

Vince Velocci

# Modeling Film DVD Sales: Which Features Matter?

# Data Source

- Movie data was scraped from [www.the-numbers.com](http://www.the-numbers.com)

# Data Source

- Movie data was scraped from [www.the-numbers.com](http://www.the-numbers.com)
- Covering 6 different genres since 1998: Action, Adventure, Horror, Comedy, Thriller/Suspense, Drama

# Data Source

- Movie data was scraped from [www.the-numbers.com](http://www.the-numbers.com)
- Covering 6 different genres since 1998: Action, Adventure, Horror, Comedy, Thriller/Suspense, Drama
- What was included?
  - \$\$\$\$\$ (DVD Sales, Opening Weekend, Domestic Gross, International Box Office, Inflation-Adjusted Gross, Production Budget)
  - Release Date, Running Time, Genre, MPAA Ratings
- Models: OLS, Ridge, Lasso, Elastic Net (all performed equally)
- 70/30 train/test split

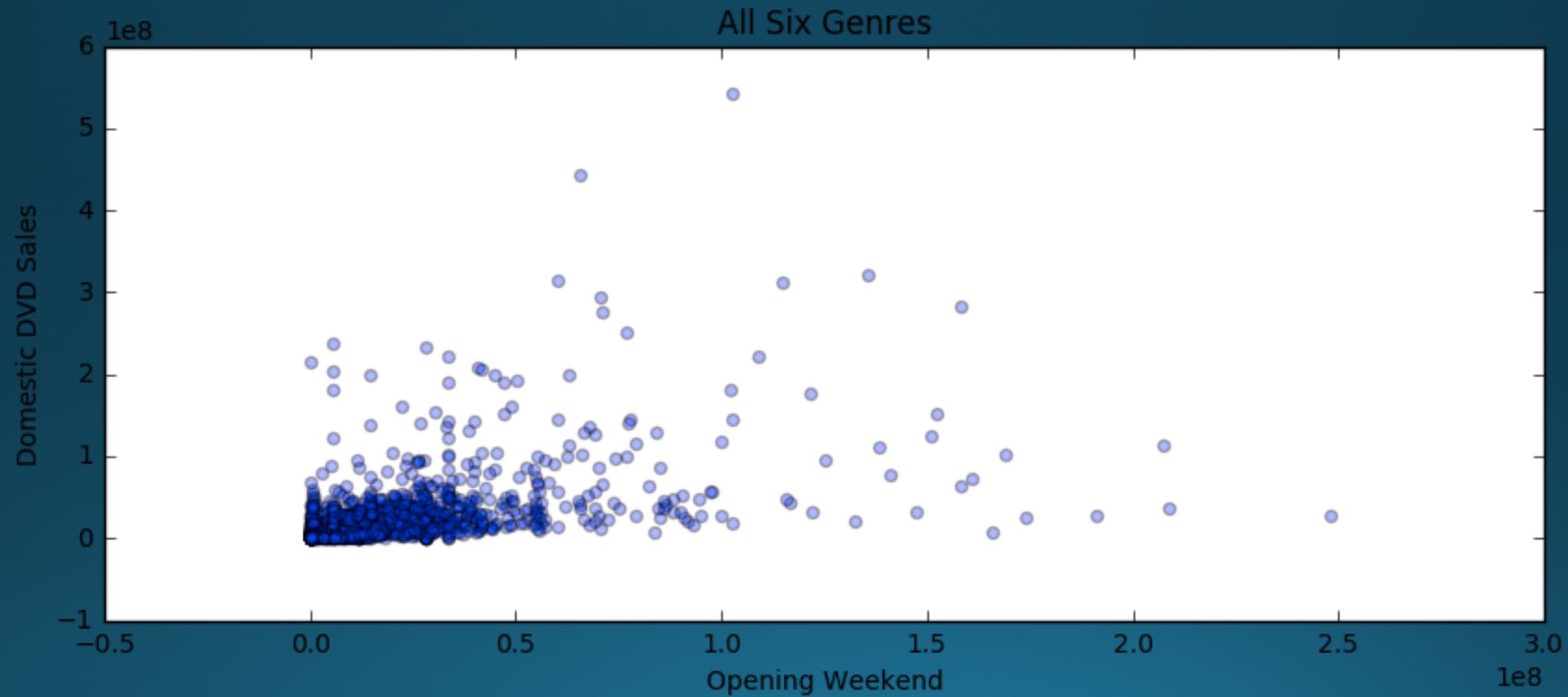
# Linear Regression for Two Genres

- Comedy: R-squared of 0.7
  - G Ratings are best (i.e. positive effect on DVD Sales)
  - Summer or Fall release is best
  - Oddly, Domestic Gross seemed to have a negative effect (not strong – random chance?)
- Horror: R-squared of 0.6
  - PG rating is best (strongest feature)
  - Running Time was strongest with this genre over all others
  - Fall or Winter release is best
  - Opening Weekend, IBO, Budget had virtually no importance

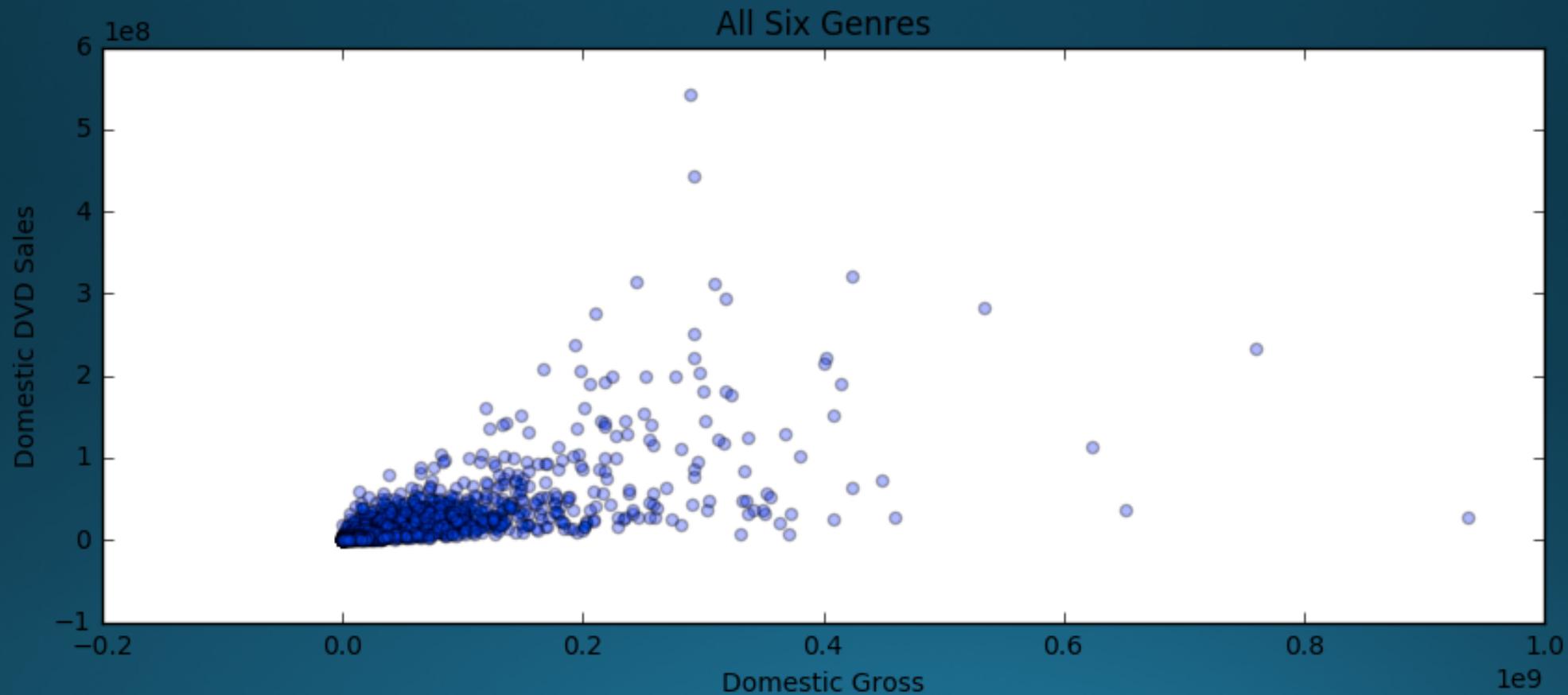
# Combine all Genres into One

- R squared of 0.64
- Adventure genre = strongest feature in the model (positive effect)
- Winter was 2<sup>nd</sup> strongest (neg effect on DVD sales – worst season)
- Summer and Fall are best
- 2<sup>nd</sup> best genre: Action
- Opening Weekend, Budget, IBO....not as important (~ 0 effect)

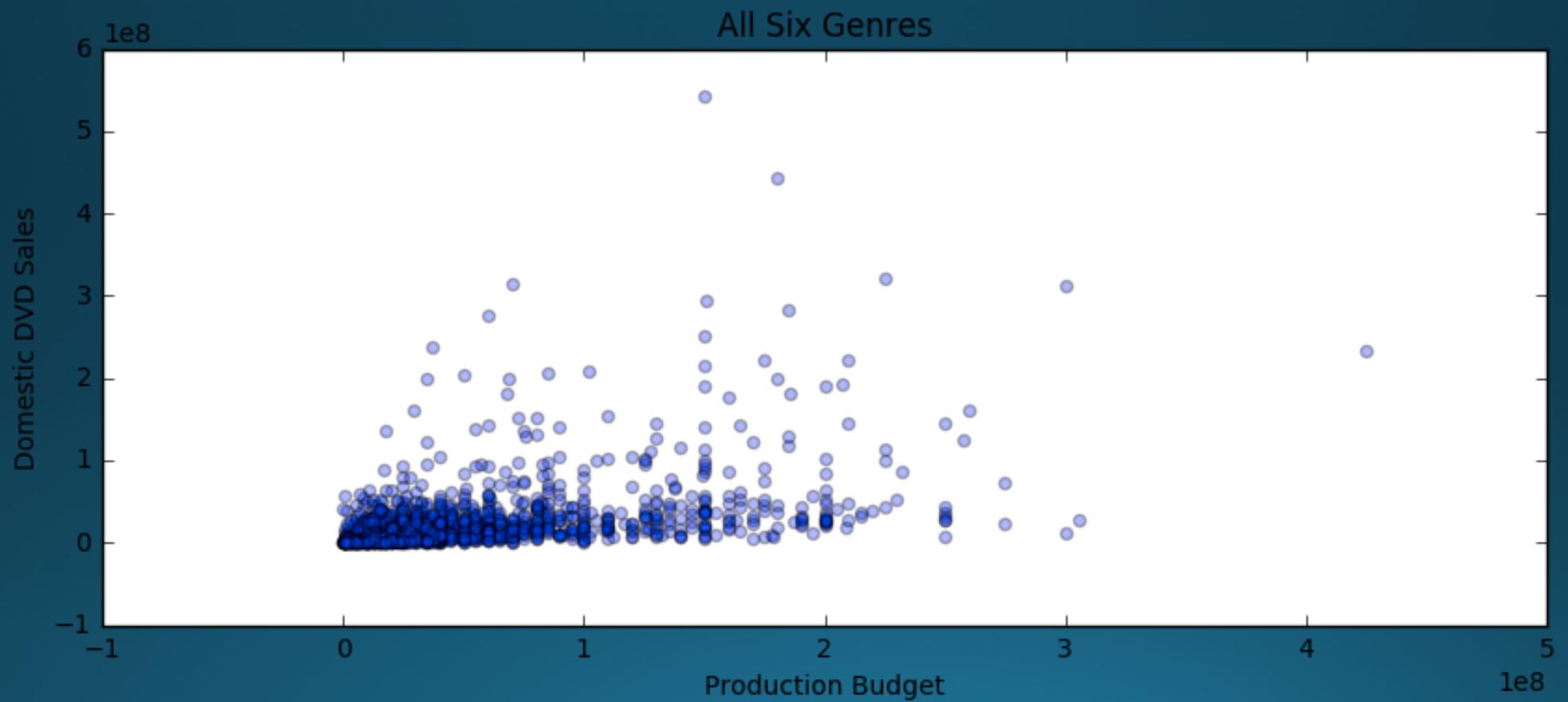
# Opening vs DVD Sales for All Genres



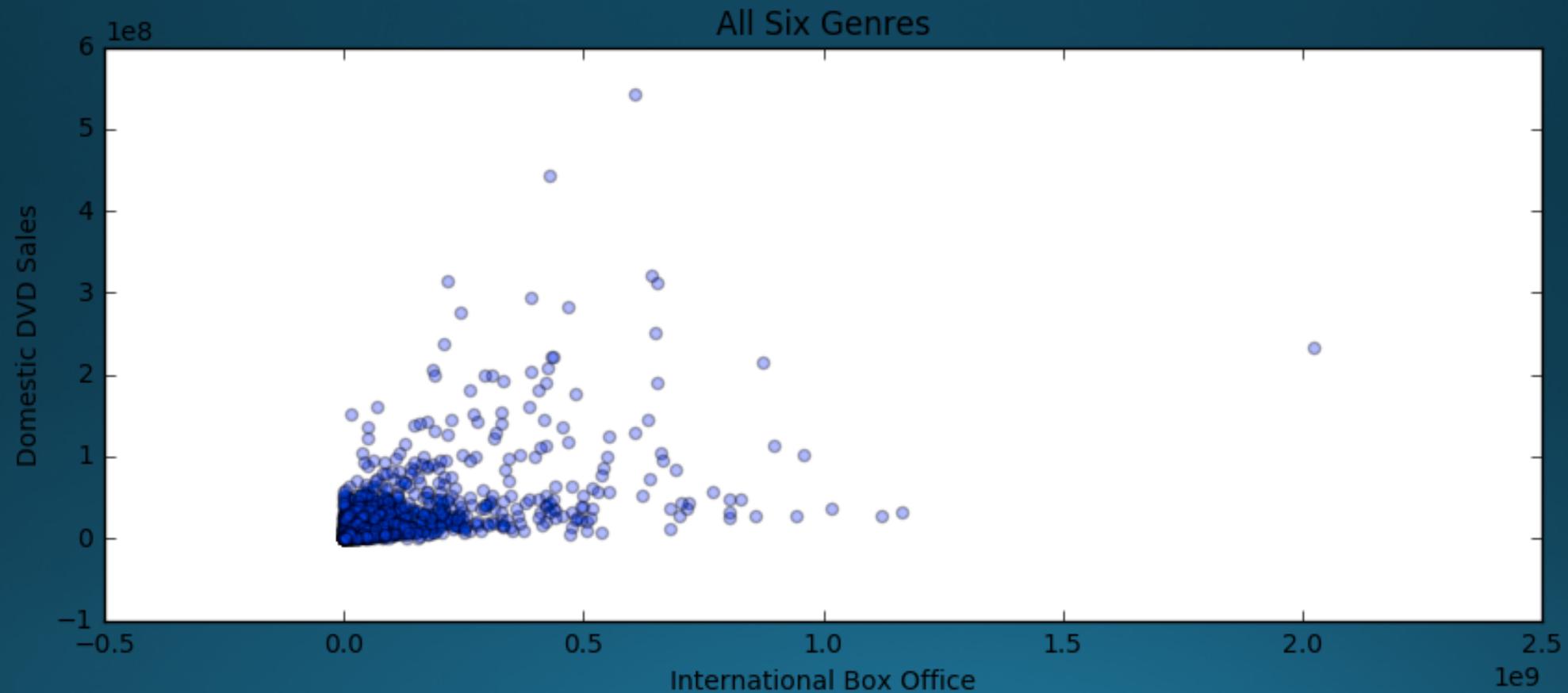
# Gross vs DVD Sales for All Genres



# Budget vs DVD Sales for all Genres



# IBO vs DVD Sales for all Genres



# Further Work

- Get more data!
- Get a different variety of data from different sources!
  - Eg. Google searches for newly released movie trailers,
  - Perhaps sentiment analysis of major movie critics?
  - Data from streaming services
- Try different models!