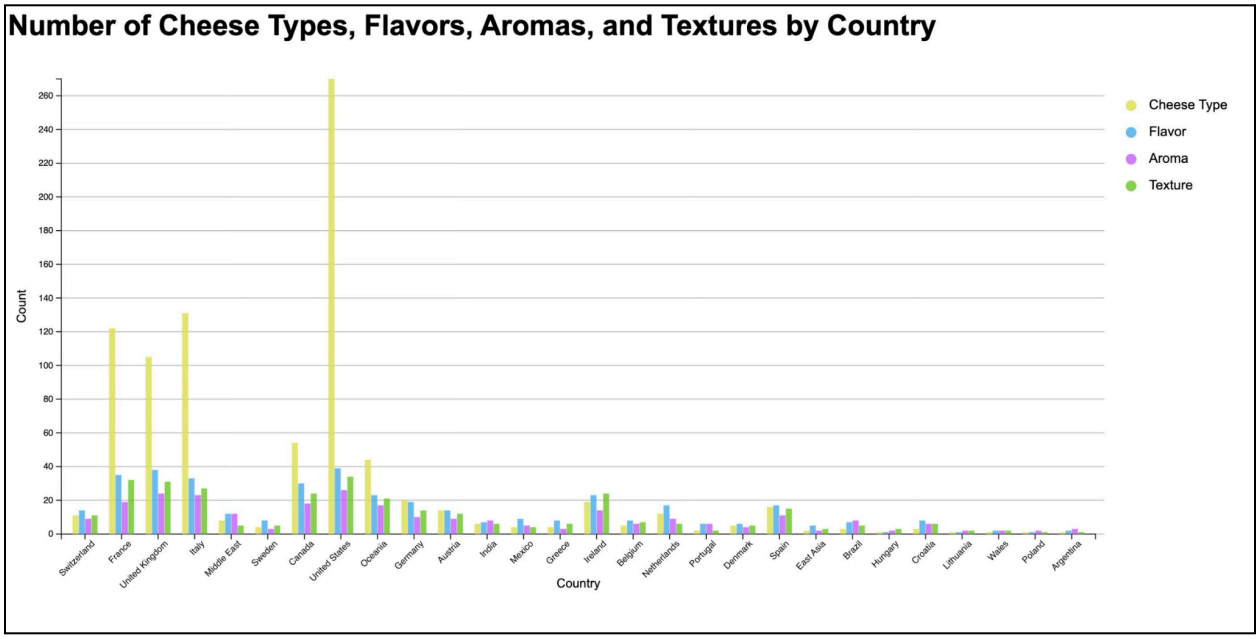
Khushi Patel (ksp67), Lauren Kwan (lmk258), Jose Vizueth (jdv72), Kody Yang (kyy23)

Final Report

Final Visualization:



Description of Data:

This is a cheese data set. It displays different kinds of cheese from all around the world. There are variables such as fat content, country, cheese type, texture, flavor, etc. It makes us wonder if there is a relationship between countries and cheese type as well as texture, flavor, or aroma. Rather than using all of the data, we selected data that was mostly not NA and had a lot of information from different countries and that said something interesting about the cheese itself, more than its manufacturing.

We got the data from

https://github.com/rfordatascience/tidytuesday/blob/master/data/2024/2024-06-04/readme.md

An overview of your design rationale:

We decided to create a bar graph visualizing the number of cheese types, aromas, textures, and flavors by country. We went with a bar graph because it gave us the best way to show comparisons between each variable in a meaningful way. We made sure each variable had a different color bar to show differentiation. We have a legend to show this differentiation and make it easy for viewers to pinpoint which color means which variable. When we first created the visual, the bar lines were very cluttered and there were too many. We further condensed some countries and got rid of some countries that had no data resulting in the graph more readable. One trade off for doing this is that we are not able to show as much specific data as it originally did since we are clustering some countries/regions together. However, we believe that the final visualization made the endpoint clearer while changing the data minimally. We decided to use drastically different colors for the bar graphs so that it would be easier to observe the bars, while putting each of the bars for each of the countries together so that the user can compare by country in an easier way. After the critique session, we added axis labels and gridlines to highlight visual channels and to make it more readable for viewers. We also had to allocate more space for the width because there are a lot of data points to be included. A smaller width was cutting it off.

The story:

This visualization provides a comparison of the diversity of cheese offerings across different countries. The United States stands out with significantly higher numbers in all categories compared to other countries. This suggests that the US has an especially diverse cheese production and consumption culture. France, The U.K., and Italy also have substantial cheese varieties, in addition to the U.S. This makes sense as it aligns with the traditional cheese-making heritage in these European regions. It was honestly surprising that there is a

significant jump in numbers for the United States, indicating either a large-scale cheese production or a broader variety of processed and artisanal cheeses. This comparison could also inspire further questions about regional preferences and how historical cheese-making traditions stack up against modern mass production in countries like the U.S. As such, we were able to argue that, while the US might not be as popular of a cheese manufacturer as other European countries, it is still able to become the leading country in cheese diversity.

Team Contributions:

ksp67: Made the initial cheese types graph, kept communication between members consistently, created the milestone deliverables

lmk258: Added Aroma types to graph, made and modified the docs used between members throughout project, created legend for graph

jdv72: Added Texture types to graph, modified the docs for correctness, added comments to code, Helped with merge conflicts.

kyy23: Added flavor types to the graph, handled merge conflicts between all github branches, polished final visualization by adding axis labels and gridlines

10-15 hours developing
The most difficult part was mapping each variable to create the multi-bar graph.