

Facebook Population During Crisis (Bing Tiles)

Last updated January 3, 2023

Getting started

Who this dataset counts

These maps show how many people on Facebook are in a geographic region while preserving their privacy. By aggregating location data, we are able to show a smoothed representation of how many people are using Facebook's app in each map grid (often referred to as tiles) for each time period. Note that this dataset is showing only a subset of the total population in an area, specifically only those Facebook users who have turned on the Location Services device setting on their mobile device. This is not the total population density.

What the dataset is most useful for

A count of Facebook users is perhaps useful, but what is more interesting is how this number is changing over time. For example, a sudden decrease in the number of people can indicate that perhaps there is a connectivity issue in an area that prevents the normal expected number of people from accessing the app, or perhaps people have left the area entirely. It also highlights areas where people may be evacuating to or congregating. This dataset is a great starting point to define a geographic area of focus over a specific time window and then develop a hypothesis regarding what is happening on the ground.

Questions the maps help answer

- How are populations reacting to a disaster? Are they leaving the impacted area?
- Where do people go when they evacuate after a disaster?
- Where should services be positioned in the aftermath of a disaster?

What you can do with these maps

- Compare the spatial distribution of the crisis population to the precrisis population
- Identify areas with population increases or decreases
- Identify areas heavily impacted by disaster
- Understand dynamics over time and estimate how many people are impacted

When these maps are generated

Facebook Population During Crisis (Tile Level) are heat maps, which show where people are located before, during and after a crisis event. The Facebook app's Safety Check feature triggers creation of the maps, which cover only the areas affected by a crisis event. They may also be generated in response to a request to the Data for Good at Meta program. You will notice that the spatial extent of the data is defined by an outer boundary box demarking the affected area.

Facebook Population During Crisis (Tile Level) vs. Facebook's population density maps

Facebook Population During Crisis (Tile Level) maps are sometimes called "Facebook population density maps," because they show the density of people who use Facebook with the Location Services device setting turned on. However, this is easy to confuse with the overall [population density maps](#) that Meta produces using estimated census data instead of user location data. These maps are publicly available. The more accurate technical name for the population density maps is the High-Resolution Settlement Layer (HRSL).

Data standards

Population sample: Facebook Population During Crisis (Tile Level) counts the number of Facebook mobile app users who have the Location Services device setting turned on.

Minimum spatial aggregation: We surface the Facebook Population During Crisis (Tile Level) counts at Bing tile level 14. This is equivalent to roughly 2.4 kilometers on a side near the equator or the size of 6 x 6 city blocks. Note: We cannot dynamically alter the tile size or time window for this metric given the complexity of our calculations.

Temporal aggregation: Facebook Population During Crisis (Tile Level) is aggregated over 3 fixed 8-hour time windows, beginning at 00:00, 08:00 and 16:00, that represent changes in population density. As all time windows are in Pacific Time, they will represent vastly different periods from region to region.

At this time we are unable to adjust the windows based on limitations of the data source from which these maps draw.

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 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 a tile has a baseline (n_baseline) or users during crisis (n_crisis) count less than a threshold value of 10, then we replace the specific count and all derived statistics with a null state.

Codebook

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

Latitude (latitude): Latitude coordinate of the center of the [Bing tile grid cell](#) for the data point. The [Presto geospatial library](#) is used to generate this field.

Longitude (longitude): Longitude coordinate of the center of the [Bing tile grid cell](#) for the data point. The [Presto geospatial library](#) is used to generate this field.

Quadkey (quadkey): The unique identifier for the grid cell 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. This means that 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.

Country (country): The 2-letter abbreviation (ISO alpha-2 code) for the data point. 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 start of the time window for recording the metric, in Pacific Time (PT).

Baseline (n_baseline): Average number of people using the Facebook app who have turned on the Location Services device setting on their mobile device that are present in the tile during the baseline period prior to the event.

The baseline period is the 90 days before 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 the 8-hour time window beginning at 08:00 PT on a Tuesday is the mean value across all 8-hour time windows beginning

at 08:00 PT Tuesdays in the 90 days prior to when the maps were first generated. (Outliers in this baseline computation are winsorized, not removed. This means the baseline is a set period of days prior to day zero on the maps, not a rolling window.)

For example, each tile in the dataset will have 7 days of 3 8-hour time windows (00:00, 08:00 and 16:00) 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.

Users during crisis (n_crisis): The number of people using the Facebook app who have turned on the Location Services device setting on their mobile device that are present in the tile during the 8-hour time window beginning at date/time (date_time).

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 crisis count to ensure that it's impossible to ascertain precise, true counts for sparsely populated locations.
- Spatial smoothing: We average counts for a location with those of surrounding locations using [inverse distance-weighted averaging](#). This gives more weight to closer locations and less weight to farther 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) 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 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 that are present during the crisis minus the average number of users present during the baseline period

Baseline ratio (density_baseline): Ratio of people using the Facebook app who have turned on the Location Services device setting on their mobile device that are present in a specific tile during the baseline period over users in all tiles in the boundary box defining the crisis

Crisis ratio (density_crisis): Ratio of people using the Facebook app who have turned on the Location Services device setting on their mobile device that are present in a specific tile during the 8-hour window over users in all tiles in the boundary box defining the crisis

Percent change (percent_change): Percent change in Facebook Location Services-enabled user population counts per tile, calculated by dividing the difference by the baseline (plus a small value, usually 1)

Z-score (z_score): 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) or sparse rural areas (which tends to be the case when using percent change). The calculated z-score is clipped to be a value between -4 and 4.

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

How does the methodology for the updated version differ from the previous version?

- **Pacific Time:** While the previous version used [Coordinated Universal Time](#) (commonly known as UTC), in the updated version all time windows are in Pacific Time. That means they will represent vastly different periods from region to region. For example, the fixed time periods will correspond to time windows beginning at 12:00, 20:00 and 04:00 when adjusted for Singapore Standard Time.
- **Resolution:** Previous versions of this dataset defined the resolution (tile size) based on the area covered by the dataset. This was due to technical limitations and mitigations that ensured calculations finished in a timely fashion. For example, when producing a dataset covering the entirety of South America, our systems would produce the dataset with larger tiles (lower resolution), whereas a dataset covering only a small city would be produced with the smallest available tiles (higher resolution). In this updated version, tile size is fixed at Bing tile level 14. This is equivalent to roughly 2.4 kilometers on a side near the equator or the size of 6 x 6 city blocks.