DSC465 Project Final Report   Group Philcago

Monica Carson
Ying Kam Chiu
Ximan Liu
Monica Mohammed
Dawei Wang

A. Introduction

According to National Insititute of Mental Health (NIMH), suicide is a major public health concern. Suicide is among the leading causes of death in the United States. Furthermore, based on recent nationwide surveys, suicide in some populations is on the rise. In this report, this health issue will be analyzed through a global lens with a dataset sourced from Kaggle (https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016).

1) Data Description

The variables analyzed in this global suicide dataset are as follows:
- Country: 101 countries
- Year: 1985-2016
- Sex: Male, Female
- Age: beginning at age 5 and categorized into 6 groups
- Suicide Rate per 100k Population
- Generation: population born between 1901-2012 categorized into 6 groups
- Income Group: range of country income level categorized into 4 groups

The ages are divided into six groups based on year born in the original data. G.I Generation is from 1901-1927; Silent Generation is from 1928-1945; Boomer Generation is from 1946-1964; Generation X is from 1965-1980; Millennials is from 1981-1996; and Generation Z is from 1997-2012.

The age group classes are assigned in the original data: 5-14, 15-24, 25-34, 35-54, 55-74, 75+. To interpret the rationale for the age groups classes, these are likely divided by life stages and its corresponding stress levels.

The country income level was not included in the original data and was obtained by joining data from the rnaturalearth R library, which contains a compilation of world data including The World Bank's Open Data Catalog by which the income levels were determined. The World Bank determines a country's discretized income level is based on the gross national income per capita using the organization's calculation method.

2) Purpose of Analysis

The goal of this project is fivefold to explore and discover patterns from: 1) the relationship of suicide rate ratio between sex by age group, 2) relationship between suicide rate, country generation, 3) relationship between suicide rate and income group and generation, 4) relationsip between suicide rate of G20 countries and economic performance, and 5) relationship between suicides rates with countries in the G20 and age range.

This report seeks to convey a story about a global issue of suicides using these 5 directions to investigate suicide rate through different time frames, diferent scopes: globally and closer look at G20 countries and different aspects.
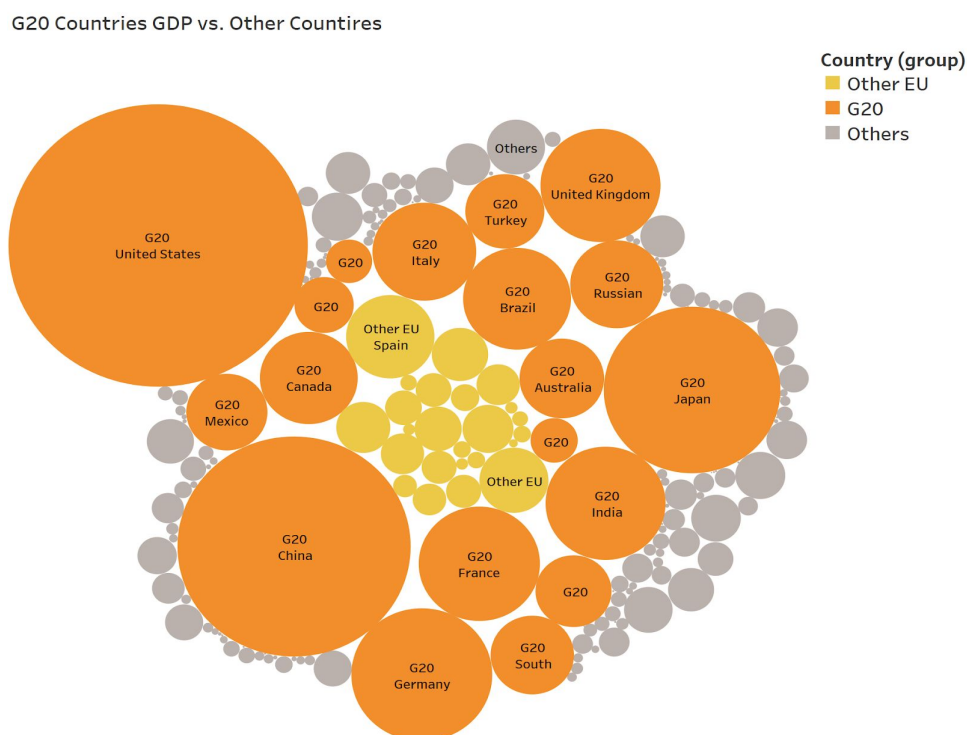
A. Exploratory Analysis

As a team we performed a vast amount of exploratory analysis. The following includes just a few sample visualizations created during this process which helped guide our explanatory visualizations.
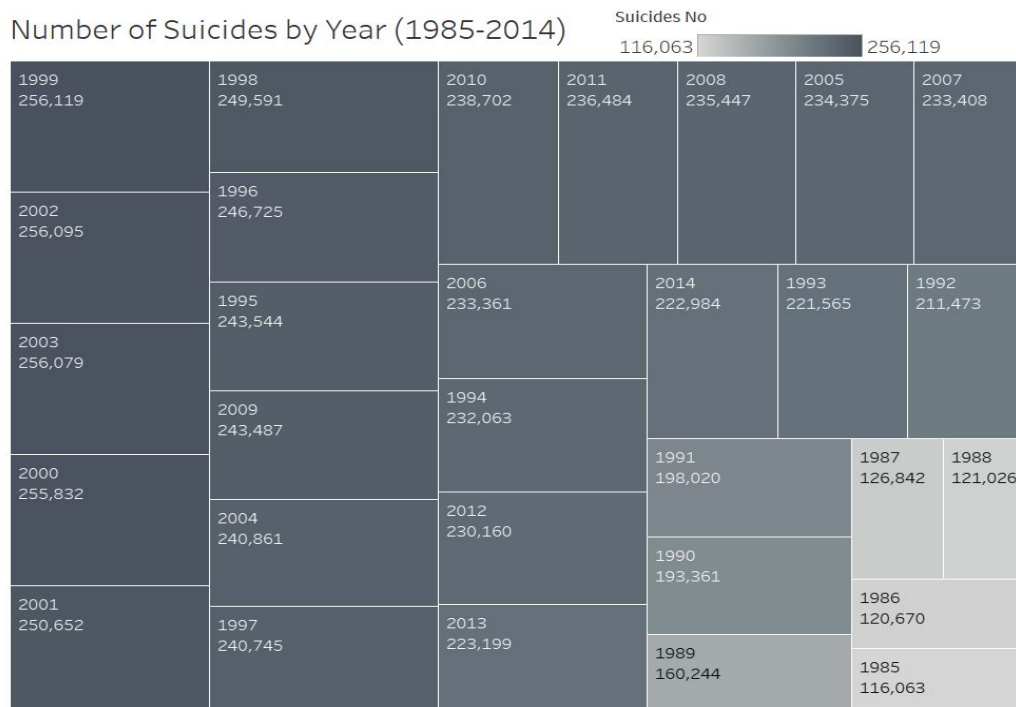
About G20:

G20 has two-thirds of the global population that represents about 90% of the global Gross Domestic Product and about 80% of international trades. It contains advanced economies, emerging economies (defined by IMF), as well as the largest developed country (U.S.), and the largest developing country (China). Variation and comprehensiveness make G20 a great representative for a detailed analysis.

This graph shows the Global GDP proportion of G20 members (2016). Please, note that G20 has 20 members which consist of 19 countries and the European Union. In this graph, individual G20 countries are colored in orange, and the countries that are a part of the European Union but not the individual members of G20 are yellow bubbles in the center.

G20 Countries GDP vs. Other Countires

**A look at the number of suicides by year:**

During our exploratory analysis the following visual was created which looks simply at the number of suicides by year across all countries within the dataset. Although our later analysis and visuals will be focusing on the rate or number of suicides per 100K in population we wanted to determine a good time frame to narrow our focus to start. Both the mid to late 90's and 2000's have a higher number of suicides. In the end we determined it would be most interesting to look at the most recent years provided in the dataset (2000-2014) and expand years when it is deemed necessary.

Number of Suicides by Year (1985-2014)

Suicides No
116,063 ▭▭▭▭▭ 256,119

1999 256,119	**1998** 249,591	**2010** 238,702	**2011** 236,484	**2008** 235,447	**2005** 234,375	**2007** 233,408
2002 256,095	**1996** 246,725					
	1995 243,544	**2006** 233,361	**2014** 222,984	**1993** 221,565	**1992** 211,473	
2003 256,079	**2009** 243,487	**1994** 232,063				
2000 255,832			**1991** 198,020	**1987** 126,842	**1988** 121,026	
	2004 240,861	**2012** 230,160	**1990** 193,361			
2001 250,652	**1997** 240,745	**2013** 223,199		**1986** 120,670		
			1989 160,244	**1985** 116,063		

A look at average suicide rates by sex across 2000-2014:
Some of later explanatory visualizations dive into suicide rates across sex but it was a natural first step while exploring the data to produce different visualizations in an effort to see if there are any noticeable variations in suicide rates between males and females. Tableau is a powerful tool that allows us to create different variations of this quickly. The following is a sample of the many exploratory visualizations that were created.

Average Suicide Rate by Sex (2000-2014)

3
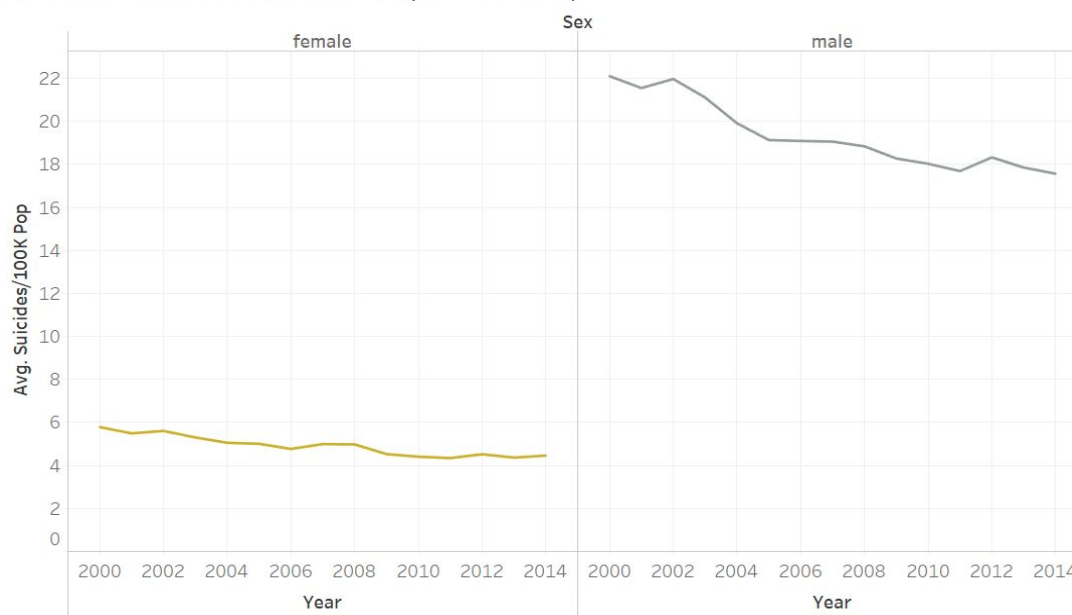
B. Explanatory Analysis

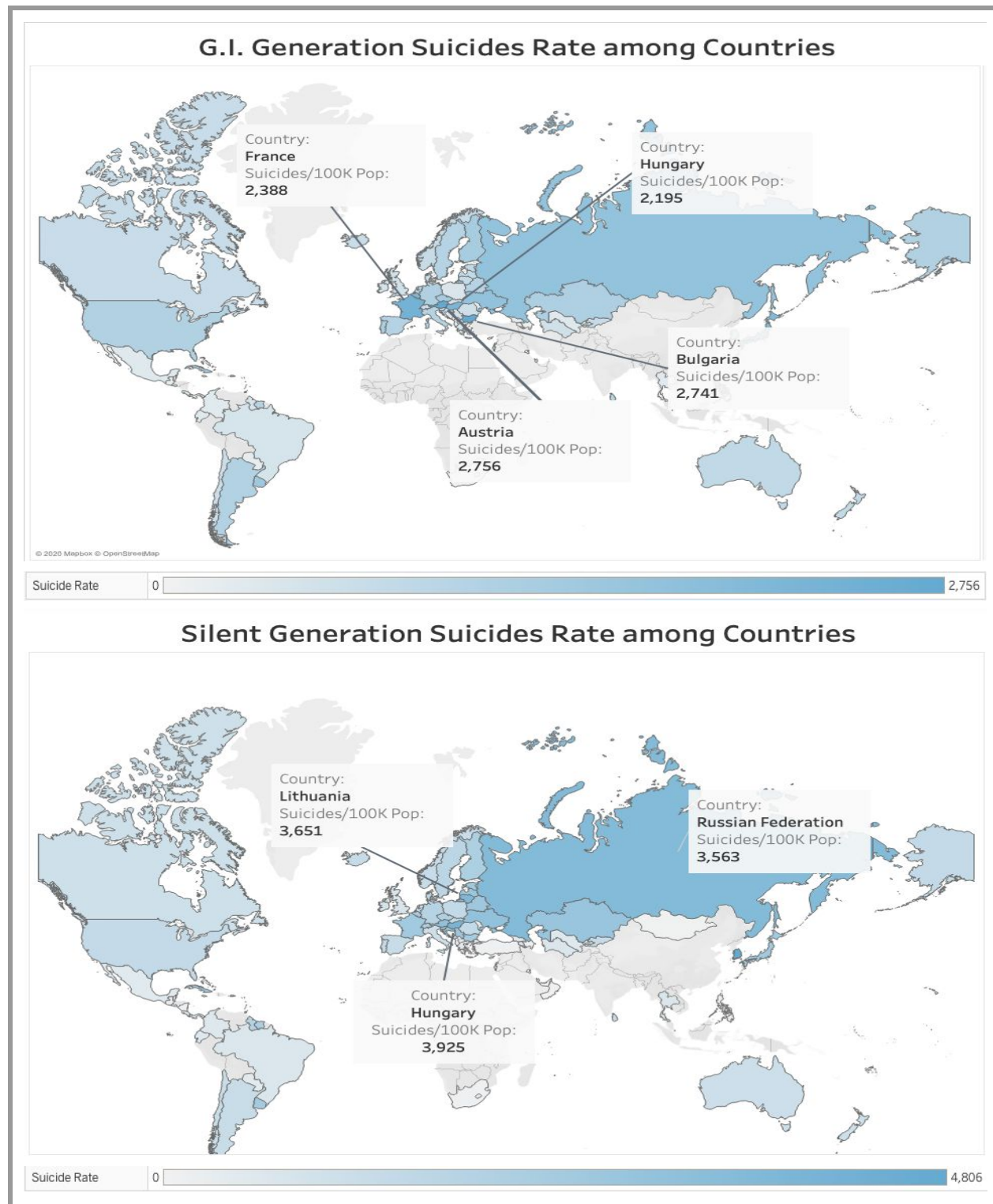1) Relationship of suicide ratio between sex by age group

In Tableau, we decided to create six 10 by 10 waffle charts to show the suicide rate ratio by sex in different chart by age group from 2000 to 2014. Within each chart, we have a hundred colored circles to indicate the represented percentage and sex in response to its relevant sucide rate ratio by sex. Each circle represents 1%. The two color choices are understandably to indicate male (greenish blue) and female (pinkish purple). To contrast the difference, the two colors are on the opposite side of the color wheel. In order to iterate the percentage per sex more clearly, we decided to mark annotations of the ratio in the center of each chart. The charts below use the size of the colored area for comparison. Color is not the best property for comparison, but it could give us a general idea of ratio. If the audience is looking for a specific percentage, the annotations would be a good help.
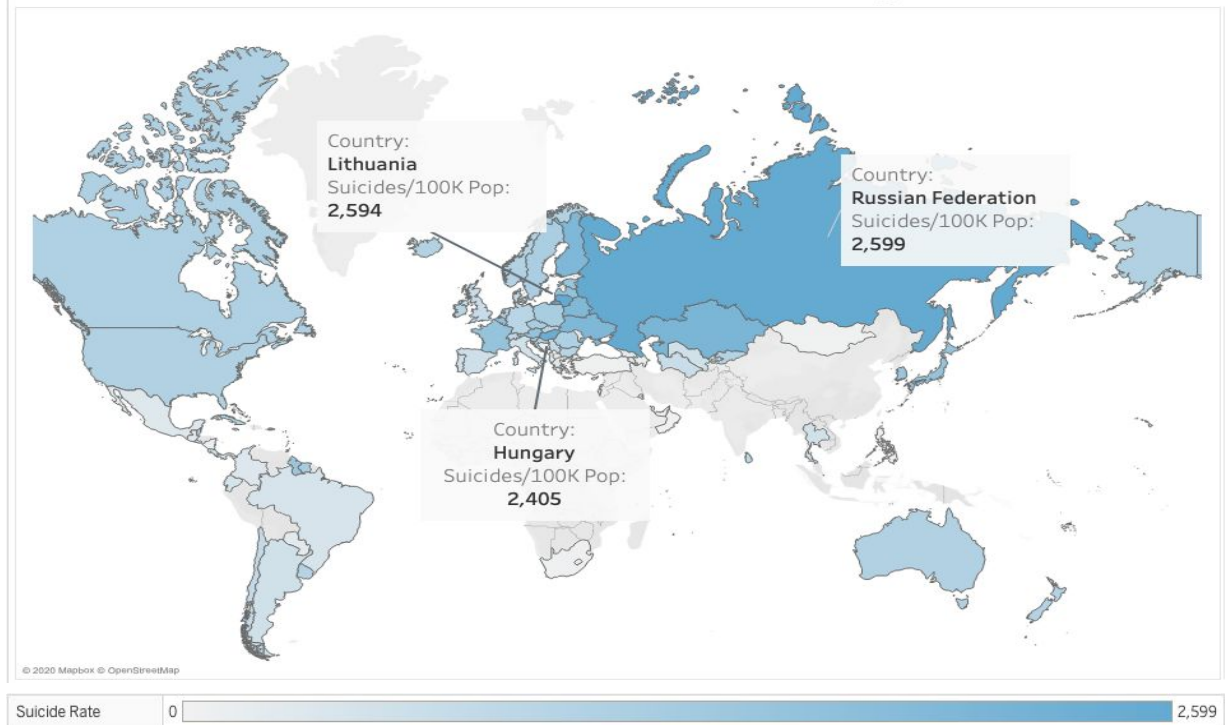

2) Relationship between suicide rate, country and generation

In order to better show the suicide rate of different countries in different years, the most intuitive way to display is to use geographic graphs. The following six graphs show the relationship between suicide rate, country and generation. In order to significantly show the suicide rate of
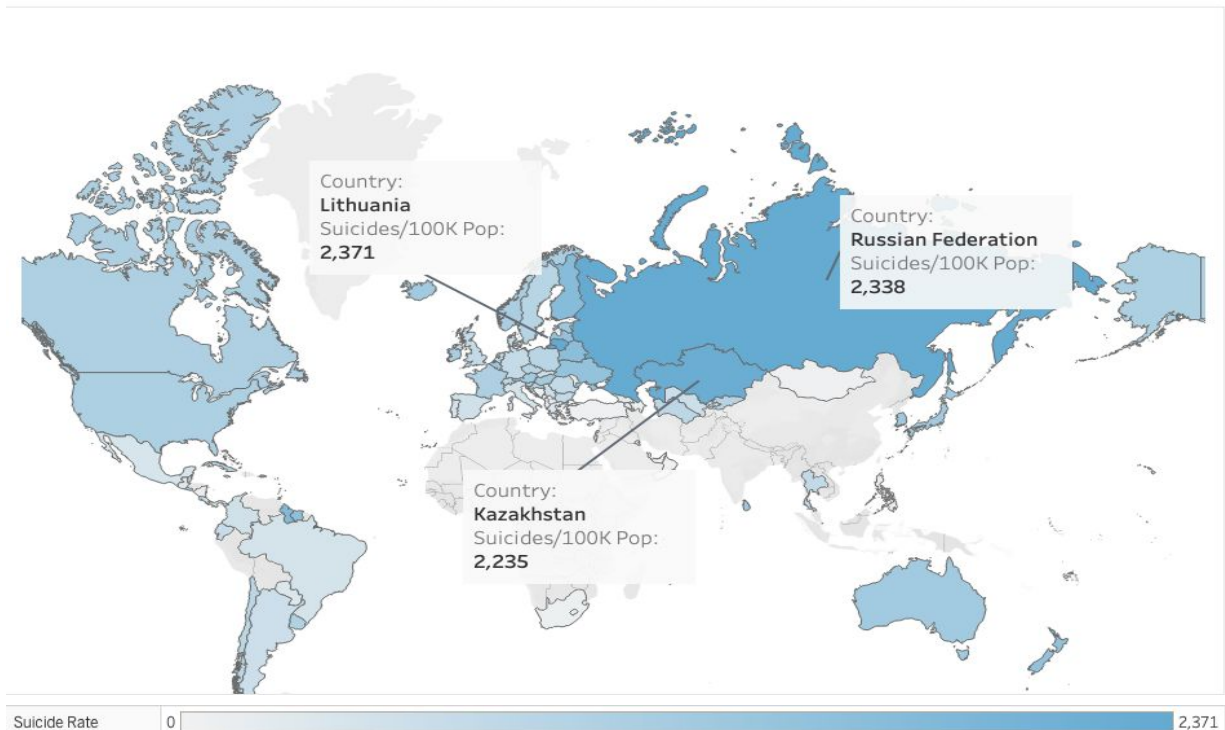
each country in different eras, the geographic map is used to arrange by country. In addition, polarized color specifications are used to more clearly express the suicide rate distribution in each country at different ages. Blue shade represents the extremes, transitioning from less severe to severe respectively.
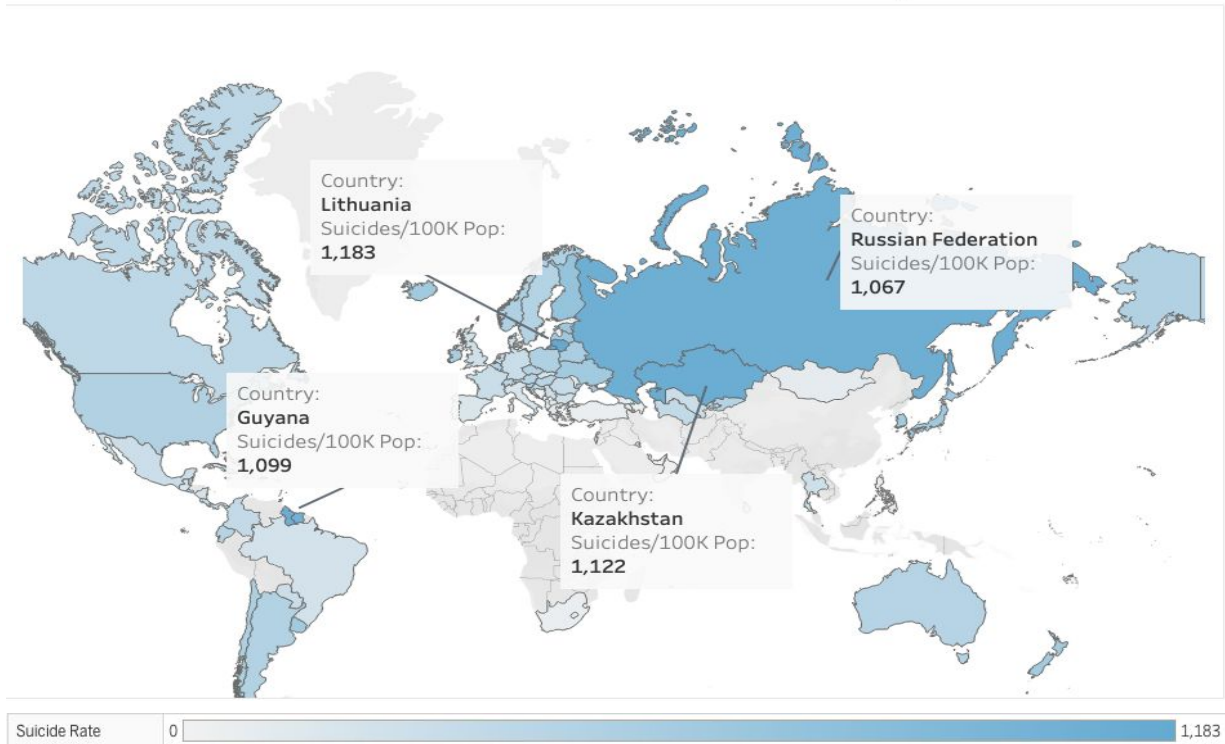
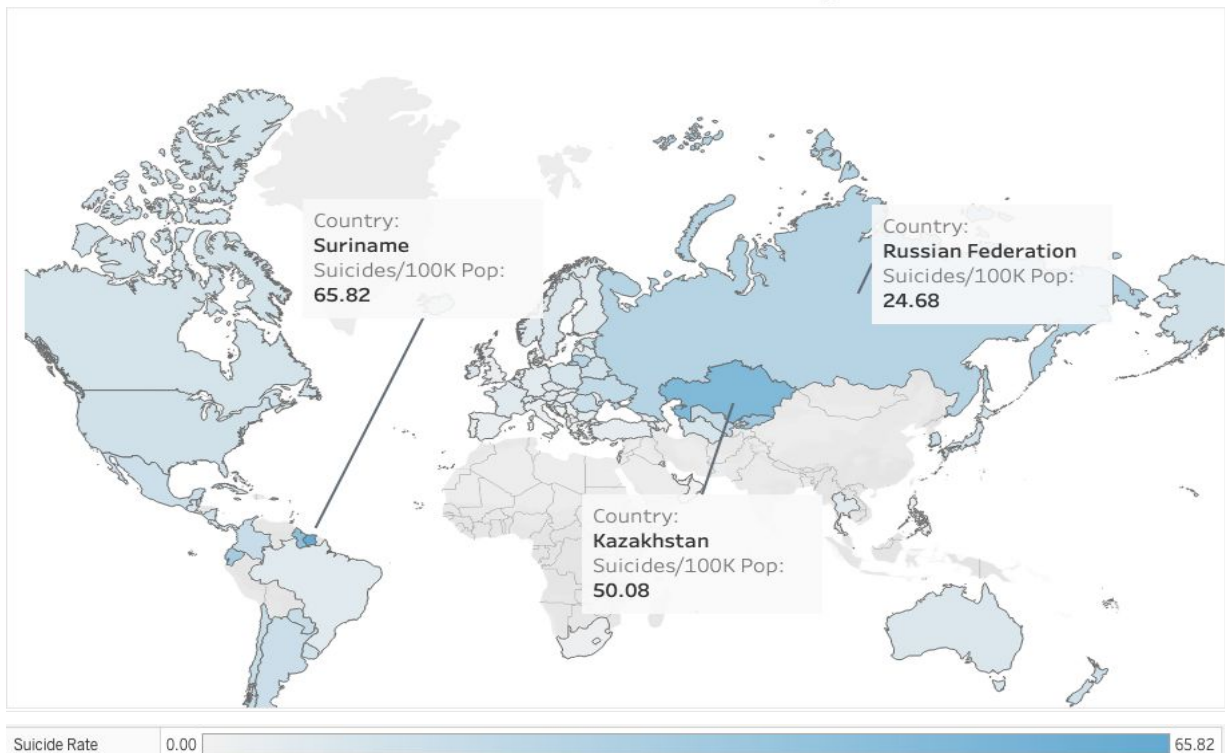**G.I. Generation Suicides Rate among Countries**

Country:
France
Suicides/100K Pop:
2,388

Country:
Hungary
Suicides/100K Pop:
2,195

Country:
Bulgaria
Suicides/100K Pop:
2,741

Country:
Austria
Suicides/100K Pop:
2,756

© 2020 Mapbox © OpenStreetMap

Suicide Rate   0   2,756

Silent Generation Suicides Rate among Countries

Country:
Lithuania
Suicides/100K Pop:
3,651

Country:
Russian Federation
Suicides/100K Pop:
3,563

Country:
Hungary
Suicides/100K Pop:
3,925

Suicide Rate   0   4,806

Boomer Generation Suicides Rate among Countries

Country:
Lithuania
Suicides/100K Pop:
2,594

Country:
Russian Federation
Suicides/100K Pop:
2,599

Country:
Hungary
Suicides/100K Pop:
2,405

© 2020 Mapbox © OpenStreetMap

Suicide Rate   0   2,599

Generation X Suicides Rate among Countries

Country:
Lithuania
Suicides/100K Pop:
2,371

Country:
Russian Federation
Suicides/100K Pop:
2,338

Country:
Kazakhstan
Suicides/100K Pop:
2,235

Suicide Rate   0   2,371

Millenials Generation Suicides Rate among Countries

Country:
Lithuania
Suicides/100K Pop:
1,183

Country:
Russian Federation
Suicides/100K Pop:
1,067

Country:
Guyana
Suicides/100K Pop:
1,099

Country:
Kazakhstan
Suicides/100K Pop:
1,122

| Suicide Rate | 0 | 1,183 |

Generation Z Suicides Rate among Countries

Country:
Suriname
Suicides/100K Pop:
65.82

Country:
Russian Federation
Suicides/100K Pop:
24.68

Country:
Kazakhstan
Suicides/100K Pop:
50.08

| Suicide Rate | 0.00 | 65.82 |

3) Relationship between suicide rate, country income level, and generation

The purpose of this Tableau-created visualization to be graded allows the viewer to make full, complete comparisons between generation groups and how the country's income level may affect this rate. Previous versions of this visualization included a violin plot, which fails to present apparent difference between the distribution of suicides of each income group within each generation using shape and area, Additionally, a bar chart with suicide trend lines also failed to covey the message due to cluttered appearance. The final visualization combines the bar chart's ability to decode specific values, compare the length of each subcategory within its parent group as well as between different groups while still providing information about the groups' changes over time.

The grouped bar chart displays aggregated suicide rates of the generational groups divided into subgroups based on the country income level. The color of the bars encodes the generational groups (ordered by oldest generation down to youngest generation) in a color scheme including color wheel-adjacent hues to show that these groups are related (by variable). Although generally this categorical scheme would require different hues of the same value, the difference in value of the purple and blue-green color was chosen to reduce overwhelming the visualization with the number of hues while still providing distinct color groups. The income groups are encoded in the position on the y-axis and the average suicide rate value is encoded in the length of the bars.

In the line chart, despite the observations being recorded by year, the data points are connected to encode a change in rate from the previous year in the angle between each dot. The G.I. Generation includes a single dot as 2000 is the only year with recorded information, likely due to the age of this population as they were born between 1901 and 1927. Since the shape and slopes of the line better tells the story, the lines are faceted by generation to view the trends separately without the lines overlapping or intersecting.
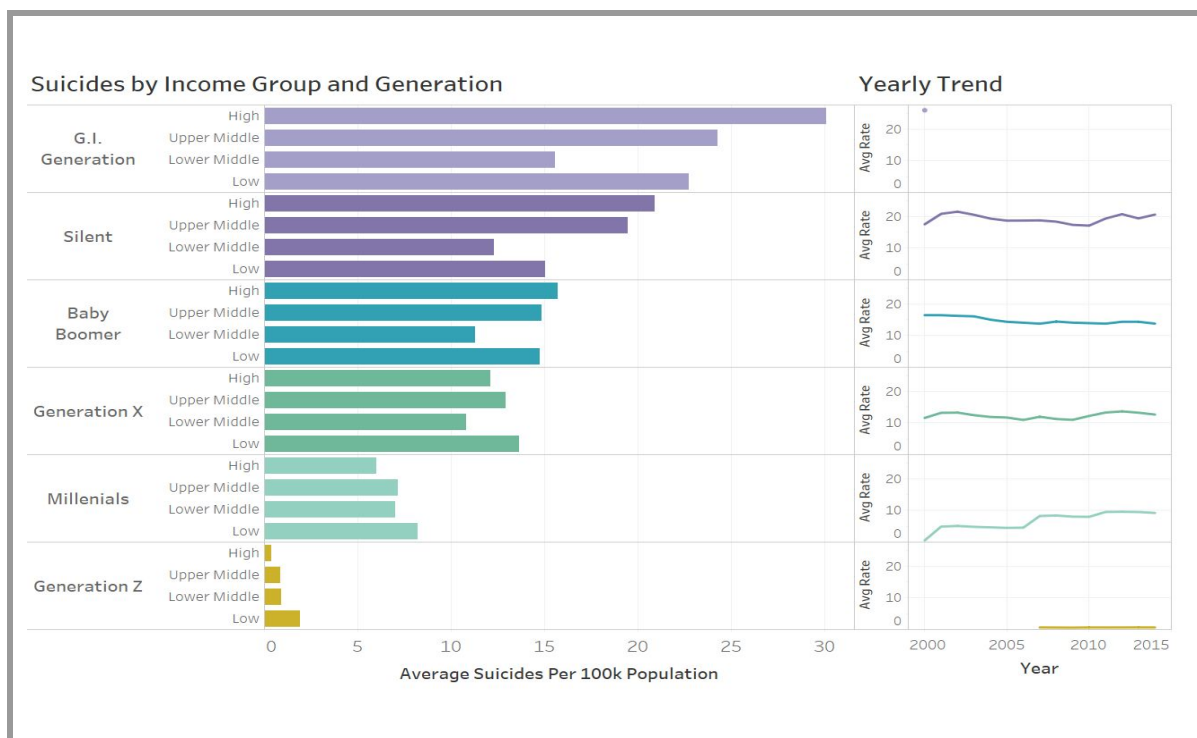
4) Relationship between suicide rate of G20 countries and economic performance

To examine every single country in G20, an interactive map was created. Due to the limitation of format, here's a plain version of this map. The map shows average suicide rate for G20 countries between 1985 to 2015. Clicking on the map and it will show the suicide rate change and GDP growth rate plot of 1985-2015 (or a timeframe that is available) for this nation in a bar/line chart.
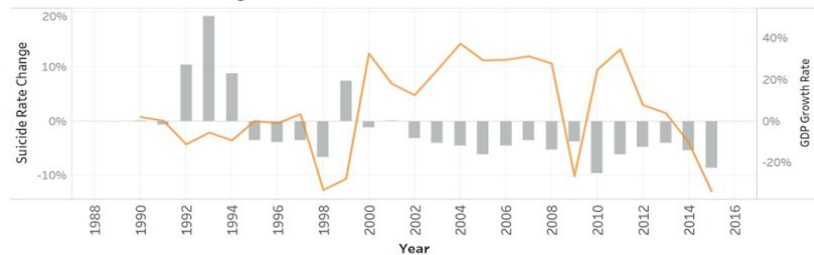

G20 Countries Suicide Rate (1985-2015 Average)

Suicides Rate (Cases/100K Pop)
0.00 — 41.00
© 2020 Mapbox © OpenStreetMap
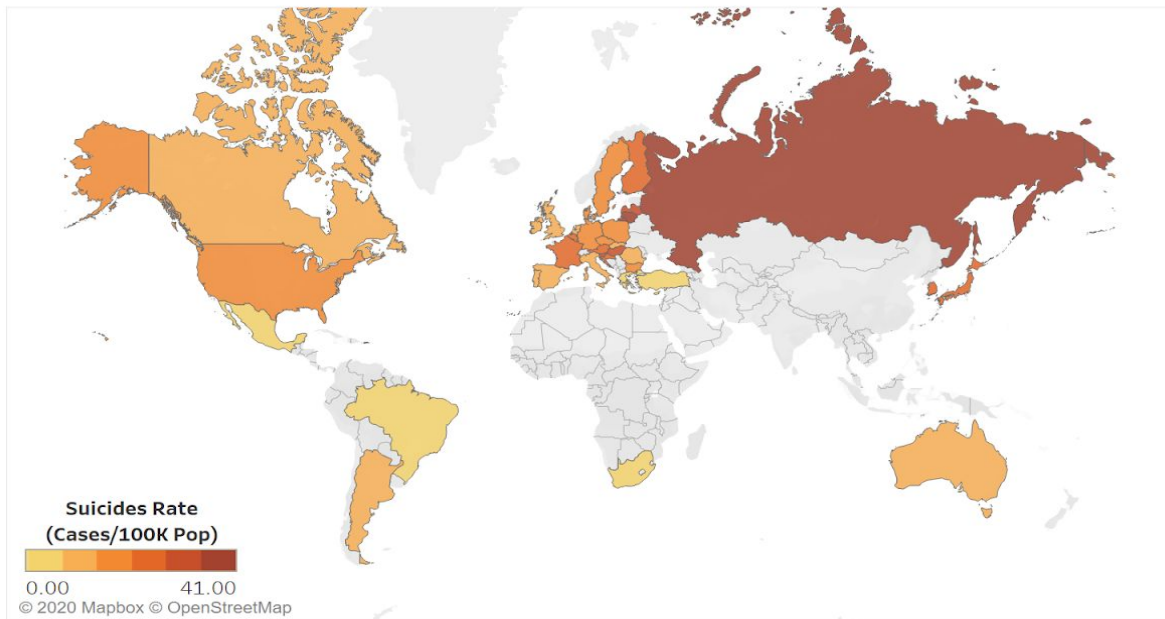
Argentina Suicide Rate Change and GDP Growth Rate (1985-2015)
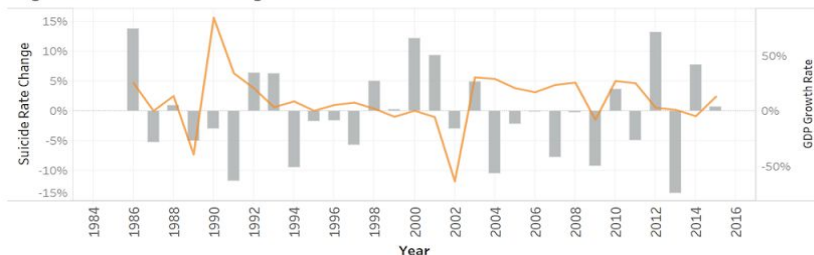
Russian Suicide Rate Change and GDP Growth Rate (1985-2015)

% Difference in Suicide Rate
GDP Growth

After a lot of exploration, there was no pattern found not only in G20 countries as a whole group but also in divided economic groups. Since grouping doesn't work, the best way to discover whether there's a pattern is mapping the graph for all counties. Besides 19 individual members, the EU consists of a lot of countries. The number of this graph is a problem for visitation. Thus, an interactive graph was introduced for a clear and straightforward layout. Since no pattern was found for most countries, the two typical countries were highlighted and present here for further discussion.

    5)   Relationship between suicides rates with countries in the G20 and age range

The following heat map was created in Tableau and provides further exploration of suicide rates within countries in the G20 between 2000 and 2014. All 6 age groups provided in the data are displayed and no further refinement was deemed necessary. The data includes quite a large number of countries to explore as displayed in some of our previous visualizations.

When looking at age groups specifically, expanding the number of countries beyond this G20 related grouping created a visualization that was difficult to digest. In the end the Russian Federation, Japan, Germany and France also displayed the most interesting pattern and are all included in the G20. With such a subject matter as suicide rates there can be a number of factors that could potentially impact this rate across countries. It was important to explore all variables included in the data. As we looked at suicides by country, economic performance, and sex it seemed like a natural progression to next look at age group.

G20 Countries - Suicide Rates/100K Population by Age - 2000 - 2014

Age	Argentina	Australia	Brazil	Canada	France	Germany	Japan	Mexico	Russian Federation	Turkey	United Kingdom	United States
5-14 years	26	12	10	23	12	9	17	22	55	4	4	20
15-24 years	354	299	141	304	203	194	402	188	774	29	157	304
25-34 years	281	433	187	341	416	274	645	179	1,137	28	275	400
35-54 years	252	454	208	464	734	422	860	160	1,169	29	328	503
55-74 years	347	328	221	365	683	524	974	155	1,118	32	222	458
75+ years	585	414	307	351	1,254	996	980	224	1,528	49	227	624

C. Analysis and Discussion

Our first line of analysis was to explore the relationship of suicide ratio between sex by age group. The age group classes are assigned from the original data. To interpret the rationale for the age groups classes, these are likely divided by life stages and its corresponding stress levels. For ages 5-14, the ratio split is comparatively more even, that male have about 60%. The stress of ages 5-14 are probably from family and school. For ages 15-24, the ratio split is getting more obvious, This group is likely in high school, college and getting into early careers. Male seem to

be more stressed in class. The peak of the male ratio of (81.45%) is from ages 25 to 34. This group is in their early to stable careers and starting their new family. The stress is understandably high in this age group. This is especially true in most countries with patrilineality. For ages 35-54, this group could be more stable in terms of not that many new changes as many as the previous group. But the ratio is similar to ages 25 to 34. For all the age groups except for 75+, the populations by sex are similar. But for ages above 75, the population of male is just almost half of the female. The suicide ratio seemed to skew to female much more. But it could be just because the population of male is just almost half of the female for ages above 75.

In another view of the age groups, somewhat surprisingly, the rates of suicides seem to be higher with the older generations. Additionally, the visualizations show that in the G.I, Silent and Baby Boomer generations, the high income group has the highest rates. This is In contrast with generations X, Millenials and Z, where the highest rate shifts to the low income group. An additional part of the story lies in the suicide rate trend for the Millenials. More data may be needed to determine a defined trend for most of the generations, but the Millenials group shows an increasing trend of suicides as time passes. A practical implication for further study is that countries considered on the lower end of income levels may investigate further how the country's economic status or health may affect the younger generations as a cause for suicide.

For each country (or G20 as a group), the suicide rate change and GDP growth rate graph shows no strong correlation between them except for Russian. The orange line shows the GDP yearly growth rate and its axis is on the right side. Lightbars show the yearly suicide growth rate and its axis is on the left side. As we can see from the graph, bars above the 0 axes indicate the suicide rate is positive and suicide cases are more than the previous year, vice versa.

Argentina has the most tremendous economic changing rate. The increasing rate during Argentina's debt crisis in 2002 was around -60% and it also has times that have over 80% rate of economic growth. But the suicide rate change has nothing to do with their economic performance. Russia tells a different story. After unsuccessful economic reform and the impact from the collapse of the Soviet Union around the early 1990s, the suicide rate was increased during this time. After they boosted their economy in the 2000s, the suicide rate was dropped constantly ever since.

Generally speaking, the suicide rate may under influence of a lot of socio-economic factors. But the economic performance of G20 members is not a universal impact element for the number of suicide cases.

The heat map above including suicide rates by age provides an impactful visuliaztion and further exploration of suicide rates within countries in the G20 between 2000 and 2014. The Russian Federation has consistently higher suicide rates across all age ranges. There are a few other points of interest within this visual and data. Japan also looks to also have consistently higher rate of suicide in comparison to the other countries except for the Russian Federation. It would be interesting to look more into any key drivers of higher suicide rates within Japan in the future. France has a high suicide rate in those 75 and older. It became apparent that physician-assisted suicide or euthaneasia laws could be a a driving factor when looking at this age category. France and Germany both have a higher rate of suicide in this age group in comparison to theother younger categories within these countries.

D. Appendix

1. Individual reports:

Ying Kam Chiu

After exploring the data, we decided to divide ourselves into 5 directions to deep dive into the data. My role was to explore the suicide ratio by sex in different age groups. The age groups were classified in the original data. There are in total 5 groups: ages 5-14, ages 15-24, ages 25-34, ages 35-54, ages 55-74 and ages 75+. I selected the data from 2000-2014 because there are least missing values during this time period.

Slicing the dataset into age groups is important because putting too much information into a visualization will confuse the audience or even to divert their attention from the message that I am going to convey. Therefore, I decided to slice the dataset into 5 groups and we are able to discover more details of suicide ratio by sex.

Before deciding on the final visualization, I was brainstorming a few possibilities and one of the possible types of graph was waffle charts. Waffle charts will give a very clear and nice comparison with small multiples or faceted design. The categorical variables: sex and age groups fit perfectly for this type of charts. With the counts of the waffles (circles), depiction of exact ratio could be very quick even without the annotations of percentage.

I used to think data visualization is very subjective as for colors, type of graphs, size and shape selections. After taking this class, I realized there could be some objective ways to differentiate a better type of graph for visualizing the data, meeting the needs of the audience and conveying a strong message. We do have a common scale to judge whether a visualization is good or there are rooms for improvement.

From this project, as a team we planned to take different angles to look at the suicides issues by using different attributes against susicide rate and different types of visualizations. As we dive further in the dataset, we found that we could have a very rich content to take about and many stories to tell. We tried to draft all the possible visualizations and have every graph rated. We exchanged ideas and made constructive comments to each other's work. The critics from others are not the easier things to take. But after digesting more opinions from others, my visualization makes more sense with the objectives of data visualizations from class.

As my personal takeaway, I think there are just too many techniques and tools for data visualization. I have to do more research on visualizations before deciding on the appropriate visualization for my data. There always could be more than one way to visualize the same type of data. The key is to learn more then I will be able to make a better choice on type of graphs, color scheme, size and scale.

If I have more time, I would probably be trying to combine another type of graph with my waffle charts. The possible type of visualization could be line graphs to see the trend or scatter plots to see the distributions. The relationship of suicide rate between sex is interesting and its the pattern

and relationship have been revealed in my visulization quite vividly. But combining another type of graph could discover more hidden patterns and relationships in the data.

Ximan Liu

After understanding the background of the data, the five members of our group analyzed the suicide rate from five perspectives. What I study is the relationship between suicide rate, country and generation. First of all, I set the variable group as generation. It is clearer to compare the suicide rates in different countries under different generation. First, I divided the generation into six groups, namely G.I Generation (1901-1927), Silent Generation (1928-1945), Boomer Generation (1946-1964), Generation X (1965-1980), Millennials (1981-1996) and Generation Z (1997 -2012).

After dividing the age groups, I was considering what kind of graphs to show the suicide rates in different countries. The first thing I tried was heatmap, which can clearly show the suicide rates of different countries from high to low. But for my research question, the way of displaying country rankings in heatmaps is too intuitive and relatively simple. Also other team members used heatmaps to display the results in the initial stage. Therefore, considering the above circumstances, I later chose a geographic map to better show the location of the country and show the suicide rate more intuitively on the map.

In the choice of colors, firstly I chose the extreme contrast of red and blue at the beginning to show the countries with the highest suicide rate to the lowest suicide rate. Later, I listened to the professor's suggestion. Because the contrast between the two colors was too dazzling and could not directly represent countries with high suicide rates, I finally chose to use gradient blue to indicate the range of suicide rates.

The suicide rate in different eras is generally increasing and dropping to low. Looking at the global suicide rate, it has been increasing since World War I until the suicide rate began to drop sharply in the 1990s. Therefore, we infer that war and regional divisions are the direct reasons for the increase in suicide rate. The Generation Z nearest to us is the era with the lowest suicide rate. It can be seen that peace is extremely important for maintaining human spiritual health.

For G.I. Generation, Austria, Hamburg, France and Belgium are the countries with the highest suicide rate. For Silent Generation, Hungary, Lithuania, and Russia are the top countries in terms of suicide rate. Overall, compared with the first era in our study, more countries have increased suicide rates, and the suicide rate is also increasing compared to the first era. We speculate that the causes of the Second World War affected the suicide rate. For Boomer Generation, the suicide rate in Russia has increased significantly compared to the previous two eras. Suicide rates in Lithuania and Hungary are also high. For Generation X, Latvia, Russia, and Kazakhstan are the top three countries with suicide rates. For Millennials, Lithuania, Kazakhstan, Guyana and Russia are the countries with the highest suicide rate. As for Generation Z, Suriname and Kazakhstan have the highest suicide rates. Generally speaking, it is a generation with a lower suicide rate.

In view of the length of the semester, this is the content of my analysis so far. If I have the opportunity to study in depth, I will choose to have a deep research about the background and generation of a country, so as to better explain some problems. For example, Russia has always been a country with a high suicide rate. The suicide rate did not decline until after the collapse of the Soviet Union. Combining history, this is a very worthy question.

All in all, I learned a lot of different data visualization methods this quarter, and I am very glad to apply these methods to the field of analysis.

Monica Mohammed

During our initial meetings, we discussed the amount of missing values and how to handle them. I calculated (using an Excel pivot table) how many countries were included in each year and found 2000 - 2014 had the most countries and could provide a good time frame with dense data for the group to better analyze the data.

Additionally, we discussed adding more information to the dataset. Since there was a geographical component to the data, we wondered if we could compare the data by continent or by development, and I researched libraries that had data we could use to join to our data and also provide geographical information if any group member wanted to use R to create maps as was shown in one of the modules. I provided code I used to clean and join the other library's data to the group although other analyses took a different direction and the additional data wasn't used by the rest of the group.

The visualization I contributed to the group to be graded was the third line of analysis which was the bar chart and line chart that shows the distribution of suicide rates by generation and income level as well as the trend over the years we decided had the least amount of missing values (2000 - 2014). I also provided the code I used to subset the data to my group members.

A key thing I learned is that it's actually easier than I thought it would be to misrepresent or distort data. Although the suicde rate was normalized to the per 100,000 rate to make comparisons between different age/generation groups (included in the data), when making my bar chart, I aggregated the data by summing. I was more focused on the story to be read from the chart rather than noticing that 45k/100k of Silent Generation-high income group in our selected countries is an incredibly high and unlikely rate. Since the scale was large, the chart wasn't able to display some of the income subgroups due to their low values, which could be important on its own, but I wanted to clearly compare the rates between generations and between income levels.

Another takeaway was that choosing an effective color scheme can be difficult. I used a Tableau palette, but choosing six out of the 10 colors proved to be a task. I noticed that there are some colors that are visibly different in some palettes (such as one warm color among all cool colors) that would make one group stand out unintentionally. The automatically assigned colors looked, as I liked to call, like a bag of skittles and it actually made the bar chart harder to read in my opinion.

Also, as I'm sure most students learned, I understood first-hand that visualizations are, indeed, an iterative process in the pursuit of storytelling. When analyzing my initial charts, I could see that there may be several ways to display the same time of data, but the "cool-looking" charts really aren't always the best choice. I wanted to use a violin plot rather than the group bar chart to show the difference between distributions, but the violin foremost showed the existence of a difference between the groups and couldn't really show by how much exactly the distributions were different. I would have missed that there appears to be a difference between the income levels with the highest suicide rate shifted between generations. Also, the visualization was originally vertical instead of horizontal and the income groups were broken down into generations, but by changing the ordering I and the viewers would be able to really see what I was trying to portray in the chart.

As an additional note, I learned to love the tidyverse library. It was intimidating at first and understanding how piping works was difficult to learn on my own in the past, but the examples and tutorials made the process very clear. Reading code that I even wrote was always a bit hard to follow, but I will certainly use tidyverse in the future and I may never analyze data in R without it again!

Dawei Wang

In the group, I actively participate in discussions and provide my insights and proposals. I was responsible for exploring the relationship between G20 countries' economic performance and suicide rate and visualizing all the discoveries along this process.

There are several visualizations I've done in the project.

Performed but not used:

- A map shows the geographic location of G20 Members was created to help people who don't know about G20. The color scheme offers advanced economies, emerging Economies, and countries that are a part of the European Union but not the individual members of the G20.
- The GDP growth rate quantified economic performance. Average suicide growth rate vs. average GDP growth rate for all of the G20 counties (not includes missing 4 Asain countries: China, India, Indonesia, Saudi Arabia) was drawn. There was no clear pattern that indicates correlations between them.
- Different economic groups may have different suicide rates. When grouping these G20 countries as advanced economies, emerging economies, and list other EU countries as another group, a similar figure was also plotted. The same result as the last visualization. A breakdown graph suggests that every single economic group could have various suicide rate distribution.
- An animation of G20 countries' suicide rate and GDP per capita between 1985-2015 was created. The sides of the dots represent the population, while colors represent countries. This animation could give the audience an idea about how the GDP per capita and suicide rate changes among these years.

Contributed to the final report:

- G20 Bubble chart that shows individual GDP of all the countries in the world and color grouped by GDP of G20 individual members, other countries in the EU but not an individual member and all the other countries. A color scheme with different colors that are close to each other on the color wheel was applied to this graph. And they represent different types of countries in the G20.
- An interactive map with average suicide rate and detailed suicide rate change vs. GDP growth rate was another graph I created. This graph could provide the above plot for every country in G20 when clicking on the country on the map.

Data visualization is definitely an iterative process: each visualization could provide insights into your dataset, and these insights could direct you to more interesting details. With plenty of discoveries, how to perform a good-looking and easy-understanding layout would be an essential task to do. Audience-oriented design is beneficial for me. Find the most comfortable way and the most evident demonstration for your audience will also lead you to the right layout, color scheme, chart selection, etc. I set the audience as public, so the graph was designed as straightforward as it could be. Though the visualization contains lavish information, the audience could quickly decode the points with a few explanations.

Another exciting part was the font size and resolution. Almost all the homework I did was edited in Office word. The resolution and appearance were not easy to adjust at the beginning. I had to change the resolution and text, legend size repeatedly, just like other components in the visualization. It became a lot easier when I learned approximate font size and best resolution for different sizes.

If I have more time, I may try to find out the missing data and put them on my map. Another thing is the diffusion cartogram. I have tried a diffusion cartogram in the States, wondering what it will look like on a world map.

Monica Carson

Our overall approach to start was utilizing some basic exploratory analysis to determine what each of us deemed interesting. This naturally brought each of us to find different interesting points in the data to build on and move in our own directions. Milestone 2 and 3 helped drive the process of creating quite a few exploratory visuals until an interesting insight was discovered. I created several iterations of exploratory and explanatory visuals and milestone 2, 3 and our final report. I always remained an active group member and hopefully provided constructive criticism using the techniques we used in class. Also assisted in any final edits or formatting before final submission whenever necessary.

More specific contributions to this final report include:

Two exploratory visualizations:

- This includes the heat map of the number of suicides per year. I thought this was a simple way for us to give some background and reasoning behind why certain visualizations

focus on the years 2000-2014. The paragraph provided below this visualization above provides insight into this decision. As a group we decided that each member can branch out of this time frame when needed to further any interesting insights in the data but this was a good starting point for all.

- The third exploratory visual is just a quick glance at average suicide rates by sex across 2000-2014. It was very simple to utilize tableau and create quite a few variations of this data. This is just a sample of one of these variations. Bringing in variables like sex, age, generation, income group and country across different timeframes was an easy way to pinpoint any variations from what you would expect visually.

G20 Countries - Heatmap on Suicide Rates/100K Population by Age (2000-2014)

- I considered quite a few different types of visualizations for this data and continually kept being drawn to this simple heat map. I felt this was the most impactful way to display this data and it became apparent that this was a very easy way for the audience to understand the message being displayed. Once I saw that the age category of those over 75 was somewhat different then what most would expect I decided to explore this area the most. I created another visualization on this topic, and it was included in the final presentation for critique. To lower the groups number of visualizations overall this was excluded from this final report. In the end this information can be seen in the heatmap as well.

When reflecting on this project one key takeaway I gained was that sometimes simple but interesting can be more impactful than something more complex that is difficult to understand in a few moments. You could say that the theme of our class was data, audience, and message. I kept this in mind throughout this project and likely will in the future. I knew that those that will be viewing this visualization would be more likely to want to comment on an interesting fact highlighted by the visualization. When it comes to visualization, I think that often less is more and creating visualizations is an iterative process. In the end one can often go in a lot of different directions and it is important to figure out what is the overall message you are trying to send with your visualization.

2. Code:

```
library(tidyverse)
library(reshape2)

data10.14 <- read.csv('master.csv') %>%
  rename(country = ï..country) %>%
  rename(gdp_for_year = gdp_for_year....) %>%
  rename(gdp_per_capita = gdp_per_capita....) %>%
  filter(year %in% c(2000:2014)) %>%
  filter(country != 'Cabo Verde') %>%
  mutate(country = ifelse(country == 'Saint Vincent and Grenadines',
        'Saint Vincent and the Grenadines', country))

library(rnaturalearth)
library(rnaturalearthdata)
library(rgeos)
```

```
world_set <- ne_countries(scale = "medium", returnclass = "sf") %>%
  select('name_long',
         'economy',
         'income_grp',
         'continent',
         'region_un',
         'subregion',
         'region_wb') %>%
  rename('country' = 'name_long')

world_suicides <- left_join(data10.14, world_set, by='country')

world_suicides <- world_suicides %>%
  mutate(generation = factor(generation,
                             ordered=TRUE,
                             levels = c('Generation Z','Millenials',
                               'Generation X', 'Boomers',
                               'Silent', 'G.I. Generation'))) %>%
  mutate(age = factor(age,
                      ordered=TRUE,
                      levels = c(
                        '5-14 years', '15-24 years', '25-34 years',
                        '35-54 years', '55-74 years', '75+ years'))) %>%
  mutate(income_grp = case_when(
    income_grp == '1. High income: OECD' ~ 'High',
    income_grp == '2. High income: nonOECD' ~'High',
    income_grp == '3. Upper middle income' ~ 'Upper Middle',
    income_grp == '4. Lower middle income' ~ 'Lower Middle',
    income_grp == '5. Low income' ~ 'Low')) %>%
  mutate(income_grp = factor(income_grp,
                             ordered=TRUE,
                             levels = c('Low', 'Lower Middle',
                                        'Upper Middle', 'High')))
```