Brian Slobotsky

GEOG788P Initial Project Proposal

My project is based on a paper published this year in Environmental Research Letters by Lin et al entitled "Food flows between counties in the United States." The paper used national and subnational data sources to create a model which spatially downscaled commodity flow data to the county scale. The data in the paper comes from Oak Ridge National Laboratory's Freight Analysis Framework (FAF) and the Commodity Flow Survey (CFS) from the US Census Bureau and US Department of Transportation, among a few other data sources. The FAF data describes freight flows between metropolitan areas. Meanwhile, the CFS data is aggregated at a national scale, and the only spatial element is distance. Lin et al used these two sources, supplemented by several other data sources, to create county flow estimates. For my project, I will use a 12.45 MB .csv file provided with the paper. This .csv file contains the county flows, identified by two columns of "origin" and "destination" counties, using US Census Bureau FIPS codes. The .csv also contains a breakdown by Standard Classification of Transported Goods (SCTG) categories 1-7 and a total value. The benefit of using this data in my project is that the data are already tidy, not in need of processing, and in a format that can be converted into spatial data via FIPS codes. Some limitations are that it is not possible to explore the data by commodity beyond SCTG codes and map the flows.

What I intend to produce is an interactive map with the folium module that has an interface where a user input can narrow the flow map to a single county (by click or by query) or a SCTG category. With folium, I can use widgets like the slider (for sliding between SCTG categories) and velocity (to give a "breathing data" appearance). A display of statistical information related to distribution would be an interesting exploration for possible inclusion.

Additionally, I intend to use exploratory spatial data analysis to investigate the spatial interaction of the model. This includes looking at the spatial weights, lags, and meaningfulness of the residuals.

I anticipate using the folium module the most, for interactive mapping via Leaflet, widgets, and potentially hovering. I will most likely use pandas and geopandas in order to prepare the data for being mapped. There is a potential need for the cenpy module because the data contains Census Bureau FIPS codes, and I can easily load a map of US counties with cenpy. I might use arcpy for some GIS functions because I am familiar with it. Finally, I could use the palettable module for different color palettes in my final visualization. I do not believe I will need to write custom code for this project.

There will be challenges working with the visualization. The mapped flows might be difficult to display in terms of line widths between counties. In addition, because this will be an interactive map, visualizations at smaller scales will appear differently than at larger scales. Another challenge is the difficulty of visualizing intra-county flows, not only from the standpoint of scale limitation, but from a visual standpoint. I expect to have issues with the “hover” functionality in folium, which seems to be quite difficult to coerce.