

Final Project: World Happiness Report

Lizeth Andrea Castellanos Beltran

12/11/2017

Introduction

The World Happiness Report is a landmark survey of the state of global happiness. The first report was published in 2012, the second in 2013, the third in 2015, and the fourth in the 2016 Update. The World Happiness 2017, which ranks 155 countries by their happiness levels, was released at the United Nations at an event celebrating International Day of Happiness on March 20th. The report continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.

Description of the Data

The happiness scores and rankings use data from the Gallup World Poll. The scores are based on answers to the main life evaluation question asked in the poll. This question, known as the Cantril ladder, asks respondents to think of a ladder with the best possible life for them being a 10 and the worst possible life being a 0 and to rate their own current lives on that scale. The scores are from nationally representative samples for the years 2013-2016 and use the Gallup weights to make the estimates representative. The columns following the happiness score estimate the extent to which each of six factors:

1. Economic production
2. Social support
3. Life expectancy
4. Freedom
5. Absence of corruption
6. Generosity

All factors contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world's lowest national averages for each of the six factors.

Download the data

I am providing the dataset for the year 2015 in the github directory <https://github.com/lacbeltran/lps/blob/master/tasks/2005.csv>

If you want to download the data from the kaggle site, you have to be logged. The URL for the CSV file is <https://www.kaggle.com/unsdsn/world-happiness/downloads/2015.csv>

Read the Data

- Read the data (If you have the file in the local directory)

```
library(readr)
df <- read_csv("2015.csv")
```

```
## Parsed with column specification:
## cols(
##   Country = col_character(),
##   Region = col_character(),
##   `Happiness Rank` = col_integer(),
##   `Happiness Score` = col_double(),
##   `Standard Error` = col_double(),
##   `Economy (GDP per Capita)` = col_double(),
##   Family = col_double(),
##   `Health (Life Expectancy)` = col_double(),
##   Freedom = col_double(),
##   `Trust (Government Corruption)` = col_double(),
##   Generosity = col_double(),
##   `Dystopia Residual` = col_double()
## )
```

The dataset contains 12 columns. The next are the descriptions of every column according to the column metadata information. The column 6 to the column 11 describe the extent to which these factors contribute in evaluating the happiness in each country. The Dystopia Residual metric actually is the Dystopia Happiness Score(1.85) + the Residual value or the unexplained value for each country as stated in the previous answer.

If you add all these factors up, you get the happiness score so it might be un-reliable to model them to predict Happiness Scores.

1. Country: Name of the country.
2. Region: Region the country belongs to.
3. Happiness Rank: Rank of the country based on the Happiness Score.
4. Happiness Score: A metric measured in 2015 by asking the sampled people the question: “How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest”.
5. Standard Error.
6. Economy (GDP per Capita): The extent to which GDP contributes to the calculation of the Happiness Score.
7. Family: The extent to which Family contributes to the calculation of the Happiness Score.
8. Health (Life Expectancy): The extent to which Life expectancy contributed to the calculation of the Happiness Score.
9. Freedom: The extent to which Freedom contributed to the calculation of the Happiness Score.
10. Trust (Government Corruption): The extent to which Perception of Corruption contributes to Happiness Score.
11. Generosity: The extent to which Generosity contributed to the calculation of the Happiness Score.
12. Dystopia Residual: The extent to which Dystopia Residual contributed to the calculation of the Happiness Score.

Goal and Question

I will try to answer this question: What countries or regions rank the highest in overall happiness and each of the six factors contributing to happiness?

Lucas said: “The countries are already ranked, as you can see in the column”Happiness Rank“. It would be indeed interesting to check whether a certain factor contribute more to the overall happiness than others. Please, be extremely rigorous in your report, justifying your choices following the LP methodology.”

Load the necessary packages:

```
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(magrittr);
library(ggplot2);
library(tidyr)

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:magrittr':
##
##   extract
```

Grafica happiness score

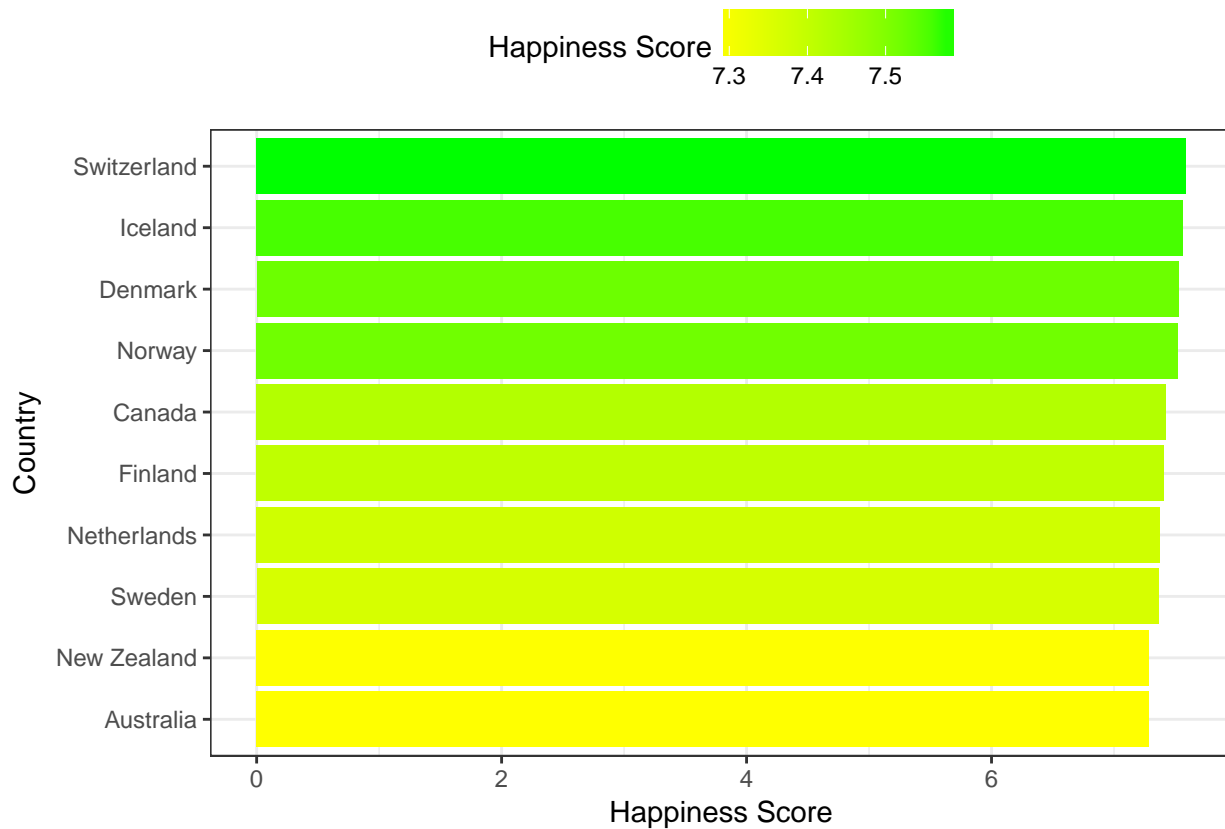
```
library(readr)
library(ggplot2)
df <- read_csv("2015.csv")

## Parsed with column specification:
## cols(
##   Country = col_character(),
##   Region = col_character(),
##   `Happiness Rank` = col_integer(),
##   `Happiness Score` = col_double(),
##   `Standard Error` = col_double(),
##   `Economy (GDP per Capita)` = col_double(),
##   Family = col_double(),
##   `Health (Life Expectancy)` = col_double(),
##   Freedom = col_double(),
##   `Trust (Government Corruption)` = col_double(),
##   Generosity = col_double(),
##   `Dystopia Residual` = col_double()
## )
df;

## # A tibble: 158 x 12
##       Country          Region `Happiness Rank`
##       <chr>          <chr>      <int>
## 1 Switzerland      Western Europe      1
## 2 Iceland           Western Europe      2
## 3 Denmark           Western Europe      3
## 4 Norway            Western Europe      4
## 5 Canada            North America      5
## 6 Finland           Western Europe      6
## 7 Netherlands       Western Europe      7
## 8 Sweden            Western Europe      8
```

```
## 9 New Zealand Australia and New Zealand          9
## 10 Australia Australia and New Zealand          10
## # ... with 148 more rows, and 9 more variables: `Happiness Score` <dbl>,
## #   `Standard Error` <dbl>, `Economy (GDP per Capita)` <dbl>,
## #   Family <dbl>, `Health (Life Expectancy)` <dbl>, Freedom <dbl>, `Trust
## #   (Government Corruption)` <dbl>, Generosity <dbl>, `Dystopia
## #   Residual` <dbl>
```

```
df %>%
  #arrange(`Happiness Rank`) %>%
  head(10) %>%
  mutate(Country = factor(Country, levels = rev(Country))) %>%
  ggplot(aes(x=Country, y=`Happiness Score`, fill = `Happiness Score`)) +
  geom_bar(stat = "identity") + #position = position_stack(reverse = TRUE)) +
  coord_flip() + theme_bw() +
  scale_fill_gradient(low = "yellow ", high = "green ") +
  theme(legend.position = "top")
```



```
library(readr)
df <- read_csv("2015.csv")

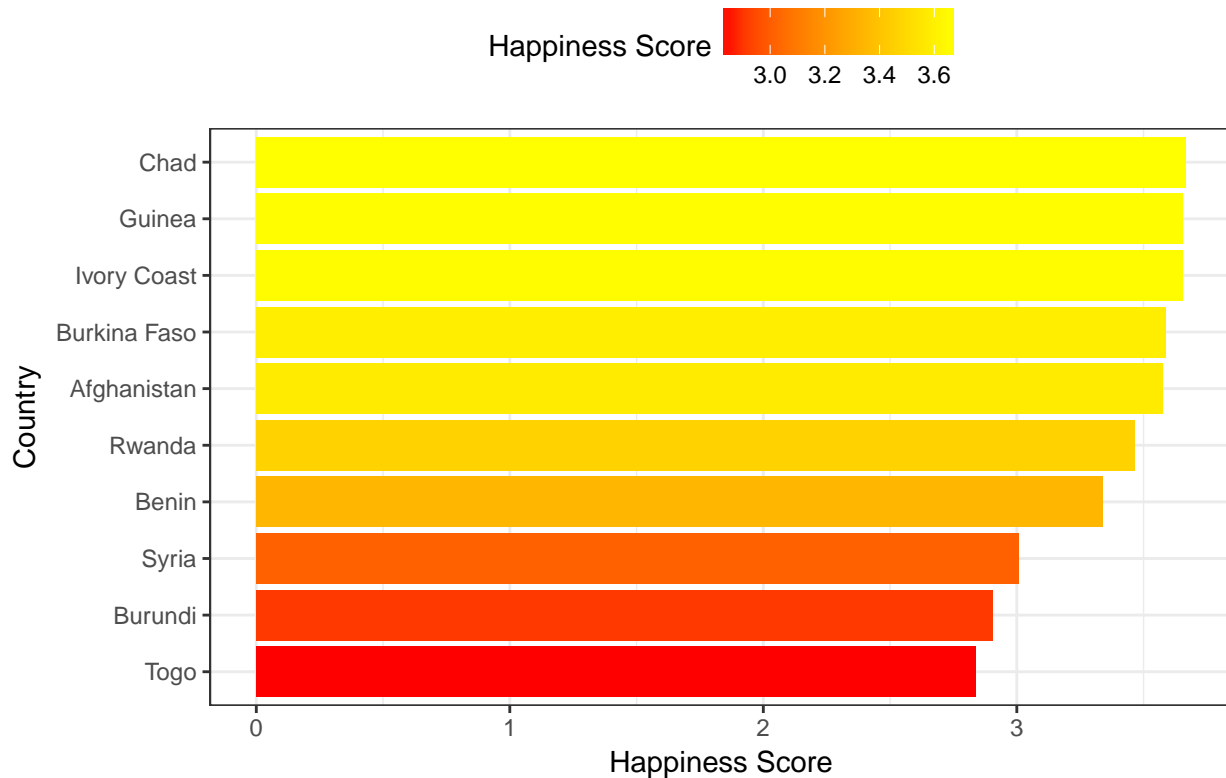
## Parsed with column specification:
## cols(
##   Country = col_character(),
##   Region = col_character(),
##   `Happiness Rank` = col_integer(),
##   `Happiness Score` = col_double(),
```

```
## `Standard Error` = col_double(),
## `Economy (GDP per Capita)` = col_double(),
## Family = col_double(),
## `Health (Life Expectancy)` = col_double(),
## Freedom = col_double(),
## `Trust (Government Corruption)` = col_double(),
## Generosity = col_double(),
## `Dystopia Residual` = col_double()
## )
df;
```

```
## # A tibble: 158 x 12
##       Country                Region `Happiness Rank`
##       <chr>                 <chr>         <int>
## 1 Switzerland             Western Europe         1
## 2 Iceland                 Western Europe         2
## 3 Denmark                 Western Europe         3
## 4 Norway                  Western Europe         4
## 5 Canada                   North America         5
## 6 Finland                 Western Europe         6
## 7 Netherlands             Western Europe         7
## 8 Sweden                  Western Europe         8
## 9 New Zealand Australia and New Zealand         9
## 10 Australia Australia and New Zealand        10
## # ... with 148 more rows, and 9 more variables: `Happiness Score` <dbl>,
## #   `Standard Error` <dbl>, `Economy (GDP per Capita)` <dbl>,
## #   Family <dbl>, `Health (Life Expectancy)` <dbl>, Freedom <dbl>, `Trust
## #   (Government Corruption)` <dbl>, Generosity <dbl>, `Dystopia
## #   Residual` <dbl>
```

```
df %>%
  #arrange(`Happiness Rank`) %>%
  tail(10) %>%
  mutate(Country = factor(Country, levels = rev(Country))) %>%
  ggplot(aes(x=Country, y=`Happiness Score`, fill = `Happiness Score`)) +
    geom_bar(stat = "identity") + #position = position_stack(reverse = TRUE)) +
    coord_flip() + theme_bw() +
    ggtitle("The 10 least happy countries of 2015") +
    scale_fill_gradient(low = "red ", high = "yellow")+
    theme(legend.position = "top")
```

The 10 least happy countries of 2015



```
df %>% head(n=2);
```

```
## # A tibble: 2 x 12
##   Country      Region `Happiness Rank` `Happiness Score`
##   <chr>        <chr>      <int>         <dbl>
## 1 Switzerland Western Europe      1         7.587
## 2 Iceland      Western Europe      2         7.561
## # ... with 8 more variables: `Standard Error` <dbl>, `Economy (GDP per
## #   Capita)` <dbl>, `Family` <dbl>, `Health (Life Expectancy)` <dbl>,
## #   Freedom <dbl>, `Trust (Government Corruption)` <dbl>,
## #   Generosity <dbl>, `Dystopia Residual` <dbl>
```

```
df %>% filter(Region == "Western Europe");
```

```
## # A tibble: 21 x 12
##   Country      Region `Happiness Rank` `Happiness Score`
##   <chr>        <chr>      <int>         <dbl>
## 1 Switzerland Western Europe      1         7.587
## 2 Iceland      Western Europe      2         7.561
## 3 Denmark      Western Europe      3         7.527
## 4 Norway        Western Europe      4         7.522
## 5 Finland       Western Europe      6         7.406
## 6 Netherlands  Western Europe      7         7.378
## 7 Sweden        Western Europe      8         7.364
## 8 Austria       Western Europe     13         7.200
## 9 Luxembourg    Western Europe     17         6.946
## 10 Ireland      Western Europe     18         6.940
```

```
## # ... with 11 more rows, and 8 more variables: `Standard Error` <dbl>,
## #   `Economy (GDP per Capita)` <dbl>, Family <dbl>, `Health (Life
## #   Expectancy)` <dbl>, Freedom <dbl>, `Trust (Government
## #   Corruption)` <dbl>, Generosity <dbl>, `Dystopia Residual` <dbl>
```

```
df %>% group_by(Region) %>% summarize(occurrence=n());
```

```
## # A tibble: 10 x 2
```

```
##           Region occurrence
##           <chr>         <int>
## 1 Australia and New Zealand      2
## 2 Central and Eastern Europe    29
## 3 Eastern Asia                   6
## 4 Latin America and Caribbean   22
## 5 Middle East and Northern Africa 20
## 6 North America                  2
## 7 Southeastern Asia              9
## 8 Southern Asia                  7
## 9 Sub-Saharan Africa             40
## 10 Western Europe                21
```

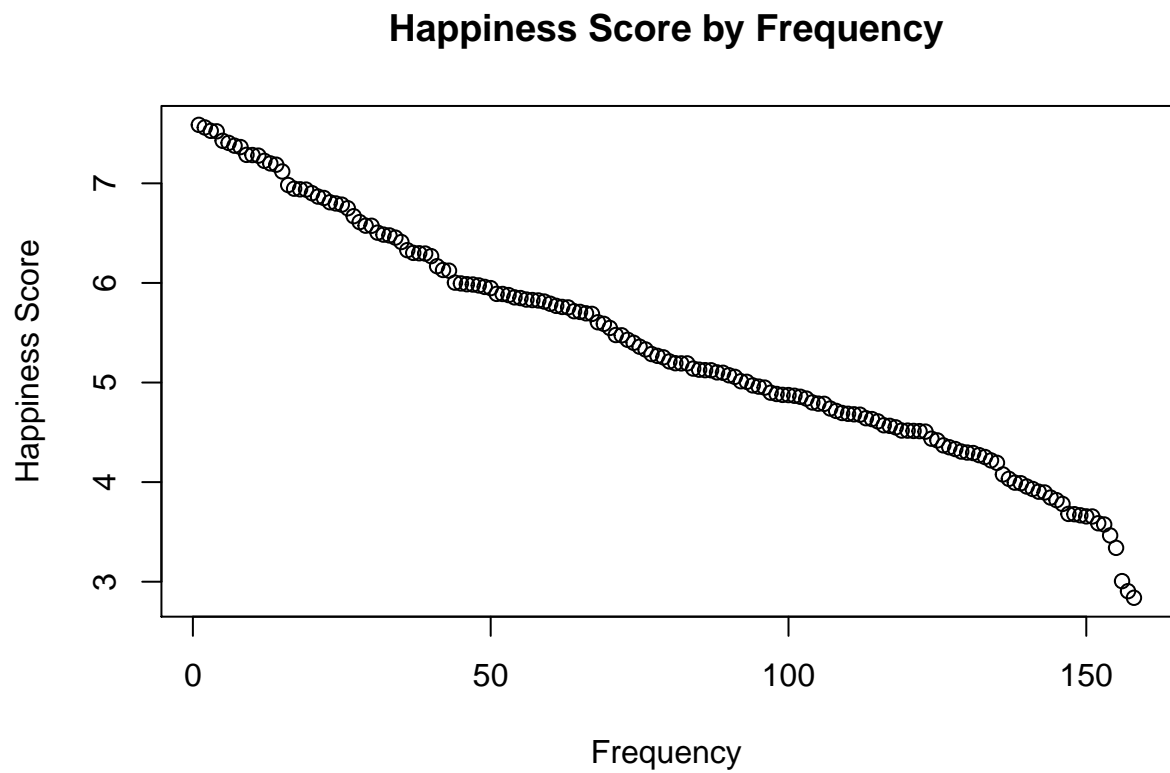
```
summary(df)
```

```
## Country           Region           Happiness Rank   Happiness Score
## Length:158        Length:158        Min.   : 1.00    Min.   :2.839
## Class :character   Class :character   1st Qu.: 40.25   1st Qu.:4.526
## Mode  :character   Mode  :character   Median : 79.50   Median :5.232
##                                     Mean  : 79.49   Mean  :5.376
##                                     3rd Qu.:118.75  3rd Qu.:6.244
##                                     Max.   :158.00  Max.   :7.587
## Standard Error     Economy (GDP per Capita)   Family
## Min.   :0.01848      Min.   :0.0000           Min.   :0.0000
## 1st Qu.:0.03727      1st Qu.:0.5458           1st Qu.:0.8568
## Median :0.04394      Median :0.9102           Median :1.0295
## Mean   :0.04788      Mean   :0.8461           Mean   :0.9910
## 3rd Qu.:0.05230      3rd Qu.:1.1584           3rd Qu.:1.2144
## Max.   :0.13693      Max.   :1.6904           Max.   :1.4022
## Health (Life Expectancy)   Freedom   Trust (Government Corruption)
## Min.   :0.0000           Min.   :0.0000   Min.   :0.00000
## 1st Qu.:0.4392           1st Qu.:0.3283   1st Qu.:0.06168
## Median :0.6967           Median :0.4355   Median :0.10722
## Mean   :0.6303           Mean   :0.4286   Mean   :0.14342
## 3rd Qu.:0.8110           3rd Qu.:0.5491   3rd Qu.:0.18025
## Max.   :1.0252           Max.   :0.6697   Max.   :0.55191
## Generosity             Dystopia Residual
## Min.   :0.0000      Min.   :0.3286
## 1st Qu.:0.1506      1st Qu.:1.7594
## Median :0.2161      Median :2.0954
## Mean   :0.2373      Mean   :2.0990
## 3rd Qu.:0.3099      3rd Qu.:2.4624
## Max.   :0.7959      Max.   :3.6021
```

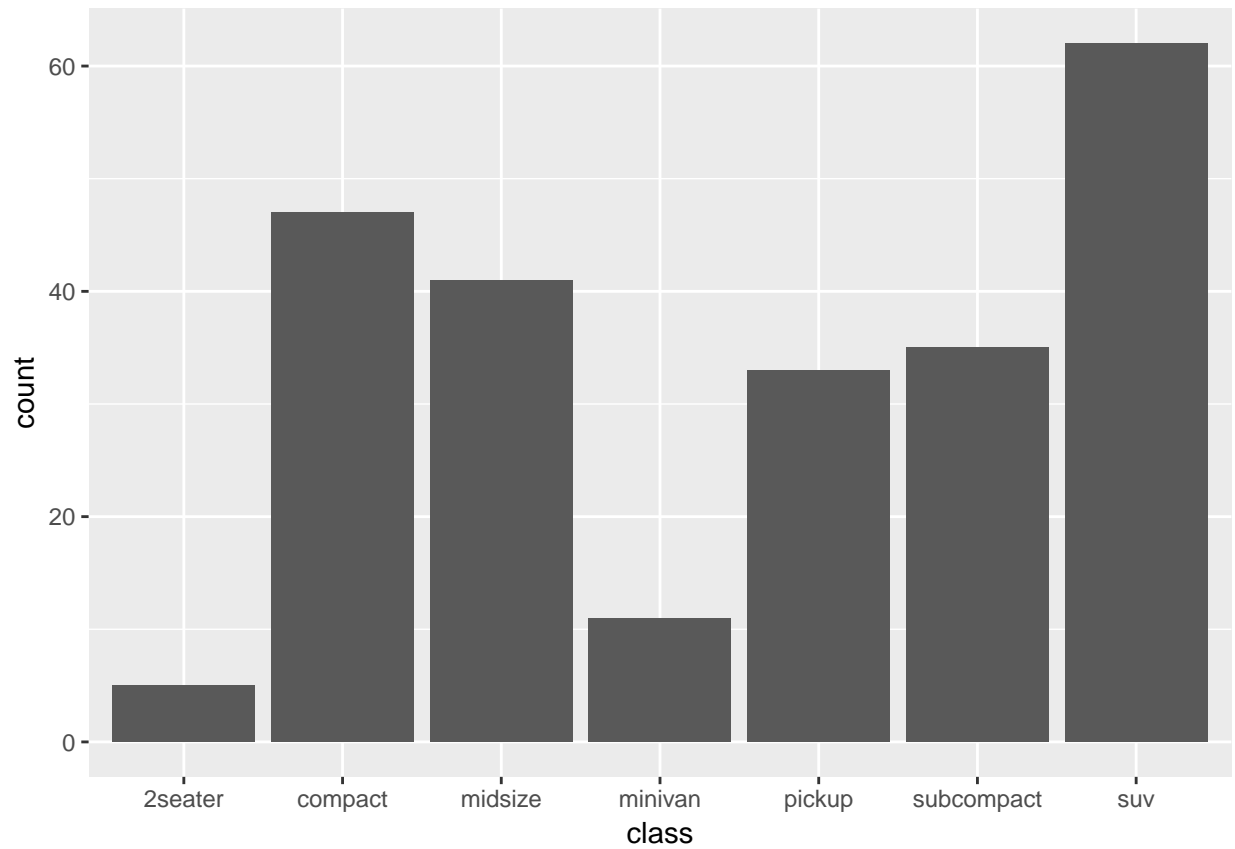
```
summary(df$`Happiness Score`)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 2.839   4.526   5.232   5.376   6.244   7.587
```

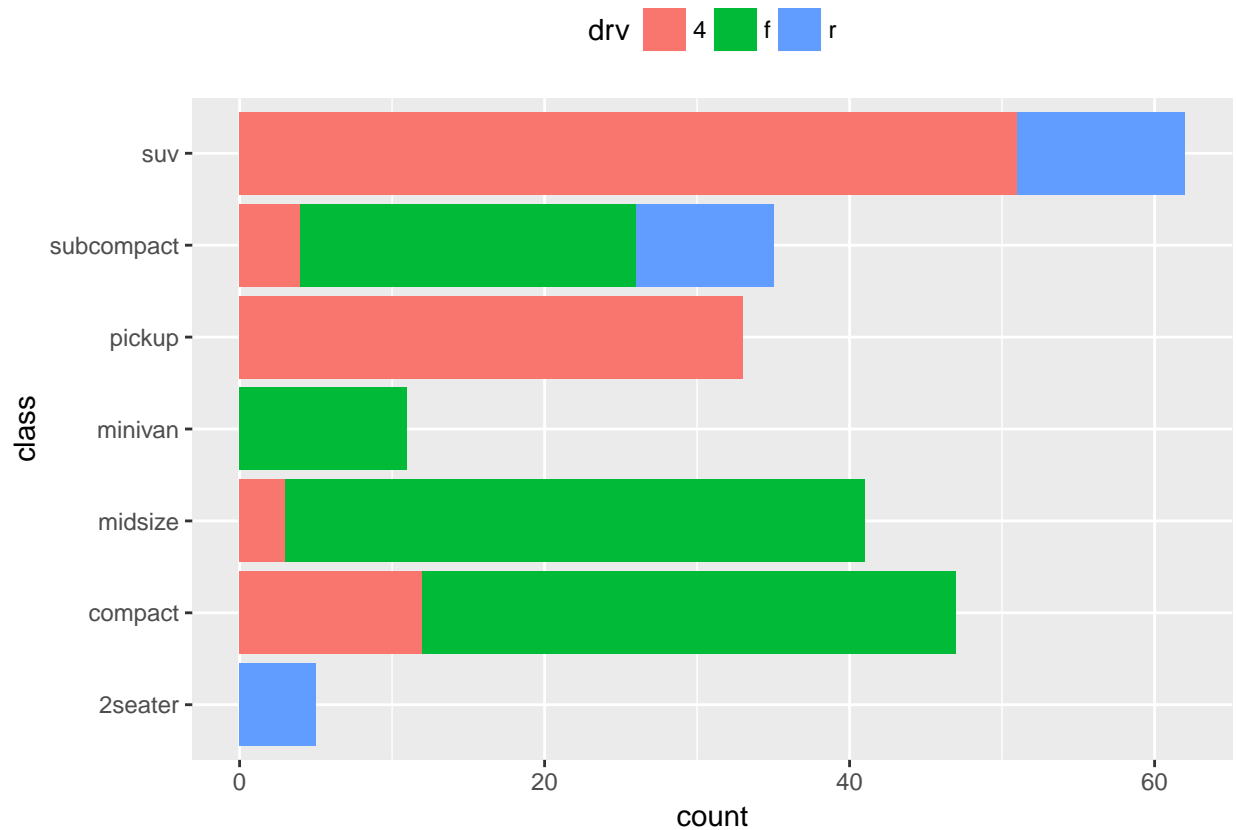
```
plot(df$`Happiness Score`, main = "Happiness Score by Frequency", ylab="Happiness Score", xlab="Frequency")
```



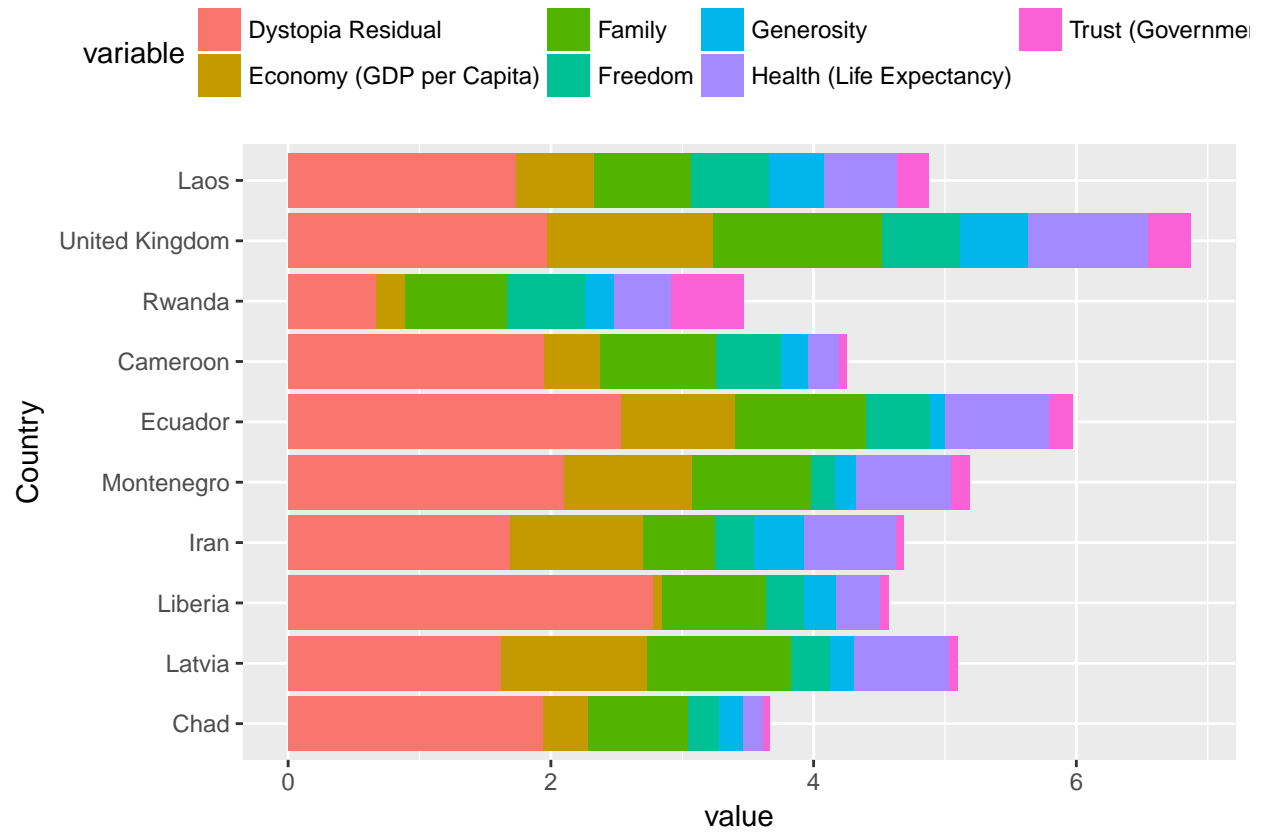
```
g <- ggplot(mpg, aes(class))  
g + geom_bar()
```

```
g +  
  geom_bar(aes(fill = drv), position = position_stack(reverse = TRUE)) +  
  coord_flip() +  
  theme(legend.position = "top")
```



```
library(tidyr)
library(magrittr)
library(dplyr)
library(ggplot2)
df %>%
  #arrange(Happiness Rank) %>%
  sample_n(10) %>%
  head(10) %>%
  mutate(Country = factor(Country, levels = rev(Country))) %>%
  select(1,6:12)%>%
  gather(variable, value, -Country) %>%
  ggplot(aes(Country))+
  geom_bar(aes(y=value, fill = variable), position = position_stack(reverse = TRUE), stat="identity") +
  coord_flip() +
  theme(legend.position = "top")
```



Answer of the Question