

Project: Visualizing Movie Data

Complete each section. When you are ready, save your file as a PDF document and submit it [here](#).

Step 1: Data Cleanup and Attribute Selection

- Clean up any missing information and choose the most important attributes you will explore further in your visualizations.
- List out the attributes (or variables) you plan to dive further with your visualizations. You should explore no more than 8 attributes.
- Please refer back to the [Data Cleanup course](#) to help you clean up your data.

Data Cleanup:

- A. Imported data into Alteryx and applied auto field to optimize field data types.
- B. Updated datatype for budget_adj and revenue_adj fields to double.
- C. Performed data investigation using tools: scatter plot and field summary and noted that for subset of movies, the fields budget_adj and revenue_adj are 0. This may potentially skew analysis e.g. average budgeted revenue for movie in specific genre.
- D. Applied filter to remove movies where budget_adj and revenue_adj is 0.
- E. Based on the questions to be answered in Step 2, it appears that we will need Genre of the movie, release year, Production Companies (Universal or Paramount) and whether movie is based on a novel or not. The fields Genre and Production Companies are delimited by pipe (|).
- F. We can identify/create these fields as –
 1. genre: by using text to column fields to split the genres vertically [Text to Columns tool]
 2. production_company: by searching keyword universal and paramount to flag each movie as either Paramount or Universal as production company. [Formula Tool]
 3. novel_based: search keyword “on novel” in field keywords to flag each movie as “Y” (based on novel) or “N” (not based on novel)

Attribute Selection:

We need to select following fields for dataset that will answer following question:

1. original_title
2. popularity
3. genre
4. release_year
5. budget_adj
6. revenue_adj
7. production_company
8. novel_based

Step 2: Tableau Visualizations

- Please make sure you follow the [rubric](#) and include Tableau Dashboards, Stories, and the appropriate visualizations (small multiples, scatter plot, bar chart, etc..) your reviewer expects your visualizations to contain. Remember: You need one Dashboard for every question (Q1-Q4) and in addition, you also need one Story, pertaining to a question of your choosing.
- Attach your visualizations as Tableau Workbooks in a zip file along with this report.

Please find tableau dashboard published on Tableau public [here](#).

IMPORTANT: Please upload the workbooks to **Tableau Public** to allow reviewers to access your workbooks. Note that simply saving your file as a “.twbx” is not enough to allow all reviewers to access. [Instructions on how to do this](#).

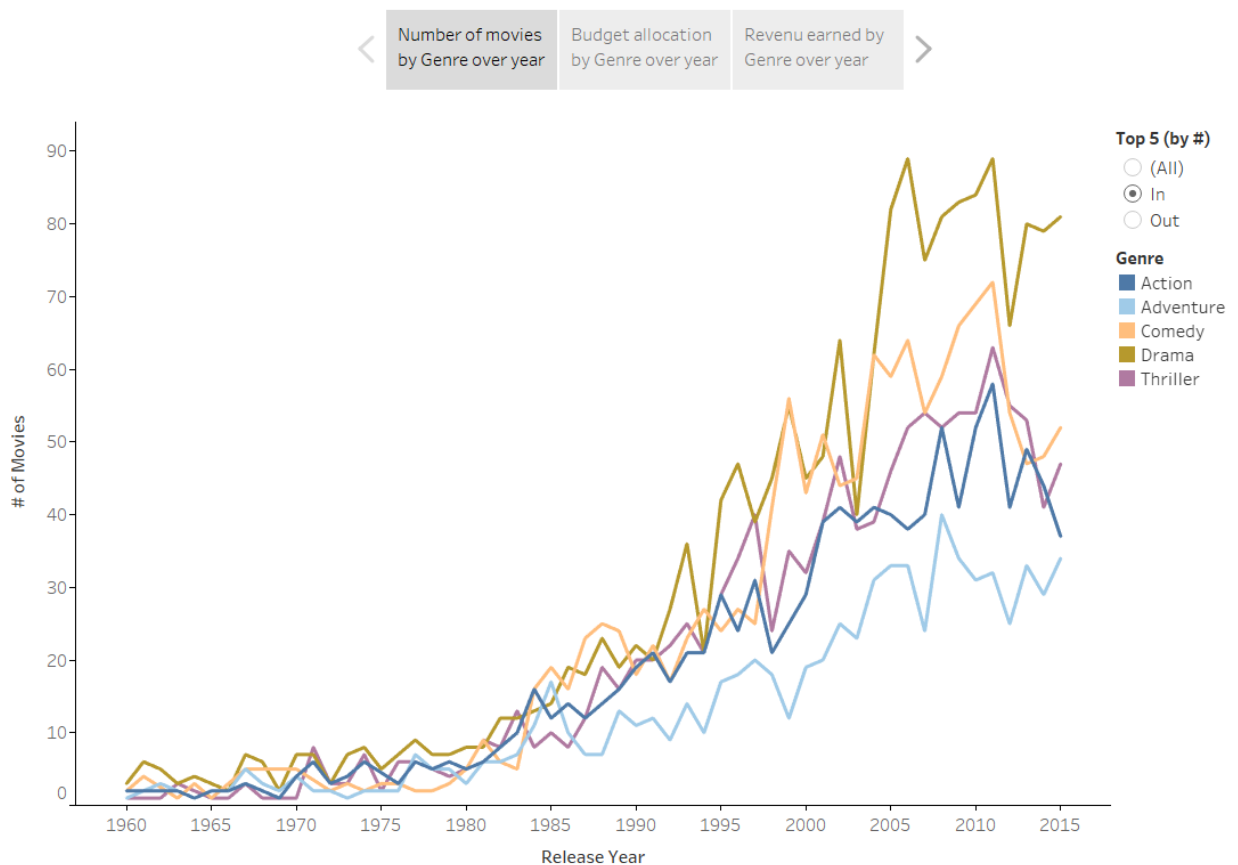
Step 3: Questions

- Answer the following questions. Refer to your online visualizations to back up your answers:
 - **Question 1:** How have movie genres changed over time?

of Movies over year:

- a) Based on visual below, the number of movies production increased dramatically after 1991. We can conclude that genres Drama, Comedy, Thriller, Action and Adventure are the top 5 genres with highest number of movies produced. The movies produced in genres Drama, Comedy, Thriller and Action were peak during year 2011. All the top 5 genres experienced a reduction in number of movies produced post 2011.

Q1: How have movie genres changed over time?

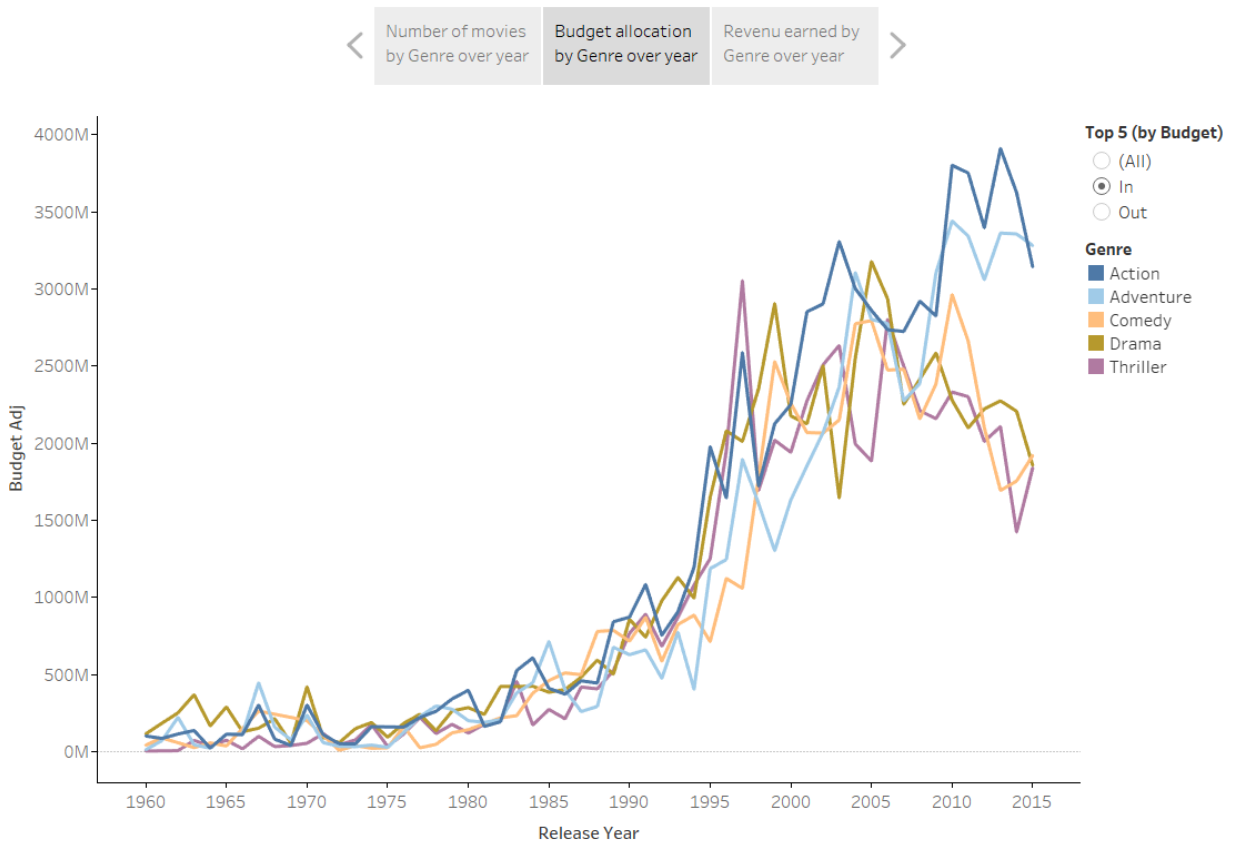


Allocated Budget by year:

As shown in the visual below, we can conclude that, movies in Action genres receives significant (higher) budget when compared to other genres. Comparing the # of movies v/s budget allocation over year, it is evident high number of movies were produced in Drama genre despite lesser budget allocation.

Additionally, we can clearly notice that, after year 2001, Action movies received higher budgets compared to other genres.

Q1: How have movie genres changed over time?



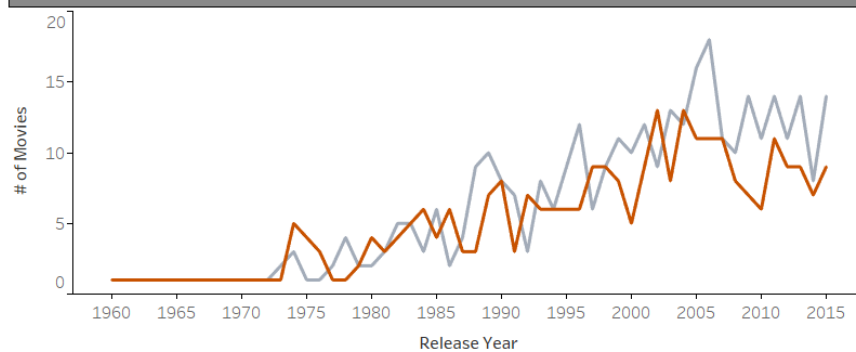
- **Question 2:** How do the attributes differ between Universal Pictures and Paramount Pictures?

Based on visual below, here are few observations (between Universal and Paramount as production house):

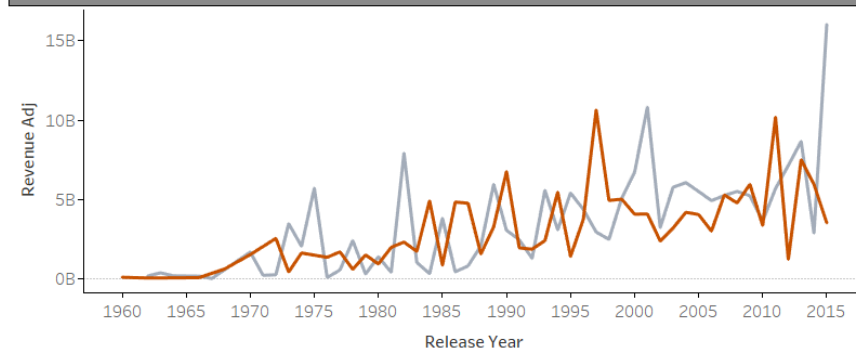
- By numbers, Universal Studios have produced higher number of movies (347 v/s 279) than Paramount Pictures
- Total revenue earned by Universal is ~ \$6B greater than Paramount
- Universal's revenue in 2015 was significantly high. Majority of this amount was contributed by the movie "Jurassic World"
- 3 out of top 5 movies by popularity were produced by Universal

Q2-How do the attributes differ between Universal Pictures and Paramount Pictures?

of Movies over Year



Revenue by Year



Production Company

Paramount

Universal

Key Statistics

	Paramount	Universal
Revenue Adj	\$157B	\$181B
Budget Adj	\$49B	\$55B
Profit	\$108B	\$126B
# of Movies	279	347

Movies by Popularity

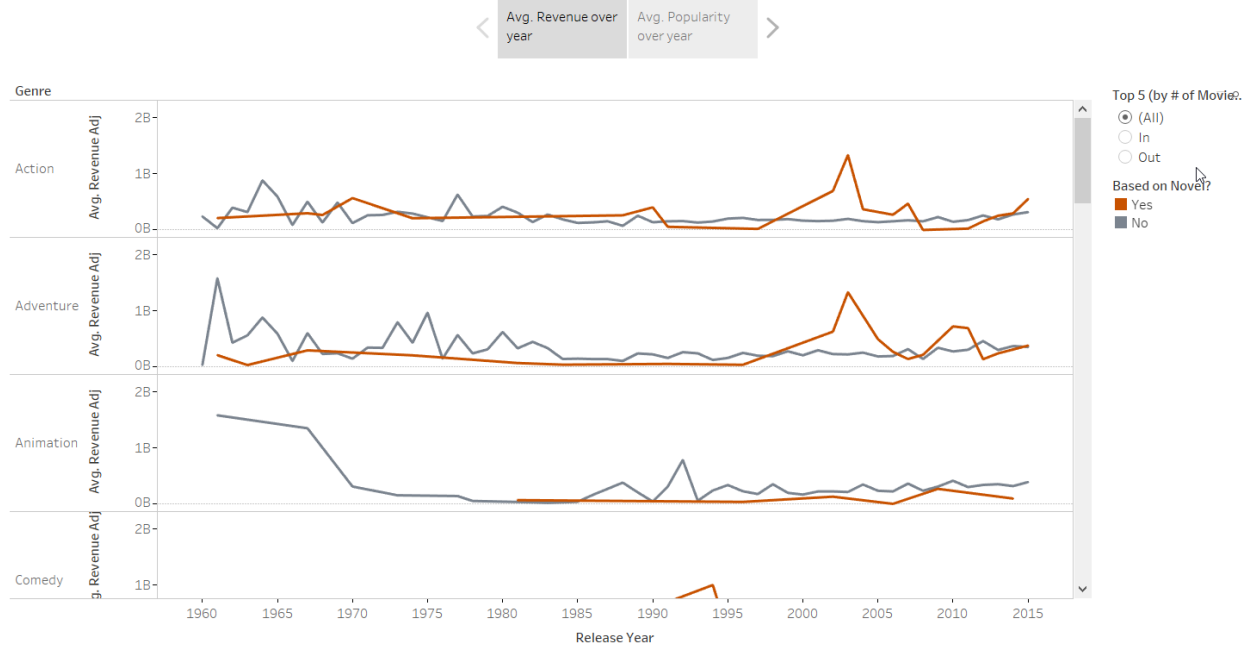
Jurassic World	131.9
Interstellar	74.8
Terminator Genisys	34.6
Minions	29.6
Teenage Mutant Nin..	28.9
Furious 7	28.0
Inglourious Basterds	24.4
Back to the Future	24.4
Forrest Gump	20.1
Oblivion	17.6
Dracula Untold	15.1
Transformers: Age o..	15.0
The Mummy	14.7
The Wolf of Wall Str..	14.6
The Bourne Identity	14.5
The Thing	13.5
Titanic	13.1
World War Z	13.0
Back to the Future P..	12.8
Gladiator	12.8
Blackhat	12.8
Scarface	12.7

- **Question 3:** How have movies based on novels performed relative to movies not based on novels?

Based on the visuals below, following are the observations associated with novel based/non-novel based movies:

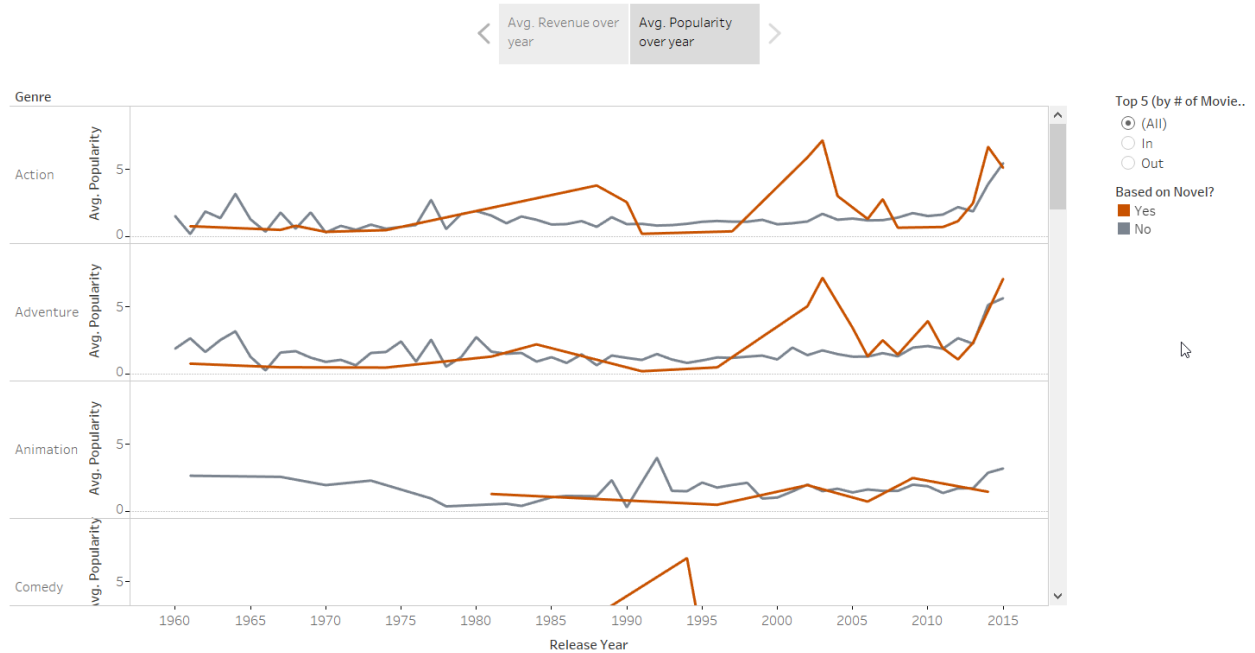
- Avg. revenue earned per movie is higher in case of movies based on novel. However, lower average revenue on non-novel based movies can be attributed to higher number of non-novel based movies produced. (**Please scroll** on the dashboard to observe trend of novel based movies earning more revenue over the year.)

Q3-How have movies based on novels performed relative to movies not based on novels?



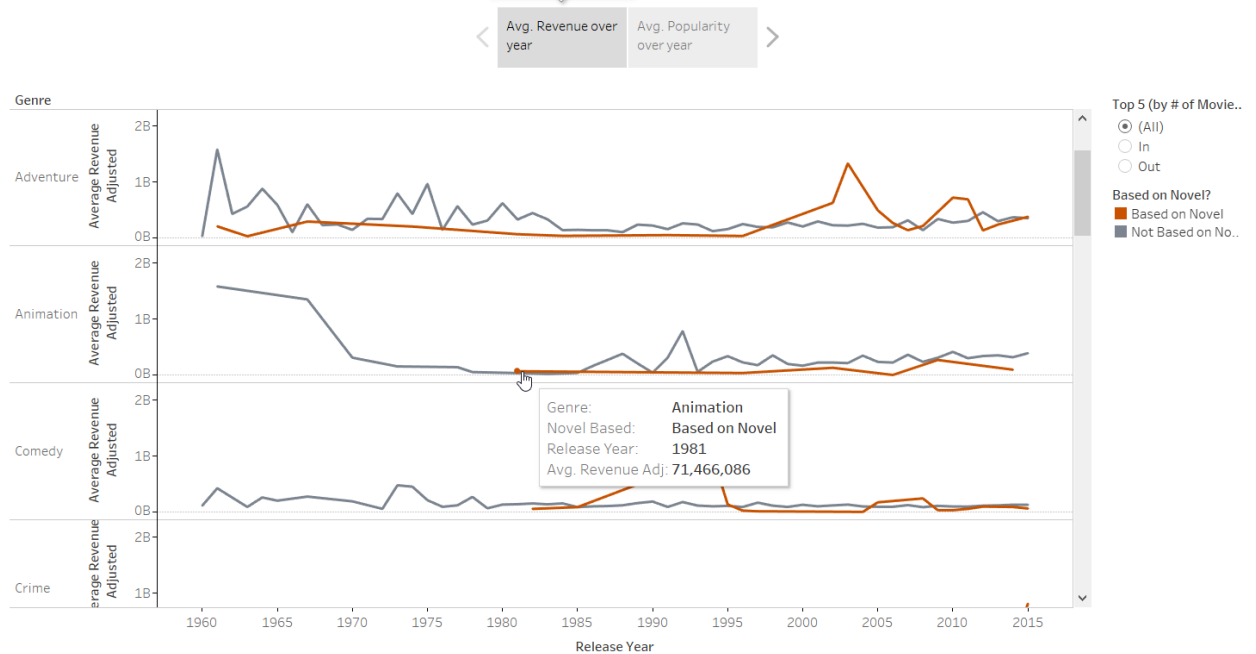
- Novel based movies appear to be more popular than non-novel based movies. **Please scroll** on the graphic or visualization to observe how novel based movies gained higher average popularity over period.

Q3-How have movies based on novels performed relative to movies not based on novels?



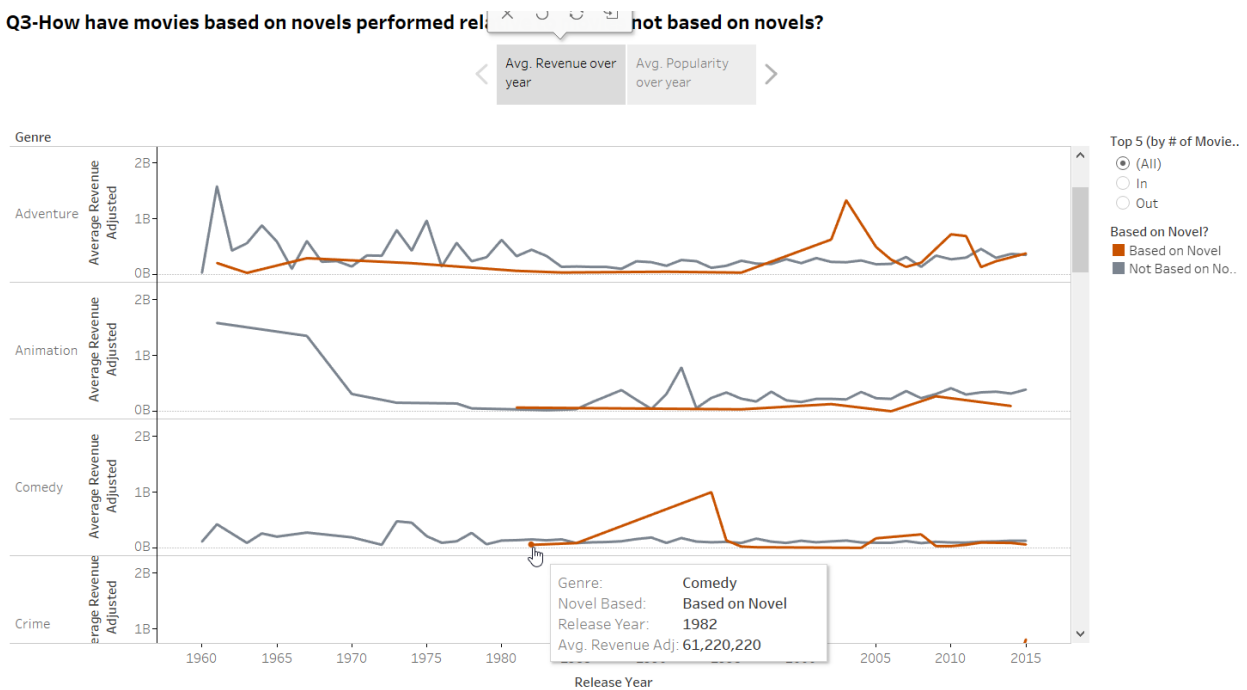
- As per the data, the world did not experience a novel based Animation movie until 1981. **If you scroll down** a bit to observe trend for Animation genre, we can conclude that first Novel Based animation movie was produced in 1981.

Q3-How have movies based on novels performed relative to movies not based on novels?



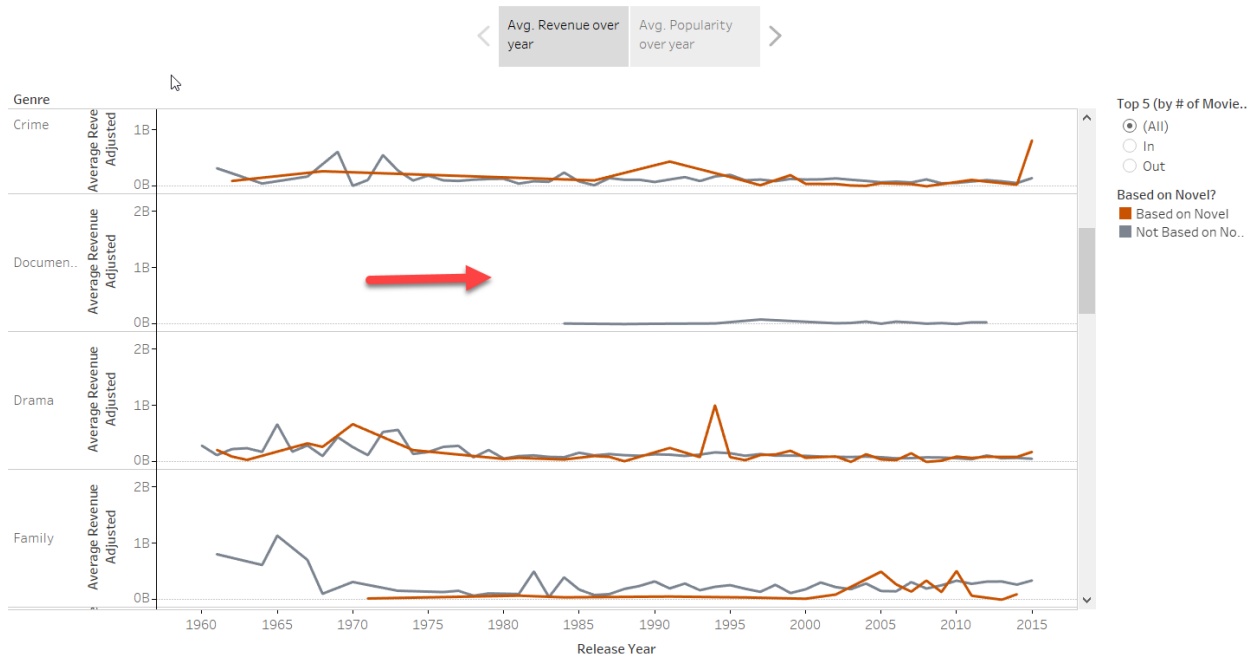
- Similarly, 1st novel based Comedy movie was produced in 1982. (**Please scroll** to Comedy genre)

Q3-How have movies based on novels performed relative to movies not based on novels?

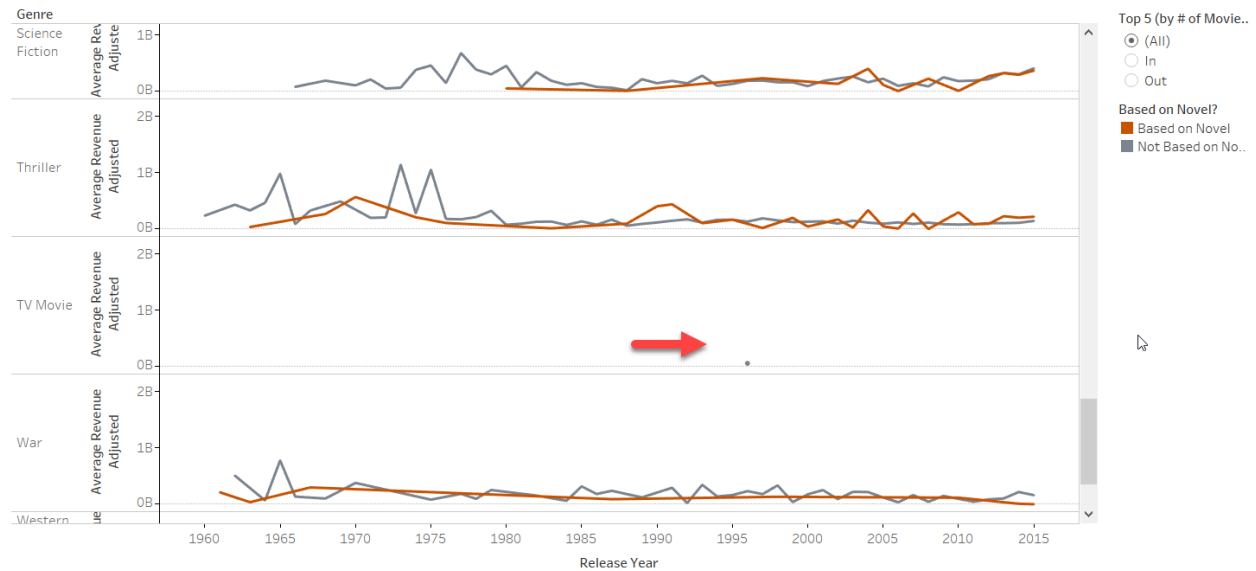


- Documentary and TV Movie are the only genres where a novel based movie does not exist. (As they are based on facts rather than imagination?). Please scroll to genre Documentary and TV Movie.

Q3-How have movies based on novels performed relative to movies not based on novels?



Avg. Revenue over year Avg. Popularity over year



- What is your additional question that you proposed? What is the answer? How did you come up with this question?

Q4 – What is the correlation between a) budget and revenue, b) profit and budget/revenue and c) popularity and revenue?

Using scatterplot on various measures, following are the observations related to correlations:

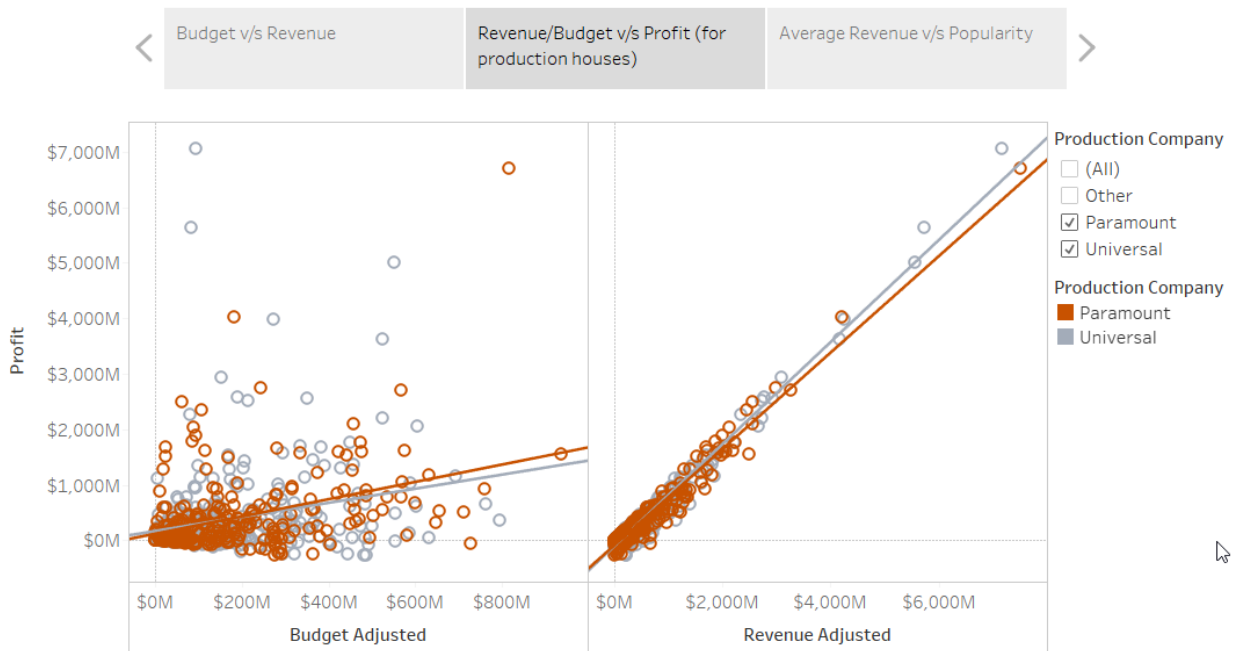
- Budget and Revenue has strong positive correlation. Higher budget movies will typically yield higher revenue. However, there are a few outliers in the data such as “The Worriers Way” which despite huge budget did not yield high revenue

Q4 – What is the correlation between a) budget and revenue, b) profit and budget/revenue and c) popularity and revenue?



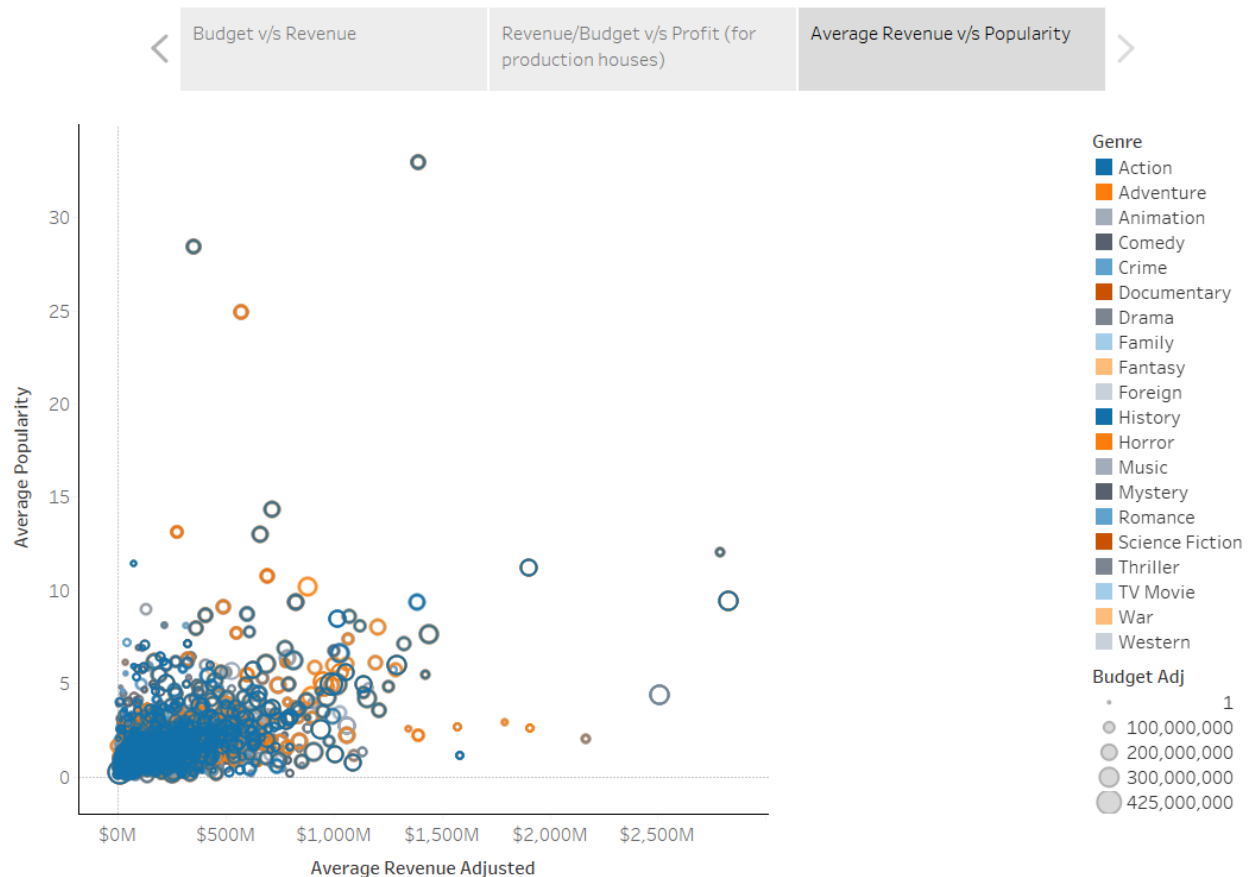
- Universal movies appear to be more profitable than Paramount movies. There is strong positive correlation between Budget and Profit.

Q4 – What is the correlation between a) budget and revenue, b) profit and budget/revenue and c) popularity and revenue?



- Typically high revenue yielding movies, tend to be more popular.

Q4 – What is the correlation between a) budget and revenue, b) profit and budget/revenue and c) popularity and revenue?



Modifications made to Tableau in version 2:

- Added screenshot for all the observations in Q3
- Bar chart was already used in dashboard Q2-Paramount / Universal
- Added size dimension to story Q4-Correlation Analysis to Average Revenue v/s Property
- Updated labels to spell out abbreviations like Avg. or Adj.