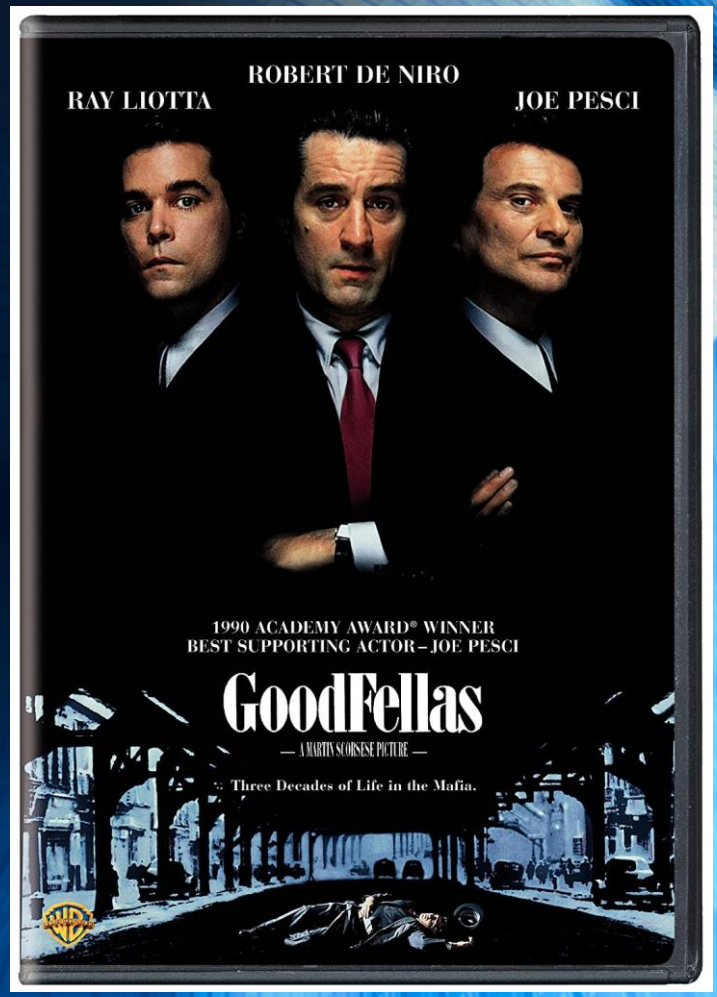
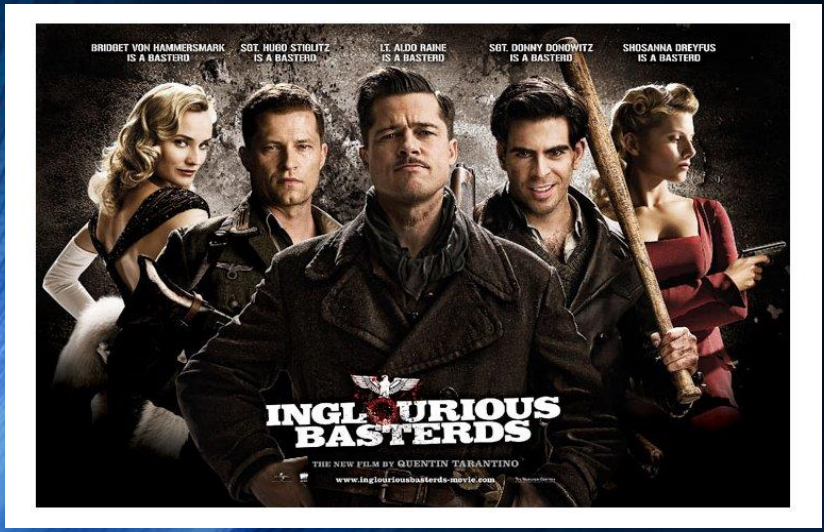
Movie Fun

SEAN WILLIAMS
RANDY TREAT
MARK WILLIAMS



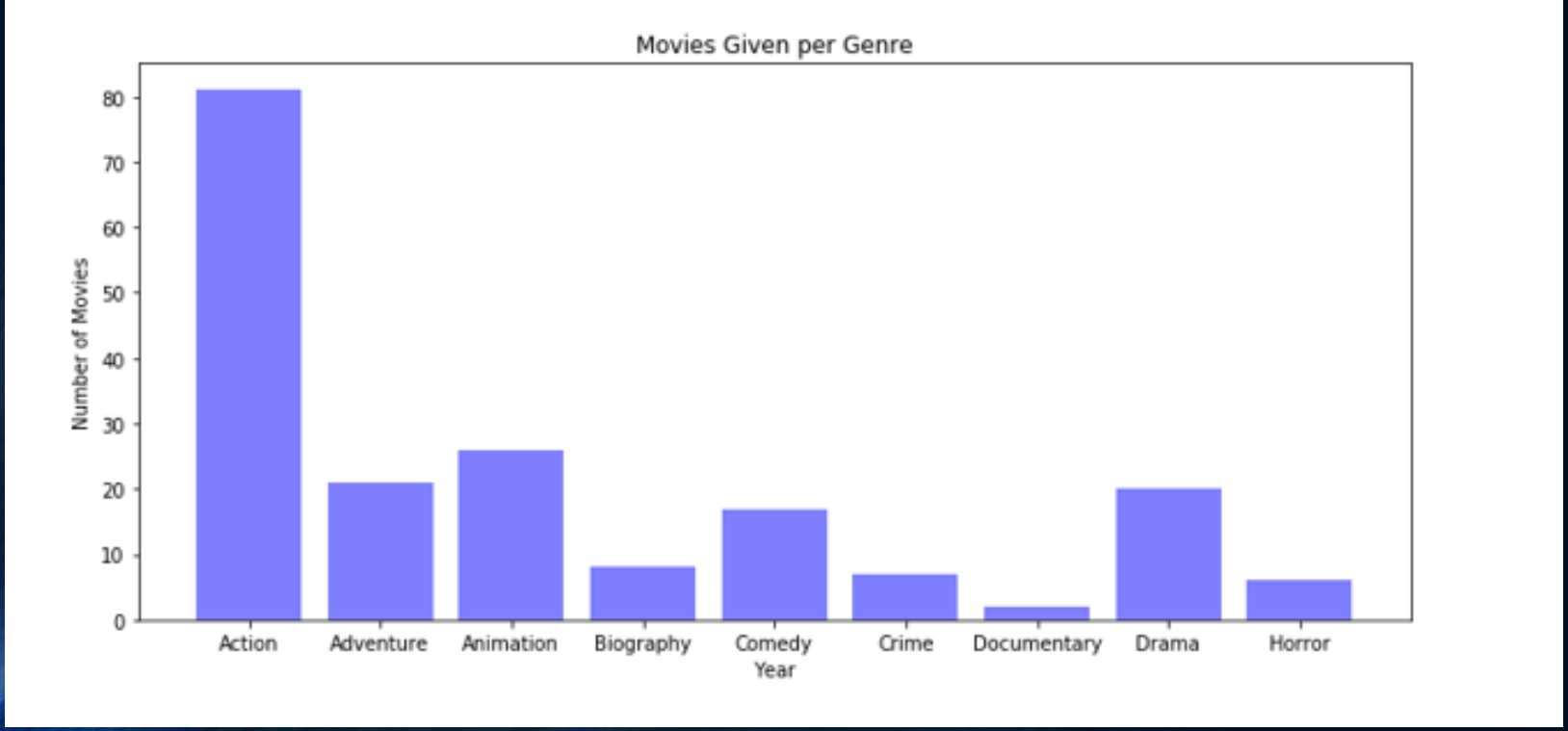
Summary

- (Enter finalized hypothesis here)
- Why we asked questions we did?
- Able to answer questions? Satisfied with results?
- Conclusion



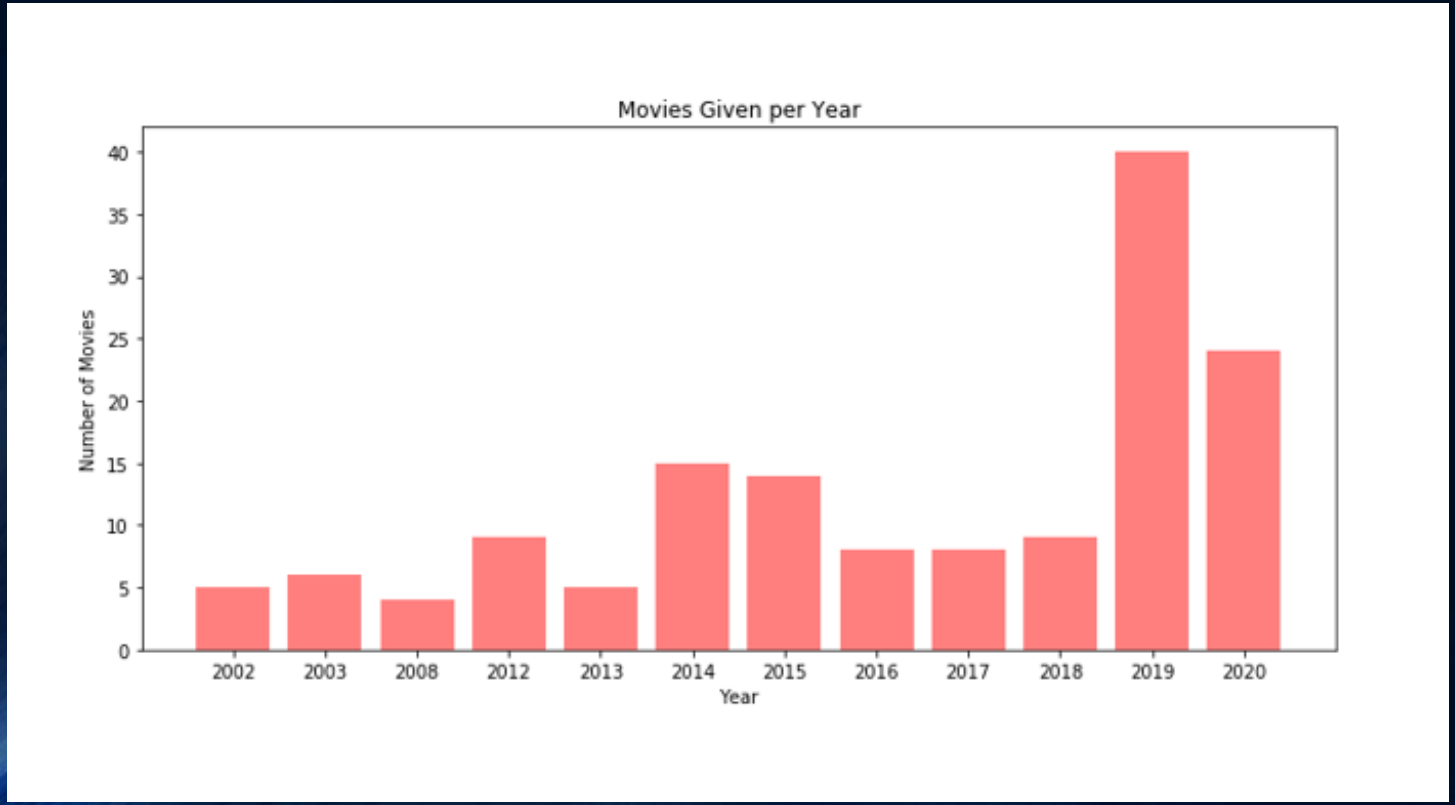
What genre of movies are generated when using Random Lists?

-Use of Random Lists data
-Narrowed scope
-Only included genres with one or more films



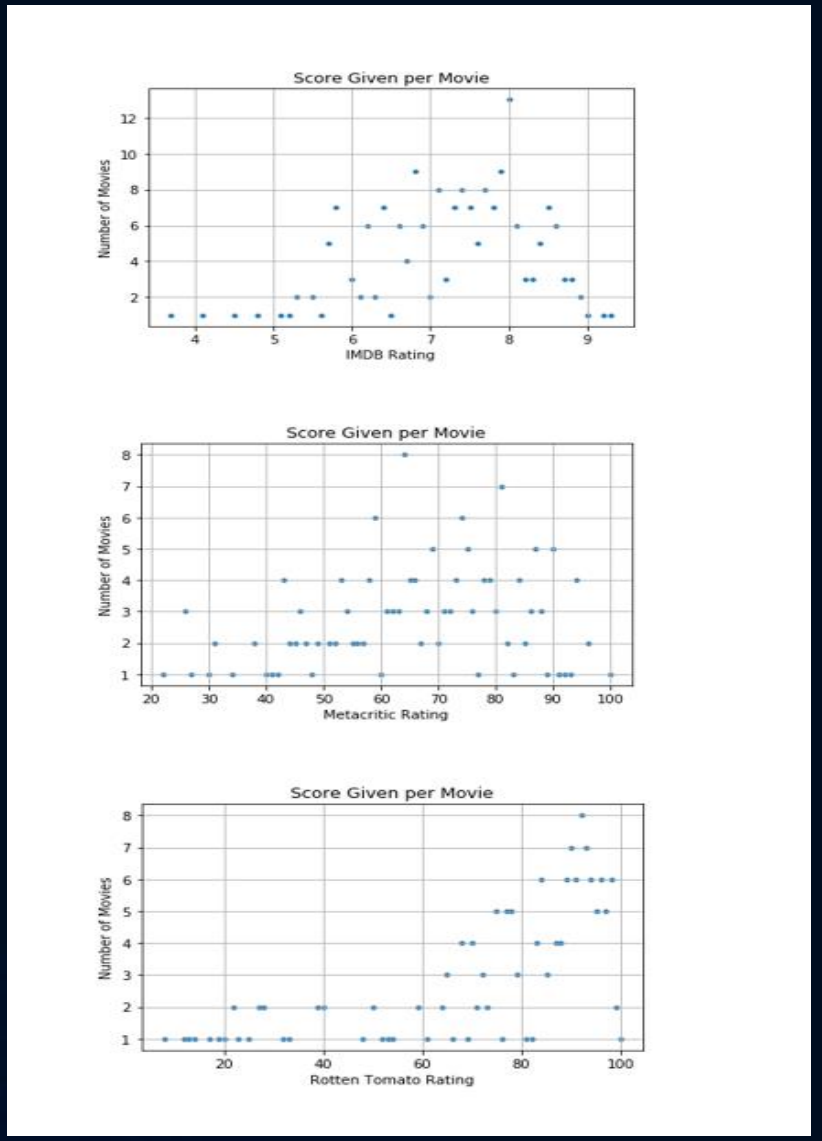
Are the movies from a specific time period? Are they newer films, older films, or is there a mix?

-Aimed to isolate the years with the most movies
-Excluded years that didn't include more than four titles
-Findings



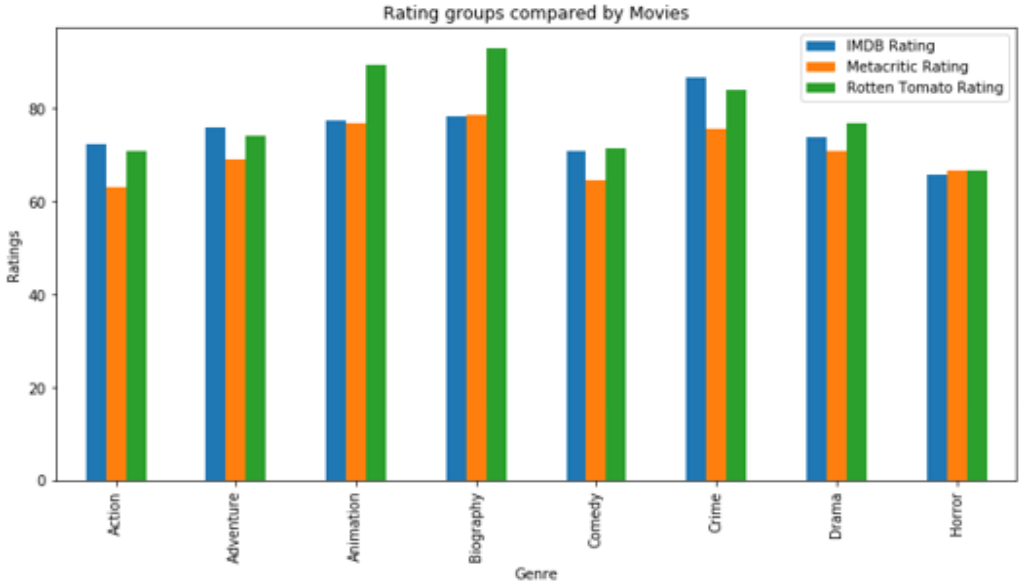
What type of movies are included in the generator? Are there any major differences in our ratings sources?

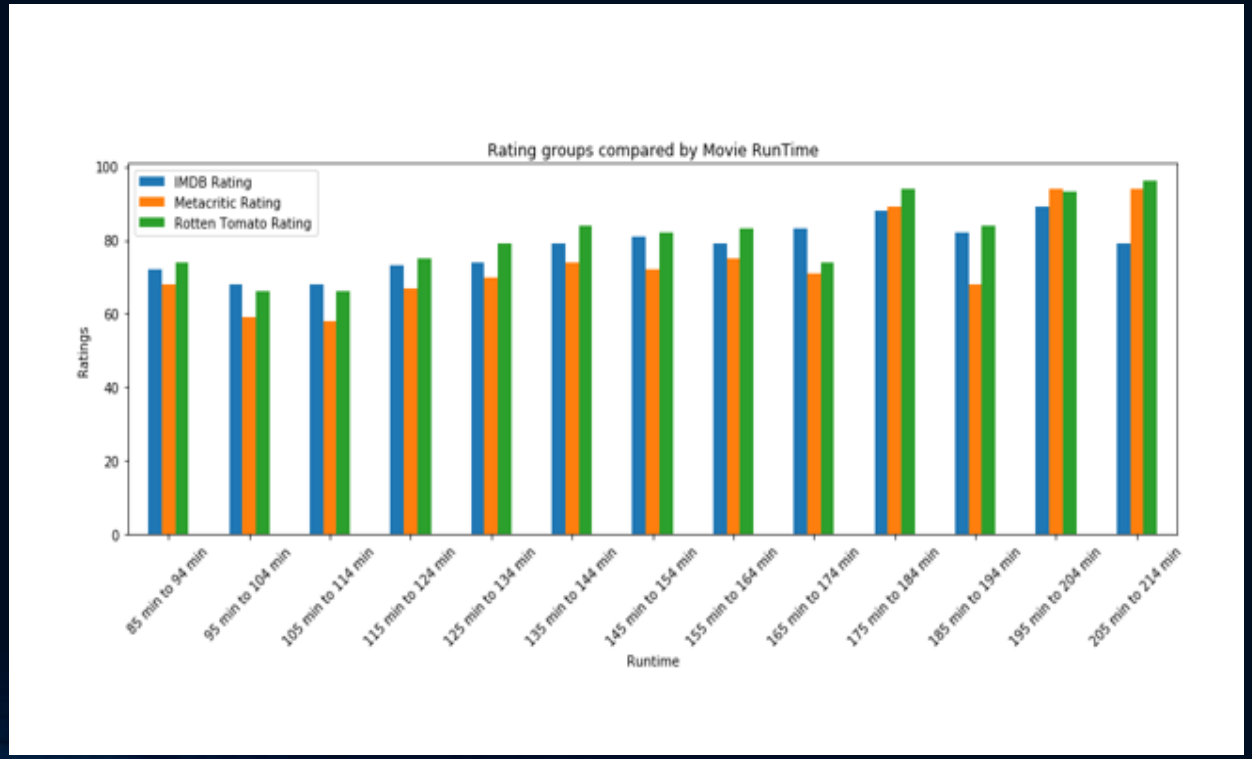
-Describe goal of models
-Pulled ratings from various sources; Why we chose those sources
-Good movie threshold; How we determined
-Findings



Is there any correlation between movie ratings based off directors, genre, and run time

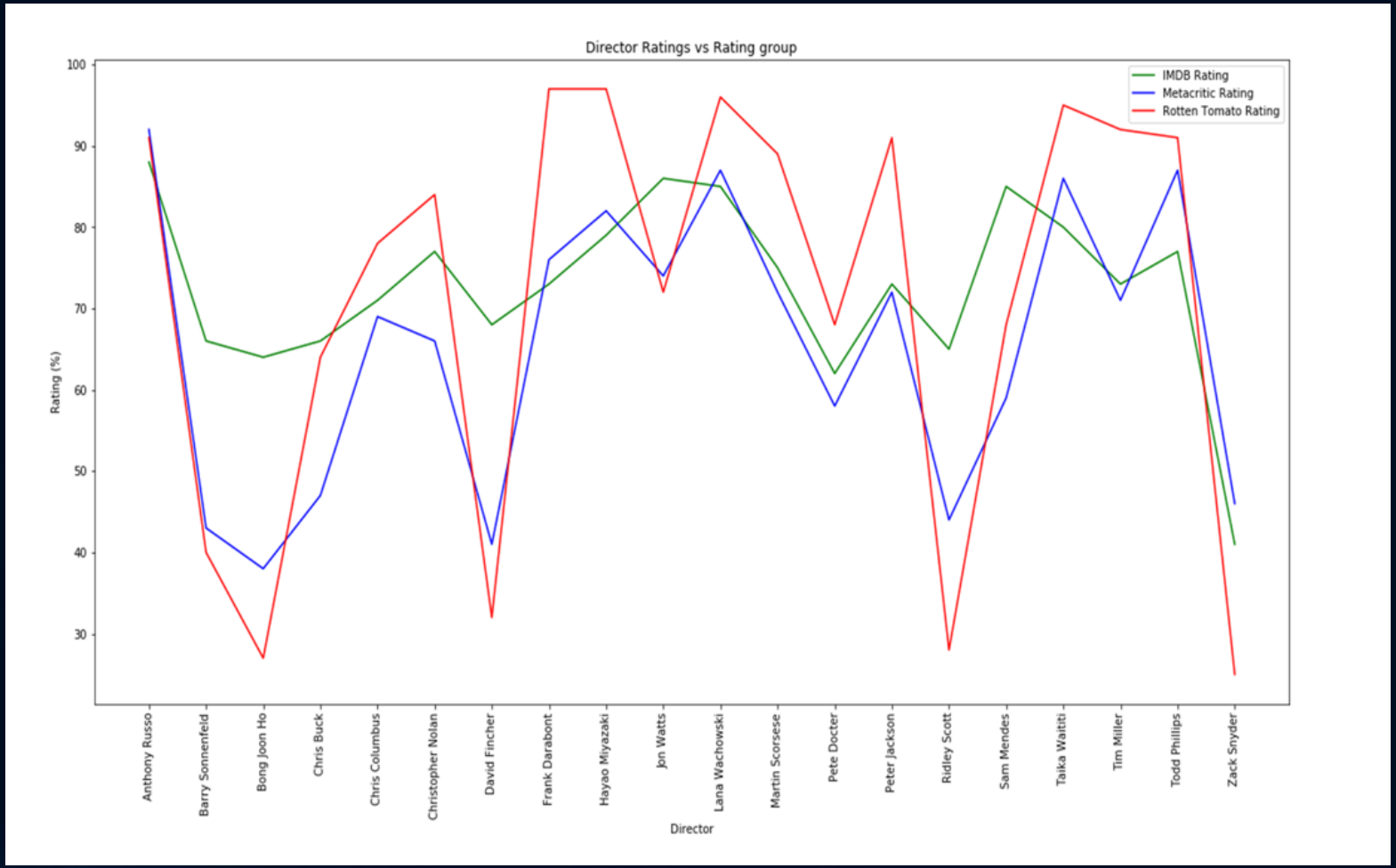
-Explain process for determining how we'd compare all three scores
-Formatting of all three ratings similarly
-Searched for patters in models
-Findings





Did Random Lists Provide Good Movies Based on Director?

- The movie directors had to be trimmed down to a more manageable number
- Directors with only 1 movie were filtered out
- Directors had to be above a 50 rating to be considered good
- 4 of the 20 director had very low average scores with Metacritic and Rotten Tomatoes . This may be due to their genre, but we were unable to correlate this
- Overall, if going by ratings from all 3 rating types 80 directors are considered good



Data Cleanup

- Exploration and cleanup
- Describe insights not anticipated:
- Problems that arose and how they were resolved
- ~~Discuss interesting figures we developed. Can use help of Jupyter notebook~~

Exploration and cleanup

- Exploration: Movie year, genre, director, ratings (Rotten Tomatoes, Metacritic, IMDB), and runtime were reviewed for correlations to see if the movies were considered good
- Clean up
  - Movie year, genre, and director had to be reduced in number due the number of values that were returned with low values, no values, N/A and none. This makes for cleaner and easier to read graphs
  - Director and genre returned multiple values in a single column of a database and had to be split and the first row from the split was kept for simplicity
  - Numeric values had to be converted to integers and floats because the were represented by strings in the data drawn from the OMDB website
  - When comparing director/ratings and genre/ratings, the mean of each rating type was calculated to keep the graphs a manageable size

Unanticipated Insight

According to Rotten Tomatoes and Metacritic, viewers do not like the horror genre

- Slide 7 show a major dip in ratings for Bong Joon Ho, David Fincher, Ridley Scott and Zack Snyder
- Due to the way the genres were split in clean up, this correlation would have been missed
- A quick internet search shows that all these directors are best known for horror movies

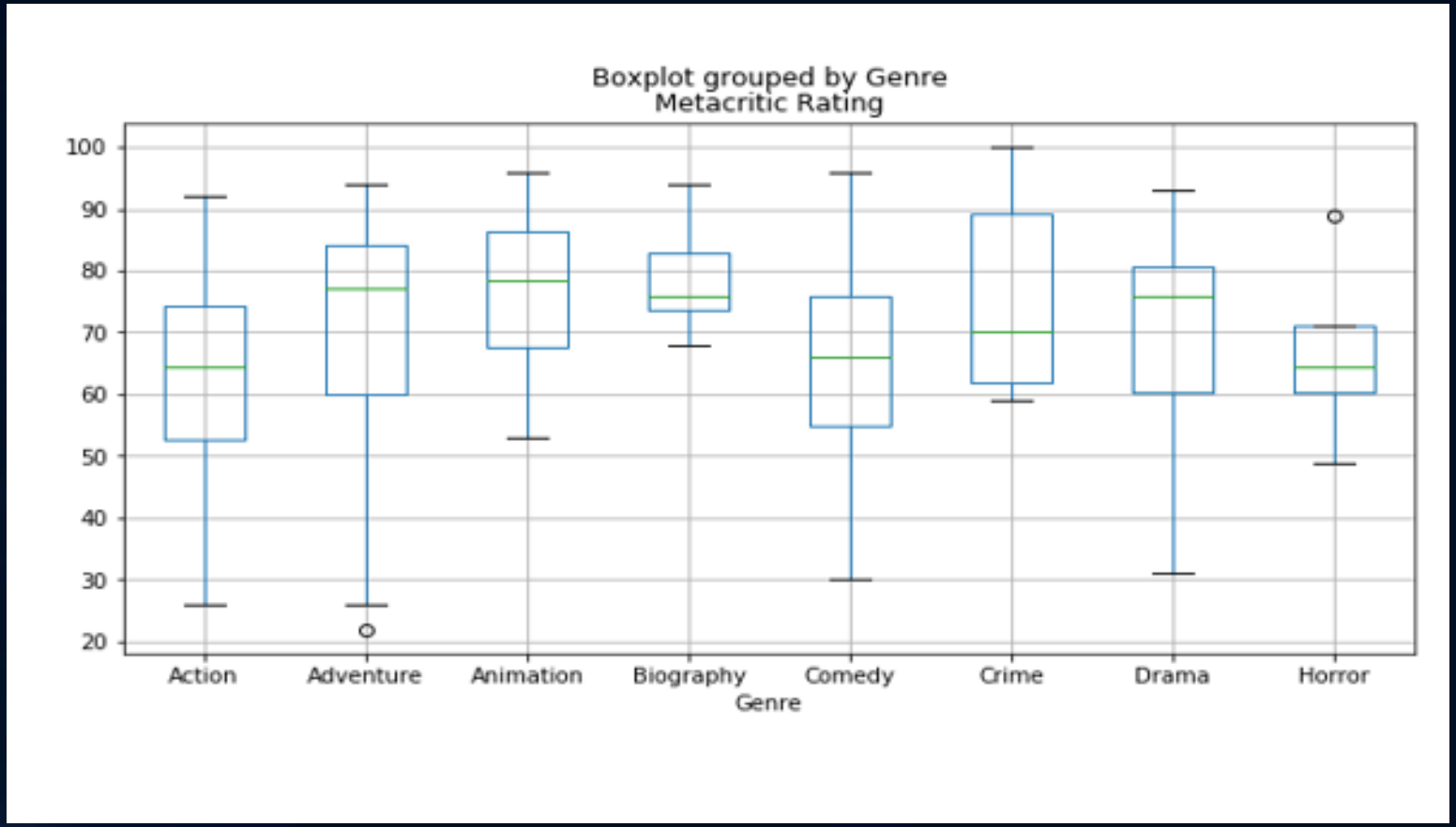
Problems

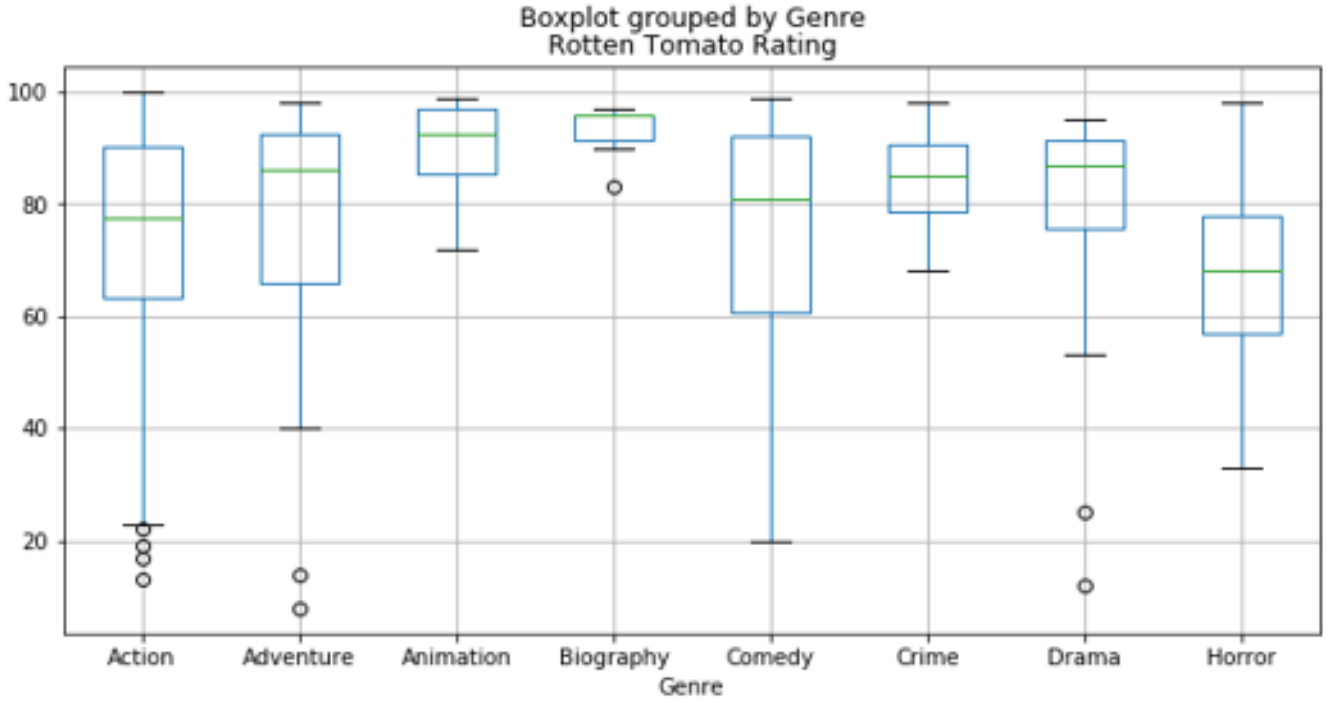
Most of the issue that were run into were coding issues:

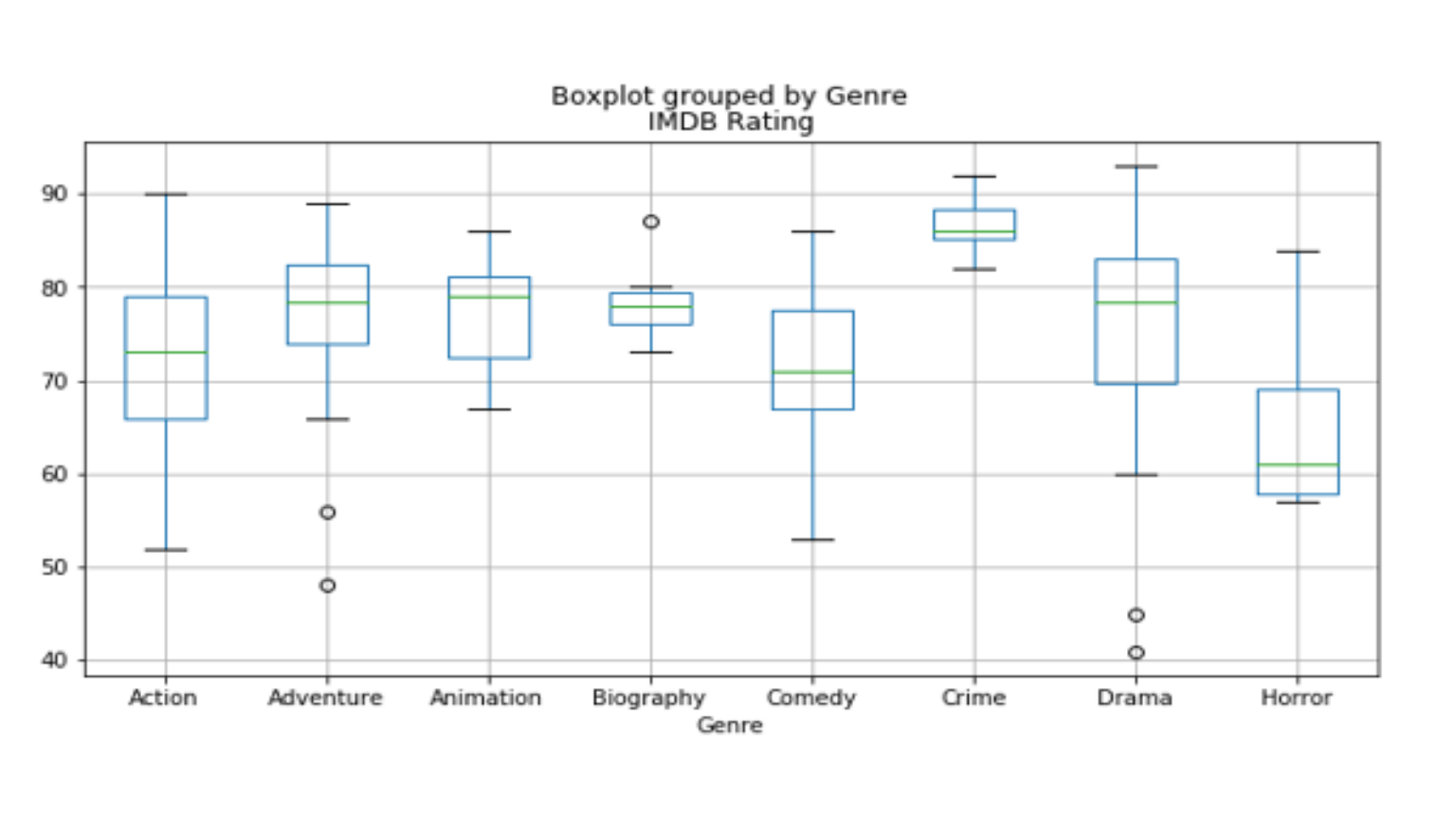
- Making sure that data values were properly converted to numbers from strings
- Using code in the correct format (ex: splitting columns)
- Making sure variables created were used in sequence and adjusted for in upstream and downstream code

Data Analysis

- Discuss the steps we took to analyze our data and answer each question
- Also use charts/figures here to help explain. Can put visualizations on add'l slides if needed


Boxplot grouped by Genre
Metacritic Rating
100
90
80
70
60
50
40
30
20
Action
Adventure
Animation
Biography
Comedy
Crime
Drama
Horror
Genre


Boxplot grouped by Genre
Rotten Tomato Rating
100
80
60
40
20
Action
Adventure
Animation
Biography
Comedy
Crime
Drama
Horror
Genre


Boxplot grouped by Genre
IMDB Rating
90
80
70
60
50
40
Action
Adventure
Animation
Biography
Comedy
Crime
Drama
Horror
Genre

Conclusion/Q&A

-Discuss our findings. Did we find what we expected? If not, why?
What conclusions can we draw from analysis?

-Describe any difficulties, and how we resolved.
What other questions would we ask, but couldn't with data?
Also, what would we ask if we had more time?
(Ex: 2 add'l weeks)

-Q/A