# Information Visualization – Final Programming Project 3

Name: Vineetha Nalamolu Student ID: 811140876
Mail: vnalamol@kent.edu

For this project, I have chosen the ***films dataset*** for developing a prototype of visualization.

Link: https://web.cs.kent.edu/~vnalamol/vineethan_IV.html

**Project Steps and Requirements**:

1. **Identify at least three visualization tasks for the dataset.**

The first step in developing the prototype is to identify the tasks for the given data set.

The main tasks I identified for the development of the prototype are as follows:

- How many films are produced over the years in each subject?
- How many films produced in Horror received awards and did not receive the awards?
- What is the average popularity over all the years for the films in a subject?
- Which subject films have the highest popularity of all?
- Which type of movies has the highest run time?
- What is the length of the films for drama in the year 1950?

2. **Design a visualization system to fulfill the identified tasks**.

In the first step, we have identified some tasks for the given dataset. Now in the second step we will design a visual system to address the above tasks. Let us now consider each task.

- **What is the count of films produced in each subject?**

  For this task, we need to group the films based on the subject. Later all the titles in that subject should be summed. This task can be visualized using different interactive charts. Here I used *Interactive Pie chart* for better interaction. In this pie chart, different subjects are colored with different colors for better understanding. When we select a subject in the pie chart, only the corresponding sector in the pie gets highlighted and the remaining sectors become dull. When we click the reset option on top of the pie the whole visualization i.e, pie chart will reappear.

- **How many films produced in Horror received awards and did not receive the awards?**

Here we will sort all the films based on the subject and after sorting them, we need to count the total awards received. I used *interactive donut chart* for displaying the number of awards received. I preferred using donut since, pie and donut complement each other and be very easy for the user to understand the relation. When we select the horror subject in the above pie chart, the total count of films that received awards will be displayed in the green color and the count of films which didn't receive the awards will be displayed in the red color.

- **What is the average popularity over all the years for the films in a subject comedy?**

The cream choice I prefer to visualize this task is by using the *interactive box plot*. The range of popularity can also be shown in this box plot along with the mean value. All the subjects are placed on the x-axis and the popularity of the films are represented in the y-axis. By looking at the box plot we can easily recognize the popularity of the films for all the subjects. The mean value is shown in the graph by a horizontal line in the box for each subject. When we click on the comedy subject, the popularity of comedy will be highlighted.

- **Which subject films have the highest popularity of all?**

For this task, we can use the same visualization we used in the above task. We have sorted the films based on the subject and used the box plot to visualize. All the subjects can be shown side by side in this visualization and will be easy for the user to understand. This type of plot will be very helpful for the movie makers to know what kind of films are more preferred by the audience so that they can make more kind of such films in future and can generate more revenue.

- **Which type of movies has the highest and the lowest run time?**

Run time is basically the length of the movie. For this task I preferred using an *interactive bar chart* to know the difference accurately. The total length of all the films in a particular genre are calculated and sorted according to the subject. The length of the bar denotes the run time for that subject. The subject with the highest bar has the highest run time and the subject with the least bar has the lowest run time.

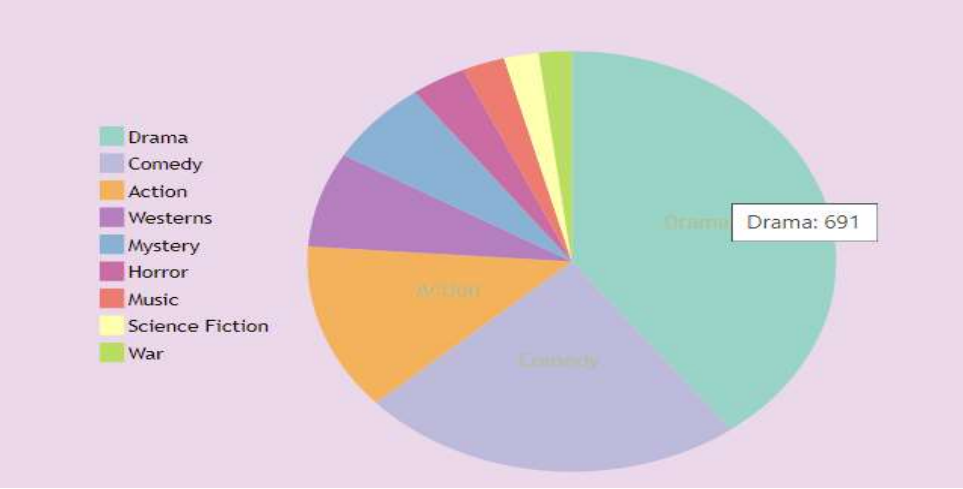
- **What is the length of the films for drama in the year 1950?**

For this task, I used *curved line chart* to depict the data. Since we have wide range of years, Line chart fits best for these attributes. Each subject is differentiated with different colors.By hovering mouse on the curve, we can see the label for the year and the total length of the films in that subject. Also, from the graph we can see that the length of the films increases over the years.

3. **Implement the Web-based visualization system with interactions.**

The last and final step in developing the prototype is to implement the tasks we have considered.

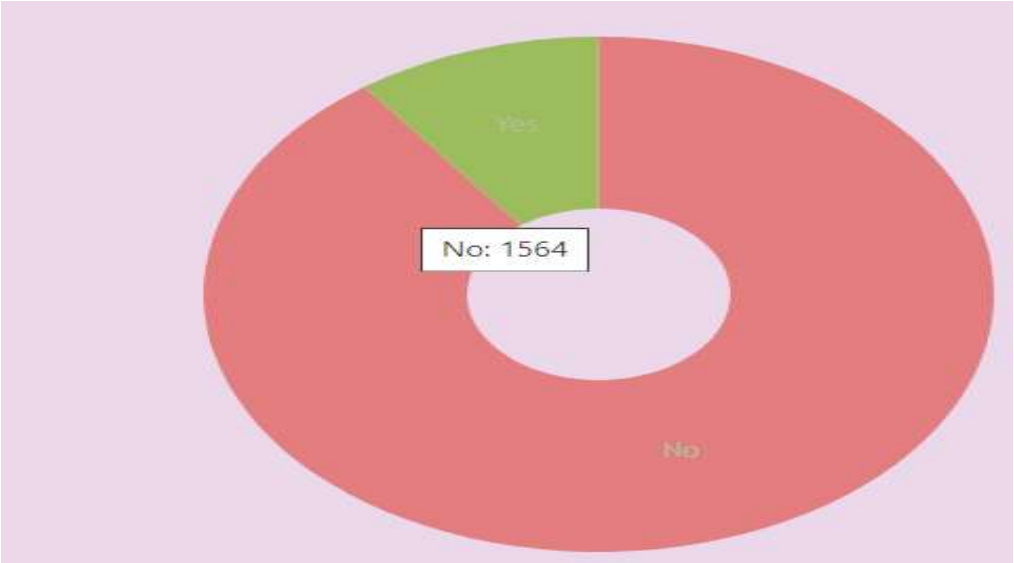
- **What is the count of films produced in each subject?**



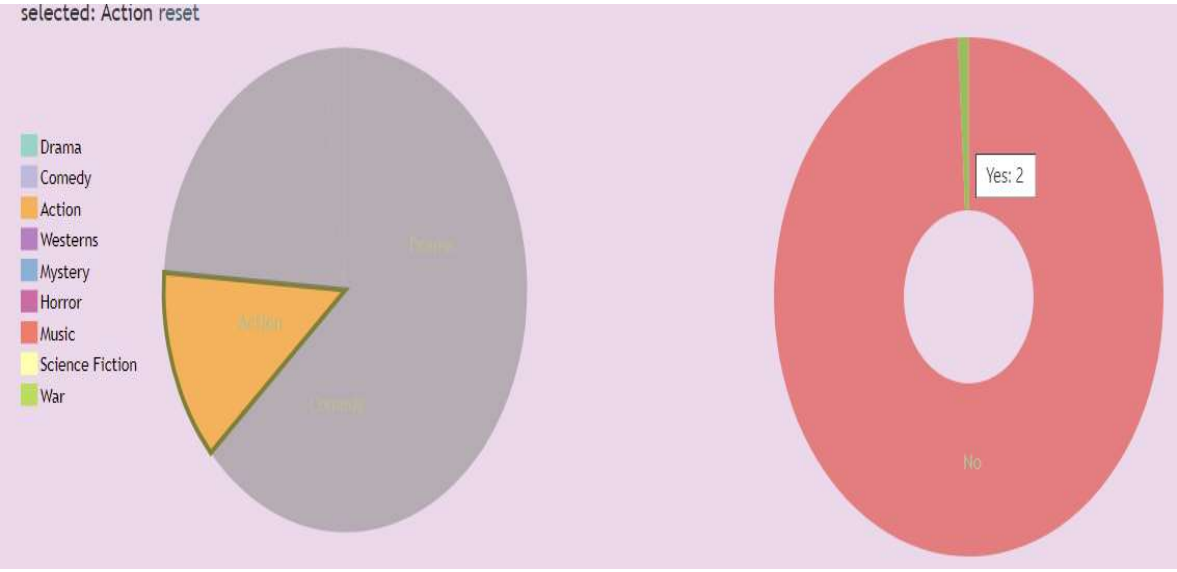

The above displayed is the pie chart showing the count of films in each subject. Here, we can see that the films in subject drama are 691.

The graph is interactive and it can be seen by hovering mouse on the chart as well as by clicking on either the legend or the sector in the pie films count in specific subject.

- **How many films produced in Horror received awards and did not receive the awards?**



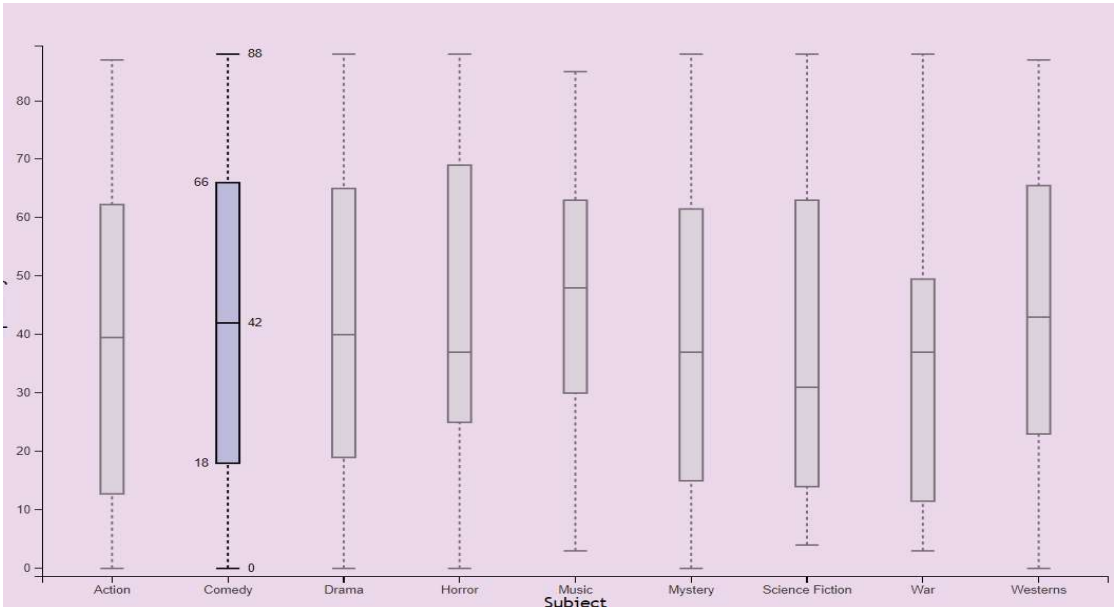

The above donut shows the total count of films that received awards and did not receive them. The total films that didn't receive any awards are 1564 and received are 176.




When we select subject in pie chart, we can see the total films that received the awards are 2 for action subject. We can also select the awards category i.e, yes or no.

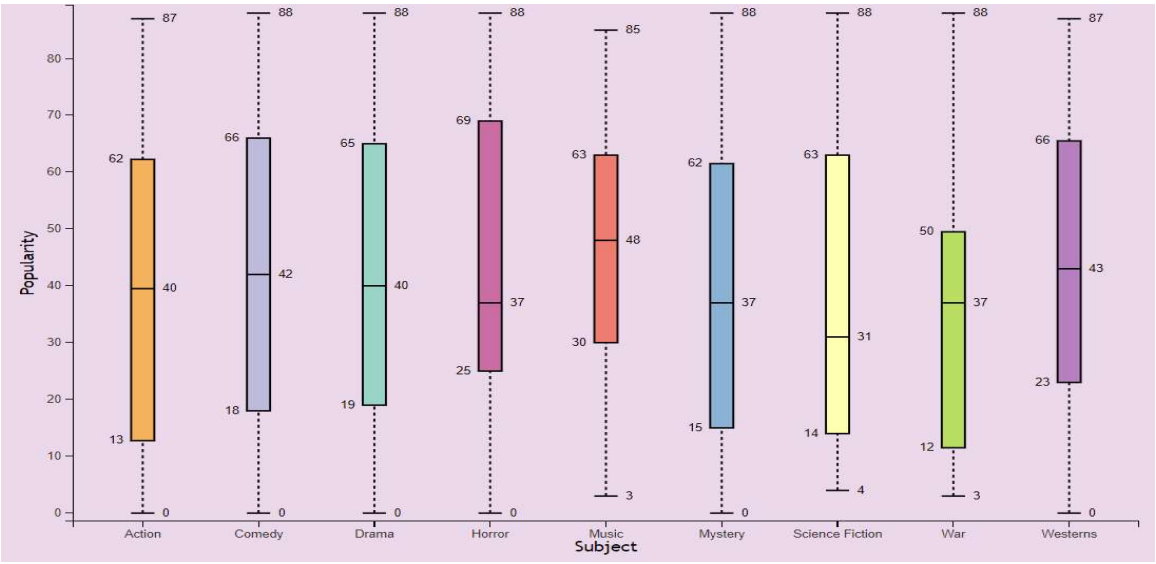
- **What is the average popularity over all the years for the films in a subject comedy?**

Interactive box plot is used for displaying the average popularity for each genre. If we select the comedy subject, only the single box will get highlighted. Thus, we can clearly see that the average popularity in comedy subject is 42 and the range of popularity ranges between 18 and 66.



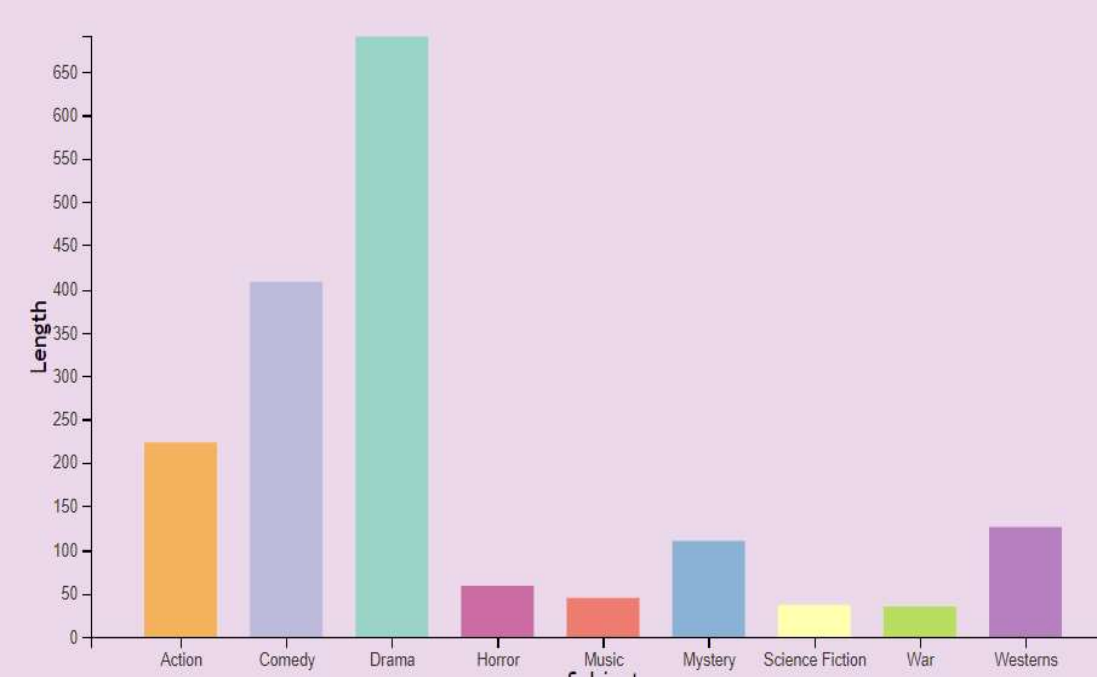

- **Which subject films have the highest popularity of all?**

In this box plot, we can see the popularity for different subject films. We can also see the range of popularity of all the subjects. We can clearly see that horror films have the highest popularity.



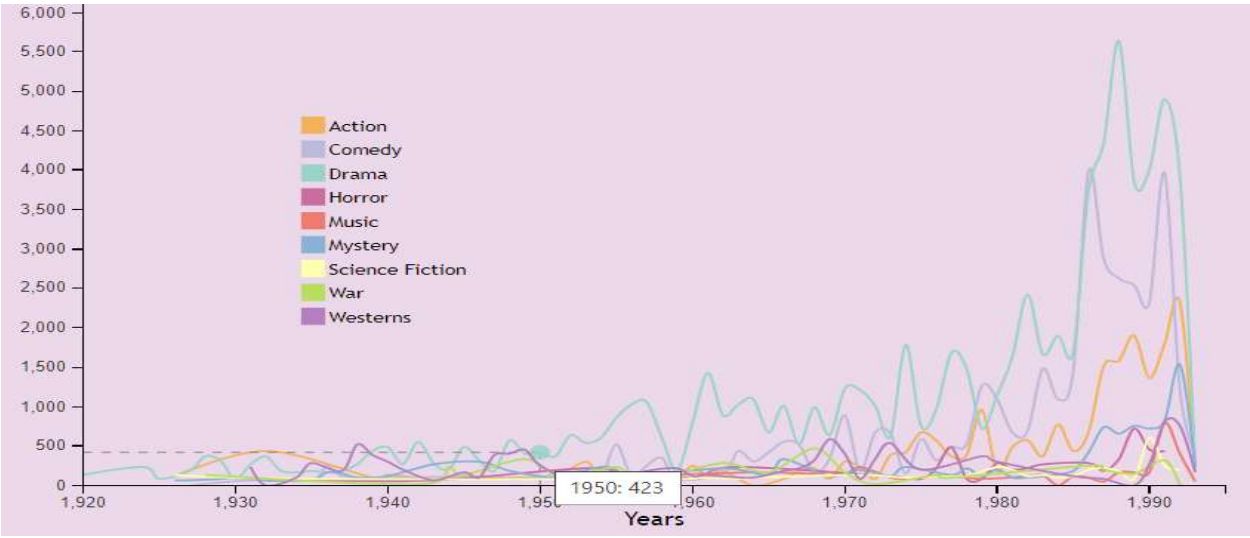
- **Which type of movies has the highest and the lowest run time?**

Run time is the total length of the films. Interactive bar chart is used to depict the highest and the lowest runtime. We can clearly see from the graph that the films in Drama has the highest runtime and the films in science fiction has the lowest run time.




- **What is the length of the films for drama in the year 1950?**

A curved line graph is used to depict the length of the films. We can get the length of the films by hovering the mouse the mouse on the particular line. In the below graph we can see that the films in drama subject has run time of 423 minutes in the year 1950.



The entire visualization can be restored by clicking the reset button on the top of the visualization.