**R Data analysis subject**

**RAW data methods for tidying and transforming TEAMS data frames**

1. Opening libraries

- tidyverse

- dplyr

- broom

1. Reading in raw files

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")

1. Tidying and transforming data
   1. Check the structure, head and tail for team1 and team 2 data
   2. Cleaning:
      1. Combining two team data frames together. Combining based on player and position, not including Age, Rank, or the 3 empty NA columns at the end
      2. Renaming variables (r renaming): changing all the variables that start with a number to having an ‘x’ in the front, and changing ‘%’ sign to ‘p’ , and change ‘/’ to \_per\_
2. Normalising team variables
   1. Normalising key variables by:
      1. ‘per game’: including 'FG' = Field goals, 'x3P' = 3-Point field goals, 'x2P' = 2-point field goals, 'FT' = Free throws, 'ORB' = Offensive rebounds, 'DRB' = Defensive rebounds, 'TRB' = Total rebounds, 'AST' = Assists, 'STL' = Steals, 'BLK' = Blocks, 'TOV' Turnovers. Each of these variables had the same starting code and added \_per\_game to the end of each new metric.
      2. ‘per minute played’ including 'FG' = Field goals, 'x3P' = 3-Point field goals, 'x2P' = 2-point field goals, 'FT' = Free throws, 'ORB' = Offensive rebounds, 'DRB' = Defensive rebounds, 'TRB' = Total rebounds, 'AST' = Assists, 'STL' = Steals, 'BLK' = Blocks, 'TOV' Turnovers. Each of these variables had the same starting code and added \_pmp to the end of each new metric
3. Checking the structure of the new data set and tail, head.

**RAW data methods for tidying and transforming PLAYER data frames**

1. Reading in raw files

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

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

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

1. Cleaning:
   1. Renaming variables (r renaming): changing all the variables that start with a number to having an ‘x’ in the front, and changing ‘%’ sign to ‘p’ , and change ‘/’ to \_per\_
   2. Summarise the multiple players on each row and put them on one row. And summarise each of the chosen stats
   3. Create a player position data frame and filter out multiple levels per player, and make sure there are only 5 single player position options.