

FINAL REPORT

Jagadeesh Bandi (z2016628)

Data Visualization (CSCI 627)

Dataset Description:

The dataset used for this project is titled Academy Awards and contains historical data about Oscar nominations and wins. It records information about films, nominees, categories, and award outcomes over many years.

Key Characteristics:

- Name: Academy Awards Dataset
- Source: Provided in the project materials
- Format: CSV (Converted from raw data)
- Data Range: Covers Oscar ceremonies from 1927 to 2024
- Data Size: The dataset contains approximately 11981 records covering nominations and wins from over 90 ceremonies.

The Academy Awards dataset used in this project was provided as part of the project materials. It was originally shared in a raw format and was converted into a structured CSV file for easier data processing and analysis.

Attribute Name	Description	Attribute Type
Year	Year when the Oscars ceremony took place	Ordinal
Film	Title of the nominated or winning film	Categorical
CanonicalCategory	Standardized name for the award category (e.g., Best Picture, Directing)	Categorical
Winner	Whether the nominee won ("True" if won, otherwise empty/null)	Categorical
FilmId	Unique identifier for the film (UUID)	Categorical
NomId	Unique identifier(s) for the nominee(s)	Categorical
Class	High-level grouping of award type (e.g., Directing, Acting, Writing)	Categorical

1. The dataset uses **UUIDs (Universally Unique Identifiers)** for films and nominees. This ensures precise tracking and avoids confusion among films or individuals with similar names.
2. The **CanonicalCategory** field standardizes award names across different years to support better analysis.
3. The dataset primarily focuses on films, rather than individual awards for actors, actresses, etc., making it suitable for analyzing trends in film-level success.

Questions:

1. The overlap between Best Picture winners and Best Director winners across decades?

The visualization analyzes how often the same film wins both Best Picture and Best Director awards, and identifies which decades had a higher or lower overlap between these two major categories.

2. What is the relationship between the number of nominations and the number of wins for films?

The visualization explores the correlation between a film's number of nominations and the number of awards it actually won, specifically focusing on Best Picture-winning and Best Director-winning films. It highlights sweepers (many wins) and snubs (many nominations but few wins).

3. What are the other Categories the Best Picture-winning films also win?

The visualization shows how successful Best Picture-winning films are in other major award categories like Editing, and Cinematography, and identifies which categories have the highest win rates among Best Picture winners.

Visualization Design And Evolution:

We will talk about each task and visualization.

Task 1:

The first task aimed to analyze how often Best Picture winners also had their directors win the Best Director award, grouped by decade.

- Filtered the dataset to get all Best Picture winners by using CanonicalCategories

- Mapped each film to its associated directors using Nomid and created a mapping of films to director IDs(nomId).
- Created a mapping of winning directors year by year for Best Director awards.
- Compared each Best Picture winner's directors against that year's winning directors to check if the director also won Best Director.

Initially considered a basic diverging bar chart where bars extended in opposite directions for "Best Picture + Director" and "Best Picture only." However, the early version made it difficult to read the exact counts and confusing for decades with fewer awards.

Finally, use the **Stacked Bar Chart**.

1. **Color encoding** was applied:
 - **Green** for "Best Picture + Director."
 - **Red** for "Best Picture only." [**Legend also added**]
2. **Opacity filtering** was added:
 - When a decade is selected, its bars stay fully visible while other decades fade.
3. **Tooltips** were added:
 - Hovering over any bar segment shows the type and the exact number of films.
4. **Axes:**
 - x-axis: A band scale was used to represent each decade clearly.
 - y-axis: A linear scale was used to represent the count of (Best Picture and Best Picture +Director) winners in each decade.
5. **Interaction:**
 - Selecting a decade dynamically highlights the relevant bars while dimming others, enhancing user interaction.
6. **Marks:**
 - Line (bars)
7. **Channels:**
 - Vertical Position on common scale (x-axis → decade)
 - Color hue (award type: Best Picture + Director vs Best Picture only)
 - Size : Length
 - Opacity (interaction to highlight selected decade).

Select a Decade

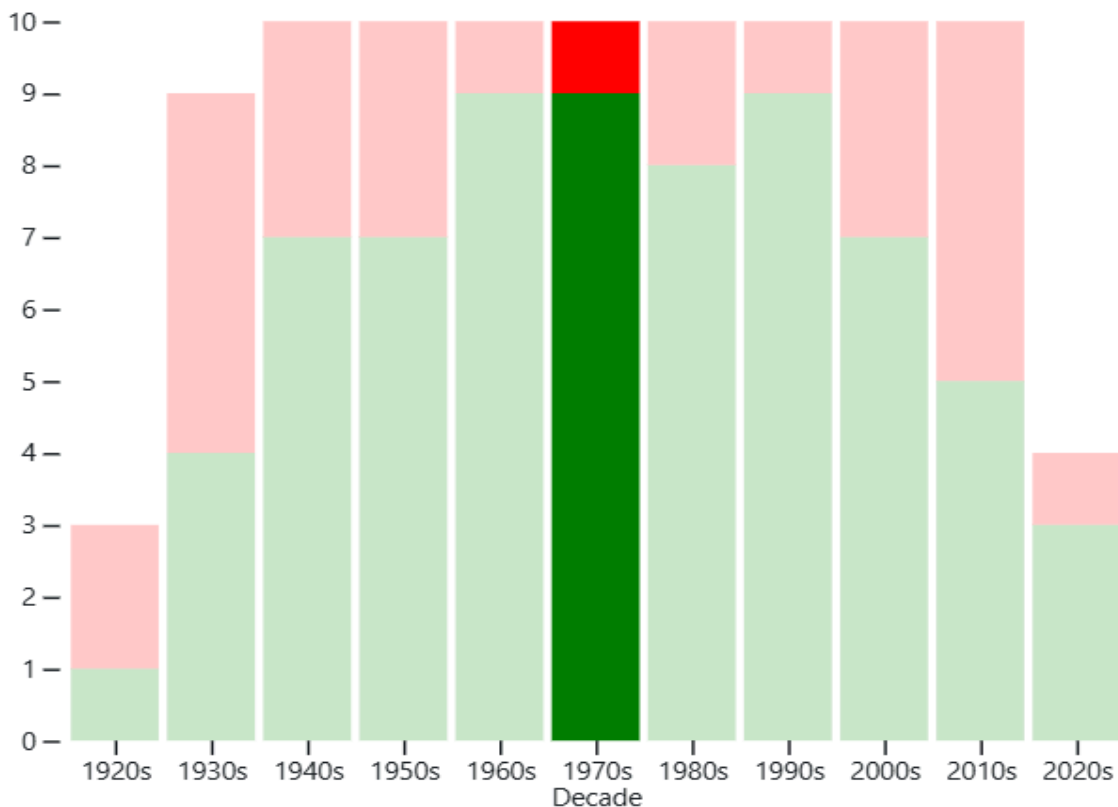
1970s



Task 1: Overlap Between Best Picture and Best Director Wins by Decade

■ Best Picture + Director ■ Best Picture only

† Best Picture Winners



Task 2:

The second task aimed to explore the relationship between the number of nominations and the number of wins for films, focusing on Best Picture winners and Best Director winners. It helps to identify:

- Sweepers (many nominations and many wins)
- Snubs (many nominations but few wins)
- Efficient performers (few nominations but good success)
- Filtered the Oscar dataset to include only films from the selected decade (based on a slider).

- Grouped the data by “FilmId” to compute for each film:
 - Total number of nominations
 - Total number of wins
 - Whether the film won Best Picture and/or Best Director
- Assigned a “category” attribute ("Best Picture + Director", "Best Picture Only", "Other").
- Calculated win ratio for each film.

Initially considered plotting simple dots in a scatterplot (nominations vs wins). However, many films had the same nominations and wins, leading to **overlapping points** that were hard to interpret.

Finally, Used a **scatterplot with additional rectangle grids (small square cells)** to show density:

- Rectangles (cells) were drawn for each nomination-wins pair, **colored based on the number of films at that spot.**
 - Overlaid circles (dots) for Best Picture and Best Director winners:
 - Color and stroke distinguish different winner types.
1. **Tooltips** were added:
 - Hover tooltips for Dots shows Film name, Year.
 - Hover tooltips for Rect(cell) shows Number of Films, Number of Nominations and number of Wins.
 2. **Interaction:**
 - Click to select a film (for linking to Task 3 visualization).
 - Selecting a decade this chart will only show the data of that decade only.
 3. **Marks:**
 - Area marks (rectangles for grid density)
 - Point marks (circles for individual important films)
 4. **Channels:**
 - Horizontal and Vertical Position on common scale (x-axis: nominations, y-axis: wins)
 - Color hue (grid color based on number of films; dot color based on film category)
 - Size (circle radius changes slightly on click interaction)

5. Overlay of Marks:

- Combined background grid rectangles with overlaid important circles for better insight without clutter.

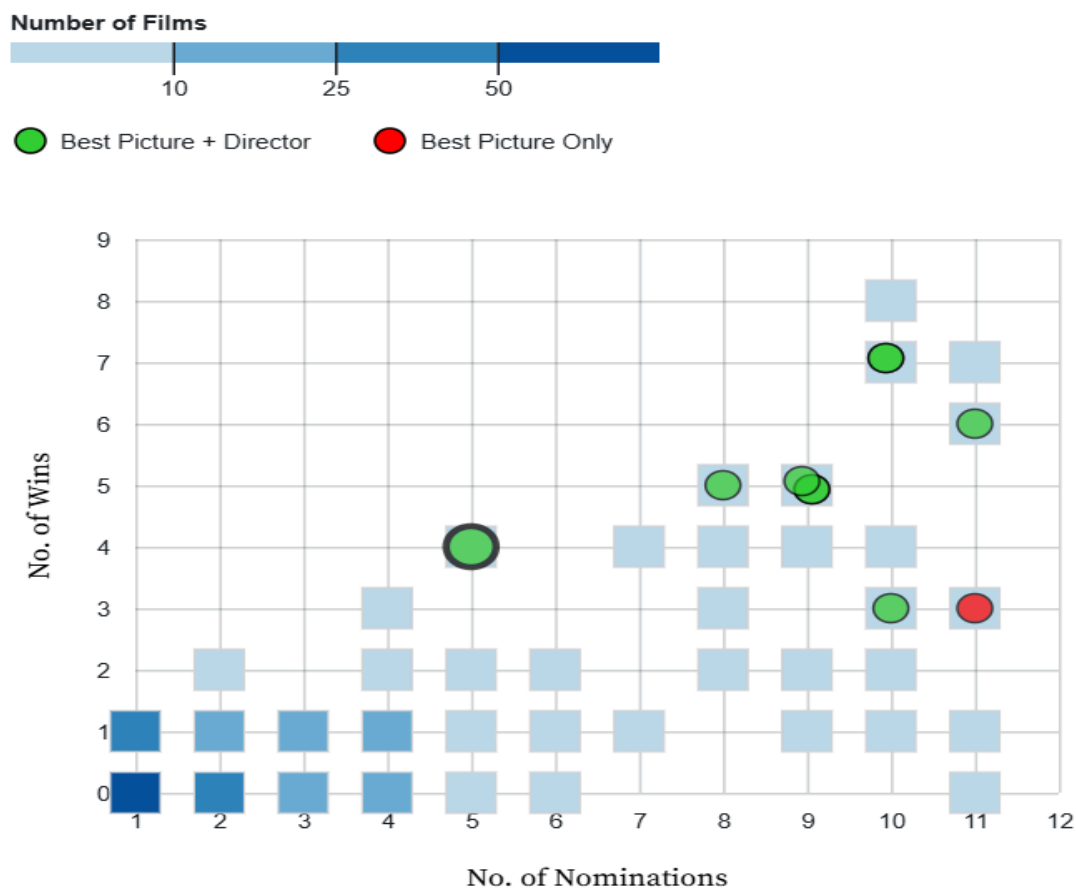
6. Legend Creation:

- Legends were created separately for both rectangles (density) and dots (categories) to make interpretation easier.

7. Jittering :

- Used Jittering to avoid overlap of dots(films).

Task 2: Film Nominations vs Wins



Task 3:

The third task aimed to analyze how successful Best Picture-winning films are in winning other major award categories like Directing, Editing, Cinematography, etc.

Filtered the Oscar dataset to include Best Picture winners from the selected decade (based on a slider).

Grouped the data by award category, and for each category:

- Calculated the number of Best Picture winners nominated.
- Calculated the number of Best Picture winners who won that category.
- Computed the win ratio for each category.

Prepared radar chart data by selecting the top 12 categories with the highest win ratios.

Initially considered using a basic donut chart to show win percentages across categories.

However, to **better show patterns across multiple categories** simultaneously, a **radar chart** was chosen. After that only the best pictures categories wanted to show. Later added click feature in task-2.

Finally, used a **Radar Chart**:

- The gold-colored area shows the **average win ratio** of Best Picture winners across categories per decade.
- When a film is selected (from Task 2), a **red radar line** is overlaid to show that specific film's win.

1. Tooltips were added:

- Hovering over category dots shows the Category Name and Win Ratio (%).
- Hovering over selected film dots shows Film Name and year.

2. Interaction:

- Selecting a film from Task 2 shows the red color line in Task 3.
- Selected film's wins are drawn as a red dashed radar shape over the golden category average.
- Filtered based on decade selected.

3. Marks:

- Line marks (radial paths connecting data points)
- Point marks (small circles at category axes)

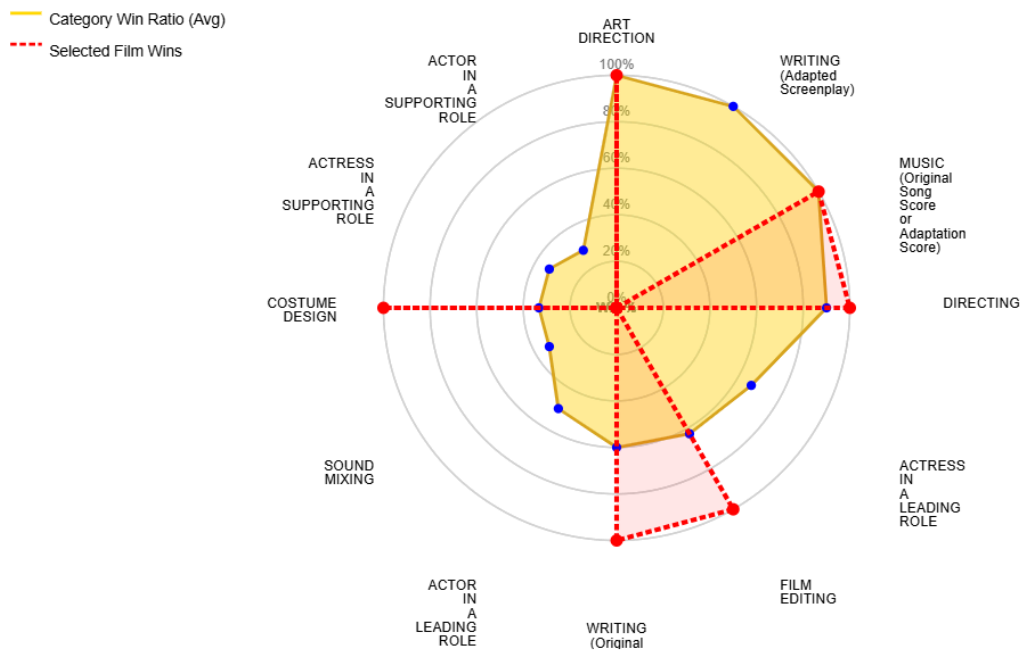
4. Channels:

- Angle (Theta Position): Represents award category (12 axes around the radar)
- Color hue:
 - Gold for overall category win ratios
 - Red for selected film wins
- Size

5. Legend Creation:

- Legend created separately for:
 - Gold line ("Category Win Ratio (Avg)")
 - Red dashed line ("Selected Film Wins")

Task 3: Award Patterns Among Best Picture Winners



Conclusion:

The stacked bar chart helps us understand how important a director's success is for a film. By showing how often Best Picture winners also won Best Director, we can see that in many decades, films with strong directors had a higher chance of winning Best Picture. This tells us that a director plays a big role in a film's overall success at the Oscars.

The scatterplot shows that getting many nominations does not always mean winning many awards. It helps identify films that performed very well (sweepers) and films that were highly nominated but won very little (snubs). It also shows that Best Picture and Best Director winners often have better win ratios compared to other films, showing how important these two awards are for a film's success.

The radar chart shows what other award categories help a film win Best Picture. Categories like Directing, Editing, and Cinematography have high success rates for Best Picture winners. This means winning in these important technical and creative categories greatly increases a film's overall strength at the Oscars. When selecting a film, users can see how it performed compared to typical Best Picture winners.