Final Project (Group 2)

Group 2

2024-05-07

- 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("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"
```

```
library(readxl)

data <- read_excel("WHR_2015.xlsx")

print(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"
```

```r
data <- data %>%
  rename(
    Economy = `Economy (GDP per Capita)`,
    Family = 'Family',
    Health = `Health (Life Expectancy)`,
    Trust = `Trust (Government Corruption)`,
    Generosity = 'Generosity',
    Happiness_Score = `Happiness Score`
  )
print(colnames(data))
```

```
##  [1] "Country"           "Region"            "Happiness Rank"
##  [4] "Happiness_Score"   "Standard Error"    "Economy"
##  [7] "Family"            "Health"            "Freedom"
## [10] "Trust"             "Generosity"        "Dystopia Residual"
```

```r
  head(
    select(data, Economy, Family, Health, Trust, Generosity, Happiness_Score)
  )
```

Economy	Family	Health	Trust	Generosity	Happiness_Score
1.39651	1.34951	0.94143	0.41978	0.29678	7.587
1.30232	1.40223	0.94784	0.14145	0.43630	7.561
1.32548	1.36058	0.87464	0.48357	0.34139	7.527
1.45900	1.33095	0.88521	0.36503	0.34699	7.522
1.32629	1.32261	0.90563	0.32957	0.45811	7.427
1.29025	1.31826	0.88911	0.41372	0.23351	7.406

```r
# colnames(data) <- c("Country", "Region", "Happiness Score",
#                     "Happiness Rank","Economy",
#                     "Health", "Freedom",
#                     "Trust", "Family", "Generosity",
#                     "Dystopia Residual", "Standard Error")

#print(colnames(selected_data))
#print(head(selected_data))
```

[Module 4: Eugene Kim - Explanatory Data Analysis]

```r
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 Europe" "Weste
```

```
##  $ Happiness Rank   : num [1:158] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Happiness_Score  : num [1:158] 7.59 7.56 7.53 7.52 7.43 ...
##  $ Standard Error   : num [1:158] 0.0341 0.0488 0.0333 0.0388 0.0355 ...
##  $ Economy          : num [1:158] 1.4 1.3 1.33 1.46 1.33 ...
##  $ Family           : num [1:158] 1.35 1.4 1.36 1.33 1.32 ...
##  $ Health           : num [1:158] 0.941 0.948 0.875 0.885 0.906 ...
##  $ Freedom          : num [1:158] 0.666 0.629 0.649 0.67 0.633 ...
##  $ Trust            : num [1:158] 0.42 0.141 0.484 0.365 0.33 ...
##  $ Generosity       : num [1:158] 0.297 0.436 0.341 0.347 0.458 ...
##  $ Dystopia Residual: num [1:158] 2.52 2.7 2.49 2.47 2.45 ...
```

```
head(data)
```

Country	Region	Happiness Rank	Happiness_Score	Standard Error	Economy	Family	Health	Freedom	Trust	Generosity	Dystopia Residual
Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678	2.51738
Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877	0.14145	0.43630	2.70201
Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139	2.49204
Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.66973	0.36503	0.34699	2.46531
Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811	2.45176
Finland	Western Europe	6	7.406	0.03140	1.29025	1.31826	0.88911	0.64169	0.41372	0.23351	2.61955

```
tail(data)
```

Country	Region	Happiness Rank	Happiness_Score	Standard Error	Economy	Family	Health	Freedom	Trust	Generosity	Dystopia Residual
Afghanistan	Southern Asia	153	3.575	0.03084	0.31982	0.30285	0.30335	0.23414	0.09719	0.36510	1.95210
Rwanda	Sub-Saharan Africa	154	3.465	0.03464	0.22208	0.77370	0.42864	0.59201	0.55191	0.22628	0.67042
Benin	Sub-Saharan Africa	155	3.340	0.03656	0.28665	0.35386	0.31910	0.48450	0.08010	0.18260	1.63328
Syria	Middle East and Northern Africa	156	3.006	0.05015	0.66320	0.47489	0.72193	0.15684	0.18906	0.47179	0.32858
Burundi	Sub-Saharan Africa	157	2.905	0.08658	0.01530	0.41587	0.22396	0.11850	0.10062	0.19727	1.83302

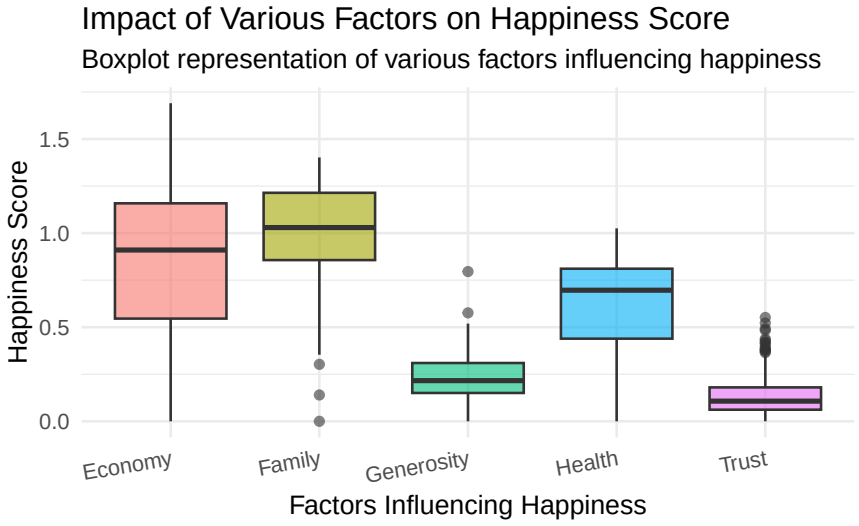
Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy	Family	Health	Freedom	Trust	Generosity	Dystopia Residual
Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.13995	0.28443	0.36453	0.10731	0.16681	1.56726

```
library(tidyr)
library(dplyr)

happiness_long <- data %>%
  pivot_longer(
    cols = c(`Economy`, `Family`, `Health`, `Trust`, `Generosity`),
    names_to = "Variable",
    values_to = "Value"
  )
```

```
library(ggplot2)

ggplot(happiness_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_boxplot(alpha = 0.6) +
  labs(title = "Impact of Various Factors on Happiness Score",
       subtitle = "Boxplot representation of various factors influencing happiness",
       x = "Factors Influencing Happiness",
       y = "Happiness Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 10, hjust = 1),
        legend.position = "none")
```



```r
library(ggplot2)

ggplot(happiness_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = TRUE, alpha = 0.6) +
  labs(title = "Impact of Various Factors on Happiness Score",
       subtitle = "Violin plots representing distribution of happiness factors",
       x = "Factors",
       y = "Happiness Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 10, hjust = 0.65, vjust = 1),
        legend.position = "none")
```