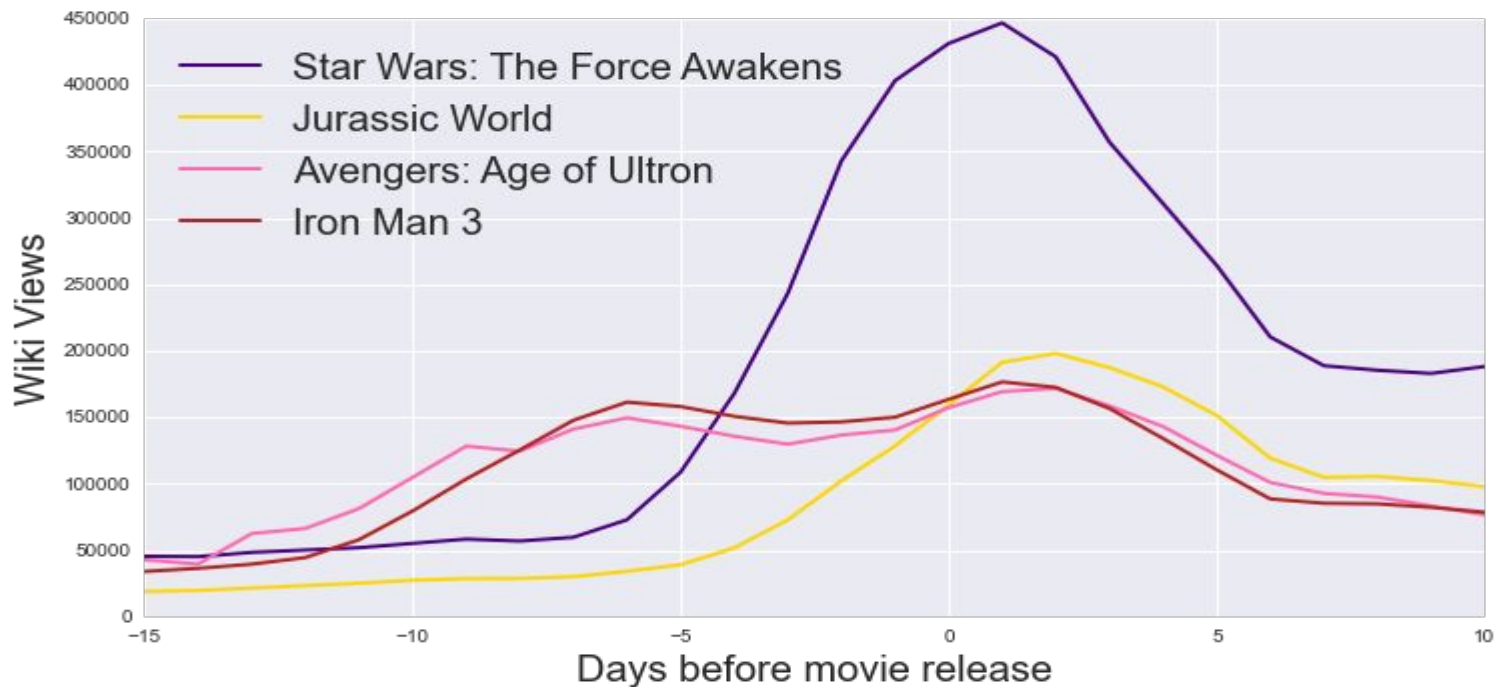


Predicting Opening Box Office Gross With Internet Activity

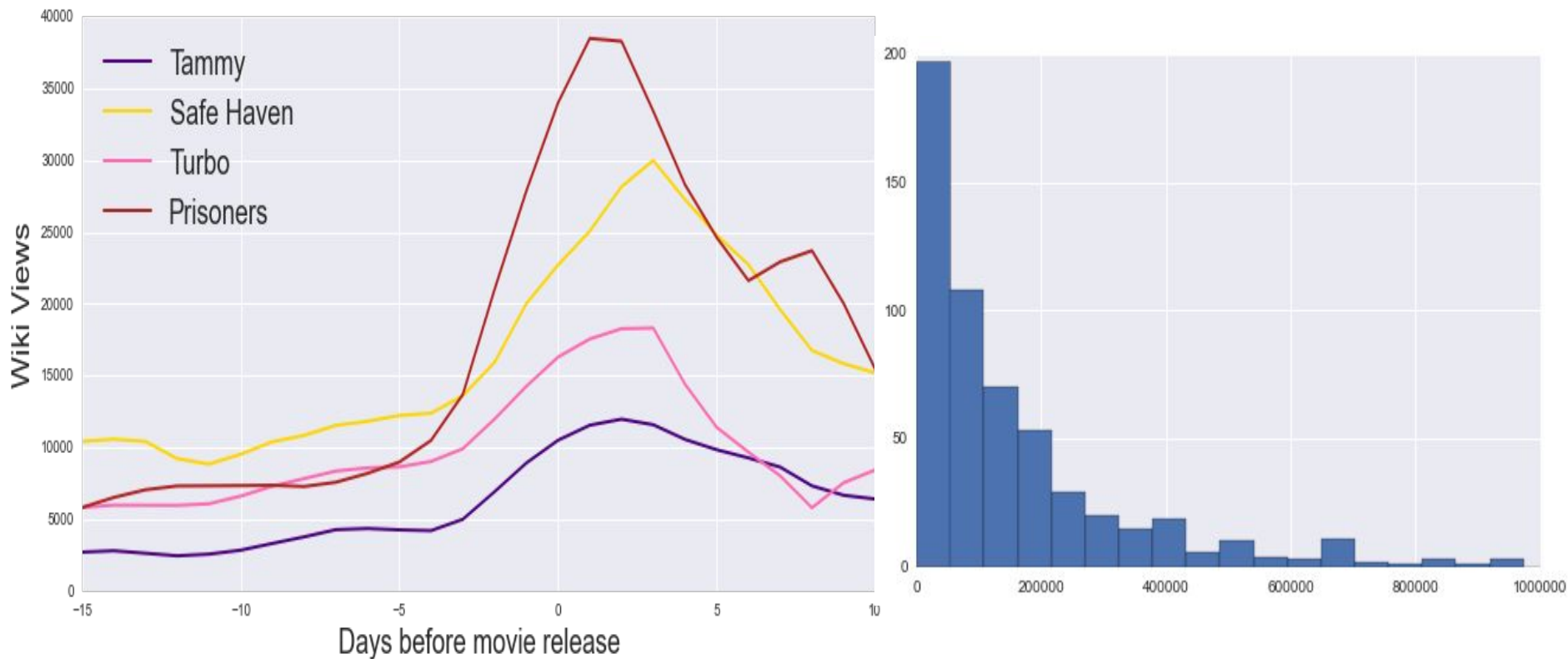
Devin Wieker

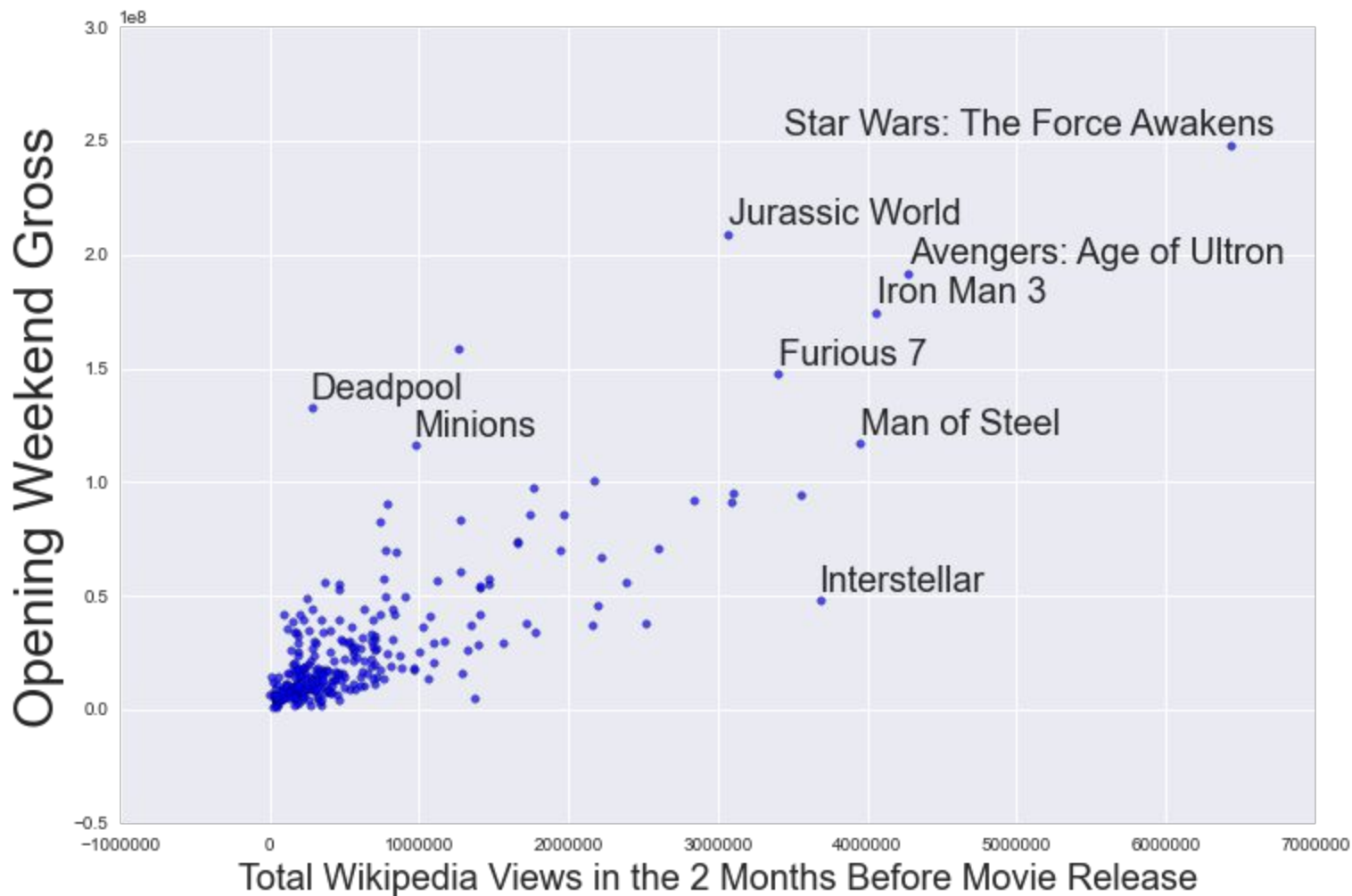
Wikipedia Page Visits

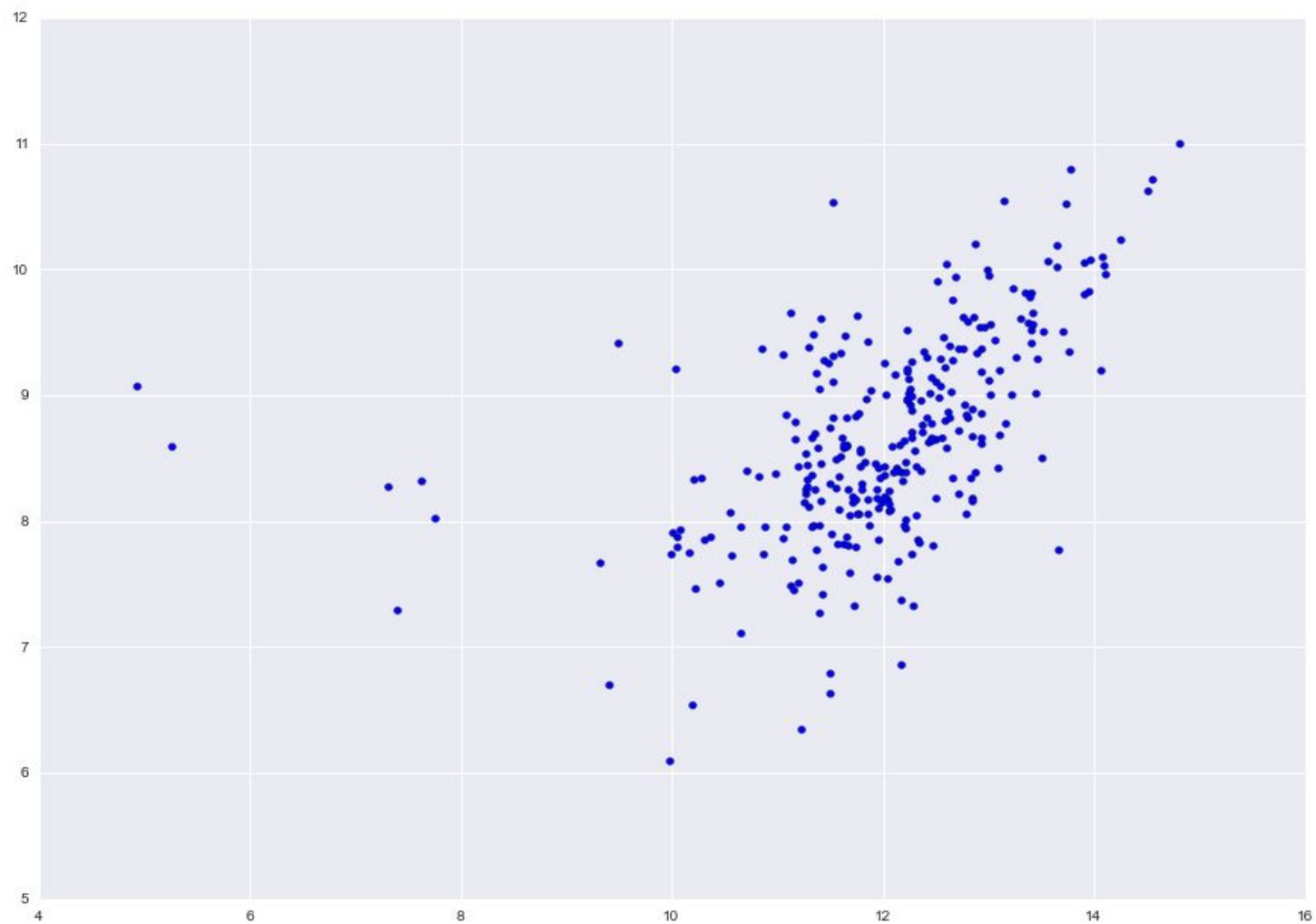
- Scrape list of movies from BoxOfficeMojo (2013 - present)
- Scrape Wikipedia page visits in months leading up to movie (daily granularity)



Wikipedia Page Visits





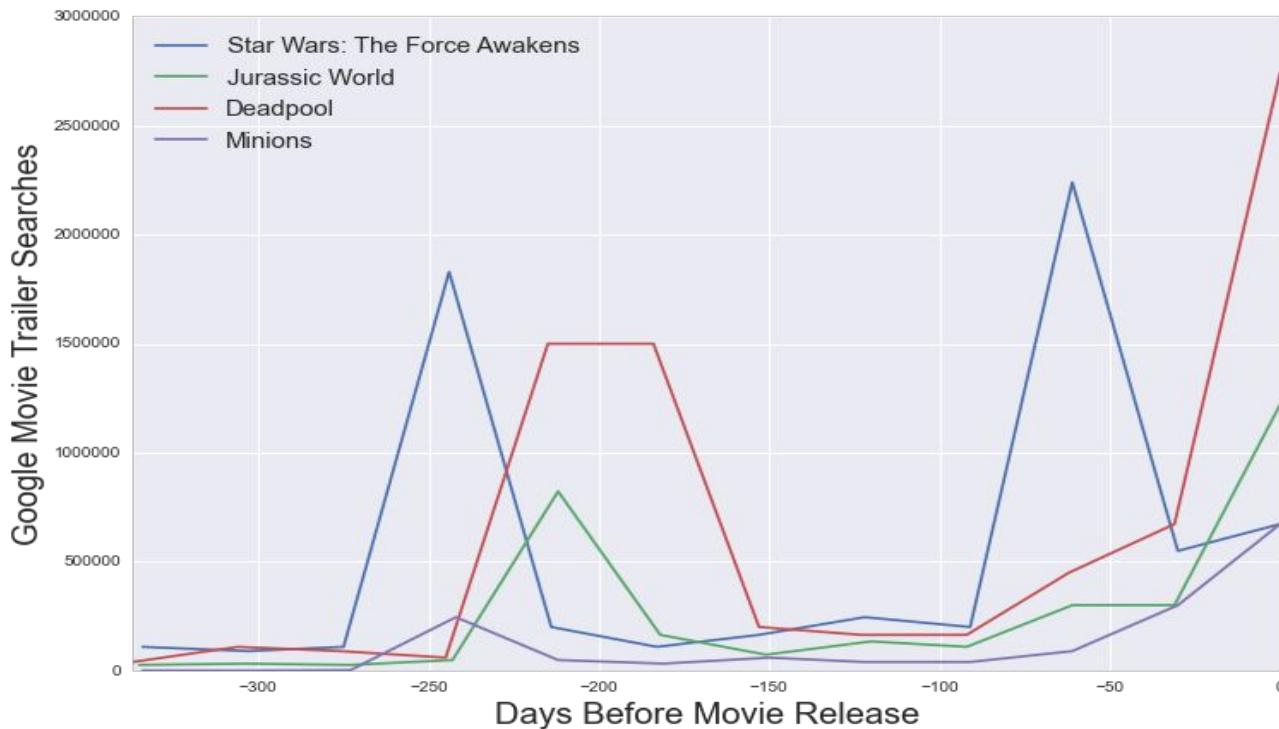




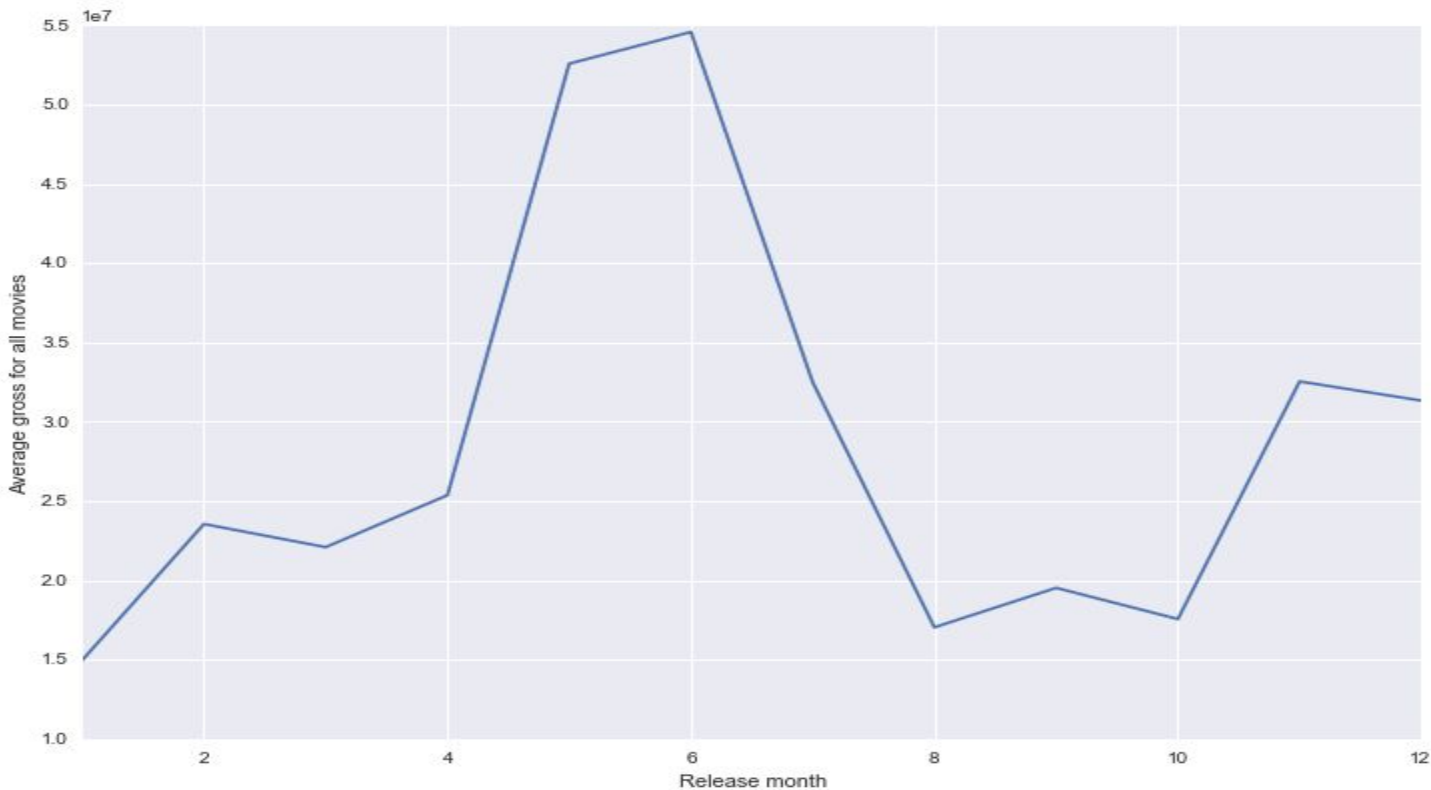
Dep. Variable:	log_opening_gross	R-squared:	0.539
Model:	OLS	Adj. R-squared:	0.537
Method:	Least Squares	F-statistic:	325.8
Date:	Thu, 28 Apr 2016	Prob (F-statistic):	8.62e-49
Time:	15:59:07	Log-Likelihood:	-290.61
No. Observations:	281	AIC:	585.2
Df Residuals:	279	BIC:	592.5
Df Model:	1		
Covariance Type:	nonrobust		

Improving the model: Google Search Traffic

- Google Keywords give monthly search traffic by keyword



Improving the Model: Seasonality



Dep. Variable:	log_opening_gross	R-squared:	0.683
Model:	OLS	Adj. R-squared:	0.675
Method:	Least Squares	F-statistic:	82.85
Date:	Thu, 28 Apr 2016	Prob (F-statistic):	6.37e-20
Time:	16:08:34	Log-Likelihood:	-78.667
No. Observations:	80	AIC:	163.3
Df Residuals:	77	BIC:	170.5
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t 	[95.0% Conf. Int.]
Intercept	5.6642	0.861	6.582	0.000	3.951 7.378
log_goog_views	0.6209	0.097	6.374	0.000	0.427 0.815
log_wiki_views	0.2461	0.100	2.464	0.016	0.047 0.445

Limitations

- Mistaken wikipedia visits (Interstellar?)
- Google search data requires EXACT keywords
 - (star war vs star wars: the force awakens)
- Google search data only has monthly granularity
- Need more data. Only ~80 movies in final regression
- Wikipedia only offers short term prediction (< ~1 week)

Conclusion

- Proof that internet activity can help predict (at least) simple outcomes
- Example: predicting the short term stock market?
- Studios can track google trailer searches far in advance (months) to gauge turnout of their movie or other competing movies.
- Wikipedia can be a useful resource for studying large scale trends of people (it's all open source and free-to-edit)