POLS 3316 Project: Analyzing International Soccer Results

Contents

1 WRITE UP

1.1 THE DATA

The data used for this project is a dataset on Kaggle that contains data on all international soccer matches throughout history, from 1872 to 2022. It can be found by clicking here: *International football results from 1872 to 2022.* The dataset is updated monthly. In total, it includes 43,170 results from various international tournaments such as the FIFA World Cup, and African Cup of Nations, etc. as well as regular friendly matches. It does not include club soccer data, which would have included matches from leagues such as the English Premier League and the UEFA Champions League, the biggest club soccer tournament in Europe. There is only data on matches that include international teams (which are mostly the national teams of various countries).

Each row contains information on one match, with each of the nine columns containing nine variables. The variables include the date of the match, the home team, the away team, the home score, the away score, the name of the tournament the match was part of, the city where the match took place, the country where the match took place, and the match's neutral status. The variables that are the most important to the project are: the date, home team, away team, home score, away score, and the tournament.

1.2 THE VARIABLES

For my project, I took two main subsets from the raw data, these data frames were subsetted based on these two elements of the tournament variable:

- FIFA World Cup
- Friendly

One contained all matches throughout the history of the FIFA World Cup and the other contained all friendly matches since 1872. They were chosen because each was ideal for studying a specific relationship between variables.

I wanted to use the FIFA World Cup dataset to study the relationship between two variables (the team's total Goal Difference and Goal Difference per matches played) and a team's total points in FIFA World Cup history (a measure of its performance). Goal difference is an important soccer statistic that measures how many goals a team scored minus how many goals it lets its opponent team score. In simpler words, goal difference is goals scored minus goals conceded.

For the friendly matches data, I wanted to study the relationship between a team's mean score as a home team and its mean score as an away team. If there was no home advantage, the mean difference between these variables should ideally be close to zero. I want to evaluate if there is a home-field advantage, that is, doo teams score better at home? The main issue with studying these variables is that the dataset does not provide variables such as every FIFA World Cup Team's total goal difference and total points or the mean home and away scores of every international team that has played friendly matches. However, this data can be calculated from the raw data.

1.3 THE DATA CLEANING PROCESS

To get the variables needed for the project, I performed some calculations and data cleaning. I will describe that process in this part of the write-up. In order to make the data cleaning process easier, I created custom functions to use in the process in Section 2.1. The comments in the code chunk explain what each function does. Then, I did some preliminary data cleaning in Section 2.3 of the project to extract a list of every international team in history. It is an interesting list to look at—the appendix contains the entire list for that reason—but it is also useful later in the code when data cleaning and calculating various team scores. The data cleaning in Sections 2.4 and 3.1 is done mostly to extract interesting information from the dataset.

1.3.1 FIFA WORLD CUP

GENERAL DATA CLEANING

For an explanation of how the variables that were used to study relationships and perform regression and hypothesis testing were extracted, skip to the next heading.

In Section 2.4, I generated a subset of every FIFA World Cup match. Naturally, for a tournament like the FIFA World Cup, the most interesting part of the tournament is the final match itself. Although this information won't be useful in the main part of the project when studying relationships and performing regression and hypothesis testing, it is interesting to look at and visualize, and I have done so in this project. The dataset does not specify which matches are final matches, so I went on the internet and collected some data. I created three vectors based on the data I collected. The world_cup vector contains the years every world cup was held, and start_date and end_date list the exact dates each world cup started and ended, respectively. Then, applying one of my custom functions, I created a subset of the FIFA World Cup subset containing only the matches that occurred on the end dates of a tournament. This created a list of 23 data frames, with each data frame containing matches on the end date. I combined this list of data frames into one data frame. This data frame listed every match occurring on the end date, but some of these matches were not the final matches. There were unique situations in 1938 and 1950 mentioned in the comments that resulted in multiple matches those years on the end date, so I deleted those rows to finally create a data frame of every FIFA World Cup final sorted by year. The entire data frame can be found in the appendix at the end of this document.

I then wanted to find the winner of each world cup final, so I created three subsets of the FIFA World Cup Final dataset that were each a data frame of matches. The first data frame contained matches when the "home team" won and the second contained matches when the "away team" won. The third data frame contained draw matches where the winner was decided by penalty shoot-outs and then by deleting columns, I only retained the winner of the shoot-out. Finally, I combined these data frames to create a new data frame that listed which team won the FIFA World Cup every year. This data frame can also be found in the appendix. Upon observing that during all the 21 FIFA World Cups, only a few teams kept winning, again and again, I created a new data frame that listed out the eight champion countries that have won the world cup in the past. This data frame can also be found in the appendix.

In Section 3.1, I perform a combination of data analysis and data cleaning. I begin by finding the mean score of the eight champion teams throughout history by combining their mean scores as "home teams" and mean scores as "away teams." I did this one by one, and the code can be viewed. Then I wanted to find the mean score of every FIFA World Cup Participant, but to do that would have been a laborious process of repeating the same code of chunk 81 times and then combining everything together, so I used custom functions and applied them to quickly return a data frame containing every team and its mean score. The scatterplots of every champion team's mean scores and every team's mean score can be found in Section 4.1 in the Data Visualization section of this document.

COMPUTING VARIABLES NEEDED FOR REGRESSION

In Section 3.1.3, I calculate different soccer stats for all the teams that have participated in the FIFA World Cup. My goal was to calculate: mathes won, matches lost, matches draw, and total points. The points are calculated by multiplying the number one wins by 3 and adding it to the number of draws.

I do this by creating a copy of the FIFA World Cup data frame and adding a column that lists the match winner and a column that lists the match loser. If the match is a draw, the word "DRAW" is listed instead for those columns. Then I mutate the data frame again and create two new columns that each list the name of the two teams if the match is a draw, or else list "NOT DRAW." Then I filter out all the rows where the value is not "NOT DRAW," creating a data frame of 119 rows that contains all the draw matches. Then based on this, I create a data frame that shows each team and how many matches it won, a data frame that shows how many matches it lost, and a dataframe that shows how many matches were a draw. Then I merged these dataframes together and calculated a new column called Pts that calculated team points by adding the number of draws to three times the number of wins.

Then, I used various custom functions and applied them to the dataset to compute Goal Difference and Average Goal Difference and then merged the result with the previous data frame to create a data frame that shows wins, losses, draws, points, goal difference, and average goal difference for every team. This is the dataset that I used to study relationships and perform regression and hypothesis testing.

1.3.2 FRIENDLY MATCHES

GENERAL DATA CLEANING

The friendly match dataset did not require much data cleaning. Through the same process I used for FIFA World Cup participants, I found the mean score for every team that has participated in friendly matches.

COMPUTING VARIABLES NEEDED FOR REGRESSION

I also found every team's mean score as a home team and its mean score as an away team. I studied the relationship between these two variables and used them to perform regression and hypothesis testing later in the project.

1.4 RELATIONSHIPS

1.4.1 FIFA WORLD CUP

I hypothesized that a team's points are dependent on its goal difference and its average goal difference. I hypothesized a positive relationship where higher goal difference and average goal difference predicts higher points for the team.

The Null Hypotheses were:

- There is no relationship between GD and Points
- There is no relationship between Average GD and Points

1.4.2 FRIENDLY MATCHES

I hypothesized that teams do not tend to have similar scores when playing at home and playing away; the mean difference is not approximately zero. I hypothesized that teams tend to score more at home than away, that there is a home-field advantage. The mean home score will be higher than the mean away score.

The Null Hypothesis was:

- The mean difference between the average home score and average away score is zero; the two scores tend to be similar

1.5 REGRESSION AND HYPOTHESIS TESTING

1.5.1 FIFA WORLD CUP

I performed two OLS regressions with both of them having points as the independent variable and either goal difference or average goal difference as the dependent variable. Based on the results of the regression, both the null hypotheses were rejected. There appears to be a very significant relationship between GD and Points and between Average Goal Difference and Points.

I also created two linear regression plots for both relationships, but I removed outliers for the plots (something I didn't do for OLS and hypothesis testing). I removed any team that had played seven or less matches in the world cup. There was a linear relationship between average goal difference team points, but for goal difference and team points, there appears to be a negative relationship but it flattens at a point where goal difference is less than zero and where most of the data points are plotted. From there, the relationship appears to be positive, with the line increasing at a decreasing rate. I believe this type of relationship exists because of outlying teams that have high points but very low goal differences (these could be teams that were once great and therefore have many points, but have recently failed to perform and keep conceding scores to their opponents). This also highlights that there are some teams that are very high performing and other teams that are not. Majority of the teams have around the same points, but some have starkly higher points and greater goal differences.

I also performed a Chi-squared test for both hypotheses. The results were very significant. The tests suggested a significant relationship between Average Goal Difference and Team Points, the p-value for this test was 0.0004998. There was also a strong, statistically-significant relationship between Goal Difference and Team Points, with a p-value of 0.008996, but it appears to be slightly weaker than between Total Goal Difference and Team Points. This might be because Average Goal Difference does not account for how much a team has participated in the FIFA World Cups, since it's an "average."

Obviously teams that tend to score higher points tend to qualify more often for the World Cup and therefore obtain higher Total Goal Difference. Amount of participation (by being good enough to qualify) might be the third variable that's making the relationship between Goal Difference and Points more significant than just Average Goal Difference. In conclusion, Higher Average Goal Difference predicts higher Team Points, with less accuracy than Total Goal Difference.

1.5.2 FRIENDLY MATCHES

In Section 3.2.2, I try to calculate how much more teams tend to score as home teams. My calculations showed that there might be a possible advantage since teams had a tendency to score 67.28% higher at home. I compare the average scores at home and away using a bar plot in Section 4.2.2 of the Data Visualization part of this document, and the home score has a much higher bar.

So in order to test this relationship and whether it was significant, I performed a Paired t-test and the null hypothesis was rejected. The test showed that home and away Scores do not tend to be similar and their mean difference is not approximately zero. In fact, there is a mean difference of 0.5212705, therefore, home scores tend to be much higher. The result was very significant, with the test producing a p-value of less than 0.00000000000000022.

1.6 CONCLUSION

I enjoyed exploring this dataset on international soccer matches. This was my first time learning and using the R language and it allowed me to learn a lot of data cleaning, data analysis, and data visualization skills. OLS Regression and Hypothesis Testing turned out to be amazing tools to test relationships between various variables.

All three of the null hypotheses were rejected. My analysis of FIFA World Cup data showed that both a team's goal difference and average goal difference predict higher points. This proves a broader point: the importance of good defense in soccer. Goal difference is in a way a measure of how good a team's defense is because it subtracts the goals it has allowed another team to score from its score. The relationship was very significant. It can be proven that teams that invest in better defenders and improve their defense will tend to perform better at the World Cup. My analysis on the friendly matches data clearly showed a home-field advantage exists. It was a very significant relationship, with the p-value very close to zero. A home-field advantage cannot be denied and this allows us to make another important argument: the importance of a neutral field in tournament matches. If the home teams have such a significant advantage, then this advantage needs to be removed for tournaments like the FIFA World Cup and the African Cup of Nations. Otherwise, it is an unfair advantage to the home team and removes the purpose of the tournament.

```r
library(rmarkdown)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(tidyr)
library(dbplyr)
```

```
##
## Attaching package: 'dbplyr'
##
## The following objects are masked from 'package:dplyr':
##
##     ident, sql
```

```r
library(vctrs)
```

```
##
## Attaching package: 'vctrs'
##
## The following object is masked from 'package:dplyr':
##
##     data_frame
##
## The following object is masked from 'package:tibble':
##
##     data_frame
```

```r
library(ggplot2)
library(ggrepel)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
library(writexl)
library(data.table)
```

```
##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
##
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(knitr)
```

2 DATA CLEANING

2.1 CREATING CUSTOM FUNCTIONS FOR THE PROJECT

```
subset.tournament <- function(column) {
  new.df <- subset(raw_data, tournament == column)
}
tournament.total.matches <- function(tournament_matrix, team) {
  new.df <- subset(tournament_matrix, home_team == team | away_team == team)
}
participating.teams <- function(tournament_matrix) {
  x <- as.matrix(count(tournament_matrix, home_team))
  y <- as.matrix(count(tournament_matrix, away_team))
  X <- x[,-2]
  Y <- y[,-2]
  z <- intersect(X,Y)
  a <- setdiff(X,Y)
  b <- setdiff(Y,X)
  p <- sort(c(z, c(a,b)))
}
home.score.mean <- function(tournament_matrix, team) {
  x <- subset(tournament_matrix, tournament_matrix$home_team == team)
  y <- mean(x$home_score)
  y
}
away.score.mean <- function(tournament_matrix, team) {
  x <- subset(tournament_matrix, tournament_matrix$away_team == team)
  y <- mean(x$away_score)
  y
}
score.mean <- function(tournament_matrix, team) {
  x1 <- subset(tournament_matrix, tournament_matrix$home_team == team)
  y1 <- subset(tournament_matrix, tournament_matrix$away_team == team)
  sum_x <- sum(x1$home_score)
  sum_y <- sum(y1$away_score)
  x_n <- count(x1)
  y_n <- count(y1)
  z <- (sum_x+sum_y) / (x_n + y_n)
}
standard.clean <- function(tournament_matrix, columns) {
   tournament_matrix[, -c(columns)] %>% rename(
    "HOME TEAM" = home_team,
    "AWAY TEAM" = away_team,
    "HOME SCORE" = home_score,
    "AWAY SCORE" = away_score
  ) %>% rename("MATCH DATE" = date)
}
simple.clean <- function(tournament_matrix) {
  colnames(tournament_matrix) <- gsub("_", " ", colnames(tournament_matrix))
  colnames(tournament_matrix) <- toupper(colnames(tournament_matrix))
  tournament_matrix
}
```

```r
world.cup.total.matches.cleaned <- function(team) {
  new.df <- subset(FIFA_WC, home_team == team | away_team == team)
  new.df1 <- standard.clean(new.df, 6:9)
}
world.cup.total.matches.v2 <- function(team) {
  new.df <- subset(FIFA_WC_GD, home_team == team | away_team == team)
}
friendly.total.matches.cleaned <- function(team) {
  new.df <- subset(Friendly_Matches, home_team == team | away_team == team)
  new.df1 <- standard.clean(new.df, 6:9)
}
subset.wc.final.matches <- function(x) {
  subset(FIFA_WC, FIFA_WC$date == x)
}
#FUNCTIONS EXPLAINED:

#subset.tournament():
#Isolates any Tournament Data into a New Data Frame by subsetting Raw Data

#tournament.total.matches():
#Creates a data frame to show matches of a given team from a specific tournament

#participating.teams():
#Creates a vector that lists all participating teams in a specific tournament

#home.score.mean()
#Calculates average home score of a given team in a given tournament

#away.score.mean()
#Calculates average away score of a given team in a given tournament

#score.mean()
#Calculates average score of a given team in a given tournament

#standard.clean()
#Modifies any data frame taken from the raw data to fit a specific standard look
#and can also delete specific columns

#simple.clean()
#Cleans up the look of any data frame taken from the raw data, mainly by making
#column names more readable to users

#world.cup.total.matches.cleaned()
#Generates a new data frame from the FIFA world cup subset of the raw data that
#only contains the matches the inputed team participated in, cleans this data frame,
#and returns it

#world.cup.total.matches.v2()
#Generates and cleans a data frame of all matches for the inputed team including
#columns giving home and away goal differences

#subset.wc.final.matches()
#Creates a subset of all FIFA matches that took place on the end dates of the tournaments
```

2.2 IMPORTING RAW DATA

```
#Raw Data Imported
raw_data <- read_csv("international_soccer_results.csv", show_col_types = FALSE)
raw_data %>% simple.clean()
```

```
## # A tibble: 43,752 x 9
##    DATE       `HOME TEAM` AWAY T~1 HOME ~2 AWAY ~3 TOURN~4 CITY  COUNTRY NEUTRAL
##    <date>     <chr>       <chr>      <dbl>   <dbl> <chr>   <chr> <chr>   <lgl>
##  1 1872-11-30 Scotland    England        0       0 Friend~ Glas~ Scotla~ FALSE
##  2 1873-03-08 England     Scotland       4       2 Friend~ Lond~ England FALSE
##  3 1874-03-07 Scotland    England        2       1 Friend~ Glas~ Scotla~ FALSE
##  4 1875-03-06 England     Scotland       2       2 Friend~ Lond~ England FALSE
##  5 1876-03-04 Scotland    England        3       0 Friend~ Glas~ Scotla~ FALSE
##  6 1876-03-25 Scotland    Wales          4       0 Friend~ Glas~ Scotla~ FALSE
##  7 1877-03-03 England     Scotland       1       3 Friend~ Lond~ England FALSE
##  8 1877-03-05 Wales       Scotland       0       2 Friend~ Wrex~ Wales   FALSE
##  9 1878-03-02 Scotland    England        7       2 Friend~ Glas~ Scotla~ FALSE
## 10 1878-03-23 Scotland    Wales          9       0 Friend~ Glas~ Scotla~ FALSE
## # ... with 43,742 more rows, and abbreviated variable names 1: `AWAY TEAM`,
## #   2: `HOME SCORE`, 3: `AWAY SCORE`, 4: TOURNAMENT
## # i Use `print(n = ...)` to see more rows
```

2.3 CREATING A LIST OF EVERY INTERNATIONAL TEAM IN HISTORY

```
#Two Matrices with 2 Columns [Column1: Home/Away Team & Column2: Team Match Count]
Home_Team_Matrix <- as.matrix(count(raw_data, home_team))
Away_Team_Matrix <- as.matrix(count(raw_data, away_team))

#Column2 Deleted to Create a List of all Home Teams and a List of all Away Teams
Home_Teams <- Home_Team_Matrix[,-2]
Away_Teams <- Away_Team_Matrix[,-2]

#List of Teams That Have Played both at Home and Away
Common_Teams <- intersect(Home_Teams, Away_Teams)

#List of Teams That Have Only Played at Home or Away
Unique_Teams <- c(setdiff(Home_Teams, Away_Teams), setdiff(Away_Teams, Home_Teams))

#Adding Common and Unique Teams to Create a Complete List of all International Teams Throughout Soccer 
International_Teams <- c(Common_Teams, Unique_Teams)
International_Teams <- data.frame(International_Teams)
head(International_Teams) %>% simple.clean()
```

```
##   INTERNATIONAL TEAMS
## 1            Abkhazia
## 2         Afghanistan
## 3       Åland Islands
## 4             Albania
## 5            Alderney
## 6             Algeria
```

2.4 FIFA WORLD CUP TOURNAMENT SUBSET

2.4.1 GENERATING SUBSET

```
#FIFA World Cup Subset
FIFA_WC <- subset.tournament("FIFA World Cup")
FIFA_WC %>% standard.clean(6:9)
```

```
## # A tibble: 900 x 5
##    `MATCH DATE` `HOME TEAM` `AWAY TEAM`   `HOME SCORE` `AWAY SCORE`
##    <date>       <chr>       <chr>                <dbl>        <dbl>
##  1 1930-07-13   Belgium     United States            0            3
##  2 1930-07-13   France      Mexico                   4            1
##  3 1930-07-14   Brazil      Yugoslavia               1            2
##  4 1930-07-14   Peru        Romania                  1            3
##  5 1930-07-15   Argentina   France                   1            0
##  6 1930-07-16   Chile       Mexico                   3            0
##  7 1930-07-17   Bolivia     Yugoslavia               0            4
##  8 1930-07-17   Paraguay    United States            0            3
##  9 1930-07-18   Uruguay     Peru                     1            0
## 10 1930-07-19   Argentina   Mexico                   6            3
## # ... with 890 more rows
## # i Use `print(n = ...)` to see more rows
```

```
FIFA_WC <- FIFA_WC %>% mutate(Match_Winner = if_else(home_score > away_score, home_team, if_else(home_s
```

2.4.2 COMBINING DATA ON FIFA WORLD CUP DATES

```
#THIS IS DONE SO THAT FINAL MATCHES CAN BE SUBSETTED LATER
world_cup <- c(1930, 1934, 1938, 1950, 1954, 1958, 1962, 1966, 1970, 1974, 1978, 1982, 1986, 1990, 1994
start_date <- as.Date(c("1930-07-03", "1934-05-27", "1938-06-04", "1950-06-24", "1954-06-16", "1958-06-
end_date <- as.Date(c("1930-07-30", "1934-06-10", "1938-06-19", "1950-07-16", "1954-07-04", "1958-06-29

FIFA_WC_Timeline <- data.frame(world_cup, start_date, end_date)
```

2.4.3 GENERATING DATA ON FIFA WORLD CUP FINAL MATCHES

```
#Using FIFA World Cup Dates Data to create a new Data Frame that only shows Matches on End Date
FIFA_WC_End_Date_Matches_List <- lapply(end_date, subset.wc.final.matches)
#list of 21 data frames
FIFA_WC_End_Date_Matches <- do.call("rbind", FIFA_WC_End_Date_Matches_List)
#combined into 1 data frame

#Data needs to be cleaned to only show Final Matches
#Rows 3 and 9 need to be deleted because:
#In 1938, there was a 3rd-place play-off on the same day
#In 1950, the world cup final was played as a round; there were 2 matches on the end date
#(I want to only included the match which that year's champion participated in)
```

```r
FIFA_WC_Final_Matches <- FIFA_WC_End_Date_Matches[-c(3,6),] #delete 2 rows
FIFA_WC_Final_Matches$world_cup <- world_cup #new column (world cup year)
FIFA_WC_Final_Matches <- relocate(FIFA_WC_Final_Matches, world_cup, .before = home_team)
FIFA_WC_Final_Matches <- relocate(FIFA_WC_Final_Matches, date, .after = away_score)
FIFA_WC_Finals <- FIFA_WC_Final_Matches[-c(6:8)]
FIFA_WC_Finals[-c(6:9)] %>% simple.clean()
```

```
## # A tibble: 21 x 5
##    `WORLD CUP` `HOME TEAM` `AWAY TEAM`    `HOME SCORE` `AWAY SCORE`
##          <dbl> <chr>       <chr>                 <dbl>        <dbl>
##  1        1930 Uruguay     Argentina                 4            2
##  2        1934 Italy       Czechoslovakia            2            1
##  3        1938 Hungary     Italy                     2            4
##  4        1950 Brazil      Uruguay                   1            2
##  5        1954 Germany     Hungary                   3            2
##  6        1958 Sweden      Brazil                    2            5
##  7        1962 Brazil      Czechoslovakia            3            1
##  8        1966 England     Germany                   4            2
##  9        1970 Brazil      Italy                     4            1
## 10        1974 Germany     Netherlands               2            1
## # ... with 11 more rows
## # i Use `print(n = ...)` to see more rows
```

2.4.4 FINDING FIFA WORLD CUP WINNERS & CURRENT CHAMPION TEAMS

```r
#Create a List of Final Winners
Home_Wins <- subset(FIFA_WC_Finals, FIFA_WC_Finals$home_score > FIFA_WC_Finals$away_score)
Home_Wins <- relocate(Home_Wins, country, .after = home_team)
Home_Wins <- Home_Wins[-c(3:6)]
Away_Wins <- subset(FIFA_WC_Finals, FIFA_WC_Finals$home_score < FIFA_WC_Finals$away_score)
Away_Wins <- relocate(Away_Wins, c(country, home_team), .after = away_team)
Away_Wins <- Away_Wins[-c(3:6)]
Shoot_Outs <- subset(FIFA_WC_Finals, FIFA_WC_Finals$home_score == FIFA_WC_Finals$away_score)
Shoot_Outs <- relocate(Shoot_Outs, c(country), .after = home_team)
#The Penalty Shoot-out winner in both observations appears to be in home_team
#Therefore, away_team will be deleted to only show winner in the data frame
Shoot_Outs <- Shoot_Outs[-c(3:6)]
colnames(Away_Wins) <- colnames(Home_Wins)
FIFA_WC_Champions <- rbind(Home_Wins, Away_Wins, Shoot_Outs)
  FIFA_WC_Champions <- arrange(FIFA_WC_Champions, world_cup) %>% rename("CHAMPION" = home_team, "WORLD C
FIFA_WC_Champions <- FIFA_WC_Champions[,-c(3:5)]
FIFA_WC_Champions
```

```
## # A tibble: 21 x 2
##    `WORLD CUP` CHAMPION
##          <dbl> <chr>
##  1        1930 Uruguay
##  2        1934 Italy
##  3        1938 Italy
##  4        1950 Uruguay
```

```
##   5        1954 Germany
##   6        1958 Brazil
##   7        1962 Brazil
##   8        1966 England
##   9        1970 Brazil
##  10        1974 Germany
## # ... with 11 more rows
## # i Use `print(n = ...)` to see more rows
```

```r
#Create a List of Teams that have won a World Cup
FIFA_WC_Champion_Teams <- FIFA_WC_Champions
FIFA_WC_Champion_Teams <- unique(sort(FIFA_WC_Champion_Teams$`CHAMPION`))
FIFA_WC_Champion_Teams <- as.data.frame(FIFA_WC_Champion_Teams)
FIFA_WC_Champion_Teams %>% rename("CHAMPIONS" = FIFA_WC_Champion_Teams)
```

```
##   CHAMPIONS
## 1 Argentina
## 2    Brazil
## 3   England
## 4    France
## 5   Germany
## 6     Italy
## 7     Spain
## 8   Uruguay
```

2.5 FRIENDLY MATCHES SUBSET

```
Friendly_Matches <- subset(raw_data, raw_data$tournament == "Friendly")
Friendly_Matches %>% standard.clean(6:9)
```

```
## # A tibble: 17,362 x 5
##    `MATCH DATE` `HOME TEAM` `AWAY TEAM` `HOME SCORE` `AWAY SCORE`
##    <date>       <chr>       <chr>              <dbl>        <dbl>
##  1 1872-11-30   Scotland    England                0            0
##  2 1873-03-08   England     Scotland               4            2
##  3 1874-03-07   Scotland    England                2            1
##  4 1875-03-06   England     Scotland               2            2
##  5 1876-03-04   Scotland    England                3            0
##  6 1876-03-25   Scotland    Wales                  4            0
##  7 1877-03-03   England     Scotland               1            3
##  8 1877-03-05   Wales       Scotland               0            2
##  9 1878-03-02   Scotland    England                7            2
## 10 1878-03-23   Scotland    Wales                  9            0
## # ... with 17,352 more rows
## # i Use `print(n = ...)` to see more rows
```

3 DATA ANALYSIS

3.1 ANALYZING FIFA WORLD CUP DATA

3.1.1 ANALYZING THE EIGHT PAST CHAMPIONS

```
FIFA_WC_ARG <- tournament.total.matches(FIFA_WC, "Argentina")
FIFA_WC_BRA <- tournament.total.matches(FIFA_WC, "Brazil")
FIFA_WC_ENG <- tournament.total.matches(FIFA_WC, "England")
FIFA_WC_FRA <- tournament.total.matches(FIFA_WC, "France")
FIFA_WC_GER <- tournament.total.matches(FIFA_WC, "Germany")
FIFA_WC_ITA <- tournament.total.matches(FIFA_WC, "Italy")
FIFA_WC_ESP <- tournament.total.matches(FIFA_WC, "Spain")
FIFA_WC_URU <- tournament.total.matches(FIFA_WC, "Uruguay")

FIFA_WC_ARG_Home_Mean <- home.score.mean(FIFA_WC_ARG, "Argentina")
FIFA_WC_ARG_Away_Mean <- away.score.mean(FIFA_WC_ARG, "Argentina")
FIFA_WC_ARG_Score_Mean <- score.mean(FIFA_WC_ARG, "Argentina")

FIFA_WC_BRA_Home_Mean <- home.score.mean(FIFA_WC_BRA, "Brazil")
FIFA_WC_BRA_Away_Mean <- away.score.mean(FIFA_WC_BRA, "Brazil")
FIFA_WC_BRA_Score_Mean <- score.mean(FIFA_WC_BRA, "Brazil")

FIFA_WC_ENG_Home_Mean <- home.score.mean(FIFA_WC_ENG, "England")
FIFA_WC_ENG_Away_Mean <- away.score.mean(FIFA_WC_ENG, "England")
FIFA_WC_ENG_Score_Mean <- score.mean(FIFA_WC_ENG, "England")

FIFA_WC_FRA_Home_Mean <- home.score.mean(FIFA_WC_FRA, "France")
FIFA_WC_FRA_Away_Mean <- away.score.mean(FIFA_WC_FRA, "France")
FIFA_WC_FRA_Score_Mean <- score.mean(FIFA_WC_FRA, "France")

FIFA_WC_GER_Home_Mean <- home.score.mean(FIFA_WC_GER, "Germany")
FIFA_WC_GER_Away_Mean <- away.score.mean(FIFA_WC_GER, "Germany")
FIFA_WC_GER_Score_Mean <- score.mean(FIFA_WC_GER, "Germany")

FIFA_WC_ITA_Home_Mean <- home.score.mean(FIFA_WC_ITA, "Italy")
FIFA_WC_ITA_Away_Mean <- away.score.mean(FIFA_WC_ITA, "Italy")
FIFA_WC_ITA_Score_Mean <- score.mean(FIFA_WC_ITA, "Italy")

FIFA_WC_ESP_Home_Mean <- home.score.mean(FIFA_WC_ESP, "Spain")
FIFA_WC_ESP_Away_Mean <- away.score.mean(FIFA_WC_ESP, "Spain")
FIFA_WC_ESP_Score_Mean <- score.mean(FIFA_WC_ESP, "Spain")

FIFA_WC_URU_Home_Mean <- home.score.mean(FIFA_WC_URU, "Uruguay")
FIFA_WC_URU_Away_Mean <- away.score.mean(FIFA_WC_URU, "Uruguay")
FIFA_WC_URU_Score_Mean <- score.mean(FIFA_WC_URU, "Uruguay")

FIFA_WC_Home_Means_Raw <- c(FIFA_WC_ARG_Home_Mean, FIFA_WC_BRA_Home_Mean, FIFA_WC_ENG_Home_Mean, FIFA_W
FIFA_WC_Home_Means <- round(FIFA_WC_Home_Means_Raw, digits = 2)

FIFA_WC_Away_Means_Raw <- c(FIFA_WC_ARG_Away_Mean, FIFA_WC_BRA_Away_Mean, FIFA_WC_ENG_Away_Mean, FIFA_W
FIFA_WC_Away_Means <- round(FIFA_WC_Away_Means_Raw, digits = 2)
```

```r
FIFA_WC_Score_Means_Raw <- as.numeric(c(FIFA_WC_ARG_Score_Mean, FIFA_WC_BRA_Score_Mean, FIFA_WC_ENG_Sco
FIFA_WC_Score_Means <- round(FIFA_WC_Score_Means_Raw, digits = 2)

# DF1 contains all the data, DF2 only shows country and its mean score
# DF2 is more relevant because home/away status does not matter as much for the world cup as it may for

FIFA_WC_Past_Champions_DF1 <- data.frame(FIFA_WC_Champion_Teams, FIFA_WC_Home_Means, FIFA_WC_Away_Means
FIFA_WC_Past_Champions_DF1 %>% rename(
  "COUNTRY" = FIFA_WC_Champion_Teams,
  "MEAN HOME SCORE" = FIFA_WC_Home_Means,
  "MEAN AWAY SCORE" = FIFA_WC_Away_Means,
  "MEAN SCORE" = FIFA_WC_Score_Means)
```

```
##     COUNTRY MEAN HOME SCORE MEAN AWAY SCORE MEAN SCORE
## 1 Argentina            1.90            1.10       1.69
## 2    Brazil            2.11            2.08       2.10
## 3   England            1.39            1.24       1.32
## 4    France            2.12            1.35       1.82
## 5   Germany            2.05            2.12       2.07
## 6     Italy            1.47            1.65       1.54
## 7     Spain            1.63            1.52       1.57
## 8   Uruguay            1.63            1.51       1.55
```

```r
FIFA_WC_Past_Champions_DF2 <- select(FIFA_WC_Past_Champions_DF1, FIFA_WC_Champion_Teams, FIFA_WC_Score_M
  "COUNTRY" = FIFA_WC_Champion_Teams,
  "MEAN SCORE" = FIFA_WC_Score_Means)
FIFA_WC_Past_Champions_DF2
```

```
##     COUNTRY MEAN SCORE
## 1 Argentina       1.69
## 2    Brazil       2.10
## 3   England       1.32
## 4    France       1.82
## 5   Germany       2.07
## 6     Italy       1.54
## 7     Spain       1.57
## 8   Uruguay       1.55
```

3.1.2 ANALYZING ALL WORLD CUP PARTICIPANTS

```
FIFA_Participating_Teams <- participating.teams(FIFA_WC)
FIFA_WC_Teams <- data.frame(FIFA_Participating_Teams)
head(FIFA_WC_Teams) %>% rename( "ALL WORLD CUP PARTICIPANTS" = FIFA_Participating_Teams) %>% simple.cle
```

```
##   ALL WORLD CUP PARTICIPANTS
## 1                    Algeria
## 2                     Angola
## 3                  Argentina
## 4                  Australia
## 5                    Austria
## 6                    Belgium
```

```
FIFA_WC_Teams_List <- as.matrix(FIFA_WC_Teams)
FIFA_WC_ALL <- apply(X=FIFA_WC_Teams, MARGIN = 1, FUN = world.cup.total.matches.cleaned)

home.score.mean.version2 <- function(x) {
X1 <- subset(FIFA_WC_ALL[[x]], FIFA_WC_ALL[[x]]$`HOME TEAM` == FIFA_WC_Teams_List[[x]])
Y1 <- subset(FIFA_WC_ALL[[x]], FIFA_WC_ALL[[x]]$`AWAY TEAM` == FIFA_WC_Teams_List[[x]])
SUM_X <- sum(X1$`HOME SCORE`)
SUM_Y <- sum(Y1$`AWAY SCORE`)
N_X <- count(X1)
N_Y <- count(Y1)
HS <- sum(SUM_X, SUM_Y) / sum(N_X, N_Y)
return(HS)} #custom function to find mean scores of all 81 teams

FIFA_WC_ALL_SCORE_MEANS_UNROUNDED <- mapply(home.score.mean.version2, 1:81) #custom function applied
FIFA_WC_ALL_SCORE_MEANS <- round(FIFA_WC_ALL_SCORE_MEANS_UNROUNDED, digits = 2)

FIFA_WC_ALL_TEAMS_SCORE_MEANS <- data.frame(FIFA_WC_Teams_List, FIFA_WC_ALL_SCORE_MEANS)
head(FIFA_WC_ALL_TEAMS_SCORE_MEANS) %>% rename(
  "TEAM" = FIFA_Participating_Teams, "MEAN SCORE" = FIFA_WC_ALL_SCORE_MEANS)
```

```
##        TEAM MEAN SCORE
## 1   Algeria       1.00
## 2    Angola       0.33
## 3 Argentina       1.69
## 4 Australia       0.81
## 5   Austria       1.48
## 6   Belgium       1.42
```

3.1.3 TEAM STATISTICS

```
FIFA_WC2 <- FIFA_WC
FIFA_WC2 <- FIFA_WC2 %>% mutate(Draw_1 = if_else(Match_Winner == "DRAW", home_team, "NOT DRAW")) %>% mu

FIFA_Draw1 <- as.data.frame(table(FIFA_WC2$home_team))
FIFA_Draw2 <- as.data.frame(table(FIFA_WC2$away_team))

FIFA_Winner <- as.data.frame(table(FIFA_WC$Match_Winner))
FIFA_Winner <- FIFA_Winner[-c(18), ]

FIFA_Loser <- as.data.frame(table(FIFA_WC$Match_Loser))
FIFA_Loser <- FIFA_Loser[-c(23), ]

FIFA_Draw <- merge(FIFA_Draw1, FIFA_Draw2, by = "Var1", all.x = TRUE, all.y = TRUE)
FIFA_Wins_Losses <- merge(FIFA_Winner, FIFA_Loser, by = "Var1", all.x = TRUE, all.y = TRUE)

FIFA_Points <- merge(FIFA_Wins_Losses, FIFA_Draw, by = "Var1", all.x = TRUE, all.y = TRUE)
FIFA_Points <- FIFA_Points %>% mutate_if(is.integer, ~replace(., is.na(.), 0)) %>% mutate(D = Freq.x.y

colnames(FIFA_Points) <- c("Team", "W","L","Dx", "Dy", "D")
FIFA_Points <- FIFA_Points[, -c(4:5)] %>% mutate(Pts = 3*W + 1*D) %>% arrange(desc(Pts))
FIFA_Points$Team <- as.character(FIFA_Points$Team)
head(FIFA_Points)
```

3.1.3.1 COMPUTING WINS, LOSSES, DRAWS, AND POINTS

```
##        Team  W  L  D Pts
## 1    Brazil 73 18 18 237
## 2   Germany 67 22 20 221
## 3     Italy 45 17 21 156
## 4 Argentina 43 23 15 144
## 5    France 34 19 13 115
## 6   England 29 19 21 108
```

```
FIFA_WC_GD <- FIFA_WC %>% mutate(HGD = home_score - away_score) %>% mutate(AGD = away_score - home_scor
FIFA_WC_GD$HGD <- FIFA_WC_GD$home_score - FIFA_WC_GD$away_score
FIFA_WC_GD$AGD <- FIFA_WC_GD$away_score - FIFA_WC_GD$home_score
FIFA_WC_GD <- FIFA_WC_GD[, -c(1,6:9)]

FIFA_WC_GD1 <- apply(X=FIFA_WC_Teams, MARGIN = 1, FUN = world.cup.total.matches.v2)

Gd.1 <- function(x) {
X1 <- subset(FIFA_WC_GD1[[x]], FIFA_WC_GD1[[x]]$home_team == FIFA_WC_Teams_List[[x]])
Y1 <- subset(FIFA_WC_GD1[[x]], FIFA_WC_GD1[[x]]$away_team == FIFA_WC_Teams_List[[x]])
SUM_X <- sum(X1$HGD)
SUM_Y <- sum(Y1$AGD)
GD <- sum(SUM_X, SUM_Y)
return(GD)} #custom function to find Goal Difference for all 81 teams
```

```r
Gd.2 <- function(x) {
X1 <- subset(FIFA_WC_GD1[[x]], FIFA_WC_GD1[[x]]$home_team == FIFA_WC_Teams_List[[x]])
Y1 <- subset(FIFA_WC_GD1[[x]], FIFA_WC_GD1[[x]]$away_team == FIFA_WC_Teams_List[[x]])
SUM_X <- sum(X1$HGD)
SUM_Y <- sum(Y1$AGD)
N_X <- count(X1)
N_Y <- count(Y1)
A_GD <- sum(SUM_X, SUM_Y) / sum(N_X, N_Y)
return(A_GD)} #custom function to find Average Goal Difference for all 81 teams

FIFA_WC_GD2 <- mapply(Gd.1, 1:81)
FIFA_WC_GD2 <- as.data.frame(FIFA_WC_GD2)
FIFA_WC_GD3 <- data.frame(FIFA_WC_Teams_List, FIFA_WC_GD2)
FIFA_WC_GD3 <- FIFA_WC_GD3 %>% rename("Team" = FIFA_Participating_Teams, "GD" = FIFA_WC_GD2) %>% arrang

FIFA_WC_A_GD <- mapply(Gd.2, 1:81)
FIFA_WC_A_GD <- as.data.frame(FIFA_WC_A_GD)
FIFA_WC_A_GD <- round(FIFA_WC_A_GD, digits = 2)
FIFA_WC_A_GD2 <- data.frame(FIFA_WC_Teams_List, FIFA_WC_A_GD)
FIFA_WC_A_GD2 <- FIFA_WC_A_GD2 %>% rename("Team" = FIFA_Participating_Teams, "Average GD" = FIFA_WC_A_GI

FIFA_WC_GD_AGD <- merge(FIFA_WC_GD3, FIFA_WC_A_GD2, by = "Team")
head(FIFA_WC_GD_AGD)
```

3.1.3.2 COMPUTING GOAL DIFFERENCE

```
##        Team  GD Average GD
## 1   Algeria  -6      -0.46
## 2    Angola  -1      -0.33
## 3 Argentina  44       0.54
## 4 Australia -18      -1.12
## 5   Austria  -4      -0.14
## 6   Belgium  -4      -0.08
```

```r
FIFA_STATS <- merge(FIFA_Points, FIFA_WC_GD_AGD, by = "Team")
head(FIFA_STATS)
```

```
##        Team  W  L  D Pts  GD Average GD
## 1   Algeria  3  7  3  12  -6      -0.46
## 2    Angola  0  1  2   2  -1      -0.33
## 3 Argentina 43 23 15 144  44       0.54
## 4 Australia  2 10  4  10 -18      -1.12
## 5   Austria 12 13  4  40  -4      -0.14
## 6   Belgium 20 19  9  69  -4      -0.08
```

3.1.4 REGRESSION MODELS

```
ols1 <- lm(`Average GD` ~ Pts, FIFA_STATS)
summary(ols1)
```

3.1.4.1 RELATIONSHIP BETWEEN AVERAGE GD AND POINTS

```
##
## Call:
## lm(formula = `Average GD` ~ Pts, data = FIFA_STATS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7855 -0.3818  0.2353  0.6812  1.4374
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.214503   0.146126  -8.311 2.15e-12 ***
## Pts          0.014940   0.002641   5.658 2.36e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.091 on 79 degrees of freedom
## Multiple R-squared:  0.2884, Adjusted R-squared:  0.2794
## F-statistic: 32.01 on 1 and 79 DF,  p-value: 2.364e-07
```

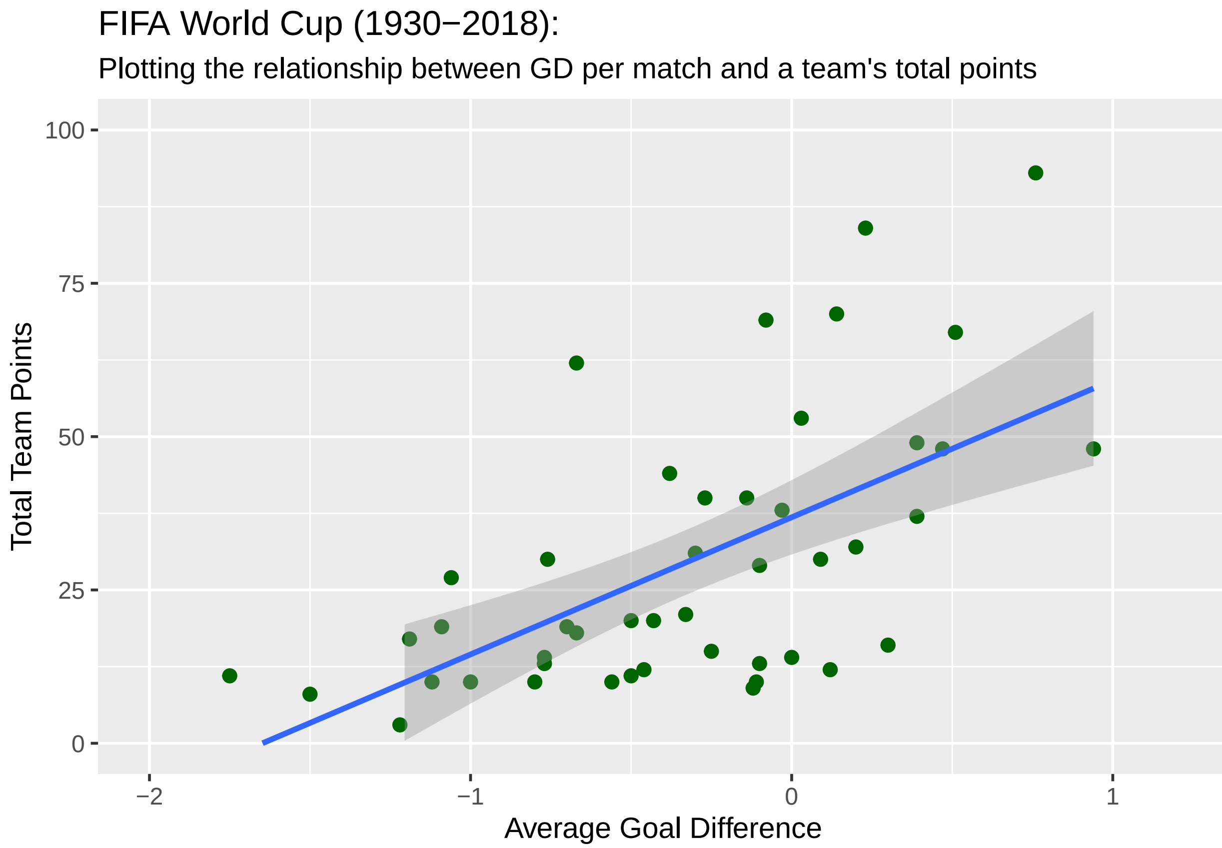
```
#REMOVED TEAMS THAT PLAYED 7 OR LESS GAMES TO REDUCE OUTLIERS
#THIS WAS EXPLAINED IN THE WRITE-UP
FIFA_STATS1 <- FIFA_STATS %>% subset(W+L+D > 7)
ggplot(data=FIFA_STATS1, aes(x=`Average GD`, y=Pts)) + geom_point(colour = "dark green", size = 2) + yli
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 7 rows containing non-finite values (stat_smooth).

## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 3 rows containing missing values (geom_smooth).
```


FIFA World Cup (1930−2018):
Plotting the relationship between GD per match and a team's total points
Total Team Points
100
75
50
25
0
−2
−1
0
1
Average Goal Difference

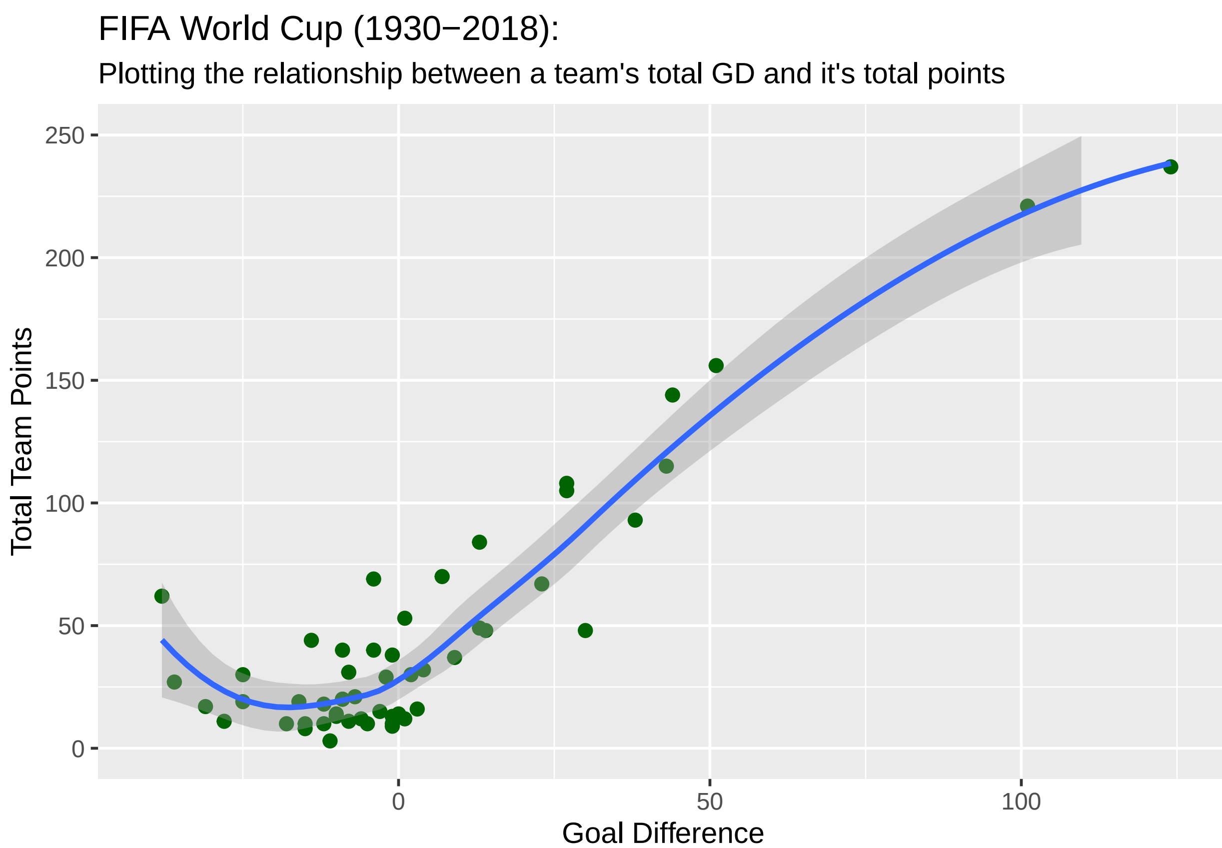
```r
ols2 <- lm(GD ~ Pts, FIFA_STATS)
summary(ols2)
```

3.1.4.2 RELATIONSHIP BETWEEN GD AND POINTS

```
##
## Call:
## lm(formula = GD ~ Pts, data = FIFA_STATS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -52.060  -6.183   2.594   8.787  30.881
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.94888    1.66679  -8.369 1.66e-12 ***
## Pts           0.45176    0.03012  14.999  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.45 on 79 degrees of freedom
## Multiple R-squared:  0.7401, Adjusted R-squared:  0.7368
## F-statistic:   225 on 1 and 79 DF,  p-value: < 2.2e-16
```

```r
#REMOVED TEAMS THAT HAD 7 OR LESS POINTS TO REDUCE OUTLIERS
#THIS WAS EXPLAINED IN THE WRITE-UP
ggplot(data=FIFA_STATS1, aes(x=`GD`, y=`Pts`)) + geom_point(colour = "dark green", size = 2) + geom_smoo
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```


FIFA World Cup (1930−2018):
Plotting the relationship between a team's total GD and it's total points
Total Team Points
250
200
150
100
50
0
0
50
100
Goal Difference

3.2 ANALYZING FRIENDLY MATCH DATA

3.2.1 MEAN SCORES OF EVERY TEAM

```
FM_Teams <- as.matrix(participating.teams(Friendly_Matches))
FM_ALL <- apply(X=FM_Teams, MARGIN = 1, FUN = friendly.total.matches.cleaned)

score.mean.version3 <- function(x) {
X1 <- subset(FM_ALL[[x]], FM_ALL[[x]]$`HOME TEAM` == FM_Teams[[x]])
Y1 <- subset(FM_ALL[[x]], FM_ALL[[x]]$`AWAY TEAM` == FM_Teams[[x]])
SUM_X <- sum(X1$`HOME SCORE`)
SUM_Y <- sum(Y1$`AWAY SCORE`)
N_X <- count(X1)
N_Y <- count(Y1)
HS3 <- sum(SUM_X, SUM_Y) / sum(N_X, N_Y)
return(HS3)} #custom function to find mean scores for all 264 teams

FM_ALL_SCORE_MEANS_UNROUNDED <- mapply(score.mean.version3, 1:264)
FM_ALL_SCORE_MEANS <- round(FM_ALL_SCORE_MEANS_UNROUNDED, digits = 2)

FM_ALL_MEAN_SCORES <- data.frame(FM_Teams, FM_ALL_SCORE_MEANS)
head(FM_ALL_MEAN_SCORES) %>% rename(
  "TEAM" = FM_Teams, "MEAN SCORE" = FM_ALL_SCORE_MEANS)
```

```
##               TEAM MEAN SCORE
## 1         Abkhazia       0.50
## 2      Afghanistan       0.90
## 3          Albania       1.21
## 4          Algeria       1.29
## 5   American Samoa       0.00
## 6        Andalusia       1.92
```

3.2.2 HOW MUCH MORE DO TEAMS TEND TO SCORE AS HOME TEAMS?

```
home.score.mean.fm.version <- function(x) {
  X1 <- subset(FM_ALL[[x]], FM_ALL[[x]]$`HOME TEAM` == FM_Teams[[x]])
  Y1 <- mean(X1$`HOME SCORE`)
  return(Y1)} #custom function to find mean score of every team when it played at home

FM_ALL_HOME_SCORE_MEANS_UNROUNDED <- mapply(home.score.mean.fm.version, 1:264)
FM_ALL_HOME_SCORE_MEANS <- round(FM_ALL_HOME_SCORE_MEANS_UNROUNDED, digits = 2)

away.score.mean.fm.version <- function(x) {
  X1 <- subset(FM_ALL[[x]], FM_ALL[[x]]$`AWAY TEAM` == FM_Teams[[x]])
  Y1 <- mean(X1$`AWAY SCORE`)
  return(Y1)} #custom function to find mean score of every team when it played away

FM_ALL_AWAY_SCORE_MEANS_UNROUNDED <- mapply(away.score.mean.fm.version, 1:264)
FM_ALL_AWAY_SCORE_MEANS <- round(FM_ALL_AWAY_SCORE_MEANS_UNROUNDED, digits = 2)

FM_ALL_HOME_AWAY_SCORES_RAW <- data.frame(FM_Teams, FM_ALL_HOME_SCORE_MEANS, FM_ALL_AWAY_SCORE_MEANS)
```

```r
#Deleting rows with missing values
FM_ALL_HOME_AWAY_SCORES_CLEANED <- na.omit(FM_ALL_HOME_AWAY_SCORES_RAW)
head(FM_ALL_HOME_AWAY_SCORES_CLEANED) %>% rename(
  "Teams" = FM_Teams,
  "Home Score (Mean)" = FM_ALL_HOME_SCORE_MEANS,
  "Away Score (Mean)" = FM_ALL_AWAY_SCORE_MEANS
) %>% simple.clean()
```

```
##         TEAMS HOME SCORE (MEAN) AWAY SCORE (MEAN)
## 1    Abkhazia              1.00              0.00
## 2 Afghanistan              1.10              0.80
## 3     Albania              1.39              0.96
## 4     Algeria              1.58              0.80
## 6   Andalusia              1.92              2.00
## 7     Andorra              0.46              0.24
```

```r
FM_AVERAGE_HOME_SCORE <- mean(FM_ALL_HOME_AWAY_SCORES_CLEANED$FM_ALL_HOME_SCORE_MEANS)
FM_AVERAGE_AWAY_SCORE <- mean(FM_ALL_HOME_AWAY_SCORES_CLEANED$FM_ALL_AWAY_SCORE_MEANS)

if(FM_AVERAGE_HOME_SCORE > FM_AVERAGE_AWAY_SCORE) {
  cat("There is a possible home advantage\n")
  FM_Home_Field_Advantage <- round(((FM_AVERAGE_AWAY_SCORE/FM_AVERAGE_HOME_SCORE) * 100), digits = 2)
  cat("In friendly matches, teams have historically scored", FM_Home_Field_Advantage,"% higher at home")
```

```
## There is a possible home advantage
## In friendly matches, teams have historically scored 67.28 % higher at home
```

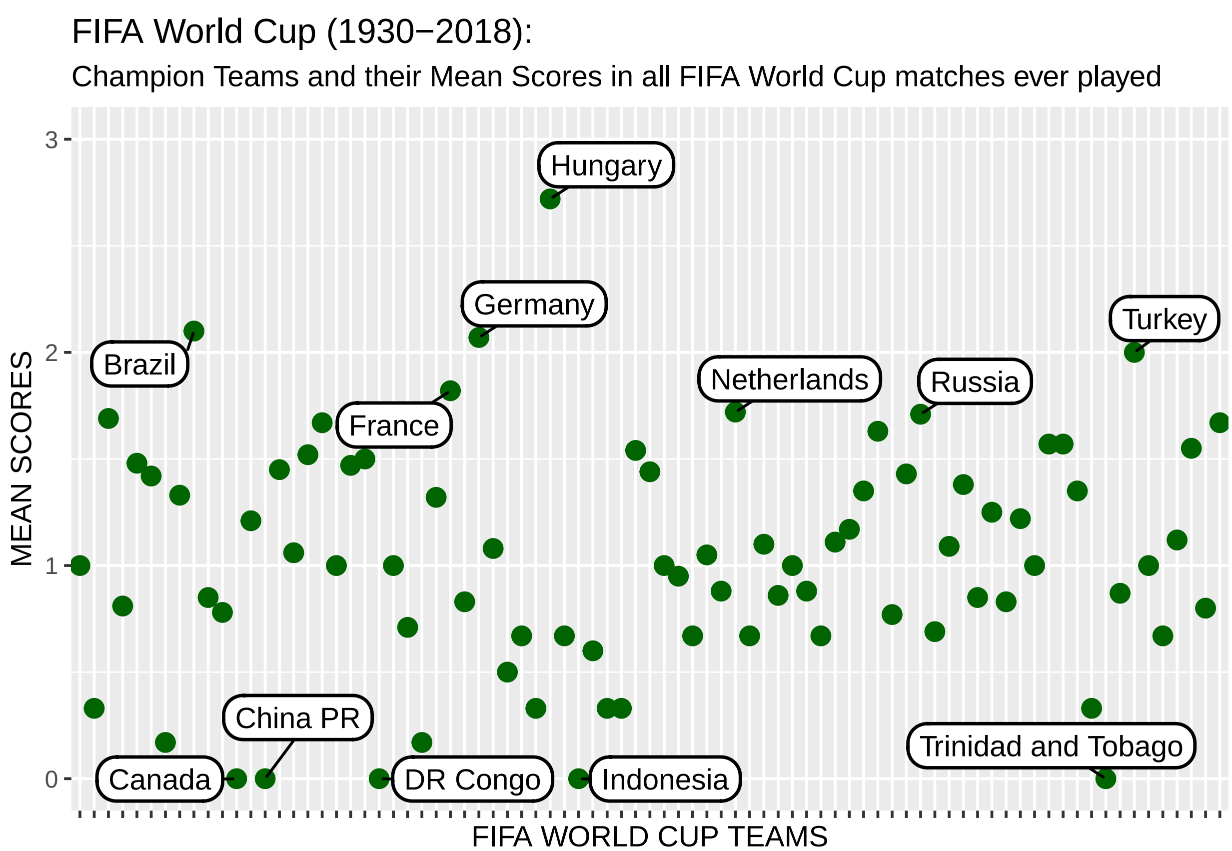
4 DATA VISUALIZATION

4.1 FIFA WORLD CUP SUBSET

4.1.1 MEAN SCORES OF EVERY TEAM

```
ggplot(data=FIFA_WC_ALL_TEAMS_SCORE_MEANS, aes(x=FIFA_Participating_Teams, y=FIFA_WC_ALL_SCORE_MEANS)) +
```

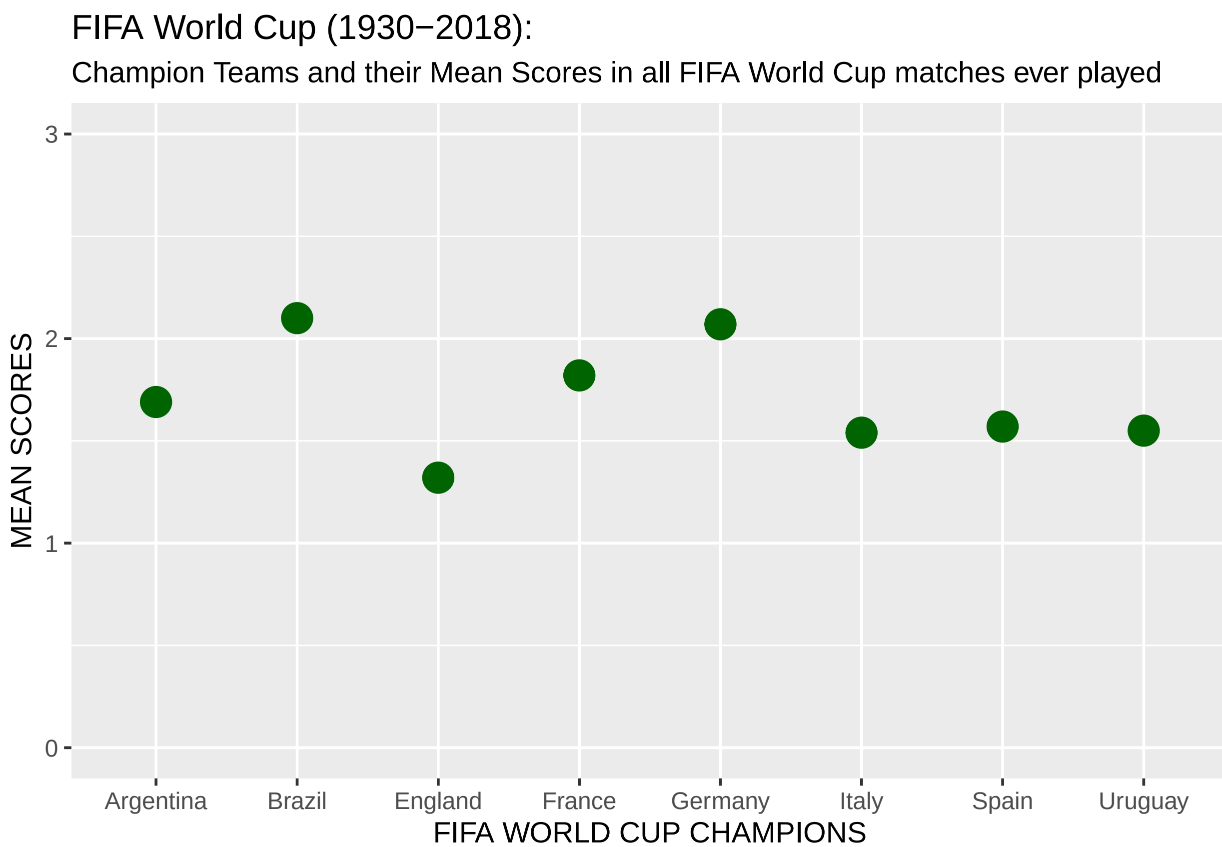
```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```



4.1.2 MEAN SCORES OF CHAMPION TEAMS

```
ggplot(data=FIFA_WC_Past_Champions_DF2, aes(x=`COUNTRY`, y=`MEAN SCORE`)) + geom_line() + geom_point(co
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```

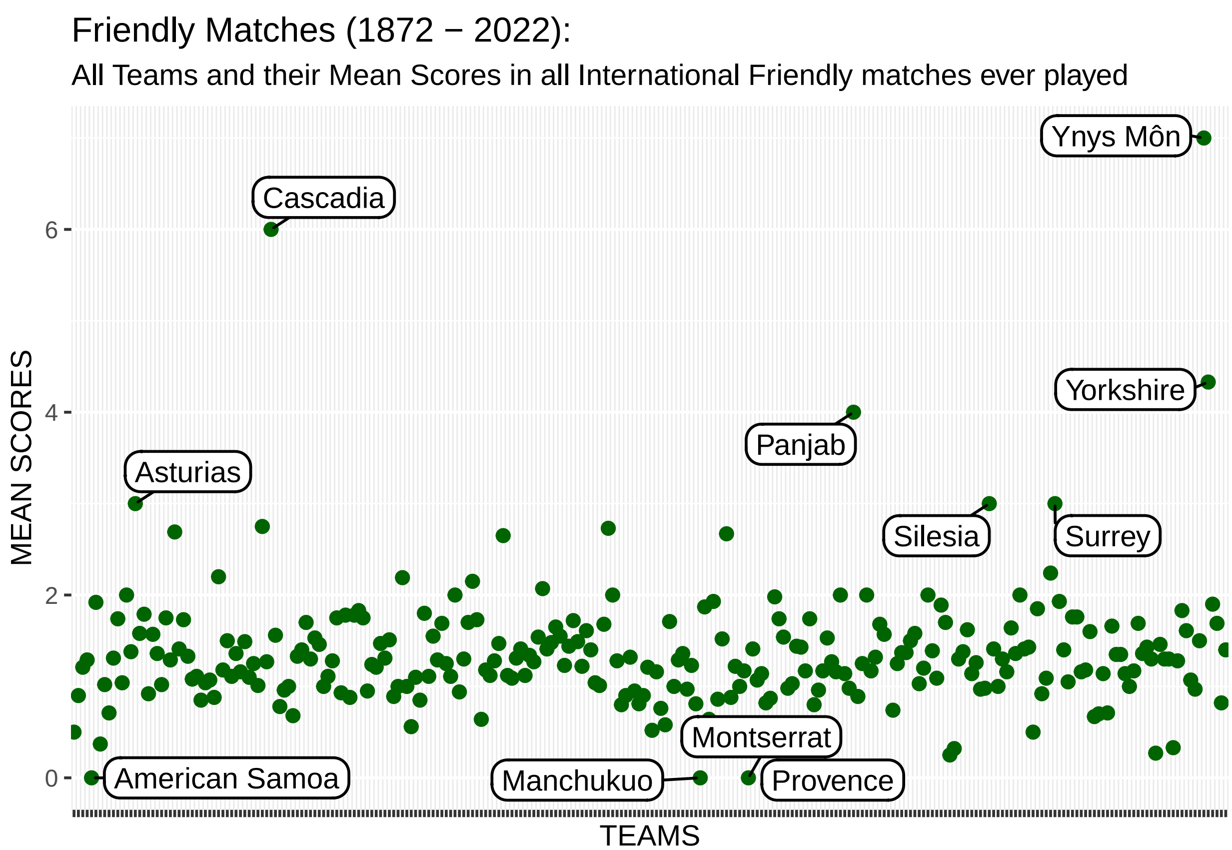


4.2 FRIENDLY MATCHES SUBSET

4.2.1 MEAN SCORES OF EVERY TEAM

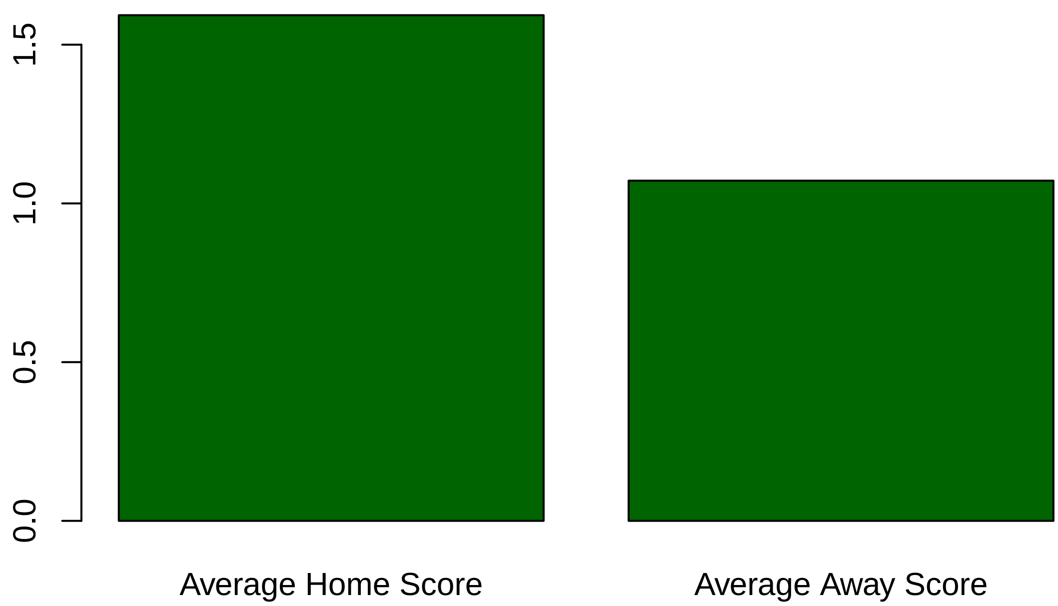
```
ggplot(data=FM_ALL_MEAN_SCORES, aes(x=FM_Teams, y=FM_ALL_SCORE_MEANS)) + geom_point(colour = "dark green
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



4.2.2 AVERAGE HOME SCORE AND AVERAGE AWAY SCORE

```
means_vector <- c(FM_AVERAGE_HOME_SCORE, FM_AVERAGE_AWAY_SCORE)
names_vector <- c("Average Home Score", "Average Away Score")
barplot(means_vector, col = "dark green", names.arg = names_vector)
```



5 HYPOTHESIS TESTING

5.1 FIFA WORLD CUP DATA

5.1.1 NULL HYPOTHESIS: NO RELATIONSHIP BETWEEN AVERAGE GD & POINTS

```
Test1 <- chisq.test(FIFA_STATS$GD, FIFA_STATS$Pts, correct = FALSE, simulate.p.value = TRUE)
Test1
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  FIFA_STATS$GD and FIFA_STATS$Pts
## X-squared = 2133, df = NA, p-value = 0.0004998
```

Null Hypothesis has been rejected.

5.1.2 NULL HYPOTHESIS: NO RELATIONSHIP BETWEEN GD & POINTS

```
Test2 <- chisq.test(FIFA_STATS$`Average GD`, FIFA_STATS$Pts, correct = FALSE, simulate.p.value = TRUE)
Test2
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  FIFA_STATS$`Average GD` and FIFA_STATS$Pts
## X-squared = 2901.9, df = NA, p-value = 0.007996
```

Null Hypothesis has been rejected.

5.2 FRIENDLY MATCHES DATA

5.2.1 NULL HYPOTHESIS: HOME AND AWAY SCORES TEND TO BE SIMILAR

```
Test3 <- t.test(FM_ALL_HOME_AWAY_SCORES_CLEANED$FM_ALL_HOME_SCORE_MEANS,  FM_ALL_HOME_AWAY_SCORES_CLEAN
Test3
```

```
##
##  Paired t-test
##
## data:  FM_ALL_HOME_AWAY_SCORES_CLEANED$FM_ALL_HOME_SCORE_MEANS and FM_ALL_HOME_AWAY_SCORES_CLEANED$FM
## t = 12.312, df = 243, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.4378754 0.6046656
## sample estimates:
## mean difference
##       0.5212705
```

Null Hypothesis has been rejected.

6 APPENDIX

6.1 ALL INTERNATIONAL TEAMS

```
International_Teams %>% simple.clean() %>% tbl_df %>% print(n=310)
```

```
## Warning: `tbl_df()` was deprecated in dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
## # A tibble: 310 x 1
##     `INTERNATIONAL TEAMS`
##     <chr>
##   1 Abkhazia
##   2 Afghanistan
##   3 Åland Islands
##   4 Albania
##   5 Alderney
##   6 Algeria
##   7 American Samoa
##   8 Andalusia
##   9 Andorra
##  10 Angola
##  11 Anguilla
##  12 Antigua and Barbuda
##  13 Arameans Suryoye
##  14 Argentina
##  15 Armenia
##  16 Artsakh
##  17 Aruba
##  18 Australia
##  19 Austria
##  20 Azerbaijan
##  21 Bahamas
##  22 Bahrain
##  23 Bangladesh
##  24 Barawa
##  25 Barbados
##  26 Basque Country
##  27 Belarus
##  28 Belgium
##  29 Belize
##  30 Benin
##  31 Bermuda
##  32 Bhutan
##  33 Bolivia
##  34 Bonaire
##  35 Bosnia and Herzegovina
##  36 Botswana
##  37 Brazil
##  38 British Virgin Islands
```

```
##  39 Brittany
##  40 Brunei
##  41 Bulgaria
##  42 Burkina Faso
##  43 Burundi
##  44 Cambodia
##  45 Cameroon
##  46 Canada
##  47 Cape Verde
##  48 Cascadia
##  49 Catalonia
##  50 Cayman Islands
##  51 Central African Republic
##  52 Central Spain
##  53 Chad
##  54 Chagos Islands
##  55 Chameria
##  56 Chile
##  57 China PR
##  58 Colombia
##  59 Comoros
##  60 Congo
##  61 Cook Islands
##  62 Corsica
##  63 Costa Rica
##  64 County of Nice
##  65 Croatia
##  66 Cuba
##  67 Curaçao
##  68 Cyprus
##  69 Czech Republic
##  70 Czechoslovakia
##  71 Darfur
##  72 Denmark
##  73 Djibouti
##  74 Dominica
##  75 Dominican Republic
##  76 DR Congo
##  77 Ecuador
##  78 Egypt
##  79 El Salvador
##  80 Ellan Vannin
##  81 England
##  82 Equatorial Guinea
##  83 Eritrea
##  84 Estonia
##  85 Eswatini
##  86 Ethiopia
##  87 Falkland Islands
##  88 Faroe Islands
##  89 Felvidék
##  90 Fiji
##  91 Finland
##  92 France
```

```
##  93 French Guiana
##  94 Frøya
##  95 Gabon
##  96 Galicia
##  97 Gambia
##  98 Georgia
##  99 German DR
## 100 Germany
## 101 Ghana
## 102 Gibraltar
## 103 Gotland
## 104 Gozo
## 105 Greece
## 106 Greenland
## 107 Grenada
## 108 Guadeloupe
## 109 Guam
## 110 Guatemala
## 111 Guernsey
## 112 Guinea
## 113 Guinea-Bissau
## 114 Guyana
## 115 Haiti
## 116 Hitra
## 117 Honduras
## 118 Hong Kong
## 119 Hungary
## 120 Iceland
## 121 India
## 122 Indonesia
## 123 Iran
## 124 Iraq
## 125 Iraqi Kurdistan
## 126 Isle of Man
## 127 Isle of Wight
## 128 Israel
## 129 Italy
## 130 Ivory Coast
## 131 Jamaica
## 132 Japan
## 133 Jersey
## 134 Jordan
## 135 Kabylia
## 136 Kárpátalja
## 137 Kazakhstan
## 138 Kenya
## 139 Kernow
## 140 Kiribati
## 141 Kosovo
## 142 Kuwait
## 143 Kyrgyzstan
## 144 Laos
## 145 Latvia
## 146 Lebanon
```

```
## 147 Lesotho
## 148 Liberia
## 149 Libya
## 150 Liechtenstein
## 151 Lithuania
## 152 Luxembourg
## 153 Macau
## 154 Madagascar
## 155 Malawi
## 156 Malaysia
## 157 Maldives
## 158 Mali
## 159 Malta
## 160 Manchukuo
## 161 Martinique
## 162 Matabeleland
## 163 Mauritania
## 164 Mauritius
## 165 Mayotte
## 166 Menorca
## 167 Mexico
## 168 Micronesia
## 169 Moldova
## 170 Monaco
## 171 Mongolia
## 172 Montenegro
## 173 Montserrat
## 174 Morocco
## 175 Mozambique
## 176 Myanmar
## 177 Namibia
## 178 Nepal
## 179 Netherlands
## 180 New Caledonia
## 181 New Zealand
## 182 Nicaragua
## 183 Niger
## 184 Nigeria
## 185 North Korea
## 186 North Macedonia
## 187 North Vietnam
## 188 Northern Cyprus
## 189 Northern Ireland
## 190 Northern Mariana Islands
## 191 Norway
## 192 Occitania
## 193 Oman
## 194 Orkney
## 195 Padania
## 196 Pakistan
## 197 Palestine
## 198 Panama
## 199 Panjab
## 200 Papua New Guinea
```

```
## 201 Paraguay
## 202 Parishes of Jersey
## 203 Peru
## 204 Philippines
## 205 Poland
## 206 Portugal
## 207 Provence
## 208 Puerto Rico
## 209 Qatar
## 210 Raetia
## 211 Republic of Ireland
## 212 Réunion
## 213 Rhodes
## 214 Romani people
## 215 Romania
## 216 Russia
## 217 Rwanda
## 218 Saare County
## 219 Saarland
## 220 Saint Helena
## 221 Saint Kitts and Nevis
## 222 Saint Lucia
## 223 Saint Martin
## 224 Saint Pierre and Miquelon
## 225 Saint Vincent and the Grenadines
## 226 Samoa
## 227 San Marino
## 228 São Tomé and Príncipe
## 229 Sápmi
## 230 Sark
## 231 Saudi Arabia
## 232 Scotland
## 233 Senegal
## 234 Serbia
## 235 Seychelles
## 236 Shetland
## 237 Sierra Leone
## 238 Singapore
## 239 Sint Maarten
## 240 Slovakia
## 241 Slovenia
## 242 Solomon Islands
## 243 Somalia
## 244 Somaliland
## 245 South Africa
## 246 South Korea
## 247 South Ossetia
## 248 South Sudan
## 249 Spain
## 250 Sri Lanka
## 251 Sudan
## 252 Suriname
## 253 Sweden
## 254 Switzerland
```

```
## 255 Syria
## 256 Székely Land
## 257 Tahiti
## 258 Taiwan
## 259 Tajikistan
## 260 Tamil Eelam
## 261 Tanzania
## 262 Thailand
## 263 Tibet
## 264 Timor-Leste
## 265 Togo
## 266 Tonga
## 267 Trinidad and Tobago
## 268 Tunisia
## 269 Turkey
## 270 Turkmenistan
## 271 Turks and Caicos Islands
## 272 Tuvalu
## 273 Uganda
## 274 Ukraine
## 275 United Arab Emirates
## 276 United Koreans in Japan
## 277 United States
## 278 United States Virgin Islands
## 279 Uruguay
## 280 Uzbekistan
## 281 Vanuatu
## 282 Vatican City
## 283 Venezuela
## 284 Vietnam
## 285 Vietnam Republic
## 286 Wales
## 287 Wallis Islands and Futuna
## 288 Western Armenia
## 289 Western Australia
## 290 Western Isles
## 291 Western Sahara
## 292 Yemen
## 293 Yemen DPR
## 294 Ynys Môn
## 295 Yorkshire
## 296 Yugoslavia
## 297 Zambia
## 298 Zanzibar
## 299 Zimbabwe
## 300 Canary Islands
## 301 Găgăuzia
## 302 Madrid
## 303 Niue
## 304 Palau
## 305 Republic of St. Pauli
## 306 Silesia
## 307 Asturias
## 308 Crimea
```

```
## 309 Surrey
## 310 Two Sicilies
```

6.2 WORLD CUP PARTICIPANTS

```r
FIFA_WC_Teams %>% rename("TEAM" = FIFA_Participating_Teams) %>% tbl_df %>% print(n=81)
```

```
## # A tibble: 81 x 1
##    TEAM
##    <chr>
##  1 Algeria
##  2 Angola
##  3 Argentina
##  4 Australia
##  5 Austria
##  6 Belgium
##  7 Bolivia
##  8 Bosnia and Herzegovina
##  9 Brazil
## 10 Bulgaria
## 11 Cameroon
## 12 Canada
## 13 Chile
## 14 China PR
## 15 Colombia
## 16 Costa Rica
## 17 Croatia
## 18 Cuba
## 19 Czech Republic
## 20 Czechoslovakia
## 21 Denmark
## 22 DR Congo
## 23 Ecuador
## 24 Egypt
## 25 El Salvador
## 26 England
## 27 France
## 28 German DR
## 29 Germany
## 30 Ghana
## 31 Greece
## 32 Haiti
## 33 Honduras
## 34 Hungary
## 35 Iceland
## 36 Indonesia
## 37 Iran
## 38 Iraq
## 39 Israel
## 40 Italy
## 41 Ivory Coast
## 42 Jamaica
## 43 Japan
## 44 Kuwait
## 45 Mexico
## 46 Morocco
```

```
## 47 Netherlands
## 48 New Zealand
## 49 Nigeria
## 50 North Korea
## 51 Northern Ireland
## 52 Norway
## 53 Panama
## 54 Paraguay
## 55 Peru
## 56 Poland
## 57 Portugal
## 58 Republic of Ireland
## 59 Romania
## 60 Russia
## 61 Saudi Arabia
## 62 Scotland
## 63 Senegal
## 64 Serbia
## 65 Slovakia
## 66 Slovenia
## 67 South Africa
## 68 South Korea
## 69 Spain
## 70 Sweden
## 71 Switzerland
## 72 Togo
## 73 Trinidad and Tobago
## 74 Tunisia
## 75 Turkey
## 76 Ukraine
## 77 United Arab Emirates
## 78 United States
## 79 Uruguay
## 80 Wales
## 81 Yugoslavia
```

6.3 WORLD CUP WINNERS

```
FIFA_WC_Champions %>% tbl_df %>% print(n=21)
```

```
## # A tibble: 21 x 2
##    `WORLD CUP` CHAMPION
##          <dbl> <chr>
##  1        1930 Uruguay
##  2        1934 Italy
##  3        1938 Italy
##  4        1950 Uruguay
##  5        1954 Germany
##  6        1958 Brazil
##  7        1962 Brazil
##  8        1966 England
##  9        1970 Brazil
## 10        1974 Germany
## 11        1978 Argentina
## 12        1982 Italy
## 13        1986 Argentina
## 14        1990 Germany
## 15        1994 Brazil
## 16        1998 France
## 17        2002 Brazil
## 18        2006 Italy
## 19        2010 Spain
## 20        2014 Germany
## 21        2018 France
```

6.4 WORLD CUP TEAMS STATS

```
FIFA_STATS[-c(7)] %>% arrange(-Pts, -GD) %>% relocate(GD, .before = Pts) %>% tbl_df %>% print(n=81)
```

```
## # A tibble: 81 x 6
##    Team                      W     L     D    GD   Pts
##    <chr>                 <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 Brazil                   73    18    18   124   237
##  2 Germany                  67    22    20   101   221
##  3 Italy                    45    17    21    51   156
##  4 Argentina                43    23    15    44   144
##  5 France                   34    19    13    43   115
##  6 England                  29    19    21    27   108
##  7 Spain                    30    18    15    27   105
##  8 Netherlands              27    11    12    38    93
##  9 Uruguay                  24    20    12    13    84
## 10 Sweden                   19    19    13     7    70
## 11 Belgium                  20    19     9    -4    69
## 12 Russia                   19    16    10    23    67
## 13 Mexico                   16    27    14   -38    62
## 14 Poland                   16    13     5     1    53
## 15 Yugoslavia               14    12     7    13    49
## 16 Hungary                  15    14     3    30    48
## 17 Portugal                 14    10     6    14    48
## 18 Switzerland              12    17     8   -14    44
## 19 Austria                  12    13     4    -4    40
## 20 Chile                    11    15     7    -9    40
## 21 Czechoslovakia           11    14     5    -1    38
## 22 Croatia                  11     8     4     9    37
## 23 Denmark                   9     6     5     4    32
## 24 Paraguay                  7    10    10    -8    31
## 25 Colombia                  9    10     3     2    30
## 26 United States             8    19     6   -25    30
## 27 Romania                   8     8     5    -2    29
## 28 South Korea               6    19     9   -36    27
## 29 Nigeria                   6    12     3    -7    21
## 30 Costa Rica                5     8     5    -9    20
## 31 Japan                     5    11     5    -9    20
## 32 Scotland                  4    12     7   -16    19
## 33 Cameroon                  4    12     7   -25    19
## 34 Peru                      5    10     3   -12    18
## 35 Bulgaria                  3    15     8   -31    17
## 36 Turkey                    5     4     1     3    16
## 37 Ghana                     4     5     3    -3    15
## 38 Republic of Ireland       2     3     8     0    14
## 39 Northern Ireland          3     5     5   -10    14
## 40 Ecuador                   4     5     1    -1    13
## 41 Serbia                    4     8     1   -10    13
## 42 Senegal                   3     2     3     1    12
## 43 Algeria                   3     7     3    -6    12
## 44 Morocco                   2     9     5    -8    11
## 45 Saudi Arabia              3    11     2   -28    11
## 46 Ivory Coast               3     5     1    -1    10
```

```
## 47 South Africa             2     3     4    -5    10
## 48 Tunisia                  2     9     4   -12    10
## 49 Iran                     2     9     4   -15    10
## 50 Australia                2    10     4   -18    10
## 51 Norway                   2     3     3    -1     9
## 52 German DR                2     2     2     0     8
## 53 Greece                   2     6     2   -15     8
## 54 Ukraine                  2     2     1    -2     7
## 55 Wales                    1     1     3     0     6
## 56 Slovakia                 1     2     1    -2     4
## 57 Slovenia                 1     4     1    -5     4
## 58 Cuba                     1     1     1    -7     4
## 59 North Korea              1     5     1   -15     4
## 60 Bosnia and Herzegovina   1     2     0     0     3
## 61 Czech Republic           1     2     0    -1     3
## 62 Jamaica                  1     2     0    -6     3
## 63 New Zealand              0     3     3   -10     3
## 64 Honduras                 0     6     3   -11     3
## 65 Angola                   0     1     2    -1     2
## 66 Israel                   0     1     2    -2     2
## 67 Egypt                    0     5     2    -7     2
## 68 Iceland                  0     2     1    -3     1
## 69 Kuwait                   0     2     1    -4     1
## 70 Trinidad and Tobago      0     2     1    -4     1
## 71 Bolivia                  0     5     1   -19     1
## 72 Iraq                     0     3     0    -3     0
## 73 Canada                   0     3     0    -5     0
## 74 Togo                     0     3     0    -5     0
## 75 Indonesia                0     1     0    -6     0
## 76 China PR                 0     3     0    -9     0
## 77 Panama                   0     3     0    -9     0
## 78 United Arab Emirates     0     3     0    -9     0
## 79 Haiti                    0     3     0   -12     0
## 80 DR Congo                 0     3     0   -14     0
## 81 El Salvador              0     6     0   -21     0
```

6.5 AVERAGE SCORES FOR WORLD CUP TEAMS

```
FIFA_WC_ALL_TEAMS_SCORE_MEANS %>% rename("TEAM" = FIFA_Participating_Teams, "MEAN SCORE" = FIFA_WC_ALL_S
```

```
## # A tibble: 81 x 2
##    TEAM                   `MEAN SCORE`
##    <chr>                         <dbl>
##  1 Algeria                        1
##  2 Angola                         0.33
##  3 Argentina                      1.69
##  4 Australia                      0.81
##  5 Austria                        1.48
##  6 Belgium                        1.42
##  7 Bolivia                        0.17
##  8 Bosnia and Herzegovina         1.33
##  9 Brazil                         2.1
## 10 Bulgaria                       0.85
## 11 Cameroon                       0.78
## 12 Canada                         0
## 13 Chile                          1.21
## 14 China PR                       0
## 15 Colombia                       1.45
## 16 Costa Rica                     1.06
## 17 Croatia                        1.52
## 18 Cuba                           1.67
## 19 Czech Republic                 1
## 20 Czechoslovakia                 1.47
## 21 Denmark                        1.5
## 22 DR Congo                       0
## 23 Ecuador                        1
## 24 Egypt                          0.71
## 25 El Salvador                    0.17
## 26 England                        1.32
## 27 France                         1.82
## 28 German DR                      0.83
## 29 Germany                        2.07
## 30 Ghana                          1.08
## 31 Greece                         0.5
## 32 Haiti                          0.67
## 33 Honduras                       0.33
## 34 Hungary                        2.72
## 35 Iceland                        0.67
## 36 Indonesia                      0
## 37 Iran                           0.6
## 38 Iraq                           0.33
## 39 Israel                         0.33
## 40 Italy                          1.54
## 41 Ivory Coast                    1.44
## 42 Jamaica                        1
## 43 Japan                          0.95
## 44 Kuwait                         0.67
## 45 Mexico                         1.05
## 46 Morocco                        0.88
```

```
## 47 Netherlands                 1.72
## 48 New Zealand                 0.67
## 49 Nigeria                     1.1
## 50 North Korea                 0.86
## 51 Northern Ireland            1
## 52 Norway                      0.88
## 53 Panama                      0.67
## 54 Paraguay                    1.11
## 55 Peru                        1.17
## 56 Poland                      1.35
## 57 Portugal                    1.63
## 58 Republic of Ireland         0.77
## 59 Romania                     1.43
## 60 Russia                      1.71
## 61 Saudi Arabia                0.69
## 62 Scotland                    1.09
## 63 Senegal                     1.38
## 64 Serbia                      0.85
## 65 Slovakia                    1.25
## 66 Slovenia                    0.83
## 67 South Africa                1.22
## 68 South Korea                 1
## 69 Spain                       1.57
## 70 Sweden                      1.57
## 71 Switzerland                 1.35
## 72 Togo                        0.33
## 73 Trinidad and Tobago         0
## 74 Tunisia                     0.87
## 75 Turkey                      2
## 76 Ukraine                     1
## 77 United Arab Emirates        0.67
## 78 United States               1.12
## 79 Uruguay                     1.55
## 80 Wales                       0.8
## 81 Yugoslavia                  1.67
```

6.6 AVERAGE SCORES FOR TEAMS IN FRIENDLY MATCHES

```
FM_ALL_MEAN_SCORES %>% rename("TEAM" = FM_Teams, "MEAN SCORE" = FM_ALL_SCORE_MEANS) %>% tbl_df %>% prin
```

```
## # A tibble: 264 x 2
##     TEAM                              `MEAN SCORE`
##     <chr>                                    <dbl>
##   1 Abkhazia                                  0.5
##   2 Afghanistan                               0.9
##   3 Albania                                   1.21
##   4 Algeria                                   1.29
##   5 American Samoa                            0
##   6 Andalusia                                 1.92
##   7 Andorra                                   0.37
##   8 Angola                                    1.02
##   9 Anguilla                                  0.71
##  10 Antigua and Barbuda                       1.31
##  11 Argentina                                 1.74
##  12 Armenia                                   1.04
##  13 Artsakh                                   2
##  14 Aruba                                     1.38
##  15 Asturias                                  3
##  16 Australia                                 1.58
##  17 Austria                                   1.79
##  18 Azerbaijan                                0.92
##  19 Bahamas                                   1.57
##  20 Bahrain                                   1.36
##  21 Bangladesh                                1.02
##  22 Barawa                                    1.75
##  23 Barbados                                  1.29
##  24 Basque Country                            2.69
##  25 Belarus                                   1.41
##  26 Belgium                                   1.73
##  27 Belize                                    1.33
##  28 Benin                                     1.08
##  29 Bermuda                                   1.11
##  30 Bhutan                                    0.85
##  31 Bolivia                                   1.04
##  32 Bosnia and Herzegovina                    1.07
##  33 Botswana                                  0.88
##  34 Brazil                                    2.2
##  35 British Virgin Islands                    1.18
##  36 Brittany                                  1.5
##  37 Brunei                                    1.11
##  38 Bulgaria                                  1.36
##  39 Burkina Faso                              1.16
##  40 Burundi                                   1.49
##  41 Cambodia                                  1.1
##  42 Cameroon                                  1.25
##  43 Canada                                    1.01
##  44 Canary Islands                            2.75
##  45 Cape Verde                                1.27
##  46 Cascadia                                  6
```

```
##  47 Catalonia                             1.56
##  48 Cayman Islands                        0.78
##  49 Central African Republic              0.96
##  50 Central Spain                         1
##  51 Chad                                  0.68
##  52 Chagos Islands                        1.33
##  53 Chile                                 1.4
##  54 China PR                              1.7
##  55 Colombia                              1.3
##  56 Comoros                               1.53
##  57 Congo                                 1.46
##  58 Cook Islands                          1
##  59 Corsica                               1.11
##  60 Costa Rica                            1.28
##  61 Croatia                               1.75
##  62 Cuba                                  0.93
##  63 Curaçao                               1.78
##  64 Cyprus                                0.88
##  65 Czech Republic                        1.78
##  66 Czechoslovakia                        1.83
##  67 Denmark                               1.75
##  68 Djibouti                              0.95
##  69 Dominica                              1.24
##  70 Dominican Republic                    1.21
##  71 DR Congo                              1.47
##  72 Ecuador                               1.31
##  73 Egypt                                 1.51
##  74 El Salvador                           0.89
##  75 Ellan Vannin                          1
##  76 England                               2.19
##  77 Equatorial Guinea                     1
##  78 Eritrea                               0.56
##  79 Estonia                               1.1
##  80 Eswatini                              0.85
##  81 Ethiopia                              1.8
##  82 Faroe Islands                         1.11
##  83 Fiji                                  1.55
##  84 Finland                               1.29
##  85 France                                1.69
##  86 French Guiana                         1.25
##  87 Gabon                                 1.11
##  88 Galicia                               2
##  89 Gambia                                0.94
##  90 Georgia                               1.3
##  91 German DR                             1.7
##  92 Germany                               2.15
##  93 Ghana                                 1.73
##  94 Gibraltar                             0.64
##  95 Greece                                1.18
##  96 Greenland                             1.12
##  97 Grenada                               1.28
##  98 Guadeloupe                            1.47
##  99 Guam                                  2.65
## 100 Guatemala                             1.12
```

```
## 101 Guernsey                                       1.09
## 102 Guinea                                         1.31
## 103 Guinea-Bissau                                  1.41
## 104 Guyana                                         1.12
## 105 Haiti                                          1.34
## 106 Honduras                                       1.27
## 107 Hong Kong                                      1.54
## 108 Hungary                                        2.07
## 109 Iceland                                        1.41
## 110 India                                          1.48
## 111 Indonesia                                      1.65
## 112 Iran                                           1.55
## 113 Iraq                                           1.23
## 114 Israel                                         1.44
## 115 Italy                                          1.72
## 116 Ivory Coast                                    1.49
## 117 Jamaica                                        1.22
## 118 Japan                                          1.61
## 119 Jersey                                         1.4
## 120 Jordan                                         1.04
## 121 Kazakhstan                                     1.01
## 122 Kenya                                          1.68
## 123 Kernow                                         2.73
## 124 Kosovo                                         2
## 125 Kuwait                                         1.28
## 126 Kyrgyzstan                                     0.8
## 127 Laos                                           0.9
## 128 Latvia                                         1.32
## 129 Lebanon                                        0.95
## 130 Lesotho                                        0.81
## 131 Liberia                                        0.9
## 132 Libya                                          1.21
## 133 Liechtenstein                                  0.52
## 134 Lithuania                                      1.16
## 135 Luxembourg                                     0.76
## 136 Macau                                          0.58
## 137 Madagascar                                     1.71
## 138 Madrid                                         1
## 139 Malawi                                         1.29
## 140 Malaysia                                       1.36
## 141 Maldives                                       0.97
## 142 Mali                                           1.23
## 143 Malta                                          0.81
## 144 Manchukuo                                      0
## 145 Martinique                                     1.87
## 146 Mauritania                                     0.64
## 147 Mauritius                                      1.93
## 148 Mayotte                                        0.86
## 149 Mexico                                         1.52
## 150 Micronesia                                     2.67
## 151 Moldova                                        0.88
## 152 Monaco                                         1.22
## 153 Mongolia                                       1
## 154 Montenegro                                     1.17
```

```
## 155 Montserrat                              0
## 156 Morocco                                 1.41
## 157 Mozambique                              1.07
## 158 Myanmar                                 1.14
## 159 Namibia                                 0.82
## 160 Nepal                                   0.87
## 161 Netherlands                             1.98
## 162 New Caledonia                           1.74
## 163 New Zealand                             1.54
## 164 Nicaragua                               0.98
## 165 Niger                                   1.03
## 166 Nigeria                                 1.44
## 167 North Korea                             1.43
## 168 North Macedonia                         1.17
## 169 North Vietnam                           1.74
## 170 Northern Cyprus                         0.8
## 171 Northern Ireland                        0.96
## 172 Northern Mariana Islands                1.17
## 173 Norway                                  1.53
## 174 Oman                                    1.27
## 175 Pakistan                                1.16
## 176 Palau                                   2
## 177 Palestine                               1.14
## 178 Panama                                  0.98
## 179 Panjab                                  4
## 180 Papua New Guinea                        0.89
## 181 Paraguay                                1.25
## 182 Parishes of Jersey                      2
## 183 Peru                                    1.17
## 184 Philippines                             1.32
## 185 Poland                                  1.68
## 186 Portugal                                1.57
## 187 Provence                                0
## 188 Puerto Rico                             0.74
## 189 Qatar                                   1.25
## 190 Republic of Ireland                     1.37
## 191 Réunion                                 1.37
## 192 Romania                                 1.5
## 193 Russia                                  1.58
## 194 Rwanda                                  1.03
## 195 Saarland                                1.2
## 196 Saint Helena                            2
## 197 Saint Kitts and Nevis                   1.39
## 198 Saint Lucia                             1.09
## 199 Saint Martin                            1.89
## 200 Saint Vincent and the Grenadines        1.7
## 201 Samoa                                   0.25
## 202 San Marino                              0.32
## 203 São Tomé and Príncipe                   1.3
## 204 Saudi Arabia                            1.38
## 205 Scotland                                1.62
## 206 Senegal                                 1.14
## 207 Serbia                                  1.26
## 208 Seychelles                              0.97
```

```
## 209 Sierra Leone                               0.98
## 210 Silesia                                    3
## 211 Singapore                                  1.41
## 212 Sint Maarten                               1
## 213 Slovakia                                   1.3
## 214 Slovenia                                   1.16
## 215 Solomon Islands                            1.64
## 216 Somalia                                    1.36
## 217 Somaliland                                 2
## 218 South Africa                               1.41
## 219 South Korea                                1.43
## 220 South Sudan                                0.5
## 221 Spain                                      1.85
## 222 Sri Lanka                                  0.92
## 223 Sudan                                      1.09
## 224 Suriname                                   2.24
## 225 Surrey                                     3
## 226 Sweden                                     1.93
## 227 Switzerland                                1.4
## 228 Syria                                      1.05
## 229 Tahiti                                     1.76
## 230 Taiwan                                     1.76
## 231 Tajikistan                                 1.16
## 232 Tanzania                                   1.18
## 233 Thailand                                   1.6
## 234 Tibet                                      0.67
## 235 Timor-Leste                                0.7
## 236 Togo                                       1.14
## 237 Tonga                                      0.71
## 238 Trinidad and Tobago                        1.66
## 239 Tunisia                                    1.35
## 240 Turkey                                     1.35
## 241 Turkmenistan                               1.14
## 242 Turks and Caicos Islands                   1
## 243 Tuvalu                                     1.17
## 244 Uganda                                     1.69
## 245 Ukraine                                    1.36
## 246 United Arab Emirates                       1.43
## 247 United States                              1.3
## 248 United States Virgin Islands               0.27
## 249 Uruguay                                    1.46
## 250 Uzbekistan                                 1.3
## 251 Vanuatu                                    1.3
## 252 Vatican City                               0.33
## 253 Venezuela                                  1.28
## 254 Vietnam                                    1.83
## 255 Vietnam Republic                           1.61
## 256 Wales                                      1.07
## 257 Yemen                                      0.97
## 258 Yemen DPR                                  1.5
## 259 Ynys Môn                                   7
## 260 Yorkshire                                  4.33
## 261 Yugoslavia                                 1.9
## 262 Zambia                                     1.69
```

```
## 263 Zanzibar                                    0.82
## 264 Zimbabwe                                    1.4
```