Colocation

Last updated August 22, 2022

Overview

Colocation Maps estimate how often people from different regions are in the same area at the same time, or “colocated.” In particular, for a pair of geographic regions x and y, these maps estimate the rate at which a randomly chosen person from x and a randomly chosen person from y are simultaneously located in the same general area during a randomly chosen time in a given week.

This dataset can help epidemiologists using metapopulation models to predict the spread of infectious disease. Metapopulation models are defined in terms of the average rates at which people from different geographic regions interact. Colocation Maps can parametrize metapopulation models by providing estimates of those rates. Metapopulation models can be contrasted with network and agent-based models, which use information about the behaviors and interactions of specific individuals. Since metapopulation models only use average rates instead of individual behaviors, we can share data that can help parametrize these models in a way that does not compromise individual privacy. Because the data adjusts to real-world changes, the dataset also reflects changes over time to how often people from different regions are near each other.

Questions the dataset helps answer

- At what rate are populations that live in different geographic regions interacting with each other?
- How does this mixing change over time based on changes to travel policies and working from home?

Features of Colocation Maps

- Updated weekly
- Available at administrative level 2 (equivalent to counties in the US) wherever there are enough users to ensure privacy, or at level 1 (equivalent to states in the US) if level 2 is unavailable
- Built using a standard methodology for the entire globe
- Available to download in csv format for analysis and input into epidemiological forecasts and models

Colocation and contact rates

Epidemiologists often estimate contact rates with data that do not directly log contacts, such as counts of people who move between locations (for example travel or commuting data). Colocation Maps are designed to offer a better estimate and to be updated more frequently over time, but they are not meant to be interpreted as direct measures of contact rates. This is an important distinction in the case of COVID-19, for which “contact” is defined as a distance of 6 feet (about 2 meters). The limitations inherent in any GPS-derived data prevent accurate counts at that distance.

Early analysis

In this paper, we explain in more detail how Colocation Maps were built during the first 2 years of the COVID-19 pandemic and report results from the first large-scale analysis of human colocation patterns across the world. The methodology with which Colocation Maps are built changed in May 2022, but this paper is still a valuable guide for how the maps can be used and analyzed. Among the findings of our study, we observe that a pair of regions can exhibit high colocation despite few people moving between them. We also find that although few pairs of people are colocated for many days over the course of a week, these pairs can contribute significantly to the total colocation time within a region or between pairs of regions.

Data standards

- **Population sample**: Facebook mobile app users who have turned on the Location Services device setting on their mobile device
- **Spatial aggregation**: US county-equivalent administrative boundaries. We use the territorial boundaries and names provided by the GADM project. Note that we do not report colocation rates between polygons in different countries.
- **Temporal aggregation**: There are 2 temporal aggregation scales used in this dataset. The first is the time window used to define a colocation event, which is 5 minutes. The second is the time window used to aggregate all colocation events and calculate the **weekly colocation rate (weekly_colocation_rate)**. This is one week. More specifically, the weekly

aggregation is calculated by averaging the daily colocation rate over the 7 days in that week.

- **Minimum counts**: Over the 7 consecutive days in the week defined by **date stamp (ds)**, both polygons must contain an average of at least 10 Facebook mobile app users who have turned on the Location Services device setting on their mobile device.
- **File format**: Data is provided in the format of a global comma-delimited text file or GeoJSON.

Codebook

- **Date (ds)**:The date in Pacific Time of the final day in the 7-day aggregation period. For example, 2022-08-08 means the dataset aggregates over the period 2022-08-02 to 2022-08-08 in Pacific Time.
- **Polygon 1 ID (polygon_1_id)**: Unique identifier from the Database of Global Administrative Areas (GADM) for polygon 1
- **Polygon 1 name (polygon_1_name)**: Name of polygon 1 based on the Database of Global Administrative Areas (GADM)
- **Polygon 2 ID (polygon_2_id)**: Unique identifier from the Database of Global Administrative Areas (GADM) for polygon 2
- **Polygon 2 name (polygon_2_name)**: Name of polygon 2 based on the Database of Global Administrative Areas (GADM)
- **Country (gadm_country**): The 2-letter abbreviation (ISO alpha-2 code) for this row. The country value is assigned according to the Database of Global

Administrative Areas (GADM) defining country boundaries (we do not report coobservation or colocation rates between polygons in different countries).

- **Polygon level (polygon_level**): Administrative level for this **country (gadm_country)** as defined in the Database of Global Administrative Areas (GADM).

Using the United States as an example:

GADM0=country (United States)

GADM1=state (Florida)

GADM2=county (Dade County)

Whenever possible, we construct the colocation matrix at level 2. For countries that don't have level 2 polygons, we fall back to level 1.

- **Weekly measured coobservation rate (weekly_measured_coobservation_rate):** The rate at which a randomly chosen person from polygon 1 and a randomly chosen person from polygon 2 were simultaneously observed in a randomly chosen 5-minute period, regardless of where they are in the world. This is different from the **weekly measured colocation rate**, in which the randomly chosen pair of people must have been observed not only in the same 5-minute period, but also in the same level-16 Bing tile.
- **Weekly measured colocation rate (weekly_measured_colocation_rate):** The rate at which a randomly chosen person from polygon 1 and a randomly

chosen person from polygon 2 were simultaneously observed in the same level-16 Bing tile in a randomly chosen 5-minute period

- **Weekly colocation rate (weekly_colocation_rate):** Of all the times that people from polygon 1 and polygon 2 could have been observed to be colocated, how often were they colocated? Mathematically, this is the **weekly measured colocation rate (weekly_measured_colocation_rate)** over the **weekly measured coobservation rate (weekly_measured_coobservation_rate)**.This adjusts the colocation rate in light of the fact that people from polygon 1 and polygon 2 were only simultaneously observed some fraction of the time, and only in those cases can we determine whether they were colocated. This adjusted value is what we propose that partners use in their analysis.
- **Is home tile colocation (is_home_tile_colocation)**: The value in this column represents the response to this statement: These pairs of people are in their shared home tile at the same time. If the value is **true**, the rate of colocation in the shared home tile is reported in the **weekly measured colocation rate (weekly_measured_colocation_rate)** and **weekly colocation rate (weekly_colocation_rate)** columns, which report raw and corrected rates, respectively. If the value is **false**, the **weekly measured colocation rate (weekly_measured_colocation_rate)** and **weekly colocation rate (weekly_colocation_rate)** columns report raw and corrected rates, respectively, at which pairs of people are colocated but at least one person in each pair is outside their home tile.