Project Report - Home of Coffee

1. Introduction

With the rise of coffee culture and the continuous increase in coffee consumption, attention to coffee beans is gradually growing. In this context, our team aims to analyze data related to coffee beans to gain insights into the coffee industry and understand market trends. We explore the diversity and quality of coffee from different regions around the world. The project report will provide a comprehensive analysis and visualization of the coffee data, encompassing information on the background, data description and processing, task abstraction, visualization design and principles, visualization results, and conclusion.

2. Data Preprocessing

In this section, we conducted data preprocessing to clean and prepare the coffee dataset for analysis. The dataset contains information on various attributes of coffee, including the country, region, altitude, year, owner, species, variety, processing method, number of bags, bag weight, aroma, flavor, aftertaste, acidity, body, balance, uniformity, sweetness, moisture, total scores, and color.

We began by renaming the columns of the dataset to ensure clarity and consistency. Next, we removed any missing values from the dataset, such as NaN values in the region, owner, variety, processing method, and color columns. We also removed any unreasonably high altitude values, as the suitable altitude for growing coffee is typically around 1000-2000. After these data cleaning steps, we were left with 796 instances in the dataset.

To facilitate visualization and analysis, we converted the qualitative variables to the character type. Additionally, we created a new column called "rating" based on the total scores, which categorized the coffee into five rating levels (1-5) for easy

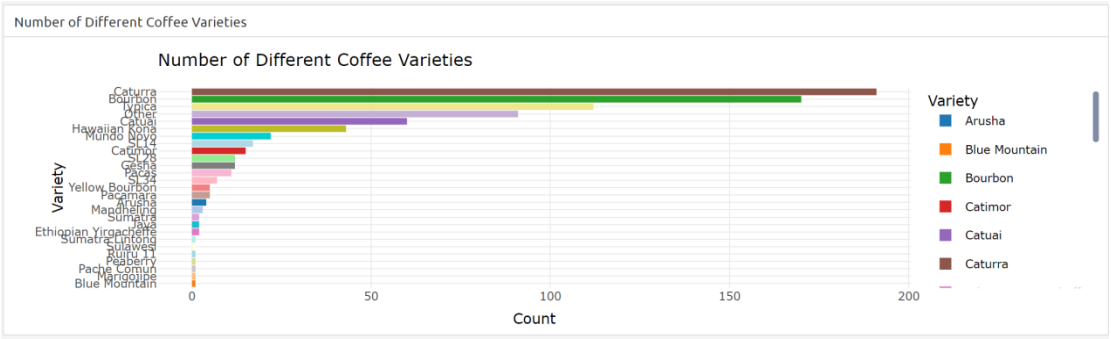
recommendation.

3. Variety Info

In this section, we explored the distribution of different coffee varieties and the coffee output by year and variety.

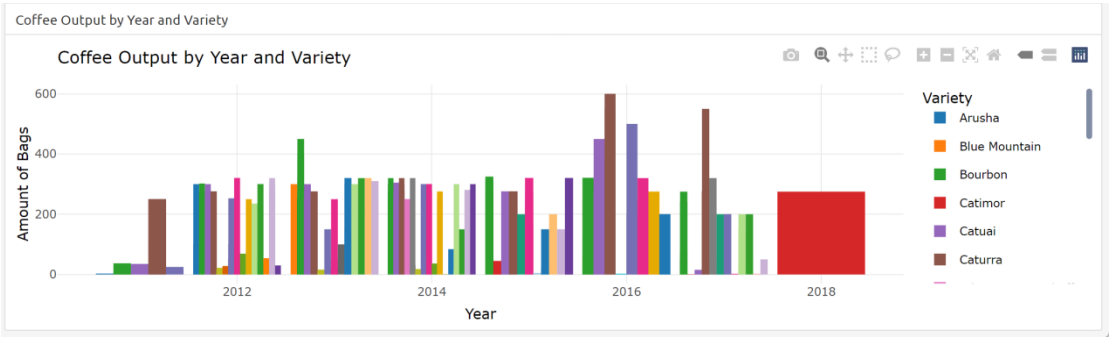
3.1 Number of Different Coffee Varieties

We visualized the number of different coffee varieties using a bar plot. Each bar represents a coffee variety, and the height of the bar represents the count. The colors of the bars are based on the coffee variety.



3.2 Coffee Output by Year and Variety

We created a bar plot to show the coffee output by year and variety. The x-axis represents the year, the y-axis represents the number of bags, and the colors represent different coffee varieties.



Principles:

Data mapping: The process of converting data values into visual elements. Different

coffee varieties and annual yields are mapped into bars of different colors and heights, making it easy to quickly identify and compare the quantities and annual yields of different varieties

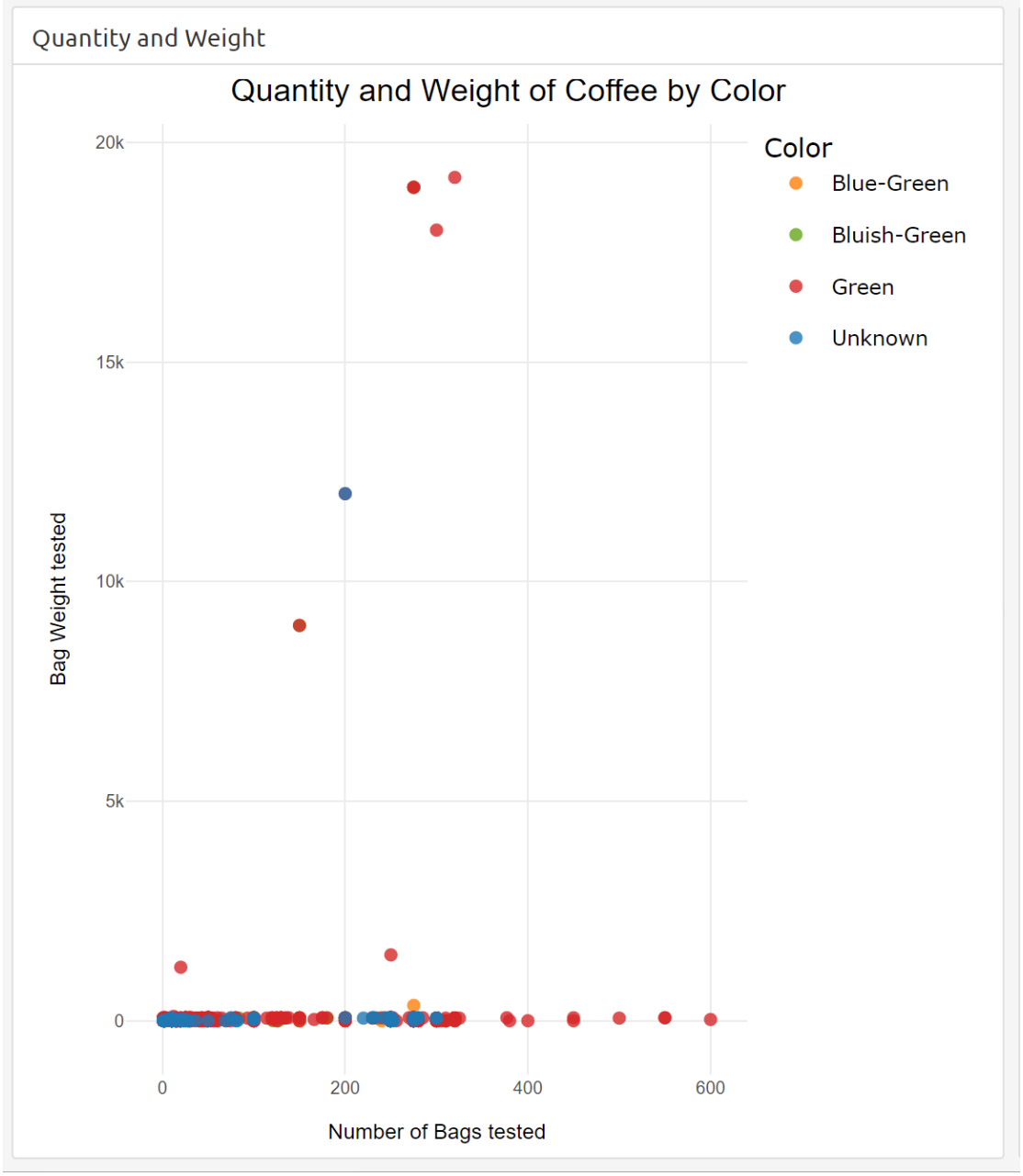
Interactivity: While these charts are static, in an interactive environment, users can be designed to click or hover to get more information

4. Color Info

In this section, we explored the quantity and weight of coffee by color and visualized the color distribution of worldwide coffees.

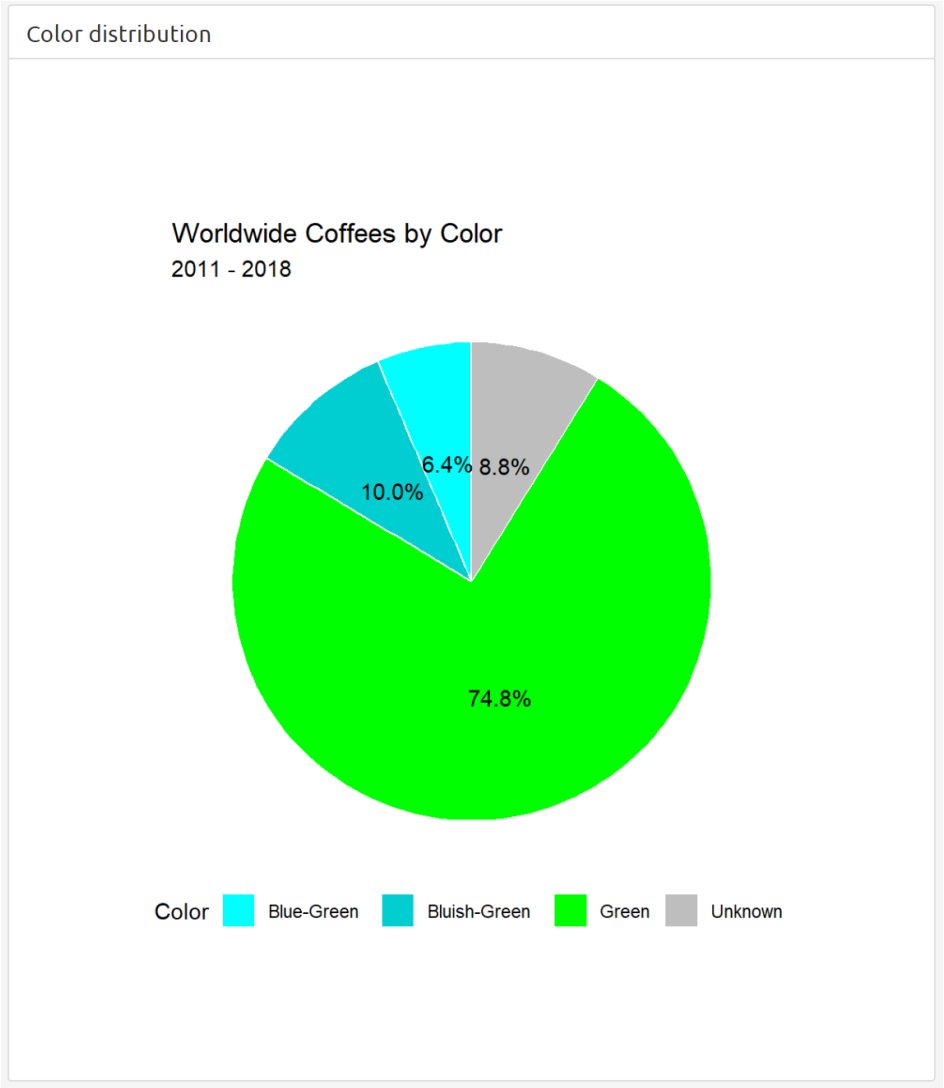
4.1 Quantity and Weight

We plotted the quantity and weight of coffee by color using a scatter plot. The x-axis represents the number of bags tested, the y-axis represents the bag weight tested, and the colors represent different coffee colors. When hovering over data points, it will display specific data of bag weight, color and number of bag. We can see that green coffee beans are relatively high both in quantity and quality, unknown-color coffee beans are all relatively low in this two value.



4.2 Color Distribution

We created a pie chart to show the color distribution of worldwide coffees. Each slice of the pie represents a color, and the size of the slice represents the proportion of coffees with that color. In addition, in the color design of this part, the corresponding real color of the coffee beans is selected, so that we can feel the proportion allocation more intuitively. In different categories of coffee beans, the most prevalent color is green.



5. Continent Info

This section focuses on analyzing coffee data at the continent level, including the number of coffee rating 5 by continent and interactive maps.

Principles:

Geographic Representation: Space information and distribution

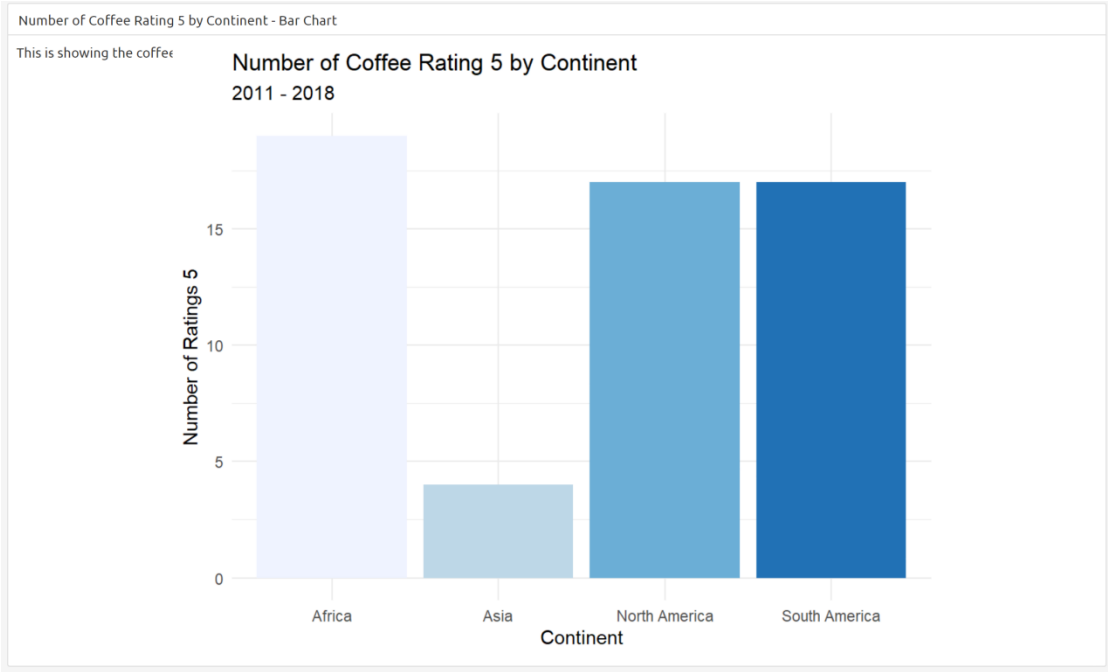
Differentiation on color mapping

Legend Guide

“Color and position are the most separate attributes!”

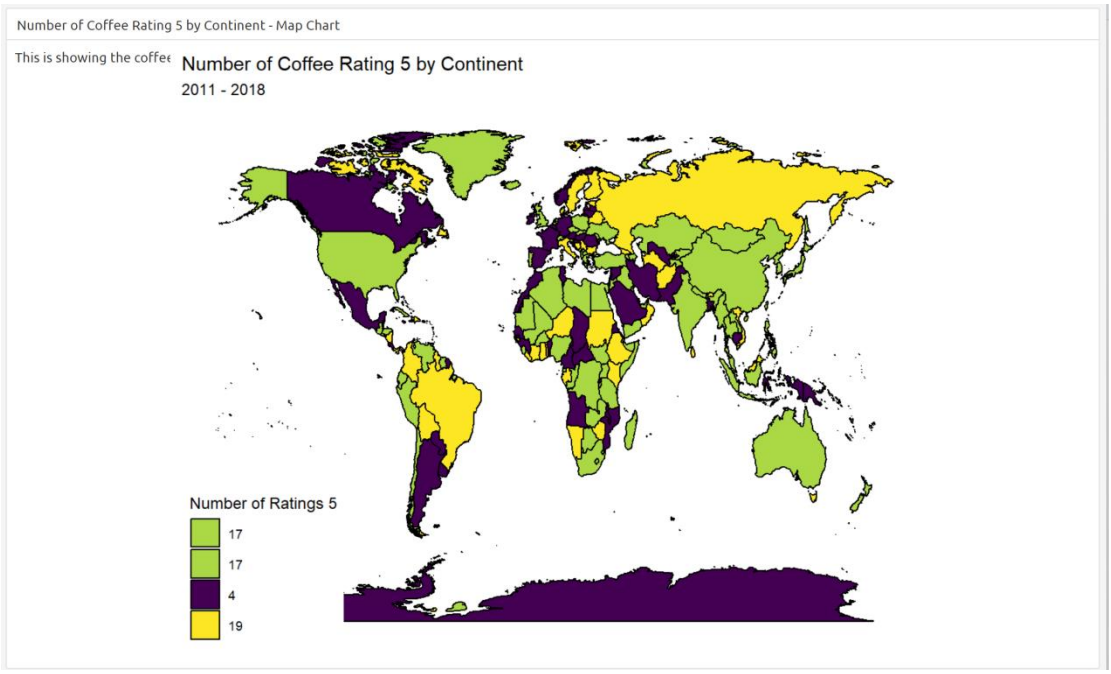
5.1 Number of Coffee Rating 5 by Continent - Bar Chart

We created a bar chart to show the number of coffee rating 5 by continent. Each bar represents a continent, and the height of the bar represents the number of ratings 5.



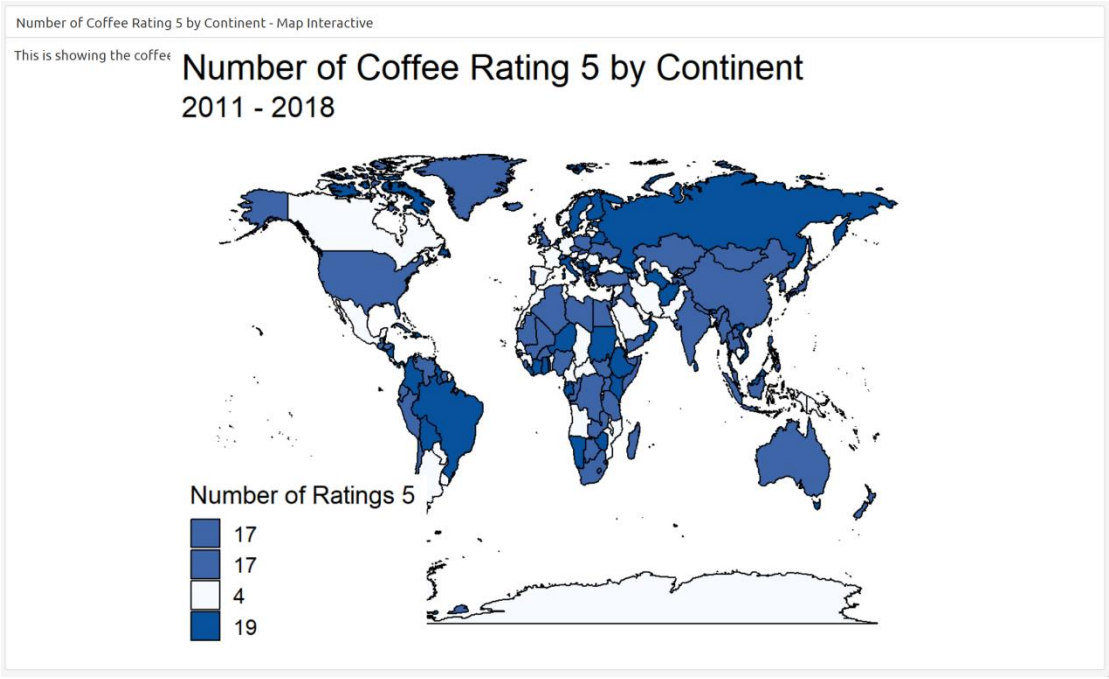
5.2 Number of Coffee Rating 5 by Continent - Map Chart

We visualized the number of coffee rating 5 by continent using a map chart. Each continent is represented by a color, with darker colors representing a higher number of ratings 5.



5.3 Number of Coffee Rating 5 by Continent - Map Interactive

We created an interactive map to show the number of coffee rating 5 by continent. Each continent is represented by a color gradient, with darker colors indicating a higher number of ratings 5.



6. Altitude Info

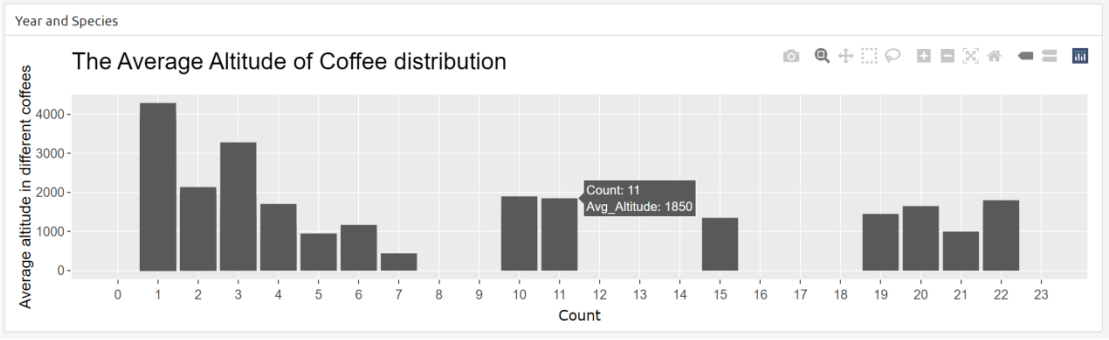
Here, we explored the average altitude of coffee distribution and the relationship between total scores, altitude, and processing method.

6.1 Average Altitude of Coffee Distribution

We analyzed and visualized the average altitude of coffee distribution using a bar chart. The x-axis represents the count, and the y-axis represents the average altitude in different coffees.

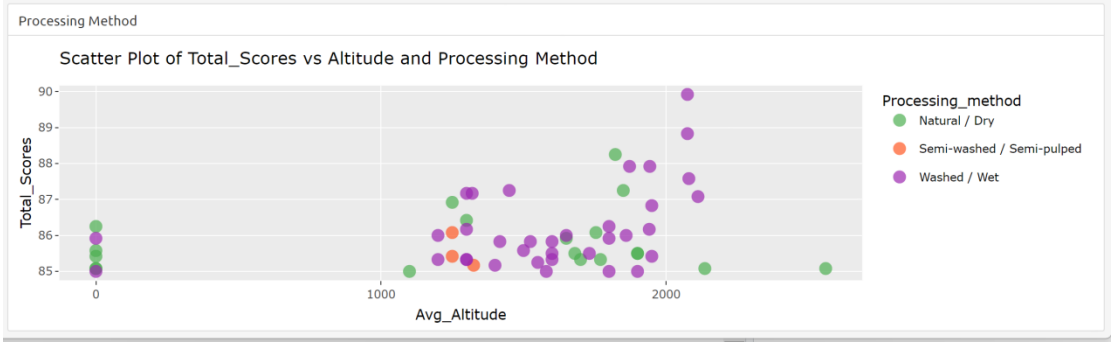
Principle:

In order to avoid the chart being too lengthy, the length of the X-axis is limited to ensure that the chart is clear and easy to read. By looking at the chart, we can conclude that most coffee beans are grown at an altitude of 2000.



6.2 Total Scores vs. Altitude and Processing Method

We examined the relationship between total scores, altitude, and processing method using a scatter plot. The x-axis represents the average altitude, the y-axis represents the total scores, and the colors represent different processing methods.



Principle:

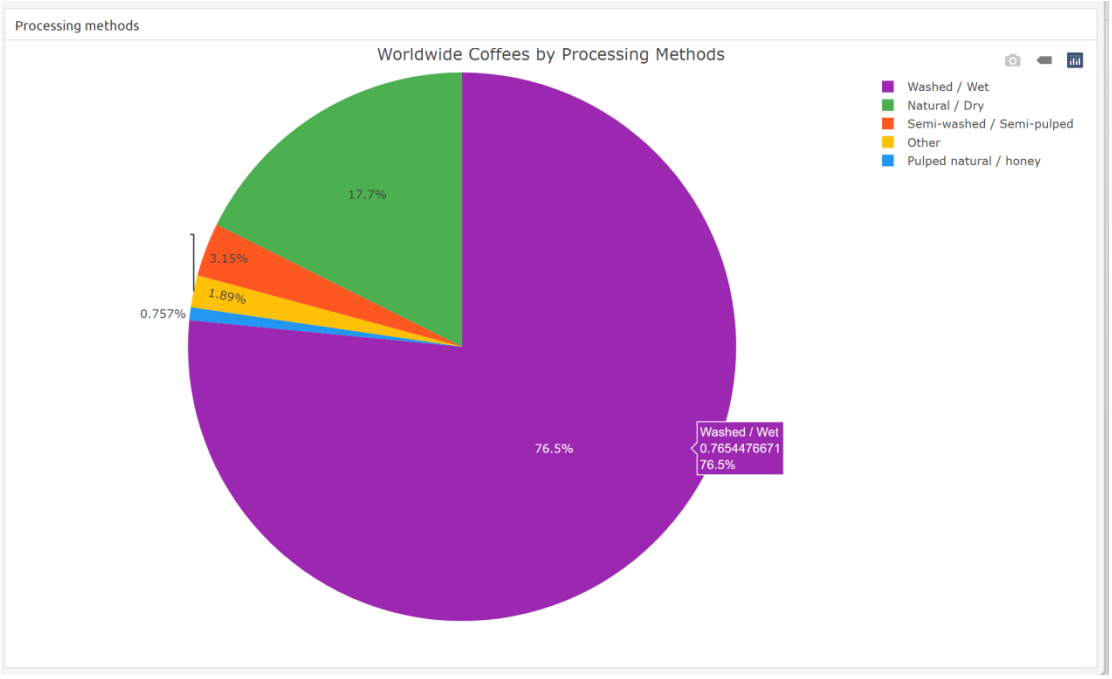
The relationship between coffee bean score and average altitude is shown in the form of scatter plot, and the processing methods are represented by different colors. The horizontal coordinate indicates the average altitude, the vertical coordinate indicates the coffee bean score, and the processing method is distinguished by color. This design allows the observer to intuitively understand the relationship between coffee bean rating, altitude, and handling method.

7. Worldwide Info

This section provides an overview of coffee data worldwide, including processing methods and average coffee scores by country.

7.1 Processing Methods

We visualized the proportion of different processing methods used in coffee production worldwide. The interactive pie chart shows the percentage of each processing method.

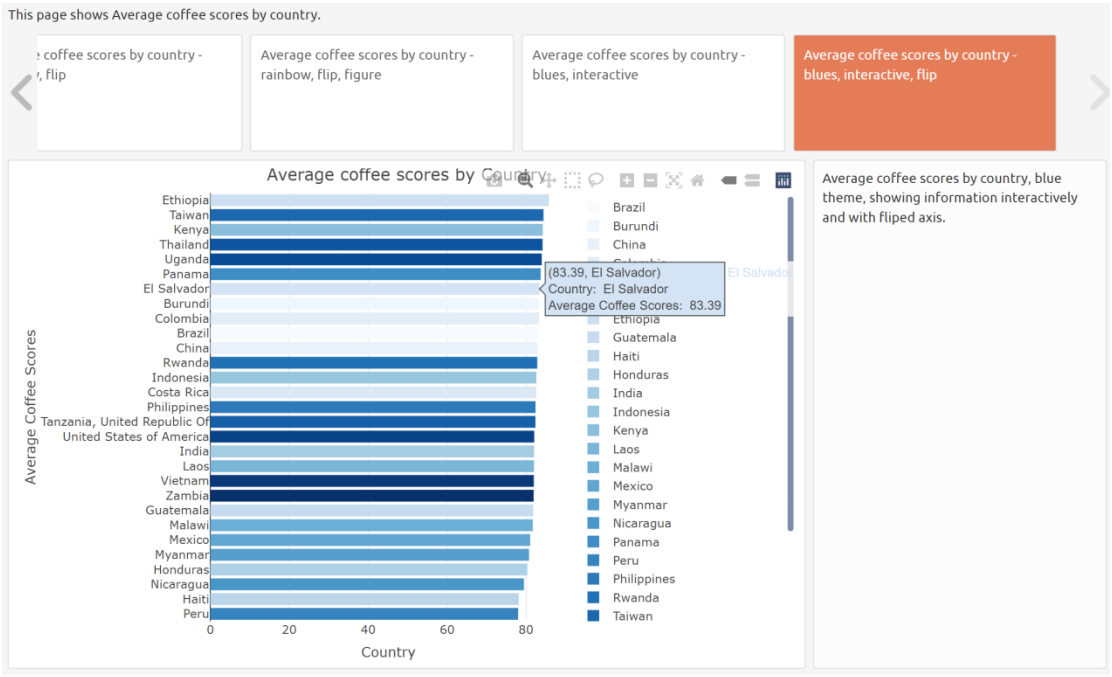


Principle:

Pie charts show the distribution of coffee bean processing methods to visually communicate information and maintain a high data-to-ink ratio. The size of each sector represents the proportion of the corresponding processing method, and the label indicates the type of processing method. The chart is designed to highlight the most common processing methods, such as immersion. In addition, by hovering the pointer over a specific area, the user can obtain detailed information at the micro level. To improve the user experience, an interactive feature has been added that allows users to hide or display the proportion of specific processing methods by clicking on the "Washed" TAB to get a clearer view of the distribution of other processing methods.

7.2 Average Coffee Scores by Country

We analyzed and visualized the average coffee scores by country. We created a bar chart and a map chart to show the average coffee scores by country. The colors represent the average scores, with darker colors indicating higher scores.



8. Ratings Info

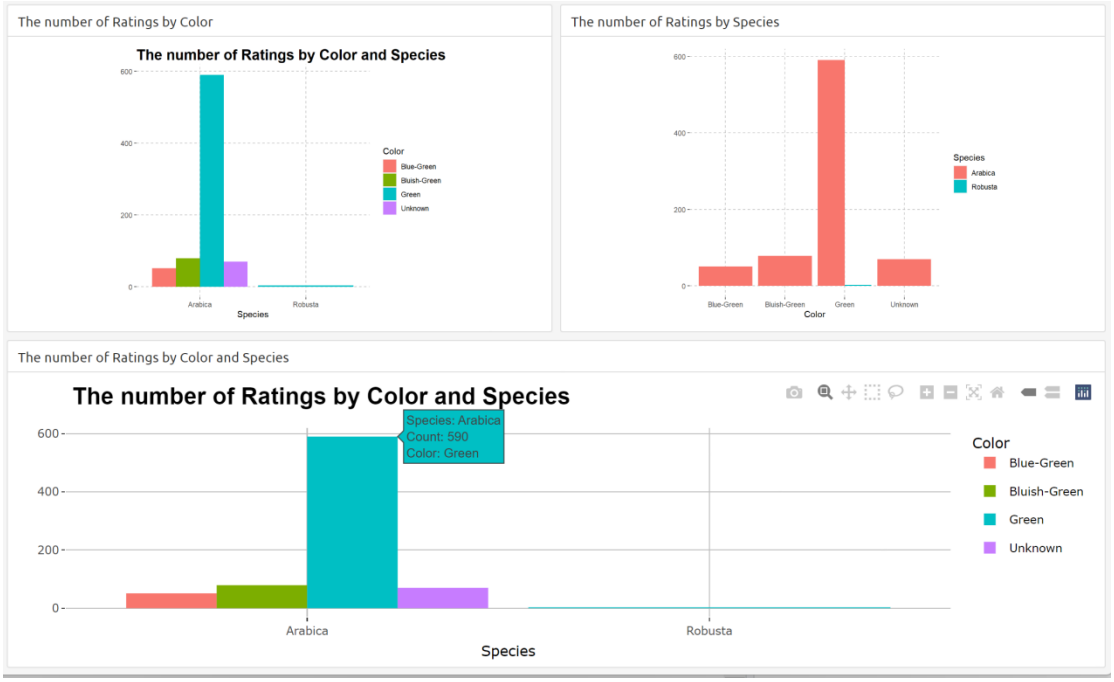
In this section, we analyzed and visualized the number of ratings by color and species.

Principles: Information compressing

Merged Graph: Both color and species information are shown interactively using one chart instead of two bar charts.

8.1 The Number of Ratings by Color and Species

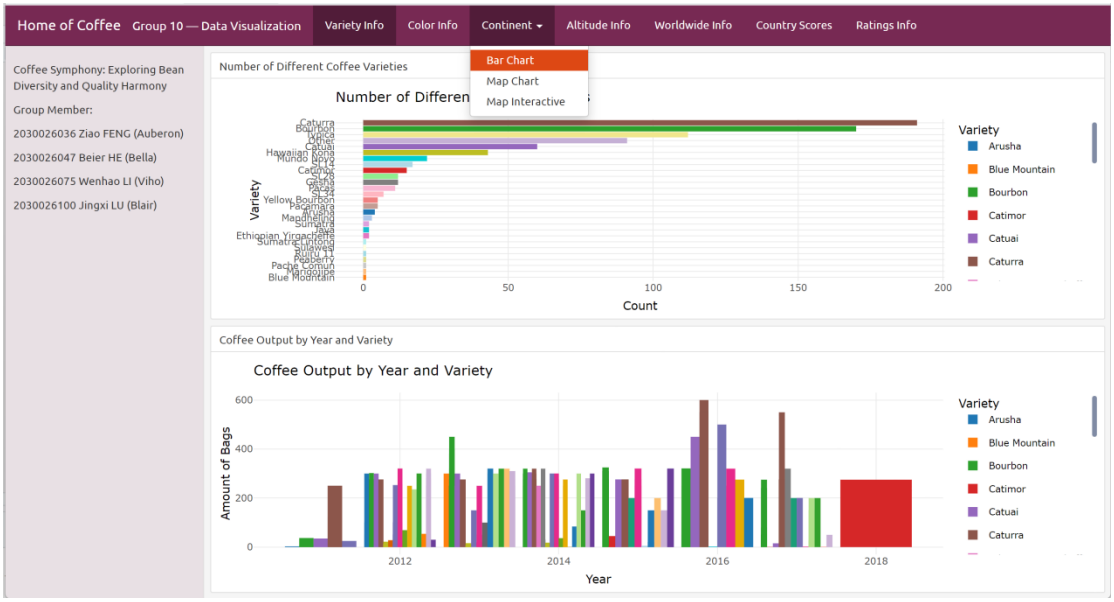
We created two bar plots to show the number of ratings by color and species. The first bar plot shows the number of ratings for each color across different species, while the second bar plot shows the number of ratings for each species across different colors.



9. Dashboard

To provide a comprehensive and visually appealing representation of the coffee data analysis, we have created a dashboard using data visualization tools. The dashboard includes various graphs and visualizations that showcase the key insights and patterns discovered throughout this project.

The dashboard is designed in a user-friendly manner, allowing stakeholders to easily navigate and interact with the data. Each graph is accompanied by appropriate labels, titles, and legends for clarity and understanding.

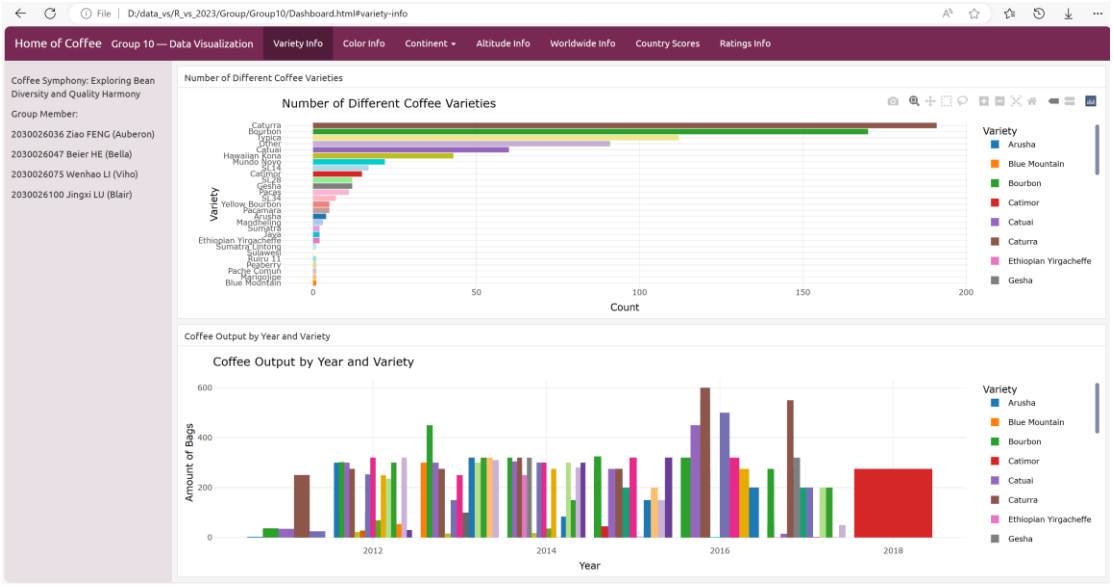


10. Conclusion

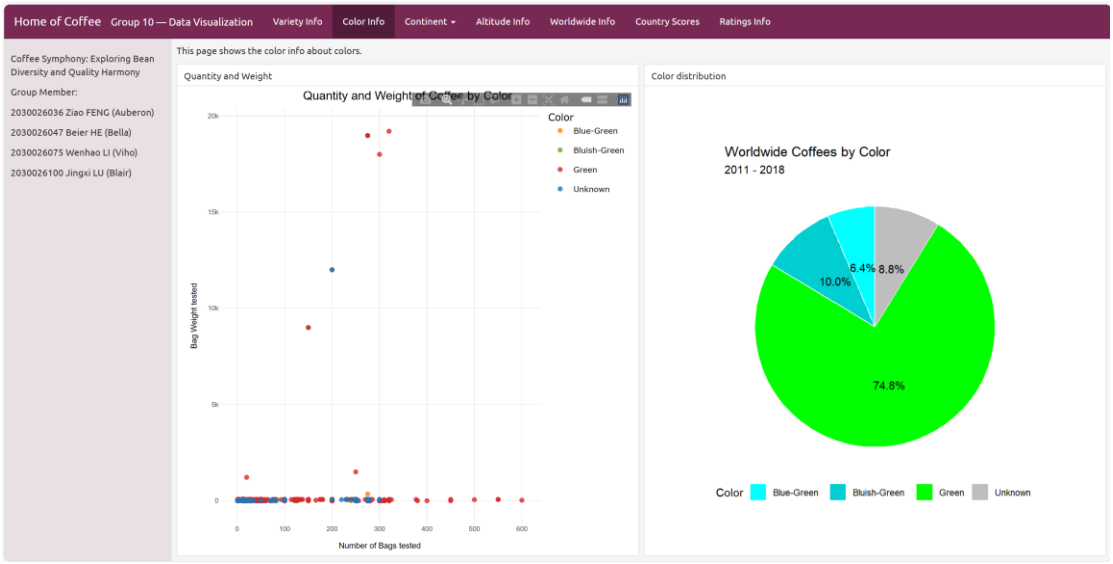
In conclusion, this project explored the diversity and quality of coffee from different regions around the world. We analyzed various attributes of coffee, including varieties, output, colors, altitudes, processing methods, and ratings. Through data preprocessing, visualization, and analysis, we gained insights into the relationships and patterns within the coffee data. The findings from this project can be used to understand and recommend different types of coffee based on their qualities and characteristics.

11. Contribution

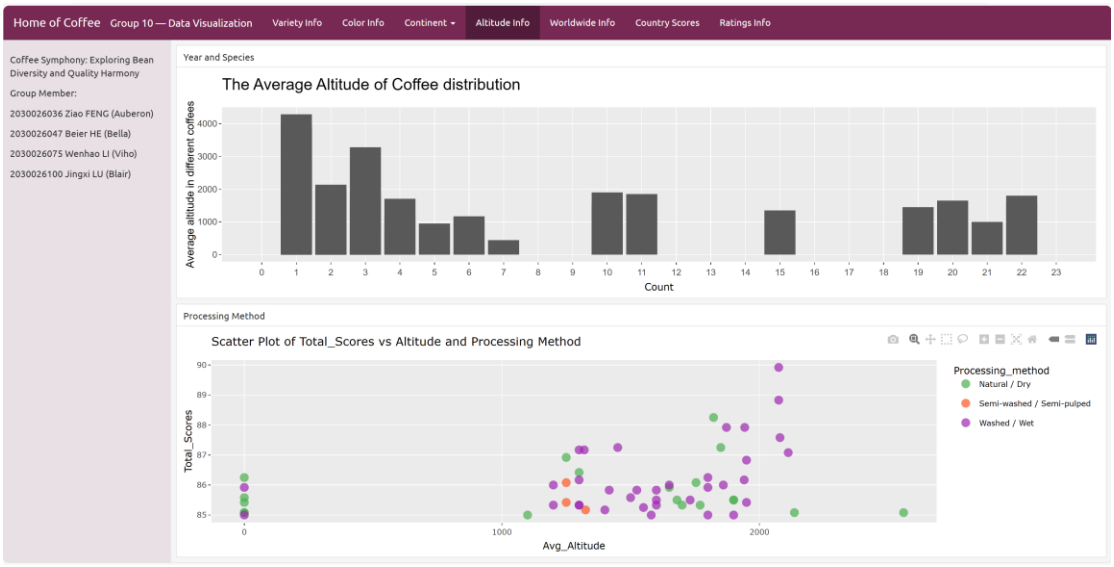
Name_studentID	Description of responsibility	Rating chosen from (30, 60, 90, 100)
Ziao FENG 2030026036	Section 5, 7.2, 8, 9	100
Jingxi LU 2030026100	Section 6, 7.1	100
Wenhao LI 2030026075	Section 3, 10	100
Beier HE 2030026047	Section 1, 2, 4	100

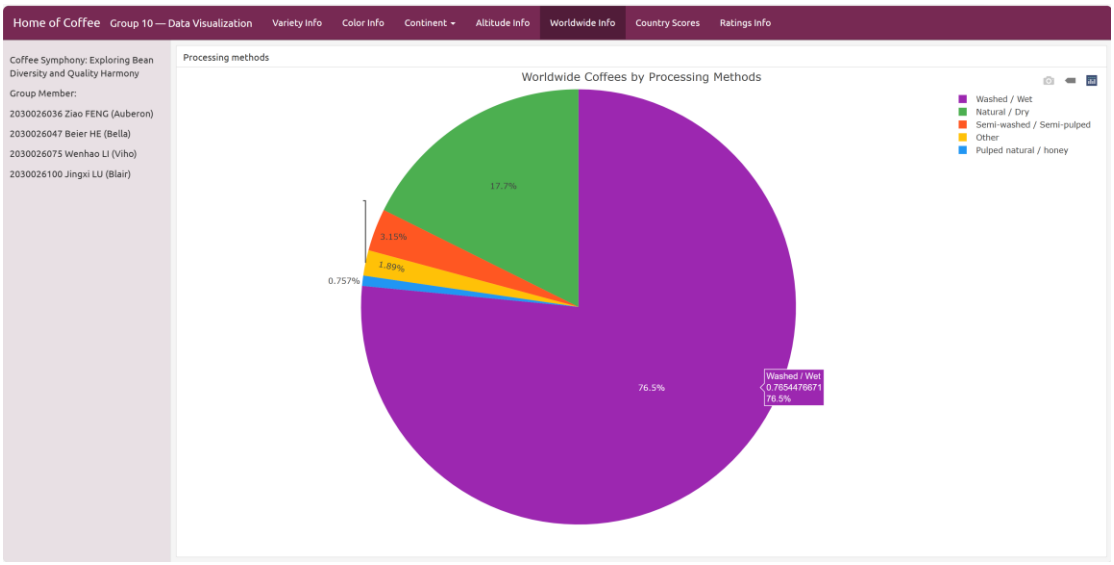


Wenhao LI 2030026075

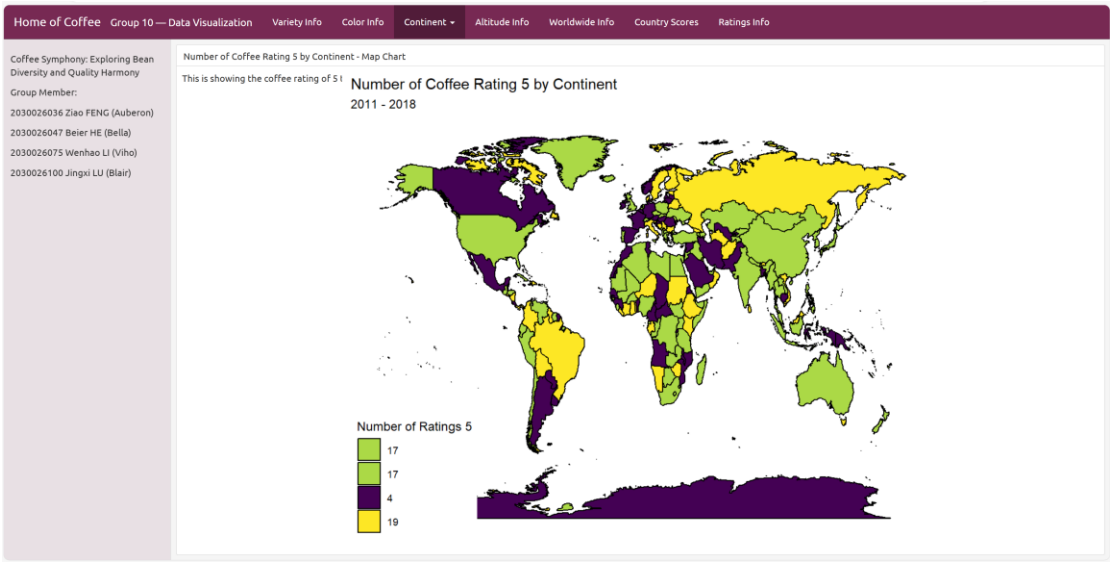


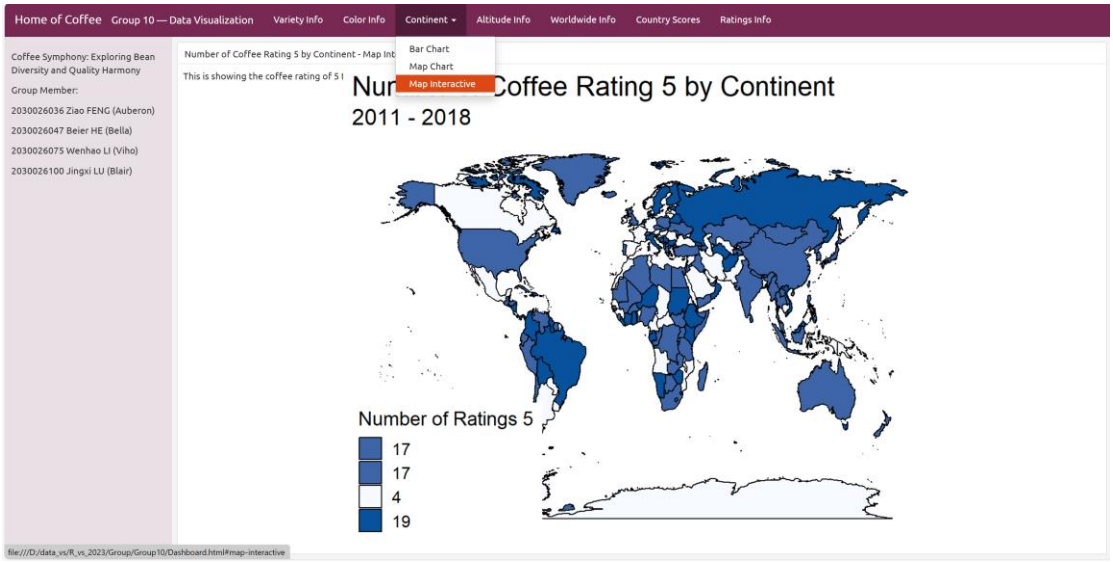
Beier HE 2030026047

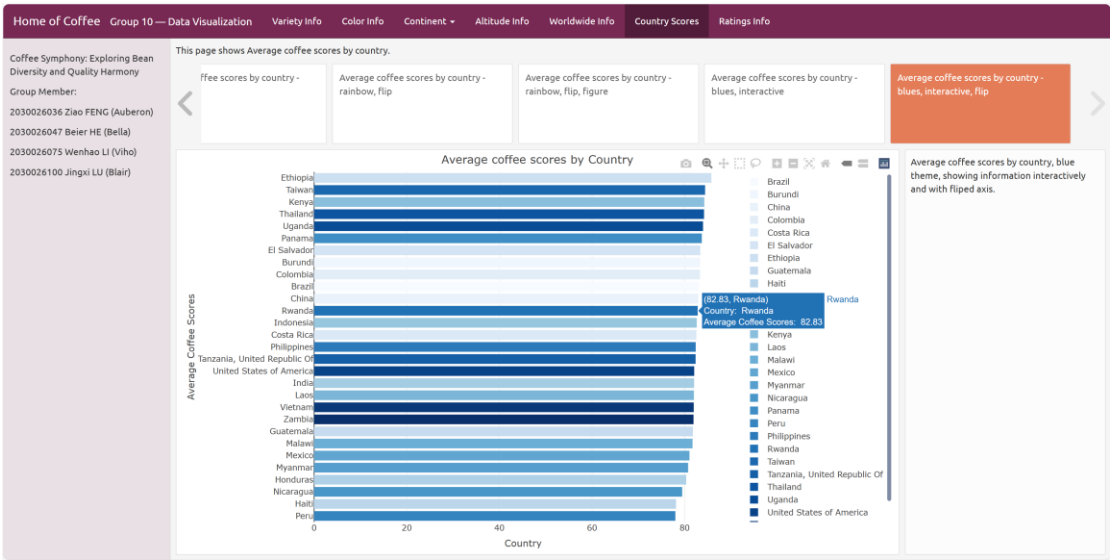


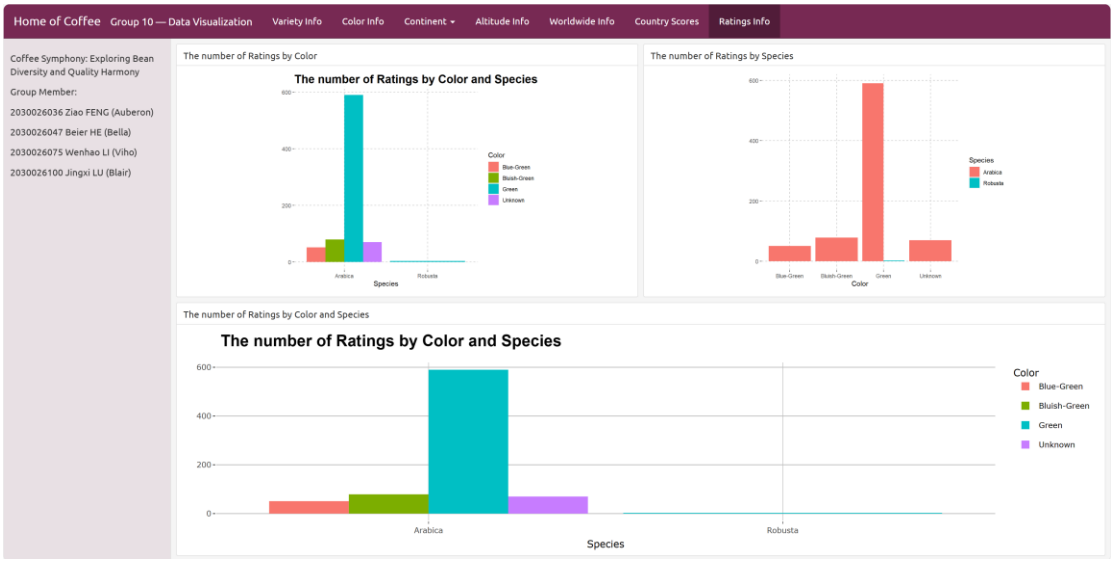


Jingxi LU 2030026100


Home of Coffee
Group 10 — Data Visualization
Variety Info
Color Info
Continent
Altitude Info
Worldwide Info
Country Scores
Ratings Info
Coffee Symphony: Exploring Bean Diversity and Quality Harmony
Group Member:
2030026036 Ziao FENG (Auberon)
2030026047 Beier HE (Bella)
2030026075 Wenhao LI (Viho)
2030026100 Jingxi LU (Blair)
Number of Coffee Rating 5 by Continent - Map Chart
This is showing the coffee rating of 5 t
Number of Coffee Rating 5 by Continent
2011 - 2018
Number of Ratings 5
17
17
4
19


Home of Coffee
Group 10 — Data Visualization
Variety Info
Color Info
Continent
Altitude Info
Worldwide Info
Country Scores
Ratings Info
Bar Chart
Map Chart
Map Interactive
Coffee Symphony: Exploring Bean Diversity and Quality Harmony
Group Member:
2030026036 Ziao FENG (Auberon)
2030026047 Beier HE (Bella)
2030026075 Wenhao LI (Viho)
2030026100 Jingxi LU (Blair)
This is showing the coffee rating of 5 t
2011 - 2018
Number of Ratings 5
17
17
4
19
file:///D:/data_vs/R_vs_2023/Group/Group10/Dashboard.html#map-interactive


Home of Coffee
Group 10 — Data Visualization
Variety Info
Color Info
Continent
Altitude Info
Worldwide Info
Country Scores
Ratings Info
Coffee Symphony: Exploring Bean Diversity and Quality Harmony
Group Member:
2030026036 Ziao FENG (Auberon)
2030026047 Beier HE (Bella)
2030026075 Wenhao LI (Viho)
2030026100 Jingxi LU (Blair)
This page shows Average coffee scores by country.
Average coffee scores by country - rainbow, flip
Average coffee scores by country - rainbow, flip, figure
Average coffee scores by country - blues, interactive
Average coffee scores by country - blues, interactive, flip
Average coffee scores by Country
Average Coffee Scores
Country
(82.83, Rwanda)
Country: Rwanda
Average Coffee Scores: 82.83
Average coffee scores by country, blue theme, showing information interactively and with fliped axis.



Ziao FENG 2030026036