

```
* Ani Laliashvili;  
* Suicide Rates Dataset: https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016;  
  
* Description and Conclusions:  
* This analysis uses the suicide rates data which contains information on the number  
of suicides in 101 countries each year between 1985-2016, broken down by gender and the age group of victims,  
in addition to the population, the suicide rate per 100,000 people and the associated generation for each group,  
as well as the HDI index, GDP and GDP per capita for each country and year. Our analysis focuses on data from Argentina,  
United States, United Kingdom, Belgium, Brazil, Greece, Mexico and Singapore and aims to summarize the highest  
risk factors for suicide and relationships between national indices and suicide rates.
```

To conduct the analysis we first print 20 observations from the data, as well as the frequencies of observations for each country and year. Then we go on to analyze suicides per 100,000 people over time for all age groups and sexes. We find that, for each country, the suicide rates are mostly stable over time. The boxplot of suicides per 100,000 people shows that the mean suicide rates are different for each country - Belgium has the highest 20% suicide rate, while Greece has the lowest - under 5%. The suicide rate for all countries seems to not show a significant upward or downward trend.

When plotting the suicide rates by sex we see a clear relationship - males have had higher suicide rates (around 15%) throughout the period in question, while the suicide rates for females have been significantly lower - under 5%.

Individuals aged 75 and up have had the highest suicide rates throughout the period across all countries, but the rate has gone down significantly since the 1990s. On the other end, the lowest suicide rate is observed among children aged 5-14. These observations hold true for every country on a national level except for the United Kingdom. In the U.K., individuals aged 35-54 are most likely to commit suicide. We also observe that over their lifetime the generation with the highest suicide rates across all nations is the Greatest Generation (G.I. Generation).

There is no clear relationship between GDP per capita and suicide rates across nations, but on a national level we see either upward (Mexico, Brazil), stable (Greece, Argentina, United States) or downward trend (Belgium, Singapore, United Kingdom) as the country's GDP per capita increases.

Finally, from the correlation coefficients we can see that population is most highly correlated with the number of suicides, but the GDP per capita and HDI are also positively correlated with it even though the relationship is not as strong in this case.

Overall, the biggest risk factor for suicide seems to be male sex and age of 75+ in an individual. The risk increases with being a member of the Greatest Generation and being an inhabitant of Belgium. Additionally, population, HDI and GDP per capita are positively correlated with the number of suicides in the country.

```
* define library;  
libname data '/folders/myfolders/data';
```

```
* import dataset;
proc import datafile='/folders/myfolders/data/master.csv'
  dbms=csv replace out=data.suicide;
  guessingrows=max;

proc sort data=data.suicide;
  by country year sex;
run;

* View dataset;
proc print data=data.suicide (obs=20);
  by country year;
run;

* View country frequency;
proc freq data=data.suicide order=freq;
  title Frequency of Observations by Country;
  tables country / nopercnt nocum;
run;

* Continue analysis on select countries with the same frequencies, but check that there are no missing
values for any year and country;
proc tabulate data=data.suicide;
  title Frequency of Observations by Country and Year;
  class year country;
  tables year, country*n;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");
  keylabel n="Freq";
run;

* Analyze suicides by country and year;
proc report data=data.suicide out=work.plot2;
  title Suicides by Year;
  column year country suicides_no population suicides_100k_pop_tot;
  define year/group;
  define country/group;
  define suicides_no/analysis format=comma12.;
  define population/analysis format=comma12.;
  define suicides_100k_pop_tot/computed;
  compute suicides_100k_pop_tot;
    suicides_100k_pop_tot = _c3_ * 100000/_c4_;
  endcomp;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
```

```
        "Greece", "Mexico", "Singapore");
run;

proc sgplot data=work.plot2;
    title Suicides by Year and Country;
    series x=year y=suicides_100k_pop_tot/ group=country;
    yaxis label = 'Suicides Per 100,000 People';
    xaxis label = 'Year';
run;

proc sgplot data=work.plot2;
    title Suicides by Country;
    vbox suicides_100k_pop_tot/ group=country;
    yaxis label = 'Suicides Per 100,000 People';
run;

proc sgplot data=work.plot2;
    title Suicides Per Year;
    vbox suicides_100k_pop_tot/ group=year;
    yaxis label = 'Suicides Per 100,000 People';
run;

* Analyze suicides by year and sex;
proc report data=data.suicide out=work.plot3;
    title Suicides by Year and Sex;
    column sex year suicides_no population suicides_100k_pop_tot;
    define sex/group;
    define year/group;
    define suicides_no/analysis format=comma12.;
    define population/analysis format=comma12.;
    define suicides_100k_pop_tot/computed;
    compute suicides_100k_pop_tot;
        suicides_100k_pop_tot = _c3_ * 100000/_c4_;
    endcomp;
    where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
        "Greece", "Mexico", "Singapore");
run;

proc sgplot data=work.plot3;
    title Suicides by Year and Sex;
    series x=year y=suicides_100k_pop_tot/ group=sex;
    yaxis label = 'Suicides Per 100,000 People';
    xaxis label = 'Year';
run;
```

---

\* Analyze suicides by age;

```
proc report data=data.suicide out=work.plot4;  
  title Suicides by Age and Year;  
  column age year suicides_no population suicides_100k_pop_tot;  
  define age/group;  
  define year/group;  
  define suicides_no/analysis format=comma12.;  
  define population/analysis format=comma12.;  
  define suicides_100k_pop_tot/computed;  
  compute suicides_100k_pop_tot;  
    suicides_100k_pop_tot = _c3_ * 100000/_c4_;  
  endcomp;  
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",  
    "Greece", "Mexico", "Singapore");  
run;
```

---

```
proc sgplot data=work.plot4;  
  title Suicides per 100,000 people by Year and Age;  
  series x=year y=suicides_100k_pop_tot/ group=age;  
  yaxis label = 'Suicides Per 100,000 People';  
  xaxis label = 'Year';  
run;
```

---

\* Analyze suicides by country and age;

```
proc report data=data.suicide out=work.plot5;  
  title Suicides by Country and Age;  
  column country age suicides_no population suicides_100k_pop_tot;  
  define age/group;  
  define country/group;  
  define suicides_no/analysis format=comma12.;  
  define population/analysis format=comma12.;  
  define suicides_100k_pop_tot/computed;  
  compute suicides_100k_pop_tot;  
    suicides_100k_pop_tot = _c3_ * 100000/_c4_;  
  endcomp;  
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",  
    "Greece", "Mexico", "Singapore");  
run;
```

---

```
proc sgpanel data=work.plot5;  
  panelby country / novarname;  
  vbar age / response=suicides_100k_pop_tot categoryorder=respasc;  
  label suicides_100k_pop_tot = 'Suicides Per 100,000 People';  
run;
```

```

label age = 'Age';
run;

* Analyze generational differences in suicide;
proc report data=data.suicide out=work.plot7;
  title Suicides by Generation;
  column country generation suicides_no population suicides_100k_pop_tot;
  define country/group;
  define generation/group;
  define suicides_no/analysis sum format=comma12. 'suicides_no_sum';
  define population/analysis sum format=comma12. 'population_sum';
  define suicides_100k_pop_tot/computed;
  compute suicides_100k_pop_tot;
    suicides_100k_pop_tot = _c3_ * 100000/_c4_;
  endcomp;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");

proc sgpanel data=work.plot7;
  panelby country / novarname;
  vbar generation / response=suicides_100k_pop_tot categoryorder=respasc;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");
  label suicides_100k_pop_tot = 'Suicides Per 100,000 People';
  label generation = "Generation";
run;

* Analyze the relationship between suicides and GDP per capita;
proc report data=data.suicide out=work.plot6;
  title Suicides by GDP Per Capita;
  column country year suicides_no population suicides_100k_pop_tot gdp_per_capita___;
  define year/group;
  define country/group;
  define gdp_per_capita___ /analysis mean format=comma12.;
  define suicides_no/analysis format=comma12.;
  define population/analysis format=comma12.;
  define suicides_100k_pop_tot/computed;
  compute suicides_100k_pop_tot;
    suicides_100k_pop_tot = _c3_ * 100000/_c4_;
  endcomp;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");
run;

```

```
proc sgplot data=work.plot6;
  scatter x=gdp_per_capita___ y=suicides_100k_pop_tot / group=country;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");
  yaxis label = 'Suicides Per 100,000 People';
  xaxis label = 'GDP Per Capita';
run;

proc sgpanel data=work.plot6;
  panelby country / novarname;
  scatter x=gdp_per_capita___ y=suicides_100k_pop_tot;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");
  label suicides_100k_pop_tot = 'Suicides Per 100,000 People';
  label gdp_per_capita___ = 'GDP Per Capita';
run;

* Determine the relationship between suicides and population, HDI and GDP;
proc corr data=data.suicide;
  title ' ';
  var suicides_no population HDI_for_year gdp_per_capita___;
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");

proc sgscatter data=data.suicide;
  matrix suicides_no population HDI_for_year gdp_per_capita___ /group=country diagonal=(histogram kernel);
  where country in ("Argentina", "United States", "United Kingdom", "Belgium", "Brazil",
    "Greece", "Mexico", "Singapore");
  label suicides_no = 'Number of Suicides';
  label population = 'Population';
  label HDI_for_year = 'HDI';
  label gdp_per_capita___ = 'GDP Per Capita';
run;
```