Introduction to GIS Methods in Economics

Giorgio Chiovelli   Sebastian Hohmann

Bonn, 27/05/2019

Table of contents

Three aims

- Give you a (selective) overview of how *Geographic Information Systems* (GIS) methods are used in current economic research.
- Introduce you to ArcGIS software and GIS scripting in python.
- Show you how you can "do GIS" for your own research through detailed examples of well published papers.

We want to make this useful for you!

- If you bring your laptop with ArcGIS installed, you can follow what we are doing in class.
- Depending on whether you have concrete questions, we can leave the last 10 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, simple examples.
2. (today) Introduction to python, GIS scripting in python.
3. (28/05/2019) replication of Michalopoulos, Stelios. (2012). "The Origins of Ethnolinguistic Diversity," American Economic Review, 102(4): 1508-1539. Q&A
4. (28/05/2019) Selected topics: (e.g. geocoding, digitizing maps, multidimensional data, interpolation, hot spot analysis)
5. (29/05/2019) 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. (29/05/2019) Optional (if time): short introduction to Google Earth Engine

Material for programming. Google drive:
`https://drive.google.com/open?id=1VIFriPOYZ84MYENXNxYK-kjuIfCabkLN`

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 5) 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)

Figure 1: Night lights in 2008

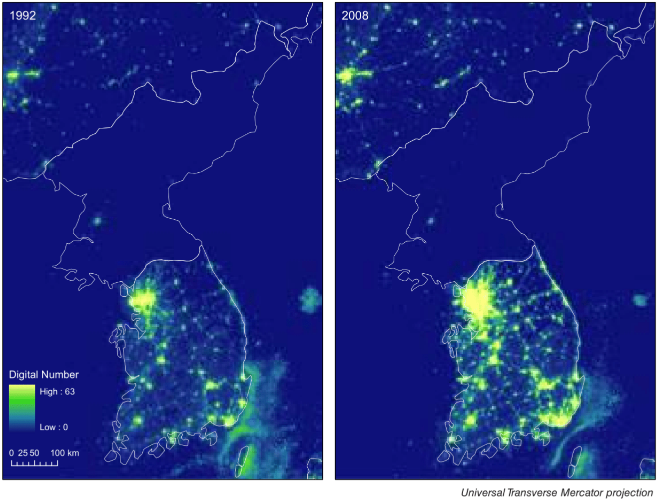
source:

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

Paper examples: satellite data

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

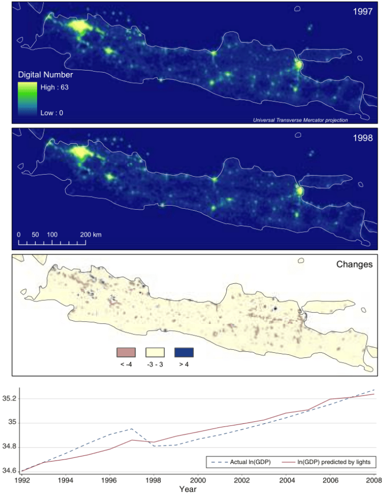
Figure 2: Long-term growth: Korea



Paper examples: satellite data

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

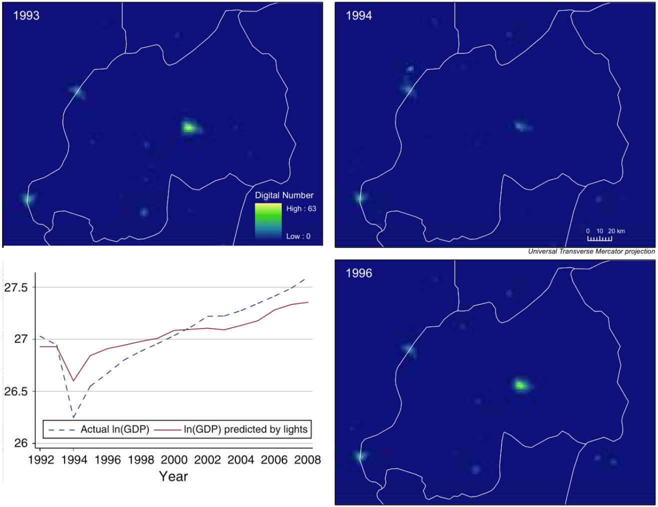
Figure 3: Short-term economic shocks: Indonesia 1997/1998



Paper examples: satellite data

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

Figure 4: 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

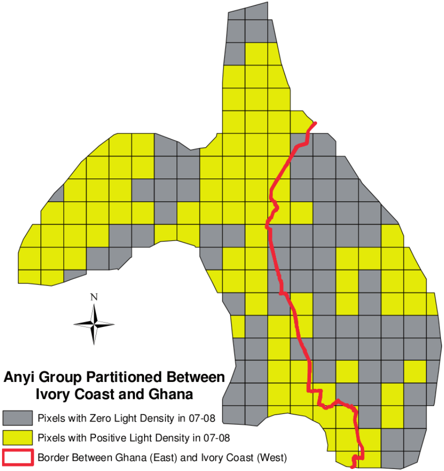
Night lights 2: Institutions and development. Michalopoulos and Papaioannou (2014)

Figure 5: Pre-colonial ethnic and modern national boundaries

Paper examples: satellite data

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

Figure 6: An example of a split ethnic group



Estimation (Linear probability model)
p = pixel
e = ethnicity
c = country

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

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)

Figure 7: Mobutu in Zaire / DRC

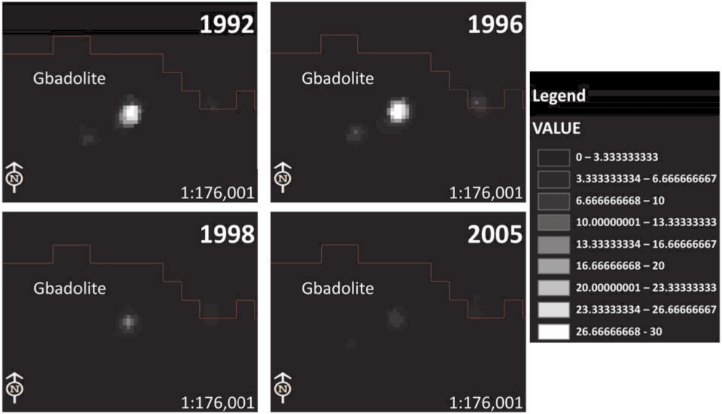


FIGURE I

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)

Figure 8: Mahinda in Sri Lanka

FIGURE II

Nighttime 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)

Figure 9: Deforestation in Riau province

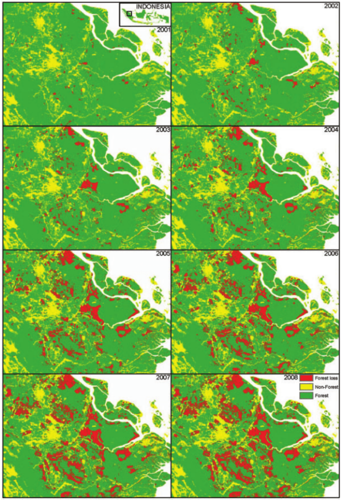


FIGURE I

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

Paper examples: satellite data

Deforestation: Burgess et al. (2012)

Figure 10: Deforestation in Indonesia 2001-2008

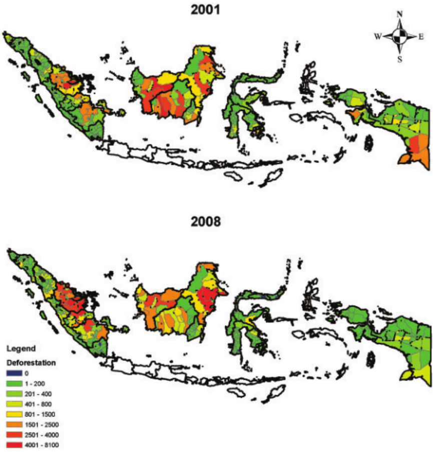


FIGURE II

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)

Figure 11: Change in pollution around the games from the paper, AOD data*

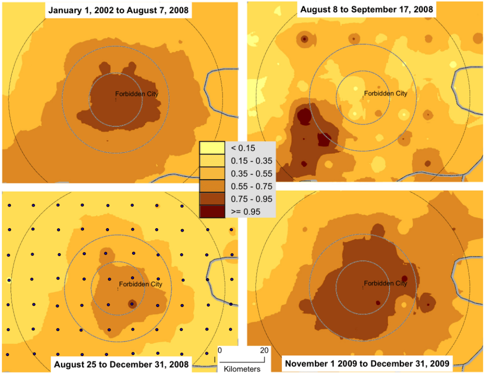


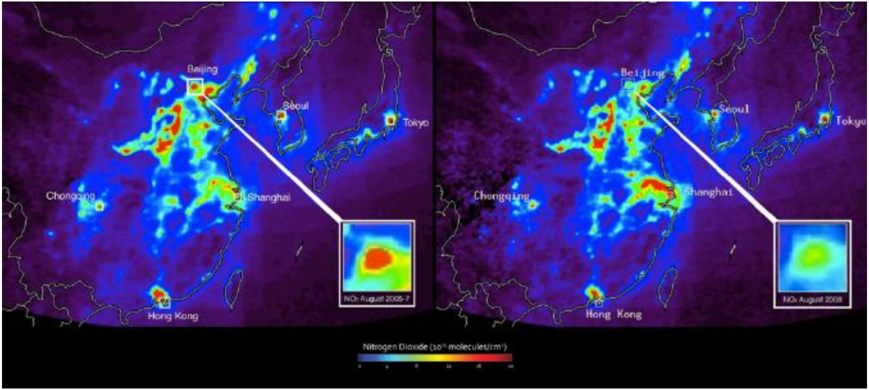
Fig. 7. Satellite-based AOD, corrected for meteorological conditions and spatiotemporal trends in and around Beijing.

* "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)

Figure 12: NO_2 concentration before and during the games



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)

Figure 13: 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)

Figure 14: 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)

Figure 15: Constructing slavery-intensity step 2

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

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

Motivation

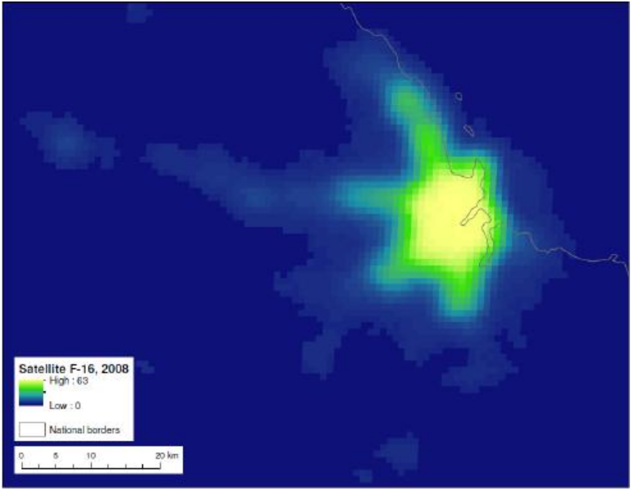
How inter-city connectivity determines the income of sub-Saharan African cities.

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).

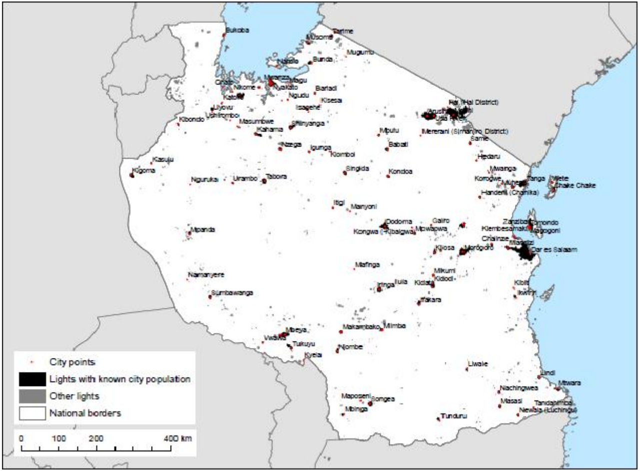
Paper examples: network

Figure 16: Luminosity (2008) in Dar el Salaam



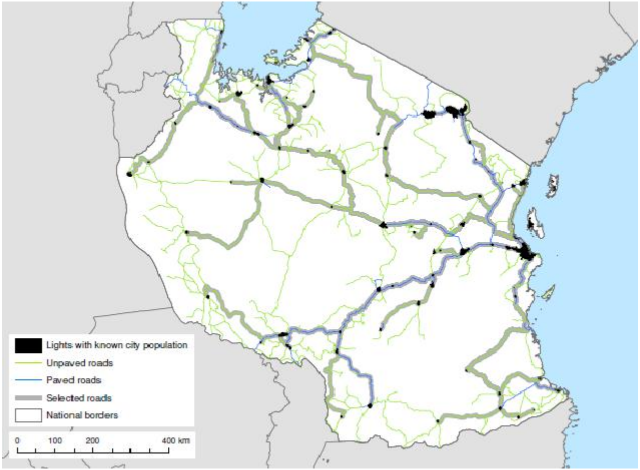
Paper examples: network

Figure 17: City Distribution in Tanzania



Paper examples: network

Figure 18: Roads connection to Dar el Salaam (Tanzania)



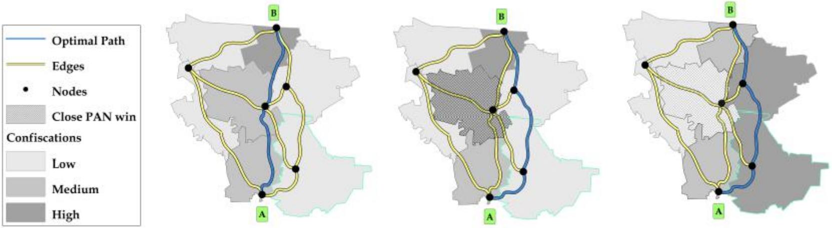
Dell, Melissa (2016). "Trafficking Networks and the Mexican Drug War" AER

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

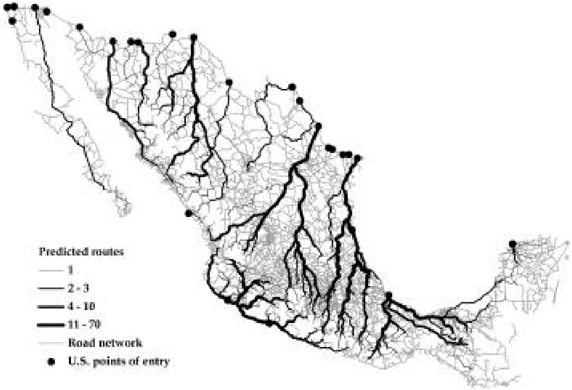
Figure 19: Spillover Methodology



(a) Legend (b) Basic environment (c) Close PAN victory increases costs; routes change (d) Do confiscation patterns also change?

Paper examples: network

Figure 20: 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 reduce form framework to quantify the spillover effect of infrastructure project.

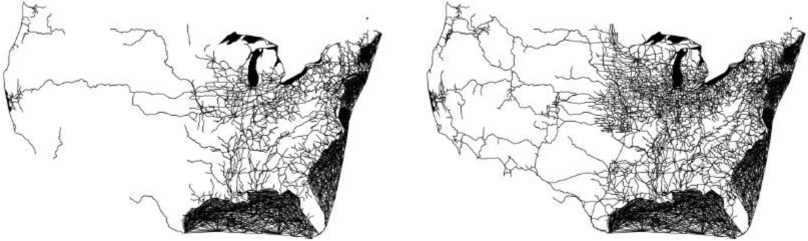
Railways expansion fostered US economic growth. Aggregate effects are considerably larger than local effects (due to higher "market access").

Paper examples: network

Figure 21: Railway expansion 1870-1890

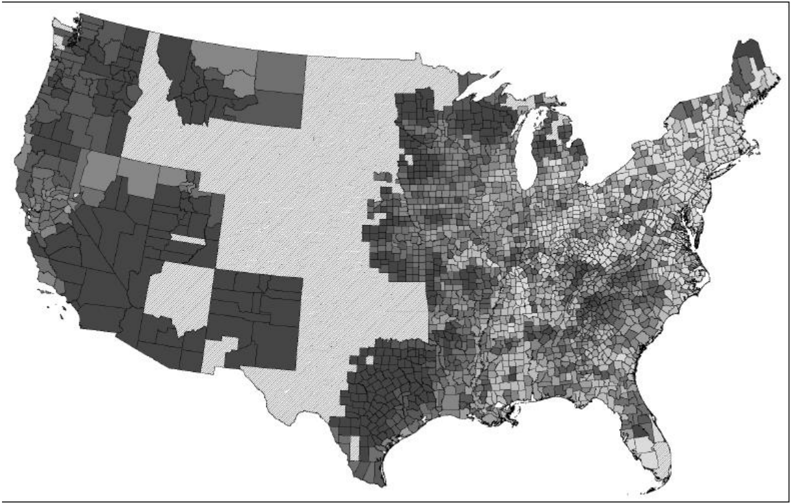
C. Natural Waterways, Canals, and 1870 Railroads

D. Natural Waterways, Canals, and 1890 Railroads



Paper examples: network

Figure 22: 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!

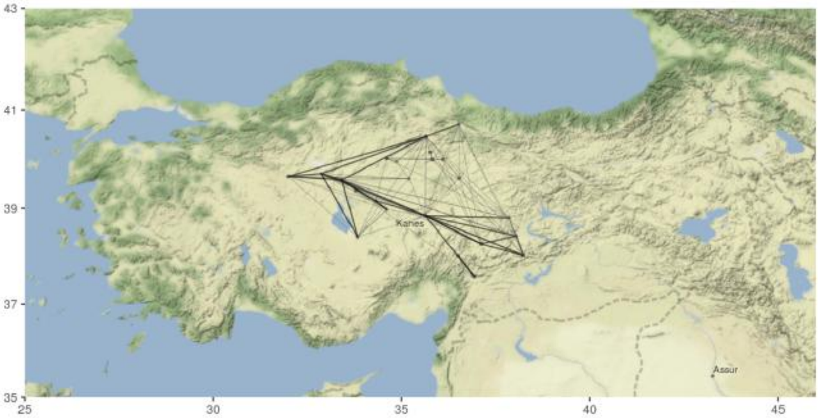
Contribution

- 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!
- model further delivers an estimate of city size.

Paper examples: network

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

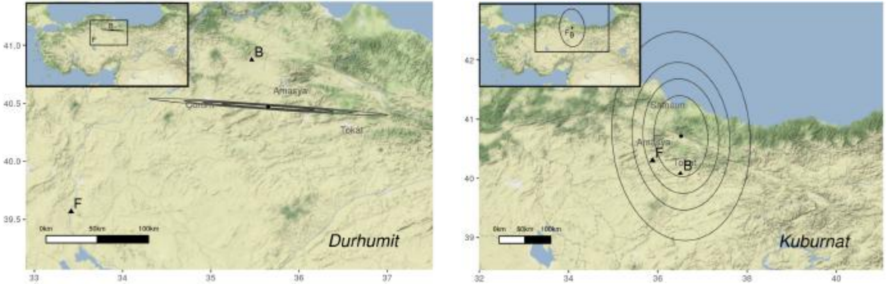
Figure 23: Ancient cities and trade routes



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

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

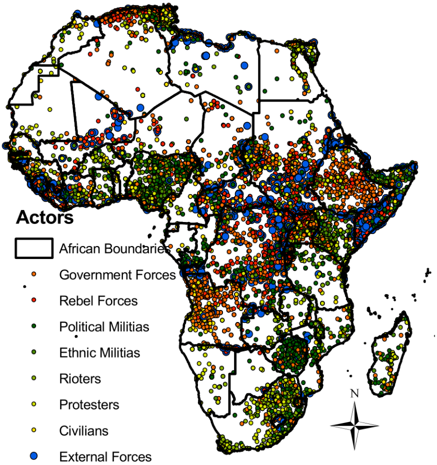
Motivation
Did the arbitrary partitioning of ethnic groups during the "Scramble for Africa" in the 19th century create a long-lasting legacy of violent conflict?

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)

Figure 24: 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?

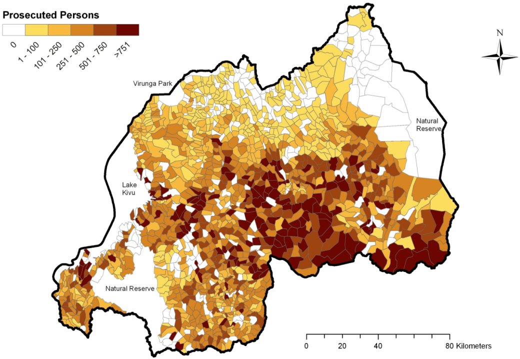
Contribution
Exploits quasi-random variation in hate radio signal strength (combination of transmitter location and Rwanda's mountainous terrain) to show that broadcasts increased participation in killings by locals. Around 10% of deaths can be attributed to radio*.

* 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)

Figure 25: Genocide violence in villages

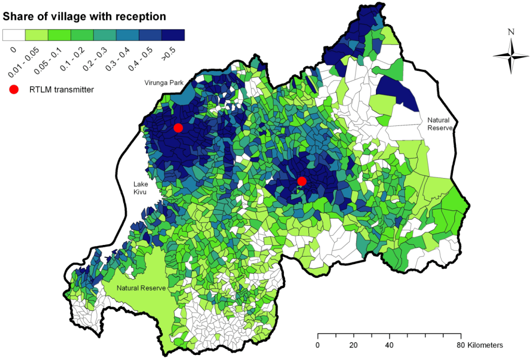


The categories represent the total number of prosecuted persons in the village (sum of militia and individual violence). White areas are missing data, either because of geography, such as parks and natural reserves, or villages that lack data in the sample.

Paper examples: conflict

Conflict 2: Propaganda. Yanagizawa-Drott (2014)

Figure 26: 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

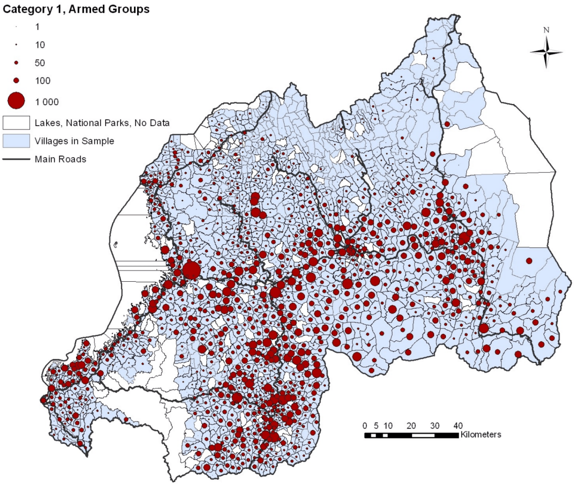
Exploits quasi-random variation in transport costs of militias in accessing villages created by different amounts of rain falling on unpaved roads during the genocide.

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)

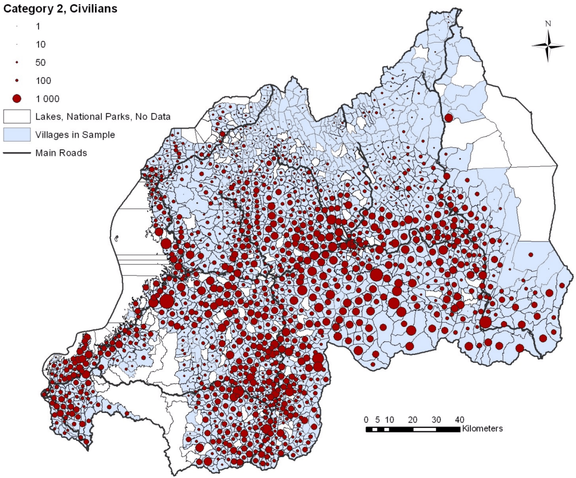
Figure 27: Violence by armed groups



Paper examples: conflict

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

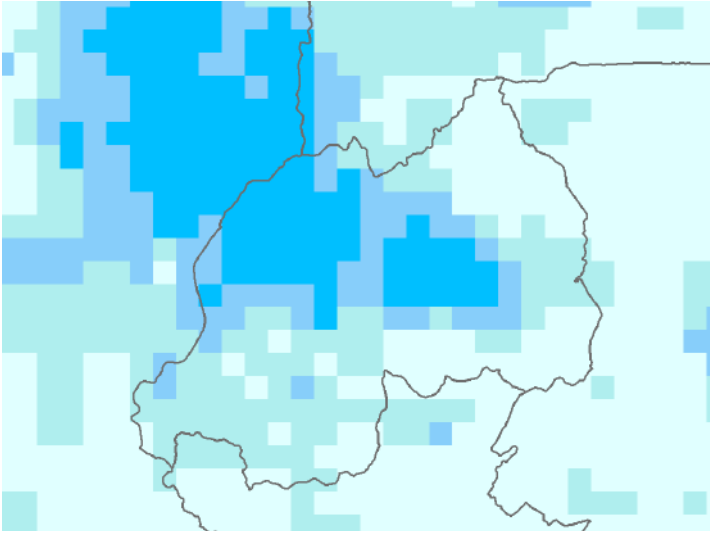
Figure 28: Violence by civilians



Paper examples: conflict

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

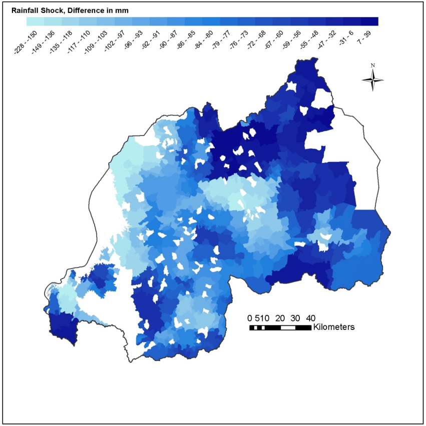
Figure 29: Rainfall Data Example - NOAA (NASA)



Paper examples: conflict

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

Figure 30: 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)

Figure 31: 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.

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)

Figure 32: Dams by District in 1970

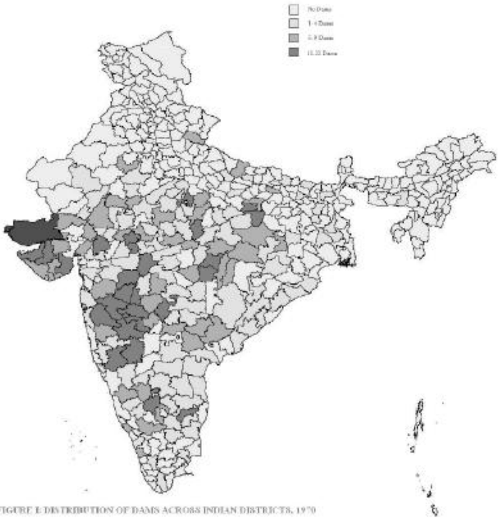


FIGURE I: DISTRIBUTION OF DAMS ACROSS INDIAN DISTRICTS, 1970

Paper examples: other applications

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

Figure 33: Dams by District in 1999

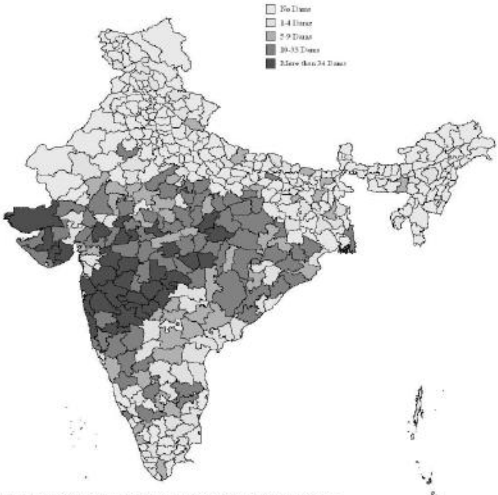


FIGURE II: THE DISTRIBUTION OF DAMS ACROSS INDIAN DISTRICTS, 1999

Figure 34: Average River Gradient (Percentage) by District

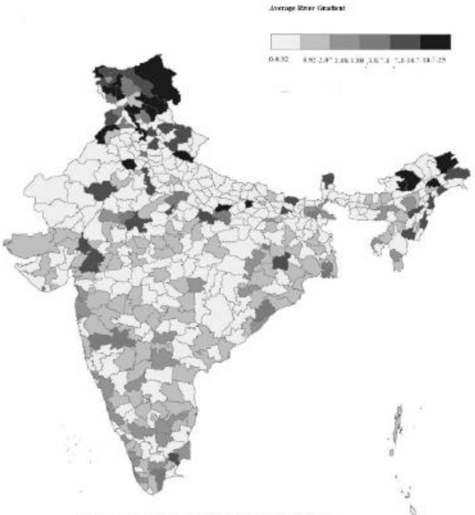


FIGURE IV: AVERAGE RIVER GRADIENT (IN PERCENTAGE), BY DISTRICT

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.

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)

Figure 35: Mita boundary

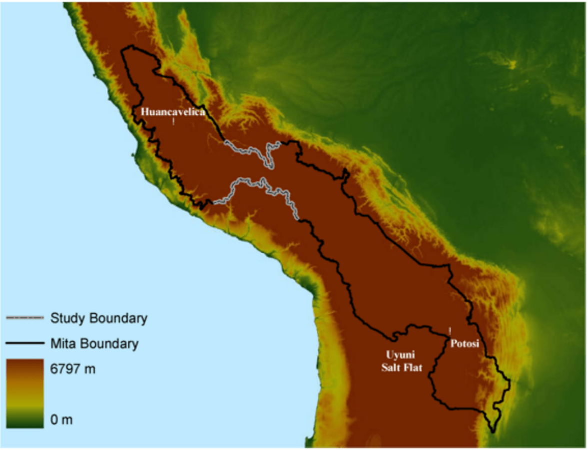


FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling inside the contiguous area formed by the *mita* boundary contributed to the *mita*. Elevation is shown in the background.

We will cover three things to start with GIS

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

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.

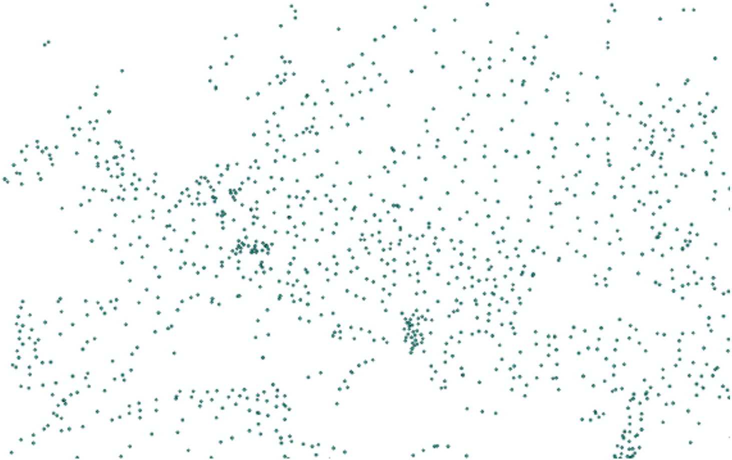
It is useful to distinguish three types of feature data

- polygon features
- polyline features
- point features

Figure 36: Countries can be treated as polygon features

Figure 37: Rivers as polylines

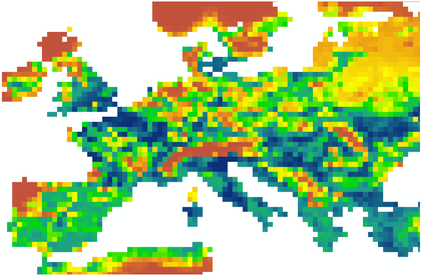
Figure 38: Cities as points



Introduction to GIS

Data types – raster data

Figure 39: 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. Bonn University is 50° 44' 2" N, 7° 6' 8" E
- **decimal degrees**
- Bonn University is 50.73389, 7.102222
- 44 arc minutes and 2 arc seconds $= 44 \times 60 + 2 = 2640$ arc seconds $= \frac{2642}{3600} = 0.733\overline{8}$ 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

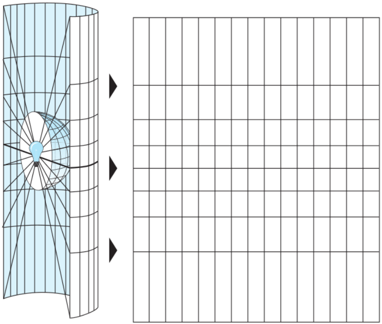
- Geographic coordinate systems are useless to calculate areas and distances along lines.
- Geographic coordinate systems distort lengths and areas: try to wrap a map around an orange!
- 1° latitude is 110.6km at the equator, 111.7km at the poles
- 1° longitude is 111.3km at the equator, 55.8km at 60°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

Figure 40: 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

Figure 41: Tangent conic projection, one standard parallel

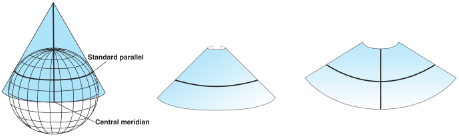
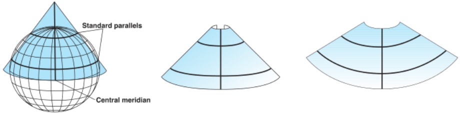
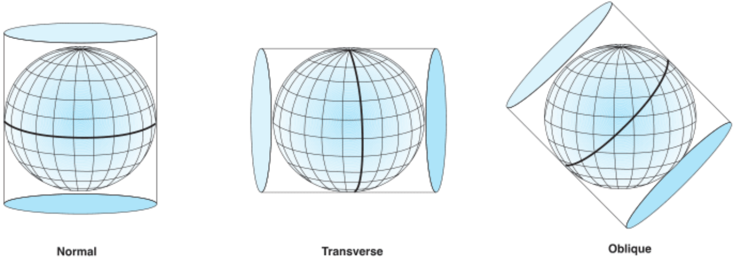


Figure 42: Secant conic projection, two standard parallels



Cylindrical projections

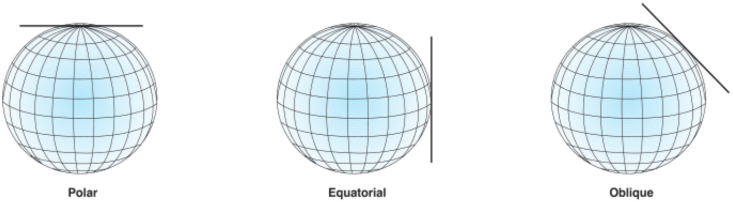
Figure 43: Cylindrical projections



Mercator projection is cylindrical with equator as line of tangency.

Planar projections

Figure 44: Planar projections with different point of tangency (*aspect*)



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

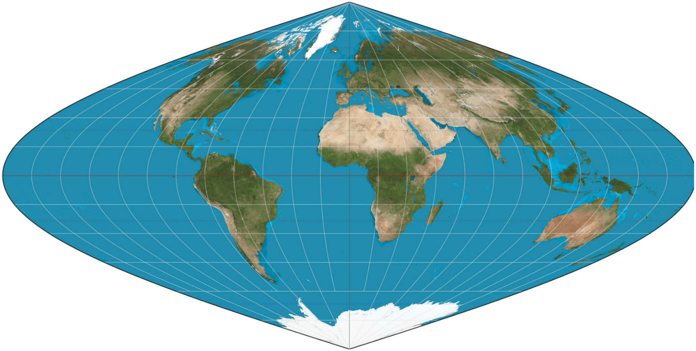
Any equal area projection

- surface area in large regions

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

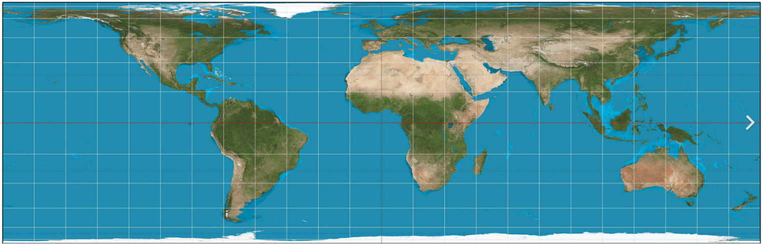
Differ just in how the world is shown

Figure 45: Sinusoidal projection



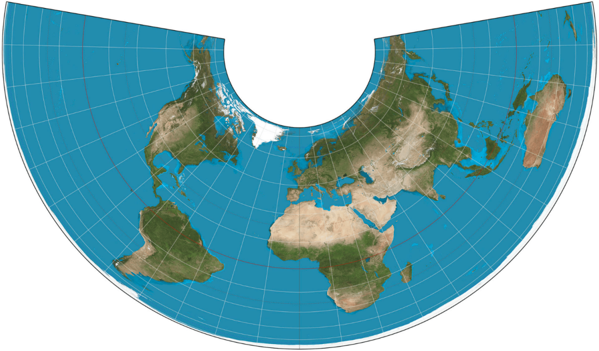
Differ just in how the world is shown

Figure 46: Lambert cylindrical equal area projection



Differ just in how the world is shown

Figure 47: Albers Equal area conic projection



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

Some of the things it can do:

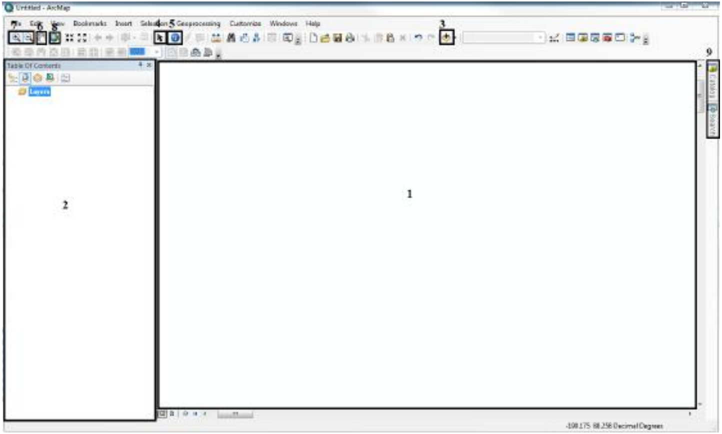
- merge and intersect shapes based on geographical relationships
- finding nearest objects according to specified metric
- calculate slope from elevation data
- take averages within polygons
- finding shortest path in a network
- make maps
- ...

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

- You will mostly use ArcMap – the software which allows you to work with geospatial data.
- There is also ArcCatalog – a type of file manager.
- As you will soon learn, doing anything in ArcMap produces a lot of garbage files at every step.
- Every shapefile, for example, usually 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 an *.shx* file (which stores a positional index of the feature geometry to allow seeking forwards and backwards quickly).
- The ArcCatalog is supposed to make managing this soup of files easier and also facilitate creation and manipulating *geo databases* and *feature datasets* (see lecture on networks).
- In practice, it is better to just automate as much as you can in python and access the ArcCatalog from within ArcMap if you absolutely have to.

Introduction to GIS

ArcMap – What the buttons mean



1. "Data frame": data is visualized here
2. "Table of contents": which datasets are loaded?
3. Load data
4. Select elements
5. Get information on geographic features
6. Pan map by dragging
7. Zoom in and out
8. Zoom to full extent (very useful)
9. Catalog (file system), Search (find geoprocessing tools)

Introduction to GIS

ArcMap – Adding data

Download some data

- go to http://www.naturalearthdata.com/downloads/10m-cultural-vectors/
- download "Admin 0 – Countries", save to some directory and unzip

Establish a "folder connection"

- before we can add data, we need to "connect" ArcMap to the folder containing the data we want to add
- On the right, click on "Catalog" (if you don't see it, click on "Windows" → "Catalog"
- In the Catalog, right click on "Folder Connections" → "Connect to Folder", find the folder in which you saved the file above, and click "Okay"
- If you connect a folder, all sub-directories will also be connected
- Connected folders stay connected until you disconnect them

Add the data

- Click , browse to the folder you just connected, and add the file *ne_10m_admin_0_countries.shp*

Introduction to GIS

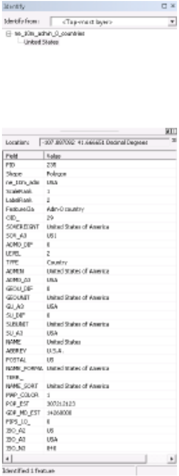
ArcMap – Inspecting the data

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

Attribute table

- right click on *ne_10m_admin_0_countries* in the table of contents → Open attribute table
- a table opens, listing a bunch of variables
- each row in the table corresponds to one polygon feature on the map
- the first two columns, "FID" and "Shape" are internal to ArcMap and are mandatory
- pick a country and right click on the small grey square at the start of its row → Select/Unselect and close the attribute table
- the country you picked is highlighted
- right click on the main window → Zoom to selected features
- click on  to zoom back out
- play around with zooming, panning, selecting, until you are comfortable

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

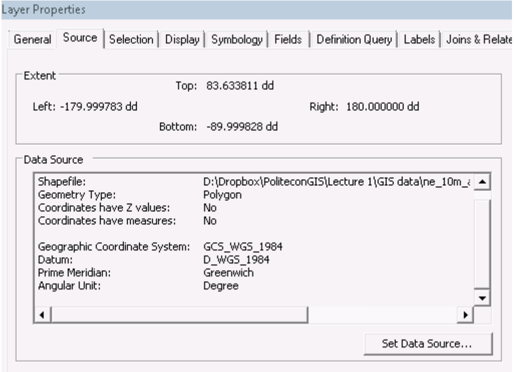


- click on and click on the US
- you should see the window on the right
- this lists all the variables in the attribute table for this particular feature
- we can select more than one feature this way (hold down the mouse and drag it over several features)
- then cycle through the attributes of the selected features in the identify window

Extent and coordinate system

- back in the table of contents, right click on “Properties” at the bottom
- select the “Source” tab

Figure 48: The source tab: extent and coordinate system



Change color of countries and outline width

- Still in the properties menu, select the "Symbology" tab
- On the left, select "Features" → Single symbol
- Click on the coloured rectangle next to "Symbol" and change fill color and outline width
- Click "Okay" and "Apply" and see how the map display changes

Label features

- Still in the properties menu, select the "Labels" tab
- Check the "Label features in this layer" box
- From the "Label Field" drop-down menu, select "ADMIN" (this is one of the variables from the attribute table)
- Click "Apply" and "Okay"

Hide a dataset

- in the Table of contents, uncheck *ne_10m_admin_0_countries*

Download some data

- go to 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
- 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 the identify tool to look up individual pixel values
- look in the source tab of the Properties menu

Change the color scheme

- Properties menu → symbology
- Select "Stretched" and change the color ramp

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 thid 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

Add data

- as before

Display XY data

- right click on the *my_points.csv* layer you just added → Display XY data
- choose "longitude" for the X field and "latitude" for the Y field
- Edit the "Coordinate System of Input Coordinates" to WGS 1984 (under Geographic Coordinate Systems → World). Note WGS 1984 may already be the default if you started with a dataset such as *ne_10m_admin_0_countries* in that coordinate system.
- dots appear on the map

Inspect the data

- look at the attribute table
- use the identify tool to get information on individual points

Model Builder

What is good for?

Making life easier

- Geoprocessing tools have many options, often need to be executed in succession.
- Clearly, we want to automate Geoprocessing somehow to make it faster.

Making research replicable

- Geoprocessing can become complex quickly.
- To make research replicable (also for ourselves!) we want to automate as much as possible of the GIS workflow.

An improvement but not the last word

- The Model Builder is a first (though not ideal) step towards automation.
- It also functions as a nice bridge towards full automation in Python (next session).

Simple example: build a model to calculate average agricultural suitability of all U.S. counties

- Download U.S. counties from `http://www.gadm.org/`.
- Download global raster data for agricultural suitability from `https://nelson.wisc.edu/sage/data-and-models/atlas/data.php?incdataset=Suitability%20for%20Agriculture`.

Model Builder

Making a simple model

Opening the model builder and adding tools

- Click on to open the Model Builder.
- To add Geoprocessing tools, find them in the Search and drag them to the Model Builder window.

Building the model for our simple example

- Copy the suitability raster, define projection to WGS 1984, resample it to a cell-size of 0.008333 using NEAREST, copy the US county features, delete fields from the copied US counties shapefile so that only state and county names remain, create a unique numeric county ID, execute Zonal Statistics to table (mean), export Zonal Statistics and attribute table of US counties to *.csv*.
- Note how using outputs from the last tool can be used as inputs for new tools.
- Use and to organize the model nicely, to verify that it is working properly and to run it.
- The model may crash at Zonal Statistics. If it does, close and restart Arc, and run the rest of the model. (One more reason for Python!)

Inspect the output

- Note that there are some counties that are missing suitability information.
- Can we track down the problem?

Model Builder

Saving models, editing models, and exporting to Python

Saving models

- Models created with the model builder must be saved inside things called "Toolboxes".
- In the model builder, click Model → Save As... →  → Choose a name for the toolbox → Double click the newly created toolbox → save the model.

Editing models

- To open an existing model, use the Catalog (on the right, above the Search) to navigate to the location of the toolbox containing it. Click on the  → right click the model you want to edit → Edit...

Exporting to Python

- Once the Model is done, in the Model Builder, click Model → Export → To Python script...
- This produces a bare-bones, (not very readable) version of a Python script for the model.
- We will make passable code out of this in the next session.

Producing the LHS in the regressions of Storeygard (2016)

- Want panel data of city-level output. Doesn't exist for Africa.
- $\Rightarrow$ Storeygard uses city-level lights as the dependent variable in his regressions.
- How can we use the lights data to get a city-level output proxy?
- Idea: Use maximum extent of contiguous visible lights over all years as "envelope" for the largest area occupied by a city during the sample period.
- Use actual city coordinates to only include envelopes corresponding to "true" cities.
- Sum all lights inside an envelope for a given year to obtain city-level lights.

Optional geoprocessing exercise:

Cities from lights (Storeygard 2016)

Inputs

- Lights data
  https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html
- Gas flare shapefiles
  https://ngdc.noaa.gov/eog/interest/gas_flares_countries_shapefiles.html
- Populated places
  http://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-populated-places/
- Political map of Africa
  https://www.arcgis.com/home/item.html?id=16da193d9bcd4ae0b74febe39730658a

Optional geoprocessing exercise:

Cities from lights (Storeygard 2016)

Detailed workflow: part 1

- Once downloaded, copy the lights raster data to *.tif* (this will save space), use 1995 and 2010 as years. For a challenge, do all years, with a loop!
- Prepare for clipping (*Dissolve* countries to get one feature per country, *Union* to create one gas-flare shapefile). Disregard flare shapefiles for Cote d'Ivoire, Ghana, and Mauritania.
- *Delete Field* to remove unnecessary attribute fields from shapefiles.
- *Polygon to raster* to turn gas flares into raster.
- *Create Fishnet* to create a clipping rectangle of the full lights raster extent.
- *Clip (Data Management)* to extend the Raster size using the fishnet.
- *Reclassify* gas flares raster to -99/0.
- *Create Constant raster* value of zero.
- *Plus* Add constant and reclassified raster.
- *Reclassify* the result to 0/NoData.
- *Plus* Add the result to the 1995 and 2010 lights raster files → gas flares now are "NoData".
- *Clip* The 1995 and 2010 rasters without the gas flares to a rough rectangle around Africa to speed up later calculations.

Optional geoprocessing exercise:

Cities from lights (Storeygard 2016)

Detailed workflow: part 2

- *Plus* Add the 1995 and 2010 lights raster files.
- *Reclassify* the result to 1/NoData.
- *Raster to Polygon* to get city "envelope" polygons.
- *Union* the envelopes with the dissolved country shapefile.
- *Select* two countries from resulting intersection. We will choose Tanzania and Kenya. Do this twice, producing two separate envelope shapefiles. To deal with coast, select missing, then *Merge*, *Dissolve*, and *Select*.
- *Multi Part to Single Part* the result.
- *Spatial Join* cities points to the envelopes.
- *Select* only the envelopes that have at least one city inside them.
- *Delete Fields* to do some intermediate clean-up.
- *Add Field* + *Calculate field* to create unique city-ID.
- *Zonal Statistics as Table* with lights and population data.
- *Table to Table* to output to *.csv*
- Merge the result with Stata (or your favorite tool).

Figure 49: The world in Mercator projection

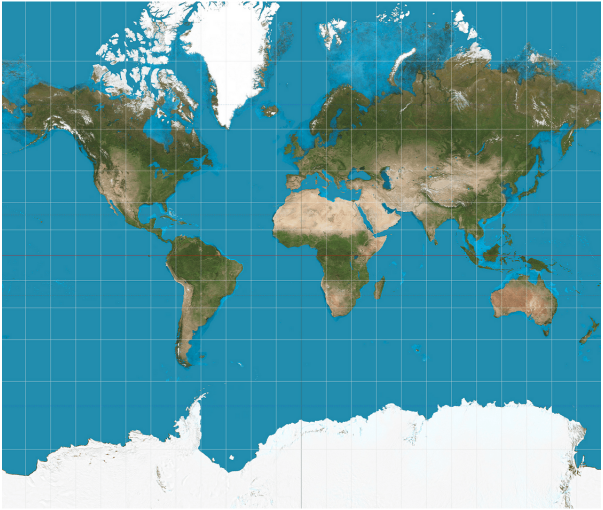


Figure 50: The actual sizes of Greenland and Australia

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`