Introduction to GIS

Raphaëlle ROFFO

Sciences Po - Urban School
Fall 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. Tutorial 1

Attendance

Student representative election

Technical setup

Technical setup

- Slides and tutorials are available on **this website**.
- Tutorial 1 will walk you through downloading QGIS and exploring the interface.
- Sessions 1, 4 and 6 will be taught in person. Sessions 2,3 and 5 will be via Zoom, but a classroom is still booked for you.

Online 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

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, Youtube videos, 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

Coursework & Participation

Final Coursework Overview

Deadline: Monday 24th October, 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 you 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 October, 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:

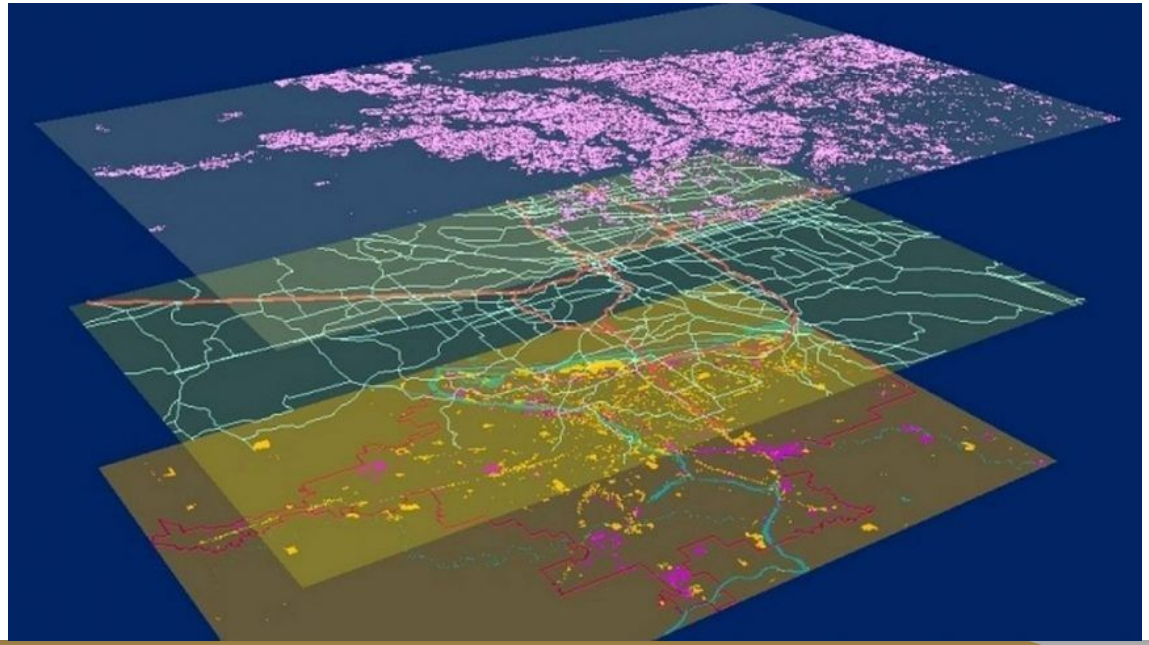
- What you've understood about the material
- What would be possible use cases for the tools / techniques introduced
- Ask questions about the things you find confusing

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

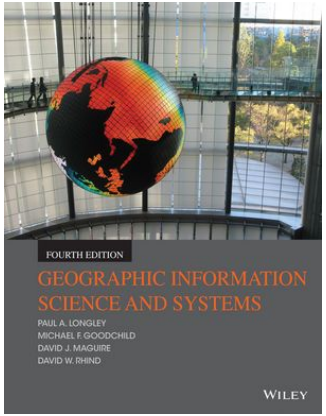
This will be followed by additional Q&A to answer any question you may have.

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.

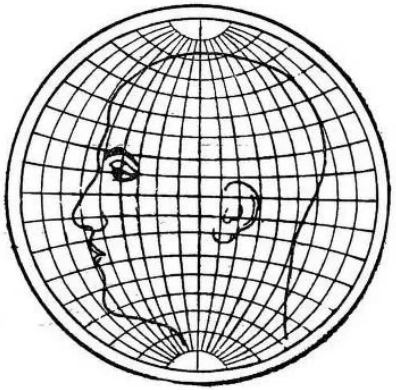


FIG. 42.—Man's head drawn on globular projection.

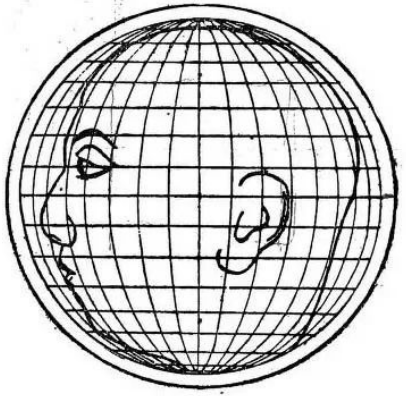


FIG. 43.—Man's head plotted on orthographic projection.

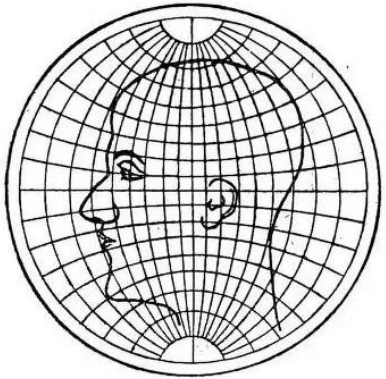


FIG. 44.—Man's head plotted on stereographic projection.

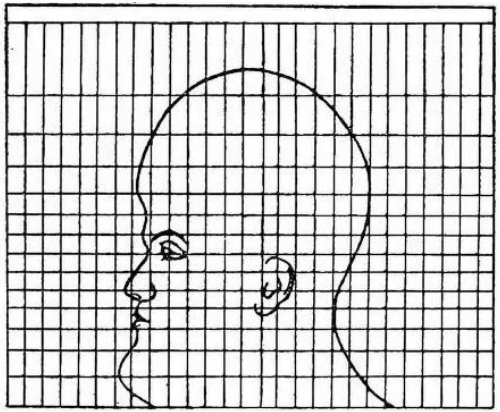
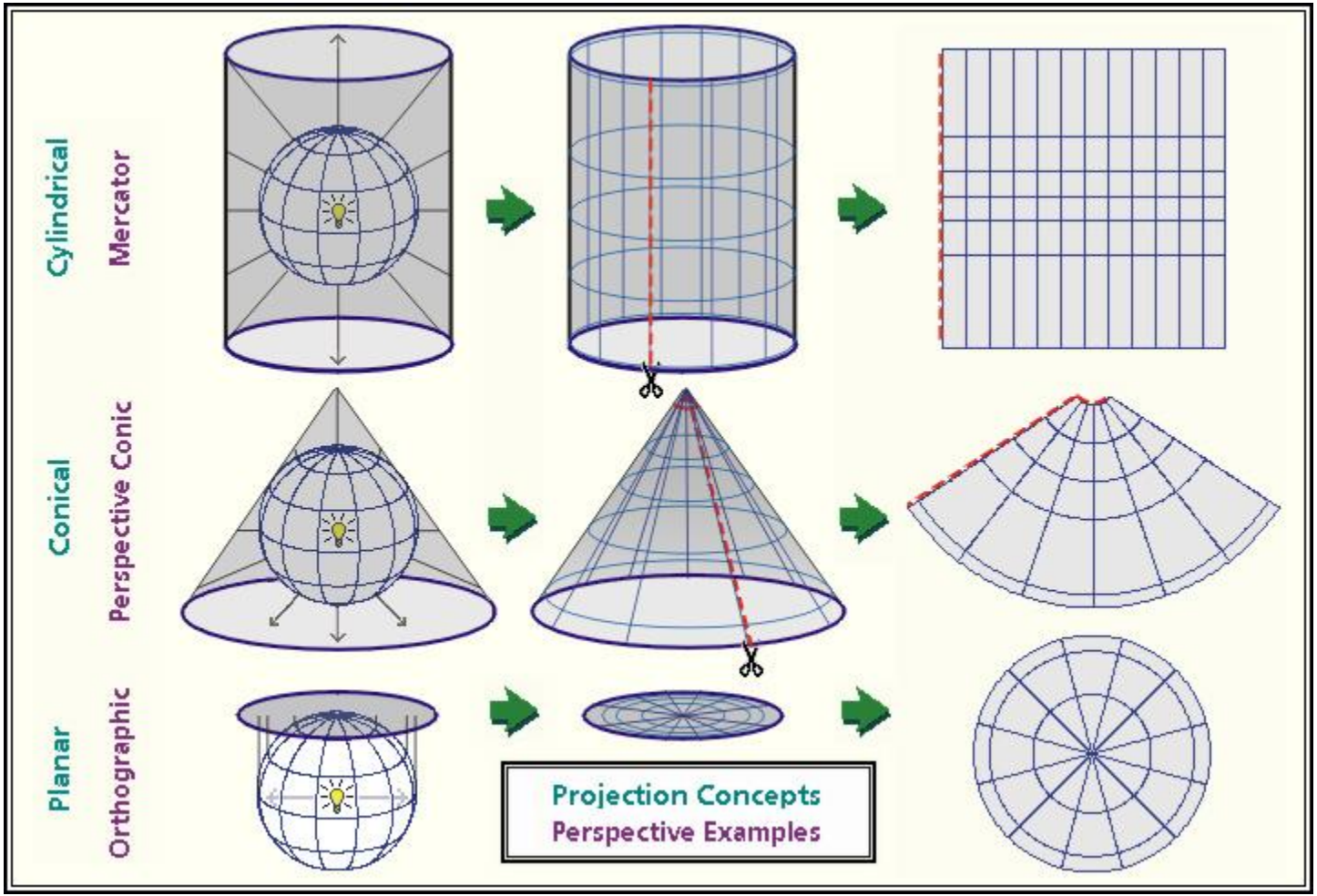
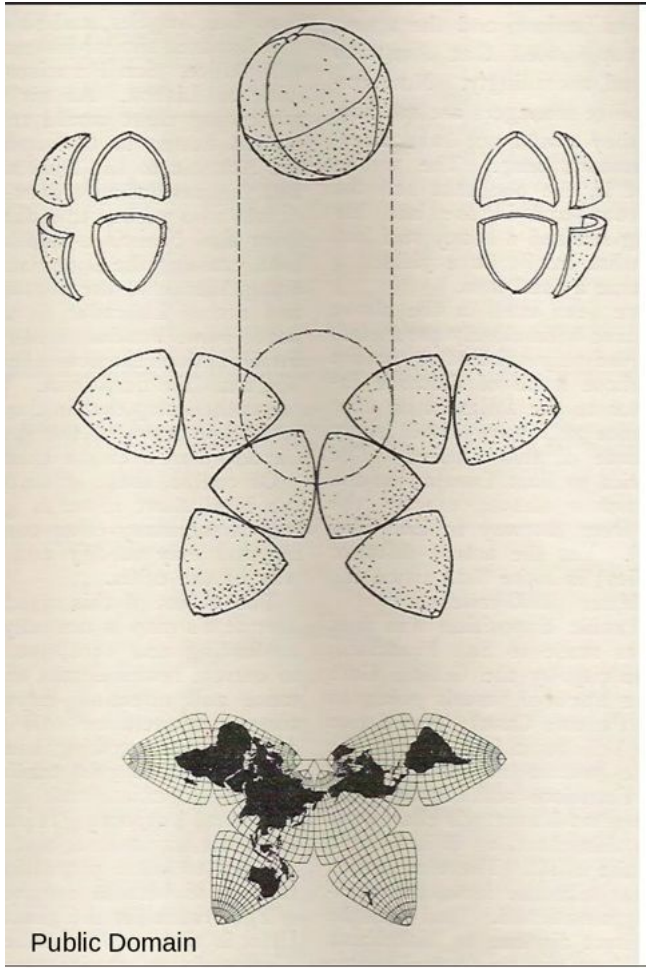
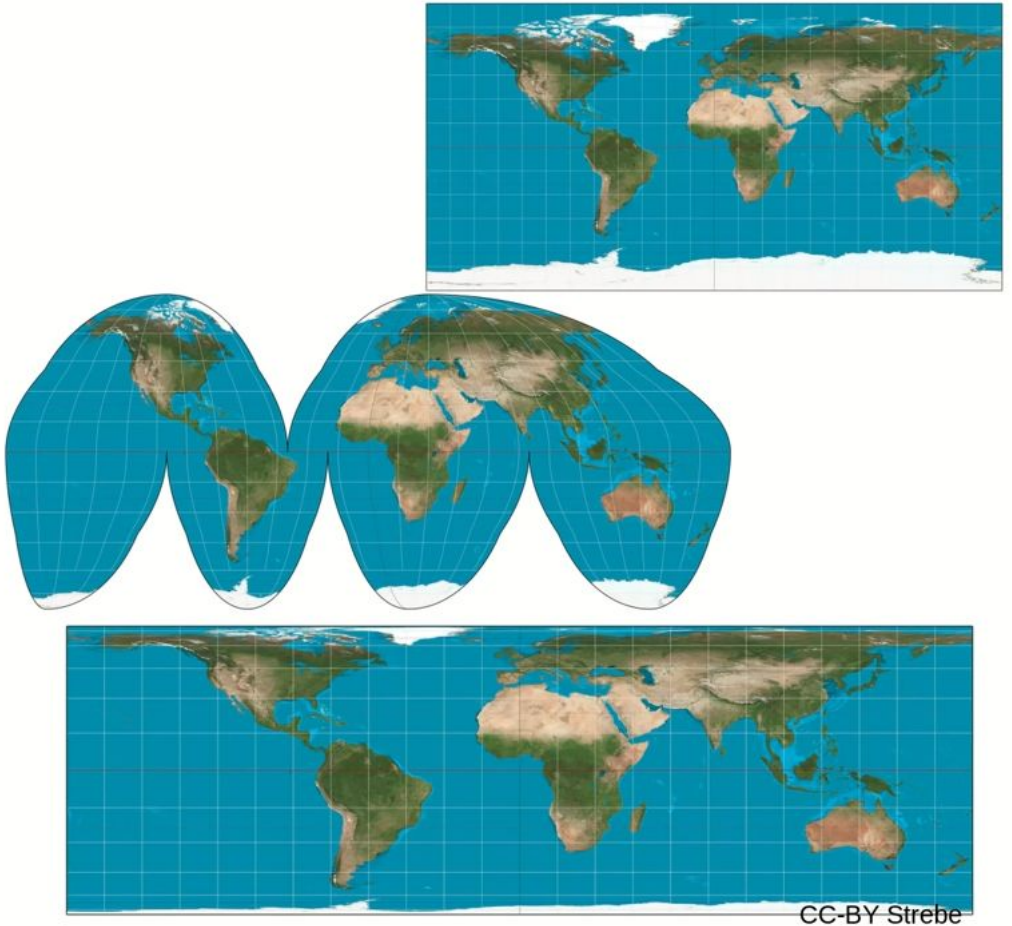


FIG. 45.—Man's head plotted on Mercator projection.

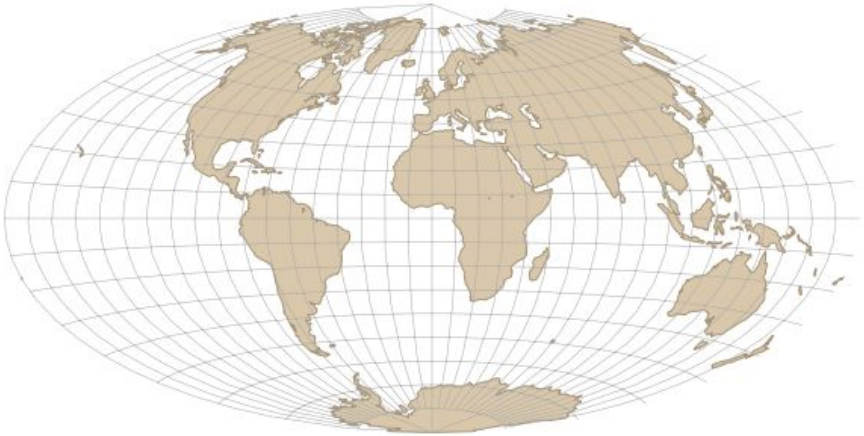

Cylindrical
Mercator
Conical
Perspective Conic
Planar
Orthographic
Projection Concepts
Perspective Examples

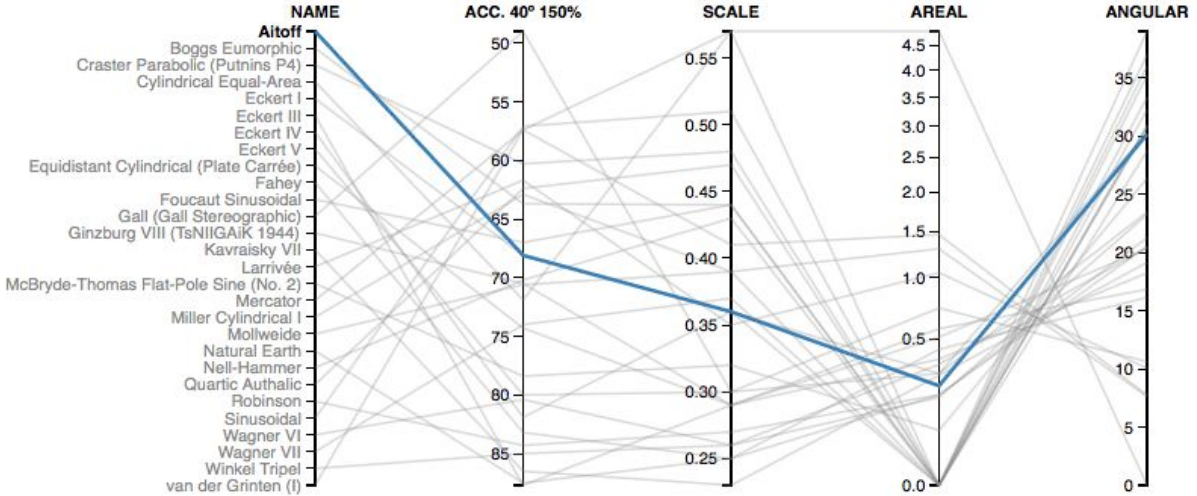


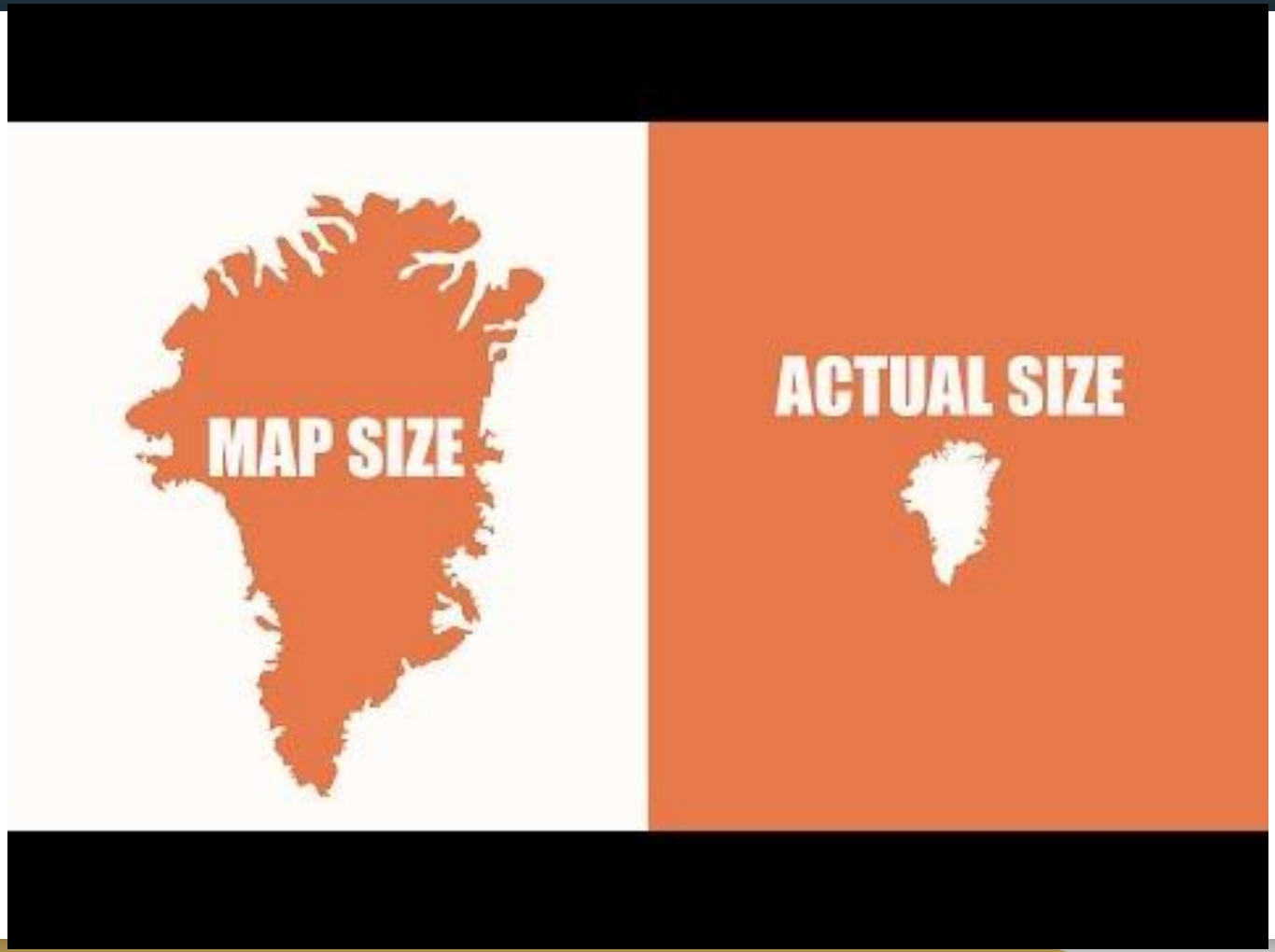
Public Domain



CC-BY Strebe




NAME
ACC. 40° 150%
SCALE
AREAL
ANGULAR
Aitoff
Boggs Eumorphic
Craster Parabolic (Putnins P4)
Cylindrical Equal-Area
Eckert I
Eckert III
Eckert IV
Eckert V
Equidistant Cylindrical (Plate Carrée)
Fahey
Foucaut Sinusoidal
Gall (Gall Stereographic)
Ginzburg VIII (TsNIIGAiK 1944)
Kavraisky VII
Larrivée
McBryde-Thomas Flat-Pole Sine (No. 2)
Mercator
Miller Cylindrical I
Mollweide
Natural Earth
Nell-Hammer
Quartic Authalic
Robinson
Sinusoidal
Wagner VI
Wagner VII
Winkel Tripel
van der Grinten (I)
50
55
60
65
70
75
80
85
0.55
0.50
0.45
0.40
0.35
0.30
0.25
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
35
30
25
20
15
10
5
0


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 phenomenons 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

(A)

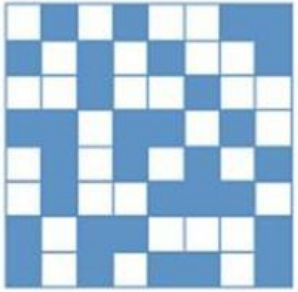
Negative

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

(B)

Dispersed

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



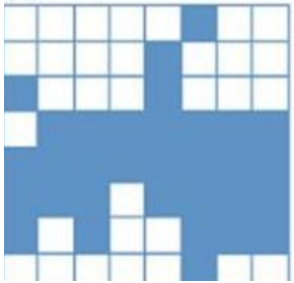
Spatial Independence

(C)

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

(D)

Spatial Clustering

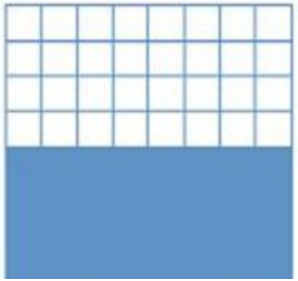


I = +0.393
$n_{BW} = 34$
$n_{BB} = 42$
$n_{WW} = 36$

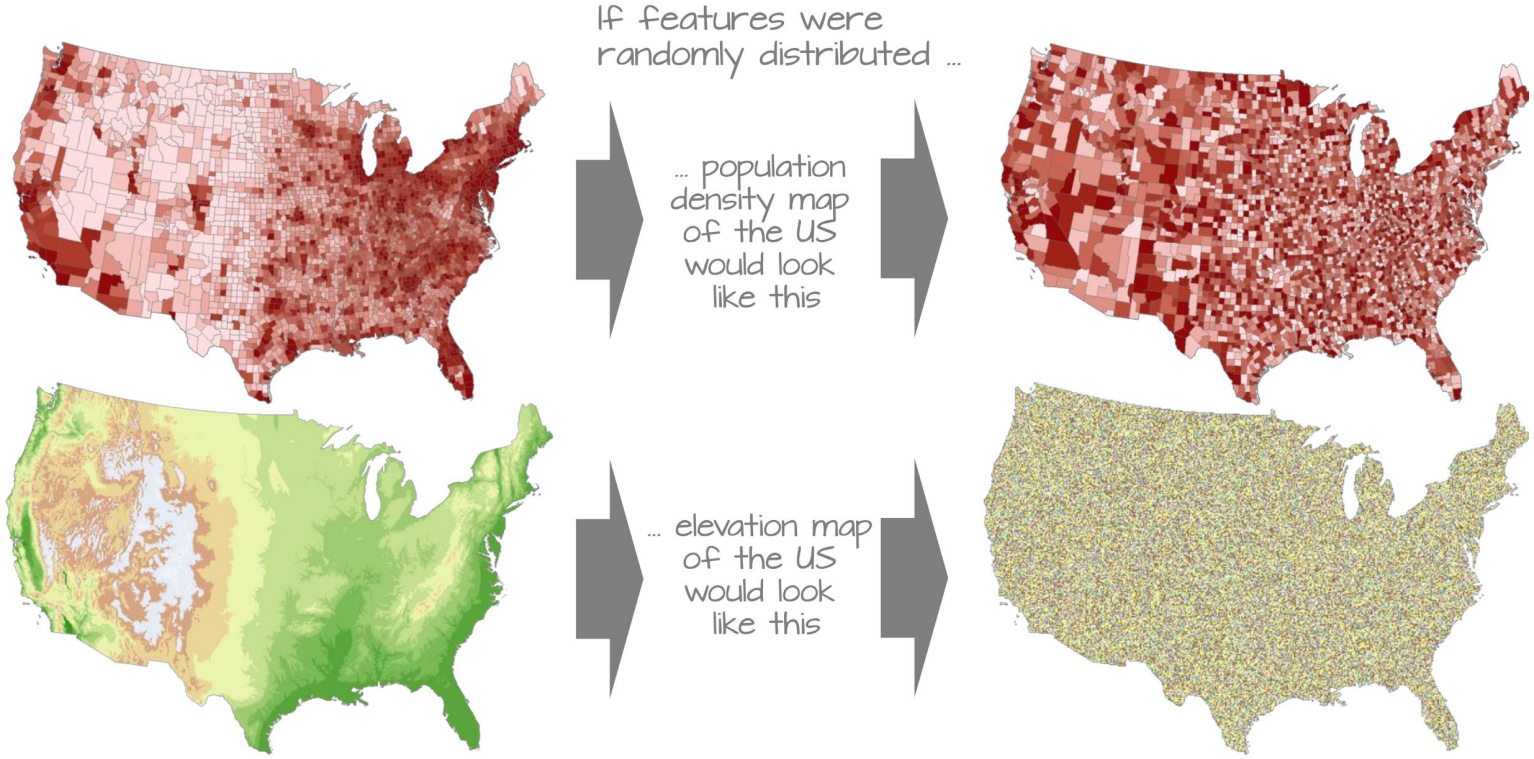
Positive

(E)

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



Spatial analysis

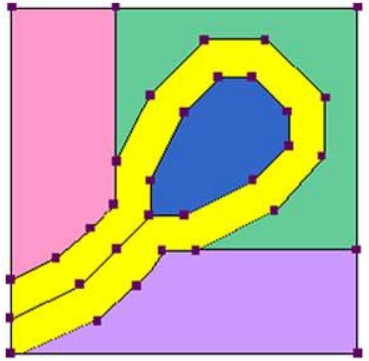


Spatial is special…

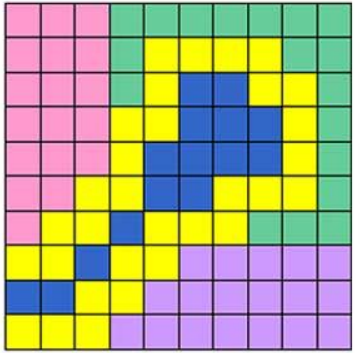
…. and can be represented in two ways:

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

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



Vector



Raster

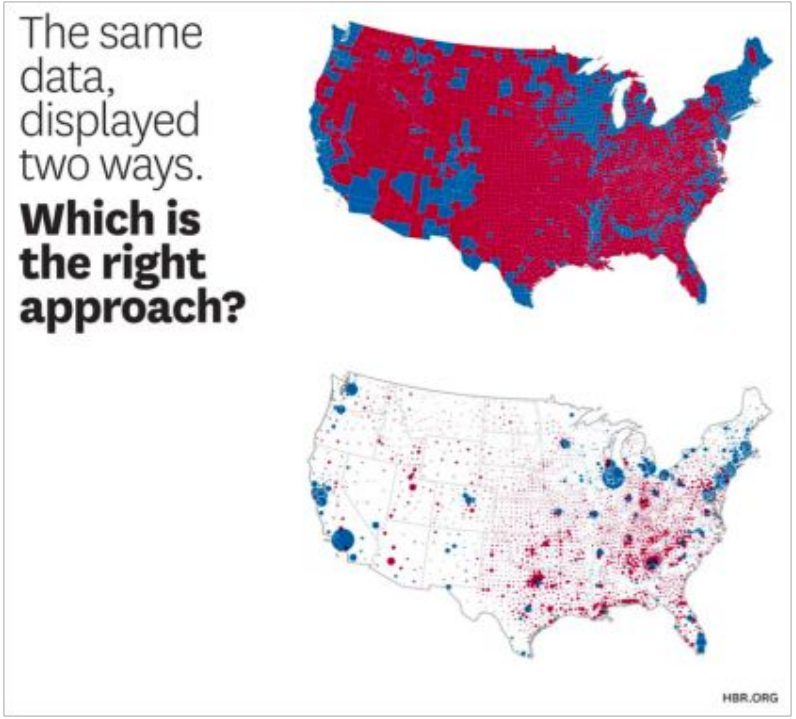
Spatial is special…

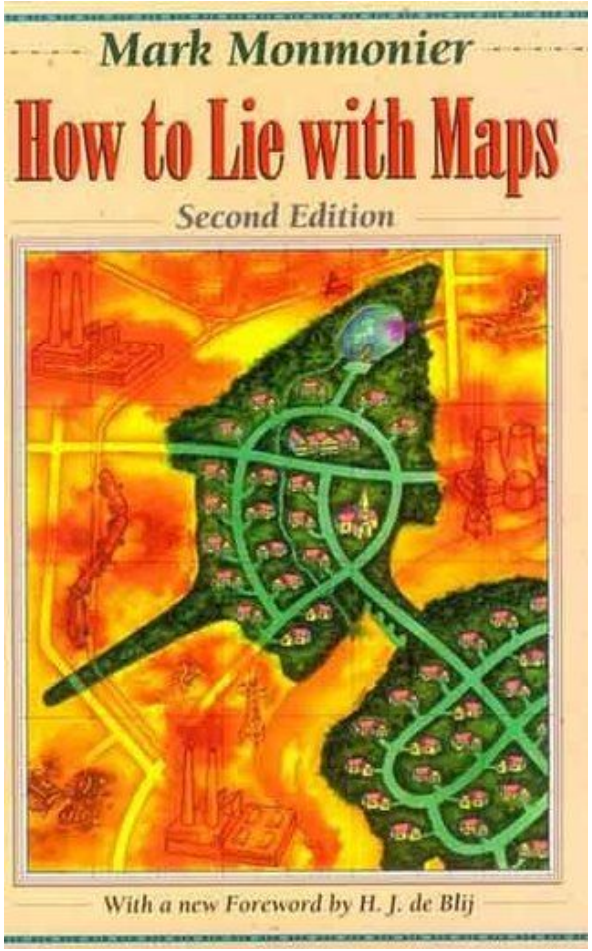
3. … because of cartography

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

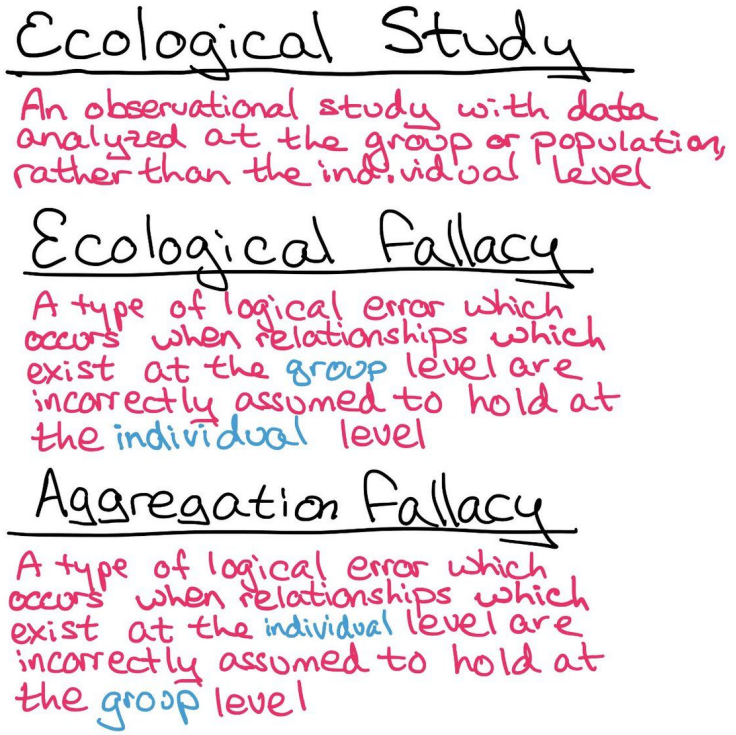




Ecological fallacy & Aggregation fallacy

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 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.

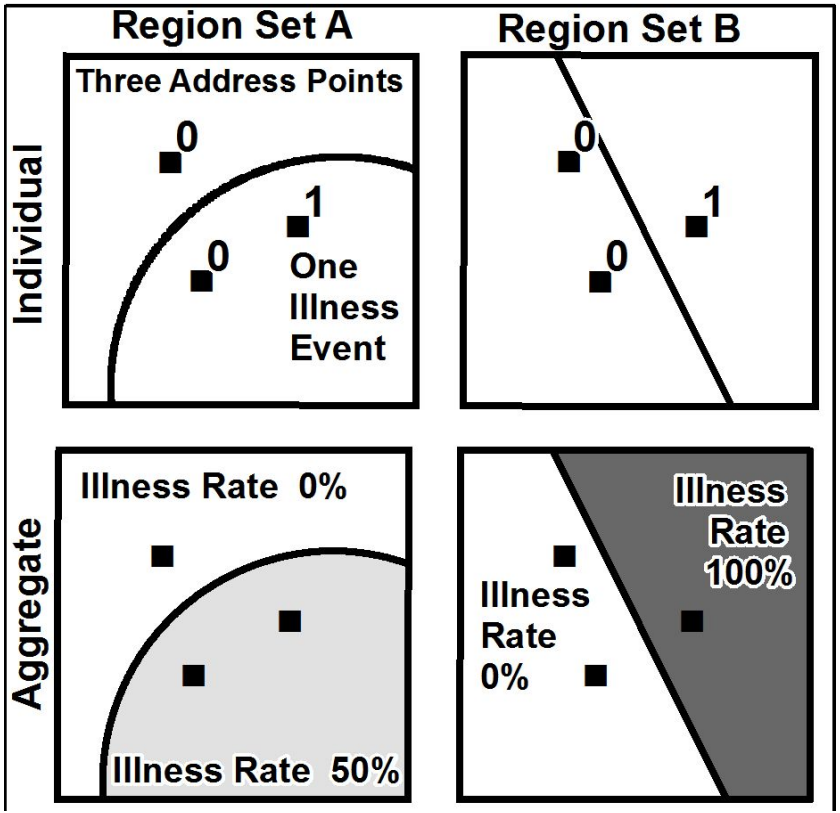
But not a single department was showing to be significantly biased against women.

How is that possible?

Credit: Nick Bearman

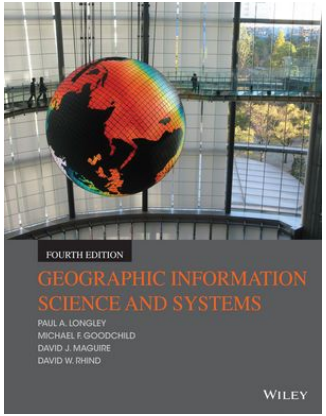
Modifiable Areal Unit Problem (MAUP)

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 (gdal, pyproj, fiona, shapely, geopandas, folium, rasterio, GDAL, sckikit-image, rasterstats, 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, but ArcGIS Pro is being rolled out and will ultimately replace it. 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

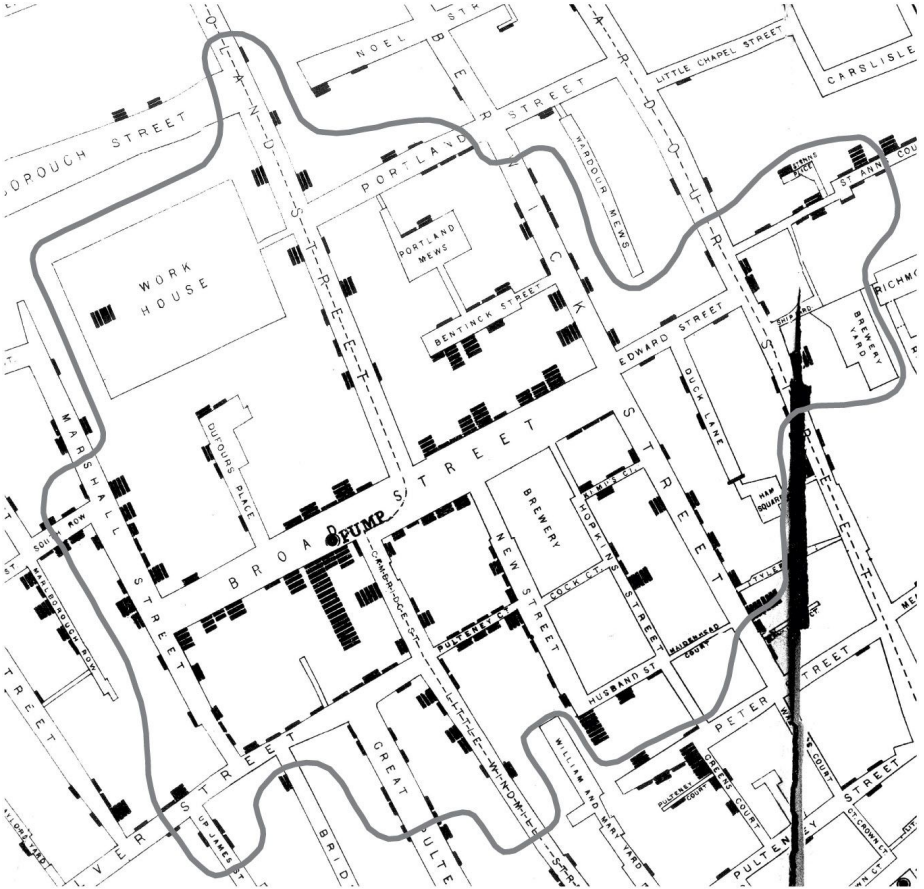
Breakout time

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

Pattern Detection

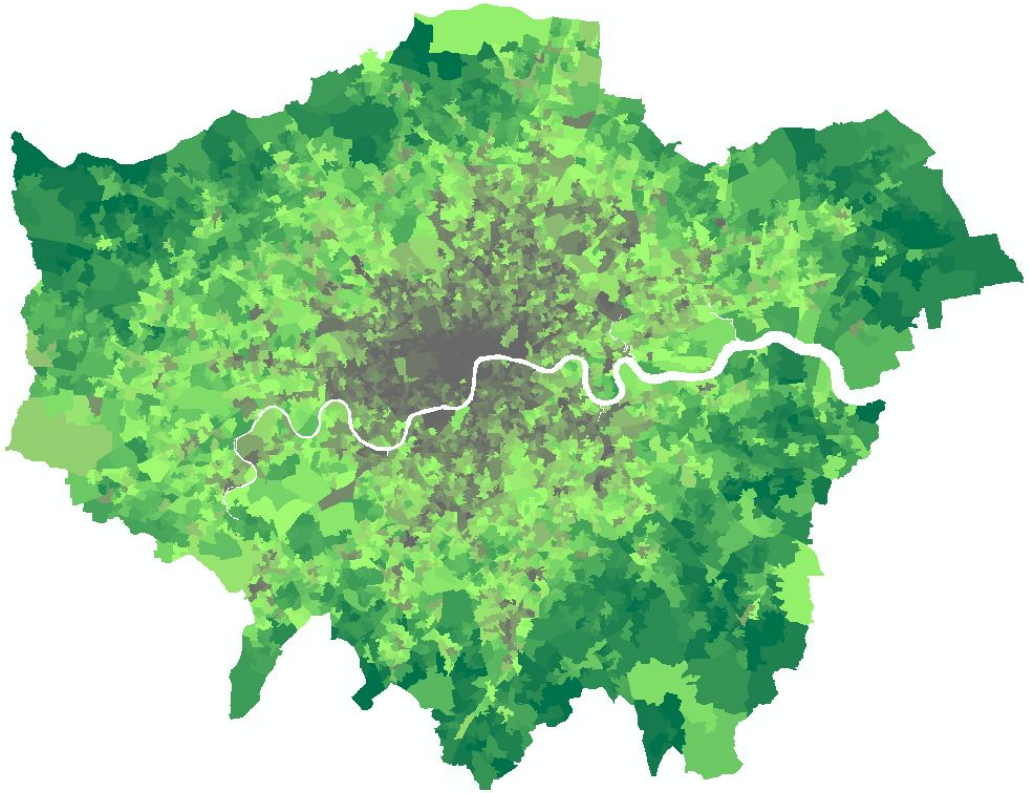
The “First GIS”:

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



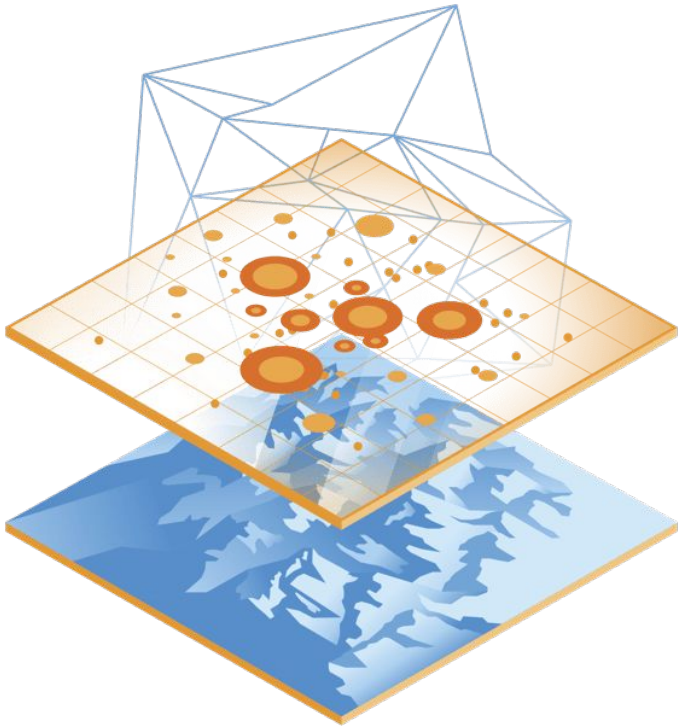
Pattern Detection

London air pollution...



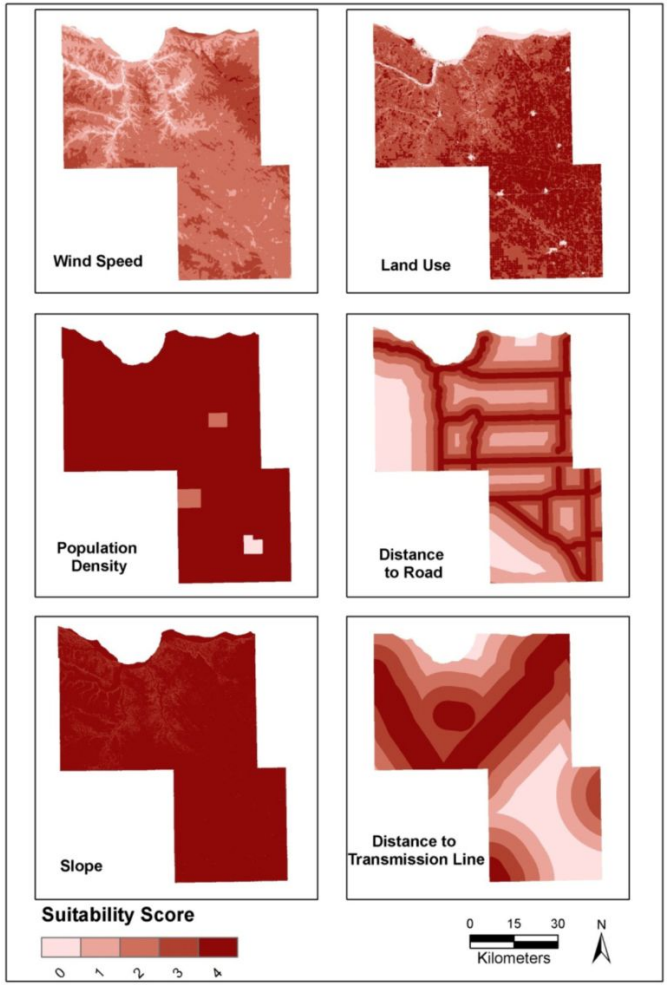
Informing decision making

- 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

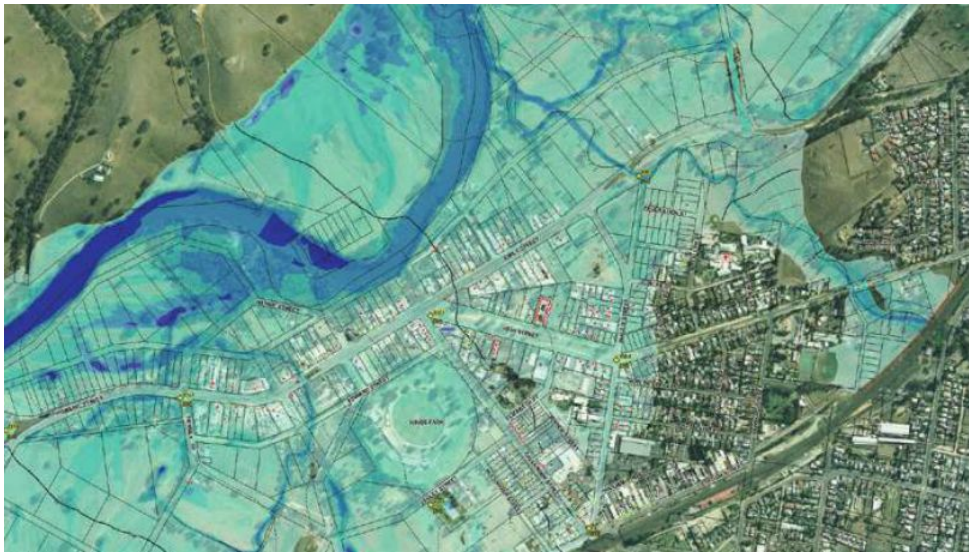


Network Analysis

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

Risk Mitigation

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

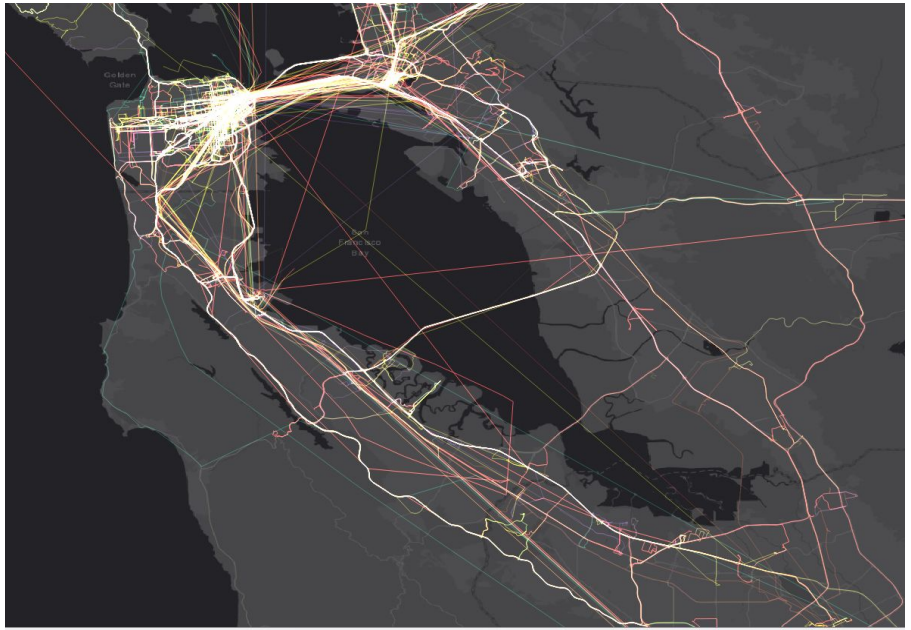


Risk Mitigation

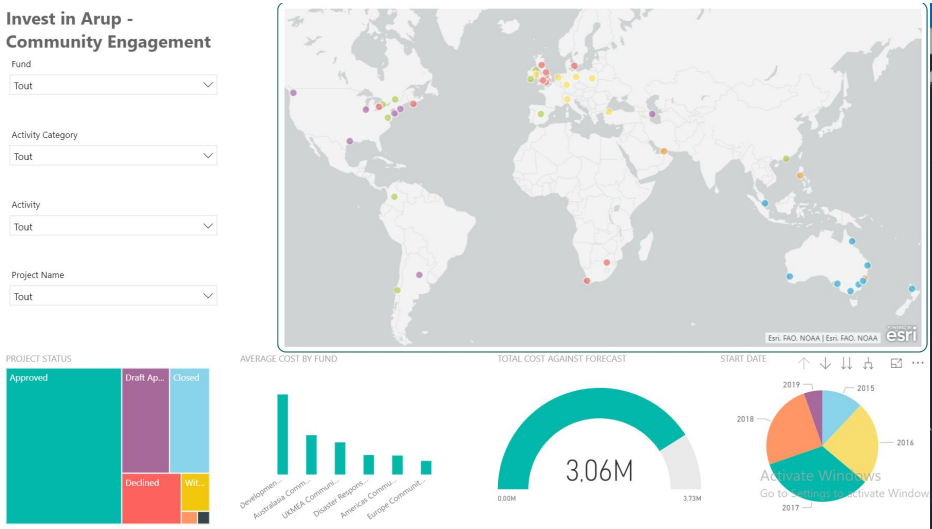
- Assessing Flood-related economic damages
- **Understanding vulnerability to climate change in Nepal**

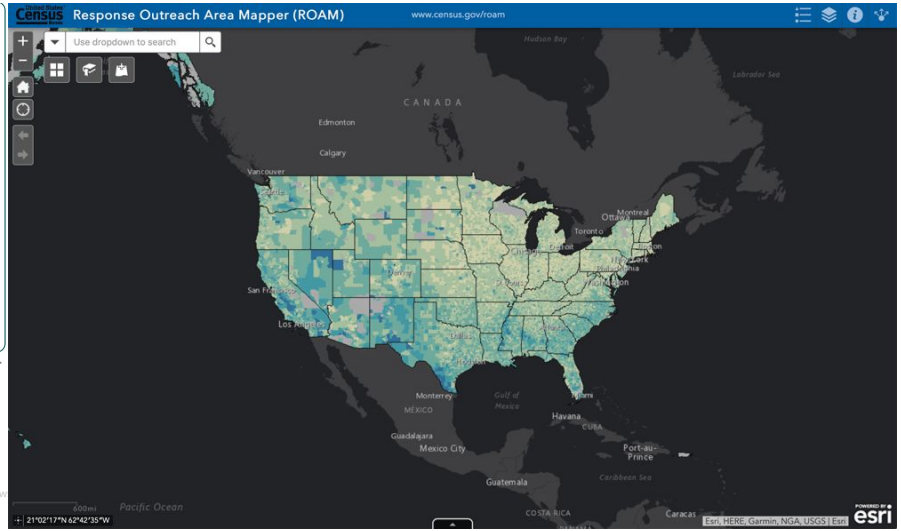
Mobility / Transport planning

Understanding urban mobility patterns



Insights (Dashboards, Webmaps etc)





A typical GIS Workflow

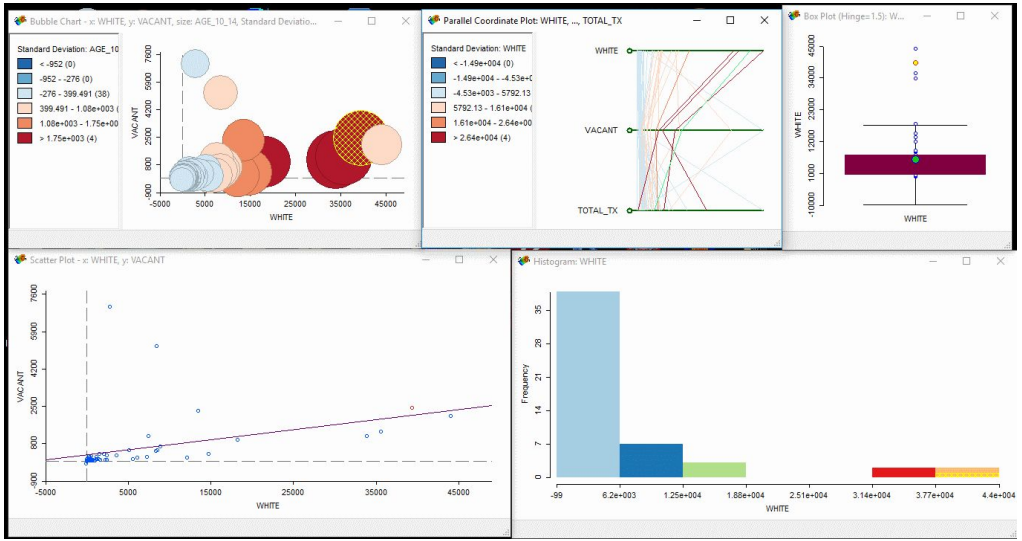
1. Requirements gathering

2. Finding the DATA

3. Exploratory Spatial Data Analysis (ESDA)

3. Exploratory Spatial Data Analysis (ESDA)

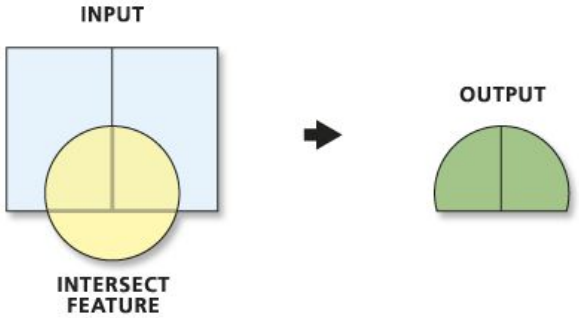
- **GeoDA** https://geodacenter.github.io/documentation.html

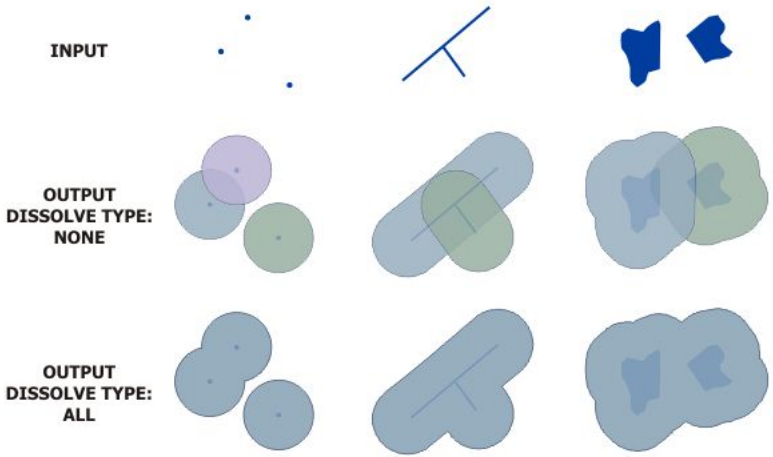


4. Refining the research objective and goals

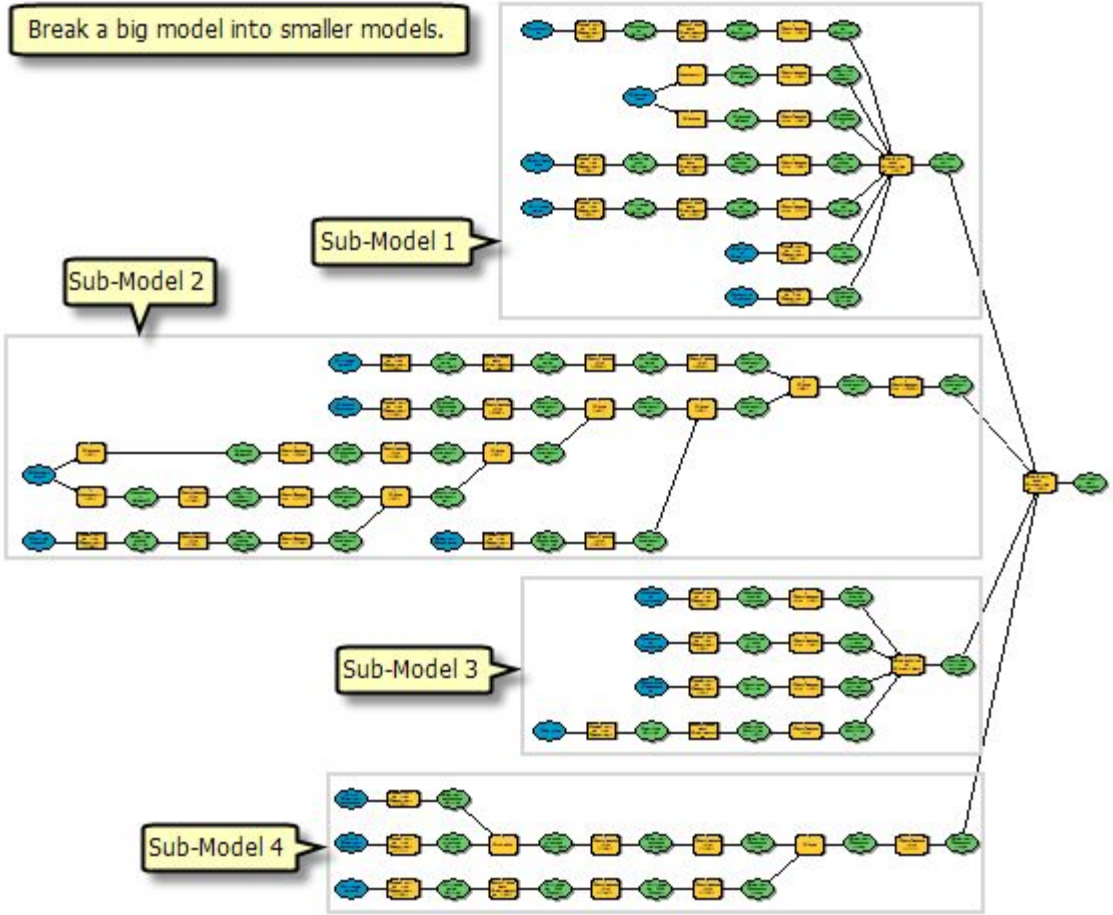
5. Reshaping your data / feature generation

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

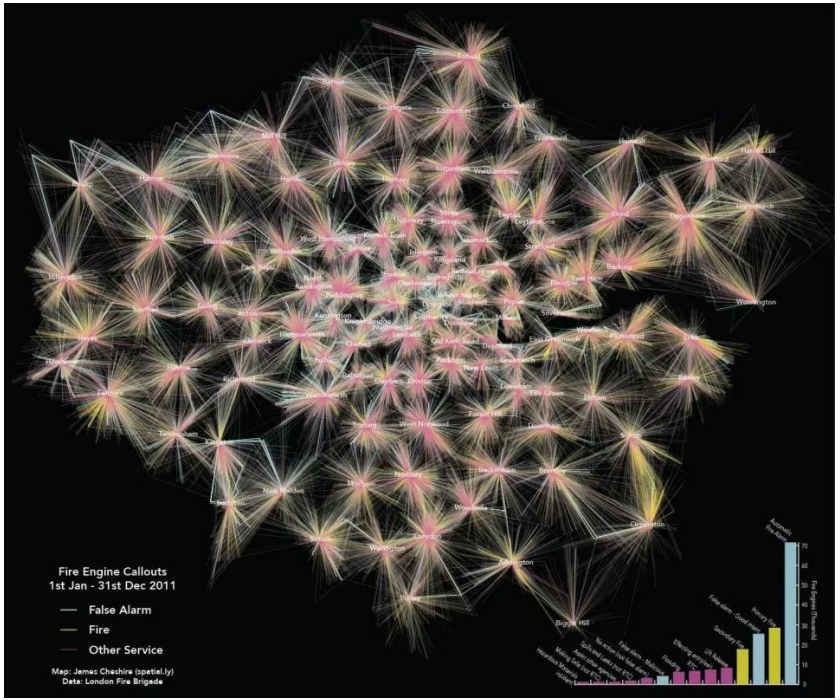


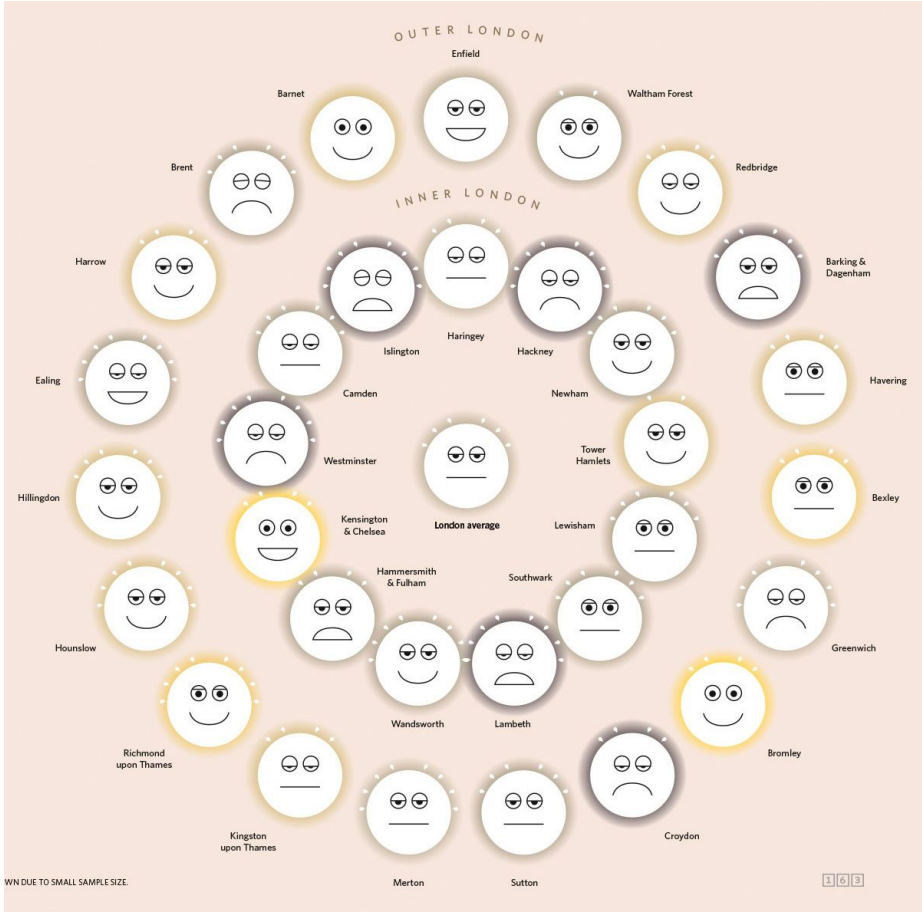


6. Analysis / Modeling


Break a big model into smaller models.
Sub-Model 1
Sub-Model 2
Sub-Model 3
Sub-Model 4

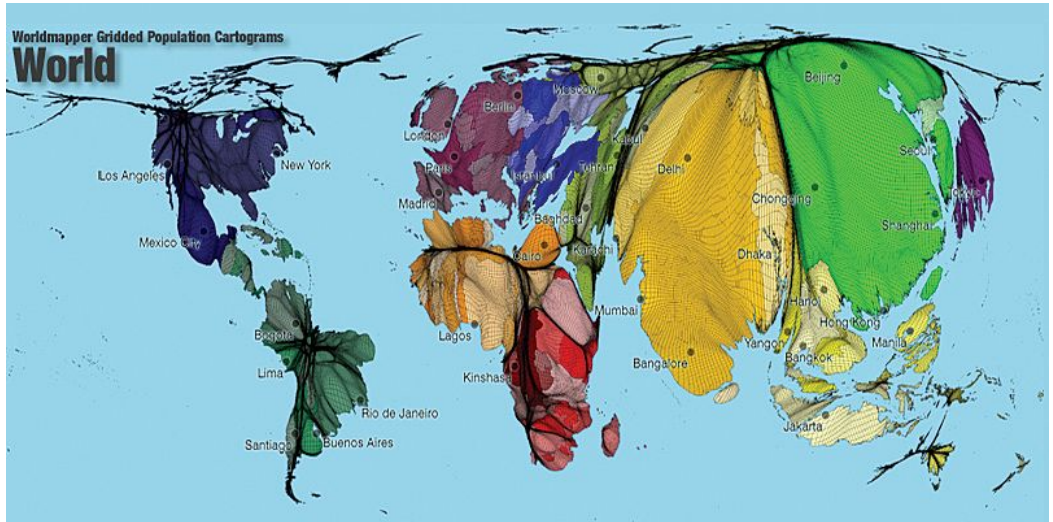
7. Visualisation

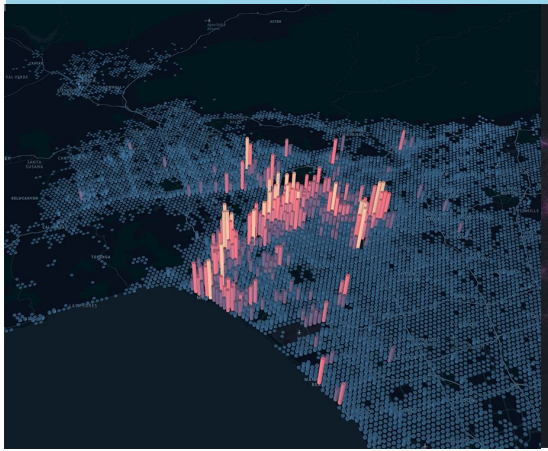


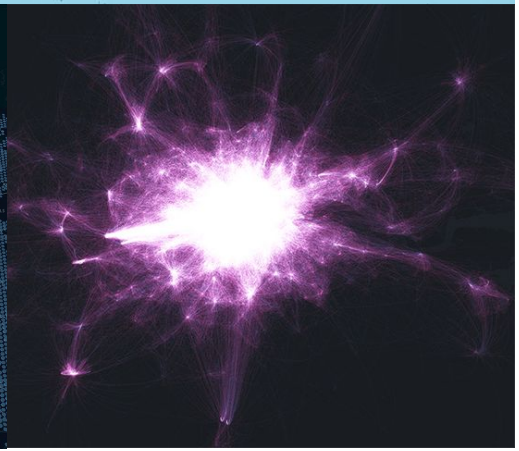



Liberty Canyon Corridor
Zoo
Hollywood Sign
SANTA MONICA MOUNTAINS NAT. REC. AREA
LOS ANGELES
Malibu
Santa Monica
Los Angeles International Airport
San Bernardino
SAN BERNARDINO NATIONAL FOREST
SAN BERNARDINO
RIVERSIDE
Plans are underway for a vegetated overpass across Highway 101. This will give mountain lions in the Santa Monica Mountains an escape route to the Simi Hills.
CHINO HILLS STATE PARK
L.A.
ORANGE
Santa Ana R.
Riverside
Anaheim
Coal Canyon Corridor
Irvine Ranch
SANTA ANA MTS.
Long Beach
Santa Ana
M122
1 June
Irvine
Mission Viejo
F50
CLEVELAND NAT. FOREST
29 Apr.
Murrieta
Temecula
25 Mar.
Palomar Linkage
PACIFIC OCEAN
Laguna Beach
M56
7 March 2010
San Clemente
CAMP PENDLETON MARINE CORPS BASE
Santa Margarita R.
Fallbrook
PALOMAR MT.
22 Mar.
3 April
Oceanside
CLEVELAND NAT. FOREST
7-13 Apr.
Escondido
Julian
Ramona
SAN DIEGO
TIJUANA
Tecate
UNITED STATES
MEXICO
SAN JACINTO MTS.
23 Jul.
Palm Springs
SAN BERNARDINO NATIONAL FOREST
5 Jul.
SANTA ROSA AND SAN JACINTO MTS. NAT. MONUMENT
RIVERSIDE
SAN DIEGO
PENINSULAR RANGES
SANTA ROSA MTS.
SAN DIEGO
IMPERIAL
RIVERSIDE
IMPERIAL
SALTON SEA
25 Jun. 2004
21 August
F18
M53
ANZA-BORREGO DESERT STATE PARK
10 May 2009
LAGUNA MTS.
30 November
CLEVELAND NAT. FOREST
M56 shot
28 Apr.
Jacumba
Parque-to-Park Linkage
25 Apr.
9 June
La Rumorosa
Laguna Salada
28 July
Mountain lions are wide-ranging animals. In Southern California, freeways restrict their movements. Here we show five of the 74 cats that researchers have tracked since 2001. Only one has ever crossed Interstate 15.
M56
On 7 March 2010, this young male left the Santa Ana Mountains in search of a new home territory. He travelled south along the beach and up Route 76 before crossing Interstate 15 to explore the Peninsular Ranges.
In late April, M56 killed eight sheep in Jatapul Valley. The landowner obtained a depredation permit. On the night of April 28, a trapper captured and shot M56.
M122
Adult males will kill or threaten males that do not leave their home ranges. M122 was born near Murrieta and dispersed to claim territory in the northwestern foothills.
F50
F50 was living with her offspring in the western Santa Ana Mountains until she was hit by a car on Highway 74 near where her daughter F62 and a male kitten were also hit.
F18
Typically, females don't disperse as far as males. F18 is an exception. She walked 100 kilometres to the San Jacintos near Palm Springs and then came back to reside in San Diego County.
M53
In summer 2009, this young male crossed into Mexico via the Parque-to-Park Linkage at least three times. He roamed 69 kilometres south of the border before he was struck and killed by a car in Mexico.
GPS tracks of mountain lions 2004-2015
Home range
Wildlife corridor
Existing
Proposed
Successful highway crossing
0
25 km
SOURCES: T. WINSTON VICKERS AND WALTER BOYCE, UNIVERSITY OF CALIFORNIA, DAVIS; BRIAN COHEN, THE NATURE CONSERVANCY; SRTM; USGS


UNITED STATES
DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY
WYOMING
(TETON COUNTY)
GRAND TETON QUADRANGLE
GRAND TETON
NATIONAL
PARK
TETON RANGE
MOUNTAINS
JACKSON LAKE
SNAKE RIVER
Main roads, names, and other information shown in red added by National Park Service, 1948
CONTOUR INTERVAL 100 FEET
FOR SALE BY U.S. GEOLOGICAL SURVEY, FEDERAL CENTER, DENVER, COLORADO OR WASHINGTON 25, D.C.


Worldmapper Gridded Population Cartograms
World
Los Angeles
New York
Mexico City
Bogota
Lima
Rio de Janeiro
Santiago
Buenos Aires
London
Paris
Madrid
Berlin
Moscow
Istanbul
Tehran
Kabul
Baghdad
Cairo
Karachi
Delhi
Mumbai
Dhaka
Bangalore
Lagos
Kinshasa
Beijing
Seoul
Chongqing
Shanghai
Hanoi
Hong Kong
Manila
Yangon
Bangkok
Jakarta





Tutorial

Homework

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