Rajwani_Final_Project

May 14, 2018

Ahil Rajwani - Michael Waugh - Data Bootcamp - Final Project

1 *How different causes of death have changed over the years.*

Principal Investigator: Ahil Rajwani **Email:** ahil.rajwani@stern.nyu.edu

"There are two things in life which are completely certain: death and taxes."

Although death is certain, the way in which we die has developed and changed over time due to various factors: GDP per capita, vehicle safety, development of medicine, and the advancement of technology just to name a few. This project will be analyzing how people die in the United States. I will be taking a look at several key correlations:

1. Suicide Rates as a function of GDP per capita over time
2. Deaths as a function of vehicle miles travelled (VMT) over time
3. Concluding correlation analyzing the share of each cause of death over time

My project will include an in depth visualization of these correlations, however, I already have a few hyoptheses about what I am going to find in my analysis and graphical representations:

1. I expect suicides to increase in times of economic distress. In times such as the Great Depression and the Great Recession when things are rough for some members of the population due to the hault in the economy, unfortunately, I predict that there will be a number of people who will resort to ending their life as opposed to living under the poverty line, starving, or being homeless.
2. The amount of vehicle related deaths are expected to go down as we have made changes to vehicle safety features, legislation, etc. But I also expect to see dips in vehicle related deaths when people don't drive as much in times of economic distress when households cannot afford to spend on gas and car maintenance.
3. I predict that diseases of the heart are going to be the leading cause of death over the years and that death by overdose of drugs has also increased in the share of the amount of deaths due to the prevalence of opiate addiction all across America. However, I do expect to see a decrease in death by diseases which have been cured such as Polio.

1.1 Suicide Rates as a function of GDP

1.1.1 Data Compilation and Manipulation

For my first correlation, (Suicide Rates as a function of GDP per capita), suicide rates was a statistic which was pretty difficult to find and some data had to be inputted manually. For the suicide statistics from 1900-1960 I had to input all the data manually from this census document (page 10) and create my own CSV file.

For suicide rates 1960 onwards, that data was a little easier to deal with thankfully. OECD has the historical data via this very comprehensive database (OECD iLibrary). I was able to pick and choose what years and for what causes of death I wanted to pinpoint (suicides) and export as CSV document, which exported as this.

Compiling and cleaning up the suicide statistics:

```
In [294]: import pandas as pd              # data package
          import matplotlib.pyplot as plt # graphics
          %matplotlib inline
          import matplotlib.ticker as plticker #axis manipulation
          import numpy as np #numerical operations and caluclations
          import seaborn as sns #aesthetic
          sns.set_style("darkgrid")
          import geopandas as gpd #for our vehicle fatalities correlation
```

```
In [295]: url1 = "https://raw.githubusercontent.com/ahilly"
          url2 = "/final_project/master/Suicide_Rate_1900_1960.csv"
          ahil_suicide = pd.read_csv(url1+url2) #read in the data found from the Census docume

          url3 = "/final_project/master/HEALTH_STAT_24042018192843590.csv"
          oecd_suicide = pd.read_csv(url1+url3) #read in the data from OECD database

          oecd_suicide = oecd_suicide.drop(["VAR", "Variable", "UNIT", "Measure", "COU","Count
                                           axis=1)
          #removed unnecessary columns/data

          oecd_suicide = oecd_suicide.rename(index=str,
                                             columns={"Value":"Suicide Deaths per 100000"})
          #renamed columns

          suicide = ahil_suicide.append(oecd_suicide, ignore_index = True)
          #append from the two sources into one dataframe
```

Compiling the GDP per capita statistics:

Moving on to the GDP per capita data, FRED only has data since 1947, but I really want to grab the great depression data as I wanted to explore the spike in suicides found in the early 1930s. I found this great website allowing me to grab the GDP per capital data with all of the data in an easily accessible csv file.

```
In [296]: url4 = "/final_project/master/USGDP_1900-2017.csv"
          gdp_per_capita = pd.read_csv(url1+url4, header=1) #first line of the file is a title

          gdp_per_capita.rename(index=str,
                                columns={"Year": "Year", "Real GDP per capita (year 2009 dolla
                                         "GDPC"},
                                inplace = True) #lengthy name
```

```
gdp_per_capita["GDPC"] = gdp_per_capita["GDPC"].str.replace(",","")
#get rid of the commas so we can convert it from an object to float

gdp_per_capita["GDPC"] = pd.to_numeric(gdp_per_capita["GDPC"])
#to_numeric is a great function in the pandas function to convert the objects into f

gdp_per_capita.dtypes
```

Out[296]:
```
Year      int64
GDPC    float64
dtype: object
```

Merge the suicide with the GDP per capita statistics:

In [297]:
```
suicide_gdp = pd.merge(suicide, gdp_per_capita)
suicide_gdp.set_index("Year", inplace = True)
suicide_gdp.head()
```

Out[297]:

Year	Suicide Deaths per 100000	GDPC
1900	10.2	6003.90
1901	10.4	6201.16
1902	10.3	6389.64
1903	11.3	6456.30
1904	12.2	6111.43

Calculate % Change Statistics:

In [298]:
```
suicide_gdp["Suicide % Change"] = suicide_gdp["Suicide Deaths per 100000"].pct_change
#pct_change method in the pandas package makes this easy for us

suicide_gdp["GDPC % Change"] = (suicide_gdp["GDPC"]).pct_change()*100
#do the same for GDP per capita

suicide_gdp.tail() #let's see how it looks
```

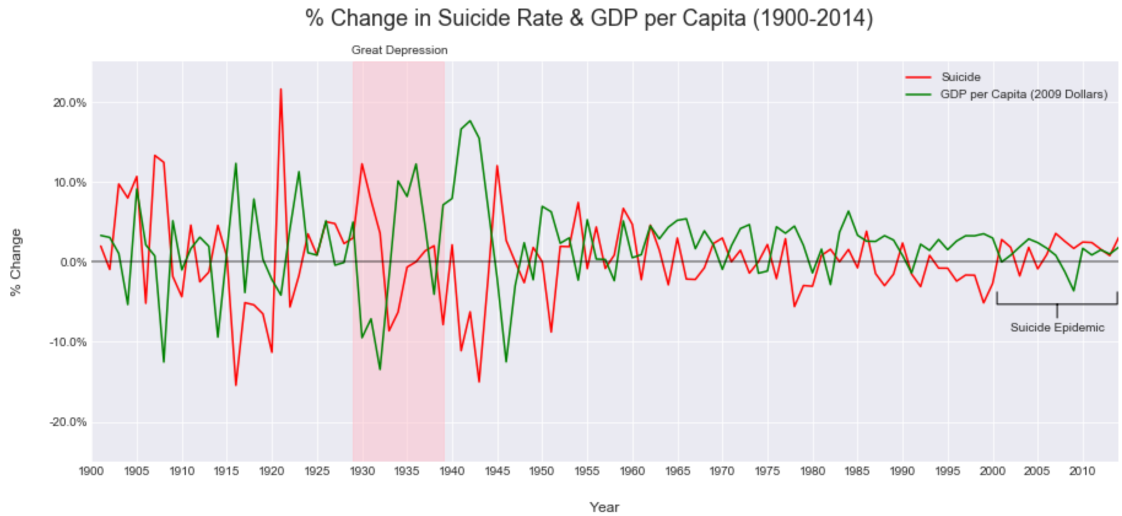
Out[298]:

Year	Suicide Deaths per 100000	GDPC	Suicide % Change	GDPC % Change
2010	12.5	47720.0	2.459016	1.683358
2011	12.8	48125.0	2.400000	0.848701
2012	13.0	48841.0	1.562500	1.487792
2013	13.1	49317.0	0.769231	0.974591
2014	13.5	50216.0	3.053435	1.822901

1.1.2 Data Presentation

I decided to go with a combined line graph mainly to highlight the inverse correlation between % change in suicides and GDP per capita. What's great about this graphical representation of the data, we can see when the opposing spikes, how suicide rates change when GDP per capita changes. When GDP per capita spiking down, we should be able to see Suicide rates spiking up and vice versa.

```
In [299]: fig, ax = plt.subplots(figsize = (15,6))

          #plot both statistics
          suicide_gdp["Suicide % Change"].plot(ax = ax, label = "Suicide", color = "red" )
          suicide_gdp["GDPC % Change"].plot(ax = ax, color = "green", label = "GDP per Capita")

          #axis manipulation
          ax.set_ylabel("% Change", labelpad=20, fontsize = 12)
          ax.set_xlabel("Year", labelpad=20, fontsize = 12)
          ax.set_xlim([1900, 2014])
          ax.set_ylim([-25, 25])
          vals = ax.get_yticks()
          ax.set_yticklabels(['{:}%'.format(x) for x in vals]) #percent sign in the axis
          loc = plticker.MultipleLocator(base=5.0)
          ax.xaxis.set_major_locator(loc) #this locator puts ticks every 5 years so we can rea

          #axis line to better visualize data
          ax.axhline(y=0, color='k', alpha = 0.5)

          #great depression annotation
          ax.axvspan(1929, 1939, color='pink', alpha=0.5)
          ax.annotate("Great Depression",(1928.77,26), xycoords='data',
                      annotation_clip=False, fontsize=10)

          #suicide epidemic annotation
          ax.annotate('Suicide Epidemic', xy=(0.94, .40), xytext=(0.94, .32), xycoords='axes f
                      fontsize=10, ha='center', va='bottom',
                      arrowprops=dict(arrowstyle='-[, widthB=4.9, lengthB=1', lw=1))

          #misc
          fig.suptitle("% Change in Suicide Rate & GDP per Capita (1900-2014)",
                       fontsize = 18, weight="light")
          ax.legend(["Suicide", "GDP per Capita (2009 Dollars)"], loc = "best")

Out[299]: <matplotlib.legend.Legend at 0x1a32410c18>
```



1.1.3 Conclusions

As we can see from the graphic above, the trend is, when GDP per capita spikes up, suicide rate spikes down and vice versa. We especially see this in the Great Depression, which I highlighted above. GDP per capita, took a heavy hit and the suicide rate spiked from 12.8 before the Great Depression to 18.1 per 100,000 people. On "Black Thursday" alone, October 24th, 1929, over 100 people attempted to or committed suicide in the NYC metropolitan area. Surprisingly enough, through further research, it was found that suicides were prominent in both sides of the financial scale, rich and poor. The change from high to low estate provoked destruction more so than poverty itself.

Following the great depression, we see GDP per capita spike up due to WWII and as a result we see suicide rates drop as well to 14.2 per 100,000 people. And then following in 1940s and 1950s we see the cyclical movements of the economy and the suicide rates seem to match the trends perfectly. Decided to run a pair-wise correlation on our dataframe to display how suicides and GDP per capita are inversely correlated:

In [300]:
```
#Pair-wise correlation table
suicide_gdp.corr("pearson")
```

Out[300]:
```
                          Suicide Deaths per 100000      GDPC  \
Suicide Deaths per 100000                  1.000000 -0.293124
GDPC                                      -0.293124  1.000000
Suicide % Change                           0.179273 -0.039365
GDPC % Change                             -0.210215 -0.008436

                          Suicide % Change  GDPC % Change
Suicide Deaths per 100000         0.179273      -0.210215
GDPC                             -0.039365      -0.008436
Suicide % Change                  1.000000      -0.498693
GDPC % Change                    -0.498693       1.000000
```

Right now, we are seeing a 50% inverse correlation between GDPC Change % and Suicide % Change, meaning that 50% of the change in suicide rate can be explained by the change in GDP. As we can conclude from this analysis, there is a high correlation between GDP per Capita and Suicide rate. Although, change in GDP doesn't explain all of the suicides, it does shed some light into how income level, or changes in income level can cause some people to contemplate suicide. Following the 40s and 50s, it seems the inverse relationship starts to get a little messy and something other than GDP per capita must be affecting the suicide rate. This brings us to the suicide epidemic, which is actually happening at this very moment. Suicide rates have have risen in every age group except the elderly, according to the National Center for Health Statistics. Among women 45 to 64 it jumped an astounding 63 percent. For men that age, it was up 43 percent. There are several reasons for this epidemic: early deaths caused by drugs and alchohol, deteriorating job prospects, loneliness & the rise of social media, as well as lack of stable relationships.

1.2 Deaths as a function of Vehicle Miles Travelled over Time

1.2.1 Data Cleaning and Manipulation

As for the the next correlation, car-related deaths over time, there is a database online called the Fatality Analysis Reporting System (FARS) run by the National Highway Traffic Safety Administration (NHTSA) which is similar to the OECD iLibrary allowing me to pick certain parameters over a given time and allowing me to export into a CSV file which I can analyze. For this analysis, I will be studying how vehicle deaths have decreased over time, and how that affected the total amount of deaths. For instance, in times like the energy crisis, or great recession, will we see a decrease in the amount of vehicle related deaths?

Data from the FARS database exported as this.

VMT, Vehicle Fatalities, Fatalities per 100 VMT:

```
In [301]: url4 = "/final_project/master/Vehicle%20Fatalities%20and%20Fatality%20Rates%201899-20
          vehicle_fatality = pd.read_csv(url1 + url4, header = 5) #read in the data from NHTSA
          vehicle_fatality = vehicle_fatality.drop(["Unnamed: 4", "Unnamed: 5", "Millions.1", "
          vehicle_fatality.columns = ["Year", "Fatalities", "VMT (Millions)", "Fatalities per 1
          vehicle_fatality.set_index("Year", inplace=True)

          #only want to analyze data where VMT is available
          x = 1899
          while x < 1921:
              vehicle_fatality = vehicle_fatality.drop([x])
              x = x + 1

          vehicle_fatality.tail()
```

Out[301]:

Year	Fatalities	VMT (Millions)	Fatalities per 100m VMT
2011	32,479	2,950,402	1.10
2012	33,782	2,969,433	1.14
2013	32,893	2,988,280	1.10
2014	32,744	3,025,656	1.08
2015	35,092	3,095,373	1.13

Total Population:

Received historical population data via the US Census database here and the data I compiled exported as this.

```
In [302]: url5 = "/final_project/master/Population%20Statistics.csv"
          population = pd.read_csv(url1 + url5) #read in the population data
          population.set_index("Year", inplace = True)
          x = 1900
          while x < 1921:
              population = population.drop([x])
              x = x + 1
          population = population.drop([2016,2017])
```

```
In [303]: vehicle_fatality.dtypes
```

```
Out[303]: Fatalities                    object
          VMT (Millions)                object
          Fatalities per 100m VMT     float64
          dtype: object
```

Data Manipulation:

We aren't quite ready yet to do our analysis, there are a lot of different metrics and data we are analyzing overtime so we need to scale the data correctly so the dataframe can work graphically.

```
In [304]: vehicle_fatality = vehicle_fatality.rename({"Fatalities per 100m VMT": "Deaths per bi
          vehicle_fatality["Deaths per billion VMT"] = vehicle_fatality["Deaths per billion VMT


          vehicle_fatality = vehicle_fatality.rename({"VMT (Millions)": "VMT (10s of Billions)"
          vehicle_fatality["VMT (10s of Billions)"] = vehicle_fatality["VMT (10s of Billions)"]
          #realized this column isn't in numeric form sp get rid of the commas so we can conve

          vehicle_fatality["VMT (10s of Billions)"] = pd.to_numeric(vehicle_fatality["VMT (10s
          #convert to numeric form

          vehicle_fatality["VMT (10s of Billions)"] = vehicle_fatality["VMT (10s of Billions)"]
          vehicle_fatality.head()
```

Out[304]:

Year	Fatalities	VMT (10s of Billions)	Deaths per billion VMT
1921	13,253	5.5027	240.8
1922	14,859	6.7697	219.5
1923	17,870	8.4995	210.2
1924	18,400	10.4838	175.5
1925	20,771	12.2346	169.8

We are making one plot with two y-axes so we need to group the dataframes in the right way, all of the dataframes with the same scale on one dataframe and then the rest in another.

In [305]:
```
vehicle_population = pd.concat([vehicle_fatality, population], axis=1)

#we want to plot fatalities on its own axis
fatalities = vehicle_fatality.drop(["VMT (10s of Billions)","Deaths per billion VMT"

#we want to plot the rest of the metrics on another
vehicle_population = vehicle_population.drop(["Fatalities"],axis=1)

#convert all the variables to the necessary data types
vehicle_population["Population"] = vehicle_population["Population"].str.replace("mil
vehicle_population["Population"] = pd.to_numeric(vehicle_population["Population"])

fatalities["Fatalities"] = fatalities["Fatalities"].str.replace(",","")
fatalities["Fatalities"] = pd.to_numeric(fatalities["Fatalities"])
vehicle_population.dtypes
```

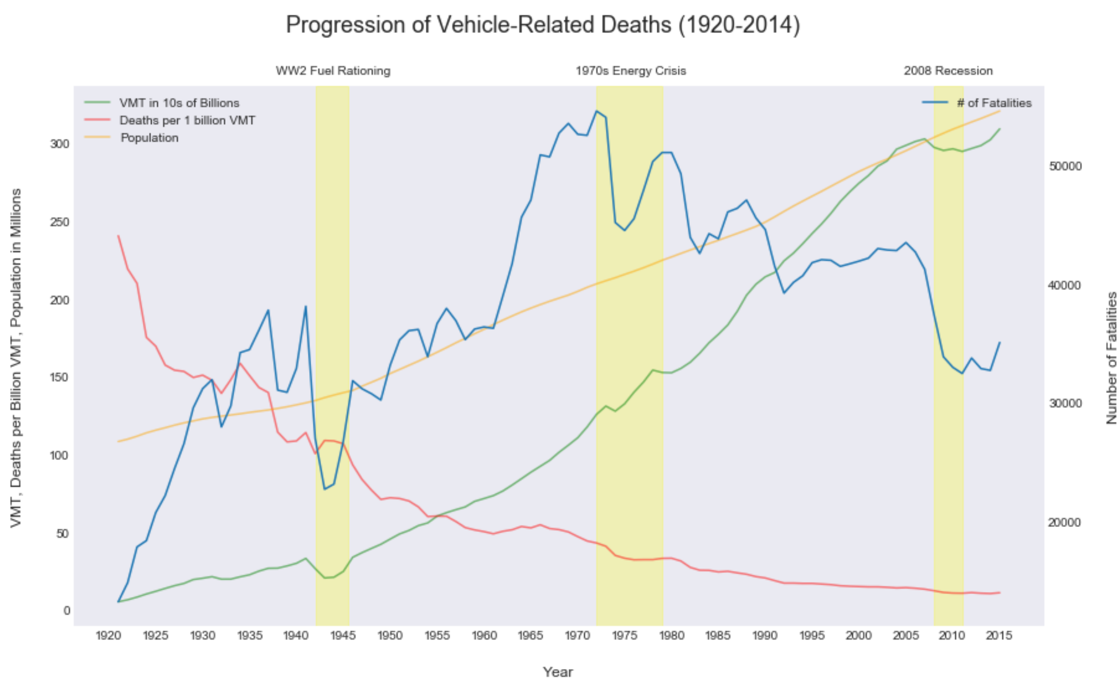
Out[305]:
```
VMT (10s of Billions)     float64
Deaths per billion VMT    float64
Population                float64
dtype: object
```

1.2.2 Data Presentation

In [306]:
```
sns.set_style("dark")#change the style for this one, the grids get messy

fig, ax1 = plt.subplots(figsize=(14,8))

#plot vmt, deaths per billion VMT,
ax1.plot(vehicle_population["VMT (10s of Billions)"],color = "green", alpha = 0.5)
ax1.plot(vehicle_population["Deaths per billion VMT"],color = "red", alpha = 0.5)
ax1.plot(vehicle_population["Population"],color = "orange", alpha = 0.5)
ax1.set_ylabel("VMT, Deaths per Billion VMT, Population in Millions", labelpad=20, fo
ax1.set_xlabel("Year", labelpad=20, fontsize = 12)


#plot # of fatalities
ax2 = ax1.twinx()
ax2.plot(fatalities)
ax2.set_ylabel("Number of Fatalities", labelpad = 20, fontsize = 12)

#year frequency
loc1 = plticker.MultipleLocator(base=5.0)
ax1.xaxis.set_major_locator(loc1)

#highlights/annotations
ax1.axvspan(1942, 1945.5, color='Yellow', alpha=0.25)
ax1.annotate("WW2 Fuel Rationing",(1937.77,345), xycoords='data',
            annotation_clip=False, fontsize=10)
```

```
ax1.axvspan(1972, 1979, color='Yellow', alpha=0.25)
ax1.annotate("1970s Energy Crisis",(1969.77,345), xycoords='data',
            annotation_clip=False, fontsize=10)
ax1.axvspan(2008, 2011, color='Yellow', alpha=0.25)
ax1.annotate("2008 Recession",(2004.77,345), xycoords='data',
            annotation_clip=False, fontsize=10)

#misc
fig.suptitle("Progression of Vehicle-Related Deaths (1920-2014)",
             fontsize = 18, weight="light")
ax1.legend(["VMT in 10s of Billions", "Deaths per 1 billion VMT","Population" ], loc
ax2.legend(["# of Fatalities" ], loc = "best")
```

Out[306]: <matplotlib.legend.Legend at 0x1a3266a080>



1.2.3 Conclusions

There is a lot going on this graph, I will admit. However, the most interesting part about this analysis are the shaded areas. According to my hypothesis, I was half right. People did die from cars less due to financial crises causing them to have lower VMT number, however, I completely forgot about the WW2 Fuel Rationing and the 1970s Energy Crisis.

Nearly a year after the Japanese raid on Pearl Harbor that brought the United States fully into World War II, the Americans get around to imposing nationwide gasoline rationing in order to fulfill the United State's rubber shortage. The 1970s energy crisis was when when the major industrial countries of the world faced substantial petroleum shortages, real and perceived, as well as elevated prices due to shocks in the Middle East.

As a result of both these events, you see a steep decline in VMT (Green), Fatalaties (Blue), and a slight decline in Deaths per Billion VMT (Red). However, another interesting point, that in the long run, although the US Population is increasing (Yellow) and VMT is increasing (Green), we are seeing a decline in fatalities (Blue) which shows us that car safety has improved, and the increased regulation advocating for car safety has kept people safer from vehicle accidents.

1.3 Cause of death over time

1.3.1 Data Cleaning and Manipulation

For this particular analysis, I will be using data from OECD's STAT Database, they only have data going back to 1964, however, that is fine enough for this statistical analysis as a lot has changed over the the past 50 years in regards to cause of mortality. It exported as this file, easy to deal with.

```
In [307]: url6 = "/final_project/master/Cause%20of%20Mortality%20-%20OECD.csv"
          mortality = pd.read_csv(url1+url6)
          mortality.set_index("Year", inplace=True)
          mortality.dtypes
```

```
Out[307]: All causes of death                                         int64
          Certain infectious and parasitic diseases                   int64
          Diseases of the blood and blood-forming organs              int64
          Endocrine, nutritional and metabolic diseases               int64
          Mental and behavioural disorders                            int64
          Diseases of the nervous system                              int64
          Diseases of the circulatory system                          int64
          Diseases of the respiratory system                          int64
          Diseases of the digestive system                            int64
          Certain conditions originating in the perinatal period      int64
          Congenital malformations and chromosomal abnormalities      int64
          Accidents                                                   int64
          Intentional self-harm                                       int64
          dtype: object
```

Put everything in terms of percentages/shares:

```
In [308]: mortality["Certain infectious and parasitic diseases"] = mortality["Certain infectio
          mortality["Diseases of the blood and blood-forming organs"] = mortality["Diseases of
          mortality["Endocrine, nutritional and metabolic diseases"] = mortality["Endocrine, nu
          mortality["Mental and behavioural disorders"] = mortality["Mental and behavioural dis
          mortality["Diseases of the nervous system"] = mortality["Diseases of the nervous sys
          mortality["Diseases of the circulatory system"] = mortality["Diseases of the circula
          mortality["Diseases of the respiratory system"] = mortality["Diseases of the respira
          mortality["Diseases of the digestive system"] = mortality["Diseases of the digestive
          mortality["Certain conditions originating in the perinatal period"] = mortality["Cer
          mortality["Congenital malformations and chromosomal abnormalities"] = mortality["Con
          mortality["Accidents"] = mortality["Accidents"]/mortality["All causes of death"]*100
          mortality["Intentional self-harm"] = mortality["Intentional self-harm"]/mortality["Al
          #the divide function was not working for these, I'm sure there is a better way to do
```

```
In [309]: mortality

Out[309]:       All causes of death  Certain infectious and parasitic diseases  \
          Year
          1960              1711982                                   1.243763
          1961              1701522                                   1.173773
          1962              1756720                                   1.120383
          1963              1813549                                   1.083070
          1964              1798051                                   1.024665
          1965              1828136                                   0.972302
          1966              1863149                                   0.926120
          1967              1851323                                   0.875482
          1968              1930082                                   0.920997
          1969              1921990                                   0.873626
          1970              1921031                                   0.866670
          1971              1927542                                   0.831473
          1972              1963944                                   0.795237
          1973              1973003                                   0.794170
          1974              1934388                                   0.812764
          1975              1892879                                   0.841575
          1976              1909440                                   0.866903
          1977              1899597                                   0.891242
          1978              1927788                                   0.935891
          1979              1913841                                   0.802575
          1980              1989841                                   0.868813
          1981              1977981                                   0.936005
          1982              1974797                                   0.986177
          1983              2019201                                   1.080824
          1984              2039369                                   1.201352
          1985              2086440                                   1.342143
          1986              2105361                                   1.502023
          1987              2123323                                   2.030355
          1988              2167999                                   2.218313
          1989              2150466                                   2.457653
          1990              2148463                                   2.588455
          1991              2169518                                   2.783936
          1992              2175613                                   2.970105
          1993              2268553                                   3.091971
          1994              2278994                                   3.304089
          1995              2312132                                   3.335796
          1996              2314690                                   2.842324
          1997              2314245                                   2.262984
          1998              2337256                                   2.192828
          1999              2391399                                   2.508323
          2000              2403351                                   2.455197
          2001              2416425                                   2.496912
          2002              2443387                                   2.618537
          2003              2448288                                   2.641029
```

2004	2397615	2.695095
2005	2448017	2.716770
2006	2426299	2.761119
2007	2423895	2.758205
2008	2471984	2.768303
2009	2437942	2.853677
2010	2469288	2.737307
2011	2515458	2.735804
2012	2543279	2.689795
2013	2596993	2.718567
2014	2626418	2.680952

	Diseases of the blood and blood-forming organs \
Year	
1960	0.194511
1961	0.192769
1962	0.193429
1963	0.187974
1964	0.195545
1965	0.193585
1966	0.185278
1967	0.186893
1968	0.259471
1969	0.256037
1970	0.262411
1971	0.269203
1972	0.266403
1973	0.266143
1974	0.276780
1975	0.278359
1976	0.275788
1977	0.283586
1978	0.289140
1979	0.324008
1980	0.320428
1981	0.321894
1982	0.332287
1983	0.354695
1984	0.358199
1985	0.368618
1986	0.369390
1987	0.382278
1988	0.383533
1989	0.399448
1990	0.420161
1991	0.429266
1992	0.434222
1993	0.427982

```
1994                                           0.426153
1995                                           0.441022
1996                                           0.432585
1997                                           0.442952
1998                                           0.455106
1999                                           0.379150
2000                                           0.387584
2001                                           0.394591
2002                                           0.395394
2003                                           0.391457
2004                                           0.393016
2005                                           0.391746
2006                                           0.364094
2007                                           0.402039
2008                                           0.399396
2009                                           0.395662
2010                                           0.393555
2011                                           0.401796
2012                                           0.409511
2013                                           0.391453
2014                                           0.406942

      Endocrine, nutritional and metabolic diseases  \
Year
1960                                          2.511417
1961                                          2.504581
1962                                          2.489071
1963                                          2.499684
1964                                          2.519895
1965                                          2.524867
1966                                          2.552077
1967                                          2.578264
1968                                          2.401660
1969                                          2.421605
1970                                          2.408030
1971                                          2.394915
1972                                          2.390598
1973                                          2.358435
1974                                          2.361884
1975                                          2.296238
1976                                          2.243956
1977                                          2.175303
1978                                          2.204236
1979                                          2.287964
1980                                          2.329834
1981                                          2.316049
1982                                          2.331531
1983                                          2.420859
```

1984	2.429379
1985	2.553057
1986	2.685763
1987	3.099434
1988	3.315085
1989	3.941843
1990	4.131512
1991	4.396276
1992	4.645909
1993	4.845952
1994	5.189965
1995	5.308304
1996	4.920616
1997	4.385145
1998	4.346293
1999	3.883584
2000	3.925561
2001	4.006746
2002	4.070456
2003	4.102336
2004	4.129646
2005	4.212552
2006	4.088820
2007	4.085903
2008	4.018068
2009	4.022286
2010	4.007714
2011	4.179915
2012	4.164938
2013	4.203284
2014	4.292424

	Mental and behavioural disorders	Diseases of the nervous system \
Year		
1960	0.262094	0.860581
1961	0.262412	0.836839
1962	0.249044	0.834965
1963	0.256734	0.802791
1964	0.263897	0.816161
1965	0.272846	0.812631
1966	0.299386	0.811261
1967	0.305889	0.795971
1968	0.343405	0.851725
1969	0.349742	0.880390
1970	0.373341	0.891709
1971	0.436566	0.829450
1972	0.441866	0.855523
1973	0.449062	0.856005

1974	0.496436	0.897235
1975	0.508326	0.883786
1976	0.505803	0.878582
1977	0.522532	0.918774
1978	0.566349	0.939574
1979	0.663796	1.072921
1980	0.704880	1.092751
1981	0.736761	1.141113
1982	0.733037	1.161335
1983	0.788282	1.289124
1984	0.813732	1.387978
1985	0.891375	1.483723
1986	0.933949	1.565765
1987	1.006630	1.690982
1988	1.107842	1.707473
1989	1.171467	1.789659
1990	1.203465	1.832380
1991	1.251384	1.844142
1992	1.333279	1.858924
1993	1.477418	1.983820
1994	1.623655	2.101805
1995	1.761059	2.216136
1996	1.901335	2.325193
1997	2.051684	2.430382
1998	2.259872	2.471360
1999	1.761647	3.520742
2000	1.914618	3.792205
2001	2.102734	3.999131
2002	2.283224	4.222540
2003	2.453061	4.482724
2004	2.566050	4.665261
2005	2.919342	4.935137
2006	3.752011	5.072664
2007	3.724543	5.244080
2008	4.232107	5.531508
2009	4.357897	5.508088
2010	4.893070	5.719543
2011	5.404543	5.766187
2012	5.815996	5.752849
2013	6.022927	5.848995
2014	5.751103	6.302272

	Diseases of the circulatory system	Diseases of the respiratory system \
Year		
1960	53.951210	5.505899
1961	54.340878	4.858944
1962	54.501400	5.137074
1963	54.230903	5.810099

1964	54.124271	5.225825
1965	54.164023	5.460589
1966	54.252880	5.564719
1967	54.129452	5.280710
1968	53.898850	6.395583
1969	53.557146	6.059189
1970	53.330790	5.860291
1971	53.633643	5.606519
1972	53.614869	5.913407
1973	53.481723	5.957264
1974	53.173148	5.650521
1975	52.199480	5.915962
1976	51.925538	6.342435
1977	51.512347	5.815760
1978	50.945540	6.331453
1979	50.319593	5.828436
1980	49.920974	6.474286
1981	49.428887	6.639548
1982	49.248201	6.469222
1983	48.837733	7.073838
1984	47.964395	7.339819
1985	47.101235	7.965194
1986	46.211410	8.119178
1987	45.609359	8.139788
1988	44.924790	8.625742
1989	43.532751	8.696162
1990	42.832713	8.946163
1991	42.427949	9.008130
1992	42.190776	8.931368
1993	41.974333	9.346046
1994	41.455791	9.281420
1995	41.329431	9.302280
1996	41.232606	9.520411
1997	40.981486	9.849519
1998	40.430402	10.225624
1999	39.907059	9.612072
2000	39.175468	9.614867
2001	38.362416	9.518566
2002	37.789306	9.602613
2003	37.053525	9.636285
2004	36.113471	9.471287
2005	35.164870	9.790537
2006	34.115993	9.341882
2007	33.430532	9.378294
2008	32.707534	9.915072
2009	32.197238	9.790389
2010	31.777743	9.616821
2011	31.146415	9.888537

2012	30.961251	9.684781
2013	30.840206	10.052819
2014	30.755196	9.835944

	Diseases of the digestive system \
Year	
1960	3.883452
1961	3.955459
1962	4.004622
1963	3.954566
1964	3.948164
1965	3.932968
1966	3.956581
1967	4.008593
1968	3.819423
1969	3.846326
1970	3.873285
1971	3.851797
1972	3.807033
1973	3.806076
1974	3.783884
1975	3.718251
1976	3.640334
1977	3.618188
1978	3.513301
1979	3.769122
1980	3.779297
1981	3.746396
1982	3.678454
1983	3.596522
1984	3.619404
1985	3.585294
1986	3.531366
1987	3.553440
1988	3.521911
1989	3.539744
1990	3.499106
1991	3.482709
1992	3.493084
1993	3.383038
1994	3.397639
1995	3.324248
1996	3.345113
1997	3.416622
1998	3.395306
1999	3.465586
2000	3.495744
2001	3.552066

```
2002                       3.548762
2003                       3.583770
2004                       3.599619
2005                       3.568848
2006                       3.610808
2007                       3.678336
2008                       3.685178
2009                       3.669242
2010                       3.699487
2011                       3.740710
2012                       3.746502
2013                       3.788343
2014                       3.845123

     Certain conditions originating in the perinatal period  \
Year
1960                                      3.919083
1961                                      3.860015
1962                                      3.654823
1963                                      3.456648
1964                                      3.354855
1965                                      3.030300
1966                                      2.771866
1967                                      2.609701
1968                                      2.271406
1969                                      2.246162
1970                                      2.249053
1971                                      1.997103
1972                                      1.720925
1973                                      1.546019
1974                                      1.488119
1975                                      1.406112
1976                                      1.299281
1977                                      1.231893
1978                                      1.142916
1979                                      1.225180
1980                                      1.149137
1981                                      1.093337
1982                                      1.052969
1983                                      0.956319
1984                                      0.925826
1985                                      0.922432
1986                                      0.873532
1987                                      0.858183
1988                                      0.840406
1989                                      0.871997
1990                                      0.822635
1991                                      0.773444
```

```
1992                                                   0.723015
1993                                                   0.665490
1994                                                   0.635675
1995                                                   0.582622
1996                                                   0.564611
1997                                                   0.565670
1998                                                   0.574520
1999                                                   0.596262
2000                                                   0.585391
2001                                                   0.574692
2002                                                   0.583411
2003                                                   0.589718
2004                                                   0.596134
2005                                                   0.597259
2006                                                   0.595228
2007                                                   0.602336
2008                                                   0.563636
2009                                                   0.533237
2010                                                   0.491235
2011                                                   0.478481
2012                                                   0.468568
2013                                                   0.465307
2014                                                   0.452936

      Congenital malformations and chromosomal abnormalities  Accidents  \
Year
1960                                                   1.276883   5.479380
1961                                                   1.288376   5.421558
1962                                                   1.206339   5.529566
1963                                                   1.147860   5.550939
1964                                                   1.128333   5.775309
1965                                                   1.067317   5.907876
1966                                                   0.974587   6.095218
1967                                                   0.935979   6.112872
1968                                                   0.870067   5.951250
1969                                                   0.884916   6.055443
1970                                                   0.875780   5.967525
1971                                                   0.827842   5.885164
1972                                                   0.748392   5.878375
1973                                                   0.712721   5.870290
1974                                                   0.699239   5.408532
1975                                                   0.699728   5.443031
1976                                                   0.680933   5.276992
1977                                                   0.683461   5.432837
1978                                                   0.672688   5.475758
1979                                                   0.706746   5.364604
1980                                                   0.700458   5.190415
1981                                                   0.683879   4.962383
```

1982	0.688881	4.633236
1983	0.652387	4.448690
1984	0.639364	4.427007
1985	0.612670	4.342564
1986	0.600277	4.379676
1987	0.580835	4.332266
1988	0.588238	4.338655
1989	0.599730	4.279817
1990	0.609040	4.149385
1991	0.580728	3.996786
1992	0.574137	3.858775
1993	0.548543	3.861404
1994	0.527864	3.889830
1995	0.514287	3.909898
1996	0.511343	3.964937
1997	0.514725	3.990632
1998	0.510599	4.035972
1999	0.434599	4.092165
2000	0.440135	4.073479
2001	0.432126	4.201951
2002	0.437385	4.368608
2003	0.429606	4.463405
2004	0.437852	4.671809
2005	0.425242	4.812426
2006	0.432305	5.041629
2007	0.430052	5.105172
2008	0.416184	4.931343
2009	0.404275	4.852125
2010	0.391773	4.902344
2011	0.385695	5.026441
2012	0.378881	5.024694
2013	0.369004	5.027237
2014	0.365860	5.180173

	Intentional self-harm
Year	
1960	1.112220
1961	1.116589
1962	1.150269
1963	1.148301
1964	1.145018
1965	1.176444
1966	1.142206
1967	1.151879
1968	1.107310
1969	1.163586
1970	1.222260
1971	1.249882

1972	1.273152
1973	1.273085
1974	1.327707
1975	1.429727
1976	1.405229
1977	1.509847
1978	1.415820
1979	1.421539
1980	1.350309
1981	1.395160
1982	1.430122
1983	1.401297
1984	1.436032
1985	1.411639
1986	1.467872
1987	1.450368
1988	1.402538
1989	1.405835
1990	1.438517
1991	1.420131
1992	1.401168
1993	1.371006
1994	1.366480
1995	1.353037
1996	1.335082
1997	1.319437
1998	1.308158
1999	1.220206
2000	1.219922
2001	1.264223
2002	1.293082
2003	1.283427
2004	1.349800
2005	1.330015
2006	1.368339
2007	1.424690
2008	1.455066
2009	1.511111
2010	1.551540
2011	1.567985
2012	1.593651
2013	1.581059
2014	1.625941

1.3.2 Data Presentation

1.3.3 Conclusions

1.4 Concluding Remarks

I chose deaths because it's an interesting topic to study and the data for most health statistics are always readily available by the OECD, WHO, or some other government/non-government organization. I couldn't be more wrong, some of this data was extremely hard to find, I was shocked to see that these organizations only started keeping track of these statistics electronically in the 1990s. This project was particular difficult because we had to do data manipulation for three different correlations, however, a learned a lot throughout the proccess and am glad I went through the troubleshooting on my own. Now on to some interesting findings from my research and analysis:

When grabbing data, I like to put it all on my github (ahilly) so its easy to view all in one place, so if you want to reference any of my data easily, just go on my github.

For the suicide correlation, I am thinking now that it would be better to do percent change in GDP per capita and percent change in suicide. I know this correlation I am going for is going to be flawed for a number of reasons but I can note those control variable in my analysis. The data is all here now, I just have to figure out how exactly I want to use it.

The OECD library really helped me out! And will especially make the last analysis/correlation a lot easier to do.

That being said, there is still a lot of work to do and I will keep you updated if I have any questions or require any assistance with the project. Also, please let me know if you have any questions for me! Thanks.