Final Project (Group 2)

Group 2

2024-05-14

- Research Question/Hypothesis: What variable in the world happiness report (family, health, trust, generosity, and economics) has the greatest effect on a nation's happiness score?
- Hypothesis: Economics plays the largest role in a nation's happiness score.

```
library(readxl)
library(dplyr)
library(ggplot2)
library(tidyr)

data <- read_excel("2019.xls")

colnames(data)
```

```
## [1] "Overall rank"                "Country or region"
## [3] "Score"                       "GDP per capita"
## [5] "Social support"              "Healthy life expectancy"
## [7] "Freedom to make life choices" "Generosity"
## [9] "Perceptions of corruption"
```

```
library(readxl)

data <- read_excel("2019.xls")

print(colnames(data))
```

```
## [1] "Overall rank"                "Country or region"
## [3] "Score"                       "GDP per capita"
## [5] "Social support"              "Healthy life expectancy"
## [7] "Freedom to make life choices" "Generosity"
## [9] "Perceptions of corruption"
```

```
data <- data %>%
  rename(
    Economy = `GDP per capita`,
    Social = 'Social support',
    Health = `Healthy life expectancy`,
    Freedom = `Freedom to make life choices`,
    Corruption = 'Perceptions of corruption',
    Happiness_Score = `Score`
  )
print(colnames(data))
```

```
## [1] "Overall rank"      "Country or region" "Happiness_Score"
## [4] "Economy"           "Social"            "Health"
## [7] "Freedom"           "Generosity"        "Corruption"
```

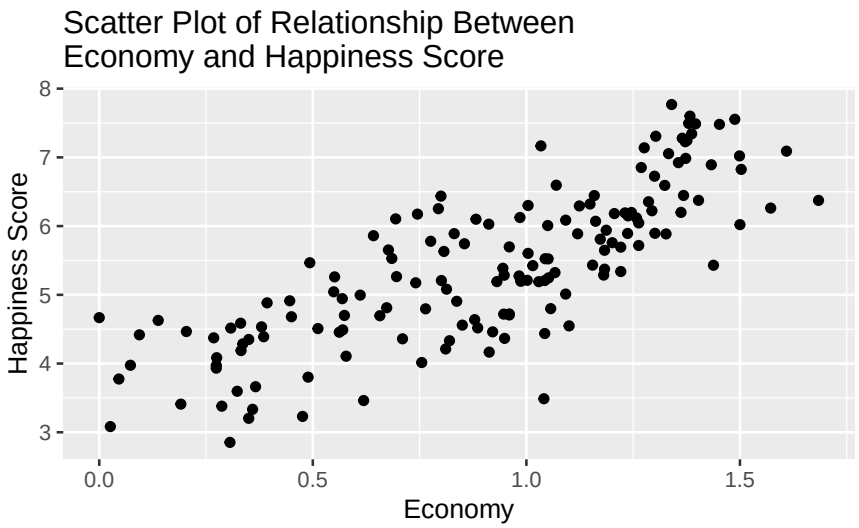
```
  head(
    select(data, Economy, Social, Health, Freedom, Corruption, Happiness_Score)
  )
```

Economy	Social	Health	Freedom	Corruption	Happiness_Score
1.340	1.587	0.986	0.596	0.393	7.769
1.383	1.573	0.996	0.592	0.410	7.600
1.488	1.582	1.028	0.603	0.341	7.554
1.380	1.624	1.026	0.591	0.118	7.494
1.396	1.522	0.999	0.557	0.298	7.488
1.452	1.526	1.052	0.572	0.343	7.480

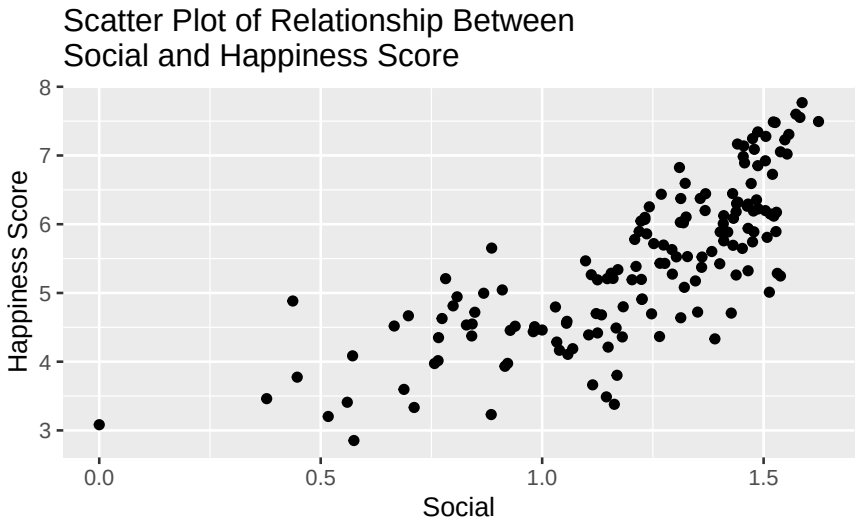
[Module 2: Junhyung Kim, Jiho Lee]

*Scatter Plot

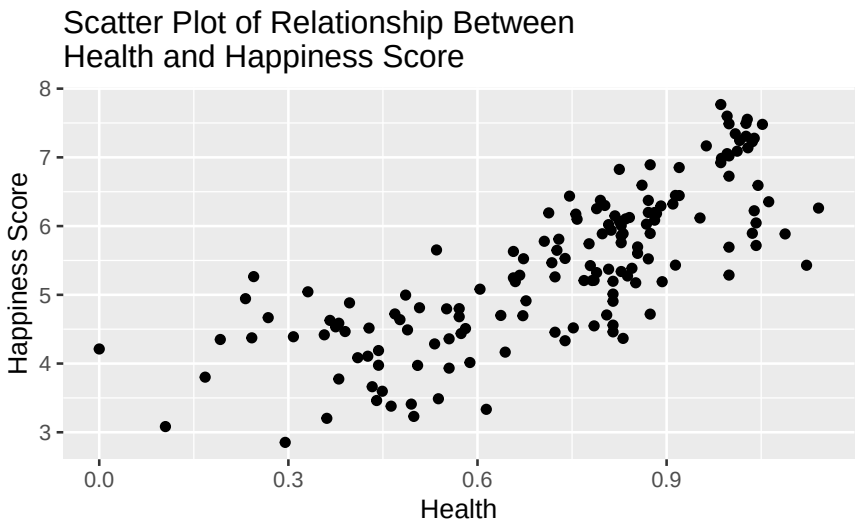
```
data %>%
ggplot() +
geom_point(mapping = aes (x = Economy, y= Happiness_Score)) +
labs(title =
       "Scatter Plot of Relationship Between
Economy and Happiness Score",
x = "Economy", y = "Happiness Score")
```




```
data %>%
ggplot() +
geom_point(mapping = aes (x = Social, y= Happiness_Score)) +
labs(title =
       "Scatter Plot of Relationship Between
Social and Happiness Score",
x = "Social", y = "Happiness Score")
```

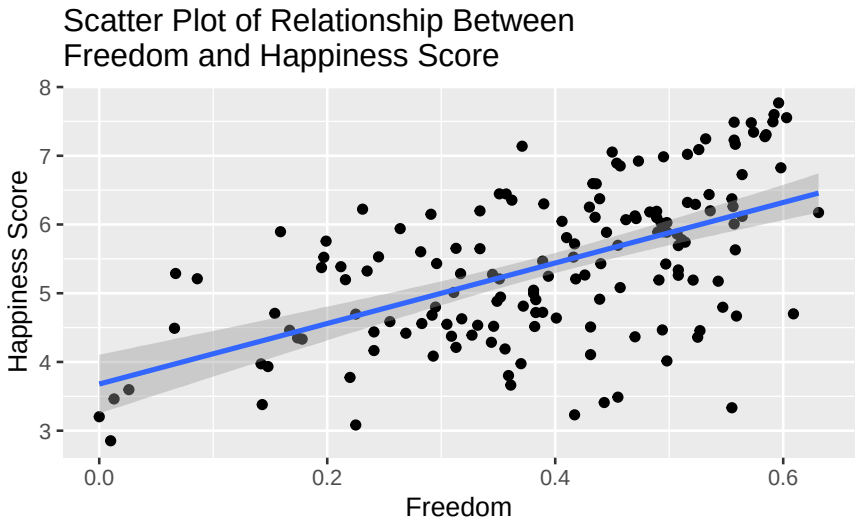



```
data %>%
ggplot() +
geom_point(mapping = aes (x = Health, y= Happiness_Score)) +
labs(title =
       "Scatter Plot of Relationship Between
Health and Happiness Score",
x = "Health", y = "Happiness Score")
```



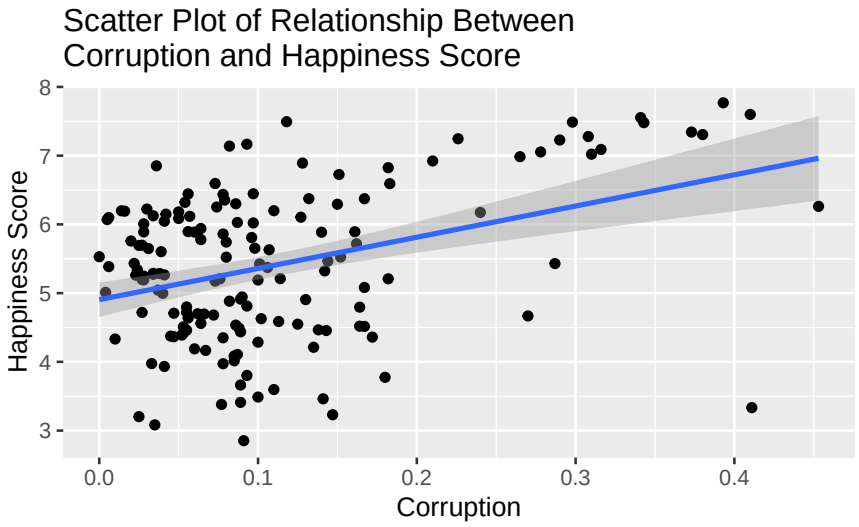

```
qplot(x = Freedom, y= Happiness_Score, data = data,
      geom = c("point", "smooth"), method = "lm") +
labs(title =
       "Scatter Plot of Relationship Between
Freedom and Happiness Score",
x = "Freedom", y = "Happiness Score")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



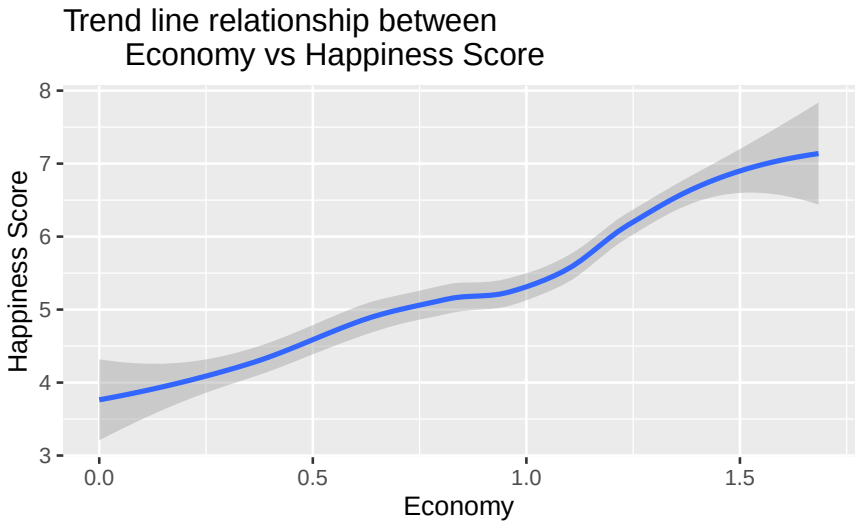

```
qplot(x = Corruption, y= Happiness_Score, data = data,
      geom = c("point", "smooth"), method = "lm") +
labs(title =
       "Scatter Plot of Relationship Between
Corruption and Happiness Score",
x = "Corruption", y = "Happiness Score")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
data %>%
  ggplot() +
  geom_smooth(mapping = aes(x = Economy, y = Happiness_Score)) +
  labs(x = "Economy", y = "Happiness Score",
       title="Trend line relationship between
       Economy vs Happiness Score")
```

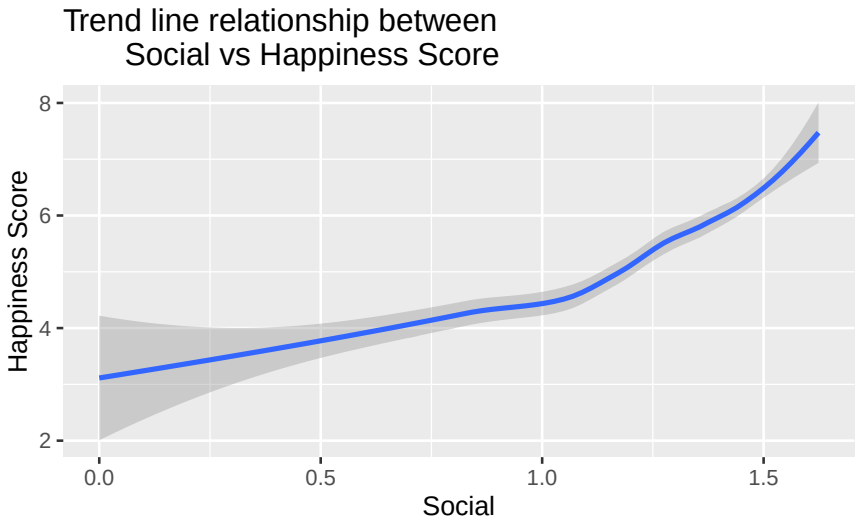
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
data %>%
  ggplot() +
  geom_smooth(mapping = aes(x = Social, y = Happiness_Score)) +
```

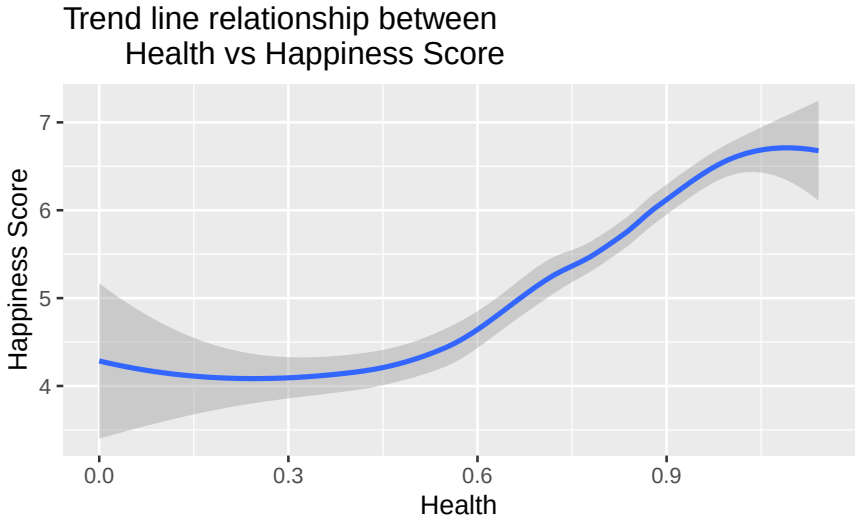
```
  labs(x = "Social", y = "Happiness Score",
       title="Trend line relationship between
       Social vs Happiness Score")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



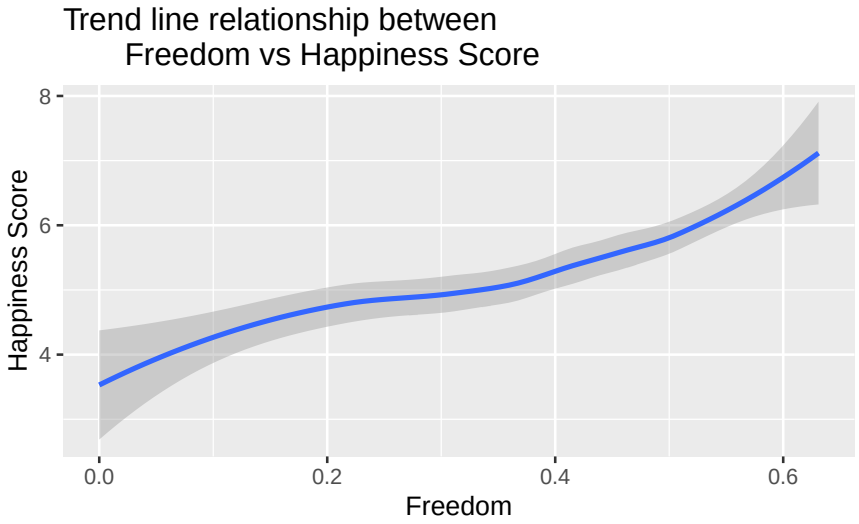

```
data %>%
  ggplot() +
  geom_smooth(mapping = aes(x = Health, y = Happiness_Score)) +
  labs(x = "Health", y = "Happiness Score",
       title="Trend line relationship between
       Health vs Happiness Score")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



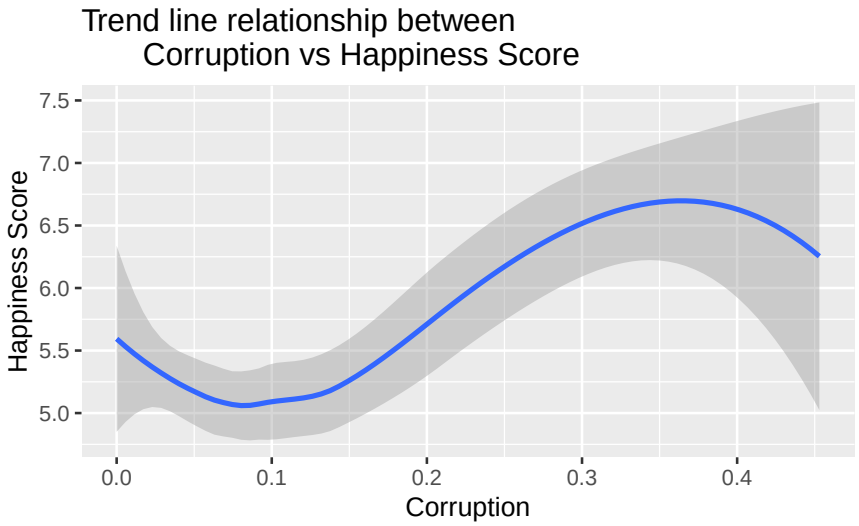
```
data %>%
  ggplot() +
  geom_smooth(mapping = aes(x = Freedom, y = Happiness_Score)) +
  labs(x = "Freedom", y = "Happiness Score",
       title="Trend line relationship between
       Freedom vs Happiness Score")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



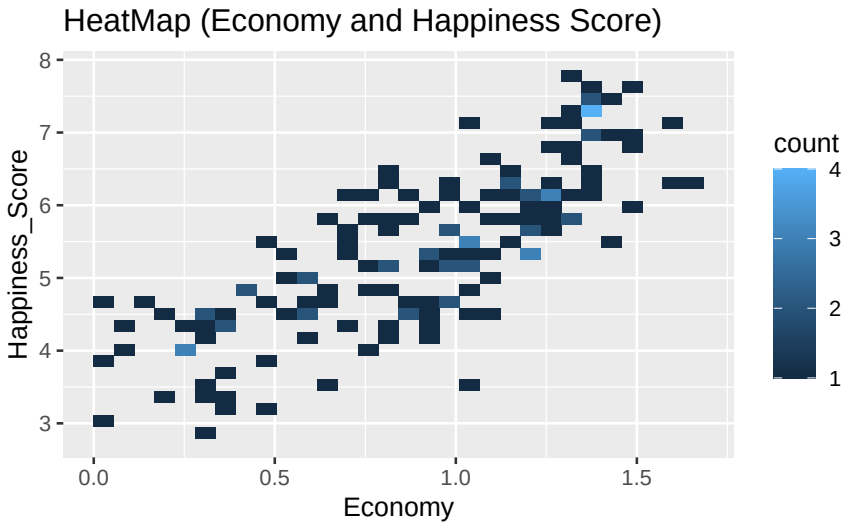

```
data %>%
  ggplot() +
  geom_smooth(mapping = aes(x = Corruption, y = Happiness_Score)) +
  labs(x = "Corruption", y = "Happiness Score",
       title="Trend line relationship between
       Corruption vs Happiness Score")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

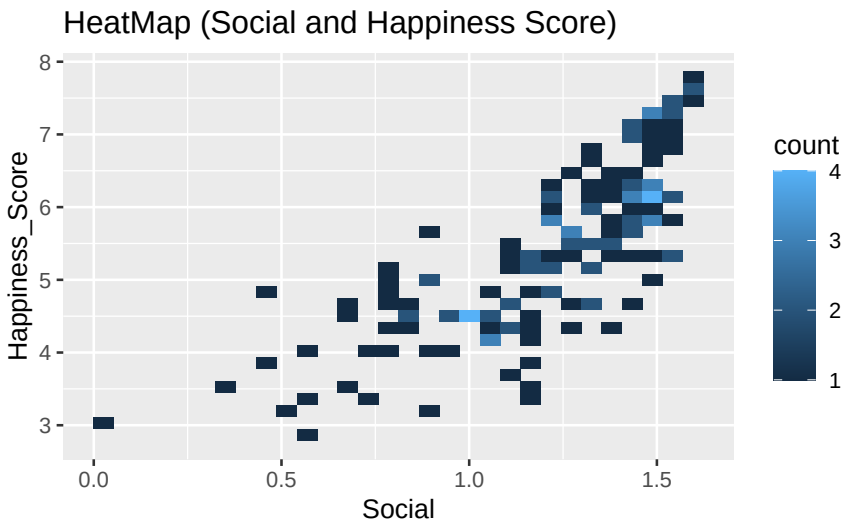


*HeatMap

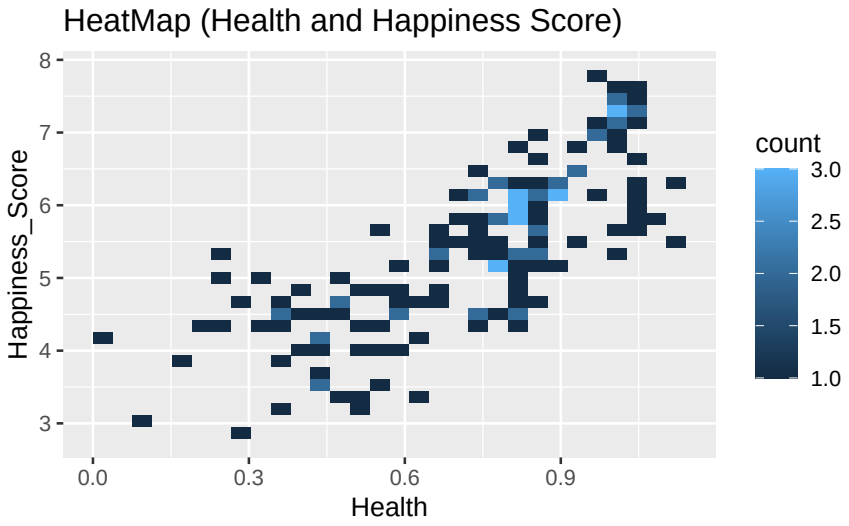
```
data %>%
ggplot() +
geom_bin2d(mapping = aes(x = Economy, y = Happiness_Score)) +
labs(title = "HeatMap (Economy and Happiness Score)",x= "Economy",
     y= "Happiness_Score")
```



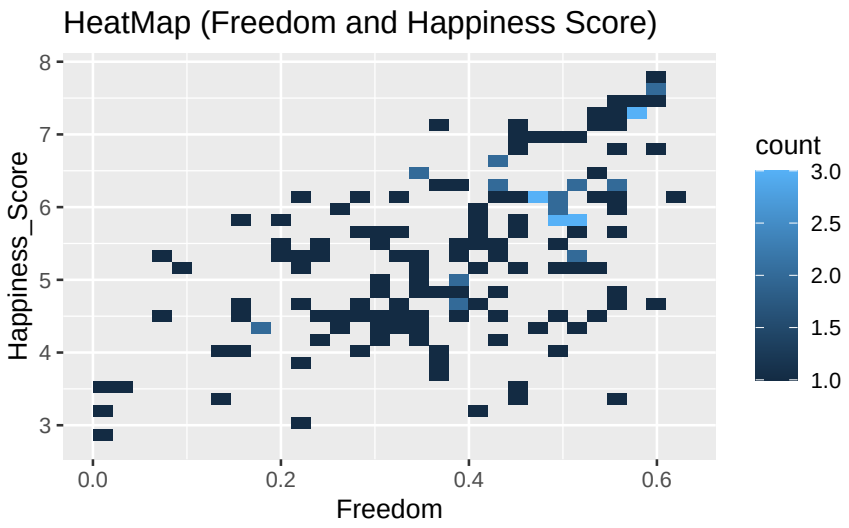

```
data %>%
ggplot() +
geom_bin2d(mapping = aes(x = Social, y = Happiness_Score)) +
labs(title = "HeatMap (Social and Happiness Score)",x= "Social",
     y= "Happiness_Score")
```



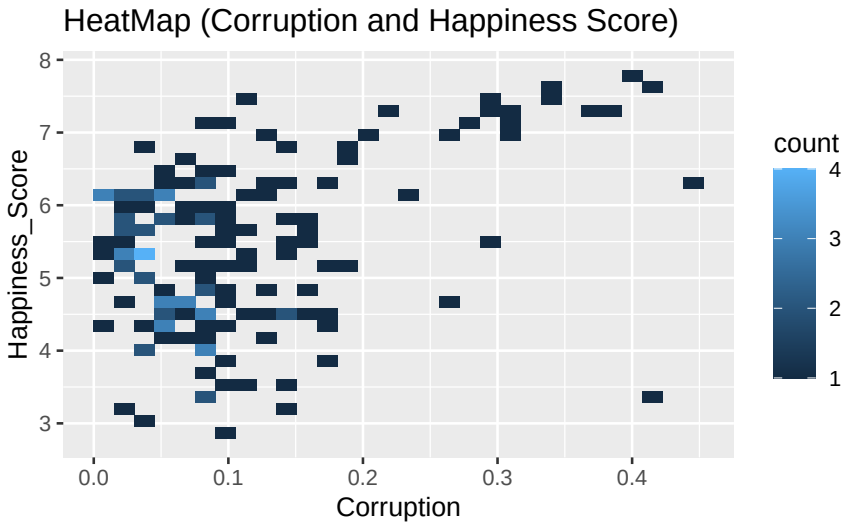

```
data %>%
ggplot() +
geom_bin2d(mapping = aes(x = Health, y = Happiness_Score)) +
labs(title = "HeatMap (Health and Happiness Score)",x= "Health",
     y= "Happiness_Score")
```



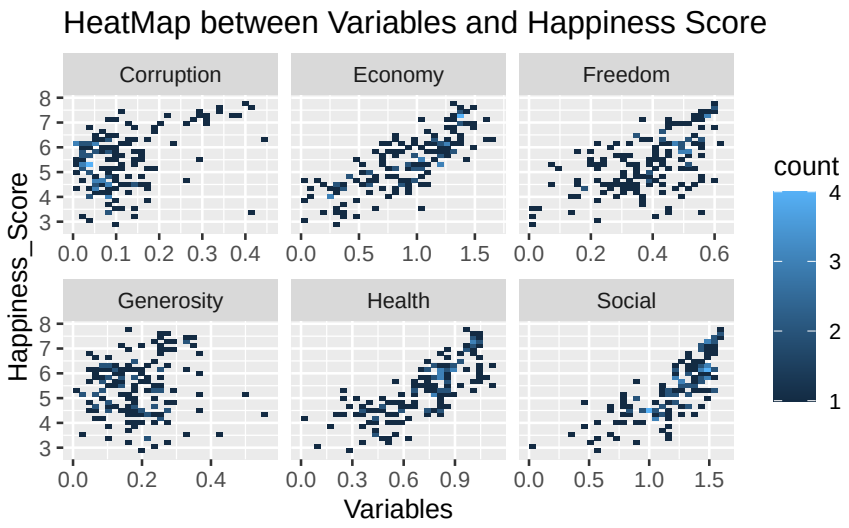

```
data %>%
ggplot() +
geom_bin2d(mapping = aes(x = Freedom, y = Happiness_Score)) +
labs(title = "HeatMap (Freedom and Happiness Score)",x= "Freedom",
     y= "Happiness_Score")
```




```
data %>%
ggplot() +
geom_bin2d(mapping = aes(x = Corruption, y = Happiness_Score)) +
labs(title = "HeatMap (Corruption and Happiness Score)",x= "Corruption",
     y= "Happiness_Score")
```




```
data %>%
pivot_longer(cols = Economy:Corruption, names_to = "Variable",
             values_to = "value") %>%
ggplot() +
geom_bin2d(mapping = aes(x = value, y = Happiness_Score)) +
labs(title = "HeatMap between Variables and Happiness Score",
     x= "Variables") +
facet_wrap(~ Variable, scales = "free_x")
```



======= [Module 4: Eugene Kim, Harold Lee - Explanatory Data Analysis]

```
str(data)
```

```
## tibble [156 x 9] (S3: tbl_df/tbl/data.frame)
##  $ Overall rank     : num [1:156] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Country or region: chr [1:156] "Finland" "Denmark" "Norway" "Iceland" ...
##  $ Happiness_Score  : num [1:156] 7.77 7.6 7.55 7.49 7.49 ...
##  $ Economy          : num [1:156] 1.34 1.38 1.49 1.38 1.4 ...
##  $ Social           : num [1:156] 1.59 1.57 1.58 1.62 1.52 ...
##  $ Health           : num [1:156] 0.986 0.996 1.028 1.026 0.999 ...
##  $ Freedom          : num [1:156] 0.596 0.592 0.603 0.591 0.557 0.572 0.574 0.585 0.584 0.53
##  $ Generosity       : num [1:156] 0.153 0.252 0.271 0.354 0.322 0.263 0.267 0.33 0.285 0.244
##  $ Corruption       : num [1:156] 0.393 0.41 0.341 0.118 0.298 0.343 0.373 0.38 0.308 0.226
```

```
head(
    select(data, Economy, Social, Health, Freedom, Corruption,
           Happiness_Score)
  )
```

Economy	Social	Health	Freedom	Corruption	Happiness_Score
1.340	1.587	0.986	0.596	0.393	7.769
1.383	1.573	0.996	0.592	0.410	7.600
1.488	1.582	1.028	0.603	0.341	7.554
1.380	1.624	1.026	0.591	0.118	7.494
1.396	1.522	0.999	0.557	0.298	7.488
1.452	1.526	1.052	0.572	0.343	7.480

```
tail(select(data, Economy, Social, Health, Freedom, Corruption,
            Happiness_Score)
  )
```

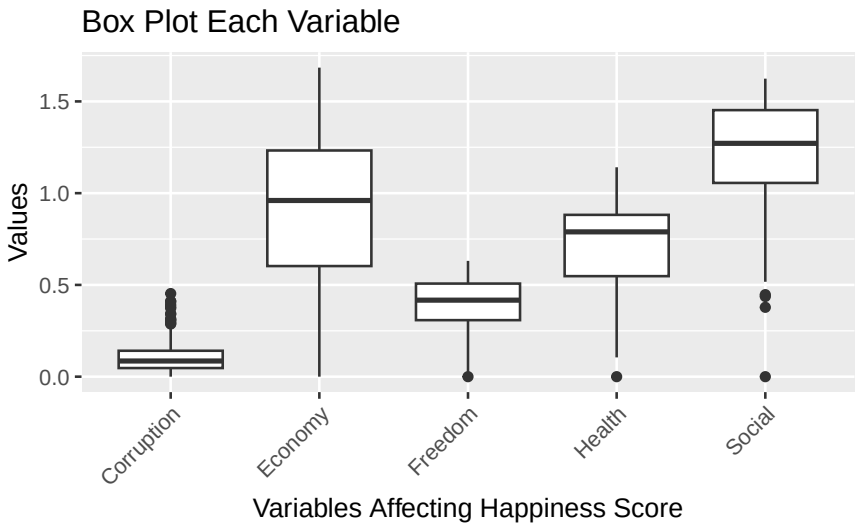
Economy	Social	Health	Freedom	Corruption	Happiness_Score
0.287	1.163	0.463	0.143	0.077	3.380
0.359	0.711	0.614	0.555	0.411	3.334
0.476	0.885	0.499	0.417	0.147	3.231
0.350	0.517	0.361	0.000	0.025	3.203
0.026	0.000	0.105	0.225	0.035	3.083
0.306	0.575	0.295	0.010	0.091	2.853

*Summary statistics

*Box Plot

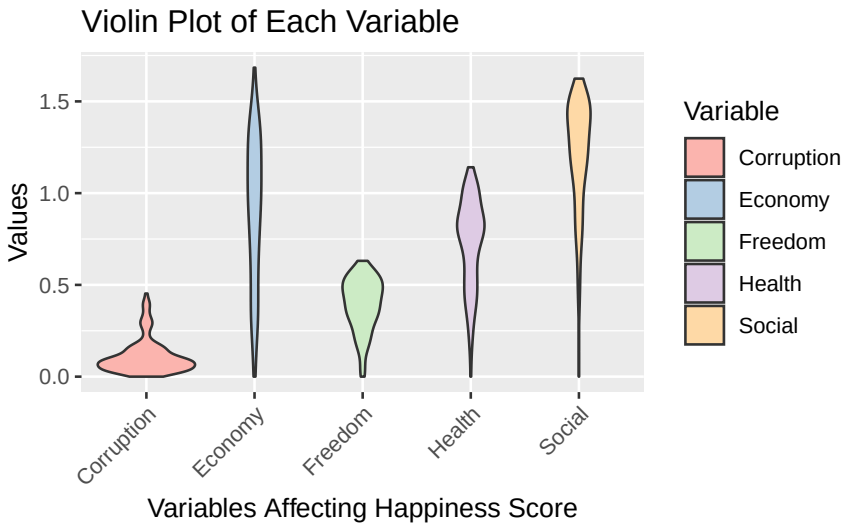
```
data_long <- data %>%
  gather(key = "Variable", value = "Score", Economy, Social, Health,
         Freedom, Corruption)

ggplot(data_long, aes(x = Variable, y = Score)) +
  geom_boxplot(width = 0.7) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Box Plot Each Variable",
       x = "Variables Affecting Happiness Score", y = "Values")
```



*Violin Plot

```
ggplot(data_long, aes(x = Variable, y = Score, fill = Variable)) +
  geom_violin(trim = TRUE) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Violin Plot of Each Variable",
       x = "Variables Affecting Happiness Score", y = "Values") +
  scale_fill_brewer(palette = "Pastel1")
```



*Summary

```
data %>%
  summarize(
    mean= mean(Economy),
    median = median(Economy),
    sd = sd(Economy),
    iqr = IQR(Economy),
    min = min(Economy),
    max = max(Economy)
 )
```

mean	median	sd	iqr	min	max
0.9051474	0.96	0.3983895	0.62975	0	1.684

```
data %>%
  summarize(
    mean= mean(Social),
    median = median(Social),
    sd = sd(Social),
    iqr = IQR(Social),
    min = min(Social),
```

```
    max = max(Social)
 )
```

mean	median	sd	iqr	min	max
1.208814	1.2715	0.2991914	0.39675	0	1.624

```
data %>%
  summarize(
    mean= mean(Health),
    median = median(Health),
    sd = sd(Health),
    iqr = IQR(Health),
    min = min(Health),
    max = max(Health)
 )
```

mean	median	sd	iqr	min	max
0.7252436	0.789	0.242124	0.334	0	1.141

```
data %>%
  summarize(
    mean= mean(Freedom),
    median = median(Freedom),
    sd = sd(Freedom),
    iqr = IQR(Freedom),
    min = min(Freedom),
    max = max(Freedom)
 )
```

mean	median	sd	iqr	min	max
0.3925705	0.417	0.1432895	0.19925	0	0.631

```
data %>%
  summarize(
    mean= mean(Corruption),
    median = median(Corruption),
    sd = sd(Corruption),
    iqr = IQR(Corruption),
    min = min(Corruption),
    max = max(Corruption)
 )
```

mean	median	sd	iqr	min	max
0.1106026	0.0855	0.0945378	0.09425	0	0.453