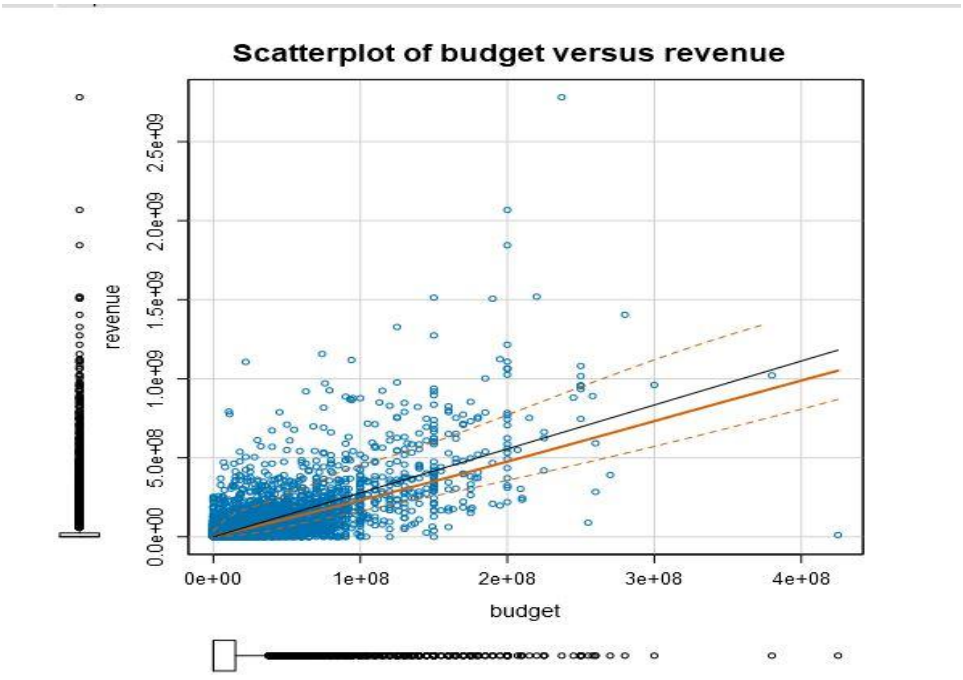
Project: Visualizing Movie Data

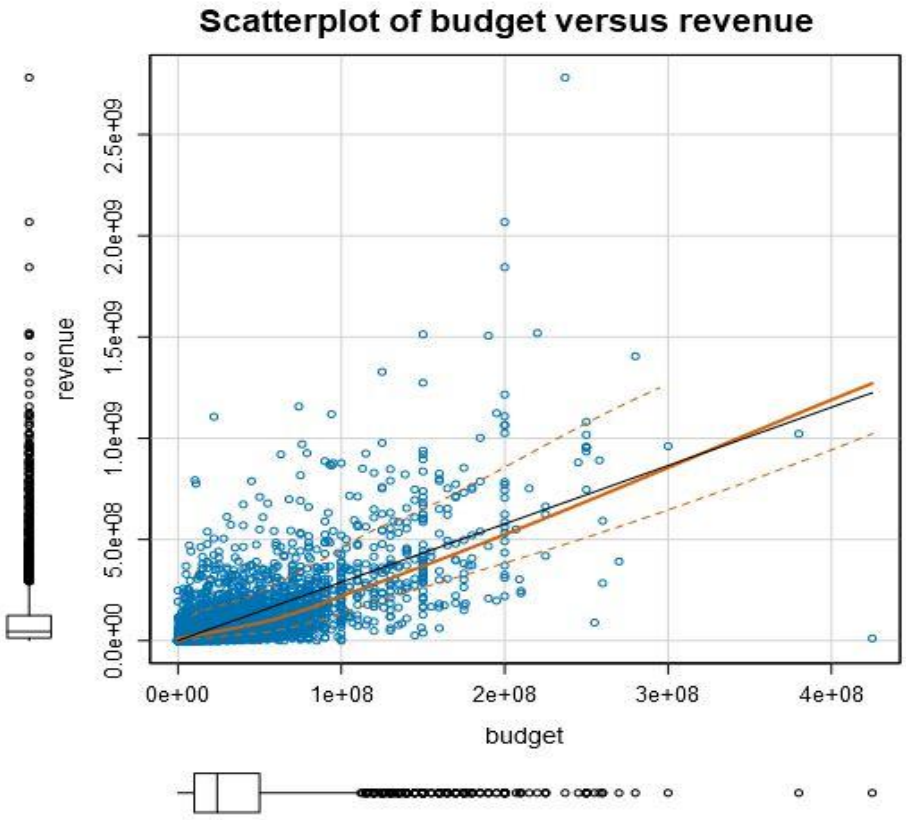
Step 1: Data Cleanup and Attribute Selection

- Clean up any missing information and choose the most important attributes you will explore further in your visualizations.
  The most important attributes that can be used to answer the questions below and my question as well are – popularity, budget, revenue, cast, genre, release date, vote average, production companies and keywords.

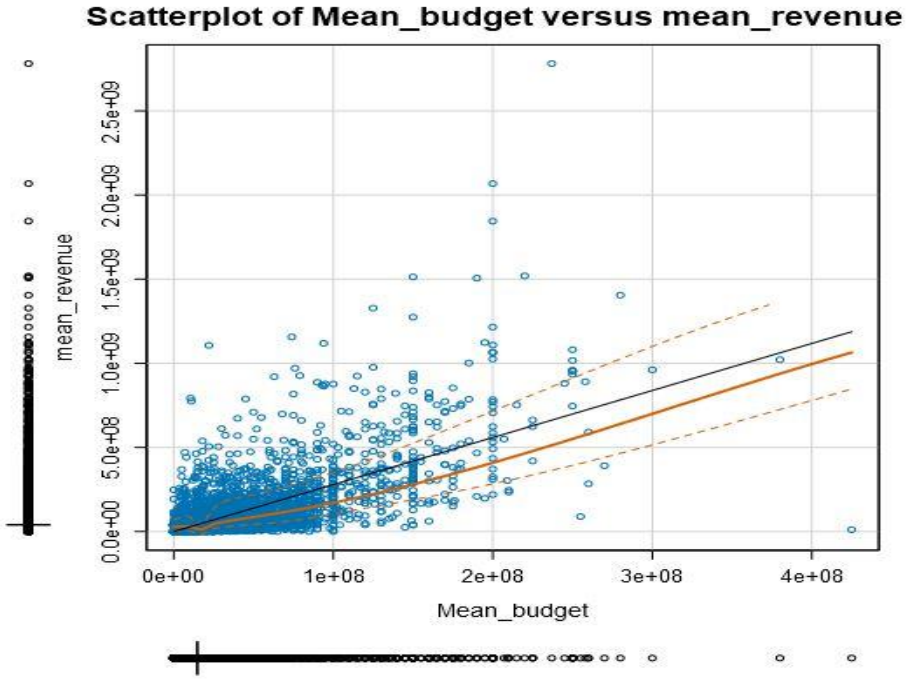
  Since we wanted budget and revenue in our data for further visualizations. There are over 4000 variables showing '0' as their value for budget and revenue which we know is not possible so deleting them is not an option. Replacing them with mean or median or mode but the median and mode for those columns return a 0 so we stick to imputing mean. Upon taking the means of budget and revenue, we can see the scatterplot is almost the same and very few changes can be observed. On deleting the missing variables, we do see quite a change in the scatterplot.



  The above graph is just running the plot without any interference with data (includes the missing data)



The above plot is after we delete the missing variables and we can see a small change in the plot.
The below plot is the one after imputing mean in place of 0s for the missing data.
The graph is almost similar to the original un-interfered data.



- List out the attributes (or variables) you plan to dive further with your visualizations. You should explore no more than 8 attributes.
  The 8 variables that are selected are - popularity, Mean budget(mean imputed values), Mean revenue(mean imputed values), genre, release date, vote average, production companies and keywords.

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.

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?

On observing the **Movie genre trend line** worksheet, we can see that the over the years- *Drama* and *Comedy* have heavily dominated the movie industry. They are more likely to receive a suitable budget to make movies for these genres, followed by *Documentary* and *Horror* movies.
Moving on to the **Bar Graph Trend** (*popularity*)Worksheet gives us a better understanding of which genres received an appropriate popularity rating. *Action|Adventure|Sciencefiction* and *Thriller* have always topped the list in the recent years from the year 2000. Drama and *Drama|Romance* can be seen in the bottom of the list.
A pie chart demonstration for an expanded understanding of how the break up for the movie genres were with respect to popularity and Vote average for the respective genres.
Budget vs Revenue with population variable shows the most popular movie genres and it shows an expected trend where the more revenue a genre collects, it has the highest popularity.

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

The **first sheet** depicts the differences between two Production companies (*Paramount* and *Universal pictures)* in terms of their total budget spent on movies and their respective total revenues collected from them. There is a slight increase in *Universal pictures* in terms of budget and revenue collected as compared to that of *Paramount pictures.* Safe to assume that *Universal pictures* are doing a little better than *Paramount Pictures.*
The **Popularity Trend** Worksheet tells us that there is a steady increase in popularity over the years, but we can spot the main differences from the year 2000. Universal pictures were doing a little better in terms of Popularity. In the year 2014 Paramount Pictures crossed a better rating that universal pictures but they dropped down again in 2015.

The **last worksheet** is based on the first worksheet where it is plotted against Revenue and Budget but in this case, it is with respect to time.

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

The **average runtime trend** worksheet - The average runtime for movies based on novels has always fluctuated but we can see a steady graph for the movies that are not based on novels. The budget and revenue data information for the movies based on novels and movies not based on novels are very similar and show an expected trend.
The **Bar graph** shows the average popularity rating between movies that are based on novels and that which are not. We can see that there is a higher popularity rating for movies that are based on novels.

- What is your additional question that you proposed? What is the answer? How did you come up with this question?
  The additional question I was keen in finding out was how well the top actors and actresses performed compared to each other with respect to popularity (over time), their voted average and how well the companies have performed (considering revenue and budget).
  The answer upon seeing the visualizations is a hard one to make as the actors don't show a clear trend on who's performing better. But, we can conclusively say that actor Tom Hardy is an upcoming actor and over the years has collected a large voted average. In the current standings in 2015, Tom hardy among the actors that we chose has the highest popularity.

  https://public.tableau.com/profile/ mohammadnadeem
  Profile name : mohammadnadeem