

BY: DATA WIZARDS



WHAT'S THE BEST STREAMING PROVIDER FOR YOU?

The goal of this project is to create an interactive website where the user can easily compare the different streaming providers with each other. Our ambition is that the website, through data visualizations, can help the user decide on which streaming provider to choose based on some of their own preferences for movies and series by parameters such as genre, actors or directors.

EPFL

THE PATH TO THE FINAL RESULT

The beginning of our journey started with finding out what we wanted to visualize. We were not sure about the particular thing that we wanted to visualize, but we knew and agreed that we wanted to create a website that we ourselves could actually take into use and that could come in handy. We also needed to make sure that we had enough data to make meaningful visualizations. After some research of potential data sets, we came across a particularly interesting data set on Kaggle. It was a data set containing information about different movies and TV shows for the different streaming providers; Netflix, Amazon Prime, Disney, and Hulu.

After we came up with the data set, we needed to figure out how to visualize the data in a meaningful way. What kind of visualizations did we want to display. We will talk about this in detail in the next section.

After the completion of the previous two milestones, it was time to get started on the visualizations themselves. It took some time to get the skeleton of the visualization to work, but we managed to do so eventually. After we were finished with the core of the visualizations, we started to add multiple features to them to make them more interactive and interesting. You can read more about this in the next part. Once the website and visualizations were all done, we finalized the process book and recorded the screencast.

THE CHALLENGES AND DESIGN DECISIONS

CHALLENGES

The following section will provide you with an insight into our challenges along the way as well as explain our design decision, which will be further talked about in the “The Sketches” part.

The first challenge we faced was the one mentioned in “The path to the final result”, finding a suitable dataset. After we decided on the dataset most of the rest of the problems were design decisions as well as coding-related.

Using the new version of D3, that is version 6 and 7, while most of the tutorials online are based on previous versions. The commandoes are different and therefore

We quite quickly came up with a bar chart. This would be relatively easy to implement while at the same time providing the reader with meaningful information about the data.

The Sankey diagram produced some trouble since there isn't a good documentation available online. The most special thing about this plugin is that it uses pass by reference instead of pass by value. This seemed to be rather unusual since most functions in JS are pass by value (or at least seem like it). Also, the multiselect tool made some trouble since the documentation doesn't specify the version of bootstrap and jQuery it requires. We found that many combinations are just not working. It was while working on the Sankey diagram when we realized that the Hulu database didn't have data about directors and actors in it. Therefore, we leveraged the IMDb and joined it to with the Kaggle csv which took some effort to get right.

THE CHALLENGES AND DESIGN DECISIONS

DESIGN DECISIONS

Bar chart

For the bar chart the idea was to just make a static non interactive simple bar chart. However, we decided that it would be a good idea to create a stacked bar chart instead as this would allow the user to easier distinguish the amount of content for each streaming provider. We also found out that it would be nice if the user could be able to see the amount of content produced per year, hence we decided to add a slider giving the user this possibility.

Sankey

Originally, our plan was to give the users of the Sankey diagram the full ability to choose from any actor in the database. We soon realized that there are just too many actors in the database and the generation of the select got slow. This impacted the user experience and hence we decided to set a minimum of 15 movies per actor. This threshold can be adapted by the user using a slider.

Similar, we originally wanted to use the whole, original dataset as an input. The drawback was that this method used a lot of resources when the website was loading. This not only impacts the user experience, but also made Firefox throw some warnings. While this loading allowed us to develop new features quickly, it impacted the user too much.

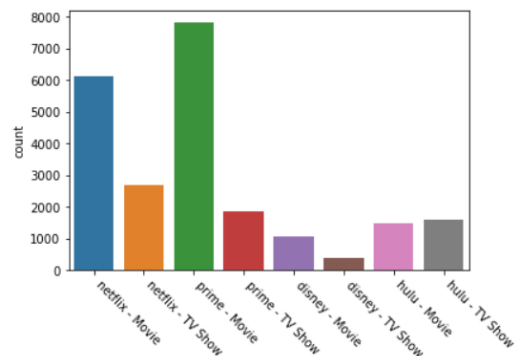
Line chart

For the line chart the idea was to display how the different providers add content to their platforms. We later decided to add some filtering possibilities so the user can filter on genre, etc., for more personal choices. The user can then see which provider that adds most content in for instance the “fantasy” category.

THE SKETCHES

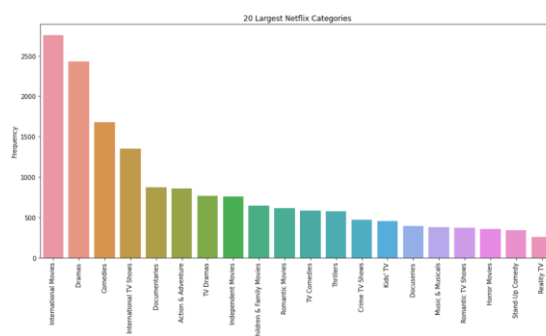
BAR CHART

The original idea was to have a normal bar chart where the bars were separated as shown in the figure below. However, after some thoughts and discussion, we decided that it would be better to have a stacked bar chart in terms of what we wanted to show with the visualization. A stacked bar chart would mean that we could sort movies and TV shows by provider, and then on top of that by type. Overall a less complex and thereby easier chart to interpret, in our opinion. We wanted to use D3 as much as possible to create this plot since it was one of the core topics of this class. We further used Bootstrap elements to give the website a modern looking and somewhat familiar touch.



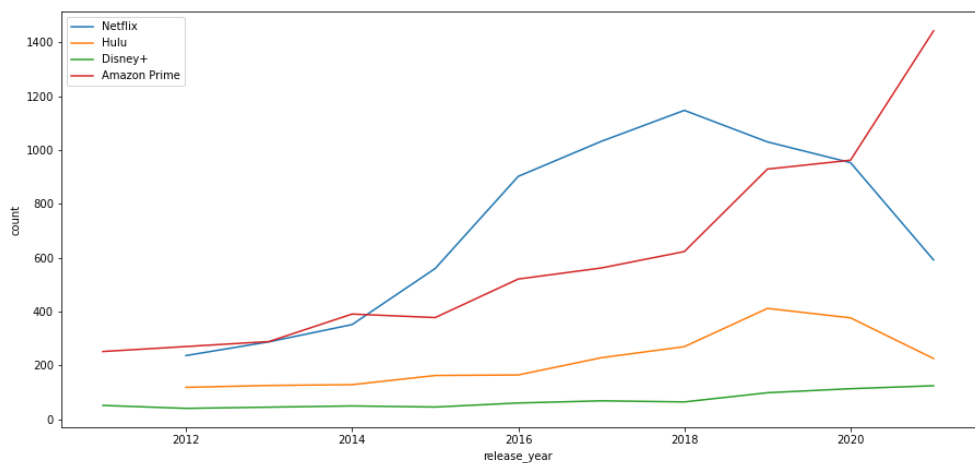
HISTOGRAM

The original idea was to do a histogram that showed the distribution of categories across the different streaming providers, so that the user could see what kind of categories the different providers provided. However, we instead decided to add this feature into the Sanky-Diagram, as this would make more sense to us. Then the user could select multiple categories at once, and then look at the distribution for all the streaming providers.



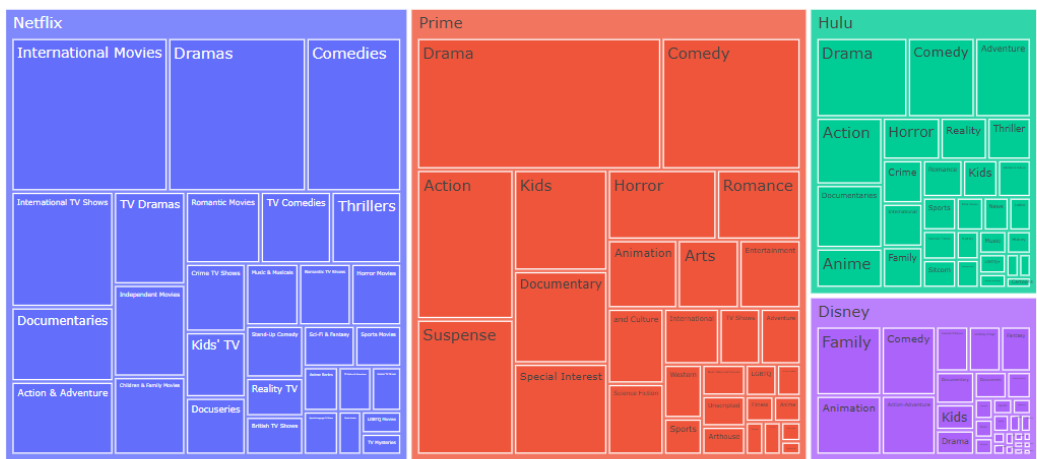
LINE CHART

The line chart is probably the visualization that has changes the smallest amount from what we originally thought of. However, we added some functionality to the visualization such as filtering and a slider so one could play with the visualization a little more.



TREE MAP

We decided not to implement the tree map and instead adding features to the existing visualizations, as we though they provided as good or even better information than what a tree map would have done.



PEER ASSESSMENT

Joël Lingg

Joël made the Sankey Diagram and contributed to the other visualizations as well by adding more features.

Ferdinand Ytteborg

Ferdinand was the main man behind the line chart and did a lot of the preprocessing.

Fredrik Hægermark

Fredrik did the website design and process book as well as the stacked bar chart.