

Objective - Examine Validity of Some “NBA Myths”:

Hot Hand Analysis

Analyzing Home Court Advantage

Performance During Regular vs Playoff Games

Project Conclusion:

# Analyzing NBA Myths

Code ▼

Samantha Kuang

12/4/2021

## Objective - Examine Validity of Some “NBA Myths”:

Many factors affect a NBA game. To understand their impacts, three different data sets will be utilized. These data points enable analysis of varying levels from examining player to team performance. Results of these descriptive analysis will measure the validity of certain “NBA Myths”. The following questions will guide our analysis.

### Questions:

1. Shooting
  - a. Are players more likely to make shots from a particular location e.g. corner three.
  - b. **Hot Hand Analysis** - the belief that a player can get “hot” and starts making every shot attempted
  - c. Warriors are known as “elite” shooters. How much farther do they shoot from compare to the league avg?
2. **Home Court Advantage** - the belief that teams perform better at home then at away games.
  - a. Is there a location that’s more or less challenging for teams to play at? For example, some believe that high elevation in Arizona gives the Nugget an advantage.
3. Regular season do not matter - the belief that elite players do not play with full effort for regular season games.

### Data Sets and Sources Information:

Conclusions from the following analyses are only as accurate as the data collected. This project assumes that the data sets being used are accurate reflections of the true historical NBA data.

1. **shots** : used python to extract from **sports\_radar** (unknown reliability)
  - Get NBA schedule for the Warrior’s last four games: ‘2021-11-30’, ‘2021-12-03’, ‘2021-12-04’, ‘2021-12-06’
  - Provided game\_id in the **playperplay database** .
  - has information about every shot attempted in all the games played in those 4 dates.
2. **allnbagames** : used python to extract from the **rapidapi website** . (unknown reliability)
  - It has information about every NBA game played between season year 2017 to 2020 inclusive.
  - “https://api-nba-v1.p.rapidapi.com/games/seasonYear/ (https://api-nba-v1.p.rapidapi.com/games/seasonYear/)

3. **stephcurry\_df** : from <https://data.world/datatouille/stephen-curry-stats>  
(<https://data.world/datatouille/stephen-curry-stats>) (unknown reliability)
  - Has Steph Curry's box score information for every game between 2009 - 2017

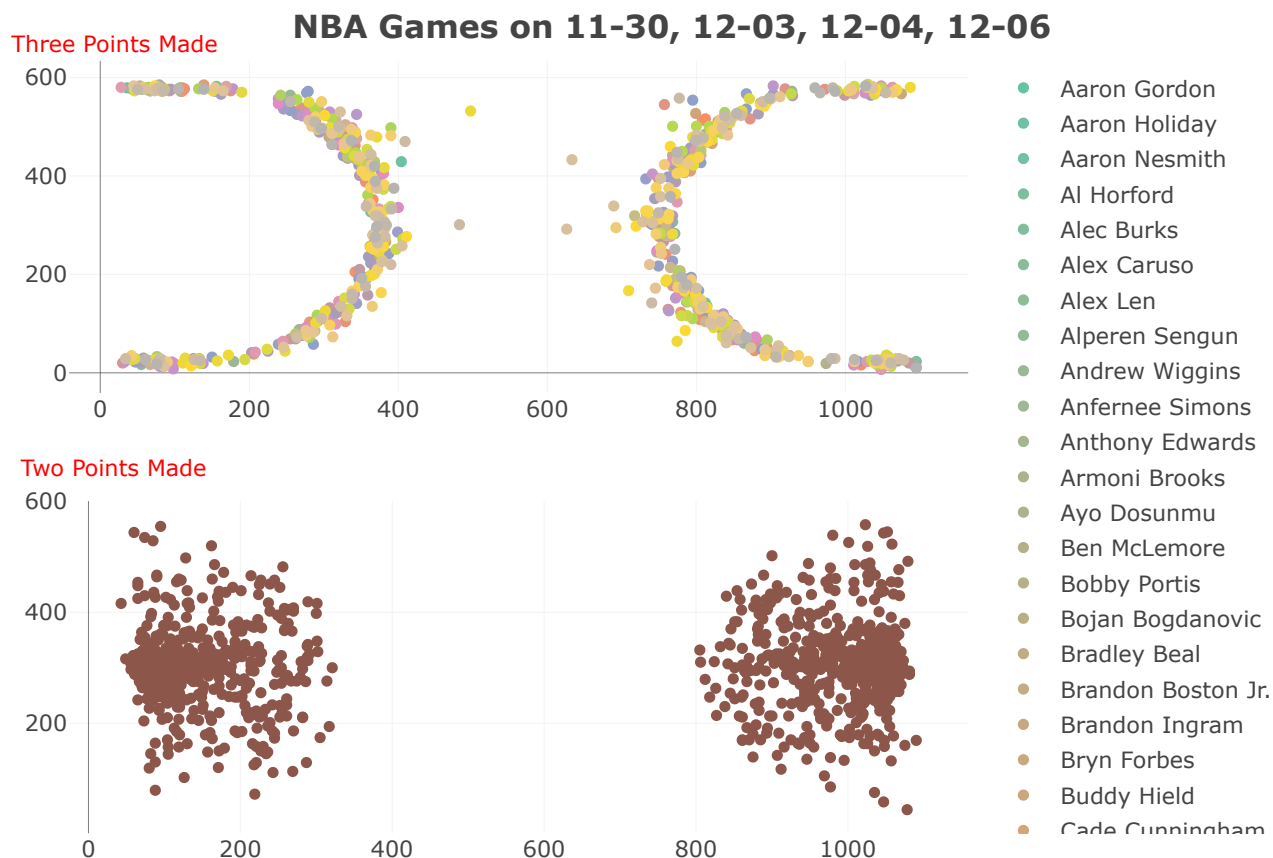
# Hot Hand Analysis

## Data Ingestion and Wrangling

In this step, data is read in via `read.csv()`. Then additional columns are added to the shots dataset for enriching the data to enable ease of use for analysis. Also created a subset of the shots data to analyze the Warriors team.

## X and Y Coordinates of Made Shots

NBA court size is 94 X 50 feet. 3 point line is 23ft and 9 inches. Graphically visual of all points made given the timeline of our data set.

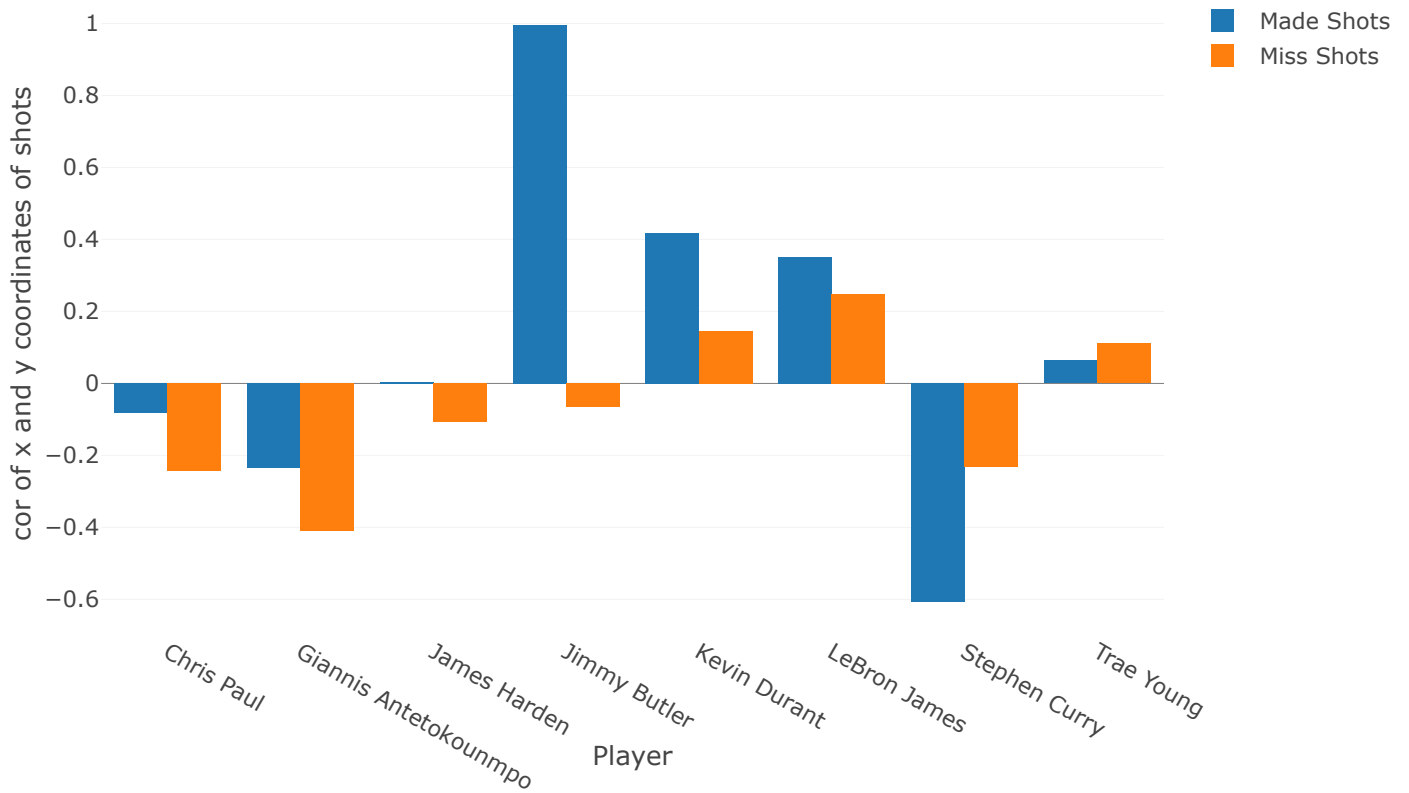


## Spatial Correlation of Top Players

Any correlations between the x and y distance of their successful shoots?

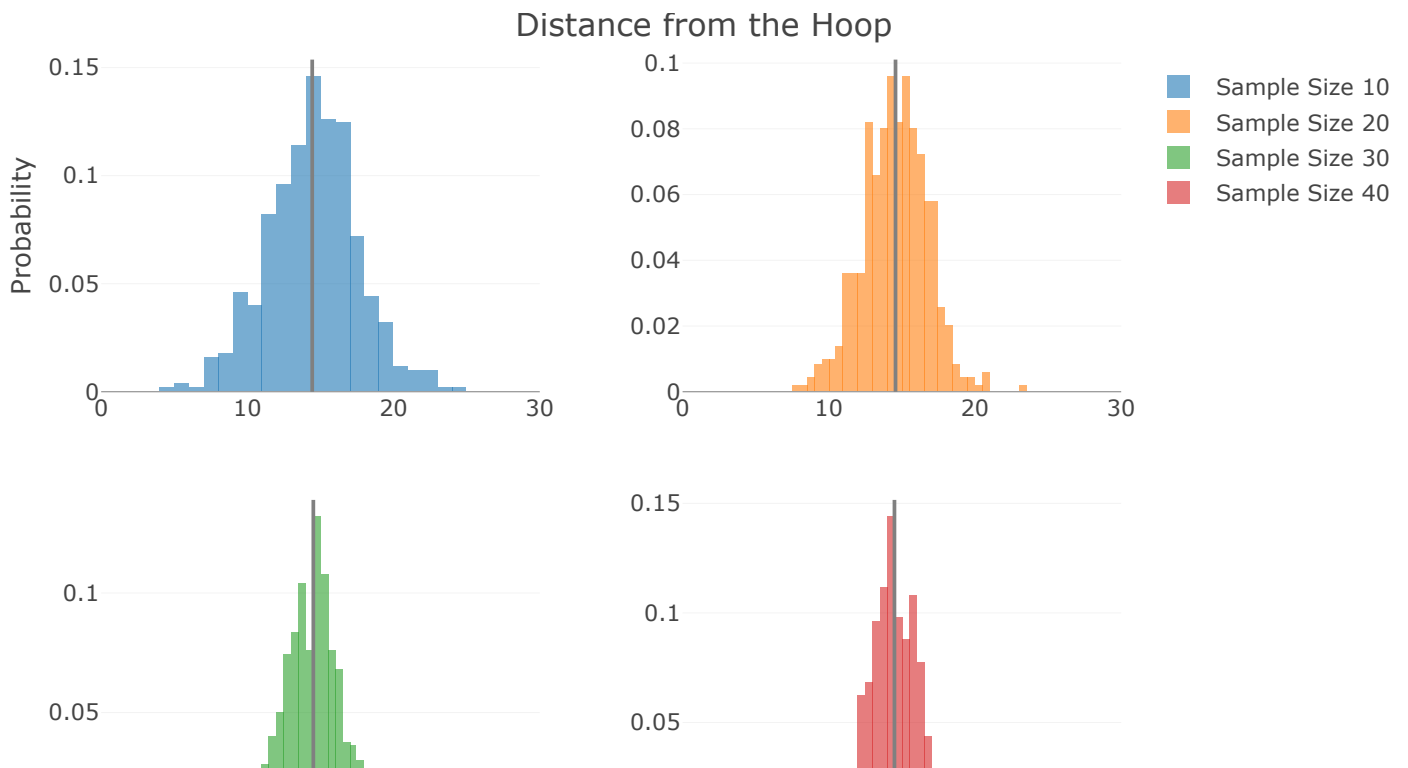
1. "Giannis Antetokounmpo"
2. "Kevin Durant"
3. "Chris Paul"
4. "Jimmy Butler"
5. "Trae Young"
6. "James Harden"
7. "LeBron James"
8. "Stephen Curry"

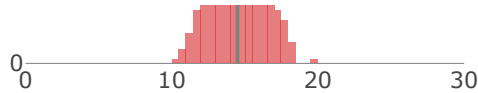
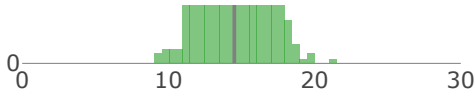
## Correlation of X and Y Location of Top Player Shots



1. This suggests that Stephen Curry takes shots where x and y are negatively correlated. Shots taken around the three point line have this effect because equation of a circle is  $(x - h)^2 + (y - k)^2 = r^2$ , and radius stays constant.
2. Jimmy Butler makes a lot of his shots by the rim -so x and y are highly correlated.

## Central Limit Theorem - Mean Shooting Distance.





```
## The Mean Shooting Distance for the League is 14.4783 feet
```

```
## The Median Shooting Distance for the League is 14.215 feet
```

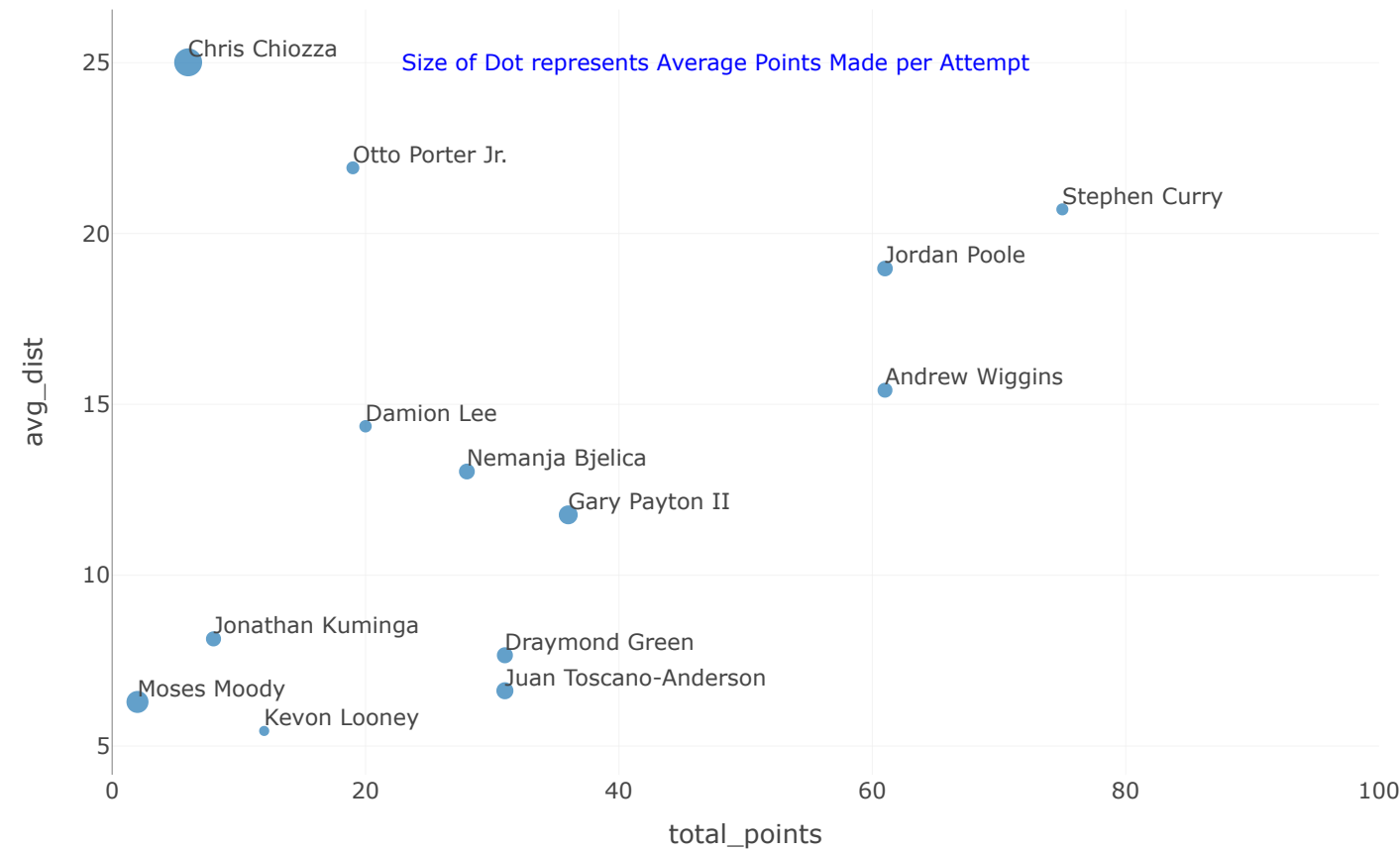
```
## sample size = 10  Mean = 14.4353 SD = 3.1277
## sample size = 20  Mean = 14.564  SD = 2.2072
## sample size = 30  Mean = 14.5105 SD = 1.8753
## sample size = 40  Mean = 14.4919 SD = 1.5672
```

The Central Limit Theorem states that as sample size increases, the sample mean will become closer to the true population mean. Furthermore, the sample variance of the sample mean will decrease and the distribution will become normally distributed. As a result, confidence in the sample mean reflecting the true population mean increases. Graphically, this is seen above. The spread of the histograms decreases (smaller variance) as sample size increases. Furthermore, the population mean is 14.49 feet, the differences decreases as sample size increases from 10 to 40.

## Examining Warriors Performance

Q. How much farther can the Warriors shoot compare to the league?

Warriors' Players - Performance Statistics



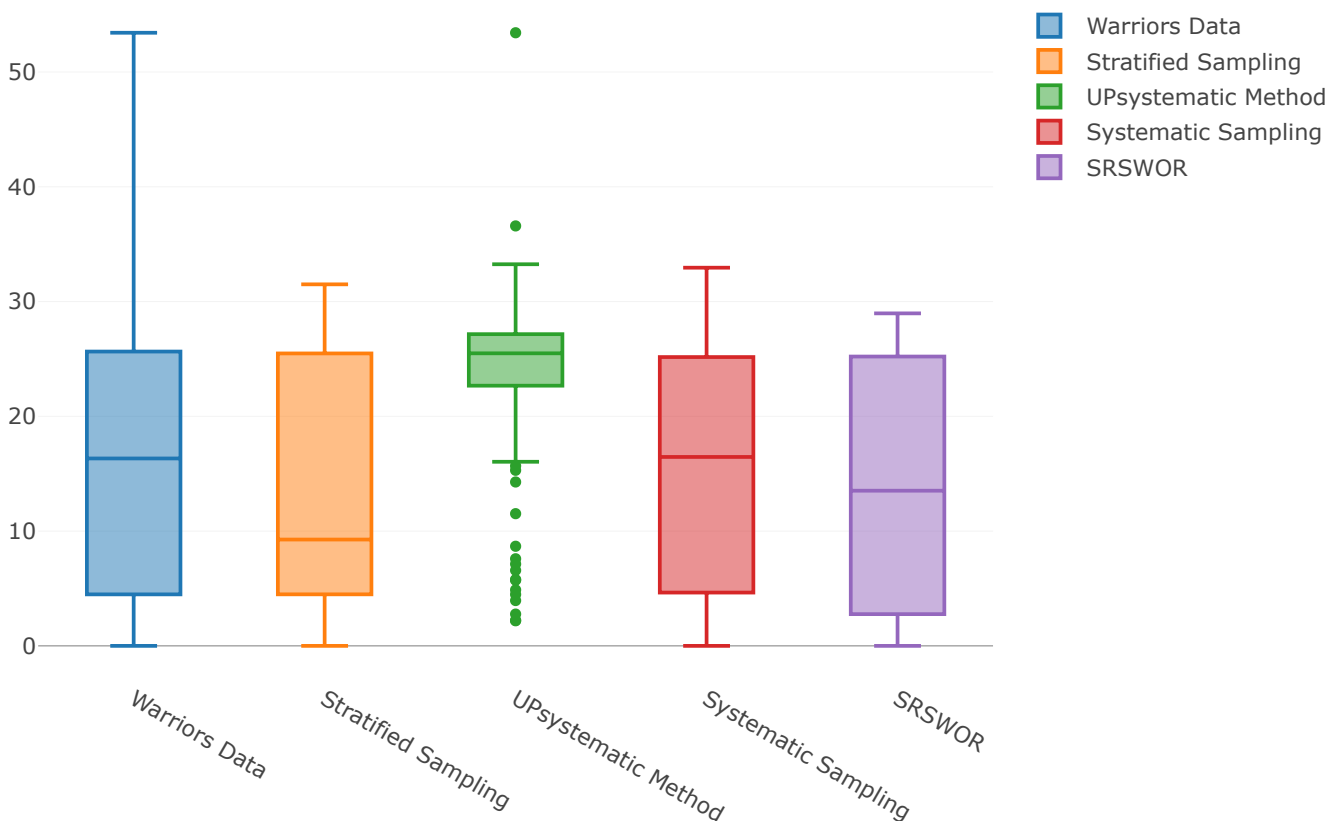
Conclusion:

1. Steph Curry generates the most points, followed by Andrew Wiggins and Jordan Poole. He has a smaller “average points made per attempt” compare to the other players, but this may be due to a multitude of reasons. e.g.
  - taking more challenging shots (guarded)
  - providing good assists to this teammates
  - shooting farther

## Sampling - Warriors Median Shooting Distance

Results from different sampling methods are used to estimate the population characteristics. Employing several sampling methods, the follow analysis aims to reveal the Warriors’ shooting distance. Comparing these numbers with the Leagues’ shooting average will quantify how strong a shooting Team the Warriors is.

Warriors Median Shooting Distance - Sampled on Players



```
## The Median Distance for the GS Warriors is 16.33 feet
## Median from Stratified Sampling of the CSG is 9.27
## Median from Sample Inclusion is 25.49
## Median from Systematic Sampling is 16.46
## Median from SRSWOR is 13.52
```

### Conclusion:

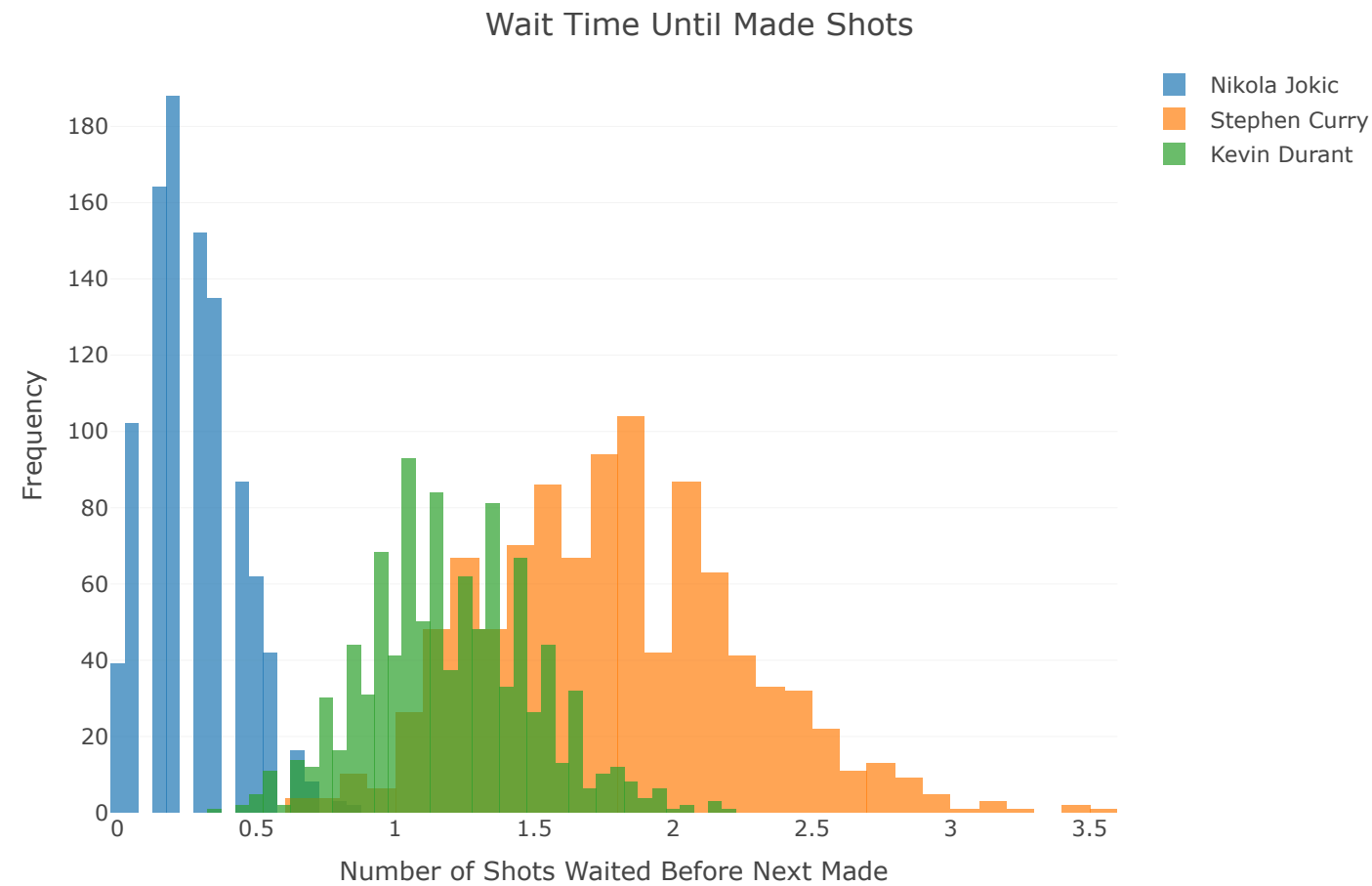
1. It looks like the most accurate is Systematic Sampling, follow by SRSWOR, then Stratified Sampling and then the Unequal Probability systematic sampling . This could be due to the fact that the data is ordered by games. Players perform differently in each game - taking more or less shots /shooting better or worst. So Systematic Sampling may have drawn an equal number of shots from each game.

2. If these samples are used instead of the whole dataset, we would either highly overestimate the Median Shooting Distance with Unequal Probability Systematic Sampling method, or highly underestimate the Median Shooting Distance with Stratified Sampling method. This large difference may be due to:
- lack of data available

• players have highly different shooting abilities
3. The league’s median shooting distance is 14.2ft and the Warriors’ median shooting distance is 16.3 ft, nearly 2feet farther.

## Hot Hand - Exponential Distribution - Wait Time until next made shot

Expectation is that there’s a peak near 0 if Hot Hand is true. If a player makes a shot, his wait time is 0 or close to 0 until his next successful shot.



```
## Mean Wait Time for Curry = 1.78 Number of Shots
## Mean Wait Time for Durant = 1.2 Number of Shots
## Mean Wait Time for Jokic = 0.29 Number of Shots
```

```
##
## Kevin Durant Stephen Curry
##      2.00      5.25
```

Conclusions:

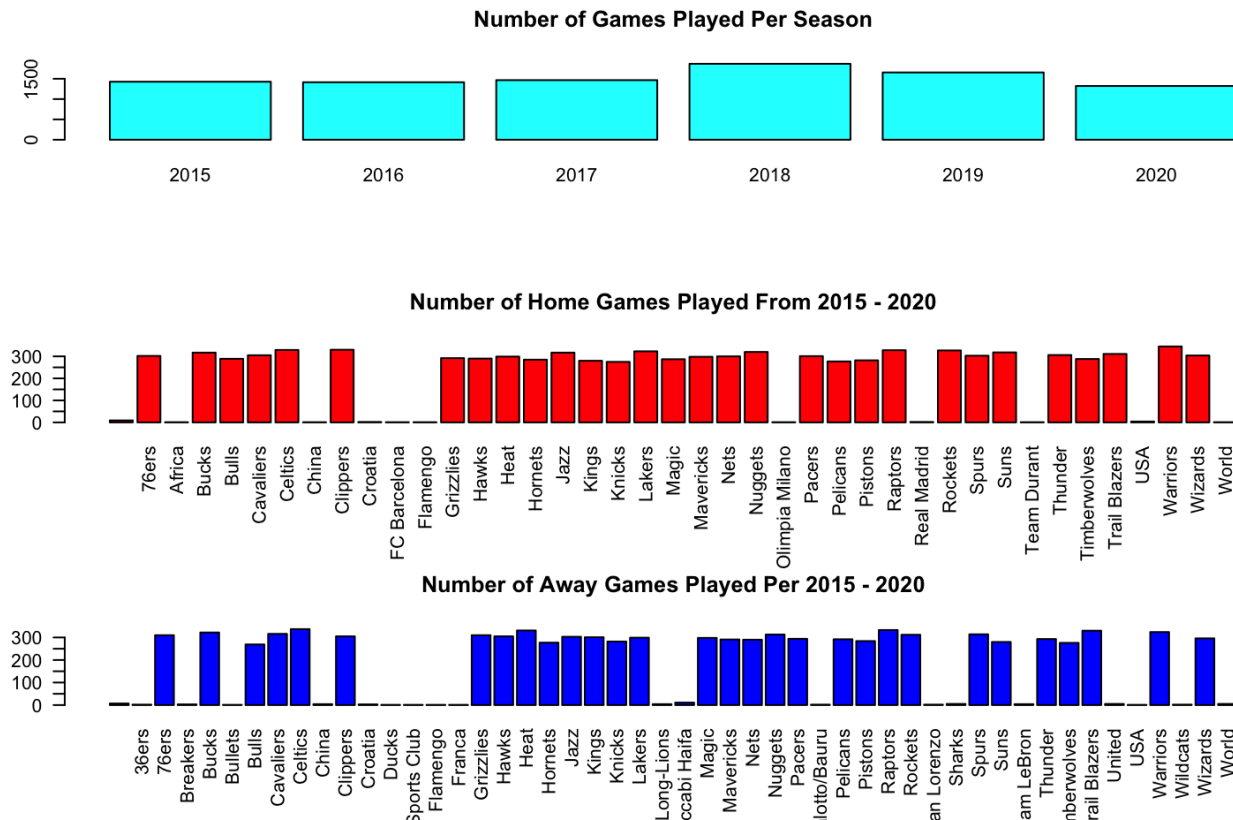
1. Durant and Jokic has a shorter wait time compare to Curry. This would imply that Curry is not a “streakly” player, but may be a “consistent” player.

- Curry takes more three points, so every successful shot that Durants and Jokic makes only amounts to 2 points while Curry can contribute 3 points.
- Jokic has a short wait. However, this analysis was based on 1 game for Jokic, explaining the white spaces between the blue bars. So, the hot hand myth is **inconclusive given the lack of data**.

Stephen Curry averages 5.25 threes per game while Kevin Durant only averages 2 threes per game.

# Analyzing Home Court Advantage

## Data Ingestion, Explore Data, Perform preprocessing and Data Wrangling

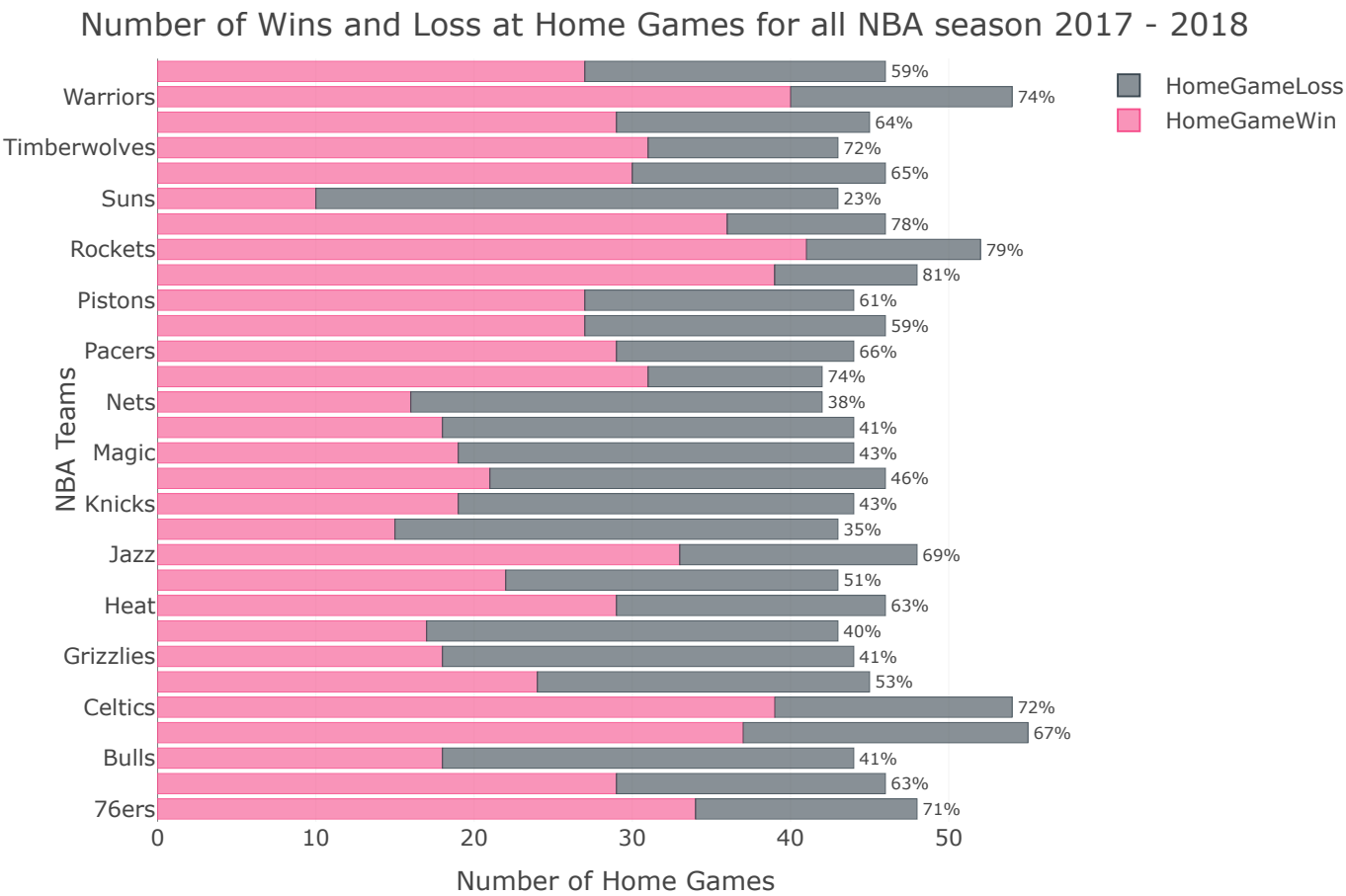


After examining the data set, there are irrelevant records:

- row filter- only keeping:
  - league = standard
  - country = US, USA, and Canada(Toronto Raptors)
  - gameStatus = Finish
- Converting data to the correct data type:
  - startTimeUTC and endTimeUTC needs to be converted to datetime
    - redefining column gameduration to endTimeUTC - startTimeUTC in minutes
    - current gameduration is unusable because it is a string in hours and minutes
  - Calculating a new column call GameDate by converting startTimeUTC to Date
- Need to filter out columns that are irrelevant or has replicated information
- Only focusing on 2017 data because data for 2015 or 2016 is incomplete,

# Categorical Variable Analysis - What's the percentage of winning at home home games?

Code



Code

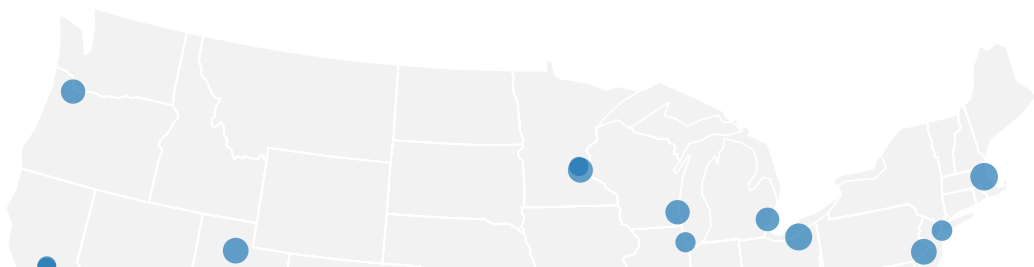
```
## On average, NBA teams has a 58% chance of winning at Home Game, giving them a slight advantage over their opponents.
```

This is only one perspective. There may be many confounding variables e.g. good teams will win at home and away - masking the advantage given by home court advantage if applicable. So we cannot make a conclusive statement.

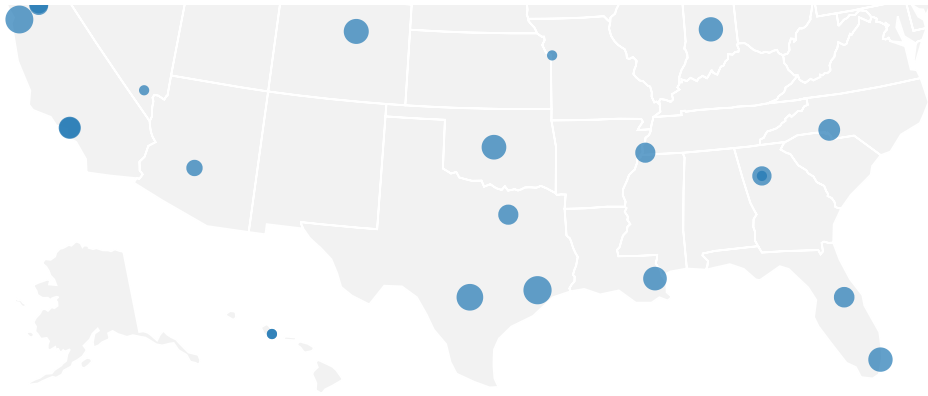
## Does the location affect outcomes?

Code

US Map of NBA locations







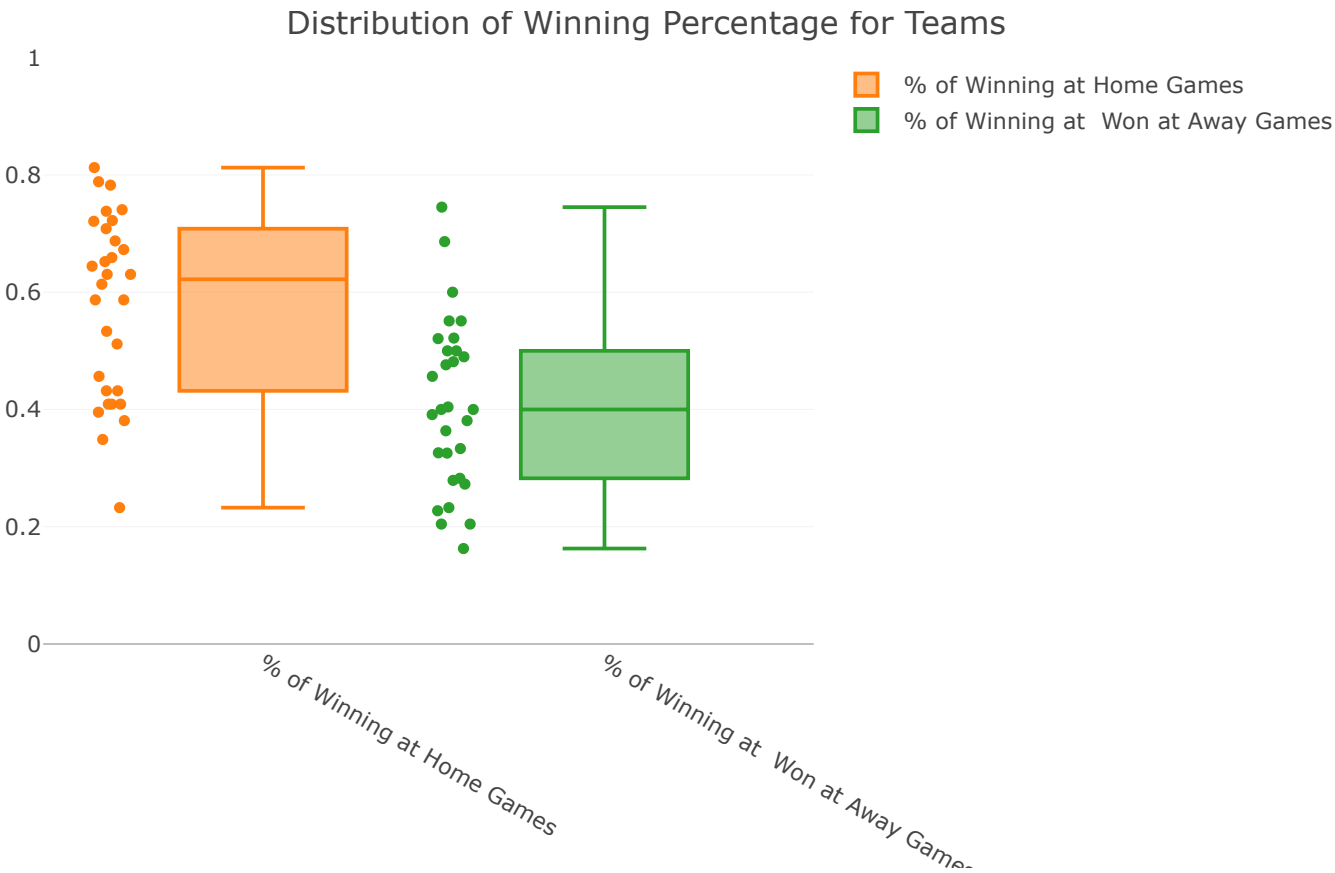
To eliminate others factors that may be affecting the impact of player’s performance in different locations. Denver is the primary focus. The objective is to calculate the difference between chance of winning for every team playing not at Denver vs playing at Denver.

Code

```
## League's likelihood of winning at Denver is 30.46%
## League's likelihood of winning at Away Game is 40.9%
## SD is 14.5%
```

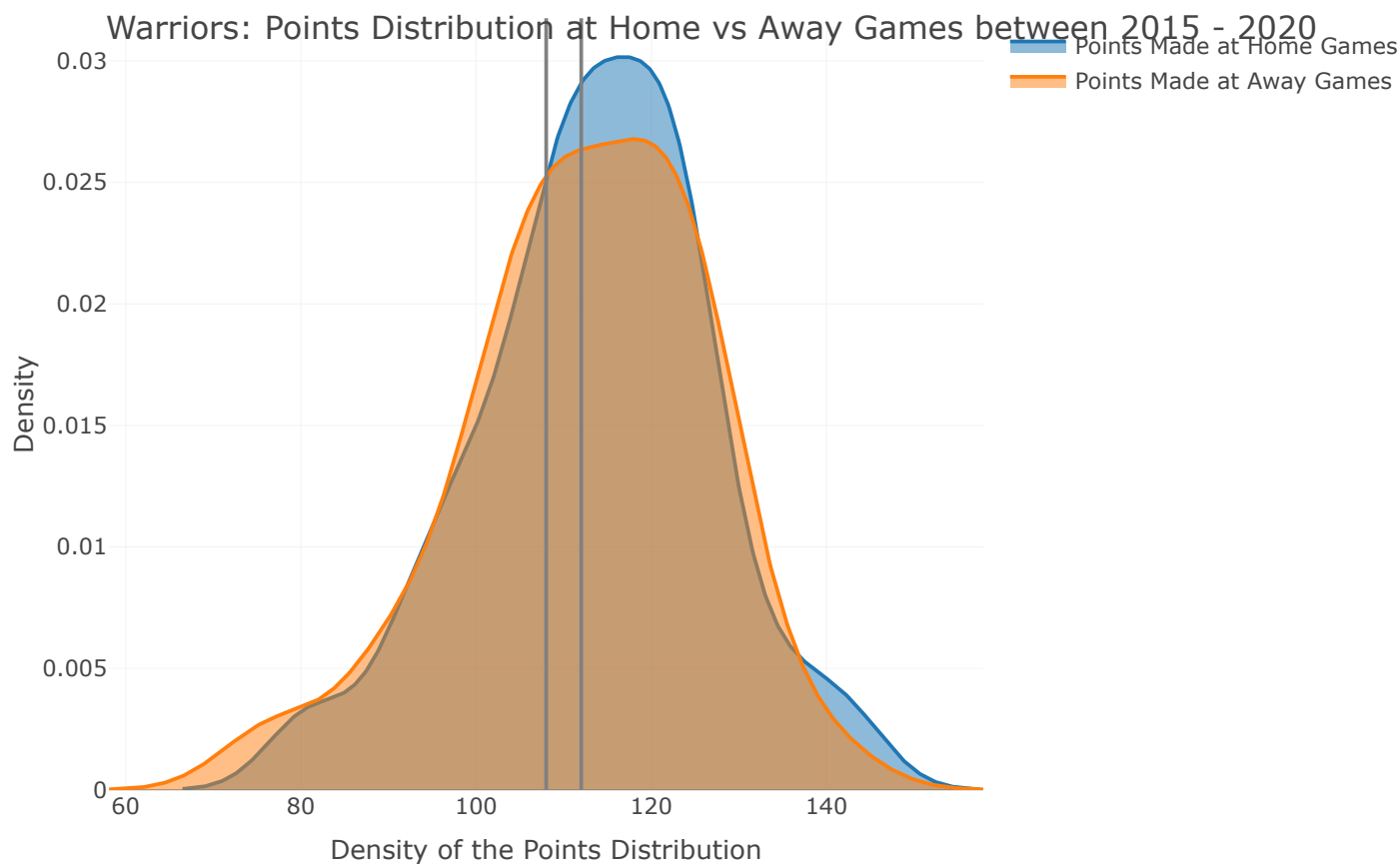
Even though teams are less likely to win at Denver, 30.46% is within 1 standard derivation of 40.9% (+/- 14.5%). This means we cannot make any conclusions with high confidence.

Code



This supports the idea that there's high variance in Winning Expectation for Teams.

## Examining just the Warriors to see if there is a drop in performance given travel and timezone difference.

[Code](#)

[Code](#)

```
## The median of the GSW Total Score is 108.
## The median of the GSW Total Score at Home Games is 112.
```

Warrior's scoring abilities doesn't seem to be affected by away games significantly. Only a 4 points difference between at home vs away.

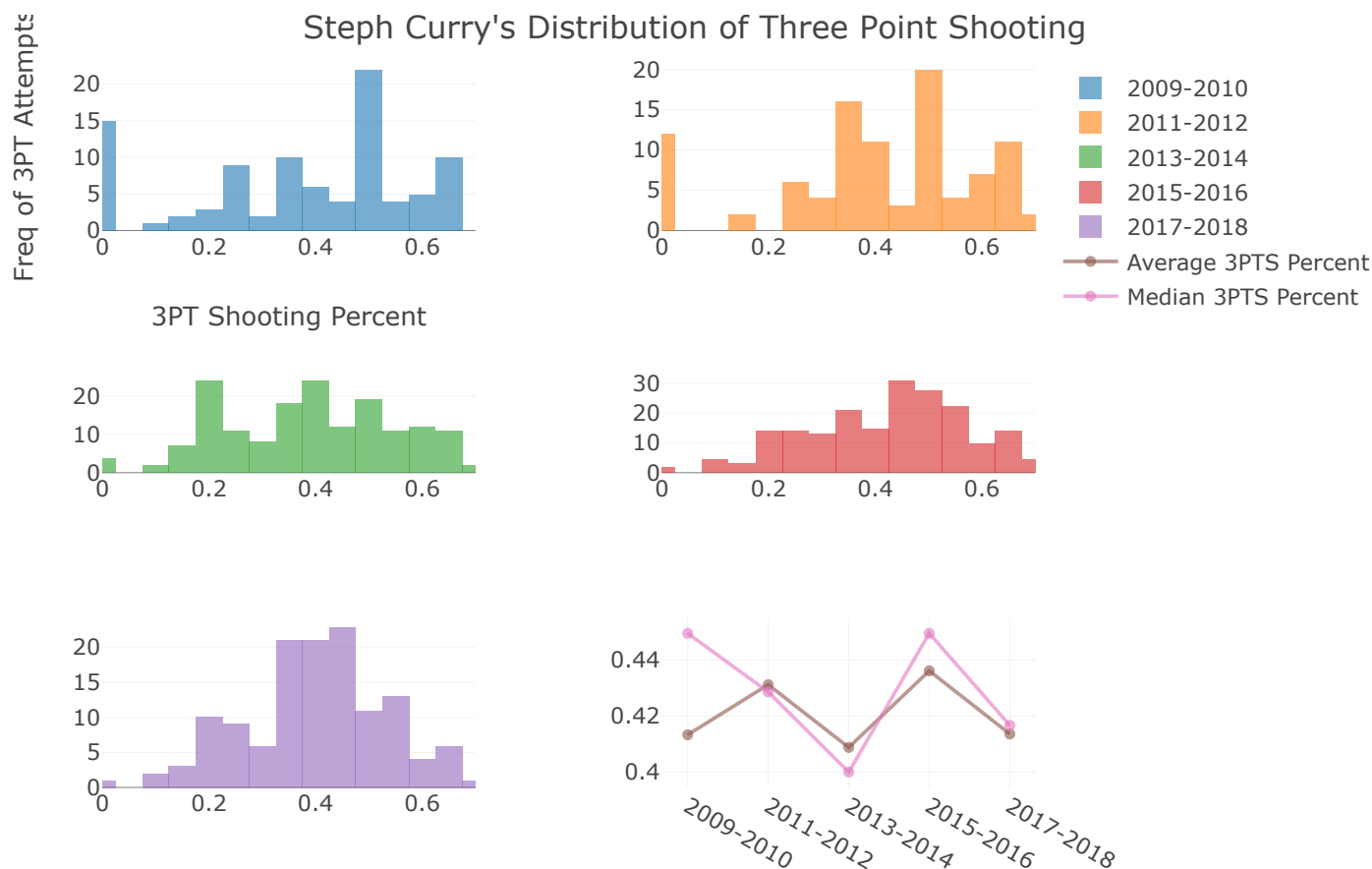
## Performance During Regular vs Playoff Games

### Data Ingestion and Preprocessing for Stephen Curry's Game stats

This data set will be used to compare Steph Curry's player performance during regular season games vs playoff games. For accuracy, the data will be filtered on seasons that the Warriors made it to the playoffs.

# Numeric Variable Analysis - 3PTS Shooting Percentage

Code



## Conclusion:

Steph Curry has always average a 40% shooting for 3PTS since season 2009 - 2018. Based on the Histograms, Curry took the most 3pts in 2015 and 2016 and averaging 44% with a median of 45%.

## Regular vs Playoff Performance

- Any performance difference between Regular and Playoff Games for elite players? Focusing only on Steph Curry for this analysis.

Code

```
## Regular season shooting percentage: 0.117206780167678 .
## Post reson shooting percentage: 0.105224210827332 .
## since the sd is relatively the same, a t-test is applied.
```

Code

```
##
## Welch Two Sample t-test
##
## data: regdf$shooting_percentage and psdf$shooting_percentage
## t = 1.7141, df = 136.98, p-value = 0.08877
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.00326292 0.04574416
## sample estimates:
## mean of x mean of y
## 0.4739225 0.4526819
```

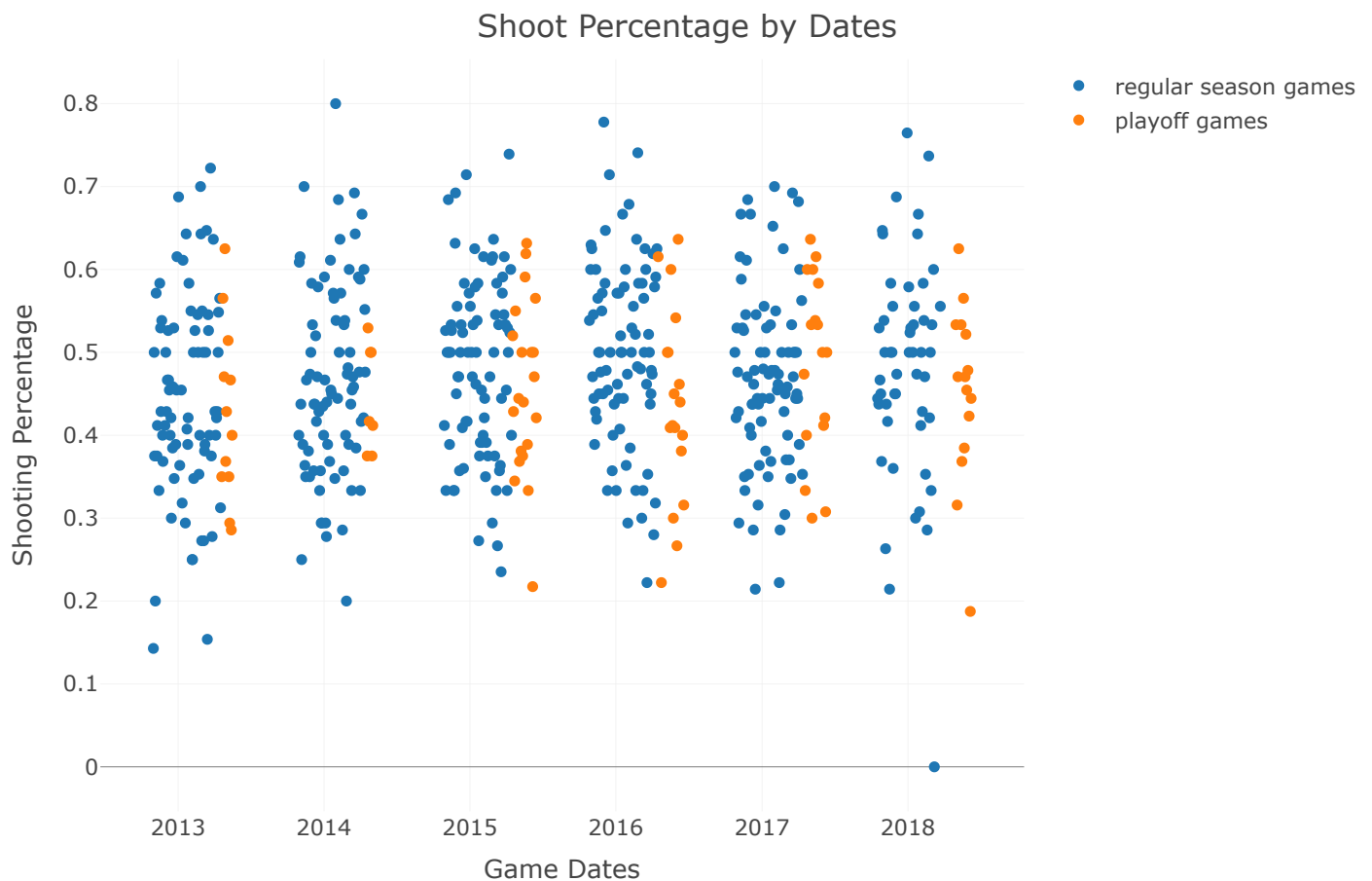
Code

```
## H0:  $\mu_1 = \mu_2$  (psdf and regdf means are equal)
## HA:  $\mu_1 \neq \mu_2$  (psdf and regdf means are not equal)
```

### Conclusion:

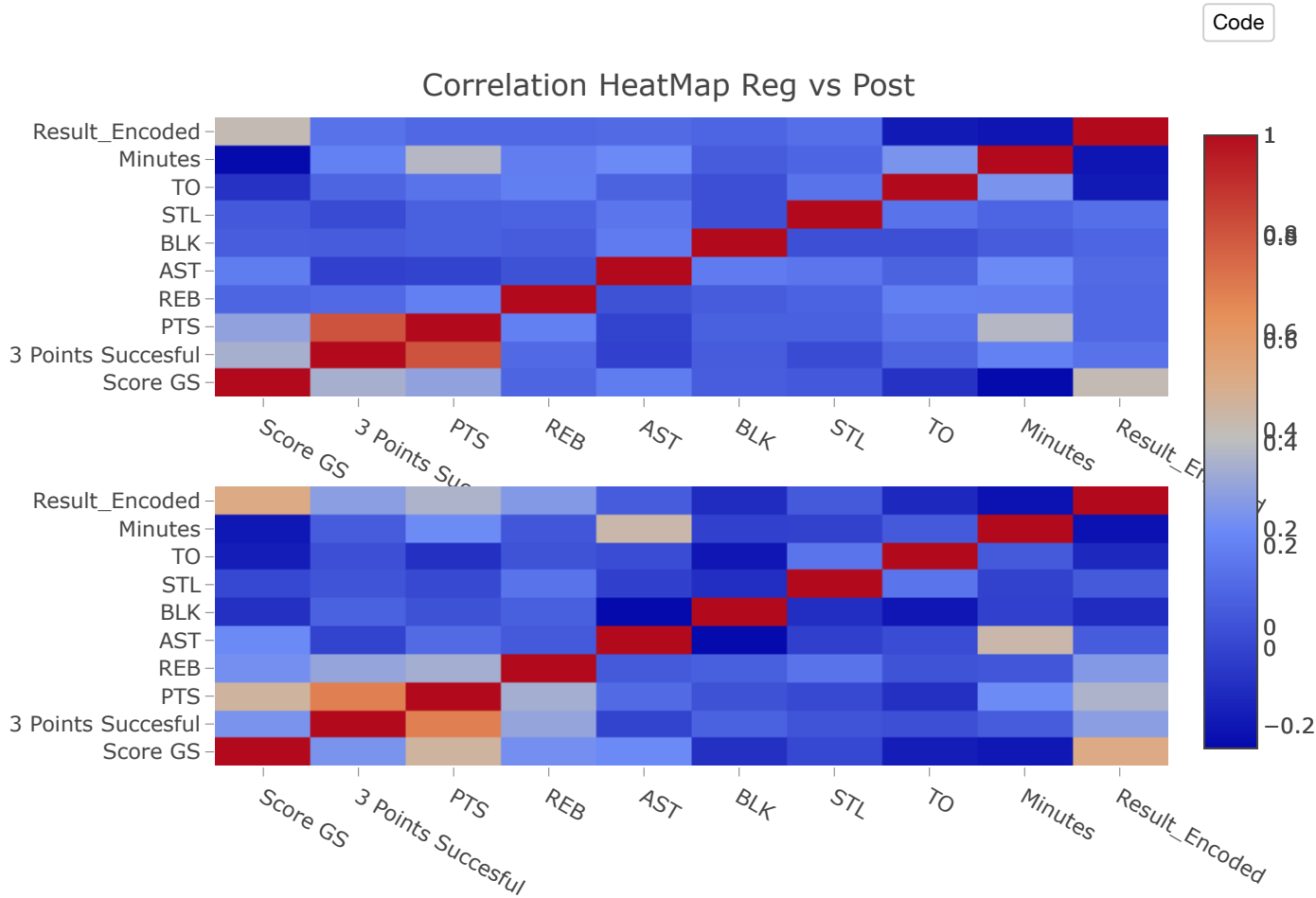
The p-value = 0.08 from the Welch Two Sample t-test. There's no significant shooting difference between regular season games and playoff games for Steph Curry.

Code



The spread of Steph Curry's shooting percentage is relatively similar between Regular Season Games and Playoff Games. However, there are more variance in the Regular Season Games.

# Examining Correlation between Steph Curry PTS Contributions and Team Score:

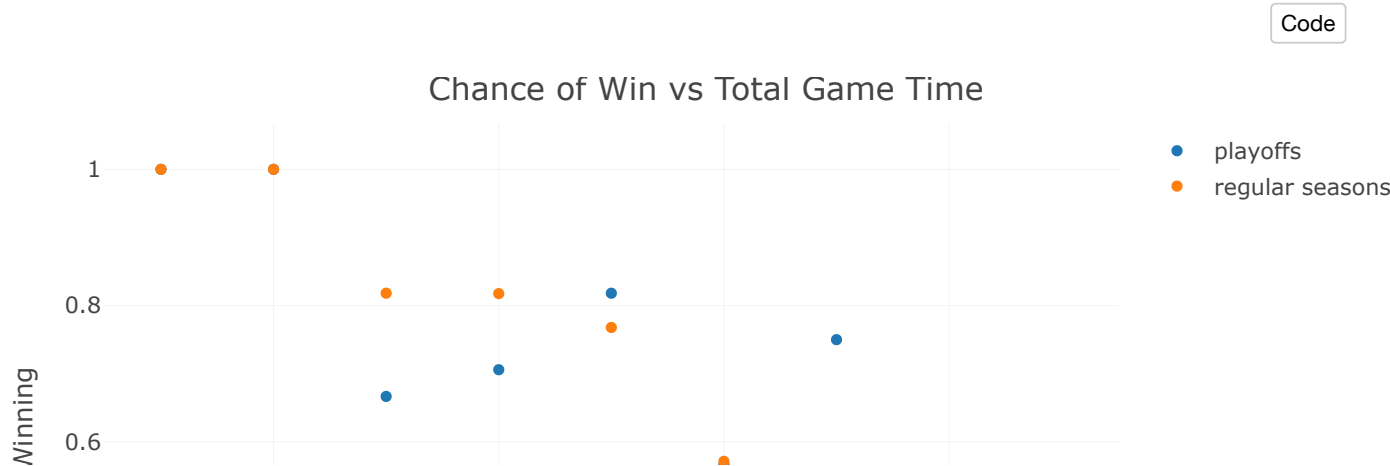


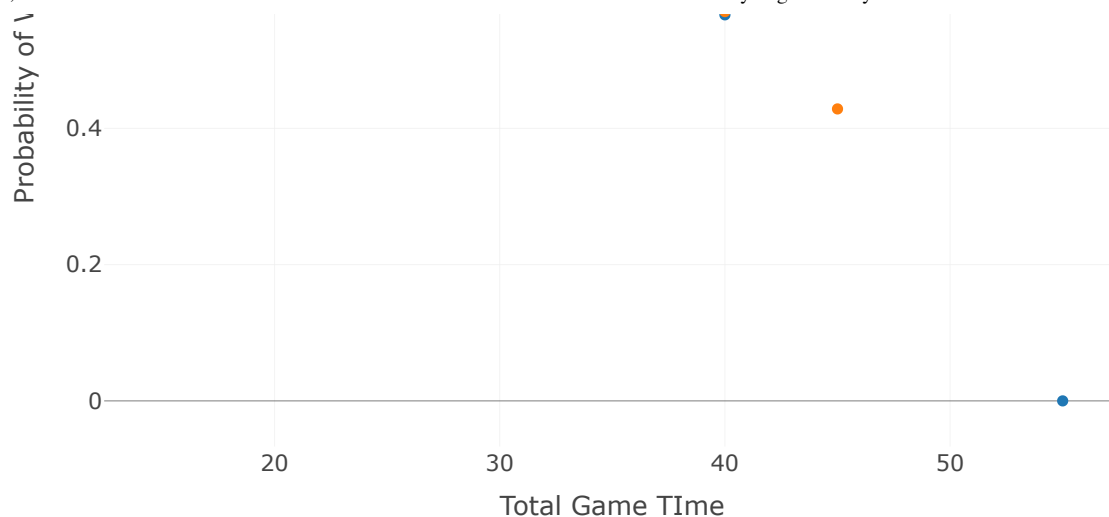
Result\_Encoded = 1 if Win, 0 if Loss.

The correlation heat map did not reveal any significant correlation between winning and other factors besides with Team Score.

Small correlations: 1. Positively correlated with Steph Curry’s points and rebounds. 2. Negatively correlated with Steph Curry’s turn Overs and playtime. \* Further analysis - check if Steph is more likely to make more mistakes as he plays longer due to exhaustion, lost of focus, or a weak bench team.

## Impact of Not Resting Top Players





There appears to be a linear negative relationships between Steph Curry's total time played and the Warriors' chance of winning. This can be caused by the lack of depth of the Warriors' Rosters. When the team has no depth, and players are injured, Steph Curry has to play more minutes which decreases their chance of winning. Needs Further Analysis

## Project Conclusion:

May factors affects a team's performance. In examining the many questions above, we attempt to verify the legitimacy of many NBA Myths.

1. Hot Hand doesn't appear to apply for Steph Curry or Kevin Durant. They appear to to be consistent players who makes successful shots on averaged every 1.2 -1.7 shot attempts. For Jokic, there was not enough data to conclusively state that he is a streaky player. He could be taking easy shots, closer to the hoop shots, and/or making smarter decisions. The x and y coordinates of Jimmy Butler's "made" shots appear to be highly correlated. A plausible explanation for this is that he's good on the offense glass, not a shooter, supported the week correlations between x and y of his "unmade".
2. The Home Team does appear to have a slight advantage. Furthermore, teams appear to have a challenging time winning at Denver during the 2017 season. However, note that the variance associated with the home court advantage analysis is high, ~ 14%. Additionally, there are many confounding variables play, e.g. dominant teams with better roster.
3. Steph Curry is known as an elite shooter/player. There is no significant difference between his performance during the regular vs post season. There appears to be a negative correlation given the Warriors' chance of winning and Steph Curry's play time. Further analysis needs to be conducted to make any conclusive statements. Could this be attributed to player exhaustion, loss of focus during long games, or injured teammates?