MA615_Final Project

Jiahao Xu

1.Abstract

In this project, I will analyze how 3 Point Field Goals Attempted and 3 Point Field Goals Percentage will improve Win Rate for each NBA Team, and whether more 3 Point Field Goals Attempted and higher 3 points percentage will increase/decrease the win share for each NBA player in different positions. Also, I will test both Team and Player data whether they follow Benford's rule. If not, I will do the normality test.

2.Introduction

2.1 Background

In 2014–15, Curry won the NBA Most Valuable Player Award and led the Warriors to their first championship since 1975. The following season, he became the first player in NBA history to be elected MVP by a unanimous vote and to lead the league in scoring while shooting above 50–40–90. That same year, the Warriors broke the record for the most wins in an NBA season. We could not imagine how crazy Golden State Warriors and Curry's three-pointer were. Curry nailed more 3s than everyone on the Bucks combined in 2016. Since the traditional basketball philosophy is that the closer you get to the basket, the easier it is to score, Curry and his Golden State Warriors definitely created a new era that NBA teams began to shot more and more 3-points instead of midrange Medium and long range 2-points jump shot.

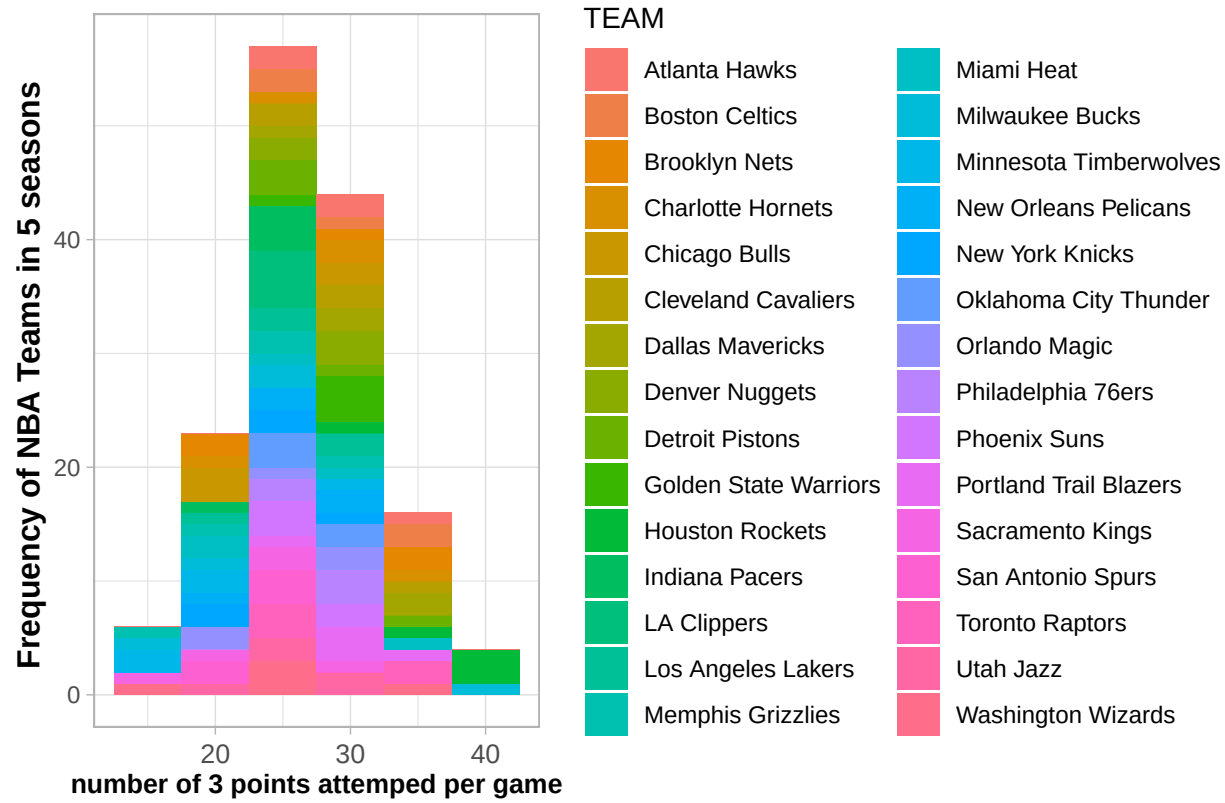
3.1 Data source

"player_data.csv" and "Players.csv" are players' season data kaggle, which was scraped from Basketball Reference. For EDA and model analysis, I only select 5 seasons' Team and Player data, since 2o14-2015 season was the first season that players in NBA were more likely to shoot a 3 than a mid-range jump shot.

3.2 Exploratory Data Analysis (EDA)

PS: Figure 1-9 are all from Team's perspective, Figure 10 is from Player's perspective.

gure 1. 3 points attemped Distribution



From figure 1, we can see that the data sample of 3 points attemped per game of NBA Teams in last 5 seasons is approximately normally distrubuted. The mean value is about 25 3-points attemped per game. And Goldent State Warriors did not unexpectedly shot the most three points, more than 40 3-points attemped per game.

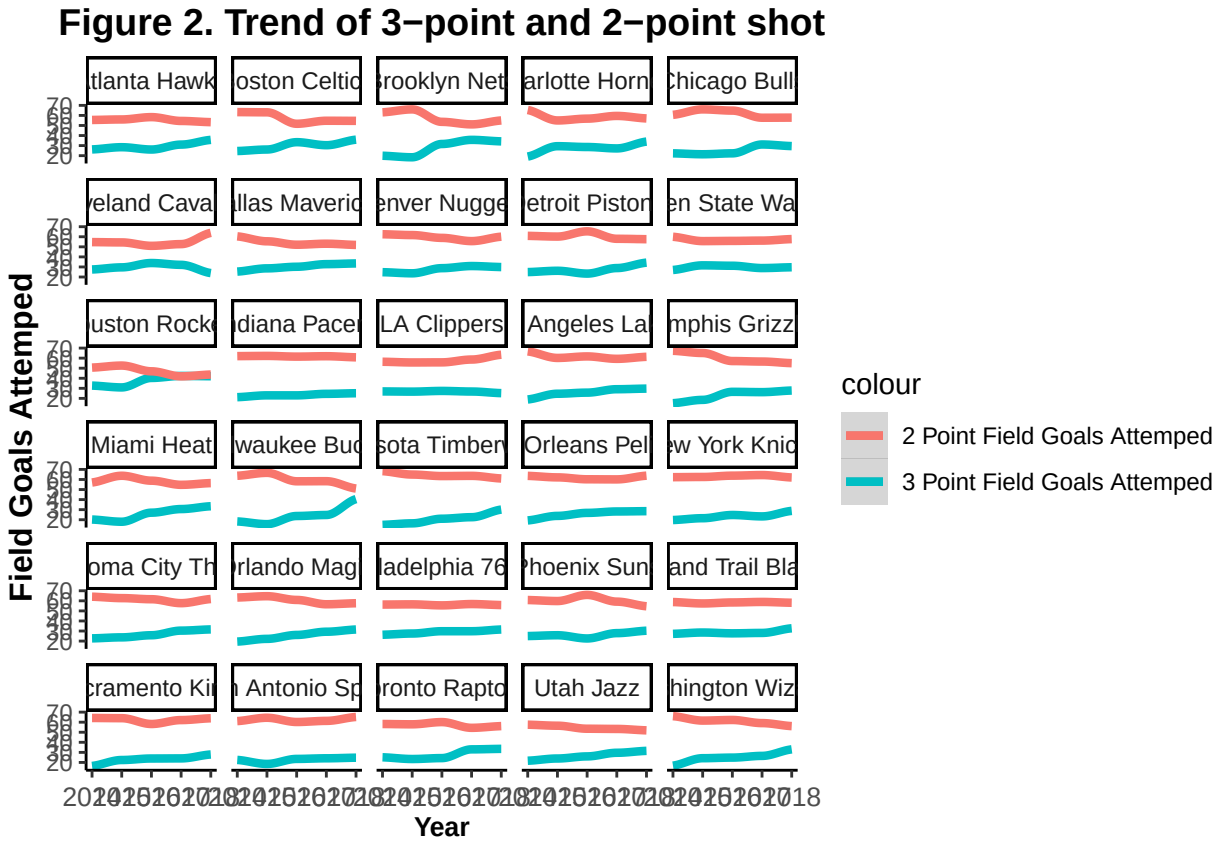
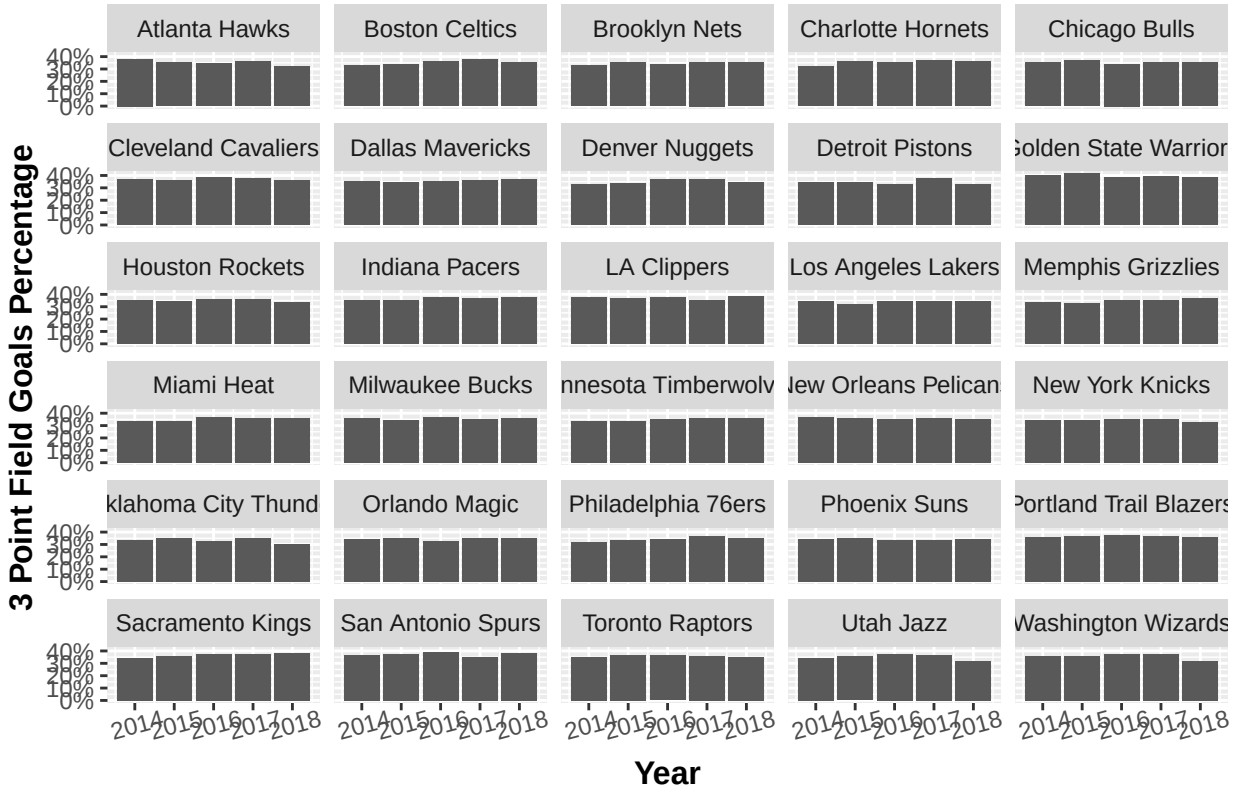


Figure 2 is a facet plot by Team to show the trend of 3 Point Field Goals Attemped and 2 Point Field Goals Attemped. From the plot, generally most NBA Team have slight decrease or keep the same in 2 point Field Goals Attemped. Only Cleveland Cavaliers and LA Clippers had obvious increase in in 2 point Field Goals Attemped. And it is obvious that most Team have increase in 3 Point Field Goals Attemped, especially Houston Rocket and Milwaukee Bucks. It is unbelievable that Houson Rocket had the same 2 point Field Goals Attemped and 3 point Field Goals Attemped last season!

Figure 3. 3−Point Field Goals Made change by year

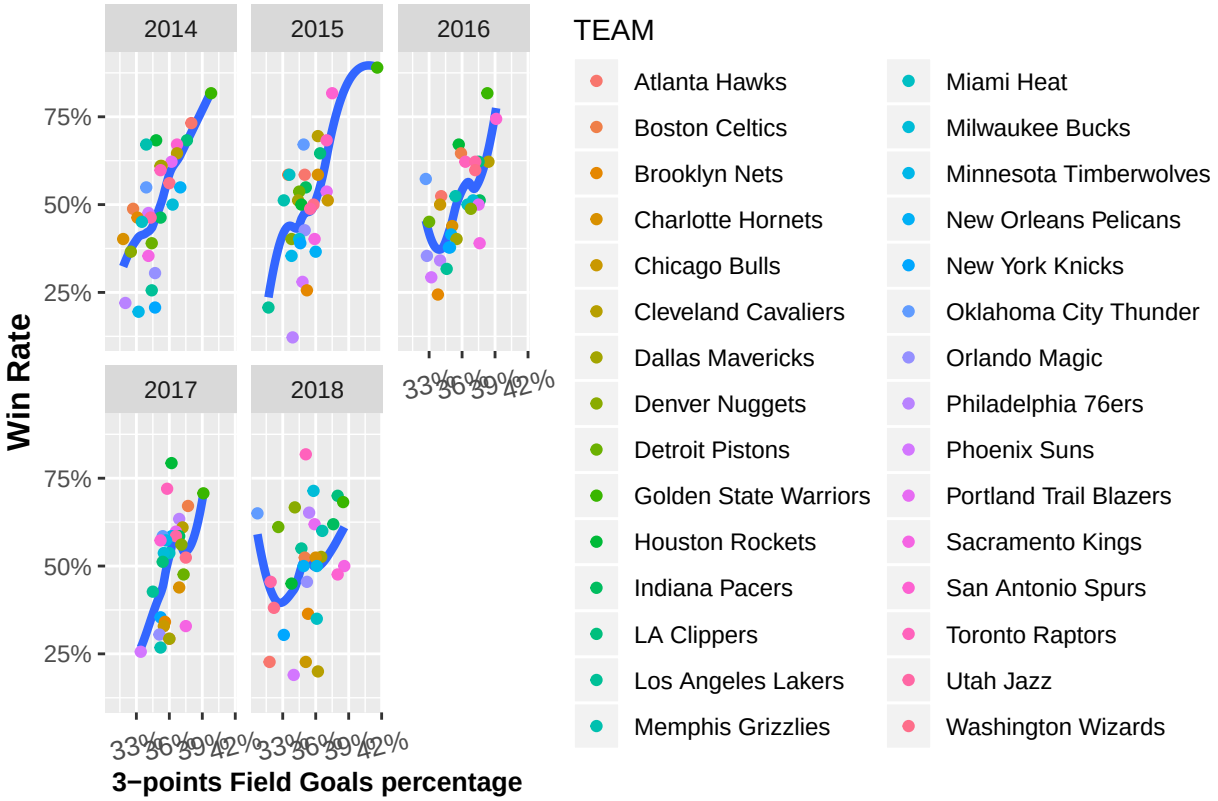
Figure 4. 3−Point Field Goal Percentage change by year



According to Figure3 and Figure4, we can figure out that most team had increase in 3 Point Field Goals

Made per game last 5 seasons. Because 3 Point Field Goals Made=3 Point Field Goals Attemped * 3 Point Goals Percentage, most team had increase in 3 Point Field Goals Attemped per game last 5 seasons, and their 3 Point Goals Percentage did not have obvious changes.

igure 5. 3 points Field Goals percentage vs. Win Rate



According to Figure 5, a facet plot by year of the relationship between 3 points Field Goals percentage and Win Rate, we can see that basically higher 3-points percentage will lead to higher Win Rate, except this season.Because so far, it only processed like 20-30 games in 2018-2019 season, the trend of this season is not so reliable.

Figure 6. 3−Point Field Goals Made vs Win Percentage

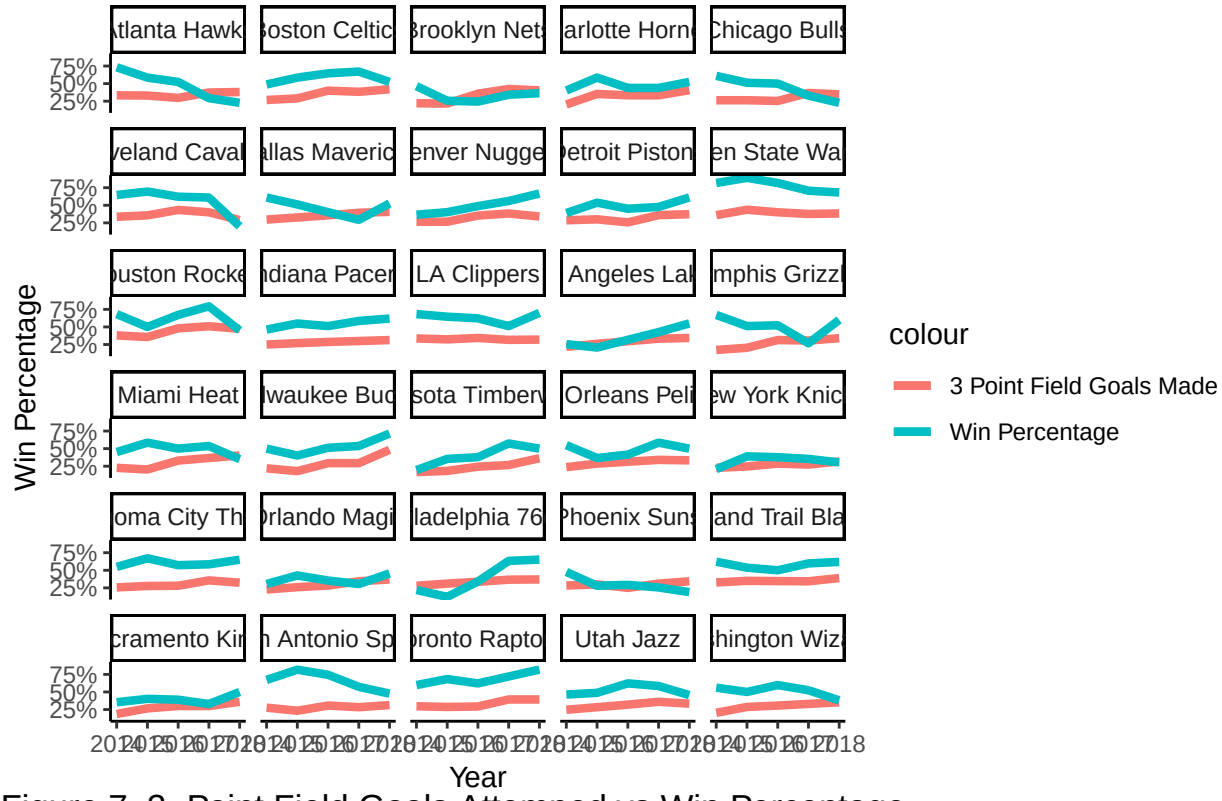
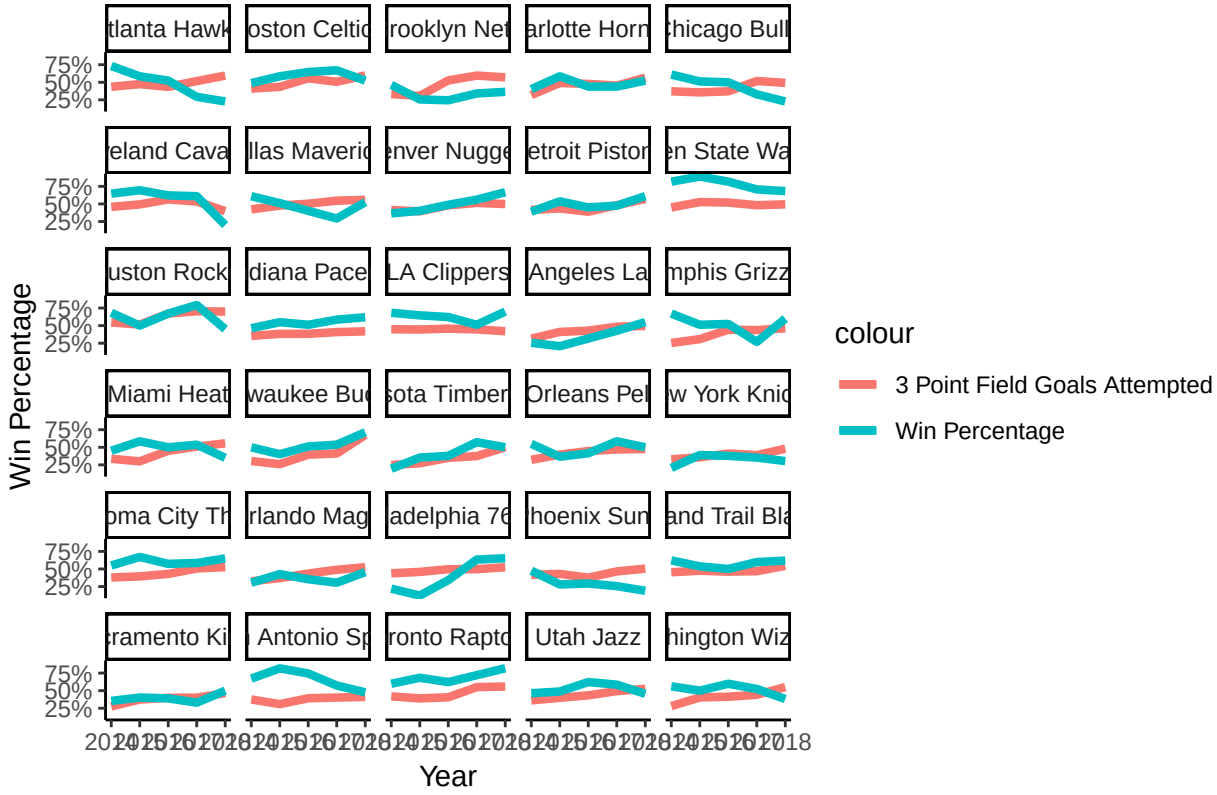
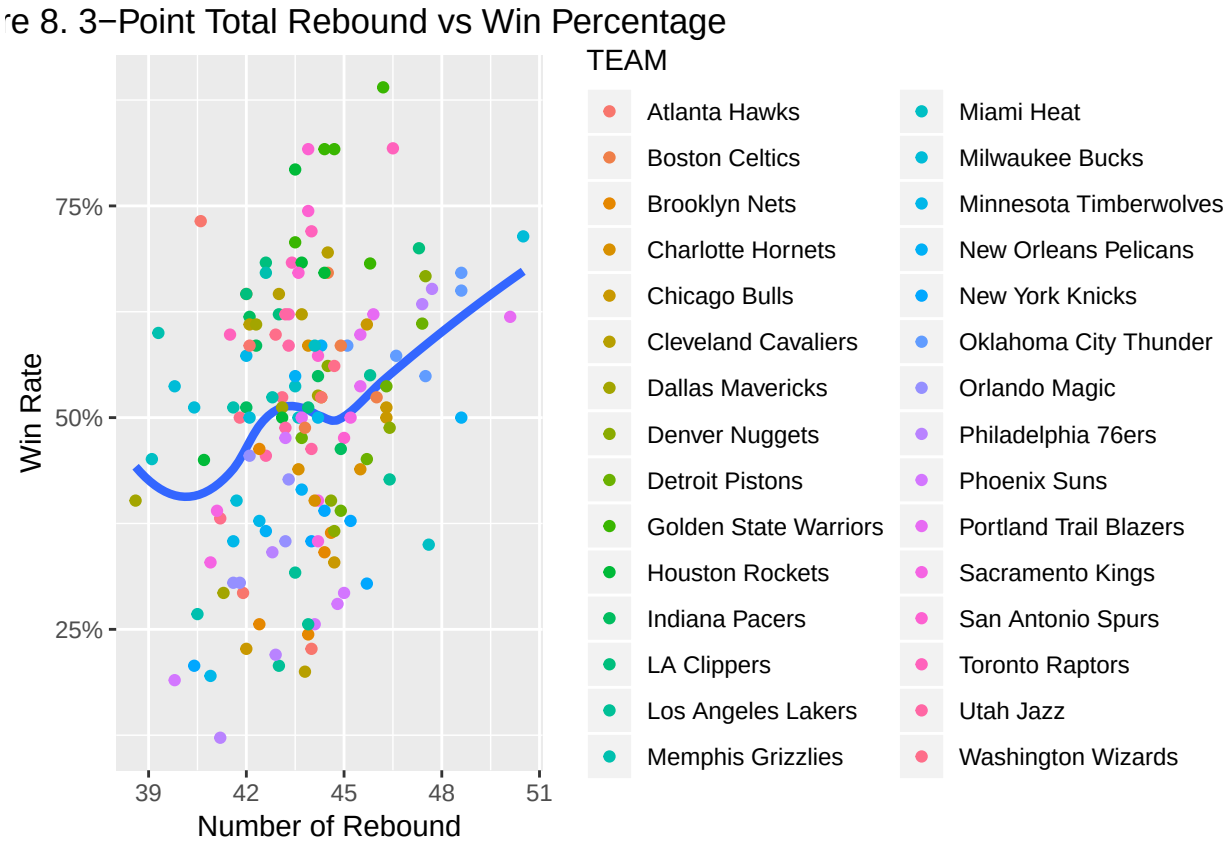
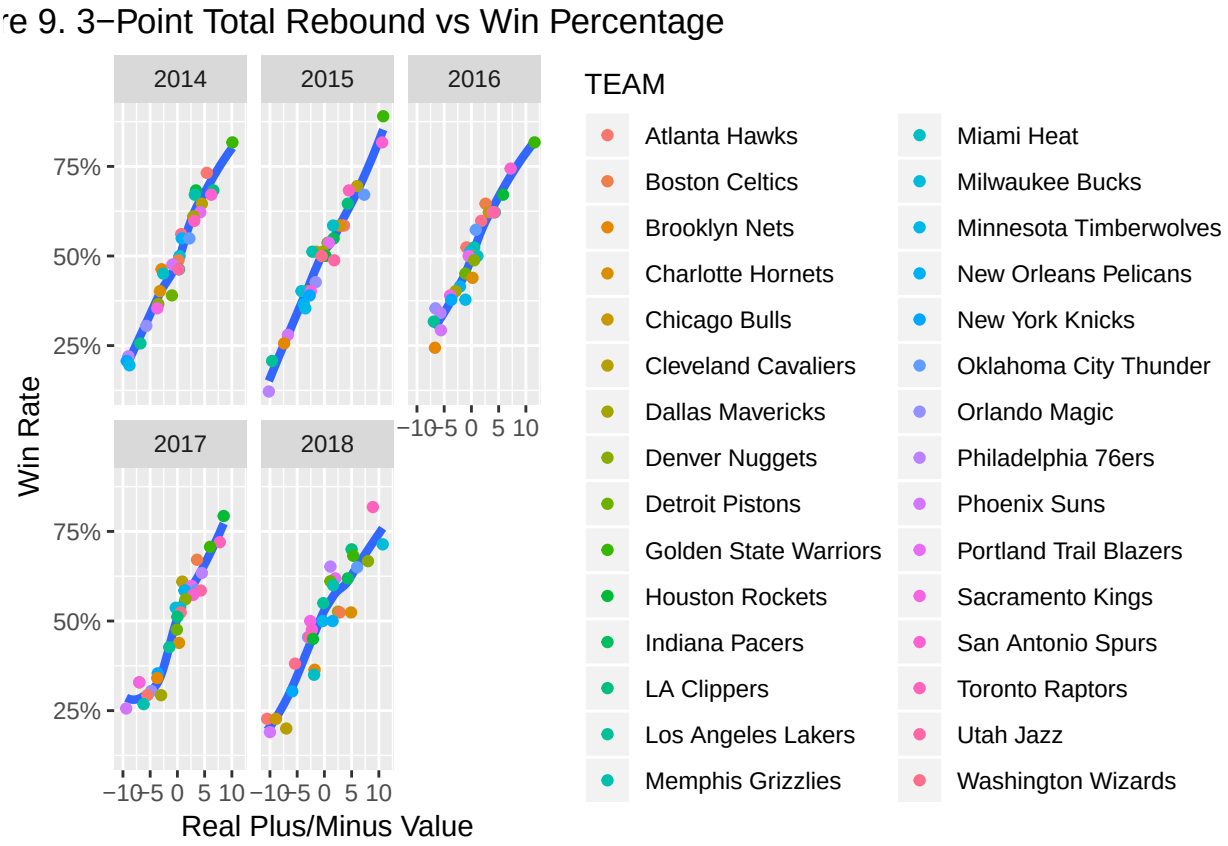


Figure 7. 3−Point Field Goals Attemped vs Win Percentage





According to figure 6 and figure 7, we can realize that the increasing of both 3-point Field Goals Attempted and Made will have obvious positive/nagetive influences to different teams. figure 8 tells us that the higher rebound, the higher win rate: just like what famous basketball player Takenori Akagi in "Slam Dunk" said, "If you master the rebound, you will master the game".



According to Figure 9, Real Plus/Minus Value has obvious positive relationship with Win Rate: higher Real Plus/Minus Value leads to higher win rate.

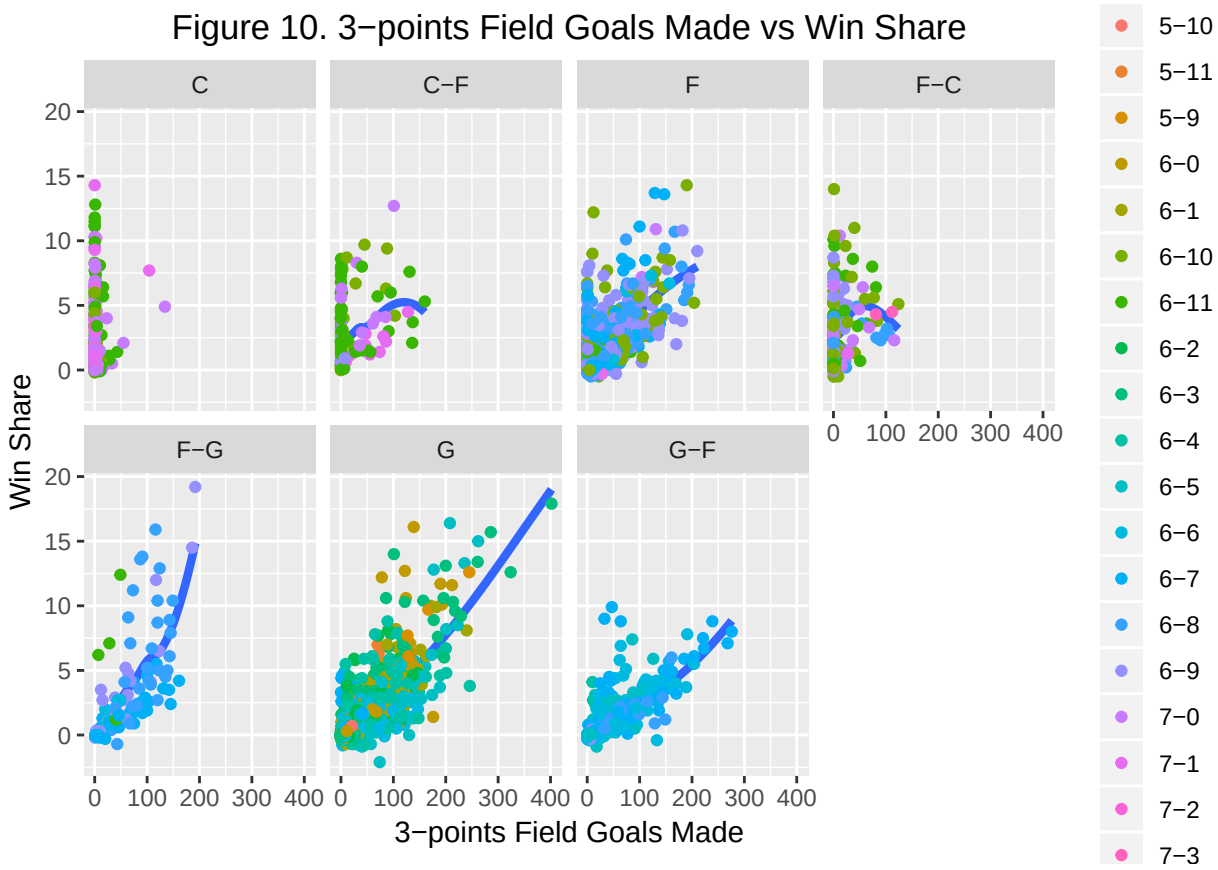
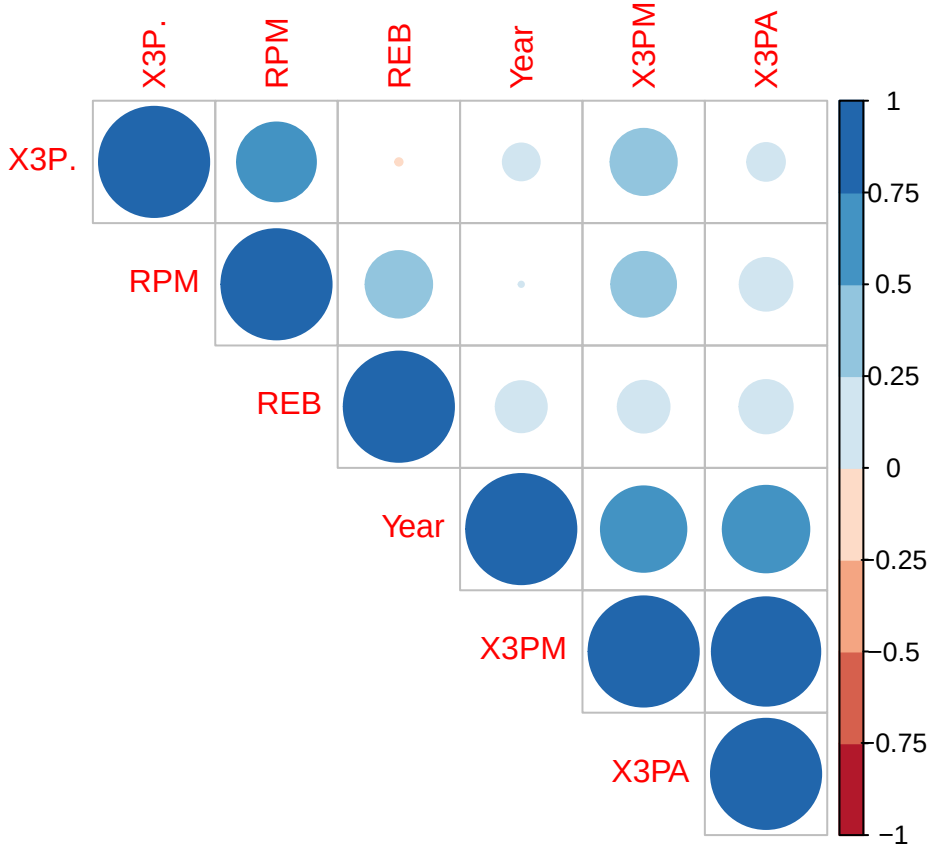


Figure 10 demonstrates the relationship between 3-points Field Goals Mad and Win Share of each player in different height and different position. From the plot, it shows that there is strong positive relationships between 3PM and WS in F-G, G, G-F and F these four positions. In position Center, almost no influences by 3PM.

3.3 Correlation Check

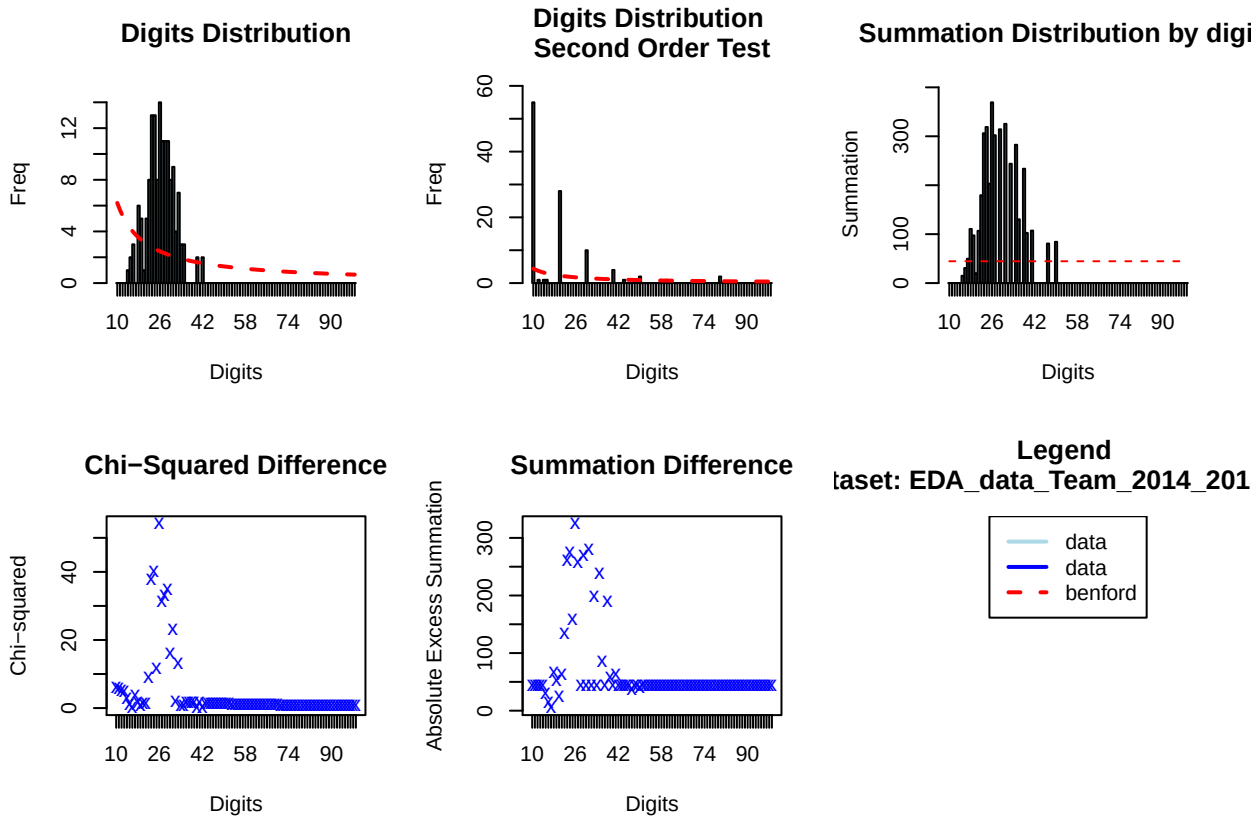
According to this correlation plot, it is obvious that X3PM, X3P., X3PA and RPM have very strong correlations.



4.1 benford analysis For Team Data

According to the benford analysis for Team data, we can easily find out that 3-points Field Goals by different Teams do not follow Benford. It is approximately normally distributed.

```
benford1 <- benford(EDA_data_Team_2014_2018$X3PA,number.of.digits = 2)
plot(benford1)
```



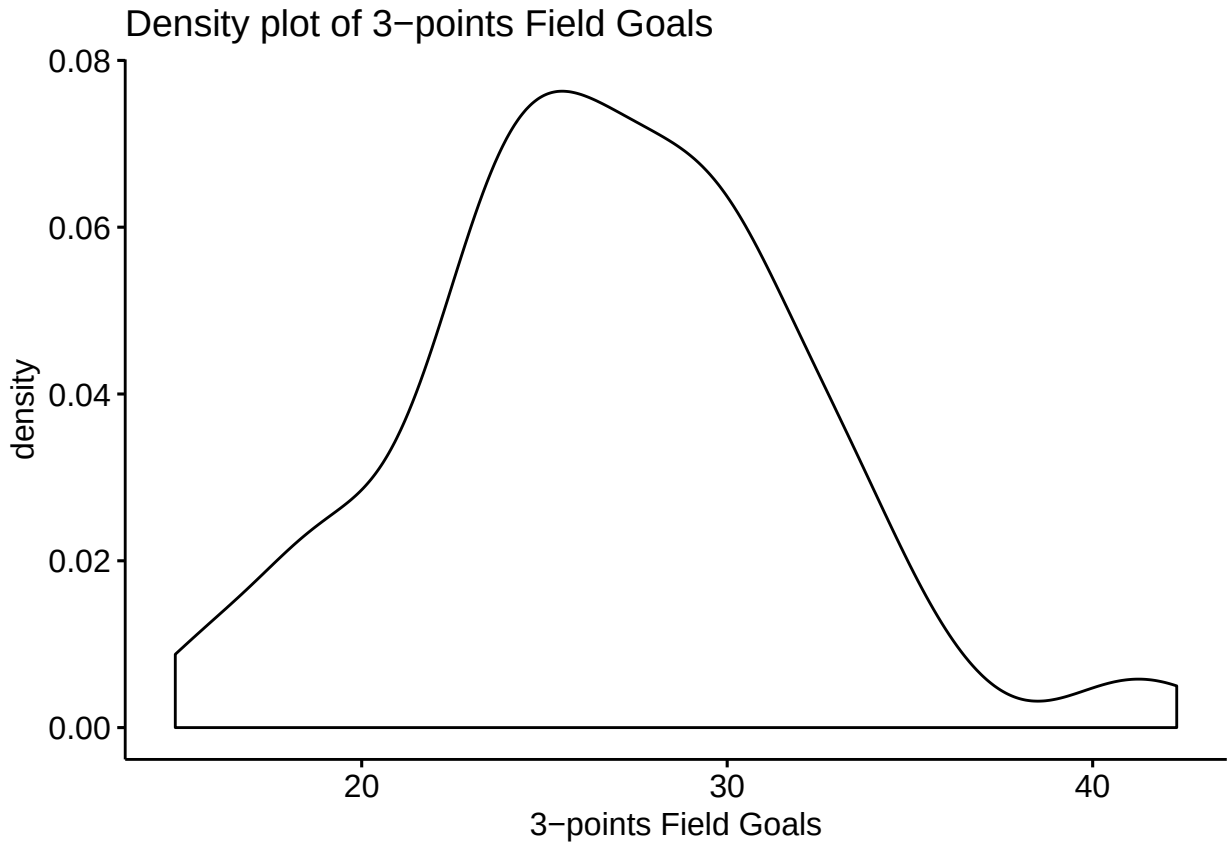
From the normality test result, the p-value > 0.05 implying that the distribution of the data are not significantly different from normal distribution. In other words, we can assume the normality.

```
# Testing for normality
shapiro.test(EDA_data_Team_2014_2018$X3PA)
```

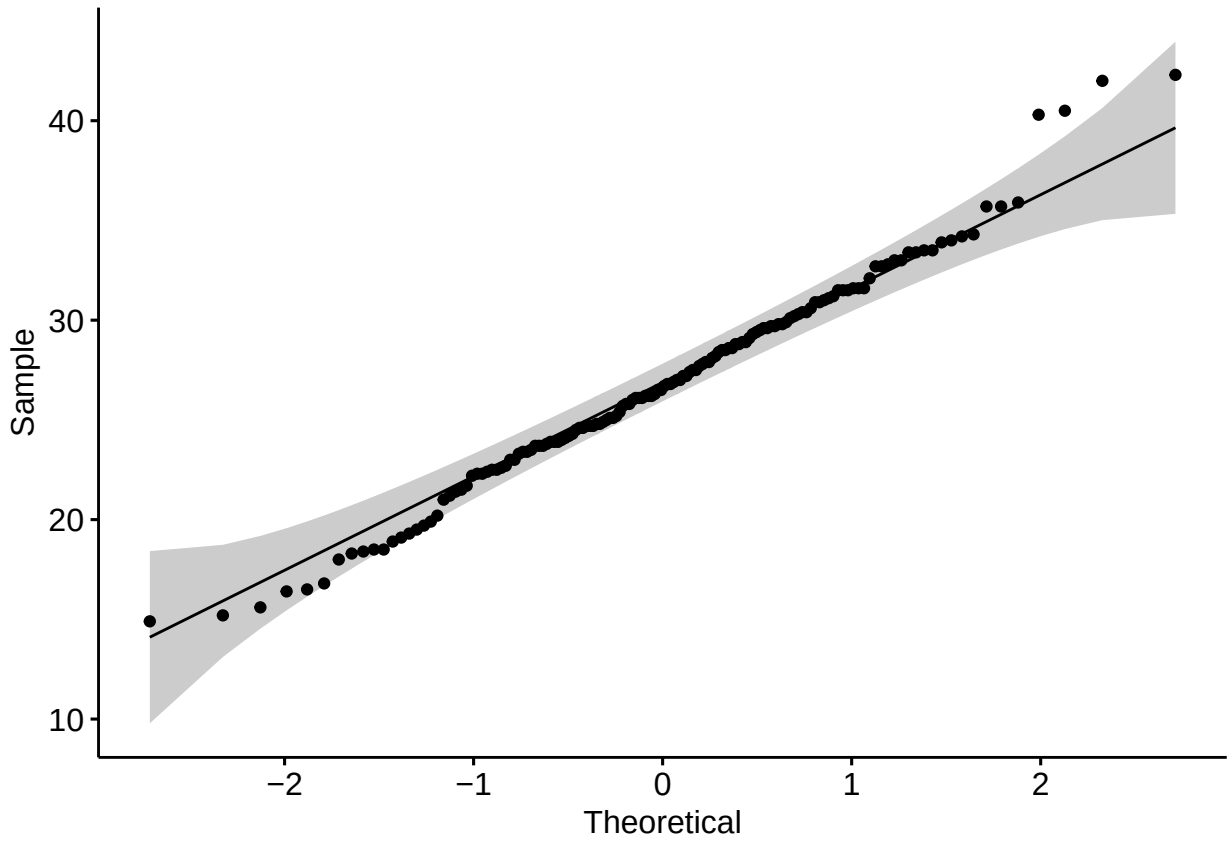
```
##
##  Shapiro-Wilk normality test
##
## data:  EDA_data_Team_2014_2018$X3PA
## W = 0.9875, p-value = 0.1979
```

Density plot to test normality and qqplot to test normality

```
#Density plot to test normality
ggdensity(EDA_data_Team_2014_2018$X3PA,
          main = "Density plot of 3-points Field Goals",
          xlab = "3-points Field Goals")
```



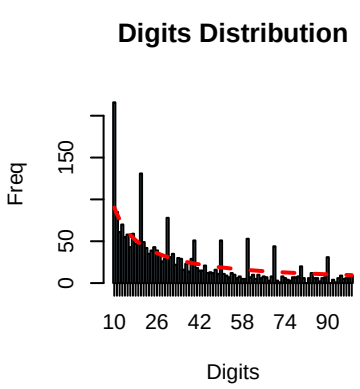

```
#qqplot to test normality
ggqqplot(EDA_data_Team_2014_2018$X3PA)
```

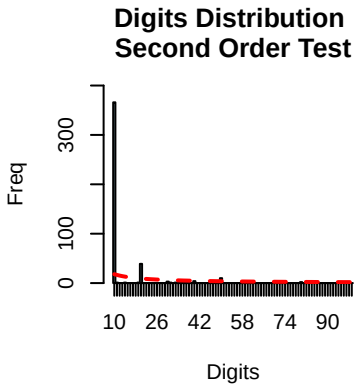


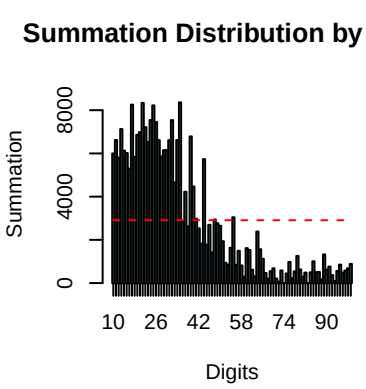
4.2 Benford analysis For Player Data

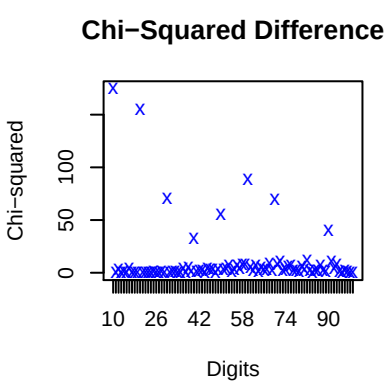
According to the benford analysis for Player data, we can easily find out that 3-points Field Goals by different Players generally follow Benford's rule, except some points like 10,20,30, and etc.Then we need to find the suspect table for these points.

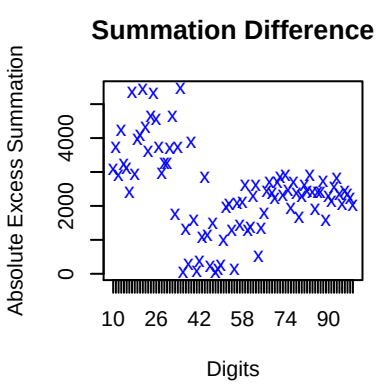
```
benford2 <- benford(EDA_data_Player_2014_2018$X3PA,number.of.digits = 2)
plot(benford2)
```

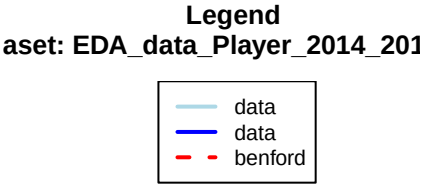













```
head(suspectsTable(benford2),10) #prints the digits by decreasing order of discrepancies
```

```
##     digits absolute.diff
##  1:     10     125.63977
##  2:     20      84.74376
##  3:     30      46.91312
##  4:     60      37.32915
##  5:     50      32.22583
##  6:     70      30.55205
##  7:     40      27.58980
##  8:     90      20.52404
##  9:     12      14.88568
## 10:     16      14.47607
```

```
duplicatesTable(benford2)#prints the duplicates by decreasing order
```

```
##       number duplicates
##    1:      1        140
##    2:      2         86
##    3:      3         52
##    4:      6         39
##    5:      7         38
##   ---
## 433:    377          1
## 434:     72          1
## 435:    384          1
## 436:    379          1
## 437:    123          1
```

```r
chisq(benford2)#gets the Chi-squared test
```

```
##
##  Pearson's Chi-squared test
##
## data:  EDA_data_Player_2014_2018$X3PA
## X-squared = 923.77, df = 89, p-value < 2.2e-16
```