Final Project Implementation
CPSC 5320 01 23SQ Visual Analytics

Group 2 : Ankita Kadam, Sartaj Bhuvaji , Jay Singhvi , Aditya Gollapalli and Madhuroopa Irukulla

Waffle Chart:

Data Transformation :

Serving size considered is 240 ml(milk) and 35-40g(cereals), By calculating the proportions of carbs, fiber, protein, etc in relation to the sum of serving sizes of respective milk or cereal types, we obtain a standardized measure that allows us to compare the relative contribution of these nutrients across different cereals and milk types.

Design Rationale :

1. Each cell corresponds to a specific proportion, and the quantity of cells assigned to a nutrient reflects its percentage, allowing for an intuitive visual comparison between nutrient proportions.
2. The waffle chart design incorporates different colors for each nutrient. Color coding enhances visual distinction and aids in quickly identifying and associating specific nutrients.
3. Dynamic waffle generation is done using 2 dropdown menu selections which allows the user to view proportions of each nutrient type in each product type.
4. Dynamic waffle creates proportions for each box as a total of all product types for that nutrient.
5. We used a dynamic legend, which is changed according to the drop down values chosen
6. Dynamic legend also allows the users to navigate to radar charts to view more details and compare further

Visual Encodings:

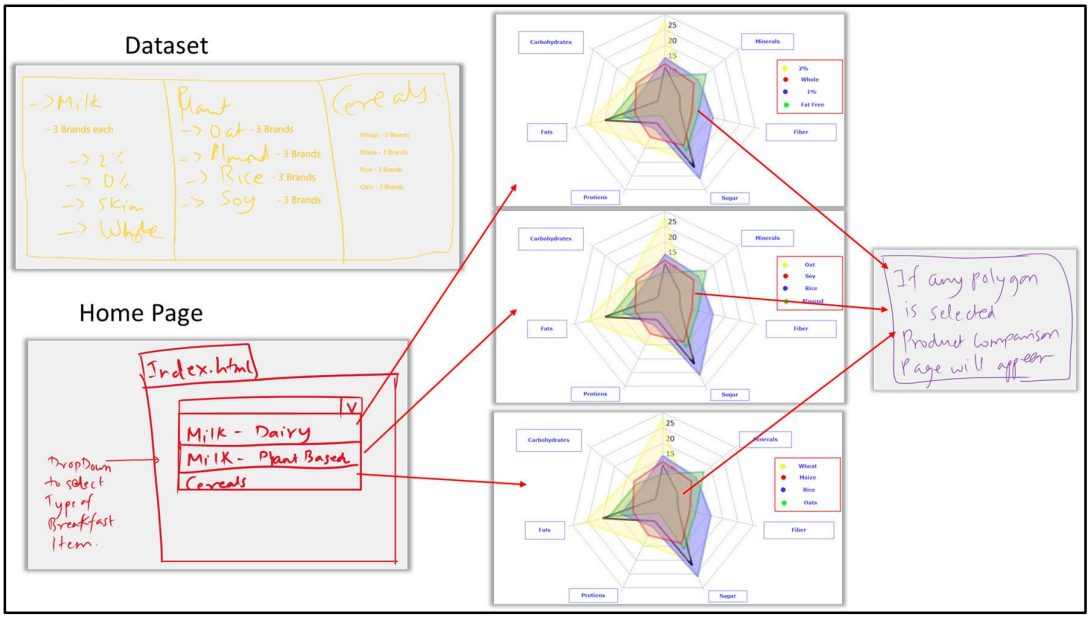
1. Cell Color: Each cell is filled with a color that represents a specific breakfast item. This visual encoding helps to distinguish between different items and enables easy identification and comparison.
2. Opacity: The opacity of each cell is set to 0.8, providing a slightly transparent appearance. This allows for a sense of depth and layering when multiple cells overlap. Additionally, when the user hovers over a cell, the opacity of other cells with the same breakfast item is reduced to 0.2, emphasizing the selected item.
3. Tooltip: When hovering over a cell, a tooltip is displayed with additional information about the specific breakfast item and its percentage. The tooltip provides a textual encoding that complements the visual encoding of the chart. It offers a precise value and enhances the understandability of the data.
4. Legend: Legend is included to provide a key for interpreting the colors used in the chart. The legend acts as a visual encoding guide, aiding in the identification and interpretation of the chart elements.

Waffle Chart Challenges :

1. Data Standardization and calculating the proportions was challenging.
2. Generating a dynamic legend for the waffle chart that gets displayed from drop down selection was a little harder on the implementation side.

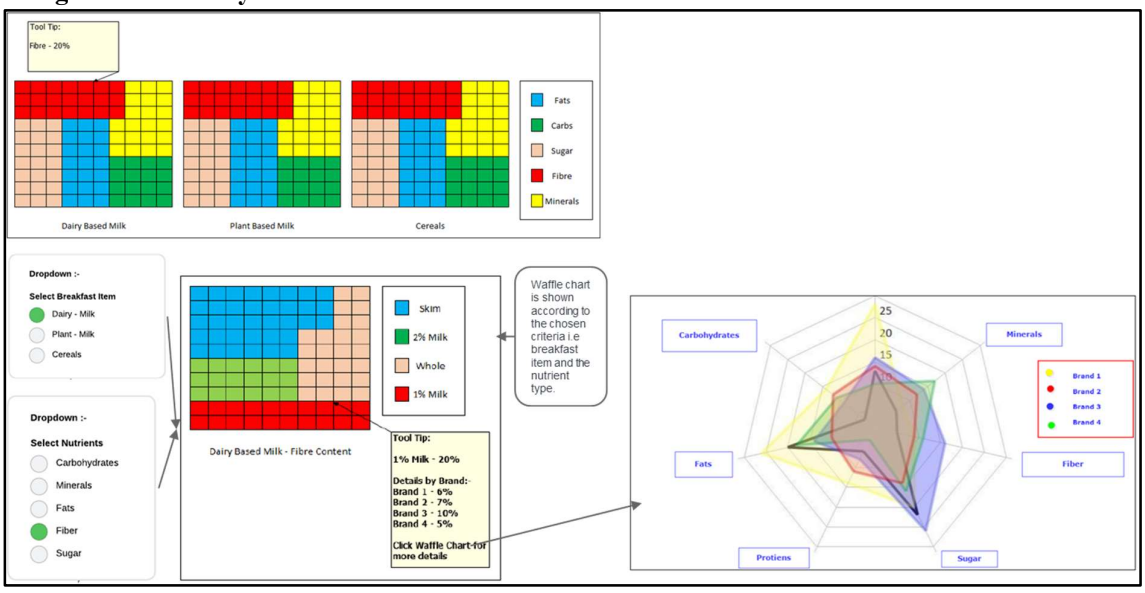
Design Iterations for Waffle chart:

Design Phase 1 Story Board:



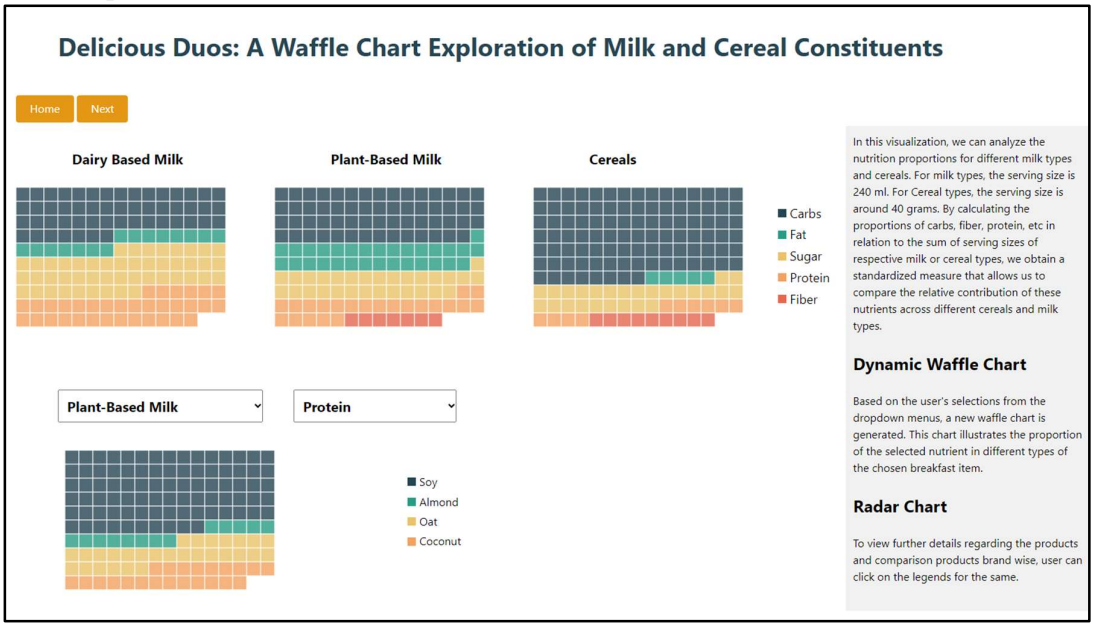
Initially in Phase 1 the waffle charts were not a part of the design phase, but however we saw our project needed more visual representation of details and information to help visualize the information. So, after getting the feedback from professor and peers regarding design phase 1, we decided to search for more ways to help with the same.

Design Phase 2 Story Board:



In phase 2 we changed the initial dropdowns from showing radar chart directly to waffle chart which will show more details about each product(i.e Dairy, Milk, Cereals) and radar to be available on drill down as it will show more details there.This will also help give us bird's eye view as well as worm's eye view of the data.

Final Implementation:



A user intuitive interface which is simple to use as well as visually appealing from the 1st instance of loading. Along with a brief description of the visualization as well as its usage and possible actions that can be taken.

Radar Chart:

Data Transformation :

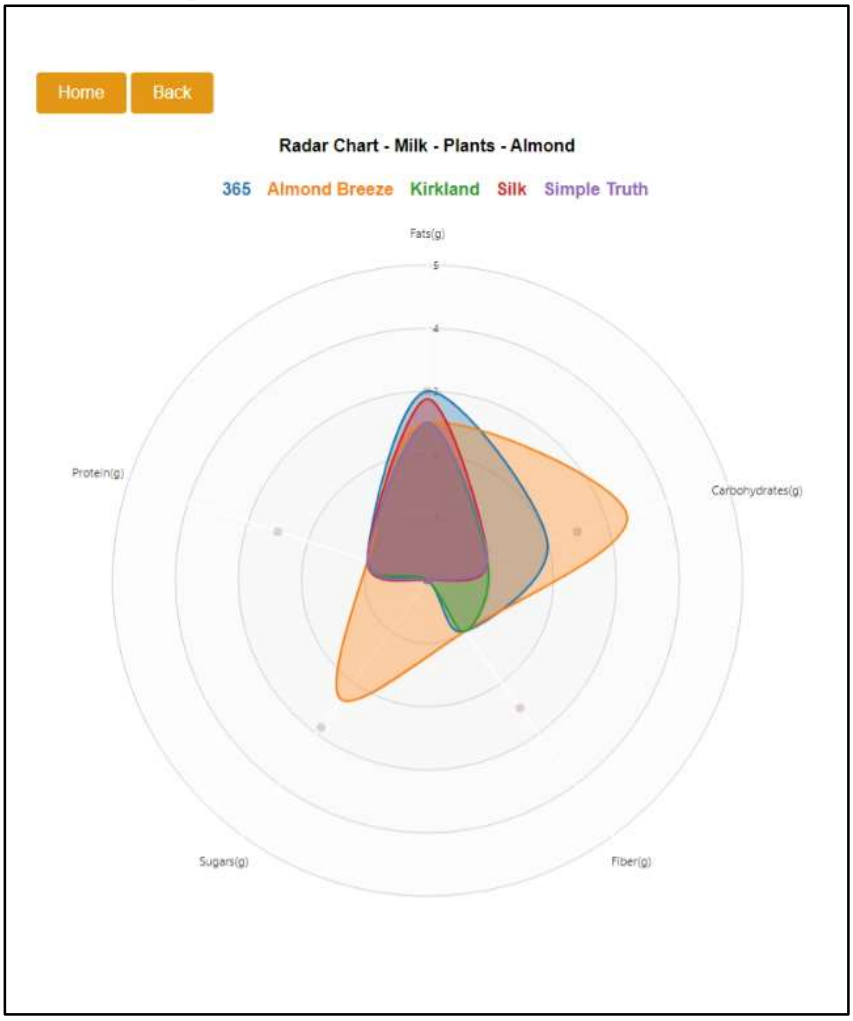
Serving size considered is 240 ml(milk) and 35-40g(cereals), By calculating the proportions of carbs, fiber, protein, etc in relation to the sum of serving sizes of respective milk or cereal types, we obtain a standardized measure that allows us to compare the relative contribution of these nutrients across different cereals and milk types.

Design Rationale :

1. Radar creates blobs polygons for each brand depending on the waffle legend selection in the previous page
2. Each spoke of the radar signifies a different nutrient and each circle on these spokes represent the value in grams, larger the circle higer the value.
3. Legends are generated dynamically too, depending on the data, i.e.if certain product type has a list of brands only those brands will appear, not all.
4. Tool tips will appear on hover of the polygon as well as the legend
5. Polygons will change their color and opacity each time the mouse hovers over them or the legends for better viewing

Visual Encodings:

1. We used colors to identfy the polygons/blobs.
2. Colors used are not gradient because the brands are not being considered as comparative directly.
3. We used same colors for hover on the dynamic legends as well as the polygons to maintain consistency.
4. Tooltip data also uses same colors as the polygons to further maintain the consistency in visual representation



Radar Chart Challenges:

1. Main challenge was with the blobs being mostly overlapping eachother this was due to the nature of data.
2. Data / Nutrients are very similar but not same so many times the visuals were not pretty enough to view but always had just noticeable difference which can be seen in all radar charts.
3. But some brands could be less in all aspects from other brands which could lead to an issue wherein the users are not able to hover over the polygon to view the data/tooltip.
4. To overcome this issue, I have added tooltip and polygon view changes on legends as well
5. This helps identify the differences as well as the data.

Stacked Bar Charts:

Data Transformations:

We merged the data from different data csvs that we collected. We had to transpose the data to fit my script requirements. For the cost, we had to scale the cost from per serving to a 30-day serving.

Reasons for changing from Circular Packing:

The professor's review helped us understand that circular packing charts would not be suitable in this case and bar charts would better represent the visualization. Switching to a stacked bar chart was also helpful as we were able to show the effect of milk and cereal in nutrient values and cost separately.

Design Rationale :

1. The stacked bar chart is helpful as it provides the total constituents of various nutrients in the meal. This helps us highlight the number of nutrients added to milk and cereal separately. Also, we decided to flip the graph to make it horizontal as it helped in adding a subplot of the cost below the nutrients.
2. The cost was scaled for 30 days to better help with comparing the monthly cost incurred on choosing a particular type of cereal and milk combination.
3. The toggle button is really helpful as the user can compare two breakfast meals and values associated with the nutrients and the cost side by side. However, if the user does not want to compare, he can toggle off the button, and the visualization changes to just one portion.
4. The visualization is flipped horizontally, and the Y axis represents different nutrients and the X axis represents their value in grams for the 1st graph. The X-axis represents the cost in dollars for the 2nd graph below the original visualization.

Visual Encoding:

We used a tooltip that displays the nutrient value and the cost value on hover. In a stacked bar chart, for the bars away from the axis, it might sometimes get difficult to get the exact values of the graph, thus adding a tooltip to display the bar value would help the user make an informed decision. Adding a toggle button helps reduce clutter. If the user wants to view visualization for just one meal type, they can turn off the toggle and would not be distracted from the comparison charts. On the other hand, if the user wanted to compare values with different cereal and milk combinations, the user can easily do so by turning the toggle on. Two different colors were used, one for cereal and one for milk. We made sure that those colors did not blend together and were appropriate for the data.

Tradeoffs and Decisions:

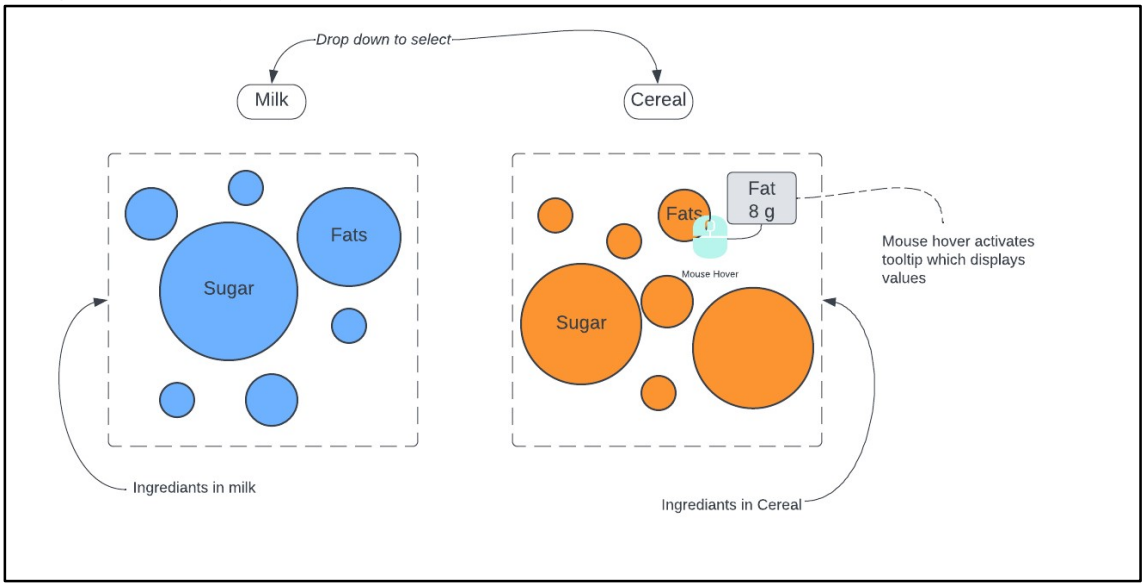
We had collected a lot of data and including 50+ options to filter from the two drop-down menus would not have been preferable. Thus we decided that we would use three of the most popular stores. Even with this, we had lots of options, this might overwhelm some users and might reduce the impact of the visualization as users would try each and every option.

Future Scope:

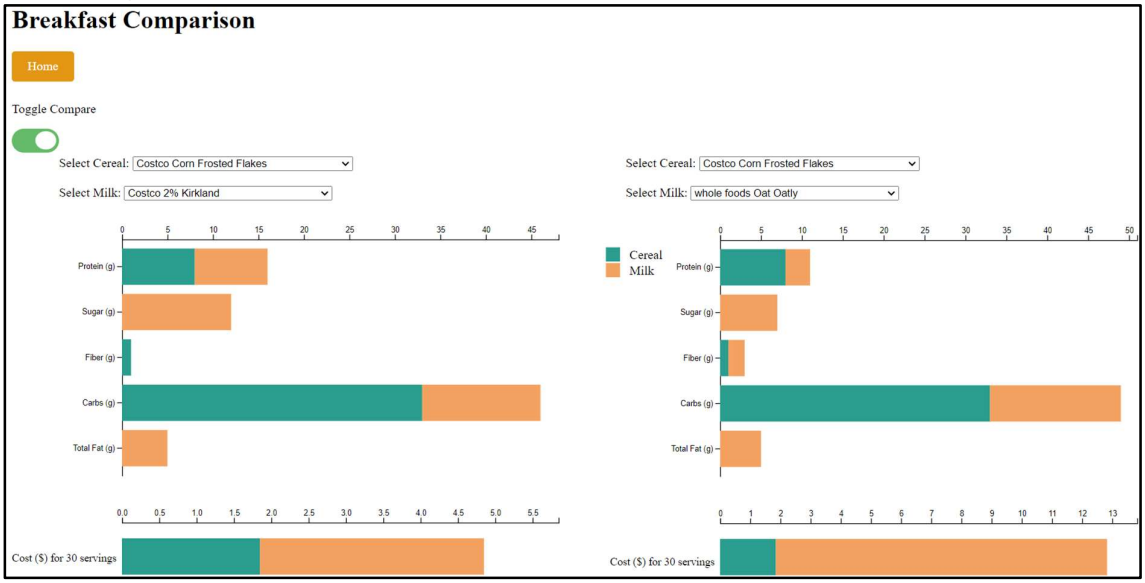
We can try building a paired stacked bar chart and visualize the data. When comparing the two meal options, the horizontal distance between the values prompts the user to make a lot of vision corrections while analyzing multiple values. This can thus be avoided. Further, we could also include a search filter so that users can search for a milk and cereal type instead of a long dropdown.

Design Iterations for Stacked Bar Plot:

Design Phase Story Board:



Final Design :



Dot Plot:

Data Transformation :

The average serving size for breakfast consists of 1 cup of milk (approximately 240 ml) and 1 cup of cereal (around 40 grams). Since milk is typically sold in gallons at the store, it needs to be converted to milliliters for accurate measurement. Additionally, the price obtained for the whole can of milk needs to be scaled down to match the 240 ml serving size.

Similarly, cereal boxes are usually measured in ounces, so they need to be converted to grams for precise measurement. The price obtained for the whole box of cereal must also be adjusted to reflect the 40 grams serving size.

Considering that Costco operates as a wholesaler, the milk and cereal are available in multiple quantities. To ensure equal distribution among all stores, the price is scaled down for one quantity.

Reason to change from the Design Phase:

Initially, during the design phase, we planned to use a scatter plot. However, after preprocessing the data, we found that there were limited data points for each category, which meant that a scatter plot would not effectively communicate our findings. Consequently, we decided on a dot plot, which allowed us to better represent our data and achieve our aim of comparing the cost of breakfast per serving at different stores.

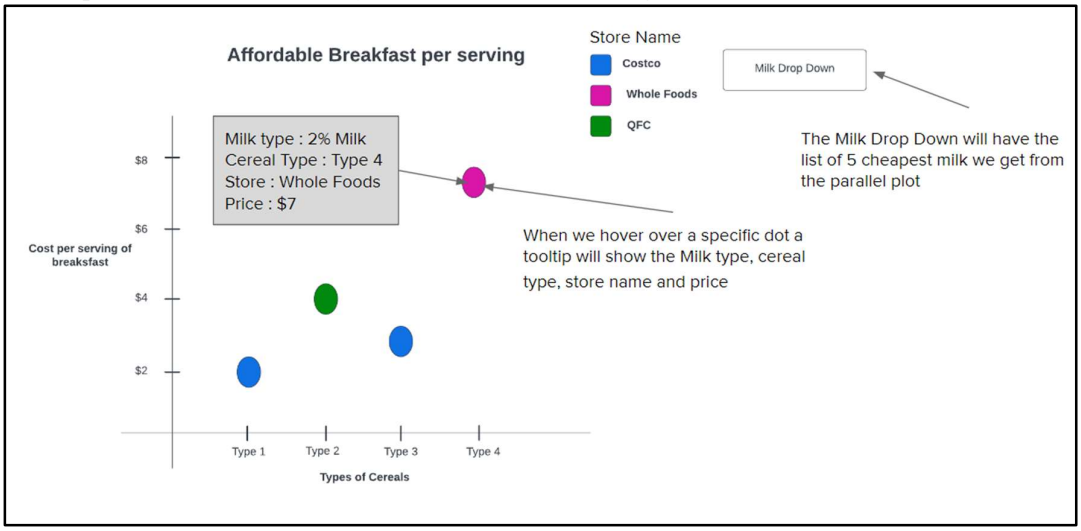
Design Rationale :

1. We used a dot plot to compare the cost of breakfast per serving. The breakfast includes 1 cup of milk (approximately 240 ml) and 1 cup of cereal (around 40 grams).
2. From our prior visualizations, we found that Whole Milk, Low Fat Milk, and Reduced Fat Milk are the three cheapest options. You can use the drop-down menu to compare the cost of each cereal with these milk options.
3. We compared the cost per serving of different cereals, such as Raisin Bran, Cheerios, Fruit Loops, and Frosted Flakes.
4. For the visual encoding, we opted to use colors that correspond to the logo colors of the respective stores. This was done to provide an intuitive link between the stores and their representation in the plot. We made an exception with Fred Mayer, assigning it a unique color, due to its logo color being similar to that of Costco's to prevent confusion.
5. Each dot on the plot represents a different breakfast option at a given store, allowing for easy comparison
6. The visualization aims to determine the most affordable store for breakfast. We compared wholesale stores like Costco and Target, as well as retail outlets like QFC, Whole Foods, and Fred Meyer.
7. The interactive legend helps compare different breakfast combinations within the same store to find the best option.
8. Tooltips have been added to provide additional information such as milk and cereal type, store name, and cost per serving when hovering over specific data points
9. This “Comparative Analysis” helps individuals on a tight budget in finding the store with the most economical breakfast prices.

Design Iterations for Dot Plot:

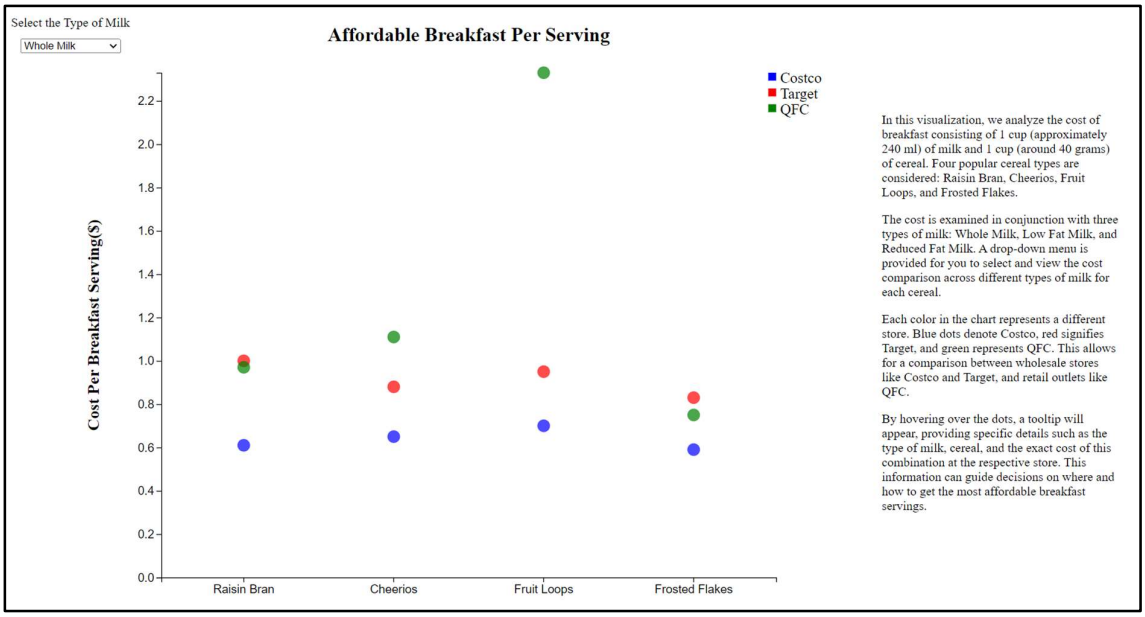
Design Phase Story Board:

The scatter plot was not used as we had limited data points for each category, which meant that a scatter plot would not effectively communicate our findings



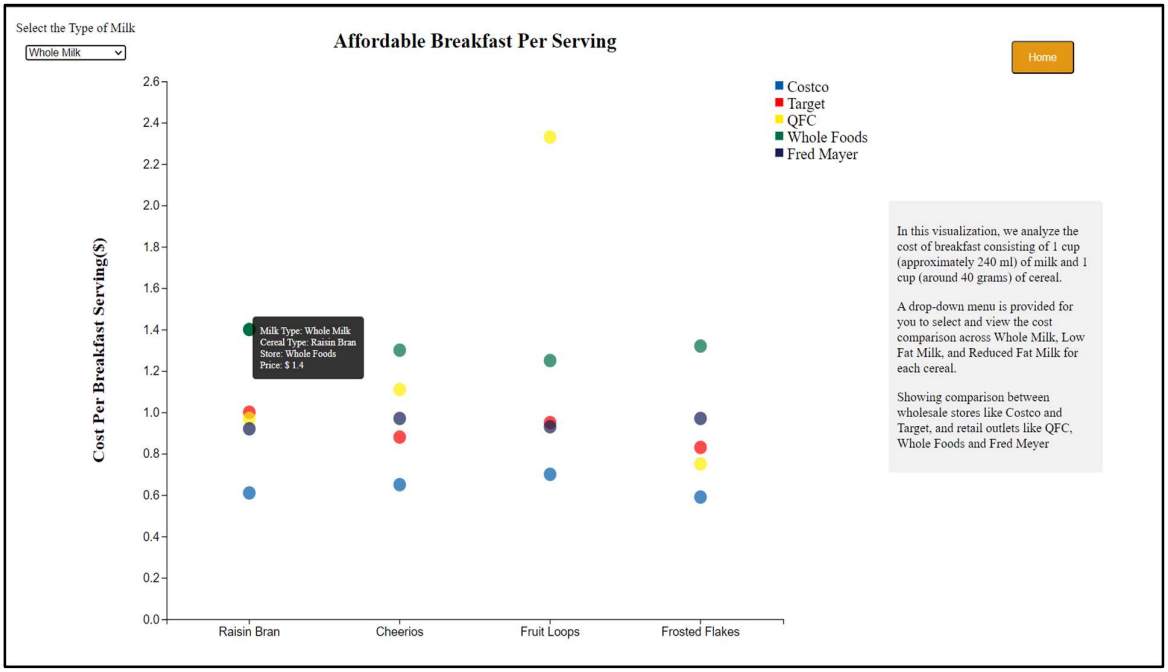
Design Iterations 1 :

Initially, we examined three stores, but for a broader comparison, we included two more: Whole Foods and Fred Mayer. The original color scheme lacked clarity, so we modified it for better differentiation. To streamline the visualization, excessive textual content was also reduced.



Final Design :

In the last version of our design, we looked at five different stores. Each store got its own color on the chart, matching the color of its logo. The only exception was Fred Mayer; we used a different color for it because its logo color was the same as Costco's. We made the chart easier to read by using less text. Plus, we added a Home button so that you can quickly go back to the main page.



Lollipop Chart

Data Transformations:

We collected the dataset from the BBC report and the data was clean. I merged the dataset for different environmental factors using Excel.

Reason to change from the Design Phase:

We initially used a Treemap, but it wasn't clear enough because Treemaps are best for showing things in a sort of family tree way. So, we changed to using a lollipop chart that you can interact with. This made the data much easier to see and understand.

Design Rationale:

The lollipop chart helps us visualize the environmental impact of a glass of milk. For this,, we chose to build a lollipop chart as it is great at showing values and also the difference in values for different labels. This helps us give a numerical value to the environmental impact and highlight how low the impact is for another type of milk.

Visual Encoding:

The chart has three buttons for visualizing different environmental impacts. The Y-axis scale changes according to the button selected and the unit on Y axis changes. The X-axis represents different types of milk and they are static

Tradeoffs and Decisions:

Deciding on the type of visualization was difficult. We could have used Barchart or s scatter plot, but by using the height of bubbles in the chart we have a better understanding of the difference in values of different types of milk.

Design Iterations for Lollipop Plot:

Design Phase Story Board:

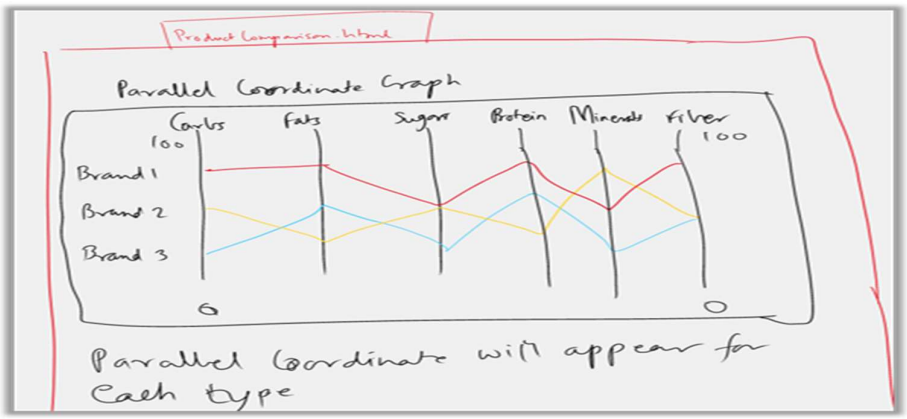
Final Design :

Parallel Plot:

Design Rationale for the Parallel Plot:

Initially, the visualization type that was chosen to represent the different milks and cereals with an adjacent price tag sorted by store and brand, along with nutrition values was a parallel coordinate plot. The parallel plot is a chart that can help us visualize multiple attributes of a product, which is our primary focus and compare based on its various variables associated with it. This is a perfect representation for data that is multi-variate. This was the initial rationale.
From this chart we could gather the cheapest milk/cereal by type/brand so that it could be utilized in the next chart for breakfast affordability comparison (Dot Plot).

The following is the proposed design (StoryBoard):



t

Explanation of the storyboard (Initially):

The keys that are on the Y-axis are brand names of individual products (Milk/Cereal). Each vertical axis represents a nutrition attribute (Eg: Fats, Carbs, Protein etc), and the Price of the product in question. Each line was color coordinated to represent a specific store that would be shown when the user hovers on with a mouse.

The Second Iteration of Parallel Chart:

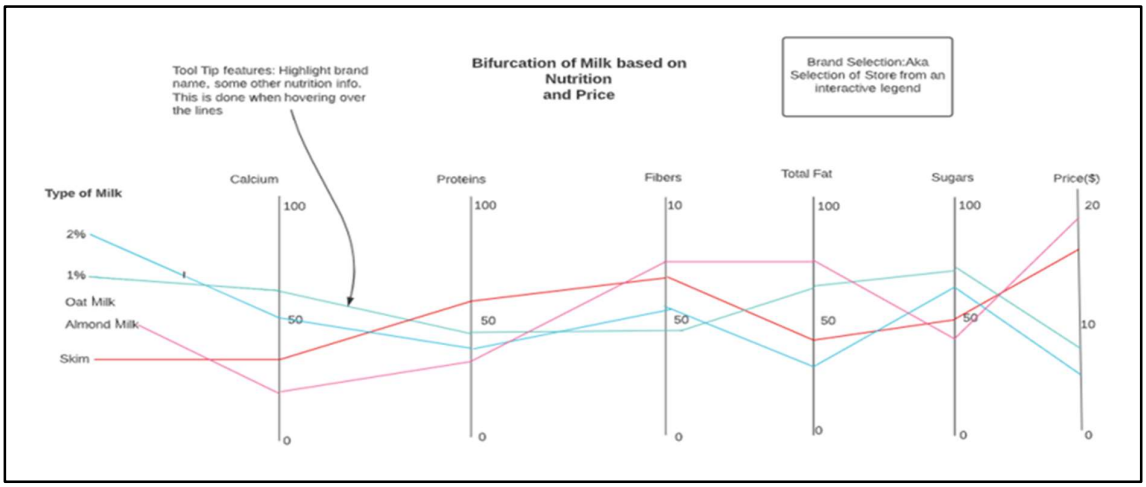


Fig: Second Iteration of the Parallel Plot

The parallel plot was remodeled from the initial phase as follows:

1. Each Vertical axis represents the attributes associated with the product, here it is that of their nutrition value and price.
2. The Y-axis represents the variants present in that selected product. (Ex: Product: Milk, Variants: Skim, 2%, 1% etc).
3. Each coloured line represents a store/brand name attached to it, it is shown when a user hovers over the line using a tool tip.
4. Legend is interactive i.e Different brands could be selected Eg: Costco: Kirkland, WholeFoods: 365 etc

The aim here is that the chart would only show the properties of different kinds of milk along with their respective store names, brand and nutrition values which would help the user transition into choosing their cereal. The visualization would be able to give the user an overview of what types of milk are present in the market along with their likely price-point.

In the dataset: The following values were scaled to make that possible:
The milk prices were scaled to correspond to a half gallon quantity of milk, as most of the retail stores sold their milk in 0.5 gallon measure and most of their prices were listed as the same in individual stores (Eg: Fred Meyer etc). This was done also because we had considered to include sellers that sold in bulk such as Costco and Target, which had their milk priced for 1 gallon / 2 gallons.

This was the basic idea of it. Which brings us to why only milk was the main target here in this visualization, as the next visualization that was to be shown was a dot plot, which showed the affordability of the breakfast per serving to the user. Previously, from the plot we would be able to

conquer the cheapest possible milk options from the parallel plot and would later use it to apply it to the dot plot visualization.

For the implementation part, this design was not chosen:

Initial Iteration: With Scaling and little processing:

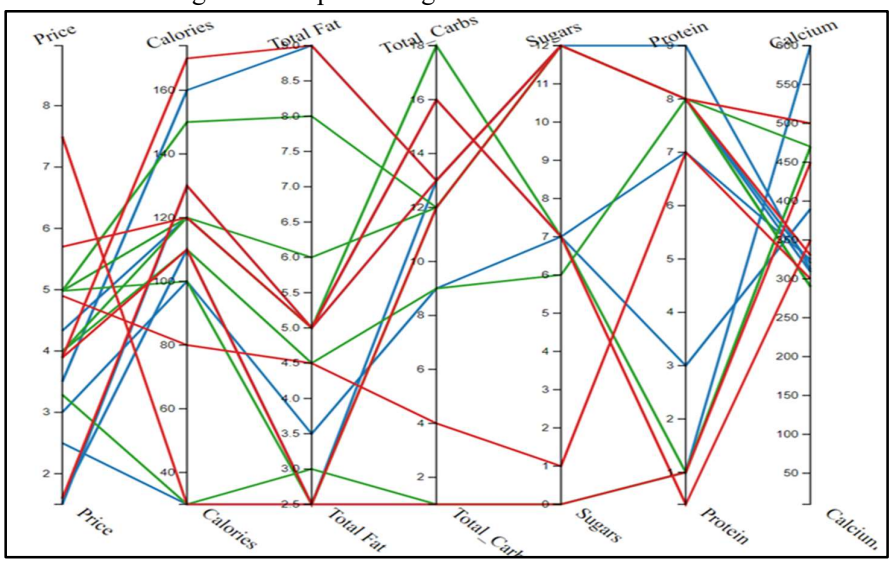


Fig: Initial Iteration without cleaning

After preprocessing and lots of data selection:

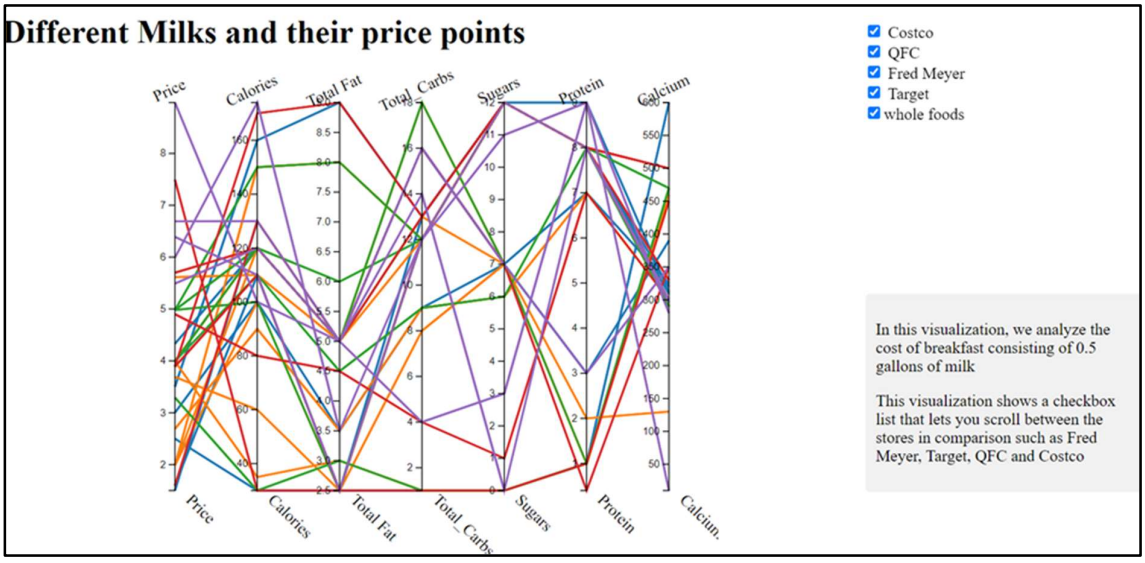


Fig: Secondary Iteration for the Implementation

1. As we can see that even after preprocessing and scaling the data, the graph is a spaghetti chart. The visual appeal misses the point, and this is the result after choosing 6 milk types for each of the selected stores. Even if the user selects the toggle with the store names, even individually the graph still was not clear.

2. Secondly, the graph seems a bit redundant as the cheap milk selection was a little obvious for the dataset. Generally, we know that vegan dietary foods are more expensive than the dairy alternative. So, for the consecutive plot, it was intuitive that the dairy alternative is much cheaper than the vegan kind. Hence, they were rightly selected in the latter graph.
3. Lastly, if the data is too processed or morphed too much, the data points or the knowledge shown becomes biased and incorrect. Ethically speaking it is not correct on our part as data scientists to portray results that are false.

Challenges

1. **Data Collection**: One of the main challenges we faced were gathering accurate and comprehensive data for each breakfast item (cereal types, milk types). It involved visiting multiple stores, conducting a little research on nutritional information. Ensuring the data was reliable was a time consuming task.
2. **Data Standardization**: The collected data was in different formats or units, making it difficult to compare and analyze. Standardizing the data by converting units, ensuring consistent measurements, and aligning nutritional categories was cumbersome. This step is crucial to ensure accurate representation and fair comparison between different breakfast items (cereal types, milk types).