

Introduction to GIS

Raphaëlle ROFFO

Sciences Po - GLM Masters
Spring Semester 2022

Session 1

Course overview
&
What is GIS ?

Today's plan

1. Logistics!
 - a. Technical setup
 - b. Attendance
 - c. Student rep election
 - d. Format of the course
 - e. Coursework
2. What's GIS / Common use cases
3. GIS and geospatial data science workflows
4. Homework (Tutorial 1)

Attendance



Student representative election



Technical setup



Technical setup

- Slides and tutorials are available on [this website](#). The sessions are recorded and will be shared on [this Google Drive](#) after each session.
- Tutorial 1 will walk you through downloading QGIS and exploring the interface.
- Room H005 is a **computer room** and is booked for you. It may be easier for you to make use of it while you work on the tutorials (display slides on the computers while doing the tutorial on your laptop, or vice versa).

Technical setup

We use **Slack** for offline discussions and to encourage you to help each other (best way to test your own understanding of the topic is to explain it to someone else!). [Join here](#) and explore the 4 channels:

- **#general** for general announcements
- **#resources** for sharing useful resources you've found (datasets, online tutorials, books...)
- **#help** for asking questions and helping each other
- **#random** for anything that doesn't fit in the 3 first channels

Slack is also the best place to contact me (better than emails)

Sessions

The course will be run in 6 sessions:

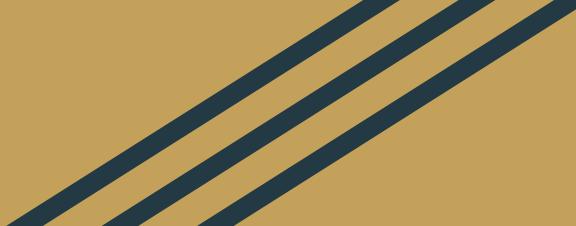
1. What is GIS?
2. Sourcing and loading data into GIS
3. Working with vector data: the attribute table
4. Cartographic design
5. Working with vector data: geoprocessing
6. Wrap up and coursework

Sessions format

For online sessions, we will limit the session to 1 hour on Zoom, leaving 1 hour of autonomous work during which you can go over the tutorials by yourselves.

Zoom sessions are designed to:

1. Introduce key concepts and theory
2. Answer any question you have about the tutorial



Coursework & Participation



Final Coursework Overview

Deadline: Monday 4th April, 23.59 Paris time.

If late: -1 point penalty for each day past the deadline.

100% Final coursework:

Working in groups of 2 or 3, students will be provided a dataset to explore and will be tasked with carrying out simple geospatial analysis and visualisation. They will produce a technical report detailing the methodology they adopted and the insights they can draw from this data.

(we will walk through the full guidelines on session 4)

Final Coursework

The final coursework is a **map production exercise**. Find a research question, carry out a simple map production workflow, and write a report summarizing your findings. You will be assessed on your capacity to:

- Frame your research question in a clear and concise manner, and ensure a few maps can provide interesting insights,
- Identify datasets that are relevant to answering your question (*technically here you may want to work backwards and use the data you already have to find your research question*),
- When appropriate, use table joins to “enrich” your vector data,
- Carefully choose your symbology, and ensure your map is accessible and colour-blind safe,
- Create map exports complete with all key cartographic elements (title, legend, data source etc).
- Analyse and interpret the patterns that emerge from your maps, explaining what this might mean in terms of policy or research outlooks.

Final Coursework

Some practical points:

- Deadline: Monday 24th April, 23.59 Paris time. *If you're late, minus 1 point for each day behind the deadline.*
- Work in groups of 2-3 students
- Work on a **European city** of your choice (*cities in the UK and other non-EU european countries are ok*)
- The report must be 3 pages minimum, 5 pages *maximum*
- You are encouraged to use the data provided for the tutorials, however if you want to challenge yourselves, you will get **+1 point bonus for working with data you have sourced yourselves.**

Final coursework: Proposed outline

You are strongly encouraged to follow this outline:

1. Executive summary (maximum ½ page, bullet points are fine)
2. Introduction / Problem / Context
3. Data sources in a table
4. High-level methodology. Keep it very short but use precise terminology. Explain your symbology choices including your choice of class breaks if you built a choropleth.
5. A minimum of 2 maps. Careful, you only have 5 pages maximum in this report so these maps must be relevant to answering your policy question
6. Analysis of the findings (half a page)
7. Conclusion / next steps for policy makers or future research outlook (half a page)

In this exercise, concision and precision are key!

Participation

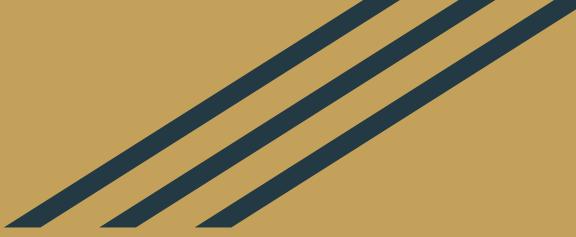
Each session will start with a 10 minute discussion on the tutorial. This shouldn't take you too much time to prepare! It is designed to strengthen your engagement with the material.

We will spend **roughly 10 minutes on this**, and 4 students will speak each time so it's about 2 minutes per student. During those 2 minutes, you can talk about:

- What you've understood about the material
- What would be possible use cases for the tools / techniques introduced
- Ask questions about the things you find confusing

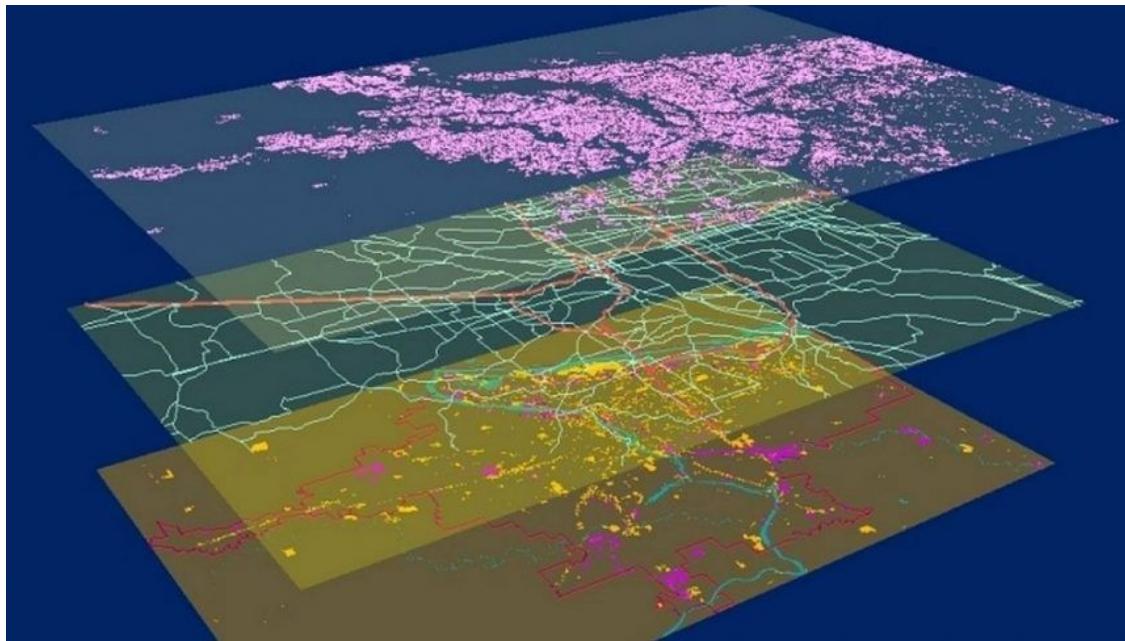
Do **not** prepare a formal presentation, and there is **no need** to prepare collectively.

This will be followed by additional Q&A to answer any question you may have.

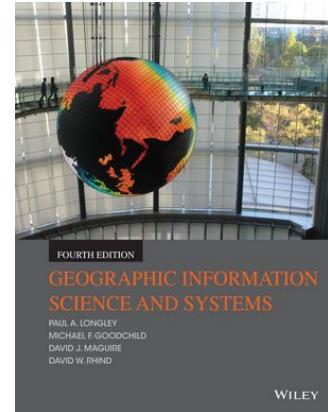


What's GIS?

Mixing geography, maths, statistics, information science, geoinformatics, computer science...



“The science of where”



“Geographic Information Systems are computer-based tools that analyze, store, manipulate and visualize geographic information, usually in a map.”

Michael Goodchild

Spatial is special



Spatial is special...

1. ... because the Earth isn't flat

and we've come up with various ways of encoding location:

- ID reference system (LSOA, MSOA, etc)
- Linear reference system (Postal addresses)
- Geographic coordinate systems: (Lat/Long from GPS coordinates)
- Projected coordinate systems: Eastings/Northings references that are centered on a specific region of the globe to avoid distortion.

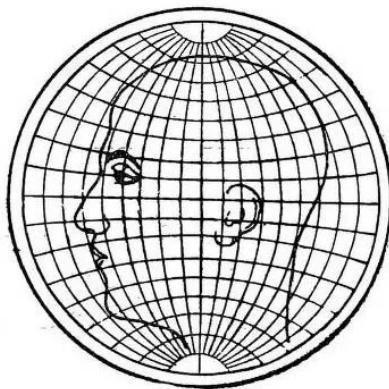


FIG. 42.—Man's head drawn on globular projection.

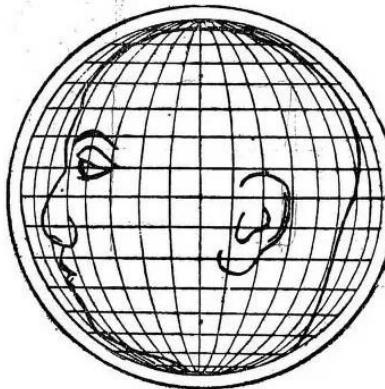


FIG. 43.—Man's head plotted on orthographic projection.

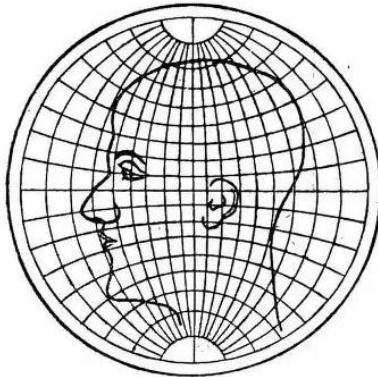


FIG. 44.—Man's head plotted on stereographic projection.

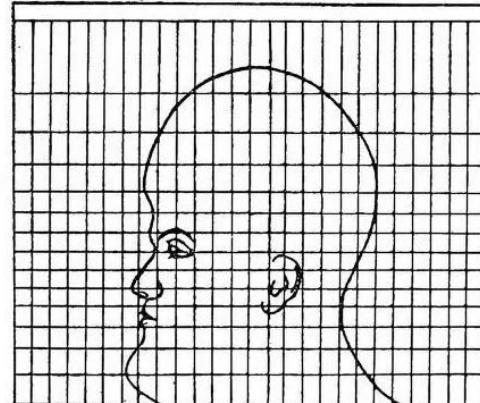
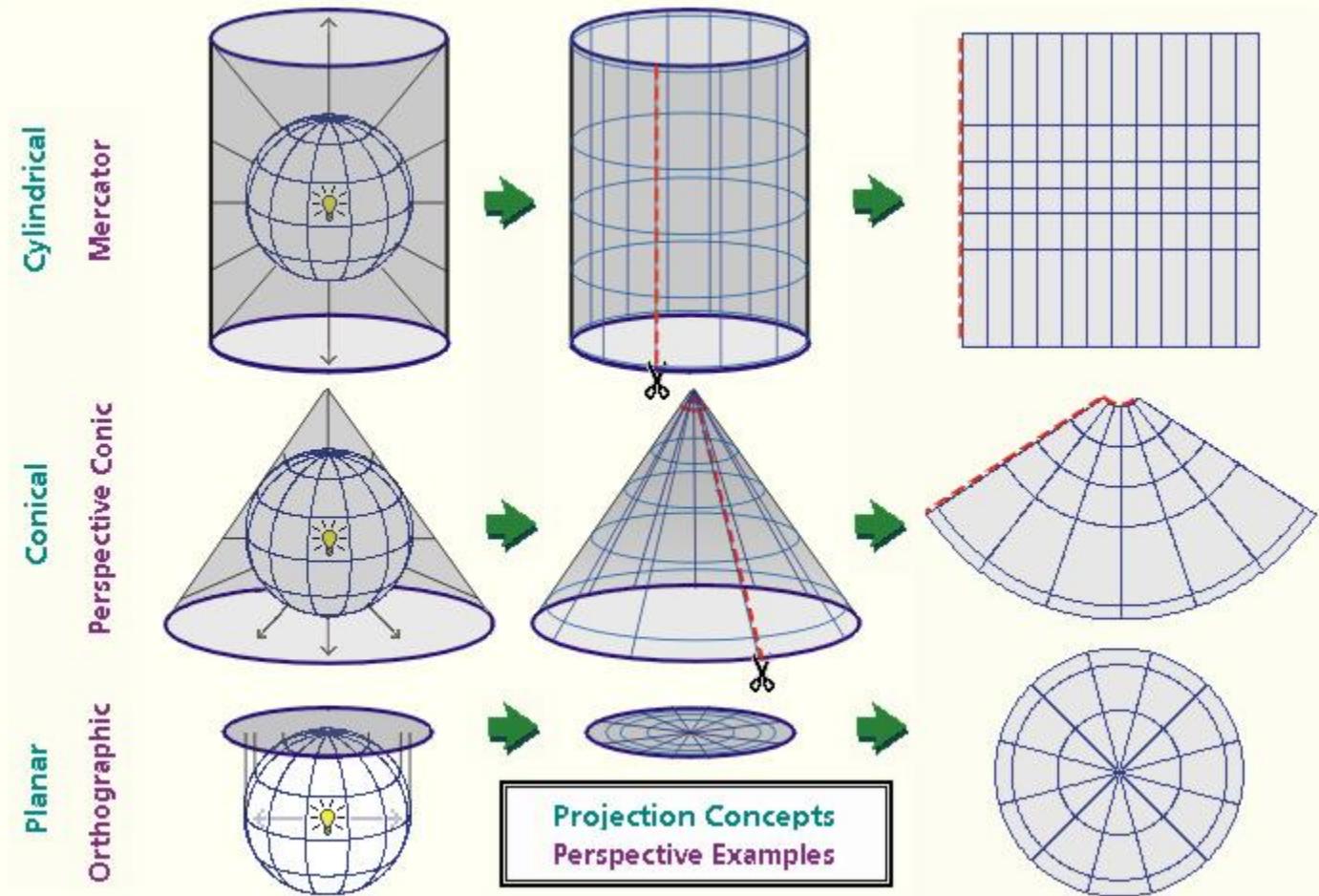
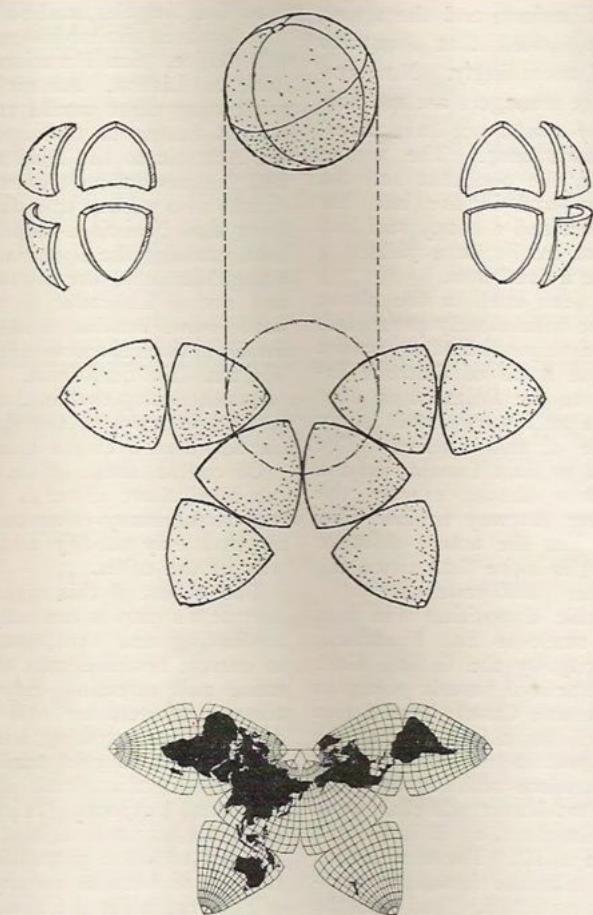
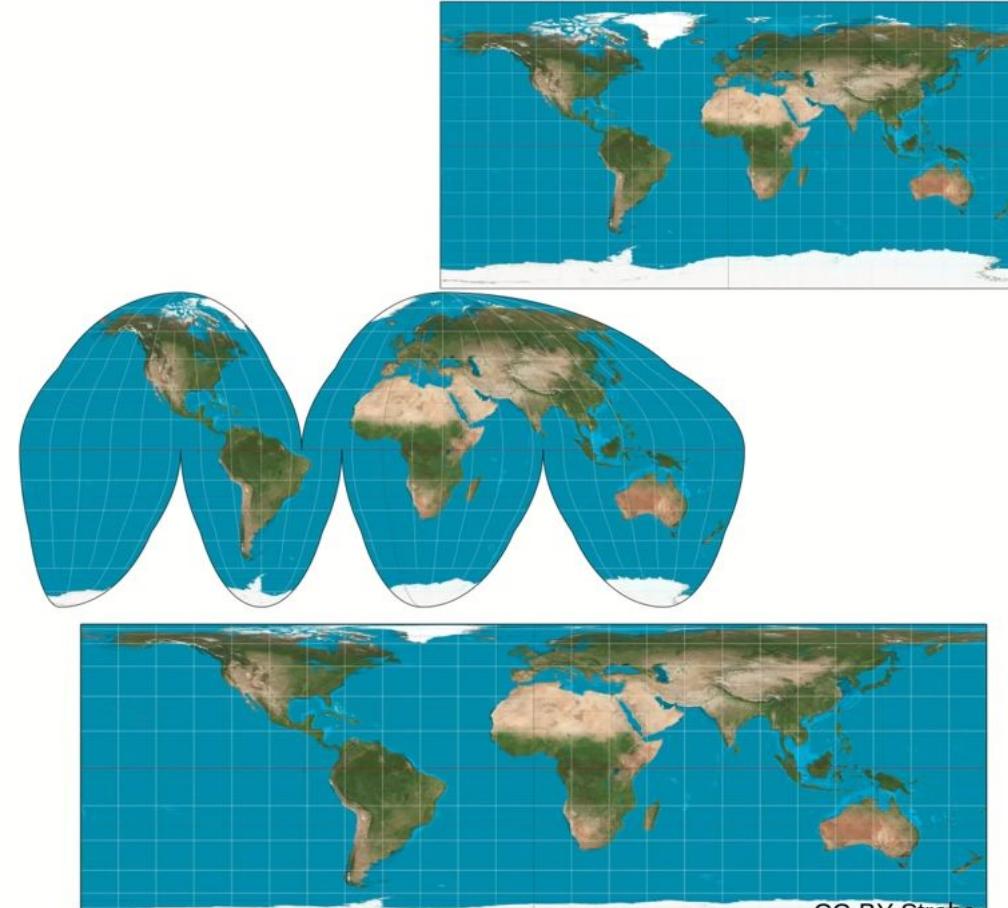


FIG. 45.—Man's head plotted on Mercator projection.

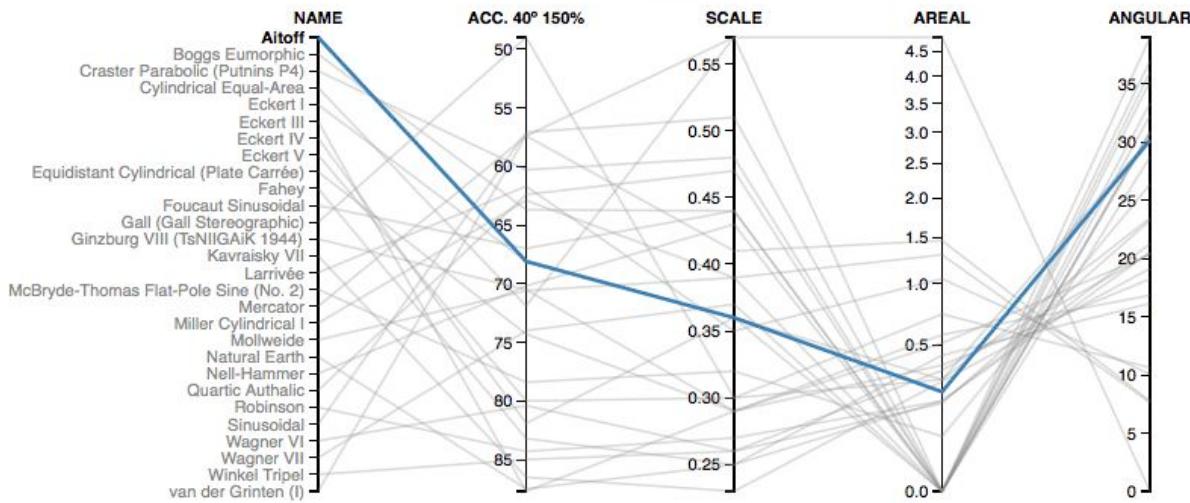
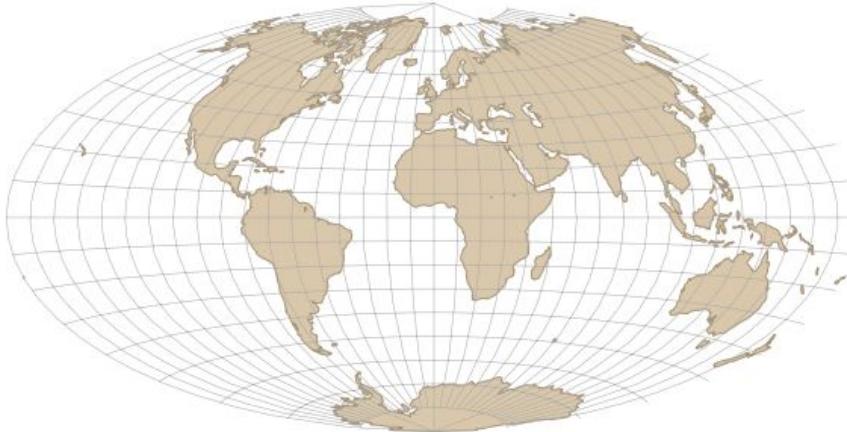




Public Domain



CC-BY Strebe





MAP SIZE

ACTUAL SIZE



Coordinate reference systems

Which coordinate reference system (CRS) should you use?

Rule of thumb:

- If you are working on a **global** map: use WGS84 (or Web Mercator)
- **France**: Lambert 1993, **US**: NAD1927, **UK**: OSGB 1936 (British National Grid)
- Each coordinate system can also be referenced by an **EPSG code**. You can find the conversion on <https://epsg.io/> (LAMB93 = EPSG 2154, WGS84 = EPSG 4326, OSGB36 = EPSG 27700, etc)

Spatial is special...

2. ... because of Geography laws and tools

You need systems to represent phenomena that happen on the surface of the planet, be them continuous or discrete.

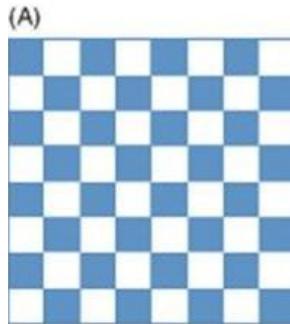
- Spatial autocorrelation
- Spatial analysis
- Spatial statistics

Tobler's (1979) First Law of Geography

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

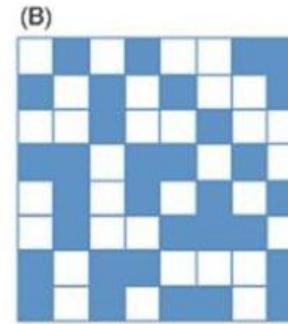


Spatial autocorrelation



Negative

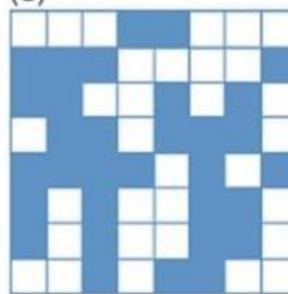
$I = -1.000$
 $n_{BW} = 112$
 $n_{BB} = 0$
 $n_{WW} = 0$



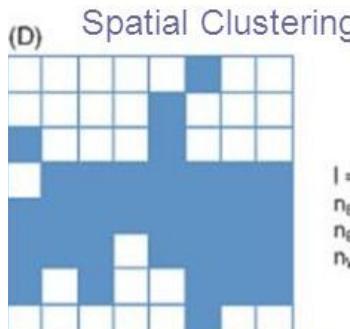
Dispersed

$I = -0.393$
 $n_{BW} = 78$
 $n_{BB} = 16$
 $n_{WW} = 18$

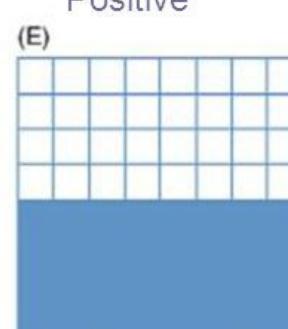
Spatial
Independence



$I = 0.000$
 $n_{BW} = 56$
 $n_{BB} = 30$
 $n_{WW} = 26$



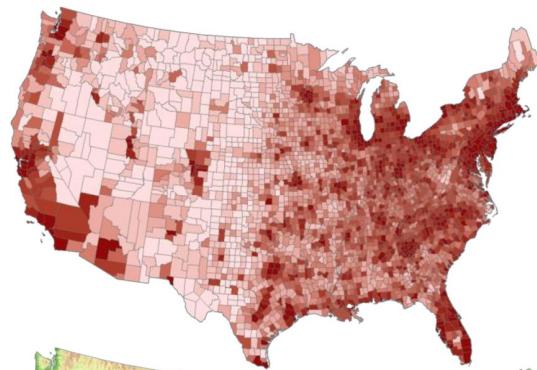
$I = +0.393$
 $n_{BW} = 34$
 $n_{BB} = 42$
 $n_{WW} = 36$



Positive

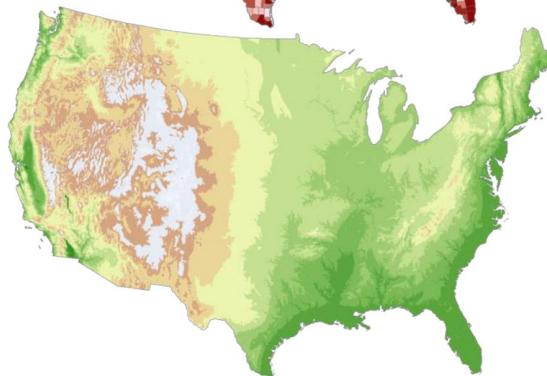
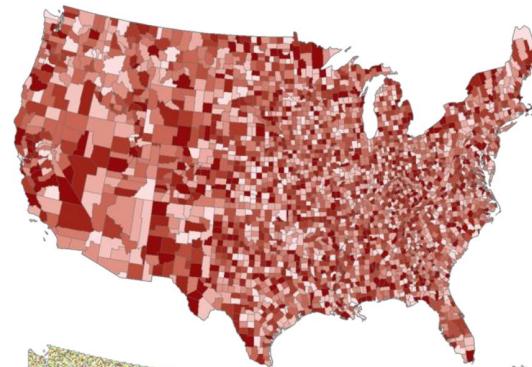
$I = +0.857$
 $n_{BW} = 8$
 $n_{BB} = 52$
 $n_{WW} = 52$

Spatial analysis

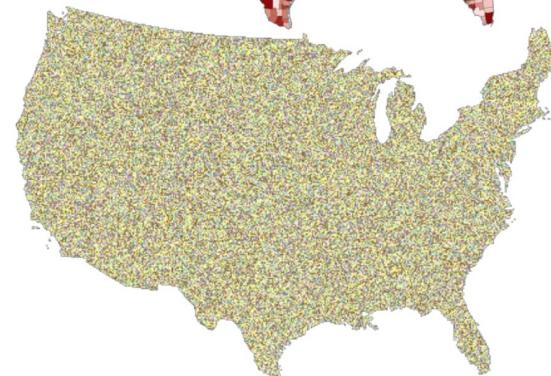


If features were
randomly distributed ...

... population
density map
of the US
would look
like this



... elevation map
of the US
would look
like this

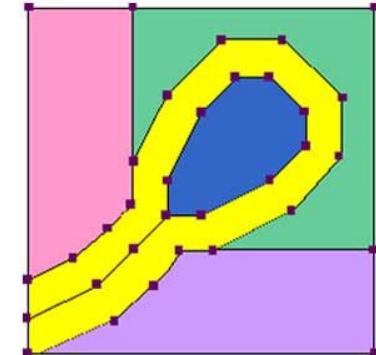


Spatial is special...

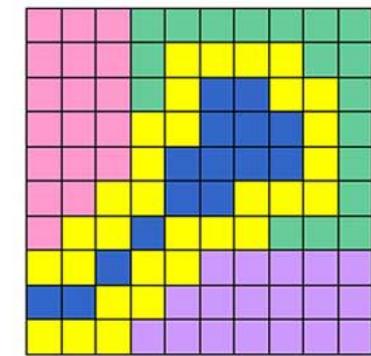
.... and can be represented in two ways:

- **Raster** : pixels, like a photo
- **Vector**: geometries: Point, Line, Polygon

Each of these data representations opens up a range of analysis tools.



Vector



Raster

Spatial is special...

3. ... because of cartography

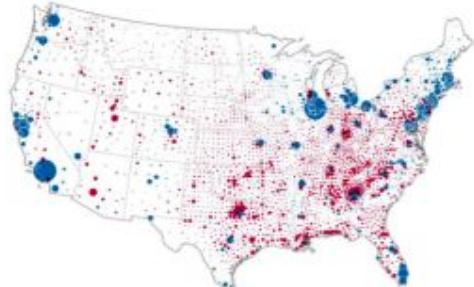
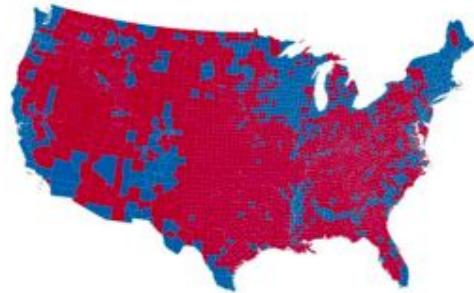
Cartography is a science and an art, and there are principles that help us better communicate with a target audience: legibility, visual contrast, figure-ground organization, hierarchical organization, and balance.

Every conscious or unconscious design choice you make will result in a map that only represents **one** of many possible stories about the spatial phenomenon your exploring.

Cartographic design

The same data,
displayed
two ways.

**Which is
the right
approach?**

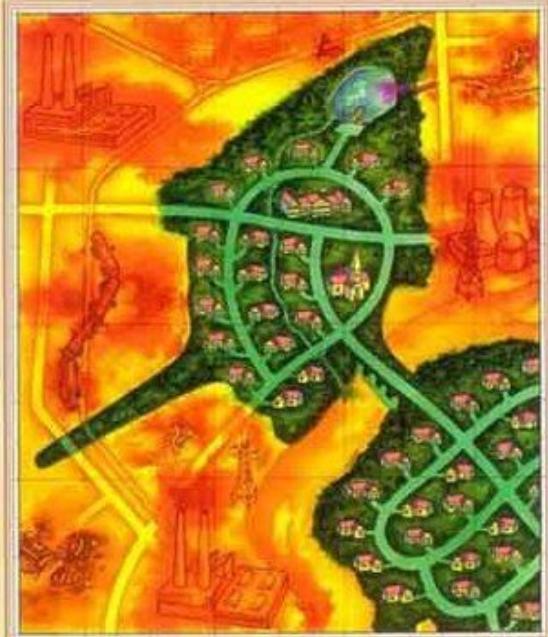


HBR.ORG

Mark Monmonier

How to Lie with Maps

Second Edition



With a new Foreword by H. J. de Blij

Ecological fallacy & Aggregation fallacy

Ecological fallacy is a logical error in the interpretation of statistical data, where you draw conclusions about *specific individuals* based on *aggregate data about a larger group*.

A useful thread on this by the American Journal of Epidemiology.

Ecological Study

An observational study with data analyzed at the group or population, rather than the individual level

Ecological Fallacy

A type of logical error which occurs when relationships which exist at the group level are incorrectly assumed to hold at the individual level

Aggregation Fallacy

A type of logical error which occurs when relationships which exist at the individual level are incorrectly assumed to hold at the group level

Ecological fallacy

There are 4 four common statistical ecological fallacies:

- confusion between group average and total average,
- confusion between higher average and higher likelihood
- confusion between ecological correlations and individual correlations,
- Simpson's paradox (when you have a trend that appears in several groups of data but disappears or reverses when the groups are combined)

Simpson's Paradox

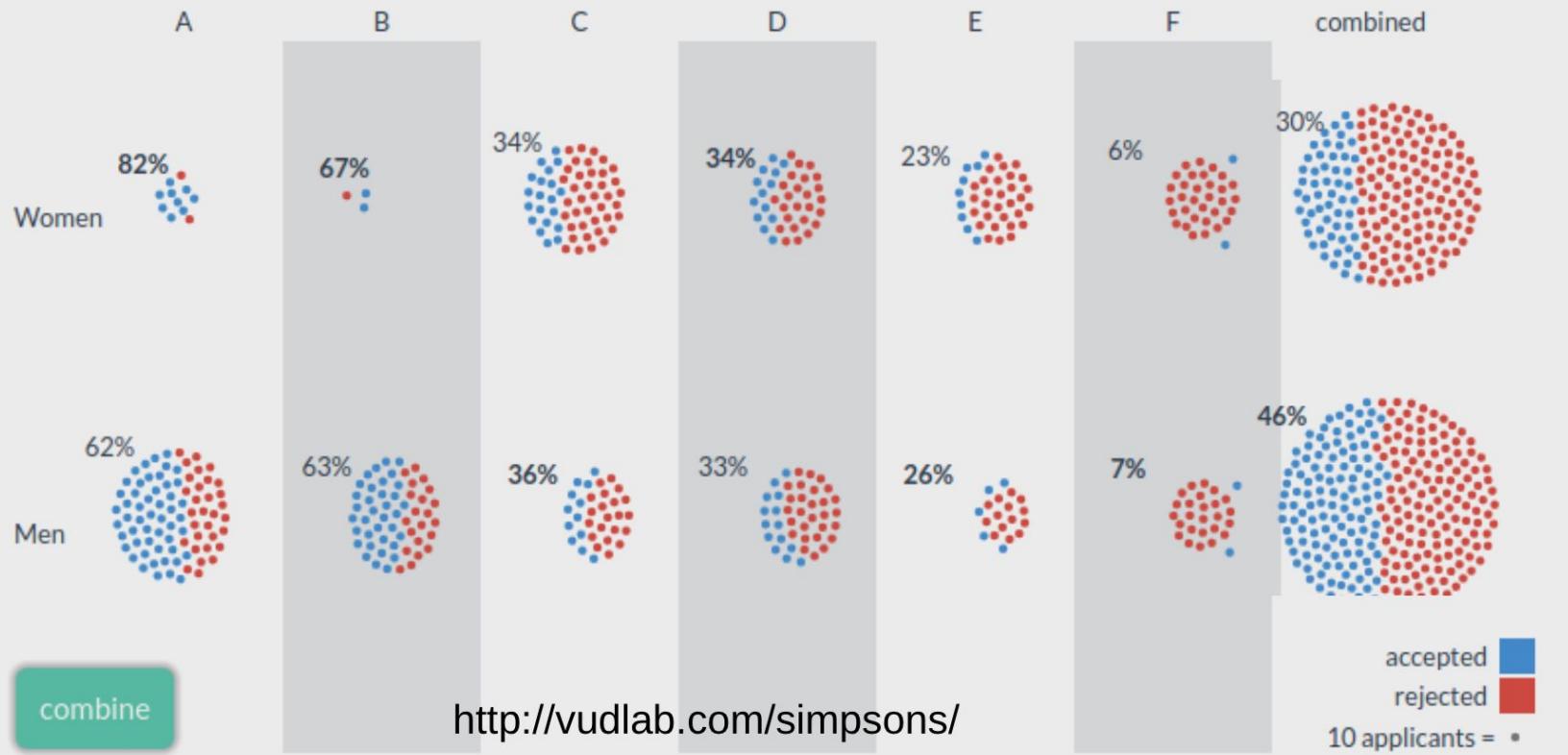
Berkeley gender bias case, 1973

46% of male applicants were accepted against 30% of female applicants.

But not a single department was showing to be significantly biased against women.

How is that possible?

Departments



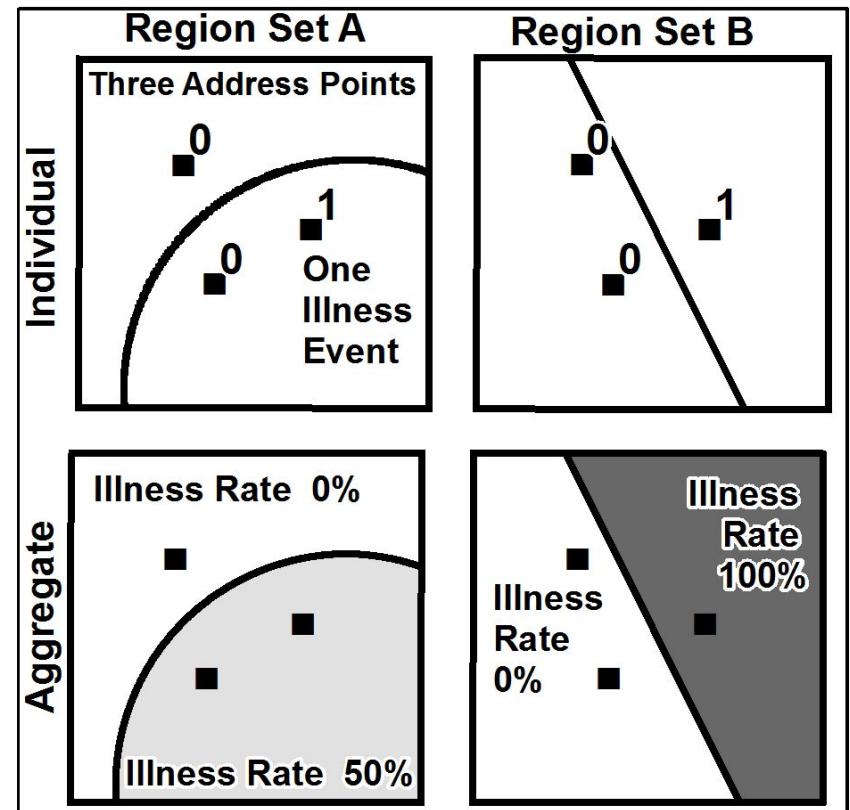
<http://vudlab.com/simpsons/>

combine

Credit: Nick Bearman

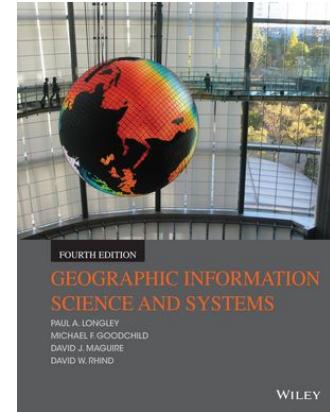
Modifiable Areal Unit Problem (MAUP)

Grouping data is **always** tricky,
and that is also true of spatial
data.



GIS Software

“The science of where”



“Geographic Information Systems are computer-based tools that analyze, store, manipulate and visualize geographic information, usually in a map.”

Michael Goodchild

Our Tools: open-source

- **QGIS** (we will be using QGIS for this course)
- **GeoDa** for data exploration
- Some **R** specialised packages (sp, rgdal, rgeos, tmap, raster, dplyr, RColorBrewer, classInt, leaflet...)
- Some **Python** libraries (shapely, geopandas, folium, rasterio, GDAL, scikit-image, rasterstats, PySAL, etc.)
- **Kepler.gl** for visualisation in browser only
- etc.

Our Tools: commercial software

- **ArcGIS:** software commercialised by ESRI. Older version of the Desktop software is called ArcMap (very clunky). Newest is called ArcGIS Pro. ESRI also offers an online suite for webmapping, location surveys etc.
Non-profit licenses are available:
<https://www.esri.com/en-us/solutions/industries/sustainability/nonprofit-program/overview>
- **FME** when dealing with complex data transformation tasks
- Some geospatial capability in **Tableau / PowerBI /Alteryx**

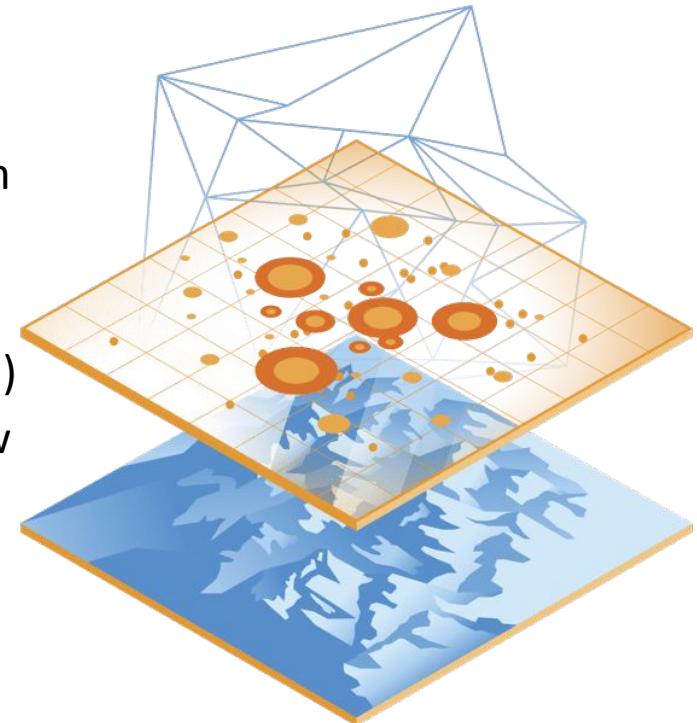
Use Cases

Breakout time

In groups of 2 to 3, discuss how GIS may be used to support ecological transitions at the city/metropolitan level.

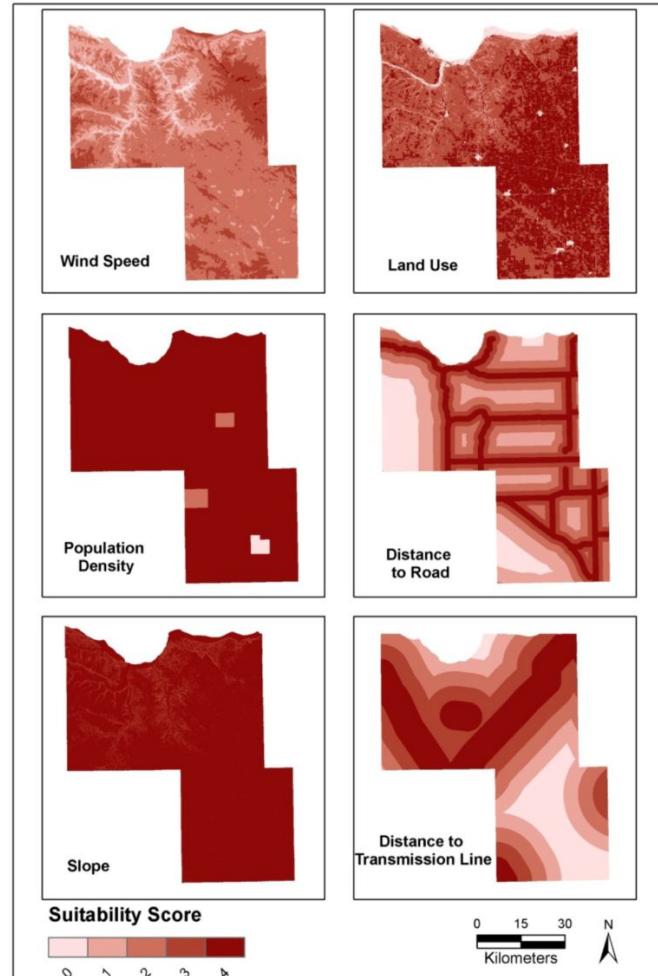
Informing decision making

- Location identification (e.g. optimal wind farm location, infrastructure planning)
- Network analysis (e.g. location/allocation problems, route optimization, service areas...)
- Trends and patterns detection (e.g. John Snow cholera outbreak)
- Modeling (e.g. flood risk mitigation, urban mobility patterns)



Site selection

Wind farms: suitability analysis



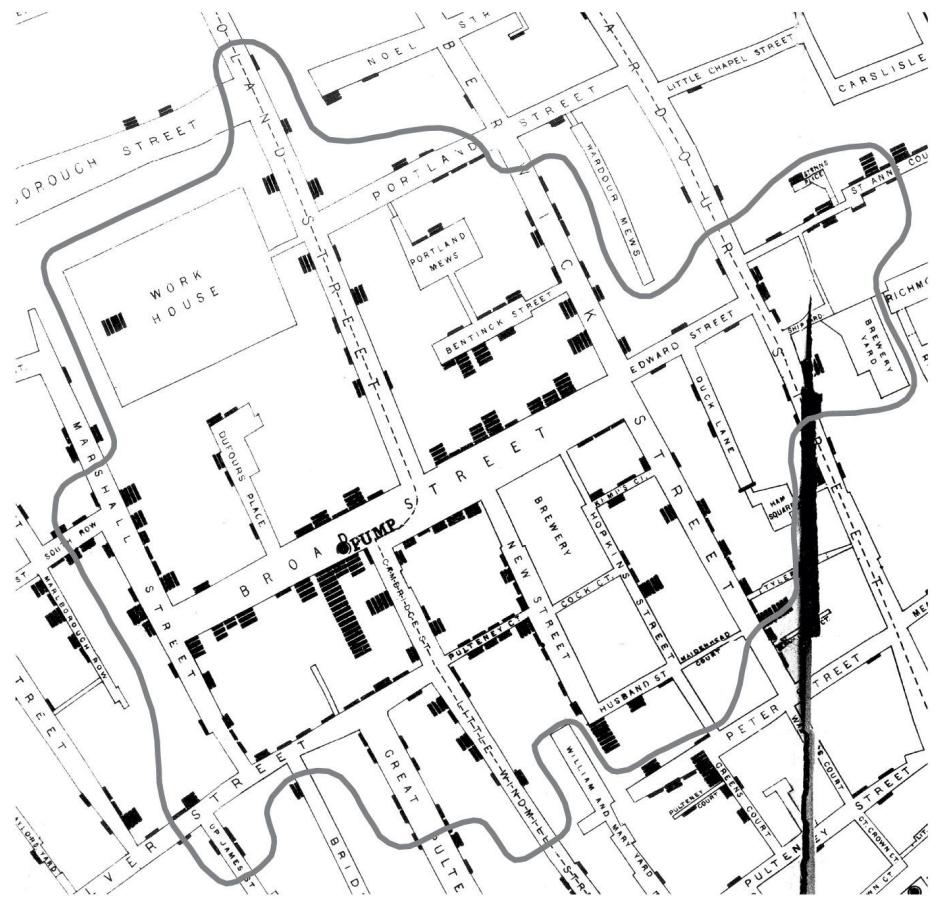
Network Analysis

- Route / Vehicle routing optimisation
- Service Areas
- Closest Facility
- Location /allocation problems
- Isochrones
- etc.



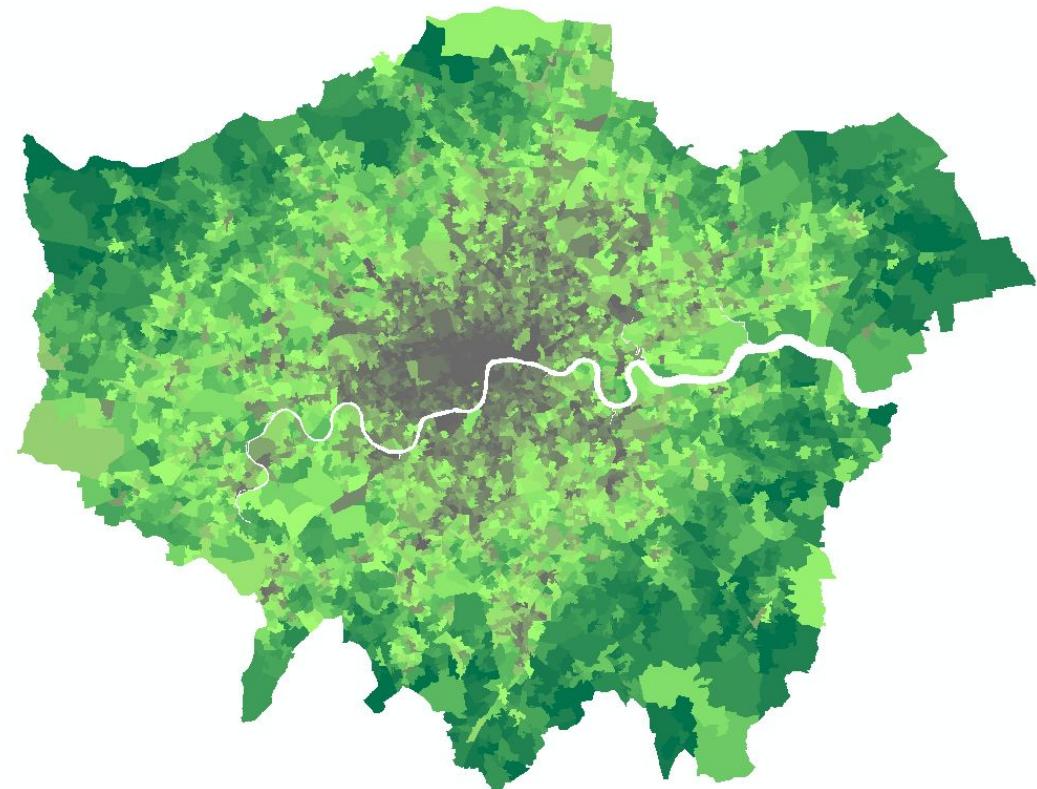
Pattern Detection

John Snow's Pump in London's
1854 Cholera outbreak



Pattern Detection

London air pollution...



Risk Mitigation

- **Assessing Flood-related economic damages**
- Understanding vulnerability to climate change in Nepal



Risk Mitigation

- Assessing Flood-related economic damages
- **Understanding vulnerability to climate change in Nepal**



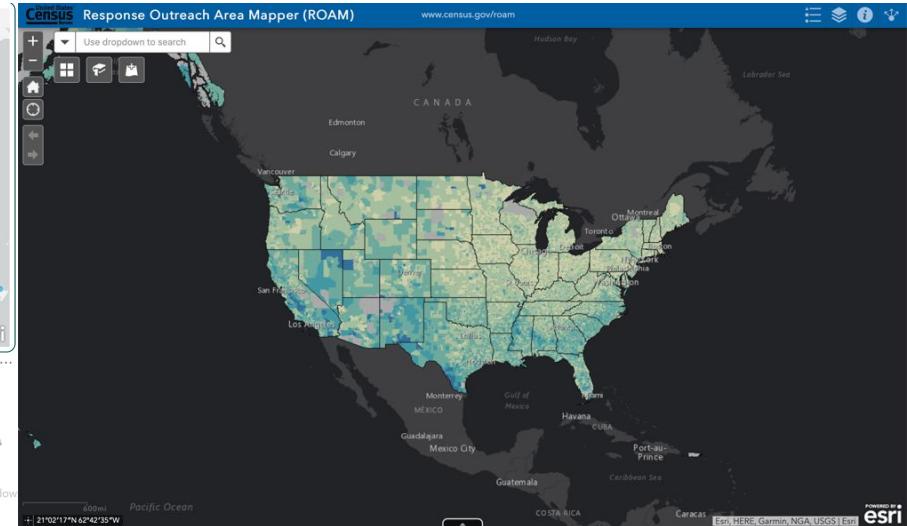
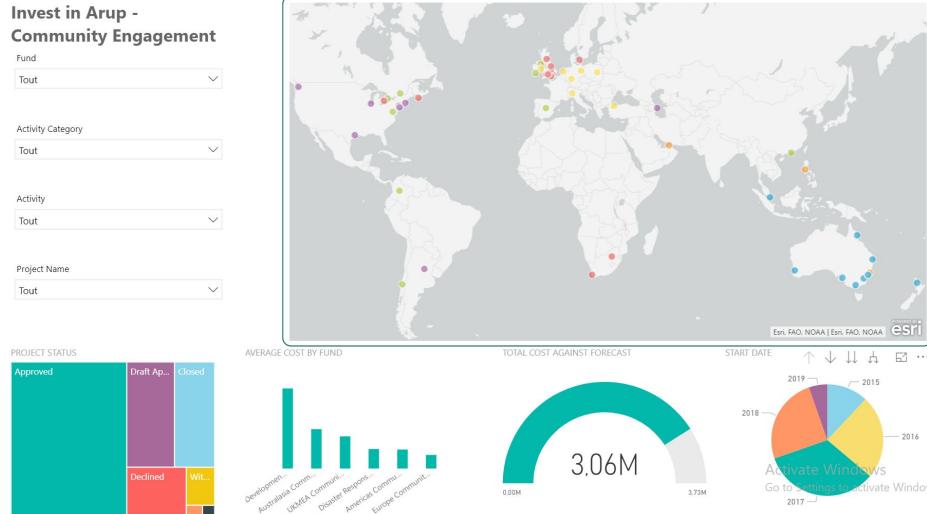
Mobility / Transport planning



Understanding urban mobility patterns



Insights (Dashboards, Webmaps etc)



A typical GIS Workflow

1. Requirements gathering



2. Finding the DATA



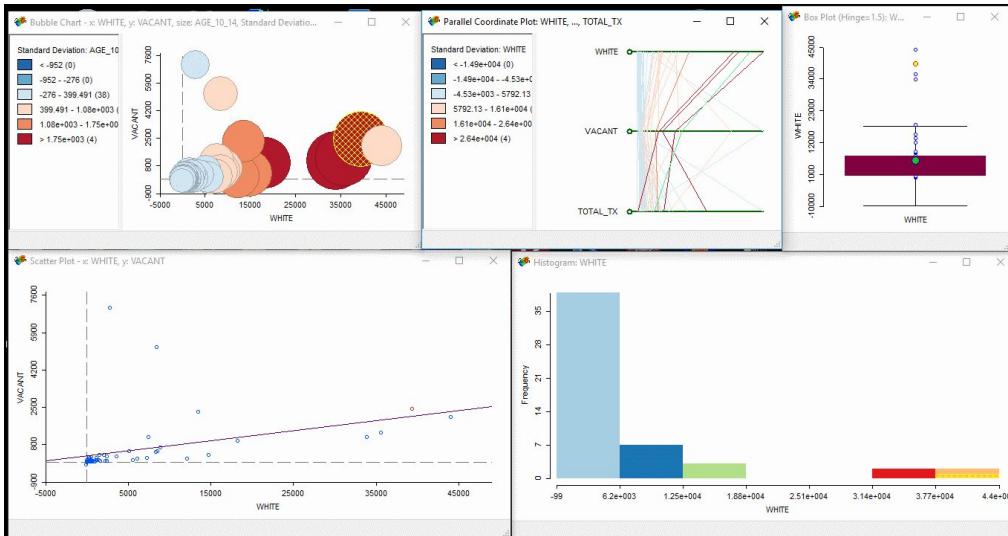
3. Exploratory Spatial Data Analysis (ESDA)



I'm going on an adventure!

3. Exploratory Spatial Data Analysis (ESDA)

- GeoDA <https://geodacenter.github.io/documentation.html>

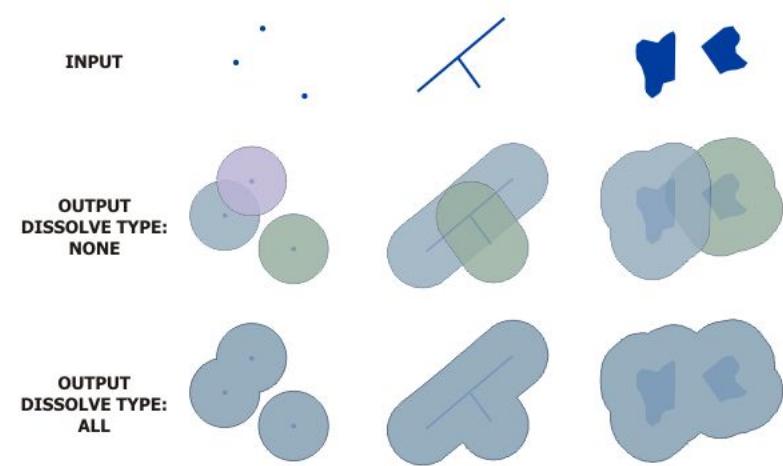
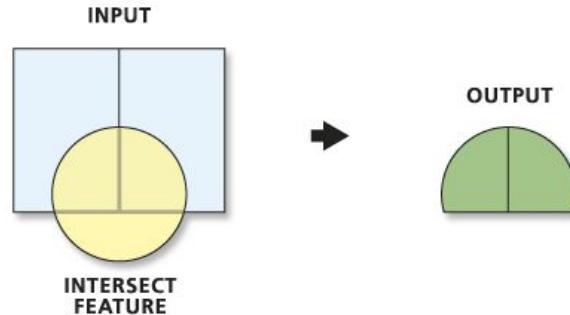


4. Refining the research objective and goals



5. Reshaping your data / feature generation

- **Geoprocessing tools** (geometry-based)
- And every other data science tool :)
(Dimensionality reduction is often useful)



6. Analysis / Modeling

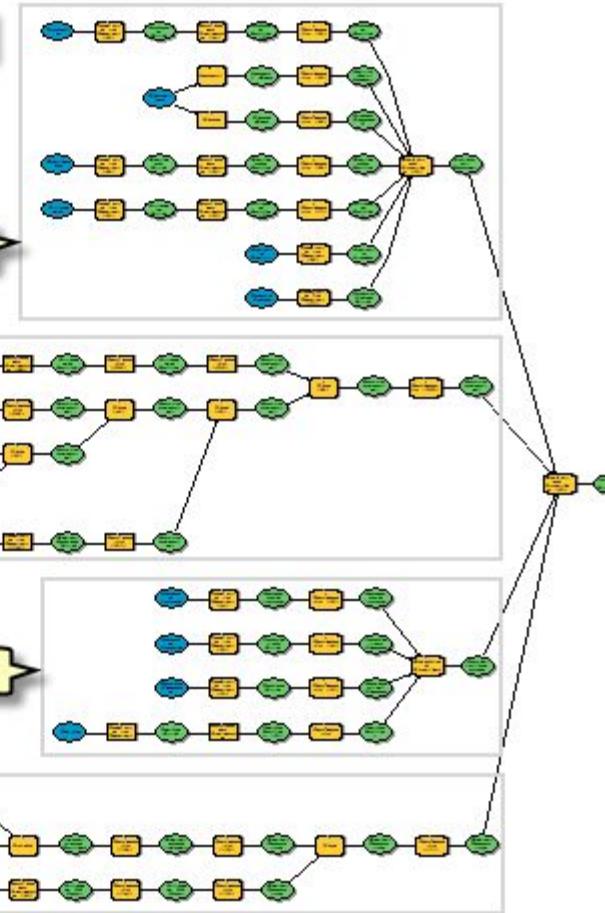


Break a big model into smaller models.

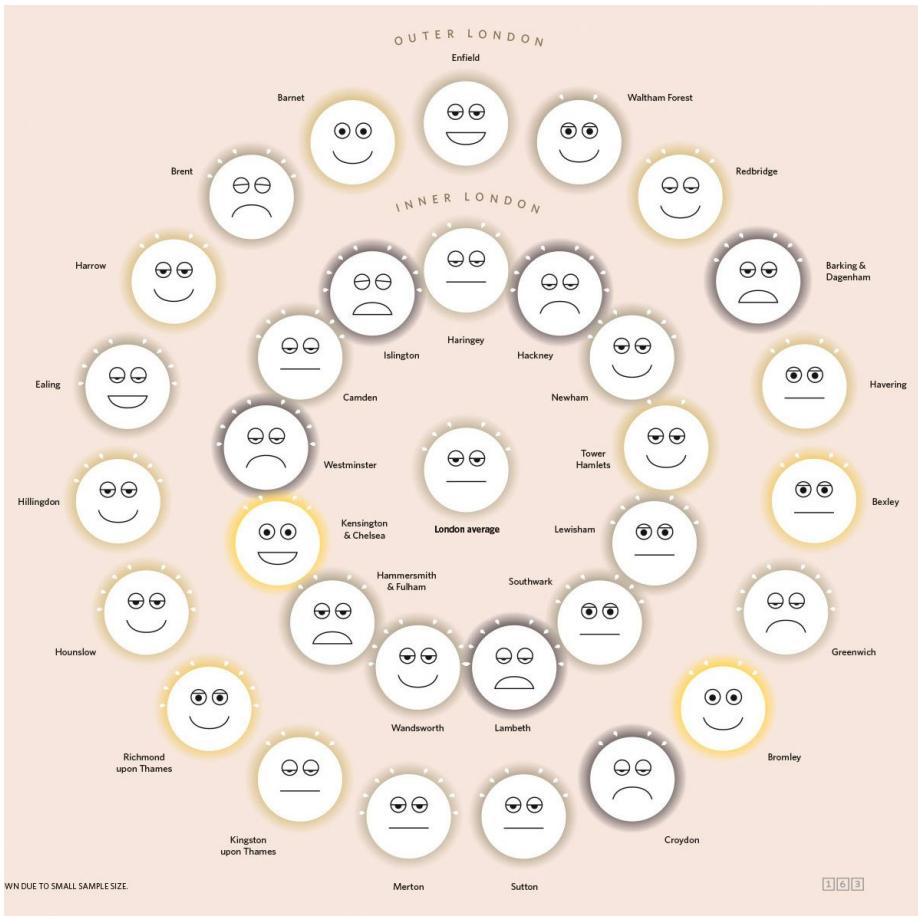
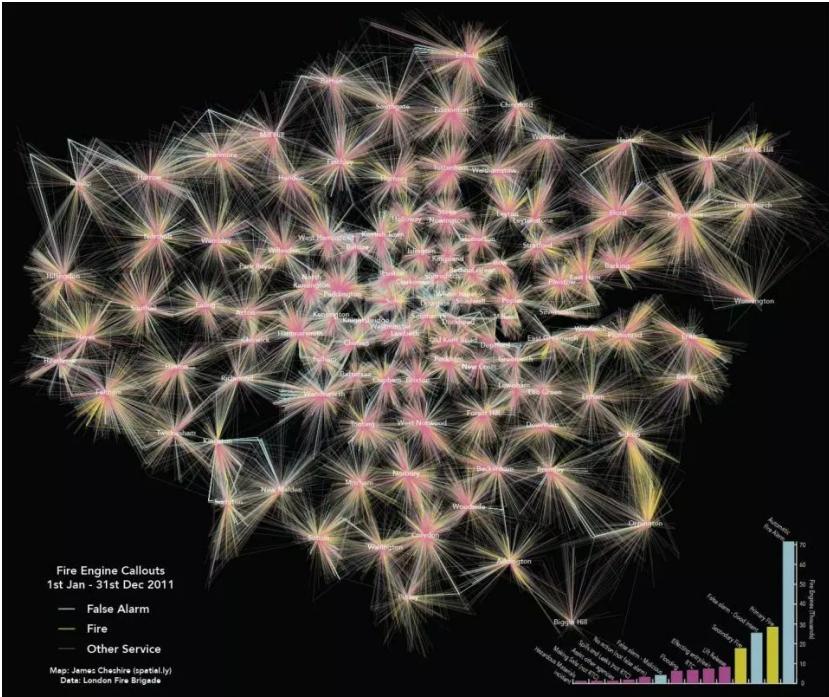
Sub-Model 2

Sub-Model 1

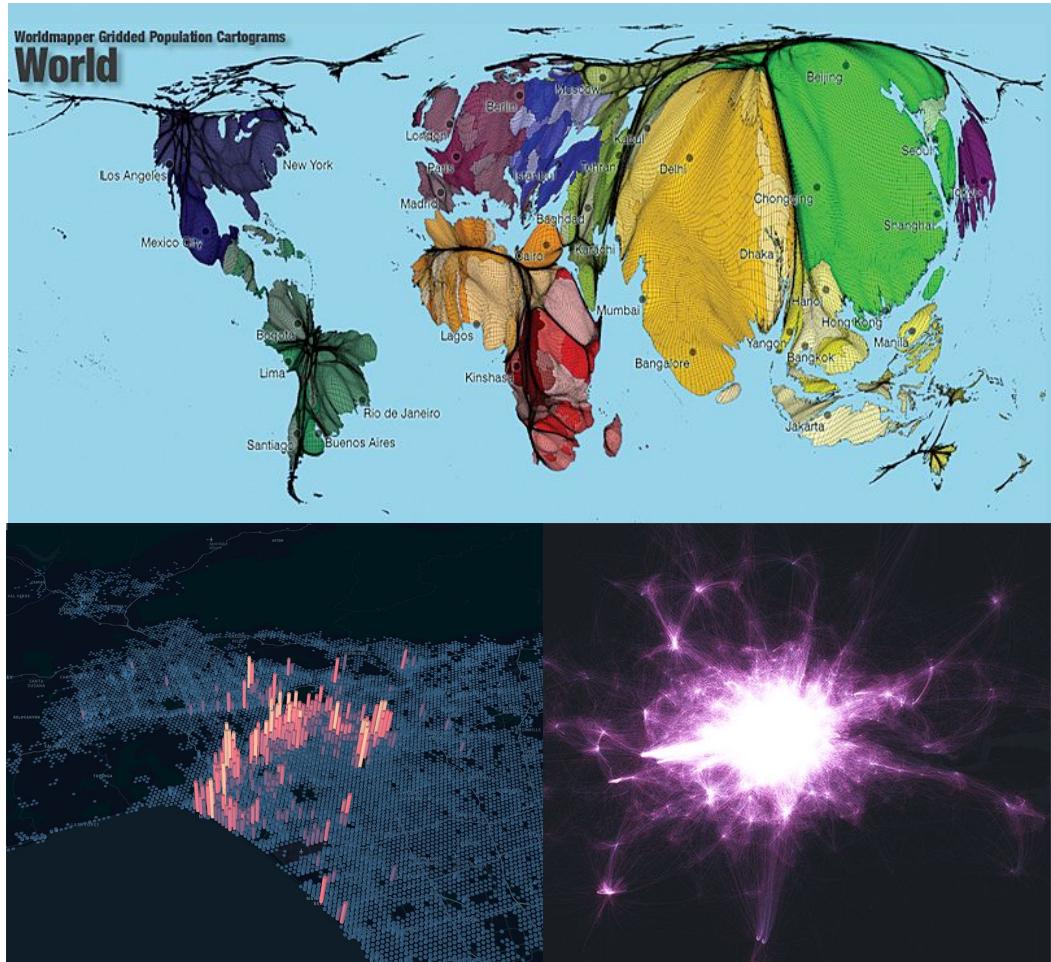
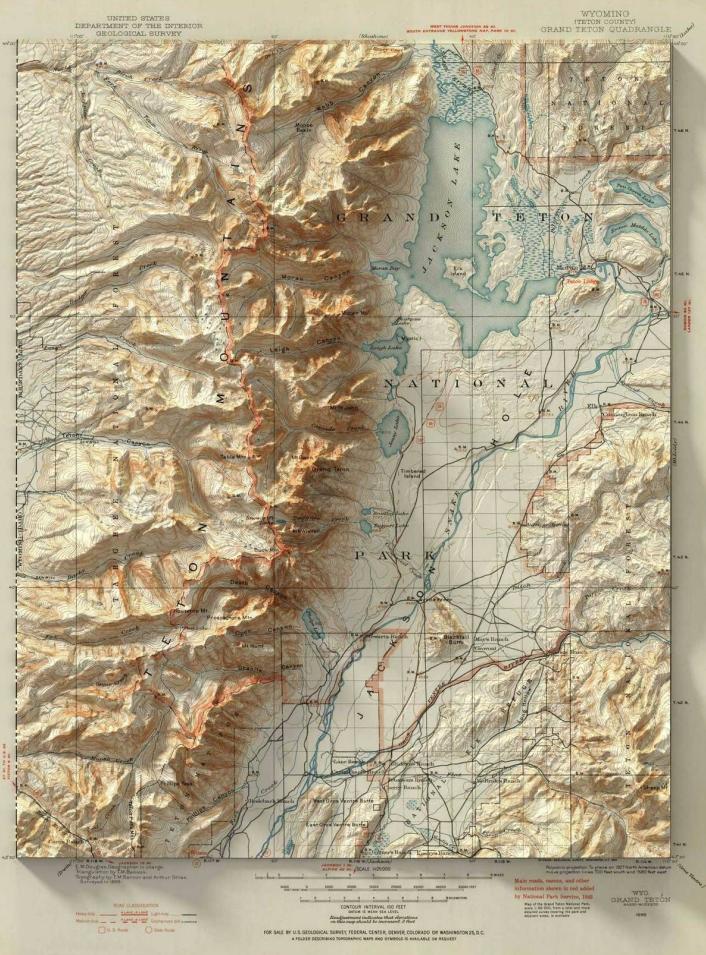
Sub-Model 4



7. Visualisation







Homework

Homework

1. Complete the [Session 1 tutorial](#)
2. Join Slack using [this link](#)
3. Re-read these slides and write down any questions you may have (you can also post them in the #help channel in Slack).