SH Movies success implies + TV Series ?

Metis SF Spring 2018 – Project Luther
Filippo Medri

IMPORTANT NOTE – do not edit this slide, builds content art elements before going to next content slide. Should go between title/section title slides and content slides.

Problem Statement

- **GOAL**

“Does the success of Super Hero Movies affects the number of Super Hero TV Series of the next year? ”

- Scraped data from 2501 Titles
  - Source: www.IMDB.com
  - From (“The Green Archer - 1920”)
  - To (“Defenders of the Gao – 2020”)

- Created 2 Different Dataset:
  - Movies ← For feature extraction
  - TV Series ← For response variable

Features

Movies

- IMDB Score
- Number of IMDB users rating
- Metascore from Metacritics
- N. of reviews from users
- N. of reviews from critics
- Opening WE gross (USA)
- Gross (USA)
- Gross (world wide)

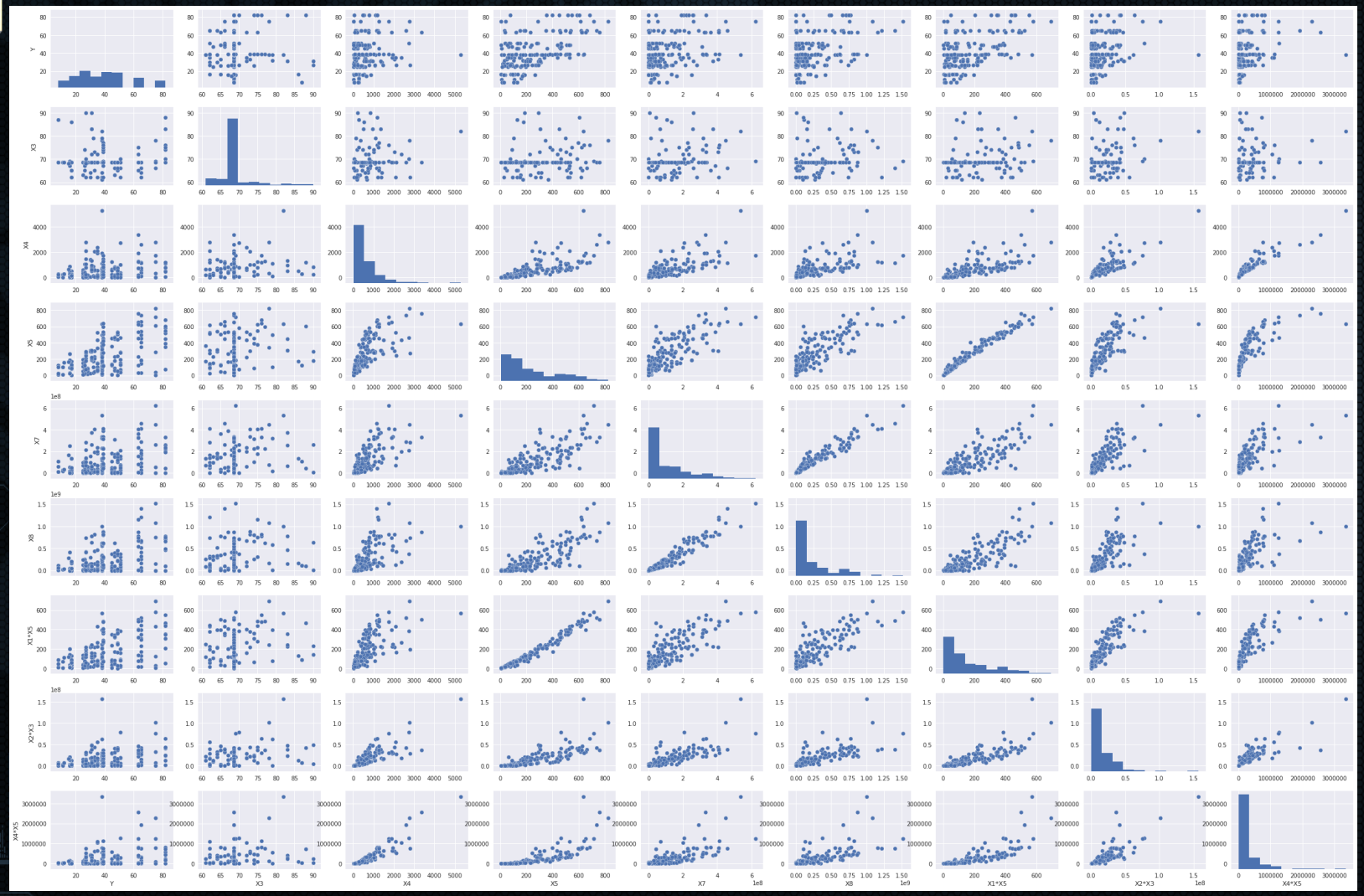
Tv Series

- Number of TV Series per year

Final Dataset

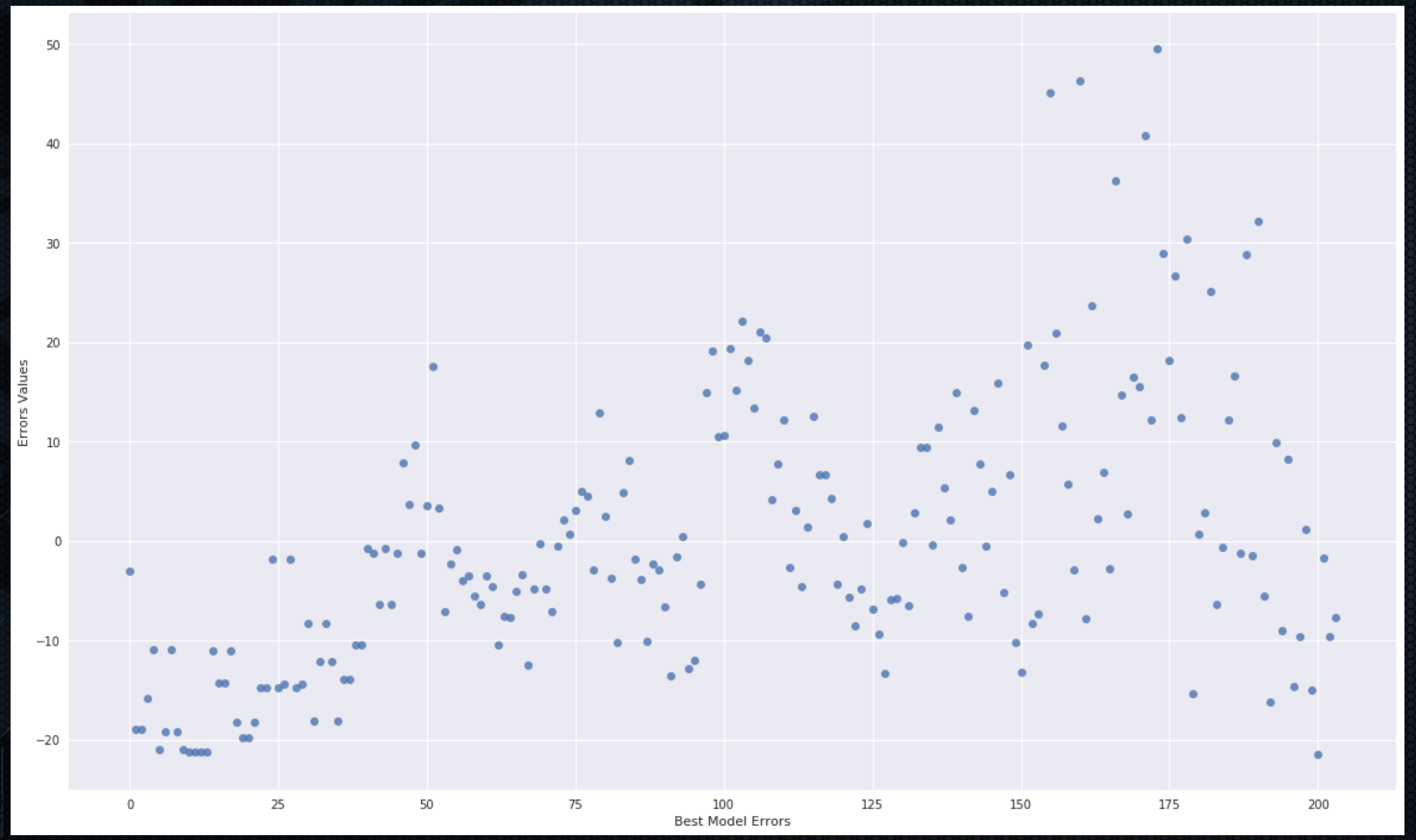
- 204 Entries
- From 1978 to 2018

Best Model



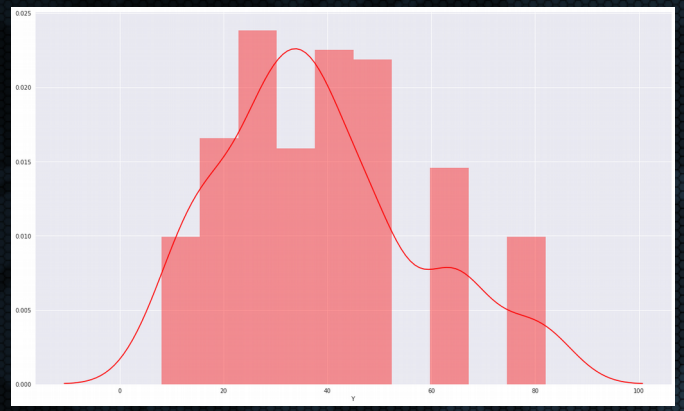
- **Facts**
- Important Variables:
  - Metascore
  - N. Reviews (User)
  - N.Reviews (Critic)
  - Gross (USA)
  - Gross (WW)
  - Score combined with N.reviews - Crit
  - User rating combined with Metascore
  - N.reviews-users combined N.reviews Critic
- Opening WE disappear !

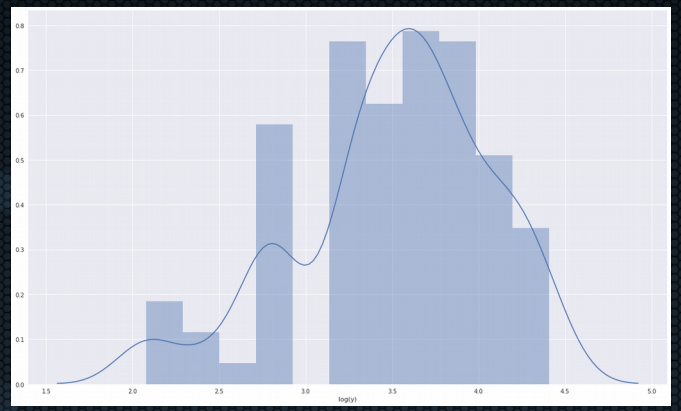
Best Model - Results

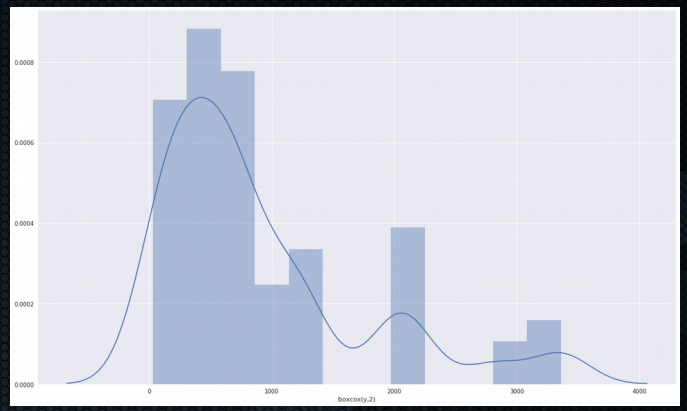


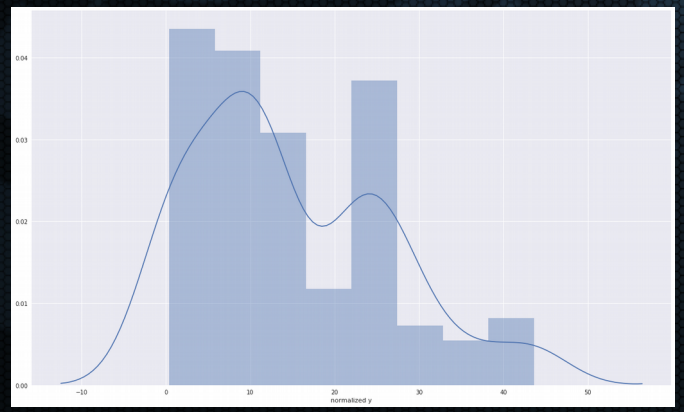
- Preliminary
- Adjusted R^2 ~ 0.4
- Train/Test Split
- Adjusted R^2 ~ 0.5
- Graph shows Heteroscedasticity

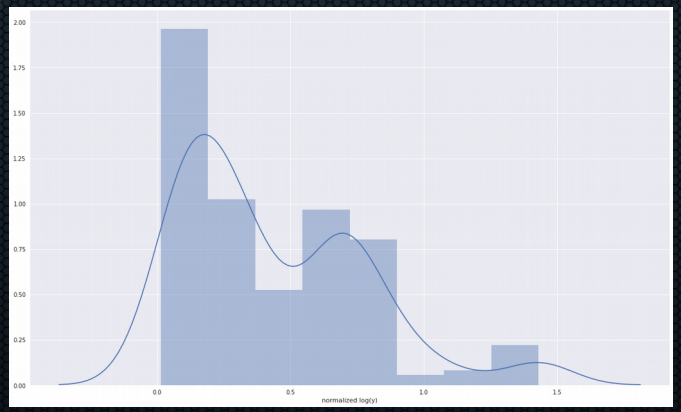
Can we improve by transform response?

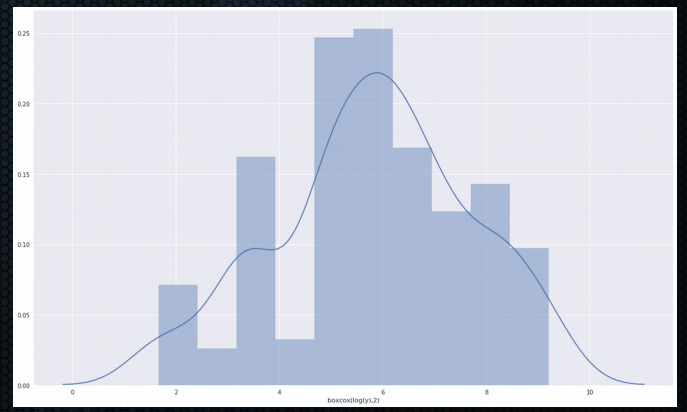






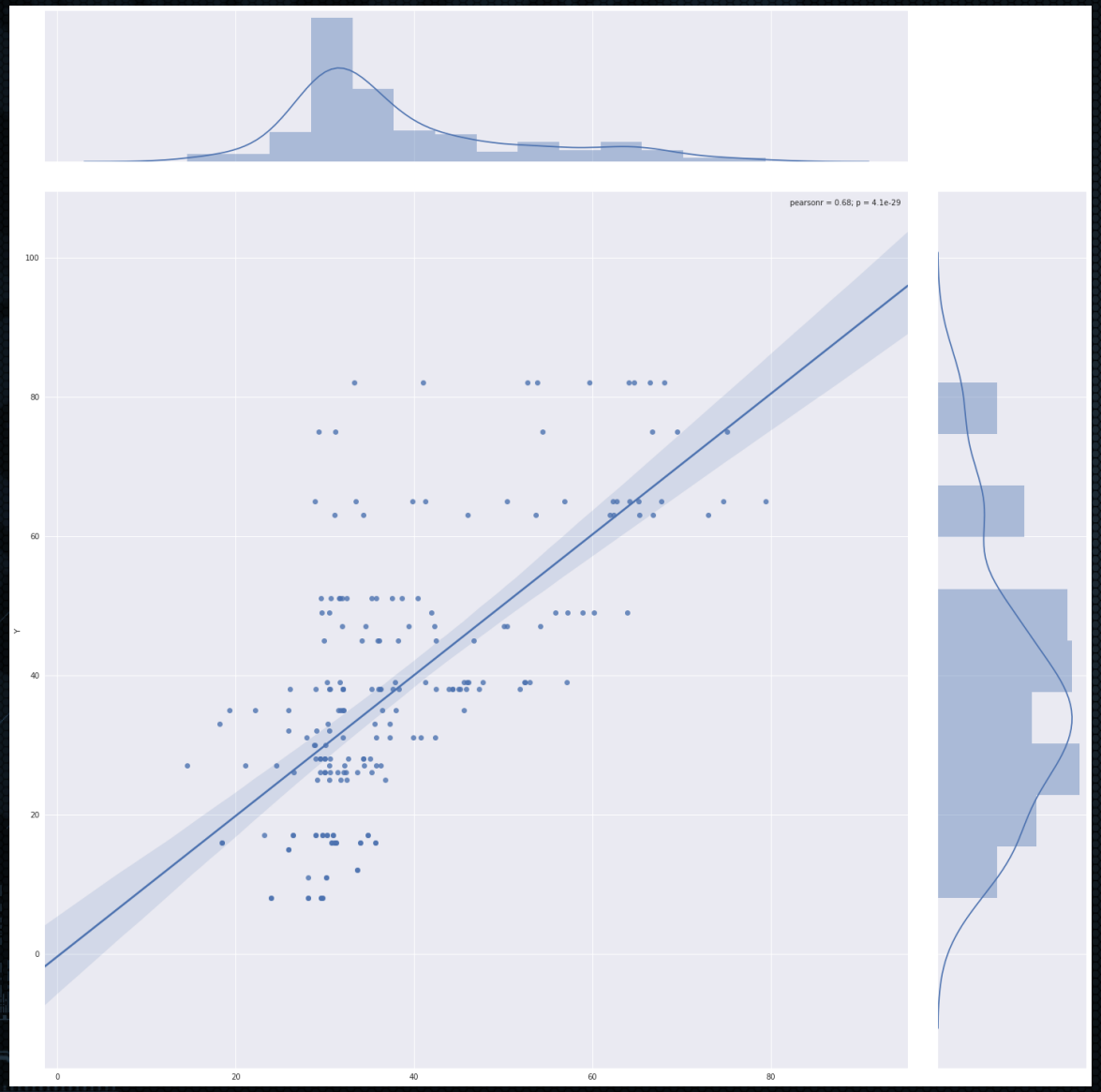






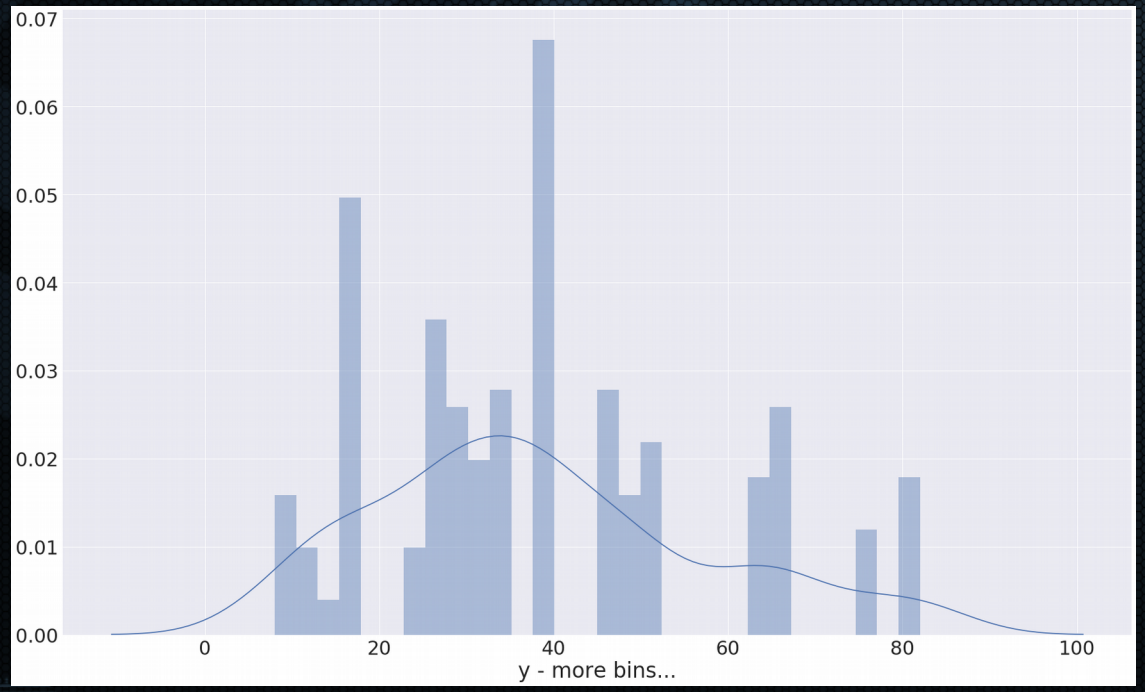
Shapiro says NO!

Model Regularization



- Lasso regularization
- Lambda 0.01
- The model is promising
- Set of features not captured

Next →



- Y not even bimodal…
- More features !
- Maybe a model not linear

Thank You!

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	56.5337	22.961	2.462	0.015	11.246	101.822
X1	-5.0685	13.506	-0.375	0.708	-31.707	21.570
X5	0.1373	0.044	3.124	0.002	0.051	0.224
X1:X5	-0.0981	0.069	-1.413	0.159	-0.235	0.039
X4	-0.0139	0.005	-2.759	0.006	-0.024	-0.004
X4:X5	1.098e-05	8.4e-06	1.308	0.193	-5.58e-06	2.75e-05
X2	-7.68e-05	5.46e-05	-1.406	0.161	-0.000	3.1e-05
X3	-0.3746	0.319	-1.173	0.242	-1.005	0.255
X2:X3	1.243e-06	7.14e-07	1.742	0.083	-1.65e-07	2.65e-06
X6	-2.493e-08	8.84e-08	-0.282	0.778	-1.99e-07	1.49e-07
X7	-1.455e-07	4.02e-08	-3.620	0.000	-2.25e-07	-6.62e-08
X8	6.082e-08	1.49e-08	4.094	0.000	3.15e-08	9.01e-08

Dep. Variable:	Y	R-squared:	0.472
Model:	OLS	Adj. R-squared:	0.441
Method:	Least Squares	F-statistic:	15.58
Date:	Thu, 26 Apr 2018	Prob (F-statistic):	1.62e-21
Time:	11:07:39	Log-Likelihood:	-823.99
No. Observations:	204	AIC:	1672.
Df Residuals:	192	BIC:	1712.
Df Model:	11		
Covariance Type:	nonrobust		