Linear Algebra and its Application to Fantasy Sports

Introduction:

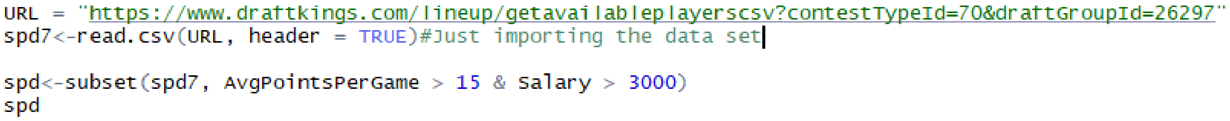
The structure of daily fantasy points is that the person with the most points within a parameter of salary is declared the winner. Declaring that variable x = the salary of the player; it would be most advantageous to optimize variable y = fantasy points. Although, there is of course many variables that contribute to how well an individual player does per game such as if the game is a back to back, or the defensive ranking of the opposition, for simplicity this variable y will be estimated by average fantasy points per game over the course of the season.

How Linear Algebra Can Solve the Problem:

Although, the actual dataset consisting of each players salaries and average points per game is complex and difficult to make meaningful analyzation from, we can use the least squares problem to answer valuable questions. Such as what are the average points per game one should be expecting at a certain salary point? As well as being able to interpret what players are valuable by having significant regression from the line of linear regression.

Using the standard equation for approximating the least squares problem: Ax=b. In this scenario our A is an m x n matrix consisting of the set of column vectors which individually define the players points, salaries, and names. The vector b will be a within the space R^m. Finally, x will be the scaling vector of the matrix A that will give the best approximation of b. The goal is to use Rstudio to find the vector x that makes the square root of (b-Ax) as small as possible, while also making sure that Ax is the closest point to b in the colA. In Rstudio we can easily find the Ax vector and the total visualization is shown below.

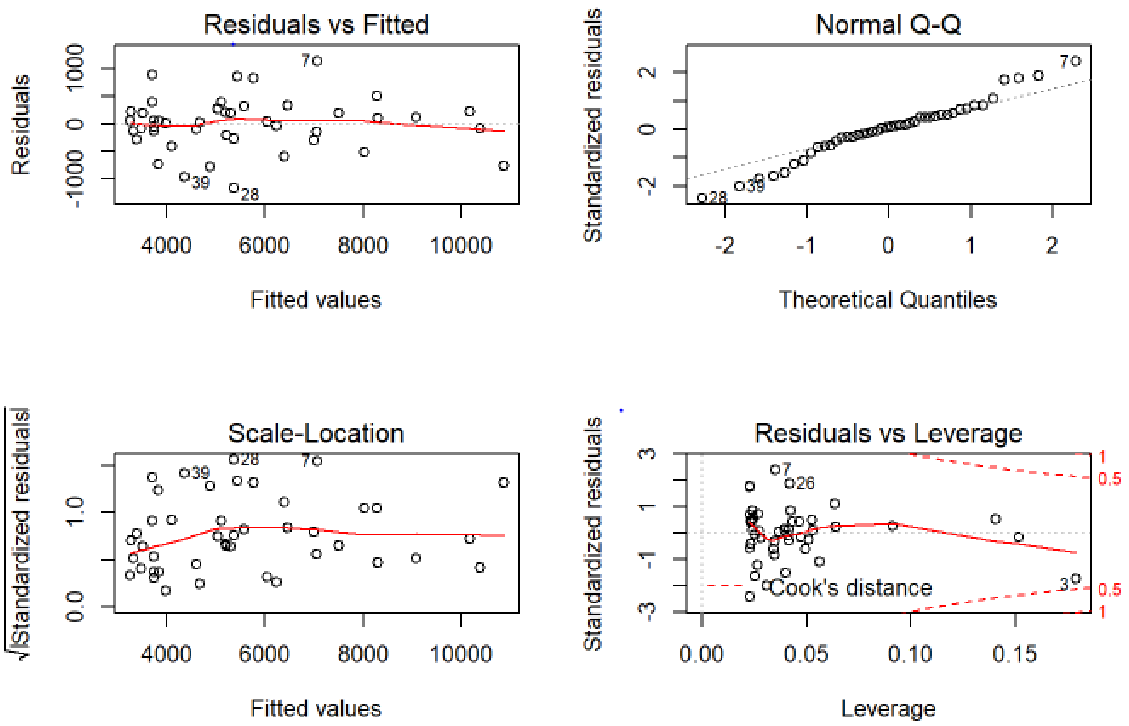
One of the benefits of simplifying the regression down to salary and average points per game is that data is easily accessible for every lineup regardless of the games playing or even the sport. The data comes in a simple matrix that is ready for use except for the fact that it includes injured players and players that hardly play, so those players need to be removed from the matrix so that the data isn’t skewed.



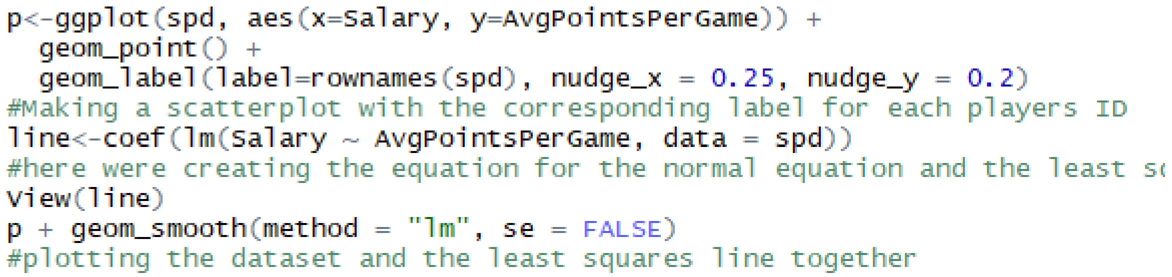
```
URL = "https://www.draftkings.com/lineup/getavailableplayerscsv?contestTypeId=70&draftGroupId=26297"
spd7<-read.csv(URL, header = TRUE)#Just importing the data set

spd<-subset(spd7, AvgPointsPerGame > 15 & Salary > 3000)
spd
```

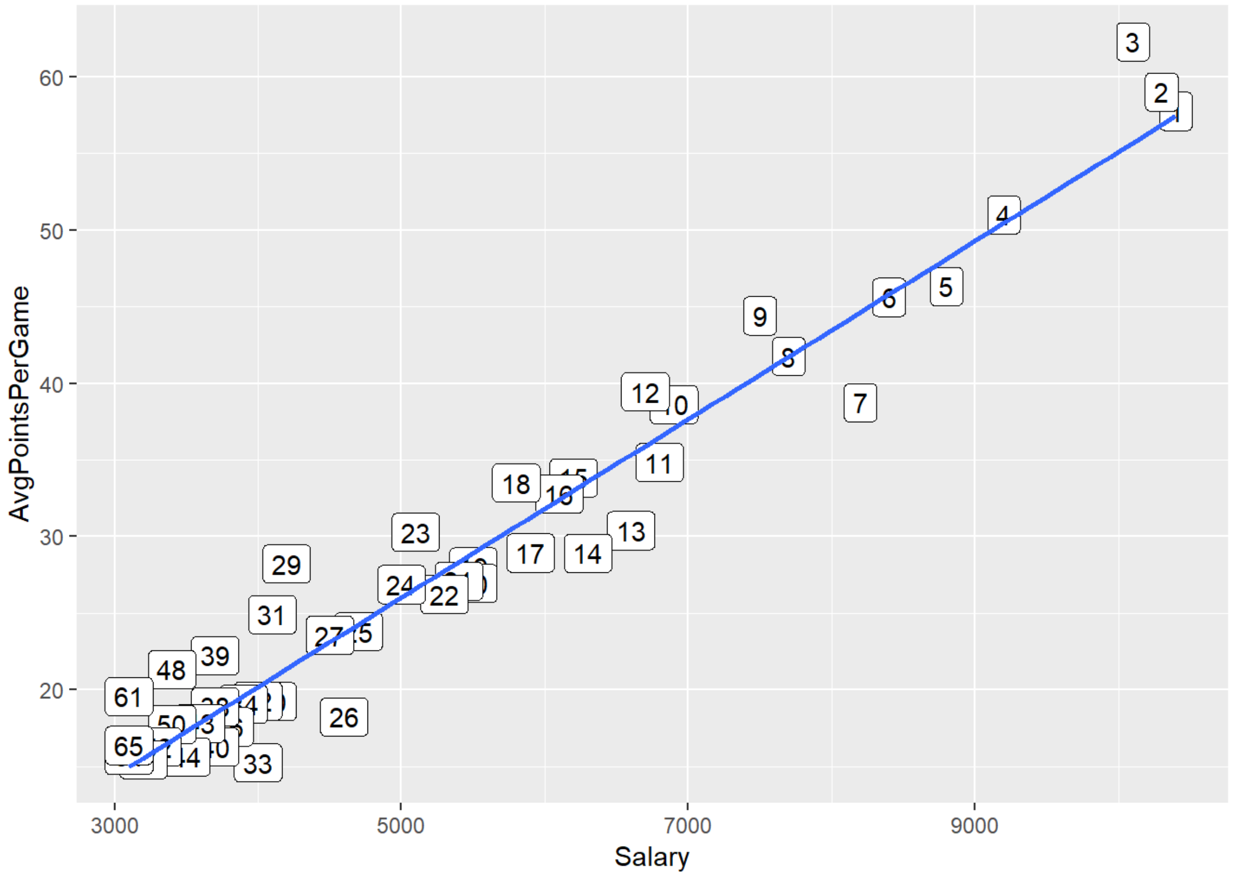
Next we check normality assumption and make sure that one residual point isn’t altering the whole data, which it shouldn’t because of the previous step.



From there we use the new data set to solve the least squares problem.

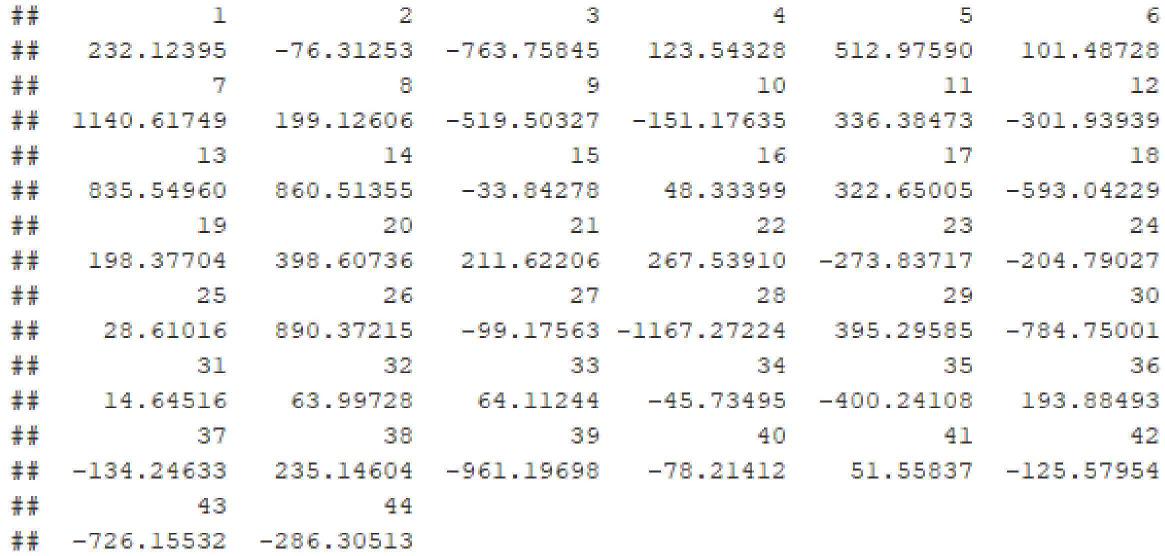


```
p<-ggplot(spd, aes(x=Salary, y=AvgPointsPerGame)) +
  geom_point() +
  geom_label(label=rownames(spd), nudge_x = 0.25, nudge_y = 0.2)
#Making a scatterplot with the corresponding label for each players ID
line<-coef(lm(Salary ~ AvgPointsPerGame, data = spd))
#here were creating the equation for the normal equation and the least s
View(line)
p + geom_smooth(method = "lm", se = FALSE)
#plotting the dataset and the least squares line together
```



The result from the least squares problem gives an extremely well fit line which is expected as the sites that formulate the salaries of players don't want to make one player significantly better or else everyone will pick that player. This is also shown in the correlation between the two variables of salary and average points per game is 0.9717 which is an extremely high correlation.

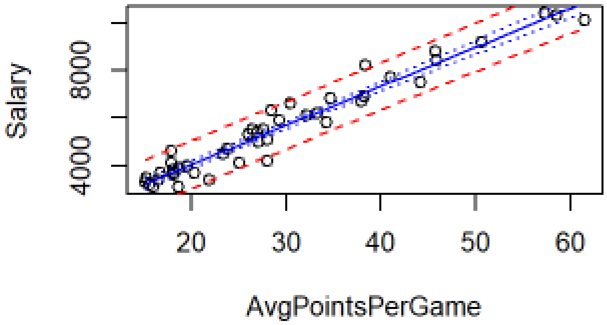
The best way to get an advantage over other players in fantasy sports is to find the individual players that deviate most from the least squares line. So, while most uses of the least squares problem attempt to find an x that alters the matrix a in order to find b, the interest of this linear regression is to find the points that deviate most from Ax=b. Since the average player follows the line of linear regression, the players whose true value deviates most by the predicting equation Ax=b are reasonably expected to have value different from that which the salary is indicating.



```
##           1           2           3           4           5           6
##   232.12395   -76.31253  -763.75845   123.54328   512.97590   101.48728
##           7           8           9          10          11          12
##  1140.61749   199.12606  -519.50327  -151.17635   336.38473  -301.93939
##          13          14          15          16          17          18
##   835.54960   860.51355   -33.84278    48.33399   322.65005  -593.04229
##          19          20          21          22          23          24
##   198.37704   398.60736   211.62206   267.53910  -273.83717  -204.79027
##          25          26          27          28          29          30
##    28.61016   890.37215   -99.17563 -1167.27224   395.29585  -784.75001
##          31          32          33          34          35          36
##    14.64516    63.99728    64.11244   -45.73495  -400.24108   193.88493
##          37          38          39          40          41          42
##  -134.24633   235.14604  -961.19698   -78.21412    51.55837  -125.57954
##          43          44
##  -726.15532  -286.30513
```

Looking at this regression chart above the players with the highest absolute value are the ones that deviate most from the regression line. By using this regression table it can estimate which players are expected to have the worst game and which ones can be expected to have the best game. As a player making a lineup you may choose to have a high ownership of players with very negative regression such as number three, while avoiding players with very positive regression numbers like the athlete number 14. By looking at this regression chart it is helpful to see which players are poorly predicted by their salary.

In a visual representation below a player would get the biggest advantage by picking players outside of the blue lines(the confidence interval) and if possible making athletes outside of the red lines (the confidence interval) athletes of interest.


Salary
AvgPointsPerGame