Are you what you eat?

Allie Collins & Erika Tyagi

Our aim is to explore the nutritional profiles of Americans and their implications

We will use dimension reduction and clustering techniques to group nutritional profiles

We will subsequently use the output to try to answer the following two questions:

- How does nutrition relate to socio-demographic profiles?
- How does nutrition relate to health outcomes?

We will leverage NHANES survey data from the CDC

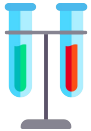
The National Health and Nutrition Examination Survey (NHANES) is a biannual, nationwide survey intended to study the health and nutritional status of adults and children in the United States

It involves five components:

Demographic

Dietary

Examination



Laboratory

Questionnaire

Methodology

Step 1: Run a sparse PCA model on our data (specifically, the dietary information)

Step 2: Run a hierarchical clustering model to gain information about the likely number of clusters

Step 3: Run partitioning clustering algorithms and internally validate

Layer on information from the other survey components, to separately investigate:

1. Socio-demographic associations –
   In the different clusters, what is the composition of age, income, race, etc?
2. Health outcomes –
   In the different clusters, what illnesses are prevalent?

Next steps

Based on the outcomes of conducting the analysis as laid out on the previous slide, we will consider additionally working through a supervised classification problem – are nutritional profiles helpful in predicting particular health statuses?