



Steroid Scandal Across Baseball Leagues

Team Big Cheese: Alice Jiang, Cassio Salge,
Justin Wood, Nicole Santolalla, Pelumi Ajayi

AGENDA

01 Introduction

Outlining the question and our goal

02 Data Exploration

Describe and explore data

03 Our Insights

Present our findings and insights

04 Next Steps

What we want to continue exploring





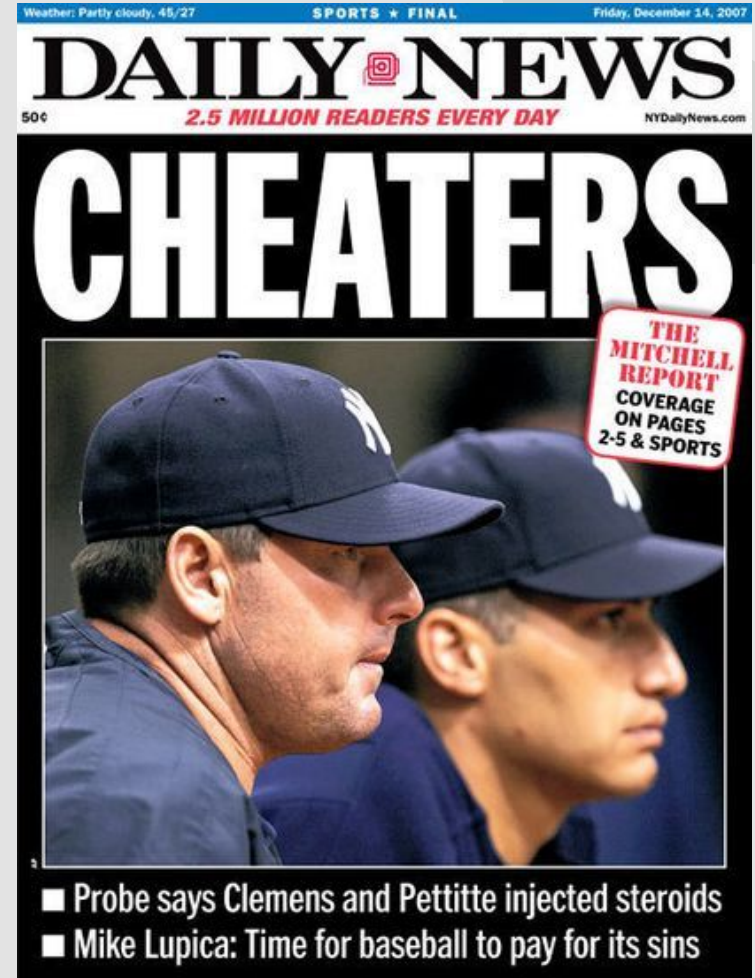
01 Introduction

Background

In the early 2000's, steroid use was becoming a major problem in the MLB and a network of highly paid and beloved players from multiple teams were accused or caught in enhancement drugs to better their performance.

However, the question remains who was tied to who and which dealers were really the big players in this distribution of illegal substances?

As Jeff Kent from the LA Dodgers says, **“Major League Baseball is trying to investigate the past so they can fix the future.”**



The questions we plan to address...

1.

Which players and teams are connected to providers?

2.

Which teams are connected to each other?

3.

How has this influenced the players and the organization performance?

Our Objective

Use social network analytics and data exploratory analysis to find the influence steroids made on baseball players and organizations.

Analytical Tools: R Studio, Tableau, Secondary Research, and domain knowledge.





02

Data Exploration

Datasets and Data-Wrangling

Canvas:

1. Players to Providers network
Transformed this dataset to edge list
2. Players to Players network
Transformed this dataset to edge list

Kaggle:

3. Players ID and their information
Connect players in our network data with their ID found on Kaggle
4. Team and games information
Connect players with teams and their teams' game info
5. Players' salary information
For analyzing the players on steroids salaries

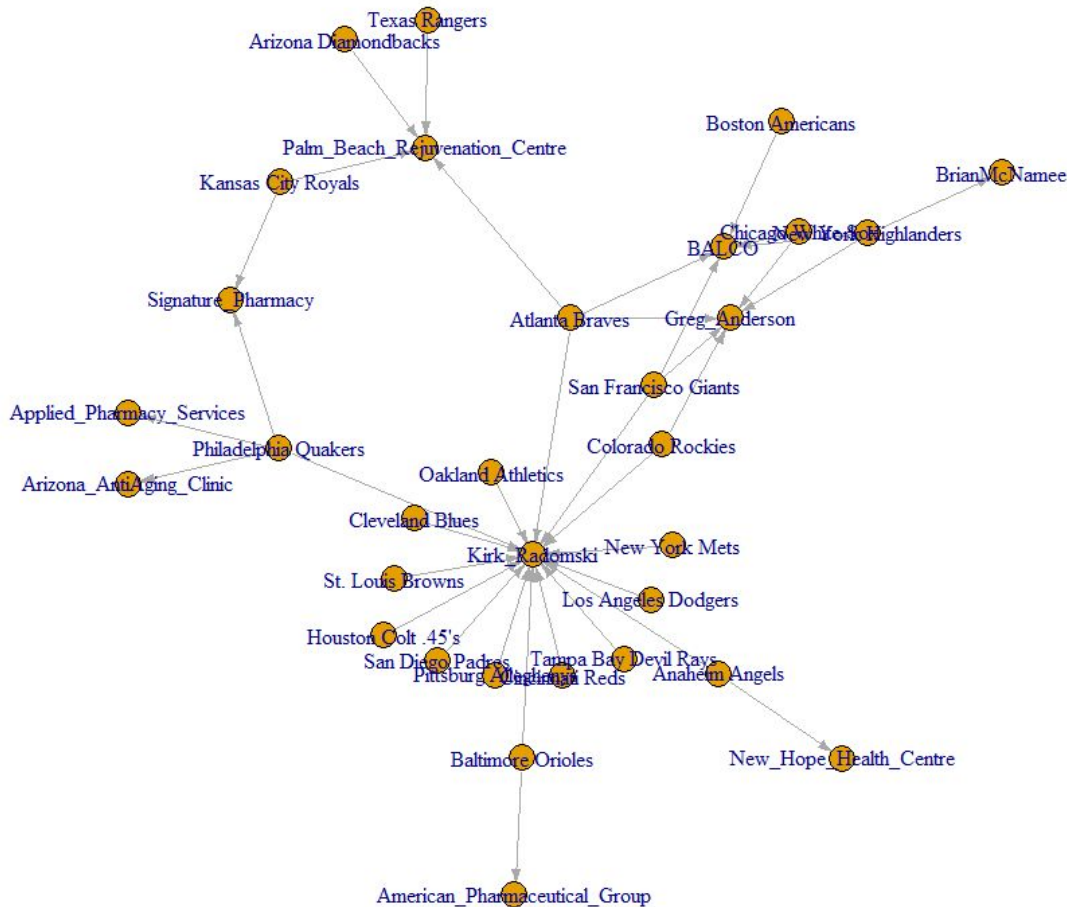
The diagram illustrates a network of relationships between individuals and organizations. The nodes are represented by yellow circles, and the edges are grey lines. The network is divided into several clusters:

- Top-Left Cluster:** Includes individuals like Steve Woodward, Gary Matthews, David Bell, John Rucker, and organizations like Applied Pharmacy Services, Signature Pharmacy, Jason Grimsley, Jose Causeco, Health Watch Clinic, Health Rejuvenation Centre, Rick Ankiel, Paul Byrd, Ismael Valdez, Jose Guillen, Darrell Holmes, and Matt Williams.
- Top-Right Cluster:** A large, dense network centered around Kirk Radomski, with many other individuals connected to him, including Dennis Neugebauer, David Segui, Mike Patton, Eric Gamme, Luis Perez, Kent Morcker, Jose Cervantes, Adam Biggs, Nook Logan, Mike Lansing, Josias Manzanillo, Jeff Williams, Bart Madich, Paul Lo Duca, Larry Diebie, Brendan Donnelly, Kevin Brown, Chris Dennels, Todd Pratt, Greg Laun, Matt Harghal, Hal Morris, Jerry Hairston, Mark Carreon, Kevin Young, Phil Hatt, Jim Pasque, Glennen Ball, Roberts, Phil Hatt, Samangelo, David Ju, Fernando Vina, Cody McKay, and Mike Stanton.
- Bottom-Left Cluster:** Includes individuals like Scott Scheneweis, Roger Clemens, Chuck Knoblauch, Brian McNamee, Andy Pettitte, Troy Glaus, and organizations like New Hope Health Centre.
- Bottom-Right Cluster:** Includes individuals like Jereny Giambi, Barry Bonds, Jason Giambi, Marvin Benard, Benito Santiago, Greg Anderson, Armando Rios, Gary Sheffield, Bobby Estalella, Randy Velarde, and organizations like BALBO and Greg Anderson.

There are very distinct components within the network.

Kirk Radomski is the most connected provider with Greg Anderson coming in second.

Team Provider Relationship

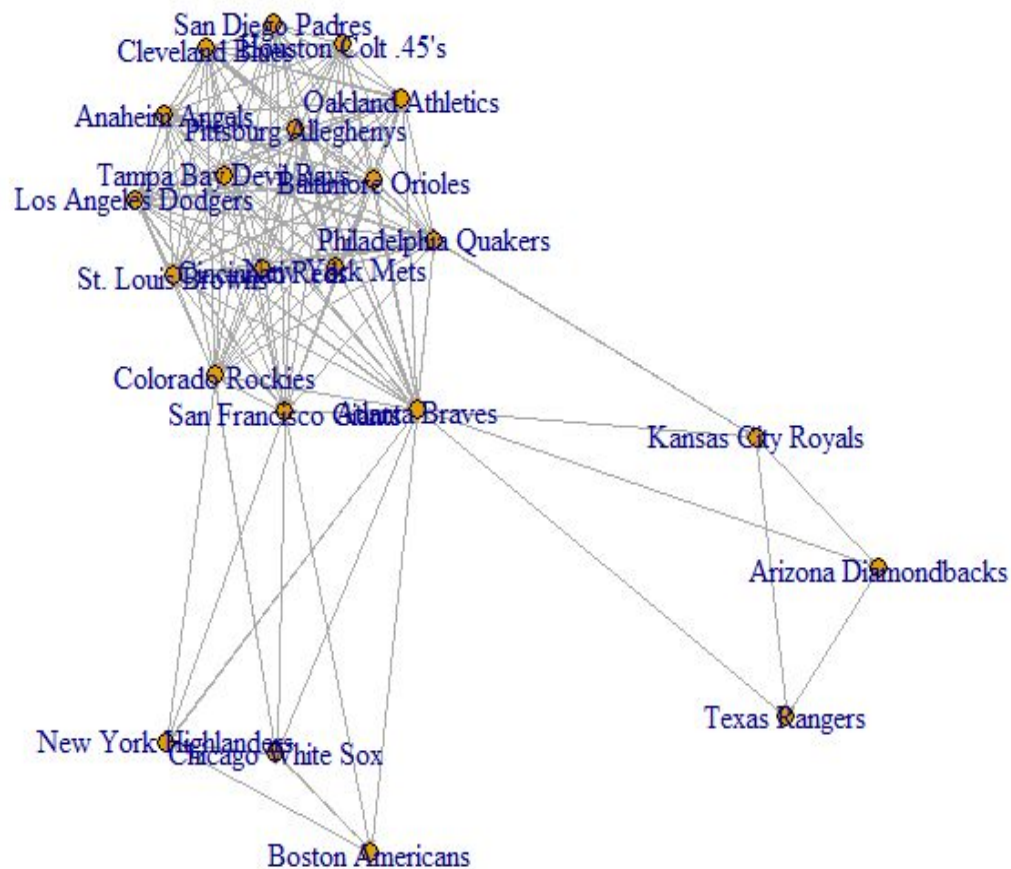


Team-Provider Network Analysis

Kirk Radomski is not only connect with players, but with majority of the MLB teams

The network shows that location is not a factor for suppliers

Teams Relationship



Team-Team Network Analysis

The whole league is connected

Ties indicate spread of information across players / teams

Centrality Measures

Player to Provider

	id	deg	in	deg	out	prank
1:	Kirk Radomski	44	0	0.264911057		
2:	Greg_Anderson	8	0	0.042553191		
3:	Palm_Beach_Rejuvenation_Centre	6	0	0.039588420		
4:	BALCO	5	0	0.024764562		
5:	Signature_Pharmacy	5	0	0.022293919		

Kirk Radomski
Degree In : 44
PageRank: 0.265

Team to Provider

	id	deg	in	deg	out	prank
:	Kirk_Radomski	16	0	0.22649573		
:	BALCO	5	0	0.06023997		
:	Greg_Anderson	5	0	0.05185733		
:	Palm_Beach_Rejuvenation_Centre	4	0	0.06582840		
:	Signature_Pharmacy	2	0	0.03229783		

Kirk Radomski
Degree In : 16
PageRank: 0.226

Team to Team

	id	deg	in	deg	out	prank
	Atlanta Braves	21	21	0.08039427		
	San Francisco Giants	18	18	0.06310591		
	Colorado Rockies	17	17	0.05859429		
	Philadelphia Quakers	16	16	0.05551910		
	Baltimore Orioles	15	15	0.05026036		

Atlanta Braves
Degree In : 21
PageRank: 0.0804



03

Our Insights

Initial Results and Insight

- 71 players and 12 providers
- 33 teams
- Average team ranking = 2.89
- 19 players won awards



Insights

Player Salary By Year

First Name Last Name	Year ID										
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Armando_Rios		200,000	225,000	310,000	925,000	450,000					
Barry_Bonds	8,916,667	9,381,057	10,658,826	10,300,000	15,000,000	15,500,000	18,000,000	22,000,000	19,331,470	15,533,970	
David_Bell	200,000	700,000	1,425,000	2,200,000	1,750,000	3,200,000	4,400,000	4,700,000	4,700,000		
Gary_Sheffield	14,936,667	9,956,667	9,916,667	9,916,667	9,916,667	11,416,667	13,000,000	13,000,000	10,756,171	10,916,071	13,326,306
Jason_Giambi	315,000	2,103,333	3,103,333	4,103,333	10,428,571	11,428,571	12,428,571	13,428,571	20,428,571	23,428,571	23,428,571
Jason_Grimsley		350,000	750,000	750,000	1,900,000	2,000,000	1,000,000	2,000,000	825,000		
Jerry_Hairston			205,000	235,000	300,000	1,550,000	1,650,000	1,800,000	2,300,000	750,000	
Jose_Canseco	2,125,000	3,325,000	3,000,000								

Barry Bonds - 1998 Salary is ~\$9M, 2005 Salary is **\$22M**

Jason Giambi - 1998 Salary \$350,000, 2007 Salary is **\$23M**

First Name Last Name	Year ID	Teams	Home runs
Armando_Rios	2003	Chicago White Sox	2
Barry_Bonds	2003	San Francisco Giants	45
	2004	San Francisco Giants	45
	2005	San Francisco Giants	5
	2006	San Francisco Giants	26
	2007	San Francisco Giants	28
David_Bell	2003	Philadelphia Quakers	4
	2004	Philadelphia Quakers	18
	2005	Philadelphia Quakers	10
	2006	Philadelphia Quakers	6
Gary_Sheffield	2003	Atlanta Braves	39
	2004	New York Highlanders	36
	2005	New York Highlanders	34
	2006	New York Highlanders	6
	2007	Detroit Tigers	25
Jason_Giambi	2003	New York Highlanders	41
	2004	New York Highlanders	12
	2005	New York Highlanders	32
	2006	New York Highlanders	37
	2007	New York Highlanders	14

Insights

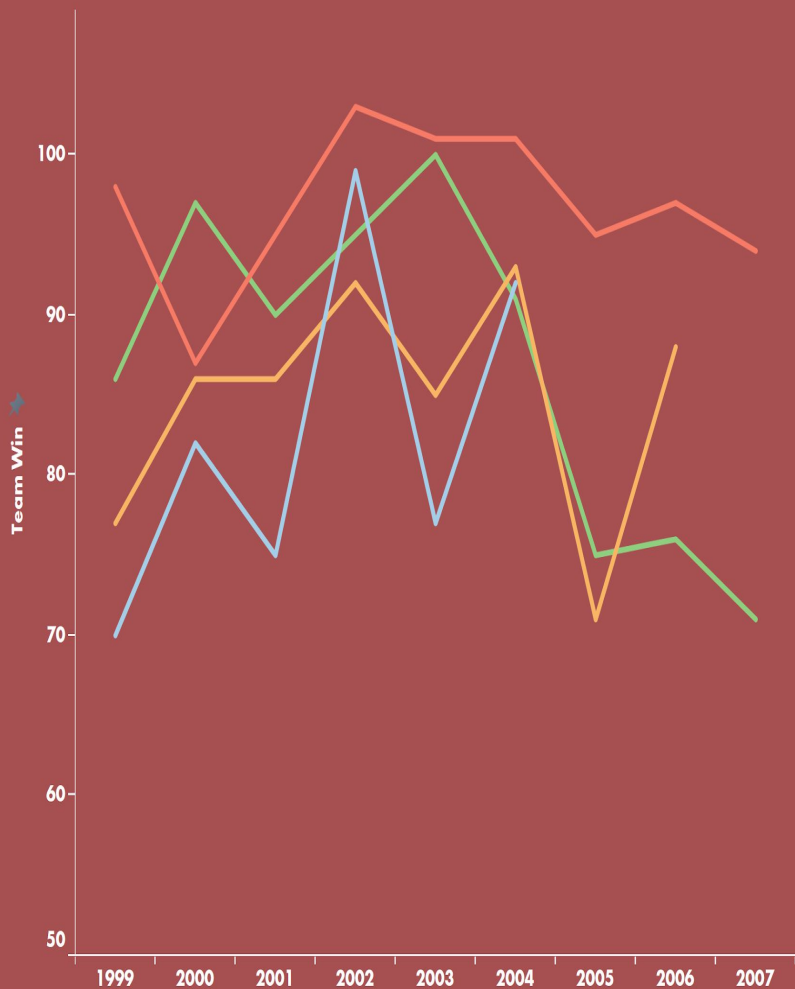
We wanted to take a deeper look at players that had ties to multiple providers

Barry Bonds, David Bell, Gary Sheffield and Jason Giambi seemed to have their peak around 2003-2007.

There are some numbers that dip from double digits to single digit home runs likely due to injuries.

We also observed some transfers amongst players which may be correlated to better performance from steroid usage

Team Wins



Teams

- Anaheim Angels
- Los Angeles Dodgers
- New York Highlande..
- San Francisco Giants

Insights

Showcasing the top 4 teams that have connections to providers have a peak of wins from 2000-2003.

Assumption: The player, team and provider network were at its high in between these years as we can tell some of the teams have a very high spike in team wins.

Key Provider Kirk Radomski

- Connected to 16 teams and 44 players
- Is one of the most connected provider who sells performing enhancement drugs
- It is beyond a player network, into the whole MLB organization
- Begs the question, did the MLB organization know Kirk's dealings prior to steroid accusations?
- Did the team managers allow providers to provide players enhancement drugs for their teams benefit?

	id	deg_in	deg_out	prank
1:	Kirk_Radomski	44	0	0.264911057
2:	Greg_Anderson	8	0	0.042553191
3:	Palm_Beach_Rejuvenation_Centre	6	0	0.039588420
4:	BALCO	5	0	0.024764562

	id	deg_in	deg_out	prank
:	Kirk_Radomski	16	0	0.22649573
:	BALCO	5	0	0.06023997
:	Greg_Anderson	5	0	0.05185733



By the end of 2005, Kirk Radomski says, he had supplied close to 300 major league players with performance-enhancing drugs.
Mike Fish/ESPN.com

But after a decade of running errands and cleaning up the clubhouse, Radomski says it was time to "make real money and stop playing," a decision that led to his role as baseball's No. 1 supplier of performance-enhancing drugs. Once he left the Mets, he held multiple jobs, most recently as co-owner of a car detailing business, while he was supplying the drugs.

He estimates that by the time federal agents showed up at the door of his Long Island home in December 2005, his client list had included more than 300 ballplayers through the years, ranging from marginal characters such as Pete Rose Jr. scuffling just to get a taste of the bigs to accomplished major leaguers such as Kevin Brown, Eric Gagne, Paul Lo Duca, Todd Hundley and -- indirectly, Radomski claims -- Miguel Tejada.

```
Call:
lm(formula = history$Salary ~ +history$team_win + history$Weight +
    history$`Year ID` + history$`Team ID`)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-7369624 -2138078 -441388  1728356 16286297
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -495535092    52979561  -9.353  < 2e-16 ***
history$team_win      6446      12432   0.519  0.60426
history$Weight      21473      7891   2.721  0.00668 **
history$`Year ID`    246130     26645  9.237  < 2e-16 ***
history$`Team ID`ARI  2830766    1198140  2.363  0.01843 *
history$`Team ID`ATL  1434927     975117  1.472  0.14162
history$`Team ID`BAL  1793269     936795  1.914  0.05602 .
history$`Team ID`BOS  2583732    1087924  2.375  0.01784 *
history$`Team ID`CAL  502511     2576555  0.195  0.84543
history$`Team ID`CHA -540784     1621164 -0.334  0.73880
history$`Team ID`CHN  1322356    1181501  1.119  0.26345
history$`Team ID`CIN  433642     1061017  0.409  0.68289
history$`Team ID`CLE  2077293    1132790  1.834  0.06714 .
history$`Team ID`COL  1482316    1137960  1.303  0.19316
history$`Team ID`DET  2580516    1454248  1.774  0.07645 .
history$`Team ID`FLO  3566493    1206483  2.956  0.00323 **
history$`Team ID`HOU  5364686    1293875  4.146  3.82e-05 ***
history$`Team ID`KCA  1098164     1135704  0.967  0.33393
history$`Team ID`LAA -1178950    1902554 -0.620  0.53569
history$`Team ID`LAN  2971547     947053  3.138  0.00178 **
history$`Team ID`MIL  383130     1516134  0.253  0.80058
history$`Team ID`MIN  1541213    1343097  1.148  0.25159
history$`Team ID`ML4  613290     1456041  0.421  0.67374
history$`Team ID`MON  613791     1061304  0.578  0.56323
history$`Team ID`NYA  4900092     898545  5.453  6.99e-08 ***
history$`Team ID`NYN  1197239     949749  1.261  0.20790
history$`Team ID`OAK  293140     1009817  0.290  0.77169
history$`Team ID`PHI  1650114    1017501  1.622  0.10534
history$`Team ID`PIT  795979     970631  0.820  0.41248
history$`Team ID`SDN  2023212    1268256  1.595  0.11113
history$`Team ID`SEA  645593     1345069  0.480  0.63141
history$`Team ID`SFN  3302632     891663  3.704  0.00023 ***
history$`Team ID`SLN  258984     1040881  0.249  0.80358
history$`Team ID`TBA  127614     1737294  0.073  0.94147
history$`Team ID`TEX  1153150    1102630  1.046  0.29603
history$`Team ID`TOR  1561538    1153576  1.354  0.17631
history$`Team ID`WAS -966310     1461192 -0.661  0.50864
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3470000 on 660 degrees of freedom
Multiple R-squared:  0.3002,    Adjusted R-squared:  0.2621
F-statistic: 7.865 on 36 and 660 DF,  p-value: < 2.2e-16
```

Regression Analysis

We built a model to predict each player's salary based on number of team wins, weight of the player, which team the player belonged to and which year.

This was our best model and only explained 30% of variation in salaries.

We found that some variables such as weight remained consistent despite the use of steroids which is abnormal. We believe that other factors may have influenced salary that can not be explained in the regression model and may also import new features in our next steps to see if the model improves.



04

Next Steps

Moving Forward

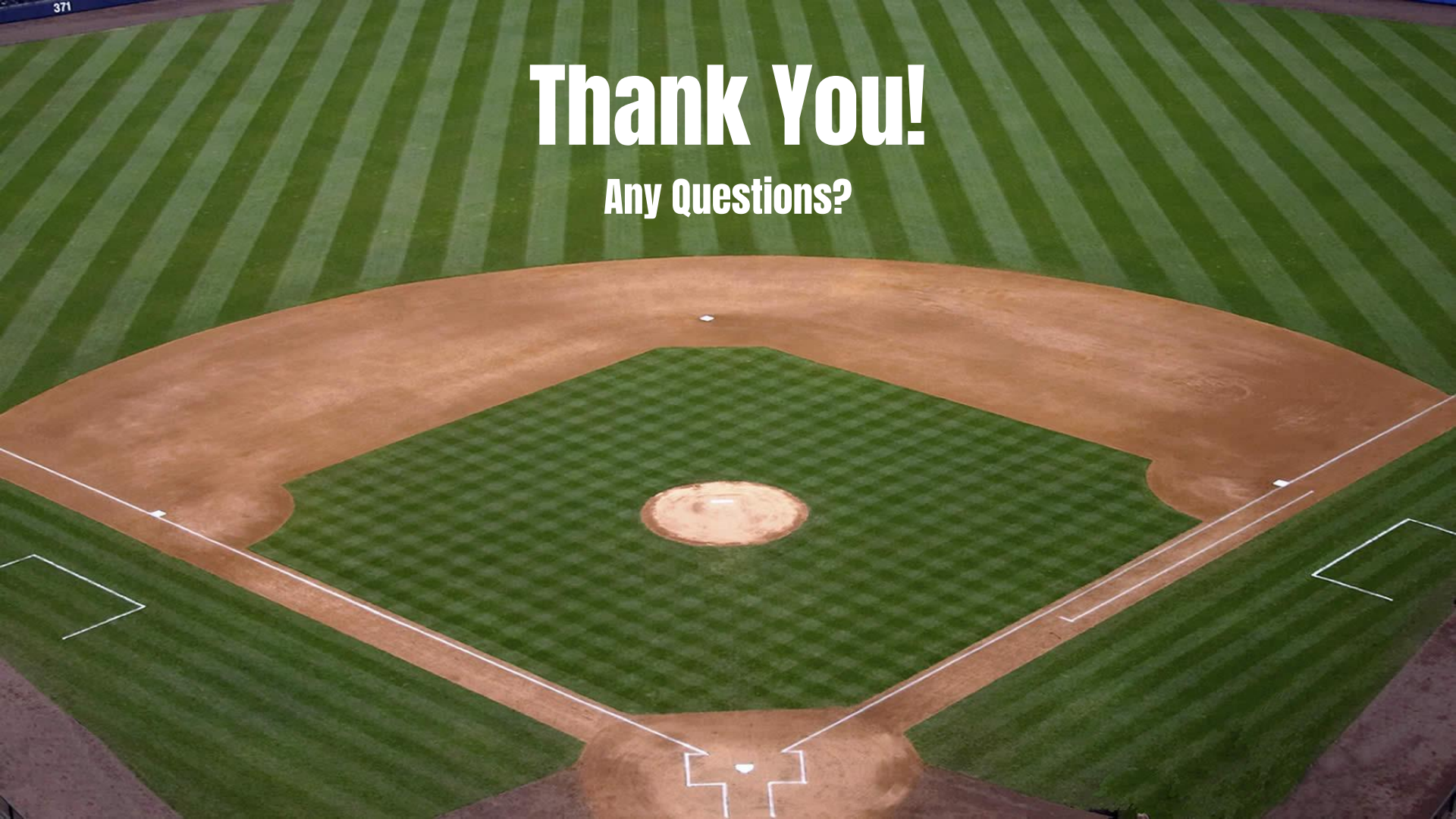
- Build on current regression models by adding attributes that may help drive insights
- Analyze diffusion of steroid usage within the network
- Dig deeper into the top performance players and analyze whether being on steroid affected their career in 5 years after the scandal



371

Thank You!

Any Questions?





05

Appendix

		id	deg_in	deg_out	prank
1:	Kirk_Radomski	44	0	0.264911057	
2:	Greg_Anderson	8	0	0.042553191	
3:	Palm_Beach_Rejuvenation_Centre	6	0	0.039588420	
4:	BALCO	5	0	0.024764562	
5:	Signature_Pharmacy	5	0	0.022293919	
6:	Applied_Pharmacy_Services	4	0	0.019329148	
7:	BrianMcNamee	3	0	0.024764562	
8:	New_Hope_Health_Centre	2	0	0.018835019	
9:	American_Pharmaceutical_Group	2	0	0.015870248	
10:	Health_Rejuvenation_Centre	2	0	0.014387862	
11:	Arizona_AntiAging_Clinic	1	0	0.008952447	
12:	Health_Watch_Clinic	1	0	0.008458319	
13:	Armando_Rios	0	2	0.006975933	
14:	Barry_Bonds	0	2	0.006975933	
15:	Gary_Sheffield	0	2	0.006975933	
16:	Jason_Giambi	0	2	0.006975933	
17:	Jeremy_Giambi	0	1	0.006975933	
18:	Benito_Santiago	0	1	0.006975933	
19:	Bobby_Estalella	0	1	0.006975933	
20:	Marvin_Benard	0	1	0.006975933	
21:	Randy_Velarde	0	1	0.006975933	
22:	Darren_Holmes	0	1	0.006975933	
23:	Ismael_Valdez	0	1	0.006975933	
24:	Jason_Grimsley	0	2	0.006975933	
25:	Jose_Guillen	0	1	0.006975933	
26:	Matt_Williams	0	1	0.006975933	
27:	Paul_Byrd	0	1	0.006975933	
28:	Scott_Schoeneweis	0	1	0.006975933	
29:	Troy_Glaus	0	1	0.006975933	
30:	David Bell	0	3	0.006975933	

	id	deg_in	deg_out	prank
:	Kirk_Radomski	16	0	0.22649573
:	BALCO	5	0	0.06023997
:	Greg_Anderson	5	0	0.05185733
:	Palm_Beach_Rejuvenation_Centre	4	0	0.06582840
:	Signature_Pharmacy	2	0	0.03229783
:	New_Hope_Health_Centre	1	0	0.02810651
:	American_Pharmaceutical_Group	1	0	0.02810651
:	Applied_Pharmacy_Services	1	0	0.02391519
:	Arizona_AntiAging_Clinic	1	0	0.02391519
:	BrianMcNamee	1	0	0.02531229

	teams	Supplier	Ties
1	Los Angeles Dodgers	Kirk_Radomski	4
1	San Francisco Giants	Greg_Anderson	3
7	Atlanta Braves	Palm_Beach_Rejuvenation_Centre	2
7	Anaheim Angels	New_Hope_Health_Centre	2
1	Baltimore Orioles	American_Pharmaceutical_Group	2
3	New York Highlanders	BrianMcNamee	2
5	Baltimore Orioles	Kirk_Radomski	2
9	Cincinnati Reds	Kirk_Radomski	2
7	Oakland Athletics	Kirk_Radomski	2
3	Atlanta Braves	BALCO	1
5	Boston Americans	BALCO	1
5	Chicago White Sox	BALCO	1
3	New York Highlanders	BALCO	1
9	San Francisco Giants	BALCO	1
5	Atlanta Braves	Greg_Anderson	1

```

Call:
lm(formula = history$Salary ~ +history$team_win + history$Weight +
    history$`Year ID` + history$`Team ID`)

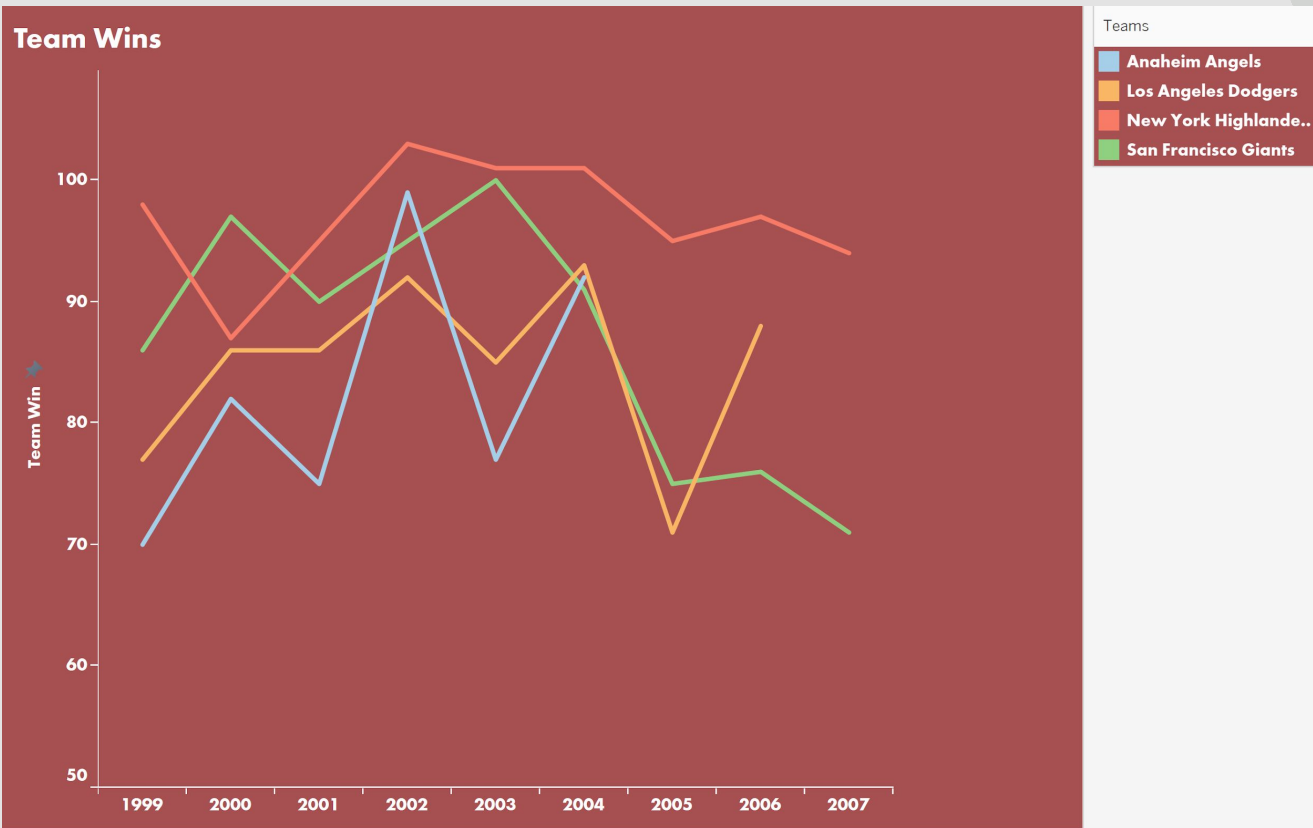
Residuals:
    Min       1Q   Median       3Q      Max
-7369624 -2138078  -441388  1728356 16286297

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -495535092    52979561  -9.353  < 2e-16 ***
history$team_win      6446       12432    0.519  0.60426
history$Weight      21473       7891    2.721  0.00668 **
history$`Year ID`    246130     26645    9.237  < 2e-16 ***
history$`Team ID`ARI  2830766    1198140   2.363  0.01843 *
history$`Team ID`ATL  1434927    975117    1.472  0.14162
history$`Team ID`BAL  1793269    936795    1.914  0.05602 .
history$`Team ID`BOS  2583732    1087924   2.375  0.01784 *
history$`Team ID`CAL  502511    2576555   0.195  0.84543
history$`Team ID`CHA -540784    1621164  -0.334  0.73880
history$`Team ID`CHN  1322356    1181501   1.119  0.26345
history$`Team ID`CIN  433642    1061017   0.409  0.68289
history$`Team ID`CLE  2077293    1132790   1.834  0.06714 .
history$`Team ID`COL  1482316    1137960   1.303  0.19316
history$`Team ID`DET  2580516    1454248   1.774  0.07645 .
history$`Team ID`FLO  3566493    1206483   2.956  0.00323 **
history$`Team ID`HOU  5364686    1293875   4.146 3.82e-05 ***
history$`Team ID`KCA  1098164    1135704   0.967  0.33393
history$`Team ID`LAA -1178950    1902554  -0.620  0.53569
history$`Team ID`LAN  2971547    947053   3.138  0.00178 **
history$`Team ID`MIL  383130    1516134   0.253  0.80058
history$`Team ID`MIN  1541213    1343097   1.148  0.25159
history$`Team ID`ML4  613290    1456041   0.421  0.67374
history$`Team ID`MON  613791    1061304   0.578  0.56323
history$`Team ID`NYA  4900092    898545   5.453 6.99e-08 ***
history$`Team ID`NYN  1197239    949749   1.261  0.20790
history$`Team ID`OAK  293140    1009817   0.290  0.77169
history$`Team ID`PHI  1650114    1017501   1.622  0.10534
history$`Team ID`PIT  795979    970631   0.820  0.41248
history$`Team ID`SDN  2023212    1268256   1.595  0.11113
history$`Team ID`SEA  645593    1345069   0.480  0.63141
history$`Team ID`SFN  3302632    891663   3.704  0.00023 ***
history$`Team ID`SLN  258984    1040881   0.249  0.80358
history$`Team ID`TBA  127614    1737294   0.073  0.94147
history$`Team ID`TEX  1153150    1102630   1.046  0.29603
history$`Team ID`TOR  1561538    1153576   1.354  0.17631
history$`Team ID`WAS -966310    1461192  -0.661  0.50864
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

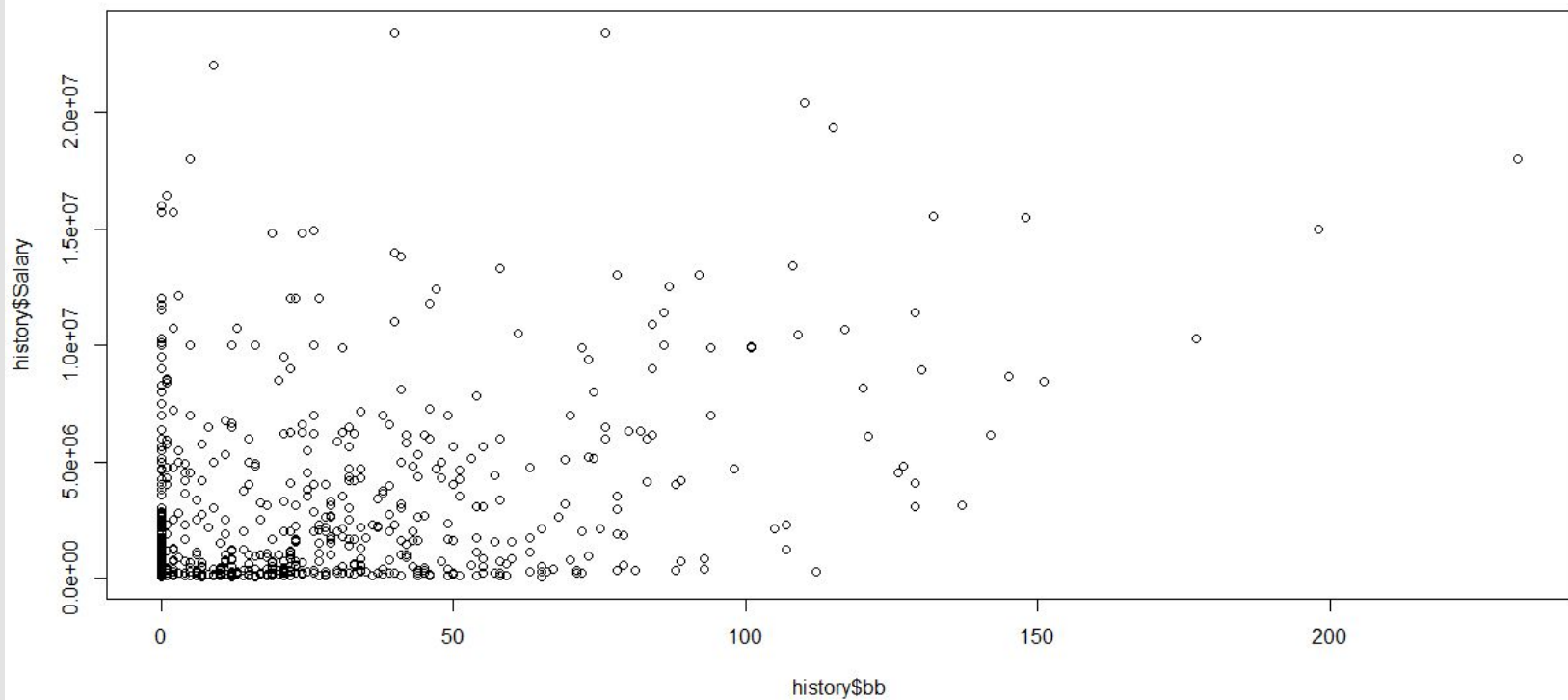
Residual standard error: 3470000 on 660 degrees of freedom
Multiple R-squared:  0.3002,    Adjusted R-squared:  0.2621
F-statistic: 7.865 on 36 and 660 DF,  p-value: < 2.2e-16

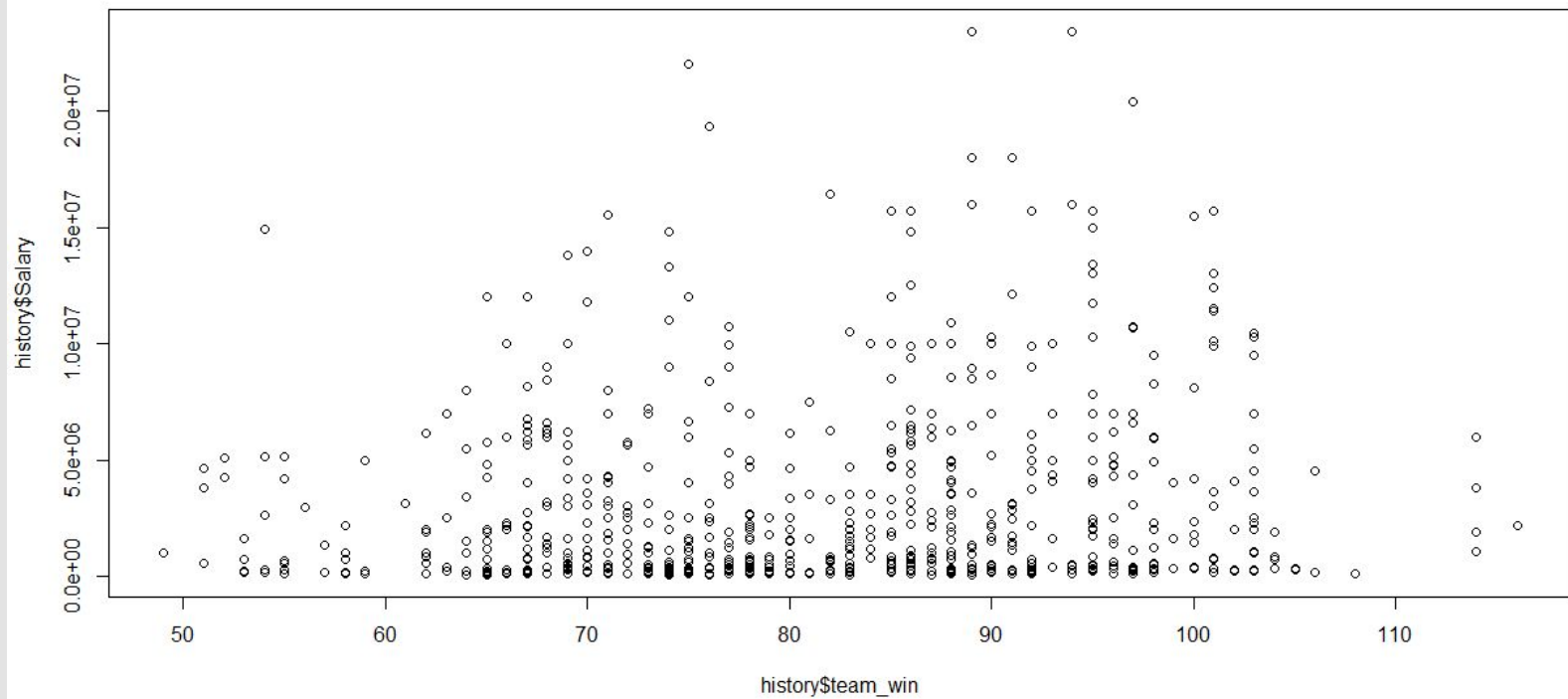
```

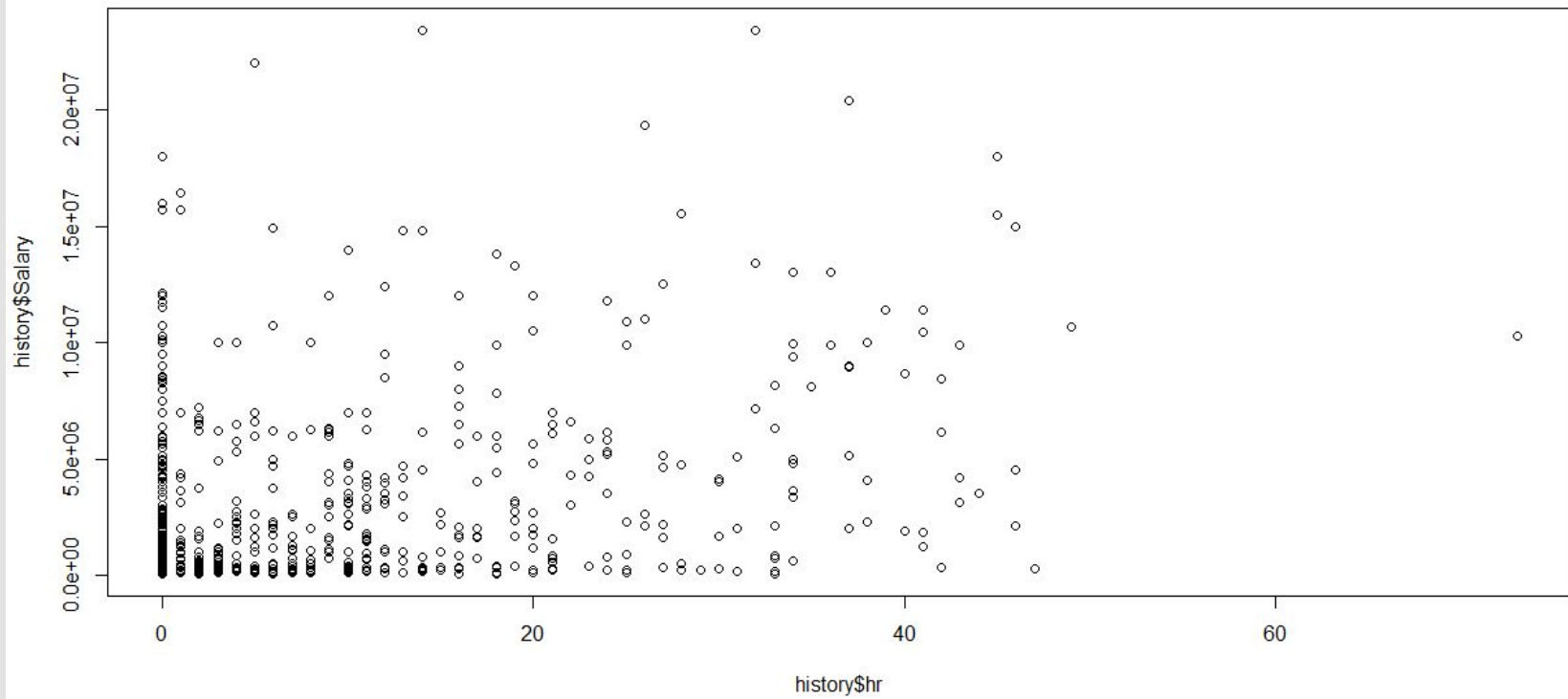
Player Salary By Year											
		Year ID									
First Name Last Name	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Armando_Rios		200,000	225,000	310,000	925,000	450,000					
Barry_Bonds	8,916,667	9,381,057	10,658,826	10,300,000	15,000,000	15,500,000	18,000,000	22,000,000	19,331,470	15,533,970	
David_Bell	200,000	700,000	1,425,000	2,200,000	1,750,000	3,200,000	4,400,000	4,700,000	4,700,000		
Gary_Sheffield	14,936,667	9,956,667	9,916,667	9,916,667	9,916,667	11,416,667	13,000,000	13,000,000	10,756,171	10,916,071	13,326,306
Jason_Giambi	315,000	2,103,333	3,103,333	4,103,333	10,428,571	11,428,571	12,428,571	13,428,571	20,428,571	23,428,571	23,428,571
Jason_Grimsley		350,000	750,000	750,000	1,900,000	2,000,000	1,000,000	2,000,000	825,000		
Jerry_Hairston			205,000	235,000	300,000	1,550,000	1,650,000	1,800,000	2,300,000	750,000	
Jose_Canseco	2,125,000	3,325,000	3,000,000								



First Name Last Name	Year ID	Teams	
Armando_Rios	2003	Chicago White Sox	2
Barry_Bonds	2003	San Francisco Giants	45
	2004	San Francisco Giants	45
	2005	San Francisco Giants	5
	2006	San Francisco Giants	26
	2007	San Francisco Giants	28
David_Bell	2003	Philadelphia Quakers	4
	2004	Philadelphia Quakers	18
	2005	Philadelphia Quakers	10
	2006	Philadelphia Quakers	6
Gary_Sheffield	2003	Atlanta Braves	39
	2004	New York Highlanders	36
	2005	New York Highlanders	34
	2006	New York Highlanders	6
	2007	Detroit Tigers	25
Jason_Giambi	2003	New York Highlanders	41
	2004	New York Highlanders	12
	2005	New York Highlanders	32
	2006	New York Highlanders	37
	2007	New York Highlanders	14







Histogram of history\$Salary

