# Team Shakespearits

## Process Book

Marie-Alix Gillyboeuf

Danaé Papadopoulos

Paula Dolores Rescala

EPFL COM-480: Data Visualization Spring 2022

# 1- The Journey

## 1.1- Brainstorming and Topic Selection

Our journey for this project began by choosing what we wanted to show. Our first meeting involved brainstorming ideas based on our common interests. We wanted to do something lighthearted involving some art form, so our first idea was to use movie and TV show data, in particular coming from Netflix. However, upon further consideration, we came to the conclusion that this topic has already been thoroughly explored and we wanted to introduce something more novel. Next, we considered doing our project on street art by making a map with the locations of graffiti in cities across the world along with some analysis of the art style itself. Unfortunately, we were not able to get any data for this, so we had to abandon the idea. Finally, we ended up considering a project that would be somehow related to literature, something we all love. We wanted to make our project specific but something a wide enough audience could enjoy, so we chose to use Shakespeare's plays as a topic since he is arguably the most famous playwright of all time. We quickly found a dataset on Kaggle including all lines of Shakespeare's plays labeled with the play name, Act and Scene number, and speaker. Additionally, we crafted our own dataset that contained character information clustered by plays, with their name, role, description, gender, and relationship to other characters.

## 1.2- The Vision

After choosing the dataset, we had to find what we wanted to visualize from our data and how we wanted to do it. Together, we discussed three main ideas that would be meaningful to share:

1. The creation of a **network** showing character relationships by play: One notable feature of Shakespeare's plays is the relationships of characters within the play. Many plays focus on themes such as love, family, betrayal, etc. all which have to do with the relationship structures of the characters. Thus, our network would show linkages between characters and provide the user with the opportunity to learn more about each character by hovering over the node that represents them, and learning more about each relationship by hovering over each edge that represents one. We also provide a drop-down menu that allows the user to select the play of interest.

2. The visualization of topics covered in Shakespeare's plays: Shakespeare wrote three types of plays: comedies, tragedies, and histories. These genres are very specific, but we wanted to break down exactly what topics were covered in each one to maybe get an idea on why they were all so famous.

3. The visualization of character personalities through **word clouds** using features of the lines they speak in the play: Often, when we enjoy a movie, show, or piece of literature, we enjoy it because we identify with, relate to, or simply admire one of the characters in the work. Thus, it is fun to analyze and visualize why we might admire them, so we thought working with character personalities would be the best way to do so. Since we had data on every line in every play, we decided to use word clouds for each character to visualize their personality. This way, we could communicate what topics and words were most frequent in the particular character's speech.

We each brainstormed sketches independently on how we could create such visualizations. We also envisioned the website to be an informative space beyond visualizations, so we planned to have Shakespeare's biography on the home page and a quiz page that would allow the user to test their knowledge on the playwright's works. We also planned to include a small visualization on the home page showing all of Shakespeare's plays divided into their genres.

## 1.3- Creation

Disclaimer: The following description is not accurate chronologically speaking since features were developed in parallel by all team members.

Once our ideas were coherent, planned, and sketched out, it was time to implement what we wanted to do. In general, the website architecture did not change much from our brainstorming and planning steps. Here, we go through each tab on our website and show both the original sketches, what we changed from them, and the final results.

**Home Page.** Our plan for the home page was as is shown in the left image in Figure 1. The actual result is what is shown in the right image. Our final result is very similar to what we had envisioned.
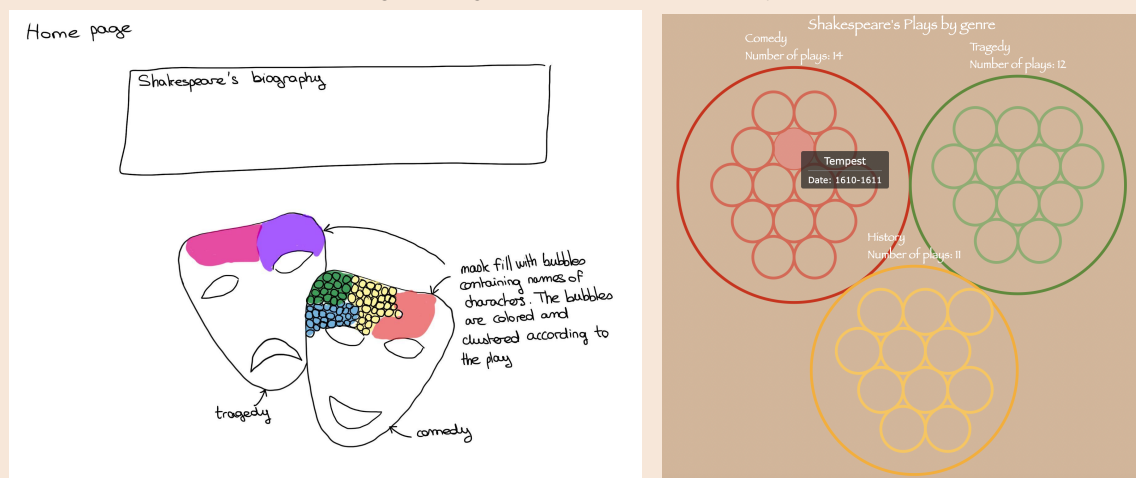


Figure 1. The sketch of the home page and the actual implemented bubble visualization. Pop-up window showing the name and the date of a play.

A small biography of Shakespeare serves as a nice introduction to the user in our final result, just as we had envisioned. Additionally, we decided to add a bubble chart containing grouped information according to each play's genre (comedy, tragedy, or history). This would allow the user to get a general and concise overview of what Shakespeare's work as a playwright entailed. This part changed a lot compared to our initial sketch, since originally we wanted to create this bubble chart in the shape of the famed comedy and tragedy masks that we also use as our logo. However, we could not figure out how to display the bubble chart in the shape of the mask and we also forgot to originally recognize that Shakespeare's work actually fell into three not two categories.

**Network Visualization.**

After visiting the home page, the user is eased into the world of Shakespeare's plays, and they can learn more about plays and their characters by playing with our network of relationships. The user can choose among eight plays ("A Midsummer Night's Dream", "Hamlet", "Julius Caesar", "King Lear", "Macbeth", "Othello", "Romeo and Juliet", and "The Merchant of Venice") and the corresponding network of relationships is displayed. For the



implementation, we decided to use Anychart library. We had first to convert our dataset into the appropriate Json file that anychart.graph requires. Once we have done that, we created the first network with all characters from all the plays. The resulting network was very messy and confusing, thus we decided to add a button that allows users to choose one particular play, and then displays the corresponding network. To add more information, and to make the network more interesting we added a pop-up window that shows the role and the description of a character on mouse hovering.
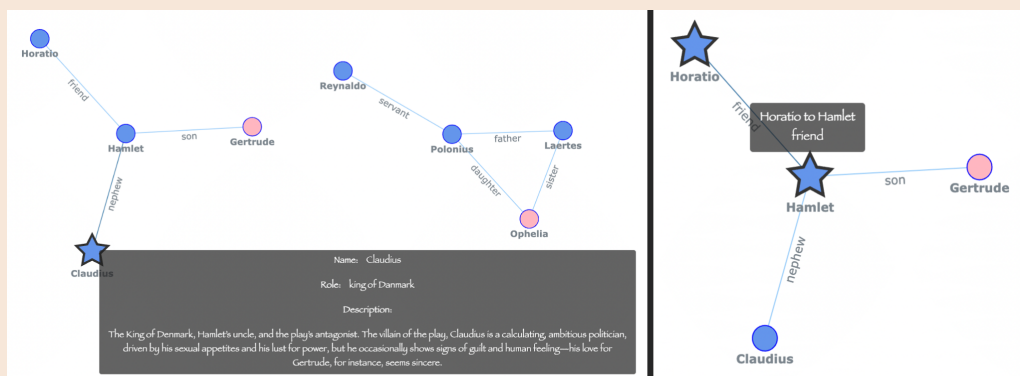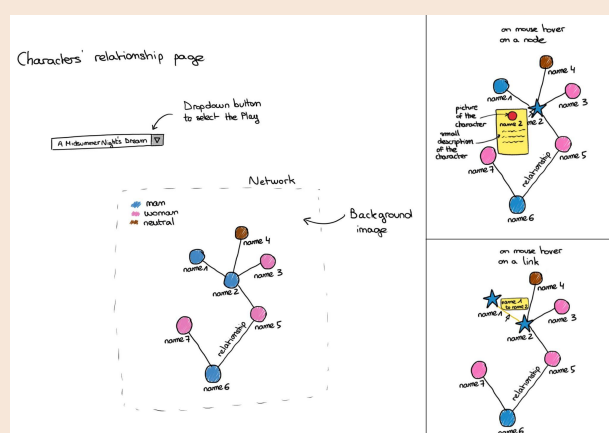


Figure 2. Network of relationships of Hamlet's characters. Pop-up window showing character's name, role and description (left image). Pop-up window showing character's link (right image).

The window contains information about the character such as its gender (man 'M', woman 'W' or none), its role in the play and a small description of its personality. Globally, apart from visual tweaks such as using nicer color and adding animations, we stuck to the original idea.

**Topic Visualization.**

Next, the user is invited to explore our topic visualizations for Shakespeare's plays. In this section, we take the most prevalent topics in the works extracted by Natural Language Processing techniques in order to trace the evolution of their appearance across plays. A few topics are covered throughout a play, however their intensity may differ in each new scene. We are especially interested in analyzing the evolution of the following four topics in each play: love, family, power and war. As we planned, we were able to create the beeswarm plot for the aforementioned topics. The radius of the bubbles varies based on the number of sentences the player expresses about a particular topic. Hovering on a circle of the plot displays a pop-up window that shows the number of the act and the scene, the name of the player as well as the number of times the topic was mentioned by the player during the same scene.
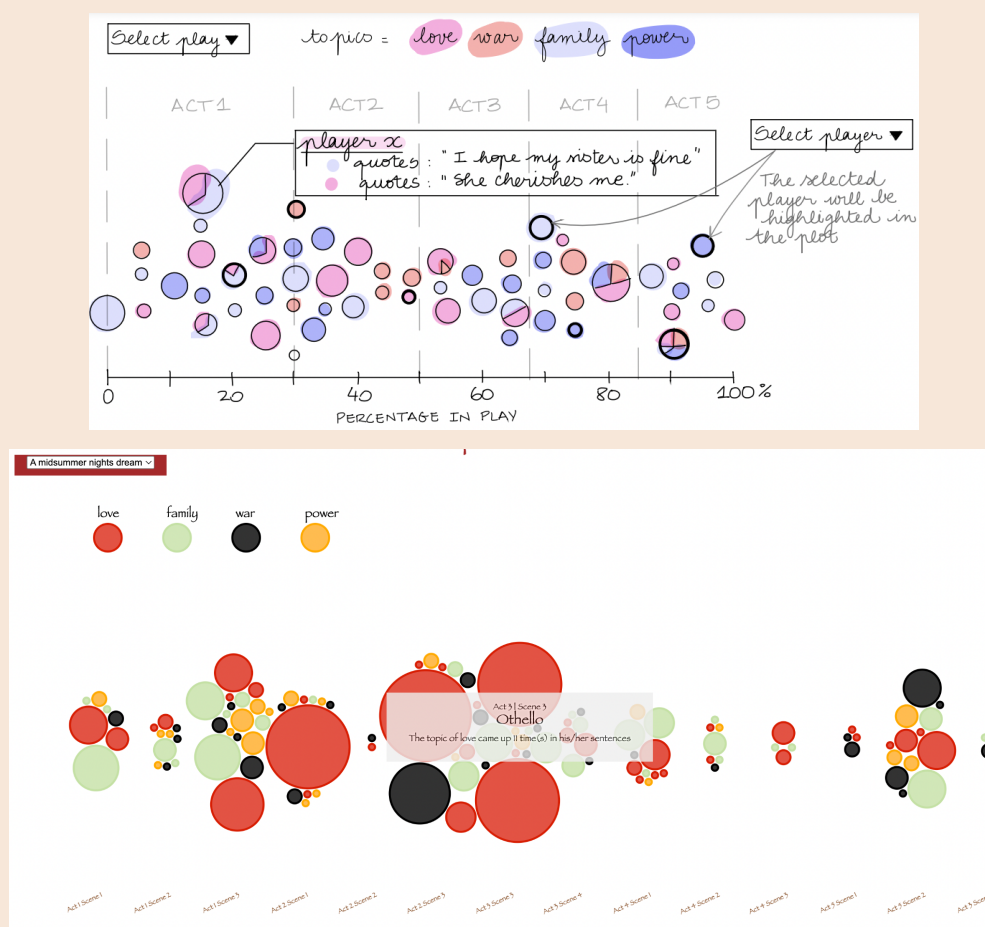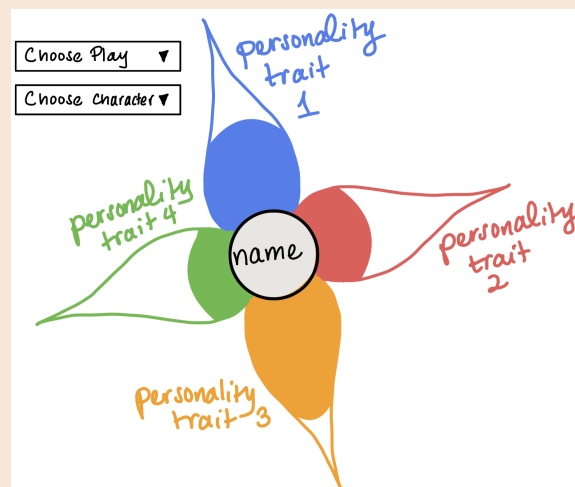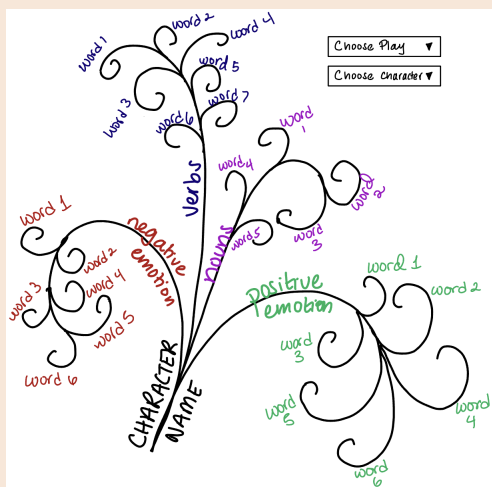


Figure 3. Beeswarm plot. Pop-up window showing the act number, the scene number, the character and the number of times the topic was mentioned by the player during the same scene.

**Personalities Visualization.**

Finally, the user enters our final visualization tab where they can see how each character's language is representative of their personality in the play they belong to. We originally had two ideas for visualizing character personalities. The first was to create something akin to word clouds for each character with words separated spatially and by color on branches according to how we classify words. The next idea was to extract personality traits using machine learning and create an equally sized flower for each character with petals filled to the extent to which each character exhibited each personality trait. However, our final visualization as displayed on the website became something much different. Using machine learning to extract personality traits from the text itself was too difficult a task for the purposes of this project. Although several methods were tried including studying different personality types such as "The Big 5" and attempting to extract features from it as well as using previously trained models to test our own corpus, we decided against using these results since they were a bit less representative of the data itself and more of a machine learning project. After doing so, we were also unable to create the word cloud in the format that we wanted with the shape and branches we originally envisioned. However, we were still able to create the word cloud with our four categories per character by using part of speech tags and sentiment analysis from NLP. We visualized the word cloud using anychart.

Figure 4. Word cloud example from our website.

## 2- Technical challenges

While working on the project, we faced some technical difficulties and we had to find out how to solve them. Below is a non-exhaustive list of those challenges.

- Visualization: The first main challenge we faced was creating the visualizations themselves. How to make Shakespeare more attractive? How to visualize a play? This was hard and we had to find a way to present plays and design interesting visualizations accessible and understandable for anyone.

- Bubble chart. Initially, we wanted to create a bubble chart such that each play's name would be contained into a bubble and the bubble would then either fill the tragedy mask or the comedy mask according to the play's type. But we faced two problems. First, we did not find how to design a bubble chart in a given shape. Secondly, we realized that there were not two types of plays but three (tragedy, comedy and history) and we only had two masks... Therefore, we decided to design a simpler bubble chart with three main groups corresponding to the three different types of plays.

- Anychart graph: We had to convert the datasets into a Json file containing the appropriate keys' name to match the ones requisite by anychart. Then we had to understand how nodes and edges were constructed to build the corresponding network.Finally, we had to figure out how tooltips worked in order to create the pop-up window we wanted.

- Web design: We had to deal with common issues of web design and web development. Indeed, dealing with a whole bunch of web libraries to create a really nice and attractive website. We are convinced that design is very important even though it is subconscious for the user.

- Beeswarm plot: Initially, we wanted each point scattered in the plot to correspond to a pie chart. This pie chart would have given us information about the proportion of topics

mentioned by each player, for each act-scene pair. However, we didn't have enough data thus the figure wasn't visually pleasing. We thus ended up representing one circle per topic for each player instead of a pie chart.

## 3- Peer assessment

During the whole project, we always tried to think and take every important decision together. Each one of us brought new ideas and we always tried to mix them thus everyone was satisfied. After the second milestone we decided to split up the coding work into individual parts, so the following is an overview of task distribution:

- Brainstorming: Danaé, Marie-Alix and Paula
- Research of the dataset: Danaé, Marie-Alix and Paula
- Design Ideas
    - Home Page: Danaé, Marie-Alix and Paula
    - Network: Marie-Alix
    - Topics: Danaé
    - Personalities: Paula
- Implementation:
    - Website Architecture: Danaé, Marie-Alix and Paula
    - Home page: Marie-Alix
    - Network: Marie-Alix
    - Topic: Danaé
    - Personalities: Paula
    - Quiz: Marie-Alix
    - About us: Marie-Alix
- Process book: Marie-Alix, Paula, Danaé
- Screencast: Paula

## Enjoy the website !

🎭