Obesity in Chicago and its relation with its venues

2. Data

Based on our problem, we have to consider the following parameters:

- Percentage of Obesity in each Community Area
- Which are the most popular venues per area regarding food and exercise
- How similar are the community areas according to their most popular venues
- How similar are the results from community areas' venues to the percentage of obesity per community area

We will need to use **data that describes the percentage of obesity per Community Area**, that is going to be retrieved from the **Chicago Health Atlas.** Data should come with the name of the community area, and the obesity percentage, **e.g. Community Area Name: Englewood, Obesity Percentage: 30%**

Then, we need to **retrieve the most popular venues per Community Area, as well as it's category** to see if the type of food, or place for exercising can affect the obesity rates. For obtaining this information we will use Foursquare API. In this case, we will only consider two types of venues:

- Food
- Outdoors and Recreation

The idea is to retrieve the 100 most popular venues with their name, category, latitude and longitude **e.g. Name: Hai Yen, Category: Vietnamese Restaurant Latitude: 41.97328297968702, longitude: -87.65705585461119**

Having this information we can compare the different categories and use the hierarchical clustering algorithm to **verify how similar are the community areas and if it exists a relation between the category of food and Outdoors and Recreation venues, and the obesity percentage per community area.**

We will also **use geographic data in GeoJson format to display the geographic limits of community areas, and use a palette to display colors according to their obesity percentages.** This data can be found in the Chicago Data Portal. For obtaining longitude and latitude per Community Area we will use Nominatim API.