William Howard

DSS 665 R Statistical Programming

Marcus Ellis

April 25, 2019

Final Project

**Introduction/Proposal**

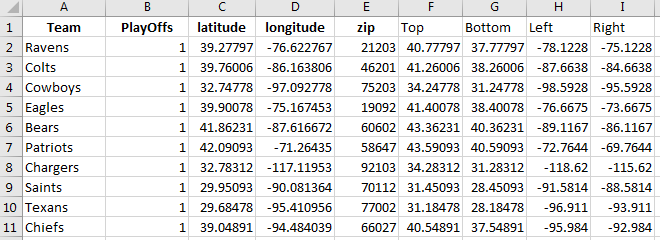
The National Football League is United States based professional sports league consisting of 32 teams. The teams are divided into two conferences contain four divisions each. Playoff teams are selected by taking the eight conference champions and two additional wild card teams. The wildcard teams are selected by taking the taking the two remain teams with the best record. Over the last decade fantasy football has also taken off in both popularity among fans and as a true marketable business segment. With organizations such as draft kings and fan duel to name a few. Leading the way in online sports betting, based on individual player fantasy points totals. What I set out to do was determine a few questions regarding the effects on geography and fantasy points have on a team’s chances of being a part of the playoffs at the end of the year.

**Data Sets**

I was able to create a few different data sets for this project. The first was to google search the longitude and latitude values of each of the thirty-two teams’ stadiums. I will use this to map each stadium geographical location on map. The next data set is called evalsPlayers A data set that provides average draft king salaries and points for the 2018 season. The third data sets are the evalsPosition data set. This data set compares salary and point average for 2018 based on their associated position. My final data set is evalsTeam. This data set provides much of the same information of salaries and point averages for the 2018 at the team level.

evalsLocation

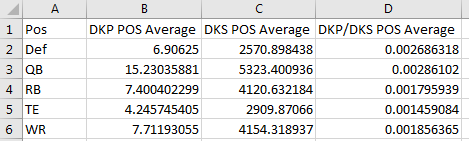
As mentioned, this data set provides geographical information for all thirty-two teams stadium locations and if the team made the playoffs or not. The number 1 represents a team that made the playoffs and the number 2 represents a team that did not.



The categories Top, Bottom, Left, and Right represent a value in addition to the longitude and latitude value. This will be used to create the area covered by markers in my graph. Since there are only going to thirty points on the map, I chose to use 1.5 to add or subtract from longitude and latitude to create a more clearly visible point.

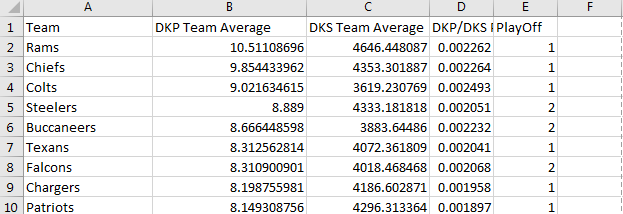
evalsPosition

The data set evalsPosition provides insight on the Draft Kings fantasy points weekly average for each position scored. As well as the average Draft Kings salaries per week for each position. Also, I thought it would be interesting to determine the average dollar spent per fantasy point earned.



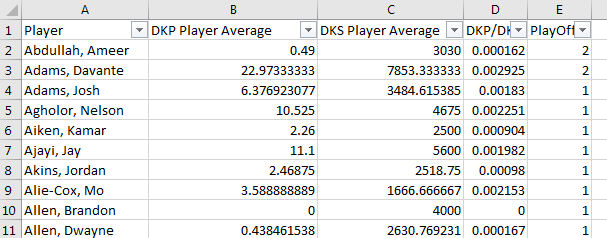
evalsTeam

The data set Evalteam provides insight on the Draft Kings fantasy points weekly average for each team scored. As well as the average Draft Kings salaries per week for each team. Also, I thought it would be interesting to determine the average dollar spent per fantasy point earned.



evalsPlayer

The data set evalPlayer provides insight on the Draft Kings fantasy points weekly average for each player scored. As well as the average Draft Kings salaries per week for each player. Also, I thought it would be interesting to determine the average dollar spent per fantasy point earned.



**Analysis**

To start my analysis, it was important to install and load a few packages.

install.packages(c("maps", "mapdata"))

install.packages(c("ggplot2", "devtools", "dplyr", "stringr"))

install.packages('ggmap')

install.packages('Rtools')

install.packages('officer')

install.packages('magrittr')

install.packages("RColorBrewer")

library(Rtools)

library(officer)

library(magrittr)

library(ggmap)

library(maps)

library(mapdata)

library(dplyr)

library(RColorBrewer)

As well as load my excel data sets.

read.csv("E:\\SJU MS Analytics\\DSS 665 - R Statistical Language\\Week 3\\evals.csv",header=TRUE)

attach(evals)

evals

evalsLocation <- read.csv("E:\\SJU MS Analytics\\DSS 665 - R Statistical Language\\Week 3\\evalsLocation.csv",header=TRUE)

attach(evalsLocation)

evalsLocation

evalsPlayer <- read.csv("E:\\SJU MS Analytics\\DSS 665 - R Statistical Language\\Week 3\\evalsPlayer.csv",header=TRUE)

attach(evalsPlayer)

evalsPLayer

evalsPosition <- read.csv("E:\\SJU MS Analytics\\DSS 665 - R Statistical Language\\Week 3\\evalsPosition.csv",header=TRUE)

attach(evalsPosition)

evalsPLayer

evalsTeam <- read.csv("E:\\SJU MS Analytics\\DSS 665 - R Statistical Language\\Week 3\\evalsTeam.csv",header=TRUE)

attach(evalsTeam)

evalsTeam

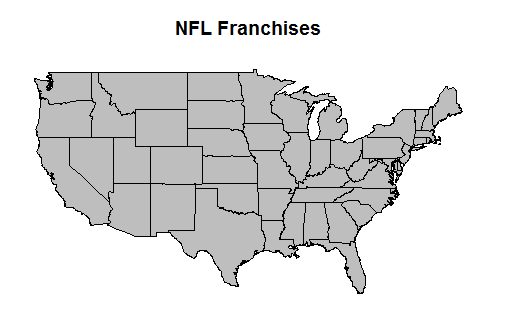
Maps

After loading in the packages and required data sets. I began testing my availability to build a map showing the continental United States.

state <- map\_data("state")

map("state", fill = TRUE, col = "grey", namesonly = TRUE)

title("NFL Franchises")



I then began calling in the longitude and latitude for each team and displaying a dot on the map

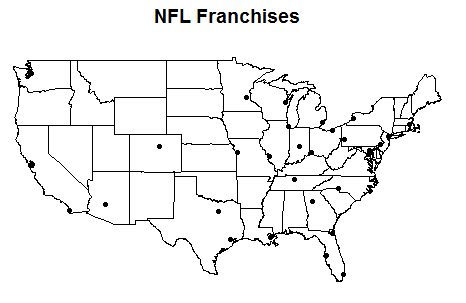
map("state", fill = TRUE, col = "white", namesonly = TRUE)

title("NFL Franchises")

for(i in 1:32){

points(longitude[i],latitude[i],pch=20)

}



As you see each of the locations for each NFL team’s stadium is located on the map by a black dot. My initial thoughts are a large percentage of the teams are located east of the Mississippi river . Next, I wanted to see the location of all twelve NFL playoff teams’ stadiums. This will be displayed on the map as yellow dots.

playoffmap = map("state", fill = FALSE, col = "GREY", namesonly = TRUE)

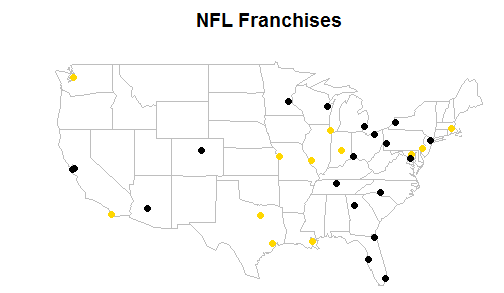
title("NFL Franchises")

for(i in 1:32){

if(PlayOffs[i]==1)points(longitude[i],latitude[i],pch=20,cex=1.5,col="Gold")

if(PlayOffs[i]==2)points(longitude[i],latitude[i],pch=20,cex=1.5,col="black")

}



All in all, I was surprised at the number of teams that were located west of the Mississippi. When compared to where most teams are located. After some deeper digging and research of the data I discovered, as mentioned in the intro, that the NFL is divided into two conference and four divisions. With each divisions champion, no matter how good or bad their record, given a play off spot. As well as two wild card spots awarded to the top teams remaining. The divisions are also somewhat based on the geographic locations of the team. This could explain why the location of playoff teams is proportionate to the overall location of all the teams. Upon further reflection I think a better approach to the data would be to plot the location of the number one seeds over the past x amount of years as well as the location of the wildcard seeds over the same amount of years. To determine if a team’s location, warm or cold state, provides an advantage when perusing the overall number one seed or wildcard seeds.

Plots

In the second part of my analyses I set out to determine if there was a relationship between the weekly average of a team’s players salaries and the weekly average of a team’s players performance.

library(ggplot2)

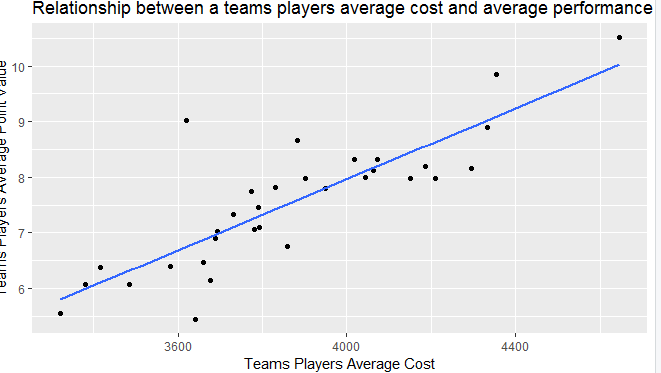
ggplot(evalsTeam, aes(x = DKS.Team.Average, y = DKP.Team.Average)) +

geom\_point() +

labs(x = "Teams Players Average Cost", y = "Teams Players Average Point Value",

title = "Relationship between a team’s players average cost and average performance") +

geom\_smooth(method = "lm", se = FALSE)



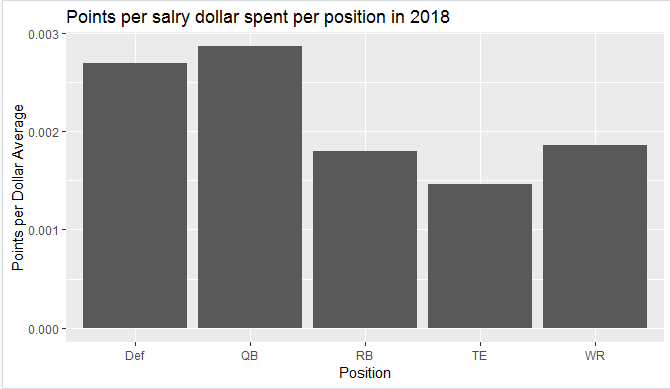
As the graph shows there is a relationship between the two variables. I was a little surprised by there weren’t more outliers, maybe if I would have performed this at the player level there would have been. As a first-time fantasy football participant last year and novice fan up to that point. I originally felt there was a certain value placed on a player brand or past performance. Players such as Tom Brady would have a few down weeks, yet their salary/cost ranked among the highest in the league. The data however, at the team, level only shows a few out liars and that on average the cost/salary of the players on those teams returns an equal/fair return of points.

The next analysis I performed was to compare the average dollar per point return amount. To determine what the cost per fantasy point was at the position level.

ggplot(evalsPosition, aes(x=Position, y=DKP.DKS.POS.Average)) + geom\_bar(stat = "identity") +

labs(x = "Position", y = "Points per Dollar Average",

title = "Points per salry dollar spent per position in 2018")



As the chart shows the QB position provides the highest return of points per dollar spent. This was as I expected the offense, largest points scoring category. Runs largely through the QB in the NFL for instance the QB gains points not only for their actions such as rushing and passing but also that of the WR and TE. Provided they pass them the ball. For instance, if the QB passes the ball to a WR and they run 15 additional yards after the catch for a touchdown. The QB receives points for the number of yards the pass went to completion, the additional yards the receiver ran, and for the receive scoring the touchdown. The biggest surprise of this graph is that defense return the second highest value per dollar. Even though the defense receives points for scoring, interceptions, and sacks. It is extremely difficult for these players to score. An explanation for this return could be defensive players are grouped in terms of cost and points as one holistic unit and the salary usually far more reasonable when compared to offensive players.

**New Package**

For a new package to utilize I chose officer. This package allows for commination between R and Microsoft Office products such as word and PowerPoint. I originally chose and different package and made a last-minute shift, big mistake. I wanted to try this package because of the new team I was tasked with leading at work. Our job is to move throughout the business and automate different teams’ monthly dashboards through vba in different Office products such as Excel, PowerPoint, and SharePoint. I started by loading in a few more packages.

install.packages("tidyverse")

install.packages("officer")

install.packages("rvg")

install.packages("gdtools")

library("tidyverse")

library("officer")

library("rvg")

library("gdtools")

reading in and establishing my word doc.

read\_docx()

my\_doc <- read\_docx()

styles\_info(my\_doc)

Recreating one of my maps to import as a picture.

src <- tempfile(fileext = ".png")

png(filename = src, width = 5, height = 4, units = 'in', res = 300)

playoffmap = map("state", fill = FALSE, col = "GREY", namesonly = TRUE)

title("NFL Franchises")

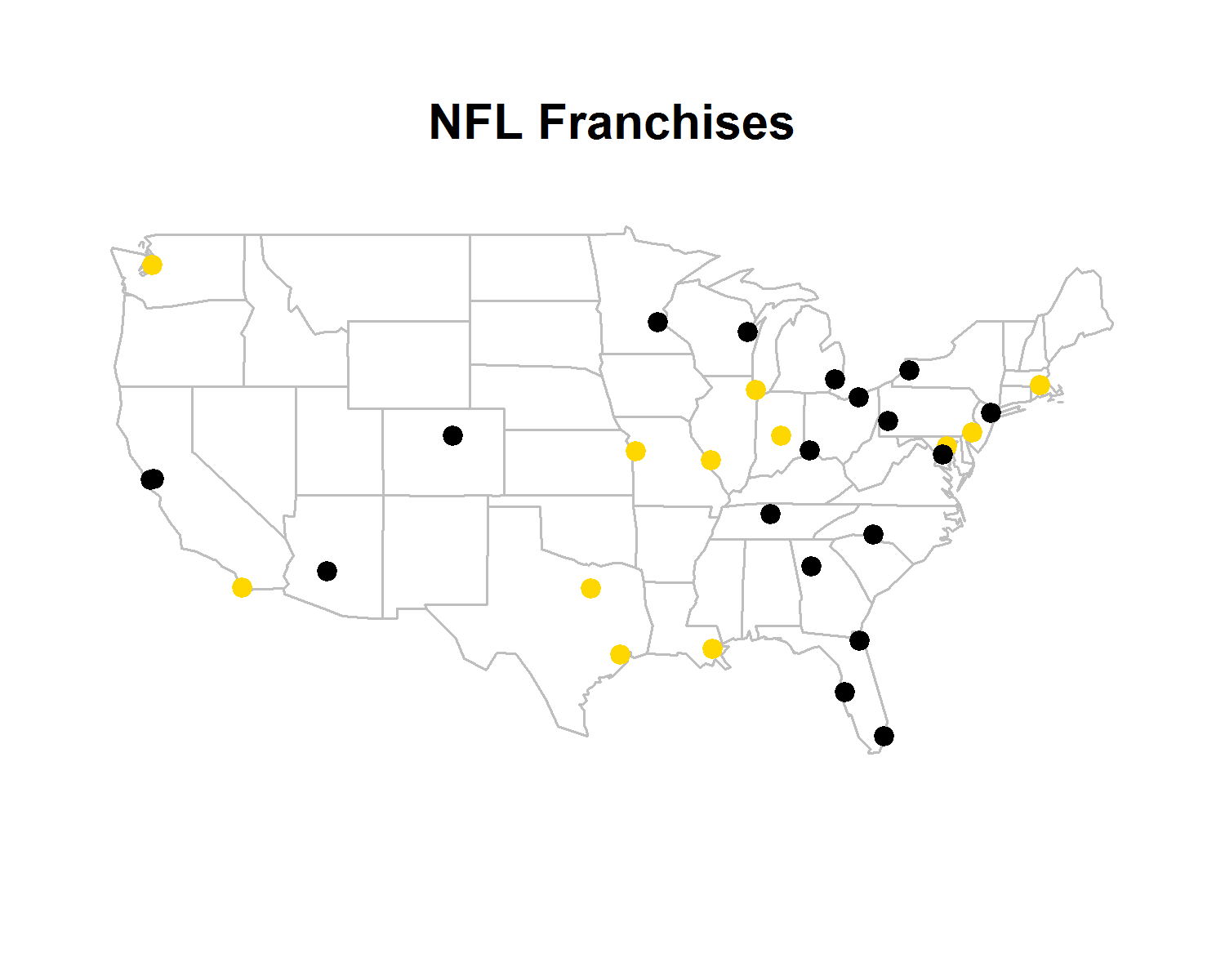
for(i in 1:32){

if(PlayOffs[i]==1)points(longitude[i],latitude[i],pch=20,cex=1.5,col="Gold")

if(PlayOffs[i]==2)points(longitude[i],latitude[i],pch=20,cex=1.5,col="black")

}

dev.off()



Then creating the header and opening paragraph you see at the beginning of the paper.

my\_doc <- my\_doc %>%

body\_add\_img(src = src, width = 5, height = 3, style = "centered") %>%

body\_add\_par("William Howard", style = "Normal") %>%

body\_add\_par("DSS 665 R Statistical Programming", style = "Normal") %>%

body\_add\_par("Marcus Ellis", style = "Normal") %>%

body\_add\_par("April 25, 2019", style = "Normal") %>%

body\_add\_par("", style = "Normal") %>%

body\_add\_par("Final Project", style = "centered") %>%

body\_add\_par("", style = "Normal") %>%

body\_add\_par("Introduction/Proposal", style = "Normal") %>%

body\_add\_par("", style = "Normal") %>%

body\_add\_par(" The National Football League is United States based professional sports league consisting of 32 teams.

The teams are dividied into two conferences contain four divisions each. Playoff teams are selected by

taking the eight conference champions and two additional wild card teams. The wildcard teams are selected by taking

the taking the two remain teams with the best record. Over the last decade fantasy football has also taken off in both

popularity among fans and as a true marketable business segment. With organizations such as draft kings and fan duel to name a few.

Leading the way in online sports betting, based on individual player fantasy points totals. What I set out to do was determine a few questions regarding

the effects on geography and fantasy points have on a team’s chances of being a part of the playoffs at the end of the year.") %>%

body\_add\_par("", style = "Normal") %>%

body\_add\_par("Data Sets", style = "Normal") %>%

body\_add\_par("", style = "Normal") %>%

body\_add\_par("I was able to create a few different data sets for this project. The first was to google search the longitude and latitude

values of each of the thirty-two teams’ stadiums. I will use this to map each stadium geographical location on map. The next data set is called evalsPlayers

A data set that provides average draft king salaries and points for the 2018 season. The third data sets are the evalsPosition data set.

This data set compares salary and point average for 2018 based on their associated position. My final data set is evalsTeam. This data

set provides much of the same information of salaries and point averages for the 2018 at the team leve.", style = "Normal") %>%

body\_add\_par("", style = "Normal") %>% # blank paragraph

**Conclusion**

At the conclusion of this project I feel I answered the questions I set out to answer. Unfortunately, I don’t feel the best answer was provided. As mentioned by determining if a team’s stadium location played a role in if that team made the playoffs, I negated the fact that playoff teams are partially determined by automatic bids from geography determined confreres. A better question would have been if the wildcard or number one overall seed showed a relation between geographical area of the team’s stadium since this data would be less dependent on the team’s conference automatic bid and more on their overall record in the league. Interestingly though was to see the relationship and almost “free” market depiction of average points to average salary relationship at both the team and position level. Also, the officer package was very helpful, even with my little experience. I wish I had more time to develop a PowerPoint presentation with the package or understand its ability to write the whole paper in word from R.