Stat 167 Final Project

NBAnalysis

By
Brandon Kim
Dinh Bui
Kaiyu Sun
Rachel Santiago
Mustafa Abdel

Research Questions

1) How has the pace of NBA games changed over the years?
    a) How has an increase in pace contributed to the change in scoring and number of 3-pointers?
    b) Did an increase in pace have an effect on the different positions in basketball?
    c) Is this new style of play efficient?
2) What impact does team location have on its **players**?
    a) Does higher attendance lead to higher average salary?
3) What impact does team location have on the **team**?
    a) Overall franchise win percentage

What are these stats?

3PA = 3 pointers attempted.

3PAR = 3 pointers/field goals attempted.

PPG = Points per game.

Pace = Possessions per 48 minutes.

PER = Player efficiency rating.

USG% = Percentage of team plays for a player while on the floor.

Coded Positions

1 = Point Guard (Average height = 6’3’’ and below)

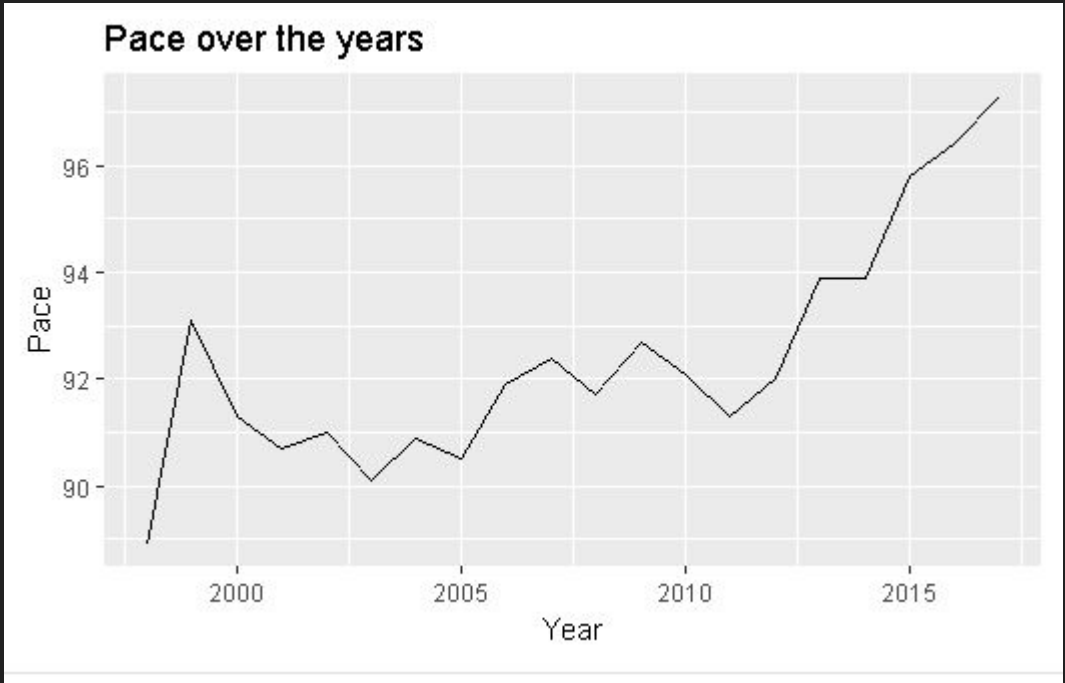
2 = Shooting Guard (Average height = 6’3’’ to 6’7’’)

3 = Small Forward (Average height = 6’6’’ to 6’10’’)

4 = Power Forward (Average height = 6’8’’ to 7’0’’)

5 = Center (Average height = 7’0’’)

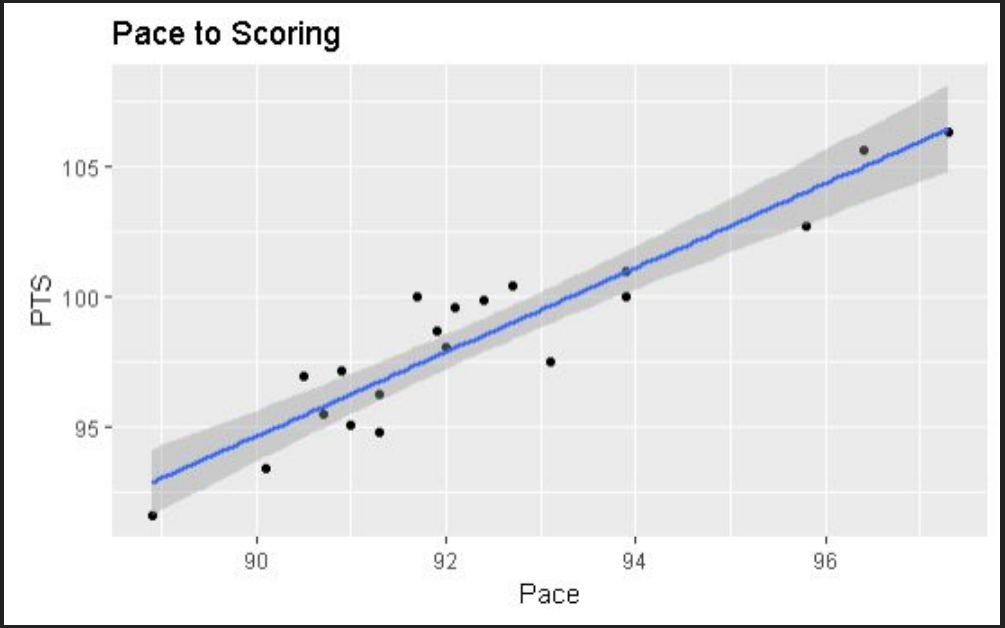
How has average pace changed over the years?



Pace = Possessions per 48 minutes

We can see a strong positive increase in pace when we plot the trend line for average league pace over the last 15+ years.

How has an increase in pace contributed to scoring?

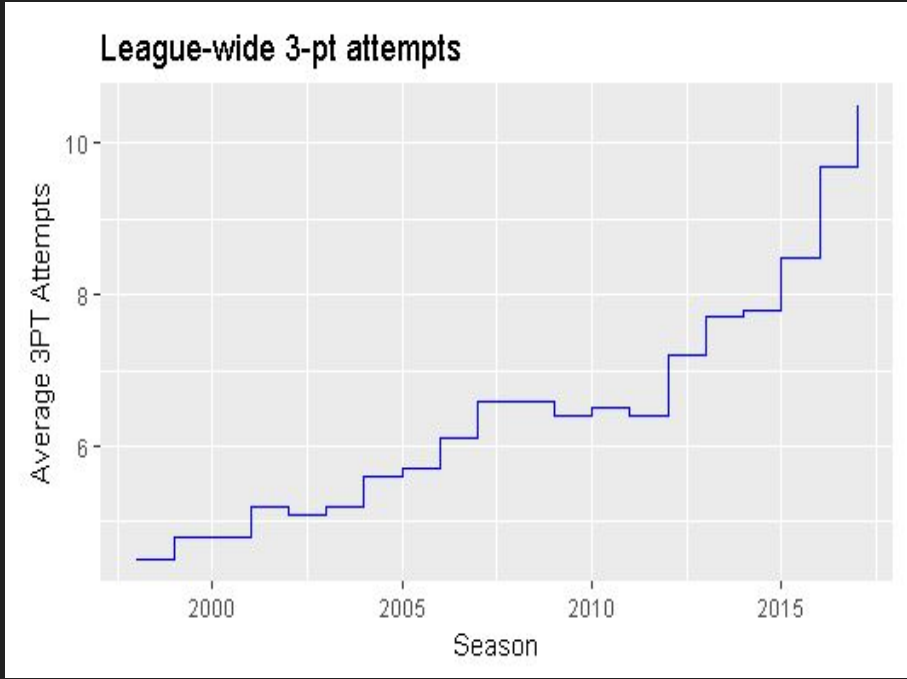


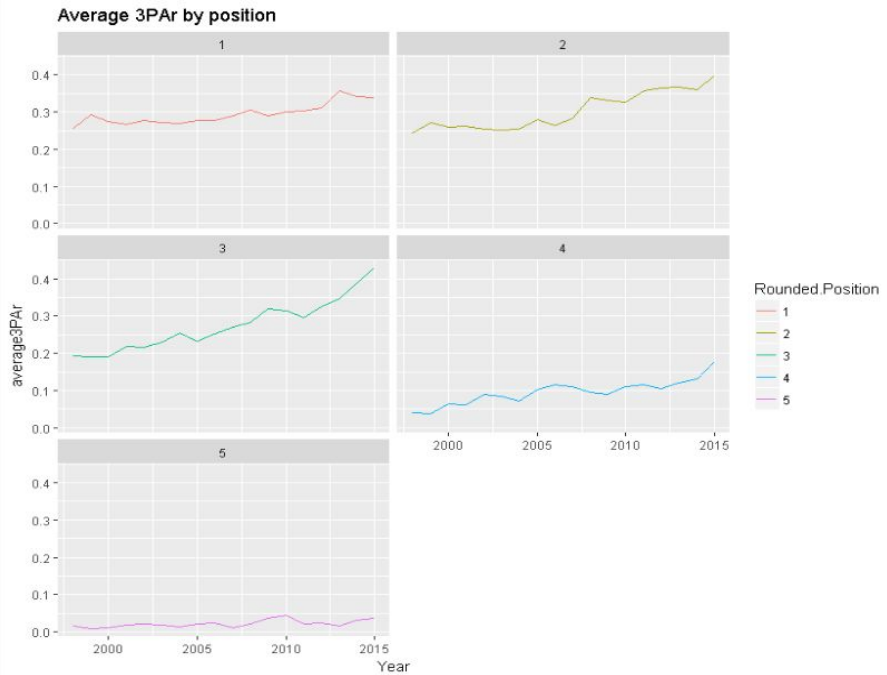

```
Residual standard error: 0.7986 on 18 degrees of freedom
Multiple R-squared:   0.87,     Adjusted R-squared:  0.8628
F-statistic: 120.5 on 1 and 18 DF,  p-value: 2.095e-09
```

It is estimated for every unit increase in average pace, average ppg increases by around .147.

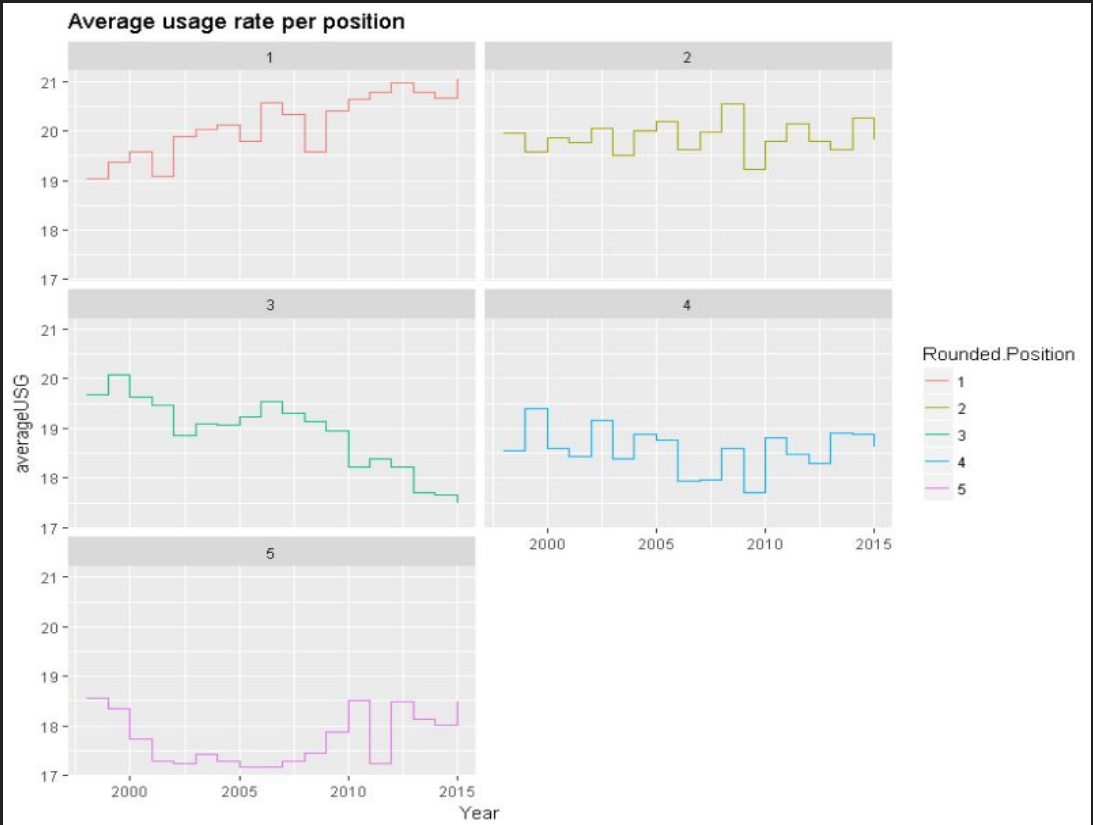
The higher the pace of the game means the higher the box score numbers which makes for exciting basketball.

How about the 3-point shots?





How are different positions being utilized over the years of increased pace?



Usage Rate: estimated percent of total plays ran for that single player for that team.

Notable positions:

- 1 (Point Guards)
- 3 (Small Forwards)

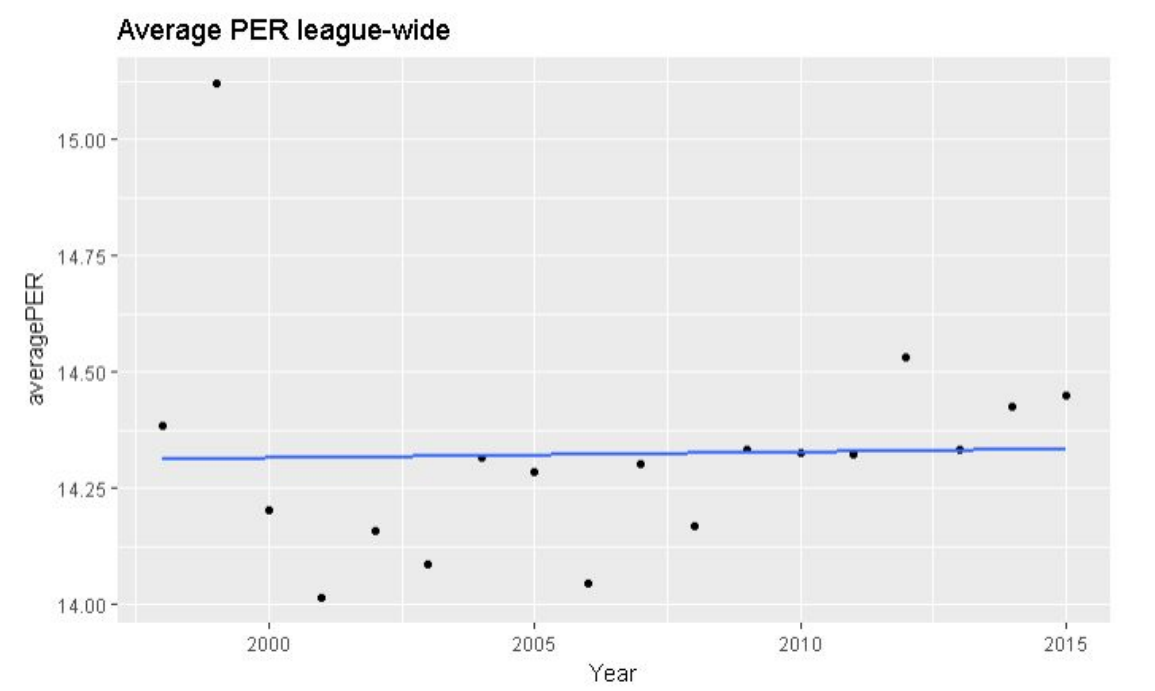
Is this faster style of play efficient?

```
uPER = (1 / MP) *
     [ 3P
     + (2/3) * AST
     + (2 - factor * (team_AST / team_FG)) * FG
     + (FT *0.5 * (1 + (1 - (team_AST / team_FG)) + (2/3) * (team_AST / team_FG)))
     - VOP * TOV
     - VOP * DRB% * (FGA - FG)
     - VOP * 0.44 * (0.44 + (0.56 * DRB%)) * (FTA - FT)
     + VOP * (1 - DRB%) * (TRB - ORB)
     + VOP * DRB% * ORB
     + VOP * STL
     + VOP * DRB% * BLK
     - PF * ((lg_FT / lg_PF) - 0.44 * (lg_FTA / lg_PF) * VOP) ]
```

What exactly is PER?

Player Efficiency Rating: The PER sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance.

Average league PER over the years

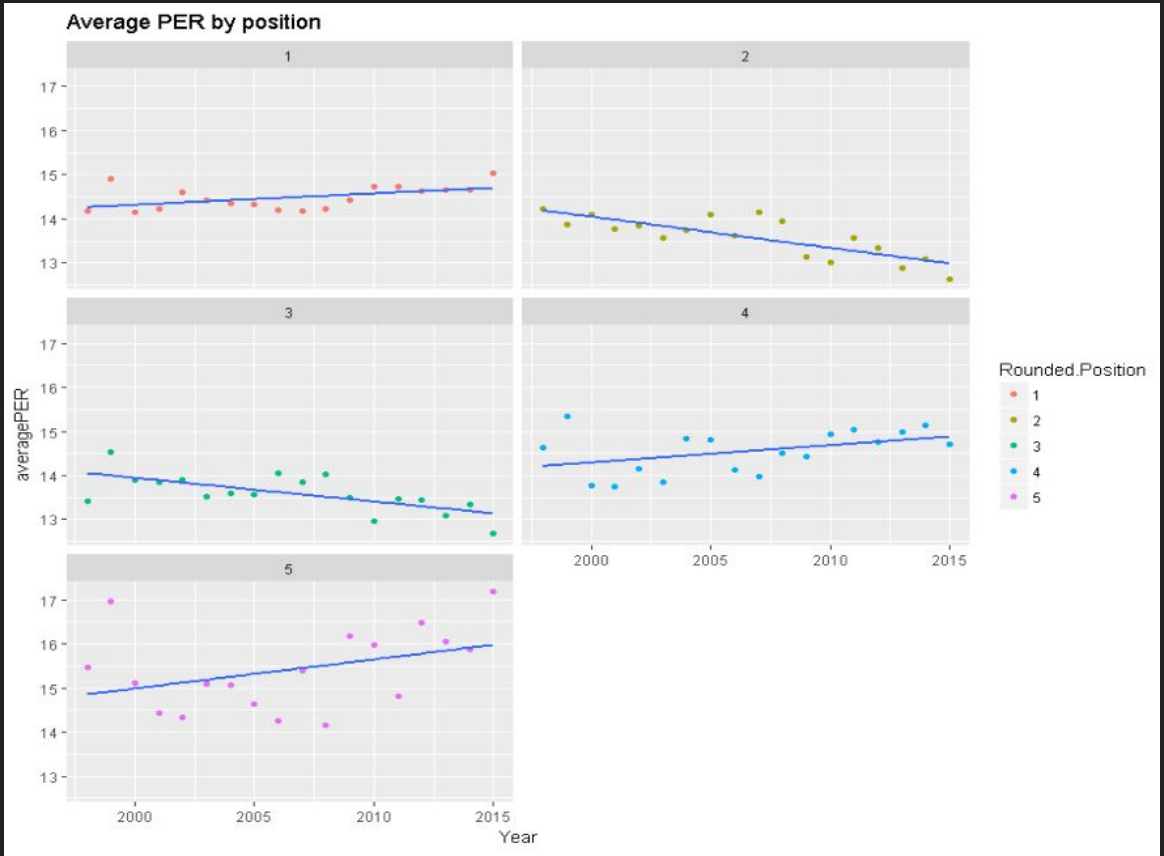



```
Residual standard error: 0.2509 on 16 degrees of freedom
Multiple R-squared:  0.000792,  Adjusted R-squared:  -0.06166
F-statistic: 0.01268 on 1 and 16 DF,  p-value: 0.9117
```

Summary:
When taking the average PER for all 5 positions and plotting it over the same stretch of years we saw increased pace, we did not see a linear relationship.

This means efficiency did not take a hit or gain while the league was transitioning towards a faster paced playstyle.

Average PER by position



Linear Regression for all 5 positions:

For each increase in year, average pace increased by an estimated

(1) .0257 for point guards (R-sq=25.21%)

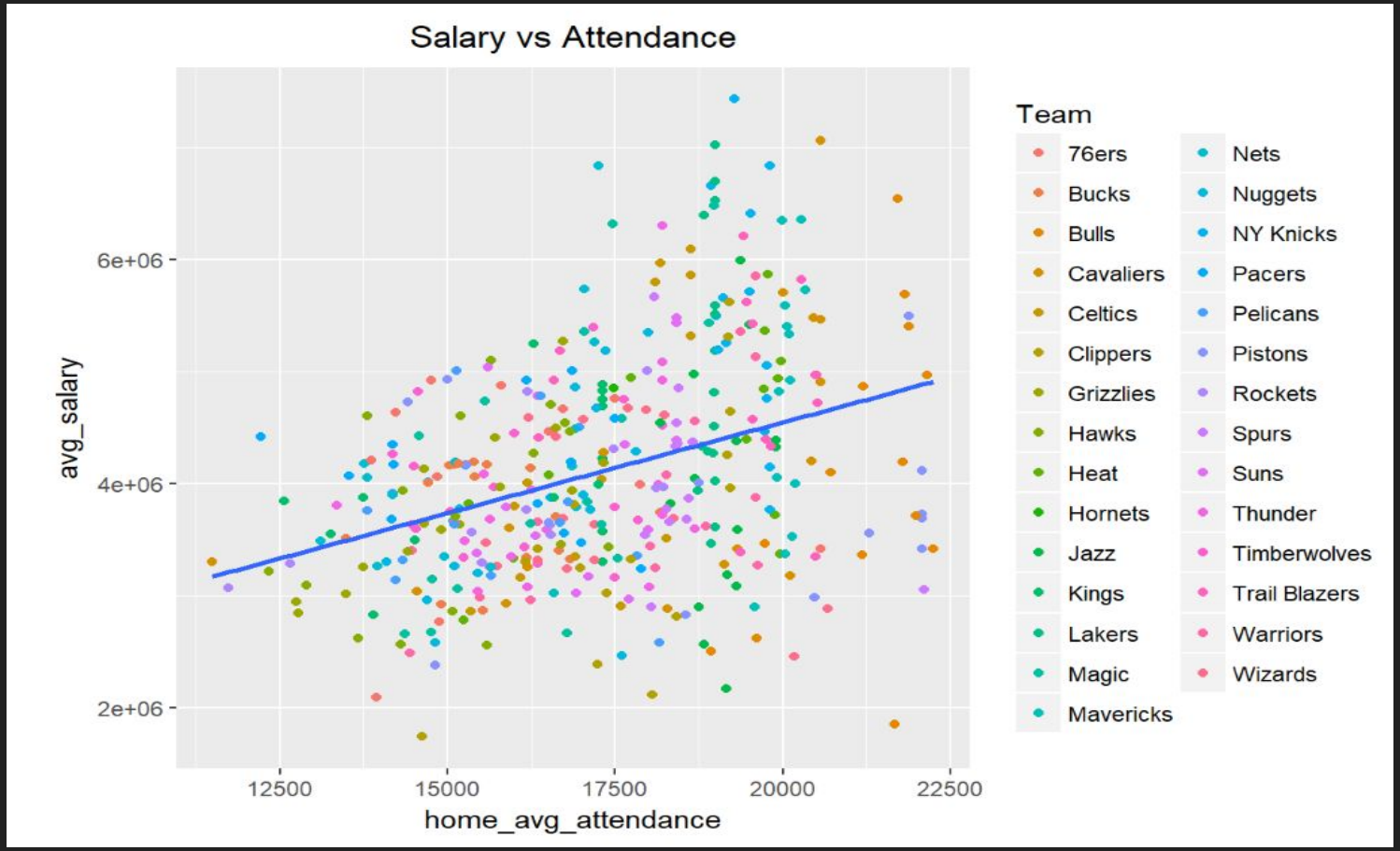
(2) -.071 for shooting guards (R-sq=63.57%)

(3) -.0537 for small forwards (R-sq=17.52%)

(4) Model not significant for power forwards.

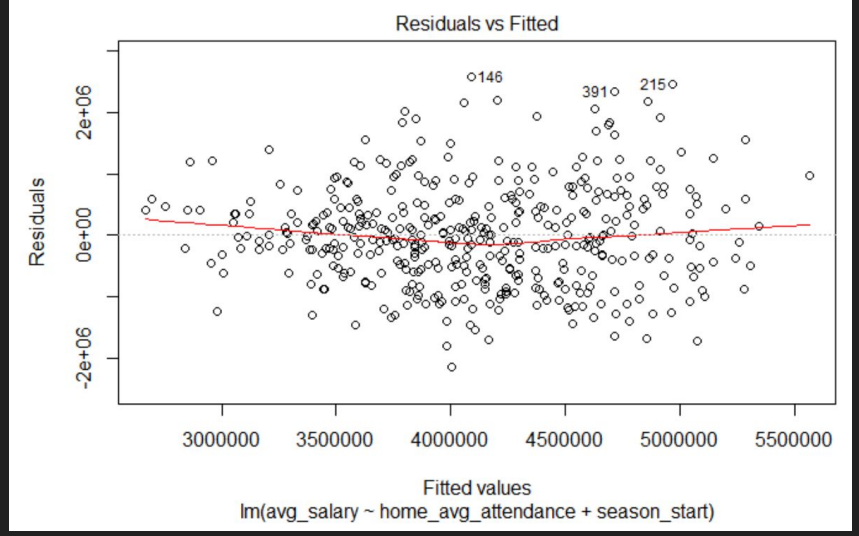
(5) Model not significant for centers.

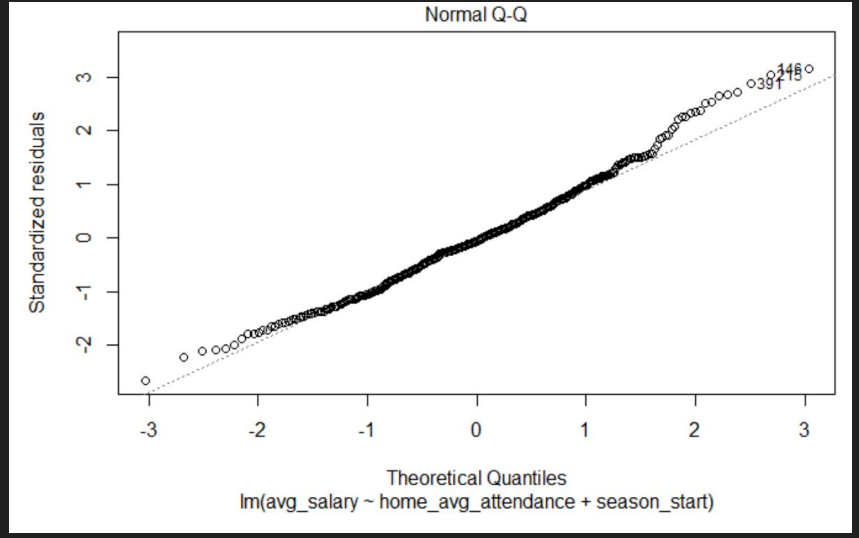
Average Salary vs Average Home Attendance



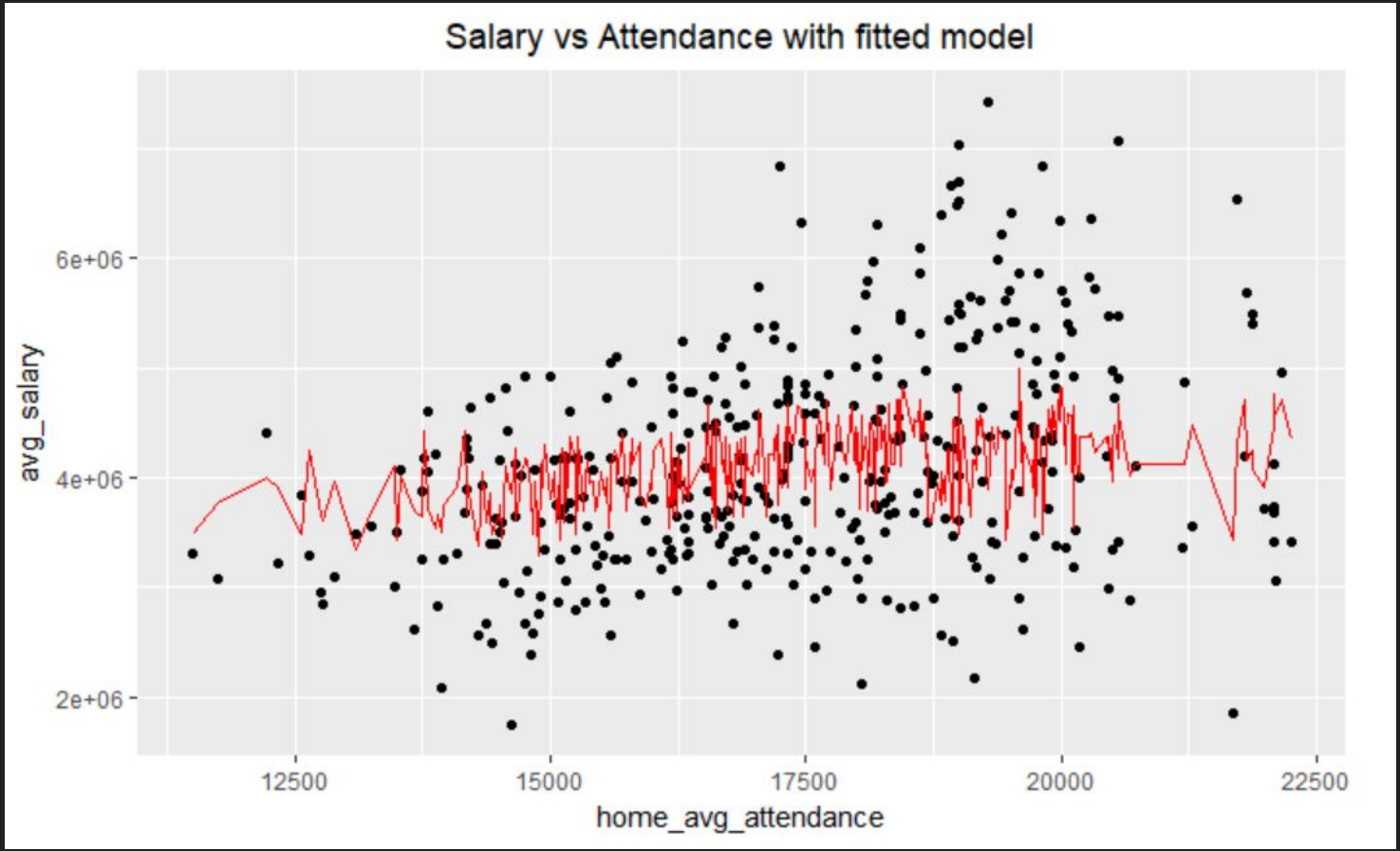
MLR Analysis

```
Call:
lm(formula = avg_salary ~ home_avg_attendance + season_start,
    data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-2146909  -563294   -49610   472568  2573130

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         8.586e+05  3.511e+05   2.445 0.014901 *
home_avg_attendance 1.451e+02  1.888e+01   7.682 1.23e-13 ***
season_start2001    1.049e+05  2.337e+05   0.449 0.653625
season_start2002    3.807e+05  2.355e+05   1.617 0.106769
season_start2003    3.473e+05  2.378e+05   1.460 0.144970
season_start2004    5.168e+05  2.383e+05   2.169 0.030707 *
season_start2005    4.869e+05  2.366e+05   2.058 0.040213 *
season_start2006    3.212e+05  2.498e+05   1.286 0.199248
season_start2007    9.157e+05  2.413e+05   3.795 0.000171 ***
season_start2008    1.312e+06  2.364e+05   5.552 5.19e-08 ***
season_start2009    1.245e+06  2.319e+05   5.370 1.34e-07 ***
season_start2010    1.079e+06  2.384e+05   4.527 7.91e-06 ***
season_start2011    9.726e+05  2.322e+05   4.189 3.45e-05 ***
season_start2012    1.014e+06  2.386e+05   4.250 2.67e-05 ***
season_start2013    1.557e+06  2.323e+05   6.701 7.12e-11 ***
season_start2014    5.157e+05  2.346e+05   2.198 0.028519 *
season_start2015    8.753e+05  2.311e+05   3.787 0.000176 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 831300 on 397 degrees of freedom
Multiple R-squared:  0.3262,    Adjusted R-squared:  0.299
F-statistic: 12.01 on 16 and 397 DF,  p-value: < 2.2e-16
```

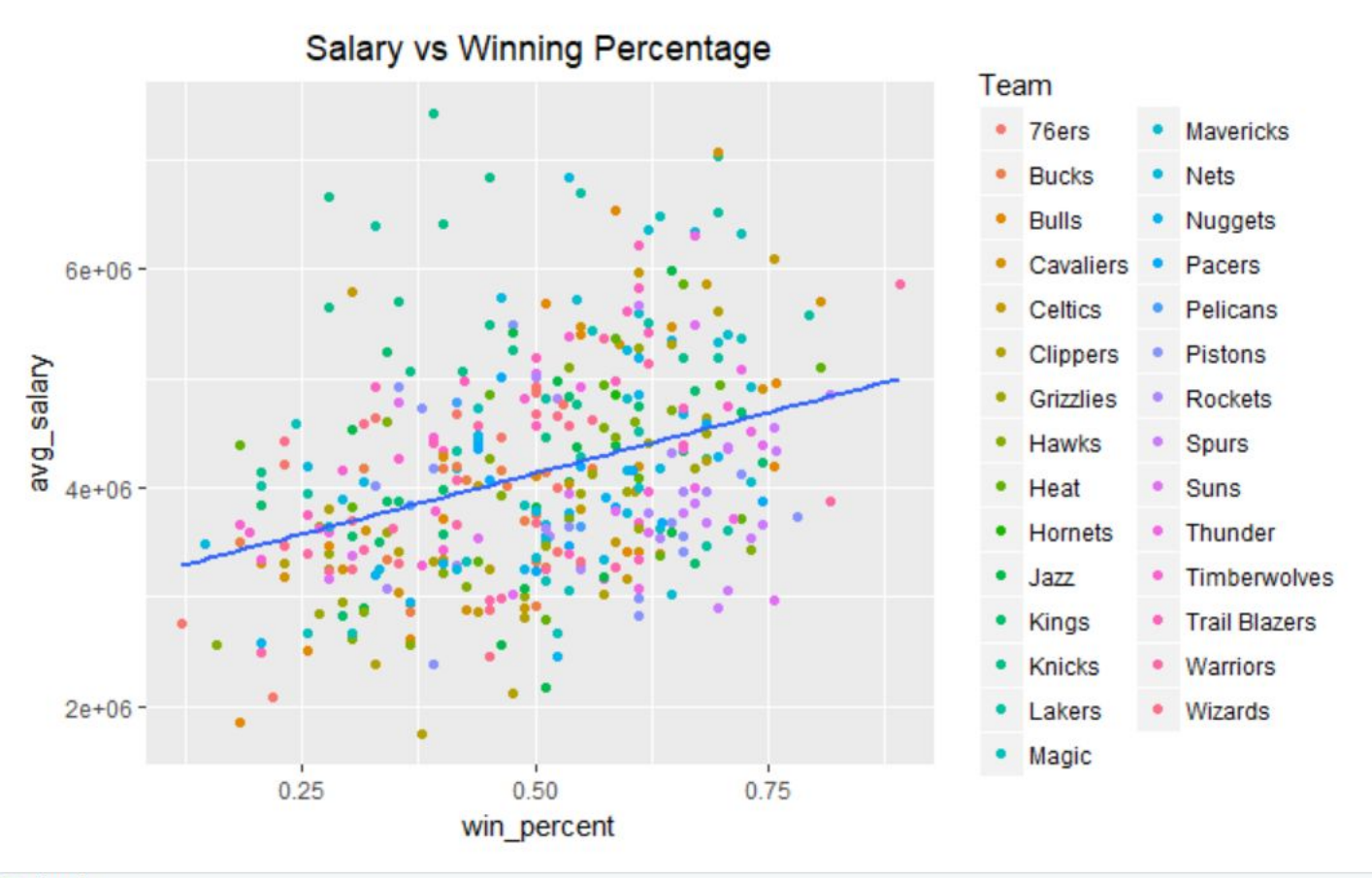




Avg Salary vs Avg Home Attendance with Fitted Model

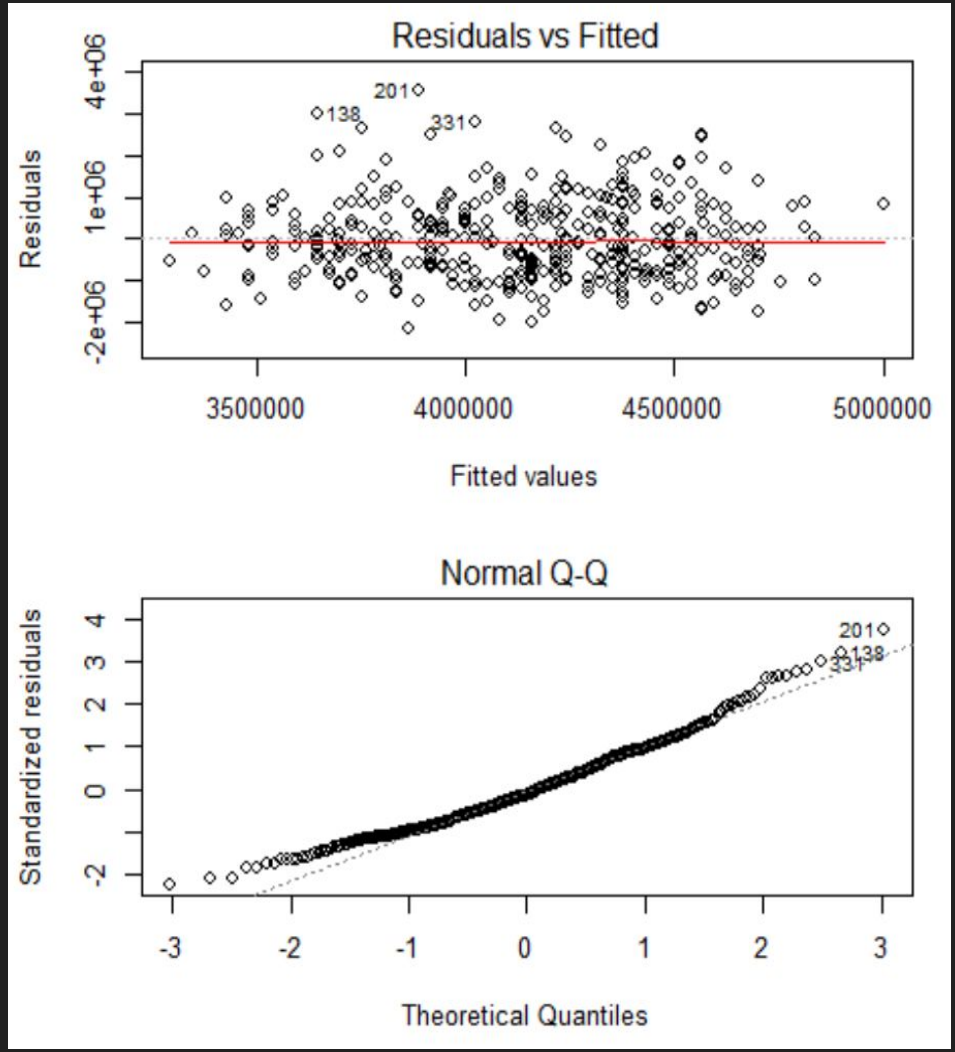


Average Salary vs Winning Percentage

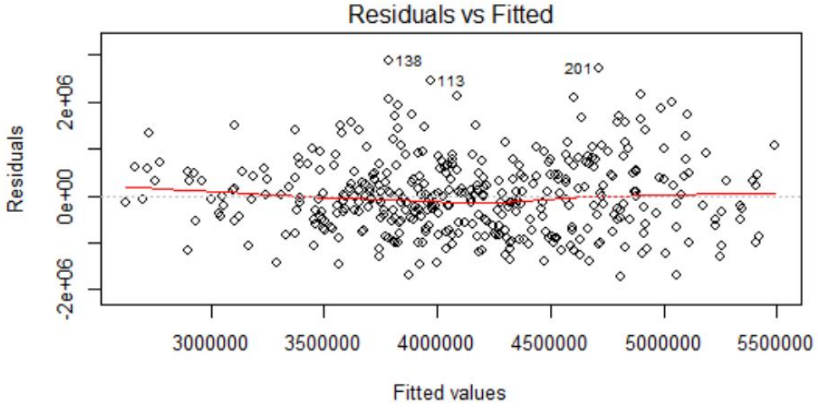


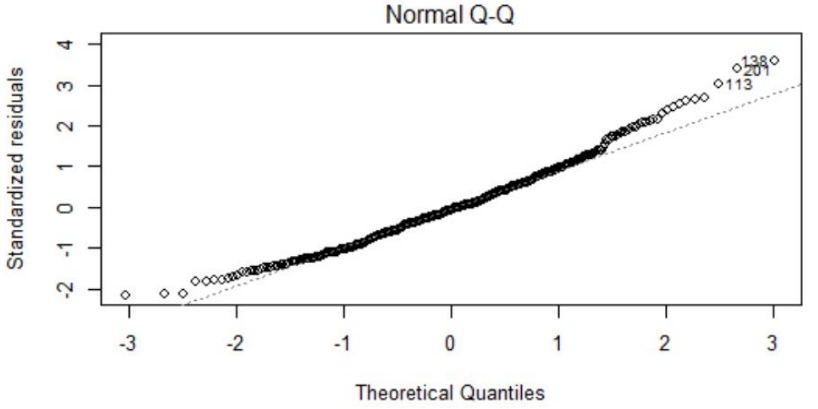
MLR Analysis

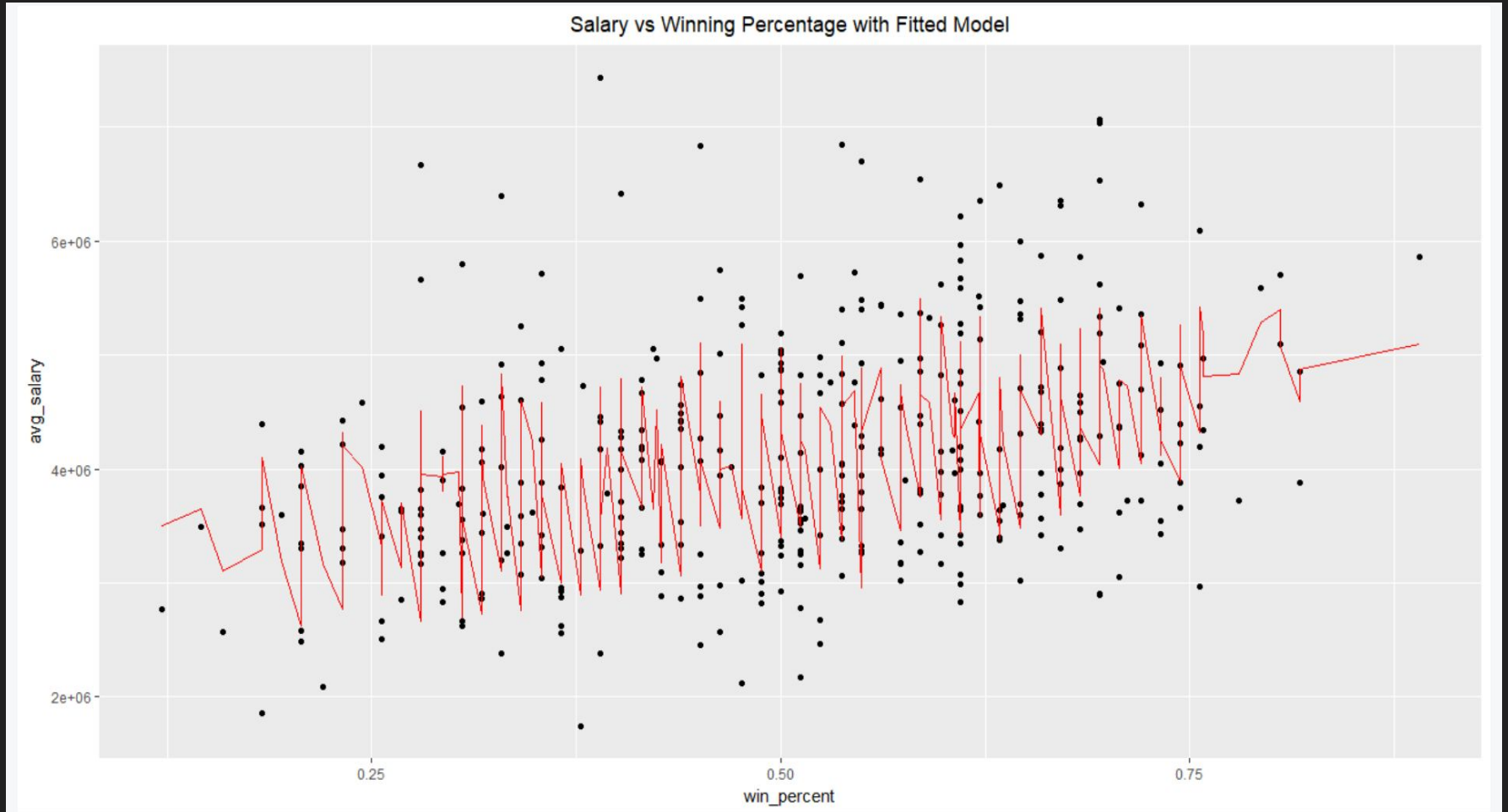
```
Call:
lm(formula = avg_salary ~ win_percent, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-2119739  -717510  -134569   627447  3546120

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3019378     162337  18.599  < 2e-16 ***
win_percent  2227254     309068   7.206 2.96e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 944800 on 393 degrees of freedom
Multiple R-squared:  0.1167,    Adjusted R-squared:  0.1145
F-statistic: 51.93 on 1 and 393 DF,  p-value: 2.961e-12
```



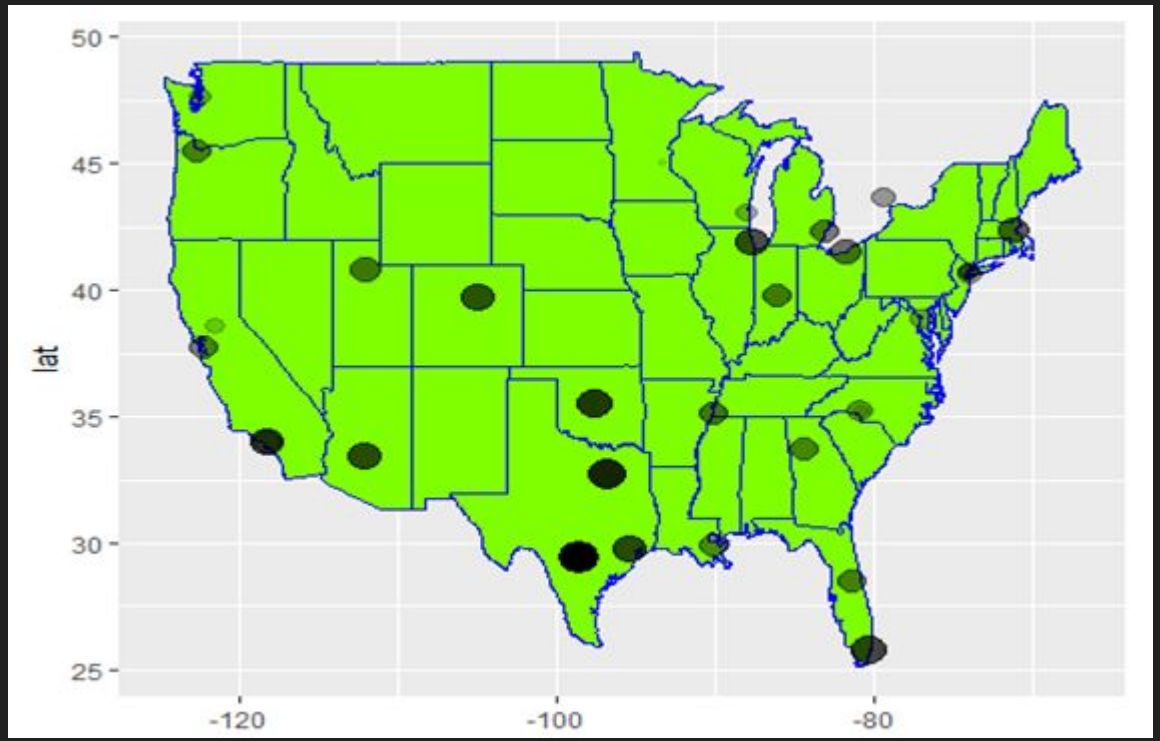
```
Call:
lm(formula = avg_salary ~ win_percent + home_avg_attendance +
    season_start, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-1738038  -551823   -52667   468597  2881486

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.214e+05  3.572e+05   2.579   0.0103 *
win_percent         1.481e+06  3.084e+05   4.802 2.27e-06 ***
home_avg_attendance 9.687e+01  2.187e+01   4.428 1.25e-05 ***
season_start2001    1.938e+05  2.341e+05   0.828   0.4083
season_start2002    3.794e+05  2.403e+05   1.579   0.1152
season_start2003    3.412e+05  2.429e+05   1.404   0.1610
season_start2004    5.653e+05  2.435e+05   2.322   0.0208 *
season_start2005    6.134e+05  2.413e+05   2.542   0.0114 *
season_start2006    4.003e+05  2.565e+05   1.560   0.1195
season_start2007    9.861e+05  2.476e+05   3.983 8.17e-05 ***
season_start2008    1.346e+06  2.387e+05   5.637 3.39e-08 ***
season_start2009    1.245e+06  2.341e+05   5.319 1.79e-07 ***
season_start2010    1.106e+06  2.410e+05   4.592 6.00e-06 ***
season_start2011    9.854e+05  2.345e+05   4.203 3.30e-05 ***
season_start2012    1.029e+06  2.413e+05   4.262 2.56e-05 ***
season_start2013    1.600e+06  2.325e+05   6.879 2.51e-11 ***
season_start2014    5.679e+05  2.348e+05   2.418   0.0161 *
season_start2015    9.595e+05  2.331e+05   4.116 4.74e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 822500 on 377 degrees of freedom
Multiple R-squared:  0.3579,    Adjusted R-squared:  0.329
F-statistic: 12.36 on 17 and 377 DF,  p-value: < 2.2e-16
```






Salary vs Winning Percentage with Fitted Model
avg_salary
6e+06
4e+06
2e+06
0.25
0.50
0.75
win_percent

Does location affect winning percentage?(2005-2015)



Top Three Teams:

San Antonio Spurs (70.75%),
Dallas Mavericks (64.61%),
Miami Heat (59.00%)

Bottom Three Teams:

Minnesota Timberwolves (33.95%)
Charlotte Bobcats/Hornets (35.94%)
Sacramento Kings (38.05%)