DS4200 Assignment 3

Anjali Tanna

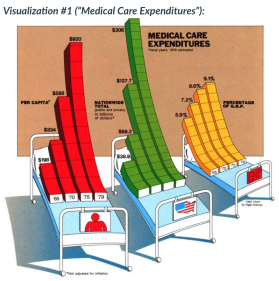
Part 1 - Service-Learning Pre-Assessment Survey

Survey completed:

Part 2 - Design Critique and Redesign

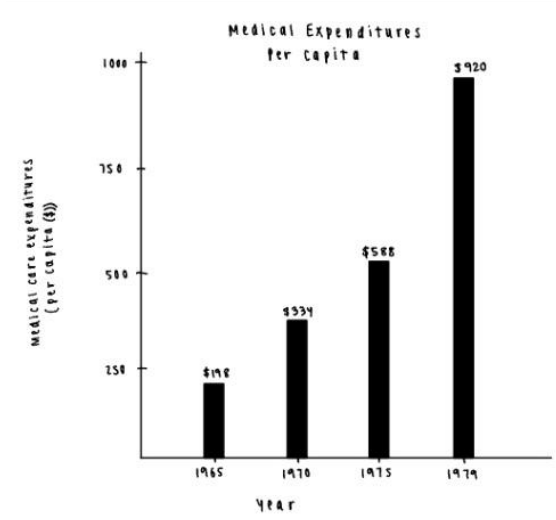
Visualization #1

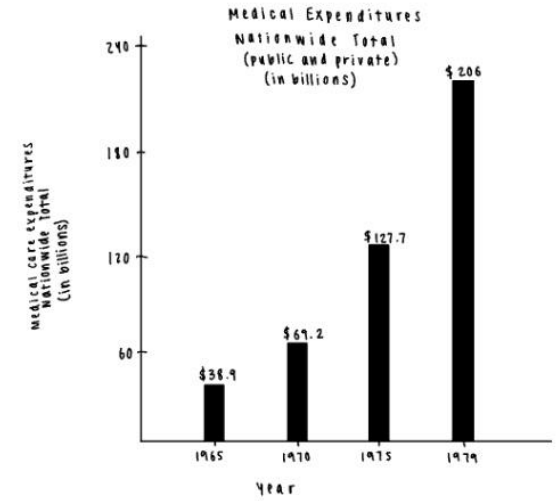
Visualization #1 ("Medical Care Expenditures"):

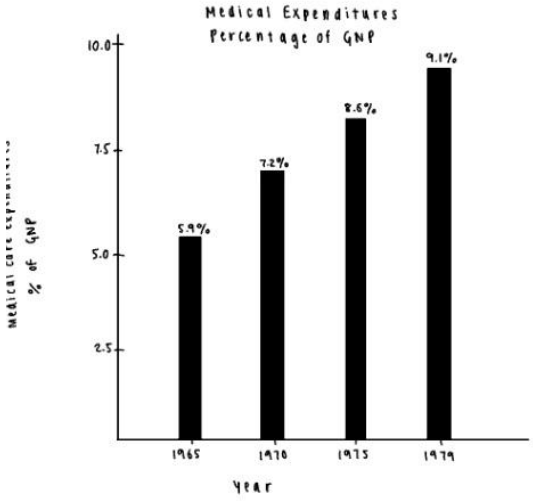


This visualization shows medical care expenditures by displaying three bar graphs in hospital beds. This visualization is not the greatest, as there is no real scale on the y-axis among the three graphs. The marks used in this visualization are lines. The channels used are position (vertical), length, color, tilt, and volume. The type of data is quantitative in both the x and y direction, as the y-axis is dealing with money and the x-axis is dealing with years.. To improve this visualization, the volume aspect could be avoided. It is not really necessary to include the volume for this bar graph. The way the bars tilt and curve is also not necessary in getting the point across. The x-axis is also not completely labeled for all three graphs. The three graphs are comparing three different things, but it is not very clear as the y-axis is labeled in a hidden sort of way. Lastly, the graph should have a consistent y-axis scale or even a more visible one since the same units are not being compared, to see an actual correlation of the data among the three graphs.

Visualization #1 Redesign:

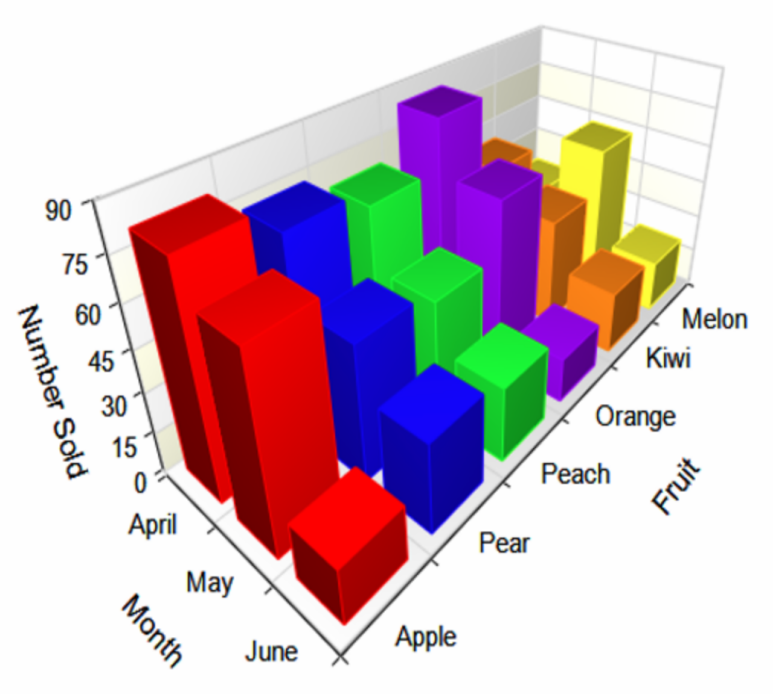






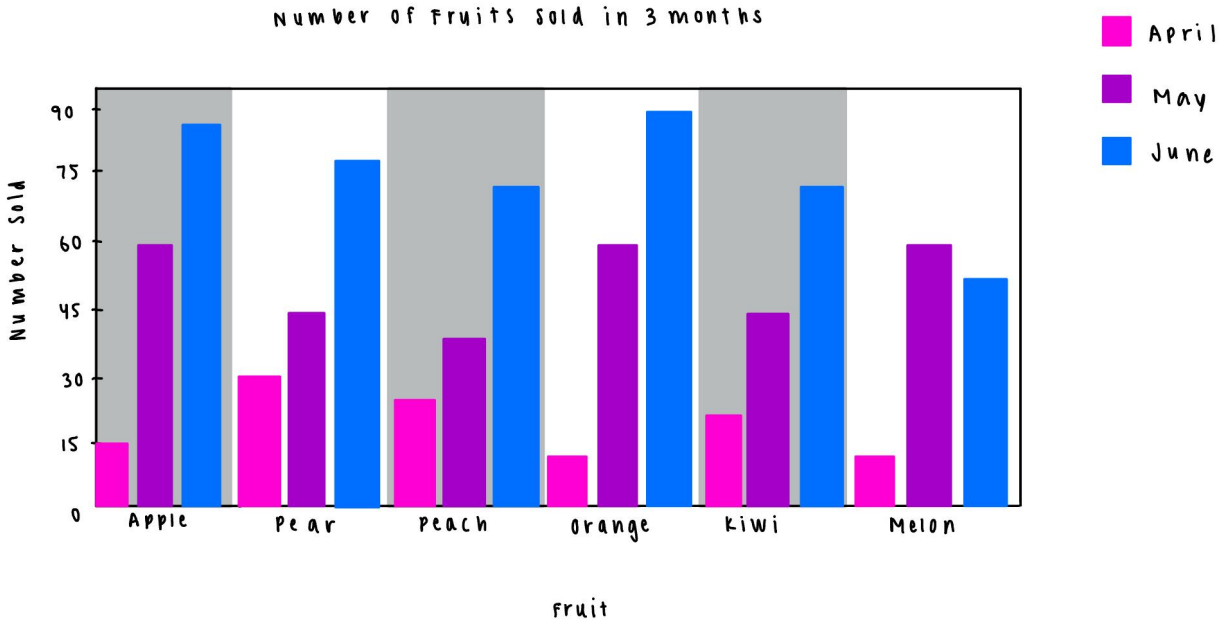
Visualization #2

Visualization #2 ("Number of fruits sold in 3 months"):



This visualization shows the number of fruits sold in the months of April, May, and June. The marks used in this visualization are lines. The channels used in this visualization are both vertical and horizontal position, color for each different type of fruit, size, and volume since it is measuring three variables. These three variables are the month, the fruit, and the number of fruit sold during that month. The number of fruits sold is a quantitative data type. The month is an ordinal data type, as it is qualitative but has some order to it, in this case it is the months. The fruit is a categorical data type as it does not matter the order. This graph is slightly effective in getting its point across, but there could be some changes that would make it easier to understand. To make this visualization better, I would eliminate the volume aspect of the graph. This visualization can be 2D, as the month axis can be changed into just using color, instead of color coding the type of fruit. This way, you can see the individual comparison of fruit sold for each month right next to each other.

Visualization #2 Redesign:



Part 3 - Tableau Practice

1. ***Data loaded into Tableau.***

2. ***Type of Data:***

Submission #	**Quantitative**
What city were you born in?	**Categorical**
Which country?	**Categorical**
What State/Province?	**Categorical**
What is your native (first) language?	**Categorical**
Other than English, what languages are you proficient in?	**Categorical**
Which is your preferred OS?	**Categorical**
Which is your preferred mobile OS?	**Categorical**
What is your favorite/preferred programming	**Categorical**

language? (Just name one)	
Rank your favorite ice cream flavors: [First choice]	**Categorical**
Rank your favorite ice cream flavors: [Second choice]	**Categorical**
Rank your favorite ice cream flavors: [Third choice]	**Categorical**
Favorite color?	**Categorical**
Favorite number?	**Quantitative**
What temperature is it outside right now? (In Fahrenheit)	**Quantitative**
On average, how many cups of coffee do you drink each day?	**Quantitative**
What time do you typically wake-up in the morning?	**Quantitative**
What time do you typically go to bed?	**Quantitative**

3. ***Data explored!***

4. ***Data Cleaning***

To clean this data, I used Excel to fix any inconsistencies and missing data. I first removed any rows that were completely “Null”. These submissions were likely mistakes, therefore it did not make sense to include them in the data set. The next step I took was to mark any empty cells as “Null”. This was done so that Tableau could easily identify any cells without a value. Next, I edited all of the cells so that they were formatted in the same way. For example, in the “Which country?” cell, there were multiple of the same answer, however they were written differently. There were different variations such as “United States”, “USA”, and “United States of America”. By changing these answers all to the same format, Tableau would have an easier time analyzing the data.

Another column I reformatted was the programming language column, by making sure all answers were written the same (ex: Python vs python). Since I explored a visualization that dealt with a world map, I made sure that each of the country, state/province, and city columns were filled out correctly so that Tableau could recognize the location on a map.

The last thing I fixed was some of the timestamps that were evidently errors. In some cases, a submission for the question “What time do you wake up?” would

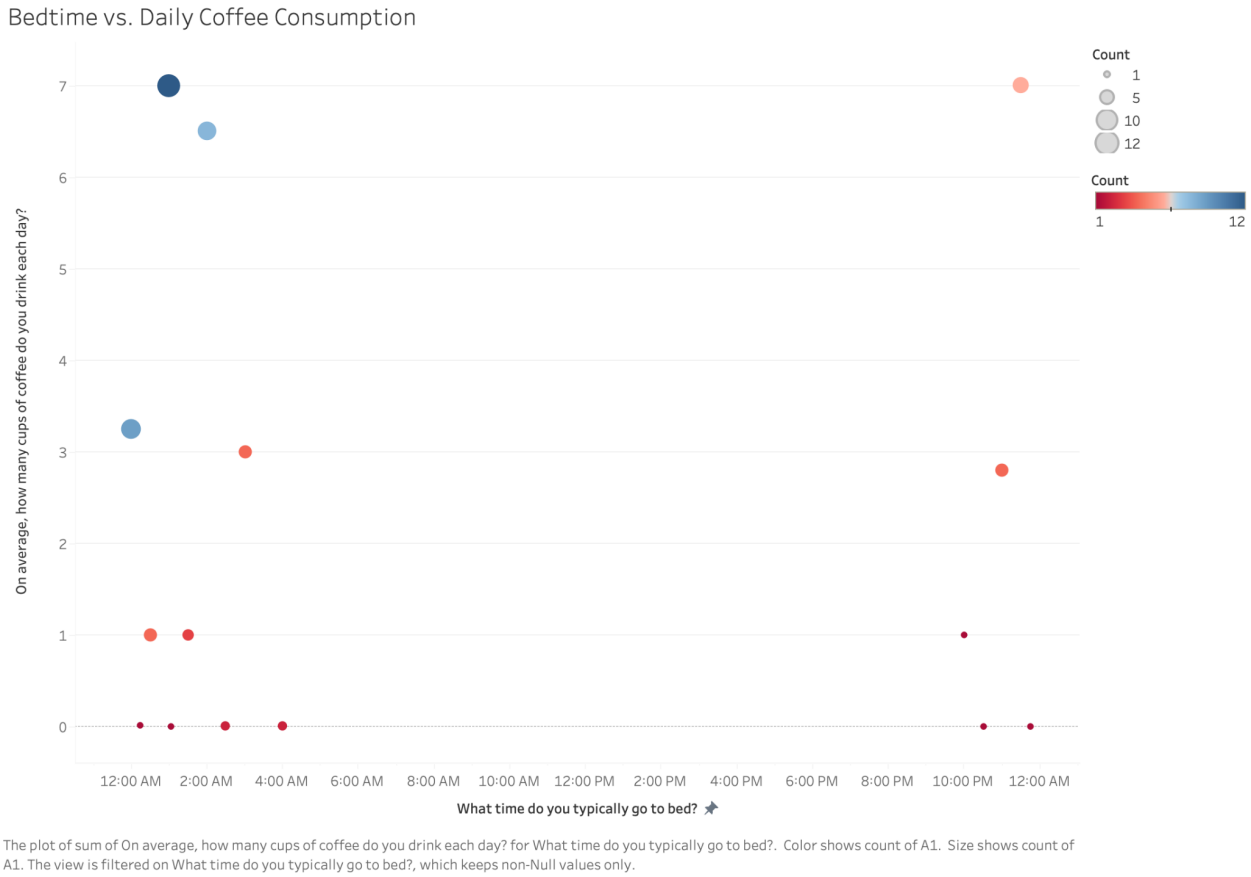
have the answer 8:00am, but then their answer for the question “What time do you go to bed?” would be 10:00am. This obviously does not make sense, and was causing outliers in the data, so I edited the 10:00am to reflect 10:00pm. The person who filled out the form likely made an error. My justification for the cleaning of the data I performed, is to provide optimal readability and analyzation for the dataset.

5. ***Clean data explored!***

6. ***Questions for the dataset:***
    a. Is there a correlation between bedtime and coffee consumption?
    b. Are there any correlations between preferred OS and preferred mobile OS?
    c. Is there a correlation between birth place (city, state/province, country) and native language?
    d. What is the most popular favorite color?
    e. What are the most popular first, second, and third choice favorite ice cream flavors?

7. ***Visualizations***

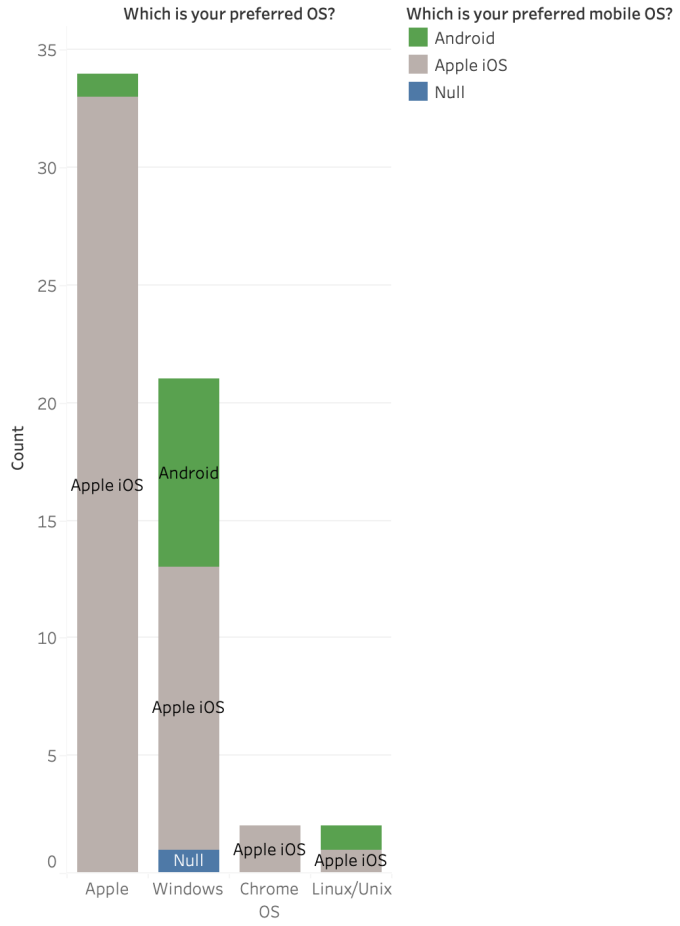
Is there a correlation between bedtime and coffee consumption?



Based on this visualization, there is a correlation between bedtime and coffee consumption. The graph depicts bedtime on the x-axis and the average amount of cups consumed of coffee on a daily basis, on the y-axis. Looking at the times from 12:00AM to 4:00AM, you can see the amount of coffee consumed is significantly higher, as there are 12 people who go to bed at 2:00AM and drink 7 cups of coffee a day. If you look over the times around 10:00PM to 12:00AM, there are less people who drink as much coffee. This visualization clearly shows that those who go to bed earlier tend to drink less coffee than those who go to sleep later. I chose to visualize the data in this way because it allows for users to see a clear timeline of bedtimes, as well as cups of coffee consumed. By using size and color, it is also easy to see how many people have consumed the number of cups of coffee at that particular time. The marks that were used are points and the channels used are position (both vertical and horizontal), color, and size.

Are there any correlations between preferred OS and preferred mobile OS?

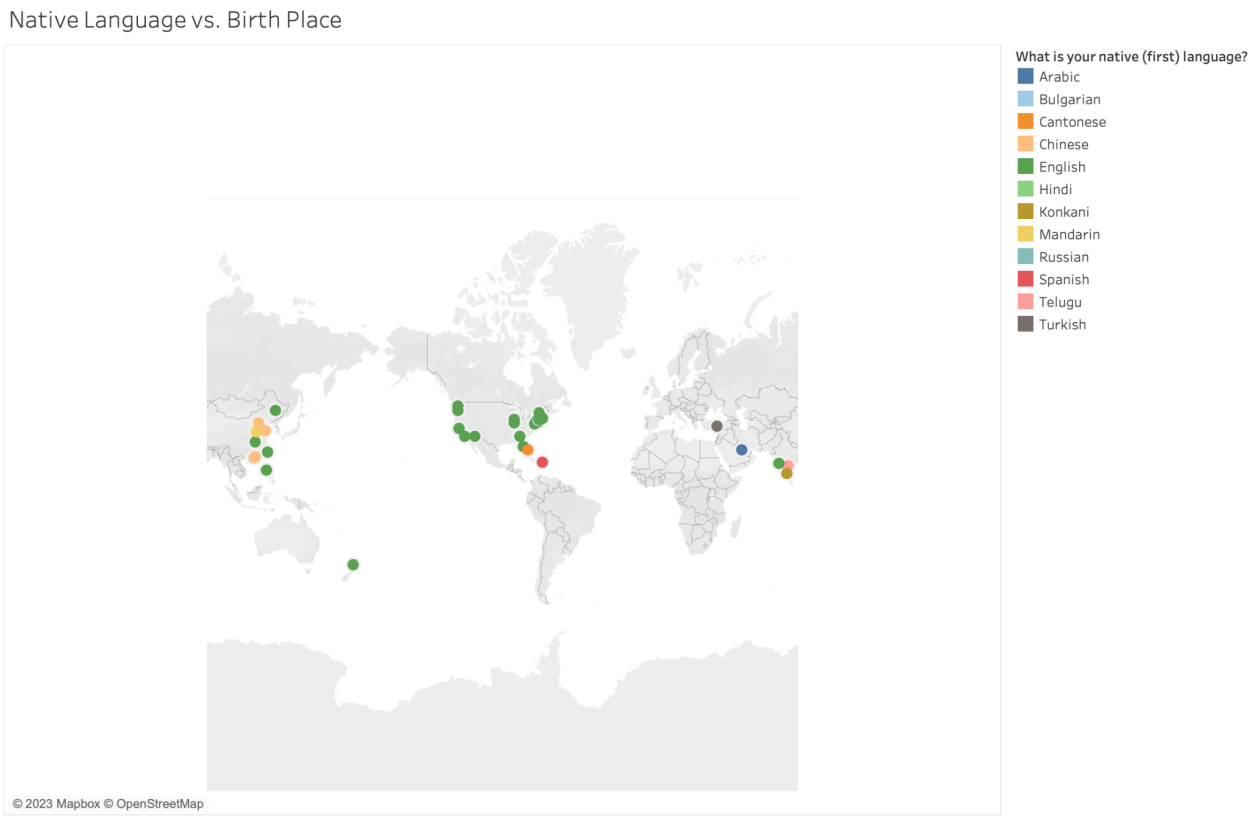
Preferred OS vs. Preferred Mobile OS



Count of A1 for each Which is your preferred OS?. Color shows details about Which is your preferred mobile OS?. The marks are labeled by Which is your preferred mobile OS?. The view is filtered on Exclusions (Which is your preferred mobile OS?,Which is your preferred OS?), which keeps 8 members.

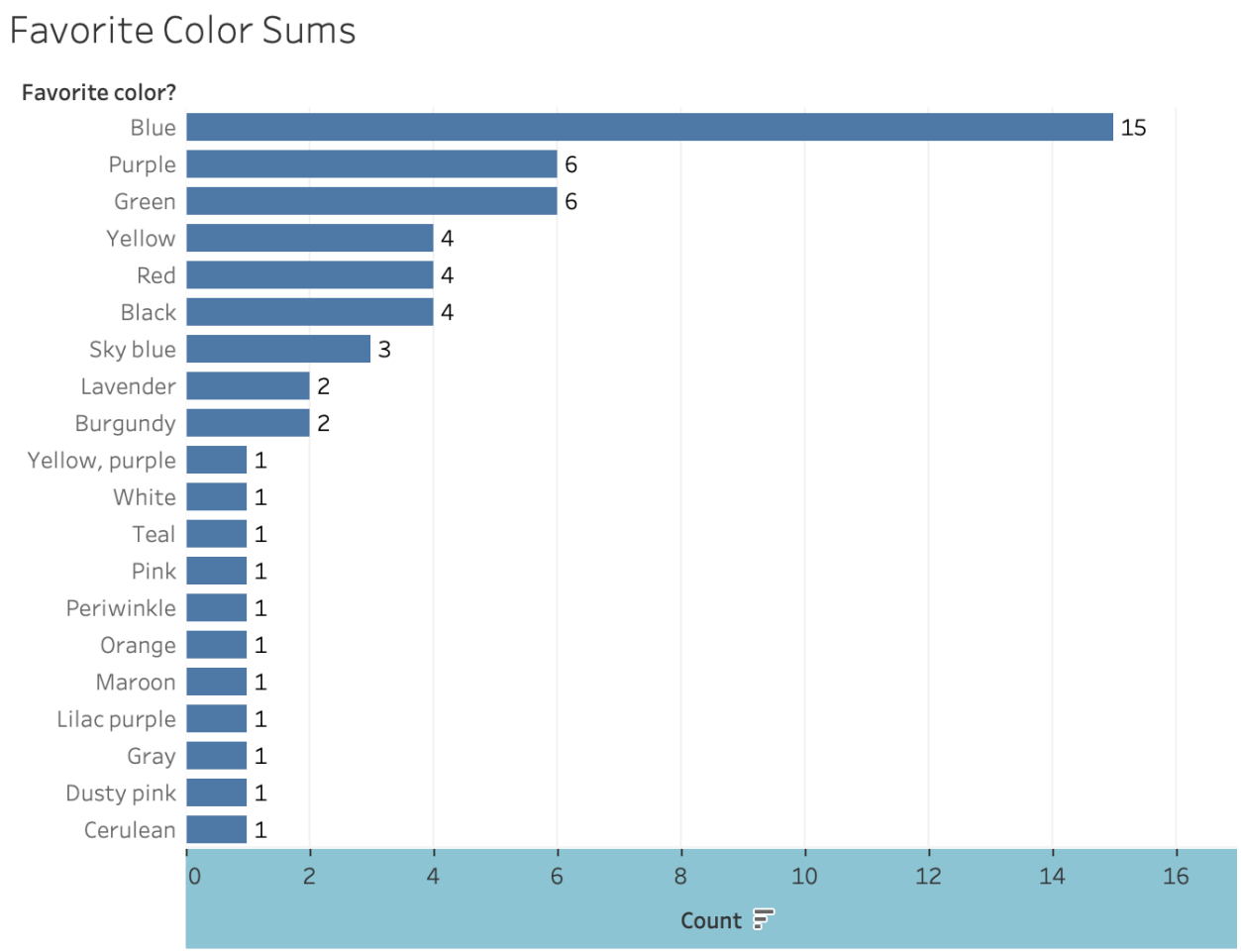
Based on this visualization, there is a slight correlation between preferred OS and preferred mobile OS. Looking at the graph, it is clear that Apple OS users prefer to use Apple iOS. However, it seems that Windows OS users seem to be a close to even split between Apple iOS and Android. It is clear based on the amount of gray that Apple iOS is a clear favorite. I chose to use a stacked bar graph because it displays the total number of people who use a respective OS while still displaying the differences in counts for preferred mobile OS. The marks used in this graph are lines and the channels are length, position, and color.

Is there a correlation between birth place (city, state/province, country) and native language?



Based on the visualization, there is not really a correlation between birth place and native language. While some people born in East Asia speak languages like Mandarin, many still speak English as their native language as well. In the United States, English seems to be the most common native language. I chose to display the graph this way so it would be easy to see which regions have a common native language. The marks used in this graph are areas and the channels used are position, color, and area.

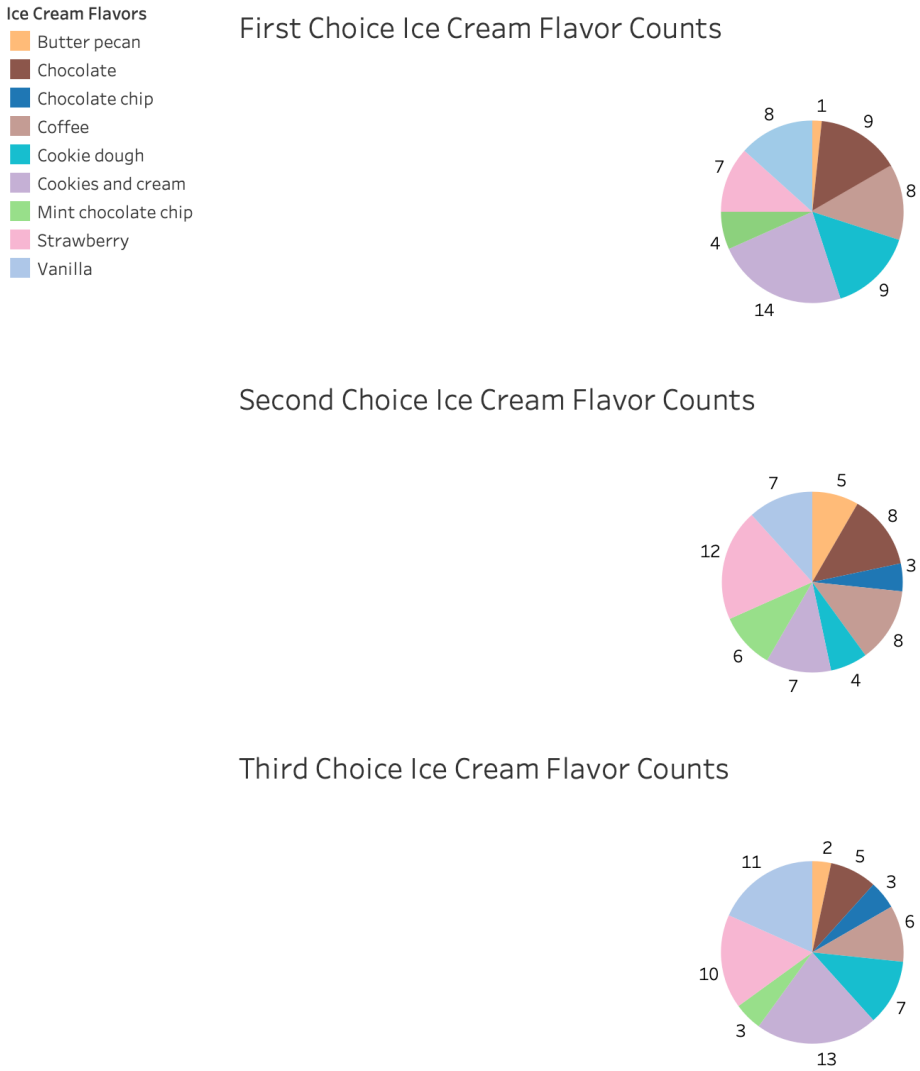
What is the most popular favorite color?



Count of A1 for each Favorite color?. The marks are labeled by count of A1. The view is filtered on Favorite color?, which excludes Null.

Based on this visualization, the most popular favorite color is blue. Using a bar graph, it is easy to see which color has the highest count. In this case, blue is the highest with 15 people having it as their favorite color. I chose to display the data in this way because it is clear to see based on the marks and channels which color is the favorite among the favorites. The marks used in this visualization are lines. The channels used are position and length.

What are the most popular first, second, and third choice favorite ice cream flavors?



Based on this visualization, the most common first favorite ice cream flavor is cookies and cream. The most common second favorite flavor is strawberry. The most common third favorite flavor is also cookies and cream. I chose to use pie charts to represent this data because it would be easy to tell which flavor was the favorite due to the label and the legend. The mark used in this graph is area. The channel used is area.

Part 4 - Practice Task Abstraction

5 Domain Tasks

- Where is there street parking available?
- Where are the trees located?
- What is the demographic of the neighborhood?
- What types of trash is thrown around?
- How many calls are being placed about the litter in the area?

Index (ID #)	“Domain” Task	Analytic Task (Low-level, “Query”)	Search Task (Mid-level)	Analyze Task (High-level)
1	Where is there street parking available?	Retrieve value	Locate	Discover
2	Where are the trees located?	Retrieve value	Locate	Discover
3	What are the demographics of different regions of the neighborhood?	Cluster	Browse	Present
4	What types of trash is thrown around?	Filter,	Browse	Annotate
5	How many calls are being placed about the litter in the area?	Compute derived value	Lookup	Derive