

Alzheimer Disease and Healthy Aging Data In US

March 27, 2024

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

0.0.1 Alzheimer Disease and Healthy Aging Data In US.csv

```
[2]: data = pd.read_csv('data/Alzheimer Disease and Healthy Aging Data In US.csv')
```

D:\DevelopTools\Anaconda\lib\site-packages\IPython\core\interactiveshell.py:3444: DtypeWarning: Columns (13,14) have mixed types.Specify dtype option on import or set low_memory=False.
exec(code_obj, self.user_global_ns, self.user_ns)

```
[3]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 214462 entries, 0 to 214461
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   YearStart                             214462 non-null int64
1   YearEnd                               214462 non-null int64
2   LocationAbbr                          214462 non-null object
3   LocationDesc                          214462 non-null object
4   Datasource                            214462 non-null object
5   Class                                 214462 non-null object
6   Topic                                 214462 non-null object
7   Question                              214462 non-null object
8   Data_Value_Unit                       214462 non-null object
9   DataValueTypeID                       214462 non-null object
10  Data_Value_Type                       214462 non-null object
11  Data_Value                            144629 non-null float64
12  Data_Value_Alt                        144629 non-null float64
13  Low_Confidence_Limit                  144472 non-null object
14  High_Confidence_Limit                  144472 non-null object
15  Sample_Size                           0 non-null      float64
16  StratificationCategory1                214462 non-null object
17  Stratification1                        214462 non-null object
18  StratificationCategory2                186793 non-null object
19  Stratification2                        186793 non-null object
```

20	Geolocation	191413	non-null	object
21	ClassID	214462	non-null	object
22	TopicID	214462	non-null	object
23	QuestionID	214462	non-null	object
24	LocationID	214462	non-null	int64
25	StratificationCategoryID1	214462	non-null	object
26	StratificationID1	214462	non-null	object
27	StratificationCategoryID2	214462	non-null	object
28	StratificationID2	214462	non-null	object

dtypes: float64(3), int64(3), object(23)
memory usage: 47.5+ MB

0.0.2

```
[4]: for name, value in data.items():
      print(value.value_counts())
```

```
2015    45980
2020    36006
2019    34354
2016    34145
2017    33429
2018    30548
Name: YearStart, dtype: int64
2020    46966
2015    35020
2019    34354
2016    34145
2017    33429
2018    30548
Name: YearEnd, dtype: int64
US      4644
WEST    4638
NRE     4614
MDW     4611
OR      4565
NY      4557
SOU     4542
UT      4222
OH      3955
GA      3951
MD      3919
HI      3907
TN      3879
MI      3796
VA      3758
FL      3753
ME      3733
```

TX	3699
NV	3696
DC	3684
WV	3682
MS	3677
PA	3648
NM	3635
AL	3633
KY	3623
AK	3611
SC	3592
NJ	3589
AZ	3582
MO	3573
IL	3571
IN	3570
WI	3569
LA	3563
MN	3555
NE	3546
CT	3543
RI	3534
OK	3526
SD	3526
ND	3514
KS	3510
ID	3507
IA	3501
AR	3498
WY	3494
CA	3447
CO	3390
NC	3349
WA	3348
MT	3348
DE	3346
NH	3284
VT	3278
MA	3174
PR	2797
GU	2703
VI	503

Name: LocationAbbr, dtype: int64

United States, DC & Territories	4644
West	4638
Northeast	4614
Midwest	4611
Oregon	4565

New York	4557
South	4542
Utah	4222
Ohio	3955
Georgia	3951
Maryland	3919
Hawaii	3907
Tennessee	3879
Michigan	3796
Virginia	3758
Florida	3753
Maine	3733
Texas	3699
Nevada	3696
District of Columbia	3684
West Virginia	3682
Mississippi	3677
Pennsylvania	3648
New Mexico	3635
Alabama	3633
Kentucky	3623
Alaska	3611
South Carolina	3592
New Jersey	3589
Arizona	3582
Missouri	3573
Illinois	3571
Indiana	3570
Wisconsin	3569
Louisiana	3563
Minnesota	3555
Nebraska	3546
Connecticut	3543
Rhode Island	3534
Oklahoma	3526
South Dakota	3526
North Dakota	3514
Kansas	3510
Idaho	3507
Iowa	3501
Arkansas	3498
Wyoming	3494
California	3447
Colorado	3390
North Carolina	3349
Washington	3348
Montana	3348
Delaware	3346

New Hampshire	3284
Vermont	3278
Massachusetts	3174
Puerto Rico	2797
Guam	2703
Virgin Islands	503
Name: LocationDesc, dtype: int64	
BRFSS	214462
Name: Datasource, dtype: int64	
Overall Health	71694
Screenings and Vaccines	46867
Nutrition/Physical Activity/Obesity	24851
Cognitive Decline	19180
Caregiving	18671
Mental Health	16600
Smoking and Alcohol Use	16599
Name: Class, dtype: int64	
Obesity	
8300	
Influenza vaccine within past year	
8300	
Physically unhealthy days (mean number of days)	
8300	
Frequent mental distress	
8300	
Current smoking	
8300	
Lifetime diagnosis of depression	
8300	
No leisure-time physical activity within past month	
8300	
Self-rated health (fair to poor health)	
8299	
Self-rated health (good to excellent health)	
8299	
Binge drinking within past 30 days	
8299	
Ever had pneumococcal vaccine	
8268	
Recent activity limitations in past month	
8233	
Disability status, including sensory or mobility limitations	
6917	
Arthritis among older adults	
5511	
Fair or poor health among older adults with arthritis	
5447	
Subjective cognitive decline or memory loss among older adults	

5088
 Diabetes screening within past 3 years
 4808
 Talked with health care professional about subjective cognitive decline or
 memory loss 4700
 Need assistance with day-to-day activities because of subjective cognitive
 decline or memory loss 4696
 Functional difficulties associated with subjective cognitive decline or memory
 loss among older adults 4696
 Fall with injury within last year
 4173
 Colorectal cancer screening
 4173
 Oral health: tooth retention
 4172
 Prevalence of sufficient sleep
 4171
 Eating 3 or more vegetables daily
 4127
 High blood pressure ever
 4127
 Cholesterol checked in past 5 years
 4127
 Eating 2 or more fruits daily
 4124
 Taking medication for high blood pressure
 4108
 Severe joint pain among older adults with arthritis
 4064
 Provide care for a friend or family member in past month
 3848
 Expect to provide care for someone in the next two years
 3797
 Provide care for someone with cognitive impairment within the past month
 3682
 Duration of caregiving among older adults
 3681
 Intensity of caregiving among older adults
 3663
 Up-to-date with recommended vaccines and screenings - Women
 3280
 Up-to-date with recommended vaccines and screenings - Men
 3271
 Mammogram within past 2 years
 3271
 Pap test within past 3 years
 3242
 Name: Topic, dtype: int64

Percentage of older adults who are currently obese, with a body mass index (BMI) of 30 or more
8300

Percentage of older adults who reported influenza vaccine within the past year
8300

Physically unhealthy days (mean number of days in past month)
8300

Percentage of older adults who are experiencing frequent mental distress
8300

Percentage of older adults who have smoked at least 100 cigarettes in their entire life and still smoke every day or some days
8300

Percentage of older adults with a lifetime diagnosis of depression
8300

Percentage of older adults who have not had any leisure time physical activity in the past month
8300

Percentage of older adults who self-reported that their health is "fair" or "poor"
8299

Percentage of older adults who self-reported that their health is "good", "very good", or "excellent"
8299

Percentage of older adults who reported binge drinking within the past 30 days
8299

Percentage of at risk adults (have diabetes, asthma, cardiovascular disease or currently smoke) who ever had a pneumococcal vaccine
8268

Mean number of days with activity limitations in the past month
8233

Percentage of older adults who report having a disability (includes limitations related to sensory or mobility impairments or a physical, mental, or emotional condition) 6917

Percentage of older adults ever told they have arthritis
5511

Fair or poor health among older adults with doctor-diagnosed arthritis
5447

Percentage of older adults who reported subjective cognitive decline or memory loss that is happening more often or is getting worse in the preceding 12 months
5088

Percentage of older adults without diabetes who reported a blood sugar or diabetes test within 3 years
4808

Percentage of older adults with subjective cognitive decline or memory loss who reported talking with a health care professional about it
4700

Percentage of older adults who reported that as a result of subjective cognitive decline or memory loss that they need assistance with day-to-day activities

4696

Percentage of older adults who reported subjective cognitive decline or memory loss that interferes with their ability to engage in social activities or household chores 4696

Percentage of older adults who have fallen and sustained an injury within last year

4173

Percentage of older adults who had either a home blood stool test within the past year or a sigmoidoscopy or colonoscopy within the past 10 years

4173

Percentage of older adults who report having lost 5 or fewer teeth due to decay or gum disease

4172

Percentage of older adults getting sufficient sleep (>6 hours)

4171

Percentage of older adults who are eating 3 or more vegetables daily

4127

Percentage of older adults who have ever been told by a health professional that they have high blood pressure

4127

Percentage of older adults who had a cholesterol screening within the past 5 years

4127

Percentage of older adults who are eating 2 or more fruits daily

4124

Percentage of older adults who have been told they have high blood pressure who report currently taking medication for their high blood pressure

4108

Severe joint pain due to arthritis among older adults with doctor-diagnosed arthritis

4064

Percentage of older adults who provided care for a friend or family member within the past month

3848

Percentage of older adults currently not providing care who expect to provide care for someone with health problems in the next two years

3797

Percentage of older adults who provided care for someone with dementia or other cognitive impairment within the past month

3682

Percentage of older adults who provided care to a friend or family member for six months or more

3681

Average of 20 or more hours of care per week provided to a friend or family member

3663

Percentage of older adult women who are up to date with select clinical preventive services

3280
Percentage of older adult men who are up to date with select clinical preventive services

3271
Percentage of older adult women who have received a mammogram within the past 2 years

3271
Percentage of older adult women with an intact cervix who had a Pap test within the past 3 years

3242
Name: Question, dtype: int64
% 197929
Number 16533
Name: Data_Value_Unit, dtype: int64
PRCTG 197929
MEAN 16533
Name: DataValueTypeID, dtype: int64
Percentage 197929
Mean 16533
Name: Data_Value_Type, dtype: int64

6.0	473
6.3	467
6.1	459
5.8	458
5.2	456
...	
99.7	2
0.5	1
0.6	1
0.3	1
0.2	1

Name: Data_Value, Length: 999, dtype: int64

6.0	473
6.3	467
6.1	459
5.8	458
5.2	456
...	
99.7	2
0.5	1
0.6	1
0.3	1
0.2	1

Name: Data_Value_Alt, Length: 999, dtype: int64

5.4	350
5.1	318
4.8	314
5.3	313

```

5          305
...
96.8      1
98.4      1
0.9       1
97.8      1
99.4      1
Name: Low_Confidence_Limit, Length: 1967, dtype: int64
6.5       216
5.8       193
6.8       192
6.7       189
7.5       186
...
1.7       1
2.8       1
2.8       1
1.5       1
1.6       1
Name: High_Confidence_Limit, Length: 1966, dtype: int64
Series([], Name: Sample_Size, dtype: int64)
Age Group    214462
Name: StratificationCategory1, dtype: int64
Overall              71919
50-64 years          71528
65 years or older    71015
Name: Stratification1, dtype: int64
Race/Ethnicity    134959
Gender              51834
Name: StratificationCategory2, dtype: int64
White, non-Hispanic    27633
Hispanic                27525
Black, non-Hispanic    26968
Native Am/Alaskan Native  26571
Asian/Pacific Islander  26262
Female                 26091
Male                   25743
Name: Stratification2, dtype: int64
POINT (-120.1550313 44.56744942)    4565
POINT (-75.54397043 42.82700103)    4557
POINT (-111.5871306 39.36070017)    4222
POINT (-82.40426006 40.06021014)    3955
POINT (-83.62758035 32.83968109)    3951
POINT (-76.60926011 39.29058096)    3919
POINT (-157.8577494 21.30485044)    3907
POINT (-85.77449091 35.68094058)    3879
POINT (-84.71439027 44.66131954)    3796
POINT (-78.45789046 37.54268067)    3758

```

POINT (-81.92896054 28.93204038)	3753
POINT (-68.98503134 45.25422889)	3733
POINT (-99.42677021 31.82724041)	3699
POINT (-117.0718406 39.49324039)	3696
POINT (-77.036871 38.907192)	3684
POINT (-80.71264013 38.6655102)	3682
POINT (-89.53803082 32.7455101)	3677
POINT (-77.86070029 40.79373015)	3648
POINT (-106.240581 34.52088095)	3635
POINT (-86.63186076 32.84057112)	3633
POINT (-84.77497105 37.64597027)	3623
POINT (-147.722059 64.84507996)	3611
POINT (-81.04537121 33.9988213)	3592
POINT (-74.27369129 40.13057005)	3589
POINT (-111.7638113 34.86597028)	3582
POINT (-92.56630005 38.63579078)	3573
POINT (-88.99771018 40.48501028)	3571
POINT (-86.14996019 39.76691045)	3570
POINT (-89.81637074 44.39319117)	3569
POINT (-92.44568007 31.31266064)	3563
POINT (-94.7942005 46.35564874)	3555
POINT (-99.36572062 41.64104099)	3546
POINT (-72.64984095 41.56266102)	3543
POINT (-71.52247031 41.70828019)	3534
POINT (-97.52107021 35.47203136)	3526
POINT (-100.3735306 44.35313005)	3526
POINT (-100.118421 47.47531978)	3514
POINT (-98.20078123 38.3477403)	3510
POINT (-114.36373 43.68263001)	3507
POINT (-93.81649056 42.46940091)	3501
POINT (-92.27449074 34.74865012)	3498
POINT (-108.1098304 43.23554134)	3494
POINT (-120.9999995 37.63864012)	3447
POINT (-106.1336109 38.84384076)	3390
POINT (-79.15925046 35.46622098)	3349
POINT (-109.4244206 47.06652897)	3348
POINT (-120.4700108 47.52227863)	3348
POINT (-75.57774117 39.00883067)	3346
POINT (-71.50036092 43.65595011)	3284
POINT (-72.51764079 43.62538124)	3278
POINT (-72.08269067 42.27687047)	3174
POINT (-66.590149 18.220833)	2797
POINT (144.793731 13.444304)	2703
POINT (-64.896335 18.335765)	503
Name: Geolocation, dtype: int64	
C01	71694
C03	46867
C02	24851

C06	19180
C07	18671
C05	16600
C04	16599

Name: ClassID, dtype: int64

TNC04	8300
TSC08	8300
TOC01	8300
TMC01	8300
TAC01	8300
TMC03	8300
TNC03	8300
TOC07	8299
TOC08	8299
TAC03	8299
TSC09	8268
TOC03	8233
TOC10	6917
TOC11	5511
TOC13	5447
TCC01	5088
TSC04	4808
TCC04	4700
TCC03	4696
TCC02	4696
TOC06	4173
TSC02	4173
TOC05	4172
TOC09	4171
TNC02	4127
TSC07	4127
TSC06	4127
TNC01	4124
TOC04	4108
TOC12	4064
TGC01	3848
TGC02	3797
TGC05	3682
TGC03	3681
TGC04	3663
TSC11	3280
TSC10	3271
TSC01	3271
TSC03	3242

Name: TopicID, dtype: int64

Q13	8300
Q18	8300
Q08	8300

Q03	8300
Q17	8300
Q27	8300
Q16	8300
Q32	8299
Q33	8299
Q21	8299
Q09	8268
Q35	8233
Q46	6917
Q43	5511
Q45	5447
Q30	5088
Q19	4808
Q42	4700
Q41	4696
Q31	4696
Q05	4173
Q15	4173
Q07	4172
Q34	4171
Q02	4127
Q22	4127
Q14	4127
Q01	4124
Q04	4108
Q44	4064
Q36	3848
Q37	3797
Q40	3682
Q38	3681
Q39	3663
Q11	3280
Q10	3271
Q12	3271
Q20	3242

Name: QuestionID, dtype: int64

59	4644
9004	4638
9001	4614
9002	4611
41	4565
36	4557
9003	4542
49	4222
39	3955
13	3951
24	3919

15	3907
47	3879
26	3796
51	3758
12	3753
23	3733
48	3699
32	3696
11	3684
54	3682
28	3677
42	3648
35	3635
1	3633
21	3623
2	3611
45	3592
34	3589
4	3582
29	3573
17	3571
18	3570
55	3569
22	3563
27	3555
31	3546
9	3543
44	3534
40	3526
46	3526
38	3514
20	3510
16	3507
19	3501
5	3498
56	3494
6	3447
8	3390
37	3349
53	3348
30	3348
10	3346
33	3284
50	3278
25	3174
72	2797
66	2703
78	503

```

Name: LocationID, dtype: int64
AGE      214462
Name: StratificationCategoryID1, dtype: int64
AGE_OVERALL    71919
5064           71528
65PLUS         71015
Name: StratificationID1, dtype: int64
RACE          134959
GENDER        51834
OVERALL       27669
Name: StratificationCategoryID2, dtype: int64
OVERALL       27669
WHT           27633
HIS           27525
BLK           26968
NAA           26571
ASN           26262
FEMALE        26091
MALE          25743
Name: StratificationID2, dtype: int64

```

0.0.3 5

5

```
[5]: data.describe()
```

```

[5]:
      count      YearStart      YearEnd      Data_Value      Data_Value_Alt  \
count      214462.000000      214462.000000      144629.000000      144629.000000
mean         2017.378477         2017.634000         37.341956         37.341956
std           1.779822           1.778926         25.183017         25.183017
min           2015.000000           2015.000000          0.000000          0.000000
25%           2016.000000           2016.000000         15.300000         15.300000
50%           2017.000000           2018.000000         32.500000         32.500000
75%           2019.000000           2019.000000         56.800000         56.800000
max           2020.000000           2020.000000        100.000000        100.000000

      count      Sample_Size      LocationID
count         0.0      214462.000000
mean          NaN          800.987821
std           NaN          2512.934094
min           NaN           1.000000
25%           NaN          18.000000
50%           NaN          33.000000
75%           NaN          49.000000
max           NaN          9004.000000

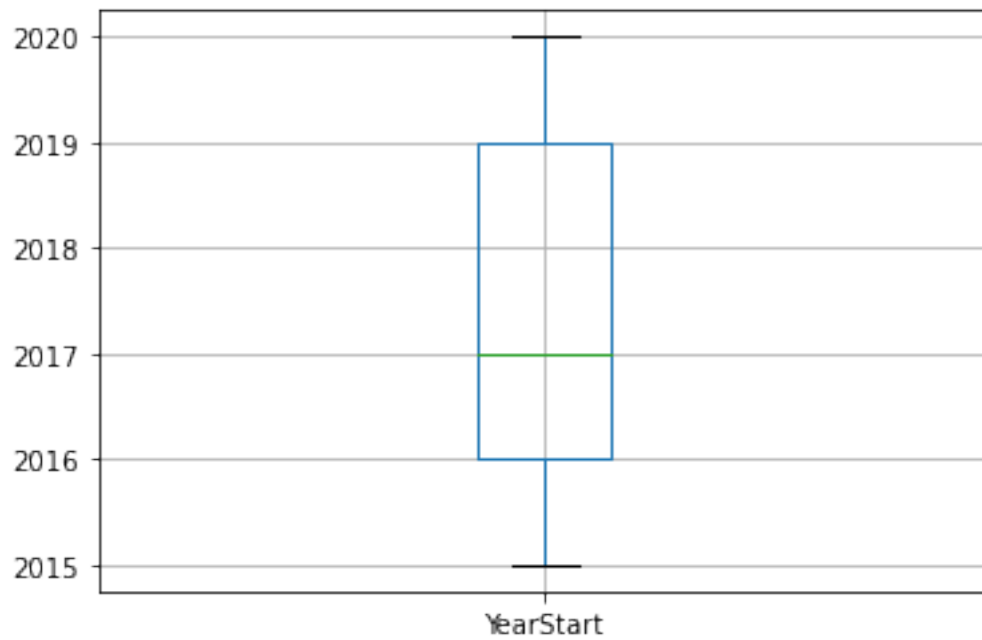
```

```
[6]: print(' ', ' ')
      for name, column in data.iteritems():
          print(name, column.isnull().sum())
```

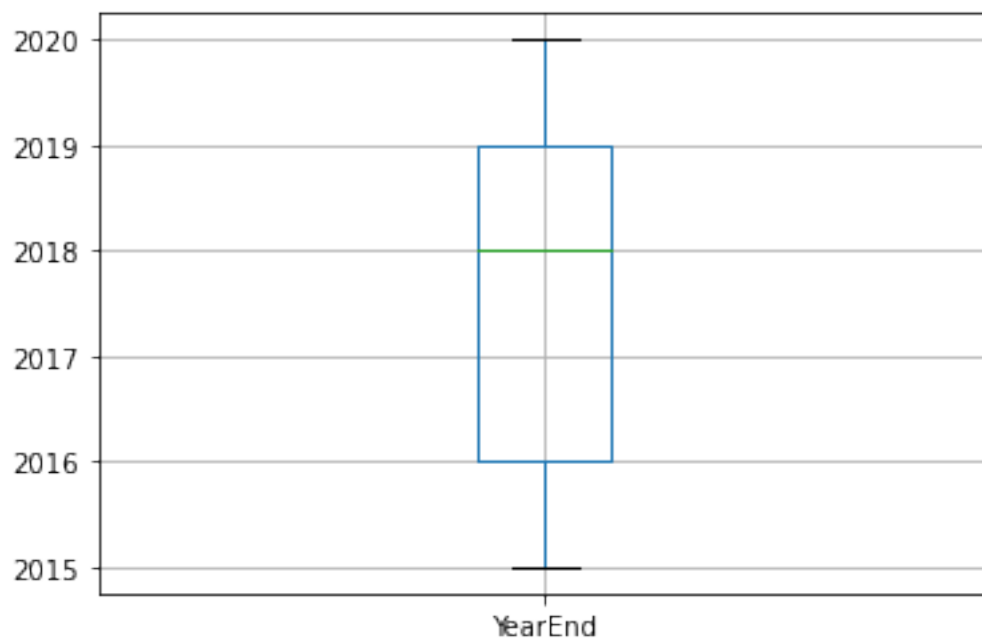
```
YearStart 0
YearEnd 0
LocationAbbr 0
LocationDesc 0
Datasource 0
Class 0
Topic 0
Question 0
Data_Value_Unit 0
DataValueTypeID 0
Data_Value_Type 0
Data_Value 69833
Data_Value_Alt 69833
Low_Confidence_Limit 69990
High_Confidence_Limit 69990
Sample_Size 214462
StratificationCategory1 0
Stratification1 0
StratificationCategory2 27669
Stratification2 27669
Geolocation 23049
ClassID 0
TopicID 0
QuestionID 0
LocationID 0
StratificationCategoryID1 0
StratificationID1 0
StratificationCategoryID2 0
StratificationID2 0
```

0.0.4

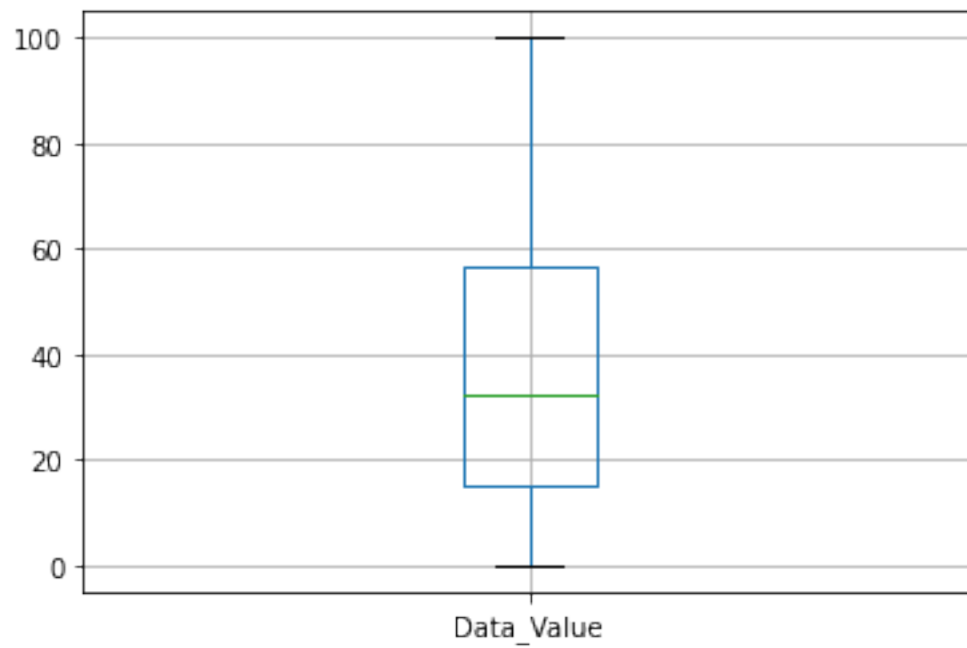
```
[7]: data.boxplot(column='YearStart')
      plt.show()
```

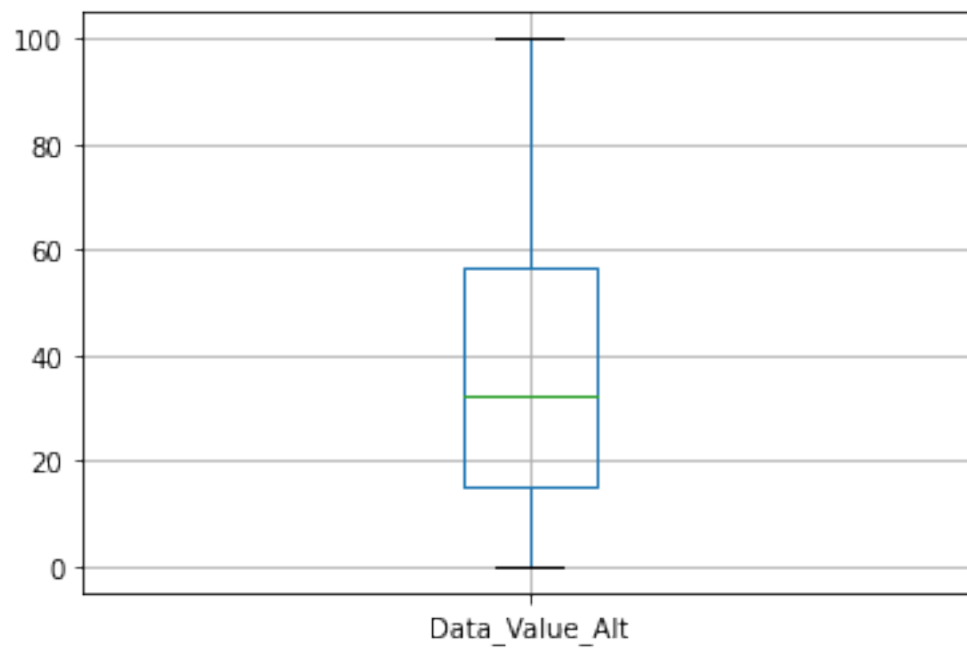
```
[8]: data.boxplot(column='YearEnd')  
plt.show()
```



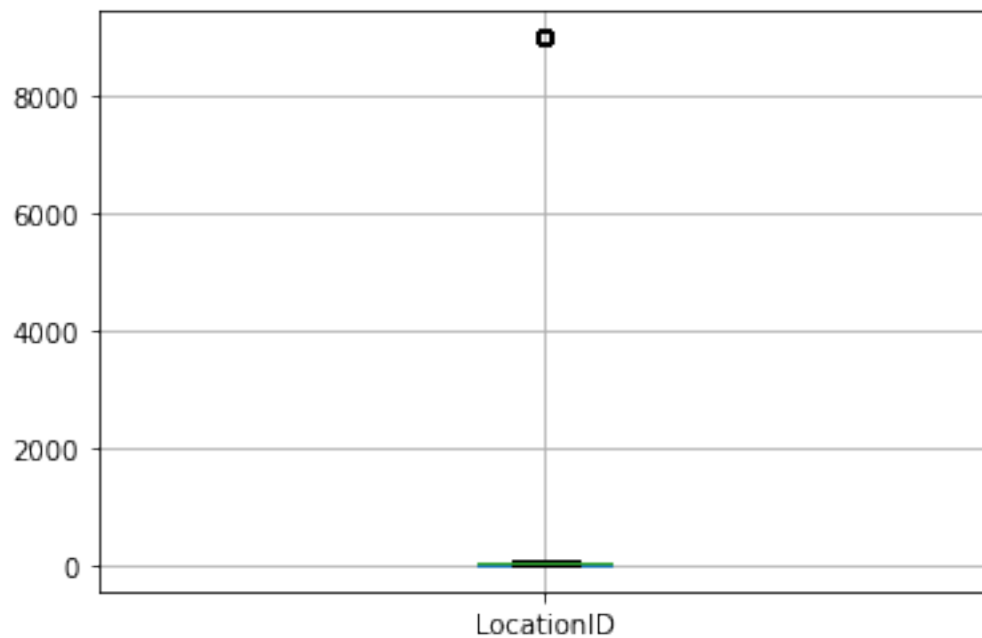
```
[9]: data.boxplot(column='Data_Value')  
plt.show()
```



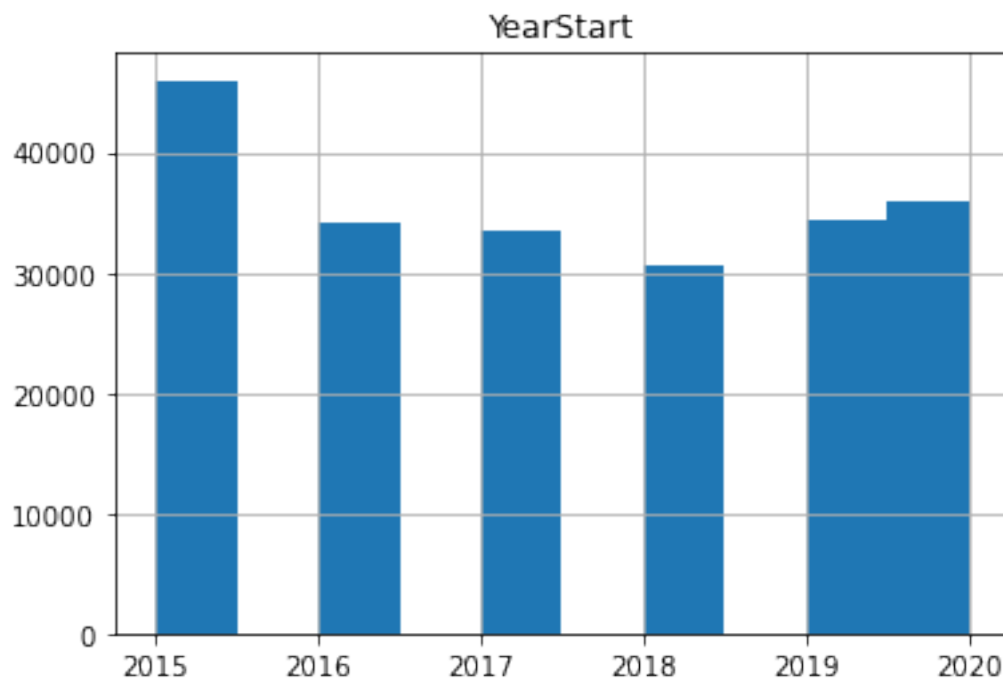
```
[10]: data.boxplot(column='Data_Value_Alt')  
plt.show()
```



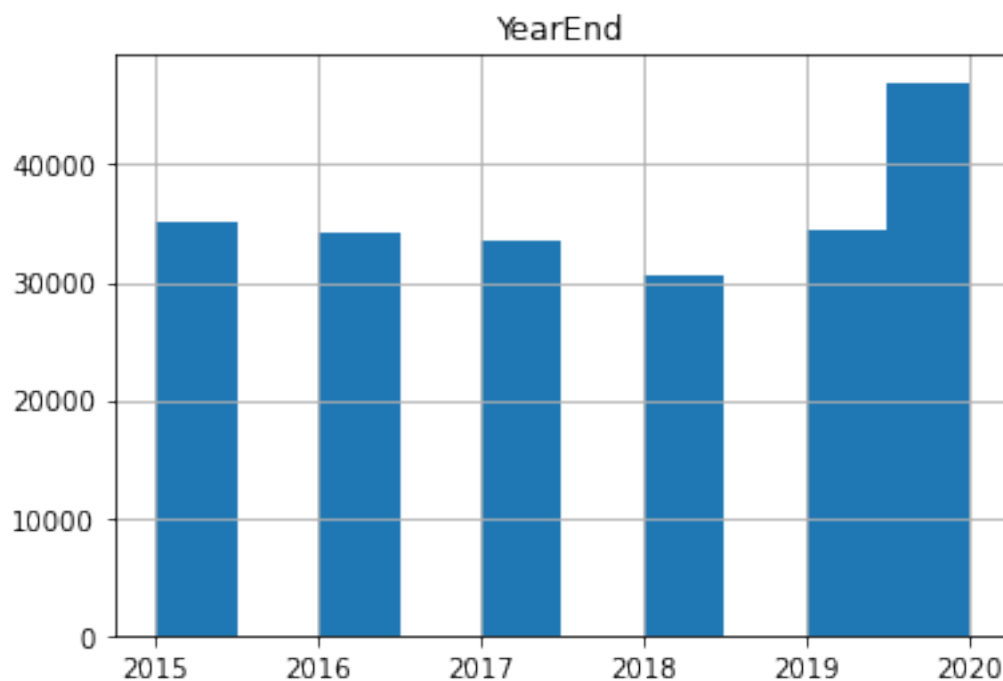
```
[11]: data.boxplot(column='LocationID')  
plt.show()
```



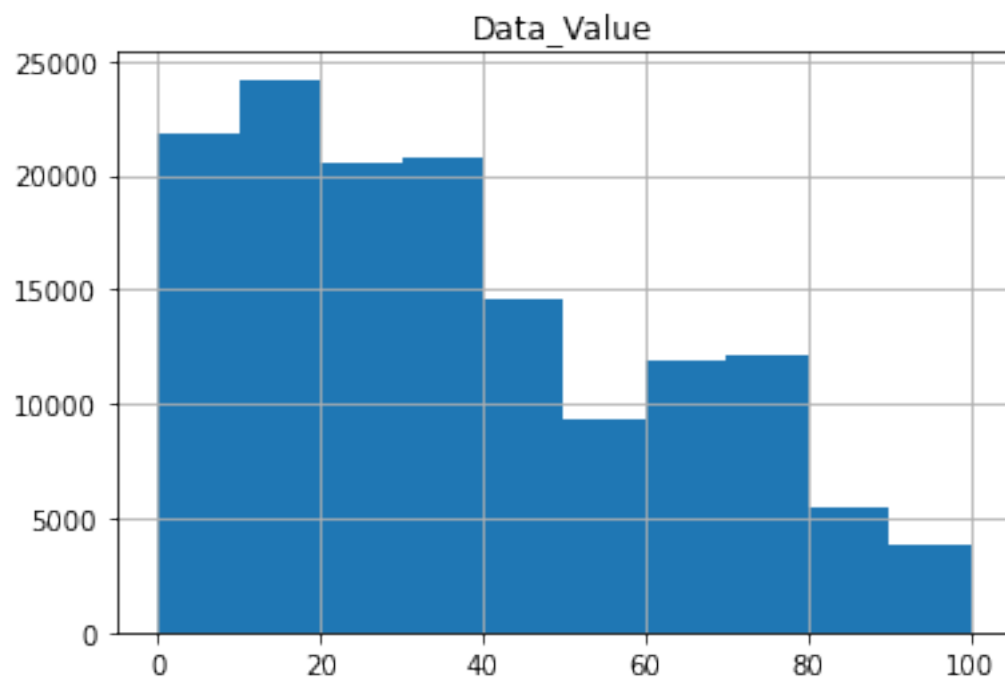
```
[12]: data.hist(column='YearStart')  
plt.show()
```



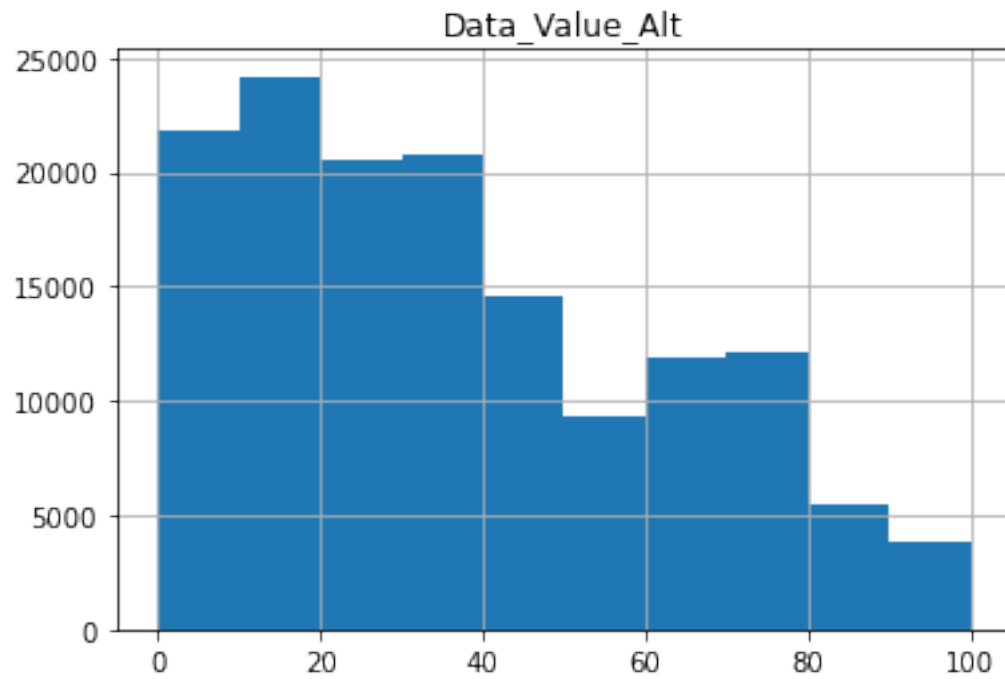
```
[13]: data.hist(column='YearEnd')  
plt.show()
```



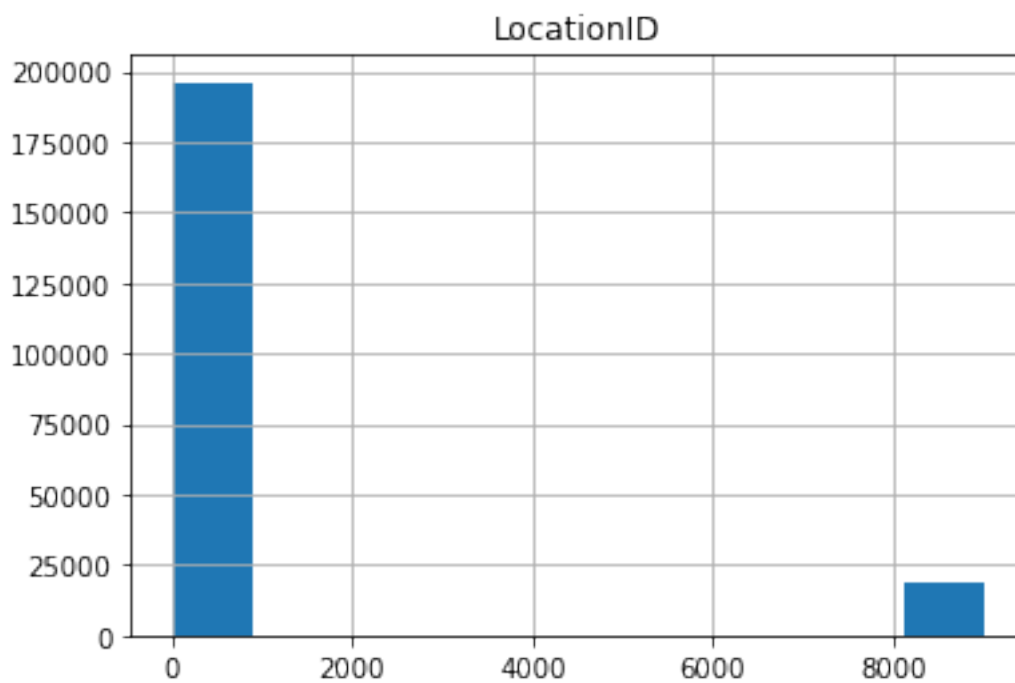
```
[14]: data.hist(column='Data_Value')  
plt.show()
```



```
[15]: data.hist(column='Data_Value_Alt')  
plt.show()
```



```
[16]: data.hist(column='LocationID')  
plt.show()
```



0.0.5

```
[17]: # Data_Value  
data['Data_Value'].isnull().sum()
```

```
[17]: 69833
```

```
[18]: data.shape
```

```
[18]: (214462, 29)
```

```
[19]: del data['Data_Value']
```

```
[20]: data.shape
```

```
[20]: (214462, 28)
```

```
[21]: data['Low_Confidence_Limit']
```

```
[21]: 0      23.9  
     1      21.2  
     2      32.5  
     3       4.0  
     4      12.1  
     ...  
    214457    NaN  
    214458    NaN  
    214459    NaN  
    214460     8.6  
    214461    NaN  
     Name: Low_Confidence_Limit, Length: 214462, dtype: object
```

```
[22]: mode = data['Low_Confidence_Limit'].value_counts().index[0]
```

```
[23]: data['Low_Confidence_Limit'].fillna(mode, inplace=True)
```

```
[24]: data['Low_Confidence_Limit']
```

```
[24]: 0      23.9  
     1      21.2  
     2      32.5  
     3       4.0  
     4      12.1  
     ...  
    214457     5.4  
    214458     5.4  
    214459     5.4  
    214460     8.6
```

```
214461      5.4
Name: Low_Confidence_Limit, Length: 214462, dtype: object
```

```
[25]: data['Data_Value_Alt']
```

```
[25]: 0      26.3
      1      24.0
      2      35.7
      3       4.8
      4      13.7
      ...
      214457   NaN
      214458   NaN
      214459   NaN
      214460   10.6
      214461   NaN
      Name: Data_Value_Alt, Length: 214462, dtype: float64
```

```
[26]: data['Data_Value_Alt'].fillna(data['Data_Value_Alt'].mean(), inplace=True)
```

```
[27]: data['Data_Value_Alt']
```

```
[27]: 0      26.300000
      1      24.000000
      2      35.700000
      3       4.800000
      4      13.700000
      ...
      214457   37.341956
      214458   37.341956
      214459   37.341956
      214460   10.600000
      214461   37.341956
      Name: Data_Value_Alt, Length: 214462, dtype: float64
```

```
[28]: data['Geolocation']
```

```
[28]: 0      POINT (-157.8577494 21.30485044)
      1      POINT (-114.36373 43.68263001)
      2      POINT (-114.36373 43.68263001)
      3      POINT (-114.36373 43.68263001)
      4      POINT (-86.14996019 39.76691045)
      ...
      214457   POINT (-108.1098304 43.23554134)
      214458   POINT (-108.1098304 43.23554134)
      214459   POINT (-108.1098304 43.23554134)
```



```
214460    POINT (-108.1098304 43.23554134)
214461    POINT (-108.1098304 43.23554134)
Name: Geolocation, Length: 214462, dtype: object
```

```
[29]: data['Geolocation'].isnull().sum()
```

```
[29]: 23049
```

```
[30]: #
      data['Geolocation'].fillna(method='pad', inplace=True)
```

```
[31]: data['Geolocation'].isnull().sum()
```

```
[31]: 0
```

```
[ ]:
```