

COM 480 DATA VISUALIZATION



F.R.I.E.N.D.S

CHART CONQUERORS

Enrico Benedettini
Mateo Echeverry
Jorge Encinas

2023

The process

When we first met to discuss potential ideas for the project, we were only coming up with **broad ideas** such as, data from football matches, from different songs, or even from Amazon products. However, **these ideas did not connect** with things we truly liked, it wasn't until we started thinking about what we enjoy that we came up with the idea of using a dataset from one of the best series in the world: [Friends](#).

Friends is a 90's Comedy TV show, based in Manhattan, NY, about 6 friends who go through **just about every life experience imaginable together**: love, marriage, divorce, children, heartbreaks, fights, job hirings and firings and all sorts of drama. We decided we wanted to **get better insights about the characters and their relationships** and share it with more Friends fans.

We searched for a **dataset** that could **fulfil our needs and goals**, and eventually we found a dataset containing all the scripts from the whole series, divided into utterances, conversations, episodes, chapters, and seasons. Not only that, but the dataset also contained some **metadata information** for every utterance such as, the **emotion**, and if **other entities** (characters) were **associated** with that utterance. This metadata was not present in all utterances, but **it was big enough to make an analysis on it**.

Once we got our data it was time to **think about what we were going to do with it**. What visualizations would Friends fans enjoy? How can we make them **engaging and fun**? Those were not easy questions to answer. After some consideration we came up with 3 potential visualizations that we could create...

After having some basic idea of what we wanted to accomplish we started working on the website using HTML, CSS, and JavaScript. Initially, **each of us worked on a separate** HTML until having a **sketch**, and then merged everything into a single file that would become the final webpage.

Once we got our review from the **second milestone**, we first started **getting the data needed** for our visualization through Python Jupyter Notebooks, and once we got our visualizations up and running, we had a stack of our three visualizations one below the other. It was time to start **writing our story** and think how we would link every visualization to keep the reader's attention. We started writing text on our visualizations, and then verified everything was **coherent and consistent**.

At that point we already had a **functional website with visualizations**, and texts explaining the visualizations and giving some of the insights gotten from them. After that we focused completely on the **design**, we made sure all the bugs in our visualizations were fixed, and that the website was **clean, neat, and minimalistic**.

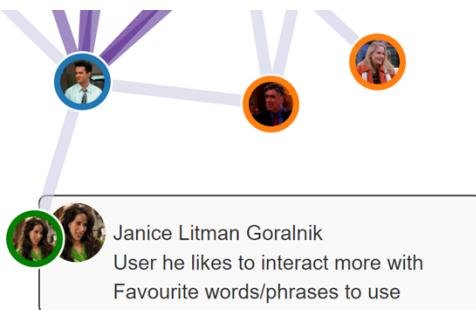
Once that was done, we were able to record the **screencast** and write this **process book**.

The sketches

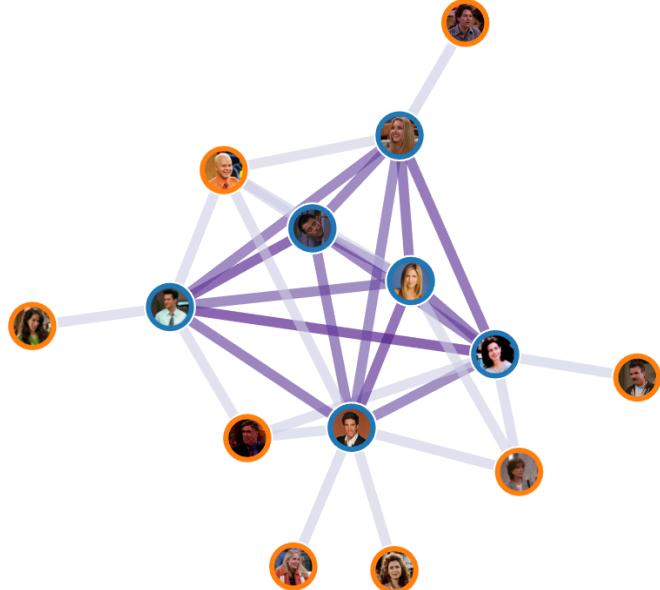
The Friendship Web

We wanted to make the Friends fans be able to reminisce about which characters **interacted with each other**. Friends has 10 seasons, and the series is written such that apart from the 6 main characters (Monica, Chandler, Phoebe, Joey, Ross, Rachel) **other secondary characters appear in the show** for one or two seasons **and then disappear**. This might lead to forgetting about important characters from past seasons.

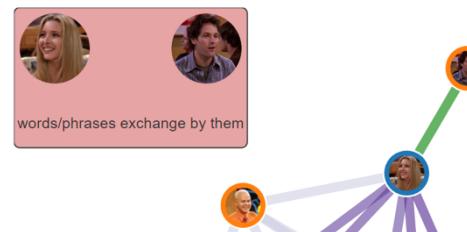
Also, we wanted to show what each character's most unique words are, and display each of these words in different seasons to visualize how their speech and expression evolved over the years.



The visualization will show the main characters and the most important second characters. The color of each edge will represent the number of interactions or gossips between each character; the darker the color is, the more interactions between the pair. When a node is clicked, general information about that character will be displayed, such as the name, maybe a description if it's a secondary character and the character they like to interact more with or gossip about, as well as words that are used especially by that character. Alternatively, when clicking an edge, we will display information related specifically about the pair, i.e.: topics do they talk about, speech patterns shared between them. With this plot we aim to give an interactive and insightful experience to users to understand and **explore more about the different relationships about the characters**.



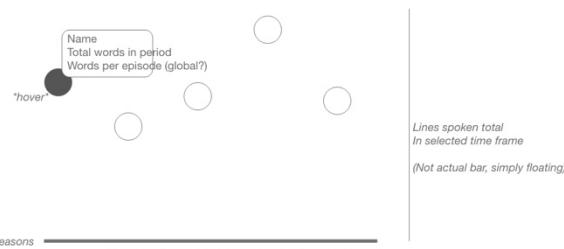
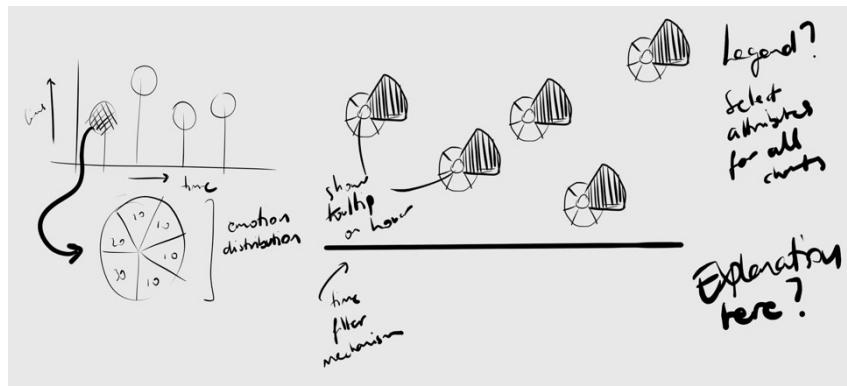
This visualization will **represent how characters interact with each other**. We want to display which characters talk more with who, which characters like to gossip and who do they like to gossip about, what vocabulary is exchanged when talking about a certain character, and the phrases commonly associated between them.



The Emotion Roulette

The goal of this visualization is to **show the emotional split in the main characters' dialogue**. Each character will be displayed as a pie chart highlighting the different emotions and their proportion over their whole dialogue, the user being able to hover over each sector to uncover additional information.

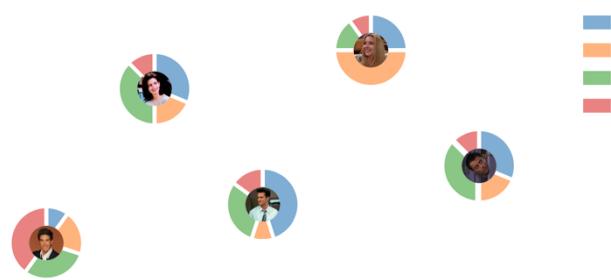
Aside from them will be a legend providing the **color coding** used as well as some additional hints and tips relating to the data being displayed. **Clicking on legend items** will enable filtering the sectors by the different emotions and observing which of these characters show more or less of it.



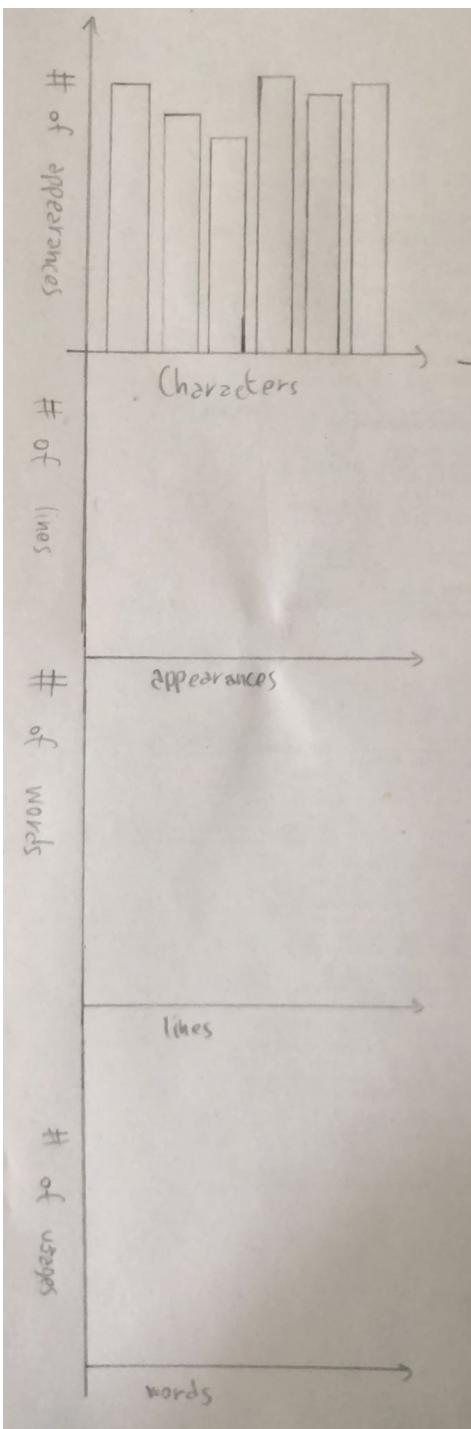
The emotional data is taken from **individual dialogue lines from the different episodes**, so the pie charts will be placed along a horizontal axis **enabling selecting from the available range of seasons**.

Vertically, the pie charts will be sorted with variable height **depending on number of lines spoken in that period** by the character. This information can be displayed upon clicking or hovering each character's image, centered in the pie chart. The sketches shown give the initial concept design for this visualization.

This prototype to the right shows a more detailed view of a possible look and behavior of the final visualization. Some of the more cosmetic aspects (including hover actions changed for static displays in suitable positions) could be modified without sacrificing the experience and meaning of the visualization.



Appearances and words



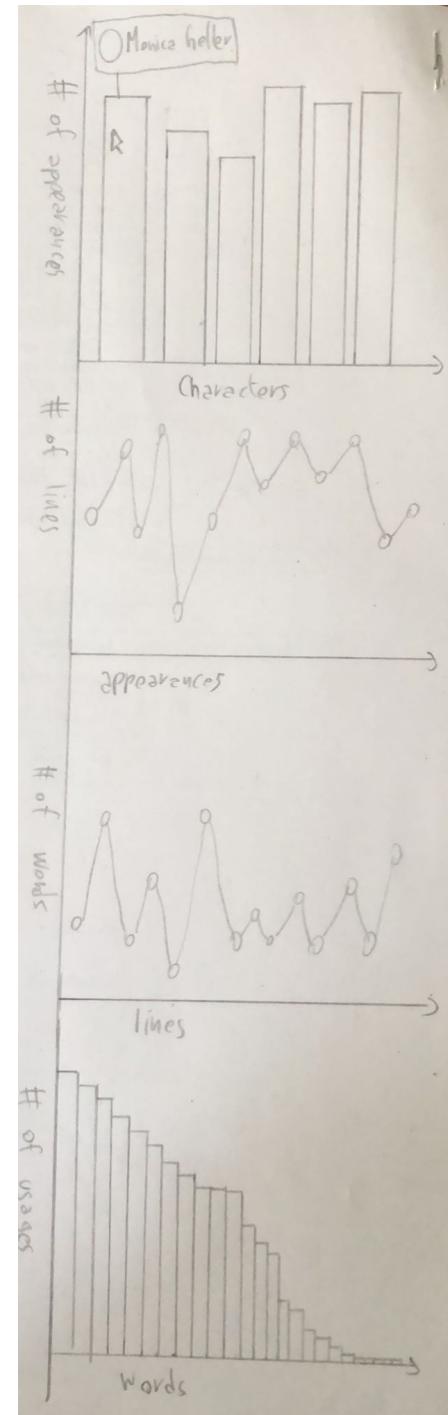
The idea for this visualization is to show **the link between a character's appearances** on screen, the **number of lines** and **numbers of words** to provide insights on how this are related between each other.

The bar chart will display the number of appearances of each character. Aligned with it, right below it, there will be **two scatterplots** initially empty that will display the number of lines on each appearance and the number of words on each line of the character we're hovering or clicking in the bar chart. Also, a bar chart with the character's favorite words will be displayed, displaying for each word their number of occurrences sorted from most to less used. This sketch shows the initial state of the visualization:

After some internal debate, we considered this visualization did not provide sufficient insights into the data in the way we wanted. It broke the overall aesthetic of the webpage and was not up to the standards of the others.

It was in this context that we decided to create a supporting visualization, more in line with the rest of the analysis during the last development days of the project. This visualization is available in the final webpage and explained in more detail in the [Final Viz](#) section below.

Since this was created as a backup, there are no sketches, and was implemented directly on the webpage.



Challenges

Our first main challenge was to understand that **the only data we had to study was text**. We didn't have any other values, such as the duration of the episode, the number of scenes per episode... So, we had to work with what was available, luckily for us there are many things you can do with text, although some of them are not trivial.

In the graph visualization, the first thing we struggled with was **how do we define an interaction?** An interaction can be when only the 2 characters are present in the conversation, or when they are both present in the conversation without considering how many more people are in the conversation, or if they mention each other's names. The way we solved this issue was to get **conversations where both characters were present independently of the number of characters in the conversation**.

Another challenge during this visualization was at the time of getting a character's, and a pair of character's **most characteristic words**. If we just get the words that have been said the most by a character, we will only get **stopwords**, such as "ohh", "sorry". Obviously, these words don't represent any character. The solution we employed was to calculate a **uniqueness score** for each word used by a character or a character pair in a conversation. This score was based on its relative frequency of use compared to other conversations or characters.

This approach allowed us to identify **words** that were **distinctly common** in one context, as opposed to words that were common across all conversations or characters. The process involved **cleaning and tokenizing** the dialogues, excluding common stop words, counting word frequencies, and finally, comparing these frequencies across different contexts to compute the uniqueness score. The top 20 words with the highest uniqueness scores were then selected as the most characteristic words for each character and conversation, providing a meaningful representation of their distinct dialogue patterns.

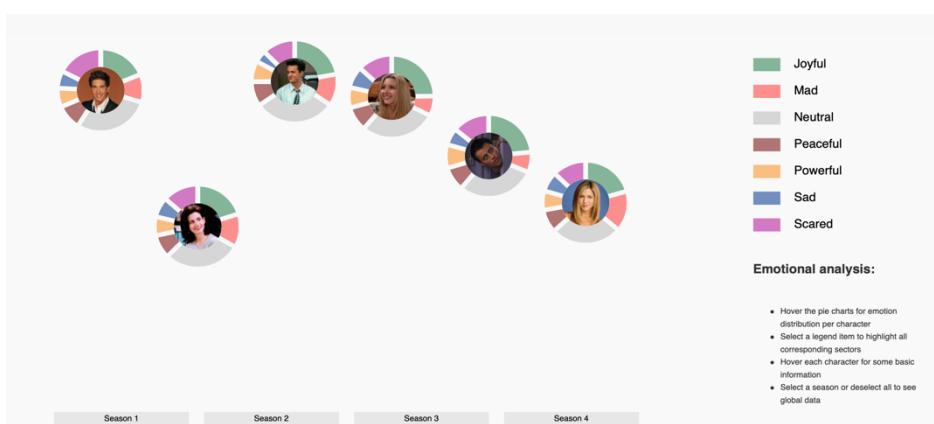
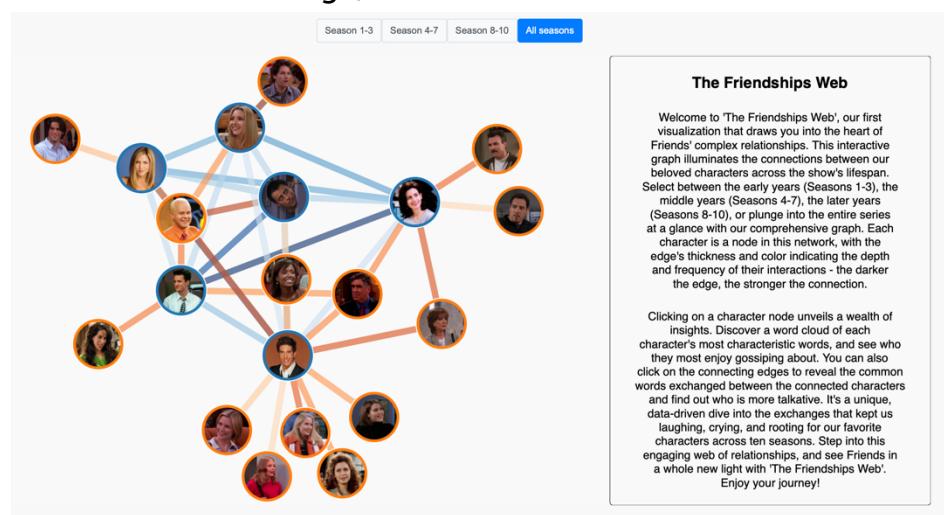
Similarly, for the emotional analysis, the available data was stripped down to that of valid entries with actual metadata. This again required a **cleaning and filtering** process. We decided to use the main emotion per line, instead of considering sub-emotions mentioned in the dataset. These were preprocessed for each main character for visualization in the different pie charts. **Identifying the different pie charts**, and the different **sectors within each chart** proved a technical challenge, but we managed to successfully replicate the original sketch with the intended functionality. Repeated toggling and selecting legend items and sectors also proved cumbersome during the development process but was implemented in the end.

As discussed, the issues with the third visualization were solved by **creating more supporting material at the last minute**.

Final VIZ

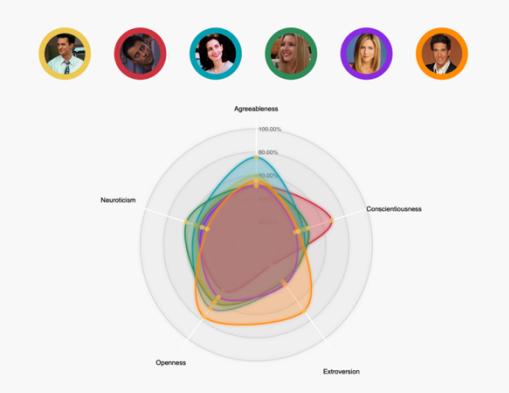
Here are the results for the visualizations, with the decided changes from the sketches presented previously:

First off, in **The Friendship Web**, instead of carrying out with the initial design of a small box when you click on a node or an edge, **we decided to make a much bigger and static box** to be able to **show more things**. When clicking on a node, we added a **brief description of that character** to refresh or introduce characters that may not be familiar, and their gossips. In the edges, we decided to add the number of words exchanged between them in a horizontal bar chart to see which character they like to talk the most. We also implemented the option to **get a graph from the early, middle, and later seasons**.

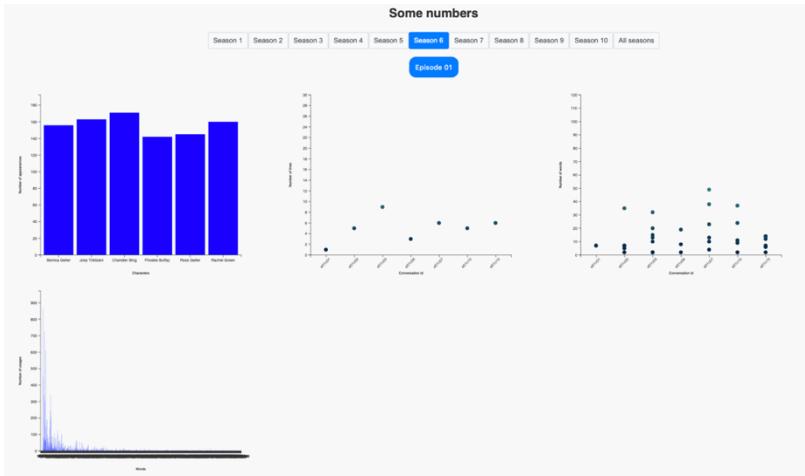


As discussed in the sketches section, this visualization was never expected to be implemented. In this visualization analyzed 5 personality traits. We thought that a good way to compare them was through a **spider chart**. In the visualization, you can **hover over the characters**, and the area corresponding to that character will increase its opacity and decrease the other areas.

Then, for the **Emotions Roulette**, additionally to the sketch, **we added more depth to the visualization by including word analysis in the form of word cloud associated to each emotion and character** in every given season. It is revealed upon clicking on any sector.



You can also do it the other way around, you can [hover over an area](#), and all the characters except the one corresponding to that area will decrease its opacity.



Finally, the final intended visualization of the third sketch came out as shown here. Upon [selecting a bar](#) from the bar chart, the rest of the plots are [updated with the corresponding data selected](#) and show, in the form of scatter plots, information on character appearances and lines of dialogue.

Here's the breakdown of the task distribution and work for the project.

Milestone 1

- *Enrico Benedettini*
 - o Dataset search
 - o Dataset description
- *Mateo Echeverry*
 - o Dataset search
 - o Exploratory data analysis
 - o Problematic
- *Jorge Encinas*
 - o Dataset search
 - o Related work

Milestone 2

- *Enrico Benedettini*
 - o Sketch for third visualization
- *Mateo Echeverry*
 - o Sketch for first visualization
 - o Visualization skeleton
- *Jorge Encinas*
 - o Sketch for second visualization
 - o Visualization skeleton

Milestone 3

- *Enrico Benedettini*
 - o Bar and scatter plot visualization
- *Mateo Echeverry*
 - o Screencast
 - o Friendship Web
 - o Additional support visualization (spider plot)
 - o Web text
- *Jorge Encinas*
 - o Process book design and composition
 - o Emotion Roulette