Using Spatial Data in Applied Economic Research

Paul A. Raschky,

University of St Gallen, October 2017

Contact details, questions and feedback

1. Lecturer: Paul Raschky
   - Office: Department of Economics,
   - email: paul.raschky@monash.edu
2. Office hours: by appointment
3. Live stream/chat consultation

Course Goals

- Provide an introduction to the use of geo-spatial data and analysis tools in applied economics.
- This course contains a mixture of general concepts of GIS and spatial analysis and practical exercises.
- It will equip economics students with an understanding of the most commonly used GIS and spatial analysis tools in applied economics.
- However, due to the nature of the course, it will NOT provide students with an in-depth understanding of all GIS and spatial analysis tools.

Course Organization

- This course is a block course that is taught in six 4-hour units.
- The course alternates between lectures and lab-practicums with in-class exercises.

Assessment

- Submit in-class exercises
- Final project ('term paper')

Unit Overview

Date/Time	Lecture	Lab
16.10.17 14:15-18:00	Introduction to GIS	Working with QGIS
17.10.17 08:15-12:00	Vector data and tools	Vector tools
18.10.17 14:15-18:00	Spatial analysis and autoregressive models in Stata	Zonal Statistics
20.10.17 14:15-18:00	Buffers and Distance	Buffers and Distance
23.10.17 14:15-18:00	Spatial Discontinuity	Spatial Discontinuity
24.10.17 08:15-12:00	Map algebra	Map Algebra

Today's Lecture

1. Introduction to GIS
2. Types of GIS data
3. Projections and Coordinate Reference Systems

1. What is GIS?

- Geographic Information Systems (and *Science*) - GIS
- A GIS is a computer-based system that allows you to ...
  - collect
  - store
  - manipulate
  - analyze
  - display
- ... spatial (locationally defined) data.

1. What is GIS?

1. Collect data - Manually

2. Store data

osmid	code	class	name	lon	lat
445368973	2303	cafe	Schwyter Uni-Beck	9.37827	47.43283

Image sources: Google Street View, garmin.com

1. Introduction - GIS and Empirical Economics

3. Visualize the data

Data sources: Open Street Map (geofabrik.de). Visualization by the author

1. What is GIS?

Collect and store data automatically: Atlas buoys and Sea Surface Temperature

buoy	Date	Temp	lon	lat
B0879	03/07/16	26.75 C	45.89	7.33

Image sources: http://www.jcommops.org/

1. What is GIS?

Manipulate (Spatial interpolation) and visualize:

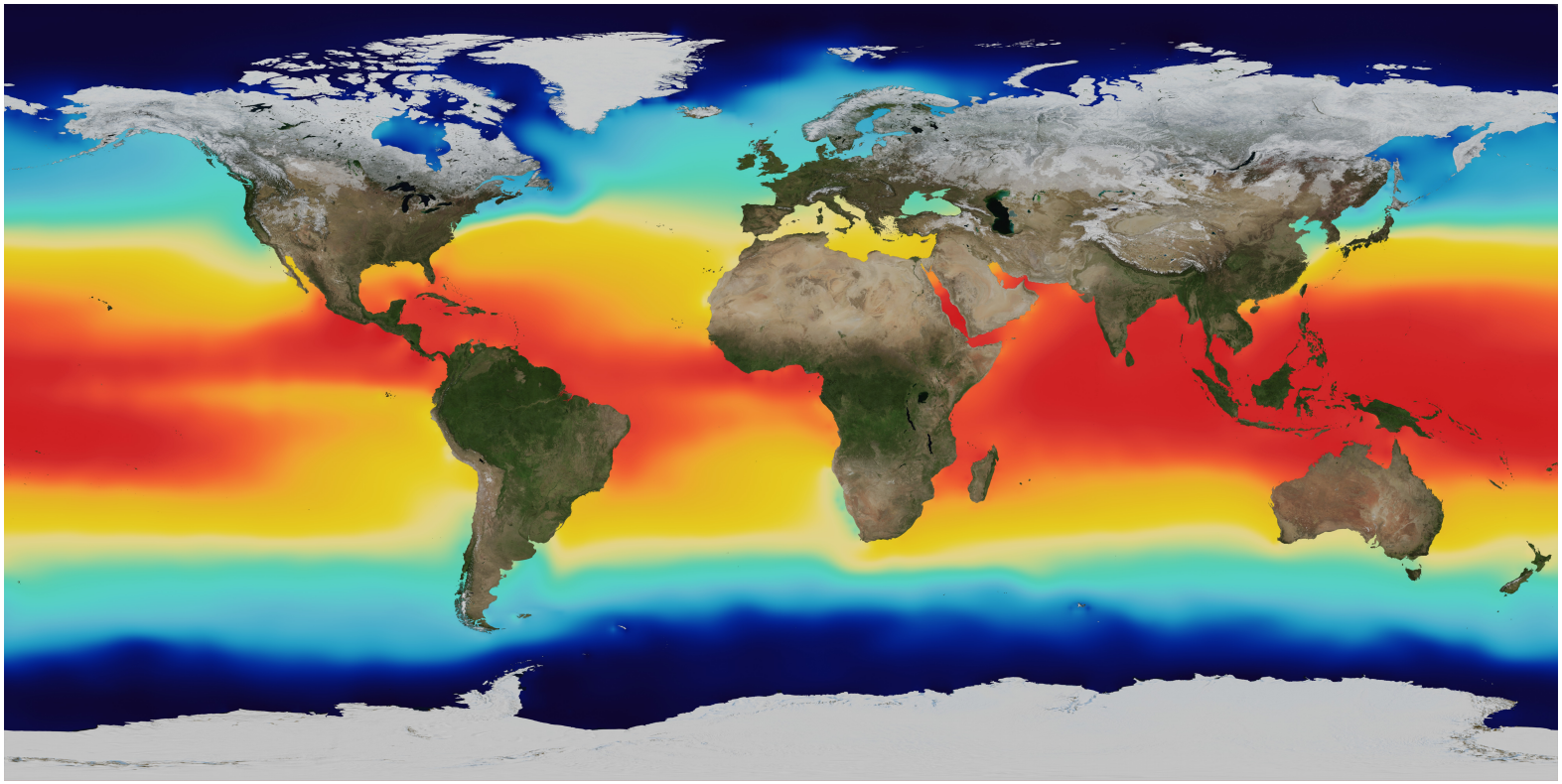


Image sources: http://www.jcommops.org/

1.1. GIS Software

ArcGIS (ArcMap)

- Widely used but requires a license.
- Windows only.
- Very accessible BUT has some quirks and bugs
- Supports python (arcpy)

QGIS

- Open Source (Free!)
- Works on all operating systems.
- Very accessible and fewer quirks and bugs
- Supports GRASS and python (arcpy) (ArcGIS is still more python friendly)

Consider python $\rightarrow$ replicability.

1.2. Statistical/Math Software with GIS tools

Stata

- A number of commands that allow you to import, analyze and visualize GIS data.
- Stata 15 includes a full suite of spatial commands (see Cameron & Trivedi 2017)

R

- rsptail.org, http://spatial-analyst.net/ (wiki)
- Tutorials: http://pakillo.github.io/R-GIS-tutorial/, http://spatial.ly/r/

MATLAB

- Mapping toolbox
- Spatial econometrics toolbox (LeSage 2009)

1.3. Other GIS software

- GRASS (open source)
- GeoServer (web map server)
- PostGIS (geospatial SQL queries)
- Geopanda (Python extension)

1.4. GIS and Empirical Economics

Why is GIS useful for Economists?

- Source of data
- Source of variation
- Statistical identification
- Toblers first law of geography:

"Everything is related to everything else, but near things are more related than distant things."
Tobler (1970)

1.5. GIS as a Source of Data

Locating events in space:

- Conflicts - ACLED, UDCP
- Development Aid projects - aiddata.org
- Global Database of Events, Language, and Tone - GDELT

1.5. GIS as a Source of Data

Passively collected data :

- Defense Meteorological Satellite Program - Operational Linescan System (DMSP-OLS) Nighttime Lights
- Elevation - DEM models (i.e. GTOPO)
- Landuse (MODIS)
- Temperature
- Air pollution

1.5. GIS as a Source of Data

Passively collected data: DMSP-OLS Nighttime Lights

Data sources: NOAA.

1.6. GIS as a Source of Variation

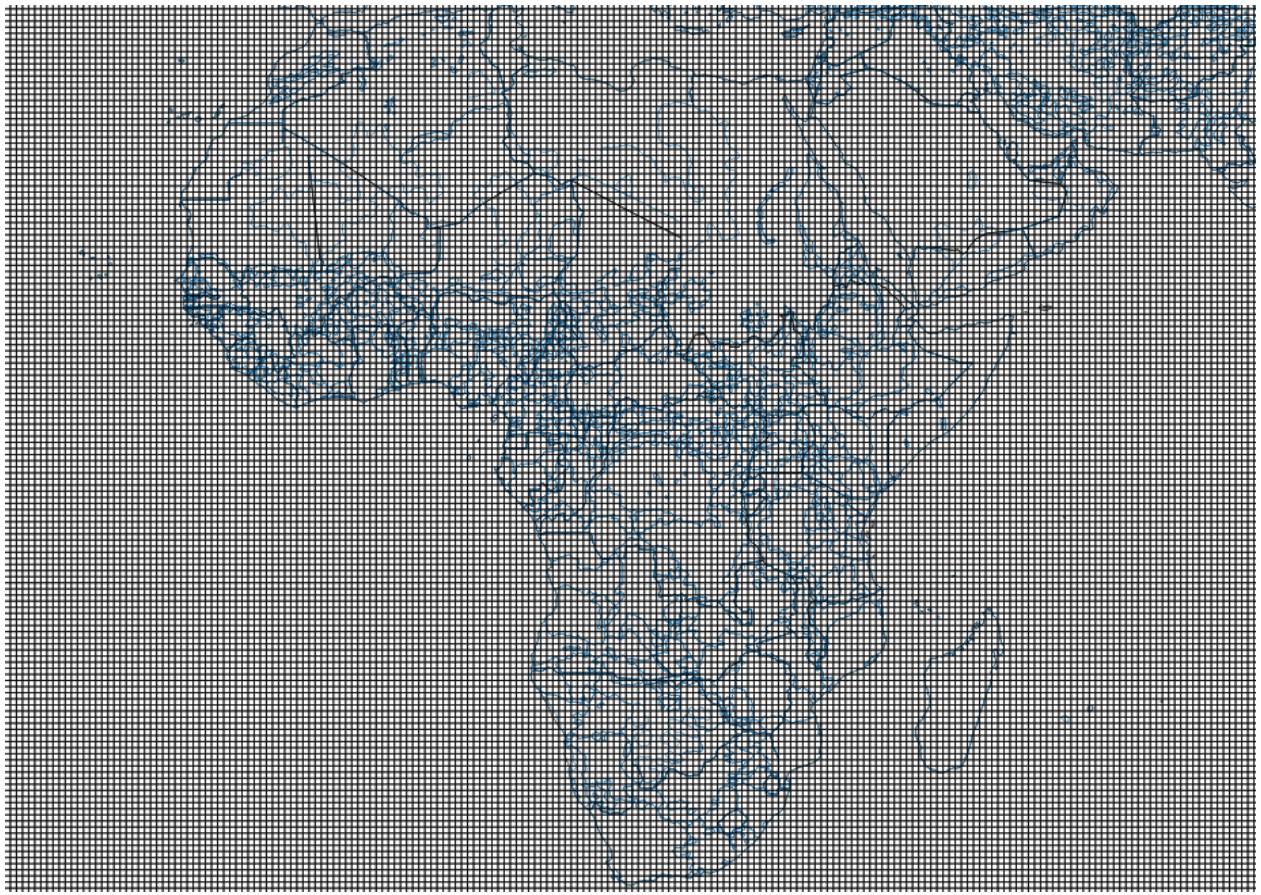
Unit of observation: Country

1.6. GIS as a Source of Variation

Unit of observation: Ethnic homelands (GREG)

1.6. GIS as a Source of Variation

Unit of observation: Grid cells (PRIO)



1.7. Using GIS for Identification

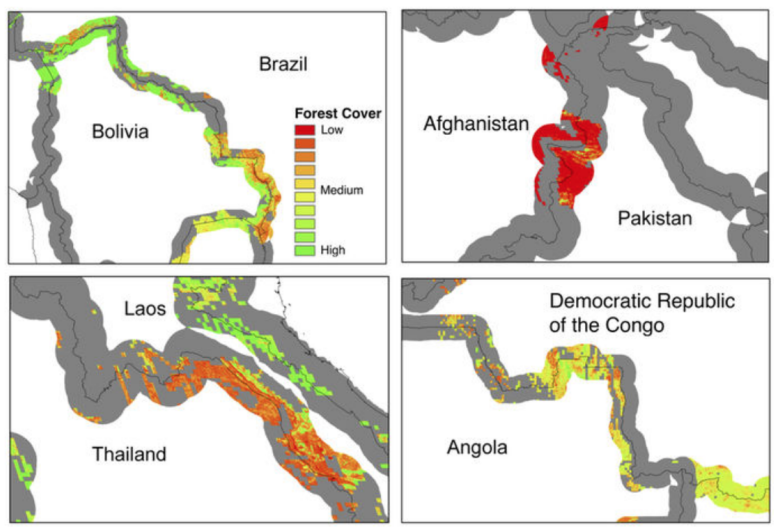
Exploiting Distances

- The Long-term Effects of Africa's Slave Trades (Nunn 2008)
- Protestant Economic History (Becker & Woessman 2009)
- Economic Development and the Global Network of Air Links (Campante & Yanagizawa-Drott forthcoming)

1.7. Using GIS for Identification

Exploiting Spatial Discontinuities

- The Persistence of Institutions - Dell (2010)
- National vs. Subnational Institutions - Michalopoulos & Papaioannou (2013)
- Economic Development and Deforestation (Crespo et al. 2017)

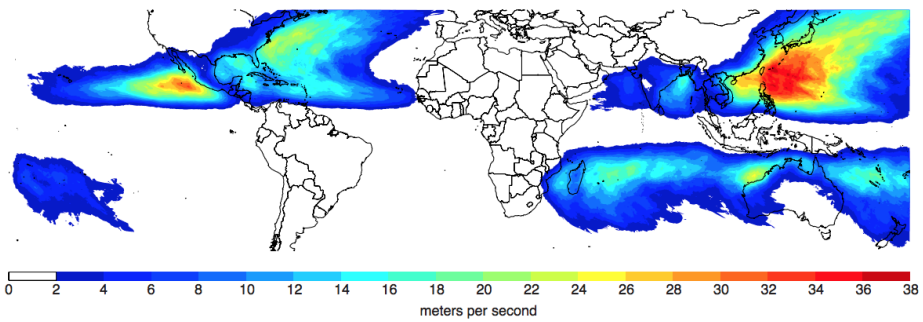


Source: Crespo et al. 2017

1.7. Using GIS for Identification

Natural Experiments

- Lightning Strikes and Mobile Phone Coverage (Manacorda & Tesei 2016)
- Cyclones and Economic Growth (Hsiang & Jina 2017)



Source: Hsiang & Jina 2017

2. GIS Data

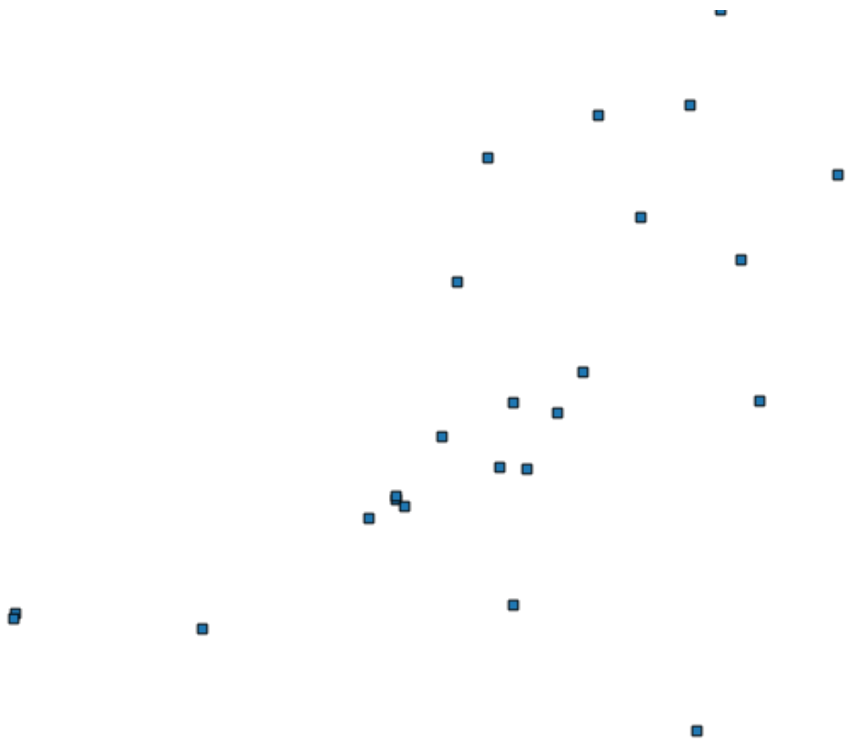
There are three main types of spatial data:

1. Vector data
2. Raster data

- Tabular data can also contain spatial information (lat/lon).
- Each type of data has its own (file) format and is edited in different ways.
- Most GIS software have different tools for vector and raster data.
- You can use GIS software to combine information from different types of spatial data.

2.1 Vector Data

Points: Post offices and post boxes in St Gallen



2.1 Vector Data

Polylines: Roads in St Gallen

2.1 Vector Data

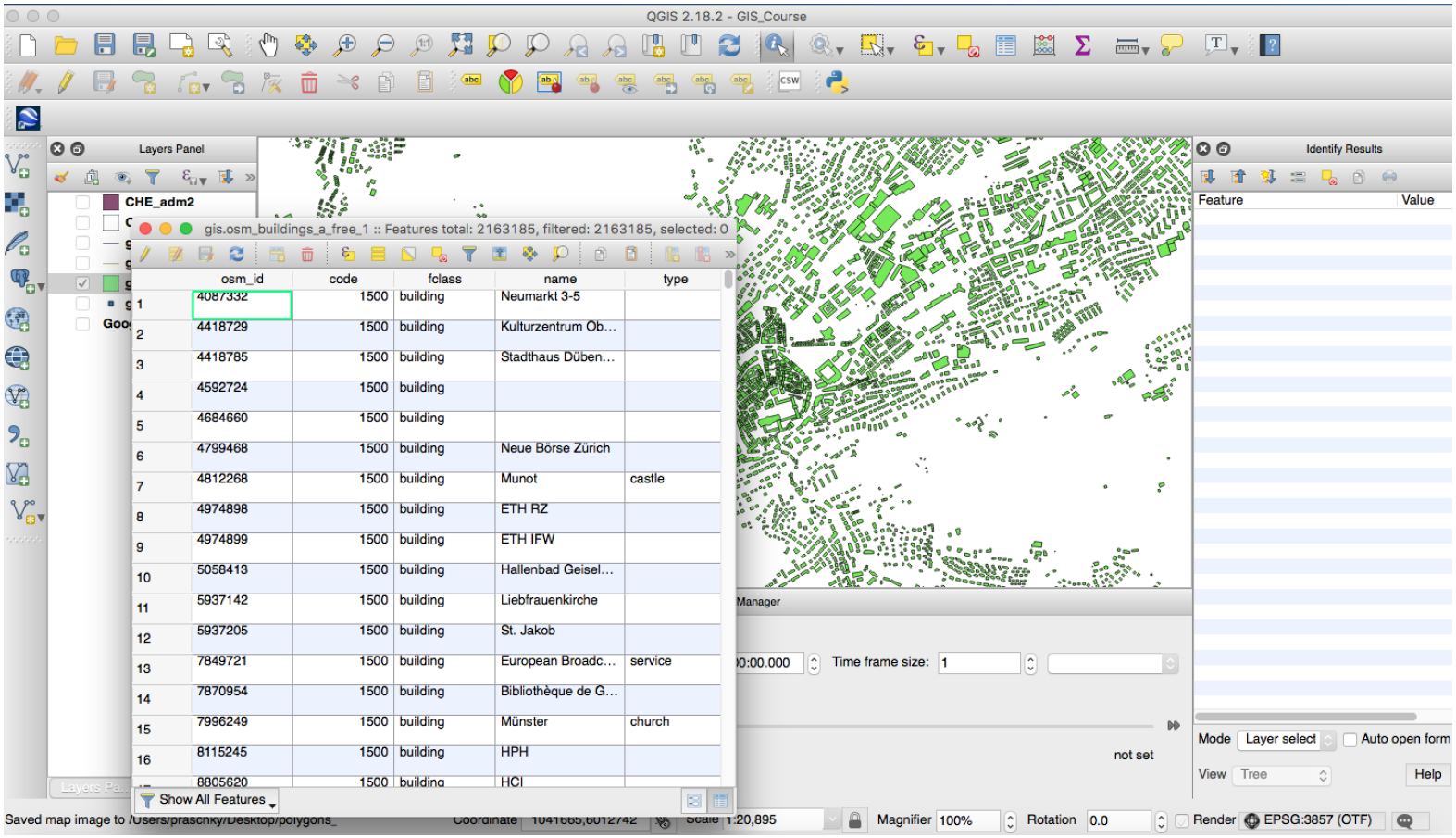
Polygones: Buildings in St Gallen

2.1 Vector Data

- Each spatial unit in a vector dataset is called feature. A set of features of the same type is called feature class.
- Vector data is stored in shapefiles. These shapefiles consist of several the following extensions:
  - **.shp, .dbf, .shx**, .prj, .cst
- Always keep the files in the same folder.
- Alternative formats are .gpx, .kml, kmz, .geojson, .osm, .bz2

2.1 Vector Data

Attribute Table



	osm_id	code	fclass	name	type
1	4087332	1500	building	Neumarkt 3-5	
2	4418729	1500	building	Kulturzentrum Ob…	
3	4418785	1500	building	Stadthaus Düben…	
4	4592724	1500	building		
5	4684660	1500	building		
6	4799468	1500	building	Neue Börse Zürich	
7	4812268	1500	building	Munot	castle
8	4974898	1500	building	ETH RZ	
9	4974899	1500	building	ETH IFW	
10	5058413	1500	building	Hallenbad Geisel…	
11	5937142	1500	building	Liebfrauenkirche	
12	5937205	1500	building	St. Jakob	
13	7849721	1500	building	European Broadc…	service
14	7870954	1500	building	Bibliothèque de G…	
15	7996249	1500	building	Münster	church
16	8115245	1500	building	HPH	
	8805620	1500	building	HCI	

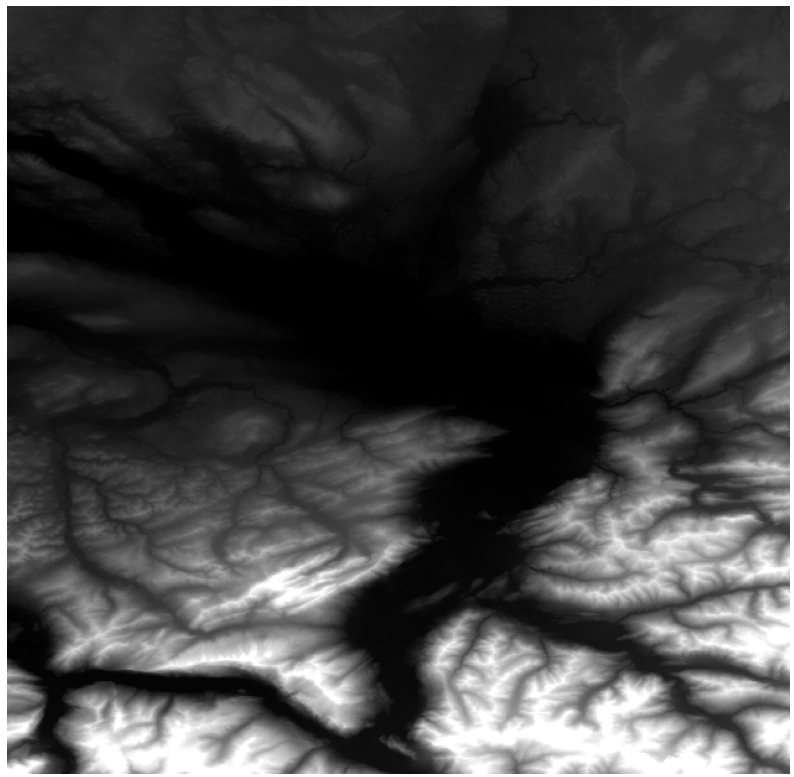
- The attribute table is stored in the .dbf file.
- This information can easily be exported to a .csv for further use in statistical software.

2.2 Raster Data

- Rasters are made up of a matrix of pixels/cells.
- Every raster layer in a GIS has pixels (cells) of a fixed size that determine its **spatial resolution.**
  - 100×100 km; 0.5 × 0.5 degrees.
- When the raster data is displayed on a screen or printed out, the red, green and blue (RGB) information is combined to show you an image that your eyes can interpret.
- Raster data can also include non-visible information (infra-red light).
- The number of bands in a raster image is referred to as its **spectral resolution.**
- File formats: (geo)TIFF (.tif), ascii, esri grid.

2.2 Raster Data

Digital Elevation Model (DEM) - St Gallen +

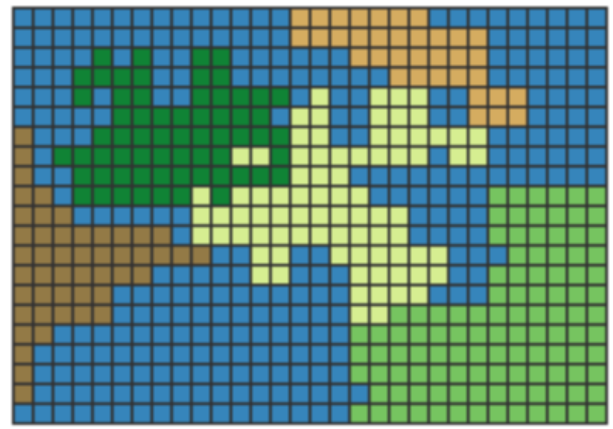


Data sources: https://github.com/lukasmartinelli/swissdem

2.2 Raster Data

Vector vs. Raster data

Polygon features

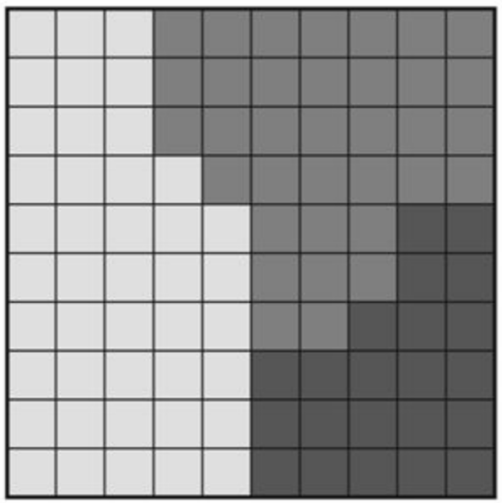


Raster polygon features

Image sources: ESRI

2.2 Raster Data

Raster data structure



Entity model

1	1	1	2	2	2	2	2	2	2
1	1	1	2	2	2	2	2	2	2
1	1	1	2	2	2	2	2	2	2
1	1	1	1	2	2	2	2	2	2
1	1	1	1	1	2	2	2	3	3
1	1	1	1	1	2	2	2	3	3
1	1	1	1	1	2	2	3	3	3
1	1	1	1	1	3	3	3	3	3
1	1	1	1	1	3	3	3	3	3
1	1	1	1	1	3	3	3	3	3

Cell values

```
10,10,3
1,1,1,2,2,2,2,2,2,2
1,1,1,2,2,2,2,2,2,2
1,1,1,2,2,2,2,2,2,2
1,1,1,1,2,2,2,2,2,2
1,1,1,1,1,2,2,2,3,3
1,1,1,1,1,2,2,2,3,3
1,1,1,1,1,2,2,3,3,3
1,1,1,1,1,3,3,3,3,3
1,1,1,1,1,3,3,3,3,3
1,1,1,1,1,3,3,3,3,3
```

File structure

Image sources: Heywood et al. 2003

2.2 Raster Data

- One problem with raster data is their (relative) size.
- A value must be recorded and stored for each cell.
- Building pyramids/overviews (QGIS)

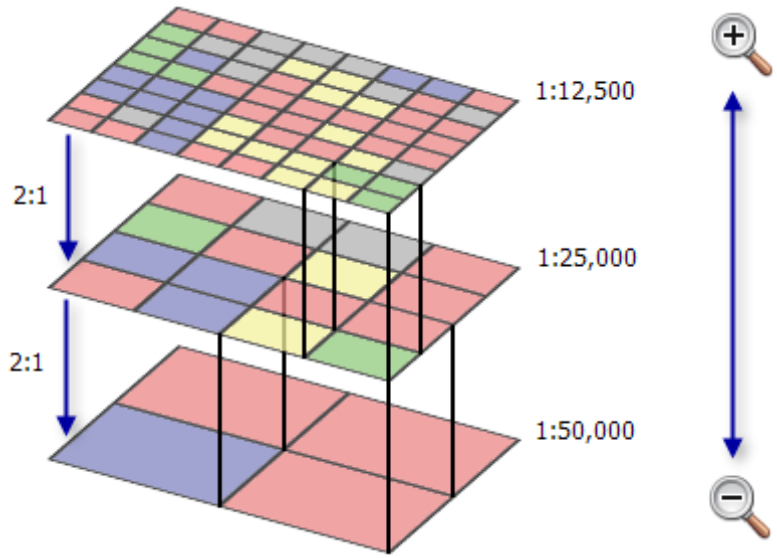


Image sources: ESRI

2.3 Tabular Data

Address	City	State	ZIP Code
1134 Massachusetts Ave	Cambridge	MA	02138
290 Main St	Cambridge	MA	02142
47 Mount Auburn St	Cambridge	MA	02138
428 Massachusetts Ave	Cambridge	MA	02139
314 3rd St	Cambridge	MA	02142
675 W Kendall St	Cambridge	MA	02142
746 Massachusetts Ave	Cambridge	MA	02139
247 Cambridge St	Cambridge	MA	02141
276 Broadway	Cambridge	MA	02139
2370 Massachusetts Ave	Cambridge	MA	02140
1687 Massachusetts Ave	Cambridge	MA	02138
1722 Massachusetts Ave	Cambridge	MA	02138

Latitude	Longitude
42.370882	-71.114246
42.362318	-71.086001
42.371406	-71.116643
42.363327	-71.100944
42.363039	-71.08254
42.364681	-71.082363
42.366312	-71.105351
42.371092	-71.080076
42.367804	-71.097311

Geo_FIPS	Total Population	Male	Female
1001020100	1900	944	956
1001020200	2342	1157	1185
1001020300	3297	1451	1846
1001020400	4272	2056	2216
1001020500	10881	5202	5679
1001020600	3782	1769	2013
1001020700	2799	1140	1659
1001020801	3096	1475	1621
1001020802	10471	5283	5188
1001020900	5637	2908	2729

2.3. Tabular Data

- XY data
- Tabular data is normally stored in .csv, .txt, .xls, .tab etc.
- Tabluar data can be joined to vector data if there is a common ID.
- Tabular data can be added as a new layer if it contains X/Y (lon/lat) information
  - It still needs to be georeferenced - converted to a shapefile for further use in GIS.

2.4. GIS Data Collection

- Primary:
  - Digital remote sensing (nighttime Lights)
  - Digital aerial photographs
  - GPS measurements (OSM)
- Secondary:
  - Data collected for other purposes converted for the use in GIS.
  - Scanning of aerial photos
  - Digitization of maps

2.4. GIS Data Collection

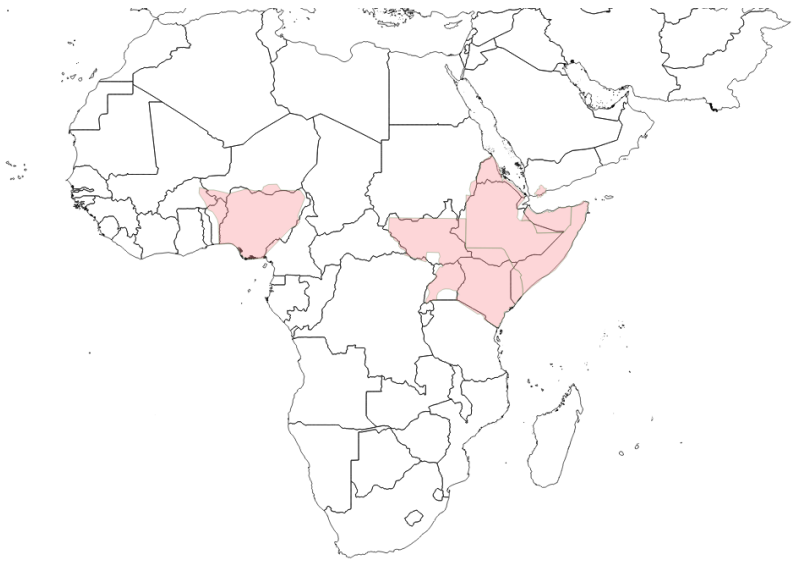


Image sources: Raschky & Phan (2015)

2.4. GIS Data Collection

- Data is collected, digitized, and/or geo-referenced by humans $\rightarrow$ human error.
- Automated data collection can also suffer from measurement errors.

3. Projections and Coordinate Reference Systems

- **Map projections** try to portray the surface of the earth or a portion of the earth on a flat piece of paper or computer screen.
- A **coordinate reference system (CRS)** then defines, with the help of coordinates, how the two-dimensional, projected map in your GIS is related to real places on the earth.
- The decision as to which map projection and coordinate reference system to use, depends on the regional extent of the area you want to work in, on the analysis you want to do and often on the availability of data.

3.1. Map Projections

a) cylindrical b) conic c) planar/azimuthal projections

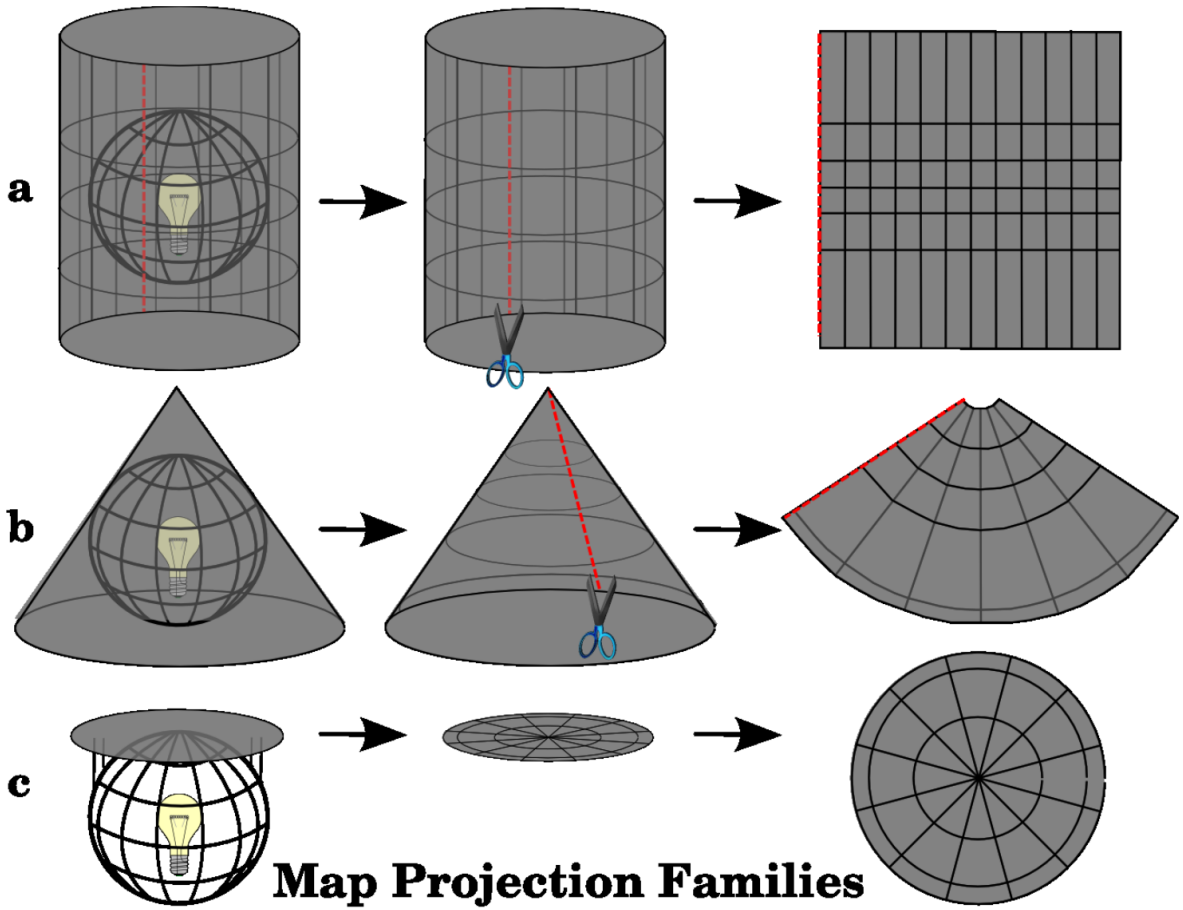


Image sources: Sutton et al. 2009

3.1. Map Projections

- Map projections are never absolutely accurate representations of the spherical earth.
- As a result of the map projection process, every map shows distortions of angular conformity, distance and area.
- It is usually impossible to preserve all characteristics at the same time in a map projection.
- Use a map projection that provides the best characteristics for your analyses.

3.1. Map Projections

Map projections with angular conformity

- The main directions of the compass rose (North, East, South and West) will always occur at 90 degrees to one another.
- Maintaining correct angular properties can be preserved on a map projection as well.
- A map projection that retains this property of angular conformity is called a **conformal** or orthomorphic projection.
- Commonly used for navigational or meteorological tasks.
- Examples:
  - Mercator projection
  - Lambert Conformal Conic projection

3.1. Map Projections

Map projections with equal distance

- projecting a map is to accurately measure distances, you should select a projection that is designed to preserve distances well. Such projections, called equidistant projections, require that the scale of the map is kept constant.
- Examples:
  - Mercator projection
  - Lambert Conformal Conic projection
  - Azimuthal Equidistant projection

3.1. Map Projections

Map projections with equal area

- When a map portrays areas over the entire map, so that all mapped areas have the same proportional relationship to the areas on the Earth that they represent, the map is an equal area map.
- Equal area projections are best used when calculations of area are the dominant calculations.
- Examples:
  - World Eckert VI
  - Lamberts Equal Area
  - Mollweide Equal Area

3.2. Coordinate Reference Systems

- The coordinate reference systems (CRS) allows you to specify every location on the earth by a set of coordinates.
- There are two types of CRS:
  1. Geographic
  2. Projected

3.2.1. Geographic Coordinate Systems

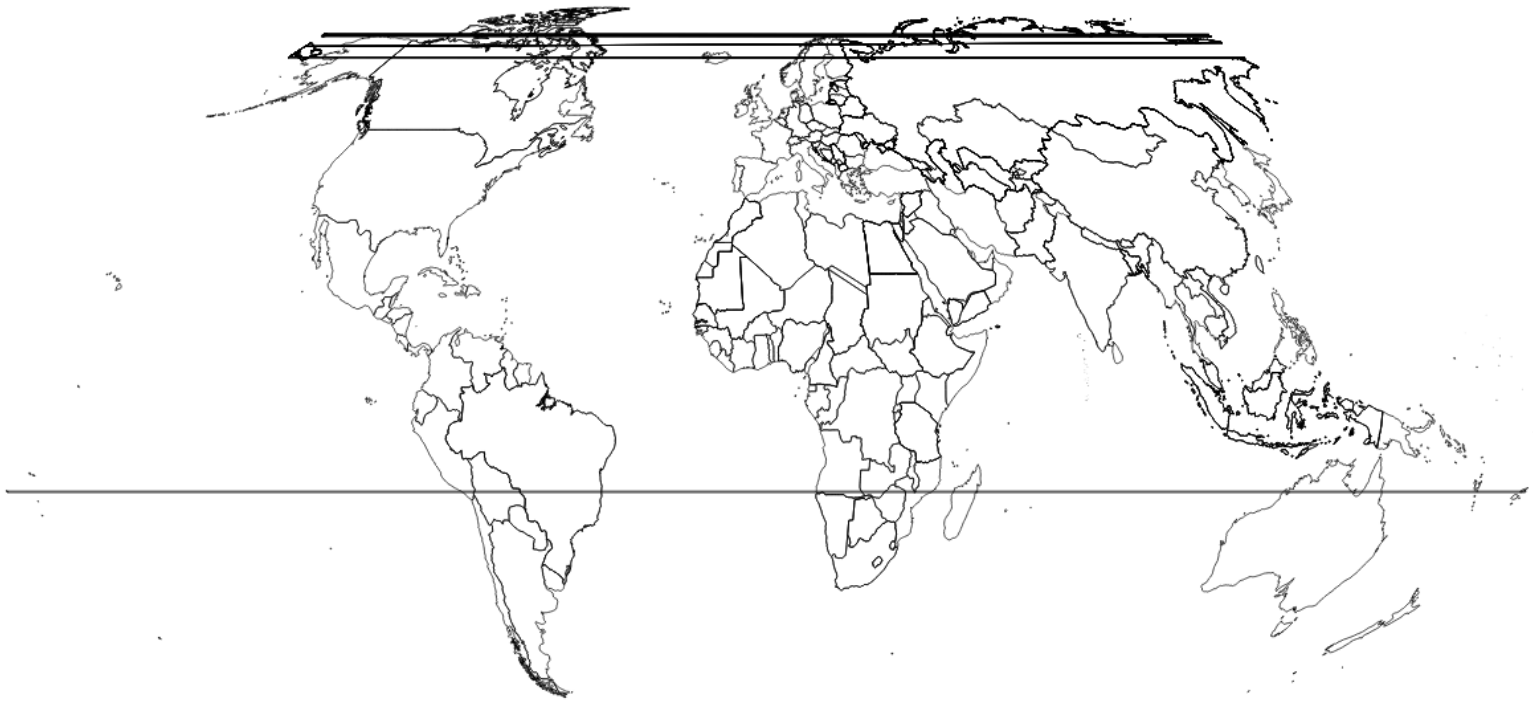
- Geographic coordinate systems are based on a 3D spherical surface and use angular units (degrees, minutes seconds.) to describe a location on the earths surface.
- A projected coordinate system (PCS) project the spheroid on a 2D surface and use linear units (meters)

3.1. Geographic Coordinate Systems

WGS 1984

3.1. Projected Coordinate Systems

World Eckert VI



3.3. Projections and CRS - Some Final Notes 1

- Projections and CRS are complex and difficult.
- IMPORTANT: Ensure that all layers are in the same coordinate system before performing calculations that involve multiple layers.
- Usually, the GIS files that you access already contain projected data (.prj file).
- Always report the CRS and projection that you have used.
- CRS often used in Econ applications:
  - Distance between points - WGS 1984
  - Length of polylines (i.e. travel distance on roads) - Universal Transverse Mercator (UTM)
  - Surface area - World Eckert VI

3.3. Projections and CRS - Some Final Notes 2

ARC GIS and QGIS have 'On the fly' reprojection.

- On the fly does not actually change the data, it just reprojects the layers as they appear on the map.
- To truly reproject datasets with different projections you need to (in QGIS):
  1. Save the respective layer as a new file
  2. Select new CRS