Exploring spatial and temporal structure of mobility in CA during COVID-19 pandemic

This exercise will help you get a taste of what it is like to work with spatial data, using various analytical tools from geography, GIS, and econometrics. Feel free to use any Internet or library resources as you see fit. You may program in any language of your choosing. I provide applicable packages in Python and in R where applicable, but you can choose any other programming languages, including Matlab, Julia, Java, C++, etc.

---

0. Formatting reports

Please submit two files: a *.pdf report presenting results and a working file with the code of analysis. You may submit one *.rmd file if you are working in R. You may also opt to submit all of the files via Github repo. Please make it private and add me (username: barguzin) to the repo.

1. Data

The raw daily data can be downloaded from Google Drive (this data set will be referred to as ***mobility data***). Depending on how much RAM you have, you might want to convert it to weekly data (instructions below), so that the computations are more manageable. Usually for large files you can iterate over chunks in pandas to filter out the data.

Please do not share the data elsewhere on the web! You may upload the data to work on Google Colab, but please do not redistribute the data set.

More information on the data set, including description of fields can be found here.

There are many resources online on spatial data analysis and visualization. I recommend you refer to the following books as a starting point, but please utilize whatever online resources you prefer:

1. Geographic Data Science with Python
2. Introduction to Spatial Data Programming with R

2. Pre-processing data

- Read in ***mobility data***, make sure it is formatted appropriately (i.e. dates should be in datetime format, and identifiers should be read as strings).
- Aggregate the data by week to make computations more manageable.
- Calculate *weighted average* for the variable called **bucketed_distance_travelled**. Please use only the information from this field, including the device count and the distance bin. Document your steps.
- Aggregate the data from census block groups to counties to make computations more manageable. You might need to read more information on the Census geographic place codes to understand how to use the codes to aggregate to a different Census level.

- Find, download and pre-process ESRI Shapefile (***geo data***) with California county borders (you will need to use it later for spatial analysis). Join the ***geo data*** to the ***mobility data***.

3. Exploring spatial structure

- Read about spatial autocorrelation measures, specifically on Global and Local Moran's I. You may refer to this page to understand the theory behind these statistical techniques.
- Calculate Global and Local Moran's I for the dataset for each week. Save both the I and p-values.

There are many tutorials online on how to do this. R geo-stack is more ripe with geospatial packages (including geostatistical packages). I would encourage you to use the "sf"package in R as it is more up to date compared to "st" for basic GIS tasks. Refer to this CRAN VIEW for a complete list of geospatial packages available in R. If you prefer working with Python, you can use the package called PySAL. Please refer to documentation in the corresponding packages to learn how to calculate Global and Local Moran's I.

4. Visualizing and mapping spatial autocorrelation

- For Global Moran's I create a line plot with scatter markers: plot time on x-axis and I on y-axis for California. Style markers depending on the p-values for Global Moran's I: use hollow markers (black edge color, white face color) for insignificant values, and use black (black edge color and black face color) for significant values. Add a vertical line for March 11, 2020.
- Comment on the Global Moran's I plot. Explain how you would interpret this graph in relation to mobility metric that you are analyzing.
- Plot Moran's scatter plot for the whole data set.
- Create a plot with 3 subplots for Local Moran's I (LISA) for Week 3, Week 12, and Week 40 (if you are using daily data, choose any single day in those three weeks). Use conventional coloring as in the documentation: hotspots - red, coldspots - blue, everything else - white.
- Compare LISA across three time snapshots. Describe what is happening to mobility at each individual week and how mobility changes over time.