I : Learning the Data from Columns

preface:

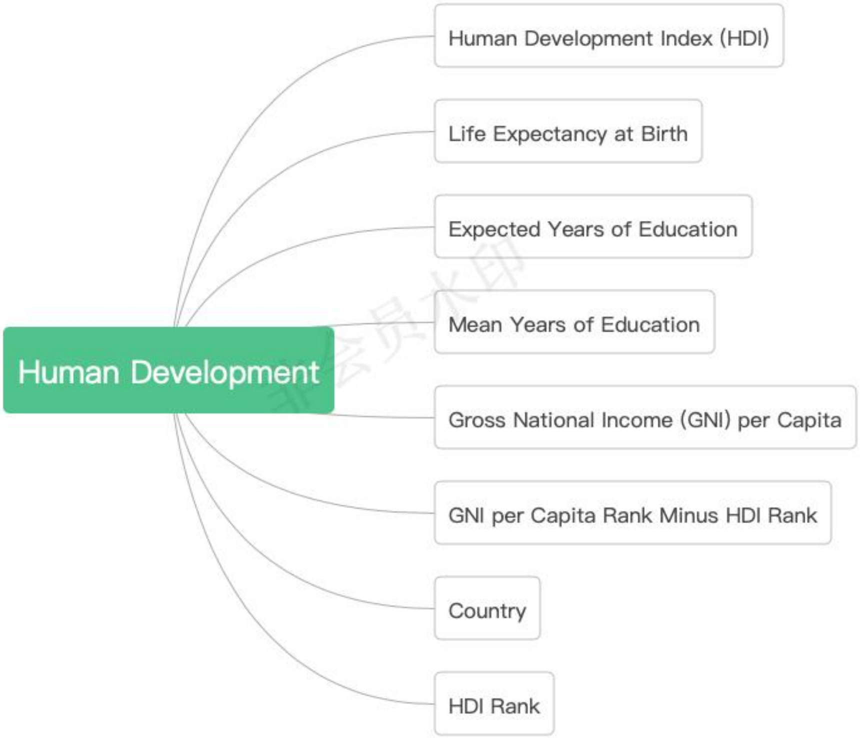
Learning the datasets we have is always a very important step. Especially we have 6 datasets this time. And the columns of the datasets are very abstract compare to the previous work. We need to learn the meaning of different indexs in order to comp up with meaningful questions.

Steps:

1. Human Development :

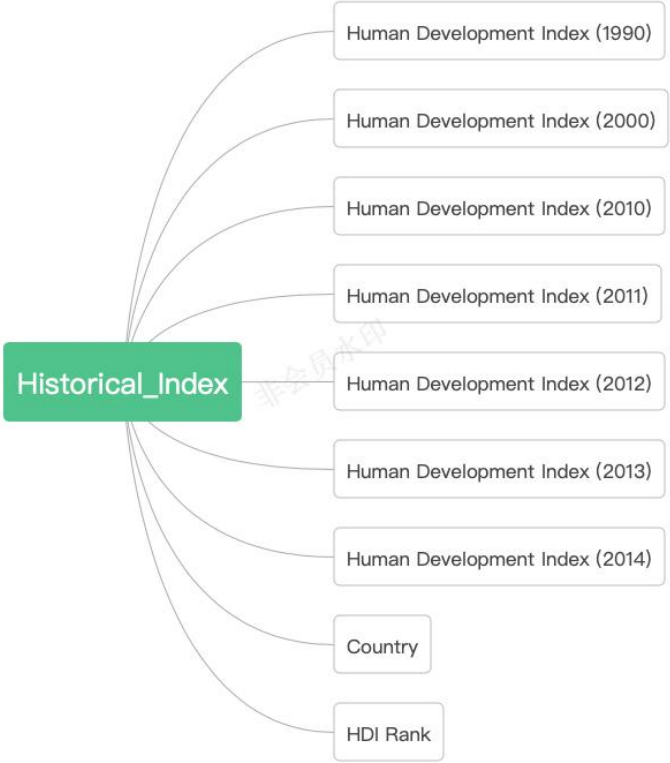
The Human Development Index (HDI) is a summary measure of achievements in key dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living. The HDI is the geometric mean of normalized indices for each of the three dimensions.

$$HDI = (IHealth \,.\, IEducation \,.\, IIncome)^{1/3}$$


Human Development
Human Development Index (HDI)
Life Expectancy at Birth
Expected Years of Education
Mean Years of Education
Gross National Income (GNI) per Capita
GNI per Capita Rank Minus HDI Rank
Country
HDI Rank

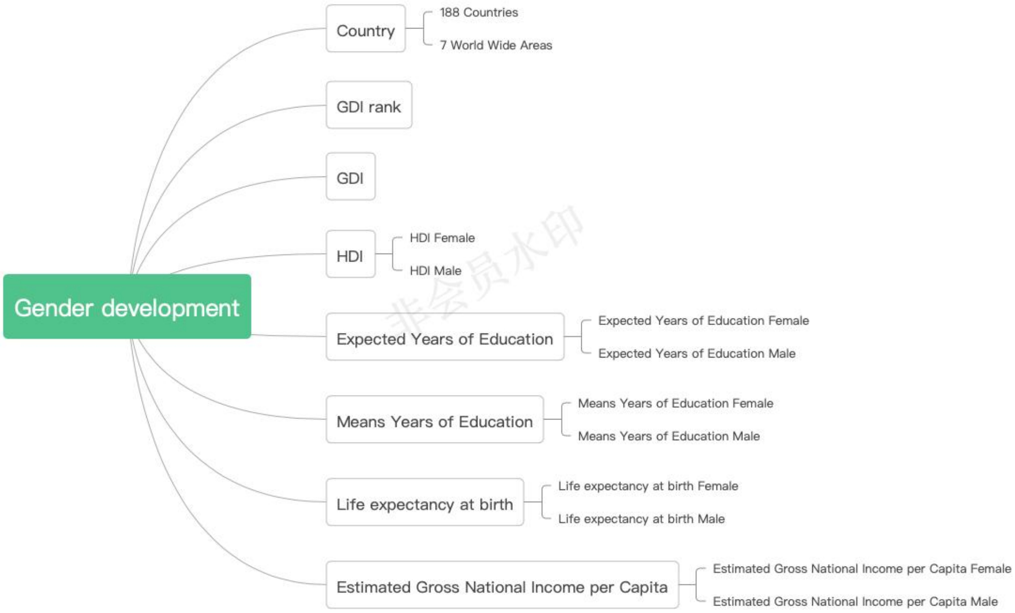
2. historical index :

The historical index are the HDI index from [1990,2000,2010,2011,2012,2013,2014]


Historical_Index
Human Development Index (1990)
Human Development Index (2000)
Human Development Index (2010)
Human Development Index (2011)
Human Development Index (2012)
Human Development Index (2013)
Human Development Index (2014)
Country
HDI Rank

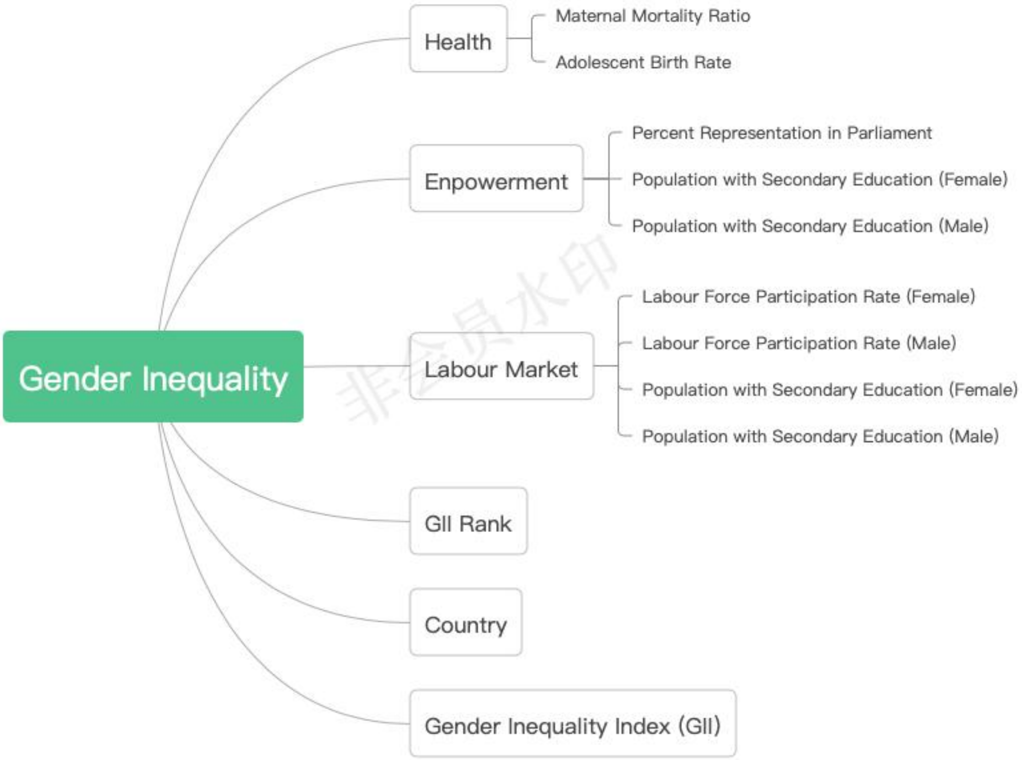
3. gender development :

The Gender Development Index (GDI) measures gender inequalities in achievement in three basic dimensions of human development: health, measured by female and male life expectancy at birth; education, measured by female and male expected years of schooling for children and female and male mean years of schooling for adults ages 25 and older; and command over economic resources, measured by female and male estimated earned income.
The GDI is simply the ratio of female HDI to male HDI: $GDI = HDI_f / HDI_m$


Gender development
Country
188 Countries
7 World Wide Areas
GDI rank
GDI
HDI
HDI Female
HDI Male
Expected Years of Education
Expected Years of Education Female
Expected Years of Education Male
Means Years of Education
Means Years of Education Female
Means Years of Education Male
Life expectancy at birth
Life expectancy at birth Female
Life expectancy at birth Male
Estimated Gross National Income per Capita
Estimated Gross National Income per Capita Female
Estimated Gross National Income per Capita Male

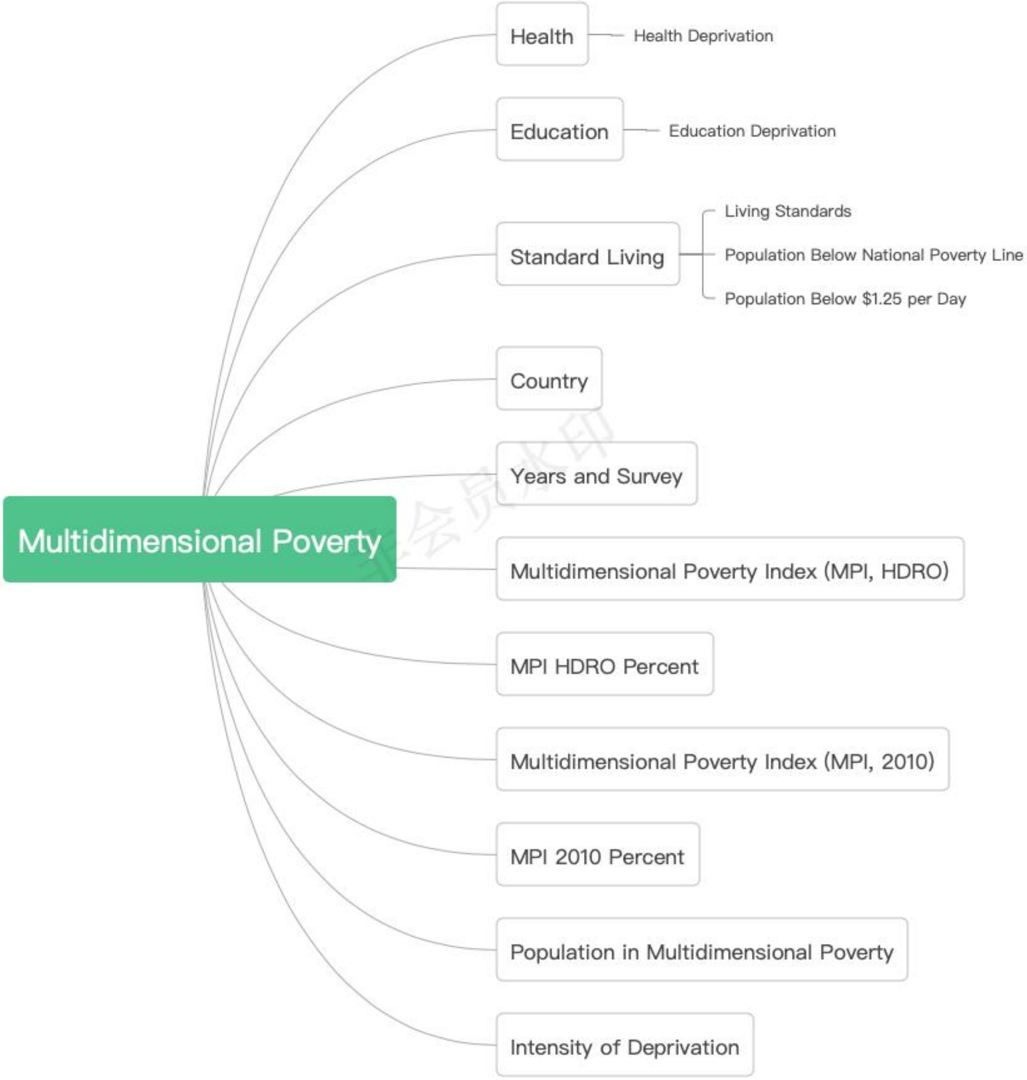
4. gender inequality :

The Gender Inequality Index (GII) reflects gender-based disadvantage in three dimensions—reproductive health, empowerment and the labour market—for as many countries as data of reasonable quality allow. It shows the loss in potential human development due to inequality between female and male achievements in these dimensions. It ranges between 0, where women and men fare equally, and 1, where one gender fares as poorly as possible in all measured dimensions.

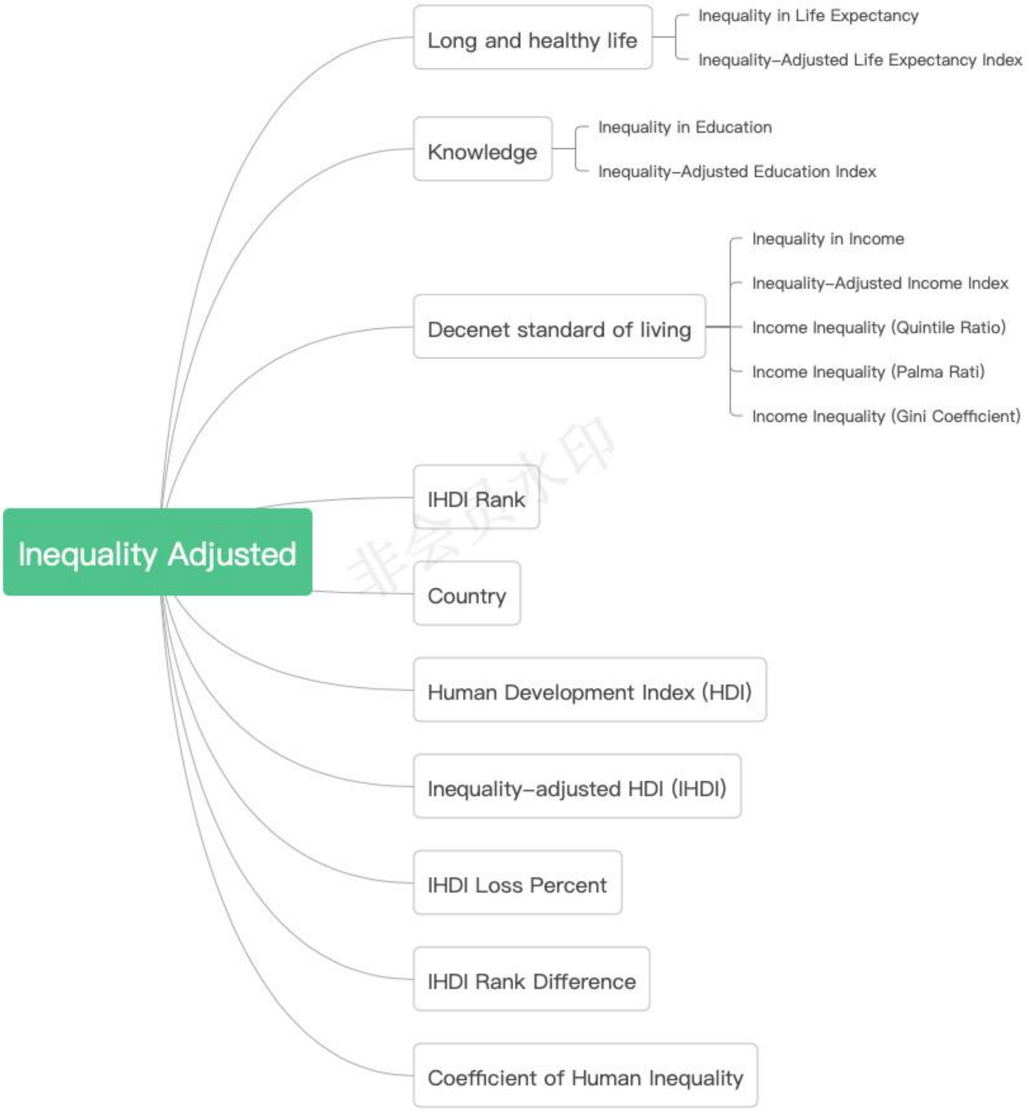

Gender Inequality
Health
Maternal Mortality Ratio
Adolescent Birth Rate
Enpowerment
Percent Representation in Parliament
Population with Secondary Education (Female)
Population with Secondary Education (Male)
Labour Market
Labour Force Participation Rate (Female)
Labour Force Participation Rate (Male)
Population with Secondary Education (Female)
Population with Secondary Education (Male)
GII Rank
Country
Gender Inequality Index (GII)

5. multidimension poverty :

The Multidimensional Poverty Index (MPI) identifies multiple deprivations at the household level in education, health and standard of living.


Multidimensional Poverty
Health
Health Deprivation
Education
Education Deprivation
Standard Living
Living Standards
Population Below National Poverty Line
Population Below $1.25 per Day
Country
Years and Survey
Multidimensional Poverty Index (MPI, HDRO)
MPI HDRO Percent
Multidimensional Poverty Index (MPI, 2010)
MPI 2010 Percent
Population in Multidimensional Poverty
Intensity of Deprivation

6. Inequality adjusted :


Inequality Adjusted
Long and healthy life
Inequality in Life Expectancy
Inequality–Adjusted Life Expectancy Index
Knowledge
Inequality in Education
Inequality–Adjusted Education Index
Decenet standard of living
Inequality in Income
Inequality–Adjusted Income Index
Income Inequality (Quintile Ratio)
Income Inequality (Palma Rati)
Income Inequality (Gini Coefficient)
IHDI Rank
Country
Human Development Index (HDI)
Inequality–adjusted HDI (IHDI)
IHDI Loss Percent
IHDI Rank Difference
Coefficient of Human Inequality

II. Learning the Data from Plots

Preface:

In the previous steps, we know what columns do we have in each datasets and their meanings, however, the columns are very abstract, in order to have a better idea about the datasets, we are going to plot some images to help us understanding the quantity of the data. Trying to find some simple relationship betweeen the columns.

Step1:

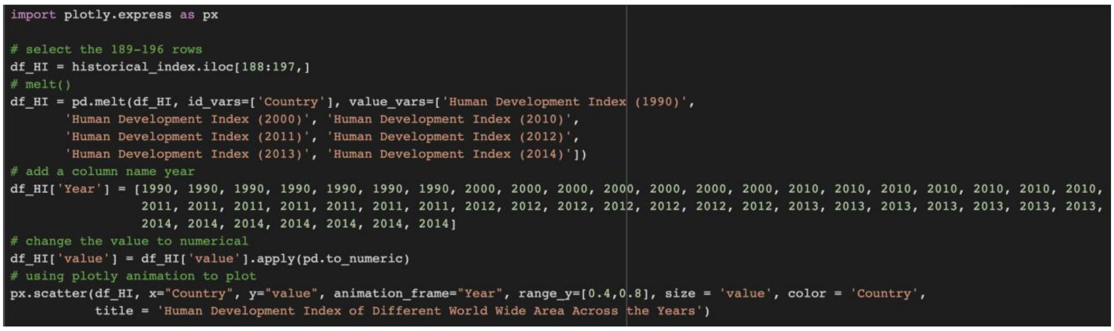
As the usual, we use pd.head( ), pd.info( ), pd.describe( ) to have a basic view of the datasets. One thing we notice is the : there are 188 different countries in the dataset and the last 7 rows of the dataset are the area in the world. They are : Arab States, East Asia and the Pacific, Europe and Central Asia, Latin America and the Caribbean, South Asia, Sub-Saharan Africa and World.

Step2:

a. Historical Index Plot

The first plot is about historical index, simply because it is the most direct dataset we have.

we select only the area of the world to plot, otherwise 188 contries will be a little bit messy. we add a columns contain the years and melt the dataset to vertical. using plotly.express to plot an animation plot.



```
import plotly.express as px

# select the 189-196 rows
df_HI = historical_index.iloc[188:197,]
# melt()
df_HI = pd.melt(df_HI, id_vars=['Country'], value_vars=['Human Development Index (1990)',
       'Human Development Index (2000)', 'Human Development Index (2010)',
       'Human Development Index (2011)', 'Human Development Index (2012)',
       'Human Development Index (2013)', 'Human Development Index (2014)'])
# add a column name year
df_HI['Year'] = [1990, 1990, 1990, 1990, 1990, 1990, 1990, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2010, 2010, 2010, 2010, 2010, 2010, 2010,
                 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2012, 2012, 2012, 2012, 2012, 2012, 2012, 2013, 2013, 2013, 2013, 2013, 2013, 2013,
                 2014, 2014, 2014, 2014, 2014, 2014, 2014]
# change the value to numerical
df_HI['value'] = df_HI['value'].apply(pd.to_numeric)
# using plotly animation to plot
px.scatter(df_HI, x="Country", y="value", animation_frame="Year", range_y=[0.4,0.8], size = 'value', color = 'Country',
           title = 'Human Development Index of Different World Wide Area Across the Years')
```

figure 2.1

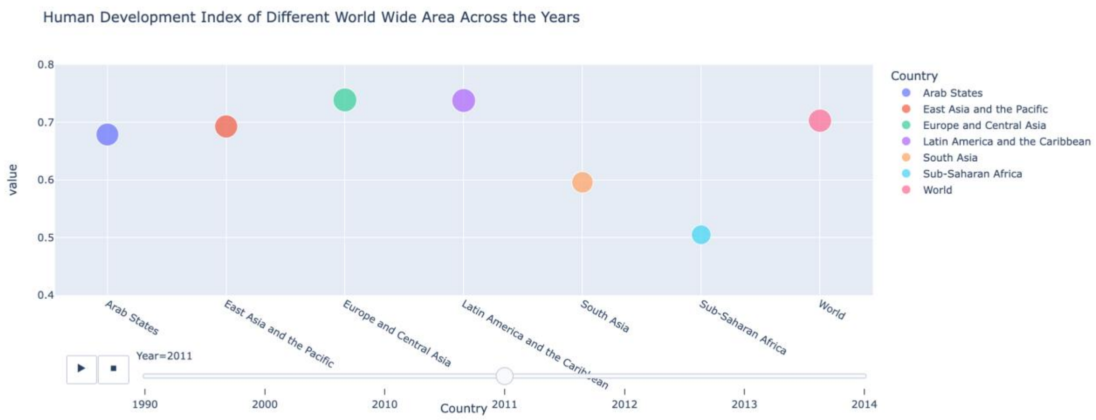
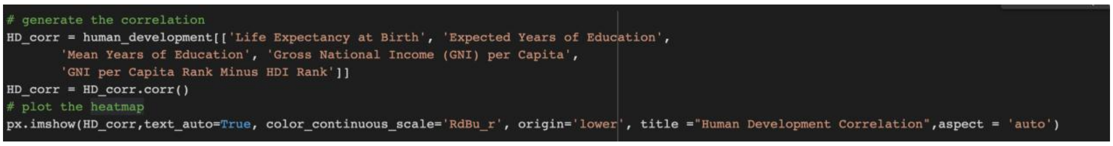


figure 2.2

As you can see above, the Europe and Central Asia, Latin America and the Caribbean are the areas with higher HDI. Sub-Saharan Africa and South Asia have lower HDI.
What you can’t see in this static plot is as the year increase, the HDI of each erea are incresing, which make perfect sense, human are development.

b. Heatmap and pair plot of the Human Development

The heatmap is very good at displaying the linear relationship of the columns and pair map can show the linear and non-linear relationship.



```
# generate the correlation
HD_corr = human_development[['Life Expectancy at Birth', 'Expected Years of Education',
       'Mean Years of Education', 'Gross National Income (GNI) per Capita',
       'GNI per Capita Rank Minus HDI Rank']]
HD_corr = HD_corr.corr()
# plot the heatmap
px.imshow(HD_corr,text_auto=True, color_continuous_scale='RdBu_r', origin='lower', title ="Human Development Correlation",aspect = 'auto')
```

figure 2.3

Life Expectancy at Birth,Mean Years of Education, Expected Years of Education and GNI per capita have strong positive linear relationship.

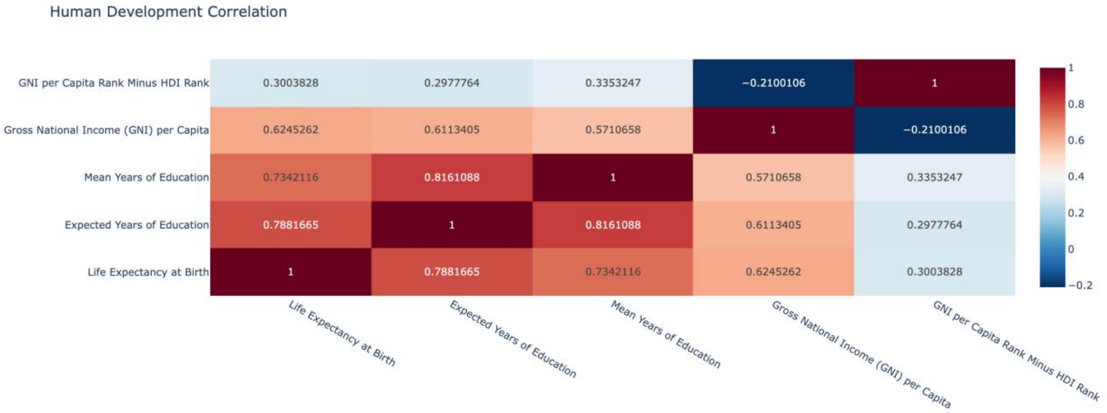
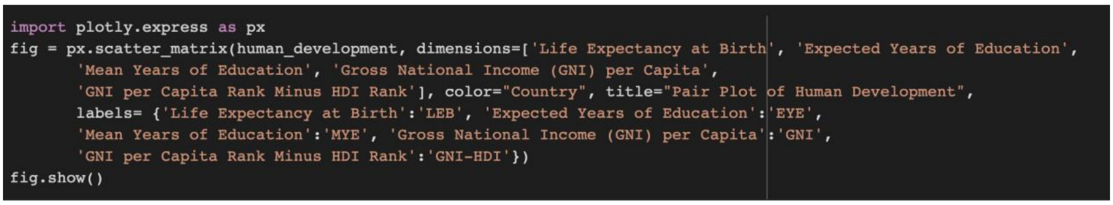


figure 2.4



```
import plotly.express as px
fig = px.scatter_matrix(human_development, dimensions=['Life Expectancy at Birth', 'Expected Years of Education',
        'Mean Years of Education', 'Gross National Income (GNI) per Capita',
        'GNI per Capita Rank Minus HDI Rank'], color="Country", title="Pair Plot of Human Development",
        labels= {'Life Expectancy at Birth':'LEB', 'Expected Years of Education':'EYE',
        'Mean Years of Education':'MYE', 'Gross National Income (GNI) per Capita':'GNI',
        'GNI per Capita Rank Minus HDI Rank':'GNI-HDI'})
fig.show()
```

figure 2.5

In the pair plot, we have the same conclution of heatmap.

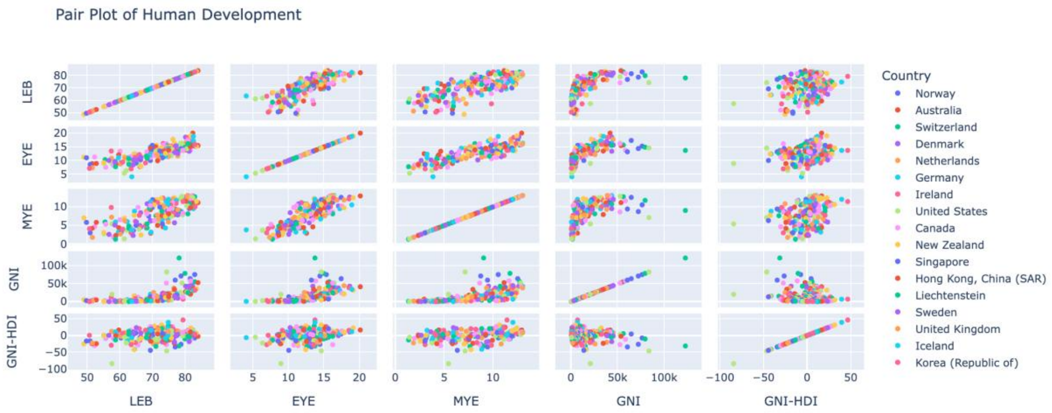

Pair Plot of Human Development
LEB
EYE
MYE
GNI
GNI-HDI
80
70
60
50
20
15
10
5
10
5
0
100k
50k
0
50
0
−50
−100
50
60
70
80
5
10
15
20
0
5
10
0
50k
100k
−100
−50
0
50
LEB
EYE
MYE
GNI
GNI-HDI
Country
Norway
Australia
Switzerland
Denmark
Netherlands
Germany
Ireland
United States
Canada
New Zealand
Singapore
Hong Kong, China (SAR)
Liechtenstein
Sweden
United Kingdom
Iceland
Korea (Republic of)

figure 2.6

III. Come up with Meaningful Questions

Question 1:

1. What are the top 20 education countries?

Question 2:

Backgroud:

There are some noticeable revolution around the world in the recent years, like HongKong protesting about the atonomy, Iran protesting about the inflation and woman right, Russia protesting about peace, China protesting about the epidemic provention policy, Myanmar protesting the junta.etc.

What we find common in these areas is these countries are essential one party county and one party or one person has the most significant power to control the whole sociaty. What we interesting are the human right and gender equality in theses countries.

Also, we want to add our list with communist contries around the world right now. Because the communist contries right now in the world are essentially one party contries, the institutions of the ruling communist party and the state have become intertwined.

Right now the communist contries in the world has only 5 countries: China, Cuba, Laos, Vietnam, North Korea, however, one thing need to mention is Russia and Iran are once bleong to communist contries.

According to Wekipidia, the current one-party states are communist contries plus Eritrea and Sahrawi Arab Democratic Republic.

2. What's the level of HDI, historical of HDI MPI, GII in these contries compare with other country?(plot these countries only and plot the whole dataset in subplots)

IV. Preprocessing and Visualization

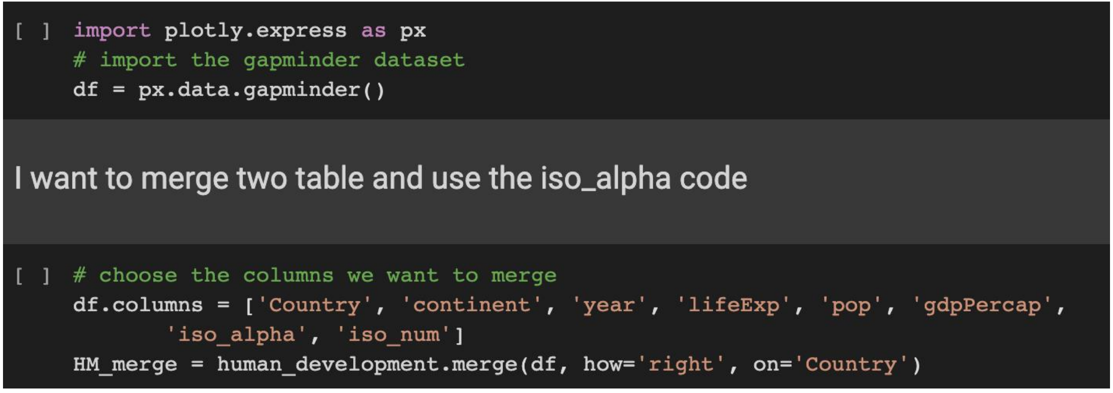
preface:

Right now we have the questions. We gonna use our visualization technics to present the relations.

Procedure:

1. Question1: What are the top 20 education countries?

Step1: merge human_development datasets with gapmider dataset.
The gapmider dataset contain the location of each countries and relevent gpd. We need to use the location information to show our education data in the world map.



```
[ ]  import plotly.express as px
     # import the gapminder dataset
     df = px.data.gapminder()
```

I want to merge two table and use the iso_alpha code

```
[ ]  # choose the columns we want to merge
     df.columns = ['Country', 'continent', 'year', 'lifeExp', 'pop', 'gdpPercap',
            'iso_alpha', 'iso_num']
     HM_merge = human_development.merge(df, how='right', on='Country')
```

figure 4.1

Step2: plot the ‘Mean Education Years’ with whole dataset, 188 countries.

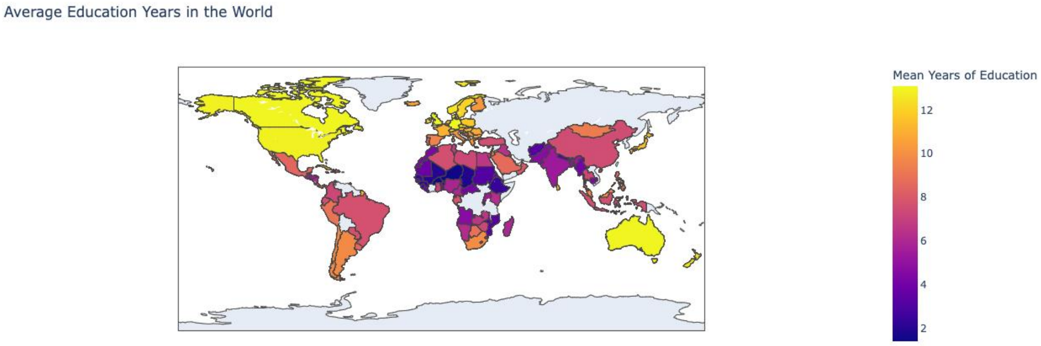
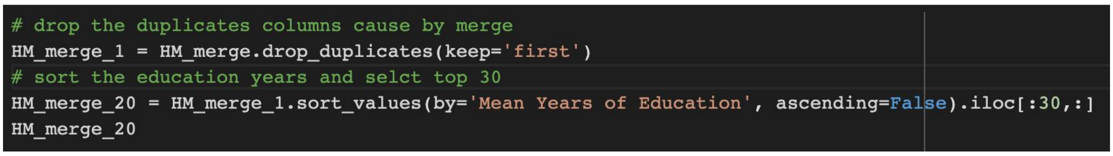


figure 4.2

Step3: Select the top 20 cuntries and plot



```
# drop the duplicates columns cause by merge
HM_merge_1 = HM_merge.drop_duplicates(keep='first')
# sort the education years and selct top 30
HM_merge_20 = HM_merge_1.sort_values(by='Mean Years of Education', ascending=False).iloc[:30,:]
HM_merge_20
```

figure 4.3

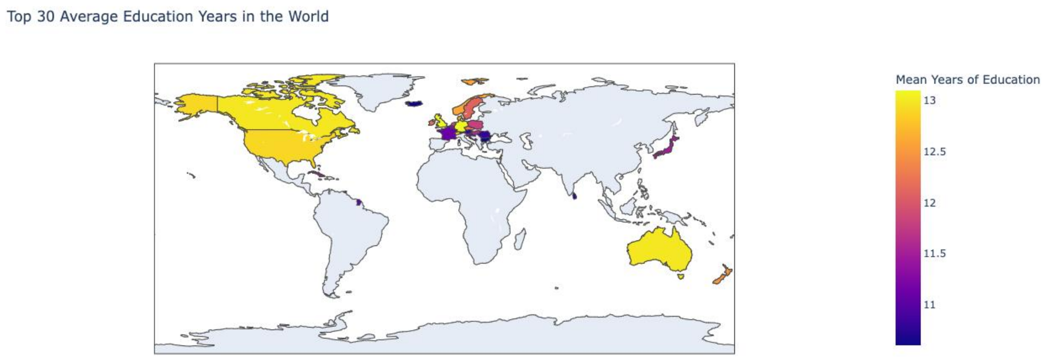


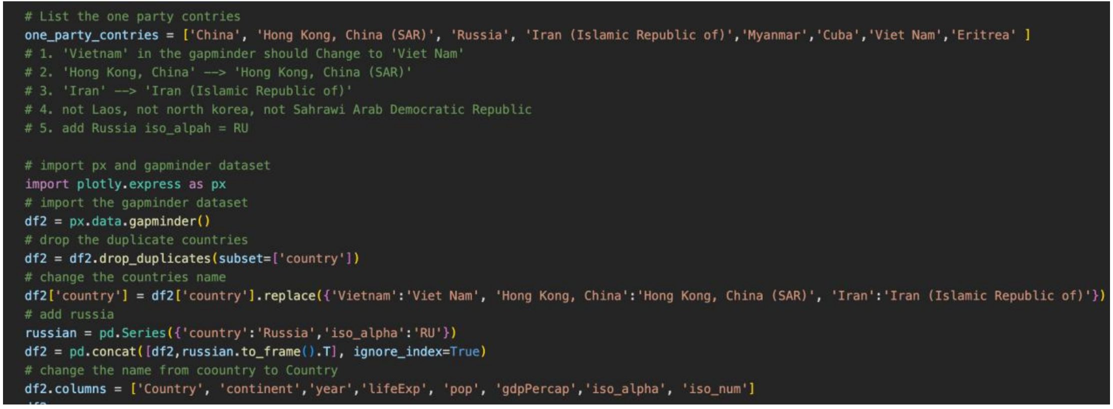
figure 4.4

As you can see, the top 13 countries' locations in the map. Canada, USA, Austrilia, some eruopian countries have long education years. The longest education year is German 13.1 years.

2. Question2:

preprocess the dataset:

In order to present the location of different countries, we need to merge our dataset with gapminder dataset, which everyone can import from the plotly. we drop some useless rows in gapminder and change the name of some countries since different dataset have different names of the countries in some cases. And finally we add Russia and its iso_alpha into gapminder and change the column name 'country' to 'Country' in order to merge on the 'Country' columns.

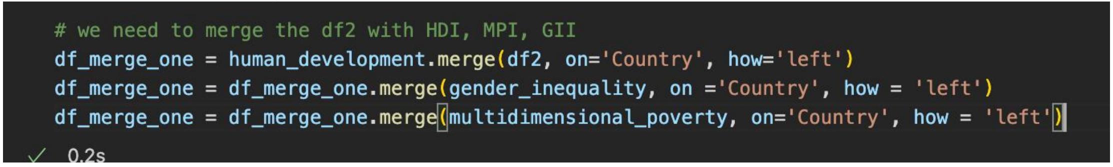


```
# List the one party contries
one_party_contries = ['China', 'Hong Kong, China (SAR)', 'Russia', 'Iran (Islamic Republic of)','Myanmar','Cuba','Viet Nam','Eritrea' ]
# 1. 'Vietnam' in the gapminder should Change to 'Viet Nam'
# 2. 'Hong Kong, China' --> 'Hong Kong, China (SAR)'
# 3. 'Iran' --> 'Iran (Islamic Republic of)'
# 4. not Laos, not north korea, not Sahrawi Arab Democratic Republic
# 5. add Russia iso_alpah = RU

# import px and gapminder dataset
import plotly.express as px
# import the gapminder dataset
df2 = px.data.gapminder()
# drop the duplicate countries
df2 = df2.drop_duplicates(subset=['country'])
# change the countries name
df2['country'] = df2['country'].replace({'Vietnam':'Viet Nam', 'Hong Kong, China':'Hong Kong, China (SAR)', 'Iran':'Iran (Islamic Republic of)'})
# add russia
russian = pd.Series({'country':'Russia','iso_alpha':'RU'})
df2 = pd.concat([df2,russian.to_frame().T], ignore_index=True)
# change the name from coountry to Country
df2.columns = ['Country', 'continent','year','lifeExp', 'pop', 'gdpPercap','iso_alpha', 'iso_num']
```

figure 4.5

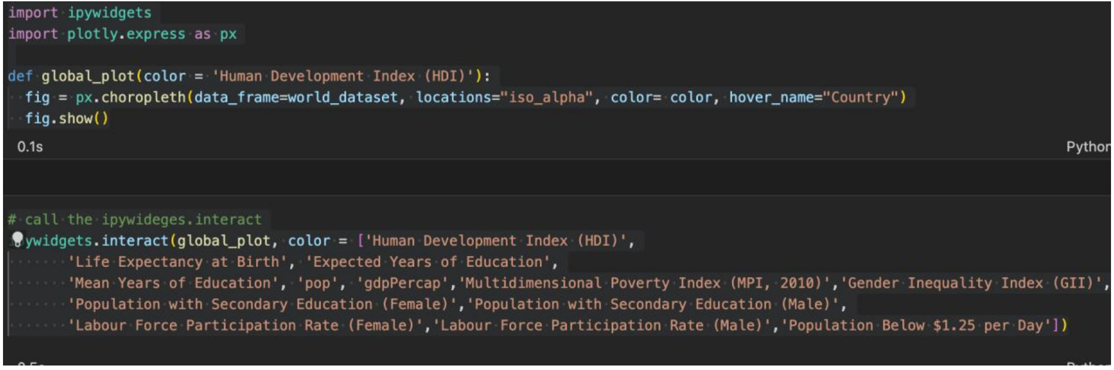
Then, we merge the gapminder dataset with Human_development, gender_inequallity and multidimension_poverty



```
# we need to merge the df2 with HDI, MPI, GII
df_merge_one = human_development.merge(df2, on='Country', how='left')
df_merge_one = df_merge_one.merge(gender_inequality, on ='Country', how = 'left')
df_merge_one = df_merge_one.merge(multidimensional_poverty, on='Country', how = 'left')
✓ 0.2s
```

figure 4.6

we use ipywideget to implement a drop-down-menu, so we can interact with the plot and select the information you want to display.



```
import ipywidgets
import plotly.express as px

def global_plot(color = 'Human Development Index (HDI)'):
  fig = px.choropleth(data_frame=world_dataset, locations="iso_alpha", color= color, hover_name="Country")
  fig.show()
```

0.1s Python

```
# call the ipywideges.interact
ipywidgets.interact(global_plot, color = ['Human Development Index (HDI)',
      'Life Expectancy at Birth', 'Expected Years of Education',
      'Mean Years of Education', 'pop', 'gdpPercap','Multidimensional Poverty Index (MPI, 2010)','Gender Inequality Index (GII)',
      'Population with Secondary Education (Female)','Population with Secondary Education (Male)',
      'Labour Force Participation Rate (Female)','Labour Force Participation Rate (Male)','Population Below $1.25 per Day'])
```

figure 4.7

we can display the whole world

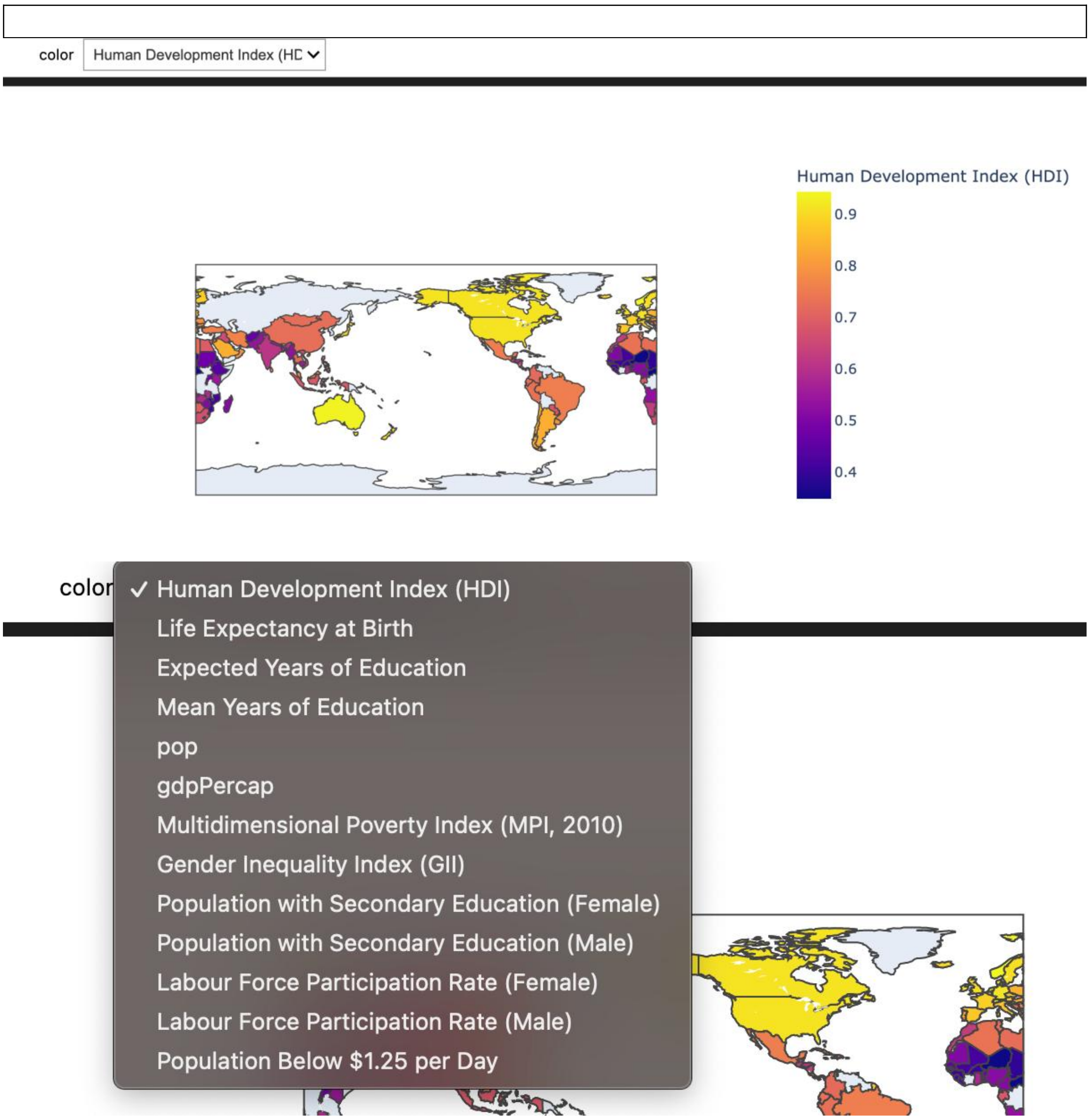
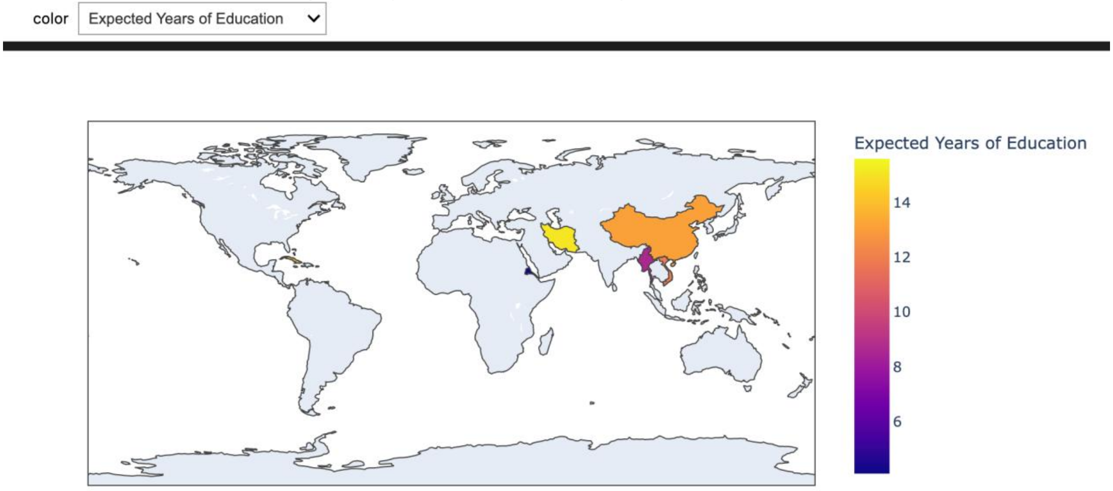
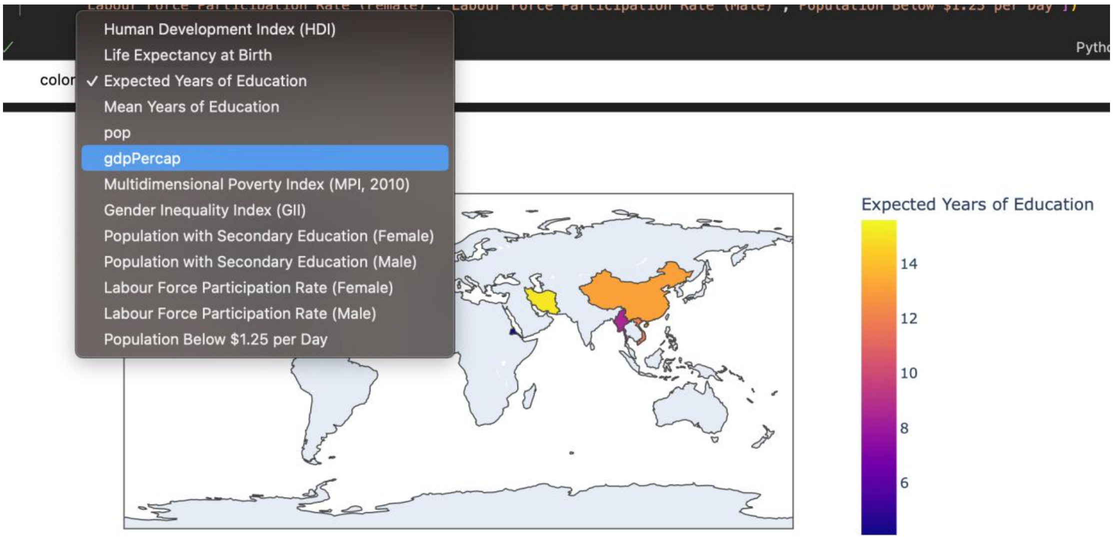

color
Human Development Index (HD
Human Development Index (HDI)
0.9
0.8
0.7
0.6
0.5
0.4
color
✓ Human Development Index (HDI)
Life Expectancy at Birth
Expected Years of Education
Mean Years of Education
pop
gdpPercap
Multidimensional Poverty Index (MPI, 2010)
Gender Inequality Index (GII)
Population with Secondary Education (Female)
Population with Secondary Education (Male)
Labour Force Participation Rate (Female)
Labour Force Participation Rate (Male)
Population Below $1.25 per Day

figure 4.8

Also we can plot the images only contains one party contries and compare.



color Expected Years of Education



```
'Labour Force Participation Rate (Female)', 'Labour Force Participation Rate (Male)', 'Population Below $1.25 per Day'])
```

color ✓ Expected Years of Education

Pyth
Human Development Index (HDI)
Life Expectancy at Birth
✓ Expected Years of Education
Mean Years of Education
pop
gdpPercap
Multidimensional Poverty Index (MPI, 2010)
Gender Inequality Index (GII)
Population with Secondary Education (Female)
Population with Secondary Education (Male)
Labour Force Participation Rate (Female)
Labour Force Participation Rate (Male)
Population Below $1.25 per Day
Expected Years of Education
14
12
10
8
6

figure 4.9

As a conclusion, we can easy to see that the one party countries have lower HDI, GII, MPI than the average level.

V. Summary and Challenges

Summary:

To Summarize, we use heatmap, pairplot, animation, map, ipywidegets to plot severeal different plots. After using these visualization technology, we can clearly see the relationship between columns and the different index in different countries.

More specific, we see the top 30 contries which have high education level and the different indexes level of one party contries in the world.

Challenges:

1. Very unfortunately, some countries we want to analysis don’t have data in our datasets.
2. In the future, we want to learn how to use ipywidgets to plot interactive subplots.