beanRating

Brewing the Best: A data driven Coffee Guide

Process Book

Team Members:

Riya Deshpande	riyarajeshde@umass.edu	34011024
Phuong Nguyen	phulnguyen@umass.edu	32597323
Kevin Dao	bdao@umass.edu	33298829

Github Link:

https://github.com/cynic10/CS_571_beanRating

I] Overview and Motivation

This project explores global coffee ratings to understand how region, species, and processing methods influence quality. The goal is to create an interactive dashboard that allows users to sort and view average coffee ratings based on key attributes like origin and processing style. Coffee is a universally consumed product with significant regional diversity, making it a compelling subject for visual exploration.

II] Related Work

While we did not reference a specific dataset or research paper, our project was born out of a shared team interest in coffee. We were curious to explore trends around coffee ratings, origins, and processing styles. The inspiration to build an interactive and engaging dashboard came after viewing the Pokémon dashboard example presented in class. It sparked the idea that data visualization can be both fun and informative, and we wanted to channel that same energy into a real-world topic we all enjoy—coffee.

III] Questions

Initially, we aimed to answer:

- What countries consistently produce high-rated coffee?
- Does the processing method impact rating?
- Are certain species or varieties more associated with high scores?

As the project evolved, we added:

- How do score distributions change by region?
- Can we visualize production trends and match them to quality?

IV] Data

We found a really good dataset on Kaggle, which encompasses a lot of things which we wanted to depict. This dataset contains detailed information on coffee quality for different attributes like acidity, sweetness, aroma, moisture, etc. along with its country of origin and production amount for different years.

Data Source:

Coffee Beans Rating Dataset

The dataset contains almost 1000 rows with both categorical as well as numerical columns. The dataset seemed pretty comprehensive but needed to be cleaned before usage.

Cleaning Steps:

- Removed irrelevant columns.
- Aggregated by country, region, species, variety, and processing method
- Calculated average scores and quality categories
- Created separate CSVs for ranking table and line charts (to be used later)

V] Exploratory Data Analysis

Our initial analysis was done manually by inspecting the dataset column by column. We looked for missing values, inconsistent entries, and columns that didn't contribute meaningfully to our visualization goals. One such column was Data.Owner, which represented the vendor or source of the data—it was inconsistent, had missing values, and didn't offer insights relevant to our analysis, so we decided to drop it.

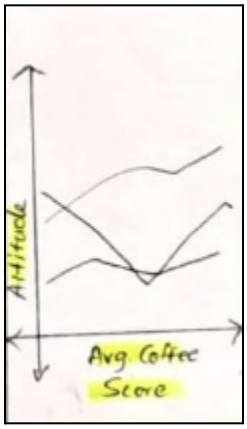
While column Location.Altitude.* was complete, we excluded it from the main visualization due to limited relevance to our core questions. However, we retained Data.Production.Number of bags and Data.Production.Bag weight, as we plan to use them later to estimate overall coffee production per country over time.

Rather than plotting charts, we used basic summaries and value distributions to guide our decisions. This helped us streamline the dataset for our ranking table and line chart visualizations, focusing on meaningful columns like country, region, species, variety, processing method, and total score.

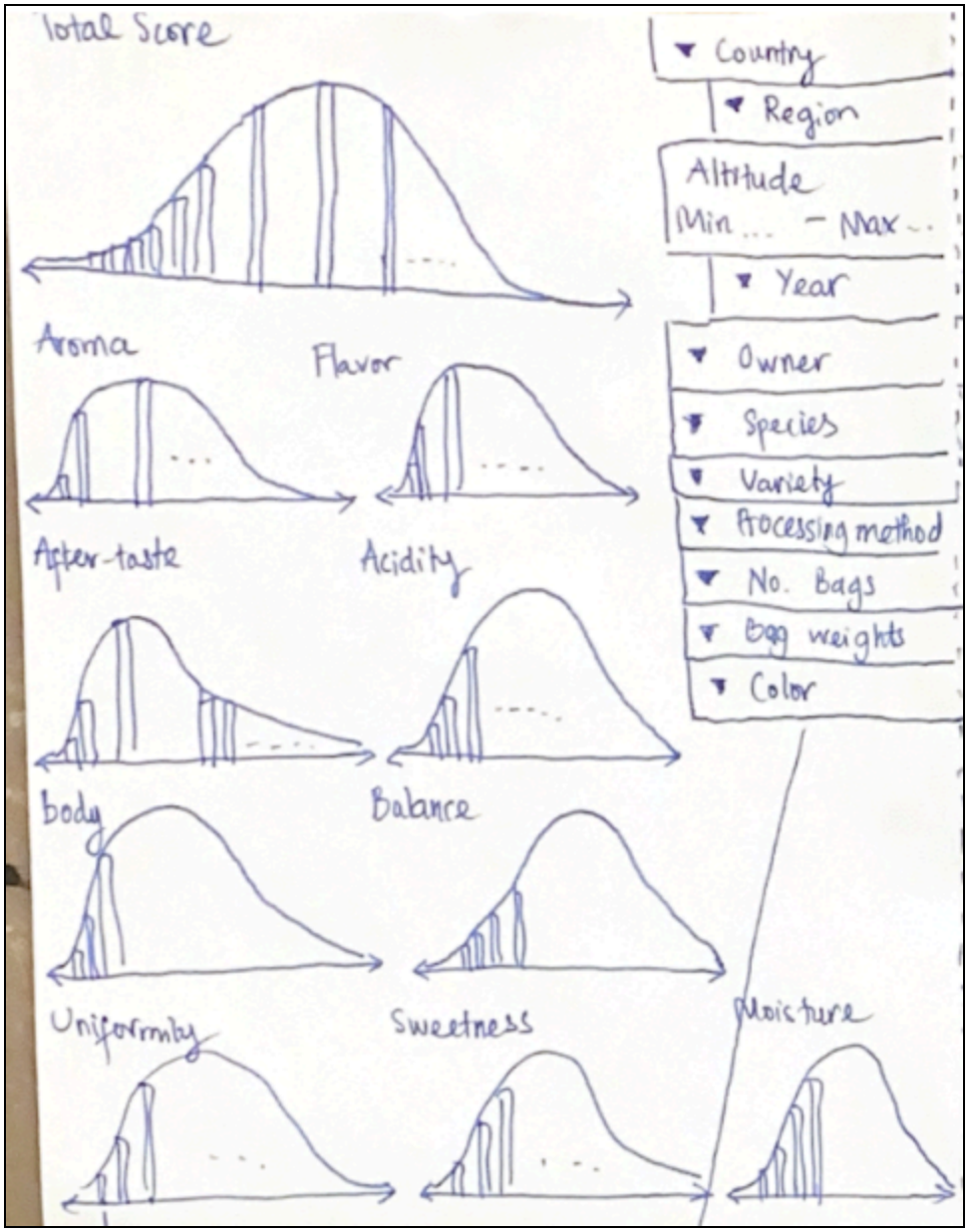
VI] Design Evolution:

Throughout our design process, we considered a wide range of visualizations to explore the coffee dataset from multiple angles. Some of our early ideas included:

- **Altitude vs. Average Score Scatter Plot**: We initially wanted to explore if altitude influenced coffee quality. However, we found that the altitude values were inconsistently populated and often missing, which limited the reliability of such a visualization. This idea was ultimately deprioritized.

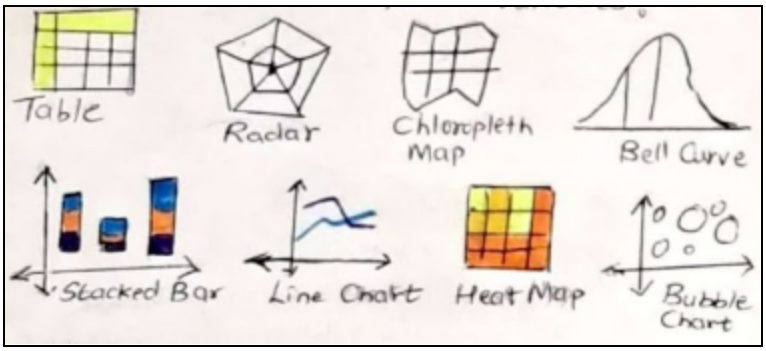


- **Bell Curve of Total Scores**: We proposed a bell curve to show the distribution of total scores across all coffees. However, as per feedback received on the proposal , we realized this might become overly comprehensive and could potentially confuse the viewer rather than offer clarity. We decided to keep our visualizations more exploratory and comparative rather than statistical.

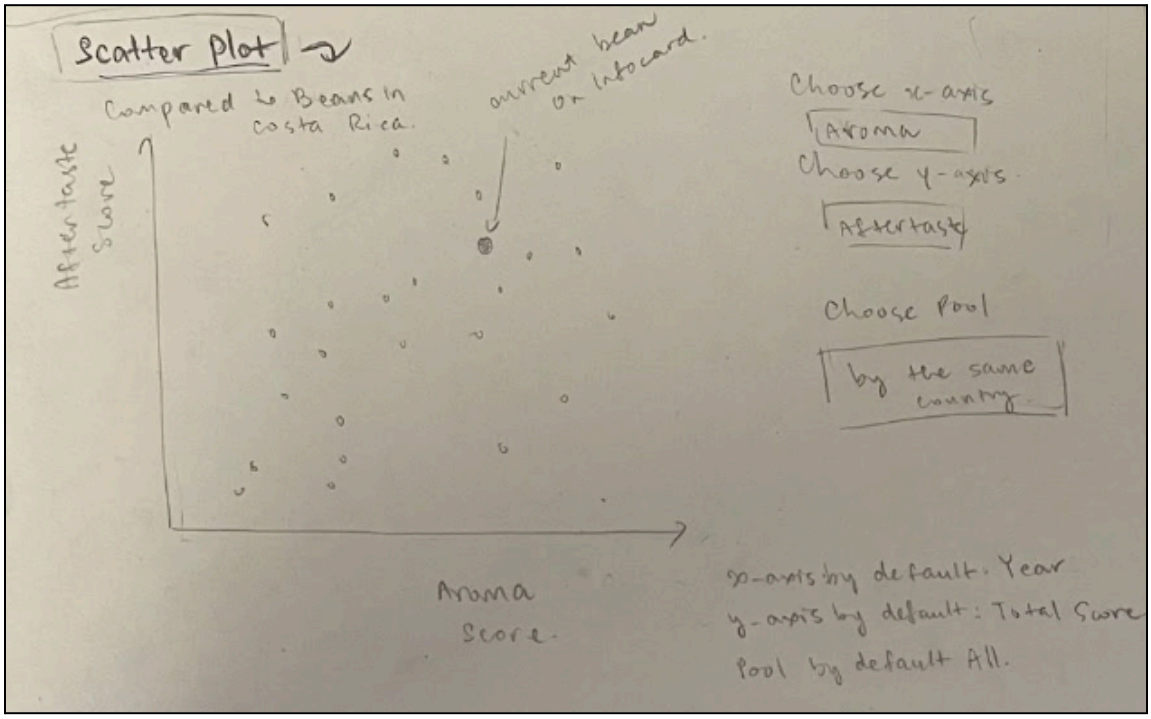


- **Choropleth Map**: A map-based visualization seemed like a natural fit for showcasing global coffee data. However, the dataset included a limited number of countries and many missing region/place fields, which made the choropleth less effective. We chose not to pursue this direction due to incomplete geospatial data.

- **Bubble Chart**: While visually engaging, a bubble chart didn't add much new insight beyond what other visualizations could already offer. It also introduced redundancy with the radar chart and scatter plot concepts, so we decided to skip it.



- **Score Comparison Scatter Plot**: We explored the idea of comparing individual score components (e.g., aroma vs. acidity), but we found no significant patterns or insights through pairwise comparisons. The visualizations became cluttered and didn't offer a clear story, so we decided against it. Moreover we also felt that it would not help users much to compare coffee characteristics with each other rather than studying individual traits for each coffee or comparing the total scores across coffee varieties.



Ultimately, we selected a combination of **interactive ranking tables**, a **radar chart**, and a **line chart**. These allowed us to cover different aspects of the data—overview, detail, and trend—while adhering to strong perceptual principles like position for precision, color for categories, and consistent encoding. They also balanced complexity with interpretability, which we found essential for user engagement.

While we deviated from some of the ideas in our original proposal, our final visualizations provide a more focused, interactive, and insightful exploration of

the coffee dataset. If time permits, we may reintroduce some of the other visualization ideas in future iterations.

VII] Implementation:

A] Radar Charts:

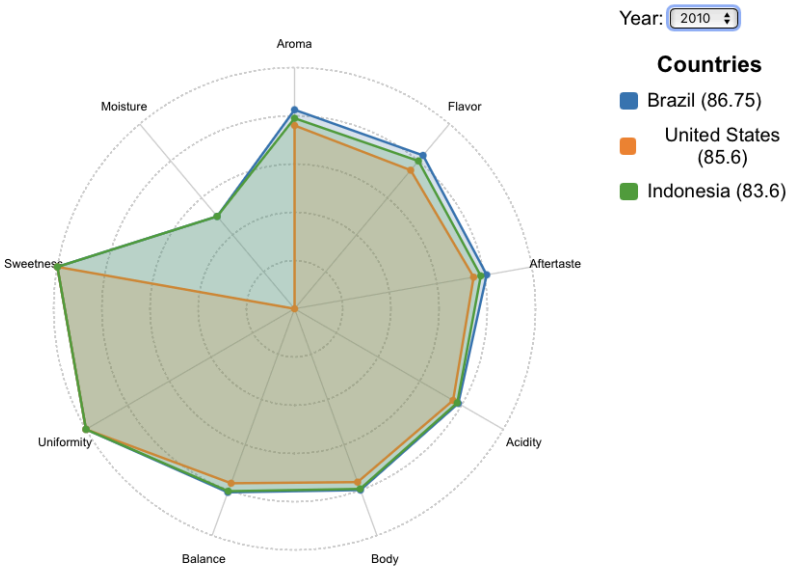


Fig 1. Current working iteration of the radar chart

For the radar chart, we have a working visualization that depicts countries and their coffee's strengths and weaknesses based on 9 different scoring categories, which are aroma, flavor, aftertaste, acidity, body, balance, uniformity, sweetness, and moisture (Fig 1). A total score across all categories for any individual country can be seen next to that country on the right legend.
The radar chart contains several options that allow a more fine-grained and isolated visualization of the rich data that we have gathered. These features include a drop-down menu that can filter the scores and countries by year and an option to toggle countries on and off in the visualization. Because the data that we have collected has different scores for many countries over different years, we have decided it would be in the viewer's best interest to isolate the year for a more insightful comparison. In addition, because we recognize that the user may also want to filter out countries for a more detailed analysis between small groups, pairs, or even individual countries. Therefore, we have added an option where countries can be filtered out by clicking on the country name in the legend. Examples of these implementations can be seen in Fig 2 and 3.

For future iterations, we aim to fully animate the graph to enhance the user's engagement and create a more enjoyable viewing experience. We also aim to allow the user to hover over data points/countries and have an in-depth report on the actual scores for that country. In addition to that, we also want to implement a dynamic radar chart, where scoring categories can also be filtered out, much like how countries can be filtered out in our current implementation.

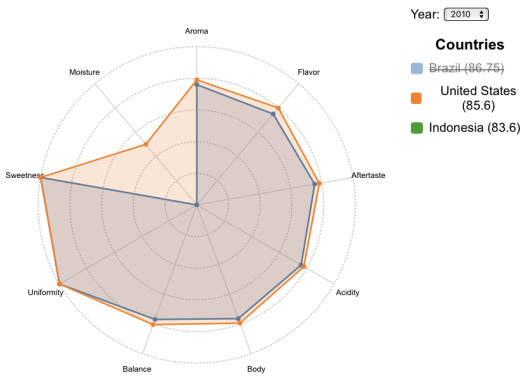


Fig 2. Countries can be filtered out

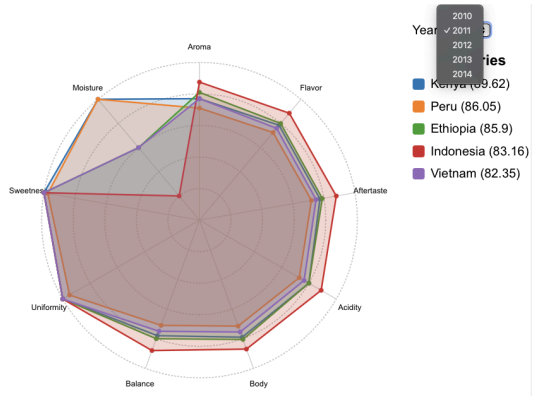


Fig 3. Different years can be selected

B] Ranking Tables:

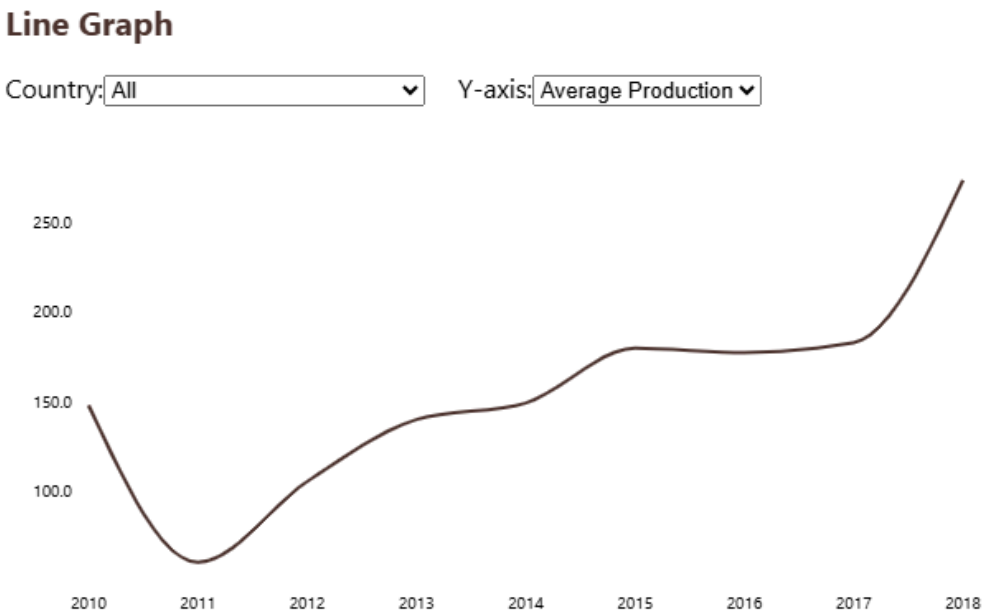
The ranking table provides an overview of all coffee beans by displaying their overall ranking. At first, the ranking table only ranked based on the total score. But later, our team realized the importance of individual score components, so we added the option to rank based on those, too. For text columns, clicking the column name will sort the rows alphabetically. For n

Ranking Table

#	Country	Year	Aroma	Flavor	Aftertaste	Acidity	Body	Balance
1	Brazil	2010	8.25	8.30	8.10	7.85	8.00	8.10
2	Ethiopia	2011	7.70	7.85	7.75	8.00	7.90	7.80
3	Colombia	2012	8.60	8.55	8.40	8.30	8.50	8.45
4	Indonesia	2010	7.60	7.50	7.55	7.60	7.65	7.70
5	Kenya	2013	8.70	8.80	8.60	8.90	8.75	8.80
6	United States	2010	7.90	8.00	7.85	7.80	7.95	8.05
7	Vietnam	2011	7.10	7.25	7.20	7.30	7.15	7.10

umber columns, clicking will sort the table in either increasing or decreasing order. Since the dataset contains hundreds of rows, we wrapped the table in a scrollable container to keep it organized and easy to navigate. We also made the header row fixed, so the column names stay visible no matter how far the user scrolls. In the next phase, I plan to improve it by adding filtering options, allowing users to view rankings based on specific regions, bean types, or processing methods for a more personalized and insightful experience.

C] Line Chart



The line graph allows users to select the Y-axis variable, either Average Production or Average Score, depending on what they wish to analyze. Users can also filter the data by selecting a specific country or viewing data from all countries combined.
In future improvements, we plan to add a clearer and more descriptive visualization key to help users better understand the axis values. Currently, the line graph only displays one line at a time. We aim to enhance this by allowing multiple lines on the graph, enabling users to compare data between selected countries and the overall average more effectively.

VI] Evaluation:

From the visualizations, several points could be made about the global coffee production industry :

- From the line graph, trends could be observed on coffee production over time across countries, where one can draw conclusions on how the key players in coffee production have shifted throughout the years. These trends can be reflected both from the average production, as well as from the average score.
- From the radar chart, a more side-by-side comparison can be made on what country's coffee tasted the best, according to experts' opinions. In addition, these scores can also reflect the coffee production, as seen in the line graph, where good tasting coffee has higher demands, and therefore resulting in high production.

The visualizations gave a good overview on the coffee production and score for each country. However, with our current implementation of the charts, not much insights can be gained from the graph without more statistical tools such as the use of trend lines, or the abilities to isolate countries/years for even more fine-grained analysis. Our hope is in future iterations, a more conclusive and insightful analysis can be done via our visualization.