ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

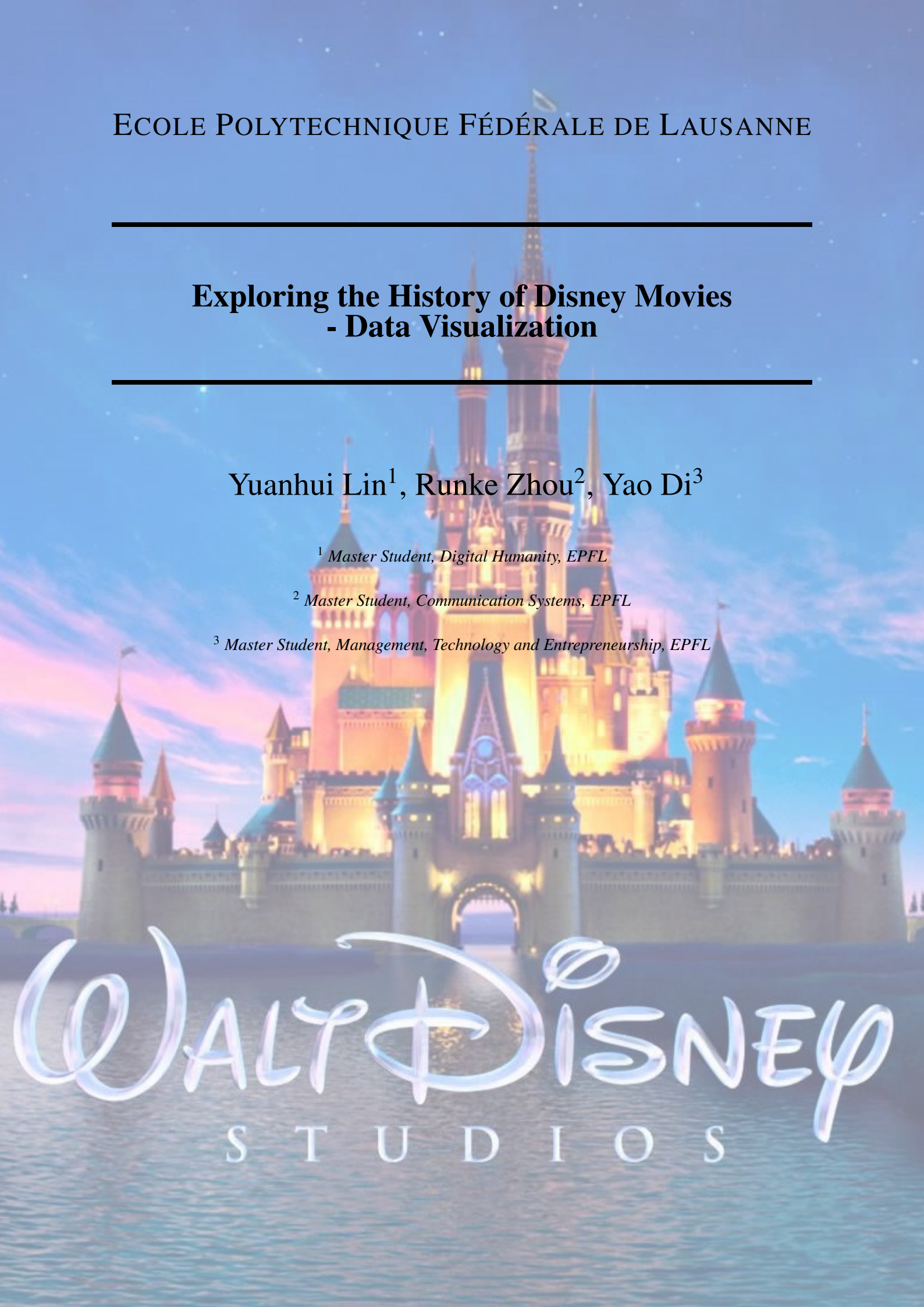
Exploring the History of Disney Movies - Data Visualization

Yuanhui Lin[1], Runke Zhou[2], Yao Di[3]

[1] *Master Student, Digital Humanity, EPFL*

[2] *Master Student, Communication Systems, EPFL*

[3] *Master Student, Management, Technology and Entrepreneurship, EPFL*



Contents

1 Motivation and Introduction **2**

1.1 Motivation 2

1.2 Introduction 2

2 Dataset and Exploratory Analysis **2**

2.1 Dataset 2

2.2 Exploratory Data Analysis 3

2.3 Further Data Processing 3

3 Design and Final Implementation **3**

3.1 Preliminary Design 3

3.2 Challenges and Changes 5

3.3 Final Implementation 6

3.3.1 Characters 6

3.3.2 Best Movies 6

3.3.3 Finance 7

3.3.4 People 7

4 Peer Assessment **8**

4.1 Lin, Yuanhui 8

4.2 Zhou, Runke 8

4.3 Di, Yao 8

1 Motivation and Introduction

1.1 Motivation

Who doesn't like Disney? With a history started from 1923, we believe Disney has a profound influence on many generations. Indeed, the three of us have to admit that Disney movies play an very important role in our childhood. They sometimes gave us inspirations and shape our mentalities. What's more, Disney has portrayed many classic characters, which people may find themselves related to. Thus, different people may have different favored characters.

We think there are many interesting aspects to dive in regarding Disney movies and the whole process of exploring data related to Disney movies will certainly be enjoyable. Therefore, we decided to analyze a large number of Disney movies and hope to provide some insightful information.

1.2 Introduction

For this project, we aim at building an interactive platform analyzing Disney movies from different aspects. Firstly, we show some famous characters from Disney movies to bring users' memory back to the time they watched Disney movies. Afterwards, the top 30 highly reviewed Disney movies throughout a history from 1940 to 2020 are presented. Other important aspects of the movies such as the budget and the box office of each movie are also analyzed in details. Last but not least, we analyze the staffs of Disney movies by showing their contributions and their cooperative relationships.

In this process book, we present the process of how we go from CSV data of Disney movies to the final website. The data exploration, detailed design and implementation of website are included. Peer assessment presents the breakdown of our project and who works on which part in detail.

2 Dataset and Exploratory Analysis

2.1 Dataset

The preliminary datasets of our website are from Kaggle. For reference to these datasets, please check: Walt Disney Movies - Soaibuzzaman,Walt Disney Movies - Diksha Bhati, Disney Movies and Films Dataset - Sameer Patel. These datasets have some overlaps regarding the contained movies but different features are included. Thus a further data cleaning to these datasets are needed. We merge these datasets together and get the final one which we used to implement our website.

2.2 Exploratory Data Analysis

- **Analysis of running time:** The running time of most Disney movies is between 80 to 120 minutes.
- **From a historical aspect:** We plotted how many movies are released each year. According to the plot, most Disney movies were released after the 1990s.
- **Analysis of Ratings:** In our final dataset, we have three ratings for a movie, IMDB, Metascore, and Rotten Tomatoes. Even though we expected a high average review of Disney movies. it is interesting to find out that there are some outliers with pretty low scores.
- **Analysis of Budget and Box Office:** We plotted the budget and box office of each movie in a time sequence. Through this plot, we can observe an obvious increase in box office concerning time.
- **Analysis of People:** Another important statistic about movies is the staffs working on the movie. People interacting in a movie include directors, producers and music directors.

After the first analysis of the dataset we have, we decided that our visualization should include the historical aspect, revealing the ratings of movies and the financial data. The related people is also of our interest.

2.3 Further Data Processing

We want to hightlight the top 30 movies through history. To analyze the 30 Disney movies with the highest review, we take into account two kinds of ratings, IMDB and metascore, to compute an average value and get the final score of movies. These two kind of ratings represents both the review of the movie from the audience and the professional critics and publications. When averaging ratings, we firstly normalizing the score range of two kinds of ratings and then get the average as the final score. By descending the final score, we get the 30 Disney movies with the highest review.

3 Design and Final Implementation

3.1 Preliminary Design

After brainstorming, we made our initial design on how our website should look like as follows:

To bring our audience a feeling of arriving at Disney wonderland, our website starts with a video of the standard Disney title sequence. As there are many aspects to describe a movie, we decided to divide our visualization to different sections.

1. **Characters:** The characters scene will be an interactive scene focused on characters in Disney movies. When the audience enters this scene, a series of characters will appear on the screen in a monochromatic format. The audience could select his/her favourite character and the selected one will be coloured. A short description describing the character will appear as well. The preliminary design of the character section is shown in Figure 1. We think this kind of interaction can make the audience more engaged and hopefully generate emotional resonance.

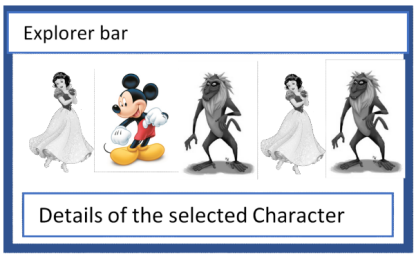


Figure 1: Character section

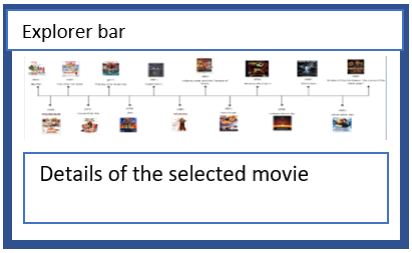


Figure 2: Best Movies section

2. **Best Movies:** The idea of the best movies scene is to present the highest rated movies from Disney. There will be a timeline presenting the issue time for the popular films. When the audience selects the icon of the specific movie, the detail of the movie will appear at the bottom. The preliminary design of the Best movies section is shown in Figure 2.

3. **Finance:** This scene will present the data in finance aspect, including the budget, the adjusted profit and the box office. We will plot the information for all movies at one plot, an efficient and intuitional way will be chosen. As we have more than 200 movies, we will use brush to make the graph scalable.The preliminary design of the finance section is shown in Figure 3.

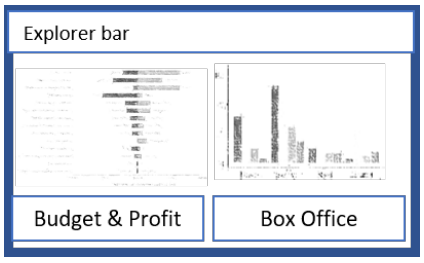


Figure 3: Finance section

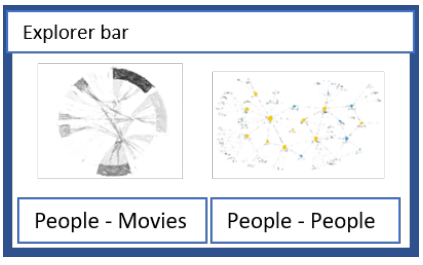


Figure 4: People section

4. **People:** This scene will present the directors, producers and music directors related to Disney movies. A viz graph will be used to show the links between the people and the films, each role will have a graph like the one shown on the left. The other will show collaboration between people, we will use d3 force directed graph to construct the network. The preliminary design of the Best movies section is shown in Figure 4.

3.2 Challenges and Changes

During the implementation of our website, some of initial designs didn't have a good visual performance, thus we faced with some challenges and made some changes to the initial design.

For some sections in our website, for example, character, finance and people sections, we wish to implement visualizations from different aspects to have a better insight on the specific topic. But if we put two visualizations in the same page, it is hard to deliver information that we want to convey to the audience. Thus we add a slider to our website. In this way, we can include two pages in the same section while keeping the website structure neat. The user can just move the slider on the bottom to go to the visualization on the next page.

In the character part, in the beginning, we planned to include the interaction of coloring character images and introduction of characters together. But the user experience turns out to be not good with long description. Therefore, we divided this section into two parts: coloring characters interaction and top characters introductions (Figure 5, 6). We made this change also because we think the audience should be very familiar with the chosen character, therefore only some interesting notes related to chosen character will be shown. Also, originally, the characters should be in SVG format in order to have animation. But we had problem finding suitable SVG file for many characters, therefore, we used PNG instead and achieved similar result.

In the best movie part, we originally designed to show posters of movies next to the timeline. But during implementation, we found that as we want to indicates the score of movies using the distance of the circle to the timeline, some of movies released in the same year might overlap with each other. Thus we changed our visualization approach by using tooltip and onclick from d3 library. In this way, when mouse hover on the point, the title of corresponding movie and the ratings will be shown. After clicking on the point, the poster and a brief introduction of this movie will be shown on the bottom of the timeline. To provide a better atmosphere for audience, theme songs for the movie will be played when user clicks on the point.

In the Finance part, we preliminary designed to analyze this section through analyzing the budget, profit, and box office of all Disney movies. However, we found the Tornado plot for 200 movies owns some aesthetic defects during the implementation. Besides, it is hard to find insight for the audiences. Therefore, we changed the scope to view the Top-10 movies according to different financial metrics, including profit, return on investment, and loss (Figure 8: Top 10 Films). Furthermore, we tried to conclude additional meaningful financial data, including the historical cash flow. Therefore, we added an additional graph (Figure 8: Historical Cashflow) which builds a more informative storyline by introducing the IMDB scores as a control variable.

In the People scene, we changed the original design of the people-movies figure as shown in Figure 4 to a radial stacked bar graph (Figure 9). We made this change because with over

200 movies and over 100 people related for each group, it is extremely hard to organise the graph in a reasonable way. Instead, we use the circular stacked bar graph to indicate the IMDB rating of the movies one person has worked on. In this way, the audience gets to know how many movies one person has contributed to and the quality of the contributions.

3.3 Final Implementation

The final implementation of our website starts with a video of the standard Disney title sequence and then is separated to different sections as we mentioned in Chapter 3.1.

3.3.1 Characters

This section starts with an interacting scene to color the character chosen by user, a brief comment will shown up after clicking. The user can follow the instruction to slide to the next page, where introductions of 6 top Disney movie characters are presented.

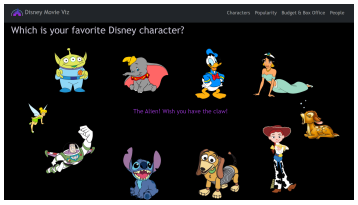


Figure 5: Coloring Characters

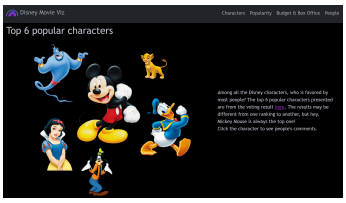


Figure 6: Top 6 Disney Characters

3.3.2 Best Movies

The core part introducing 30 best Disney movies is delivered by a timeline. The timeline contains the period from 1940 to 2020 which covers the whole history of Disney movies in our database. This timeline is based on the lolipop chart of d3.js. To show which movie has higher score, if the movie with a final score greater than 86, the circle for this movie will be located above the timeline, and vice versa. The distance from the circle to the timeline indicates how much the score of the specific movie is greater or less than 86.

When the mouse is hovered on the circle, information including movie title and movie ratings are shown in the tooltip. While clicking on the circle, the poster of the movie and a short generalization of the story are shown. What's more, the theme song of the specific movie will be played along. To pause the audio, just click the small icon on the top of the page.

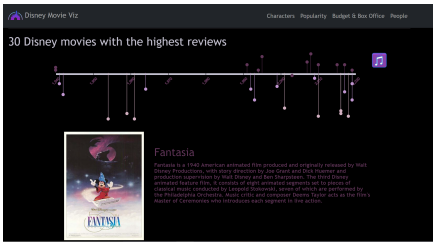


Figure 7: Final implementation of Best Movie Section

3.3.3 Finance

The finance section is based on the financial data (budget and box office). We changed the original plots shown in Figure 3 in order to better convey the information. What's more, we added a spiral plot to show the box office in time.

The first scene of the section presents the relationship between the budget of the movie and the final IMDB score. As the budget value could be biased by inflation, we add a range-slider so that the user could filter out the movies released during a specific period.

The second scene focuses on the ability of the movies to generate cash flow. A spiral plot was used to visualize the box office of all selected movies. The user could monitor both the historical cash flow and the frequency to release movies.

The scope of the third scene is to view the Top-10 movies according to different financial metrics, including profit, return on investment, and loss. The user could view the target list by clicking on the button for certain metrics.

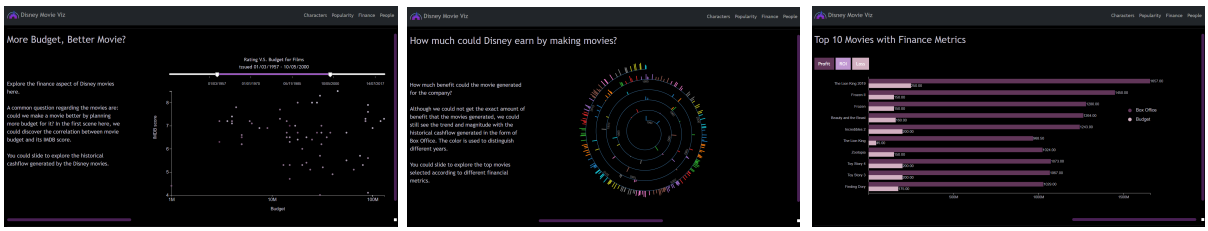


Figure 8: Different scenes for Finance data visualization. From left to right: Budget vs Score; Historical Cashflow; Top 10 Films

3.3.4 People

The people scene is divided into 2 different aspects with regards to directors, producers and music directors. Firstly, we analyze people's contributions to Disney films. For each of the staff groups, we extracted how many movies each of them has worked on and the IMDB score of each movies. IMDB scores are divided into different classes with a score of 8-9 being considered as high ranking and a score of 2-3 being considered as low ranking. Different class has a different color. A longer bar indicates more movies the person has contributed to and the user can see clearly the quality of the contribution by looking at the color of the stacked bar.

Then, we analyze the cooperative relationship between people. People with more contributions is shown to has larger node in the network and the thickness of the links indicates how strong the connection is. If the two person have worked together for more than two times, of course there is a thick link in between.

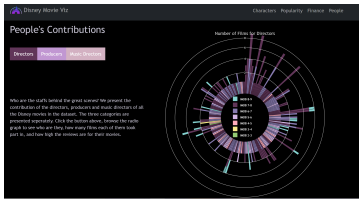


Figure 9: People's Contributions

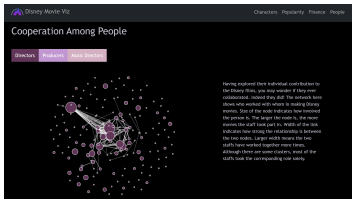


Figure 10: Cooperation Among People

4 Peer Assessment

For each milestones in our project, our group members had meetings on how we would finish the task and brainstormed many ideas that could make our project better. Afterwords, we divided these works to different parts and each of us handled one of them. The division of work between our group members is shown as following.

4.1 Lin, Yuanhui

M1: Exploratory data analysis
M2: Initializing the website skeleton
M3: Character section, people section and writing the process book

4.2 Zhou, Runke

M1: Data exploration and second half of the report
M2: Writing report of Milestone 2
M3: Best Movies section and writing the process book

4.3 Di, Yao

M1: Data explore and first half of report
M2: Draft and report MVP
M3: Finance section, Screencast of Viz