

Introduction

The World Happiness Report is a yearly survey of the state of global happiness. The report are used global from governments and organization to use happiness indicators to help with policy-making decisions.

```
# installs the necessary libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.1.2      v stringr 1.4.0
## v tidyr   1.1.3      v forcats 0.5.1
## v readr   1.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(readr)
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following object is masked from 'package:purrr':
##
##   transpose

## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

```
library(corrplot)
```

```
## corrplot 0.89 loaded
```

```
library(corrgram)
library(forcats)
```

Overview of the Data

The data that used here is the survey from 2019. They are 9 variables in this survey. 9 Variables are as follows

Overall rank: The over all rank in which country has been ranked.

Country or region: The name of the country in the survey.

Score: The score that each country has been given in the survey.

GDP per capita: Per capita gross domestic product (GDP) is the country's economic output per person.

Social Support: Social support is rated by how much a citizen of the country gets support from family and friends.

Healthy life expectancy: The rating of health and life expectancy of the country.

Freedom to make life choices: The rating of the country's citizen to make their own choices in life.

Generosity: The rating of the country's citizen of giving their time to others.

Perceptions of corruption: The country's perceived levels of public sector corruption.

```
# reads the 2019 csv file and shows the first 5 rows of the data set
data <- read.csv("2019.csv")
head(data)
```

```
## Overall.rank Country.or.region Score GDP.per.capita Social.support
## 1 1 Finland 7.769 1.340 1.587
## 2 2 Denmark 7.600 1.383 1.573
## 3 3 Norway 7.554 1.488 1.582
## 4 4 Iceland 7.494 1.380 1.624
## 5 5 Netherlands 7.488 1.396 1.522
## 6 6 Switzerland 7.480 1.452 1.526
## Healthy.life.expectancy Freedom.to.make.life.choices Generosity
## 1 0.986 0.596 0.153
## 2 0.996 0.592 0.252
## 3 1.028 0.603 0.271
## 4 1.026 0.591 0.354
## 5 0.999 0.557 0.322
## 6 1.052 0.572 0.263
## Perceptions.of.corruption
## 1 0.393
## 2 0.410
## 3 0.341
## 4 0.118
## 5 0.298
## 6 0.343
```

Structure of the Data

Below will show the structure of the data set. The output shows there are 156 columns and 9 rows.

```
# shows the structure of the data set
str(data)
```

```
## 'data.frame': 156 obs. of 9 variables:
## $ Overall.rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Country.or.region : chr "Finland" "Denmark" "Norway" "Iceland" ...
## $ Score : num 7.77 7.6 7.55 7.49 7.49 ...
## $ GDP.per.capita : num 1.34 1.38 1.49 1.38 1.4 ...
## $ Social.support : num 1.59 1.57 1.58 1.62 1.52 ...
## $ Healthy.life.expectancy : num 0.986 0.996 1.028 1.026 0.999 ...
## $ Freedom.to.make.life.choices: num 0.596 0.592 0.603 0.591 0.557 0.572 0.574 0.585 0.584 0.532 ...
## $ Generosity : num 0.153 0.252 0.271 0.354 0.322 0.263 0.267 0.33 0.285 0.244 ...
## $ Perceptions.of.corruption : num 0.393 0.41 0.341 0.118 0.298 0.343 0.373 0.38 0.308 0.226 ...
```

Summary of the Data set

Below will shows the summary of the data. For the numeric variables it show Min, 1st Quarter, Median, Mean, 3rd Quarter, and Max. For Country.or.region it shows that it is a character class.

```
# shows the summary of the data set
summary(data)
```

```
## Overall.rank Country.or.region Score GDP.per.capita
## Min. : 1.00 Length:156 Min. :2.853 Min. :0.0000
## 1st Qu.: 39.75 Class :character 1st Qu.:4.545 1st Qu.:0.6028
## Median : 78.50 Mode :character Median :5.380 Median :0.9600
## Mean : 78.50 Mean :5.407 Mean :0.9051
## 3rd Qu.:117.25 3rd Qu.:6.184 3rd Qu.:1.2325
## Max. :156.00 Max. :7.769 Max. :1.6840
## Social.support Healthy.life.expectancy Freedom.to.make.life.choices
## Min. :0.000 Min. :0.0000 Min. :0.0000
## 1st Qu.:1.056 1st Qu.:0.5477 1st Qu.:0.3080
## Median :1.272 Median :0.7890 Median :0.4170
## Mean :1.209 Mean :0.7252 Mean :0.3926
## 3rd Qu.:1.452 3rd Qu.:0.8818 3rd Qu.:0.5072
## Max. :1.624 Max. :1.1410 Max. :0.6310
## Generosity Perceptions.of.corruption
## Min. :0.0000 Min. :0.0000
## 1st Qu.:0.1087 1st Qu.:0.0470
## Median :0.1775 Median :0.0855
## Mean :0.1848 Mean :0.1106
## 3rd Qu.:0.2482 3rd Qu.:0.1412
## Max. :0.5660 Max. :0.4530
```

Renaming the columns

Here I renamed the columns for efficient use through out the report when calling them.

```
# renames the columns in the data set and shows the first 5 rows with the change of column names
data <- data %>% rename(country = Country.or.region, gdp = GDP.per.capita, support=Social.support, life_
head(data)
```

	rank	country	Score	gdp	support	life_expectancy	make_choices	Generosity
## 1	1	Finland	7.769	1.340	1.587	0.986	0.596	0.153
## 2	2	Denmark	7.600	1.383	1.573	0.996	0.592	0.252
## 3	3	Norway	7.554	1.488	1.582	1.028	0.603	0.271
## 4	4	Iceland	7.494	1.380	1.624	1.026	0.591	0.354
## 5	5	Netherlands	7.488	1.396	1.522	0.999	0.557	0.322
## 6	6	Switzerland	7.480	1.452	1.526	1.052	0.572	0.263

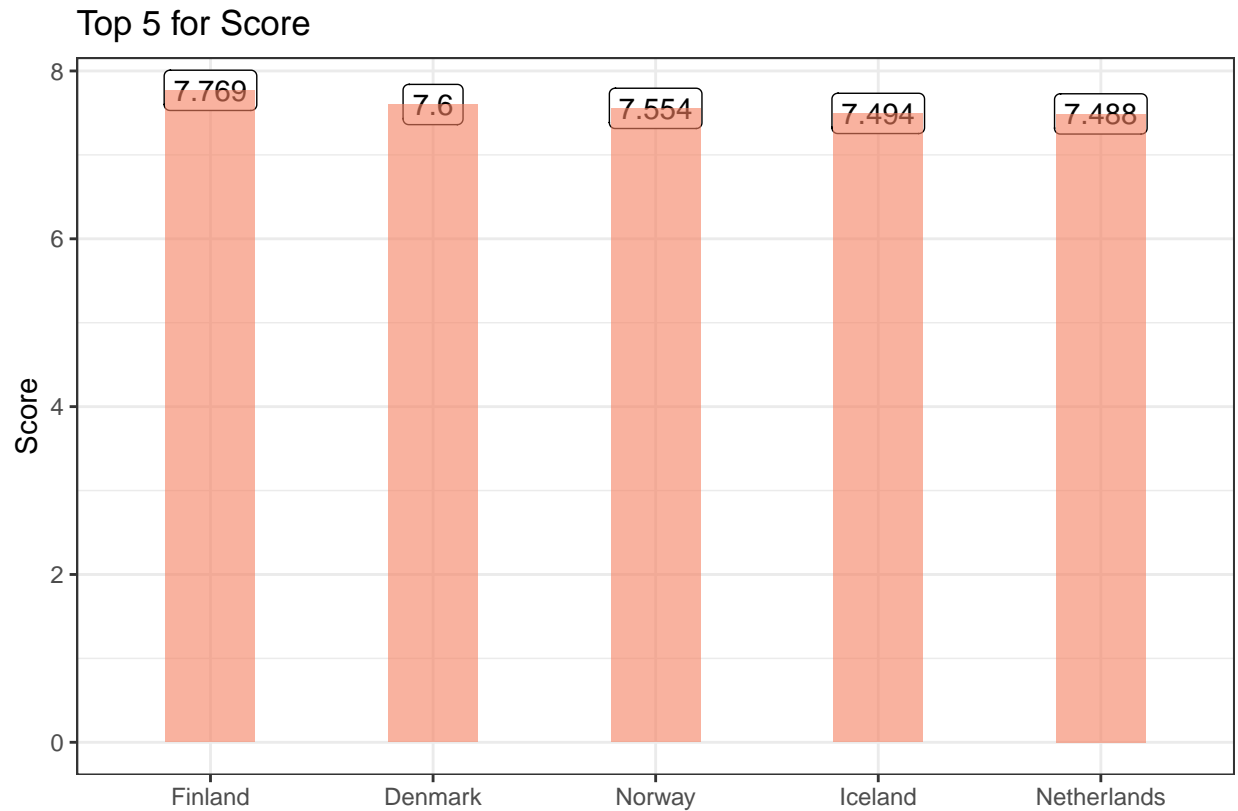
##	corruption
## 1	0.393
## 2	0.410
## 3	0.341
## 4	0.118
## 5	0.298
## 6	0.343

Graph for Top 5 countries with the highest Scores

Below will output a bar chart of the top 5 countries with the highest score

```
# outputs bar chart of the top 5 countries with the highest scores
top_n(data,n=5,Score) %>% mutate(country = fct_reorder(country, desc(Score))) %>%
  ggplot( aes(x=country, y= Score)) + geom_label(aes(label = Score))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +

  xlab("") +
  theme_bw()+labs(title="Top 5 for Score",y="Score")
```

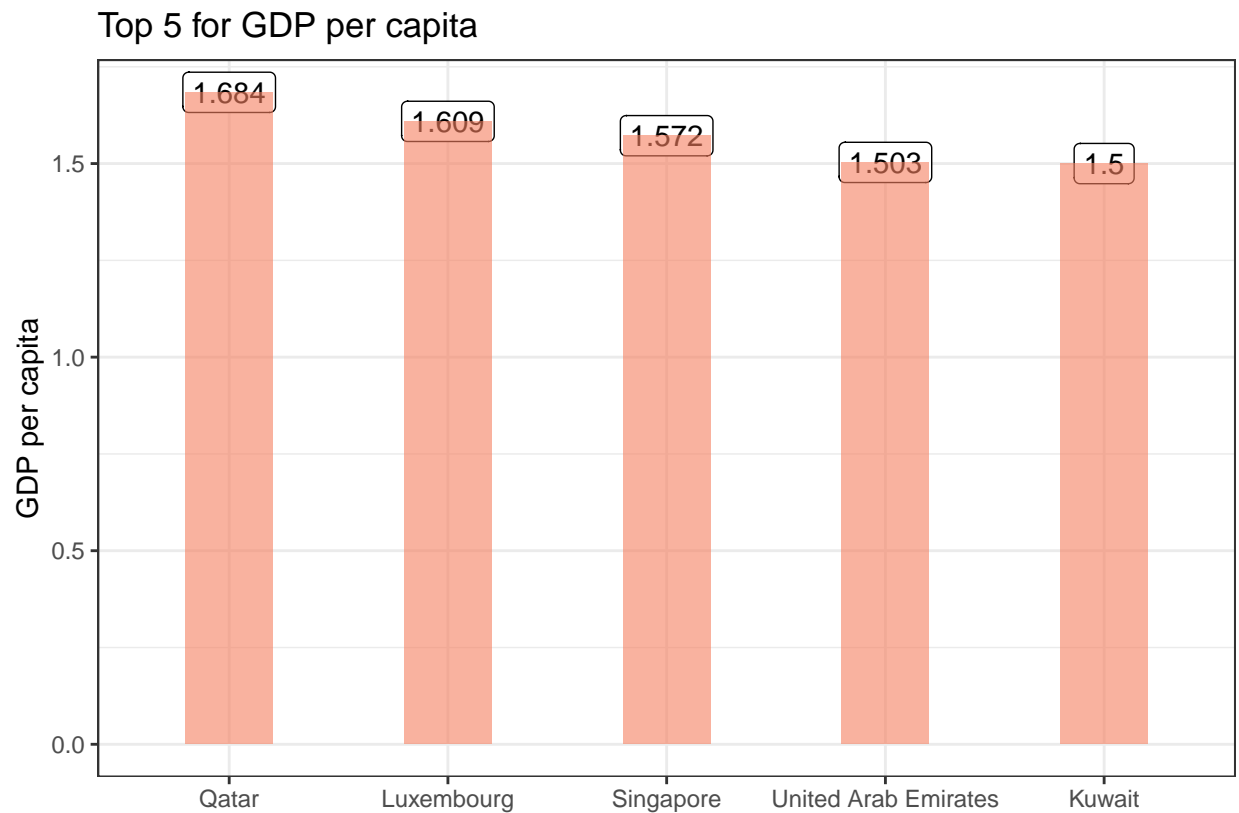


Graph for Top 5 countries with the highest GDP per Capita

Below will output a bar chart of the top 5 countries with the highest GDP per Capita

```
# outputs the top 5 countries with highest GDP per capita
top_n(data,n=5, gdp) %>%mutate(country = fct_reorder(country, desc(gdp))) %>%
  ggplot( aes(x=country, y=gdp)) + geom_label(aes(label = gdp))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +

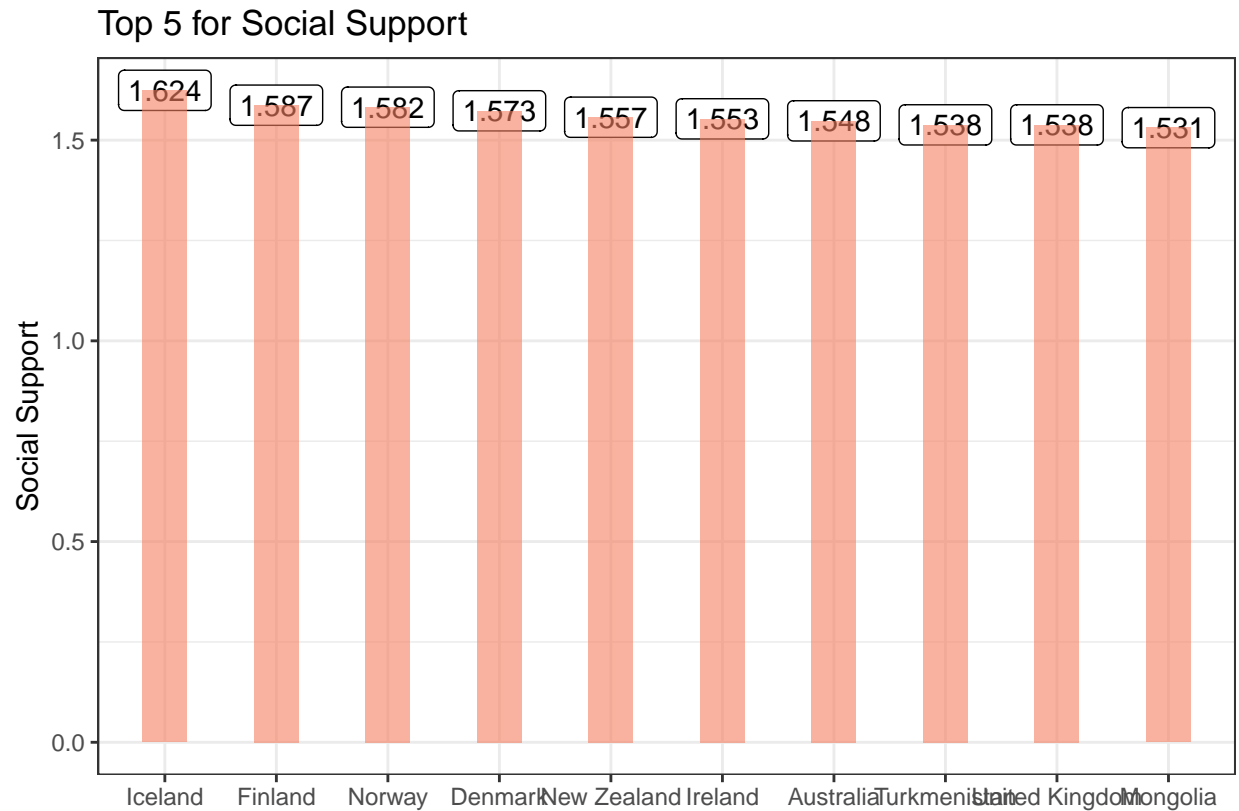
  xlab("") +
  theme_bw()+labs(title="Top 5 for GDP per capita ", y="GDP per capita")
```



Graph for Top 5 countries with the highest Social Support

Below will output a bar chart of the top 5 countries with the highest Social Support

```
# outputs the top 5 countries with the highest Social Support
top_n(data,n=10, support) %>%mutate(country = fct_reorder(country, desc(support))) %>%
  ggplot( aes(x=country, y=support)) + geom_label(aes(label = support))+ geom_bar(stat="identity", fill=
  xlab("") +
  theme_bw() + labs(title="Top 5 for Social Support",y="Social Support")
```

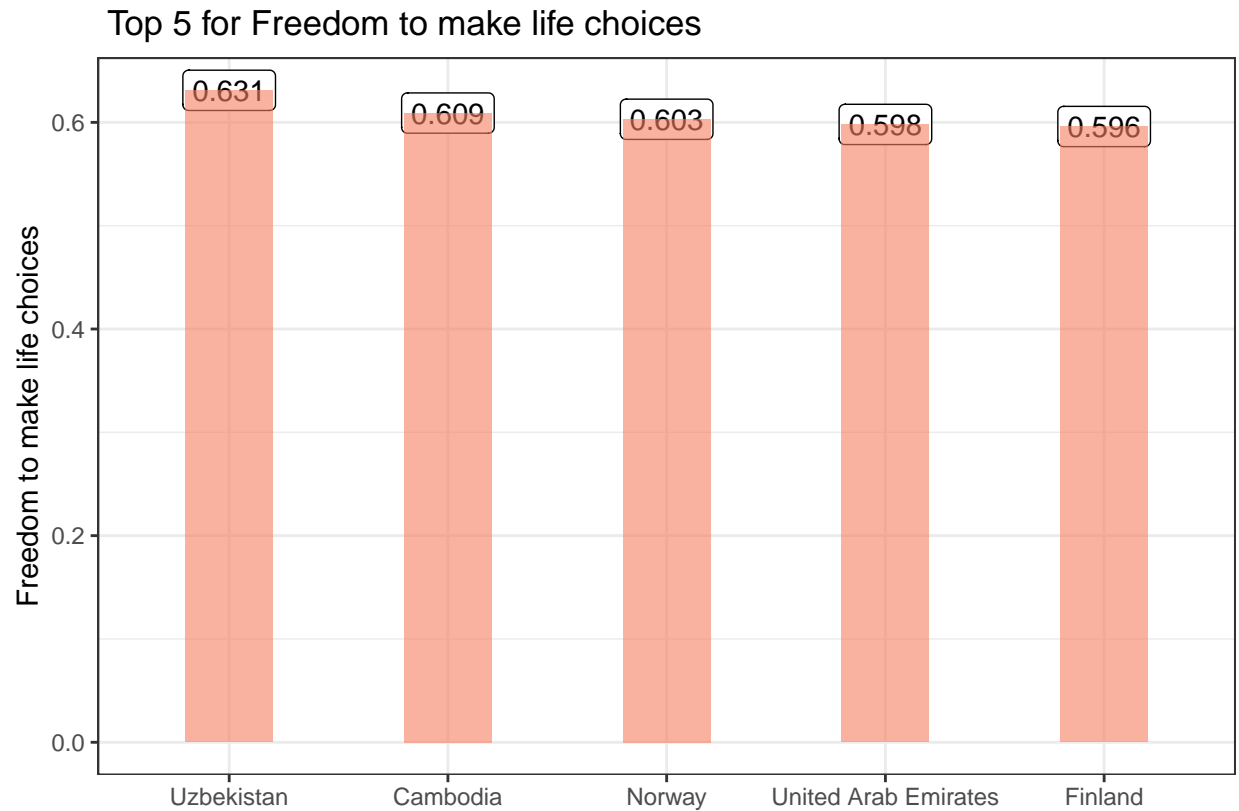


Graph for Top 5 countries with the highest Freedom to make life choices

Below will output a bar chart of the top 5 countries with the highest Freedom to make life choices

```
# outputs the top 5 countries with the highest Freedom to make life choices
top_n(data,n=5,make_choices) %>% mutate(country = fct_reorder(country, desc(make_choices))) %>%
  ggplot( aes(x=country, y = make_choices)) + geom_label(aes(label = make_choices))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +

  xlab("") +
  theme_bw() + labs(title=" Top 5 for Freedom to make life choices ",y="Freedom to make life choices")
```



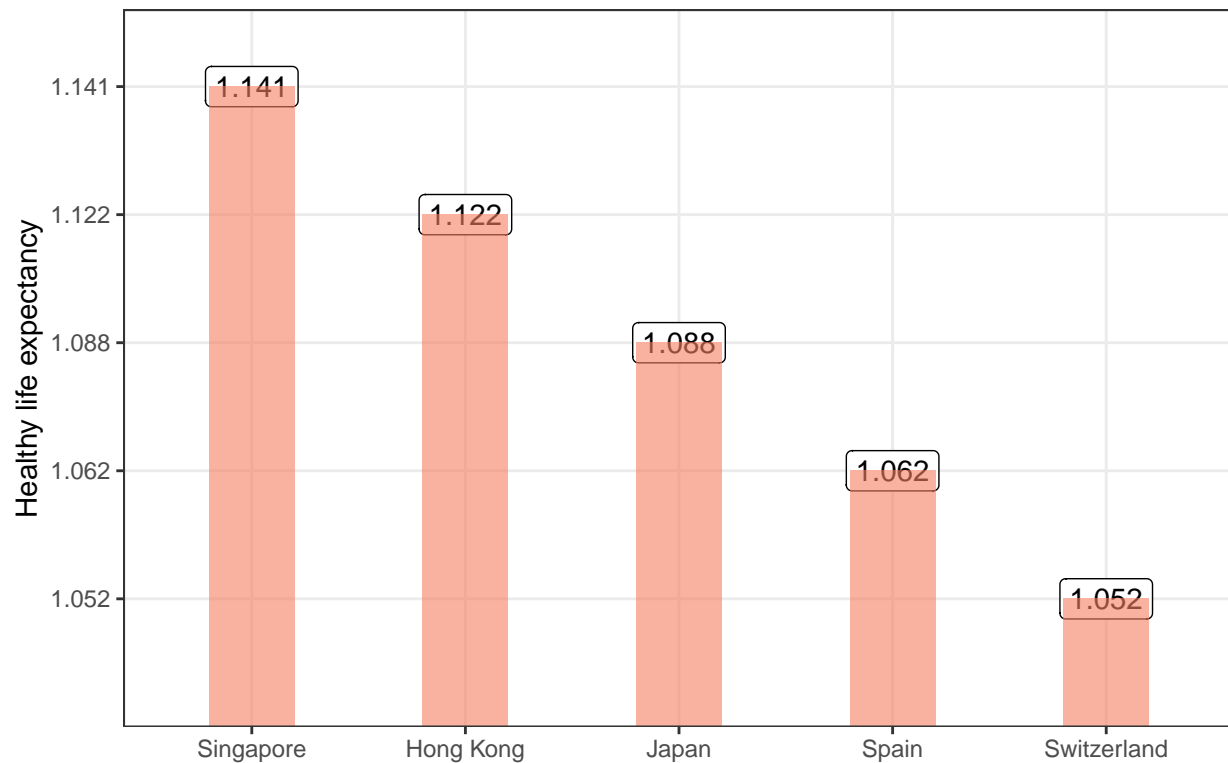
Graph for Top 5 countries with the highest Healthy life expectancy

Below will output a bar chart of the top 5 countries with the highest Healthy life expectancy

```
# outputs the top 5 countries with the highest Healthy life expectancy
top_n(data,n=5,life_expectancy) %>% mutate(country = fct_reorder(country, desc(life_expectancy))) %>%
  ggplot( aes(x= country, y= as.character(life_expectancy))) + geom_label(aes(label = life_expectancy)) +
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +

  xlab("") +
  theme_bw() + labs(title="Top 5 for Healthy life expectancy", y="Healthy life expectancy")
```


Top 5 for Healthy life expectancy

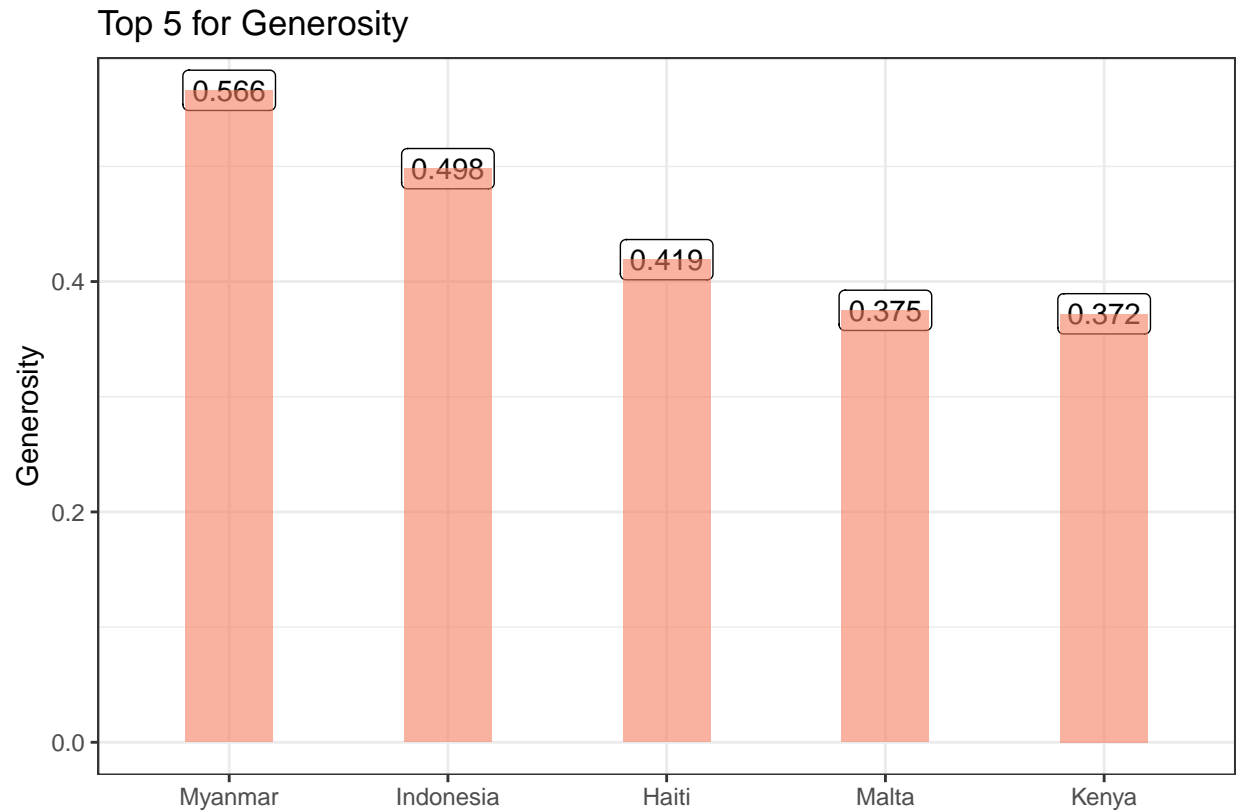


Graph for Top 5 countries with the highest Generosity

Below will output a bar chart of the top 5 countries with the highest Generosity

```
# outputs the top 5 countries with the highest Generosity
top_n(data,n=5,Generosity) %>% mutate(country = fct_reorder(country, desc(Generosity))) %>%
  ggplot( aes(x=country, y=Generosity)) + geom_label(aes(label = Generosity))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +

  xlab("") +
  theme_bw() + labs(title="Top 5 for Generosity",y="Generosity")
```



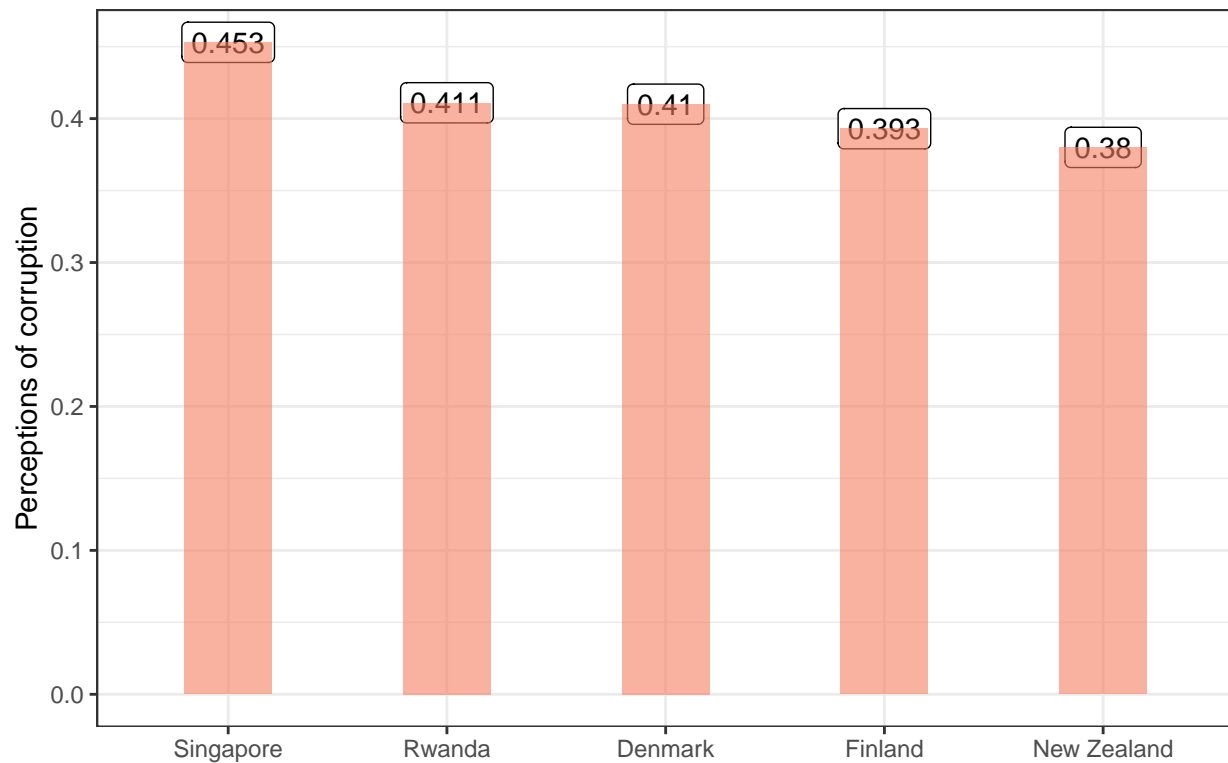
Graph for Top 5 countries with the highest Perceptions of corruption

Below will output a bar chart of the top 5 countries with the highest Perceptions of corruption

```
# outputs the top 5 countries with the Perceptions of corruption
top_n(data,n=5,corruption) %>% mutate(country = fct_reorder(country, desc(corruption))) %>%
  ggplot( aes(x=country, y=corruption)) + geom_label(aes(label = corruption))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +

  xlab("") +
  theme_bw() + labs(title="Top 5 for Perceptions of corruption", y="Perceptions of corruption")
```

Top 5 for Peceptions of corruption



Dropping Columns

Here I dropped the columns country and rank to only have columns with numeric values and put them into the variable num_data

```
# drops rows that do not contain numeric data
num_data <- data %>% subset(select=-c(country,rank))
head(num_data)
```

```
##   Score   gdp support life_expectancy make_choices Generosity corruption
## 1 7.769 1.340  1.587           0.986         0.596    0.153    0.393
## 2 7.600 1.383  1.573           0.996         0.592    0.252    0.410
## 3 7.554 1.488  1.582           1.028         0.603    0.271    0.341
## 4 7.494 1.380  1.624           1.026         0.591    0.354    0.118
## 5 7.488 1.396  1.522           0.999         0.557    0.322    0.298
## 6 7.480 1.452  1.526           1.052         0.572    0.263    0.343
```

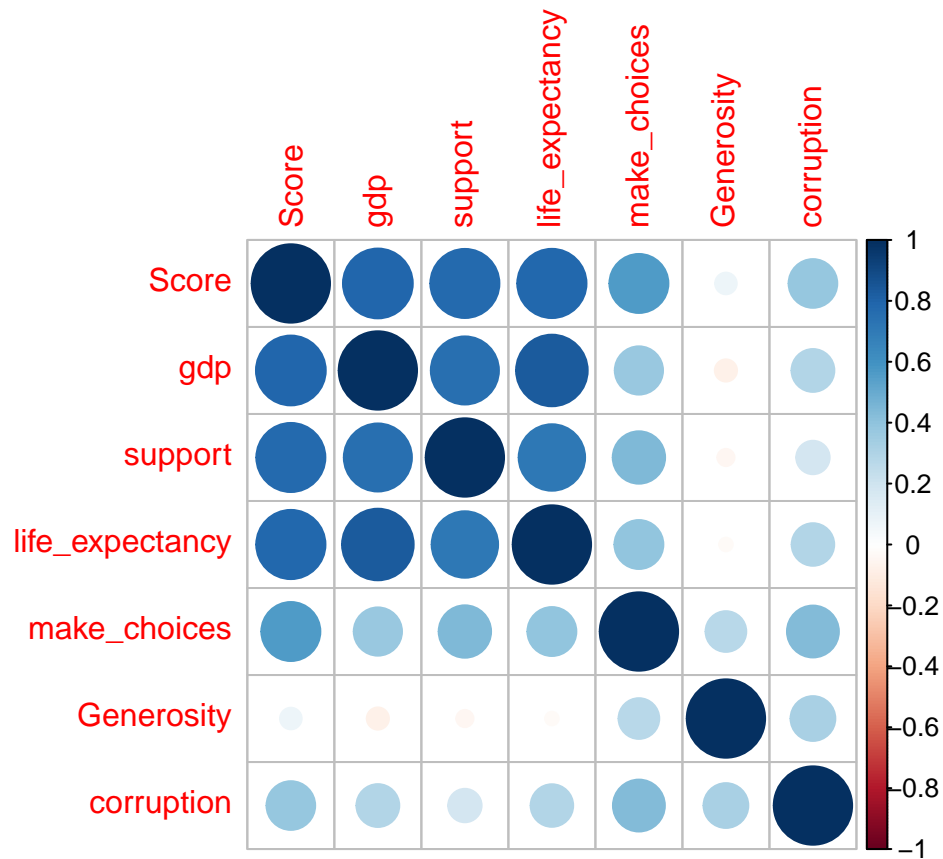
Correlation of the variables

Below will show the correlation between the variables. This will show which variables has the highest correlation to the score variable.

```
# calculates the correlation of the data
cordata = num_data[,c(1,2,3,4,5,6,7)]
corr <- round(cor(cordata), 3)
corr
```

```
##          Score    gdp support life_expectancy make_choices Generosity
## Score      1.000  0.794  0.777          0.780      0.567      0.076
## gdp        0.794  1.000  0.755          0.835      0.379     -0.080
## support    0.777  0.755  1.000          0.719      0.447     -0.048
## life_expectancy 0.780 0.835 0.719          1.000      0.390     -0.030
## make_choices 0.567 0.379 0.447          0.390      1.000      0.270
## Generosity  0.076 -0.080 -0.048         -0.030      0.270      1.000
## corruption  0.386 0.299 0.182          0.295      0.439      0.327
##
##          corruption
## Score            0.386
## gdp              0.299
## support          0.182
## life_expectancy  0.295
## make_choices     0.439
## Generosity       0.327
## corruption       1.000
```

```
corrplot(corr, method = "circle")
```



Conclusion

With the data has been collected and shown we are now able to see which variables have a high correlation to the Score of the country. From the Corrplot we see the two highest variables that affects Score is GDP per capita and Social Support. While corruption and genrosity had a negative correlation to the Score.