ASSESSMENT 2

Ryan Needham, 19156892

April 2021

Word Count: 1656

1 Introduction

The following report will be based on the data set "Suicide Rates Overview 1985 to 2016", that compares socio-economic info with suicide rates by year and country. The report will answer four non-trivial questions pertaining to the data set and through a series of data visualisations will be answered. Data visualisations enables someone with a large or small statistical background to be able to comprehend large tables of data(such as the one being explored in this report) and turn them into a graphical representation that can easily communicate a trend or relationship of the data. In the report the aim is to make the visualisations easy to understand and also display a substantial amount of information in both the captions and the graph itself.It is important because it allows for questions that would otherwise be hard to answer by looking at blocks of numbers to be computerised and answered instantaneously. The report is organised into sections including: Introduction, Motivation and Objectives, Experimental Results, Summary. It can be argued that in the report the four non trivial questions have successfully been answered with graphs and explanations.

2 Motivations And Objectives

The next section will be talking about the motivation for choosing the data set and the non-trivial questions intended to be answered. The data set that was chosen initially had 12 features and over 20,000 different observations. The data set was sourced from 'Kaggle' and has been recognised as a reliable source from others who have downloaded it. The data set's observations and features are related towards global suicides throughout the world. The reason t his dataset is being explored is because it met the prerequisites for number of observations and features, while also being an easy data set to clean using 'dyplr' package. Unlike other data sets it seemed that all the features had a correlation with each other. When initially looking at the data set the idea was to choose four non-trivial questions that pertained to the variables and answering them with graphs that use these variables with different aesthetics using the 'ggplot' package in

RStudio. The objective of this report is to try and explore as much of this data set as possible whilst answering four non-trivial questions that each go in depth on a different feature, for example 'age', 'generation' or 'gender'.

The first non-trivial question that needed to be answered was "In the past 20 years what was the trend for global suicides, if one exists?", the reason this question was chosen was because it relates to all of the observations in the dataset and has the potential to give great visual oversight of the past 20 years. This question cannot be simply answered without providing the correct visualisation and features together, thus making it a non-trivial question. The idea for questions involving the year feature is to create a line graph as that would be the best to show the passage of time in a visualisation.

The next question that was to be answered was "How do suicides differ in each continent and gender?". The following question was chosen because it picks out some of the more weighted features that affect the data set more so than others, following that the question is deemed to be non-trivial due to the fact it cannot be answered by simply looking at the data set and requires computerised calculations to answer effectively.

The third non-trivial question is "How does each generation differ in suicides throughout the past 20 years?". Again, this question was non-trivial for its need to be computerised for an effective answer. The question itself is interesting for the data set because it tackles another feature from the data set not used already that being the 'generation' column. Due to the fact we are focusing on a categorical variable for this question it would be likely to show a bar chart rather than a scatter plot.

The final non-trivial question that was to be answered was "Is there a noticeable trend between age and suicide throughout the past 20 years?". This question is like the first however it tries to get a focus on the age feature of the data set, instead of creating a broad overview. Like many of the other questions, this question would also need to be computerised and given a data visualisation to try identifying a trend, this will make it void of being a trivial question.

3 Experimental Results

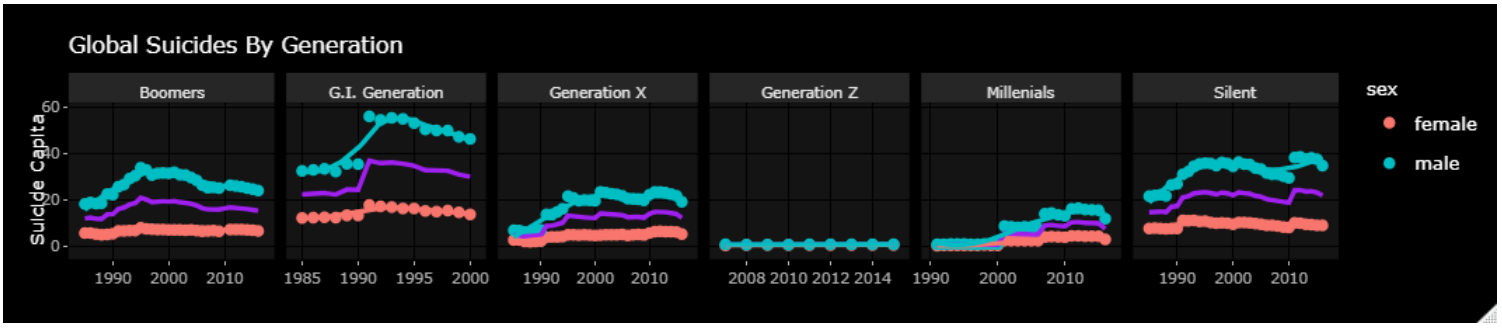

Figure 1: This graph shows the features of year and Suicide Capita that has been have been filtered by gender and going through each different generation that was within the dataset

This graph shows the Global suicides filtered by generation and sex in the form of a line graph and scatter plot. We can identify that the highest peak of suicide was located in the 'G.I' Generation between 1990 and 1993, that is if we're referring to the mean of both genders(Purple line). We can also identify that for most generations there aren't any major relationships, however when looking at the 'Silent' Generation we can see that it has a positive linear trend, conveying that this generation has seemingly become more suicidal over the years, although this is a weak relationship.

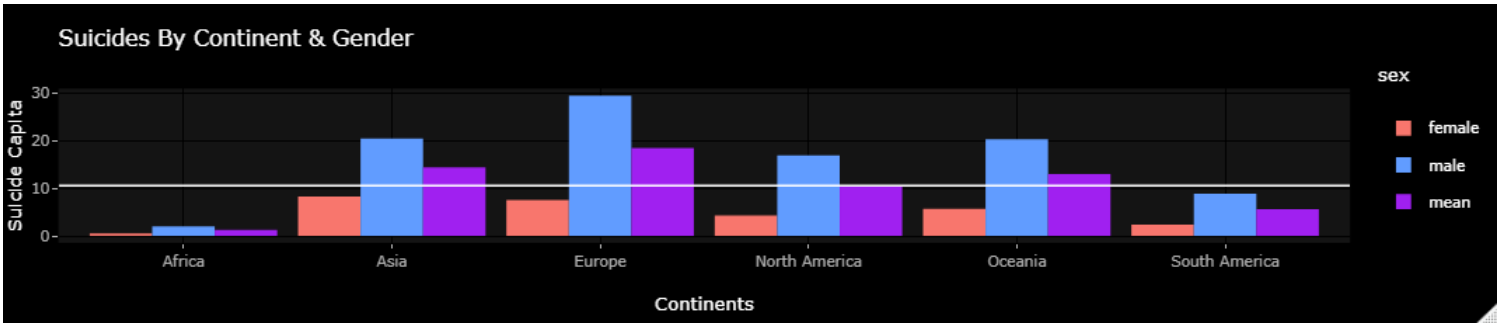

Figure 2: This is a graph that is depicting how the continent feature and Suicide Capita interact with one another using a bar plot. The white going across the graph is the total mean

This graph shows that in all continents the male gender(blue) has a higher rate of suicide than someone who is female(Pink). Europe is host of the highest male suicide rates in the world. However, the highest female suicide fatalities are located in Asia. The lowest Suicide Capita for both genders are in Africa,

suggesting that there is a weak correlation between something like GDP and Suicide. If we look at the mean bars(purple) we can see that Africa and Europe are the ones that stray the furthest from the grand total mean.

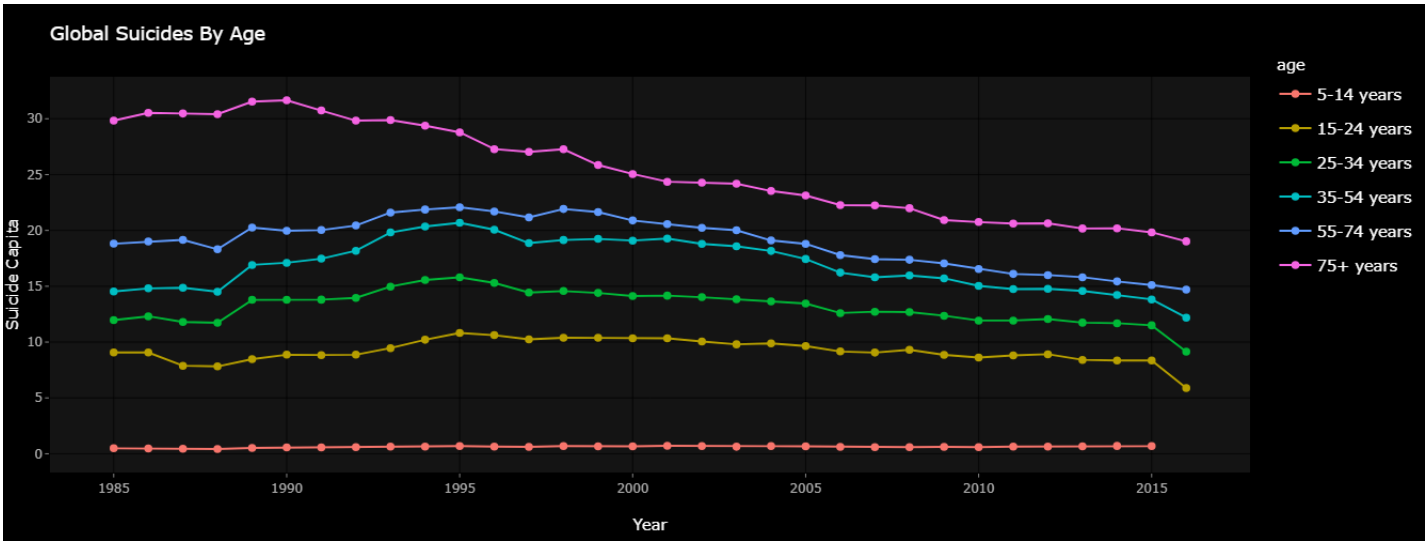

Figure 3: The graph shows a line and scatter plot that is representing the features year and Suicide Capita, is has been filtered through age.

When looking at the graph we can easily notice that suicide seems to have a strong correlation with age. It shows that throughout the past 20 years people who are older will have a larger Suicide Capita. Another key thing to notice is that 2016 has a substantial decrease in Suicide Capita compared to that of 1985. The peak Suicide Capita was located in the 75+ age category, with a number of '31.65'. If we exclude the outlier of ages '5-14' the lowest Suicide Capita would be in the ages 14-24 at a number of '5.89', in the year 2016.

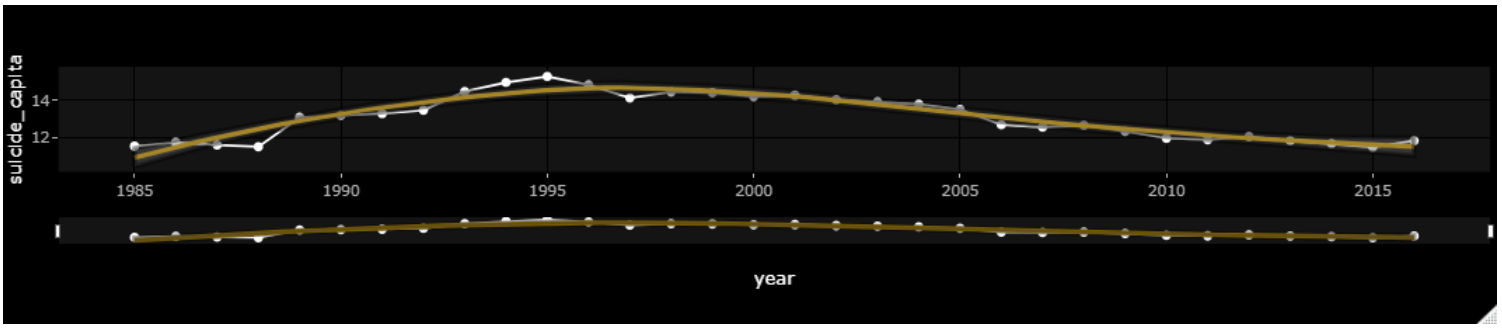

Figure 4: The graph shows the total amount of global suicides with the features year and Suicide Capita. The graph is also paired with a smooth linear regression line.

For this graph it shows the total global Suicide Capita from 1985 to 2016, unlike others there are only two variables, 'Suicide Capita' and 'Year'. Something we can identify is that 1995 has the highest Suicide Capita but looking at the 'loess' regression line it shows that around 1993 would be the highest point of Suicide Capita. We can also identify that from 1995 the regression line shows a negative linear relationship indicating that the suicide rate's over the years are progressively getting lower.

4 Conclusion

The first question that should be addressed in the conclusion is "In the past 20 years what was the trend for global suicides, if one exists?". If we look at figure 4 "Global Suicide by Year" we can identify that the ongoing trend shows the Global Suicide Capita slowly increasing from 1985 all the way up to 1995(Highest Peak), from then on we can see a negative linear relationship between year and Suicide Capita. In short the trend from 1985 to 2016 is an arc that ultimately ends with a negative relationship, thus a decrease in suicides only from 1995. However from 1985 to 2016, Suicide Capita is the same. The main finding of this graph was that suicides have neither increased or decreased massively over the last 20 years. Similar to figure 3 there was a line and scatter plot graph that used some smooth linear regression that would give a rough estimate of where the points would be placed.

The second question that needs to be answered is "How do suicides differ in each continent and gender?". When looking back at figure 2 we can identify that Europe has the largest peak mean of Suicide Capita. Then if we look at Africa it has the lowest mean of Suicide Capita compared to the rest of the continents. Gender suicide for males is also highest in Europe and the lowest being in Africa. A conclusion we can make about the graph as a whole is that countries with a lower GDP income or even life expectancy seem to have a lower

Suicide Capita, this is emphasised by the massive juxtaposition in Africa and Europe. Another key thing to pick up is that females have a lower Suicide Capita for every continent, Africa again being the lowest.

When looking at Figure 1 we can answer the non-trivial question "How does each generation differ in suicides throughout the past 20 years?". We can quickly learn that the G.I. Generation has the largest Suicide Capita on average especially in the year 1993. For the most part each generation seems to have a relatively different mean of suicide capita, except for Generation Z. So to answer the question we can say that there are 'Yes' the generations do differ throughout the 20 years, with 'Silent' and 'Boomers' being the most similar.

The last non-trivial question to look at is "Is there a noticeable trend between age and suicide throughout the past 20 years?". The way to answer this graph is by looking at figure 3. The most obvious thing to pinpoint is that Suicide Capita has a heavy relationship with age. The relationship being that the older people have a higher Suicide Capita. The reason that a line/scatter plot graph was chosen was because it allows for the x axis to be a representative of time whilst easily understanding it.

From this report we can draw a few conclusions from the dataset; we can acknowledge that Suicide Capita has a heavy correlation with age, throughout the whole 20 years with a higher age typically meaning a Suicide Capita. Suicide Capita globally peaked in the year 1995 and has since then been going back down slowly. The continent with the lowest suicide capita is Africa with the highest being Europe.