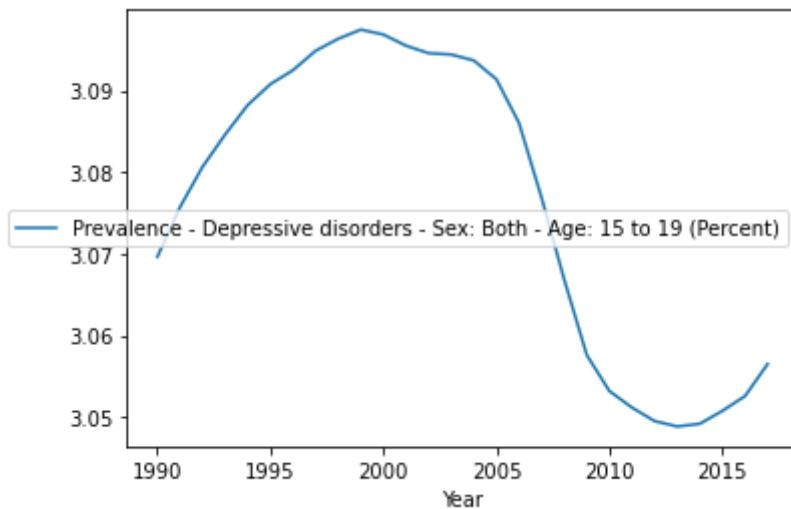
```
In [1584]: df78.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1584]: <AxesSubplot:xlabel='Year'>
```



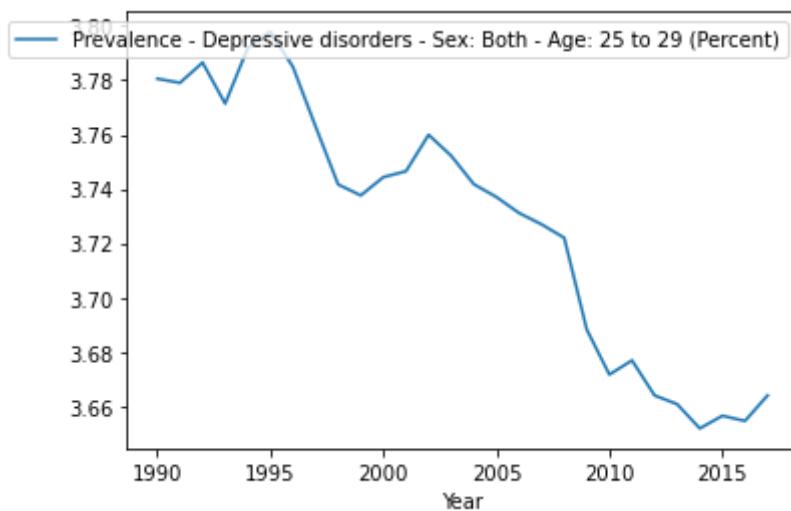
```
In [1585]: df78.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1585]: <AxesSubplot:xlabel='Year'>
```



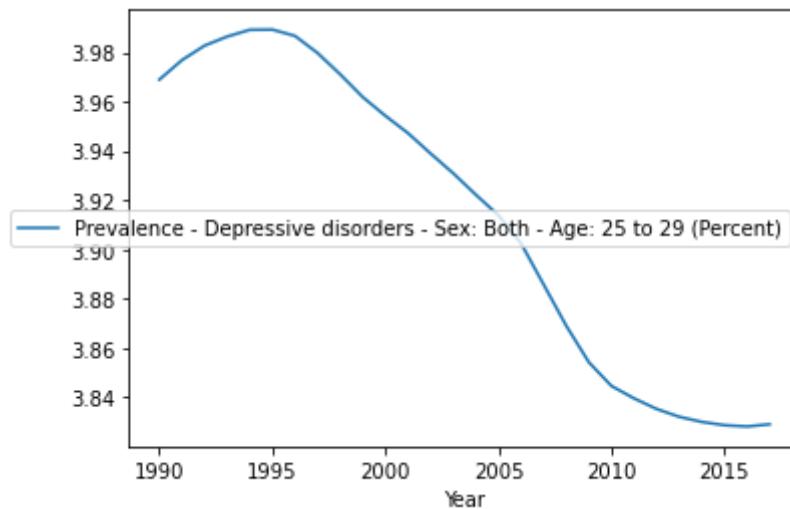
```
In [1586]: df78.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1586]: <AxesSubplot:xlabel='Year'>
```



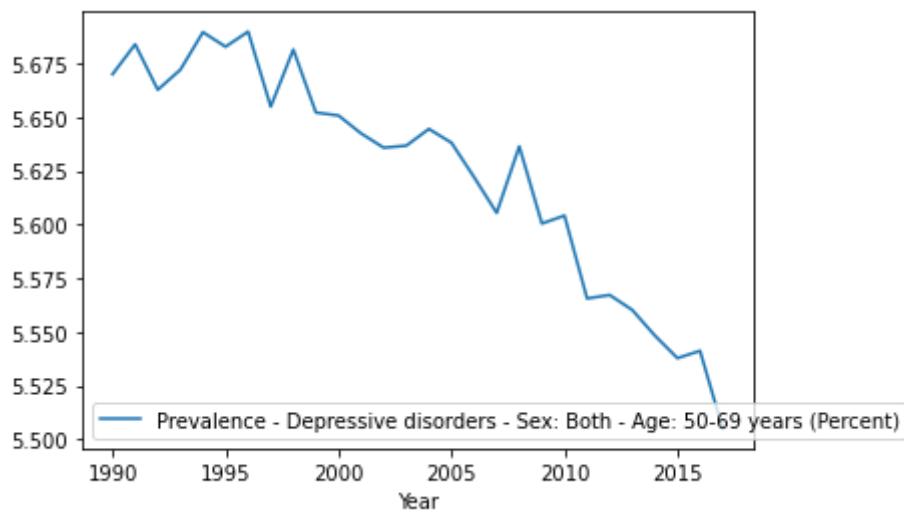
```
In [1587]: df78.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1587]: <AxesSubplot:xlabel='Year'>
```



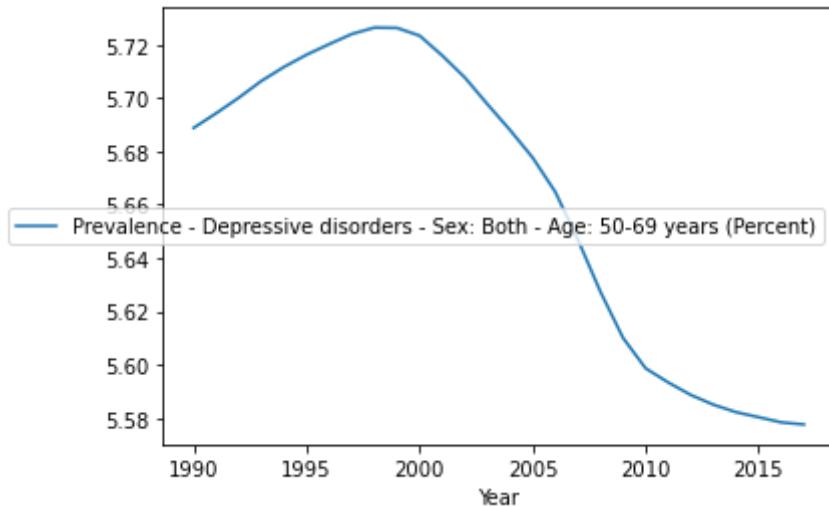
```
In [1588]: df78.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[1588]: <AxesSubplot:xlabel='Year'>
```



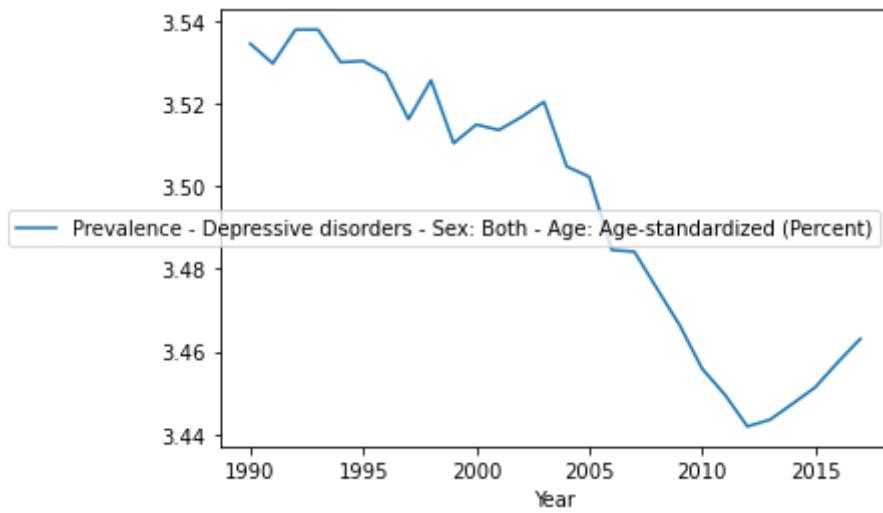
```
In [1589]: df78.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[1589]: <AxesSubplot:xlabel='Year'>
```



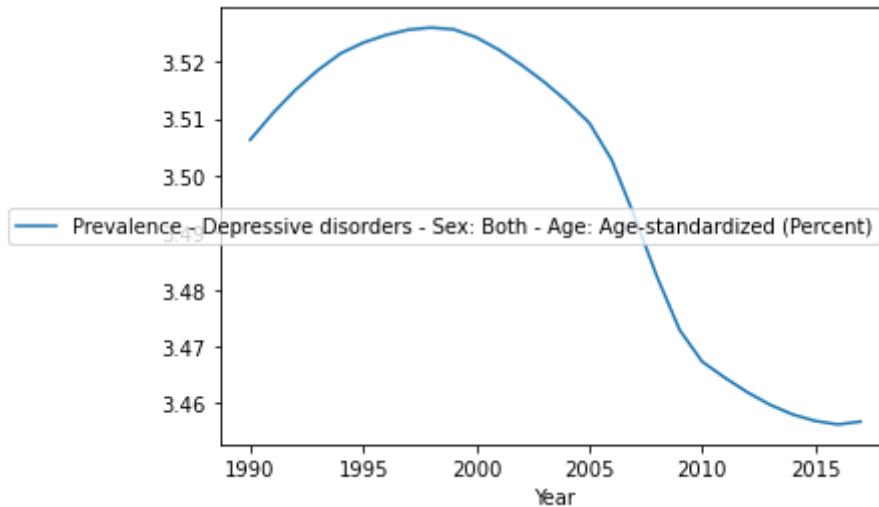
```
In [1590]: df78.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[1590]: <AxesSubplot:xlabel='Year'>
```



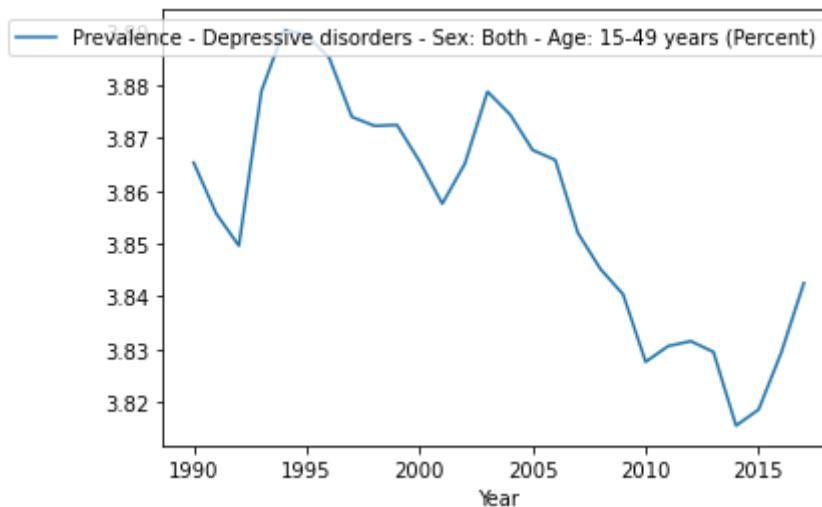
```
In [1591]: df78.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[1591]: <AxesSubplot:xlabel='Year'>
```



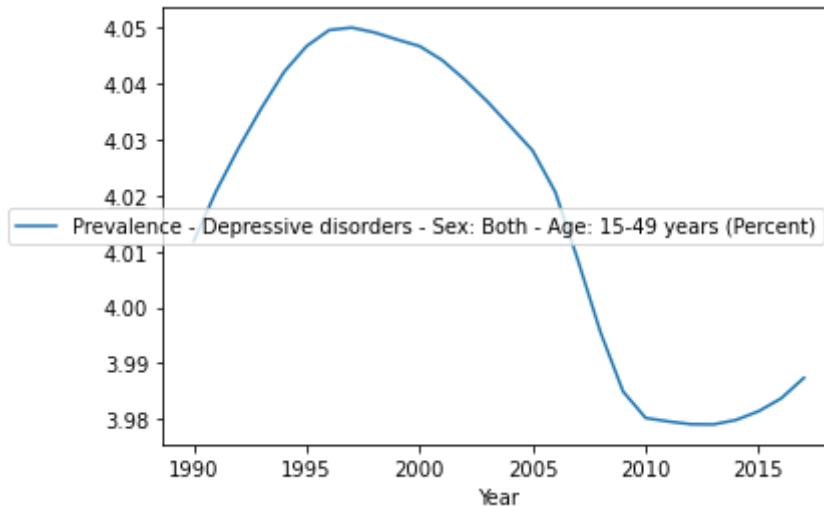
```
In [1592]: df78.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[1592]: <AxesSubplot:xlabel='Year'>
```



```
In [1593]: df78.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[1593]: <AxesSubplot:xlabel='Year'>
```



Eating Disorders

In [1594]: df16.info()
df16.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 9 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Bot
h - Age: All Ages (Rate)           6468 non-null   float64
4   DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Bot
h - Age: 70+ years (Rate)          6468 non-null   int64
5   DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Bot
h - Age: 50-69 years (Rate)        6468 non-null   int64
6   DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Bot
h - Age: Age-standardized (Rate)  6468 non-null   float64
7   DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Bot
h - Age: 15-49 years (Rate)        6468 non-null   float64
8   DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Bot
h - Age: 5-14 years (Rate)         6468 non-null   float64
dtypes: float64(4), int64(3), object(2)
memory usage: 454.9+ KB
```

Out[1594]:

	Entity	Code	Year	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age- standardized (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)
0	Afghanistan	AFG	1990	17.487272	0	0	20.754767	38.803801	6.607866
1	Afghanistan	AFG	1991	17.401168	0	0	20.238001	38.161588	6.301823
2	Afghanistan	AFG	1992	18.097456	0	0	19.708028	37.933909	6.039508
3	Afghanistan	AFG	1993	18.498291	0	0	19.237173	37.602625	5.801409
4	Afghanistan	AFG	1994	18.131521	0	0	18.863183	37.177799	5.684285

Checking for missing values:

```
In [1595]: missing = pd.concat([df16.isnull().sum(), 100 * df16.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1595]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age-standardized (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)		0	0.000000
Code	980	15.151515	

```
In [1596]: v1='DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both -'
v2='DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both -'
v3='DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both -'
v4='DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both -'
v5='DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both -'
v6='DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both -'
```

In [1597]: df16.describe()

Out[1597]:

Year	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age- standardized (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)
count	6468.000000	6468.000000	6468.0	6468.0	6468.000000	6468.000000
mean	2003.500000	48.836225	0.0	0.0	48.261493	91.569072
std	8.078372	29.203942	0.0	0.0	30.295637	55.255736
min	1990.000000	15.224833	0.0	0.0	15.690928	30.790943
25%	1996.750000	26.006396	0.0	0.0	25.538960	49.343837
50%	2003.500000	38.437277	0.0	0.0	37.273714	71.437059
75%	2010.250000	65.698769	0.0	0.0	59.739773	116.516397
max	2017.000000	164.615012	0.0	0.0	178.584448	337.597559

In [1598]: df16.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1598]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate)
334	Australia	2016
333	Australia	2015
335	Australia	2017
332	Australia	2014
331	Australia	2013

In [1599]: df16.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1599]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)
0	Afghanistan	1990	0
4309	Norway	2015	0
4319	Oceania	1997	0
4318	Oceania	1996	0
4317	Oceania	1995	0

In [1600]: df16.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1600]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)
0	Afghanistan	1990	0
4309	Norway	2015	0
4319	Oceania	1997	0
4318	Oceania	1996	0
4317	Oceania	1995	0

In [1601]: df16.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1601]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age-standardized (Rate)
335	Australia	2017	178.584448
334	Australia	2016	178.546085
333	Australia	2015	178.176283
332	Australia	2014	177.610418
331	Australia	2013	177.065506

In [1602]: df16.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1602]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	
335	Australia	2017	337.597559
334	Australia	2016	337.168266
333	Australia	2015	336.191620
332	Australia	2014	334.535364
331	Australia	2013	332.894977

In [1603]: df16.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1603]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)	
5424	Spain	2010	49.154747
5423	Spain	2009	49.057458
5425	Spain	2011	49.051415
5422	Spain	2008	49.000541
5426	Spain	2012	48.635102

In [1604]: df16.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1604]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate)	
3253	Liberia	1995	15.224833
3867	Mozambique	1993	15.327080
3865	Mozambique	1991	15.350980
3868	Mozambique	1994	15.353369
3866	Mozambique	1992	15.358847

In [1605]: `df16.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()`

Out[1605]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)
0	Afghanistan	1990	0
4318	Oceania	1996	0
4317	Oceania	1995	0
4316	Oceania	1994	0
4315	Oceania	1993	0

In [1606]: `df16.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()`

Out[1606]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)
0	Afghanistan	1990	0
4318	Oceania	1996	0
4317	Oceania	1995	0
4316	Oceania	1994	0
4315	Oceania	1993	0

In [1607]: `df16.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()`

Out[1607]:

	Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age-standardized (Rate)
3893	Myanmar	1991	15.690928
3894	Myanmar	1992	15.722576
3892	Myanmar	1990	15.723885
3895	Myanmar	1993	15.765951
3896	Myanmar	1994	15.789158

In [1608]: `df16.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()`

Out[1608]:

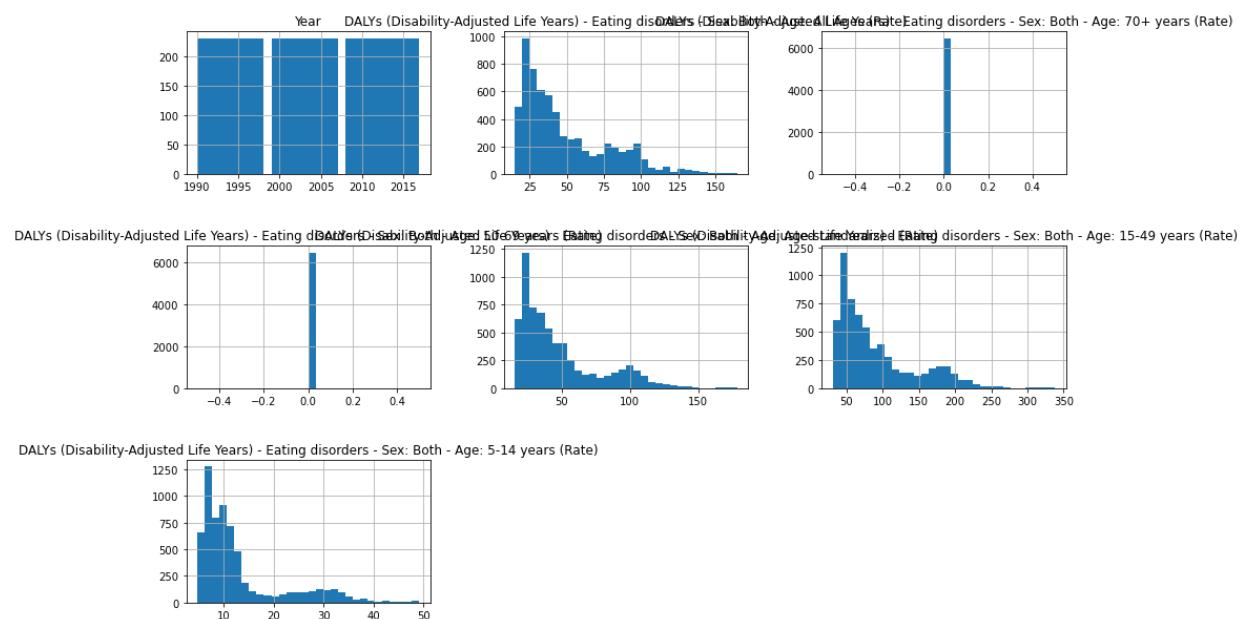
Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	
3893	Myanmar	1991	30.790943
3894	Myanmar	1992	30.818041
3897	Myanmar	1995	30.825962
3896	Myanmar	1994	30.829248
3895	Myanmar	1993	30.840212

In [1609]: `df16.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()`

Out[1609]:

Entity	Year	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)	
10	Afghanistan	2000	4.648002
9	Afghanistan	1999	4.658829
5169	Somalia	2007	4.693983
5168	Somalia	2006	4.701205
5170	Somalia	2008	4.706316

In [1610]: `df16.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [1611]: df16.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1611]: Entity
Somalia      15.812351
Liberia      17.346909
Niger        17.396366
Mozambique   17.490650
Ethiopia     17.575235
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1612]: df16.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1612]: Entity
Afghanistan      0
Nigeria          0
North Africa and Middle East 0
North America    0
North Korea      0
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate), dtype: int64
```

```
In [1613]: df16.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1613]: Entity
Afghanistan      0
Nigeria          0
North Africa and Middle East 0
North America    0
North Korea      0
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate), dtype: int64
```

```
In [1614]: df16.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1614]: Entity
Somalia      16.458345
Liberia      17.573552
Ethiopia     18.148662
Mozambique   18.199337
Democratic Republic of Congo 18.444730
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [1615]: df16.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1615]: Entity
Somalia      32.096642
Liberia      34.217422
Ethiopia     35.599156
Mozambique   35.767336
Democratic Republic of Congo 36.047674
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1616]: df16.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1616]: Entity
Somalia      5.002864
Mozambique   5.218170
Niger         5.258165
Ethiopia      5.301831
Liberia       5.344406
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

```
In [1617]: df16.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1617]: Entity
Andorra      128.523362
Brunei        129.978497
Luxembourg   131.667316
Australasia   147.579073
Australia     153.901055
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate), dtype: float64
```

```
In [1618]: df16.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1618]: Entity
Greenland     0
Grenada       0
Guam          0
Guinea        0
Zimbabwe      0
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate), dtype: int64
```

```
In [1619]: df16.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1619]: Entity
Greenland     0
Grenada       0
Guam          0
Guinea        0
Zimbabwe      0
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate), dtype: int64
```

```
In [1620]: df16.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1620]: Entity
Austria       122.478996
Spain          128.929136
Luxembourg    136.118065
Australasia    152.320885
Australia      158.445172
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age-standardized (Rate), dtype: float64
```

```
In [1621]: df16.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1621]: Entity
New Zealand      224.555911
Spain            233.450172
Luxembourg       248.975644
Australasia      286.113288
Australia        298.075217
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate), dtype: float64
```

```
In [1622]: df16.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1622]: Entity
Andorra          36.339473
Finland          36.660581
Luxembourg       38.447100
Italy             41.328041
Spain             46.643578
Name: DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate), dtype: float64
```

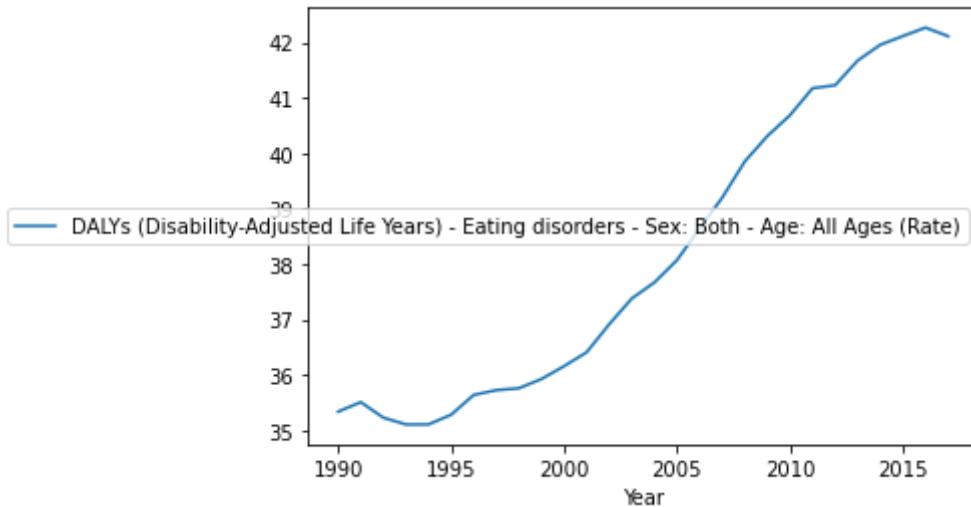
```
In [1623]: df16_mean = df16.groupby('Year').mean()
df16_mean.head()
```

```
Out[1623]:
```

DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age- standardized (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)
Year					
1990	45.919771	0	0	44.818384	86.790896
1991	46.011156	0	0	44.857442	86.786200
1992	46.092496	0	0	44.917443	86.753376
1993	46.194877	0	0	45.017861	86.763582
1994	46.312914	0	0	45.140539	86.809653

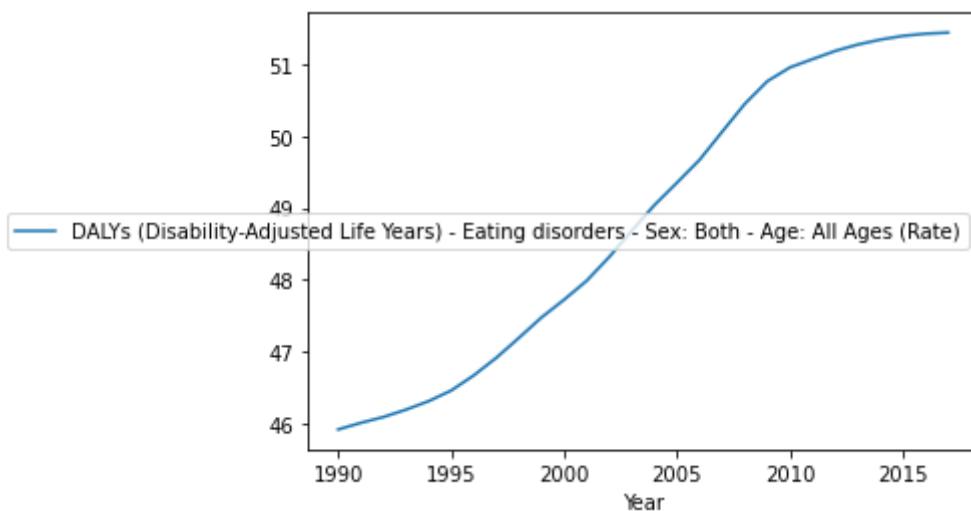
```
In [1624]: df16.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1624]: <AxesSubplot:xlabel='Year'>
```



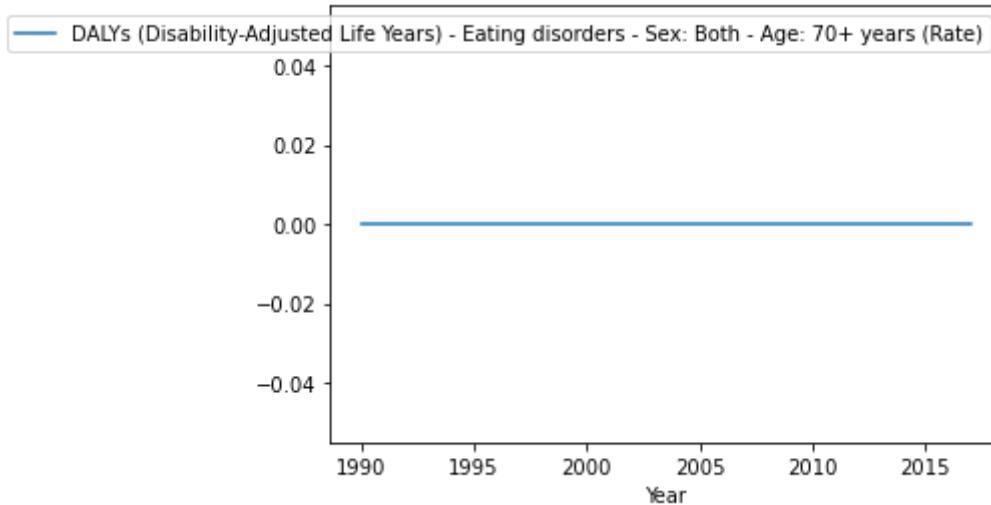
```
In [1625]: df16.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1625]: <AxesSubplot:xlabel='Year'>
```



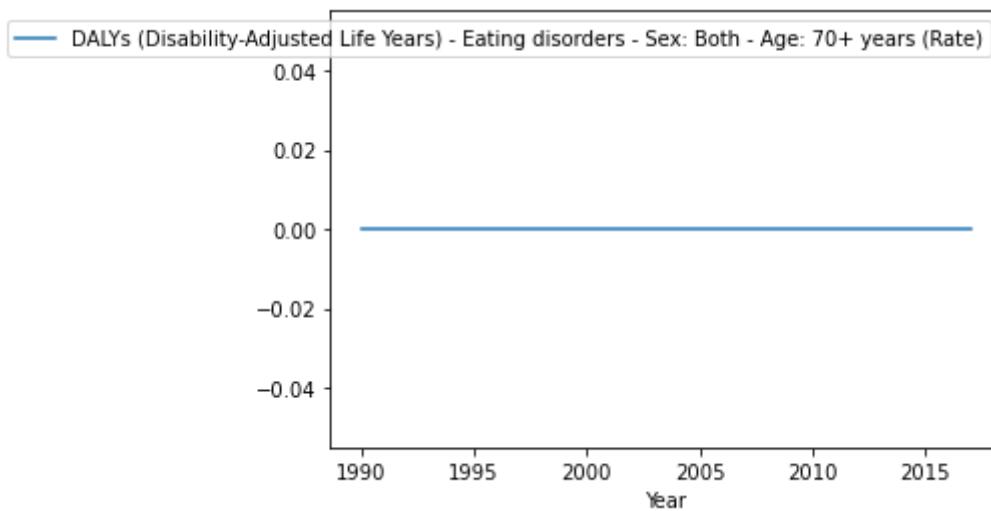
```
In [1626]: df16.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1626]: <AxesSubplot:xlabel='Year'>
```



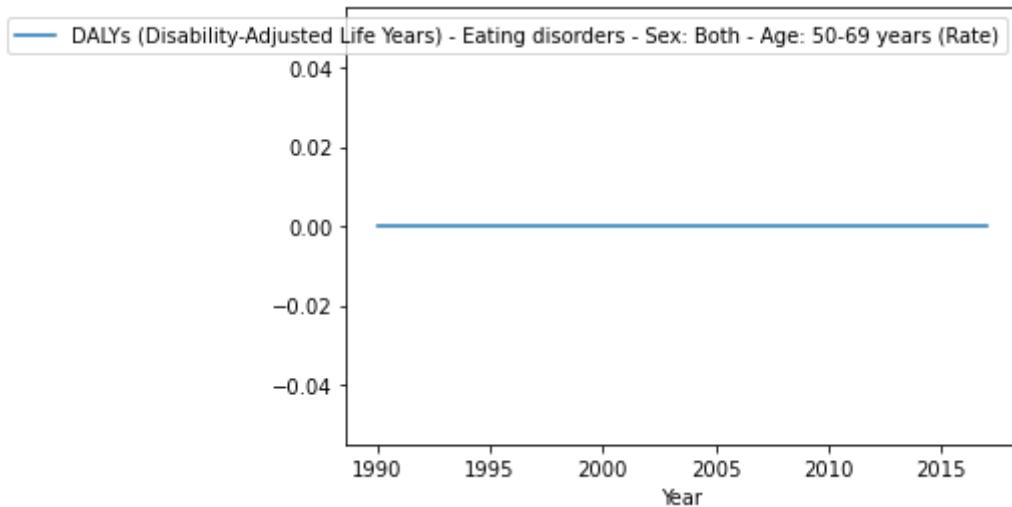
```
In [1627]: df16.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1627]: <AxesSubplot:xlabel='Year'>
```



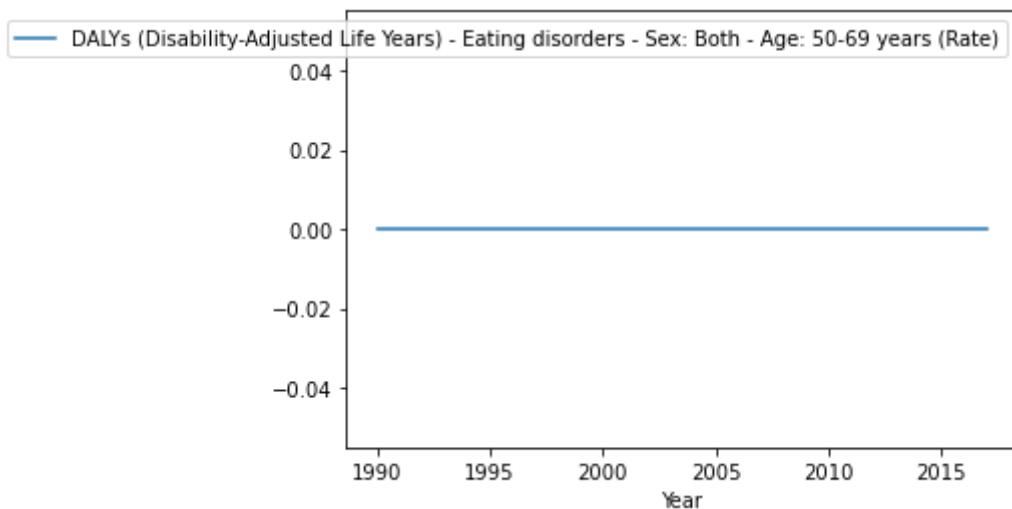
```
In [1628]: df16.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1628]: <AxesSubplot:xlabel='Year'>
```



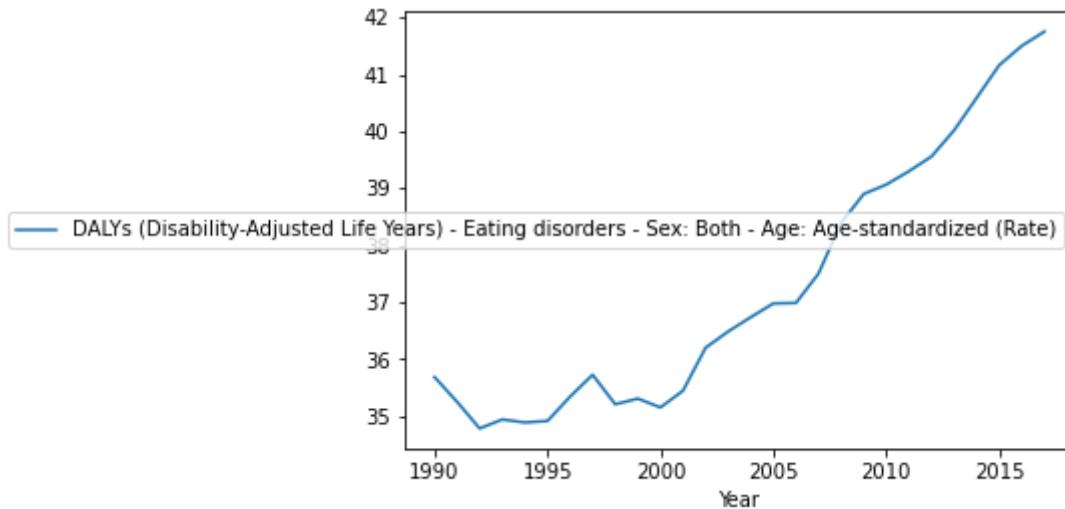
```
In [1629]: df16.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1629]: <AxesSubplot:xlabel='Year'>
```



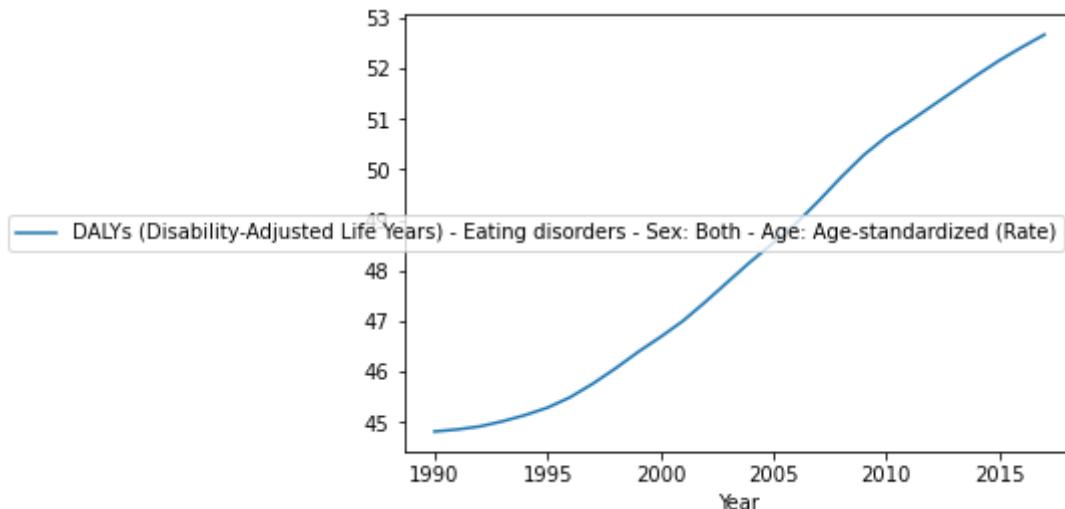
```
In [1630]: df16.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1630]: <AxesSubplot:xlabel='Year'>
```



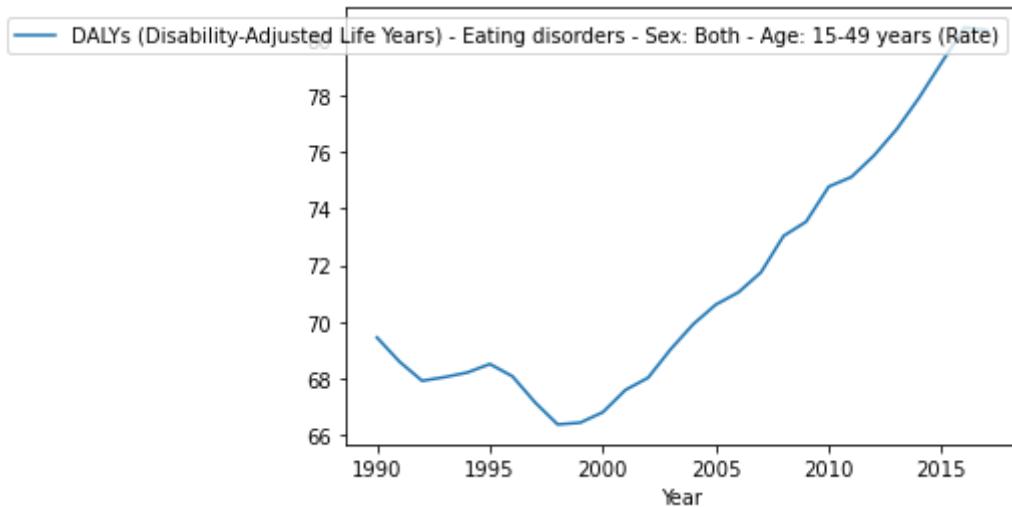
```
In [1631]: df16.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1631]: <AxesSubplot:xlabel='Year'>
```



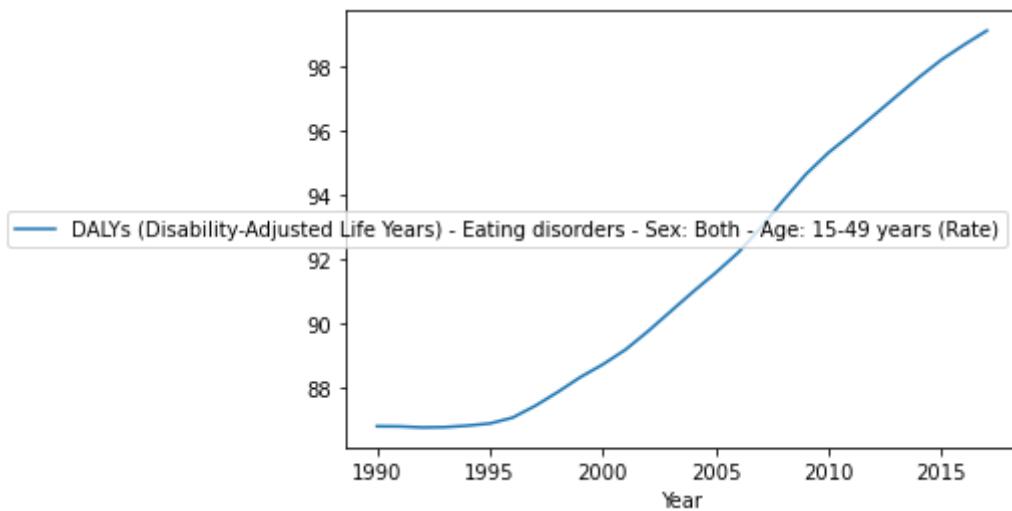
```
In [1632]: df16.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1632]: <AxesSubplot:xlabel='Year'>
```



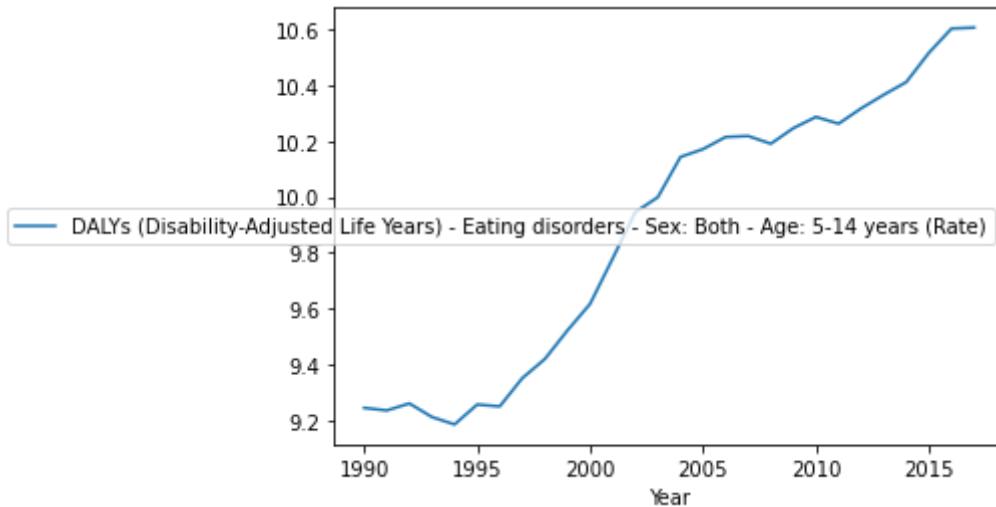
```
In [1633]: df16.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1633]: <AxesSubplot:xlabel='Year'>
```



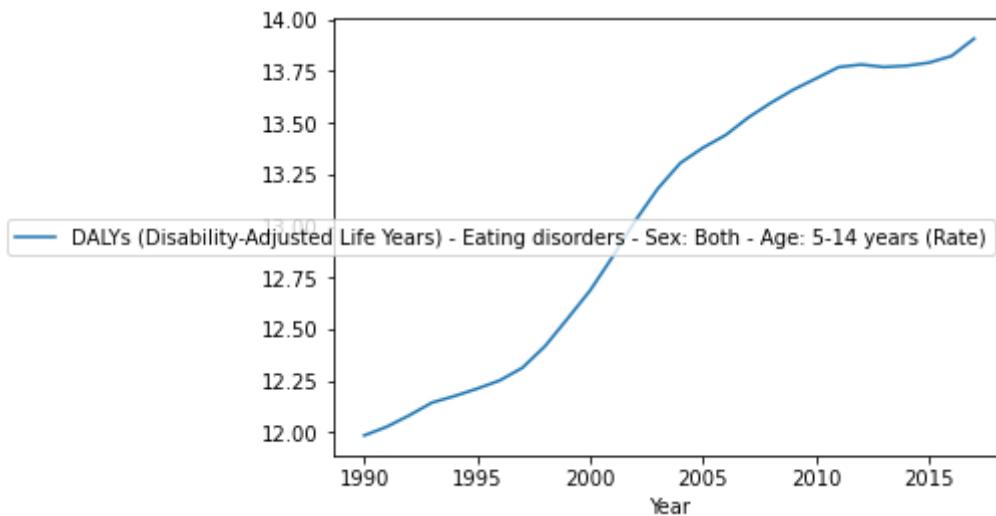
```
In [1634]: df16.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1634]: <AxesSubplot:xlabel='Year'>
```



```
In [1635]: df16.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1635]: <AxesSubplot:xlabel='Year'>
```



In [1636]: df34.info()
df34.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 4 columns):
 #   Column           Non-
 Null  Count  Dtype
 ---  --    --
 0   Entity      6468
 non-null  object
 1   Code        5488
 non-null  object
 2   Year        6468
 non-null  int64
 3   Deaths - Eating disorders - Sex: Both - Age: All Ages (Number) 6468
 non-null  float64
dtypes: float64(1), int64(1), object(2)
memory usage: 202.2+ KB
```

Out[1636]:

	Entity	Code	Year	Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	1990	0.009930
1	Afghanistan	AFG	1991	0.010376
2	Afghanistan	AFG	1992	0.012854
3	Afghanistan	AFG	1993	0.015958
4	Afghanistan	AFG	1994	0.016801

Checking for missing values:

In [1637]: missing = pd.concat([df34.isnull().sum(), 100 * df34.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1637]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)		0	0.000000
Code	980	15.151515	

In [1638]: v1='Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)'

In [1639]: df34.describe()

Out[1639]:

Year Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)		
count	6468.000000	6468.000000
mean	2003.500000	5.654176
std	8.078372	26.023061
min	1990.000000	0.000106
25%	1996.750000	0.016237
50%	2003.500000	0.100280
75%	2010.250000	0.934696
max	2017.000000	326.471123

In [1640]: df34.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1640]:

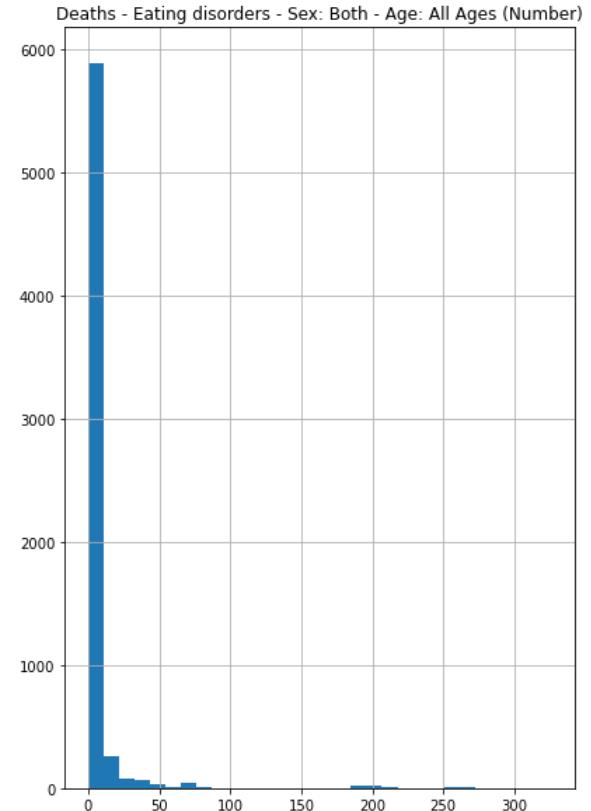
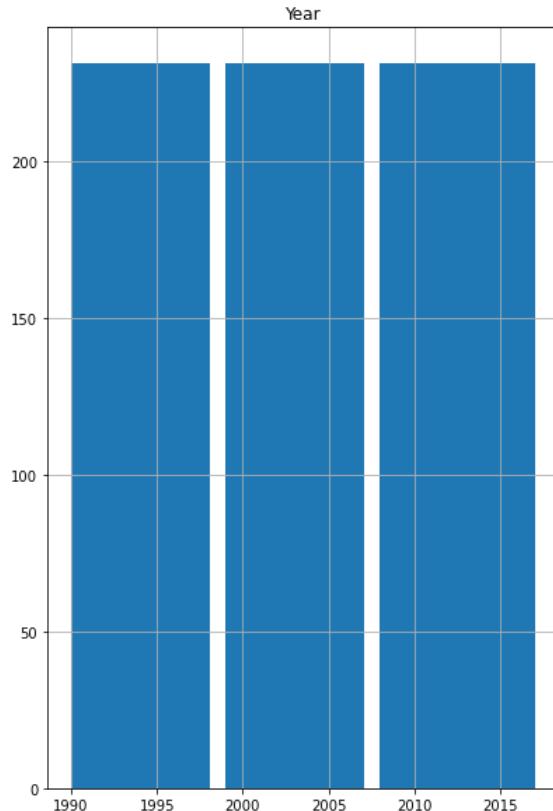
Entity Year Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)			
6383	World	2017	326.471123
6382	World	2016	319.494896
6381	World	2015	312.838532
6380	World	2014	308.079177
6379	World	2013	299.684466

In [1641]: df34.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1641]:

Entity Year Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)			
3500	Maldives	1990	0.000106
3501	Maldives	1991	0.000108
3502	Maldives	1992	0.000109
3503	Maldives	1993	0.000113
3504	Maldives	1994	0.000118

```
In [1642]: df34.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1643]: df34.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1643]: Entity
United States Virgin Islands      0.000218
Maldives                          0.000282
Seychelles                         0.000296
Saint Vincent and the Grenadines   0.000523
Grenada                           0.000600
Name: Deaths - Eating disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

```
In [1644]: df34.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1644]: Entity
High-income Asia Pacific      64.561814
Western Europe                 76.148341
High-income                     184.884534
High SDI                        191.423568
World                           262.029608
Name: Deaths - Eating disorders - Sex: Both - Age: All Ages (Number), dtype: float64
```

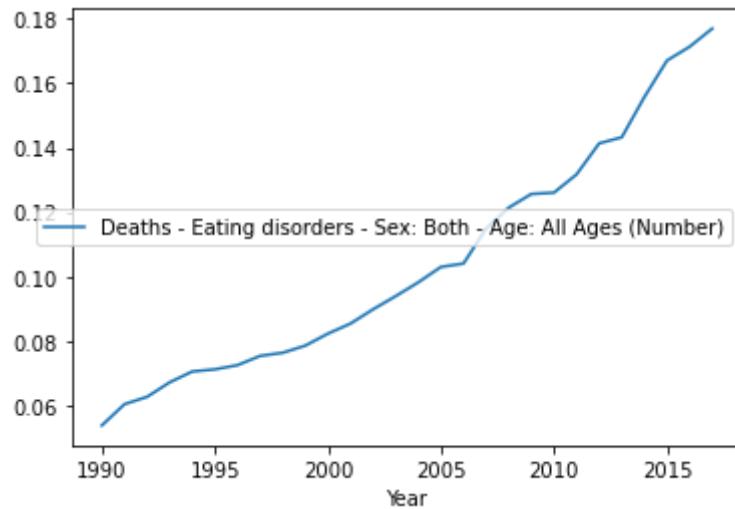
```
In [1645]: df34_mean = df34.groupby('Year').mean()
df34_mean.head()
```

```
Out[1645]:
Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)
```

Year	
1990	3.890324
1991	4.007735
1992	4.194992
1993	4.510057
1994	4.750777

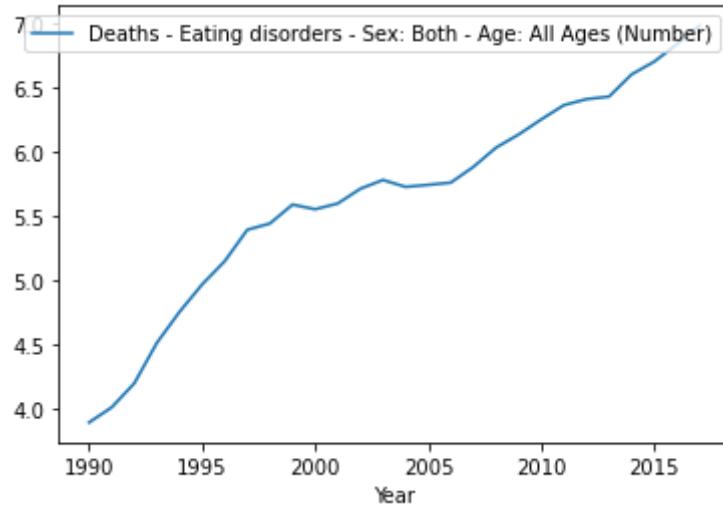
```
In [1646]: df34.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1646]: <AxesSubplot:xlabel='Year'>
```



```
In [1647]: df34.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1647]: <AxesSubplot:xlabel='Year'>
```



In [1648]: df55.info()
df55.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)
6468 non-null   float64
4   Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)
6468 non-null   float64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1648]:

	Entity	Code	Year	Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)
0	Afghanistan	AFG	1990	6077.853495	2351.689470
1	Afghanistan	AFG	1991	6257.514020	2458.831153
2	Afghanistan	AFG	1992	7624.503607	3092.250253
3	Afghanistan	AFG	1993	8978.220955	3800.781447
4	Afghanistan	AFG	1994	9207.362771	3955.342909

Checking for missing values:

In [1649]: missing = pd.concat([df55.isnull().sum(), 100 * df55.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1649]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)		0	0.000000
Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)		0	0.000000
Code	980	15.151515	

```
In [1650]: v1='Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)'
v2='Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)'
```

```
In [1651]: df55.describe()
```

```
Out[1651]:
```

	Year	Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)
count	6468.000000	6.468000e+03	6.468000e+03
mean	2003.500000	1.809780e+05	7.882217e+04
std	8.078372	7.045898e+05	3.056102e+05
min	1990.000000	2.912341e+01	1.475009e+01
25%	1996.750000	2.817867e+03	1.348587e+03
50%	2003.500000	1.163604e+04	5.290655e+03
75%	2010.250000	6.527186e+04	2.785540e+04
max	2017.000000	1.092640e+07	4.875281e+06

```
In [1652]: df55.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

```
Out[1652]:
```

	Entity	Year	Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)
6383	World	2017	1.092640e+07
6382	World	2016	1.077195e+07
6381	World	2015	1.061189e+07
6380	World	2014	1.045570e+07
6379	World	2013	1.029700e+07

```
In [1653]: df55.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()
```

```
Out[1653]:
```

	Entity	Year	Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)
6383	World	2017	4.875281e+06
6382	World	2016	4.791807e+06
6381	World	2015	4.706558e+06
6380	World	2014	4.624731e+06
6379	World	2013	4.542386e+06

```
In [1654]: df55.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[1654]:

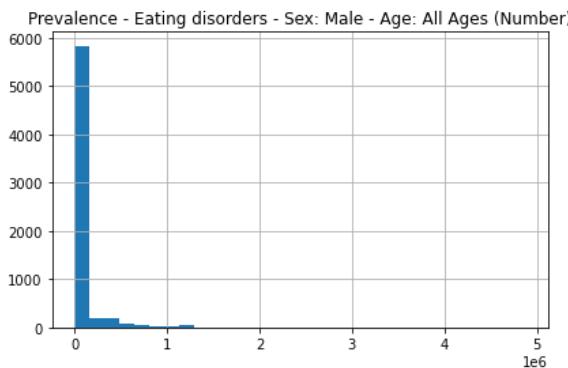
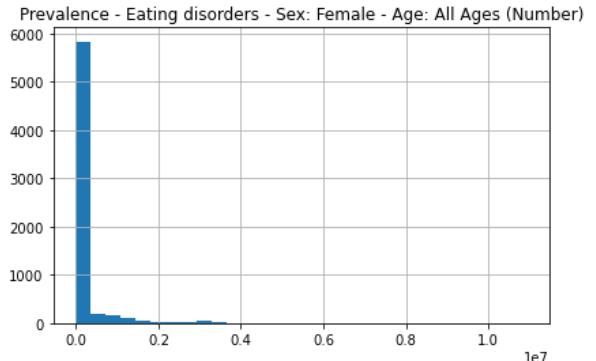
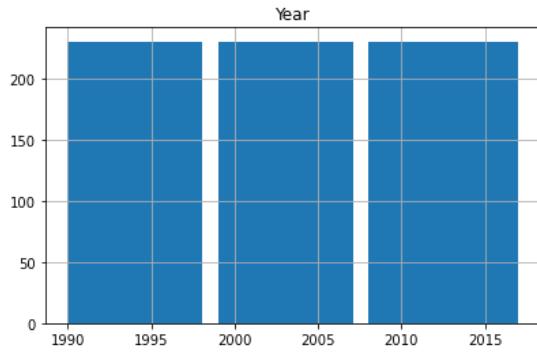
	Entity	Year	Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)
3584	Marshall Islands	1990	29.123410
3585	Marshall Islands	1991	30.386863
3586	Marshall Islands	1992	31.589804
3587	Marshall Islands	1993	32.749016
3588	Marshall Islands	1994	33.784494

```
In [1655]: df55.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()
```

Out[1655]:

	Entity	Year	Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)
3584	Marshall Islands	1990	14.750091
3585	Marshall Islands	1991	15.418601
3586	Marshall Islands	1992	15.962217
3587	Marshall Islands	1993	16.466788
3588	Marshall Islands	1994	17.029511

```
In [1656]: df55.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1657]: df55.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1657]: Entity
Marshall Islands      38.642474
Kiribati              61.411551
American Samoa       66.648951
Micronesia (country) 75.978745
Tonga                 78.060673
Name: Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number), dtype: float64
```

```
In [1658]: df55.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1658]: Entity
Marshall Islands      19.603499
Kiribati              28.938786
American Samoa       32.065387
Micronesia (country) 38.057318
Tonga                 38.616211
Name: Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1659]: df55.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1659]: Entity
Southeast Asia, East Asia, and Oceania  1.639942e+06
Middle SDI                            2.003634e+06
High SDI                             3.135389e+06
High-income                           3.155140e+06
World                                8.708968e+06
Name: Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number), dtype: float64
```

```
In [1660]: df55.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1660]: Entity
Southeast Asia, East Asia, and Oceania  8.296313e+05
Middle SDI                            9.225588e+05
High-income                           1.153459e+06
High SDI                             1.162530e+06
World                                3.823162e+06
Name: Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number), dtype: float64
```

```
In [1661]: df55_mean = df55.groupby('Year').mean()  
df55_mean.head()
```

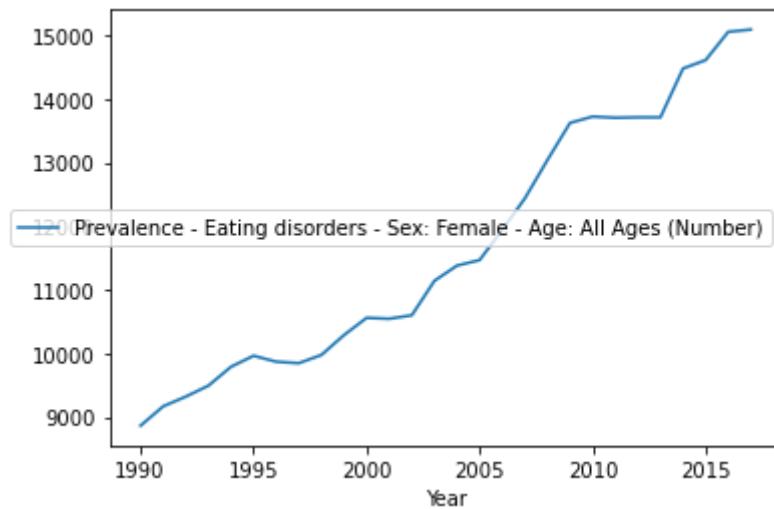
Out[1661]:

Year	Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)
------	--	--

Year	Prevalence - Eating disorders - Sex: Female - Age: All Ages (Number)	Prevalence - Eating disorders - Sex: Male - Age: All Ages (Number)
1990	140652.613153	61415.173409
1991	143112.414441	62532.735849
1992	145576.255266	63595.045107
1993	148038.344316	64661.275520
1994	150536.126073	65718.769509

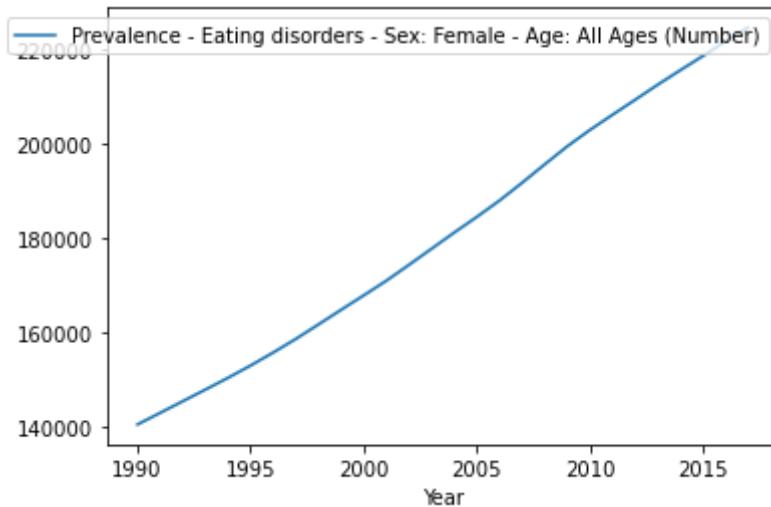
```
In [1662]: df55.groupby('Year')[v1].median().plot(legend=True)
```

Out[1662]: <AxesSubplot:xlabel='Year'>



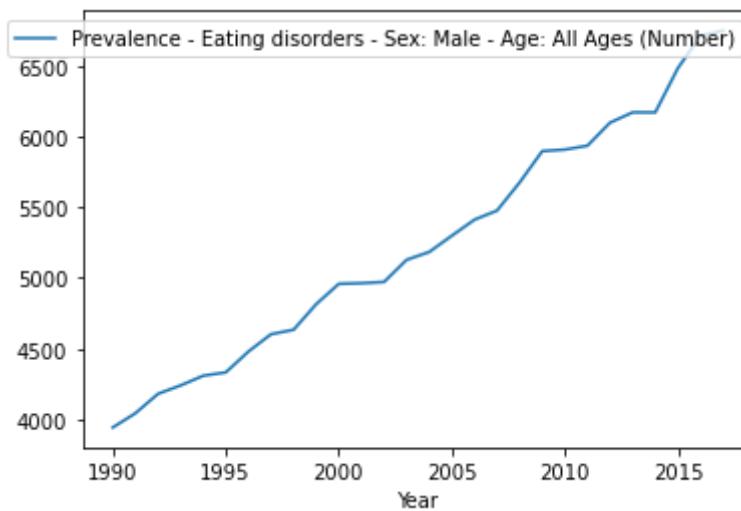
```
In [1663]: df55.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1663]: <AxesSubplot:xlabel='Year'>
```



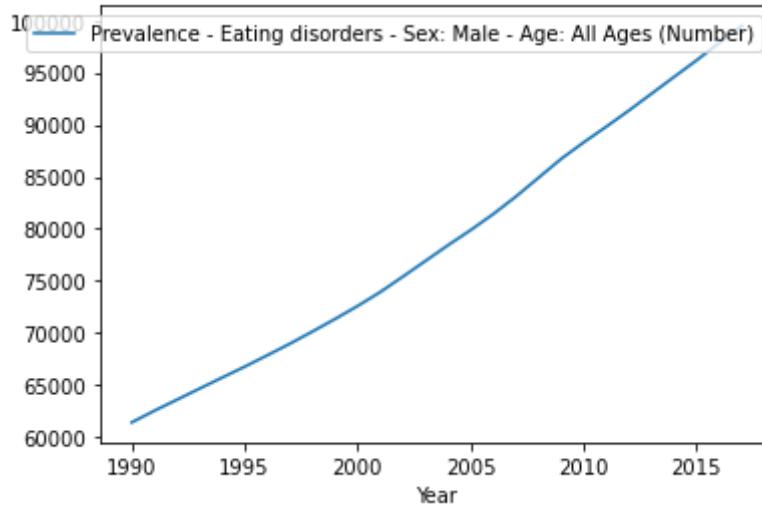
```
In [1664]: df55.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1664]: <AxesSubplot:xlabel='Year'>
```



```
In [1665]: df55.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1665]: <AxesSubplot:xlabel='Year'>
```



In [1666]: df56.info()
df56.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
4   Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)
6468 non-null   float64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1666]:

	Entity	Code	Year	Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)	Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	1990	1869.815372	6671.754883
1	Afghanistan	AFG	1991	1940.434799	6880.802091
2	Afghanistan	AFG	1992	2379.431771	8447.597072
3	Afghanistan	AFG	1993	2815.379527	10093.391198
4	Afghanistan	AFG	1994	2868.647733	10426.222061

Checking for missing values:

In [1667]: missing = pd.concat([df56.isnull().sum(), 100 * df56.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1667]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)		0	0.000000
Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)		0	0.000000
	Code	980	15.151515

In [1668]: v1='Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)'
v2='Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)'

In [1669]: df56.describe()

Out[1669]:

	Year	Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)	Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)
count	6468.000000	6.468000e+03	6.468000e+03
mean	2003.500000	5.844110e+04	2.026547e+05
std	8.078372	2.288044e+05	7.858983e+05
min	1990.000000	1.270034e+01	3.171380e+01
25%	1996.750000	8.822921e+02	3.343499e+03
50%	2003.500000	3.444009e+03	1.349250e+04
75%	2010.250000	1.761223e+04	7.542155e+04
max	2017.000000	3.360273e+06	1.250966e+07

In [1670]: df56.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1670]:

	Entity	Year	Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)
6383	World	2017	3.360273e+06
6382	World	2016	3.325280e+06
6381	World	2015	3.288404e+06
6380	World	2014	3.253157e+06
6379	World	2013	3.217087e+06

In [1671]: df56.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1671]:

	Entity	Year	Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)
6383	World	2017	1.250966e+07
6382	World	2016	1.230544e+07
6381	World	2015	1.209630e+07
6380	World	2014	1.189363e+07
6379	World	2013	1.168949e+07

In [1672]: `df56.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()`

Out[1672]:

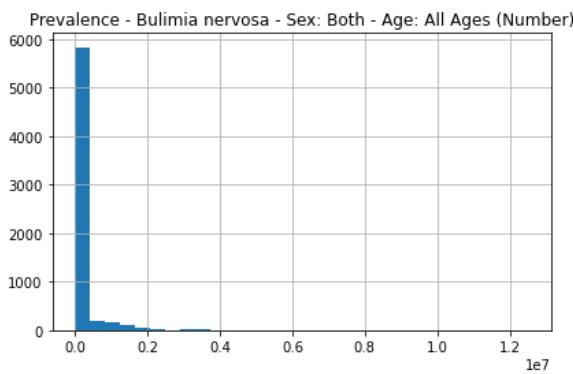
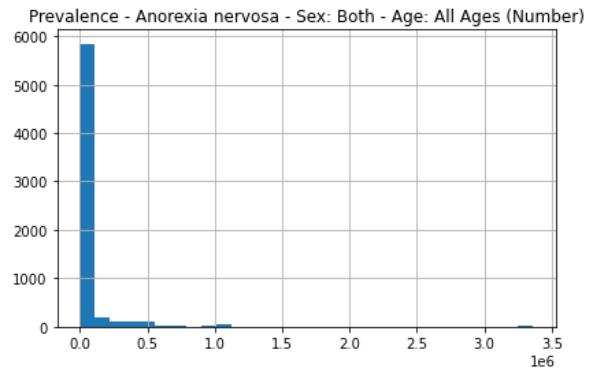
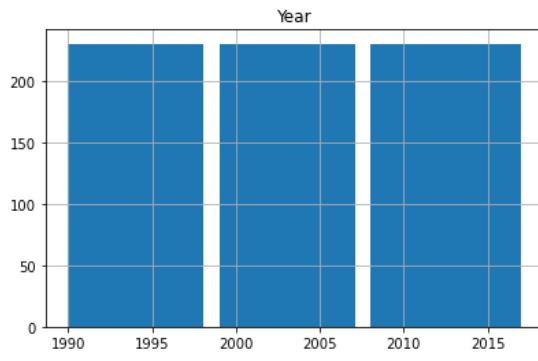
	Entity	Year	Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	12.700335
3585	Marshall Islands	1991	13.186776
3586	Marshall Islands	1992	13.665141
3587	Marshall Islands	1993	14.129362
3588	Marshall Islands	1994	14.569969

In [1673]: `df56.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()`

Out[1673]:

	Entity	Year	Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)
3584	Marshall Islands	1990	31.713796
3585	Marshall Islands	1991	33.059605
3586	Marshall Islands	1992	34.376537
3587	Marshall Islands	1993	35.642452
3588	Marshall Islands	1994	36.832068

In [1674]: `df56.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [1675]: df56.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1675]: Entity
Marshall Islands      15.978751
Kiribati              24.714784
American Samoa       25.497151
Dominica              26.780446
Micronesia (country) 31.750680
Name: Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number),
dtype: float64
```

```
In [1676]: df56.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1676]: Entity
Marshall Islands      42.841602
Kiribati              66.591107
American Samoa       73.578549
Micronesia (country) 83.427398
Tonga                 85.680034
Name: Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number),
dtype: float64
```

```
In [1677]: df56.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1677]: Entity
Middle SDI            6.295449e+05
Southeast Asia, East Asia, and Oceania 6.370675e+05
High SDI               1.038173e+06
High-income            1.043900e+06
World                  2.804543e+06
Name: Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number),
dtype: float64
```

```
In [1678]: df56.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1678]: Entity
High-middle SDI       1.890186e+06
Middle SDI             2.313065e+06
High SDI               3.271819e+06
High-income            3.276613e+06
World                  9.791403e+06
Name: Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number),
dtype: float64
```

```
In [1679]: df56_mean = df56.groupby('Year').mean()  
df56_mean.head()
```

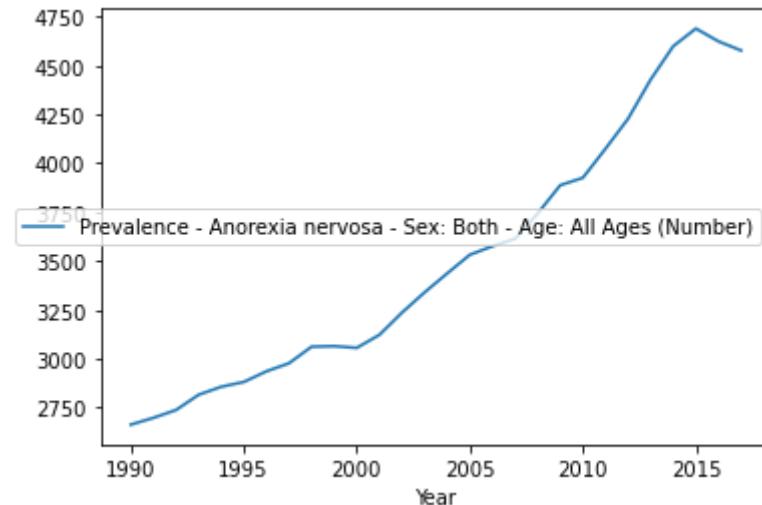
Out[1679]:

Year	Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)	Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)
------	--	---

Year	Prevalence - Anorexia nervosa - Sex: Both - Age: All Ages (Number)	Prevalence - Bulimia nervosa - Sex: Both - Age: All Ages (Number)
1990	47595.004946	155592.328433
1991	48243.723646	158531.083730
1992	48895.894391	161449.292777
1993	49544.971617	164365.396287
1994	50197.657604	167281.611530

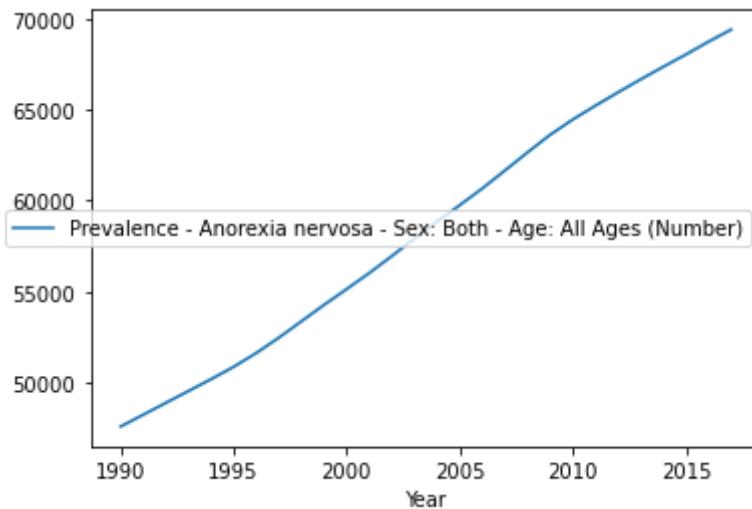
```
In [1680]: df56.groupby('Year')[v1].median().plot(legend=True)
```

Out[1680]: <AxesSubplot:xlabel='Year'>



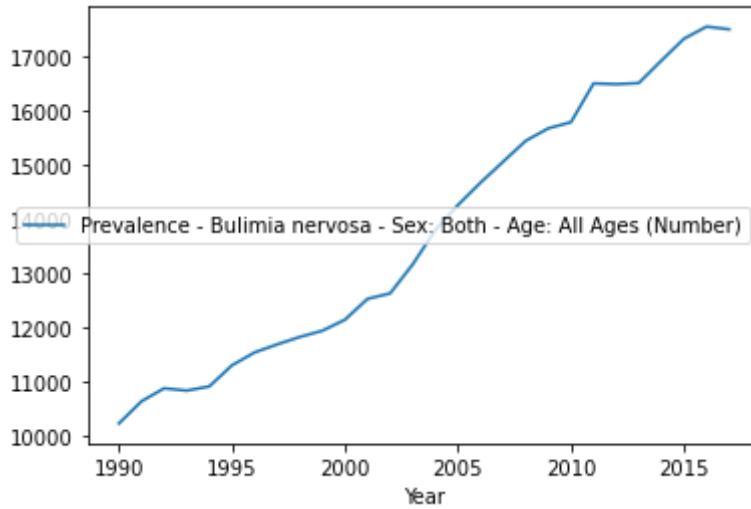
```
In [1681]: df56.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1681]: <AxesSubplot:xlabel='Year'>
```



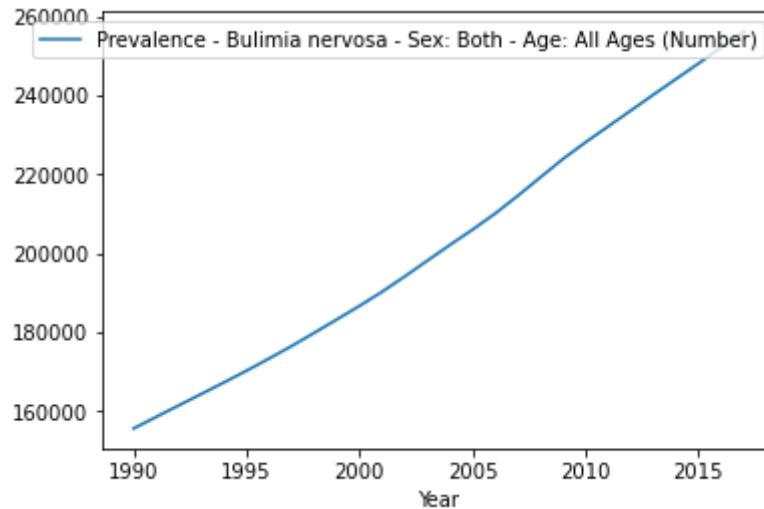
```
In [1682]: df56.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1682]: <AxesSubplot:xlabel='Year'>
```



```
In [1683]: df56.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1683]: <AxesSubplot:xlabel='Year'>
```



In [1684]: df82.info()
df82.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 13 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)
6468 non-null   float64
4   Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)
6468 non-null   float64
5   Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
6   Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)
6468 non-null   float64
7   Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)
6468 non-null   float64
8   Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)
6468 non-null   float64
9   Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)
6468 non-null   int64
10  Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)
6468 non-null   float64
11  Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)
6468 non-null   int64
12  Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)
6468 non-null   float64
dtypes: float64(8), int64(3), object(2)
memory usage: 657.0+ KB
```

Out[1684]:

	Entity	Code	Year	Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)
0	Afghanistan	AFG	1990	0.205061	0.057963	0.085209	0.253637	0.225103	0.029186
1	Afghanistan	AFG	1991	0.200435	0.056773	0.084860	0.247479	0.217095	0.027694
2	Afghanistan	AFG	1992	0.197078	0.055832	0.088429	0.240508	0.209636	0.026495
3	Afghanistan	AFG	1993	0.192601	0.054793	0.090556	0.235083	0.203330	0.025313
4	Afghanistan	AFG	1994	0.188914	0.054836	0.088791	0.229719	0.199129	0.024626

Checking for missing values:

```
In [1685]: missing = pd.concat([df82.isnull().sum(), 100 * df82.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count',ascending=True)
```

Out[1685]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)		0	0.000000
Code	980	15.151515	

```
In [1686]: v1='Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)'
v2='Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)'
v3='Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)'
v4='Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)'
v5='Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)'
v6='Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)'
v7='Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)'
v8='Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)'
v9='Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)'
v10='Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)'
```

In [1687]: df82.describe()

Out[1687]:

	Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.531695	0.126384	0.240193	0.555195	0.547019
std	8.078372	0.366585	0.090696	0.147939	0.326691	0.359661
min	1990.000000	0.164833	0.043624	0.072877	0.178863	0.168182
25%	1996.750000	0.261458	0.066412	0.124929	0.303241	0.277352
50%	2003.500000	0.381149	0.091149	0.188156	0.445838	0.407413
75%	2010.250000	0.653783	0.128287	0.323992	0.701805	0.679370
max	2017.000000	2.120488	0.527088	0.843046	2.056389	0.265432

In [1688]: df82.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1688]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)
334	Australia	2016	2.120488
335	Australia	2017	2.120414
333	Australia	2015	2.113774
332	Australia	2014	2.109945
331	Australia	2013	2.101620

In [1689]: df82.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1689]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)
5424	Spain	2010	0.527088
5425	Spain	2011	0.527066
5426	Spain	2012	0.524078
5427	Spain	2013	0.522544
5423	Spain	2009	0.522124

In [1690]: df82.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1690]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)
333	Australia	2015	0.843046
332	Australia	2014	0.842953
331	Australia	2013	0.842940
330	Australia	2012	0.842378
334	Australia	2016	0.842091

In [1691]: df82.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1691]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)
335	Australia	2017	2.056389
334	Australia	2016	2.053816
333	Australia	2015	2.048208
332	Australia	2014	2.039884
331	Australia	2013	2.033183

In [1692]: df82.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1692]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)
335	Australia	2017	2.148368
334	Australia	2016	2.143965
333	Australia	2015	2.136231
332	Australia	2014	2.129615
331	Australia	2013	2.120497

In [1693]: df82.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1693]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)
5424	Spain	2010	0.265432
5423	Spain	2009	0.265157
5422	Spain	2008	0.264965
5425	Spain	2011	0.264652
5421	Spain	2007	0.263175

In [1694]: df82.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1694]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)
0	Afghanistan	1990	0
4309	Norway	2015	0
4319	Oceania	1997	0
4318	Oceania	1996	0
4317	Oceania	1995	0

In [1695]: df82.sort_values(by=v8, ascending=False)[['Entity', 'Year', v8]].head()

Out[1695]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)
334	Australia	2016	0.943991
333	Australia	2015	0.943206
335	Australia	2017	0.943081
332	Australia	2014	0.941661
331	Australia	2013	0.939176

In [1696]: df82.sort_values(by=v9, ascending=False)[['Entity', 'Year', v9]].head()

Out[1696]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)
0	Afghanistan	1990	0
4309	Norway	2015	0
4319	Oceania	1997	0
4318	Oceania	1996	0
4317	Oceania	1995	0

In [1697]: df82.sort_values(by=v10, ascending=False)[['Entity', 'Year', v10]].head()

Out[1697]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)
333	Australia	2015	1.558513
334	Australia	2016	1.556749
332	Australia	2014	1.556637
335	Australia	2017	1.553368
331	Australia	2013	1.551718

In [1698]: df82.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1698]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)
3893	Myanmar	1991	0.164833
3892	Myanmar	1990	0.165046
3897	Myanmar	1995	0.165092
3894	Myanmar	1992	0.165279
3895	Myanmar	1993	0.165579

In [1699]: df82.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1699]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)
1993	Ethiopia	1995	0.043624
3869	Mozambique	1995	0.043927
3868	Mozambique	1994	0.043938
1997	Ethiopia	1999	0.044174
3866	Mozambique	1992	0.044175

In [1700]: df82.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1700]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)
3867	Mozambique	1993	0.072877
3865	Mozambique	1991	0.072913
3866	Mozambique	1992	0.072969
3868	Mozambique	1994	0.073017
3253	Liberia	1995	0.073096

In [1701]: df82.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1701]:

Entity Year Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)			
3894	Myanmar	1992	0.178863
3893	Myanmar	1991	0.178979
3892	Myanmar	1990	0.179363
3895	Myanmar	1993	0.179881
3896	Myanmar	1994	0.180658

In [1702]: df82.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1702]:

Entity Year Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)			
3893	Myanmar	1991	0.168182
3894	Myanmar	1992	0.168413
3895	Myanmar	1993	0.169108
3892	Myanmar	1990	0.169460
3896	Myanmar	1994	0.170437

In [1703]: df82.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[1703]:

Entity Year Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)			
10	Afghanistan	2000	0.019302
9	Afghanistan	1999	0.019406
11	Afghanistan	2001	0.019624
5168	Somalia	2006	0.019916
5169	Somalia	2007	0.019917

In [1704]: df82.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[1704]:

Entity Year Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)			
0	Afghanistan	1990	0
4318	Oceania	1996	0
4317	Oceania	1995	0
4316	Oceania	1994	0
4315	Oceania	1993	0

In [1705]: df82.sort_values(by=v8, ascending=True)[['Entity', 'Year', v7]].head()

Out[1705]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)
3893	Myanmar	1991	0
3892	Myanmar	1990	0
3894	Myanmar	1992	0
3895	Myanmar	1993	0
3896	Myanmar	1994	0

In [1706]: df82.sort_values(by=v9, ascending=True)[['Entity', 'Year', v7]].head()

Out[1706]:

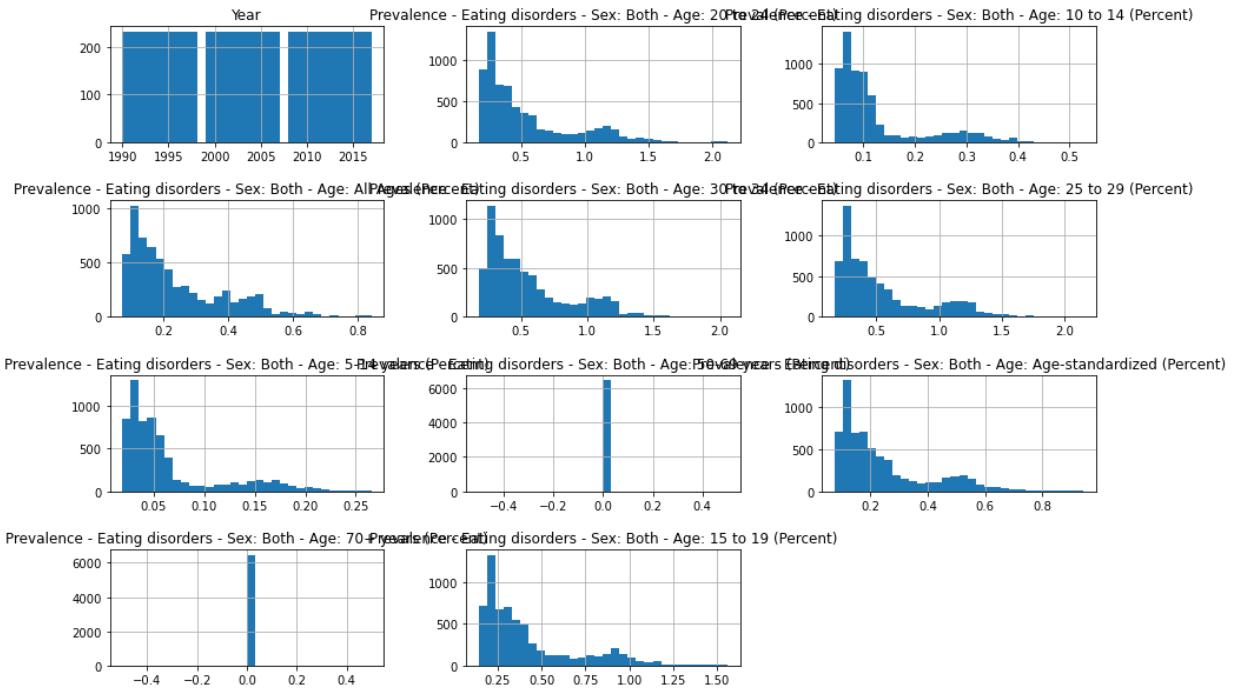
	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)
0	Afghanistan	1990	0
4318	Oceania	1996	0
4317	Oceania	1995	0
4316	Oceania	1994	0
4315	Oceania	1993	0

In [1707]: df82.sort_values(by=v10, ascending=True)[['Entity', 'Year', v7]].head()

Out[1707]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)
3869	Mozambique	1995	0
3867	Mozambique	1993	0
5166	Somalia	2004	0
5168	Somalia	2006	0
3865	Mozambique	1991	0

```
In [1708]: df82.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1709]: df82.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1709]: Entity
Somalia           0.171000
Liberia           0.186419
Ethiopia          0.186966
Mozambique        0.191577
Democratic Republic of Congo  0.191754
Name: Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1710]: df82.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1710]: Entity
Somalia           0.046382
Ethiopia          0.047963
Liberia           0.048728
Mozambique        0.049349
Democratic Republic of Congo  0.050099
Name: Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1711]: df82.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1711]: Entity
Somalia          0.075427
Niger            0.082526
Liberia          0.083562
Mozambique       0.083635
Ethiopia          0.083718
Name: Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [1712]: df82.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1712]: Entity
Somalia          0.194446
Liberia          0.211575
Cambodia         0.215284
Mozambique       0.215943
Solomon Islands  0.219425
Name: Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [1713]: df82.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1713]: Entity
Somalia          0.179063
Liberia          0.197106
Ethiopia          0.200429
Mozambique       0.200637
Cambodia         0.200794
Name: Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1714]: df82.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1714]: Entity
Somalia          0.021492
Mozambique       0.022185
Niger            0.022237
Ethiopia          0.022609
Democratic Republic of Congo  0.022950
Name: Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent), dtype: float64
```

```
In [1715]: df82.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1715]: Entity
Afghanistan          0
Nigeria              0
North Africa and Middle East 0
North America         0
North Korea            0
Name: Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent), dtype: int64
```

```
In [1716]: df82.groupby('Entity')[v8].mean().sort_values().head()
```

```
Out[1716]: Entity
Somalia             0.078287
Liberia              0.084655
Ethiopia             0.086306
Mozambique           0.086969
Democratic Republic of Congo 0.088121
Name: Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1717]: df82.groupby('Entity')[v9].mean().sort_values().head()
```

```
Out[1717]: Entity
Afghanistan          0
Nigeria              0
North Africa and Middle East 0
North America         0
North Korea            0
Name: Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent), dtype: int64
```

```
In [1718]: df82.groupby('Entity')[v10].mean().sort_values().head()
```

```
Out[1718]: Entity
Somalia             0.145687
Ethiopia             0.156196
Liberia              0.156368
Mozambique           0.159785
Democratic Republic of Congo 0.161527
Name: Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [1719]: df82.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1719]: Entity
New Zealand        1.451112
Spain               1.508388
Luxembourg          1.600262
Australasia         1.786203
Australia            1.851924
Name: Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent), dtype: float64
```

```
In [1720]: df82.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1720]: Entity
Australia      0.369397
Andorra        0.375310
Luxembourg    0.397500
Italy          0.408410
Spain          0.468679
Name: Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent), dtype: float64
```

```
In [1721]: df82.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1721]: Entity
Andorra        0.638032
Luxembourg    0.647572
Brunei         0.658154
Australasia   0.760236
Australia     0.791415
Name: Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent), dtype: float64
```

```
In [1722]: df82.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1722]: Entity
Spain          1.330248
New Zealand   1.393269
Luxembourg    1.449949
Australasia   1.760471
Australia     1.828657
Name: Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent), dtype: float64
```

```
In [1723]: df82.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1723]: Entity
Spain          1.437625
New Zealand   1.460683
Luxembourg    1.589257
Australasia   1.813776
Australia     1.877656
Name: Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent), dtype: float64
```

```
In [1724]: df82.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1724]: Entity
Finland        0.196585
Australia      0.197361
Luxembourg    0.205382
Italy          0.218324
Spain          0.252461
Name: Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent), dtype: float64
```

```
In [1725]: df82.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1725]: Entity
Greenland      0
Grenada        0
Guam           0
Guinea          0
Zimbabwe        0
Name: Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent), dtype: int64
```

```
In [1726]: df82.groupby('Entity')[v8].mean().sort_values().tail()
```

```
Out[1726]: Entity
New Zealand    0.643235
Spain           0.664637
Luxembourg     0.686169
Australasia    0.806650
Australia       0.838270
Name: Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1727]: df82.groupby('Entity')[v9].mean().sort_values().tail()
```

```
Out[1727]: Entity
Greenland      0
Grenada        0
Guam           0
Guinea          0
Zimbabwe        0
Name: Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent), dtype: int64
```

```
In [1728]: df82.groupby('Entity')[v10].mean().sort_values().tail()
```

```
Out[1728]: Entity
Luxembourg     1.201432
Italy           1.212252
Spain           1.309574
Australasia    1.319776
Australia       1.375250
Name: Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent), dtype: float64
```

```
In [1729]: df82_mean = df82.groupby('Year').mean()
df82_mean.head()
```

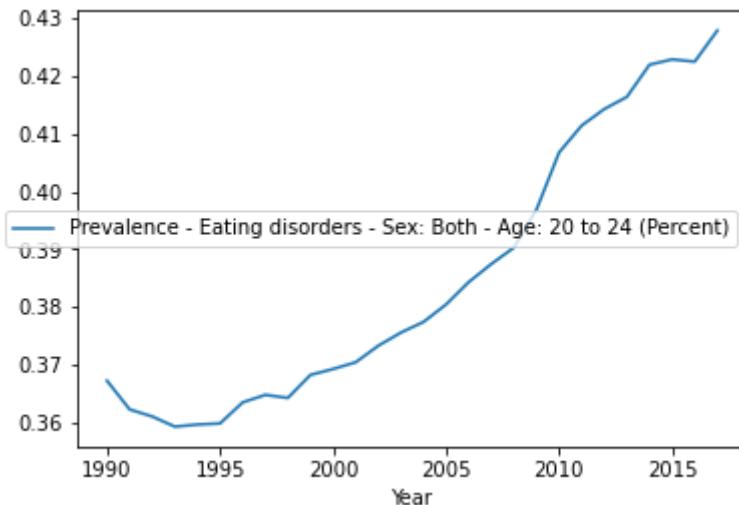
/opt/anaconda3/lib/python3.8/site-packages/IPython/core/displayhook.py:27
5: UserWarning: Output cache limit (currently 1000 entries) hit.
Flushing oldest 200 entries.
warn('Output cache limit (currently {sz} entries) hit.\n'

Out[1729]:

Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)	Prevalence - Eating disorders - Sex: Both - Age: All standard (Percent)
Year							
1990	0.492228	0.116429	0.226040	0.516996	0.507238	0.058461	0 0.221
1991	0.492978	0.116772	0.226539	0.517090	0.508064	0.058779	0 0.222
1992	0.493862	0.117149	0.226979	0.517576	0.508901	0.059171	0 0.222
1993	0.495007	0.117601	0.227451	0.518110	0.509675	0.059500	0 0.222

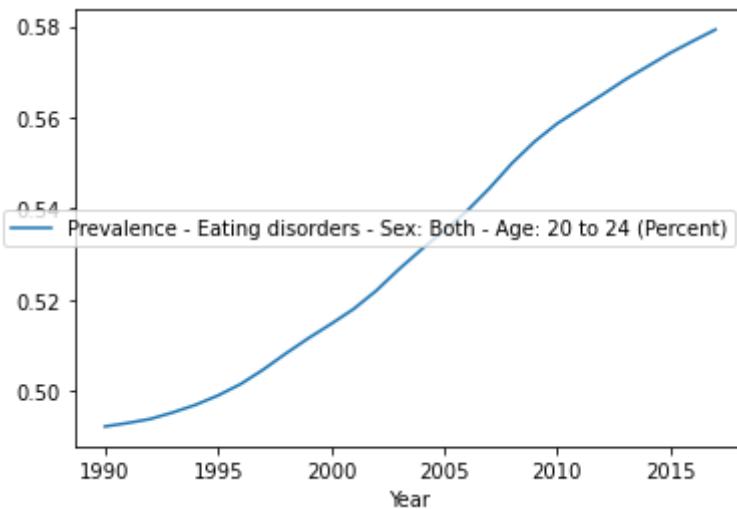
```
In [1730]: df82.groupby('Year')[v1].median().plot(legend=True)
```

Out[1730]: <AxesSubplot:xlabel='Year'>



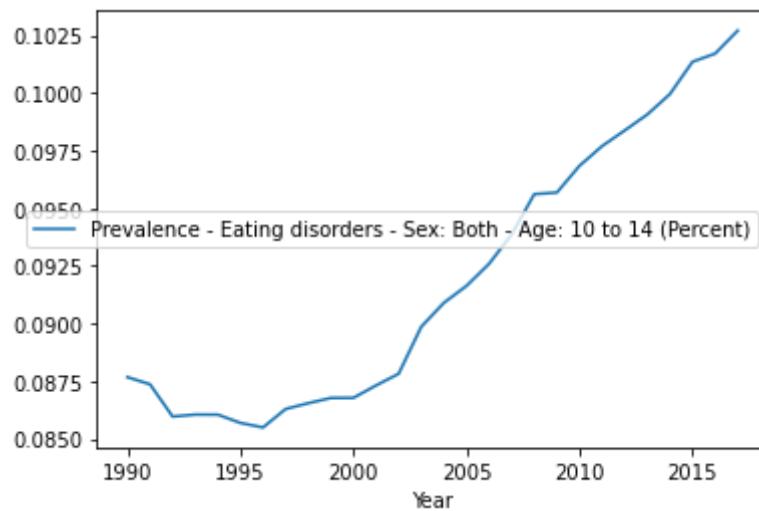
```
In [1731]: df82.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1731]: <AxesSubplot:xlabel='Year'>
```



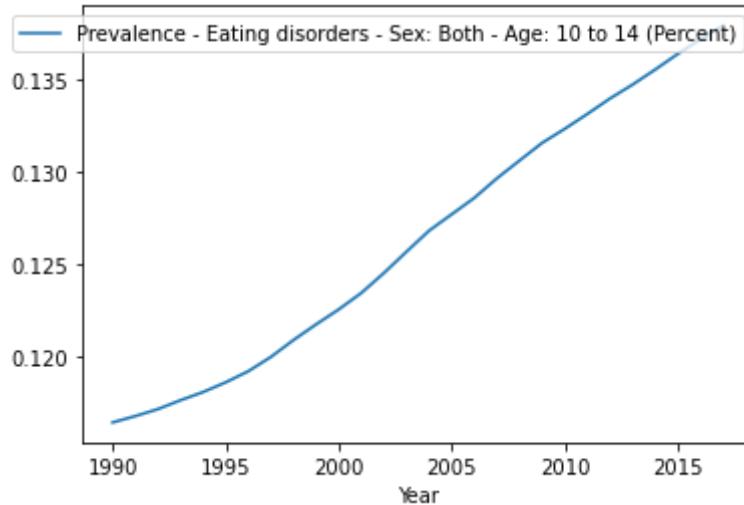
```
In [1732]: df82.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1732]: <AxesSubplot:xlabel='Year'>
```



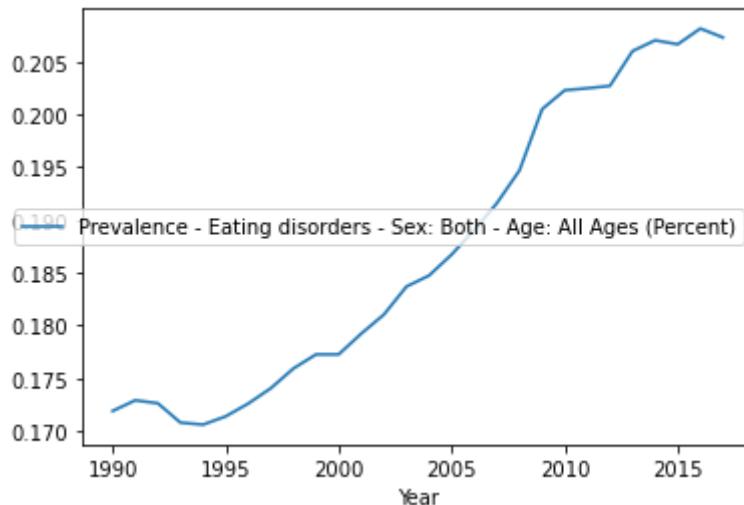
```
In [1733]: df82.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1733]: <AxesSubplot:xlabel='Year'>
```



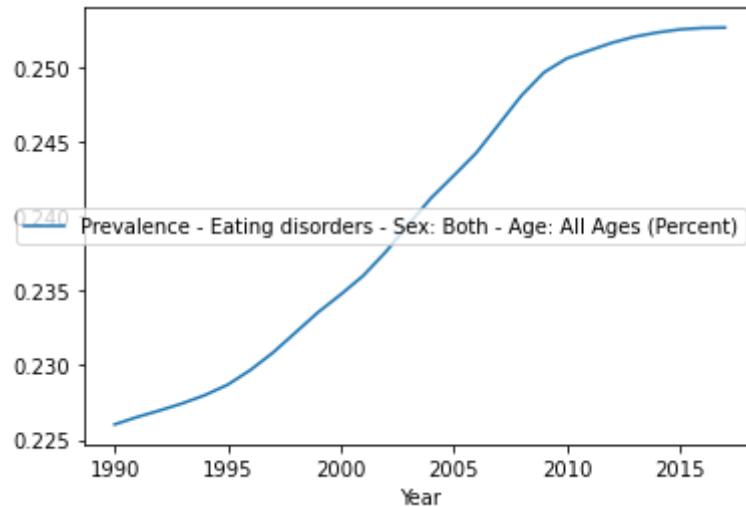
```
In [1734]: df82.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1734]: <AxesSubplot:xlabel='Year'>
```



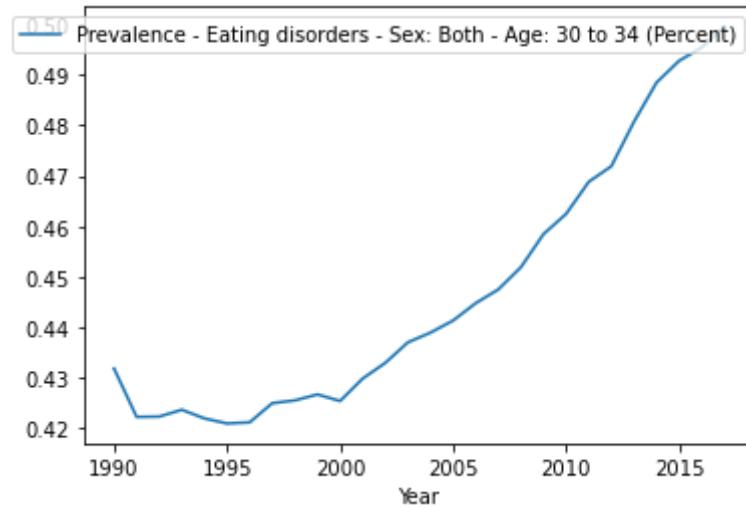
```
In [1735]: df82.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1735]: <AxesSubplot:xlabel='Year'>
```



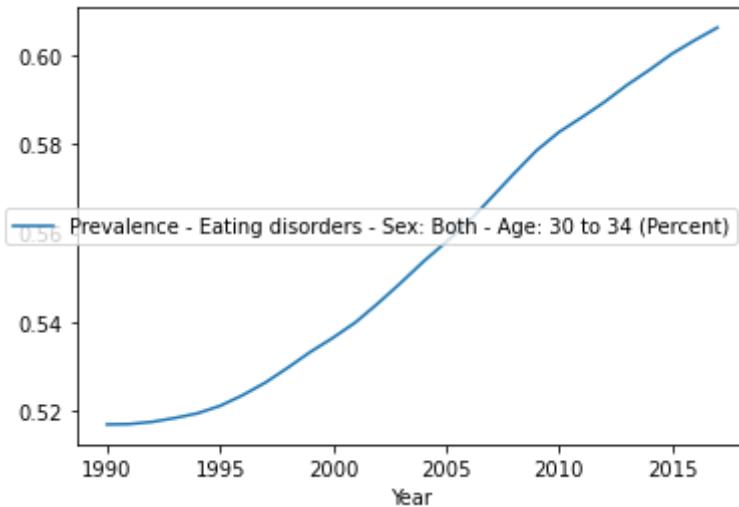
```
In [1736]: df82.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1736]: <AxesSubplot:xlabel='Year'>
```



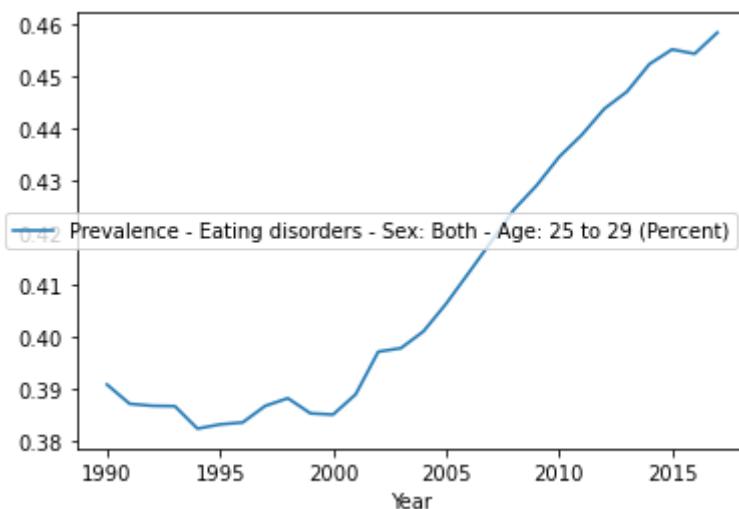
```
In [1737]: df82.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1737]: <AxesSubplot:xlabel='Year'>
```



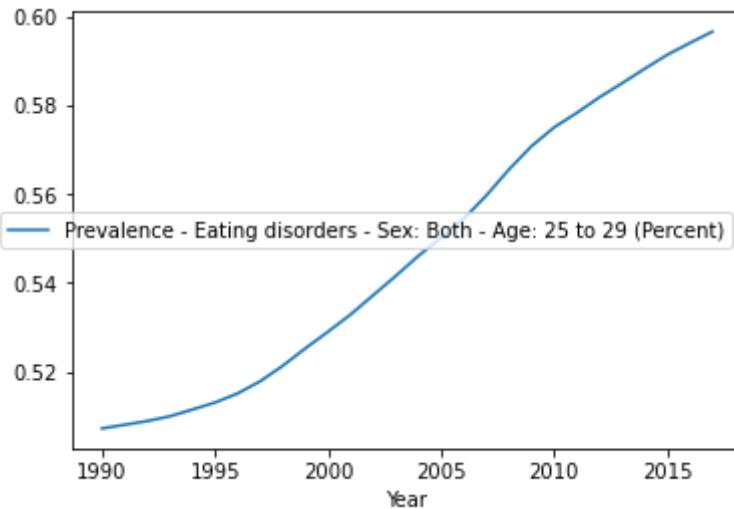
```
In [1738]: df82.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1738]: <AxesSubplot:xlabel='Year'>
```



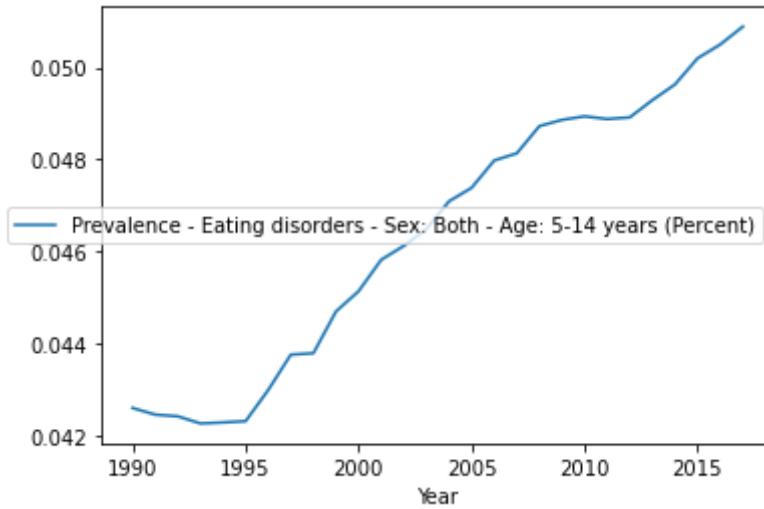
```
In [1739]: df82.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1739]: <AxesSubplot:xlabel='Year'>
```



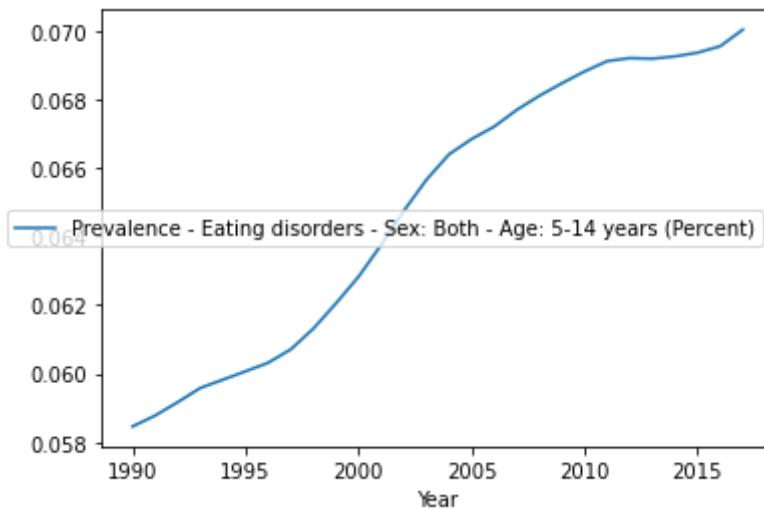
```
In [1740]: df82.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1740]: <AxesSubplot:xlabel='Year'>
```



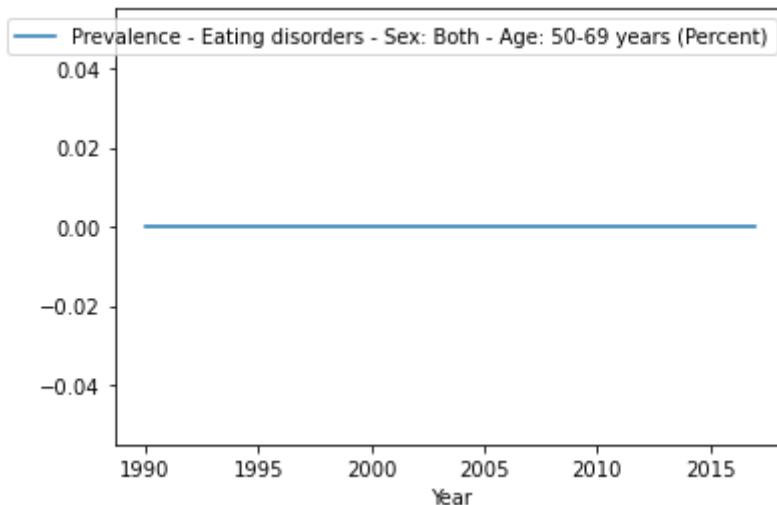
```
In [1741]: df82.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1741]: <AxesSubplot:xlabel='Year'>
```



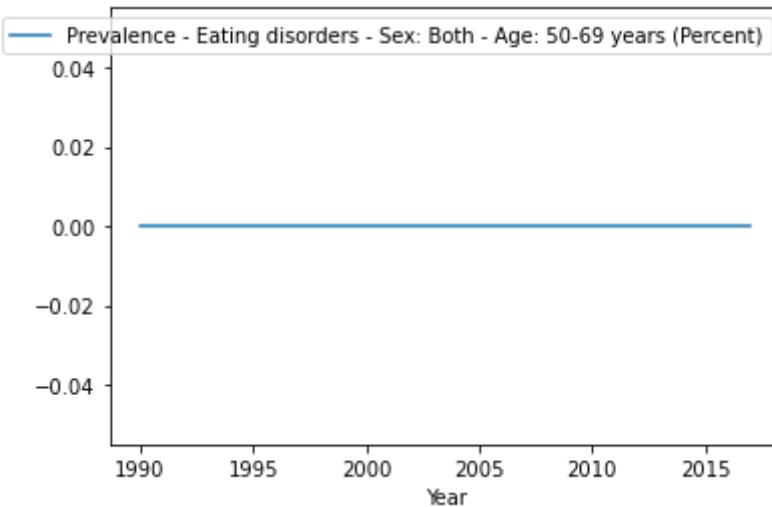
```
In [1742]: df82.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1742]: <AxesSubplot:xlabel='Year'>
```



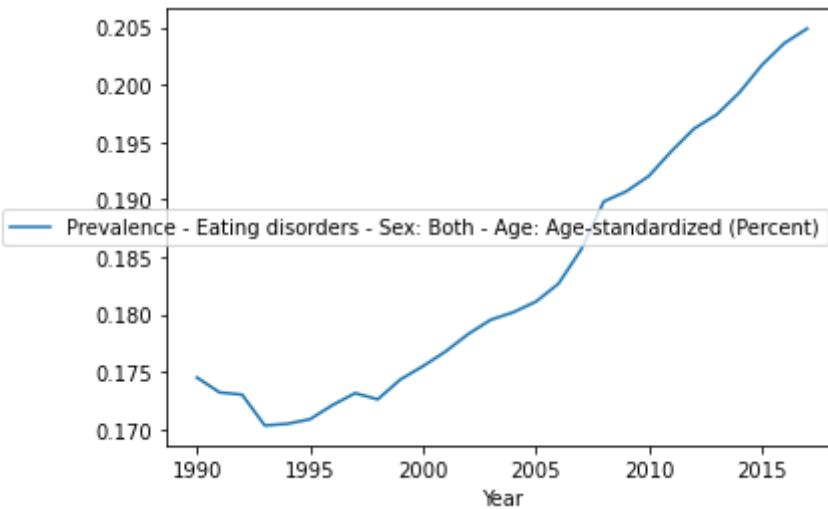
```
In [1743]: df82.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1743]: <AxesSubplot:xlabel='Year'>
```



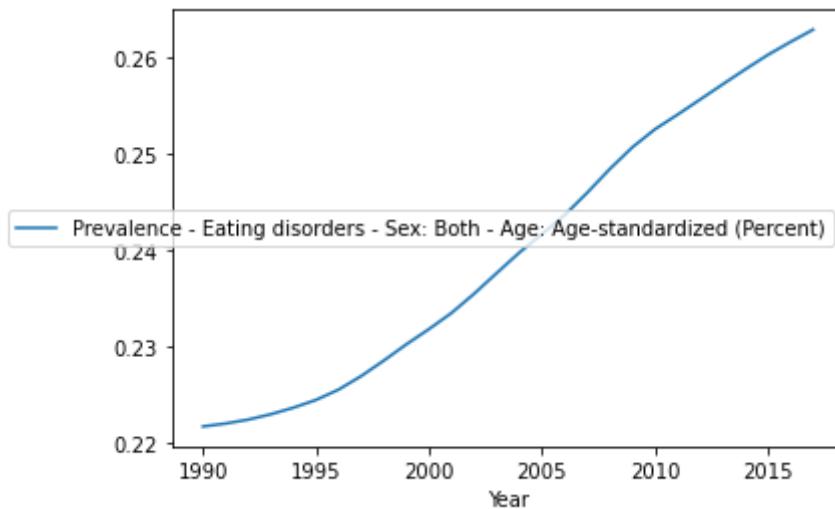
```
In [1744]: df82.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[1744]: <AxesSubplot:xlabel='Year'>
```



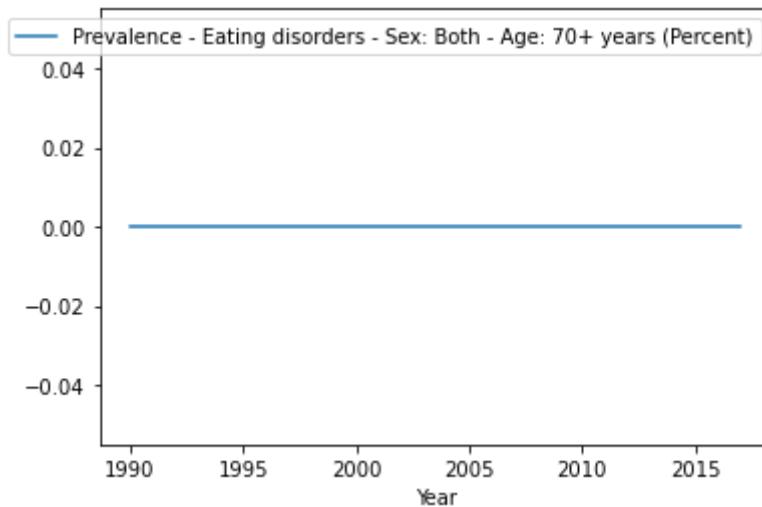
```
In [1745]: df82.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[1745]: <AxesSubplot:xlabel='Year'>
```



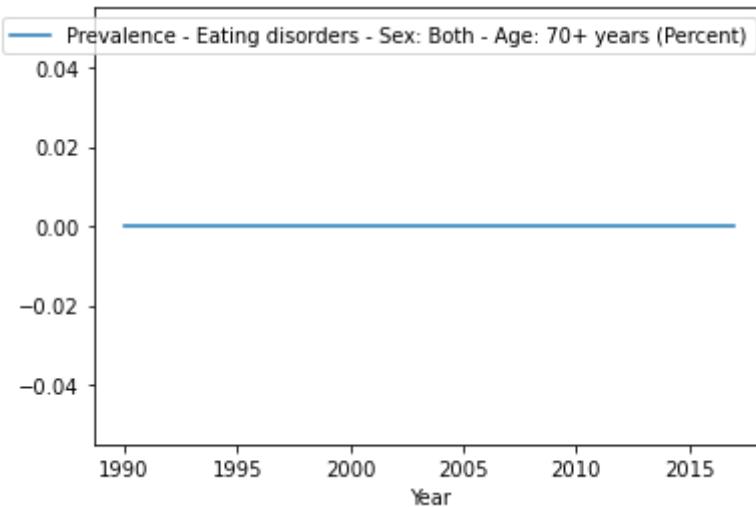
```
In [1746]: df82.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[1746]: <AxesSubplot:xlabel='Year'>
```



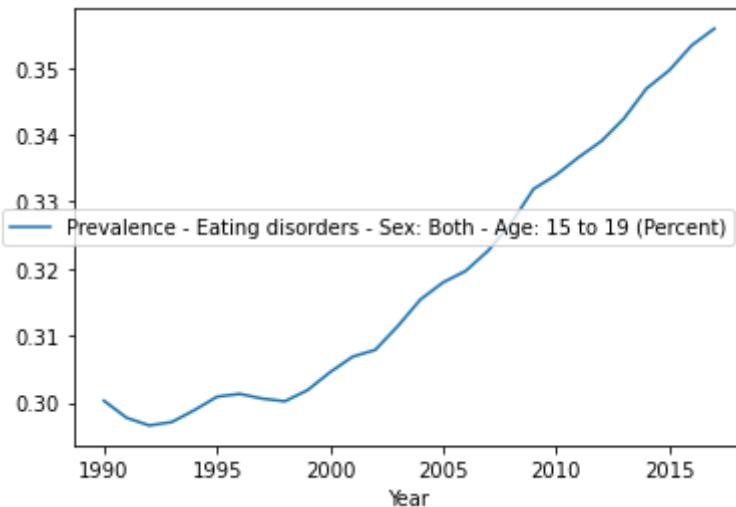
```
In [1747]: df82.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[1747]: <AxesSubplot:xlabel='Year'>
```



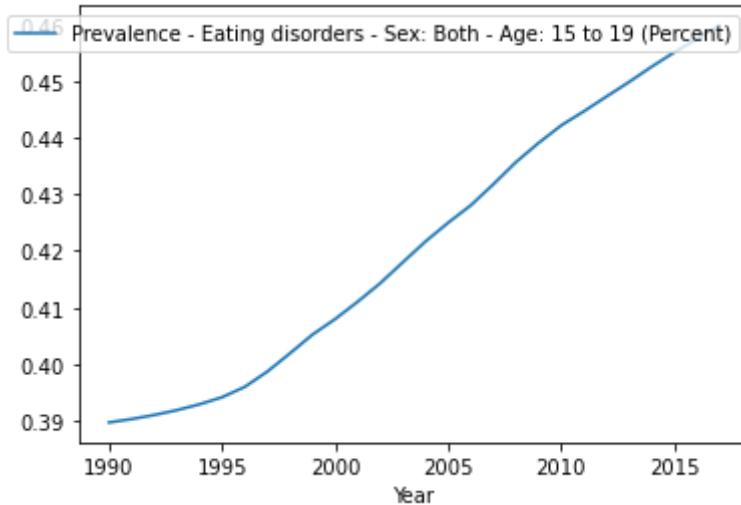
```
In [1748]: df82.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[1748]: <AxesSubplot:xlabel='Year'>
```



```
In [1749]: df82.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[1749]: <AxesSubplot:xlabel='Year'>
```



ADHD

In [1750]: df53.info()
df53.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 5 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity    object
6468 non-null   object
1   Code      object
5488 non-null   object
2   Year      int64
6468 non-null   int64
3   Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female
- Age: All Ages (Number) 6468 non-null   float64
4   Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male -
Age: All Ages (Number) 6468 non-null   float64
dtypes: float64(2), int64(1), object(2)
memory usage: 252.8+ KB
```

Out[1750]:

	Entity	Code	Year	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
0	Afghanistan	AFG	1990	36372.093241	84023.829459
1	Afghanistan	AFG	1991	37618.541437	87172.725113
2	Afghanistan	AFG	1992	44314.463114	103926.014285
3	Afghanistan	AFG	1993	51144.300550	121174.627507
4	Afghanistan	AFG	1994	52949.136913	125597.544292

Checking for missing values:

```
In [1751]: missing = pd.concat([df53.isnull().sum(), 100 * df53.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1751]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)		0	0.000000
Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)		0	0.000000
Code		980	15.151515

```
In [1752]: v1='Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)'
v2='Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)'
```

In [1753]: df53.describe()

Out[1753]:

Year	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
count	6468.000000	6.468000e+03
mean	2003.500000	4.014413e+05
std	8.078372	1.564105e+06
min	1990.000000	1.060946e+02
25%	1996.750000	7.740858e+03
50%	2003.500000	3.005130e+04
75%	2010.250000	1.308604e+05
max	2017.000000	2.165715e+07

```
In [1754]: df53.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[1754]:

Entity	Year	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)
6383	World	2017
6382	World	2016
6381	World	2015
6380	World	2014
6379	World	2013

In [1755]: df53.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1755]:

	Entity	Year	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
6383	World	2017	5.166036e+07
6382	World	2016	5.132780e+07
6381	World	2015	5.101141e+07
6380	World	2014	5.070490e+07
6379	World	2013	5.038945e+07

In [1756]: df53.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1756]:

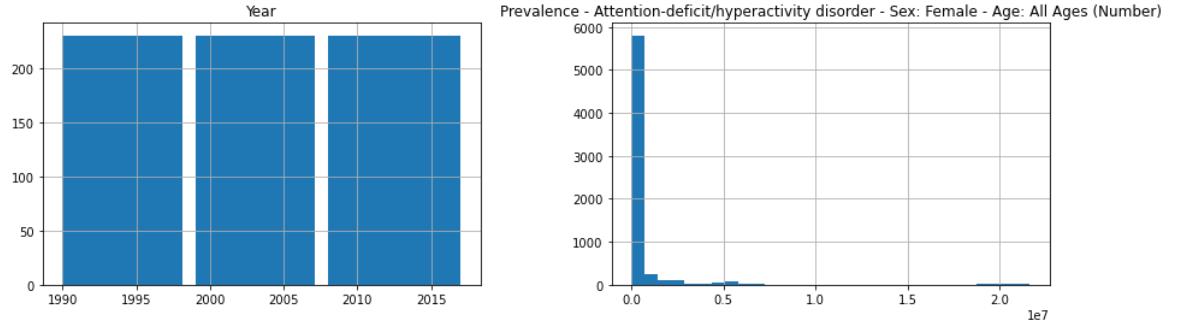
	Entity	Year	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)
4283	Northern Mariana Islands	2017	106.094556
4282	Northern Mariana Islands	2016	110.584525
4281	Northern Mariana Islands	2015	114.803278
4280	Northern Mariana Islands	2014	118.468194
140	Andorra	1990	121.294664

In [1757]: df53.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

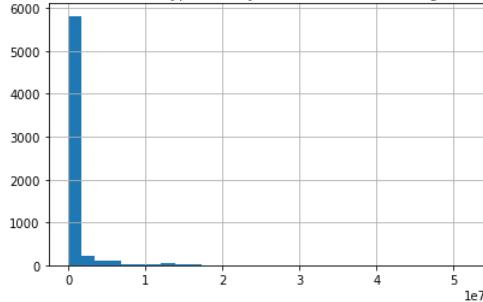
Out[1757]:

	Entity	Year	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
4283	Northern Mariana Islands	2017	258.848566
4282	Northern Mariana Islands	2016	267.546556
4281	Northern Mariana Islands	2015	277.483468
4280	Northern Mariana Islands	2014	288.492357
4256	Northern Mariana Islands	1990	297.548061

```
In [1758]: df53.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)



```
In [1759]: df53.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1759]: Entity
Andorra           144.975832
Northern Mariana Islands 147.412430
Marshall Islands 152.845272
American Samoa   154.112796
Seychelles        156.330321
Name: Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female
- Age: All Ages (Number), dtype: float64
```

```
In [1760]: df53.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1760]: Entity
Northern Mariana Islands 335.044620
Seychelles               355.761891
Marshall Islands         383.022368
American Samoa          385.894886
Andorra                  423.241453
Name: Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male -
Age: All Ages (Number), dtype: float64
```

```
In [1761]: df53.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1761]: Entity
China                               5.117918e+06
East Asia                           5.373103e+06
Middle SDI                          5.396560e+06
Southeast Asia, East Asia, and Oceania 6.586842e+06
World                                1.935280e+07
Name: Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female
- Age: All Ages (Number), dtype: float64
```

```
In [1762]: df53.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1762]: Entity
China                               1.332079e+07
Middle SDI                          1.347812e+07
East Asia                           1.398759e+07
Southeast Asia, East Asia, and Oceania 1.662976e+07
World                                4.690309e+07
Name: Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male -
Age: All Ages (Number), dtype: float64
```

```
In [1763]: df53_mean = df53.groupby('Year').mean()
df53_mean.head()
```

```
Out[1763]:
```

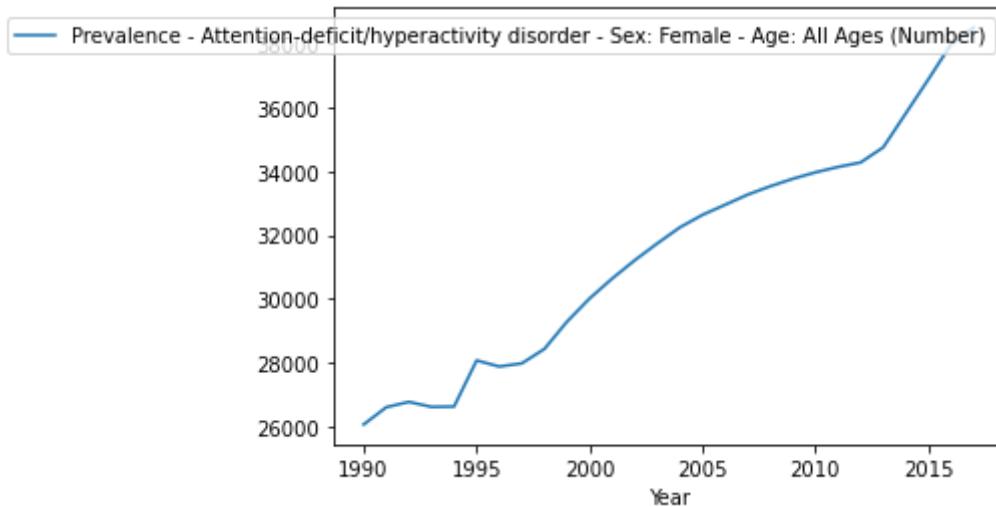
Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)

Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)

Year	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
1990	343817.374176	845413.591757
1991	348668.236653	858700.905123
1992	353608.349019	871978.313066
1993	358514.498113	884666.117014
1994	363436.028518	896580.094198

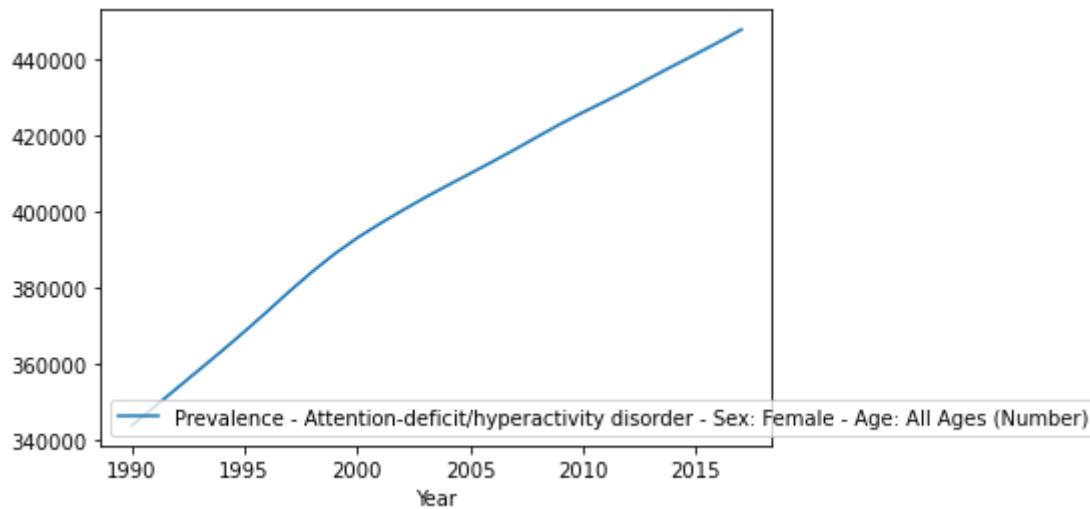
```
In [1764]: df53.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1764]: <AxesSubplot:xlabel='Year'>
```



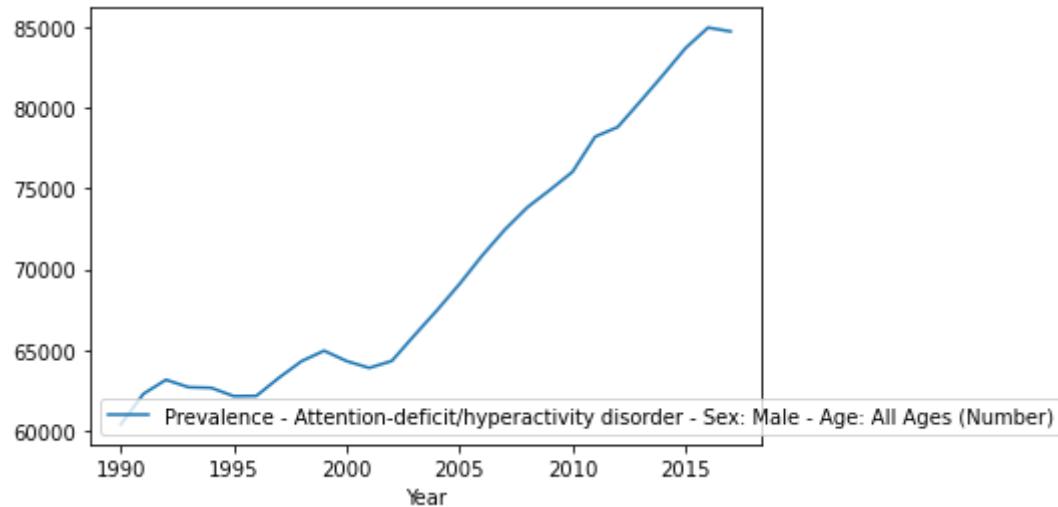
```
In [1765]: df53.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1765]: <AxesSubplot:xlabel='Year'>
```



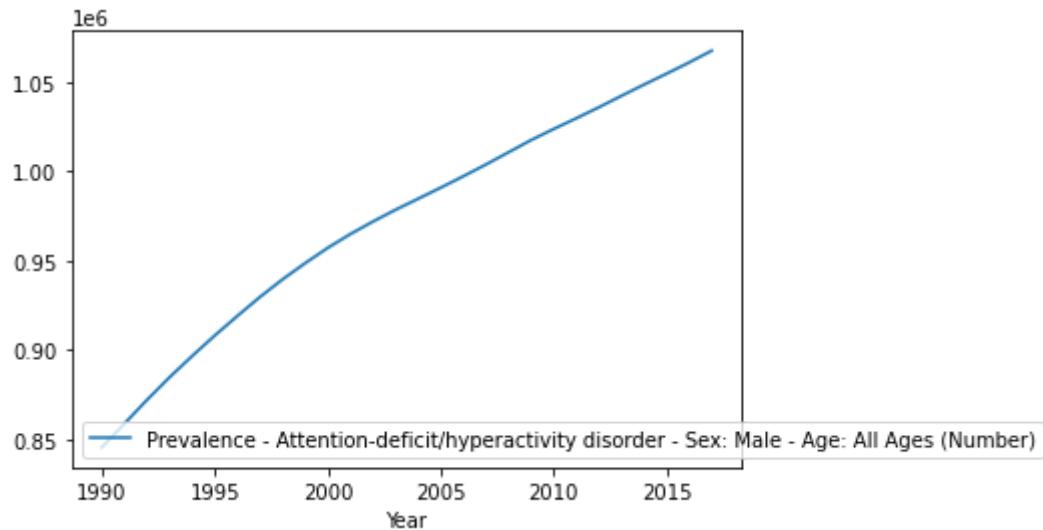
```
In [1766]: df53.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1766]: <AxesSubplot:xlabel='Year'>
```



```
In [1767]: df53.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1767]: <AxesSubplot:xlabel='Year'>
```



Mental Health Disorders

In [1768]: df65.info()
df65.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6156 entries, 0 to 6155
Data columns (total 5 columns):
 #   Column           Non-Null Count
 t   Dtype            dtype: object
 ---  -----
 -   -----
 0   Entity          6156 non-null
 object
 1   Code             5292 non-null
 object
 2   Year             6156 non-null
 int64
 3   Prevalence - Mental health disorders: Male (Number) 6156 non-null
 float64
 4   Prevalence - Mental health disorders: Female (Number) 6156 non-null
 float64
 dtypes: float64(2), int64(1), object(2)
 memory usage: 240.6+ KB
```

Out[1768]:

	Entity	Code	Year	Prevalence - Mental health disorders: Male (Number)	Prevalence - Mental health disorders: Female (Number)
0	Afghanistan	AFG	1990	8.783771e+05	7.913653e+05
1	Afghanistan	AFG	1991	9.366272e+05	8.377000e+05
2	Afghanistan	AFG	1992	1.013077e+06	9.006442e+05
3	Afghanistan	AFG	1993	1.098308e+06	9.695527e+05
4	Afghanistan	AFG	1994	1.183058e+06	1.038032e+06

Checking for missing values:

In [1769]: missing = pd.concat([df65.isnull().sum(), 100 * df65.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)

Out[1769]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
	Prevalence - Mental health disorders: Male (Number)	0	0.000000
	Prevalence - Mental health disorders: Female (Number)	0	0.000000
	Code	864	14.035088

In [1770]: v1='Prevalence - Mental health disorders: Male (Number)'
v2='Prevalence - Mental health disorders: Female (Number)'

In [1771]: df65.describe()

Out[1771]:

	Year	Prevalence - Mental health disorders: Male (Number)	Prevalence - Mental health disorders: Female (Number)
count	6156.000000	6.156000e+03	6.156000e+03
mean	2003.000000	7.233899e+06	7.557368e+06
std	7.789514	3.090349e+07	3.198344e+07
min	1990.000000	2.424608e+03	2.155059e+03
25%	1996.000000	1.442822e+05	1.403570e+05
50%	2003.000000	5.359235e+05	5.795400e+05
75%	2010.000000	2.446451e+06	2.610463e+06
max	2016.000000	4.616241e+08	4.856486e+08

In [1772]: df65.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1772]:

	Entity	Year	Prevalence - Mental health disorders: Male (Number)
6074	World	2016	461624118.3
6073	World	2015	456343688.0
6072	World	2014	450998468.4
6071	World	2013	445928760.7
6070	World	2012	440959579.5

In [1773]: df65.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1773]:

	Entity	Year	Prevalence - Mental health disorders: Female (Number)
6074	World	2016	485648639.6
6073	World	2015	479948775.9
6072	World	2014	474149534.3
6071	World	2013	468684720.8
6070	World	2012	463354573.7

```
In [1774]: df65.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[1774]:

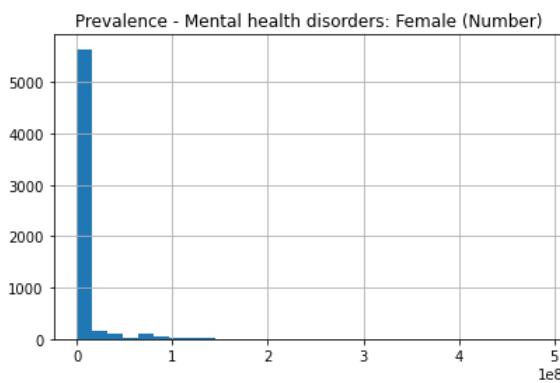
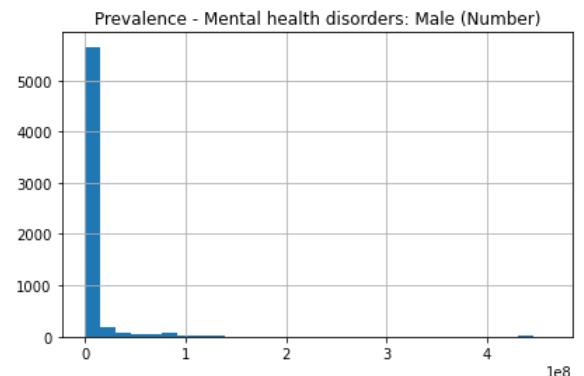
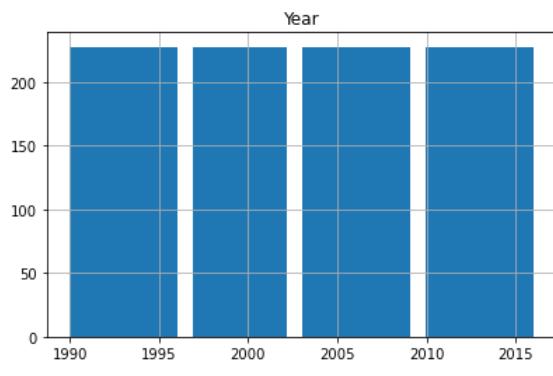
	Entity	Year	Prevalence - Mental health disorders: Male (Number)
3402	Marshall Islands	1990	2424.608477
81	American Samoa	1990	2465.052386
3403	Marshall Islands	1991	2486.142767
82	American Samoa	1991	2515.356582
3404	Marshall Islands	1992	2528.300632

```
In [1775]: df65.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()
```

Out[1775]:

	Entity	Year	Prevalence - Mental health disorders: Female (Number)
3402	Marshall Islands	1990	2155.059413
4050	Northern Mariana Islands	1990	2175.680512
3403	Marshall Islands	1991	2209.063287
81	American Samoa	1990	2243.445486
3404	Marshall Islands	1992	2257.810265

```
In [1776]: df65.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1777]: df65.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1777]: Entity
Marshall Islands    3094.845314
American Samoa     3250.909443
Greenland          3842.358671
Bermuda            3849.972585
Dominica           4228.805495
Name: Prevalence - Mental health disorders: Male (Number), dtype: float64
```

```
In [1778]: df65.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1778]: Entity
Marshall Islands      2814.812772
American Samoa        3074.303332
Greenland             3977.133079
Bermuda               4148.328952
Northern Mariana Islands 4310.547655
Name: Prevalence - Mental health disorders: Female (Number), dtype: float64
```

```
In [1779]: df65.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1779]: Entity
East Asia            8.451308e+07
South Asia           8.984385e+07
Low-middle SDI       1.094689e+08
Middle SDI          1.262261e+08
World                3.934162e+08
Name: Prevalence - Mental health disorders: Male (Number), dtype: float64
```

```
In [1780]: df65.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1780]: Entity
East Asia            8.249695e+07
South Asia           8.576008e+07
Low-middle SDI       1.065225e+08
Middle SDI          1.270422e+08
World                4.110676e+08
Name: Prevalence - Mental health disorders: Female (Number), dtype: float64
```

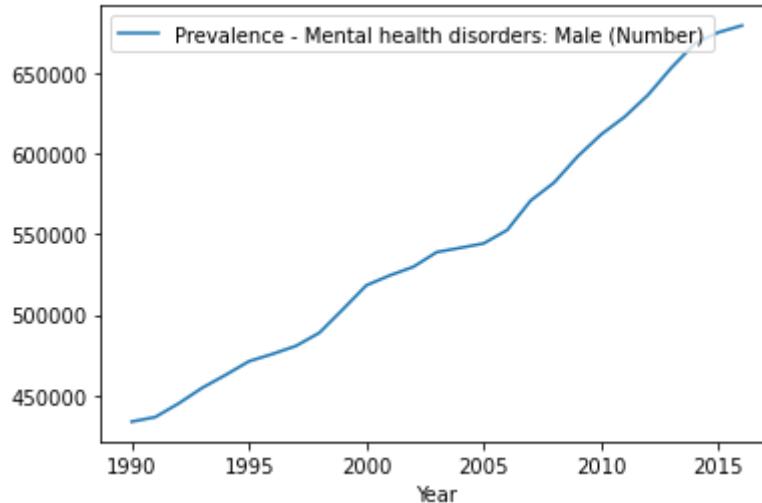
```
In [1781]: df65_mean = df65.groupby('Year').mean()  
df65_mean.head()
```

Out[1781]:

Year	Prevalence - Mental health disorders: Male (Number)	Prevalence - Mental health disorders: Female (Number)
1990	5.860018e+06	6.066180e+06
1991	5.957550e+06	6.170525e+06
1992	6.064214e+06	6.284241e+06
1993	6.175059e+06	6.401561e+06
1994	6.288587e+06	6.522248e+06

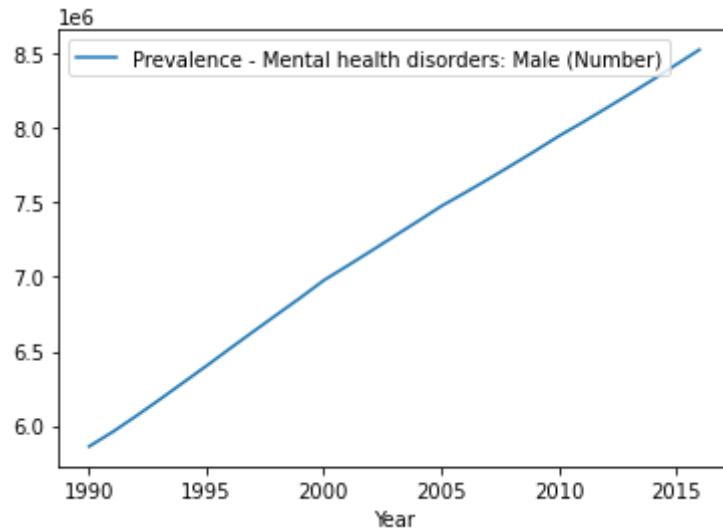
```
In [1782]: df65.groupby('Year')[v1].median().plot(legend=True)
```

Out[1782]: <AxesSubplot:xlabel='Year'>



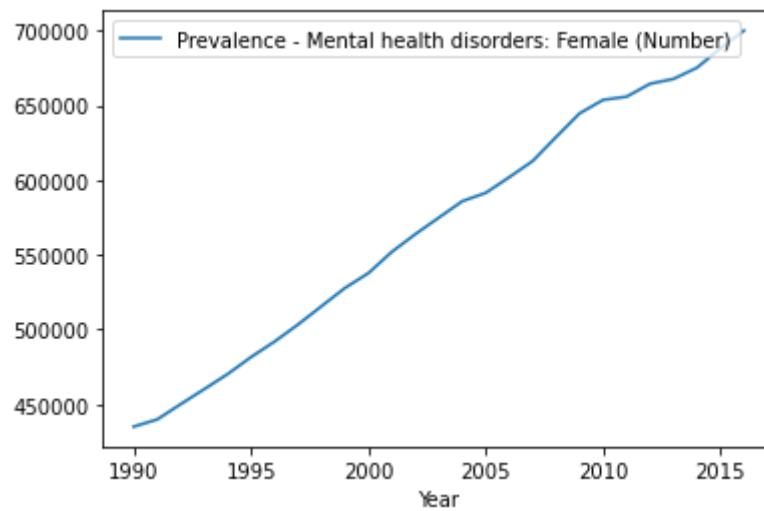
```
In [1783]: df65.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1783]: <AxesSubplot:xlabel='Year'>
```



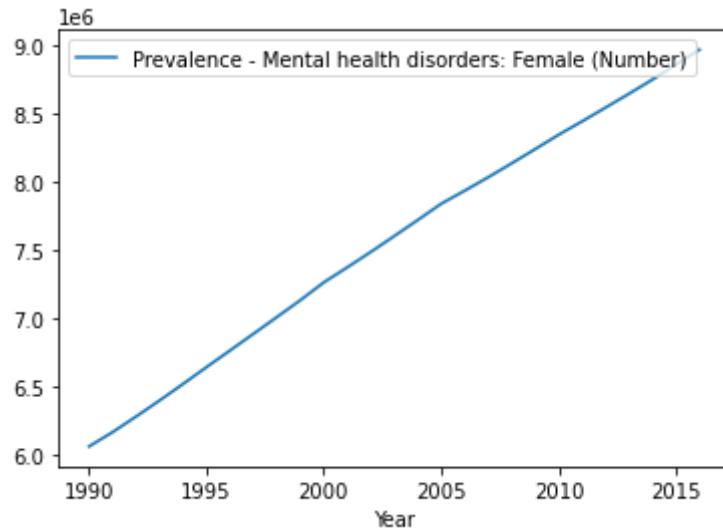
```
In [1784]: df65.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1784]: <AxesSubplot:xlabel='Year'>
```



```
In [1785]: df65.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1785]: <AxesSubplot:xlabel='Year'>
```



In [1786]: df69.info()
df69.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 10 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year        int64
6468 non-null   int64
3   Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)       6468 non-null   float64
4   Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)       6468 non-null   float64
5   Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)       6468 non-null   float64
6   Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)       6468 non-null   float64
7   Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)       6468 non-null   float64
8   Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)       6468 non-null   float64
9   Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)       6468 non-null   float64
dtypes: float64(7), int64(1), object(2)
memory usage: 505.4+ KB
```

Out[1786]:

	Entity	Code	Year	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)
0	Afghanistan	AFG	1990	0.160560	0.697779	0.101855	4.828830	1.677082
1	Afghanistan	AFG	1991	0.160312	0.697961	0.099313	4.829740	1.684746
2	Afghanistan	AFG	1992	0.160135	0.698107	0.096692	4.831108	1.694334
3	Afghanistan	AFG	1993	0.160037	0.698257	0.094336	4.830864	1.705320
4	Afghanistan	AFG	1994	0.160022	0.698469	0.092439	4.829423	1.716069

Checking for missing values:

```
In [1787]: missing = pd.concat([df69.isnull().sum(), 100 * df69.isnull().mean()], axis=1)
missing.columns=['count','%']
missing.sort_values(by='count', ascending=True)
```

Out[1787]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
	Code	980	15.151515

```
In [1788]: v1='Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent'
v2='Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent'
v3='Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent'
v4='Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent'
v5='Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent'
v6='Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent'
v7='Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent'
```

In [1789]: df69.describe()

Out[1789]:

	Year	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	0.211644	0.719145	0.239998	3.989921	0.862278	3.49
std	8.078372	0.044253	0.171589	0.158141	1.167526	0.460679	0.65
min	1990.000000	0.146902	0.314535	0.073908	2.023393	0.383650	2.13
25%	1996.750000	0.181530	0.615532	0.122387	3.188824	0.535064	3.00
50%	2003.500000	0.199563	0.693134	0.182525	3.554373	0.726430	3.49
75%	2010.250000	0.236365	0.835063	0.292666	4.682163	0.940157	3.91
max	2017.000000	0.375110	1.206597	0.943991	8.967330	3.452476	6.60

In [1790]: df69.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()

Out[1790]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)
3992	Netherlands	2006	0.375110
3991	Netherlands	2005	0.375092
3993	Netherlands	2007	0.375087
3994	Netherlands	2008	0.374991
3990	Netherlands	2004	0.374961

In [1791]: df69.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1791]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)
4027	New Zealand	2013	1.206597
4026	New Zealand	2012	1.206549
4028	New Zealand	2014	1.206502
4025	New Zealand	2011	1.206429
4029	New Zealand	2015	1.206423

In [1792]: df69.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1792]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)
334	Australia	2016	0.943991
333	Australia	2015	0.943206
335	Australia	2017	0.943081
332	Australia	2014	0.941661
331	Australia	2013	0.939176

In [1793]: df69.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1793]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)
4016	New Zealand	2002	8.967330
4017	New Zealand	2003	8.966183
4015	New Zealand	2001	8.965461
4018	New Zealand	2004	8.962565
4014	New Zealand	2000	8.959512

In [1794]: df69.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1794]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)
6103	United States	2017	3.452476
6102	United States	2016	3.353142
4171	North America	2017	3.338079
6101	United States	2015	3.262674
4170	North America	2016	3.245661

In [1795]: df69.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1795]:

	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)
2273	Greenland	1995	6.602754
2274	Greenland	1996	6.601250
2275	Greenland	1997	6.598258
2276	Greenland	1998	6.586734
2272	Greenland	1994	6.582469

In [1796]: df69.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1796]:

	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)
524	Belarus	2010	5.474668
525	Belarus	2011	5.473536
526	Belarus	2012	5.467508
523	Belarus	2009	5.461357
527	Belarus	2013	5.455420

In [1797]: df69.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1797]:

	Entity	Year	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)
3866	Mozambique	1992	0.146902
3867	Mozambique	1993	0.146907
3865	Mozambique	1991	0.146938
3868	Mozambique	1994	0.146949
3869	Mozambique	1995	0.147017

In [1798]: df69.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1798]:

	Entity	Year	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)
1260	China	1990	0.314535
1261	China	1991	0.314987
1680	East Asia	1990	0.315198
1262	China	1992	0.315466
1681	East Asia	1991	0.315645

In [1799]: df69.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1799]:

	Entity	Year	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)
3893	Myanmar	1991	0.073908
3892	Myanmar	1990	0.074078
3894	Myanmar	1992	0.074080
3895	Myanmar	1993	0.074227
3896	Myanmar	1994	0.074309

In [1800]: df69.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1800]:

	Entity	Year	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)
6259	Vietnam	2005	2.023393
6260	Vietnam	2006	2.023866
6258	Vietnam	2004	2.024291
6261	Vietnam	2007	2.024573
6262	Vietnam	2008	2.025459

In [1801]: df69.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1801]:

	Entity	Year	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)
705	Bosnia and Herzegovina	1995	0.383650
704	Bosnia and Herzegovina	1994	0.384059
706	Bosnia and Herzegovina	1996	0.384598
703	Bosnia and Herzegovina	1993	0.384720
702	Bosnia and Herzegovina	1992	0.386291

In [1802]: `df69.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()`

Out[1802]:

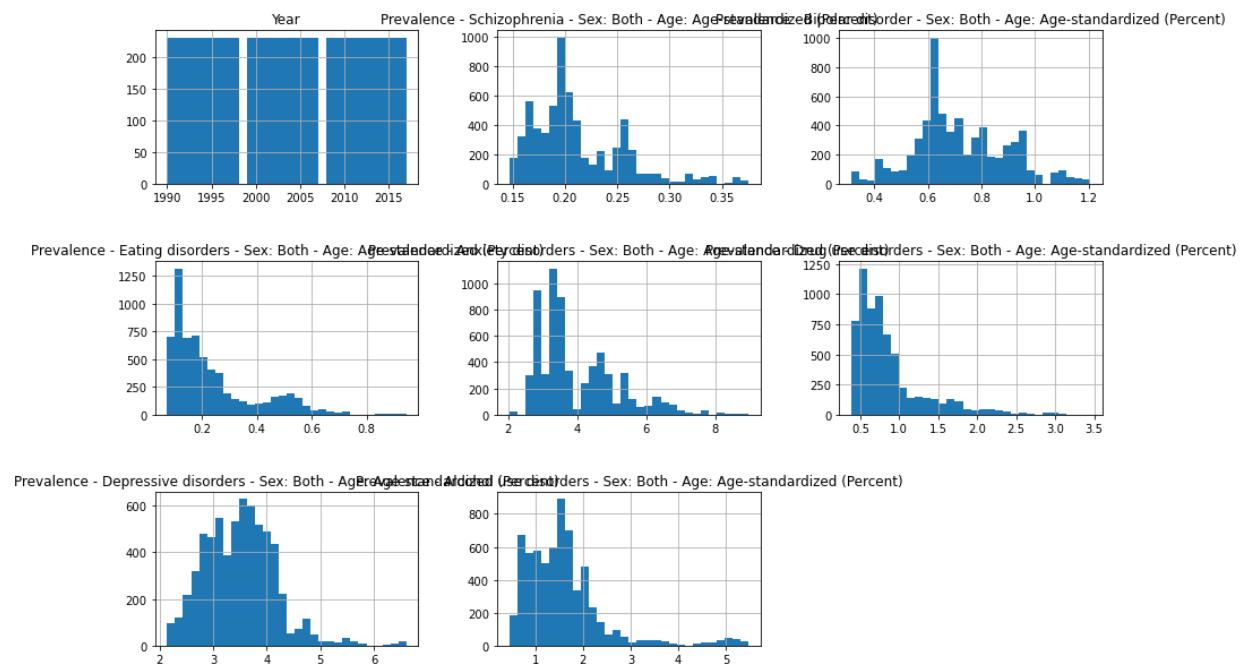
Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)				
	Entity	Year	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)	
28	Albania	1990	2.139903	
29	Albania	1991	2.141201	
30	Albania	1992	2.143395	
31	Albania	1993	2.145263	
32	Albania	1994	2.150070	

In [1803]: `df69.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()`

Out[1803]:

Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)				
	Entity	Year	Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)	
2861	Italy	1995	0.446940	
2862	Italy	1996	0.447098	
2860	Italy	1994	0.447662	
2863	Italy	1997	0.447723	
2864	Italy	1998	0.448558	

In [1804]: `df69.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);`



```
In [1805]: df69.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1805]: Entity
Central African Republic      0.148887
Mozambique                     0.150924
Malawi                         0.152340
Somalia                        0.153407
Burundi                        0.153603
Name: Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1806]: df69.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1806]: Entity
China                          0.319829
East Asia                      0.320392
North Korea                    0.325294
Taiwan                         0.350378
Southeast Asia, East Asia, and Oceania 0.379516
Name: Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1807]: df69.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1807]: Entity
Somalia                        0.078287
Liberia                         0.084655
Ethiopia                        0.086306
Mozambique                     0.086969
Democratic Republic of Congo   0.088121
Name: Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1808]: df69.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1808]: Entity
Vietnam                         2.035264
Tajikistan                      2.512893
Mongolia                        2.526100
Uzbekistan                     2.541930
Kyrgyzstan                      2.553457
Name: Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1809]: df69.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1809]: Entity
Bosnia and Herzegovina        0.412833
Burkina Faso                  0.416361
Mali                           0.426686
Guinea                        0.429939
Chad                           0.437192
Name: Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1810]: df69.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1810]: Entity
Albania      2.190442
Myanmar      2.252113
Poland       2.284375
Colombia     2.311194
Peru         2.317500
Name: Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1811]: df69.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1811]: Entity
Italy        0.456138
Singapore    0.479829
Israel       0.533354
Morocco      0.563468
Japan        0.565542
Name: Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1812]: df69.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1812]: Entity
United States 0.338117
New Zealand   0.340113
Australasia   0.360540
Australia     0.364498
Netherlands   0.372330
Name: Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1813]: df69.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1813]: Entity
England       1.105841
Northern Ireland 1.142926
Australia     1.144660
Australasia   1.153425
New Zealand   1.199014
Name: Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1814]: df69.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1814]: Entity
New Zealand   0.643235
Spain          0.664637
Luxembourg    0.686169
Australasia   0.806650
Australia     0.838270
Name: Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent), dtype: float64
```

```
In [1815]: df69.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1815]: Entity
Australasia      6.895850
Iran             6.932546
Norway           7.637363
Northern Ireland 7.826927
New Zealand      8.651309
Name: Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized
(Percent), dtype: float64
```

```
In [1816]: df69.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1816]: Entity
Australia        2.194212
Libya            2.197783
United Arab Emirates 2.631126
North America    2.777057
United States    2.844713
Name: Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized
(Percent), dtype: float64
```

```
In [1817]: df69.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1817]: Entity
Finland          5.183080
Uganda           5.227817
Lesotho          5.423811
Morocco          5.610298
Greenland        6.459109
Name: Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized
(Percent), dtype: float64
```

```
In [1818]: df69.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1818]: Entity
Estonia          4.917451
Ukraine          5.059726
Eastern Europe   5.106033
Russia           5.158491
Belarus          5.286263
Name: Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized
(Percent), dtype: float64
```

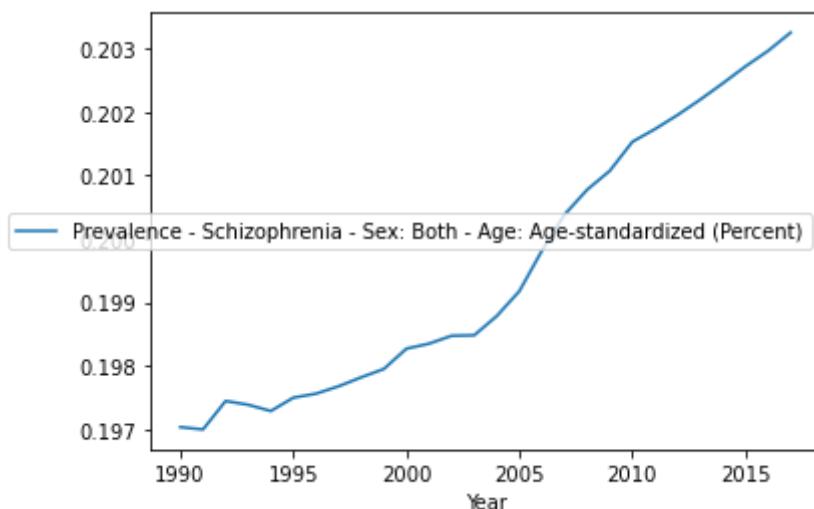
```
In [1819]: df69_mean = df69.groupby('Year').mean()
df69_mean.head()
```

Out[1819]:

Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Alcohol disorders - Sex: Both - Age: Age-standardized (Percent)
Year						
1990	0.209548	0.715392	0.221774	3.957269	0.808283	3.506288
1991	0.209586	0.715740	0.222076	3.960009	0.813466	3.510948
1992	0.209634	0.716091	0.222481	3.962778	0.818692	3.515033
1993	0.209690	0.716430	0.223033	3.965405	0.823780	3.518531
1994	0.209751	0.716755	0.223710	3.967976	0.828389	3.521437

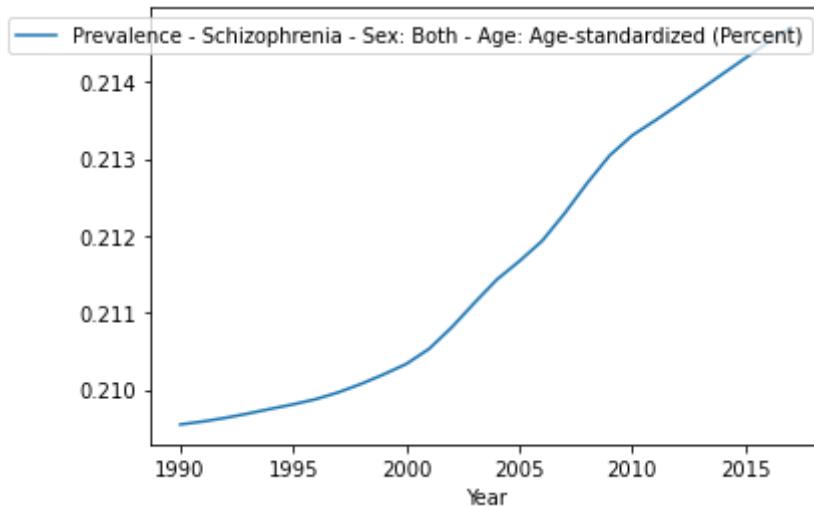
```
In [1820]: df69.groupby('Year')[v1].median().plot(legend=True)
```

Out[1820]: <AxesSubplot: xlabel='Year'>



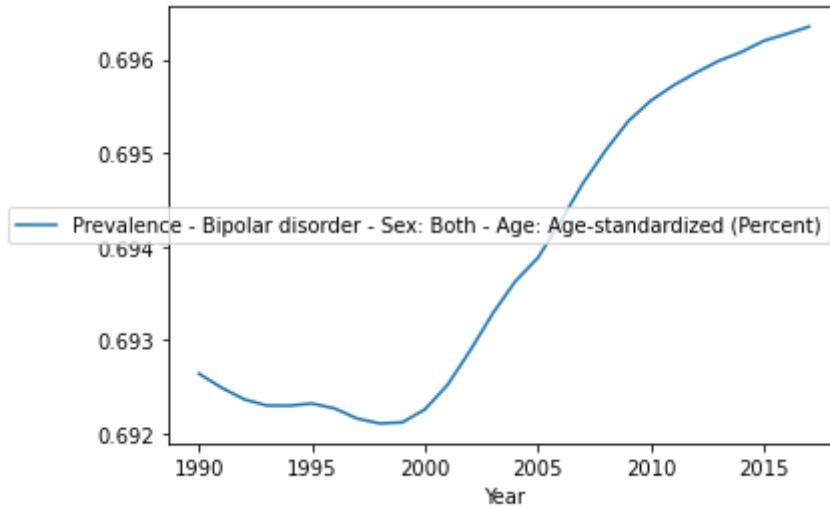
```
In [1821]: df69.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1821]: <AxesSubplot:xlabel='Year'>
```



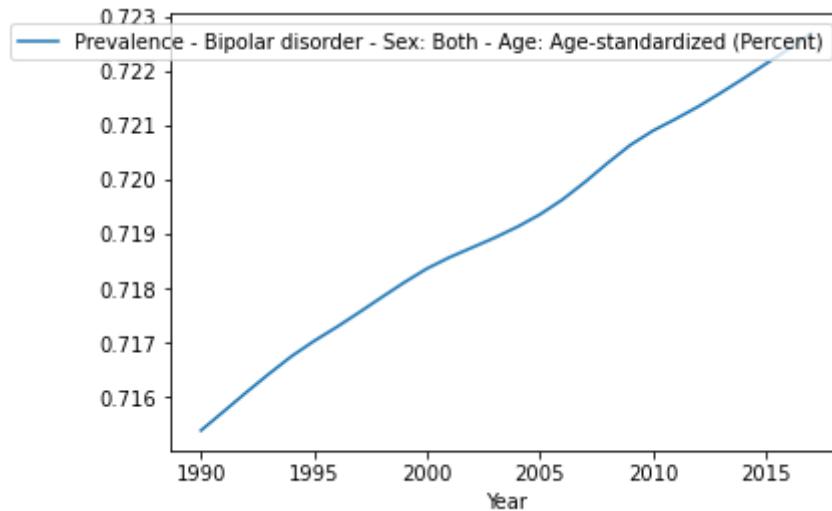
```
In [1822]: df69.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1822]: <AxesSubplot:xlabel='Year'>
```



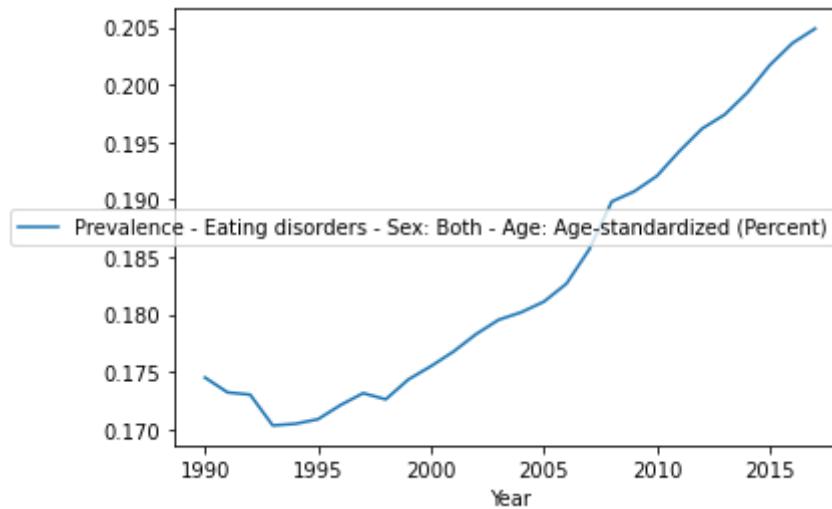
```
In [1823]: df69.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1823]: <AxesSubplot:xlabel='Year'>
```



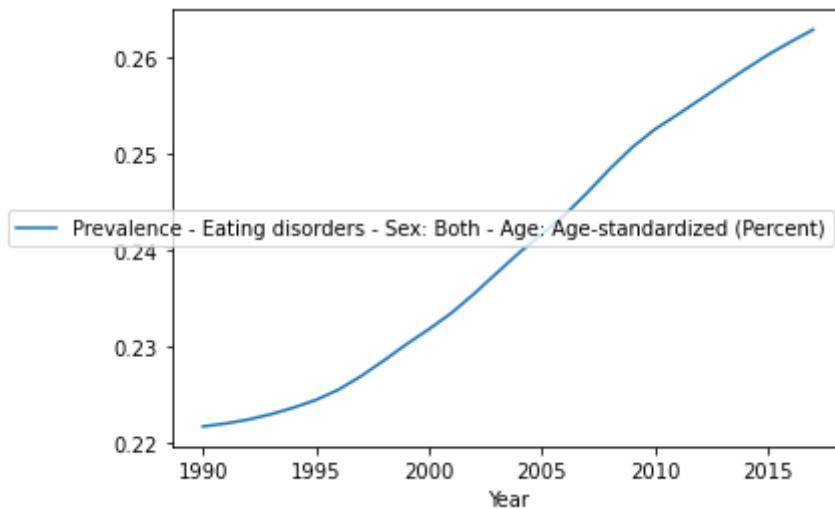
```
In [1824]: df69.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1824]: <AxesSubplot:xlabel='Year'>
```



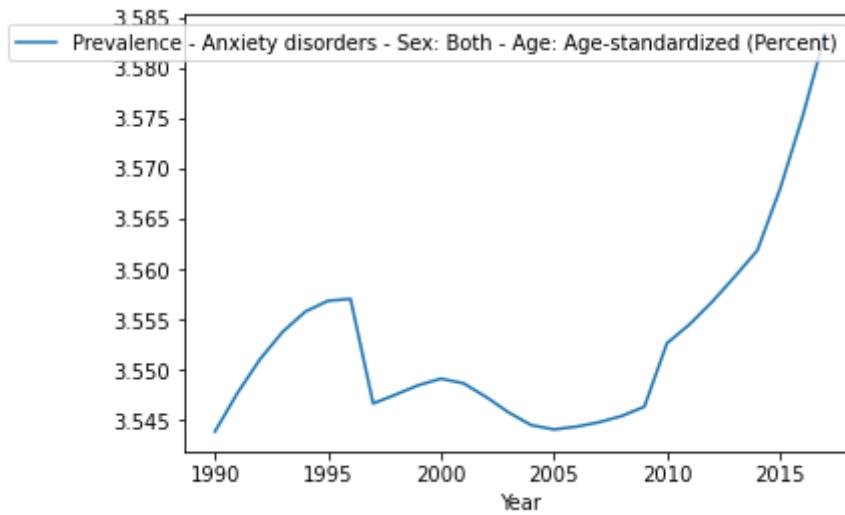
```
In [1825]: df69.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1825]: <AxesSubplot:xlabel='Year'>
```



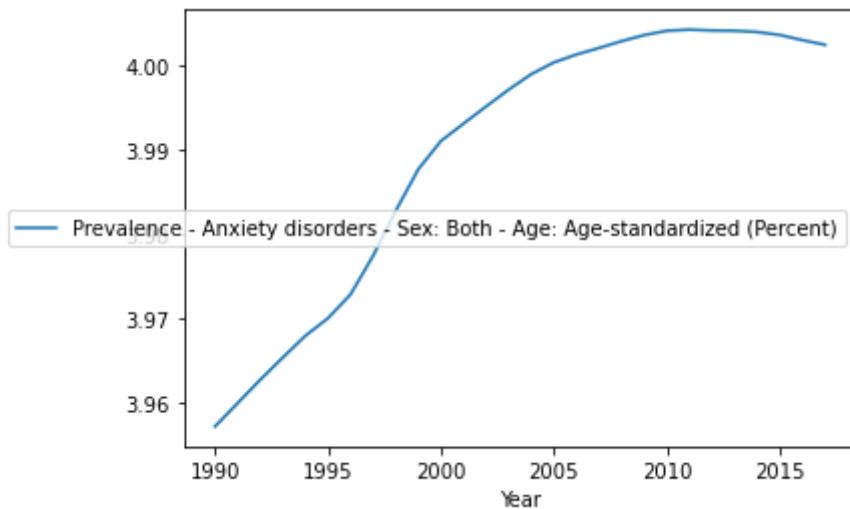
```
In [1826]: df69.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1826]: <AxesSubplot:xlabel='Year'>
```



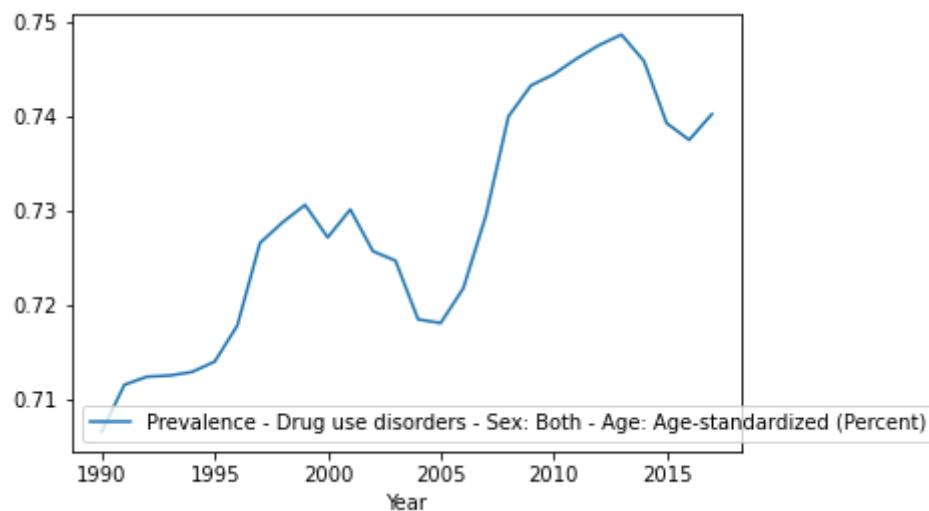
```
In [1827]: df69.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1827]: <AxesSubplot:xlabel='Year'>
```



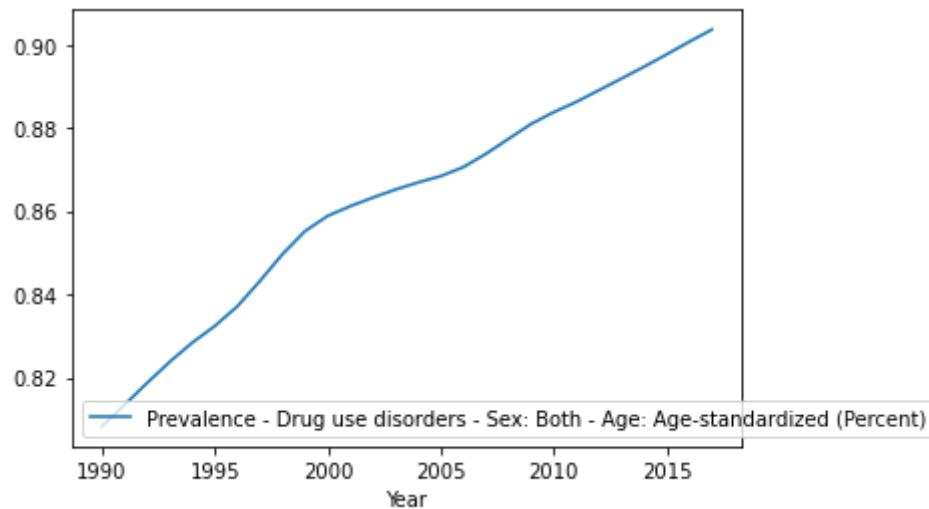
```
In [1828]: df69.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1828]: <AxesSubplot:xlabel='Year'>
```



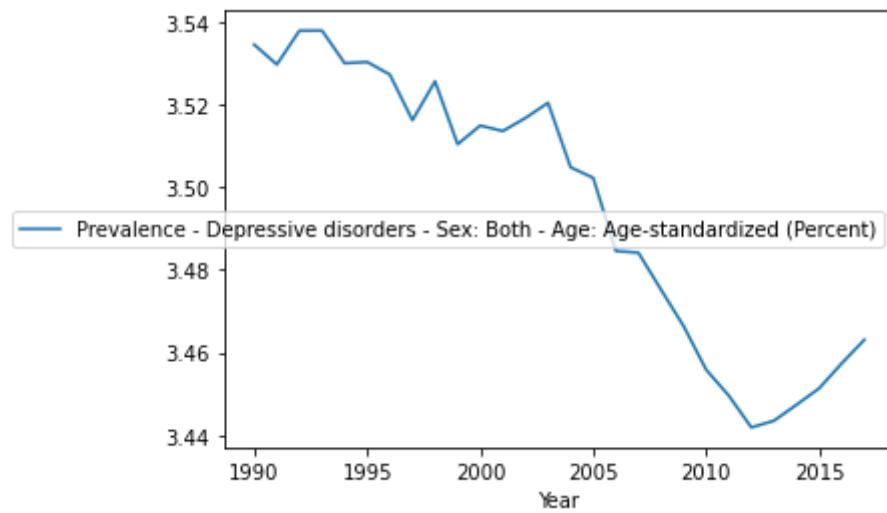
```
In [1829]: df69.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1829]: <AxesSubplot:xlabel='Year'>
```



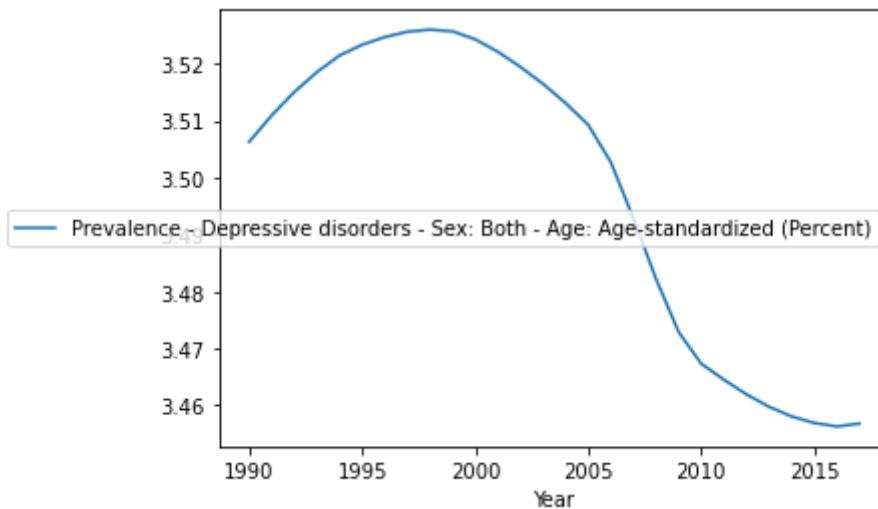
```
In [1830]: df69.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1830]: <AxesSubplot:xlabel='Year'>
```



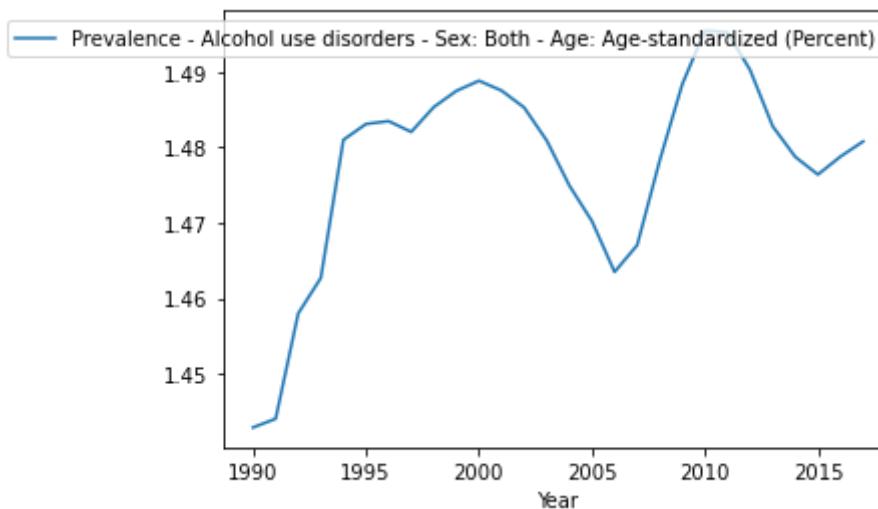
```
In [1831]: df69.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1831]: <AxesSubplot:xlabel='Year'>
```



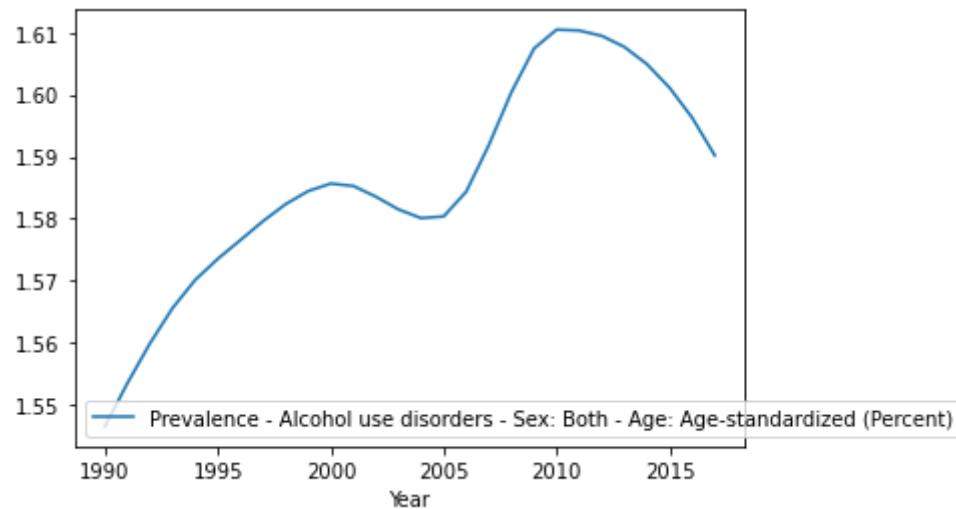
```
In [1832]: df69.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1832]: <AxesSubplot:xlabel='Year'>
```



```
In [1833]: df69.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1833]: <AxesSubplot:xlabel='Year'>
```



In [1834]: df84.info()
df84.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 14 columns):
 #   Column
Non-Null Count Dtype
---  -----
0   Entity      object
6468 non-null   object
1   Code        object
5488 non-null   object
2   Year         int64
6468 non-null   int64
3   Prevalence - Mental and substance use disorders - Sex: Both - Age: 1
0 to 14 (Percent)       6468 non-null   float64
4   Prevalence - Mental and substance use disorders - Sex: Both - Age: 2
5 to 29 (Percent)       6468 non-null   float64
5   Prevalence - Mental and substance use disorders - Sex: Both - Age: 2
0 to 24 (Percent)       6468 non-null   float64
6   Prevalence - Mental and substance use disorders - Sex: Both - Age: 3
0 to 34 (Percent)       6468 non-null   float64
7   Prevalence - Mental and substance use disorders - Sex: Both - Age: All
Ages (Percent)          6468 non-null   float64
8   Prevalence - Mental and substance use disorders - Sex: Both - Age: 1
5 to 19 (Percent)       6468 non-null   float64
9   Prevalence - Mental and substance use disorders - Sex: Both - Age: 7
0+ years (Percent)     6468 non-null   float64
10  Prevalence - Mental and substance use disorders - Sex: Both - Age: 1
5-49 years (Percent)   6468 non-null   float64
11  Prevalence - Mental and substance use disorders - Sex: Both - Age: 5
-14 years (Percent)    6468 non-null   float64
12  Prevalence - Mental and substance use disorders - Sex: Both - Age: 5
0-69 years (Percent)   6468 non-null   float64
13  Prevalence - Mental and substance use disorders - Sex: Both - Age: All
ge-standardized (Percent) 6468 non-null   float64
dtypes: float64(11), int64(1), object(2)
memory usage: 707.6+ KB
```

Out[1834]:

	Entity	Code	Year	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence
				- Mental and substance use					
0	Afghanistan	AFG	1990	17.963929	20.968224	20.394054	21.363230	16.471540	20.746551
1	Afghanistan	AFG	1991	18.308466	21.226918	20.690527	21.628881	16.750416	21.059514
2	Afghanistan	AFG	1992	18.618200	21.437639	20.975017	21.855734	17.022630	21.360625

Entity	Code	Year	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence	Prevalence
			- Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)	- Mental and substance use disorders - Sex: Both - Age: 25 to 29 (Percent)	- Mental and substance use disorders - Sex: Both - Age: 20 to 24 (Percent)	- Mental and substance use disorders - Sex: Both - Age: 30 to 34 (Percent)	- Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)	- Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)
3	Afghanistan	AFG	1993	18.854464	21.595253	21.184427	22.046316	17.208295
4	Afghanistan	AFG	1994	19.022019	21.725390	21.313204	22.158546	17.230492

Checking for missing values:

```
In [1835]: missing = pd.concat([df84.isnull().sum(), 100 * df84.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1835]:

		count	%
	Entity	0	0.000000
	Year	0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 25 to 29 (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 20 to 24 (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 30 to 34 (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 70+ years (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Percent)		0	0.000000
Prevalence - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Percent)		0	0.000000
	Code	980	15.151515

```
In [1836]: v1='Prevalence - Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)'
v2='Prevalence - Mental and substance use disorders - Sex: Both - Age: 25 to 34 (Percent)'
v3='Prevalence - Mental and substance use disorders - Sex: Both - Age: 20 to 29 (Percent)'
v4='Prevalence - Mental and substance use disorders - Sex: Both - Age: 30 to 39 (Percent)'
v5='Prevalence - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)'
v6='Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)'
v7='Prevalence - Mental and substance use disorders - Sex: Both - Age: 70+ (Percent)'
v8='Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 49 (Percent)'
v9='Prevalence - Mental and substance use disorders - Sex: Both - Age: 5 to 14 (Percent)'
v10='Prevalence - Mental and substance use disorders - Sex: Both - Age: 50 to 64 (Percent)'
v11='Prevalence - Mental and substance use disorders - Sex: Both - Age: Age Standardized (Percent)'
```

```
In [1837]: df84.describe()
```

Out[1837]:

	Prevalence - Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 25 to 34 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 20 to 29 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 30 to 39 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)
count	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000	6468.000000
mean	2003.500000	13.843204	14.829490	14.865825	15.090616	12.779463
std	8.078372	2.133417	2.586980	2.676551	2.530750	2.127102
min	1990.000000	9.804068	10.196966	9.851233	10.736103	9.304492
25%	1996.750000	12.304926	12.717292	12.692713	13.016868	11.009757
50%	2003.500000	13.169728	14.164913	14.191402	14.420812	12.241130

```
In [1838]: df84.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[1838]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)
4019	New Zealand	2005	21.570803
4018	New Zealand	2004	21.551895
4020	New Zealand	2006	21.550192
4017	New Zealand	2003	21.515708
4021	New Zealand	2007	21.489496

In [1839]: df84.sort_values(by=v2, ascending=False)[['Entity', 'Year', v2]].head()

Out[1839]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 25 to 29 (Percent)
4015	New Zealand	2001	24.480835
4014	New Zealand	2000	24.477077
4017	New Zealand	2003	24.476785
4016	New Zealand	2002	24.475263
4018	New Zealand	2004	24.466427

In [1840]: df84.sort_values(by=v3, ascending=False)[['Entity', 'Year', v3]].head()

Out[1840]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 20 to 24 (Percent)
4014	New Zealand	2000	25.081546
4015	New Zealand	2001	25.067472
4016	New Zealand	2002	25.054657
4013	New Zealand	1999	25.047430
4017	New Zealand	2003	25.041765

In [1841]: df84.sort_values(by=v4, ascending=False)[['Entity', 'Year', v4]].head()

Out[1841]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 30 to 34 (Percent)
4019	New Zealand	2005	23.860110
4018	New Zealand	2004	23.856325
4017	New Zealand	2003	23.847496
4016	New Zealand	2002	23.824051
4015	New Zealand	2001	23.806772

In [1842]: df84.sort_values(by=v5, ascending=False)[['Entity', 'Year', v5]].head()

Out[1842]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)
4018	New Zealand	2004	19.121172
4019	New Zealand	2005	19.120496
4017	New Zealand	2003	19.109256
4016	New Zealand	2002	19.094076
4015	New Zealand	2001	19.083385

In [1843]: df84.sort_values(by=v6, ascending=False)[['Entity', 'Year', v6]].head()

Out[1843]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)
4019	New Zealand	2005	24.957525
4018	New Zealand	2004	24.956186
4017	New Zealand	2003	24.948015
4015	New Zealand	2001	24.930757
4016	New Zealand	2002	24.929691

In [1844]: df84.sort_values(by=v7, ascending=False)[['Entity', 'Year', v7]].head()

Out[1844]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 70+ years (Percent)
5976	Uganda	2002	17.860393
5975	Uganda	2001	17.858938
5977	Uganda	2003	17.848988
5974	Uganda	2000	17.848758
5978	Uganda	2004	17.838659

In [1845]: df84.sort_values(by=v8, ascending=False)[['Entity', 'Year', v8]].head()

Out[1845]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Percent)
4014	New Zealand	2000	23.275827
4015	New Zealand	2001	23.275161
4016	New Zealand	2002	23.269874
4017	New Zealand	2003	23.269592
4018	New Zealand	2004	23.265574

In [1846]: df84.sort_values(by=v9, ascending=False)[['Entity', 'Year', v9]].head()

Out[1846]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Percent)
323	Australia	2005	17.027704
324	Australia	2006	17.020454
295	Australasia	2005	17.019664
296	Australasia	2006	17.012759
322	Australia	2004	17.003319

In [1847]: df84.sort_values(by=v10, ascending=False)[['Entity', 'Year', v10]].head()

Out[1847]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Percent)
2765	Iran	2011	19.423087
2764	Iran	2010	19.423018
2766	Iran	2012	19.414361
2767	Iran	2013	19.398764
2763	Iran	2009	19.392922

In [1848]: df84.sort_values(by=v11, ascending=False)[['Entity', 'Year', v11]].head()

Out[1848]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Percent)
4019	New Zealand	2005	19.115460
4018	New Zealand	2004	19.108603
4017	New Zealand	2003	19.094315
4016	New Zealand	2002	19.076815
4020	New Zealand	2006	19.074092

In [1849]: df84.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()

Out[1849]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)
6267	Vietnam	2013	9.804068
6268	Vietnam	2014	9.805702
6266	Vietnam	2012	9.807570
6269	Vietnam	2015	9.820463
6265	Vietnam	2011	9.821922

In [1850]: df84.sort_values(by=v2, ascending=True)[['Entity', 'Year', v2]].head()

Out[1850]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 25 to 29 (Percent)
6270	Vietnam	2016	10.196966
6271	Vietnam	2017	10.199592
6269	Vietnam	2015	10.221545
6268	Vietnam	2014	10.253529
6267	Vietnam	2013	10.272351

In [1851]: df84.sort_values(by=v3, ascending=True)[['Entity', 'Year', v3]].head()

Out[1851]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 20 to 24 (Percent)
6271	Vietnam	2017	9.851233
6270	Vietnam	2016	9.854900
6269	Vietnam	2015	9.872112
6268	Vietnam	2014	9.886086
6267	Vietnam	2013	9.906359

In [1852]: df84.sort_values(by=v4, ascending=True)[['Entity', 'Year', v4]].head()

Out[1852]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 30 to 34 (Percent)
6271	Vietnam	2017	10.736103
6270	Vietnam	2016	10.754024
6269	Vietnam	2015	10.781917
6268	Vietnam	2014	10.801679
6267	Vietnam	2013	10.818112

In [1853]: df84.sort_values(by=v5, ascending=True)[['Entity', 'Year', v5]].head()

Out[1853]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)
2604	Honduras	1990	9.304492
2605	Honduras	1991	9.343707
2606	Honduras	1992	9.376895
2607	Honduras	1993	9.413546
2608	Honduras	1994	9.448293

In [1854]: df84.sort_values(by=v6, ascending=True)[['Entity', 'Year', v6]].head()

Out[1854]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)
6267	Vietnam	2013	10.998515
6268	Vietnam	2014	10.999883
6270	Vietnam	2016	11.008862
6269	Vietnam	2015	11.014723
6266	Vietnam	2012	11.021167

In [1855]: df84.sort_values(by=v7, ascending=True)[['Entity', 'Year', v7]].head()

Out[1855]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 70+ years (Percent)
1315	Colombia	2017	9.662779
1314	Colombia	2016	9.666067
1313	Colombia	2015	9.670499
1312	Colombia	2014	9.674853
1311	Colombia	2013	9.681401

In [1856]: df84.sort_values(by=v8, ascending=True)[['Entity', 'Year', v8]].head()

Out[1856]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Percent)
6270	Vietnam	2016	10.873140
6271	Vietnam	2017	10.876976
6269	Vietnam	2015	10.885297
6268	Vietnam	2014	10.896684
6267	Vietnam	2013	10.912898

In [1857]: df84.sort_values(by=v9, ascending=True)[['Entity', 'Year', v9]].head()

Out[1857]:

	Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Percent)
6269	Vietnam	2015	7.111707
6270	Vietnam	2016	7.113488
6268	Vietnam	2014	7.126671
6271	Vietnam	2017	7.138508
6267	Vietnam	2013	7.158568

```
In [1858]: df84.sort_values(by=v10, ascending=True)[['Entity', 'Year', v10]].head()
```

Out[1858]:

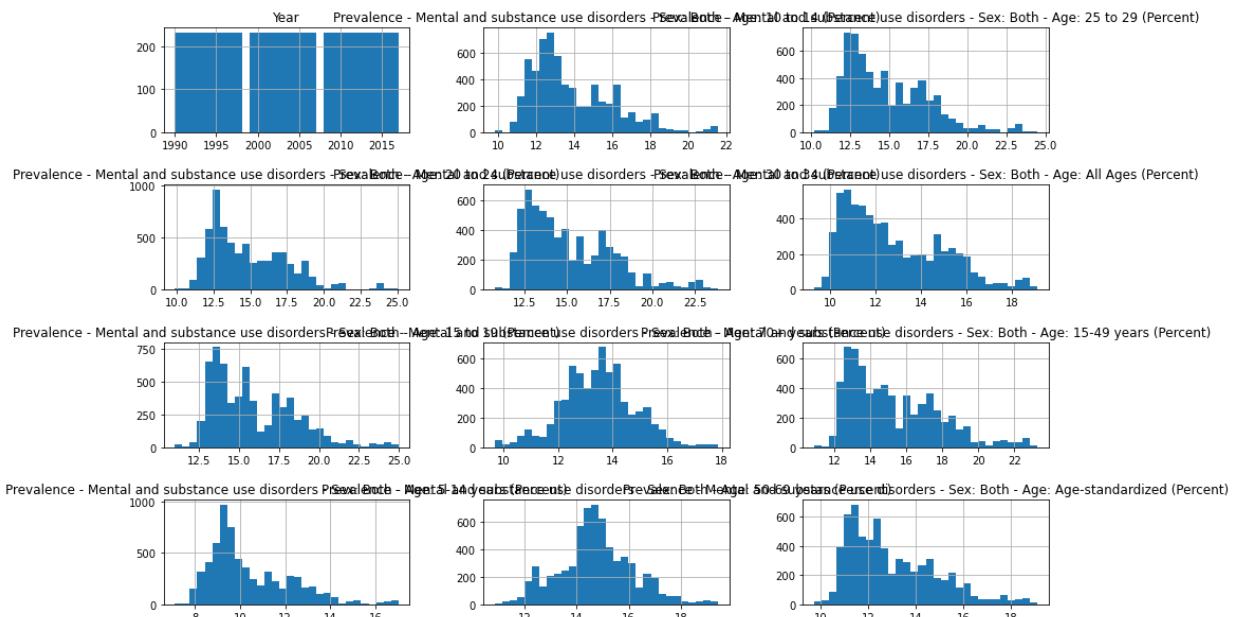
Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Percent)	
1315	Colombia	2017	10.836164
1314	Colombia	2016	10.838537
1313	Colombia	2015	10.845660
1312	Colombia	2014	10.859117
1311	Colombia	2013	10.872313

```
In [1859]: df84.sort_values(by=v11, ascending=True)[['Entity', 'Year', v11]].head()
```

Out[1859]:

Entity	Year	Prevalence - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Percent)	
6270	Vietnam	2016	9.715255
6271	Vietnam	2017	9.715823
6269	Vietnam	2015	9.726278
6268	Vietnam	2014	9.737802
6267	Vietnam	2013	9.754010

```
In [1860]: df84.hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1861]: df84.groupby('Entity')[v1].mean().sort_values().head()
```

```
Out[1861]: Entity
Vietnam      10.176164
Colombia    10.969129
Bhutan       10.994251
Honduras     11.025915
Mexico        11.051133
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
10 to 14 (Percent), dtype: float64
```

```
In [1862]: df84.groupby('Entity')[v2].mean().sort_values().head()
```

```
Out[1862]: Entity
Vietnam      10.757359
Seychelles   11.445830
American Samoa 11.574340
Azerbaijan   11.575041
Taiwan        11.589039
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
25 to 29 (Percent), dtype: float64
```

```
In [1863]: df84.groupby('Entity')[v3].mean().sort_values().head()
```

```
Out[1863]: Entity
Vietnam      10.418287
Seychelles   11.260056
Indonesia    11.362460
Taiwan        11.433606
Azerbaijan   11.505024
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
20 to 24 (Percent), dtype: float64
```

```
In [1864]: df84.groupby('Entity')[v4].mean().sort_values().head()
```

```
Out[1864]: Entity
Vietnam      11.275817
Azerbaijan   11.818769
American Samoa 11.908028
Tonga         11.965652
Seychelles    11.992333
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
30 to 34 (Percent), dtype: float64
```

```
In [1865]: df84.groupby('Entity')[v5].mean().sort_values().head()
```

```
Out[1865]: Entity
Honduras     9.865474
Mali          9.871461
Vietnam      10.022407
Colombia     10.066137
Tajikistan    10.170773
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
All Ages (Percent), dtype: float64
```

```
In [1866]: df84.groupby('Entity')[v6].mean().sort_values().head()
```

```
Out[1866]: Entity
Vietnam           11.438094
Seychelles        12.459693
Indonesia         12.633285
Philippines       12.692561
Southeast Asia    12.784996
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
15 to 19 (Percent), dtype: float64
```

```
In [1867]: df84.groupby('Entity')[v7].mean().sort_values().head()
```

```
Out[1867]: Entity
Colombia          9.820737
Brunei            9.896426
Japan              10.270327
High-income Asia Pacific 10.393529
Nicaragua          10.649932
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
70+ years (Percent), dtype: float64
```

```
In [1868]: df84.groupby('Entity')[v8].mean().sort_values().head()
```

```
Out[1868]: Entity
Vietnam            11.344128
Azerbaijan        11.945262
Armenia            12.146840
Georgia            12.172411
Colombia           12.203750
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
15-49 years (Percent), dtype: float64
```

```
In [1869]: df84.groupby('Entity')[v9].mean().sort_values().head()
```

```
Out[1869]: Entity
Vietnam            7.678530
Mexico             7.948963
Honduras           8.039466
Colombia           8.046080
Seychelles          8.105304
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
5-14 years (Percent), dtype: float64
```

```
In [1870]: df84.groupby('Entity')[v10].mean().sort_values().head()
```

```
Out[1870]: Entity
Colombia           11.164292
Brunei            11.540830
Japan              11.828373
High-income Asia Pacific  11.895286
Panama             12.056992
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
50-69 years (Percent), dtype: float64
```

```
In [1871]: df84.groupby('Entity')[v11].mean().sort_values().head()
```

```
Out[1871]: Entity
Colombia           10.146276
Vietnam            10.162393
Azerbaijan        10.491635
Honduras           10.619798
Central Latin America  10.666722
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
Age-standardized (Percent), dtype: float64
```

```
In [1872]: df84.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1872]: Entity
Spain               18.754521
Northern Ireland   19.596173
Australia          21.069148
Australasia        21.104724
New Zealand         21.271165
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
10 to 14 (Percent), dtype: float64
```

```
In [1873]: df84.groupby('Entity')[v2].mean().sort_values().tail()
```

```
Out[1873]: Entity
Iran                21.753228
Australia          22.956184
Australasia        23.084285
Greenland           23.130499
New Zealand         23.747508
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
25 to 29 (Percent), dtype: float64
```

```
In [1874]: df84.groupby('Entity')[v3].mean().sort_values().tail()
```

```
Out[1874]: Entity
Iran                21.238034
Australia          23.754850
Greenland           23.819164
Australasia        23.871663
New Zealand         24.459045
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
20 to 24 (Percent), dtype: float64
```

```
In [1875]: df84.groupby('Entity')[v4].mean().sort_values().tail()
```

```
Out[1875]: Entity
Iran           22.215898
Australia      22.499743
Australasia    22.605592
Greenland      22.783500
New Zealand    23.127719
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
30 to 34 (Percent), dtype: float64
```

```
In [1876]: df84.groupby('Entity')[v5].mean().sort_values().tail()
```

```
Out[1876]: Entity
Northern Ireland 18.151573
Greenland        18.430469
Australia        18.437124
Australasia      18.497070
New Zealand      18.797728
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
All Ages (Percent), dtype: float64
```

```
In [1877]: df84.groupby('Entity')[v6].mean().sort_values().tail()
```

```
Out[1877]: Entity
Northern Ireland 22.093731
Greenland        23.340912
Australia        24.135807
Australasia      24.220525
New Zealand      24.624464
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
15 to 19 (Percent), dtype: float64
```

```
In [1878]: df84.groupby('Entity')[v7].mean().sort_values().tail()
```

```
Out[1878]: Entity
Estonia         16.121839
South Asia      16.138418
India           16.312168
Lesotho          17.248295
Uganda          17.606749
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
70+ years (Percent), dtype: float64
```

```
In [1879]: df84.groupby('Entity')[v8].mean().sort_values().tail()
```

```
Out[1879]: Entity
Iran           21.725414
Australia      22.304898
Australasia    22.388173
Greenland      22.654296
New Zealand    22.804802
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
15-49 years (Percent), dtype: float64
```

```
In [1880]: df84.groupby('Entity')[v9].mean().sort_values().tail()
```

```
Out[1880]: Entity
Northern Ireland    14.993887
Spain                15.078240
New Zealand          16.602702
Australasia          16.660496
Australia            16.673548
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
5-14 years (Percent), dtype: float64
```

```
In [1881]: df84.groupby('Entity')[v10].mean().sort_values().tail()
```

```
Out[1881]: Entity
India                17.702876
Afghanistan          17.990487
Northern Ireland     18.506918
Morocco              18.779953
Iran                 19.212029
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
50-69 years (Percent), dtype: float64
```

```
In [1882]: df84.groupby('Entity')[v11].mean().sort_values().tail()
```

```
Out[1882]: Entity
Greenland            17.746121
Iran                 17.811353
Australia            18.436703
Australasia          18.499496
New Zealand          18.819918
Name: Prevalence - Mental and substance use disorders - Sex: Both - Age:
Age-standardized (Percent), dtype: float64
```

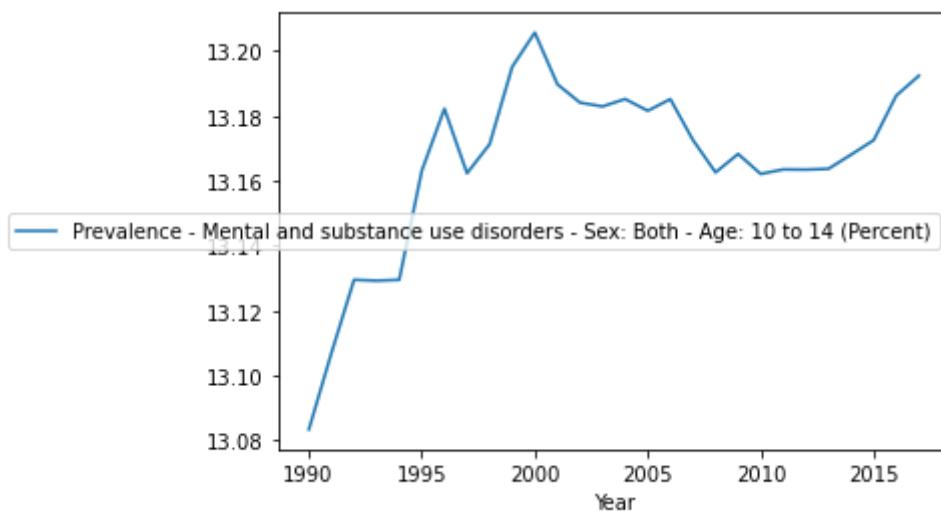
```
In [1883]: df84_mean = df84.groupby('Year').mean()
df84_mean.head()
```

Out[1883]:

	Prevalence - Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 25 to 29 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 20 to 24 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 30 to 34 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)
Year								
1990	13.792973	14.915774	14.927280	15.179913	12.518224	15.892375	13.537861	15.2936
1991	13.798817	14.921970	14.940712	15.183216	12.543190	15.897942	13.539874	15.2986
1992	13.804283	14.925606	14.950233	15.184791	12.568096	15.902002	13.541843	15.3016
1993	13.808488	14.927787	14.958451	15.184553	12.594012	15.906823	13.543761	15.3036
1994	13.812067	14.928909	14.962120	15.182701	12.619545	15.909549	13.544275	15.3046

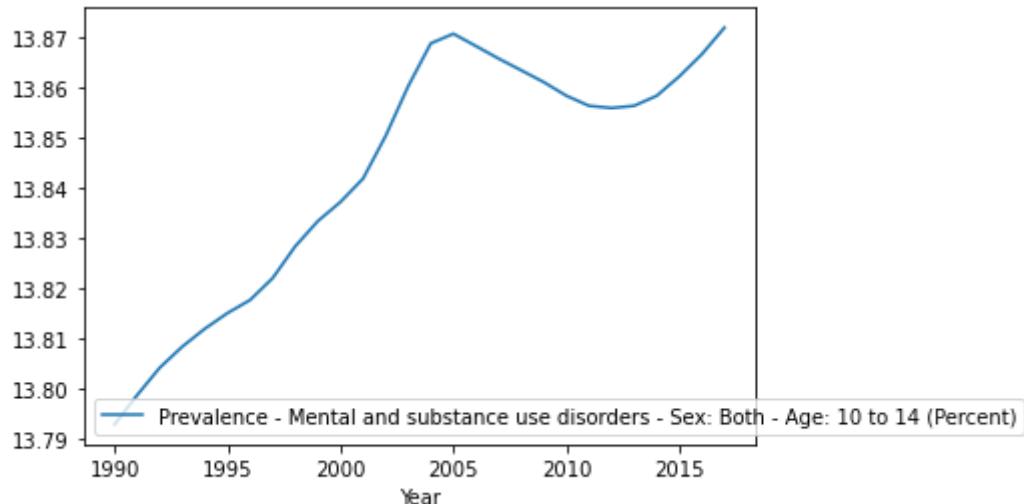
```
In [1884]: df84.groupby('Year')[v1].median().plot(legend=True)
```

Out[1884]: <AxesSubplot:xlabel='Year'>



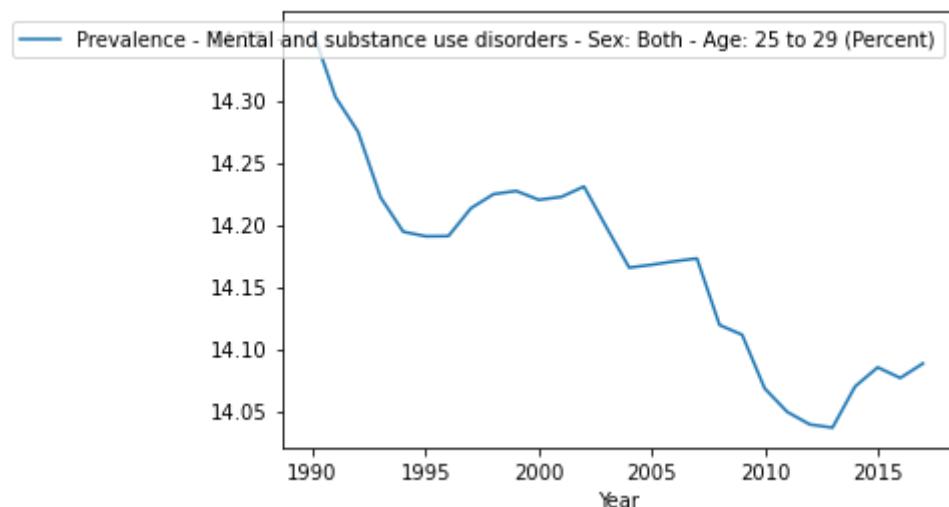
```
In [1885]: df84.groupby('Year')[v1].mean().plot(legend=True)
```

```
Out[1885]: <AxesSubplot:xlabel='Year'>
```



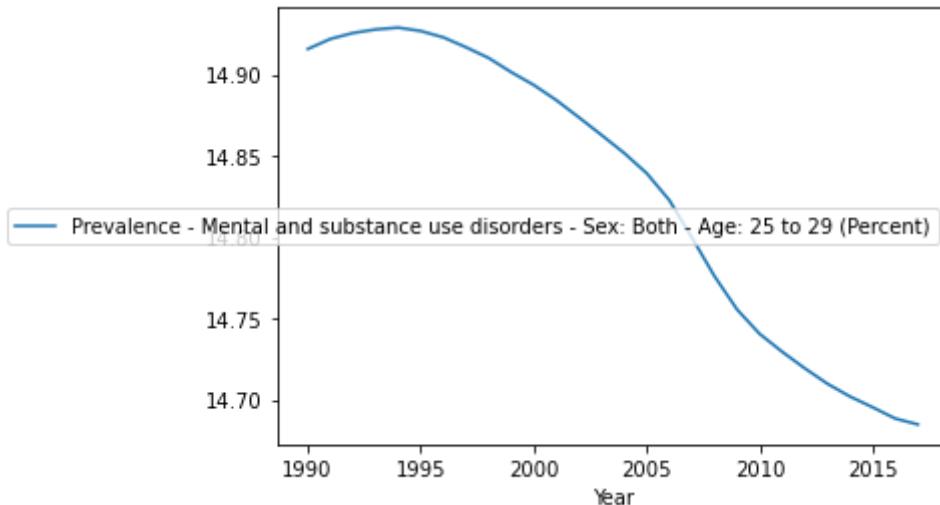
```
In [1886]: df84.groupby('Year')[v2].median().plot(legend=True)
```

```
Out[1886]: <AxesSubplot:xlabel='Year'>
```



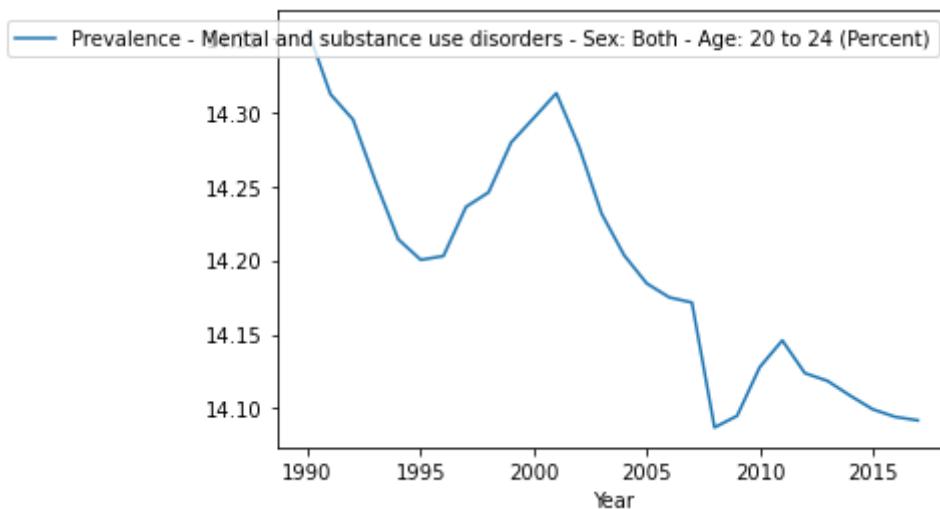
```
In [1887]: df84.groupby('Year')[v2].mean().plot(legend=True)
```

```
Out[1887]: <AxesSubplot:xlabel='Year'>
```



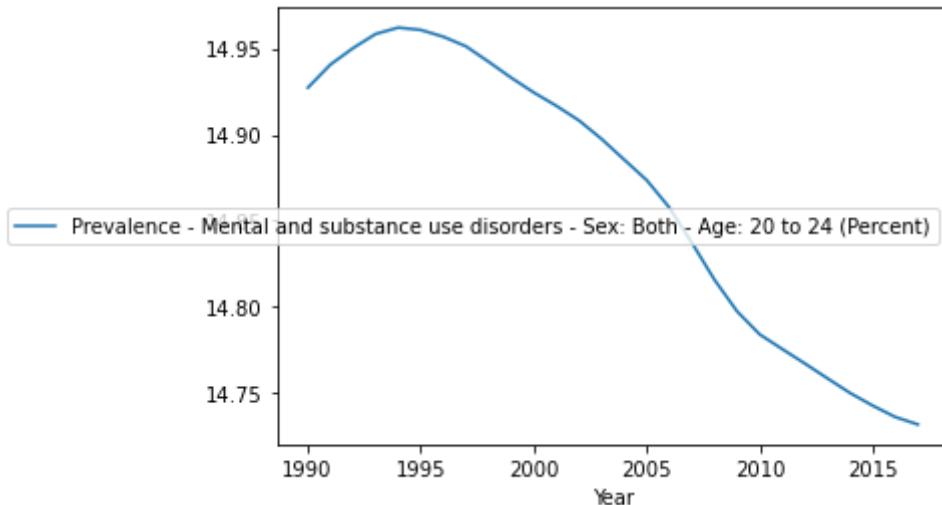
```
In [1888]: df84.groupby('Year')[v3].median().plot(legend=True)
```

```
Out[1888]: <AxesSubplot:xlabel='Year'>
```



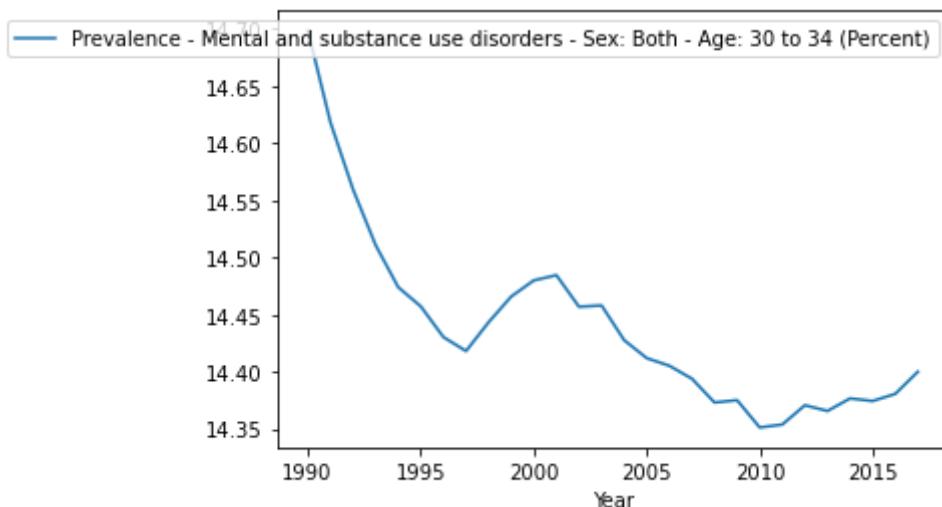
```
In [1889]: df84.groupby('Year')[v3].mean().plot(legend=True)
```

```
Out[1889]: <AxesSubplot:xlabel='Year'>
```



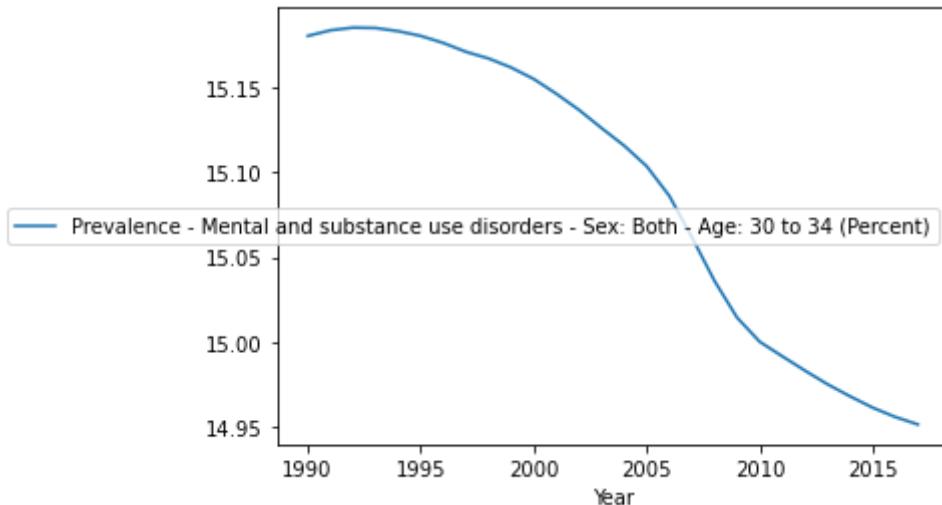
```
In [1890]: df84.groupby('Year')[v4].median().plot(legend=True)
```

```
Out[1890]: <AxesSubplot:xlabel='Year'>
```



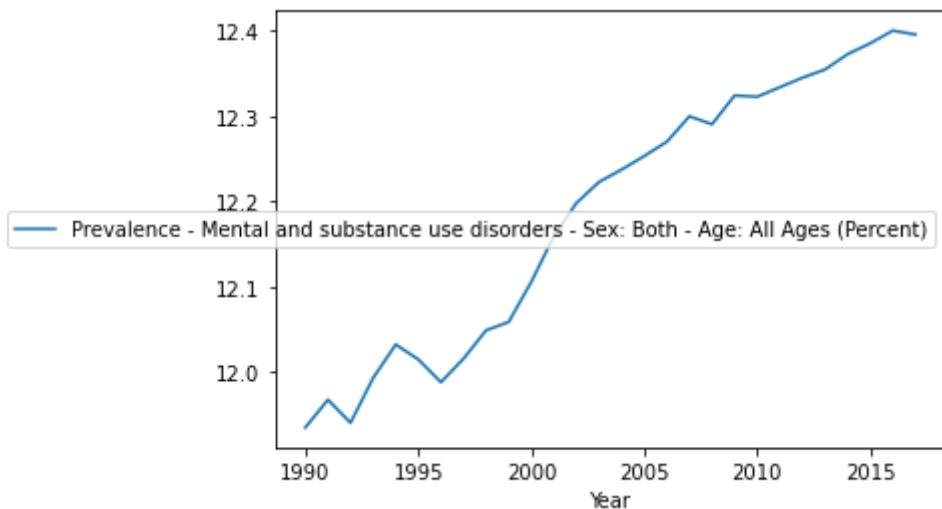
```
In [1891]: df84.groupby('Year')[v4].mean().plot(legend=True)
```

```
Out[1891]: <AxesSubplot:xlabel='Year'>
```



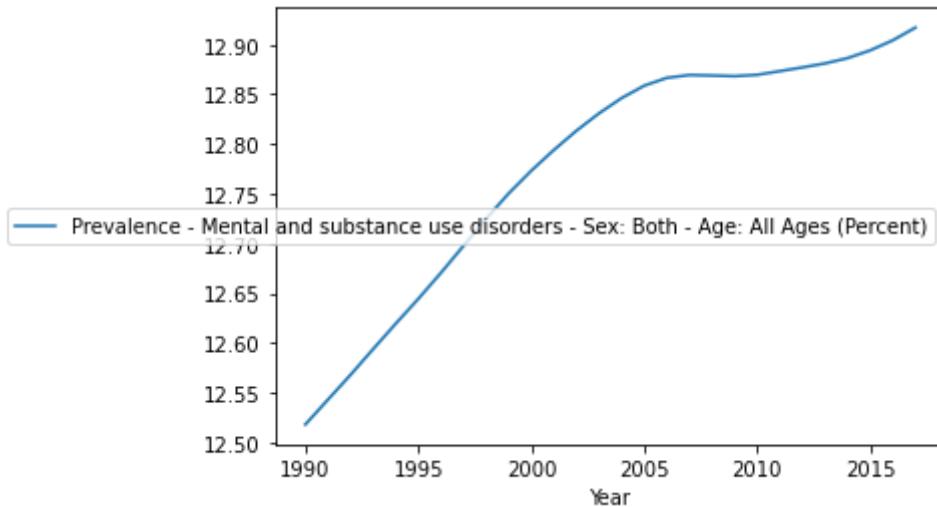
```
In [1892]: df84.groupby('Year')[v5].median().plot(legend=True)
```

```
Out[1892]: <AxesSubplot:xlabel='Year'>
```



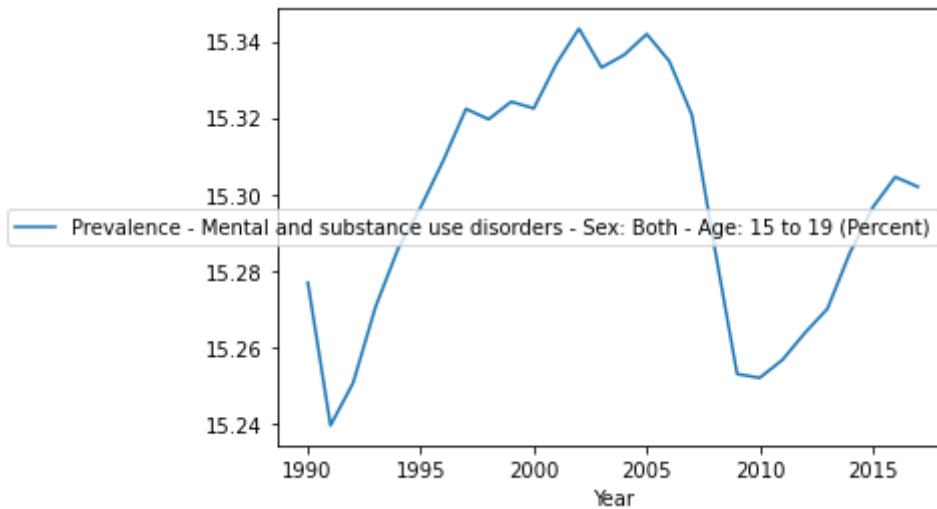
```
In [1893]: df84.groupby('Year')[v5].mean().plot(legend=True)
```

```
Out[1893]: <AxesSubplot:xlabel='Year'>
```



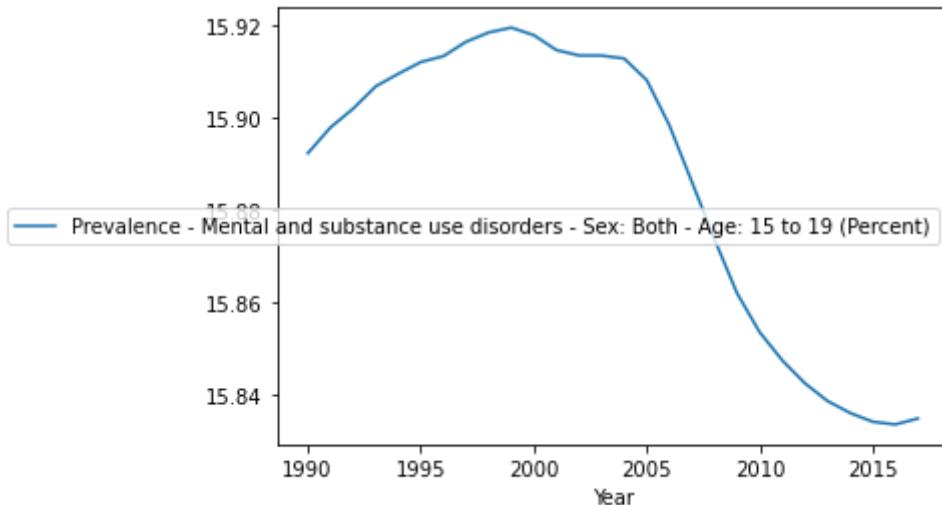
```
In [1894]: df84.groupby('Year')[v6].median().plot(legend=True)
```

```
Out[1894]: <AxesSubplot:xlabel='Year'>
```



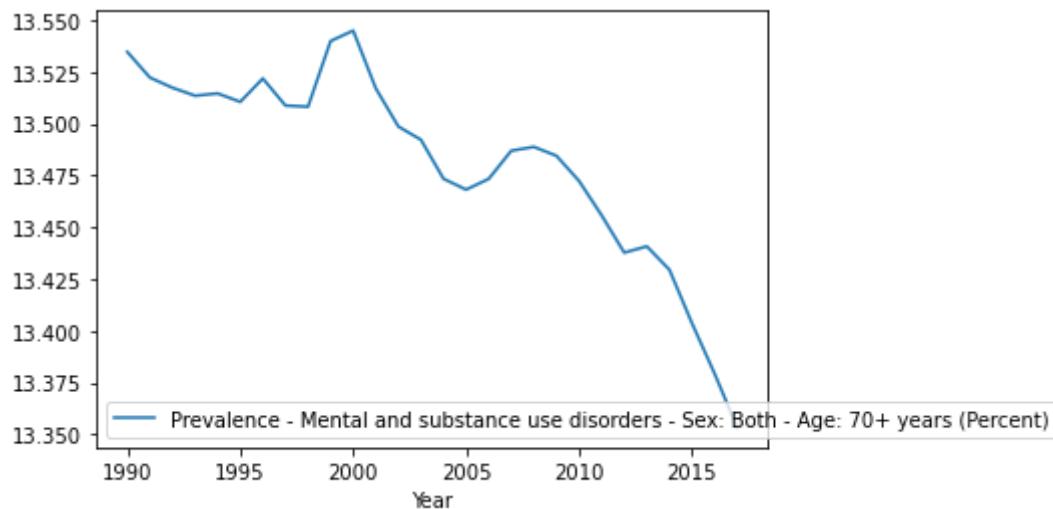
```
In [1895]: df84.groupby('Year')[v6].mean().plot(legend=True)
```

```
Out[1895]: <AxesSubplot:xlabel='Year'>
```



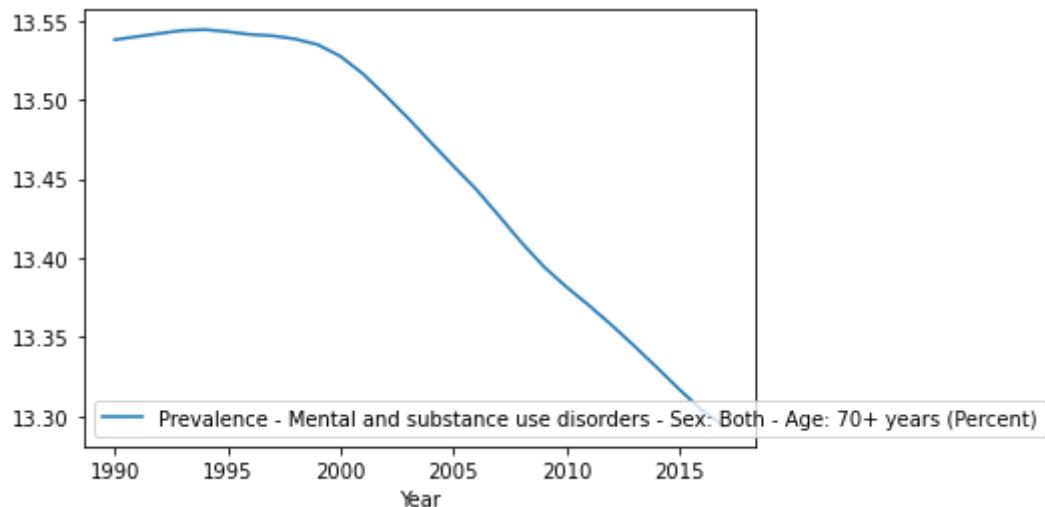
```
In [1896]: df84.groupby('Year')[v7].median().plot(legend=True)
```

```
Out[1896]: <AxesSubplot:xlabel='Year'>
```



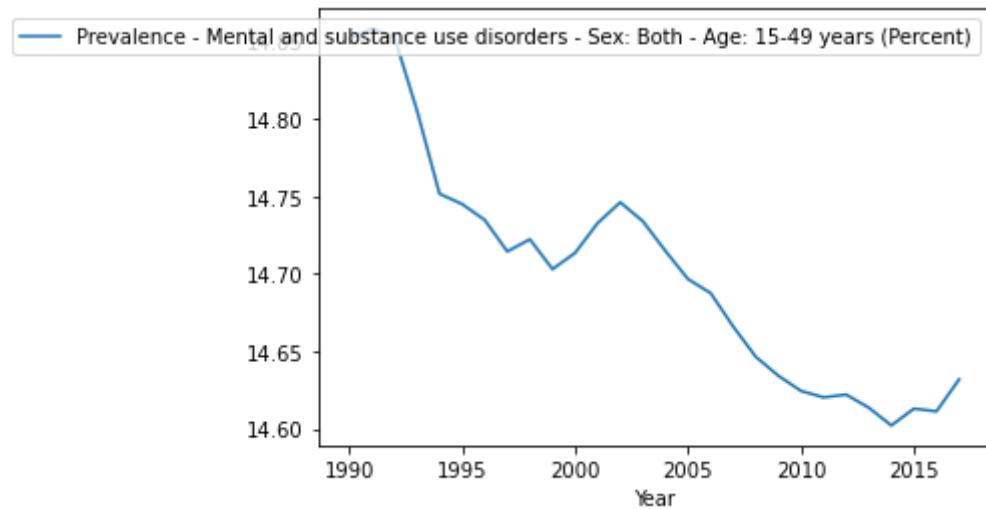
```
In [1897]: df84.groupby('Year')[v7].mean().plot(legend=True)
```

```
Out[1897]: <AxesSubplot:xlabel='Year'>
```



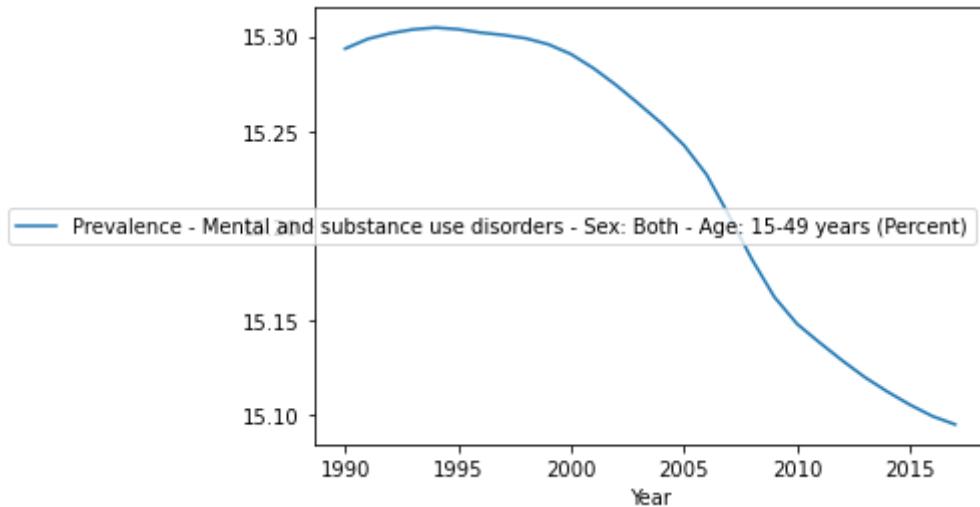
```
In [1898]: df84.groupby('Year')[v8].median().plot(legend=True)
```

```
Out[1898]: <AxesSubplot:xlabel='Year'>
```



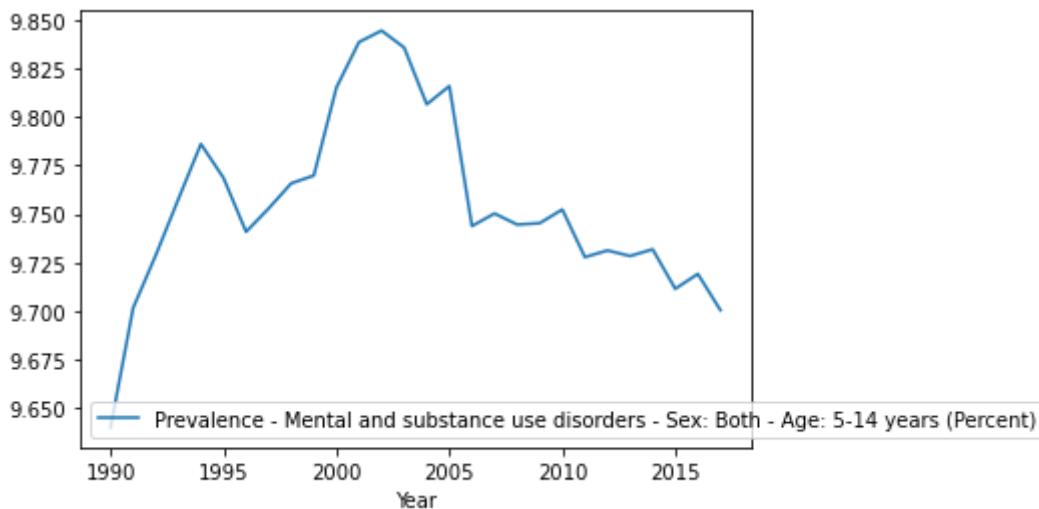
```
In [1899]: df84.groupby('Year')[v8].mean().plot(legend=True)
```

```
Out[1899]: <AxesSubplot:xlabel='Year'>
```



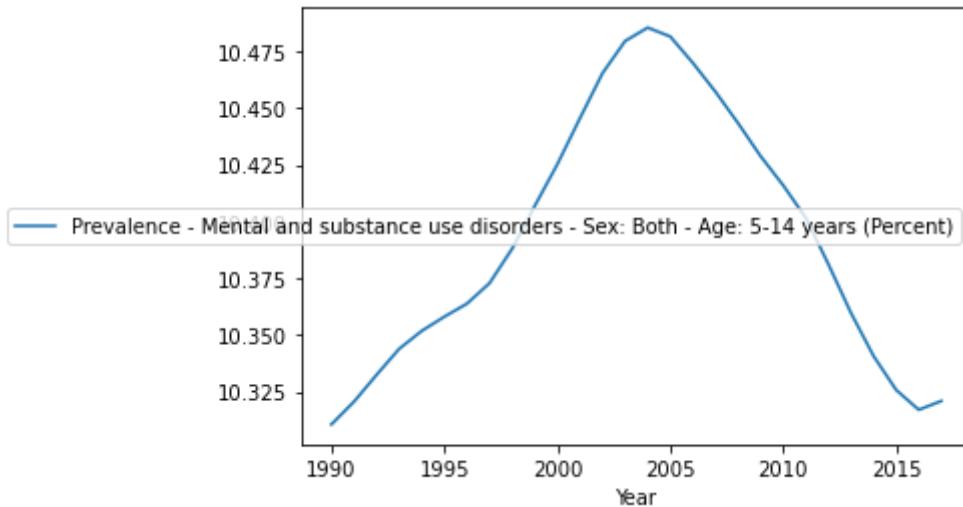
```
In [1900]: df84.groupby('Year')[v9].median().plot(legend=True)
```

```
Out[1900]: <AxesSubplot:xlabel='Year'>
```



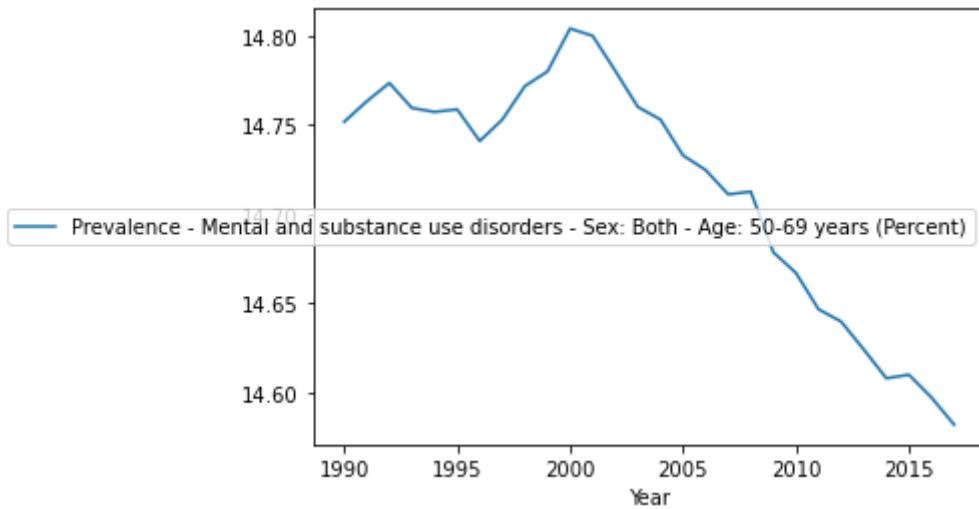
```
In [1901]: df84.groupby('Year')[v9].mean().plot(legend=True)
```

```
Out[1901]: <AxesSubplot:xlabel='Year'>
```



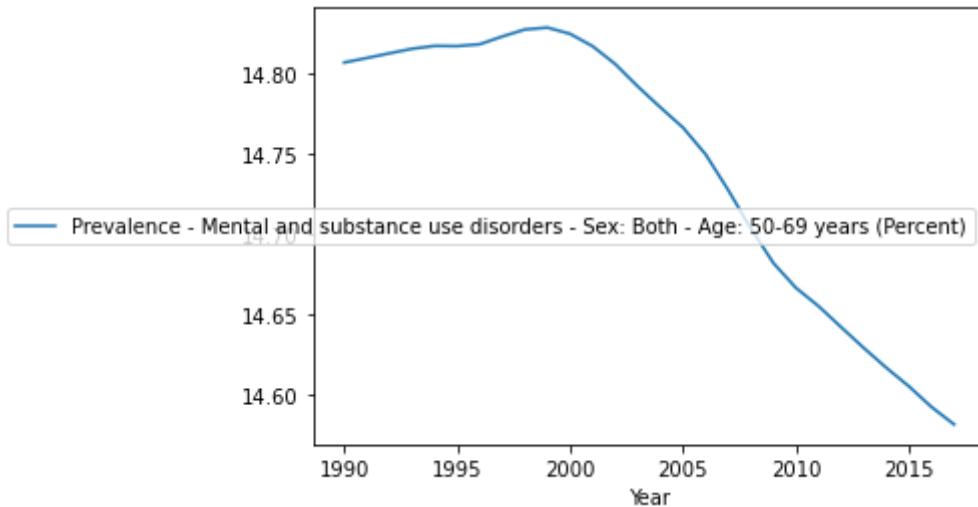
```
In [1902]: df84.groupby('Year')[v10].median().plot(legend=True)
```

```
Out[1902]: <AxesSubplot:xlabel='Year'>
```



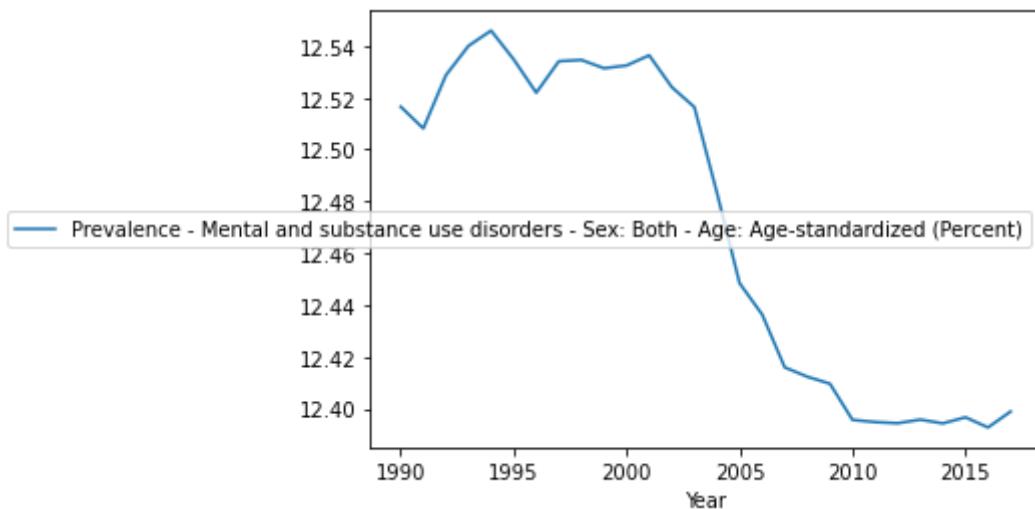
```
In [1903]: df84.groupby('Year')[v10].mean().plot(legend=True)
```

```
Out[1903]: <AxesSubplot:xlabel='Year'>
```



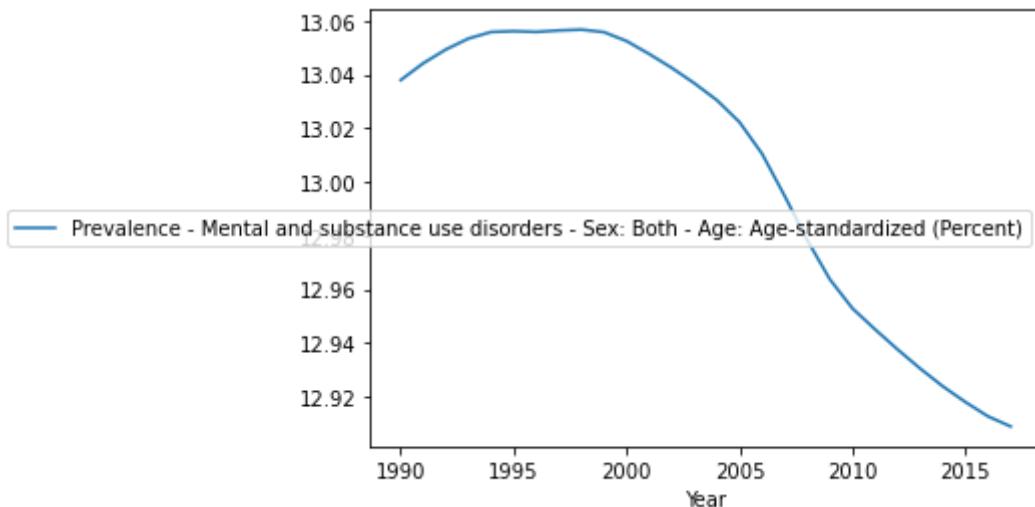
```
In [1904]: df84.groupby('Year')[v11].median().plot(legend=True)
```

```
Out[1904]: <AxesSubplot:xlabel='Year'>
```



```
In [1905]: df84.groupby('Year')[v11].mean().plot(legend=True)
```

```
Out[1905]: <AxesSubplot:xlabel='Year'>
```



```
In [1906]: df107.info()
df107.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6156 entries, 0 to 6155
Data columns (total 4 columns):
 #   Column          Dtype  
 --- 
 0   Entity          object  
 1   Code            object  
 2   Year            int64  
 3   Prevalence - Mental health disorders: Both (age-standardized percent)  float64 
dtypes: float64(1), int64(1), object(2)
memory usage: 192.5+ KB
```

```
Out[1906]:
```

	Entity	Code	Year	Prevalence - Mental health disorders: Both (age-standardized percent)
0	Afghanistan	AFG	1990	16.046473
1	Afghanistan	AFG	1991	16.095666
2	Afghanistan	AFG	1992	16.165110
3	Afghanistan	AFG	1993	16.215425
4	Afghanistan	AFG	1994	16.271298

Checking for missing values:

```
In [1907]: missing = pd.concat([df107.isnull().sum(), 100 * df107.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1907]:

	count	%
Entity	0	0.000000
Year	0	0.000000
Prevalence - Mental health disorders: Both (age-standardized percent)	0	0.000000
Code	864	14.035088

```
In [1908]: v1='Prevalence - Mental health disorders: Both (age-standardized percent)'
```

```
In [1909]: df107.describe()
```

Out[1909]:

	Year	Prevalence - Mental health disorders: Both (age-standardized percent)
count	6156.000000	6156.000000
mean	2003.000000	13.504531
std	7.789514	1.568068
min	1990.000000	10.195587
25%	1996.000000	12.089270
50%	2003.000000	13.506910
75%	2010.000000	14.600790
max	2016.000000	18.196191

```
In [1910]: df107.sort_values(by=v1, ascending=False)[['Entity', 'Year', v1]].head()
```

Out[1910]:

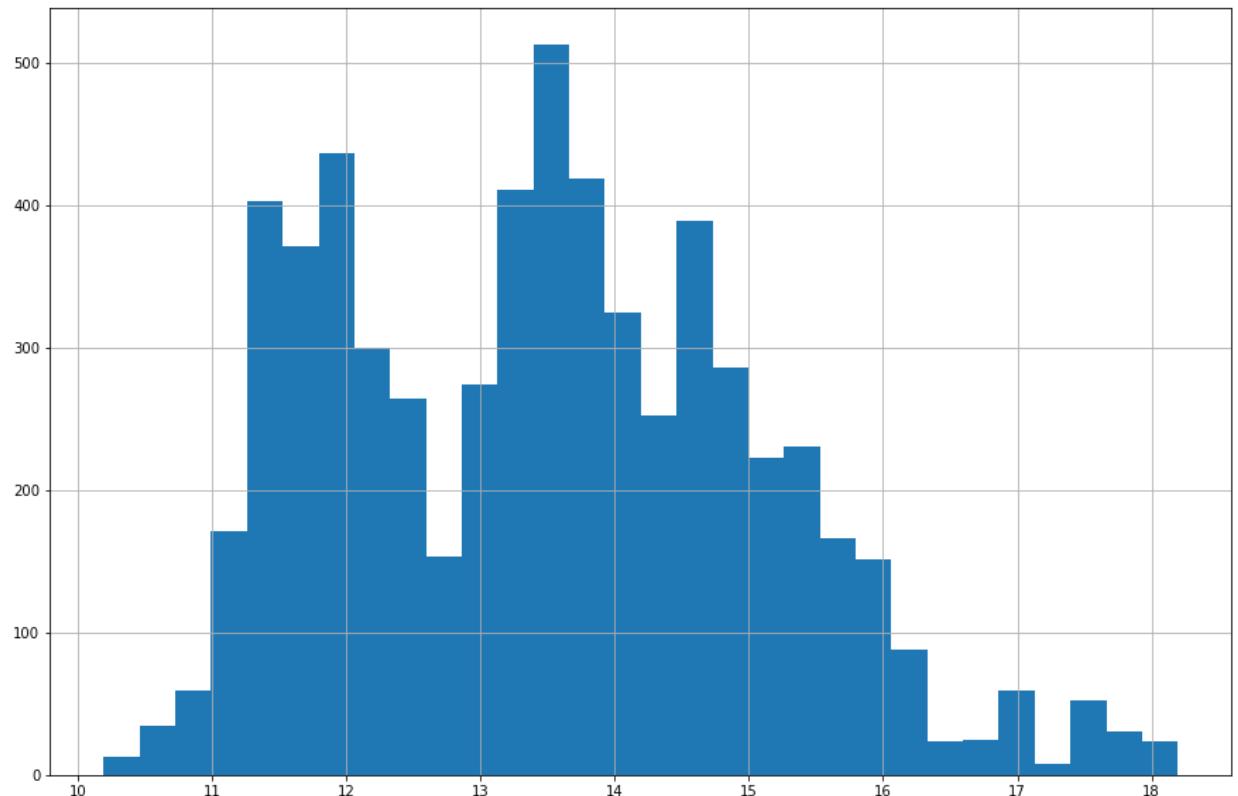
	Entity	Year	Prevalence - Mental health disorders: Both (age-standardized percent)
312	Australia	2005	18.196191
313	Australia	2006	18.171638
285	Australasia	2005	18.150350
2618	Iran	2016	18.143560
314	Australia	2007	18.140135

```
In [1911]: df107.sort_values(by=v1, ascending=True)[['Entity', 'Year', v1]].head()
```

Out[1911]:

	Entity	Year	Prevalence - Mental health disorders: Both (age-standardized percent)
1241	Colombia	2016	10.195587
1240	Colombia	2015	10.209335
1239	Colombia	2014	10.214928
1238	Colombia	2013	10.220622
1237	Colombia	2012	10.230566

```
In [1912]: df107[v1].hist(bins=30, figsize=(15,10))
plt.subplots_adjust(hspace=0.5);
```



```
In [1913]: df107.groupby('Entity')[v1].mean().sort_values().head()
```

Out[1913]: Entity

Colombia	10.540669
Vietnam	10.545437
Azerbaijan	10.857853
Mexico	10.993439
Central Latin America	11.005715
Name: Prevalence - Mental health disorders: Both (age-standardized percent), dtype: float64	

```
In [1914]: df107.groupby('Entity')[v1].mean().sort_values().tail()
```

```
Out[1914]: Entity
Greenland      16.956065
Iran           17.610517
New Zealand    17.652422
Australasia    17.819939
Australia      17.854488
Name: Prevalence - Mental health disorders: Both (age-standardized percent), dtype: float64
```

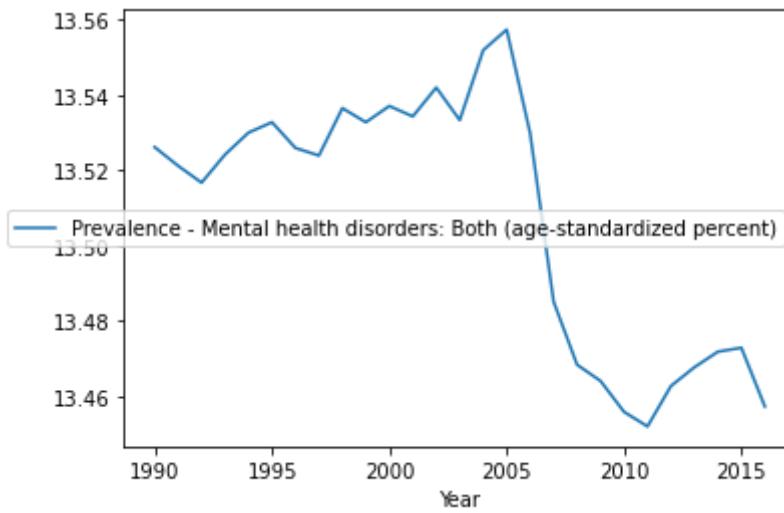
```
In [1915]: df107_mean = df107.groupby('Year').mean()
df107_mean.head()
```

```
Out[1915]:
Prevalence - Mental health disorders: Both (age-standardized percent)
```

Year	
1990	13.496756
1991	13.502705
1992	13.510082
1993	13.516313
1994	13.522379

```
In [1916]: df107.groupby('Year')[v1].median().plot(legend=True)
```

```
Out[1916]: <AxesSubplot:xlabel='Year'>
```



DALY by Cause

In [1917]:

df92.info()

df92.head()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 26 columns):
 #   Column          Non-Null Count  Dtype  
---  --  
 0   Entity          6468 non-null    object 
 1   Code             5488 non-null    object 
 2   Year             6468 non-null    int64  
 3   DALYs (Disability-Adjusted Life Years) - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)
 4   DALYs (Disability-Adjusted Life Years) - Neoplasms - Sex: Both - Age: All Ages (Percent)
 5   DALYs (Disability-Adjusted Life Years) - Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)
 6   DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)
 7   DALYs (Disability-Adjusted Life Years) - Chronic respiratory diseases - Sex: Both - Age: All Ages (Percent)
 8   DALYs (Disability-Adjusted Life Years) - Other non-communicable diseases - Sex: Both - Age: All Ages (Percent)
 9   DALYs (Disability-Adjusted Life Years) - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Percent)
 10  DALYs (Disability-Adjusted Life Years) - Diarrhea, lower respiratory, and other common infectious diseases - Sex: Both - Age: All Ages (Percent)
 11  DALYs (Disability-Adjusted Life Years) - Neonatal disorders - Sex: Both - Age: All Ages (Percent)
 12  DALYs (Disability-Adjusted Life Years) - Maternal disorders - Sex: Both - Age: All Ages (Percent)
 13  DALYs (Disability-Adjusted Life Years) - Musculoskeletal disorders - Sex: Both - Age: All Ages (Percent)
 14  DALYs (Disability-Adjusted Life Years) - Unintentional injuries - Sex: Both - Age: All Ages (Percent)
 15  DALYs (Disability-Adjusted Life Years) - Neurological disorders - Sex: Both - Age: All Ages (Percent)
 16  DALYs (Disability-Adjusted Life Years) - HIV/AIDS and tuberculosis - Sex: Both - Age: All Ages (Percent)
```

```

6156 non-null    float64
17 DALYs (Disability-Adjusted Life Years) - Transport injuries - Sex: B
oth - Age: All Ages (Percent)
6468 non-null    float64
18 DALYs (Disability-Adjusted Life Years) - Self-harm - Sex: Both - Ag
e: All Ages (Percent)
6468 non-null    float64
19 DALYs (Disability-Adjusted Life Years) - Interpersonal violence - Se
x: Both - Age: All Ages (Percent)
6468 non-null    float64
20 DALYs (Disability-Adjusted Life Years) - Conflict and terrorism - Se
x: Both - Age: All Ages (Percent)
6468 non-null    float64
21 DALYs (Disability-Adjusted Life Years) - Exposure to forces of natur
e - Sex: Both - Age: All Ages (Percent)
6468 non-null    float64
22 DALYs (Disability-Adjusted Life Years) - Neglected tropical diseases
and malaria - Sex: Both - Age: All Ages (Percent)
6468 non-null    float64
23 DALYs (Disability-Adjusted Life Years) - Nutritional deficiencies - 
Sex: Both - Age: All Ages (Percent)
6468 non-null    float64
24 DALYs (Disability-Adjusted Life Years) - Digestive diseases - Sex: B
oth - Age: All Ages (Percent)
6468 non-null    float64
25 DALYs (Disability-Adjusted Life Years) - Other communicable, materna
l, neonatal, and nutritional diseases - Sex: Both - Age: All Ages (Percen
t) 6156 non-null    float64
dtypes: float64(23), int64(1), object(2)
memory usage: 1.3+ MB

```

Out[1917]:

Entity	Code	Year	DALYs	(Disability-Adjusted Life Years)	DALYs	(Disability-Adjusted Life Years)	DALYs	(Disability-Adjusted Life Years)	DALYs	(Disability-Adjusted Life Years)	(Dis
			(Disability-Adjusted Life Years) - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)	- Sex: Both - Age: All Ages (Percent)	Neoplasms - Sex: Both - Age: All Ages (Percent)	Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)	- Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)	respiratory diseases - Sex: Both - Age: All Ages (Percent)	commu dis Sex Age: A (P		
0	Afghanistan	AFG	1990	9.392936	2.472793	2.897539	1.320975	1.682207	6.		
1	Afghanistan	AFG	1991	9.301676	2.448247	2.892036	1.354445	1.668920	6.		
2	Afghanistan	AFG	1992	9.084693	2.429226	2.953674	1.506504	1.665093	6.		
3	Afghanistan	AFG	1993	8.080855	2.213392	3.001952	1.487670	1.514913	6.		
4	Afghanistan	AFG	1994	7.464999	2.069922	2.995463	1.389687	1.411217	6.		

5 rows × 26 columns

Checking for missing values:

```
In [1918]: missing = pd.concat([df92.isnull().sum(), 100 * df92.isnull().mean()], axis=1)
missing.columns=['count', '%']
missing.sort_values(by='count', ascending=True)
```

Out[1918]:

		count	%
	Entity	0	0.000000
DALYs (Disability-Adjusted Life Years) - Nutritional deficiencies - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Neglected tropical diseases and malaria - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Exposure to forces of nature - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Conflict and terrorism - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Interpersonal violence - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Self-harm - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Transport injuries - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Neurological disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Unintentional injuries - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Musculoskeletal disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Digestive diseases - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Maternal disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Other non-communicable diseases - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Chronic respiratory diseases - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Neoplasms - Sex: Both - Age: All Ages (Percent)		0	0.000000
DALYs (Disability-Adjusted Life Years) - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)		0	0.000000
	Year	0	0.000000
DALYs (Disability-Adjusted Life Years) - Neonatal disorders - Sex: Both - Age: All Ages (Percent)		0	0.000000

		count	%
DALYs (Disability-Adjusted Life Years) - Diarrhea, lower respiratory, and other common infectious diseases - Sex: Both - Age: All Ages (Percent)	312	4.823748	
DALYs (Disability-Adjusted Life Years) - HIV/AIDS and tuberculosis - Sex: Both - Age: All Ages (Percent)	312	4.823748	
DALYs (Disability-Adjusted Life Years) - Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)	312	4.823748	
DALYs (Disability-Adjusted Life Years) - Other communicable, maternal, neonatal, and nutritional diseases - Sex: Both - Age: All Ages (Percent)	312	4.823748	
Code	980	15.151515	

Compiling Available Health Data

There are three main types of information available: prevalence, disability adjusted life years, and deaths. Some of this is available by sex and by age for each condition.

Prevalence

```
In [1919]: prev_frames=[df71, df80, df64, df84, df65, df107, df69, df58, df73, df50, d
on=['Entity', 'Year']
prevalence=pd.merge(df4, df94, how='outer', on=on)
for frame in prev_frames:
    prevalence=pd.merge(prevalence, frame, how='outer', on=on)
```

```
In [1920]: prevalence.head()
```

```
Out[1920]:
```

	Entity	Code_x	Year	Prevalence -	Prevalence -	Continent	Code_y	Prevalence -
				Alcohol and substance use disorders: Male (age-standardized percent)	Alcohol and substance use disorders: Female (age-standardized percent)			Alcohol and substance use disorders: Both (age-standardized percent)
0	Abkhazia	OWID_ABK	2015	NaN	NaN	Asia	NaN	NaN
1	Afghanistan	AFG	1990	1.740499	1.314530	NaN	AFG	1.530549
2	Afghanistan	AFG	1991	1.742372	1.310338	NaN	AFG	1.530105
3	Afghanistan	AFG	1992	1.741842	1.307518	NaN	AFG	1.529283
4	Afghanistan	AFG	1993	1.745838	1.305118	NaN	AFG	1.530944

5 rows × 138 columns

```
In [1921]: prevalence.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6558 entries, 0 to 6557
Columns: 138 entries, Entity to Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
dtypes: float64(114), int64(1), object(23)
memory usage: 7.0+ MB
```

```
In [1922]: prevalence.shape
```

```
Out[1922]: (6558, 138)
```

In [1923]: prevalence.describe().T

Out[1923]:

	count	mean	std	min	25%	50%
Year	6558.0	2003.657823	8.133552e+00	1990.000000	1997.000000	2004.000000
Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)	6156.0	3.121449	1.260856e+00	1.480319	2.102214	2.843404
Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)	6156.0	1.532282	7.093454e-01	0.806518	1.144971	1.375345
Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)	6156.0	2.311523	9.106291e-01	1.196984	1.651095	2.054947
Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)	6468.0	0.105648	4.076594e-02	0.043191	0.070354	0.104129
...
Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)_y	6468.0	0.239998	1.581412e-01	0.073908	0.122387	0.182525
Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)	6468.0	0.000000	0.000000e+00	0.000000	0.000000	0.000000
Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)	6468.0	0.421181	2.784197e-01	0.140185	0.218446	0.315261
Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)	6468.0	401441.299273	1.564105e+06	106.094556	7740.857540	30051.302500
Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)	6468.0	972830.379091	3.818551e+06	258.848566	19718.996389	67314.612554

115 rows × 8 columns

```
In [1924]: print(list(prevalence.columns))
```

```
['Entity', 'Code_x', 'Year', 'Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)', 'Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)', 'Continent', 'Code_y', 'Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)', 'Code_x', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)', 'Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)_x', 'Code_y', 'Prevalence - Drug use disorders - Sex: Both - Age: 30 to 34 (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: 10 to 14 (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: 15 to 19 (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: 25 to 29 (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: All Ages (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: 20 to 24 (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: 70+ years (Percent)', 'Prevalence - Drug use disorders - Sex: Male - Age: 5-14 years (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: 15-49 years (Percent)', 'Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)_x', 'Code_x', 'Prevalence - Mental and substance use disorders - Sex: Male - Age: All Ages (Number)', 'Prevalence - Mental and substance use disorders - Sex: Female - Age: All Ages (Number)', 'Code_y', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 10 to 14 (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 25 to 29 (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 20 to 24 (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 30 to 34 (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 15 to 19 (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 70+ years (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Percent)', 'Prevalence - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Percent)', 'Code_x', 'Prevalence - Mental health disorders: Male (Number)', 'Prevalence - Mental health disorders: Female (Number)', 'Code_y', 'Prevalence - Mental health disorders: Both (age-standardized percent)', 'Code_x', 'Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)_x', 'Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)_x', 'Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)_x', 'Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)_x', 'Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)_y', 'Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)_x', 'Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)_y', 'Code_y', 'Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)', 'Prevalence - Anxiety disorders - Sex: Female - Age: All Ages (Number)']
```

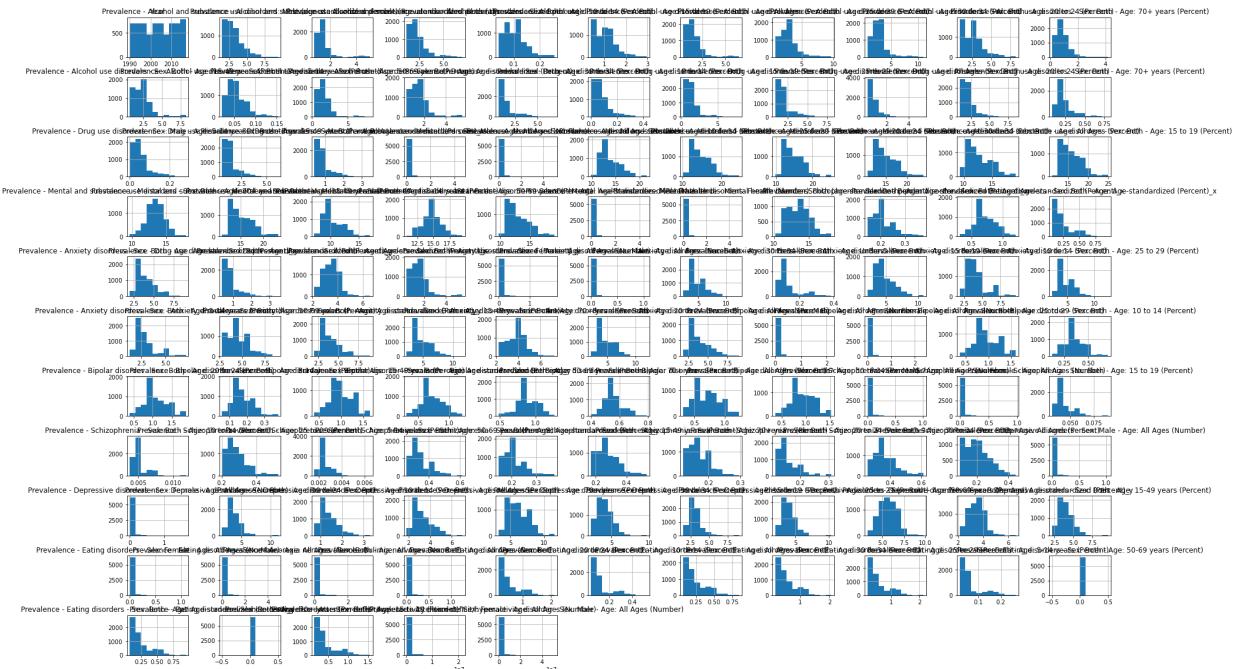

nce - Eating disorders - Sex: Both - Age: All Ages (Percent)', 'Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)', 'Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)', 'Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)', 'Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)', 'Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)_y', 'Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)', 'Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)', 'Code', 'Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)', 'Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)']

```
In [1925]: labels=['Code', 'Code_x', 'Code_y']
prevalence.drop(labels=labels, axis=1, inplace=True)
```

```
In [1926]: prevalence.shape
```

```
Out[1926]: (6558, 117)
```

```
In [1927]: prevalence.hist(figsize=(30,20))
plt.subplots_adjust(hspace=0.5, wspace=0.5)
```



In [1928]: *#checking out the fourth and sixth from last figures*

```
prevalence.iloc[:, -6: ].describe()
```

Out[1928]:

	Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)_y	Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
count	6468.0	6468.000000	6468.0	6468.000000	6.468000e+03	6.468000e+03
mean	0.0	0.239998	0.0	0.421181	4.014413e+05	9.728304e+05
std	0.0	0.158141	0.0	0.278420	1.564105e+06	3.818551e+06
min	0.0	0.073908	0.0	0.140185	1.060946e+02	2.588486e+02
25%	0.0	0.122387	0.0	0.218446	7.740858e+03	1.971900e+04
50%	0.0	0.182525	0.0	0.315261	3.005130e+04	6.731461e+04
75%	0.0	0.292666	0.0	0.501833	1.308604e+05	3.181667e+05
max	0.0	0.943991	0.0	1.558513	2.165715e+07	5.166036e+07

In [1929]: prevalence.iloc[:, -6:].head(15)

Out[1929]:

	Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)_y	Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)	Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - deficit/hyperactivity disorder - Sex: Female - Age: All Ages (Number)	Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: All Ages (Number)
0	NaN	NaN	NaN	NaN	NaN	NaN
1	0.0	0.101855	0.0	0.211466	36372.093241	84023.829459
2	0.0	0.099313	0.0	0.206605	37618.541437	87172.725113
3	0.0	0.096692	0.0	0.199407	44314.463114	103926.014285
4	0.0	0.094336	0.0	0.193288	51144.300550	121174.627507
5	0.0	0.092439	0.0	0.189605	52949.136913	125597.544292
6	0.0	0.090980	0.0	0.188190	54719.437913	129877.650634
7	0.0	0.089709	0.0	0.186838	56790.184986	134959.835838
8	0.0	0.088372	0.0	0.184621	58875.967844	140055.344472
9	0.0	0.087330	0.0	0.183258	60936.767862	144936.661591
10	0.0	0.086267	0.0	0.182827	61302.045036	145277.142088
11	0.0	0.086021	0.0	0.182202	62296.725524	147301.164376
12	0.0	0.086517	0.0	0.183154	64209.552220	151761.412990
13	0.0	0.087023	0.0	0.184214	69209.093404	164349.476236
14	0.0	0.087189	0.0	0.183249	74710.940259	178223.497623

In [1930]: #dropping fourth and sixth from last, lack of data
prevalence.drop(labels=['Prevalence - Eating disorders - Sex: Both - Age: 5
#dropping data for children <5
prevalence.drop(labels=['Prevalence - Anxiety disorders - Sex: Both - Age:

In [1931]: prevalence.shape

Out[1931]: (6558, 114)

```
In [1932]: prevalence_mean = prevalence.groupby('Year').mean()
prevalence_mean.head()
```

Out[1932]:

Prevalence - Alcohol and substance use disorders: Male (age-standardized percent)	Prevalence - Alcohol and substance use disorders: Female (age-standardized percent)	Prevalence - Alcohol and substance use disorders: Both (age-standardized percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)	Prevalence - Alcohol use disorders - Sex: All Ages (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)	Pr
Year							
1990	3.072778	1.526307	2.285541	0.103682	1.046755	1.456609	2.955716
1991	3.081035	1.525942	2.289356	0.103985	1.049101	1.467102	2.959101
1992	3.089954	1.526076	2.293757	0.104270	1.051481	1.476878	2.962774
1993	3.099274	1.526165	2.298235	0.104516	1.053740	1.486171	2.967044
1994	3.108663	1.526574	2.302819	0.104709	1.055523	1.494500	2.971078

5 rows × 111 columns

```
In [1933]: prevalence.Entity.value_counts()
```

```
Out[1933]: Algeria      28
Papua New Guinea    28
East Asia          28
Taiwan              28
Cambodia           28
..
Western Sahara     1
Isle of Man         1
Bavaria             1
West Germany        1
South Ossetia       1
Name: Entity, Length: 321, dtype: int64
```

```
In [1934]: #prevalence.to_csv(r'Data/prevalence.csv')
```

DALY

```
In [1935]: daly_frames=[df43, df17, df18, df19, df48, df12, df13, df14, df39, df16]
on=['Entity', 'Year']
daly=pd.merge(df92, df3, how='outer', on=on)
for frame in daly_frames:
    daly=pd.merge(daly, frame, how='outer', on=on)
```

```
In [1936]: daly.head()
```

Out[1936]:

Entity	Code_x	Year	DALYs (Disability- Adjusted Life Years)		DALYs (Disability- Adjusted Life Years)		DALYs (Disability- Adjusted Life Years)		DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders	DALYs (Disability- Adjusted Life Years) - Chronic respiratory diseases - Sex: Both - Age: All Ages (Percent)	comm c Age
			Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)	Neoplasms - Sex: Both - Age: All Ages (Percent)	Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)	Neoplasms - Sex: Both - Age: All Ages (Percent)	Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)	Neoplasms - Sex: Both - Age: All Ages (Percent)	Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)	(I Adj Year	
0	Afghanistan	AFG	1990	9.392936	2.472793	2.897539	1.320975	1.682207			
1	Afghanistan	AFG	1991	9.301676	2.448247	2.892036	1.354445	1.668920			
2	Afghanistan	AFG	1992	9.084693	2.429226	2.953674	1.506504	1.665093			
3	Afghanistan	AFG	1993	8.080855	2.213392	3.001952	1.487670	1.514913			
4	Afghanistan	AFG	1994	7.464999	2.069922	2.995463	1.389687	1.411217			

5 rows × 87 columns

```
In [1937]: daly.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6468 entries, 0 to 6467
Data columns (total 87 columns):
 #   Column
Non-Null Count Dtype  
---  --  
0   Entity      object 
6468 non-null   object 
1   Code_x      object 
5488 non-null   object 
2   Year        int64  
6468 non-null   float64
3   DALYs (Disability-Adjusted Life Years) - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
4   DALYs (Disability-Adjusted Life Years) - Neoplasms - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
5   DALYs (Disability-Adjusted Life Years) - Diabetes - Sex: Both - Age: All Ages (Percent)
6468 non-null   float64
```

```
In [1938]: daly.shape
```

Out[1938]: (6468, 87)

In [1939]: daly.describe().T

Out[1939]:

	count	mean	std	min	25%	50%	75%
Year	6468.0	2003.500000	8.078372	1990.000000	1996.750000	2003.500000	2010.250000
DALYs (Disability- Adjusted Life Years) - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)	6468.0	14.201126	7.900404	0.557580	7.788590	13.905230	18.716578
DALYs (Disability- Adjusted Life Years) - Neoplasms - Sex: Both - Age: All Ages (Percent)	6468.0	8.972946	5.675025	0.349207	3.968375	7.570153	13.962118
DALYs (Disability- Adjusted Life Years) - Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)	6156.0	5.467008	3.040279	0.269139	3.410626	4.777504	6.604851
DALYs (Disability- Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)	6468.0	4.642531	2.233705	0.174323	2.825027	4.512848	6.342170
...
DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)	6468.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	count	mean	std	min	25%	50%	75%
DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)	6468.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age- standardized (Rate)	6468.0	48.261493	30.295637	15.690928	25.538960	37.273714	59.739773
DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	6468.0	91.569072	55.255736	30.790943	49.343837	71.437059	116.516397
DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)	6468.0	13.040452	8.608827	4.648002	7.201464	9.905871	13.630086

74 rows × 8 columns

In [1940]: `print(list(daly.columns))`

```
['Entity', 'Code_x', 'Year', 'DALYs (Disability-Adjusted Life Years) - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Neoplasms - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Diabetes, urogenital, blood, and endocrine diseases - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Chronic respiratory diseases - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Other non-communicable diseases - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Diarrhea, lower respiratory, and other common infectious diseases - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Neonatal disorders - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Maternal disorders - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Musculoskeletal disorders - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Unintentional injuries - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Neurological disorders - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - HIV/AIDS and tuberculosis - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Transport injuries - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Self-harm - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Interpersonal violence - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Conflict and terrorism - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Exposure to forces of nature - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Neglected tropical diseases and malaria - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Nutritional deficiencies - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Digestive diseases - Sex: Both - Age: All Ages (Percent)', 'DALYs (Disability-Adjusted Life Years) - Other communicable, maternal, neonatal, and nutritional diseases - Sex: Both - Age: All Ages (Percent)', 'Code_y', 'DALYs (Disability-Adjusted Life Years) - Alcohol and substance use disorders (Age-standardized percent)', 'Code_x', 'DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)', 'Code_y', 'DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)', 'DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)', 'DALYs (Disability-Adjusted Life Years) - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)', 'Code_x', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Number)', 'DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: All Ages (Number)', 'DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Number)', 'DALYs (Disability-Adjusted Life Years) - Alcohol
```

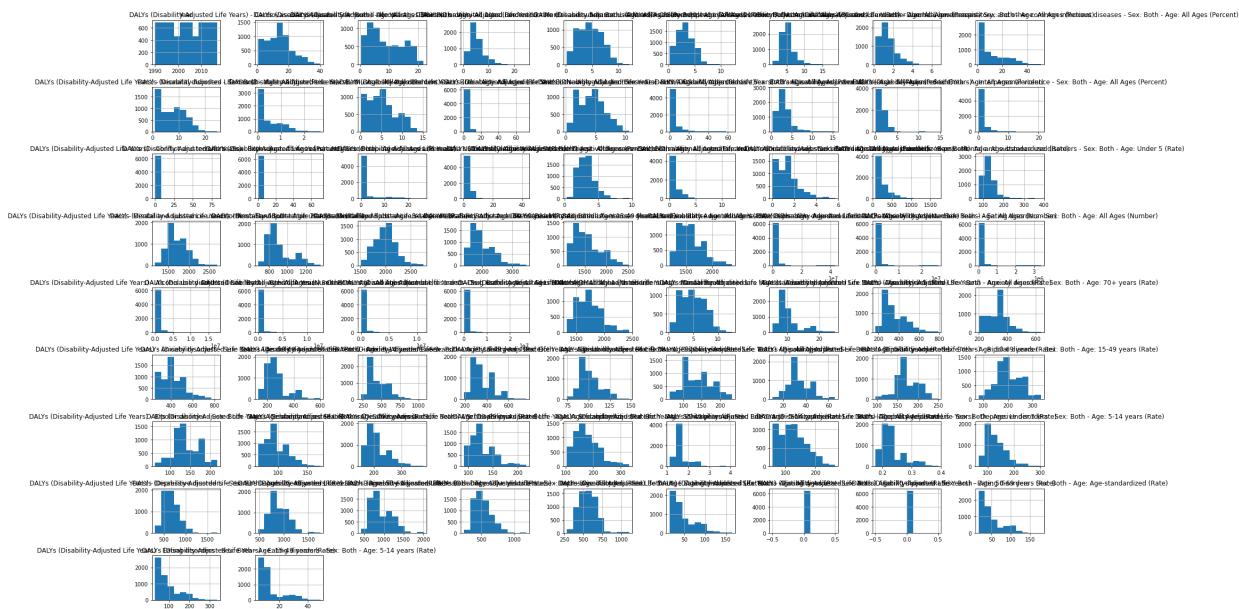
use disorders - Sex: Both - Age: All Ages (Number)', 'DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Number)', 'DALYs (Disability-Adjusted Life Years) - Other mental and substance use disorders - Sex: Both - Age: All Ages (Number)', 'DALYs (Disability-Adjusted Life Years) - Drug use disorders - Sex: Both - Age: All Ages (Number)', 'Code_y', 'DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized rate)', 'Code_x', 'DALYs (Disability-Adjusted Life Years) - Mental health disorders (Age-standardized percent)', 'Code_y', 'DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: Under 5 (Rate)', 'DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 5-14 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 15-49 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 50-69 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Anxiety disorders - Sex: Both - Age: 70+ years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 70+ years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: All Ages (Rate)', 'DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 5-14 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 50-69 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: 15-49 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Bipolar disorder - Sex: Both - Age: Age-standardized (Rate)', 'Code_y', 'DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 70+ years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 50-69 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: Age-standardized (Rate)', 'DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 15-49 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: 5-14 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Schizophrenia - Sex: Both - Age: All Ages (Rate)', 'Code_x', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Under 5 (Rate)', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 5-14 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 15-49 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 50-69 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: 70+ years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: All Ages (Rate)', 'DALYs (Disability-Adjusted Life Years) - Depressive disorders - Sex: Both - Age: Age-standardized (Rate)', 'Code_y', 'DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: All Ages (Rate)', 'DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age-standardized (Rate)', 'DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)', 'DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)']

```
In [1941]: labels=['Code_x', 'Code_y']
daly.drop(labels=labels, axis=1, inplace=True)
```

```
In [1942]: daly.shape
```

```
Out[1942]: (6468, 75)
```

```
In [1943]: daly.hist(figsize=(30,20))
plt.subplots_adjust(hspace=0.5, wspace=0.5)
```



```
In [1944]: #checking out the 19th, 20th, and 21st figures
daly.iloc[:, 19:21].describe()
```

```
Out[1944]:
```

DALYs (Disability-Adjusted Life Years) -
Conflict and terrorism - Sex: Both - Age: All
Ages (Percent)

DALYs (Disability-Adjusted Life Years) - Exposure
to forces of nature - Sex: Both - Age: All Ages
(Percent)

count	6468.000000	6468.000000
mean	0.509023	0.117686
std	2.723289	1.353555
min	0.000000	0.000000
25%	0.000940	0.000580
50%	0.013015	0.007133
75%	0.168618	0.031976
max	82.752299	70.419667

In [1945]: *#checking out the fourth and fifth from last figures*
 daly.iloc[:, -5:].describe()

Out[1945]:

	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age- standardized (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability- Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)
count	6468.0	6468.0	6468.000000	6468.000000	6468.000000
mean	0.0	0.0	48.261493	91.569072	13.040452
std	0.0	0.0	30.295637	55.255736	8.608827
min	0.0	0.0	15.690928	30.790943	4.648002
25%	0.0	0.0	25.538960	49.343837	7.201464
50%	0.0	0.0	37.273714	71.437059	9.905871
75%	0.0	0.0	59.739773	116.516397	13.630086
max	0.0	0.0	178.584448	337.597559	49.154747

```
In [1946]: daly.iloc[:, -5: ].head(15)
```

Out[1946]:

DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 70+ years (Rate)	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 50-69 years (Rate)	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: Age-standardized (Rate)	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 15-49 years (Rate)	DALYs (Disability-Adjusted Life Years) - Eating disorders - Sex: Both - Age: 5-14 years (Rate)
0	0	0	20.754767	38.803801
1	0	0	20.238001	38.161588
2	0	0	19.708028	37.933909
3	0	0	19.237173	37.602625
4	0	0	18.863183	37.177799
5	0	0	18.572830	36.978924
6	0	0	18.325176	36.874517
7	0	0	18.051890	36.594568
8	0	0	17.849554	36.528506
9	0	0	17.645544	36.321056
10	0	0	17.602620	36.318737
11	0	0	17.705911	36.613846
12	0	0	17.813673	36.997150
13	0	0	17.851164	36.993734
14	0	0	18.050296	37.276372

```
In [1947]: daly.drop(labels=['DALYs (Disability-Adjusted Life Years) - Eating disorder
```

```
In [1948]: #dropping data for children <5
daly.drop(labels=['DALYs (Disability-Adjusted Life Years) - Mental and subs
```

```
In [1949]: daly.shape
```

Out[1949]: (6468, 70)

```
In [1950]: daly_mean = daly.groupby('Year').mean()
daly_mean.head()
```

Out[1950]:

	DALYs (Disability- Adjusted Life Years) - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent)	DALYs (Disability- Adjusted Life Years) - Other non- communicable diseases - Sex: Both - Age: All Ages (Percent)	DALYs (Disability- Adjusted Life Years) - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Percent)	DALYs (Disability- Adjusted Life Years) - In di Se (
Year									
1990	13.906756	7.760885	4.399636	3.962344	3.128041	5.791973	1.286397	11	
1991	13.956143	7.857225	4.457569	4.022536	3.140292	5.733164	1.304674	16	

```
In [1951]: daly.Entity.value_counts()
```

```
Out[1951]: Algeria          28
Papua New Guinea      28
Lithuania            28
Andean Latin America  28
Afghanistan          28
..
Portugal              28
Bolivia               28
Brazil                28
Benin                 28
North Macedonia       28
Name: Entity, Length: 231, dtype: int64
```

```
In [1952]: #daly.to_csv(r'Data/daly.csv')
```

Deaths

```
In [1953]: death_frames=[df23, df20, df21, df22, df24, df31, df33, df35, df36, df37, df49]
on=['Entity', 'Year']
deaths=pd.merge(df6, df49, how='outer', on=on)
for frame in death_frames:
    deaths=pd.merge(deaths, frame, how='outer', on=on)
```

In [1954]: `deaths.head()`

Out[1954]:

	Entity	Code_x	Year	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)	Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)	Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)	Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)
0	Afghanistan	AFG	2007	9121.085992	29066.442137	461.195202	1846.996686	53532.680495
1	Afghanistan	AFG	2008	8387.057275	26623.480551	437.718960	1681.270324	53402.322328
2	Afghanistan	AFG	2009	7318.273004	24792.335792	415.776634	1568.095029	53024.450772
3	Afghanistan	AFG	2011	6919.757958	23115.144836	299.758258	1468.203744	52815.737495
4	Afghanistan	AFG	2012	6631.942601	22155.754481	302.255411	1406.214788	52961.704530

5 rows × 109 columns

In [1955]: `deaths.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6686 entries, 0 to 6685
Columns: 109 entries, Entity to Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)
dtypes: float64(92), int64(1), object(16)
memory usage: 5.6+ MB
```

In [1956]: `deaths.shape`

```
/opt/anaconda3/lib/python3.8/site-packages/IPython/core/displayhook.py:27
5: UserWarning: Output cache limit (currently 1000 entries) hit.
Flushing oldest 200 entries.
warn('Output cache limit (currently {sz} entries) hit.\n'
```

Out[1956]: (6686, 109)

In [1957]: `deaths.describe().T`

Out[1957]:

	count	mean	std	min	25%	50%	
Year	6686.0	2003.469488	8.076979	1990.000000	1996.000000	2003.000000	2010.000000
Deaths - Meningitis - Sex: Both - Age: All Ages (Number)							
Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)	6468.0	58527.671862	232696.669663	6.462032	610.654008	4313.073393	18629.87
Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)							
Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)	6468.0	7968.171555	35566.546169	0.035425	8.939738	139.132563	2337.48
...							
Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)							
Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)	6156.0	9.555160	11.518568	0.647502	3.314448	5.838692	10.69

	count	mean	std	min	25%	50%
Deaths - Mental and substance use						
disorders -	6468.0	5.014424	9.795962	0.034565	0.356321	1.254213
Sex: Both - Age: Age- standardized (Rate)						
Deaths - Mental and substance use						
disorders -	6468.0	5.106939	9.895266	0.028276	0.351129	1.278754
Sex: Both - Age: All Ages (Rate)						
Deaths - Eating disorders -						
Sex: Both -	6468.0	5.654176	26.023061	0.000106	0.016237	0.100280
Age: All Ages (Number)						

93 rows × 8 columns

```
In [1958]: print(list(deaths.columns))
```

```
['Entity', 'Code_x', 'Year', 'Deaths - Meningitis - Sex: Both - Age: All Ages (Number)', 'Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)', 'Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)', 'Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)', 'Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)', 'Deaths - Alzheimer disease and other dementia - Sex: Both - Age: All Ages (Number)', 'Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number)', 'Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages (Number)', 'Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Number)', 'Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number)', 'Deaths - Hepatitis - Sex: Both - Age: All Ages (Number)', 'Deaths - Neoplasms - Sex: Both - Age: All Ages (Number)', 'Deaths - Parkinson disease - Sex: Both - Age: All Ages (Number)', 'Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages (Number)', 'Deaths - Malaria - Sex: Both - Age: All Ages (Number)', 'Deaths - Drowning - Sex: Both - Age: All Ages (Number)', 'Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number)', 'Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Number)', 'Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)_x', 'Deaths - Tuberculosis - Sex: Both - Age: All Ages (Number)', 'Deaths - Road injuries - Sex: Both - Age: All Ages (Number)', 'Deaths - Maternal disorders - Sex: Both - Age: All Ages (Number)', 'Deaths - Neonatal disorders - Sex: Both - Age: All Ages (Number)', 'Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)_x', 'Deaths - Exposure to forces of nature - Sex: Both - Age: All Ages (Number)', 'Deaths - Diarrheal diseases - Sex: Both - Age: All Ages (Number)', 'Deaths - Environmental heat and cold exposure - Sex: Both - Age: All Ages (Number)', 'Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number)', 'Deaths - Self-harm - Sex: Both - Age: All Ages (Number)', 'Deaths - Conflict and terrorism - Sex: Both - Age: All Ages (Number)', 'Deaths - Diabetes mellitus - Sex: Both - Age: All Ages (Number)', 'Deaths - Poisonings - Sex: Both - Age: All Ages (Number)', 'Code_y', 'Deaths - Unsafe water source - Sex: Both - Age: All Ages (Number)', 'Deaths - Unsafe sanitation - Sex: Both - Age: All Ages (Number)', 'Deaths - No access to handwashing facility - Sex: Both - Age: All Ages (Number)', 'Deaths - Household air pollution from solid fuels - Sex: Both - Age: All Ages (Number)', 'Deaths - Secondhand smoke - Sex: Both - Age: All Ages (Number)_x', 'Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)_x', 'Deaths - Drug use - Sex: Both - Age: All Ages (Number)_x', 'Deaths - Diet low in fruits - Sex: Both - Age: All Ages (Number)', 'Deaths - Diet low in vegetables - Sex: Both - Age: All Ages (Number)', 'Deaths - Unsafe sex - Sex: Both - Age: All Ages (Number)', 'Deaths - Low physical activity - Sex: Both - Age: All Ages (Number)', 'Deaths - High fasting plasma glucose - Sex: Both - Age: All Ages (Number)', 'Deaths - High body-mass index - Sex: Both - Age: All Ages (Number)', 'Deaths - High systolic blood pressure - Sex: Both - Age: All Ages (Number)', 'Deaths - Smoking - Sex: Both - Age: All Ages (Number)', 'Deaths - Iron deficiency - Sex: Both - Age: All Ages (Number)', 'Deaths - Vitamin A deficiency - Sex: Both - Age: All Ages (Number)', 'Deaths - Low bone mineral density - Sex: Both - Age: All Ages (Number)', 'Deaths - Air pollution - Sex: Both - Age: All Ages (Number)', 'Deaths - Outdoor air pollution (all ages) (IHME)', 'Deaths - Diet high in sodium - Sex: Both - Age: All Ages (Number)', 'Deaths - Diet low in whole grains - Sex: Both - Age: All Ages (Number)', 'Deaths - Diet low in nuts and seeds - Sex: Both - Age: All Ages (Number)', 'Code_x', 'Deaths - Alcohol use disorders - Sex: Both - Age: Age-standardized (Rate)', 'Code_y', 'Deaths - Amphetamine use disorders -
```

Sex: Both - Age: Age-standardized (Rate)', 'Code_x', 'Deaths - Opioid use disorders - Sex: Both - Age: Age-standardized (Rate)', 'Code_y', 'Deaths - Cocaine use disorders - Sex: Both - Age: Age-standardized (Rate)', 'Code_x', 'Deaths - Drug use disorders - Sex: Both - Age: Age-standardized (Rate)', 'Code_y', 'Deaths - Alcohol and substance use disorders', 'Code_x', 'Deaths - Drug use disorders - Sex: Both - Age: Under 5 (Number)', 'Deaths - Drug use disorders - Sex: Both - Age: 70+ years (Number)', 'Deaths - Drug use disorders - Sex: Both - Age: 15-49 years (Number)', 'Deaths - Drug use disorders - Sex: Both - Age: 50-69 years (Number)', 'Deaths - Drug use disorders - Sex: Both - Age: 5-14 years (Number)', 'Code_y', 'Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)_y', 'Deaths - Substance use disorders - Sex: Both - Age: 70+ years (Number)', 'Deaths - Substance use disorders - Sex: Both - Age: 50-69 years (Number)', 'Deaths - Substance use disorders - Sex: Both - Age: 15-49 years (Number)', 'Deaths - Substance use disorders - Sex: Both - Age: Under 5 (Number)', 'Deaths - Substance use disorders - Sex: Both - Age: 5-14 years (Number)', 'Code_y', 'Deaths - Cocaine use disorders - Sex: Both - Age: All Ages (Number)', 'Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)_x', 'Deaths - Opioid use disorders - Sex: Both - Age: All Ages (Number)', 'Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)_y', 'Deaths - Other drug use disorders - Sex: Both - Age: All Ages (Number)', 'Deaths - Amphetamine use disorders - Sex: Both - Age: All Ages (Number)', 'Code_x', 'Deaths - Tobacco - Sex: Both - Age: All Ages (Number)', 'Deaths - Alcohol use - Sex: Both - Age: All Ages (Number)_y', 'Deaths - Drug use - Sex: Both - Age: All Ages (Number)', 'Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)_y', 'Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)', 'Code_y', 'Deaths - Mental and substance use disorders - Sex: Both - Age: Under 5 (Rate)', 'Deaths - Mental and substance use disorders - Sex: Both - Age: 70+ years (Rate)', 'Deaths - Mental and substance use disorders - Sex: Both - Age: 5-14 years (Rate)', 'Deaths - Mental and substance use disorders - Sex: Both - Age: 15-49 years (Rate)', 'Deaths - Mental and substance use disorders - Sex: Both - Age: 50-69 years (Rate)', 'Deaths - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)', 'Deaths - Mental and substance use disorders - Sex: Both - Age: All Ages (Rate)', 'Code', 'Deaths - Eating disorders - Sex: Both - Age: All Ages (Number)']

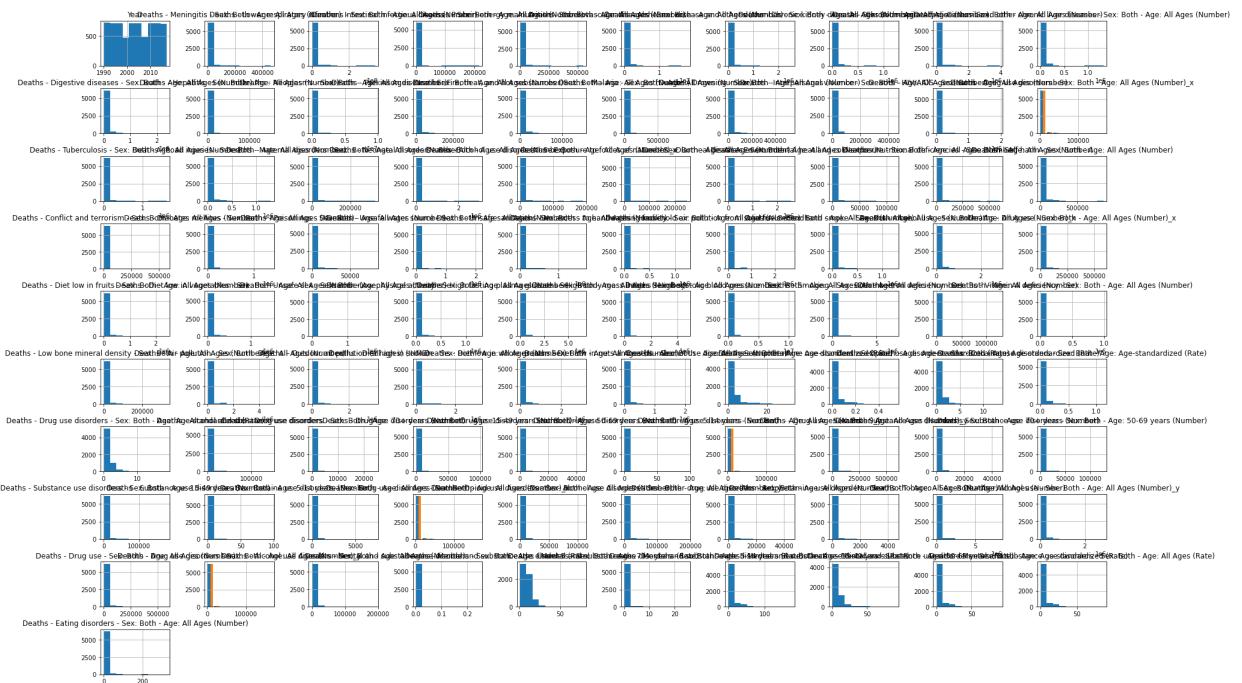
```
In [1959]: labels=['Code_x', 'Code_y']
deaths.drop(labels=labels, axis=1, inplace=True)
```

```
In [1960]: #dropping data for children <5
deaths.drop(labels=['Deaths - Drug use disorders - Sex: Both - Age: Under 5',
```

```
In [1961]: deaths.shape
```

```
Out[1961]: (6686, 93)
```

```
In [1962]: deaths.hist(figsize=(30,20))
plt.subplots_adjust(hspace=0.5, wspace=0.5)
```



```
In [1963]: deaths_mean = deaths.groupby('Year').mean()
deaths_mean.head()
```

Out[1963]:

	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)	Deaths - Intestinal infectious diseases - Sex: Both - Age: All Ages (Number)	Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)	Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)	Deaths - Alzheimer disease and other dementias - Sex: Both - Age: All Ages (Number)	Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number)
Year							
1990	9315.515696	68796.991468	3886.642079	10343.740269	249393.947809	21745.965432	12516.418
1991	9282.691062	68232.511287	3884.848451	10227.505913	252961.761354	22546.504184	12748.981
1992	9286.065424	67599.619178	3883.872305	10056.585540	257417.300819	23287.027013	13074.676
1993	9235.644602	66898.636450	3870.870643	9880.162598	265107.130545	24139.956184	13372.735

```
In [1964]: deaths.Entity.value_counts()
```

```
Out[1964]: Algeria          28
New Zealand        28
Burundi            28
Uganda             28
Lebanon            28
..
East Germany (GDR)    1
Saint Kitts and Nevis 1
French Polynesia     1
New Caledonia       1
West Germany (FRG)   1
Name: Entity, Length: 257, dtype: int64
```

```
In [1965]: #deaths.to_csv(r'Data/deaths.csv')
```

Continued in EnvironMentalHealth_02_DataWranglingEDAPart2.ipynb