**03/06/2023** - Mario and Nathan dataset downloaded - on the hunt for outside source

**03/10/2023** - plan to have datasets merged and looking nice by wednesday 🙂

**03/12/2023** - drop columns we find to be less relevant;

Teams and games are same dataset, so got rid of teams

Players was not modified here

ranking1 = ranking %>% select(-c(LEAGUE\_ID, SEASON\_ID, RETURNTOPLAY, CONFERENCE))

games1 = games %>% select(-c(GAME\_STATUS\_TEXT, TEAM\_ID\_away, TEAM\_ID\_home)) rename to date

details1 = details %>% select(-c(COMMENT, TEAM\_ABBREVIATION)) %>% drop\_na() - got rid of players that dnp

[~~https://www.rdocumentation.org/packages/nbastatR/versions/0.1.10131~~](https://www.rdocumentation.org/packages/nbastatR/versions/0.1.10131) ~~i put this into R’~~

Joined together rankings and games into one dataset by game\_id and joined details and players into another by player\_id; one is more of player-level and the other is game-level

For the player and rank, nullify any previous seasons/teams the player played with; only counting them for the team they play for now

**03/16/2023** -

ranking1 = ranking %>% select(-c(LEAGUE\_ID, SEASON\_ID, RETURNTOPLAY, CONFERENCE)) %>% rename("DATE" = "STANDINGSDATE") #get rid of G, W, L ??

games1 = games %>% select(-c(GAME\_STATUS\_TEXT, TEAM\_ID\_away, TEAM\_ID\_home)) %>% rename("DATE" = "GAME\_DATE\_EST", "TEAM\_ID" = "HOME\_TEAM\_ID")

details1 = details %>% select(-c(COMMENT, TEAM\_ABBREVIATION)) %>% drop\_na() #deleted players who dnp as well

player\_join = inner\_join(details1, players, by = c("PLAYER\_ID", "PLAYER\_NAME"))

rank\_data = inner\_join(ranking1, games1, by = c("DATE", "TEAM\_ID"))

player\_data = player\_join %>% select(-c(TEAM\_ID.y, SEASON)) %>% rename("TEAM\_ID" = "TEAM\_ID.x") %>% distinct()

**3/20/2023** -

* Narrowed down data selection to 2022-2023 season

**Data Scraping from Five Thirty Eight**

**3/14/2023 thru 3/15/2023 -** Looking through and grabbing data from betting websites.

**3/16/2023 thru 3/19/2023** - Used microsoft's Power Automate to scrape spread predictions from Five Thirty Eight website and put into excel spreadsheet. Data will be scraped daily to note any changes in predictions.

**3/19/2023 -** Joined spread predictions with the IDs of home and away teams instead of team Nicknames

**3/20/2023 thru 3/21/2023 -** Added extra variables regarding average points per game, average steals per game, and average rebounds per game and how it changes as the season goes on.