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.

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.

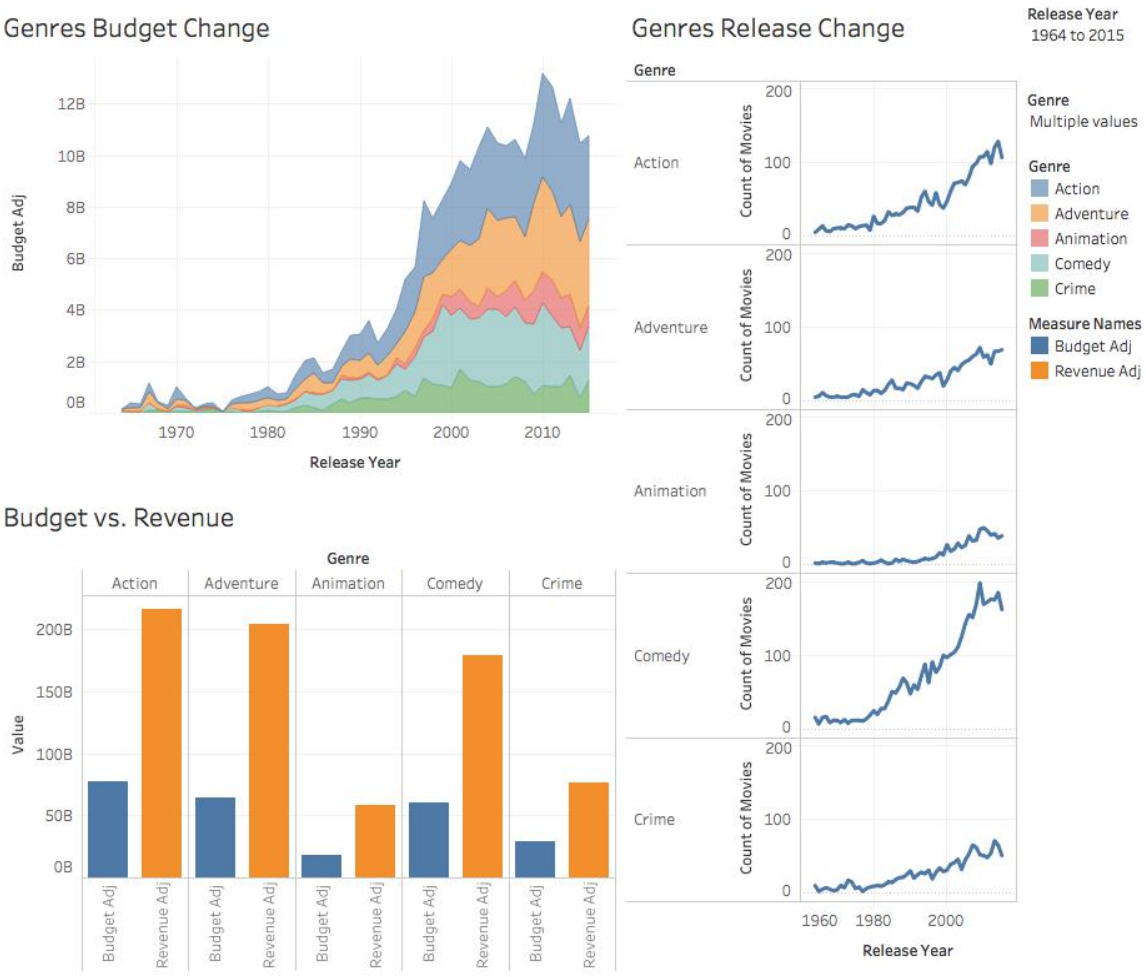
IMPORTANT: Please upload the workbooks to **Tableau Public** to allow reviewers to access your workbooks. Note that simply saving your file as a ".twbx" is not enough to allow all reviewers to access. Instructions on how to do this.

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?

For question 1 data, I cleaned it up in Alteryx by parsing all the genres, transposing, filtering out the null data in genres, and then selecting only seven relevant columns and renaming some of them. I did not filter out the nulls in adjusted budget and adjusted revenue columns because I only need a general trend of how different genres changes over time. I did budget trend and revenue trend using sums. And when doing sum, especially after transposing, one movie could be calculated in several different genres, and then some nulls are not important or meaningful to filter out.

The attributes include id, popularity, release year, budget, revenue, and genre.



Take five different genres as examples to answer the question. The release number of all five genres, action, adventure, animation, comedy and crime are increasing during the years. Among them, action and comedy experienced a substantial increase since 1990. Comedy is the one with the most dramatic increase curse. Except adventure, all other four genres suffered a slightly decrease in the past a few years.

For budget invested, crime and animation are less than the other three genres. For all five genres, the budget increased substantially around the year of 2000. Among them, action and adventure enjoy the most abundant budget.

For each genre, to compare its budget and revenue, they all have positive profits. Action, adventure, and comedy have higher revenue. But action and adventure have better revenue/budget rate than comedy. For animation, though its revenue and budget looks lower than others, the revenue/budget rate is actually very good.

Link of Q1

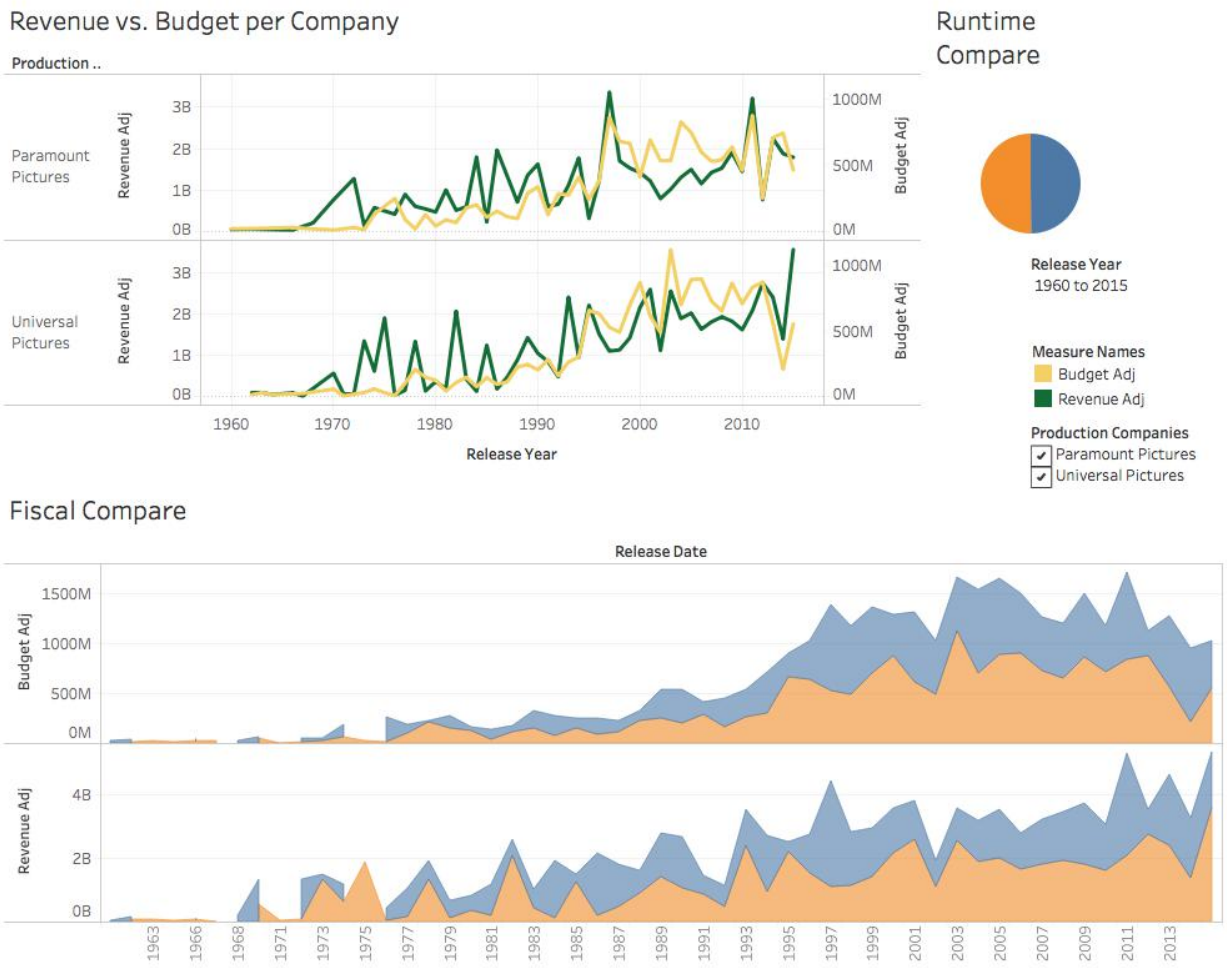
https://public.tableau.com/views/Q1_253/DashboardQ1?:embed=y&:display_count=yes&publish=yes

Link of Q1 Story
https://public.tableau.com/views/Q1_253/Q1story?:embed=y&:display_count=yes&publish=yes

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

For question 2 data, I cleaned it up in Alteryx by selecting only relevant attributes, parsing all the production companies, transposing, filtering out the null data in production companies, and then further filtering out 0 value in budget column or revenue column. I have seven final attributes left. By filtering out 0 value in budget or revenue, 599 out of 953 rows left. Though it seems almost one third of the data was left behind, it makes sense. Since I want to compare attributes of two companies', only data with full information is useful. It is hard to imagine a movie without budge or revenue. Therefore, analysis based on no budget or revenue movies is not meaningful. The clean process is necessary.

The attributes include id, runtime, release date, release year, budget, revenue, and production companies.



Paramount Pictures and Universal Pictures have many similar aspects. For example, their average movie run time is almost the same. But at the same time, they have several different attributes. To compare the revenue and budget of each of them, they all experienced years of revenue higher than budget and years of revenue lower than budget. And then, both Paramount's revenue and budget went down in recent years. At the same time, both Universal's revenue and budget went up. And its revenue enjoyed a great lead compared to its budget.

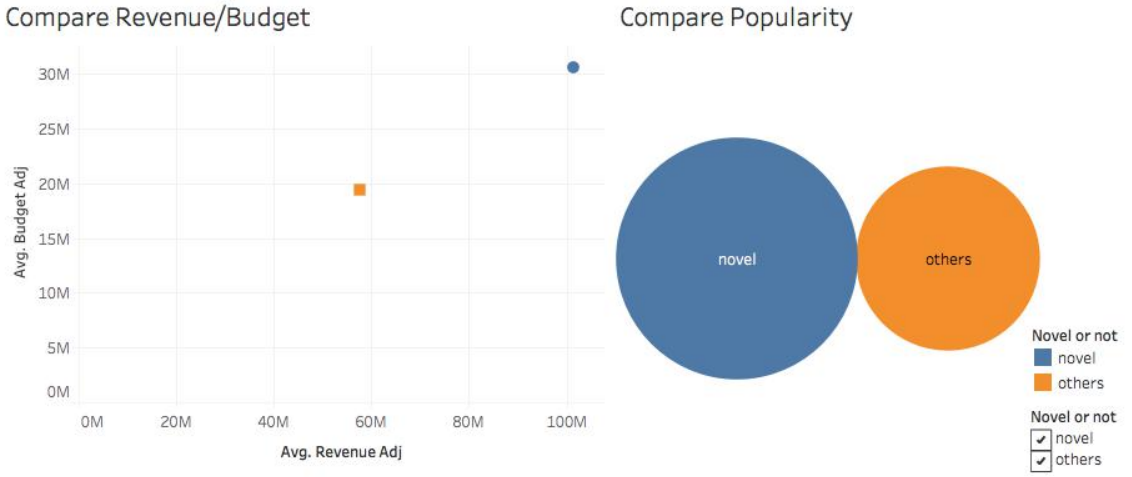
To compare revenue, Universal almost always enjoys more revenue than Paramount. And for budget, previously, Universal also has more budget than Paramount. But in recent years, Universal's budget dropped lower than Paramount.

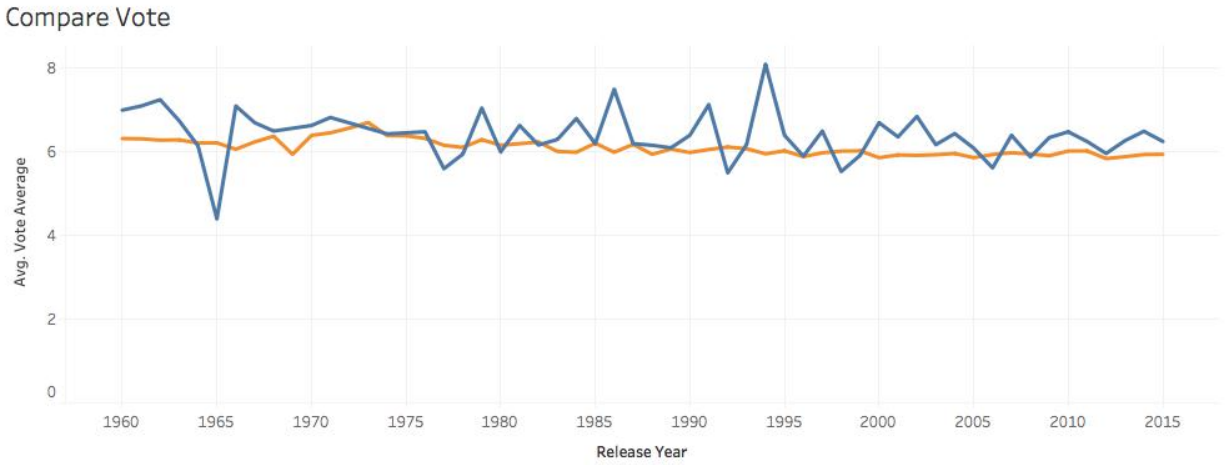
Link of Q2
https://public.tableau.com/views/Q2_225/Dashboard1?:embed=y&:display_count=yes&publish=yes

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

For question 3 data, I cleaned it up in Alteryx by selecting only relevant attributes, parsing keywords, transposing, using formula to mark “novel or not”, filtering out the ‘null’s, and then separating “novel” and “others” by filtering. I further used “join” to find out those who have both novel and others. I summarized ones only with “novel”, ones with both “novel” and “others”, and ones only with “others”. I did the summarize to eliminate duplicates in the three ones. And then, I used union to combine all the three. Now we have single list of movies marked with “novel” or “others”, but no other keywords.

The attributes include id, popularity, average vote, release year, budget, revenue, and novel or not.





It is obviously, that the average popularity of movies based on novels is better than those not based on novels. At the same time, the movies based on novels always have more budget and more revenue than their counterparts. But the rate of revenue and budget for the two kinds are similar.

The average votes for movies not based on novels are basically not changed over the years. At the same time, movies based on novels changed a lot. The average votes of movies based on novels were much lower than those of others around 1965. There were other four time periods, that the average votes of movies based on novels were slightly lower than those of others. All other time, the average votes of movies based on novels were higher. The highest point is around 1995.

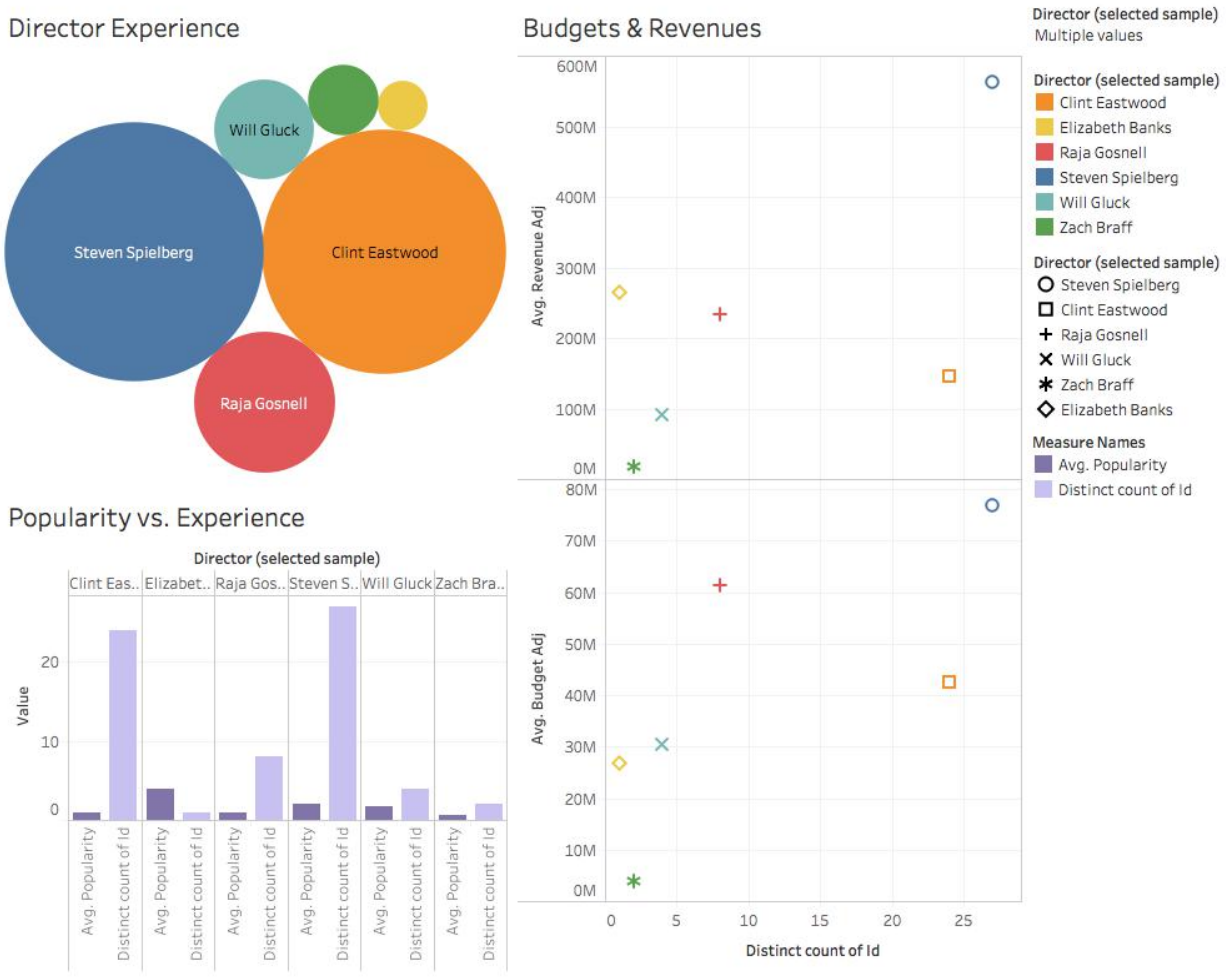
Link of Q3
https://public.tableau.com/views/Q3_213/Dashboard1?:embed=y&:display_count=yes&publish=yes

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

How have movies directed by directors with different experience performed?

For question 3 data, I cleaned it up in Alteryx by selecting only relevant attributes, filtering out the ‘null’s in director column and zero content in budget and revenue.

The attributes include id, popularity, director, average vote, release year, budget, and revenue.



I first listed all directors in order or their number of released movies. And then I picked up six sample directors with different movie release number as a sample group.

For budget and revenue, basically, the more experience directors enjoy more revenue and budget. But the ratio is not so stable. For example, the second experienced director Clint Eastwood has a lower average budget and average revenue compared to the third experienced director Raja Gosnell. Further, directors with only one movie may have data statistically not so meaningful.

Then I compared the experience and popularity of directors. Seems there is no strong relationship between one's experience and average popularity.

Q4 Link
https://public.tableau.com/shared/Y37Q5N384?:display_count=yes