Visualization assignment

Pedro Silva 93011 João Soares 93078
Information Visualization, 2022
Master in Informatics Engineering, University of Aveiro

***Abstract* –This report presents the development of a prototype for the application Cinema-Goers. This report begins by indicating the objectives and motivation behind the application also explaining the target users and the questions that the application should answer.**
Then presents the dataset used, and finally addresses the visualization solutions developed, by first explaining the Low Fidelity Prototype and the user feedback received and after that explains the Prototype including the implementation challenges and the evaluation method used and its results.

I. Motivation and objectives

In an era in which there has been a huge increase in movies released yearly while at the same time these are more interconnected (cinematic universes) and easier to access, we found that an application focused on finding easily the insights in cinema will have a wide user base of curious fans about the intricacies of the medium.

II. Users and the Questions

With the huge increase in users due to easier access to movies and the increase of *cinematicuniverses*, (a set of creative works where more than one writer independently contributes a work that can stand alone but fits into the joint development of the story line, characters, or world of the overall project [1]), in the industry more people get attracted and curious about certain movies. Taking this into account we focused our work on some interesting questions we think this demographic would find interesting.

A. Characterization of the users and their context

The users in mind for the application could be normal people curious about movies that occasionally want to know details about a certain movie.
However we expect that the main users of the application will consist of *Cinephiles*[1]. A persona representing the user description would be Luis.
Luis is a 17 year old student that lives in Aveiro, loves watching movies and has the ambition of one day direct a blockbuster movie. He tries to follow and study, the different aspects of the movie industry, like analyzing the genre of movies, that successful directors like Martin Scorsese, direct to better manage is future career, but this task gets complicated due to the lack of ways to visualize the different information's related to the movie industry, so he would enjoy an application that offers that solution.

[1]a person who is very interested in and enthusiastic about cinema as an art form, and knows a lot about films [2]

B. Questions to Answer

The following questions were elaborated by combining our own suggestions and the suggestions of the users interested in the movie industry:

1. Most popular movies?
2. Where are movies produced?
3. What is the most spoken language in movies?
4. Is there any relationship between budget and popularity?
5. In which movies Production Companies worked together to make it?
6. Production Companies that produced the most movies on a certain year?
7. Relation of film language and revenue obtained?
8. Director speciality regarding the gender of the movies?
9. What are the great centers of cinema and how have they changed over time?

III. Dataset

The dataset used in the prototype is The Movies Dataset, that contains information on 45,000 movies featured in the Full MovieLens dataset. It was decided to only considered the movies with a budget higher than 0, ($movie.budget > 0$), which results in a final dataset with 8903 movies.

The dataset contains numerous movie-related metadata, among which we used:

- Budget
- Genres
- Original Language
- Spoken Languages
- Title
- Overview
- Popularity
- Production Companies
- Production Countries
- Release Date
- Revenue
- Director (throw the Cast data by pre-processing)

IV. VISUALIZATION SOLUTION

A. Low fidelity prototype and user feedback

The Low fidelity prototype (Fig. 1, Fig. 2, Fig. 3 Fig. 4, Fig. 5, Fig. 6 and Fig. 7) developed with balsamiq presented a way to answer the questions mentioned previously.

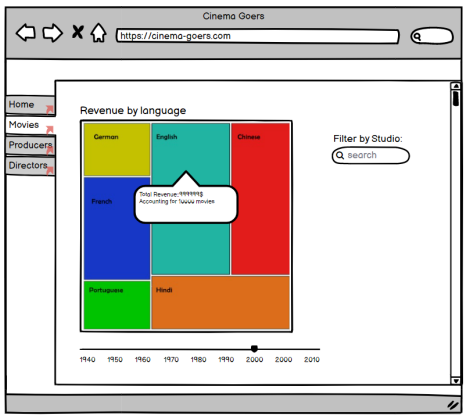

Fig. 1: Low fidelity Treemap to answer question Q7

After users tested the low fidelity prototype their feedback was:

- Compress more information and visualization options in the different views to reduce the number of them. For example the visualization in Figure 2 was remodeled to be a option in the Stacked Area Chart (Fig. 10).
- Instead of having a simple Treemap (Fig. 1) , we could implement something of a Hierarchical Treemap , thus benefiting from a greater clarity in the data visualization.

Fig. 2: Low fidelity Stacked Bar Plot to answer question Q6

Fig. 3: Low fidelity Stacked Area Chart to answer question Q3

Do to time limitation we also reduced the number of visualizations in the prototype.

Fig. 4: Low fidelity Scatter Plot to answer question Q1 and Q4

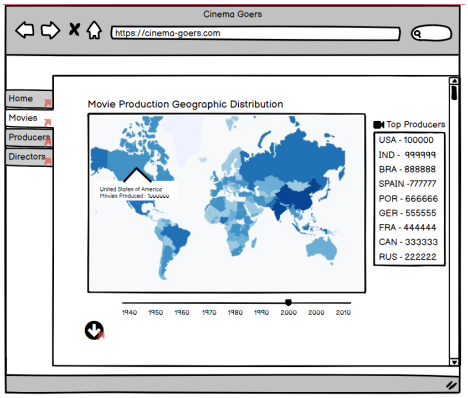

Fig. 5: Low fidelity Choropleth Map to answer question Q9

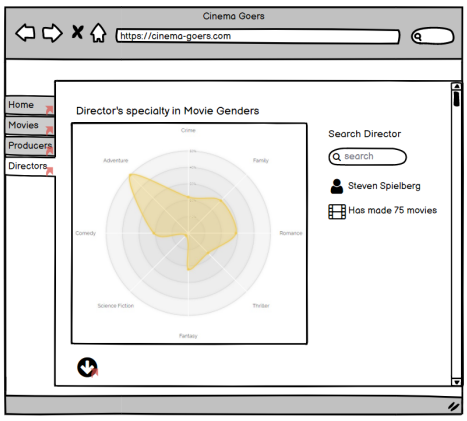

Fig. 6: Low fidelity Spider Chart to answer question Q8

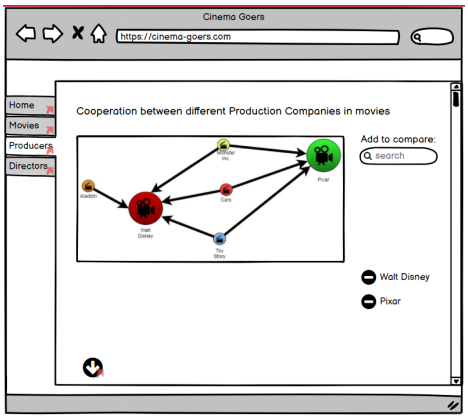

Fig. 7: Low fidelity Network Graph to answer question Q2 and Q5

B. *Functional Prototype*

The following Functional Protoype has already some modifications due to the results of the Usability Test that we will address later in the report.

The Scatterplot (Fig.8) displays the relationship between the Budget of a Movie and the Popularity Score of a Movie, addressing questions Q1 and Q4. For each data point (Movie), the value of its Popularity Score is represented on the X axis, the Budget on the Y axis.

It is possible to filter by Production Company so that only Movie points from the searched Production Company appear. To reset, search by nothing (empthy search bar).

The implementation was based of D3.js Gallery from Yan Holtz [3]. The Hierarchical Tree Map (Fig. 9) displays the relationship between the original Language of a Movie and it's revenue (questions Q3 and Q7), using squared representations of the percentage of revenue each language has for all movies in the dataset.

By clicking on a determined language the Tree Map descends a level below and shows the a Tree Map of movie gender by revenue for movies of said determined language. By hovering over a square in both levels it's possible to see how much exact revenue it represents.

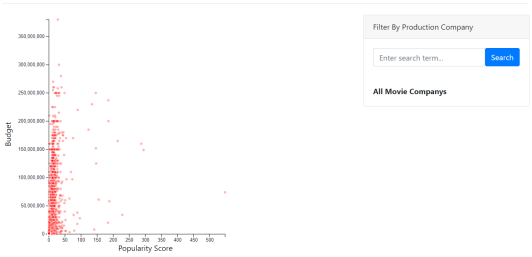

Fig. 8: Scatter Plot Visualization to answer question Q1 and Q4

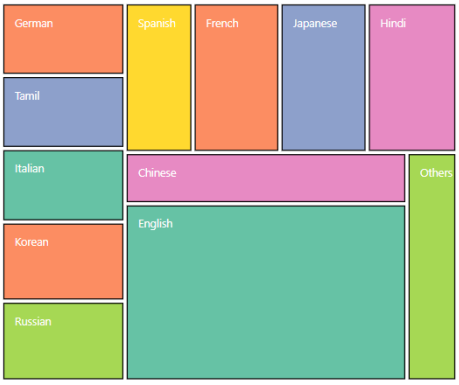

Fig. 9: Hierarchical Tree Map Visualization to answer question Q3 and Q7

The Stacked Area Chart (Fig. 10 displays how the number of Movies produced worldwide evolved by using either its original language or its Production Companies (question Q3 and Q6).

It's possible to visualise this evolution by either entering the languages to compare or the name of the production companies.

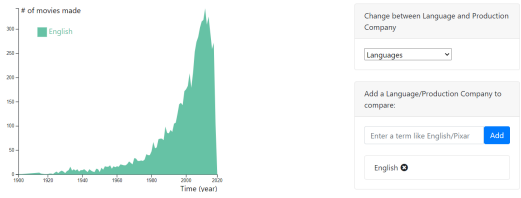

Fig. 10: Stacked Area Chart Visualization to answer question Q3 and Q6

The Spider Chart (Fig. 11) is a two-dimensional chart type designed to plot one or more series of values over multiple quantitative variables. In this case it will plot the percentage of Movies of a certain genre from a Director, being possible to check which genres of Movies the Director directs the most, addressing question Q8.It also allows a direct comparison between directors.

The implementation was adapted from Nadieh Bre-

mer "Radar Chart Redesign" [4].

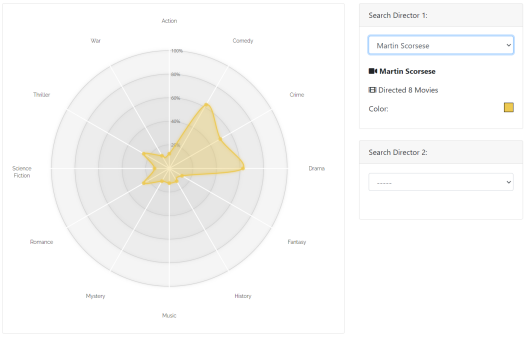

Fig. 11: Spider Chart Visualization to answer question Q8

Choropleth Map (Fig. 12) displays divided geographical areas that are colored in relation the number of Movies Produced in the Country. It allows to study the evolution of the great centers of cinema and how have they changed over the last decades (question Q9).

The implementation was based from of D3.js Gallery from Yan Holtz [5]. The file World.geojson was also adapted and rectified within the needs of the prototype.

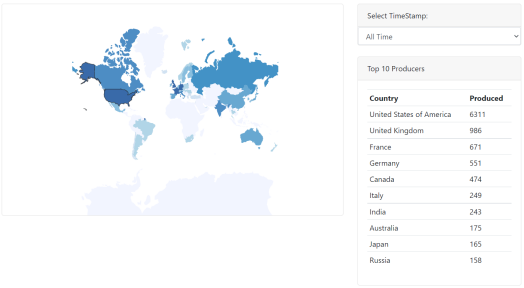

Fig. 12: Choropleth Map Visualization to answer question Q9

Finally, the Network Graph (Fig. 13) between movie studios and the movies they contributed in making, each movie is connected to all studios who worked on it, addressing questions Q2 and Q5 .

It allows to filter which studios the user wants to see movies of and pan around the graph, can also hover a movie or a studio to see them in detail.

C. Implementation challenges

aki coisas sobre dificuldades (falta de documentação. problemas de dados? cenas de js etc)

A implementation challenge found was the many differences between the versions of D3. Although every visualization (except Spider Chart in version 3 that was having color issues in V6) are using version 6 of D3, many examples and documentation are better explained in previous versions and when trying to adapt or even learn, we found that many functions had different syntax or even different functionalities. This made the implementation process slow in the start but after

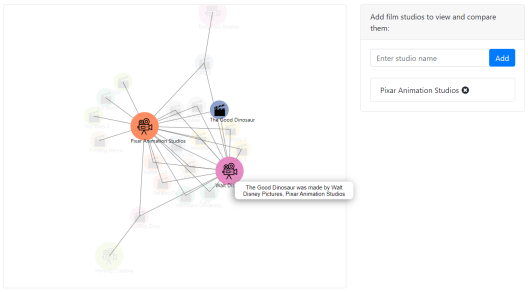

Fig. 13: Network Graph Visualization to answer question Q2 and Q5

that we were better adapted to the differences between the versions of D3.

In the Scatter Plot graph we had to remove completely the "points" when updating the search because the optional solution of removing the "points" that didn't fit the search, keeping the ones that fit and add new points, was not working properly.

Because directors have done between them Movies with different types of genres sometimes the visualization in the SpiderChart became very confusing, especially when comparing two very different type of directors. To fix and mitigate this problem, we decided to only use the combination of genres between the directors instead of all genres like we were doing and order them alphabetically in the chart.

In the Network Graph chart, we weren't able to provide the ability to drag and drop nodes from one position to another due to conflicts between the drag function and the force used to maintain the nodes centered on screen, information about this topic seems to vary a lot between different versions of D3.

D. Evaluation and changes in the protoype

To evaluate the prototype, we elaborated a Usability Test with several parts:

- 10 sets of questions to be answered using the prototype.
- A set of questions to collect data about the profile of the user
- A set of questions about the experience with the prototype
- In the end, a standard usability questionnaire (SUS- System Usability Scale) to collect the user opinion

The users that took the Usability Test were all very interested in Movies and Cinema in general, all had some degree of computer literacy with the exception of one user that consequently, provided us with important feedback about the prototype that otherwise would have probably been ignored.

Overall, we gather good results with all the users giving correct answers to questions to be answered using the prototype and most of the users found the question easy/very easy to answer. In the end users expressed

some problems and some negative aspects about the prototype some of which were already addressed in the prototype, due to being very common among the users:

- Add an indication of "what color is the director" when comparing the genres of a Director in the SpiderChart Graph.
- In the Tree Map we scaled the "rectangles" to improve the visualization, improved the "Go Back" button and added a tooltip to make it clear to the user, that to filter by language the user had to click in the "rectangle" of the language.
- Add an indication of what color corresponded to what language/studio and make the scale end in 2020 in the Stack Area Chart.
- Add hover effects to improve, (and help to distinct), the visualization of the movies made by a production company and tooltips throw out the Network Graph to make it easier to know how many movies a production company produced.

The remaining feedback was left to be address in Future Work as it was only mentioned by some users.

V. Conclusion and Future Work

We were able to develop an application that provides for curious people and cinephiles a variety of insightful visualizations that cover a large range of attributes about the Movie Industry of the dataset we used.

Taking into the account the input provided by the Usability Test results we were able to fine tune some of the details in our work and add new functionalities that improved upon them. For future work we could continue tuning our visualizations, with remaining feedback provided by the users: Possibility to select different range of years in the Choropleth Map, not just decades. Filter by more than one Production Company in the Scatterplot Visualization. In the SpiderChart add the option of suggestive Search rather than the select options. To avoid performance issues in the Network Graph when searching for large movie companies will add the option to only render a certain amount of items indicated by the user.

References

[1] Wikipedia, "Shared universe", 2021, [Online; accessed 10-January-2022].
URL: `https://en.wikipedia.org/wiki/Shared_universe`

[2] Cambridge Dictionary, "Cinephile", 2022, [Online; accessed 4-January-2022].
URL: `https://dictionary.cambridge.org/pt/dicionario/ingles/cinephile`

[3] Yan Holtz, "Scatterplot with tooltip in d3.js", 2022, [Online; accessed 2-January-2022].
URL: `https://www.d3-graph-gallery.com/graph/scatter_tooltip.html`

[4] Nadieh Bremer, "Radar chart redesign", 2022, [Online; accessed 2-January-2022].
URL: `http://bl.ocks.org/nbremer/21746a9668ffdf6d8242`

[5] Yan Holtz, "Most basic choropleth map in d3.js", 2022, [Online; accessed 3-January-2022].
URL: `https://www.d3-graph-gallery.com/graph/choropleth_basic.html`