Data 413: Classwork/Lab 4

Frankie Tyndall

2022-07-07

## Using the web address <https://en.wikipedia.org/wiki/ITF_Rankings> and the R coding structure presented in class to web scrape the following table found on the page.

wikiurl <- read\_html(  
 "https://en.wikipedia.org/wiki/ITF\_Rankings")  
ITF\_Rankings <- wikiurl%>%  
 html\_table(., fill = T)  
  
ITF\_Rankings[[3]] -> ITF  
ITF

## # A tibble: 7 × 3  
## `Opponent Nation Ranking1` `Bonus Points` `Bonus Points`  
## <chr> <chr> <chr>   
## 1 Opponent Nation Ranking1 Away Home   
## 2 1 to 2 125 100   
## 3 3 to 4 112.5 90   
## 4 5 to 8 93.75 75   
## 5 9 to 16 62.5 50   
## 6 17 to 32 50 40   
## 7 33 to 64 31.25 25

## Using the web address <https://www.mlb.com/stats/2018> and the R coding structure presented in class, web scrape the table found on the page.

wikiurl <- read\_html("https://www.mlb.com/stats/2018")  
baseballdata2018 <- wikiurl%>%  
 html\_table(., fill = T)  
  
baseballdata2018[[1]] -> BD2018  
BD2018

## # A tibble: 25 × 18  
## PLAYERPLAYER TEAMTEAM GG ABAB RR HH `2B2B` `3B3B` HRHR RBIRBI  
## <chr> <chr> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 1MikeM TroutTrou… LAA 140 471 101 147 24 4 39 79  
## 2 2MookieM BettsBe… BOS 136 520 129 180 47 5 32 80  
## 3 3J.D.J MartinezM… BOS 150 569 111 188 37 2 43 130  
## 4 4ChristianC Yeli… MIL 147 574 118 187 34 7 36 110  
## 5 5JoseJ RamírezRa… CLE 157 578 110 156 38 4 39 105  
## 6 6NolanN ArenadoA… COL 156 590 104 175 38 2 38 110  
## 7 7AlexA BregmanBr… HOU 157 594 105 170 51 1 31 103  
## 8 8PaulP Goldschmi… ARI 158 593 95 172 35 5 33 83  
## 9 9TrevorT StorySt… COL 157 598 88 174 42 6 37 108  
## 10 10AnthonyA Rendo… WSH 136 529 88 163 44 2 24 92  
## # … with 15 more rows, and 8 more variables: BBBB <int>, SOSO <int>,  
## # SBSB <int>, CSCS <int>, AVGAVG <dbl>, OBPOBP <dbl>, SLGSLG <dbl>,  
## # `caret-upcaret-downOPScaret-upcaret-downOPS` <dbl>

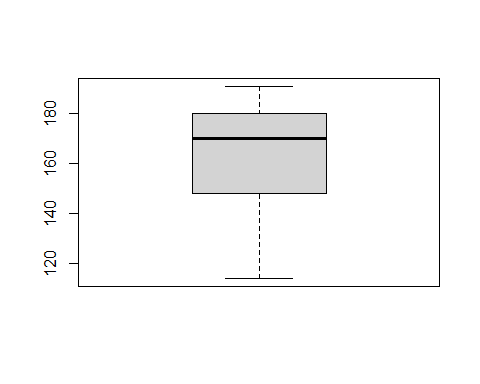
## Use and show R code to find the average number of hits for all players in the table from number 4

hit <- BD2018 %>%  
 select(HH)  
  
mean(hit$HH)

## [1] 163.52

## Use and show R code to produce a boxplot for the number of hits (use tidyverse/ggplot coding). Use the data table from number 4.

boxplot(hit$HH)

 ## Use and show dplyr coding to determine which player had the greatest number of strikeouts using the data table from number 4.

strikeout <- BD2018 %>%  
 select(c(PLAYERPLAYER, SOSO)) %>%  
 arrange(desc(SOSO))  
  
strikeout

## # A tibble: 25 × 2  
## PLAYERPLAYER SOSO  
## <chr> <int>  
## 1 20KhrisK DavisDavisDH20‌‌‌ 175  
## 2 8PaulP GoldschmidtGoldschmidt1B8‌‌‌ 173  
## 3 16BryceB HarperHarperRF16‌‌‌ 169  
## 4 9TrevorT StoryStorySS9‌‌‌ 168  
## 5 19JavierJ BáezBaez2B19‌‌‌ 167  
## 6 12MattM CarpenterCarpenter1B12‌‌‌ 158  
## 7 25MitchM HanigerHanigerRF25‌‌‌ 148  
## 8 3J.D.J MartinezMartinezDH3‌‌‌ 146  
## 9 23MattM ChapmanChapman3B23‌‌‌ 146  
## 10 15JesusJ AguilarAguilar1B15‌‌‌ 143  
## # … with 15 more rows

## Use and show dplyr coding to show the batting averages for Washington Nationals players and Colorado Rockies players using the data table from number4

batting\_average <- BD2018 %>%  
 filter(TEAMTEAM == "WSH"|TEAMTEAM == "COL") %>%  
 select(c(TEAMTEAM, AVGAVG)) %>%  
 group\_by(TEAMTEAM)  
   
  
batting\_average

## # A tibble: 5 × 2  
## # Groups: TEAMTEAM [2]  
## TEAMTEAM AVGAVG  
## <chr> <dbl>  
## 1 COL 0.297  
## 2 COL 0.291  
## 3 WSH 0.308  
## 4 WSH 0.249  
## 5 COL 0.291