Analysis of NBA Champion Based on Regular Season and Playoffs Performance

Yu-Lin Shih (CIN305857869)

Department of Information Systems

College of Business and Economics

California State University – Los Angeles

Object of study

The 2018 NBA playoff just began few days ago. People are excited to see the games and wondering stories like how far will the young 76ers go or will LeBron James lead the new Cavaliers into the finals and most of all will the golden state worriers win their next championship?

Each year nearly 22 million people in U.S. [1] watch the playoff serious and it creates millions of revenue on commercial, jerseys and tickets. Which it causes the meaning of winning the championship is not only glory but also means earning more money. However, there are many factors may influence the final result such as chemistry, player heath, coach and etc. To find out the keys to win the championship, many teams have embraced the statistic [2] on training players or finding others disadvantages. In fact, Professor Harish Bhat in UC Merced used big data to predict NBA winners and the accuracy is up to 80 percent. [3] The object of this project is to find out the pattern of NBA champions according to their regular season and playoffs performance.

On the other hand, many fans and players also care about NBA 2k series. This series convert the performance of each player into numbers and rank each team according to these numbers. This project will also consider the data in NBA 2k series and analysis the relationship between reality and NBA 2k.

Data Sets

1. NBA Enhanced Box Score and Standings Stats:

<https://www.kaggle.com/pablote/nba-enhanced-stats/data>

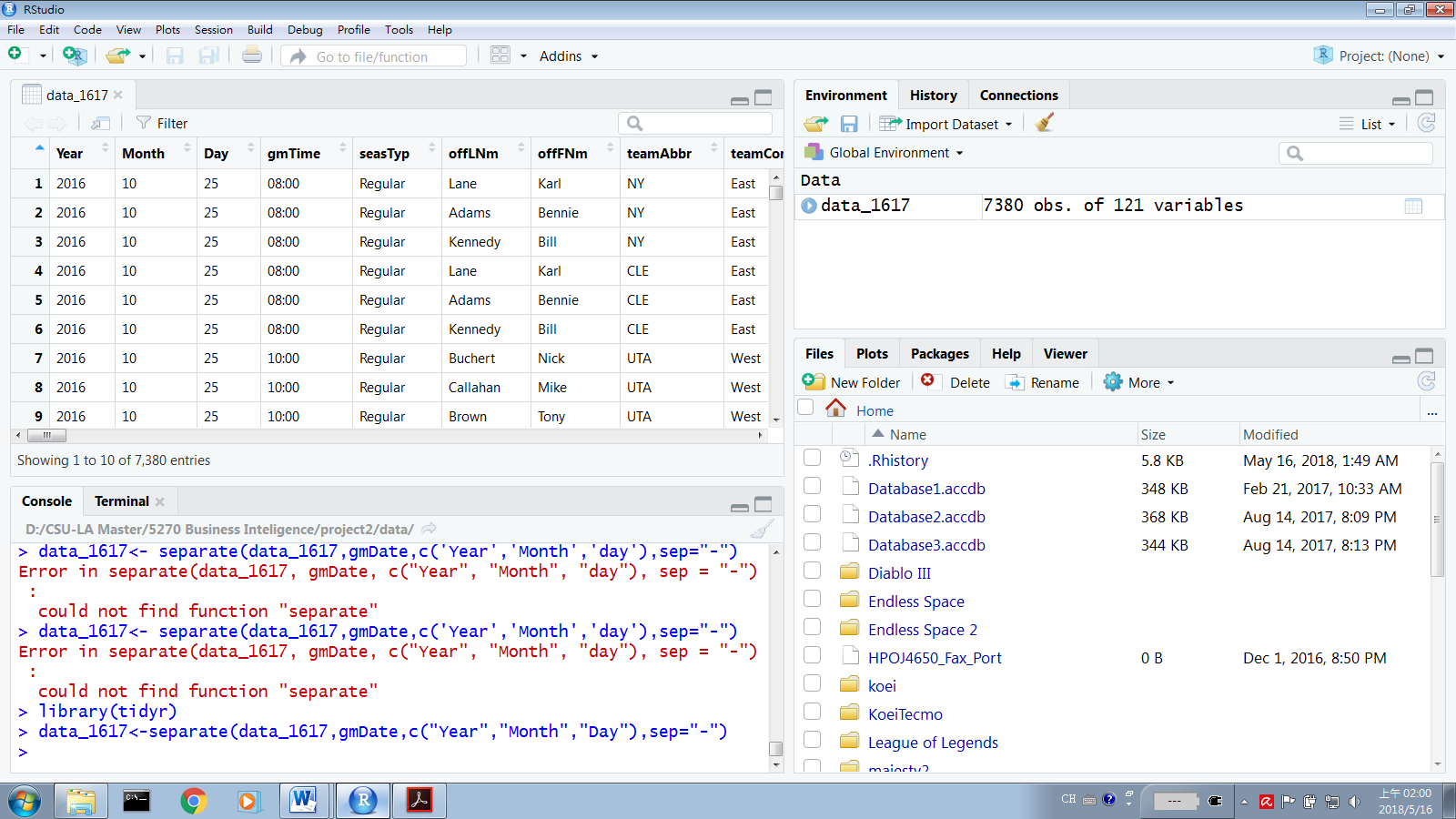
Data Description

This project will mainly use 2017-17\_officalBoxScore.csv and 2017-18\_officalBoxScore.csv.

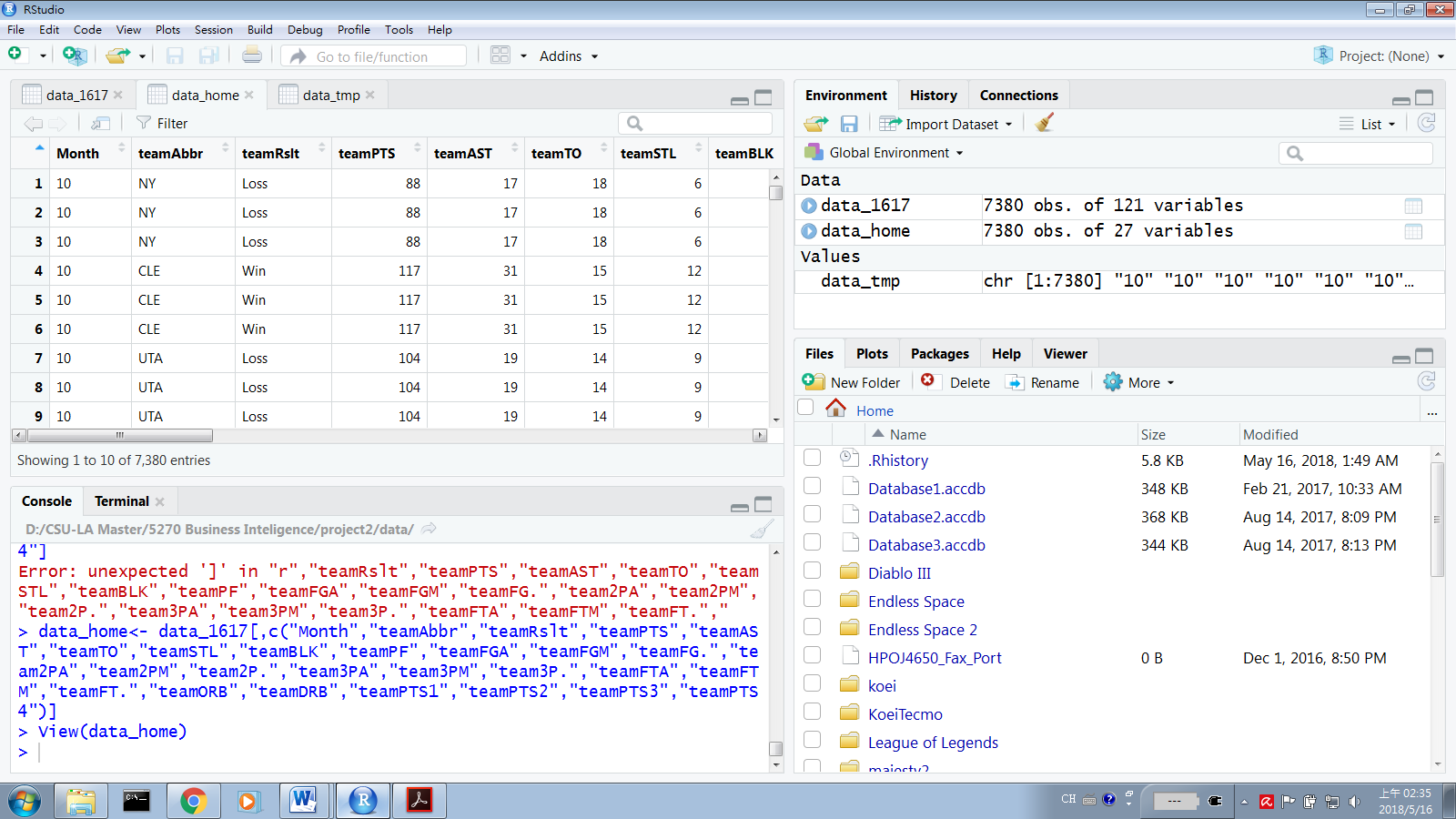
These datasets are in CSV format and contains box scores of each game and standing stats from 2016 to 2018. The amount of this datasets will be 8.52 MB and more than 100 columns.

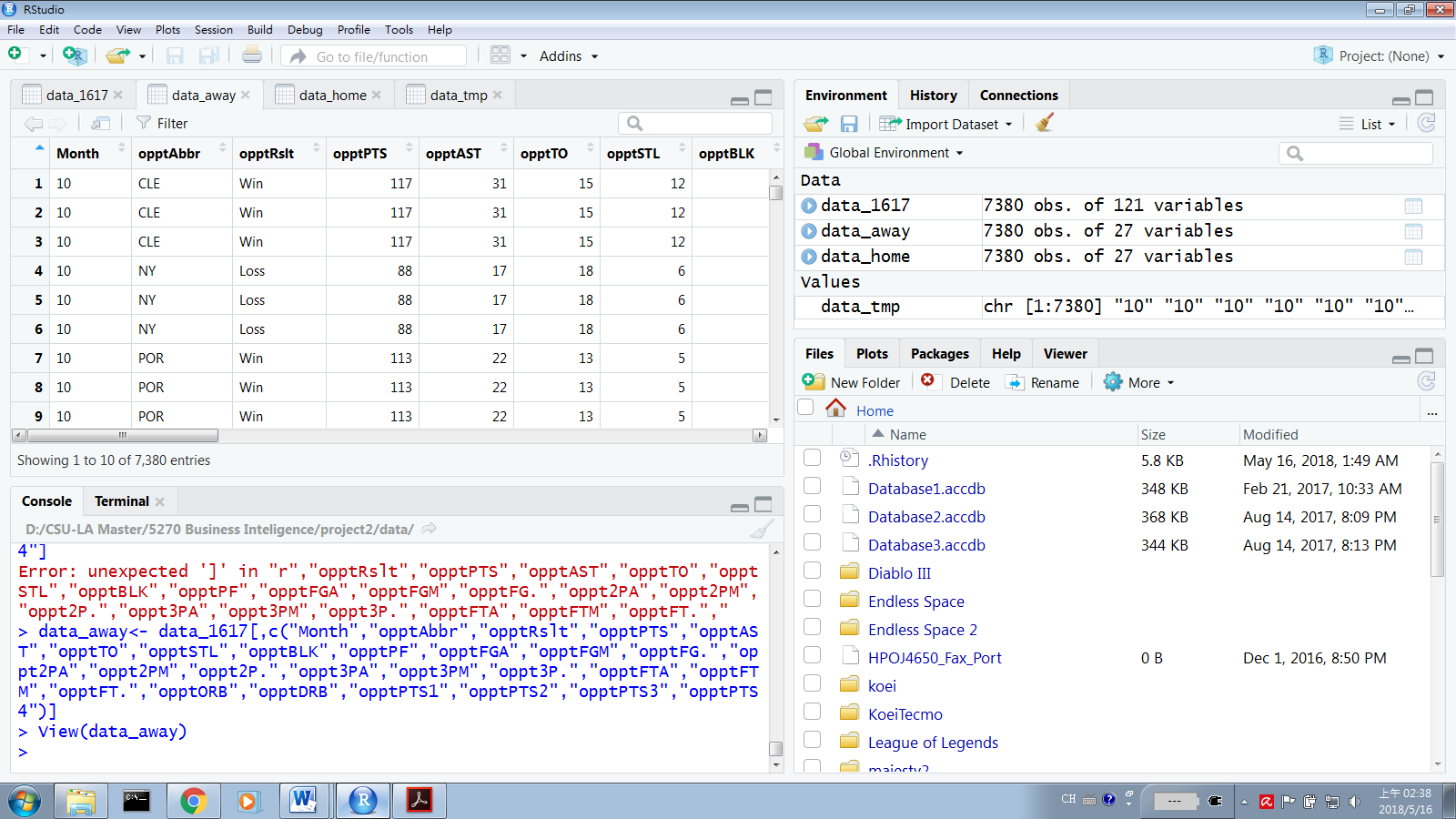
Data Cleaning

1. Separate columns: In order to analysis data by month. The gmDate will be separate into Year, Month and day.

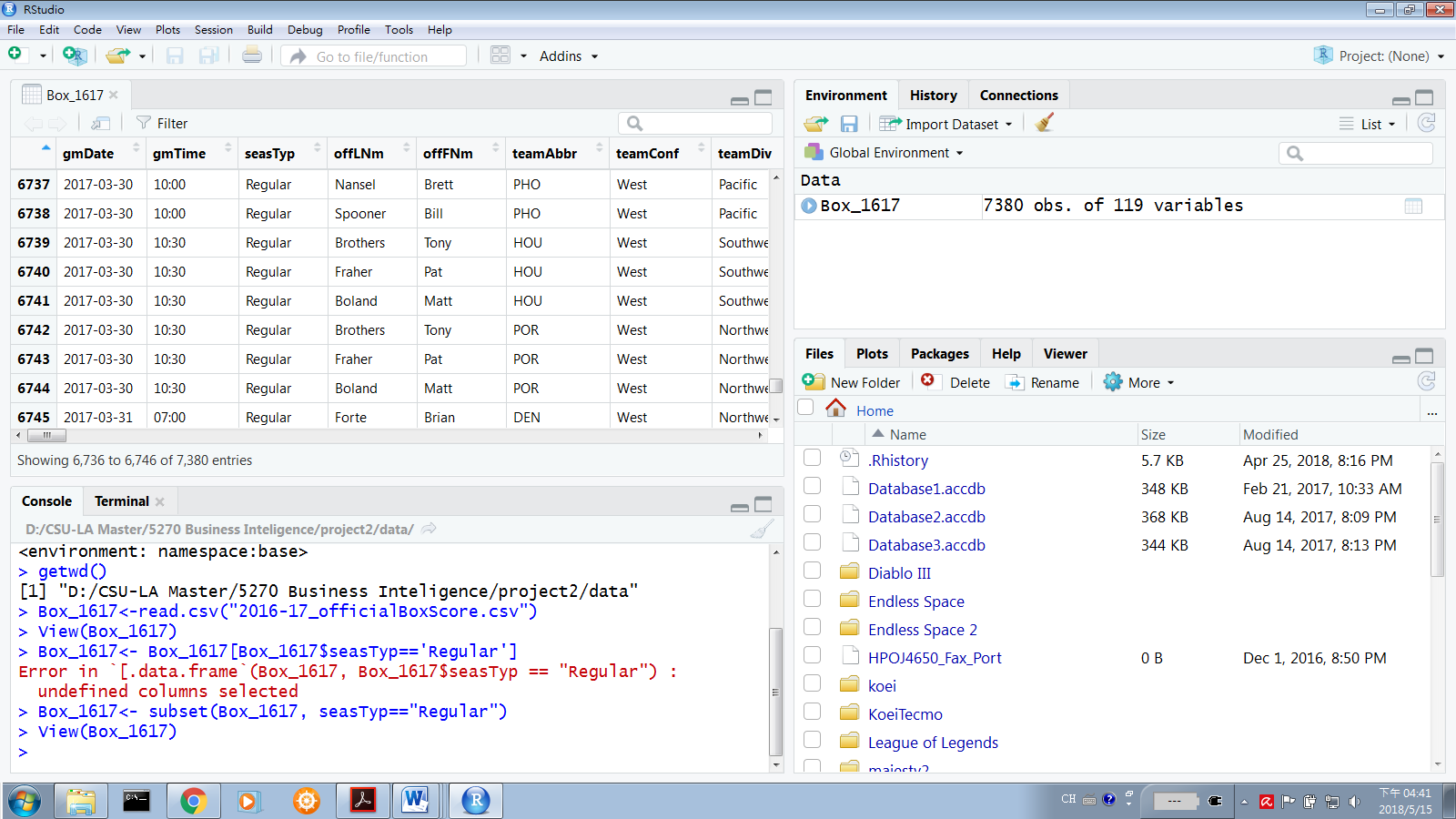


1. Remove irrelevant columns: The dataset combine both teams’ data in one game. To analysis more easily, I separate the dataset into home and away.





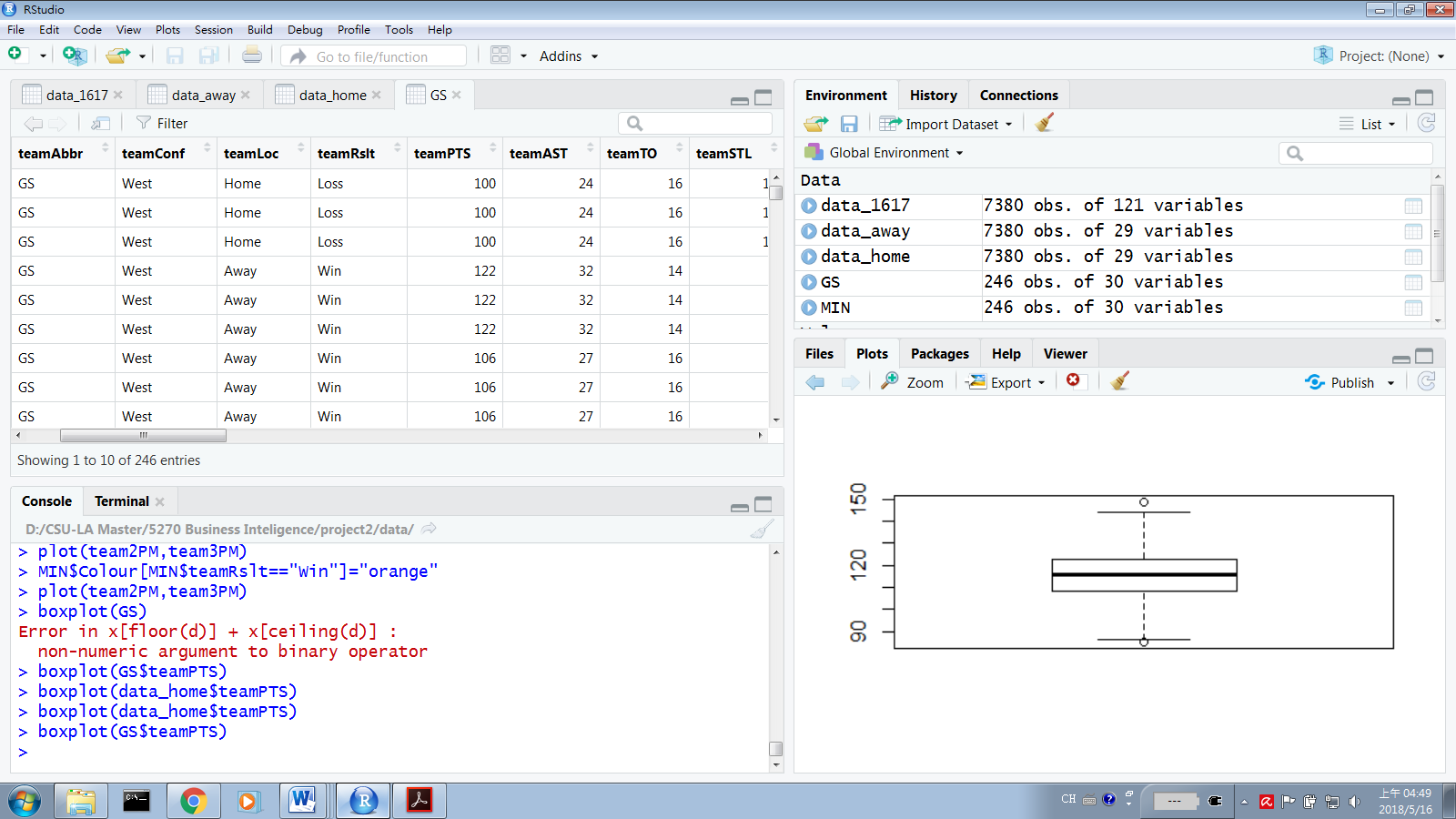
1. Remove irrelevant rows: This project focus on the games in regular season. Games on play-off will be delete.

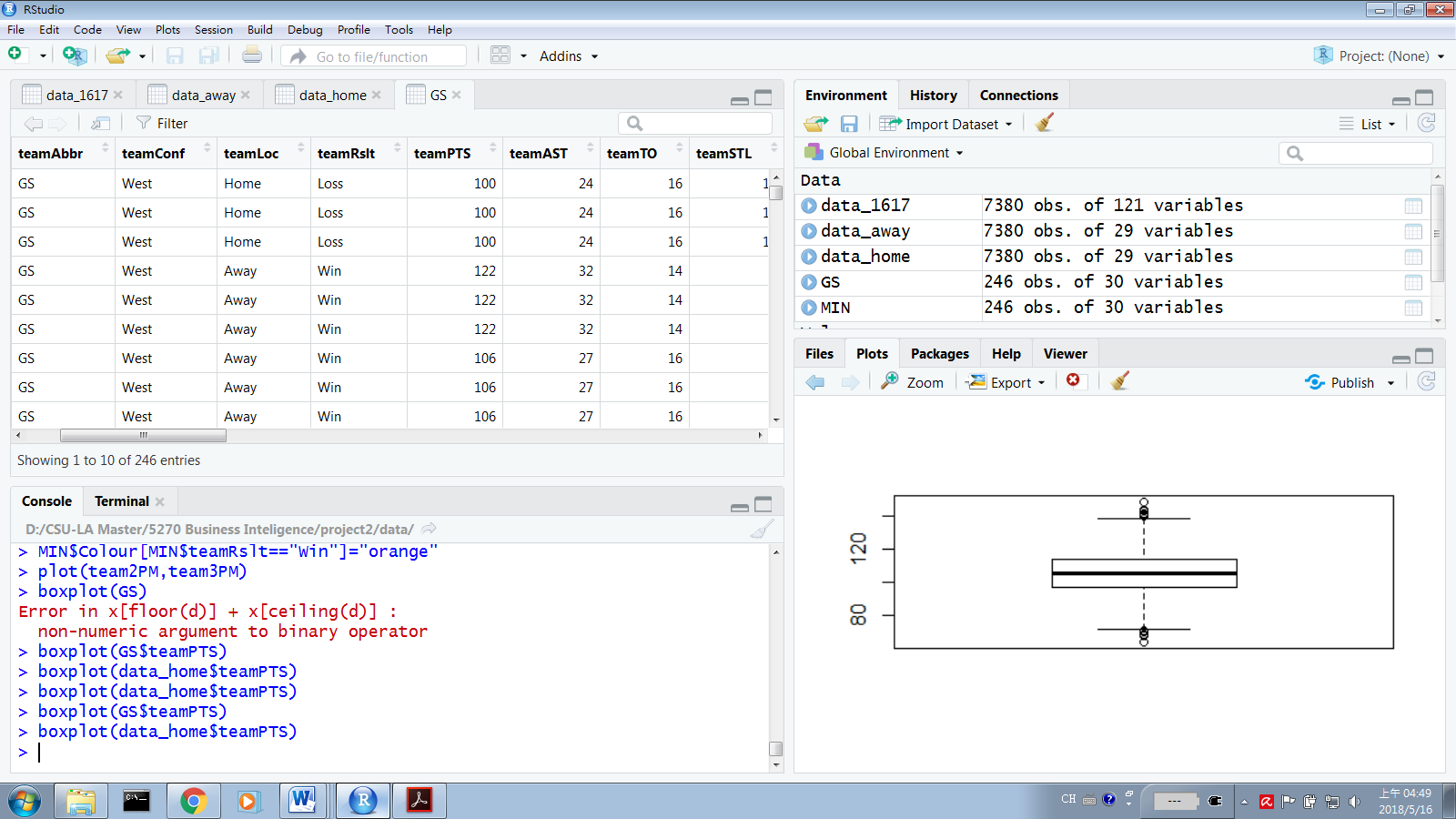


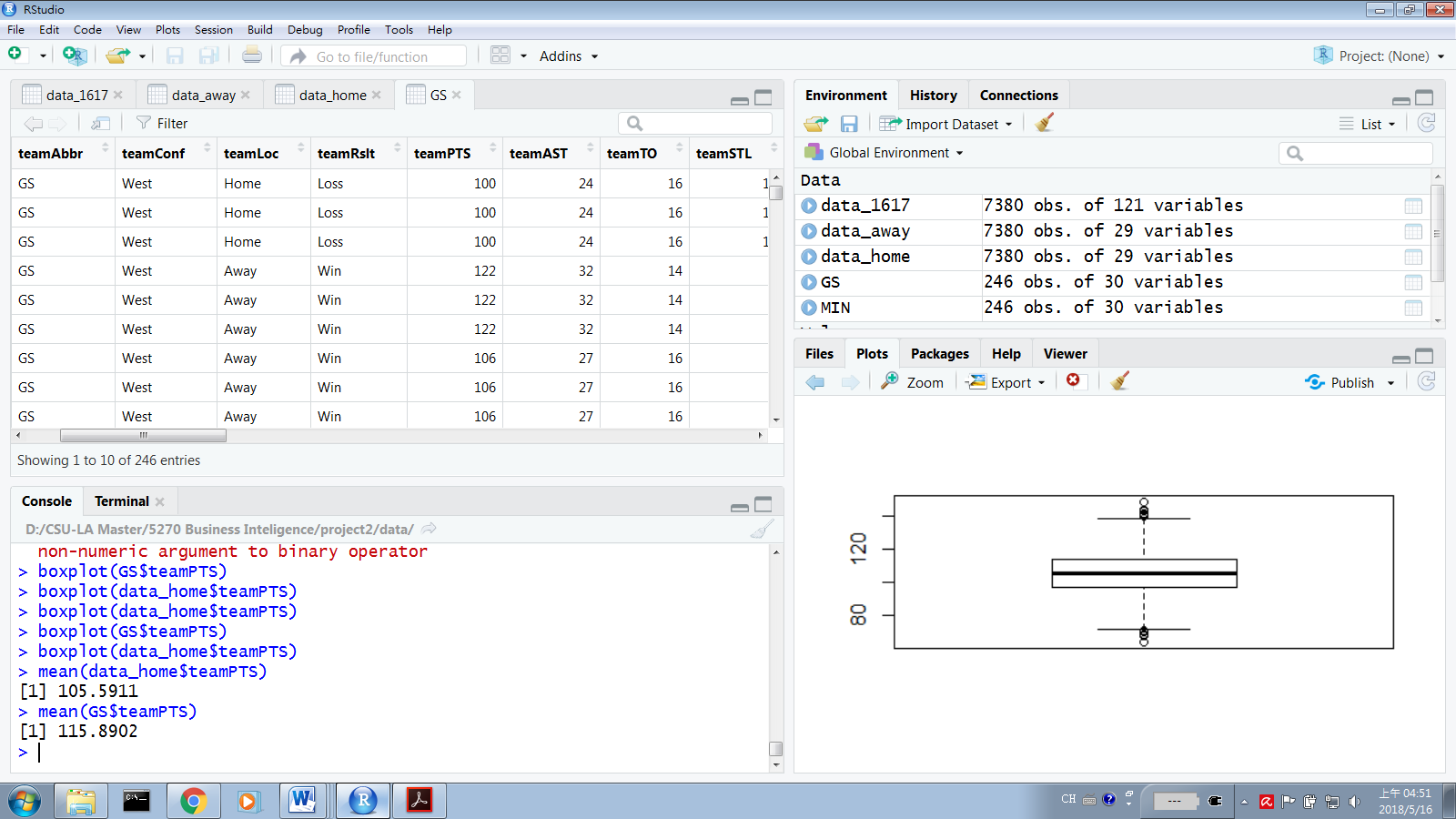
Project Description

This project will focus on 3 questions. The first one is what kind of performance do those teams which enter the playoffs have in regular seasons? The second one will be what are the different between the champion and the other teams? The third question is what are the attribute did the champion have? These data will be processed and analysis by R language.

1. Performance of teams that enter playoffs.

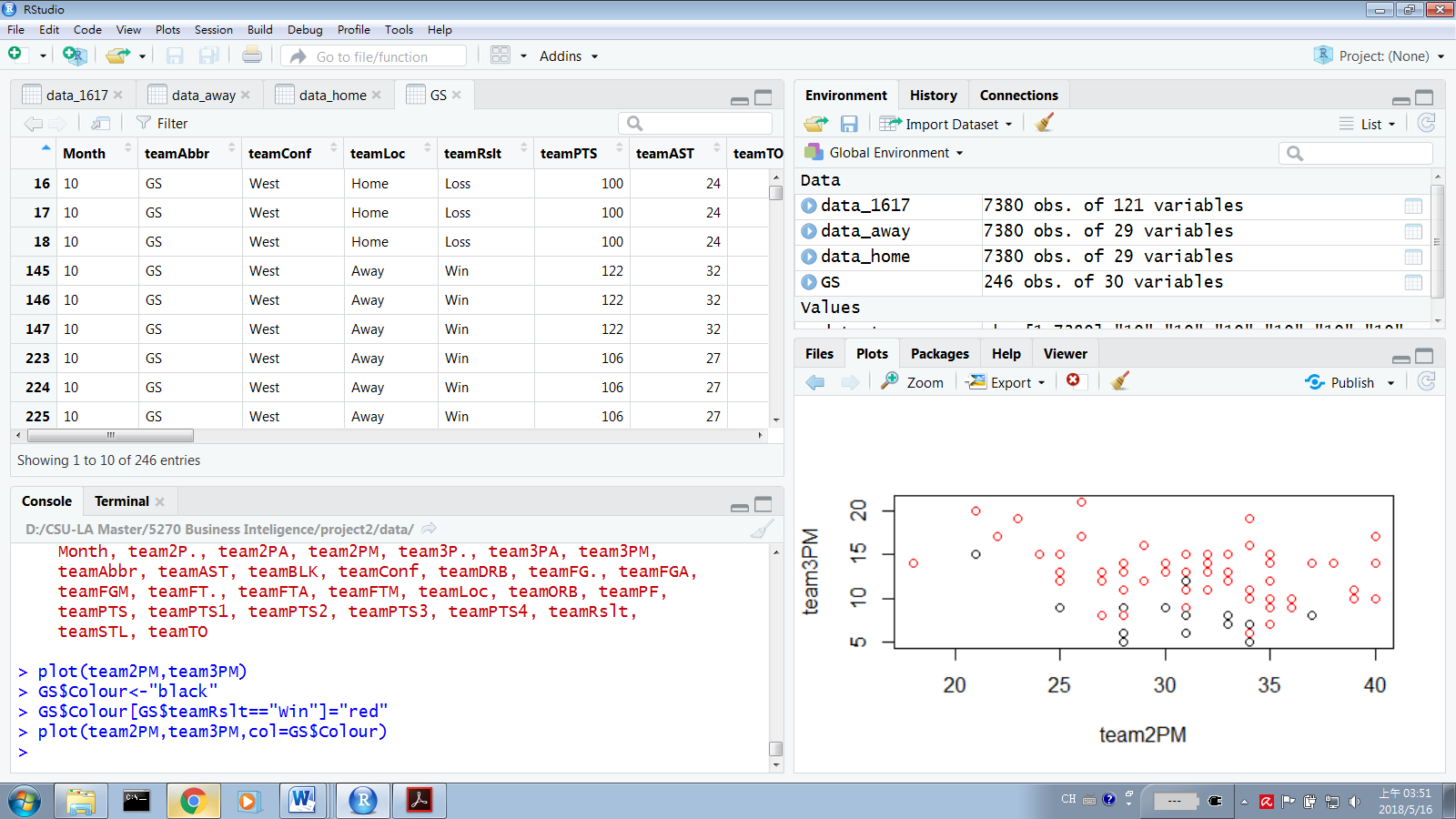






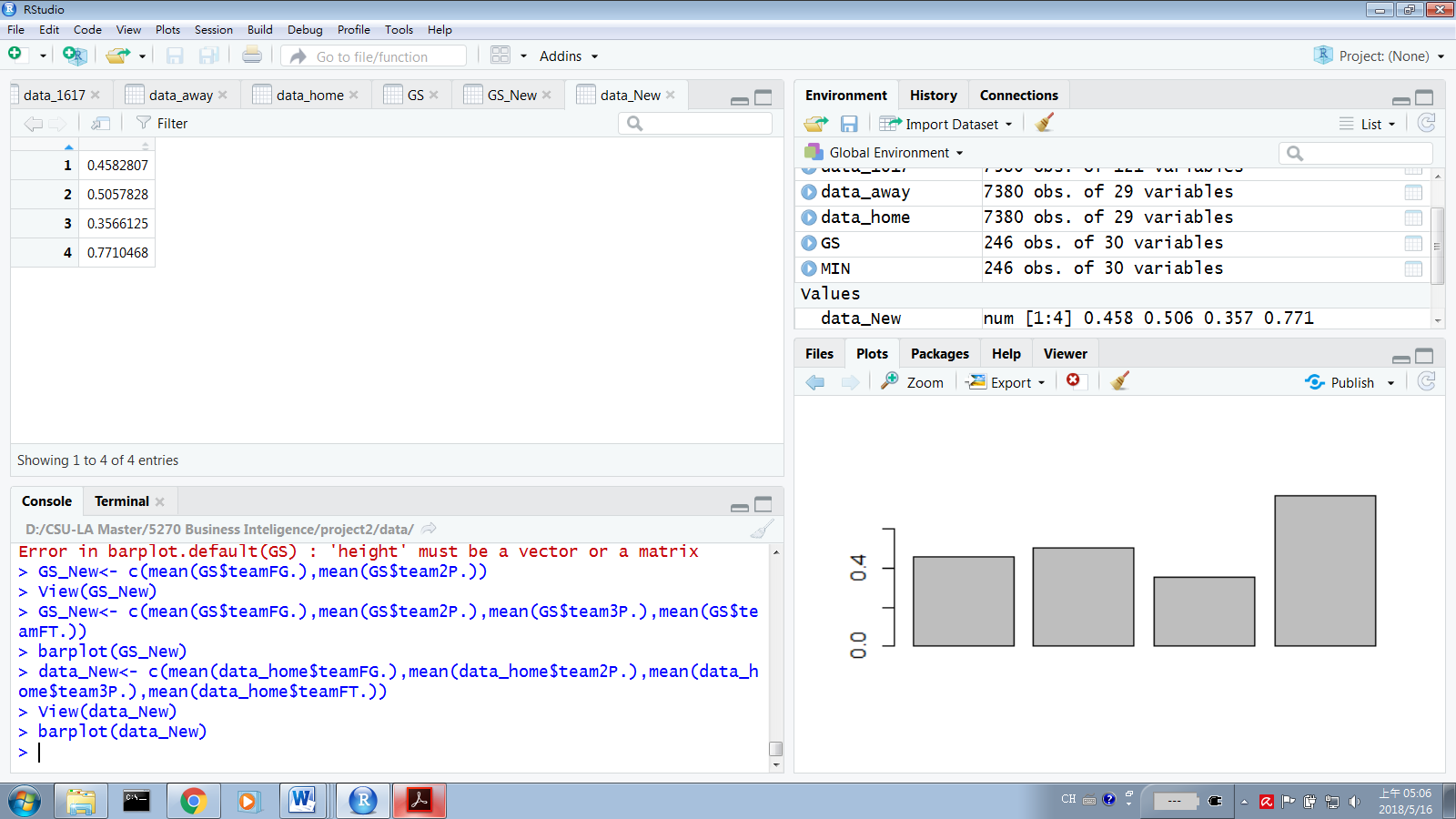
The average point of each team in regular seasons is 105. The Golden State Warriors’ average point is 115. Even the best team in NBA, the point they got are similar to the rest of teams. As the result, the ability of getting point are similar for each team.

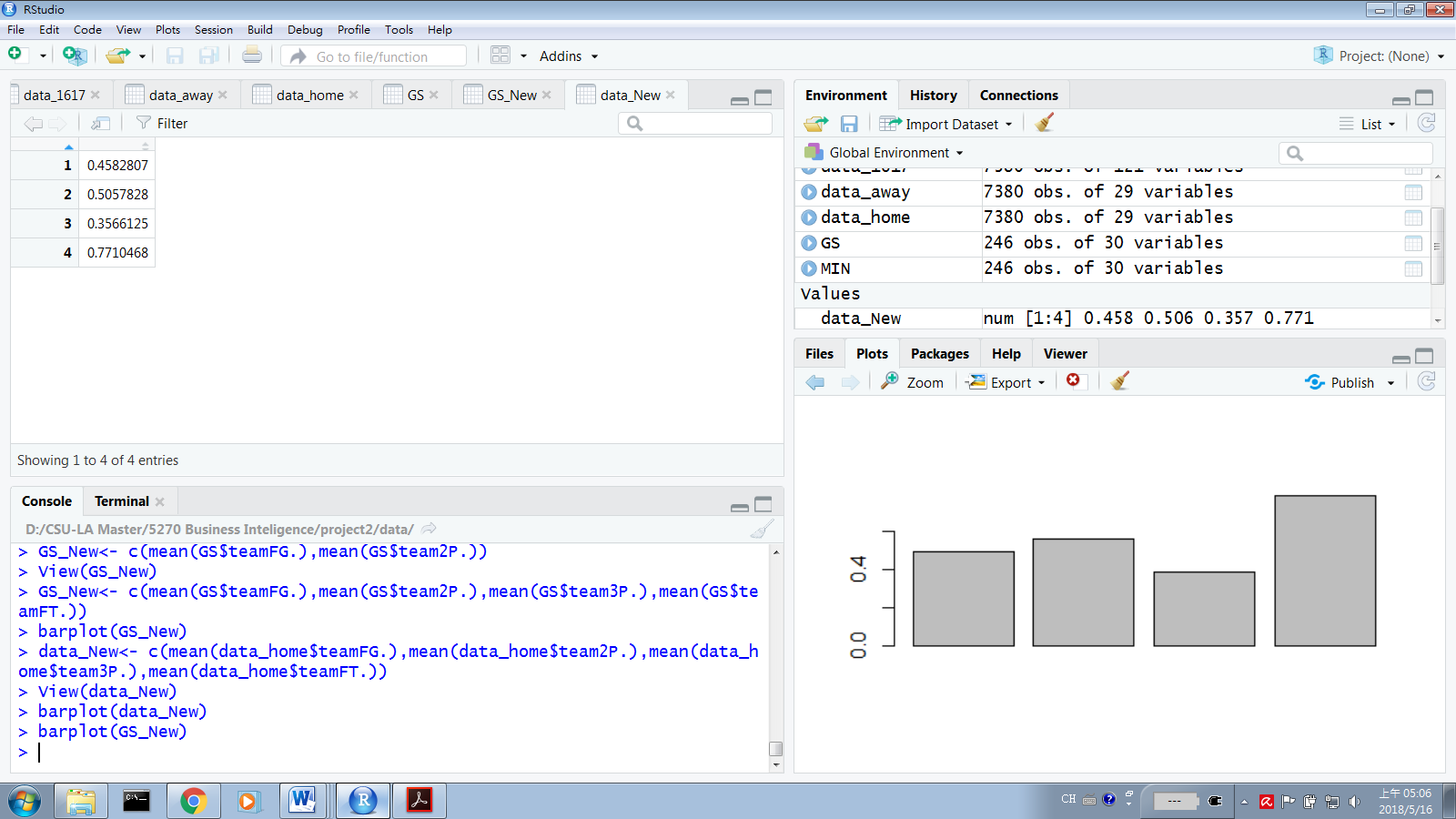
1. Different between the champion and the other teams



Every team has it attribute. For example, Golden State Warriors is good at shooting 3 points. The scatter plot showing that those games they won, which are the red spots, because they shoot more threes. On the other hand we can see in the plot that they have shoot less 2 points compare with the league.

1. What are the attribute did the champion have.

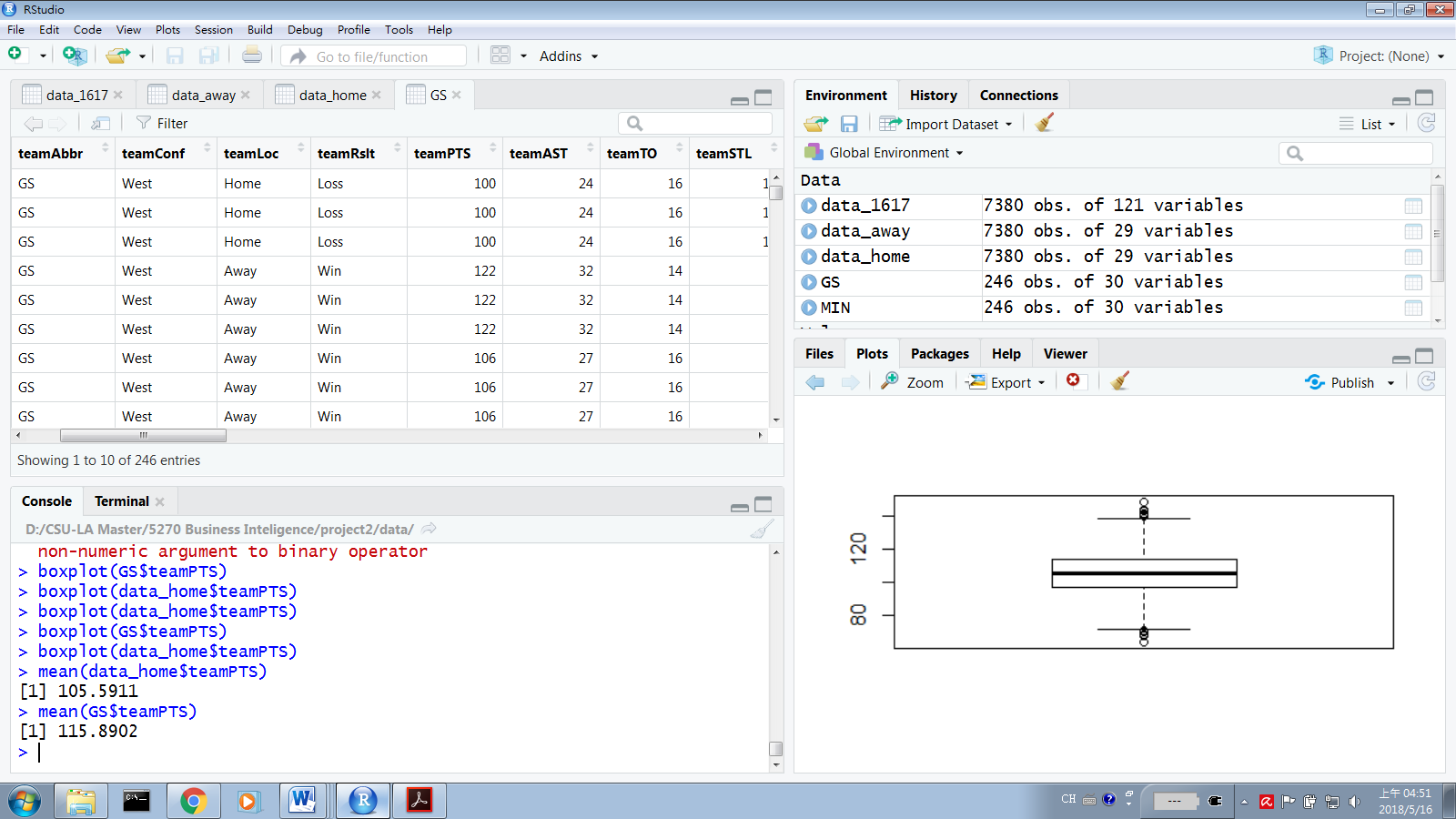




Compare with other teams and Golden State Warriors, all of the number are better than the league. As the result, the champion has better 2 points and 3 points ability, less turnover and more free throw shooting.

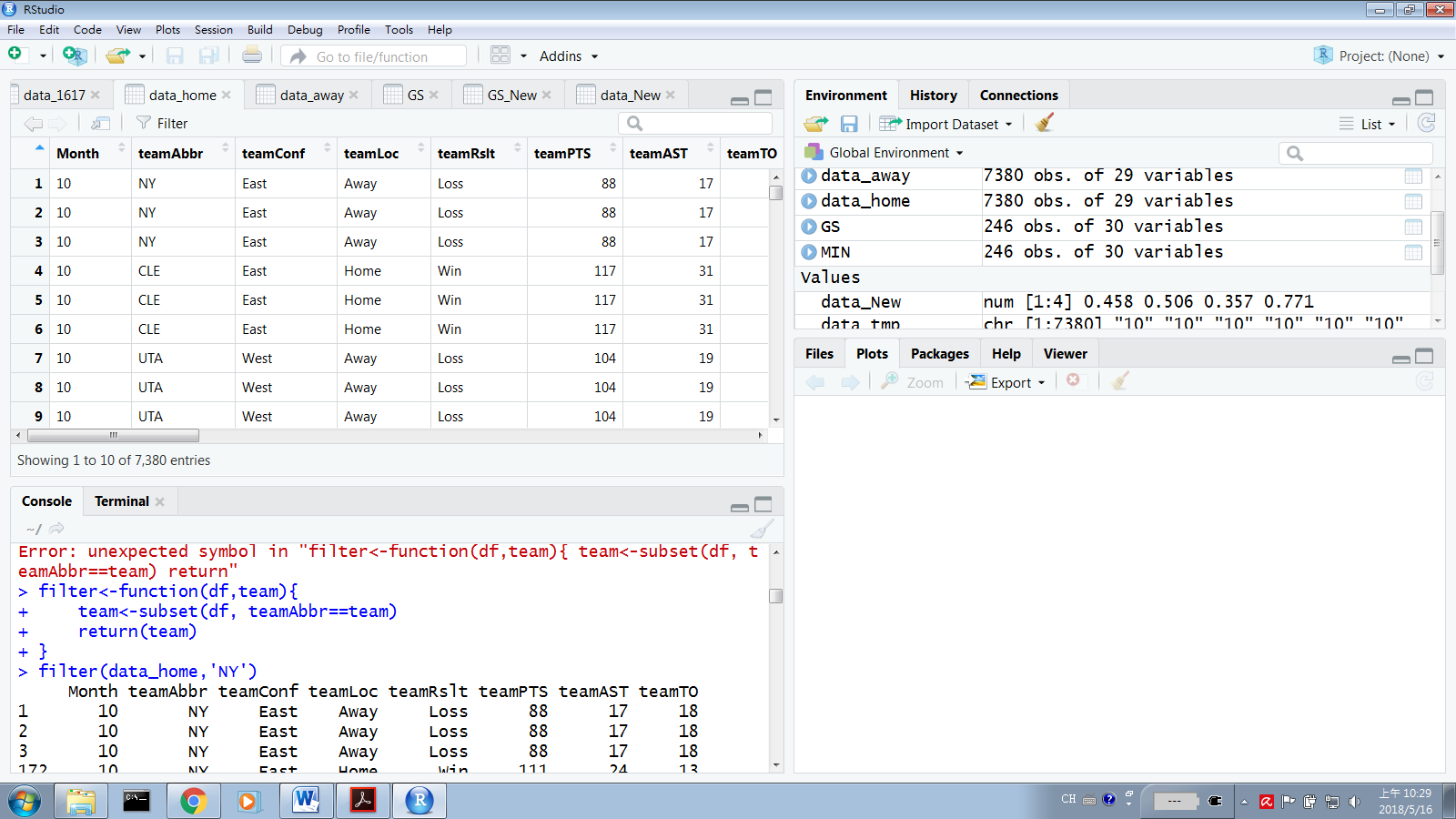
Statistics & Function

1. Mean



The point they got are similar to the rest of teams. As the result, the ability of getting point is similar for each team.

1. Function



This function helps people to separate all data from a specific team. In this case, the function filter all the data about New York Nicks.

Reference

1. “Average TV viewership of NBA finals games in United States from 2002 to 2017(in millions)”, retrieved from <https://www.statista.com/statistics/240377/nba-finals-tv-viewership-in-the-united-states/>
2. Dan Kopf “Data analysis have made the NBA unrecognizable”, Oct 18, 2017 retrieved from <https://qz.com/1104922/data-analytics-have-revolutionized-the-nba/>
3. “Researchers Dive Into Big Data to Predict NBA Winners”, Apr 27, 2016 retrieved from <https://www.ucmerced.edu/news/2016/researchers-dive-big-data-predict-nba-winners>

Code

data\_1617<- separate(data\_1617,gmDate,c(“Year”,”Month”,”day”),sep=”-”)

data\_home<- data\_1617[,c(“Month”,”teamAbbr”,”teamRslt”,”teamPTS”,”teamAST”,”teamTO”,”teamSTL”,”teamBLK”,”teamPF”,”teamFGA”,”teamFGM”,”teamFG”,”team2PA”, ”team2PM”, ”team2P”, ”team3PA”, ”team3PM”, ”team3p”, ”teamFTA”, ”teamFTM”, ”teamFT”, ”teamORB”, ”teamPTS1”, ”teamPTS2”, ”teamPTS3”, ”teamPTS4”)]

Box\_1617<-subset(Box\_1617, seasTyp==”Regular”)

Attach(GS)

GS$Colour = “black”

GS$Color[GS$teamRslt==”win”]=red

Plot(team2PM, team3PM)

Boxplot(GS$teamPTS)

Mean(GS$teamPTS)

GS\_New<-c(mean(GS$teamFG.), mean(GS$team2P.), mean(GS$team3P.), mean(GS$teamFT.))

Barplot(GS\_New)

Filter<-function(df,team){

Team<-subset(df, teamAbbr==team)

Return(team)  
}