Project proposal

Statguys: Lewie Eatherton, Charlie Bonetti, Chris Yang

10/9/2020

Section 1. Introduction

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
## v tibble  3.0.3     v purrr   0.3.4
## v tidyr   1.1.1     v dplyr   1.0.1
## v readr   1.3.1     v forcats 0.5.0
## -- 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()
```

```
library(broom)
library(knitr)
```

```
shots <- read_csv("data/shot_logs.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   MATCHUP = col_character(),
##   LOCATION = col_character(),
##   W = col_character(),
##   GAME_CLOCK = col_time(format = ""),
##   SHOT_RESULT = col_character(),
##   CLOSEST_DEFENDER = col_character(),
##   player_name = col_character()
## )
```

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

For our STA 210 Final Project, our group is interested in investigating how certain variables may have an influence on a basketball players' shot efficiency in the National Basketball Association (NBA). We are all avid NBA fans and express great curiosity in what makes a good shot versus a bat shot. Thus, we are excited by the opportunity to explore whether or not certain factors exist which surprisingly have an effect on shot success versus other factors that surprisingly don't have an effect. An 2015 article titled "Basketball

Shot Types and Shot Success in Different Levels of Competitive Basketball" published by Frane Erčulj and Erik Štrumbel piqued our interest in this subject matter. We became intrigued by how different shot types affected shot success across different levels of competition. Additionally, it was insightful to learn that there were no discernable differences between different situational variables on shot type and shot success between levels. Since the effect of situation variables on shot success seemed to be constant throughout all levels of competition, we wanted to see what exactly the situation variables might be, and to what degree of influence they had on an individual's shot. As a result, we formulated our research question: "Do certain situational variables during a game have an effect on an NBA player's shot success?" Based on our prior knowledge of the game of basketball, we recognize that both players and coaches deem certain shots as "good shots" and some as "bad". Furthermore, through our own intuition after watching countless games and playing the sport ourselves, we hypothesize that some situational variables (e.g. shot distance or the distance of the closest defender) will have a greater effect on an NBA player's shot success than other situational variables.

Section 2. Data description

Data description:

Each observation in our data set is a unique shot that was taken in an NBA basketball game during the 2014-2015 season. Features of each observation include qualities such as the game location, matchup, player name, and shot result. There are a total of 128,069 observations within the data set. The response variable we want to record is shot percentage. This value determines the likelihood of a successful shot when all predictor values are taken into consideration. The data was originally collected via NBA's rest API, which scrapes historical data of past games from the official NBA website. All of the data was either collected by player tracking (via computer vision techniques) or by manual entry at an NBA scoring table (via humans).

Variables:

Shot number –> The shot number for that player in that specific game (8 would mean it was the player's 8th shot in that game) Shot_clock –> How many seconds were left on the shot clock when the player shot the basketball Shot_distance –> How far away the player was from the basketball when shooting Dribbles –> The number of dribbles the player took before shooting Touch_time –> How long the player touched the ball for before shooting (after being passed to) PTS_TYPE –> whether the shot was 2 or 3-pointer Closest_Defender –> Who the closest defender was to the shooting player Close_Def_Distance –> How far away said player was when the ball was shot Player_name –> Who shot the ball FGM –> Whether the shot went in or not (response variable)

Citations:

Erčulj F, Štrumbelj E (2015) Basketball Shot Types and Shot Success in Different Levels of Competitive Basketball. PLoS ONE 10(6): e0128885. https://doi.org/10.1371/journal.pone.0128885

Section 3. Glimpse of data

```
glimpse(shots)
```

```
## Rows: 128,069
## Columns: 21
## $ GAME_ID                   <dbl> 21400899, 21400899, 21400899, 21400899, ...
## $ MATCHUP                   <chr> "MAR 04, 2015 - CHA @ BKN", "MAR 04, 201...
## $ LOCATION                  <chr> "A", "A", "A", "A", "A", "A", "A", "A", ...
## $ W                         <chr> "W", "W", "W", "W", "W", "W", "W", "W", ...
## $ FINAL_MARGIN              <dbl> 24, 24, 24, 24, 24, 24, 24, 24, 24, 1, 1...
## $ SHOT_NUMBER               <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 2, 3, 4, 1...
## $ PERIOD                    <dbl> 1, 1, 1, 2, 2, 2, 4, 4, 4, 2, 2, 4, 4, 4...
## $ GAME_CLOCK                <time> 01:09:00, 00:14:00, 00:00:00, 11:47:00,...
## $ SHOT_CLOCK                <dbl> 10.8, 3.4, NA, 10.3, 10.9, 9.1, 14.5, 3....
## $ DRIBBLES                  <dbl> 2, 0, 3, 2, 2, 2, 11, 3, 0, 0, 8, 14, 2,...
```

```
## $ TOUCH_TIME                <dbl> 1.9, 0.8, 2.7, 1.9, 2.7, 4.4, 9.0, 2.5, ...
## $ SHOT_DIST                 <dbl> 7.7, 28.2, 10.1, 17.2, 3.7, 18.4, 20.7, ...
## $ PTS_TYPE                  <dbl> 2, 3, 2, 2, 2, 2, 2, 2, 3, 3, 3, 2, 2, 3...
## $ SHOT_RESULT               <chr> "made", "missed", "missed", "missed", "m...
## $ CLOSEST_DEFENDER          <chr> "Anderson, Alan", "Bogdanovic, Bojan", "...
## $ CLOSEST_DEFENDER_PLAYER_ID <dbl> 101187, 202711, 202711, 203900, 201152, ...
## $ CLOSE_DEF_DIST            <dbl> 1.3, 6.1, 0.9, 3.4, 1.1, 2.6, 6.1, 2.1, ...
## $ FGM                       <dbl> 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0...
## $ PTS                       <dbl> 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 2, 2, 0...
## $ player_name               <chr> "brian roberts", "brian roberts", "brian...
## $ player_id                 <dbl> 203148, 203148, 203148, 203148, 203148, ...
```