Proposal:

The goal of our project will be to analyze the relationship between obesity rates and various socioeconomic factors. Obesity is associated with a range of chronic health conditions and conducting research is essential for understanding what could potentially be contributing to the epidemic and how it could be improved. Utilizing publicly available datasets from [Data.gov](http://data.gov/) and Kaggle, we will be using a structured data science workflow to explore how factors such as median state income, or the number of fast-food restaurants per state correlate with obesity rates. We will be collaborating through slack, and a shared Github repository.

First data set .csv: Fast food restaurants in the United States

<https://www.kaggle.com/datasets/thedevastator/fast-food-restaurants-in-the-united-states>

Second data set: JSON: National Obesity by State

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

\*\*(<https://cran.r-project.org/web/packages/USA.state.boundaries/readme/README.html>

R data package contains a map of the USA, could possibly use latitude and longitude)\*\*

Action Items:

1. Create a data set with full State Name, it’s abbreviated name and the population for each of the states. Seyi “The Creator” will be taking on this role.
2. For the fast food restaurant data:

This data set has abbreviated state names, use Seyi’s dataset to replace them with their full names (NY -> New York). This will make it easier to work with the National Obesity dataset. Adriana will do this.

Write code that counts how many times each province/state occurs.

Final outcome:

2 columns: 1. State and 2. number of fast food restaurants

Mutate dataset to include a third column: fast food restaurants per capita.

Adriana will do this.

1. For the Obesity JSON file:

Use jsonlite to convert the JSON-formatted data to a data frame

Keith: Can extract the obesity percentage from the JSON

1. Merge both Obesity percentage Final and the Fast food Final datasets