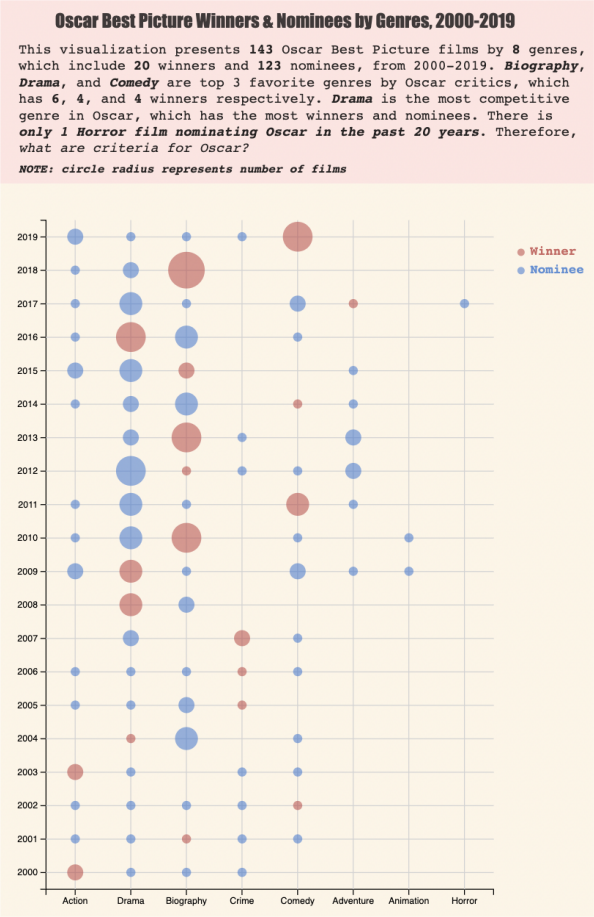
INFO 5100

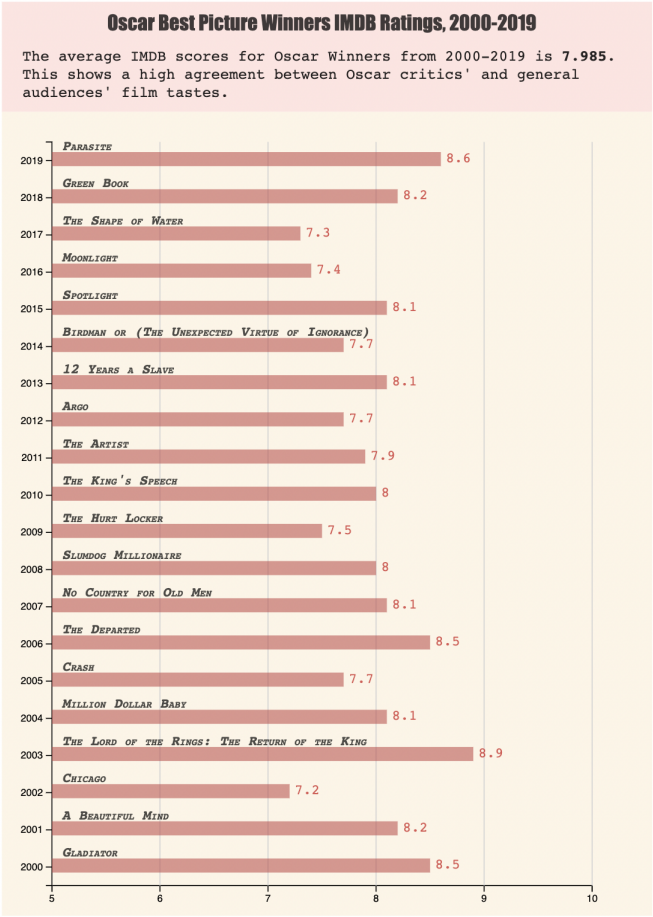
Project 1 report

Team members: Han Gao, Yuxin Kang, Pengwei Tian, Haixin Tang

Our outcome: Our team focused on data visualization of *Oscar Best Picture Winners and Nominees* from 2000 to 2019, as well as the IMDB score of those selected films. In detail, we were able to utilize the data to find whether there is a disagreement between average audiences and critics of prizes or not based on IMDB scores, the correlations between numbers of winners/nominees of each genre, and how competitive each of them is in the past 20 years.

We went through multiple design discussions and coding iterations to achieve our final design, please checkout the screen shot of these two visualizations that we created below.





Data Selection Process: We went through a traditional design-thinking process. Unlike other people who tried to start the whole process by finding examples of existing visualizations for different data sources, each of the team members began by locating their personal interest and looking for data that might have a potential in developing.

Different ideas were brought up, including *Changes to Forest Coverage Rate in Recent 10 Years Globally*, *World Happiness Report 2015-2019*, *StockX Sneaker 2019 Data Contest*, and so on. We gathered around and analyzed the pros and cons of each of the possibilities considering the complexity of the dataset, the cohesiveness of each individual data, and the possible visualization variations. We chose *Oscar Best Picture Winners/Nominees and IMDB Rating*, since those data are more organized and will bring unique perspectives of Oscar films. For example, unlike the StockX Sneaker 2019 Data Contest, most of the Oscar data are categorized by year and genre. It was way more convenient for us to retrieve the data and organize them into those 2 categories.

Our Desires: Through team corporations and observations, we found the desire to do this visualization as well. Looking at winners from each year made us wonder about the winner of the future. Which genre does it belong to? What's the possibility of it? How many of them would be nomineed? And What's people's reaction to those films compared to Oscar's selection? Do people see Oscar's choices as representatives?

Data Source: https://www.kaggle.com/martinmraz07/oscar-movies

Data Details: Oscar Best Picture Movie contains data categories like film, Oscar year, studio/production company, award, year of releases, movie time, movie genre, and their IMDB ratings. Data ranged from 1934 to 2019, with 570 indexes with over 564 unique values in film names. It was kind of sporadic when we first got the dataset. However, because of the fact that we already constructed a problem statement and our potential goal, we filtered some of the variables to make the dataset more manageable. We believed the filtered data was good enough to answer our question and those variables were effective enough to support our drawings.

Data Filtering Process: To make visualizations more visually acceptable but still with a wide range of details, we decided to only visualize films in a 20-year range, from 2000-2019.

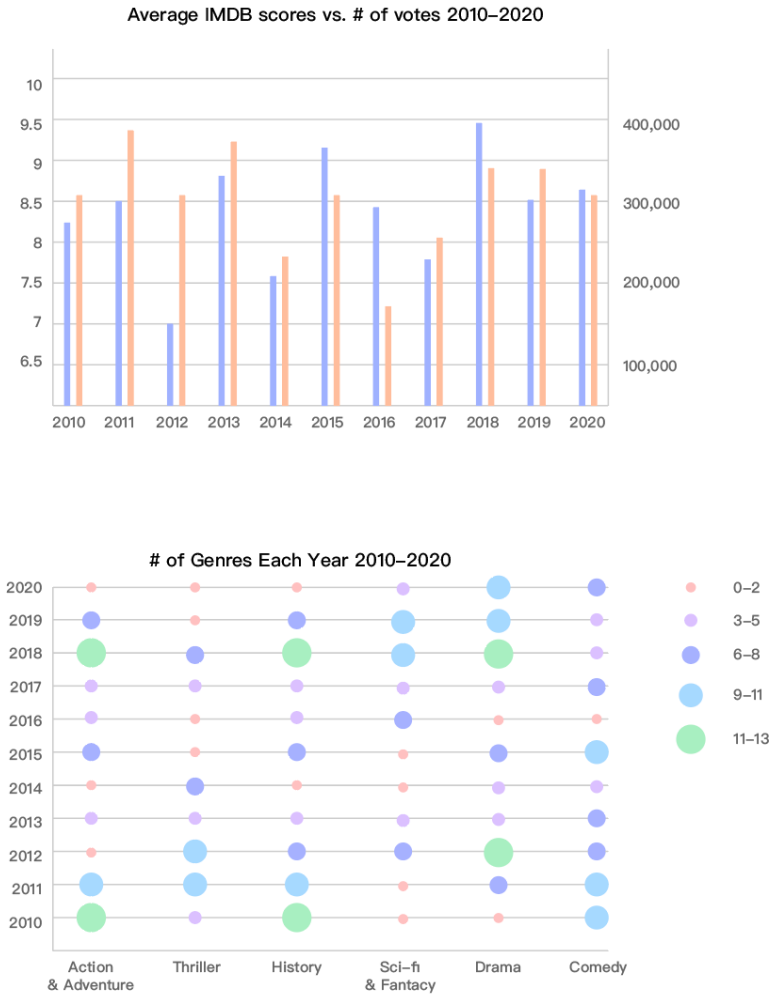
```
// Filter films from 2000-2019
data = data.filter( (d) => {return d["Oscar Year"]>=2000
                                && d["Oscar Year"]<=2019});

console.log(data);
// Get winners from 2000-2019
winners = data.filter( (d) => {return d["Award"] === "Winner"});
console.log("Winner: ", winners);
// Get nominees from 2000-2019
nominees = data.filter( (d) => {return d["Award"] === "Nominee"});
console.log("Nominee: ", nominees);
```

We used the ***d3.filter()*** function to select data from 2000 to 2019. Then, we created two array objects to hold filtering results of Oscar ***winners*** and ***nominees*** respectively, which help us get easy access to winners and nominees in the later stage of coding.

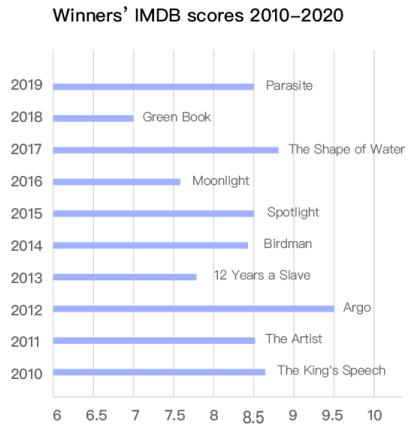
Design Sketching Process: Our actual design came in a sequential order, apparently. Eva started the initial design sketches to ideate how the visualization would look like. For that iteration, multiple design factors were brought up during the group brainstorming session, such as circle radius (r), transparency of different dots, color palette, and the placement of text. We believed our first version was on the right track, but we thought we were unclear on the definition of circle sizes, its relationship with the winners and nominees, the exact rating of the chosen IMDB Oscar winners, and grid line weights. Also, we believe using a 10 year span was not clear enough to visualize the complex data. Please checkout our iteration 1 below.

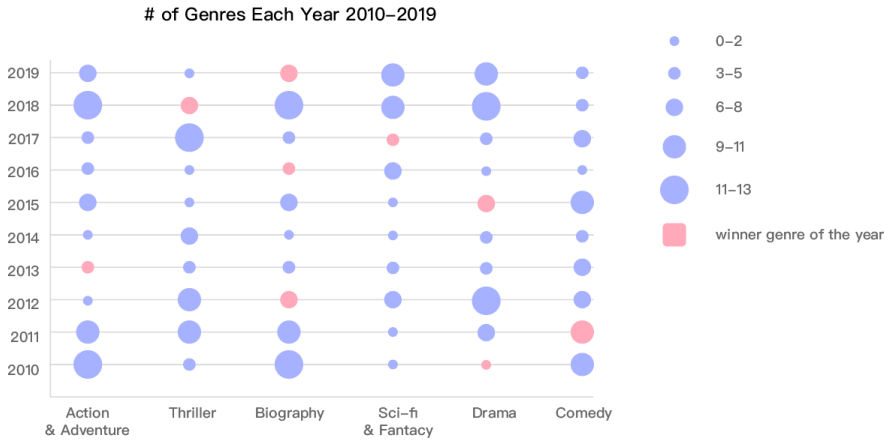
Iteration 1 Sketch



After the discussion, we decided to only present IMDB ratings of winners with the film title in the first plot. For the second plot, we utilized **circles** as **visual marks**. We selected several visual channels to convey variations of information: **vertical aligned positions** vary corresponding to Oscar years, **horizontal aligned positions** vary according to film genres, **color variations** distinguish between winners and nominees, and **circle radius** varies based on the number of films. Please checkout our iteration 2 sketch below.

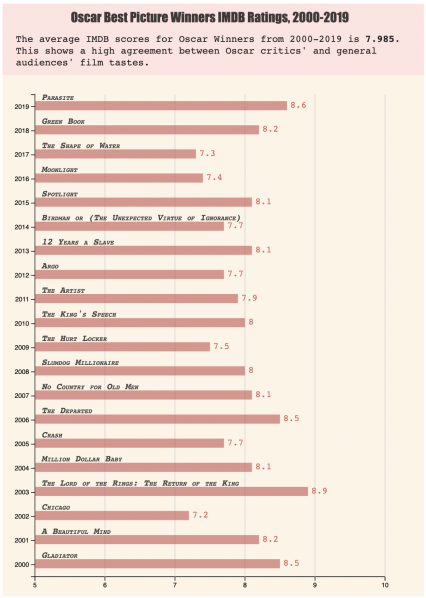
Iteration 2 Sketch





Final Outcome Rationales - First Plot: The final outcome of the first plot shows the IMDB Ratings of Oscar Best Picture Winners from 2000-2019.

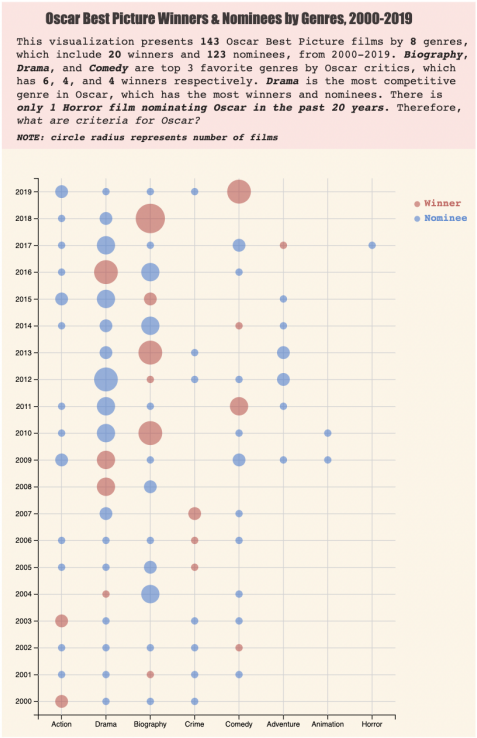
First Plot Final Version



- **Marks:** rectangles, gridlines
- **Channels:**
    - Vertical aligned positions vary according to the film's Oscar year.
    - Horizontal aligned lengths vary according to the film's IMDB ratings.
- **Colors:**
    - ***Canvas Background:*** we chose pink and yellow to color our canvas background to make the plot more visually appealing. We selected the low-saturations to ensure that viewers can read all labels and data easily.
    - ***Axes & Gridlines:*** we want to keep the graph simple, so we selected black as the axes' color. We want to ensure that data are visually visible, so we selected light grey as the gridlines' color.
    - ***Rectangle:*** we chose the pink color, #C14242, to color all rectangles and set the opacity to 0.55 to bring viewers a sense of unity.
- **Labels:**
    - ***Film Titles:*** we decide to include film titles as labels, so viewers can understand the corresponding film to each rectangle. If they want to watch the film with a high IMDB rating, they can easily find the film title above each rectangle.
    - ***IMDB Ratings:*** we also include the IMDB rating beside each rectangle to show the exact rating score of each film. We believe that a 0.5 difference in ratings means a lot. Since all winners have similar ratings, we want to present this subtle difference.
- **IMDB Extent:**
    - We didn't use [0, 10] as the IMDB scale domain, since viewers cannot visually see the difference among all winners. Instead, we set the domain as ***[5, 10]*** to visually show difference in winners' IMDB ratings.

Final Outcome Rationales - Second Plot: the final outcome of the second plot shows Oscar Best Picture films by genres from 2000-2019 with highlighted winners for each year.

Second Plot Final Version



- **Marks:** circles, gridlines
- **Channels:**
    - Vertical aligned positions vary corresponding to the film's Oscar year.
    - Horizontal aligned positions vary corresponding to the film's genre.
    - Circle colors vary to distinguish winners and nominees.
    - Circle radius varies to present how many films in the corresponding genre and year.
- **Colors:**
    - ***Canvas Background:*** we used the same background color as the first plot to give viewers a sense of unity.
    - ***Axes & Gridlines:*** we want to keep the graph simple, so we selected black as the axes' color. We want to ensure that data are visually visible, so we selected light grey as the gridlines' color.
    - ***Circles:*** we chose the same pink color, #C14242, as the first plot to color all winners, which gives the sense of unity. We selected a blue color, #2C77D2, to color all nominees to distinguish nominees with more visually visible contrast. We set opacity to 0.55 for aesthetic consideration.
- **Labels:**
    - ***Winner & Nominee Label:*** we add two labels on the right side of the plot, so viewers can easily understand which color represents winners and nominees respectively.
    - ***Note Label:*** we add a note label on the top of the plot in the description to instruct viewers that radius varies according to the number of films to prevent confusion.
- **Genres:**
    - ***Data Processing:*** in the original database, each film is assigned with multiple genres in a string format, such as ***"Action, Adventure, Drama"***. Originally, we decided to take all genres of a film into consideration. However, via the discussion, we found that it was confusing to mark multiple genres for a winner in a year. Eventually, we decided to only take the primary genre into consideration. We used the function ***d3.map()*** to select all primary genres and store them into an array object.

```
// select all available genres
let genre = d3.map(data, d => d["Movie Genre"].split(',')[0]);
```

The Story: *Chicago* received the lowest rating at 7.2, and *The Lord of the Ring: The Return of the King* received the highest rating at 8.9. The average IMDB rating that we got for Oscar winners from that period is 7.985, which shows a high agreement between Oscar critics' and general audiences' film tastes. So, from those analysts, people do see Oscar winners as representative of great movies.

The second visualization presents 143 Oscar best picture films by 8 genres, which include 20 winners and 123 nominees, from 2000-2019. Biography, Drama, and Comedy are the top 3 favorite genres by Oscar critics, which has 6,4, and 4 winners respectively. Drama is the most competitive genre in Oscar, which has the most winners and nominees. Therefore, in order to win in this genre, the film producers really need to try hard. There is only 1 horror film that was nominated for an Oscar in the past 20 years. These analysts brought an interesting question: ***what are the criteria for an Oscar?***

The Next Step: It was an enjoyable first project. All of the team members were pleased to find a desire of what to do when doing the design iterations while looking into possible data variations. Also, fortunately, we were able to visualize the data on the graphs. In our final meeting, we talked about the possibility of enlarging the range of variations of the dataset. We may look into movie time, and find its relationship with the winning percentages. We believe that's the next step to take if all of us have the desire to keep looking into it.

Team Contribution:

- **Team Roles:**
  - Eva Kang -- Data Search & Design Sketch
  - Han Gao -- Coding & Presentation
  - Haixin Tang -- Visual Coding
  - Pengwei Tian -- Drafting Report

- **Time Distribution:**
  - Coding: 10 hrs
  - Drafting Report: 4 hr
  - Data Search & Design Sketch: 4 hr

- **Project Timeline:**

Date	Task	Team Member
Monday, 09/20/2021	1st Group Meeting Location: *Rhodes Hall* ● Get to know each other ● Assign tasks	Eva Kang, Han Gao, Haixin Tang, Pengwei Tian
Wednesday, 09/22/2021	Idea Brainstorm: ● Each person comes up with at least one project idea ● Write the idea *Title*, *Data Source Link*, and *Description* in the document	Eva Kang, Han Gao, Haixin Tang, Pengwei Tian

Thursday, 09/23/2021	Milestone 1 Document Editing & Submission	Han Gao
Friday, 09/24/2021	2nd Group Meeting *Location: Rhodes Hall* ● Choose the final idea for Project 1 ● Consider design elements	Eva Kang, Han Gao, Haixin Tang, Pengwei Tian
Saturday, 09/25/2021	Design Sketching	Eva Kang
Sunday, 09/26/2021		
Monday, 09/27/2021 2pm	3rd Group Meeting Location: *Rhodes Hall* ● Finalize Design Sketching ● Start working on the Milestone 2 Document	Eva Kang, Han Gao, Haixin Tang, Pengwei Tian
Tuesday, 09/28/2021	Coding	Haixin Tang Han Gao
Wednesday 09/29/2021		
Thursday 09/30/2021	Editing Milestone 2 Document & Submission	Han Gao
Friday 10/01/2021	Coding	Haixin Tang Han Gao
Saturday 10/02/2021		
Sunday 10/03/2021		
Monday 10/04/2021	4th Group Meeting Location: *Rhodes Hall* ● Weekly check-in with coding progress ● Discuss rationale for design elements	Eva Kang, Han Gao, Haixin Tang, Pengwei Tian

Tuesday 10/05/2021	Coding	Haixin Tang Pengwei Tian
Wednesday 10/06/2021	Finalize Coding Report Writing, Editing & Submission	Han Gao Eva Kang Pengwei Tian
Thursday 10/07/2021		
Friday 10/08/2021	Presentation	Han Gao