Visualize Suicidal Rate and Trend

Sanzida Parvin

When we lived in Seattle, me and my husband struggled a lot to cope up with the rainy gloomy days during the winter. Then we heard that the suicide rate in USA is highest in Seattle due to the gloomy rainy weather. That was a kind of shocking news for us.

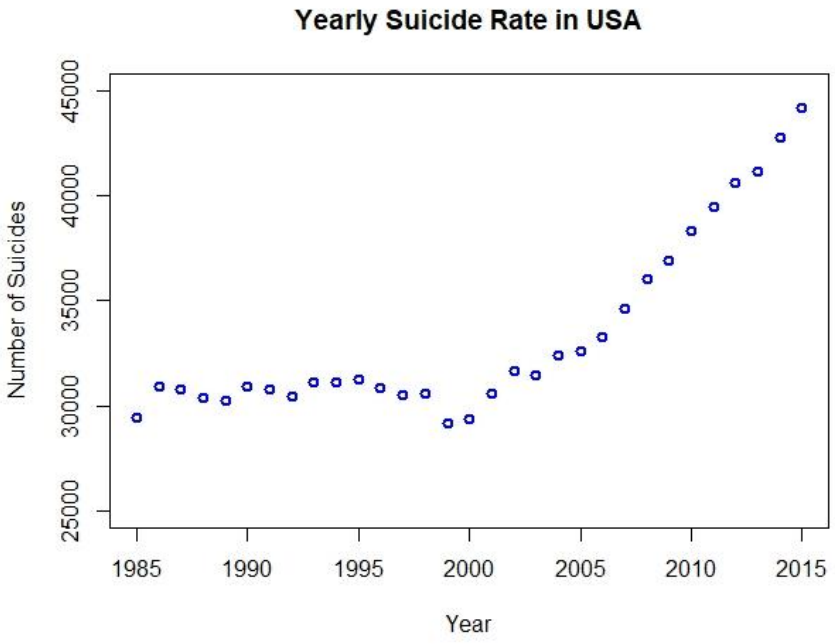
For this visualization project I have found the suicide data set from Kaggle site and thought to see the actual rate and trend of suicide in USA. In my visualization I have shown the yearly suicide rate of USA. For this purpose, I have used the years from 1985 to 2015 as my X-axes and the frequency if the suicide on my Y-axes. The link of the data source will be given at the end of this report.

For this visualization project I have chosen the suicide data set from Kaggle. The full data set has 27820 observations with 12 variables. The observations consist of the data about 100 countries all over the world. For this project I have filtered out only the USA data with relevant six variables. The original data set has no missing or unusual values, so I did not need any cleaning procedure to prepare it. Finally, I have used R programming language to visualize my findings. The R code for my visualization project is given below.

```
suicide = read.csv(file = "C:/Users……../suicide.csv")

head(suicide)

names(suicide)

colnames(suicide)[which(names(suicide) == "ï..country")] <- "country"

USA_data = suicide[(which(suicide$country == "United States")), ]

USA_data[1:5,]
```

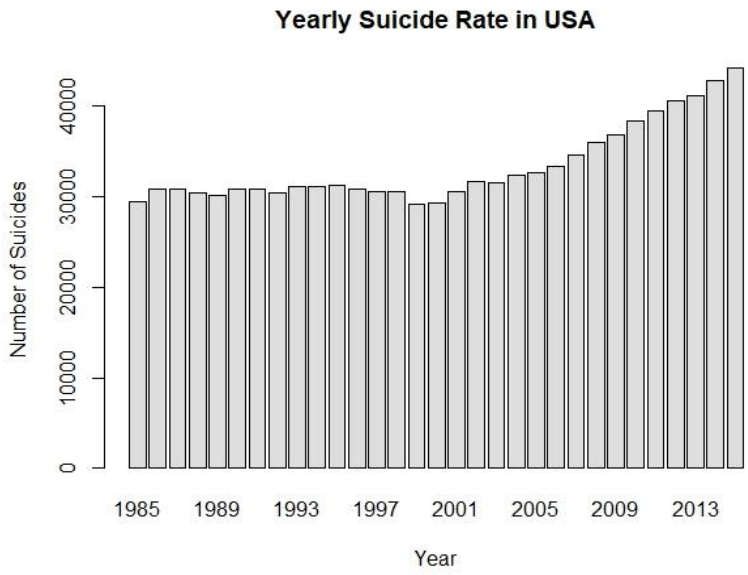
```
USA_data = data.frame(USA_data[,1:6])

USA_data[1:5,]

attach(USA_data)

num <- aggregate(suicides_no ~ year, USA_data, sum)

plot(num, main = "Yearly Suicide Number in USA", xlab = "Year", ylab = "Number of
Suicides", col = "blue", lwd = 2, ylim = c(25000,45000))

barplot(height = num$suicides_no, names.arg = num$year, main = "Yearly Suicide Number in
USA", xlab = "Year", ylab = "Number of Suicides", col = "gray87")

plot(num$year,num$suicides_no,type = "l", main = "Yearly Suicide Number in USA", xlab =
"Year", ylab = "Number of Suicides", col = "red", lwd = 2, ylim = c(25000,45000))
```

First of all, I have summed up all the 'suicides_no' column values according to the year. Then I made a scatter plot to see how it looks like. The scatter plot is



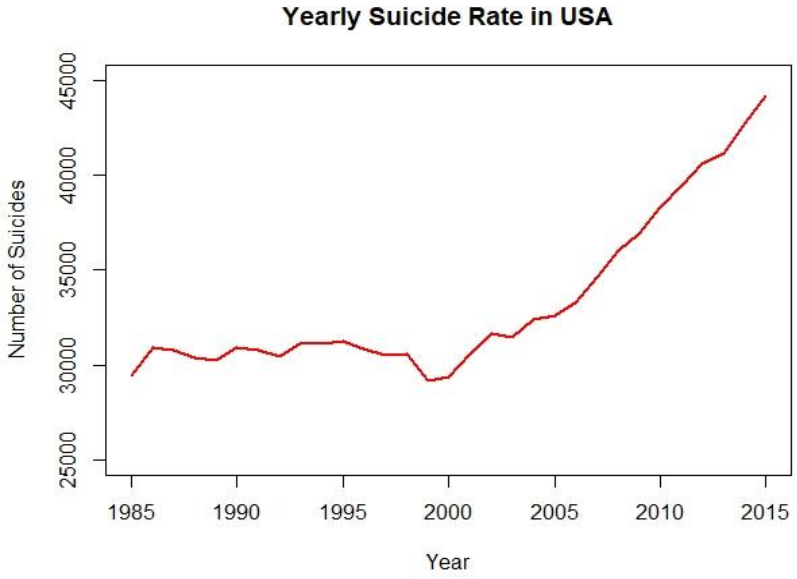
The scatter plot is showing the resulting points very clearly but if we want to compare any one trend with another trend then a scatter points cannot differentiate the overlapping points very easily.

Then I made a bar plot to see how the graph looks like.



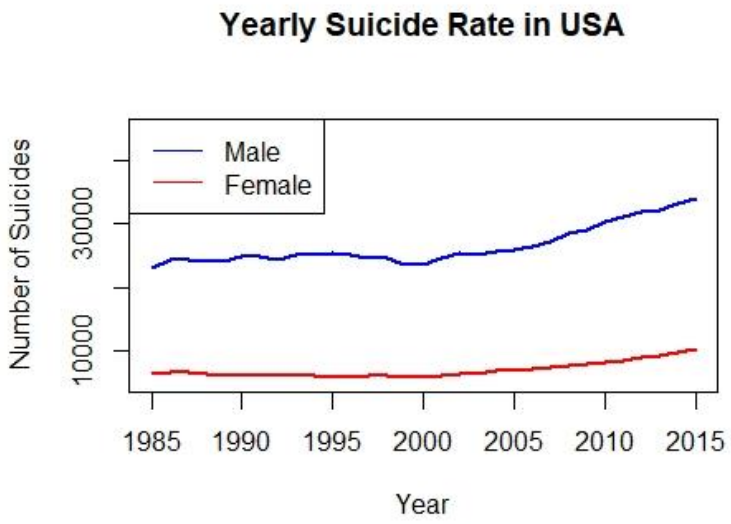
The bar graph is looked good. It shows the ups and downs of the data frequency very clearly. But if the X-axes has more data in future then it might be a problem to accommodate the bar of the years through the X-axes.

As a third and better option I made a scatter plot with line and found that the third one has less objection than the previous two.



We can easily see the trend of the data. If we need to compare any other dataset in future with this one, we can easily add lines with it to compare. Also adding more X or Y axes values will not be a problem to show any trend.

The comparison between male and female are shown below



Statistics show that the suicide rate is higher in male than in female.

Link of the suicide dataset

https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/version/1