Process Book

Anime Data Visualization

Alexandre CHAU, Pedro TORRES DA CUNHA, Joachim DUNANT
May 28, 2020

1 Introduction

Animes are a way to entertain each and everyone, by showing different stories in beautiful or dreadful universes. With them, one can escape in a different World for a period of time, learn about life by the morality shown in them, or even get to see diverse cultures. They have grown exponentially in popularity over the years, to the point that today almost everyone has seen at least in their lifespan.

Thinking about how popular it is nowadays, being able to see different visualizations about them is more important now than ever. LOOK AT IT NOW ! LOOK AT THE VISUALIZATIONS ! DO IT ! JUST, DO IT ! From every anime depending on their airing date, to the top

2 Problem statement

The main concern is to give insights and the ability to learn more about different aspects of the world of anime.

How has anime evolved from its earliest years in 1917 until now? Did the number of produced animes increase through time? Is the average number of episodes now bigger or smaller?

How are the genres distributed across the animes? What is the genre that is the most represented and what are the most popular genre per anime?

What are the most popular studios? Which genres are the most prevalent for each of them and what are the ratings of their productions?

What are the most popular voice actors for each language? Who has already worked together for an anime? What does the network of anime dubbing look like?

3 Dataset

The dataset used to answer those question is the MyAnimeList Dataset.

It contains data about animes from MyAnimeList which is one of the biggest platforms to see and rate animes or movies. Precise and clean information about each anime can be gathered including the genres, studio, popularity and more.

The voice actors are unfortunately not included in this dataset. However, we can gather them through Jikan API, an unofficial MyAnimeList API from which almost everything from the site can be scrapped.

4   Project structure

4.1   Tools of the trade

4.2   Architecture

5   Visualizations

5.1   Histogram: 100 years of anime

5.2   Bubble chart: 43 genres to classify them all

The goal was to show the distribution of the genres across the animes. Given the non-negligeable number of genres, we thought a bubble chart would be perfect as it quickly shows the most represented genres and how they compare to each other.

From the main anime dataset, we were able to directly access the genres of each anime. Inspired by some d3 example, we were able to generate the positions and radii for the bubbles.

To illustrate the genres, we decided to display the image of the anime with the most favorites for each genre. However by doing so, if two genres have the same anime as the most favorited, then they both will have the image of that anime. For example, all the genres of the most favorited anime will share same image. To prevent that we select the image from the most represented genre to the least by taking the anime that was the most favorited and not yet selected by another genre. Then one genre might not display the most favorited anime, but the anime diversity will be way bigger.
Some of the image links

We also wanted to give the ability to filter the genres that are displayed. We do that by displaying a dropdown list of toggable buttons, each corresponding to one genre. We also provide an `ALL` button and a `NONE` button to quickly show all the bubbles again or unselect everything so that one can then selectively choose the bubbles he is interested in. Each time list of genres to display changes, the bubble positions and radii are recomputed. The bubbles then smoothly move and scale from their previous position to the new one.

Another feature we wanted to include is the display of the most popular animes for each genre on demand. In order to turn the visualization very interactive, we decided to do it by zooming into the selected genre, making it cover most of the screen and creating another bubble chart inside it consisting of the most popular animes. The user could then jump from one genre bubble to another by clicking on neighbouring genres. If there are animes that appear on both genre, then they would also smoothly travel from the first genre to the second. A first implementation of it was just to focus on the selected genre and scale every genre bubble proportionally to it. However, this would cause performance issues when selecting the small genres as it would scale the most represented animes to enormous sizes. The issue was almost unnoticable on firefox but google-chrome had a hard time dealing with it.
To address the issue, instead of scaling up the other bubbles, we just push/offset them by the size of the bubble chart so that they lie close to the boundary and remain at the original size.
The difference between both implementations is shown below.

Bubble chart: version 1	Bubble chart: version 2 (final)
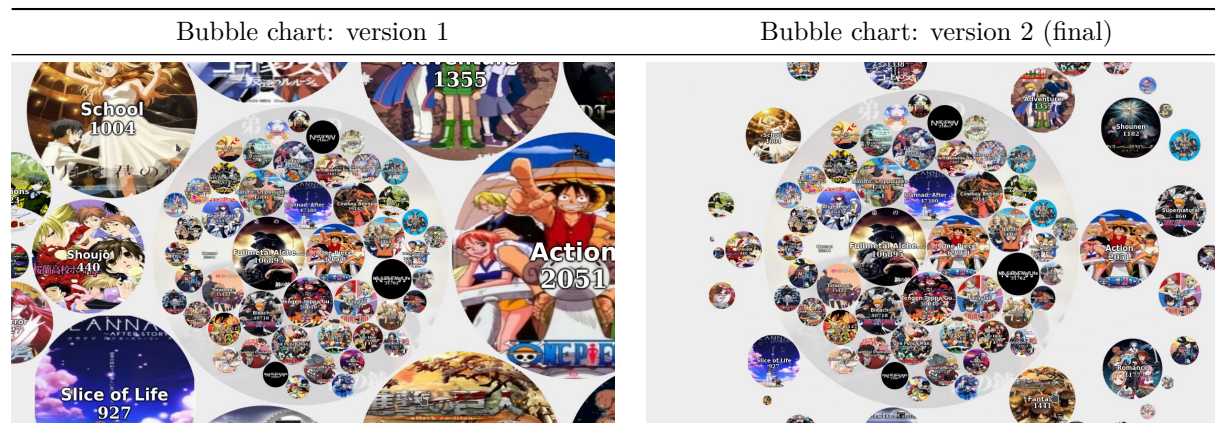
Finally, depending on what the user is focusing on, information about the chart/genre/anime is displayed on the side.

5.3 Sankey diagram: 711 studios captivating the world

For this diagram, the aim was to show informations about the most consequential anime studios. We thought about doing a Sankey diagram to show a nice link between the number of animes produced by the studio per genre, and then sorted per score's mean (rounded to the closest integer). This provides an insightful visualization showing directly the influence of the top studios over anime's world.

To have a nicely interactive diagram, we used `amcharts` library, as shown here. Its API brings a lot of different ways to interact easily with the representation and offers many useful possibilities with animations handled automatically.

The main difficulty here was to show groups of links, going from a studio to the score, passing through the wanted genre. to handle it, we used an `id` for each link, with a concatanation from the name of the studio (2 first letters and last one), with the genre, and a number depending on if it goes to a genre or to a note (for example, the link going from Toei Animation to Action will have its `id=TonAc-0` and for the same studio going to a note `id=TonAc-1`). Then we can use the first part of the `id` to know if it belongs to the same group, and highlight all of them when hovering one of its links.

We had then to choose the number of studios and genres to show up in the diagram, and decided to go for a balance between information and performance. Indeed, each new node adds a lot of links, affecting performance directly. We went then for 10 studios (chosen on the number of animes produced) and 5 genres (depending on the number of animes), to still have enough information to make the diagram relevant.

Like the other diagrams, we put information on the sidebar depending on what is hovered, and decided to show pertinant information for each studio and a short description for each genre.

5.4 Chord diagram: 11292 voices giving life to characters

Finally, the last diagram is about the voice actors, their popularity and links to other actresses or actors. This was done by using a chord diagram, so that we could easily have a node per actor and a link if they played in the same anime or movie at least once.

Again, we used `amcharts` library with the documation of the chord diagram here.

We sorted the voice actors depending on the language they dub with a button to be able to go from one language to another, and chose the top ones after the number of `member favorites` they have. We also created a dropdown list to change the number represented, by using a dictonary on the dataset.

The issue with this diagram was on the script to fetch and sort the data, to take into account all possibilities of dropdowMenus. We had to fetch every anime information to find the actors per language and character, to

be able to retrieve then information per actor, before sorting them, taking the different top ones per language, checking who played in the same anime than another one and putting informations on a JSON file for each node and link each pair. Another file was made for the sidebar which stores the informations for each voice actor, after finding the top 3 characters they dubbed (depending on the popularity of the anime the character was in).

Then, we thought about showing more information directly on the diagram, to have a more general view of it all. A circleBullet to show the popularity and number of anime/movies dubbed by each, as well as the possibility ot highlight every actress or actor linked to the one hovered. With all of that, someone looking at the diagram can directly see insightful information in a single glance. Finally, we added buttons to change how the diagram is sorted or which top and language we want to show, which was done by getting the right data from its dictionary (for example `top 10 -> "English" -> data`) and sorting it in a different way, as `amcharts` will create nodes depending on their appearance on the dataset.

6   Future work and limitations

7   Conclusion