

# Milestone III - Process Book

Vera Fristedt Andersson, Mari Sofie Lerfaldet & Thomas Bech Madsen  
*Data Visualization, COM-486, EPFL, Switzerland*

May 28, 2020

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Dataset</b>	<b>1</b>
2.1	Exploring the data . . . . .	1
2.2	Cleaning the data . . . . .	2
<b>3</b>	<b>The Webpage</b>	<b>2</b>
3.1	The Map of France . . . . .	2
3.1.1	Design . . . . .	3
3.2	The Home Page . . . . .	3
3.3	The Varieties Page . . . . .	3
3.4	The Search Engine Page . . . . .	3
3.4.1	Helping the user . . . . .	4
3.5	The Data Page . . . . .	4
<b>4</b>	<b>The Database</b>	<b>4</b>
<b>5</b>	<b>Challenges</b>	<b>5</b>
5.1	Design, D3.js, TopoJSON and SVG . . . . .	5
5.2	The Search Engine . . . . .	5
5.3	Browser and screen size . . . . .	5
<b>6</b>	<b>Changes</b>	<b>6</b>
6.1	The data . . . . .	6
6.2	Design . . . . .	6
6.2.1	Map of France . . . . .	6
6.2.2	Varieties . . . . .	6
6.2.3	Search Engine . . . . .	7
6.2.4	The Data . . . . .	7
<b>7</b>	<b>Peer Assessment</b>	<b>8</b>

# 1 Introduction

This visualization project is created to make it easier for amateur wine lovers to explore different wines in France. As there are many people out there who are curious about wine, but who do not have any particular knowledge about it. We wanted to make an interactive and fun way of exploring the different wines in France, by making it easier to explore a wine's origin, price and quality. With our visualization the user can get inspiration for the next time they choose to purchase wine.

To visualize the data we have chosen the following methods:

- Two interactive maps over the different provinces in France.
  - The first map will give the user general information about a province
  - The second map will give the user more detailed information about the wines that originate within a chosen province.
- A search engine where the user can explore the wines from different provinces, regions, varieties and within a chosen price range.

By creating the visualization as a story, in which the user will go through different chapters, our aim is that the target group will find it inspiring to explore the different wines that are made across France. The start page is the most simple and general page, the second page includes more detailed information and the third page is very detailed. By dividing the website in this way, we hope that the user will feel that it is easy to go through the website, while at the same time being educational and fun.

The website is optimized for use in Google Chrome, with a screen size of 13”.

# 2 The Dataset

In this project we use the dataset, Wine Reviews [4], about wine from different parts of the world, where the data is gathered from *Wine Enthusiast* [2]. Wine Enthusiast only includes wines rated 80/100 or higher by sommeliers.

## 2.1 Exploring the data

We knew from the start that we wanted to work with only one country. Therefore, our first step was to explore the dataset to be able to decide on what country to work with. We already had a clear idea on what visualization we wanted to do, and that was an interactive map of a large wine producing country. Italy, France and the United States were the three countries that were of interest to us, as all wine drinking people have most likely come across wines from these three countries. In order to make a decision, we looked at how much data that was contained in the dataset for each of the three countries. We also looked at how much data that was missing for each country. Finally, we looked at how many provinces each country has in total and how many of these provinces that were represented in the dataset. We realized that it was important for us to have a large percentage of the country's total provinces represented in the dataset because we felt that it would be more fun to visualize a map where many provinces have data on wine, rather than just a few. As it turned out,

France had the largest percentage of represented provinces, there was a lot of data on the different french wines, and there was little information missing.

## 2.2 Cleaning the data

Upon choosing France it was time to look into the data in more detail. We explored the data and did some data cleaning in Milestone 1 that we finally did not end up using. This is described in Section 6. The data cleaning that we did use, was done after Milestone 2. During Milestone 2, we realized that we did not have as many provinces as we thought we did. It turned out that four of the provinces in the dataset were not actual provinces so we were not able to include them in our skeleton visualization. However, we realized that two of the false provinces included gathered data from several other provinces so if we were to clean this data we could actually get even more provinces than what we initially had planned. We cleaned this data by looking at the regions included in the false provinces, googling them and looking up their actual province. These false provinces were then replaced by the correct province in the dataset. The other two false provinces were actually regions part of existing provinces so we edited them to be part of the correct province. Thanks to this cleaning we went from having 7 to 12 wine producing provinces.

## 3 The Webpage

To begin creating the webpage, we used quite some time to develop sketches as an illustration, as we believed that it would make working together easier. When the sketches were done, it proved to be an important step as it helped in developing the design faster, since we all had something tangible to base the changes on for the actual implementation. Throughout the development of the page, we kept iterating over the design into a better solution.

In order to set up the webpage, we chose to use *GitHub Pages* and a template from *Bootstrap Temple* [3]. The template was only used to get the fonts and the shell of the webpage. To further adjust, we added our own *CSS* rules for tables, the map visualizations and different objects, e.g. the navigation bar that leads to the different pages.

### 3.1 The Map of France

The map of France has been created using *TopoJSON*. The *TopoJSON* map of France and its provinces was gathered from *Khartis* [1], a open-source thematic map service, licensed under the *MIT* license. The provinces in the map are from the 2015 revision with 22 provinces. Today, France includes 18 provinces. However, we felt that this change did not affect our visualization due to three reasons. First, since the new provinces are simply larger groupings of the previous provinces, the borders are still the same when it comes to the map. Second, some of the new province names would lead to confusion when it comes to wine. For example Champagne-Ardenne and Alsace are well know wine provinces that today go under the common province name Grand Est. Third, the provinces in our dataset were grouped according to the 2015 provinces.

### **3.1.1 Design**

The map takes up half of the Home page and the Varieties page. This is to catch the users attention as we want them to interact with it. The provinces that we have wine data on, are colored in light beige and the provinces that we do not have wine data on, are colored in white. When the user hovers over one of the wine provinces, it will change in color to burgundy red and when hovering on any province, the tooltip will display the province name. This color change is used to indicate that the province can be clicked. When clicked, the map will zoom in on the chosen province and color it in light pink. The colors that are used on the map are colors of different varieties of wine: white, red and rosé. We chose these colors to create a stronger association with the topic at hand - wine.

The text box to the right of the map will display different information depending on which page the user is on and also depending on if a province has been selected or not. When there is no province selected, the text box will show instructions on what the user should do, and once a province is selected, the text box will show information about the chosen province.

## **3.2 The Home Page**

The home page, as mentioned before, is an introduction to the topic at hand. When selecting a province, the text box on the right will appear with some statistics on the chosen province. These statistics include the number of regions, number of varieties, average price and average points of the province. Apart from this, there is also a link to the next page that informs the user that they can click there to see the varieties in the chosen province. For three of the provinces, there is also a fun fact included that informs the user on how these provinces differ from the rest.

## **3.3 The Varieties Page**

The varieties page is very similar to the home page. The difference is how the data is displayed. When a province is selected, a table appears underneath the map. One column displays the varieties of wine produced in the province and the other column displays the regions in which the wine variety is produced. The varieties in the table function as links, in the same way as the link on the home page does. When clicking on a variety, the user is navigated to the third page where the inputs for the search engine will be updated according to the variety that was selected.

Further data is displayed in the text box to the right of the map. Here, the user is informed of the province selected and that they can scroll down to see the table. In addition to this, we have added a bar plot, that shows how many varieties that each province produces, which can be seen in Figure 2. When a province is selected on the map, it will also appear as selected in the bar plot, by using the same color scheme as the map does. The bar plot is also interactive such that the user can hover on or select one of the bars. When doing so, the map will be updated to do the same, and vice versa.

## **3.4 The Search Engine Page**

The search engine page is different from the other two pages. On this page there is no map displayed, instead the user is met with four text fields followed by a large "Search" button. In these text fields, it is possible for the user to enter a province, a region, a variety and a

price range. This is for the users that want to explore the wines more in detail. Depending on the search query that a user chooses, different results will be shown. There will always be a summary of the search and a table that includes all of the wines in the dataset that matched the query. The table includes the provinces, the regions, the varieties, the price, the points, the wine title and the wine year. Along with the wine title, there is an information icon that the user can click on to see the wine review written by the sommelier that rated the wine.

#### 3.4.1 Helping the user

We have decided to incorporate helping functionality and feedback into the search engine page in order to avoid mistakes and confusion. When a user clicks on a text field, a dropdown menu will appear with the possible alternatives for the chosen text field. The province, region and variety dropdowns are adjusted according to each other, such that a user can only choose the alternatives that match the already selected inputs in the other text fields. For example, say that a user has already selected the province Alsace. Then when the user clicks on the region or variety text fields, the dropdown options will only contain regions or varieties that exist in Alsace. We decided to not link the price range dropdown options with the other input fields as we believe that it is more fun for the user to explore the price part on their own.

All of the dropdowns are scrollable if the number of options exceed a certain amount. The dropdown options for region and variety are sorted according to the wine title count, in descending order, to make the more common alternatives visible when opening the dropdown. It is also possible for a user to write in the text field. When doing so, the dropdown alternatives will be filtered such that the visible options all contain the user entered text string. If a search query is unsuccessful, an error message will be shown. When the search query is gathering matching data from the database, a loading icon will appear, informing the user that their request is being processed.

### 3.5 The Data Page

For those that are interested in the data that is used for the visualizations, it is possible to navigate to the data page. This page, and all other pages, is accessible through the hamburger menu in the top right corner. The data page includes information regarding where the dataset is from, what it includes, the project goal and information about us.

## 4 The Database

Setting up the database proved quite challenging and modifications were made throughout the project. We used the service *Firebase* for our data. *Firebase* is a *NoSQL* database service, which means that complex queries like the ones in *SQL ALike*, are not possible. The client must request the entire document and then do all processing of the data locally. We were worried that this would be too slow for our purposes, but we managed to speed up the queries by having multiple perspectives of the data. The trade-off of this approach is high redundancy in the data. For future development of the project, we recommend using an *SQL* database that will allow for complex queries of the data, to speed up loading time and reduce data redundancy.

## 5 Challenges

### 5.1 Design, D3.js, TopoJSON and SVG

As we have little experience using *D3.js* and *SVG*, a big challenge was to get familiar with the technology, while making sure that our visualizations were feasible to make. The designing of the webpage also took some time, as we wanted to be sure that we were making something that would feel clean and simplistic, but at the same time have visualizations that would interest a user and make the experience of the webpage fun and intriguing.

Another challenge was the visualizations of the map. Never having worked with *TopoJSON* before, it was difficult to get the map working. It was quite easy to get it showing on the website, but making it interactive turned out to be very difficult for us. We couldn't figure out how to treat the different provinces as separate objects. Thanks to a lot of googling and reading of online tutorials on how to implement *TopoJSON* using *D3.js*, we finally managed to access the provinces separately through mouse hovering or mouse clicking.

It also proved quite difficult to connect the bar plot, on the varieties page, with the map of France. We wanted them to behave in the same way, such that they mimic each other's behavior when it comes to what is selected or hovered. We finally managed to get it working, but it took quite some time and it required everyone to help out.

### 5.2 The Search Engine

While implementing the search engine page, there were many aspects that proved to be difficult. Adjusting the dropdown options according to the other text fields took a long time. A lot of if-statements were needed in order to handle every case, which does not feel like the best solution. The same applies to the handling of the search query results. It took a long time to implement as there were so many cases to consider. Once again, a lot of if-statements were needed which does not feel like the most optimal solution. However, it works and we are happy with that. The time that it takes to load the table data is another problem. Since we have to load the entire data from *Firebase* and then loop through everything, it takes a while for the data to show. As none of us possessed the knowledge on how to optimize it, we chose to add the loading icon in order to give feedback to the user, such that they do not get too frustrated about the search query taking long.

### 5.3 Browser and screen size

Our webpage has only been tested on *Google Chrome* and is therefore not developed for other browsers. It is also not possible to make the browser window small as this will distort the map of France. We were not able to figure out how to re-scale the map of France and this is why the map is static and the window cannot be too small. As all of us have 13-inch laptops, we know that the webpage looks good and works for this screen size. We do not know how it will behave if it is used on a different screen size.

## 6 Changes

### 6.1 The data

In Milestone 1, we did some data cleaning where we explored the different columns in the dataset. When doing this we discovered that there were some wineries and varieties that had very few occurrences. In order to adjust the data to the visualization, we decided to set a threshold of a minimum of 30 occurrences which included 35 different varieties and 95 different wineries. This reduction also made the data more complete as the amount of missing data was reduced. In the end we chose to not use this cleaning for the website, as it did not have an impact on how the data would be displayed to the user. As we have a search engine, we wanted all of the data to be available to the user.

### 6.2 Design

During this project we have made several changes to the general design of the website, with regards to how it initially was. One of the important changes was to change the way of navigating through the site. We replaced having arrows at the bottom of each page, by having a navigation line at the top, which leads to the three pages: *Wines of France*, *Varieties* and *Search engine*. In addition, we also added a hamburger menu, to add the link to the data page which describes the data used for the website. The color scheme was also specified further, by having shades of beige, pink and red which symbolize the different colors of wine.

#### 6.2.1 Map of France

For the map of France, we chose to zoom in on the selected province, instead of showing it as an entity itself which was our initial idea (see Figure 1). We also added a link so that the user can decide to see information about that province on the varieties page directly, instead of having to navigate to the varieties page and selecting it again.

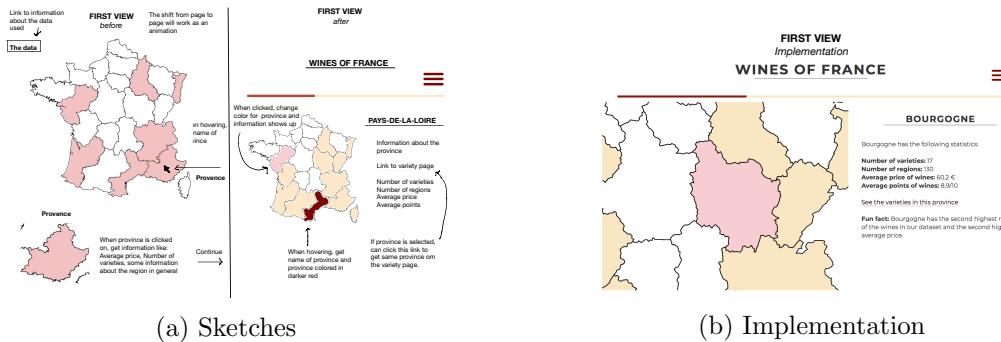


Figure 1: Sketches and implementation for the map of France

#### 6.2.2 Varieties

On the varieties page, we decided to not implement our initial idea of having dots on the map, representing a variety each. This decision was made because we felt that it did not add any significant value. Instead we chose to add a bar plot next to the map, displaying the

number of varieties produced in every province (see Figure 2). This is to help the user to easily understand which provinces that produce more wine varieties. When clicking on a bar, the map will zoom in and select the province that corresponds to the bar. That way, the bar plot also serves as a means of navigation.

Providing links in the varieties table to the search engine was a decision to help the user find the wines they are interested in. Before, the user would have to memorize the variety when going to the search engine.

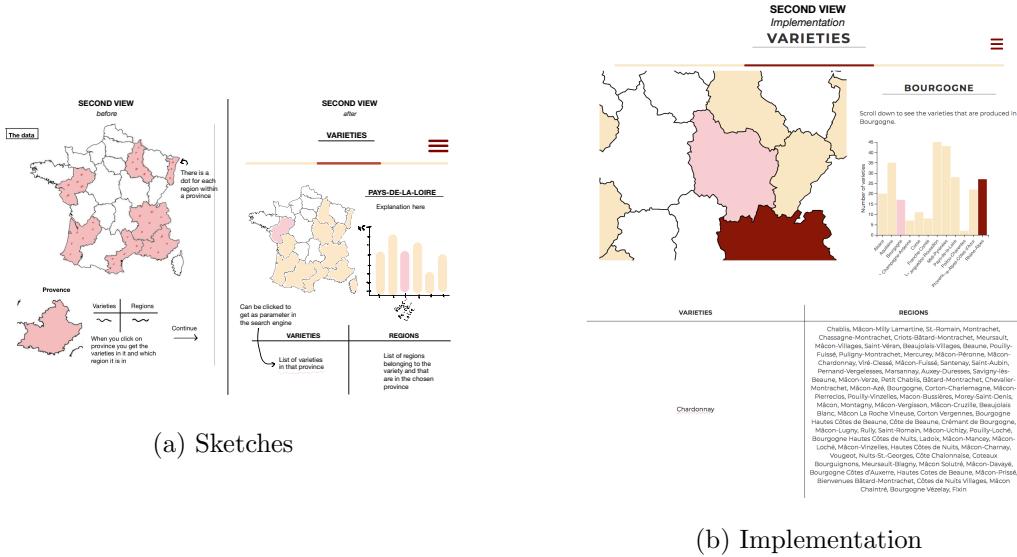


Figure 2: Sketches and implementation for varieties

### 6.2.3 Search Engine

For the search engine page, we did not make any drastic changes. We decided to not display the province map, as we had initially planned, since this would not be beneficial if a user selects input which generates multiple provinces. We also changed the input to be more consistent with the information that the user has gone through. This was done by focusing on regions instead of wineries. We decided to remove wineries completely as this information is most likely not relevant for a wine amateur. For the output of a search, the chosen columns can be seen in the sketch provided in Figure 3b.

### 6.2.4 The Data

On the data page, we replaced the initial idea of having a home button with a hamburger menu. This provides a more coherent design throughout our website. In addition we also added some information about us and the goal of the project (see Figure 4).

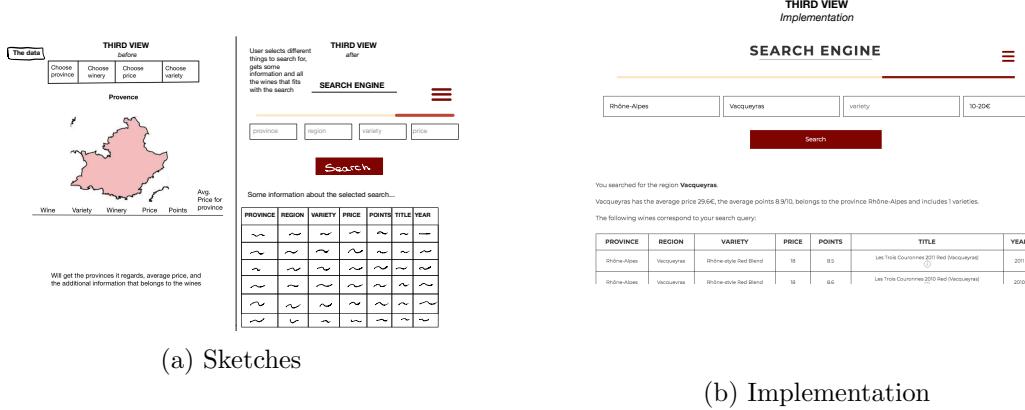


Figure 3: Sketches and implementation for the search engine

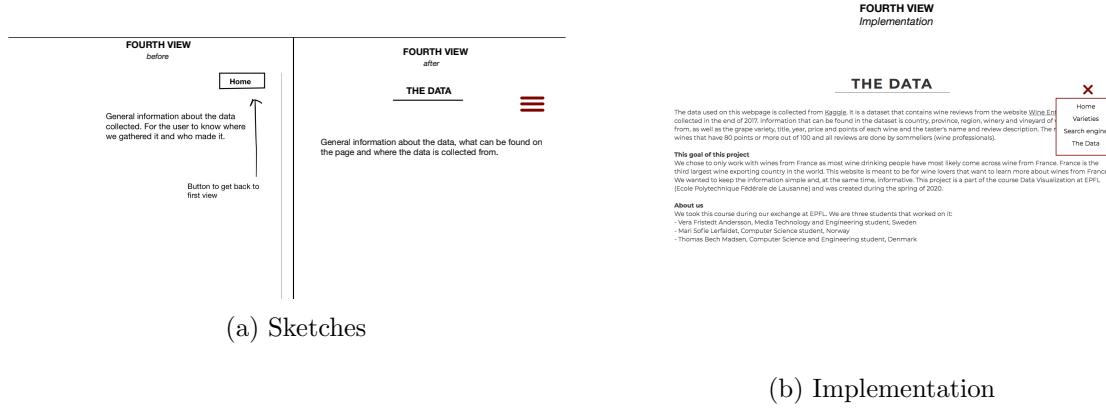


Figure 4: Sketches and implementation for the data

## 7 Peer Assessment

Below follows a description on what each group member has worked on during the scope of the project.

Mari: Data cleaning, set up of the webpage and its formatting/styling, creation of barplot, cleaning and structuring of code/repository, page navigation, creation of sketches, handling dropdown menus and report writing.

Vera: Data exploration and cleaning, map and province interactions, home page province info, contents of the dropdown menus, created the search queries and its outputs, loading icon, info on the data page, css styling and report writing.

Thomas: Data exploration, database setup, TopoJSON incorporation, barplot interactions and linking to map, varieties table, code refactoring and report writing.

## References

- [1] AtelierCartographie. *Khartis*. URL: <https://github.com/AtelierCartographie/Khartis>. (accessed: 27.05.2020).
- [2] Wine Enthusiast. *Wine Enthusiast*. URL: <https://www.winemag.com/>. (accessed: 27.05.2020).
- [3] Bootstrap Temple. *Bootstrap Temple*. URL: <https://bootstraptemple.com/p/black-and-white>. (accessed: 28.05.2020).
- [4] Zackthoutt. *Wine Reviews*. URL: <https://www.kaggle.com/zynicide/wine-reviews>. (accessed: 28.05.2020).