Big Data for Development in Sri Lanka

LIRNEasia Big Data for Development Team

http://lirneasia.net/projects/bd4d/

University of Sri Jayewardenepura

18 October 2016

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..

LIRNEasia is a regional non-profit think tank. Our mission is that of

Catalyzing policy change through research to improve people's lives in the emerging Asia Pacific by facilitating their use of hard and soft infrastructures through the use of knowledge, information and technology.

LIRNEasia
Pro-poor. Pro-market.

Where we work

Big data work only in Sri Lanka in 2012-16
Is being extended to Bangladesh in 2016-17

LIRNEasia
Pro-poor. Pro-market.

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.
- Challenges include: analysis, capture, curation, search, sharing, storage, transfer, visualization, and privacy violations.
- Examples:
  - 100 million Call Detail Records per day generated by Sri Lanka mobile operators
  - 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

LIRNEasia
Pro-poor. Pro-market.

Mobile Network Big Data is only option at this time

Country	Mobile SIMs/100	Internet users/100	Facebook users/100
	2015	2014	2016
Myanmar	69	2	20.4
Bangladesh	82	10	13.7
Pakistan	67	14	14.3
India	76	18	11.5
Sri Lanka	125	26	19.6
Philippines	120	40	54.6
Indonesia	131	17	33.8
Thailand	125	35	58.9

LIRNEasia Pro-poor. Pro-market.

Sources: https://www.gsmaintelligence.com/; http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf; Facebook advertising portal; http://data.worldbank.org/indicator/SP.POP.TOTL

Mobile network big data + other data ➔ rich, timely insights that serve private **as well** as public purposes

Mobile network big data (CDRs, Internet access usage, airtime recharge records)

Other Data Sources

1. Data from Dept. of Census & Statistics
2. Transportation data
3. Health data
4. Financial data
5. Etc.

Construct Behavioral Variables

1. Mobility variables
2. Social variables
3. Consumption variables

Dual purpose insights

Private purposes

1. Mobility & location based services
2. Financial services
3. Richer customer profiles
4. Targeted marketing
5. New VAS

Public purposes

1. Transportation & Urban planning
2. Crises response + DRR
3. Health services
4. Poverty mapping
5. Financial inclusion

Big data used in the research

- Multiple mobile operators in Sri Lanka provided four different types of meta-data
  - Call Detail Records (CDRs): Records of calls, SMS, Internet access
  - Airtime recharge records
  - No Visitor Location Registry (VLR) data, because they are written over & not stored
- 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
- Historical, not real time; therefore analyzed in batch mode in a hardware stack costing < USD 30k
- 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
- Now also using CCTV footage as well as satellite imagery

The technology used

- We built our own internal Apache Hadoop cluster:
  - 2 Master Nodes & 8 Slave Nodes
  - Total of 30 TB disk space with a replication factor of 2
- Distributed processing frameworks:
  - Apache Pig
  - Apache Hive
  - Apache Spark
  - Apache Giraph
  - Apache Hadoop Yarn
- Tools & libraries:
  - Java & Python
  - R
  - Processing
  - QGIS

LIRNEasia
Pro-poor. Pro-market.

The rest of the presentation

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Predicting spatial spread of dengue in Sri Lanka
  - Implications for public policy
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

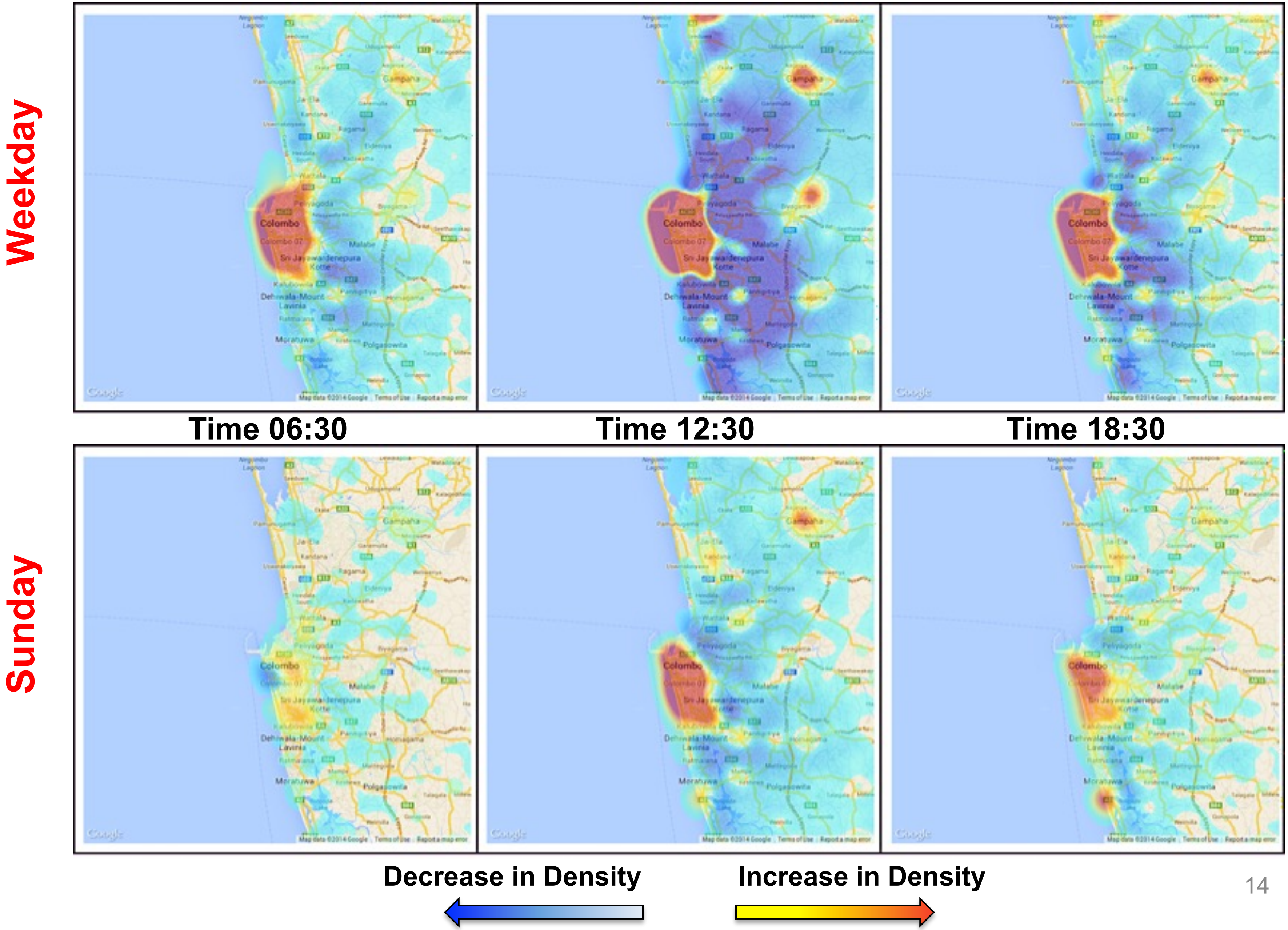
- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

LIRNEasia
Pro-poor. Pro-market.

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

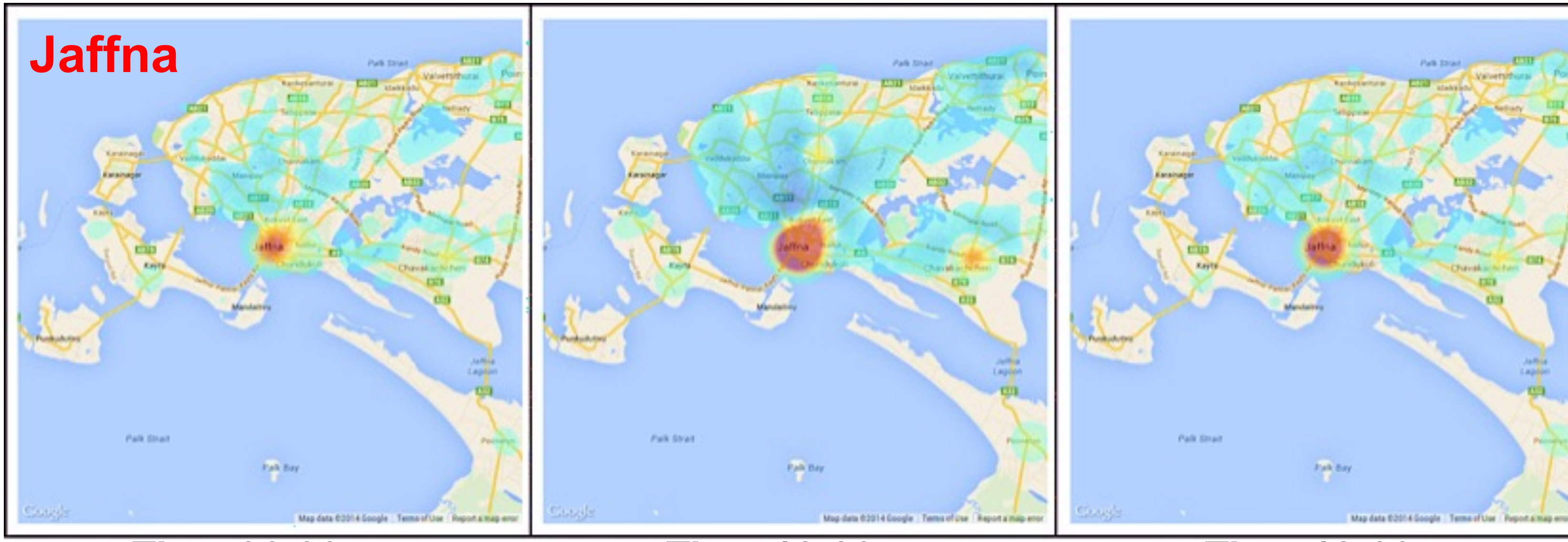
Population density changes in Colombo region: weekday/ weekend

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

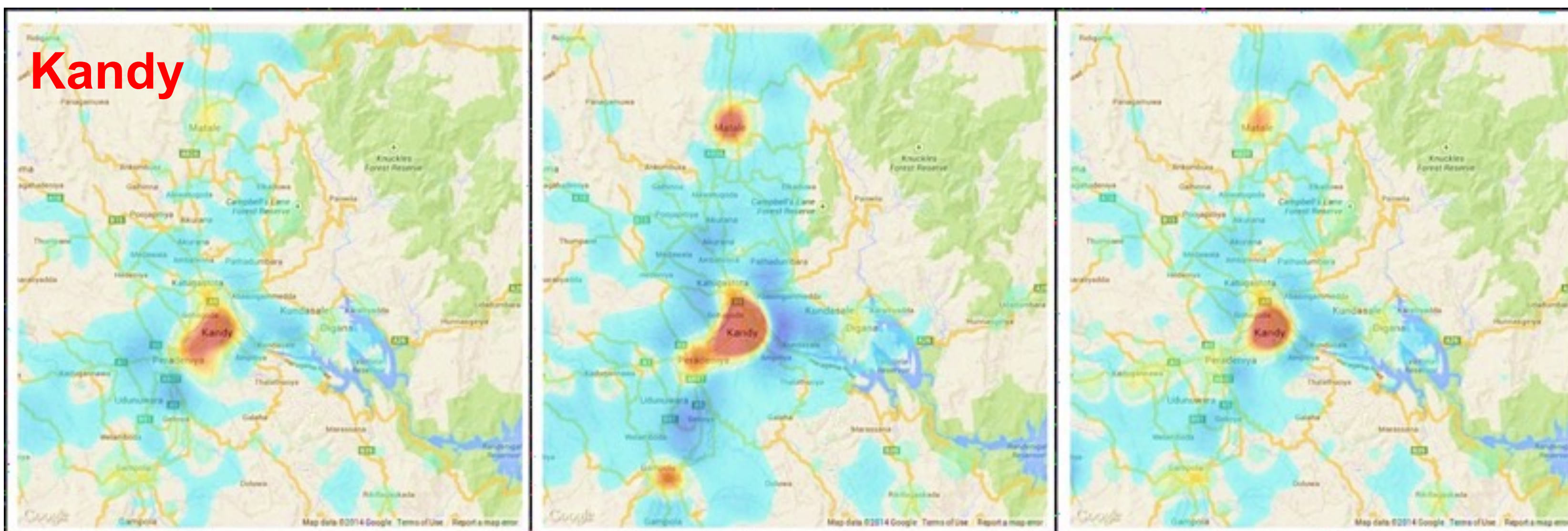


Population density changes in Jaffna & Kandy regions on a normal weekday

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

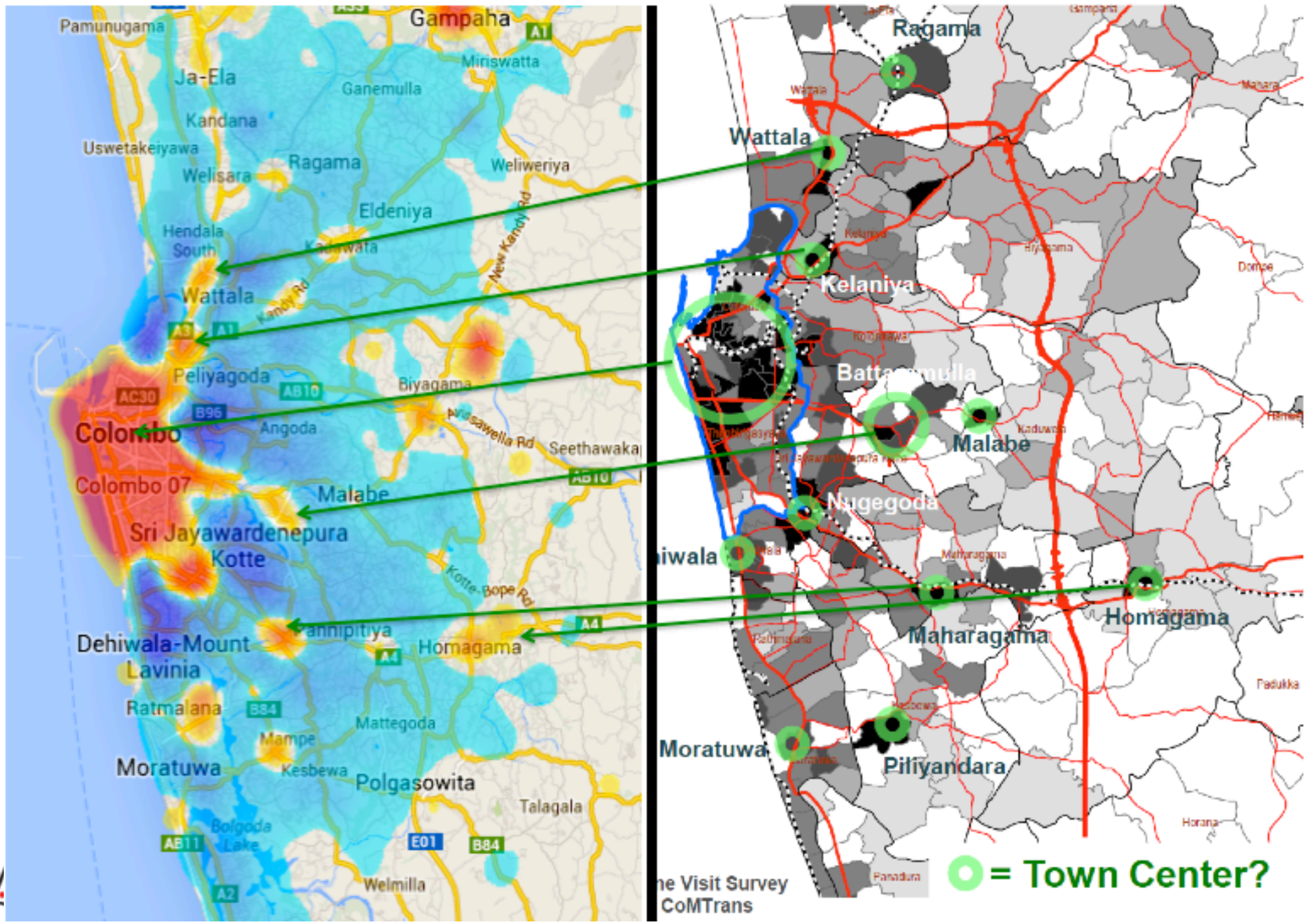


Time 06:30 | **Time 12:30** | **Time 18:30**



Decrease in Density | **Increase in Density**

Our findings closely match results from expensive & infrequent transportation surveys



Estimating population at fine spatial levels

- We first estimate the mobile phone user density in a DSD based on home locations identified for users:

$$\sigma_{c_i} = \frac{1}{A_{c_i}} \sum_{v_j} \sigma_{v_j} A_{(c_i \cap v_j)}$$

c_i	i^{th} DSD
A_c	Area of DSD
σ_{vj}	Mobile phone user density in j^{th} voronoi
$A_{(ci \cap vj)}$	Area of intersection between c_i and v_j

- We modeled the relationship between MP user density in a DS ($\boldsymbol{\sigma}_c$) with actual population density (ρ_c) as follows:

$$\rho_c = \alpha \sigma_c^{\beta}$$

- β is an adjustment for the variation in mobile phone ownership between urban and rural regions w.r.t population
- We transform the model to the following format and carry out population weighted linear regression:

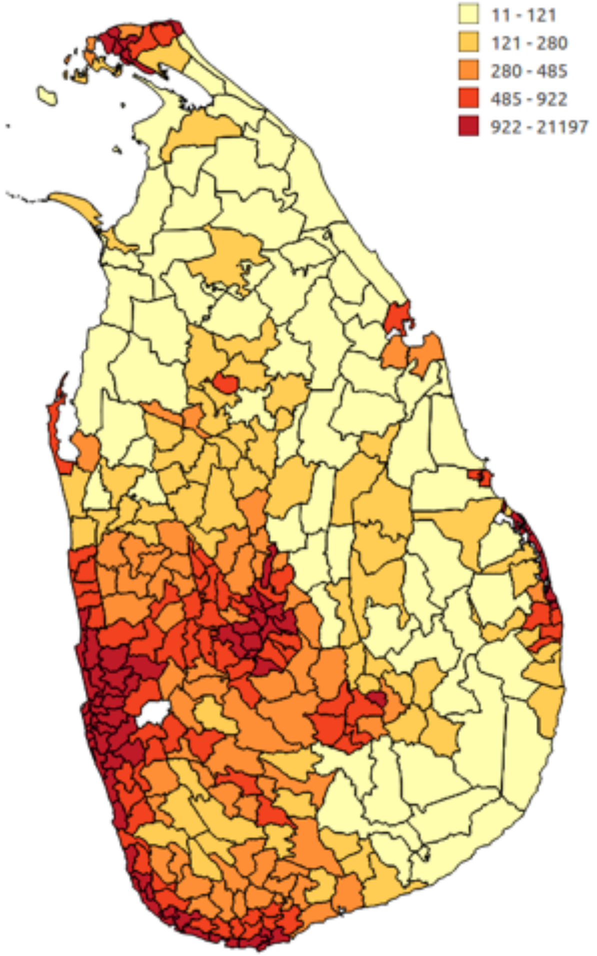
$$\log(\rho_c) = \log(\alpha) + \beta \log(\sigma_c)$$

R^2 - 0.93 , α - 8.6 , β - 0.805

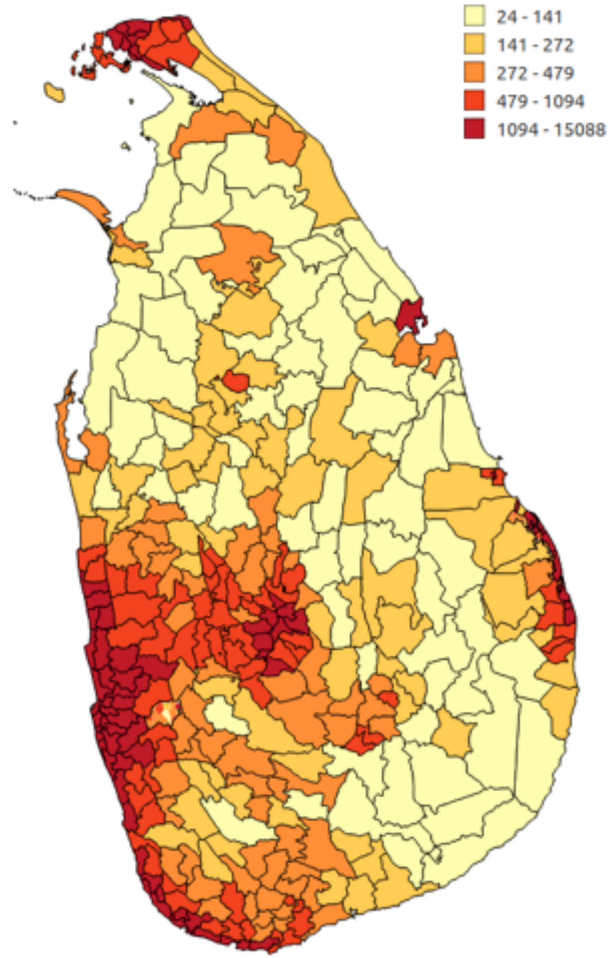
LIRNEasia Pro-poor. Pro-market.

MNBD data can give us granular & high-frequency estimates of population density

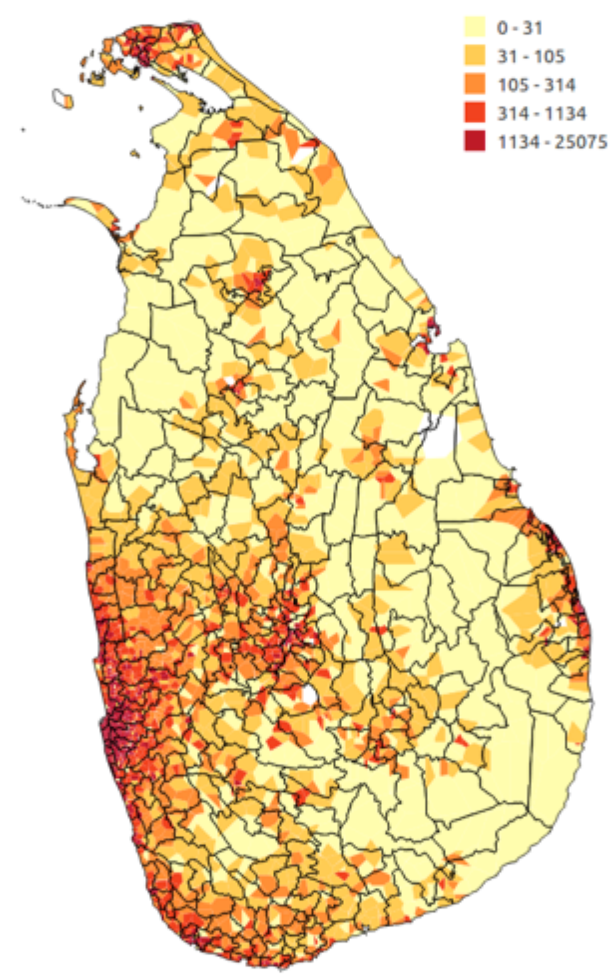
DSD population density from 2012 census



DSD population density estimate from MNBD



Voronoi cell population density estimate from MNBD



- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

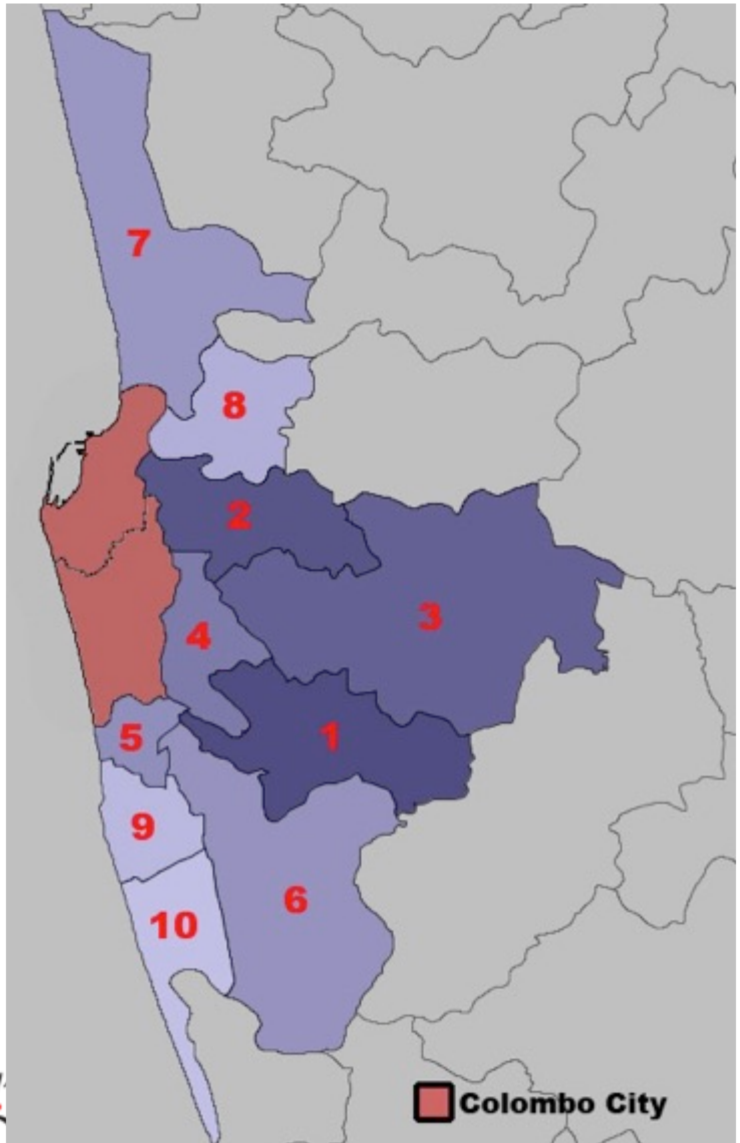
LIRNEasia Pro-poor. Pro-market.

Methodology

- Based on extracted average diurnal mobility pattern for population, choose time frames for home and work
  - Home time: 21:00 to 05:00
  - Work time: 10:00 to 15:00
- Calculate a home and work location for each user:
  - Match cell towers to Divisional Secretariat Division (DSD)
  - Count each DSD at most once per *day*.
  - Pick the DSD with the largest number of "hits"
  - For work consider only weekdays that are not public holidays

LIRNEasia
Pro-poor. Pro-market.

46.9% of **Colombo City's** daytime population comes from the surrounding regions



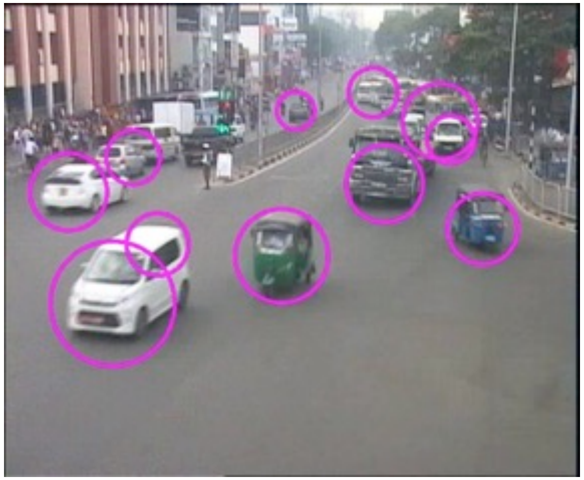
Colombo city is made up of Colombo and Thimbirigasyaya DSDs

Home DSD	%age of Colombo's daytime population
Colombo city	53.1
1. Maharagama	3.7
2. Kolonnawa	3.5
3. Kaduwela	3.3
4. Sri Jayawardanapura Kotte	2.9
5. Dehiwala	2.6
6. Kesbewa	2.5
7. Wattala	2.5
8. Kelaniya	2.1
9. Ratmalana	2.0
10. Moratuwa	1.8

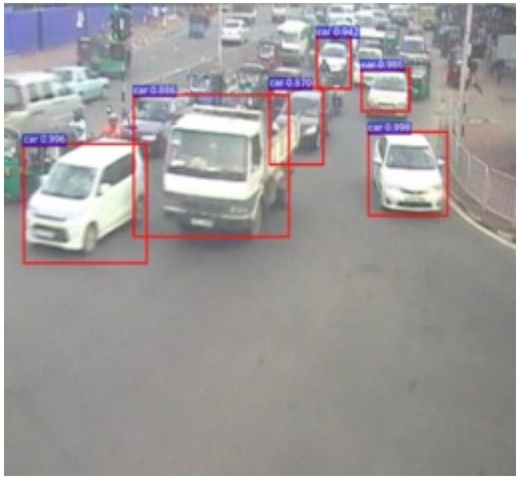
- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

Understanding traffic conditions

- CDR data isn't good enough to understand traffic conditions
- CCTV footage gives us a better chance of understanding traffic flow (volume, speeds)



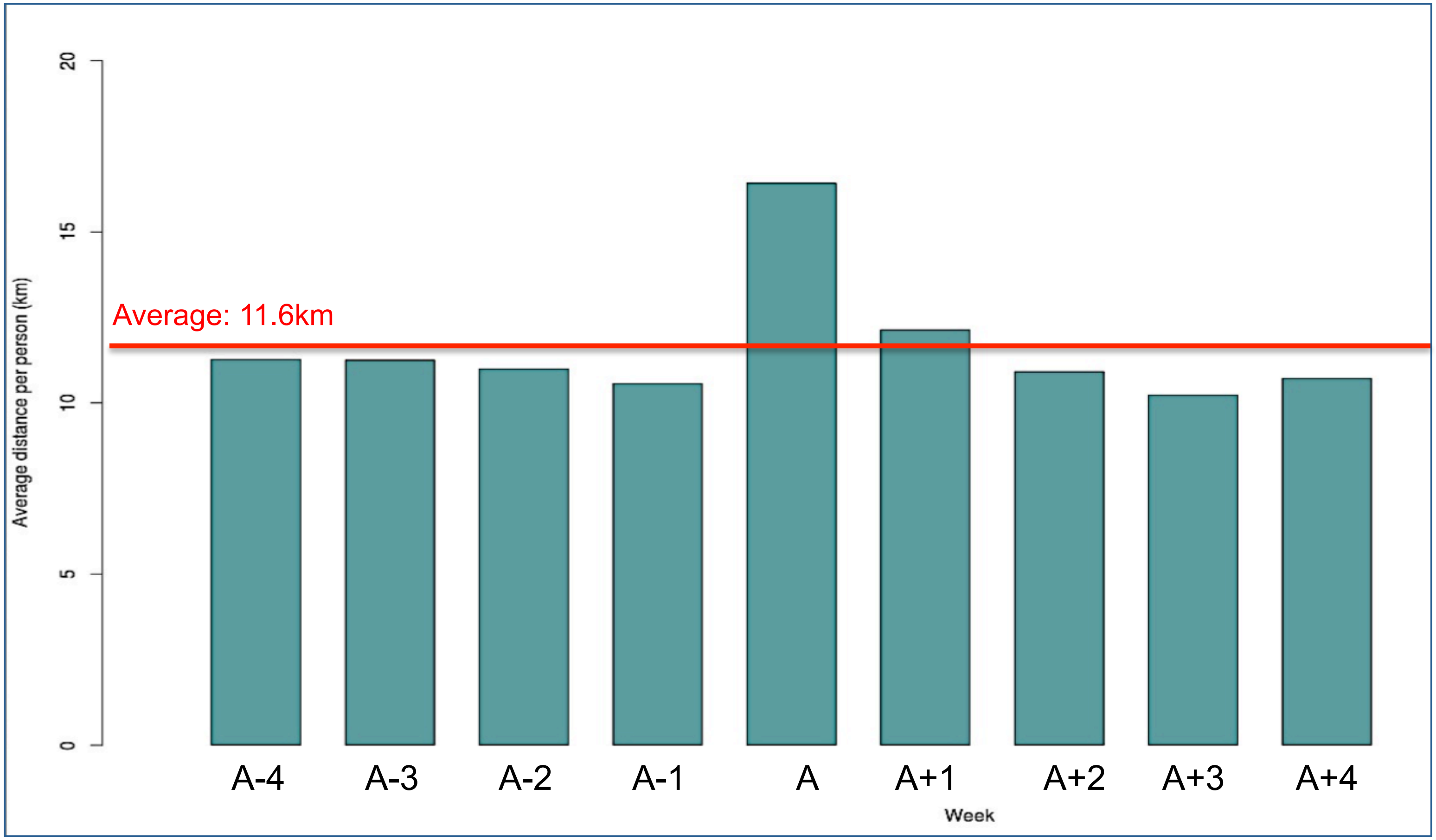
Haar-feature classification



Deep learning based classifier

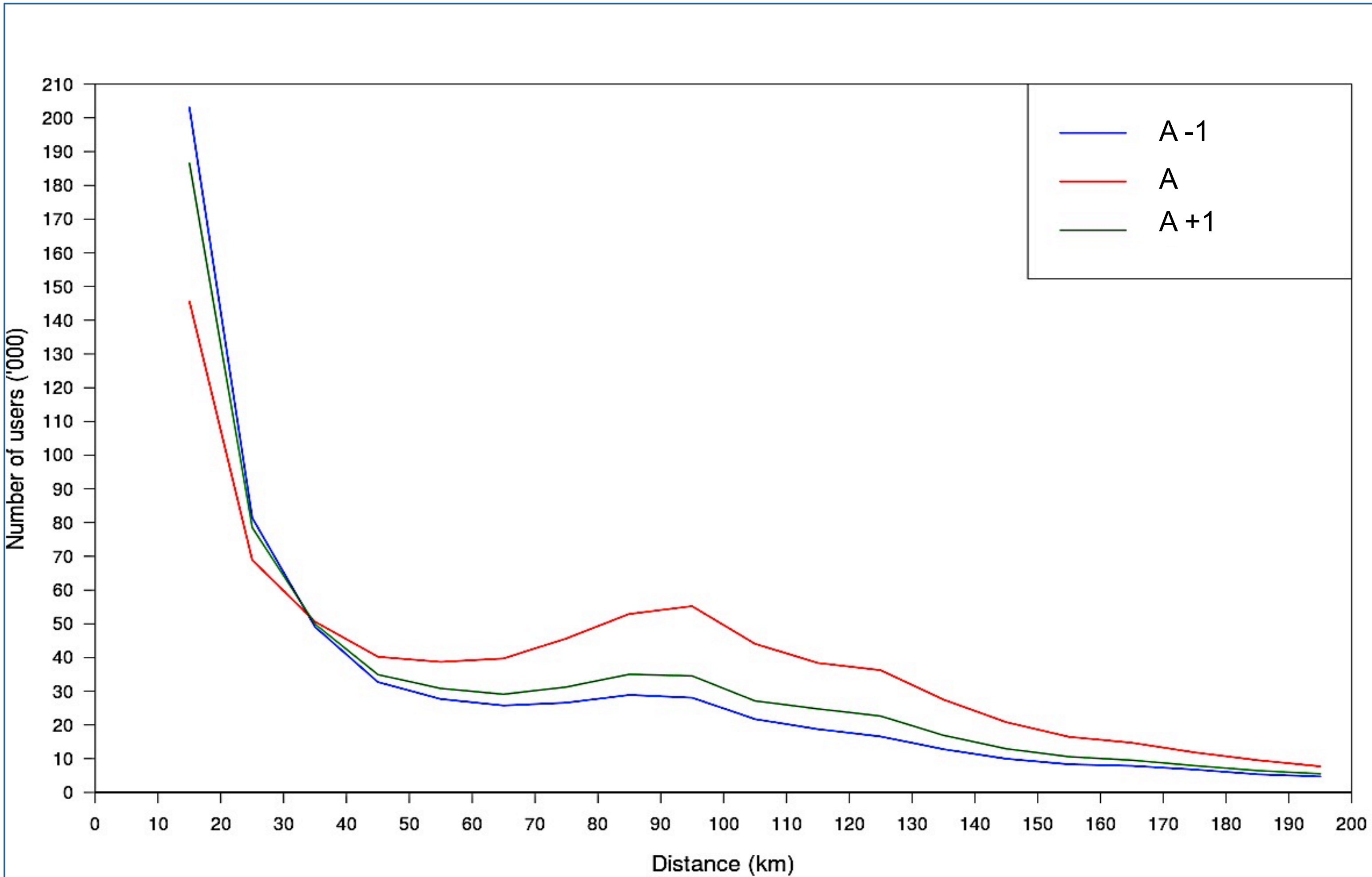
- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

People travel greater distances during Avurudu

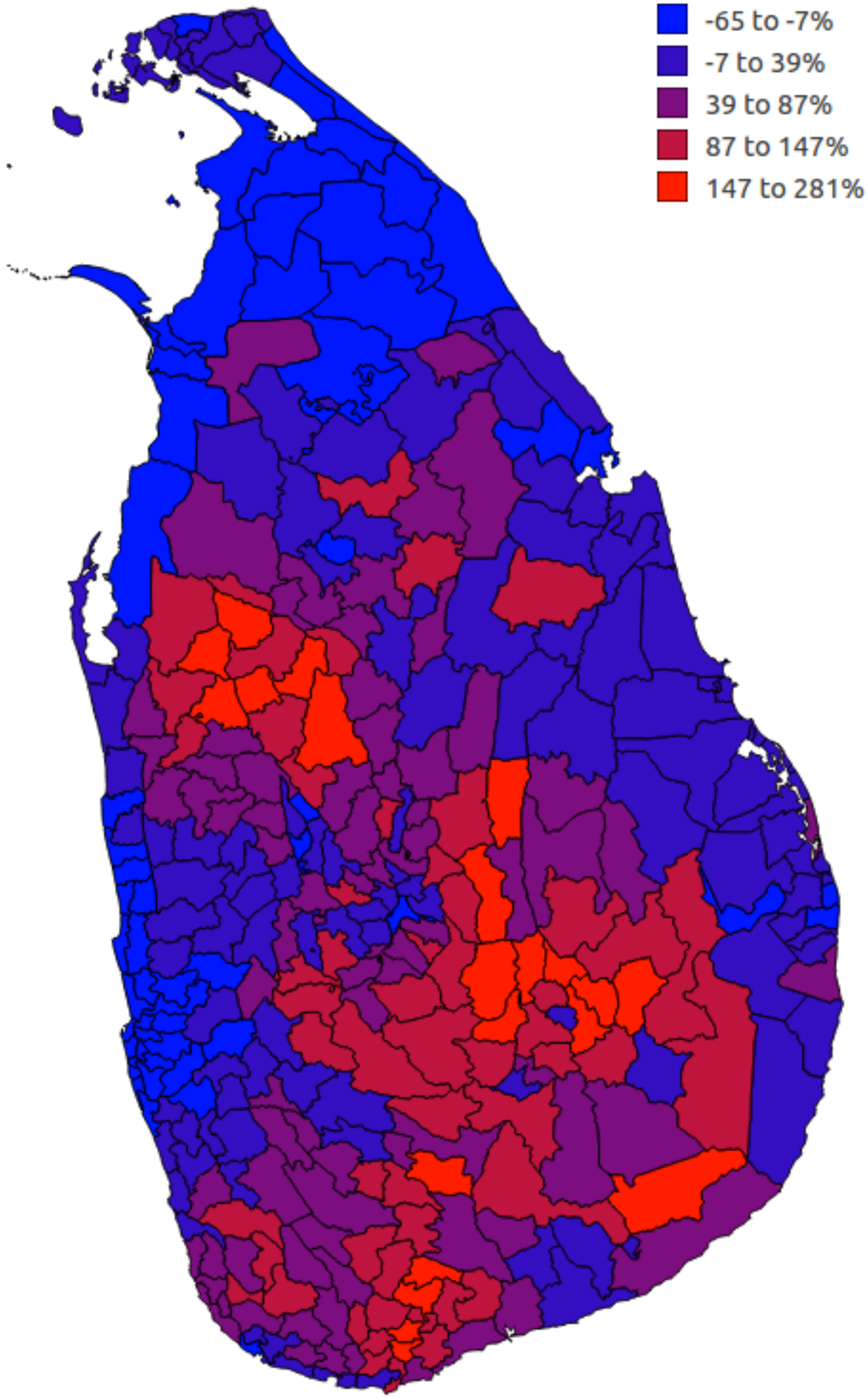


LIRNEasia
Pro-poor. Pro-market.

More people travel greater distances during Avurudu



LIRNEasia Pro-poor. Pro-market.

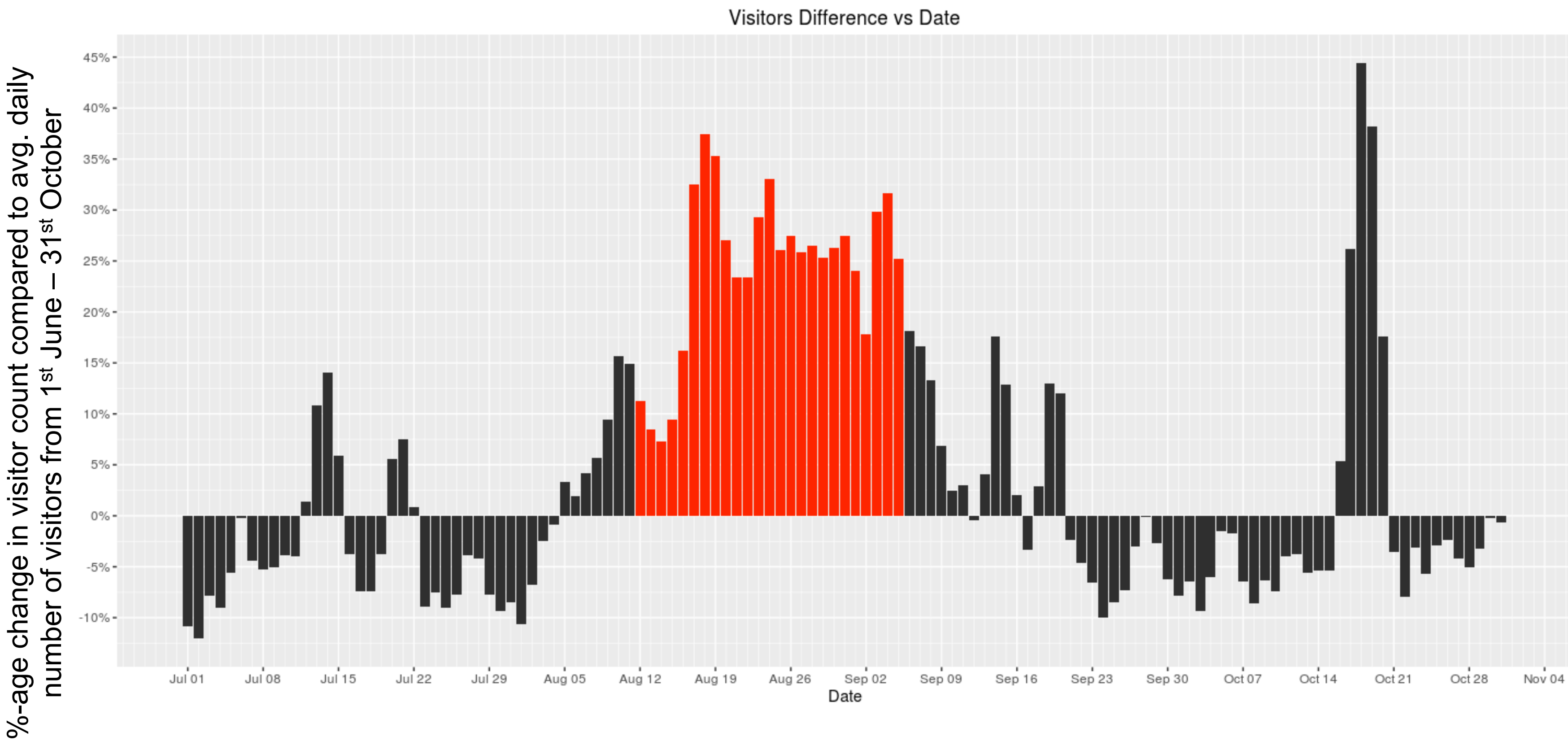


Net inflow during Avurudu weekend

With the exception of Ampara, more Colombo city residents travelled to other DSDs during Avurudu, as compared to other weekends (some examples below):

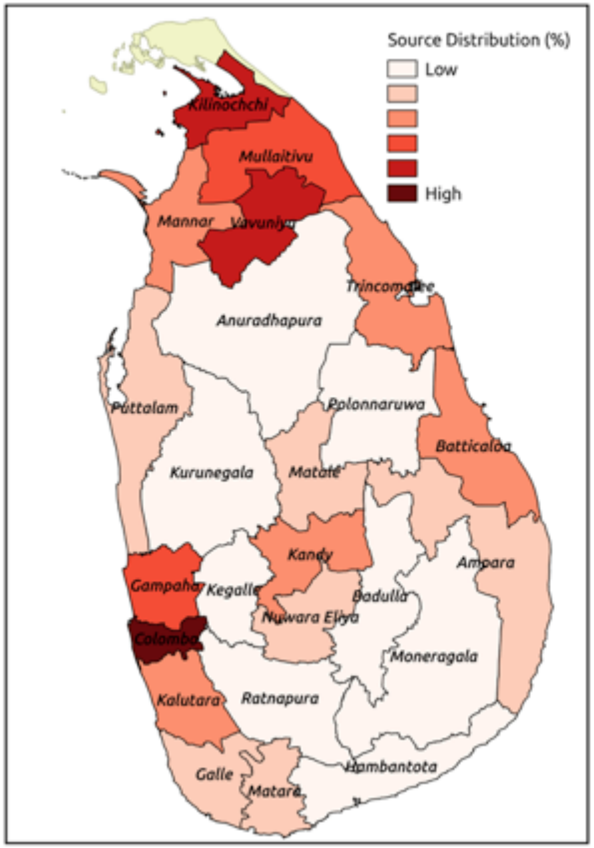
- Nuwara Eliya: 315%
- Kotmale: 233%
- Town & Gravets (Trincomalee): 100%
- Udunuwara: 93%
- Kalutara: 90%
- Jaffna: 80%
- Galle Four Gravets: 77%
- Gangawata Korale: 71%
- Attanagalla: 71%
- Mirigama: 66%

Nallur festival in Jaffna



LIRNEasia
Pro-poor. Pro-market.

Visitors during Nallur festival: where they came from

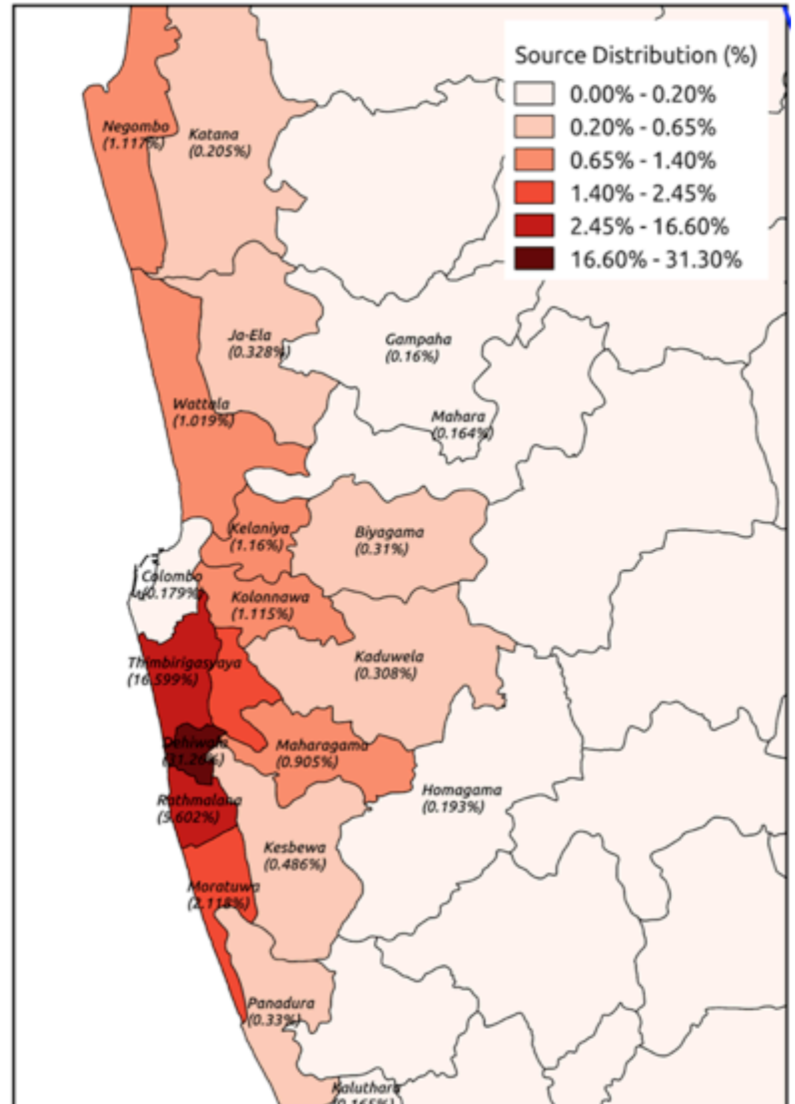


	Kilinochchi	Colombo	Vavuniya	Mullaitivu
%-age of visitors	23.53	22.99	11.83	7.57

LIRNEasia
Pro-poor. Pro-market.

Going deeper: where they came from

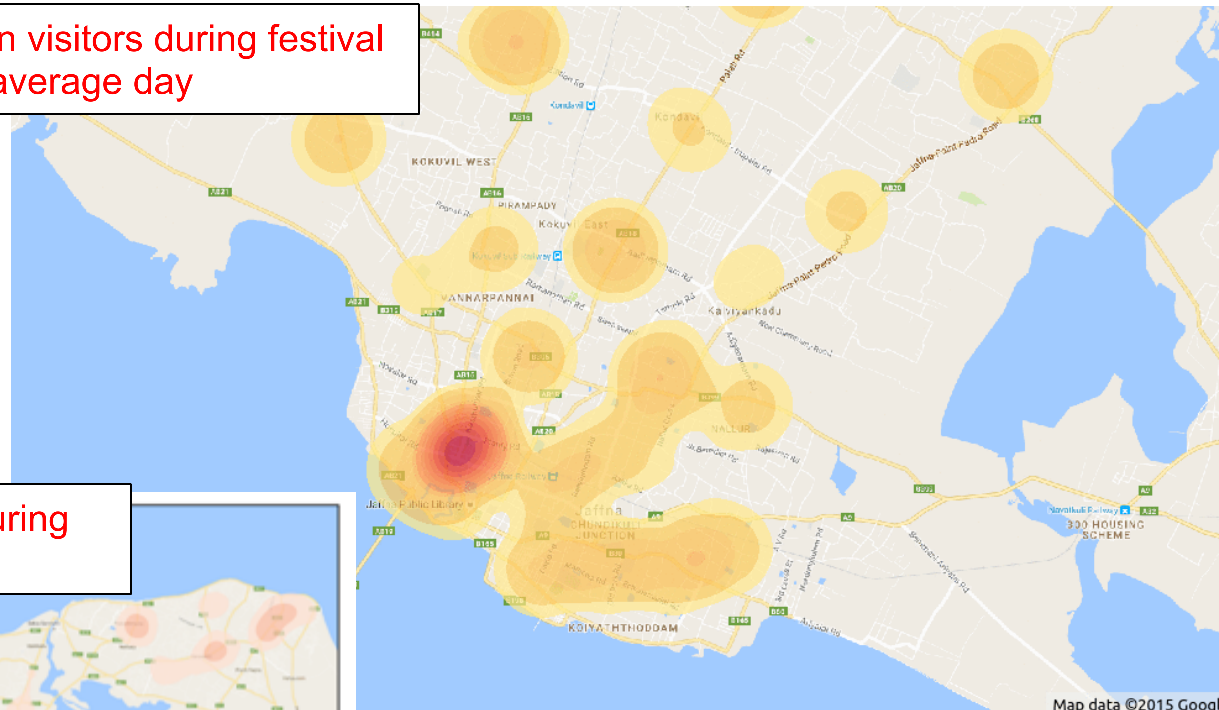
DSD	Average # of visitors (%) b/w 1st June – 31st Oct	Increase in visitors (%)
Thimbirigasyaya	6.88%	53.78%
Colombo	3.08%	53.22%
Dehiwala	4.25%	50.71%
Rathmalana	1.24%	43.65%
Mannar	2.38%	33.14%



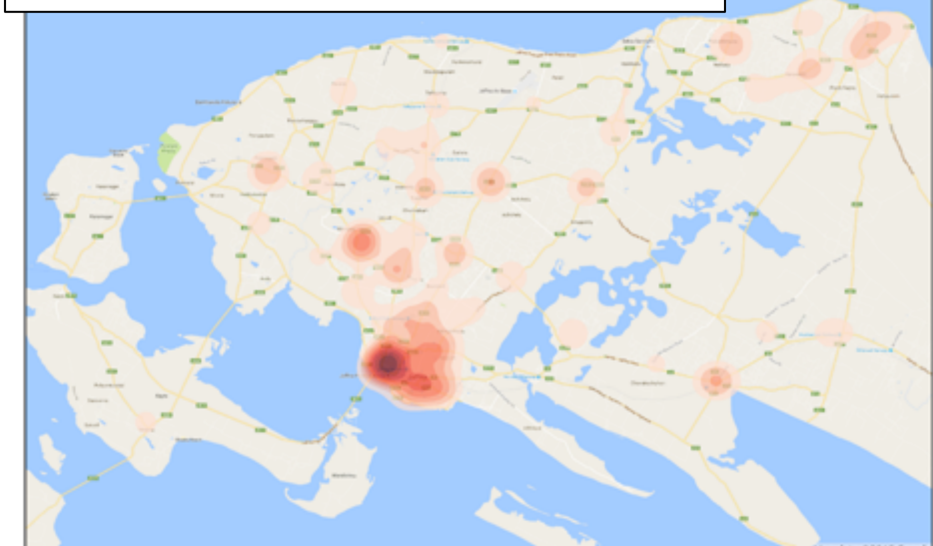
LIRNEasia
Pro-poor. Pro-market.

Where were they during the festival?

%-age increase in visitors during festival compared to an average day



People days density distribution during festival period



- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

LIRNEasia
Pro-poor. Pro-market.

Implications for public policy

- Population maps and mobility/migration patterns are essential policy tools
  - Included in most census and household surveys
- MNBD allows us to *improve* & *extend* measurement
  - Large sample sizes
  - Very frequent measurement
- More precise measures
  - Commuting/ seasonal/ long-term migration

LIRNEasia
Pro-poor. Pro-market.

Implications for public policy

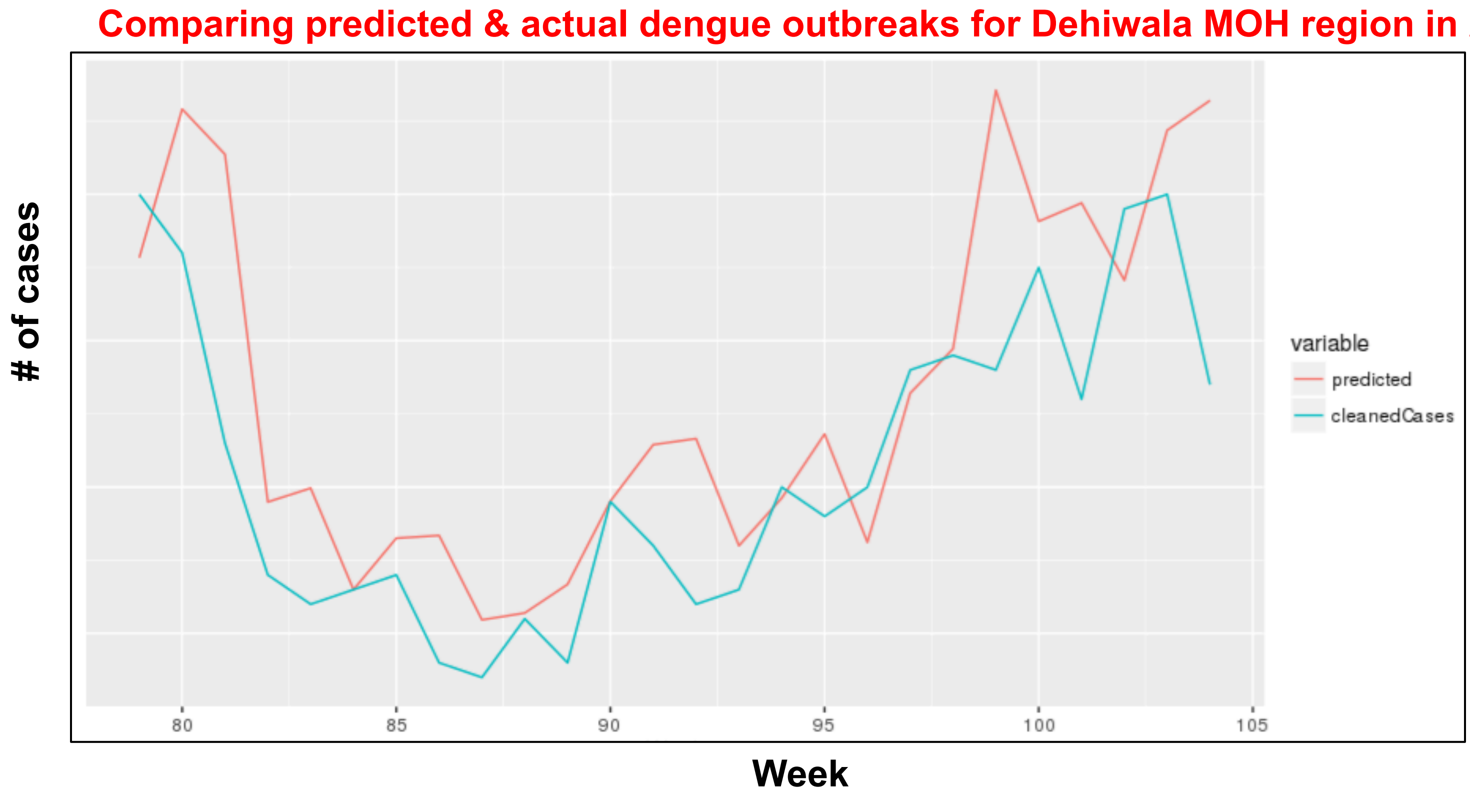
- Improve planning for “spiky” events that create pressure on government and privately supplied services
  - Humanitarian response to disasters
  - Special events/holidays
- Urban & transportation planning
  - Can identify high volume transport corridors to prioritize for provision of mass transit
  - Map de facto municipal boundaries
- Health policy
  - Mobility patterns that can help respond to spread of infectious diseases (e.g., dengue)

LIRNEasia Pro-poor. Pro-market.

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

LIRNEasia
Pro-poor. Pro-market.

Predicting spatial spread of dengue in Sri Lanka

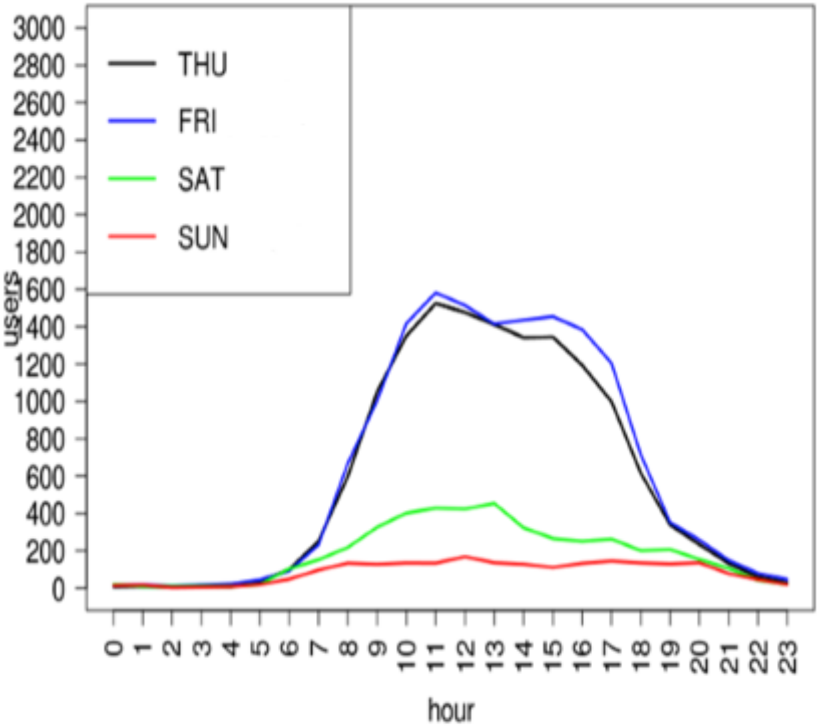


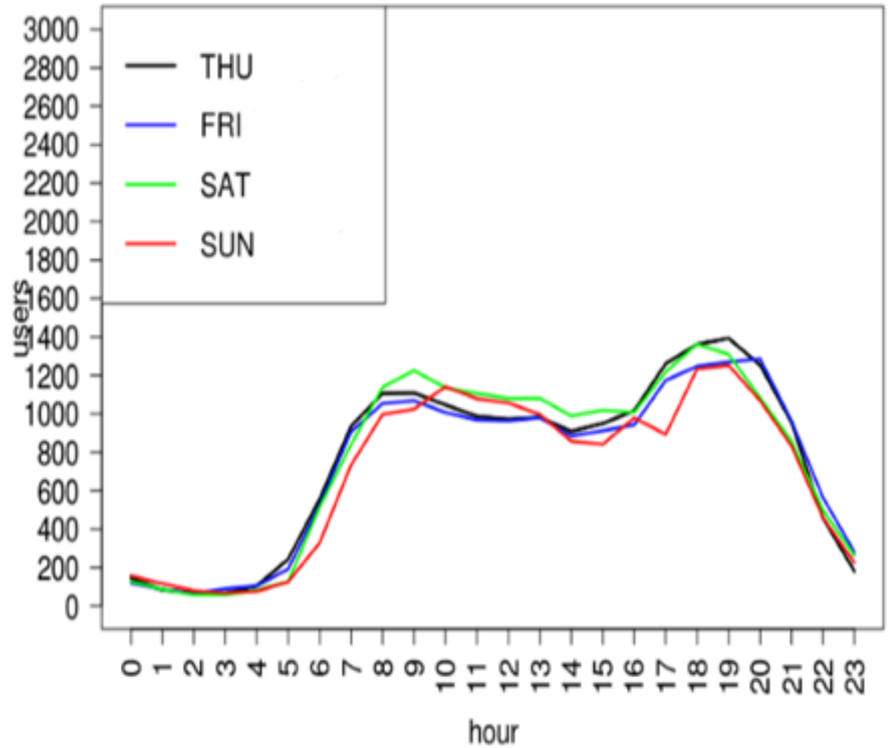
- Such analyses will be very important should any cases of zika be detected in Sri Lanka
- Initials results are being improved further in partnership with the University of Moratuwa and the Epid Unit of the Ministry of Health

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

LIRNEasia
Pro-poor. Pro-market.

Hourly loading of base stations reveals distinct patterns





Type X: ?

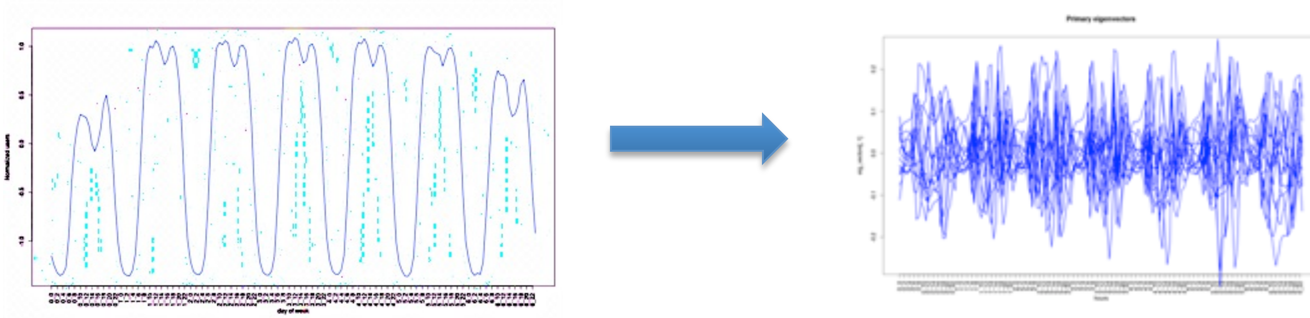
Type Y: ?

- We can use this insight to group base stations into different groups, using unsupervised machine learning techniques

LIRNEasia
Pro-poor. Pro-market.

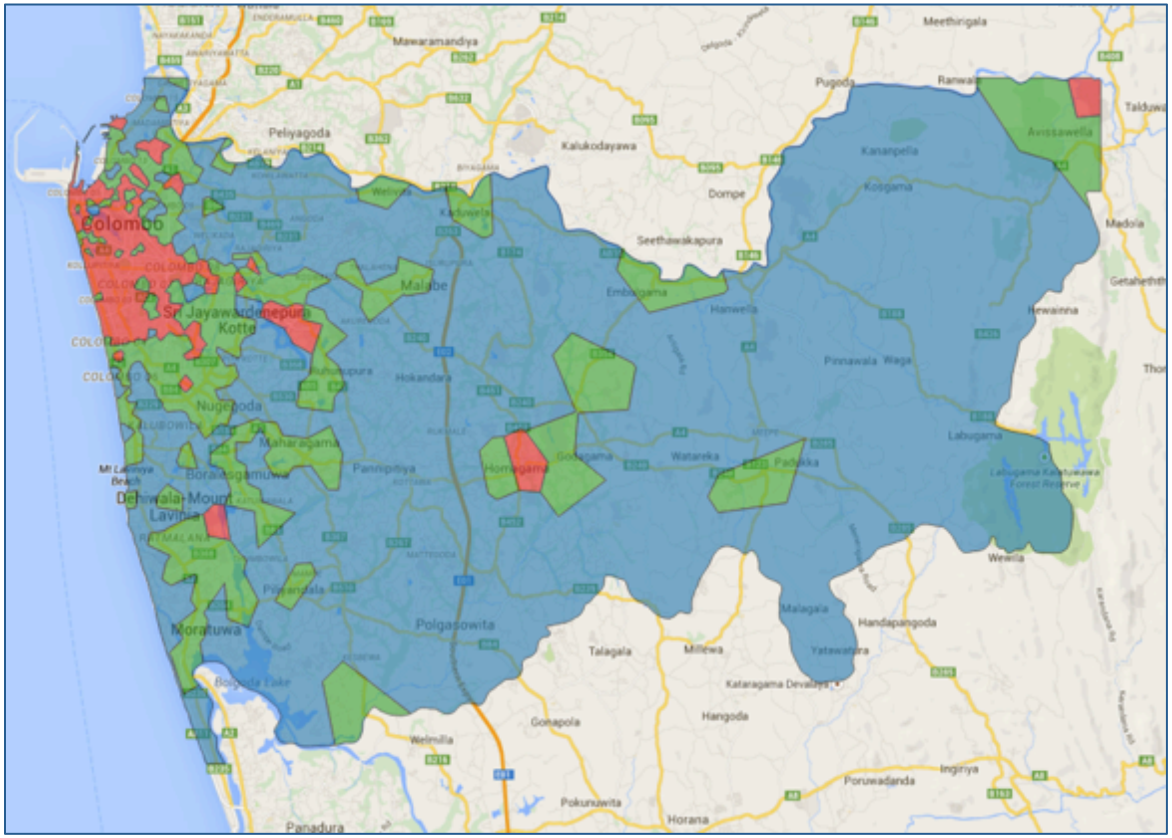
Methodology

- The time series of users connected at a base station contains variations, that can be grouped by similar characteristics
- A month of data is collapsed into an indicative week (Sunday to Saturday), with the time series normalized by the z-score
- Principal Component Analysis(PCA) is used to identify the discriminant patterns from noisy time series data



  - Each base station's pattern is filtered into 15 principal components (covering 95% of the data for that base station)
- Using the 15 principal components, we cluster all the base stations into 3 clusters in an unsupervised manner using k-means algorithm

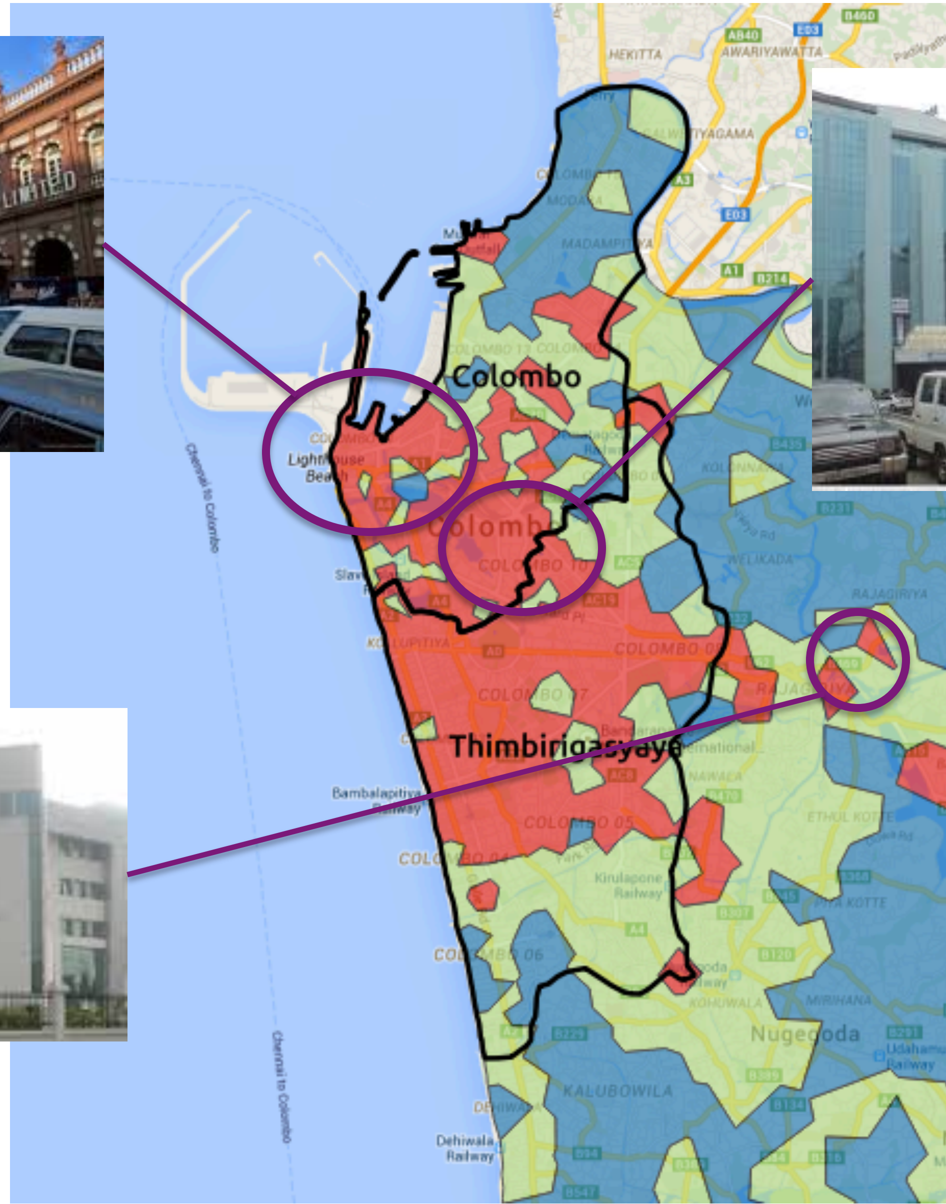
Three spatial clusters in Colombo District



- Cluster-1 exhibits patterns consistent with commercial area
- Cluster-3 exhibits patterns consistent with residential area
- Cluster-2 exhibits patterns more consistent with mixed-use

LIRNEasia
Pro-poor. Pro-market.

Our results show Central Business District (CBD) in Colombo city has expanded



LIRNEasia
Pro-poor. Pro-market.

Small area in NE corner of Colombo District classified as belonging to Cluster 1?

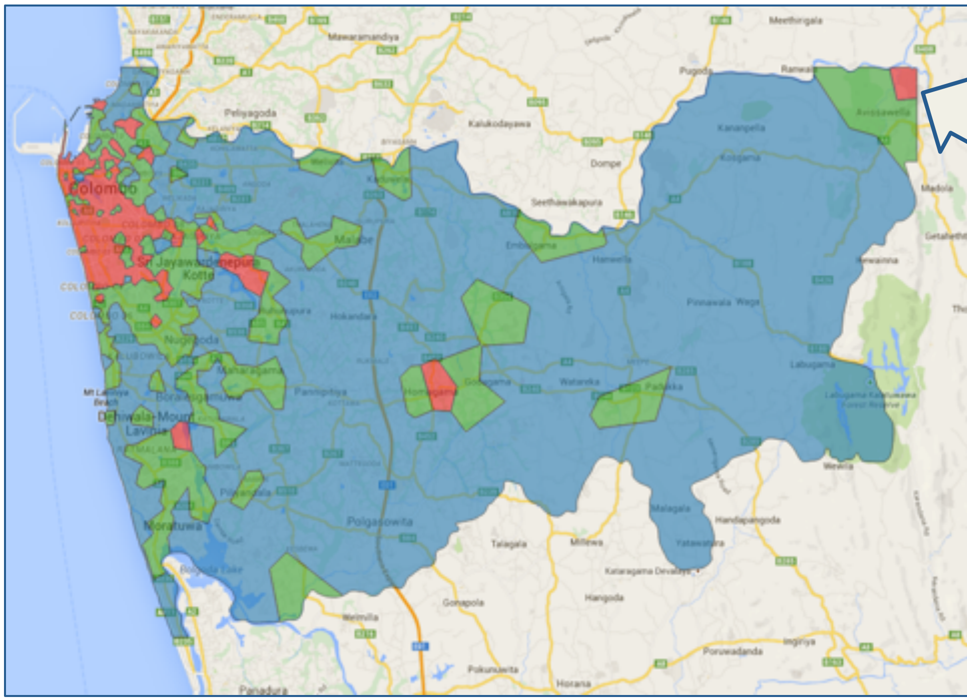


Photo ©Senanayaka Bandara - Panoramio

Seethawaka Export Processing Zone

LIRNEasia
Pro-poor. Pro-market.

We use silhouette coefficients to understand the quality of the clustering

- Silhouette coefficient indicates quality of clustering

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

a(i) - average distance of i with all other data within the same cluster

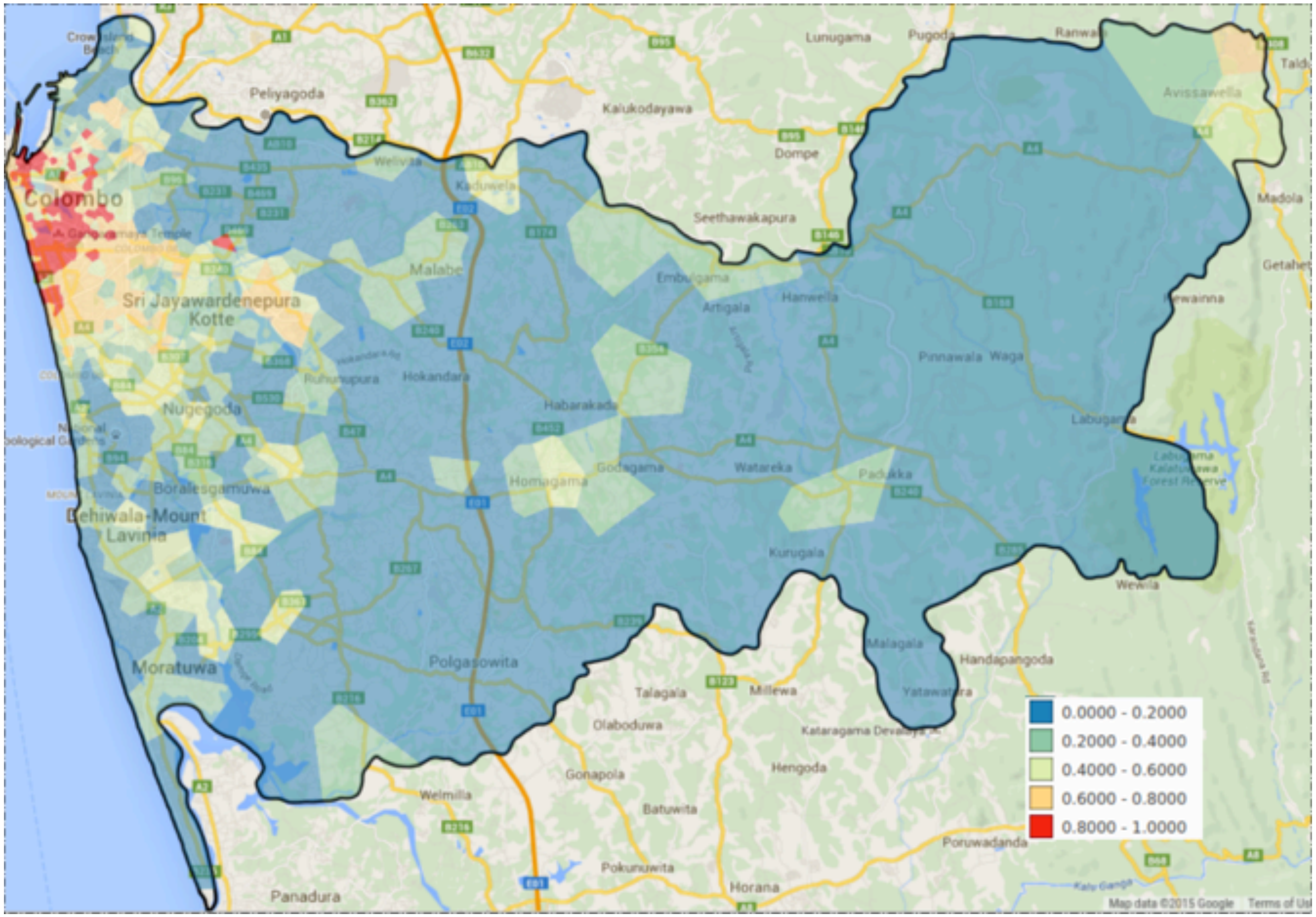
b(i) - average distance of i with all other data within the neighboring cluster

- Based on the s-values, Cluster 3 is the least coherent amongst the three

Cluster	Avg. Silhouette Coefficient
1 – Commercial	0.46
2 – Residential	0.36
3 – Mixed-use	0.22

LIRNEasia Pro-poor. Pro-market.

Commercial to Residential spectrum



Highly residential → Highly commercial

LIRNEasia
Pro-poor. Pro-market.

But using just MNBD gets us only so far

- Analyses currently being improved using machine vision techniques on satellite data, as well as using Foursquare data
- 4 different sources of data are being used to collectively give a better timely picture of land use
  - Existing ground truth survey data (data often old and geographically sparse)
  - MNBD
  - Satellite imagery
  - Foursquare

LIRNEasia
Pro-poor. Pro-market.

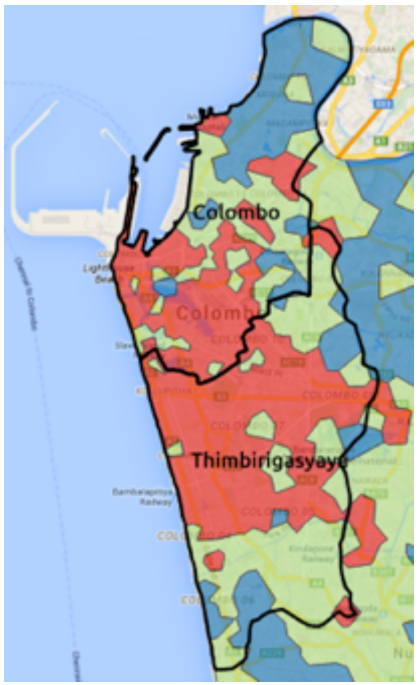
Implications for urban policy

- Almost real-time monitoring of urban land use
  - We are currently working on understanding finer temporal variations in zone characteristics (especially the mixed-use areas)
- Can complement infrequent surveys & align master plan to reality

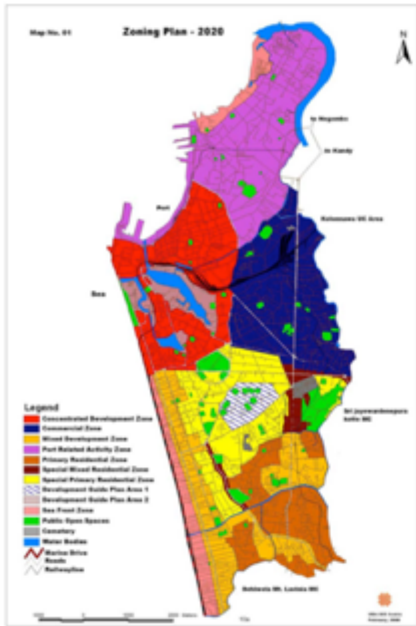
LIRNEasia
Pro-poor. Pro-market.

Final results will help a better comparison of reality against plans

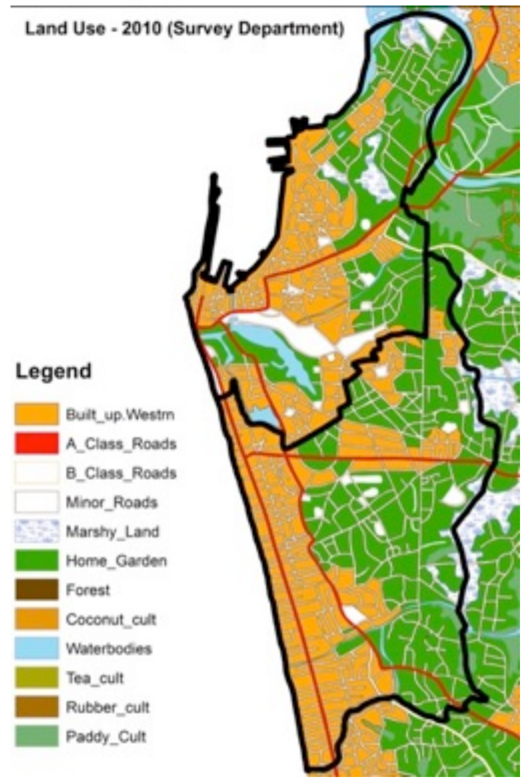
2013 MNBD analyses



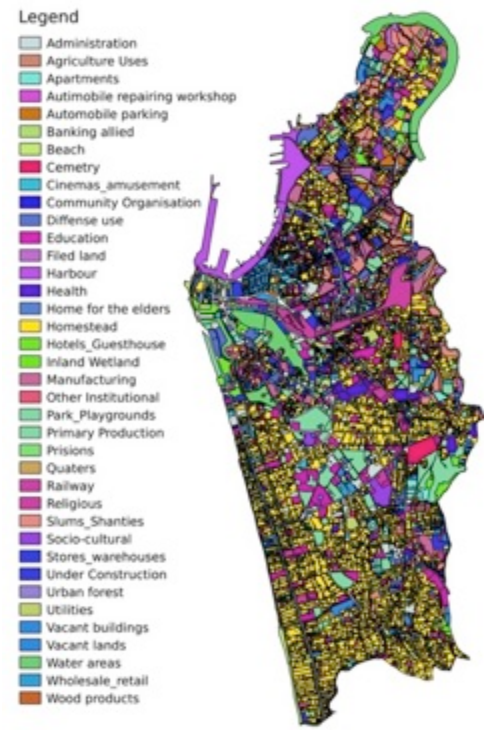
2020 UDA plan



2010 Survey Dept. Land Use Map



2000 UDA land use survey



LIRNEasia
Pro-poor. Pro-market.

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

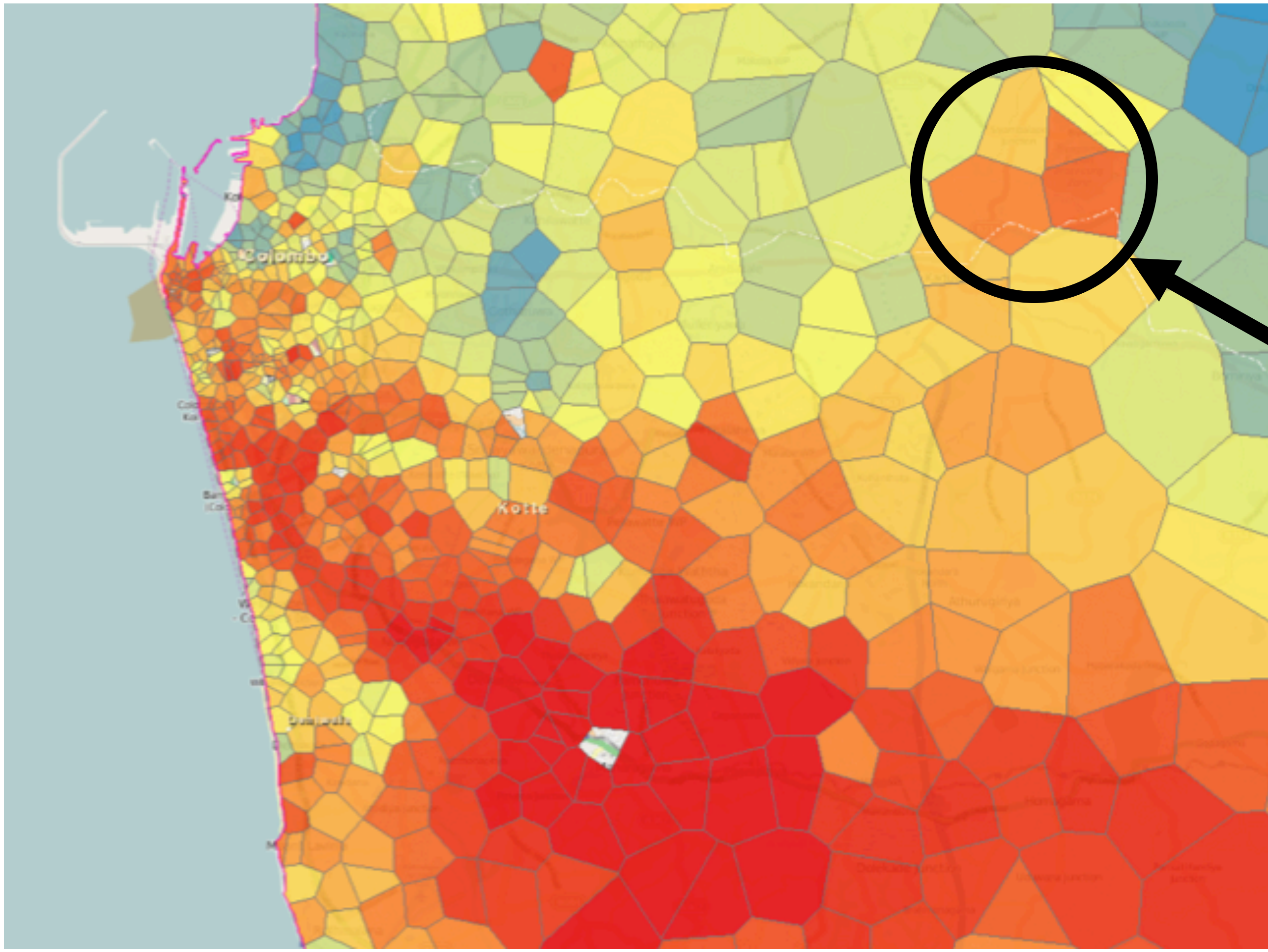
What does mobility tell us about economic activity?

Economic activity =
(number of workers) x **(productivity per worker)**

Observed — **Must be inferred**

- We assume more productive regions are more attractive destinations
- Commuting patterns emerge from the trade-off between attractiveness of a workplace and the cost of getting there

Example of commuting flows from one origin location



Biyagama Export Processing Zone

Low Commuting Flows

High Commuting Flows

Theoretical model outline

Agent ω at residential location i chooses work location j offering wage w_j and at distance d_{ij}, and ω has effective income at j

$$y_{ij\omega} = \frac{w_j z_{ij\omega}}{d_{ij}}$$

where $z_{ij\omega}$ is iid Fréchet-distributed random productivity shock.

Commuting flow probabilities:

$$\pi_{ij} = \frac{\left(w_j/d_{ij}\right)^{\epsilon}}{\sum_s (w_s/d_{is})^{\epsilon}}$$

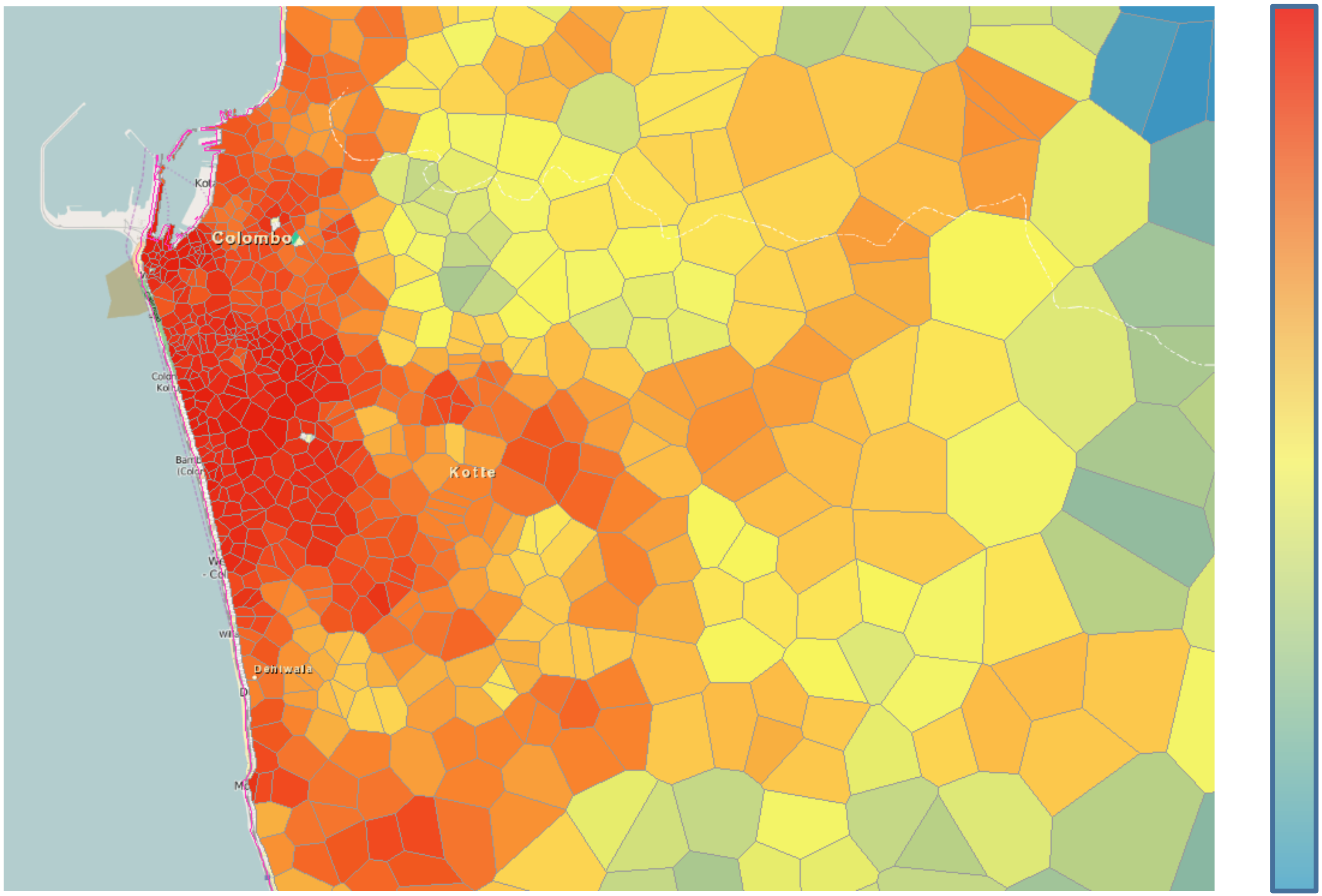
We estimate origin-constrained gravity model:

$$\log(\pi_{ij}) = \psi_j + \epsilon \log(d_{ij}) - \mu_i + \varepsilon_{ij}$$

LIRNEasia
Pro-poor. Pro-market.

We can develop new proxy measures of economic activity

Economic activity

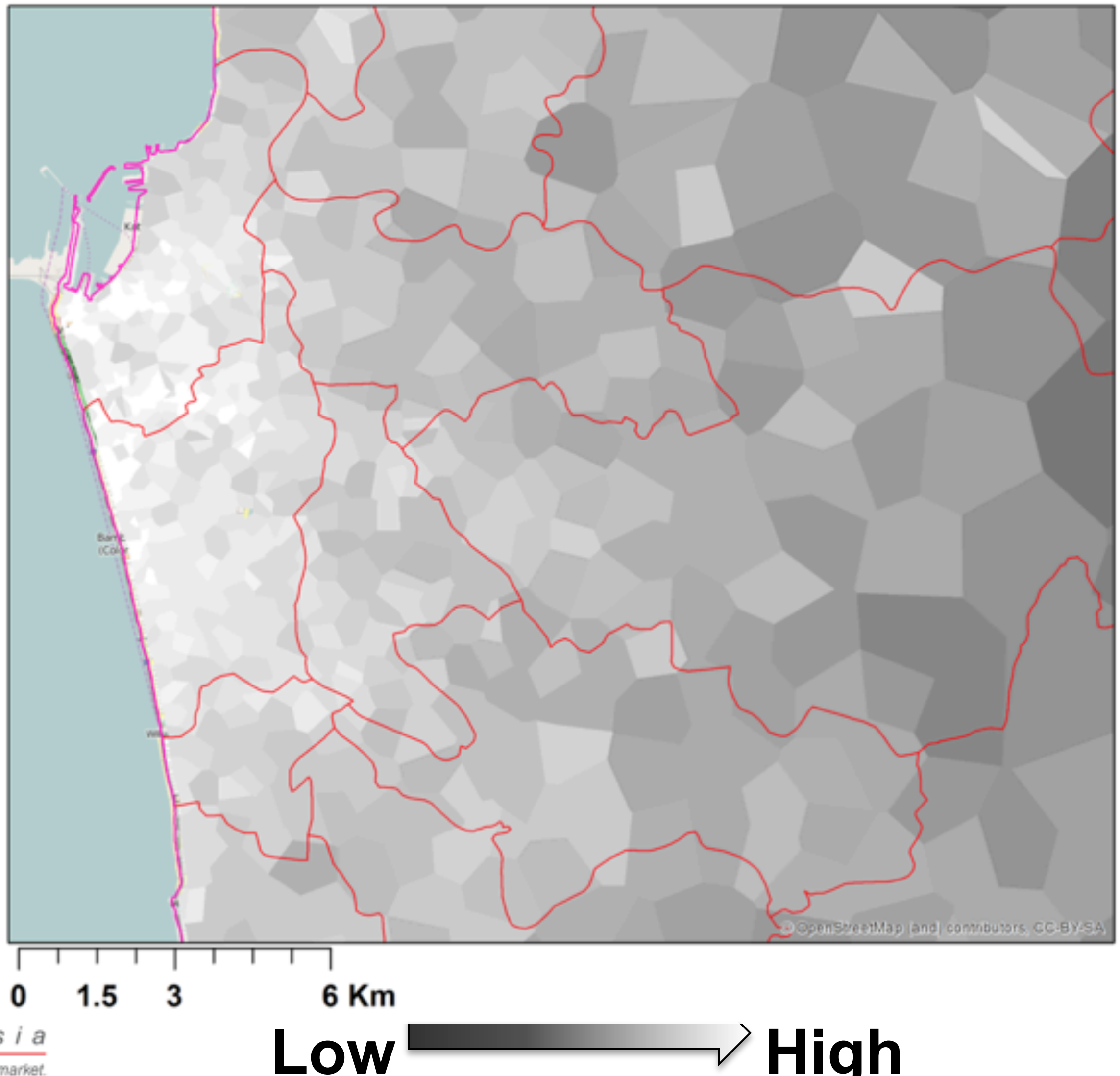


High economic activity

Low economic activity

LIRNEasia
Pro-poor. Pro-market.

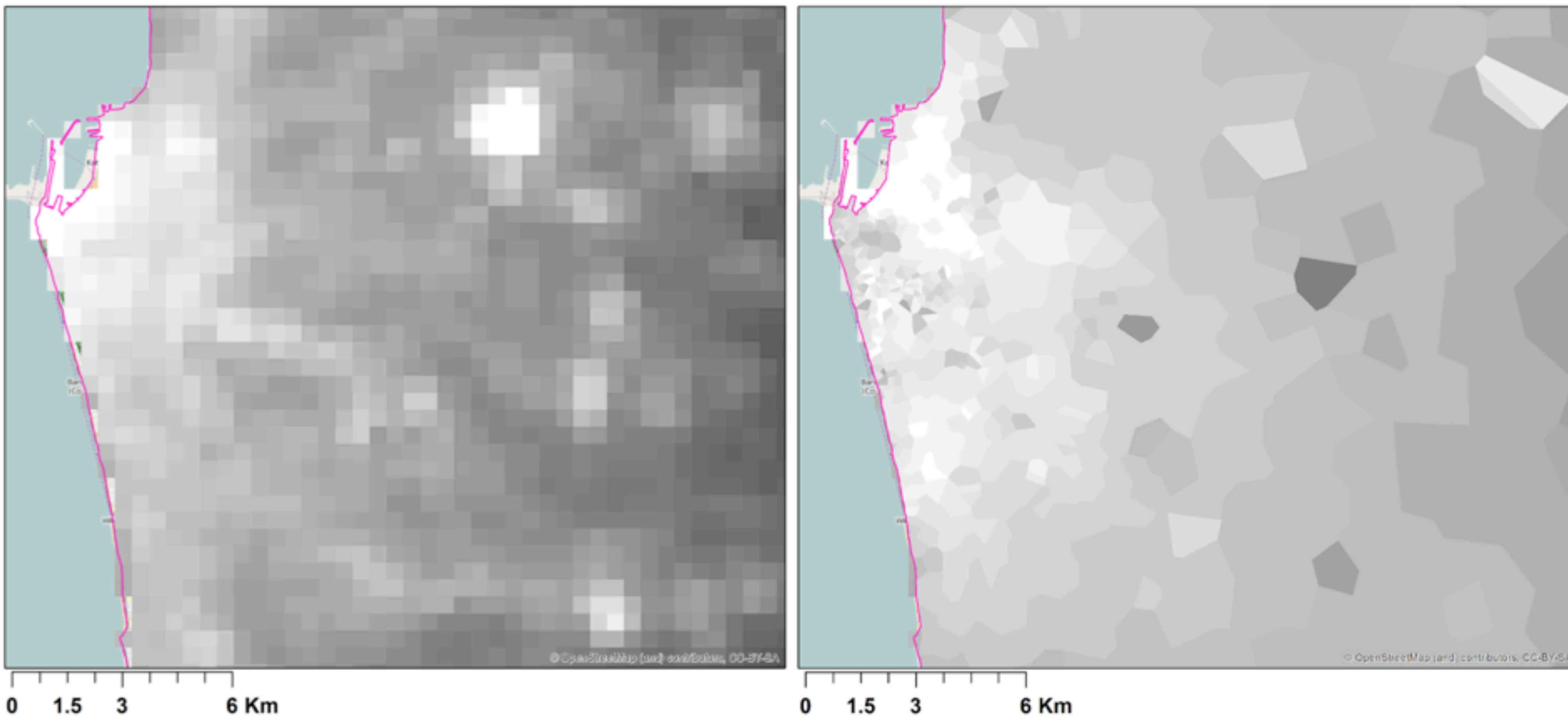
Economic activity/km^2



LIRNEasia
Pro-poor. Pro-market.

Model validation using nightlight data from satellites

Nightlights | Mean income



Low → **High**

LIRNEasia
Pro-poor. Pro-market.

Incorporating other data can give further insights

Household data: Census/HIES/LFS
Industrial data: ASI, Industrial Census

	Nightlights	Household data	Industrial Data
Geographic variation	✓	✓	✓
Time variation	yearly	quarterly/ 2-3yrs/decade	yearly/decade
Relevant variables		**Education, (un)employment, skill levels**	**Employment, capital intensity**
Ideal for:		Improving Measure	Improving & Validation

LIRNEasia Pro-poor. Pro-market.

Benefit of an improved framework for modeling economic activity

- **Increase the coverage** of existing surveys (both temporal and geographic)
  - By **calibrating** with household, industry census and survey data, when available
  - Then, mobile data can be used to **predict/extrapolate** for time periods and regions without survey data
- Can capture **informal economic activity**
  - Other research suggests informal economy is almost 30% of GDP in Sri Lanka

LIRNEasia
Pro-poor. Pro-market.

Implications

- A **new measure** of economic activity from commuting flows (from mobile phone data)
  - Significant potential use for policy and research.
  - Fine temporal and spatial resolution.

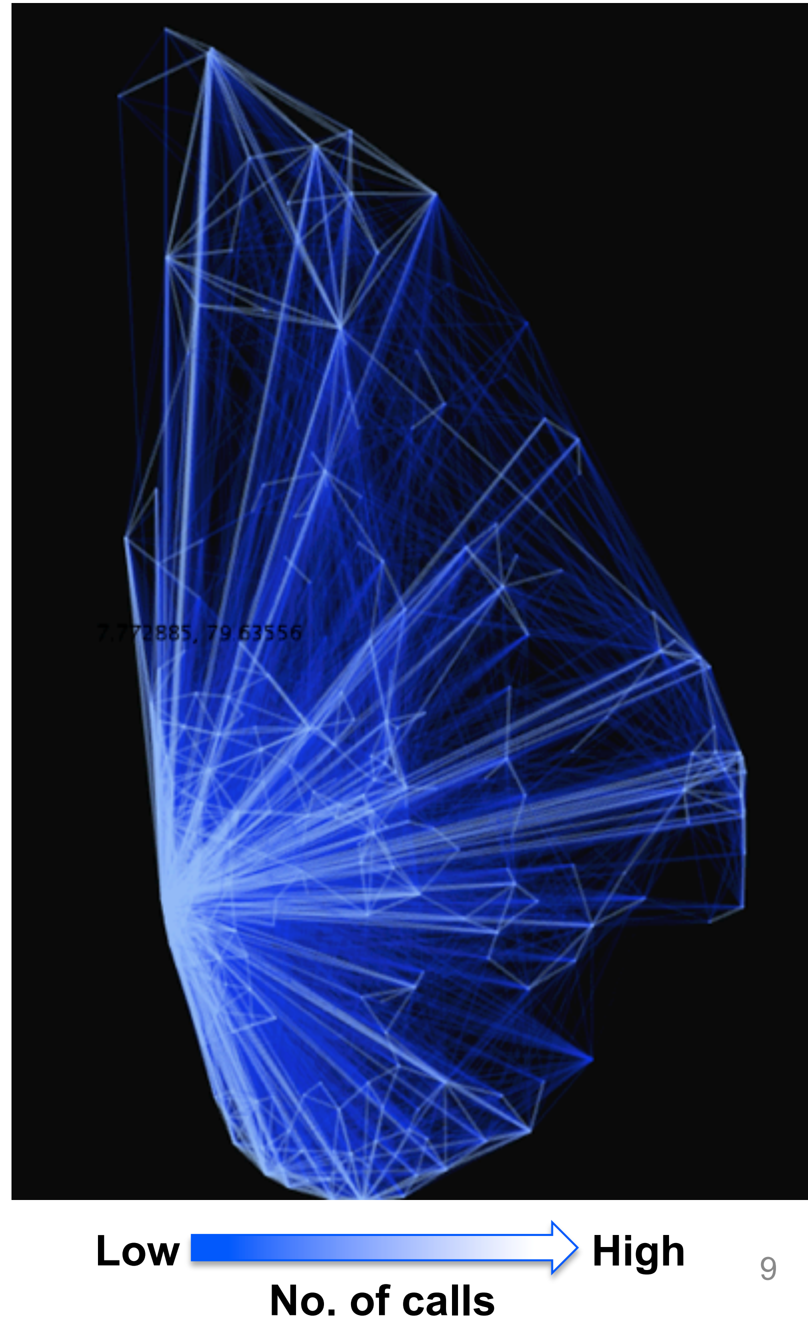
- **Preliminary validation** with the best available data looks promising

- **Additional data** (Industrial, Household) will allow measure to be taken to next level

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- **Understanding Sri Lankan communities**
- Analytical challenges
- Team & collaborators

Prima facie, Colombo city (Colombo & Thimbirigasyaya DSDs) seems to be the center of Sri Lanka’s social network

- Each link represents the raw number of outgoing and incoming calls between two DSDs
  - Divisional Secretariat Division (DSD) is a third level administrative division; 331 in total in LK

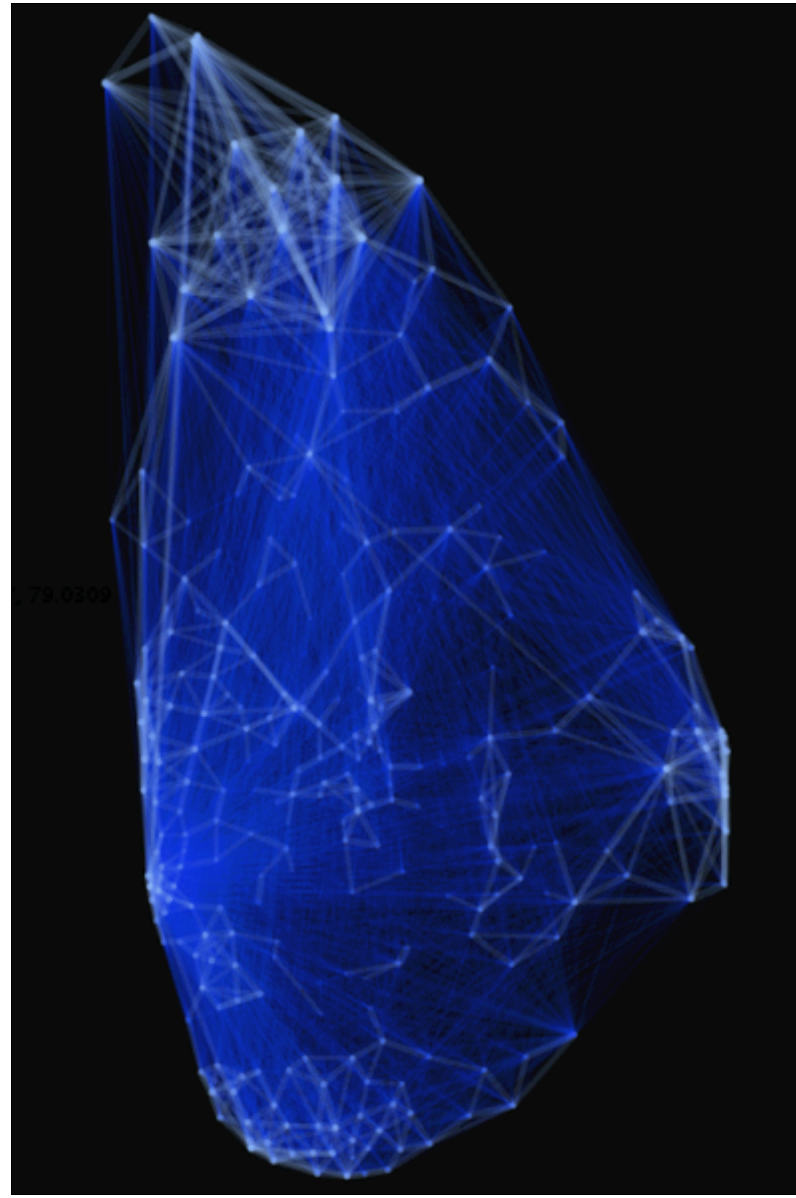


LIRNEasia
Pro-poor. Pro-market.

A different picture emerges when call volume is normalized by population

$$Normalized\ calls\ (DSD_1, DSD_2) = \frac{No. of\ calls\ (DSD_1, DSD_2)}{Population\ (DSD_1) \times Population\ (DSD_2)}$$

- Strongly connected regional networks become visible



Low → High

No. of calls

Identifying communities: methodology

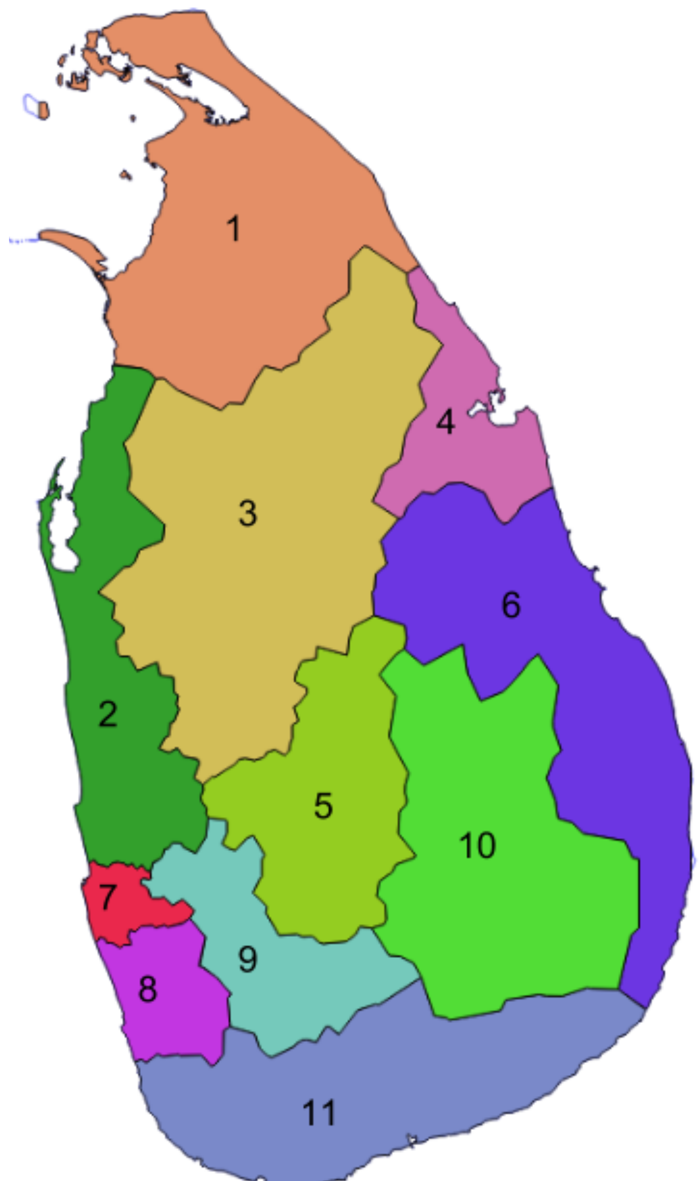
- The social network is segregated such that overlapping connections between communities are minimized
- Strength of a community is determined by *modularity*
  - Modularity Q = (edges inside the community) – (expected number of edges inside the community)

$$Q = \frac{1}{2m} \sum_{a,b} (A_{a,b} - \frac{k_a k_b}{2m}) \delta(c_a, c_b)$$

M. E. J.-Newman, Michele-Girvan, "Finding and evaluating community structure in networks", Physical Review E, APS, Vol. 69, No. 2, p. 1-16, 204.

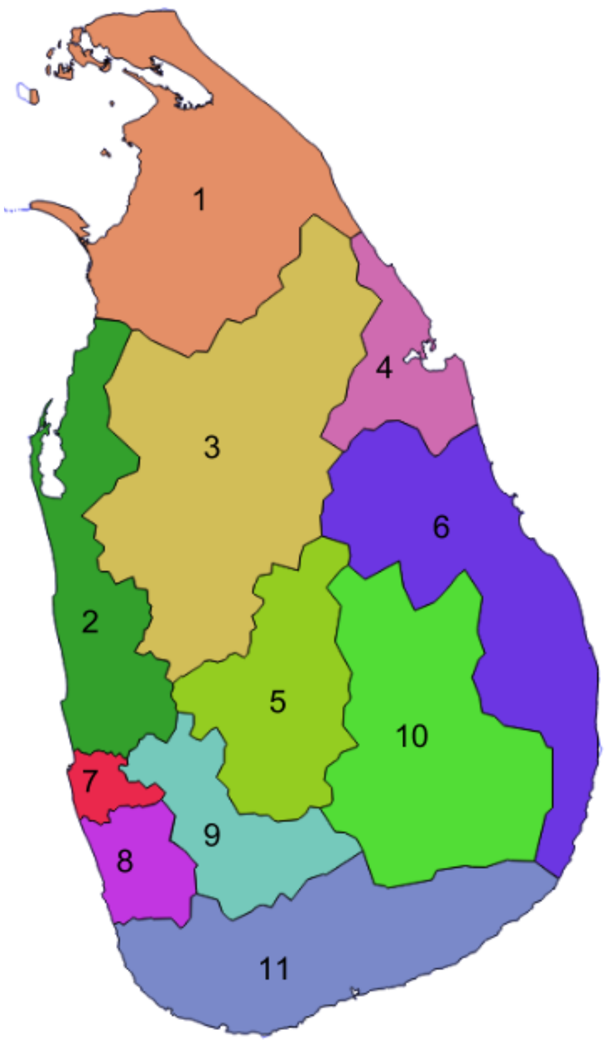
LIRNEasia Pro-poor. Pro-market.

We find Sri Lanka is made up of 11 communities

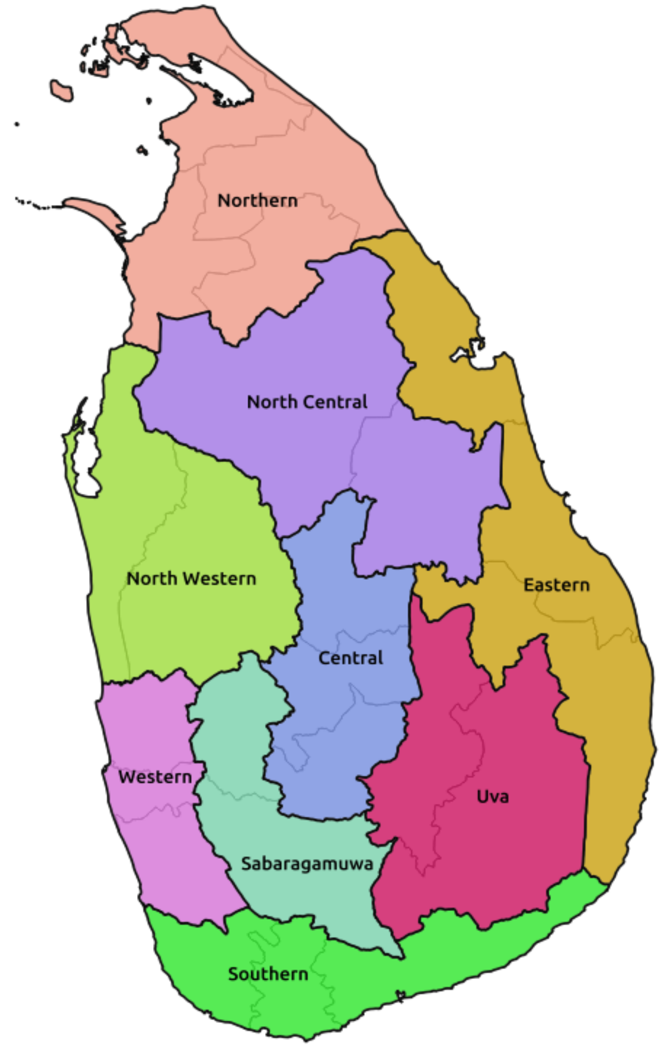


LIRNEasia
Pro-poor. Pro-market.

How do communities match existing administrative divisions?



The 11 detected communities



The 9 provinces

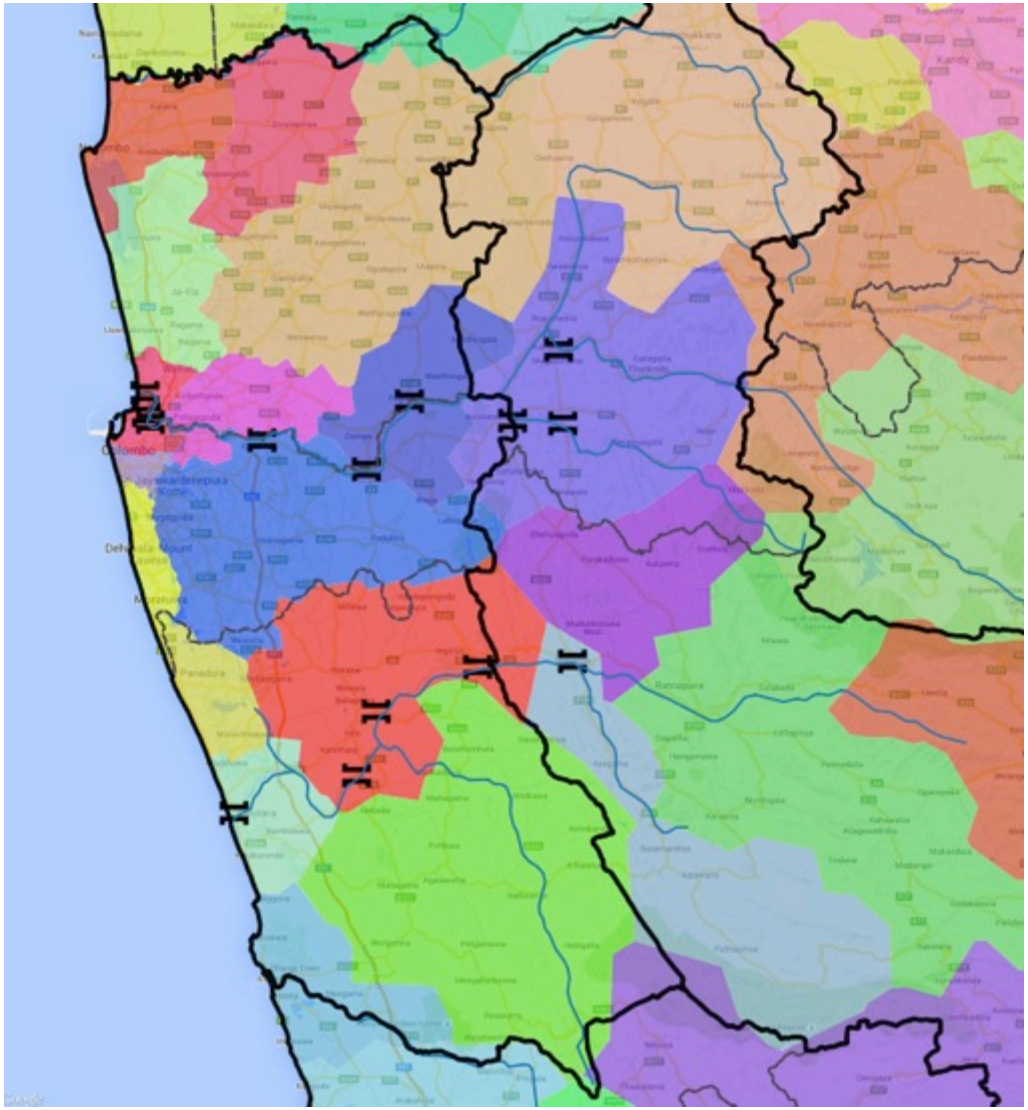
LIRNEasia

With some exceptions, boundaries of communities differ from existing administrative divisions

- Northern (1), Uva (10) and Southern (11) communities most similar to existing provincial boundaries; but 11 takes Embilipitiya and Kataragama
- Colombo district is clustered as a single community (7)
- Gampaha merges with coastal belt of North Western Province (2) and Kalutara (8) is its own community
  - What does this mean for Western Province Megapolis?
- Batticaloa & Ampara districts of the Eastern Province merge with the Polonnaruwa district of North Central Province to form its own distinct community (6)
  - Possibly reflective of economic linkages since this is the rice belt of Sri Lanka
  - Does economics override ethnicity?

LIRNEasia
Pro-poor. Pro-market.

More differences appear when we zoom in further



Bridge

- The littoral regions form their own distinct sub-communities
- The northern part of Colombo city forms a community with Wattala, across the Kelani river
- In general, rivers no longer form natural boundaries of communities

LIRNEasia
Pro-poor. Pro-market.

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

LIRNEasia
Pro-poor. Pro-market.

Addressing analytical challenges

Challenge	Solution(s)
Data is biased towards frequent users	• Understand and adjust for selection bias
Data sparsity	• Interpolation techniques • Probability based models
Different tower densities	• Different scale of analyses depending on region
Validating results	• Using other data sources, e.g., data from Dept. of Census and Statistics, transportation survey data

- Understanding population density & mobility
  - Population density
  - Commuting patterns: where do people live and work
  - Understanding traffic
  - Mobility changes during important events: Avurudu & Nallur festival
  - Implications for public policy
- Predicting spatial spread of dengue in Sri Lanka
- Understanding land use characteristics
- Measuring urban economic activity
- Understanding Sri Lankan communities
- Analytical challenges
- Team & collaborators

Staff

- Danaja Maldeniya (Computer Science, Statistics)
  - Now at U of Michigan but still collaborating
- CD Athuraliya (Computer Science)
- Dedunu Dhananjaya (Software Engineer, Systems Administrator)
  - Moved to private sector, but still collaborating
- Isuru Jayasooriya (Computer Science, Machine Vision)
- Kaushalya Madhawa (Computer Science, Statistics)
  - Now at Tokyo Institute of Technology
- Keshan de Silva (Computer Science, Agent Based Simulations)
- Lasantha Fernando (Computer Science)
- Madhushi Bandara (Computer Science)
  - Now at U of New South Wales but still collaborating
- Nisansa de Silva (Computer Science)
  - Now at U of Oregon, but still collaborating
- Robert Galyean (Mathematics, Physics)
- Prof. Rohan Samarajiva (Public Policy)
- Sriganesh Lokanathan (Public Policy, Computer Science)
- Shazna Zuhyle (Research Manager)
- Thavisha Gomez (Research Manager)

Collaborators

- Prof. Amal Kumarage (Dept. of Transport & Logistics, UoM)
  - Transportation, Urban Planning
- Dr Amal Shehan Perera (Dept. of CSE, UoM)
  - Data Mining
- Fields of View
  - Indian research institute specializing in games and simulations for public policy issues
- Gabriel Kreindler (MIT)
  - Economics, Statistics
- Prof Joshua Blumenstock (UC Berkley, School of Information)
  - Data Science, Economics, Statistics
- Prof Moinul Zaber (U of Dhaka)
  - Data Science, Public Policy
- Shibasaki & Sekimoto Laboratory, University of Tokyo
  - Big Data for Development research practice
- Yuhei Miyauchi (MIT)
  - Economics, Statistics

Selected Publications & Reports

- Lokanathan, S., Kreindler, G., de Silva, N. D., Miyauchi, Y., Dhananjaya, D., & Samarajiva, R. (2016). Using Mobile Network Big Data for Informing Transportation and Urban Planning in Colombo. *Information Technologies & International Development*
- Samarajiva, R., Lokanathan, S., Madhawa, K., Kriendler, G., & Maldeniya, D. (2015). Big data to improve urban planning. *Economic and Political Weekly*, Vol L. No. 22, May 30
- Maldeniya, D., Lokanathan, S., & Kumarage, A. (2015). Origin-Destination matrix estimation for Sri Lanka using mobile network big data. 13th International Conference on Social Implications of Computers in Developing Countries. Colombo
- Kreindler, G. & Miyauchi, Y. (2015). Commuting and Productivity: Quantifying Urban Economic Activity using Cell Phone Data. LIRNEasia
- Lokanathan, S & Gunaratne, R. L. (2015). Mobile Network Big Data for Development: Demystifying the Uses and Challenges. *Communications & Strategies*.
- Lokanathan, S. (2014). The role of big data for ICT monitoring and for development. In *Measuring the Information Society 2014*. International Telecommunication Union.

More information:

http://lirneasia.net/projects/bd4d/

LIRNEasia
Pro-poor. Pro-market.