Satellite Vision

Satellite and Spatial Data for Public Policy

Instances from Indian Economic Survey

Parth Khare

IDFC Training session

September 5th, 2019

Section I: Satellite Imagery
Remote Sensing to Computer Vision

Policy & Data: Emerging economies

Leveraging data externality

- Satellite data free from human error and intentional field survey biases
- Volumes of private data (behavioral consumer) being collected holding high potential for public policy
- Census is conducts in 10 years span, so significant on poverty cannot be captured
- Dynamic changes need to be understood for effective policy
- Tools and platforms to learn data sciences R, Python
- Big Data, lesser time, need for dynamic visual tools

4857 Satellites

- Untapped resource
- LUC to Machine Learning to Computer Vision
- Emerging countries

Source: NASA

Build & Scale

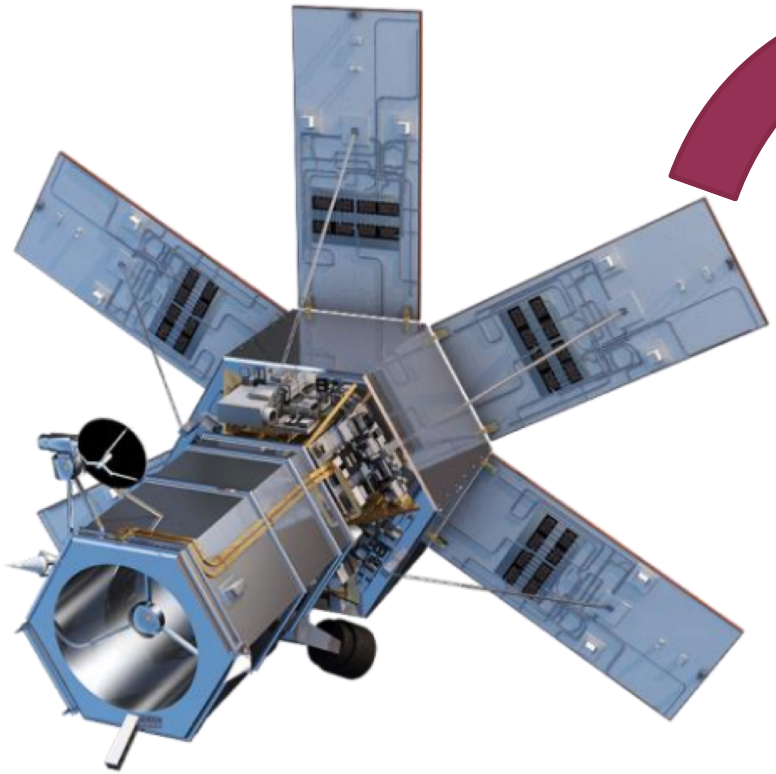


Photo Credit: DigitalGlobe

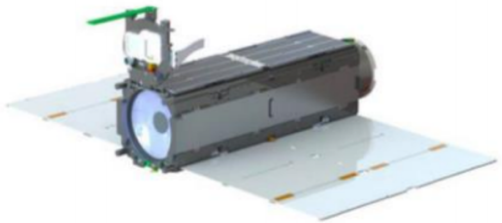


Photo Credit: Planet

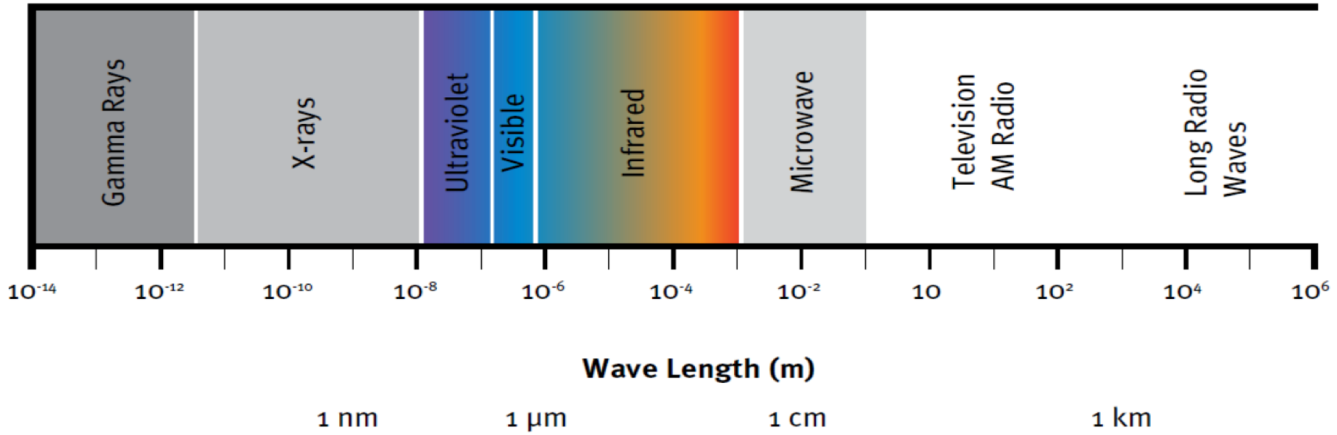
- Release 1972 Landsat (1990)
- Worldview
- Leaner satellites ~5,000 lb to ~10 lb
- Improved frequency
- But resolution compromised (0.5m – 3m)

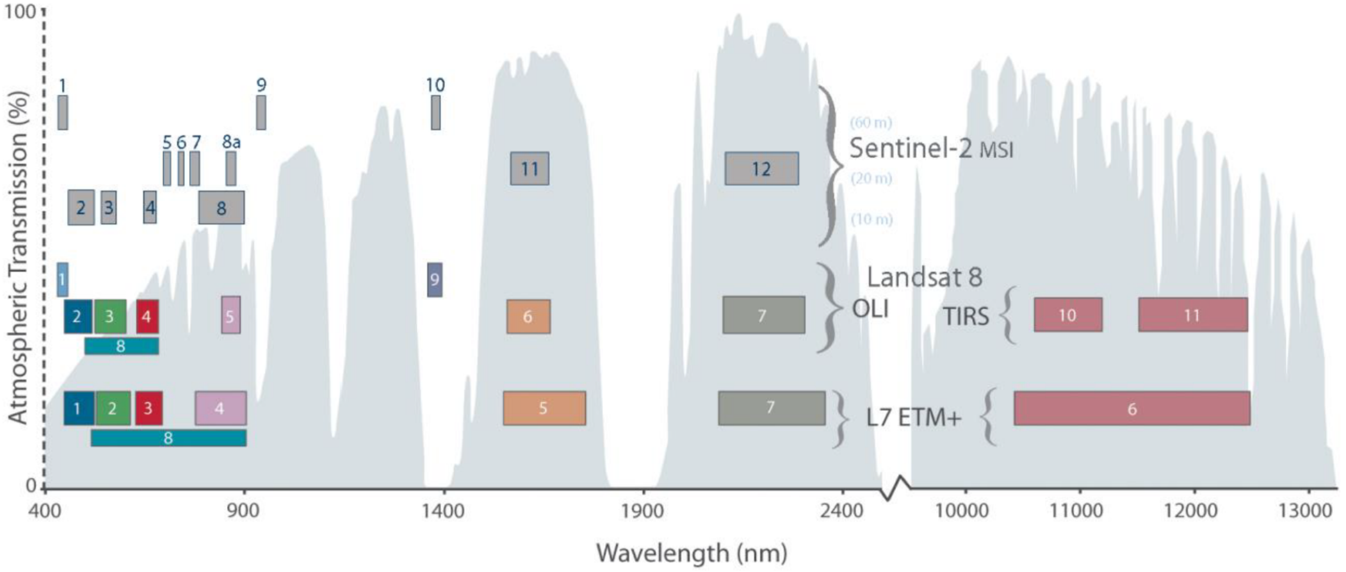
World War: Photometry/ Connect the pixels?

Source: BBC

Source: Eurosense

Deconstructing imagery to data: pixel



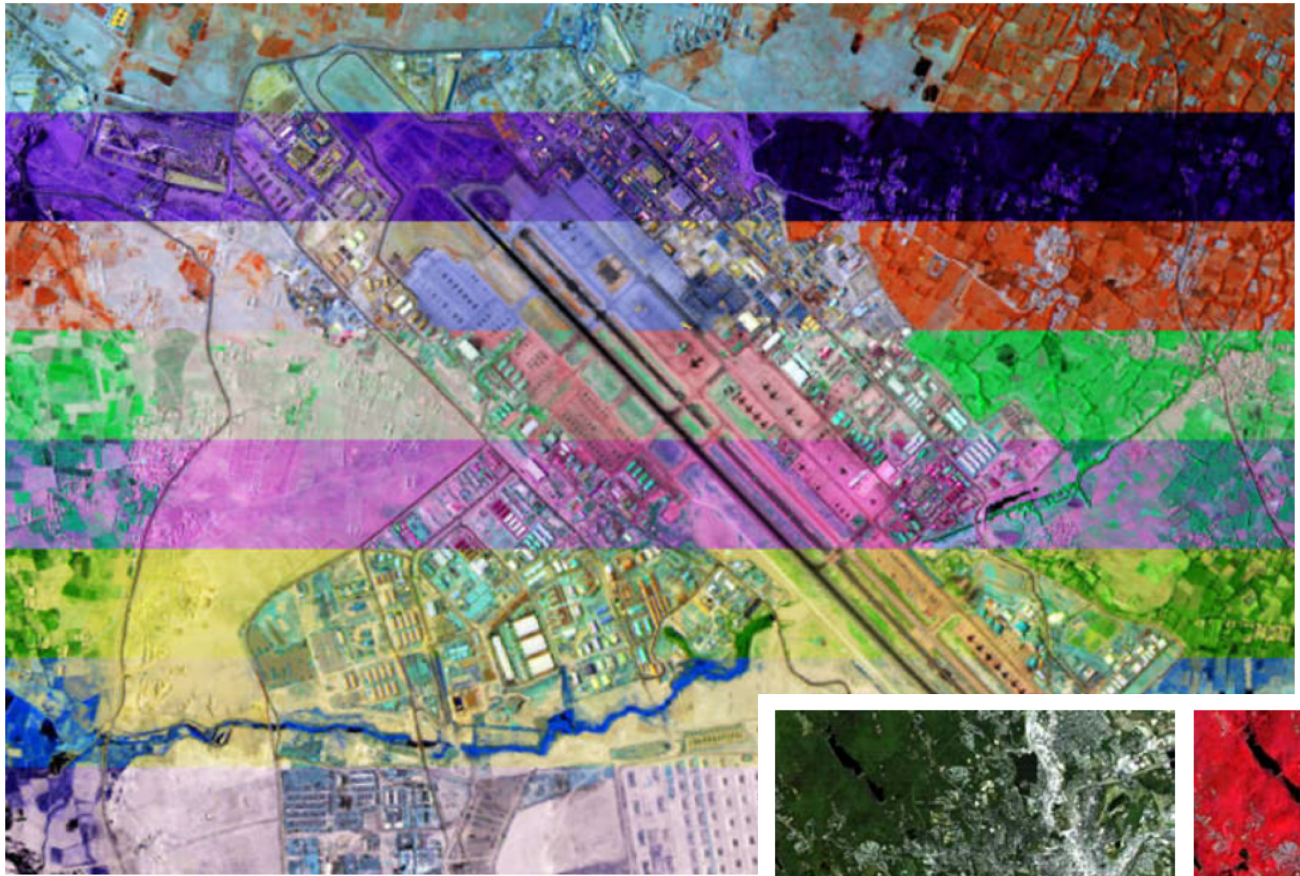


Source: NASA, LANDSAT

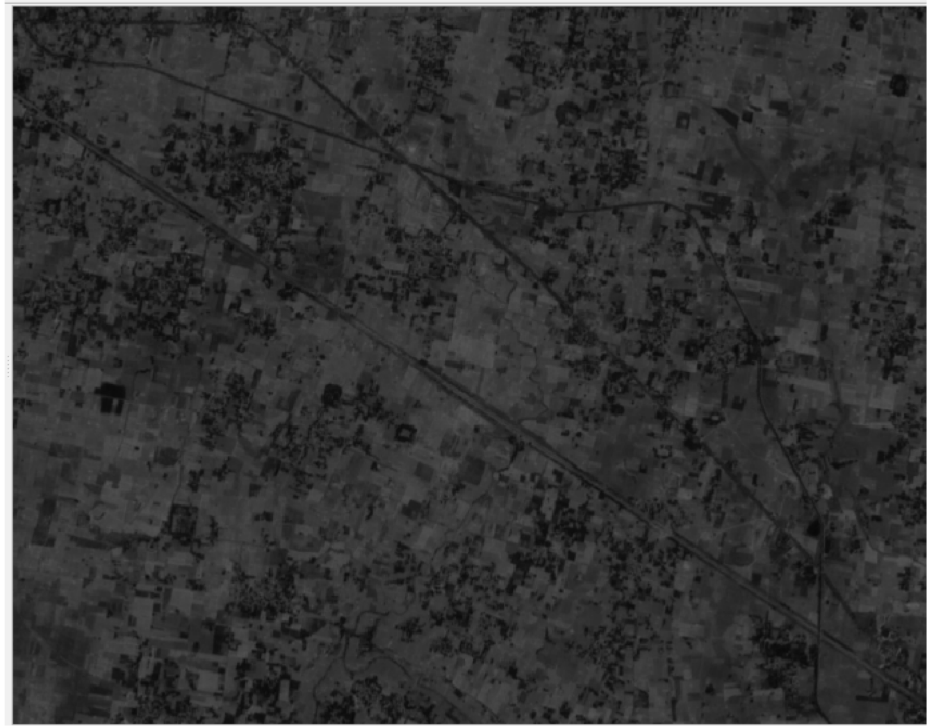
General | Projection | Histogram | Pixel data

Row	0	1	2	3	4	5	6
0	010.408	011.313	011.313	010.106	010.709	010.860	01
1	010.106	010.408	010.559	009.654	011.162	011.313	01
2	010.106	009.050	009.503	009.654	010.408	010.408	01
3	009.654	009.654	009.503	009.654	010.408	011.313	01
4	010.257	009.955	011.162	011.011	011.011	010.559	01
5	009.804	010.257	010.860	010.408	011.463	010.559	01
6	009.804	009.654	010.559	010.559	010.860	010.559	01
7	010.106	010.408	010.106	010.559	011.463	010.106	01
8	011.162	009.955	010.408	010.860	010.559	010.860	01
9	010.709	010.709	010.408	011.313	011.313	010.408	01
10	010.408	010.860	010.408	010.257	010.709	010.559	01
11	010.860	010.559	010.257	010.709	010.709	011.011	01
12	011.313	010.106	010.559	010.559	010.860	010.860	01
13	011.313	010.559	010.559	010.709	010.559	010.709	01
14	011.162	010.709	011.011	011.011	011.463	010.408	01
15	011.463	011.313	011.011	011.162	010.408	010.709	01
16	011.463	011.313	010.860	010.408	009.804	010.709	01
17	011.614	011.313	010.559	009.654	010.408	011.011	01
18	011.614	010.106	011.011	010.860	009.955	010.408	01
19	011.162	011.313	011.463	010.257	010.559	010.408	01
20	011.313	011.313	010.408	010.709	010.106	010.106	00
21	010.559	010.106	010.106	011.162	011.162	009.955	00
22	010.709	010.408	010.106	010.860	010.559	010.709	01

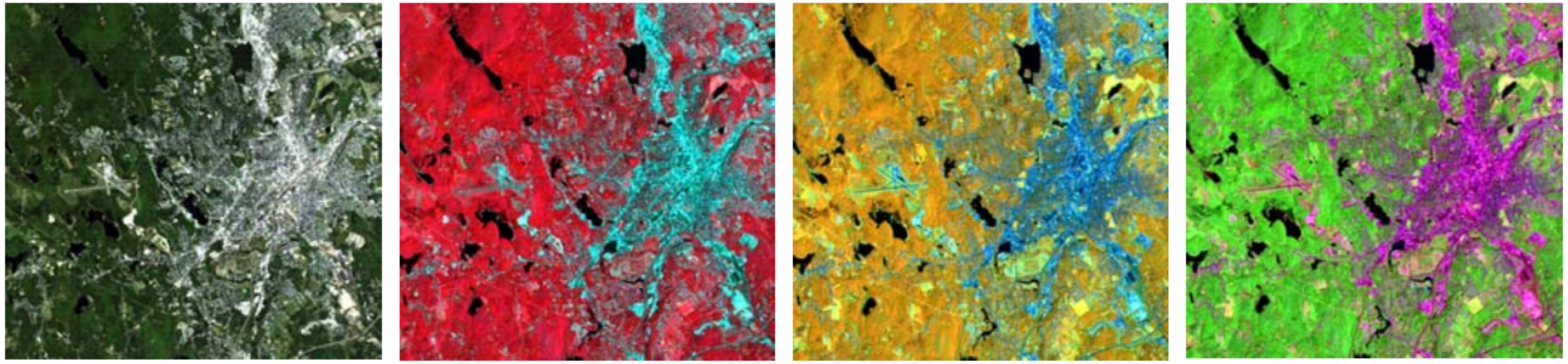
Spectral Bands



Source: DigitalGlobe



Azamgarh Red Band



RGB=bands 3,2,1 RGB=bands 4,3,2 RGB=bands 4,5,3 RGB=bands 7,4,2

Cleaning & Preprocessing

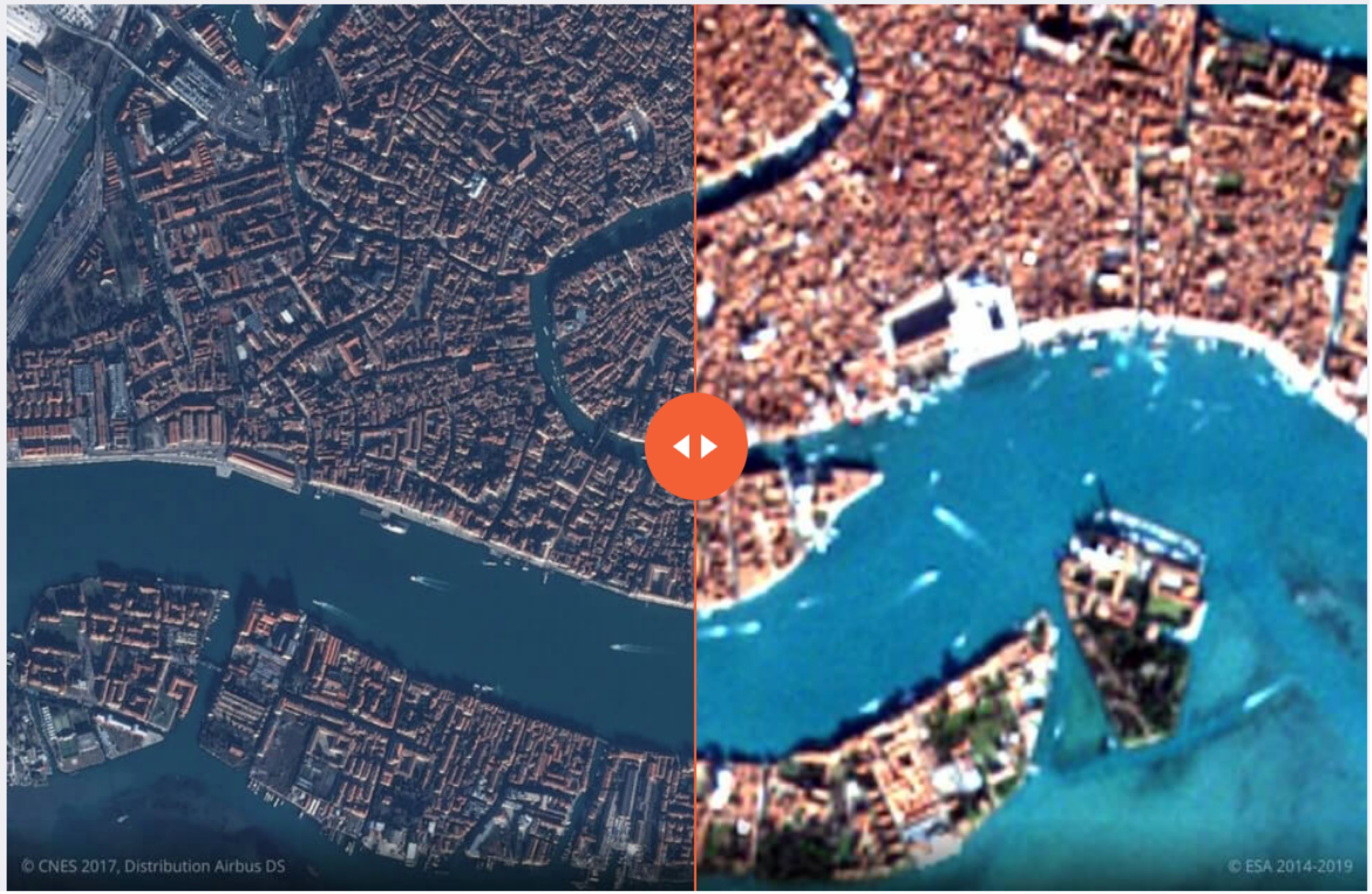
IMAGE BEFORE ACOMP PROCESS

IMAGE AFTER ACOMP PROCESS

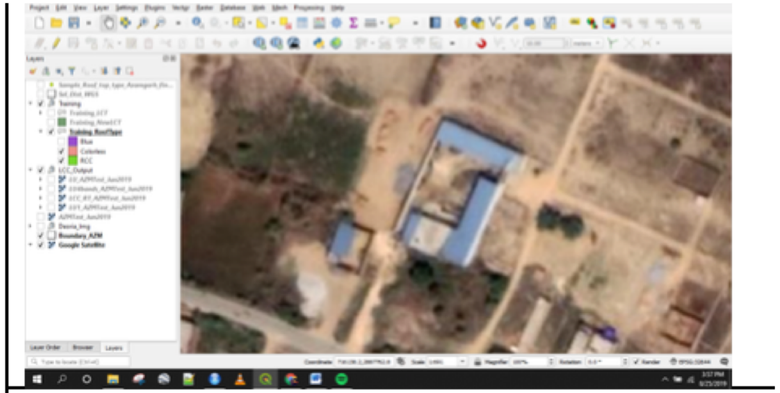
Source: Digital Globe

- Atmospheric Correction
- Radiometric calibration
- Further: SAR Lidar, DTM

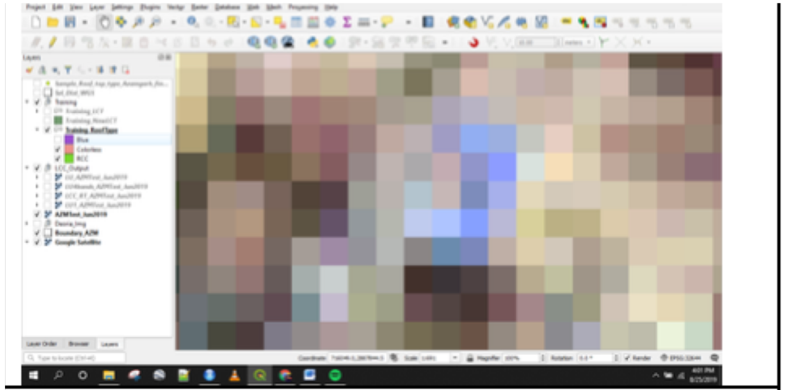
Use Cases & Resolution



Source: EOS



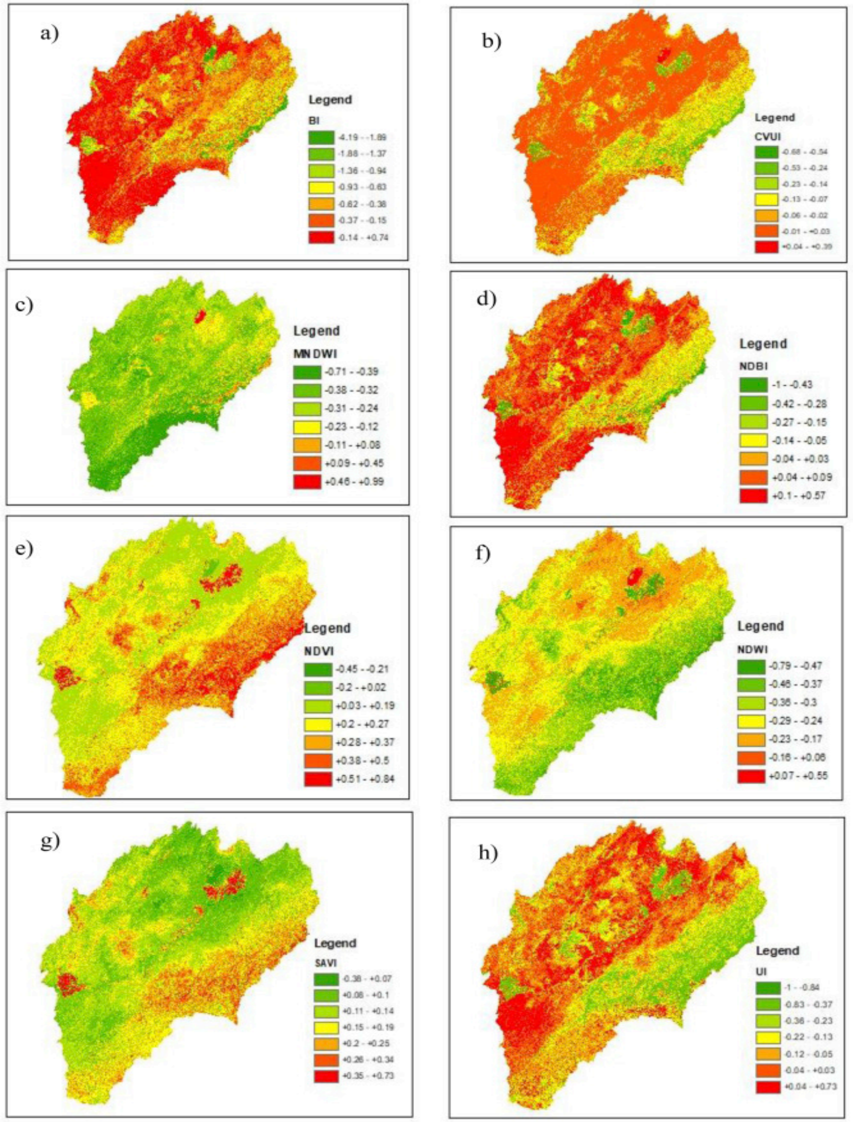
Google Sat Imagery



Sentinel-2 Image in True Color (RGB)

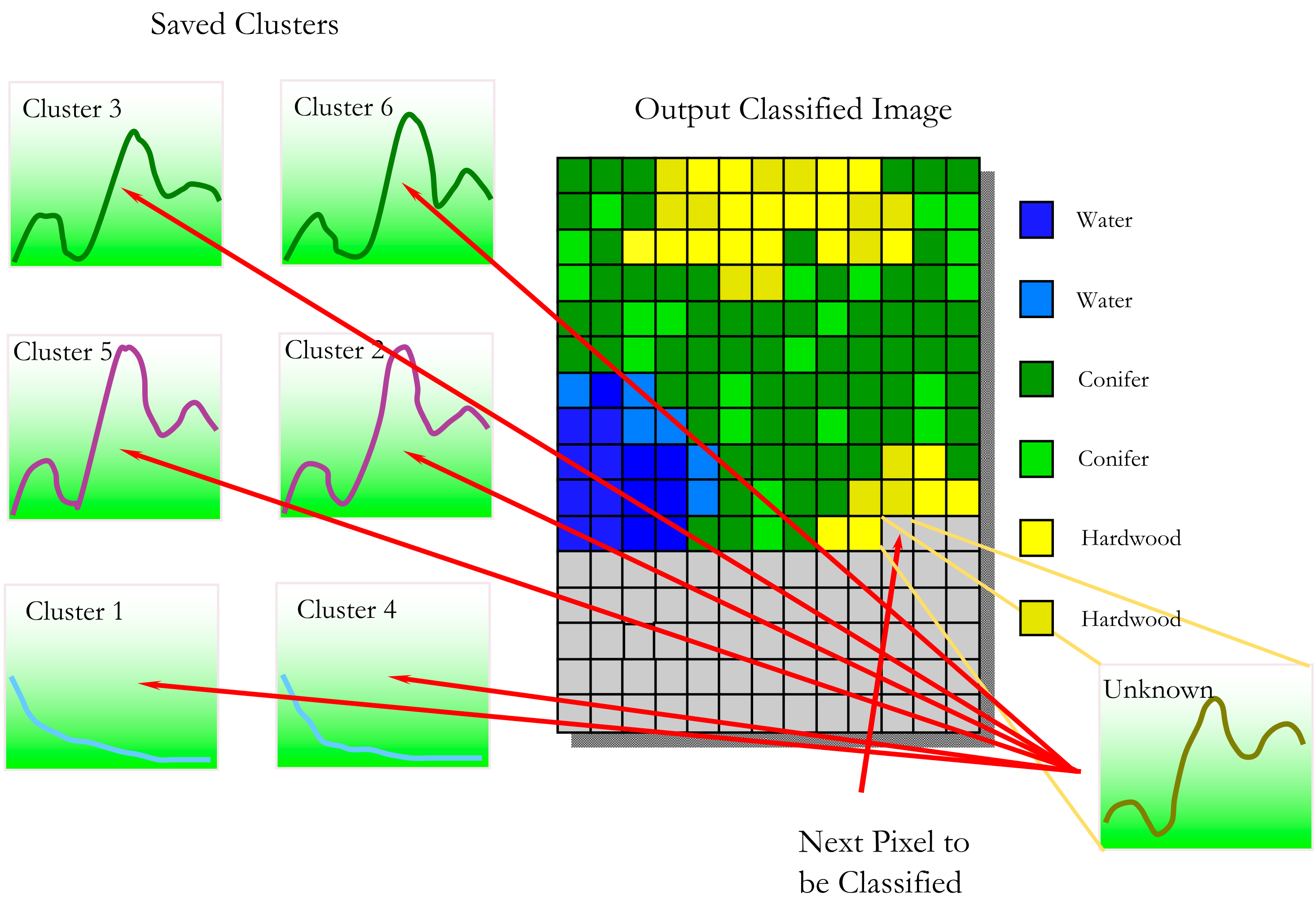
- Nature of task
- Resolution, Bands & Pricing
- Pan sharpening/SRM (R&D)
- Free resources: end

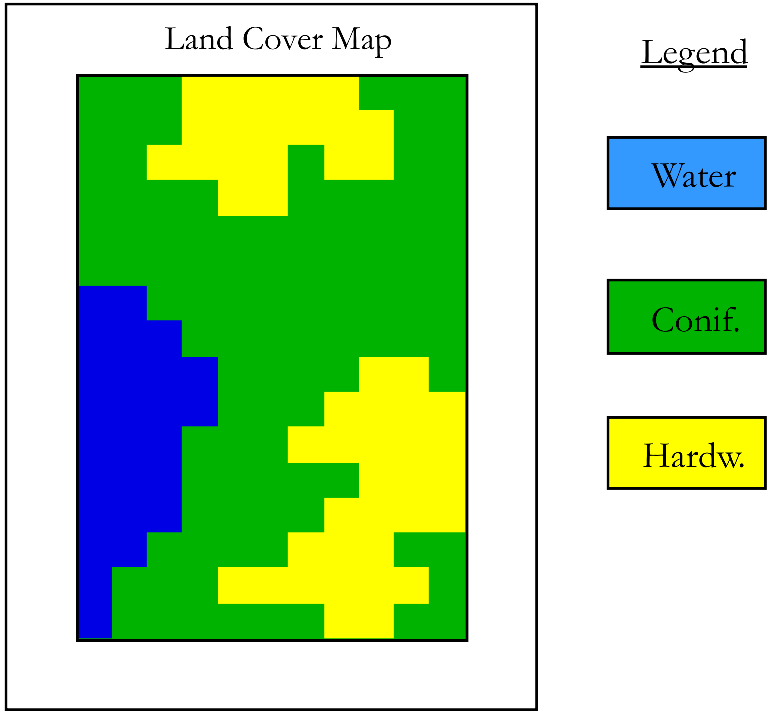
Appropriation of Remote Sensing Data Science



- Conventional LUC
- Lillesand and Kefer 1987
- Lower Resolution
- Non parametric methods and distribution

Unsupervised & Supervised





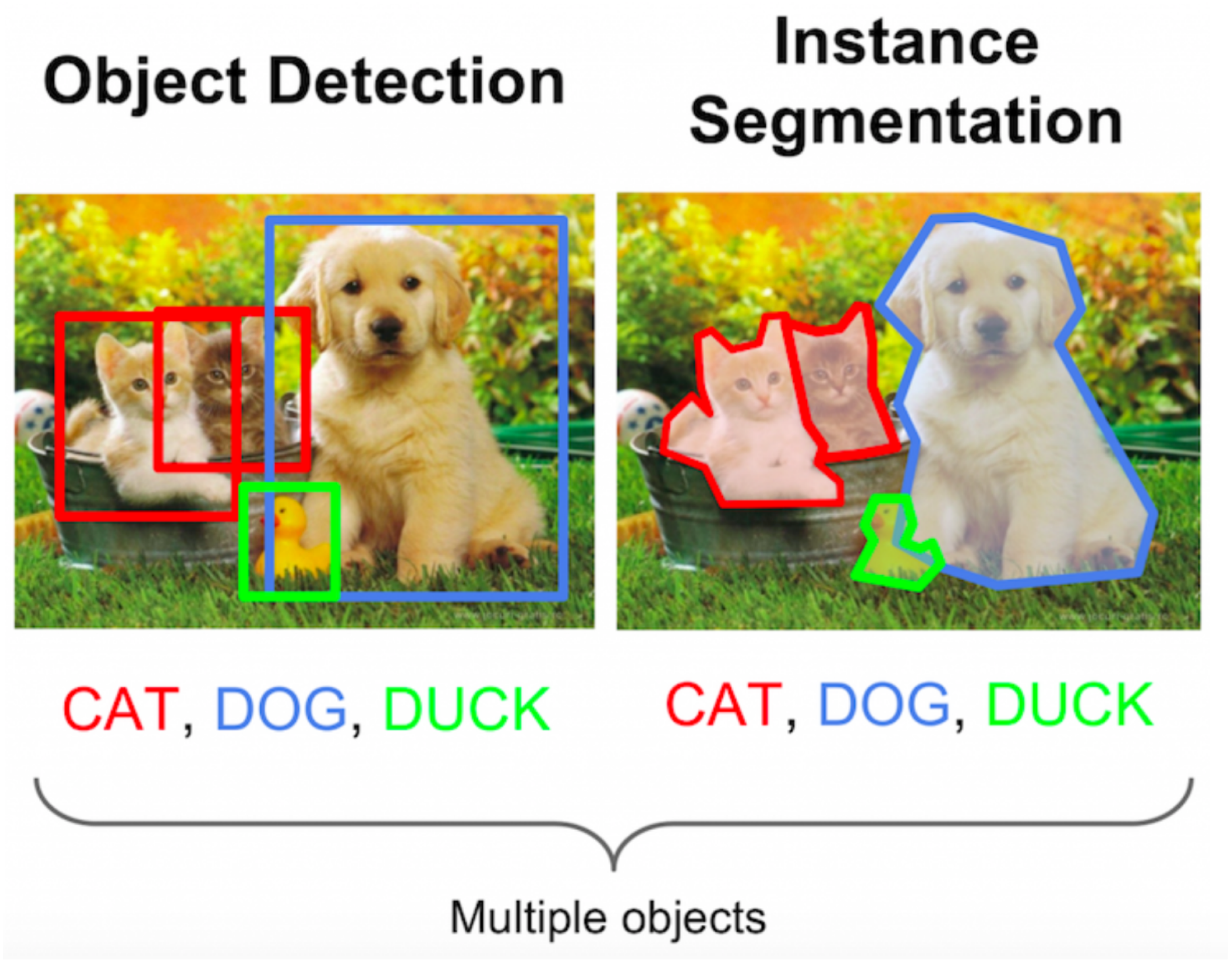
- Evolution Stages
  - Stage I: NDBI to Supervised
    - Training costly
  - Stage II: Parallel (Unsupervised)
  - Stage III: CNN

The Bank of Now
IDFC BANK

THE UNIVERSITY OF CHICAGO

Computer Vision: 101

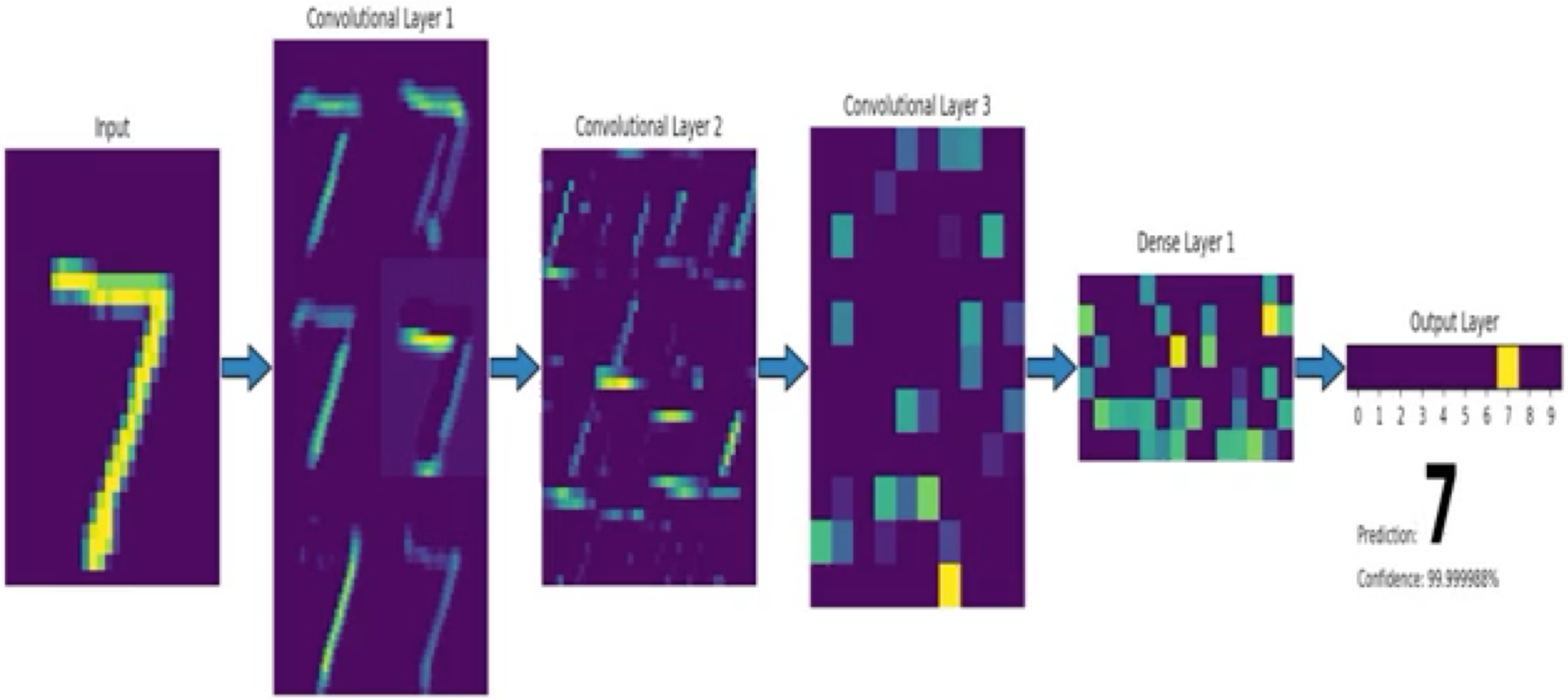
- Classification is passe
- Classification models
  - Identifying/filtering object of interest
- Detection models
  - Bounding box over specified objects
- Segmentation models
  - Exact outline of the object



Source: Clerify

Deep Learning: pixel is dead

peek into CNN black-box



Source: KDNuggets

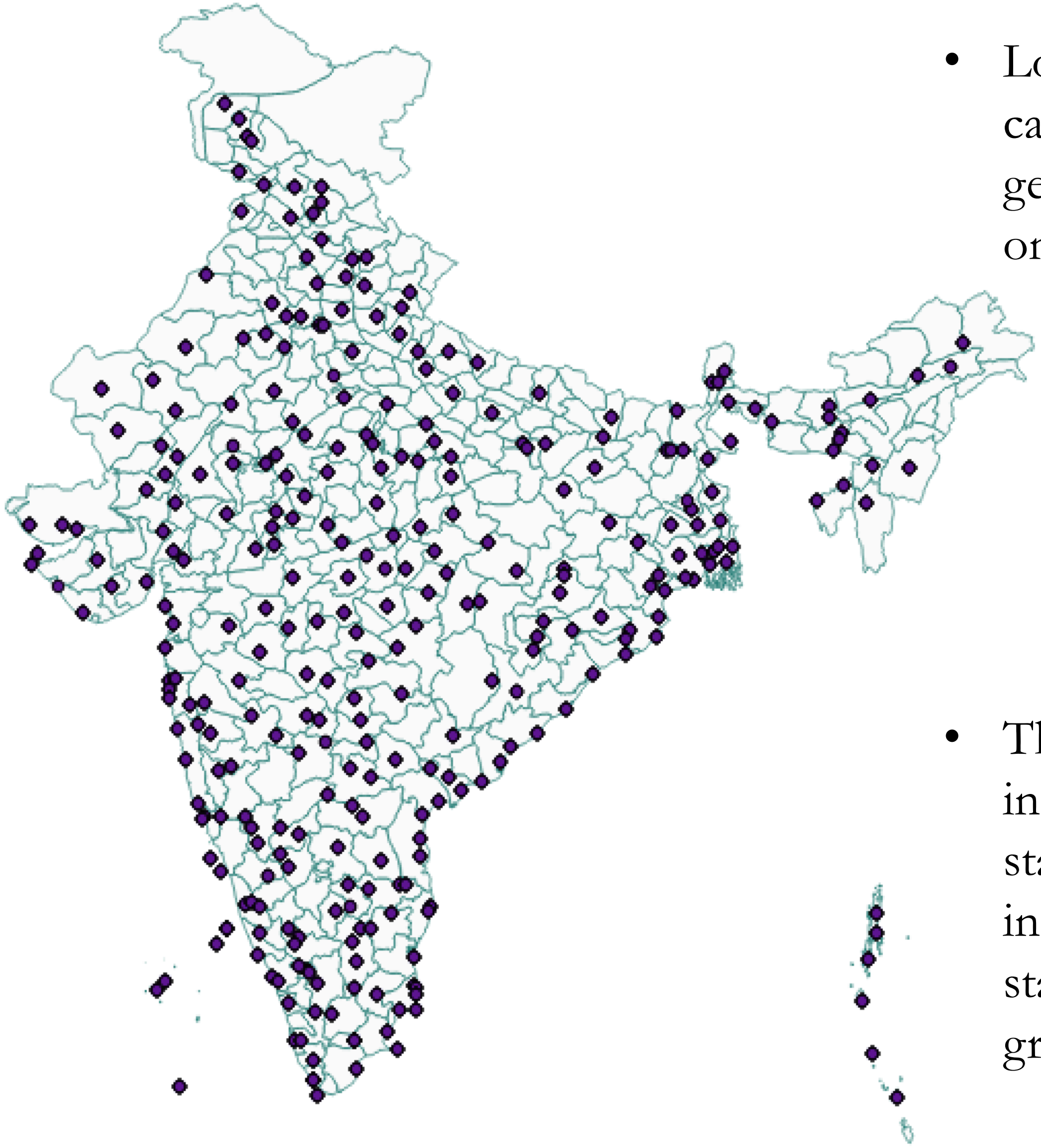
So what is spatial data?

- Raster: (*.tiff, *.png)
  - Ready to use: Nightlights, Aerosol Optical Depth (AOD: ppm)
  - Extraction

- Vector: (*.shp, *.geojson)
  - Shapefiles: points, line, polygons

- Tools:
  - Operational: ArcGis, Qgis
  - Programming: Python/R
    - R > : raster ease
    - Python >: scaling/automation
- Cloud: GPU's/Tensorflow, Keras,

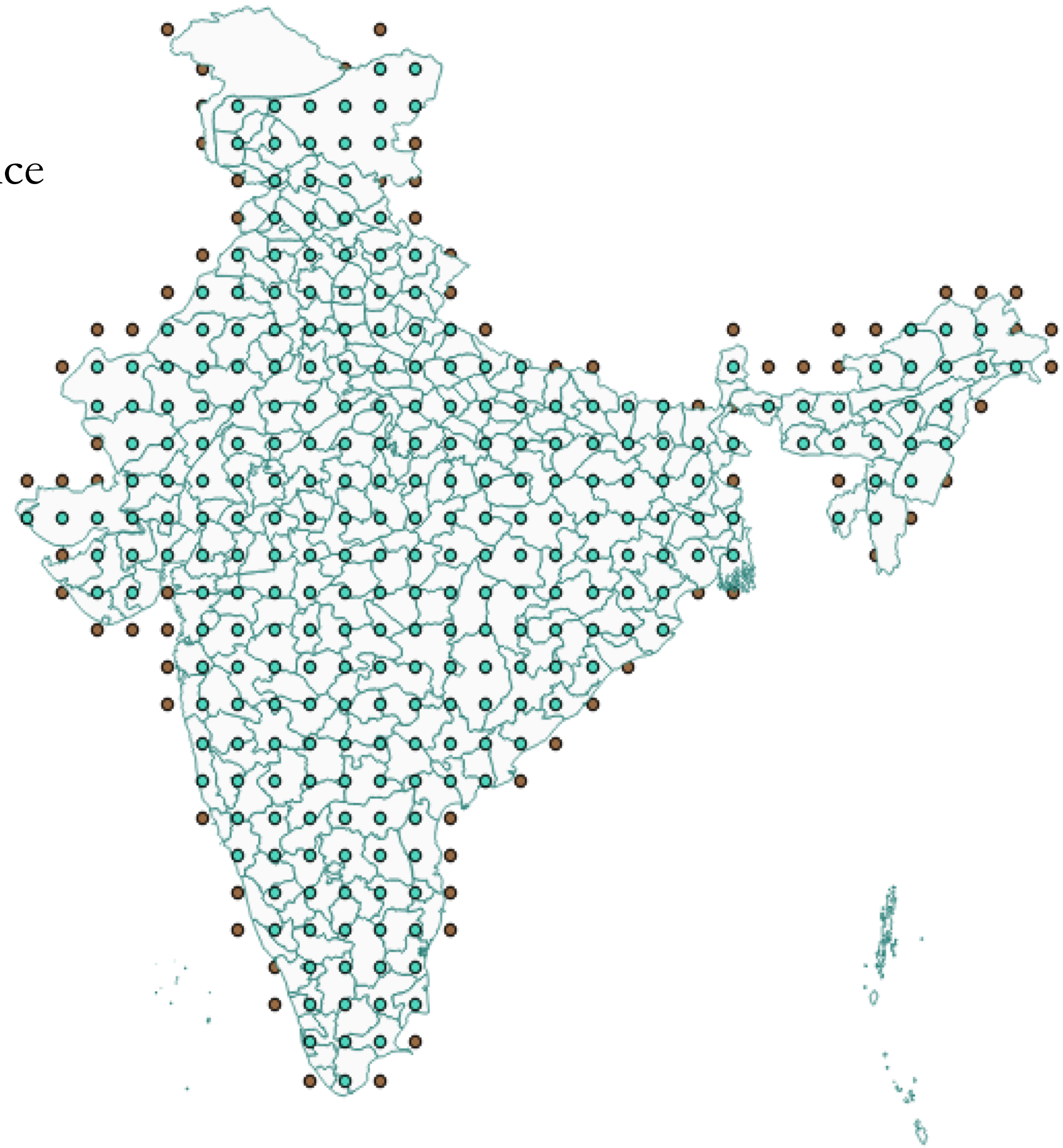
Power of location in public analytics



Source: parthkhare.github.io

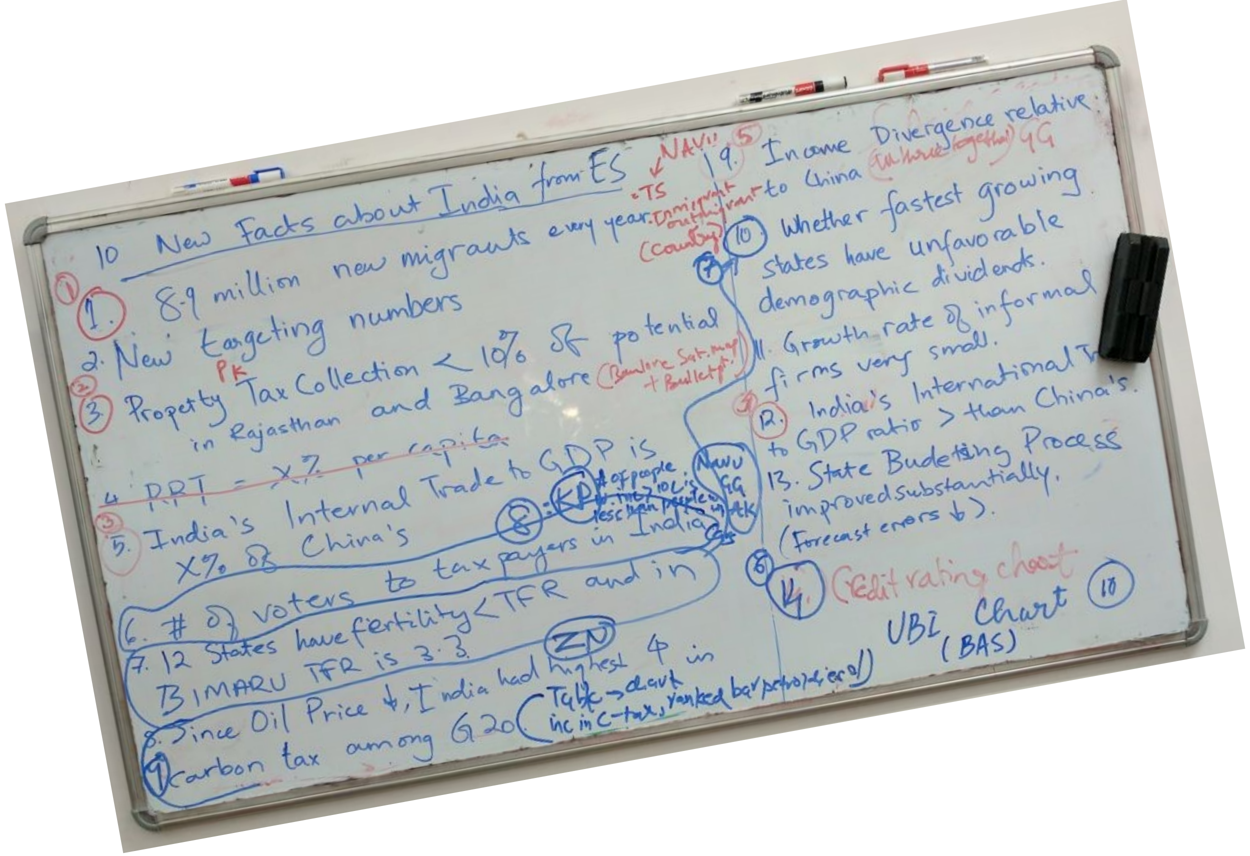
- Location of station can bias their geographical influence on crops

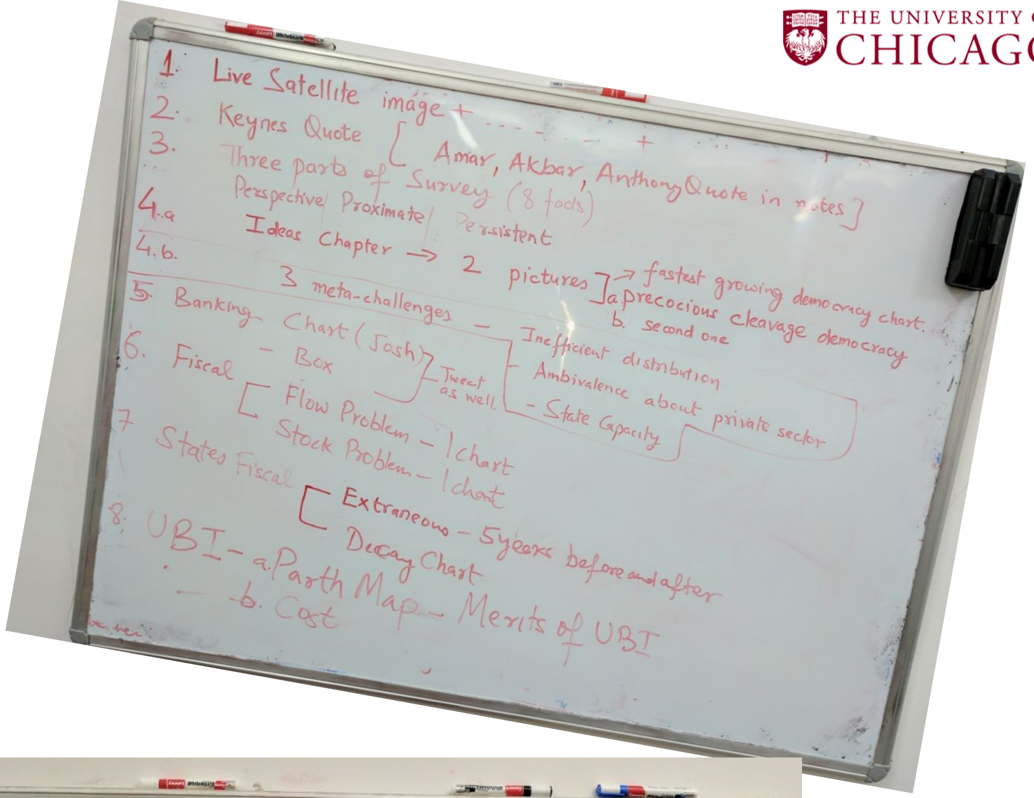
- Therefore, after increasing the stations, we spatially interpolated the stations on an even grid

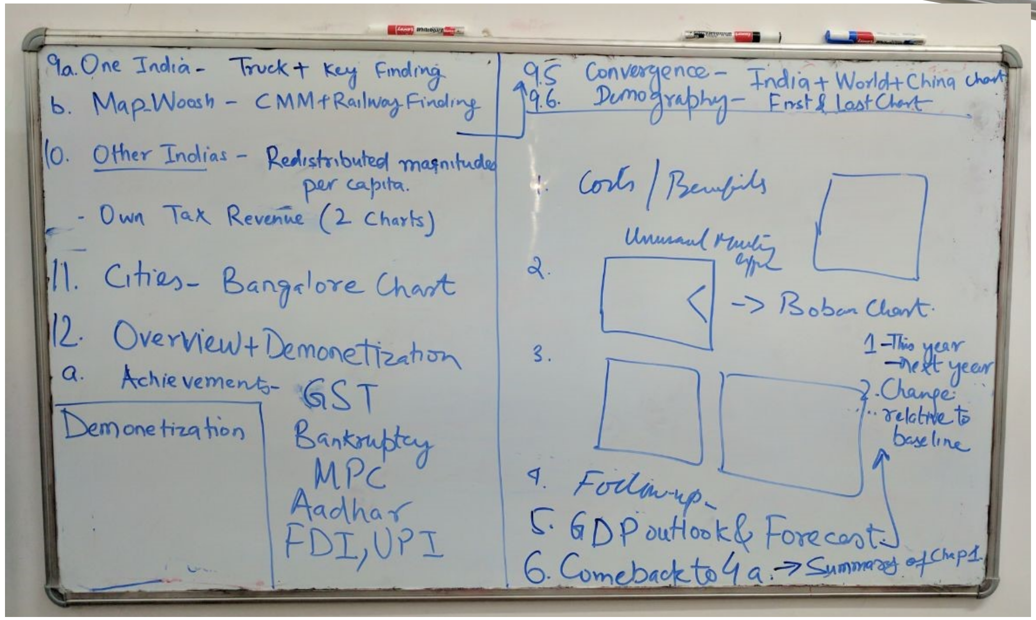


Section II. A: Public Policy
Deployed Applications at MoF

Deconstructing Economic Survey



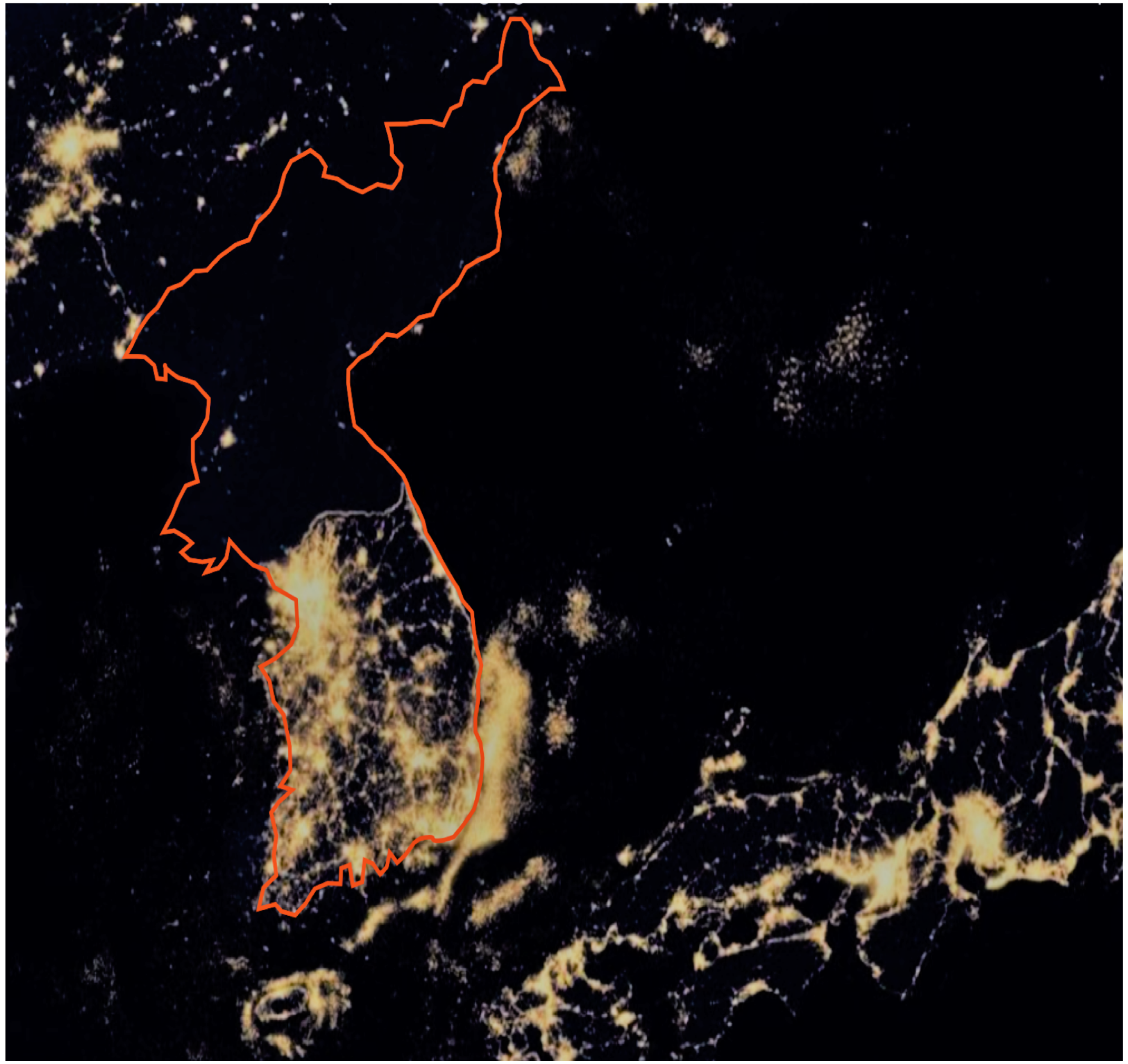




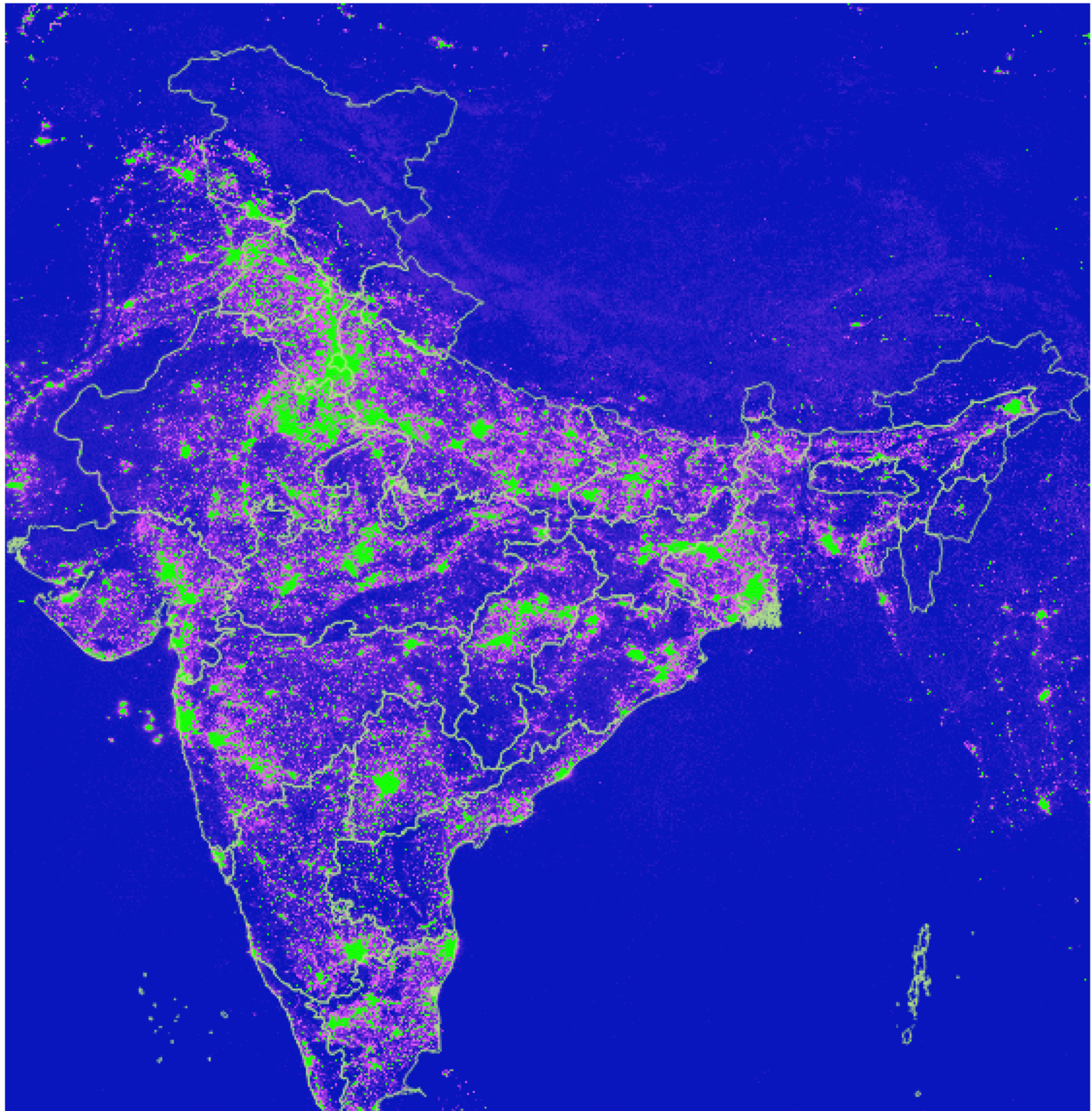
Policy making & Data: Economic Survey

- *New Data:* Satellite Imagery
  - Realizing Fiscal Space: Property Tax
  - Implications CSS: Urbanisation Rate {migration/flux}
- *Improving Data Capacity:* Ground Stations: Weather
  - Ramifications for IPCC: RCP protocol
- *Out of Box use*: Passenger
  - Dynamic assessment

Visualizing Data: Impact & Audience

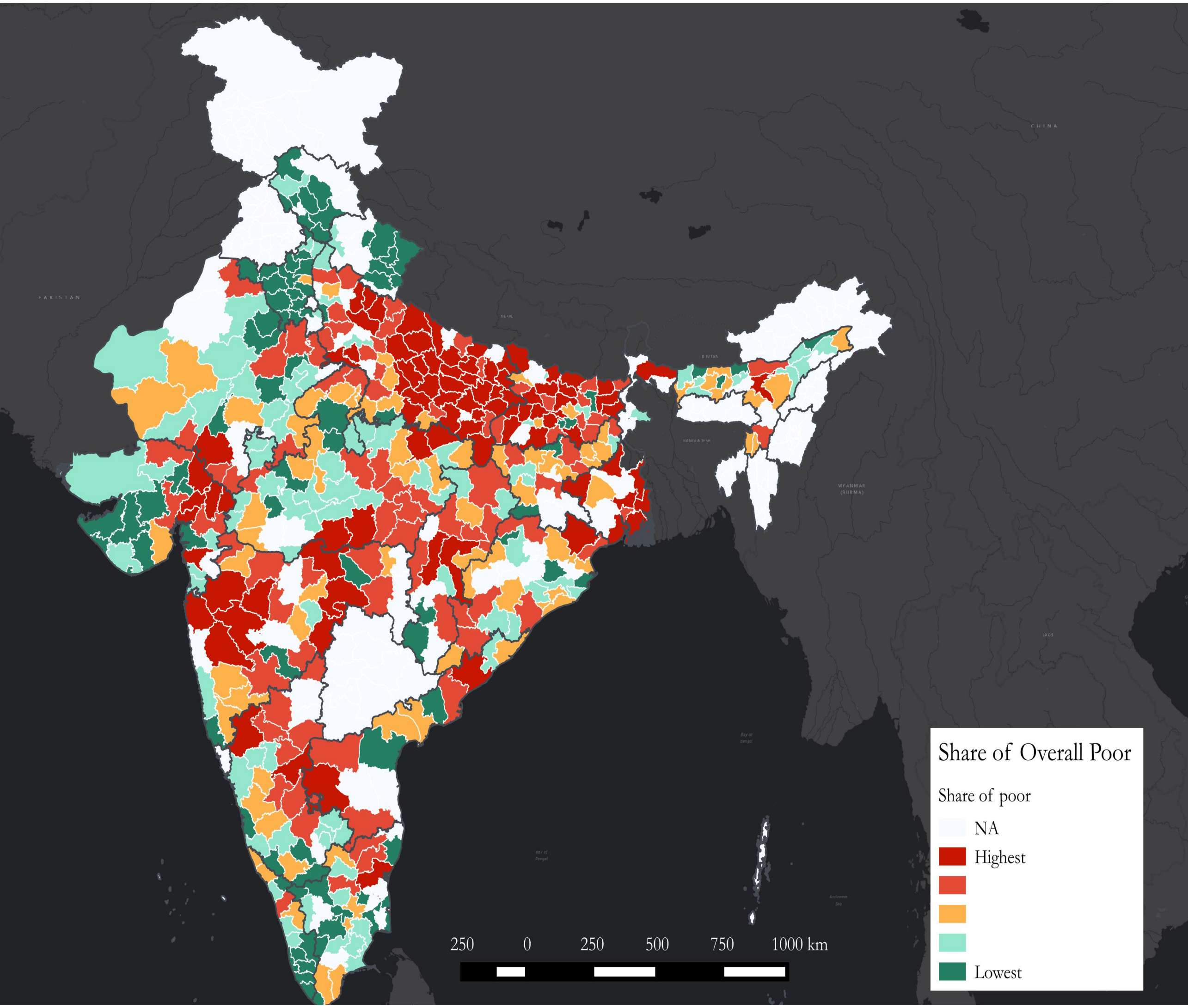


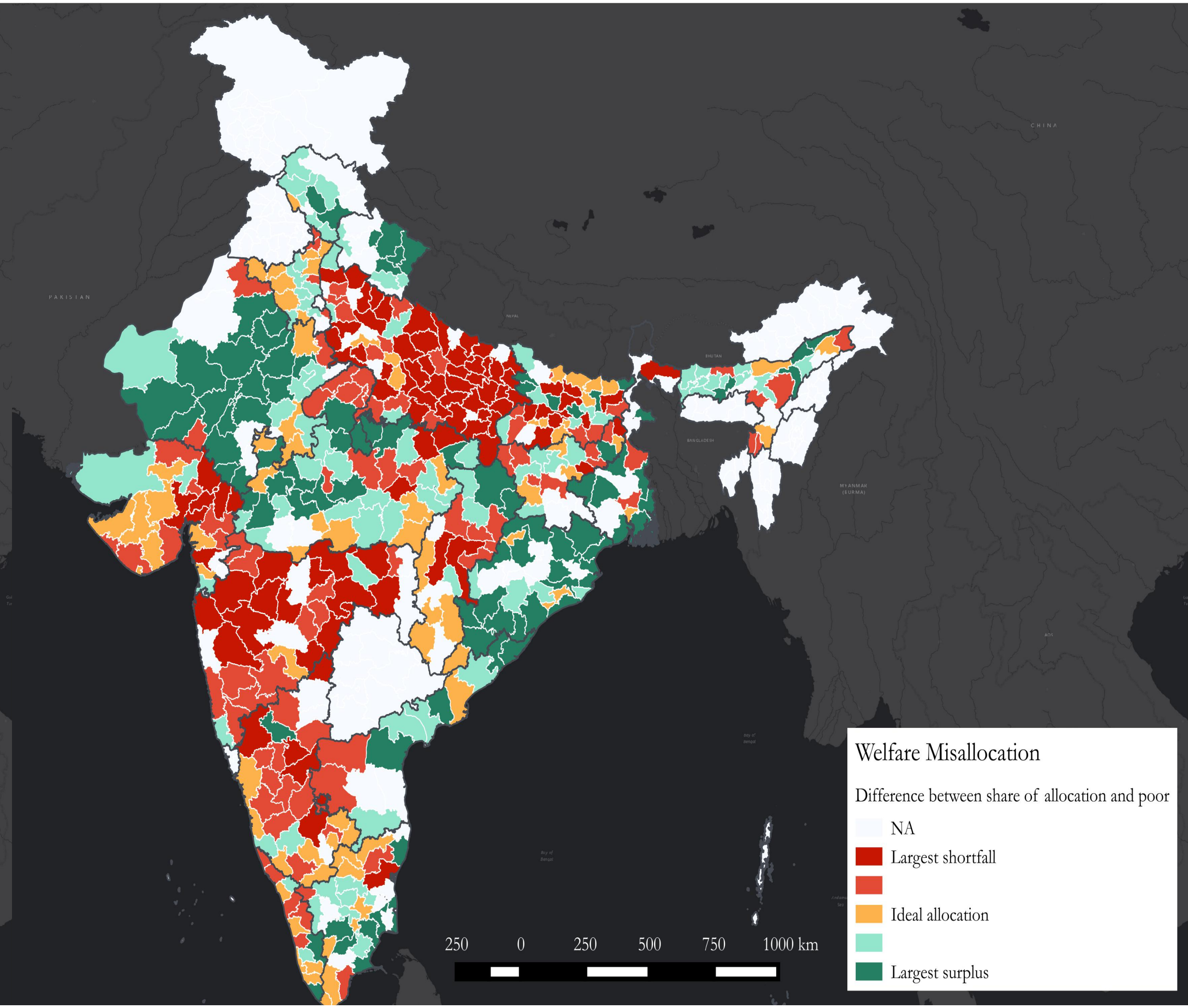
Source: NOAA



Source: parthkhare.github.io

Where are the poor & what are the districts of misallocation





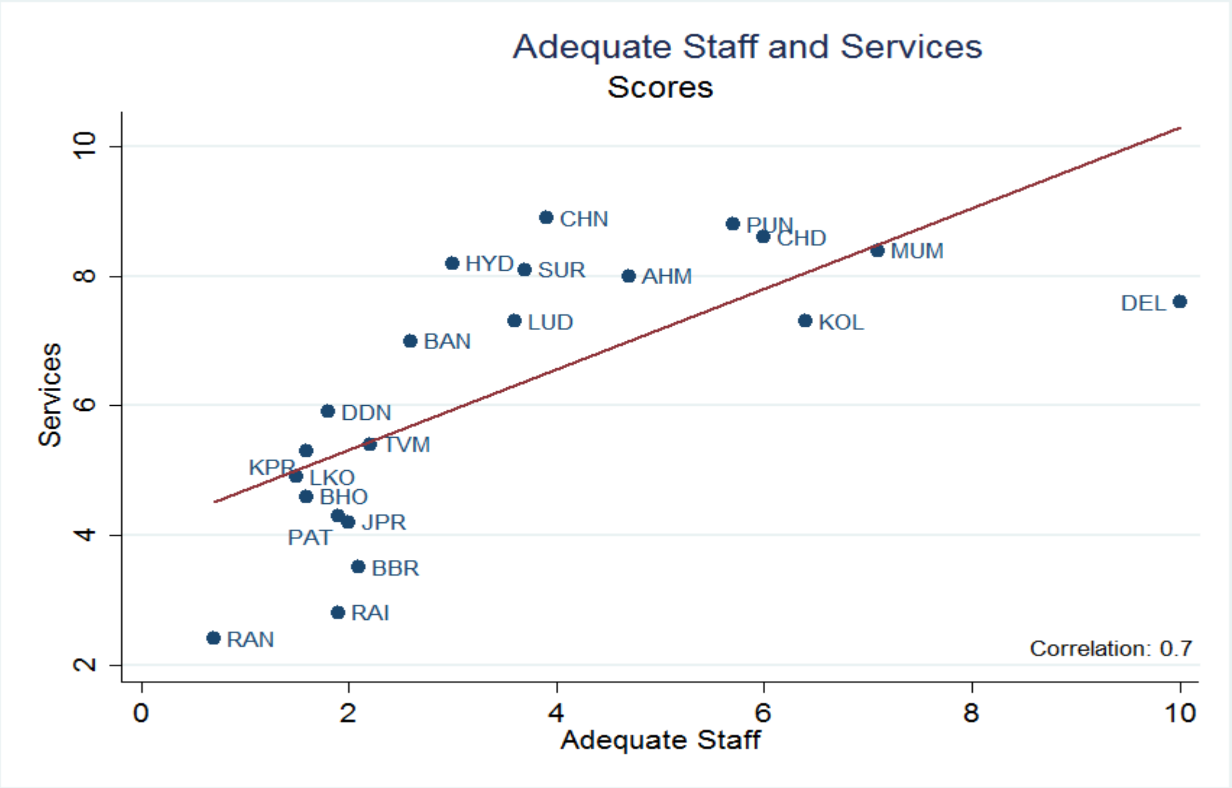
Source: Economic Survey 2017

Section I: Satellite Data

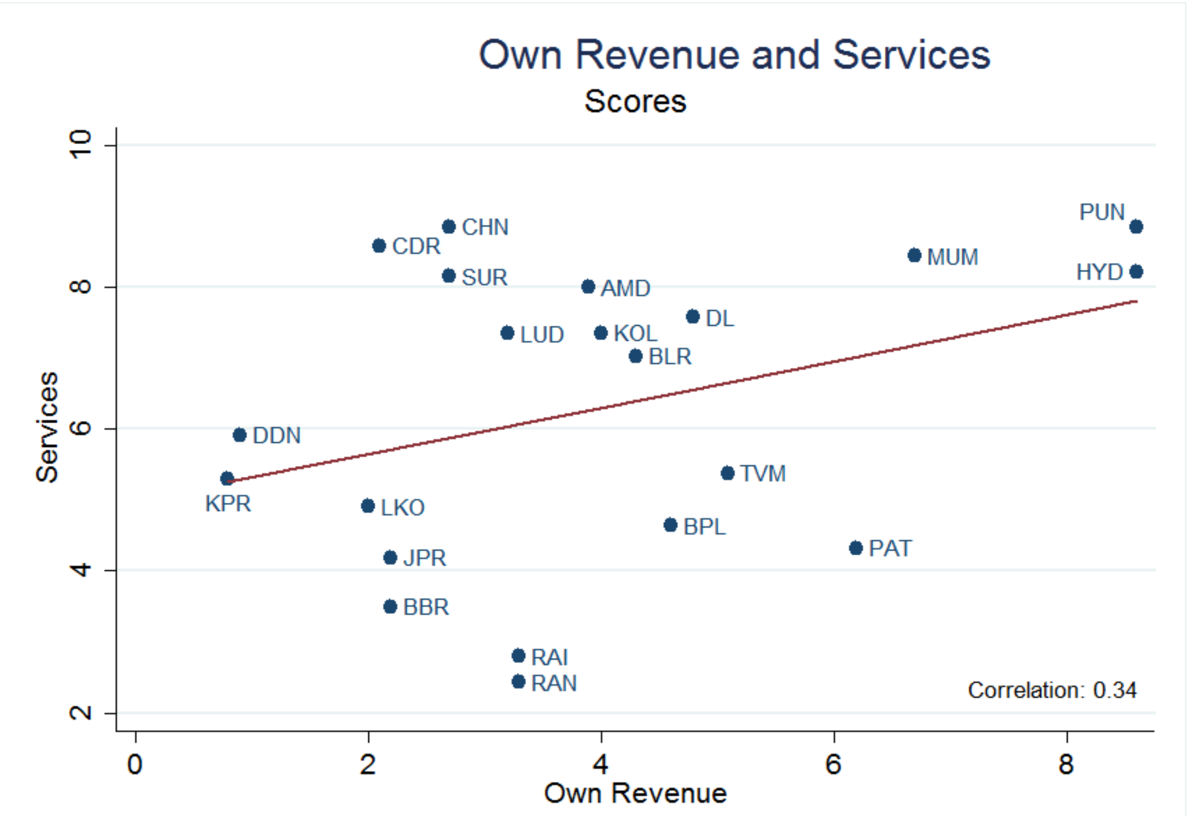
- Free from human error
- Free from intentional field survey biases
- Dynamically updated
  - Census estimates in 10 years, thereby
  - Fails to capture the spurts and flux through demography mobilization
- Better assessment of Social Sector Funding/Grants to States, ULB's, Panchayati Raj
  - Helps planned expenditure and revenue collection

=> *BMC's implementation, 15th Finance Commission proposal forest detection,*

Property Tax I: Motivation [Revenue Capacity]



- Cities do not maximize their tax potential
- Power to impose more taxes does not imply more revenue



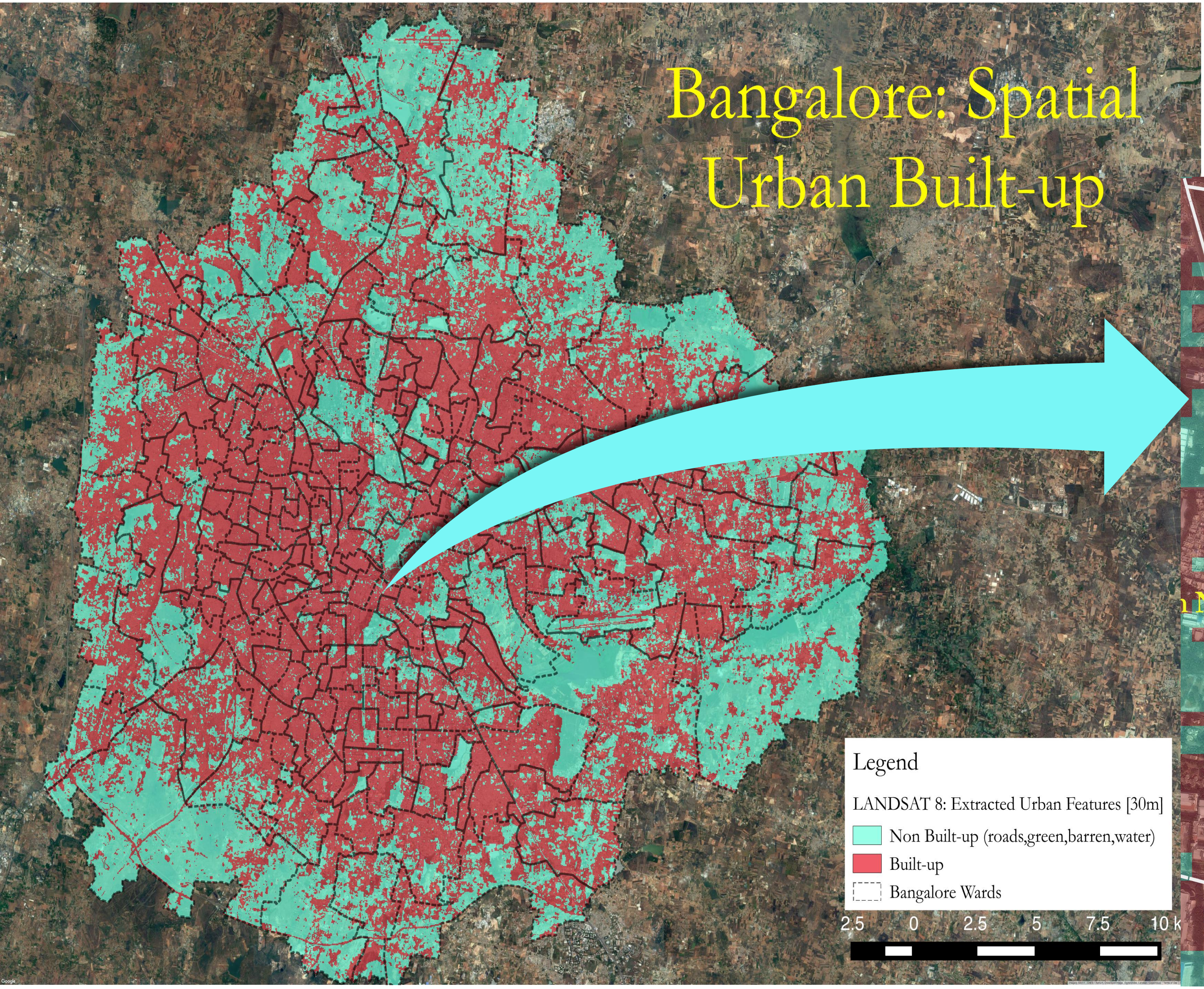
Source: Annual Survey of Cities Janagraha - 2016

- Efficiency from Self Financing: Cities that generate their revenue tend to provide better services
- Driven by
  - Size of the tax base,
  - Efficiency in tax collection and
  - Level of economic activity in the city area

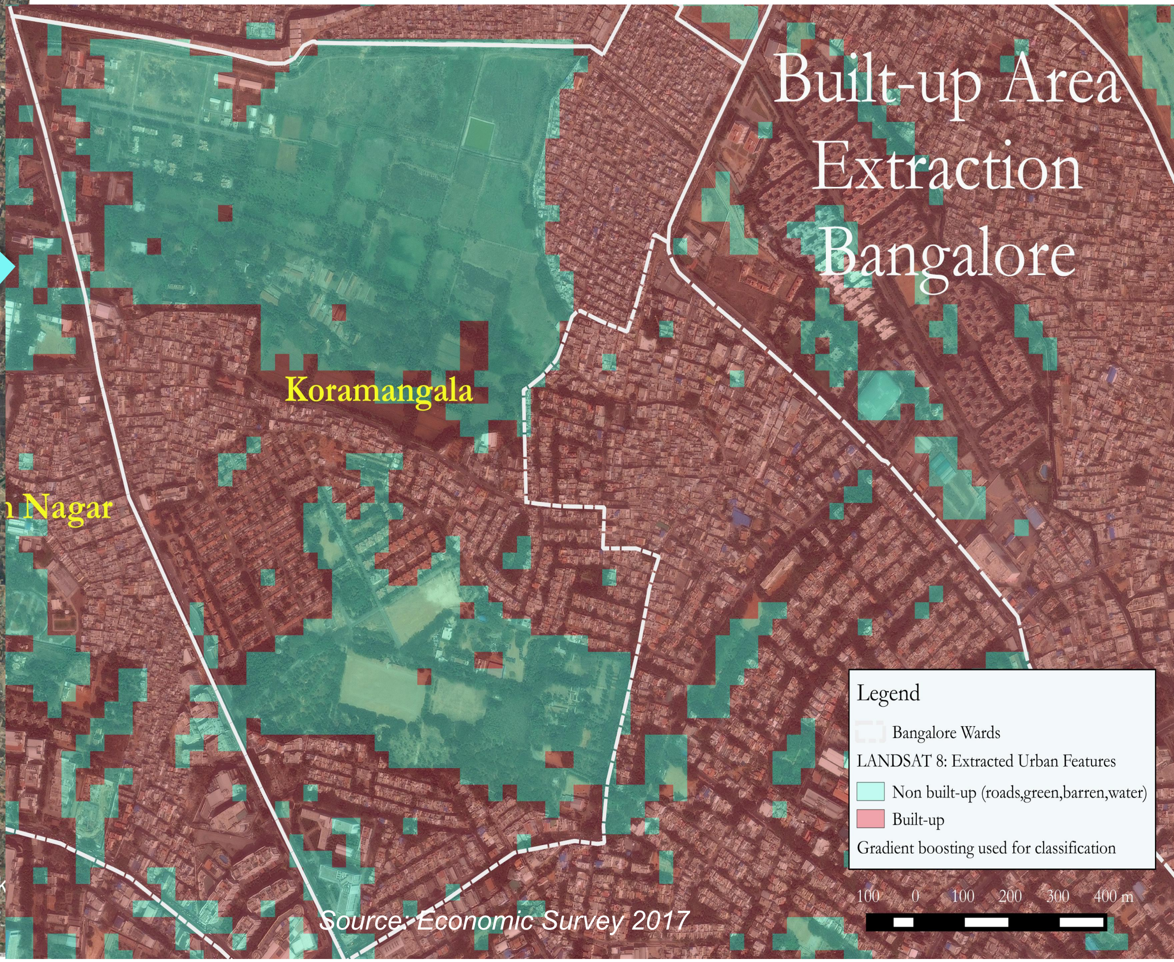
The Bank of Now IDFC BANK

THE UNIVERSITY OF CHICAGO

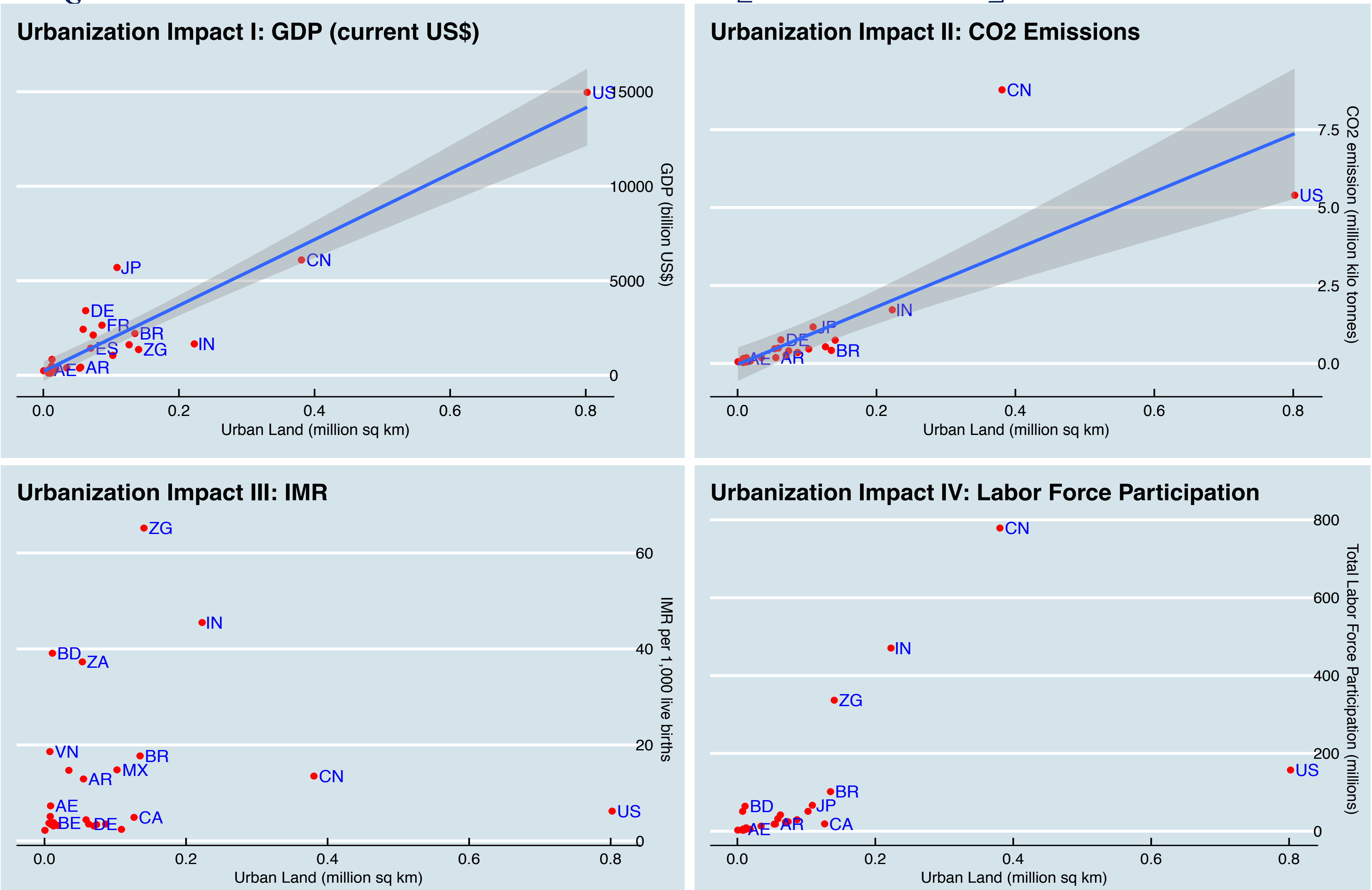
Property Tax II: Open Source Data + Municipality



- Jaipur can collect **more than 90%** of its current property tax
- Bangalore can collect **more than 80%** of its current property tax



Urbanization Rate I: Motivation [Bi-causal]

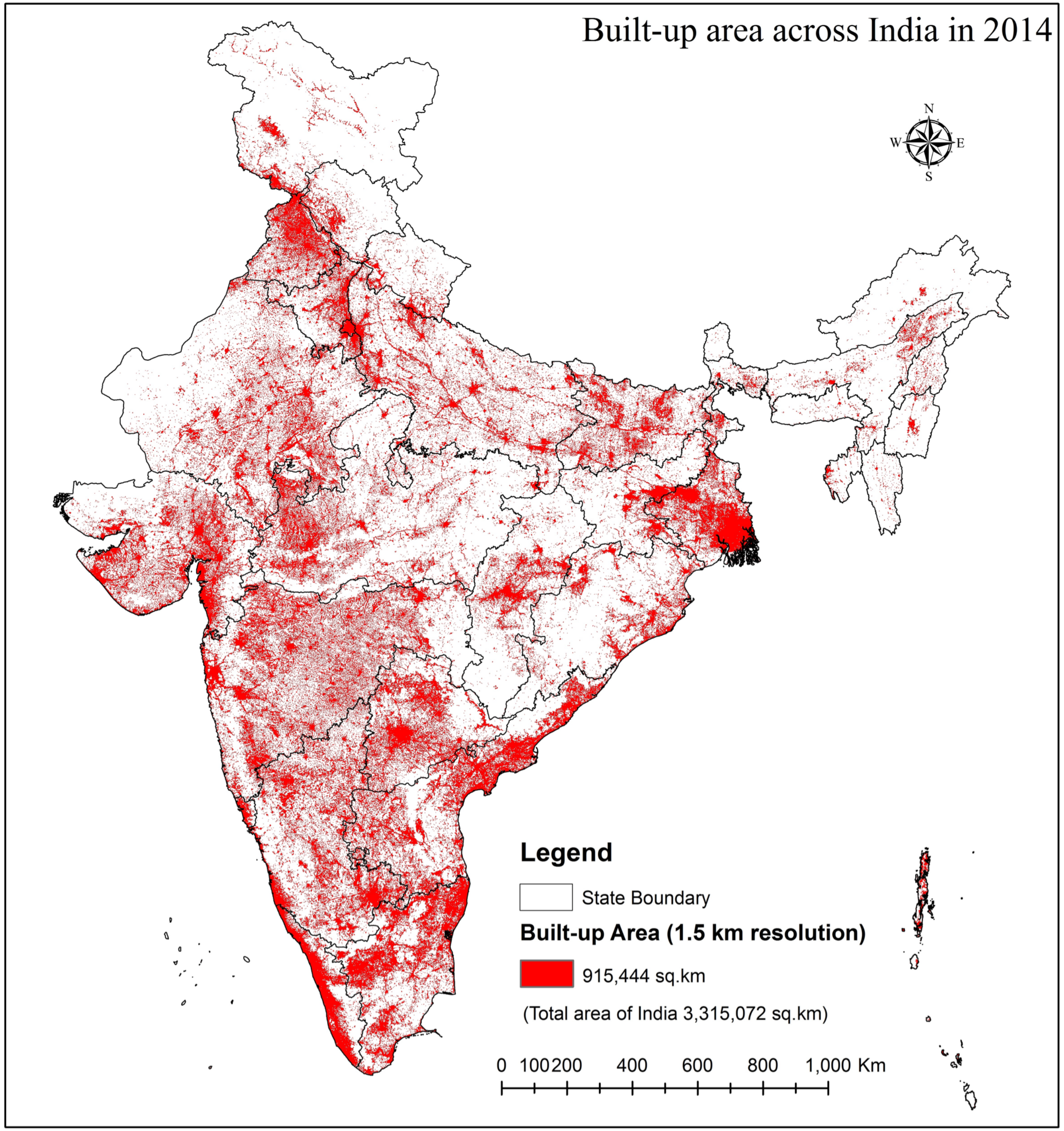


Source: WDI 2010 Data

Urbanization Rate II: Re-define

Census 2011 [31.16%]	Population of at least 5000
	Density of at least 400 persons per sq. km
	75 percent of male working population is engaged in non-agricultural activities

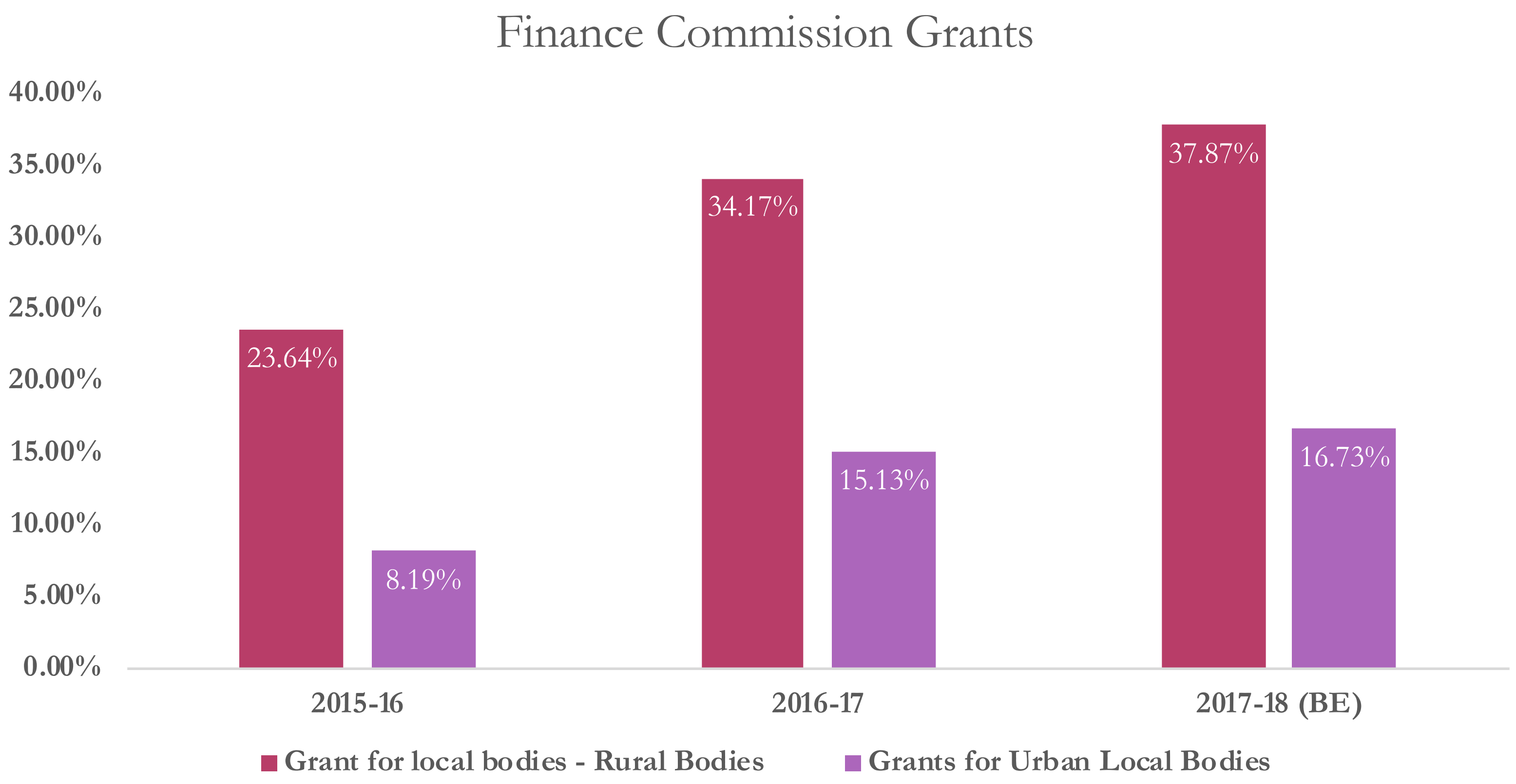
- Need for renewal (Census definition unchanged since 1961)
- Inconsistency among different governmental definitions
- Underestimation of extent of urbanisation (when compared to definitions of other countries)
- Survey based
- Migration Bias
- 27.1% – 31.2% in 10 years with 2 digit growth



Source: Economic Survey 2017 II

Urbanization Rate III: Policy Implication(s)

- Economic growth closely linked to urbanisation
  - North America/ Europe were 50% in 1940s, 80% now
- Cities the gateways to globalization
  - 80% of all FDI into China in 1990 went to top 20 cities

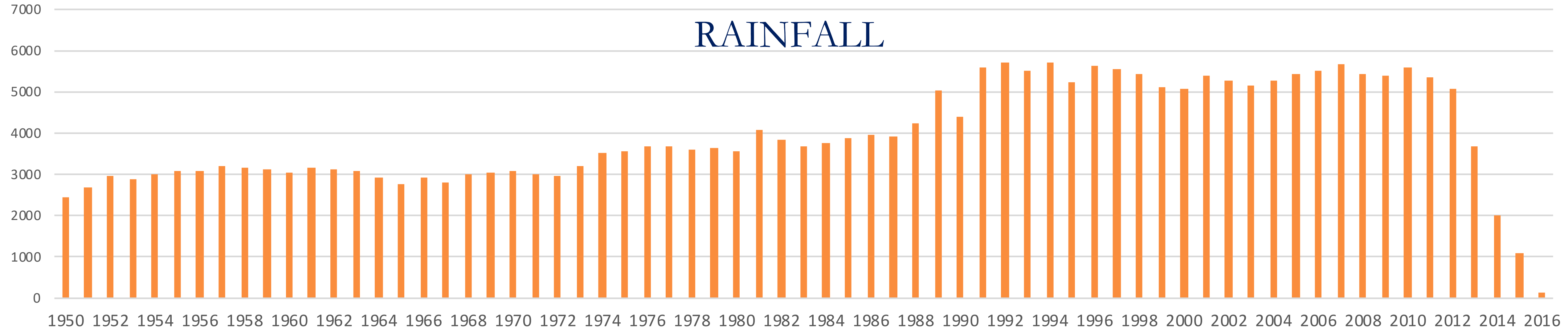


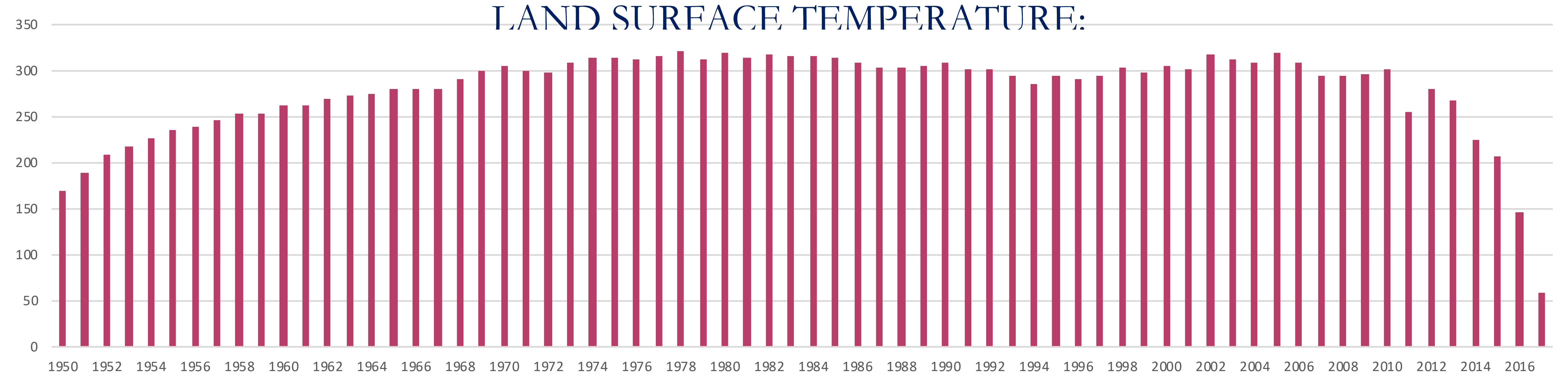
Section II: Climatology Data

- Ground Stations
- Climate Data on Rainfall & Temperature
    - One of earliest sources of recorded data in India
    - Yet observations from all ground stations not leveraged properly
- Divergence with International Results
    - also seen in groundwater/air quality data

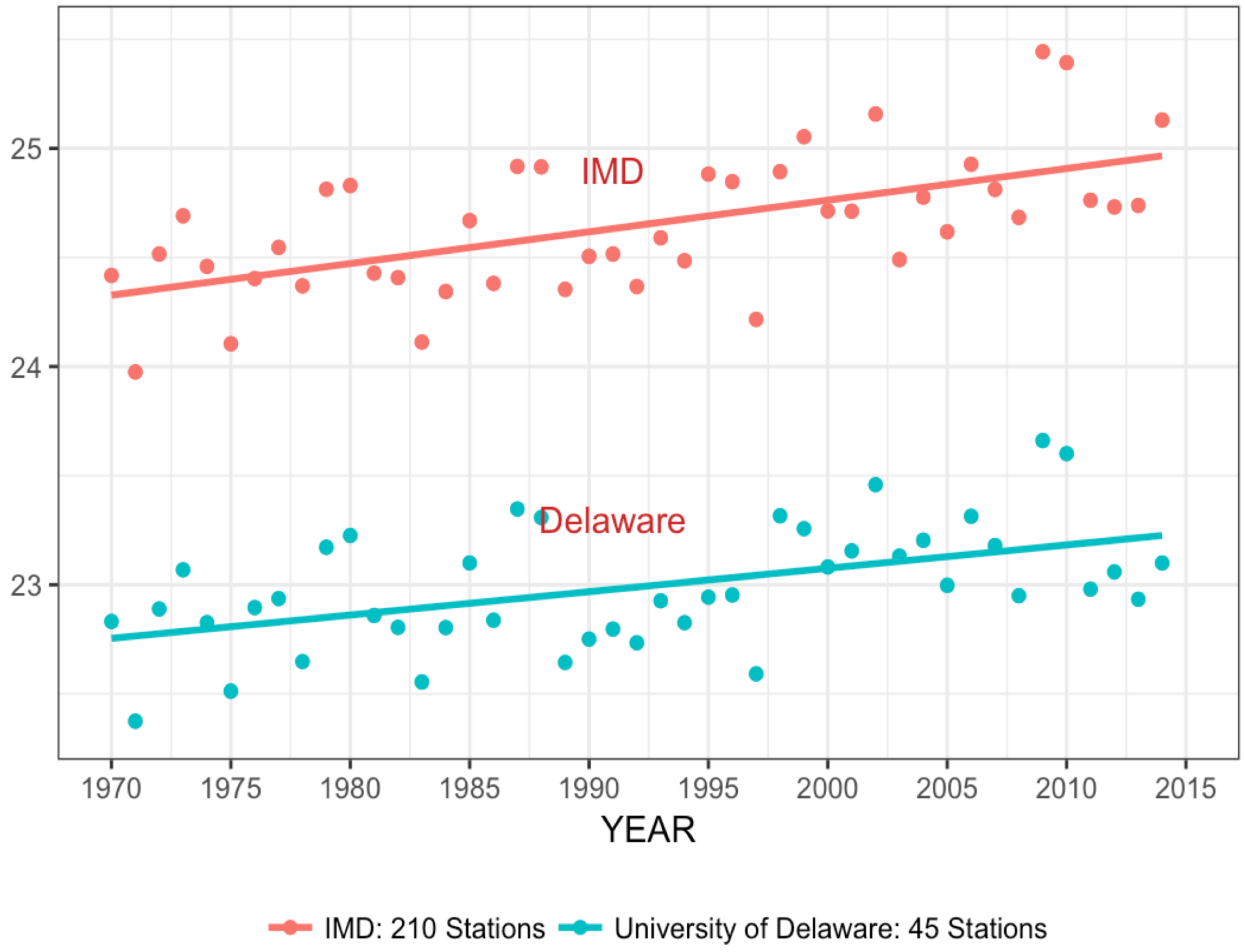
=> 6 months on data, global divergence IPCC

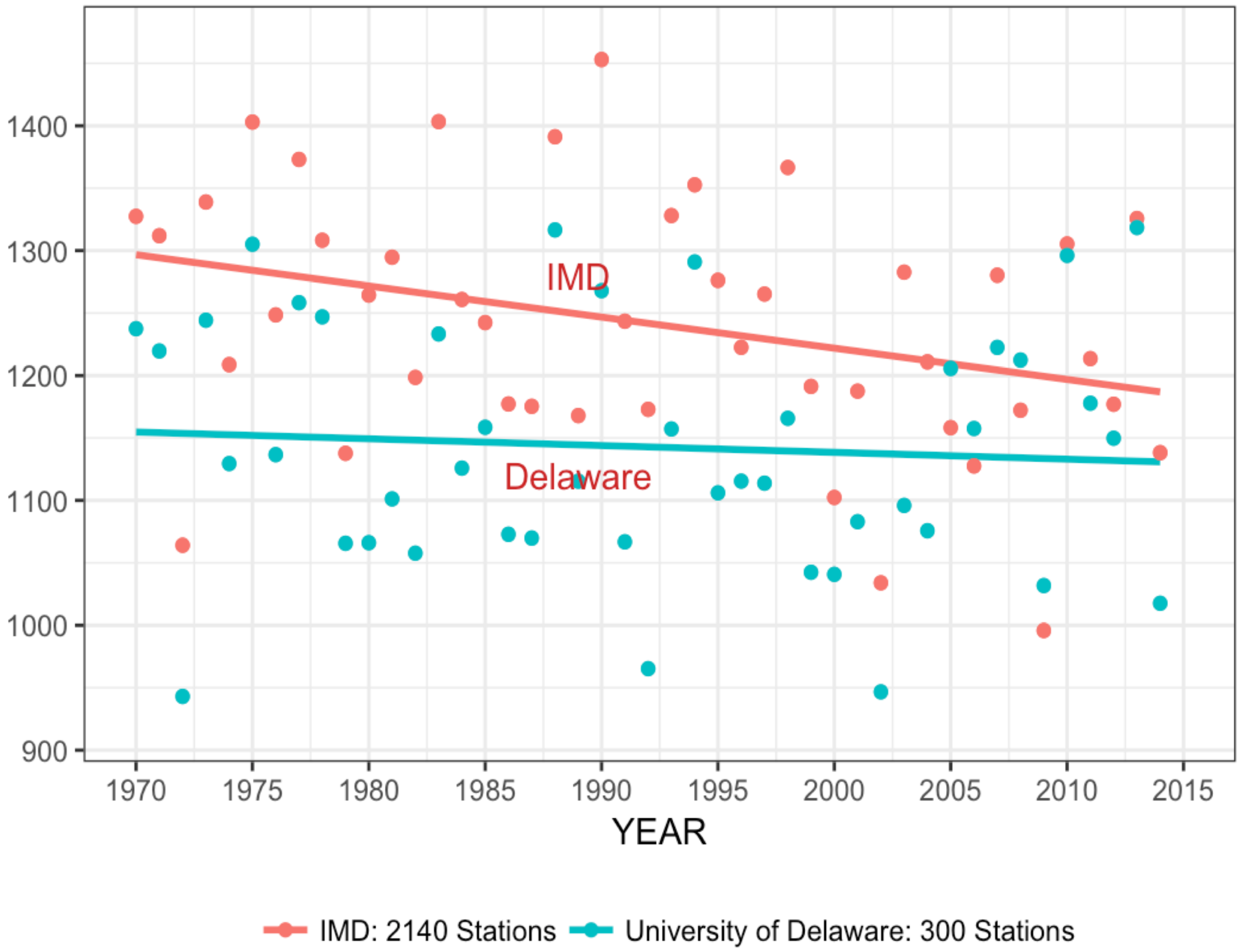
Climate Change'd' I: Motivation [Ground Stations]





Climate Change'd' I: Motivation [Leverage Data Density]

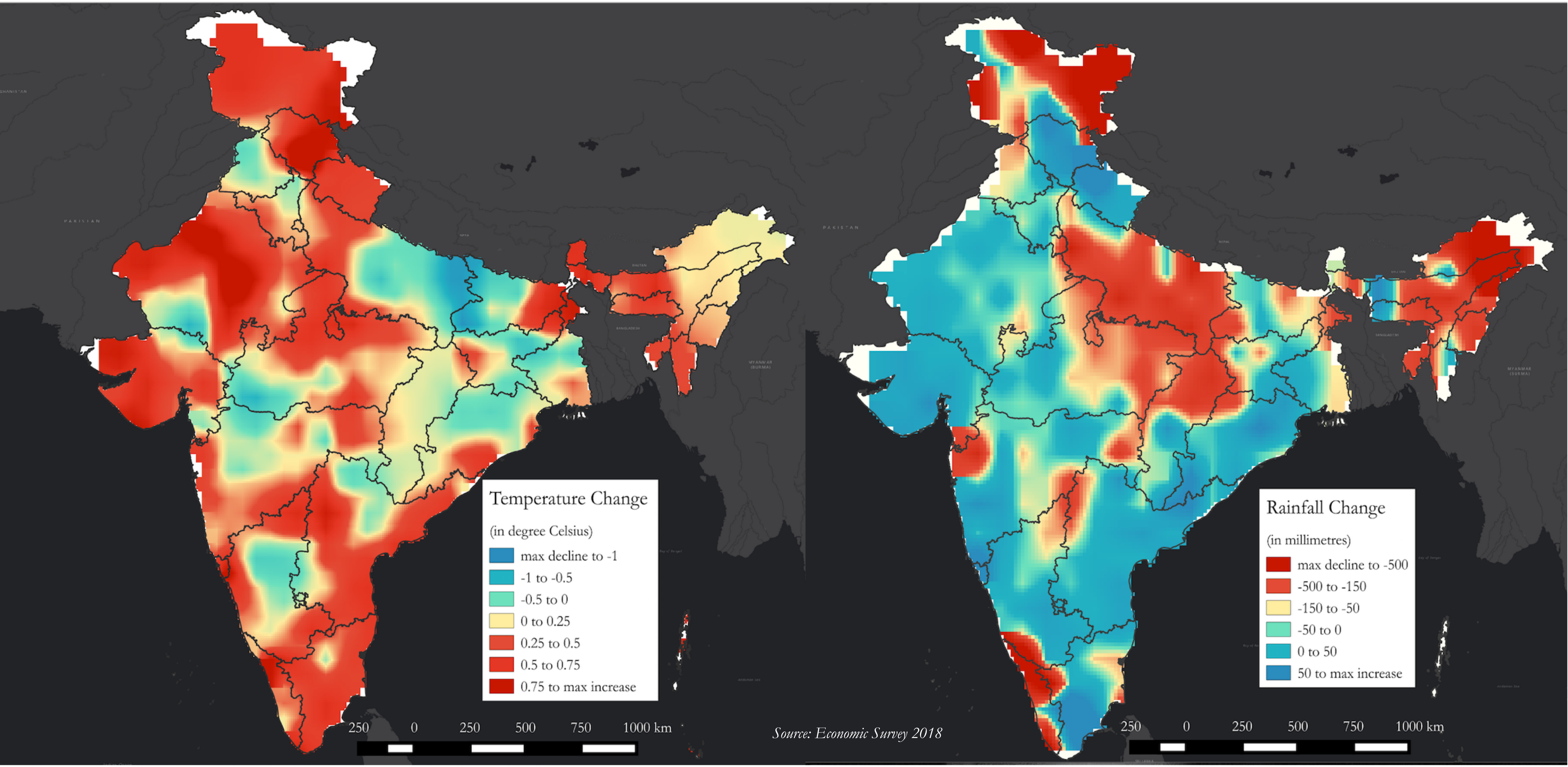




- International source of Indian Data with far fewer stations
- Upkeep: sourcing from magnetic tapes/Fortan

Source: Economic Survey 2018

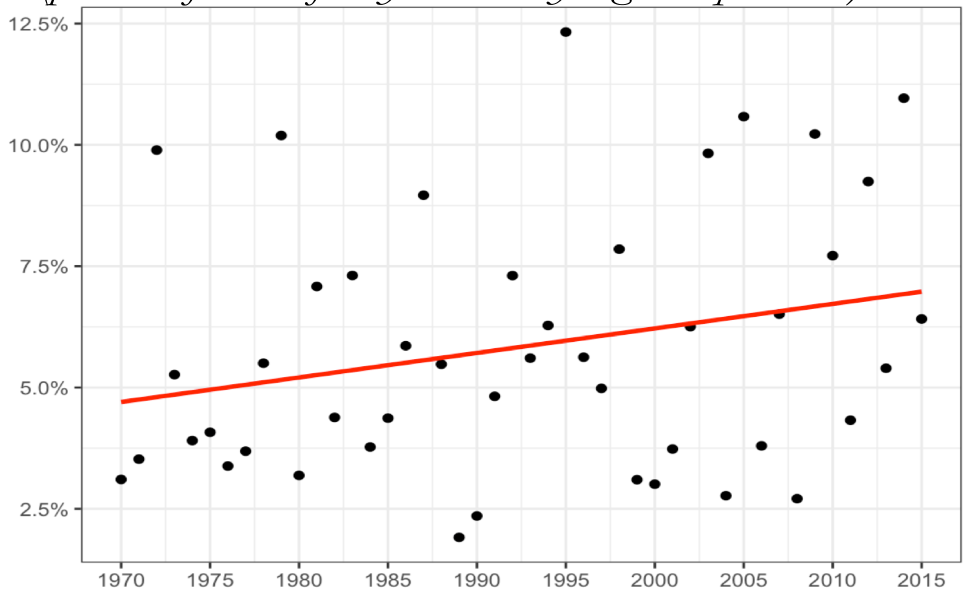
Climate Change'd' II: Change in weather patterns between the last decade and 1950-1980



Source: Economic Survey 2018

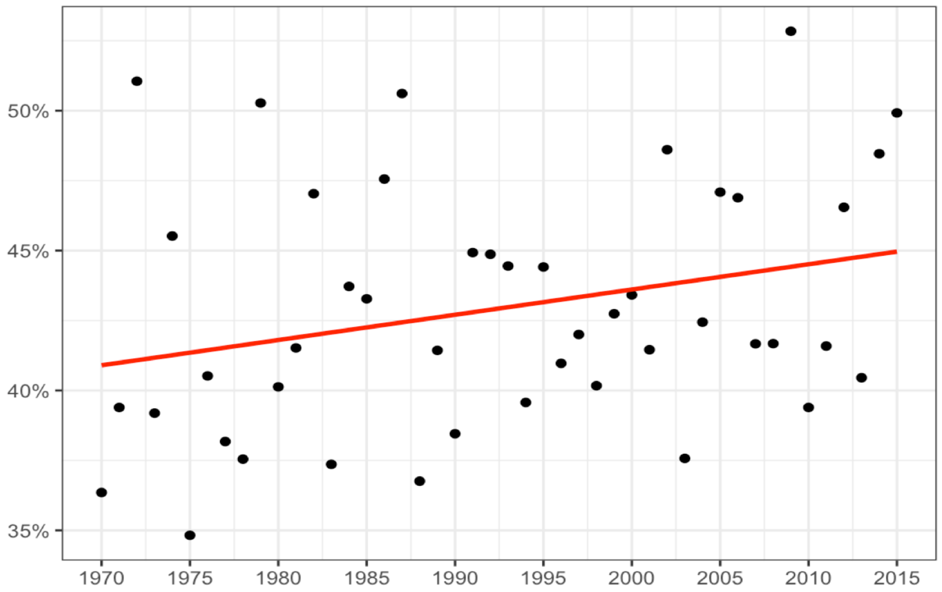
Hot Days

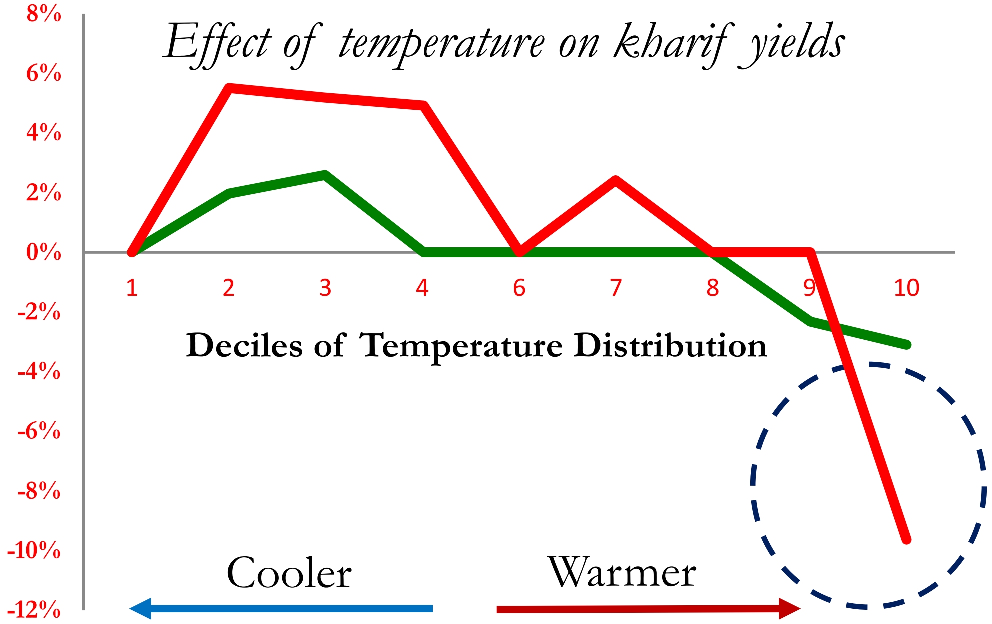
(percent of kharif days with very high temperature)

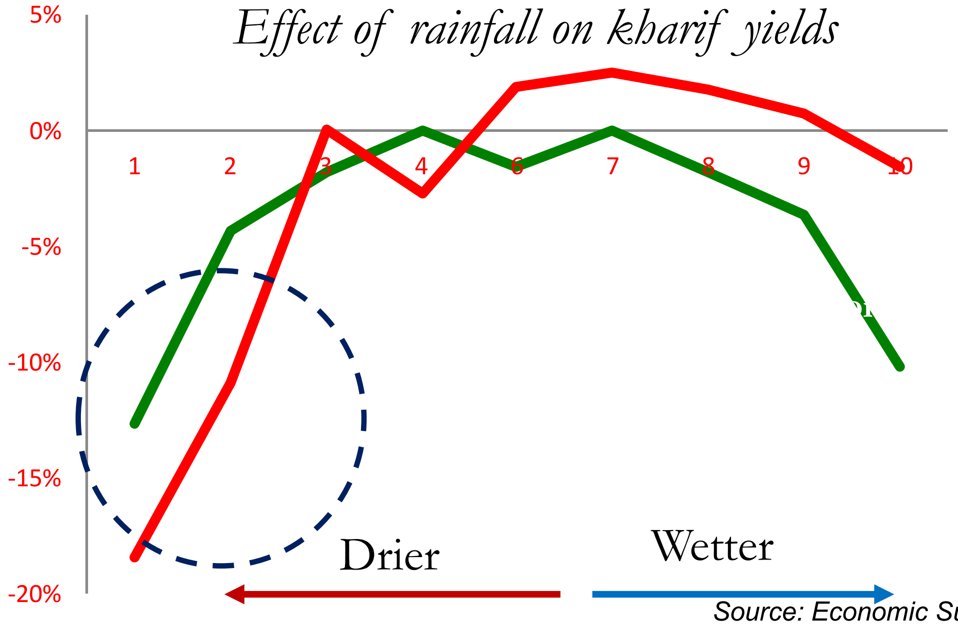


Dry Days

(percent of kharif days with less than 0.1mm rainfall)







Source: Economic Survey 2018

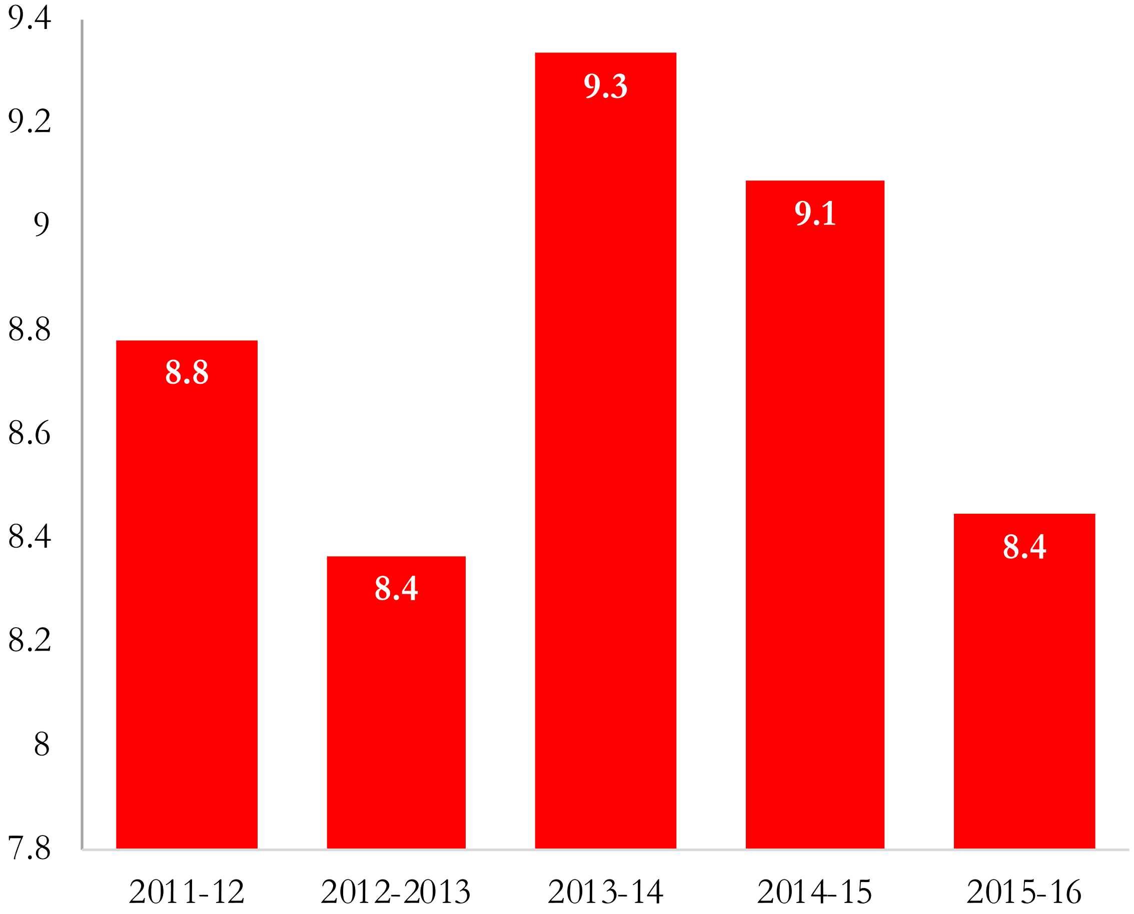
Section II: Passenger Migration

- Unreserved passenger traffic data between every origin destination pair for the years 2011-16 was obtained
- Railway 5 years data density:
  - 4,57,88,175 points for 1 years of unreserved tickets, which showed the annual mobility in India from 4 million to 9 million
  - 250 million rows
- Passenger movement {break-journey}
- Discerning day journeys
- Validation with Holiday festival season

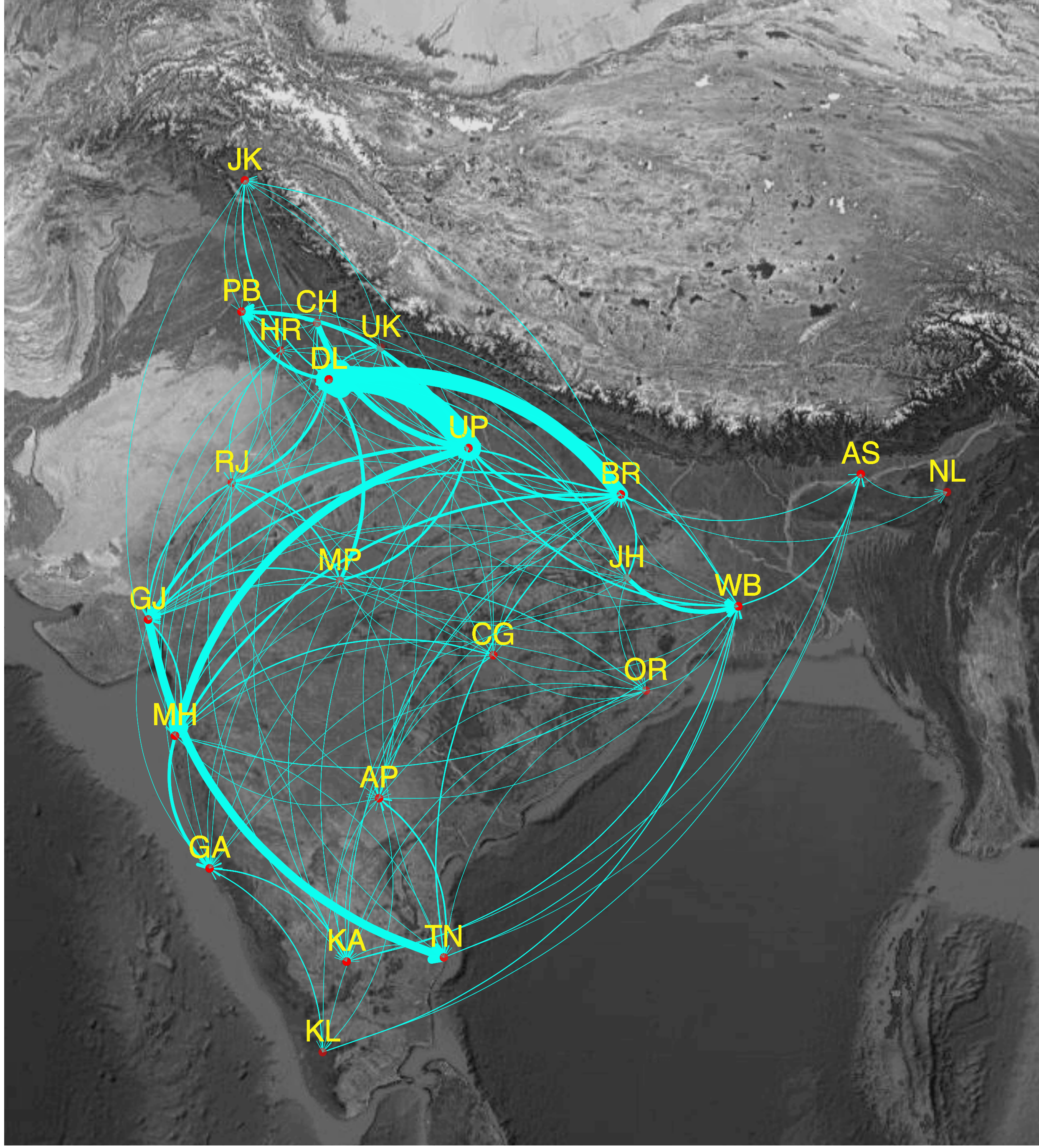
=> *BMC's implementation, 15th Finance Commission proposal forest detection,*

Migration I: Motivation

Migration II: Dynamic



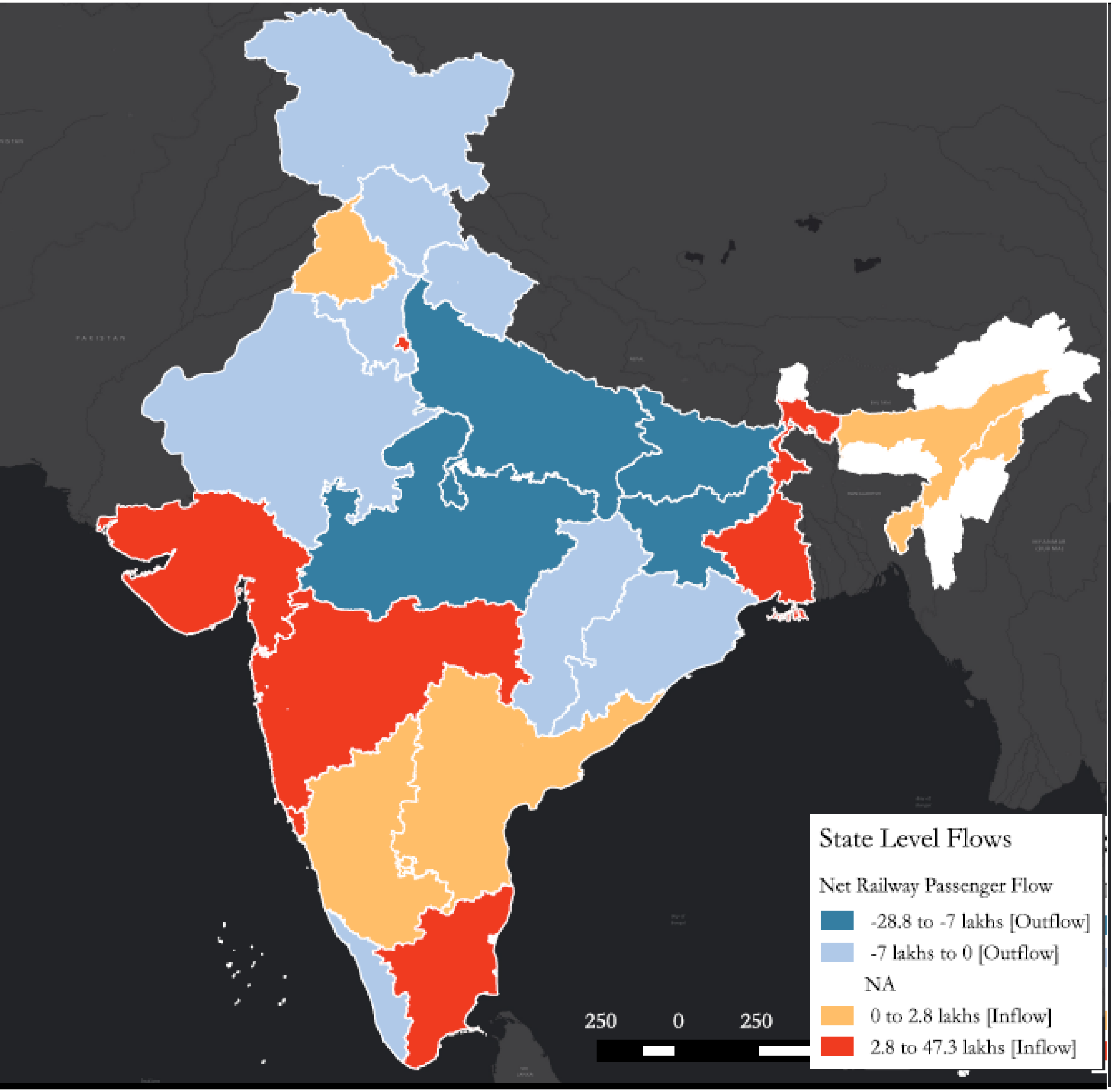
Source: Economic Survey 2017

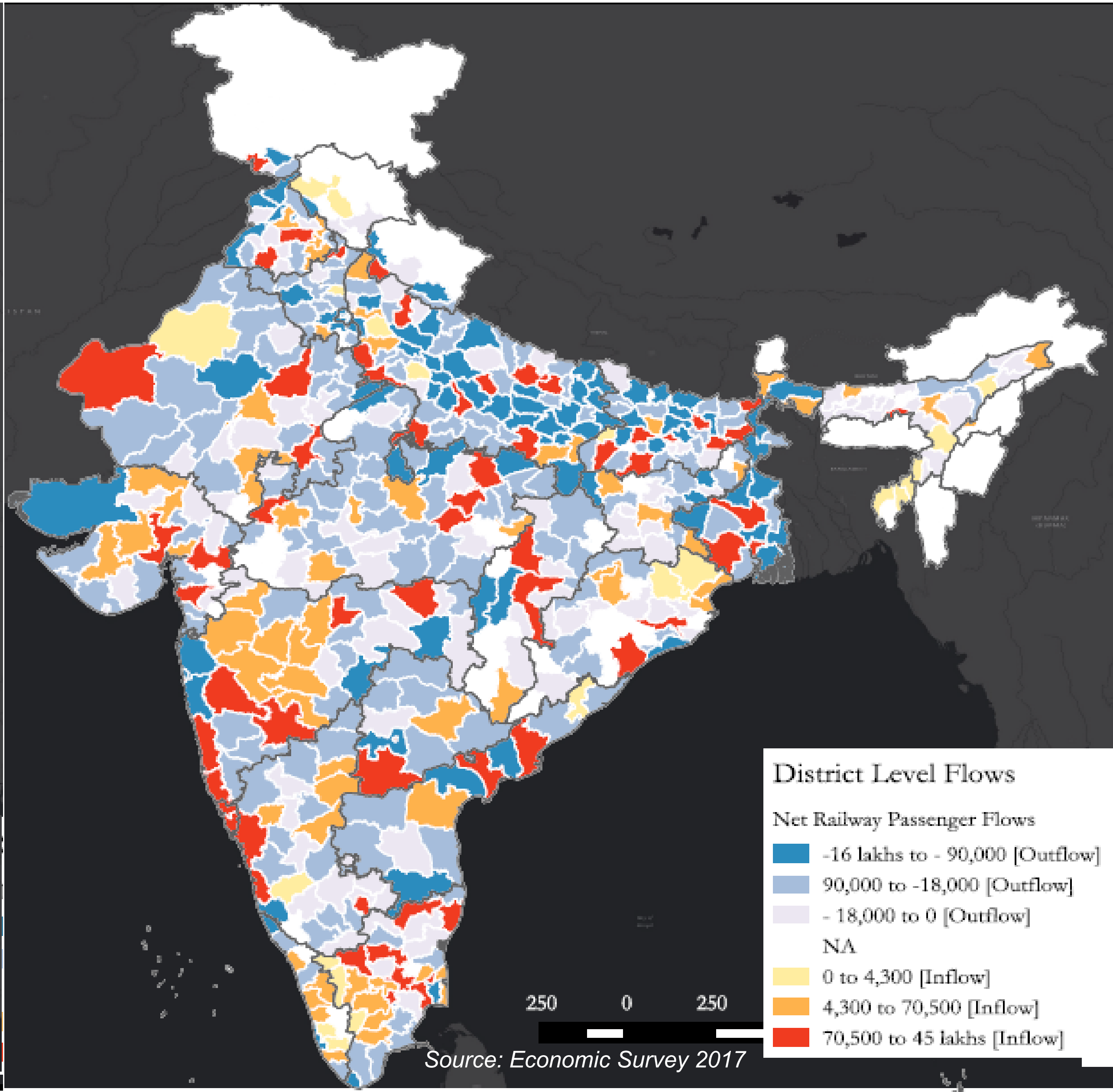


The Bank of Now IDFC IDFC BANK

Migration III: Railway Passenger Traffic

THE UNIVERSITY OF CHICAGO





Source: Economic Survey 2017

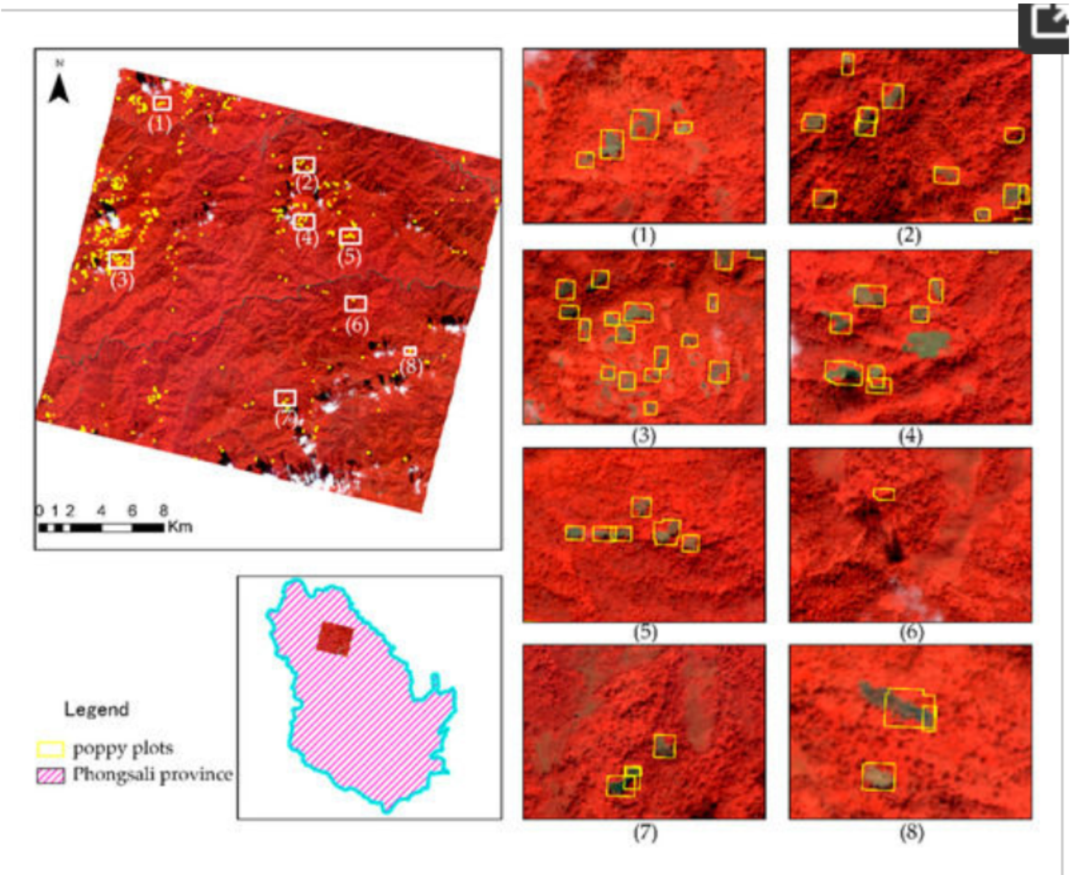
Section II. B: Public Policy

Applications: Under research & external

Active Projects I Tin (Rooftops)

Poverty assessment: Census roof-top

Active Projects III: Opium & Conflict



Source: Yiu, Tian, 2018 Myanmar

- More than 90% of illicit heroin globally, and more than 95% of the European supply

- Provides about 400,000 jobs in Afghanistan

- Area: Helmand, South Kandahar

FIN

follow up

email: pkhare@uchicago.edu, parthakhare@gmail.com

website: https://parthkhare.github.io/