The length of a film should be causally related to the endurance of the Human Bladder

Movie Trend Analysis

Lab: 05-P1

Tutor: Farah Kabir

Manmeet Singh: 30749476

Table of Contents

Introduction **2**

Design **2**

SHEET 1: **2**

SHEET 2: **3**

SHEET 3: **3**

SHEET 4: **3**

SHEET 5: **3**

Implementation **4**

User Guide **5**

Conclusion **10**

Bibliography **11**

Appendix **12**

Introduction

Film industry has become a major part of our life. In 20th century no one would have believed that film entertainment will be a billion dollars industry one day. Investment by major film production companies say a lot. Let us take Walt Disney, they have acquired production companies such as Marvel Studios, Pixar, 20th Century Studios and many more. So, it is important to analyse how film industry has changed over the period.

Every generation has their own taste and preferences. It is important for different production companies to analyse their audience to generate maximum revenue. Based on past data they want to understand the current trend and genre people prefer.

We will analyse the how popularity of different genre has changed over the period, the revenue they have generated and their budget. How production companies have competed in terms of movie released, budget and revenue they have generated. We also look at the movies based on novel and how they have performed compare to movies not based on novel. Which country tops the list with most production and which genre is more preferred in that country.

The intended audiences for this project are production companies, directors, media-services provider such as Amazon Prime, Hotstar, and Netflix and online YouTube channels. Also, those who focus mainly on short films to send a message to the society.

Design

Design planning for visualisation start with detail analysis of the dataset. Determining the content of dataset and what to be visualised is the first step in this process. Next step following the analysis of data is to find the visual options. What kind of graph to be used, whether it should be interactive, and convey your message in an interactive format? For this purpose, we use Five Sheet Design Methodology. It will help us consider different designs and help us finalize the design which goes best with the dataset and the information to be conveyed to the intended audience.

SHEET 1:

This is the Brainstorm section or the Ideation where we randomly select all the potential graphs and design which suits our dataset. Such as bar and line graph, interactive graphs and informative. In this section we filter, categorize, combine & refine, and put questions. Graphs that are chosen initially for the project are:

i. Pie chart to visualize different types of genre. Which will provide detail regarding their proportion accordingly.
ii. Bar graphs to count of number of movies released by different countries and based on that plot bar graph of genre for that country.
iii. Line graph to show relation between different production countries and movies that are based on Novel or not.

Sheets 2,3,4 are alternative designs. These three design sheets enable the user to record detailed alternative concepts. The basic idea of these three sheets is to provide Information, Layout, Focus/Parti, and Discussion.

SHEET 2:

Sheet 2 is the part of Initial design. The sheet 2 short list the preferences. Which helps in simplifying the decisions regarding to the visualisation designs. Line graph was finalised in this section, best suit for projecting count of number of Novel vs Non-Novel movies released over the years. Choosing between Novel vs Non-Novel movies will provide bar and line graph of number of movies released and revenue generated over the years, respectively.

Alternatives considered: The idea of line chart is changed, as the count difference between Novel vs Non-Novel movies was too big, I had to drop the line chart as it was not conveying much of a information, instead I kept individual bar and line graph for movies count and revenue generated respectively. Also, these are interactive graphs when you hover over the graphs it will provide the relevant information.

SHEET 3:

Sheet 3 is also part of Initial design. Pie chart was finalised in this phase. Pie chart was chosen because of different type of genres. As it represents very precisely the proportion of each and every genre in all the dataset. Then we select top five genres and compare how they had done over the years in term of revenue.

Alternatives considered: Initially pie chart was supposed to be non-interactive. But I have created an interactive pie chart which shoe their count and percentage, respectively. Then, instead of plotting only revenue line graph, I chose to show relation between revenue and budget of top five genres. Also, these are interactive graphs when you hover over the graphs it will provide the relevant information.

SHEET 4:

Sheet 3 is also part of Initial design. Bar graph to show total number of movies released by different countries and line graph was chosen for movies released by different production companies and represent comparison between top production companies. Based on the bar graph we have option to chose from the top ten countries that what genre type is preferred in that country.

Alternatives considered: Instead of line graph to show comparison between top production companies, I have created a bar chart of total number of movies released by these top five companies (as there are many companies I have decided to show top five) and a line chart to show relation between revenue and budget for these companies. Also, these are interactive graphs when you hover over the graphs it will provide the relevant information.

SHEET 5:

In the final section of Five Sheet Design Methodology, developer considers a final concept before delivery. It contains information to outline the final design, how can a user operate it (the application), main idea behind the visualisation, and detail to implement the design. The design on the final sheet could be the same as one developer has created on sheet 2,3, and 4 or it can be an combination of the concepts.

Alternatives considered: Changes that were made in sheets 2,3,4 are added in final design and so the explanations for the graphs will also changed accordingly. To create informative story

telling via visualisation, I have used tab format. Tab 1 display two graphs, pie chart and line graph. Tab 2 has bar and line graph. Tab 3 and 4 have bar graph and line graph, respectively.

Implementation

The libraries used in the application are:

1. Library (shiny): Shiny is an R package use to create interactive web apps. It runs the Shiny server having a web framework created for using web applications. Shiny lets you do analysis in an interactive web environment. It can be extend with CSS themes, htmlwidgets, and JavaScript actions. (https://shiny.rstudio.com/)

2. Library (reshape2): Flexibly restructure and aggregate data using just two functions: melt and 'dcast' (or 'acast'). (https://cran.r-project.org/web/packages/reshape2/index.html).

3. library (ggplot2): A system for 'declaratively' creating graphics, based on "The Grammar of Graphics". You provide the data, tell 'ggplot2' how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details. (https://cran.r-project.org/web/packages/ggplot2/index.html)

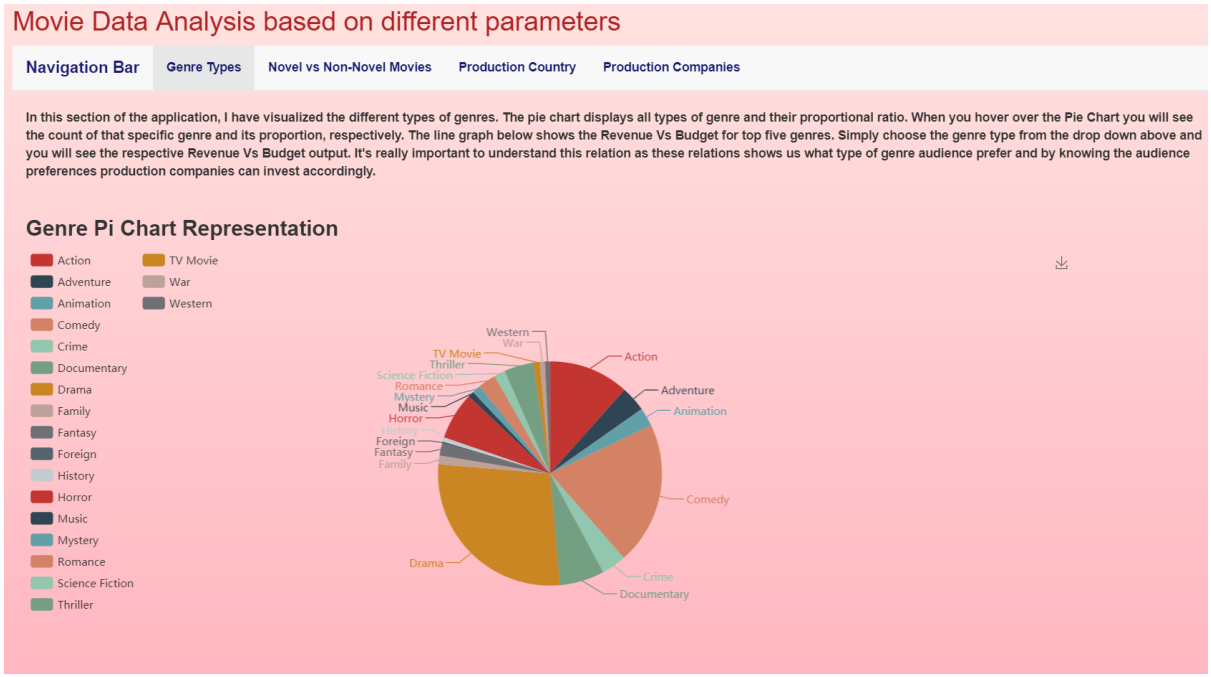
4. library (plotly): Create interactive web graphics from 'ggplot2' graphs and/or a custom interface to the (MIT-licensed) JavaScript library 'plotly.js' inspired by the grammar of graphics. (https://cran.r-project.org/web/packages/plotly/index.html)

5. library (ECharts2Shiny): Embed interactive charts to their Shiny applications. These charts will be generated by ECharts library (https://cran.r-project.org/web/packages/ECharts2Shiny/index.html)

6. library (shinyWidgets): Collection of custom input controls and user interface components for 'Shiny' applications. Give your applications a unique and colorful style. (https://cran.r-project.org/web/packages/shinyWidgets/index.html)

User Guide

When the intended audience runs the application, landing page is:

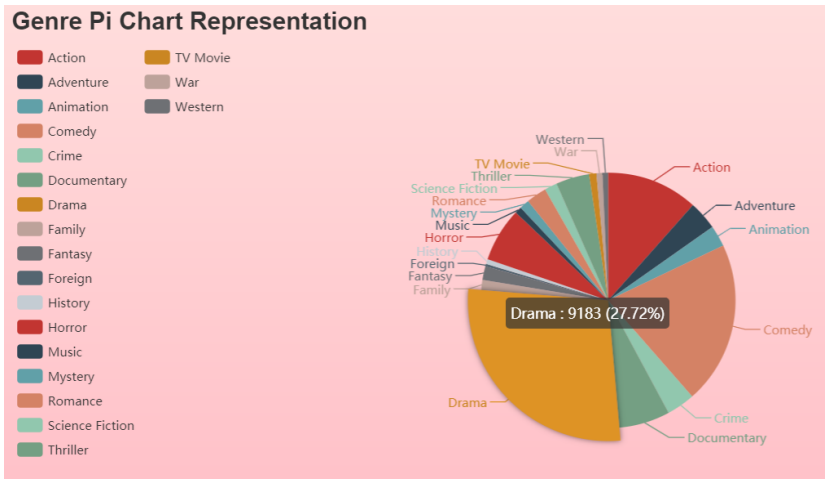


On the navigation bar there are four tabs, click on any the to enter that page.

Let us go step by step how to use this application for all the tabs.

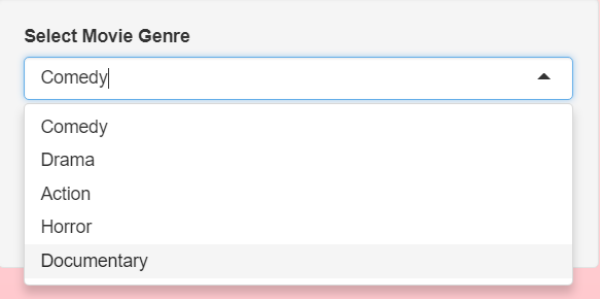
Genre Types:

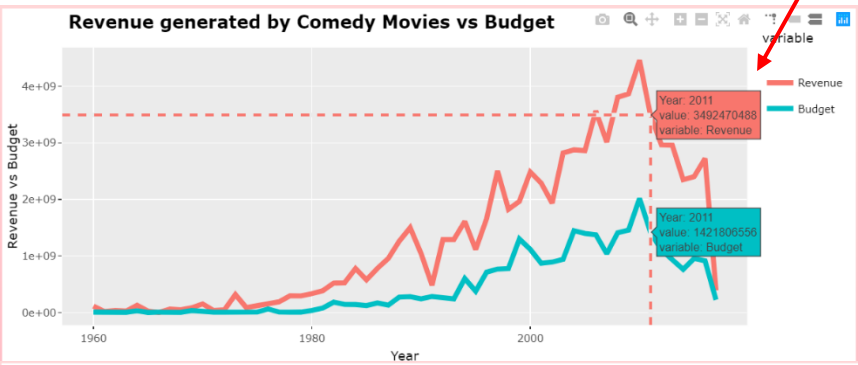
This the landing page having information regarding genre types. First graph is a pie chart,



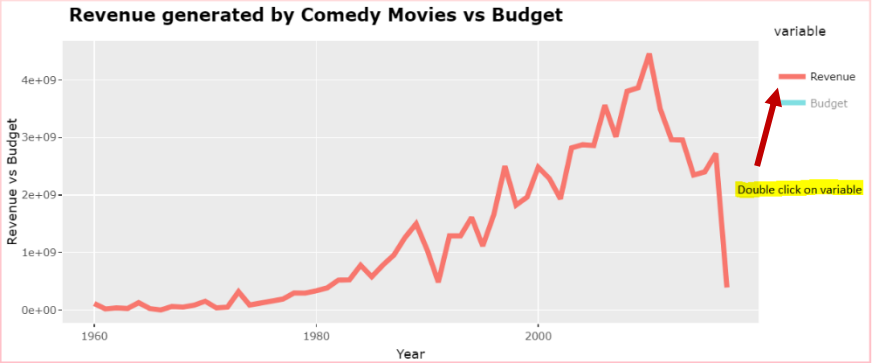
when you hover over a specific genre, we can see the genre count and its percentage. Here, for example when we hover over genre Drama, we can its count 9183 and proportion 27.72%.

Then scroll down the page where we have drop down menu. You can select any genre type from the drop down and line chart will generate depicting Revenue vs Budget relation for each genre type.



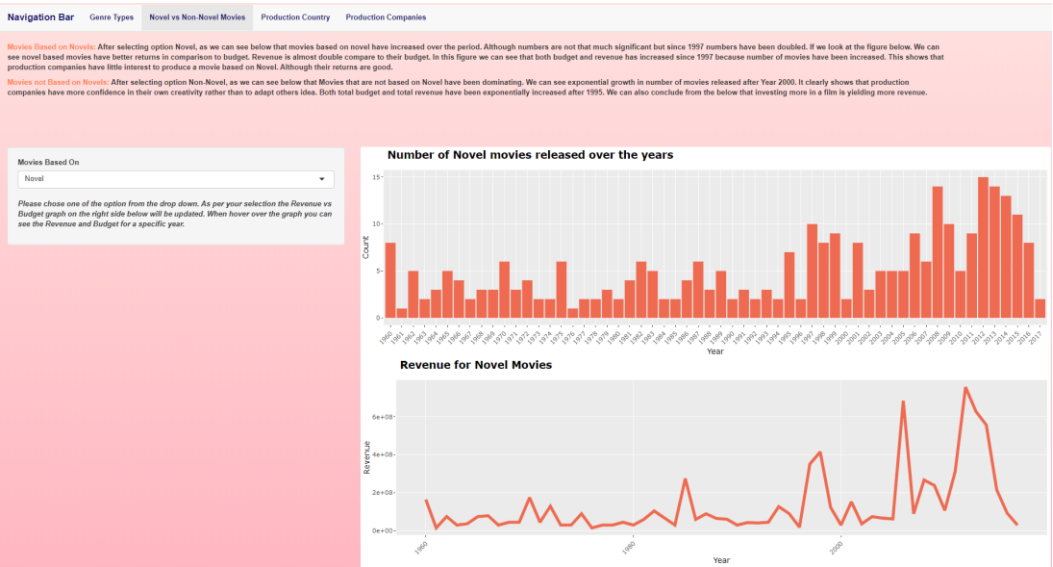


When you hover over the line chart you can see the respective value such as Year, value and variable name.

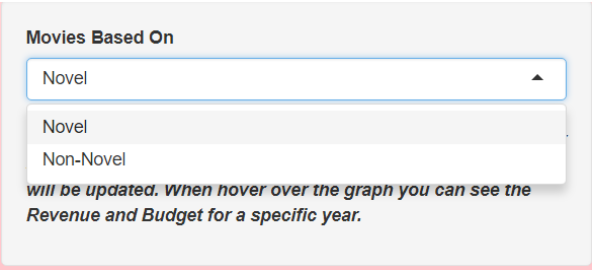


If you double click on the variable line chart for only that variable will be visible

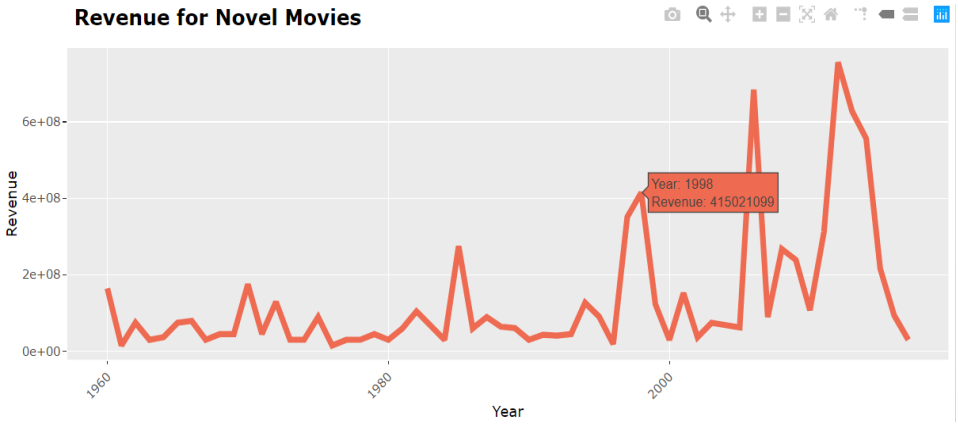
Novel vs Non-Novel Movies



Under this tab again we have a drop-down widget. When you chose from the drop down you



we will get a line graph for revenue generated by either Novel or Non-Novel type movies.

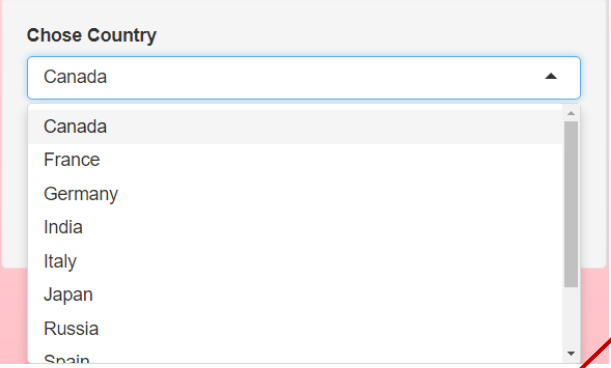


It's an interactive plot, when you hover over the line chart, we can see the Year and Revenue generated in that year.

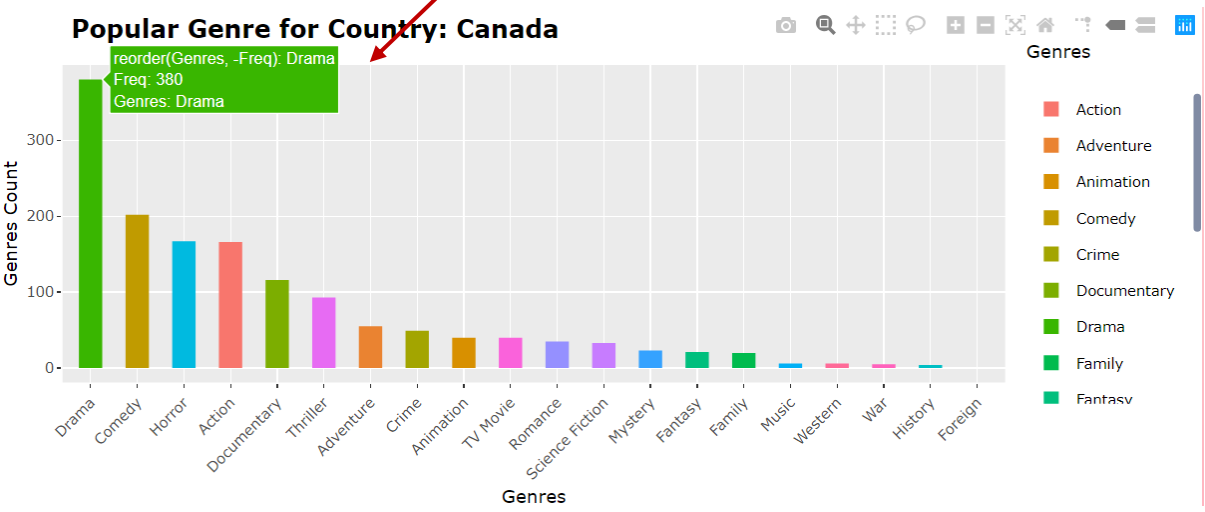
Production Country

This is design for this tab where we have bar chart for movies released by different countries and genre popularity in those countries

Like other tabs in this section we have from down menu to select a country, which generate bar graph for genre based on the popularity.



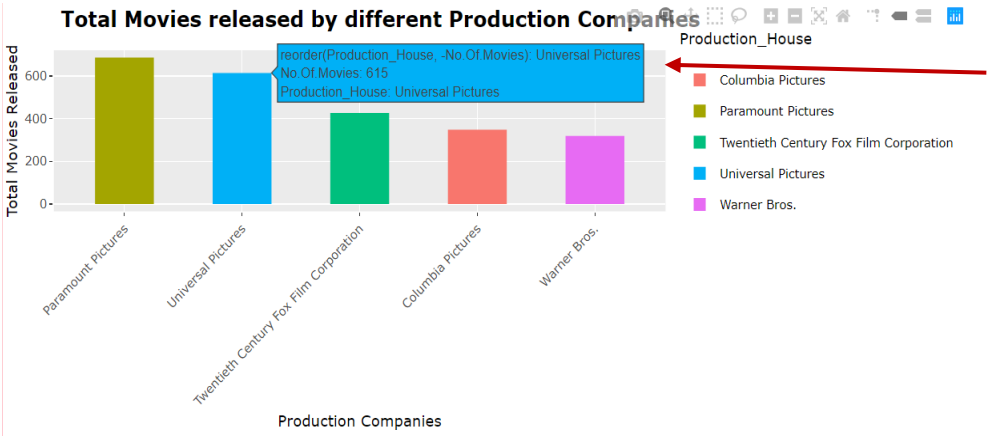
After selecting a country bar graph generated will look like this. It's a interactive bar graph which provide valuable information.



Production Companies

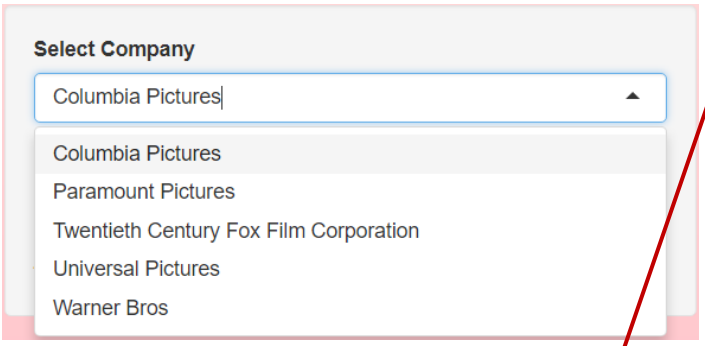
This page contains information on production companies' revenue and total movies they have produced.

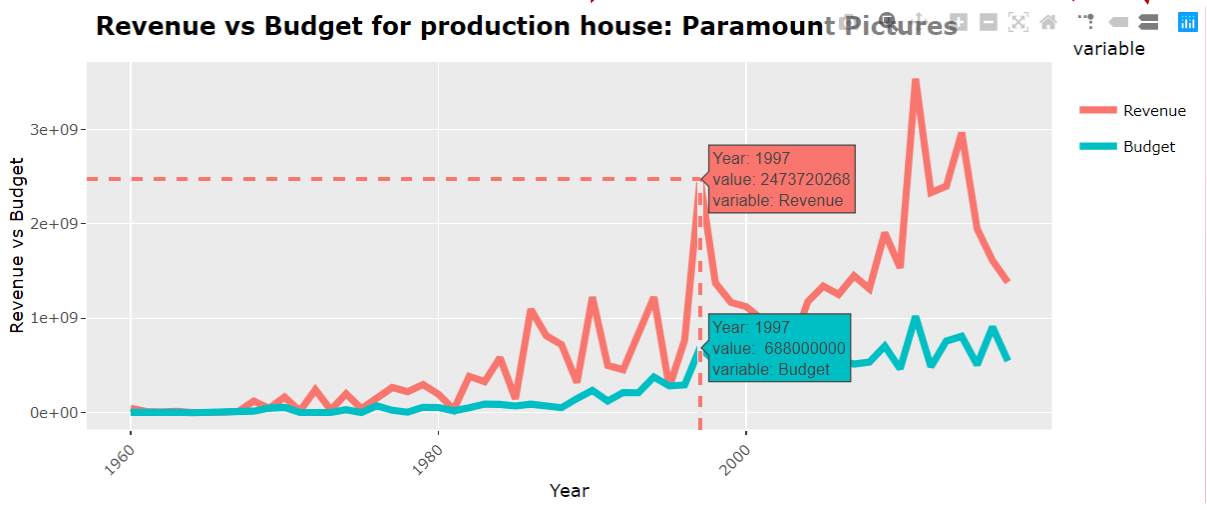
Bar graph is interactive and provide information on total movies produced by the production house.



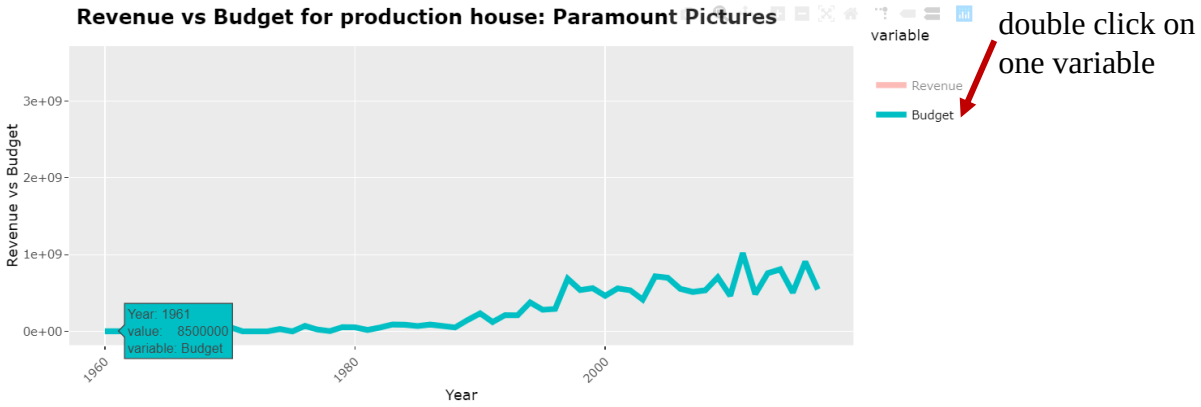
From the drop-down menu on this page, you can select a production house and it will render an interactive line graph to show relation between revenue and budget

To toggle spikes line and compare data on hover, click to option on the graph





And if you double click on one variable it will provide information regarding only that variable, here for example budget



Conclusion

This visualization project taught me how to use R programming effectively building an interactive website with Shiny Package and other packages. There were different variables in the dataset and different type of information to be visualised. Information regarding revenue and budget, number of movies produced based on country and production houses and different genre types. Using all these information and correct visualisation we can find the relation between all these variables.

For every data type, a specific graph is suitable for it. It is important to find the right graph which fit the dataset, or we can say which can represent data visually precisely. Layout selection somehow depend upon the intended audience as well and how to convey the story on your website as well.

There are many scopes of improvements in this project, such as adding interactive graphs with animation. Creating hyperlinks which will take you to another part of your application

Bibliography

http://pdritsos.com/files/Roberts-et-al-FDS-2016.pdf

https://shiny.rstudio.com/

https://cran.r-project.org/web/packages/reshape2/index.html

https://cran.r-project.org/web/packages/ggplot2/index.html

https://cran.r-project.org/web/packages/plotly/index.html

https://cran.r-project.org/web/packages/ECharts2Shiny/index.html

https://cran.r-project.org/web/packages/shinyWidgets/index.html)

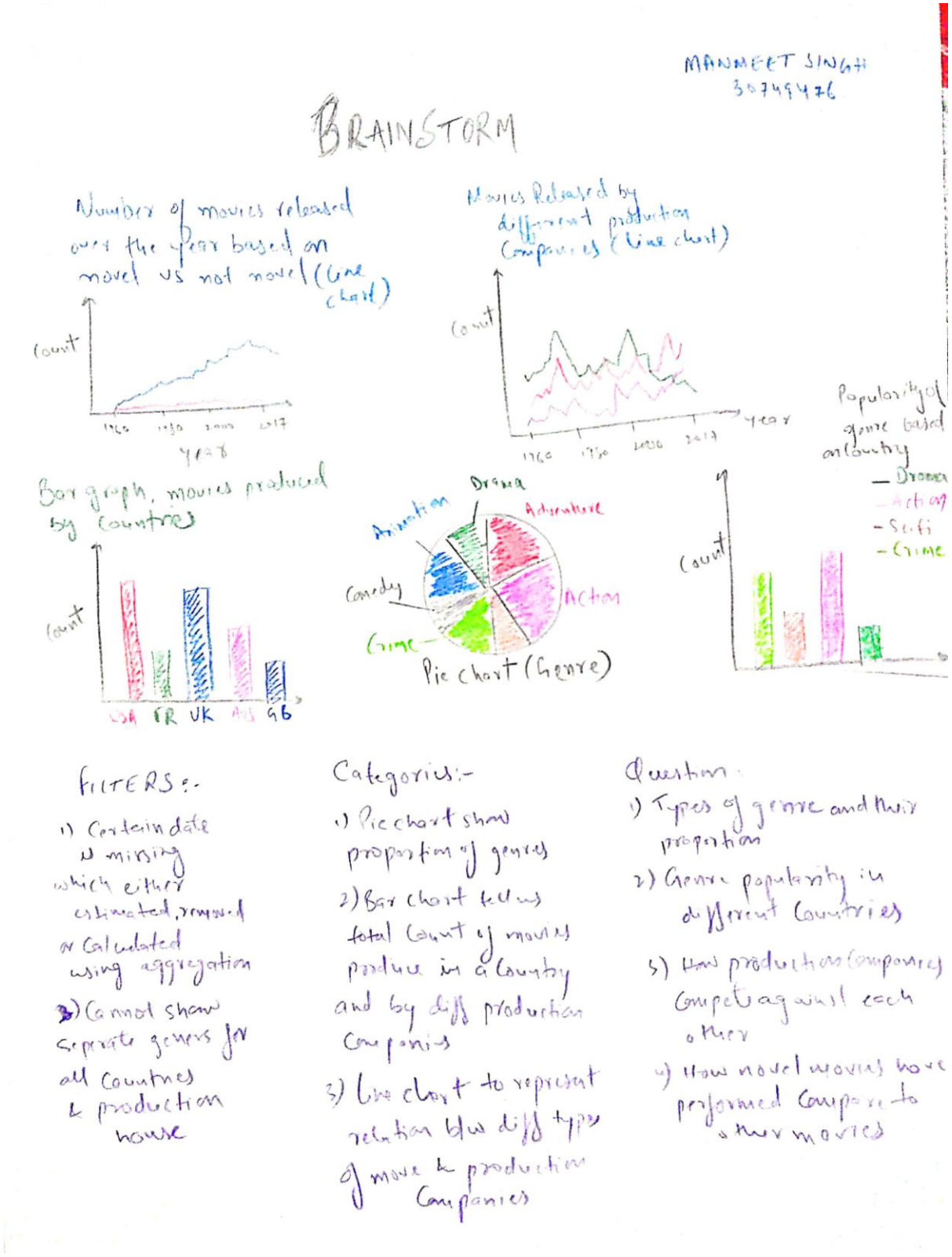
https://shiny.rstudio.com/tutorial/written-tutorial/lesson1/

https://mastering-shiny.org/advanced-ui.html#generating-html

https://bookdown.org/paulcbauer/idv2/8-7-ui-html-tag-functions.html

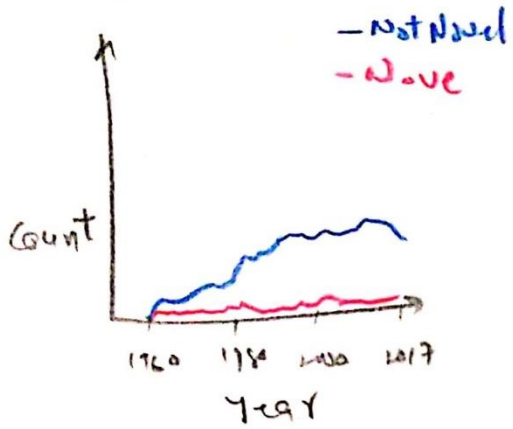
https://www.kaggle.com/rounakbanik/the-movies-dataset#movies_metadata.csv

Appendix


MANMEET SINGH
3074476
BRAINSTORM
Number of movies released over the year based on novel vs not novel (Line chart)
count
year
Movies Released by different production Companies (Line chart)
count
year
Popularity of genre based on Country
Drama
Action
Sci-fi
Crime
count
Bar graph, movies produced by Countries
count
USA FR UK AUS GB
Animation
Drama
Adventure
Comedy
Action
Crime
Pie chart (Genre)
FILTERS :-
1) Certain date is missing which either estimated, removed or Calculated using aggregation
2) Cannot show seperate genres for all Countries & production house
Categories:-
1) Pie chart show proportion of genres
2) Bar chart tells total Count of movies produce in a Country and by diff production Companies
3) Line chart to represent relation btw diff type of movie & production Companies
Question :
1) Types of genre and their proportion
2) Genre popularity in different Countries
3) How production Companies Compete against each other
4) How novel movies have performed Compare to other movies

Sheet 2: Intial Design

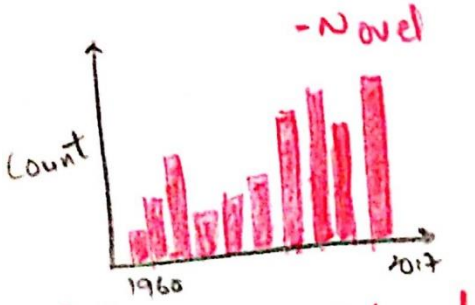
Layout



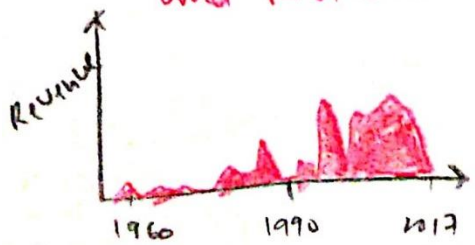
Operation

- User can Select the the type of movie based on novel or not
- After Selecting movie type, Count of Novel or non-novel movies will be displayed with there Revenue

Focus:



Selecting movies based on Novel will display barchart as Count and Revenue



Positive:

- Interactive
- informative

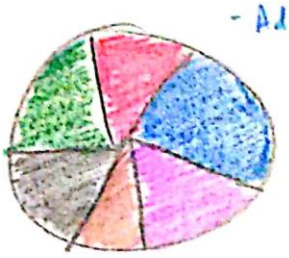
Negative:

- Cannot display information regarding revenue at same time
- Tedious to implement

Sheet3: Intial Design

Layout

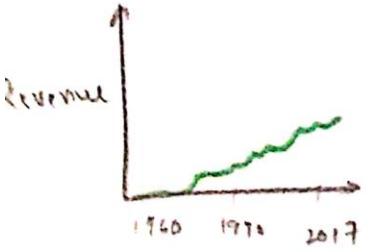
- Comedy
- Crime
- Action
- Adventure



Operation:

- Pie chart of genre type will be static
- Once user Select genre type, Line chart of Revenue generated

Focus:



Choosing a genre type will provide Line chart of Revenue

Positive:

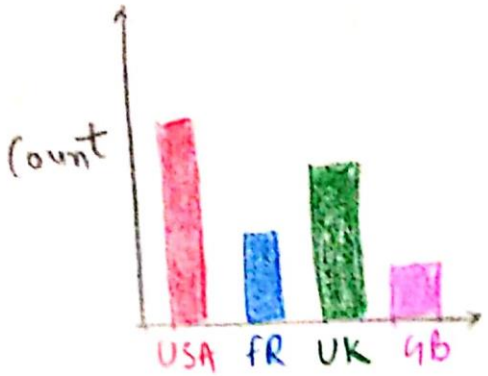
- Interactive
- informative

Negative:

- Cannot display revenue of genre type at once

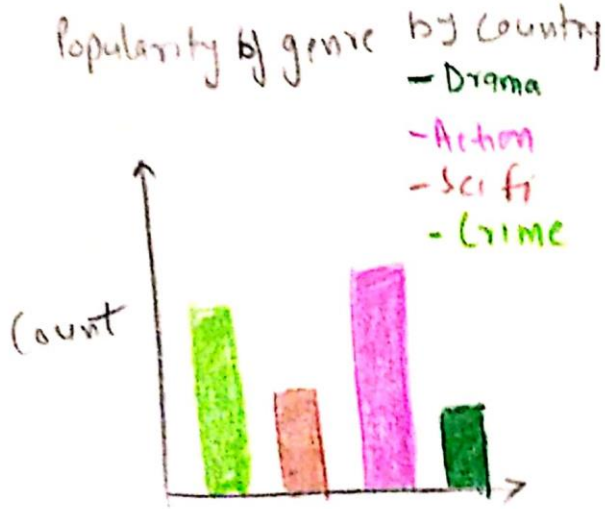
Study : Initial Design

Design layout:

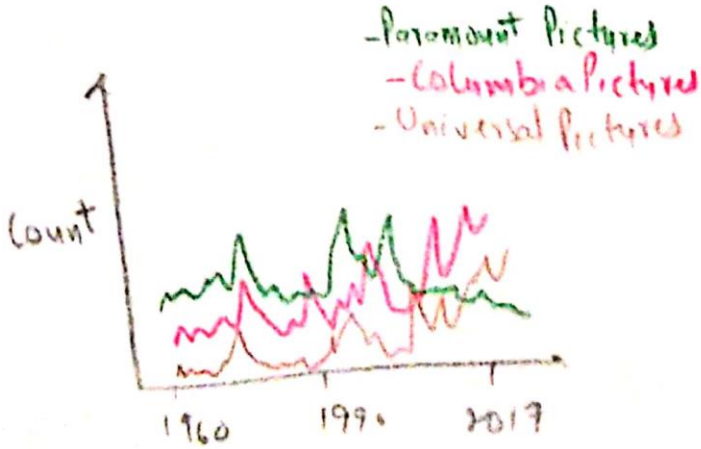


Operation :-

- Number of movies released by countries will be displayed directly. No user input required



- Popularity of genre based on country will be intuitive
  User can select country and genre popularity will be displayed



Positives

- Informative
- Easy to display

Negative:

- Simple design

Sheet 5. Realization
Layout
-Novel
-Not Novel
Count
Legend
-Genre type
Legend
-Country
Count
-Country
-Sci fi
-Action
-Drama
Legend
production
Companies
Count
Focus
Count
Revenue
Grave
Revenue
Operation.
• User can filter the preferences and visualisation are generated as per preference
Details.
• Libraries.
1) Shinny
2) ggplotz
• Tool
1) R-studio
• Estimated build time
- 4:7 days