Alzheimer Disease and Healthy Aging Data In US

March 18, 2024

[1]:
```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list␣
 ↪all files under the input directory

import os
for dirname, _, filenames in os.walk('input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
input\Alzheimer Disease and Healthy Aging Data In US.csv
```

[23]:
```
file_path = 'input/Alzheimer Disease and Healthy Aging Data In US.csv'
data = pd.read_csv(file_path)
data.columns
```

[23]:
```
Index(['YearStart', 'YearEnd', 'LocationAbbr', 'LocationDesc', 'Datasource',
       'Class', 'Topic', 'Question', 'Data_Value_Unit', 'DataValueTypeID',
       'Data_Value_Type', 'Data_Value', 'Data_Value_Alt',
       'Low_Confidence_Limit', 'High_Confidence_Limit', 'Sample_Size',
       'StratificationCategory1', 'Stratification1', 'StratificationCategory2',
       'Stratification2', 'Geolocation', 'ClassID', 'TopicID', 'QuestionID',
       'LocationID', 'StratificationCategoryID1', 'StratificationID1',
       'StratificationCategoryID2', 'StratificationID2'],
      dtype='object')
```

3. 3.1

   5

[14]:
```
Nominal_Attribute = ['LocationAbbr', 'LocationDesc', 'Class', 'Topic',
                         'Question', 'DataValueTypeID', 'Stratification1',
                         'StratificationCategory2', 'ClassID', 'QuestionID',
                         'StratificationCategoryID2', 'StratificationID2']
print('  :')
for attribute in Nominal_Attribute:
```

```
    print('-------------------------------------------------')
    print(attribute + ":")
    print(data[attribute].value_counts())
```

```
  :
-------------------------------------------------
LocationAbbr:
LocationAbbr
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: count, dtype: int64
------------------------------------------------
LocationDesc:
LocationDesc
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: count, dtype: int64
------------------------------------------------
Class:
Class
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: count, dtype: int64
------------------------------------------------
Topic:
```

```
Topic
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: count, dtype: int64
--------------------------------------------------
Question:
Question
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: count, dtype: int64
------------------------------------------------
DataValueTypeID:
```

```
DataValueTypeID
PRCTG    197929
MEAN      16533
Name: count, dtype: int64
------------------------------------------------
Stratification1:
Stratification1
Overall              71919
50-64 years          71528
65 years or older    71015
Name: count, dtype: int64
------------------------------------------------
StratificationCategory2:
StratificationCategory2
Race/Ethnicity    134959
Gender             51834
Name: count, dtype: int64
------------------------------------------------
ClassID:
ClassID
C01    71694
C03    46867
C02    24851
C06    19180
C07    18671
C05    16600
C04    16599
Name: count, dtype: int64
------------------------------------------------
QuestionID:
QuestionID
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
```

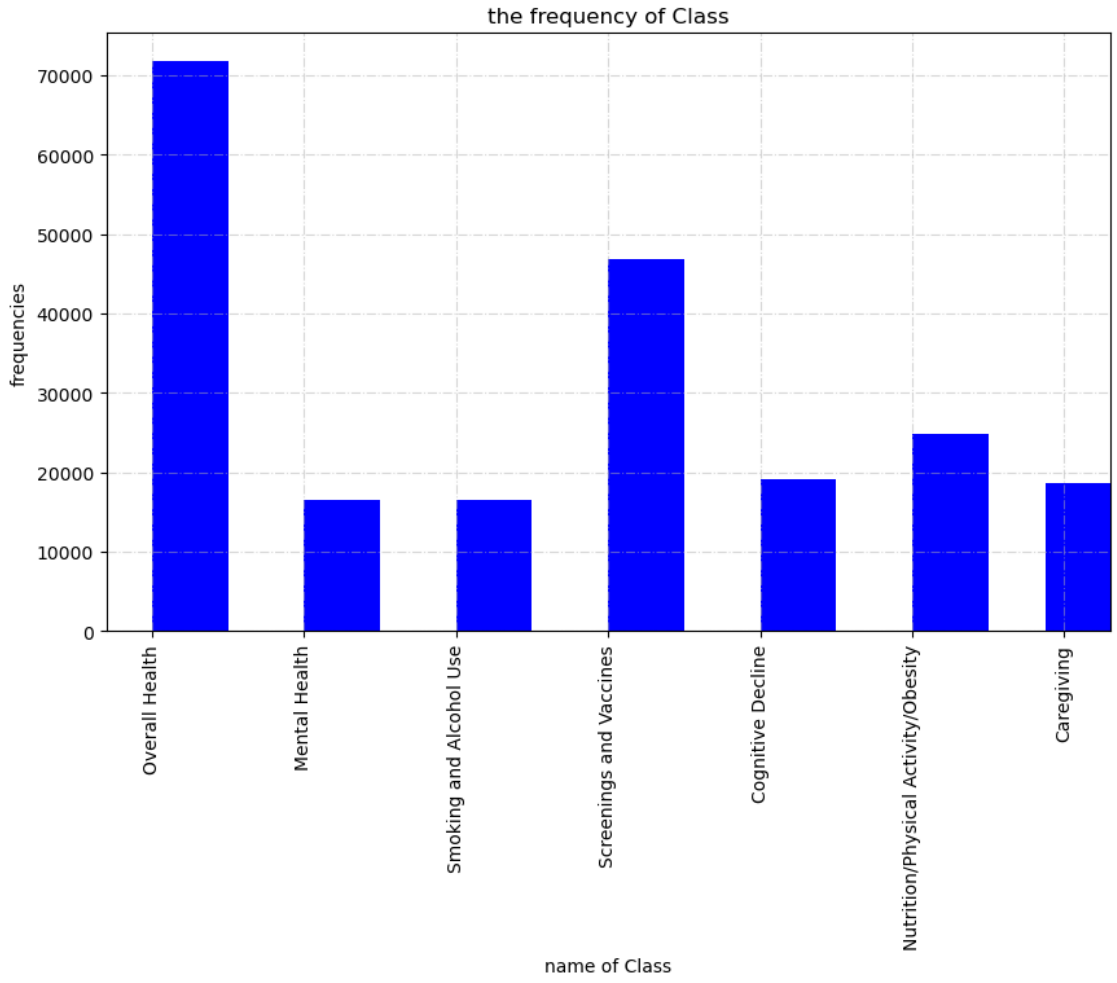
```
Q42    4700
Q41    4696
Q31    4696
Q05    4173
Q15    4173
Q07    4172
Q34    4171
Q02    4127
Q22    4127
Q14    4127
Q01    4124
Q04    4108
Q44    4064
Q36    3848
Q37    3797
Q40    3682
Q38    3681
Q39    3663
Q11    3280
Q10    3271
Q12    3271
Q20    3242
Name: count, dtype: int64
--------------------------------------------------
StratificationCategoryID2:
StratificationCategoryID2
RACE       134959
GENDER      51834
OVERALL     27669
Name: count, dtype: int64
--------------------------------------------------
StratificationID2:
StratificationID2
OVERALL    27669
WHT        27633
HIS        27525
BLK        26968
NAA        26571
ASN        26262
FEMALE     26091
MALE       25743
Name: count, dtype: int64
```

[16]:
```python
def five_number(data):
    Min_num = min(data)
    Max_num = max(data)
    Q1_num = np.percentile(data, 25)
```

```
    Median_num = np.median(data)
    Q3_num = np.percentile(data, 75)
    return Min_num, Max_num, Q1_num, Median_num, Q3_num

print('  :')

for attribute in Number_Attribute:
    print('-------------------------------------------------')
    print(attribute + ":")
    print('    ')
    print(data[attribute].isnull().sum())
    print('  :')
    print(five_number(data.dropna(subset=[attribute])[attribute]))
```

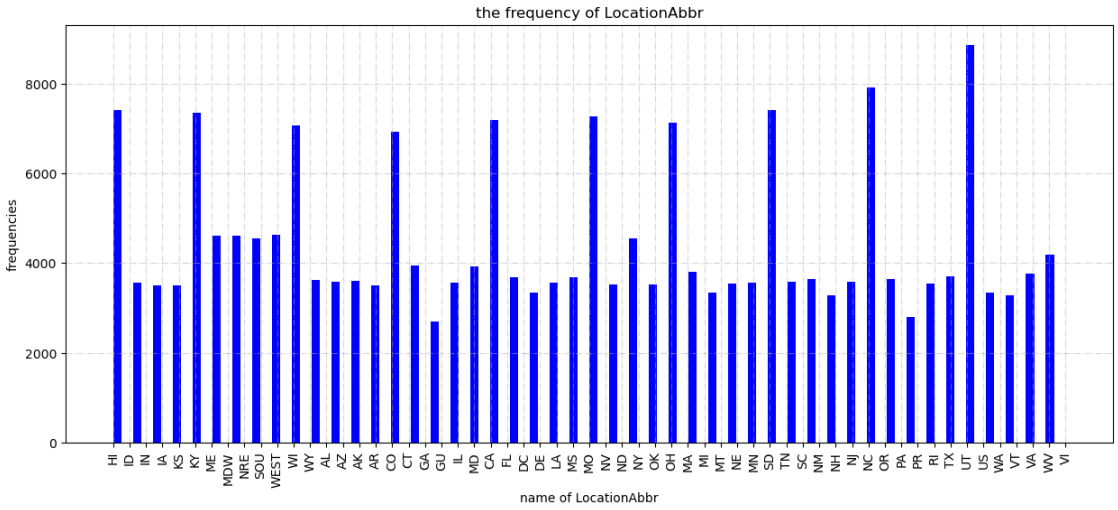
```
  :
-------------------------------------------------
YearStart:

0
  :
(2015, 2020, 2016.0, 2017.0, 2019.0)
-------------------------------------------------
YearEnd:

0
  :
(2015, 2020, 2016.0, 2018.0, 2019.0)
-------------------------------------------------
Data_Value:

69833
  :
(0.0, 100.0, 15.3, 32.5, 56.8)
-------------------------------------------------
Data_Value_Alt:

69833
  :
(0.0, 100.0, 15.3, 32.5, 56.8)
-------------------------------------------------
LocationID:

0
  :
(1, 9004, 18.0, 33.0, 49.0)
```

[17]:
```
# histogram
# Class
plt.figure(num=1, figsize=(10,6))
sub_data = data.dropna(subset=['Class'])['Class']
plt.xticks(rotation=90)
plt.hist(sub_data, bins=48, width=0.5, color='blue')
plt.grid(alpha=0.5, linestyle='-.')
plt.xlabel('name of Class')
plt.ylabel('frequencies')
plt.title(r'the frequency of Class')
plt.show()
```

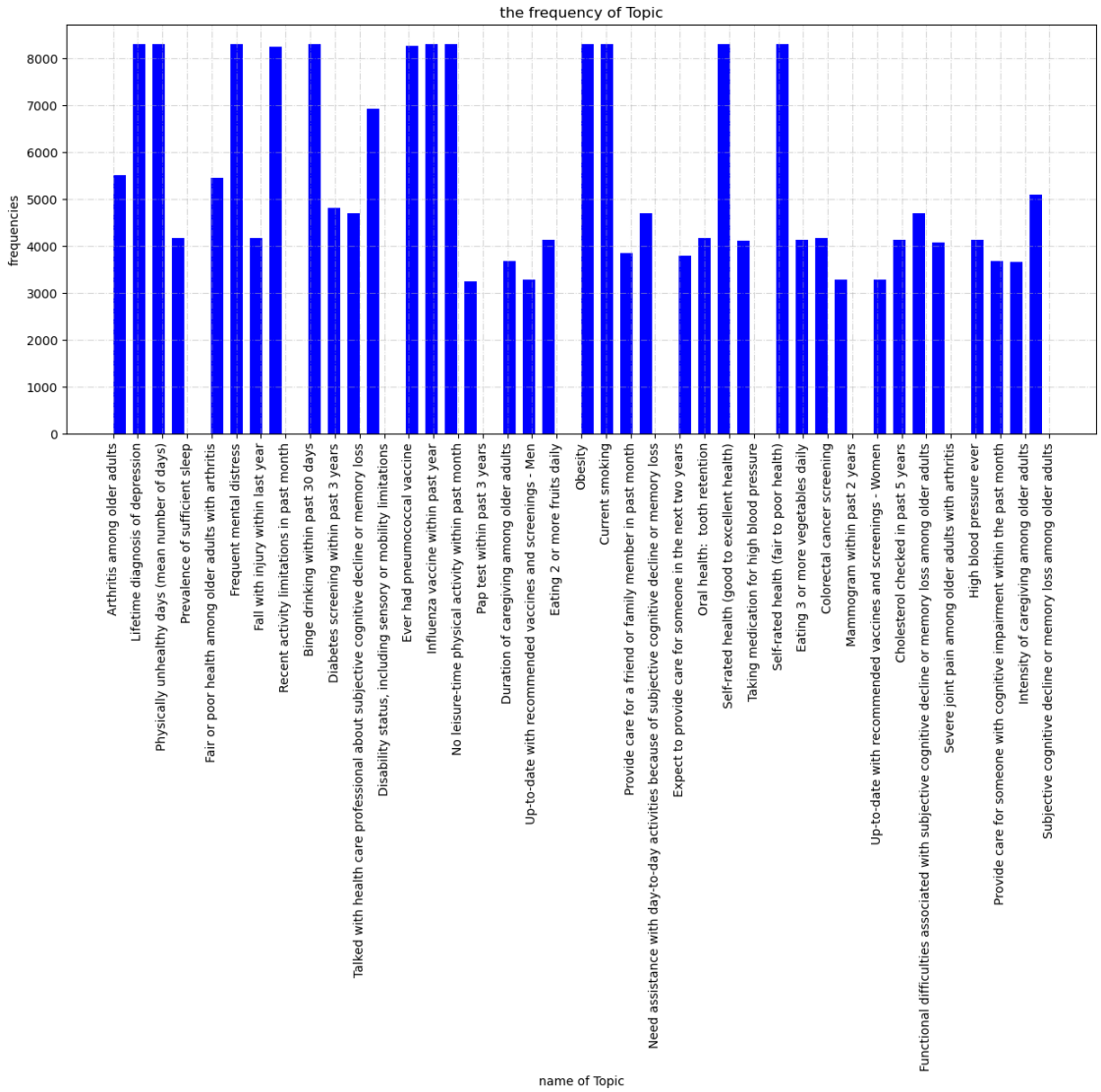


[18]:
```
# LocationAbbr
plt.figure(num=2, figsize=(15,6))
sub_data = data.dropna(subset=['LocationAbbr'])['LocationAbbr']
plt.xticks(rotation=90)
```

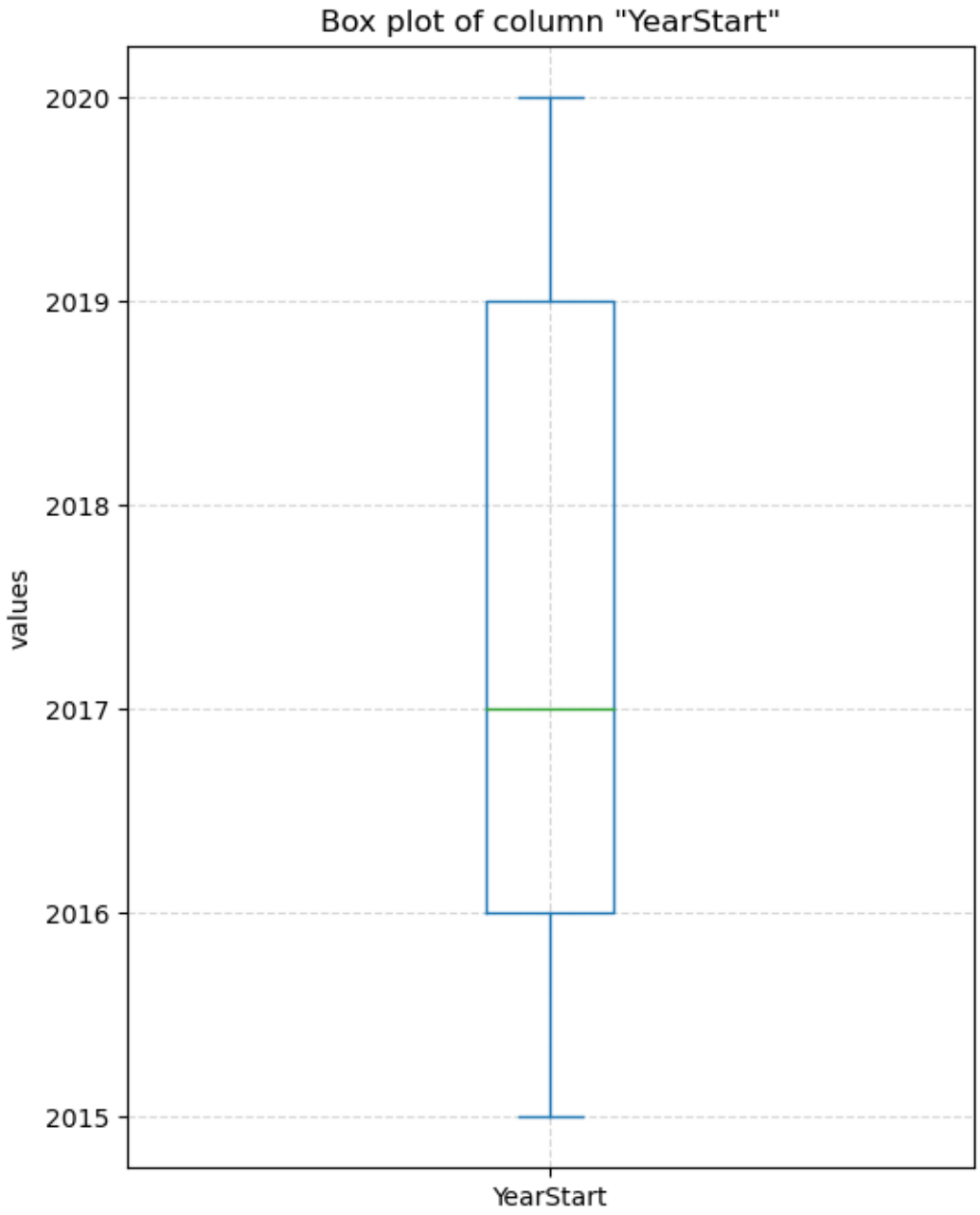
```
plt.hist(sub_data, bins=48, width=0.5, color='blue')
plt.grid(alpha=0.5, linestyle='-.')
plt.xlabel('name of LocationAbbr')
plt.ylabel('frequencies')
plt.title(r'the frequency of LocationAbbr')
plt.show()
```



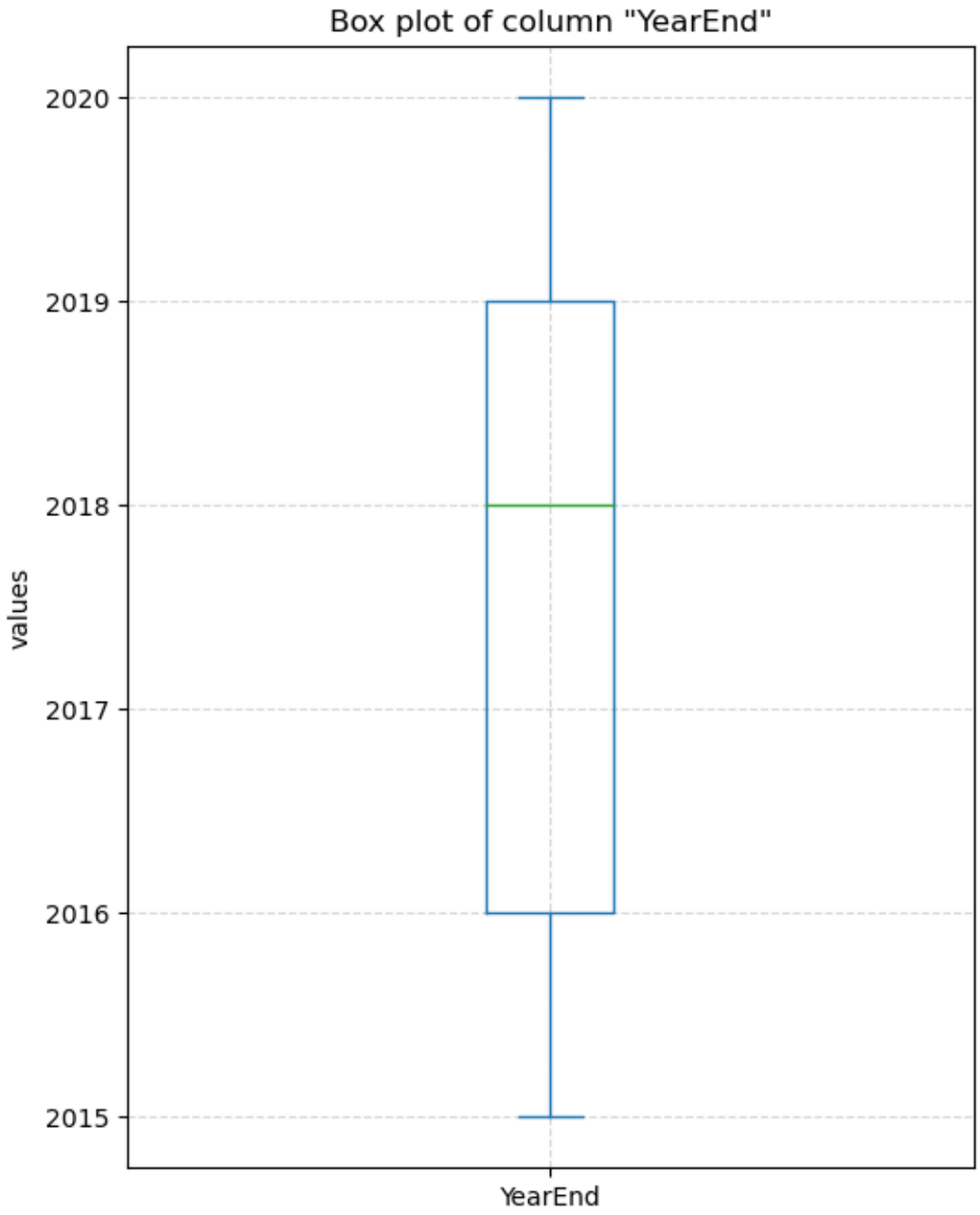
[19]:
```
# Topic
plt.figure(num=3, figsize=(15,6))
sub_data = data.dropna(subset=['Topic'])['Topic']
plt.xticks(rotation=90)
plt.hist(sub_data, bins=48, width=0.5, color='blue')
plt.grid(alpha=0.5, linestyle='-.')
plt.xlabel('name of Topic')
plt.ylabel('frequencies')
plt.title(r'the frequency of Topic')
plt.show()
```



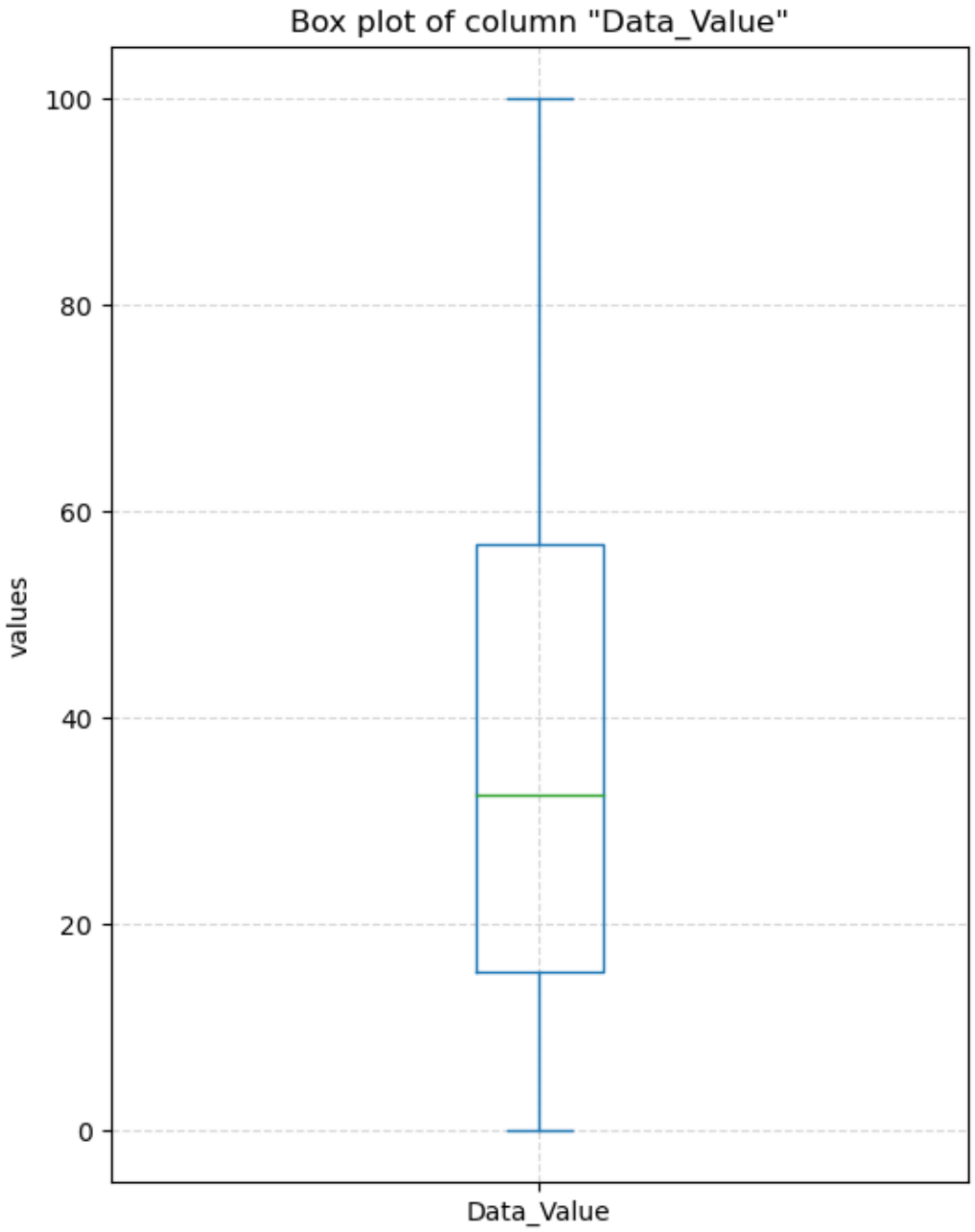
[20]:
```
# box-plot picture
# YearStart
plt.figure(num=4, figsize=(6, 8))
sub_data = data.dropna(subset=['YearStart'])['YearStart']
sub_data.plot.box(title='box plot')
plt.grid(linestyle="--", alpha=0.5)
plt.ylabel('values')
plt.title(r'Box plot of column "YearStart"')
plt.show()
```



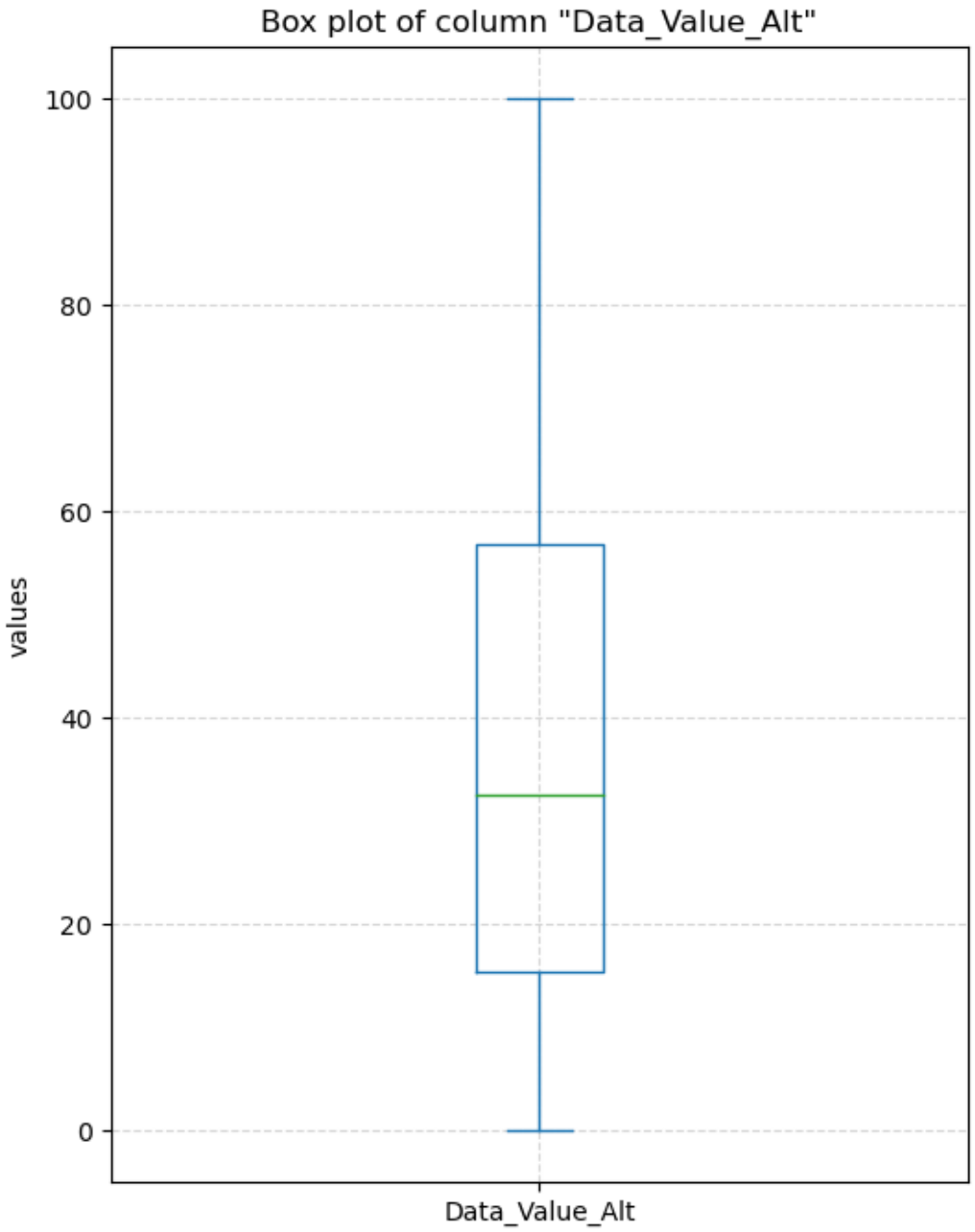
[21]:
```
# YearEnd
plt.figure(num=5, figsize=(6, 8))
sub_data = data.dropna(subset=['YearEnd'])['YearEnd']
sub_data.plot.box(title='box plot')
plt.grid(linestyle="--", alpha=0.5)
plt.ylabel('values')
plt.title(r'Box plot of column "YearEnd"')
plt.show()
```



[22]:
```
# Data_Value
plt.figure(num=6, figsize=(6, 8))
sub_data = data.dropna(subset=['Data_Value'])['Data_Value']
sub_data.plot.box(title='box plot')
plt.grid(linestyle="--", alpha=0.5)
plt.ylabel('values')
plt.title(r'Box plot of column "Data_Value"')
plt.show()
```



[24]:
```
# Data_Value_Alt
plt.figure(num=7, figsize=(6, 8))
sub_data = data.dropna(subset=['Data_Value_Alt'])['Data_Value_Alt']
sub_data.plot.box(title='box plot')
plt.grid(linestyle="--", alpha=0.5)
plt.ylabel('values')
plt.title(r'Box plot of column "Data_Value_Alt"')
plt.show()
```



3.2 :

[51]:
```
#
print('    ')
data[data[['Data_Value']].isnull().T.any()].iloc[:10,10:]
```

[51]:

	Data_Value_Type	Data_Value	Data_Value_Alt	Low_Confidence_Limit	High_Confidence_Limit	Sample_Size	StratificationCategory1	Stratification1	StratificationCategory2	Stratification2	Geolocation	ClassID	TopicID	QuestionID	LocationID
32	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	65 years or older	Race/Ethnicity	Native Am/Alaskan Native	POINT (-111.7638113 34.86597028)	C01	TOC13	Q45	4
33	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	65 years or older	Race/Ethnicity	Native Am/Alaskan Native	POINT (-86.63186076 32.84057112)	C06	TCC04	Q42	1
42	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	Overall	Race/Ethnicity	Hispanic	POINT (-86.63186076 32.84057112)	C03	TSC03	Q20	1
47	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	50-64 years	Race/Ethnicity	Hispanic	POINT (-86.63186076 32.84057112)	C02	TNC01	Q01	1
48	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	Overall	Race/Ethnicity	Hispanic	POINT (-86.63186076 32.84057112)	C03	TSC09	Q09	1
49	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	65 years or older	Race/Ethnicity	Hispanic	POINT (-147.722059 64.84507996)	C02	TNC04	Q13	2
63	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	50-64 years	Race/Ethnicity	Asian/Pacific Islander	POINT (-111.7638113 34.86597028)	C02	TNC01	Q01	4
64	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	Overall	Race/Ethnicity	Native Am/Alaskan Native	POINT (-86.63186076 32.84057112)	C01	TOC04	Q04	1
66	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	Overall	Race/Ethnicity	Asian/Pacific Islander	POINT (-86.63186076 32.84057112)	C01	TOC07	Q32	1
67	Percentage	NaN	NaN	NaN	NaN	NaN	Age Group	50-64 years	Race/Ethnicity	Asian/Pacific Islander	POINT (-111.7638113 34.86597028)	C03	TSC09	Q09	4

```
   StratificationCategoryID1 StratificationID1 StratificationCategoryID2  \
32                       AGE            65PLUS                       RACE
33                       AGE            65PLUS                       RACE
42                       AGE       AGE_OVERALL                       RACE
47                       AGE              5064                       RACE
48                       AGE       AGE_OVERALL                       RACE
49                       AGE            65PLUS                       RACE
63                       AGE              5064                       RACE
64                       AGE       AGE_OVERALL                       RACE
66                       AGE       AGE_OVERALL                       RACE
67                       AGE              5064                       RACE

   StratificationID2
32               NAA
33               NAA
42               HIS
47               HIS
48               HIS
49               HIS
63               ASN
64               NAA
66               ASN
67               ASN
```

[56]:
```
#
data1 = data.copy(deep=True)
data1 = data1.dropna()
print('    ')
data1[data[['Data_Value']].isnull().T.any()].iloc[:10,10:]
```

[56]:
```
Empty DataFrame
Columns: [Data_Value_Type, Data_Value, Data_Value_Alt, Low_Confidence_Limit,
High_Confidence_Limit, Sample_Size, StratificationCategory1, Stratification1,
StratificationCategory2, Stratification2, Geolocation, ClassID, TopicID,
QuestionID, LocationID, StratificationCategoryID1, StratificationID1,
StratificationCategoryID2, StratificationID2]
Index: []
```

[53]:
```
#
data2 = data.copy(deep=True)
data2['Data_Value'] = data2['Data_Value'].fillna(np.median(data2.
 ↪dropna(subset=['Data_Value'])['Data_Value']),inplace=False)
print('      ')
data2[data[['Data_Value']].isnull().T.any()].iloc[:10,10:]
```

[53]:

	Data_Value_Type	Data_Value	Data_Value_Alt	Low_Confidence_Limit
32	Percentage	32.5	NaN	NaN
33	Percentage	32.5	NaN	NaN
42	Percentage	32.5	NaN	NaN
47	Percentage	32.5	NaN	NaN
48	Percentage	32.5	NaN	NaN
49	Percentage	32.5	NaN	NaN
63	Percentage	32.5	NaN	NaN
64	Percentage	32.5	NaN	NaN
66	Percentage	32.5	NaN	NaN
67	Percentage	32.5	NaN	NaN

	High_Confidence_Limit	Sample_Size	StratificationCategory1
32	NaN	NaN	Age Group
33	NaN	NaN	Age Group
42	NaN	NaN	Age Group
47	NaN	NaN	Age Group
48	NaN	NaN	Age Group
49	NaN	NaN	Age Group
63	NaN	NaN	Age Group
64	NaN	NaN	Age Group
66	NaN	NaN	Age Group
67	NaN	NaN	Age Group

	Stratification1	StratificationCategory2	Stratification2
32	65 years or older	Race/Ethnicity	Native Am/Alaskan Native
33	65 years or older	Race/Ethnicity	Native Am/Alaskan Native
42	Overall	Race/Ethnicity	Hispanic
47	50-64 years	Race/Ethnicity	Hispanic
48	Overall	Race/Ethnicity	Hispanic
49	65 years or older	Race/Ethnicity	Hispanic
63	50-64 years	Race/Ethnicity	Asian/Pacific Islander
64	Overall	Race/Ethnicity	Native Am/Alaskan Native
66	Overall	Race/Ethnicity	Asian/Pacific Islander
67	50-64 years	Race/Ethnicity	Asian/Pacific Islander

	Geolocation	ClassID	TopicID	QuestionID	LocationID
32	POINT (-111.7638113 34.86597028)	C01	TOC13	Q45	4
33	POINT (-86.63186076 32.84057112)	C06	TCC04	Q42	1
42	POINT (-86.63186076 32.84057112)	C03	TSC03	Q20	1
47	POINT (-86.63186076 32.84057112)	C02	TNC01	Q01	1
48	POINT (-86.63186076 32.84057112)	C03	TSC09	Q09	1
49	POINT (-147.722059 64.84507996)	C02	TNC04	Q13	2
63	POINT (-111.7638113 34.86597028)	C02	TNC01	Q01	4
64	POINT (-86.63186076 32.84057112)	C01	TOC04	Q04	1

```
66  POINT (-86.63186076 32.84057112)     C01   TOC07        Q32           1
67  POINT (-111.7638113 34.86597028)     C03   TSC09        Q09           4

   StratificationCategoryID1 StratificationID1 StratificationCategoryID2  \
32                       AGE            65PLUS                       RACE
33                       AGE            65PLUS                       RACE
42                       AGE       AGE_OVERALL                       RACE
47                       AGE              5064                       RACE
48                       AGE       AGE_OVERALL                       RACE
49                       AGE            65PLUS                       RACE
63                       AGE              5064                       RACE
64                       AGE       AGE_OVERALL                       RACE
66                       AGE       AGE_OVERALL                       RACE
67                       AGE              5064                       RACE

   StratificationID2
32               NAA
33               NAA
42               HIS
47               HIS
48               HIS
49               HIS
63               ASN
64               NAA
66               ASN
67               ASN
```

[57]:
```
#
data3 = data.copy(deep=True)
data3['Data_Value'] = data3['Data_Value'].interpolate()
print('      ')
data3[data[['Data_Value']].isnull().T.any()].iloc[:10, 10:]
```

[57]:
```
   Data_Value_Type  Data_Value  Data_Value_Alt Low_Confidence_Limit  \
32      Percentage   60.333333             NaN                  NaN
33      Percentage   52.166667             NaN                  NaN
42      Percentage   36.300000             NaN                  NaN
47      Percentage   36.475000             NaN                  NaN
48      Percentage   31.450000             NaN                  NaN
49      Percentage   26.425000             NaN                  NaN
63      Percentage   26.366667             NaN                  NaN
64      Percentage   46.033333             NaN                  NaN
66      Percentage   61.666667             NaN                  NaN
67      Percentage   57.633333             NaN                  NaN

   High_Confidence_Limit  Sample_Size StratificationCategory1  \
```

```
32                  NaN          NaN               Age Group
33                  NaN          NaN               Age Group
42                  NaN          NaN               Age Group
47                  NaN          NaN               Age Group
48                  NaN          NaN               Age Group
49                  NaN          NaN               Age Group
63                  NaN          NaN               Age Group
64                  NaN          NaN               Age Group
66                  NaN          NaN               Age Group
67                  NaN          NaN               Age Group

      Stratification1 StratificationCategory2          Stratification2  \
32  65 years or older          Race/Ethnicity  Native Am/Alaskan Native
33  65 years or older          Race/Ethnicity  Native Am/Alaskan Native
42            Overall          Race/Ethnicity                  Hispanic
47         50-64 years          Race/Ethnicity                  Hispanic
48            Overall          Race/Ethnicity                  Hispanic
49  65 years or older          Race/Ethnicity                  Hispanic
63         50-64 years          Race/Ethnicity    Asian/Pacific Islander
64            Overall          Race/Ethnicity  Native Am/Alaskan Native
66            Overall          Race/Ethnicity    Asian/Pacific Islander
67         50-64 years          Race/Ethnicity    Asian/Pacific Islander

                        Geolocation ClassID TopicID QuestionID  LocationID  \
32  POINT (-111.7638113 34.86597028)     C01   TOC13        Q45           4
33  POINT (-86.63186076 32.84057112)     C06   TCC04        Q42           1
42  POINT (-86.63186076 32.84057112)     C03   TSC03        Q20           1
47  POINT (-86.63186076 32.84057112)     C02   TNC01        Q01           1
48  POINT (-86.63186076 32.84057112)     C03   TSC09        Q09           1
49   POINT (-147.722059 64.84507996)     C02   TNC04        Q13           2
63  POINT (-111.7638113 34.86597028)     C02   TNC01        Q01           4
64  POINT (-86.63186076 32.84057112)     C01   TOC04        Q04           1
66  POINT (-86.63186076 32.84057112)     C01   TOC07        Q32           1
67  POINT (-111.7638113 34.86597028)     C03   TSC09        Q09           4

   StratificationCategoryID1 StratificationID1 StratificationCategoryID2  \
32                       AGE            65PLUS                       RACE
33                       AGE            65PLUS                       RACE
42                       AGE       AGE_OVERALL                       RACE
47                       AGE              5064                       RACE
48                       AGE       AGE_OVERALL                       RACE
49                       AGE            65PLUS                       RACE
63                       AGE              5064                       RACE
64                       AGE       AGE_OVERALL                       RACE
66                       AGE       AGE_OVERALL                       RACE
67                       AGE              5064                       RACE
```

```
   StratificationID2
32               NAA
33               NAA
42               HIS
47               HIS
48               HIS
49               HIS
63               ASN
64               NAA
66               ASN
67               ASN
```

[58]:
```
#
data4 = data.copy(deep=True)
data4['Data_Value'] = data4['Data_Value'].fillna(np.
 ↪mean(data4['Data_Value']),inplace=False)
print('      : ')
data4[data[['Data_Value']].isnull().T.any()].iloc[:10,10:]
```

```
     :
```

[58]:
```
   Data_Value_Type  Data_Value  Data_Value_Alt Low_Confidence_Limit  \
32      Percentage   37.341956             NaN                  NaN
33      Percentage   37.341956             NaN                  NaN
42      Percentage   37.341956             NaN                  NaN
47      Percentage   37.341956             NaN                  NaN
48      Percentage   37.341956             NaN                  NaN
49      Percentage   37.341956             NaN                  NaN
63      Percentage   37.341956             NaN                  NaN
64      Percentage   37.341956             NaN                  NaN
66      Percentage   37.341956             NaN                  NaN
67      Percentage   37.341956             NaN                  NaN

   High_Confidence_Limit  Sample_Size StratificationCategory1  \
32                   NaN          NaN               Age Group
33                   NaN          NaN               Age Group
42                   NaN          NaN               Age Group
47                   NaN          NaN               Age Group
48                   NaN          NaN               Age Group
49                   NaN          NaN               Age Group
63                   NaN          NaN               Age Group
64                   NaN          NaN               Age Group
66                   NaN          NaN               Age Group
67                   NaN          NaN               Age Group

      Stratification1 StratificationCategory2           Stratification2  \
32  65 years or older          Race/Ethnicity  Native Am/Alaskan Native
33  65 years or older          Race/Ethnicity  Native Am/Alaskan Native
```

```
42            Overall          Race/Ethnicity                   Hispanic
47         50-64 years         Race/Ethnicity                   Hispanic
48            Overall          Race/Ethnicity                   Hispanic
49   65 years or older         Race/Ethnicity                   Hispanic
63         50-64 years         Race/Ethnicity    Asian/Pacific Islander
64            Overall          Race/Ethnicity  Native Am/Alaskan Native
66            Overall          Race/Ethnicity    Asian/Pacific Islander
67         50-64 years         Race/Ethnicity    Asian/Pacific Islander

                        Geolocation ClassID TopicID QuestionID  LocationID  \
32  POINT (-111.7638113 34.86597028)     C01   TOC13        Q45           4
33  POINT (-86.63186076 32.84057112)     C06   TCC04        Q42           1
42  POINT (-86.63186076 32.84057112)     C03   TSC03        Q20           1
47  POINT (-86.63186076 32.84057112)     C02   TNC01        Q01           1
48  POINT (-86.63186076 32.84057112)     C03   TSC09        Q09           1
49   POINT (-147.722059 64.84507996)     C02   TNC04        Q13           2
63  POINT (-111.7638113 34.86597028)     C02   TNC01        Q01           4
64  POINT (-86.63186076 32.84057112)     C01   TOC04        Q04           1
66  POINT (-86.63186076 32.84057112)     C01   TOC07        Q32           1
67  POINT (-111.7638113 34.86597028)     C03   TSC09        Q09           4

   StratificationCategoryID1 StratificationID1 StratificationCategoryID2  \
32                       AGE            65PLUS                       RACE
33                       AGE            65PLUS                       RACE
42                       AGE       AGE_OVERALL                       RACE
47                       AGE              5064                       RACE
48                       AGE       AGE_OVERALL                       RACE
49                       AGE            65PLUS                       RACE
63                       AGE              5064                       RACE
64                       AGE       AGE_OVERALL                       RACE
66                       AGE       AGE_OVERALL                       RACE
67                       AGE              5064                       RACE

   StratificationID2
32               NAA
33               NAA
42               HIS
47               HIS
48               HIS
49               HIS
63               ASN
64               NAA
66               ASN
67               ASN
```

[ ]: