INF552 Data Visualization: Project Report

Mengfei Gao

December 2023

1 Introduction

Coffee, an integral part of everyday routines and a key player in the world's economy, embodies a complex journey from bean to brew, encompassing extensive networks of production, distribution, and consumption. In our current era, where data shapes decisions, the significance of data visualization emerges as a crucial tool for decoding the intricacies of the coffee sector. This report introduces the Coffee Quality Visualization project, which leverages the capabilities of D3.js to transform detailed coffee production data into visually striking and informative displays, providing a unique perspective on the global coffee landscape.

2 Dataset and Motivation

2.1 Dataset Overview

The data for this project comes from the Coffee Quality Institute (CQI), a non-profit entity committed to enhancing the quality and worth of coffee globally. CQI engages in a range of activities including research, training, and certification, all aimed at raising the bar for coffee quality, fostering sustainable practices, and supporting the growth of the specialty coffee sector. Their online database serves as a treasure trove of information for those involved in the coffee industry or simply passionate about it. It offers detailed insights into various aspects of coffee production and processing, as well as sensory evaluations, delving into elements such as the impact of genetics and soil on coffee quality.

This dataset offers a comprehensive snapshot of the coffee industry across numerous countries, painting a vivid picture of the global coffee scene. It includes information on:

- Production by country or region.
- Total cup points, a measure of coffee quality.
- The geographic spread of coffee-growing areas.
- Detailed sensory evaluations of coffee, covering characteristics like aroma, flavor, aftertaste, acidity, body, balance, uniformity, clean cup, and overall sweetness.

2.2 Motivation

The motivation behind the Coffee Quality Visualization project stems from the increasing significance of coffee in the worldwide market. Recognized as both a key economic asset and a beverage enjoyed by many, coffee's impact is extensive. This initiative aims to:

- Deepen insights into the distribution of coffee quality across various regions.
- Aid key players in the coffee sector, encompassing producers, exporters, and consumers, by offering a resource for enlightened decision-making.

The visualization tools created for this project transcend traditional data presentation methods. They are designed to be both interactive and captivating, enabling users to delve into and comprehend the diverse aspects that influence coffee production and its quality.

3 Visualization

This project is made with D3(short for Data-Driven Documents) . This JavaScript library is renowned for its ability to craft dynamic, interactive visualizations directly in web browsers. D3 leverages a variety of web standards, including Scalable Vector Graphics (SVG), HTML5, and Cascading Style Sheets (CSS), to bring data to life in a visually appealing and interactive manner.

What sets D3.js apart is its comprehensive suite of functionalities. It allows for meticulous selection and manipulation of document elements, crafting of SVG objects, and application of styles. Transitions and dynamic effects are seamlessly integrated, adding a layer of sophistication to the visualizations. This is particularly advantageous when dealing with extensive datasets, as D3.js can efficiently bind these large volumes of data to SVG objects, facilitating the creation of diverse charts and diagrams that are both informative and visually striking.

We use GeoJSON for the map-creating. GeoJSON is a widely accepted standard for geospatial data interchange. This format is adept at representing simple geographical features, along with their associated non-spatial attributes. Its versatility is evident in its support for various geometries like points, line strings, polygons, and their multipart collections. By incorporating GeoJSON, the project gains an enhanced capability to represent geographical data in a manner that is both accurate and easily interpretable, further enriching our visualizations with geographical context and precision.

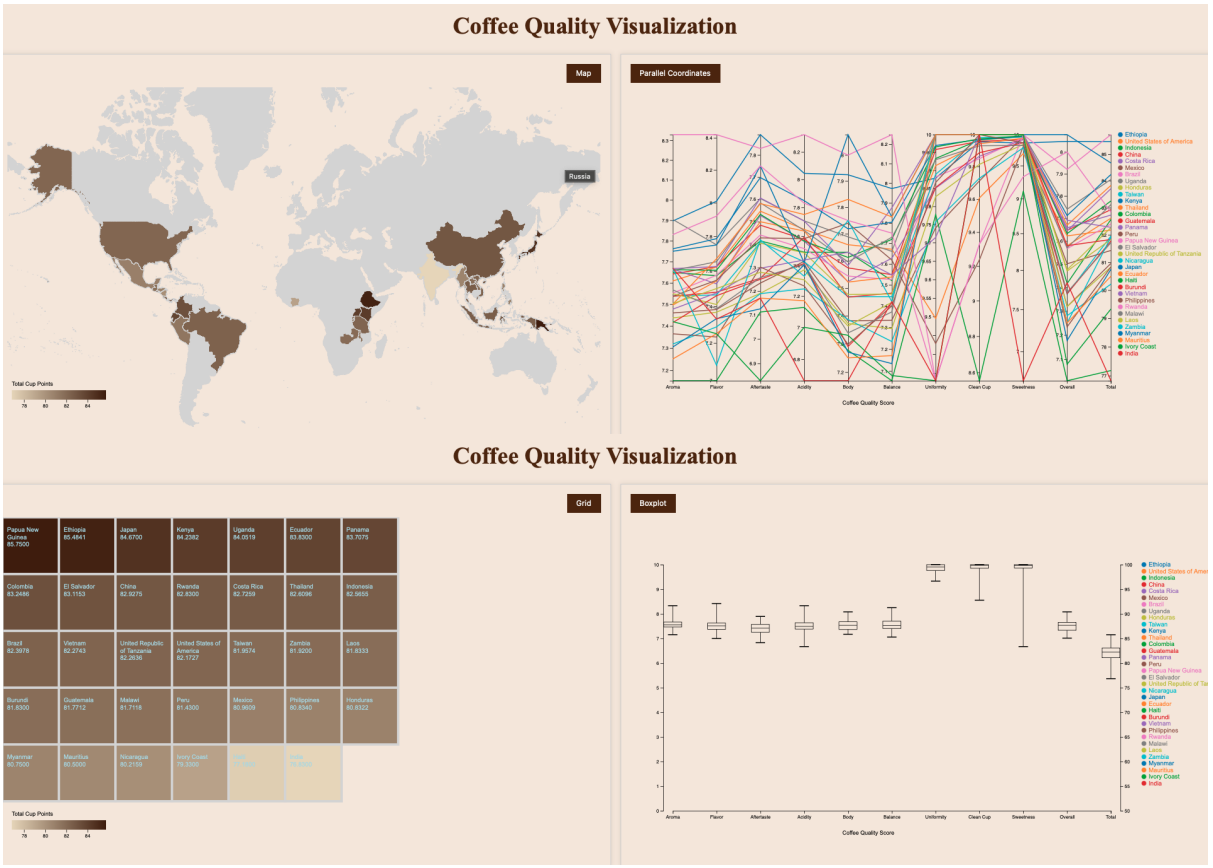

Figure 1: Coffee Quality Visualization includes the World Map, Grid Layout, Parallel Coordinates Plot, and Box Plot.

3.1 Key Features

The project features several interactive visualizations, each serving a specific purpose:

3.1.1 World Map Visualization

The World Map serves as a cornerstone in our visualization, leveraging the innate human ability to understand and interpret geographical data. To construct this map, the project utilizes a GeoJSON file, which accurately represents the geographical boundaries of each country involved in coffee production.

The choice of color encoding for the map is both aesthetically pleasing and functionally informative. It employs a gradient transition from a light beige to a deep, dark coffee hue, symbolically aligning with the range of coffee beans' colors. This color scheme is meticulously selected to depict the average Total Cup Points for each country, where the intensity of the color directly correlates with the quality of coffee produced. Lighter shades indicate lower average quality, while darker tones represent higher quality beans. This use of color not only creates a visually striking map but also serves a crucial analytical purpose. It allows viewers to quickly grasp the quality distribution of coffee across the globe, making it

easy to identify regions that excel in coffee production. The map can provide an immediate visual summary of global coffee quality at a glance.

3.1.2 Grid Map Layout

In our visualization, the grid layout adopts a straightforward yet effective approach to presenting coffee quality data. Countries are organized in a left-to-right, top-to-bottom sequence based on their Total Cup Points. This method aligns each nation within a uniform grid, simplifying the comparison of coffee quality across different countries. Unlike traditional map-based representations, this grid format strips away geographical biases and focuses solely on the ranking of coffee quality. Each country, represented by a cell in the grid, is positioned according to its quality score, making it easy for viewers to discern the hierarchy in global coffee production quality. This layout offers a clear, linear progression from the highest-scoring countries in the top left to those with lower scores towards the bottom right, thereby providing an intuitive and accessible way to visualize and compare the coffee quality of various nations.

3.1.3 Parallel Coordinates Plot

The Parallel Coordinates Plot in our project is a striking tool for analyzing the average quality scores of coffee from different countries. It presents a multidimensional perspective, where each axis corresponds to a distinct attribute of coffee quality, such as aroma, flavor, body, and acidity. In this plot, every country is depicted by a polyline, with each segment intersecting the axes at points that reflect their national average scores for these attributes. Each country's polyline has a random color. While the colors do not carry specific means, they aid in distinguishing between lines, making it easier for viewers to trace and compare the quality profiles of various countries.

3.1.4 Box Plot

The Box plot in our project is designed to represent country-specific analysis of coffee quality attributes. The detailed representation is pivotal for understanding the distribution and range of coffee qualities within individual countries. Each box in the figure corresponds to a specific country and depicts the statistical distribution of scores for a specific quality attribute. The x-axis is organized by different coffee quality scores. This arrangement allows for a focused examination of how each performs across various quality aspects of coffee.

The box itself illustrates the interquartile range of the scores for that attribute in the country, with a line inside each box denoting the median value. The whiskers extending from the boxes highlight the full spectrum of scores, from the lowest to the highest. This visualization is particularly beneficial for those invested in the coffee industry, including producers, and exporters. It provides a clear and concise representation of where each country stands in terms of different quality attributes of coffee. By comparing the boxes across different

attributes, stakeholders can gain insights into the strengths and weaknesses of coffee production in various regions.

3.2 Interactivity

Our project incorporates brushing and linking techniques, enabling dynamic interaction across different visualizations. When users select a specific country on the World Map or the Grid Layout, this action triggers an automatic highlighting of the corresponding data in the Parallel Coordinates Plot. Additionally, it can also display the Box plot for selected countries, offering a detailed view of coffee quality scores.

This interactivity not only enriches the user experience but also facilitates a more intuitive and comprehensive exploration of the data. For instance, selecting a country on the map immediately brings up its coffee quality profile in the Parallel Coordinates Plot, visually displaying the average scores across various attributes. This referencing allows users to easily compare and contrast different countries, enhancing their understanding of global coffee quality trends.

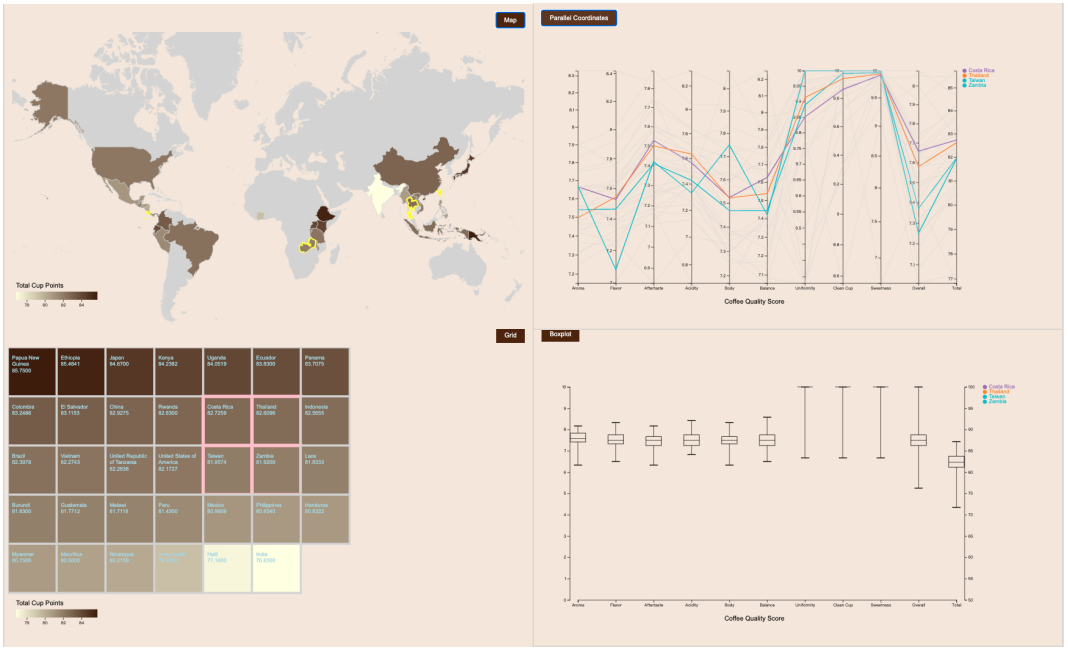

Figure 2: When clicking or using a brush to select, the charts update

4 Conclusion

This project translates complex coffee data into visualizations. By utilizing D3.js, the project has created a suite of interactive and dynamic visual tools that illuminate the world of coffee quality. From the World Map's intuitive geographical representation to the detailed insights offered by the Box Plot, each element of the visualization works in concert to deliver a comprehensive understanding of global coffee quality.