DS 413/613 LAB Webscraping for tables KEY

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

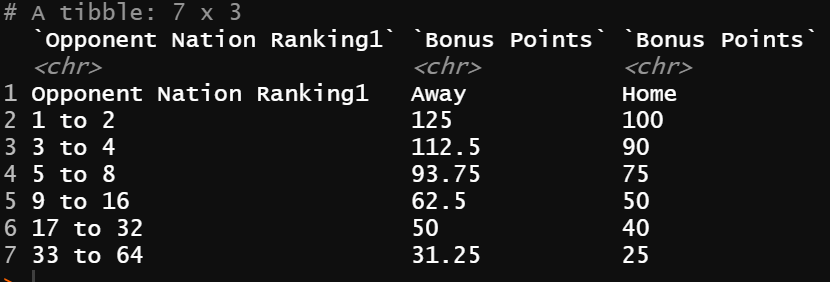
wikiurl <- read\_html("https://en.wikipedia.org/wiki/ITF\_Rankings")

tennisdataITF <- wikiurl%>%

html\_table(., fill = T)

tennisdataITF[[3]] -> TDITF

TDITF



2) 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

3) Use and show R code to find the average number of hits for all players in the table from number 2

mean(BD2018$HH)

4) Use and show R code to produce a boxplot for the number of

hits (use tidyverse/ggplot coding). Use the data table from number 2.

ggplot(data = BD2018) +

geom\_boxplot(mapping = aes(y = HH))

5) Use and show dplyr coding to determine which player had the

greatest number of strikeouts from the data table from number 2.

BD2018%>%

select(PLAYERPLAYER, SOSO)%>%

arrange(desc(SOSO))

6) Use and show dplyr coding to show the batting averages for

Washington Nationals players and Colorado Rockies players.

BD2018%>%

select(PLAYERPLAYER,TEAMTEAM, AVGAVG)%>%

filter(TEAMTEAM == "WSH"| TEAMTEAM == "COL")%>%

arrange(PLAYERPLAYER)