Spring 2021: Applied Spatial Econometrics Midterm Exam (Due 03/28 @ midnight)

For the 3 questions below, in each case you should submit a map and the code used to generate the map along with any interpretations as they are called for. All three questions focus on the use of R and Python to create simple choropleth maps and, in the case of the final question, to produce new spatial data for visualization and further analysis.

You are welcome to use the classroom on 03/25 to work as a group to troubleshoot issues and questions associated with the exam. However, everyone should work through the mechanics of code development and implementation on their own computers.

For Q1 and Q2, I have given you the raw data from the Robert Wood Johnson County Health Reports. Note that the data are on multiple tabs in the excel sheet and that you will need to make use of data from different tabs to complete the exam.

Question 1:

Using the CA County Health Rankings Data (2018 County Healt h Rankings California Data - v3.xls) create a multi-panel including figure of the following deliverables:

- A county level map of "Health Factors" by Rank, where a higher score means healthier .
- A county level map of "Residential Segregation; Black/ White", where higher score means more segregated
- A county level map of "Income Inequality", where higher scores mean more inequality
- A scatterplot (with a fit/trend line)of "Income Inequality" by " Residential Segregation; Black/White"

To complete the multi-panel figure, use the ggplot2 package to create each item and present in a single figure that includes 1) the paneled maps/graphs, 2) title and attribution, 3) subtitles, and 4) legends with meaningful titles and classification labels.

Question 2:

Use the geopandas and matplotlib packages in Python to produce a 6 panel map of the following “Ranked Measure Data” by county in CA indicating the % of the population with the following characteristics:

1. Poor or Fair Health
2. Adult Obesity
3. Physical Inactivity
4. Access to Exercise Opportunities
5. Excessive Drinking

6. Adult Smoking

Your final map should include a map title with attribution and 6 sub panel maps with subtitles and meaningful labels.

Question 3:

For this question, I have given you two files:

- "Trees_NTA.csv" is the total number of trees within each of the NYC Neighborhood Tabulation Areas (NTA)
- "NYC_Tract_lat_long.csv" is the center geometry of every census tract in NYC

To complete this question:

1. Create a raster interpolation of the NTA trees file using an IDW interpolator (hint: be sure the extent of the interpolation is larger than the extent of the tract centroids).
2. Next, in R sample the tree interpolated raster with the point locations of the census tract centroid file.
3. Extract and attach the % of the population white at the census tract level from the US Census Bureau and join it to the Census Tract centroid file… if you need to review the census api extraction tool see this video.
4. Estimate a GWR model examining the relationship between the %white and Exposure to Trees
5. Join the GWR results to a census tract shapefile for the 5 counties in NYC and present the Local R2 and Relationship between %white and Tree Exposure in a 2 panel map using GGplot.