

Final Project (Group 2)

Group 2

2024-05-04

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

```
data <- read_excel("WHR_2015.xlsx")
```

```
colnames(data)
```

```
## [1] "Country"           "Region"
## [3] "Happiness Rank"    "Happiness Score"
## [5] "Standard Error"    "Economy (GDP per Capita)"
## [7] "Family"            "Health (Life Expectancy)"
## [9] "Freedom"           "Trust (Government Corruption)"
## [11] "Generosity"        "Dystopia Residual"
```

```
colnames(data) <- c("Country", "Region", "Happiness Score",
                    "Happiness Rank", "Economy (GDP per Capita)",
                    "Health (Life Expectancy)", "Freedom",
                    "Trust (Government Corruption)", "Family", "Generosity",
                    "Dystopia Residual", "Standard Error")
```

```
str(data)
```

```
## tibble [158 x 12] (S3: tbl_df/tbl/data.frame)
## $ Country           : chr [1:158] "Switzerland" "Iceland" "Denmark" "Norway" ..
## $ Region            : chr [1:158] "Western Europe" "Western Europe" "Western Eu
## $ Happiness Score    : num [1:158] 1 2 3 4 5 6 7 8 9 10 ...
## $ Happiness Rank     : num [1:158] 7.59 7.56 7.53 7.52 7.43 ...
## $ Economy (GDP per Capita) : num [1:158] 0.0341 0.0488 0.0333 0.0388 0.0355 ...
## $ Health (Life Expectancy) : num [1:158] 1.4 1.3 1.33 1.46 1.33 ...
## $ Freedom           : num [1:158] 1.35 1.4 1.36 1.33 1.32 ...
## $ Trust (Government Corruption): num [1:158] 0.941 0.948 0.875 0.885 0.906 ...
## $ Family            : num [1:158] 0.666 0.629 0.649 0.67 0.633 ...
```

```
## $ Generosity           : num [1:158] 0.42 0.141 0.484 0.365 0.33 ...
## $ Dystopia Residual    : num [1:158] 0.297 0.436 0.341 0.347 0.458 ...
## $ Standard Error      : num [1:158] 2.52 2.7 2.49 2.47 2.45 ...
```

```
head(data)
```

Country	Region	Happiness Score	Happiness Rank	Economy (GDP per Capita	Health (Life Ex- pectancy)	Freedom	Trust (Govern- ment Corrup- tion)	Family Generosity	Dystopia Resid- ual	Standard Er- ror
Switzerland	Western Eu- rope	1	7.587	0.03411	1.39651	1.34951	0.94143	0.665574	1978.29678	2.51738
Iceland	Western Eu- rope	2	7.561	0.04884	1.30232	1.40223	0.94784	0.628071	14146.43630	2.70201
Denmark	Western Eu- rope	3	7.527	0.03328	1.32548	1.36058	0.87464	0.649084	18350.34139	2.49204
Norway	Western Eu- rope	4	7.522	0.03880	1.45900	1.33095	0.88521	0.669033	16500.34699	2.46531
Canada	North Amer- ica	5	7.427	0.03553	1.32629	1.32261	0.90563	0.632073	2950.45811	2.45176
Finland	Western Eu- rope	6	7.406	0.03140	1.29025	1.31826	0.88911	0.641694	11370.23351	2.61955

```
tail(data)
```

Country	Region	Happiness Score	Happiness Rank	Economy (GDP per Capita	Health (Life Ex- pectancy)	Freedom	Trust (Govern- ment Corrup- tion)	Family Generosity	Dystopia Resid- ual	Standard Er- ror
Afghanistan	Southern Asia	153	3.575	0.03084	0.31982	0.30285	0.30335	0.234044	109710.36510	1.95210
Rwanda	Sub- Saharan Africa	154	3.465	0.03464	0.22208	0.77370	0.42864	0.592015	15190.22628	0.67042
Benin	Sub- Saharan Africa	155	3.340	0.03656	0.28665	0.35386	0.31910	0.484600	108010.18260	1.63328

Country	Region	Happiness Score	Happiness Rank	Economy (GDP per Capita)	Health (Life Expectancy)	Trust (Government Corruption)	Family Generosity	Dystopia Residual	Standard Error
Syria	Middle East and Northern Africa	156	3.006	0.05015	0.66320	0.474890.72193	0.1568418900	0.47179	0.32858
Burundi	Sub-Saharan Africa	157	2.905	0.08658	0.01530	0.415870.22396	0.1186010062	0.19727	1.83302
Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.139950.28443	0.3646310730	0.16681	1.56726

*Center

```
data %>%
  summarize(center = median(`Happiness Score`, na.rm = TRUE))
```

center
79.5

```
data %>%
  summarise(mean_happiness_score = mean(`Happiness Score`, na.rm = TRUE))
```

mean_happiness_score
79.49367

```
data <- read_excel("WHR_2015.xlsx")

medians <- sapply(data[c('Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)',
                        'Trust (Government Corruption)', 'Generosity')], median, na.rm = TRUE)

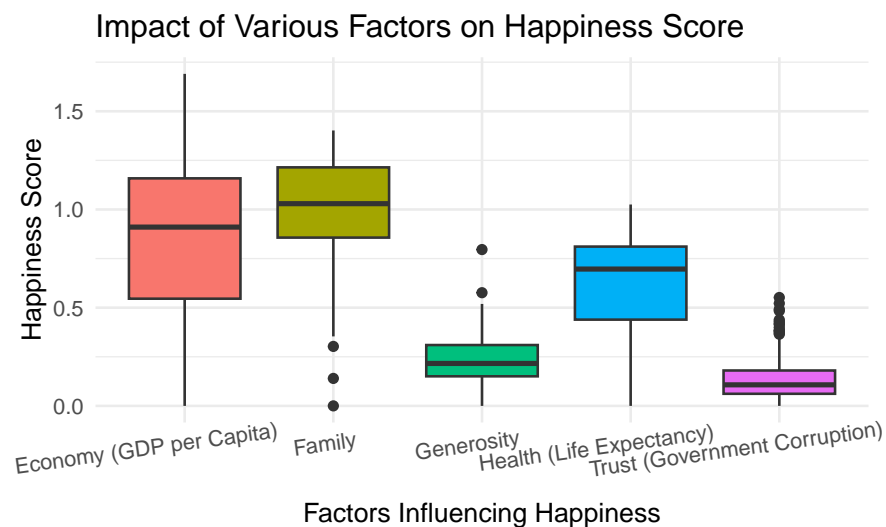
print(medians)
```

```
##      Economy (GDP per Capita)      Family
##              0.910245              1.029510
##      Health (Life Expectancy) Trust (Government Corruption)
##              0.696705              0.107220
##              Generosity
##              0.216130
```

```
library(readxl)
library(dplyr)
library(tidyr)
library(ggplot2)
happiness_data <- read_excel("WHR_2015.xlsx")
```

```
happiness_long <- happiness_data %>%
  pivot_longer(
    cols = c(`Economy (GDP per Capita)`, Family, `Health (Life Expectancy)`, `Trust (Government`
    names_to = "Variable",
    values_to = "Value"
  )
```

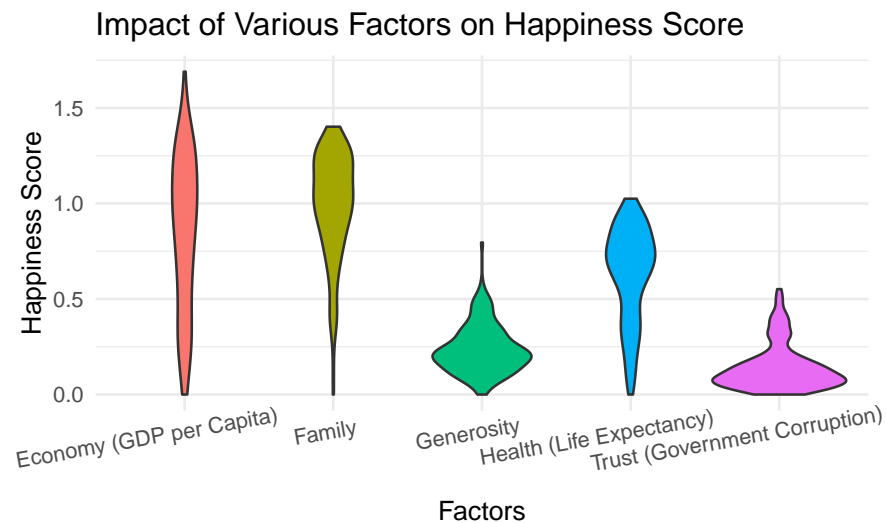
```
ggplot(happiness_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_boxplot() +
  labs(title = "Impact of Various Factors on Happiness Score",
       x = "Factors Influencing Happiness",
       y = "Happiness Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 8, hjust = 0.65),
        legend.position = "none")
```



- The boxplot for the economy shows a high median, but family has the highest.

```
ggplot(happiness_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin() +
  labs(title = "Impact of Various Factors on Happiness Score",
       x = "Factors",
       y = "Happiness Score") +
  theme_minimal() +
```

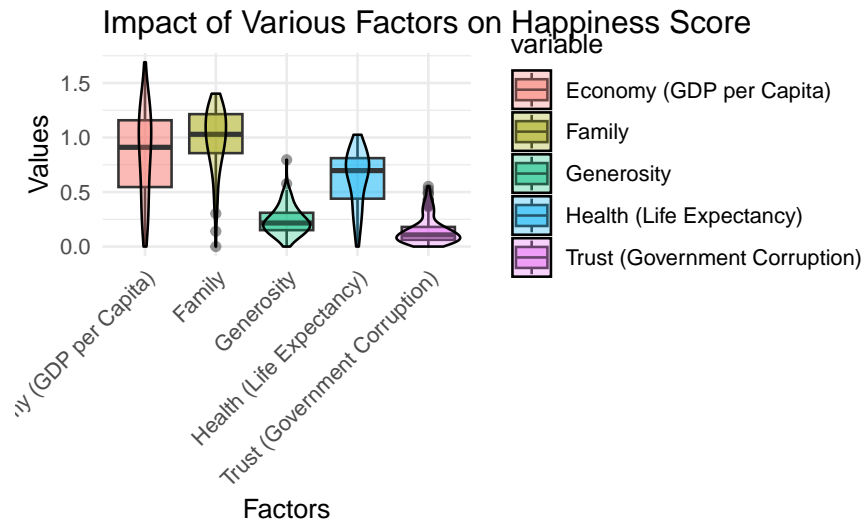
```
theme(axis.text.x = element_text(angle = 10, hjust = 0.65, vjust = 1),
      legend.position = "none")
```



```
data <- read_excel("WHR_2015.xlsx")

data_long <- pivot_longer(data, cols = c('Economy (GDP per Capita)', 'Family',
                                          'Health (Life Expectancy)', 'Trust (Government Corruption)',
                                          'Generosity'), names_to = "variable", values_to = "value")

ggplot(data_long, aes(x = variable, y = value, fill = variable)) +
  geom_boxplot(alpha = 0.5) +
  geom_violin(alpha = 0.3, color = "black", adjust = 1.5) +
  labs(title = "Impact of Various Factors on Happiness Score",
       x = "Factors",
       y = "Values") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
```



- Family has the highest median of 1.029510 and Economy had 0.910245. Both are located in the high portion compared to other factors.

```
library(reshape2)
```

```
##
```

```
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
## smiths
```

```
data_selected <- data %>%
```

```
  select(`Happiness Score`, `Economy (GDP per Capita)`, Family, `Trust (Government Corruption)`,
```

```
  cor_matrix <- cor(data_selected, use = "complete.obs")
```

```
cor_data <- melt(cor_matrix)
```

```
heatmap_plot <- ggplot(cor_data, aes(Var1, Var2, fill = value)) +
```

```
  geom_tile(color = "white") +
```

```
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0, limit = c(-1, 1)) +
```

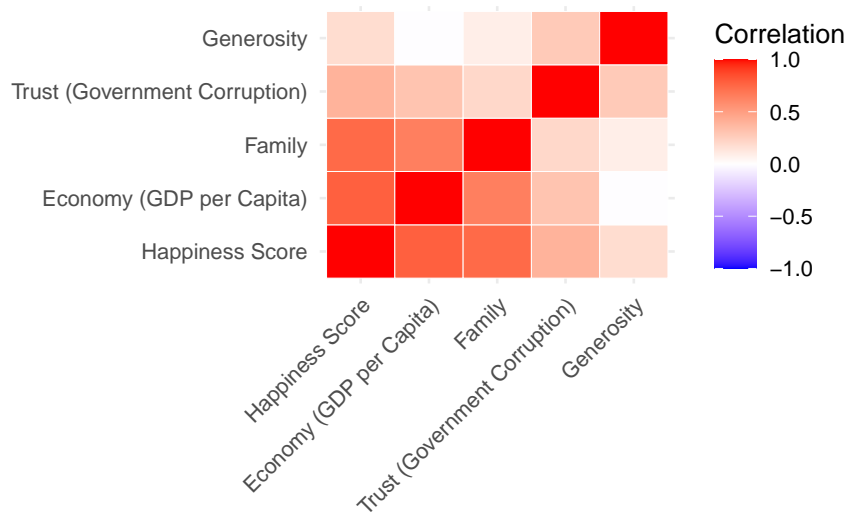
```
  theme_minimal() +
```

```
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
```

```
        axis.title = element_blank()) +
```

```
  labs(fill = "Correlation")
```

```
print(heatmap_plot)
```



- Red shows the strongest positive correlation, blue shows the strongest negative correlation, and white shows no relationship close to 0.
- Most Influential: Economy (GDP per Capita), closely followed by Family and Health.
- Moderately Influential: Trust (Government Corruption).
- Least Influential: Generosity.

```
library(corrplot)
```

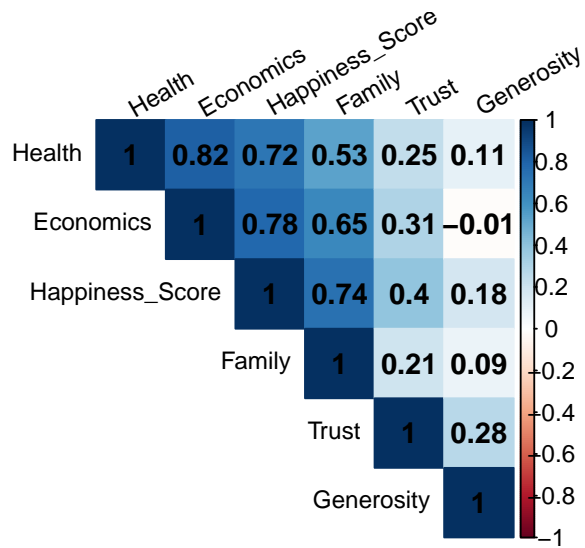
```
## corrplot 0.92 loaded
```

```
selected_data <- data %>%
  select(Happiness_Score = `Happiness Score`, Family, Health = `Health (Life Expectancy)`,
         Trust = `Trust (Government Corruption)`, Generosity, Economics = `Economy (GDP per Cap)`)

cor_matrix <- cor(selected_data, use = "complete.obs")

par(mar = c(5, 5, 5, 5))

corrplot(cor_matrix, method = "color", type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 30, addCoef.col = "black",
         tl.cex = 0.8)
```



- Happiness Score and Economics: Correlation coefficient 0.78
- Happiness Score and Health: Correlation coefficient 0.72
- Happiness Score and Family: Correlation coefficient 0.74
- Happiness Score and Trust: Correlation coefficient 0.4
- Happiness Score and Generosity: Correlation coefficient 0.18

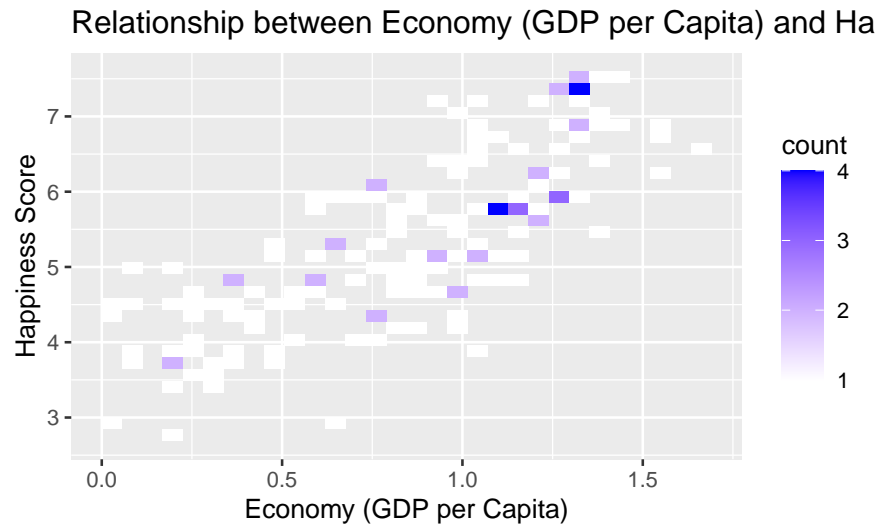
```
library(readxl)
library(ggplot2)

data <- read_excel("WHR_2015.xlsx")

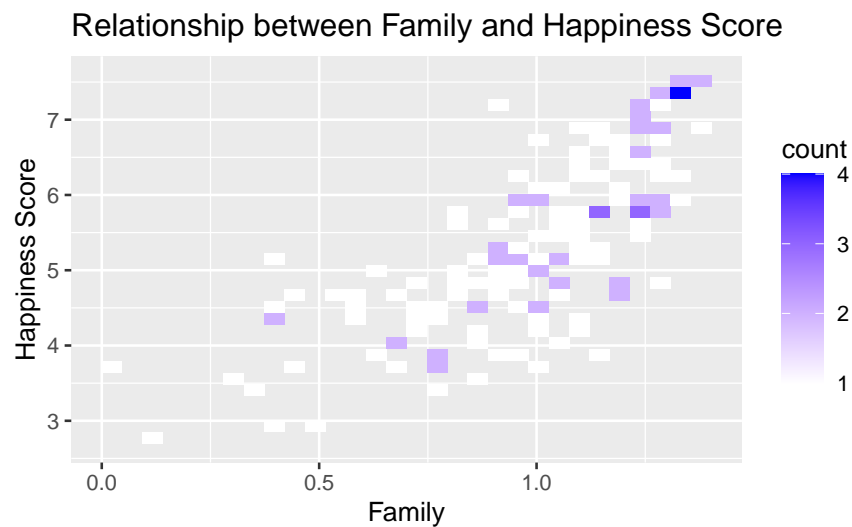
data <- data[, c("Economy (GDP per Capita)", "Family", "Health (Life Expectancy)", "Trust (Govt)", "Generosity", "Happiness Score")]

data <- na.omit(data)

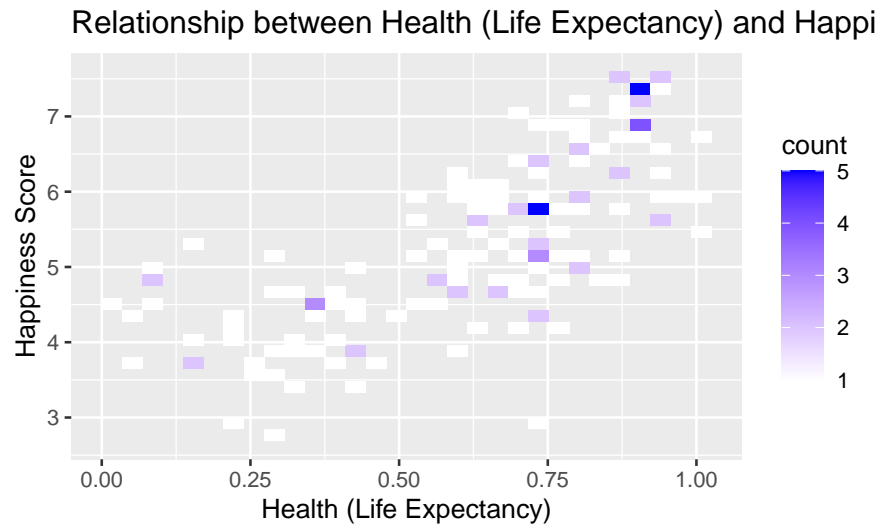
ggplot(data, aes(x = `Economy (GDP per Capita)`, y = `Happiness Score`)) +
  geom_bin2d() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Relationship between Economy (GDP per Capita) and Happiness Score")
```

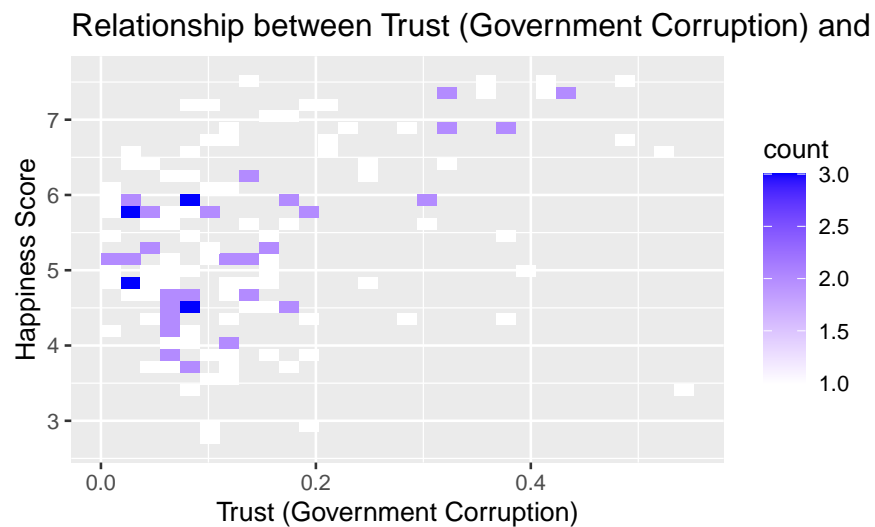
```
ggplot(data, aes(x = Family, y = `Happiness Score`)) +
  geom_bin2d() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Relationship between Family and Happiness Score")
```



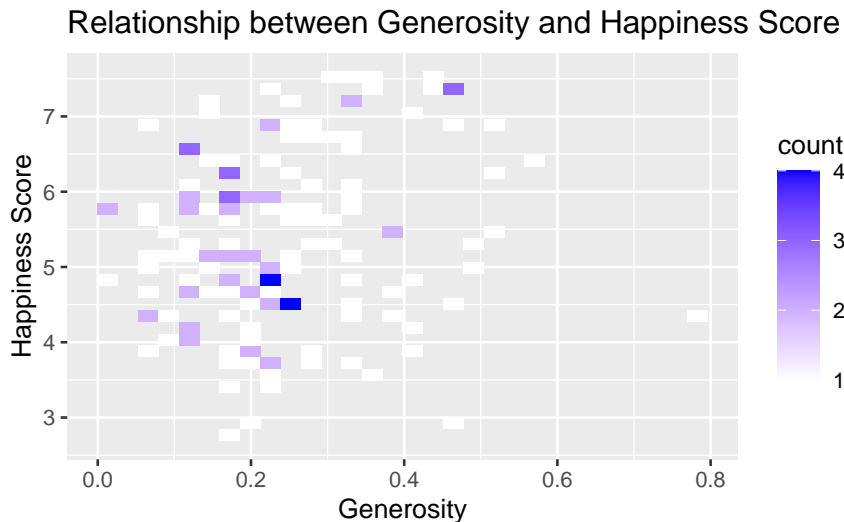
```
ggplot(data, aes(x = `Health (Life Expectancy)`, y = `Happiness Score`)) +
  geom_bin2d() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Relationship between Health (Life Expectancy) and Happiness Score")
```



```
ggplot(data, aes(x = `Trust (Government Corruption)`, y = `Happiness Score`)) +
  geom_bin2d() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Relationship between Trust (Government Corruption) and Happiness Score")
```



```
ggplot(data, aes(x = Generosity, y = `Happiness Score`)) +
  geom_bin2d() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Relationship between Generosity and Happiness Score")
```



```
data <- read_excel("WHR_2015.xlsx")

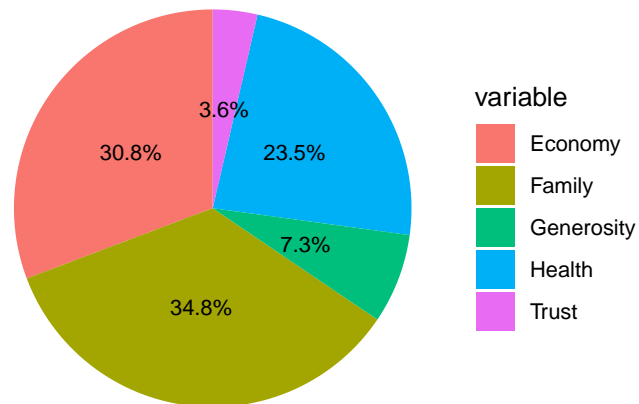
medians <- data %>%
  summarise(
    Economy = median(`Economy (GDP per Capita)`, na.rm = TRUE),
    Family = median(Family, na.rm = TRUE),
    Health = median(`Health (Life Expectancy)`, na.rm = TRUE),
    Trust = median(`Trust (Government Corruption)`, na.rm = TRUE),
    Generosity = median(Generosity, na.rm = TRUE)
  )

total <- sum(medians)
medians <- medians / total * 100

pie_data <- data.frame(
  variable = names(medians),
  value = as.numeric(medians),
  label = sprintf("%.1f%%", as.numeric(medians))
)

ggplot(pie_data, aes(x = "", y = value, fill = variable)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  theme_void() +
  labs(title = "Contribution of Various Factors to Happiness Score") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5), size = 3)
```

Contribution of Various Factors to Happiness Score



- The size of each sector is calculated based on the median value of each variable relative to the total sum of medians. Family variable occupies the largest sector, indicating it has the most significant impact on happiness scores, contributing 34.8% of the total.