

Movement Between Places During Crisis (Bing Tiles)

Updated May 9, 2023

Getting started

Questions the maps help answer

- How much is movement disrupted by a crisis event?
- Where are populations going after they evacuate due to a disaster?
- Where should services be positioned in the aftermath of a disaster?

What you can do with these maps

- Identify both origin and destination tiles with abnormal amounts of population movement
- Identify areas heavily impacted by disaster
- Understand dynamics over time and estimate how many people are impacted

Who this dataset counts

These maps show how many Facebook app users (who have turned on the Location Services device setting on their mobile device) have moved from tile A to tile B between 8-hour windows while preserving the users' privacy. By aggregating this location data, we are able to show a representation of how many Location Services-enabled users have moved from one tile to another for each time period.

Note that this dataset is showing only a subset of the total population in an area, specifically only those Facebook app users who have enabled the Location Services device setting on their mobile device. This is not the total population moving.

These counts are also compared to precrisis movement levels between the 2 tiles. This can point to areas that are more impacted by a crisis, describing how evacuations are occurring or how regular movement patterns have been disrupted.

When these maps are generated

The Facebook app's Safety Check feature triggers creation of the maps, which cover only the areas affected by a crisis event. This dataset is generated for each area and time experiencing a crisis. Partners are also welcome to request dataset generation.

Movement Between Places During Crisis (Bing Tiles) vs. other mobility maps

Movement Between Places During Crisis (Bing Tiles) is sometimes called Facebook Movement Maps (Tile Version) or Movement Vectors because it shows the count of people moving from one tile to another. A very similar dataset is called Movement Between Places During Crisis (Administrative Regions). Several additional datasets describe people's mobility, including Movement Distribution and Colocation Maps. The naming of this dataset causes some confusion.

Common mistakes to avoid with this dataset

When new users see the name Movement Places During Crisis, they think this is the dataset they want to use because they are interested in human mobility. This is unfortunate because the visualization and interpretation of Facebook Population During Crisis (Bing Tiles) is a much more beginner-friendly dataset to view.

The inclination to dive first into movement maps can lead partners to a common mistake. People may try to add up all the inbound and outbound movement for a single tile in order to calculate a net flux for the tile of interest. This is not advised, as any movement vector with fewer than 10 users is excluded to protect privacy.

Understanding vector values

Vector values are significantly more complex to comprehend and visualize than simple population counts per tile. A vector has a direction defined by the starting and ending tile, and metrics comparing the crisis event to a precrisis baseline ($n_{baseline}$) can have both positive and negative values. So for example the vector from tile A to tile B can be a value of $+X$, but that does not mean that the vector from tile B to tile A is $-X$.

An easy way to understand the nature of this dataset is to think of commuting. These maps can tell you how many people went from tile A in San Francisco to tile B in Menlo Park this morning and whether that is more or fewer than a normal day before the crisis. They can also tell you

how many people went from tile B in Menlo Park to tile A in San Francisco during the same time period. Of course the magnitude of those counts will be very different.

Calculating movement between tiles

The Movement Between Places During Crisis (Bing Tiles) data represents the most common location (Bing tile) of the user within the first time window, and then the most common location within the second time window. This defines the starting and ending of the vector. The vector time is then assigned to the starting time (in Pacific Time) of the second window.

So for example an individual's start coordinates for time window A are the tile center for their modal Bing tile (the most common Bing tile among all their locations in the time period). The end coordinates are obtained the same way, but for time window B, which is 8 hours later. The user's start and end coordinates are then used to assign them to a vector. We then aggregate all users assigned to the same vector to yield the values in the dataset.

People who remain within a single tile

The Movement Between Places During Crisis (Bing Tiles) data also shows people who remained within a single tile, which is a common point of confusion. Data is provided for the number of people who "moved" from tile A to tile A between 8-hour windows.

Data standards

Population sample: Movement Between Places During Crisis (Bing Tiles) is derived from Facebook mobile app users who have enabled the Location Services setting on their mobile device.

Minimum spatial aggregation: We surface the Movement Between Places During Crisis (Bing Tiles) counts at Bing tile level 14, the minimum tile size allowed for privacy protections. This is equivalent to roughly 2.4 kilometers on a side near the equator or the size of 6 x 6 city blocks.

Temporal aggregation: Movement Between Places During Crisis (Bing Tiles) shows people moving from the place they spend the most time during one time period to the place they spend the most time during the next time period. The dataset is aggregated over adjacent 8-hour periods, beginning at 00:00, 08:00 and 16:00 Pacific Time, with 2 transitions between periods per day. Each row in the dataset represents one of these transitions. The first 8-hour time

window ends and the second 8-hour window begins at the value shown in date/time (date_time), with 2 possible times: 08:00 or 16:00 Pacific Time.

The time windows are fixed and do not change based on local time zones or the geography of the dataset. For example, a dataset generated in response to a typhoon in Manilla would have time windows that start at 15:00, 23:00 and 07:00 Philippine Standard Time. This is a known limitation of the current version of this dataset due to internal intraday data-retention policies that are applied based on Pacific Time (PT).

Update frequency and dataset start date: Dataset generation is triggered by either a Safety Check on the Facebook platform or a request to the Data for Good at Meta program. By default the dataset is updated within 2 8-hour time windows for 14 days, but additional time can be requested. The dataset is accessible for 90 days after the final update day and then is discarded.

Minimum counts: If both the baseline (n_baseline) and users during crisis (n_crisis) time count are less than a threshold value of 10 users, then the entire row is dropped. If one is greater than 10, we keep that count but null the other and all 3 derived metrics.

Crossing international borders: If the starting quadkey (start_quadkey) and ending quadkey(end_quadkey) represent different countries, then we drop that vector from the dataset as this would indicate crossing an international border.

Codebook

These take the form **human-friendly metric name (metric name in csv)**.

Starting latitude (start_latitude): Latitude coordinate of the center of the [Bing tile](#) grid cell for the origin or start of the movement vector. We use the [Presto geospatial library](#) to generate this field.

Starting longitude (start_longitude): Longitude coordinate of the center of the [Bing tile](#) grid cell for the origin or start of the movement vector. We use the [Presto geospatial library](#) to generate this field.

Ending latitude (end_latitude): Latitude coordinate for the center of the Bing tile that represents the destination or end of the movement vector. We use the [Presto geospatial library](#) to generate this field.

Ending longitude (end_longitude): Longitude coordinate of the center of the [Bing tile](#) that represents the destination or end of the movement vector. We use the [Presto geospatial library](#) to generate this field.

Length in kilometers (length_km): The distance in kilometers between the starting and ending latitude and longitude points that define the vector

Starting quadkey (start_quadkey): The unique identifier for the grid cell defining the start of the vector in the [Bing tile system](#). Each quadkey uniquely identifies a single tile at a particular level of detail. Quadkeys provide a one-dimensional index key that usually preserves the proximity of tiles in latitudinal and longitudinal space. In other words, 2 tiles that have nearby latitudinal and longitudinal coordinates usually have quadkeys that are relatively close together. Counting the number of characters in a quadkey string is a quick way to determine the tile's spatial granularity.

Note: Leading zeros are important and must be preserved in order to properly identify the tile. Although the values appear numeric, we recommend converting them into a string. Common spreadsheet tools like Excel and Google Sheets will not do this by default.

Ending quadkey (end_quadkey): The unique identifier for the grid cell that represents the destination or end of the movement vector in the [Bing tile system](#). Each quadkey uniquely identifies a single tile at a particular level of detail. Quadkeys provide a one-dimensional index key that usually preserves the proximity of tiles in latitudinal and longitudinal space. In other words, 2 tiles that have nearby latitudinal and longitudinal coordinates usually have quadkeys that are relatively close together. Counting the number of characters in a quadkey string is a quick way to determine the tile's spatial granularity.

Country (country): The 2-letter abbreviation (ISO alpha-2 code) for the starting and ending country (any vector where the start and end country are different is excluded). The country value is assigned according to a tile's intersection with country administrative boundaries derived from [GADM](#).

Date/Time (date_time): The time period represented by the data reporting. This value is the end time in Pacific Time of the first 8-hour segment and thus also the start time of the second 8-hour segment that defines the vector.

Baseline (n_baseline): Average number of users:

- Who have turned on the Location Services device setting on their mobile device) and

- Whose most common tile during the 8-hour segment ending at date/time (date_time) was starting quadkey (start_quadkey) and
- Whose most common tile during the 8-hour segment beginning at date/time (date_time) was ending quadkey (end_quadkey)

The baseline period is the 45 days going back from the day the maps were started (you can see this by looking at the day the first data is generated). The baseline value is segmented by the day of the week and the time window for the metric. Thus the baseline value for 00:00 PT on a Tuesday is the mean value (outliers are winsorized, not removed) across all the 00:00 PT plus Tuesdays in the 45 days prior to when the maps first generated. This means the baseline is a set period of days prior to day zero on the maps, not a rolling window. So for example each region in the dataset will have 7 days of 3 8-hour windows (00:00 PT, 08:00 PT and 16:00 PT) for a total of 21 unique baseline values across the dataset.

If the same person appeared at multiple tiles in a time interval, we only count their most frequent tile, choosing the latest of their most frequent tiles in the event of a tie. If this number is below 10, the value will appear as \N as this is below the level required for privacy protection.

Note: We can't backfill this dataset to previous dates because the baseline metric calculation already includes the entire time period that Meta retains Facebook Location Services data (in this case, 45 days). From the moment these maps are kicked off, they are already using all the past data that we have available to create the dataset.

Users during crisis (n_crisis): The number of people:

- Who have turned on the Location Services device setting on their mobile device)
- Whose most common tile during the 8-hour segment ending at date/time (date_time) was starting quadkey (start_quadkey) and
- Whose most common tile during the 8-hour segment beginning at date/time (date_time) was ending quadkey (end_quadkey)

If the same person appeared at multiple tiles in either time interval, we only count their most frequent tile, choosing the latest of their most frequent tiles in the event of a tie.

To protect privacy, we use a number of techniques to obfuscate the exact count. These are all typical of methods used to protect privacy in geospatial population datasets like the census.

- Random noise: We add a small amount of random noise to the users during crisis (n_crisis) count to ensure that it's impossible to ascertain precise, true counts for sparsely populated locations.
- Dropping small counts: We drop locations with small counts from the final datasets. If both the baseline (n_baseline) and users during crisis (n_crisis) time count are less than a threshold value of 10 users, then the entire row is dropped. If one is greater than 10, we keep that count but null the other and all 3 derived metrics.

Difference between baseline and crisis (n_difference): The number of people using the Facebook app who have turned on the Location Services device setting on their mobile device and are present during the crisis minus the average number of users present during the baseline period

Percent change (percent_change): Percent change in population counts of Facebook app users who have turned on the Location Services device setting on their mobile device and are along the vector defined by the starting quadkey (start_quadkey) and ending quadkey (end_quadkey), calculated by dividing the difference by the baseline (n_baseline) (plus a small value, usually 1).

Z-score (z_score): The number of standard deviations the users during crisis (n_crisis) value is above or below the baseline (n_baseline).

The z-score (sometimes called the standard score) highlights the areas with the most significant differences between what is observed during the crisis and what is typically seen during the baseline. Our recommended metric, it prevents the measured change from being dominated by dense areas (which tends to be the case when using only **difference between baseline and crisis/n_difference**) or sparse rural areas (which tends to be the case when using percent change/percent_change). The calculated z-score is clipped to be a value between -4 and 4.

Date stamp (ds): The date in Pacific Time for which data is being reported