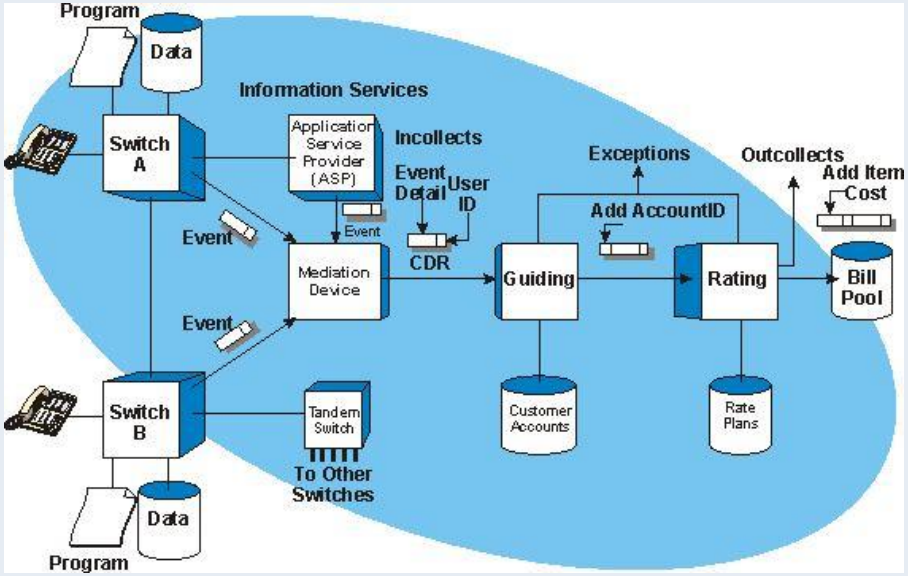
Call Data Records

Lecture 27: CSE 490c



Topics

- Data Science for Development
- AI for Social Good

- Today
  - Mobile Phone Data

What data does T-Mobile know about you?

Telco Information

- Call Data Records (Call Detail Records)
  - Meta data on individual calls
- Cell Tower Logs
  - Information of handset connections with towers

- These are generally managed separately
- Call Data Records are maintained for billing

Other information sources

- Cell tower locations
  - Mapping from Tower IDs to geo coordinates
  - Proprietary or crowd sourced
- Mapping of IEMI numbers to handset models
  - Either exact model or class of phone

Access to cell phone data

- Proprietary to Telco
- Provide competitive advantage
  - Possibly for marketing or data services
  - Linkage with mobile money
- Subject to government privacy regulations
- Possible access to aggregated data

Data protection

- Risk in disclosure of personal information
- Various approaches to anonymization
  - Hashing of phone numbers
  - Aggregating cell towers
- Bucketing of data

What is the distribution of phone types in rural areas

- Debate on smart phone vs. basic phone prevalence
- Who owns 2G vs 3G vs 4G phones
- Cell tower logs
  - Record of every phone connection to base station
  - Type Allocation Code (TAC) indicates model
- Track simcard and phone separately
- Analysis
  - Phones associated with the community
  - Sharing of phones
  - Upgrades and downgrades

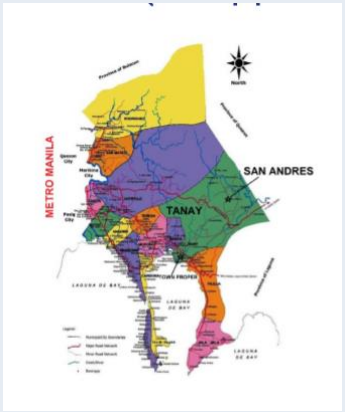
Access to data

- Vast amount of data with Telcos
- But what if you aren't a telco?
- Analysis of community cellular logs from Philippines and Indonesia

Shah et al. (2017) An Investigation of Phone Upgrades in Remote Community Cellular Networks

TIMESTAMP	IMEI (International Mobile Equipment Identity)	IMSI (International Mobile Subscriber Identity)

Cell Phone Study

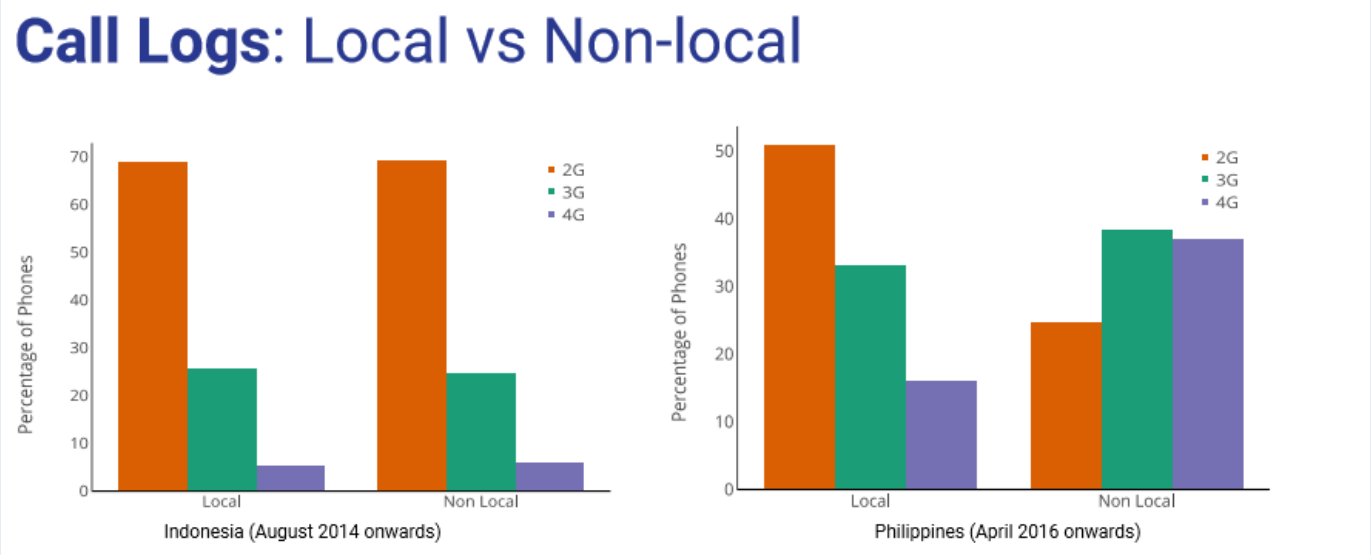


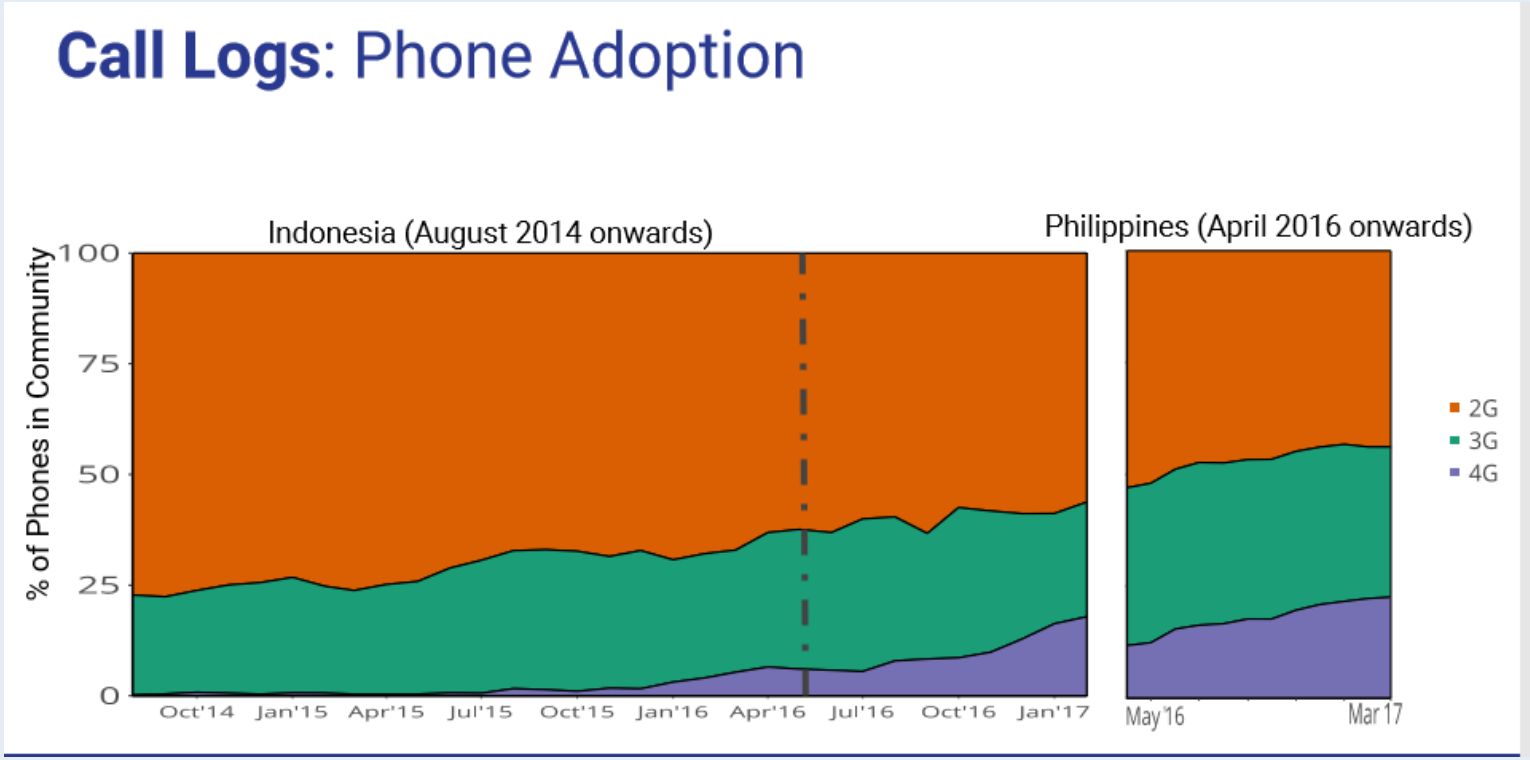
- Kushal Shah, Kurtis Heimerl
- Call records from community networks
  - Philippines, 300K records
  - Papua, 115K records
- Log data from all users
  - Even from users not in the network

Definitions

- Local user
  - SIM registered with local telecom or in the community for at least 10 days
- Primary phone
  - Phone which connects to the network most often during the day
- Phone sharing
  - If phone changes users with operation returning to earlier user
- Phone upgrade or downgrade
  - If a user changes primary phone but does not switch back

Results


Call Logs: Local vs Non-local
Percentage of Phones
70
60
50
40
30
20
10
0
2G
3G
4G
Local
Non Local
Indonesia (August 2014 onwards)
Percentage of Phones
50
40
30
20
10
0
2G
3G
4G
Local
Non Local
Philippines (April 2016 onwards)

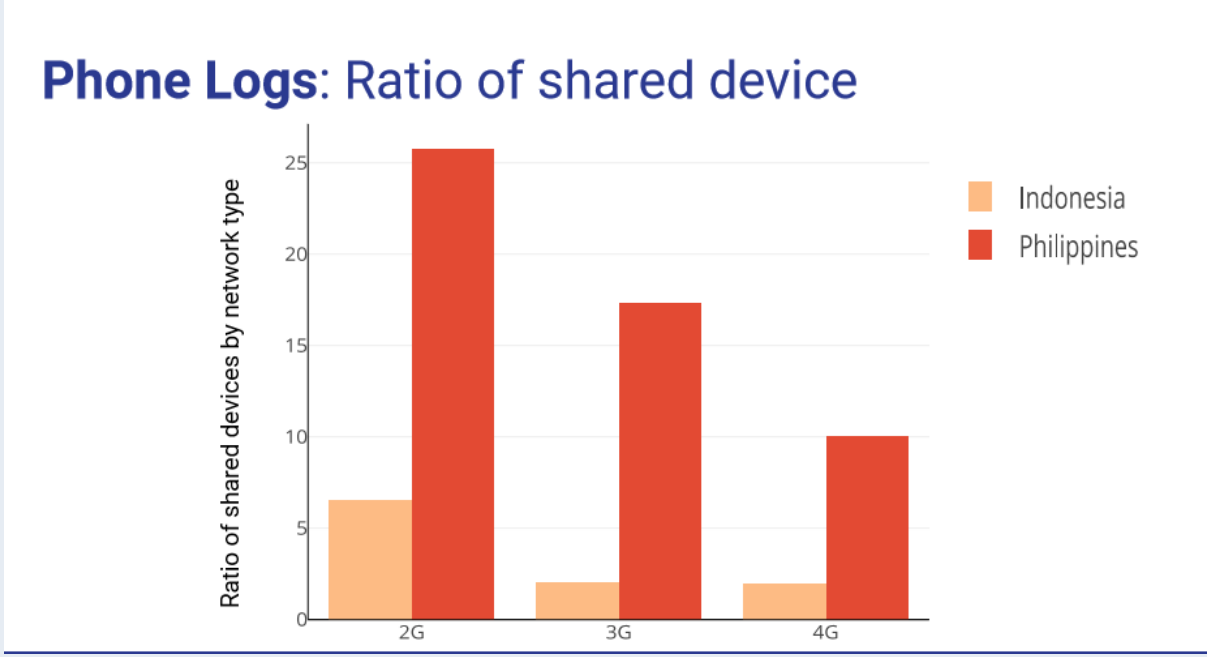

Call Logs: Phone Adoption
Indonesia (August 2014 onwards)
Philippines (April 2016 onwards)
% of Phones in Community
100
75
50
25
0
Oct'14
Jan'15
Apr'15
Jul'15
Oct'15
Jan'16
Apr'16
Jul'16
Oct'16
Jan'17
May'16
Mar'17
2G
3G
4G

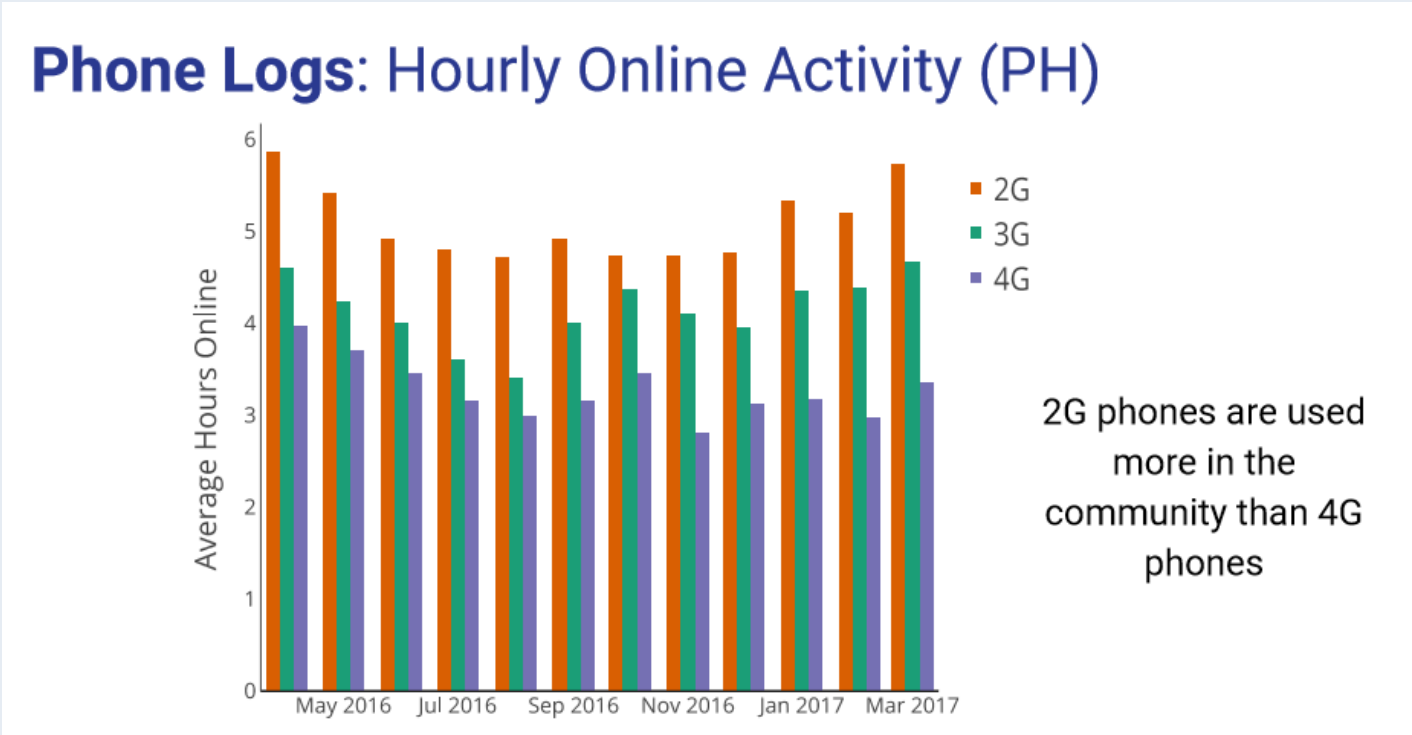
Phone Upgrades

Call Logs: Phone Upgrades

From:	To: 2G	To: 3G	To: 4G
Philippines			
2G	39.75%	3.56%	13.79%
3G	4.70%	3.58%	3.18%
4G	12.38%	3.66%	15.51%
Indonesia			
2G	55.38%	1.32%	13.26%
3G	1.60%	1.65%	0.85%
4G	14.73%	0.94%	10.25%

Other results





Call Data Records

- Meta data associated with calls
- Source number
- Destination number
- Source Tower (ID)
- Destination Tower (ID) [might be missing]
- Time
- Duration
- Status

Call Detail Records

LATITUDE	LONGITUDE	DATE	TIME	NUMBER	NAME	DURATION
44.50880 N	73.18223 W	1/28/2008	0917	802-555-1234	Chittenden Bank	0:10:17
44.50880 N	73.18223 W	1/28/2008	0942	802-555-8673	Poopsie LaRue	0:01:03
44.50880 N	73.18223 W	1/28/2008	0945	802-555-9201	Hanley Strappman	0:05:32
44.27834 N	73.21263 W	1/29/2008	2205	802-555-7758	Verizon Voice Mail	0:01.13
44.27834 N	73.21263 W	1/29/2008	1532	802-555-4492	Widgets LLC	0:03:47
44.27834 N	73.21263 W	1/29/2008	2209	802-555-7758	Verizon Voice Mail	0:00.36
44.50880 N	73.18223 W	1/30/2008	0830	202-555-1818	British Embassy	0:18:12
44.27834 N	73.21263 W	1/30/2008	2208	802-555-7758	Verizon Voice Mail	0:00.53
44.27834 N	73.21263 W	1/30/2008	2211	802-555-8673	Poopsie LaRue	0:06:18
44.50880 N	73.18223 W	1/31/2008	0903	202-555-1843	British Embassy	0:03:21
44.50880 N	73.18223 W	1/31/2008	0908	416-555-9834	British Embassy	0:22:04
44.4143 N	73.03561 W	1/31/2008	1047	802-555-9201	Hanley Strappman	0:01:02
44.4143 N	73.03561 W	1/31/2008	1050	213-555-2761	M. Fendell	0:09:06
44.25295 N	72.58229 W	1/31/2008	1127	802-555-9201	Hanley Strappman	0:05:38

Types of studies

- How people use technology
- Populations studies (where people are)
- Event studies (what happens when)
- Epidemiology studies
- Economic studies

Working with CDRs

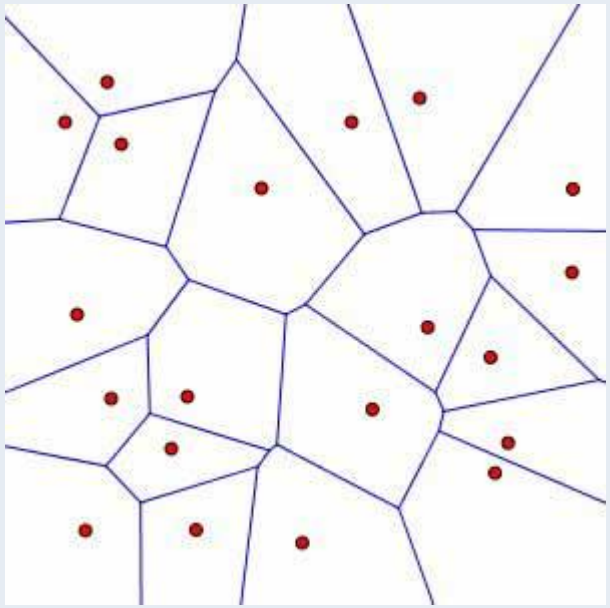
- Preprocess data for higher level structure
- Align data with other sources
  - Tower data
  - Economic / Population Data

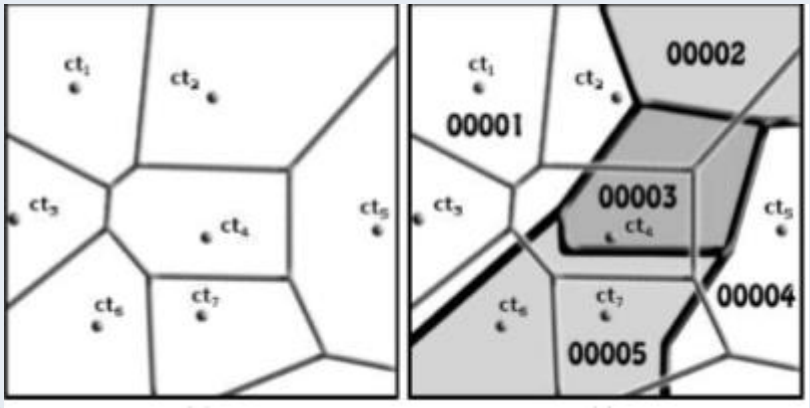
Location

- Tower information
- Computing home location
  - Multiple algorithms and heuristics
- Change in home location
- Commuters
- Truckers
- Migrants

Matching locations to areas

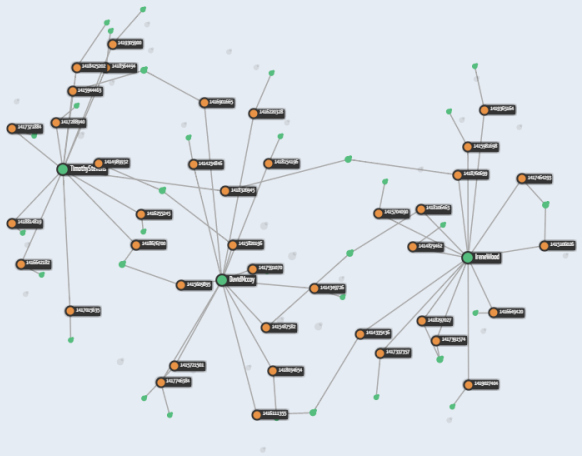
- Voronoi Diagram
  - Closest point regions to each tower
- Mapping regions to towers
  - Area based methods
    - Pro-rate overlap
  - Centroid matching





Call Graph

- Directed or undirected graph on calls
- Measurement of call volume
- Detection of high in-degree and out-degree nodes
- Identification of social network



Phone use studies

- Given demographic information, study phone use behavior
  - Call volume, call timings (time of day, date), number of contacts
- Studies of Sim Card Churn
- Inference of demographics from behavior

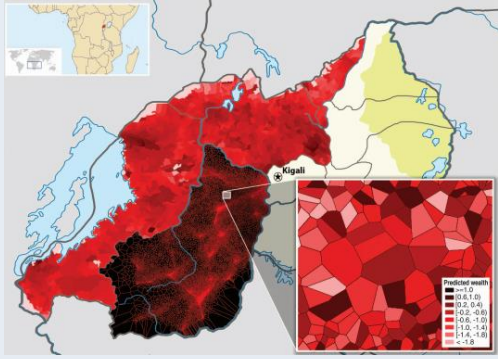
Mehrotra et al. (2012), Differences in Phone Use Between Men and Women: Quantitative Evidence from Rwanda.

Migration Studies

- Movement of people is an area of significant study
- Lack of census data make this hard to study
- Short term migration
  - What are the patterns
  - Is it possible to distinguish between shorter term and permanent migration
- Forced migration and droughts
  - Match to climate data
  - Question on local versus long distance migration
- Technical issues in definitions of movement

Blumenstock et al. (2015), Predicting poverty and wealth from mobile phone metadata

Economic Studies



- Predict economic status at a local (e.g. District) level
- Household surveys are expensive. Idea is to use Cell Phone Data to expand surveys
- Correlate household surveys with CDR
  - Compute wide range of properties of CDR
  - Construct machine learning model to predict household assets (from survey data)
  - Apply to all call records in the data set

Epidemiology

- Correlating human movement data and disease frequency
- Substantial work on Malaria and Call Data Records
- Key use case is malaria elimination
  - Understanding if cases are local infections or from other regions
  - Understand movement from high incidence to low incidence areas
- Technical modelling work that combines migration and economic studies

Event studies

- Look at impact of events in data sets
- High volume of calls related to disasters, elections, holidays
- Spike in call volumes has been observed associated with earth quakes
- Significant interest in call data records and disaster response
  - Technical issues related to infrastructure and economic displacement