Project subject area:

Obesity in the United States. I want to explore what geographic areas in the US are obese and attempt to ascertain what factors may be to play.

Data sources

Can be used via Api:

National obesity percentages by state.

<https://catalog.data.gov/dataset/national-obesity-by-state-d765a>

Csv:

National data on Obesity risk factors

<https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system>

Data chart:

States obesity ranks

<https://en.wikipedia.org/wiki/Obesity_in_the_United_States#:~:text=Over%2070%20million%20adults%20in,%2Dadjusted%20rate%20of%2041.1%25.>

Relationships

I will pull this data together by linking the state involved. This will require some trimming of certain data sets that are international vs just us based but for the scope of this project we will just be focusing on the states withing the U.S.

Project Approach:

I plan to do as I would with any data project. First, we will gather and clean the data, next we will dig in and explore the data to make sense of it. Then build a model or visualization to represent the data and its findings. That said, I do not know how best to do this using python, so that portion will take some time to figure out.

Concerns/Challenges:

I am concerned with getting data that does not seem consistent with each other. utilizing data from multiple sources there is a chance we will see the data look differently based on how it was gathers and other factors in each data collections gathering process but hopefully when we get to the data exploration phase this will become apparent and we can evaluate options at that time.

Ethical implications:

I think discrimination will be the largest ethical issue. This has a chance to become apparent in the data that one or more groups of people are reflected in a poor light in these data sets. For research purposes this is not a big deal but for instance if a certain demographic group or geographic location is more prone to obesity and say an insurance company utilized this data to set rates for those groups higher based on these finding that could result in some level of discrimination.