**fINFO 5709 PROJECT 2**

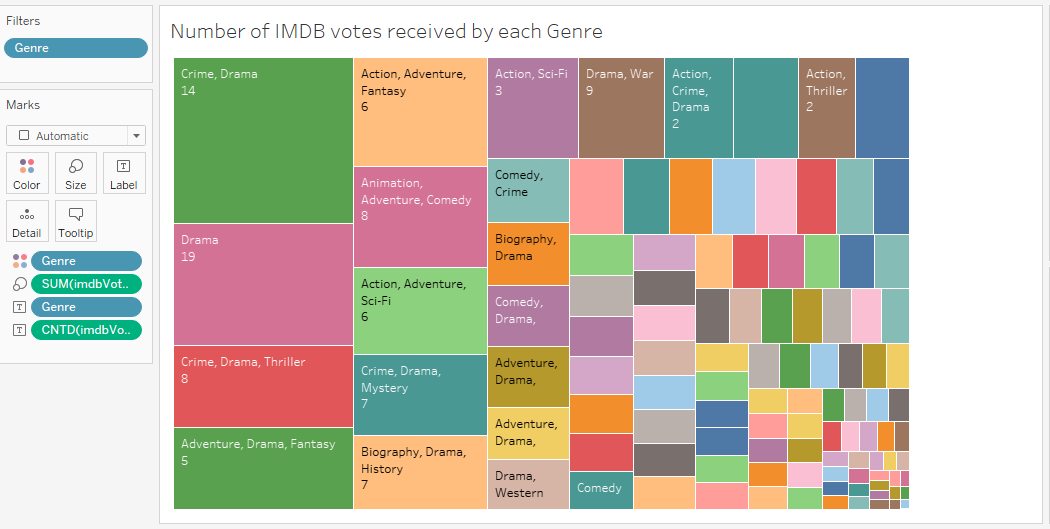
**MULTIDIMENSIONAL DATA VISUALIZATION DESIGN**

**Chandrika Yalamarthy**

**11364624**

Data visualization has been done using the IMDB top 250 movies dataset. Data cleaning has been performed in Tableau using the data interpreter. The following are the insights formed from the data visualization.

1. **Which Genre has received the highest count of IMDB votes?**

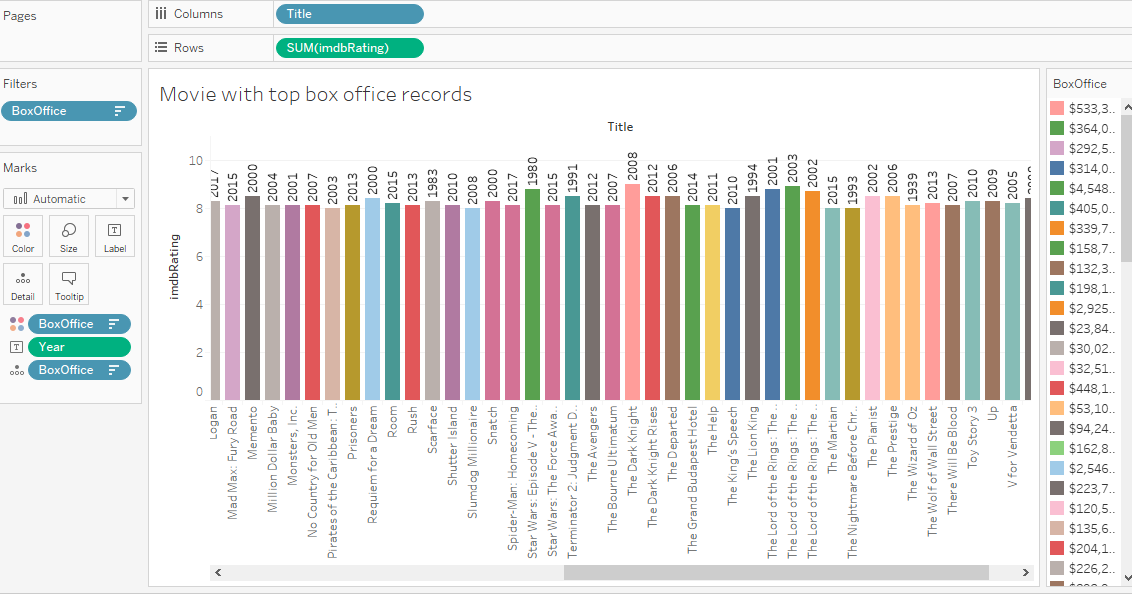


The above visualization shows the count of IMBD votes for each Genre. I have placed Genre under Filters pane and marks pane, I have placed Genre under color, SUM of Imdb votes undersize and I have used CNT ( count) to show the number of count of IMDB votes and I have placed it under label tile. I have used TreeMap to create this visualization.

Treemap displays data in nested rectangles, the dimensions define the structure of the treemap and measures define the size or color. Nested Rectangles means that the hierarchy levels in the data are expressed by larger rectangles above in the hierarchy containing the smaller ones below in the hierarchy.

From the above visualization we can conclude that Drama Genre has the highest number of count and the Crime, Drama genre recorded 96,84,911 votes and Action, Crime, Music genre recorded the lowest number of 2872 votes. The color shows details about the genre , the best-suited measure to color a treemap is a performance measure. The size of the rectangle is determined by a quantitative measure the values of which should sum up along the hierarchical structure of data. The measure must always be more or equal to zero as we cannot express a negative number with size.

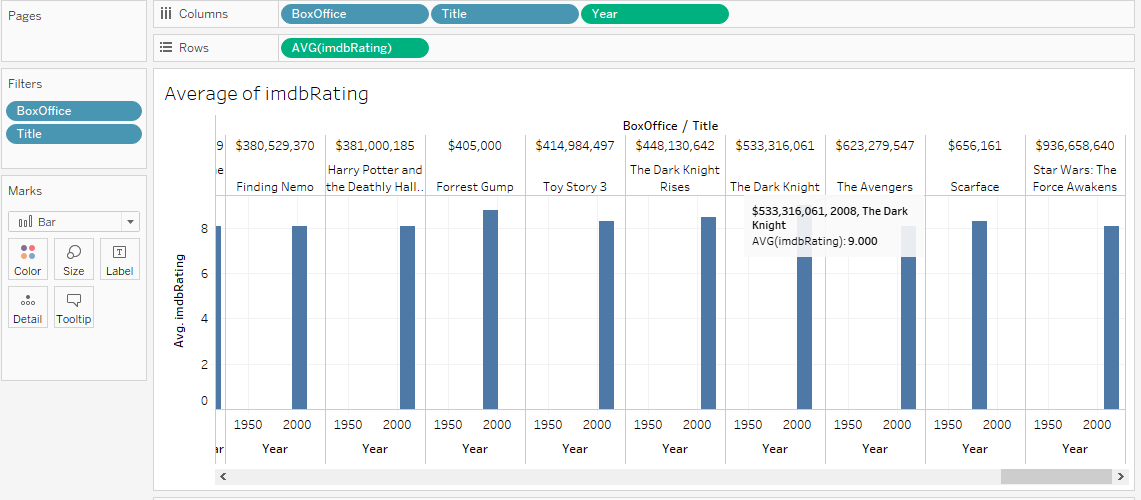
1. **Which movie holds the box-office record?**



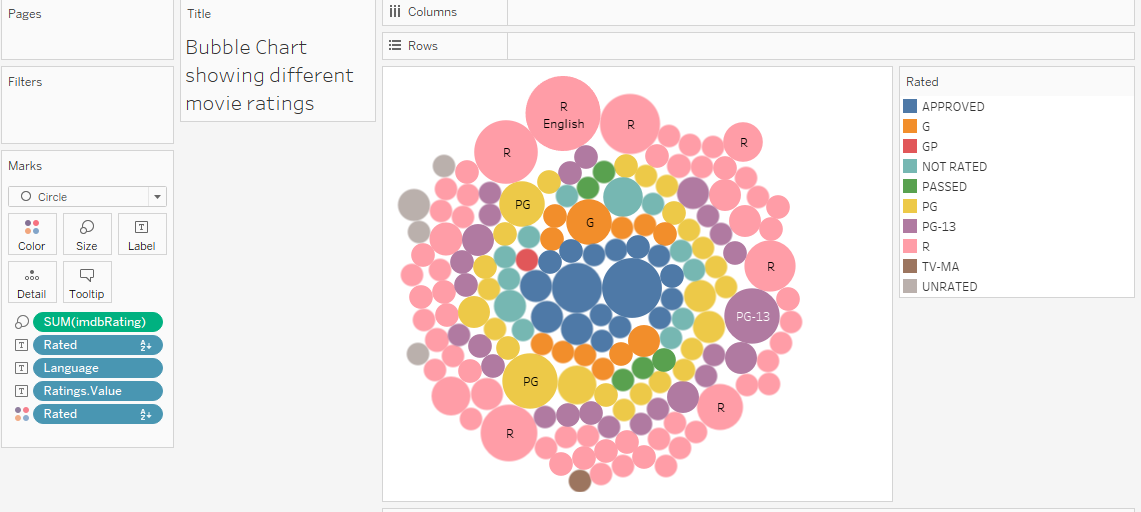
The above visualization depicts movies with a top box office collection, I have used a horizontal bar chart for this visualization. A bar chart represents data in rectangular bars with the length of the bar proportional to the value of the variable. The bar chart with color change we can apply colors to the bars based on their ranges, the longer bars get darker shades and the smaller bars get the lighter shades.

From the above visualization, we could see that the highest box office record of 936,658,640$ million was made by the movie Star Wars The Force Awakens in the year 2015 with an IMDB rating of 8.1. The lowest box office record of 135,620$ was made by 2001: A Space Odyssey in the year 1968 with IMDB rating 8.3.

In marks pane, BoxOffice dimension is used for colors as we can see they are different colors for each dimension. Columns has title and rows contain sum of IMDB ratings , year is labelled on the top of the bar chart and the details pane shows the Box Office collections. The view is filtered on BoxOffice and Title. The BoxOffice filter excludes N/A values. The below visualization shows the average of IMDB rating , The Dark Knight marks an average IMDB rating of 9.00, the higher than expected average of IMDB rating for this mark increasing the sum of IMDB rating.



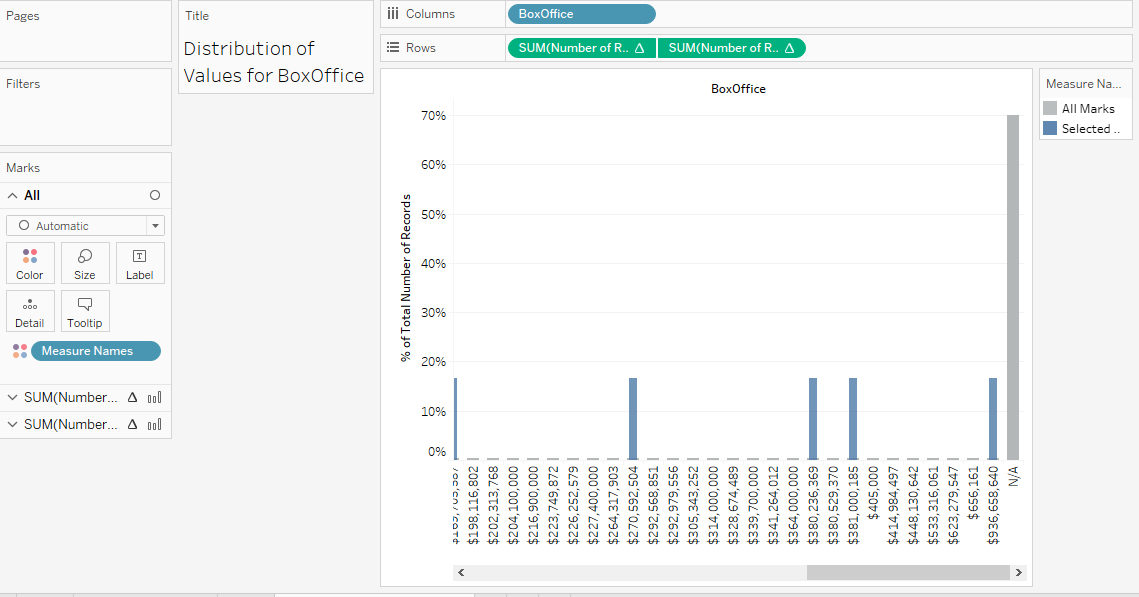
1. **Different Movie Ratings**



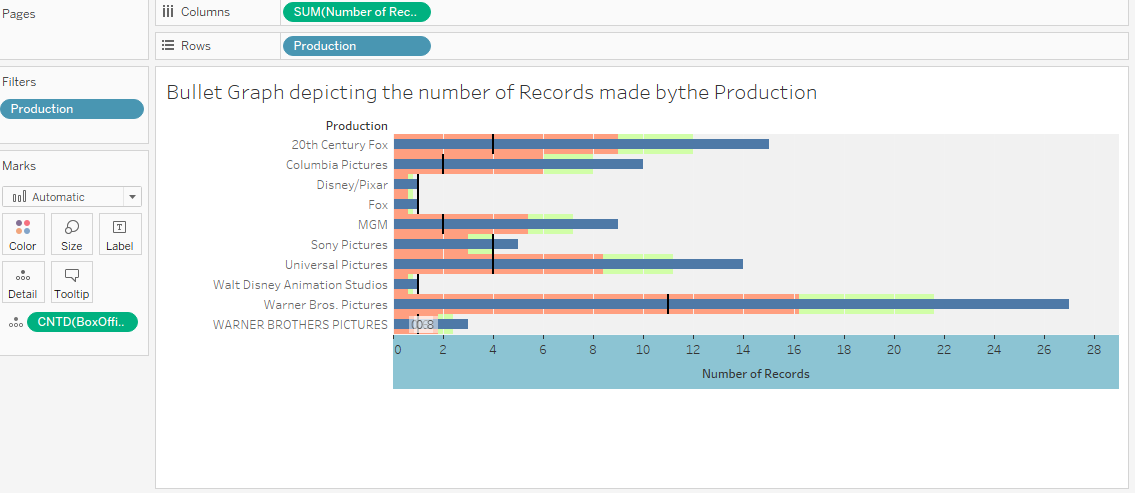
The above visualization is a bubble chart showing different movie ratings , bubble chart display data as a cluster of circles each of the values in the dimension field represents a circle whereas the value of the measure represent the size and color of the circles. The basic building blocks for a packed bubble chart are the mark type must be circle, detail as dimension , size as measure, color as dimension or measure and label can be dimension or measure.

The movie ratings are classified into Approved, G, GP, NOT RATED, PASSED, PG, PG-13, R, TV- MA and UNRATED. We could see from the visualization that most of the movies are R rated which means those movies are restricted and requires accompanying parent or adult guardian, the pink color shows the movies that are restricted. There is only one television program that is rated as TV-MA which means it is only for mature audience and there is one movie that is rated as GP which indicates that it is suggested for General Audiences, but Parental Guidance is Advised.

The color shows details about Rated , the size shows sum of IMDB rating and the marks are labeled by Rated, Language and Ratings Value. I have selected PG-13 circle and this mark has an interesting of values for BoxOffice , the below chart shows how the selected mark has a different distribution of values within BoxOffice compared to the distribution of values within BoxOffice for all records in the source visualization.

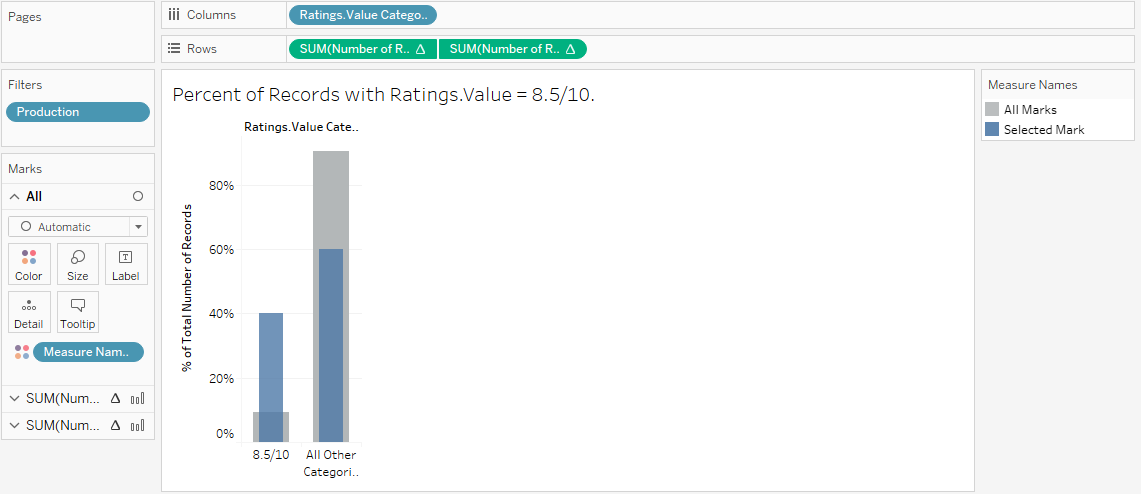


1. **Number of Records made by the Production**



The above visualization shows the number of records made by the production. I have used Bullet Graph for this visualization , bullet graph is a variation of Bar Chart. We compare the value of one measure with another measure in the context of finding the variation in the first measure within a variation in the second measure. It is like two bars drawn upon one another to indicate their individual values at the same position in the graph. Bullet graphs are used to compare one value represented by a horizontal bar to another value represented by a vertical line. The horizontal bar represents the value of interest and the vertical line represents a comparative value, the color ranges can be set to define different segments.

In this visualization the sum of number of records is placed in the columns shelf and production is used in rows shelf. The view is filtered on Production which keeps 10 of 89 members. We could see that the horizontal bar represents the distinct count of Box Office and number of records for the Production house. The vertical line represents Average distinct count of box office. We could see Warner Bros Pictures has maximum number of records. The color ranges has been set as stoplight to define different segments. The below chart shows the percentage of records with ratings value within the selected mark compared to the percentage of all records with ratings value in the source visualization. The selected mark is Sony Pictures and it occupies 60% of the records.



**PART 2**

**Critique the visualization tool**

Critique is an analytical evaluation of what works and what doesn’t, critique is valuable and helps to grow as we decide to seek it out.

**Tools Strengths and Weaknesses**

**Strengths:**

Tableau has remarkable visualization capabilities, it converts unstructured information into comprehensive logical results that are fully functional, interactive and appealing dashboards.

Tableau has an intuitive manner of creating graphics and a user-friendly interface allow non-development users to utilize the basic application functionality to the fullest.

Lay users can enjoy the capabilities that Tableau offers for stats parsing such as dashboard development. Tableau operates fast even on big data which makes its powerful performance as robust and reliable.

**Weaknesses:**

Tableau fails to provide centralized data-level security, it just allows establishing row-level security which stipulates that every user his/her account.

Tableau lacks functionality required for a full-fledged business intelligence tool such as large-scale reporting, the building of data tables and static layouts.

Tableau’s solution’s capabilities do not allow for a smooth embedment, seamless Tableau’s integration into a company’s product could be a real challenge from each of the money and technical points.

**Characteristics of UI:**

User Experience is important in making sure the extension is well adopted and the user has an easy time learning to use it as this will enhance the value of the extension , the content in the extension should clearly indicate to the user how to interact with it.

UI patterns come in handy as they are drawn from best practices within UI design and they are reusable solution to commonly occurring problems.

Color is important to use with purpose and clarity , we have to use it minimally to highlight functionality in the extension using effective colors will communicate to the user what to focus on. The font should be a clear hierarchy of content distinguished by using different font sizes , weight , color and spacing.