STAT 345 Midterm Project

Due March 31

```
library(devtools)
```

```
## Loading required package: usethis
```

```
devtools::install_github("abresler/nbastatR", force = TRUE)
```

```
## Downloading GitHub repo abresler/nbastatR@HEAD
```

```
##
## -- R CMD build -----------------------------------------------------------------
##      checking for file '/private/var/folders/qy/s8n6lb591n38q4jsqllm1vyw0000gn/T/RtmpItRyNa/remotes2
##   -  preparing 'nbastatR':
##      checking DESCRIPTION meta-information ...  v  checking DESCRIPTION meta-information
##   -  checking for LF line-endings in source and make files and shell scripts
##   -  checking for empty or unneeded directories
##      Omitted 'LazyData' from DESCRIPTION
##   -  building 'nbastatR_0.1.152.tar.gz'
##
##
```

```
## Warning in i.p(...): installation of package
## '/var/folders/qy/s8n6lb591n38q4jsqllm1vyw0000gn/T//RtmpItRyNa/file240d6b2a9ae3/nbastatR_0.1.152.tar.g
## had non-zero exit status
```

```
tinytex::install_tinytex(force = TRUE)
```

```
## The directory /usr/local/bin is not writable. I recommend that you make it writable. See https://git
```

```
## tlmgr install tlgpg
```

```
## tlmgr update --self
```

```
## tlmgr install tlgpg
```

```
## tlmgr --repository http://www.preining.info/tlgpg/ install tlgpg
```

```
## tlmgr option repository 'https://mirror.math.princeton.edu/pub/CTAN/systems/texlive/tlnet'
```

```
## tlmgr update --list
```

```r
Sys.setenv("VROOM_CONNECTION_SIZE" = 131072 * 10000)
team_2010 <- teams_shots (teams= "Chicago Bulls", seasons =2010)
```

```
## Warning: 'funs()' was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with 'tibble::lst()': tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## i The deprecated feature was likely used in the nbastatR package.
##   Please report the issue at <https://github.com/abresler/nbastatR/issues>.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## Chicago Bulls 2009-10 shot data
```

```r
team_2012 <- teams_shots (teams= "Chicago Bulls", seasons =2012)
```

```
## Chicago Bulls 2011-12 shot data
```

```r
team_2014 <- teams_shots (teams= "Chicago Bulls", seasons =2014)
```

```
## Chicago Bulls 2013-14 shot data
```

```r
team_2016 <- teams_shots (teams= "Chicago Bulls", seasons =2016)
```

```
## Chicago Bulls 2015-16 shot data
```

```r
team_2018 <- teams_shots (teams= "Chicago Bulls", seasons =2018)
```

```
## Chicago Bulls 2017-18 shot data
```

```r
team_2020 <- teams_shots (teams= "Chicago Bulls", seasons =2020)
```

```
## Chicago Bulls 2019-20 shot data
```

```r
shots3 <- rbind(team_2010,team_2012,team_2014,team_2016,team_2018,team_2020)
court + geom_point(data=shots3, aes(x=locationX,y=locationY, color=zoneRange, shape = typeEvent),alpha=
  scale_color_brewer(palette = "RdGy")# + theme_null()
```
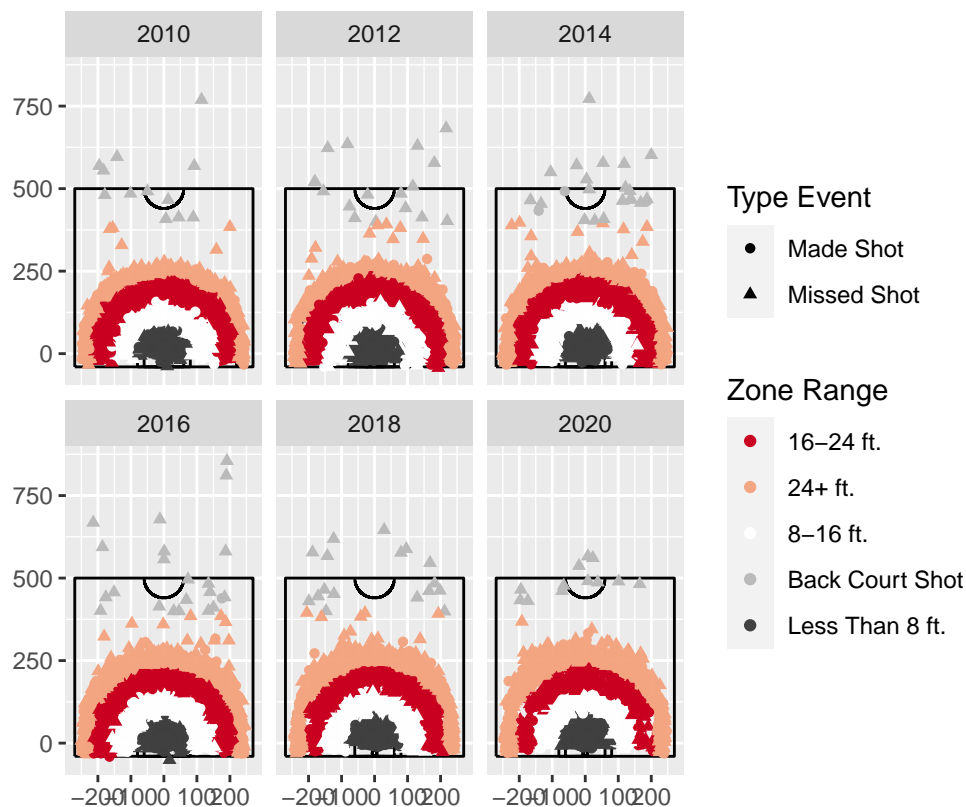
Report:

Six minature graphics of the shots made and missed

-The court graphic looks at the team shots of the Chicago Bulls from six seasons: 2010, 2012, 2014, 2016, 2018, and 2020.

-The shots made is identified by a circle and the shots missed is identified by a triangle.

-From where on the court the player made the shot is represented by RdGy theme coloring. The different zone ranges and colors are identified in the legend made for the courts graphic.

-As the years go on, it appears that the players are shooting closer toward the hoop. The light red region on each graphic (the more than 24 feet), appears to get more dense as the years proceed.

-The missed shots seem to get closer to the hoop as the years go on.

The first court graphic looks at six different seasons of the Chicago Bulls team shots. The seasons used in this graphic are 2010, 2012, 2014, 2016, 2018, and 2020.By creating this graphic, it shows team shots and if they changed throughout the years for any trends. Overall, looking for changes in distance regarding players shooting throughout the years and how that impacted made and missed shots.From the court graphic, it appears that the players are shooting closer as the seasons go on. The more than 24 feet gets more dense as the years go on. The amount of back court shots steadily increases then approaches a decrease again as the years go on.

```
#Plots of one season of the Chicago Bulls
# Chicago Bulls #red
# Chicago Bulls is #black
first_graph <- ggplot(data=team_2014, aes(x = locationX, y=locationY - 400)) + # the points were brough
  geom_point(alpha = 0.2, aes(color = team_2014$typeEvent)) +  # changing opacity and color
```

```
    scale_color_manual(values = c("red","black")) + # setting colors to Chicago Bulls
    labs(x = "", y = "", color = "Type of Shot") +
    xlim(-300,300) +
    ylim(-450,0) +
    ggtitle("Chicago Bulls In 2014")
first_graph
```
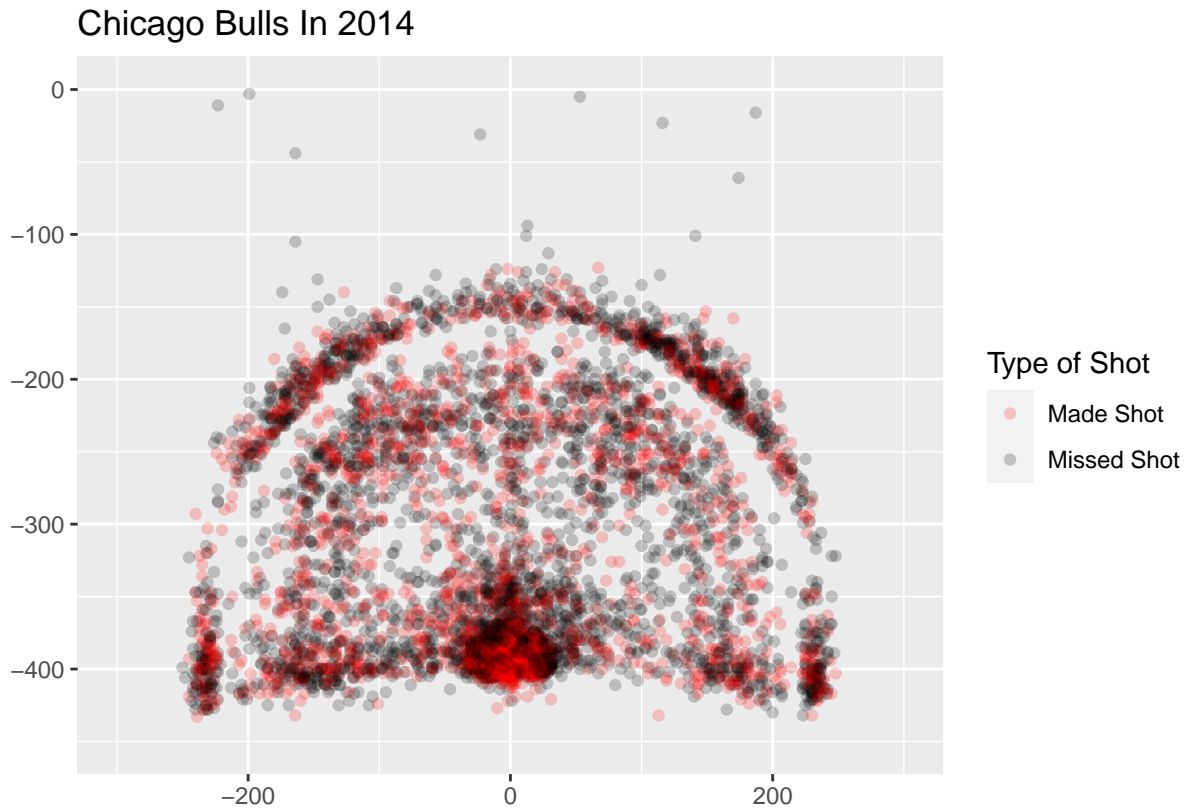

Report:

Huge graph of one of the years of made and missed shots

-The Graph shows Chicago Bulls made and missed shots in the year 2020.

-Made shots are the red points.

-Missed shots are the black points.

-The greater density areas of red and black depicts locations where the number of shots taken were more heavy.

-The three point line and in front of the hoop seems to have the highest density of made and missed shots.

-The graph is not too significant in showing where the highest amount of made and missed shots are.

```
df <- data.frame(read.csv('~/Downloads/shot_logs.csv'))
str(df)
```

```
## 'data.frame':    128069 obs. of  21 variables:
##  $ GAME_ID                    : int  21400899 21400899 21400899 21400899 21400899 21400899 21400899 2
```

4

```
##  $ MATCHUP                  : chr  "MAR 04, 2015 - CHA @ BKN" "MAR 04, 2015 - CHA @ BKN" "MAR 04, 20
##  $ LOCATION                 : chr  "A" "A" "A" "A" ...
##  $ W                        : chr  "W" "W" "W" "W" ...
##  $ FINAL_MARGIN             : int  24 24 24 24 24 24 24 24 24 1 ...
##  $ SHOT_NUMBER              : int  1 2 3 4 5 6 7 8 9 1 ...
##  $ PERIOD                   : int  1 1 1 2 2 2 4 4 4 2 ...
##  $ GAME_CLOCK               : chr  "1:09" "0:14" "0:00" "11:47" ...
##  $ SHOT_CLOCK               : num  10.8 3.4 NA 10.3 10.9 9.1 14.5 3.4 12.4 17.4 ...
##  $ DRIBBLES                 : int  2 0 3 2 2 2 11 3 0 0 ...
##  $ TOUCH_TIME               : num  1.9 0.8 2.7 1.9 2.7 4.4 9 2.5 0.8 1.1 ...
##  $ SHOT_DIST                : num  7.7 28.2 10.1 17.2 3.7 18.4 20.7 3.5 24.6 22.4 ...
##  $ PTS_TYPE                 : int  2 3 2 2 2 2 2 2 3 3 ...
##  $ SHOT_RESULT              : chr  "made" "missed" "missed" "missed" ...
##  $ CLOSEST_DEFENDER         : chr  "Anderson, Alan" "Bogdanovic, Bojan" "Bogdanovic, Bojan" "Brown,
##  $ CLOSEST_DEFENDER_PLAYER_ID: int  101187 202711 202711 203900 201152 101114 101127 203486 202721 20
##  $ CLOSE_DEF_DIST           : num  1.3 6.1 0.9 3.4 1.1 2.6 6.1 2.1 7.3 19.8 ...
##  $ FGM                      : int  1 0 0 0 0 0 0 1 0 0 ...
##  $ PTS                      : int  2 0 0 0 0 0 0 2 0 0 ...
##  $ player_name              : chr  "brian roberts" "brian roberts" "brian roberts" "brian roberts"
##  $ player_id                : int  203148 203148 203148 203148 203148 203148 203148 203148 203148 20
```

summary(df)

```
##      GAME_ID          MATCHUP            LOCATION              W
##  Min.   :21400001   Length:128069      Length:128069      Length:128069
##  1st Qu.:21400233   Class :character   Class :character   Class :character
##  Median :21400449   Mode  :character   Mode  :character   Mode  :character
##  Mean   :21400452
##  3rd Qu.:21400673
##  Max.   :21400908
##
##   FINAL_MARGIN       SHOT_NUMBER         PERIOD        GAME_CLOCK
##  Min.   :-53.0000   Min.   : 1.000   Min.   :1.000   Length:128069
##  1st Qu.: -8.0000   1st Qu.: 3.000   1st Qu.:1.000   Class :character
##  Median :  1.0000   Median : 5.000   Median :2.000   Mode  :character
##  Mean   :  0.2087   Mean   : 6.507   Mean   :2.469
##  3rd Qu.:  9.0000   3rd Qu.: 9.000   3rd Qu.:3.000
##  Max.   : 53.0000   Max.   :38.000   Max.   :7.000
##
##    SHOT_CLOCK        DRIBBLES         TOUCH_TIME         SHOT_DIST
##  Min.   : 0.00    Min.   : 0.000   Min.   :-163.600   Min.   : 0.00
##  1st Qu.: 8.20    1st Qu.: 0.000   1st Qu.:   0.900   1st Qu.: 4.70
##  Median :12.30    Median : 1.000   Median :   1.600   Median :13.70
##  Mean   :12.45    Mean   : 2.023   Mean   :   2.766   Mean   :13.57
##  3rd Qu.:16.68    3rd Qu.: 2.000   3rd Qu.:   3.700   3rd Qu.:22.50
##  Max.   :24.00    Max.   :32.000   Max.   :  24.900   Max.   :47.20
##  NA's   :5567
##    PTS_TYPE      SHOT_RESULT        CLOSEST_DEFENDER
##  Min.   :2.000   Length:128069      Length:128069
##  1st Qu.:2.000   Class :character   Class :character
##  Median :2.000   Mode  :character   Mode  :character
##  Mean   :2.265
##  3rd Qu.:3.000
##  Max.   :3.000
```

```
##
##  CLOSEST_DEFENDER_PLAYER_ID CLOSE_DEF_DIST       FGM                PTS
##  Min.   :   708             Min.   : 0.000  Min.   :0.0000  Min.   :0.0000
##  1st Qu.:101249             1st Qu.: 2.300  1st Qu.:0.0000  1st Qu.:0.0000
##  Median :201949             Median : 3.700  Median :0.0000  Median :0.0000
##  Mean   :159038             Mean   : 4.123  Mean   :0.4521  Mean   :0.9973
##  3rd Qu.:203079             3rd Qu.: 5.300  3rd Qu.:1.0000  3rd Qu.:2.0000
##  Max.   :530027             Max.   :53.200  Max.   :1.0000  Max.   :3.0000
##
##  player_name           player_id
##  Length:128069     Min.   :   708
##  Class :character  1st Qu.:101162
##  Mode  :character  Median :201939
##                    Mean   :157238
##                    3rd Qu.:202704
##                    Max.   :204060
##
```

```r
histogram <- hist(df$SHOT_DIST, plot=FALSE)
shot_counts <- histogram$counts
made_shots <- subset(df, FGM == 1)
made_hist <- hist(made_shots$SHOT_DIST, plot=FALSE)
made_counts <- made_hist$counts
breaks <- made_hist$breaks
distance_breaks <- head(breaks, length(breaks) - 1)
percs <- made_counts / shot_counts
barplot(shot_counts, names.arg=distance_breaks, main="Shot Frequency Using Distance", col="Black", xlab
```

Shot Frequency Using Distance
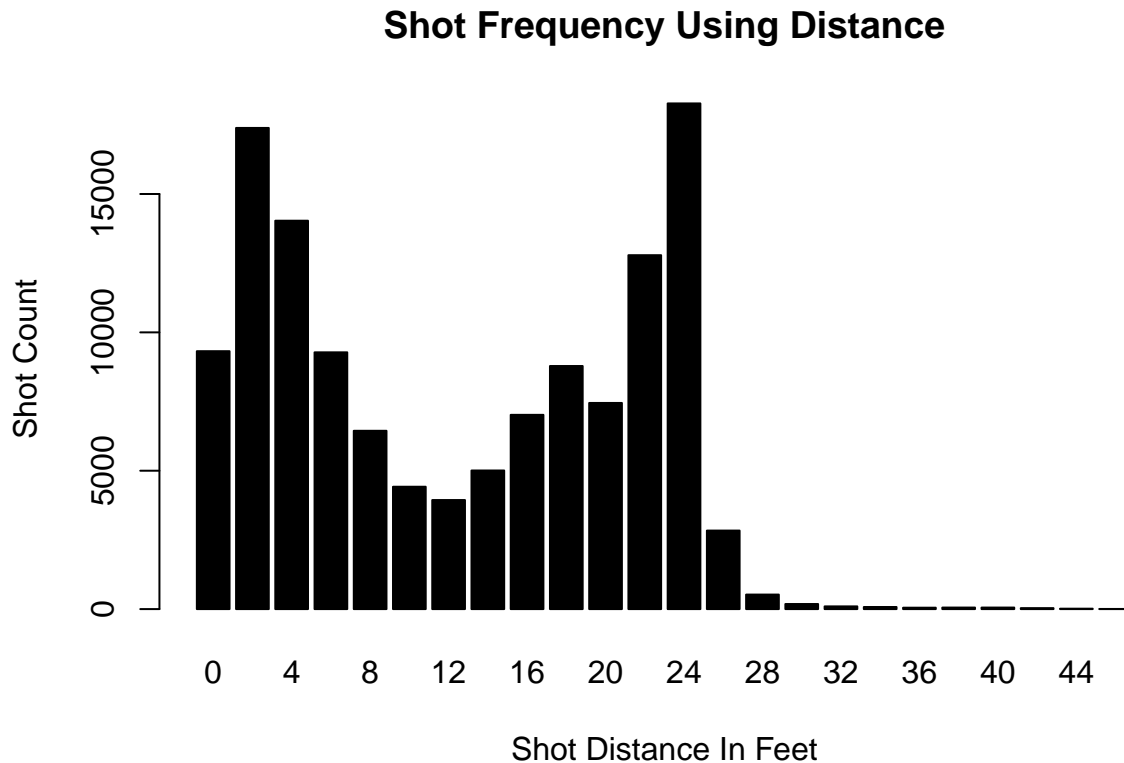
Report:

Shot Frequency Using Distance Histogram

-The NBA has record of shot logs containing every shot taken in 2014-2015 regular season, which was downloaded through Kaggle -Source URL:https://www.kaggle.com/datasets/dansbecker/nba-shot-logs

-The data set contained an array of 21 variables, however, the graphs only pull data from shot distance and shot counts.

-The first barplot shows shot distance in feet on the x-axis and number of shots made from those area on the y-axis.

-The most shots were taken at 2 feet and 24 feet away. The 24 feet is expected because the average 3 point line is 22 feet and that is close to there.

-After the 24 feet there is a huge drop in distance regarding shots being thrown from there.

This graphic is showing the shot distance in feet and amount of time players shot from those locations. It does not correlate with the top graphs, but it is significant because the big graph I created for Chicago Bulls 2014 data is included within this one. Those players, alongside many other teams who competed in the NBA can be looked at to see what there shot counts look like depending on distance when they play.