Mapping the Margins

Geospatial tools for measuring informal and illicit economies

Melissa (Mel) Pavlik

Yale University | June 2025

Informal economies and industries

The politics of informality are a key interest for social scientists

We ask questions addressing:

- *Informality's causes*
- *Informality's consequences*
- *Effects of policies targeting informality*

About informal sectors like:

- *Informal transit and transport*
- *Street vending, hawking, informal markets*
- *Artisanal mining*

However: informal industries are inherently difficult to measure

1. By nature "in the shadows"
2. Small-scale and context-dependent
3. Spatially fragmented; not bounded by formal administrative units

This project

Key purpose

Present strategies for making use of **fine-grained geospatial data** to overcome some of these measurement challenges

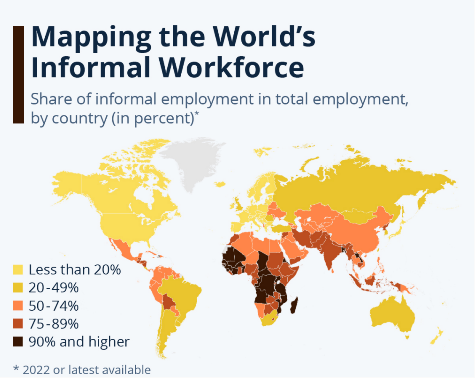
- Integrate **Very High Resolution (VHR) remote sensing data**, augmented via processing computer vision processing and spatiotemporal fusion
- 'Shrink the scale' at which we measure these industries to overcome modifiable areal unit problem (MAUP) and other issues
- Allowing insights beyond what we can learn from coarser data

Empirical example

Measurement of informal transit industry in Lagos, Nigeria and fine-grained enforcement patterns of a 2022 ban on unregistered motorcycle taxis

Current Ways of Measuring Informal Industries

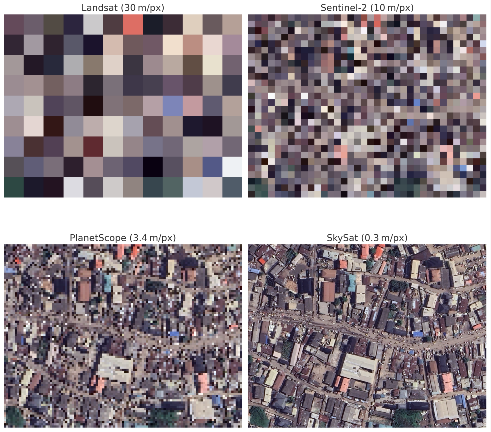
- **Ethnographies, fieldwork, surveys**
  Limited in scope and temporally static
- **Official estimates and administrative data**
  Selectively collected and incomplete in non-random ways
- **Nighttime lights and remote sensing**
  Promising, but insufficient resolution for micro-scale dynamics



New opportunities using emerging forms of geospatial data

Very High Resolution (VHR) remote sensing data

Satellite imagery with spatial resolutions of up to ≈ 25-30cm per pixel



- Remote sensing imagery increasingly used; primarily coarse resolution (≈ 30m per pixel) and across large swaths
- Recently loosened regulations by the U.S. & European Space Agency (ESA) → higher spatial resolution data available

Advantages (and disadvantages) of these data

The good news:

- Extremely precise, highly disaggregated
- Will only become more fine-grained and more available
- Fosters context-specific and highly flexible measurement

The bad news:

- Expensive!
- Varying availability and potential non-random missingness
- Computationally intensive analysis

HOWEVER: Some of these disadvantages can be mitigated with certain types of pre-processing and integration of other types of data!

Basic methodological framework

1. **Shoe leather** Understand what to look for based on context
2. **Unit-aware spatial referencing** Aggregate into theoretically-informed units
3. **Hand-code** Manually label a selection of images (familiar "image-as-data" pipelines)
4. **Object detection** Train and predict with computer vision models
   - *Bonus* Use adversarial de-biasing models to correct bias in predictions (Gordon et. al, 2025)
5. **Integrate coarser data:** Use spatiotemporal fusion to infer distributions from more widely available but less precise data
   - *Bonus* Use field observations and alternative data sources to add "observed anchors" informing the observation likelihood
6. **Analyze and model**
   - *Bonus* Use geographically weighted machine learning (GWML) to reveal which features matter where (Wang et al., 2024)

Empirical example: Informal motorcycle taxis in Lagos, Nigeria

- Lagos is Africa's largest metropolitan area (population $>$ 23 million), with a poorly connected road network; Lagos traffic is infamous
- Lagosians rely on huge informal transit industry of unregistered minibuses, tricycles, and 'okada' motorcycle taxis to navigate the city
  - Fostered emergence of Lagos transport unions, mafia-like organizations that coerce taxation from commercial vehicles that pass through their 'motorparks'
  - Left alone by the state, who rely on the union for election interference and thuggery
- Lagos began to enforce an 'okada ban' in June 2022, allegedly to reduce traffic congestion and crime

Research question

What explains uneven state enforcement of the Lagos okada ban?

The data

- VHR remote sensing data along Lagos roadways from commercial providers and Google Earth historical imagery
- Open Street Map (OSM) road data, transformed into connected spatial network
- Original data collection of various potential points of interest along the road network: police presence, traffic chokepoints, motorpark locations, etc.

The method

1. **Shoe leather** Fieldwork-based information on spatial patterns of okada
2. **Unit-aware spatial referencing:** Aggregate okada counts not into local government areas or randomly-defined grid-cells, but into segments of road along a network
3. **Hand-code** Manually label a 4,000 (familiar "image-as-data" pipelines)

The method (continued)

4. **Object detection** Train and predict okada presence on over 50,000 road segment images
5. **Integrate coarser data:** Use Bayesian hierarchical spatial disaggregation to include additional lower-resolution images to fill in gaps and interpolate presence across unsampled areas and times
    - *Bonus* Use additional hyper-precise (but temporally sparse) data from Google Streetview imagery to bound observed likelihood

The results: Road-segment level

Key takeaway

The main effect of okada ban enforcement was to push operators into areas controlled by extortionate state allies who benefit from okada presence in their motorparks.

	Okada presence			
	(1)	(2)	(3)	(4)
Motorparks × Post	3.050***			
	(1.009)			
Police stations × Post		-0.4451		
		(0.9001)		
Traffic × Post			0.1534	
			(0.2227)	
Pre-ban presence × Post				-0.1380
				(0.2003)
Date FE	Yes	Yes	Yes	Yes
Road segment FE	Yes	Yes	Yes	Yes

The results: Aggregated up

Key takeaway

We don't see this effect when running the same analysis at the grid-cell level (below) or higher.

	Okada presence			
	(1)	(2)	(3)	(4)
Motorparks × Post	106.8			
	(65.44)			
Police stations × Post		-61.41		
		(81.11)		
Traffic × Post			79.92**	
			(24.73)	
Pre-ban presence × Post				-6.875
				(45.52)
Date FE	Yes	Yes	Yes	Yes
Road segment FE	Yes	Yes	Yes	Yes

Implications

- This analysis may imply that the politics of ban enforcement may center on a corrupt relationship with a third-party ally, rather than other motivations; increasing enforcement so as to drive okada towards areas of union extortion
- **We cannot come to this conclusion by doing analysis at the grid-square level, or at the local government level; even if we could measure the informal transport industry at those higher levels of aggregation**
- Using micro-level data and very finely disaggregated geospatial data, we can come to insights about the patterns of informal industries that are not otherwise possible

Thank you!

melissa.pavlik@yale.edu