NBA Player Salary & Metrics

Jordan Pflum, Namit Agrawal,
Kessiena Ofunrein, Shifan Hu,
Timothy Cheng, Jenny Robinson

Process

Identify scope
Using player metrics as a predictor of player salary

Variable selection
LASSO regression and Stepwise

Analysis
Compared usefulness and accuracy of models

Create all data frames
Grouped by positions, age, and no grouping

Models & prediction
Variety of models used, including regression, random forests, boosting

Interpretation
What did we conclude, and why does this matter?

Data Wrangling

CSVs

- NBA Salary: 1985-2017
- NBA Salary Cap: 1985-2017
- Advanced NBA Player Metrics: 1985-2017

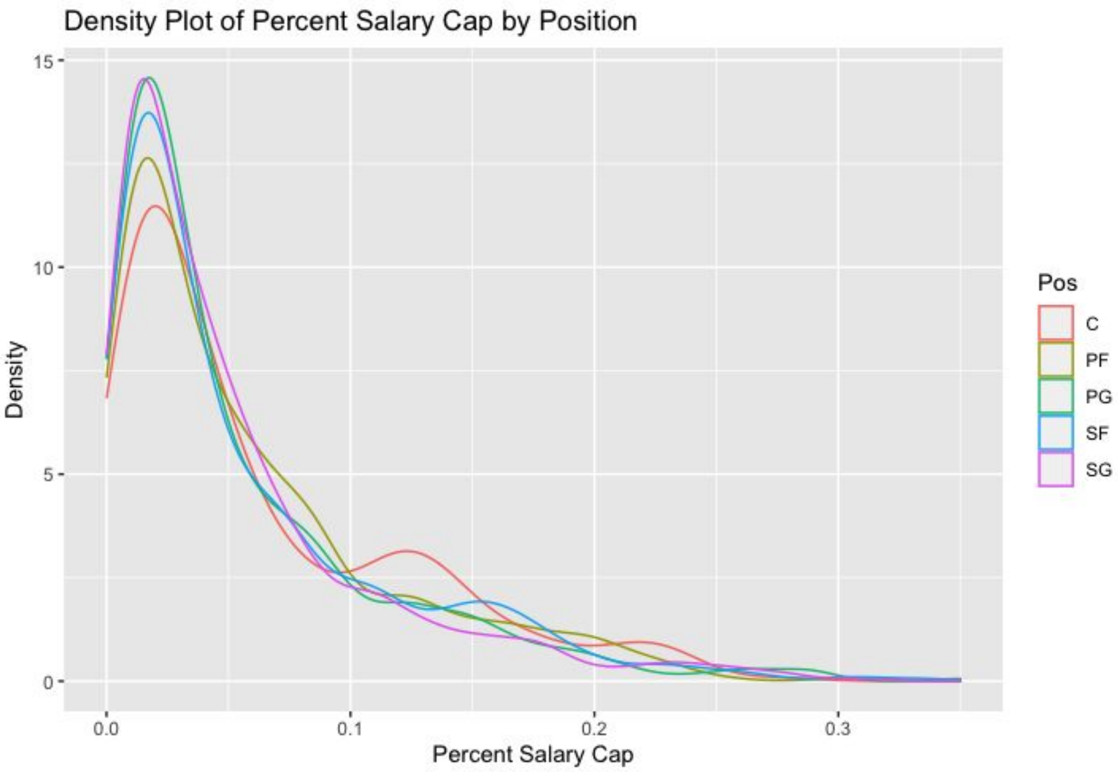
Manipulation

- Traded Players
  - Summarized Stats per Season
- Time Frame
  - 2010-2017

Final Dataframe

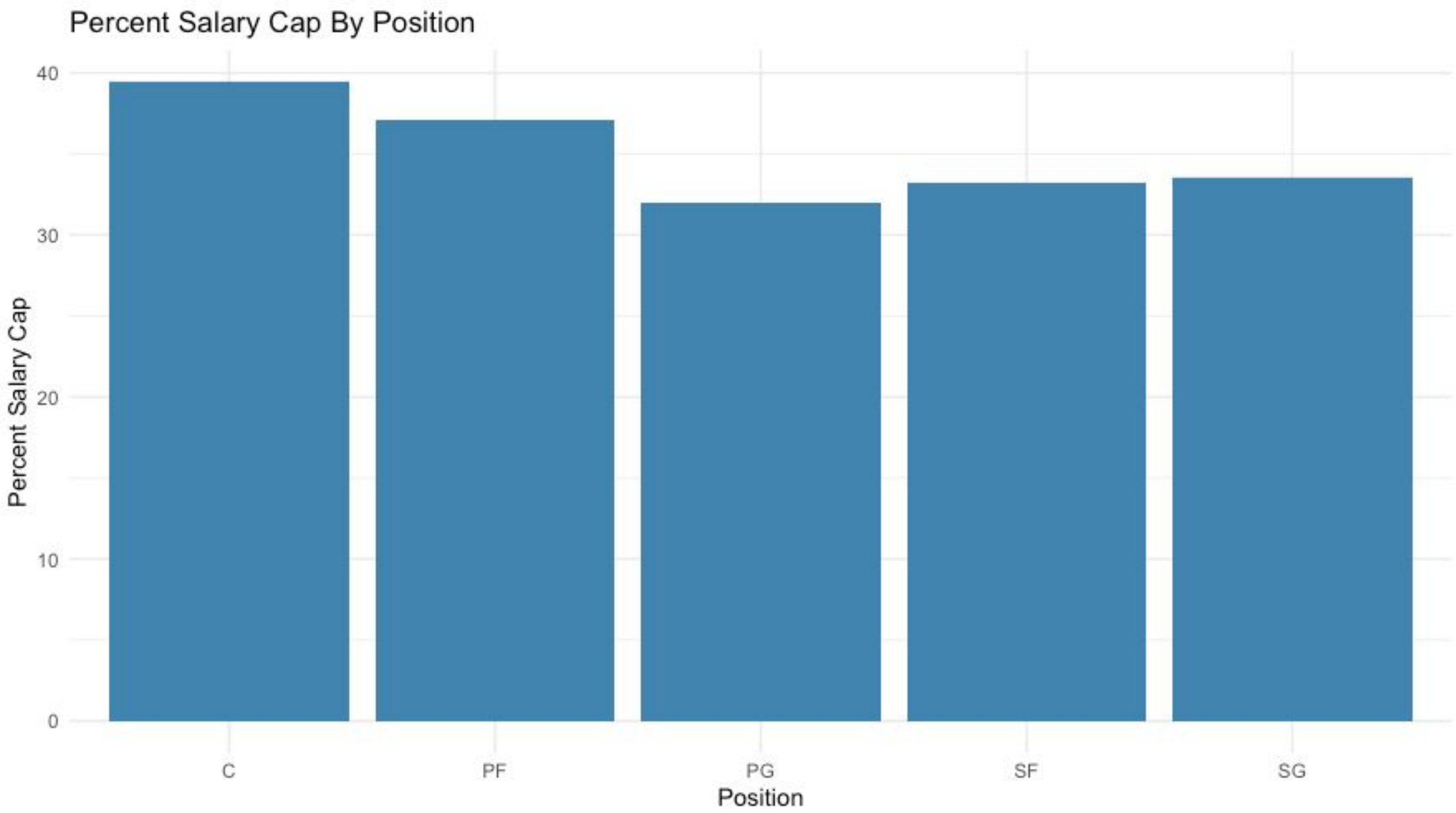
- 3028 Observations
- 50 Predictors
  - PTS
  - TS%
  - OWS
- Response Variable
  - Player Salary % Salary Cap

Analysis of Data



Identify scope → **Data frame creation** → Variable selection → Models/Analysis → Interpretation

Analysis of Data



Identify scope → **Data frame creation** → Variable selection → Models/Analysis → Interpretation

Model Subsetting

Complete

- Complete salary and metric data frame

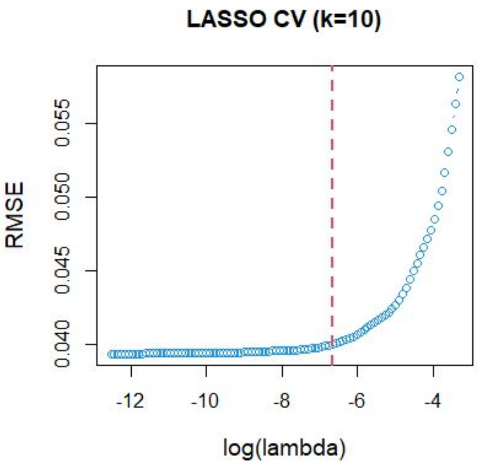
Position

- Shooting guard; Center; Power Forward; Small Forward; Point Guard
- Position indicates different value placed on metrics

Age

- 18-22; 23-26; 27-30; 31-35; >35
- Age indicates experience and reputation

Variable Selection



ex. (3/20 coefficients)

Intercept	0.058
Age	0.01
G	-0.01

ex. (5/19 coefficients)

Coefs	**Estimate**	**p-value**
Age	2.58e-3	<2e-16
G	-4.38e-4	1.57e-13
DRB	5.62e-5	5.51e-7
PF	-9.3e-5	1.14e-4
TSA	4.227e-5	9.35e-16

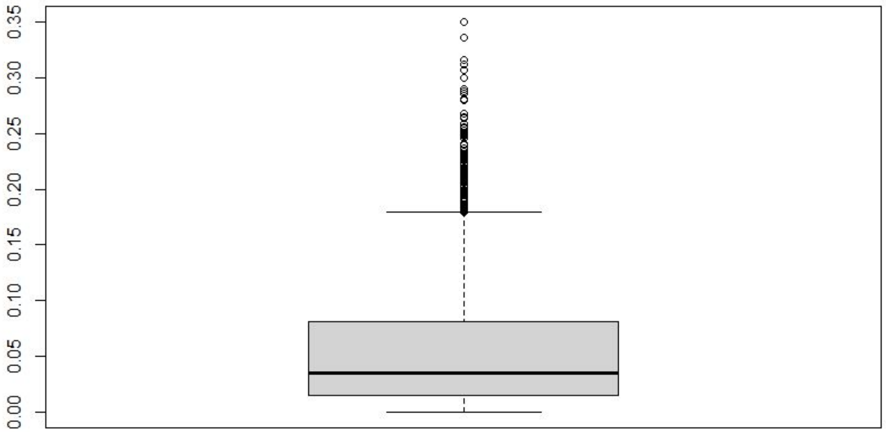
LASSO RMSE: 0.04001212 % salary cap
(3.96 $million)
LASSO adjR^2 = 0.535

Stepwise RMSE: 0.03 % salary cap
(3.1 $million)
Stepwise adjR^2 = 0.537

Model Comparison (Out-Of-Sample RMSE)

(in $millions)	No Grouping	Shooting Guard	Center	Power Forward	Small Forward	Point Guard
Multiple Linear Regression	3.06	3.58	4.04	2.29	3.58	3.62
Random Forest	3.08	3.34	3.93	2.23	3.63	3.37
Boosting	2.85	3.19	3.88	2.14	3.50	3.02

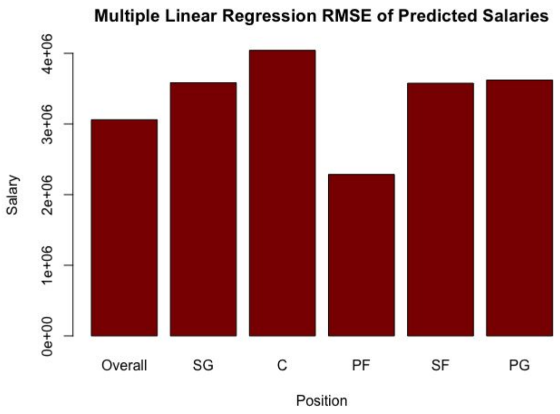
Salary % of Salary Cap Boxplot

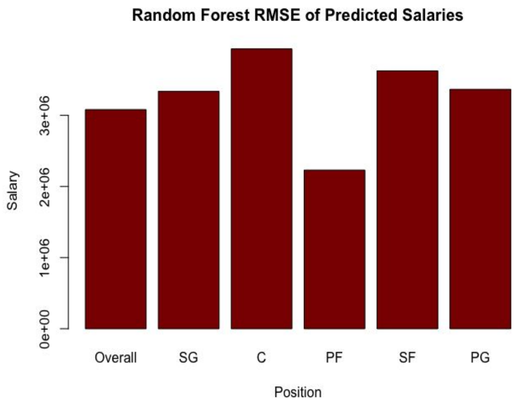


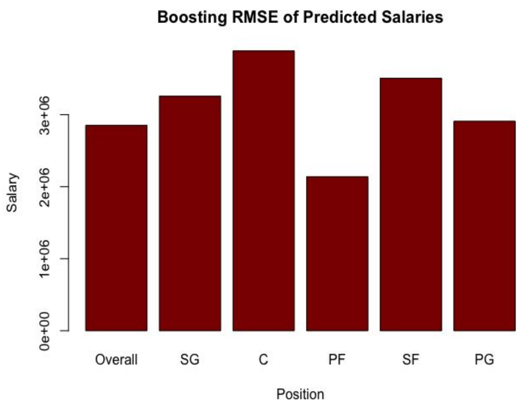
Boxplot Summary

- Min: 0.0000748
- 1st Q: 0.0148
- Median: 0.035
- 3rd Q: 0.081
- Max: 0.35

RMSE Graphs







Identify scope → Data frame creation → Variable selection → **Models/Analysis** → Interpretation

Model Comparison (Adj R2)

(adjR^2)	No Grouping	Shooting Guard	Center	Power Forward	Small Forward	Point Guard
Multiple Linear Regression	0.5373	0.5271	0.5277	0.6025	0.5461	0.5436
Random Forest	0.5466	0.2336	0.1175	0.4179	0.2719	0.4397
Boosting	0.6131	0.3022	0.1444	0.4635	0.3181	0.5495

Test Case (Boosting - Pos Grouping [Center])

Clint Capela

Predicted $6,750,428.20

Actual $2,334,520.00

Difference Attributable To:

- ❑ Young Player (First Contract)

What's the Point?

- Player salary possibly based on more than performance than outside factors
- Really about the GM's bottom line - revenue, endorsements, etc.
- Position doesn't matter in the long run