## Assignment 3 part 2 (50 pts)

## due October 25th 2022

## Project Preparation

We will prepare the project in 4 parts developed in the assignments. In this assignment we start with part I and ii, the first step is to make your selection on area and data. Keeping in mind that for the purposes of the class you need to adapt your expectations and interests to data available.

**Part I: Search and Selection (30 pts)**

In urban planning and policy, a ‘wicked problem’ is a problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize.

While data science informed policy does not bring the ultimate solution, it, however, can help. Identify an area of your interest to do a data science research project for this class. Examples are vulnerability of transportation infrastructures, housing crisis, traffic prediction, epidemic spreading, emergency responses, adoption of sustainable technologies, among others.

Motivate your research project answering the three questions below :

* What is the problem?
* Why should we care?
* What do you want to do applying data analysis and modeling? (identify your data sources, ask us in Ed Discussions what you are looking for)
* Do you already have the data sources to pursue your idea?

Studies of “food environment” using geographic methods have revealed disparities in exposure and access to food. GIS methods have been popular to analyze food access using five dimensions: availability, accessibility, affordability, accommodation, acceptability. Availability and accessibility have been well-captured using geographic methods, but by using open data (e.g., OpenStreetMap) there may exist ways to apply a simple methodology at scale to improve upon measures like those gathered for the USDA’s Food Atlas. The Food Atlas is a national mapping tool that reports statistics related to accessibility and availability, but the geographic granularity is relatively coarse as it aggregates data by county, which may not capture the full variability in food environments within a single given county. Moreover, there may exist proxies for dimensions of affordability, accommodation, and acceptability via data from reviews and ratings (e.g., Yelp data).

We care about this project because developing a flexible tool that relies on open data can provide information to the public and policymakers about the quality of food access to different people. It can inform public health measures related to diet, exercise, urban planning, and land use, since it is likely that some populations and demographics are not being served as well as others by their food environments. In general, we already know that diabetes and other health conditions that may be influenced by diet disproportionately impact Americans of minority racial groups, and much of this may have to do with food access, as already proposed by the now-dated concept of food deserts.

Since the literature is mature on this topic, there are newer concepts to pay attention to that we could incorporate into our mapping tool, well summarized by [Brookings Institute](https://www.brookings.edu/research/beyond-food-deserts-america-needs-a-new-approach-to-mapping-food-insecurity/). For instance, it has been shown that people typically purchase food not near their home but near their workplace, so incorporating employment data and home–work flows is important. There is also the claim that mapping technology is limited by the edge effect, but we aim to prove this to be untrue with our mapping tool, which is why we use spatial radius from any given point rather than being confined to a single city/county. Newer discourse also encourages foodscape mappers to incorporate less-captured food security strategies, such as farmers’ markets, community and school gardens, and small grocers. While the data may not be as thorough for these destinations, there is no reason that volunteer mapping (such as through OSM) cannot account for these, so our tool would be robust to future incorporation of these map features into OSM.

We have data from OSM to locate all grocery stores, convenience stores, fast-food restaurants, etc., within our areas of interest. We also have access to Google Maps API to provide routing information between neighborhoods and these points of interest.

**Part IIa: Data Science Story (now or more guidelines later based on Part I) (10 pts)**

Select a research paper related with the selected dataset. It does not have to be exactly the same data. It is only important that you read and refer here something published about an analysis of the similar kind of data or method you choose. After you select a data set of your interest, we can help you to find an appropriate article for you to write about in your project written report.

**Part IIb: Data Analysis (more guidelines later based on Part I and IIa) (10 pts)**

Find a data set to work with it and describe it here. Can you supplement the data with a model?

**Part III: Visualization (more guidelines later based on Part I - III)**