

Information Visualization

CHECKPOINT I: Visualization Proposal

G14

1. Domain

Our goal in this project is to study the impact of the political and social situation in the United States over the Oscars winners and nominees, from 1927 to 2015, by relating characteristics of the actors (gender, ethnicity) and movies (movie genre) with social and political variables such as the political party of the president, if there was a military conflict at the time, racial conflicts, and homicide rate.

2. Dataset

We are going to use the following datasets (for the Wikipedia datasets we will use web scrapers to perform data retrieval):

1. <https://www.kaggle.com/theacademy/academy-awards#database.csv> - this dataset contains Academy Award winners and nominees since 1927 to 2015.
2. https://www.kaggle.com/harshitagpt/us-presidents/#us_presidents.csv – this dataset contains US presidents since 1789 and their political parties.
3. <https://www.kaggle.com/grouplens/movielens-20m-dataset#movie.csv> – this dataset contains movie names and their genres.
4. <https://data.world/crowdfunder/academy-awards-demographics/workspace/file?filename=Oscars-demographics-DFE.csv> – this dataset contains the demographics of all Oscar winners, including their ethnicity
5. <http://labs.time.com/story/oscar-diversity/?fbclid=IwAR3ogD4C4yiDdetf2WMTnPhcQsQkF52Zy4l7BR3a7-vc066vRuGkgYpe-Y0> – this web page contains a chart showing the ethnicity (white or non-white) of Oscars winners and nominees for the awards of Best Actor, Best Actress, Best Supporting Actor and Best Supporting Actress, since 1928.
6. https://en.wikipedia.org/wiki/List_of_countries_by_intentional_homicide_rate_by_decade – this web page contains the homicide rate in the US per year
7. https://en.wikipedia.org/wiki/List_of_wars_involving_the_United_States – this web page contains a list of wars involving the US and corresponding years
8. https://www.history.com/topics/black-history/black-history-milestones?fbclid=IwAR3YODpITj_NPr-fw_BJIE2lkKTEpjMnfM8Wt2jF5NRpYLA6emX1vy2NGM – this web page contains a list of black history milestones per year.

3. Example Questions

To answer these concrete questions, we'll take the following into account:

- Regarding ethnicity, we will consider the winner and the nominees separately to study the ratio of non-white winners vs nominees. Regarding gender, we will only

analyse gender inequality in the Best Director award, since all the other awards that go to a single person distinguish gender

- We consider the social and political events that took place in the year to which the Oscars ceremony refers.

Our questions:

- **Question 1:** How are military conflicts involving the USA at the time reflected in the movie genre of the Oscars winner for Best Picture?
 - For each year, from dataset 1 we find the Best Picture winner for each year; from dataset 3 we will find the genres of the best picture winner; from dataset 7 we will find the wars per year
- **Question 2:** How is the homicide rate reflected in the movie genre of the Best Picture winner?
 - For each year, from dataset 6 we have the homicide rate; from dataset 1 we find the Best Picture winner; from dataset 3 we will find the genres of the best picture winner;
- **Question 3:** How does the ethnicity of the Oscars winners and nominees change over time?
 - For each year, from dataset 1 we get the winners and nominees, from dataset 4 we get the ethnicity of Best Director winners and nominees; from dataset 5 we get the ethnicity of the winners and nominees for the awards of Best Actor, Best Actress, Best Supporting Actor and Best Supporting Actress
- **Question 4:** How does the ethnicity of the Oscars winners and nominees change with the party in power?
 - Same as question 3 plus from dataset 2 we get the party in power in that year
- **Question 5:** How does the ethnicity of the Oscars winners and nominees change with racial conflicts?
 - Same as question 3 plus racial conflicts or achievements in that year from dataset 8
- **Question 6:** How does the gender ratio of the Oscars nominees for Best Director change over time?
 - For each year, from dataset 1 we get the Best Director nominees

4. Data Sample

(from "database.csv") from dataset 1
year; ceremony; Award; Winner; Name; Film
1927/1928; 1; Actor; ; Richard Barthelmes; The Noose

(from "us_presidents.csv") from dataset 2
S.No.; start; end; president; prior; party; vice
1; April 30,1789; March 4,1797; George Washington; Commander-in-Chief of the Continental Army (1775-1783); Nonpartisan; John Adams

(from "movie.csv") from dataset 3
movieId; title; genres;
1; Toy Story(1995); Adventure|Animation|Children|Comedy|Fantasy

(from "Oscars-demographics-DFE.csv") from dataset 4
race_ethnicity; ...; person;
White; ...; Tommy Lee Jones;