**Mapping Ancestry and Restaurants in New York City**

Anqi Li, Yue Guo

**Description**

We built a webpage and several graphs to explore people’s reported ancestry in New York City neighborhoods and the locations of restaurants labeled as serving that group of people on Yelp to see if there is a correlation between the two. For example, we wish to explore if restaurants labeled as “Russian” on Yelp are primarily located in regions with significant population of Russian ancestry.

The sources of our data are:

1. People’s reported ancestry by census tract from ACS data, pulled using the Census API.

2. Restaurants labeled with specific categories from Yelp, pulled using Yelp’s API.

**Data Collection**

We downloaded relevant data from the two APIs. We cross-referenced the available ancestries in ACS with the available category labels published in the Yelp API’s documentation. In the end, we found that the following groups/categories appeared in both documentations:

*Russian, Ukrainian, Polish, German, Italian, Romanian, French, Arabic*

However, after some trials we found that there are no restaurants in New York City labeled as “Romanian” (the Yelp API returned 0 results), so we dropped Romanian as a category. We also found that the Census API was very slow at downloading data for some unknown reasons, with or without geographies, so in the end we downloaded ACS data using the API in R instead.

**Data Manipulation and Display**

We first joined the ACS data (in spreadsheet form) to the shapefile of census tracts.

We also performed a spatial join using the GeoPandas package to aggregate tract-level demographics data to neighborhood level.

We created a folium layer for each ancestry group. We displayed the neighborhood choropleth based on what percentage of the total population in that neighborhood has reported an ancestry of each group.

Restaurant data pulled from the Yelp API was added as dots, with one folium layer for each category. The icon of each restaurant was color-coded based on the star ratings of that restaurant on Yelp.

Finally, we created several plots showing the star ratings of the restaurants in relationship of their price level (the number of “$” signs) as reported on Yelp.

**Results**

Overall, the Yelp API seemed to report significant fewer restaurants than we expected. In addition, the limitation of folium (being unable to group layers) also limited map visualization.

That said, we found that the location of a few ancestry groups, namely people with German and Polish ancestry, highly correspond with the location of restaurants. For example, there is a high concentration of both Polish people and Polish restaurants in Greenpoint.

We believe that the popularity of French food led to the fact that French restaurants are distributed throughout Manhattan.

Of all the racial groups, we found that there is significant concentration of people with Italian ancestry on Staten Island, but there are no Italian restaurants on Staten Island. In fact, all Italian restaurants are concentrated in Manhattan. Perhaps the business nature of Manhattan as well as the popularity of Italian food (at least when compared to Arabic) led to the concentration of such restaurants in the area.