

Data Visualization. Milestone 3.

Roxane Burri, Jean-Luc Stoupy, Mark Tropin

May 2022

1 The project

1.1 Final product

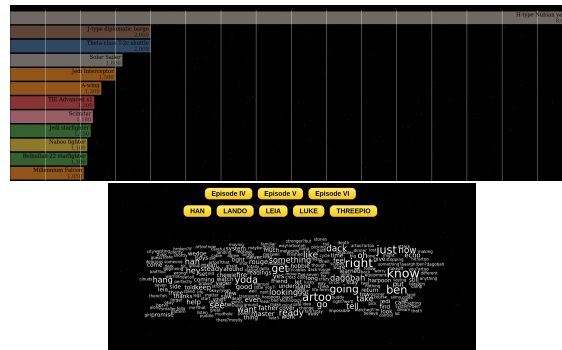
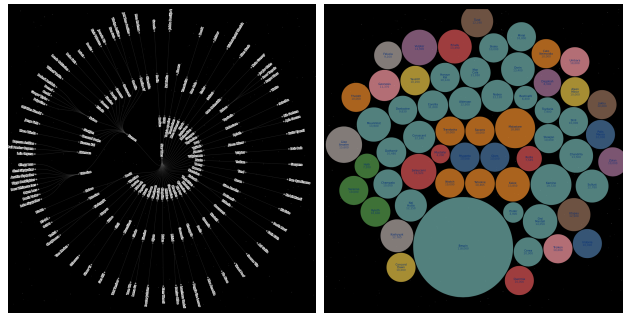
In a current context where most of the information is sad, we decided to choose a joyful universe allowing to have fun and so we chose the cinematographic universe of Star Wars.

Through different datasets, we want to show some facts about the Star Wars universe. These visualizations are simple to capture the interest of any kind of knowledge about Star Wars. This universe is vast and complex, so the goal of all the work is to bring fun facts that can bring new people into the universe or that can entertain people who already know Star Wars. We are primarily targeting Star Wars fans that would like to learn more about the Star Wars universe as well as newcomers that want to get to know the Star Wars saga, in a visual and entertaining manner. With the objective of immersing the Internet user in the star wars universe, we decided for the final project to use a front end style like in the Star Wars episodes.



Figure 1: Left : the title in a Star Wars style. Right : Source code to make the text like in the episodes

Here are images to represent the different graphics on the website.



2 Challenges

2.1 Visual style

One of the main design decisions that we had to take was the visual aspect of the web-site in general. Our first prototype was white and used simple colors and fonts. While this style was intuitive and understandable for the end user (which was one of our main requirements), it wasn't fun and exciting: looking at this interface, the user had no idea they were about to explore the Star Wars universe. Therefore, we decided to focus on improving this aspect of our web-site and in the end we were very satisfied with the result.

2.2 Radial tidy tree

With this graphic we wanted to represent the distribution of the different species, planet and characters. We started with a representation with the planets in the center, then in second level the species on each planets, and the character of the species on the planet on third level. But this visualization did not bring much because it did not look like a tree. We then decided to exchange the first level of planets with the second level of species. This allowed us to obtain a visualization with a shape that looks more like a tree.

2.3 Bubble chart

This visualization was slightly different from the one that we proposed in the previous milestone. At first, we tried to use a smooth zooming visualization that would zoom in on a particular planet and show its characteristics. We have encountered numerous technical issues with this visualization and were unable to change the sizes of the circles which represented the planets; we thus decided to replace this visualization with a bubble chart.

We had to adapt the visualization to represent all the data that we wanted to show, and we succeeded: our bubble chart represents the climate of the planet with colors, and the sizes of the bubbles represent the diameter of the planets. The user can also see the name and the size of each planet depicted in the chart.

2.4 Bar chart race

The goal of this visualization is to represent the different maximum speeds of the starships. The bar chart race allows to create a small race to show these speeds. The biggest problem here is that some bars were not aligned on the left but started in the middle of the graph. After research in the code we were able to solve this problem.

2.5 The "word cloud"

One of the challenges that we have faced was deciding how to visualize the dataset with the character lines. It was clear to us that we wanted to use a

”word cloud” visualization, however we were unsure about how to tackle the various data sources. To recap, the character lines in our dataset are first grouped by episode (the classic trilogy consisting of Episodes IV, V and VI) and then by character.

We decided to provide the end user with the choice of episode and then offer them the N most talkative characters in each respective episode (these were not always the same characters: for instance, Luke appears in all three movies while Lando can only be seen in Episode V). We decided to set $N = 5$ for the sake of visual appearance (the user probably does not want to see 20+ buttons with *all* the characters in a given episode) and because of the amount of data we had (it makes no sense to plot a word cloud for the least talkative characters simply because there isn’t enough data for them).

The implementation of the ”word cloud” visualization also presented several challenges. We were unsure how to display the buttons and where to put the visualization so that the interface appears intuitive to the user. Our first idea was to ask the user for the number of the episode, then replace the first set of buttons by a different set, now asking the user for the choice of character. Finally, when the episode and the character is chosen, the buttons would be replaced by the visualization itself. This approach presented a ”logistical” challenge: how does the user backtrack if they wish to change the episode or the character? Do they refresh the web page, scroll down to the visualization and re-submit their new choice? This seemed unintuitive and tedious; we want our users to see the visualization they want to see with minimal effort. Therefore, we decided to keep the episode buttons and dynamically update the characters and the visualization itself when the user confirms their choice.

3 Comparison with the first milestone

Our final deliverable is very similar to what we proposed in the preceding milestones; we have included all of the visualizations that we anticipated to produce. Moreover, we have embellished our web-site with the famous Star Wars rolling intro and a title stylized with the Star Wars font. The background was also updated: it is currently black and covered with bright stars which can be frequently seen in the Star Wars saga.

We should also note that the box office dataset, which was mentioned in the first milestone, does not appear in our milestones 2 and 3. This dataset was not related to the Star Wars movies *per se*: it contained ratings and box office statistics for a wide range of movies, and only a few data points were relevant to the topic of Star Wars. We decided to remain focused on presenting data relevant to the Star Wars lore itself: therefore, this dataset was not visualized.

As for the visualizations proposed in the second milestone, we can say that all of them were implemented with minor modifications. For example, in the case of the ”word cloud” visualization, we have decided not to reuse the space occupied by the buttons for the actual visualization, but to display all of the button sets (episode and character) and the visualization in sequence: we have

concluded that this layout will be more intuitive and easier to use.

To compare our final web-site with the "Minimal Viable Product" discussed in the previous milestones, we can say that we have implemented all of the proposed visualizations (the smooth zooming visualization was replaced by the bubble chart for technical reasons). As for the proposed enhancements of our visualizations, we decided not to add any new features (we don't want to confuse the user for the sake of showing more data), and we concentrated on the visual aspect of our visualizations and the web-site in general. Moreover, we decided to concentrate on this part to create a more authentic Star Wars experience for our users.

4 Peer assessment

4.1 Roxane

I worked on :

- Radial tidy tree : I took over the work of Jean-Luc, and I changed the data to change the planet - species - character into species - planet - character.
- Bubble chart : I imported the bubble chart code from d3 (we did not put it during milestone 2), and I preprocessed the data to represent the planets depending on their climate and their diameter.
- Bar chart race : I preprocessed the data of the ships, and I added random values to simulate a race.
- I also wrote all the texts except those of the words coulds on the site (text inspiration for intro).

4.2 Jean-Luc

- Radial tidy tree : I made the graph with our first idea of making a tree organized as : planet - species - character. After this try, we were not convinced and Roxane inverse it to have something more visual.
- All Charts : I managed to import all templates on an easy way during milestone 2 to work on it in a way that the code stay cleaned and well managed.
- Bar chart race / Bubble chart : I resolve last problem on it such as alignment or broken links.
- I made all the general design with the StarWars theme (font and starry sky) and create the general shape of the webpage.

4.3 Mark

I have worked on the "word cloud" visualization.

First of all, I was tasked with preprocessing the dataset in a way that would be suitable for visualizing with the tools provided by D3.js. This was done by the means of a Jupyter notebook which loaded the CSV file with the character lines, grouped all dialogue lines by episode and character, combined them into a coherent string of text and produced a text file for each (episode, character) pair individually.

I have then worked on the button functionality and the visualization itself. I have faced difficult design decisions that I was able to resolve by communicating with my teammates and researching the Web to find what approach works best (this is described in the "Challenges" section). With the help of Roxane and Jean-Luc, we were also able to figure out a way of updating the visualizations which was time-efficient and at the same time made sense to the end user.

After the buttons were implemented, I focused on the "word cloud" itself. I made sure that the buttons provided the correct filenames to the word cloud, and that the visualization would be updated appropriately.

Finally, I have worked on the cosmetic aspect of the word cloud: I styled the buttons and the font of the cloud that would correspond to our topic of the Star Wars universe. To wrap up, I added a small description of my visualization to guide the user.