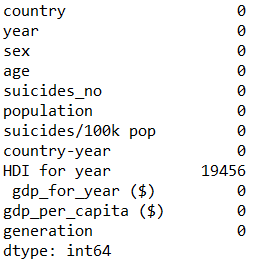
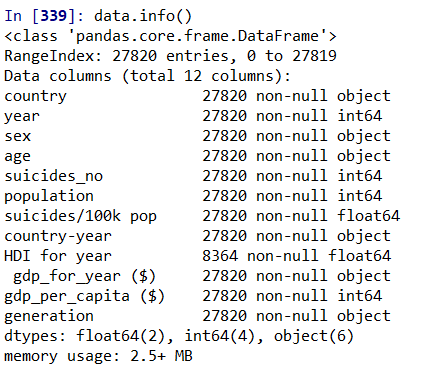
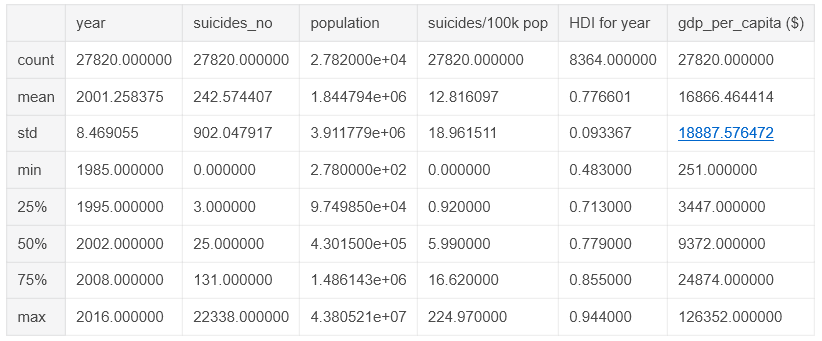
**Suicide Rates Overview 1985-2016**

**About the Data:**

The master file compares socio-economic info with suicide rates by year and country. The compiled dataset was pulled from four other datasets linked by time and place and was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum. The variables include country, year, sex, age group, count of suicides, population, suicide rate, country-year composite key, HDI for year, gdp\_for\_year, gdp\_per\_capita, generation (based on age grouping average). Human Development Index (HDI) is used to measure whether a particular country is underdeveloped, developing or fully developed. The index relies on levels of education, life expectancy, literacy and gross domestic product to rank countries into groups. A quick analysis of the variables show what steps will be needed to clean and transform the data for analysis.



<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>

**Analysis Questions:**

Question 1: What residents are most likely to commit suicide?

Question 2: What time frame did most suicides occur?

Question 3: Are men or women more likely to commit suicide?

Question 4: What age group is most likely to commit suicide?

**Cleaning and Transforming the Data:**

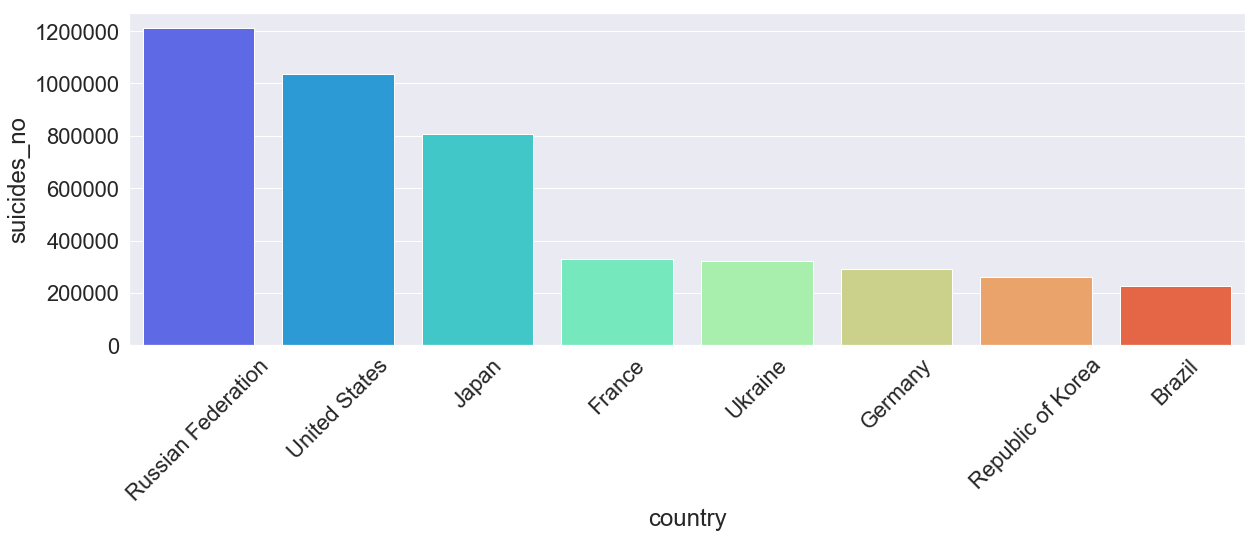
The “HDI for year” and “country-year” columns can be dropped. To provide clear analysis results, the years can be grouped in buckets of ten years by performing “elif” loop commands and apply the new mapping to the “year” column.

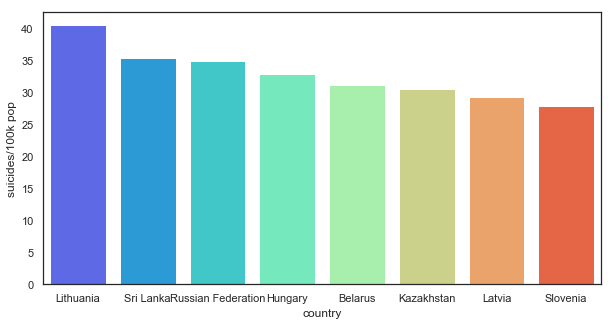
**Country Analysis:**

About the Code:

* Creates a data frame to sum the number of suicides by country then sorts in descending order and plots the results
* Sorts the suicides per 100K population means in descending order, grouped by countries

Results/Output:





Conclusion:

* The countries with the highest amount of suicides are Russian Federation, US, Japan, France, Ukraine, Germany, Republic of Korea, and Brazil.
* The top 8 countries based on the mean of suicides/100K are Lithuania, Sri Lanka, Russian Federation, Hungary, Belarus, Kazakhstan, Latvia, and Slovenia.

Next Steps:

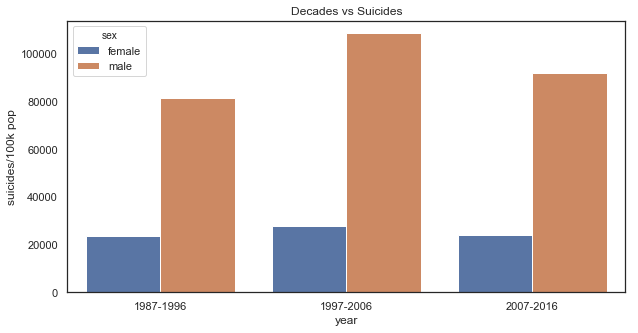
* It would be interesting to compare the results to happiness data.

**Decade Analysis:**

About the Code:

* Created a loop function to put the years into buckets.
* Plots the suicides per 100K pop by year group and breaks out male vs female

Results/Output:



Conclusion:

* 1997-2006 had the highest suicide rate
* That decade has the highest rate based on actual counts and suicides/100K

Next Steps:

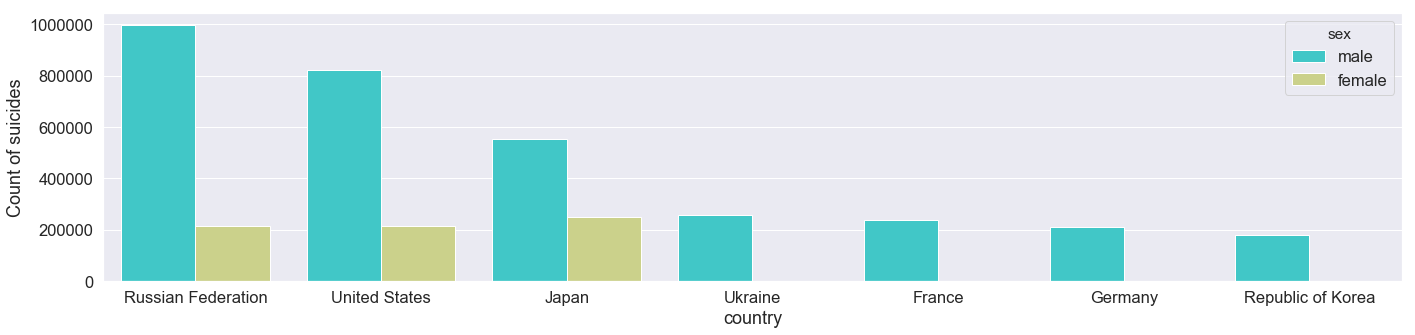
* Analyze world events that may explain the reason for the spike.

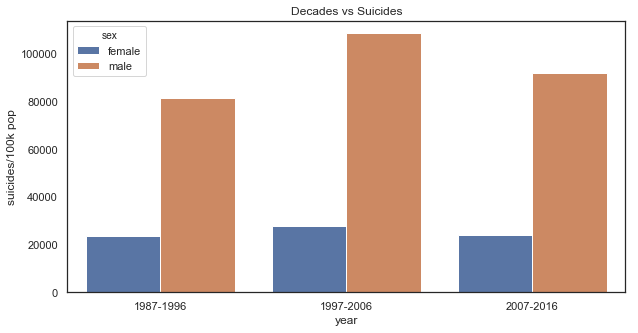
**Gender Analysis:**

About the Code:

* Creates a data frame to sum the number of suicides by country then sorts in descending order and plots the results to compare male vs. female
  + Because the numbers are so large for male totals I Russian Federation the female results do not show for Ukraine, France, Germany, and Republic of Korea on the chart, but there are female suicides in those countries
* Plots the suicides per 100K pop by year group and breaks out male vs female
  + Easier to see male vs. female results so I reused this.

Results/Output:





Conclusion:

* Men are much more likely to commit suicide, this holds consistent by timeframe and country.

Next Steps:

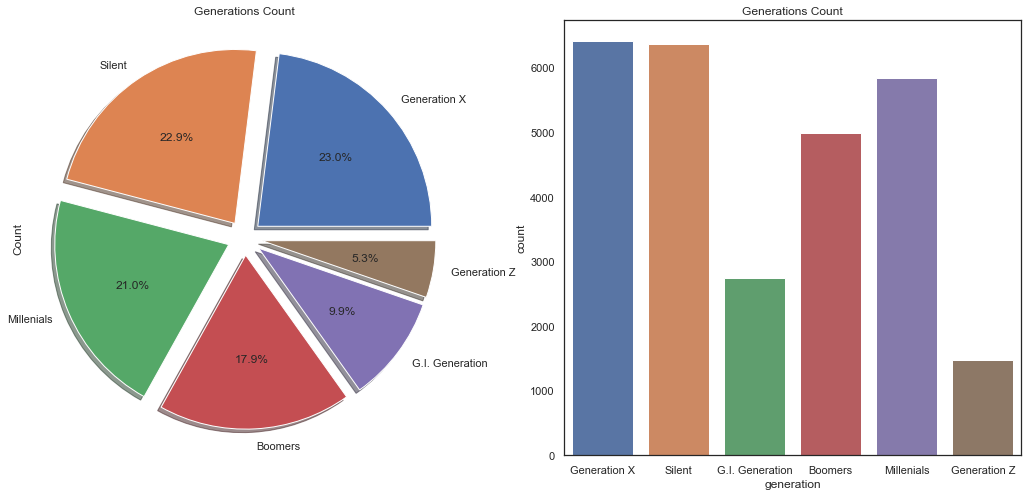
* Try to determine the reason for this – job, family, biological, or societal pressures?

**Generation Analysis:**

About the Code:

* Plots the total suicides by generation in a pie chart and column.

Results/Output:



Conclusion:

* Generation X, Silent, and Millennials make up 66.9% of the total suicides.
* Generation Z is currently under 18 years old which is the reason for the low percentage.

Next Steps:

* Why is Generation GI suicide rate so low, is it because they avoided debt, had strong loyalty, or was it not documented?