

# Movie Data Analysis

Tim Hügele and Mark Subra





# Data Sets

- IMDb (5000 total movies)
  - Obtained from Kaggle  
(<https://www.kaggle.com/carolzhangdc/imdb-5000-movie-dataset>)
- Box Office Mojo
  - Obtained by Python web scraper: 1400 highest grossing movies

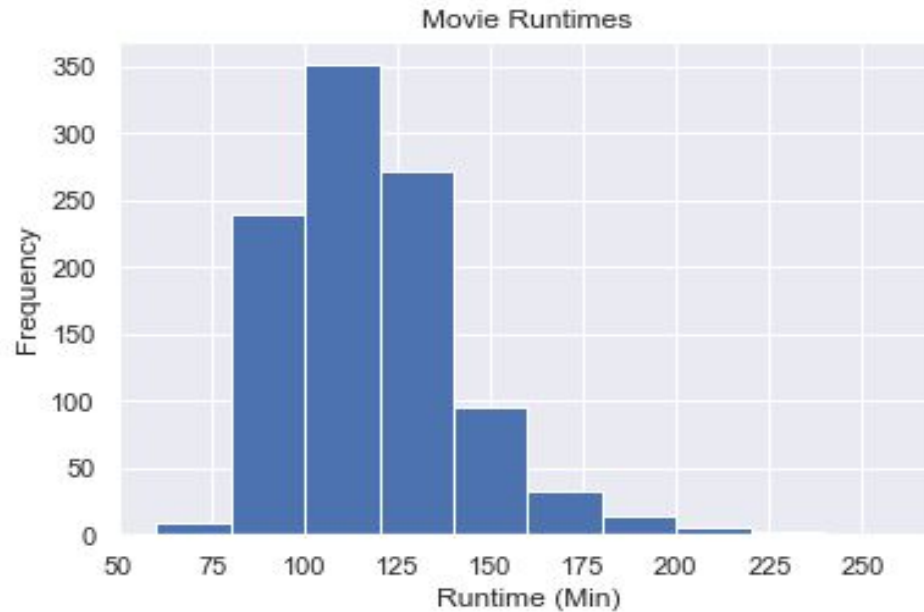


## Question

Does a movie's runtime affect its success at the box office or its rating on IMDb?

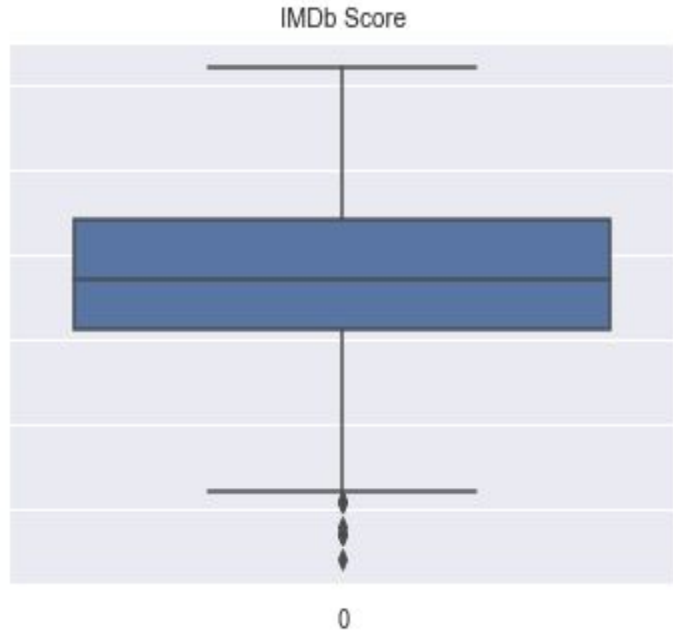
What kinds of movies make the most at the box office, and do they make more domestically or abroad?

# Movie Runtimes

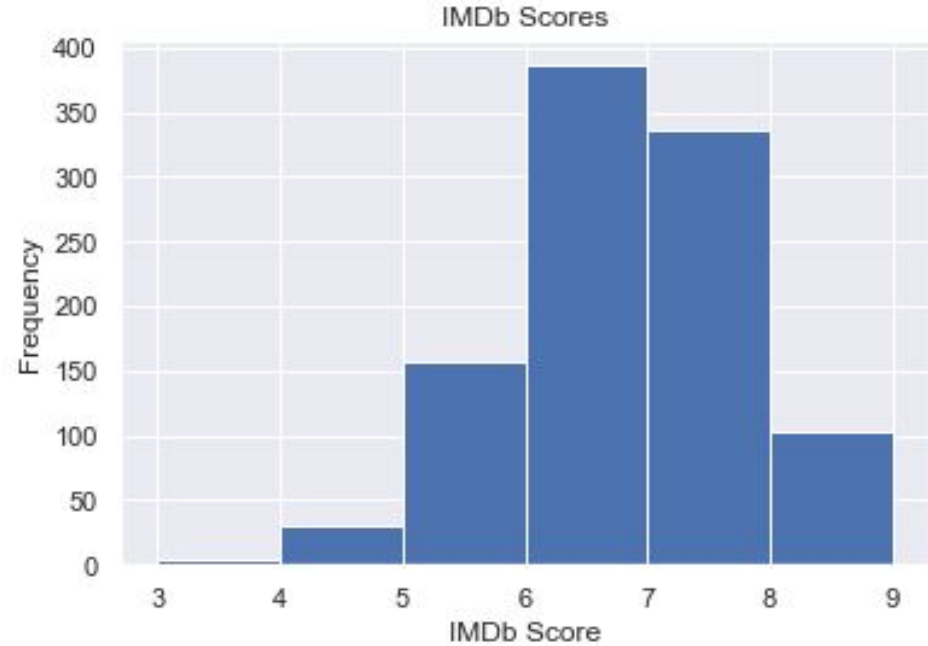


Average Runtime = 117 minutes

# IMDb Ratings



Some extreme values, don't significantly affect data



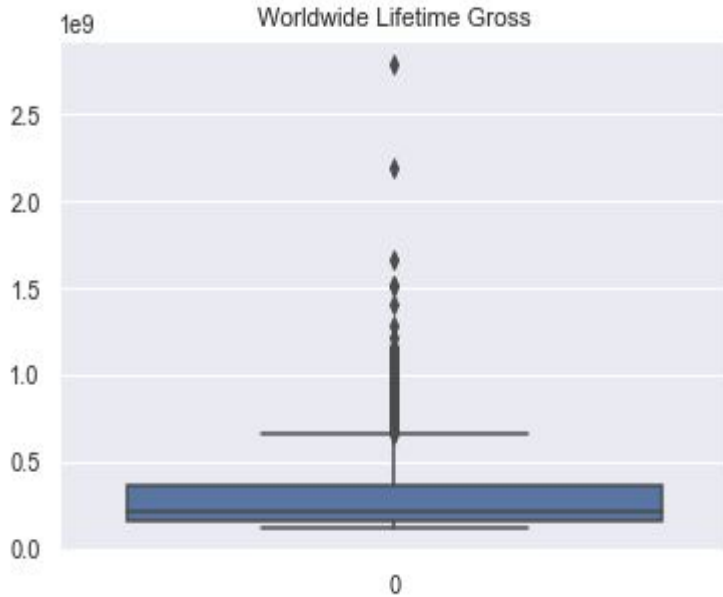
Average = 6.75

# Runtimes vs. IMDb Ratings

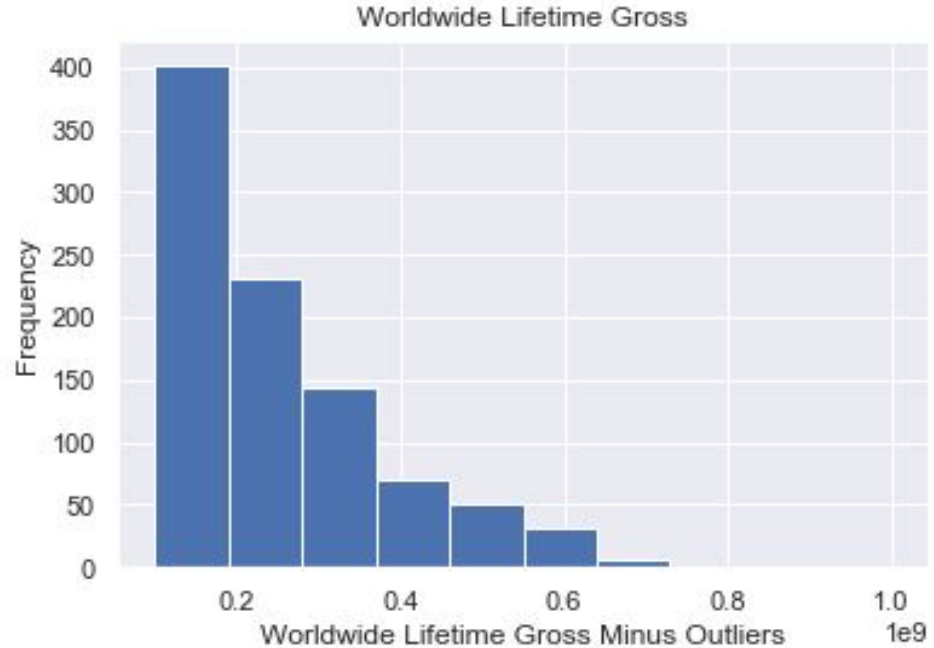
- Corr. = 0.407
- Corr. - Outliers = .411
- < Mean Runtime = IMDb Score 6.45
- > Average Mean = IMDb Score 7.13



# Worldwide Lifetime Gross Revenue

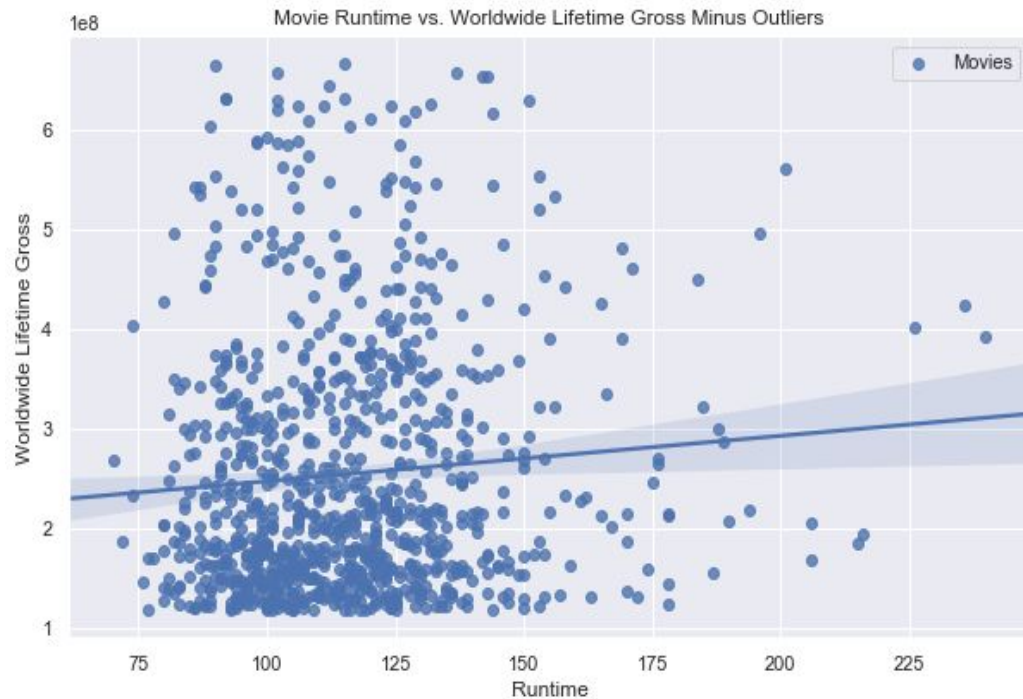


Data with Extreme Values



Average w/o Extremes = \$254 Million

# Runtimes vs. Worldwide Lifetime Gross Revenue



- Corr. = 0.261
- Corr. - Outliers = 0.0813
- < Mean Runtime = \$248 Mil.
- > Mean Runtime = \$261 Mil





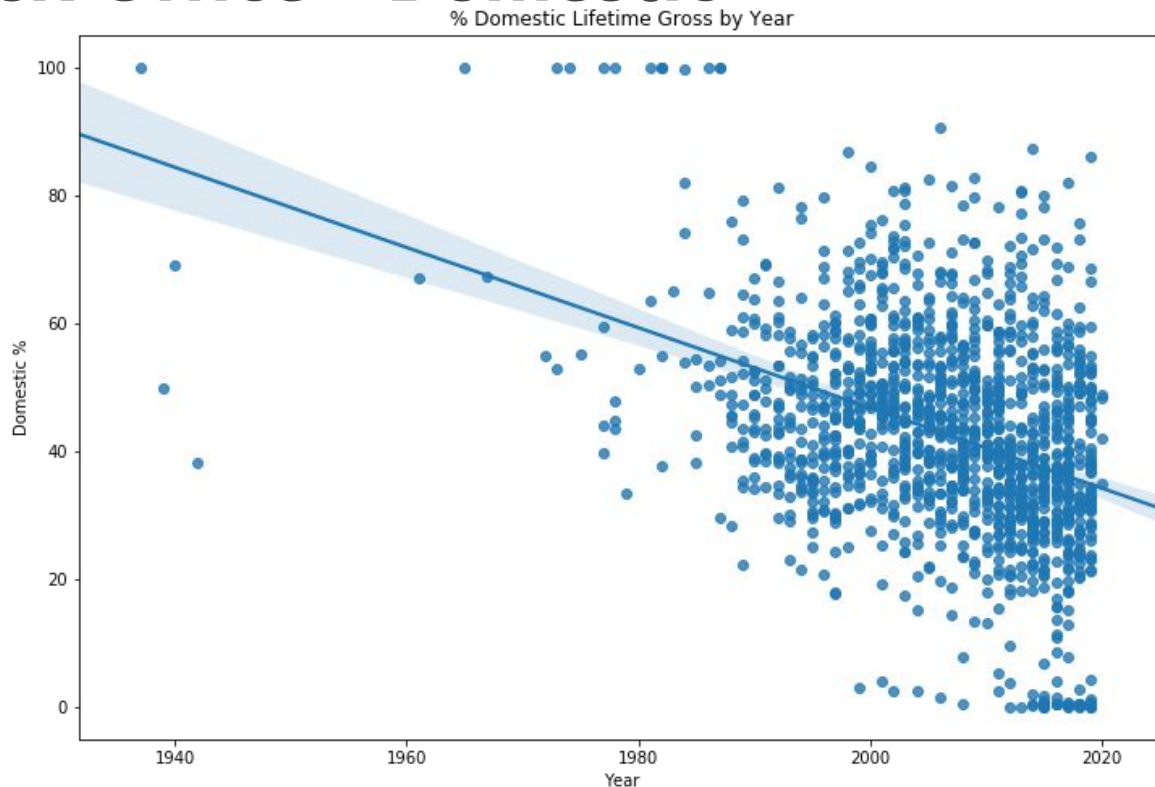
## Runtime Conclusions

- Longer runtimes tend to mean higher IMDb Ratings
- Longer movies have a weak relationship with Gross Revenue
- Further study - Would like to look into Budget Data. Longer movies might be correlated with higher budgets. If longer movies don't generate more revenue, longer movies might not be a good investment.



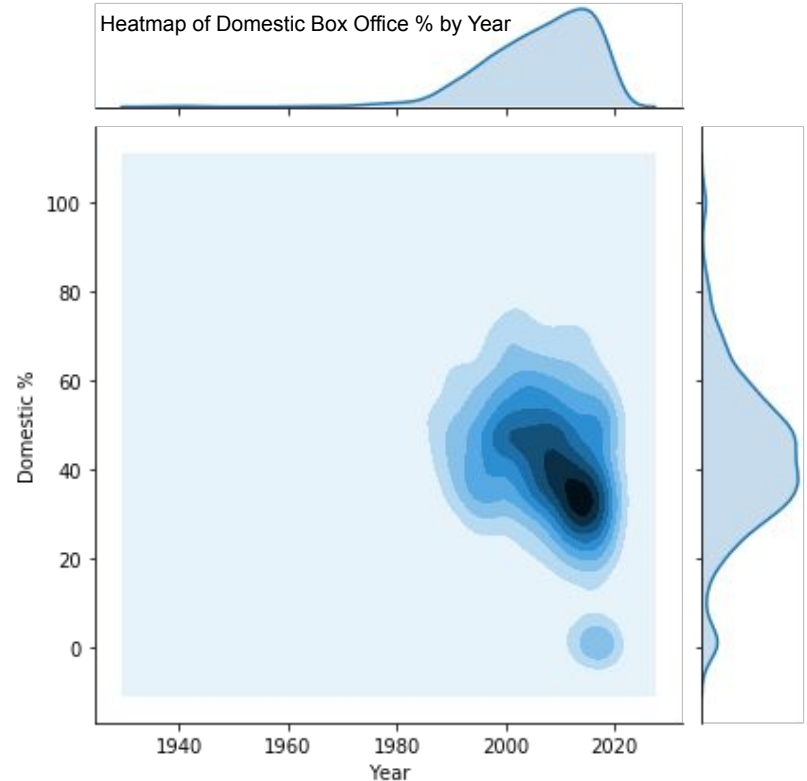
# Box Office - Domestic

- Percent of domestic total revenue declining by time
- Negative correlation
- Domestic revenue and year are correlated by  $-0.37$



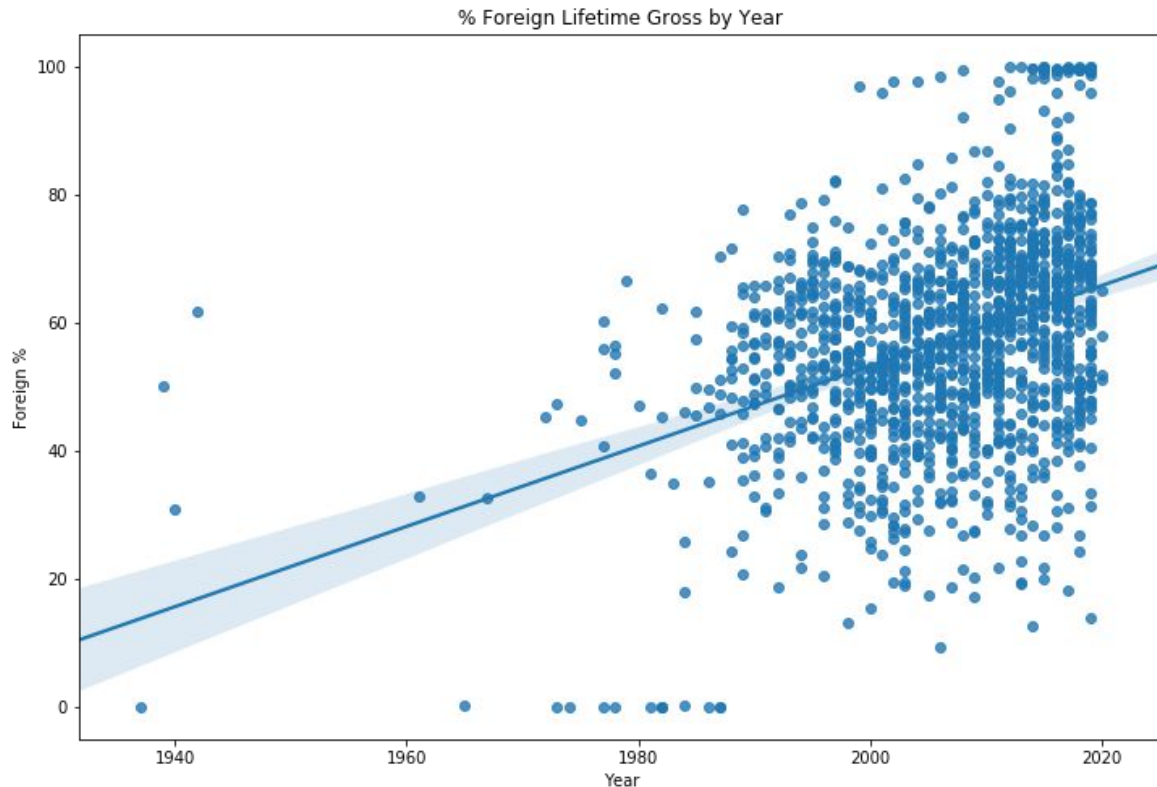
# Box Office - Domestic

- Percent of domestic total revenue declining by time
- Negative correlation
- Domestic revenue and year are correlated by -0.37
- Older films had higher domestic percentage of revenue
- Newer films are making less percentage domestically



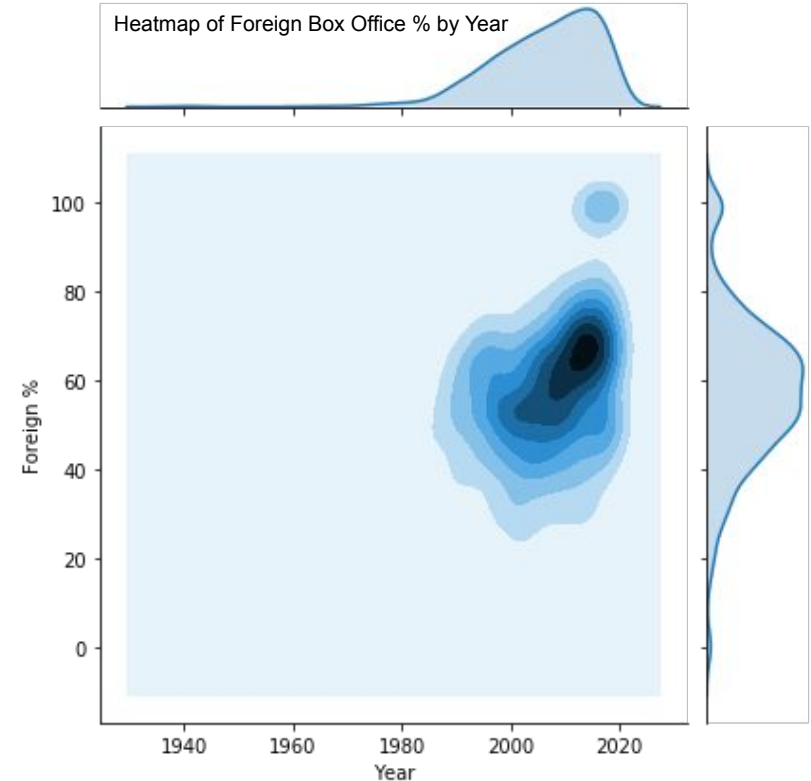
# Box Office - Foreign

- Percent of foreign total revenue increasing over time
- Positive correlation
- Foreign revenue and year are correlated by  $+0.37$



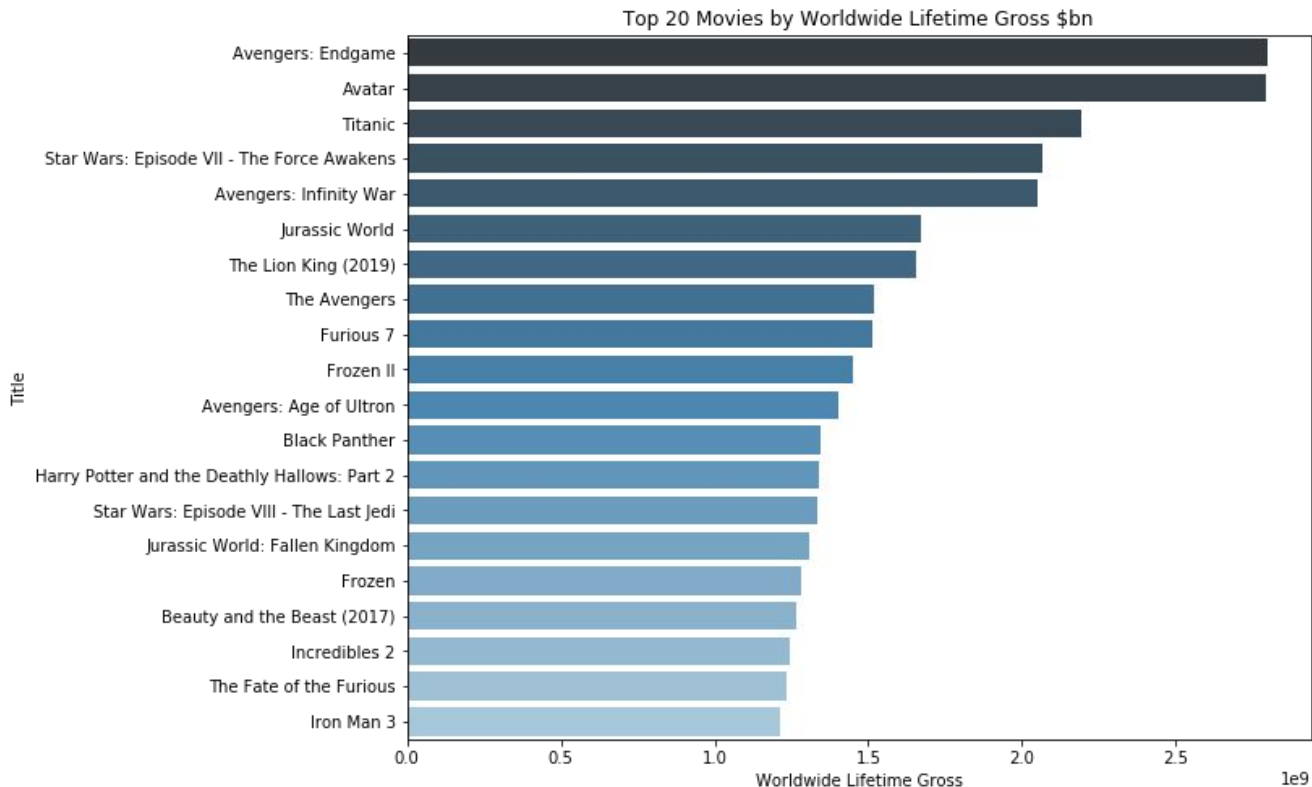
# Box Office - Foreign

- Percent of foreign total revenue increasing over time
- Positive correlation
- Domestic revenue and year are correlated by  $+0.37$
- New films have higher foreign percentage of revenue
- Newer films are making more percent of revenue abroad



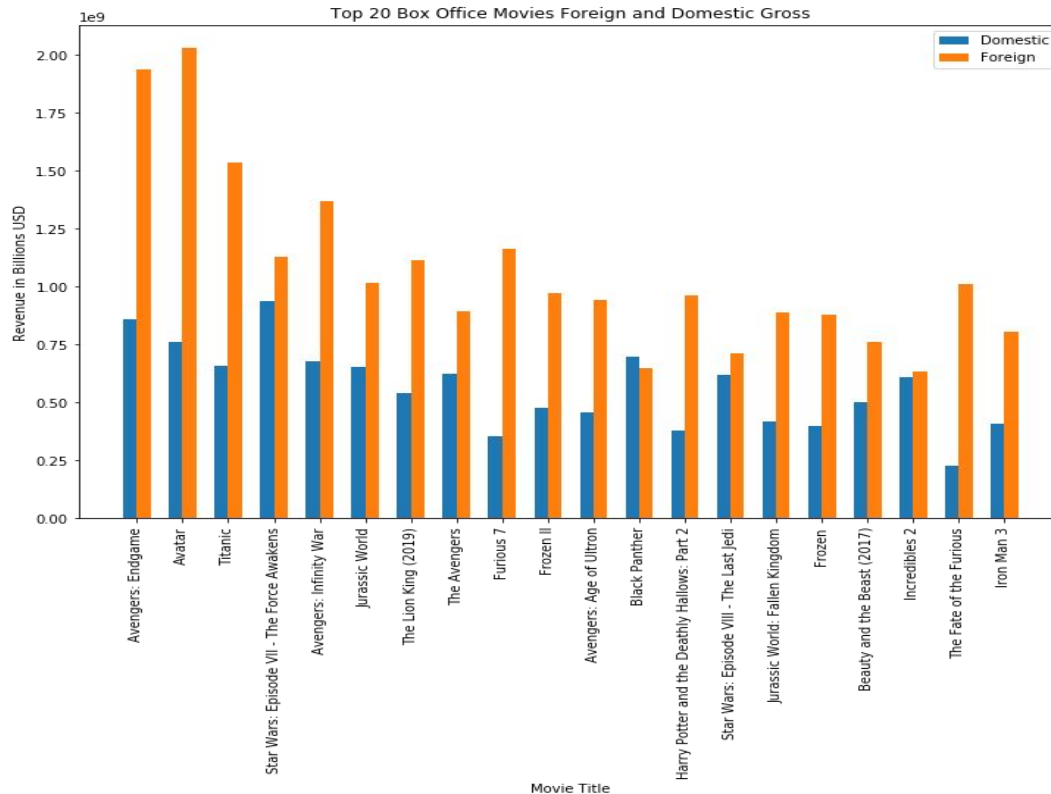


# Box Office - Top 20 Movies of All Time

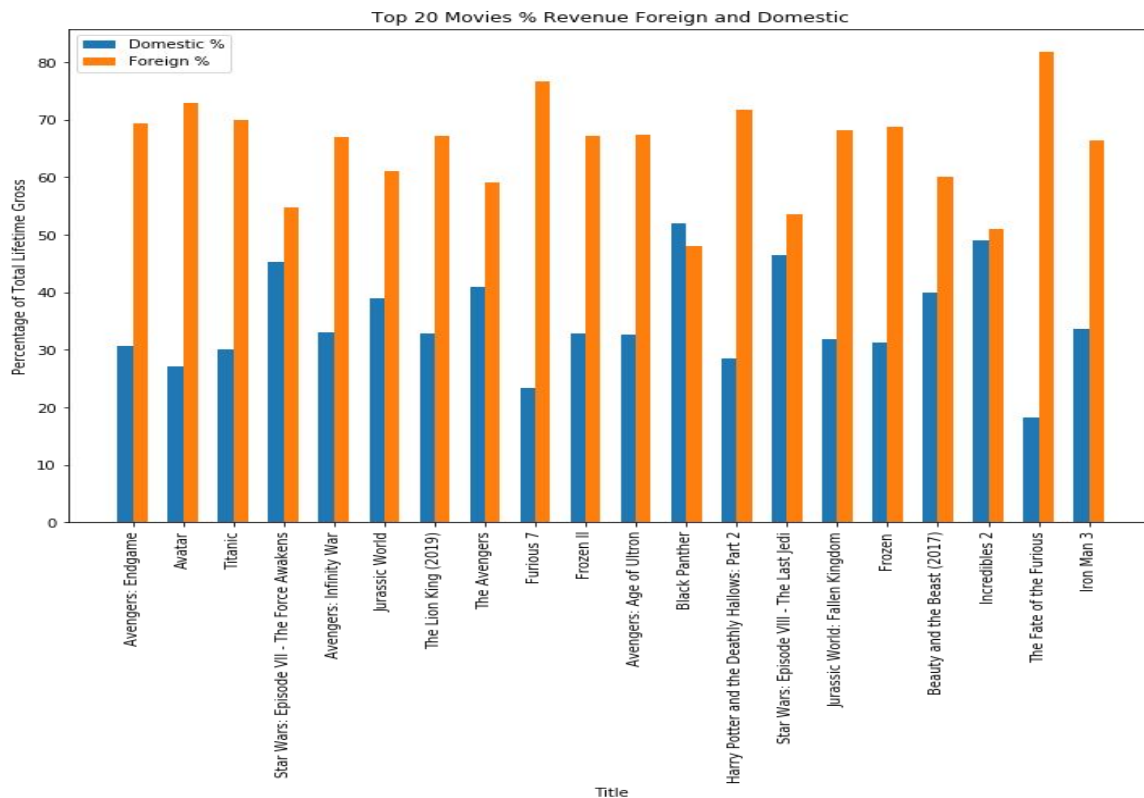


# Box Office - Top 20 Movies of All Time

- Sequels
- Remakes
- Major franchises
- Recognizable to foreign audiences
- PG-13 or lower MPAA rating
- Released 2009 or after, except Titanic



# Box Office - Top 20 Movies of All Time

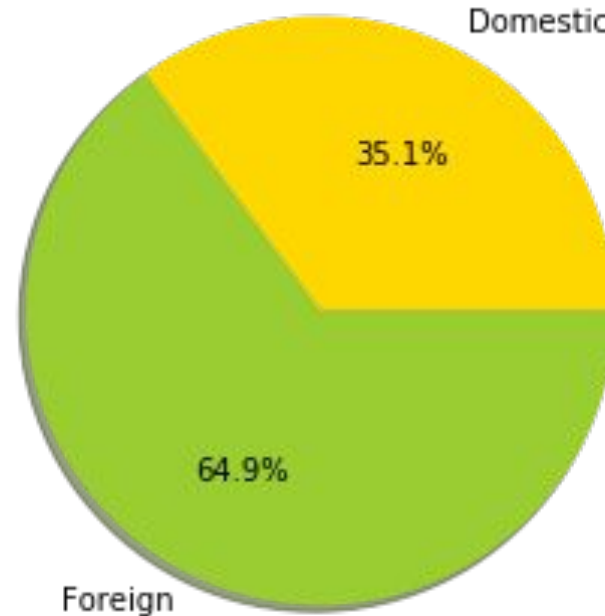




# Box Office - Top 50

Box Office Revenue of Top 50 Grossing Movies

- Worldwide lifetime gross combined of top 50 box office movies
  - \$64,555,201,083
- Domestic
  - \$22,693,185,633
- Foreign
  - \$41,862,015,450





## Box Office Conclusions

- Foreign box office much more significant now than prior
- Highest grossing films: Sequels, Remakes, and Major Franchises
  - Recognizable to foreign audiences
- PG-13 or lower - suitable for families and children
- Special effects, visually appealing
  - Enhanced theater experience
    - More tickets sold, higher price per ticket