

Sociopolitical Impact At The Oscars

João Henriques

Lisbon, Portugal

j.tiago.henriques@tecnico.ulisboa.pt

Leonor Llansol

Almada, Portugal

leonor.llansol@tecnico.ulisboa.pt

Pedro Esteves

Lisbon, Portugal

pedrofonsecaesteves@gmail.com

Instituto Superior Técnico

ABSTRACT

In this report we present a project done for the Information Visualization (InfoVis) course at Instituto Superior Técnico. The project consists in developing an infoVis that allows users to understand and compare how different factors affect a chosen theme. We decided to study the sociopolitical impact at the Oscars, because of all the influence this industry has in our lives, this being one of the most viewed ceremonies worldwide.

Using programming languages as HTML, CSS, JS and the d3.js framework we developed our infoVis, that will help people comparing the impact of these factors in the Oscars.

Author Keywords

Information Visualization, InfoVis, Oscars.

INTRODUCTION

The goal of this project is to study the impact of sociopolitical features over the Oscars winners and nominees, from 1928 to 2015, by relating characteristics of the actors (ethnicity, sexual orientation and birthplace), with social and political variables such as the political party of the president and Black History milestones.

Today, in the 21st century, we live in an era where human rights and equality among all people are widely discussed, regardless of their ethnicity, gender or age. Being Oscars, a ceremony followed all around the world, it is important to understand if there is no discrimination in the movie industry.

In the current market there are tools that make possible to see patterns in Oscars movies or compare the winners of certain categories along with the budget of the movie. In our InfoVis it is possible to analyze each year, or a group of years, and understand if the change of political party, or a “Black History” milestone changed in an impactful way the ethnicity of the nominees. With that we pretend to answer the following questions:

- **Question 1:** How does the political party in power in the USA influence the sexual orientation of nominees?
- **Question 2:** How does the political party in power in the USA influence the birthplace of Oscars winners?

- **Question 3:** How does the ethnicity of the Oscars winners and nominees change over time?
- **Question 4:** How does the ethnicity of the Oscars winners and nominees change with the party in power?
- **Question 5:** How does the ethnicity of the Oscars winners and nominees change with Black History Milestones?

These were all the questions defined in the Checkpoint 1, plus the question “How does the gender ratio of the Oscars nominees for Best Director change over time?”, this last one we decided to take out, because it was the only question talking about Best director category.

Currently, to answer these questions, is necessary to go Wikipedia and compare different tables and datasets by hand. We present here, a much more fast and enjoyable way of getting these results.

RELATED WORK

There are some visualizations available on the Internet which concern the Oscars. There is one, “Box office or best picture? Can you really have both?” [15], that uses information visualization to find if there is a correlation between movies that win the Best Picture Award and movies that made the highest amount of money. Also related to profit, [16] made a visualization to find out if a movie having an Oscar nomination affects its profit. We also saw visualizations regarding the racial diversity of the Oscars, which we studied in this project, such as [17], which shows how the Oscars are dominated by white people. This visualization was made on 2016, where all 10 nominees for the awards of Best Actor and Best Actress were white. Another one is [18], which answers the question “How diverse is the Oscars?” by showing the ethnicity of all winners and nominees in the main categories since 1927.

To get inspiration for the idioms we wanted to use to answer our questions, we looked through examples of idioms done using d3.js [2-4], which is the JavaScript library that we use to develop our project. We also got some inspiration by looking at the class’s Hall of Fame [1], in particular from the projects “Death and Its Friends”, “Which country fits your personality the best?” and “Global Peace Index”, which we thought had the closest design to what we wanted to accomplish.

THE DATA

Being this project about the Oscars, we thought that would be easy to get a dataset or a few datasets, with all the information necessary. However, we quickly understood that wasn't the case. All the necessary information was available on the internet, but not all could be accessed in "a click". It was necessary to merge different datasets, scrape Wikipedia pages and search some information about Oscars nominees by hand, like birthplace and ethnicity. All the data selected, was available online from reliable sources. The following table contains the data selected, where and how we got it.

Website Name	Dataset Content	Way of getting the content
Kaggle [5]	Academy Award winners and nominees from 1927 to 2015	Csv Download
Kaggle [6]	US presidents since 1789 and their political parties	Csv Download
History Chanel [8]	Black history milestones per year	Copied by hand
Wikipedia [9]	List of LGBT Academy Award winners and nominees	Web scrapped
Google / Wikipedia	Birthplace of each nominee	Searched and filled by hand
Data.world[7]	Demographics of all Oscar winners, including their ethnicity and birthplace	Csv Download
Time Labs [10]	Ethnicity of each nominee (White)	Copied by hand
Google / Wikipedia	Ethnicity of each nominee	Searched and filled by hand

Table 1. These table contains the Websites and the respective content used to our project. It also describes how we get the content.

After getting all the needed information stored in csv files, it was time to clean and merge the datasets. We accomplished that through the use of Pentaho Data Integration (PDI) transformations, to remove columns, sort the dataset by a column, rename column names, merge datasets through common columns, create new datasets, added a sentinel values to empty cells, and trim strings.

After having all the data merged, small problems in the datasets still had to be fixed, like some mistakes in Wikipedia list of presidents, and some strings that stayed unformatted due to special characters. We had the birthplace of the nominees, including the country and the state of the USA nominees but we needed that information grouped by continent, that, we did by hand with help of PDI transformations.

Due to data restriction we divided the ethnicity in only four categories: White, Black, Asian and Hispanic. We also only consider four sexual orientations: Straight, Bisexual, Gay and Lesbian. And we only have the data from the following categories: best actor/actress, best supporting actor/actress and best director.

For the checkpoint two we also acquired more data to answer the question 6, that after it was removed for the final solution as explained in the Introduction.

VISUALIZATION

In this section, it is described with detail the InfoVis, as well as explained each design and implementation choice.

Overall Description

Our visualization has the layout showed in Figure 1.

The layout is composed by one control section and five idioms, that are briefly described in the list below:

- A **Multi-Line Chart**, that shows the percentage of the nominees ethnicity by year, in the top right corner;
- A **TreeMap**, that encodes the birthplace of Oscars nominees, in the top middle;
- A **Grouped Bars Chart**, that displays the sexual orientation by political party in power, in the top left corner.
- A customized version of a **Lollipop** chart, that shows the Black History milestones, in bottom left corner;
- A **Radar Chart**, that displays the ethnicity percentage per political party in power

The layout works as a static dashboard, in terms of resize to screens device dimensions, but all the idioms can be changed by interact with other idioms, or by using the controls section. The following subsections will present in detail the individual idioms and how they interact with each other.

To have a better contrast between the colors used and the background, we decided to use a dark background and consequently a white color for text and axis. We also decided to use pastel colors to facilitate the long-term use of our application.



Figure 1. Overview of our layout.

Ethnicity Percentage Over the Years

To see how the ethnicity of the Oscars winners and nominees' changes over time, a Multi-Line Chart was implemented (Figure 2).

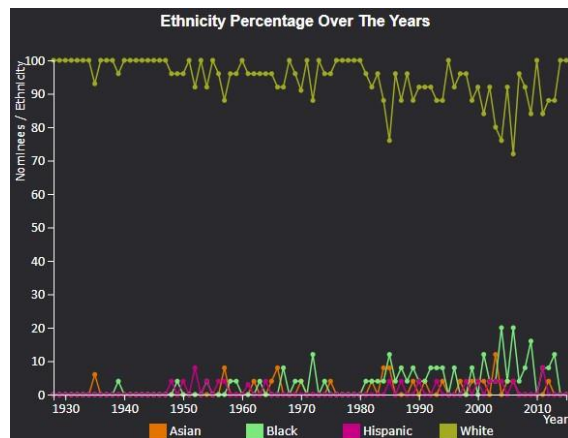


Figure 2. Multi-Line Chart

We chose this idiom because it is well suited for representing the variation of information over time, and can be used to compare different ethnicities simultaneously. It also shows very well the discrepancy in the values between white people and the other ethnicities.

In this idiom is possible to mouseover one line, highlighting the selected line and changing the opacity of other lines to improve visibility of the one selected, it does the same to the legend. It is also possible to mouseover one circle, increasing the size of the circle and appearing one tooltip with information (year and percentage of nominees of the respective ethnicity) about this point. Besides that, it is possible to select one ethnicity or more, by clicking in the line or in the legend square, respectively, that will change the info showed in other idioms.

The X axis of this chart scales according to time range year selected in controls.

Oscar's Nominees Birthplace

To see the birthplace of the nominees, we decided to use a TreeMap (Figure 3).



Figure 3. TreeMap

This idiom shows the information in a hierarchical way, allowing us to see the birthplace of the winners/nominees with different granularities (state/country/continent). The areas of each rectangle help us compare the difference in the amount people chosen between different locations. Since there is an overwhelming difference from certain places to others, those that had a small amount were included in the rectangle "Others". Another chart that would allow us to see the same type of information, like a choropleth map, would have a significant portion of unused space.

In this idiom, it is possible to select each rectangle (if it is not a leaf) to access a nested view. By using this idiom, it is possible to see the difference between continents, countries,

and US states. A mouseover will create a tooltip with the precise value of nominee's birthplace of the respective rectangle. To go one level up, just click on the bar with the darker purple.

Political Parties in Power Impact On Sexual Orientation

We wanted to see the difference in the sexual orientation of the nominees, and if it is influenced by political party in power. So, we decided to use a grouped bars chart (Figure 4).

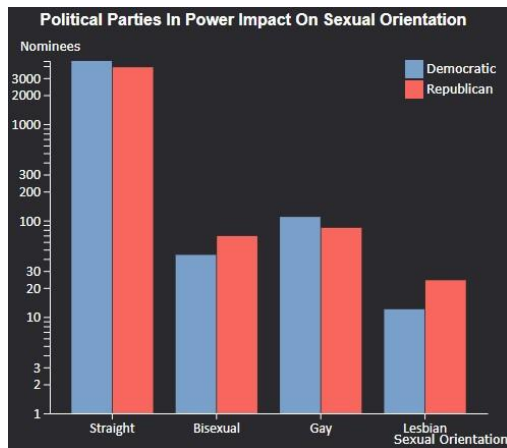


Figure 4. Grouped bars

This chart was chosen because it facilitates the visualization of different perspectives for the same variable, where the different perspectives are the political parties and the variable is the sexual orientation of the nominees. In this chart each color represents the color of a United States party. A mouseover over each bar will show a tooltip, giving the number of nominees encoded by the respective bar. Clicking on a bar will hide the other bar and increase the width of the bar selected. It is also possible to select or unselect one party by clicking in the legend square. Changing the party in this idiom will affect all the idioms.

Selecting a party and the time range to one interval where the selected party wasn't in power, will make the idiom without bars. This is not a bug, it is due to the fact that there is no data to show on this year. Also, if the selected interval has only one party in power, unselecting the party is disabled.

Since there are a lot more straights than the other sexual orientation classes, the Y axis has a logarithmic scale. The axis scales with the biggest value presented on the idiom.

Black History Milestones

To showcase the milestones that mark the struggle and the evolution of ethnical acceptance/conflicts in the USA regarding black people, we decided to make our own version of a Lollipop Chart (Figure 5).

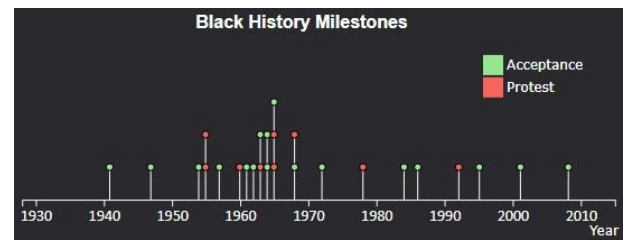


Figure 5. Lollipop

There are two types of milestones, the “Acceptance”, represented by a green circle, and the “Protest”, represented by a red circle.

The most important thing of this idiom is the position in our layout. Positioned right below the Multi-line Chart, it allows a better comparison of whether the milestones affect the percentage of nominees of a respective ethnicity in the same or the next years. Also, to facilitate this comparison, it is possible to make a mouseover that will create a vertical white line in the Multi-Line Chart in the respective year. It will also appear a tooltip, with the date and name of the milestone, as it is possible to see in figure 6.

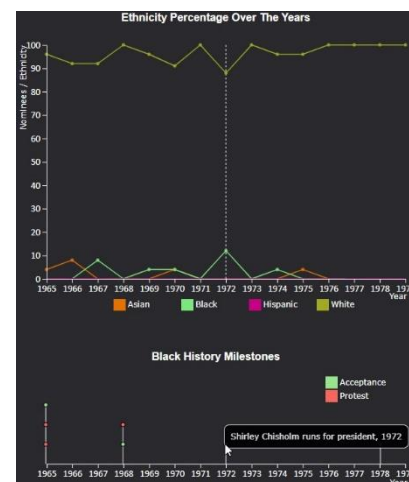


Figure 6. Lollipop mouse over, tooltip, and vertical line in Multi-Line Chart, selected years: 1970 – 1991.

This chart also scales with the time range of years selected in the controls.

Ethnicity Percentage Per Political Party in Power

To visualize the impact of each political party in the ethnicity of nominees, we decided to implement a Radar Chart (Figure 7).

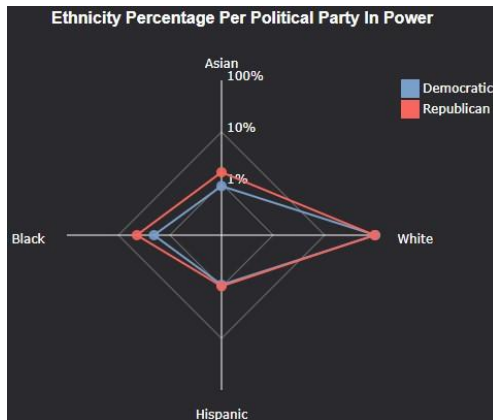


Figure 7. Radar Chart

This chart allows a quick comparison between each ethnicity, since it is only necessary to compare the position of each point. To improve the visibility, we didn't add a color to the polygon area. In this idiom, it is possible to mouseover each circle to get the precise percentage of each ethnicity. It is also possible to select the party by clicking on the legend square, that will affect this and the other idioms. Again, since there are a lot more white people in this domain, this idiom has a logarithmic scale, but since every axis is divided in the same values, which are 1%, 10% and 100%, we only write the scale in one axis, to improve visibility.

Controls

We position controls (Figure 7) in the middle of our layout, since interacting with this affects all the idioms.

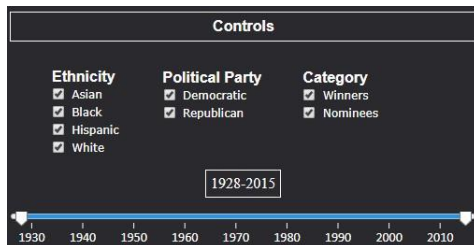


Figure 8. Controls

The Ethnicity and the Political Party can be selected here or in the respective idioms. But the Category and the time range can only be selected in this section. Interacting with this section will affect all the idioms. When a checkbox can't be selected, it gets disabled.

Rationale

To make our visualization, we tried to use simple and intuitive idioms that would not pose a threat to our user's understanding. Almost as important as the idioms are the mechanisms we provide to interact with the idioms. Using common utilities like a slider bar and checkboxes we believe that these are a powerful tool, especially for non-expert computer users. For the more expert users, we give the option to interact with the idiom itself. Providing a simple and interactive way of interacting with simple

graphics creates a very powerful way of visualizing information.

With that in mind, we drew our first sketch (Figure 9):

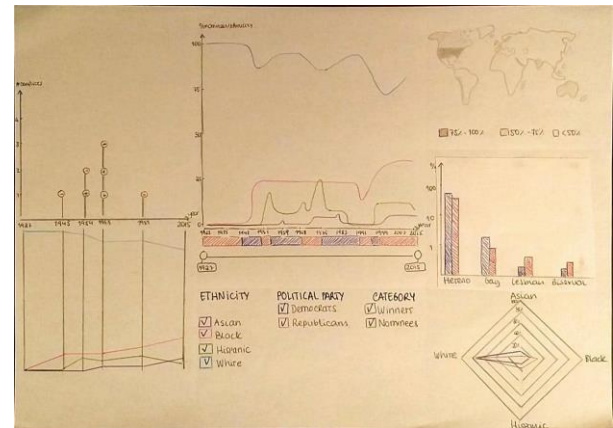


Figure 9. First Sketch

From this to our first prototype (Figure 10), we changed the Choropleth Map to a Tree Map, which in our opinion works better for this purpose, since it reduces the amount of unused space that we would have with the Choropleth Map.

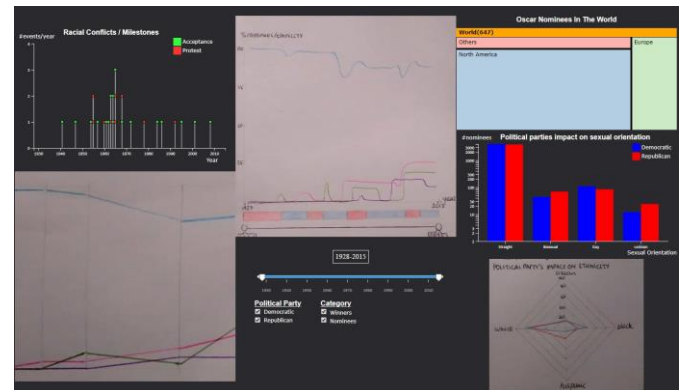


Figure 10. First prototype

From this to our final delivery (Figure 1), we changed several things:

- We removed the Slopegraph, since it would represent the same information that we get now with the multi-line chart and the line that appears on mouseover over a lollipop event.
- We changed the Layout, to position the idioms according to what they allow to view. One of the changes was having the Radar Chart below the Grouped Bars, since they both allow to visualize information according to the party in power. Another change to the layout was to place the Lollipop below the Multi-line Chart, since they both scale the X axis according to the selected years and the purpose of the lollipop is to relate the events with the percentage of nominees per ethnicity, so it makes sense to have them together.

- We changed the color scheme. We decided to use a pale color scheme, since the colors were too loud and might hurt the wearer's eyes.
- Regarding the colors of the treemap, we decided to use the same color for each rectangle, since they don't encode anything.

DEMONSTRATE THE POTENTIAL

In this section, we are going to describe the potential of our InfoVis, for which we are going to answer the following questions:

- **Question 3:** How does the ethnicity of the Oscars winners and nominees change over time?
- **Question 4:** How does the political party in power influence the birthplace of Oscars winners?

In the end of this section, we are going to show a couple of gold insights that we discovered will using our won InfoVis.

How does the ethnicity of the Oscars winners and nominees change over time?

To answer this question, we can just look to the Multi-Line Chart (Figure 2) and get an overall information. But if you want to zoom in, in a certain time range, for example to see the time when Barack Obama was in power (2008-2015), and only see the **black** ethnicity winners. It is also possible, just change the year range in the slider bar, unselect the nominee's checkbox, and select the black ethnicity. The result is present on the Figure 11.

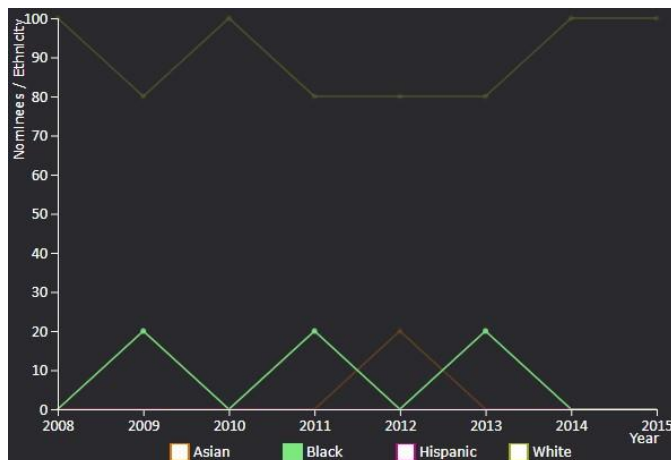


Figure 11. Multi-Line Chart, from 2008-1025, with Winners with black ethnicity selected.

How does the political party in power influence the birthplace of Oscars winners?

This question is answered in the TreeMap. Imagine that we want to compare the birthplace of USA Oscars nominees when there was a Republican party in Power, and when there was Democratic party in Power. For this we need to interact with the TreeMap until we have reached the USA States (North America -> USA). After this we can select the

desired party. The result of this is present the Figures below (Figure 12 / 13):



Figure 12. USA States with Republican party selected

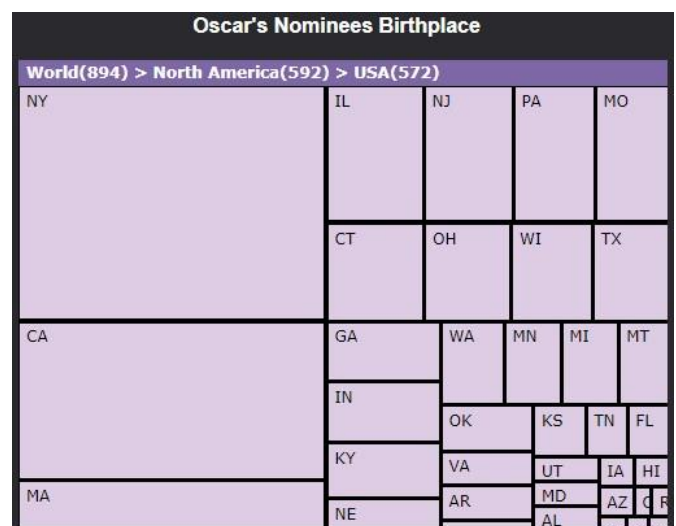


Figure 13. USA States with Democratic party selected

We notice that New York and California are the states with more nominees in both cases. Another thing that is possible to see, is that the when the Democrats were in power, there are more nominees from Massachusetts compared to when there is a Republican party in power. The same thing, in reversal, happens with Texas, when there is a Republican party in power there are more nominees from Texas.

Gold Insight

When using our visualization, we notice that in the year of 2010, when Barack Obama was the president, there was only White and Straight people nominee. As you can see in the Figure 14.

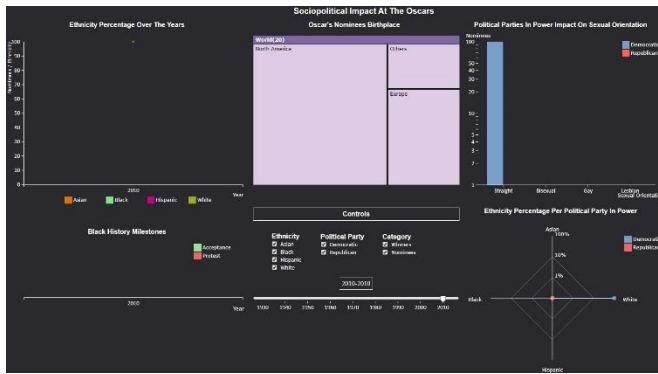


Figure 14. Our InfoVis in 2010

Another interesting fact it was presented in Figure 6. It's possible to see that Shirley Chisholm, the first black woman elected at the United States Congress, runs for the US presidency. In the same year, 1972, there was clear change in the black nominee percentage, changing from 0% to 12%.

From images 12 and 13 we can see there are some changes in the nominees selected by party in power, especially in the Texas and Massachusetts states. Historically, Texas is a Republican party and Massachusetts is Democratic. This is at least an interesting coincidence, leaving one interesting question up in the air: "Is the birthplace of Actors taken in count, when choosing the Oscars Nominees?".

Besides the five questions that we intended to answer with our visualization, there are many others which can be answered, such as:

- "How many black gay men have won an Oscar?"
 - Selecting black ethnicity and winner category, and looking at the grouped bars chart, the answer is 0.
- "Were more Asians nominated / won when the Democrats or the Republicans were in power?"
 - Selecting Asian ethnicity and looking at the radar chart, the answer is Republicans.
- "What is the birthplace of most Hispanic winners/nominees while the Democrats were in power?"
 - Selecting Hispanic ethnicity and Democratic political party, navigating through the TreeMap we find that the answer is Puerto Rico, with 4 winners / nominees.

IMPLEMENTATION DETAILS

To facilitate the work and the communication between the team, we used a Messenger group to communicate, and a GitHub repository to share the work done.

For web scraping we used a script available on a GitHub repository [14], that allows to web scrap Wikipedia pages,

in entirety, even when these have multiple tables, with multi rows per column, or multi columns per row. As explained in the data section, we used the PDI to take care of our csv files.

To facilitate the developing process, we had one idiom per file, and a main.js file that has the common variables such as the time of the animation. Due to the new Js version [12], it is possible to import and export variables of different files, and make general changes, as quick as changing the value of a single variable. Also, to have a consistent project we defined different CSS classes that make easy to style and change the elements. A good example of this are the tooltips. We couldn't use the title tag, since this can't be stylized. So, we created our own way to generate tooltips, using CSS classes, so that all the tooltips have the same style, and imported and exported variables to help us in the creation process.

To implement the slider range bar, we used a file available on GitHub [11], that also uses the version 5 of d3. This file allows to style the bar as we want, and creating a functional simple slider, just using a few lines of code. So, we didn't have to worry about creating a slider from scratch, using the saved time to improve the interaction between the slider and the idioms.

To implement each idiom, we took inspiration from the examples found online and the ones given in the labs. However, it was not a copy and paste, since it was necessary to adapt those examples to: implement the animations; link the different idioms; in some cases change the axis type and scale the axis. To implement the treemap, it was also necessary to process the data to be in a json array. Since fast response is an important thing on this type of projects, we measured the time of the most durable functions using js techniques [13], after which we could identify the problems and find ways to solve them. To implement the link between the views, we used the d3 dispatch, that allows to send a single dispatch to multiple files, using the same event with different namespaces (we used the name of each idiom as namespace).

In the end of our project, we only had encountered one bug, that occurs in Mult-Line chart when, the user changes the time range very quick and multiple times. This happens because we decided that our animation would take 1 second to end, which is, in our opinion, the best value because it gives a smooth visualization. So, when a user, in one second, selects different time ranges (for example, changes the time range ten times), the animation stops and starts over. After doing this for so many times in a row, there is a bug, where some circles are not hidden. D3.js doesn't offer any way to solve this problem, the only solution was decreasing the time duration of our animation. We didn't do that, because it made the transitions less smooth and that goes against the purpose of our visualization, since we don't want to harm the look & feel for the users.

CONCLUSION & FUTURE WORK

With this project we were able to learn the theoretical and practical concepts behind an InfoVis. We also deepen our knowledge and expertise in HTML, CSS and JS, furthermore we learn to work in a new framework: d3.js. The idioms that we implemented can answer all the questions we wanted it to answer at the beginning, and allow to see interesting information about the Oscars, as explained in the section above.

If we were to start this project from the scratch, undoubtedly more importance would have been given to the data selection/cleaning/processing step since on the final stretch some time was spent adding new information to the datasets. If we had more time, we would like to implement a bar below the Multi-Line chart, that that would show us how long each party has been in power, as you can see in our first sketch (Figure 9). We also would benefit if we had financial funds allocated to the project. Hiring people that could find the data necessary to show how many actors exist by sexuality and ethnicity, or to extract necessary to calculate the winner/nominee ratio would give us more time to focus on the Viz. We also would like to acquire a designer, that could help us improving our layout and our scheme color, with a more experienced opinion.

ACKNOWLEDGMENTS

We thank our teachers for taking the time to teach us and giving constructive feedback that has undoubtedly improved this project. We also thank our colleagues who provided constructive criticism so that we could improve our layout.

REFERENCES

1. <https://fenix.tecnico.ulisboa.pt/disciplinas/VII13/2019-2020/1-semester/hall-of-fame>, last accessed in 12/12/2019
2. <https://github.com/d3/d3/wiki/Gallery>, last accessed in 12/12/2019
3. <https://bl.ocks.org/>, last accessed in 12/12/2019
4. <https://www.d3-graph-gallery.com/>, last accessed in 12/12/2019
5. <https://www.kaggle.com/theacademy/academy-awards#database.csv>, last accessed in 12/12/2019
6. https://www.kaggle.com/harshitagpt/us-presidents/#us_presidents.csv, last accessed in 12/12/2019
7. <https://data.world/crowdflower/academy-awardsdemographics/workspace/file?filename=Oscars-demographics-DfE.csv>, last accessed in 12/12/2019
8. https://www.history.com/topics/black-history/black-historymilestones?fbclid=IwAR3YODpITj_NPrfw_BJIE2lkKTEpjgMnfM8Wt2jF5NRpYLA6emX1vy2NGM, last accessed in 12/12/2019
9. https://en.wikipedia.org/wiki/List_of_LGBT_Academy_Award_winners_and_nominees?fbclid=IwAR1GI6NSf9BAuZsYJfrEBhA7oE5bmKknwllhlQ9-j1O9Dc15c541PmBjck8, last accessed in 12/12/2019
10. http://labs.time.com/story/oscars-diversity/?fbclid=IwAR32LRwO2XVa2eWFdXtzZc3m3tzp75883ilKEoM_8IA4IubXqLnhFM19GVk, last accessed in 12/12/2019
11. <https://github.com/johnwalley/d3-simple-slider>, last accessed in 12/12/2019
12. <http://es6-features.org/#Constants>, last accessed in 12/12/2019
13. <https://developer.mozilla.org/en-US/docs/Web/API/Performance/now>, last accessed in 12/12/2019
14. <https://github.com/rocheio/wiki-table-scrape>, last accessed in 13/12/2019
15. Box office or best picture? Can you really have both? <https://nyu-vis-fall2018.github.io/storytelling-group-5/>, last accessed in 14/12/2019
16. Impact of the Oscars nominations <http://krisztinaszucs.com/viz/oscars/>, last accessed in 14/12/2019
17. Adding Up the White Oscar Winners <https://www.bloomberg.com/graphics/2016-oscar-winners/>, last accessed in 14/12/2019
18. How diverse is the Oscars? <https://www.pinterest.pt/pin/482237072584681311/visual-search/?cropSource=6&h=671&w=530&x=22&y=10>, last accessed in 14/12/2019