Chess_Transfers

josef_waples

8/27/2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
# Week 1 Assignment - Loading Data into a Data Frame
# Choose one dataset on this url: https://data.fivethirtyeight.com/
# "American Chess Is Great Again" - Chess transfer data
chess_transfer_data <- read.csv('~/Downloads/transfers_csv.csv')
# First thing first - look at the structure of the dataframe. Also, look at the first 3 and last 3 rows
str(chess_transfer_data)
```

```
## 'data.frame':    932 obs. of  5 variables:
##  $ url          : Factor w/ 18 levels "https://ratings.fide.com/fedchange.phtml?year=2000",..: 1 1 1
##  $ id           : int  2019221 14401754 14401762 2019221 14401754 14401762 6700284 1613782 2600536 2
##  $ federation   : Factor w/ 105 levels "aho","alb","alg",..: 101 14 14 101 14 14 29 9 8 18 ...
##  $ form.fed     : Factor w/ 101 levels "","aho","alb",..: 70 23 100 70 23 100 40 47 18 100 ...
##  $ transfer.date: Factor w/ 545 levels "1/10/14","1/11/07",..: 110 32 32 110 32 32 84 464 102 334 ..
```

```
head(chess_transfer_data, n = 3)
```

```
##                                                 url       id federation
## 1 https://ratings.fide.com/fedchange.phtml?year=2000  2019221        usa
## 2 https://ratings.fide.com/fedchange.phtml?year=2000 14401754        bih
## 3 https://ratings.fide.com/fedchange.phtml?year=2000 14401762        bih
##   form.fed transfer.date
## 1      phi      12/15/00
## 2      cro       1/31/00
## 3      yug       1/31/00
```

```
tail(chess_transfer_data, n = 3)
```

```
##                                                   url       id federation
## 930 https://ratings.fide.com/fedchange.phtml?year=2017  2002515        pol
## 931 https://ratings.fide.com/fedchange.phtml?year=2017   407747        sco
## 932 https://ratings.fide.com/fedchange.phtml?year=2017 13900820        ger
##     form.fed transfer.date
```

```
## 930      usa        1/9/17
## 931      eng       1/12/17
## 932      fid       3/29/17
```

```
# After looking at the structure, I think the variable 'transfer.date' should be a date, not a factor
chess_transfer_data$transfer.date <- as.Date(chess_transfer_data$transfer.date, "%m/%d/%y")
str(chess_transfer_data)
```

```
## 'data.frame':    932 obs. of  5 variables:
##  $ url          : Factor w/ 18 levels "https://ratings.fide.com/fedchange.phtml?year=2000",..: 1 1 1
##  $ id           : int  2019221 14401754 14401762 2019221 14401754 14401762 6700284 1613782 2600536 2
##  $ federation   : Factor w/ 105 levels "aho","alb","alg",..: 101 14 14 101 14 14 29 9 8 18 ...
##  $ form.fed     : Factor w/ 101 levels "","aho","alb",..: 70 23 100 70 23 100 40 47 18 100 ...
##  $ transfer.date: Date, format: "2000-12-15" "2000-01-31" ...
```

```
# Now I want to look at transfer date as a year only. Substring works because the format is the same fo
chess_transfer_data$transfer.date <- substring(chess_transfer_data$transfer.date,1,4)
str(chess_transfer_data)
```

```
## 'data.frame':    932 obs. of  5 variables:
##  $ url          : Factor w/ 18 levels "https://ratings.fide.com/fedchange.phtml?year=2000",..: 1 1 1
##  $ id           : int  2019221 14401754 14401762 2019221 14401754 14401762 6700284 1613782 2600536 2
##  $ federation   : Factor w/ 105 levels "aho","alb","alg",..: 101 14 14 101 14 14 29 9 8 18 ...
##  $ form.fed     : Factor w/ 101 levels "","aho","alb",..: 70 23 100 70 23 100 40 47 18 100 ...
##  $ transfer.date: chr  "2000" "2000" "2000" "2000" ...
```

```
# loading dplyr so I can use the pipe operator - %>%
library(dplyr)
```

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

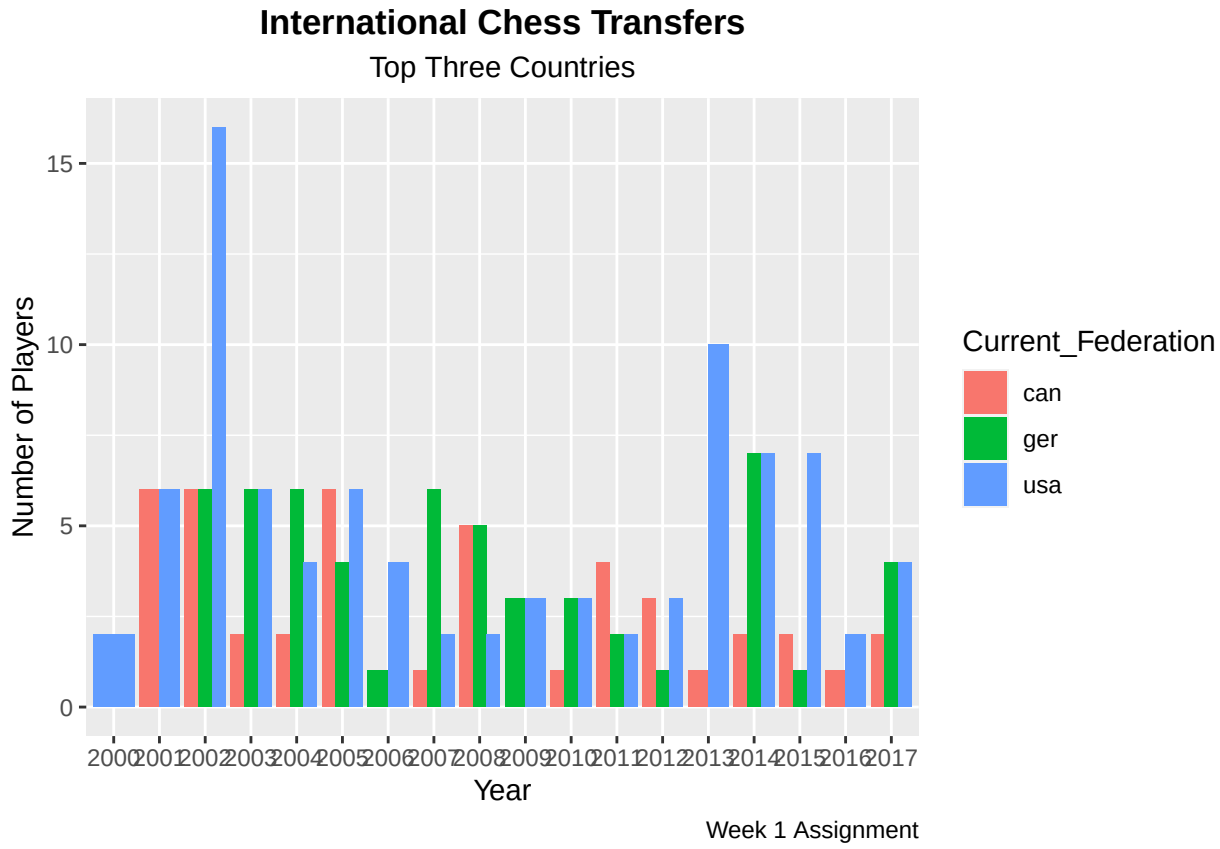
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
# how many transfers happen for each federation each year?
chess_transfers_clean <- chess_transfer_data %>%
  group_by(federation, transfer.date) %>%
  count()
# which country has the most transfers total?
count_of_total_transfers_by_country_table <- chess_transfers_clean %>%
  group_by(federation) %>%
  summarize(sum = sum(n))
# there are probably too many countries to graph - I'm choosing top three
top_three_countries_total <- count_of_total_transfers_by_country_table %>%
  top_n(3)
```

```
## Selecting by sum
```

```
# let's merge the dataframes so everything is in one place
totals_chess_transfers_table <- merge(chess_transfers_clean, count_of_total_transfers_by_country_table,
# I'm going to rename the headers
names(totals_chess_transfers_table) <- c("Current_Federation", "Transfer_Year", "Number_of_Tranfers_in_
# take the main data frame and filter to include only the countries whose total number of transfers wil
chess_transfers_top_3 <- totals_chess_transfers_table %>%
  filter(Current_Federation %in% top_three_countries_total$federation)
# try to graph this as a scatterplot
library(ggplot2)
ggplot(data = chess_transfers_top_3, aes(x = Transfer_Year, y = Number_of_Tranfers_in_Year, fill = Curr
  geom_col(position = "dodge") +
    labs(title="International Chess Transfers", subtitle = "Top Three Countries") +
   theme(plot.title = element_text(hjust = 0.5, face = "bold")) +
  theme(plot.subtitle = element_text(hjust = 0.5)) +
  labs(caption = "Week 1 Assignment") +
  xlab("Year") +
  ylab("Number of Players")
```




```
# this graph seems to support the meaning of the article. The United States does seem to be trying to a
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.