**Methodological Note**

This file contains supplementary information for the following paper:

**Banks, alternative institutions, and the spatial-temporal ecology of racial inequality**

Mario L. Small, Armin Akhavan, Mo Torres, and Qi Wang

Corresponding Author: Mario L. Small (mariosmall@fas.harvard.edu)

**1. Using Google Maps and Google Places API to find AFIs**

In this section, we describe in greater detail the process through which we acquired data on conventional banks and alternative financial institutions (AFIs).

*Data source*

We used Google Maps and Google Places API. Google Maps collects large amounts of data on many kinds of commercial and non-commercial establishments, including conventional banks and alternative financial institutions, such as check cashing places, payday lenders, etc.

Our rationale for using Google data instead of alternatives was the comprehensiveness and high quality of the data. The establishment data is based on at least three sets of sources. First, Google uses publicly available administrative data sources from official and private sources. Second, as the volume and quality of its street imaging capacity have ballooned, Google has also developed advanced algorithms to detect places based on Google Street View imagery, which allow the detection and updating of establishments with high accuracy.[[1]](#footnote-1),[[2]](#footnote-2),[[3]](#footnote-3) Google essentially reads signs from storefronts and shops about what an establishment is and feeds it to its database, through a process that, given the company’s vast amount of data, refined algorithms, and computational power, results in remarkable accuracy.[[4]](#footnote-4) Third, Google employs crowdsourcing. Business owners can easily submit information on their establishment if their data are missing. Google also encourages “local guides,” people who know their neighborhoods well, to independently post their own data on places.[[5]](#footnote-5) At this juncture, anyone can propose the existence of an establishment: “[From] Local Guides, active Google users, and business owners via Google My Business, [Google] receives more than 20 million contributions from users every day.”[[6]](#footnote-6)

The sources, however, are not taken at face value. Each submission from individuals to add or correct information about an establishment is reviewed for accuracy before posting. In addition, automated data are also checked. Google has employed what has been described as “a small army of people to manually correct the information collected from these different sources.”[[7]](#footnote-7) Indeed, at this juncture, the process of improving establishment and place data at the local level has been “gamified,” through a social media platform that rewards input and databased improvement through public recognitions, perks, and prizes for contributions. The Local Guides platform is, in fact, both for the submission and the continuous correction of site information.

The combination of administrative data, proprietary imagery, and crowdsourcing, all continuously updated and reviewed for accuracy, has meant that Google has produced an extraordinary database. We performed additional validations. Before conducting our analyses, we compared data on conventional banks and ATMs from Google Places to those from Microsoft Bing Maps in each of the 20 cities we studied.[[8]](#footnote-8) In all but two cities, Google identified more existing locations.[[9]](#footnote-9) We also examined the Google data against some known neighborhoods in our local city, Boston, and found a high degree of accuracy, with known establishments present in the data, and none of the establishments incorrectly labeled.

The Google data are ultimately the optimal source for our purposes. They are more accurate and appropriate than other sources that might appear better, such as an official list of banks or AFIs from a government source. While the latter have obvious advantages, they are already part of Google’s database. More importantly, they have two disadvantages for our purposes. First, official data are typically released annually, rather than updated continuously. Since establishments are continuously either opening or going out of business, any yearly or semi-annual snapshot will inevitably be dated. Second, official records are often inaccurate. They are inaccurate not merely due to clerical errors but, more importantly, because they may not reflect reality as it is experienced—establishments may exist only on paper, constituting an address and a mailbox; be long dead but still on the books; or be misclassified by malfeasant actors. The upshot of these possibilities is that, as in many areas of life, the government’s record of an establishment may not match the reality that someone walking down the block on their neighborhood experiences. Our data come closest to approximating that experience, since it takes that official record, confirms it with a Street View car that has driven down the block and taken a picture of the building and signage at address, and re-confirms it with a crowdsourced platform wherein an individual, within a social media context of continuous evaluation for accuracy, reports that an establishment is in fact at that location. While no data source is perfect, ours comes closest to that ideal for the objectives of our analysis.

*Extraction*

Each of the establishments in Google Maps’ database is stored as an “object” with a set of attributes. Google’s Places API “is a service that returns information about places using HTTP requests. Places are defined within this API as establishments, geographic locations, or prominent points of interest.”[[10]](#footnote-10) Through Google Places API, one can access the Google Maps data directly. The data returned from the Google APIs are stored as json “objects,” which we collected and then parsed into our own database for original analysis.

The current paper was part of a larger study for which we collected data on a large number of establishments. We collected establishment data using the Google Places API from December 2016 to January 2017 in the top 20 cities in the U.S. We split the area of each city into grids and used the “nearby” function of the Places API to collect any places in every grid. If a search returned more than 60 (i.e., the maximum numbers of places that can be returned) places, we split the grid into four equal smaller grids and searched again in each of the sub-grids. The algorithm stopped whenever a search returned less than 60 places and moved on to the next grid. The codes and algorithms can be found here: [https://github.com/urbaninformaticsandresiliencelab/gmaps\_scraper](about:blank).[[11]](#footnote-11) All place data collected from a city were stored in a combined .json file.

Of particular importance to our analysis was the place attribute “type,” which describes the function of the establishment (see [https://developers.google.com/places/supported\_types](about:blank)). On Google Places API, conventional banks are assigned the “type” *bank*. By accessing the Google Maps database through the Places API, we were able to simply and efficiently retrieve data that identified conventional banks in the twenty major cities in our analysis. However, not all Google Maps data are available through the Places API; some must be retrieved using a web browser. For example, searching for an establishment on Google Maps using a web browser or mobile phone returns data such as hours of operation that are not retrievable from the Places API. Similarly, while banks are easily accessible through the Places API, alternative financial institutions are not, because there is currently no Places API “type” for AFIs.

Still, AFIs are indeed readily identifiable through a simple Google Maps search using a web browser. Doing so requires deciding on search terms. For example, while there is no Places API “type” for “check cashing establishment,” a browser search on Google Maps for the term “check cashing” retrieves the data from the same database used by the Places API, and lists the full results, including the “name,” “rating,” and “vicinity” attributes from each place “object.” In fact, the results also include a line indicating that the establishment is of the “Check cashing service” category (see below). This establishment category as indicated on the web browser is not one of the Places API listed “types.”

A screenshot of a cell phone

Description automatically generated

On a browser, it is possible to find AFIs under different searches, and a particular AFI may be listed under different categories in different searches. For example, since many payday lenders cash checks, the same establishment may be listed as a “Loan agency” in one search and as a “Check cashing service” in another. However, some categories are broader than others. While payday lenders are rather consistently categorized under check cashing services, the reverse is not necessarily the case, as many places where one could cash a check for a fee do not offer loans.

We used the term “check cashing” to identify check cashing service categories, which is appropriate given the ubiquity of cashing checks as a basic service across a large swath of different AFIs, including not only check cashing establishments but also payday lenders and currency exchanges. Indeed, in some cities where payday lenders are not legally allowed, alternative financial institutions are simply known as “check cashing stores.” While our approach is appropriately comprehensive, it may nonetheless exclude some alternative institutions, such as pawn shops that do not cash checks or are otherwise not classified as doing so. Thus, our process may undercount alternative financial services, particularly more obscure ones. If so, then unless institutions such as pawnshops are especially likely in predominantly white neighborhoods, our results would likely understate the extent of racial inequality.

To create an intermediate dataset of AFIs, we first used Google Maps to identify all available establishments marked “Check cashing service.” We then matched these establishments to the “name” for each location in the Places API, and extracted the place “objects” and all related attributes. This process provided AFI data that are as rich as the data we had collected on conventional banks. To scrape the AFI data, we programmed a Python script to open a Firefox browser, navigate to the Google Maps URL, click on the search box, type in “check cashing,” and click on the search button. From there, we processed through the results, storing the names of all places categorized as “Check cashing service.” The script then matched the names to the place “object” from our dataset built from the Places API.

We experimented with four ways to collect these data with Google Maps on a web browser. Using http://www.google.com/maps/ as the base URL, we tried each of the following:

1. Using coordinates, in latitude/longitude format, and passing the coordinates to the search end point with a search query. For example, to access Chicago: http://www.google.com/maps/search/<search query, e.g. check cashing>/@40.7274826,-74.0902511,11z
2. Using the place endpoint and navigating to city, e.g., Chicago, using the URI of `https://www.google.com/maps/place/Chicago,+IL/` and then clicking on the element for NEARBY search `section-action-button-icon maps-sprite-pane-action-ic-searchnearby` and search for `check cashing`
3. Searching for the place with the search endpoint, meaning first using, e.g., Chicago to search for Chicago, which returns the place endpoint, and continuing with the steps above.
4. Combining the two searches and searching, e.g., for `Chicago,+IL+check+cashing`

The script we settled on uses the fourth method to search. It goes through the list of results, paginating and matching names between the browser-based Google Maps and the Places API. This process, for Chicago alone, returned 254 places we then added to our database. We then made additional rounds of browser scraping with different zooming scales, and looked at the rate of finding additional places as a determining factor to stop the operations. We were able to scrape the rest of the places that the first round did not return by making a loop/iterator for a zoom/drag operation with randomized values for zoom levels and drag directions, ensuring the completeness of our data.

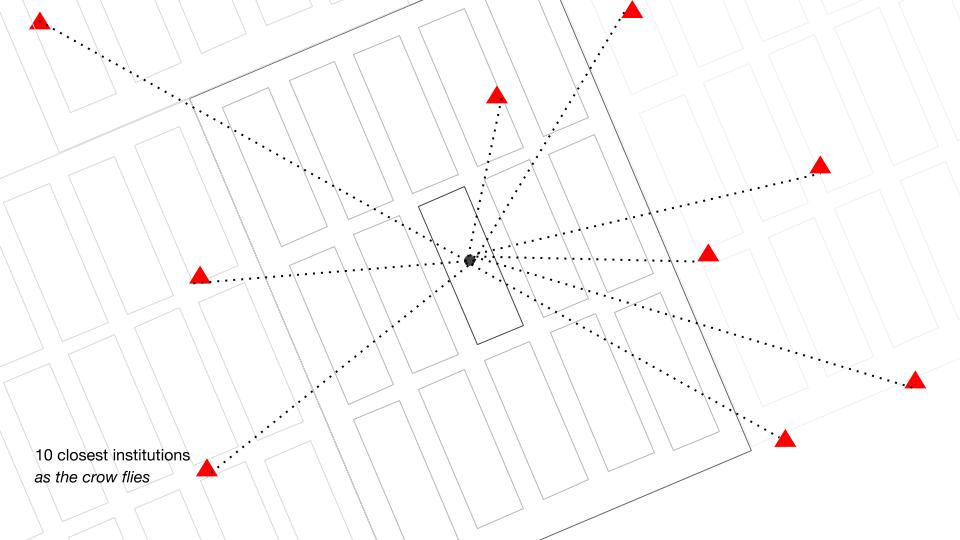
**2. Calculating access**

An important contribution of this paper is our improvement in the measurement of access in the literature on the spatial distribution of conventional banks versus alternative financial institutions. In this section, we describe the process we used to calculate access in our analysis.

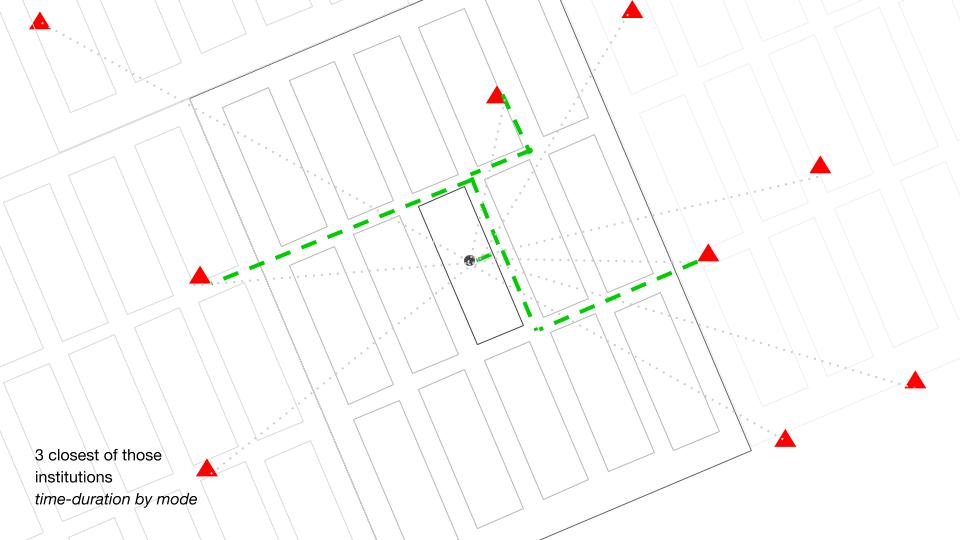
To minimize spatial aggregation error associated with accessibility measure calculations, we calculated our measures at the most granular census geography available: census blocks. We then summarized the block-level data for block groups when performing our analysis in relation to other socio-demographic attributes from census data.



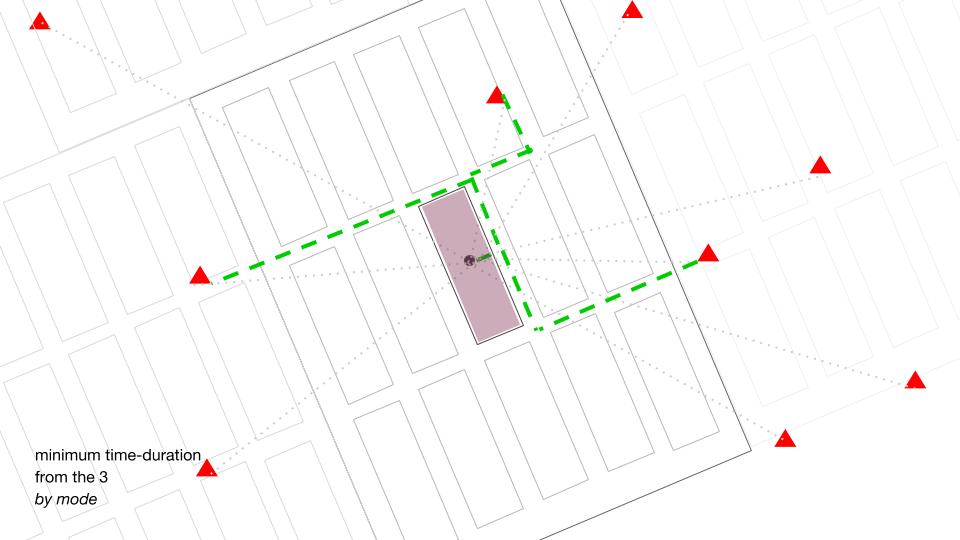
Calculating travel times for all possible combinations of blocks and amenities is not computationally feasible. We instead limited our calculations to the ten closest amenities (conventional banks and AFIs) based on their linear (“as the crow flies”) distance to each individual block’s centroid. We calculated linear distances using Postgres/PostGIS database’s ST\_Distance\_Sphere function.



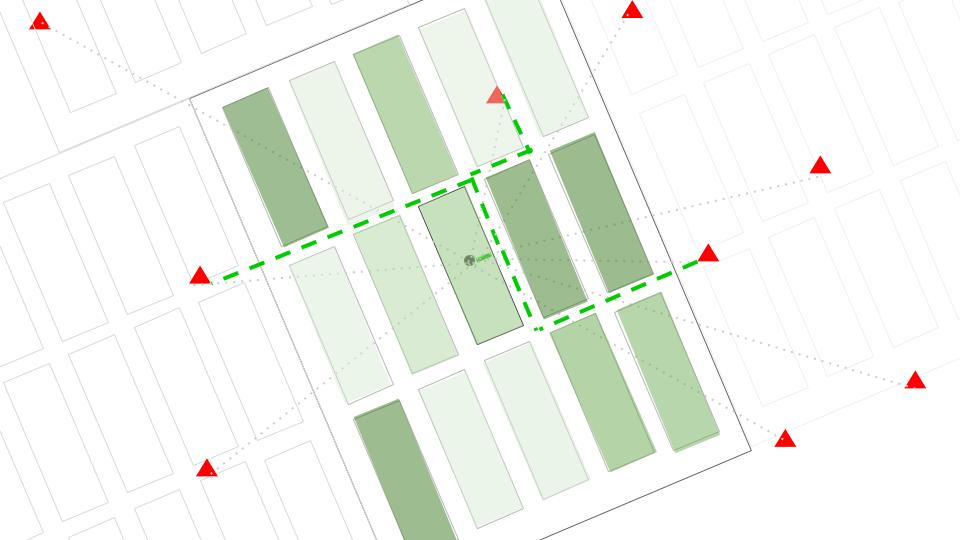
We then used open-source commercial routing engine Graphhopper to estimate the driving and walking times from each block centroid to the ten closest amenities. We repeated this process for all 20 cities in our analysis. We used OTP (Open Trip Planner) to calculate travel times for public transportation. We repeated this process for 19 of the 20 cities in our analysis. Public transportation data were not available for Memphis at the time of our analysis, and thus this city was excluded from public transit calculations.



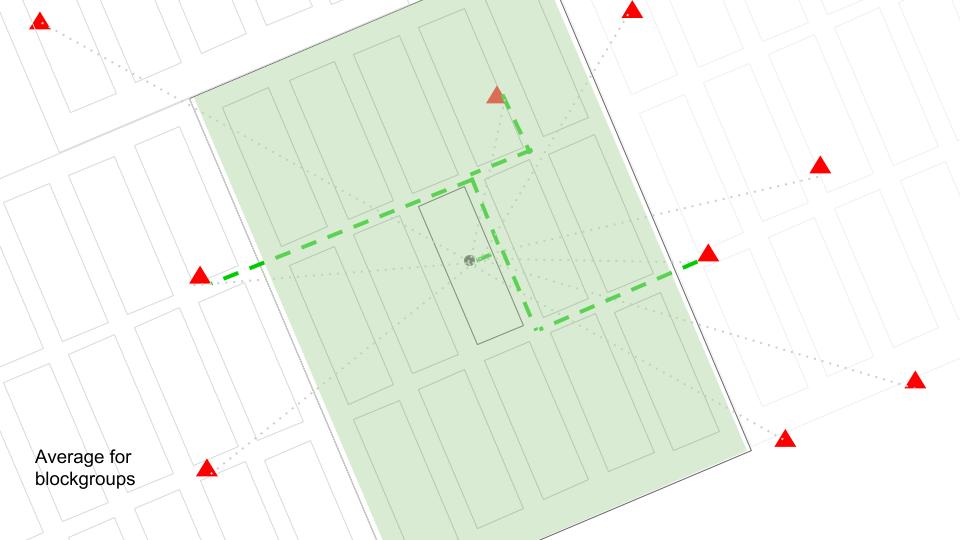
We then selected the three closest amenities using the duration-time measure.



We assigned the minimum of the three distances (vs. average of the three) to each block.



Finally, we calculated the average values of the blocks for assigning the time-duration to each block group.



**3 Estimating adjusted time to nearest AFI vs. nearest bank**

To determine whether the adjusted time to the block group’s nearest AFI was closer than the nearest bank, we estimated random effects models with a binary dependent variable. The model was essentially a two-level hierarchical generalized linear model for block group

*i* in city *j* predicting the log of the odds of that the nearest AFI was closer. The link function was, , where = 1 if time to nearest AFI < time to nearest bank and 0 otherwise. The model takes the following form:

Level 1

where

is the expected log odds that the AFI is closer for block group *i* in city *j*

is the average rate at which the AFI is closer in city *j*

are coefficients associated with the block-group-level variables

are the independent variables

is the specific variation associated with a given block group

Level 2

where

is the average rate at which AFI is closer

is the random “city effect,” or variation associated with a given city

After simple substitution,

The model was estimated in Stata. In Stata, melogit estimates the equation above. However, we used xtlogit, as it produces nearly identical results and allows for more straightforward production of Figures 1 and 2. For code, see [https://github.com/urbaninformaticsandresiliencelab/bnk\_afi\_si/tree/master/scripts/stata](about:blank)

**4 Variables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Description** | **Source** | **Alternative field names** |
| **cityname** | City name. |  | CITYNAME |
| **wht15** | White alone as proportion of total population. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B03002. (July 2019) | WHT15 |
| **blc15** | Black or African American alone as proportion of total population. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B03002. (July 2019) | BLC15 |
| **lat15** | Hispanic or Latino/a American alone as proportion of total population. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B03002. (July 2019) | LAT15 |
| **asi15** | Asian alone as proportion of total population. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B03002. (July 2019) | ASIA15 |
| **oth15** | Other as proportion of total population. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B03002. (July 2019) |  |
| **pov15** | Income in the Past 12 Months Below Poverty Level as proportion of total population for whom poverty status Is determined. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B17021. (July 2019) | POV15 |
| **frn15** | Foreign born as proportion of total population. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B99051. (July 2019) | FRN15 |
| **ump15** | Employment rate as civilian labor force employed as proportion of total civilian labor force population. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B23025. (July 2019) | UMP15 |
| **edu15** | Education level as proportion of population 25 years and over with doctorate degree. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B15003. (July 2019) | EDU15 |
| **own15** | Home owner ship as proportion of population in owner occupied housing, from the total population living in occupied housing units. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B25008. (July 2019) | OWN15 |
| **blb00** | Built before 2000 as proportion of total housing units built before 2000. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B25036. (July 2019) | BLB00 |
| **hu15** | Housing units. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B25001. (July 2019) | HU15 |
| **hu15sqk** | Housing units per square kilometer. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B25001. (July 2019) |  |
| **vacrat15** | Proportion of housing units not occupied. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Tables B25004, B25001 . (July 2019) | VACRAT15 |
| **ppdnl15** | Population density; Total population, expressed as natural log. | U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B03002. (July 2019) | POPDEN\_NATLOG15 |
| **cmdnpcpt** | Commercial density; Number of commercial establishments per 1,000 total population of occupied housings. | Google Places API, U.S. Census Bureau (2016). 2015-2011 American Community Survey 5 year estimate. Table B03002. (July 2019) | COMDENPERCAPT |

**5 Data and replication code links**

Dataset S1 (external link): Data and Replication Code for Calculating Access

[https://github.com/urbaninformaticsandresiliencelab/bnk\_afi\_si/tree/master/scripts/python](about:blank)

Dataset S2 (external link): Data and Replication Code for the Random Effects Analysis of Time-Based Proximities to Banks and AFIs

[https://github.com/urbaninformaticsandresiliencelab/bnk\_afi\_si/tree/master/scripts/stata](about:blank)

1. Yu, Qian, Christian Szegedy, Martin C. Stumpe, Liron Yatziv, Vinay Shet, Julian Ibarz, and Sacha Arnoud. “Large scale business discovery from street level imagery.” arXiv preprint arXiv:1512.05430 (2015). [Accessed July 2, 2020]. [↑](#footnote-ref-1)
2. Movshovitz-Attias, Yair, Qian Yu, Martin C. Stumpe, Vinay Shet, Sacha Arnoud, and Liron Yatziv. “Ontological supervision for fine grained classification of street view storefronts.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1693-1702. 2015. [↑](#footnote-ref-2)
3. Wojna, Zbigniew, Alexander N. Gorban, Dar-Shyang Lee, Kevin Murphy, Qian Yu, Yeqing Li, and Julian Ibarz. “Attention-based extraction of structured information from street view imagery.” In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 844-850. IEEE, 2017. [↑](#footnote-ref-3)
4. [https://www.theatlantic.com/technology/archive/2012/09/how-google-builds-its-maps-and-what-it-means-for-the-future-of-everything/261913/](about:blank) [Accessed July 3, 2020] [↑](#footnote-ref-4)
5. [https://maps.google.com/localguides](about:blank) [Accessed July 3, 2020] [↑](#footnote-ref-5)
6. https://cloud.google.com/blog/products/maps-platform/beyond-the-map-how-we-build-the-maps-that-power-your-apps-and-business [Accessed July 3, 2020] [↑](#footnote-ref-6)
7. Nahar, Anish. 2017. “Google Maps: The Most Expansive Data Machine.” In *HBS Digital Initiative*. [https://digital.hbs.edu/platform-digit/submission/google-maps-the-most-expansive-data-machine/](about:blank) [Accessed July 3, 2020] [↑](#footnote-ref-7)
8. We thank Markus Mobius for the acquisition and extraction of the Microsoft data. [↑](#footnote-ref-8)
9. Depending on the city, there were between 5% and 721% more places in the Google data, with the starkest differences deriving from Google’s effective identification of ATMs. [↑](#footnote-ref-9)
10. Google Maps Platform (2019). Overview: Places API. Available online at [https://developers.google.com/places/web-service/intro](about:blank). [Accessed June 26, 2020] [↑](#footnote-ref-10)
11. Updated June 29, 2020. The data were collected using a previous version of Google Places APIs, and at the time collecting the data was free of charge. Google has updated its Place API since, and one can now incur a high cost by using the “nearby” function. [↑](#footnote-ref-11)