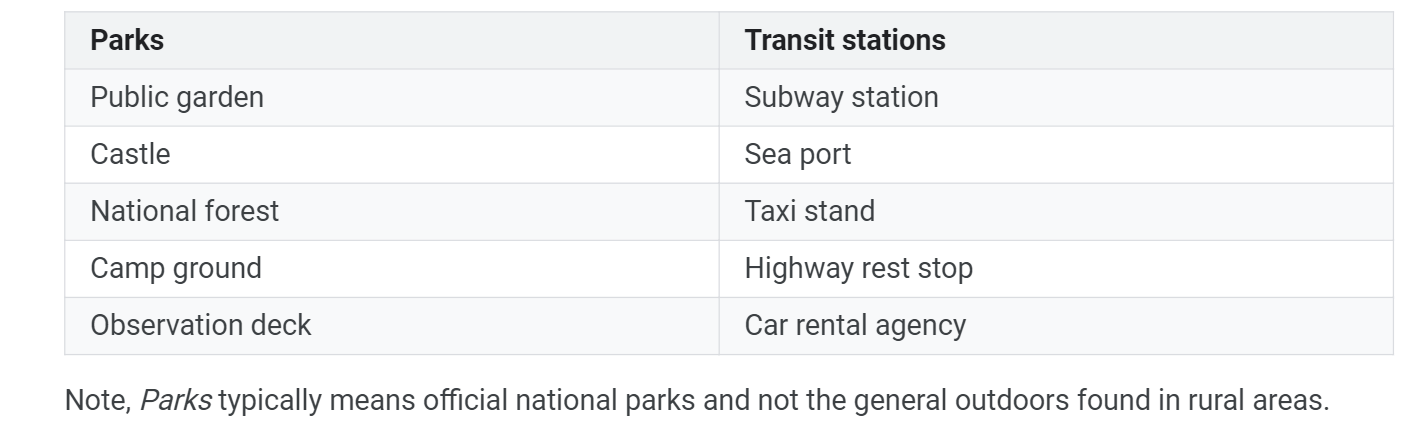
To summarize how Google produces the data:

1. **Set A Baseline**

The baseline is the median value of visits and length of stay at different high-level categories of places including retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential, for the corresponding day of the week, during the 5-week period Jan 3–Feb 6, 2020. To help with tracking week-to-week changes, the baseline days never change.

The *Residential* category is measured by average duration (hours) whereas the other categories are measured by total visitors. The table below shows some of the wide range of places included in 2 example categories:



1. **Collect Data**

Starting from Feb 15, 2020, sets of data are collected from users who have opted-in to Location History**1** for their Google Account, which is off by default.

Google gets information about users’ location from user’s device’s IP address, user’s activity (e.g. search for “Cafes in Paris”), and user’s device’s location if granted.**2**

To know where users live and work, it is inferred that**3** either users have added their workplace and home locations on Google Maps when it prompted users to do so or Google Maps signed in on to a phone of users that is capable of GPS, knowing that users visit their workplace and stay at home everyday at the same time, so that Google added them as users' workplace and home.

To get the number of visits to stores, parks and workplaces, it is inferred that**4** Google uses Points-of-Interest (POI) building footprints (polygons) that helps with knowing exactly the location, shape, & size of a store. Google then uses machine learning to join the GPS data against the building footprints to derive store visits (a process known as [store visit attribution](https://blog.safegraph.com/revealing-safegraphs-secret-method-for-getting-accurate-store-visits-from-gps-data-db9af9660585?utm_campaign=poi_questions&utm_medium=referral&utm_source=quora)**5**).

For transit stations, Google counts the number of visits to participating transit agencies and other places providing a publicly accessible service, like car rental agencies. Google Transit partners have already shared their static transit data (e.g. routes, stops, and

schedules) with Google Transit, which are integrated with Google Maps and accessible to Google users. Once static route and schedule information are shared, participating agencies can add live transit updates via Realtime Transit which enhance riders’ experience with the latest departure and arrival times, service alerts, and vehicle positions. Google Transit participating partners must make sure that service covers at least 4 weeks from and including the data that’s made available to Google Transit.**6**

1. **Add Random Noise To Protect Privacy7 and Aggregate8**

To ensure that no individual user’s journey can be identified, Google creates representative models of aggregate data by employing a technique called differential privacy**9**, together with k-anonymity, to aggregate population flows over time. Initially implemented in 2014, this approach to differential privacy intentionally adds

artificial random “noise” to the datasets in a way that maintains both users' privacy and the data's accuracy at an aggregate level. The project of

differential privacy contains a set of libraries of ε- and (ε, δ)-differentially private

algorithms, which can be used to produce aggregate statistics over numeric data sets containing private or sensitive information.

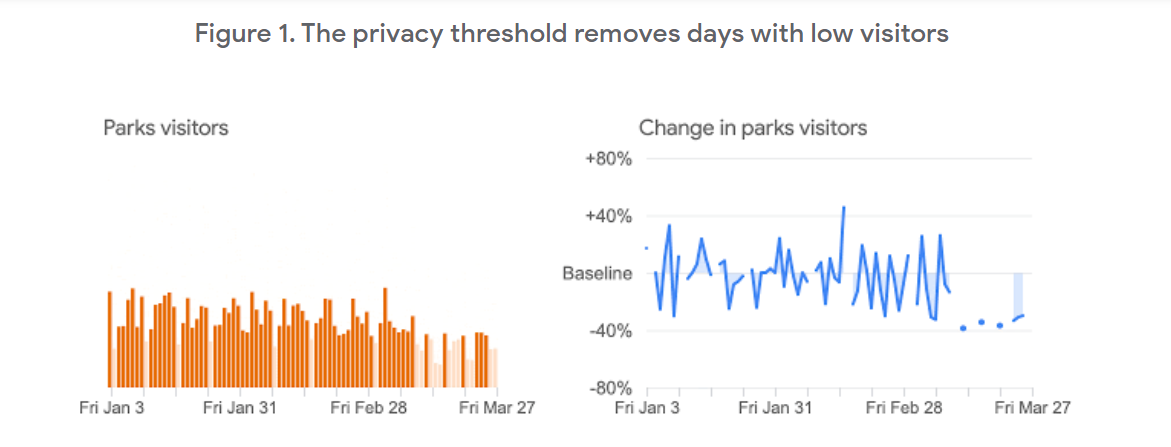
1. **Calculate**

The percentage change in number of visits and duration at the places compared to a baseline value for that day of the week is calculated each day.

1. **Coverage and Others about The Data**

Google calculates these insights based on data from users who have opted-in to Location History for their Google Account, so the data represents a sample of our users. As with all samples, this may or may not represent the exact behavior of a wider population.

What data is included in the calculation depends on users’ settings, connectivity, and whether it meets Google’s privacy threshold**10**. When the data doesn't meet quality and privacy thresholds, we might see empty fields for certain places and dates. In this case, we should treat gaps as *true unknowns* instead of assuming that gaps mean places weren’t busy. See the figure**11** below.



**OTHER REFERENCES**

<https://support.google.com/covid19-mobility/answer/9825414?hl=en&ref_topic=9822927>

<https://www.blog.google/technology/health/covid-19-community-mobility-reports?hl=en>