CPSC 4030
Dr. Federico Iuricich
**Project Title:** Amazon Customer Trends
Names:
- Lucas Boyer (lrboyer@clemson.edu C91887948)
- Garrett Versteegh (gverste@clemson.edu C77126911)
- Trevor Rizzo (trizzo@clemson.edu C49093535)

**Overview and Motivation:**

Our dataset contains behavioral survey data from Amazon shoppers. We found the data particularly interesting because it was able to demonstrate, with a large sample size, various usage habits of shoppers and demographics on one of the world's leading e-commerce sites. This dataset asks what people tend to buy, what demographic they are, what they review, how reviews impact consumer perception, and many other questions related to various consumer behaviors. We determined that after the completion of this project, we could gain a better understanding of the demographic composition, habits, and user types prevalent among Amazon consumers. Our dataset comprises behavioral survey data from individuals engaging with Amazon. What we believe we can learn from the visualization and analysis of it could prove beneficial to companies and customers long-term. The insights we gain from the data could also prove to be interesting to see what other shoppers' habits are like.

**Related Work:**

The inspiration for using our dataset stemmed from a curiosity about how a service as large and developed as Amazon's e-commerce platform was utilized and perceived. When searching for a dataset, we were determined to find a set with business and technological implications, and the intersection of user experiences with different features and functionality with Amazon's website seemed like a perfect match.

**Questions:**

The questions we are trying to answer:
- What does the age and gender distribution of Amazon purchasers look like, and how do these demographics relate to one another in response to purchase frequency?
- How do age and gender demographics play a role in the overall profile of how users interact with Amazon's e-commerce features?
- What is the distribution of each gender and age range that uses any of Amazon's features?

Initially, we started with a handful of surface-level questions, more representative of direct cause and effect relationships. Our other initial questions involved very vague or subjective ideas that were not answerable within the scope of our dataset. One of these questions included "How does frequency of app/purchase correlate to familiarity with the app?" for which we determined we would not be able to answer that question accurately and effectively with our data. As we worked with our preliminary visualizations and prototypes, we asked ourselves what hadn't become obvious but felt close to being answered, which is how we landed on our final questions.

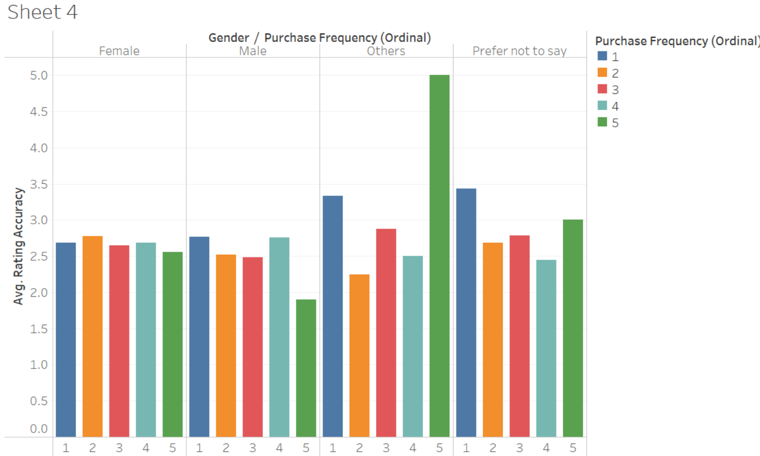
**Data:**

We obtained our data from a dataset site kaggle.com. Link:
[https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset](https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset)
Our dataset came from a consumer behavior survey conducted in June of 2023. The survey features 601 unique responses, with 24 attributes ranging from demographic information to responses centered around various aspects of the Amazon consumer experience. There are ordinal prompts, such as purchase frequency, browsing frequency, cart completion, frequency, personalized recommendation usage, and shopping satisfaction. There are also categorical prompts, like purchase categories, product search methods, cart abandonment factors, and service appreciation aspects. We did not have to do significant data cleanup, as the ordinal responses had pre-selected answer choices, and most of the relevant data for our questions was clean and organized. We did have to add an additional column named "Age Group" which put the responses into age groups ranging from 0-20, 21-30, 31-40, 41-50, and 51-65. We did have to delete 2 responses as their ages did not have a value.

**Exploratory Data Analysis:**

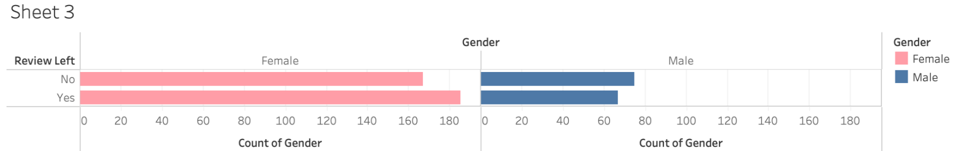
We explored many different visualizations using Tableau:

Sheet 4

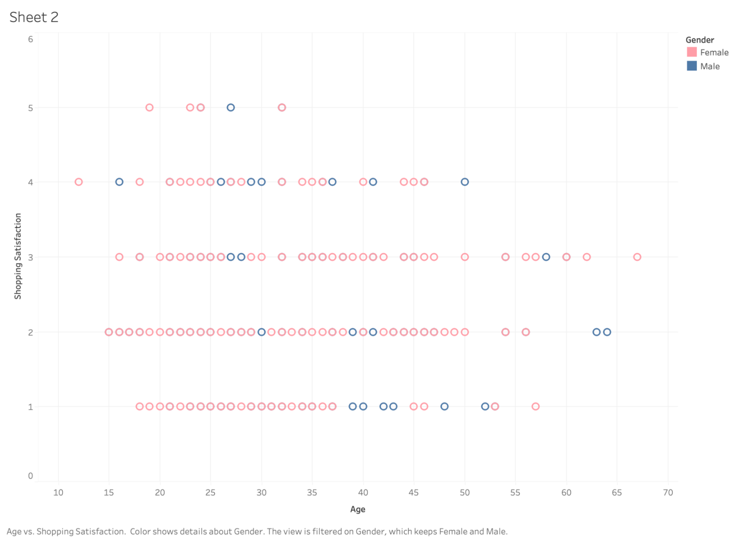



Average of Rating Accuracy for each Purchase Frequency (Ordinal) broken down by Gender. Color shows details about Purchase Frequency (Ordinal).

Sheet 3



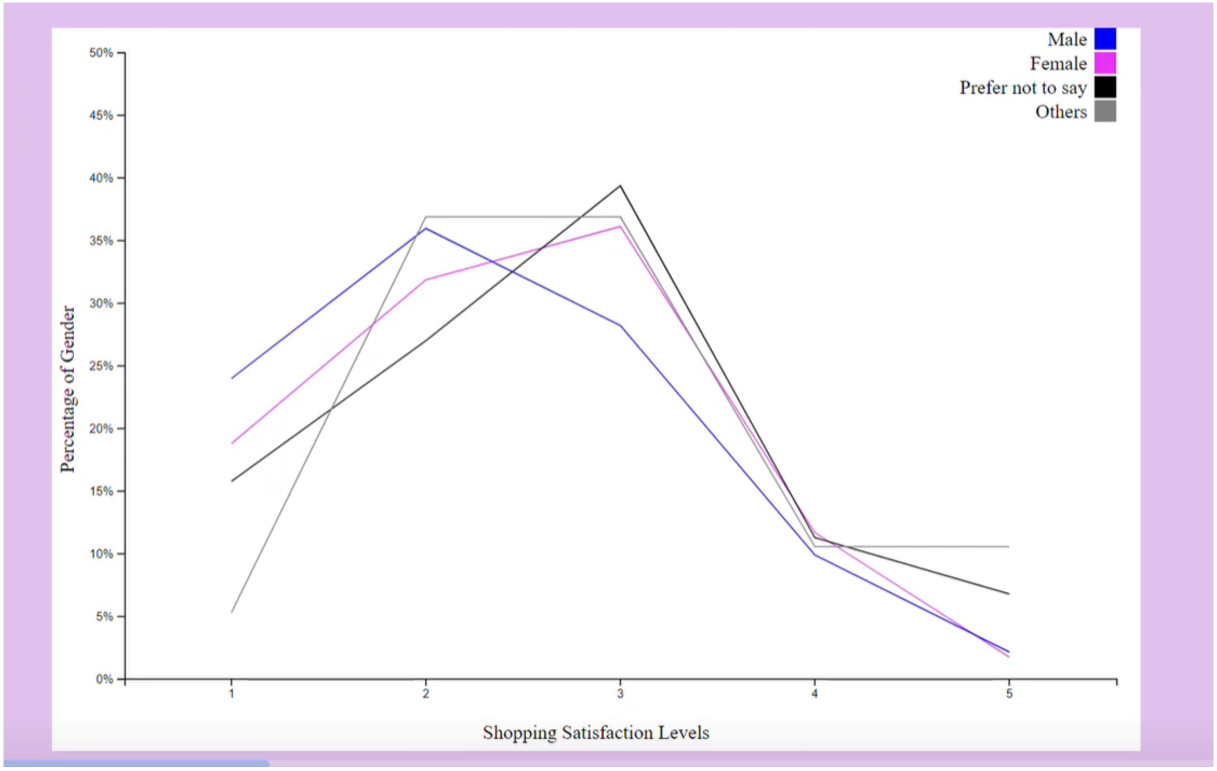

Count of Gender for each Review Left broken down by Gender. Color shows details about Gender. The view is filtered on Gender, which keeps Female and Male.

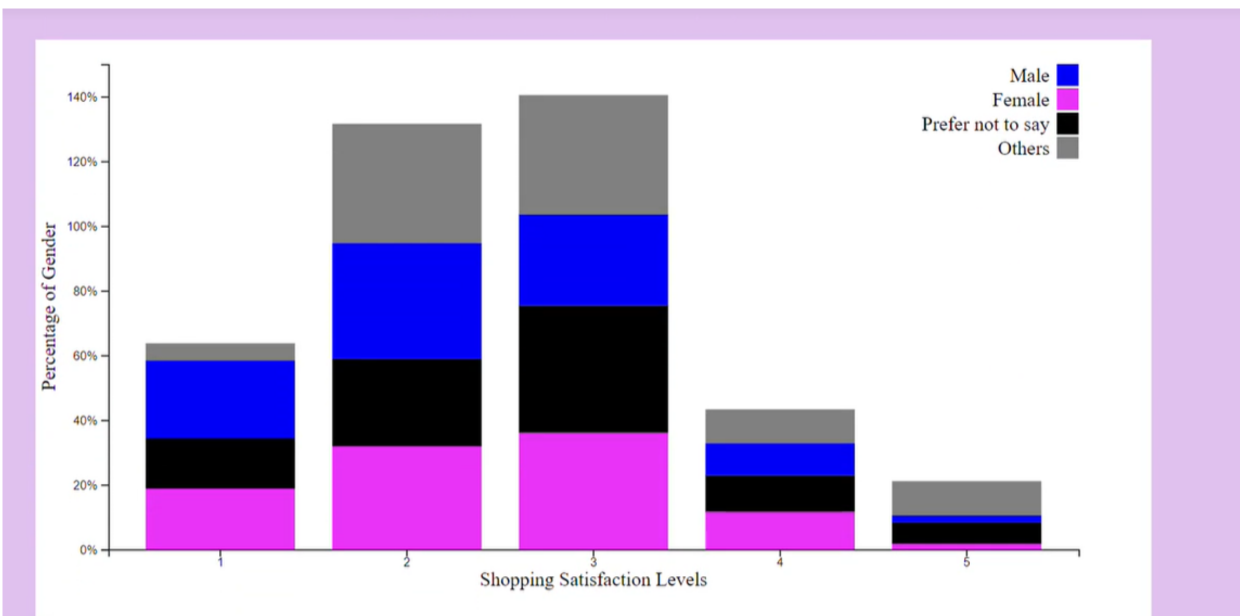




We gleaned valuable insights, focusing particularly on gender, age, and the frequency of shopping, browsing, and cart completion. We also discovered the age data would cause problems using it as is. We decided it would be easier if the age data were in an ordinal format of age ranges. We also discovered that 2 of the entries in the data had a missing value for age. Given that gender, age, and shopping behaviors were crucial to answering our questions, we made our visualizations to include these variables and allow for filtering between them.

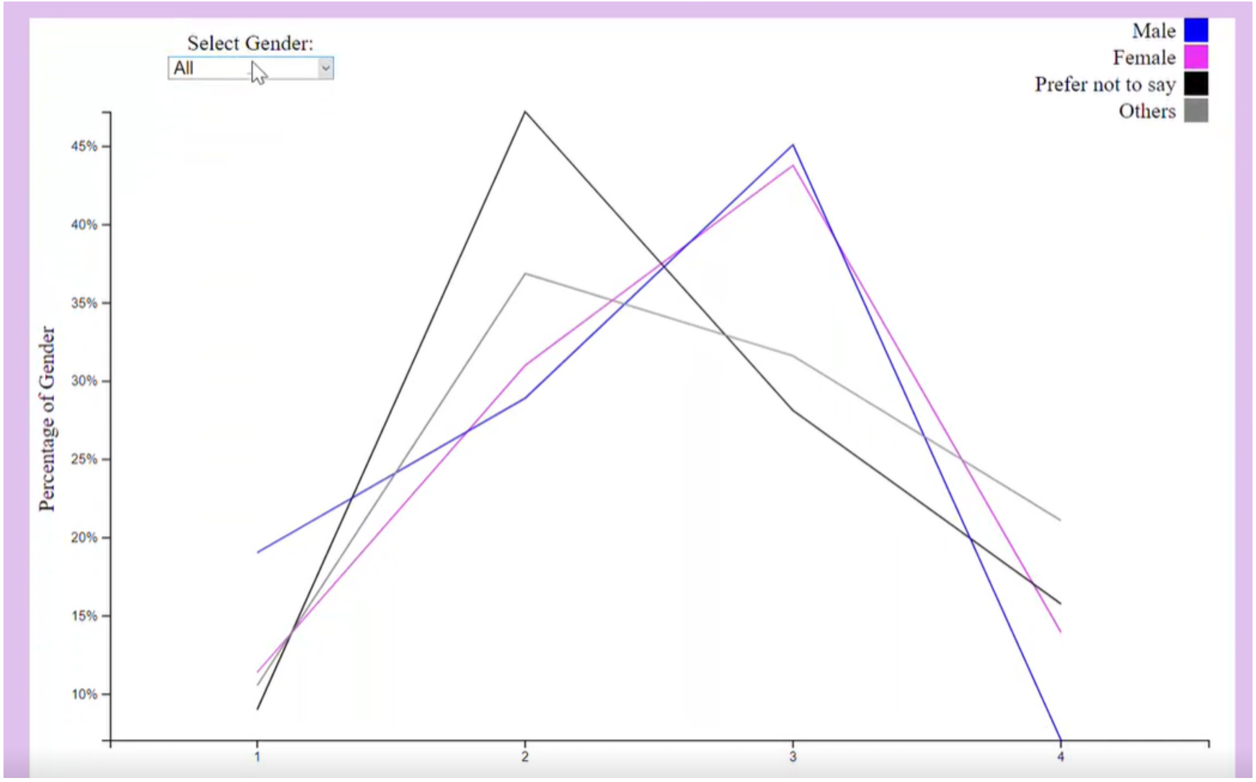
**Design Evolution:**

Seen below are our original visualizations. Our primary focus initially was solely looking at gender and its interaction with customer perception and usage. The first two visualizations show gender compared to shopping satisfaction, the third shows browsing percentage by gender, and the pie charts show shopping frequency by gender. We quickly realized that first, gender was not ideal as many of the interactions showed minimal difference by gender. Further, our choice of marks with the line charts was not effective as it was not able to visually indicate a clear distinction between genders. Our marks of color and distance had proven to be somewhat effective and helpful, and we decided to continue with them as we moved forward. We also realized that, as mentioned above, some of our questions were unanswerable, and our visualizations were weak as a result. We deviated from our initial proposal and created 4 new proposals.

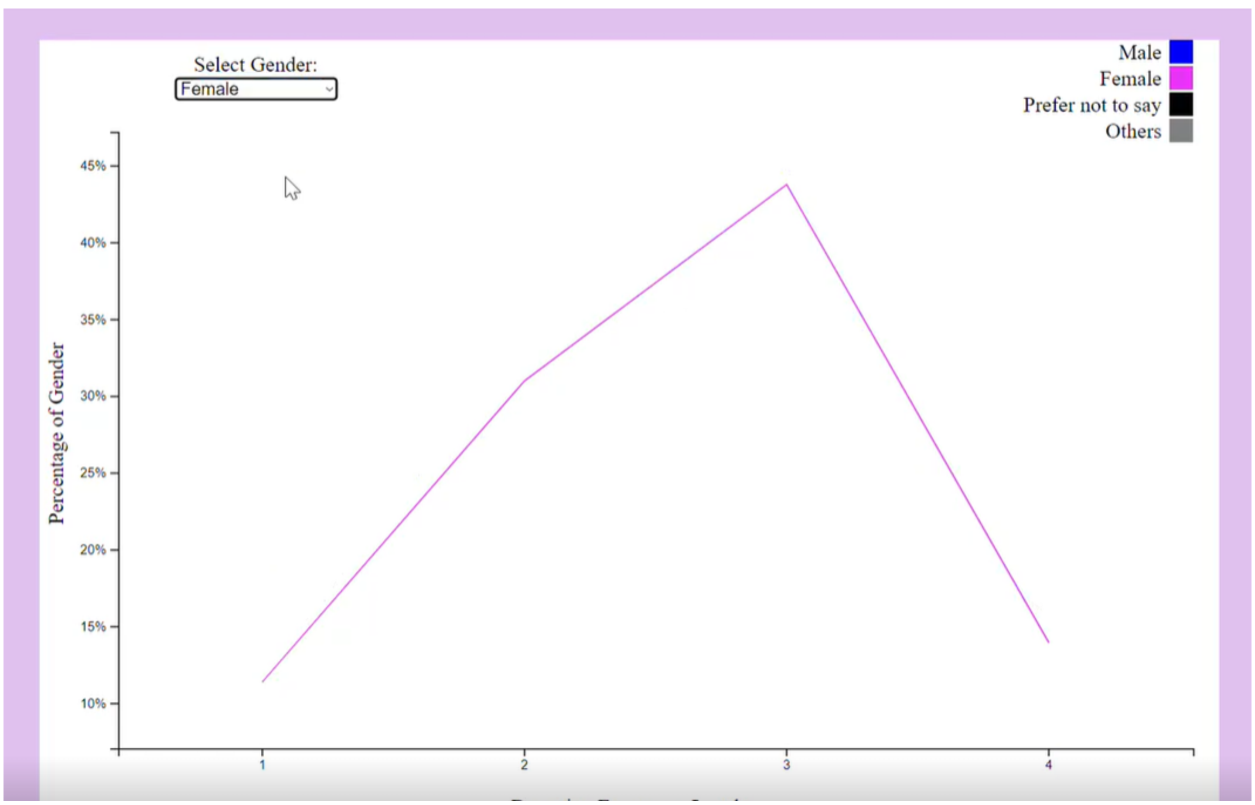



User Shopping Satisfaction by Gender (stacked bar).

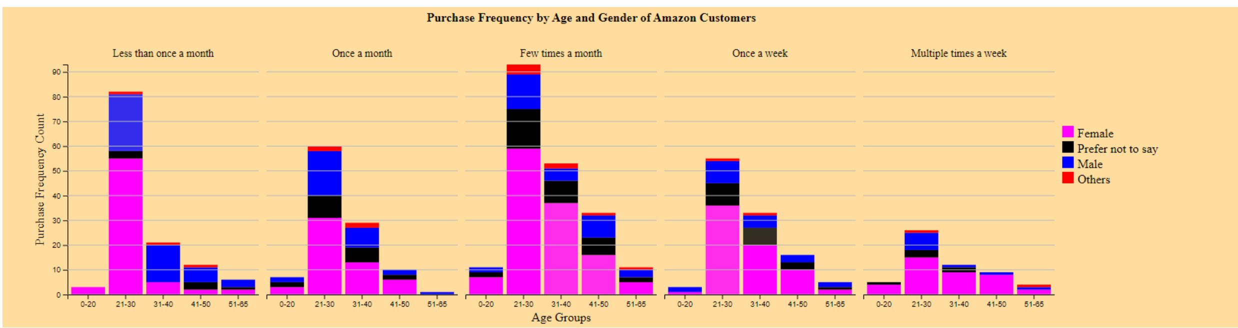


Select Gender:
All
Male
Female
Prefer not to say
Others
Percentage of Gender
45%
40%
35%
30%
25%
20%
15%
10%
1
2
3
4

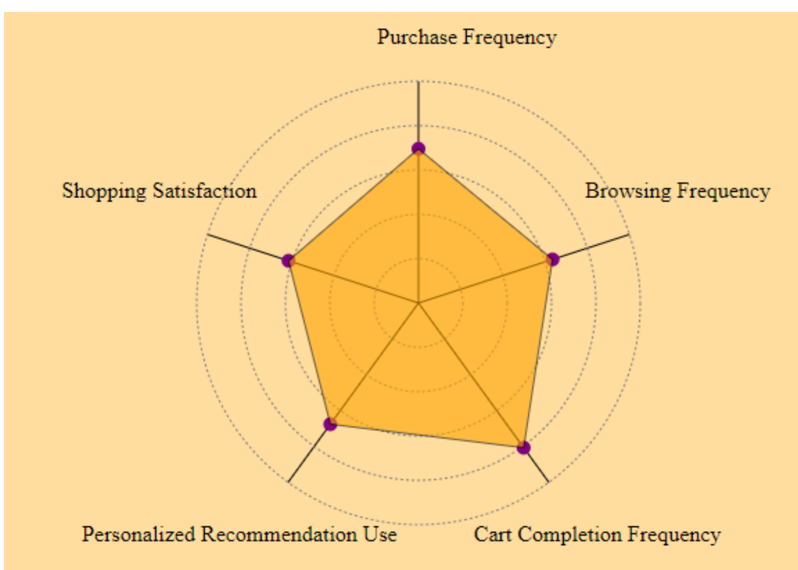



Select Gender:
Female
Male
Female
Prefer not to say
Others
Percentage of Gender
45%
40%
35%
30%
25%
20%
15%
10%
1
2
3
4

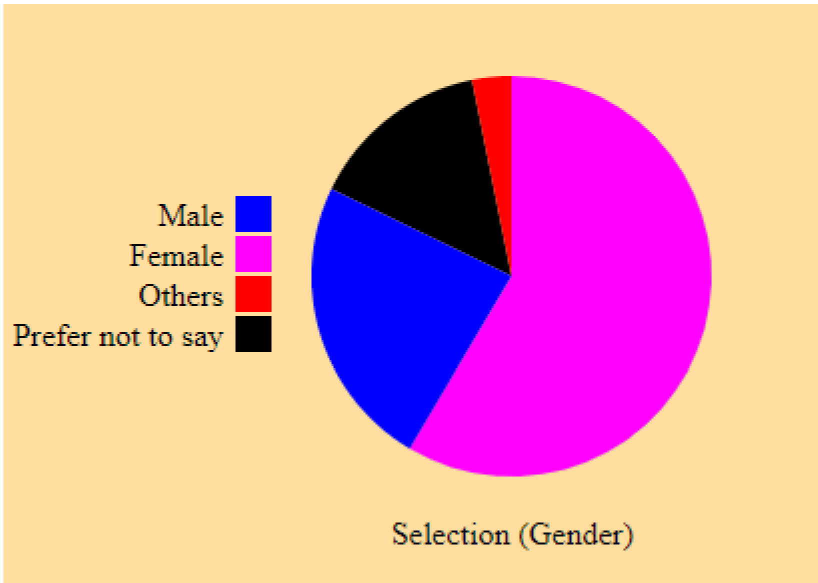

For the second half of the project, we shifted gears and tried out a few different visualizations. The stacked bar chart was able to show more clearly how both gender and age alters consumer habits.




The radar chart was able to build a visual profile of how different demographics thought of their experience, utilized features of the platform, and how they chose to shop. This profile shows clear distinctions at a glance, and the interactive component provides the ability to explore how age and gender alters each of these factors both comprehensively and individually.

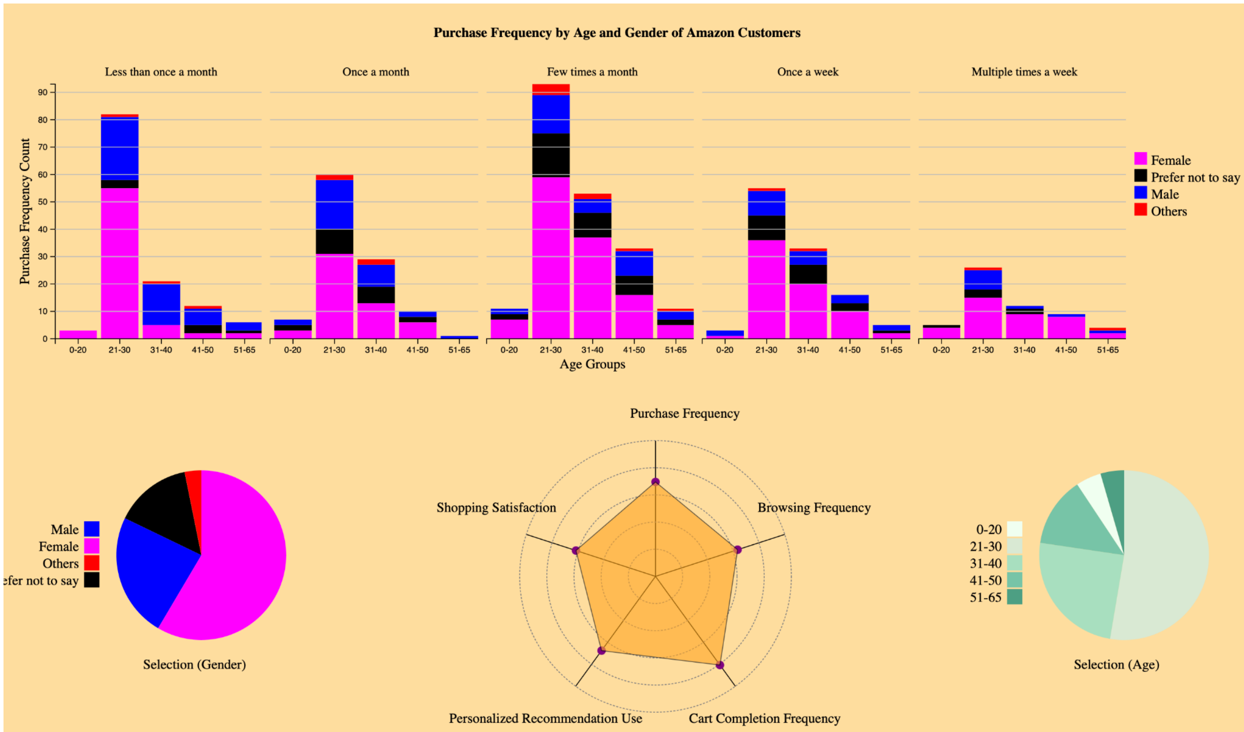


Finally, the updated pie charts correlated much better to the relationships present in the other visualizations, while also telling their own story, presenting a generalized idea of population proportions for each demographic.

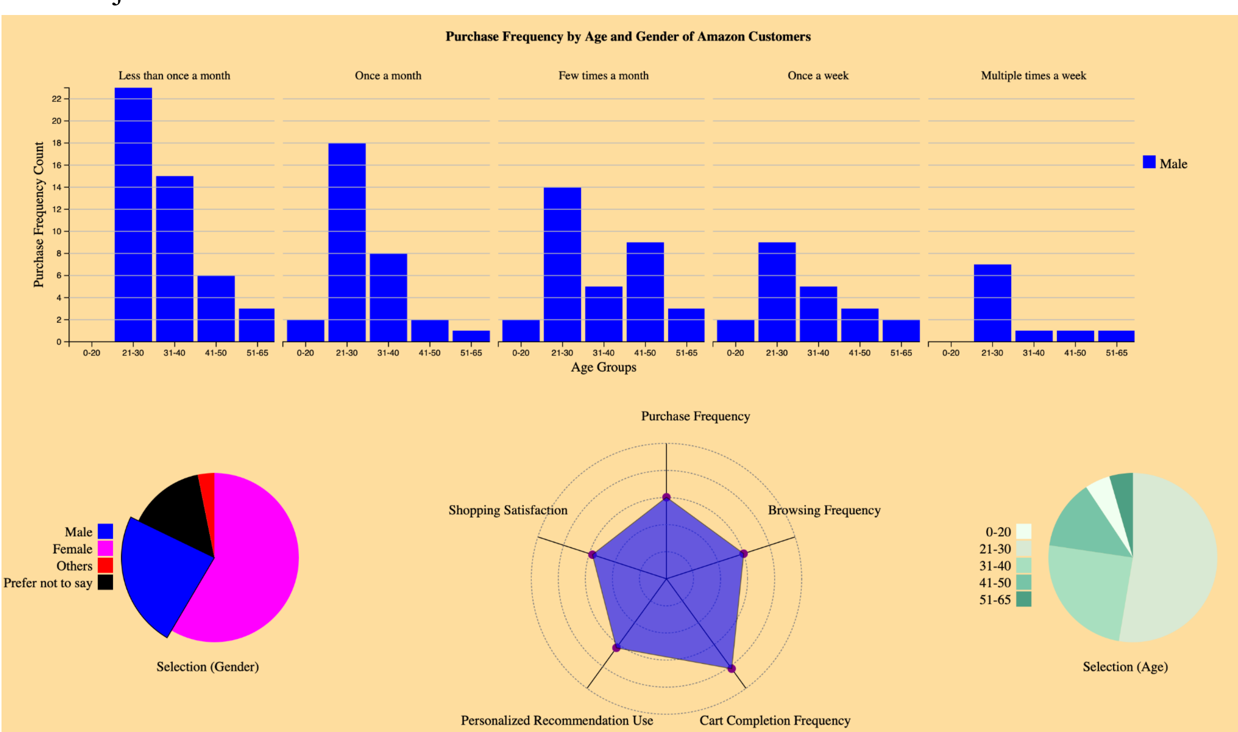



**Implementation:**

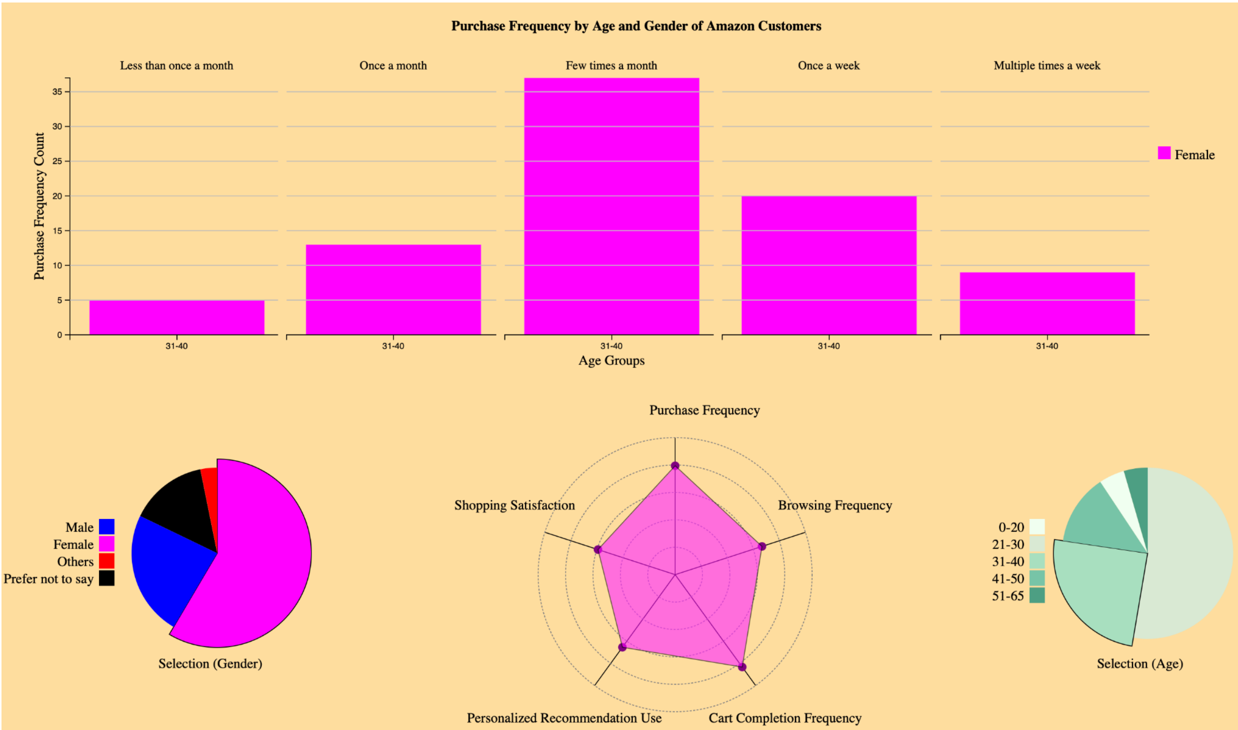
We have 2 pie charts that we are using for filter selection. You can select up to one slice from one or both pie charts and the radar chart and bar chart will be filtered accordingly. We ended up going with these visualizations because we thought they looked the best and provided the most value and insight. We already had experience with pie charts and bar charts from our initial attempts, so we just made them better. The bar chart gives a good comparison between different genders by age group, and the radar visualization is particularly pleasing to look at and easy to understand, although it is a bit more difficult to draw conclusions from. The pie chart filtering the other two visualizations is what really made the other two work so well since they can be filtered at will. We also managed to get the colors to match between the charts, and this made our visualizations significantly more intuitive and pleasing to interact with. With nothing selected in the pie charts for filtering, our final version looks like this:



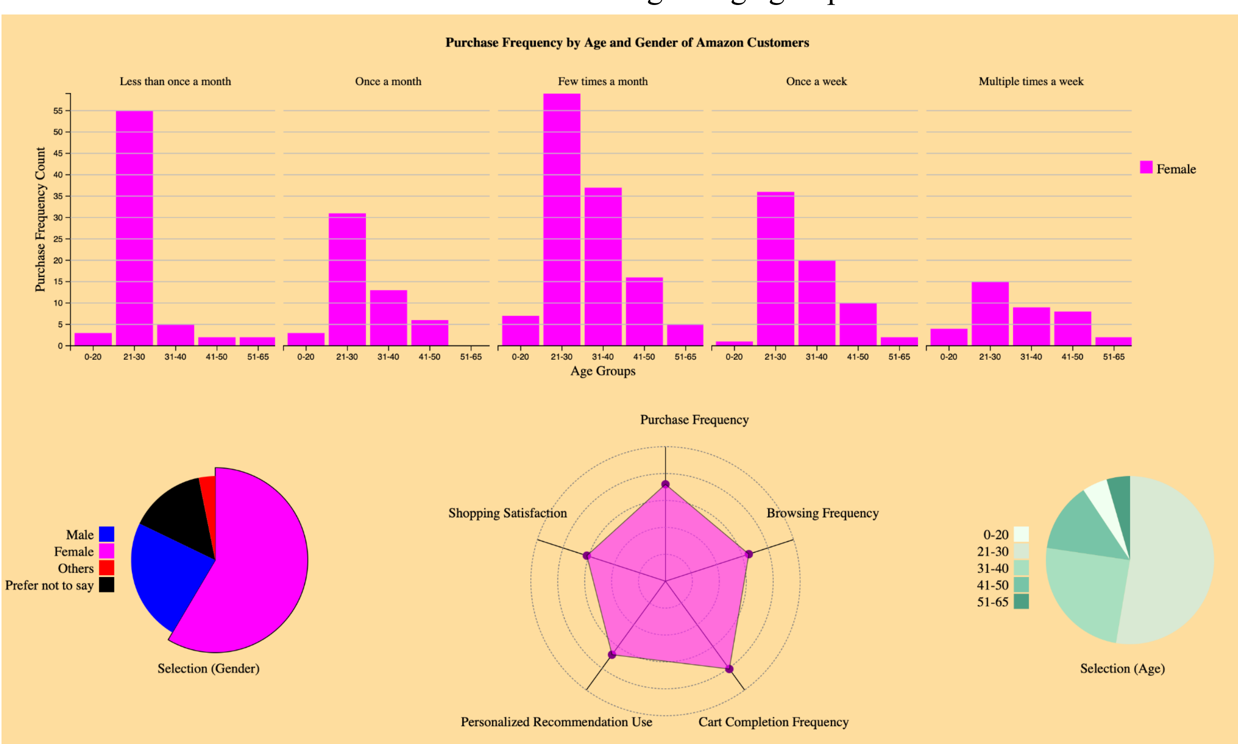

The following examples will be different combinations of filter selections. It will be easy to see which options are chosen because selected pie slices are enlarged and outlined.
Here are just males selected:



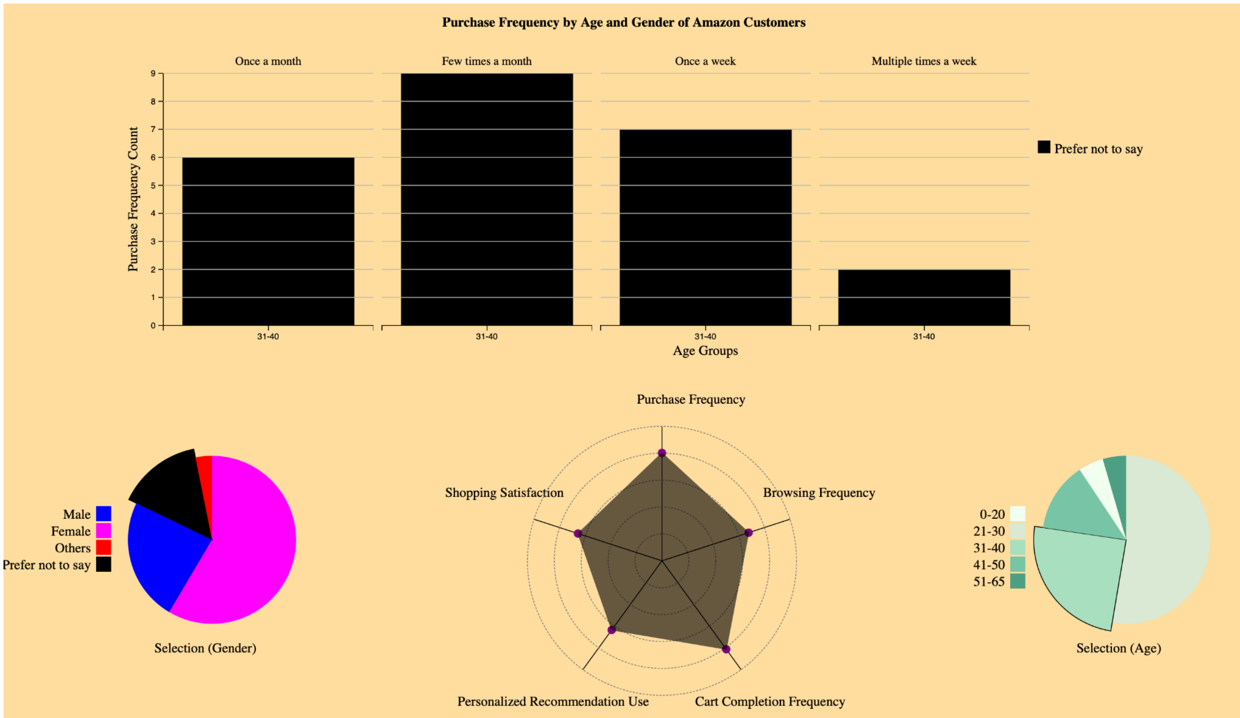
And here are females selected and also filtered for only the 31-40 year old age group.




And here is that same visualization without filtering for age group.



And this is filtering for people who answered with “Prefer not to say” and are in the 31-40 age group.




**Evaluation:**

**What did you learn about the data by using your visualizations?**

We learned that the majority of Amazon users are female. It is about ⅔ female and ⅓ male and others. We learned that the majority of users are between ages 21-30. We also learned that for people aged 31-40, males tend to make purchases alot more frequently than females.

We also discovered that women in the age range of 21-30 exhibit a higher frequency of shopping once a month compared to other age demographics. Finally, we found that males 41-50 browse, make purchases, and use recommendations less than most users, but complete their cart much more frequently on average and are generally less satisfied by their shopping.

**How did you answer your questions? :**

We answered our first question of “What does the age and gender distribution of Amazon purchasers look like, and how do these demographics relate to one another in response to purchase frequency?” with the first visualization of the stacked bar chart. The interactivity component as it relates to this visualization helps to demonstrate the relationship between the demographics, as you can see clear differences in purchasing frequency both by age and gender.

Our second question is answered by the radar chart, which shows a clear profile of each demographic over 5 distinct areas. This chart clearly demonstrates that by age, gender, and combinations of both, that each distinct demographic has its own different ways to interact with Amazon's features. Our final question regarding the overall proportions of demographics is answered with our interactivity component, which allows users to explore the ways in which these groups are different, and how they relate to each other.

**How well does your visualization work, and how could you further improve it?**

Our visualizations work quite well because the filtering with the pie charts is easy and intuitive to use. This causes the two main charts to change accordingly and their colors change as necessary. A small improvement could be that we make the bar chart match whatever shade of green is selected for the age pie filter (assuming no gender is selected). We could also potentially improve our visualizations by allowing multiple filter selections, but this would create a new problem of not knowing which color to make the visualizations if they are a combination of filters. Overall, for what our visualizations aim to do, we think they work very well and are both effective and intuitive.