Lab: Inferring locations and mobility from mobile phone metadata

Introduction
In this lab you will:

- Explore the format of mobile phone metadata using synthetic data for the country of Togo
- Design and implement several algorithms for inferring home locations from mobile phone metadata
- Adapt home location algorithms to infer mobility traces
- Reflect on benefits, limitations, and risks of using mobile phone metadata to track mobility

The data analysis portion of the lab can be completed in any programming language or data analysis tool, including python, R, Stata, or a combination of google sheets and google maps. This PDF contains hints for completing the lab in Python and google sheets/google maps. If using python, the file lab.ipynb provides a Jupyter notebook template with prompts.

All data used in this lab are synthetic, but they are made to "look" like real mobile phone metadata, both in their format and in the underlying behavioral patterns.

This lab is loosely based on a number of papers using information from mobile phone data to infer locations and mobility, including:

- Chi, G., Lin, F., Chi, G., & Blumenstock, J. (2020). A general approach to detecting migration events in digital trace data. *PloS One*, 15(10).
- Warren, R., Aiken, E., & Blumenstock, J. (2022). Home Location Detection from Mobile Phone Data: Evidence from Togo. In ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (pp. 685-692).

Part 1: Read in the data and calculate summary statistics [10 minutes]

- Read in the data files from the 'data' folder: synthetic transaction data ('cdr.csv'), tower locations ('towers.csv'), and a shapefile of Togo's prefectures ('prefectures.geojson').
    - The file ‘cdr.csv’ contains all calls placed on a fake network of **149** mobile subscribers in Togo between January 1 and March 31, 2020 (90 days).
    - The file ‘towers.csv’ contains the GPS coordinate for each of the **103** towers through which calls are placed.
    - The file ‘prefectures.geojson’ is a shapefile of Togo’s **40** prefectures.
    - Python hint: Use pandas to read the two .csv files and geopandas to read the .geojson file.
    - Google sheets/google maps hint: Just upload the cdr.csv file into google sheets – you won’t need to work with the towers.csv or prefectures.geojson file in sheets.
- How many transactions are there?
- How many calls does the average subscriber place per day?
    - Google sheets/google maps hint: Create a pivot table! You will be using a lot of these in this lab.

Part 2: Map the towers [5 minutes]

- Create a map of the prefectures of Togo, and overlay the locations of each cell tower.
    - Python hint: Use geopandas for spatial manipulation. Use the file ‘prefectures.geojson’ rather than ‘prefectures.kml’.
    - Google sheets/google maps hint: Visit https://www.google.com/maps/d/u/0/. Click on “on “create a new map”, and then use “add layer” to upload towers.csv and prefectures.kml as data layers. For towers.csv, use ‘latitude’ and ‘longitude’ columns as placemarks, and the ‘tower_id’ column as the title for the markers.

- Are towers distributed evenly throughout the country, or do they tend to be concentrated in certain area(s)? What might explain the spatial trends you observe?

Part 3: Inferring home prefectures. [25 minutes]

- Design and code up an algorithm for inferring which prefecture each subscriber lives in.
    - Google sheets/google maps hint: This is another good time for a pivot table!
- Does your algorithm infer a home prefecture for every subscriber in the dataset? Why or why not?
- What are the strong points of this algorithm? What are the weak points?
- Identify a way to improve your algorithm, and code it up. Does your algorithm infer a home prefecture for more or fewer subscribers than your original algorithm? Why or why not? What are the strong points of this algorithm? What are the weak points?

Part 4: Inferring mobility [15 minutes]

- Using one of the algorithms above, what is your best guess for each subscriber's home *tower* in each week?
- Identify five subscribers living in different parts of Togo (based on different home prefectures identified in the previous section). Plot the mobility traces of these five subscribers on a map of Togo.
    - Google sheets/google maps hint: To our knowledge this would be prohibitively challenging in google sheets, involving both a join and mapping line segments. If you're working on google sheets, you can skip the map.

Part 5: Reflection [5 minutes]

- What are different ways phone-based mobility maps like this could be used to inform development or humanitarian decision-making? What would they do well? How might they fall short?
- What concerns might you have about using phone-based mobility maps like these in humanitarian and development policy? Pick one of the concerns that was raised above and brainstorm several strategies to mitigate the potential harm. You might think about technical mitigations as well as legal, policy, or social mitigations.