World Happiness Exploratory Data Analysis

Wahyu Widiawati

2023-03-19

1. Introduction

1.1 Metadata

Data Source : https://worldhappiness.report/
Data Range : 2018 - 2022

1.2 Definition

World happiness is the collective state of well-being and contentment of individuals and societies globally. This happiness level is evaluated and presented in a World Happiness Report, which includes country rankings and overall scores. A country's happiness rank is determined by its happiness score, which is measured by several parameters, including but not limited to:

- Gross Domestic Product (GDP) per capita: the monetary value of all finished goods and services produced within a country's borders in a specific period of time, divided by the number of people living in that country. It reflects how well the economy of the country is performing.
- Social support: the support gained from someone when they encounter problems.
- Healthy life expectancy: the expectancy regarding physical and mental health.
- Freedom to make life choices: the satisfaction of the freedom to choose what people do with their lives.
- Generosity: the act of giving to others without expecting anything in return. It involves showing kindness, compassion, and willingness to help others in need.
- Perceptions of corruption: refers to the question like, "do people trust their governments and have trust in the benevolence of others?"

1.3 The Analysis Questions

In this project, we will explore the world happiness datasets from 2018-2022 and examine the following questions:
1. How have the rankings changed over time?
2. What factors are most strongly correlated with happiness scores?

2. Preparation

2.1 Load Libraries

We need to load some libraries here, such as dplyr (for data manipulation), ggplot2 (to create a plot), and corrplot (to reflect a correlations between the parameters).

```
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(ggplot2)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

2.2 Importing Datasets

```
getwd()
```

```
## [1] "C:/Users/organizer/OneDrive/Documents/Data Analytics/Portofolio/EDA"
```

```
df_2018 <- data.frame(read.csv(file = "dataset_2018.csv", header = TRUE, sep = ","))
df_2019 <- data.frame(read.csv(file = "dataset_2019.csv", header = TRUE, sep = ","))
df_2020 <- data.frame(read.csv(file = "dataset_2020.csv", header = TRUE, sep = ","))
df_2021 <- data.frame(read.csv(file = "dataset_2021.csv", header = TRUE, sep = ","))
df_2022 <- data.frame(read.csv(file = "dataset_2022.csv", header = TRUE, sep = ","))
```

3. Data Cleaning and Formatting

3.1 List of Fields

```
names(df_2018)
```

```
## [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"
```

```
names(df_2019)
```

```
## [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"
```

```
names(df_2020)
```

```
##  [1] "Overall.Rank"
##  [2] "Country.name"
##  [3] "Regional.indicator"
##  [4] "Ladder.score"
##  [5] "Standard.error.of.ladder.score"
##  [6] "upperwhisker"
##  [7] "lowerwhisker"
##  [8] "Logged.GDP.per.capita"
##  [9] "Social.support"
## [10] "Healthy.life.expectancy"
## [11] "Freedom.to.make.life.choices"
## [12] "Generosity"
## [13] "Perceptions.of.corruption"
## [14] "Ladder.score.in.Dystopia"
## [15] "Explained.by..Log.GDP.per.capita"
## [16] "Explained.by..Social.support"
## [17] "Explained.by..Healthy.life.expectancy"
## [18] "Explained.by..Freedom.to.make.life.choices"
## [19] "Explained.by..Generosity"
## [20] "Explained.by..Perceptions.of.corruption"
## [21] "Dystopia...residual"
```

```
names(df_2021)
```

```
##  [1] "Overall.Rank"
##  [2] "Country.name"
##  [3] "Regional.indicator"
##  [4] "Ladder.score"
##  [5] "Standard.error.of.ladder.score"
##  [6] "upperwhisker"
##  [7] "lowerwhisker"
##  [8] "Logged.GDP.per.capita"
##  [9] "Social.support"
## [10] "Healthy.life.expectancy"
## [11] "Freedom.to.make.life.choices"
## [12] "Generosity"
## [13] "Perceptions.of.corruption"
## [14] "Ladder.score.in.Dystopia"
## [15] "Explained.by..Log.GDP.per.capita"
## [16] "Explained.by..Social.support"
## [17] "Explained.by..Healthy.life.expectancy"
## [18] "Explained.by..Freedom.to.make.life.choices"
```

```
## [19] "Explained.by..Generosity"
## [20] "Explained.by..Perceptions.of.corruption"
## [21] "Dystopia...residual"
```

```
names(df_2022)
```

```
##  [1] "Overall.Rank"
##  [2] "Country"
##  [3] "Happiness.score"
##  [4] "Whisker.high"
##  [5] "Whisker.low"
##  [6] "Dystopia..1.83....residual"
##  [7] "Explained.by..GDP.per.capita"
##  [8] "Explained.by..Social.support"
##  [9] "Explained.by..Healthy.life.expectancy"
## [10] "Explained.by..Freedom.to.make.life.choices"
## [11] "Explained.by..Generosity"
## [12] "Explained.by..Perceptions.of.corruption"
```

3.2 Removing Unnecessary Fields

In this EDA, we have different data structures for the 2020-2022 datasets, which have more fields than 2018 and 2019 datasets. Therefore, we need to adjust them by removing some fields. It's important to consider whether the field is necessary or not. The fields like regional indicator, upperwhisker, lowerwhisker, etc. were ensured to be unnecessary for this EDA. Hence, we can remove these fields.

```
df_2020 <- df_2020 [ ,-c(3,5:14,21)]
names(df_2020)
```

```
## [1] "Overall.Rank"
## [2] "Country.name"
## [3] "Ladder.score"
## [4] "Explained.by..Log.GDP.per.capita"
## [5] "Explained.by..Social.support"
## [6] "Explained.by..Healthy.life.expectancy"
## [7] "Explained.by..Freedom.to.make.life.choices"
## [8] "Explained.by..Generosity"
## [9] "Explained.by..Perceptions.of.corruption"
```

```
df_2021 <- df_2021 [ ,-c(3,5:14,21)]
names(df_2021)
```

```
## [1] "Overall.Rank"
## [2] "Country.name"
## [3] "Ladder.score"
## [4] "Explained.by..Log.GDP.per.capita"
## [5] "Explained.by..Social.support"
## [6] "Explained.by..Healthy.life.expectancy"
## [7] "Explained.by..Freedom.to.make.life.choices"
## [8] "Explained.by..Generosity"
## [9] "Explained.by..Perceptions.of.corruption"
```

```
df_2022 <- df_2022 [ ,-c(4:6)]
names(df_2022)
```

```
## [1] "Overall.Rank"
## [2] "Country"
## [3] "Happiness.score"
## [4] "Explained.by..GDP.per.capita"
## [5] "Explained.by..Social.support"
## [6] "Explained.by..Healthy.life.expectancy"
## [7] "Explained.by..Freedom.to.make.life.choices"
## [8] "Explained.by..Generosity"
## [9] "Explained.by..Perceptions.of.corruption"
```

3.3 Changing Field Name

Once we have the same data structures for all the datasets, we need to make the column names consistent. This will help us to bind them together, as the next step will requires the data to have same column names.

```
colnamesconv <- names(df_2018)
colnames(df_2020) <- colnamesconv
colnames(df_2021) <- colnamesconv
colnames(df_2022) <- colnamesconv
names(df_2018)
```

```
## [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"
```

```
names(df_2019)
```

```
## [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"
```

```
names(df_2020)
```

```
## [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"
```

```
names(df_2021)
```

```
## [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"
```

```
names(df_2022)
```

```
## [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"
```

3.4 Checking Missing Value

Checking for missing values is important to ensure that we have accurate data. Missing values can cause issues in data analysis and can affect the validity of our results. Therefore, we need to check for missing values and handle them appropriately.

```
sum(is.na(df_2018))
```

```
## [1] 0
```

```
sum(is.na(df_2019))
```

```
## [1] 0
```

```
sum(is.na(df_2020))
```

```
## [1] 0
```

```
sum(is.na(df_2021))
```

```
## [1] 0
```

```
sum(is.na(df_2022))
```

```
## [1] 7
```

3.5 Delete Missing Value

In the previous step, we observed that we have 7 missing values in 2022 dataset. We need to determine if we can be remove them or not.

```
which(is.na(df_2022), arr.ind = TRUE, useNames = TRUE)
```

```
##      row col
## [1,] 147   3
## [2,] 147   4
## [3,] 147   5
## [4,] 147   6
## [5,] 147   7
## [6,] 147   8
## [7,] 147   9
```

Since all the missing values are in the same row and there are more missing values than filled values in that row, we can remove the entire row.

```
df_2022 <- na.omit(df_2022)
sum(is.na(df_2022))
```

```
## [1] 0
```

Now, we have cleaned the data and there are no missing value anymore.

3.6 Adding Year Field

In the next step, we will bind all the datasets together. Therefore, we need to add a year column to differentiate the data from different datasets.

```
df_2018$Year <- rep("2018", nrow(df_2018))
df_2019$Year <- rep("2019", nrow(df_2019))
df_2020$Year <- rep("2020", nrow(df_2020))
df_2021$Year <- rep("2021", nrow(df_2021))
df_2022$Year <- rep("2022", nrow(df_2022))
names(df_2018)
```

```
##  [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"    "Year"
```

3.7 Binding Datasets

In this step, we combined all of the datasets into a single dataframe named "dfAll".

```
dfAll <- rbind(df_2018,df_2019,df_2020,df_2021,df_2022)
```

4. Exploratory Data Analysis

4.1 Statistic Descriptive

4.1.1 Overview

This is the overview of the datasets we have.

```
glimpse(dfAll)
```

```
## Rows: 760
## Columns: 10
## $ Overall.rank                <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13~
## $ Country.or.region           <chr> "Finland", "Norway", "Denmark", "Iceland"~
## $ Score                       <dbl> 7.632, 7.594, 7.555, 7.495, 7.487, 7.441,~
## $ GDP.per.capita              <dbl> 1.305, 1.456, 1.351, 1.343, 1.420, 1.361,~
## $ Social.support              <dbl> 1.592, 1.582, 1.590, 1.644, 1.549, 1.488,~
## $ Healthy.life.expectancy     <dbl> 0.874, 0.861, 0.868, 0.914, 0.927, 0.878,~
## $ Freedom.to.make.life.choices <dbl> 0.681, 0.686, 0.683, 0.677, 0.660, 0.638,~
## $ Generosity                  <dbl> 0.202, 0.286, 0.284, 0.353, 0.256, 0.333,~
## $ Perceptions.of.corruption   <chr> "0.393", "0.340", "0.408", "0.138", "0.35~
## $ Year                        <chr> "2018", "2018", "2018", "2018", "2018", "~
```

```
print(unique(dfAll$Country.or.region))
```

```
##   [1] "Finland"                 "Norway"
##   [3] "Denmark"                 "Iceland"
##   [5] "Switzerland"             "Netherlands"
##   [7] "Canada"                  "New Zealand"
##   [9] "Sweden"                  "Australia"
##  [11] "United Kingdom"          "Austria"
##  [13] "Costa Rica"              "Ireland"
##  [15] "Germany"                 "Belgium"
##  [17] "Luxembourg"              "United States"
##  [19] "Israel"                  "United Arab Emirates"
##  [21] "Czech Republic"          "Malta"
##  [23] "France"                  "Mexico"
##  [25] "Chile"                   "Taiwan"
##  [27] "Panama"                  "Brazil"
##  [29] "Argentina"               "Guatemala"
##  [31] "Uruguay"                 "Qatar"
##  [33] "Saudi Arabia"            "Singapore"
##  [35] "Malaysia"                "Spain"
##  [37] "Colombia"                "Trinidad & Tobago"
##  [39] "Slovakia"                "El Salvador"
##  [41] "Nicaragua"               "Poland"
##  [43] "Bahrain"                 "Uzbekistan"
##  [45] "Kuwait"                  "Thailand"
##  [47] "Italy"                   "Ecuador"
##  [49] "Belize"                  "Lithuania"
##  [51] "Slovenia"                "Romania"
##  [53] "Latvia"                  "Japan"
##  [55] "Mauritius"               "Jamaica"
##  [57] "South Korea"             "Northern Cyprus"
##  [59] "Russia"                  "Kazakhstan"
##  [61] "Cyprus"                  "Bolivia"
##  [63] "Estonia"                 "Paraguay"
##  [65] "Peru"                    "Kosovo"
##  [67] "Moldova"                 "Turkmenistan"
##  [69] "Hungary"                 "Libya"
```

```
##  [71] "Philippines"                "Honduras"
##  [73] "Belarus"                    "Turkey"
##  [75] "Pakistan"                   "Hong Kong"
##  [77] "Portugal"                   "Serbia"
##  [79] "Greece"                     "Lebanon"
##  [81] "Montenegro"                 "Croatia"
##  [83] "Dominican Republic"         "Algeria"
##  [85] "Morocco"                    "China"
##  [87] "Azerbaijan"                 "Tajikistan"
##  [89] "Macedonia"                  "Jordan"
##  [91] "Nigeria"                    "Kyrgyzstan"
##  [93] "Bosnia and Herzegovina"     "Mongolia"
##  [95] "Vietnam"                    "Indonesia"
##  [97] "Bhutan"                     "Somalia"
##  [99] "Cameroon"                   "Bulgaria"
## [101] "Nepal"                      "Venezuela"
## [103] "Gabon"                      "Palestinian Territories"
## [105] "South Africa"               "Iran"
## [107] "Ivory Coast"                "Ghana"
## [109] "Senegal"                    "Laos"
## [111] "Tunisia"                    "Albania"
## [113] "Sierra Leone"               "Congo (Brazzaville)"
## [115] "Bangladesh"                 "Sri Lanka"
## [117] "Iraq"                       "Mali"
## [119] "Namibia"                    "Cambodia"
## [121] "Burkina Faso"               "Egypt"
## [123] "Mozambique"                 "Kenya"
## [125] "Zambia"                     "Mauritania"
## [127] "Ethiopia"                   "Georgia"
## [129] "Armenia"                    "Myanmar"
## [131] "Chad"                       "Congo (Kinshasa)"
## [133] "India"                      "Niger"
## [135] "Uganda"                     "Benin"
## [137] "Sudan"                      "Ukraine"
## [139] "Togo"                       "Guinea"
## [141] "Lesotho"                    "Angola"
## [143] "Madagascar"                 "Zimbabwe"
## [145] "Afghanistan"                "Botswana"
## [147] "Malawi"                     "Haiti"
## [149] "Liberia"                    "Syria"
## [151] "Rwanda"                     "Yemen"
## [153] "Tanzania"                   "South Sudan"
## [155] "Central African Republic"   "Burundi"
## [157] "North Macedonia"            "Gambia"
## [159] "Swaziland"                  "Comoros"
## [161] "Taiwan Province of China"   "Trinidad and Tobago"
## [163] "North Cyprus"               "Hong Kong S.A.R. of China"
## [165] "Maldives"                   "Luxembourg*"
## [167] "Czechia"                    "Guatemala*"
## [169] "Kuwait*"                    "Belarus*"
## [171] "Turkmenistan*"              "North Cyprus*"
## [173] "Libya*"                     "Azerbaijan*"
## [175] "Gambia*"                    "Liberia*"
## [177] "Congo"                      "Niger*"
```

```
## [179] "Comoros*"                  "Palestinian Territories*"
## [181] "Eswatini, Kingdom of*"     "Madagascar*"
## [183] "Chad*"                     "Yemen*"
## [185] "Mauritania*"               "Lesotho*"
## [187] "Botswana*"                 "Rwanda*"
```

By using this overview, we can quickly inspect the structure and contents of the earliest and most recent dataset.

```
head(df_2018)
```

```
##   Overall.rank Country.or.region Score GDP.per.capita Social.support
## 1            1           Finland 7.632          1.305          1.592
## 2            2            Norway 7.594          1.456          1.582
## 3            3           Denmark 7.555          1.351          1.590
## 4            4           Iceland 7.495          1.343          1.644
## 5            5       Switzerland 7.487          1.420          1.549
## 6            6       Netherlands 7.441          1.361          1.488
##   Healthy.life.expectancy Freedom.to.make.life.choices Generosity
## 1                   0.874                        0.681      0.202
## 2                   0.861                        0.686      0.286
## 3                   0.868                        0.683      0.284
## 4                   0.914                        0.677      0.353
## 5                   0.927                        0.660      0.256
## 6                   0.878                        0.638      0.333
##   Perceptions.of.corruption Year
## 1                     0.393 2018
## 2                     0.340 2018
## 3                     0.408 2018
## 4                     0.138 2018
## 5                     0.357 2018
## 6                     0.295 2018
```

```
head(df_2022)
```

```
##   Overall.rank Country.or.region Score GDP.per.capita Social.support
## 1            1           Finland 7.821          1.892          1.258
## 2            2           Denmark 7.636          1.953          1.243
## 3            3           Iceland 7.557          1.936          1.320
## 4            4       Switzerland 7.512          2.026          1.226
## 5            5       Netherlands 7.415          1.945          1.206
## 6            6       Luxembourg* 7.404          2.209          1.155
##   Healthy.life.expectancy Freedom.to.make.life.choices Generosity
## 1                   0.775                        0.736      0.109
## 2                   0.777                        0.719      0.188
## 3                   0.803                        0.718      0.270
## 4                   0.822                        0.677      0.147
## 5                   0.787                        0.651      0.271
## 6                   0.790                        0.700      0.120
##   Perceptions.of.corruption Year
## 1                     0.534 2022
## 2                     0.532 2022
## 3                     0.191 2022
```

```
## 4                    0.461 2022
## 5                    0.419 2022
## 6                    0.388 2022
```

4.1.2 Mean

This is the mean value of the world happiness score, year by year.

```
dfAll_mean_by_year <- data.frame(dfAll %>%
  group_by(Year) %>%
  summarize(mean(Score))) %>%
  print()
```

```
##   Year mean.Score.
## 1 2018    5.375917
## 2 2019    5.407096
## 3 2020    5.473240
## 4 2021    5.532839
## 5 2022    5.553575
```

Additionally, this is the calcuation of the overall mean value of the world happiness score from 2018-2022.

```
dfAll_mean <- dfAll %>%
  summarize(mean(Score)) %>%
  print()
```

```
##   mean(Score)
## 1    5.466804
```

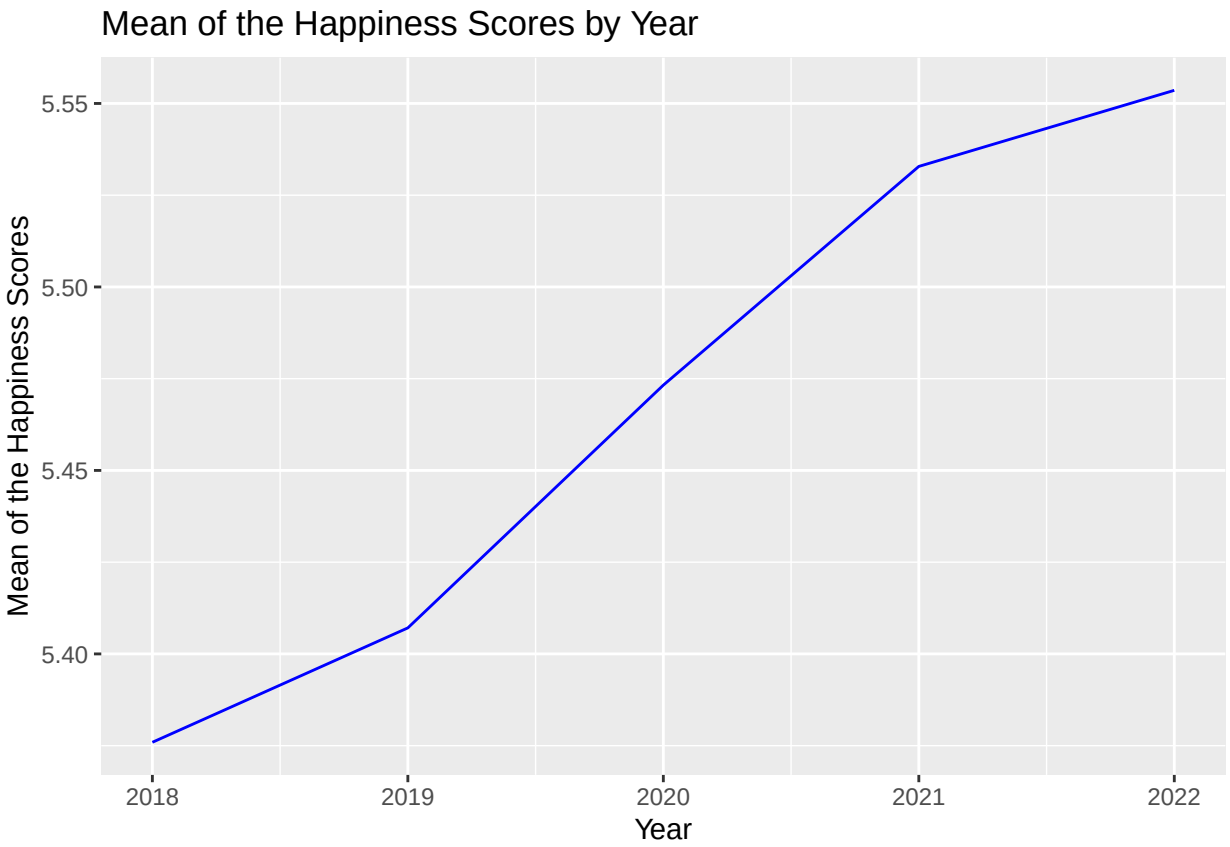
Then, to create a visualization using ggplot2, we need to ensure that the data we want to plot is in a numeric data type.

```
str(dfAll_mean_by_year)
```

```
## 'data.frame':    5 obs. of  2 variables:
##  $ Year       : chr  "2018" "2019" "2020" "2021" ...
##  $ mean.Score.: num  5.38 5.41 5.47 5.53 5.55
```

As we can see, the year field is in a character data type. Therefore, we need to convert it to numeric.

```
dfAll_mean_by_year$Year <- as.numeric(dfAll_mean_by_year$Year)
ggplot(data = dfAll_mean_by_year, aes(x = Year, y = mean.Score.)) +
  geom_line(color = "blue") +
  labs(title = "Mean of the Happiness Scores by Year", x = "Year", y = "Mean of the Happiness Scores")
```



Based on this plot, we can see that the mean hapiness score increases year by year. It means, the level of happiness of people from all over the world increases.

4.1.3 Median

This is the median value of the world happiness score, year by year.

```
dfAll_median_by_year <- data.frame (dfAll %>%
  group_by(Year) %>%
  summarize(median(Score)))%>%
  print()
```

```
##   Year median.Score.
## 1 2018        5.3780
## 2 2019        5.3795
## 3 2020        5.5150
## 4 2021        5.5340
## 5 2022        5.5685
```

Additionally, this is the median value of world happiness score from 2018-2022.

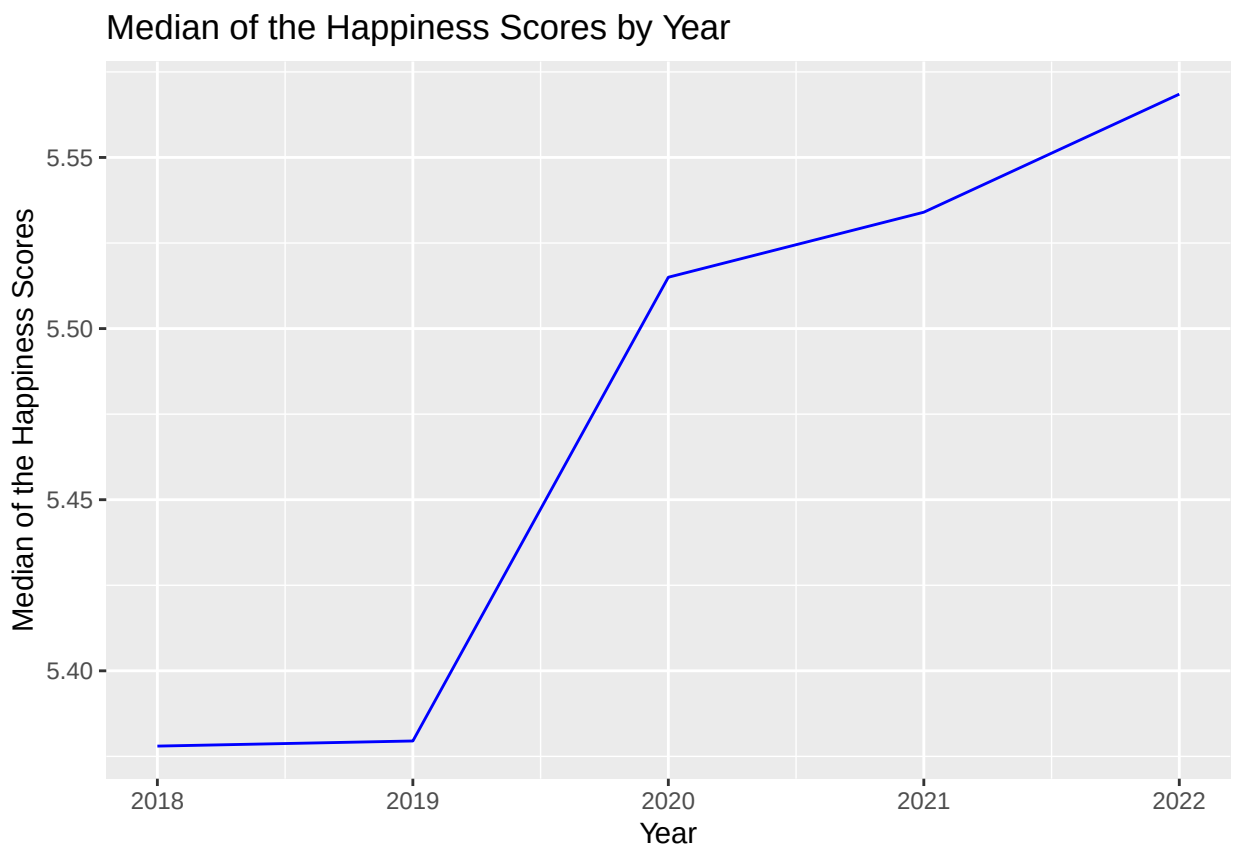
```
dfAll_median <- dfAll%>%
  summarize(median(Score)) %>%
  print()
```

```
##   median(Score)
## 1         5.477
```

```
str(dfAll_median_by_year)
```

```
## 'data.frame':    5 obs. of  2 variables:
##  $ Year         : chr  "2018" "2019" "2020" "2021" ...
##  $ median.Score.: num  5.38 5.38 5.51 5.53 5.57
```

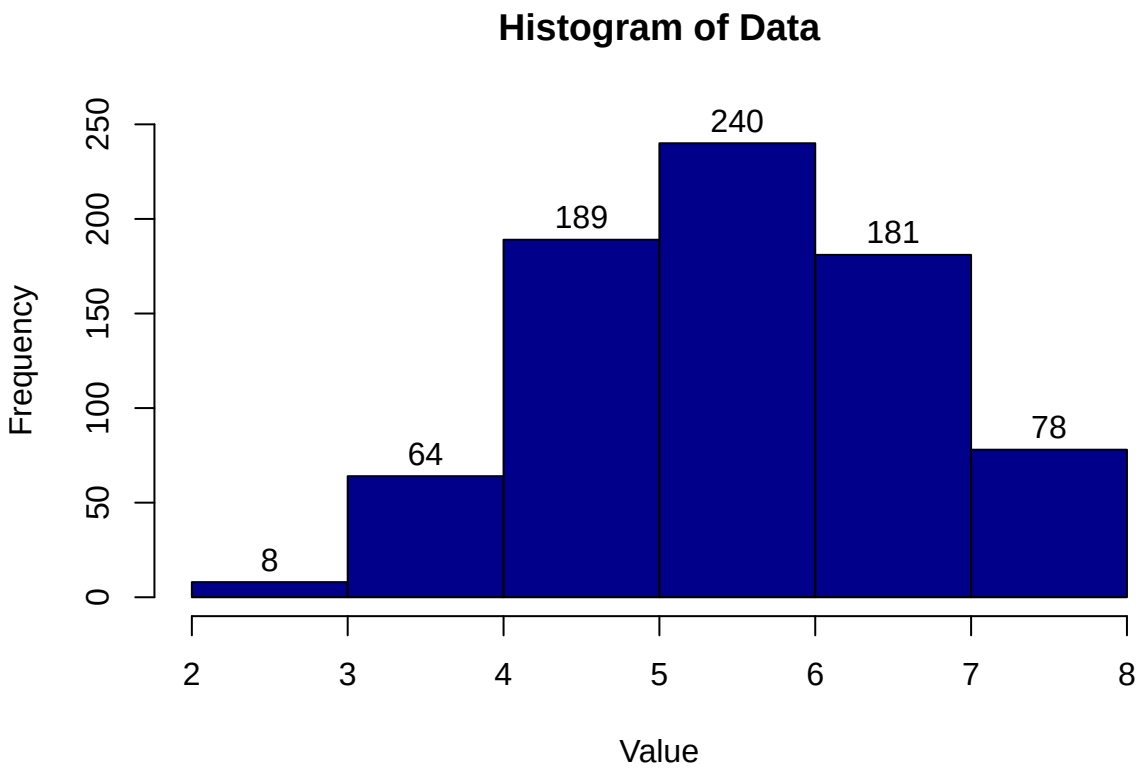
```
dfAll_median_by_year$Year<-as.numeric(dfAll_median_by_year$Year)
ggplot(data = dfAll_median_by_year, aes(x = Year, y = median.Score.)) + geom_line(color = "blue") +labs
              x = "Year", y = "Median of the Happiness Scores")
```



4.1.4 Mode

This is the mode value of the world happiness score, year by year.

```
hist(dfAll$Score, breaks = 5, col = "darkblue",
     labels = TRUE, ylim = c(0,250),
     xlab = "Value", ylab = "Frequency",
     main = "Histogram of Data" )
```



According to this histogram, we can see the distribution of the happiness score is mostly distributed in the range of 4-7.

4.2 Multivariate Analysis

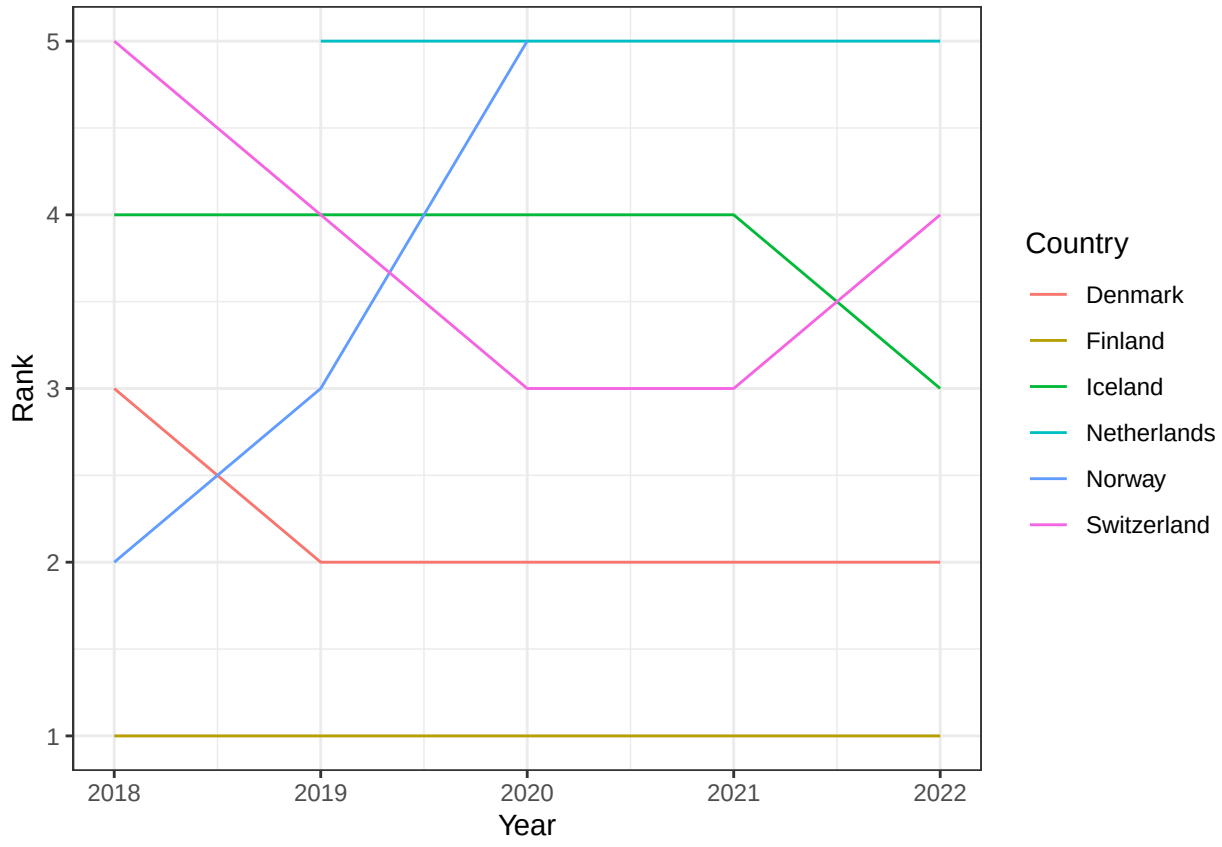
4.2.1 How Have the Rankings Changed Over Time?

```
dfAll$Year <- as.numeric(dfAll$Year)
dfAll_rank <- dfAll %>%
            select(Year, Country.or.region, Overall.rank) %>%
            filter(Overall.rank <= 5) %>%
            print()
```

```
##    Year Country.or.region Overall.rank
## 1  2018           Finland            1
## 2  2018            Norway            2
## 3  2018           Denmark            3
## 4  2018           Iceland            4
## 5  2018       Switzerland            5
## 6  2019           Finland            1
## 7  2019           Denmark            2
## 8  2019            Norway            3
## 9  2019           Iceland            4
## 10 2019       Netherlands            5
```

```
## 11 2020           Finland           1
## 12 2020           Denmark           2
## 13 2020       Switzerland           3
## 14 2020           Iceland           4
## 15 2020            Norway           5
## 16 2021           Finland           1
## 17 2021           Denmark           2
## 18 2021       Switzerland           3
## 19 2021           Iceland           4
## 20 2021       Netherlands           5
## 21 2022           Finland           1
## 22 2022           Denmark           2
## 23 2022           Iceland           3
## 24 2022       Switzerland           4
## 25 2022       Netherlands           5
```

```
ggplot(dfAll_rank, aes(x = Year, y = Overall.rank, color = Country.or.region)) +
  geom_line() +
  labs(x = "Year", y = "Rank", color = "Country") +
  theme_bw()
```



The line plot reveals that Denmark, Finland, Iceland, Netherlands, Norway, and Switzerland are the 6 countries that consistently rank in the top 5 during 2018-2022. Interestingly, only Finland, Denmark, and Iceland managed to maintain a consistently high rank in happiness score throughout the years. Moreover, the plot also shows some fluctuations in rankings over the years, implying that happiness levels in different countries are not always stable. In addition to the insights gained from the line plot, we can also explore

the correlation between parameters that contribute to the happiness score. For instance, we can analyze the correlation between GDP per capita and life expectancy, or the correlation between social support and generosity. By examining these correlations, we can gain a better understanding of the factors that have the strongest influence on a country's happiness score, and how they may have changed over time.

4.2.2 Correlation Between Parameters (What Factors Affect the Most to Happiness Score?

```
subset_dfAll <- dfAll[, c("Score", "GDP.per.capita", "Social.support", "Healthy.life.expectancy",
                "Freedom.to.make.life.choices", "Generosity", "Perceptions.of.corruption")]
str(subset_dfAll)
```

```
## 'data.frame':    760 obs. of  7 variables:
##  $ Score                       : num  7.63 7.59 7.55 7.5 7.49 ...
##  $ GDP.per.capita              : num  1.3 1.46 1.35 1.34 1.42 ...
##  $ Social.support              : num  1.59 1.58 1.59 1.64 1.55 ...
##  $ Healthy.life.expectancy     : num  0.874 0.861 0.868 0.914 0.927 0.878 0.896 0.876 0.913 0.91 ...
##  $ Freedom.to.make.life.choices: num  0.681 0.686 0.683 0.677 0.66 0.638 0.653 0.669 0.659 0.647 ...
##  $ Generosity                  : num  0.202 0.286 0.284 0.353 0.256 0.333 0.321 0.365 0.285 0.361 ...
##  $ Perceptions.of.corruption   : chr  "0.393" "0.340" "0.408" "0.138" ...
```

```
subset_dfAll$Perceptions.of.corruption <- as.numeric(subset_dfAll$Perceptions.of.corruption)
```

```
## Warning: NAs introduced by coercion
```

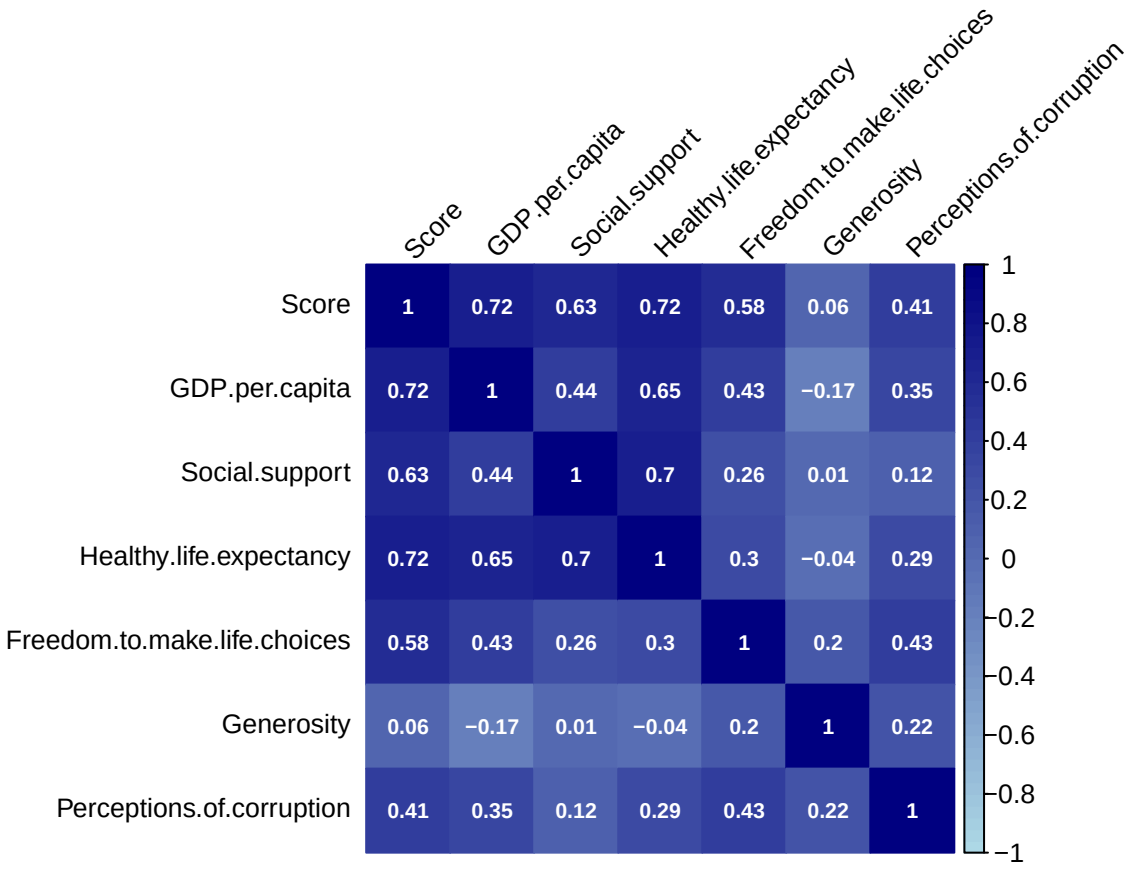
```
sum(is.na(subset_dfAll))
```

```
## [1] 1
```

```
subset_dfAll <- na.omit(subset_dfAll)

corr_matrix <- cor(subset_dfAll, method = "pearson")

corrplot(corr_matrix, method = "color", type = "full", tl.col = "black",
         tl.srt = 45, tl.cex = 0.8, col = colorRampPalette(c("#ADD8E6", "#000080"))(50),
         addCoef.col = "white", number.cex = 0.7)
```



According to the correlation plot above, We can observe a strong correlation between the GDP per capita and healthy life expectancy, as well as between GDP per capita and social support. This indicates that a country with a higher GDP is likely to have better outcomes in terms of both health and social support, which in turn contributes to greater overall happiness in that country. Therefore, a higher GDP tends to be associated with a higher happiness score.

Based on the correlation analysis conducted, it appears that increasing a country's GDP per capita may contribute to a higher happiness level. Additionally, improving factors such as healthy life expectancy and social support may also positively impact a country's happiness level. Therefore, to increase a country's happiness level, policies and initiatives that focus on improving the economy, healthcare, and social welfare may be effective. However, it is important to note that happiness is a complex construct that cannot be solely attributed to these factors, and other cultural, social, and psychological factors may also play a significant role.