1. The first thing required for this project is to gather the data. Fortunately for us, the USDA National Nutrients Reference data was available in an *almost* convenient format from https://github.com/schirinos/nutrient-db

2. After the git repository has been pulled or downloaded, the second sub task was to extract a JSON file from the nutrients DB file. This was also convenient because there is a python file (which requires the installation of pymongo) that does this for us with a single command, namely, nutrientdb.py -e > nutrients.json

3. The most challenging part of this program was to load the data. There are two ways to do so:

- By using Pandas' inbuilt reader, by calling read_json, with the most important parameter, called as line=True, which reads every individual line as a spearate JSON
- The other method was to use a 'with open(...)' block and read the line 1-by-1 adding it to the DataFrame

Even though the Project Description states using the second method, the first method was applied because it was much more effortless. In a research environment, where one has to produce 'alphas' in quantity and quality, which portfolio managers can utilize, cutting down on non-value adding activities such as reading line by line, then converting to a DataFrame is much more desirable.

4. Also, using the in-built reader not only helps us avoid convoluted code of reading JSON line by line, it also helps us get up an running with Step 2 of the project, that is, loading it into a DataFrame consisting of Food Names, Groups, IDs and Manufactures, since we already have a superset.

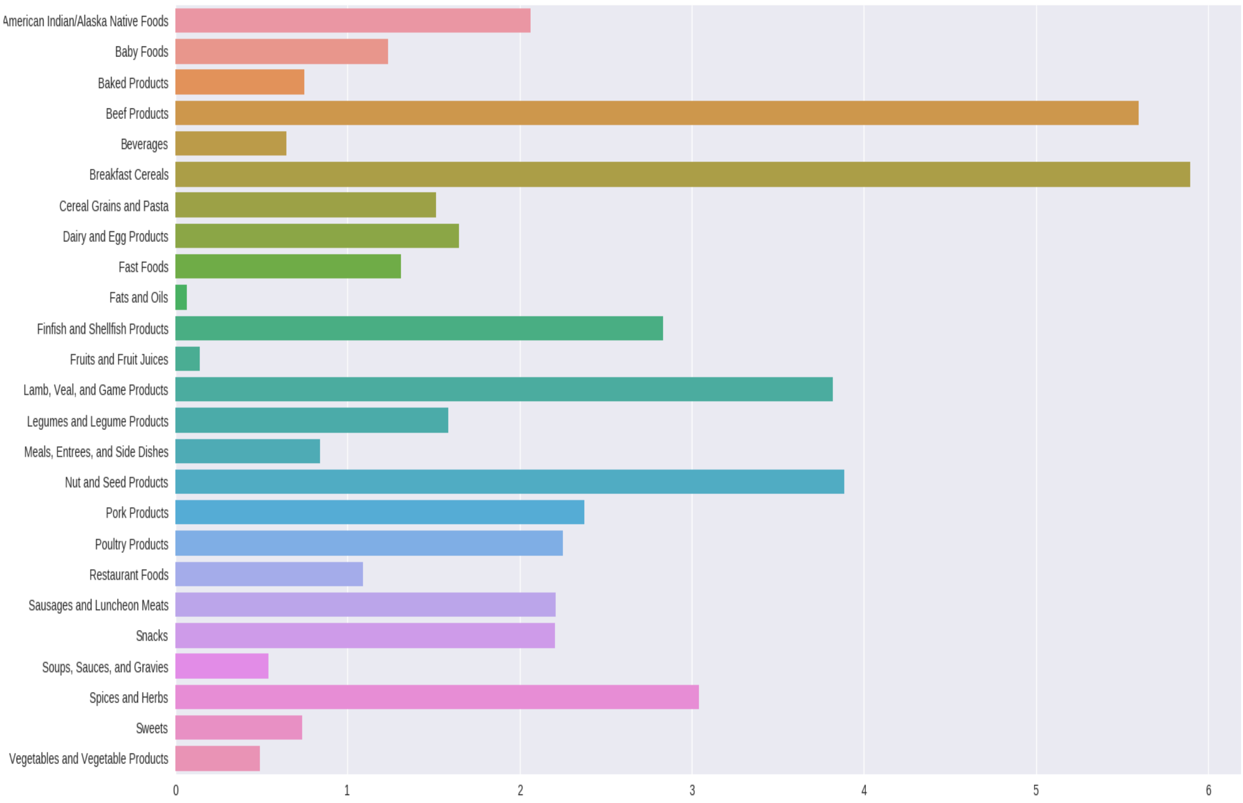
5. The purpose of this project was to bring some structure to semi-structured data (JSON), so that we could not only navigate it easily, but analyze it with ease too. Also, it proves that not necessarily we may be able to convert all the semi-structure to a structured dataset, but will have to deal with the nested semi-structured data along the way. This helps us understand that Data can be in any form and one must be prepared to do Data Massaging and Wrangling to get the relevant fields and analyze them

6. The amnio acids are stored nested deep within the initial DataFrame which we create, in the form of a Dictionary. Therefore, we create a mapping dictionary, that uses the information of all the amino acids from Wikipedia - https://en.wikipedia.org/wiki/Amino_acid#Table_of_standard_amino_acid_abbreviations_and_properties

This gives us the list of Amino Acids which we can use for our Analysis. This also proves that Data can be fetched from other sources, and it's not necessary that all the information may be contained in our initial source. Such 'assistive' data (also called as Metadata), helps in bringing more meaning to the Data which is under analysis.

7. By using a series of for loops, we map the Data of Amino Acids to the Food Groups. We then tabulate it in the form of a DataFrame, and just print the initial 5 entires and the last 5 entires. We also spit out a CSV so that the tabulated data can be analyzed further. This maintenance of state is sometimes essential, because the O(n^2) operation on roughly 8.8k Dataset can be taxing on time and computation resources. Next time, we can just load the CSV (though a pickle format would be faster, a CSV is preferred in case a non programmer is looking at the dataset)

8. We do similar for-loop searching and mapping for gathering all the Data about food groups that contain Zinc, and add its values to the map. Later we construct a DataFrame out of this map, so that we can do central tendency analysis on this data. We demonstrate a small analysis by calculating the median for each Food Type and visualizing it to make further sense.



From the image (appears blurred in the doc, but fine once the program is run) we can infer a lot of information, such as Beef and BreakFast Cereals could easily be recommended to people who lack Zinc content in their body, and not Fats and Oils, since they give the least amont of Zinc. Also we can infer that Non-Vegetarian foods (Such as Beef, Lamb, FinFish etc) tend to give more Zinc content than Vegetarian food such as Fruits and Cereal Grains and Pastas. However, we do have outliers in the form of BreakFast Cereals and Nut and Seed products.