PROCESS BOOK

BY AHMAD ELROUBY,
CARLO REFICE, LUDOVICA SCHAERF

PROCESS BOOK

THE JOURNEY

Our original idea came while browsing Kaggle for datasets. Upon finding the chocolate datasets, and looking at the information within, we immediately came up with the idea for the "World of Chocolate" visualization, visually highlighting the relationships between chocolate producers and harvesting.

Some more research led us to the dataset's original source, the website Flavors of Cacao, which had a more complete dataset and valuable explanations of the various features. This, plus inspiration from previous years' projects, gave us the idea to focus the site on pairing interesting visual information with a textual explanation of the process of chocolate making and the subtleties involved.

We first came up with an overall list of visualizations we'd like to implement and a very basic sketch of what they would look like. We then started working on the two main sections (the chocolate globe and the favorite chocolate finder) in parallel, pre-processing the dataset as needed based on the visualizations' requirements.

After the second intermediate deadline was completed, with the basic infrastructure for the visualizations complete, we decided to focus on creating a consistent visual language and style for the project. We settled on a "Charlie and the Chocolate Factory" theme, which we found fun and familiar enough for users to make them immediately interested in reading on.

We then focused on improving the visualizations with more features and explanations for how to interact with them. The globe visualization was augmented with more information on chocolate ratings by country, and the chocolate finder was accompanied by an analysis of how the selected characteristics of the chocolate influence its ratings.

Finally, we revamped the website layout to follow a simple top-to-bottom flow, with visualizations followed by in-depth explanations.

CHALLENGES AND DECISIONS

The chocolate factory

The logo

The logo was, without a doubt, the most time consuming and least worthy part of the project. It was made using Inkscape to be later animated. Making the SVG with all the different layers and components, plus including the colour effects on the title and the pieces of chocolate took so much time that we were worn out before starting to animate it. Since this seemed to be also taking quite some time, we gave up on the endeavour and only left the logo without animation.

The styling

The styling of the website was kept quite simple. After a few efforts to make something more complex, we realised it did not come together on the website, so we decided on a palette of chocolate colours (a darker and a lighter shade) and used it for all texts and visualizations. We decided also on two fonts that seemed to work well together: the Willie Wonka font found online and converted to a reusable font using font-face and the classic Helvetica Arial. We used bootstrap for subdividing the space and the rest was all done in CSS. We implemented a first version of the website with scrolling laterally and vertically but we then reduced it to only a vertical scroll for clarity and simplicity.

The narrative

The flavours of cacao website from which we acquired the data had a notable amount of information that we wanted to exploit. The text is rather long and very serious, therefore it was quite hard to summarise and make engaging. We decided to give a storyline to the text so that every visualisation would be accompanied by some curiosities and also some text to better understand what they could see from each visualization.

A world of chocolate

The implementation of the chocolate globe came together pretty smoothly overall. We started by adapting a D3 example on how to display geographical data on a globe with flying arcs between locations to our data set, then implemented new features on top of that, such as selecting locations on a map with a mouse click, and coloring countries by the average rating of bars based on either production or bean harvesting.

This latter feature gave us a bit of trouble, as the data set used to draw countries had slightly different country names compared to our chocolate data set. This required some manual conversion between the two to fix.

The method we chose to visualize the globe and arcs, using SVG, is not actually three-dimensional, but just relies on 2D projections. This meant that hiding elements on the side of the globe that is not facing the user has to be done manually, by computing their "depth" and manually filtering out from the data set those that exceed the halfway point.

We initially intended to make the visual indicators for chocolate producers and bean harvesters be a chocolate bar or a cocoa bean icon respectively. However, this led to lots of visual clutter and difficulties with selection, so we decided to make the icon simple circles with different colors, and we added an explanatory legend to the side of the visualization instead.

Finally, after completing the final version of the visualization with all the features, we found that the user was faced with a large amount of visual information with no clear way of determining how to interact with it. We thus decided to add a short textual explanation of what each element represents to the left of the visualization, as well as a small info icon that, when hovered, gives additional tips on how to manipulate it.

Your favourite chocolate finder

The data processing

The data processing for the chocolate finder section was the most problematic. While the ingredients and chocolate percentages were easy to represent (respectively having low cardinality and being numeric), the possible bar flavours were about 1200. We originally thought of reducing the number to around 200 using a word embedding based clustering. However, this did not seem to work well as the word embeddings for flavours such as cherry, strawberry, ... were very reasonably very close to each other, leading to clusters that were reducing the overall possible flavour profile. We, therefore, decided to keep the original flavours.

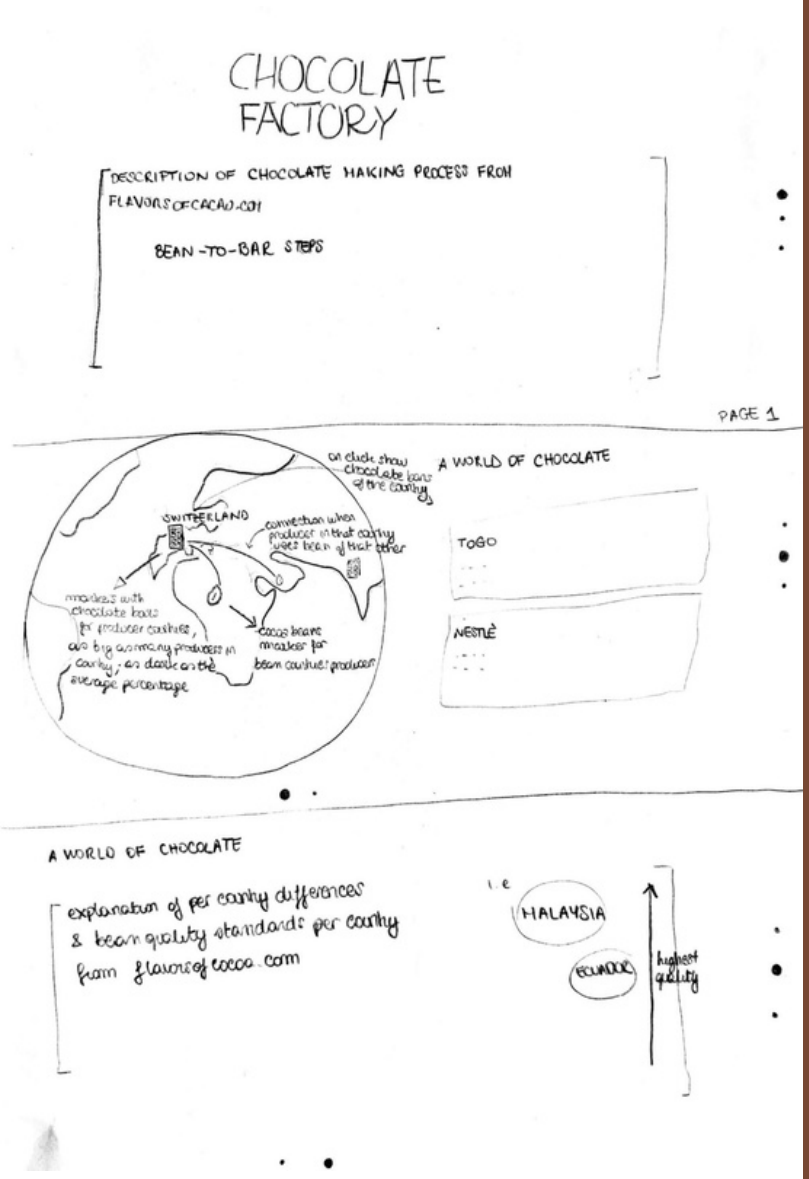
To be able to return a significant amount of results, including therefore also close matches, we decided to adopt, this time, word embeddings. While clustering the fruity flavours into one and using only that one among the choices was influencing badly the results, returning cherry flavoured bars when searching for strawberry is not. Thus, we mapped each chocolate bar to the average vector of its ingredients and each ingredient is also represented by its word embedding (using spacy to retrieve the word embeddings). The ingredients selected by the user are averaged to obtain the overall flavour. Finally, the cosine similarity with the vectors corresponding to the chocolate bars is computed to sort the results.

Library mismatch

To be able to provide the user with a memorable experience, we opted for an easy and intuitive design while picking their favorite flavors and ingredients. Trying to achieve that was somehow difficult since we couldn’t find a single UI library that provides all the components required. One of the libraries that we wanted to use that enabled multiple selections in list format with the capability to search required the usage of bootstrap v4. While that proved to be useful, other components were depending on bootstrap v5 instead. This introduced a discrepancy for the UI components and the layout was unfortunately inconsistent. To mitigate that we had to manually intervene by specifying different styling attributes to make sure the website would look organized and visually appealing.

PROCESS IN SKETCHES

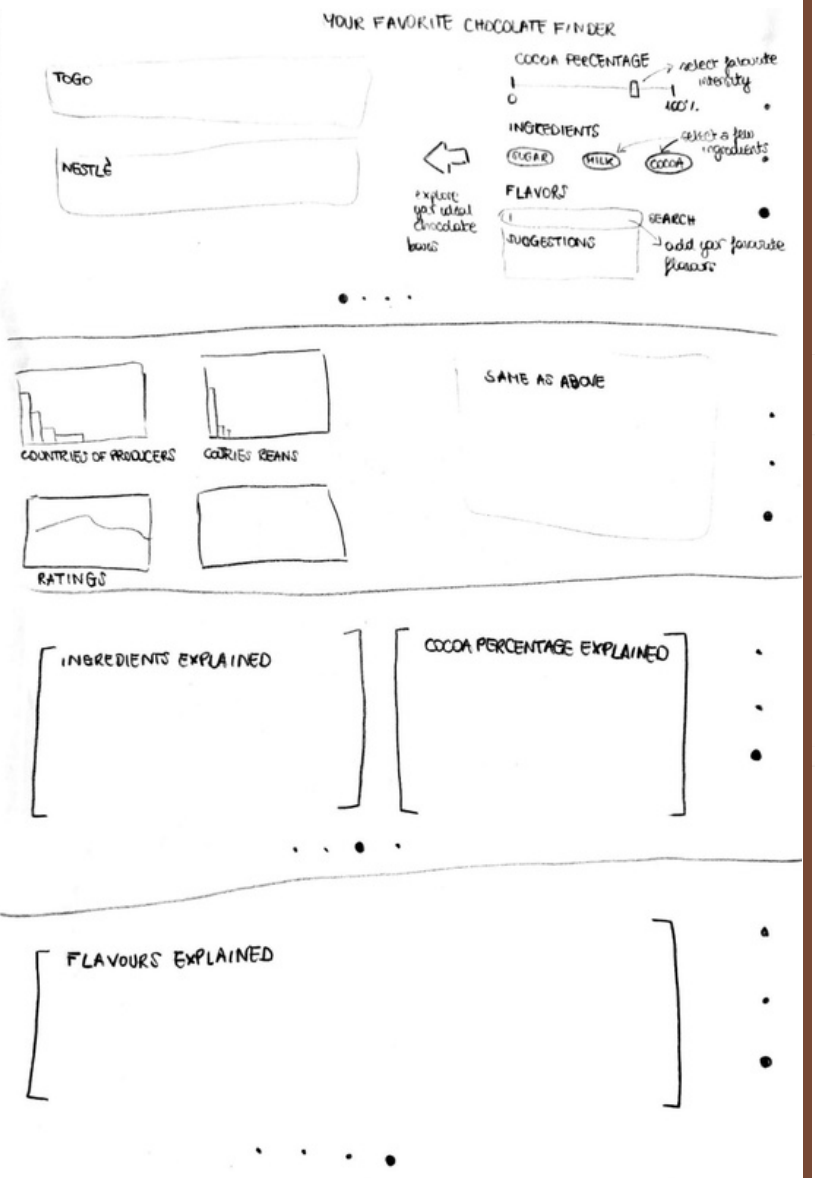
Ideas



As one can see, we remained quite true to our original plan. The main elements and layout remained the same. We changed the title into a chocolate icon, the font to Willie Wonka style and we kept the narrative simple, without the visual effects we anticipated.

Realization

Ideas



Realization

Even for the second part, we kept the same original idea. We condensed the final text into one page and made the four visualizations of the second page as being consecutive instead of grids.

PEER ASSESSMENT

AHMAD ELROUBY

- "Chocolate Finder" section
  - designed and implemented the finder using bootstrap components
  - used processed data and their embeddings to pick the chocolates closest to the user's choices
  - aggregated the different impacting factors for the choice of the chocolate bar
  - utilized leafletJS to map out results according to the country of origin
  - utilized plots/graphs (implemented by Ludovica) to display analysis for the choices made by the user.

CARLO REFICE

- "A world of chocolate" visualization
  - designed and implemented the entire visualization using D3's geographical data modules
  - used preprocessed data (handled by Ludovica) to determine and display geographical locations of chocolate producers, harversters, and the links between them.
  - implemented interactivity using D3 click, drag and hover events
  - helped refine the layout of the website using CSS.

LUDOVICA SCHAERF

- Data analysis and data processing for all parts of the website
- Prepared some of the visualizations of the "Chocolate Finder"
- "The chocolate factory"
  - made the logo
  - set up the styling of the website
  - created the narrative