DATS 6401 - Individual Project - Joseph Valle (Data Cleanup)

February 18, 2021

0.1 1. Installing Our Data

[1]:
```
import pandas as pd
import numpy as np
from numpy import array
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
gdp_whole = pd.read_csv("gdp-whole.csv", header=4)
mil_whole = pd.read_csv("military-whole.csv", header=4)
pop_whole = pd.read_csv("population-whole.csv", header=4)
educ_pct = pd.read_excel("education-pct.xlsx", header=0)
heal_pct = pd.read_csv("health-pct.csv", header=4)
china_educ = pd.read_excel("china-educ.xlsx")
us_educ = pd.read_csv("us-educ.csv")
```

[3]:
```
mil_whole
```

[3]:
```
     Country Name Country Code                      Indicator Name  \
0           Aruba          ABW  Military expenditure (current USD)
1     Afghanistan          AFG  Military expenditure (current USD)
2          Angola          AGO  Military expenditure (current USD)
3         Albania          ALB  Military expenditure (current USD)
4         Andorra          AND  Military expenditure (current USD)
..            …            …                                  …
259        Kosovo          XKX  Military expenditure (current USD)
260   Yemen, Rep.          YEM  Military expenditure (current USD)
261  South Africa          ZAF  Military expenditure (current USD)
262        Zambia          ZMB  Military expenditure (current USD)
263      Zimbabwe          ZWE  Military expenditure (current USD)

     Indicator Code          1960          1961          1962          1963  \
0    MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
1    MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
2    MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
3    MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
4    MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
```

```
..               …             …             …             …             …
259  MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
260  MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
261  MS.MIL.XPND.CD  6.999997e+07  1.137500e+08  1.861999e+08  1.889999e+08
262  MS.MIL.XPND.CD           NaN           NaN           NaN           NaN
263  MS.MIL.XPND.CD           NaN           NaN           NaN           NaN

             1964          1965  …          2012          2013  \
0             NaN           NaN  …           NaN           NaN
1             NaN           NaN  …  2.385834e+08  2.171941e+08
2             NaN           NaN  …  4.144635e+09  6.090752e+09
3             NaN           NaN  …  1.832047e+08  1.800155e+08
4             NaN           NaN  …           NaN           NaN
..              …             …  …             …             …
259           NaN           NaN  …  4.339040e+07  4.859768e+07
260           NaN           NaN  …  1.618840e+09  1.648751e+09
261  2.715999e+08  2.894499e+08  …  4.489590e+09  4.118208e+09
262           NaN           NaN  …  3.463014e+08  3.813458e+08
263           NaN  1.560000e+07  …  3.182720e+08  3.567000e+08

             2014          2015          2016          2017          2018  \
0             NaN           NaN           NaN           NaN           NaN
1    2.682271e+08  1.995186e+08  1.858783e+08  1.914071e+08  1.980747e+08
2    6.846249e+09  3.608299e+09  2.764055e+09  3.062873e+09  1.983614e+09
3    1.781204e+08  1.323507e+08  1.308532e+08  1.443827e+08  1.758867e+08
4             NaN           NaN           NaN           NaN           NaN
..              …             …             …             …             …
259  5.357579e+07  4.998416e+07  5.193762e+07  5.726263e+07  6.334407e+07
260  1.714831e+09           NaN           NaN           NaN           NaN
261  3.892469e+09  3.488868e+09  3.169756e+09  3.637677e+09  3.670401e+09
262  4.436044e+08  3.724476e+08  2.995048e+08  3.396645e+08  3.780254e+08
263  3.681000e+08  3.766770e+08  3.580650e+08  3.405220e+08  4.203640e+08

             2019  2020  Unnamed: 65
0             NaN   NaN          NaN
1    2.268744e+08   NaN          NaN
2    1.470939e+09   NaN          NaN
3    1.975172e+08   NaN          NaN
4             NaN   NaN          NaN
..              …     …            …
259  6.570533e+07   NaN          NaN
260           NaN   NaN          NaN
261  3.465133e+09   NaN          NaN
262  2.932758e+08   NaN          NaN
263  5.469390e+08   NaN          NaN

[264 rows x 66 columns]
```

[4]:
```
educ_pct
```

[4]:
```
                                    Source                Country     1970  \
0    WB, World Development Indicators (WDI)              Australia      NaN
1                                      NaN      Brunei Darussalam  6.23915
2                                      NaN               Cambodia      NaN
3                                      NaN                   Fiji      NaN
4                                      NaN   Hong Kong SAR, China      NaN
..                                     …                      …        …
173                                    NaN            Gambia, The      NaN
174                                    NaN                   Togo  2.11380
175                                    NaN                 Uganda      NaN
176                                    NaN                 Zambia  4.38781
177                                    NaN               Zimbabwe      NaN

         1990     2000     2005     2010     2011     2012     2013     2014  \
0     4.67634  4.89147  4.91030  5.55006  5.07451  4.86900  5.22974  5.16477
1     3.95996  3.70591      NaN  2.04661  3.32210  2.88959      NaN  3.35319
2         NaN  1.65599      NaN  1.53379  1.51069  1.56090  2.05054  1.90939
3         NaN  5.85725  5.12197      NaN  4.17306      NaN  3.88289      NaN
4     2.46609      NaN  4.13463  3.51003  3.41822  3.50961  3.76032  3.57255
..        …        …        …        …        …        …        …        …
173   2.82584  1.46587  0.69188  2.56322  2.50117  2.63757  1.81839  2.23094
174       NaN  3.92090  3.15978  4.09635  4.30930  4.71920  4.42365  4.78455
175       NaN  2.46167      NaN  1.74151  2.31769  1.81216  1.88643  1.93036
176       NaN  1.78819  1.73553      NaN      NaN      NaN      NaN      NaN
177  12.45426      NaN      NaN  1.54406      NaN  6.07021  5.99598  6.13835

        2015     2016     2017     2018     2019
0    5.31127  5.27678  5.12425      NaN      NaN
1        NaN  4.42541      NaN      NaN      NaN
2        NaN      NaN      NaN  2.16286      NaN
3        NaN      NaN      NaN      NaN      NaN
4    3.26212  3.29265  3.31010  3.32652  3.81058
..       …        …        …        …        …
173  2.16003  2.02443      NaN  2.42030      NaN
174  5.10860  4.95558  4.98976  5.36957      NaN
175  2.34049  2.15536  2.27205  2.13052      NaN
176  4.62432  3.74792  3.72965  4.61809      NaN
177      NaN      NaN  5.81878  5.87135      NaN

[178 rows x 16 columns]
```

[5]:
```
china_educ
```

[5]:
```
    Unnamed: 0                                          Unnamed: 1  Unnamed: 2  \
0          NaN                                                 NaN         NaN
```

```
1           NaN  Share of GDP spent on public education in Chin…        NaN
2           NaN  Public expenditure on education as a share of …        NaN
3           NaN                                              NaN          NaN
4           NaN                                             2009         3.59
5           NaN                                             2010         3.65
6           NaN                                             2011         3.93
7           NaN                                             2012         4.28
8           NaN                                             2013         4.16
9           NaN                                             2014         4.10
10          NaN                                             2015         4.26
11          NaN                                             2016         4.22
12          NaN                                             2017         4.14
13          NaN                                             2018         4.11
14          NaN                                             2019         4.04

   Unnamed: 3
0         NaN
1         NaN
2         NaN
3         NaN
4        in %
5        in %
6        in %
7        in %
8        in %
9        in %
10       in %
11       in %
12       in %
13       in %
14       in %
```

[6]:
```
us_educ
```

[6]:
```
          DATE  G160291A027NBEA
0   1959-01-01           14.484
1   1960-01-01           15.902
2   1961-01-01           17.329
3   1962-01-01           18.567
4   1963-01-01           20.260
..         …                …
56  2015-01-01          904.776
57  2016-01-01          925.797
58  2017-01-01          954.364
59  2018-01-01          992.997
60  2019-01-01         1021.492
```

```
[61 rows x 2 columns]
```

[7]:
```python
col_list = ['Country Name','2013','2014','2015','2016','2017']
col_list_gdp = ['Country Name','2013 GDP','2014 GDP','2015 GDP','2016␣
↪GDP','2017 GDP']
col_list_mil = ['Country Name','2013 Mil','2014 Mil','2015 Mil','2016␣
↪Mil','2017 Mil']
col_list_pop = ['Country Name','2013 Pop','2014 Pop','2015 Pop','2016␣
↪Pop','2017 Pop']
col_list_educ = ['Country Name','2013 Educ Pct','2014 Educ Pct','2015 Educ␣
↪Pct','2016 Educ Pct','2017 Educ Pct']
col_list_heal = ['Country Name','2013 Heal Pct','2014 Heal Pct','2015 Heal␣
↪Pct','2016 Heal Pct','2017 Heal Pct']

#Our columns will cover the years 2013 to 2017 for each variable.
```

0.2 2A. Cleaning Up Our Data (China Education)

[8]:
```python
china_educ = china_educ.transpose()
china_educ.columns = china_educ.iloc[1]
china_educ
```

[8]:
```
Unnamed: 1  NaN Share of GDP spent on public education in China 2009-2019  \
Unnamed: 0  NaN                                                       NaN
Unnamed: 1  NaN  Share of GDP spent on public education in Chin…
Unnamed: 2  NaN                                                       NaN
Unnamed: 3  NaN                                                       NaN

Unnamed: 1 Public expenditure on education as a share of gross domestic product
(GDP) in China from 2009 to 2019  \
Unnamed: 0                                                  NaN
Unnamed: 1  Public expenditure on education as a share of …
Unnamed: 2                                                  NaN
Unnamed: 3                                                  NaN

Unnamed: 1  NaN  2009  2010  2011  2012  2013  2014  2015  2016  2017  2018  \
Unnamed: 0  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN
Unnamed: 1  NaN  2009  2010  2011  2012  2013  2014  2015  2016  2017  2018
Unnamed: 2  NaN  3.59  3.65  3.93  4.28  4.16   4.1  4.26  4.22  4.14  4.11
Unnamed: 3  NaN  in %  in %  in %  in %  in %  in %  in %  in %  in %  in %

Unnamed: 1  2019
Unnamed: 0   NaN
Unnamed: 1  2019
Unnamed: 2  4.04
Unnamed: 3  in %
```

[9]:
```
china_educ = china_educ.iloc[2,[8,9,10,11,12]]
china_educ = pd.DataFrame(china_educ)
china_educ = china_educ.transpose()
china_educ

#We isolate the row showing China's percentage figures on educational spending.
```

[9]:
```
Unnamed: 1  2013 2014  2015  2016  2017
Unnamed: 2  4.16  4.1  4.26  4.22  4.14
```

[10]:
```
china_educ = china_educ.reset_index(drop=True, inplace=False)
china_educ['Country Name'] = 'China'
china_educ = china_educ[['Country Name','2013','2014','2015','2016','2017']]
china_educ
```

[10]:
```
Unnamed: 1 Country Name  2013 2014  2015  2016  2017
0                 China  4.16  4.1  4.26  4.22  4.14
```

[11]:
```
china_educ = pd.DataFrame(china_educ, index=range(1), columns=list(col_list))
china_educ.columns = col_list_educ
china_educ

#We rename the columns so that they can align with the other countries later.
```

[11]:
```
  Country Name 2013 Educ Pct 2014 Educ Pct 2015 Educ Pct 2016 Educ Pct  \
0        China          4.16           4.1          4.26          4.22

  2017 Educ Pct
0          4.14
```

0.3 2B. Cleaning Up Our Data (US Education)

[12]:
```
us_educ = us_educ.transpose()
us_educ.columns = us_educ.iloc[0]
us_educ
```

[12]:
```
DATE             1959-01-01  1960-01-01  1961-01-01  1962-01-01  1963-01-01  \
DATE             1959-01-01  1960-01-01  1961-01-01  1962-01-01  1963-01-01
G160291A027NBEA      14.484      15.902      17.329      18.567       20.26

DATE             1964-01-01  1965-01-01  1966-01-01  1967-01-01  1968-01-01  \
DATE             1964-01-01  1965-01-01  1966-01-01  1967-01-01  1968-01-01
G160291A027NBEA      22.264      24.848      28.506       32.26      36.376

DATE             …  2010-01-01  2011-01-01  2012-01-01  2013-01-01  \
DATE             …  2010-01-01  2011-01-01  2012-01-01  2013-01-01
G160291A027NBEA  …     813.238     821.745     828.477     860.947
```

```
DATE            2014-01-01  2015-01-01  2016-01-01  2017-01-01  2018-01-01  \
DATE            2014-01-01  2015-01-01  2016-01-01  2017-01-01  2018-01-01
G160291A027NBEA    883.586     904.776     925.797     954.364     992.997

DATE            2019-01-01
DATE            2019-01-01
G160291A027NBEA    1021.49

[2 rows x 61 columns]
```

[13]:
```
us_educ = us_educ.iloc[1,54:59]
us_educ = pd.DataFrame(us_educ)
us_educ = us_educ.transpose()
us_educ

#Similarly like we did with China, we isolate the row showing the US's absolute␣
↪figures for educational spending.
```

[13]:
```
DATE            2013-01-01 2014-01-01 2015-01-01 2016-01-01 2017-01-01
G160291A027NBEA    860.947    883.586    904.776    925.797    954.364
```

[14]:
```
us_educ = us_educ.reset_index(drop=True, inplace=False)
us_educ['Country Name'] = 'United States'
us_educ = us_educ[['Country␣
↪Name','2013-01-01','2014-01-01','2015-01-01','2016-01-01','2017-01-01']]
us_educ = us_educ.rename(columns={'2013-01-01':'2013','2014-01-01':
↪'2014','2015-01-01':'2015',
                                  '2016-01-01':'2016','2017-01-01':'2017'})
us_educ
```

[14]:
```
DATE   Country Name     2013     2014     2015     2016     2017
0     United States  860.947  883.586  904.776  925.797  954.364
```

[15]:
```
gdp_whole.iloc[240:250,:] #The US is located on the row with index number 249␣
↪in our GDP dataset.
```

[15]:
```
            Country Name Country Code     Indicator Name  Indicator Code  \
240  Trinidad and Tobago          TTO  GDP (current US$)  NY.GDP.MKTP.CD
241              Tunisia          TUN  GDP (current US$)  NY.GDP.MKTP.CD
242               Turkey          TUR  GDP (current US$)  NY.GDP.MKTP.CD
243               Tuvalu          TUV  GDP (current US$)  NY.GDP.MKTP.CD
244             Tanzania          TZA  GDP (current US$)  NY.GDP.MKTP.CD
245               Uganda          UGA  GDP (current US$)  NY.GDP.MKTP.CD
246              Ukraine          UKR  GDP (current US$)  NY.GDP.MKTP.CD
247  Upper middle income          UMC  GDP (current US$)  NY.GDP.MKTP.CD
248              Uruguay          URY  GDP (current US$)  NY.GDP.MKTP.CD
```

```
249        United States         USA  GDP (current US$)  NY.GDP.MKTP.CD

             1960          1961          1962          1963          1964  \
240  5.356701e+08  5.849612e+08  6.193192e+08  6.782354e+08  7.118934e+08
241           NaN           NaN           NaN           NaN           NaN
242  1.399507e+10  7.988889e+09  8.922222e+09  1.035556e+10  1.117778e+10
243           NaN           NaN           NaN           NaN           NaN
244           NaN           NaN           NaN           NaN           NaN
245  4.230084e+08  4.415241e+08  4.490126e+08  5.161478e+08  5.890566e+08
246           NaN           NaN           NaN           NaN           NaN
247  2.237883e+11  2.039626e+11  2.147155e+11  2.243569e+11  2.537432e+11
248  1.242289e+09  1.547389e+09  1.710004e+09  1.539681e+09  1.975702e+09
249  5.433000e+11  5.633000e+11  6.051000e+11  6.386000e+11  6.858000e+11

             1965  …          2012          2013          2014  \
240  7.365689e+08  …  2.576322e+10  2.726848e+10  2.761584e+10
241  9.910476e+08  …  4.504411e+10  4.625106e+10  4.763233e+10
242  1.196667e+10  …  8.805560e+11  9.577994e+11  9.389344e+11
243           NaN  …  3.767177e+07  3.750908e+07  3.729061e+07
244           NaN  …  3.965053e+10  4.568053e+10  4.996479e+10
245  8.848739e+08  …  2.718867e+10  2.879163e+10  3.247236e+10
246           NaN  …  1.757814e+11  1.833101e+11  1.335034e+11
247  2.816171e+11  …  2.078827e+13  2.203138e+13  2.273000e+13
248  1.890769e+09  …  5.126439e+10  5.753123e+10  5.723601e+10
249  7.437000e+11  …  1.619701e+13  1.678485e+13  1.752716e+13

             2015          2016          2017          2018          2019  \
240  2.506289e+10  2.228478e+10  2.247483e+10  2.380815e+10  2.426971e+10
241  4.317348e+10  4.180121e+10  3.980243e+10  3.977030e+10  3.879669e+10
242  8.643143e+11  8.696831e+11  8.589886e+11  7.783819e+11  7.614255e+11
243  3.549207e+07  3.654780e+07  4.061925e+07  4.258816e+07  4.727146e+07
244  4.737860e+10  4.977402e+10  5.332063e+10  5.800120e+10  6.317707e+10
245  3.224812e+10  2.907872e+10  3.075647e+10  3.291615e+10  3.516516e+10
246  9.103096e+10  9.335599e+10  1.121904e+11  1.309019e+11  1.537811e+11
247  2.135217e+13  2.133891e+13  2.358067e+13  2.527329e+13  2.580776e+13
248  5.327430e+10  5.268761e+10  5.953009e+10  5.959689e+10  5.604591e+10
249  1.822470e+13  1.871496e+13  1.951935e+13  2.058016e+13  2.143323e+13

     2020  Unnamed: 65
240   NaN          NaN
241   NaN          NaN
242   NaN          NaN
243   NaN          NaN
244   NaN          NaN
245   NaN          NaN
246   NaN          NaN
247   NaN          NaN
```

```
248   NaN          NaN
249   NaN          NaN

[10 rows x 66 columns]
```

[16]:
```
us_gdp = gdp_whole.iloc[249,:]
us_gdp = pd.DataFrame(us_gdp)
us_gdp = us_gdp.transpose()
us_gdp

#By taking the US's absolute GDP figures, we can use these to calculate the␣
 ↪percentage amounts in educational
#spending over each year.
```

[16]:
```
      Country Name Country Code     Indicator Name  Indicator Code       1960  \
249  United States          USA  GDP (current US$)  NY.GDP.MKTP.CD  5.433e+11

          1961       1962       1963       1964       1965  …        2012  \
249  5.633e+11  6.051e+11  6.386e+11  6.858e+11  7.437e+11  …  1.6197e+13

            2013         2014         2015        2016         2017  \
249  1.67848e+13  1.75272e+13  1.82247e+13  1.8715e+13  1.95194e+13

            2018         2019 2020 Unnamed: 65
249  2.05802e+13  2.14332e+13  NaN         NaN

[1 rows x 66 columns]
```

[17]:
```
us_gdp = us_gdp.iloc[:,np.r_[0,57:62]]
us_gdp

#Only the years 2013 through 2017 will be included in our data.
```

[17]:
```
      Country Name         2013         2014         2015        2016  \
249  United States  1.67848e+13  1.75272e+13  1.82247e+13  1.8715e+13

            2017
249  1.95194e+13
```

[18]:
```
us_educ.iloc[:,1:6] = us_educ.iloc[:,1:6]*(1*10**9)
us_educ

#Our educational figures for the US need to be converted into billions like our␣
 ↪GDP figures.
```

[18]:
```
DATE   Country Name         2013         2014         2015         2016  \
0     United States  8.60947e+11  8.83586e+11  9.04776e+11  9.25797e+11
```

```
DATE          2017
0     9.54364e+11
```

[19]:
```python
print(us_educ.dtypes)
print()
print(us_gdp.dtypes)
```

```
DATE
Country Name    object
2013            object
2014            object
2015            object
2016            object
2017            object
dtype: object

Country Name    object
2013            object
2014            object
2015            object
2016            object
2017            object
dtype: object
```

[20]:
```python
us_educ['2013'] = us_educ['2013'].astype(float) #Our data must be converted as␣
↪floats with the rest for later.
us_educ['2014'] = us_educ['2014'].astype(float)
us_educ['2015'] = us_educ['2015'].astype(float)
us_educ['2016'] = us_educ['2016'].astype(float)
us_educ['2017'] = us_educ['2017'].astype(float)

us_gdp['2013'] = us_gdp['2013'].astype(float)
us_gdp['2014'] = us_gdp['2014'].astype(float)
us_gdp['2015'] = us_gdp['2015'].astype(float)
us_gdp['2016'] = us_gdp['2016'].astype(float)
us_gdp['2017'] = us_gdp['2017'].astype(float)
```

[21]:
```python
us_data = pd.concat([us_educ, us_gdp])
us_data = us_data.reset_index(drop=True)
us_data
```

[21]:
```
DATE   Country Name          2013          2014          2015          2016  \
0     United States  8.609470e+11  8.835860e+11  9.047760e+11  9.257970e+11
1     United States  1.678485e+13  1.752716e+13  1.822470e+13  1.871496e+13

DATE          2017
```

```
0     9.543640e+11
1     1.951935e+13
```

[22]:
```
us_data['2013 Educ Pct'] = us_data.iloc[0,1]/us_data.iloc[1,1]*100
us_data['2014 Educ Pct'] = us_data.iloc[0,2]/us_data.iloc[1,2]*100
us_data['2015 Educ Pct'] = us_data.iloc[0,3]/us_data.iloc[1,3]*100
us_data['2016 Educ Pct'] = us_data.iloc[0,4]/us_data.iloc[1,4]*100
us_data['2017 Educ Pct'] = us_data.iloc[0,5]/us_data.iloc[1,5]*100
us_data

#To find the percentage amounts in educational spending, we divide the␣
↪educational figures by the GDP ones and
#multiply these by 100.
```

[22]:
```
DATE   Country Name          2013          2014          2015          2016  \
0     United States  8.609470e+11  8.835860e+11  9.047760e+11  9.257970e+11
1     United States  1.678485e+13  1.752716e+13  1.822470e+13  1.871496e+13

DATE          2017  2013 Educ Pct  2014 Educ Pct  2015 Educ Pct  \
0     9.543640e+11        5.12931       5.041238       4.964558
1     1.951935e+13        5.12931       5.041238       4.964558

DATE  2016 Educ Pct  2017 Educ Pct
0          4.946828       4.889322
1          4.946828       4.889322
```

[23]:
```
us_data = us_data.iloc[0,np.r_[0,6:11]]
us_data = pd.DataFrame(us_data)
us_data = us_data.transpose()
us_data
```

[23]:
```
DATE   Country Name 2013 Educ Pct 2014 Educ Pct 2015 Educ Pct 2016 Educ Pct  \
0     United States       5.12931       5.04124       4.96456       4.94683

DATE 2017 Educ Pct
0          4.88932
```

0.4 2C. Cleaning Up the Rest of Our Data

[24]:
```
print(gdp_whole.loc[:,'2013':'2017'].dtypes) #This confirms we are using floats␣
↪to represent our data properly.
print()
print(mil_whole.loc[:,'2013':'2017'].dtypes)
print()
print(pop_whole.loc[:,'2013':'2017'].dtypes)
print()
print(educ_pct.loc[:,'2013':'2017'].dtypes)
```

```
print()
print(heal_pct.loc[:,'2013':'2017'].dtypes)
```

```
2013    float64
2014    float64
2015    float64
2016    float64
2017    float64
dtype: object

2013    float64
2014    float64
2015    float64
2016    float64
2017    float64
dtype: object

2013    float64
2014    float64
2015    float64
2016    float64
2017    float64
dtype: object

2013    float64
2014    float64
2015    float64
2016    float64
2017    float64
dtype: object

2013    float64
2014    float64
2015    float64
2016    float64
2017    float64
dtype: object
```

[25]:
```
country_list1 = ['Australia','Brazil','France','Germany','Italy','Japan',
                 'Korea, Rep.','Mexico','Russian Federation','United Kingdom']
country_list2 =␣
 ↪['Australia','Brazil','China','France','Germany','Italy','Japan','Korea, Rep.
 ↪',
                 'Mexico','Russian Federation','United Kingdom','United States']
mil_whole = mil_whole.rename(columns={'Country':'Country Name'})
educ_pct = educ_pct.rename(columns={'Country':'Country Name'})
```

```
#Our first country list does not include China and the US because their␣
↪educational figures were not fully reported
#alongside the other countries at first.
```

[26]:
```
gdp_whole = gdp_whole[col_list]
mil_whole = mil_whole[col_list]
pop_whole = pop_whole[col_list]
educ_pct = educ_pct[col_list]
heal_pct = heal_pct[col_list]

gdp_whole.columns = col_list_gdp
mil_whole.columns = col_list_mil
pop_whole.columns = col_list_pop
educ_pct.columns = col_list_educ
heal_pct.columns = col_list_heal

#We first isolate in our datasets the Country Name column as well as our years␣
↪of interest (2013-2017).  After that,
#we rename the year columns to correspond with their respective variables.
```

[27]:
```
gdp_whole = gdp_whole[gdp_whole['Country Name'].isin(country_list2)]
mil_whole = mil_whole[mil_whole['Country Name'].isin(country_list2)]
pop_whole = pop_whole[pop_whole['Country Name'].isin(country_list2)]
educ_pct = educ_pct[educ_pct['Country Name'].isin(country_list1)]
heal_pct = heal_pct[heal_pct['Country Name'].isin(country_list2)]

#Recall that our original dataset reporting the shares of GDP toward␣
↪educational spending did not include China and
#the US, but this will be fixed soon.
```

[28]:
```
mil_whole
```

[28]:
```
           Country Name      2013 Mil      2014 Mil      2015 Mil  \
11            Australia  2.482526e+10  2.578371e+10  2.404621e+10
27               Brazil  3.287479e+10  3.265961e+10  2.461770e+10
38                China  1.798805e+11  2.007722e+11  2.144715e+11
53              Germany  4.486569e+10  4.421606e+10  3.702007e+10
75               France  5.200146e+10  5.313475e+10  4.564747e+10
79       United Kingdom  5.686176e+10  5.918286e+10  5.386219e+10
114               Italy  2.995745e+10  2.770103e+10  2.218085e+10
117               Japan  4.902393e+10  4.690347e+10  4.210610e+10
124         Korea, Rep.  3.431122e+10  3.755230e+10  3.657077e+10
152              Mexico  6.473144e+09  6.758668e+09  5.468838e+09
200  Russian Federation  8.835290e+10  8.469650e+10  6.641833e+10
249       United States  6.792290e+11  6.477890e+11  6.338296e+11

         2016 Mil      2017 Mil
```

```
11   2.638295e+10  2.769111e+10
27   2.422475e+10  2.928305e+10
38   2.164043e+11  2.284663e+11
53   3.972491e+10  4.236577e+10
75   4.737059e+10  4.919566e+10
79   4.811894e+10  4.643330e+10
114  2.503303e+10  2.644789e+10
117  4.647129e+10  4.538703e+10
124  3.688528e+10  3.917068e+10
152  5.336876e+09  5.062077e+09
200  6.924531e+10  6.652730e+10
249  6.398564e+11  6.467529e+11
```

[29]:
```
educ_pct
```

[29]:
```
          Country Name  2013 Educ Pct  2014 Educ Pct  2015 Educ Pct  \
0            Australia        5.22974        5.16477        5.31127
6                Japan        3.66538        3.59059            NaN
8          Korea, Rep.        5.24758        5.26613        5.25271
42              France        5.50028        5.51206        5.46424
44             Germany        4.96111        4.93962        4.83498
49               Italy        4.14367        4.06128        4.07363
63  Russian Federation        3.76383        4.01388        3.83403
75      United Kingdom        5.52980        5.59900        5.55616
80              Brazil        5.83885        5.94848        6.24106
95              Mexico        4.69605        5.26062        5.23524

    2016 Educ Pct  2017 Educ Pct
0         5.27678        5.12425
6         3.18618        3.18218
8         4.33310            NaN
42        5.43259        5.45160
44        4.84022        4.90512
49        3.81579        4.04494
63        3.76044        4.68991
75        5.42670        5.44209
80        6.31404        6.32255
95        4.90949        4.52282
```

[30]:
```
educ_pct.iloc[1,3] = educ_pct.iloc[1,[1,2,4,5]].mean(axis=0)
educ_pct.iloc[2,5] = educ_pct.iloc[2,[1,2,3,4]].mean(axis=0)
educ_pct

#To find the best estimates for Japan and South Korea's shares of GDP toward
 ↪education in 2015 and 2017, respectively,
#we calculate them as the means over the rest of the time span.
```

[30]:
```
            Country Name  2013 Educ Pct  2014 Educ Pct  2015 Educ Pct  \
0              Australia        5.22974        5.16477       5.311270
6                  Japan        3.66538        3.59059       3.406083
8            Korea, Rep.        5.24758        5.26613       5.252710
42                France        5.50028        5.51206       5.464240
44               Germany        4.96111        4.93962       4.834980
49                 Italy        4.14367        4.06128       4.073630
63    Russian Federation        3.76383        4.01388       3.834030
75        United Kingdom        5.52980        5.59900       5.556160
80                Brazil        5.83885        5.94848       6.241060
95                Mexico        4.69605        5.26062       5.235240

    2016 Educ Pct  2017 Educ Pct
0         5.27678        5.12425
6         3.18618        3.18218
8         4.33310        5.02488
42        5.43259        5.45160
44        4.84022        4.90512
49        3.81579        4.04494
63        3.76044        4.68991
75        5.42670        5.44209
80        6.31404        6.32255
95        4.90949        4.52282
```

[31]:
```python
educ_pct = pd.concat([educ_pct, china_educ, us_data], axis=0)
educ_pct = educ_pct.rename_axis(None, axis=1)
educ_pct

#We fix our dataset reporting the shares of GDP toward educational spending to␣
↪include China and the US.
```

[31]:
```
            Country Name 2013 Educ Pct 2014 Educ Pct 2015 Educ Pct  \
0              Australia       5.22974       5.16477       5.31127
6                  Japan       3.66538       3.59059       3.40608
8            Korea, Rep.       5.24758       5.26613       5.25271
42                France       5.50028       5.51206       5.46424
44               Germany       4.96111       4.93962       4.83498
49                 Italy       4.14367       4.06128       4.07363
63    Russian Federation       3.76383       4.01388       3.83403
75        United Kingdom        5.5298         5.599       5.55616
80                Brazil       5.83885       5.94848       6.24106
95                Mexico       4.69605       5.26062       5.23524
0                  China          4.16           4.1          4.26
0          United States       5.12931       5.04124       4.96456

   2016 Educ Pct 2017 Educ Pct
0        5.27678       5.12425
```

```
6        3.18618       3.18218
8         4.3331       5.02488
42       5.43259        5.4516
44       4.84022       4.90512
49       3.81579       4.04494
63       3.76044       4.68991
75        5.4267       5.44209
80       6.31404       6.32255
95       4.90949       4.52282
0           4.22          4.14
0        4.94683       4.88932
```

[32]:
```
gdp_whole = gdp_whole.sort_values(by='Country Name',ascending=True)
mil_whole = mil_whole.sort_values(by='Country Name',ascending=True)
pop_whole = pop_whole.sort_values(by='Country Name',ascending=True)
educ_pct = educ_pct.sort_values(by='Country Name',ascending=True)
heal_pct = heal_pct.sort_values(by='Country Name',ascending=True)

data = gdp_whole.merge(mil_whole, on=['Country Name'])
data = data.merge(pop_whole, on=['Country Name'])
data = data.merge(educ_pct, on=['Country Name'])
data = data.merge(heal_pct, on=['Country Name'])
data

#We arrange each of the countries into alphabetical order and consolidate all␣
↪of the datasets together into one.
```

[32]:
```
          Country Name      2013 GDP      2014 GDP      2015 GDP  \
0            Australia  1.576184e+12  1.467484e+12  1.351694e+12
1               Brazil  2.472807e+12  2.455994e+12  1.802214e+12
2                China  9.570406e+12  1.047568e+13  1.106155e+13
3               France  2.811078e+12  2.852166e+12  2.438208e+12
4              Germany  3.732743e+12  3.883920e+12  3.356236e+12
5                Italy  2.141315e+12  2.159134e+12  1.835899e+12
6                Japan  5.155717e+12  4.850414e+12  4.389476e+12
7          Korea, Rep.  1.370795e+12  1.484318e+12  1.465773e+12
8               Mexico  1.274443e+12  1.315351e+12  1.171868e+12
9   Russian Federation  2.292473e+12  2.059242e+12  1.363481e+12
10      United Kingdom  2.786023e+12  3.063803e+12  2.928591e+12
11       United States  1.678485e+13  1.752716e+13  1.822470e+13

        2016 GDP      2017 GDP      2013 Mil      2014 Mil      2015 Mil  \
0   1.208847e+12  1.329188e+12  2.482526e+10  2.578371e+10  2.404621e+10
1   1.795700e+12  2.062831e+12  3.287479e+10  3.265961e+10  2.461770e+10
2   1.123328e+13  1.231041e+13  1.798805e+11  2.007722e+11  2.144715e+11
3   2.471286e+12  2.595151e+12  5.200146e+10  5.313475e+10  4.564747e+10
4   3.467498e+12  3.682602e+12  4.486569e+10  4.421606e+10  3.702007e+10
```

```
5   1.875797e+12  1.961796e+12  2.995745e+10  2.770103e+10  2.218085e+10
6   4.922538e+12  4.866864e+12  4.902393e+10  4.690347e+10  4.210610e+10
7   1.500112e+12  1.623901e+12  3.431122e+10  3.755230e+10  3.657077e+10
8   1.078491e+12  1.158913e+12  6.473144e+09  6.758668e+09  5.468838e+09
9   1.276787e+12  1.574199e+12  8.835290e+10  8.469650e+10  6.641833e+10
10  2.694283e+12  2.666229e+12  5.686176e+10  5.918286e+10  5.386219e+10
11  1.871496e+13  1.951935e+13  6.792290e+11  6.477890e+11  6.338296e+11

        2016 Mil  …  2013 Educ Pct  2014 Educ Pct  2015 Educ Pct  \
0   2.638295e+10  …        5.22974        5.16477        5.31127
1   2.422475e+10  …        5.83885        5.94848        6.24106
2   2.164043e+11  …           4.16            4.1           4.26
3   4.737059e+10  …        5.50028        5.51206        5.46424
4   3.972491e+10  …        4.96111        4.93962        4.83498
5   2.503303e+10  …        4.14367        4.06128        4.07363
6   4.647129e+10  …        3.66538        3.59059        3.40608
7   3.688528e+10  …        5.24758        5.26613        5.25271
8   5.336876e+09  …        4.69605        5.26062        5.23524
9   6.924531e+10  …        3.76383        4.01388        3.83403
10  4.811894e+10  …         5.5298          5.599        5.55616
11  6.398564e+11  …        5.12931        5.04124        4.96456

    2016 Educ Pct  2017 Educ Pct  2013 Heal Pct 2014 Heal Pct 2015 Heal Pct  \
0         5.27678        5.12425       8.761164      9.035971      9.323312
1         6.31404        6.32255       7.977212      8.396250      8.870910
2            4.22           4.14       4.710022      4.773228      4.888723
3         5.43259         5.4516      11.436080     11.580735     11.466167
4         4.84022        4.90512      10.991872     11.015840     11.178020
5         3.81579        4.04494       8.775438      8.867919      8.856892
6         3.18618        3.18218      10.791594     10.832049     10.885507
7          4.3331        5.02488       6.247894      6.474381      6.652717
8         4.90949        4.52282       5.940777      5.630386      5.797087
9         3.76044        4.68991       5.079809      5.180229      5.295604
10         5.4267        5.44209       9.978552      9.957584      9.904352
11        4.94683        4.88932      16.210423     16.406448     16.710752

   2016 Heal Pct 2017 Heal Pct
0       9.200713      9.205448
1       9.207422      9.469337
2       4.981881      5.151192
3      11.501345     11.390502
4      11.230436     11.322296
5       8.725330      8.677740
6      10.834610     10.796343
7       6.914327      7.106949
8       5.615698      5.518895
9       5.265220      5.343881
```

```
10      9.866712      9.825112
11     17.048983     17.003614

[12 rows x 26 columns]
```

[33]:
```python
print(data.dtypes)
```

```
Country Name      object
2013 GDP         float64
2014 GDP         float64
2015 GDP         float64
2016 GDP         float64
2017 GDP         float64
2013 Mil         float64
2014 Mil         float64
2015 Mil         float64
2016 Mil         float64
2017 Mil         float64
2013 Pop         float64
2014 Pop         float64
2015 Pop         float64
2016 Pop         float64
2017 Pop         float64
2013 Educ Pct     object
2014 Educ Pct     object
2015 Educ Pct     object
2016 Educ Pct     object
2017 Educ Pct     object
2013 Heal Pct    float64
2014 Heal Pct    float64
2015 Heal Pct    float64
2016 Heal Pct    float64
2017 Heal Pct    float64
dtype: object
```

[34]:
```python
data['2013 Educ Pct'] = data['2013 Educ Pct'].astype(float) #This is a quick␣
 ↪fix to ensure we are working with floats.
data['2014 Educ Pct'] = data['2014 Educ Pct'].astype(float)
data['2015 Educ Pct'] = data['2015 Educ Pct'].astype(float)
data['2016 Educ Pct'] = data['2016 Educ Pct'].astype(float)
data['2017 Educ Pct'] = data['2017 Educ Pct'].astype(float)
```

[35]:
```python
data.iloc[:,1:16] = data.iloc[:,1:16].apply(lambda x: x/(1*10**9))
data.iloc[:,0:16]

#We convert our figures for GDP, military spending, and population into␣
 ↪billions.
```

[35]:

```
           Country Name      2013 GDP      2014 GDP      2015 GDP  \
0             Australia   1576.184467   1467.483705   1351.693985
1                Brazil   2472.806920   2455.993625   1802.214374
2                 China   9570.406236  10475.682921  11061.553080
3                France   2811.077726   2852.165761   2438.207896
4               Germany   3732.743446   3883.920155   3356.235704
5                 Italy   2141.315327   2159.133920   1835.899237
6                 Japan   5155.717056   4850.413536   4389.475623
7           Korea, Rep.   1370.795200   1484.318220   1465.773246
8                Mexico   1274.443085   1315.351184   1171.867608
9    Russian Federation   2292.473247   2059.241965   1363.481063
10       United Kingdom   2786.022873   3063.803240   2928.591002
11        United States  16784.849196  17527.163695  18224.704440

        2016 GDP      2017 GDP    2013 Mil    2014 Mil    2015 Mil  \
0    1208.846994   1329.188476   24.825263   25.783709   24.046206
1    1795.700169   2062.831046   32.874787   32.659614   24.617702
2   11233.276537  12310.409371  179.880451  200.772204  214.471496
3    2471.285607   2595.151045   52.001462   53.134751   45.647472
4    3467.498002   3682.602480   44.865692   44.216058   37.020073
5    1875.797464   1961.796197   29.957446   27.701034   22.180845
6    4922.538141   4866.864410   49.023932   46.903467   42.106103
7    1500.111596   1623.901497   34.311221   37.552299   36.570769
8    1078.490652   1158.913036    6.473144    6.758668    5.468838
9    1276.786979   1574.199387   88.352896   84.696505   66.418327
10   2694.283210   2666.229180   56.861760   59.182859   53.862185
11  18714.960538  19519.353692  679.229000  647.789000  633.829639

      2016 Mil    2017 Mil  2013 Pop  2014 Pop  2015 Pop  2016 Pop  2017 Pop
0    26.382947   27.691112  0.023128  0.023476  0.023816  0.024191  0.024602
1    24.224747   29.283050  0.201036  0.202764  0.204472  0.206163  0.207834
2   216.404283  228.466270  1.357380  1.364270  1.371220  1.378665  1.386395
3    47.370590   49.195662  0.065999  0.066312  0.066548  0.066724  0.066864
4    39.724907   42.365774  0.080646  0.080982  0.081687  0.082349  0.082657
5    25.033028   26.447893  0.060234  0.060789  0.060731  0.060627  0.060537
6    46.471288   45.387032  0.127445  0.127276  0.127141  0.126995  0.126786
7    36.885283   39.170682  0.050429  0.050747  0.051015  0.051218  0.051362
8     5.336876    5.062077  0.118827  0.120355  0.121858  0.123333  0.124777
9    69.245309   66.527304  0.143507  0.143820  0.144097  0.144342  0.144497
10   48.118944   46.433303  0.064128  0.064602  0.065116  0.065612  0.066059
11  639.856443  646.752927  0.315994  0.318301  0.320635  0.322941  0.324986
```

[36]:

```python
data['2013 GDP Cap'] = data['2013 GDP']/data['2013 Pop']
data['2014 GDP Cap'] = data['2014 GDP']/data['2014 Pop']
data['2015 GDP Cap'] = data['2015 GDP']/data['2015 Pop']
data['2016 GDP Cap'] = data['2016 GDP']/data['2016 Pop']
data['2017 GDP Cap'] = data['2017 GDP']/data['2017 Pop']
```

```
data['2013 Mil Cap'] = data['2013 Mil']/data['2013 Pop']
data['2014 Mil Cap'] = data['2014 Mil']/data['2014 Pop']
data['2015 Mil Cap'] = data['2015 Mil']/data['2015 Pop']
data['2016 Mil Cap'] = data['2016 Mil']/data['2016 Pop']
data['2017 Mil Cap'] = data['2017 Mil']/data['2017 Pop']

data['2013 Mil Pct'] = data['2013 Mil']/data['2013 GDP']*100
data['2014 Mil Pct'] = data['2014 Mil']/data['2014 GDP']*100
data['2015 Mil Pct'] = data['2015 Mil']/data['2015 GDP']*100
data['2016 Mil Pct'] = data['2016 Mil']/data['2016 GDP']*100
data['2017 Mil Pct'] = data['2017 Mil']/data['2017 GDP']*100

#We calculate the GDP per-capita and military spending per-capita figures for␣
↪each country, and report the shares of
#GDP toward military spending as whole numbers.
```

[37]:
```
data['2013 Educ'] = (data['2013 Educ Pct']/100)*data['2013 GDP']
data['2014 Educ'] = (data['2014 Educ Pct']/100)*data['2014 GDP']
data['2015 Educ'] = (data['2015 Educ Pct']/100)*data['2015 GDP']
data['2016 Educ'] = (data['2016 Educ Pct']/100)*data['2016 GDP']
data['2017 Educ'] = (data['2017 Educ Pct']/100)*data['2017 GDP']

data['2013 Educ Cap'] = data['2013 Educ']/data['2013 Pop']
data['2014 Educ Cap'] = data['2014 Educ']/data['2014 Pop']
data['2015 Educ Cap'] = data['2015 Educ']/data['2015 Pop']
data['2016 Educ Cap'] = data['2016 Educ']/data['2016 Pop']
data['2017 Educ Cap'] = data['2017 Educ']/data['2017 Pop']

data['2013 Heal'] = (data['2013 Heal Pct']/100)*data['2013 GDP']
data['2014 Heal'] = (data['2014 Heal Pct']/100)*data['2014 GDP']
data['2015 Heal'] = (data['2015 Heal Pct']/100)*data['2015 GDP']
data['2016 Heal'] = (data['2016 Heal Pct']/100)*data['2016 GDP']
data['2017 Heal'] = (data['2017 Heal Pct']/100)*data['2017 GDP']

data['2013 Heal Cap'] = data['2013 Heal']/data['2013 Pop']
data['2014 Heal Cap'] = data['2014 Heal']/data['2014 Pop']
data['2015 Heal Cap'] = data['2015 Heal']/data['2015 Pop']
data['2016 Heal Cap'] = data['2016 Heal']/data['2016 Pop']
data['2017 Heal Cap'] = data['2017 Heal']/data['2017 Pop']

#We calculate the bulk expenditures and per-capita spending on education and␣
↪healthcare.
```

[38]:
```
data.loc[:,['2013 Mil Pct','2014 Mil Pct','2015 Mil Pct','2016 Mil Pct','2017␣
↪Mil Pct']]
```

[38]:

	2013 Mil Pct	2014 Mil Pct	2015 Mil Pct	2016 Mil Pct	2017 Mil Pct
0	1.575023	1.757001	1.778968	2.182489	2.083310
1	1.329452	1.329792	1.365970	1.349042	1.419556
2	1.879549	1.916555	1.938891	1.926457	1.855879
3	1.849876	1.862962	1.872173	1.916840	1.895676
4	1.201950	1.138439	1.103024	1.145636	1.150430
5	1.399021	1.282970	1.208173	1.334527	1.348147
6	0.950865	0.966999	0.959251	0.944051	0.932572
7	2.503016	2.529936	2.494981	2.458836	2.412134
8	0.507919	0.513830	0.466677	0.494847	0.436795
9	3.854043	4.112994	4.871232	5.423403	4.226104
10	2.040965	1.931679	1.839184	1.785965	1.741535
11	4.046679	3.695915	3.477860	3.418957	3.313393

[39]:
```
data.loc[:,['2013 Mil Pct','2014 Mil Pct','2015 Mil Pct','2016 Mil Pct','2017␣
↪Mil Pct']].mean(axis=1)
```

[39]:
```
0     1.875358
1     1.358762
2     1.903466
3     1.879505
4     1.147896
5     1.314568
6     0.950748
7     2.479781
8     0.484014
9     4.497555
10    1.867866
11    3.590561
dtype: float64
```

[40]:
```
data['13-17 GDP Change'] = data['2017 GDP']-data['2013 GDP']
data['13-17 Pop Change'] = data['2017 Pop']-data['2013 Pop']
data['13-17 Mil Change'] = data['2017 Mil']-data['2013 Mil']
data['13-17 Educ Change'] = data['2017 Educ']-data['2013 Educ']
data['13-17 Heal Change'] = data['2017 Heal']-data['2013 Heal']

data['13-17 GDP Pct Change'] = data['13-17 GDP Change']/data['2013 GDP']*100
data['13-17 Pop Pct Change'] = data['13-17 Pop Change']/data['2013 Pop']*100
data['13-17 Mil Pct Change'] = data['13-17 Mil Change']/data['2013 Mil']*100
data['13-17 Educ Pct Change'] = data['13-17 Educ Change']/data['2013 Educ']*100
data['13-17 Heal Pct Change'] = data['13-17 Heal Change']/data['2013 Heal']*100

data['13-17 Mil Average'] = data.loc[:,['2013 Mil','2014 Mil','2015 Mil',
                                        '2016 Mil','2017 Mil']].mean(axis=1)
data['13-17 Mil Pct Average'] = data.loc[:,['2013 Mil Pct','2014 Mil Pct','2015␣
↪Mil Pct',
```

```
                                                 '2016 Mil Pct','2017 Mil Pct']].
↪mean(axis=1)
data['13-17 Educ Average'] = data.loc[:,['2013 Educ','2014 Educ','2015 Educ',
                                         '2016 Educ','2017 Educ']].mean(axis=1)
data['13-17 Educ Pct Average'] = data.loc[:,['2013 Educ Pct','2014 Educ␣
↪Pct','2015 Educ Pct',
                                             '2016 Educ Pct','2017 Educ Pct']].
↪mean(axis=1)
data['13-17 Heal Average'] = data.loc[:,['2013 Heal','2014 Heal','2015 Heal',
                                         '2016 Heal','2017 Heal']].mean(axis=1)
data['13-17 Heal Pct Average'] = data.loc[:,['2013 Heal Pct','2014 Heal␣
↪Pct','2015 Heal Pct',
                                             '2016 Heal Pct','2017 Heal Pct']].
↪mean(axis=1)

data.iloc[7,0] = 'South Korea'
data.iloc[9,0] = 'Russia'
data = data.sort_values(by='Country Name',ascending=True)
data = data.set_index('Country Name')

data.loc[:,'13-17 GDP Change':'13-17 Heal Pct Average']

#Finally, we calculate the fixed and percentage changes for GDP, military,␣
↪educational, and healthcare spending.  As
#an additional measure, we determine the average amounts of spending and shares␣
↪of GDP within each category over our
#time span.  We also set the index as the countries themselves.
```

[40]:

```
                13-17 GDP Change  13-17 Pop Change  13-17 Mil Change  \
Country Name
Australia            -246.995991          0.001474          2.865850
Brazil               -409.975874          0.006798         -3.591737
China                2740.003135          0.029015         48.585819
France               -215.926681          0.000866         -2.805800
Germany               -50.140966          0.002011         -2.499919
Italy                -179.519130          0.000303         -3.509553
Japan                -288.852647         -0.000659         -3.636901
Mexico               -115.530049          0.005950         -1.411068
Russia               -718.273860          0.000990        -21.825592
South Korea           253.106297          0.000933          4.859461
United Kingdom       -119.793693          0.001931        -10.428456
United States        2734.504496          0.008992        -32.476073

                13-17 Educ Change  13-17 Heal Change  13-17 GDP Pct Change  \
Country Name
Australia              -14.319412         -15.734346            -15.670500
```

Brazil	-13.959967	-1.924628	-16.579373
China	111.522049	183.364505	28.629956
France	-13.139890	-25.876366	-7.681277
Germany	-4.549446	6.656783	-1.343274
Italy	-9.375564	-17.670230	-8.383592
Japan	-34.104240	-30.940662	-5.602570
Mexico	-7.432936	-11.752630	-9.065140
Russia	-12.456261	-32.329930	-31.331832
South Korea	9.665525	29.764022	18.464195
United Kingdom	-8.962899	-16.044725	-4.299810
United States	93.417000	598.100667	16.291505

	13-17 Pop Pct Change	13-17 Mil Pct Change \
Country Name		
Australia	6.372029	11.544087
Brazil	3.381450	-10.925506
China	2.137574	27.010060
France	1.311681	-5.395618
Germany	2.494119	-5.572005
Italy	0.502642	-11.715128
Japan	-0.517245	-7.418623
Mexico	5.007410	-21.798799
Russia	0.689684	-24.702747
South Korea	1.850166	14.162893
United Kingdom	3.010507	-18.340017
United States	2.845571	-4.781314

	13-17 Educ Pct Change	13-17 Heal Pct Change \
Country Name		
Australia	-17.371529	-11.394095
Brazil	-9.668673	-0.975676
China	28.011543	40.678218
France	-8.498339	-8.049210
Germany	-2.456696	1.622425
Italy	-10.566510	-9.403570
Japan	-18.046803	-5.561026
Mexico	-12.419589	-15.522846
Russia	-14.436218	-27.762150
South Korea	13.436736	34.752445
United Kingdom	-5.817741	-5.771385
United States	10.850494	21.981762

	13-17 Mil Average	13-17 Mil Pct Average	13-17 Educ Average \
Country Name			
Australia	25.745847	1.875358	72.382754
Brazil	28.731980	1.358762	129.351961
China	207.998941	1.903466	456.509856

```
France              49.469987          1.879505        144.158367
Germany             41.638501          1.147896        177.556067
Italy               26.264049          1.314568         80.427044
Japan               45.978364          0.950748        164.871512
Mexico               5.819921          0.484014         59.151628
Russia              75.048068          4.497555         68.611581
South Korea         36.898051          2.479781         74.738590
United Kingdom      52.891810          1.867866        155.926059
United States      649.491402          3.590561        905.894000

                13-17 Educ Pct Average  13-17 Heal Average  \
Country Name
Australia                     5.221362          126.059289
Brazil                        6.132996          184.803872
China                         4.176000          537.065275
France                        5.472154          302.235934
Germany                       4.896210          403.935157
Italy                         4.027862          175.178552
Japan                         3.406083          523.676211
Mexico                        4.924844           68.445868
Russia                        4.012418           89.336049
South Korea                   5.024880           99.678494
United Kingdom                5.510750          280.188133
United States                 4.994251         3030.334259

                13-17 Heal Pct Average
Country Name
Australia                     9.105322
Brazil                        8.784226
China                         4.901009
France                       11.474966
Germany                      11.147693
Italy                         8.780664
Japan                        10.828020
Mexico                        5.700569
Russia                        5.232948
South Korea                   6.679254
United Kingdom                9.906462
United States                16.676044
```

[41]:
```python
print(data.shape)
```

```
(12, 76)
```

[42]:
```python
data.iloc[:,10:15] = data.iloc[:,10:15].apply(lambda x: x*(1*10**9))
data.iloc[:,10:15]
```

```
#We convert the population columns back to their original figures.
```

```
[42]:                   2013 Pop      2014 Pop      2015 Pop      2016 Pop  \
Country Name
Australia       2.312813e+07  2.347569e+07  2.381600e+07  2.419091e+07
Brazil          2.010359e+08  2.027637e+08  2.044718e+08  2.061631e+08
China           1.357380e+09  1.364270e+09  1.371220e+09  1.378665e+09
France          6.599869e+07  6.631207e+07  6.654827e+07  6.672410e+07
Germany         8.064560e+07  8.098250e+07  8.168661e+07  8.234867e+07
Italy           6.023395e+07  6.078914e+07  6.073058e+07  6.062750e+07
Japan           1.274450e+08  1.272760e+08  1.271410e+08  1.269945e+08
Mexico          1.188272e+08  1.203551e+08  1.218583e+08  1.233334e+08
Russia          1.435070e+08  1.438197e+08  1.440969e+08  1.443424e+08
South Korea     5.042889e+07  5.074666e+07  5.101495e+07  5.121780e+07
United Kingdom  6.412827e+07  6.460230e+07  6.511622e+07  6.561159e+07
United States   3.159937e+08  3.183010e+08  3.206352e+08  3.229413e+08

                    2017 Pop
Country Name
Australia       2.460186e+07
Brazil          2.078338e+08
China           1.386395e+09
France          6.686438e+07
Germany         8.265700e+07
Italy           6.053671e+07
Japan           1.267858e+08
Mexico          1.247773e+08
Russia          1.444967e+08
South Korea     5.136191e+07
United Kingdom  6.605886e+07
United States   3.249855e+08
```

```
[43]: list(data.columns)
```

```
[43]: ['2013 GDP',
       '2014 GDP',
       '2015 GDP',
       '2016 GDP',
       '2017 GDP',
       '2013 Mil',
       '2014 Mil',
       '2015 Mil',
       '2016 Mil',
       '2017 Mil',
       '2013 Pop',
       '2014 Pop',
       '2015 Pop',
```

```
'2016 Pop',
'2017 Pop',
'2013 Educ Pct',
'2014 Educ Pct',
'2015 Educ Pct',
'2016 Educ Pct',
'2017 Educ Pct',
'2013 Heal Pct',
'2014 Heal Pct',
'2015 Heal Pct',
'2016 Heal Pct',
'2017 Heal Pct',
'2013 GDP Cap',
'2014 GDP Cap',
'2015 GDP Cap',
'2016 GDP Cap',
'2017 GDP Cap',
'2013 Mil Cap',
'2014 Mil Cap',
'2015 Mil Cap',
'2016 Mil Cap',
'2017 Mil Cap',
'2013 Mil Pct',
'2014 Mil Pct',
'2015 Mil Pct',
'2016 Mil Pct',
'2017 Mil Pct',
'2013 Educ',
'2014 Educ',
'2015 Educ',
'2016 Educ',
'2017 Educ',
'2013 Educ Cap',
'2014 Educ Cap',
'2015 Educ Cap',
'2016 Educ Cap',
'2017 Educ Cap',
'2013 Heal',
'2014 Heal',
'2015 Heal',
'2016 Heal',
'2017 Heal',
'2013 Heal Cap',
'2014 Heal Cap',
'2015 Heal Cap',
'2016 Heal Cap',
'2017 Heal Cap',
```

```
 '13-17 GDP Change',
 '13-17 Pop Change',
 '13-17 Mil Change',
 '13-17 Educ Change',
 '13-17 Heal Change',
 '13-17 GDP Pct Change',
 '13-17 Pop Pct Change',
 '13-17 Mil Pct Change',
 '13-17 Educ Pct Change',
 '13-17 Heal Pct Change',
 '13-17 Mil Average',
 '13-17 Mil Pct Average',
 '13-17 Educ Average',
 '13-17 Educ Pct Average',
 '13-17 Heal Average',
 '13-17 Heal Pct Average']
```

[44]:

```
data.iloc[:,60:]
```

[44]:

```
                13-17 GDP Change  13-17 Pop Change  13-17 Mil Change  \
Country Name
Australia            -246.995991          0.001474          2.865850
Brazil               -409.975874          0.006798         -3.591737
China                2740.003135          0.029015         48.585819
France               -215.926681          0.000866         -2.805800
Germany               -50.140966          0.002011         -2.499919
Italy                -179.519130          0.000303         -3.509553
Japan                -288.852647         -0.000659         -3.636901
Mexico               -115.530049          0.005950         -1.411068
Russia               -718.273860          0.000990        -21.825592
South Korea           253.106297          0.000933          4.859461
United Kingdom       -119.793693          0.001931        -10.428456
United States        2734.504496          0.008992        -32.476073

                13-17 Educ Change  13-17 Heal Change  13-17 GDP Pct Change  \
Country Name
Australia              -14.319412         -15.734346            -15.670500
Brazil                 -13.959967          -1.924628            -16.579373
China                  111.522049         183.364505             28.629956
France                 -13.139890         -25.876366             -7.681277
Germany                 -4.549446           6.656783             -1.343274
Italy                   -9.375564         -17.670230             -8.383592
Japan                  -34.104240         -30.940662             -5.602570
Mexico                  -7.432936         -11.752630             -9.065140
Russia                 -12.456261         -32.329930            -31.331832
South Korea              9.665525          29.764022             18.464195
United Kingdom          -8.962899         -16.044725             -4.299810
```

United States	93.417000	598.100667	16.291505

	13-17 Pop Pct Change	13-17 Mil Pct Change \
Country Name		
Australia	6.372029	11.544087
Brazil	3.381450	-10.925506
China	2.137574	27.010060
France	1.311681	-5.395618
Germany	2.494119	-5.572005
Italy	0.502642	-11.715128
Japan	-0.517245	-7.418623
Mexico	5.007410	-21.798799
Russia	0.689684	-24.702747
South Korea	1.850166	14.162893
United Kingdom	3.010507	-18.340017
United States	2.845571	-4.781314

	13-17 Educ Pct Change	13-17 Heal Pct Change \
Country Name		
Australia	-17.371529	-11.394095
Brazil	-9.668673	-0.975676
China	28.011543	40.678218
France	-8.498339	-8.049210
Germany	-2.456696	1.622425
Italy	-10.566510	-9.403570
Japan	-18.046803	-5.561026
Mexico	-12.419589	-15.522846
Russia	-14.436218	-27.762150
South Korea	13.436736	34.752445
United Kingdom	-5.817741	-5.771385
United States	10.850494	21.981762

	13-17 Mil Average	13-17 Mil Pct Average	13-17 Educ Average \
Country Name			
Australia	25.745847	1.875358	72.382754
Brazil	28.731980	1.358762	129.351961
China	207.998941	1.903466	456.509856
France	49.469987	1.879505	144.158367
Germany	41.638501	1.147896	177.556067
Italy	26.264049	1.314568	80.427044
Japan	45.978364	0.950748	164.871512
Mexico	5.819921	0.484014	59.151628
Russia	75.048068	4.497555	68.611581
South Korea	36.898051	2.479781	74.738590
United Kingdom	52.891810	1.867866	155.926059
United States	649.491402	3.590561	905.894000

```
                   13-17 Educ Pct Average  13-17 Heal Average  \
Country Name
Australia                        5.221362          126.059289
Brazil                           6.132996          184.803872
China                            4.176000          537.065275
France                           5.472154          302.235934
Germany                          4.896210          403.935157
Italy                            4.027862          175.178552
Japan                            3.406083          523.676211
Mexico                           4.924844           68.445868
Russia                           4.012418           89.336049
South Korea                      5.024880           99.678494
United Kingdom                   5.510750          280.188133
United States                    4.994251         3030.334259

                   13-17 Heal Pct Average
Country Name
Australia                        9.105322
Brazil                           8.784226
China                            4.901009
France                          11.474966
Germany                         11.147693
Italy                            8.780664
Japan                           10.828020
Mexico                           5.700569
Russia                           5.232948
South Korea                      6.679254
United Kingdom                   9.906462
United States                   16.676044
```

[45]:
```
data = data.iloc[:,[0,1,2,3,4,10,11,12,13,14,5,6,7,8,9,40,41,42,43,44,
                   50,51,52,53,54,25,26,27,28,29,30,31,32,33,34,45,46,47,48,49,
                   55,56,57,58,59,35,36,37,38,39,15,16,17,18,19,20,21,22,23,24,
                   60,65,61,66,62,67,70,71,63,68,72,73,64,69,74,75]]
data.iloc[:,60:]
```

[45]:
```
                   13-17 GDP Change  13-17 GDP Pct Change  13-17 Pop Change  \
Country Name
Australia               -246.995991            -15.670500          0.001474
Brazil                  -409.975874            -16.579373          0.006798
China                   2740.003135             28.629956          0.029015
France                  -215.926681             -7.681277          0.000866
Germany                  -50.140966             -1.343274          0.002011
Italy                   -179.519130             -8.383592          0.000303
Japan                   -288.852647             -5.602570         -0.000659
Mexico                  -115.530049             -9.065140          0.005950
Russia                  -718.273860            -31.331832          0.000990
```

South Korea	253.106297	18.464195	0.000933
United Kingdom	-119.793693	-4.299810	0.001931
United States	2734.504496	16.291505	0.008992

Country Name	13-17 Pop Pct Change	13-17 Mil Change	13-17 Mil Pct Change \
Australia	6.372029	2.865850	11.544087
Brazil	3.381450	-3.591737	-10.925506
China	2.137574	48.585819	27.010060
France	1.311681	-2.805800	-5.395618
Germany	2.494119	-2.499919	-5.572005
Italy	0.502642	-3.509553	-11.715128
Japan	-0.517245	-3.636901	-7.418623
Mexico	5.007410	-1.411068	-21.798799
Russia	0.689684	-21.825592	-24.702747
South Korea	1.850166	4.859461	14.162893
United Kingdom	3.010507	-10.428456	-18.340017
United States	2.845571	-32.476073	-4.781314

Country Name	13-17 Mil Average	13-17 Mil Pct Average	13-17 Educ Change \
Australia	25.745847	1.875358	-14.319412
Brazil	28.731980	1.358762	-13.959967
China	207.998941	1.903466	111.522049
France	49.469987	1.879505	-13.139890
Germany	41.638501	1.147896	-4.549446
Italy	26.264049	1.314568	-9.375564
Japan	45.978364	0.950748	-34.104240
Mexico	5.819921	0.484014	-7.432936
Russia	75.048068	4.497555	-12.456261
South Korea	36.898051	2.479781	9.665525
United Kingdom	52.891810	1.867866	-8.962899
United States	649.491402	3.590561	93.417000

Country Name	13-17 Educ Pct Change	13-17 Educ Average \
Australia	-17.371529	72.382754
Brazil	-9.668673	129.351961
China	28.011543	456.509856
France	-8.498339	144.158367
Germany	-2.456696	177.556067
Italy	-10.566510	80.427044
Japan	-18.046803	164.871512
Mexico	-12.419589	59.151628
Russia	-14.436218	68.611581
South Korea	13.436736	74.738590
United Kingdom	-5.817741	155.926059

United States	10.850494	905.894000

	13-17 Educ Pct Average	13-17 Heal Change \
Country Name		
Australia	5.221362	-15.734346
Brazil	6.132996	-1.924628
China	4.176000	183.364505
France	5.472154	-25.876366
Germany	4.896210	6.656783
Italy	4.027862	-17.670230
Japan	3.406083	-30.940662
Mexico	4.924844	-11.752630
Russia	4.012418	-32.329930
South Korea	5.024880	29.764022
United Kingdom	5.510750	-16.044725
United States	4.994251	598.100667

	13-17 Heal Pct Change	13-17 Heal Average \
Country Name		
Australia	-11.394095	126.059289
Brazil	-0.975676	184.803872
China	40.678218	537.065275
France	-8.049210	302.235934
Germany	1.622425	403.935157
Italy	-9.403570	175.178552
Japan	-5.561026	523.676211
Mexico	-15.522846	68.445868
Russia	-27.762150	89.336049
South Korea	34.752445	99.678494
United Kingdom	-5.771385	280.188133
United States	21.981762	3030.334259

	13-17 Heal Pct Average
Country Name	
Australia	9.105322
Brazil	8.784226
China	4.901009
France	11.474966
Germany	11.147693
Italy	8.780664
Japan	10.828020
Mexico	5.700569
Russia	5.232948
South Korea	6.679254
United Kingdom	9.906462
United States	16.676044

[46]:
```python
china_us = data.loc[['China','United States']]
china_us

#For a case study that we will apply later to make predictions on China and the␣
↪US, we isolate those countries'
#figures into a separate dataframe.
```

[46]:
```
                 2013 GDP      2014 GDP     2015 GDP      2016 GDP  \
Country Name
China          9570.406236  10475.682921  11061.55308  11233.276537
United States  16784.849196  17527.163695  18224.70444  18714.960538

                 2017 GDP      2013 Pop      2014 Pop      2015 Pop  \
Country Name
China          12310.409371  1.357380e+09  1.364270e+09  1.371220e+09
United States  19519.353692  3.159937e+08  3.183010e+08  3.206352e+08

                 2016 Pop      2017 Pop  …  13-17 Mil Average  \
Country Name                             …
China          1.378665e+09  1.386395e+09  …         207.998941
United States  3.229413e+08  3.249855e+08  …         649.491402

               13-17 Mil Pct Average  13-17 Educ Change  \
Country Name
China                       1.903466         111.522049
United States               3.590561          93.417000

               13-17 Educ Pct Change  13-17 Educ Average  \
Country Name
China                      28.011543          456.509856
United States              10.850494          905.894000

               13-17 Educ Pct Average  13-17 Heal Change  \
Country Name
China                        4.176000         183.364505
United States                4.994251         598.100667

               13-17 Heal Pct Change  13-17 Heal Average  \
Country Name
China                      40.678218          537.065275
United States              21.981762         3030.334259

               13-17 Heal Pct Average
Country Name
China                        4.901009
United States               16.676044
```

[2 rows x 76 columns]

[47]:
```
china_us['13-14 GDP Ab'] = china_us['2014 GDP']-china_us['2013 GDP']
china_us['14-15 GDP Ab'] = china_us['2015 GDP']-china_us['2014 GDP']
china_us['15-16 GDP Ab'] = china_us['2016 GDP']-china_us['2015 GDP']
china_us['16-17 GDP Ab'] = china_us['2017 GDP']-china_us['2016 GDP']

china_us['13-14 GDP Pct'] = (china_us['2014 GDP']-china_us['2013 GDP'])/
 ↪china_us['2013 GDP']*100
china_us['14-15 GDP Pct'] = (china_us['2015 GDP']-china_us['2014 GDP'])/
 ↪china_us['2014 GDP']*100
china_us['15-16 GDP Pct'] = (china_us['2016 GDP']-china_us['2015 GDP'])/
 ↪china_us['2015 GDP']*100
china_us['16-17 GDP Pct'] = (china_us['2017 GDP']-china_us['2016 GDP'])/
 ↪china_us['2016 GDP']*100

china_us['13-14 Mil Ab'] = china_us['2014 Mil']-china_us['2013 Mil']
china_us['14-15 Mil Ab'] = china_us['2015 Mil']-china_us['2014 Mil']
china_us['15-16 Mil Ab'] = china_us['2016 Mil']-china_us['2015 Mil']
china_us['16-17 Mil Ab'] = china_us['2017 Mil']-china_us['2016 Mil']

china_us['13-14 Mil Pct'] = (china_us['2014 Mil']-china_us['2013 Mil'])/
 ↪china_us['2013 Mil']*100
china_us['14-15 Mil Pct'] = (china_us['2015 Mil']-china_us['2014 Mil'])/
 ↪china_us['2014 Mil']*100
china_us['15-16 Mil Pct'] = (china_us['2016 Mil']-china_us['2015 Mil'])/
 ↪china_us['2015 Mil']*100
china_us['16-17 Mil Pct'] = (china_us['2017 Mil']-china_us['2016 Mil'])/
 ↪china_us['2016 Mil']*100

#We calculate the absolute and relative changes in GDP and military spending␣
 ↪for both countries.
```

[48]:
```
china_us['13-14 Educ Ab'] = china_us['2014 Educ']-china_us['2013 Educ']
china_us['14-15 Educ Ab'] = china_us['2015 Educ']-china_us['2014 Educ']
china_us['15-16 Educ Ab'] = china_us['2016 Educ']-china_us['2015 Educ']
china_us['16-17 Educ Ab'] = china_us['2017 Educ']-china_us['2016 Educ']

china_us['13-14 Educ Pct'] = (china_us['2014 Educ']-china_us['2013 Educ'])/
 ↪china_us['2013 Educ']*100
china_us['14-15 Educ Pct'] = (china_us['2015 Educ']-china_us['2014 Educ'])/
 ↪china_us['2014 Educ']*100
china_us['15-16 Educ Pct'] = (china_us['2016 Educ']-china_us['2015 Educ'])/
 ↪china_us['2015 Educ']*100
china_us['16-17 Educ Pct'] = (china_us['2017 Educ']-china_us['2016 Educ'])/
 ↪china_us['2016 Educ']*100
```

```
china_us['13-14 Heal Ab'] = china_us['2014 Heal']-china_us['2013 Heal']
china_us['14-15 Heal Ab'] = china_us['2015 Heal']-china_us['2014 Heal']
china_us['15-16 Heal Ab'] = china_us['2016 Heal']-china_us['2015 Heal']
china_us['16-17 Heal Ab'] = china_us['2017 Heal']-china_us['2016 Heal']

china_us['13-14 Heal Pct'] = (china_us['2014 Heal']-china_us['2013 Heal'])/
↪china_us['2013 Heal']*100
china_us['14-15 Heal Pct'] = (china_us['2015 Heal']-china_us['2014 Heal'])/
↪china_us['2014 Heal']*100
china_us['15-16 Heal Pct'] = (china_us['2016 Heal']-china_us['2015 Heal'])/
↪china_us['2015 Heal']*100
china_us['16-17 Heal Pct'] = (china_us['2017 Heal']-china_us['2016 Heal'])/
↪china_us['2016 Heal']*100

#We do the same thing for educational and healthcare spending.
```

[49]:
```
china_us
```

[49]:
```
                    2013 GDP      2014 GDP     2015 GDP      2016 GDP  \
Country Name
China            9570.406236  10475.682921  11061.55308  11233.276537
United States   16784.849196  17527.163695  18224.70444  18714.960538

                    2017 GDP      2013 Pop      2014 Pop      2015 Pop  \
Country Name
China           12310.409371  1.357380e+09  1.364270e+09  1.371220e+09
United States   19519.353692  3.159937e+08  3.183010e+08  3.206352e+08

                    2016 Pop      2017 Pop  …  15-16 Educ Pct  \
Country Name                                …
China           1.378665e+09  1.386395e+09  …        0.598891
United States   3.229413e+08  3.249855e+08  …        2.323337

               16-17 Educ Pct  13-14 Heal Ab  14-15 Heal Ab  15-16 Heal Ab  \
Country Name
China                7.511256      49.259916      40.740481      18.859753
United States        3.085666     154.690087     169.900190     145.225118

               16-17 Heal Ab  13-14 Heal Pct  14-15 Heal Pct  15-16 Heal Pct  \
Country Name
China              74.504356       10.927991        8.147637        3.487582
United States     128.285272        5.685265        5.908369        4.768538

               16-17 Heal Pct
Country Name
China               13.313183
```

```
United States        4.020587

[2 rows x 108 columns]
```

0.5 3. Exporting Our Data

[50]:
```python
data.to_excel(r'valle-individual-project-data.xlsx',index=True)
china_us.to_excel(r'valle-china-us-data.xlsx',index=True)

#We export our dataframes to Excel for our visualizations later.  The smaller␣
↪dataframe will be merged inside the
#larger one in a separate sheet soon.  This Excel file will then be downloaded␣
↪into Google Sheets.
```