

DTSA 5304 Final - Cereal Nutrition Comparison Visualization

Introduction

Whether it is a nostalgic treat or daily routine, cereal holds a special place at the breakfast table. For health-conscious parents, there is the age-old concern that cereal for children is excessively sugary, but does data support this claim?

This analysis reviews 68 different cereal brands to understand the relationship between the nutritional macros (sugar, fat, protein) and calories for adult-targeted cereal vs. children-focused cereals. The analysis is based around this dataset (<https://www.kaggle.com/datasets/crawford/80-cereals>) where I removed some cereals and labeled them children vs. adult from my experience.

The goal of this visualization is to aid parents and health-conscious adults explore various cereal options to make informed decisions. It also provides larger insight into similarities across brands, most notably brands marketed to children and to adults.

Users of this visualization might have tasks as follows:

- Identify the children's breakfast cereal with the least sugar
 - Purpose: Reduce sugar in a child's diet
 - Means: Compare sugar per serving of cereal
 - Data Characteristics: minimum sugar (low-level), understand relationship between sugar per serving and total calories per serving size (high-level).
 - Target Data: Absolute reference frame of calories and g per serving.
 - Workflow: Initial analysis to answer a relatively simple question
- Find cereal to maximize protein while considering sugar, fat, and calories to meet diet macro goals
 - Purpose: Identify cereals that balance all three macros (sugar, protein, fat) and weigh in accordance with nutritional goals
 - Means: Select high protein cereals and consider calories, sugar, and fat composition
 - Data Characteristics: maximum protein (low-level), weighing other macro factors and serving size (high-level)
 - Target Data: Absolute reference frame of measurement, relative subjective macros
 - Workflow: Performed with protein and macro goals in mind

Visualization

Please use this link (<https://github.com/nvastine/cereal-nutrition-visualization/blob/main/Cereal-Nutritional-Comparison.html>) for the html file to interact with the visualization. Below is a series of screenshots that show the initial state and following a selection. Note the screenshot does not include the tooltip when hovering over a point.

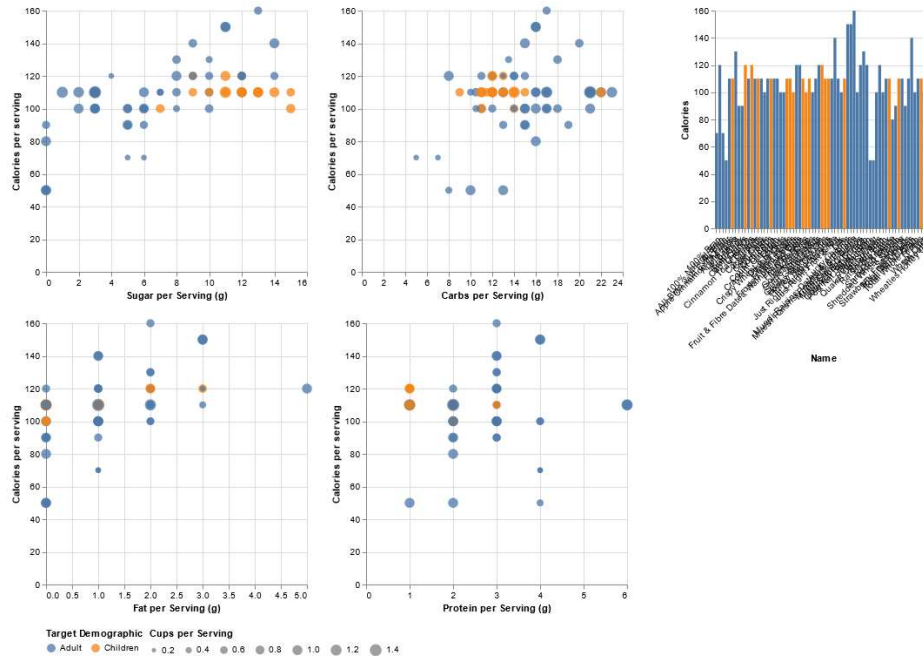


Figure 1: Initial Visualization

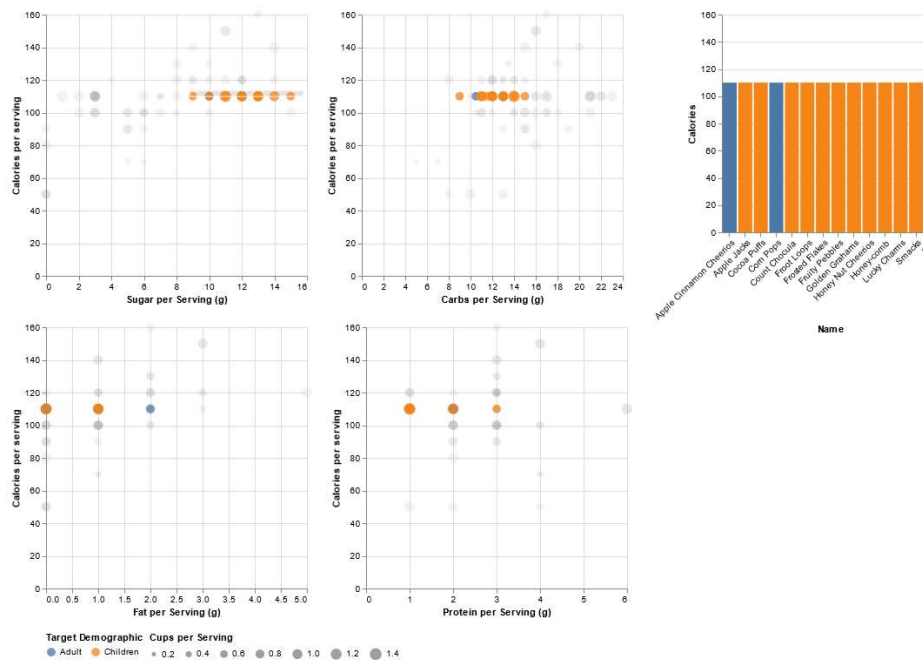


Figure 2: After Selection

Summary of key elements

This visualization prioritizes encoding the main metric (calories and macros) as *position* to make them easiest to interpret.

The target demographic is encoded as *color* since there is a clear categorical distinction between for-children and for-adult cereals. The points are set to a reduced opacity, which allows color the secondary purpose of showing which areas are distinctly for children, distinctly for adults, or somewhere in between.

Size encodes serving size as it isn't critical to this analysis but might be valuable for more detailed review. *Size* is used as a larger point intuitively represents a larger serving size. Thus, a very small point with low calories may help the user consider if the cereal is low calorie only because of a small serving size.

The small multiples also *interact*, so a user can select multiple brands on one plot and compare them on the other plots, making the relevant information pop. There is another bar plot which filters to the selection to help identify the relevant cereals. Unfortunately, the secondary plot is overwhelming with the list of cereals when there is no selection, and it may be unclear to users what the plot is intended for.

A *tooltip* is included for users to identify each point and provide more precise figures. Unfortunately, the individual cereal points overlap due to their similar nutritional value, exacerbated by the FDA rounding 'g per serving' macros. The secondary bar plot should be used for final identification because of this overlapping point flaw.

Evaluation approach

Procedure

This visualization was evaluated using *journals* though supplemented with questions. The study participants were given the tool and asked to provide their initial impression. They were then asked to answer the following questions to help guide discovery of the tool. The questions begin very targeted (low-level) and progress to more complex subjective responses (high-level).

1. Which cereal(s) have the most sugar per serving? How much sugar do they have per serving?
2. Which cereal(s) have the smallest serving size?
3. Which cereal seems the most calorie dense?
4. If you were picking a cereal for bodybuilding (high protein, low carbs, low fat), which cereal would you pick and why?
5. Do you notice any trends across children's cereal?

Finally, they're asked to record any suggestions or points of confusion.

People Recruited

The following people evaluated the design:

1. My sister Bailey, a physical therapy assistant
2. My cousin Rian, a scientist
3. My friend Scott, a software developer

Results

All participants were able to answer all 5 questions thoughtfully, but none of the participants figured out the selection process. Thus, they relied on tooltips for point-by-point comparison across the various plots which would be taxing and harder to filter the data than intended.

In general, they found the color encoding clear and the serving size encoding intuitive. Also many mentioned it would be better to highlight the same point on multiple graphs, which again was encoded as the selection process but was not discovered by the users.

Unsurprisingly the participants had similar concerns in the design as I did, such as:

- Overlap of points makes it difficult to analyze
- Initially being overwhelmed by the secondary plot when all cereals are listed
 - This is especially bad since the users didn't figure out how to simplify the plot through selection
- Lack of context with no overall chart title and unclear secondary plot
 - This too could help guide users to actively select on the graph instead of passively hover for tooltips

Some unique insights they reported included:

- Surprise by how much serving size varied by volume
 - All three suggested the plots should be normalized by volume instead of serving size, which is interesting from a nutritional reporting perspective
- Children cereal effectively all follow the same calories per serving with different sugar content and serving size
- Children cereal tends to have more sugar than adult cereals

They also provided the following suggestions:

- Consider listing multiple cereals in the tooltip
- Providing a filter for adult and children cereals or a set macro limit (ex. max sugar)
- Normalizing by volume instead of by serving size since most consumers pour the same volume of cereal regardless of serving
- Consider other cross plots specific to the questions, for example calories vs. serving size or carbs vs. protein.

Synthesis

The visualization's use of position, color, size, and tooltips communicated the necessary information for the users to gather insights. Once users were made aware of the selection option, they found it much easier to use to identify options for comparison. The use of small multiples made the intent of the analysis clear even with limited provided context, and naturally users wanted to understand which cereal was healthiest.

The largest shortcoming of the visualization was poor communication of the selection methodology. I intentionally did not mention the selection process as I was curious if users would discover it naturally. I believe the participants did not find the selection process since interactive visualizations are far less common than static visualizations, and drawing a selection is less natural than hovering over points.

Some elements to improve in future design include:

- Consider a hover selection because it passive responds to the user instead of requiring action by the user to draw the selection, which is less intuitive.
- Improve labeling and instructions within the plot itself (for example adding a title for 'Cereal Nutritional Comparison').
- Add 'Selected Cereals' title to secondary plot explaining it is selected cereals for comparison, hopefully indicating to the user they can select from the graph.
- Make secondary plot much wider so even if all cereals are selected it isn't overwhelming.
- Jitter the points to reduce the effects of overlap.
- Add filters to restrict the data set for adult vs. children cereals.
- Consider per-volume reporting instead of per-serving reporting.
- Consider additional specific plots depending on intended questions from consumers.
 - Ex: Do manufacturers intentionally reduce serving size to reduce calories? (serving size vs. calories), Do manufacturers reduce serving size to reduce sugar? (serving size vs. sugar), Which cereal maximizes protein while limiting carbs? (carbs vs. protein)

Conclusion

While the visualization was not perfect, I am glad the participants were able to engage with it and answer the provided questions. I believe this visualization captures the right amount of information without being overwhelming, though it relies on user knowledge to better interact with the tool. It is also important to recognize that any flaws identified as the developer (ex: overlapping points, dense initial secondary plot) are very likely to be identified by users. In this case the developed selection solution to these problems was not discovered by the users, and as such the flaws seem entirely unaddressed. It was very valuable to collect feedback and be forced to evaluate the design critically, especially with a diverse audience.