

Nutritions

Hypotheses:

Mothers working with food nutrition data want to generate a list of healthy food for their children's meals. As there are too many kinds of nutrition data, it's impossible for mothers to evaluate all of them. Therefore, only several influential nutrition data should be considered. Mineral is necessary for children to build strong bones, muscle and healthy red blood cells. Ash, combination of essential minerals and toxic materials, should also be taken into account. What's more, children are supposed to get enough energy from meals.

Tasks:

1. Does high lipid mean much energy?
2. What is the relationship between ash and minerals (e.g. Sodium)?
3. In terms of breakfast cereals, identify cereals with the most ash;
4. What cereals would be the best for children's breakfast?.
5. Does the nutritional value of cereals(e.g. ash) vary a lot? if so, how?

Analysis:

1. In terms of relationship and trend of lipid and energy, I would like to use the line plot, which is a quick and simple way to organize data.

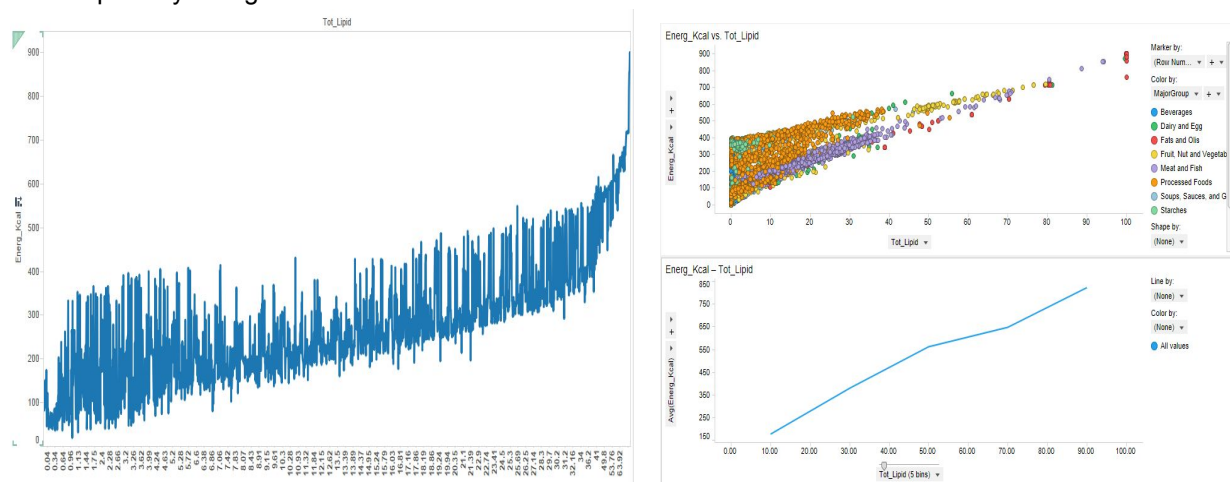


Figure 1 Energy vs. Lipid

2. Same as task one, I would like to show the relationship between the amount of ash and sodium.

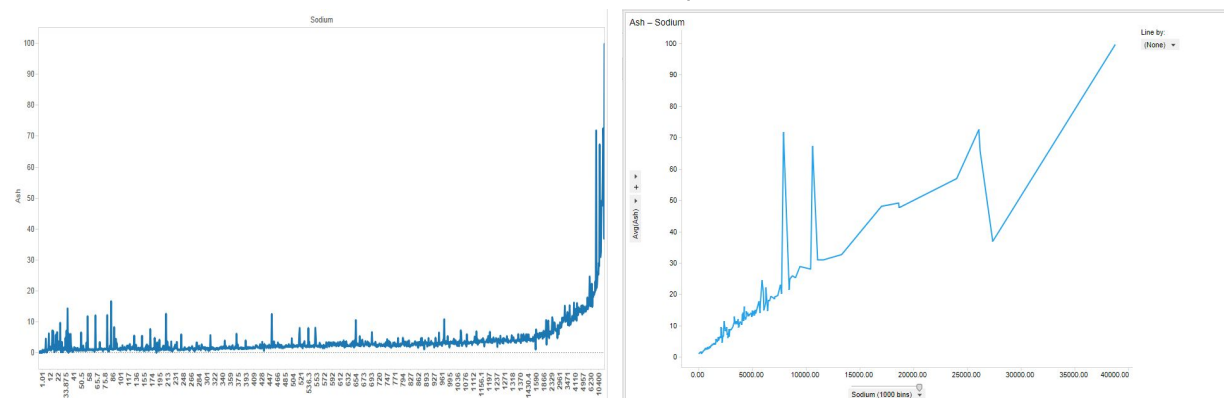


Figure 2 Ash vs. Sodium

3、Make a line plot about ash and short_desc, and then keep only shot_descs which belong to breakfast cereals. Finally, sort short_desc descending by ash. In this case, the short_desc in first column is the cereal with most ash.

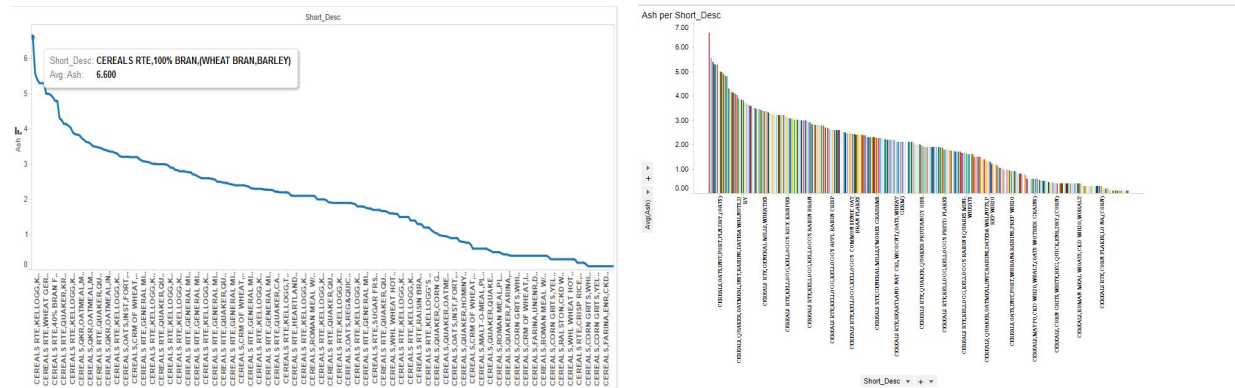


Figure 3 Ash vs. Cereals

4、A balanced breakfast includes carbohydrates(provides energy), protein and fiber(provides a feeling of fullness). Therefore, I need to make three line plots and calculate the sum of carbohydrates, protein and fiber. And then keep only short_descs belong to breakfast cereals. Finally, draw a line plot based on the sum and sort it descending by the sum.

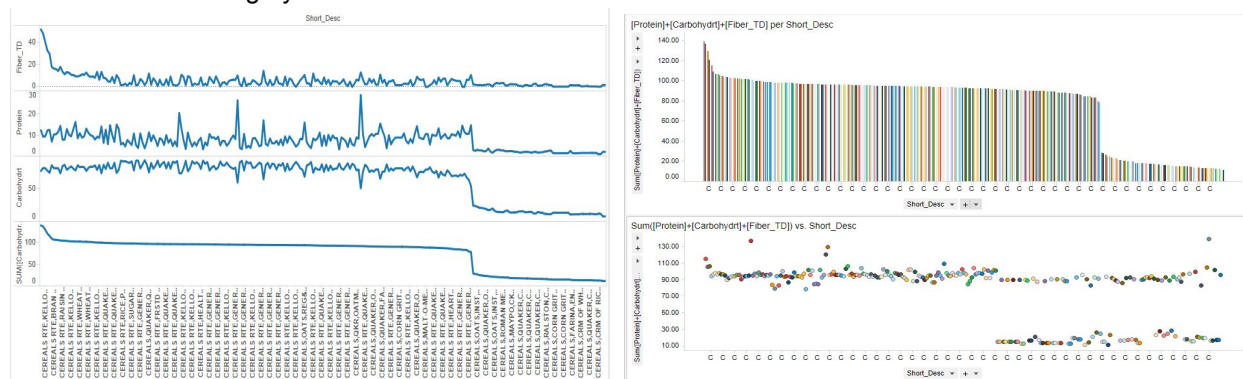


Figure 4 Sum of carbohydrates, protein and fiber vs. cereals

5、It is difficult for Tableau and Spotfire to calculate all nutrition values' variances and put them into one diagram just like Excel. Therefore, I chose ash as the value on y-axis and short_descs belong to breakfast cereals as the value on x-axis. And then draw the average line in the line plot.

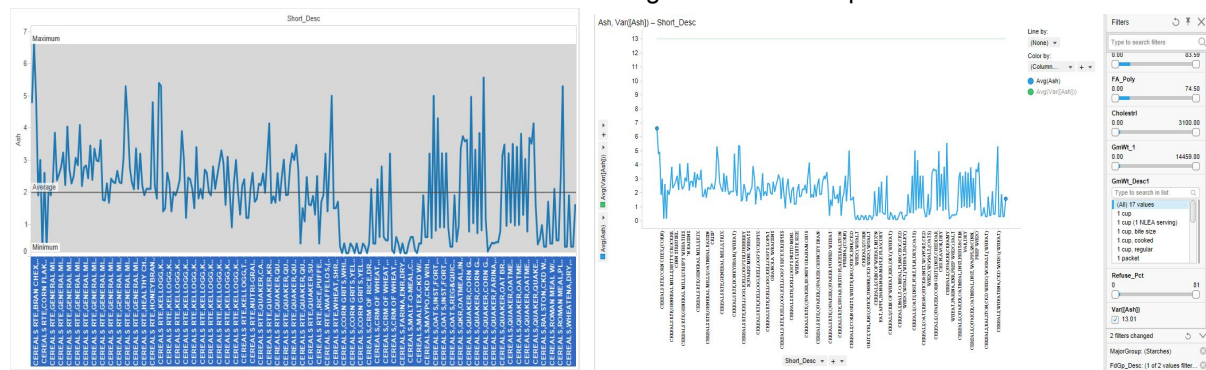


Figure 5 variation of ash vs. cereals

Colleges

Hypotheses:

High students who will go to university want to choose the most suitable college after working with these college information.

Tasks:

1. Does high graduation rank mean high reputation?
2. What is the relationship between % Full-time Faculty and Faculty Residence Rank?
3. Identify colleges with the highest 25th SAT/ACT;
4. Which college would be best for the students who want small class with high quality?
5. Does the value of top 10% of HS Class vary a lot? if so, how?

Analysis:

1. It's simple and effective to use a line plot for showing the relationship and trend of graduation rank and reputation.

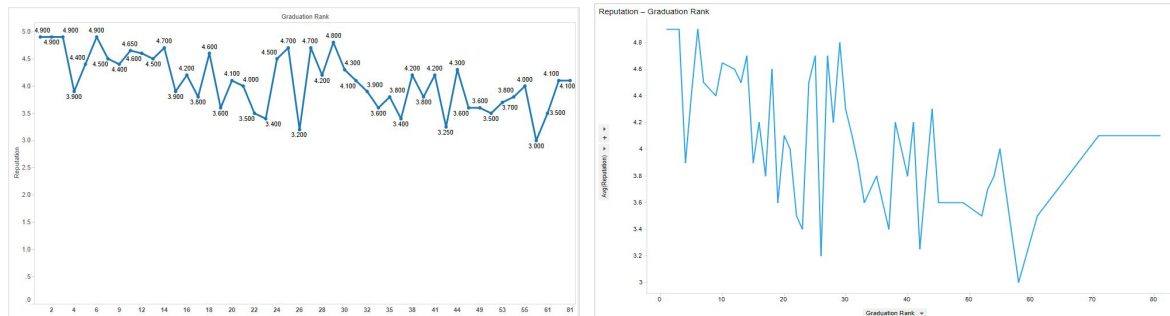


Figure 6 Reputation vs. Graduation Rank

2. Same as task one, draw a line plot which shows the relationship between % Full-time Faculty and Faculty Residence Rank.



Figure 7 Relationship between % Full-time Faculty and Faculty Residence Rank

3. I need to draw a bar chart to identify the value of 25th SAT/ACT, And then choose the school with highest value according to the height of the bar.

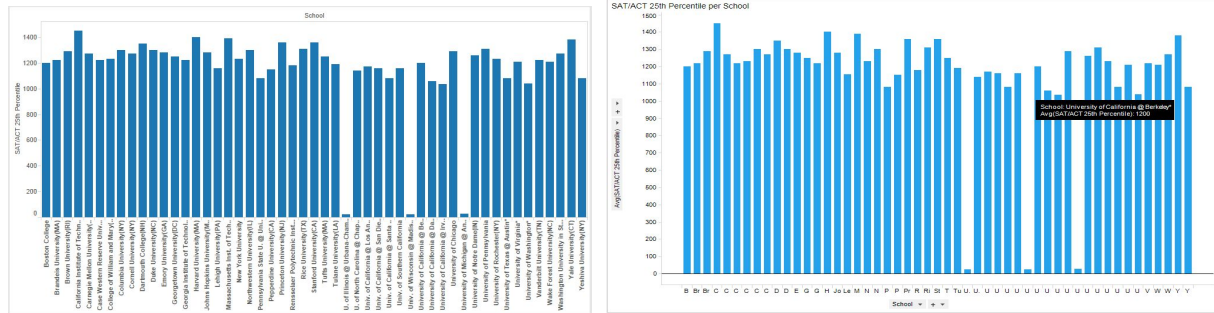


Figure 8 SAT/ACT 25th Percentile vs. School

4、Based on the requirements of students, I'd like to take % classes under 20, students/faculty ratio and financial res. rank into account. I need to draw three line plots for these three factors. And then calculate the index of ranking the schools. The value of index = value of % Classes Under 20 - Student/Faculty Ratio - value of Financial Res. Rank. Finally, I draw a line plot based on the index and sort it descending by the index.

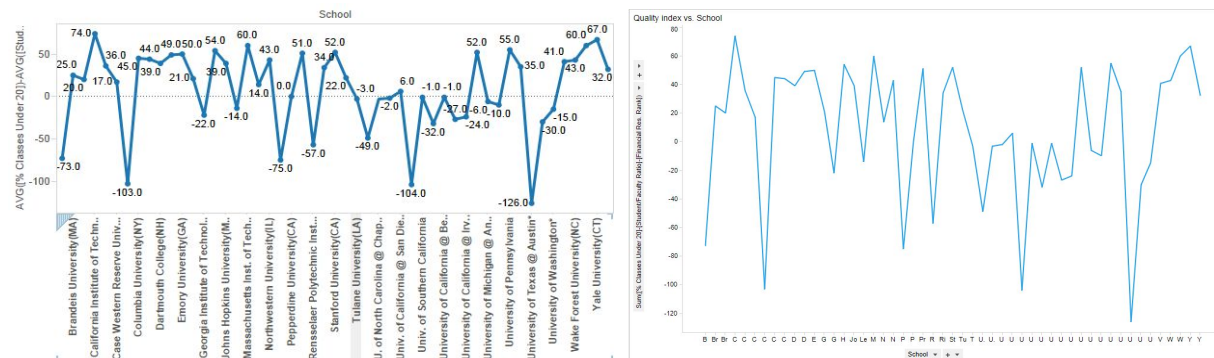


Figure 9 Quality index vs. School

5、As mentioned above, I'd like choose the value of top 10% of HS Class as the y-axis and schools as x-axis. And then draw an average line in the line plot.

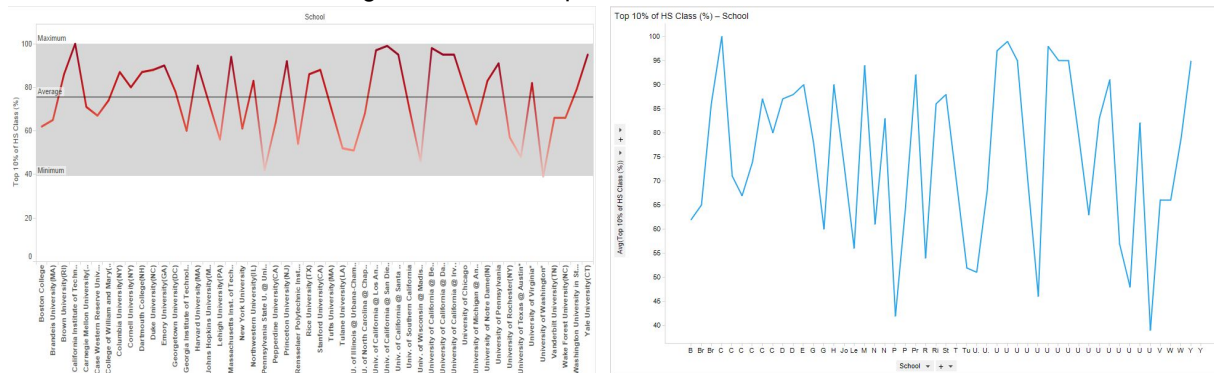


Figure 10 Variation of Top 10% of HS Class vs. School

Results:

With regards to nutrition data, I have found that lipid is related to energy, in other words, high lipid means much energy. The second finding is that more sodium means more ash as well. It seems that two lines from different tools are not similar. Actually it is the binning that causes the false appearances. Binning is a way to

group a number of more or less continuous values into a smaller number of "bins". In the third task, the result is that Cereals RTE, 100% BRAN,(WHEAT BRAN,BARLEY) is the one with most ash. According to the fourth diagram about nutrition, the best cereal for children's breakfast is Cereal RTE, general mills, fiber one. The final outcome of fifth task is that the value of ash in cereals varies a lot.

In terms of colleges information, the outcomes are that graduation rank has nothing to do with reputation, Faculty Residence Rank is not related to % Full-time Faculty, California Institute of Technology is the one with highest 25th SAT/ACT Percentile, and the value of Top 10% HS Class doesn't vary a lot. The most interesting finding is that is the best choice for students who require small classes and high educational quality.

Summary:

Based on the tutorial videos of Tableau and Spotfire, I can accomplish some basic tasks like drawing a line plot for relationship, a bar chart for comparison and color points or lines to differentiate all kinds of objects. However, when it comes to calculation, I felt difficulties to make use of values. For example, as for colleges information, students have to evaluate the quality via % classes under 20, students/faculty ratio and financial res. rank. However, these three factors cannot just be summed simply. Due to the different characteristic of student/faculty ratio, I have to subtract it from the sum of two other factors. To achieve this function, I can either create calculated field with programming language or input formula into the "Rows" blanket. What's more, the type of data is essential, because different types will cause different shapes of diagrams. I was deeply impressed by interactive diagrams from these two tool. For one thing, the tools can add some tooltips and details to the diagram automatically, which makes me understand data better and act upon that understanding. For another the relationship among different types of diagrams is amazing. For example, in the dashboard, if I select one point in the line plot, the corresponding bar in bar charts and the corresponding pie in pie charts will show up, at the same time, the other elements will dim.

During exploring datasets with Tableau, I often cannot put dimensions or measures into "Rows" or "Columns" blankets, and the warning "The field being added is much larger than the recommended maximum for this shelf. " often came out. However, if I drag the tag and drop it into the white field directly, it will work, which surprises and confuses me a lot.

As for Spotfire, I found "bins" when I drew a line plot for nutrition data. Every time I changed the value of "bins", it seemed that the line's shape changed but its trend didn't. After that, I realized the function of binning. Binning is a way to group a number of more or less continuous values into a smaller number of "bins".

In general, I think Tableau is better than Spotfire. Tableau has more powerful mapping and charting abilities, because it provides more options to display data, such as geographic maps, world bubbles, histogram and pie charts. Another reason is that it's more useful. We can customize the color, shape, size, details and tooltips, and then comparing multiple levels of information becomes easy. Furthermore, multiple data types and sources are acceptable by Tableau, and the superb dashboard is fairly extraordinary, because it is versatile and can be exported as PPT, which is much easier to tell stories about data. The last but not least, I am able to create new visualization using the "Show Me" tab easily. There are still some weakness in Tableau. During my exploration, I cannot put multiple metrics on the secondary vertical axis. Some may work, but much less than the one I expected. For Spotfire, it offers a wide range of formula types including statistical ones, which in turn allows getting more deeper insight into data. But combination chart doesn't allow stacked bars along with a line. And data transformation cannot be changed once implemented. If we want to change something like data types, we have to redo the entire data diagram. What's more, when I draw a line plot, I cannot sort data descending or ascending by the value on y-axis, which restricts the analysis.

In my opinion, for simple operations Tableau is much easy to use, but for large and complex operations Spotfire may be more appropriate. Also, Tableau is more suitable for reasonable denormalized data.