

Final_Project_pt3.R

young

2022-06-04

#Introduction.

I set out to see trends among different possible factors of suicide including if the happiness score, various policy or laws had an effect on the number of suicides that occurred.

#The problem statement you addressed.

*# Is there correlations between policy or happiness score of a country and suicide.
Does GDP and Happiness have a correlation
does policy or law make a difference when it comes to the number of suicides?*

#How you addressed this problem statement

I looked at the different variables and tried to see if there was any significance between the different variables and the questions I was asking. I utilized both plots and models in order to try to get the insight from the data that was displayed.

#Analysis.

```
setwd("C:/Users/young/Desktop/Classes/DSC520/GIT/")
```

#import data

```
library(readxl)
```

```
library(ggplot2)
```

```
happiness_df <- read_excel("Final Project/Data/Appendix_2_Data_for_Figure_2.1.xls")
```

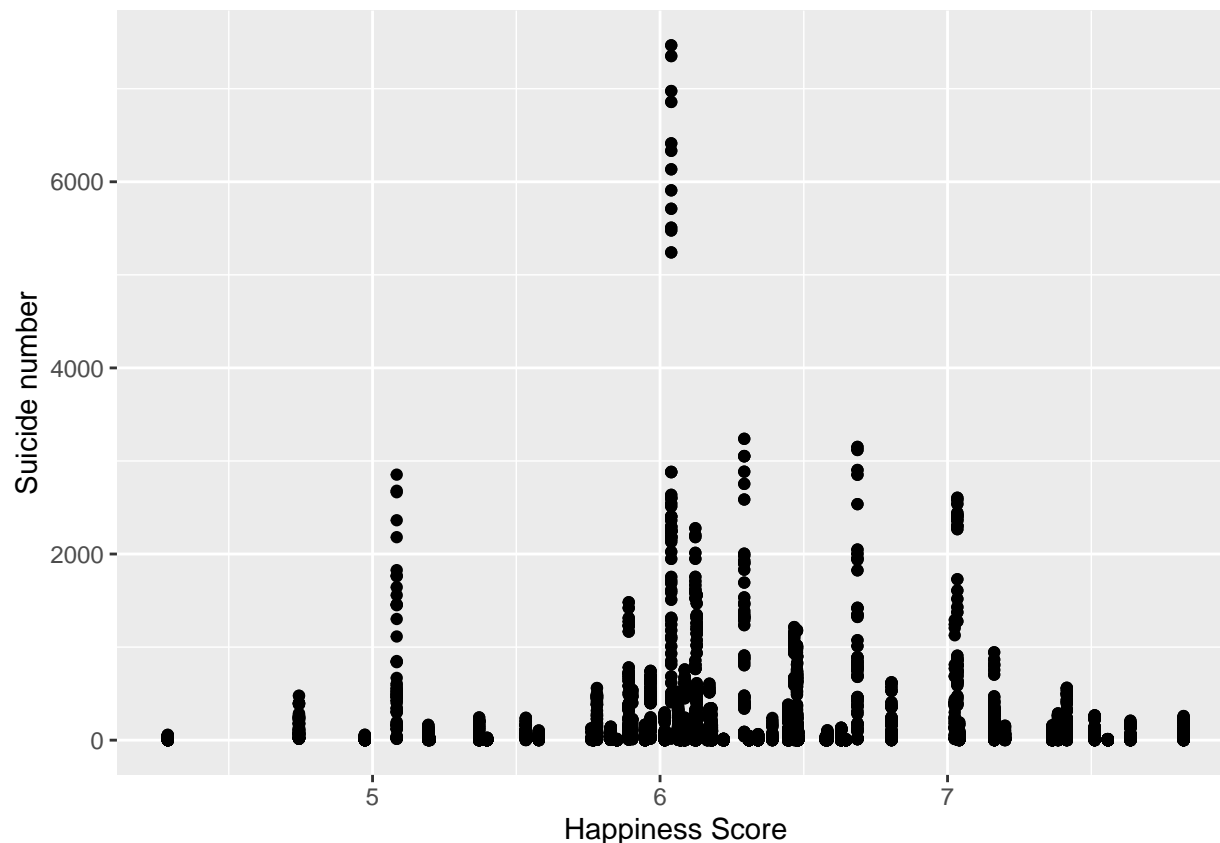
```
suicide_policy_df <-read_excel("Final Project/Data/suicide_policy.xls")
```

```
who_df <- read_excel("Final Project/Data/who_suicide_statistics.xlsx")
```

```
happiness_policy_df <- merge(happiness_df, suicide_policy_df,by = 'country')
```

```
combined_df <- merge(happiness_policy_df, who_df, by = 'country')
```

```
ggplot(combined_df, aes(x=happiness_score, y=suicides_no)) + labs(x ="Happiness Score", y="Suicide number")
```



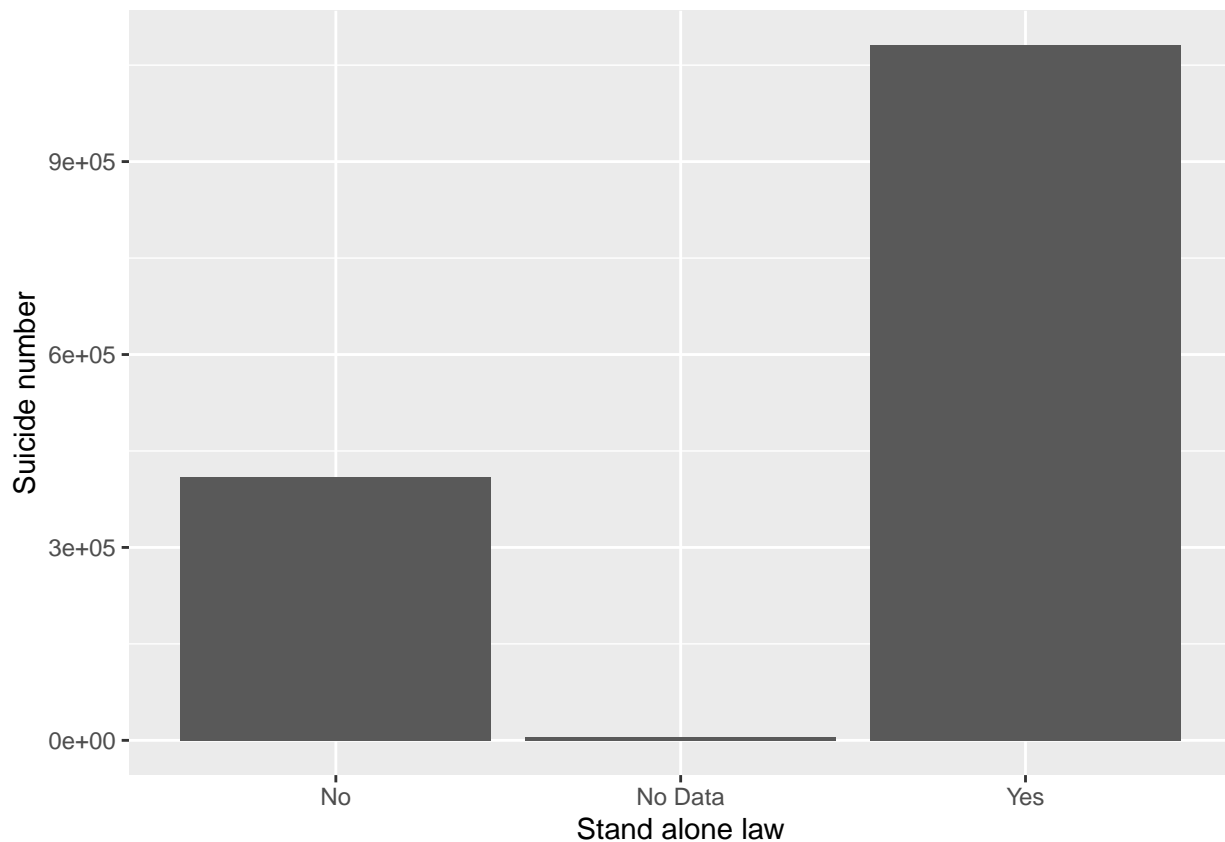
```
happiness_suicide_lm <- lm(happiness_score ~ suicides_no, combined_df)
summary(happiness_suicide_lm)
```

```
##
## Call:
## lm(formula = happiness_score ~ suicides_no, data = combined_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0485 -0.3680 -0.1153  0.4688  1.4877
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.336e+00  8.809e-03  719.253  <2e-16 ***
## suicides_no -1.127e-05  1.637e-05  -0.689    0.491
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7435 on 8074 degrees of freedom
## Multiple R-squared:  5.877e-05, Adjusted R-squared:  -6.508e-05
## F-statistic: 0.4745 on 1 and 8074 DF, p-value: 0.4909
```

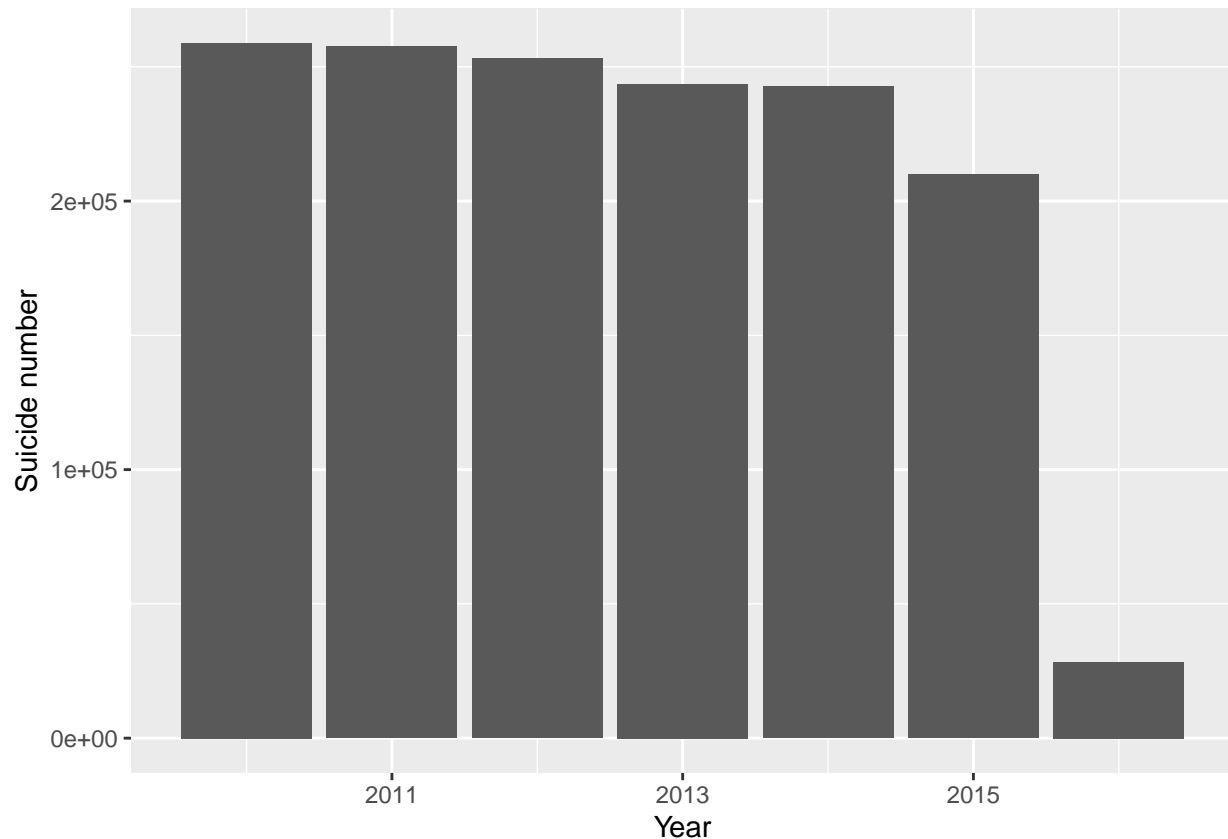
```
happiness_gdp_lm <- lm(happiness_score ~ GDP, combined_df)
summary(happiness_gdp_lm)
```

```
##
## Call:
## lm(formula = happiness_score ~ GDP, data = combined_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.61581 -0.25745 -0.03753  0.37306  1.14259
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.56704    0.04371   58.73  <2e-16 ***
## GDP          2.22223    0.02555   86.99  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5342 on 8074 degrees of freedom
## Multiple R-squared:  0.4838, Adjusted R-squared:  0.4837
## F-statistic: 7567 on 1 and 8074 DF, p-value: < 2.2e-16
```

```
ggplot(combined_df, aes(x=law_stand_alone, y=suicides_no)) + labs(x = "Stand alone law", y = "Suicide number")
```



```
ggplot(combined_df, aes(x=year.y, y=suicides_no)) + labs(x = "Year", y = "Suicide number") + geom_bar(stat="summary")
```



Looking at the plot between the suicide and happiness it actually becomes a pretty decent bell curve. That means that there can't be that much of a relation since we would think that if you have a lower happiness score then there would be higher suicides. The P Value of it was .4909 so it was not a significant value. The adjusted r-squared was $-6.508e-05$ which means that as one goes up the other goes down which is what we think. Likewise, looking at the GDP and happiness models, the p value was greater than 2.2 which means there is no relation between the happiness score and the gdp. The adjusted r-square was .48 which is fairly close to significance. The plots for the law with number of suicides, there was a much higher number of suicides when there was a policy and same with the laws.

#Implications.

#The implications of these statistics is that there is no implications between the different variables. The significance between the several variables and data was fairly low. The overall trend of the number of suicides over the years was downward with 2015 being significantly less than 2010.

#Limitations.

There are a few number of limitations one being the data itself. The number of suicides that are recorded may not be totally accurate in all countries, especially some countries where it is still looked down upon to commit suicide and it may be misrepresented or reported wrongly. It is also incomplete in a few different ways and other data had to be cleaned or entered to make it valid which leaves a lot of holes in accuracy.

#Concluding Remark

#There seems to be no reasonable significance between either the number of suicides and the happiness, or the number of suicides and the law or policies that a country has enacted. It shows a large number of suicides for

*# but the actual statistical significance seems fairly low. Looking at the number of suicides per year,
to have incomplete data or it was a very good year. I am leaning towards the data being incomplete as*