**R Data analysis subject**

Need to be presented in an R markdown file

Methods:

“The methods in the report have been thoroughly and effectively described, providing logical justifications for the analytical decisions made.”

List all the changes I make to the R document.

1. Opening libraries

- tidyverse

- dplyr

- broom

2. Reading in files

sal <- read\_csv("data/raw/2018-19\_nba\_player-salaries.csv")

stat\_pl <- read\_csv("data/raw/2018-19\_nba\_player\_statistics.csv")

stat\_tm1 <- read\_csv("data/raw/2018-19\_nba\_team\_statistics\_1.csv")

stat\_tm2 <- read\_csv("data/raw/2018-19\_nba\_team\_statistics\_2.csv")

pay <- read\_csv("data/raw/2019-20\_nba\_team-payroll.csv")

3. Tidying and transforming data

a. Cleaning:

i. Combining variables from two different data frames (R Binding)

Adding stats from team1 data frame to team data frame 2. Did not include rank, team or age, or the last 3 columns of team1 data in binding of two data frames.

ii. Renaming variables (r renaming)

a. We want to find the average minutes played her game. Therefore we want to create a new variable: AMPG (average minutes played per game).

b. Renaming variables: changing variables in new ‘df\_teams’ data frame to not start with a number or not include a ‘%’ sign. Added a ‘x’ in front of number variables and changed ‘%’ sign to ‘p’ , and change ‘/’ to \_per\_

c. Using this new variable: win percentage (WINp) = wins / games.

Explore the relationship between this and points which we believe to be the outcome that will result in wins. But we want to see how statistically likely this will be.

* Win percentage per game
* 2PFGA
* FTA
* Steals
* Blocks
* Turnovers

Per minute played per game

d. We also want to get the number of :

* Point
* Assists
* FGA
* 3PFGA
* 2PFGA
* FTA
* Steals
* Blocks
* Turnovers

Per game.

Exploring data:

1. Compare points per game / minutes played compared to :

* FGA
* Assists
* 3PFGA
* 2PFGA
* FTA
* Steals
* Blocks
* Turnovers
* ORB
* DRB
* TOB

# Found that only 3P, 2P, Assists, FT have positive relationships with scoring points. All the other variables the CI cross over zero so we don’t have any certainty that those variables help to score points.

Compared all the variables that can help score points.

Methods overall:

1. Normalise all the variables to minutes played and per game // intercept, slope and CI very similar between the two.
   1. Compared all these variables to points // Assist, 3P, 2P, FT are related to scoring the most amount of points.
   2. Compare all of these variables to winning.

Results

This criterion is linked to a learning outcomeReport - description of results

The results have been thoroughly and effectively described using both text and visual aids.  
  
Visualisations provide all relevant aspects required for the correct interpretation.

Recommendations

This criterion is linked to a learning outcomeReport - recommendations

The results of the analysis have been interpreted appropriately, and the recommendations made are supported by analytical evidence.  
  
The recommendations have been described and presented effectively and clearly.