



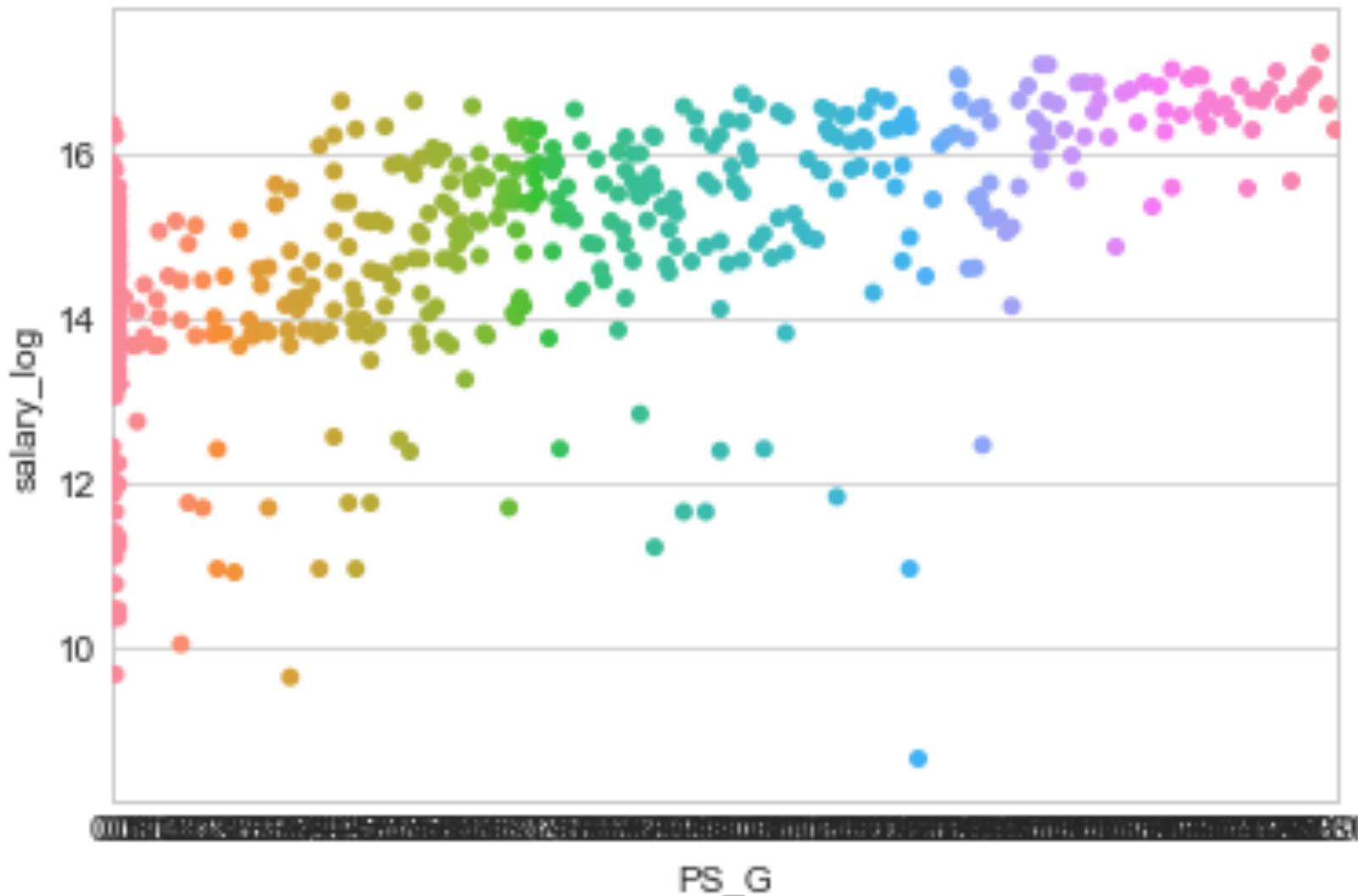
NBA | f | @ | d | t | #NBAAllStar



# 3 EASY WAYS TO REDUCE YOUR R SQUARED WITHOUT ANY EXERCISE!

PREDICTING NBA SALARIES FROM INDIVIDUAL PLAYER  
STATISTICS

Points Per Game and Salary



## OUR MODEL: OLS LINEAR REGRESSION TO PREDICT SALARIES

1

Minutes  
played

2

Points per  
game

3

Free throw  
percentage

4

Age

5

Games  
Started

## LOG OF SALARY

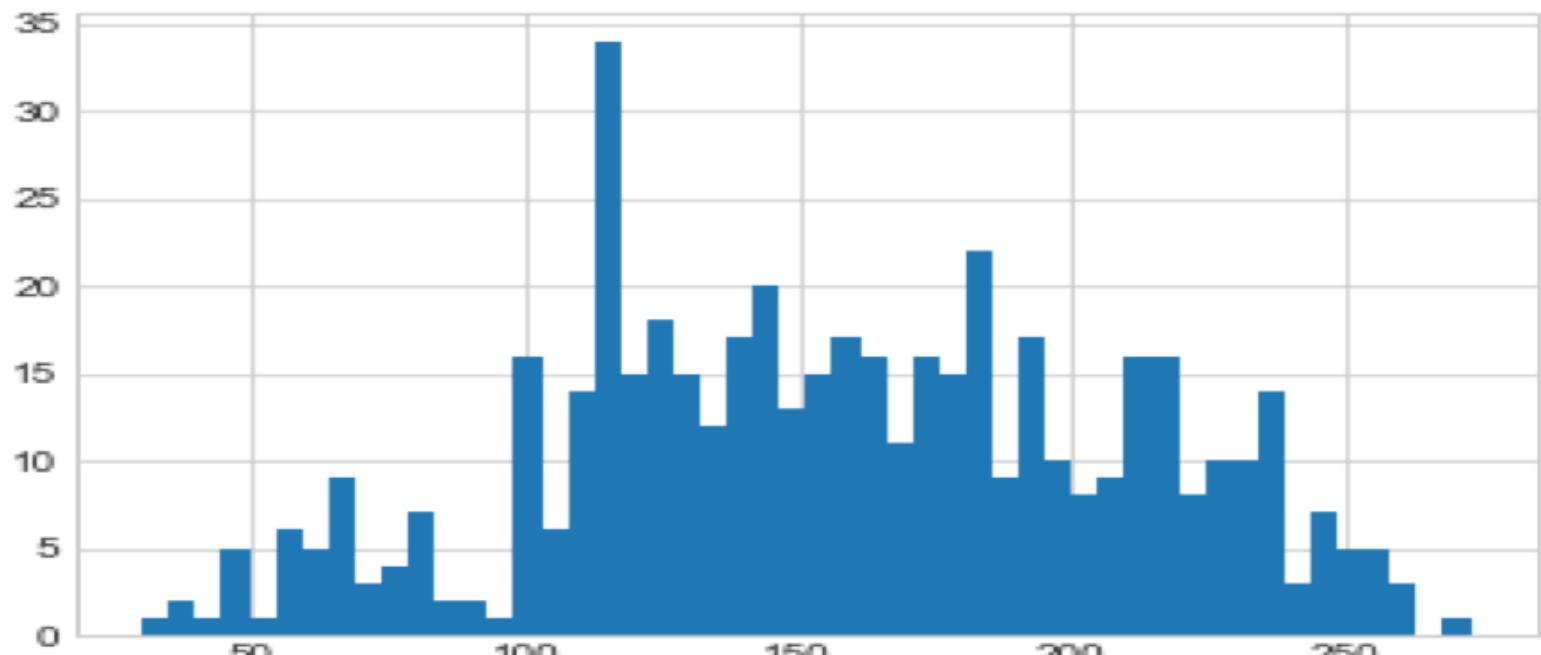
- R-squared .355
- Adjusted R-squared : .349

# FIRST EASY WAY TO IMPROVE R SQUARED: BOXCOX ON RESPONSE

OLS Regression Results

Dep. Variable:	salary_bx	R-squared:	0.459			
Model:	OLS	Adj. R-squared:	0.453			
Method:	Least Squares	F-statistic:	82.34			
Date:	Thu, 25 Jan 2018	Prob (F-statistic):	1.56e-62			
Time:	23:37:15	Log-Likelihood:	-2485.8			
No. Observations:	492	AIC:	4984.			
Df Residuals:	486	BIC:	5009.			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	115.6678	3.494	33.106	0.000	108.803	122.533
MP	-0.0010	0.505	-0.002	0.998	-0.993	0.992
PS_G	4.8378	0.762	6.353	0.000	3.342	6.334
FT_PCT	-24.3571	12.094	-2.014	0.045	-48.120	-0.595
Age	0.8563	0.337	2.541	0.011	0.194	1.518
GS	0.2080	0.100	2.085	0.038	0.012	0.404

# BOX COX OF SALARY





SECOND EASY  
WAY!

ADD MORE  
FEATURES

- Age+Games+Games  
Started+Minutes Field Goals  
Turnovers, Blocks, Points  
Scored...etc.
- 26 Features in all.

### OLS Regression Results

Dep. Variable:	salary_bx	R-squared:	0.530
Model:	OLS	Adj. R-squared:	0.503
Method:	Least Squares	F-statistic:	20.14
Date:	Thu, 25 Jan 2018	Prob (F-statistic):	2.55e-60
Time:	23:37:16	Log-Likelihood:	-2451.2
No. Observations:	492	AIC:	4956.
Df Residuals:	465	BIC:	5070.
Df Model:	26		

## R SQUARED

- Increase of 18 percentage points when combined with box cox.

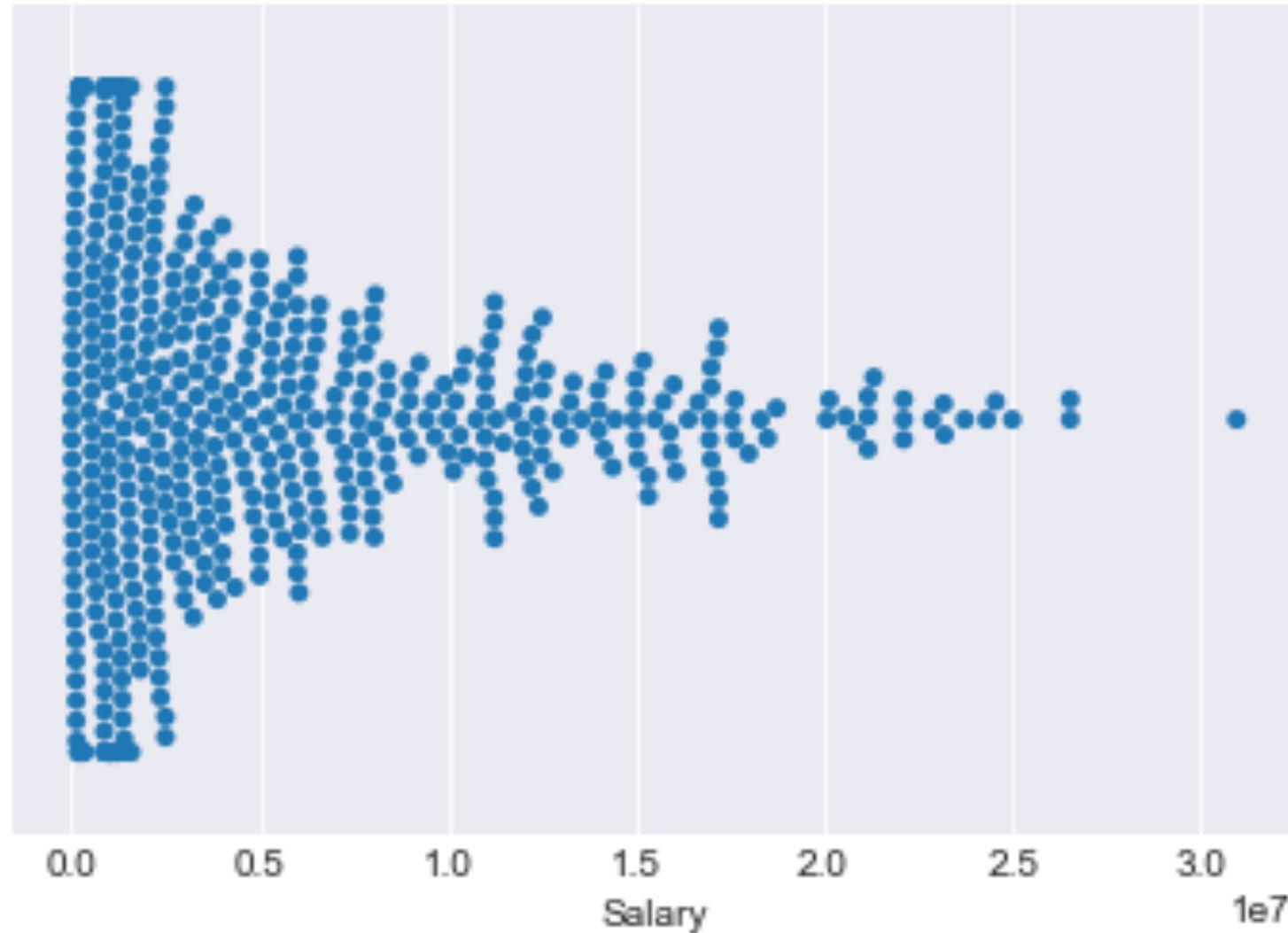
## CROSS VALIDATION

```
[ -21.91876382 -77.81449695 -40.15629648 -76.36186083 -39.98180271  
-55.50534172 -31.88129181 -42.18273678 -10.28833926 -36.61898295 ]
```

## STEP 3 FOR INCREASING R SQUARED.

Cut outliers: Salaries only above  
\$500k and top 20 players

### Swarm Plot NBA Salaries



### OLS Regression Results

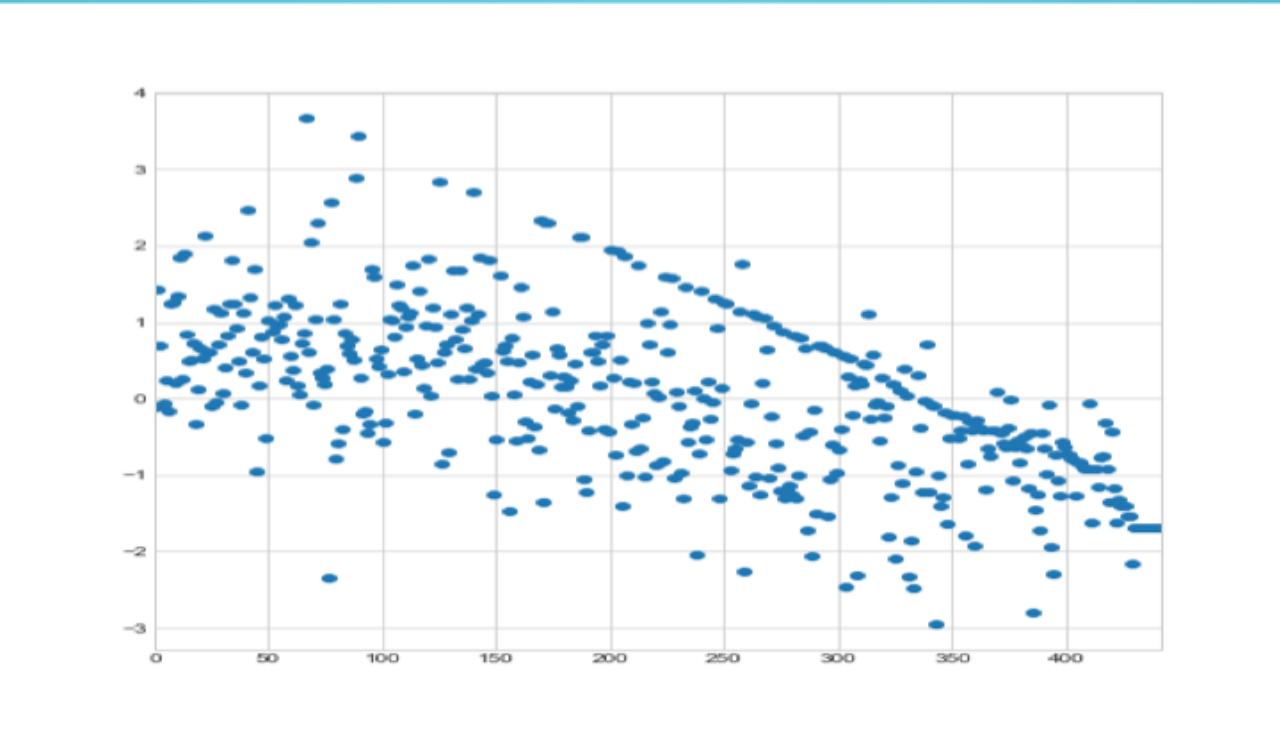
<b>Dep. Variable:</b>	salary_bx	<b>R-squared:</b>	0.626
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.602
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	26.75
<b>Date:</b>	Fri, 26 Jan 2018	<b>Prob (F-statistic):</b>	1.20e-72
<b>Time:</b>	09:32:16	<b>Log-Likelihood:</b>	-669.04
<b>No. Observations:</b>	443	<b>AIC:</b>	1392.
<b>Df Residuals:</b>	416	<b>BIC:</b>	1503.
<b>Df Model:</b>	26		
<b>Covariance Type:</b>	nonrobust		

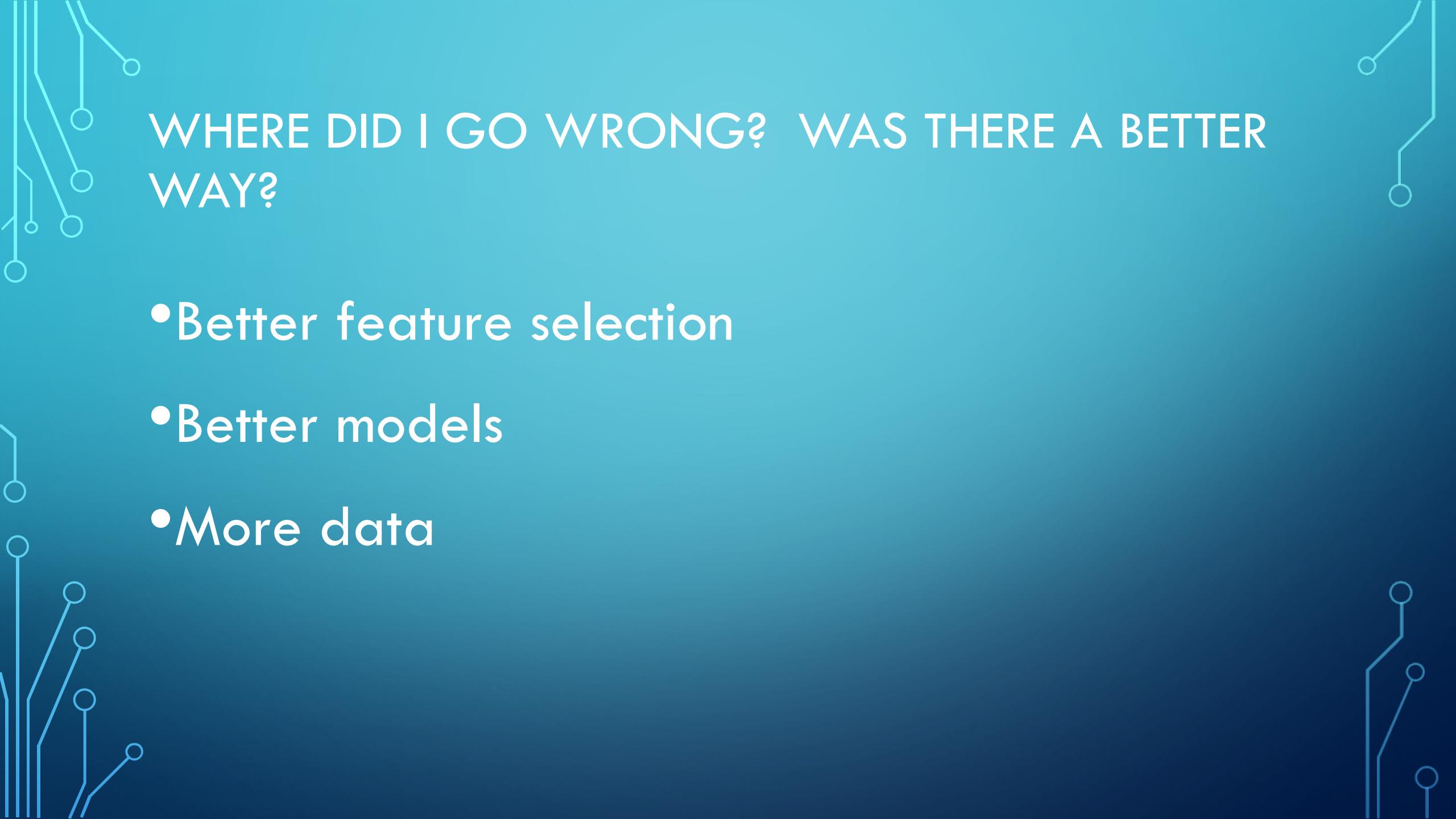
R sq = .626

# NEGATIVE R-SQUAREDS!

```
[ -84.44429321 -49.02014531 -73.34595448 -37.18750537 -46.88348304  
-63.09030911 -52.058176 -55.92835573 -11.02507753 -45.49065468 ]
```

# RESIDUAL PLOT





WHERE DID I GO WRONG? WAS THERE A BETTER WAY?

- Better feature selection
- Better models
- More data