Assignment03. Exploring Urban Mobility with Spatial Data

Due by 11:00 am November 23, 2021

__All__ operations must be performed in Python!

In this assignment, you are asked to explore a dataset that contains latitude and longitude data, thereby meeting a minimal definition of spatial data. You will use this within the Python ecosystem to assess the impact of new transportation systems we have seen arise in the city. You will use your existing knowledge of spatial data manipulation to assess how well these systems serve a greater societal good.

Assignment Steps:

1. Choose whether you will look at historic NYC Citibike trip data or Chicago rideshare trip data.

    - Citi Bike System Data. (Smaller, and maybe less interesting)
      Please choose a one or two months period (if you want, you can do a comparative analysis).
      Downloadable from https://www.citibikenyc.com/system-data. Click "Downloadable files of Citi Bike trip data" or "Download Citi Bike trip history data".
      Please use NYC Citibike data, not Jersey City one.
    - Chicago Transportation Network Providers (TNP) data. (Larger, but most complete!)
      Please choose a one or two months period (if you want, you can do a comparative analysis).
      Downloadable from
      https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p

2. Explore, munge, and visualize this information for your own reference, and assess the spatial characteristics, descriptive statistics and attributes of what the data you have chosen to analyze. Use this to help formulate your thinking for step 3. You can bring additional data such as other geographic entities, Census, NYC or Chicago Open Data, etc.

3. Pick your research question and a hypothesis around the use of your chosen system. For example, does Uber (rideshare) provide more mobility services for wealthier areas than lower income areas? Do New Yorkers use CitiBike differently based on their neighborhoods? Your research question and a corresponding hypothesis should include a spatial aspect.

4. Perform evaluations to test your hypothesis. You are invited to use public resources (Census, Open Street Map, NYC or Chicago Open Data, APIs, etc.) to bring in additional information (GIS layers, other geo-located tables, etc.) to test your hypothesis. Also, you are invited to use related quantitative methods (statistical tests, modeling, spatial analysis, etc.) to perform your analysis. All operations must be performed in Python.

5. Your deliverable will be another Medium blog entry that address the following:
    a. Framing your question and a short write-up on the data you are using
    b. Research question, hypothesis, and evidence.
    Describe your findings and the evidence that supports your first hypothesis. Use well-curated maps, scatter plots, tables, and numerical analysis to argue this point. This should be no more than 1,000 words. (Tip: you can create a table in a markdown setting in Jupyter notebook)

In some cases, you will be presented with an unfortunate situation where you will not find a result that supports your hypothesis. This is acceptable, and in many cases, inevitable. Should you refute your own hypothesis, present your findings and note that you find opposing evidence. You can state that you unfortunately did not find evidence supporting your hypothesis, and articulate why.

Deliverables

You should work on two deliverables:

- **"Medium" style post** (PDF format)
    a. No more than 1,000 words (excluding visuals and tables)
    b. Include your appendix about additional data (if any) as well.
- **Code** (Jupyter notebook and PDF)
    Please submit your Python code as Jupyter notebook (.ipynb format) and PDF with outputs. Please submit a single notebook, not multiple. Make sure I can run your code and create the same results without any error. Also, think about how to write your code in parts - a collection section, and an analysis section.

Metrics for Success

- Framing a research question and hypothesis
- Completeness of how you provided **spatial** data analysis
- Attractiveness and effectiveness of your visualization
- Completeness and clarity of your script
- Clarity of your article

An Extra Credit

If you actually post your assignment to your personal blog or create a Medium blog post in a non-anonymized way (in that you are willing to put your name to your work and bravely share it publicly), note the course, and pledge to keep it up, you will receive 1 point added to your grade for this assignment. Add the URL at the end of the assignment for evaluation. It should be noted that when you're applying for jobs, people will search for you. This assignment is testament to your analytical chops and therefore, I highly recommend you use this as a way of increasing your digital footprint.

Note

- You should use Python based coding.
- Remember, this course is not an Excel class.
- Don't forget to practice good file management.
- If you do something in Python but don't know how to code or you have an error, Google it! Again, this is not a joke.
- If you feel lost, use your colleagues and instructor as resources. But please keep in mind that your code cannot be the exact same as your colleagues'.