WHERE TO ADD RELEASES

- How to determine where to add sequel movie releases

- Chose between areas for new or limited releases based on rating and up-voting

DATA

- IMDB Sequel Keyword Search - web scraped

- Billboard Soundtrack Chart - web scraped

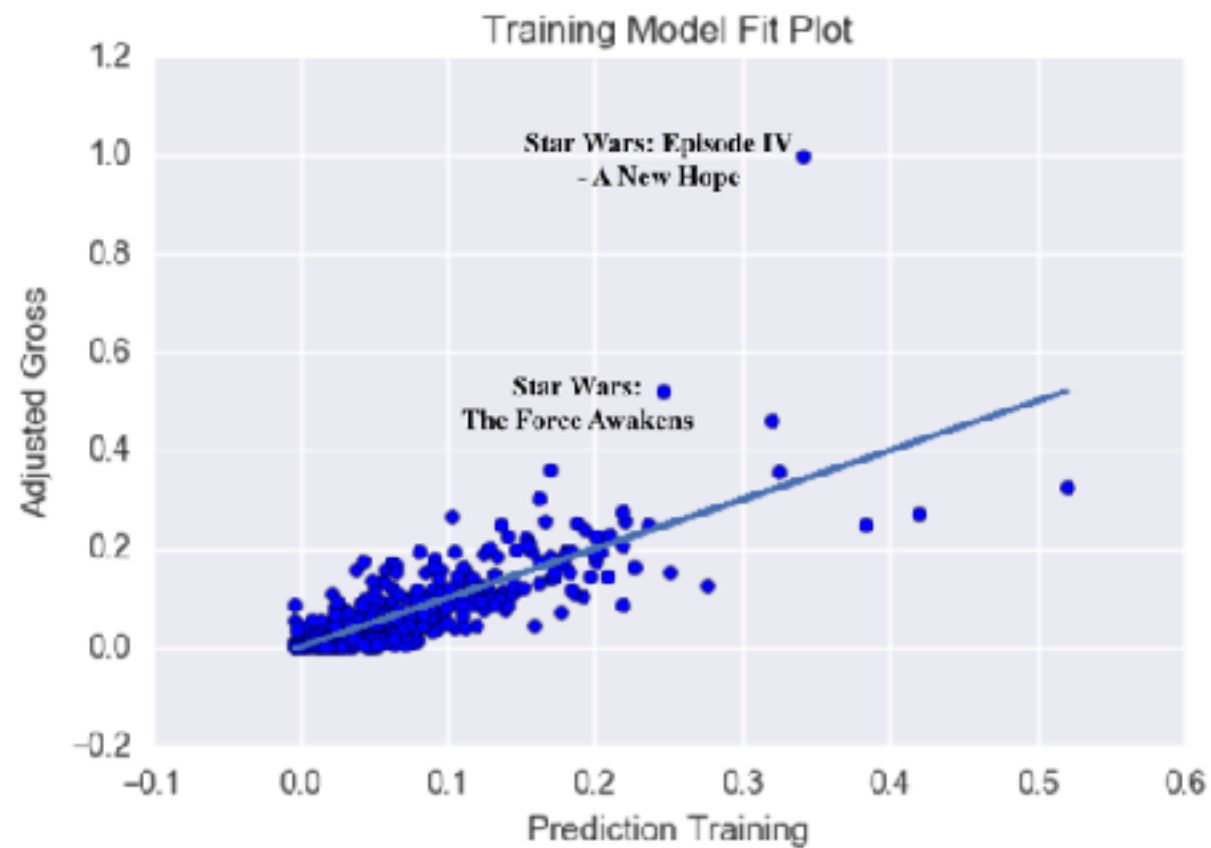- Inflation Adjustment, Consumer Price Index

THE LORD OF THE RINGS
THE RETURN OF THE KING

THE JOURNEY ENDS DECEMBER 17TH

MODEL

Dep. Variable:	adj_gross	R-squared:	0.762
Model:	OLS	Adj. R-squared:	0.758
Method:	Least Squares	F-statistic:	198.3
Date:	Thu, 26 Jan 2017	Prob (F-statistic):	6.91e-131
Time:	23:19:40	Log-Likelihood:	657.67
No. Observations:	441	AIC:	-1301.
Df Residuals:	434	BIC:	-1273.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
votes	0.5154	0.027	18.854	0.000	0.462 0.569
runtime	0.0100	0.029	0.344	0.731	-0.047 0.067
adventure	0.0205	0.006	3.271	0.001	0.008 0.033
animation	0.0275	0.011	2.610	0.009	0.007 0.048
fantasy	0.0183	0.008	2.369	0.018	0.003 0.034
drama	-0.0192	0.007	-2.667	0.008	-0.033 -0.005
user_score	0.0158	0.023	0.692	0.490	-0.029 0.061

- Votes most predictive

- User Score not significant

- OLS feature elimination by p-value

MODEL: TEST VS TRAIN
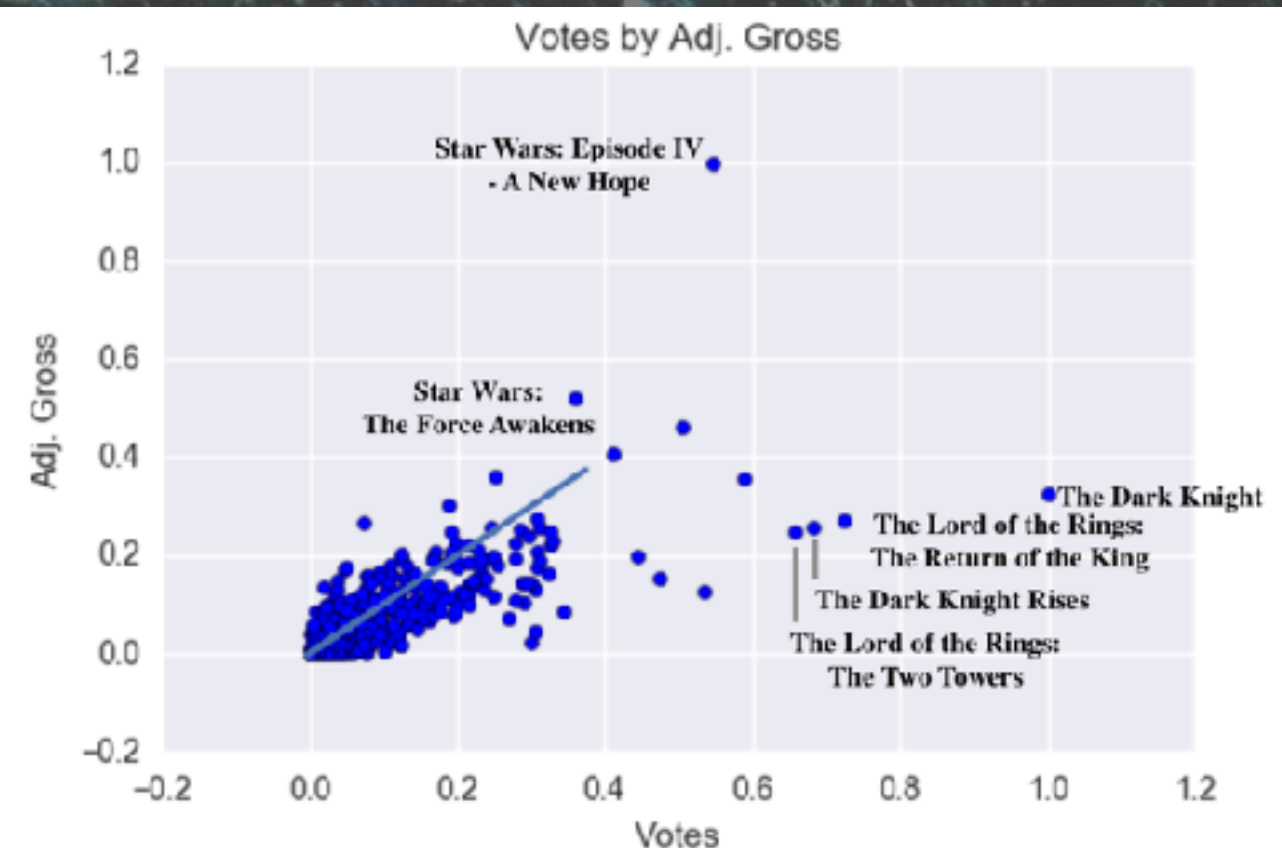
- Model: OLS
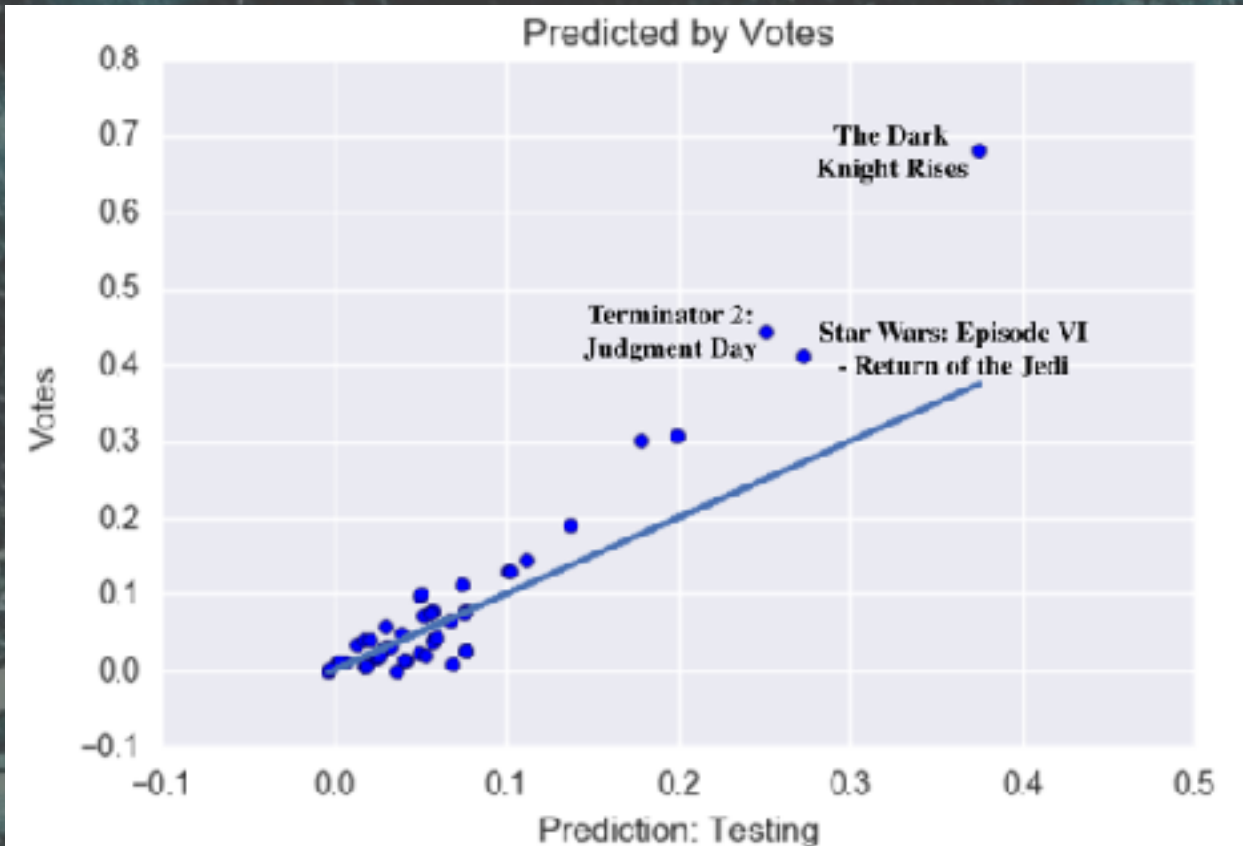
- Adj. R-squared: 0.762

- Up-Votes

- User Scores

- Runtime

- Genres

MODEL BY VOTES

- Up votes help predict

- Skewed by select few series
  - Star Wars
  - Dark Knight
  - Lord of the Rings

Predicted by Votes

Votes by Adj. Gross

REMOVE OUTLIERS

```
==============================================================================
Dep. Variable:              adj_gross   R-squared:                       0.817
Model:                            OLS   Adj. R-squared:                  0.815
Method:                 Least Squares   F-statistic:                     316.0
Date:                Fri, 27 Jan 2017   Prob (F-statistic):          5.07e-153
Time:                        12:23:37   Log-Likelihood:                 787.95
No. Observations:                 430   AIC:                            -1564.
Df Residuals:                     424   BIC:                            -1540.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
votes          0.5502      0.029     19.271      0.000       0.494      0.606
animation      0.0367      0.008      4.875      0.000       0.022      0.051
runtime        0.0903      0.021      4.351      0.000       0.050      0.131
adventure      0.0161      0.005      3.537      0.000       0.007      0.025
user_score    -0.0521      0.017     -3.135      0.002      -0.085     -0.019
drama         -0.0058      0.005     -1.109      0.268      -0.016      0.005
==============================================================================
Omnibus:                       72.368   Durbin-Watson:                   1.975
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              219.306
Skew:                           0.772   Prob(JB):                     2.39e-48
Kurtosis:                       6.140   Cond. No.                         16.0
==============================================================================
```
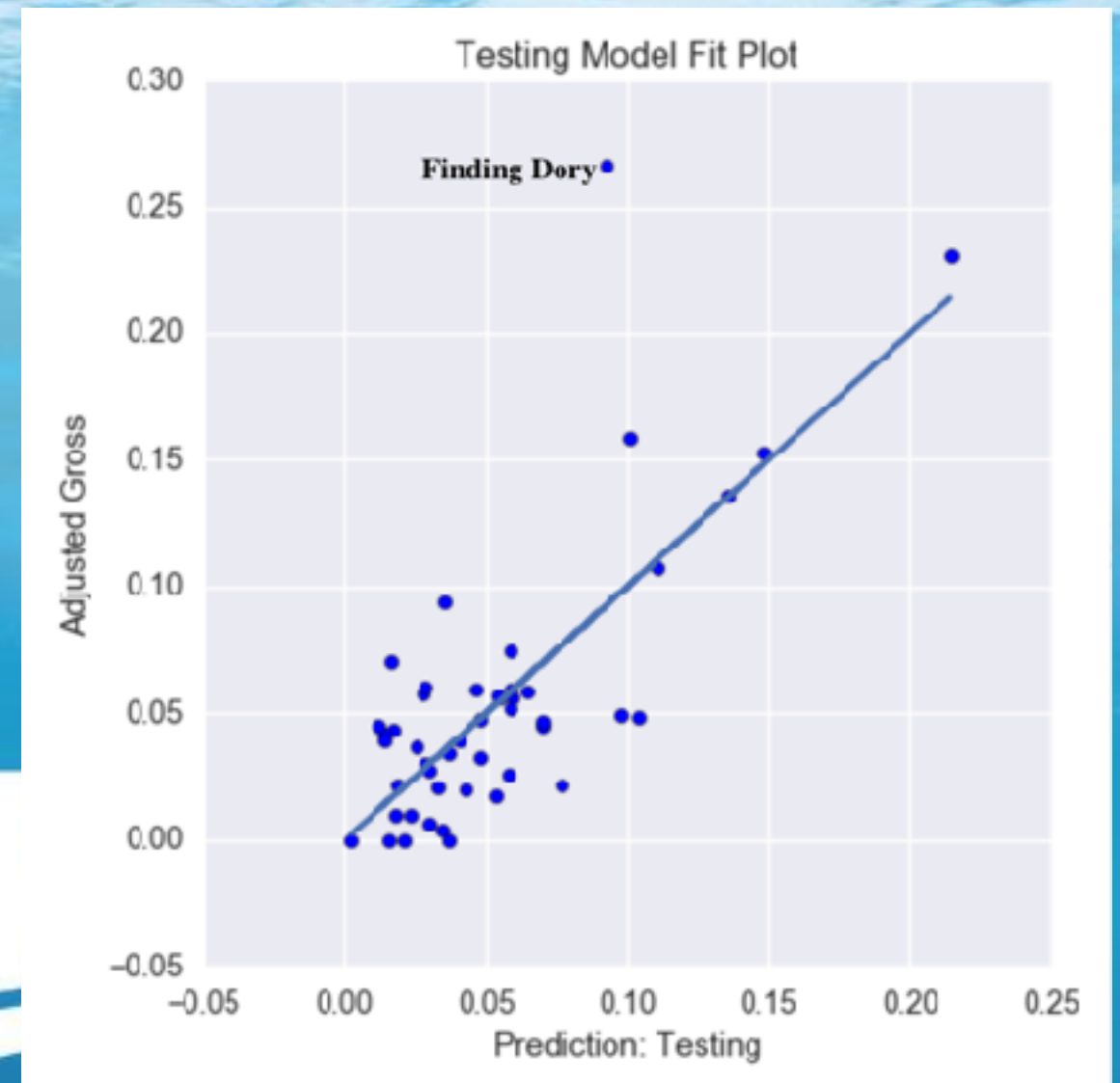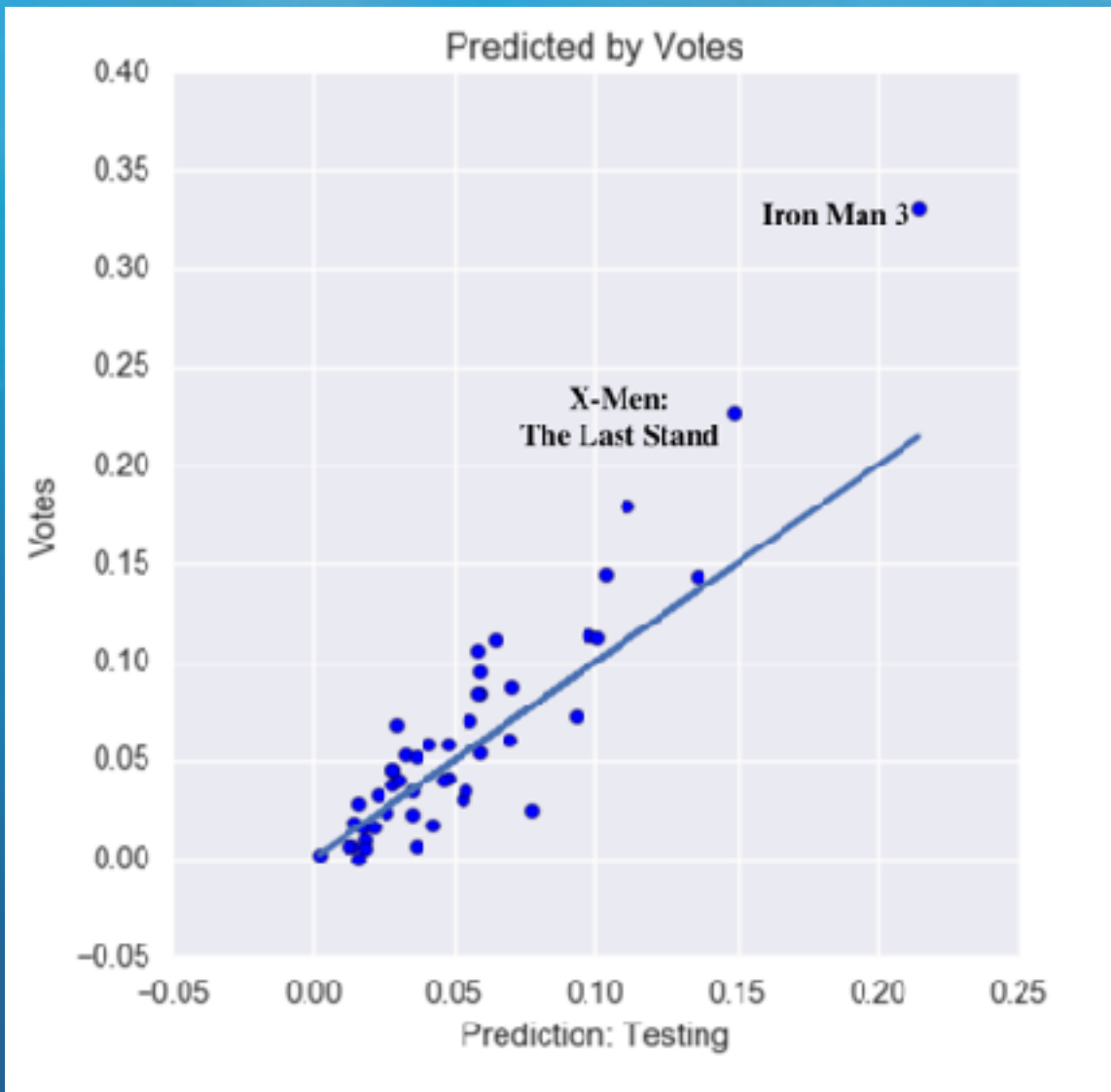
- R-squared goes up: 0.758 to 0.815

- Votes have the highest coefficient

- User score coefficient is small and negative, but significant

Testing Model Fit Plot

Predicted by Votes

REMOVE OUTLIERS

FINDINGS AND FUTURE QUESTIONS

- Up votes more important than user scores to determine where to expand releases

- Quality of first in series vs. other movies

- Will people see the third if the second is bad?

NOTES AND LIMITATIONS

- More outliers in movies in series

- Reduced features to those found in over 5 movies

- Looked at Billboard soundtrack scores, but was not predictive