Data Visualization Final Project

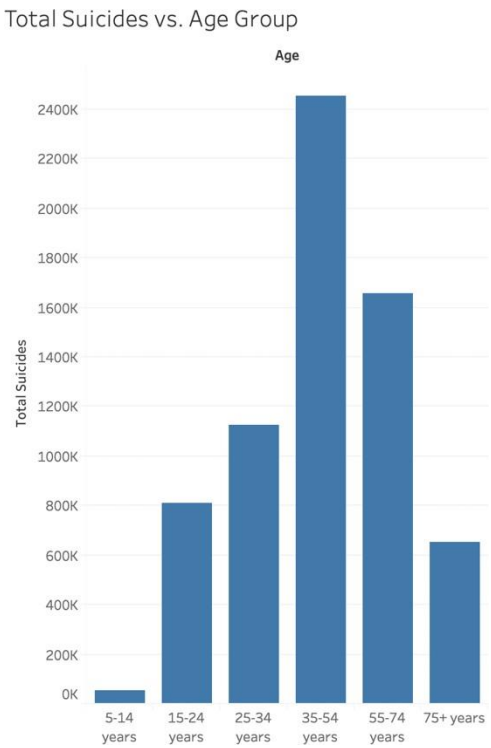
Dataset source: https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016
GitHub repo link: https://github.com/skhaniyur/dataviz_final

Introduction

For my final project, I chose the dataset called: Suicide Rates Overview 1985-2016. I obtained this dataset from Kaggle. Nowadays, there are many studies on world happiness and the happiest countries in the world. I wanted to use this dataset to explore the other side of the narrative which examines the countries where people are most in pain or depressed. This dataset includes individual-level information like age group and generation as well as country-level information like GDP per capita and HDI per year. Using this dataset, I wanted to understand which countries suffer from the greatest number of suicides and dig deeper into the people who are committing suicide. Initially, my hypothesis is that location of country (e.g. cold climate vs. tropical climate) and economic factors affect when and who commits suicide.

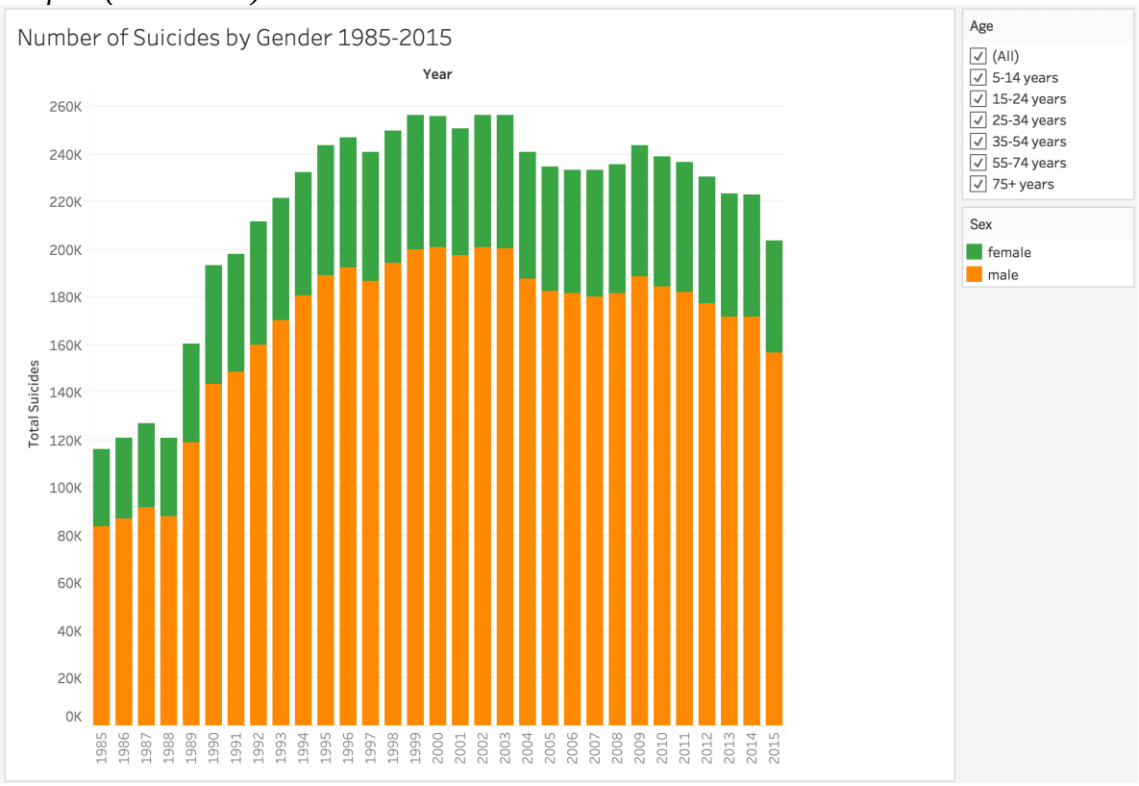
Summary

Histogram



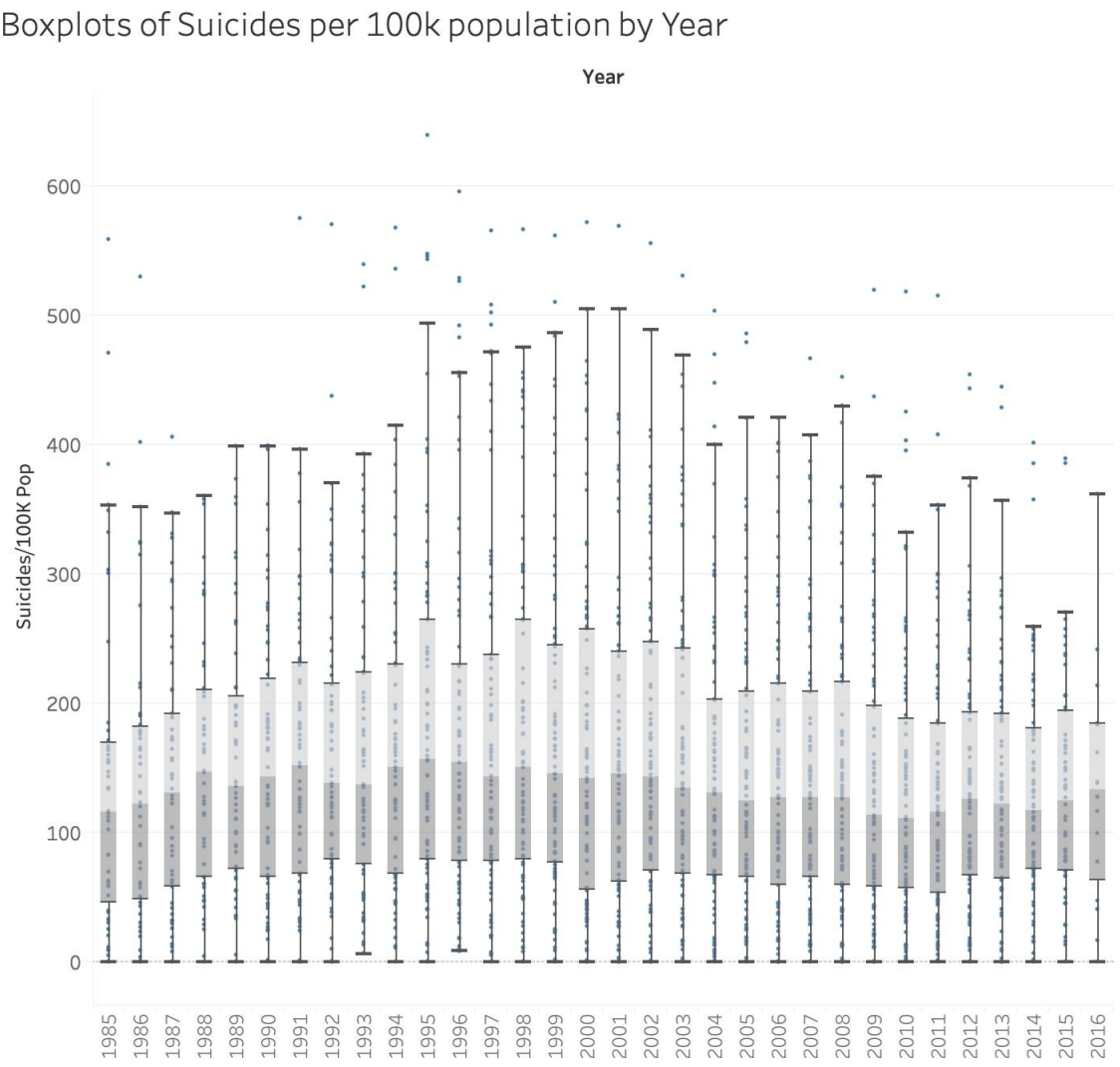
The histogram on the left shows the total number of suicides committed by each age group over the entire period of time (1985-2015). Surprisingly, people between the ages of 35-74 commit the greatest number of suicides. My initial thought was that younger people were more likely to take their lives because the news reports mainly these instances of suicide more frequently. However, I was surprised to learn that middle-aged people were much more depressed. This could be due to negative economic factors since these have a much greater impact at this age.

Barplot (Interactive)



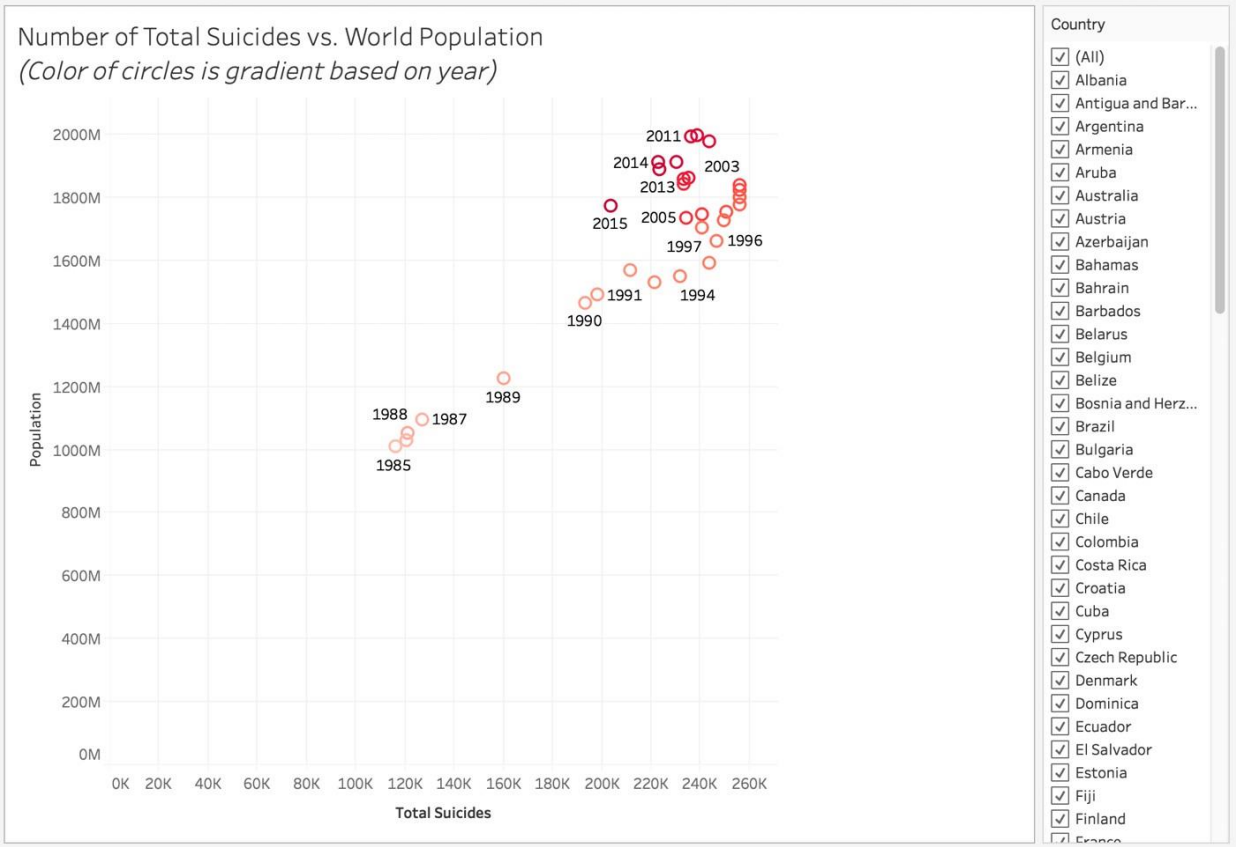
The barplot above breaks down the total number of suicides over time further by gender. It is interactive, so a user can filter by age group as well if needed. (Please see .twb file in github repository for interactive features.) Females tend to commit suicide more than men regardless of year or age group. This could be attributed to the additional social and economic pressures felt by women. There seems to be an overall dip in suicides between the years of 2004-2007, and this might be because of the positive global economy during this time.

Boxplot



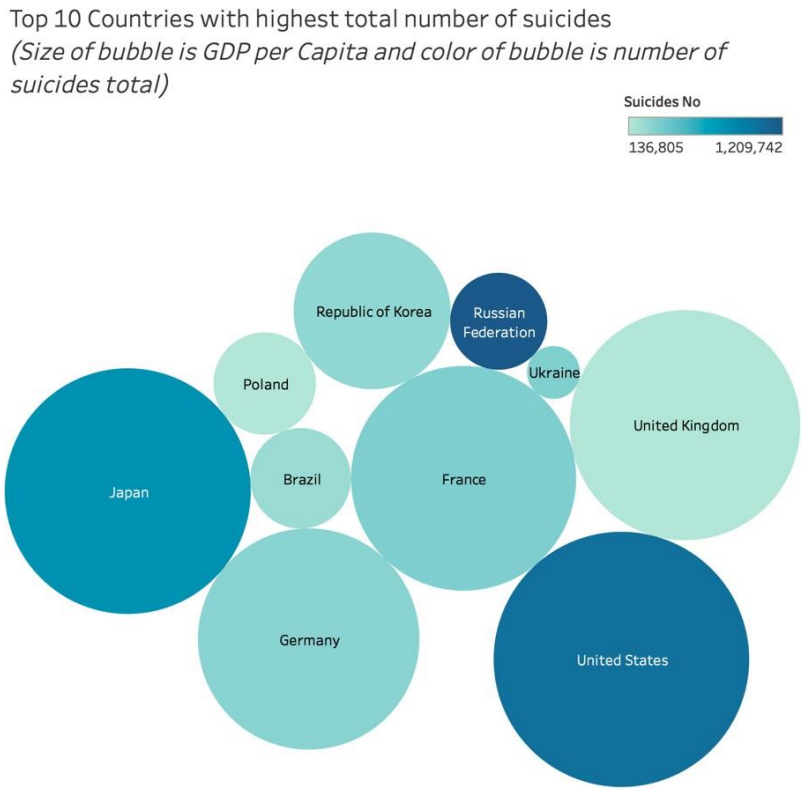
Wanting to further investigate the rate of suicides over time, I created the boxplot above. As seen in the last graph, there is a peak in the rate of suicides per 100K population around 2000-2001. This could again be attributed to the internet bubble bursting at this time which caused a global economic downturn. Again, it looks like there are less suicides during times of economic prosperity.

Scatterplot (Interactive)



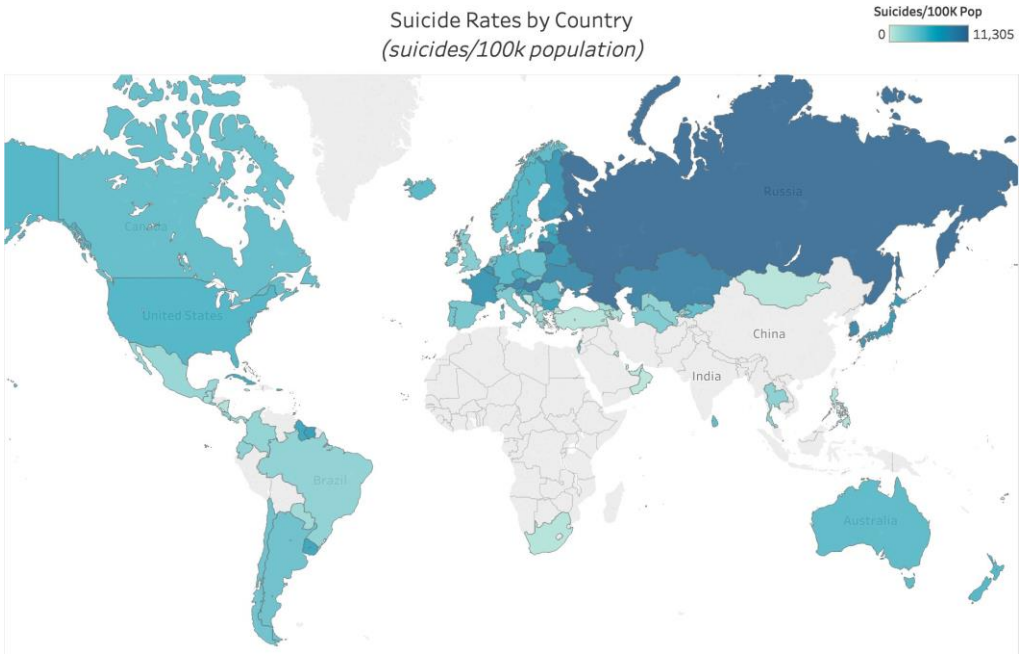
As a sanity check, I wanted to see if the number of total suicides increased as the population increased. As shown above, there is a clear positive relationship between the two numbers. The points on the graph plotted by year and are colored by a gradient based on year as well. This graph is also interactive, so a user could select a specific country they wanted to examine more closely using the filter on the side. (Please see .twb file in github repository for interactive features.)

Bubble Map



The bubble map shows the top 10 countries with the highest total number of suicides. I wanted to explore the relationship between a country's GDP and suicides because GDP is a common indicator of a country's economic status. For example, Russia has the highest number of suicides, but it has a relatively low GDP compared to other countries. On the other hand, the UK has a lower number of suicides, but has a higher GDP per capita.

Chloropleth Map



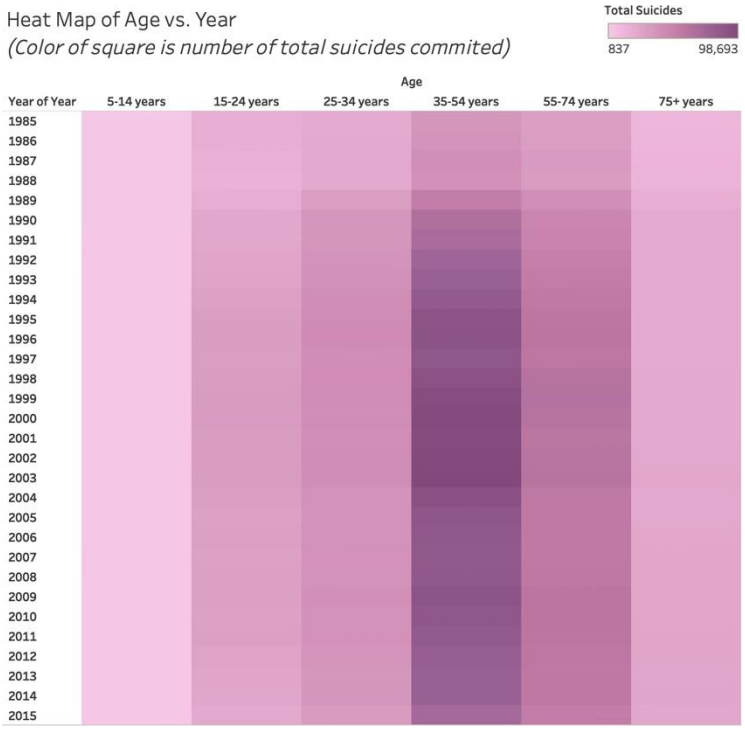
The chloropleth map above shows the suicide rate by country. The darker the color, the higher the rate of suicide. (Please see the Storyline section for a more detailed description.)

Connection Map

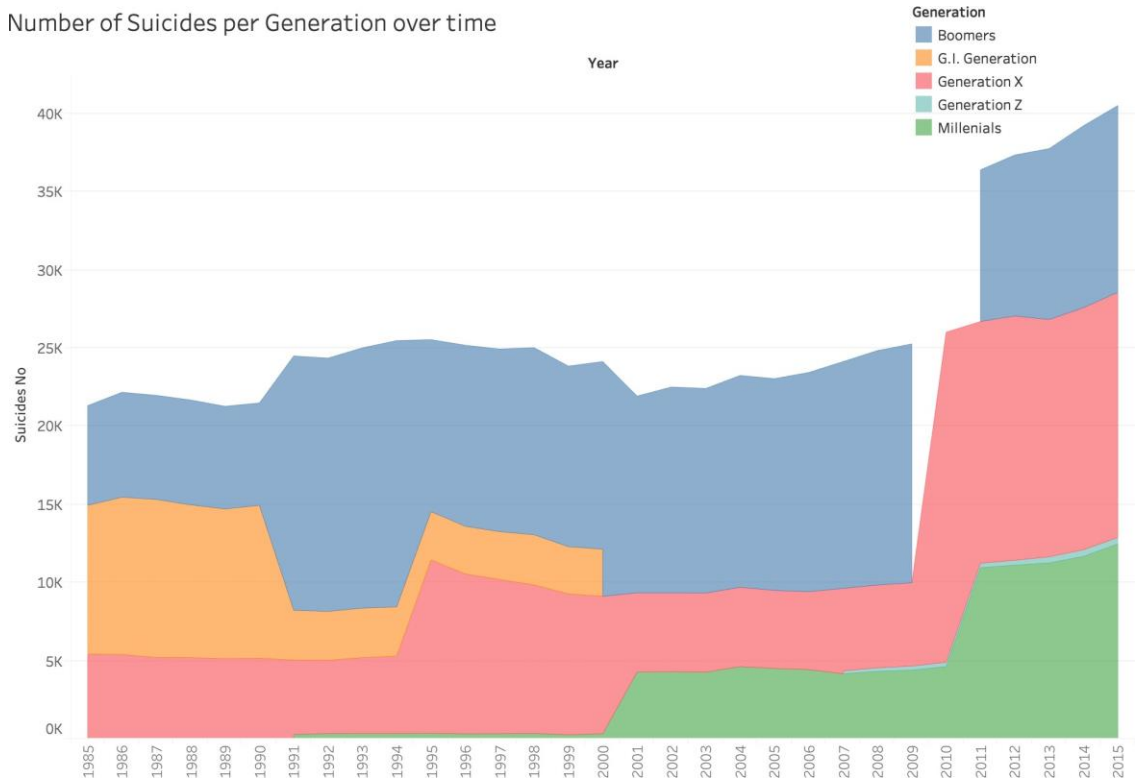
The connection map does not have any real meaning. It just shows the countries connected to each other using the data from two years (2014-2015).

Heat Map

The heat map shows the total number of suicides committed by each age group over time. Again, we can see that the 35-54 age group has the highest number of suicides across the board. However, there are some years where the color is lighter than others, and it is usually correlated with positive global economic growth.

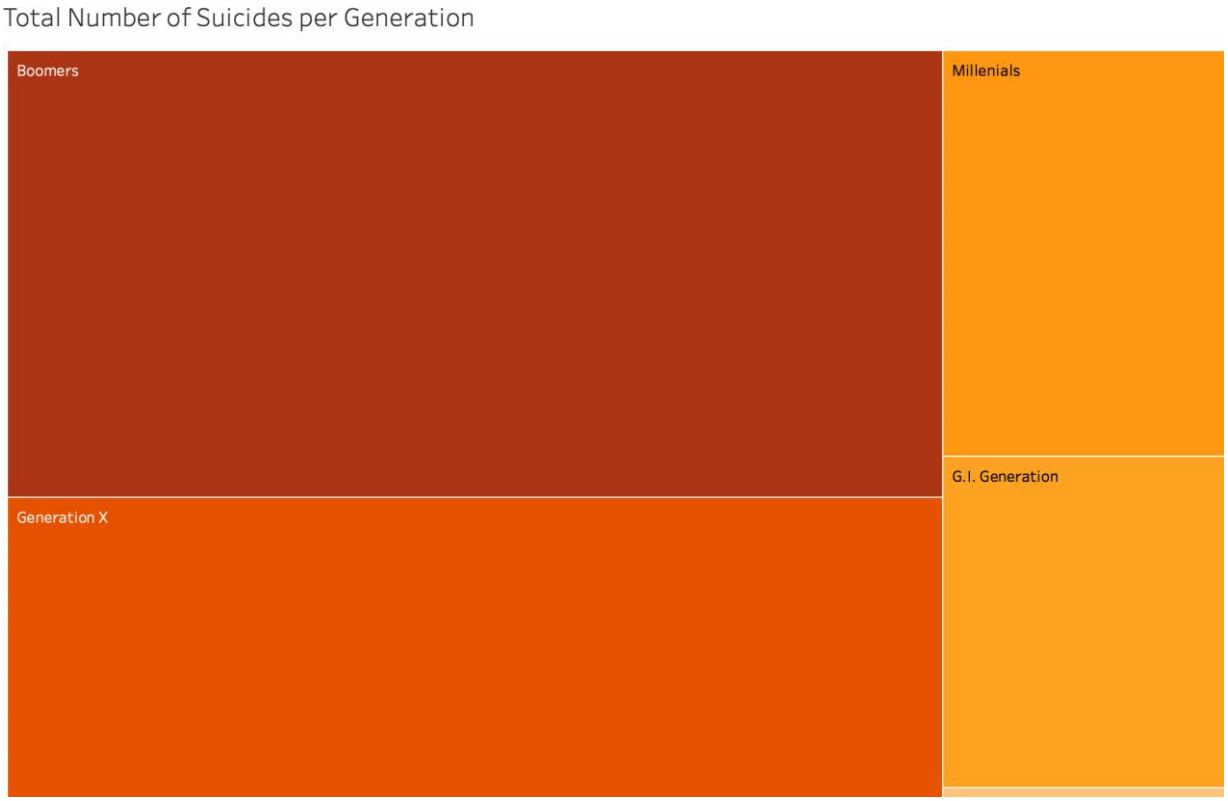


Stacked Area Graph



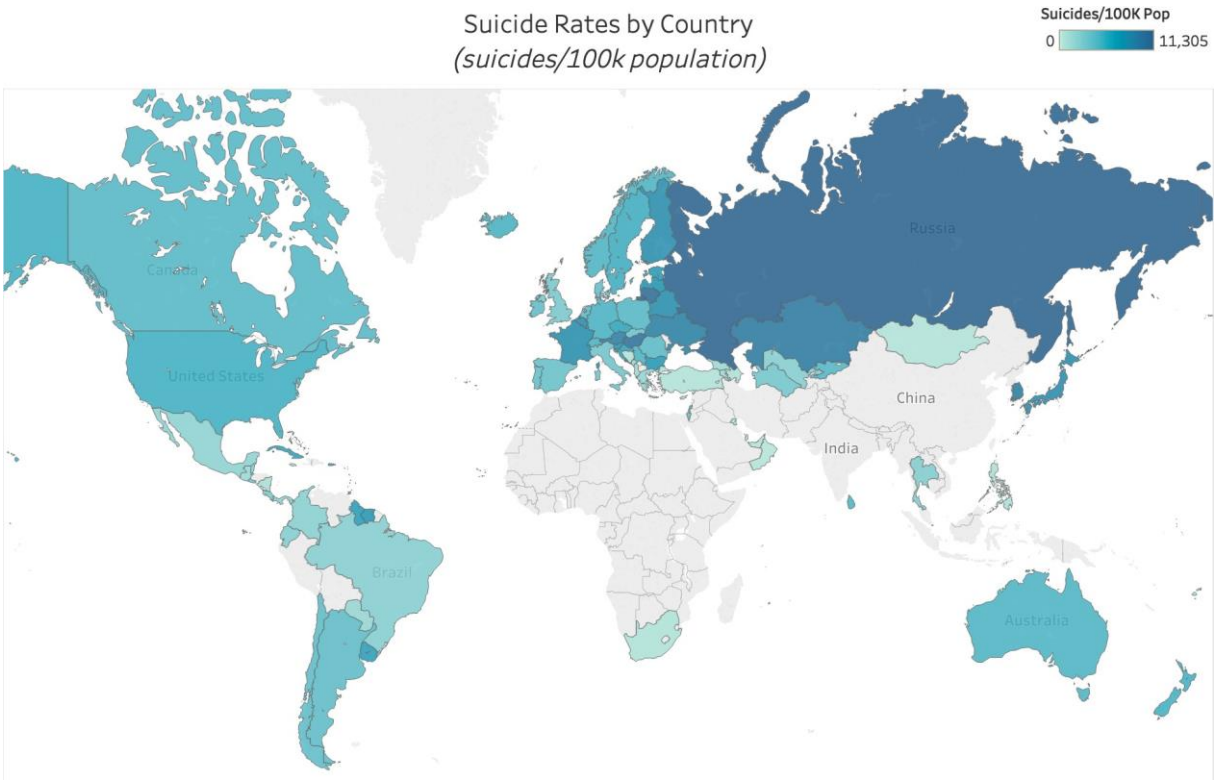
Once again related to age, the stacked area graph above shows the total number of suicides by generation over time. Interestingly, boomers are the generation with the highest number of suicides. There has also been a sharp increase in this generation's suicide number in the last 5 years of the graph as well. It would be interesting to look into causes of this phenomenon.

Tree Map



The tree map shows almost the same information as the previous graph. However, since this chart compares total number of suicides, we see that Generation Z is virtually non-existent in comparison to the other generations. But this makes sense because Gen Z is currently the youngest generation.

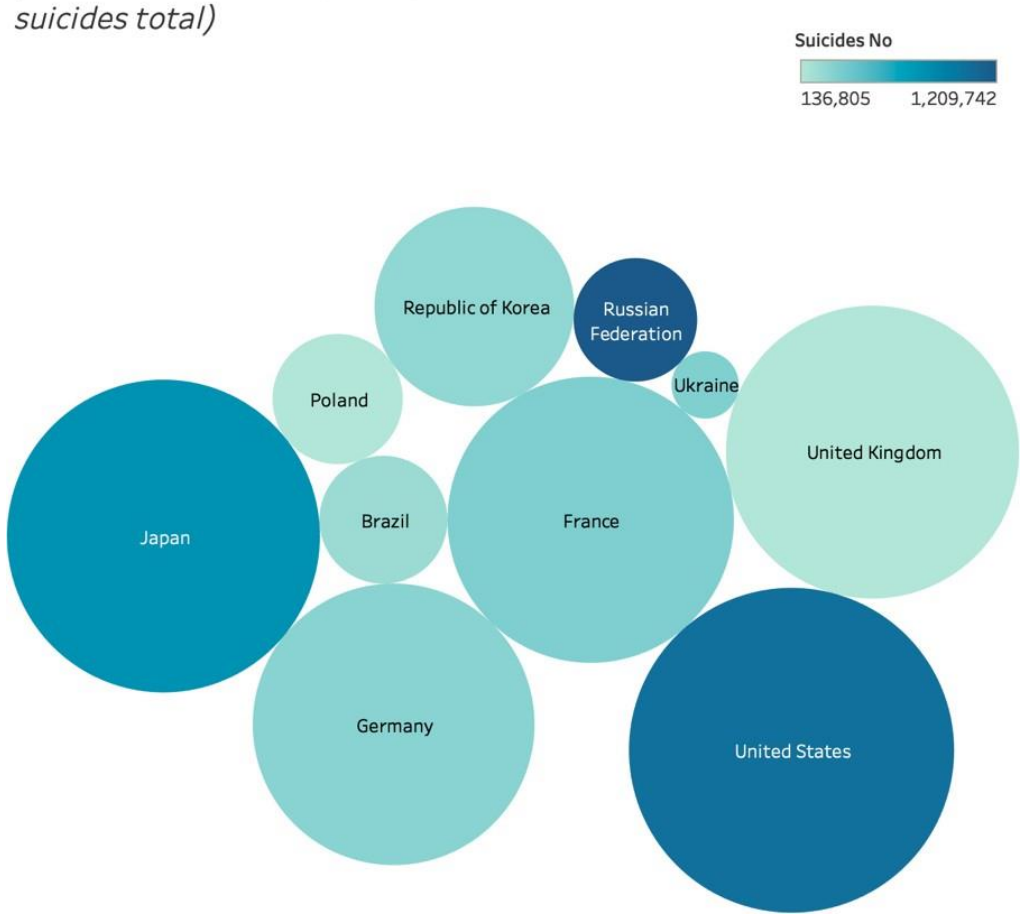
Storyline



As mentioned in the introduction, I was curious to know whether climate affects the rate of suicides in a country. From the map above, it looks like there could be some weak positive correlation between colder climates and suicide rates. Russia and some other Eastern European countries have high suicide rates and cold climates, but I think there is not enough data to make a strong conclusion. It is important to note this dataset is missing data for many major countries such as China and India which are a large part of the global population. If more information from these countries as well as African countries was included, then maybe we would see a stronger trend.

Combined with the insights from the bubble plot below, we can see that Russia is a country where there is cold climate and low GDP per capita when compared to other countries in the world. This combination of factors might be a reason why there are so many suicides in Russia.

Top 10 Countries with highest total number of suicides
(Size of bubble is GDP per Capita and color of bubble is number of suicides total)



Conclusion

There has been a steady increase in the number of suicides alongside the growth in population. Middle-aged people seem to be the most likely to take their own life, and specifically females are more likely to commit suicide across the board than males. Economic and climate factors seem to have a slight effect on suicide rates as well. Colder climates and weaker economies may cause increased depression and stress on people which drives them to take their lives.

Code

Not applicable because all the graphs were done in Tableau. See .twb file if needed.