World Happiness And Corruption 2015-2020

Or are we too corrupt to be happy?

Saikat Bera, MDS202228, saikatb@cmi.ac.in

2022-10-30

Abstract :-

Both corruption and happiness are of concern to academics, governments, and policymakers. The World Bank and UNDP define corruption as "the misuse or the abuse of public office for private gain". World happiness is all about prosperity and economic growth. The real obstacle to world economic growth and stability are corrupt international leaders. Corruption occurs in developed as well as developing countries due to the opposing power of authorities and private interests. Therefore, the danger of corruption has become an attractive issue of debate in recent years and I chose to dig into this.

Introduction :-   Finding correlations between other variables (like government trust, freedom, family size, social support, CPI score etc.) and happiness score as well as visualizing which variables mostly affect the happiness rank would be our primary challenge.

Our Hypothesis :- World happiness is correlated with government corruption. In particular they have a strong negative correlation. That means high government corruption results in lower happiness in their citizens.

Our Data <- Link to the Data Set

Data-Set Description :-   ELIAS TURK preprocessed and aggregated all valid data for countries with existent data from the year 2015 to 2020 using Pandas and he added the corruption perception index taken from transparency. Joining CPI(Corruption Perception Index) was just a curiosity to find underlying answers and findings. ELIAS TURK then merged the data sets based on Country and Year to transform the data into a long data format. Government trust may already be taken from CPI scores, but CPI scores make it clearer to understand and contrast it with the dependent variables in our case, happiness score.

```
data = read.csv('E:\\Coding\\Sourish Das\\Visualisation Project\\Data\\WorldHappiness_Corruption_2015_20
# Loading the data
df = as.data.frame(data)
attach(df)
```

```
head(df)
summary(df)
```

Variables of Interest :-

1. **Country :**   There is data on 132 unique countries over years.
2. **Happiness Score :**   An average of responses to the primary life evaluation question from the Gallup World Poll (GWP). (Scale : 0-10)
3. **GDP per Capita :**   Gross Domestic Product (GDP) per capita shows a country's GDP divided by its total population.
4. **Family :**   The extent to which Family contributes to the calculation of the Happiness Score.

5. **Health :** The extent to which Health(Life Expectancy) contributes to the calculation of the Happiness Score.
6. **Freedom :** The extent to which Freedom contributes to the calculation of the Happiness Score.
7. **Generosity :** A numerical value calculated based on poll participants' perceptions of generosity in their country.
8. **Government Trust :** The extent to which Perception of Corruption contributes to Happiness Score.
9. **Dystopia Residual :** A score based on a hypothetical comparison to the world's saddest country.
10. **Continent :**
11. **Year :** The data is collected across 6 years, i.e., 2015-2020.
12. **Social Support :** Social support is the perception and actuality that one is cared for, has assistance available from other people, and most popularly, that one is part of a supportive social network.
13. **CPI Score :** Corruption perception index (CPI) is an index which ranks countries by their perceived levels of public sector corruption, as determined by expert assessments and opinion surveys.

Note that- The missing data is filled with 0's.

Graphical Analysis :- Here are some graphs to understand the data better.

```
ggplotbycont = df %>%
  ggplot(aes(happiness_score, fill = continent)) +
  geom_histogram(bins = 20, col = 'black', alpha = 0.5 ) +
  facet_wrap(~continent) +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5)) +
  labs(title = "Histogram Of Happiness Score",
       x = "Happiness Score", y = "Count")
```

```
ggplotwithcont = df %>%
  ggplot(aes(happiness_score, fill = continent, group=continent)) +
  geom_histogram(bins = 20, col = 'black', alpha = 0.5) +
  theme(legend.background = element_rect(fill = "white", size = 1, colour = "white"),
    legend.justification = c(0, 1),
    legend.position = c(0, 1),
    legend.key.size = unit(1.25, 'mm'),
    legend.title = element_text(size=10),
    legend.text = element_text(size=8),
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5)) +
  labs(title = "Histogram Of Happiness Score \n(Grouped By Continent)",
       x = "Happiness Score", y = "Count", fill = "Continent")
```

```
grid.arrange(ggplotbycont, ggplotwithcont, ncol = 2)
```

```
df %>%
  group_by(continent, Year) %>%
  summarise(mean_happy = mean(happiness_score)) %>%
  ggplot(aes(as.double(Year), mean_happy, color = continent)) +
    geom_line() +
    geom_point(size = 4, alpha = 0.5) +
    labs(title = "Change of mean Happiness over Years 2015-2020",
         subtitle = "For different Continents",
         x = "Year", y = "Mean Happiness Score") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```
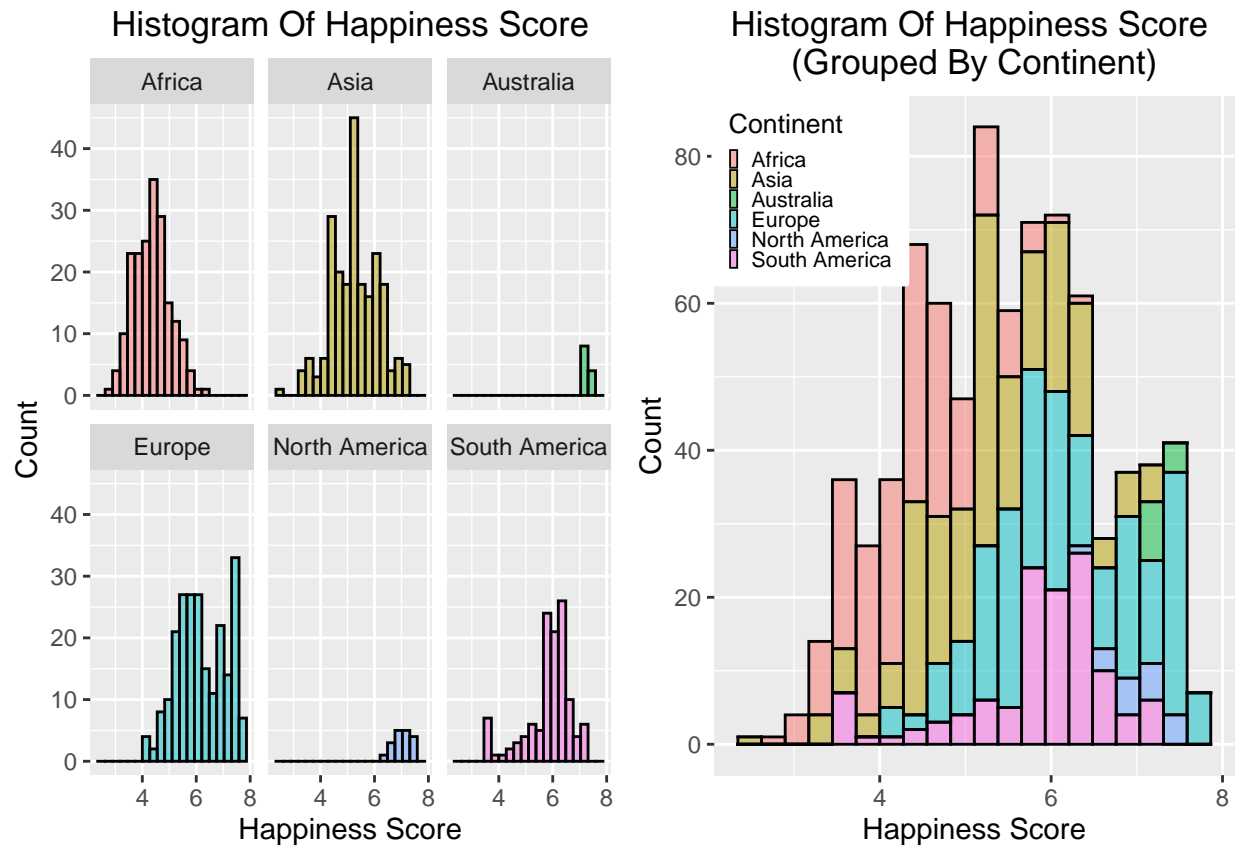
Figure 1: In this figure we explore Histogram of Happiness Score for different Continent and also we see an agreegated histogram of Happiness Score with different Continent indexed by different color
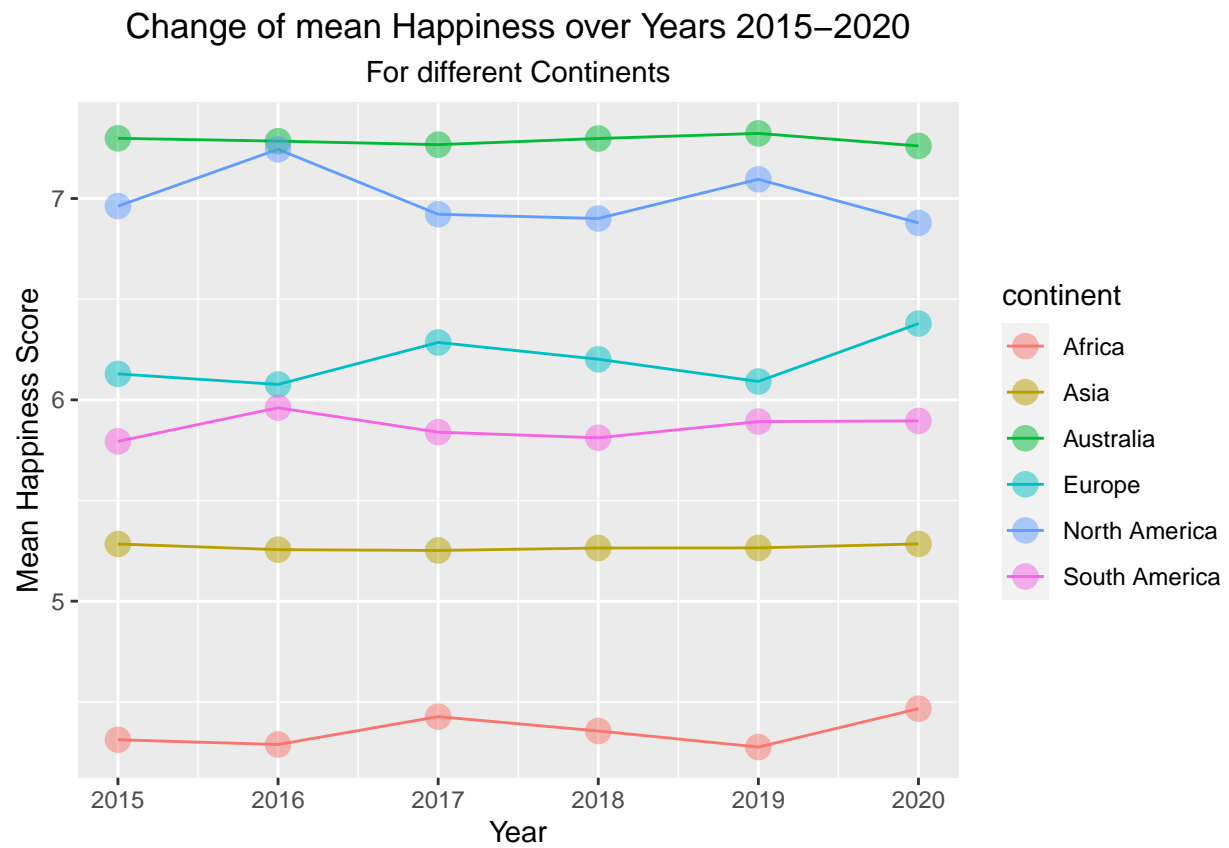
Figure 2: By this figure we see the cange of Mean Happiness Score over the years 2015-2020 for different Continents

```
df %>%
  filter(Year %in% c(max(Year), min(Year))) %>%
  select(Country, happiness_score, Year) %>%
  pivot_wider(names_from = Year, values_from = happiness_score, names_prefix = "year_") %>%
  group_by(Country) %>%
  summarise(change_in_happiness = year_2020 - year_2015) %>%
  mutate(Country = fct_lump(Country, 10, w = abs(change_in_happiness))) %>%
  mutate(Country = fct_reorder(Country, -change_in_happiness)) %>%
  filter(Country != "Other") %>%
  ggplot(aes(change_in_happiness, Country, fill = Country)) +
  geom_col() +
  scale_x_continuous(labels = comma) +
  labs(title = "Change in Happiness Score from 2015 to 2020",
       subtitle = "For Top 10 countries with most change",
       x = "Change in Happiness Score") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```
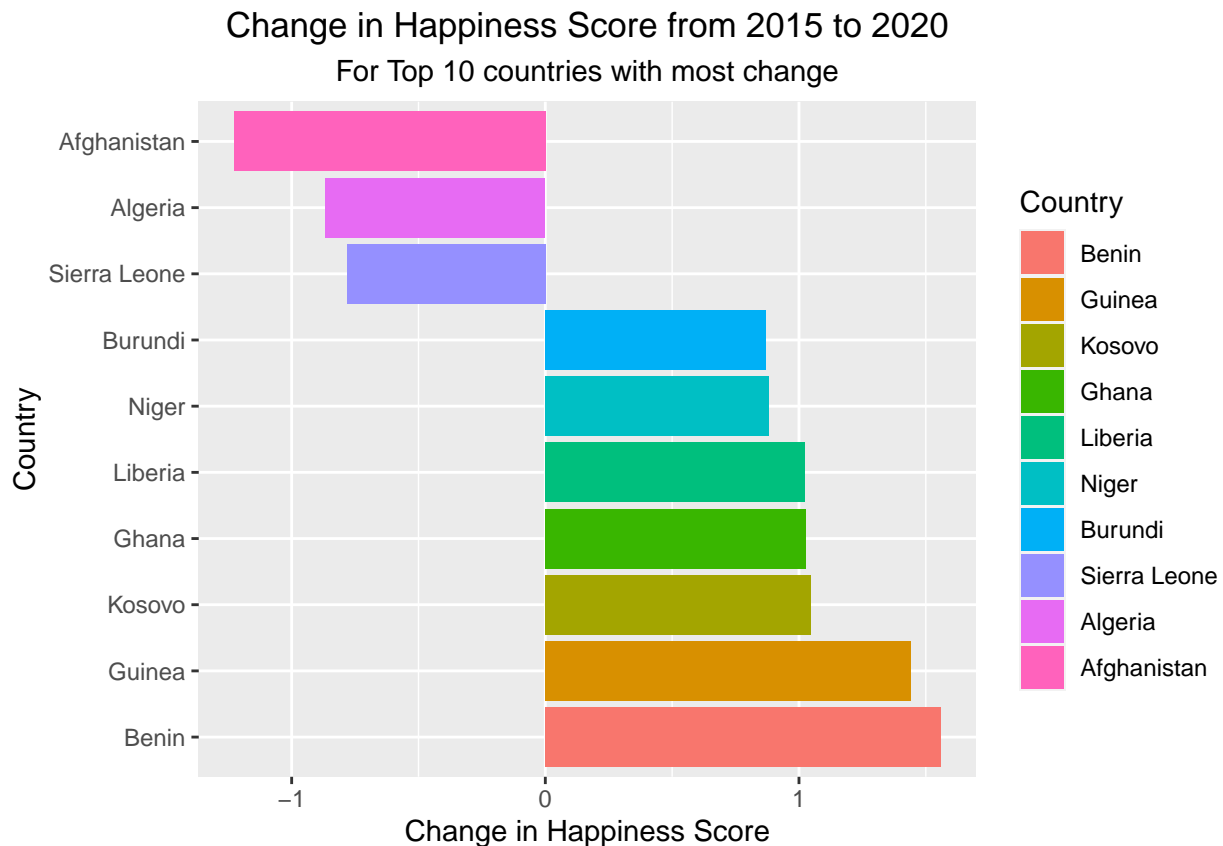

Figure 3: In this bar diagram we see the absolute change in Happiness Score from 2015 to 2020 for top 10 countries with most absolute change in Happiness Score

```
scatterplotlegend = function(col1, col2){
  df %>%
    ggplot(aes({{col1}}, {{col2}}, color = continent)) +
    geom_point()}
```

```r
scatterplot = function(col2, str1){
  df %>%
    ggplot(aes(happiness_score, {{col2}}, color = continent)) +
    geom_point() +
    theme(legend.position = "none") +
    labs(x = "Happiness Score", y = str1)}

s1 = scatterplot(gdp_per_capita, "GDP per capita")
s2 = scatterplot(health, "Health")
s3 = scatterplot(freedom, "Freedom")
s4 = scatterplot(generosity, "Generosity")
s5 = scatterplot(government_trust, "Government Trust")
legend <- get_legend(scatterplotlegend(happiness_score, gdp_per_capita))
grid.arrange(s1, s2, s3, s4, s5, legend,
             layout_matrix = matrix(c(1, 2, 6, 3, 4, 5), byrow = TRUE, ncol = 3),
             top = textGrob("Relationship between Happiness Score and Different Variables (w.r.t. Conti
```
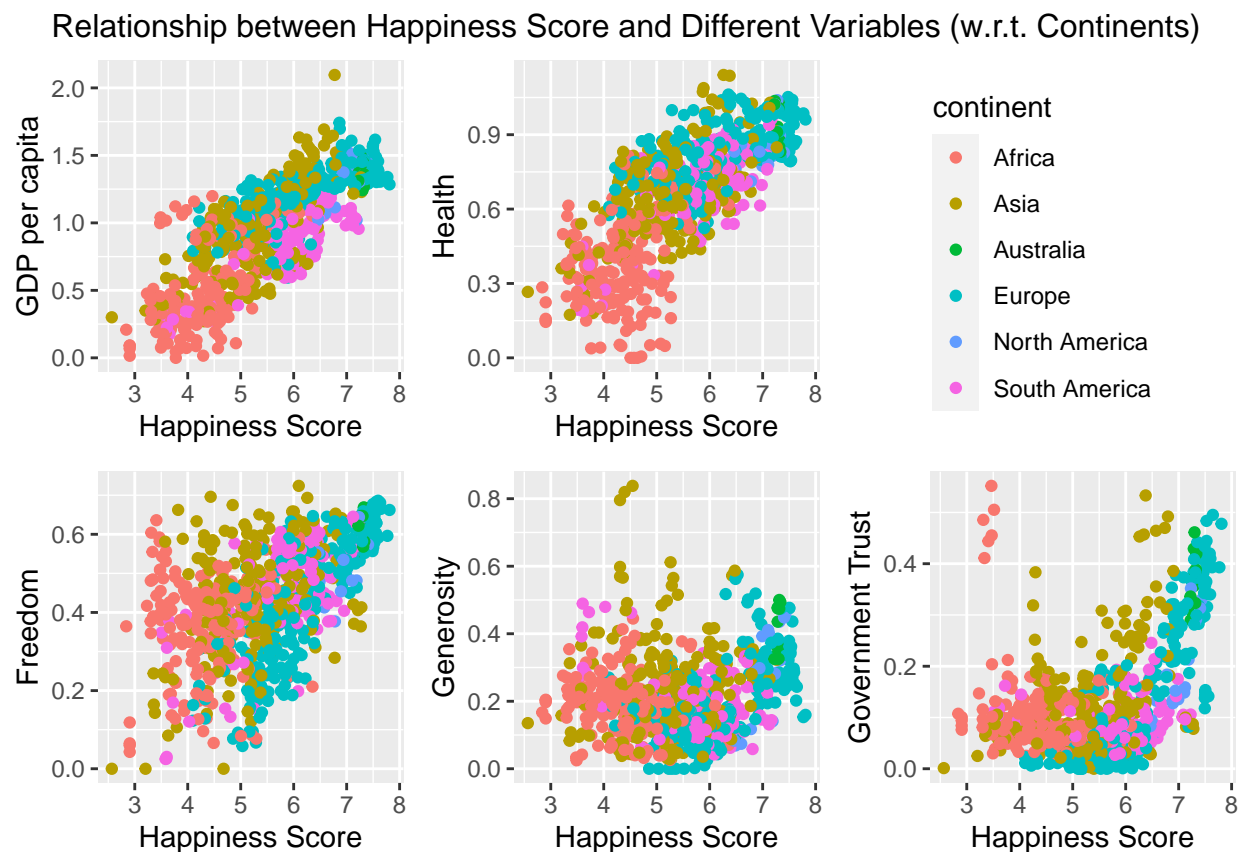

Figure 4: In this figure we visualise simple scatterplot between Happiness Score and different other factors like GDP per capita, Health, Freedom, Generosity and Government Trust

```r
first = function(x) {df %>%
  arrange(desc(gdp_per_capita)) %>%
  filter(Year == x) %>%
  slice(1:10) %>%
  ggplot(aes(gdp_per_capita, happiness_score, fill = gdp_per_capita)) +
```

```
    geom_point(aes(size = gdp_per_capita, col = viridis(10))) +
    geom_line(col = 'black') +
    geom_smooth(col = 'purple') +
    theme(legend.position = "none", plot.title = element_text(hjust = 0.5, size = 8)) +
    labs(title = paste("Year", x, sep = " "), x = "Gdp per Capita", y = "Happiness Score")}
grid.arrange(first(2015), first(2016), first(2017),
             first(2018), first(2019), first(2020), ncol = 3,
             top = textGrob("Relationship between Gdp per Capita And Happiness Score"))
```
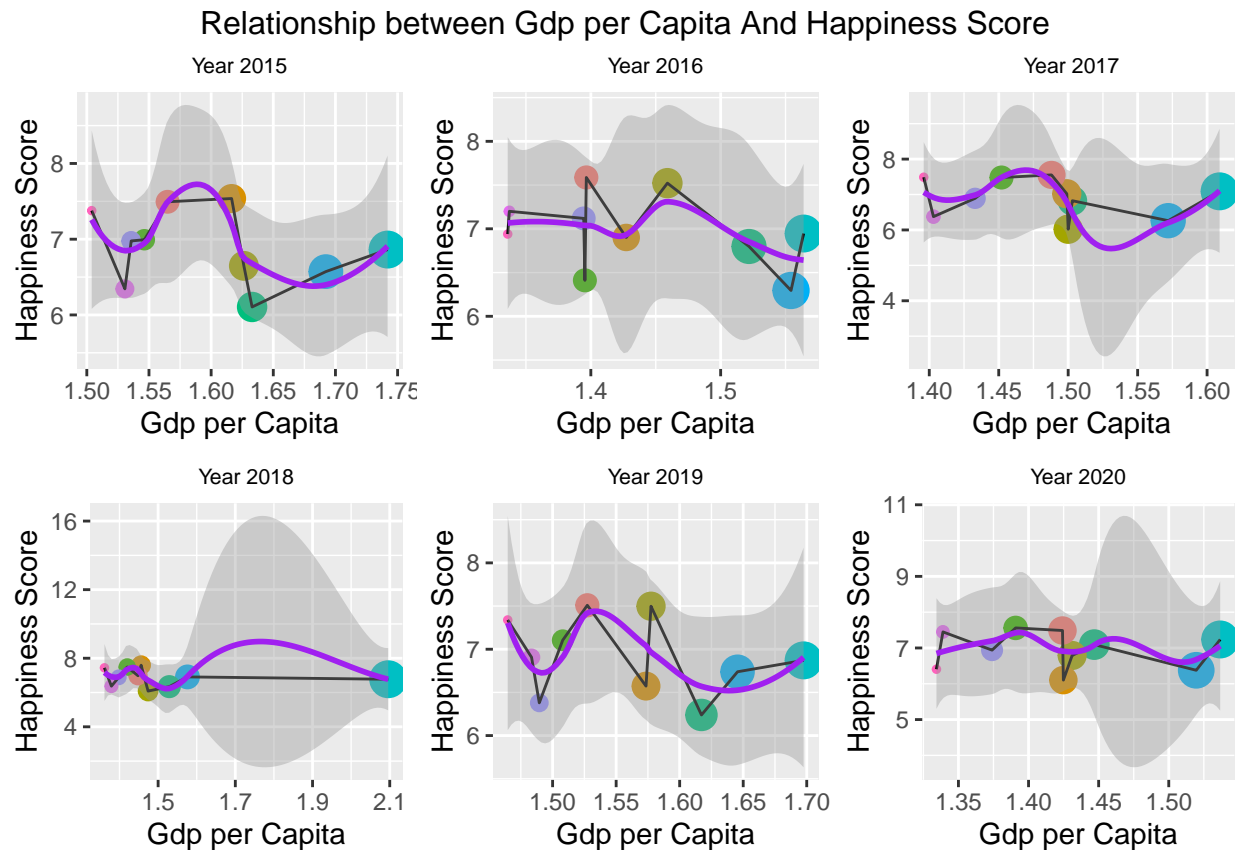

Figure 5: In this figure we see relationship between GDP per capita and Happiness score for top top 10 countries with most GDP per capita each year from 2015 to 2020

Summary of Graphical Analysis :- From figure-1 we see that for the continent of Africa the Happiness Scores are concentrated in lower end of the spectrum, at the same time European countries have Happiness scores at the higher end of the scale. From figure-2 we see the same thing that over the years, the mean happiness score for European and North American countries is much higher than for Asian or African countries. From the figure-3 we see that in countries like Afghanistan Happiness score reduces from 2015 to 2020 whereas countries like Benin and Guinea saw significant improvement in Happiness score from 2015 to 2020. From figure-4 we see Happiness score and GDP per capita have an almost linear relationship, and Happiness score and Government trust have a polynomial relationship with some outliers present.

Conclusion :- In conclusion, I want to say that this is only a preliminary analysis of the Data set. I will perform a more detailed analysis using R Shiny in the future.

I would also like to thank professor Sourish Das for his guidance.