GIS for Economists 1

Giorgio Chiovelli, Sebastian Hohmann

16/03/2020

Table of contents

Introduction

What are these sessions about?

Three aims

- Give you a (selective) overview of how *Geographic Information Systems* (GIS) methods are used in current economic research.
- Introduce you to QGIS software and GIS scripting in python.
- Show you how you can "do GIS" for your own research through detailed examples of well published papers.
- **If time**: show you how to process large-scale satellite imagery in the cloud with Google Earth Engine.

We want to make this useful for you!

- If you bring your laptop with QGIS installed, you can follow what we are doing in class.
- Depending on whether you have concrete questions, we can leave the last 20-30 minutes at the end of each session to try to answer them.

What we will do in each session

1. (today) Overview of GIS applications in economic research, setting up QGIS, simple examples
2. (today) Introduction to python
3. (17/03/2020) Replication of Michalopoulos, Stelios. (2012). "The Origins of Ethnolinguistic Diversity," American Economic, Review, 102(4): 1508-1539.
4. (17/03/2020) Selected topics: geocoding via google / geonames API, looping over raster data, multidimensional data, interpolation, hot spot analysis, digitizing an old map
5. (18/03/2020) Network analysis in GIS. Replication of Hornbeck, Richard, and Dave Donaldson. (2016). "Railroads and American Economic Growth: A Market Access Approach," Quarterly Journal of Economics (2016), 131(2): 799-858.
6. (18/03/2020) Optional: Intro to Google Earth Engine

Material for programming. Google drive

Introduction

Why use GIS in Economic Research?

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
- old kingdoms
- artificial units

More credible identification strategies. Examples include

- account for many geographic covariates
- create instruments (distance from certain locations, like Mainz (spread of printing press) from Wittenberg (spread of Protestantism))
- conduct spatial RD-design exploiting historical accidents (Mita in Peru, colonial border drawing in Africa, extent of historical Empires in Europe etc.)
- Explore level at which pattern uncovered prevails. Invariance to spatial aggregation? This is where virtual country specifications (see lecture 3) are useful

Based on Stelios Michalopoulos's slides on GIS for Economic History

Paper examples: satellite data

Night lights 1: "Pure measurement". Henderson, Storeygard and Weil (2012)

J. Vernon Henderson, Adam Storeygard David N. Weil (2012) "Measuring Economic Growth from Outer Space," AER 102(2): 994–1028.

Motivation

Income / output is poorly measured, especially in developing countries and at small scales.

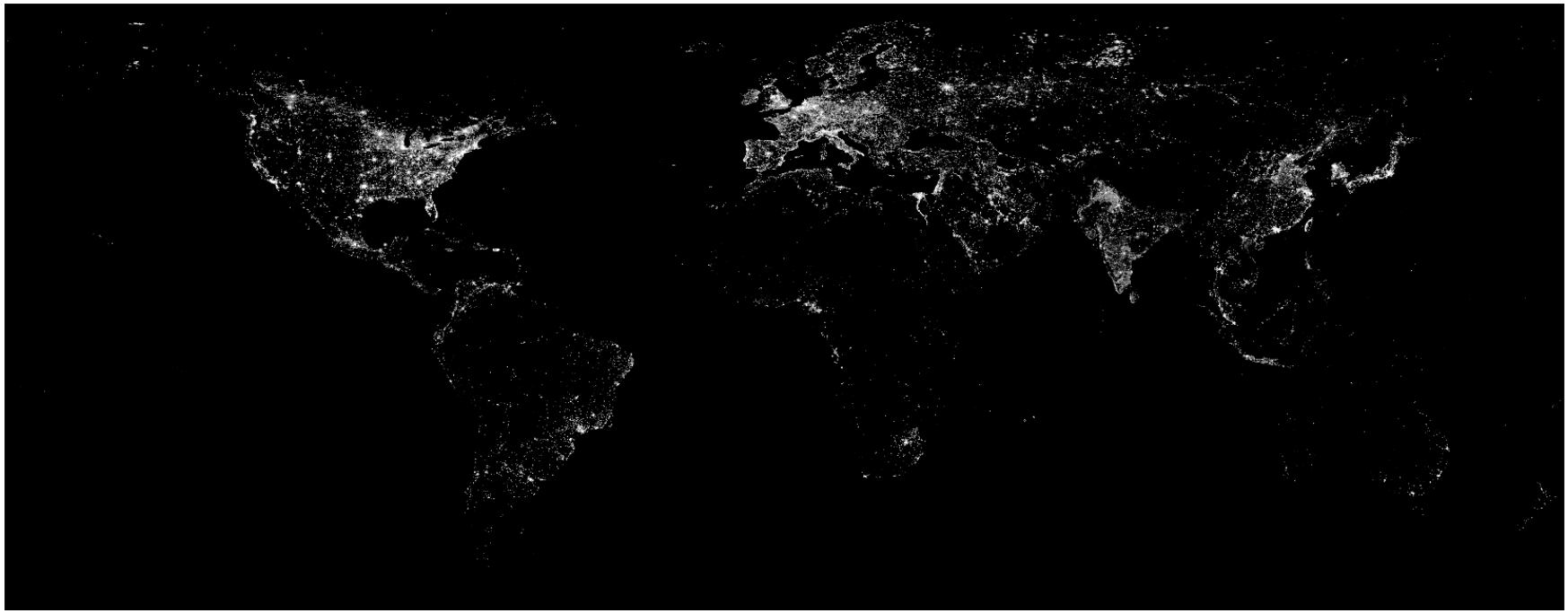
Contribution

Use data from NASA satellites capturing light of human settlements at night and develop method to aggregate these with official growth statistics.

Paper examples: satellite data

Night lights 1: "Pure measurement". Henderson, Storeygard and Weil (2012)

Night lights in 2008



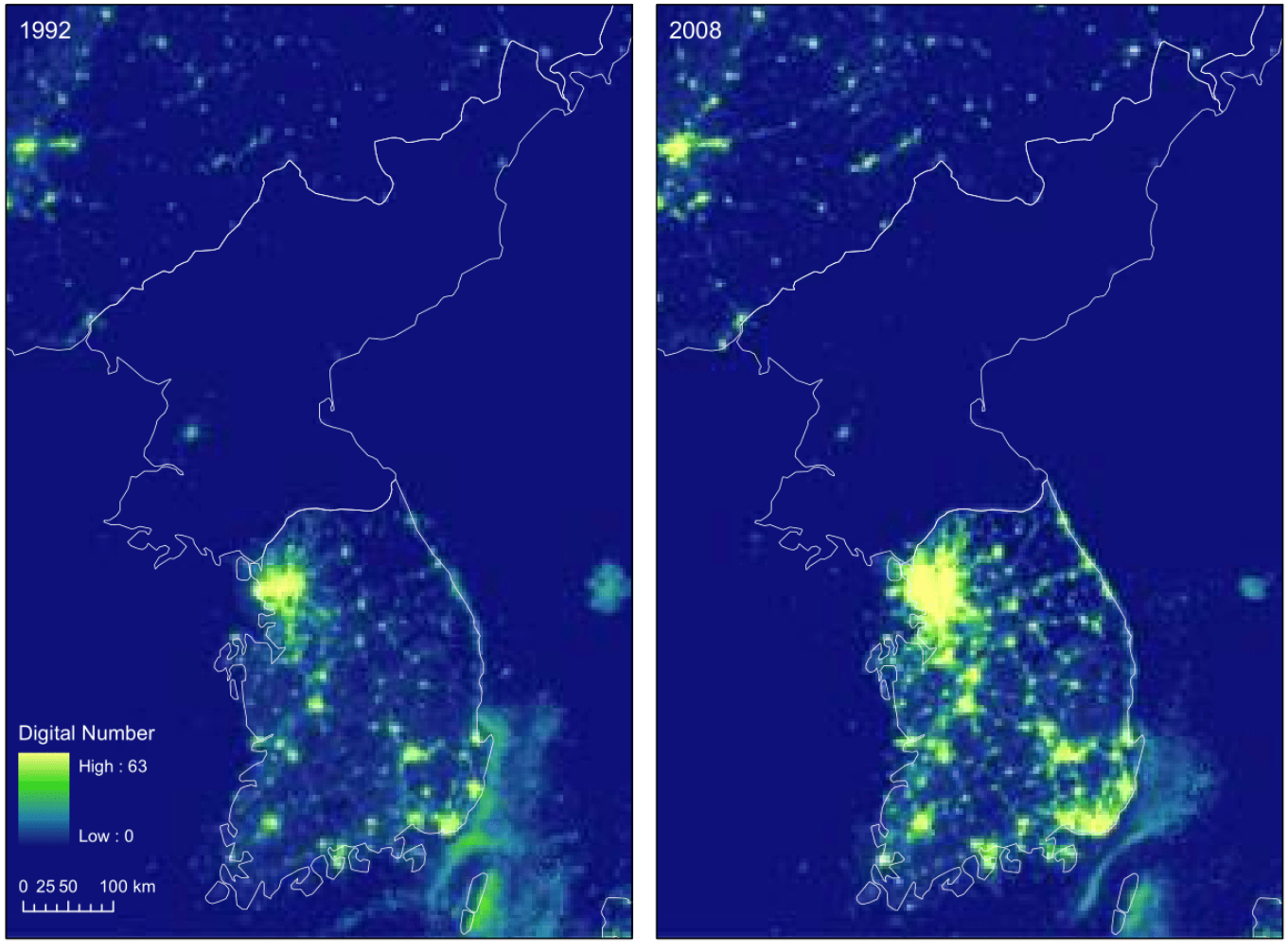
source:

`https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`

Paper examples: satellite data

Night lights 1: "Pure measurement". Henderson, Storeygard and Weil (2012)

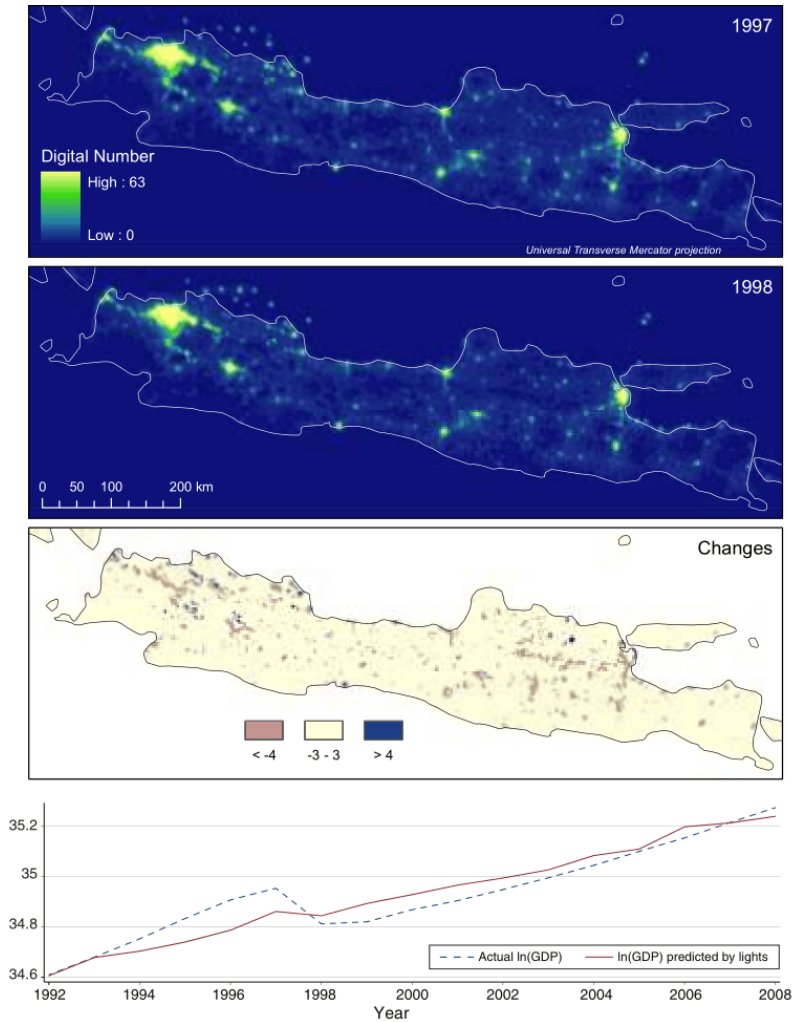
Long-term growth: Korea



Paper examples: satellite data

Night lights 1: "Pure measurement". Henderson, Storeygard and Weil (2012)

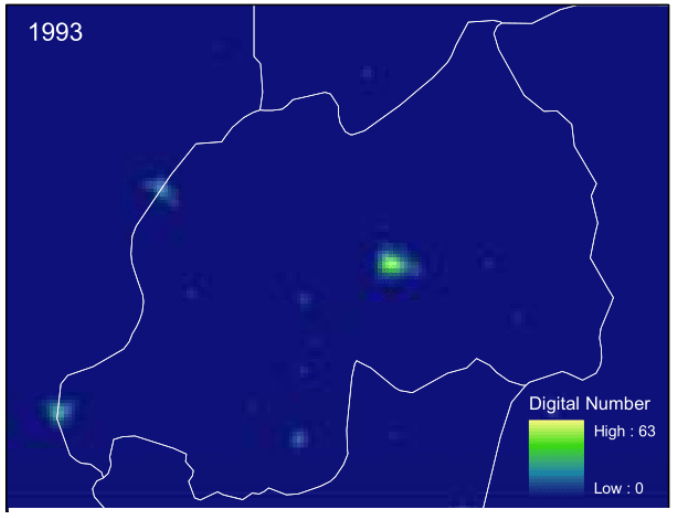
Short-term economic shocks: Indonesia 1997/1998

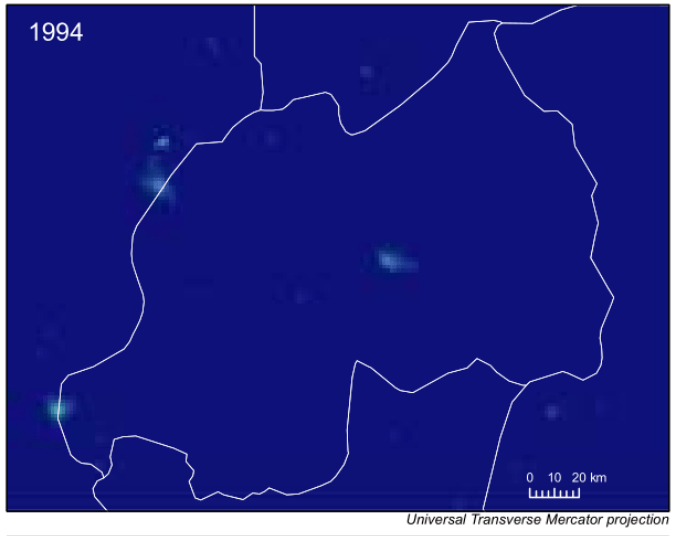


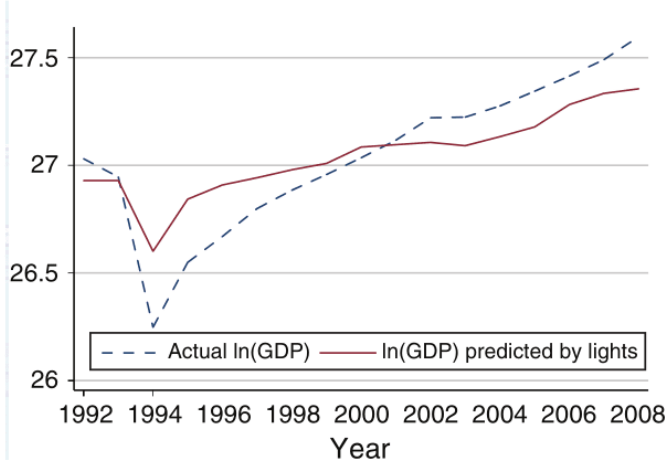
Paper examples: satellite data

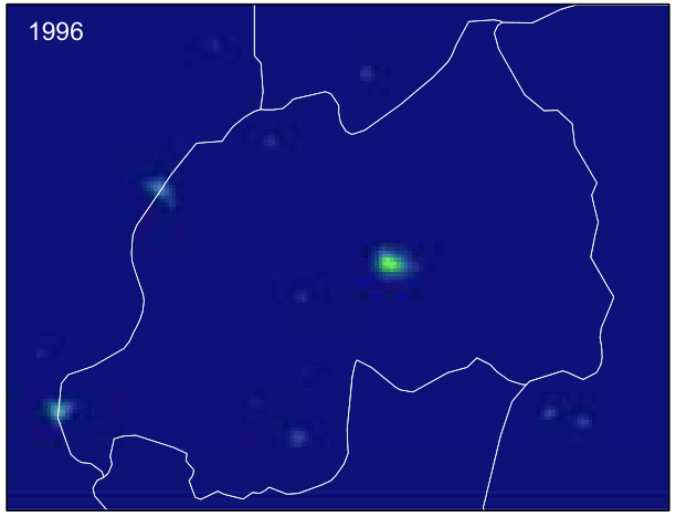
Night lights 1: "Pure measurement". Henderson, Storeygard and Weil (2012)

The impact of war: Rwanda 1994









Paper examples: satellite data

Night lights 2: Institutions and development. Michalopoulos and Papaioannou (2014)

Elias Papaioannou and Stelios Michalopoulos (2014) "National Institutions and Sub-national Development in Africa," QJE 129(1): 151-213.

Motivation

Big literature on the role of institutions in economic growth (Acemoglu et al. (2001) etc.; La Porta et al (1997) etc.; Engerman and Sokoloff (2000) etc.) but identification challenge since institutions are endogenous.

Contribution

Combine Satellite data on night lights, old maps of ethnic groups ancestral homelands, and modern national borders to identify effect of national institutions on local development in Africa.

Paper examples: satellite data

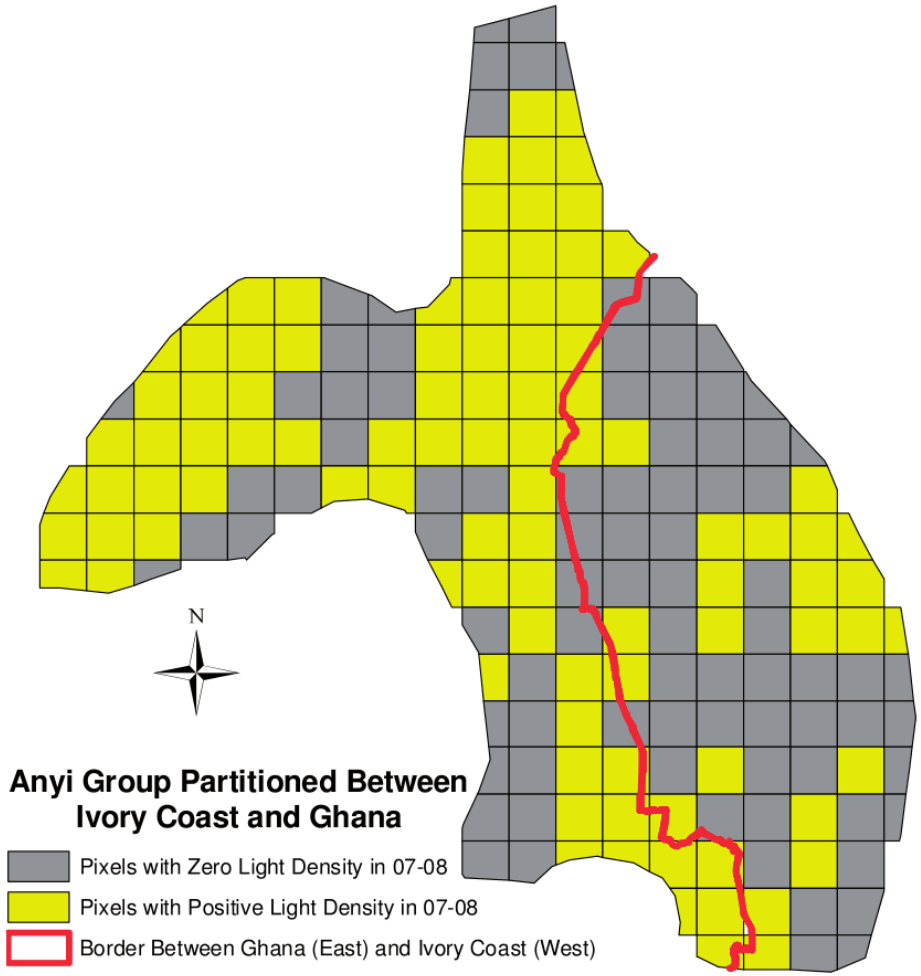
Night lights 2: Institutions and development. Michalopoulos and Papaioannou (2014)

Pre-colonial ethnic and modern national boundaries

Paper examples: satellite data

Night lights 2: Institutions and development. Michalopoulos and Papaioannou (2014)

An example of a split ethnic group



Paper examples: satellite data

Night lights 2: Institutions and development. Michalopoulos and Papaioannou (2014)

Estimation (Linear probability model)

p = pixel

e = ethnicity

c = country

$$\mathbb{I}(pix\_lit_{p,e,c}) = \alpha_e + \beta \times inst\_qual_c + controls \qquad (1)$$

Focussing on partitioned ethnicities only, including α_e kills the coefficient on institutional quality.

Paper examples: satellite data

Night lights 3: Favoritism. Hodler and Raschky (2014)

Roland Hodler and Paul Raschky (2014). "Regional Favoritism," QJE 129 (2): 995-1033.

Motivation

Large literature on favoritism from Bates (1974) onwards, but much of it anecdotal, case studies. Limited evidence across countries shows large heterogeneity across policy outcomes (Kramon and Posner (2013)).

Contribution

Systematically assess favoritism in large sample of countries: 38,427 sub-national regions from 126 countries, spanning democracies and autocracies. Exploit change in political leadership and information about birthplaces of countries' political leaders.

Paper examples: satellite data

Night lights 3: Favoritism. Hodler and Raschky (2014)

Mobutu in Zaire / DRC

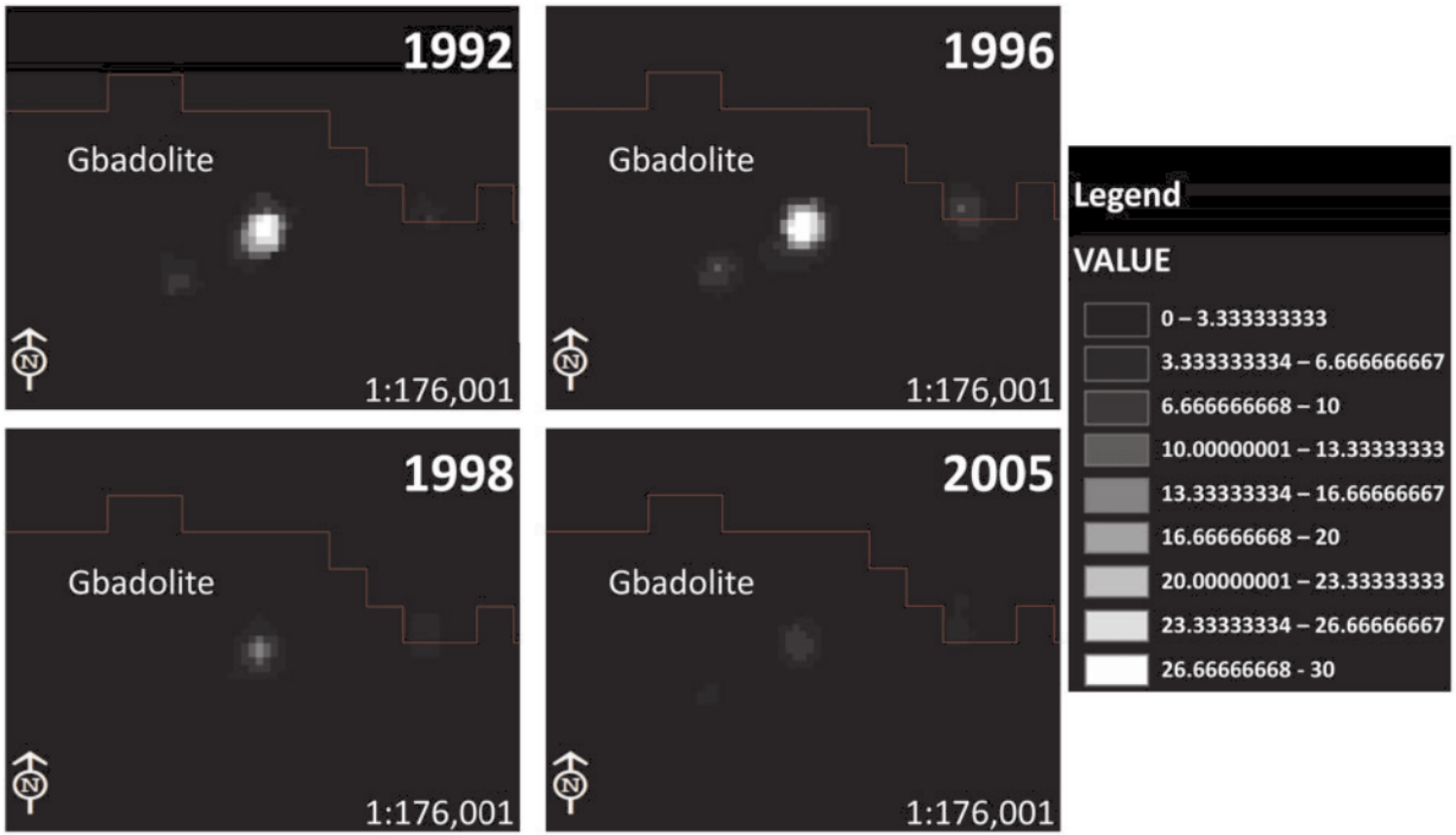


FIGURE I

Nightime Light Intensity in Gbadolite in 1992, 1996, 1998, and 2005

Mobutu Sese Seko was president of Zaire until 1997.

Paper examples: satellite data

Night lights 3: Favoritism. Hodler and Raschky (2014)

Mahinda in Sri Lanka

FIGURE II

Nightime Light Intensity in Hambantota in 2003 and 2006

Mahinda Rajapaksa became prime minister of Sri Lanka in 2004 and president in 2005.

Paper examples: satellite data

Deforestation: Burgess et al. (2012)

Robin Burgess, Matthew Hansen, Benjamin Olken, Peter Potapov, and Stefanie Sieber (2012). "The Political Economy of Deforestation in the Tropics," QJE 127 (4): 1707-1754.

Motivation

Tropical deforestation accounts for one fifth of greenhouse gas emissions. Much of it comes from illegal logging. What are the economic incentives driving this?

Contribution

Use satellite data on changes in forest cover to examine how local officials respond to changing incentives by allowing more or less logging.

Consistent with Cournot model: creation of new districts leads to higher deforestation and lower timber prices.

Paper examples: satellite data

Deforestation: Burgess et al. (2012)

Deforestation in Riau province

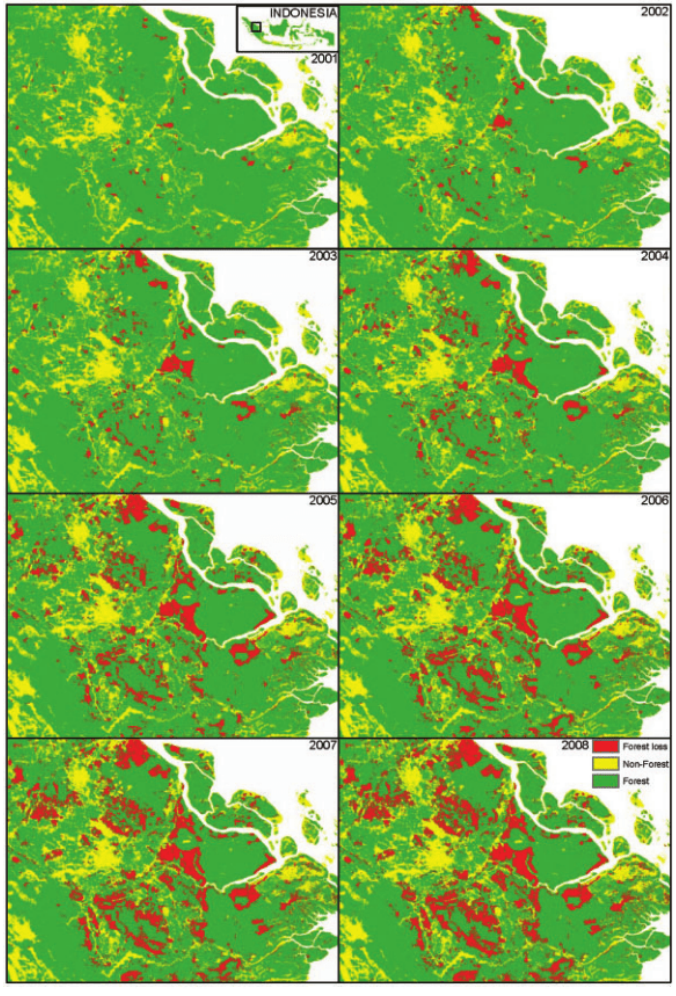


FIGURE I

Forest Cover Change in the Province of Riau, 2001–2008

Paper examples: satellite data

Deforestation: Burgess et al. (2012)

Deforestation in Indonesia 2001-2008

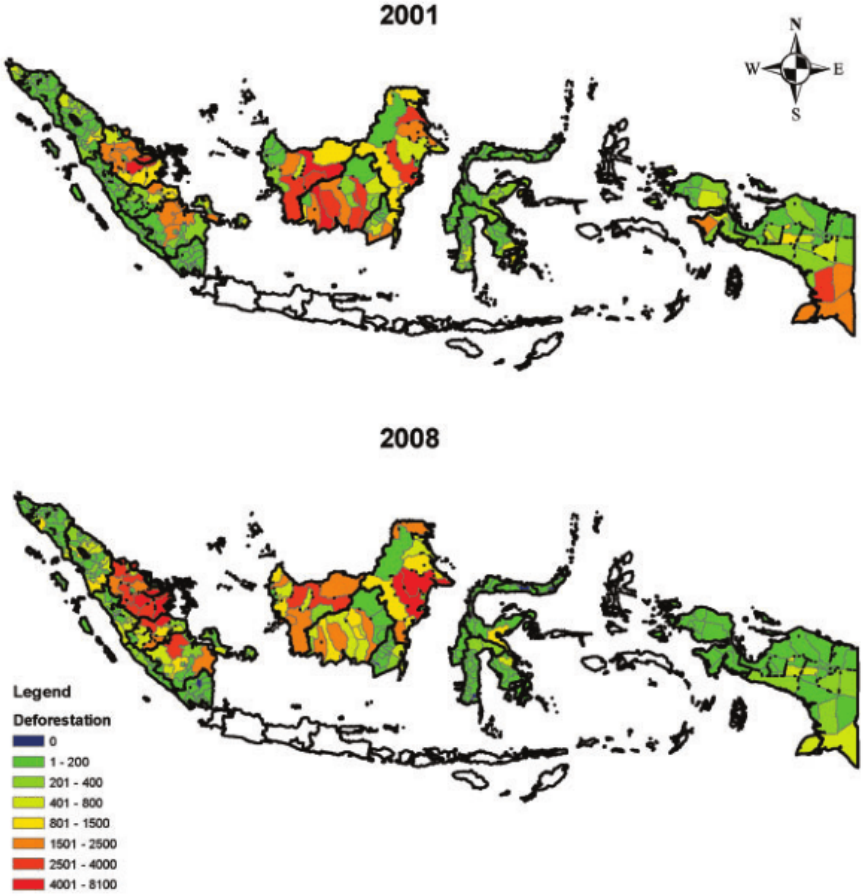


FIGURE II

District-Level Logging in Indonesia Using the 2008 District Boundaries, 2001 and 2008

Paper examples: satellite data

Pollution: Chen et al. (2013)

Yuyu Chen, Ginger Zhe Jin, Naresh Kumar, and Guang Shi (2013). "The Promise of Beijing: Evaluating the impact of the 2008 Olympic Games on air quality," JEEM 66 424-443.

Motivation

Pollution travels. Can drastic local policy interventions improve air quality? Sustainably?

Contribution

Show that measures to close roads and factories around the time of the 2008 Olympics improved air quality – both according to official statistics and to data from NASA satellites. Pollution quickly reverted to original levels after the games.

Paper examples: satellite data

Pollution: Chen et al. (2013)

Change in pollution around the games from the paper, AOD data*

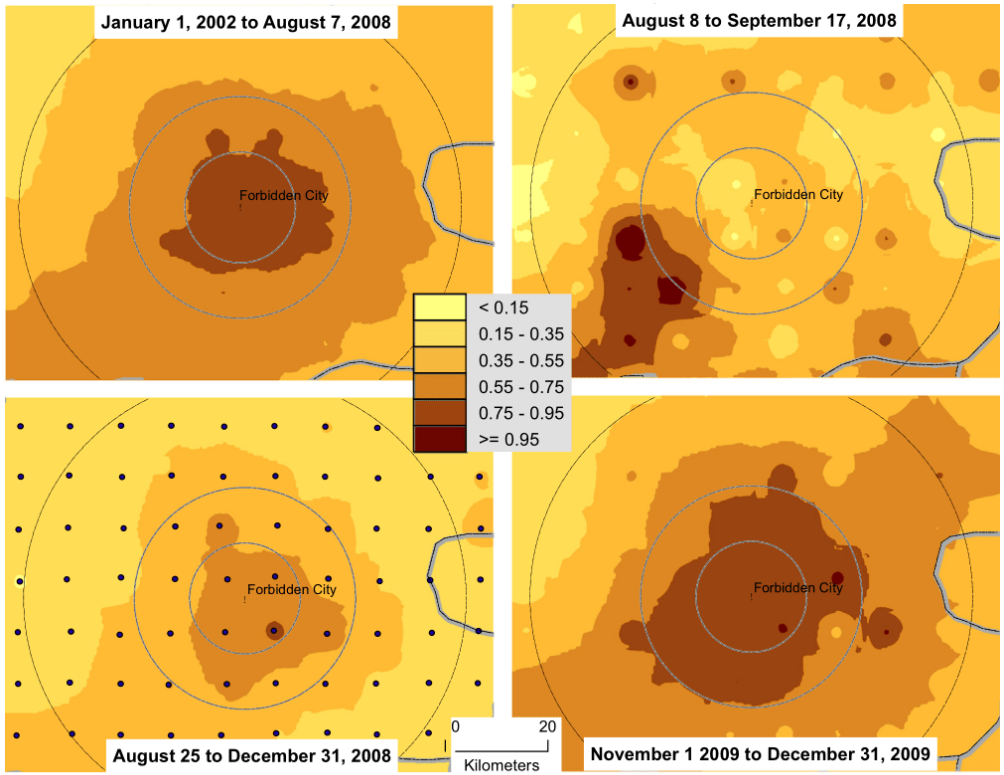


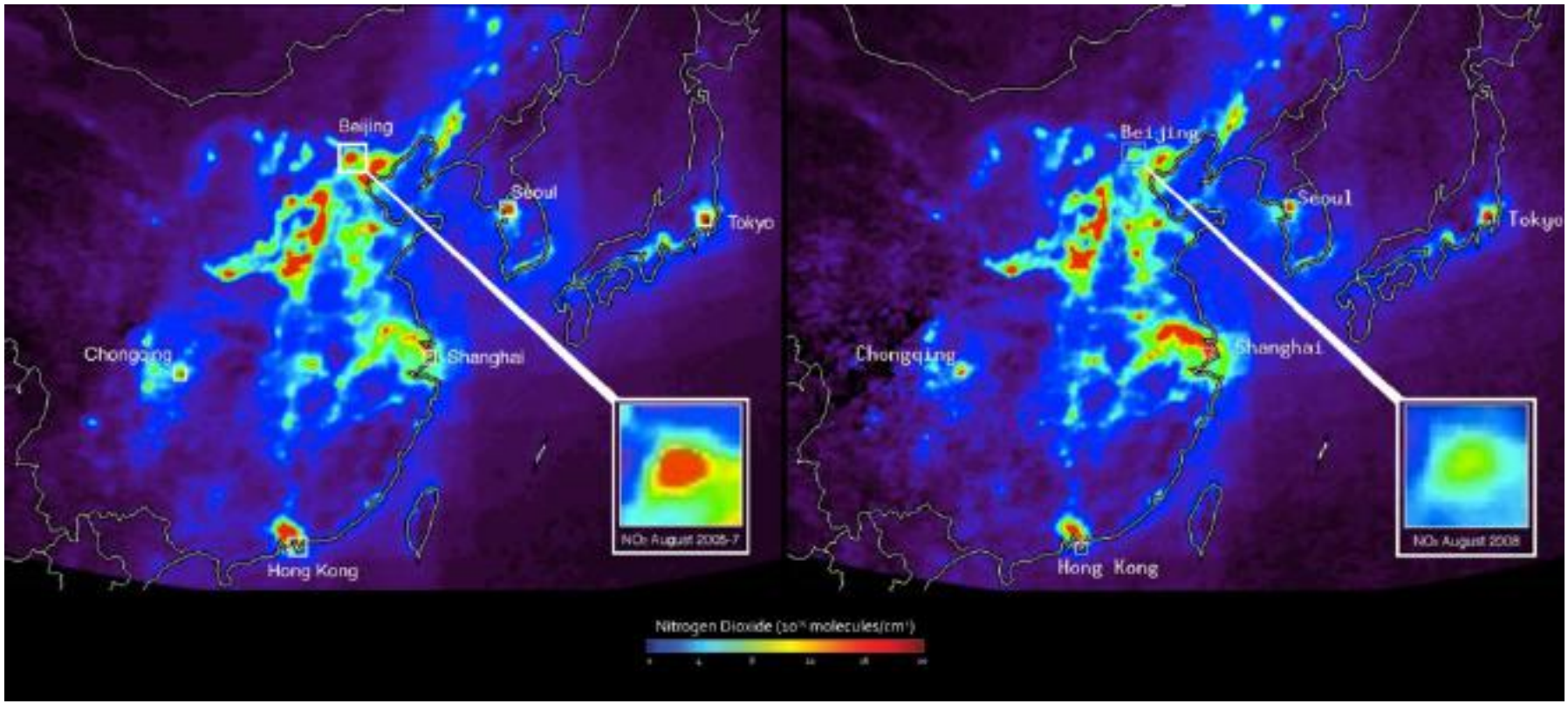
Fig. 7. Satellite-based AOD, corrected for meteorological conditions and spatiotemporal trends in and around Beijing.

* "AOD captures the amount of radiation absorbed, reflected, and scattered due to the presence of solid and liquid particulates suspended in the atmospheric colum."

Paper examples: satellite data

Pollution: Chen et al. (2013)

NO_2 concentration before and during the games



Paper examples: old maps

The impact of slavery: Nunn (2008)

Nathan Nunn (2008). "The Long-term Effects of Africa's Salve Trades," QJE 123(1): 139-176.

Motivation

Can Africa's underdevelopment be explained by the slave trade?

Contribution

Using archival research and GIS techniques, constructs estimates of slavery-intensity at the country-level. Uses these estimates, together with distance-based IV (also constructed with GIS) to show that slavery had negative impact on development among African countries.

Paper examples: old maps

The impact of slavery: Nunn (2008)

Constructing slavery-intensity step 1

TABLE I
SLAVE ETHNICITY DATA FOR THE TRANS-ATLANTIC SLAVE TRADE

Location	Years	Num. ethnic.	Num. obs.	Record type
Valencia, Spain	1482–1516	77	2,675	Crown records
Puebla, Mexico	1540–1556	14	115	Notarial records
Dominican Republic	1547–1591	26	22	Records of sale
Peru	1548–1560	16	202	Records of sale
Mexico	1549	12	80	Plantation accounts
Peru	1560–1650	30	6,754	Notarial records
Lima, Peru	1583–1589	15	288	Baptism records
Colombia	1589–1607	9	19	Various records
Mexico	1600–1699	28	102	Records of sale
Dominican Republic	1610–1696	33	55	Government records
Chile	1615	6	141	Sales records
Lima, Peru	1630–1702	33	409	Parish records
Peru (Rural)	1632	25	307	Parish records
Lima, Peru	1640–1680	33	936	Marriage records
Colombia	1635–1695	6	17	Slave inventories
Guyane (French Guiana)	1690	12	69	Plantation records
Colombia	1716–1725	33	59	Government records
French Louisiana	1717–1769	23	223	Notarial records
Dominican Republic	1717–1827	11	15	Government records
South Carolina	1732–1775	35	681	Runaway notices
Colombia	1738–1778	11	100	Various records
Spanish Louisiana	1770–1803	79	6,615	Notarial records
St. Dominique (Haiti)	1771–1791	25	5,413	Sugar plantations
Bahia, Brazil	1775–1815	14	581	Slave lists
St. Dominique (Haiti)	1778–1791	36	1,280	Coffee plantations
Guadeloupe	1788	8	45	Newspaper reports
St. Dominique (Haiti)	1788–1790	21	1,297	Fugitive slave lists
Cuba	1791–1840	59	3,093	Slave registers
St. Dominique (Haiti)	1796–1797	56	5,632	Plantation inventories
American Louisiana	1804–1820	62	223	Notarial records
Salvador, Brazil	1808–1842	6	456	Records of manumission
Trinidad	1813	100	12,460	Slave registers
St. Lucia	1815	62	2,333	Slave registers
Bahia, Brazil	1816–1850	27	2,666	Slave lists
St. Kitts	1817	48	2,887	Slave registers
Senegal	1818	17	80	Captured slave ship
Berbice (Guyana)	1819	66	1,127	Slave registers
Salvador, Brazil	1819–1836	12	871	Manumission certificates
Salvador, Brazil	1820–1835	11	1,106	Probate records
Sierra Leone	1821–1824	68	605	Child registers
Rio de Janeiro, Brazil	1826–1837	31	772	Prison records
Anguilla	1827	7	51	Slave registers
Rio de Janeiro, Brazil	1830–1852	190	2,921	Free africans' records
Rio de Janeiro, Brazil	1833–1849	35	476	Death certificates

TABLE I
(CONTINUED)

Location	Years	Num. ethnic.	Num. obs.	Record type
Salvador, Brazil	1835	13	275	Court records
Salvador, Brazil	1838–1848	7	202	Slave registers
St. Louis/Goree, Senegal	1843–1848	21	189	Emancipated slaves
Bakel, Senegal	1846	16	73	Sales records
d'Agoué, Benin	1846–1885	11	70	Church records
Sierra Leone	1848	132	12,425	Linguistic and British census
Salvador, Brazil	1851–1884	8	363	Records of manumission
Salvador, Brazil	1852–1888	7	269	Slave registers
Cape Verde	1856	32	314	Slave census
Kikoneh Island, Sierra Leone	1896–1897	11	185	Fugitive slave records

Paper examples: old maps

The impact of slavery: Nunn (2008)

The Murdock (1959) tribal map as digitized by Nunn (2008)

Also used by Michalopoulos and Papaioannou (ECMA 2013, QJE 2014, AER 2016).

Paper examples: old maps

The impact of slavery: Nunn (2008)

Constructing slavery-intensity step 2

FIGURE II
Ethnic Boundaries Defined by Murdock (1959) and Modern Political Boundaries

Paper examples: network

Trasport Costs and Economic Development: Storeygard (2016)

Storeygard, Adam (2016). "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa" RES 123(1): 139-176.

Motivation

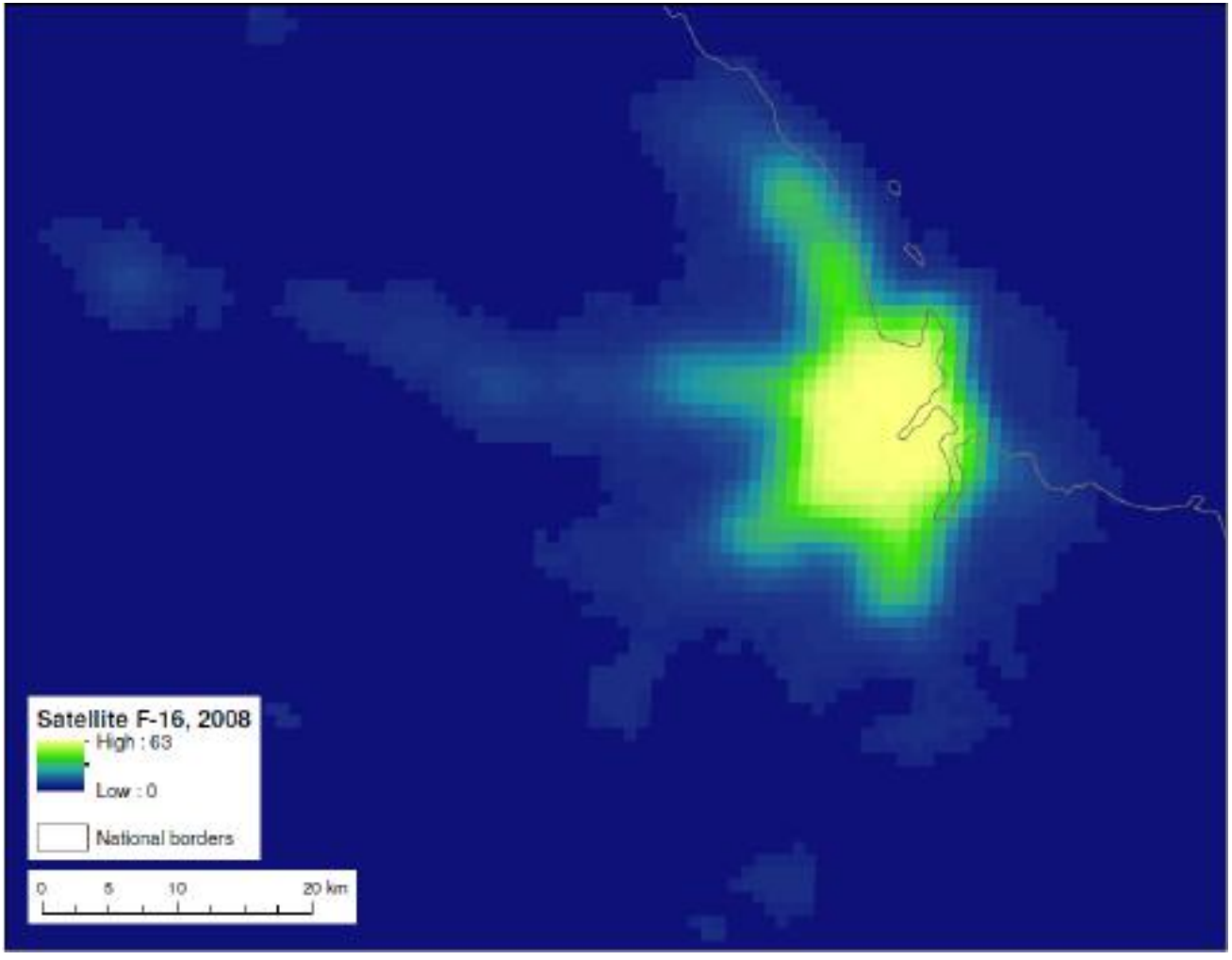
How inter-city connectivity determines the income of sub-Saharan African cities.

Contribution

Using road quality/type, lights, and oil price shock, estimates how growth is propagated far away from the city port. Negative elasticity of city economic activities wrt transport costs. Effect is heterogeneous in road quality (paved/unpaved).

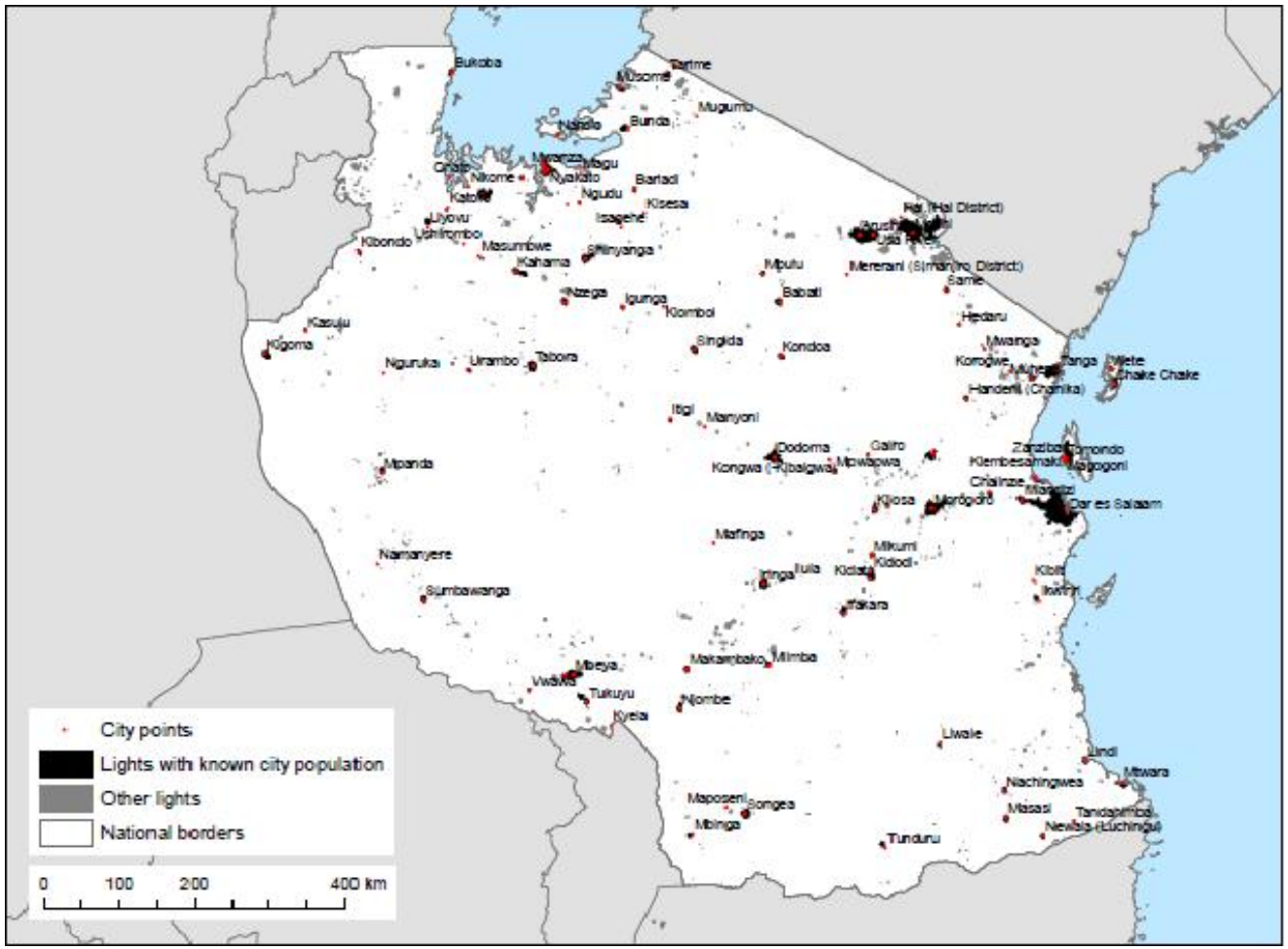
Paper examples: network

Luminosity (2008) in Dar el Salaam



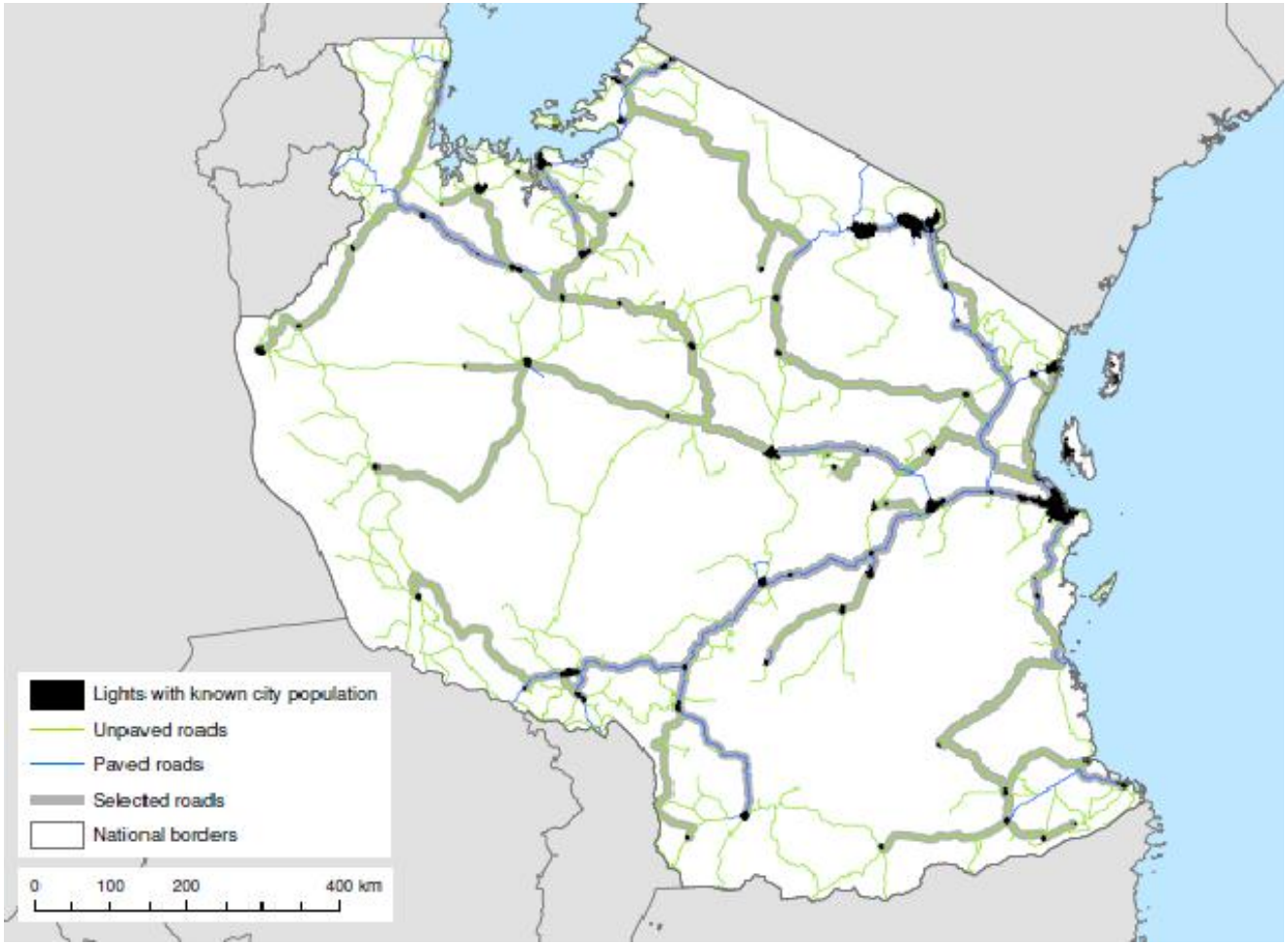
Paper examples: network

City Distribution in Tanzania



Paper examples: network

Roads connection to Dar el Salaam (Tanzania)



Paper examples: network

Trafficking Networks and the Mexican Drug War: Dell (2016)

Dell, Melissa (2016). "Trafficking Networks and the Mexican Drug War" AER

Motivation

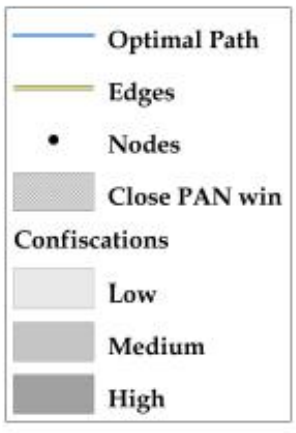
How crime repression policy affects spatial pattern of violence and crime .

Contribution

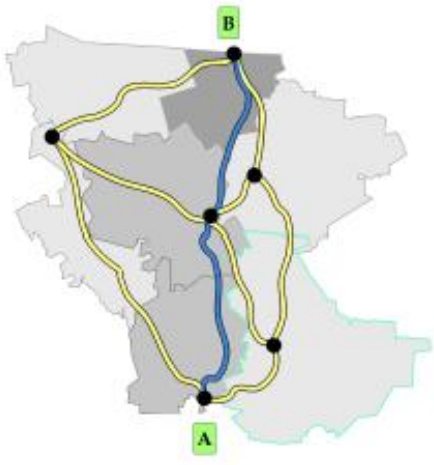
Using close elections, disaggregated crime data, and road network, to identify how crime reacts to crackdown policy. Quantification of externalities and unintended consequences of crime repression.

Paper examples: network

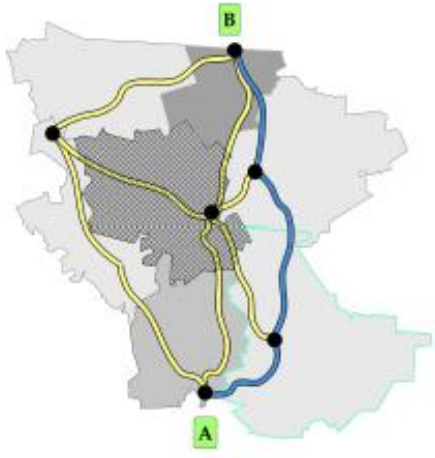
Spillover Methodology



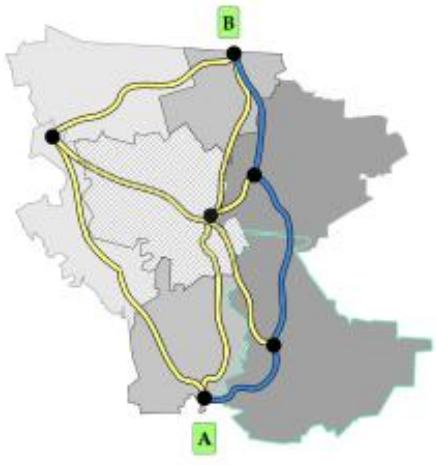
(a) Legend



(b) Basic environment



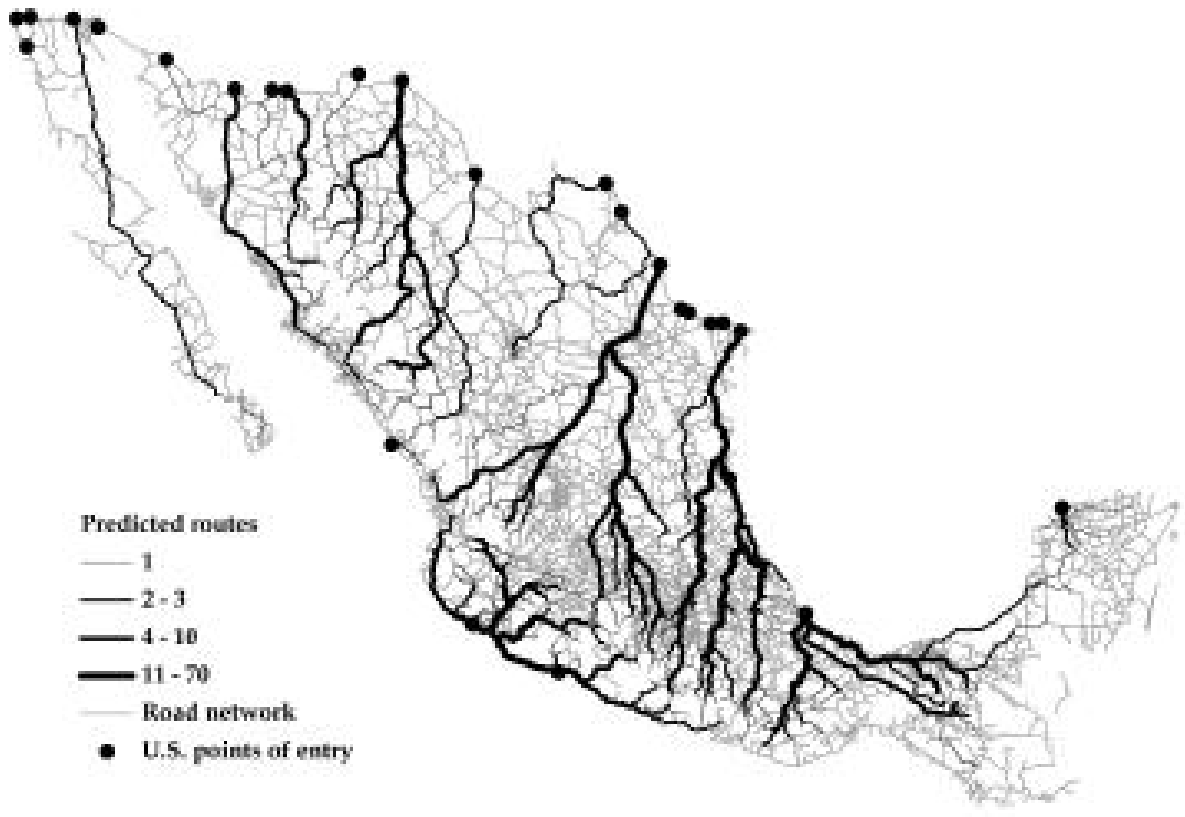
(c) Close PAN victory increases costs; routes change



(d) Do confiscation patterns also change?

Paper examples: network

Roads Network and Predicted Trafficking Routes



Paper examples: network

Railroads and American Economic Growth: Donaldson and Hornbeck (2015)

Donaldson, Dave and Hornbeck, Richard (2015). "Railroads and American Economic Growth: A "Market Access" Approach" QJE

Motivation

Evaluate the role of railway construction for US economic growth.

Contribution

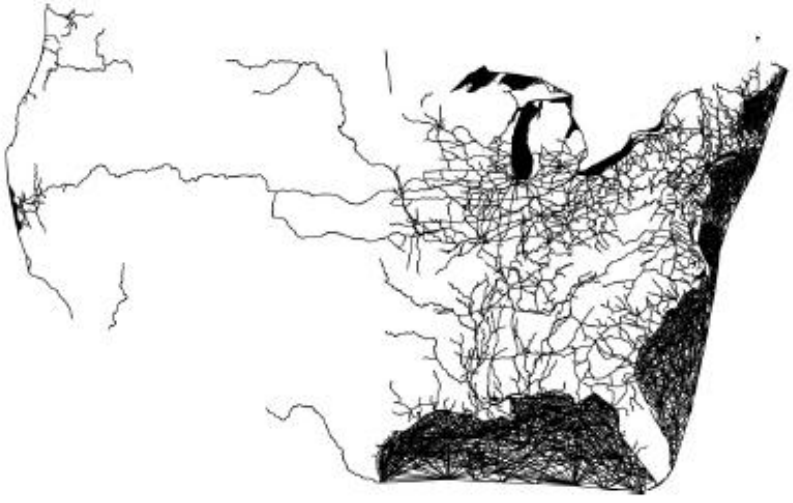
Combine transportation network over time and census data to show that expansion of the railway network fostered both local and aggregated economic growth. Theory-based application of intra-country trade model (Eaton and Kortum, 2002) in a reduced form framework to quantify the spillover effect of infrastructure project.

Railways expansion fostered US economic growth. Aggregate effects are considerably larger than local effects (due to higher "market access").

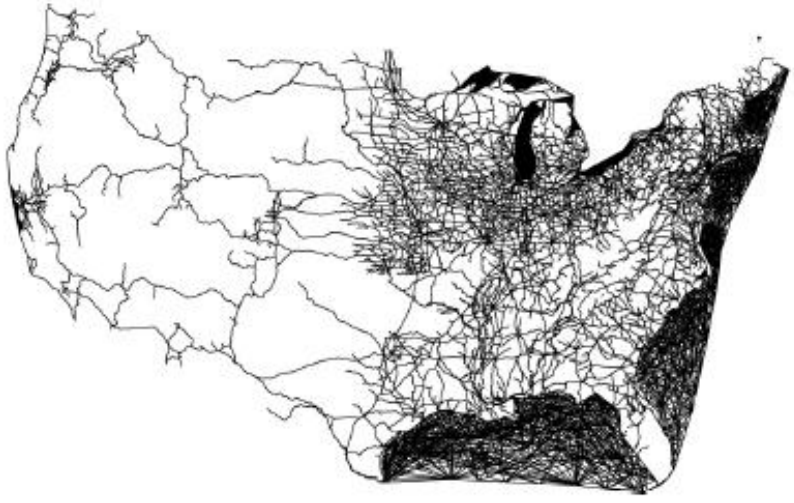
Paper examples: network

Railway expansion 1870-1890

C. Natural Waterways, Canals, and 1870 Railroads

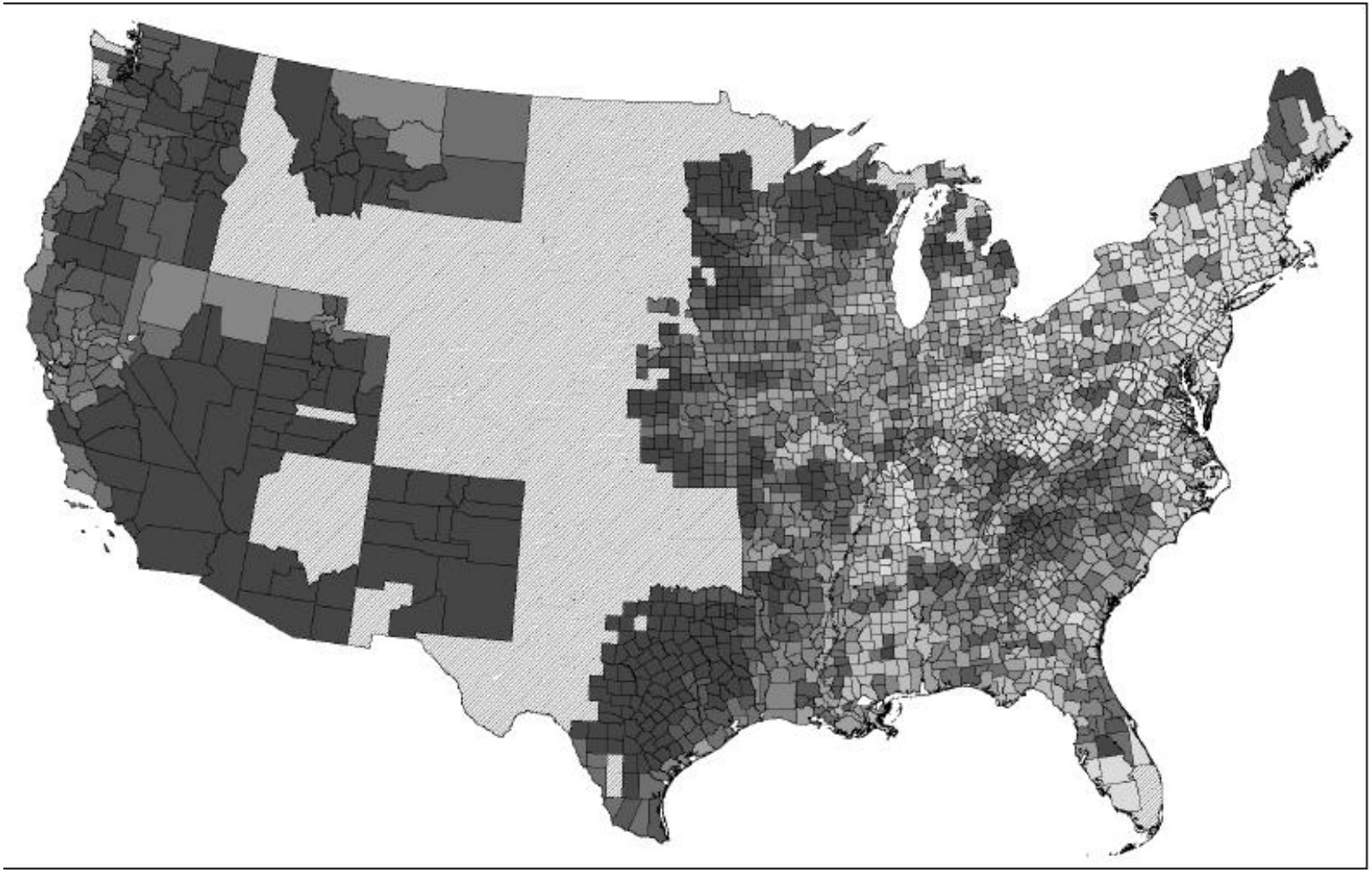


D. Natural Waterways, Canals, and 1890 Railroads



Paper examples: network

Change in Market Access 1870-1890



Paper examples: network

Archeology!: Barjamovic, Chaney, Cosar and Hortacsu (2019)

Gojko Barjamovic, Thomas Chaney, Kerem Cosar, and Ali Hortacsu (2019). "Trade, Merchants and the Lost Cities of the Bronze Age" QJE

Motivation

Test gravity model of trade, find lost cities!

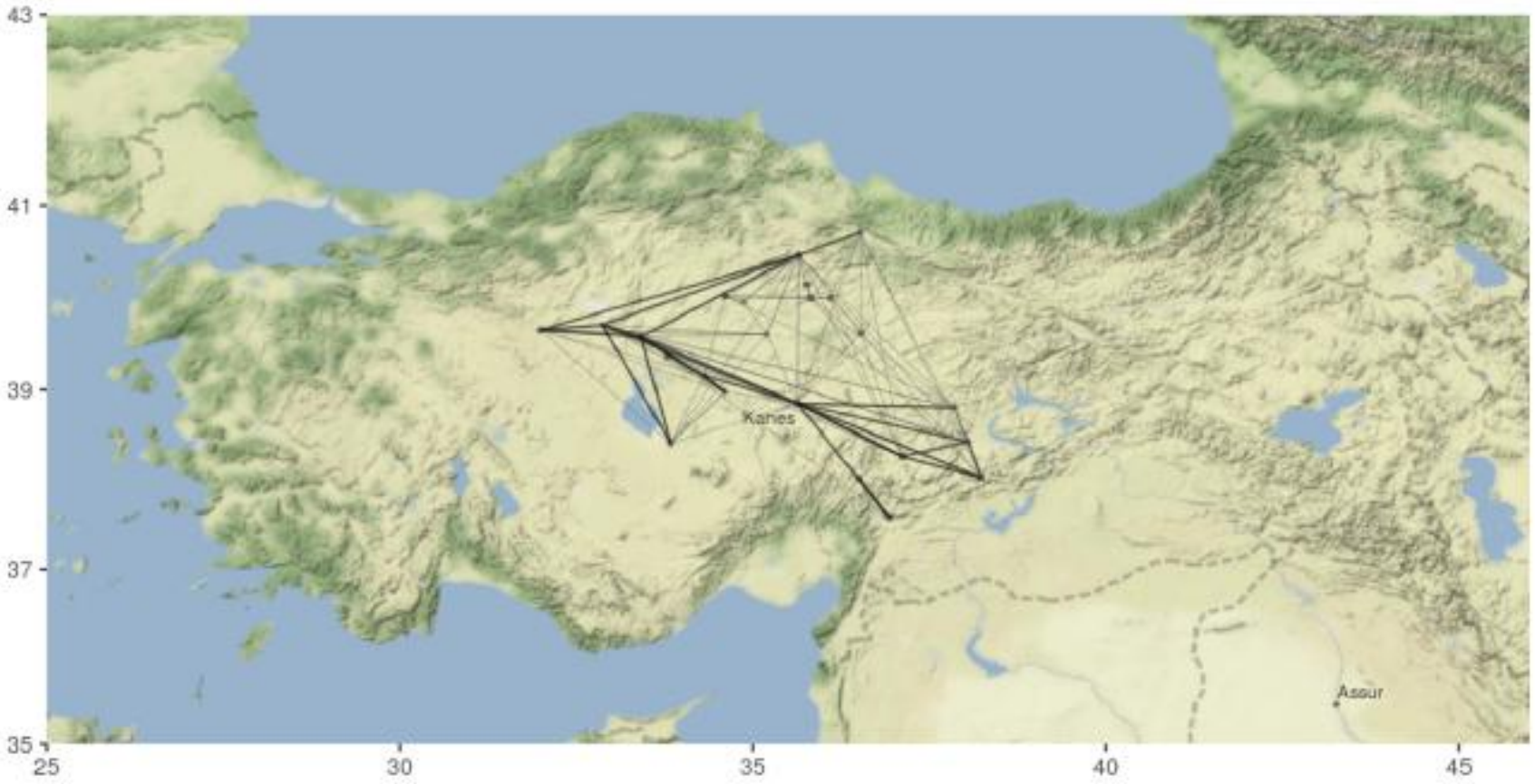
Contribution

- extract information on trade linkages between Bronze-age Assyrian cities: how many transactions between cities?
- specify structural gravity model of trade that can be estimated purely with: (1) trade flows between cities, (2) geographic distance between cities
- can "invert" the gravity model to pinpoint the likely locations of lost cities!

Paper examples: network

Archeology!: Barjamovic, Chaney, Cosar and Hortacsu (2019)

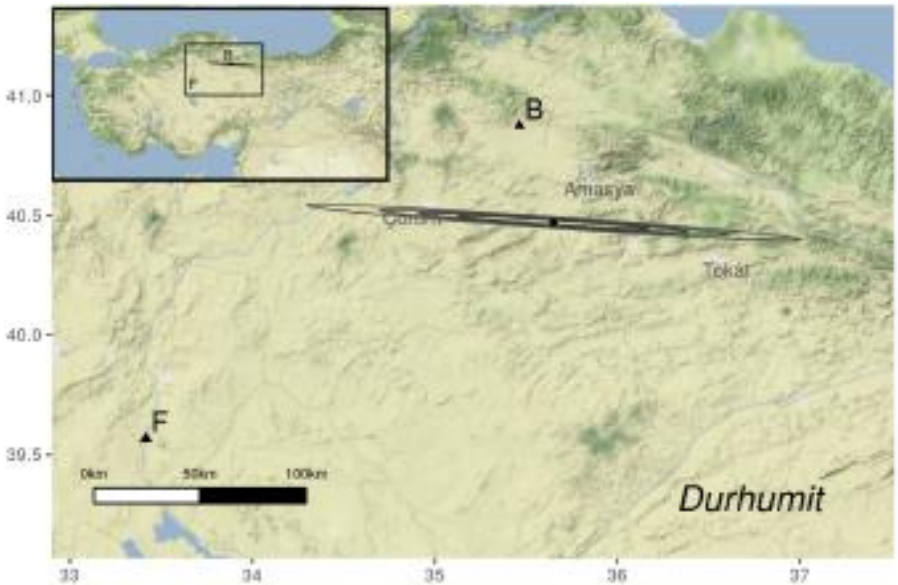
Ancient cities and trade routes

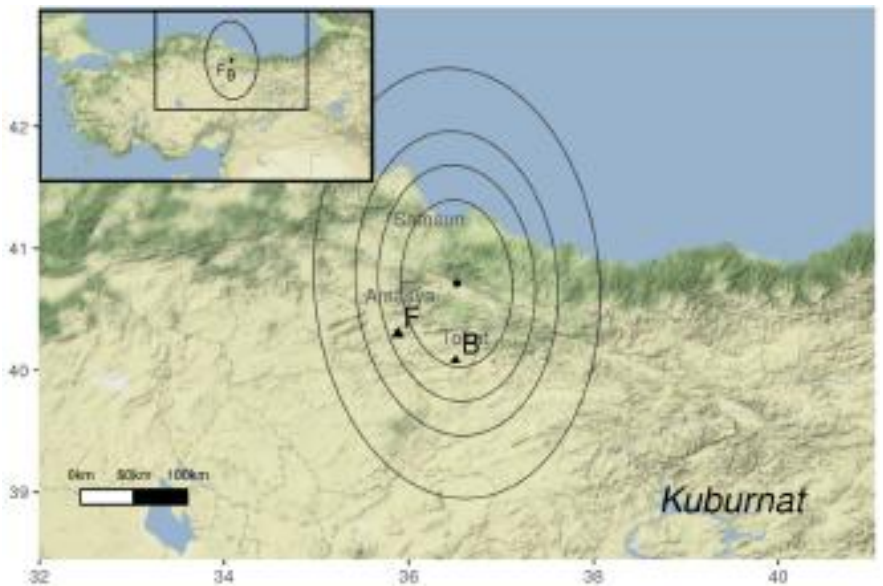


Paper examples: network

Archeology!: Barjamovic, Chaney, Cosar and Hortacsu (2019)

Locating lost cities





"dot" = point estimate of lost city location, "B/F" = Archeologist estimates of city locations

Paper examples: conflict

Conflict 1: Ethnic partitioning. Michalopoulos and Papaioannou (2016)

Stelios Michalopoulos and Elias Papaioannou (2016). "The Long-Run Effects of the Scramble for Africa," AER 106(7). 1802-1848

Motivation

Did the arbitrary partitioning of ethnic groups during the "Scramble for Africa" in the 19th century create a long-lasting legacy of violent conflict?

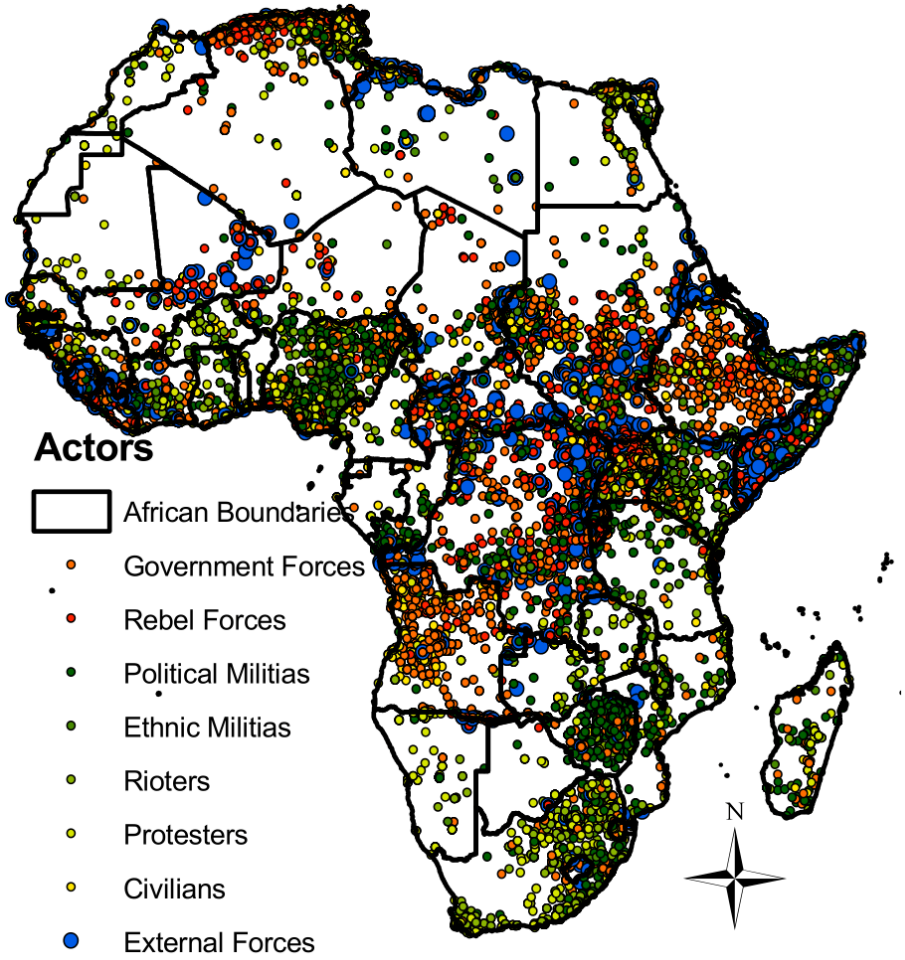
Contribution

Combine the digitized Murdock (1959) map with modern political boundaries and geo-referenced data on conflict events to show that ethnicities split between two modern states experience more violent conflict (incidence, severity, and duration). Identification credible since drawing of colonial borders "quasi-random" with respect to existing ethnic groups.

Paper examples: conflict

Conflict 1: Ethnic partitioning. Michalopoulos and Papaioannou (2016)

Conflict events as reported by ACLED*



* http://www.acleddata.com/

Paper examples: conflict

Conflict 2: Propaganda. Yanagizawa-Drott (2014)

David Yanagizawa-Drott (2014). "Propaganda and Conflict: Evidence from the Rwandan Genocide," QJE 129(4). 1947-1994

Motivation

Rwandan genocide (1994) killed 800,000-1,000,000 people. Quickest mass extermination in history. Prominent role of "hate radio" in mobilizing killers. $\rightarrow$ Can mass media be used to incite violence against civilians?

Contribution

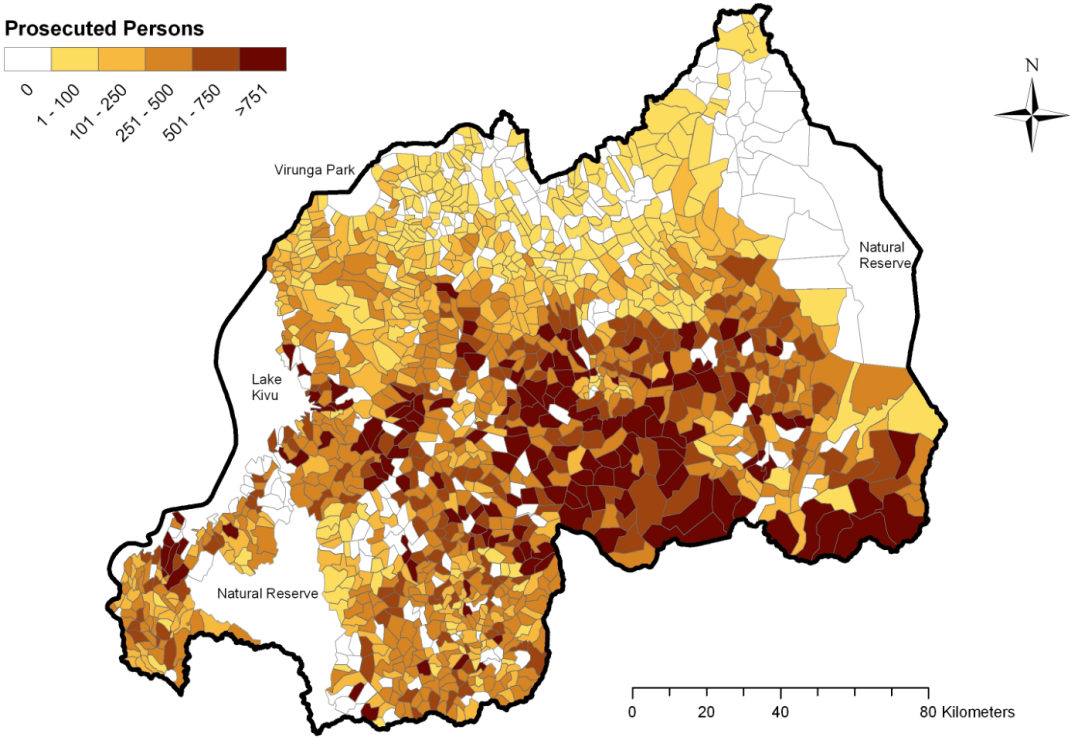
Exploits quasi-random variation in hate radio signal strength (combination of transmitter location and Rwanda's mountainous terrain) to show that broadcasts increased participation in killings by locals. Around 10% of deaths can be attributed to radio*.

* For similar identification strategy exploiting variation in signal strength see Enikolopov, Ruben, Maria Petrova and Ekaterina Zhuravskaya (2011) "Media and Political Persuasion: Evidence from Russia". AER, 101(7): 3253-85.

Paper examples: conflict

Conflict 2: Propaganda. Yanagizawa-Drott (2014)

Genocide violence in villages

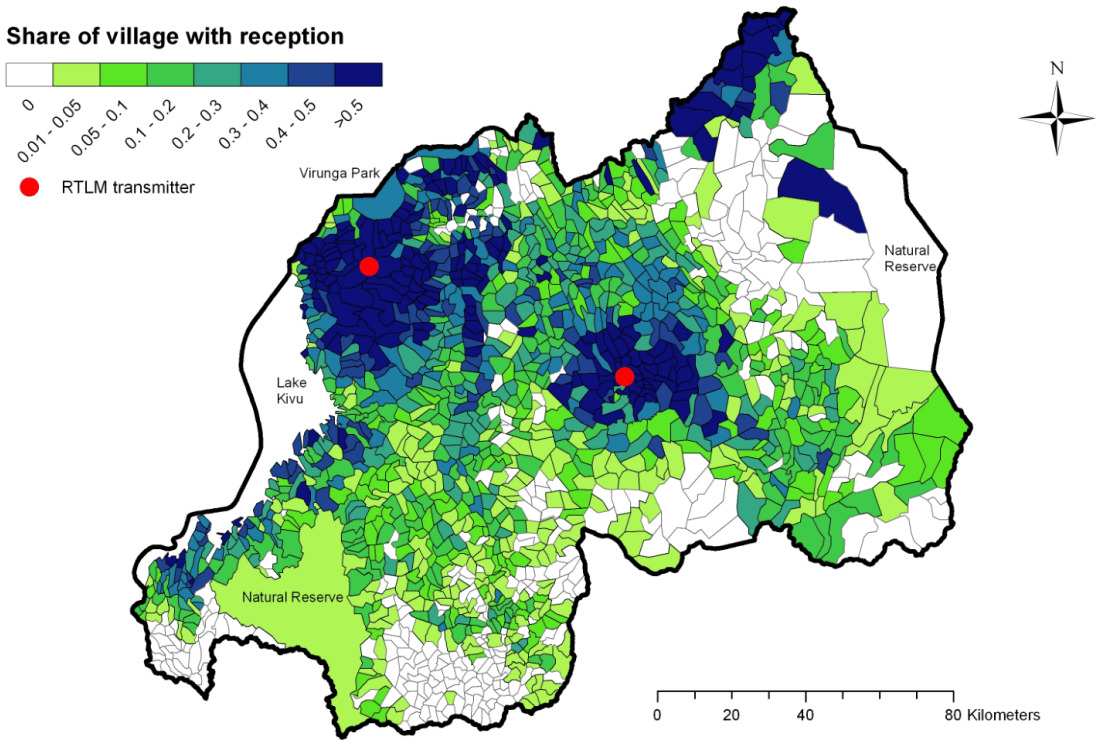


The categories represent the total number of prosecuted persons in the village (sum of militia and individual violence). White areas are missing data, either because of geography, such as parks and natural reserves, or villages that lack data in the sample.

Paper examples: conflict

Conflict 2: Propaganda. Yanagizawa-Drott (2014)

RTLM radio coverage



The figure shows the radio coverage in villages (share of village area with sufficient radio reception) based on the Longley-Rice propagation model. Source: Author's calculations in ArcGIS using the Longley-Rice Propagation Model.

Paper examples: conflict

Conflict 3: The role of militias as motivators. Rogall (2014)

Thorsten Rogall, (2014). "Mobilizing the masses for genocide," unpublished.

`https://sites.google.com/site/thorstenrogall/`

Motivation

Did organized militias matter in as organizers and motivators for civilians killing other civilians?

Contribution

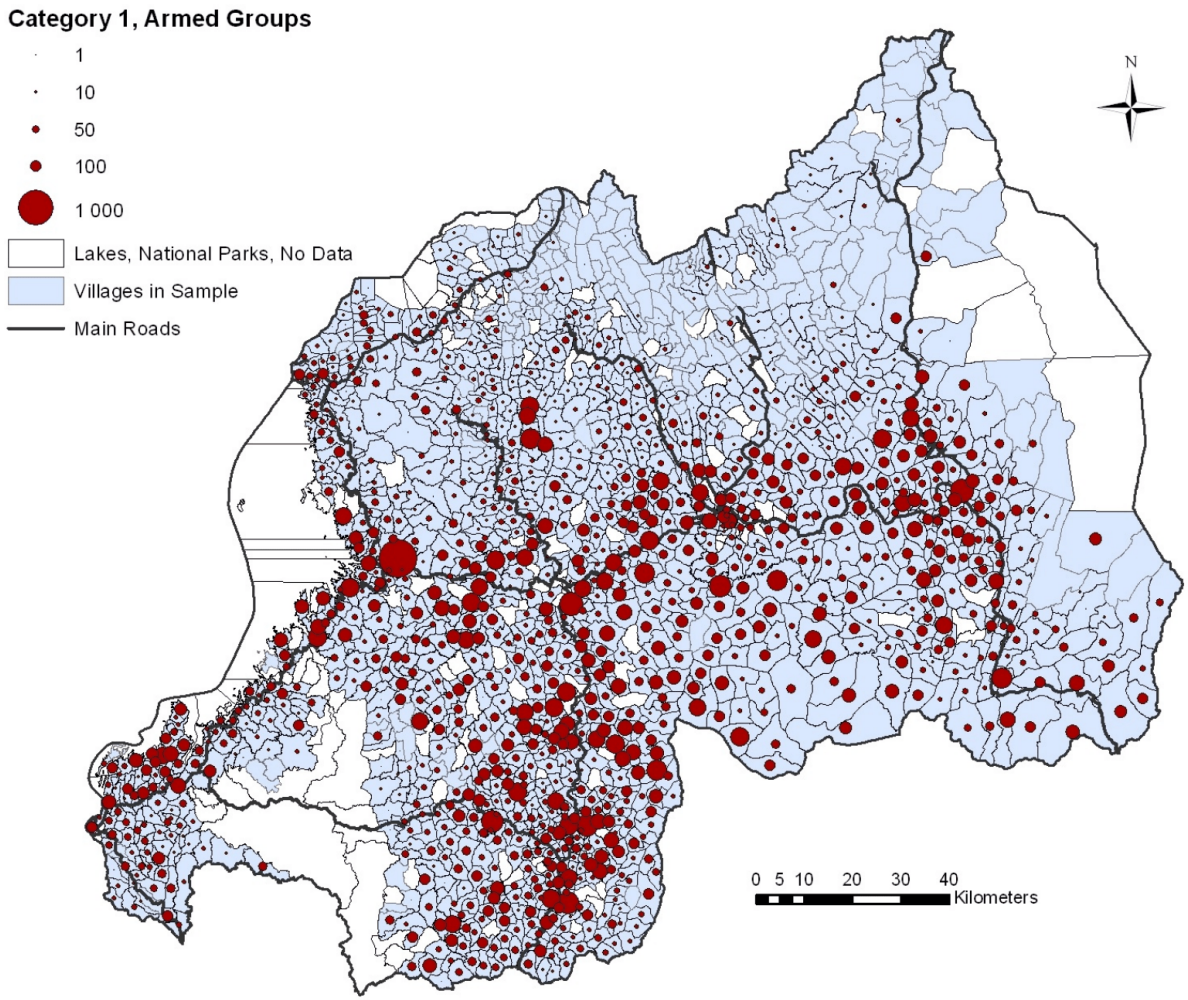
Exploits quasi-random variation in transport costs of militias in accessing villages created by different amounts of rain falling on unpaved roads during the genocide.

Main Finding: Militia presence did increase killing rates. One additional militiaman resulted in 13 additional deaths.

Paper examples: conflict

Conflict 3: The role of militias as motivators. Rogall (2014)

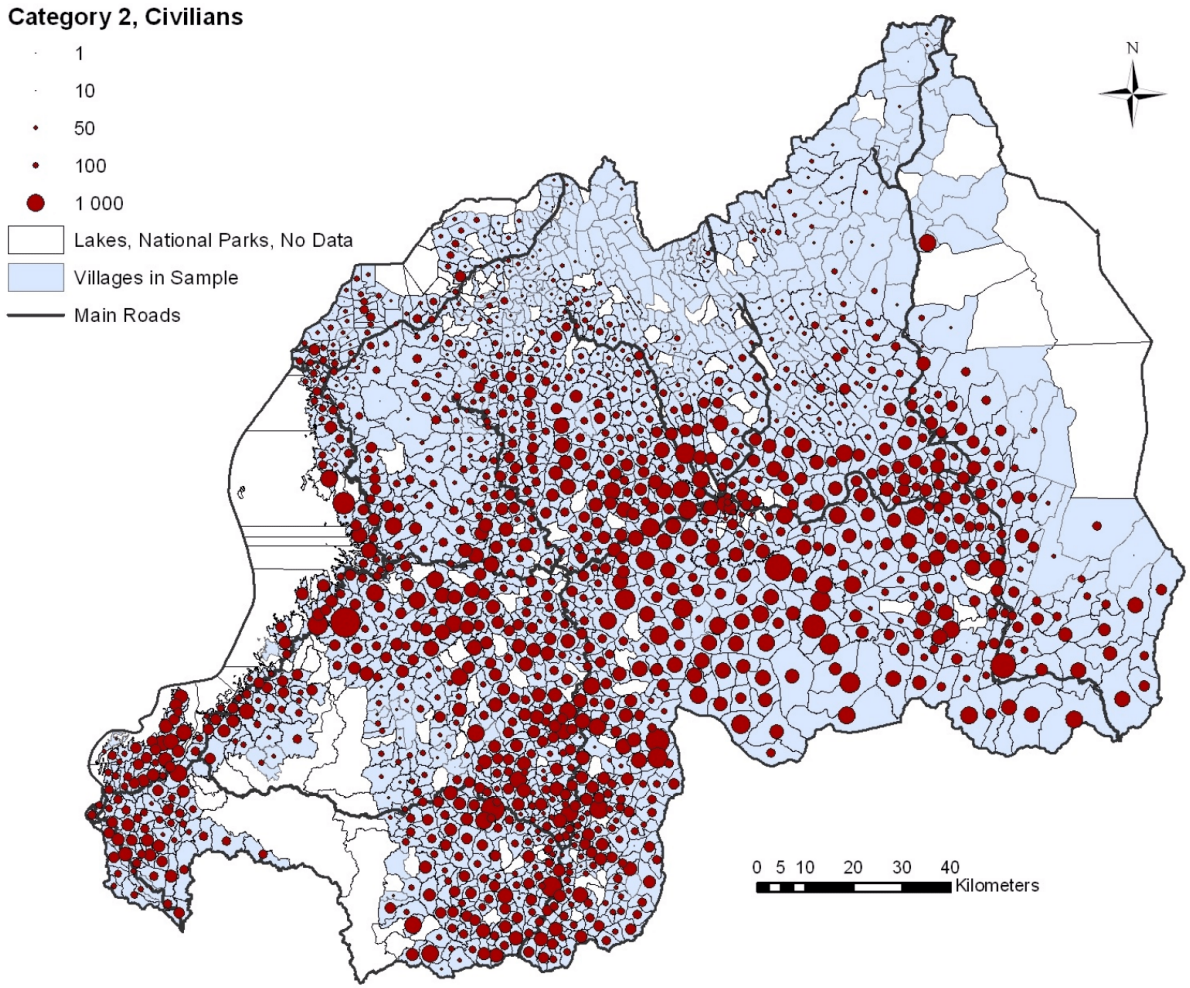
Violence by armed groups



Paper examples: conflict

Conflict 3: The role of militias as motivators. Rogall (2014)

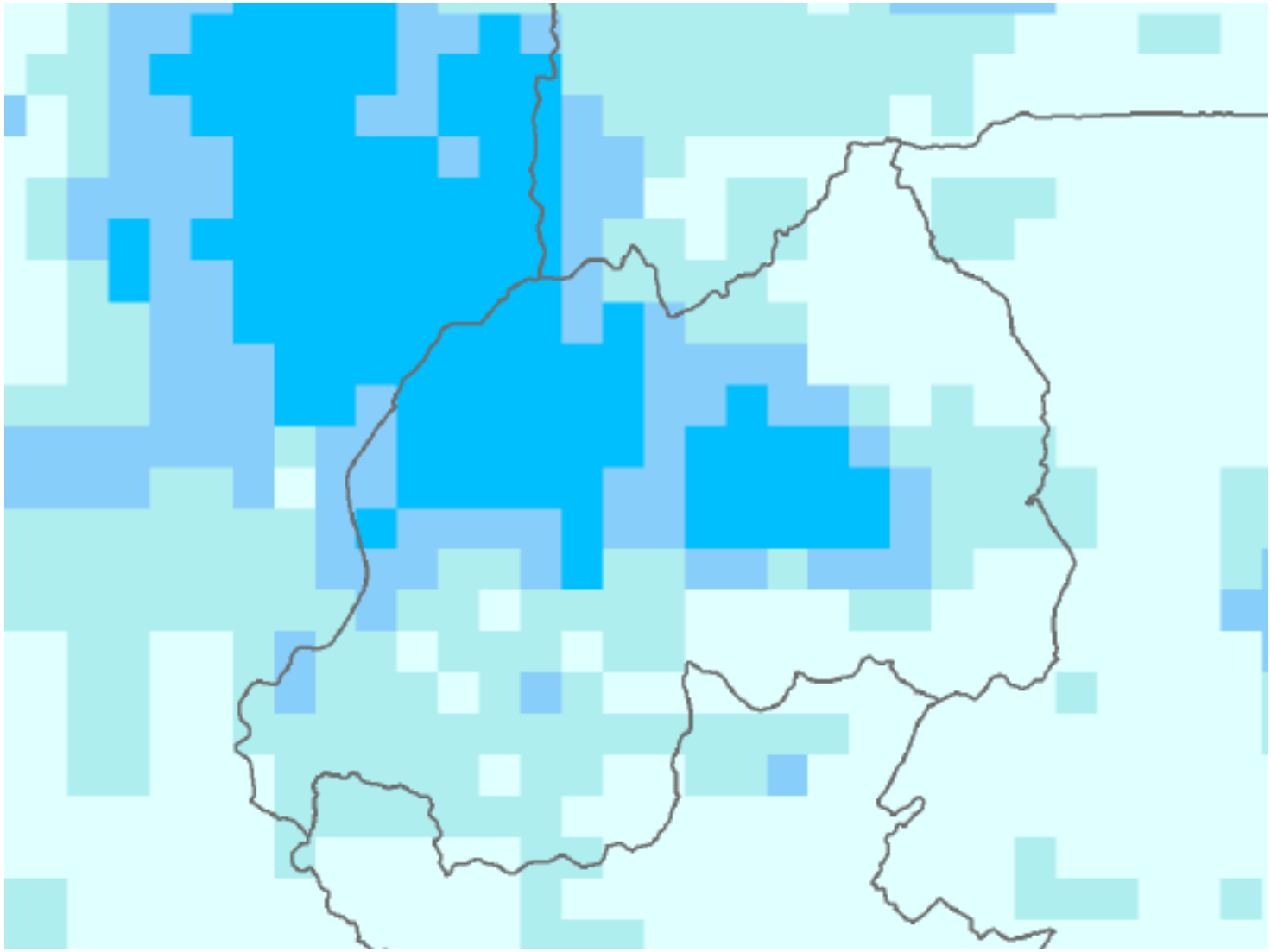
Violence by civilians



Paper examples: conflict

Conflict 3: The role of militias as motivators. Rogall (2014)

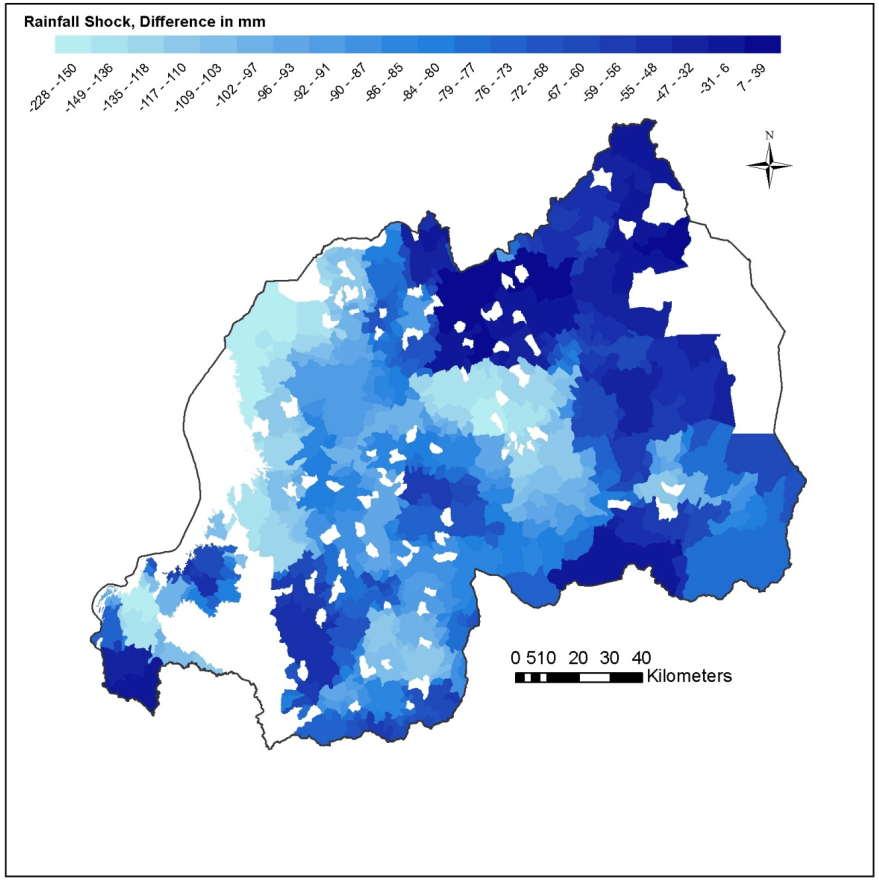
Rainfall Data Example - NOAA (NASA)



Paper examples: conflict

Conflict 3: The role of militias as motivators. Rogall (2014)

Detailed rainfall data

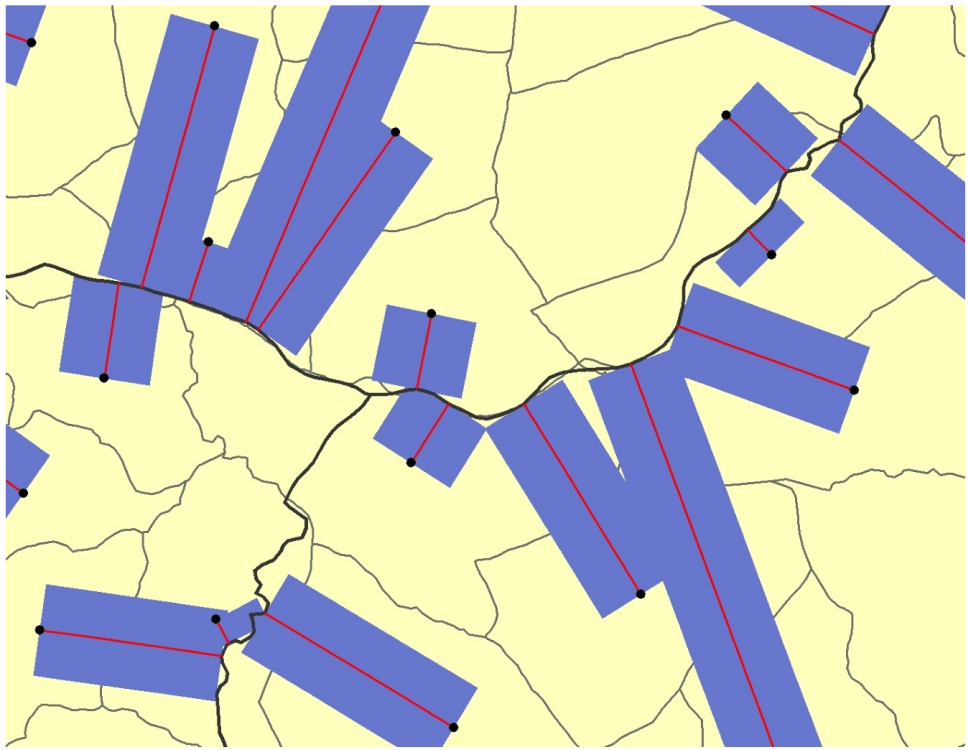


Note: This map shows rainfall along the way between main road and village during the period of the genocide in 1994 for each village, subtracting rainfall between main road and village during the 100 calendar days of the genocide of an average year (years 1984-1993). White areas are either national parks, Lake Kivu or villages not in the sample.

Paper examples: conflict

Conflict 3: The role of militias as motivators. Rogall (2014)

Instrument construction



Instrument: Interaction of the length of the red line and amount of rain falling on the area of the blue rectangle during the period of the genocide.

Paper examples: other applications

Other applications 1: Topography as an instrument. Pande and Duflo (2007)

Rohini Pande and Esther Duflo (2007). "Dams," QJE, 122(2) :601-646.

Motivation

Construction of dams for irrigation and power generation is (a) widespread (45,000 large dams, 1/2 of rivers in the world have one), (b) thought to lead to development and poverty reduction.

Contribution

Overcome endogeneity problem of dam construction using river gradient as instrument for dam suitability.
In panel of Indian districts, find distributional effects:
downstream districts benefit (output mean $\nearrow$, output volatility, poverty $\searrow$),
dam districts lose (output mean $\rightarrow$, output volatility, poverty $\nearrow$).

Paper examples: other applications

Other applications 1: Topography as an instrument. Pande and Duflo (2007)

Dams by District in 1970

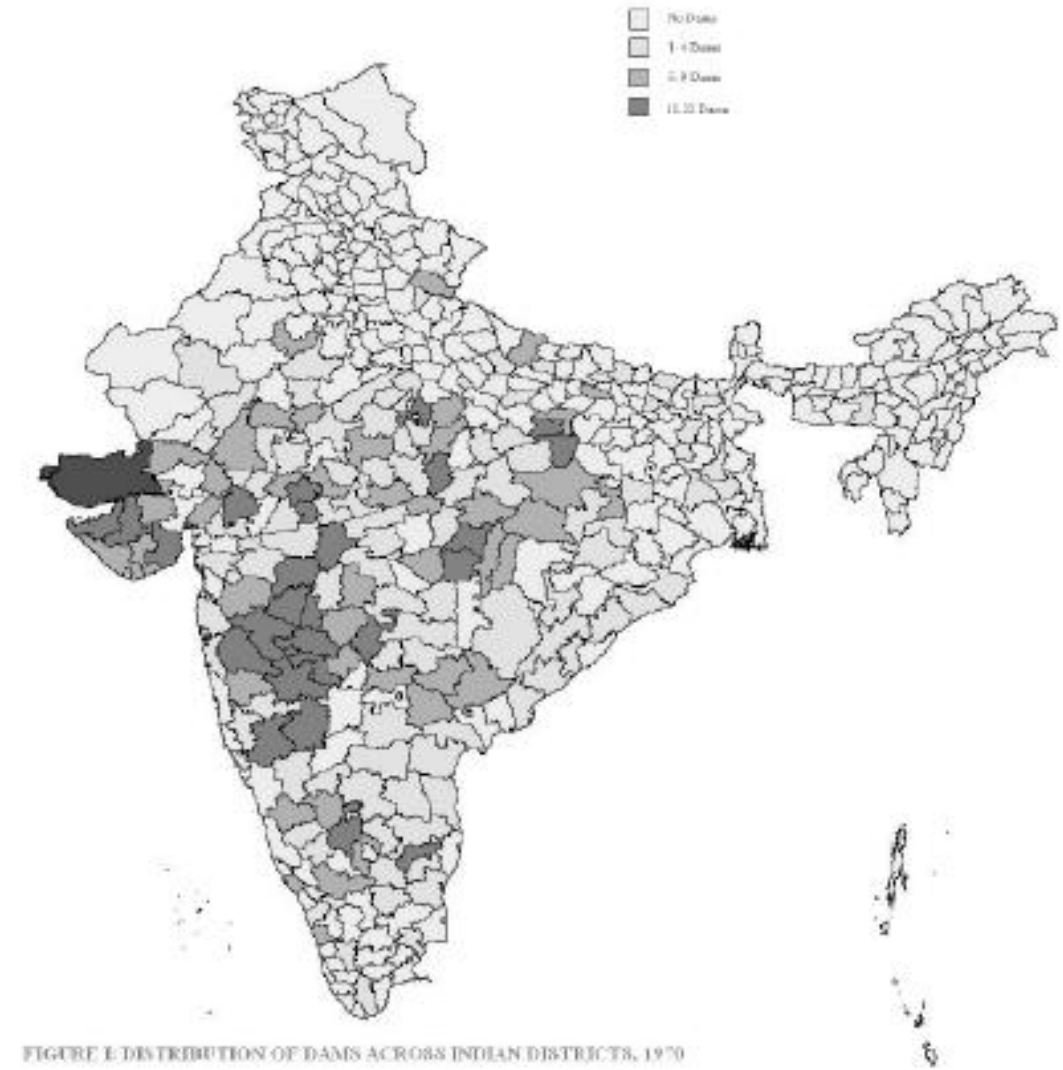


FIGURE I: DISTRIBUTION OF DAMS ACROSS INDIAN DISTRICTS, 1970

Paper examples: other applications

Other applications 1: Topography as an instrument. Pande and Duflo (2007)

Dams by District in 1999

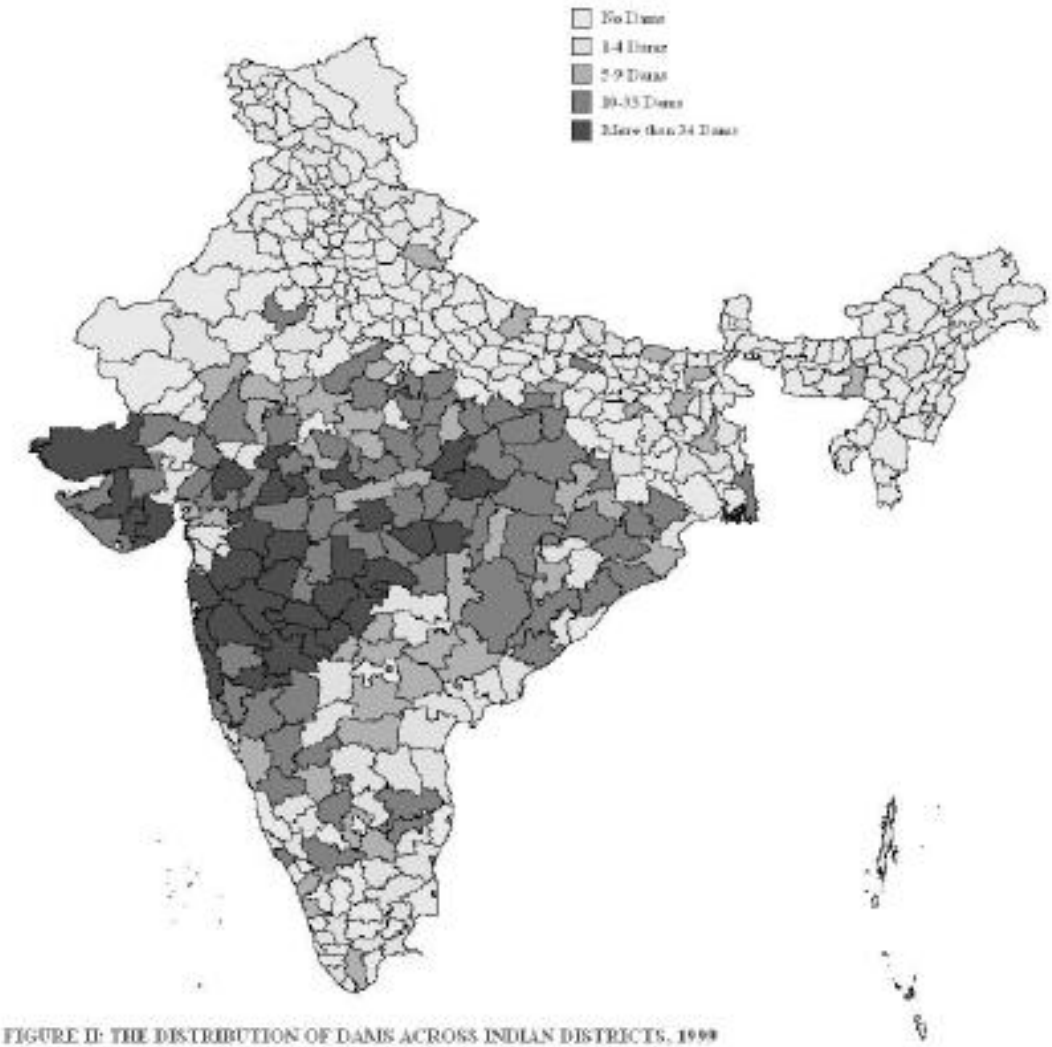


FIGURE II: THE DISTRIBUTION OF DAMS ACROSS INDIAN DISTRICTS, 1999

Paper examples: other applications

Other applications 1: Topography as an instrument. Pande and Duflo (2007)

Average River Gradient (Percentage) by District

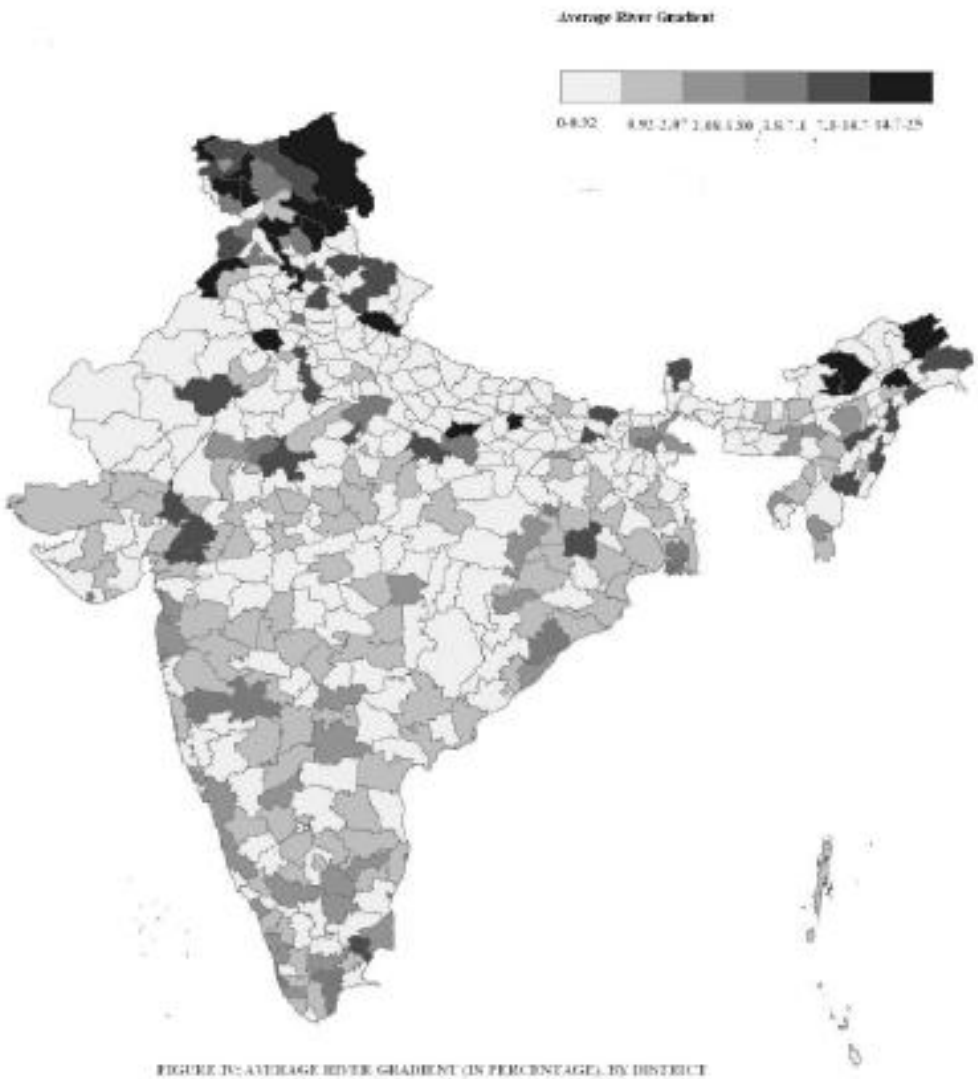


FIGURE IV: AVERAGE RIVER GRADIENT (IN PERCENTAGE) BY DISTRICT

Paper examples: other applications

Other applications 1: Topography as an instrument. Other papers

Other examples of topography instruments

- Dinkelman (2011): land gradient as IV for electrification.
- Qian (2008): land gradient as IV for tea production.
- Olken (2009): signal strength (blocked by mountains) as IV for TV availability.
- Durante, Pinotti, and Tesei (2017): similar strategy as Olken (2009)
- Yanagizawa-Drott (2014): Same idea (see above).
- Lipscomb et al. (2013): Topography measures (river gradient etc.) as IV for hydropower.

Paper examples: other applications

Other applications 2: Spatial RDD. Dell (2010)

Melissa Dell (2010). "The Persistent Effects of Peru's Mining Mita," ECMA. 78 (6) :1863-1903.

Motivation

Are there long-run effects of historical institutions (the forced-labor system of Peru's mining "Mita") on today's development? If so, what are the channels?

Contribution

One of the first to use spatial RDD to cleanly identify effect. Finds economically meaningful effects (25% lower household consumption, 6% increase in childhood stunting). Speaks to channels: "Mita districts historically had fewer large landowners and lower educational attainment. Today, they are less integrated into road networks and their residents are substantially more likely to be subsistence farmers." (persistence through land tenure, public goods, market participation)

Paper examples: other applications

Other applications 2: Spatial RDD. Dell (2010)

Mita boundary

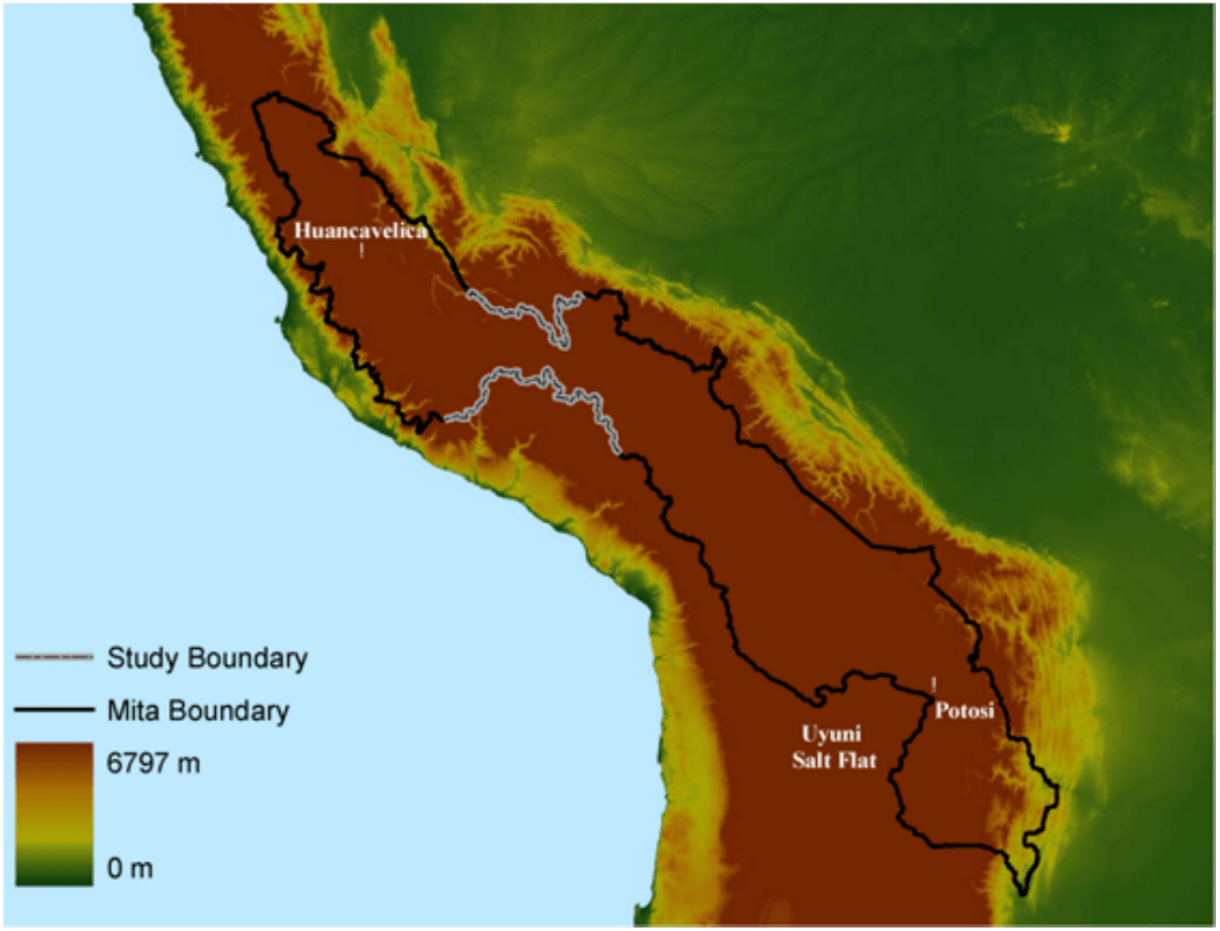


FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling inside the contiguous area formed by the *mita* boundary contributed to the *mita*. Elevation is shown in the background.

Introduction to GIS

Outline

We will cover three things to start with GIS

- Types of geographic data
- Coordinate systems and projections
- Introduction to QGIS

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 common.

It is useful to distinguish three types of feature data

- polygon features
- polyline features
- point features

Introduction to GIS

Data types – polygon features

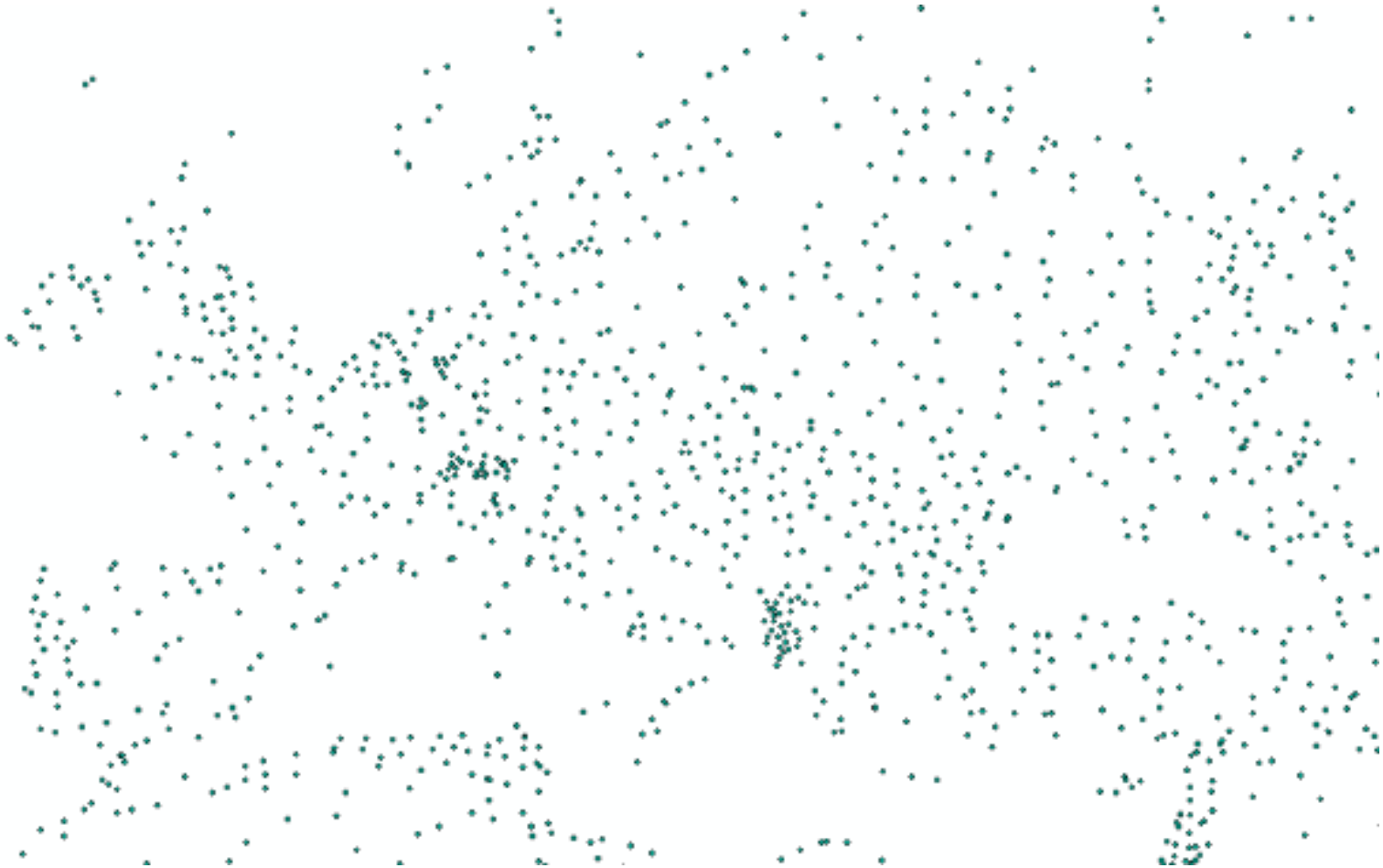
Countries can be treated as polygon features

Rivers as polylines

Introduction to GIS

Data types – point features

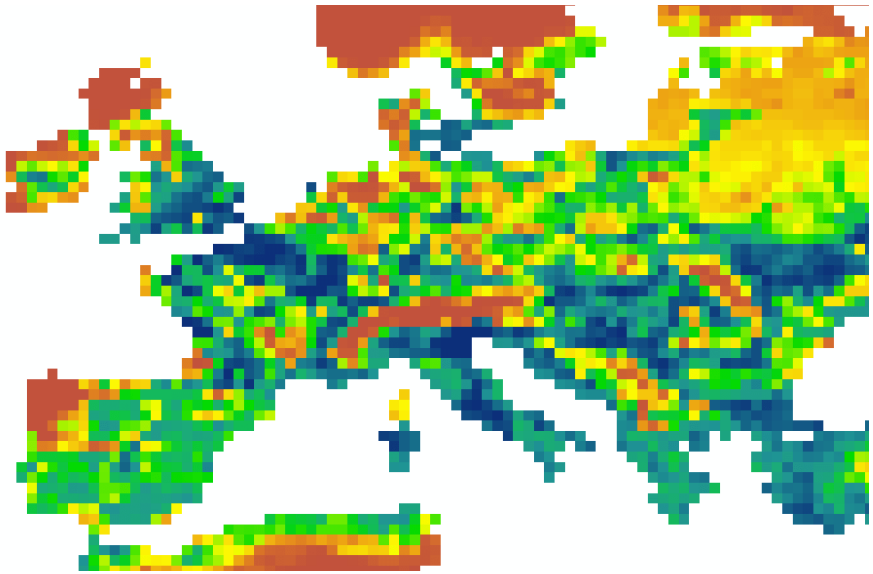
Cities as points



Introduction to GIS

Data types – 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)

Why this matters

- Geospatial data are data with geographic identifiers attached to each data point.
- To view, manipulate, and make calculations with geographic data, the identifiers need to be referenced with respect to a coordinate system.
- Each coordinate system represents the Earth's sphere (or a part of it) in two dimensions.
- Results of calculations depend on the coordinate system used. Using the correct one is crucial!

There are two basic types of coordinate systems

- Geographic coordinate systems.
- Projected coordinate systems.

Based on Melissa Dell's GIS notes.

Geographic coordinate systems represent locations in spherical coordinates .

It is standard to write information for geographic coordinates systems in two ways

- **degrees, minutes, seconds**
- e.g. London is 51° 30' 26" N, 0° 7' 39" W
- **decimal degrees**
- London is 51.5072, -0.1275
- 30 arc minutes and 26 arc seconds $= 30 \times 60 + 26 = 1826$ arc seconds $= \frac{1826}{3600} = 0.5072$ decimal degrees.

The standard geographic coordinate system is WGS 1984 ("World Geodetic System"). It is the only one ever used by economists.

For distances between two points along the Earth's surface, can use **geodesic distance formula**: need only $((lat_1, lon_1), (lat_2, lon_2))$.

Problem

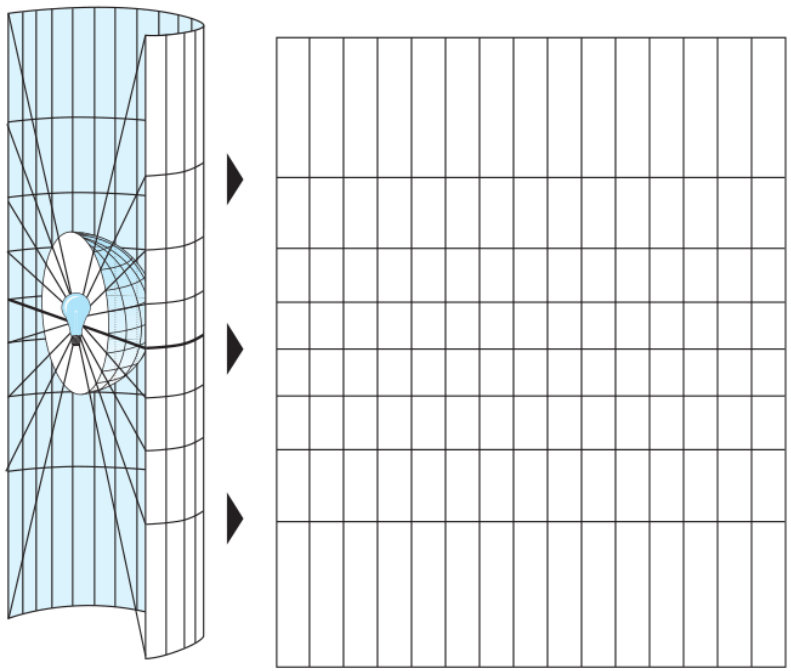
- Geographic coordinate systems are useless to calculate areas and distances along lines.
- Geographic coordinate systems distort lengths and areas: try to wrap a map around an orange!
- 1° latitude is 110.6km at the equator, 111.7km at the poles
- 1° longitude is 111.3km at the equator, 55.8km at 60° N/S

▸ Mercator distortion example

To make accurate area and (some) distance calculations, need to **project** the data from a geographic coordinate system.

One can think of shining a light, placed on the centre of the earth, through the earth surface, and casting a shadow on a projection surface of a certain shape.

Projecting onto a cylindrical projection surface



note stretching of data near the poles

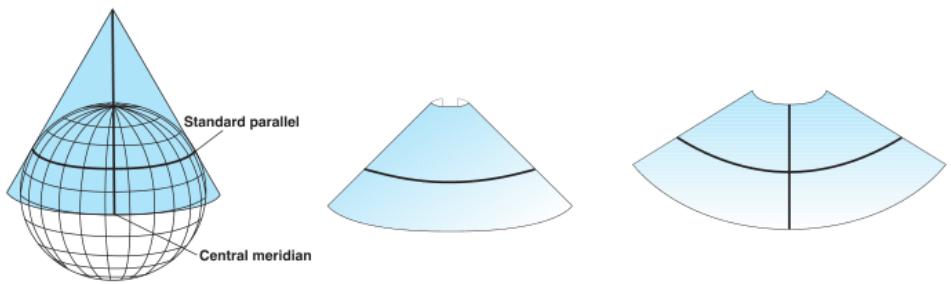
Section based on Melita Kennedy's notes on understanding projections

Useful three-way classification of simple projections

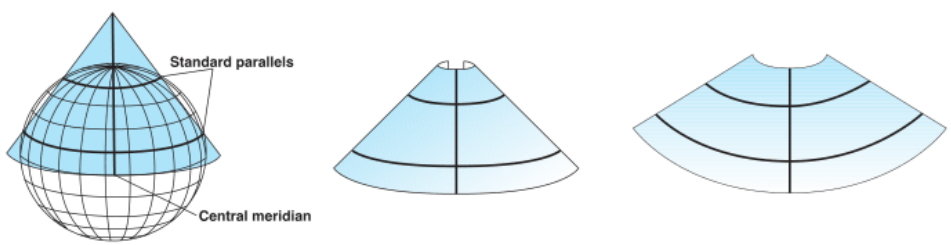
- conic
- cylindrical
- planar

Conic projections

Tangent conic projection, one standard parallel

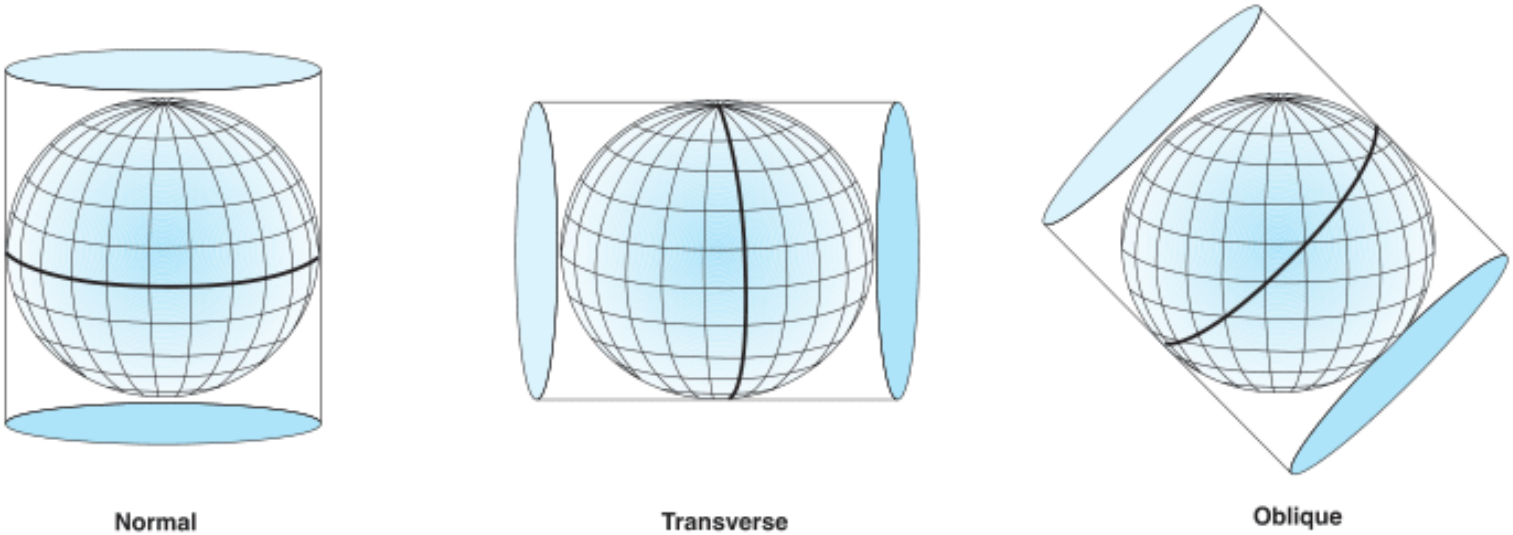


Secant conic projection, two standard parallels



Cylindrical projections

Cylindrical projections



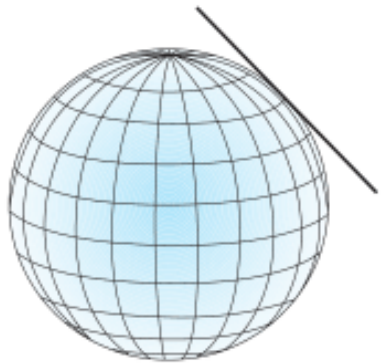
Mercator projection is cylindrical with equator as line of tangency.

Planar projections

Planar projections with different point of tangency (*aspect*)

Polar

Equatorial



Oblique

WGS 1984

- for distance between two points

UTM: divides surface of the earth into many regions, each gets its own projection

- distance/surface area in small regions
- length of polylines

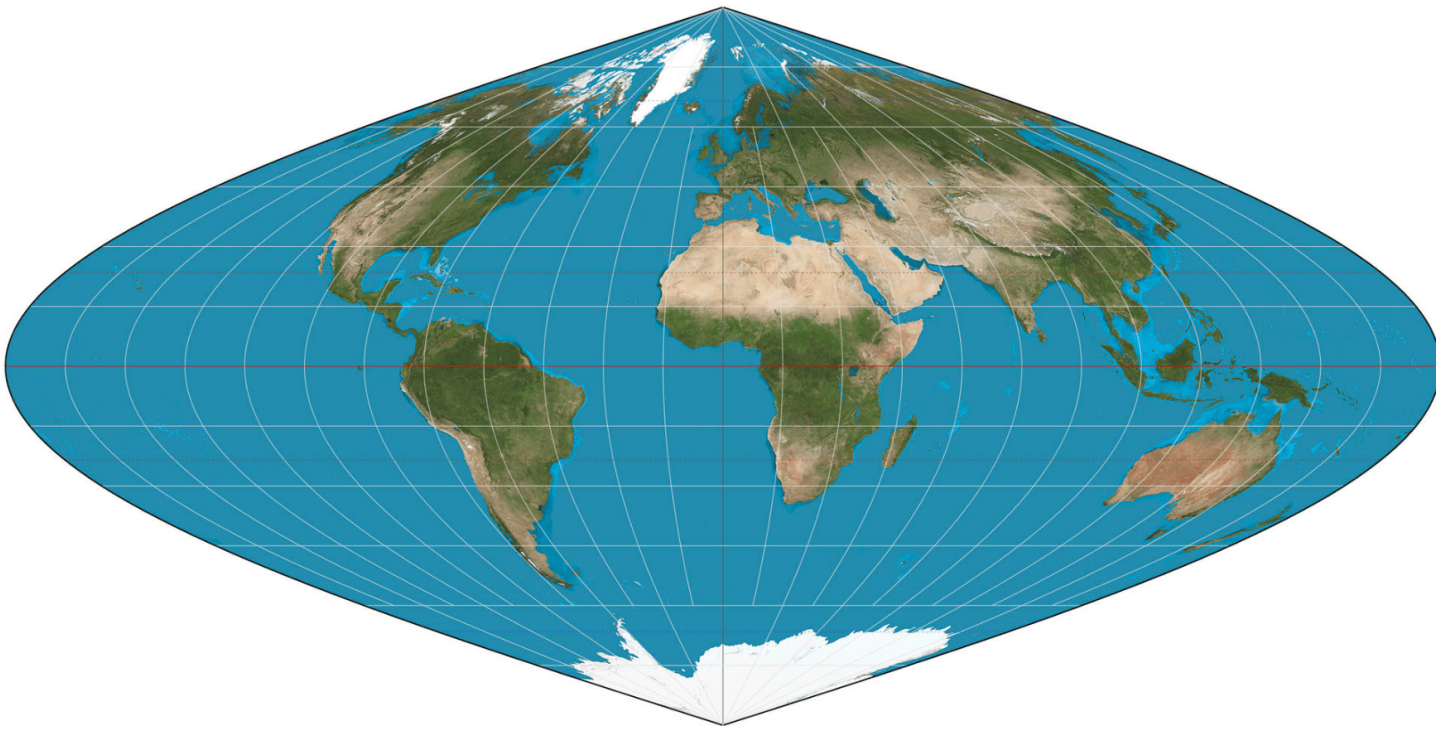
Any equal area projection

- surface area in large regions

This and the following three slides are based on Masayuki Kudamatsu's GIS course

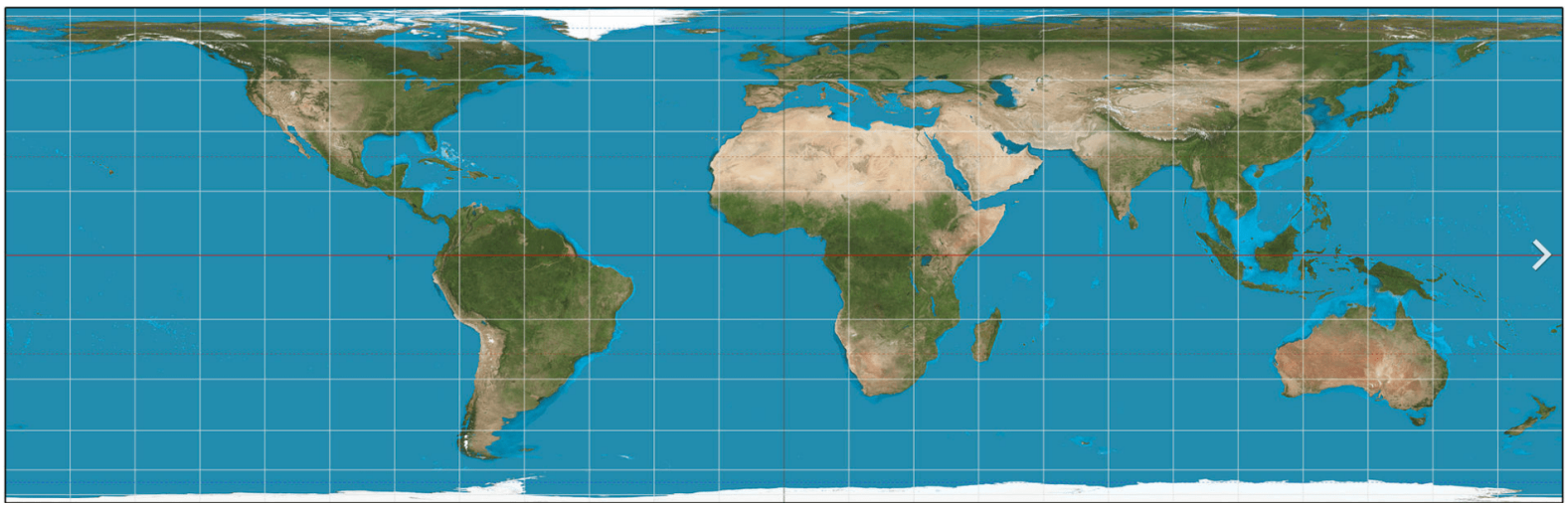
Differ just in how the world is shown

Sinusoidal projection



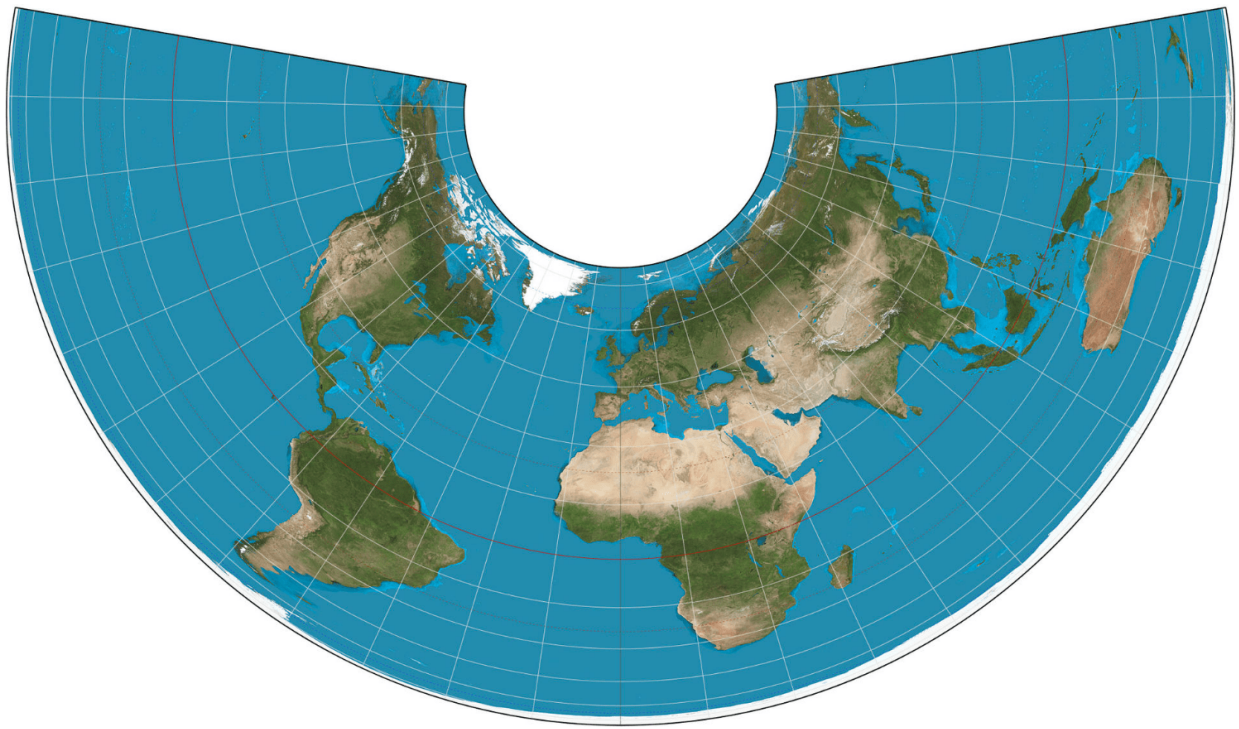
Differ just in how the world is shown

Lambert cylindrical equal area projection



Differ just in how the world is shown

Albers Equal area conic projection



Introduction to GIS

QGIS – What is it good for?

QGIS is a programme for analysing, modifying, and creating geo-spatial data.

Some of the things it can do:

- merge and intersect shapes based on geographical relationships
- finding nearest objects according to specified metric
- calculate slope from elevation data
- take averages within polygons
- finding shortest path in a network
- make maps
- ...

Go here:
`https://qgis.org/en/site/forusers/download.html`

- Go to the tab for your operating system
- Choose the appropriate standalone installer for your system
- We recommend the "Latest release (richest on features):"
- Execute the installer, choose default options, don't install the datasets

Mac users may face an "unidentified developer" problem. See here for solutions:
`https://kb.wisc.edu/helpdesk/page.php?id=25443`

Mac users may also encounter an issue with `GDAL` geoprocessing tools (we will need this later). See here for a solution:

`https://gis.stackexchange.com/questions/276853/gdal-scripts-not-found-in-qgis-3-on-osx`

Introduction to GIS

QGIS – Installing QGIS, to run pyqgis from the command line (tested only on Windows)

Optional

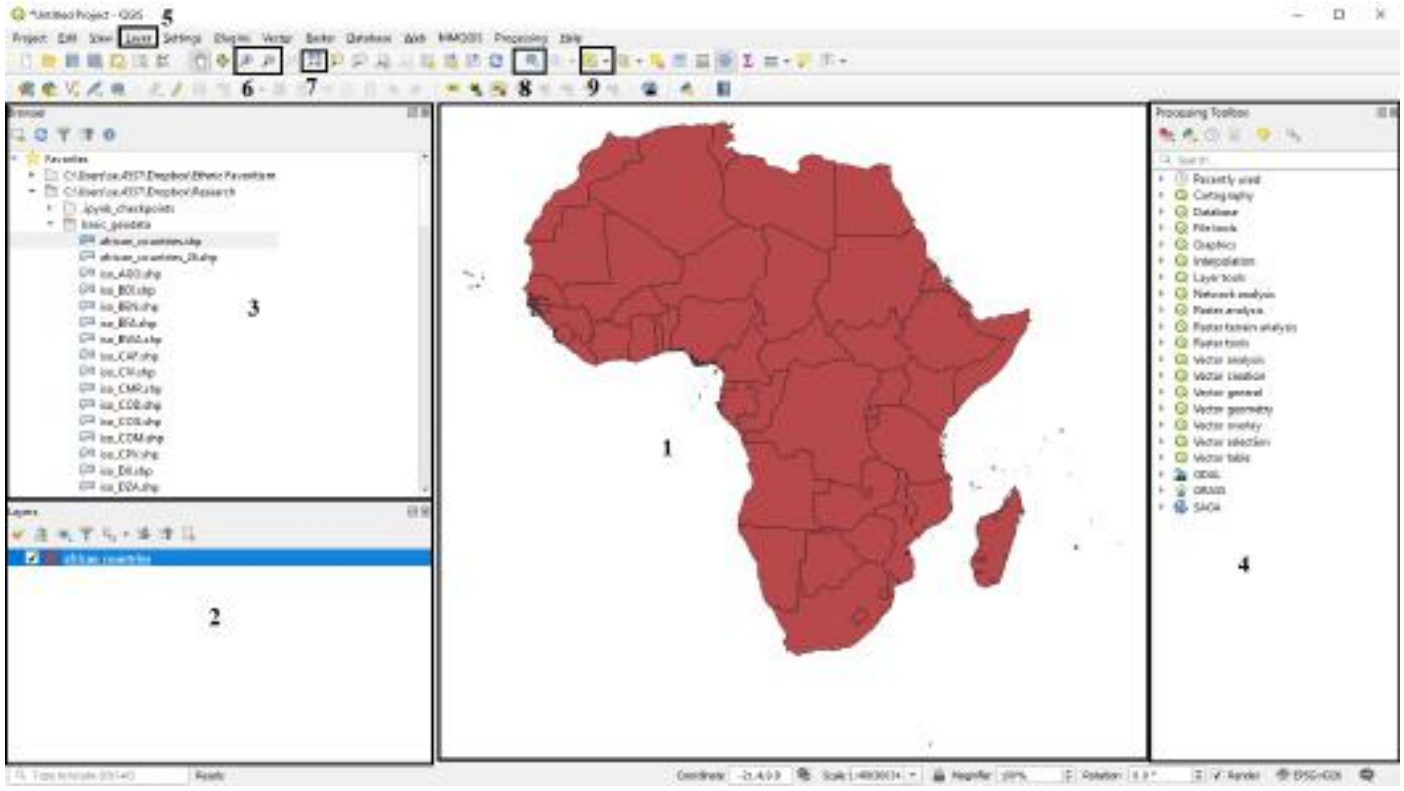
- instead choose OSGeo4W Network Installer (64 bit) or (32 bit), depending on your system
- Choose Express Desktop Install
- keep the default checks (QGIS, GDAl, GRASS GIS) and install all dependencies
- agree with all the license terms
- this will put a folder *OSGeo4W* on your desktop
- inside the OSGeo4W folder, double-click on OSGeo4W Shell
- you may get a message that says "The system cannot find the path specified"
- if so, go do `YOUR_INSTALLATION_DIRECTORY/OSGeo4W64/etc/ini` and delete `rbatchfiles.bat`
- reopen the OSGeo4W Shell (this should have taken care of the error)
- type: `python-qgis`
- type: `import qgis`
- type: `import pandas as pd`
- type: `quit()`
- in order to run **GRASS GIS** algorithms, you will also need to set the environment for that correctly:
  open `OSGeo4W.bat` in `YOUR_INSTALLATION_DIRECTORY/OSGeo4W64` with a text editor and paste the line
  `call "%~dp0\apps\grass\grass78\etc\env.bat"`
  right underneath the line
  `call "%~dp0\bin\o4w_env.bat"`

Intalling QGIS will put a bunch of programmes on your computer.

- We will introduce you to QGIS – the software which allows you to work with geospatial data.
- QGIS has a built in "Browser" – a type of file viewer.
- As you will soon learn, doing anything in QGIS can produce a lot of intermediate files at every step.
- Every shapefile, for example, comes with an *.shp* file (which stores the feature geometry itself), a *.dbf* file (which stores the attribute table, see below), a *.prj* file (which stores the projection), and often with an *.shx* file (which stores a positional index of the feature geometry to allow seeking forwards and backwards quickly).
- The Browser is can make managing this soup of files easier as you always just see one file.
- In practice, it is better to just automate as much as you can in python and access the Browser from within QGIS if you absolutely have to.

Introduction to GIS

QGIS – What the buttons mean



1. "Canvas": data is visualized here
2. "Layers": which datasets are loaded?
3. "Browser": quickly load datasets from disk locations, delete data
4. Processing toolbox

   Note: showing 2-4 is a good default configuration, but other windows can be shown (see "View" → "Panels")

5. Among other things, add layers (data) to canvas
6. Zoom in and out
7. Zoom to full extent (very useful)
8. Get information on geographic features
9. Select elements

Introduction to GIS

QGIS – Adding data

Download some data

- from the google drive or from `http://www.naturalearthdata.com/downloads/10m-cultural-vectors/`
- download "Admin 0 – Countries", save to some directory and unzip

Add directory to "Favorites"

- facilitates loading data
- navigate to the folder containing the file you downloaded and add it to favorites
- sometimes may have to refresh ( ) to see new data

Add the data

- In the folder you just added, right-click, double-click (or do the mac thing) to "Add Layer to Project" or single click and click on "add layer" ( )
- Can also go more complicated: "Layer" → "Add Layer" → "Add Vector Layer" → "Browser" → navigate to your folder and add the data.

You should see a political map of the world in the main window and an entry in "Layers" listing the dataset.
We have loaded a file with **polygon features**. Features come with "attributes".

Attribute table

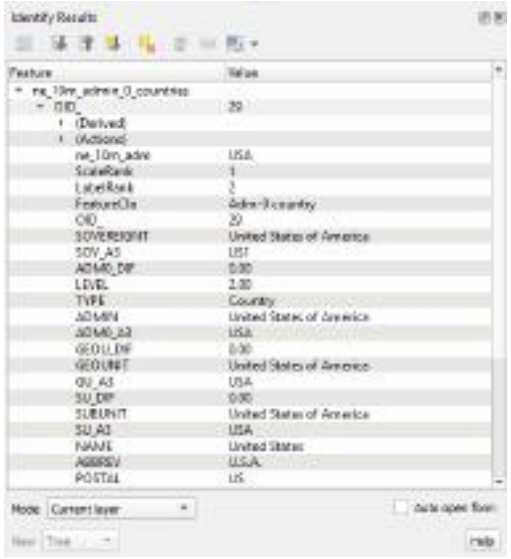
- right click on *ne_10m_admin_0_countries* in Layers → Open Attribute Table
- a table opens, listing a bunch of variables
- each row in the table corresponds to one polygon feature on the map
- pick a country and click on the small grey square at the start of its row with the number to select the row and close the attribute table
- the country you picked is highlighted
- click on Zoom to Selection ()
- click on  to zoom back out
- click on  to de-select all features
- in select mode (), you can also control-click on a feature on the canvas to un-select it
- play around with zooming, panning, selecting, until you are comfortable

Introduction to GIS

QGIS – Inspecting the data

We can also use **Identify Features** to view the information in the attribute table for one or more features

- click on  and click on the US
- you should see the window on the right
- this lists all the variables in the attribute table for this particular feature
- we can select more than one feature this way (hold down the mouse and drag it over several features)
- then expand and collapse individual feature attribute lists to find the one(s) we are interested in



Extent and coordinate system

- back in Layers, right click on the layer and "Properties" at the bottom
- select the "Information" tab and check "Information from provider"
- select the "Source" tab

Change color of countries and outline width

- Back under Layers, right click the *ne_10m_admin_0_countries* layer → Properties → Symbology
- Change Fill color, outline ("Stroke") color, and outline width
- Click "OK" and "Apply" and see how the canvas display changes

Label features

- Back to Layer Properties window, select the "Labels" tab
- From the drop-down menu at the top, choose "Single Labels"
- From the "Label with" drop-down menu, select "ADMIN" (this is one of the variables from the attribute table)
- Click "OK" and "Apply" and see how the canvas display changes

Hide a dataset

- in Layers, uncheck *ne_10m_admin_0_countries*

Introduction to GIS

QGIS – Adding and inspecting some raster data

Download some data

- from the google drive or from `https://nelson.wisc.edu/sage/data-and-models/atlas/data.php?incdataset=Suitability%20for%20Agriculture`
- download the data, save to some directory and unzip

Add the data

- exactly as with the feature data (the data are under *suit/hdr.adf*)
- ignore the warning message (we will deal with this in a later lecture)

Inspect the data

- zoom in closely so you can make out individual cells
- use  to look up individual pixel values
- look in the source tab of the Properties menu (note the absence of a CRS)

Change the color scheme

- Layer → Properties → Symbology
- Band Rendering → Render Type → Singleband pseudocolor → choose Linear Interpolation and your favorite Color Ramp → OK + Apply

So far we have added data downloaded from the web. There is one type of spatial data that we can easily create ourselves: **Point features**.

Create point features in text editor

- open a text editor
- in the first line, type: *point_name, latitude, longitude*
- in the second line, type: *some name, -12.3, 117.2*
- in the third line, type: *some other name, 65.2, -56.3*
- keep adding as many points as you like, until you're bored.
- make sure latitude $\in [-90, 90]$, longitude $\in [-180, 180]$
- save the data under *my_points.csv* in the directory where you saved the files you downloaded

Introduction to GIS

QGIS – Adding and inspecting latitude/longitude data

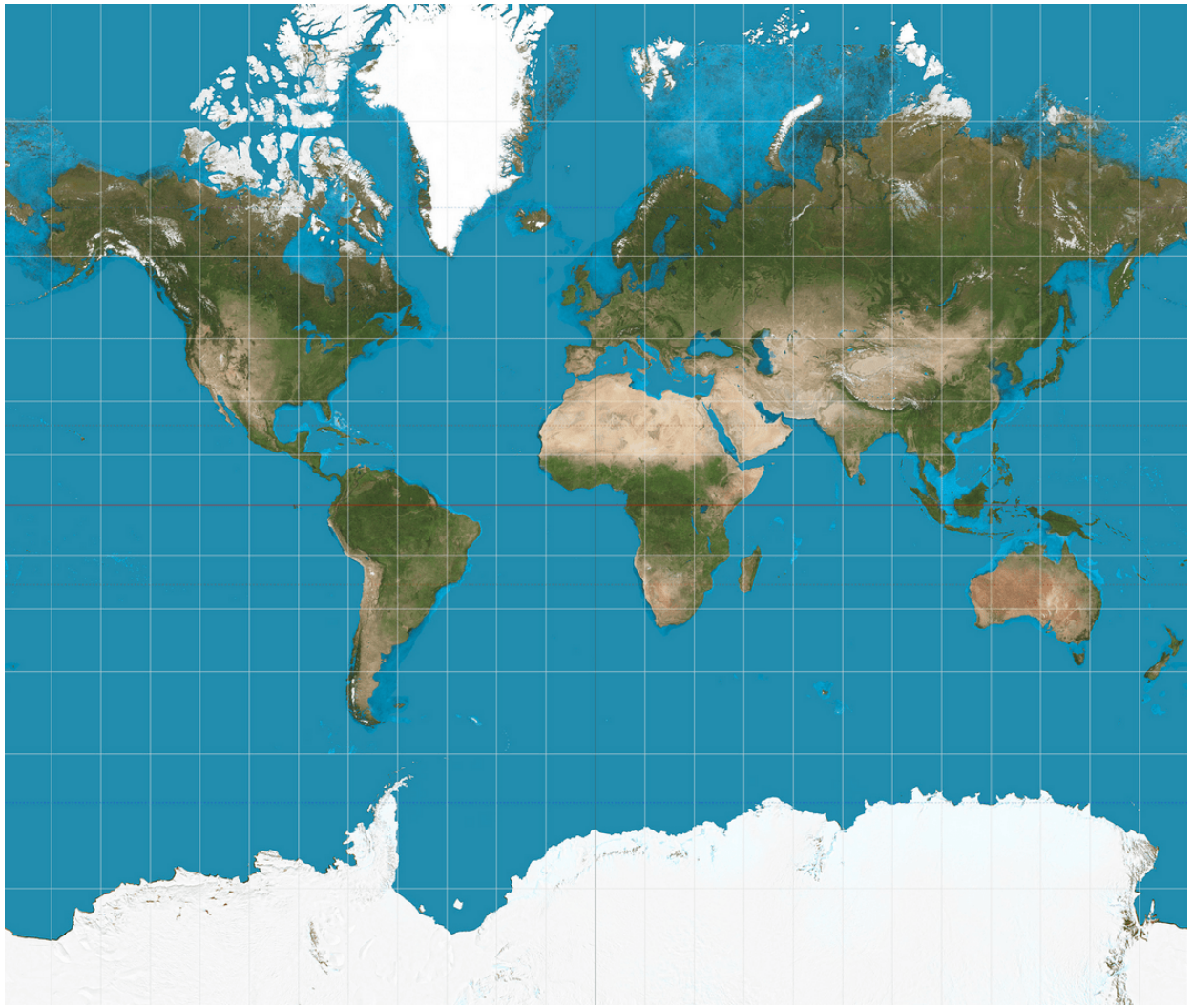
Add and display data

- Layer → Add Layer → Add Delimited Text Layer → Under File name, browse to the point layer; Pick a layer name
- File Format → CSV
- Geometry Definition → Point coordinates
  - X field: longitude
  - Y field: latitude
  - Geometry CRS: Project CRS: EPSG: 4326 – WGS 84
- Click "Add"
- dots appear on the map

Inspect the data

- look at the attribute table
- use Identify Features () to get information on individual points

The world in Mercator projection



The actual sizes of Greenland and Australia

nice *The Economist* article on map projections:

`https://www.economist.com/blogs/graphicdetail/2016/12/daily-chart-1?fsrc=scn/fb/te/bl/ed/misleadingmapsandproblematicprojections`

back