Data Wrangling Summary

Curtis Higa

November 8, 2017

Project Abstract

The goal of this project is to find a correlation between different three-point statistics and win percentage throughout a NBA regular season.

Wrangling the Data

An online database called http://www.mysportsfeeds.com/ allows users to pull numerous sports statistics from their API and into their chosen programming language. To pull this data in R, they've posted a github repository and package that includes a function that helps specify what data you're looking for. I've chosen to take the gamelogs of each NBA team for the 2016-2017 regular season. I then created the following function to help go through all 30 teams and pull their gamelogs from the API.

```
nba_gamelog <- function (nba_team) {
  team_gamelog <- msf_get_results(league='nba',
                                  season='2016-2017-regular',
                                  feed='team_gamelogs',
                                  params=list(team=nba_team))
  return(team_gamelog)
}

t_gamelog_df <- t_gamelog$api_json$teamgamelogs$gamelogs
```

The last line of code is used for pulling the actual data frame from their API and is the source of my data for this project.

To make things simpler, I renamed all the relevant columns to shorter and concise names. After pulling this data, the total number of points needed to be calculated per each team for each game. All relevant scoring columns were multiplied and added up per standard NBA scoring rules using the *mutate* function.

Next was to use a *for* loop to consolidate every team gamelog into one data frame. I created a vector of all 30 NBA teams, an empty initialized data frame, and a dummy column to add to the data frame to store the team abbreviation corresponding to the data. Using the following *merge* function, every time a team gets run through the for loop, it gets appended to the data frame **all_gamelogs**.

```
all_gamelogs <- merge(all_gamelogs,relevant,all=TRUE)
```