```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------- tidyverse 1.2.1 --

## tibble  1.4.2     purrr   0.2.5
## tidyr   0.8.1     dplyr   0.7.6
## readr   1.1.1     forcats 0.3.0

## -- Conflicts ------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x readr::guess_encoding()  masks rvest::guess_encoding()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x purrr::pluck()           masks rvest::pluck()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```r
library(broom)
library(stringr)
library(knitr)
```

```r
nba_social_power <- read_csv("data/nba_2016_2017_100.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   PLAYER_NAME = col_character(),
##   TEAM_ABBREVIATION = col_character(),
##   W_PCT = col_double(),
##   MIN = col_double(),
##   OFF_RATING = col_double(),
##   DEF_RATING = col_double(),
##   NET_RATING = col_double(),
##   AST_PCT = col_double(),
##   AST_TO = col_double(),
##   AST_RATIO = col_double(),
##   OREB_PCT = col_double(),
##   DREB_PCT = col_double(),
##   REB_PCT = col_double(),
##   TM_TOV_PCT = col_double(),
##   EFG_PCT = col_double(),
##   TS_PCT = col_double(),
##   USG_PCT = col_double(),
##   PACE = col_double(),
##   PIE = col_double(),
##   FGM_PG = col_double()
##   # ... with 8 more columns
## )

## See spec(...) for full column specifications.
```

ERASE ALL INSTRUCTIONS BEFORE SUBMITTING THE PROPOSAL

Project Description

- What is the research question you wish to explore? *This includes an introduction to the subject matter you're investigating, the motivation for the for your research question (citing any relevant literature), and your hypotheses regarding the research question of interest.*
- Describe the dataset you wish to explore. *This includes describing the observations in the data, a general description of the variables, and how the data was originally collected (not how you found the data but how the original curator of the data collected it.)*

The data set we wish to explore includes on-court performance data for NBA players in the 2016-2017 season, along with their salary, Twitter engagement, and Wikipedia traffic data. Because we are examining the relationship between player stats and the number of twitter followers, we also filtered for players who had an active twitter account, by filtering for values where TWITTER_HANDLE is not n/a. There are 100 observations. The data was originally collected from ESPN, Basketball-Reference, Twitter, Five-ThirtyEight, and Wikipedia. We assume that basketball players in the dataset were not chosen randomly because while the creators did not specify how they chose which players to include, many big name basketball players are in the dataset. We assume the creators chose the best basketball players and the rest, other random smaller name basketball players. More information about the data set can be found at this link: https://www.kaggle.com/noahgift/social-power-nba While the dataset has 63 variables, we chose to only use 15 by selecting stats we thought were indicative of a player's performance and twitter followers because our research question attempts to predict the number of twitter followers. Below, we've listed the variable we're planning to use and a general description of the variables:

PLAYER_NAME: Player's name (categorical) TEAM_ABBREVIATION: Abbreviation for the team the player is on (categorical) AGE: Player age (quantitative) W_PCT: Percentage of games played won (quantitative) OFF_RATING: Player offensive rating calculated using the formula - Offensive Production Rating = (Points Produced / Individual Possessions) x OAPOW × PPG + FTM/FT * 3pt% + FG% (quantitative) DEF_RATING: Player defensive rating - Defensive Player Rating = (Players Steals*Blocks) + Opponents Differential= 1/5 of possessions - Times blown by + Deflections* OAPDW(Official Adjusted Players Defensive Withstand) (quantitative) NET_RATING: Average of the offensive/defensive rating (quantitative) AST_RATIO: Assists-to-turnovers ratio (quantitative) REB_PCT: Total rebounds (quantitative) USG_PCT: Usage percentage, an estimate of how often a player makes team plays (quantitative) PIE Player impact factor, a statistic roughly measuring a player's impact on the games that they play that's used by nba.com (quantitative) SALARY_MILLIONS: Salary in millions (quantitative) TWITTER_FOLLOWER_COUNT_MILLIONS: Number of Twitter followers (quantitative) TWITTER_HANDLE: Twitter handle (categorical) PTS: Points scored (quantitative)

The Data

Finally, we will examine the relevant variables in the dataset after removing players without twitter handles (since social power cannot be measured):

```r
glimpse(nba_social_power) # entire dataset
```

```
## Observations: 100
## Variables: 63
## $ PLAYER_ID                     <int> 201566, 1626246, 1627743, 2030...
## $ PLAYER_NAME                   <chr> "Russell Westbrook", "Boban Ma...
## $ TEAM_ID                       <int> 1610612760, 1610612765, 161061...
## $ TEAM_ABBREVIATION             <chr> "OKC", "DET", "BOS", "NOP", "H...
## $ AGE                           <int> 28, 28, 22, 24, 27, 28, 32, 32...
## $ GP                            <int> 81, 35, 5, 75, 81, 62, 74, 61,...
## $ W                             <int> 46, 16, 1, 31, 54, 51, 51, 43,...
## $ L                             <int> 35, 19, 4, 44, 27, 11, 23, 18,...
## $ W_PCT                           <dbl> 0.568, 0.457, 0.200, 0.413, 0....
## $ MIN                             <dbl> 34.6, 8.4, 3.4, 36.1, 36.4, 33...
## $ OFF_RATING                      <dbl> 107.9, 104.3, 124.2, 104.2, 11...
## $ DEF_RATING                      <dbl> 104.6, 102.4, 117.8, 102.5, 10...
## $ NET_RATING                      <dbl> 3.3, 1.9, 6.3, 1.7, 6.3, 16.0,...
## $ AST_PCT                         <dbl> 0.543, 0.054, 0.300, 0.110, 0....
## $ AST_TO                          <dbl> 1.92, 0.90, 0.00, 0.87, 1.95, ...
## $ AST_RATIO                       <dbl> 23.4, 5.1, 31.1, 7.3, 27.6, 18...
## $ OREB_PCT                        <dbl> 0.053, 0.166, 0.091, 0.067, 0....
## $ DREB_PCT                        <dbl> 0.279, 0.313, 0.118, 0.269, 0....
## $ REB_PCT                         <dbl> 0.167, 0.239, 0.103, 0.170, 0....
## $ TM_TOV_PCT                      <dbl> 12.2, 5.7, 0.0, 8.4, 14.1, 8.5...
## $ EFG_PCT                         <dbl> 0.476, 0.545, 0.875, 0.518, 0....
## $ TS_PCT                          <dbl> 0.554, 0.606, 0.753, 0.580, 0....
## $ USG_PCT                         <dbl> 0.408, 0.248, 0.172, 0.326, 0....
## $ PACE                            <dbl> 102.31, 97.20, 87.46, 100.19, ...
## $ PIE                             <dbl> 0.230, 0.196, 0.194, 0.192, 0....
## $ FGM                             <int> 824, 72, 3, 770, 674, 551, 736...
## $ FGA                             <int> 1941, 132, 4, 1526, 1533, 1026...
## $ FGM_PG                          <dbl> 10.2, 2.1, 0.6, 10.3, 8.3, 8.9...
## $ FGA_PG                          <dbl> 24.0, 3.8, 0.8, 20.3, 18.9, 16...
## $ FG_PCT                          <dbl> 0.425, 0.545, 0.750, 0.505, 0....
## $ GP_RANK                         <int> 18, 365, 458, 111, 18, 244, 12...
## $ W_RANK                          <int> 62, 345, 464, 196, 15, 21, 21,...
## $ L_RANK                          <int> 330, 149, 34, 419, 221, 79, 19...
## $ W_PCT_RANK                      <int> 143, 270, 468, 316, 56, 14, 45...
## $ MIN_RANK                        <int> 22, 432, 475, 9, 7, 40, 1, 64,...
## $ OFF_RATING_RANK                 <int> 115, 260, 2, 271, 27, 9, 18, 1...
## $ DEF_RATING_RANK                 <int> 168, 82, 475, 88, 296, 57, 287...
## $ NET_RATING_RANK                 <int> 95, 139, 53, 144, 55, 9, 45, 1...
## $ AST_PCT_RANK                    <int> 1, 412, 32, 230, 2, 79, 7, 4, ...
## $ AST_TO_RANK                     <int> 140, 381, 466, 387, 132, 89, 9...
## $ AST_RATIO_RANK                  <int> 96, 467, 28, 442, 48, 168, 68,...
## $ OREB_PCT_RANK                   <int> 171, 4, 81, 139, 242, 320, 220...
## $ DREB_PCT_RANK                   <int> 22, 8, 317, 27, 73, 55, 81, 21...
## $ REB_PCT_RANK                    <int> 53, 5, 201, 48, 144, 106, 125,...
## $ TM_TOV_PCT_RANK                 <int> 358, 30, 1, 129, 426, 131, 347...
## $ EFG_PCT_RANK                    <int> 339, 100, 2, 189, 167, 31, 32,...
## $ TS_PCT_RANK                     <int> 181, 52, 4, 112, 40, 12, 35, 3...
## $ USG_PCT_RANK                    <int> 2, 63, 279, 9, 7, 31, 15, 72, ...
## $ PACE_RANK                       <int> 47, 328, 483, 121, 34, 17, 230...
## $ PIE_RANK                        <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,...
## $ FGM_RANK                        <int> 1, 344, 461, 3, 10, 26, 4, 79,...
## $ FGA_RANK                        <int> 1, 360, 476, 5, 4, 35, 19, 78,...
## $ FGM_PG_RANK                     <int> 2, 306, 458, 1, 15, 9, 3, 51, ...
## $ FGA_PG_RANK                     <int> 1, 356, 480, 3, 9, 22, 15, 58,...
## $ FG_PCT_RANK                     <int> 293, 47, 3, 95, 253, 53, 45, 1...
## $ CFID                            <int> 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, ...
## $ CFPARAMS                        <dbl> 2.015662e+15, 1.626246e+16, 1....
## $ WIKIPEDIA_HANDLE                <chr> "Russell_Westbrook", "Boban_Ma...
## $ TWITTER_HANDLE                  <chr> "russwest44", "0", "d_jay11", ...
## $ SALARY_MILLIONS                 <dbl> 26.54, 7.00, 1.45, 22.12, 26.5...
## $ PTS                             <dbl> 31.6, 5.5, 2.0, 28.0, 29.1, 25...
## $ ACTIVE_TWITTER_LAST_YEAR        <int> 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ TWITTER_FOLLOWER_COUNT_MILLIONS <dbl> 4.500, 0.000, 0.049, 1.220, 4....
```

```
nba_social_power_mod <- nba_social_power %>%
  filter(TWITTER_HANDLE != "0") %>%
  select(PLAYER_NAME,
         TEAM_ABBREVIATION,
         AGE,
         W_PCT,
         OFF_RATING,
         DEF_RATING,
         NET_RATING,
         AST_RATIO,
         REB_PCT,
         USG_PCT,
         PIE,
         SALARY_MILLIONS,
         ACTIVE_TWITTER_LAST_YEAR,
         TWITTER_FOLLOWER_COUNT_MILLIONS,
         PTS)

glimpse(nba_social_power_mod) # modified dataset for analysis
```

```
## Observations: 95
## Variables: 15
## $ PLAYER_NAME                     <chr> "Russell Westbrook", "Demetriu...
## $ TEAM_ABBREVIATION               <chr> "OKC", "BOS", "NOP", "HOU", "G...
## $ AGE                             <int> 28, 22, 24, 27, 28, 32, 32, 26...
## $ W_PCT                           <dbl> 0.568, 0.200, 0.413, 0.667, 0....
## $ OFF_RATING                      <dbl> 107.9, 124.2, 104.2, 113.6, 11...
## $ DEF_RATING                      <dbl> 104.6, 117.8, 102.5, 107.3, 10...
## $ NET_RATING                      <dbl> 3.3, 6.3, 1.7, 6.3, 16.0, 7.7,...
## $ AST_RATIO                       <dbl> 23.4, 31.1, 7.3, 27.6, 18.4, 2...
## $ REB_PCT                         <dbl> 0.167, 0.103, 0.170, 0.123, 0....
## $ USG_PCT                         <dbl> 0.408, 0.172, 0.326, 0.341, 0....
## $ PIE                             <dbl> 0.230, 0.194, 0.192, 0.190, 0....
## $ SALARY_MILLIONS                 <dbl> 26.54, 1.45, 22.12, 26.50, 26....
## $ ACTIVE_TWITTER_LAST_YEAR        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, ...
## $ TWITTER_FOLLOWER_COUNT_MILLIONS <dbl> 4.500, 0.049, 1.220, 4.470, 16...
## $ PTS                             <dbl> 31.6, 2.0, 28.0, 29.1, 25.1, 2...
```