

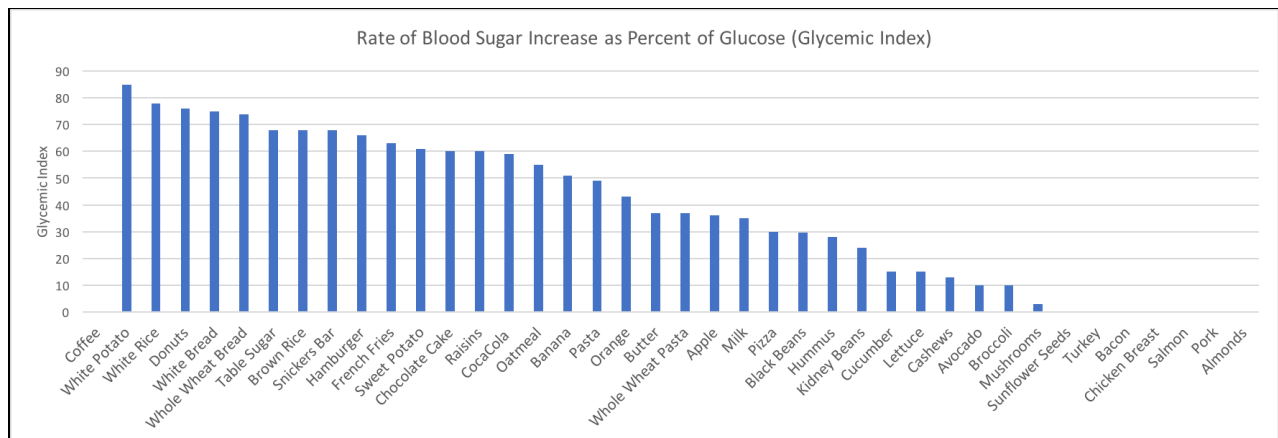
Using Calculations on Nutrition Facts Data to Compare Foods

Jesse Murray

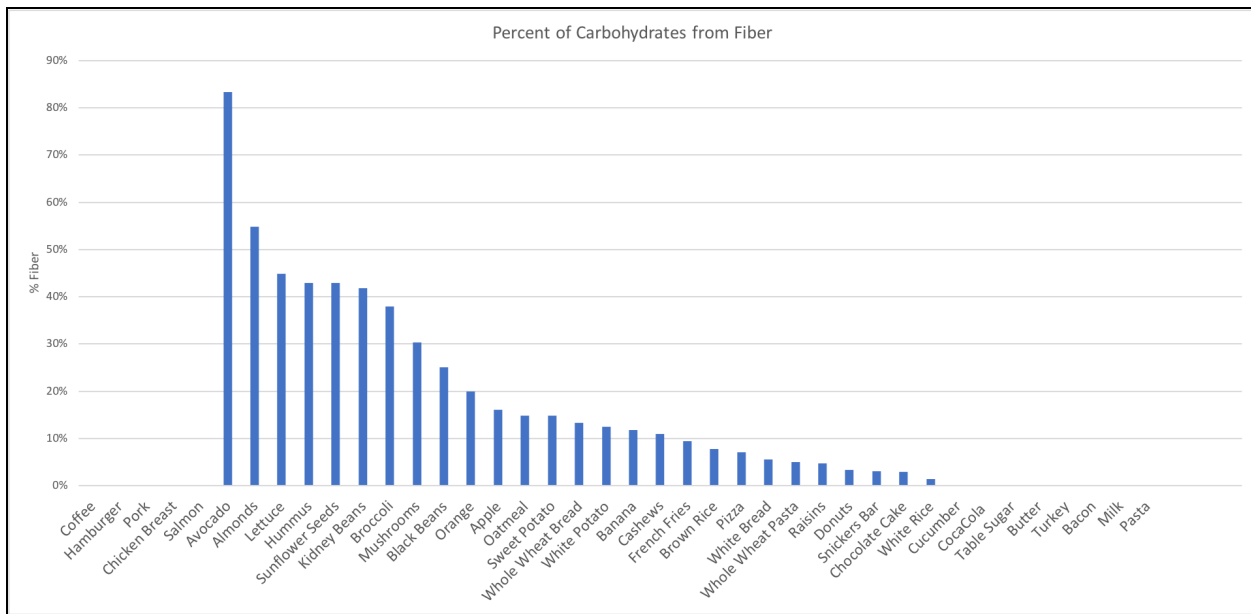
In this relatively simple data-analysis project, I compare different foods using USDA data and a variety of different sources for the glycemic index and price measurements. The metrics I use include price, mass, calories, total fat, saturated fat, polyunsaturated fat, monounsaturated fat, trans fat, carbohydrates, fiber, protein, and glycemic index. I then calculate an overall score for the healthiness of the different foods. One metric I often use is to compare foods using energy as a standard. The reason is that we each consume a consistent amount of energy per day, about 2,000 calories per day for the average person in the population, and thus you can use energy as a standard to compare how much good stuff/bad stuff or how much you'd have to pay for a days-worth of food.

Carbohydrates

Here we can compare the 'sugary-ness' of foods, which is measured by glycemic index. Some foods have zero listed either because they have no carbohydrates or because their carbohydrates are in very long, complex chains such that the glycemic index is zero.

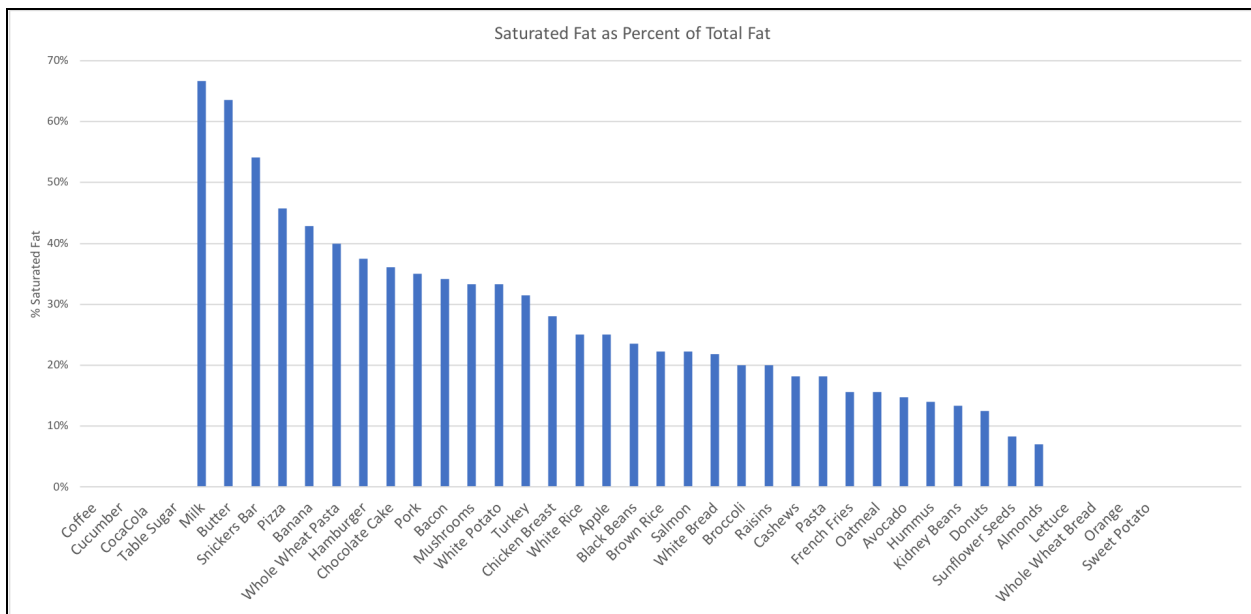


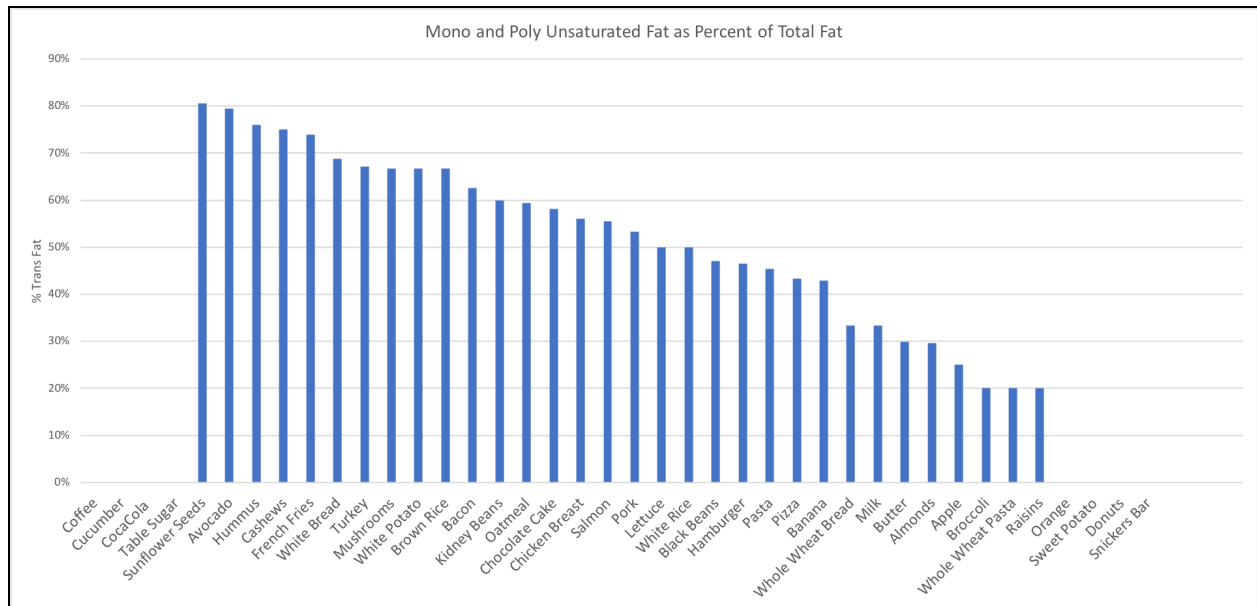
Here we can compare foods by the percent of the carbohydrates that are fiber, which is known to be a healthy carbohydrate.



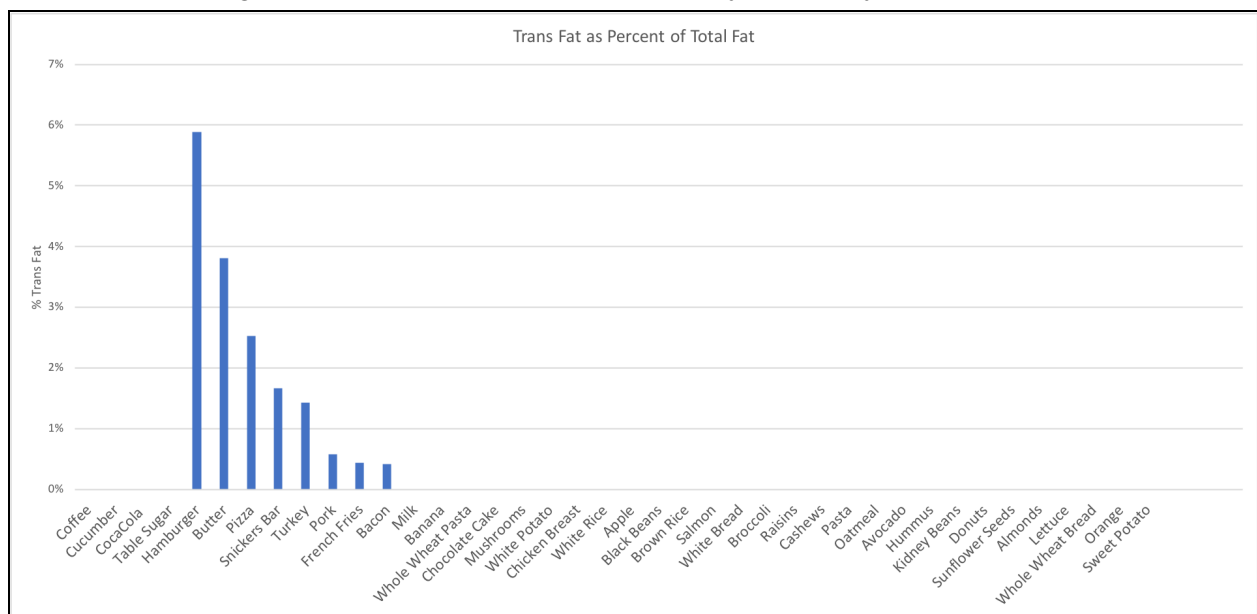
Fat

Similar, but with fat, we can compare foods by the percentage of fat that is saturated fat, often considered to be a less healthy fat, and the percentage of fat that is mono and polyunsaturated fat, often considered to be a more healthy fat.



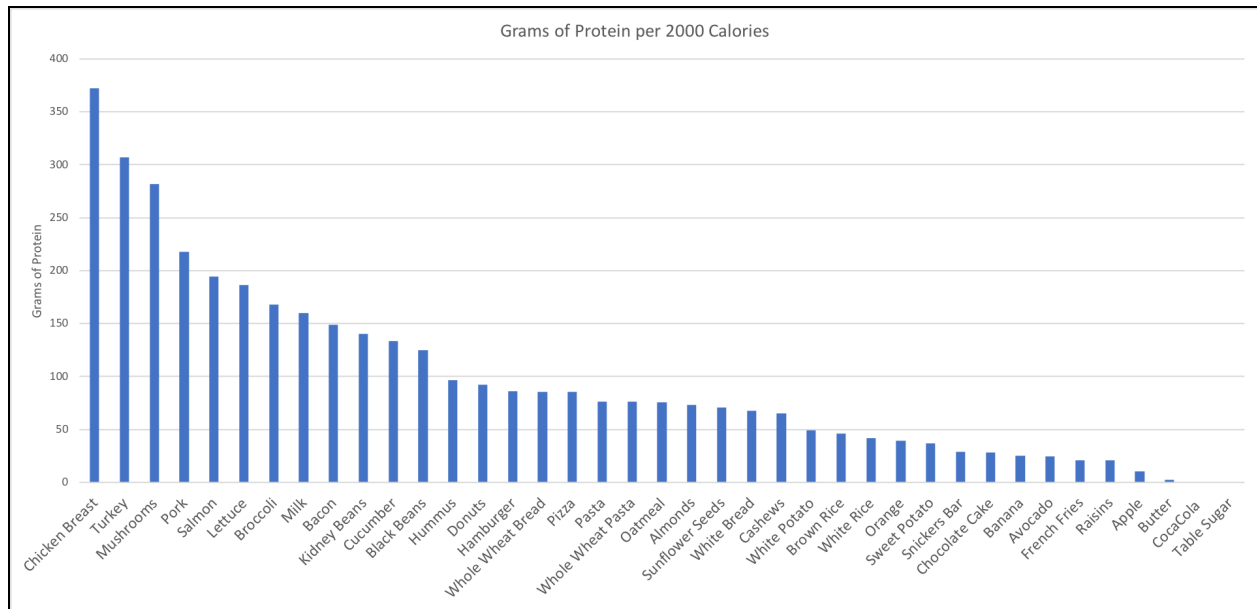


Also by the percentage of fat that is trans fat. Note that the USDA considers pizza and hamburgers to have trans fat, although I'm not entirely sure if this is the case for all produced pizza and hamburgers. Trans fat is often considered very unhealthy.



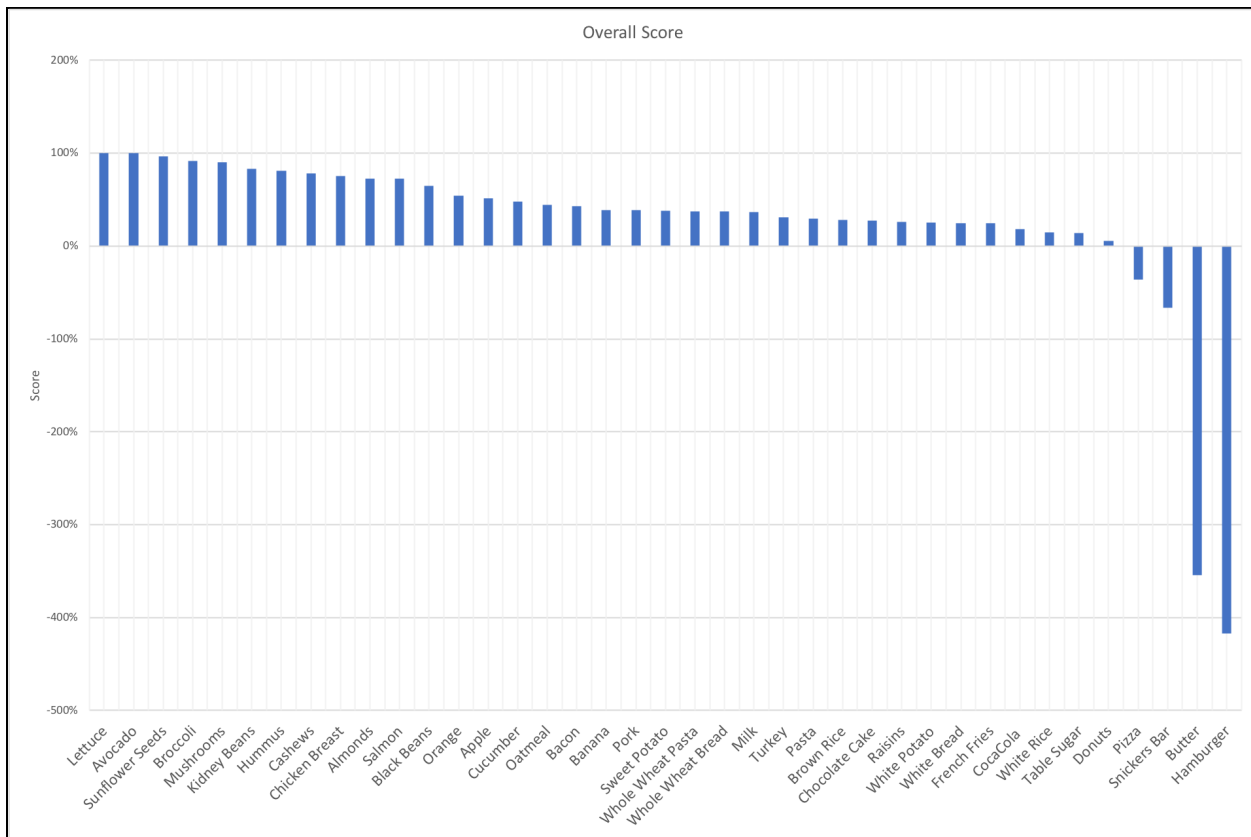
Protein

Here we can compare foods by the amount of protein provided by a days-worth of the food. It's surprising to see that there's a much higher protein density (grams of protein per calorie) in mushrooms, lettuce, broccoli, and milk than in hamburgers.

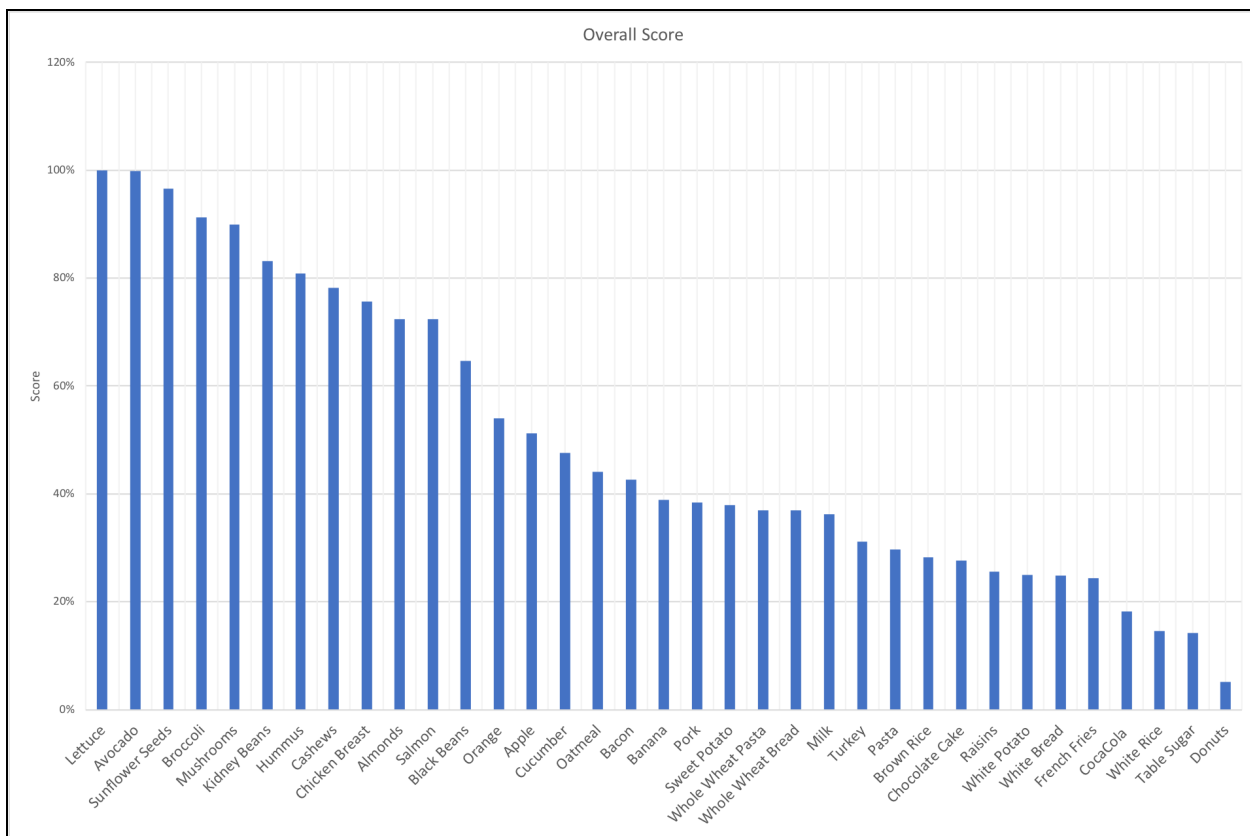


Overall Healthiness

Here we can compare the overall healthiness of different foods with the simple metric of adding - as a percentage of the maximum value for each of them - an inverse of the glycemic index (because lower is better), grams of fiber per 2000 calories, grams of mono and polyunsaturated fat per 2000 calories, grams of protein per 2000 calories, and subtracting from that grams of saturated fat per 2000 calories and ten times the grams of trans fat per 2000 calories (subtracting because those foods are healthy and thus should contribute negatively). Then, dividing that total value by the maximum. Using this method, we get the below plot. The negative values result from those foods having trans fat. We can show the plot more clearly by removing the negative values...

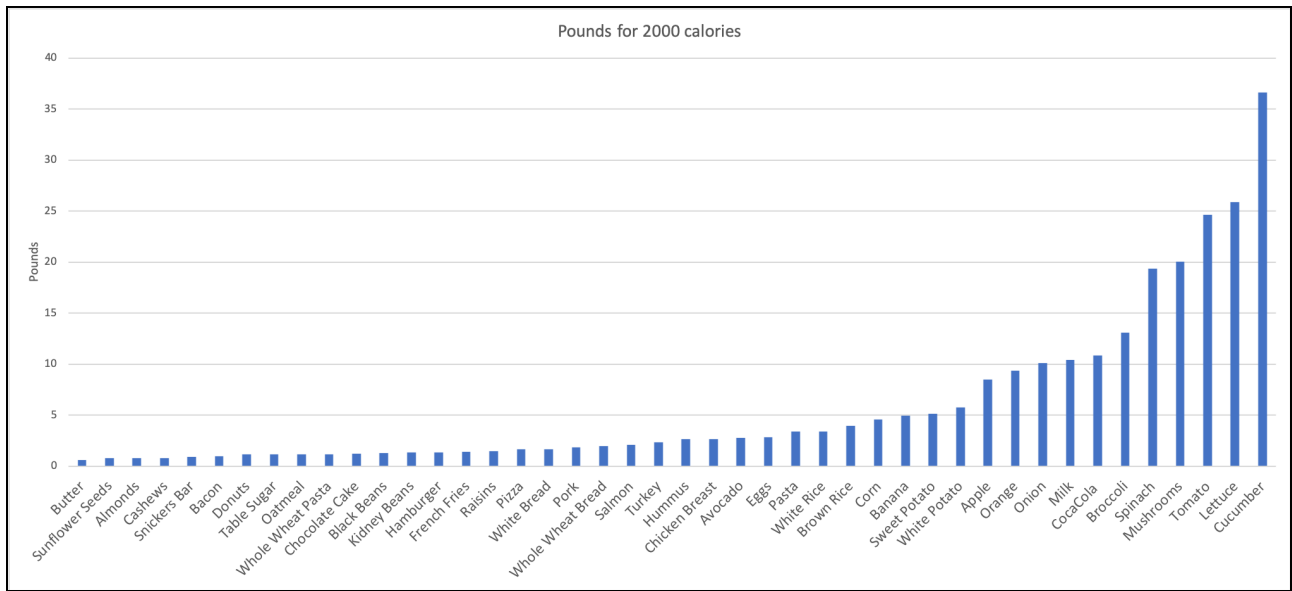


Here is the graph of the overall score with the negative scores removed. In many ways, the overall score seems to confirm basic intuitions about healthy versus less healthy foods. Note the meat foods: salmon and chicken, in the top 15 healthy foods. It would be interesting to compare the average overall score of the different food categories: fruits, vegetables, grains, nuts, etc.

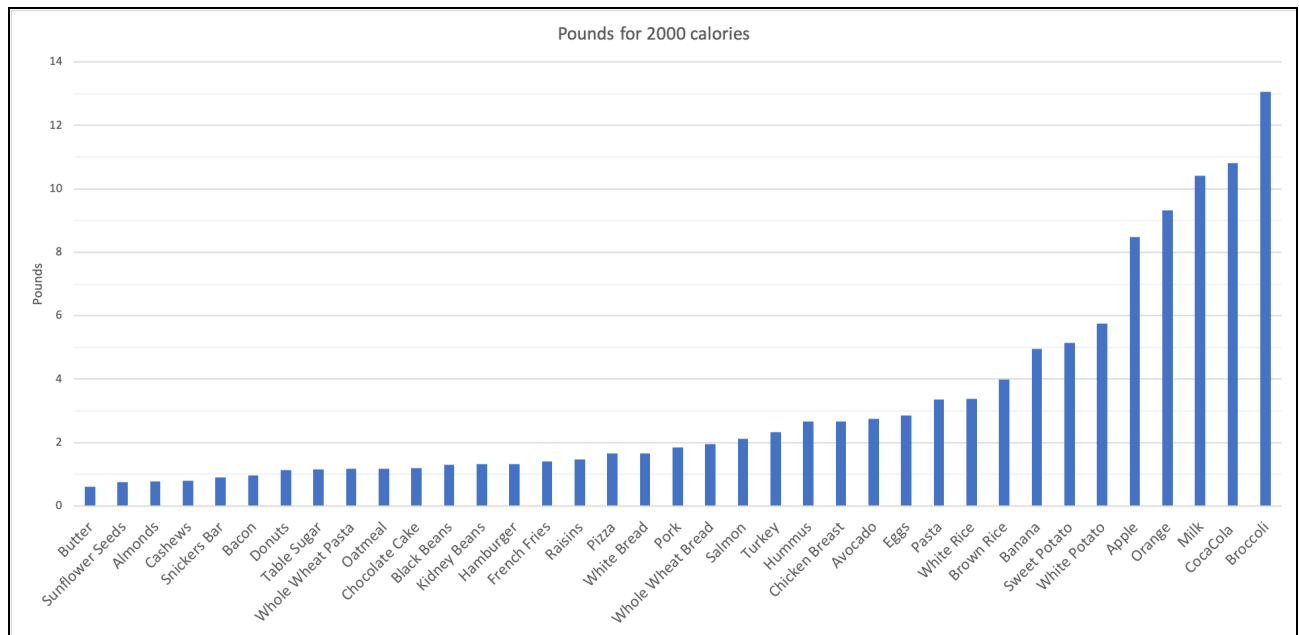


Energy Density

Here we can compare the energy density of the different foods in terms of the number of pounds for a days-worth of food. You can pretty clearly see a major problem for some vegetables, which is that they have very few calories. To exist off of cucumbers you would need to eat about 36 pounds each day, which is clearly undesirable. We can make it easier to compare the foods in this graph by removing the very large values...

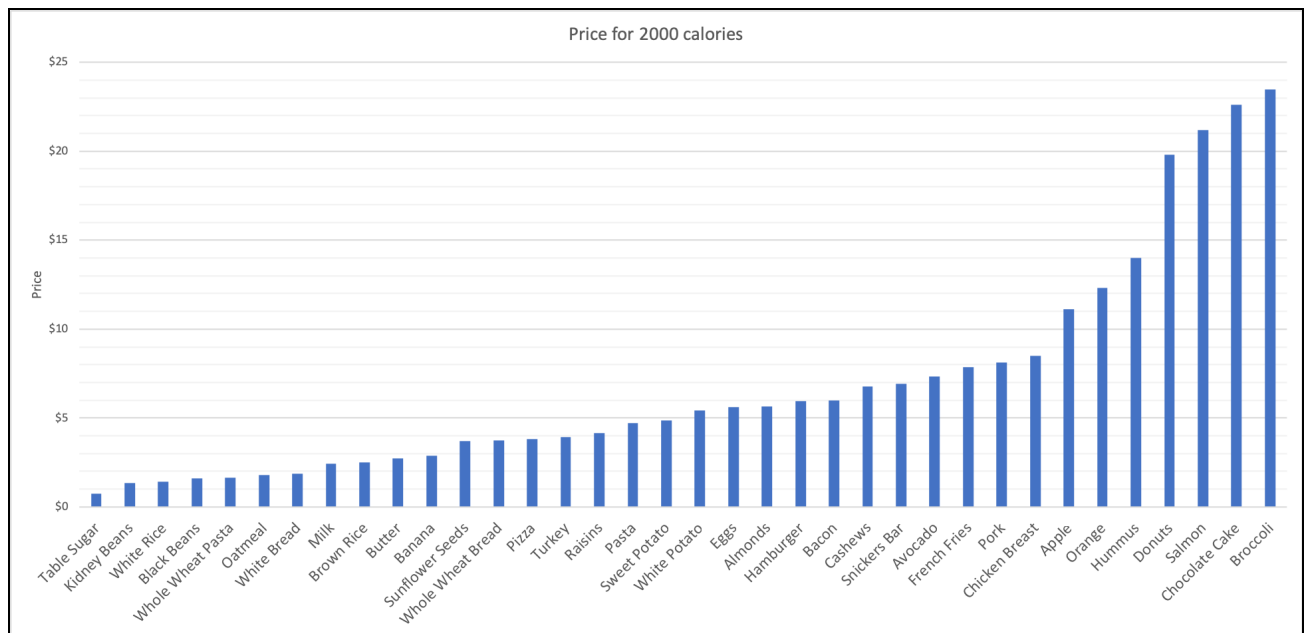
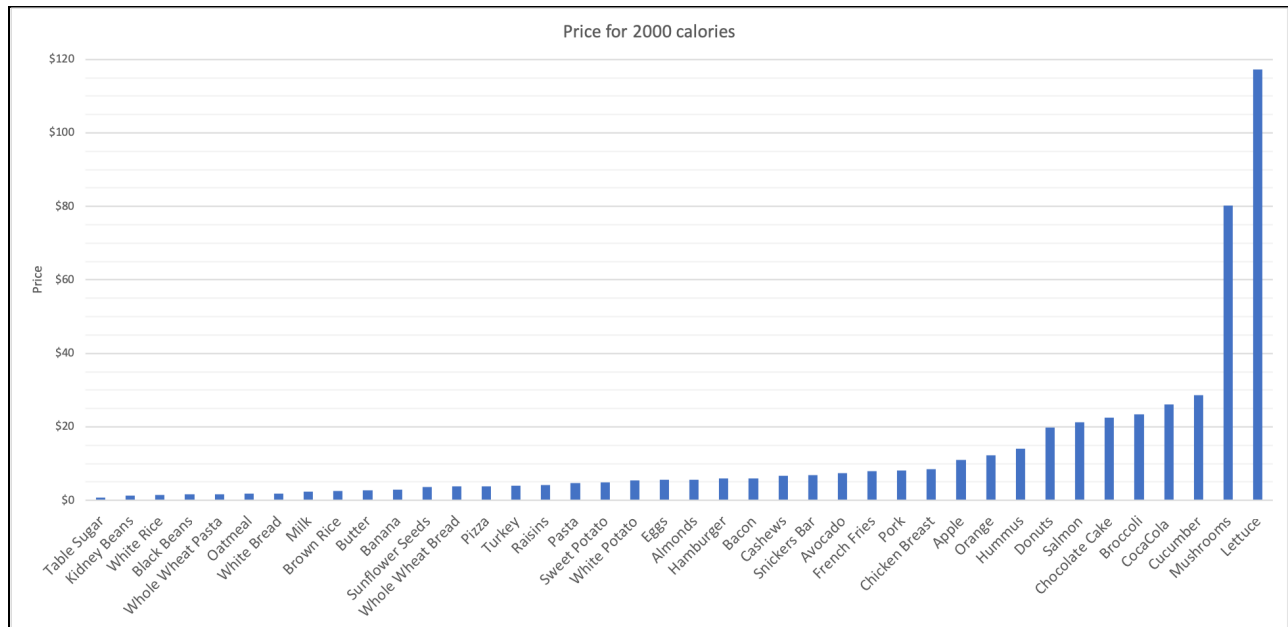


This is the same graph as before, but with some of the very large values removed. Note that if you were going on a backpacking or hiking trip, you would want to bring foods that weigh very little for the same number of calories. That is, you would want to bring the calorie dense foods more towards the left of the graph such as seeds, nuts, and oatmeal.



Cost

Finally, we can compare the price of different foods by considering the cost of a days-worth of the food. You can see that a days worth of food in lettuce would be highly expensive but a days-worth of sugar, beans, pasta, and oatmeal would be rather cheap. This graph can be made easier to read by removing the very large values...



This was a rather simple project. It would be interesting to revisit the project using *all* the foods in the USDA nutrition facts dataset, finding correlation coefficients between different metrics, dividing them into food types, and applying statistical or machine learning techniques.