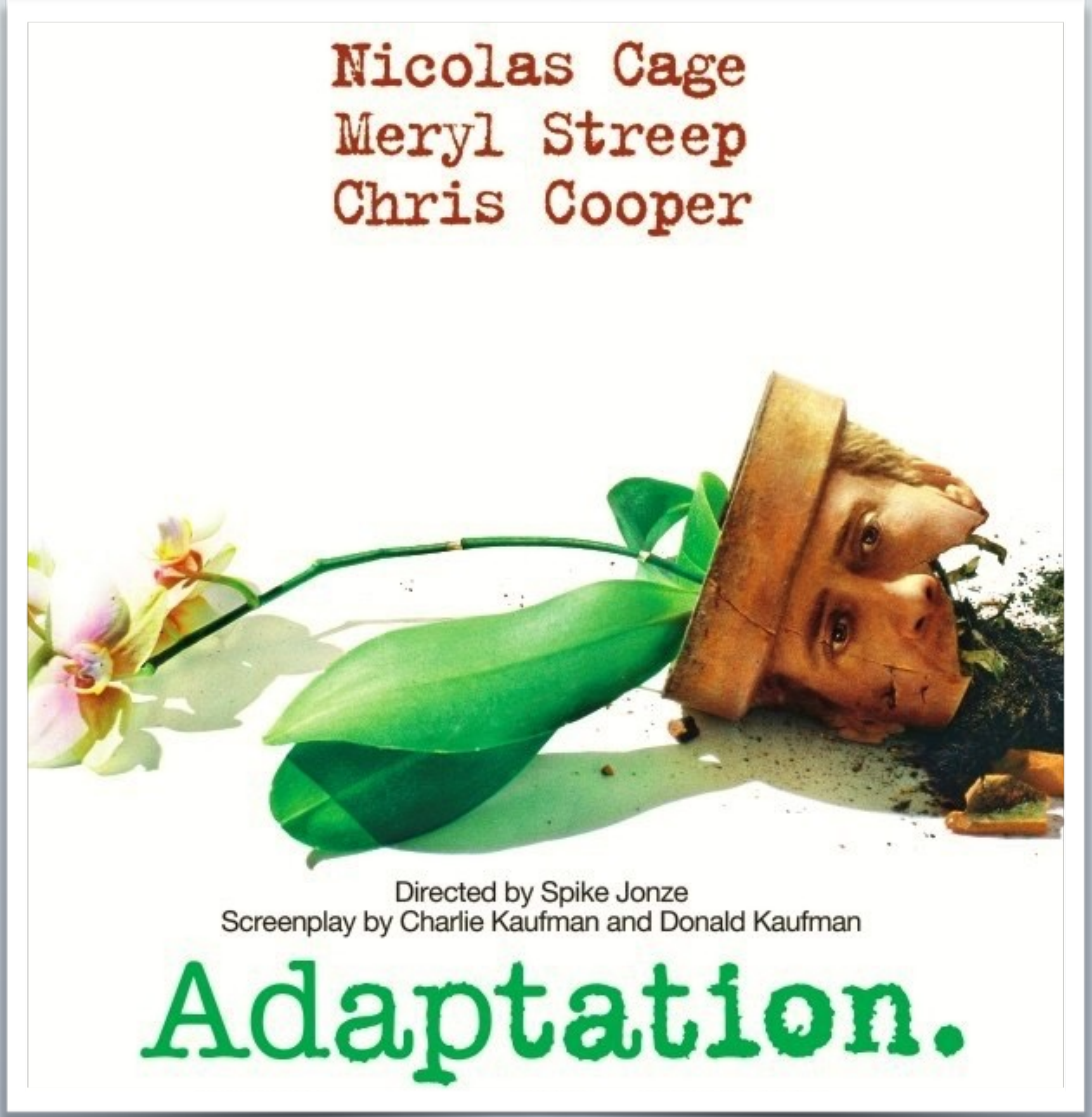
Low-Budget Movies at the Box Office

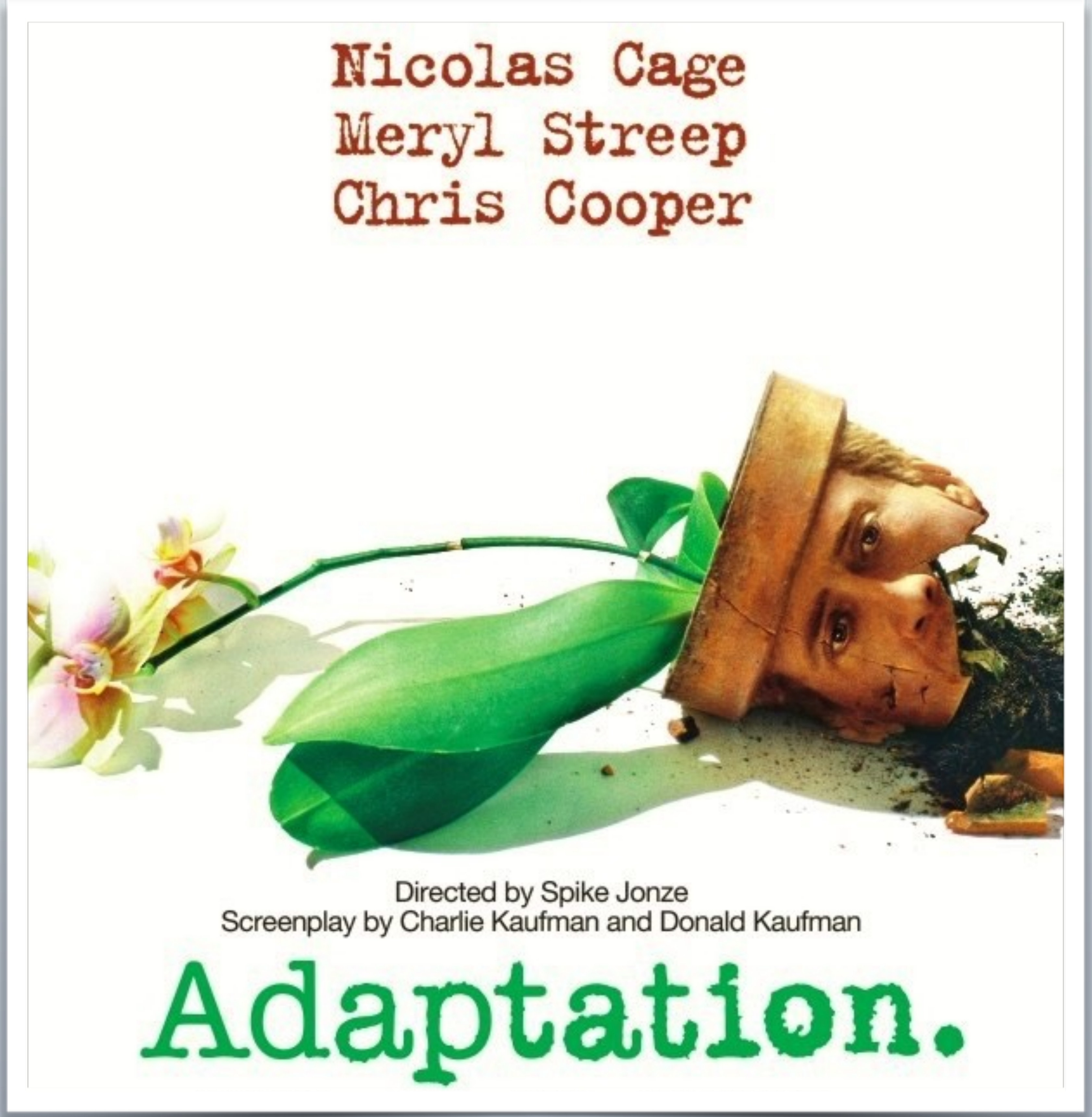
Project Luther: Revenge of Benson

Directed by: Steven Bierer

Produced in cooperation with Metis Data Science (Seattle Unit)



Released: February 14, 2003
Budget: $19 million
Box Office: $22.2 million



Released: February 14, 2003
Budget: $19 million
Box Office: $22.2 million

Released: November 15, 2002
Budget: $100 million
Box Office: $262.0 million

Overview

Overview

- Studios must generate revenue to keep producing
- <u>My question</u>: What determines box office success?

Overview

- Studios must generate revenue to keep producing
- <u>My question</u>: What determines box office success?
    - ➡ Especially for lower-budget films

Overview

- Studios must generate revenue to keep producing
- <u>My question</u>: What determines box office success?
    - ➡ Especially for lower-budget films
        - ➡ Using known factors available at release

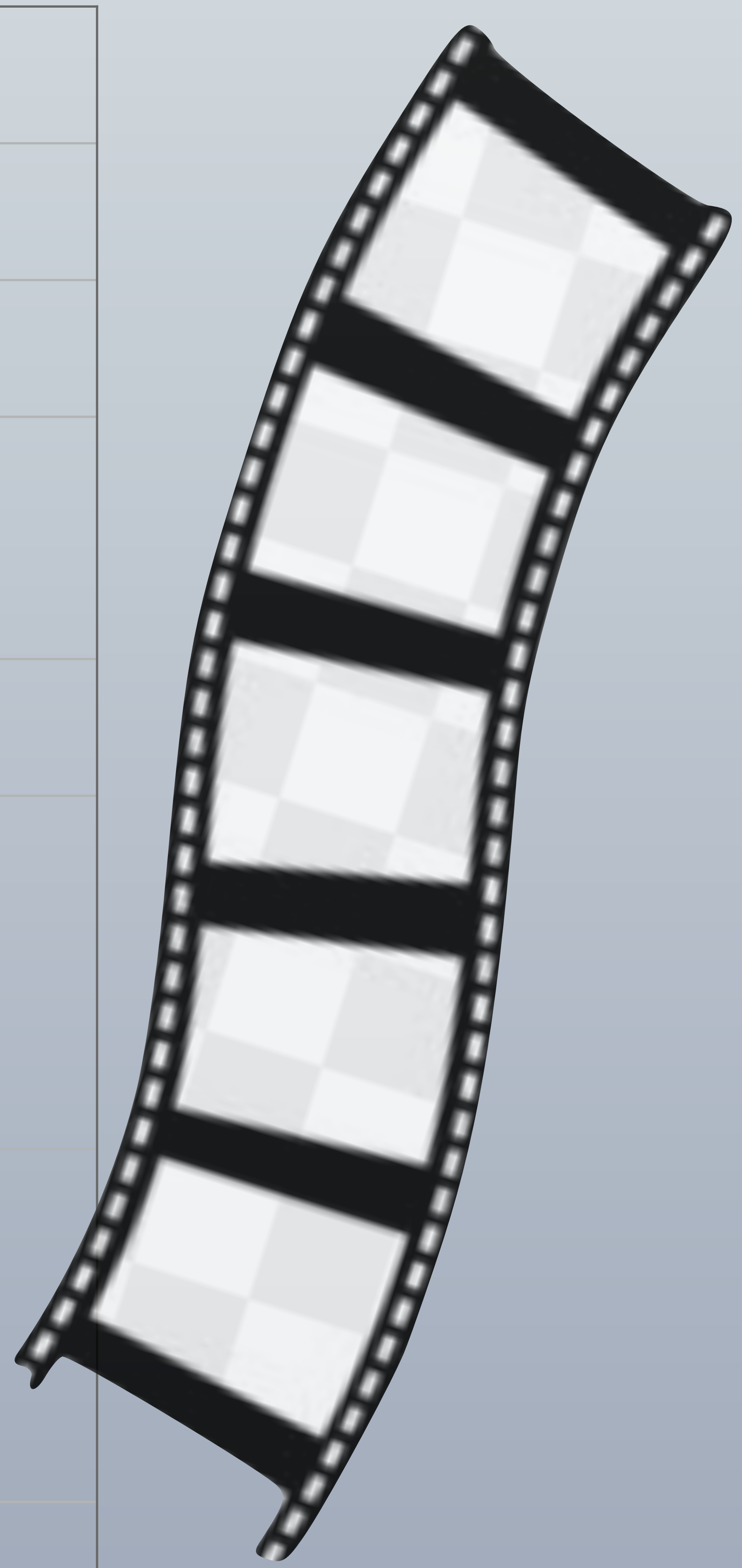
Overview

- Studios must generate revenue to keep producing
- My question: What determines box office success?
  - ➡ Especially for lower-budget films
    - ➡ Using known factors available at release
      - ➡ (But not whether Meryl Streep will be in it)

Methods - Scraping and Filtering

- Feature films (2008-2018) from IMDb.com

Data Element	*Type*	*Transformation*
Gross Box Office Sales	Numerical	Log
Opening Weekend Sales	Numerical	Log
Budget	Numerical	Log
Release Date	Numerical	Time filtering and averaging
Genre: Family, Comedy	Categorical	Dummy var
MPAA Rating: G, PG, etc	Categorical	Assigned rank (0 to 4)
User Rating and Count	Numerical	Not used

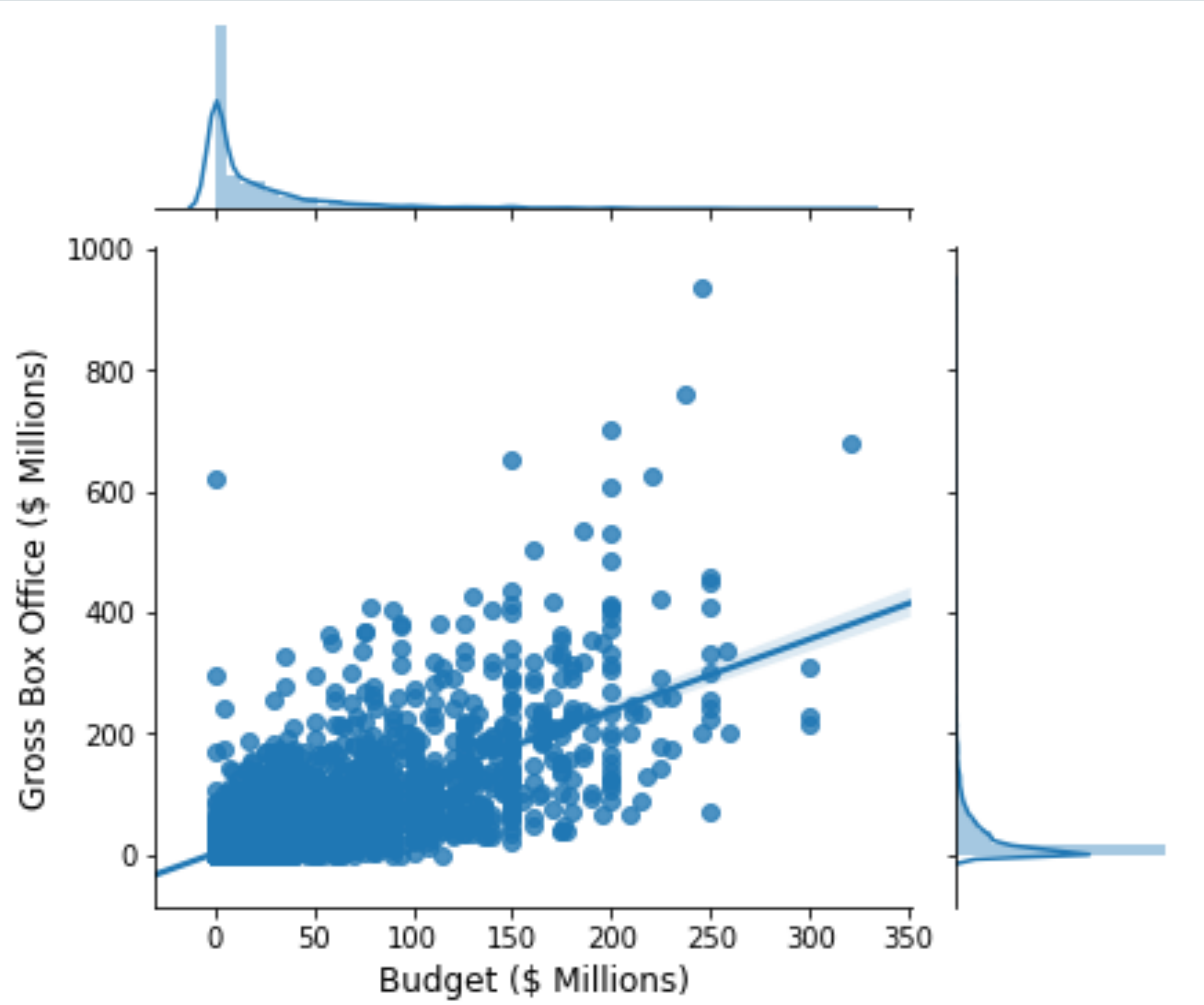


- Removed movies with no budget or sales data

Analysis

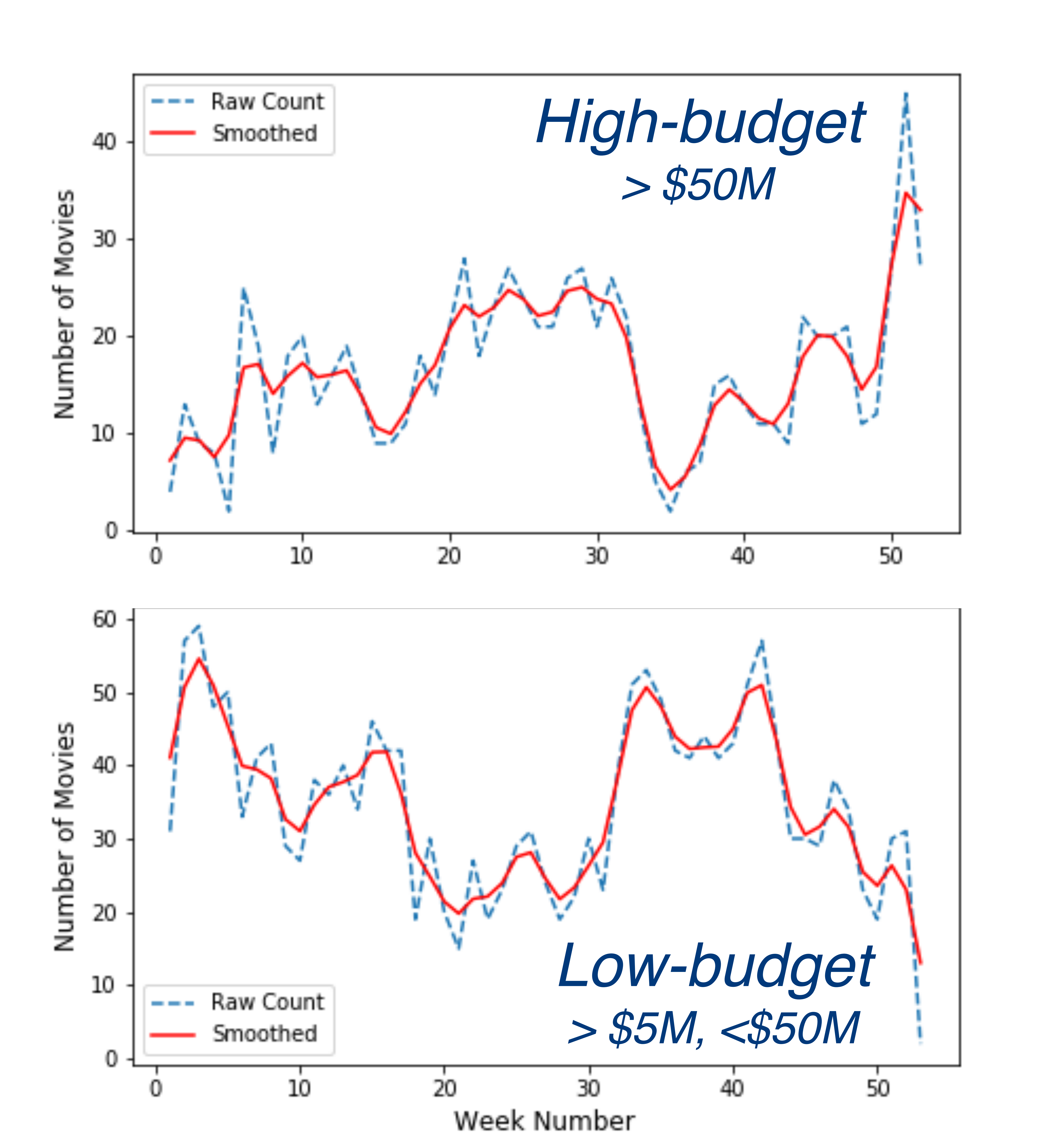
- Wide **variability** in box office across all films
- Good prediction with 1st-order model
  - Dominated by budget
  - R^2 = .483 (log)
  - R^2 = .421 (original)
- Same model on low-budget movies <u>not</u> good

Sales versus budget for *all movies* (n = 5887)



Analysis

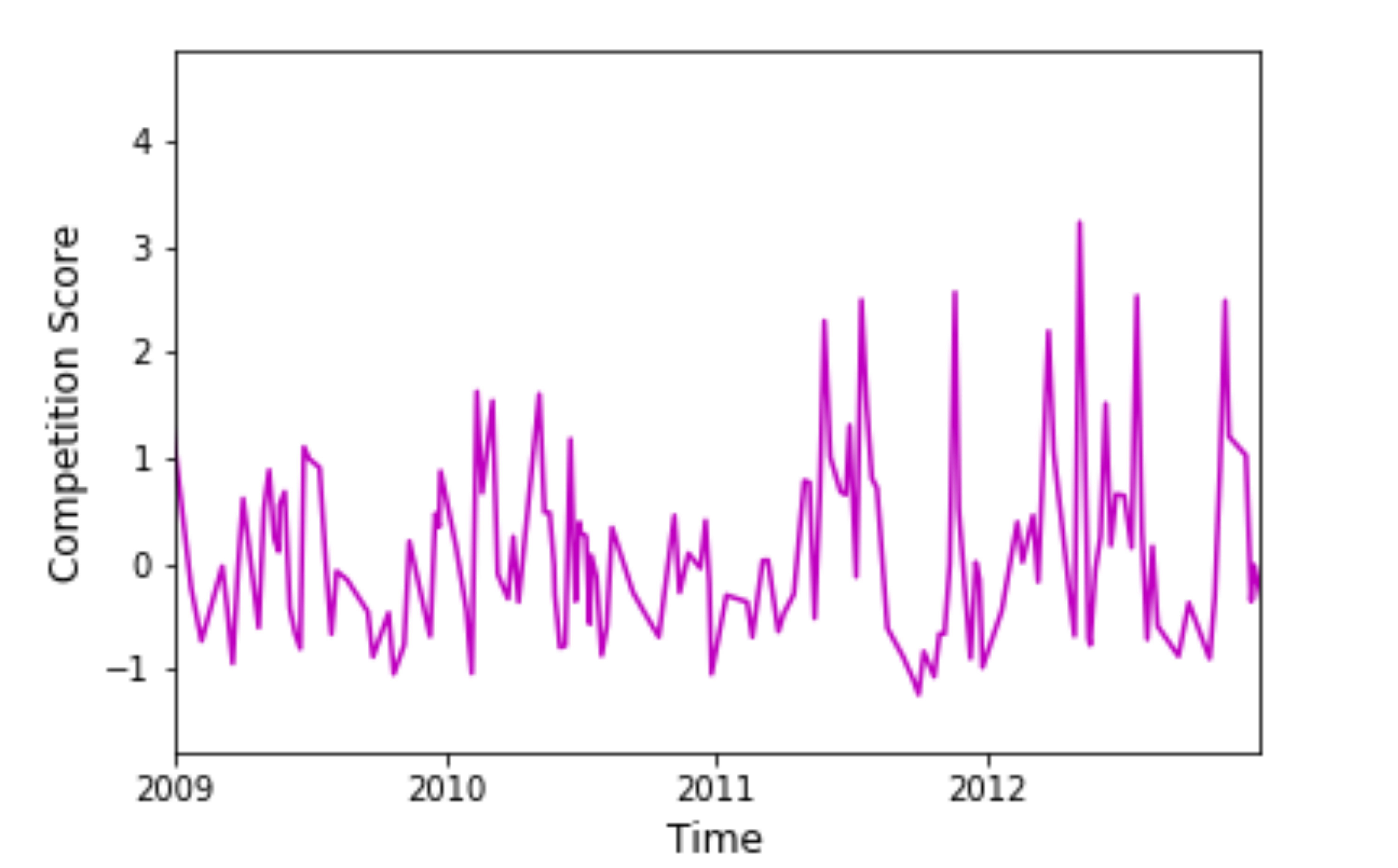
- **Seasonal trends** for release date are nearly *opposite* for low-budget (n=1848) and high-budget movies (n=866)
- Normalized “seasonality scores” were added as features



Analysis

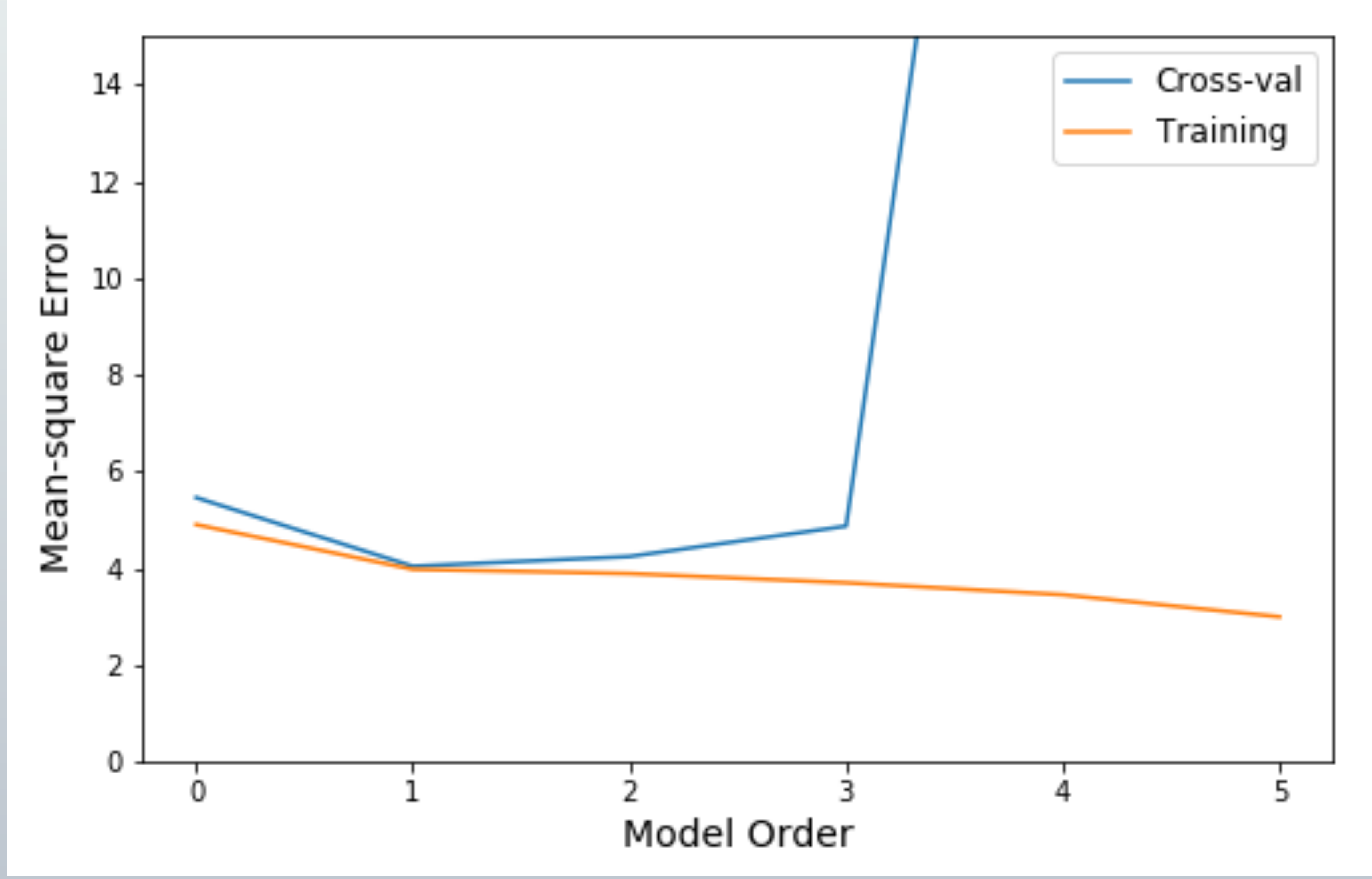
- **Competition** with high-budget movies was quantified based on opening week ticket sales

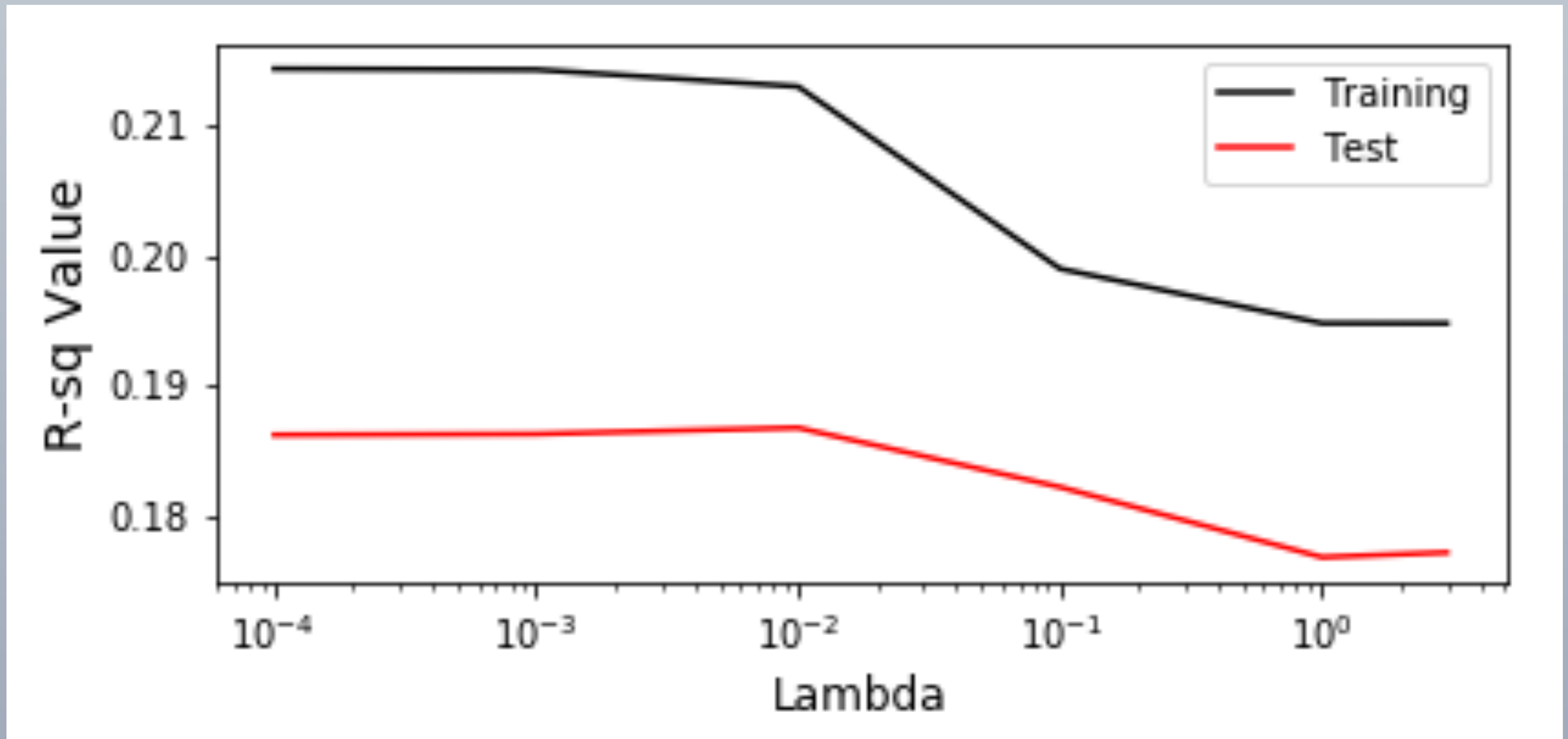
High-budget "Competition" Score



Analysis

Low-budget data

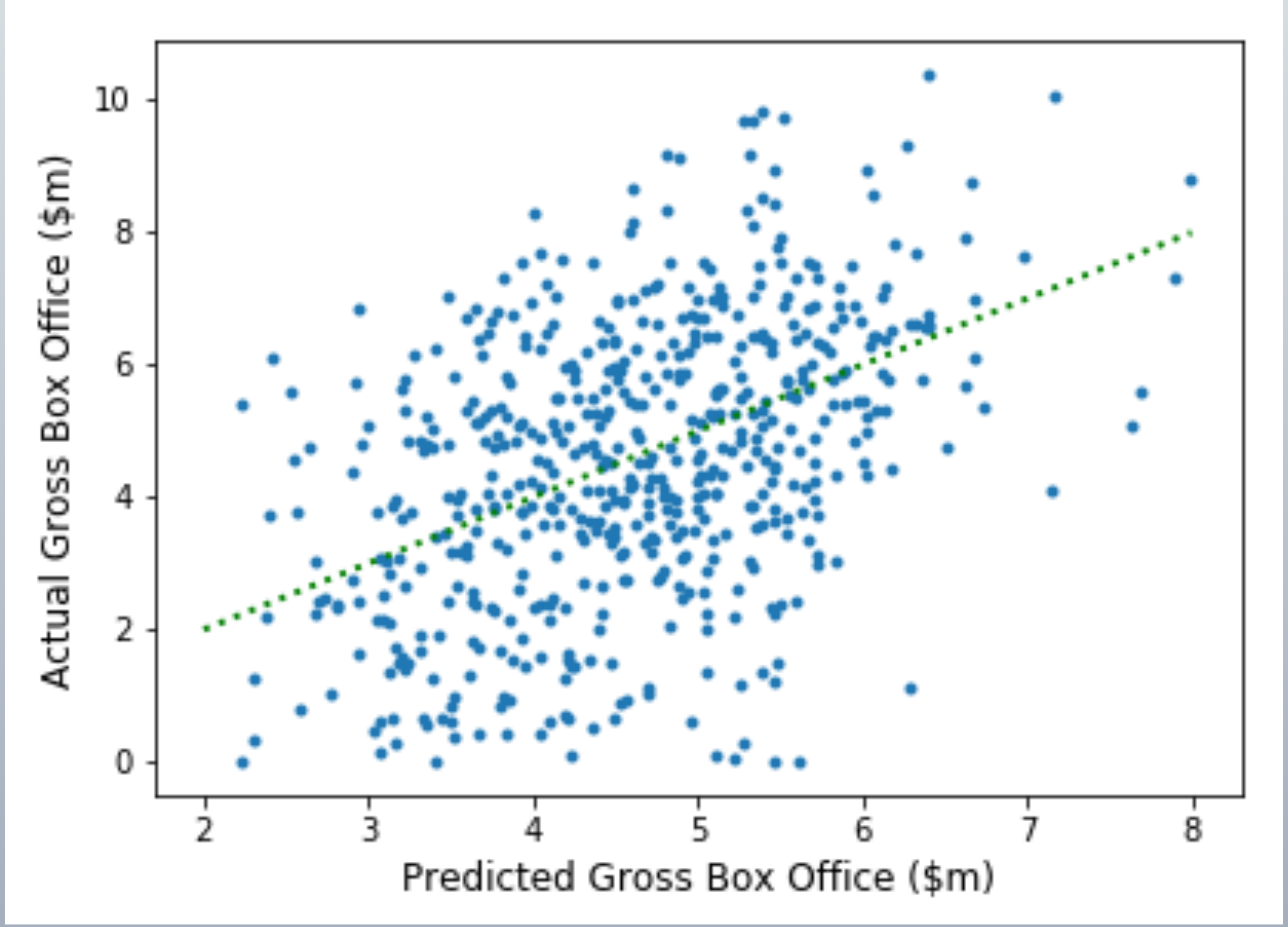




- **Cross-validation** analysis suggested a 2nd- or 3rd-order model

- **Lasso Regularization** for 2nd-order model refined feature selection

Conclusions - Low-budget data



- A final **R-squared value** (untrained, transformed data) of only 0.19 was achieved.

- Important features:

Budget * LB-Season (-)

Budget * Comedy (+)

LB-Season * HB-Competiton (-)

Comedy * MPAA-Lvl (-)

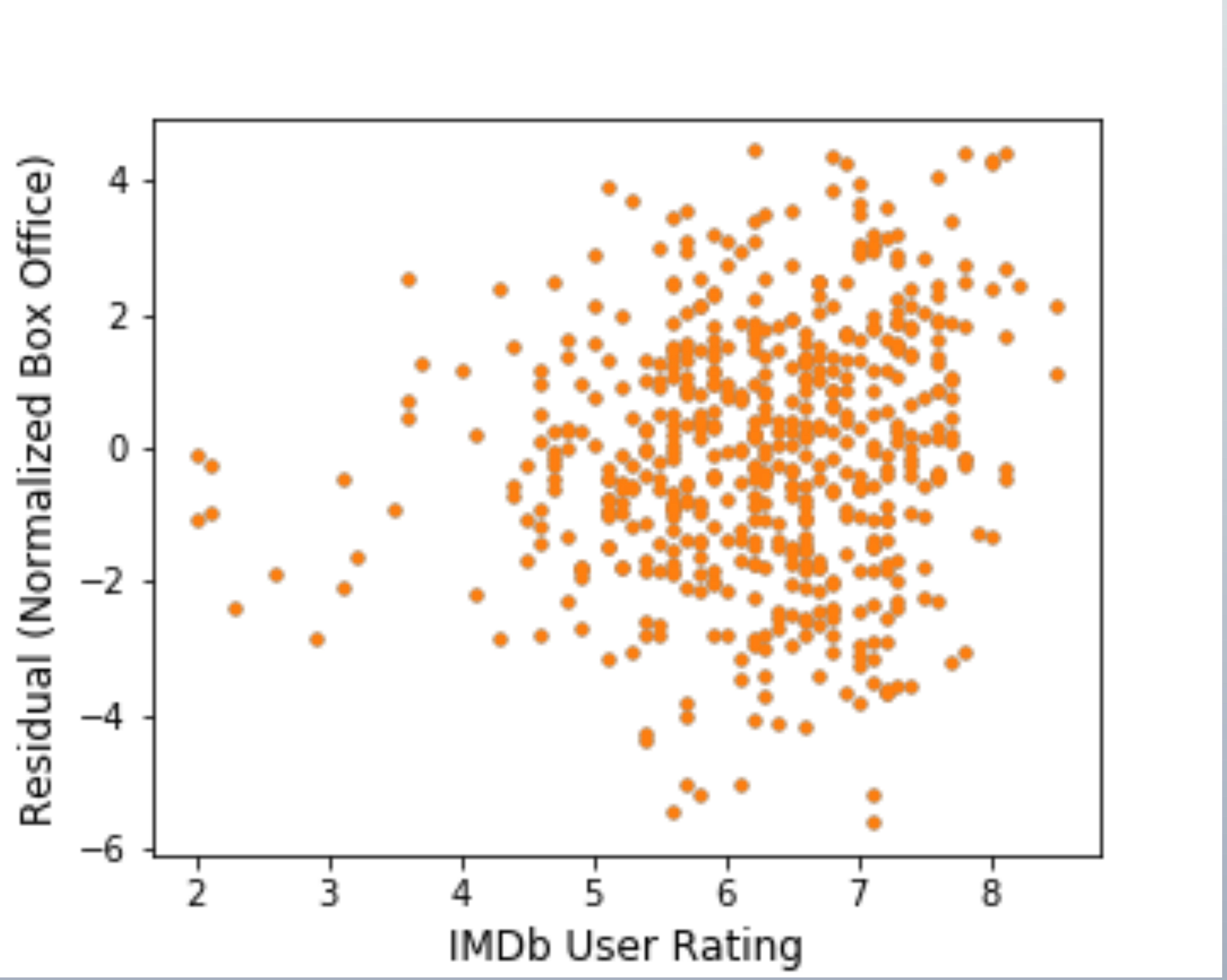
Future Analysis

- Evaluate influence of **genre**
  - Genre-specific competition, week-by-week
- Include measure of pre-release “buzz”
  - Early proxy for critical reception??
- Implement as **mixed model**
  - Random effects of year, genre, etc
- More data!


THE END
WB
A Warner Bros.
PICTURE

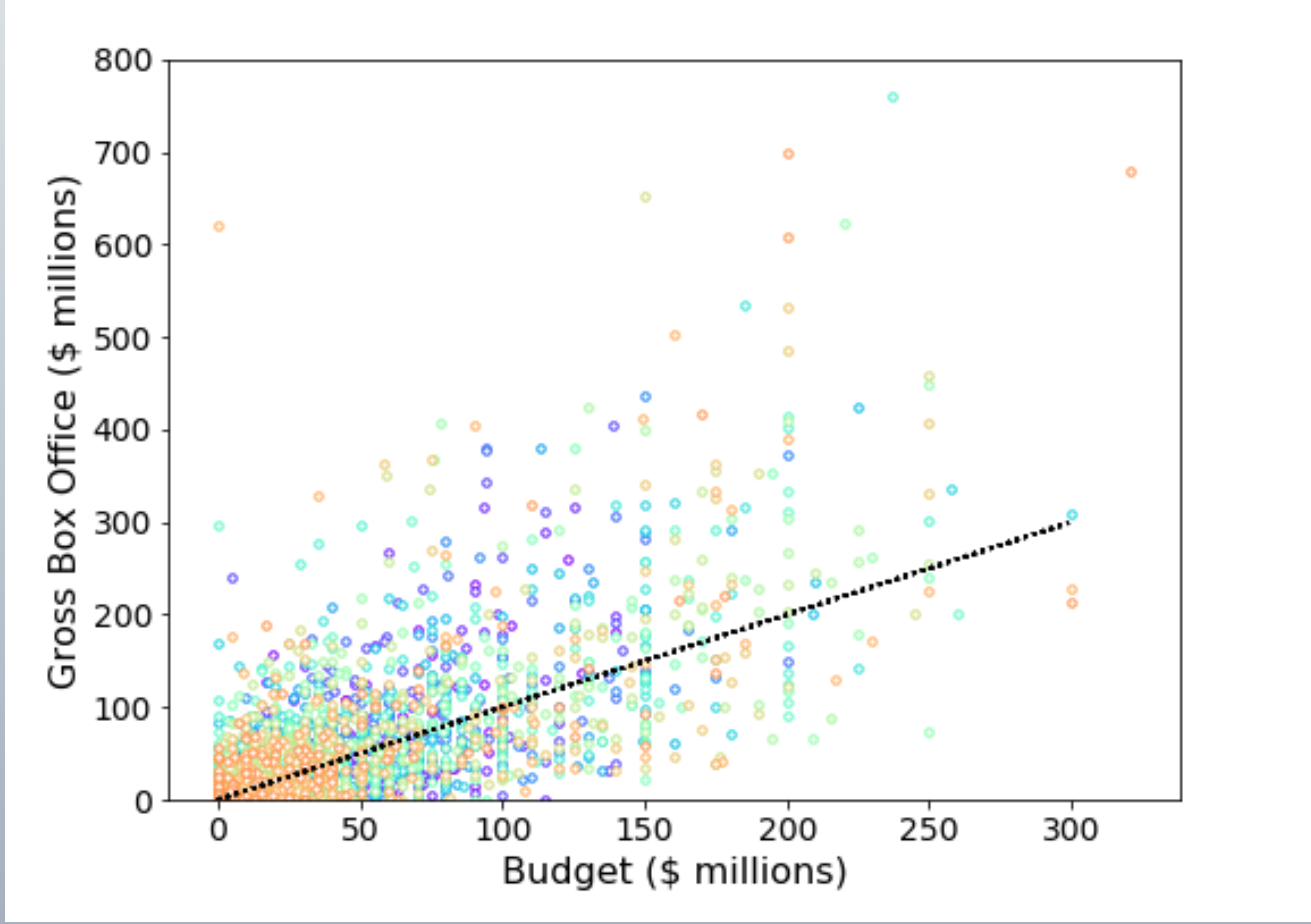
Appendix

Residual Error vs IMDb Movie Rating



Analysis

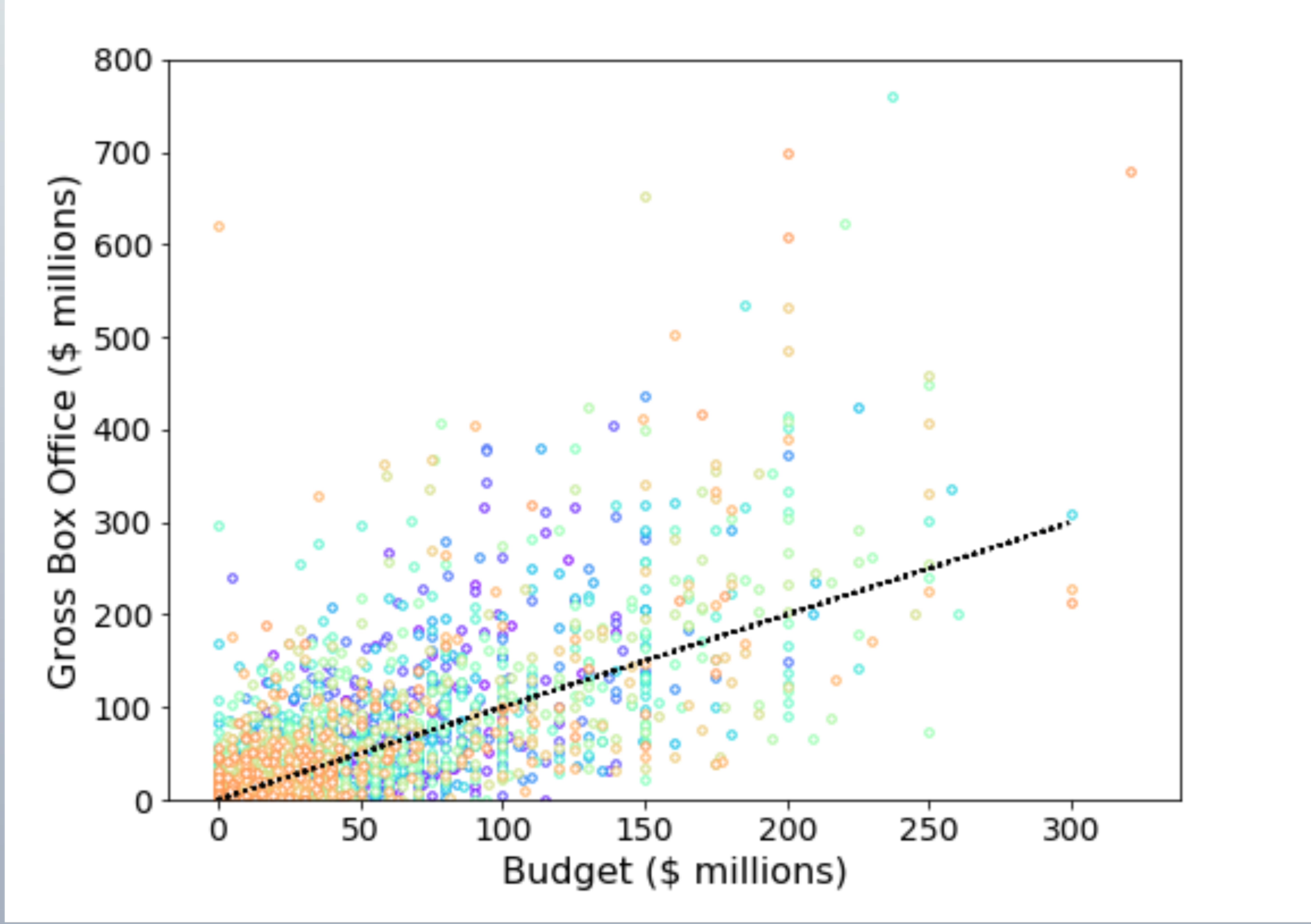
Sales versus budget for *all movies* (n = 5887)



Analysis

- Wide **variability** in box office across all films
- Good prediction with 1st-order model

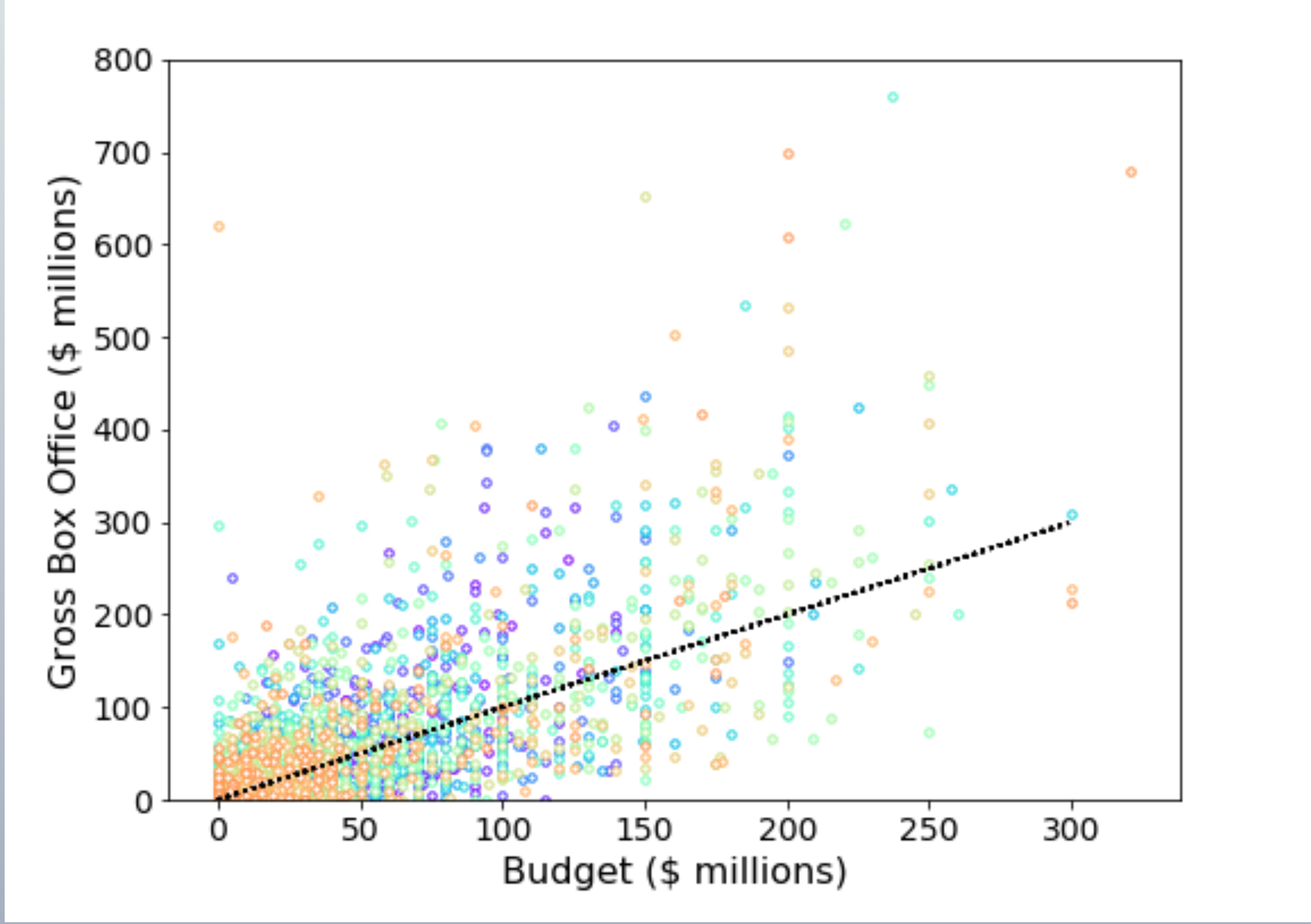
Sales versus budget for *all movies* (n = 5887)



Analysis

- Wide **variability** in box office across all films
- Good prediction with 1st-order model
  - Dominated by budget

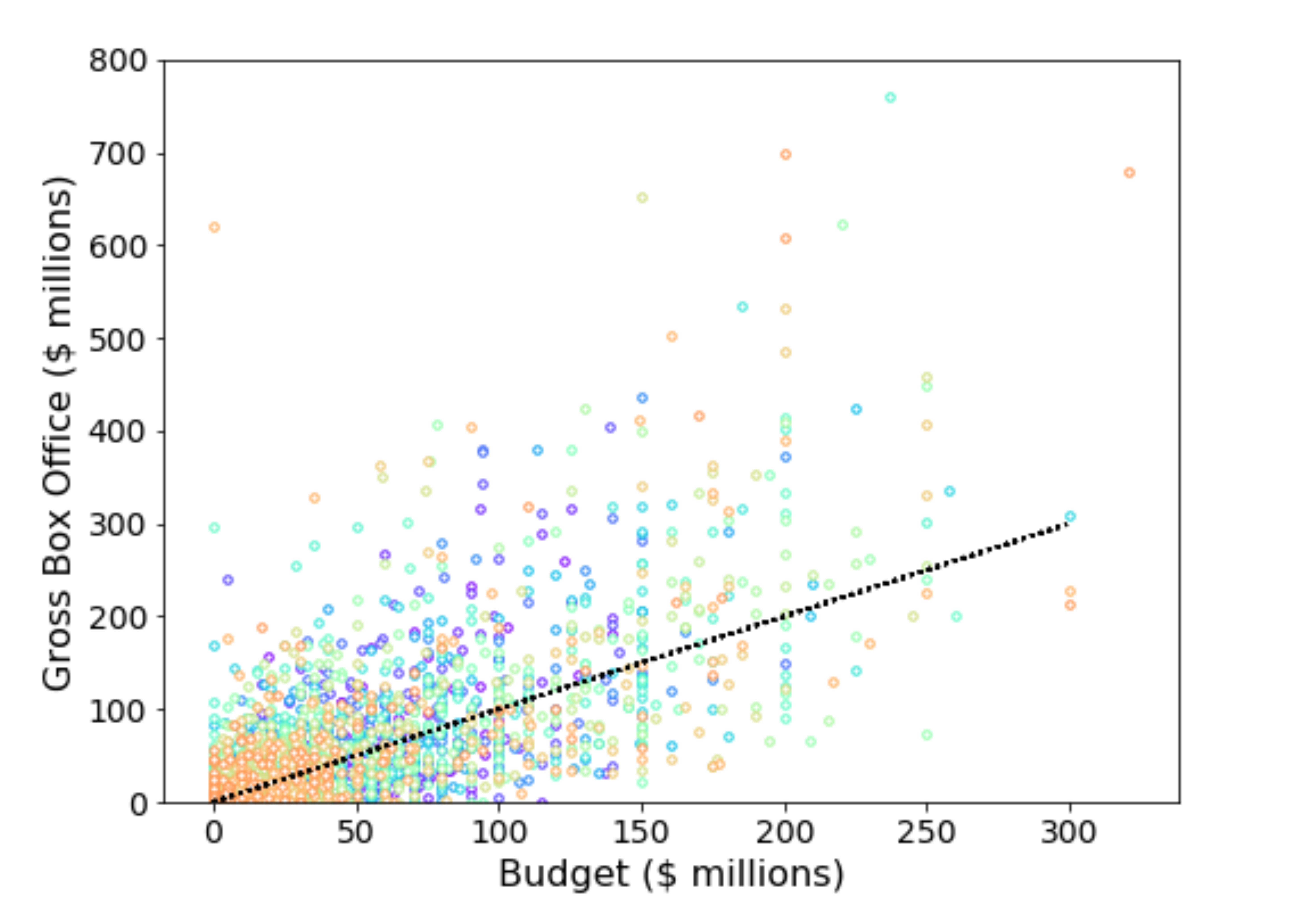
Sales versus budget for *all movies* (n = 5887)



Analysis

- Wide **variability** in box office across all films
- Good prediction with 1st-order model
    - Dominated by budget
    - $R^2 = .483$ (log)

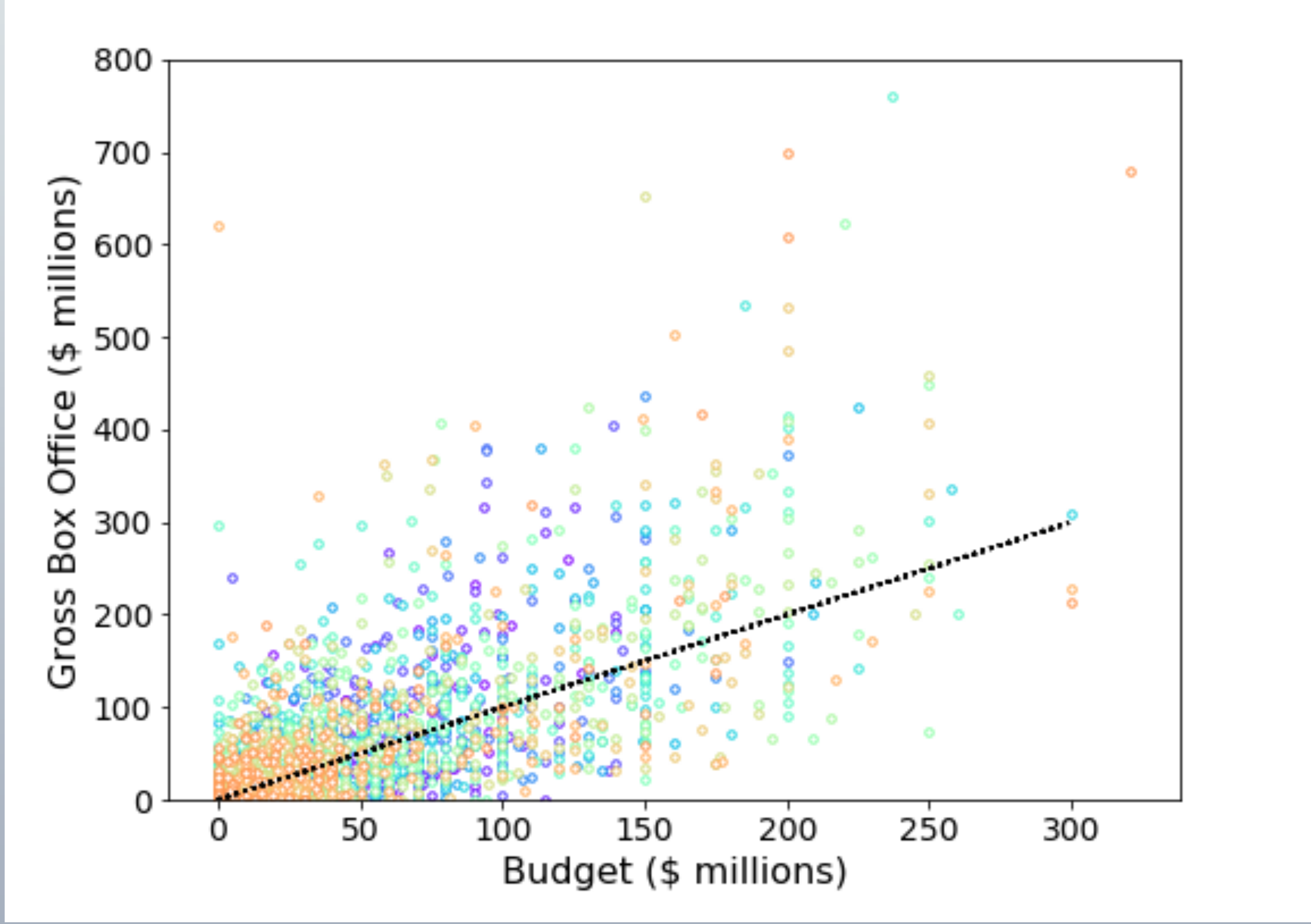
Sales versus budget for *all movies* (n = 5887)



Analysis

- Wide **variability** in box office across all films
- Good prediction with 1st-order model
  - Dominated by budget
  - R^2 = .483 (log)
  - R^2 = .421 (original)

Sales versus budget for *all movies* (n = 5887)



Analysis

- Wide **variability** in box office across all films
- Good prediction with 1st-order model
  - Dominated by budget
  - $R^2 = .483$ (log)
  - $R^2 = .421$ (original)
- Same model on low-budget movies <u>not</u> good

Sales versus budget for *all movies* (n = 5887)