

B and Other Horror Movies

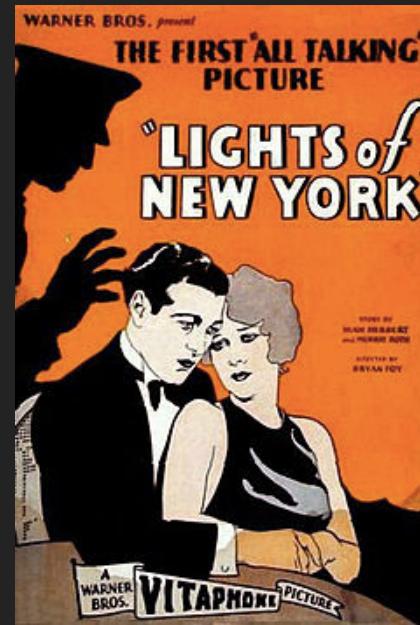
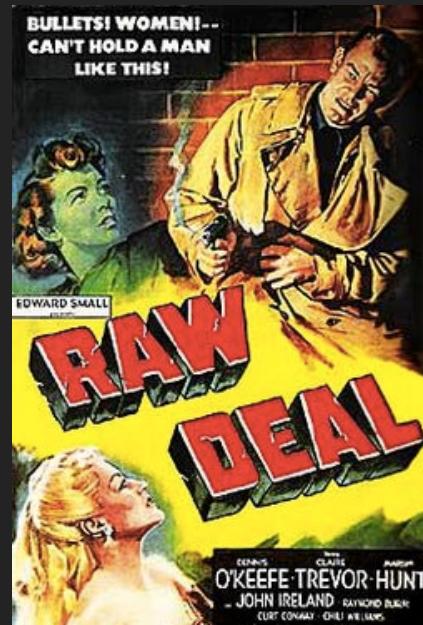
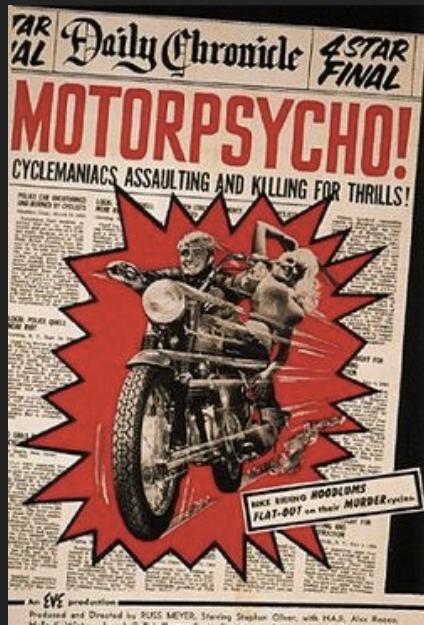
Factors That May Impact Profit Margins

Part I

Peter Rasmussen
April 22, 2016

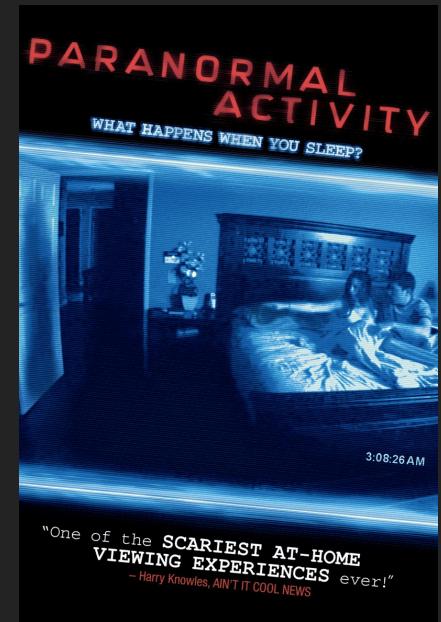
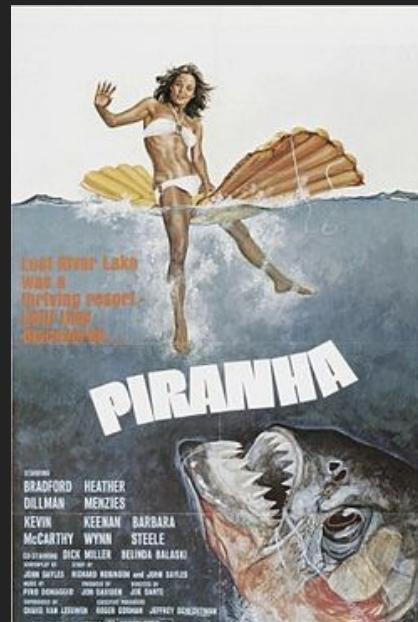
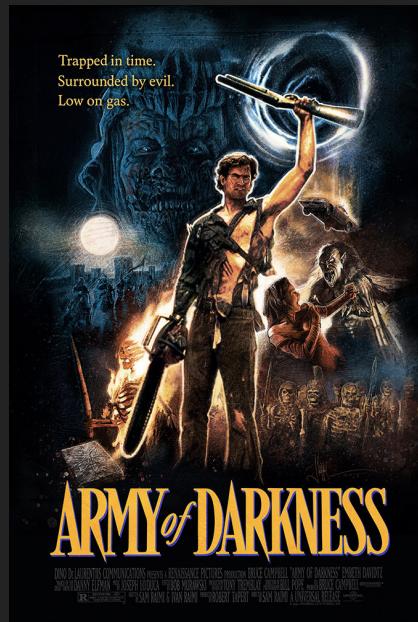
What is a B Movie?

- In the 1950s, movie theaters often put on double features, and the second one was typically a lower budget, lower quality film, a B movie
- Today, we can consider a B movie to be a low-budget, non-arthouse film



B Horror Movies

- Many B movies in the horror genre have been extremely successful
- Case in point: Paranormal Activity had \$0.5M budget but grossed \$214M
- This presentation looks at factors that affect profit margin for B & other horror flicks



Methodology & Workflow

Data acquisition

Sources

- The Numbers: budget, box office, and time data
- Box Office Mojo: user and meta scores

Web scraping tools

- Selenium for The Numbers
- BeautifulSoup for Box Office Mojo

Data set:

- 165 samples, 1970-2015, American flix

Analysis

Wrangle data

- Clean data, adjust for CPI, derive dependent variable
- Only 165 observations, budget data limiting factor

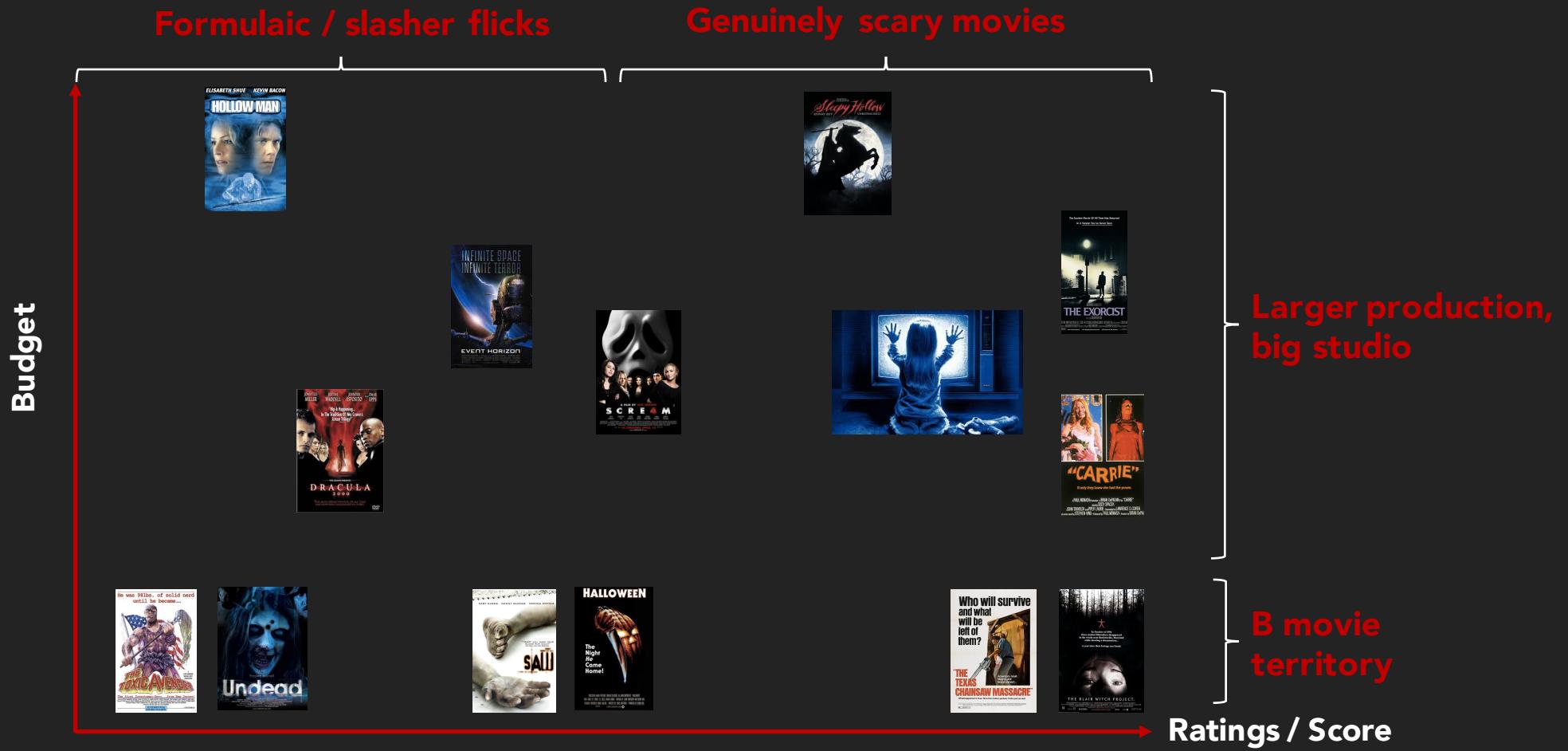
Cull features

- Employ Lasso LARS method

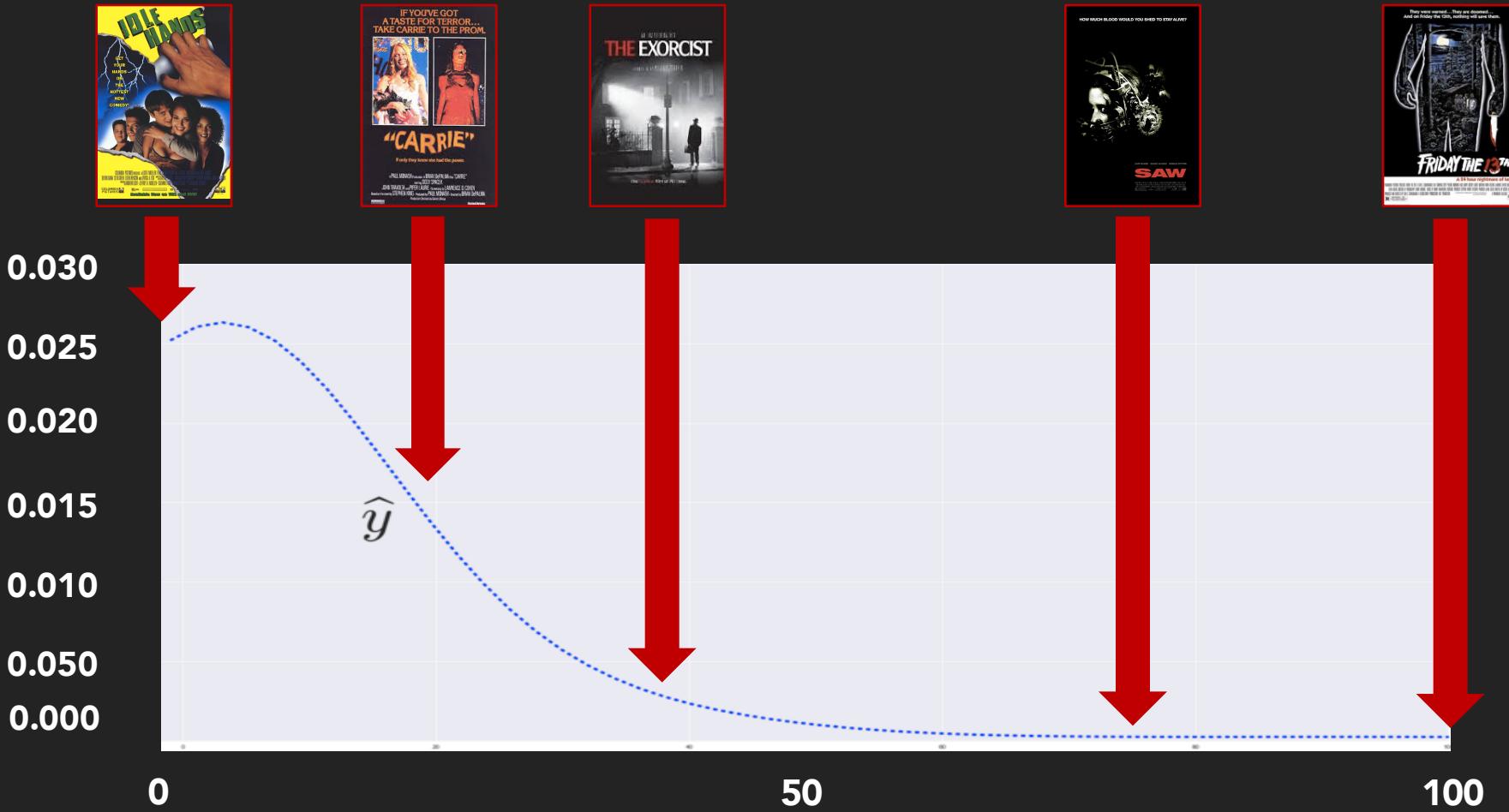
Perform regression

- OLS on features retained features

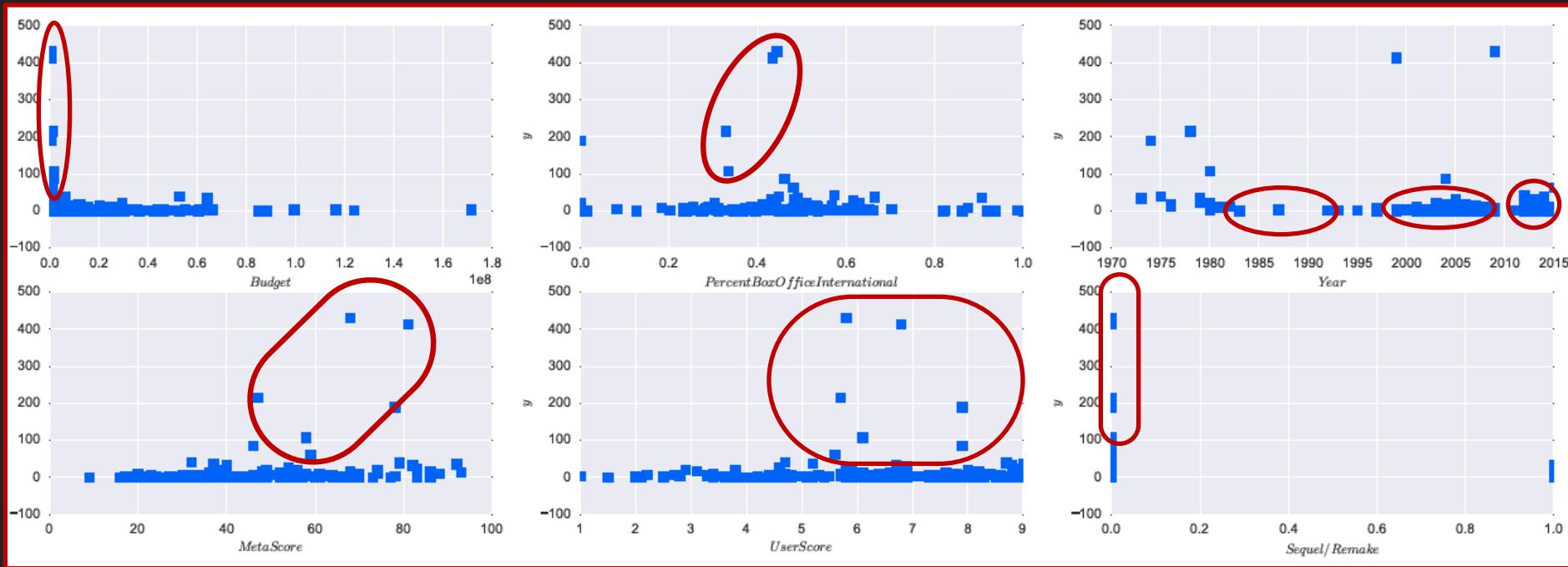
Horror Movies: Qualitative FW



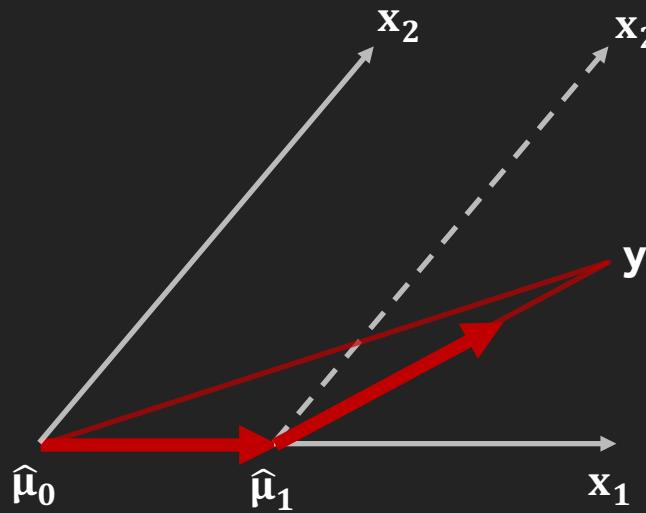
Analysis: PMF of Profit Margins



Analysis: Features v Margins



Analysis: Lasso LARS



- Chose Lasso LARS over Lasso b/c former included more features
- Selected features
 - Budget
 - Year
 - Meta score
- Excluded features
 - User score
 - Percent revenue that's international
 - Quarter of release
 - Sequel / remake

Figure adapted from <https://www.quora.com/What-is-Least-Angle-Regression-and-when-should-it-be-used>

Analysis: Regression Statistics

Dep. Variable:	margin	R-squared:	0.153
Model:	OLS	Adj. R-squared:	0.137
Method:	Least Squares	F-statistic:	9.735
Date:	Thu, 21 Apr 2016	Prob (F-statistic):	6.08e-06
Time:	23:19:31	Log-Likelihood:	-877.66
No. Observations:	165	AIC:	1761.
Df Residuals:	162	BIC:	1771.
Df Model:	3		
Covariance Type:	nonrobust		

Model's predictive power is low

	coef	std err	t	P> t	[95.0% Conf. Int.]
budget_i	-3.627e-07	1.54e-07	-2.359	0.020	-6.66e-07 -5.91e-08
year	-0.0046	0.006	-0.812	0.418	-0.016 0.007
metascore	0.6844	0.214	3.196	0.002	0.262 1.107

Analysis: Regression Statistics

Dep. Variable:	margin	R-squared:	0.153
Model:	OLS	Adj. R-squared:	0.137
Method:	Least Squares	F-statistic:	9.735
Date:	Thu, 21 Apr 2016	Prob (F-statistic):	6.08e-06
Time:	23:19:31	Log-Likelihood:	-877.66
No. Observations:	165	AIC:	1761.
Df Residuals:	162	BIC:	1771.
Df Model:	3		
Covariance Type:	nonrobust		

Highly unlikely that the model outputs could be reproduced if null hypothesis were true

	coef	std err	t	P> t	[95.0% Conf. Int.]
budget_i	-3.627e-07	1.54e-07	-2.359	0.020	-6.66e-07 -5.91e-08
year	-0.0046	0.006	-0.812	0.418	-0.016 0.007
metascore	0.6844	0.214	3.196	0.002	0.262 1.107

Analysis: Regression Statistics

Dep. Variable:	margin	R-squared:	0.153
Model:	OLS	Adj. R-squared:	0.137
Method:	Least Squares	F-statistic:	9.735
Date:	Thu, 21 Apr 2016	Prob (F-statistic):	6.08e-06
Time:	23:19:31	Log-Likelihood:	-877.66
No. Observations:	165	AIC:	1761.
Df Residuals:	162	BIC:	1771.
Df Model:	3		
Covariance Type:	nonrobust		

Profit margin:

- Decreases w/ budget and year
- Increases w/ meta score

	coef	std err	t	P> t	[95.0% Conf. Int.]
budget_i	-3.627e-07	1.54e-07	-2.359	0.020	-6.66e-07 -5.91e-08
year	-0.0046	0.006	-0.812	0.418	-0.016 0.007
metascore	0.6844	0.214	3.196	0.002	0.262 1.107

Analysis: Regression Statistics

Dep. Variable:	margin	R-squared:	0.153
Model:	OLS	Adj. R-squared:	0.137
Method:	Least Squares	F-statistic:	9.735
Date:	Thu, 21 Apr 2016	Prob (F-statistic):	6.08e-06
Time:	23:19:31	Log-Likelihood:	-877.66
No. Observations:	165	AIC:	1761.
Df Residuals:	162	BIC:	1771.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[95.0% Conf. Int.]
budget_i	-3.627e-07	1.54e-07	-2.359	0.020	-6.66e-07 -5.91e-08
year	-0.0046	0.006	-0.812	0.418	-0.016 0.007
metascore	0.6844	0.214	3.196	0.002	0.262 1.107

95% confidence that:

- Budget has a negative impact
- Meta score has a positive impact

Conclusions and Next Steps

- Further analysis is needed before we can conclude more about B movies in particular
- Even so, below are features of statistical significance:
 - There is a slightly negative correlation between budget and profit margin among all surveyed horror movies, owing perhaps to the wild success of B horror movies
 - There is a positive correlation between meta score and profit margin, which makes sense intuitively
- Don't leave out the outliers because that's what I'm most interested in here!!
- Things I didn't expect / was surprised by:
 - Lasso LARS method did not find significant correlation between users' scores and profit margin
 - The lack of budget data on The Numbers site really limited my sample size
- Going forward:
 - Bigger sample size would make analysis more robust
 - Normalize variables like year (thanks Hyatt!)
 - This would give us a more complete picture of the profit margins of horror movies in general and B horror movies in particular
 - Include other features such as ratings, directors, actors, etc.

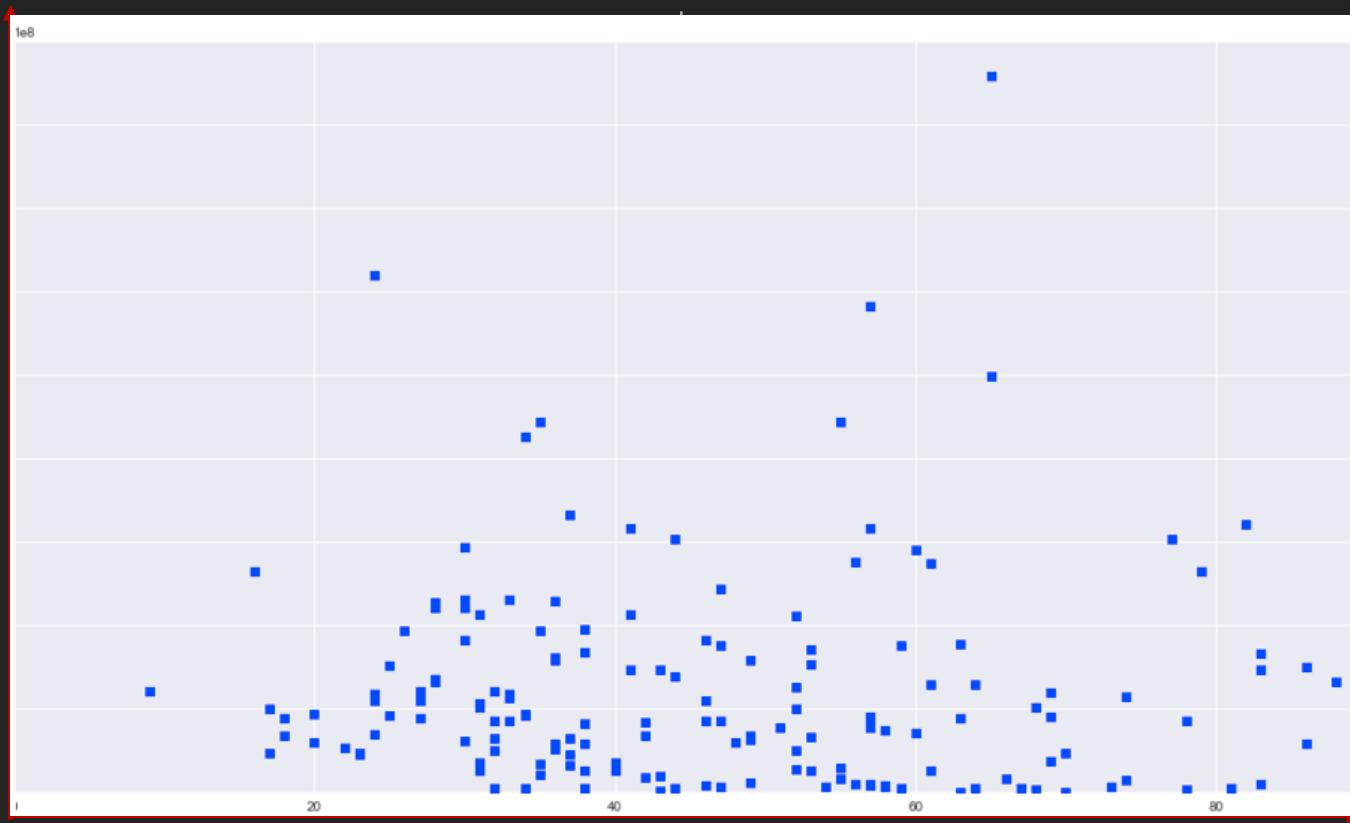
Appendix

Regression Statistics

Omnibus:	228.009	Durbin-Watson:	1.971
Prob(Omnibus):	0.000	Jarque-Bera (JB):	12346.006
Skew:	5.820	Prob(JB):	0.00
Kurtosis:	43.747	Cond. No.	1.60e+10

Horror Movies: Scatter Plot

Budget



Ratings / Score