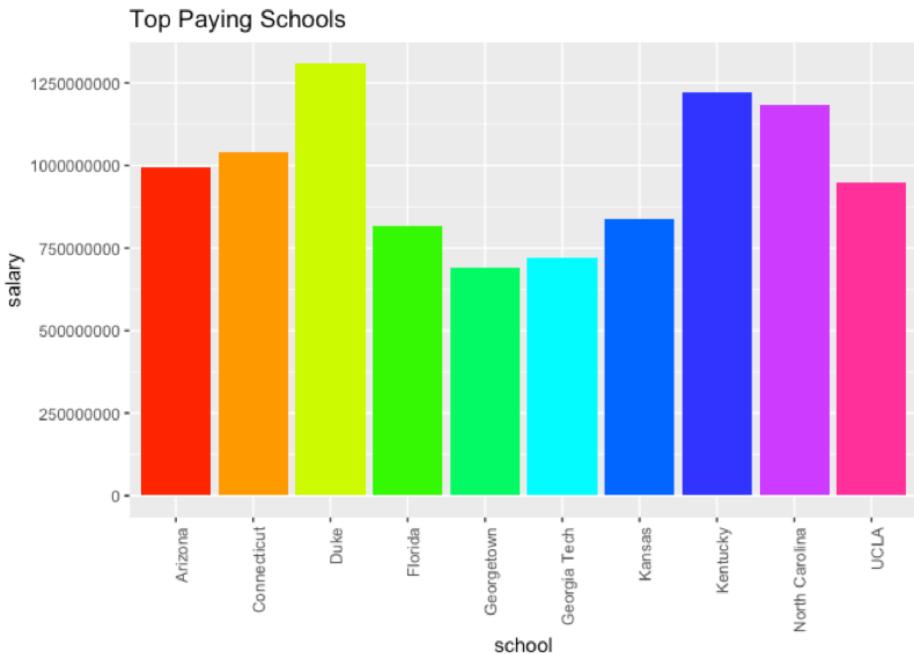


NBA SCOUTING

- Top college programs
- Where is the scouting money going internationally?
- Number 1 overall pick vs a second round gem

Colleges in the United States

```
## # A tibble: 11 x 3
##   school      count_school    salary
##   <fct>          <int>     <dbl>
## 1 None            1285 7407695183
## 2 Kentucky        261 1220596079
## 3 Duke             252 1308479631
## 4 North Carolina  241 1182185814
## 5 UCLA             206 947159451
## 6 Kansas            194 840093830
## 7 Arizona           190 995023686
## 8 Connecticut       184 1038212650
## 9 Florida            143 817700613
## 10 Georgia Tech      129 718884989
## 11 Texas             116 627431018
```



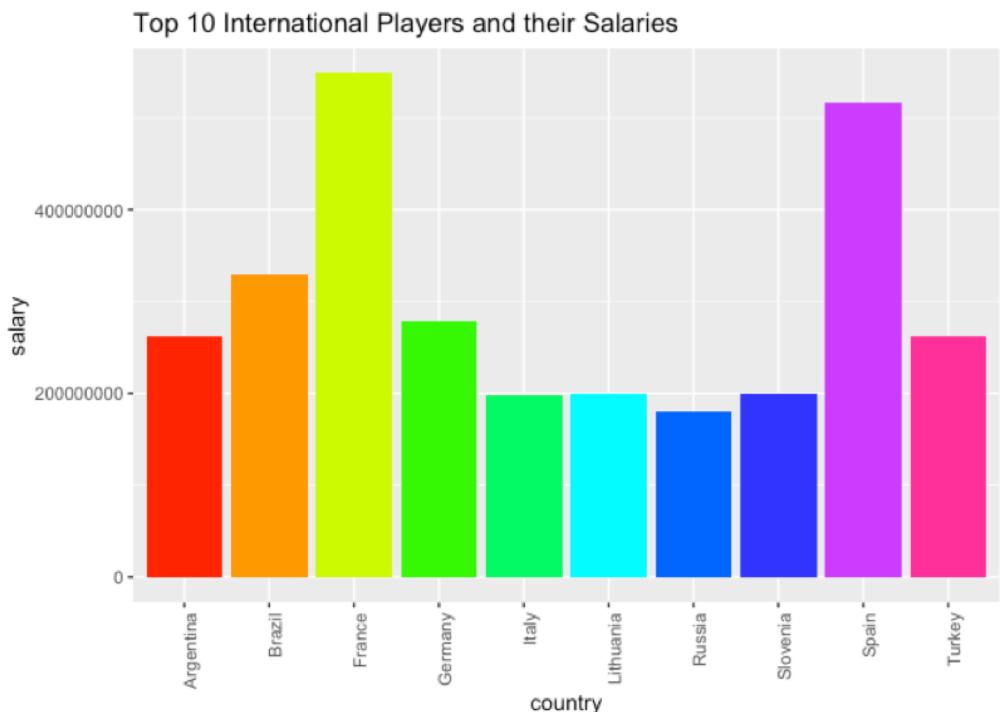
- None row is high school players, internationals and players that did not go to college.

- Top colleges are:
 - Kentucky
 - Duke
 - North Carolina
- Best college programs in the United States producing NBA players.
- Most of these schools scout Big men who play 1 year in college before transitioning to the NBA



Overseas

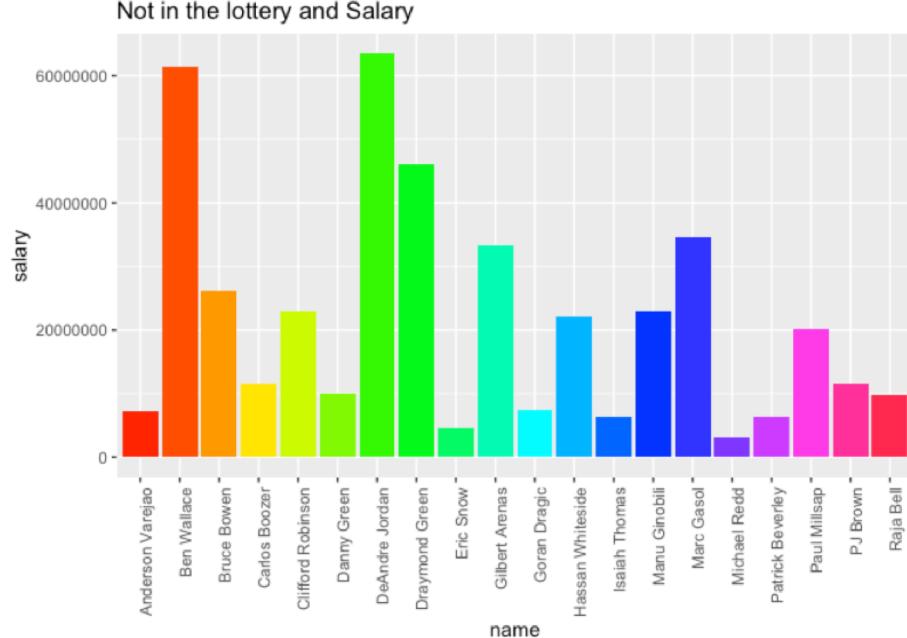
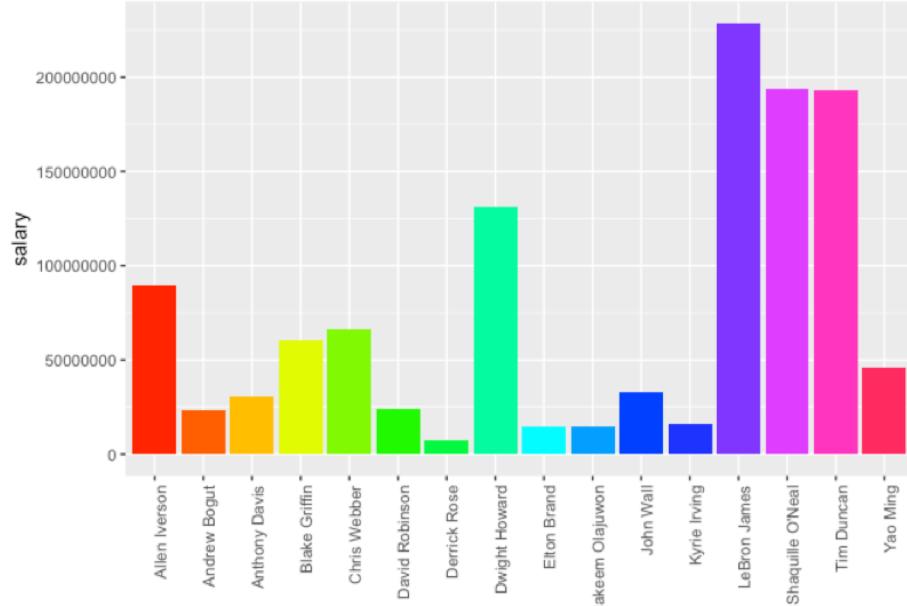
```
## # A tibble: 10 x 3
##   country  count_country    salary
##   <fct>          <int>     <dbl>
## 1 France            102 548647323
## 2 Spain              68 515601566
## 3 Brazil             62 329170240
## 4 Germany            29 278303863
## 5 Argentina          55 262352074
## 6 Turkey             45 261965589
## 7 Slovenia            52 199606417
## 8 Lithuania           27 199502355
## 9 Italy               32 197172659
## 10 Russia            32 180585067
```



- Should look at the European leagues.
- Most NBA players can be drafted but are legally obligated to stay in Europe till their European league contract expires.
- Young European players might be a trend since they are more mature players.
 - Play amongst men at age 15
 - No American Money culture
 - Learn team basketball instead of star play basketball



Number 1 overall vs Second round Gem



- Strategy of tanking to get a good NBA lottery pick.
 - Good examples
 - 76ers
 - Boston Celtics
 - Bad Examples
 - Knicks
 - Phoenix
- Playing to compete and drafting smart
 - San Antonio Spurs
- Number 1 overall picks will get paid more just entering the league. Most awards are media determined. More coverage on these players will lead to bigger contracts.
- A second rounder won't get as much media coverage and thus awards might be harder to come by and thus a lower salary overall.



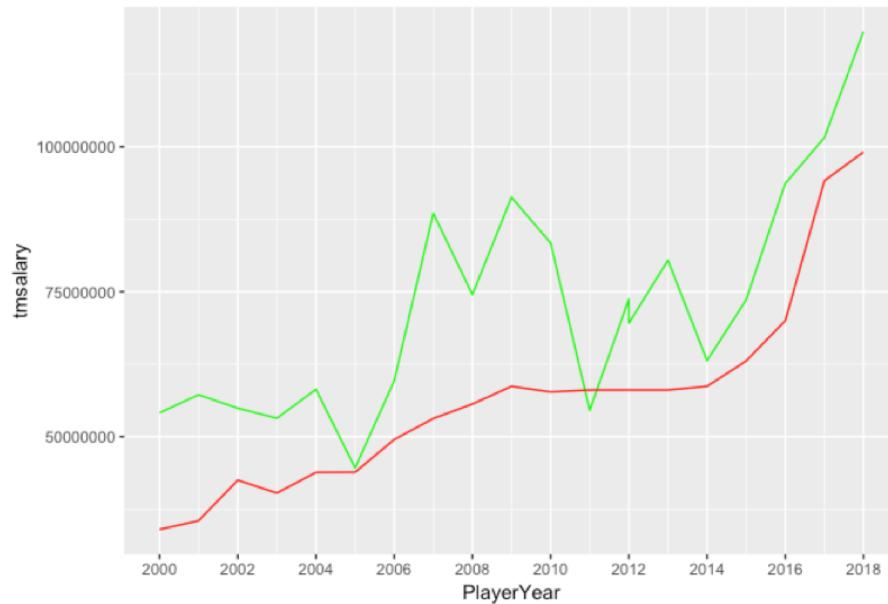
NBA SALARY REFERENCE

- Winning teams Vs Losing Teams
- The Best team in NBA history vs the Worst team in NBA history

Best vs Worst teams

Avoid making mistakes like the worst.

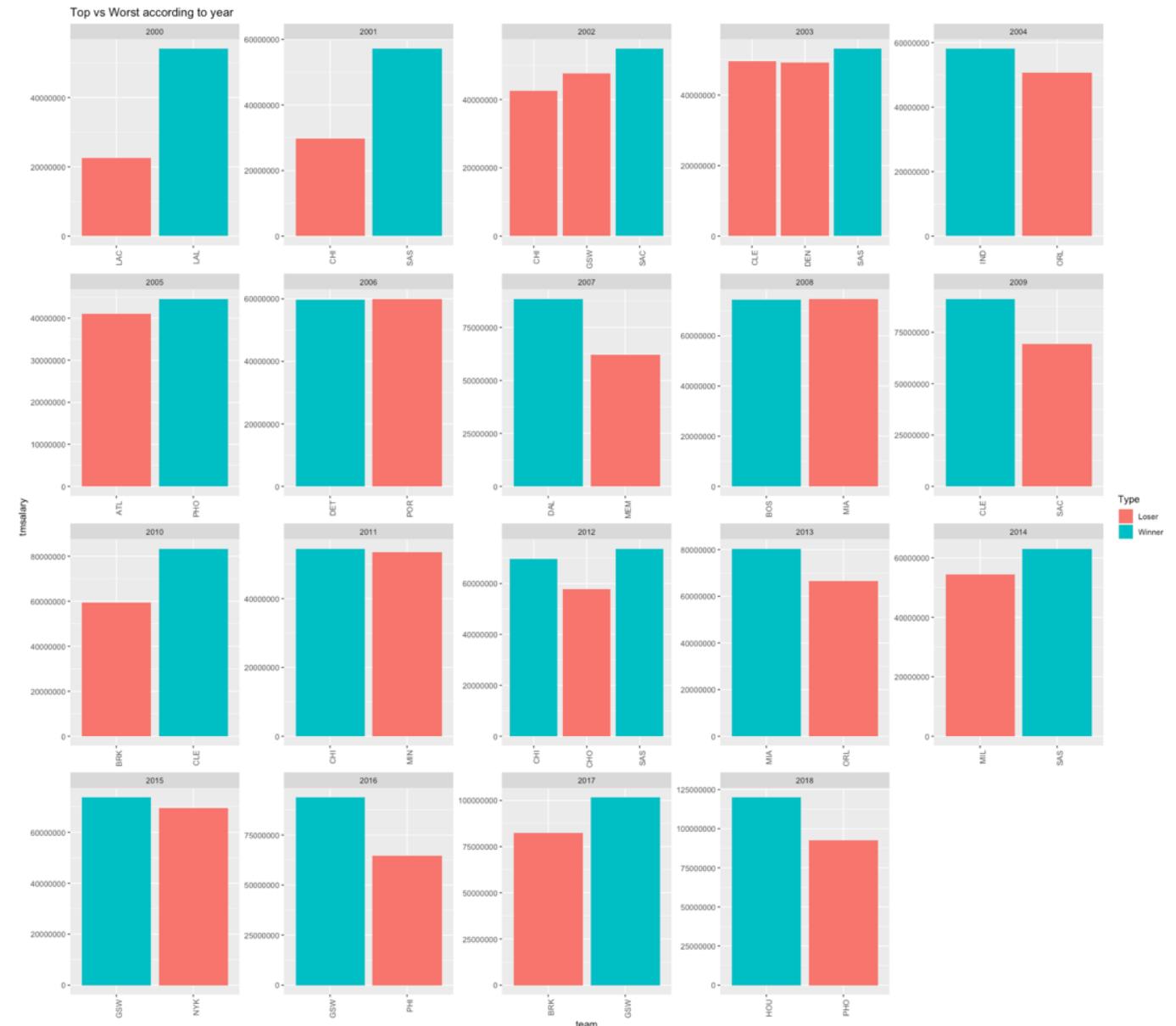
Winning Teams and their Salary Cap



Losing Teams and their salary cap



The best teams vs the worst teams

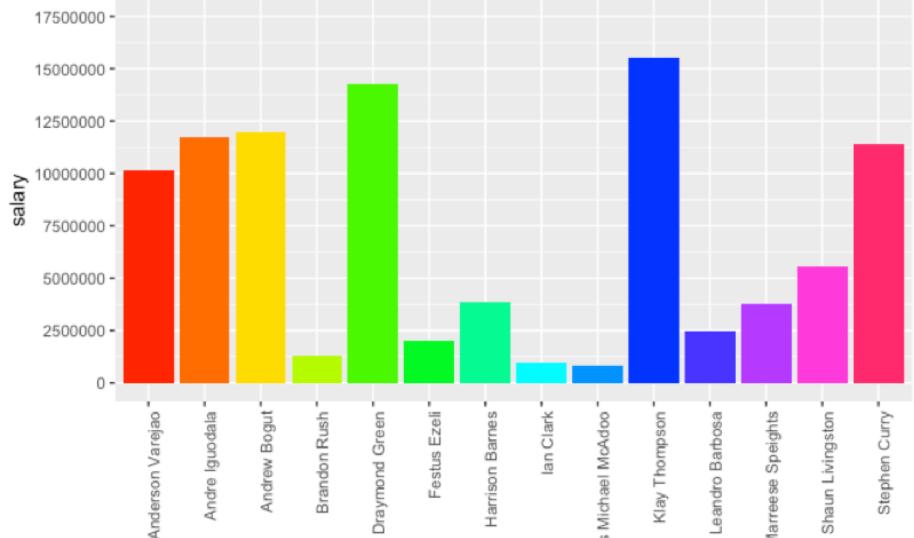


- Best teams over the years have spent more, salary cap has a pro rata basis hence losing teams still spend near the cap.

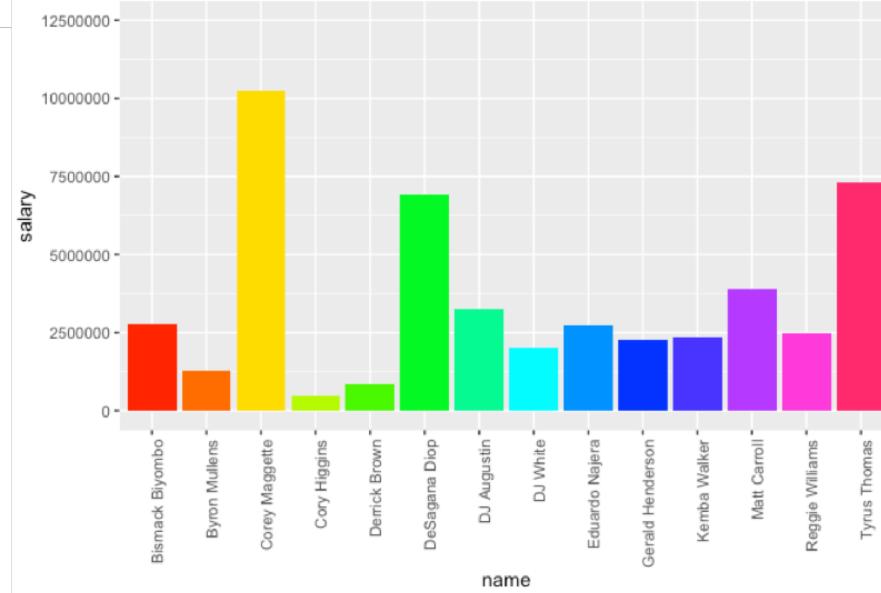


Best vs Worst in NBA History

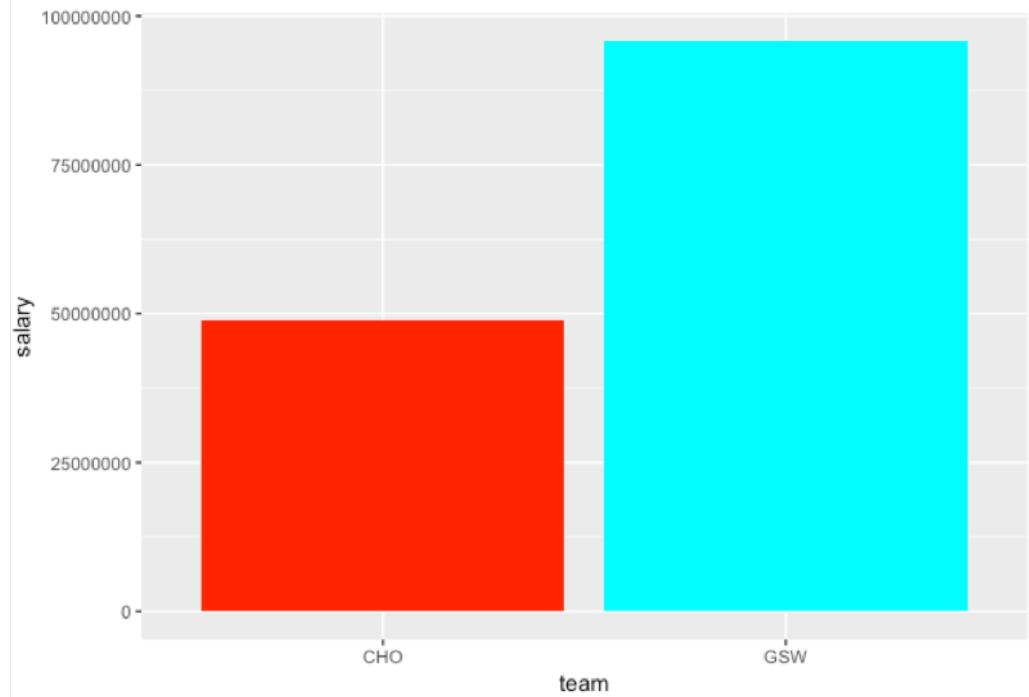
GSW 2016 Player Salary Spread



CHO 2012 salary spread



worst vs the best



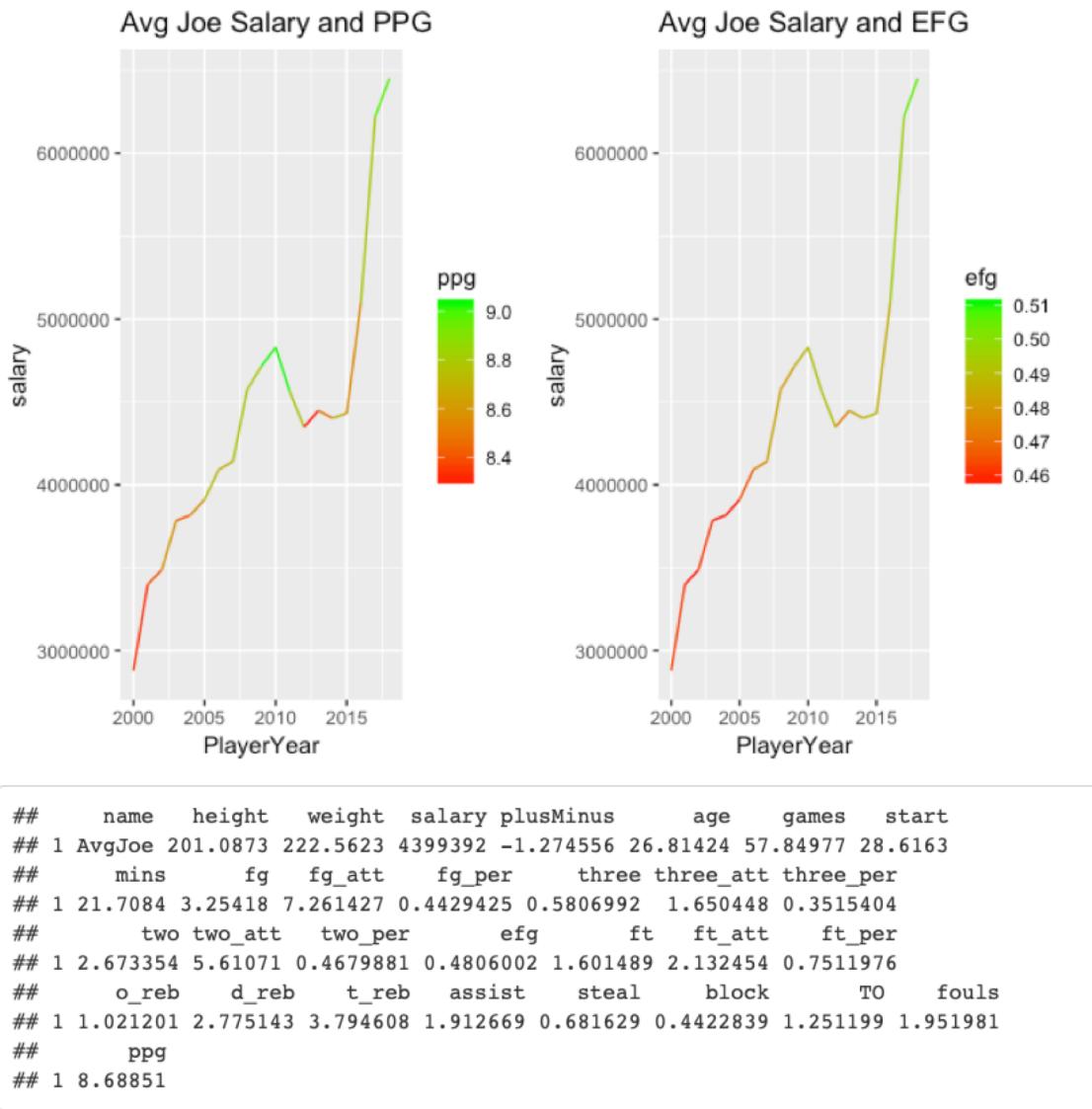
- Both were somewhat anomalies for the perfect storm
 - Bobcats
 - Shortened NBA season, with no practice
 - Rebuild mode signed players just to meet salary cap
 - Warriors
 - Good drafting
 - MVP player was on a cheap discount due to previous injuries



AVERAGE JOE VS

- How does the average joe compare to:
 - Superstar
 - Star
 - RolePlayer
 - MVP
 - DPOY
 - ROY
 - Mid-level exception player

What to expect from an Average Joe

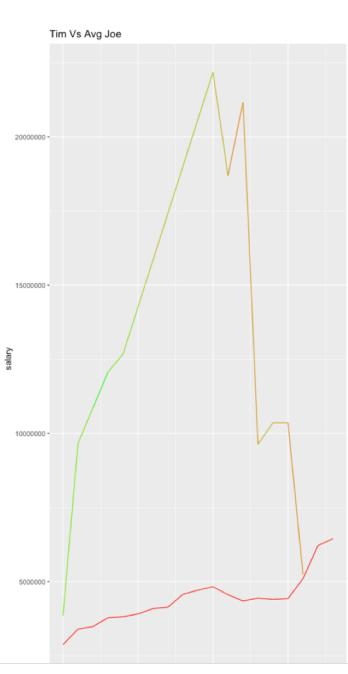
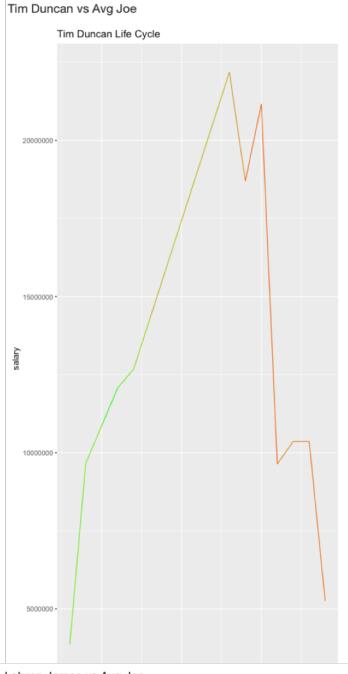


- The most average player for a team these are the players you want to fill in as role players.
- Can expect to pay a salary of 4.4 mil approximately
- Averaging about 8.6 points a game
- We can see the trend in Effective Field Goal percentage is going up with the number of years.
- Huge salary cap increase after 2015, need to collect more data until the next CBA opt out in 2023

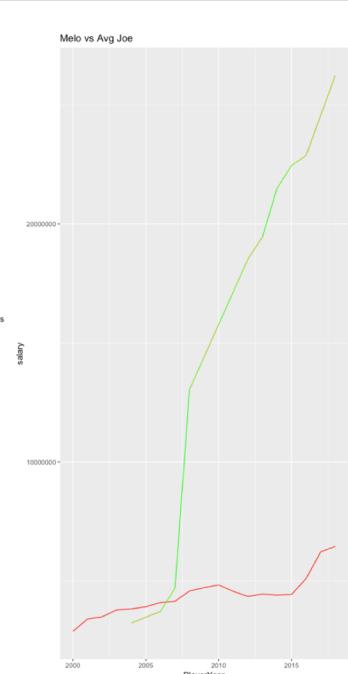
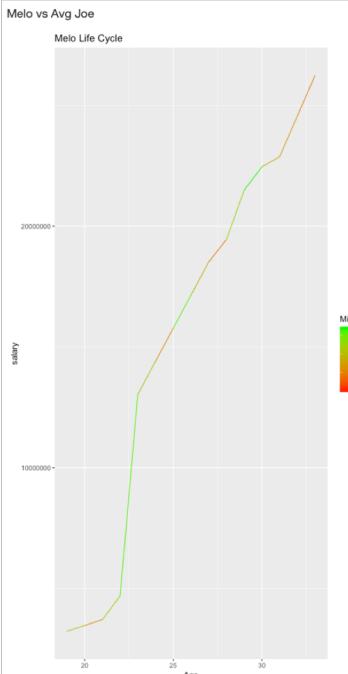


Superstars, Stars and Role Players vs the Average Joe

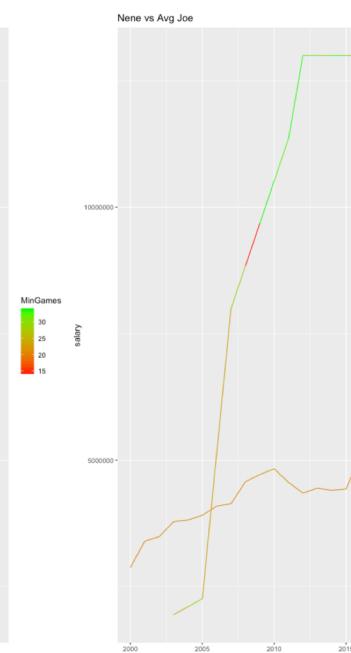
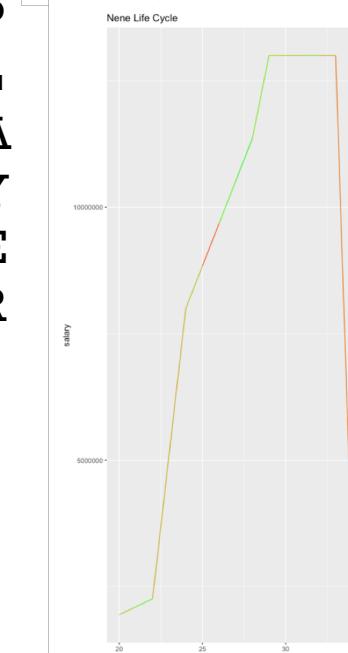
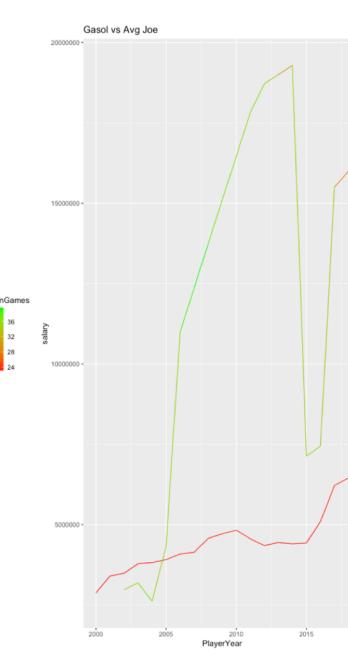
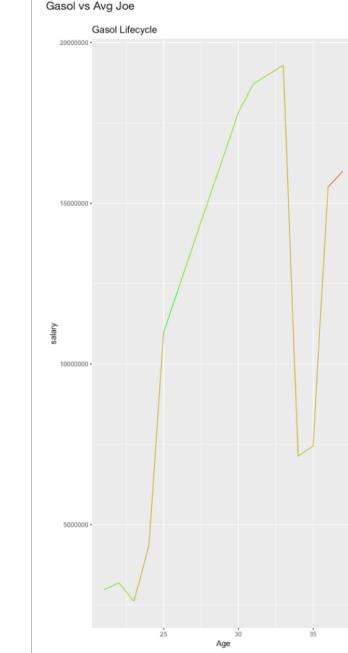
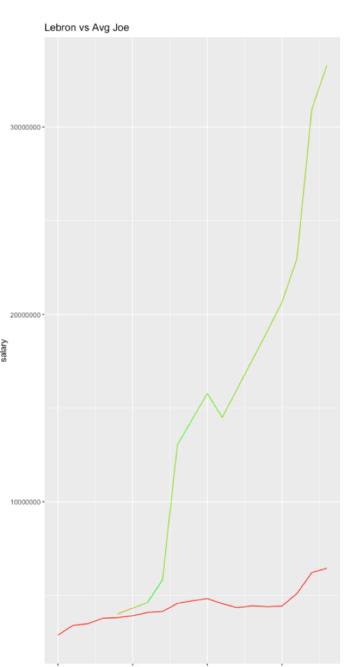
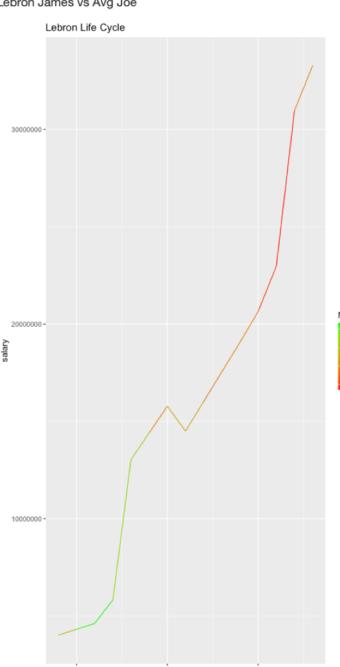
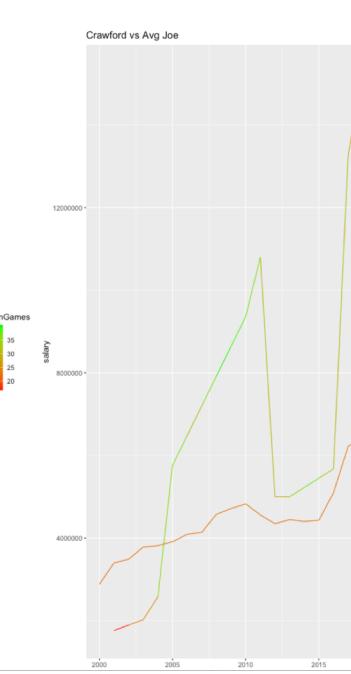
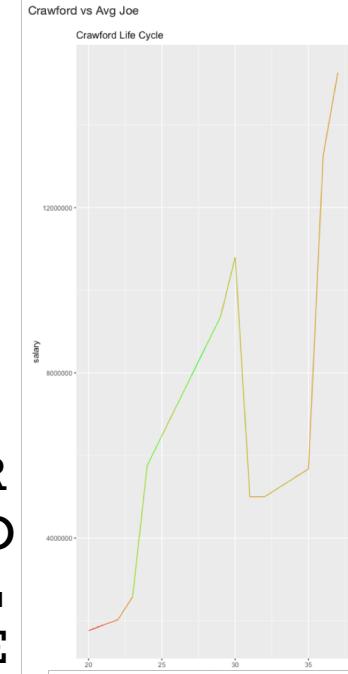
S U P E R S T A R



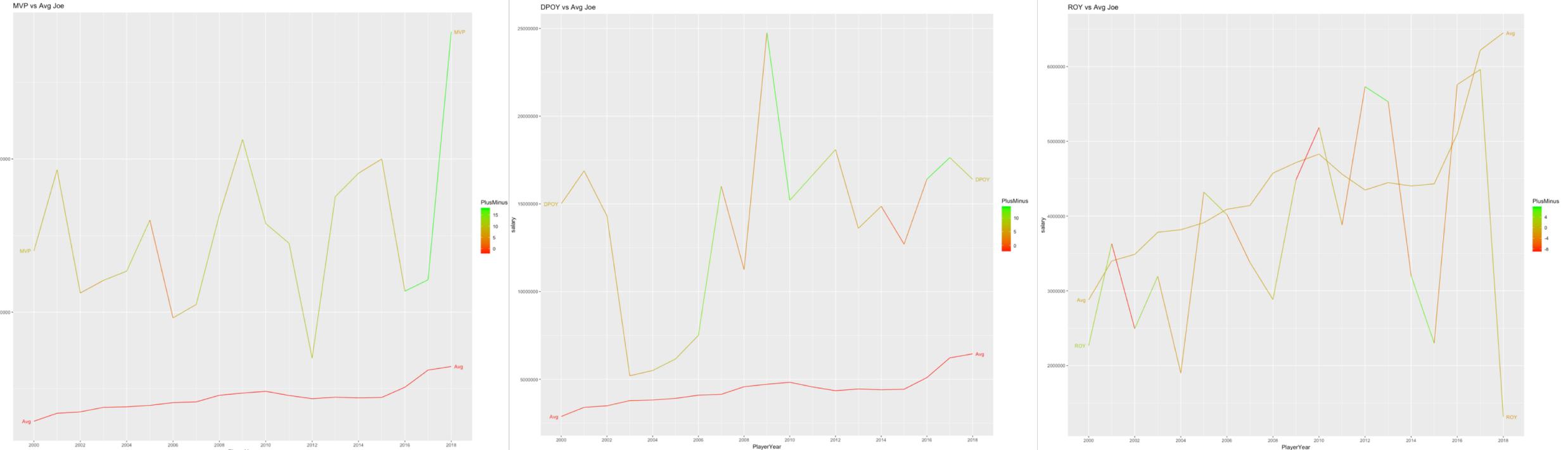
S T A R



R O L E P L A Y E R



Award Winners vs Average Joe



- MVP and DPOY winners have a higher salary than the average joe.
 - This is because of the structure of the CBA unlocking more salary that can be paid to the player
- Rookie of the year data is above or below. First pick is usually paid the most, winners drafted later have a lower initial salary.
- 2018 winner Ben Simmons was actually drafted in 2017 so his contract was before the new CBA with the average increasing overall thus a huge difference.

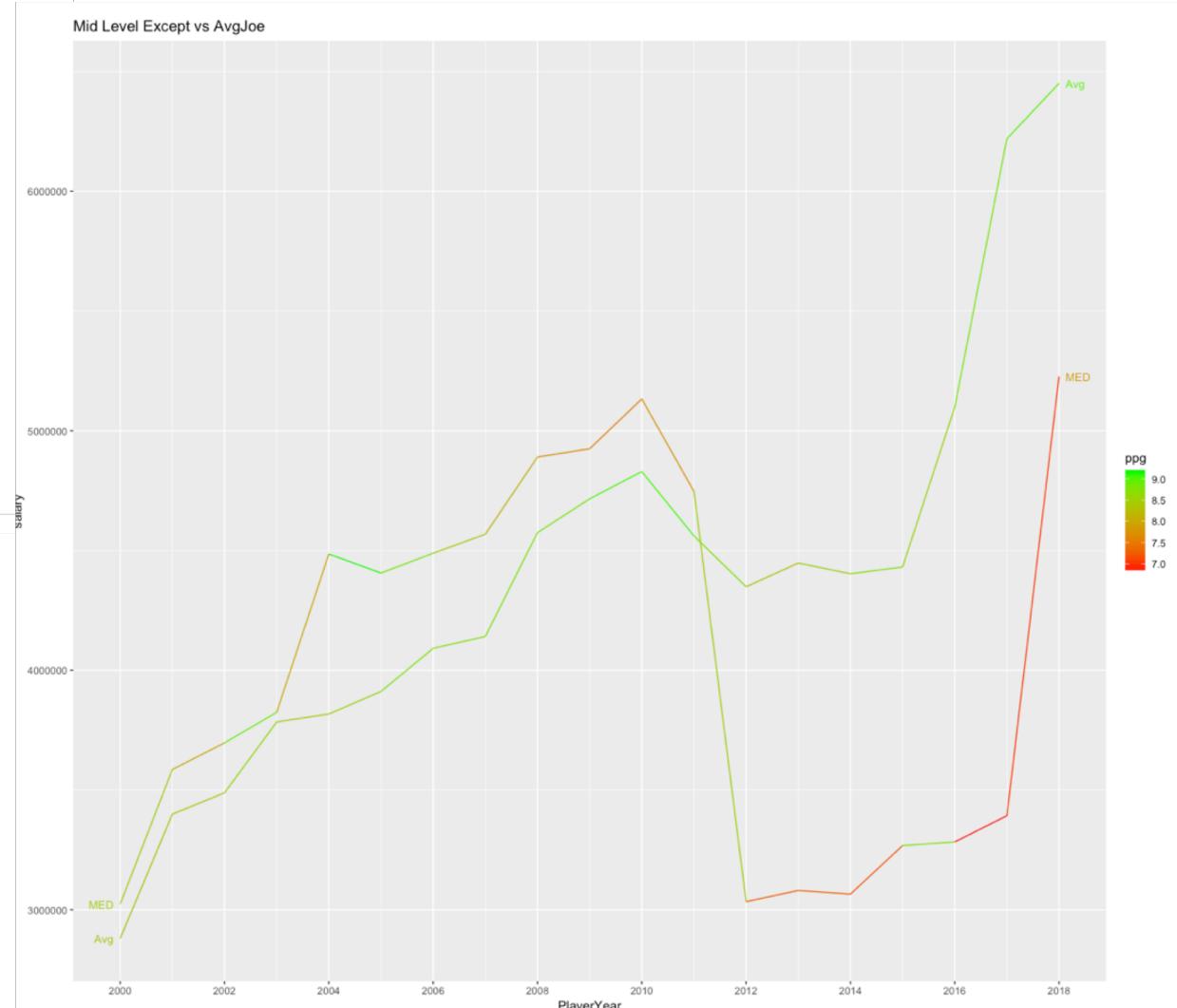


Out of money thank god for exceptions – Mid-level exception

What to Expect with a mid-level exception player.

```
##   PlayerYear salary height_cm PlusMinus MinGames     ppg      EFG
## 1    2018 5226548 199.8919 -1.45945946 19.97838 7.862162 0.5142168
## 2    2017 3392485 199.9000 -1.43000000 18.22250 7.077500 0.5026139
## 3    2016 3282897 201.0238 -1.26666667 18.58095 6.902381 0.4972249
## 4    2015 3267768 199.6735 -0.24693878 21.51633 8.748980 0.4935115
## 5    2014 3064909 202.5106 -0.62978723 19.58936 7.225532 0.5069429
## 6    2013 3080277 200.3958 -0.99166667 19.80833 7.329167 0.4928712
## 7    2012 3033485 202.2500 -1.41923077 21.43654 7.290385 0.4944566
## 8    2011 4746809 200.3542 -1.43958333 22.56667 8.477083 0.4938972
## 9    2010 5132612 200.6364 -0.35454545 22.12182 7.790909 0.4983103
## 10   2009 4924562 201.4000 -0.03200000 22.65800 7.588000 0.4923068
## 11   2008 4890182 200.2000 -0.30000000 22.48364 7.709091 0.4937362
## 12   2007 4567807 200.0189 -1.11320755 22.99623 8.175472 0.4912060
## 13   2006 4488497 199.0476 -0.06031746 24.19841 8.488889 0.4902785
## 14   2005 4405707 203.0217 -1.96086957 23.15000 8.804348 0.4673055
## 15   2004 4485209 202.4651 0.67441860 24.23023 9.153488 0.4651711
## 16   2003 3824031 200.2083 1.52500000 22.42708 8.054167 0.4753164
## 17   2002 3696504 200.1746 -0.57777778 23.50159 8.868254 0.4737335
## 18   2001 3585560 199.4237 -1.44915254 23.21017 8.038983 0.4646160
## 19   2000 3023264 199.9796 -0.65306122 22.12245 8.469388 0.4679259
```

- Set criteria for mid level exception as 25 percent over and under that years salary cap average player pay.
- Before 2011 it was the average salary of teams over the cap.
- The mid level players were better signings before the 2011 CBA
- 2011 CBA is when David Stern implemented rules to stop super teams
- Mid level exception can also be used on multiple players
- Other things to see is the biannual exception



NBA REGRESSION ANALYSIS

- Multiple Linear Regression 2000-2018
- Multiple Linear Regression considering the Collective Bargaining Agreements

2000-2018 Multiple Linear Regression

```
##  
## Call:  
## lm(formula = salary ~ t_reb + TO + ppg + FG + Age + started +  
##     three + two + assist + block, data = data, subset = +steal +  
##     NumYears)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -1890342 -719888  143796  410398 1786027  
##  
## Coefficients:  
##             Estimate Std. Error t value    Pr(>|t|)  
## (Intercept) -4924622.4  133245.1 -36.959 < 2e-16 ***  
## t_reb        69609.1   12584.1   5.532 0.00000003275268 ***  
## TO          713760.0   102179.4   6.985 0.00000000000307 ***  
## ppg         1608869.0   22628.6  71.099 < 2e-16 ***  
## FG          -2058458.2  305883.4  -6.730 0.00000000001820 ***  
## Age          185457.7   6119.3   30.307 < 2e-16 ***  
## started     -11621.6    594.4  -19.550 < 2e-16 ***  
## three        -3808624.1  270022.4 -14.105 < 2e-16 ***  
## two          -1421161.9  278863.9  -5.096 0.00000035439981 ***  
## assist        804.6    26500.8   0.030     0.976  
## block        1714923.3  36829.5   46.564 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 793700 on 7967 degrees of freedom  
## Multiple R-squared:  0.9239, Adjusted R-squared:  0.9238  
## F-statistic: 9678 on 10 and 7967 DF,  p-value: < 0.000000000000022
```

- **Multiple Linear Regression with the variables that are most correlated with Salary**
 - Total Rebounds
 - Turnover
 - Points per game
 - Field Goal
 - Age
 - Started
 - Three made
 - Two made
 - Assists
 - Blocks
 - Steals
 - Number of Years in the league
- Removed EFG from the first tested model.
- High Adj R squared, high F-stat and low p-value showing that it is significant

2000-2018 split over 3 CBAs (1999 CBA, 2005 CBA, 2011 CBA)

1999 CBA

```
## 
## Call:
## lm(formula = salary ~ Age + t_reb + TO + ppg + FG + started +
##     three + two + assist + block, data = cbal, subset = +steal +
##     NumYears)
## 
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -1852106 -615220   85809  333425 1768193 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 4722618   246846 -19.132 < 2e-16 ***
## Age          176655    11456  15.420 < 2e-16 ***
## t_reb        63148    23090   2.735  0.006286 **  
## TO           814801   188083   4.332  0.0000154 ***
## ppg          1665340   44007  37.843 < 2e-16 ***
## FG           -1974900  559603  -3.529  0.000425 *** 
## started      -11539    1023  -11.277 < 2e-16 ***
## three        -4136394  490758  -8.429 < 2e-16 ***
## two          -1645545  505201  -3.257  0.001141 **  
## assist       -18589    48975  -0.380  0.704310  
## block        1650149   67529  24.436 < 2e-16 *** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 801600 on 2389 degrees of freedom
## Multiple R-squared:  0.9239, Adjusted R-squared:  0.9235 
## F-statistic: 2898 on 10 and 2389 DF,  p-value: < 0.0000000000000022
```

2005 CBA

```
## 
## Call:
## lm(formula = salary ~ Age + t_reb + TO + ppg + FG + started +
##     three + two + assist + block, data = cba2, subset = +steal +
##     NumYears)
## 
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -518211 -92081  -25344   93906 1346827 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1449376.3  52711.8 -27.496 < 2e-16 ***
## Age          73775.1   1833.4  40.239 < 2e-16 ***
## t_reb        176073.3  12343.7 14.264 < 2e-16 ***
## TO           -3466135.5  58297.0 -59.456 < 2e-16 ***
## ppg          -1705038.4  37376.3 -45.618 < 2e-16 ***
## FG           -94003.7   221599.3 -0.424   0.671  
## started      36550.8    947.3  38.582 < 2e-16 ***
## three        6614905.4  189533.8 34.901 < 2e-16 ***
## two          5778475.5  188073.0 30.725 < 2e-16 ***
## assist       540546.5   17210.5 31.408 < 2e-16 ***
## block        784881.1   44083.0 17.805 < 2e-16 *** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 249800 on 2434 degrees of freedom
## Multiple R-squared:  0.9891, Adjusted R-squared:  0.9891 
## F-statistic: 2.212e+04 on 10 and 2434 DF,  p-value: < 0.0000000000000022
```

2011 CBA

```
## 
## Call:
## lm(formula = salary ~ Age + t_reb + TO + ppg + FG + started +
##     three + two + assist + block, data = cba3, subset = +steal +
##     NumYears)
## 
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -4406673 -998940  -382705   2056766 4672091 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 18569818   466372 -39.818 < 2e-16 ***
## Age          344673    14620  23.575 < 2e-16 ***
## t_reb        1172011   42364  27.665 < 2e-16 ***
## TO           5956972   242854  24.529 < 2e-16 ***
## ppg          -4766604   114015 -41.807 < 2e-16 ***
## FG           2041864   1495278   1.366  0.172181 
## started      4608      3814   1.208  0.227112 
## three        19046415   1676491  11.361 < 2e-16 ***
## two          9376796   1622094   5.781  0.00000000817 ***
## assist       -414443    111963  -3.702  0.000218 *** 
## block        957182    168833   5.669  0.00000001564 *** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2177000 on 3122 degrees of freedom
## Multiple R-squared:  0.8894, Adjusted R-squared:  0.8891 
## F-statistic: 2511 on 10 and 3122 DF,  p-value: < 0.0000000000000022
```

- Regime Bias possibility
 - Could the different rules and especially the different ways the contracts are structured over the years.
 - We can see the model is best during the 2005 CBA at 0.9891
 - Worst with the new CBA 0.8891
 - Interesting to see 2017 CBA with more data in the future



FINAL THOUGHTS

- Can we make a team with Moneyball?
 - We can use the regression model we made as a reference point as well as information from our data exploration.
 - What is far more important is the state of the CBA every 6-10 years.
 - Basketball Related Income is increasing with more popularity of basketball and globalization
 - More power to the players association
- The limits of our data
 - Team specific compensation structure
 - Playing a number of games
 - Holding an average points, rebounds or assists
 - Player pensions
 - Social influences
 - Brand building outside of the team
 - Not a video game
 - Team and player chemistry
 - Human error in trades
 - Injuries and stuck with a bad contract
- What to explore in the Future
 - Player options and Team options
 - Home town discounts
 - Social media influence and media coverage paying an average player the max contract – Andrew Wiggins
 - How modern day science is extending an NBA players life cycle.
- Conclusion
 - Basketball unlike baseball can be dominated by one good player and with the right conditioning can play the whole 48 minutes.
 - Salary cap can be flexible allowing teams to sign more for what they should be able to.
 - The CBA agreement is a huge determining factor and understanding it first and Player contract rules and regulations should be emphasized first while player data analyzation should only assist in salary talks