

DaigleInClassLab_Wk5D2.R

2011home

Wed Feb 14 17:00:01 2018

```
## Christopher Daigle
## Wk5D2InClassLab
```

```
## Exercise 1
```

```
# How many days are between Jan 13 1903 and Feb 14 2018?
```

```
dates1 <- c("01/13/1903", "02/14/2018")
dates1 <- as.Date(dates1, format = "%m/%d/%Y")
daysBtwn1 <- dates1[2] - dates1[1]
daysBtwn1
```

```
## Time difference of 42036 days
```

```
# Create a sequence of dates, by day, from Jan 1 2018 to Feb 14 2018
```

```
dates2 <- as.Date(c("01/01/2018", "02/14/2018"), format = "%m/%d/%Y")
daysBtwn2 <- dates2[2] - dates2[1]
dateSeq <- dates2[1] + 0:daysBtwn2
```

```
## Exercise 2
```

```
# Download "complete.csv" from "International football results from 1872 to 2018" from Kaggle
```

```
setwd("/Users/2011home/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/R/Week5")
dir()
```

```
## [1] "complete.csv"           "district_rev_exp_readtable.txt"
## [3] "district_rev_exp.csv"   "district_rev_exp.txt"
## [5] "district_rev_exp.xlsx"  "results.csv"
## [7] "school15doc.pdf"
```

```
data <- read.csv("results.csv")
head(data)
```

```
##      date home_team away_team home_score away_score tournament  city
## 1 1872-11-30  Scotland   England         0         0   Friendly Glasgow
## 2 1873-03-08   England  Scotland         4         2   Friendly  London
## 3 1874-03-07  Scotland   England         2         1   Friendly Glasgow
## 4 1875-03-06   England  Scotland         2         2   Friendly  London
## 5 1876-03-04  Scotland   England         3         0   Friendly Glasgow
## 6 1876-03-25  Scotland    Wales         4         0   Friendly Glasgow
##      country
## 1 Scotland
## 2  England
## 3 Scotland
## 4  England
## 5 Scotland
## 6 Scotland
```

```
str(data)
```

```
## 'data.frame':   38759 obs. of  8 variables:
## $ date      : Factor w/ 14772 levels "1872-11-30","1873-03-08",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ home_team : Factor w/ 244 levels "Afghanistan",...: 187 67 187 67 187 187 67 237 187 187 ...
```

```

## $ away_team : Factor w/ 244 levels "Afghanistan",...: 66 183 66 183 66 236 183 183 66 236 ...
## $ home_score: int 0 4 2 2 3 4 1 0 7 9 ...
## $ away_score: int 0 2 1 2 0 0 3 2 2 0 ...
## $ tournament: Factor w/ 95 levels "ABCS Tournament",...: 48 48 48 48 48 48 48 48 48 48 ...
## $ city : Factor w/ 1798 levels "Atele", "6th of October City",...: 585 919 585 919 585 585 919 ...
## $ country : Factor w/ 268 levels "Afghanistan",...: 204 70 204 70 204 204 70 259 204 204 ...

# Convert the variable "date" to variable using as.Date
data$date <- as.Date(data$date, format = "%Y-%m-%d")
str(data)

## 'data.frame': 38759 obs. of 8 variables:
## $ date : Date, format: "1872-11-30" "1873-03-08" ...
## $ home_team : Factor w/ 244 levels "Afghanistan",...: 187 67 187 67 187 187 67 237 187 187 ...
## $ away_team : Factor w/ 244 levels "Afghanistan",...: 66 183 66 183 66 236 183 183 66 236 ...
## $ home_score: int 0 4 2 2 3 4 1 0 7 9 ...
## $ away_score: int 0 2 1 2 0 0 3 2 2 0 ...
## $ tournament: Factor w/ 95 levels "ABCS Tournament",...: 48 48 48 48 48 48 48 48 48 48 ...
## $ city : Factor w/ 1798 levels "Atele", "6th of October City",...: 585 919 585 919 585 585 919 ...
## $ country : Factor w/ 268 levels "Afghanistan",...: 204 70 204 70 204 204 70 259 204 204 ...

# Compare the average number of goals in 1930-1939 and in 2005-2014 and see whether it is increased or decreased
year <- format(data$date, format = "%Y")
period1 <- data[year >= 1930 & year <= 1939, ]
period2 <- data[year >= 2005 & year <= 2014, ]
goals1 <- period1$home_score + period1$away_score
goals2 <- period2$home_score + period2$away_score
avgGoals1 <- mean(goals1)
avgGoals2 <- mean(goals2)

avgGoals2 > avgGoals1

## [1] FALSE

# The average number of goals decreased

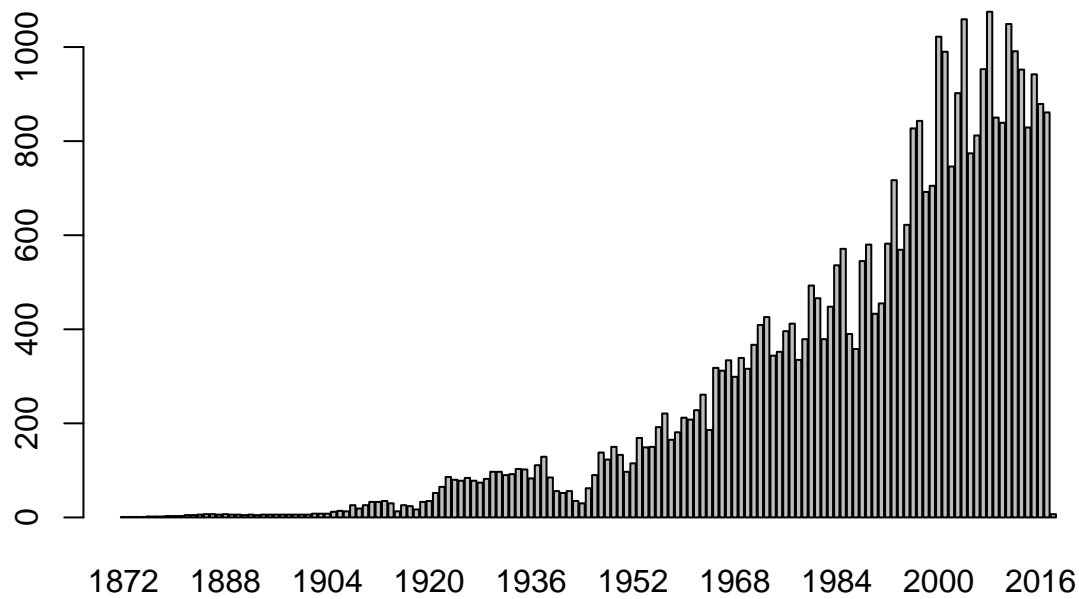
# What is the ratio of home winning games to the total number of games in the 1950s?
game50 <- data[year >= 1950 & year <= 1959, ]
homeWin <- data[data$home_score > data$away_score, ]
ratio <- length(homeWin$home_score) / length(data$home_score)
ratio

## [1] 0.4824686

# The ratio is approximately 0.48

# Plot the number of games over each year and see whether there was a decrease in games in WW1 and WW2
data$year <- format(data$date, format = "%Y")
plot(factor(data$year))

```



```
# This is extra, an experiment in playing with tables and plots of sort
# WW1 <- data[year >= 1914 & year <= 1918,]
# # Segment
# WW1yr <- format(WW1$date, format = "%Y")
# # Simplify date
# WW1freq <- table(WW1yr)
# # Distinguish each year and the number of times they occur (counting the number of games in the year)
# WW1freq
#
# WW2 <- data[year >= 1939 & year <= 1945,]
# WW2yr <- format(WW2$date, format = "%Y")
# WW2freq <- table(WW2yr)
# WW2freq
#
# plot(c(WW1freq,WW2freq), type = "b")
# WW1freq
# WW2freq
```