COM-480 Data Visualization

Spring 2024

Coffee BEaN

PROCESS BOOK

A journey to the rich world of coffee beans and people's coffee habits

Created by

Burcu Özer

Elif Kurtay

Nikhen Sanjaya Nyo

WHY COFFEE?

As three dedicated programmers, we often found ourselves fueled by countless cups of coffee while coding during midnight. Those late-night coding sessions, filled with the rich aroma of our favorite brews, sparked a deeper appreciation for the beverage that kept us going. We started to wonder about the diverse world of coffee, its widespread popularity, and its special significance for people like us who rely on it to stay focused and inspired.

Driven by our passion for both programming and coffee, we were thrilled to explore the rich and varied world of global coffees, investigating the factors behind its immense popularity and its importance for individuals like us. This led us to embark on a journey to create a platform for coffee lovers. Our aim was to uncover a wealth of information about coffee and its drinkers, helping enthusiasts find their ideal coffee variety based on taste and aroma preferences.

WHAT HAVE WE EXPLORED?

As we searched through the Internet, we discovered a wealth of data that would help us to create our platform. Sourcing our datasets from Kaggle, we found four key resources that formed the backbone of our visualizations.

Our journey began with the Coffee Quality Data (CQI May-2023), obtained from the Coffee Quality Institute. This dataset was designed to enhance the understanding and appreciation of coffee quality worldwide. It provided detailed sensory evaluations and quality scores, giving us a rich foundation to explore the origins and unique characteristics of different coffee types. It also involves a score named “total cup points” to quantify the quality of coffee beans based on 10 sensory parameters: Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup (lackness of defects), Sweetness and Overall Impression. A variety that has a score higher than 80 is considered as "Specialty Coffee".

Next, we came across Coffee Tastes & Survey Data, an extensive collection of insights into the habits and preferences of coffee drinkers. This dataset was a bit of a challenge at first, but after some meticulous preprocessing, it revealed fascinating details about who drinks coffee, why they drink it, and their general preferences. We focused on key aspects such as age, gender, consumption habits, and favorite coffee drinks, which added a rich context to our understanding of coffee's universal appeal.

Lastly, to discover the unique relationship between coders and coffee, we found the Coffee Coding and Data dataset. This dataset was particularly exciting because it offered a glimpse into the coffee habits of fellow programmers. It allowed us to visualize how coffee fuels the coding routines of developers around the world in a fun and interactive way.

Together, these datasets provided us with a comprehensive foundation to create engaging visualizations, demonstrating the diverse and captivating world of coffee.

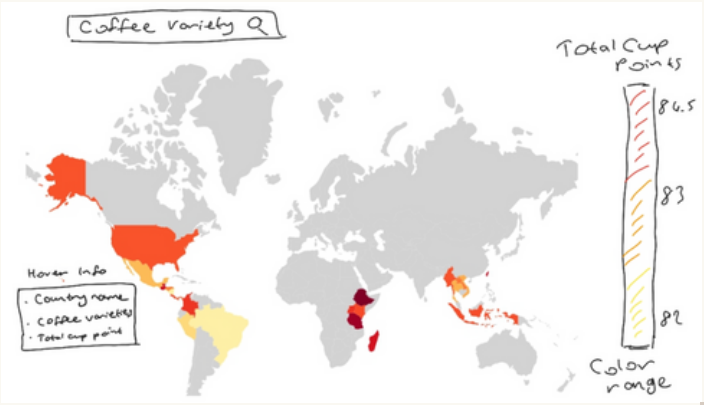
OVERVIEW OF THE VISUALIZATIONS

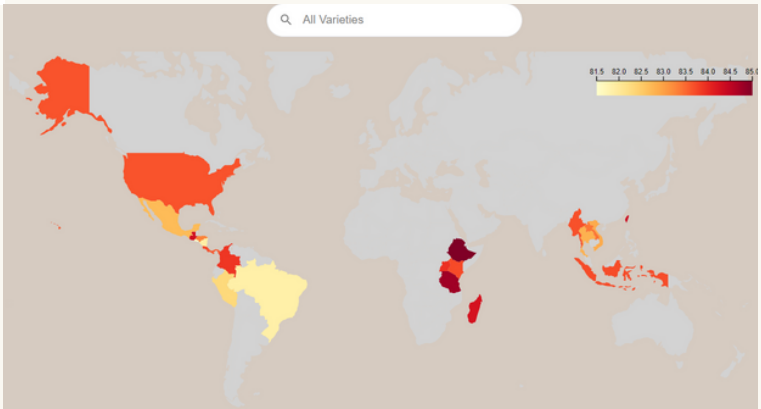
To get the best out of the datasets and showcase them effectively, we identified four core visualizations. Our story began with the coffee itself, where we displayed the origins of different coffee varieties from around the world in our first visualization. The second visualization showed the taste profiles of these various coffee types, allowing users to explore and compare flavors. Next, we delved into the habits of coffee drinkers, providing insights into their daily interactions with coffee through two separate visualizations. One presented the coffee habits of people of all ages and demographics in a pyramid format, illustrating the broad appeal of coffee across different groups. The other was a dedicated section for coders, where we illustrated their coffee habits in a parallel coordinate map, highlighting the unique relationship between programming and coffee consumption.

IMPLEMENTATION

VIZ 1: COFFEE VARIETIES IN A WORLD MAP

To allow people to learn where the different coffee varieties come from, we've implemented an interactive world map feature. All the functionalities in the initial sketch of this visualization were applied with some placement adjustments for aesthetic purposes.



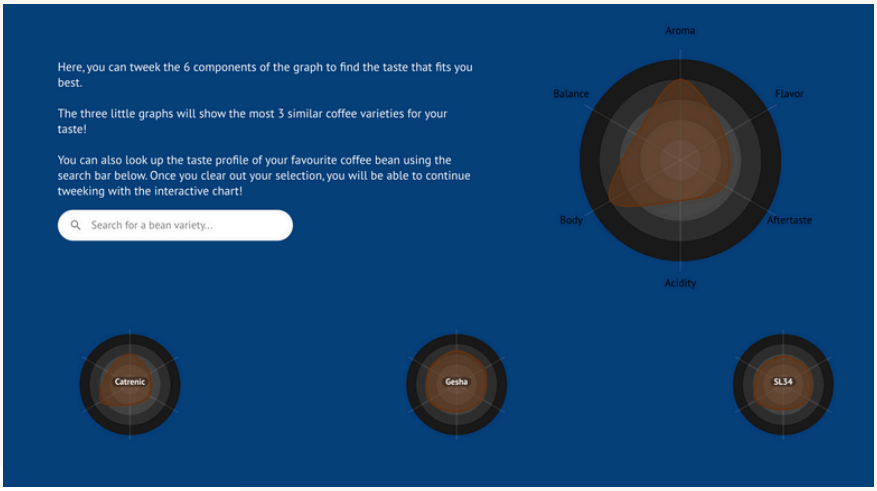


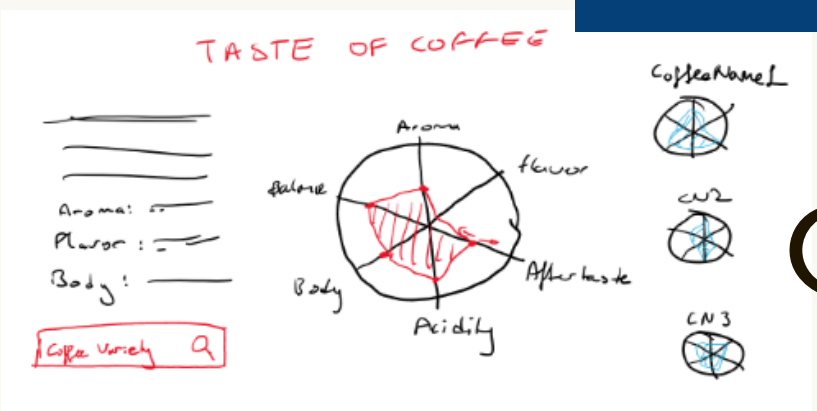
In this section, users can explore coffee-harvesting countries by hovering over them to reveal the different varieties they produce. To implement the map, we used Mercator projection since it makes it easier to distinguish between countries in 2D. Additionally, a search bar allows users to filter the displayed coffee variety, offering the flexibility to focus on specific types. There are some coffee blends that can be selected too. The coloring of countries on the map is based on the quality of the variety selected, represented by "total cup points," with a corresponding legend displaying the color range. This interactive visualization provides users with an engaging and informative way to explore the origins and qualities of various coffee varieties globally.

VIZ 2: COFFEE TASTES IN A RADAR CHART

In this visualization, we've implemented four radar charts to display the taste profiles of coffee. Among these charts, one is interactive, allowing users to manipulate the taste profile by adjusting values directly on the chart. The remaining three radar charts are static and display the taste profiles of coffee varieties closely resembling that of the interactive chart. Therefore, any modifications made to the interactive chart will also reflect in the profiles shown in the other three. Additionally, we've incorporated a search bar, enabling users to look up specific coffee varieties and view their taste profiles directly on the interactive chart. This setup offers users an interactive and intuitive way to explore and compare the taste characteristics of different coffee varieties.

In Milestone 2, we used a different framework to implement the radar charts. However, that framework did not have proper documentation so we had a lot of trouble trying to achieve the interactibility and styles we wanted. At the end, we decided to move to D3.js and changed all the code. Our model D3 code was based on Nadieh Bremer's implementation.



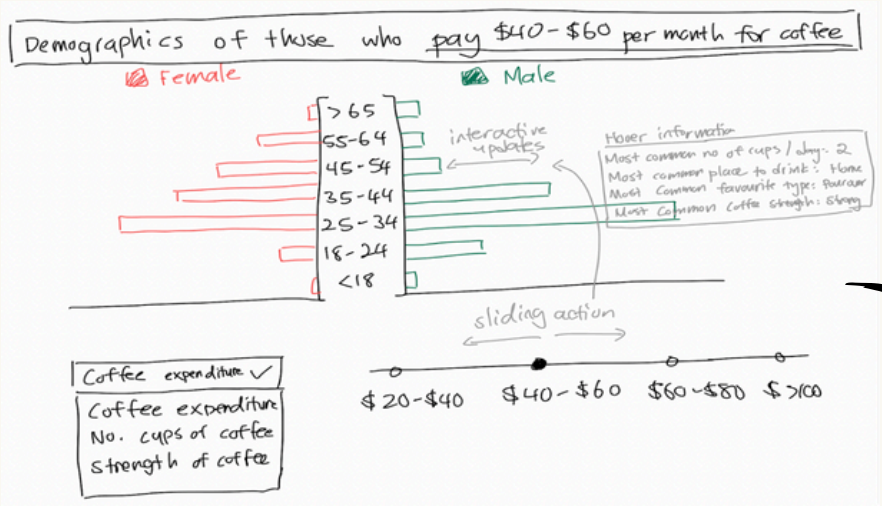


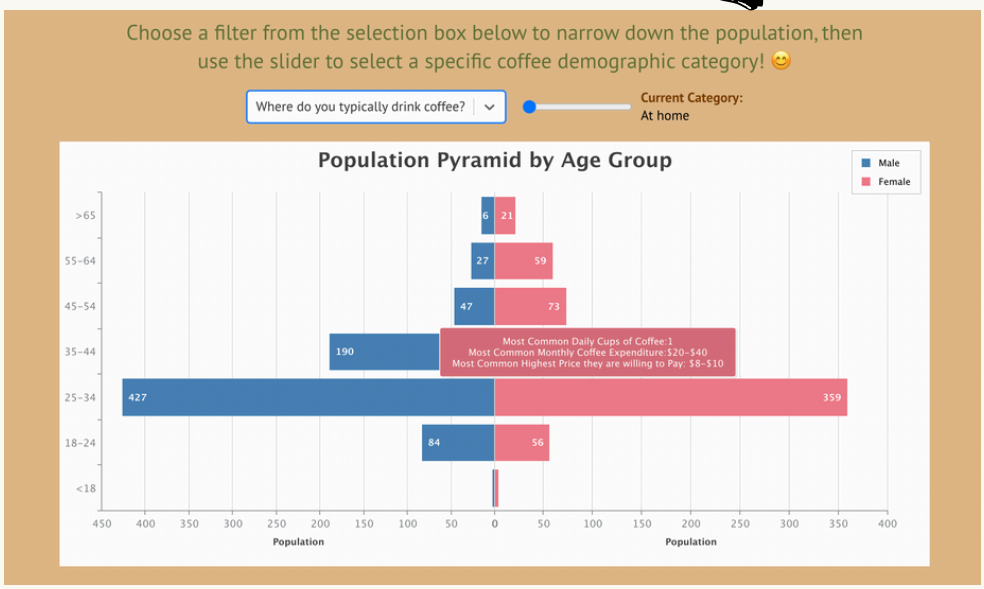
VIZ 3: COFFEE DRINKING HABITS BY DEMOGRAPHIC IN A POPULATION PYRAMID

We've added an interactive Population Pyramid chart to explore coffee drinkers' habits. The chart includes a population filter and an interactive slider to display categories like types of sweeteners used, reasons for drinking coffee, and where coffee is purchased and consumed.

When users select a filter and category with the slider, the chart updates to show relevant statistics for the chosen population. This includes hover information on daily cups of coffee, monthly coffee expenditure, and the highest price they are willing to pay.

We included the main components from Milestone 2 but updated the user-selectable filters. Now, users can view coffee expenditure, daily cups of coffee, and the highest price an age group is willing to pay through hover information. They can also select interesting habits like where people drink or buy their coffee.

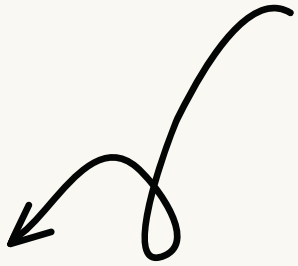


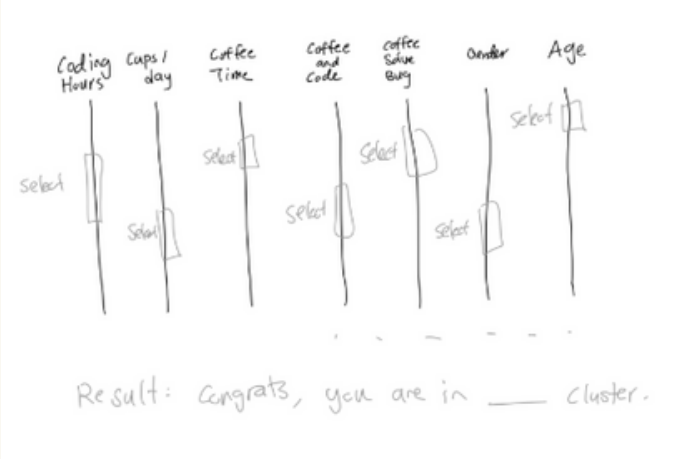


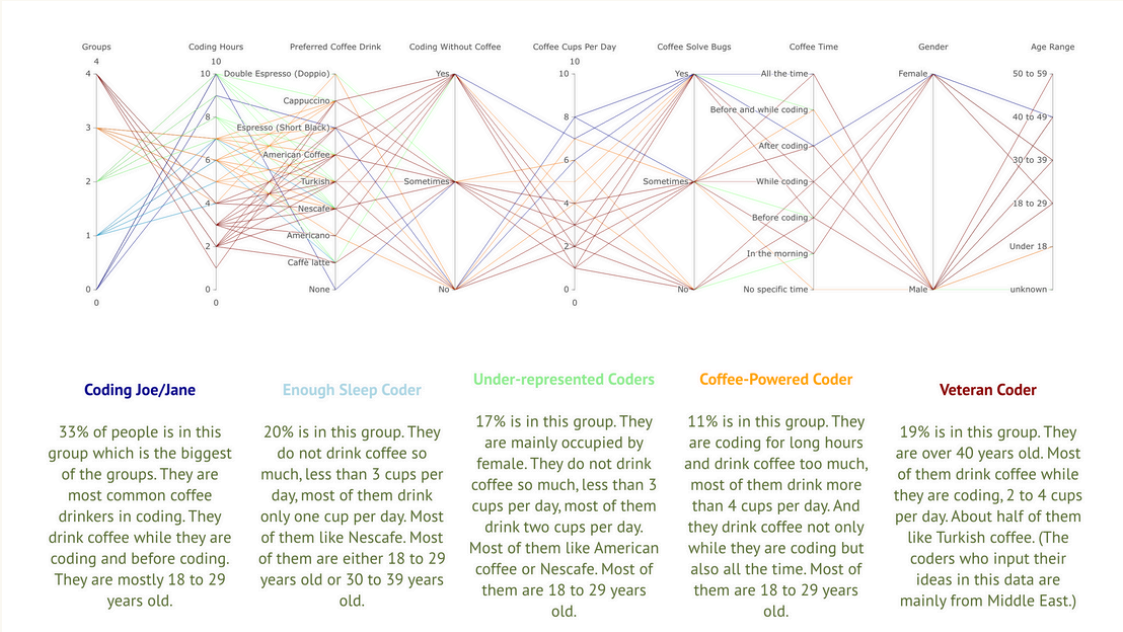
VIZ 4: COFFEE CHOICE & HABITS OF CODERS IN PARALLEL COORDINATE GRAPH

This final section is dedicated to coders and their coffee habits. It consists of a dynamic parallel coordinate graph illustrating each coder's demographic information, coffee consumption, and beliefs according to several characteristics. The graph is interactive, enabling coders to choose ranges and learn about different cluster types, such as "Coding Joe/Jane", "Enough Sleep Coder", "Under-represented Coders", "Coffee-Powered Coder", and "Veteran Coder". These clusters are groups of people who have similar habits with coffee and they were identified through K-means clustering in our exploratory data analysis.

Our initial plan was to find the cluster of the user when they selected ranges in each column in the graph. However, we later thought that this feature was not related to visualizing the dataset and it would need us to attach an ML model to our website. Therefore, we decided to not continue with this approach but simply let people see all the cluster, and cluster information. They can still select their answers for different characteristics and see which cluster they are more close to. They can also observe their differences to other clusters in more detail.







Coding Joe/Jane

33% of people is in this group which is the biggest of the groups. They are most common coffee drinkers in coding. They drink coffee while they are coding and before coding. They are mostly 18 to 29 years old.

Enough Sleep Coder

20% is in this group. They do not drink coffee so much, less than 3 cups per day, most of them drink only one cup per day. Most of them like Nescafe. Most of them are either 18 to 29 years old or 30 to 39 years old.

Under-represented Coders

17% is in this group. They are mainly occupied by female. They do not drink coffee so much, less than 3 cups per day, most of them drink two cups per day. Most of them like American coffee or Nescafe. Most of them are 18 to 29 years old.

Coffee-Powered Coder

11% is in this group. They are coding for long hours and drink coffee too much, most of them drink more than 4 cups per day. And they drink coffee not only while they are coding but also all the time. Most of them are 18 to 29 years old.

Veteran Coder

19% is in this group. They are over 40 years old. Most of them drink coffee while they are coding, 2 to 4 cups per day. About half of them like Turkish coffee. (The coders who input their ideas in this data are mainly from Middle East.)

TOOLS USED

Python	Used during data analysis, as well as transforming the public datasets into the format we needed for the visualizations. We mainly used their “pandas” library.
HTML, CSS, JavaScript	Used for the visual structure, interactivity, and layout of the website.
React	React is the foundation of our website. It provides ways to structure code into smaller components while keeping track of state information for each component. Thanks to React, it was easy to divide the work between us as we could each work on different components without interfering with each other's work. Some of our members were familiar with React which decreased our time for setting up the basics for the website.
D3.js, Plotly	These frameworks were new for all of us. We discovered their tools to transform our data into beautiful and interactive visualizations.
Github Pages	We used Github to host our website. This way we did not have to work to create a production build and find a source to host our web server.

CHALLENGES

Finding a Dataset

One of the primary challenges we faced was finding a comprehensive dataset. The Coffee Quality Data provided insights into coffee origins and characteristics but lacked consumer preferences and habits. Conversely, the Coffee Tastes & Survey Data offered demographic insights but missed specific taste profiles and correlations with coding habits. Thus, we combined multiple datasets to cover all aspects of our project.

Styling Issues

Since all the visualizations had different components, styling the website required careful considerations to ensure aesthetics and prevent any overlaps between the components. Therefore, we needed to apply frequent adjustments to CSS properties such as positioning, margins, and padding to achieve the desired aesthetics while preserving functionality.

Data Preparation

During our data preparation, we found that the Coffee Survey Dataset had significantly more male data points, skewing the graphs and making it difficult to draw accurate conclusions. To balance the representation, we randomly selected a subset of male data points equal to the number of female data points, making it easier to draw insights from the data which would no longer be heavily skewed.

Version and Integration Issues

We had to rely on some public frameworks in order to achieve interactive visualizations. This caused some version and integration issues during development. For example, the interactive radar chart code was published in D3.js by Nadieh Bremer in 2015. Hence, while trying to incorporate it into our repository, we came across many depracated tools that needed to get updated according to the rules of newer version. Similarly, as we used different JavaScript tools, it was also a challenge to integrate everything into the React workspace.

PEER ASSESSMENT

BURCU ÖZER

I worked on the exploratory data analysis and data preparation of the Coffee Quality dataset. Then, I worked on the visualization of coffee bean harvesting countries on world map using a search bar and creating a color bar.

ELIF KURTAY

I worked on the data preprocessing of Coffee Quality, Coders, and Cocktails datasets. Then, I created a template website for all of us to work on with React. Afterwards, I mainly worked on visualizing the taste profiles of coffee beans with the radar charts and the search bar. Additionally, I also helped wrap up the implementation of the parallel coordinates graph.

NIKHEN SANJAYA NYO

I worked on preparing the data for the Coffee Survey dataset, to visualize the habits of coffee drinkers through the population pyramid. Additionally, I worked on the initial implementation of the habits of coders in parallel coordinates graph on Milestone 2.

Click here to see our website

Click here to see our code