

# **Georeferenced Data for Geoconflict Analysis**

**CC2Conflicts Webinar Series**

**Giorgio Chiovelli**  
Universidad de Montevideo

# Structure

## 1. Introduction: Climate and Conflict

### 2. Georeferenced and Satellite Data

- Why GIS
- Data Format: Vector and Raster data
- Softwares
- Google Earth Engine
- Common operations

### 3. Main Conflict and Climate Datasets

- Traditional Conflict Data: UCDP-GED and ACLED
- Big conflict data: ICEWS and GEDTL
- Climate Data: Rainfall, Temperature, Droughts

### 4. Additional Data:

- Economic Activities
- Urbanization
- Education and Wealth
- Migration/Displacement
- Agriculture
- Geography

### 5. Summary and Takeaways

# Climate. Definition.

Source: Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. "Climate and conflict." *Annu. Rev. Econ.* 7.1 (2015): 577-617.

- Observations of climatic variables: **Temperature, Rainfall, Water availability**
  - Climate indices: [El Niño Southern Oscillation Index, Palmer Drought Severity Index]
- Temporal Dimension:
  - Variables may be averaged over different time frames.
  - Short averaging periods (e.g., annual) still provide valuable insights into climate variability.
- Importance of Short-Lived Events:
  - Societies respond to both short-lived and long-lived changes in climatic variables. **[conflict]**
  - Frequency of short-lived events is economically relevant. **[riots, protests]**

# Conflict. Definition.

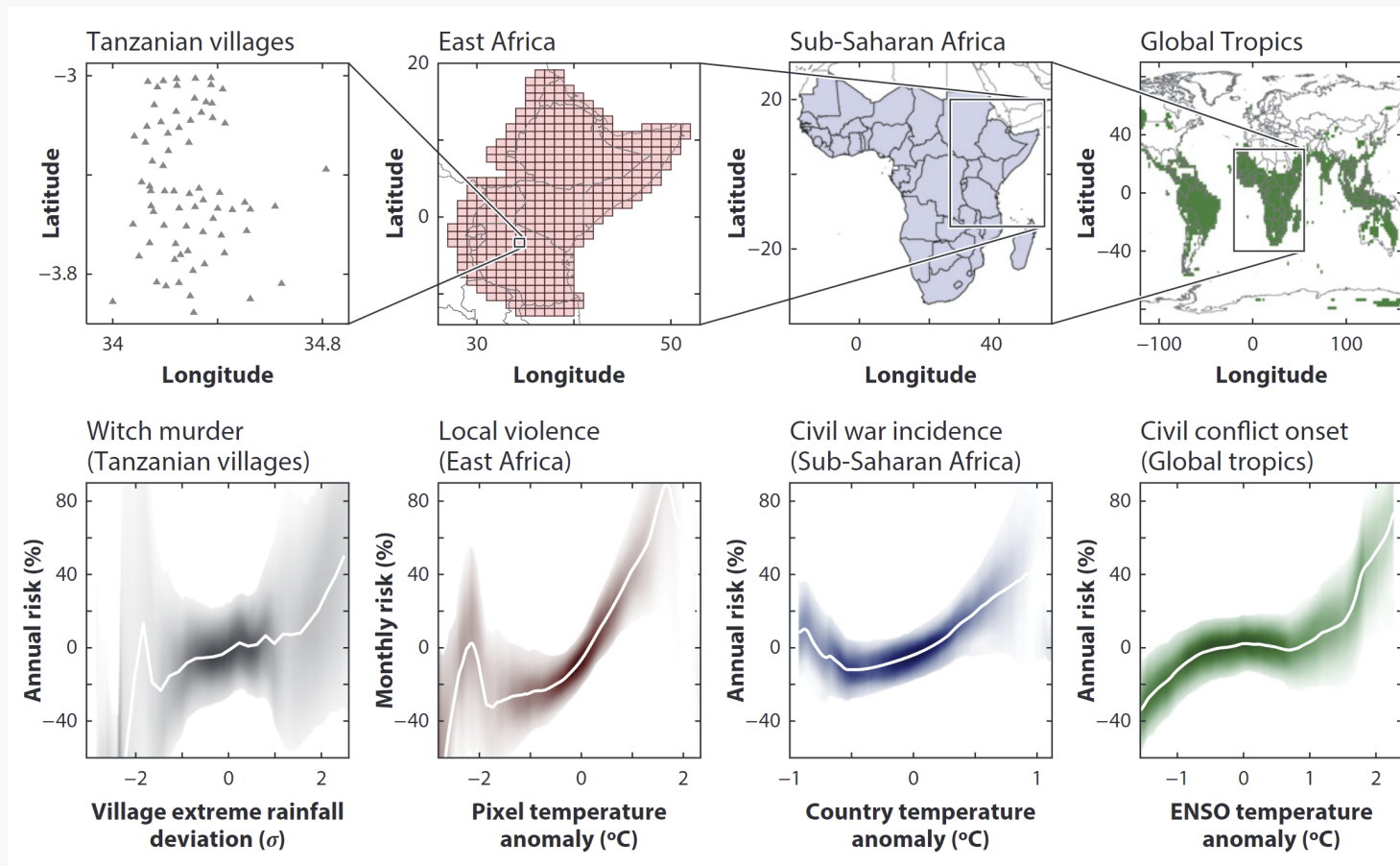
**Source: Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. "Climate and conflict." *Annu. Rev. Econ.* 7.1 (2015): 577-617.**

- Definition of Conflict:
  - Events where standard dispute resolution fails.
  - Often violent, but not always.
  - Involves individuals or groups.
  - Can be organized or disorganized.
  - Motivated personally, politically, or otherwise.
- Comprehensive Approach:
  - Different types of conflict may be interconnected.
  - Responses to climate might share commonalities across various conflicts.



# Climate and Conflict.

Source: Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. "Climate and conflict." *Annu. Rev. Econ.* 7.1 (2015): 577-617.



# Structure

## 1. Introduction: Climate and Conflict

## 2. Georeferenced and Satellite Data

- Why GIS
- Data Format: Vector and Raster data
- Softwares
- Google Earth Engine

## 3. Main Conflict and Climate Datasets

- Traditional Conflict Data: UCDP-GED and ACLED
- Big conflict data: ICEWS and GEDTL
- Climate Data: Rainfall, Temperature, Droughts

## 4. Additional Data:

- Economic Activities
- Urbanization
- Education and Wealth
- Migration/Displacement
- Agriculture
- Geography

## 5. Summary and Takeaways

# GIS in Economics Research.

## — Geographic Information System (GIS)

### — Without GIS, spatial units of analysis are limited to

- countries
- administrative districts in some developed countries (NUTS in Europe, counties in the US)
- some villages in developing countries

### — With GIS, the unit of analysis can be any level of spatial

- aggregation
- administrative units across all countries
- All populated territories globally
- locations of ethnic/linguistic groups
- artificial units

# GIS in Economics Research.

- **More credible identification strategies. [not the focus of this lecture!]**
  - account for many geographic covariates
  - create instruments (e.g., distance from specific locations)
  - conduct spatial Regression Discontinuity design
  - explore the level at which the pattern uncovered prevails. Invariance to spatial aggregation?
  
- **Accessing data with spatial structure [today's lecture]**
  - information about where observations are in relation to each other.
  
- **[check our free [GIS online course at the Wheeler Institute at LBS](#). May 2024]**

# Introduction to GIS. Data Type.

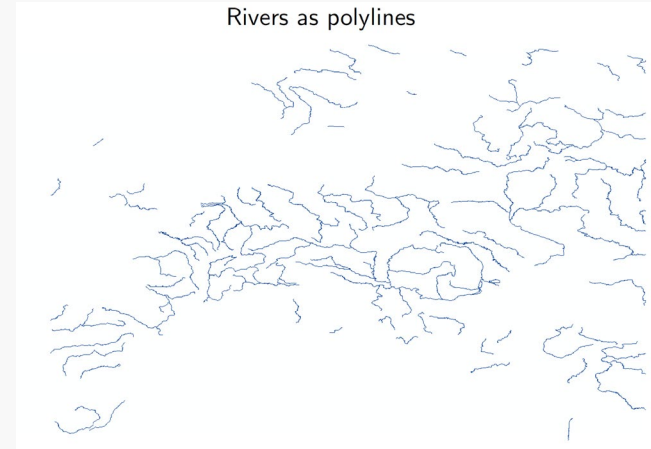
- **Geographic data comes in a vast number of formats.**
- For 99.9% of all applications, you will only need two:
  - **Feature (vector) data**, files end in .shp (shapefiles)
  - **Raster (cell) data**, files (typically) end in .tif, but other formats are also standard.
- Working with climatic data, you might also encounter NetCDF:
  - **NetCDF** (network Common Data Form) is a file format for storing multidimensional scientific data (variables) such as temperature, humidity, pressure, wind speed, and direction. [less relevant with Google Earth Engine (more soon)]
- It is helpful to distinguish three types of feature data
  - polygon features
  - polyline features
  - point features

# GIS. Vector (Shapefile) Data.

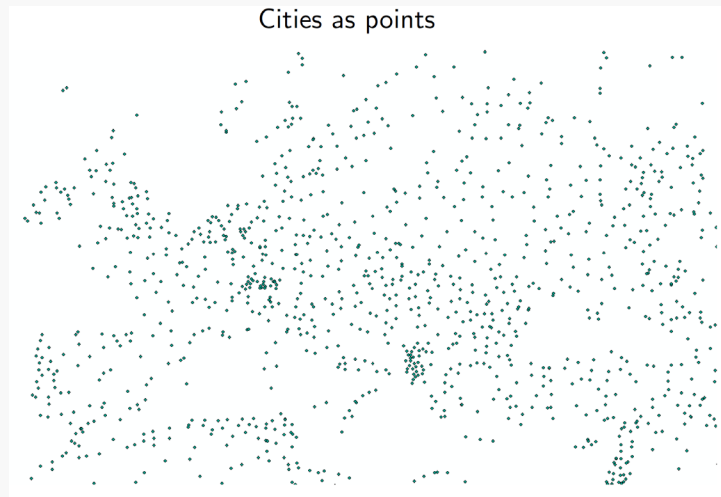
Countries can be treated as polygon features



Rivers as polylines

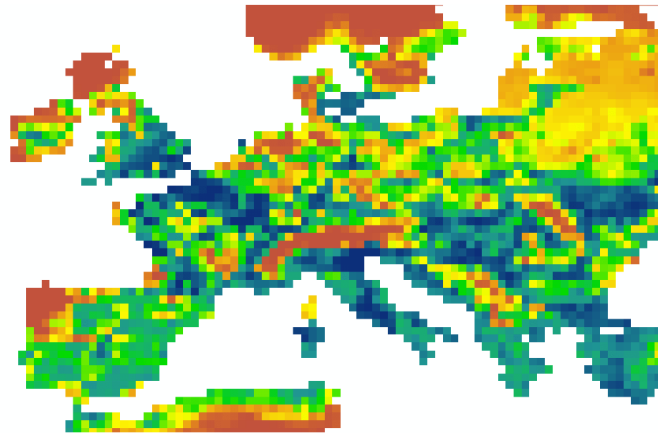


Cities as points



# GIS. Raster Data.

Agricultural suitability in raster format



Each cell in the raster has a value

1	1	0	0
	1	2	2
4	0	0	2
4	0	1	1

- cells can be empty
- data does not need to be integer (can be float)

# GIS operations, software, and programming languages.

## Operations

### — **Gazillion data manipulations/operations through GIS**

- Most common in econ:
  - Zonal Statistics: computing statistics of rasters upon shapefiles [e.g. rainfall averages at district level]

## Softwares and coding languages

### — **Older generations [like myself] started with:**

- ArcGIS [ERSI software, licensed]
- Then, switched to QGIS [Mac users, open-source and free]

### — Newer generations directly coding in **Python** (background of ArcGIS and QGIS) and/or **R**.

### — Big spatial data query usually done in SQL or Postgres

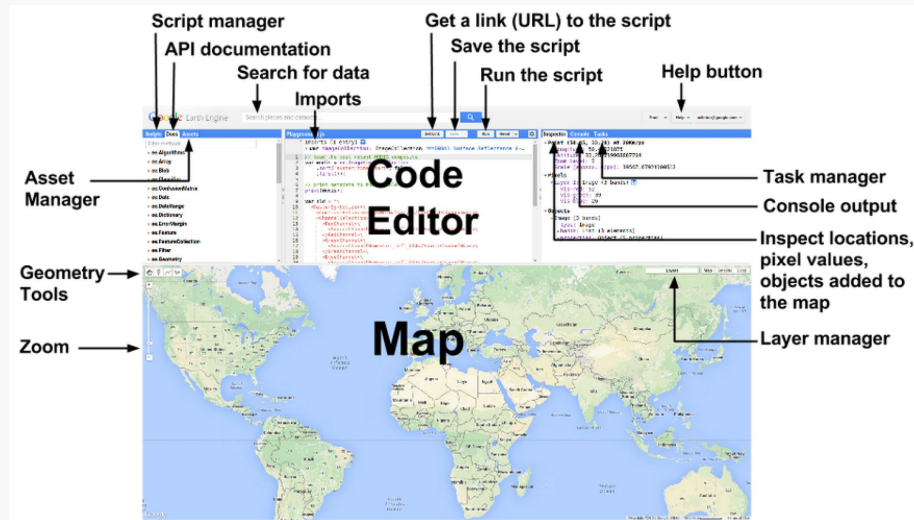


# Google Earth Engine.

**Google Earth Engine (GEE) is a cloud computing platform combining**

- Vast collection of satellite images and other geospatial data, together
  - with great documentation:
  - <https://developers.google.com/earth-engine/datasets/>
- You can also upload your data (like shapefiles)
- Convenient coding environment (Java) that runs within the browser and allows you
  - to visualize data as well as analyze and transform them with a broad
  - range of algorithms
- Ability to run analysis on Google servers! analysing huge datasets becomes possible (though there are still query limits)

# Google Earth Engine.



— source: GEARS lab Github page:  
<https://github.com/geospataleco/GEARS>

— If you want to get deeper into GEE, this (in addition to the excellent official documentation) is a great starting point

# Structure

## 1. Introduction: Climate and Conflict

## 2. Georeferenced and Satellite Data

- Why GIS
- Data Format: Vector and Raster data
- Softwares
- Google Earth Engine
- Common operations

## 3. Main Conflict and Climate Datasets

- Traditional Conflict Data: UCDP-GED and ACLED
- Big conflict data: ICEWS and GEDTL
- Climate Data: Rainfall, Temperature, Droughts

## 4. Additional Data:

- Economic Activities
- Urbanization
- Education and Wealth
- Migration/Displacement
- Agriculture
- Geography

## 5. Summary and Takeaways

# Conflict.

## Overview

### 1. UCDP-GED

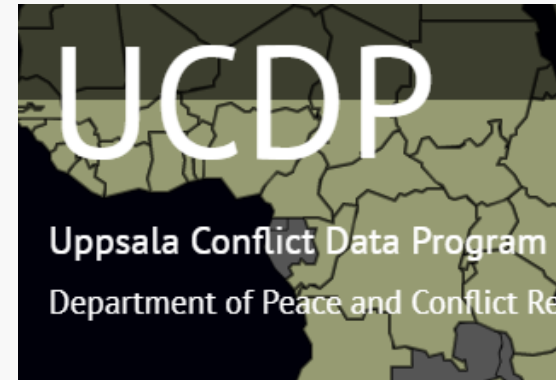
- Major and minor civil wars, civilian violence. 1989-2019.
- Standard definition of conflict
- Number of casualties

### 2. ACLED

- Riots, protests, battles, etc. 1997-2022
- Aggregation of news sources.

— **Source: Uppsala University & ACLED**

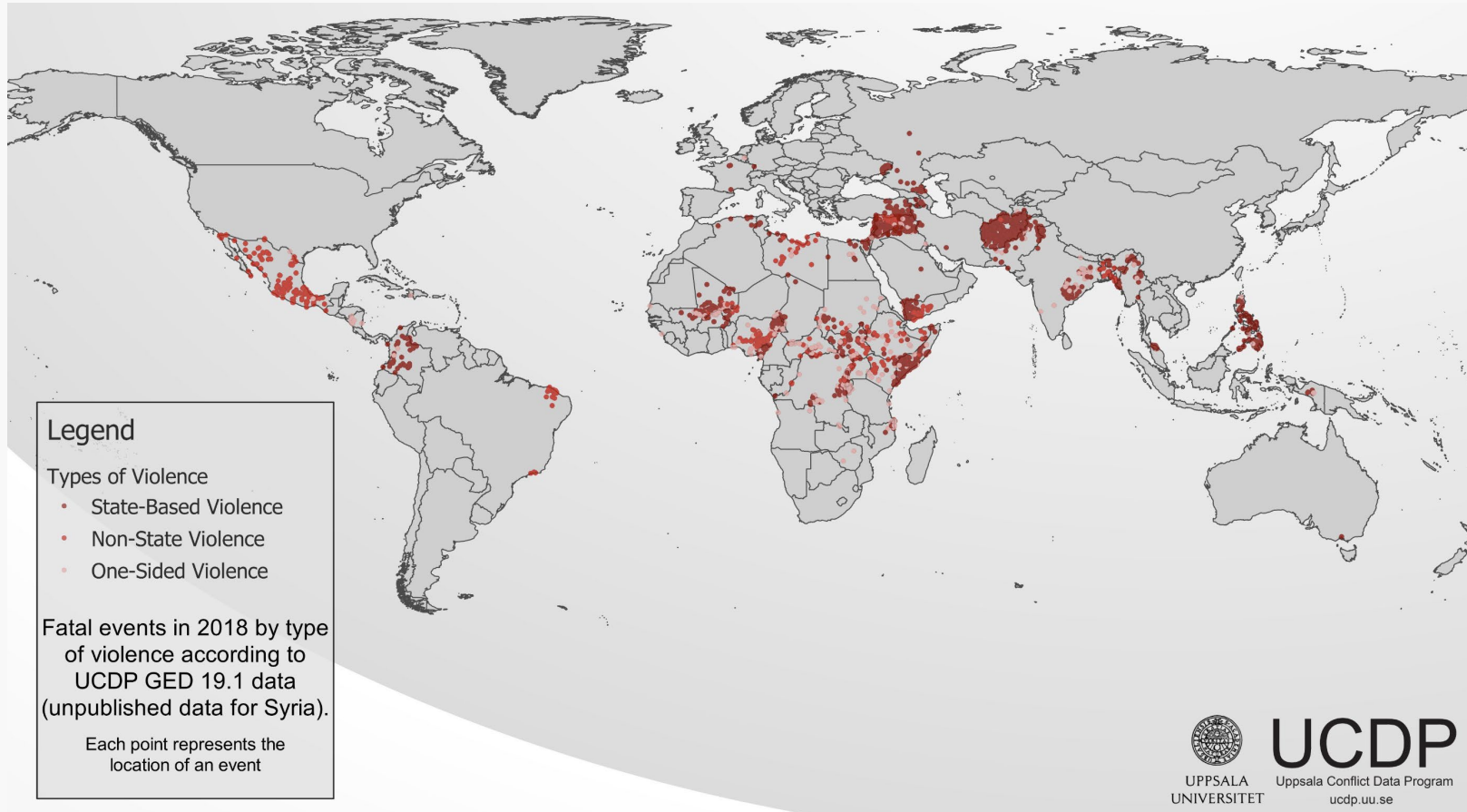
- Download [UCDP-GED](#)
- Download [ACLED](#)



**ACLED**  
Bringing clarity to crisis

# Conflict.

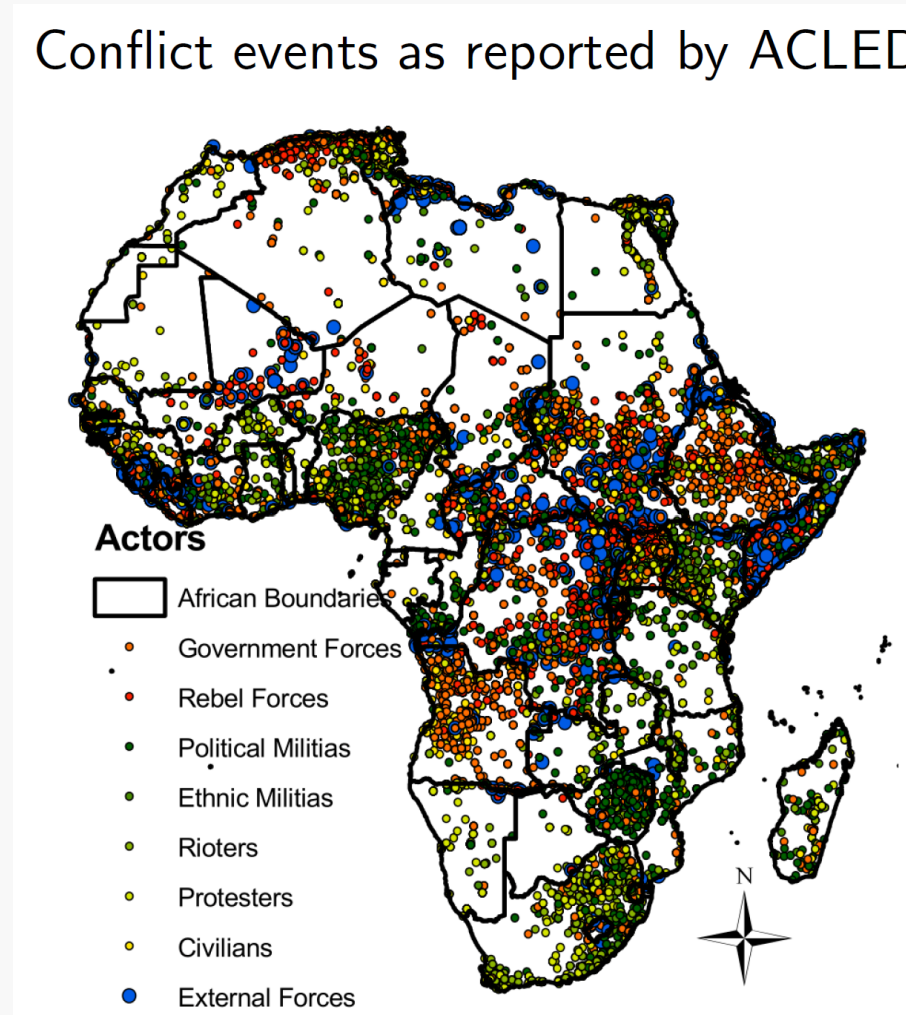
## Example 1. UCDP-GED conflict events



# Conflict.

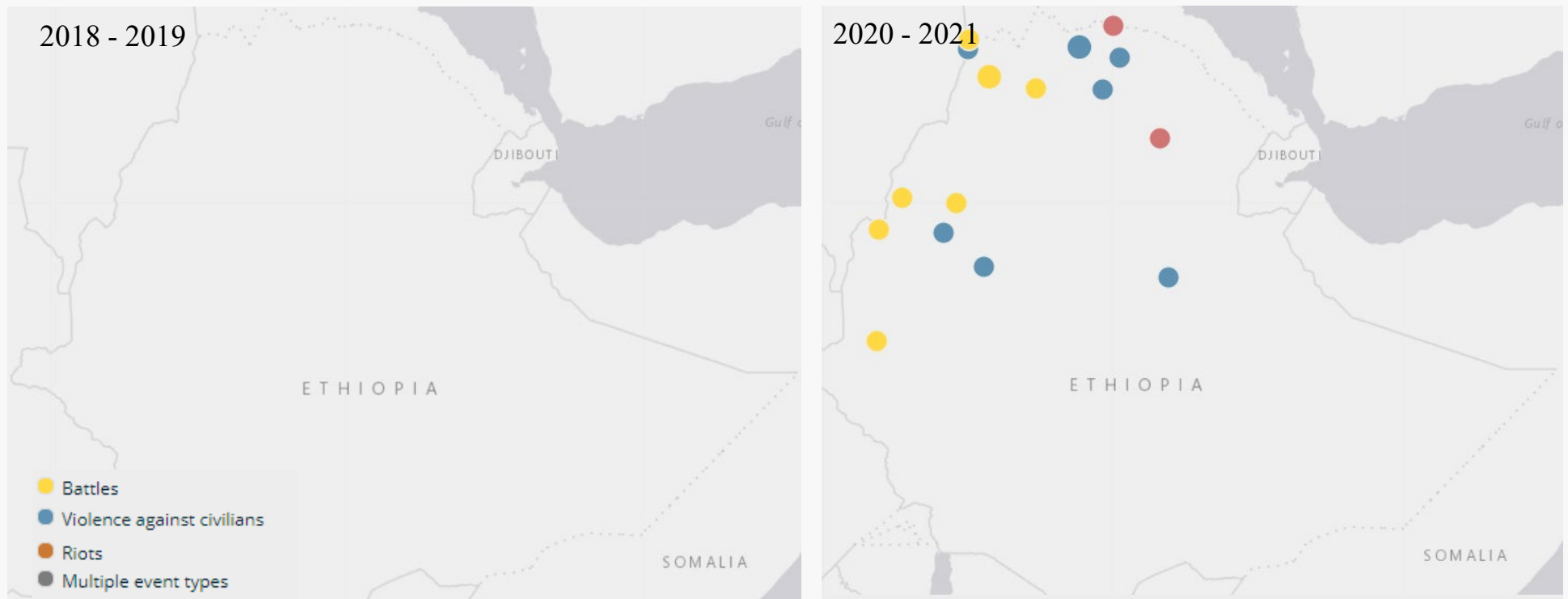
## Example 2. ACLED conflict events

Conflict events as reported by ACLED



# Conflict (ACLED).

## Example 3. Ethiopia Recently



# Conflict. SCAD & ICEWS

## 3. Social Conflict Analysis Database (SCAD)

- Social conflicts from 1990-2017
- Lower intensity conflict
- Africa and now also Mexico, Central America, and the Caribbean

— Source: University of Austin & ICEWS

- Download [SCAD](#)





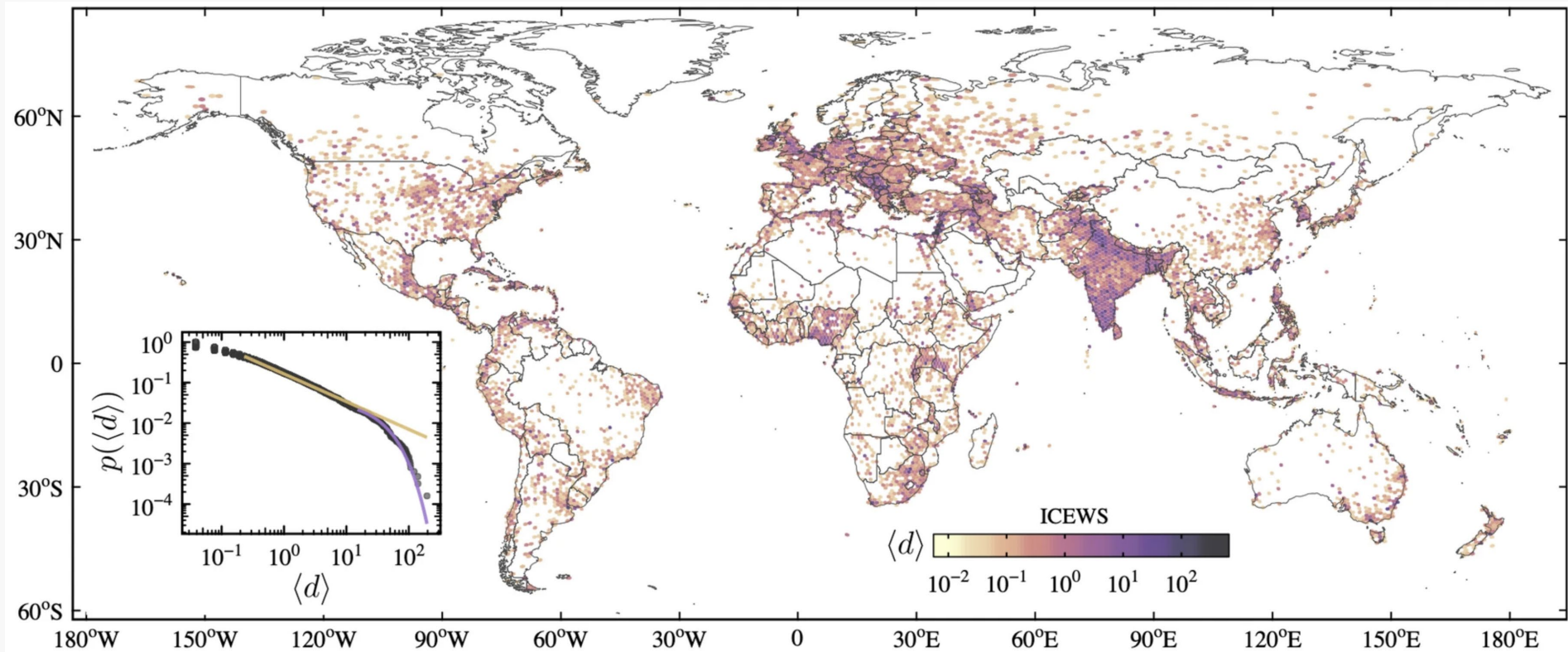
# Conflict. ICEWS

- Integrated Crisis Early Warning System
- Covers the period from 1995 to 2022 (20M events).
- Unlike other datasets (e.g., ACLED), ICEWS focuses on a broader range of political interactions.
- Includes political statements, accusations of crime or corruption, and human rights abuses.
- Information on the source and target of the interaction and the level of hostility or cooperation (-10 to 10).
- Events are automatically identified and extracted from news articles, geo-referenced, and time-stamped.



Download [ICEWS](#)

# ICEWS. Worldwide 1995-2022



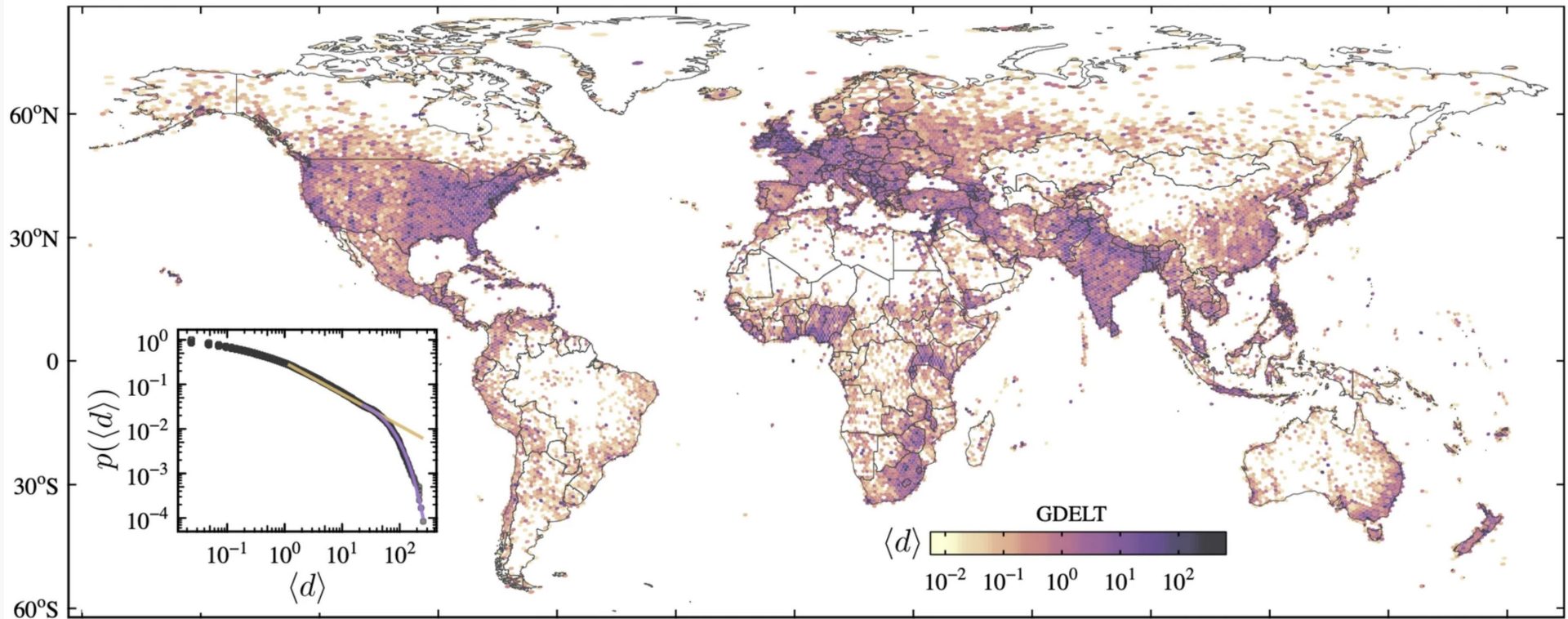
Source: Ferreira, L. N., Hong, I., Rutherford, A., & Cebrian, M. (2021). The small-world network of global protests. *Scientific Reports*, 11(1), 19215.

# Towards Big Data. GDELT Project.

- GDELT Event Database records over 300 categories of physical activities around the world
- from riots and protests to peace appeals and diplomatic exchanges
- Highly georeferenced accuracy across the entire planet
- January 1, 1979 to present.
- 2.5 TB in 2022.

**The GDELT Project**

# GDELT. Worldwide 1979-2022



Source: Ferreira, L. N., Hong, I., Rutherford, A., & Cebrian, M. (2021). The small-world network of global protests. *Scientific Reports*, 11(1), 19215.

# Comparing ICEWS and GDELT

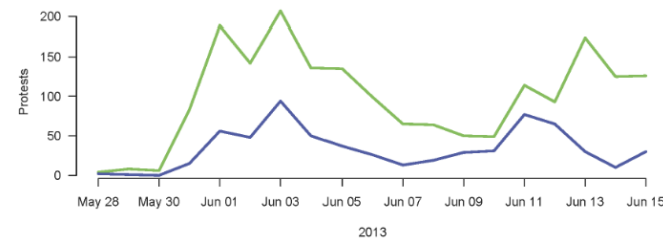
Figure 6: Map of ICEWS Protests in Turkey



Figure 5: Map of GDELT Protests in Turkey



Figure 4: ICEWS (blue) and GDELT (green) plots of protests during May and June 2013.



Source: Ward, Michael D., et al. "Comparing GDELT and ICEWS event data." *Analysis* 21.1 (2013): 267-297.

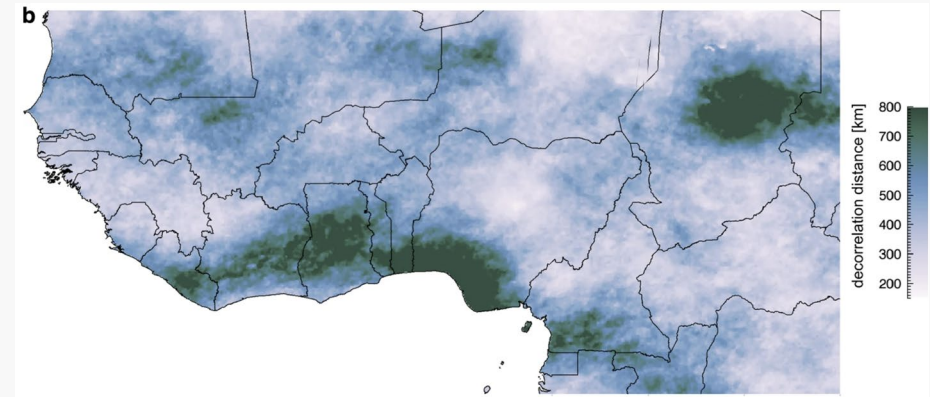


# Rainfall CHIRPS

- Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset
- incorporates satellite imagery with in-situ station data to create rainfall time series
- 0.05-degree resolution grid level (5x5km at the equator)
- Worldwide coverage from 1981 to present (monthly)

— **Source: UCSB**

- Download [CHIRPS](#)

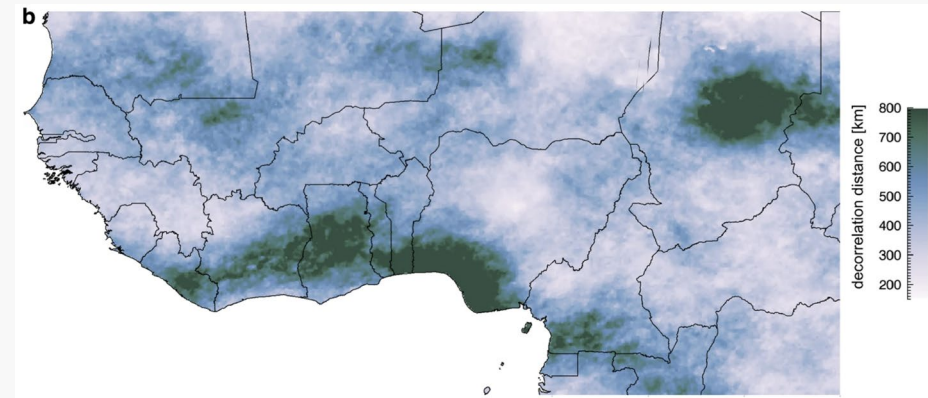


Source: Funk, Chris, et al. "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes." *Scientific data* 2.1 (2015): 1-21.

# Rainfall CHIRPS

- Dependent on weather station locations
- Interpolated data:

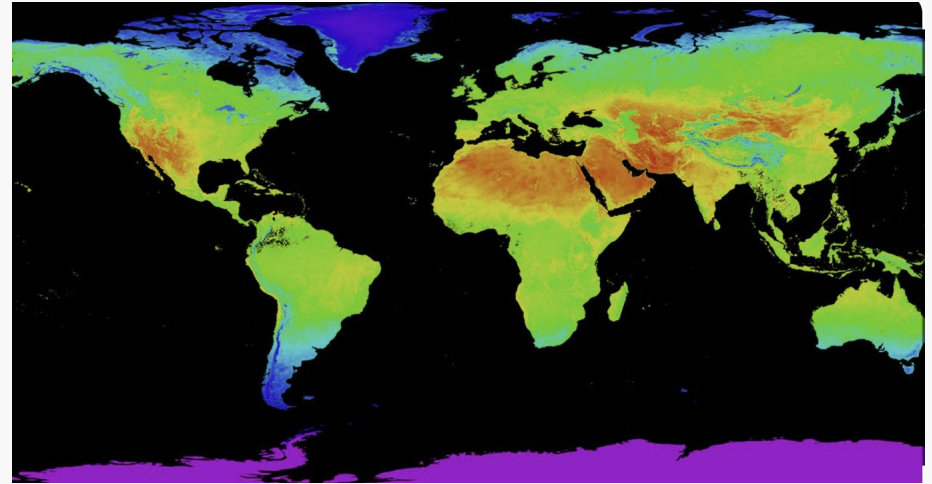
“For any given pixel, the CHIRPS blending procedure is based on a weighted average of the ratios between the five closest stations and the CHIRP”



Source: Funk, Chris, et al. "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes." *Scientific data* 2.1 (2015): 1-21.

# Temperature. MODIS Terra 6

- Monthly Land Surface Temperature and Emissivity (LST&E) values
- 2000 to present
- 0.05 degree latitude/longitude
- Combination of daytime and nighttime observations
- Download [MODIS Terra 6](#)





# Droughts. SPEI

- Global Standardised Precipitation-Evapotranspiration Index (SPEI)
- information about drought conditions at the global scale
- 0.5 degrees spatial resolution and a monthly time resolution from 1901-present
- based on monthly precipitation and potential evapotranspiration data from the CRU [coarse resolution than MODIS]
- R package to compute SPEI time series under various data scenarios.



# ACLED and SPEI.

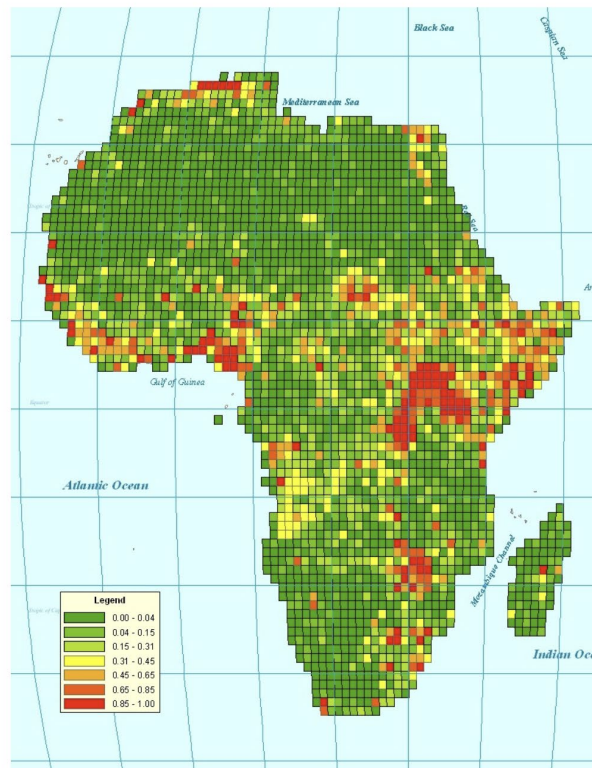


Figure 1A:

Fraction of years with at least one conflict event (1997-2011)

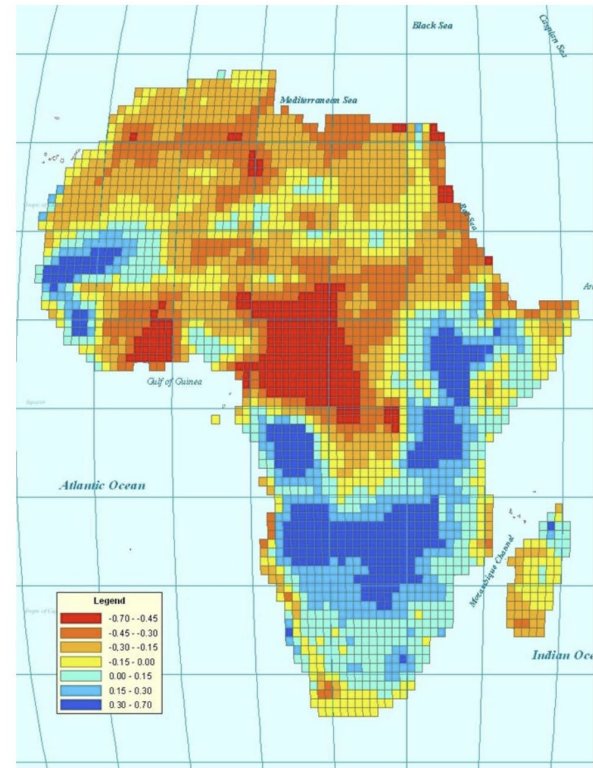


Figure 1B:

Average SPEI (1997-2011)

Harari, Mariaflavia, and Eliana La Ferrara. "Conflict, climate and cells: A disaggregated analysis." (2013).

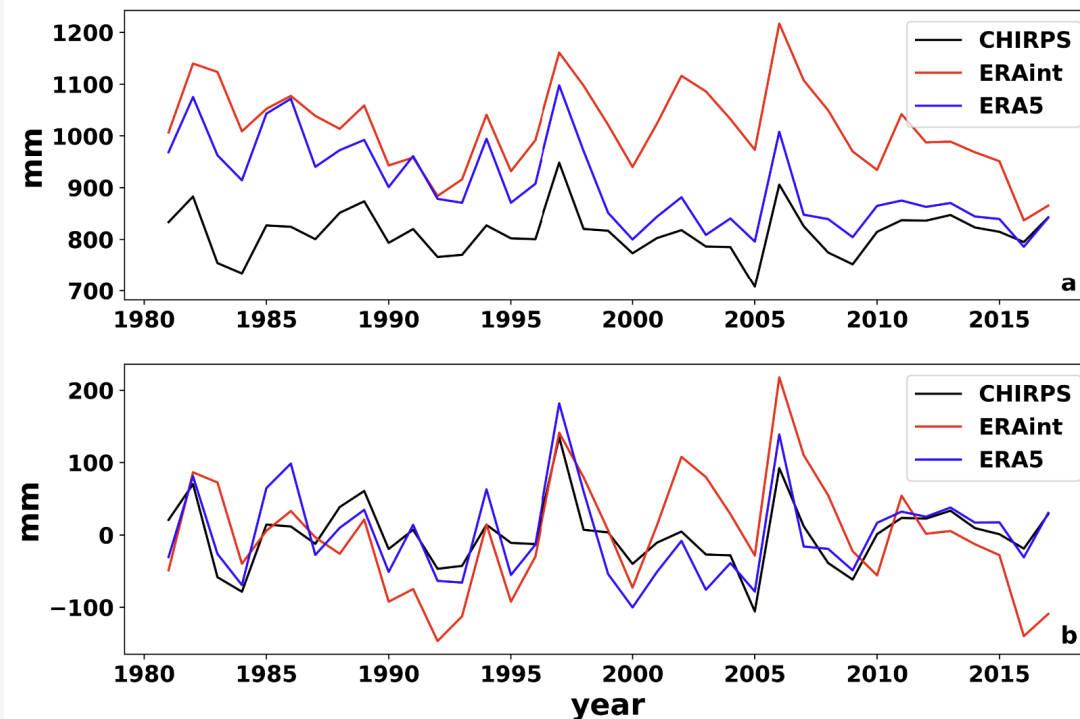
# Climate. ERA5: Rainfall and Temperature

- CMWF / Copernicus Climate Change Service
- Reanalysis combines model data with worldwide observations into a complete and consistent dataset.
- ERA5 data is available from 1940 to three months to present
- Lat-Lon grid of 0.25 degrees

Description
Average air temperature at 2m height (monthly average)
Minimum air temperature at 2m height (monthly minimum)
Maximum air temperature at 2m height (monthly maximum)
Dewpoint temperature at 2m height (monthly average)
Total precipitation (monthly sums)
Surface pressure (monthly average)
Mean sea level pressure (monthly average)
10m u-component of wind (monthly average)
10m v-component of wind (monthly average)

## Comparison CHIRPS vs. ERA5

**Figure 10.** Annual precipitation averaged over longitude band 25° E–50° E and 13° S–17° N (East Africa box) in observations (CHIRPS, black), ERA-interim (red), and ERA5 (blue) from 1981 to 2017 as annual sums (a) and detrended annual anomalies (b).



Gleixner, Stephanie, Teferi Demissie, and Gulilat Tefera Diro. "Did ERA5 improve temperature and precipitation reanalysis over East Africa?." *Atmosphere* 11.9 (2020): 996.

- The “best” dataset is the one that fits your application
- Several considerations:
  - Temporal coverage of your analysis
  - Spatial dimensions of your units of observation [e.g. country vs pixels]
- All datasets have limitations.

# Structure

## 1. Introduction: Climate and Conflict

## 2. Georeferenced and Satellite Data

- Why GIS
- Data Format: Vector and Raster data
- Softwares
- Google Earth Engine
- Common operations

## 3. Main Conflict and Climate Datasets

- Traditional Conflict Data: UCDP-GED and ACLED
- Big conflict data: ICEWS and GEDTL
- Climate Data: Rainfall, Temperature, Droughts

## 4. Additional Data:

- Economic Activities
- Urbanization
- Education and Wealth
- Migration/Displacement
- Agriculture
- Geography

## 5. Summary and Takeaways

# Additional Outcomes and Potential Mechanisms

- **Development Proxies:**

- Economic Activities [Luminosity]
- Urbanization [Landcover data]
- Education and Wealth [DHS surveys and IPUMS census data]
- Migration/Displacement and Mobility [IPUMS census data]
  
- Agriculture [FAO GAEZ crop-specific data]
- Growing Season [MIRCA]
- Geographic Data [Elevation, Agricultural Suitability]

# GIS data. Luminosity / Light Density at Night

## Overview

### — Data Overview

- Proxy: Income / output / development [level and changes]
- Universal coverage
- Every 1 km<sup>2</sup> across the world
- Every year since 1992 [1992 – 2013, DMSP; 2013-2021 VIIRS]

### — Source: NOA/NASA

- Download DMSP at the Earth Observation Group
- Download VIIRS at the Earth Observation Group
- New Time Series linking DMSP and VIIRS will be available soon.



Source: NASA



# GIS data. Luminosity / Light Density at Night

## Overview

### — Caveats

- Broad notion of “development” [mixing population and income/output]
- Technical issues [e.g., blooming, top-coding, etc.]
- Many unlit areas

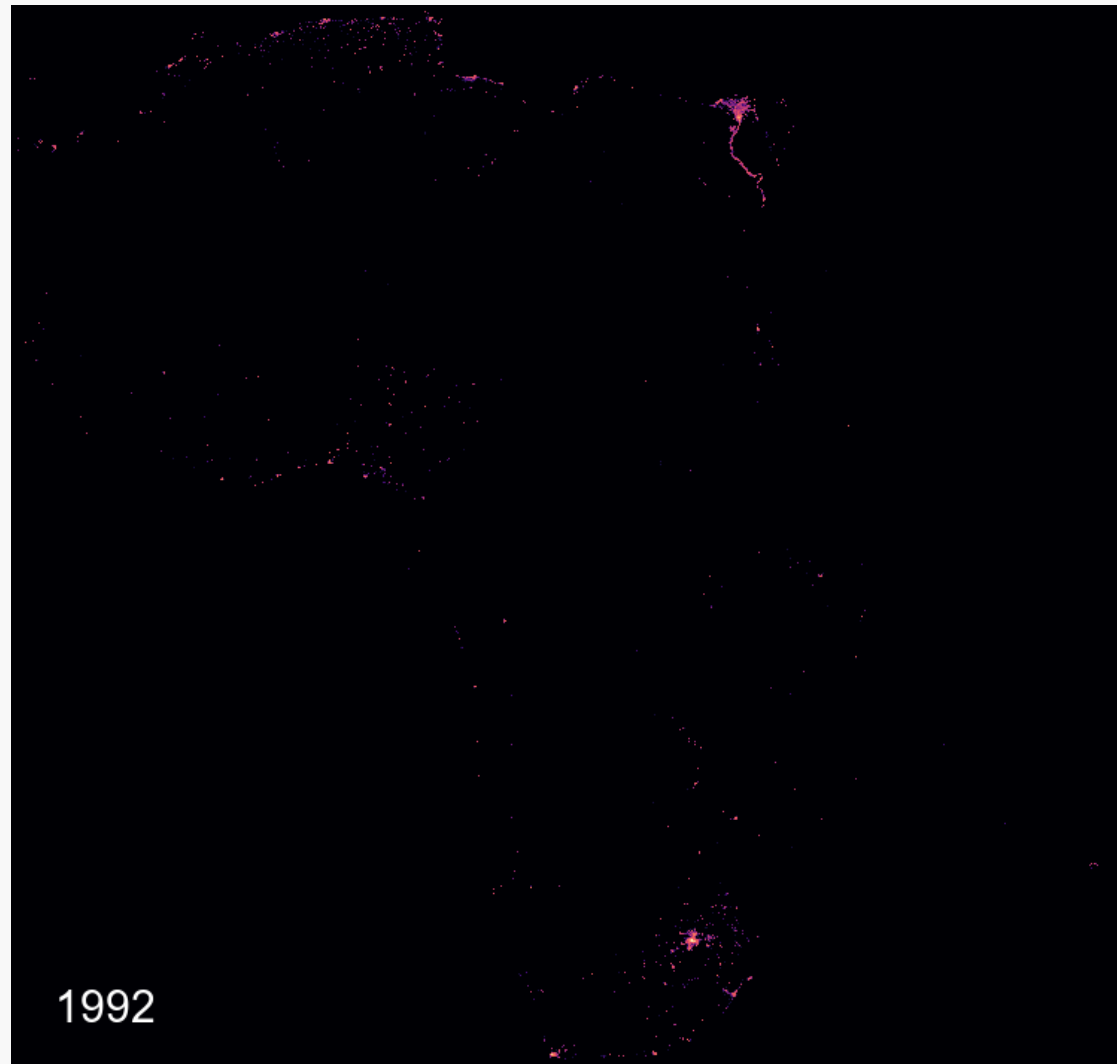


Source: NASA



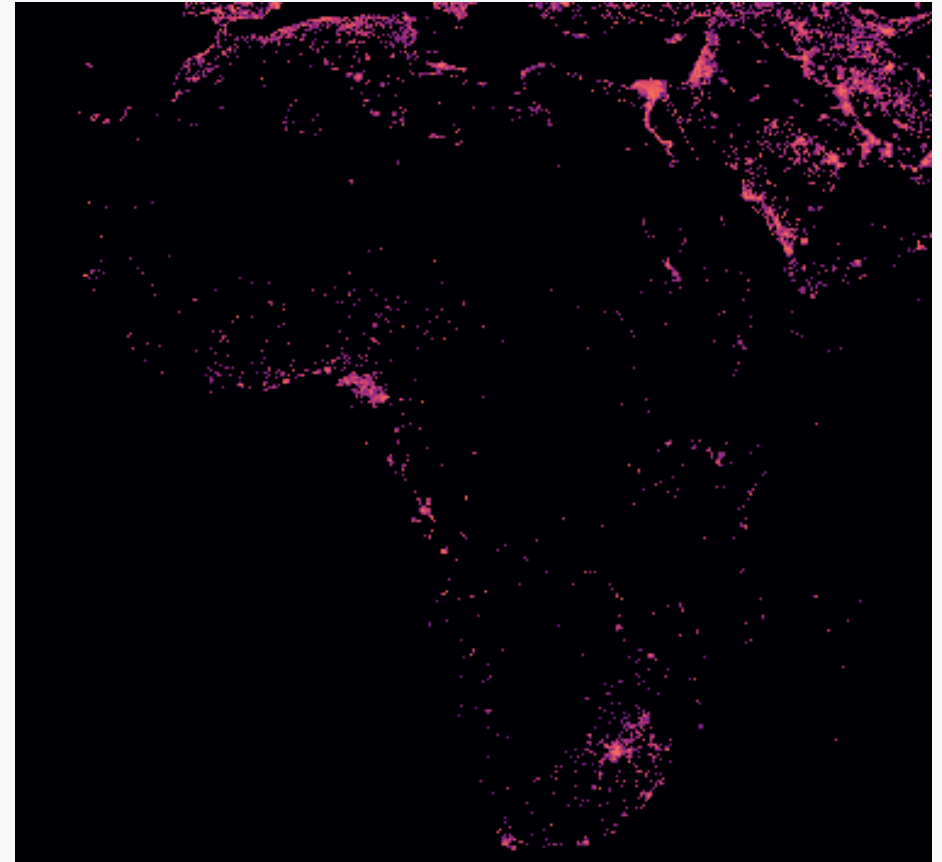
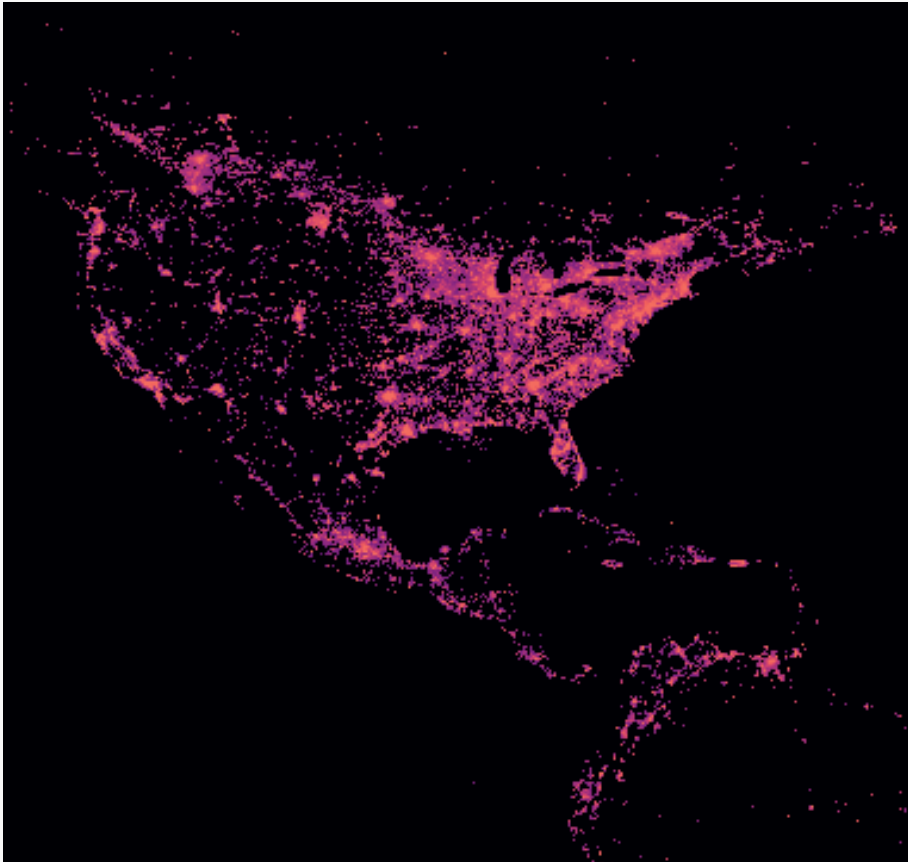
# GIS data. Luminosity / Light Density at Night

## Africa at Night



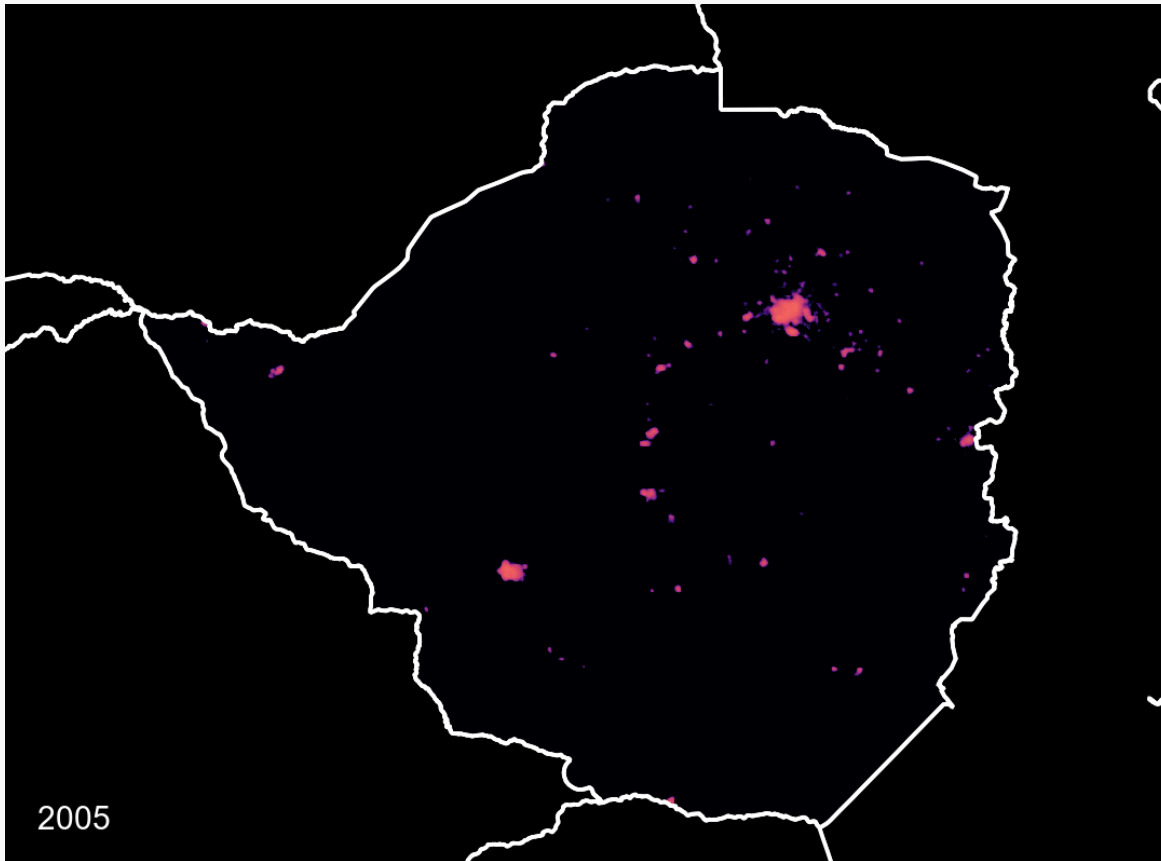
# GIS data. Luminosity / Light Density at Night

## A Contrast with the North America



# GIS data. Luminosity as Proxy of Spatial Development.

## Example 1. Zimbabwe. 2005-2010 Hyperinflation



### Hyperinflation in Zimbabwe

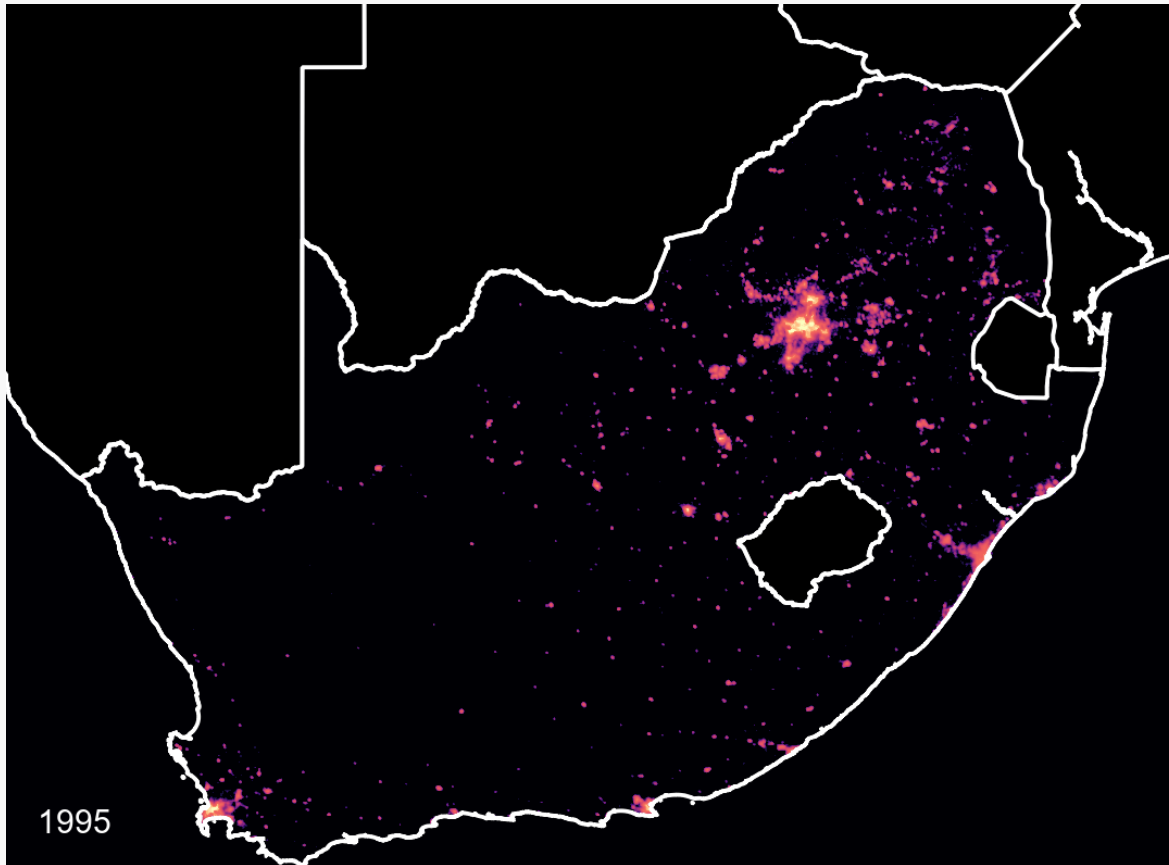
- Inflation increased rapidly, peaking in 2008 [79,600,000,000% in Nov.]
- Official statistics stop in 2008

### Nightlights show

- declining light to 2008
- Slight rebound afterwards

# GIS data. Luminosity as Proxy of Spatial Development.

## Example 2. South Africa. Democratic Era



### South Africa Post-Apartheid

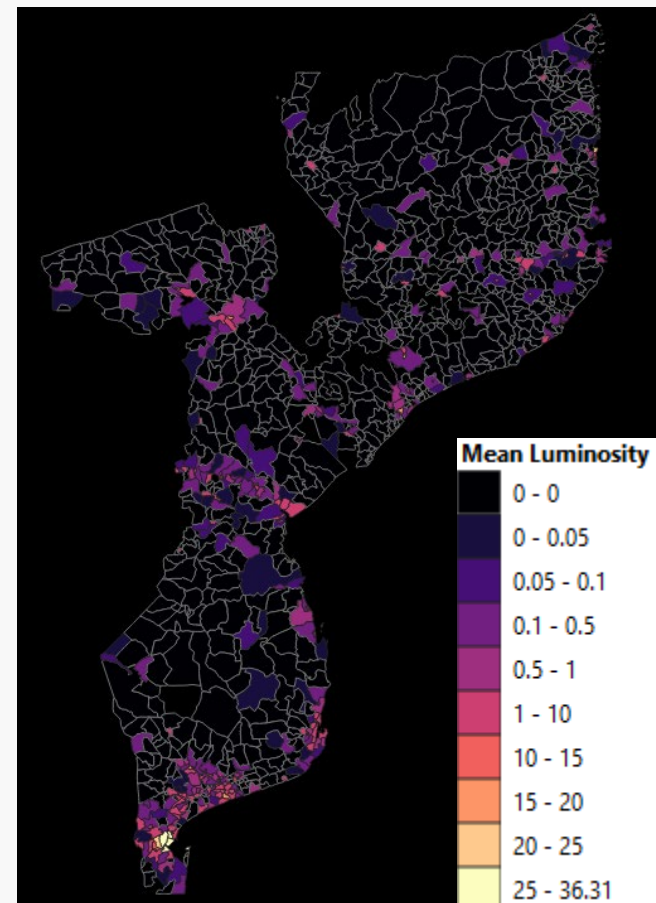
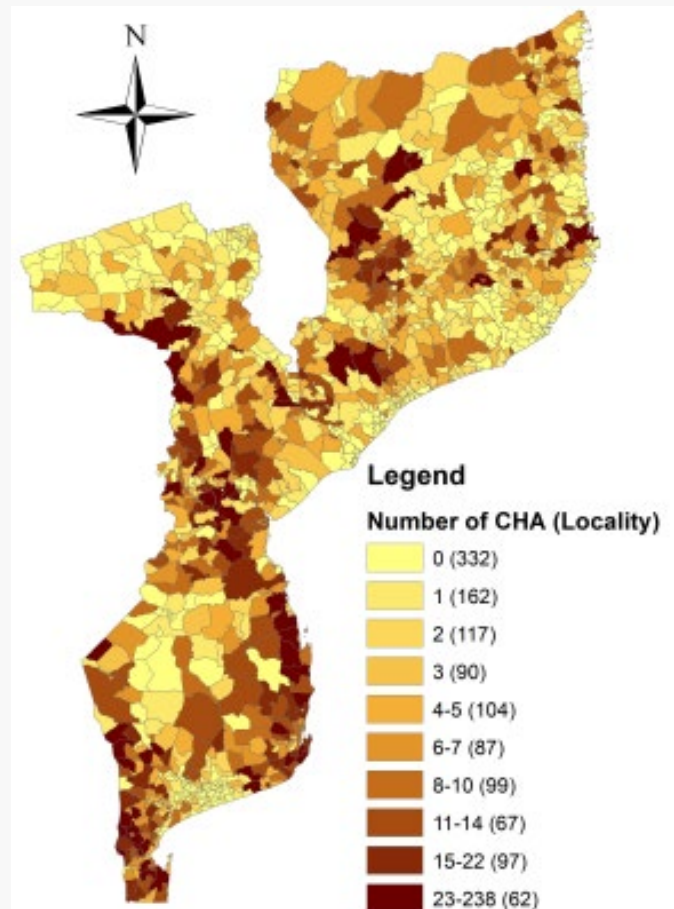
- Population growth 44% since 1995
- GDP growth 97% since 1995

### Nightlights show

- Growth of Johannesburg/Durban
- Growth in the hinterlands
- Drop in spatial inequality

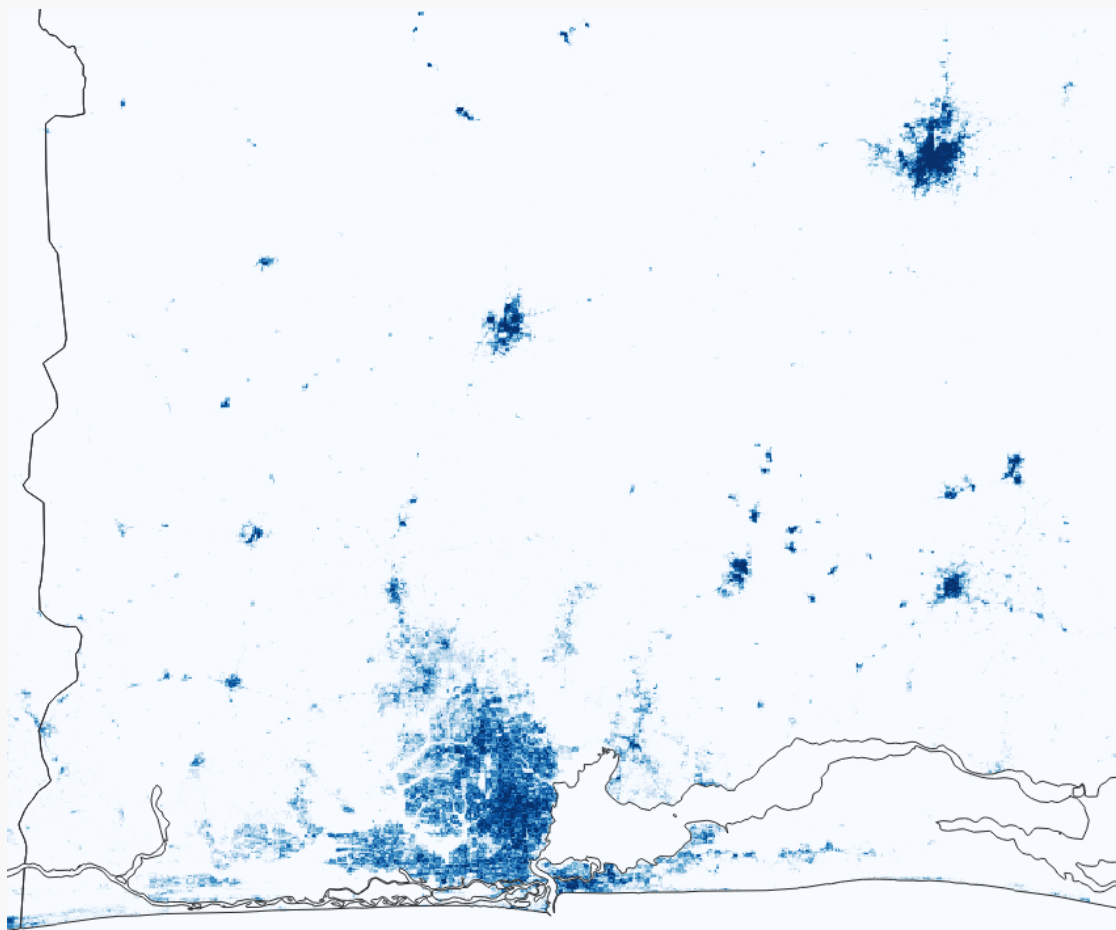
# GIS data. Luminosity Application.

[“Landmines and Spatial Development”. Chiovelli, Michalopoulos, and Papaioannou. (2021).]



# GIS data. Landcover (GHSL).

## Brief summary



1975

- **Share of land that is ‘built-up’**
  - Worldwide
  - Epochs: 1975, 1990, 2000, 2014
- **Built space at a very fine level**
  - Growth of cities and towns
  - Pattern of urban development (‘leap-frogging’)
- **Application:**
  - “Colonial Legacies: Shaping African Cities”. Baruah, Henderson, and Peng (JEGeo 2017).
- **Source: European Commission**
  - Download at [GHSL](#)

# Surveys. Demographic and Health Survey (DHS). Overview

## — Aspects:

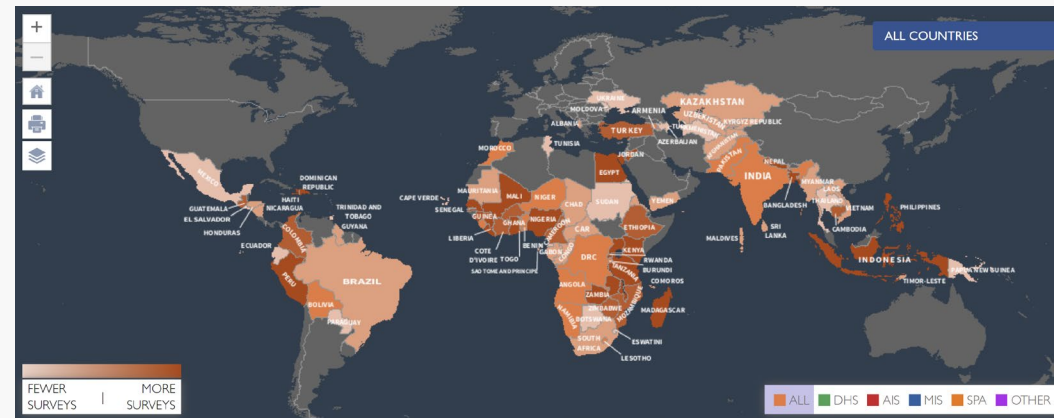
- Fertility, education, household structure, assets and wealth, immunization, access to public goods, ...

## — Coverage (see map)

- Samples of the population for a given country and year [from 1985]

## — Source: USAID, DHS Program

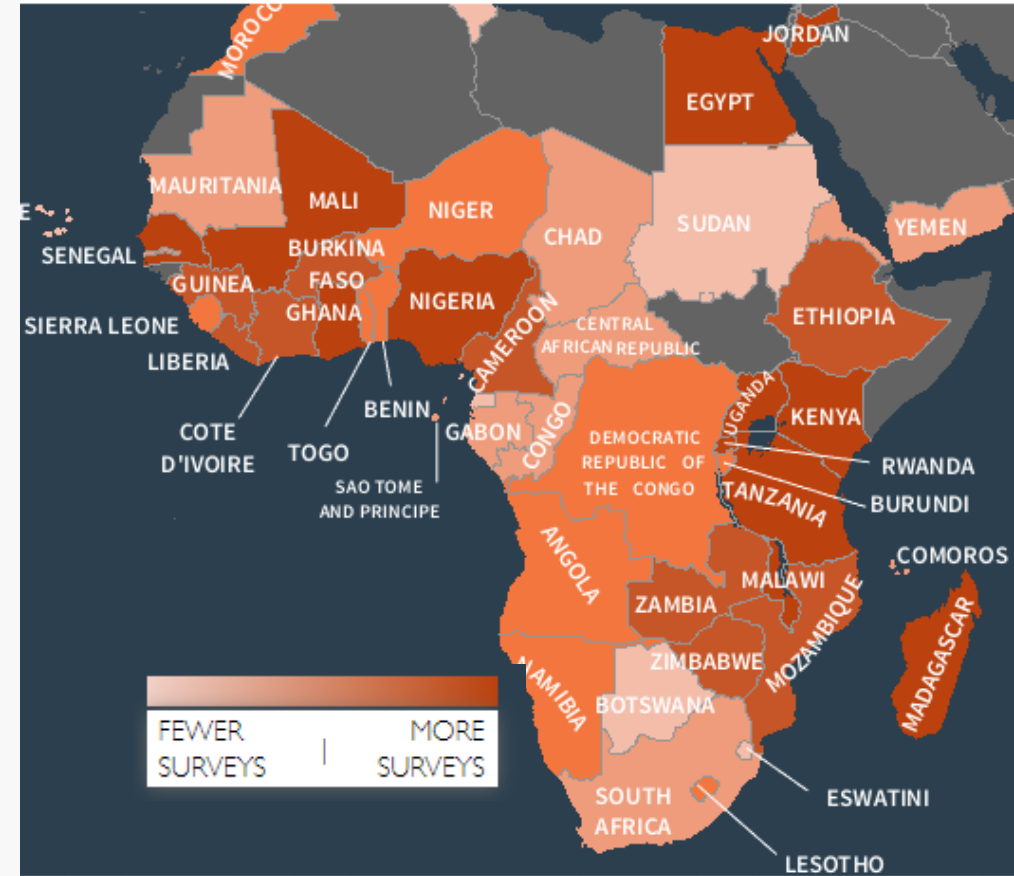
- Download at the [DHS Program](#)



# Surveys. Demographic and Health Survey (DHS).

## Caveats

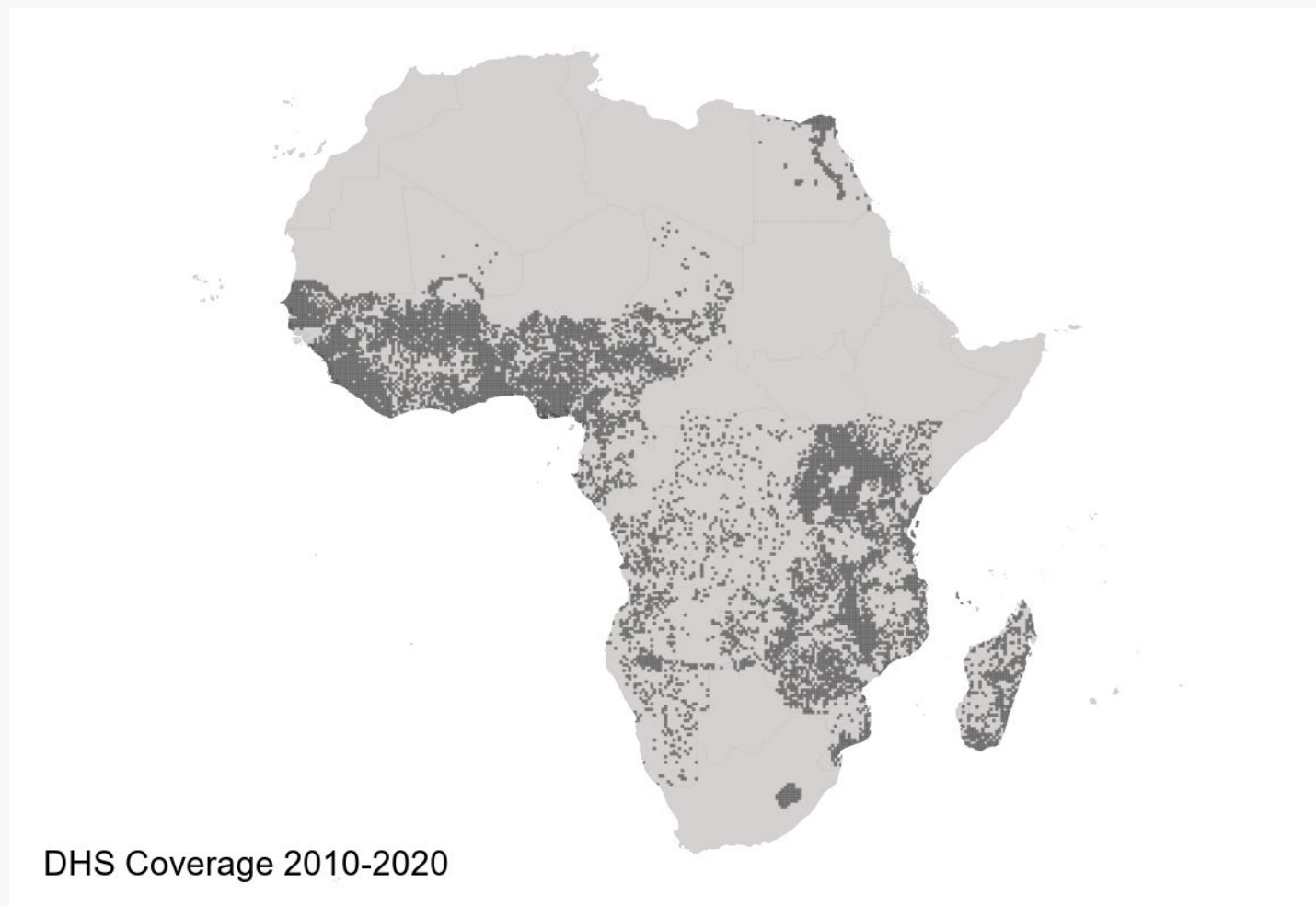
- Same locations not necessarily repeated over time
- Sample may differ (only female in some year, all citizens in others, not all surveys give location)
- Oversampling of proximate to capitals and other big cities areas





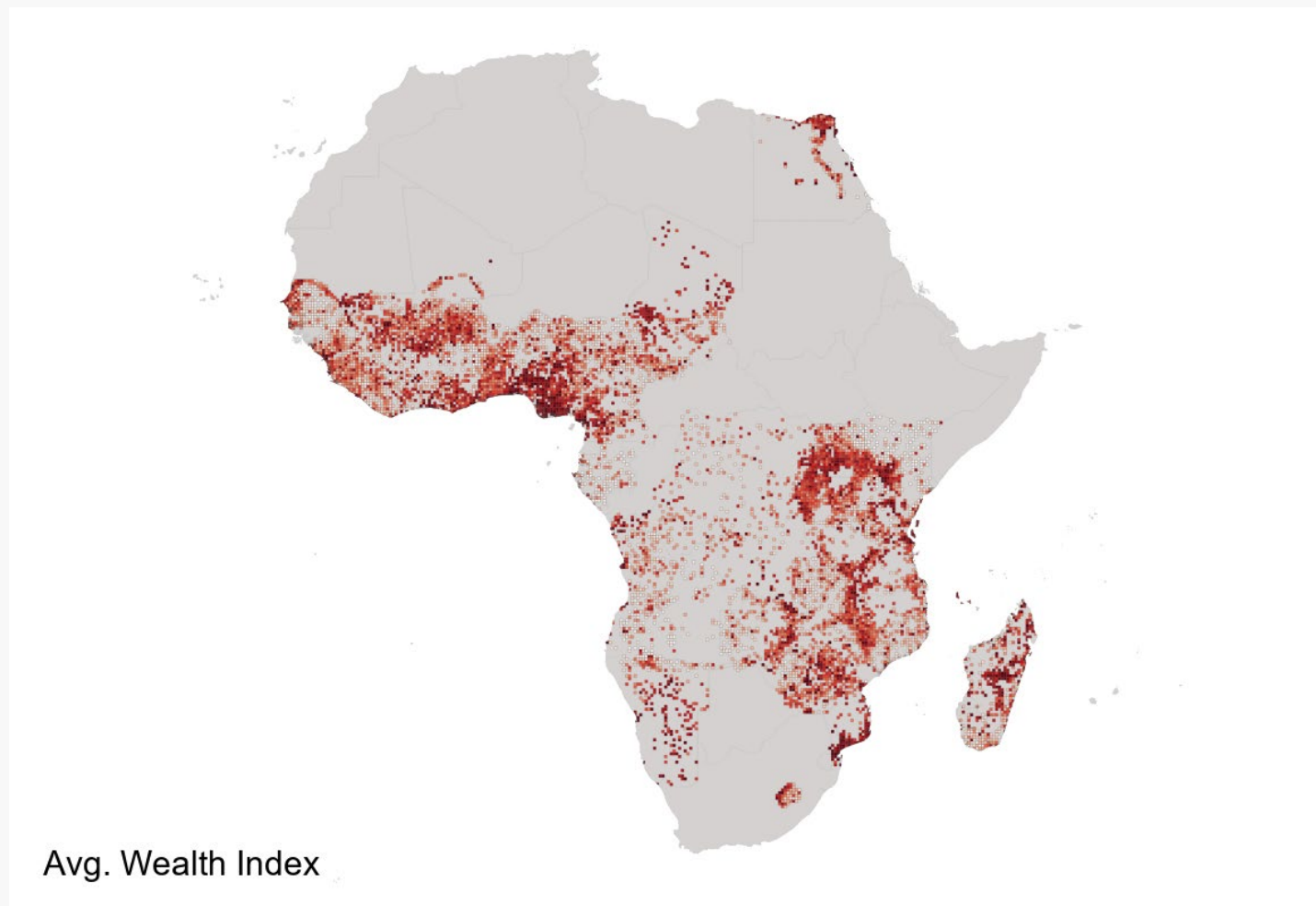
# Surveys. Demographic and Health Survey (DHS).

## Example 1. Pan-African coverage.



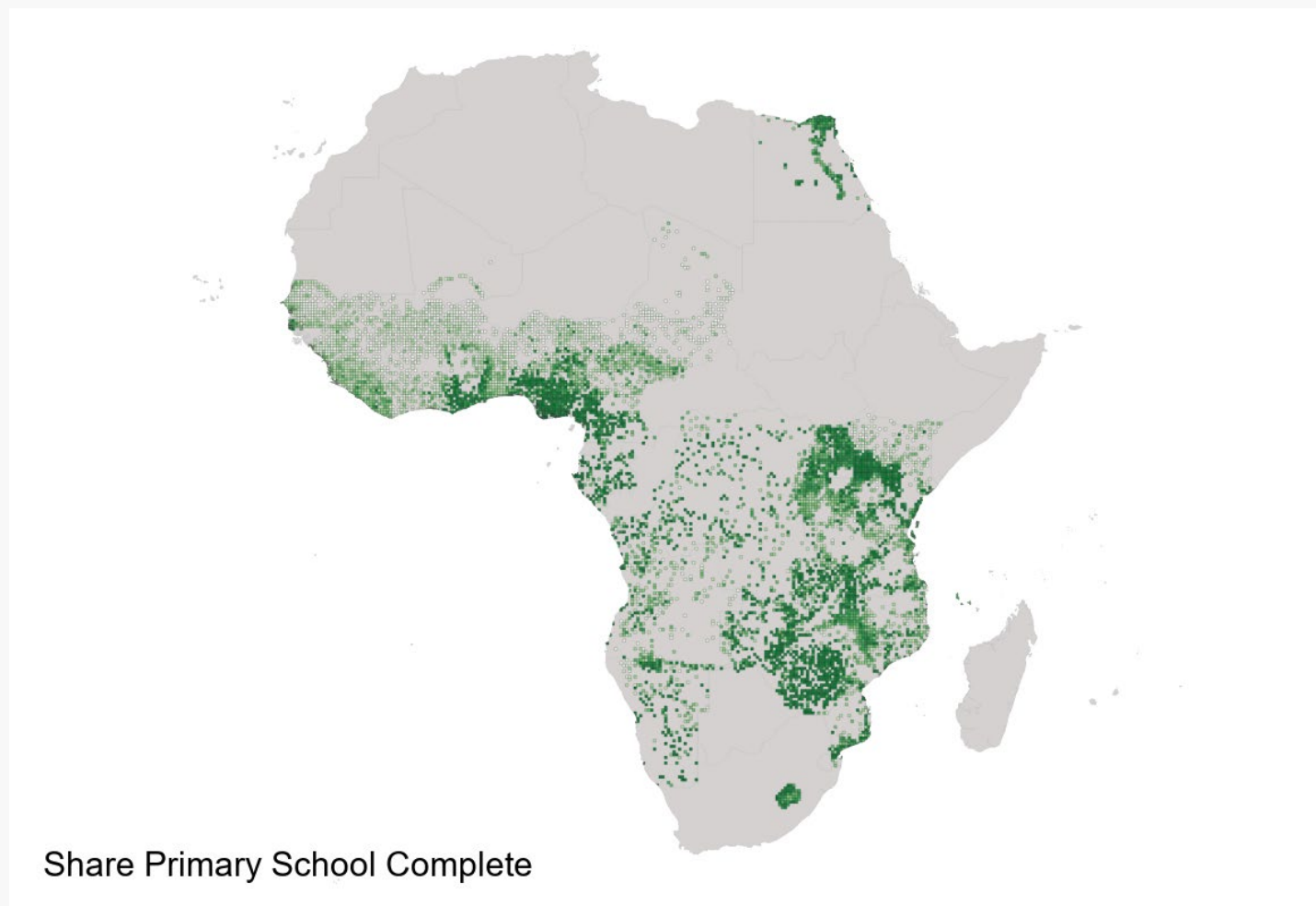
# Surveys. Demographic and Health Survey (DHS).

## Example 1. Pan-African coverage: Wealth.



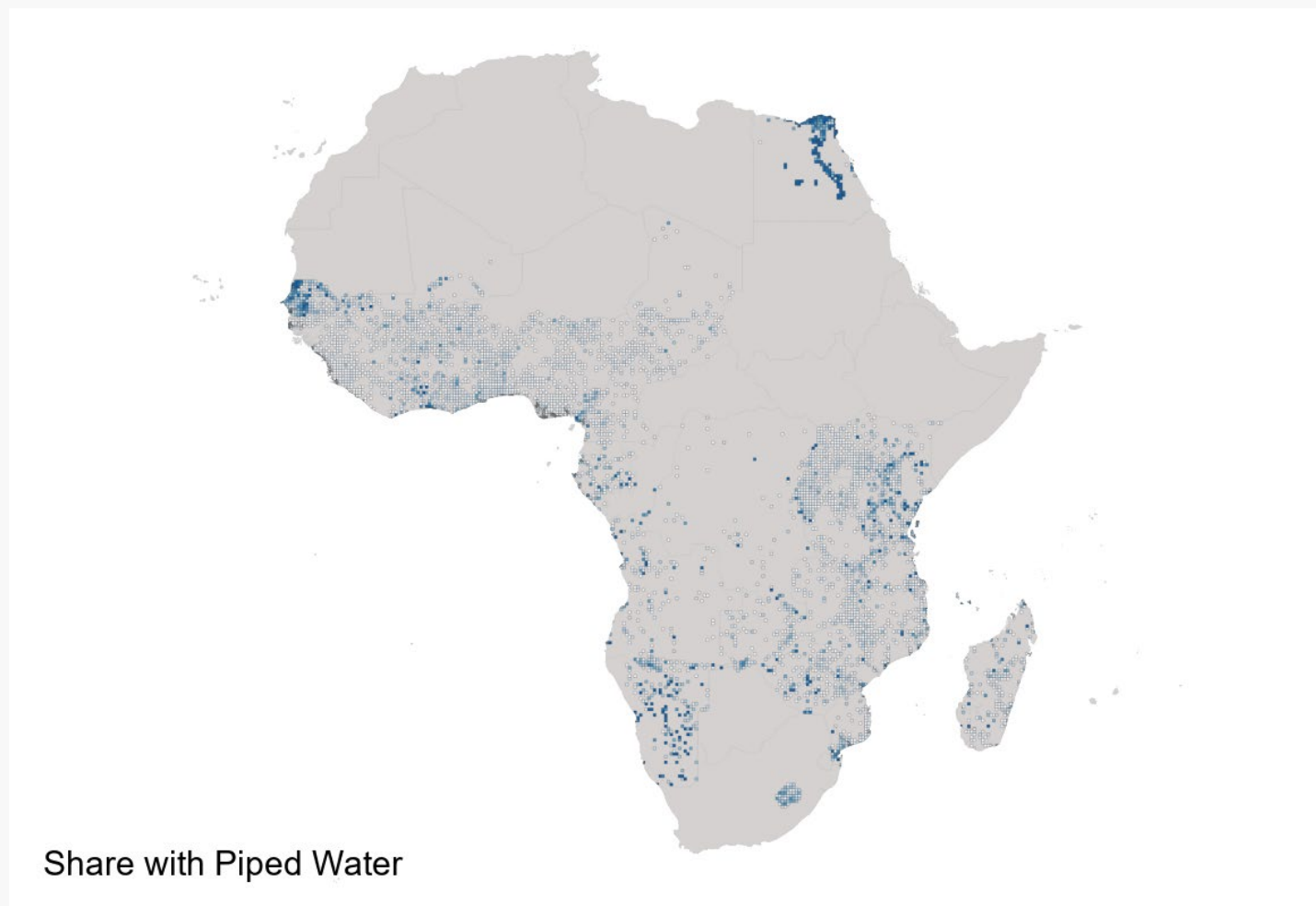
# Surveys. Demographic and Health Survey (DHS).

## Example 1. Pan-African coverage: Education.

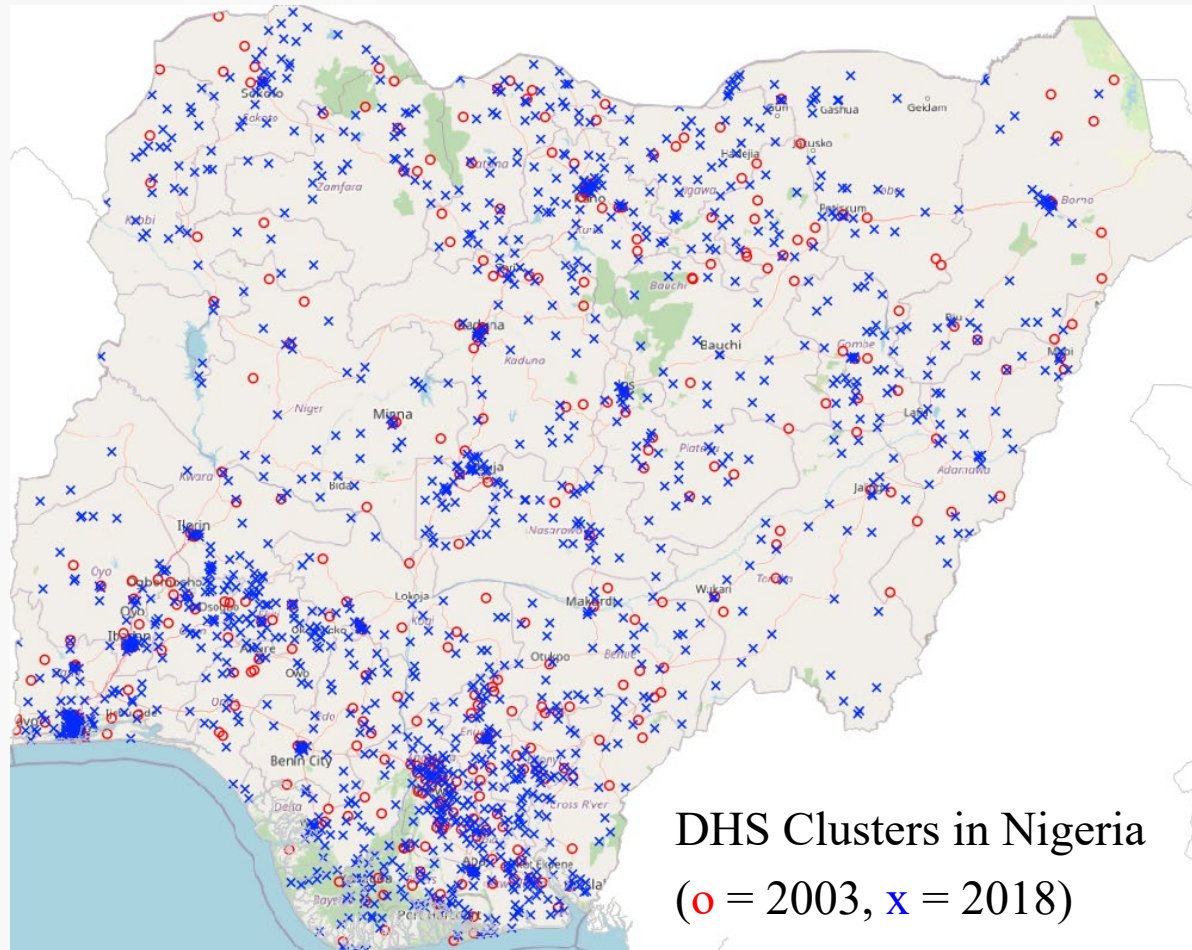


# Surveys. Demographic and Health Survey (DHS).

## Example 1. Pan-African coverage: Water Access.



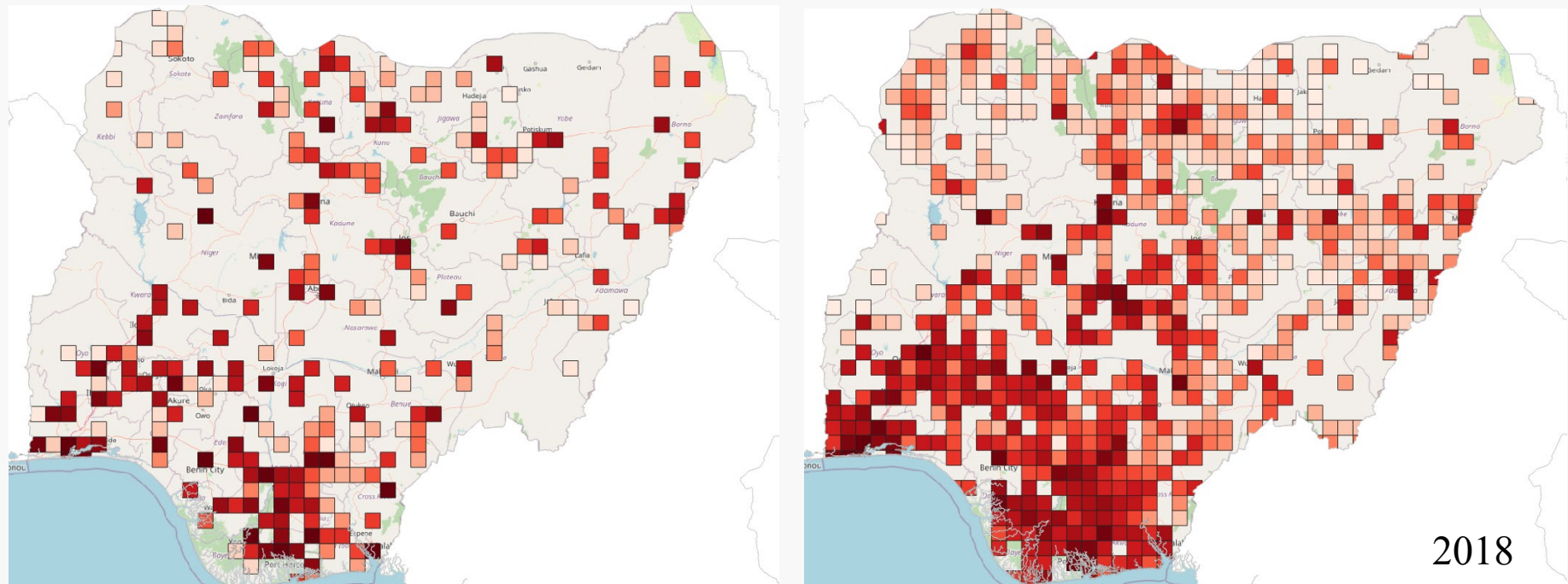
# Surveys. Demographic and Health Survey (DHS). Example 2. Nigeria coverage.





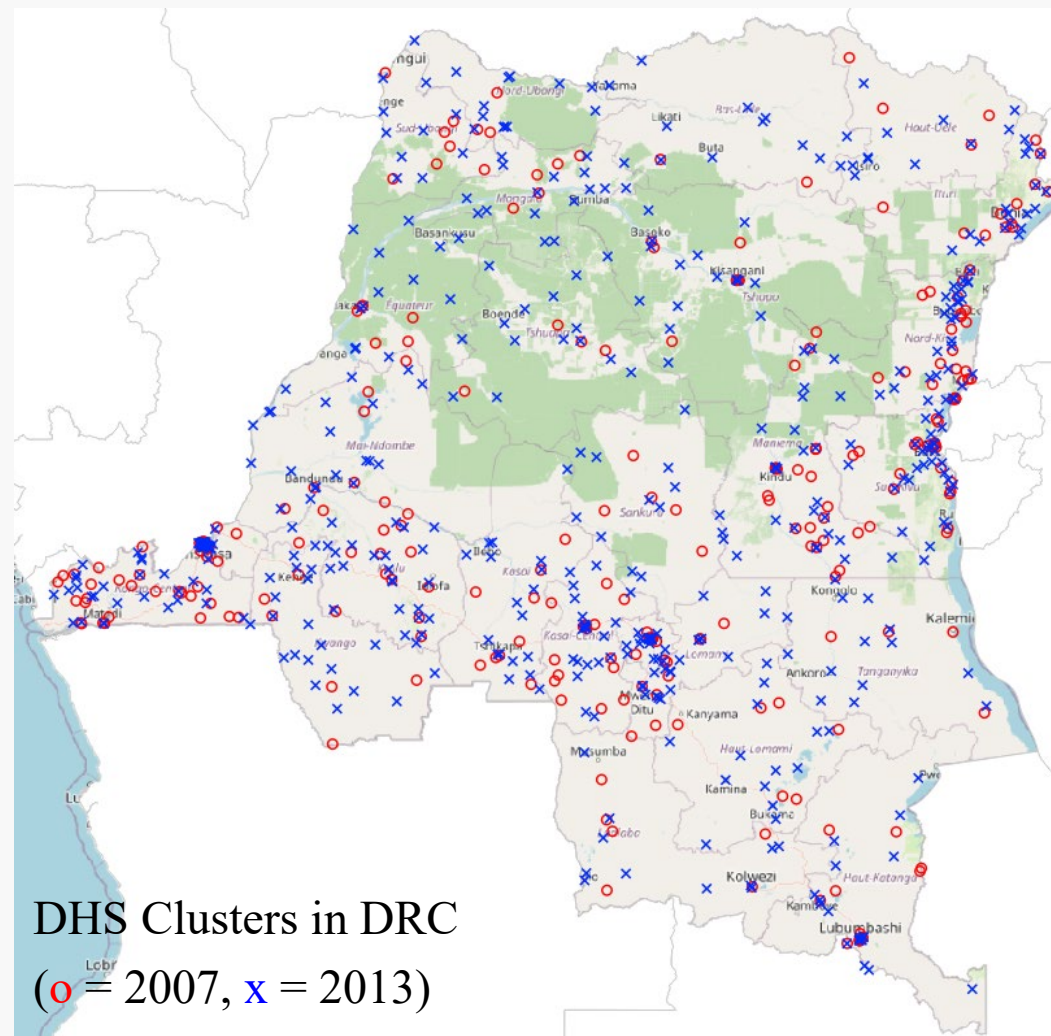
# Surveys. Demographic and Health Survey (DHS).

## Example 2. Nigeria Composite Wealth Index.



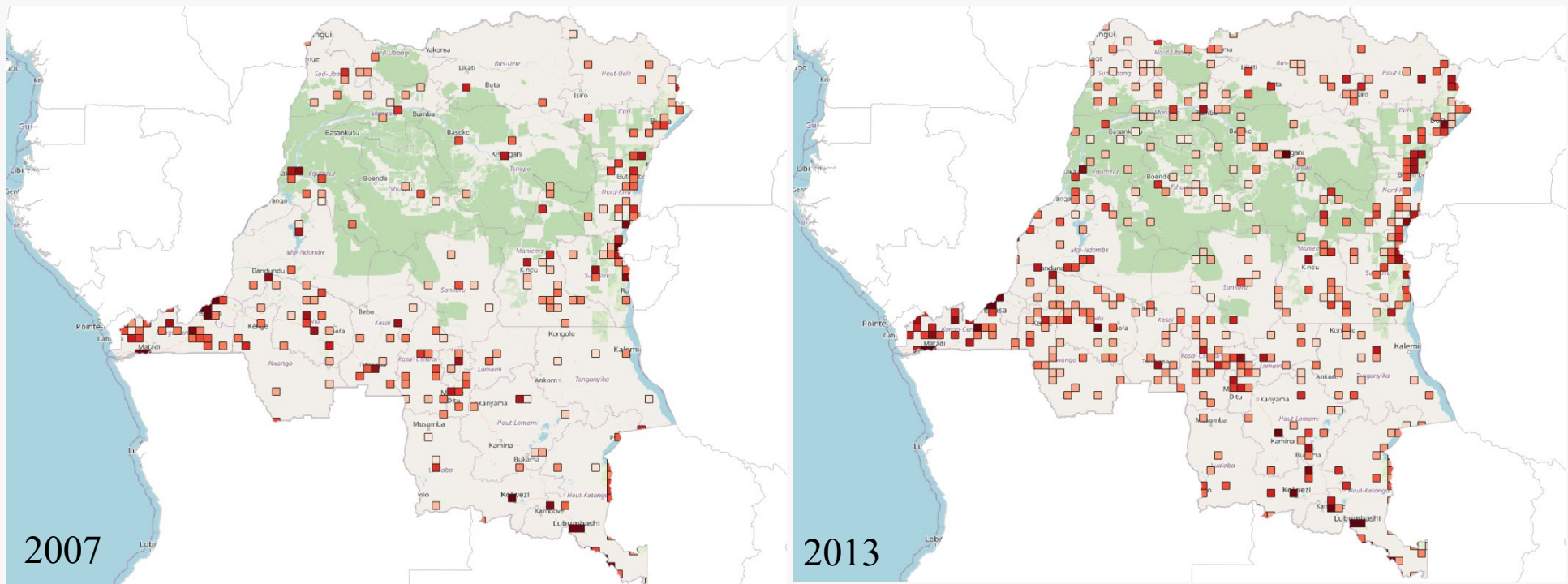
# Surveys. Demographic and Health Survey (DHS).

## Example 3. DRC coverage.



# Surveys. Demographic and Health Survey (DHS).

## Example 3. DRC Composite Wealth Index.

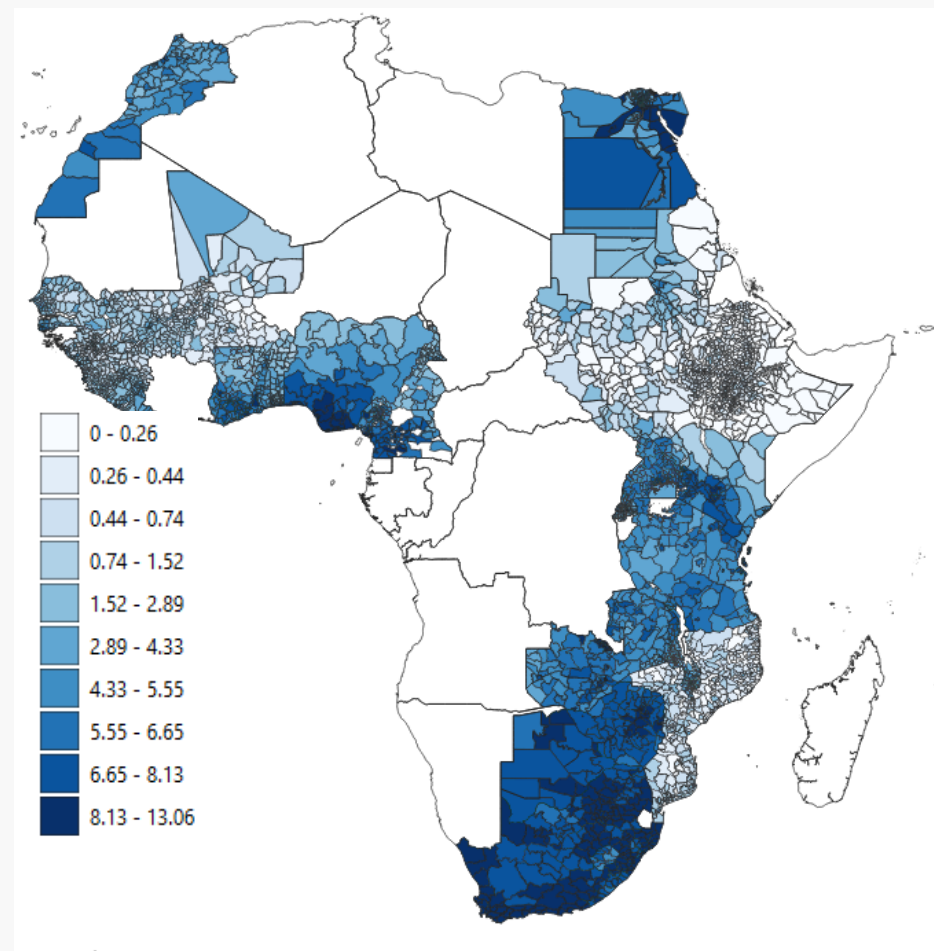




# Admin data. IPUMS.

## Overview

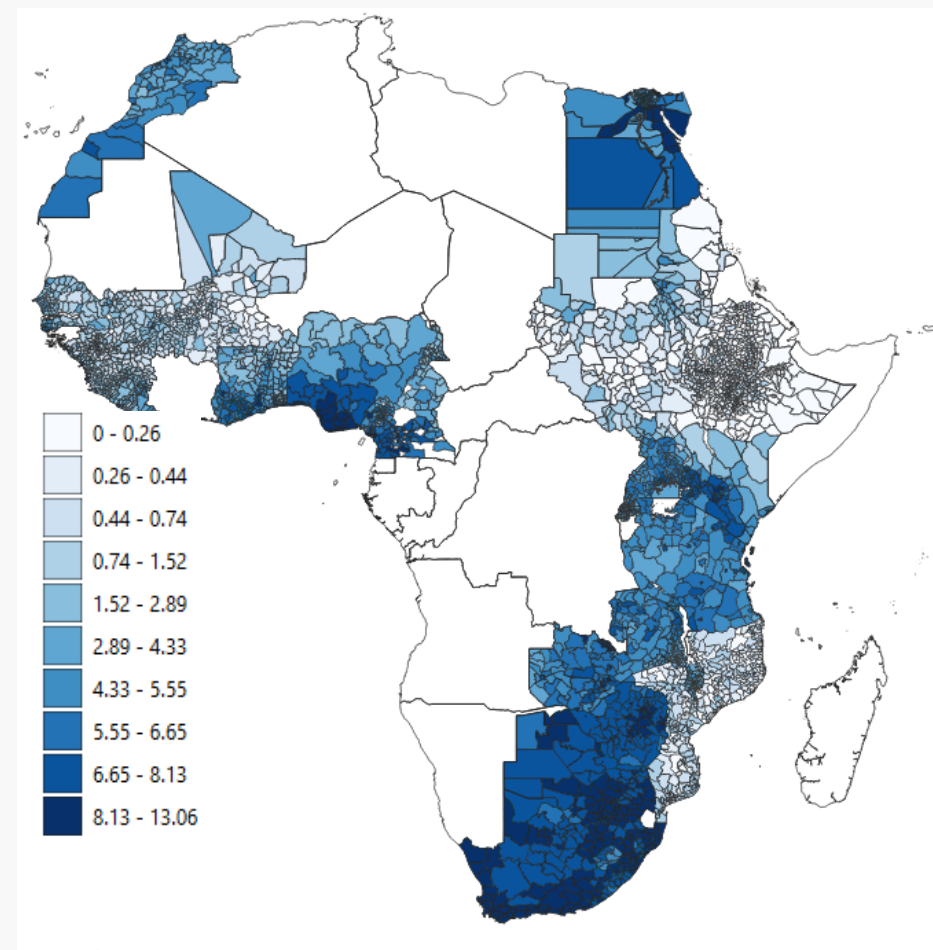
- “Integrated Public Use Microdata Series”
- Standardized censuses from many countries [global coverage]
- Aspects:
  - Education, household structure, employment, birthplace and nativity, births and deaths, housing materials.
- Repeats every 10 years (typically)
- Source: University of Minnesota
  - Download at [IPUMS international](https://www.ipums.org/)



# Admin data. IPUMS.

## Caveats

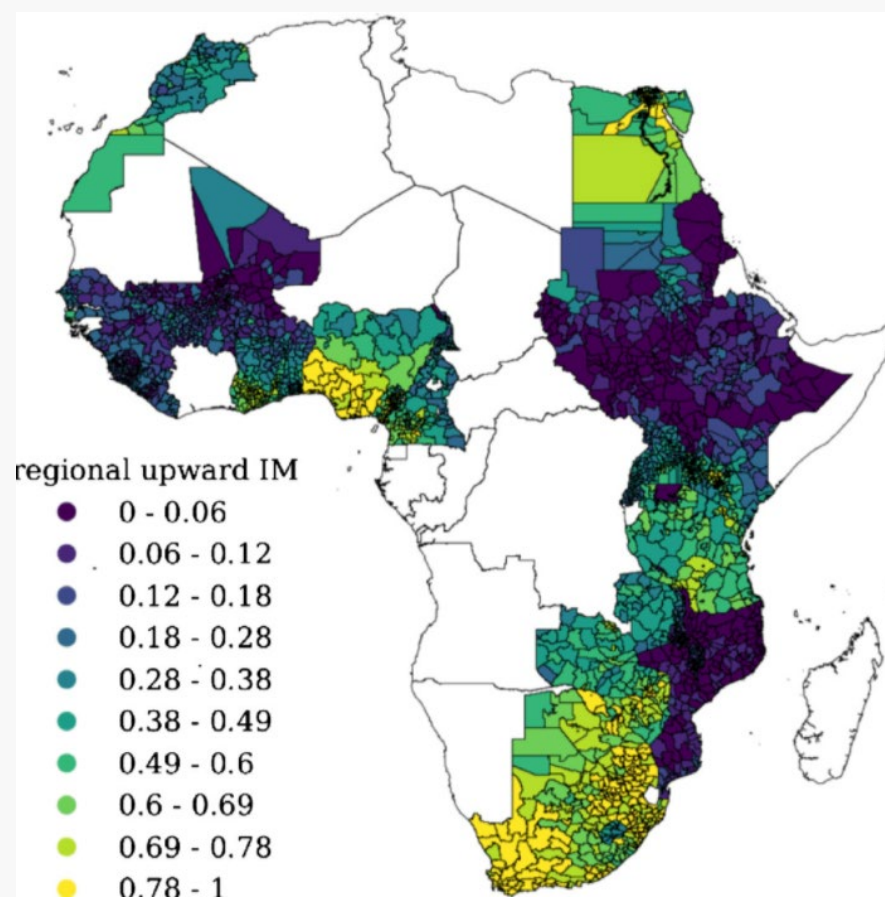
- **Sampled (~10%) from censuses**
- **Not many censuses for many countries**
- **Sometimes loses details from the country censuses**
  - E.g. location of household
- **Not *everything* is standardised**
  - E.g. locations and their geographical boundaries



## Admin data. IPUMS.

### Example 1. Intergenerational Mobility

[“Intergenerational Mobility in Africa”. Alesina, Hohmann, Michalopoulos, and Papaioannou. (2021).]



#### — Intergenerational mobility in education

- household structure → how well children do relative to their parents

#### — Effect of places

- Children of uneducated parents more likely to complete school if mobility in their region is high



# Admin data. Country Censuses.

## Caveats

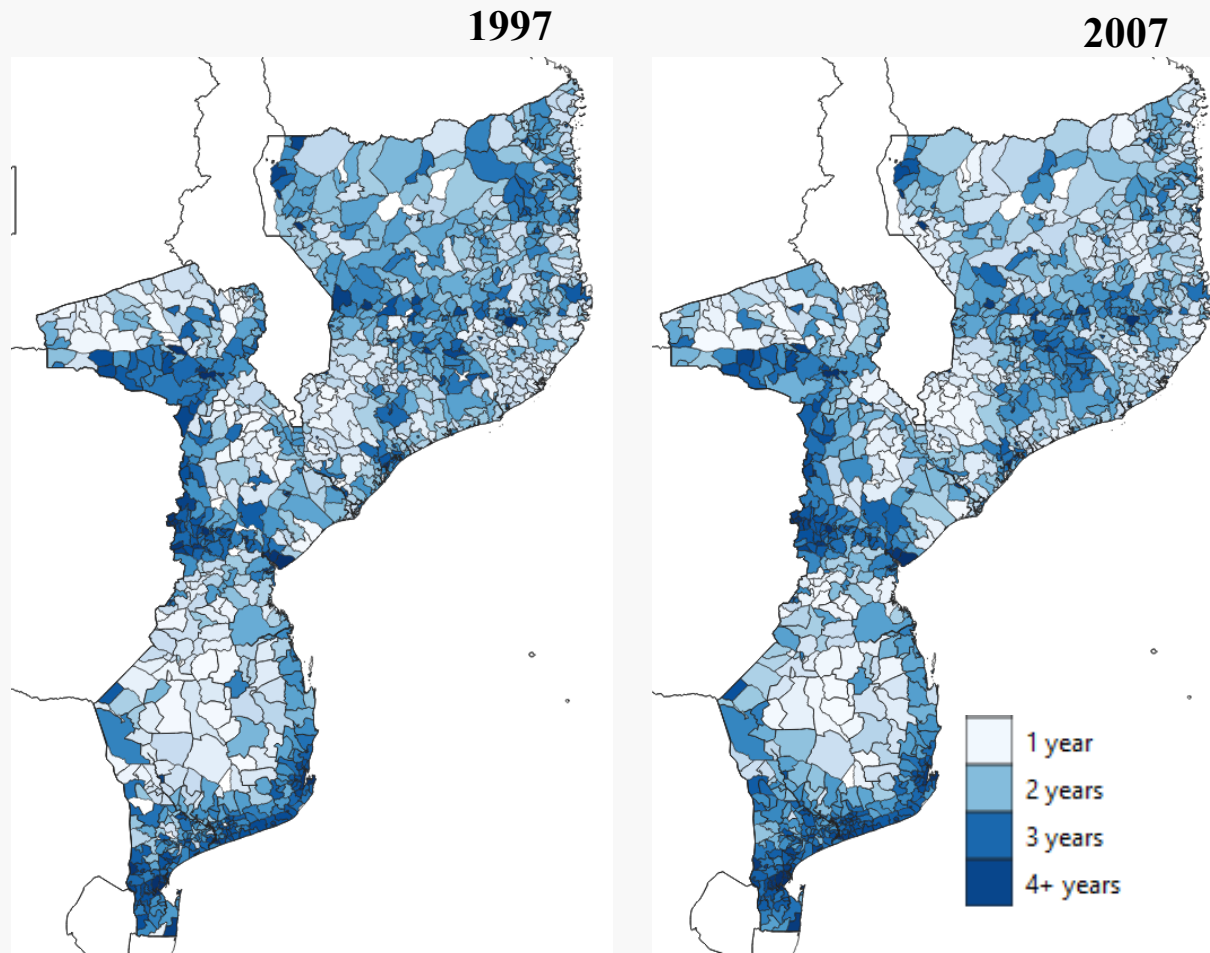
- Spatially coarse
- Infrequent (each 10 years)
- Not standardized across countries
- Can be difficult to gain access (not publicly available)





# Admin data. Country Censuses.

## Example 1. Education in Mozambique



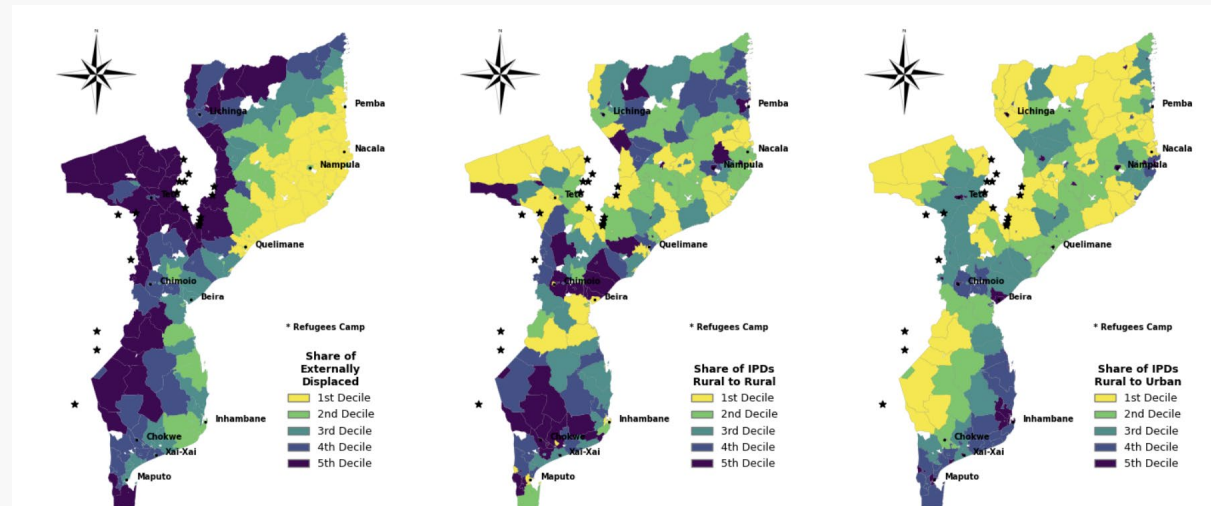
— Mean years of school for adults 15+

- Education is rising
- Some areas gain
  - Some losing

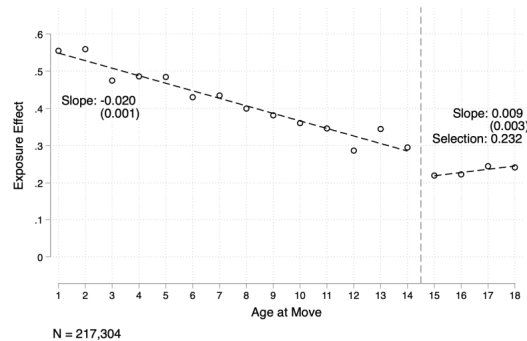
# Admin data. Country Censuses.

## Example 2. Forced displacement in Mozambique

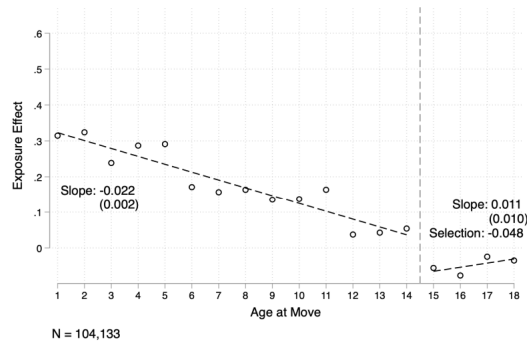
- Reconstructing migration matrix based on birth and residency
- Linking timing of conflict exposure at cohort level to retrieve displacement



Forced Displacement by District of Birth



Panel C: Moved  $\pm 2$  Years of Conflict



Panel D: Top 50% and Moved  $\pm 2$  Years of Conflict

Source: Chiovelli, G., Michalopoulos, S., Papaioannou, E., & Sequeira, S. (2021). Forced Displacement and Human Capital: Evidence from Separated Siblings (No. w29589). National Bureau of Economic Research.

# GIS data. Elevation (SRTM).

## Brief summary

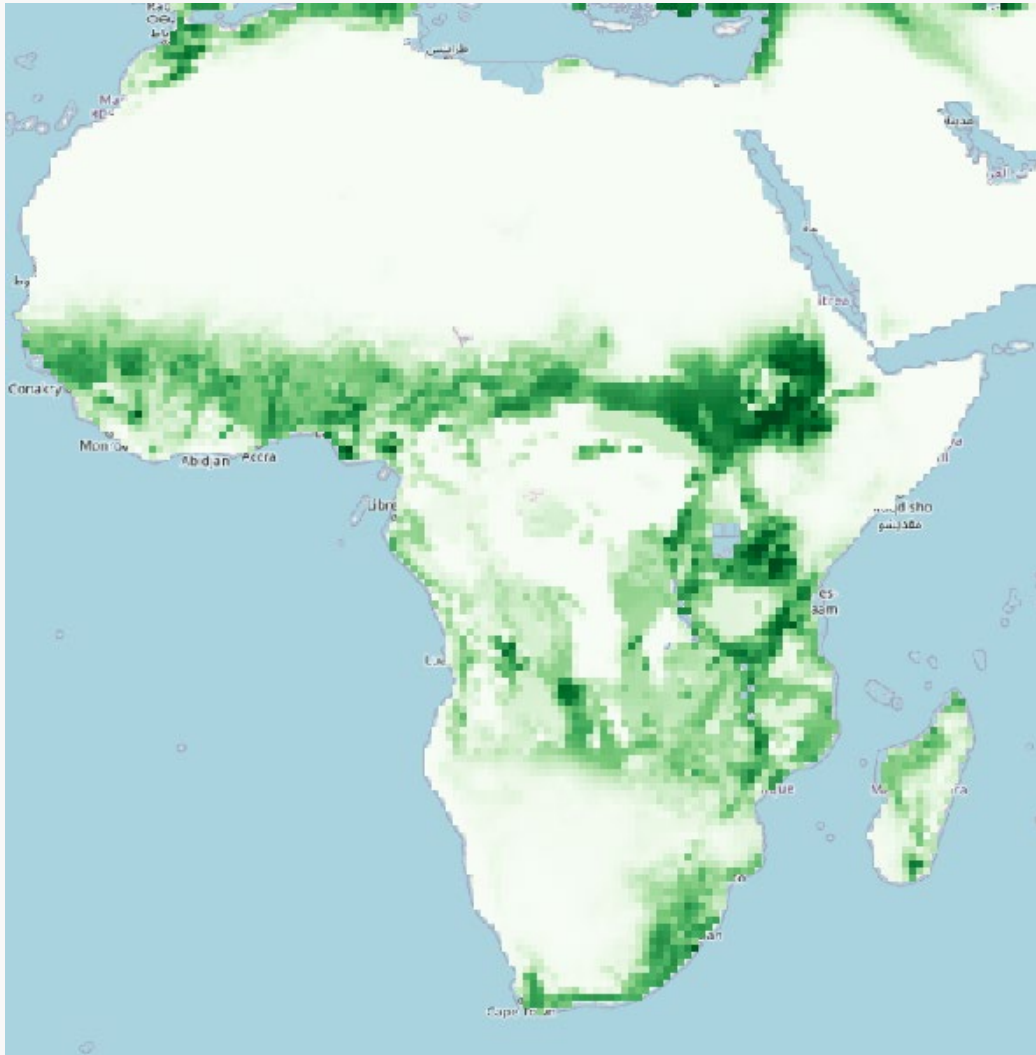


- **Height above sea level**
  - Worldwide
  - One point in time
- **Measure features of the terrain**
  - Important for agriculture, flooding, conflict, etc.
- **Application:**
  - “Ruggedness: The Blessing of Bad Geography in Africa”. Nunn and Puga (REStat 2017).
- **Source: USGS**
  - Download from the [EROS Center](#)



# GIS data. Agricultural Suitability.

## Brief summary



- **Fraction of land suitable for agriculture.**
  - Worldwide
  - One point in time
- **Application:**
  - “The Origins of Ethnolinguistic Diversity”. Michalopoulos (AER 2012).
- **Source: University of Wisconsin-Madison**
  - Download from the [CSGE](#)

# Crop-Specific Potential Yield and Suitability. FAO GAEZ.

- The Global Agroecological Zones (GAEZ)
- Agronomic-based knowledge to evaluate land suitability and potential yields (in tons) for 51 crops.
- 9km x 9km grid cells, with data providing information on suitability and potential yield for each main crop at this level.
- Different scenarios: rain-fed to irrigated agriculture.
- Production capacity determined solely by agro-climatic conditions [unaffected by changes in agricultural production technology.]



Food and Agriculture Organization  
of the United Nations

| GAEZ Data Portal

Download [FAO GAEZ](#)

# Crops Growing Season. MIRCA

- Monthly Irrigated and Rainfed Crop Areas
- Monthly growing areas of 26 irrigated and rainfed crops with related crop calendars for 402 spatial units around the world
- Differentiate between Northern and Southern hemisphere
- Download [MIRCA](#)

# Other data. High Resolution Satellite Images (VHR).

## Summary

Kibera, Nairobi VHR image for 2004 (left) and 2015 (right).



Kibera, Nairobi classified buildings for 2004 (left) and 2015 (right)  
- unchanged (blue), demolished (red), redeveloped (green).

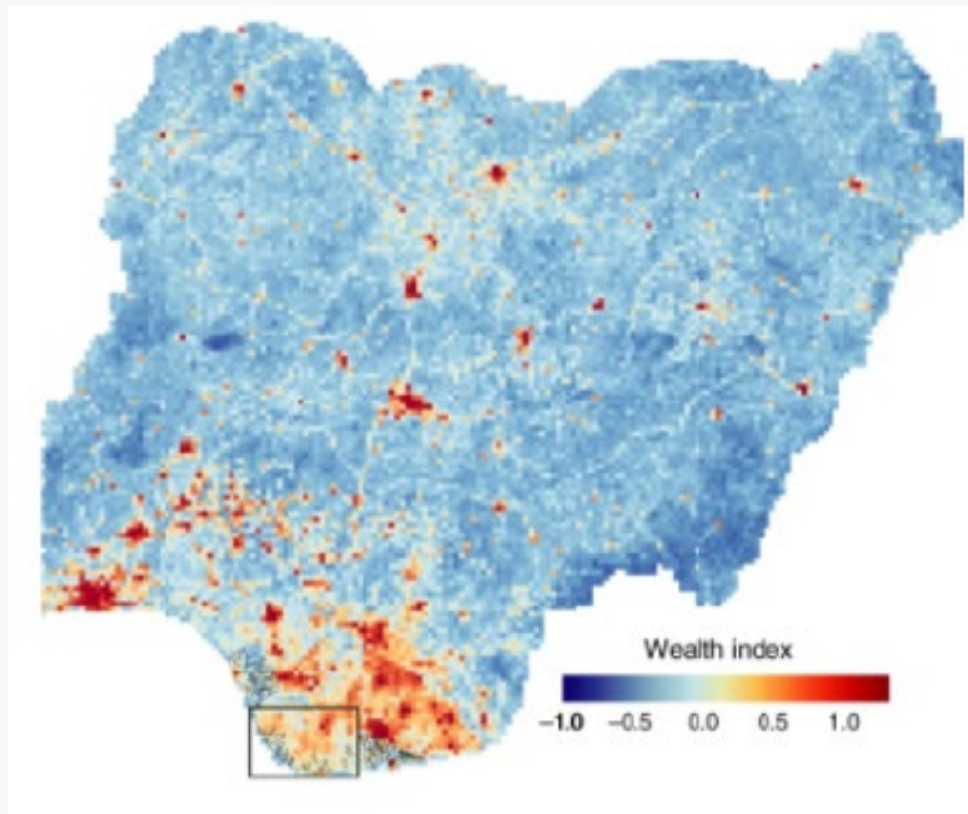


- **Modern satellites can take very accurate photos**
  - Measure build capital allocation at a very fine scale
- **Caveats**
  - Data is expensive
  - Difficult computationally to scale across broad areas
- **Application**
  - “Building the City: From Slums to a Modern Metropolis”. Henderson, Regan and Venables (RES 2021).

# Other data. Modelled wealth and income.

## Summary

### Predicted wealth across Nigeria



- **Machine learning + GIS + survey data**
  - Complete coverage spatially and annually
  - Asset wealth/Spending
- **Caveats**
  - Proprietary data
  - Not easily implementable from first principles
- **Implementation**
  - “Using publicly available satellite imagery and deep learning to understand economic well-being in Africa”. Yeh et al. (Nature 2020).



# Structure

## 1. Introduction: Climate and Conflict

## 2. Georeferenced and Satellite Data

- Why GIS
- Data Format: Vector and Raster data
- Softwares
- Google Earth Engine
- Common operations

## 3. Main Conflict and Climate Datasets

- Traditional Conflict Data: UCDP-GED and ACLED
- Big conflict data: ICEWS and GEDTL
- Climate Data: Rainfall, Temperature, Droughts

## 4. Additional Data:

- Economic Activities
- Urbanization
- Education and Wealth
- Migration/Displacement
- Agriculture
- Geography

## 5. Summary and Takeaways

# Summary

## — This has been a non-exhausting overview

- GIS [Data types and softwares]
- Conflict and Climate Data
- Additional auxiliary on Development, Mobility, Agriculture, etc

## — There are many data out there

- Can you think of data sources not mentioned here?
- What kind of questions are poorly answered by typical data in Economics?
- Can you find novel data sources to fill that gap?