FLOWMINDER.ORG

UN Mobile Data Training Workshop:

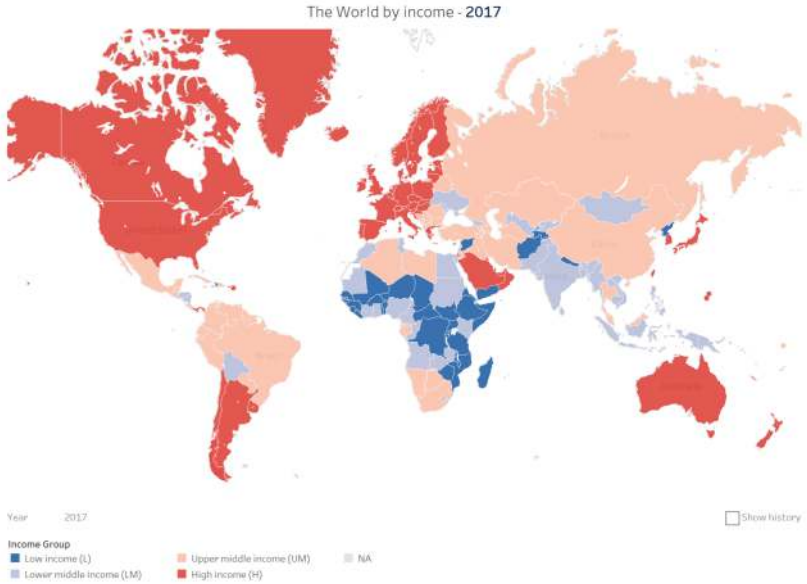
Mobile data analysis in low-income country settings

Tracey Li, PhD
tracey.li@flowminder.org

Mobile data analysis in LMICs vs HICs

Mobile data analysis may be very different in low-income countries than in high-income countries.

Analysis methods used in north America and western Europe may not be appropriate in sub-Saharan Africa.



http://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html

What are the differences?

1. The types of data that the operator collects can be different.
2. The mobile network infrastructure may be different.
3. People use phones in a different way.
4. The behaviours that we study (e.g. mobility) are different.

1. Types of data

Mobile operators in high-income countries often retain signalling data.

This is extremely high-resolution data. The location of any phone that is switched on is recorded once every few seconds or minutes.

Huge volumes of data are generated. A lot of storage space is required, and a lot of processing power if the data are to be analysed.

By Argonne National Laboratory's Flickr page - originally posted to Flickr as Blue Gene. From Argonne National Laboratory. Uploaded using F2ComButton, CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=6412306

1. Types of data

Signalling data can be used for e.g. traffic analysis.

https://www.researchgate.net/publication/275273390_The_Cellular_Network_as_a_Sensor_From_Mobile_Phone_Data_to_Real-Time_Road_Traffic_Monitoring (A. Janacek et al, IEEE Transactions on Intelligent Transportation Systems, 2015)

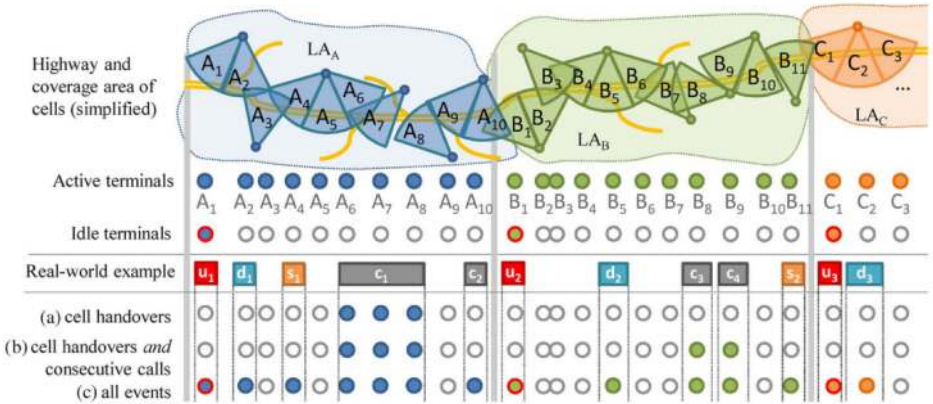


Fig. 3. Schematic overview of signaling event generation on a generic highway—u: location update; d: data connection; s: SMS; c: call; (a) using only cell handover events. (b) extending cell handover events with information created by two consecutive calls. (c) using all event types.

1. Types of data

Signalling data is expensive and resource-intensive for an operator to store and process. Many operators therefore do not retain this data as it is of low value to them too.

Only data that is essential to the operational running of the business is retained. But they always have CDRs as these are used for billing.

Therefore, ‘mobile data analysis’ in low-income countries usually means ‘CDR analysis’.

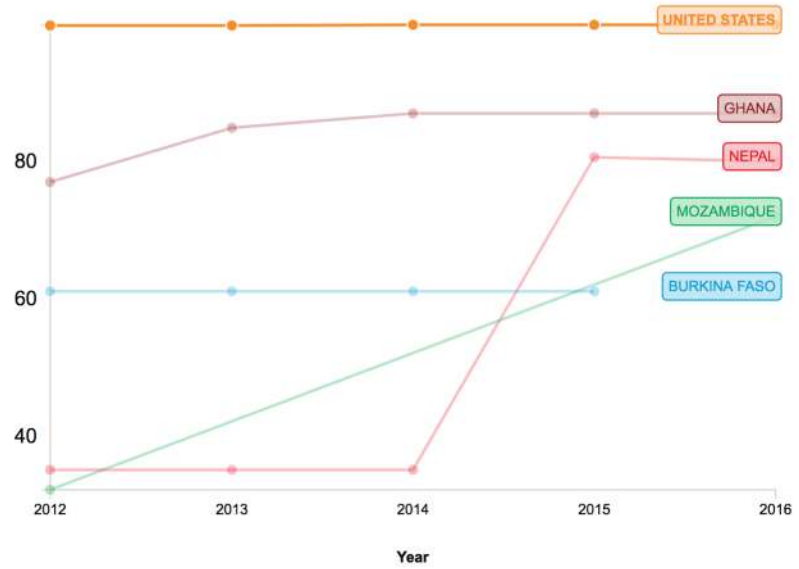
MSISDN	MSISDN_COUNTERPART	CELL_ID	REGION	EVENT_TYPE	TIMESTAMP
AA204V1542DCA00	VEWV782AS945GJE	451154211	north	voice	2016-10-10 15:35:25
AA204V1542DCA01	GNBE72BEA00HE51	451354312	north	voice	2016-10-10 20:03:45
AA204V1542DCA02	EYB470HRAK504EC	451354312	north	voice	2016-10-10 21:21:56
AA204V1542DCA03	PET3621WOZI6220	451354312	north	voice	2016-10-10 21:59:32
AA204V1542DCA04	VEWV782AS945GJE	452616792	central	voice	2016-10-10 22:42:23
B45QHV45CAEVA5	ETG942BCVAEH36L	476126941	south	sms	2016-10-10 08:13:21
B45QHV45CAEVA6	ETG942BCVAEH36L	476126941	south	sms	2016-10-10 08:14:15
B45QHV45CAEVA7	ETG942BCVAEH36L	476126941	south	sms	2016-10-10 08:14:59
B45QHV45CAEVA8	RBY25BAC942HCE4	476126941	south	sms	2016-10-10 12:41:01
B45QHV45CAEVA9	RBY25BAC942HCE5	476126941	south	sms	2016-10-10 13:10:45
B45QHV45CAEVA10	EVG365BCAL246OF	476126941	south	sms	2016-10-10 15:20:43
B45QHV45CAEVA11	PRA19HXME36P64B	413579554	south	voice	2016-10-10 18:08:32
B45QHV45CAEVA12	RVC830RMC29EBB7	413579554	south	voice	2016-10-10 18:54:39
B45QHV45CAEVA13	DOB402VRM70GIBE	413579554	south	sms	2016-10-10 20:53:32
B45QHV45CAEVA14	DOB402VRM70GIBE	413579554	south	sms	2016-10-10 21:21:51
CZW926NRV43WEP1	EBI69BCA033KKK6	486201511	east	voice	2016-10-10 09:01:10
CZW926NRV43WEP2	EBG663JJEB234PM	492500516	east	voice	2016-10-10 21:58:20
CZW926NRV43WEP3	TTBE206B67FDWUT	420594230	central	voice	2016-10-10 12:01:29
CZW926NRV43WEP4	TTBE206D67FDWUT	420594230	central	voice	2016-10-10 15:46:18
DBT396BCW22YTVR	CBQB506BBCLR38Y	455193201	central	sms	2016-10-10 16:28:28

2. Mobile network infrastructure

World Bank data show that > 99% of people in high-income countries have mobile coverage. It is lower for middle and low-income countries.

The number of cell towers per person is usually higher in HICs.

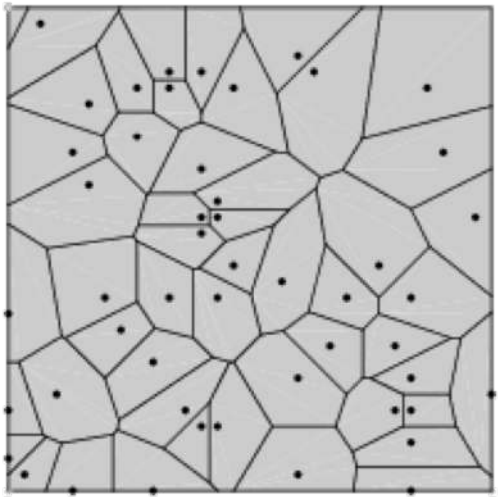
Percent of population with mobile network coverage



https://tcdata360.worldbank.org/indicators/entrp.mob.cov?country=USA

2. Mobile network infrastructure

The density of cell towers affects the uncertainty of the location estimate if the location of the recording cell tower is the only available information.



2. Mobile network infrastructure

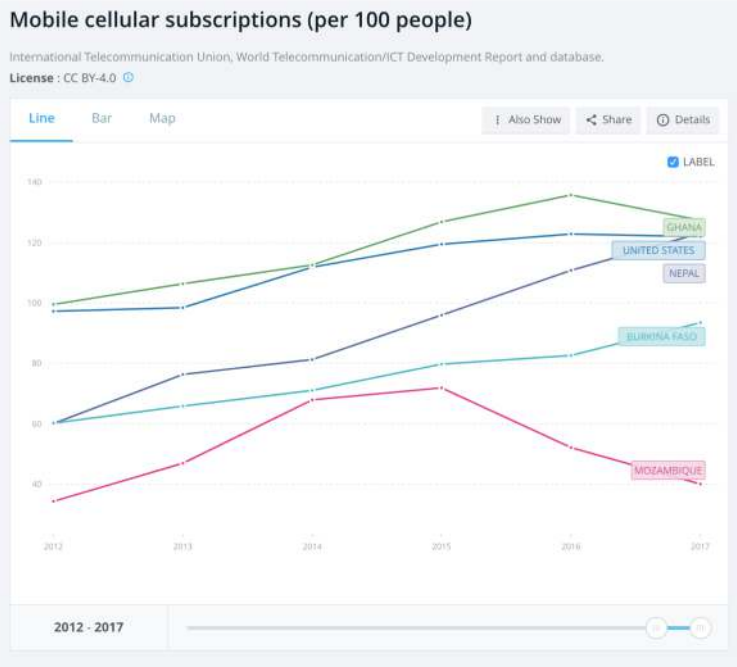
In e.g. Haiti, the uncertainty of the location estimate ~ distance between cell towers. In rural areas this is several tens of km.

If triangulation can be used, then estimates are much better - in Boston (USA), estimates from AirSage have an average uncertainty of 200-300 metres.

http://oro.open.ac.uk/35088/1/hbu2010_final.pdf

3. Differences in phone usage

The number of people with mobile phones, and the number of phones per person, is different.



https://data.worldbank.org/indicator/IT.CEL.SETS.P2?end=2017&locations=US-GH-NP-MZ-BF&start=2012

3. Differences in phone usage

It is more common for people in LMICs to use multiple SIMs.

https://www.gsmaintelligence.com/research/2017/01/variable-network-quality-a-key-driver-of-multi-sim-ownership/597/

“The GSMA Intelligence Consumer Survey 2016 shows that variable network quality is the second biggest driver of multi-SIM ownership.”

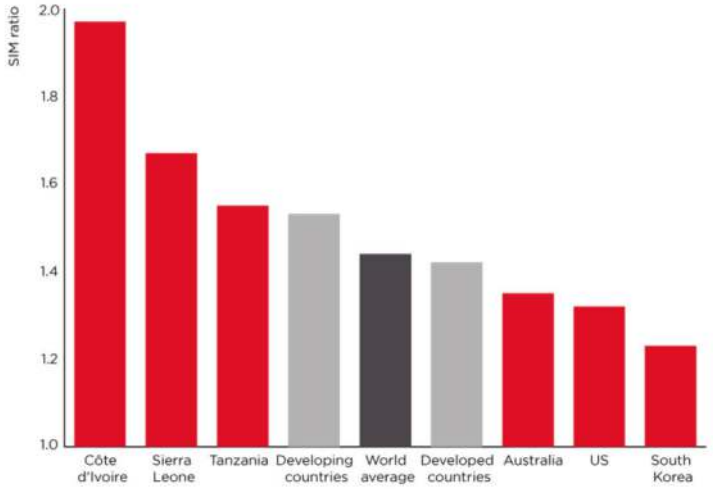


Figure 1: Average number of SIM cards per subscriber

Note: Developing countries average excludes China; Developed countries average excludes Russia.

Source: GSMA Intelligence Consumer Survey 2016

3. Differences in phone usage

https://www.gsmaintelligence.com/research/2017/01/variable-network-quality-a-key-driver-of-multi-sim-ownership/597/

“In developing countries however, more than a third of multi-SIM users claimed they switch between different operators to make use of the best call quality in certain locations.”

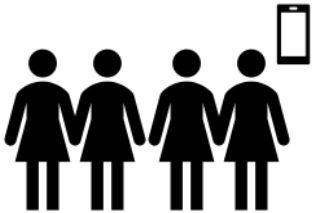
“Price sensitivity also remains a strong factor in countries such as Côte d’Ivoire, DRC, Tanzania and other developing economies, where up to a third of multi-SIM users say they regularly buy new SIM cards to take advantage of discounts and promotions”.

3. Differences in phone usage

People often use multiple SIMs.

And we know from ICT surveys that family members/friends often share a SIM.

Therefore the assumption of “1 SIM = 1 person” isn’t always valid.



3. Differences in phone usage – SIM sharing

We know that many family members often share a phone. Why is this important?

If we have a sequence of locations:

Did 1 person visit all 4 locations [A B C D]?

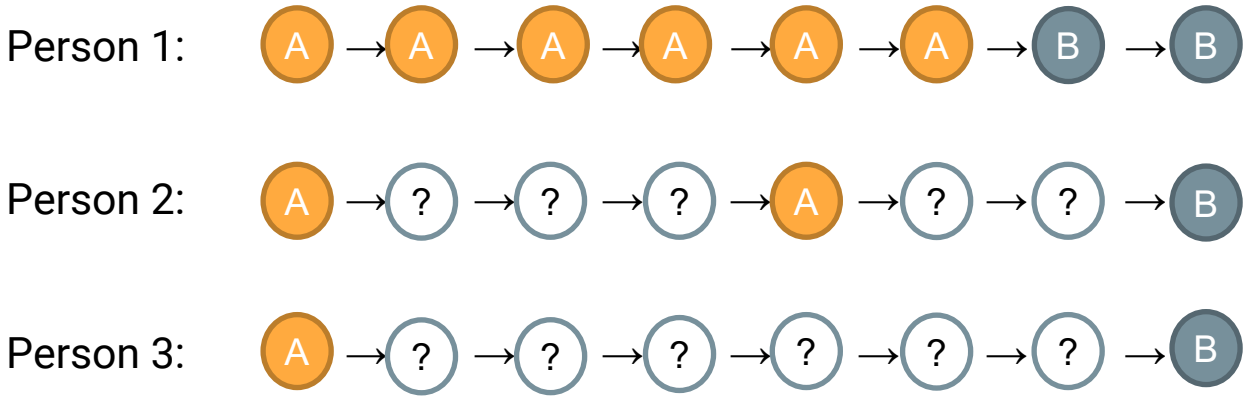
Or did person 1 visit [A B C] and another visit [A B D]?

3. Differences in phone usage – calling frequency

A lot of people use their phones infrequently (e.g. less than once a week) in LMICs.

Imagine you want to know whether someone has been displaced from their home.

Here are the locations that you see people at each day:



3. Differences in phone usage – calling frequency

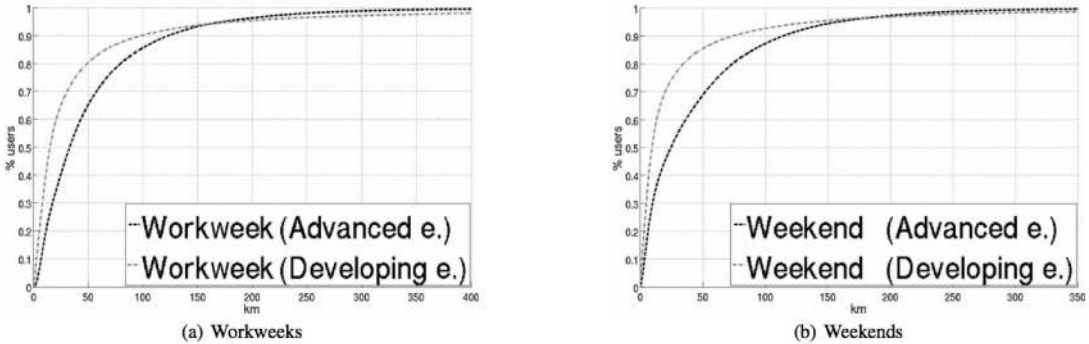
What would you assume about the days when there is no data?

What conclusions would you draw? How many people were displaced?

4. Differences in mobility behaviour

Mobile phone data allows us to observe mobility behaviour. This may be inherently different in different countries.

https://www.aaai.org/ocs/index.php/SSS/SSS10/paper/viewFile/1095/1357 (A.Rubio, Artificial Intelligence for Development, 2010)



(a) Workweeks (b) Weekends

Figure 4: CDF of the average diameter of the area of influence during (a) workweeks and during (b) weekends.

4. Differences in mobility behaviour

It is important to think whether observed differences are really due to inherent differences in behaviour, or whether they are artefacts of reduced temporal or spatial resolution.

Example: if Person 1 has been seen to visit 2 locations, and Person 2 has been seen to visit 4 locations, does Person 1 travel less than Person 2, or do they just use their phone less?

Example analyses that are relevant in LMICs

Mobility in post-disaster scenarios

Disease transmission e.g. malaria

Dynamic population mapping – fluctuating demand for public services

Poverty-mapping

FlowKit:

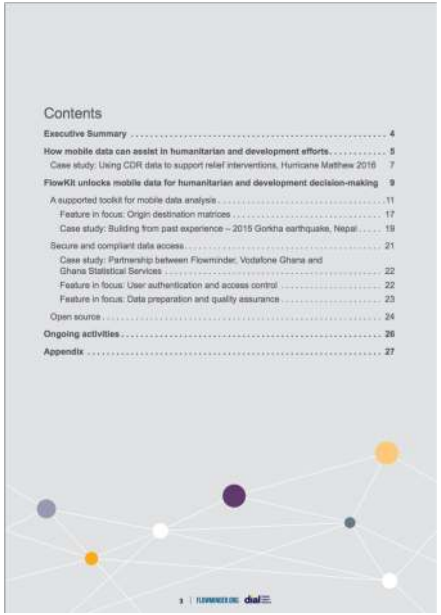
An open-source tool to enable access to, and analysis of, mobile phone data

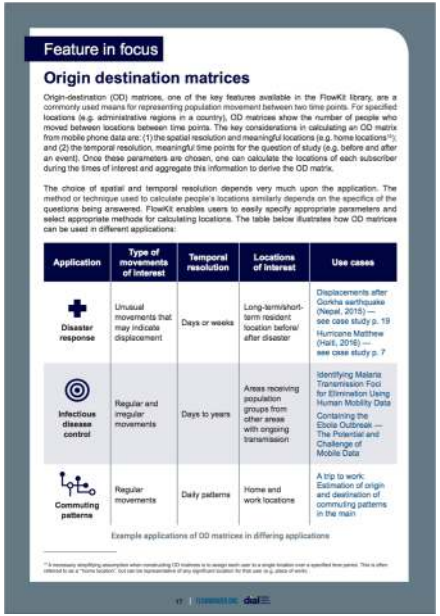
What is FlowKit?

- An open-source (free, publicly available) set of tools developed by Flowminder.
- Supported by Digital Impact Alliance.
- Provides an interface between the mobile operator’s data and the analyst.
- Provides operator with fine control and visibility into who can access what data.
- Enables people with no previous experience to quickly begin basic analysis.
- Enables experienced analysts to obtain more flexible access to the data.

FlowKit white paper

https://digitalimpactalliance.org/wp-content/uploads/2019/02/FlowKit_UnlockingthePowerofMobileData.pdf


Contents


Feature in focus

Origin destination matrices

Origin-destination (OD) matrices, one of the key features available in the FlowKit library, are a commonly used means for representing population movement between two time points. For specified locations (e.g. administrative regions in a country), OD matrices show the number of people who moved between locations between time points. The key considerations in calculating an OD matrix from mobile phone data are: (1) the spatial resolution and meaningful locations (e.g. home locations[*]); and (2) the temporal resolution, meaningful time points for the question of study (e.g. before and after an event). Once these parameters are chosen, one can calculate the locations of each subscriber during the times of interest and aggregate this information to derive the OD matrix.

The choice of spatial and temporal resolution depends very much upon the application. The method or technique used to calculate people's locations similarly depends on the specifics of the questions being answered. FlowKit enables users to easily specify appropriate parameters and select appropriate methods for calculating locations. The table below illustrates how OD matrices can be used in different applications:

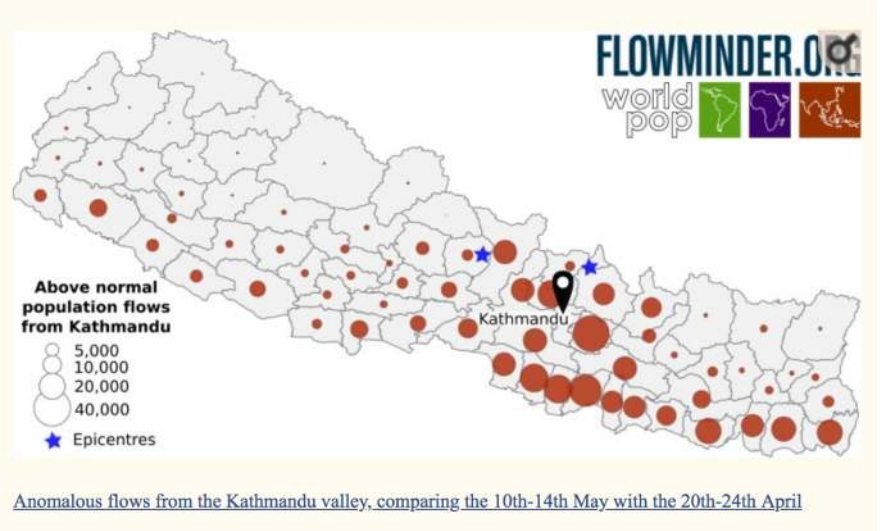
Application	Type of movements of interest	Temporal resolution	Locations of interest	Use cases
Disaster response	Unusual movements that may indicate displacement	Days or weeks	Long-term/short-term resident location before/after disaster	Displacements after Gorkha earthquake (Nepal, 2015) — see case study p. 19 Hurricane Matthew (Haiti, 2016) — see case study p. 7
Infectious disease control	Regular and irregular movements	Days to years	Areas receiving population groups from other areas with ongoing transmission	Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data Containing the Ebola Outbreak — The Potential and Challenge of Mobile Data
Commuting patterns	Regular movements	Daily patterns	Home and work locations	A trip to work: Estimation of origin and destination of commuting patterns in the main

Example applications of OD matrices in differing applications

[*] A necessary simplifying assumption when constructing OD matrices is to assign each user to a single location over a specified time period. This is often referred to as a "home location", but can be representative of any significant location for that user (e.g. place of work).

Demo – example analysis

Analyse the effect of a disaster by measuring the change in behaviour (flows) after a disaster.



Anomalous flows from the Kathmandu valley, comparing the 10th-14th May with the 20th-24th April

R.Wilson et al, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4779046/