Assessing Bias in Smartphone Mobility Estimates in Low Income Countries

Sveta Milusheva
World Bank
Washington, DC, USA
smilusheva@worldbank.org

Daniel Björkegren
Brown University
Providence, RI, USA
dan@bjorkegren.com

Leonardo Viotti
World Bank
Washington, DC, USA
lviotti@worldbank.org

ABSTRACT

It has become common for governments and practitioners to measure mobility using data from smartphones, especially during the COVID-19 pandemic. Yet in countries where few people have smartphones, or use mobile internet, the movement of smartphones may not be a good indicator of the movement of the population. This paper develops a framework for approaching potential bias that can arise when measuring mobility with smartphones. Using mobile phone operator records in Uganda, we compare the mobility of smartphones and the basic and feature phones that are more common. Smartphones have different travel patterns, and decrease mobility substantially more in response to a COVID-19 lockdown. This suggests caution when interpreting smartphone mobility estimates in contexts with low adoption.

KEYWORDS

mobile phone data, mobility, COVID-19, data bias, smartphones

ACM Reference Format:
Sveta Milusheva, Daniel Björkegren, and Leonardo Viotti. 2021. Assessing Bias in Smartphone Mobility Estimates in Low Income Countries. In *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '21), June 28-July 2, 2021, Virtual Event, Australia.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3460112.3471968

1 INTRODUCTION

Understanding the mobility of populations is crucial for transportation [15, 19, 23, 35], the spread of disease [2, 7, 25, 28, 31, 37–39, 45, 46], natural disasters [8, 14, 22, 36], and–during a pandemic–measuring social contact [1, 18, 20]. A wide array of recent work uses data collected from the motion of smartphones to infer how people in a society move [9, 11, 12, 27, 29, 30]. Under the COVID-19 pandemic this type of analysis has crossed into the mainstream, with a proliferation of analysis using providers like Google Mobility Reports, Facebook, Unacast, Cuebiq, SafeGraph, and Baidu. But this raises the question: do smartphones move in the same way as the population? This is a concern particularly in societies where few people own smartphones. Smartphone owners are likely to be wealthier, may live in different areas, and may move differently. If so, smartphone mobility estimates may be misleading about how populations move [10]. This paper assesses this question, by comparing how smartphones move to how other types of mobile phones move, in a baseline month, and in response to the shock of the arrival of COVID-19.

Smartphone mobility data has several advantages for measuring mobility: many smartphones have GPS which can provide precise locations and can collect data passively at a high frequency. Additionally, mobility can be measured by independent apps. However, there are many countries where few people own smartphones, and even among adopters, usage is low due to high costs of data and sparse wifi coverage. Another possibility is to measure mobility from operator records, which note the cell towers used to transmit transactions. This is harder for researchers and policymakers to access, but includes the mobility of both smartphones and basic/feature phones. An issue is that location measurement is active, not passive–operators typically only record the locations of individuals when they make a transaction.

In this paper we develop a framework for thinking about two biases that can arise when inferring mobility from digital data: selection into ownership of the technology and selective use of the technology. We use data from a major mobile network operator in Uganda, a lower-income country that has similar phone ownership patterns as other countries in sub-Saharan Africa. We identify likely smartphone users in the dataset, and compare the behaviors of this sample of users to non-data users.

We find that data users (smartphones) have different mobility patterns from users who do not regularly use data packets. Data users have more longer-distance travel, with 13% to 22% more daily trips to non-neighboring counties at baseline on average. Additionally, they decrease mobility more after the COVID-19 lockdown policies relative to non-data users, particularly in the counties most affected by COVID-19 policies. That means that inferring mobility based on smartphones could lead policymakers to erroneously believe that population mobility has dropped more than it actually has. This is in line with research from developed country settings that has found larger decreases in mobility among higher income populations [17, 21, 42].

1.1 Related literature

This paper joins a large literature that uses internal data from mobile phone operators (Call Detail Records, or CDR) to measure the mobility of mobile phones in developing country contexts [3, 4, 6, 16, 19, 25, 44–46]. It is challenging to access these data, however, and these examples are difficult to replicate across countries [26]. As more people have adopted smartphones, it has become

COMPASS '21, June 28-July 2, 2021, Virtual Event, Australia

ACM ISBN 978-1-4503-8453-7/21/06…$15.00
https://doi.org/10.1145/3460112.3471968

common to measure mobility using smartphone apps in developed countries, where many people own smartphones. There is an emerging literature using these measures in developing country contexts [20, 27, 29, 34]. We study a low income context where few people have smartphones.

This paper builds on work that studies biases that arise from measuring mobility using data from mobile phones [13]. These works have primarily assessed two types of bias:

Mobility estimates may not be representative when they are measured on a subset of a population that has adopted a digital technology. [43] finds that mobility differs little by demographics among mobile phone owners in Kenya during an early period of adoption, 2008-2009.

Many digital technologies collect digital trace data only when particular software and features are enabled (e.g., GPS and apps that collect user data), or actions are taken (e.g., calls are placed, for operator data, or app check-ins or posts). [32, 47] find that mobility as measured by where actions are taken can differ from more passively collected GPS measures.

[10] assesses the net of both biases in the US, finding that smartphone mobility measures undercount older demographics when compared to voting records in a national election.

This paper evaluates whether smartphones have different mobility patterns from basic/feature phones in a developing country. In settings like ours, smartphone penetration is still low and the selection bias may be larger because the devices are costly relative to average income. Additionally, often people pre-pay for phone services, and given the high cost of data, this may further limit the population that is captured with smartphone data. This is a setting with limited data to validate indicators; we provide evidence by comparing those with access to a smartphone to those without access within the same dataset and study how the generated mobility indicators differ. In this way we limit any differences that might arise due to the data sources and can focus on the differences in behaviors that are measured by the same source for different types of users.

2 BACKGROUND

Mobile phone subscriptions have grown dramatically in the last two decades, even in low-income countries. In 2005, there were 23 mobile phone subscriptions per 100 people in developing countries; by 2019 there were 103 subscriptions per 100 people [40]. However, in lower income countries few of these phones are smartphones connected to the internet. In Africa, there are only 33 mobile broadband subscriptions per 100 people (Figure 1). Given the lack of data in many African countries, though, there is substantial interest in statistics generated using smartphones. We consider what biases may arise when studying population mobility based on smartphones.

We focus on Uganda, which has similar rates of adoption to other countries in sub-Saharan Africa. Per 100 people, Uganda has 57 mobile phones, but only 34 mobile broadband connections (10th percentile and 16th percentile respectively, out of 184 countries reported by ITU in 2018 or 2019). The proportion of mobile phones that have broadband connections (around 0.59), is close to the median across sub-Saharan African countries (0.67).

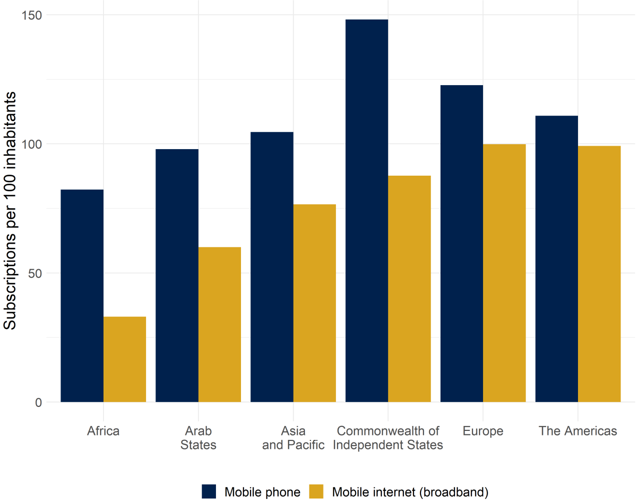


Figure 1: Mobile Phone Subscriptions by Region, 2020

Notes: Regions are based on the regional grouping of the ITU's Telecommunication Development Bureau. Values are June 2020 estimates for 2020 and the data was updated November 2020. Source: ITU World Telecommunication/ICT Indicators database.

Households with smartphones are wealthier and more educated than those without. Table 1 shows demographic characteristics by phone type, as collected by Research ICT Africa (RIA)'s 2018 After Access ICT Access and Use Survey. People without a mobile phone are less likely to have electricity, have fewer years of education, they have lower log household income and a lower number of assets on average. While those with a basic phone show higher values for all of these characteristics they are still lower than those with a feature phone and much lower than those with a smartphone. These patterns in relation to the characteristics of different types of phone owners are consistent across the other eight countries in sub-Saharan Africa where RIA conducted this survey in 2018 (see Appendix for table of statistics). This suggests that smartphones will tend to track the behavior of high income people.

People with internet access self report about twice as much travel as those without access, in the 2016 Demographic and Health Survey (DHS) [41]. Male internet users reported taking an average 11 trips in the last 12 months, but non internet users only 6 (for females this difference was 4 versus 2).[1] Phone owners move more as well: male phone owners took 9 trips compared to nonowners who report taking only 4 (for females, this difference was 3 compared to 2).

We focus on early 2020, with data that brackets the first case of COVID in Uganda (March 21, 2020). A number of strict measures were put in place with the goal of reducing transmission of COVID-19 and that had important effects for mobility. These included suspension of public gatherings on March 18th, the closing of schools and a ban on travel to countries labeled as high risk due to their case numbers. This was followed by a suspension of public transport and a ban on international travel starting on March 25th, and then a lockdown and nationwide curfew from 7pm to 6:30am

[1] Individual level, nationally representative weights are used.

Table 1: Demographics by Phone Owned in Uganda

	Type of Phone Owned			
	(1) Basic	(2) Feature	(3) Smart	(4) None
Percent of Individuals	34.7%	5.9%	8.0%	51.4%
Percent of Phone Owners	71.3%	12.2%	16.5%	-
Years of Education	7.6	10.1	13.1	5.3
Has Electricity	0.6	0.8	0.9	0.5
Number of Major Assets	0.9	1.1	2.1	0.5
Log HH Income	11.8	12.1	12.8	10.9
HH Size	4.9	5.6	4.7	5.2
Observations	685	131	247	801

Notes: Data come from the After Access Africa 2018 survey conducted by Research ICT Africa (RIA). Nationally representative individual weights were applied to produce mean values for characteristics. Number of assets was calculated by summing how many of the following assets were owned by the household: landline, refrigerator, radio, TV, car, motorcycle.

on March 30th. The lockdown was extended past May 5th, but with some restrictions easing after this date, and while the curfew was extended on May 18th, shops, public transport and schools started to reopen at that time in a limited way [24].

The COVID-19 pandemic provides an important example of how these type of mobility indicators generated from mobile phones can be relevant and timely for policymakers in a crisis. Additionally, the context provides an opportunity to study how different measures of mobile phone data portray mobility, in baseline conditions and in response to new policies and shocks.

3 METHODS

3.1 Theory: A sampling problem

Each individual i within the population of interest $\mathcal{N}$ has true mobility given by their full sequence of locations over time, that is,

$$\mathcal{L}_i = (l_{it})_{t \in \mathcal{T}}$$

for each location l_{it} visited at every moment in time $t \in \mathcal{T}$.

Any digital device of type d captures only a sample of this mobility. Measured mobility may differ from the population in two ways:

(1) Device d records only individuals $N_d \subseteq \mathcal{N}$ who have adopted that device. If adopters have different mobility patterns, they may not be representative of the population.

(2) Device d captures locations only at particular times $\mathbf{t}_d \subseteq \mathcal{T}$. For example, smartphones may record a location every few minutes when the GPS is on, or CDR records the tower used when a call is placed. If sampling times $\mathbf{t}_d$ are correlated with location, mobility measures may be biased.[2] The frequency of sampling $\mathbf{t}$ can also affect measures of mobility. If an identical movement pattern is tracked with different devices, with d sampled less frequently than d' ($\mathbf{t}_d \subset \mathbf{t}_{d'}$), d may appear to have less mobility because more location observations are missing.

In our setting, adoption of smartphones is far lower than of any mobile phone, so that $N_{smartphone} \subset N_{anymobilephone} \subset \mathcal{N}$. Our aim is to assess the first potential bias for smartphone owners, relative to any mobile phone, while holding fixed the second type of bias. To do this, we attempt to comparably measure the mobility of smartphone and non-smartphone users within data that includes both.

This approach will not uncover the bias resulting from omitting people who do not have mobile phones at all. This is in contrast to [10], who instead take a particular time t corresponding to the U.S. election and compare smartphone mobility estimates to ground truth poll statistics. Such ground truth data is rare in low income countries.

3.2 Data

We work with the largest mobile network operator (MNO) in Uganda, which supported COVID-19 efforts by allowing access to anonymized, aggregated data to understand mobility and the epidemiology of the disease. As a side effect of operation, MNOs store Call Detail Records (CDR) for billing purposes, which contain a record for each call and internet data use for each account, including a timestamp and the location of the closest cell phone tower.[3] When a user makes calls or uses internet data in different locations, these records reveal that the person has connected with different towers, and thus that the user has moved. We use voice call and data observations for February and April 2020.

This mobility data from operators differs in several respects from commonly used smartphone mobility data [13, 32]. First, crucially, it captures the mobility of both smartphones and basic/feature phones. This allows us to compute the mobility of people with smartphones, who would appear in these common datasets, and the mobility of people with basic/feature phones who are omitted from those datasets. Second, it tends to be less precise, since towers can be spaced out far apart, especially in rural areas. Third, it collects location data only under active use (when a call or data packet is sent), while smartphone mobility data may be collected more passively. (In this context it is fairly common for users to turn off GPS to conserve battery, so smartphone mobility data may be more actively selected than in other contexts.) A fourth difference is that smartphone mobility data is typically reported only for people who have particular apps installed, regardless of their operator (though these apps tend to be common). Our data does not restrict based on apps, but it is only for a single operator which has a large share of the market and has a similar fraction of smartphones as the national average.

3.3 Methods

We identify smartphones based on use of mobile data. We define a user as a *data user* (likely smartphone) if they have at least one internet data transaction per day on average, in at least one of

[2] For example, if smartphone users keep location sensing on when traveling, but turn it off in their neighborhood to conserve battery, they will appear to be away from home more than they actually are. Or, if a user places calls only while at home, they will appear in CDR data to remain stationary at home, regardless of their actual travel.

[3] The data are de-identified, with account numbers replaced by a random ID that can be followed over time.

the three months: February, March or April; and a *non-data user* otherwise.[4] In February, there were a total of 11,818,038 unique subscribers with at least one observation. Of these, 4,299,886 were defined as data users (36% of mobile phones).[5] We define a trip by two consecutive calls from the same user using mobile phone towers located in different counties.

4 ANALYSIS

4.1 Geographic Representation

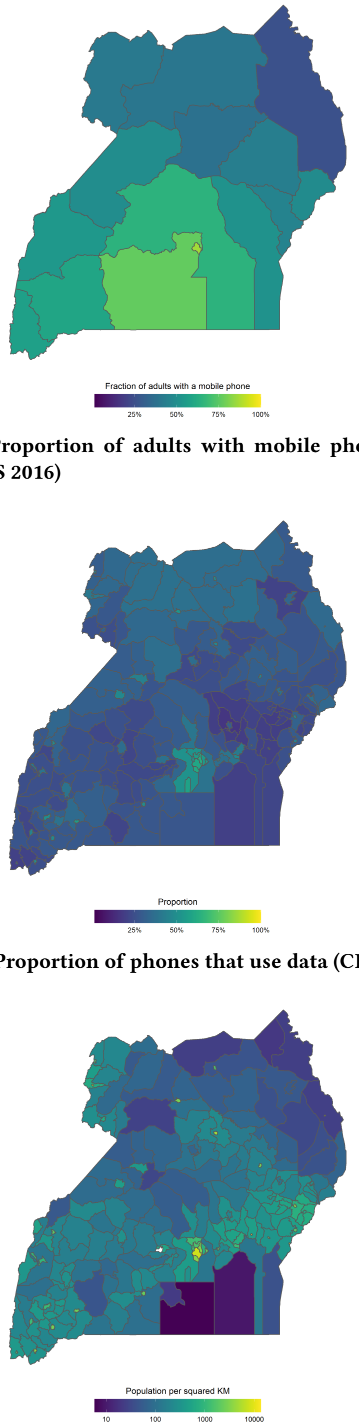
Mobile phone ownership is moderate across Uganda, as shown in Figure 2 Panel a. 86% of adults in the capital of Kampala own mobile phones, and ownership remains high in nearby regions, as reported by the DHS [41]. However, ownership is lower in the periphery, with only 25% of adults owning a phone in Karamoja in the northeast.

Few Ugandans own smartphones, and those that do are predominantly in urban areas. We present the proportion of mobile phones that are smartphones in our CDR data in panel b of Figure 2.[6] In some counties, only 16% of phone subscribers use data; while in others, they are as high as 62%. Comparing the map in panel b of the smartphone proportion in our data with a map of population density in Uganda (panel c) shows that the areas with the highest percentage of data users are the urban, denser areas and the greater Kampala area. These spatial differences suggest that smartphone data is likely to underweight rural areas (see Figure 7 in the Appendix for a comparison of population density versus proportion of data users by county).

4.2 Mobility

We compare the mobility of data users (which would be observed in smartphone mobility data) with non-data users (which would not), to assess whether smartphone mobility is representative of the mobility of all phone owners. We first consider baseline mobility in February, to analyze differences during typical conditions, and then the change during a crisis when multiple policies affected mobility.

4.2.1 Sampling procedure. In order to compare mobility across user types, it is necessary to address the differential use bias: locations are captured only at specific times; in our data, only when transactions are placed. Since we aim to compare data users and non-data users, we aim to equalize bias, but do not claim to remove all bias, by ensuring that the timing and frequency of location observations is comparable across the two groups. In our data for February, we find that each data user has their location observed 29 times per day on average, but each non-data user is observed only 6. This imbalance would inherently lead to differences in the


Fraction of adults with a mobile phone
25%
50%
75%
100%
Proportion
25%
50%
75%
100%
Population per squared KM
10
100
1000
10000

(a) Proportion of adults with mobile phones (DHS 2016)

(b) Proportion of phones that use data (CDR)

(c) Population density (Census 2014)

Figure 2: Geographic Representation

Notes: Panel a uses regionally representative weights. In panel b, proportions were calculated by dividing number of data users by the total number of users, for a given home location in February based on data from the main mobile phone provider that are used in this paper.

[4] We allow a user to qualify in any of the months, as COVID may have affected data usage.

[5] Note that over 6,546,047 million users have at least one data observation during the sample period, but having just one observation is unlikely to be indicative of owning a smartphone. There is a large group of users that have only one or two data observations (potentially they may have a feature phone that allows some limited data use, but is unlikely to have mobile apps which collect smartphone location data).

[6] These proportions were calculated by dividing number of data users by the total number of users, for a given home location in February. Home location is the mode of the location of the last observation of each day that month [25, 33].

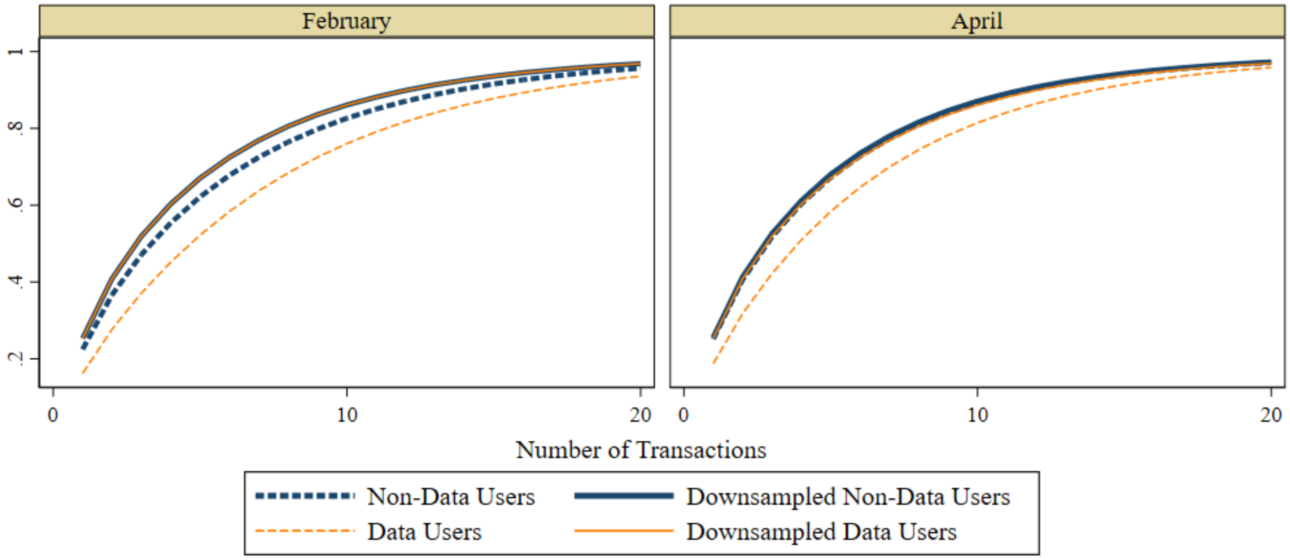


Figure 3: Cumulative Distribution Function of Average Daily Transactions Before and After Downsampling

Notes: Number of users with each daily frequency of voice calls were averaged across the days of the month. For the downsampled figures, the number of voice calls per user were randomly reduced to align the number of daily calls for data users and non-data users and to produce the same daily average number of calls per user of around 5.4.

measurement of mobility between the two groups: we would mechanically see more movement among data users as a function of the higher number of observations.

We explore one approach to correct for differences in active usage between smartphone and other phone users. We subsample location observations, computing mobility using only observations that are deemed more comparable: $\tilde{\mathbf{t}}_{datauser} \subseteq \mathbf{t}_{datauser}$ and $\tilde{\mathbf{t}}_{nondatauser} \subseteq \mathbf{t}_{nondatauser}$. The sampling of locations could differ between data users and non-data users in many subtle ways. The success of this approach will depend on the interaction of usage and mobility, which more work is needed to explore. We demonstrate a proof of concept here, which has two steps:

First, we restrict consideration to locations observed during voice calls, which are less likely to be differentially used between the two types of handsets. We omit data transactions, which are used primarily by smartphones, and SMS, because in instances where a person would send a text, smartphone users may substitute to WhatsApp or other chat apps.[7] This step helps to mitigate a large part of the different phone usage: in February data users have 8 calls per day on average, while non-data users have 6.

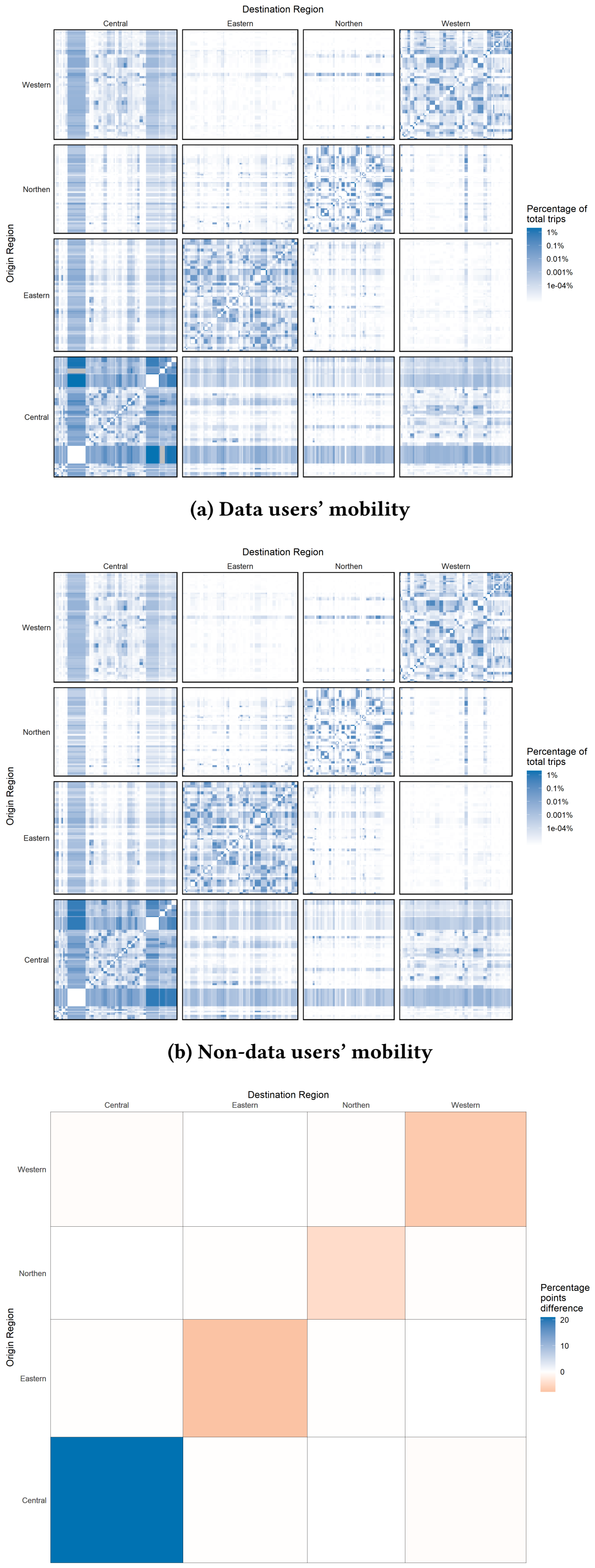
Second, we downsample to account for different temporal resolution between groups, since even after restricting observations to voice calls, data users have more transactions (see Figure 3 for the cumulative distribution function). The downsampling is conducted so as to match the daily distribution of voice call frequency for data users and non-data users. With N_d data users and N_n non-data users, the larger non-data user population is broken up into 2000 bins in order of their daily number of transactions and the number of transactions is averaged per bin. The data user sample is similarly grouped into 2000 bins based on daily call frequency. For each data user, we randomly draw as many calls from their set of calls as the average number of calls in the corresponding bin of non-data users.

We base the daily transaction distribution on a day that is the 10th percentile of daily calls per subscriber for February and April.[8] We apply this distribution for non-data users to downsample both data users and non-data users for each day in February and April. This helps to mitigate any differences in phone usage both across data and non-data users as well as across time.

Prior to downsampling, data and non-data users on average have 7.6 and 6.2 daily observations in February and 6.8 and 5.6 daily observations respectively in April. After the downsampling, both data users and non-data users have 5.4 calls per day on average in February and April and the distributions across months and users are the same (Figure 3).

[7] After we limit to voice observations, the number of subscribers reduces as there are some users that only have data observations. 11,439,718 users remain in total, and 4,074,677 data users remain.

[8] We did not choose the lowest day out of all the days in February and April for calls per subscriber because the lowest day would likely be an outlier and therefore the distribution of calls might be atypical.



(a) Data users' mobility

(b) Non-data users' mobility

(c) Regional difference: data users - non-data users

Figure 4: Origin/Destination Matrices for February 2020

NOTES: Number of trips as proportion of total trips made by group during the month. Cell dimensions are scaled by county population. In panel c the population weighted county differences are summed at the regional level

Downsampling reduces the number of trips we infer for data users from 0.83 to 0.67 daily on average in February, lower than the average of 0.73 trips for non-data users (down from 0.80). This difference could arise from an actual difference in mobility, or if data users' calls do not have the same distribution with respect to travel that nondata users do. Uniform random downsampling would not resolve the temporal sampling problem if, for example, data users are less likely to call businesses for information while traveling. We further break down movement into close and far, looking at trips to neighboring counties versus non-neighboring counties. We find that when focusing on trips further away, both the downsampled dataset and the dataset with all voice observations show that data users have more trips per person (between 12.7% for downsampled and 22% for the full dataset more trips compared to non-data users). For closer trips to neighboring counties, the pattern when downsampling the data is different from the non-downsampled data with trips per person in February for data users being above non-data users for the full dataset and below non-data users for the downsampled dataset (see Appendix Figure 11).

We compare the probability of traveling from an origin location to a destination location across data and non-data users in a baseline time period (February). We then compare their responses to the sudden implementation of policies that limited mobility due to the COVID-19 pandemic.

4.2.2 Baseline mobility. We visualize baseline movement between counties using an origin/destination (OD) matrix, for February 2020 (Figure 4). The y axis represents origins, and the x axis destinations, ordered by county ID and clustered by the four main regions in the country. Each cell is colored by an intensity corresponding to the fraction of trips from that origin to that destination, out of all trips made by that group in the country. The dimensions of each cell are scaled by county population. Panel a shows the OD matrix for data users, while Panel b shows it for non-data users. Because this measure compares the relative amounts of mobility within each group, this will minimize exposure to any remaining transaction frequency bias.

The mobility patterns of data and non-data users share many features. Movement is highly clustered within region, as we see the four regional squares along the diagonal are a darker color. Kampala, in the Central Region, is a mobility hub and the widest cell (given it has the highest population). It is the only row and column that are almost entirely darker, representing links with most other counties.

However, data users' trips are much more concentrated in the central region. Panel c shows the percentage point difference between the two matrices.[9] Trips to Kampala account for only 4% of non-data users' trips but 10% of data users' trips. Data users are relatively less likely to travel within the eastern region, as shown by the higher proportion of negative values (this region had lower smartphone adoption in Figure 2). More generally, they have relatively less travel between all other regions since so much of their travel is concentrated in the central region.

4.2.3 Change in mobility in response to COVID-19. There is often interest in how mobility changes in response to new policies

[9] We compute the population weighted difference over all counties within each region.

[15, 21, 30]. During the COVID pandemic, there was substantial interest in both monitoring changes from baseline, and understanding how policies aimed at mitigating spread of the disease reduced mobility and curbed disease transmission. We compute how mobility changed from baseline before COVID-19 disrupted Uganda (February 2020) to a month under lockdown (April 2020).

Data users decrease mobility 64% more than non-data users (downsampled measure) or 40% (using measure based on all voice calls). Non-data users decrease their daily trips by 14% on average, but data users decrease their trips by 23%, in the downsampled mobility measures.[10] If we instead measure mobility using all voice calls, non-data users decrease their trips by 19%, and data users by 27%.[11] See Appendix Figure 9 for the daily values for trips per subscriber as a percent of the baseline.

We compare the percent decrease in number of trips per person, at the county level, for data users (y axis) versus non-data users (x axis) (Figure 5). If movement decreases equally for data users and non-data users, then the circles on the figure would fall along the red 45 degree line. Instead, data users deviate from the line at the top right. The same conclusion holds when we calculate and plot the values using all voice observations without downsampling (Appendix Figure 8). Given we saw in Table 1 that smartphone owners in Uganda are on average higher income, this aligns with research from developed country settings that has found larger decreases in mobility among higher income populations [17, 21, 42]. This suggests that mobility as measured by smartphones may have subtle but systematic biases in low income populations.

We break down changes in mobility further in the OD matrices in Figure 6. Generally movement has decreased between most county pairs for both data users and non-data users. It is important to note that there are a few county pairs where we see increases in mobility. When we compare the decrease in mobility for data users and non-data users, panel c shows that the largest difference is in movement to and from the Central region. The decreases in movement to and from this region, as well as between counties in this region, is much larger. The ability to break down mobility into such a fine level and analyze the heterogeneity is one of the important benefits of working with mobile phone or smartphone data like this.

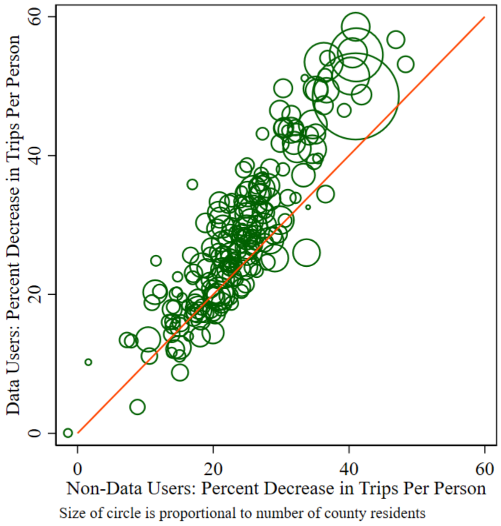


Figure 5: Percent Change in Number of Trips per Person for Data Users and for Non-Data Users by County

Notes: The daily trips per person indicator is calculated by taking an average of the total number of subscribers entering a county on a given day divided by the number of subscribers whose home location is that county that month.

5 DISCUSSION

Digital technologies provide opportunities to better understand populations about which little data has been gathered. However, how an individual is represented in these data depends on whether they have adopted, and how they use these technologies. In particular, there is concern that the poorest segments of the population may be omitted [5].

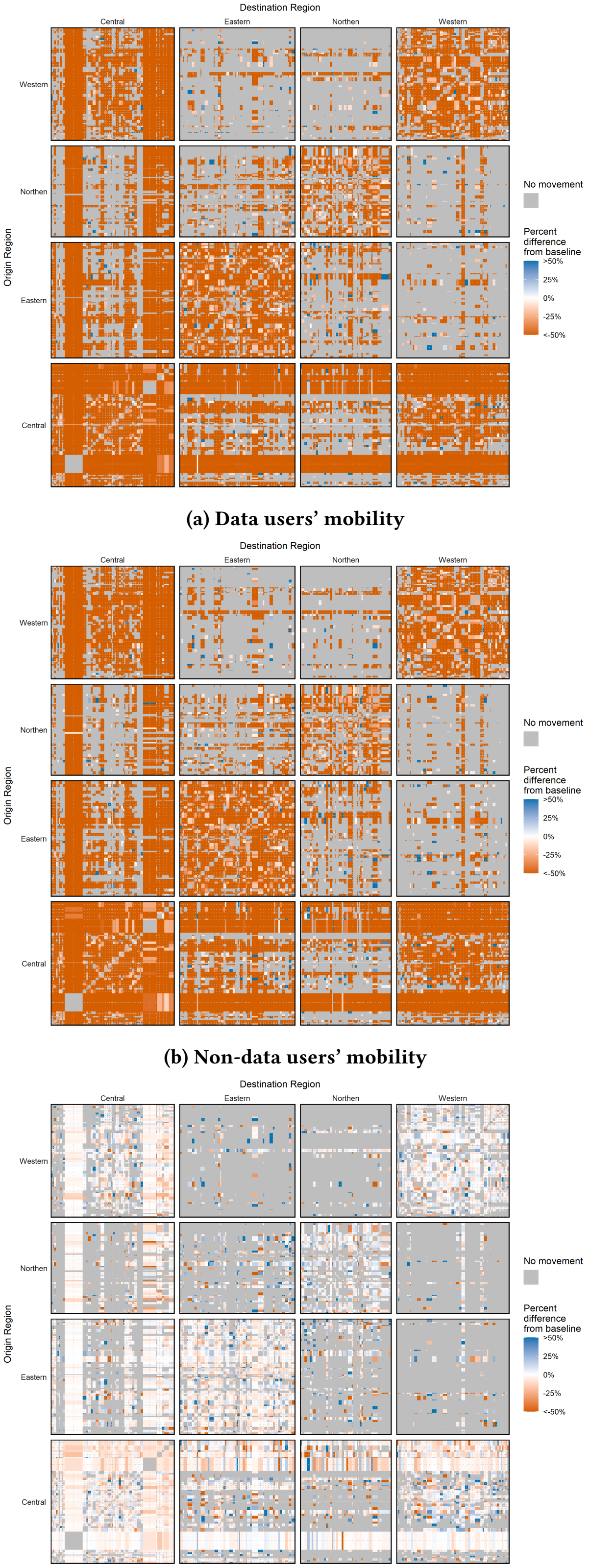
In these cases, ideally one would be able to compare digital measures to an authoritative ground truth [10]. However, in many developing countries, such ground truth data is not available. This paper shows how one digital source of data can be used to better understand what is measured by another.

We use data on an operator that serves all types of phones in Uganda to better understand what can be captured in smartphone mobility data. We find that smart phone adoption is concentrated in urban areas. We find that data that is sampled actively, when a technology is being used, can lead to biases. We demonstrate one simple possible adjustment that involves downsampling, but more work needs to be done to better understand the origin of biases in different types of data. We are optimistic that this process can lead to both a better understanding of biases in new forms of data, and feasible corrections that allow them to more inclusively measure behavior.

We find a number of takeaways, regardless of whether the data is downsampled or not. Data users travel to non-neighboring counties more on average during the baseline period, and decreased movement much more in response to COVID-19 lockdowns. Given that those users made more long distance trips on average at baseline, and react the most to the restrictive measures, this potentially means important decreases in the likelihood of spread of the infection within a country. At the same time, the results point to the need for recognizing the limitations of data coming solely from wealthier users. When using data from digital devices, policymakers should consider potential omissions of low-income people.

[10] Decreases are calculated by comparing daily trips per user for each day in April to the average value for the corresponding day of the week in February. We then calculate the average value across days in April.

[11] This aligns with the fact that voice observations per subscriber per day decline from February to April; therefore, it is possible that some of the decrease measured when looking at all observations arises from true decrease in movement and some arises mechanically from a decrease in observations. The downsampling procedure aims to correct for this by equalizing the number of observations per subscriber in February and April, but it may be overcompensating if user mobility behavior is correlated with phone usage behavior.



(a) Data users' mobility

(b) Non-data users' mobility

(c) Difference: data users - non-data users

Figure 6: Origin/Destination Matrices: Percent Change from February to April 2020

Notes: Change in mobility is calculated as the difference in average daily trips in April versus in February.

ACKNOWLEDGMENTS

This research would not have been possible without the support of MTN Uganda, who provided the data as well as their time and technical resources for processing the data in order to support research that could help with the ongoing COVID-19 pandemic. We are grateful to Aidan Coville for valuable comments and Ruiwan Zhang for research assistance. The research has been funded with UK aid from the UK government through the ieConnect for Impact program; with support from the Trust Fund for Statistical Capacity Building III (TFSCB-III), which is funded by the United Kingdom's Foreign, Commonwealth & Development Office, the Department of Foreign Affairs and Trade of Ireland, and the Governments of Canada and Korea; as well as support from the Research Support Budget in the Development Economics Vice-Presidency.

The findings, interpretations and conclusions expressed in this paper do not necessarily reflect the views of the World Bank, the Executive Directors of the World Bank or the governments whom they represent. The World Bank does not guarantee the accuracy of the data included in this work.

REFERENCES

[1] Hamada S Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M Squire, and Lauren M Gardner. 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* 20, 11 (2020), 1247–1254.

[2] Duygu Balcan, Vittoria Colizza, Bruno Goncalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. 2009. Multiscale Mobility Networks and the Spatial Spreading of Infectious Diseases. *Proceedings of the National Academy of Sciences* 106, 51 (2009), 21484–21489.

[3] Linus Bengtsson, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, Kankoe Sallah, Stanislas Rebaudet, and Renaud Piarroux. 2015. Using mobile phone data to predict the spatial spread of cholera. *Scientific reports* 5 (2015), 8923.

[4] Linus Bengtsson, Xin Lu, Anna Thorson, Richard Garfield, and Johan Von Schreeb. 2011. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS medicine* 8, 8 (2011).

[5] Joshua Blumenstock. 2018. Don't forget people in the use of big data for development.

[6] Joshua E Blumenstock. 2012. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. *Information Technology for Development* 18, 2 (2012), 107–125.

[7] Isaac I Bogoch, Oliver J Brady, MU Kraemer, Matthew German, Marisa I Creatore, Manisha A Kulkarni, John S Brownstein, Sumiko R Mekaru, Simon I Hay, Emily Groot, et al. 2016. Anticipating the International Spread of Zika Virus from Brazil. *Lancet* 387, 10016 (2016), 335–336.

[8] Leah Platt Boustan, Matthew E Kahn, and Paul W Rhode. 2012. Moving to higher ground: Migration response to natural disasters in the early twentieth century. *American Economic Review* 102, 3 (2012), 238–44.

[9] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 7840 (2021), 82–87.

[10] Amanda Coston, Neel Guha, Derek Ouyang, Lisa Lu, Alexandra Chouldechova, and Daniel E Ho. 2021. Leveraging Administrative Data for Bias Audits: Assessing Disparate Coverage with Mobility Data for COVID-19 Policy. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 173–184.

[11] Corentin Cot, Giacomo Cacciapaglia, and Francesco Sannino. 2021. Mining Google and Apple mobility data: temporal anatomy for COVID-19 social distancing. *Scientific reports* 11, 1 (2021), 1–8.

[12] Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, Jake Kruse, Dorte Dopfer, Ajay K Sethi, Juan Francisco Mandujano Reyes, Brian S Yandell, and Jonathan A Patz. 2020. Association of mobile phone location data indications of travel and stay-at-home mandates with covid-19 infection rates in the us. *JAMA network open* 3, 9 (2020), e2020485–e2020485.

[13] Kyra H Grantz, Hannah R Meredith, Derek AT Cummings, C Jessica E Metcalf, Bryan T Grenfell, John R Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, et al. 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature communications* 11, 1 (2020), 1–8.

[14] Clark L Gray and Valerie Mueller. 2012. Natural disasters and population mobility in Bangladesh. *Proceedings of the National Academy of Sciences* 109, 16 (2012),

6000–6005.

[15] Rema Hanna, Gabriel Kreindler, and Benjamin A Olken. 2017. Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science* 357, 6346 (2017), 89–93.

[16] Felana Angella Ihantamalala, Vincent Herbreteau, Feno MJ Rakotoarimanana, Jean Marius Rakotondramanga, Simon Cauchemez, Bienvenue Rahoilijaona, Gwenaëlle Pennober, Caroline O Buckee, Christophe Rogier, Charlotte Jessica Eland Metcalf, et al. 2018. Estimating sources and sinks of malaria parasites in Madagascar. *Nature communications* 9, 1 (2018), 3897.

[17] Jonathan Jay, Jacob Bor, Elaine O Nsoesie, Sarah K Lipson, David K Jones, Sandro Galea, and Julia Raifman. 2020. Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. *Nature human behaviour* 4, 12 (2020), 1294–1302.

[18] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Louis Du Plessis, Nuno R Faria, Ruoran Li, William P Hanage, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368, 6490 (2020), 493–497.

[19] Gabriel E Kreindler and Yuhei Miyauchi. 2021. *Measuring commuting and economic activity inside cities with cell phone records.* Technical Report. National Bureau of Economic Research.

[20] Shengjie Lai, Nick W Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R Floyd, Amy Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, et al. 2020. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* 585, 7825 (2020), 410–413.

[21] Minha Lee, Jun Zhao, Qianqian Sun, Yixuan Pan, Weiyi Zhou, Chenfeng Xiong, and Lei Zhang. 2020. Human mobility trends during the early stage of the COVID-19 pandemic in the United States. *PLoS One* 15, 11 (2020), e0241468.

[22] Xin Lu, Linus Bengtsson, and Petter Holme. 2012. Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences* 109, 29 (2012), 11576–11581.

[23] Glenn Lyons and John Urry. 2005. Travel time use in the information age. *Transportation Research Part A: Policy and Practice* 39, 2-3 (2005), 257–276.

[24] Federica Margini, Anooj Pattnaik, Tapley Jordanwood, Angellah Nakyanzi, and Sarah Byakika. 2020. *Case Study: The Initial COVID-19 Response in Uganda.* Technical Report. ThinkWell and Ministry of Health Uganda.

[25] Sveta Milusheva. 2020. Managing the spread of disease with mobile phone data. *Journal of Development Economics* 147 (2020), 102559.

[26] Sveta Milusheva, Anat Lewin, Tania Begazo Gomez, Dunstan Matekenya, and Kyla Reid. 2021. Challenges and Opportunities in Accessing Mobile Phone Data for COVID-19 Response in Developing Countries. *Data & Policy* Forthcoming (2021).

[27] Pierre Nouvellet, Sangeeta Bhatia, Anne Cori, Kylie EC Ainslie, Marc Baguelin, Samir Bhatt, Adhiratha Boonyasiri, Nicholas F Brazeau, Lorenzo Cattarino, Laura V Cooper, et al. 2021. Reduction in mobility and COVID-19 transmission. *Nature communications* 12, 1 (2021), 1–9.

[28] Emily Oster. 2012. Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa. *Journal of the European Economic Association* 10, 5 (2012), 1025–1058.

[29] Pedro S Peixoto, Diego Marcondes, Cláudia Peixoto, and Sérgio M Oliva. 2020. Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PloS one* 15, 7 (2020), e0235732.

[30] Emanuele Pepe, Paolo Bajardi, Laetitia Gauvin, Filippo Privitera, Brennan Lake, Ciro Cattuto, and Michele Tizzoni. 2020. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Scientific data* 7, 1 (2020), 1–7.

[31] R Mansell Prothero. 1977. Disease and Mobility: a Neglected Factor in Epidemiology. *International Journal of Epidemiology* 6, 3 (1977), 259–267.

[32] Gyan Ranjan, Hui Zang, Zhi-Li Zhang, and Jean Bolot. 2012. Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review* 16, 3 (2012), 33–44.

[33] Nick W Ruktanonchai, Patrick DeLeenheer, Andrew J Tatem, Victor A Alegana, T Trevor Caughlin, Elisabeth zu Erbach-Schoenberg, Christopher Lourenço, Corrine W Ruktanonchai, and David L Smith. 2016. Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. *PLoS Computational Biology* 12, 4 (2016), e1004846.

[34] Nick Warren Ruktanonchai, Corrine Warren Ruktanonchai, Jessica Rhona Floyd, and Andrew J Tatem. 2018. Using Google Location History data to quantify fine-scale human mobility. *International journal of health geographics* 17, 1 (2018), 1–13.

[35] Andreas Schafer and David G Victor. 2000. The future mobility of the world population. *Transportation Research Part A: Policy and Practice* 34, 3 (2000), 171–205.

[36] Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, Ryosuke Shibasaki, Nicholas Jing Yuan, and Xing Xie. 2016. Prediction and simulation of human mobility following natural disasters. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 2 (2016), 1–23.

[37] David Stuckler, Sanjay Basu, Martin McKee, and Mark Lurie. 2011. Mining and Risk of Tuberculosis in Sub-Saharan Africa. *American journal of public health* 101, 3 (2011), 524–530.

[38] Clarence C Tam, Mishal S Khan, and Helena Legido-Quigley. 2016. Where Economics and Epidemics Collide: Migrant Workers and Emerging Infections. *The Lancet* 388, 10052 (2016), 1374–76.

[39] Andrew J Tatem and David L Smith. 2010. International population movements and regional Plasmodium falciparum malaria elimination strategies. *Proceedings of the National Academy of Sciences* 107, 27 (2010), 12222–12227.

[40] ITU World Telecommunications. 2021. ICT Indicators database. *https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx* (2021).

[41] Uganda Bureau of Statistics (UBOS) and ICF. 2018. *Uganda Demographic and Health Survey 2016.* Technical Report. BOS and ICF.

[42] Joakim A Weill, Matthieu Stigler, Olivier Deschenes, and Michael R Springborn. 2020. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences* 117, 33 (2020), 19658–19660.

[43] Amy Wesolowski, Nathan Eagle, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee. 2013. The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface* 10, 81 (2013), 20120986.

[44] Amy Wesolowski, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee. 2012. Quantifying the Impact of Human Mobility on Malaria. *Science* 338, 6104 (2012), 267–270.

[45] Amy Wesolowski, CJE Metcalf, Nathan Eagle, Janeth Kombich, Bryan T Grenfell, Ottar N Bjørnstad, Justin Lessler, Andrew J Tatem, and Caroline O Buckee. 2015. Quantifying Seasonal Population Fluxes Driving Rubella Transmission Dynamics Using Mobile Phone Data. *Proceedings of the National Academy of Sciences* 112, 35 (2015), 11114–11119.

[46] Amy Wesolowski, Taimur Qureshi, Maciej F Boni, Pål Roe Sundsøy, Michael A Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O Buckee. 2015. Impact of Human Mobility on the Emergence of Dengue Epidemics in Pakistan. *Proceedings of the National Academy of Sciences* 112, 38 (2015), 11887–11892.

[47] Zengbin Zhang, Lin Zhou, Xiaohan Zhao, Gang Wang, Yu Su, Miriam Metzger, Haitao Zheng, and Ben Y Zhao. 2013. On the validity of geosocial mobility traces. In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks.* 1–7.

6 APPENDICES

Table 2: Demographics by Phone Owned in sub-Saharan Africa

Kenya	(1) No Phone	(2) Basic Phone	(3) Feature Phone	(4) Smartphone
Years of Educ	7.3	9.7	11.4	14.3
Has Electricity	0.5	0.6	0.6	0.9
Log HH Income	8.0	8.9	9.0	9.8
Number of Assets	0.9	1.1	1.3	2.3
Observations	134	512	153	409
Mozambique				
Years of Educ	3.6	5.7	8.3	10.7
Has Electricity	0.4	0.6	0.7	1.0
Log HH Income	6.8	7.5	8.1	8.6
Number of Assets	0.7	1.1	2.5	2.7
Observations	504	427	54	186
Ghana				
Years of Educ	4.5	7.1	7.0	12.3
Has Electricity	0.6	0.9	0.8	1.0
Log HH Income	4.5	5.5	5.9	6.0
Number of Assets	1.3	1.8	2.2	2.5
Observations	266	502	123	309
Nigeria				
Years of Educ	4.1	8.6	10.4	13.5
Has Electricity	0.5	0.9	0.8	1.0
Log HH Income	8.9	9.7	10.0	9.8
Number of Assets	1.1	2.0	2.3	2.6
Observations	628	425	456	299
Rwanda				
Years of Educ	4.1	6.2	7.2	13.0
Has Electricity	0.2	0.5	0.4	1.0
Log HH Income	9.6	10.2	10.6	11.8
Number of Assets	0.4	1.0	1.0	2.1
Observations	551	387	144	129
South Africa				
Years of Educ	7.2	8.6	10.9	12.1
Has Electricity	0.9	0.9	1.0	1.0
Log HH Income	7.1	7.3	7.1	7.8
Number of Assets	2.2	2.5	2.7	3.2
Observations	263	623	133	796
Tanzania				
Years of Educ	5.5	7.2	8.5	11.9
Has Electricity	0.3	0.5	0.6	0.9
Log HH Income	10.5	11.3	11.9	12.4
Number of Assets	0.7	1.1	1.2	2.4
Observations	402	468	77	244
Senegal				
Years of Educ	2.3	3.3	9.3	10.5
Has Electricity	0.7	0.9	0.9	1.0
Log HH Income	10.3	10.8	10.5	11.2
Number of Assets	1.6	1.8	2.4	2.9
Observations	233	564	112	324

Notes: Data come from the After Access Africa 2018 survey conducted by Research ICT Africa (RIA). Nationally representative individual weights were applied to produce mean values for characteristics. Number of assets was calculated by summing how many of the following assets were owned by the household: landline, refrigerator, radio, TV, car, motorcycle.

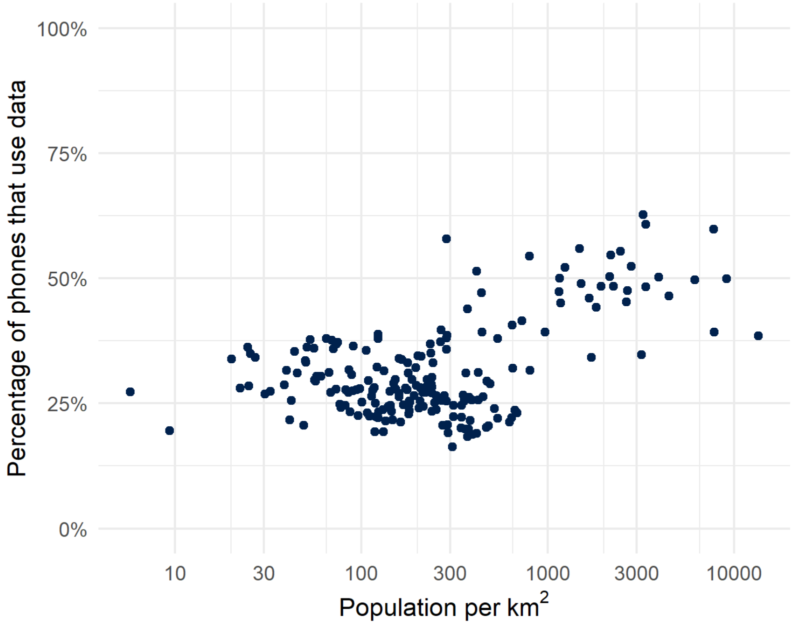


Figure 7: Population Density versus Proportion of Phones that use Data at the County Level

Notes: Values are calculated for 202 counties. Population data to calculate density come from the 2014 Uganda Census. Proportions were calculated by dividing number of data users by the total number of users, for a given home location in February based on the data from the main mobile phone provider that are used in this paper.

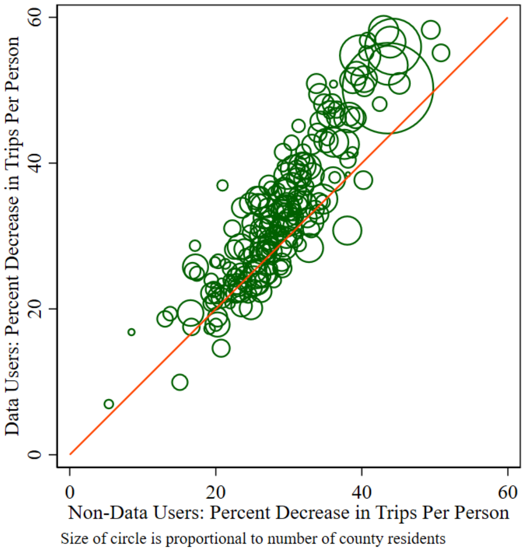


Figure 8: Percent Decrease in Number of Trips per Person for Data Users and for Non-Data Users by County, Using All Voice Observations

Notes: All voice observations are used for data users and non-data users without downsampling. The daily trips per person indicator is calculated by taking total number of subscribers entering a county on a given day and dividing by the number of subscribers whose home location is that county that month.

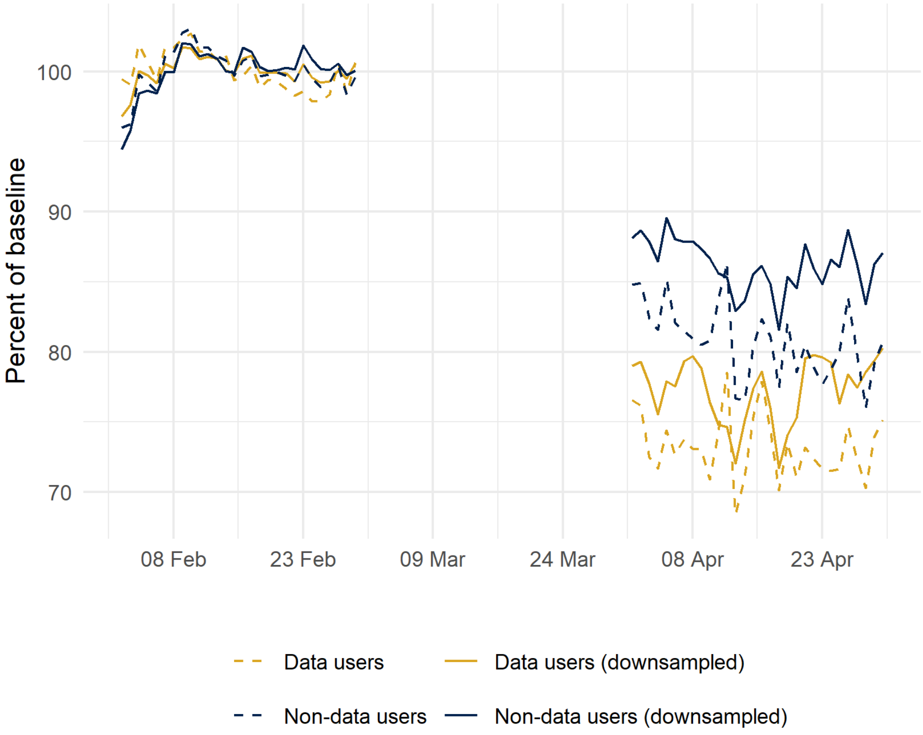
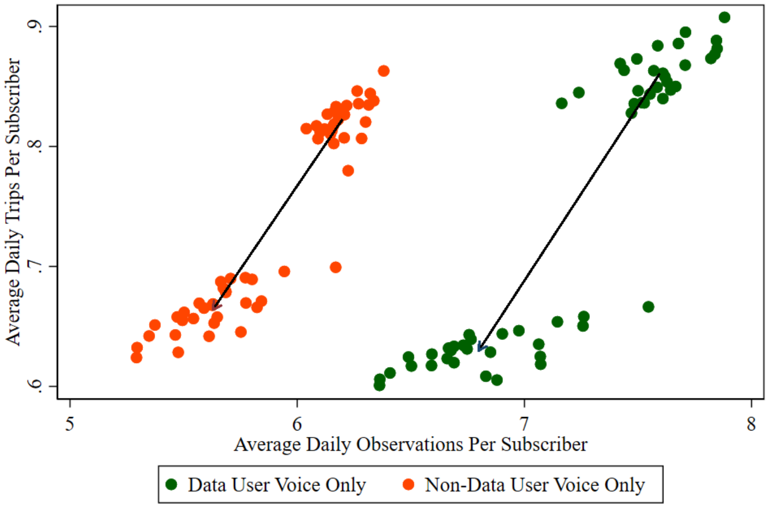


Figure 9: Daily Trips Per Person as a Percent of the Baseline Daily Trips Per Person

Notes: The baseline is defined for each day of the week by averaging days in February.



Black arrows point from the centroid of the cluster of days
in February to the centroid of the cluster of days in April

Figure 10: Average Daily Moves Per Subscriber versus Average Daily Observations Per Subscriber

Notes: Observations and trips per subscriber are calculated at the national level. The values are calculated with all voice observations. The arrows start at the centroids of the February clusters and point to the centroids of the April clusters.

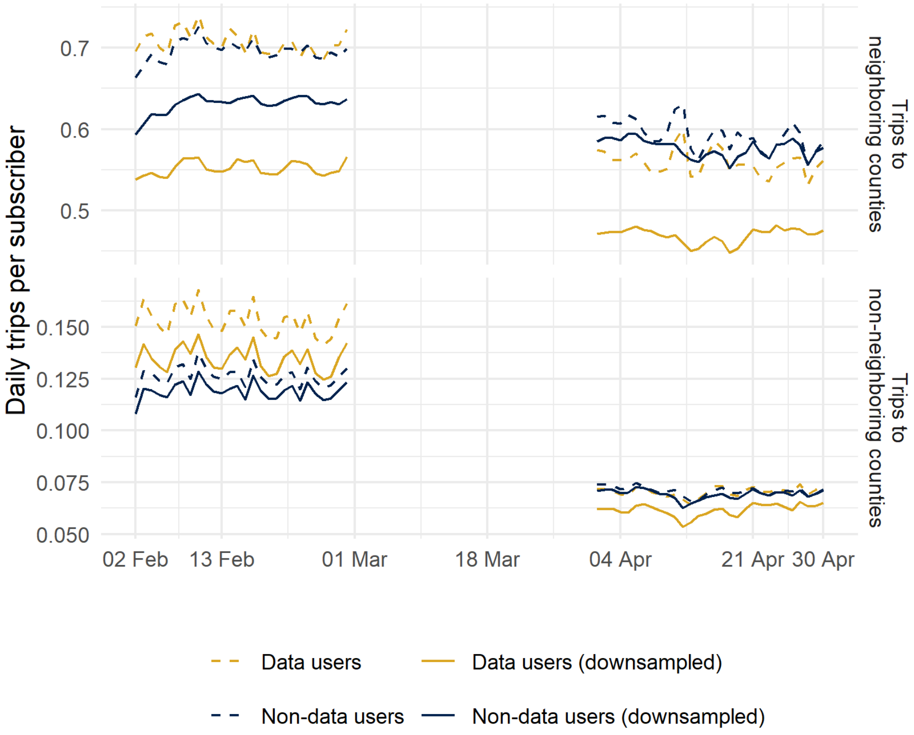
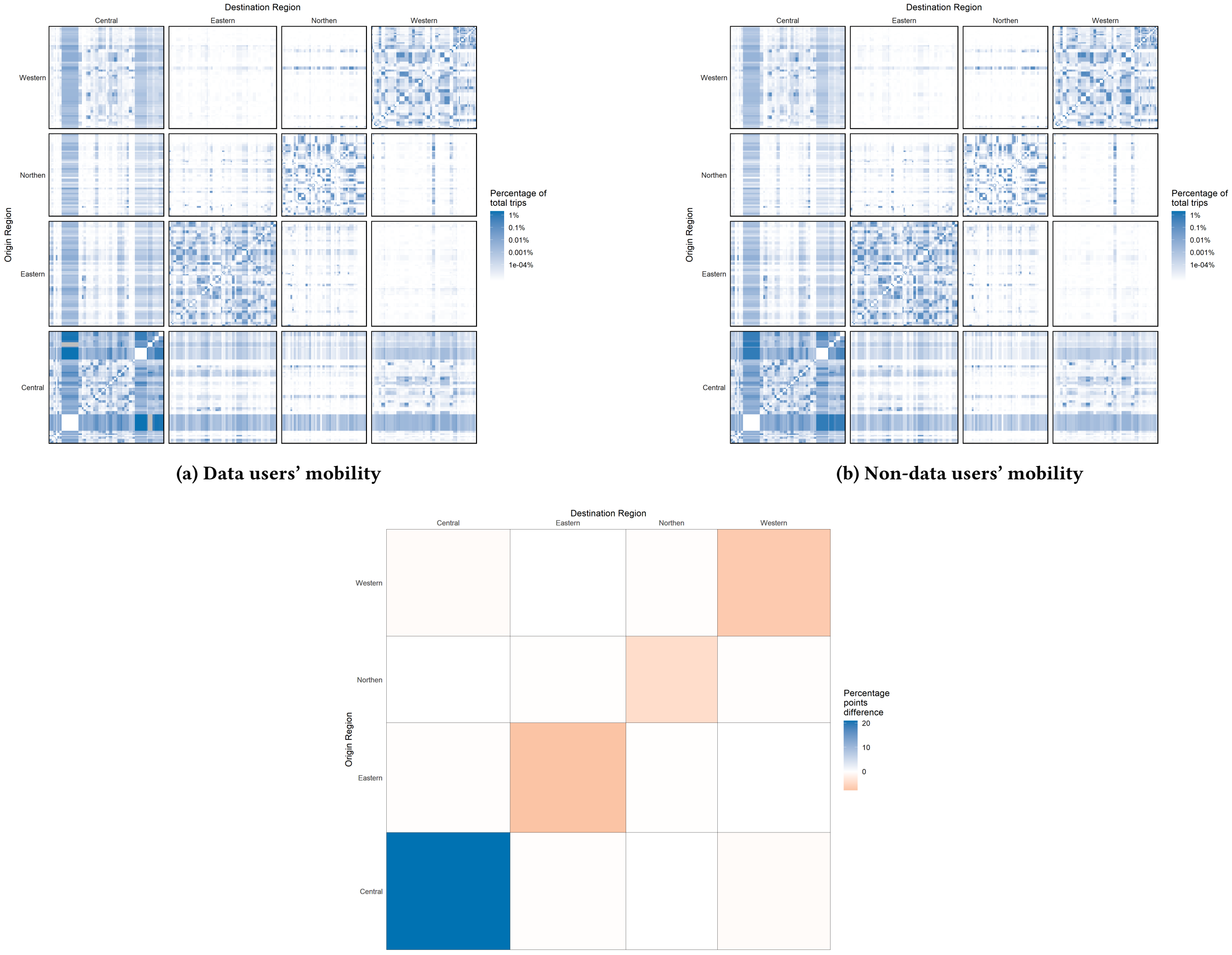


Figure 11: Average Daily Trips Per Day Per Subscriber for Neighboring and Non-Neighboring Counties

Notes: Neighboring county is defined as a county that shares any part of a border with the origin county. Non-neighboring counties are all other counties.

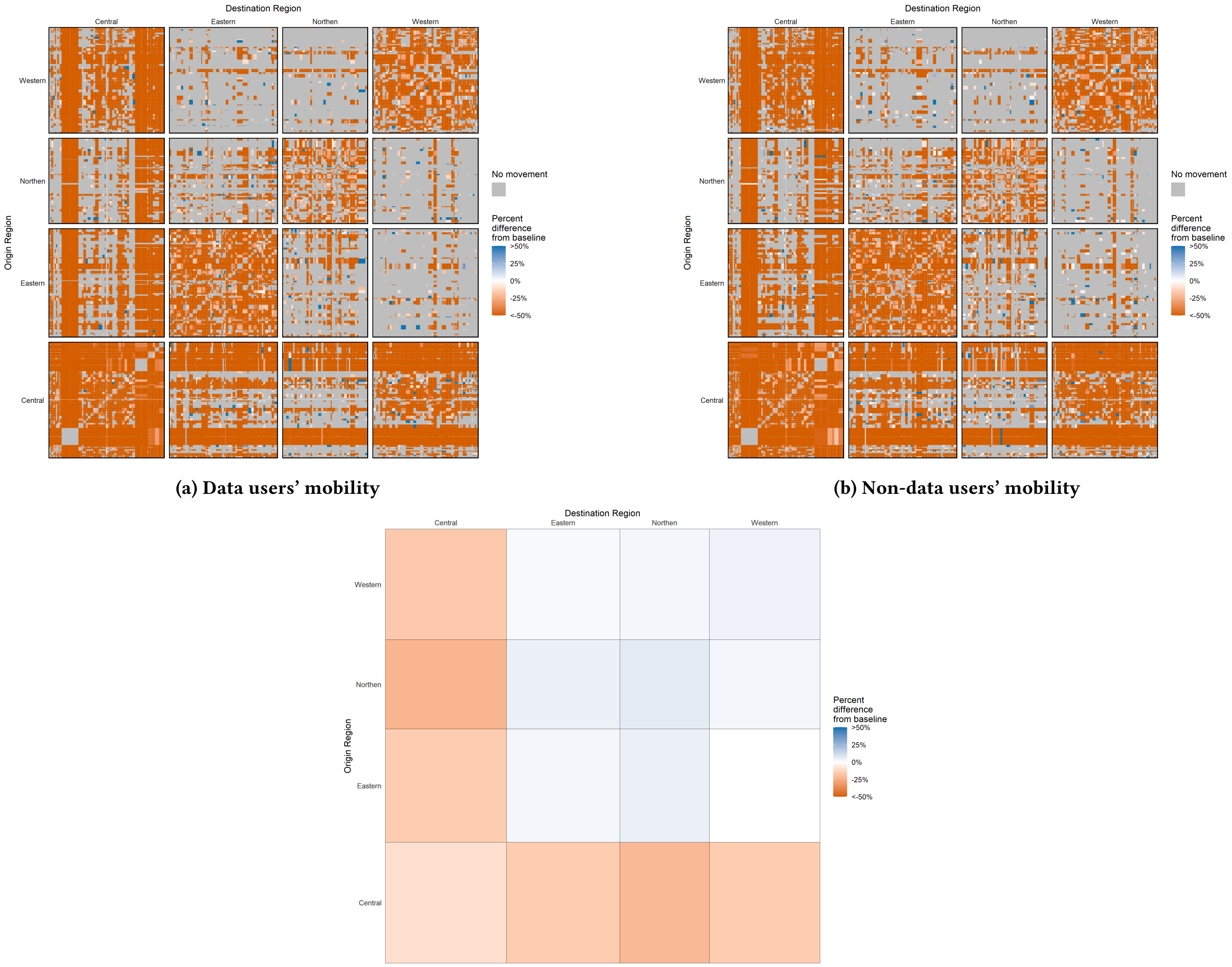


(a) Data users' mobility

(b) Non-data users' mobility

(c) Regional difference: data users - non-data users

Figure 12: Origin/Destination Matrices for February 2020 (No Downsamping)



(a) Data users' mobility

(b) Non-data users' mobility

(c) Regional difference: data users - non-data users

Figure 13: Origin/Destination Matrices: Percent Change from February to April 2020 (No Downsamping)