**Problem Statement**

**Organization Name:** U.S. Department of Agriculture FoodData Central (FDC)

**Challenge Name:** USDA FDC

**Difficulty:**Level 2: Participants with some data analysis background.

Background

For more than 100 years, USDA’s Agricultural Research Service (ARS) has analyzed foods and determined nutrient/component values for foods consumed in the U.S. To continue to provide transparent analytical data to Americans, USDA’s FoodData Central (FDC) was launched in 2019. FDC is an integrated data system that provides expanded nutrient and food component data in one location. This database provides a snapshot in time of nutrients and other components found in a wide variety of foods and food products. FDC includes five distinct data types, each of which has a unique purpose and unique attributes. For this project, we will focus on and compare two of the data types, Foundation Foods and SR Legacy.

**Foundation Foods** includes values derived from analyses for food components, including nutrients on a diverse range of foods and ingredients as well as extensive underlying metadata. These metadata include the number of samples, sampling location, date of collection, and analytical approaches used. These values are the most current dataset and include a min, max, and median value for each sample analyzed.

**SR Legacy** has been the primary food composition data type in the United States for decades. It is considered historic food composition data and will not be updated further. It provides a comprehensive list of values for food components, including nutrients derived from analyses, imputations, and the published literature. This data type has provided the values for most other public and private food composition databases and has supported a wide range of public policy initiatives, research studies, and diet planning and education activities. These values were generated before 2019 and will have a mean value and depending on the number of data points may have a min, max, and standard error.

As analytical sampling continues and resources and staff are becoming increasingly limited, FDC is establishing new ways to incorporate data. The old paradigm of careful targeted analysis of a complete set of nutrients does not work at a time when the analysis of 1 food can exceed $50,000. It would be beneficial to look back at historical data and see how the data has changed and remained the same. This insight will provide a better understanding of which nutrients or components should be analyzed and which ones remain constant.

# Questions

Have food components changed over time for equivalent foods in the two data types?

Is there a difference between food components recently analyzed in Foundation Foods compared to matching foods in the historical SR Legacy dataset?

When comparing nutrients/components in equivalent foods from the two different data types, does the mean value for SR Legacy fall between the minimum and maximum values of the Foundation Foods?

Which nutrients/components fall out of range the most and which fall in range more frequently?

The dataset can be filtered by food categories as well.