Cereal Nutrition Design Study

Dataset

The dataset I have chosen is “80 Cereals”, provided on Kaggle by Chris Crawford. It can be found here:

https://www.kaggle.com/datasets/crawford/80-cereals

The dataset provides nutrition information on 80 different breakfast cereals produced by major breakfast cereal manufacturers. There are 80 rows with 16 columns. Each column, after the name of the cereal, its manufacturer, and its type (hot or cold), is a nutrition fact about the product: calories, fat, etc; followed by serving size information.

At the most abstract, what I am interested in here is the relationship between different nutrition facts in these cereals and what, if any, correlations can be found between them.

The previously provided visualizations for this dataset are histograms for each of the nutrition facts. These visualizations do a fine job of showing relative frequency of values for these things across cereals but do not help to illuminate the relationship between the values in given cereals or even to show which cereals have what levels of the relevant nutrient. To bolster this information I will create visualizations that help to compare these values and explore their relationships.

Question of Interest

The specific question I am interested in is, given the rise of low carbohydrate, high protein focused nutrition advice, what cereals are most worthwhile for a consumer concerned with their protein intake? That is, what is the relationship, if any, between protein content, carbohydrate content, and sugars, and can a general guideline be provided which would allow a consumer to select a breakfast cereal with a relatively high protein content and a relatively low sugar and carbohydrate content?

Thus, the goal for the visualizations is to provide clearer information comparing the different cereals and their nutritional values. The task will be conducted by comparing the protein content of the cereals to other salient features of their nutritional profile: fat, sugar, carbohydrates, vitamins, etc.

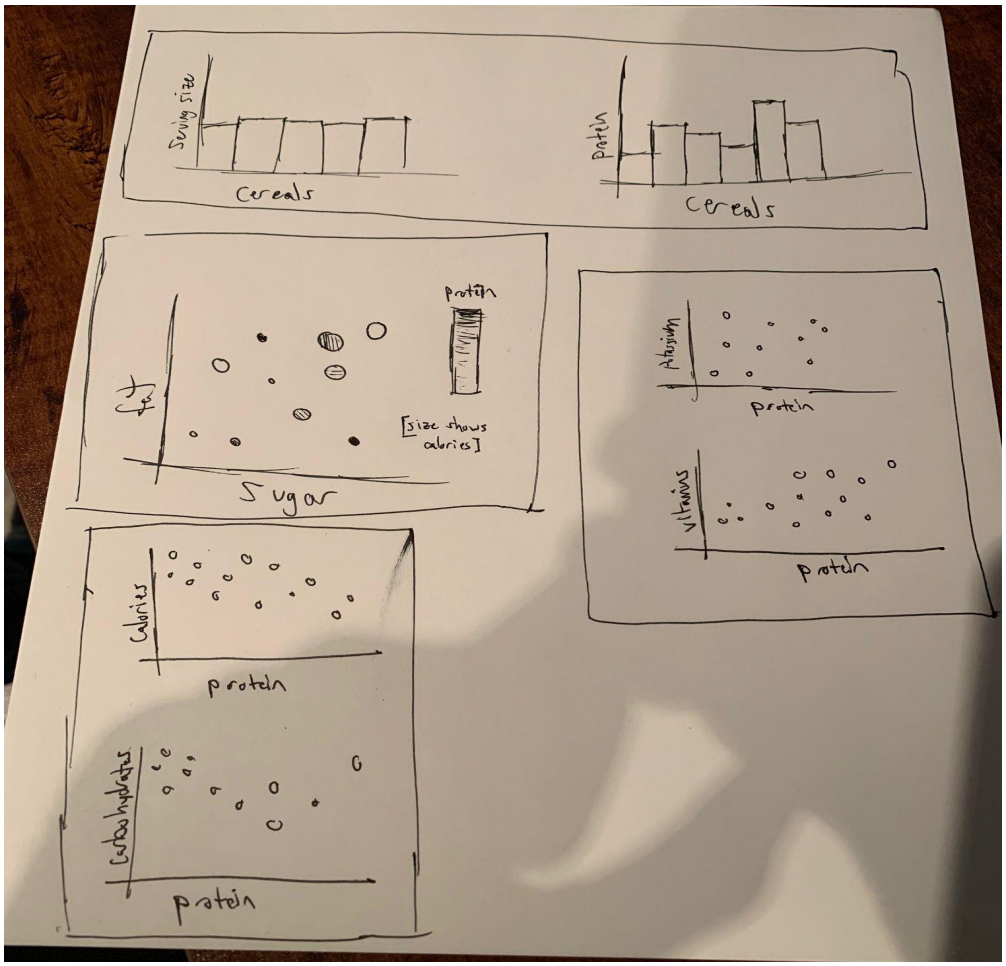
Low-Fidelity Prototypes

The visualizations thus seek to provide clear guidance about which cereals, if any, provide the highest protein content relative to their other nutritional content.

To this end, I will create the following visualizations:

1. A pair of histograms of serving sizes by weight and protein content of cereals
2. A graph with a distribution of cereals showing their relative protein, calories, sugars and fat
3. A pair of graphs with a distribution of cereals showing their relative protein and vitamins and relative protein and potassium
4. A pair of graphs with a distribution of cereals showing their relative protein and calories and relative protein and carbohydrates

I created the following rough sketches of the visualizations that I planned to create:



Study

In order to evaluate the efficacy of these visualizations, a group of ordinary consumers should be recruited. For this project, friends and acquaintances were used. Groups were shown the created visualizations as well as the ordinary nutrition labels of a group of cereals. These

participants were then asked whether they feel the visualizations helped them to understand the relevant nutrition information for these cereals and if they feel that these visualizations might affect their buying behaviors in the future. They were also asked about which, if any, of the visualizations they felt were most helpful to them and which, if any, they felt were not helpful or unnecessary. These interviews were done in a “think aloud” fashion.

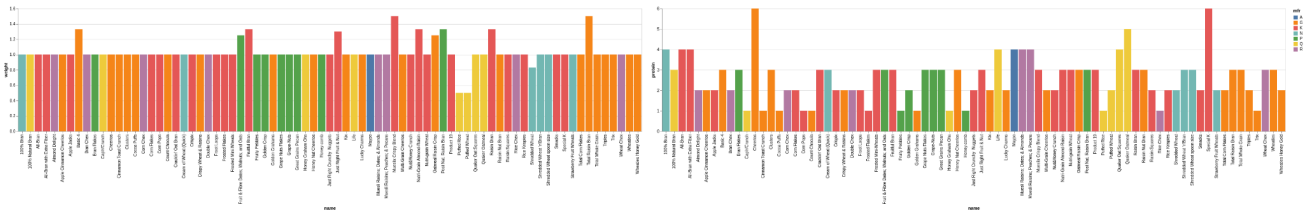
For a larger-scale study, participants might be found at grocery stores or other food retail locations. Participants recruitment and interviews could be conducted by the same individual. Interview results could then be given to a design team to implement improvements. This process could then be repeated as necessary.

The visualizations would be considered successful if the consumers did feel that they had a better sense of the nutrition information looking at the visualizations than from looking at nutrition labels alone and if the consumers reported that this increased understanding would affect their future purchasing behavior.

Visualizations and Results

My final visualizations were:

1. A pair of histograms of serving sizes by weight and protein content of cereals

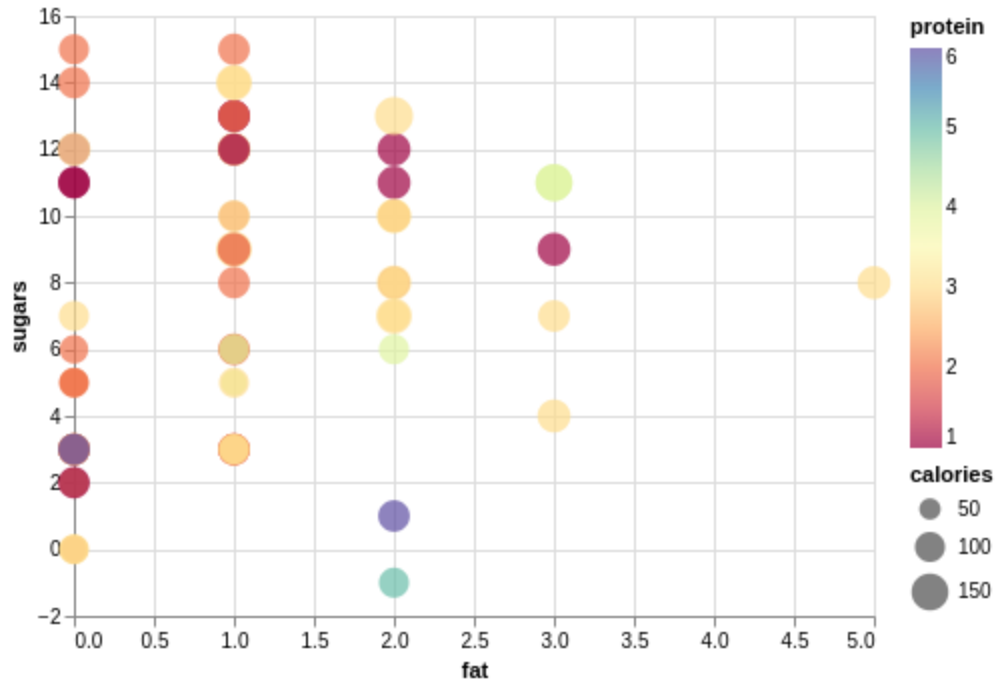


Note, this screenshot has been shrunk substantially from the version presented to subjects. Colors correspond to manufacturers in order to help consumers find the brands they are most likely to buy. The histogram to the left depicts the serving sizes of the cereals; the one to the right shows the protein contents of the cereals.

Subjects were particularly interested in the second histogram, showing relative protein content. This graph was perhaps the most effective in altering consumer behavior in subjects concerned about protein content. The simplicity and clarity of the graph were especially effective here.

Many subjects reported finding the color-coding unhelpful and/or distracting. In future iterations this should be removed.

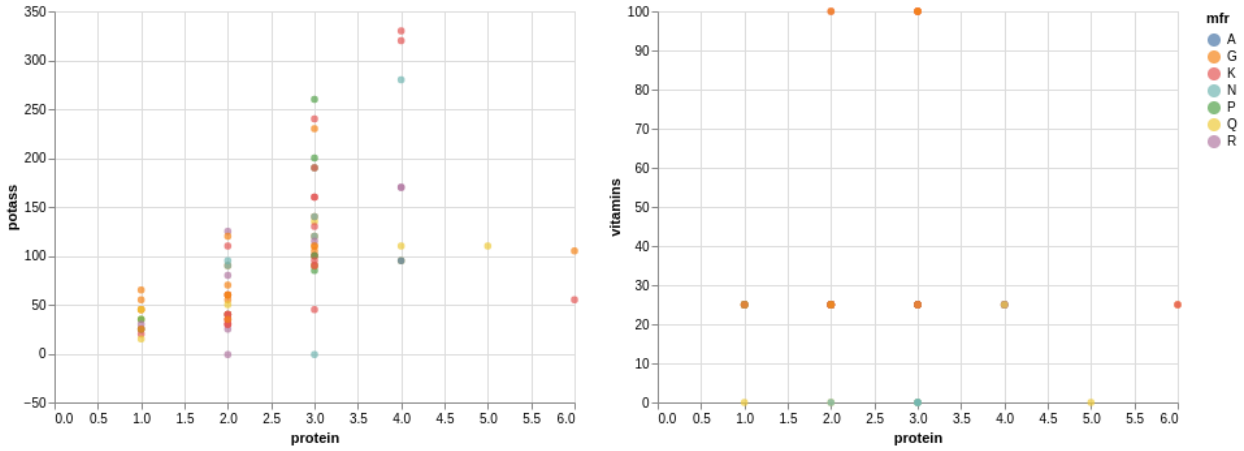
2. A graph with a distribution of cereals showing their relative protein, calories, sugars and fat



This is the most complex visualization. In the version presented to subjects the points can be hovered over to receive the name of the cereal as a ‘tooltip’. This visualization is the most visually complex design. It is aimed at being more visually appealing and at providing the highest density of information.

As hoped, subjects did enjoy the aesthetic appeal of this design. However, it was also somewhat difficult for subjects to parse. This graph is perhaps the most deserving of repeated iterations. Moving forward, multiple versions should be produced in order to maximize its legibility while preserving its level of viewer engagement.

3. A pair graphs with a distribution of cereals showing their relative protein and vitamins and relative protein and potassium

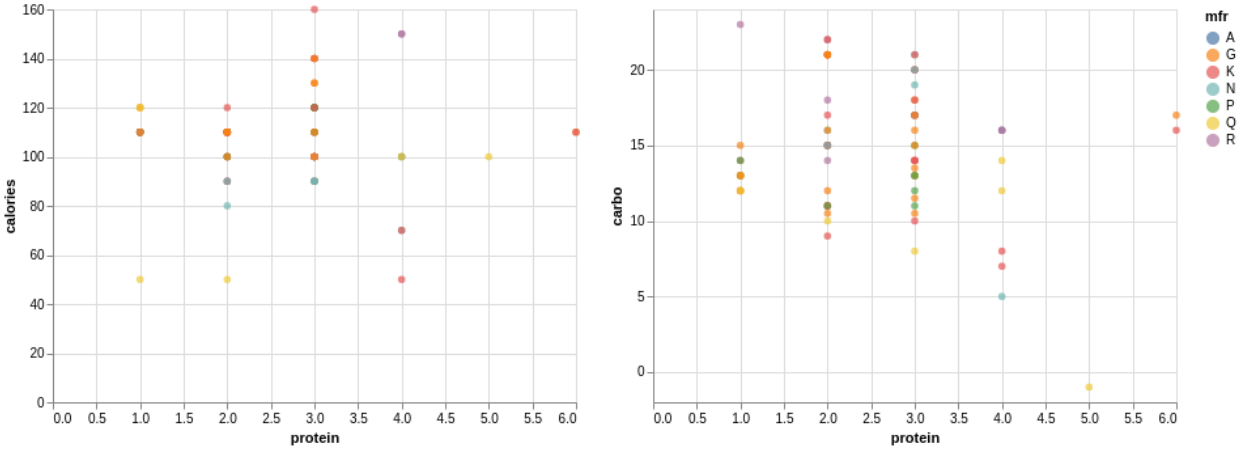


As before, color corresponds to manufacturer and the name of each cereal can be received as tooltip when hovered over with the mouse in the version viewed by subjects. The simplicity of this pair and the one following were intended to be clear to subjects.

Subjects again reported color-coding of manufacturer data to be unhelpful. Future designs should omit manufacturer data, perhaps across all designs. It simply does not appear to be important to consumers.

This pair was the least engaging to the subjects. Potassium content was not a salient product feature for them and vitamin content has a peculiar distribution that ultimately makes very little difference to self-reported consumer behaviors.

4. A pair graphs with a distribution of cereals showing their relative protein and calories and relative protein and carbohydrates



As before, color corresponds to manufacturer and the name of each cereal can be received as tooltip when hovered over with the mouse in the version viewed by subjects.

This pair and the pair previous were designed to be simple in order to be clear to subjects. This pair, calorie and carbohydrate count, aimed to provide more obviously salient information to consumers than potassium or vitamin counts.

As expected, subjects were very interested in these visualizations. Subjects who were most concerned with protein content were also very concerned with carbohydrate content and maximizing protein to carbohydrate ratios was important to them.

Subjects said that while they did not feel that the graphs provided substantially different information to them on these topics than the nutrition labels, having the graphs made it much easier to compare a larger sample of products simultaneously and easily select the ‘best’ options.

Conclusions

The clearest outcomes here are that consumers need important, simple information conveyed in a visually appealing way. In general, manufacturer data was not relevant to subjects and was found to be distracting in visualizations. This information should be dropped from future iterations of the designs. Color coding more generally was not necessarily disliked

but must be easy to understand. This is most pressing in future iterations of the sugar-fat-protein-calories visualization. Subjects enjoyed the aesthetic appeal of the visualization but found it to be too much information to effectively process at a glance.

The careful balance between engaging design and effective communication is especially important here. For consumer information, in particular, viewers are unlikely to make a great effort to understand visualizations without a certain amount of visual appeal.

The most effective visualizations were the simplest. A direct one-to-one graph of protein to carbohydrates and the simple histogram of protein content were by far the visualizations that the subjects reported as being most likely to alter their future purchasing behaviors. Moving forward, simplicity should be emphasized while attempting to establish appealing designs.