Policy-relevant evidence from mobile network big data

15 May 2015

LIRNEasia

Pro-poor. Pro-market.

IDRC CRDI

Canada

This work was carried out with the aid of a grant from the International Development Research Centre, Canada and the Department for International Development UK..

UKaid

from the British people

Big data

- An all-encompassing term for any collection of **data** sets so **large** or complex that it becomes difficult to process using traditional **data** processing applications. The challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and privacy violations. Examples:
  - 100 million Call Detail Records a day generated by Sri Lanka companies (transaction generated data (TGD) ➔ behavioral ➔ relevant to development)
  - 45 Terabytes of data from Hubble Telescope

Why big data? Why now?

- Proximate causes
  - Increased “datafication”: Very large sets of schema-less (unstructured, but processable) data now available
  - Advances in memory technology: No longer is it necessary to archive most data and work with small subset
  - Advances in software: MapReduce, Hadoop

If we want comprehensive coverage of the population, what are the sources of big data in developing economies?

- Administrative data
  - E.g., digitized medical records, insurance records, tax records
- Commercial transactions (transaction-generated data)
  - E.g., Stock exchange data, bank transactions, credit card records, supermarket transactions connected by loyalty card number
- Sensors and tracking devices
  - E.g., road and traffic sensors, climate sensors, equipment & infrastructure sensors, mobile phones communicating with base stations, satellite/ GPS devices
- Online activities/ social media
  - E.g., online search activity, online page views, blogs/ FB/ twitter posts

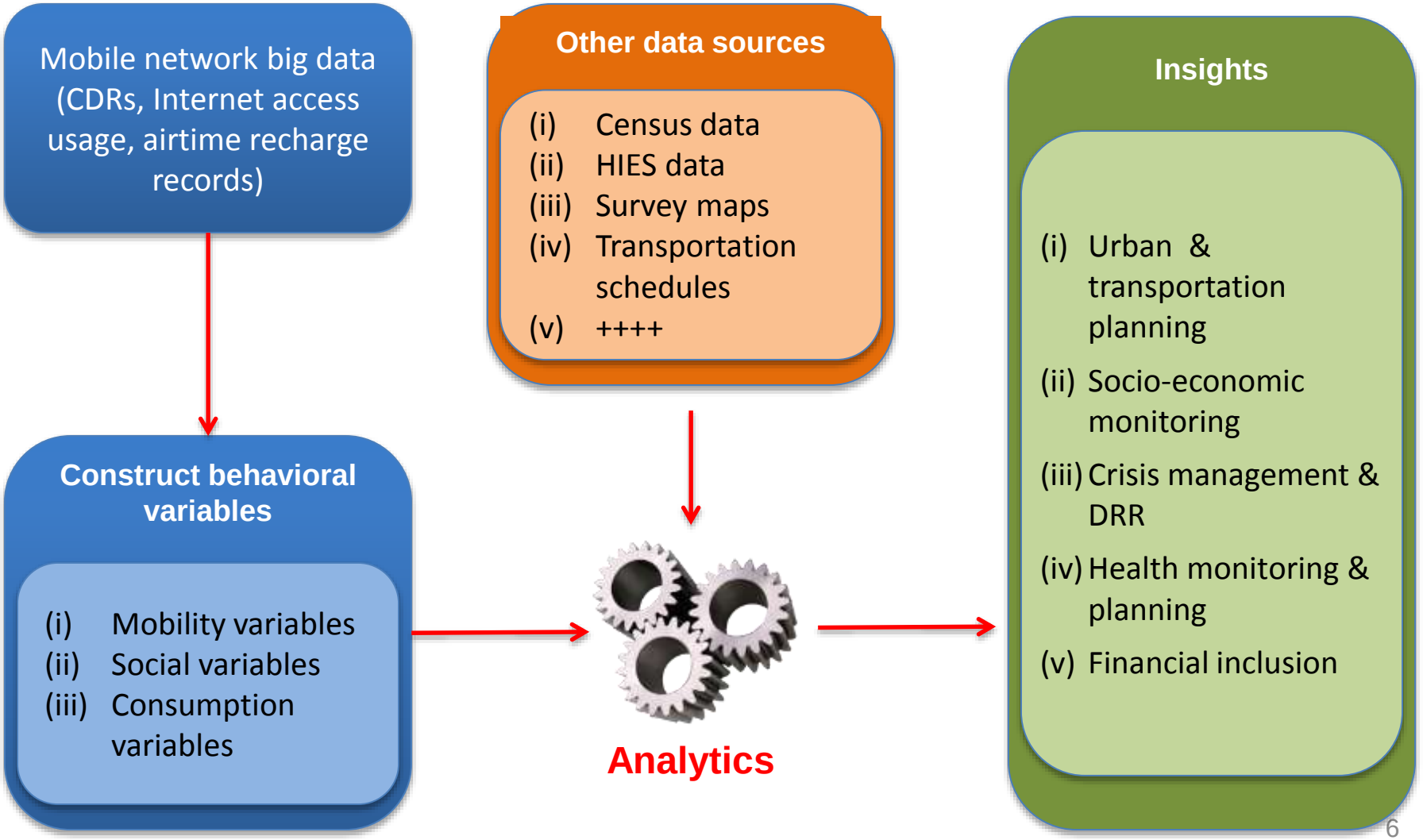
Currently only mobile network big data has broad population coverage

	Mobile SIMs/100	Internet users/100	Facebook users/100
Myanmar	13	1	4
Bangladesh	67	7	6
Pakistan	70	11	8
India	71	15	9
Sri Lanka	96	22	12
Philippines	105	39	41
Indonesia	122	16	29
Thailand	138	29	46

Source: ITU Measuring Information Society 2014; Facebook advantage portal

LIRNEasia
Pro-poor Pro-market

Mobile network big data + other data ➔ rich, timely insights



Data used in the research

- Multiple mobile operators in Sri Lanka have provided four different types of meta-data
  - Call Detail Records (CDRs)
    - Records of calls
    - SMS
    - Internet access
  - Airtime recharge records
- Data sets do not include any Personally Identifiable Information
  - All phone numbers are pseudonymized
  - LIRNE*asia* does not maintain any mappings of identifiers to original phone numbers
- Cover 50-60% of users; very high coverage in Western (where Colombo the capital city in located) & Northern (most affected by civil conflict) Provinces, based on correlation with census data

Contents

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Mobility changes during important events: Case study of Avurudu
  - Implications for public policy
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding consumption behavior
- Challenges
  - Analytical challenges
  - Other challenges
- Team structure
- What can MNBD do for official statistics in Sri Lanka

LIRNEasia
Pro-poor Pro-market

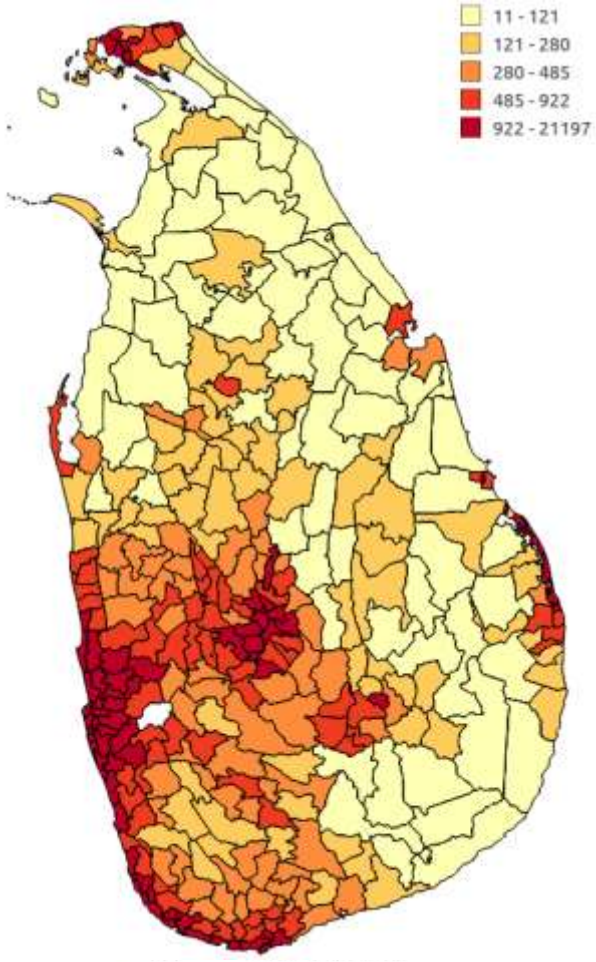
Contents

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Mobility changes during important events: Case study of Avurudu
  - Implications for public policy
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding consumption behavior
- Challenges
  - Analytical challenges
  - Other challenges
- Team structure
- What can MNBD do for official statistics in Sri Lanka

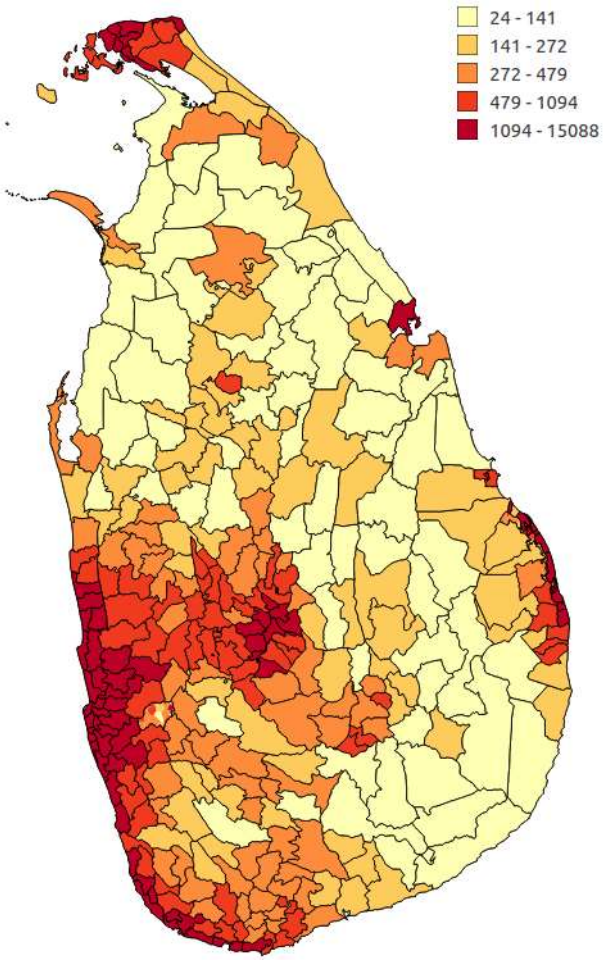
LIRNEasia
Pro-poor Pro-market

MNBD data can give us geospatially fine-grained & high-frequency estimates of population density

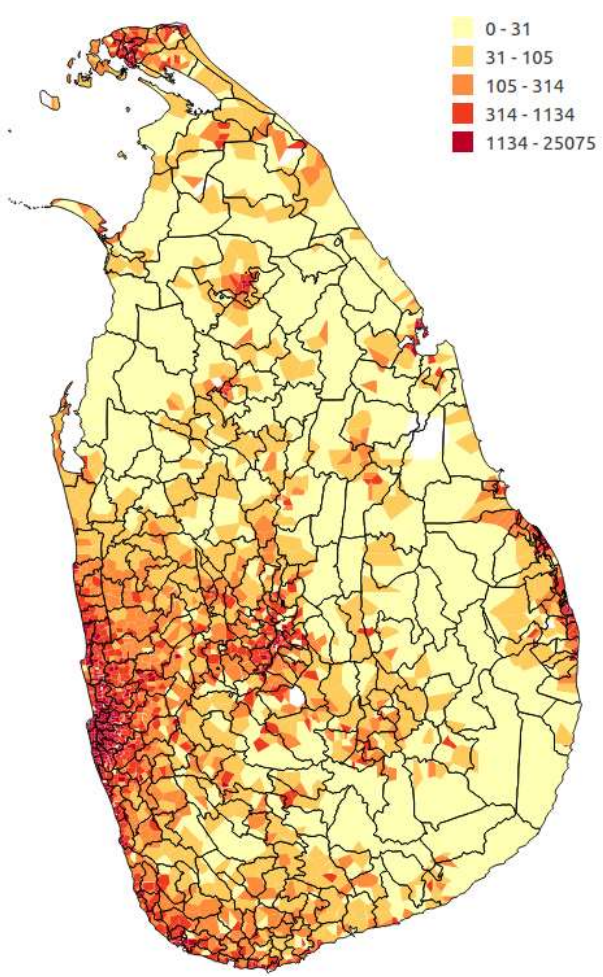
DSD population density from 2012 census



DSD population density estimate from MNBD



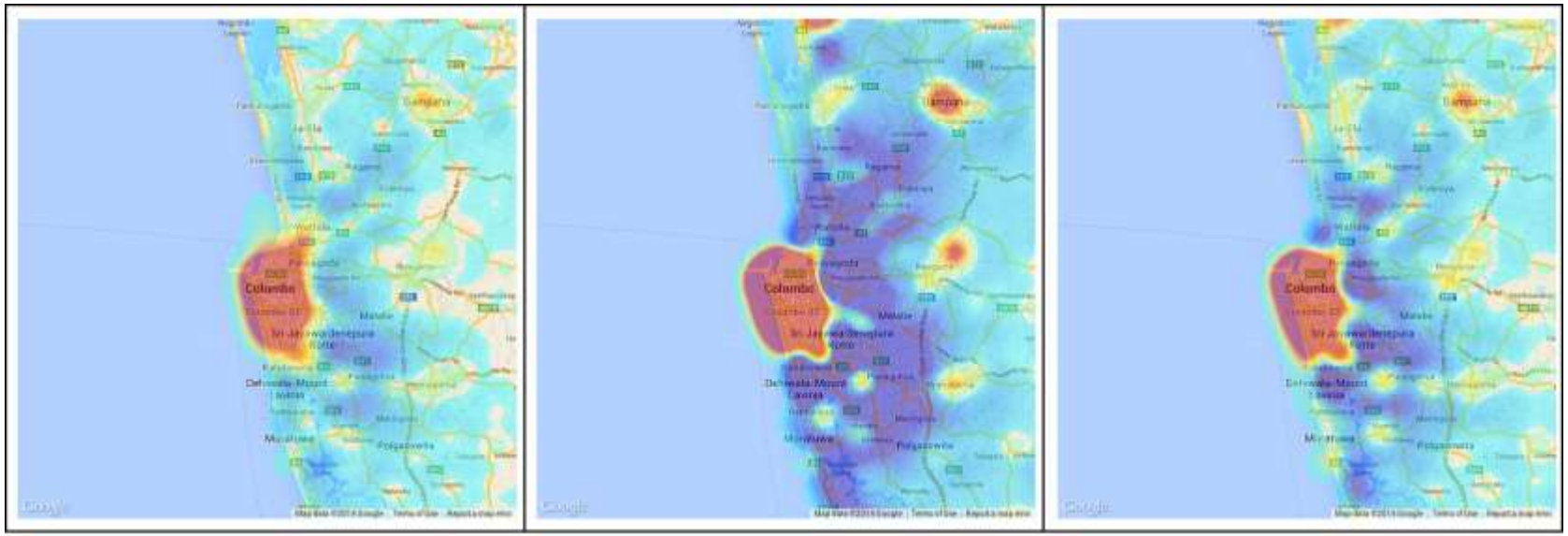
Voronoi cell population density estimate from MNBD



Population density changes in Colombo region: weekday/ weekend

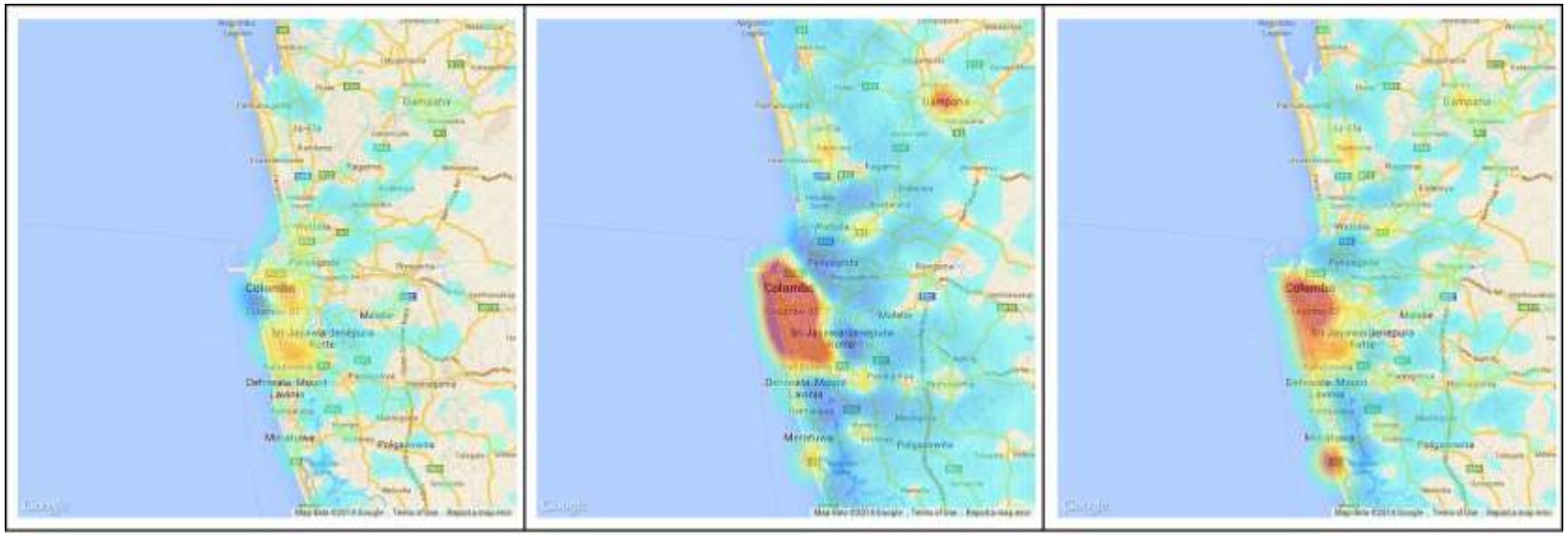
Pictures depict the change in population density at a particular time relative to midnight

Weekday



Time 06:30 Time 12:30 Time 18:30

Sunday

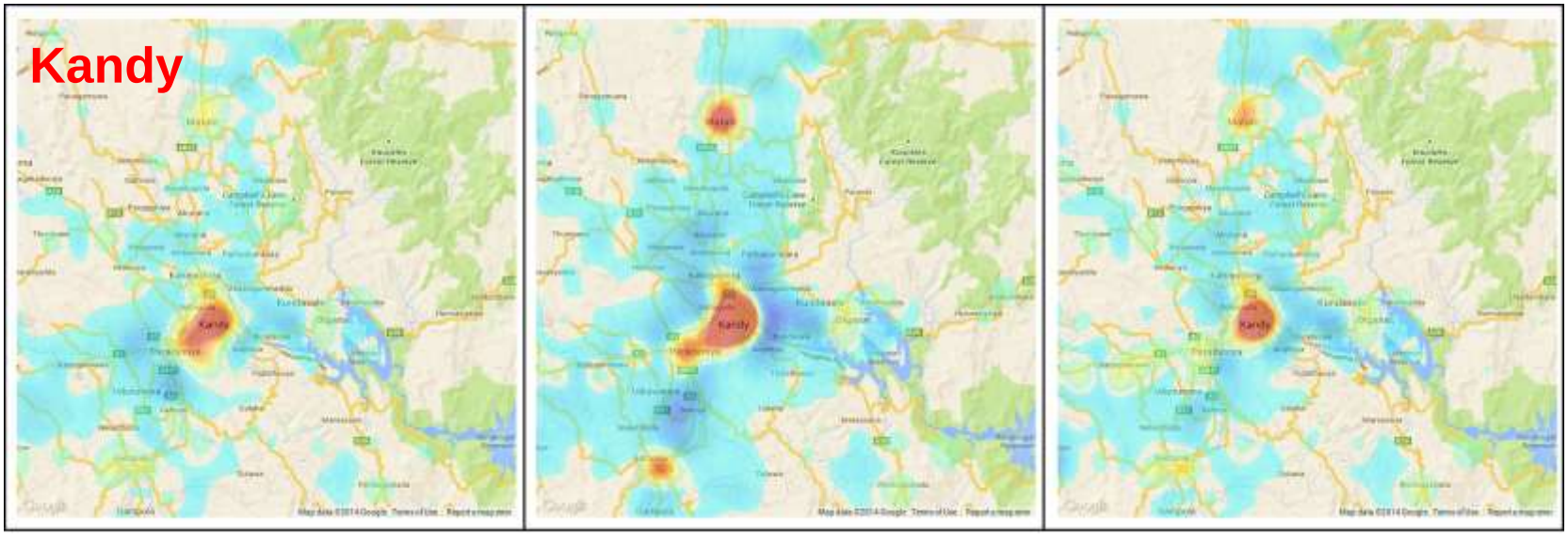


Decrease in Density Increase in Density

Population density changes in Jaffna & Kandy regions on a normal weekday

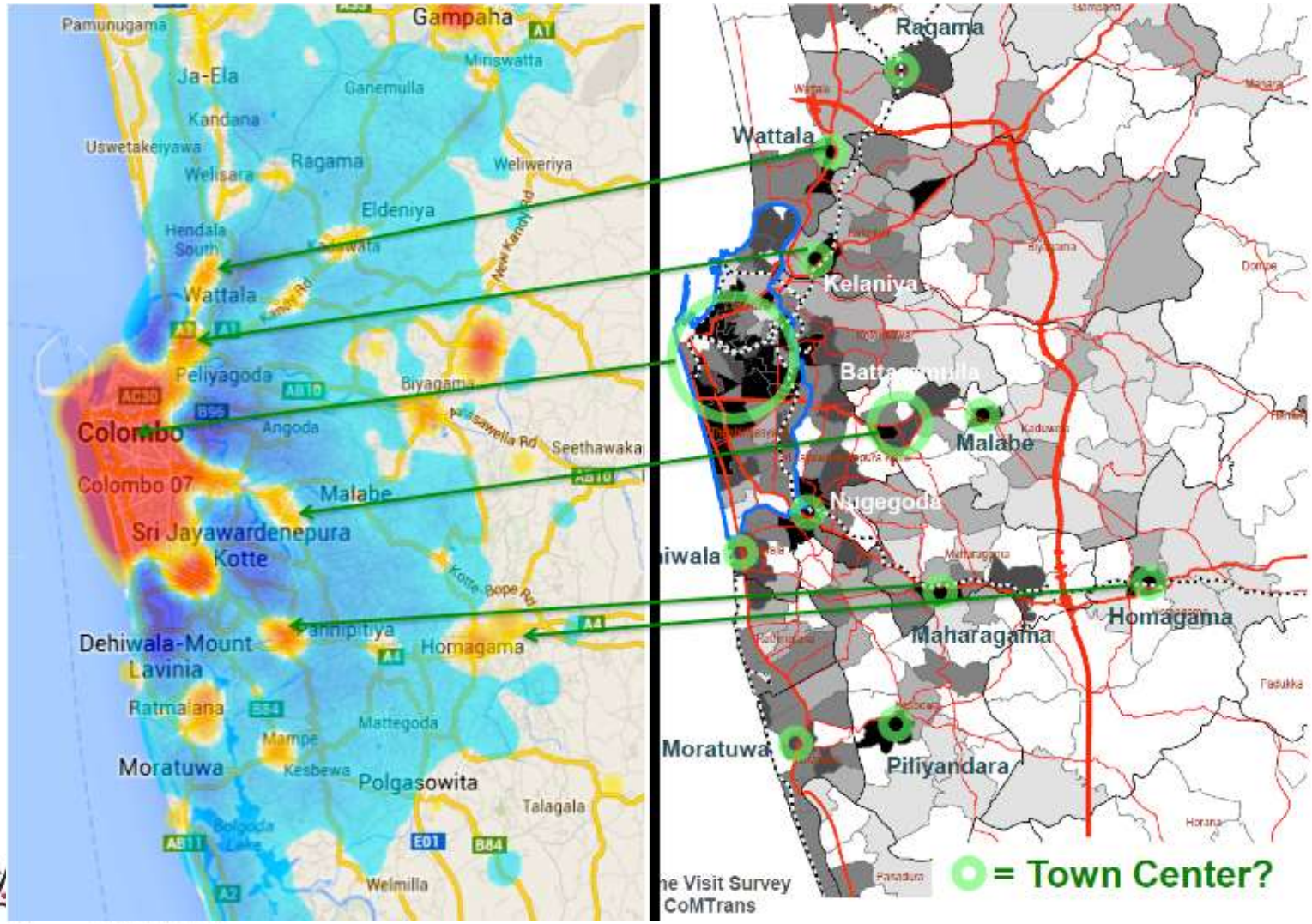
Pictures depict the change in population density at a particular time relative to midnight

Time 06:30 Time 12:30 Time 18:30



Decrease in Density Increase in Density

Our findings closely match results from expensive & infrequent transportation surveys



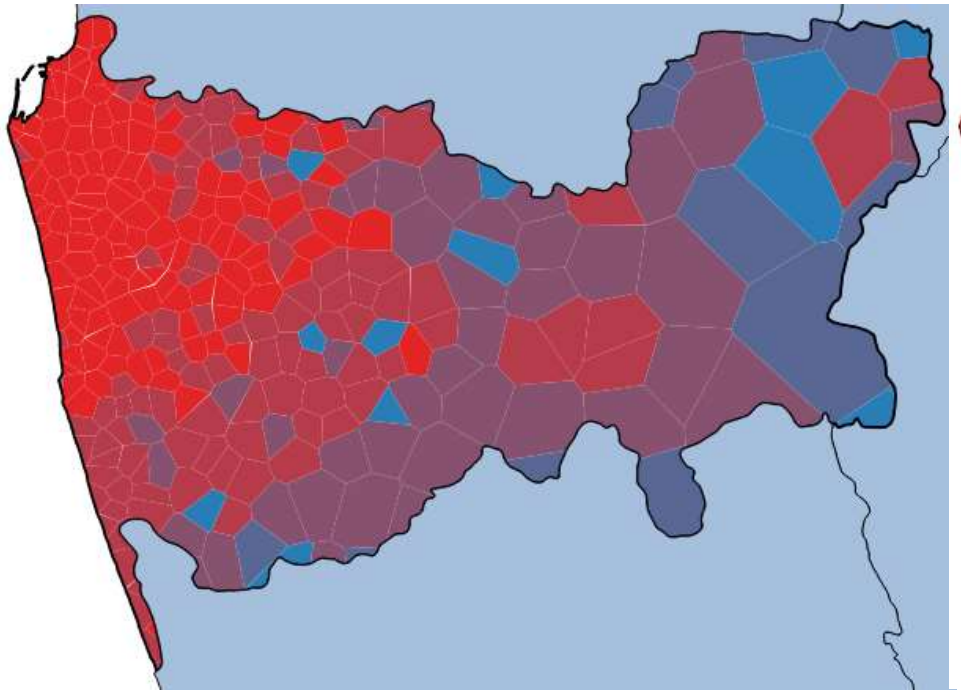
Contents

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Mobility changes during important events: Case study of Avurudu
  - Implications for public policy
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding consumption behavior
- Challenges
  - Analytical challenges
  - Other challenges
- Team structure
- What can MNBD do for official statistics in Sri Lanka

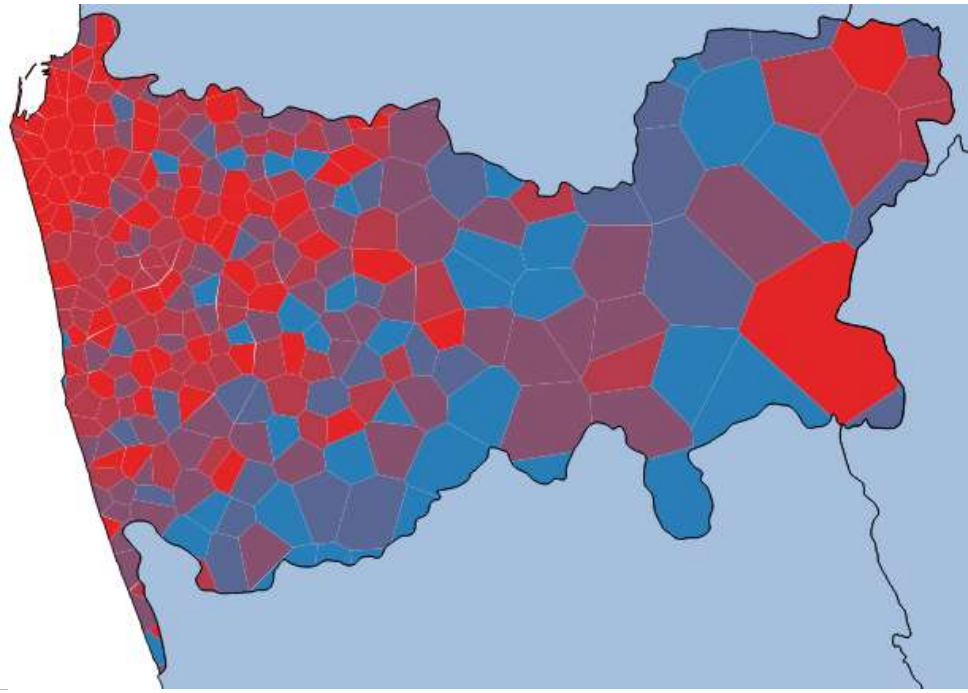
LIRNEasia
Pro-poor Pro-market

Average monthly reload by Colombo & adjacent region residents is high, as is variability

Average monthly reload amount



Co-efficient of Variation (COV)

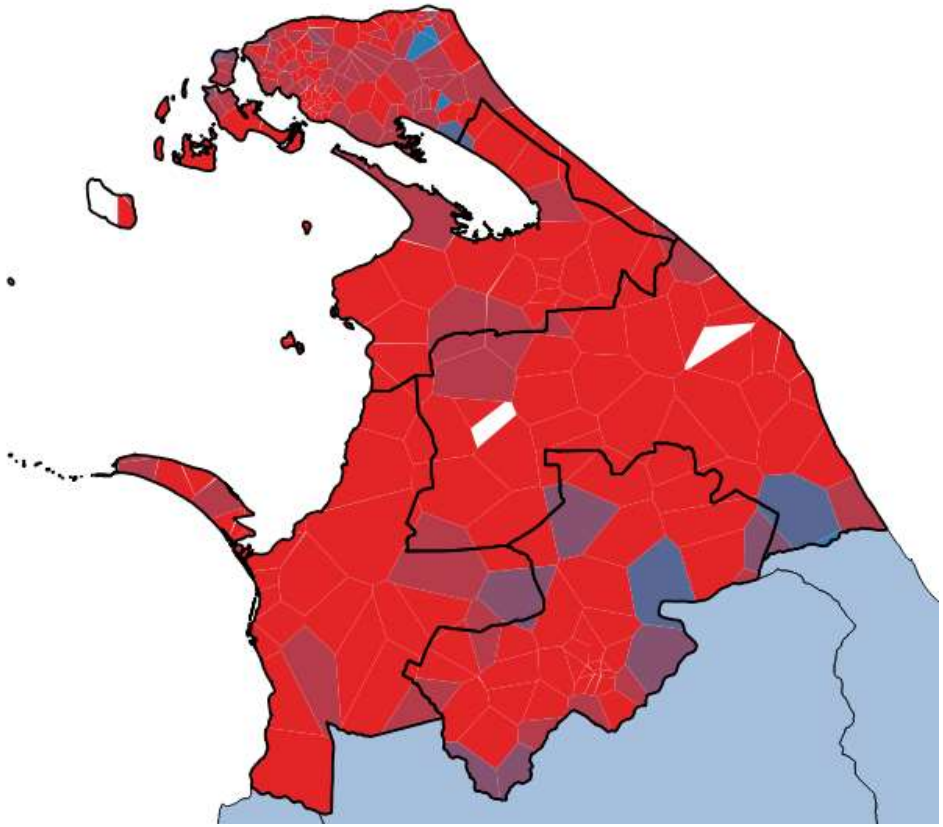


Low → **High**

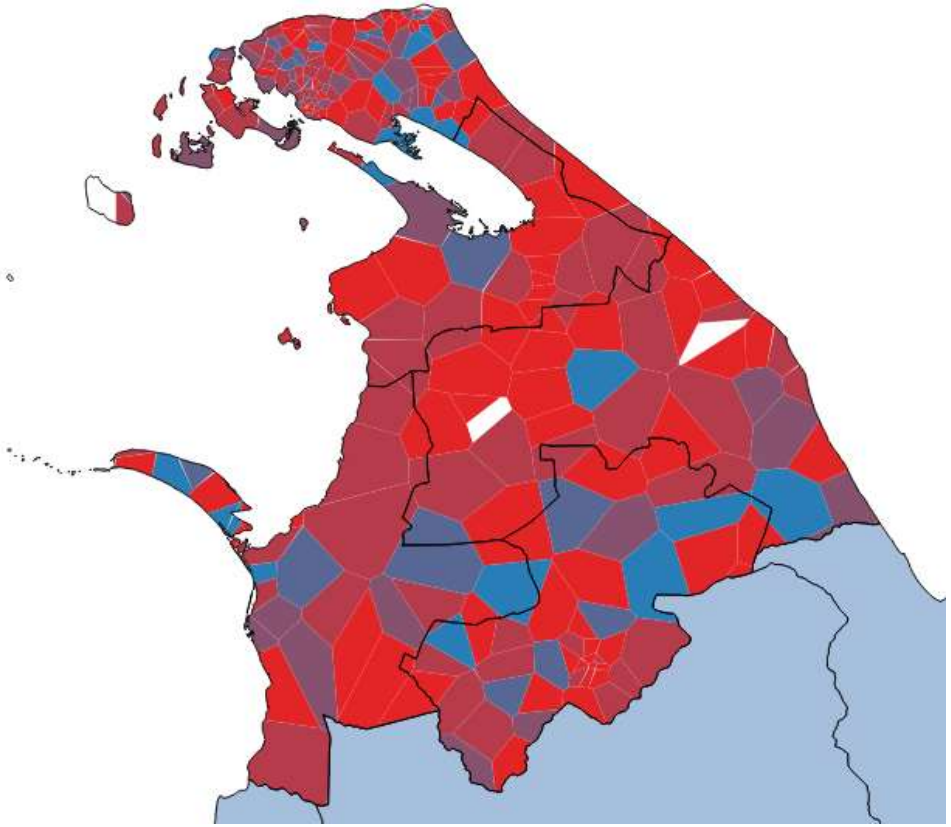
LIRNEasia
Pro-poor Pro-market

Similar story for Northern Province residents, but average monthly reload is higher than Colombo district

Average monthly reload amount



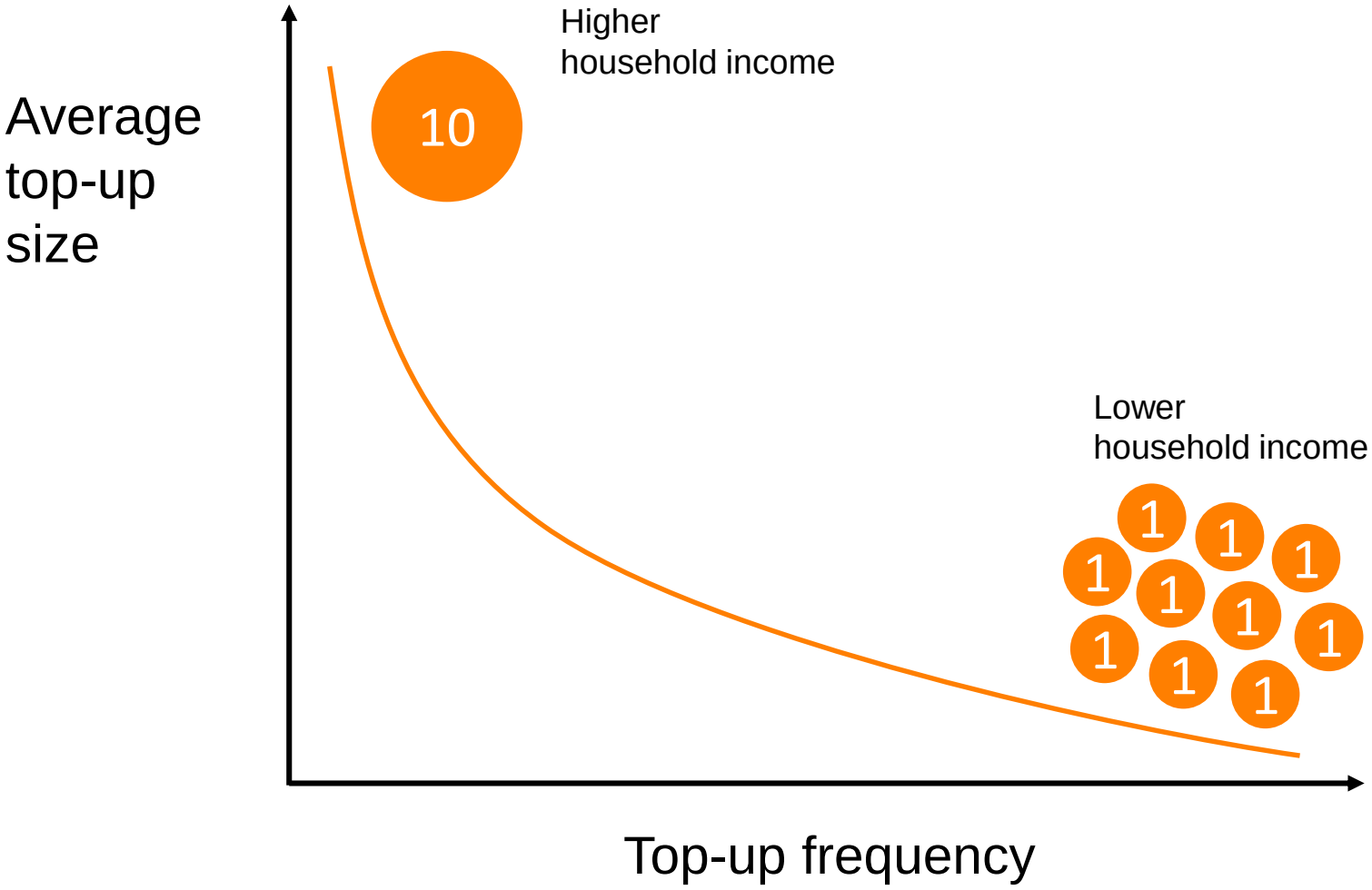
Co-efficient of Variation (COV)



LIRNEasia
Pro-poor Pro-market

Low → **High**

Research in other countries suggest differential reload behavior may be correlated to household income



LIRNEasia
Pro-poor Pro-market

Source: Gutierrez, T., Krings, G., & Blondel, V. D. (2013). Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets, 1–6. Retrieved from http://arxiv.org/abs/1309.4496

But preliminary LK results are inconclusive

- Large majority of LK mobile users reload using recharge cards and higher denomination cards are not easy to come by
- High reload spending in Northern Province meshes with
  - Findings from Department of Census and Statistics
  - LIRNEasia research on high communication expenditures
- Addition work required:
  - Ideally we would have user-level socio-economic data through phone surveys, but difficult to implement
  - Co-relate fine grained MNBD behavioral data (not just reload but also mobility and social) with census-block level data from HIES and Census
    - V. Soto, V. Frias-Martinez, J. Virseda, and E. Frias-Martinez. 2011. Prediction of socioeconomic levels using cell phone records. In Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization (UMAP'11). Springer-Verlag, Berlin, 377–388

Contents

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Mobility changes during important events: Case study of Avurudu
  - Implications for public policy
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding consumption behavior
- Challenges
  - Analytical challenges
  - Other challenges
- Team structure
- What can MNBD do for official statistics in Sri Lanka

LIRNEasia
Pro-poor Pro-market

So what can mobile network big data do for official statistics in Sri Lanka?

Feature of MNBD	Benefit
High frequency	• Can complement infrequent surveys and census • "New" findings (e.g. hourly population density)
Near real time	• Create near real-time indicators (e.g., economic activity, socio-economic levels) at fine geographic and temporal resolution • Respond to real-time event (e.g., Avurudu)
(Almost) universal coverage	• Allow extrapolation from sample surveys (e.g. HIES, LFS) • Capture informal (economic) activities
Cheaper	• Marginal cost is minimal • Can reduce expenditures on some current surveys (e.g., transport surveys)

LIRNEasia
Pro-poor Pro-market