**Choropleth Maps of CDC Data**

**Project Proposal**

The objective of the project is to generate choropleth visualizations of various health-related data. It is anticipated that typical visualizations would entail shading collections of ‘regions’ in accordance with shading color determined from our quantities of interest. Here are some examples:

a) Shade all the states in the US in accordance with average life expectancy.

b) Shade all the counties in a given state in accordance with average income.

c) Shade all the census tracts in a given county in accordance with per-capita murder rate.

A census tract is a geographic region defined for the purpose of taking a census. The underlying region will tend to have a population of between 2,500 and 8,000.

Potential quantities of interest might also include; averaged socio-economic status, education level, ethnicity/migrant status, per-capita HIV rate.

**Data Sources**

There is a wide variety of health-related data available from the CDC website. By way of example, life expectancy estimates are made available through a SODA API call to the CDC website - <https://data.cdc.gov/resource/5h56-n989.json> which returns life expectancy data in JSON format for census tracts within each county for the majority of US states.

There are two sources for the geometric data required to define and shade the reasons. Eric Celeste's website is: <https://eric.clst.org/tech/usgeojson>. It contains state and county boundary data stored as polygons in GeoJSON format. The Cartographic Boundary Files supplied by the United States Census Bureau contain the polygons that define the boundary of the census tracts: <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>. The data stored on this site is not typically in GeoJSON format. However, it can be easily converted using the easy-to-use website Also, data has been sourced from the Cartographic Boundary Files supplied by the United States Census Bureau <https://mapshaper.org>.

**Output**

The objective is to develop a Flask application that uses relevant information found in a pre-populated database (possibly MongoDB or Postgres) in order to supply GeoJSON data to Flask routes. This information can then be processed using calls to a JavaScript library such as D3 and Leaflet.

For a given quantity of interest (for example life expectancy), the application will provide the user with drop-down menus containing for example; lists of relevant states, lists of relevant counties, lists of census tracts. For instance, selection of Mecklenburg County , NC will provide a visualization of census tracts that would look like the following: albeit with different colors if a variation in life expectancy is reported !

A close up of a map

Description automatically generated

The following illustrates what the counties of North Carolina look like after plotting using geopandas. The data (GeoJSON format) came from Eric Celeste’s website.

A close up of a map

Description automatically generated