LEARNING ABOUT COUNTRIES THROUGH CINEMA

Data Visualization project

By Gonxhe Idrizi, Yann Bouquet, Emma Lanctuit

GitHub: https://github.com/com-480-data-visualization/com-480-project-lesbigleux

EPFL

TABLE OF CONTENTS

RESEARCH PROJECT

Motivation

We decided to turn ourselves to the theme of cinema, a theme that the three of us particularly enjoy. Cinema is for us one of the many fascinating environments which is constantly evolving. It is such a powerful tool that allows storytellers to bring their beautiful narratives to life. If a picture is worth a thousand words, surely a film is worth a million. By the late 19th century, the movie camera was born and allowed the creation of the many moving films that can be seen on screen. It has been evolving through a lot of genres, allowing many movie lovers to feel many emotions by having different experiences. As this environment has evolved through a whole century, many countries have been having their own movie trends. As this area includes many subjects, we wanted to get away from the visualizations that are overlooked and off the beaten track, such as discrimination or gender inequalities in the distribution of awards.

We then thought of using the world of cinema as a tourism indicator for each country. Through various articles, we were able to observe a correlation between a country's cinematic influence and its tourism. This phenomenon of fan tourism, where people have learned about the existence of a landscape of one country through a particular movie, is now a trend that increases the popularity of the concerned country. Furthermore, this cinematic influence leads to an increase in tourism of the country, which is often beneficial to its economy and development.

Constraints

We were unable to find a dataset that would allow us to link these two aspects of cinema: the country of production to the tourism brought by its publication and its recognition within the many different prestigious film ceremonies. That is why we chose, for this project, to work on the correlation between the country of production and the ceremonies where it has been present (and sometimes rewarded). The part showing the temporal links between cinematographic rewards and the growth of one country's tourism caused by the cinema industry is therefore not part of our visualization.

Our final idea

Based on these questions and analysis of our datasets, we decided to redefine our main question. Therefore, our different visualizations answer the following question:

How have a country's film trends evolved over time through the prism of ceremonies and rewards?

OUR DATASETS

Our visualizations are explorations of the IMDB awards and movies datasets. Most frequent and relevant cinematographic events are part of the first dataset, which contains awards from 1886 to 2020. These events will be joined with more than 80'000 films information contained in the second dataset.

During the exploration of our datasets, multiple questions were raised that we tried to answer through some visualizations. There are two main endpoints in our visualization: film producing countries and events with their respecting awards.
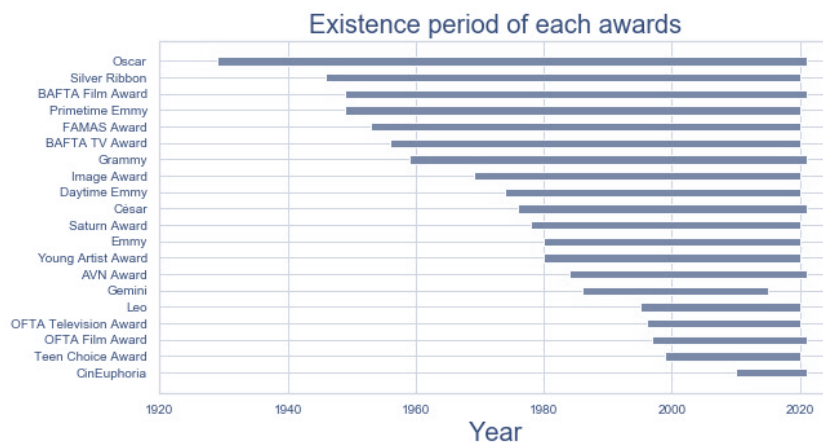

The feasibility of our visualization considers one century of evolution over Cinema Awards.

What is the existence period of each event pre-selected till these days?

We decided to keep both winners and nominees as a nominated movie will still have more visibility than one without a price. As we link our award dataset to our film dataset in purpose to match award to film-producing countries, we have to make sure that we have enough countries represented in our IMDB dataset to conduct an interesting study.

How many movies are represented per country in the IMDB dataset?
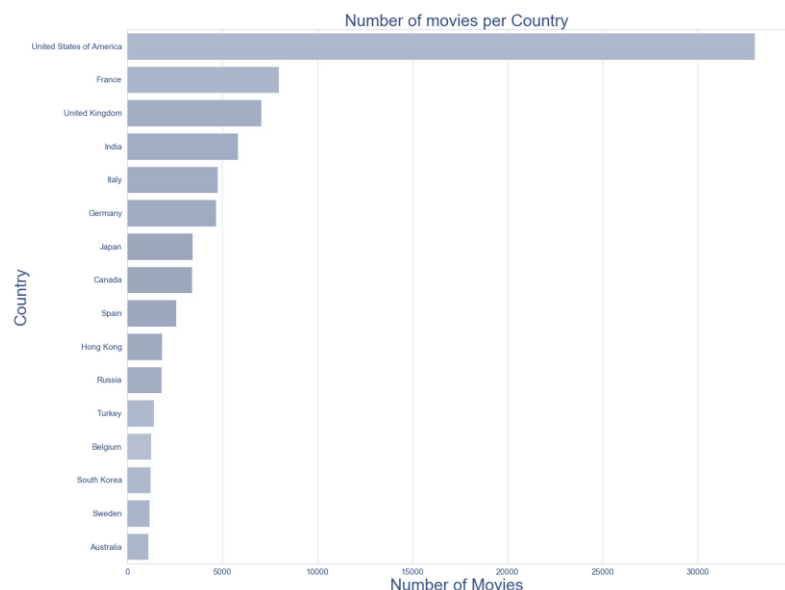
DESIGN PROCESS

Goal: Analyze a country's film trends based on the prestigious awards it has received over the years.

CONTEXT
1

Which countries are represented in the film world and at which ceremony? This question will be our context setup (first page of the website) in order to understand the distribution of the countries' film production awarded by the most famous ceremonies throughout time.

EXPLANATORY
2

After choosing a country, what movie genres have been produced by a country over time? Has there been an emergence of new film genres per country over time? Are these changes linked to new awards received: is there any correlation?

CONCLUSION
3

After the general information accumulated over all countries and awards, one may ask himself how the repartition of countries linked with the film genres did evolve through time based on if the movie was nominated/has won in prestigious film ceremonies.

Part 1 - Context

Goal

The context of our website is dedicated to the following questions: **How are the countries and film award ceremonies related? What kind of awards are given at which ceremony? In other words, how is the ceremony defined?**

To answer the first question, we wanted to **represent the relation by year between countries and events through a Bipartite Graph**. Indeed, the first sub-part of nodes represents 8 of the most recognized events on an international scale and the second sub-part represents the 20 countries most represented by the films competing in these events. A link exists between an event and a country if at least one film from this country has been nominated at this event. Moreover, we have a **selector to generate the graph** of any year between 1929 and 2020.

To be able to answer the second question, we propose to the user **to click on any node of the graph** (in any pre-selected year). If this node corresponds to an event, **a Sequences Sunburst appears on the right and represents all prizes awarded during this event at the selected year.**

We thus analyze distributions of these prizes, the winners among the nominees, the films compared to the winners, and finally the categories. This allows the user to understand what the event is all about and how the films of the countries are represented in the event and at a particular year.

The user can interact in several ways with these two visualizations:
- a node can be **moved** to be isolated in the area provided for this purpose. If it is in the appropriate area (or next to it) the node will not move after dragging it. If you try to move it to the side of its initial group (an 'event' node on the side of the 'event' nodes for example) it will be **automatically placed** next to the nodes of its group once the mouse click is released.
- **Passing the mouse** over a sunburst generated via an event node will allow to **select** a sequence that will be summarized above the sunburst.

In this way the user fully understands through these visualizations how countries and events are connected through films and why we wonder how these events can reflect the film trends of the countries. This contextualization thus leads us to wonder how the film trend of a country has evolved over time in comparison to the evolution of its recognition in film events.

Challenges

Initially, we didn't want to restrict the number of countries represented by the network. Moreover, in order to be able to select the year, a timeline allowing a *"brushing"* to select a year or a period was proposed. After some analysis, we realized that the number of countries represented may be too large to have a readable network on the website. This is the reason why we decided to limit the number of countries to 20 by considering the most represented countries (in number of nominations in the events) in each defined period. This solved the problem of readability over one year. However, if we take a period of ten years for example, we could see that the graph was therefore almost fully connected. So, by abandoning this idea of period, we focused on selecting networks by individual years. Consequently, modeling a timeline was of too little interest for the time available. We finally opted for an HTML selector to select a year.

The number of awards given depends on the ceremony and the year. So, having a fixed list of colors to represent these awards brought a lot of complication. Indeed, taking two close colors for two different awards makes sense if a lot of awards are involved. On the other hand, it doesn't make sense when we don't have many. Therefore, this color palette must be adapted each time to not have something disturbing for the eyes. We decided to use Chroma js to generate those dynamic palettes.

One thing to notice is that we decided to provide information over **every year**, meaning since the creation of the Academy Awards ceremony. Hence if the user is willing to observe the **evolutions**, he can go through different years. That is why those visualizations don't propose any way to compare networks or sunbursts through time. Indeed, our desire to keep the visualization clear for any user did not allow us to find a solution that would satisfy everyone's curiosity.

Part 2 - Explanatory

Goal

The purpose of this section is to push the user to discover by himself **the correlations between genres produced in a country and the awards**. We want him to be the actor in the discovery and let him wonder and ask himself questions. We push him in this direction by giving him some examples of readable stories from the visualization.

We have therefore decided to create **two correlated visualizations**:
- Genres produced by country over time to explore trends
- Evolution of number of awards owned by the same films

For the pre-selection, we have chosen to show all the films and therefore the trend of the genres and awards in the world through time. By this, we realized that it was not as readable as we expected. Therefore, we added **a brush** allowing to "zoom" on a desired period. In addition, we allowed the user to **highlight** a specific genre and its trend.

Finally, we thought it would be useful to give the user the possibility to **observe the proportion** of one genre in relation to the others over time and, at the same time, the proportion of the awards received. That is why we have added **two buttons** - ratio type - which allow to select the desired type of graph.

Challenges

We have created a **selector** to let the user choose his country. This had two main challenges:
- **Retrieving** the countries present in the dataset as to present them in the selector.
- Making the **link** between the selection and the update of the graphics.

In the visualization, we had to reduce the number of genres represented as it made the visualization less clear. We have chosen to keep all the countries that have received at least one award in a famous event.

Part 3 - Conclusion

Goal

This visualization is intended to allow one user to see how the pre-selected countries linked with the film genres did evolve through time based on if the movie was nominated or has won in at least one prestigious film ceremony. **This way one user will be able to capture a country's prevailing film trend throughout ranges of time.**

One funny and interactive way to explore these countries trends is by using the parallel sets visualization. It will allow us to visualize and compare the proportions of the evolution of genres

and success of movies per country. Three different attributes are used in this graph to show the relations between them: the countries, the genres of movies and if the movie has won a price and/or if it was nominated. The user can **hover** over the different links to get the proportion values he wants and can, as well, **reorder** the groups per attribute to get other insights.

To be able to see the evolution of these relations through time, the user has the **possibility to use a timeline to select a specific time range** he wants to inspect and therefore compare the proportions of the relations with other time ranges.

Challenges

Many challenges have been encountered during the creation of this visualization. Indeed, the datasets we used included 180 countries and 26 different genres. We tried to present at least one impacting country in the cinema per continent so that the graph was sufficiently visible. This gave us 7 interesting countries to analyse. Indeed, having to many coloured crossing-lines would have been unfavourable for the good visualization of the proposed model. We also **pre-selected the genres** that stood out the most in the trends, meaning 7 of them, of each period as to have a better and more correct comparison per time ranges. To be able to see the same information in all the visualizations, we had to reduce the number of genres in the same way as in the exploratory part.

We planned at first to present the movies that didn't take part in the prestigious film ceremonies, but no dataset had all the information that we needed, meaning the genre as well as the country of production of a movie.

Tools for all visualizations

D3 disjoint force-directed graph, D3 sequences sunburst, Chroma js, D3 parallel sets, D3 brushing,

PEER ASSESSMENT

To conclude, we are very happy of the development process we have shared together. We've all participated in the creation of our core goal, presenting each of us different ways of analysing these datasets. We've been keeping in touch every week, even while being confined, to see how each of us was making progress and what difficulties we've been encountering. This gave us the possibility to share our own experiences by helping each other when having different issues.

Each one of us developed one individual visualization to answer the goal of our project. Yann implemented mostly the context part, Emma implemented the explanatory one as well as the skeleton of the website and Gonxhe implemented the conclusion visualization. We've had a critical eye on each visualization to be sure that it answered the primary idea and helped each other to add more value to its graph. We respected each other's ideas and had fun during all the development of this project.

RESOURCES

Datasets :

- IMDB Awards: https://www.kaggle.com/iwooloowi/film-awards-imdb
- IMDB Movies: https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset