

Midterm Project Description
SLIS/IGPI-6155: Information Visualization

Instructor: Chaoqun Ni

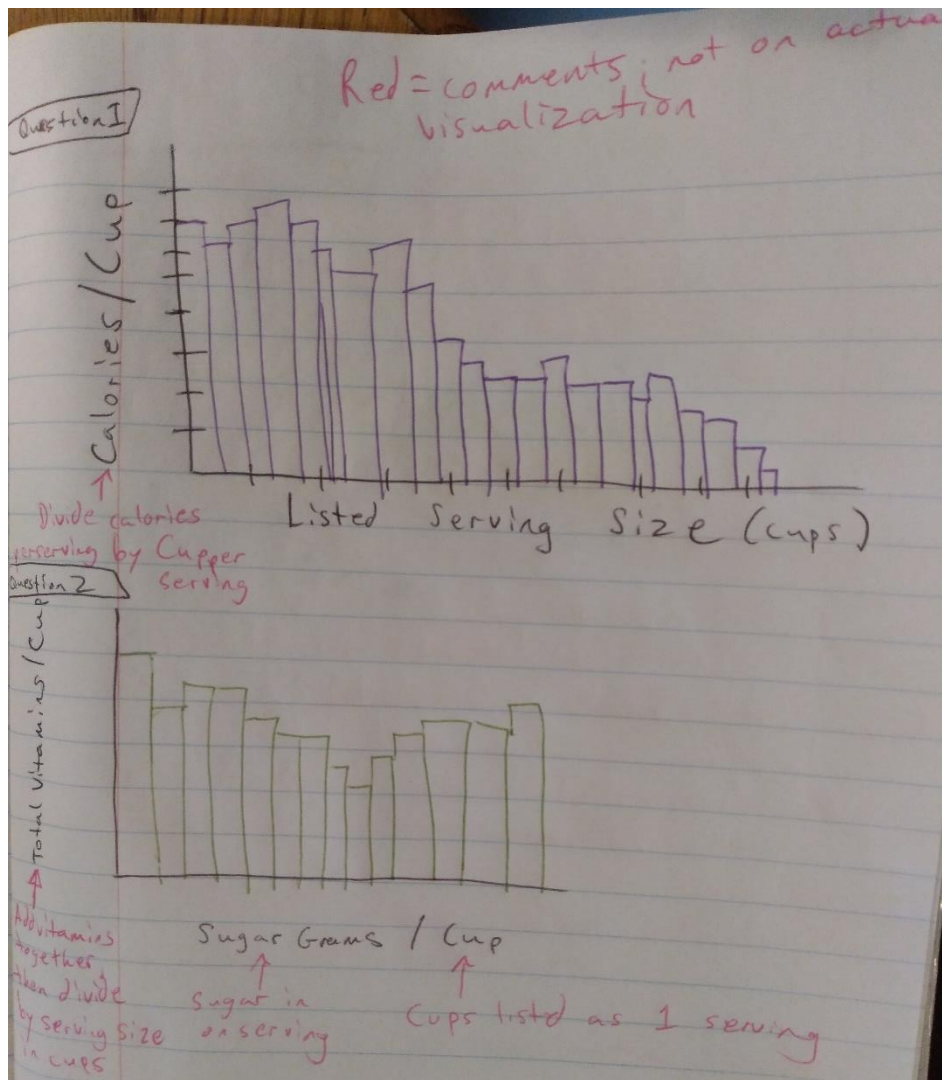
Student: Michael Hart

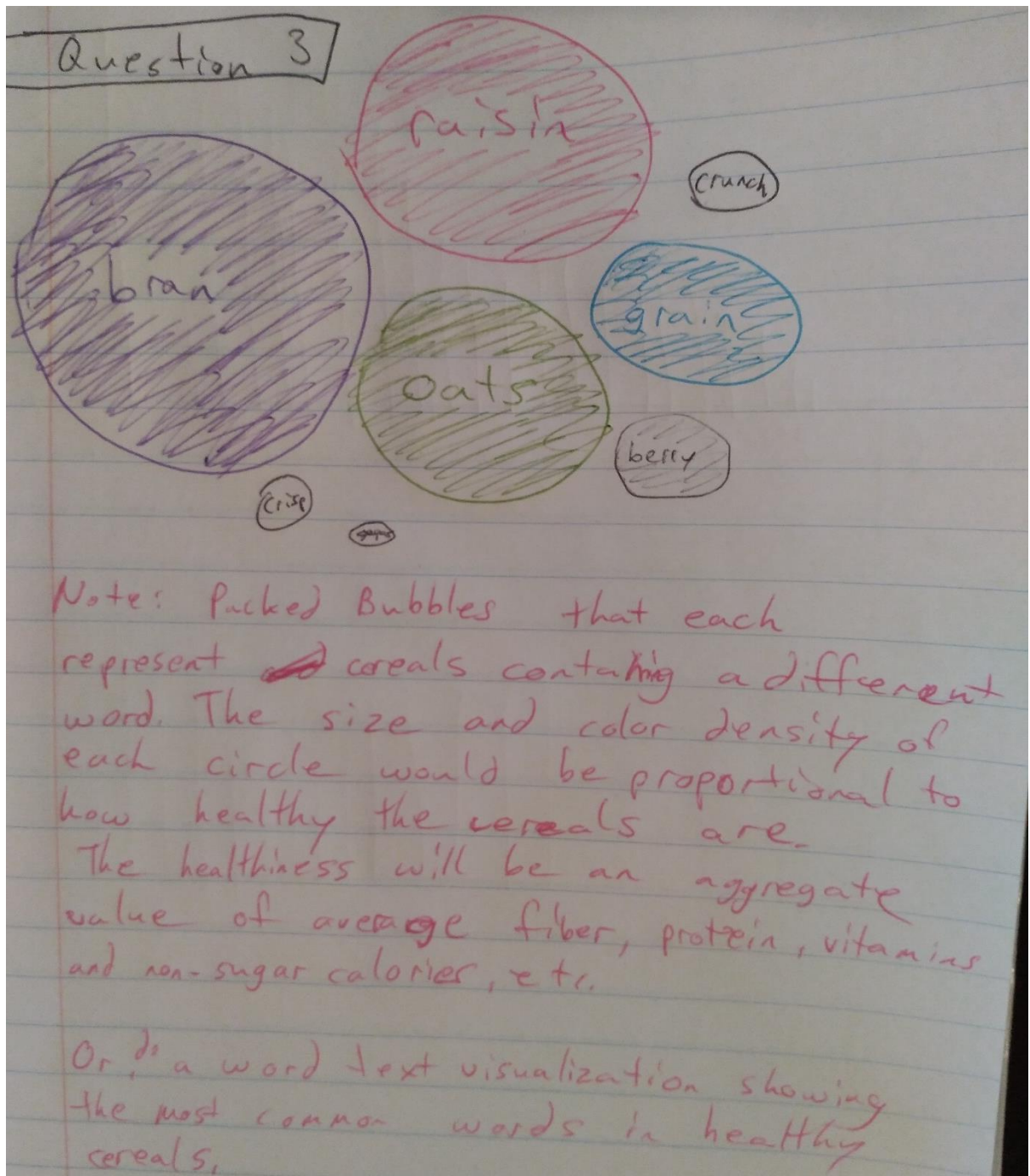
- List your three analytic questions or queries.
 1. *Do cereals with higher caloric density (defined by calories/cup) list a smaller portion as the default serving size on the cereal boxes?*

I would assume that most cereal producers would do this to hide that a cereal is high-calorie, as most shoppers probably just glance at the calories-per-serving and don't actually check the serving size and perform the division necessary to calculate the caloric density of different cereals.
 2. *Is there a relationship between sugar content and vitamin content?*

My hypothesis is that sugar-dense cereals are generally less vitamin-dense. However, the inverse could be true if the manufacturers of sugar-dense cereals fortify them with vitamins to make the cereal appear "healthy" to skeptical parents. Higher vitamin content per serving allows manufacturers to market high-sugar candy as containing "80% of daily required vitamins C and D", for example.
 3. *Are cereals with healthy, wholesome, or natural-sounding words in their name, such as "grain", "raisin", "bran", "oats" etc. actually healthier than other cereals?*

I will make a list of natural-sounding words that appear in the cereal names to establish this category, and check for a relationship to indisputably healthy cereal content such as fiber, protein, low-sugar, and vitamins. I will exclude "fat" in this measure, since the long-term health effects of dietary fat consumption are still widely disputed. Caloric density will also be disregarded for this question, since many calorie-dense foods are healthy and many low-calorie foods are unhealthy.
- Draw/sketch/show your design on a separate piece of paper. Feel free to annotate the sketch with small comments or captions to explain what it is and how it would work.

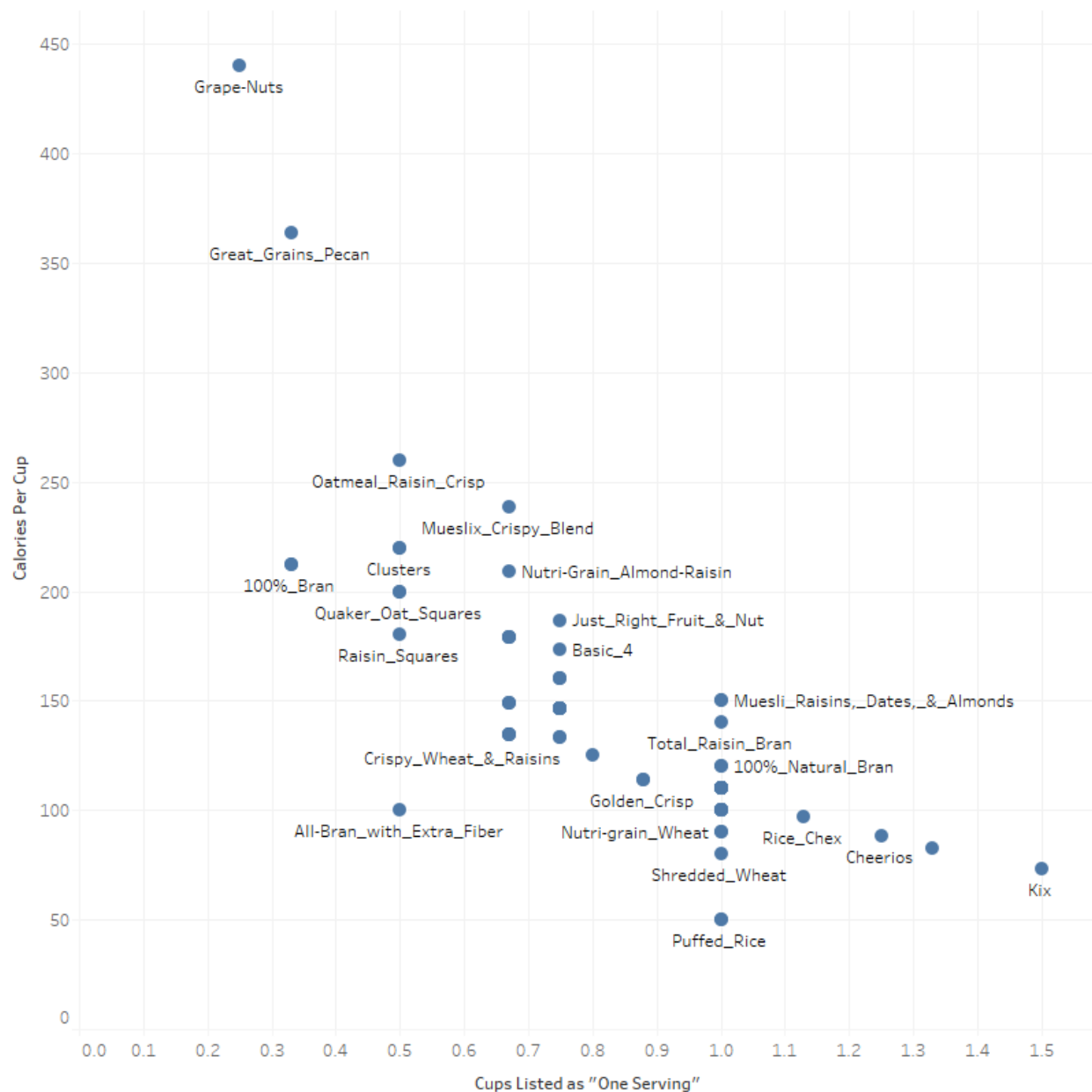




- On a separate page, explain your visualization design in a paragraph or two. Feel free to turn in scanned paper-based sketches.
- For the question "Do cereals with higher caloric density (defined by calories/cup) list a smaller portion as the default serving size on the cereal boxes?", I want to use a bar graph or scatterplot to show the precise numbers that are used. A bar graph or scatterplot will also allow me to include all 78 cereals in one visualization, which would be difficult with a treemap, packed bubbles, or pie chart. Moreover, I expect the correlation to be strongly negative, so a bar graph or scatterplot is the most straightforward way to convey this relationship to the viewer.

2. For the question *“Is there a relationship between sugar content and vitamin content?”*, I am planning to use a bar graph or scatterplot for the same reasons listed in question 1. For question 2, I don’t know if the correlation will be negative or positive, so a bar graph or scatterplot will help figure that out. For both questions 1 and 2, I am looking at the relationship between the nutrient content of the cereals, and am not overly concerned with the values of individual cereals, so I will only label the names of the individual cereals if the visualization is not already oversaturated with information.
 3. For the question *“Are cereals with healthy, wholesome, or natural-sounding words in their name, such as “grain”, “raisin”, “bran”, “oats” etc. actually healthier than other cereals?”*, I am focusing on a small number of keywords that appear in the cereal titles, so I can easily include all of them in the visualization. Because I am not seeking to establish a positive or negative correlation between two numerical values, a bar graph or scatterplot would not be helpful. Instead, I will use a treemap or packed bubbles, which emphasize the relative difference in values between the cereal name keywords.
- Provide snapshots and the discussion of your Tableau visualization and be able to demonstrate interactively how you arrived at this visualization.

Cups Per Serving vs Calories Per Cup in Common Breakfast Cereals



Sum of Cups Listed as "One Serving" vs. sum of Calories Per Cup. The marks are labeled by cereal name. Details are shown for cereal name.

This visualization plots the number of cups listed as "one serving" versus the calories per cup. As I predicted, there appears to be a strong negative correlation between the cups listed as "one serving" and the calories per cup.

I chose to use a scatterplot rather than the bar graph that I originally sketched because scatterplots are better for displaying positive or negative correlations. Furthermore, the bars on a bar graph take up a lot of space unnecessarily. Using a scatterplot allowed me to use that white space to include the names of

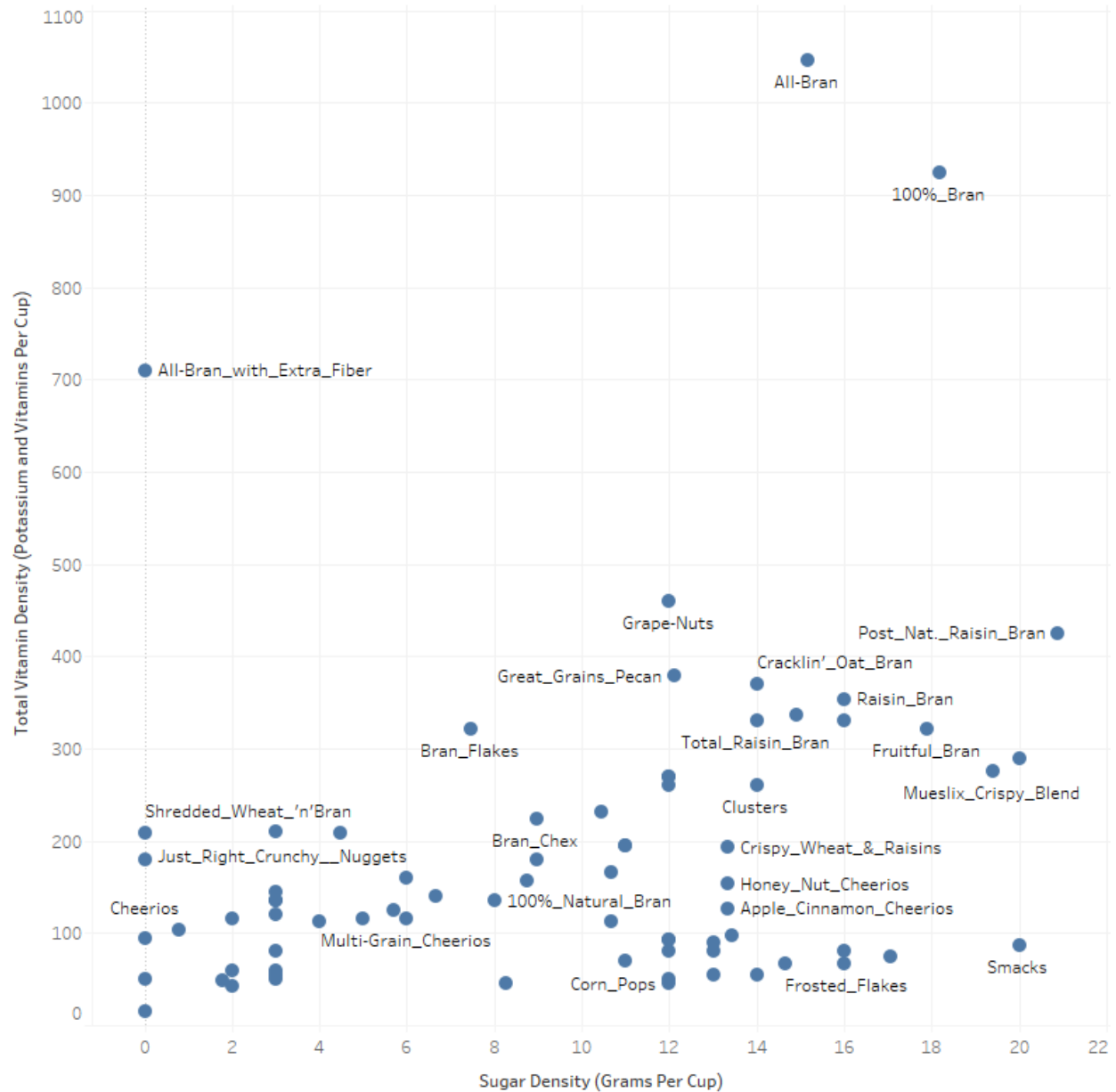
most of the cereals next to the dots that represent their place on the scatterplot, so more information is provided to the viewer at a glance.

Most of the cereal names are listed next to the dot that represents them, but in the cases where they aren't, the user can hover their mouse over the dot on Tableau Public to see the cereal name as well as the precise values of "Cups Listed as 'One Serving'" and "Calories Per Cup".

It is interesting to note that this correlation does not necessarily reflect the healthiness of the cereal. Grape-Nuts have the highest caloric density and list a smaller number of cups as "one serving", but the cereal is high in fiber and other nutrients, and is rather tasteless and bland. Conversely, Kix, Cheerios, and Rice Chex all have low caloric density and list a larger number of cups as "one serving", but those cereals are not particularly vitamin-dense or fiber-dense.

This suggests to me that cereal manufacturers are responding to the market forces of consumer choices/preferences to keep the calories-per-serving listed on cereal box nutrition facts relatively low. If the serving sizes on the boxes were uniform, calorie-dense yet healthy cereals like Grape-Nuts would display a shockingly high number of calories in one serving. That may cause fewer shoppers to purchase the cereal, regardless of how healthy it is.

Sugar Density vs Vitamin Density in Common Breakfast Cereals



Sum of Sugar Density (Grams Per Cup) vs. sum of Total Vitamin Density (Potassium and Vitamins Per Cup). The marks are labeled by cereal name. Details are shown for cereal name.

This visualization plots the sugar density of cereals against the total vitamin density. There seems to be a slightly positive correlation, with high-sugar cereals generally having more vitamins.

I used a scatterplot instead of a bar chart here for the same reasons that I listed for the first visualization. In this case, a scatterplot is even more beneficial, because there are several extreme outliers which are much easier to spot than they would be in a bar chart.

SLIS/IGPI6155-Information Visualization

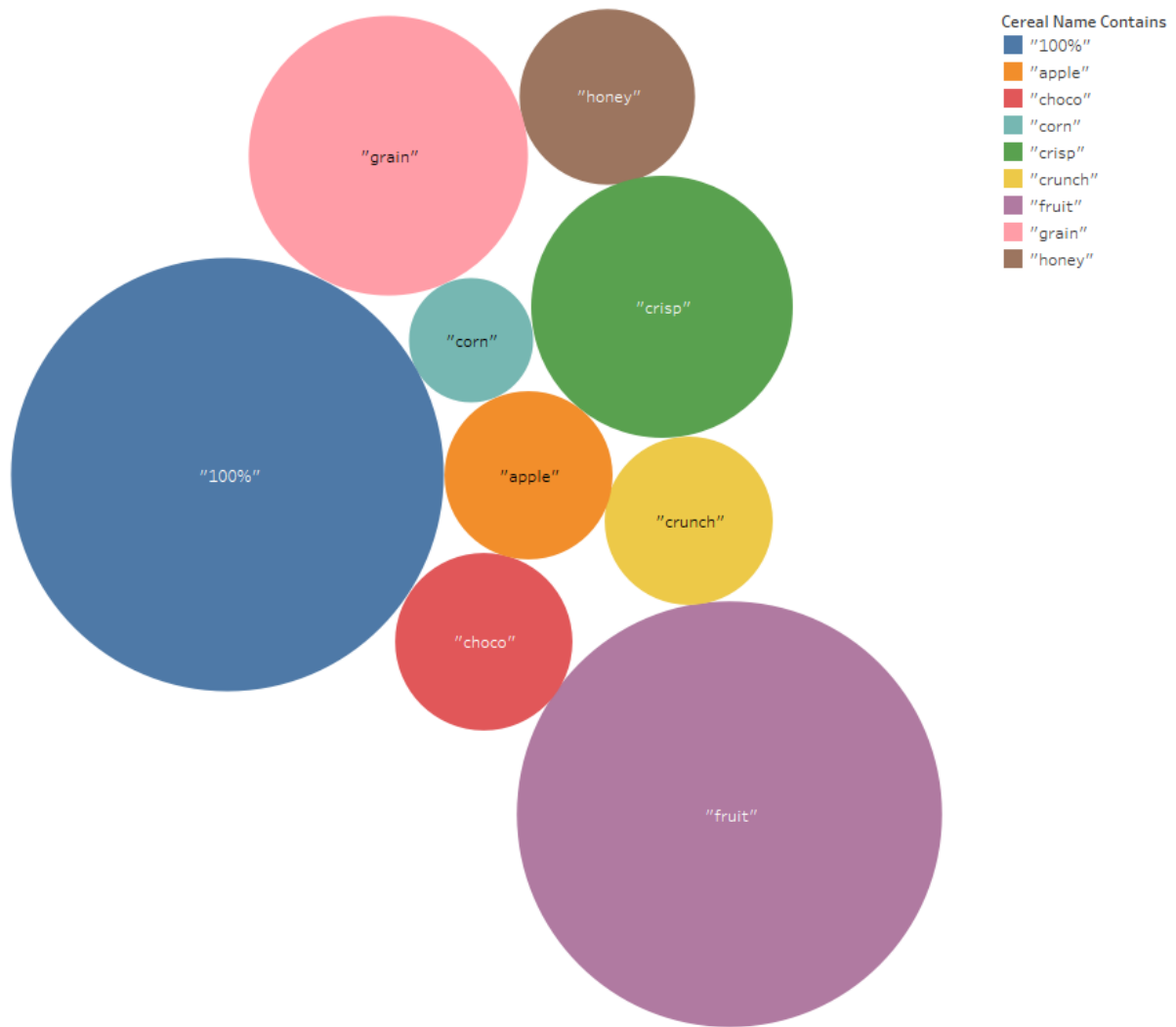
Most of the cereal names are listed next to the dot that represents them, but in the cases where they aren't, the user can hover their mouse over the dot on Tableau Public to see the cereal name as well as the precise values of "Sugar Density" and "Total Vitamin Density".

Without looking at the detailed ingredient listings for each of the cereals, it is difficult to determine if this correlation is due to the manufacturers adding vitamins to high-sugar cereals in order to coax parents into buying the cereals for their children. It could also be the case that many naturally healthy cereals that are high in vitamins would be flavorless without added sugar.

The outliers of this visualization are also interesting. All-Bran with Extra Fiber has very high vitamin density combined with very low sugar density, suggesting that it is one of the healthiest cereals. 100% Bran has both high vitamin density and high sugar density. Smacks have extremely low vitamin density and extremely high sugar density, which would help us to nominate this cereal as one of the unhealthiest.

Average Healthiness of Cereals With Keywords in Name

Healthiness = $\frac{[\text{Fiber}]}{[\text{Cups}]} + \frac{[\text{Protein}]}{[\text{Cups}]} + \frac{[\text{Potassium}]}{[\text{Cups}]} - \frac{[\text{Sugars}]}{[\text{Cups}]}$



Cereal Name Contains. Color shows details about Cereal Name Contains. Size shows sum of Healthiness. The marks are labeled by Cereal Name Contains.

This visualization shows the relative healthiness of cereals that contain specific keywords. I created a "healthiness" metric by adding the fiber, protein, potassium, and vitamin density, and subtracting the sugar density from that.

Each circle only displays the keyword, but if the user hovers their mouse over the circle on Tableau Public they will be shown the exact "Healthiness" value that was calculated. I chose to not display the Healthiness value by default on the circles since it is a derived value, and won't make intuitive sense to viewers. Instead, I want the viewer to focus on the relative sizes of the circles.

I chose to visualize healthiness with the packed bubbles because my healthiness metric is a multivariate calculation and it would be difficult to visualize the precise numbers that went into it (i.e. the graph/chart would be multi-dimensional).

The visualization could be useful for cereal marketers if they want to avoid naming a new cereal with a word that is often associated with unhealthy cereals. It could also be useful for health-conscious shoppers who don't want to be tricked by the name of cereals; just because a cereal contains the wholesome-sounding words "corn", "apple", or "grain", it may not necessarily be a healthy choice. Shoppers can also learn that they should consider avoiding cereals with "crisp", "crunch" or "choco" in their name, although those are less of a surprise.

This method does not account for misspellings, such as the "Froot" in "Froot Loops". The visualization's power is also somewhat weakened by the low sample size; there are only 1-2 cereals for some keyword categories.

- Provide a link to your Tableau Public upload for this purpose.

First Two Visualizations

https://public.tableau.com/views/MichaelHartMidterm12/CupsPerServingvsCaloriesPerCup?:display_count=y&publish=yes&:origin=viz_share_link

Third Visualization

(I couldn't figure out how to filter cereal names based on contained words within Tableau, so I had to make a separate Tableau file sourcing a different Excel workbook)

https://public.tableau.com/views/MichaelHartMidterm3/Sheet1?:display_count=y&publish=yes&:origin=viz_share_link

