ActiveXchange

Movement Data Module

Frequently Asked Questions

Short Summary Video Click Here
Longer Explainer Video Click Here

mapbox

WHAT IS MOVEMENT DATA?

This service is enabled through a partnership between ActiveXchange and global location service and mapping provider, Mapbox.

Movement is a privacy-forward dataset of device density, activity, and movement over time drawn from 700M+ users globally contributing 20B+ live location updates daily, drawn from 45,000+ smartphone apps. These apps provide 'event pings' on an ongoing basis, which is the sharing of GPS locations (even if the app isn't open on the screen, depending on app/phone settings). GPS data also means devices can provide location data where there is no cellular coverage. Due to the number and breadth of these apps, the data becomes representative of the whole population and removes the influence of different demographics or different geographic areas (84%+ accuracy when validated against known counts and economic activity).

The dataset captures significant driving and non-driving mobile device activity aggregated into geographic tiles of 100-metre resolution (quadkeys) that cover the entire region. Activity within these quadkeys or across collections of quadkeys can then be tracked on a rolling basis, with data provided back to January 2022 to allow for trends.

The data is all anonymised to be data privacy compliant, and normalised to create a linear index, where 1.0 was the busiest quadkey per country in January 2020 (the anchor month), so an index of 0.1 would be 10% as busy (level of activity/visits) as this anchor location and month. Data is updated for the previous month by 7th of the preceding month.

This does not include dwell time data or data on individuals, but it does provide nationwide coverage across all types of infrastructure and environments, from which significant relative insights can be drawn. For an individual to generate an 'event' (app shares GPS) which is linked to a quadkey varies by mobile app and there is no minimum dwell time. Across all apps aggregated this provides a robust overall measurement.

84%+ correlation between UK movement (activity) index and known (reported) economic activity [1]

1 - As an example, in the UK, when correlated with known counts (e.g. footfall in a town park), 1.0 when aggregated up to a month equates to c200,000 visits.

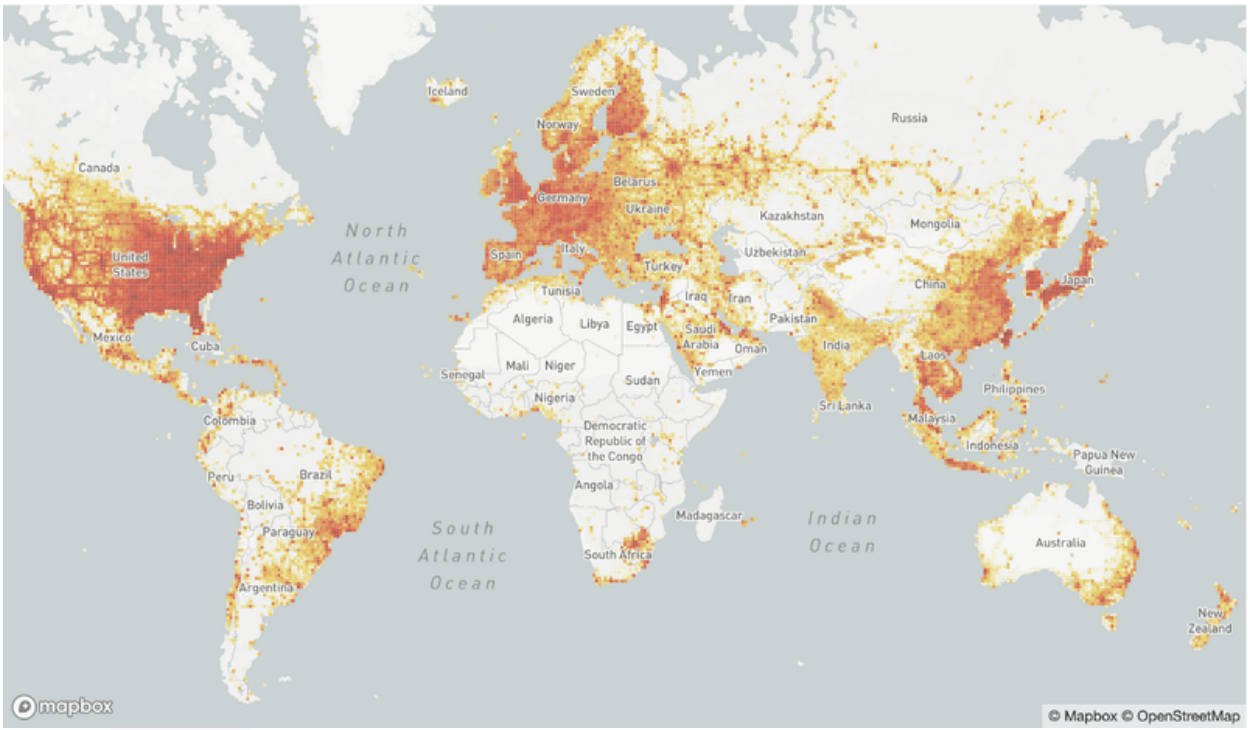
WHAT CAN I DO WITH THIS DATA?

Movement data provides a high-definition view of how, where, and when people move through a specific geography. In particular, this data allows you to:

- Understand the relative changes in mobile device activity in space and across time.
- Visualise changes in device density by hour of day, across multiple months or years.
- Observe significant anomaly events that occur and correlate changes in activity before and after – this provides an alert of the potential need for action e.g. movement suddenly drops at a facility/space/path.

FEATURES

- Hi-resolution location accuracy sourced from on-device GPS (other datasets are aggregated from cellular network carriers with limited location accuracy and high amounts of noise). This also means the data is less affected by areas with no cellular signal.
- An underlying panel of millions of monthly active users in every major market across multiple travel modalities (legacy data sets are limited to relatively small sample sizes or based on fixed sensors with limited insights and no coverage outside urban areas).
- Sourced from a diverse mix of thousands (45,000+) of different apps (other data sets are tied to a small number of sources which causes bias in the data).
- The most comprehensive privacy-forward global dataset (other data sets track advertising identifiers and personal information, which may fail to be compliant with regional and local privacy laws).



WHAT IS THE ACTIVITY INDEX?

The Activity Index is the density metric which reflects the level of activity in the time window and spatial regions defined above. It is a decimal value ranging from 0 to 1. A normalisation process is applied to smooth out any effects of extreme outlier activity and to reflect the general population (not just app and smart phone users). ActiveXchange is working towards also equating indexes to an estimation of people count, accommodating different settings.

1.0 on the index was the grid in the UK with the highest device density per hour through January 2021 (the baseline reference month - central London).

Everything else is worked back from this point in a linear format. i.e. 0.01 is half the number of devices (people) per hour than 0.02 etc.

WHAT FORMAT IS THE DATA PROVIDED IN?

The standard csv export, which also powers the Movement Data dashboard, has the following formats:

- NAME: either the name of the National Database site that is being monitored (this should match Sport England's Active Places database), or this is the custom name added by the user within the Mapping part of the platform.
- TYPE: as above, how the 'movement collection' was generated.
- YEAR-MONTH: 202201 = 2022 January; 202202= 2022 February etc.
- DAY-PERIOD-ID: 0 = weekday; 1 = weekend.
- TIME-PERIOD-ID: 24hr clock i.e. 7 = 7am; 13 = 1pm
- AVERAGE ACTIVITY INDEX: the average across the quadkeys selected within each collection

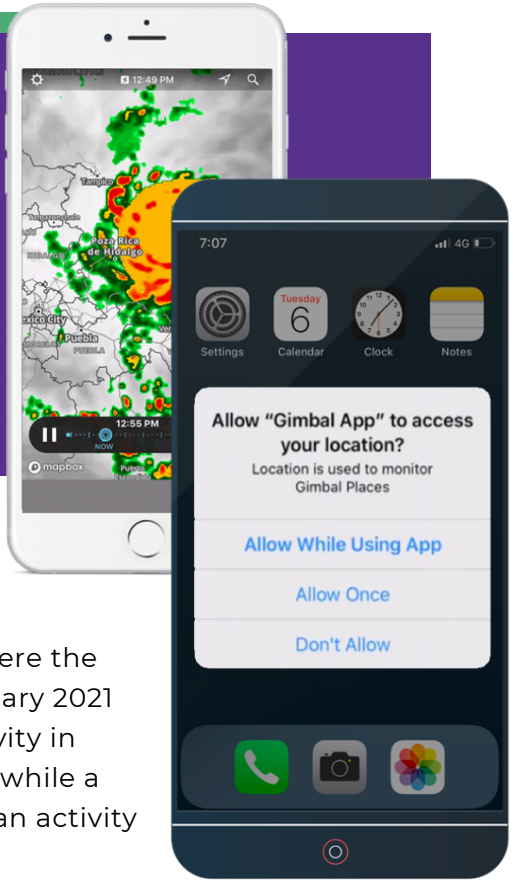
Name	Type	YearMonth	DayPeriodId	TimePeriodId	AverageActivityIndex
Edinburgh Playing Fields	NIDSite	202201	0	7	0.000764
Edinburgh Playing Fields	NIDSite	202201	0	8	0.000282
Edinburgh Playing Fields	NIDSite	202201	0	12	0.000508
Edinburgh Playing Fields	NIDSite	202201	0	13	0.000263

ActiveXchange also has access to specific day of week data (not featured in the platform). This is aggregated across the month and available within 4-hour time periods. i.e. Mondays in January from 00:00-0400, 0400-0800, 0800-1200 etc. This data is available on request, subject to additional data processing charges.

HOW IS THE DATA VALIDATED?

- Localised adjustments (urban vs rural) through use of an auto-scaling adaptive model.
- Normalcy index and the use of 'ground truth measurements' such as known sports attendance, time at home, traffic congestion, retail footfall, office occupancy, flights, film box office and public transport (gathered by the Economist).
- ActiveXchange undertakes correlation and validation checks with a range of registered sport and leisure known activity. Using this, ActiveXchange will also be supporting clients to equate the index to approximate visit numbers.

- 88% of UK adults had a smartphone in 2019, increasing to 94% by 2025 (Statista)
- 95% of smartphone owners use their mobile every day
- In 2020, there were 65 million mobile internet users in the UK
- 50% of 10-year-olds in the UK owned a smartphone in 2019



HOW CAN YOU COMPARE MOVEMENT ACROSS GEOGRAPHIC REGIONS AND TIME?

The activity per day and tile is normalised relative to a baseline, where the baseline represents activity in the 99.9th percentile day/tile of January 2021 in each country. In other words, a city block that had the most activity in January in the UK would have an activity index value of around 1.0, while a city block that had half that activity on a day in March would have an activity index value of 0.5.

The de-identified telemetry data that provides the foundation for Movement Data is closely correlated with the movement of people, but does not provide a direct measure of the absolute number of people moving. To emphasise this point, we have chosen to present this data as a unit-less activity index. Decisions made from Movement data should be informed by a comparison, whether that's a comparison of the activity difference between two facilities/spaces on any given day, or the change in activity for a given location over time, or a combination of the above. As ActiveXchange receives known registered data an approximate guide between likely users and activity index can be provided.

For any single place and time, the actual value of the activity index is not the only thing that matters. We also care about the difference in activity at that place and time relative to some other place or time. In fact, it makes more sense to talk about our confidence level for a comparison made from Movement data over space or time than of an individual data point. As we develop a mechanism to provide confidence levels for any given comparison, this general rule applies: comparisons made over short timescales and short distances are more trustworthy than those made over long timescales and distances.

HOW IS THE DATA NORMALISED TO BETTER TRACK TRENDS?

Sometimes it's useful to normalise the data regionally, to understand how local trends differ from other regions. In other cases it may be useful to smooth out the timeline data, to avoid having too much volatility in the data.

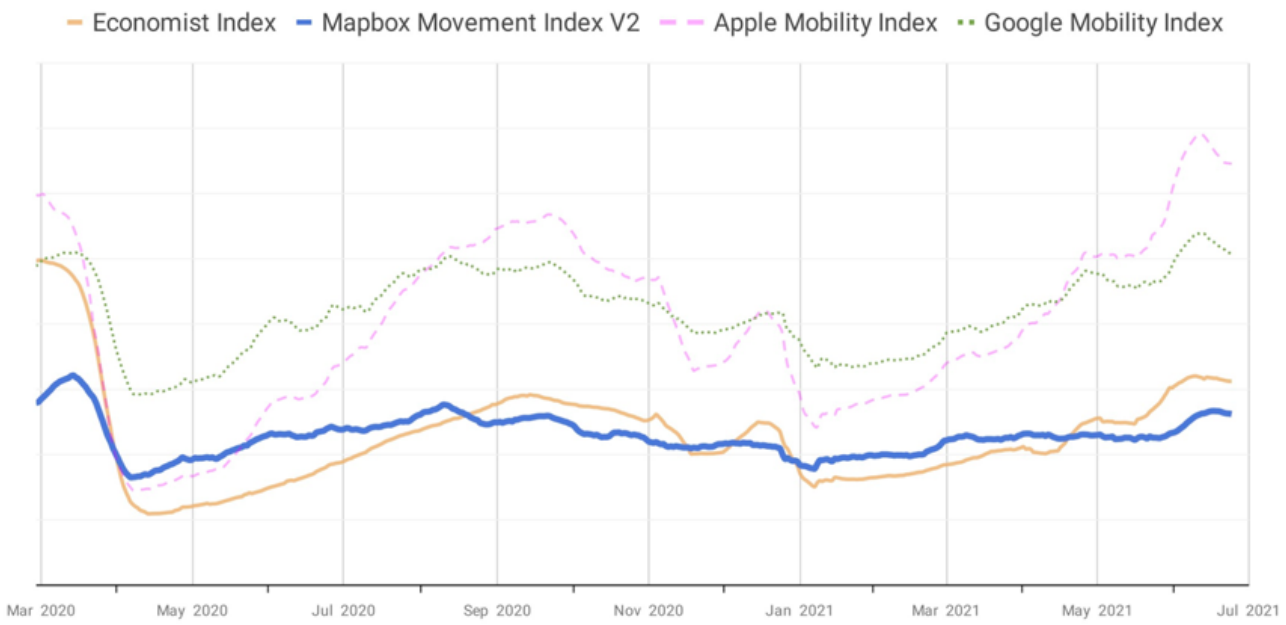
When you are interested in smooth timelines, we recommend working with moving averages with a window of seven days, or a multiple of seven. This will help smooth out some natural weekend effects. You can do this at any zoom level or boundary aggregation. For example, you can pick four weeks in February and define a generic Monday as the average of those to compare against for the future Mondays. This normalisation would help when trying to identify changes over time of individual days of week. When you want to compare individual areas (like local authorities) that have

a different average activity level, it might be useful to normalise each area independently if you are interested in their relative change more than in their absolute difference.

HOW ARE ESTIMATES CALIBRATED?

Mobile device adoption and usage patterns vary meaningfully across urban, suburban, and rural residents. In addition, many mobile apps are designed for use exclusively in urban areas (e.g. ridesharing, food delivery, community/neighbourhood activism apps). This generates a significant imbalance in mobile device activity levels measured in urban vs. rural areas, artificially inflating urban activity levels and depressing rural activity patterns. To address this inconsistency, an auto-scaling adaptive model is applied that identifies and measures the urban/rural imbalance and applies a set of "corrections" at the city level on a daily basis. This self-adapting system is fully automated.

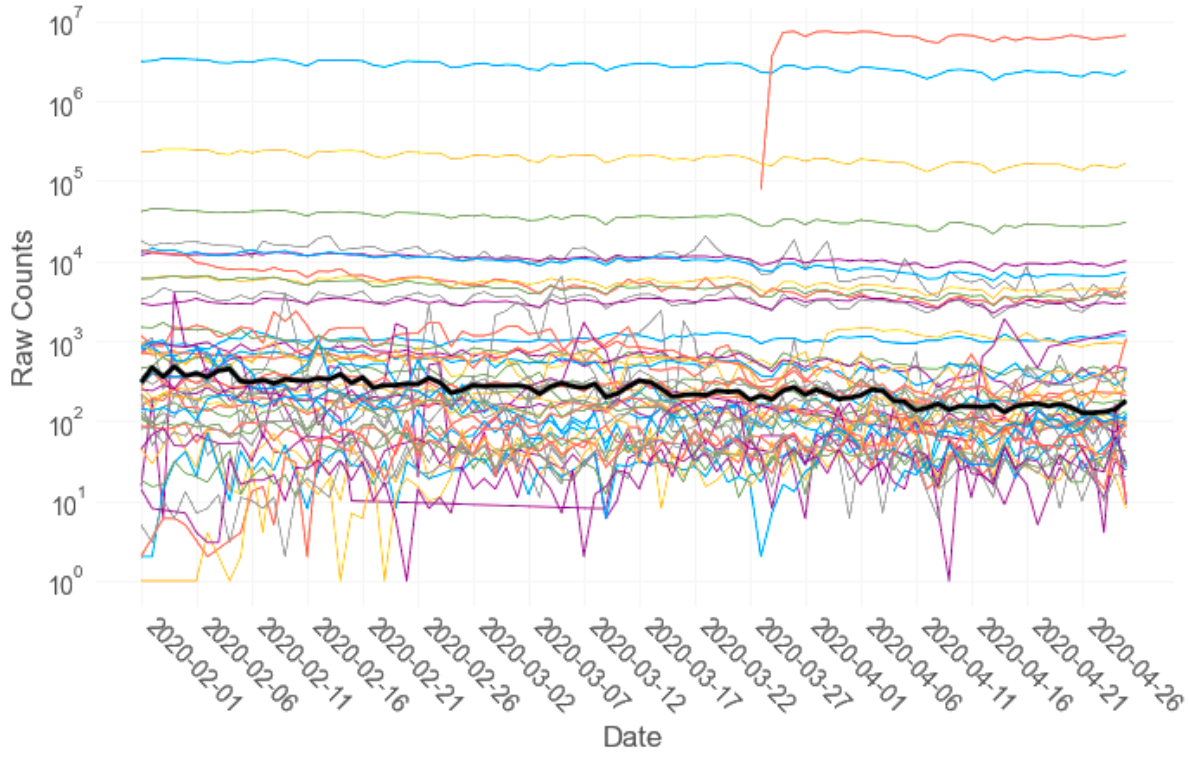
Good calibrations are always difficult and that's especially true here: actual ground-truth measurements of that density-of-humans number simply do not exist at the very fine spatial and temporal scales required. Other similar data sets exist from other providers and we compare Mapbox (ActiveXchange) data to the Economist's Normalcy Index (known attendances, spend etc.) – top right graph below. Importantly, ActiveXchange undertakes further correlation validation with known local community sport and leisure users and attendance patterns in different settings and communities to ensure confidence of use.



This Movement data has a significant advantage over our competitors in the "calibrating device density" game: Mapbox-powered maps are used by a great number of different apps, each targeting a different subset of the population, running on both major mobile device platforms (iOS and Android). Using the law of averages it is possible to calibrate the data against itself.

HOW DO WE AVOID SKEWING FROM NEW HIGH USE APPS?

If we assume that each app samples a different, uncorrelated subsample of the overall population, then we can calculate device-density timelines of activity for each app, and then take the median curve across all of those timelines. That median curve will be a much more reliable measure of trends in human density than any raw aggregation of device density could be. Mapbox powers everything from weather apps to in-car navigation to fitness apps and location-based games, each of which really do sample very different subsets of real humans – bottom right graph (noting the curve isn't skewed by the new app).



This diversity-of-apps concept is a double-edged sword: the apps that use Mapbox occasionally churn. Sometimes large volumes of data begin flowing from a new app that recently started using Mapbox, or an app may suddenly go dark, or maybe, some app just goes viral for a few days in one city. If any of these individual apps are popular enough, these sorts of changes will significantly affect the population-sampling rates (and thereby total data counts). Median-to-the-rescue again: the median-across-providers mechanism works great to handle app churn, in addition to handling biased-subpopulation effects.

Thinking about time and space: The median-across-apps curve is effective, but it only works when enough data from enough different apps exists to calculate the median curves. The median-across-apps idea is data-hungry; if a data set strictly requires enough telemetry to take a median across apps, the resulting data product would only be available over relatively large areas (e.g. a city) and not available at much finer granularities (e.g. single city block) required for most customer use cases. Even if actual activity varies hour-by-hour and block-by-block, the demographic and app-churn effects calibrated out of the signal really only vary over longer timescales (e.g. day-to-day or month-to-month) and larger spatial scales (e.g. city-to-city, or city-vs-suburban-vs-rural). We can take advantage of this fact to calculate a unit-less correction factor F, by calculating the median-by-app curve on those larger scales and measuring the fractional difference between that curve and the actual device density curve. Then we can smoothly interpolate that F down to the smaller spatial scales required so as to apply it to uncalibrated device densities resulting in calibrated activity measures at finer spatial granularities.

AGGREGATION ACROSS MULTIPLE DAYS

Quadkeys with zero activity in a given day are excluded from that day's data. As a result, when trying to compare activity aggregated over multiple days, we do not recommend trying to take an average of the activity (since this value would be biased by the omitted quadkeys). Instead, we recommend either using the sum of activity index across multiple days as the point of comparison, or weighting the average appropriately to account for the omitted quadkeys.

Due to Mapbox's strict privacy requirements, quadkeys that do not pass minimum activity thresholds over the period of aggregation are omitted. As a result, the daily data will have more zero-activity areas, even when aggregated over a month, than monthly data.

COMPARISON ACROSS GEOGRAPHIC REGIONS AND TIME

The data for each country is normalised within that country, while the normalisation factors are different between countries. As a result, the levels of activity indices are not directly comparable across different countries. The data can be used to compare trends in activity levels over time within each country, though, with comparisons being inherently slightly more exact within closer geographic proximity. This is described in more detail below.

The de-identified telemetry data that provide the foundation for the Movement product are well correlated with the movement of people about the world, but do not provide a direct measure of the absolute number of people moving. To emphasise this point, we have chosen to present these data as the unit-less activity index. Any decision made from Movement data should be informed by a comparison, whether that's a comparison of the activity difference between two areas of a city on any given day, of the change in activity for a given location over some time span, or some combination of the above.

For any single place and time, the actual value of the activity index is not the only thing that matters. We also care about the difference in activity at that place and time relative to some other place or time. In fact, it makes more sense to talk about our confidence level for a comparison made from Movement data over space or time than of an individual data point.

As we develop a mechanism to provide confidence levels for any given comparison, this general rule applies: Comparisons made over short timescales and short distances are more trustworthy than those made over long timescales and distances.

We are always working to better calibrate our underlying data, but we expect some drift to accumulate. Day-to-day comparisons are almost always quite exact, while year-over-year comparisons can be trusted to provide general trends but not detailed examination. Similarly, any comparison made within a single city will almost always be quite exact, while comparisons made across cities should be used to provide directional trends and small differences may not carry a lot of signal.

Also note that the activity index measurement is more stable for areas in which there is a relatively high average activity, since it is less likely that an individual event could influence the scores in the area. As we work through a quantified confidence level for absolute values, this general rule also applies: Comparisons made over areas that have a high average activity are more trustworthy than those made over areas that have a low average activity.

Any specific questions or support please don't hesitate to reach out to the ActiveXchange team: intelligence@ActiveXchange.co.uk