Commuting and Productivity: Quantifying Urban Economic Activity using Cellphone Data

Gabriel Kreindler
Yuhei Miyauchi
Economics Department, MIT

Netmob, April 8th 2015

LIRNEasia
Pro-poor. Pro-market.

IDRC CRDI
Canada

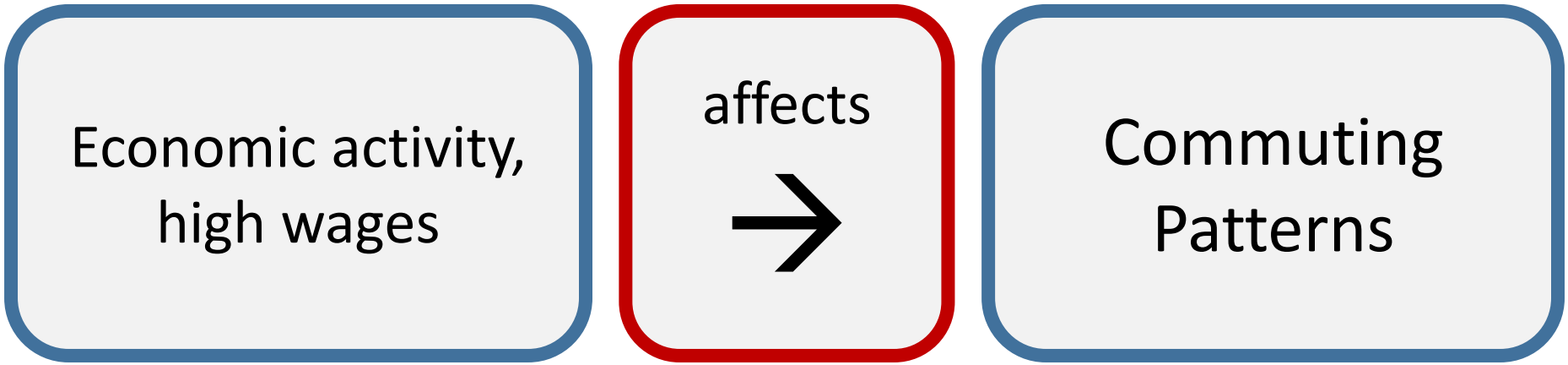
This work was carried out with the aid of a grant from the International Development Research Centre, Canada and the Department for International Development UK..

UKaid
from the British people

Source: https://macetology.wordpress.com/

Usual Research Approach:

- *How does GDP **affect** migration patterns?*
- *How do higher wages **influence** commuting?*

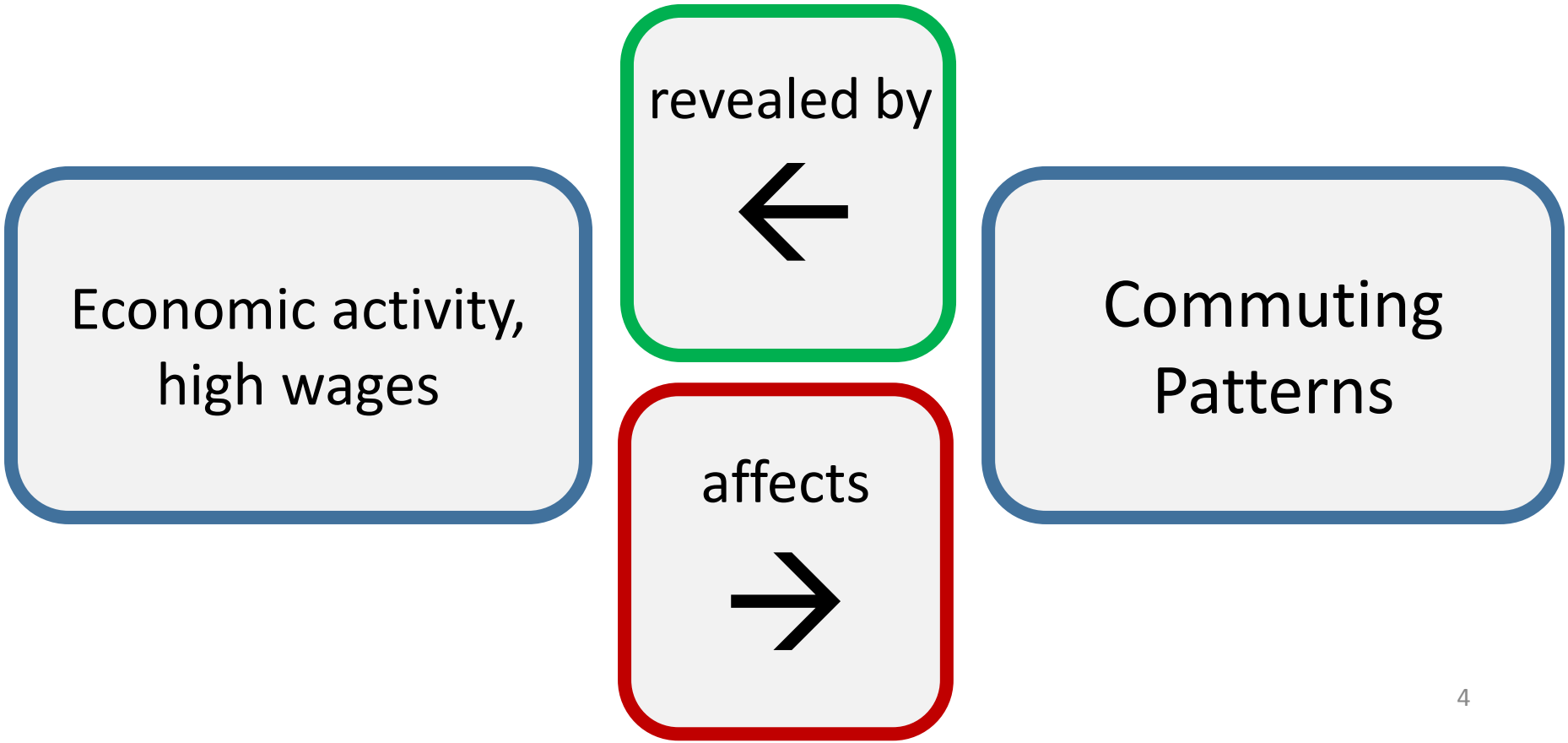


Usual Research Approach:

- *How does GDP **affect** migration patterns?*
- *How do higher wages **influence** commuting?*

Our Research Question:

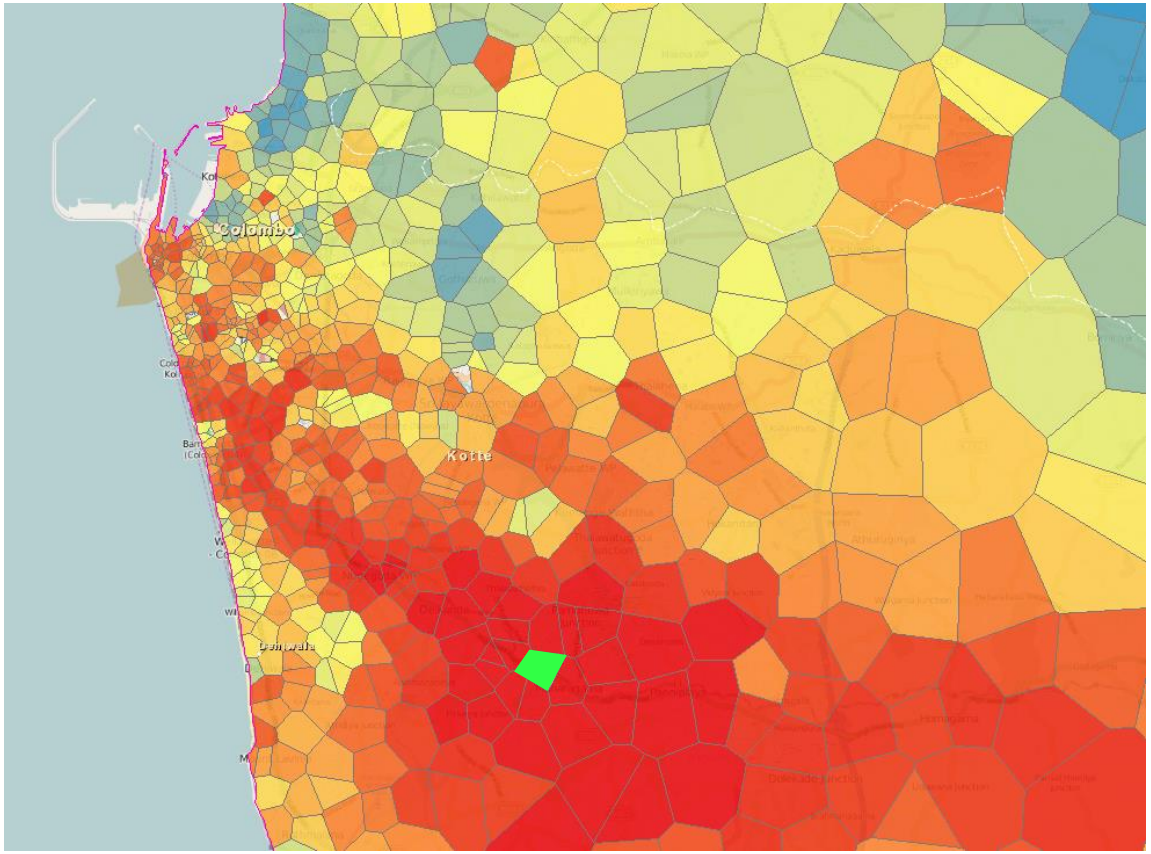
- *Can we **infer** economic activity from commuting?*



What do we mean by economic activity?

- $\text{Output}_i = (\text{number of workers @ } i) \times \left(\frac{\text{output}}{\text{worker}}\right)_i$
- $\left(\frac{\text{output}}{\text{worker}}\right)_i$ is productivity
- Assumption: more productive workers paid more

How are commuting flows informative about productivity?



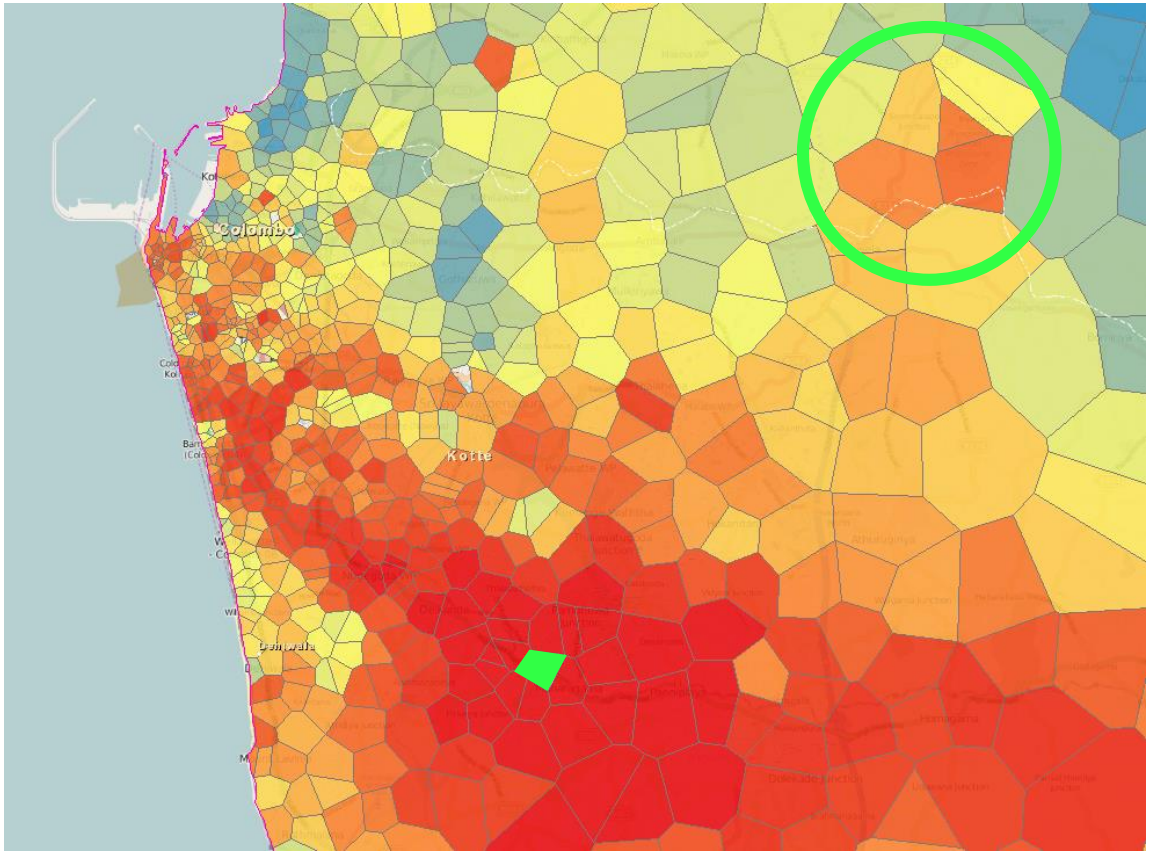
High Commuting Flows

Low Commuting Flows

Commuting flows defined for each cell phone owner:

- first location b/w 5am & 10am => origin location
- last location b/w 10am & 3pm => destination location

How are commuting flows informative about productivity?



High Commuting Flows

Low Commuting Flows

Commuting flows defined for each cell phone owner:

- first location b/w 5am & 10am => origin location
- last location b/w 10am & 3pm => destination location

What we do:

1. Set up an agent-based model of job location choice
   - Implies commuting flows satisfy a gravity equation
   - Destination attractiveness is a function of wage

2. Estimate destination attractiveness using commuting flows extracted from cell phone data

3. (*In progress*) We validate the model using economic activity data from a separate data source

From the Theoretical Model to the Data

- Agent ω at residential location i chooses work location j that maximizes:

$$\underbrace{y_{ij\omega}}_{\text{income}} = \underbrace{w_j}_{\text{wage}} \underbrace{z_{ij\omega}}_{\substack{\text{random un-modeled} \\ \text{factors}}} / \underbrace{d_{ij}}_{\text{distance}}$$

- Commuting flow probabilities follow origin-constrained gravity model:

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

wage w_j related to destination attractiveness: $\psi_j = \log(w_j)$.

- Model makes (strong) assumptions:
  - (Almost) identical agents
  - All commuting is work related (no education, shopping, etc.)
  - Choice of work only depends on wages and commuting distance
  - Specific functional form

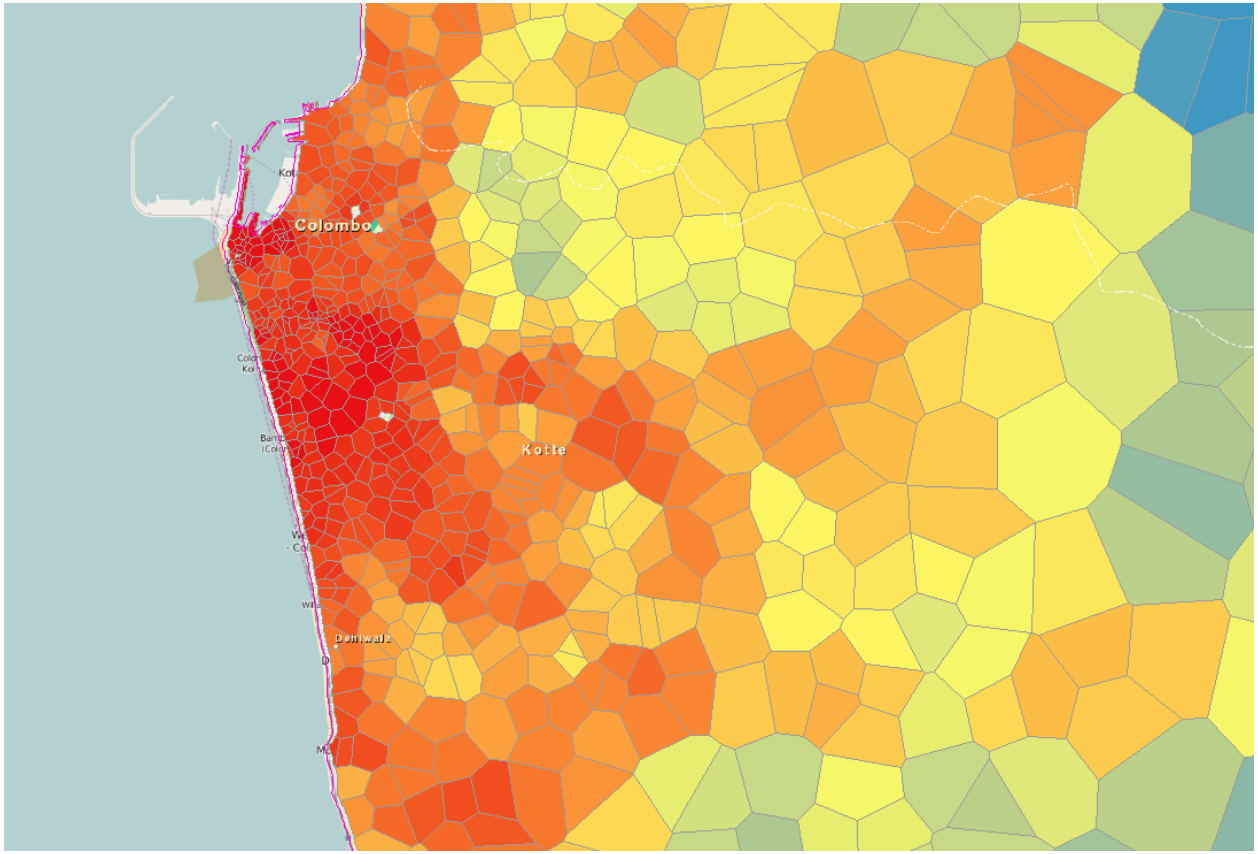
Our take: how far can we get with a very simple model?

Data and Estimation

- Commuting flows extracted from CDR data
  - Defined at (cell phone) × (day level):
    - origin location ← first location b/w 5am & 10am
    - destination location ← last location b/w 10am & 3pm

- Estimation:
  - Currently OLS with origin and destination factor variables (fixed effects)
  - Adapted for high dimensional factor variables
  - Ideally: impose origin constraint (non-linear)

Results: estimated log(wage)



High estimated wage

Low estimated wage

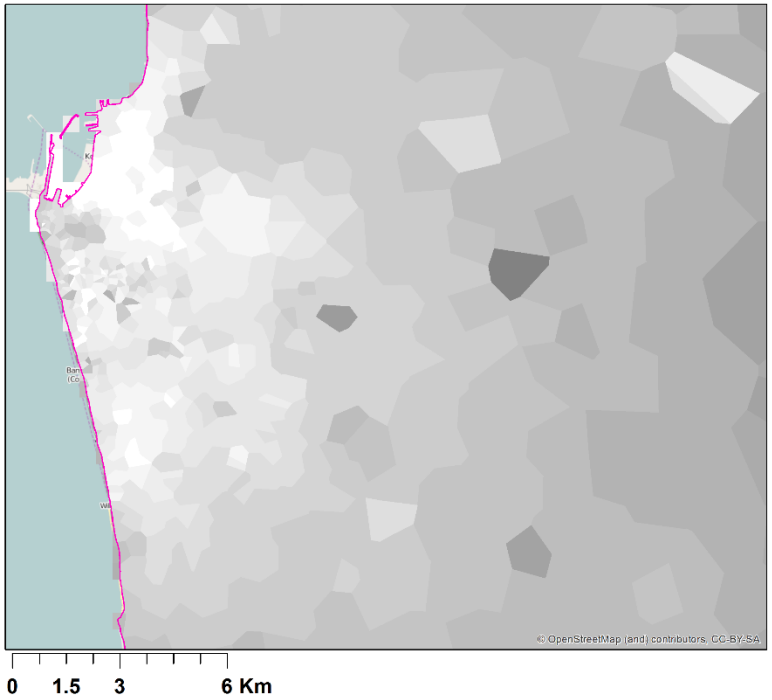
Intuition: log(wage) is estimated as employment *in excess* of what is predicted by distance to residential population.

Preliminary Validation

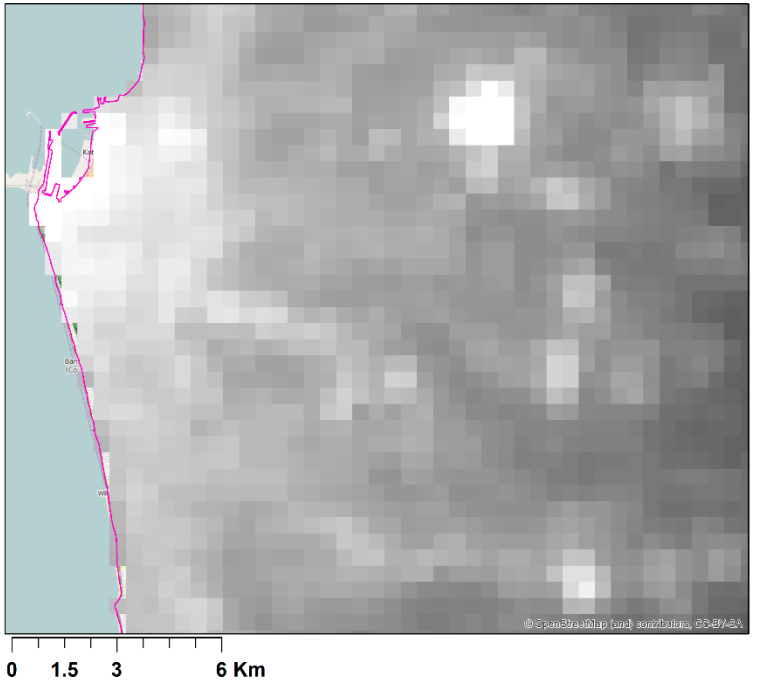
- **Ideal validation:** using independent wage data at fine spatial resolution.
  - Alternatively: commercial electricity consumption data.

- Today: use nighttime lights data (VIIRS)
  - Indirect way to test the model
  - Good proxy of *residential income* (Mellander et al. 2013)
  - We use the model to predict residential income.

Predicted Mean Income



Nighttime lights (VIIRS)



R-squared: 0.71

- Correlated **after** controlling for population, tower size, etc.

Next Steps

- Exploit rich time variation in cell phone data – dynamics
- Validation of the model with better data.

- Applications:
  1. Measure impact of reducing fuel subsidies. (e.g. Jakarta)
  2. Effects of *Hartals / Oborodh* in Bangladesh (strikes that shut down transportation and econ activity)
  3. Study impact of transport restrictions (*Cordon Sanitaire*) in Sierra Leone due to Ebola.
- Complement traditional data collection on economic activity (GDP)

Thank You!