Location exposure index based on PlaceIQ data[*]

Victor Couture   Jonathan I. Dingel   Allison Green   Jessie Handbury

Berkeley Haas   Chicago Booth   Princeton   Wharton

Kevin Williams

Yale SOM

April 2, 2020

version 0.5

We use PlaceIQ data that describe smartphone devices "pinging" in a given geographic unit on a given day. We use these data to compute a location exposure index ("LEX"). The LEX is a matrix that answers the following query: Among smartphones that pinged in a given location today, what share of those devices pinged in another location at least once during the previous 14 days?

We are making these indices publicly available to all researchers in the context of the spread of COVID-19. The indices, which describe human exposure varying across locations and time within the United States, could be useful in a variety of applications. The notes below precisely define how we compute these indices so that researchers can use them appropriately.

Notation

- Index geographic units by g and g'. The number of geographic units is G.

- Index dates by d and t.

- Index devices by i.

- Let $p_{igt} \in \{0,1\}$ be an indicator that is equal to one if device i pinged in geography g on date t.

- Let $\mathfrak{p}_{it} \in \{0,1\}$ be an indicator that is equal to one if device i pinged anywhere on date t, $\mathfrak{p}_{it} \equiv \max_g p_{igt}$.

- Define the set of devices that pinged in g' on date d by $\mathcal{I}_{g',d} \equiv \{i : p_{ig'd} = 1\}$.

Devices covered

We restrict the set of devices included in our calculations. Not all devices in the raw data ping regularly. Ping frequency reflects a device's applications, settings, and movements. To focus on devices whose (non-)movements can be reliably characterized (particularly after the onset of COVID-19), we restrict the set of devices included in our calculation of $\mathbf{LEX}_d$ (defined below) to those that pinged on at least 11 days

over any 14-day period from November 1, 2019 through date d. In order to be included in our calculation on date d, a device must satisfy the following inclusion criterion:

$$\exists d' : \left(\text{November 1, 2019} \leq d' \leq d - 13 \ \& \ \sum_{t=d'}^{d'+13} \mathfrak{p}_{it} \geq 11 \right)$$

Since the earliest date for which we report $\textbf{LEX}_d$ is January 20, 2020, this criterion selects a set of devices on the basis of more than two months of prior activity. This window should be sufficient to select devices that ping regularly. For later dates, this criterion selects devices based on 14-day periods within a longer window of potential activity. Given the reduced movement associated with the ongoing pandemic, we worry that a criterion with a fixed window of potential activity would exclude devices that temporarily reduced their movements. Our criterion's growing window continues to include these devices based on their movements observed earlier.

Definition of location exposure index

LEX definition: What fraction of active phones in geographic unit g' on day d have been active in geographic unit g at any point in the last 14 days?

$$\text{LEX}_{gg'd} \equiv \frac{\sum_{i \in \mathcal{I}_{g',d}} \mathbf{1}\left\{ \sum_{t=d-14}^{d-1} p_{igt} > 0 \right\}}{\sum_{i \in \mathcal{I}_{g',d}} \mathbf{1}\left\{ \sum_{t=d-14}^{d-1} \mathfrak{p}_{it} > 0 \right\}} = \frac{\sum_i \mathbf{1}\left\{ i : \left(p_{ig'd} = 1 \ \& \ \sum_{t=d-14}^{d-1} p_{igt} > 0 \right) \right\}}{\sum_i \mathbf{1}\left\{ i : \left(p_{ig'd} = 1 \ \& \ \sum_{t=d-14}^{d-1} \mathfrak{p}_{it} > 0 \right) \right\}}$$

Thus $\textbf{LEX}_d$ is $G \times G$ matrix in which each cell reports, among devices that pinged on day d in the column location g' and pinged anywhere during the previous 14 days, the share of devices that pinged in the row location g at least once during the previous 14 days.

Locations reported

Many US counties have few residents and therefore few devices in the PlaceIQ data. The LEX matrices we report are restricted to counties with reasonably large device samples.

To implement this restriction, we assign each device to a unique "residential county" on each date that it pings. In particular, on each day, the assigned residential county of a device is the county in which that device had the highest (cumulative) duration of time at residential locations on that date.[1] These "residential" assignments are not used in any LEX calculation except the determination of which counties to report. The counties that appear in our reported LEX matrices had at least 1,000 devices assigned to them as their "residential county" on every day of the week of January 6-12, 2020. Thus, we report LEX values for 2,018 of the more than 3,000 counties in the United States.

Device coverage over time

In the context of the COVID-19 pandemic, a potential concern about the $\text{LEX}_{gg'd}$ defined above is that devices may not generate pings when "sheltering in place", due to lack of movement. Both the numerator

---

[1]The PlaceIQ data identifies visits to residential location separately. In the event of a tie in duration, the device is assigned to the county in which it made more distinct visits to residential locations. In the event of a tie in distinct visits, the device is assigned to the county in which the device had the highest duration of time at non-residential locations. In the event of a tie in non-residential duration, the device is assigned to the county in which it made more distinct visits to non-residential locations.

and denominator restrict to attention to devices that ping in g' on day d ($i \in \mathcal{I}_{g',d}$). Since early March 2020, there has been a general decline in the number of devices generating pings, presumably due to individuals restricting their movements in response to the pandemic.

For applications in which one wants a notion of non-local exposure associated with "active" devices that are moving around within location g', this may not be a problem. $\text{LEX}_{gg'd}$ captures the locational histories of devices that are "out and about" in geographic unit g' on date d and does not capture the locational histories of devices that do not generate any pings because they are isolated and not moving. For applications in which one wants to describe exposure in terms of the population of devices, regardless of whether they are generating pings or not, this may be a problem. We have posted the number of devices that ping in each county on each date (the denominator of $\text{LEX}_{gg'd}$) to help researchers use these measures appropriately.

Device coverage over space

The full PlaceIQ device sample is meant to be nationally representative. However, our sample selection criteria mean that the devices used to compute exposure indices are unlikely to be fully representative of the residential population or daytime population on a county-by-county basis.

A reminder

Phone movement and human movements are correlated but not synonymous. These data offer the opportunity to use the former as a proxy for the latter, with the necessary caveats.