Introduction to GIS

Raphaëlle ROFFO

Sciences Po - GETEC Masters
Fall Semester 2021

Session 1

Course overview

&

What is GIS ?

Today’s plan

Attendance

Student representative election

Technical setup

Technical setup

- Slides, tutorial material and sessions recordings are available on the **GitHub** after each session.
- 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!**
- You will need to **download QGIS** (tutorial coming up).
- Room J307 (13 rue de l'Université) is a **computer room** and is booked for you **Wednesdays 08.00-12.30**. 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).

Sessions

The course will be run in 6 sessions:

1. Overview of GIS: GIS as a field of research, typical workflows
2. Sourcing and loading data into GIS
3. Visualising spatial data
4. Working with vector data
5. Processing raster data
6. Advanced GIS + coursework

Sessions

This semester being taught online, each session will last 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

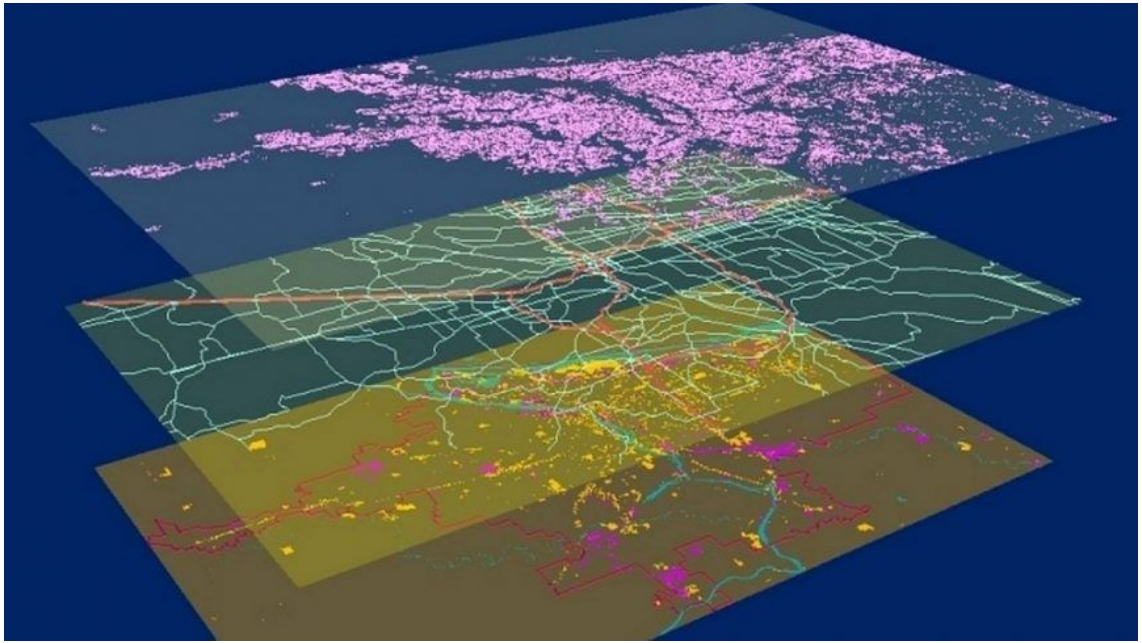
Coursework

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.
(full guidelines will be provided on session 4)

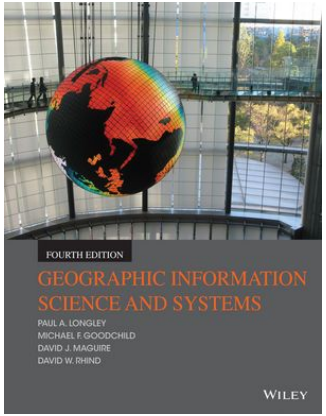
Deadline for submission: Wednesday 24th November, 23:59.

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.

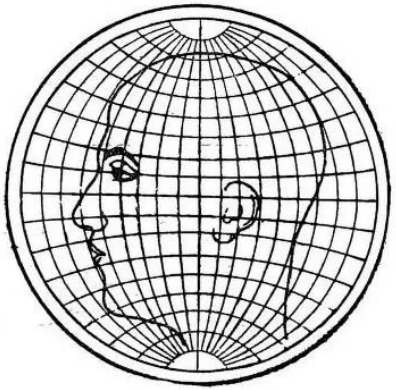


FIG. 42.—Man's head drawn on globular projection.

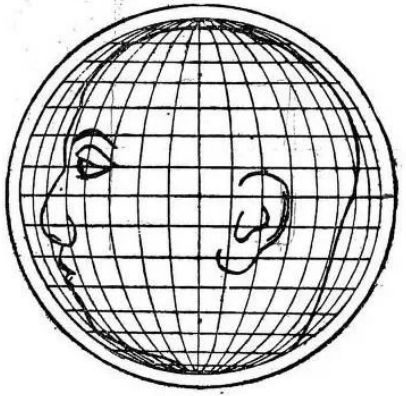


FIG. 43.—Man's head plotted on orthographic projection.

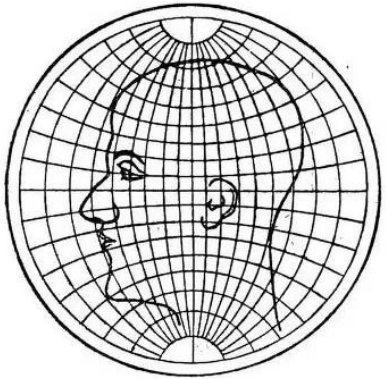


FIG. 44.—Man's head plotted on stereographic projection.

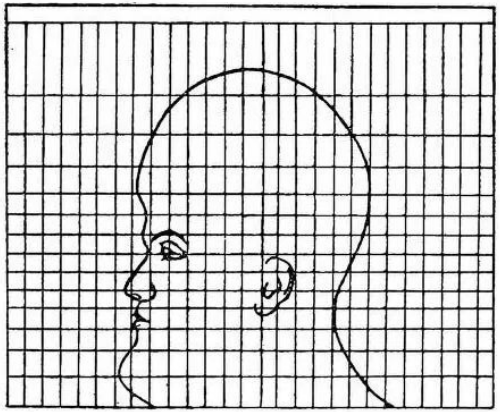
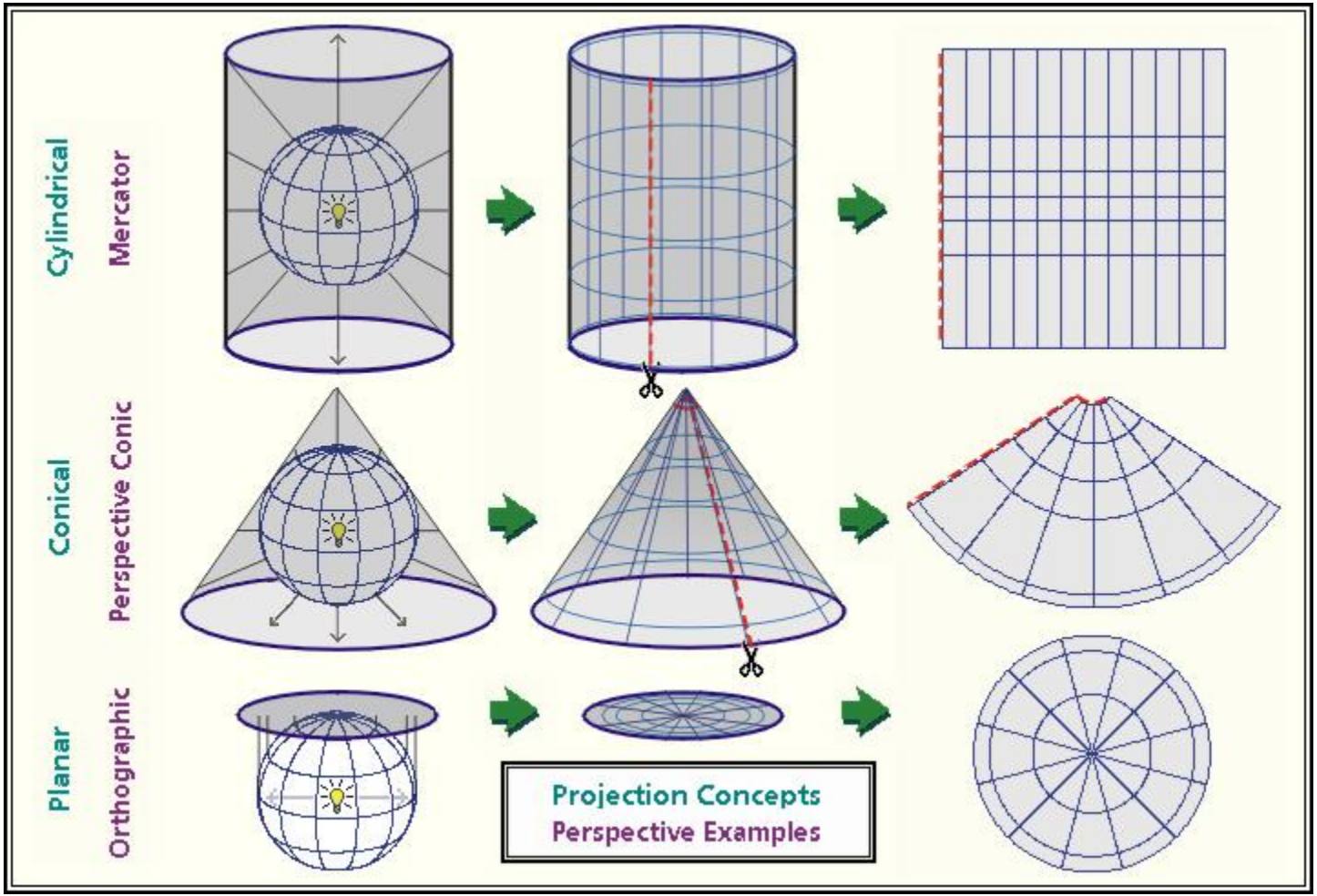
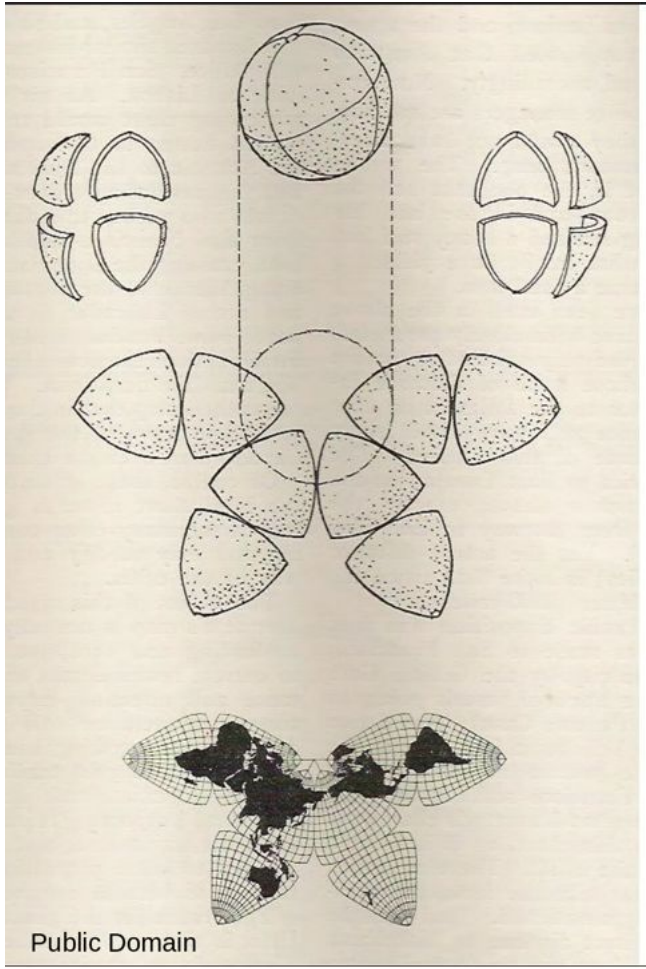
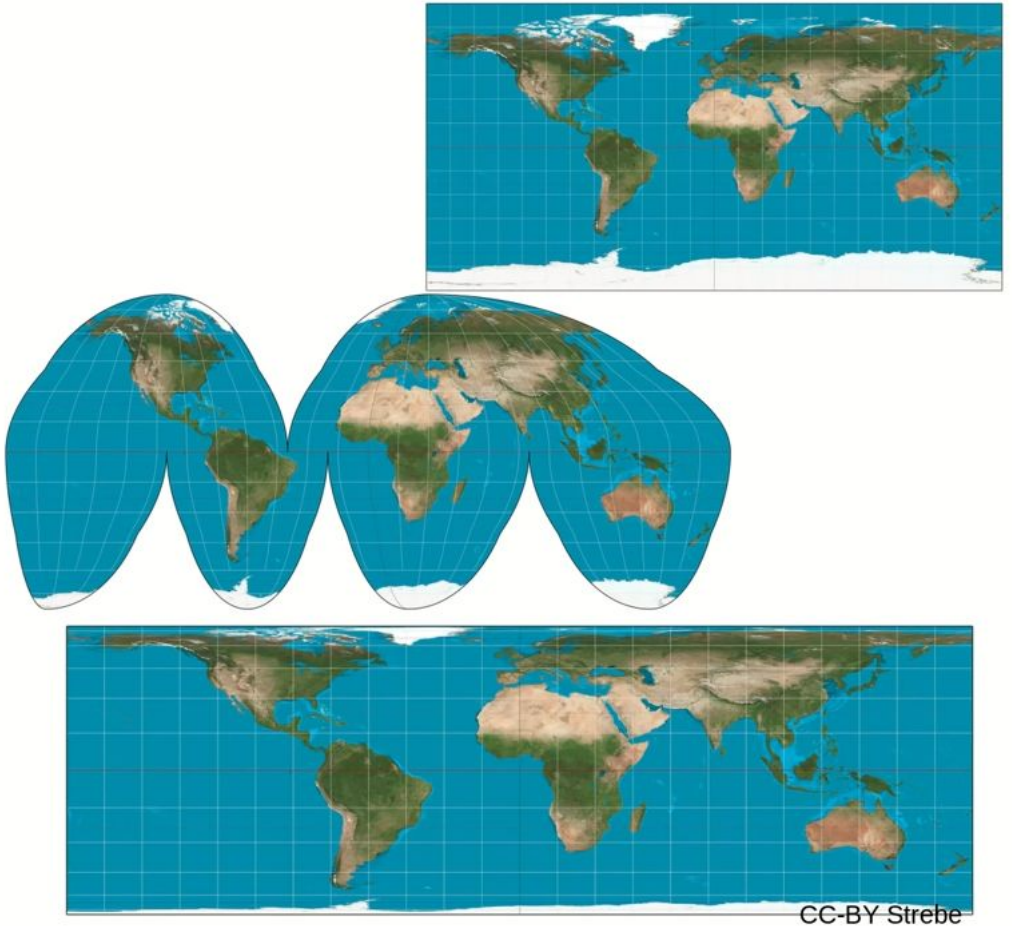


FIG. 45.—Man's head plotted on Mercator projection.

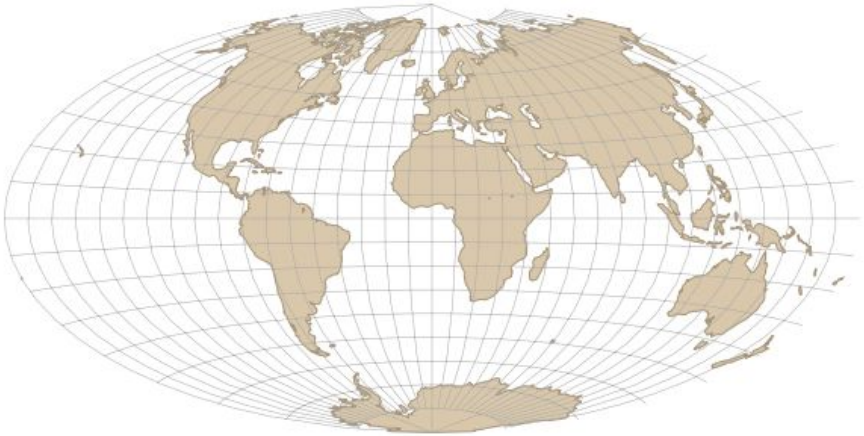

Cylindrical
Mercator
Conical
Perspective Conic
Planar
Orthographic
Projection Concepts
Perspective Examples

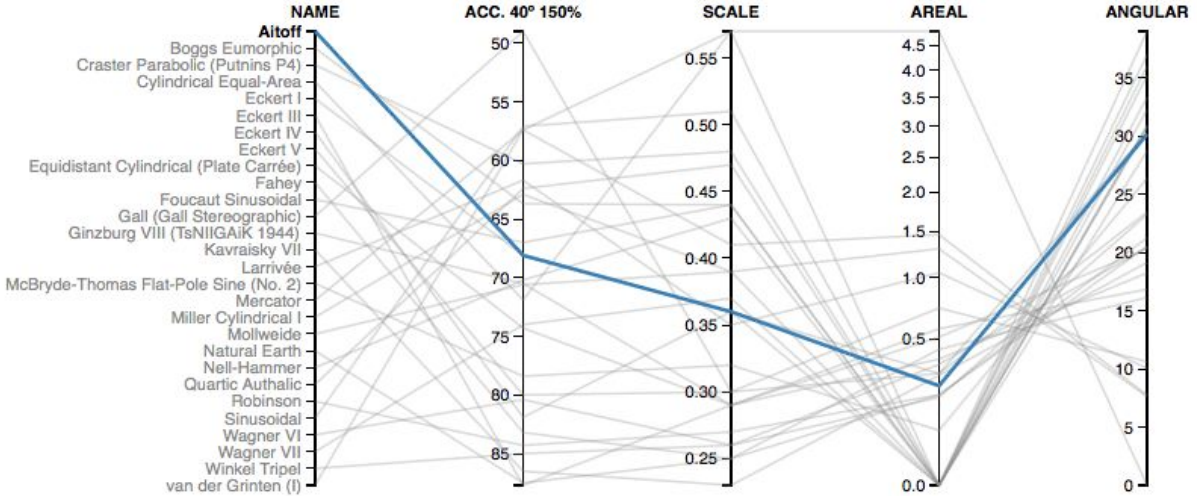


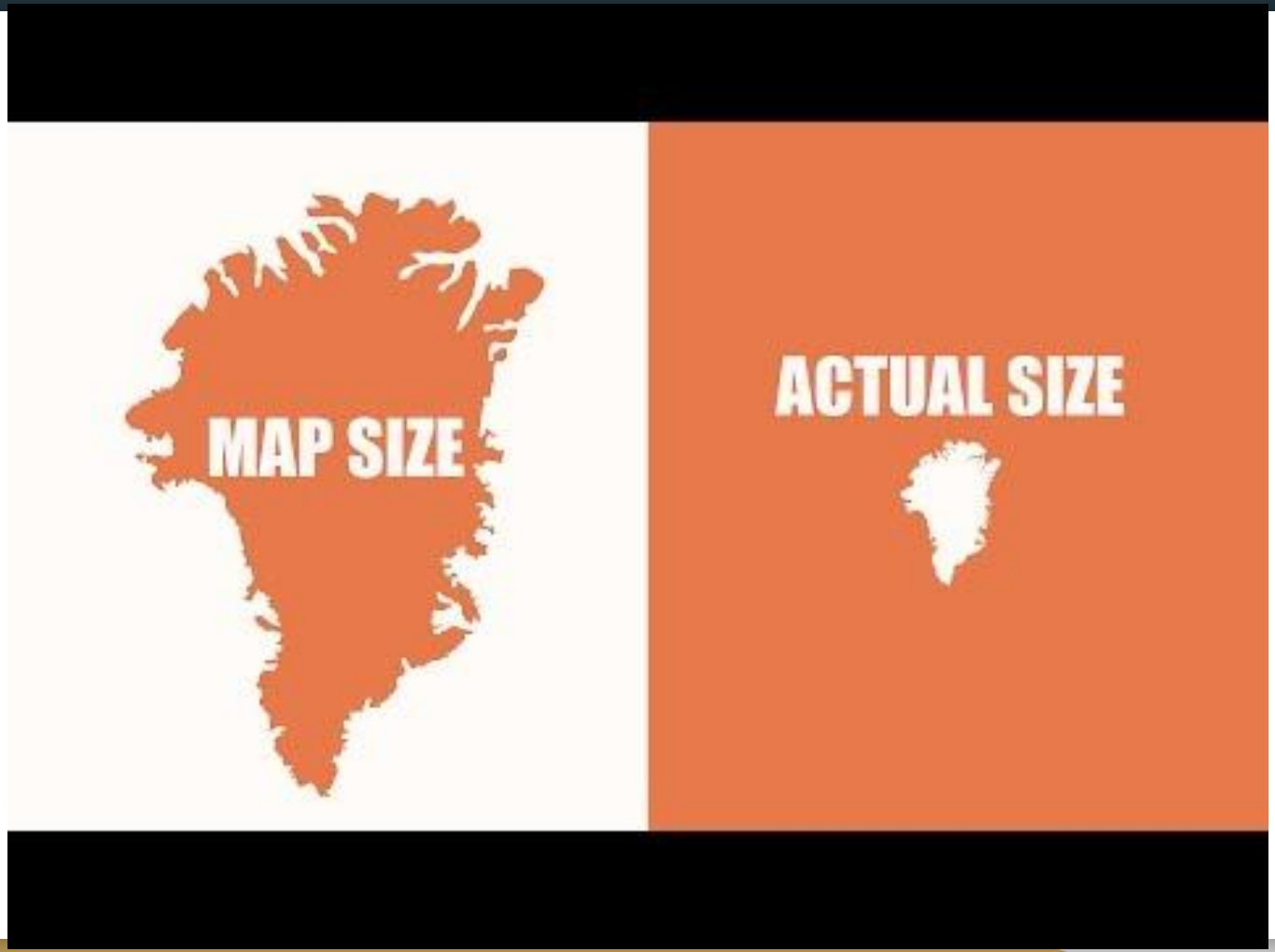
Public Domain



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
$n_{BW} = 112$
$n_{BB} = 0$
$n_{WW} = 0$

Dispersed

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

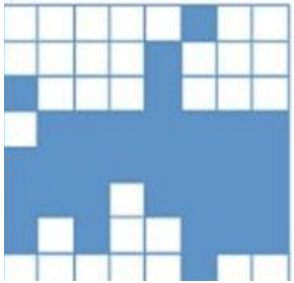
(B)

Spatial Independence

(C)

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

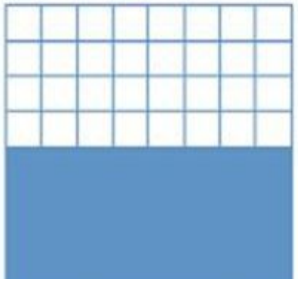
(D) Spatial Clustering



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

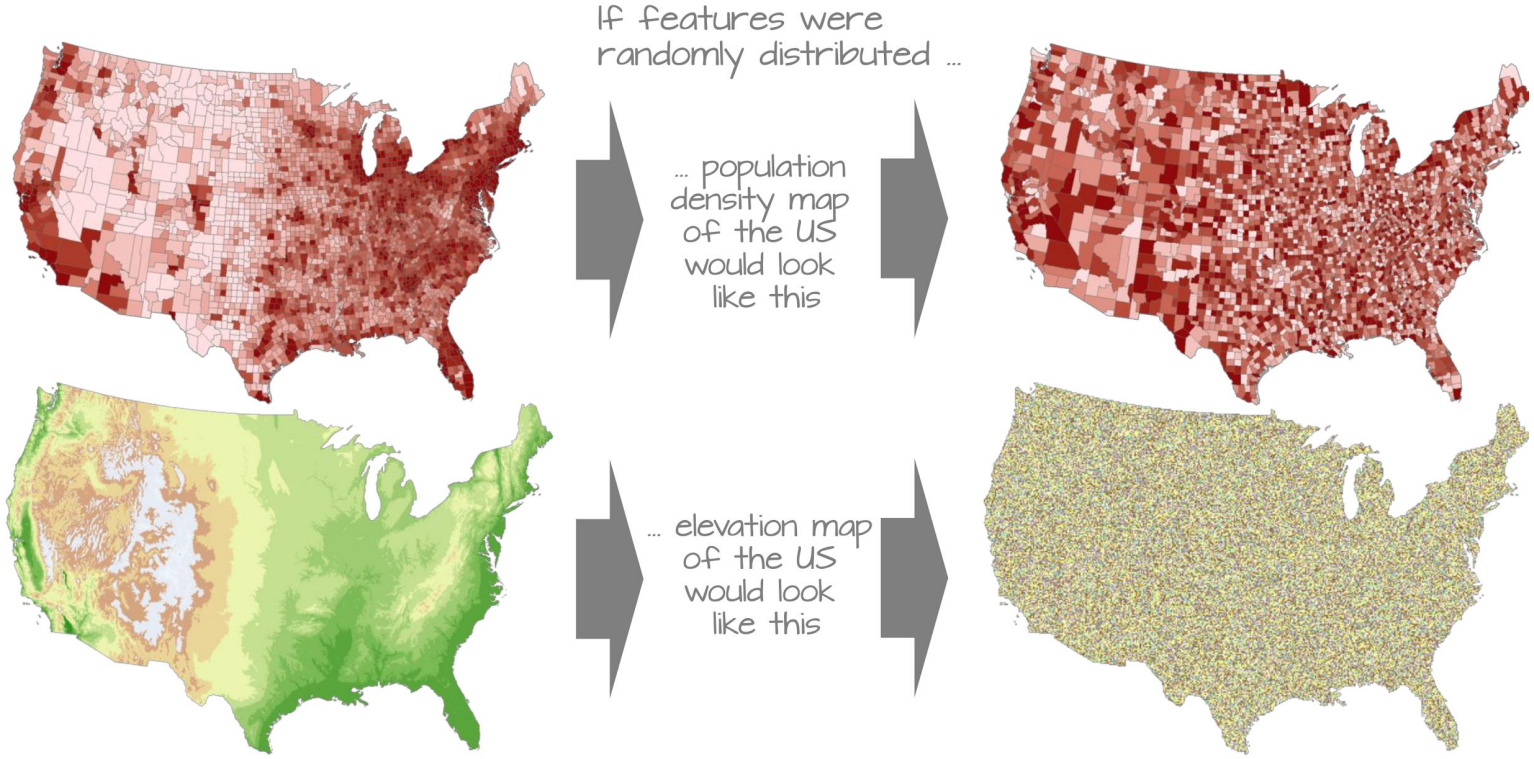
Positive

(E)



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

Spatial analysis

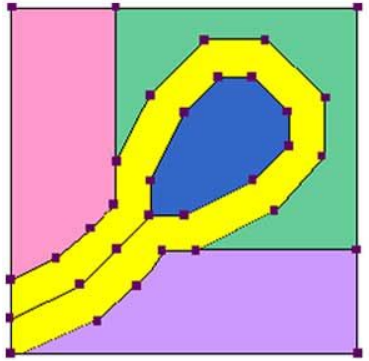


Spatial is special…

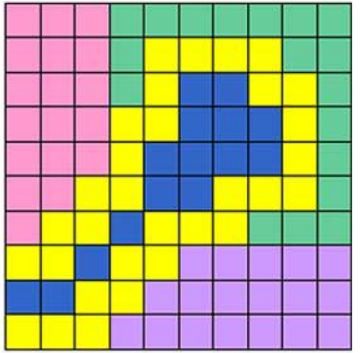
…. and can be represented in two ways:

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

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



Vector



Raster

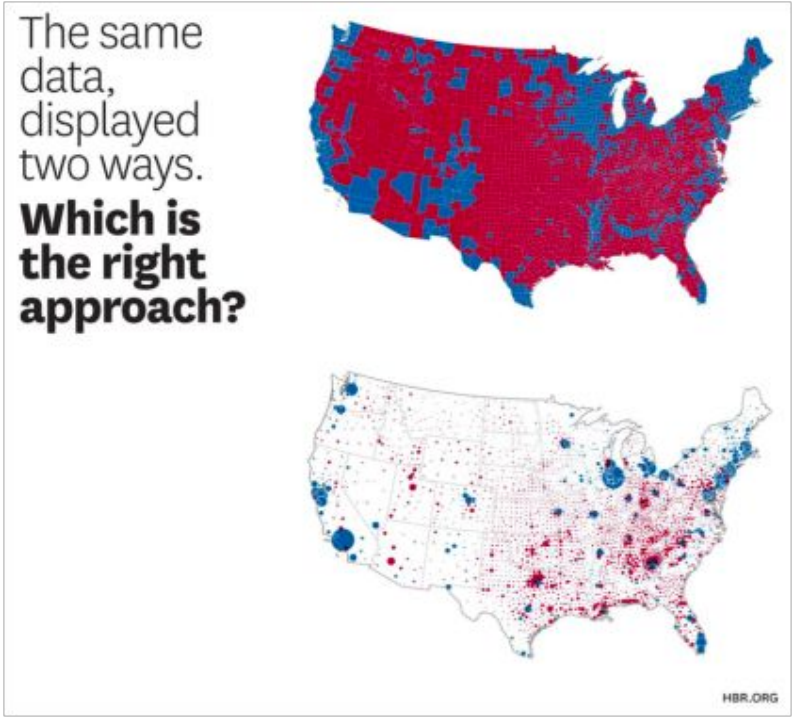
Spatial is special…

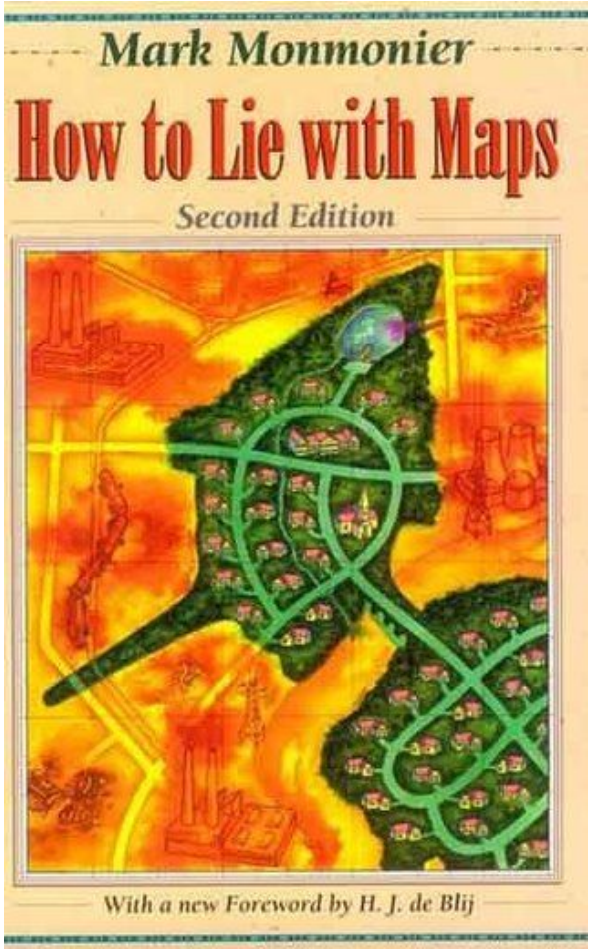
3. … because of cartography

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

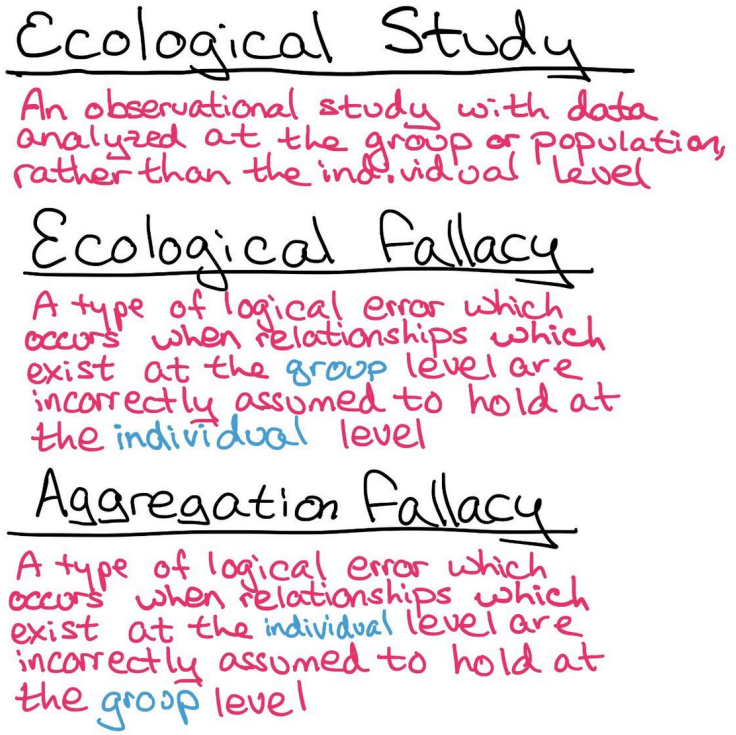




Ecological fallacy & Aggregation fallacy

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.

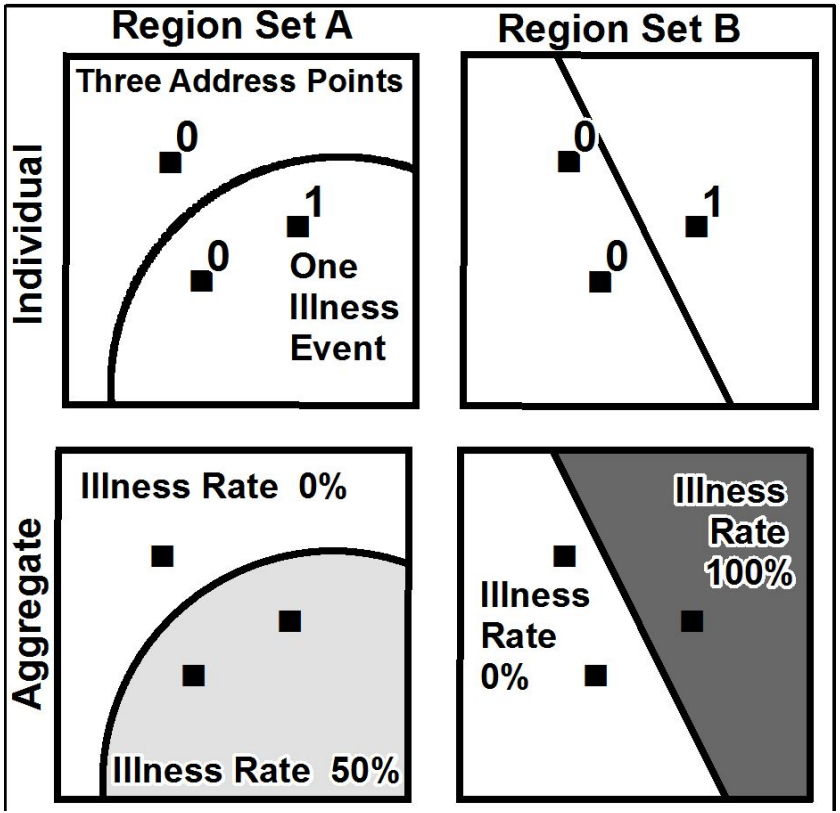
But not a single department was showing to be significantly biased against women.

How is that possible?

Credit: Nick Bearman

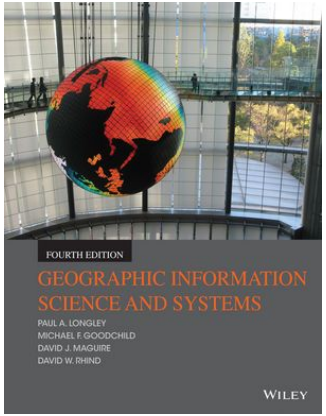
Modifiable Areal Unit Problem (MAUP)

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 (shapely, geopandas, folium, rasterio, GDAL, sckikit-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**

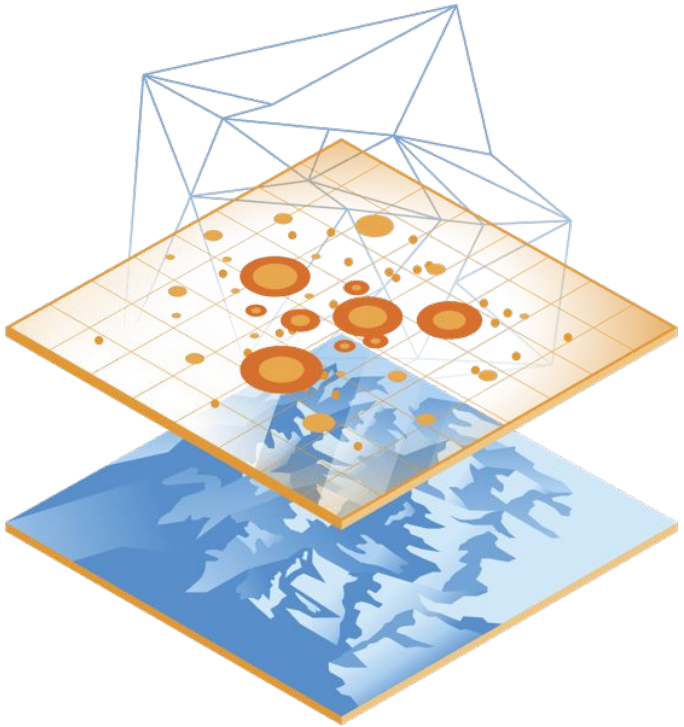
Use Cases

Breakout time

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

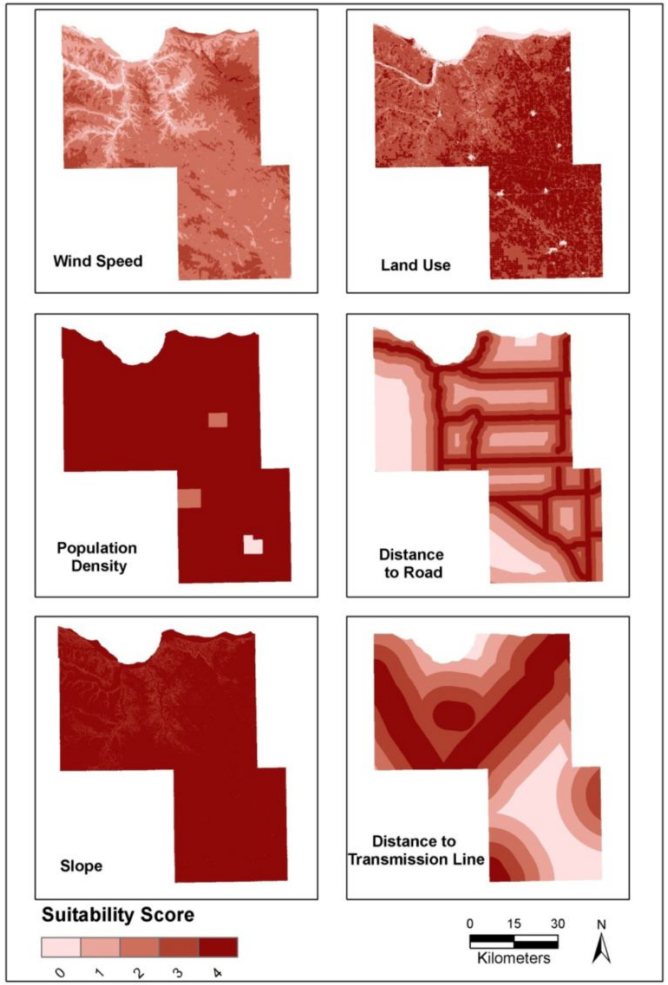
Informing decision making

- 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

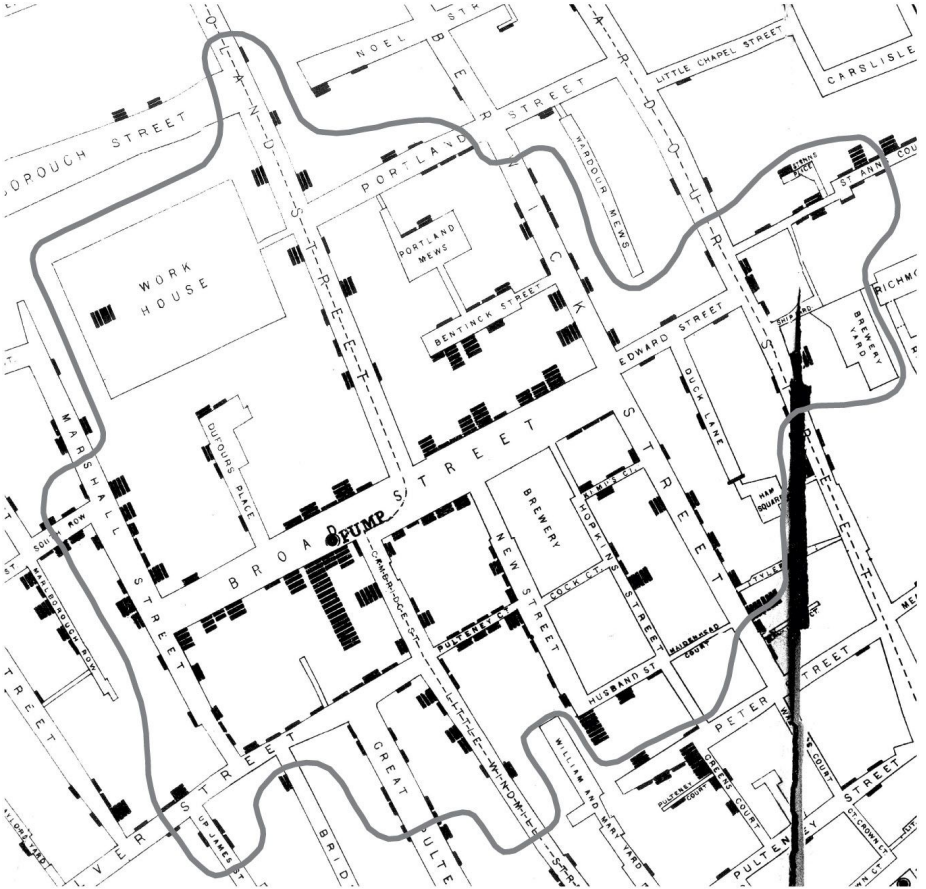


Network Analysis

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

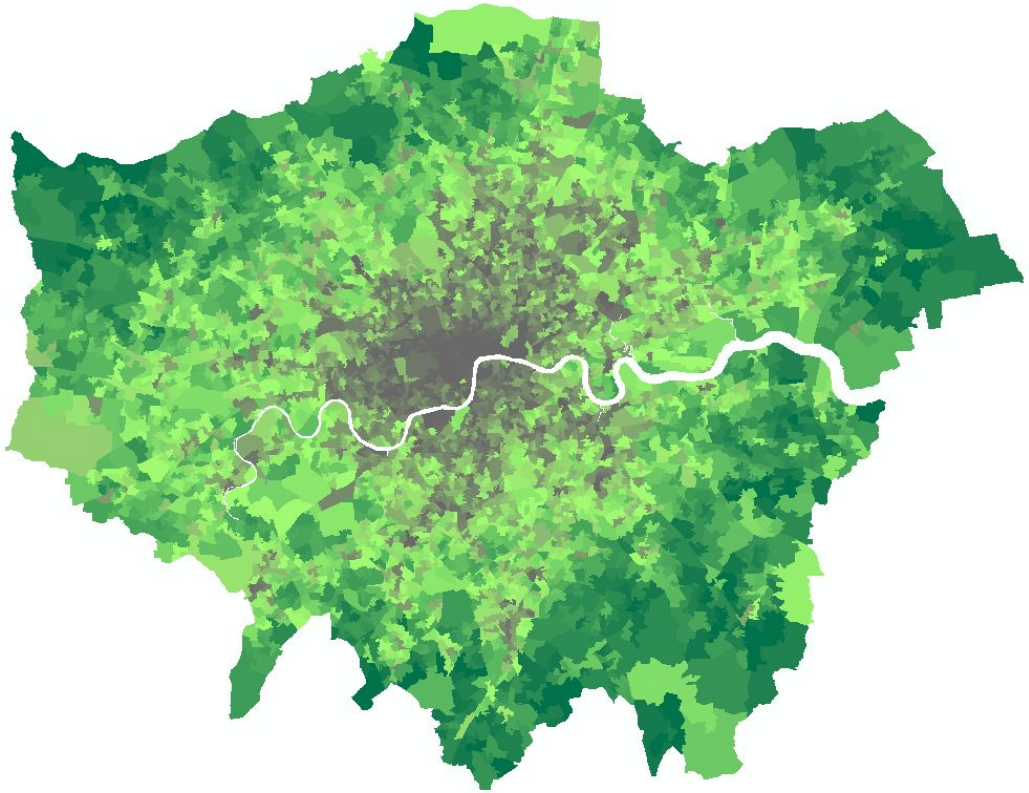
Pattern Detection

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



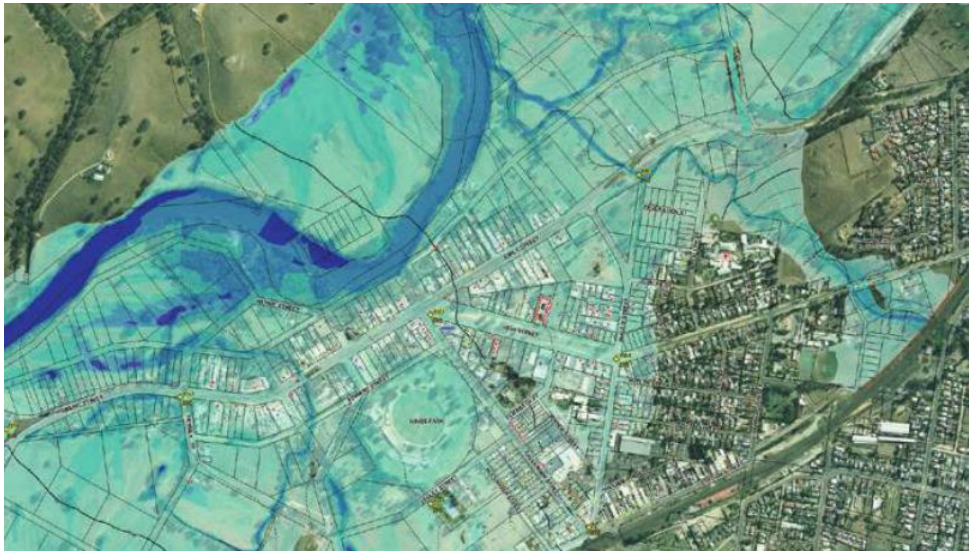
Pattern Detection

London air pollution...



Risk Mitigation

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



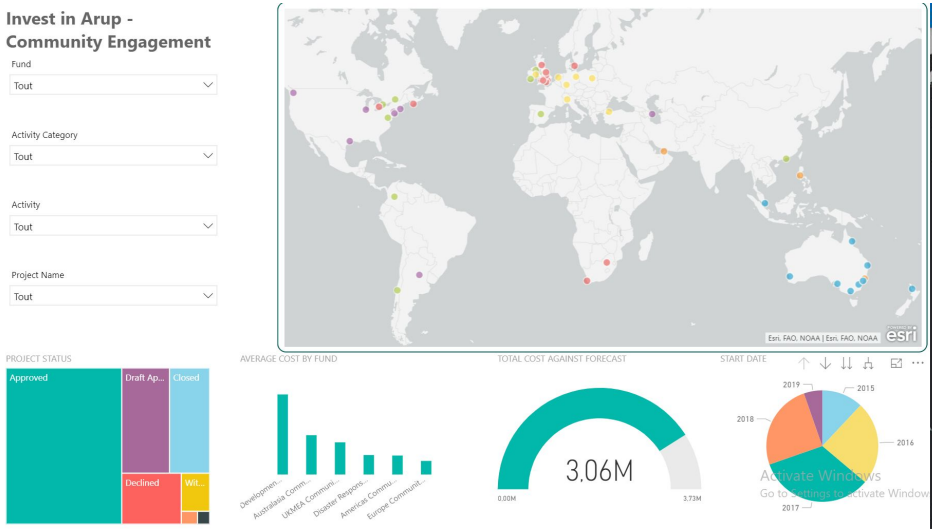
Risk Mitigation

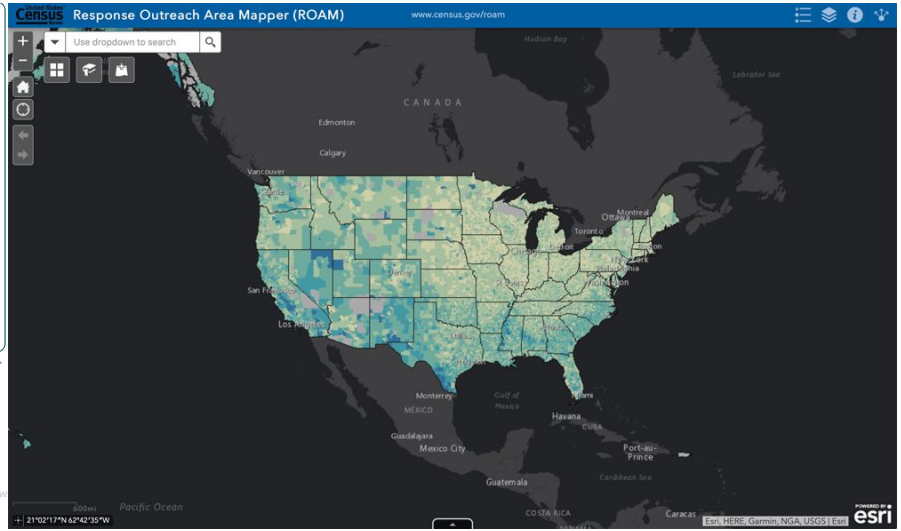
- Assessing Flood-related economic damages
- **Understanding vulnerability to climate change in Nepal**

Mobility / Transport planning

Understanding urban mobility patterns

Insights (Dashboards, Webmaps etc)





A typical GIS Workflow

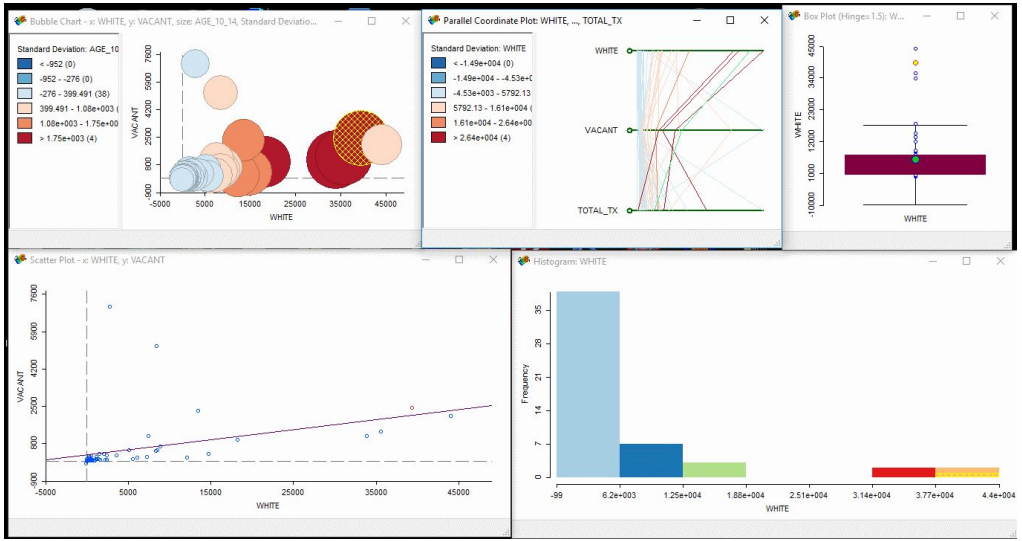
1. Requirements gathering

2. Finding the DATA

3. Exploratory Spatial Data Analysis (ESDA)

3. Exploratory Spatial Data Analysis (ESDA)

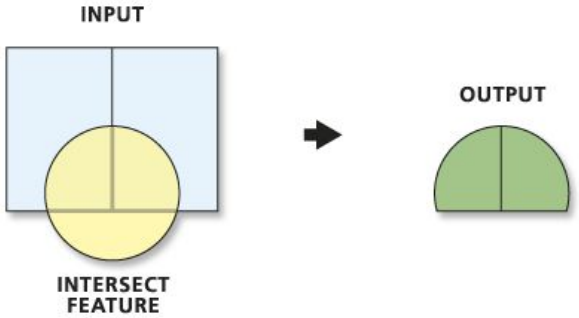
- **GeoDA** https://geodacenter.github.io/documentation.html

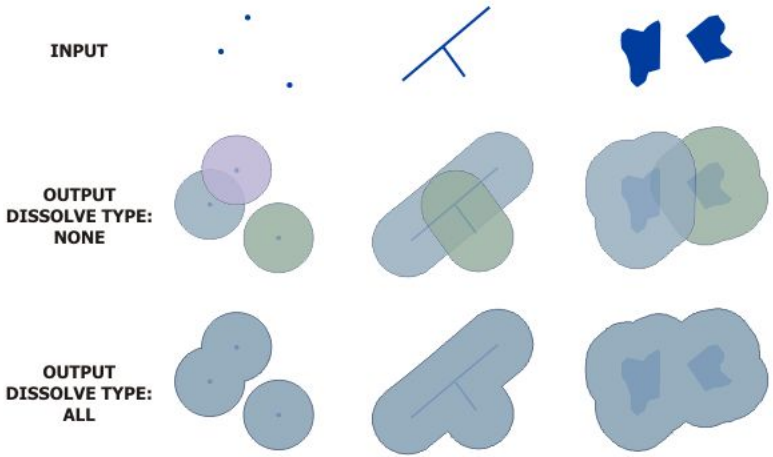


4. Refining the research objective and goals

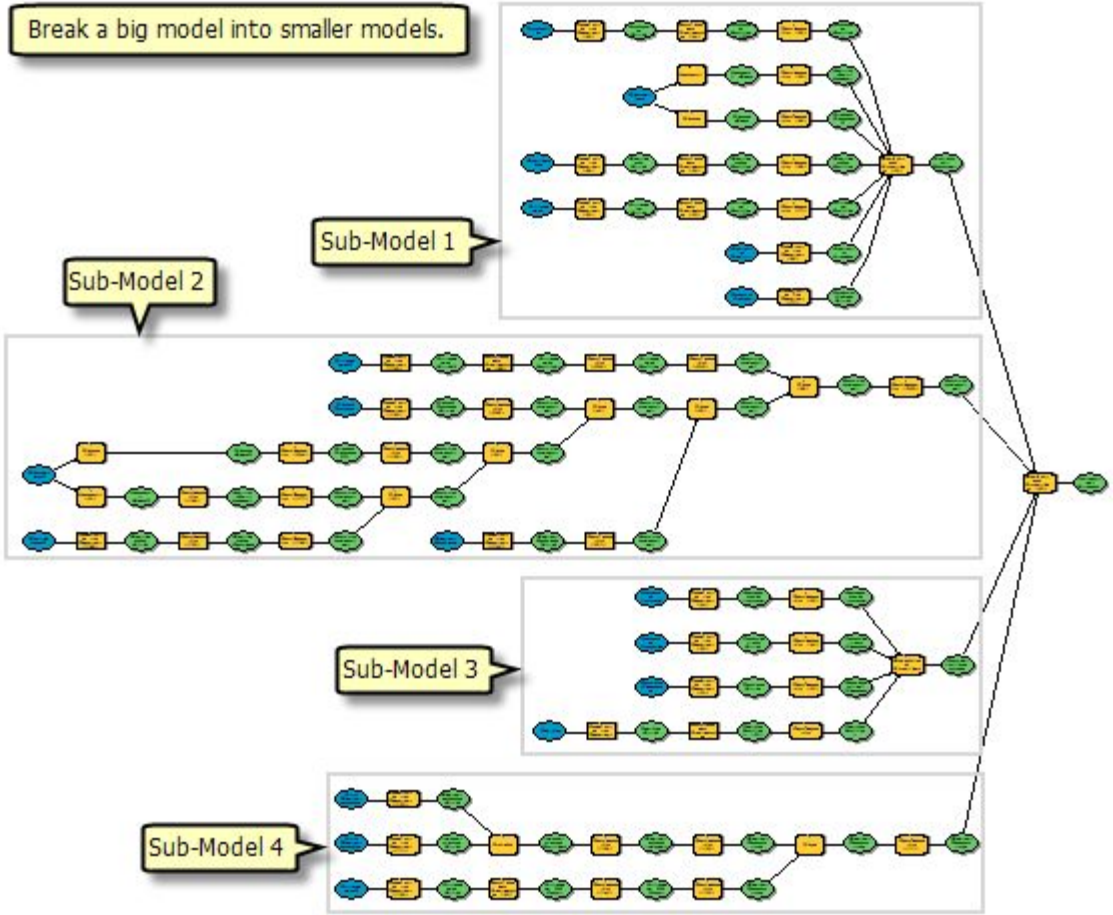
5. Reshaping your data / feature generation

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

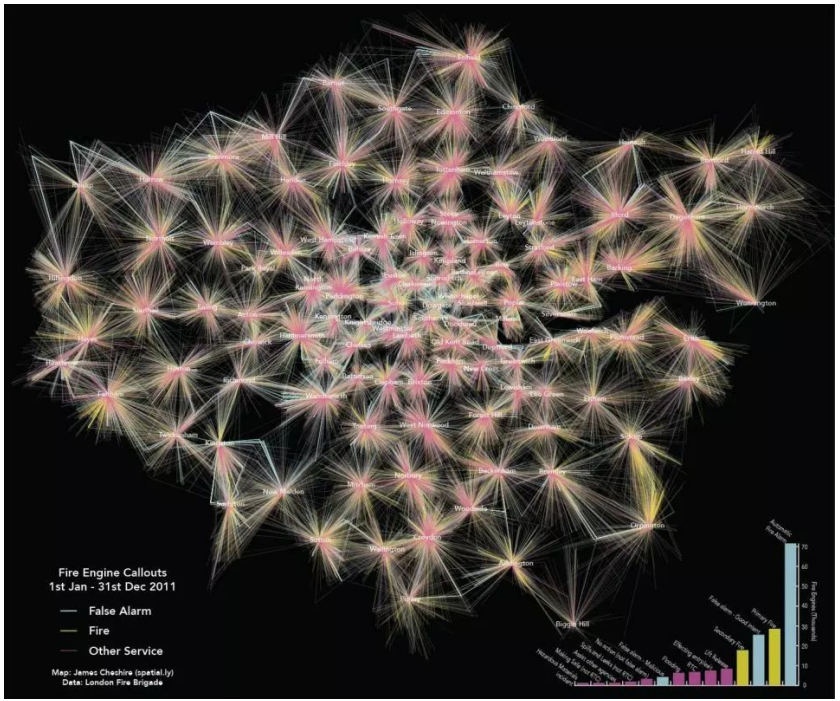


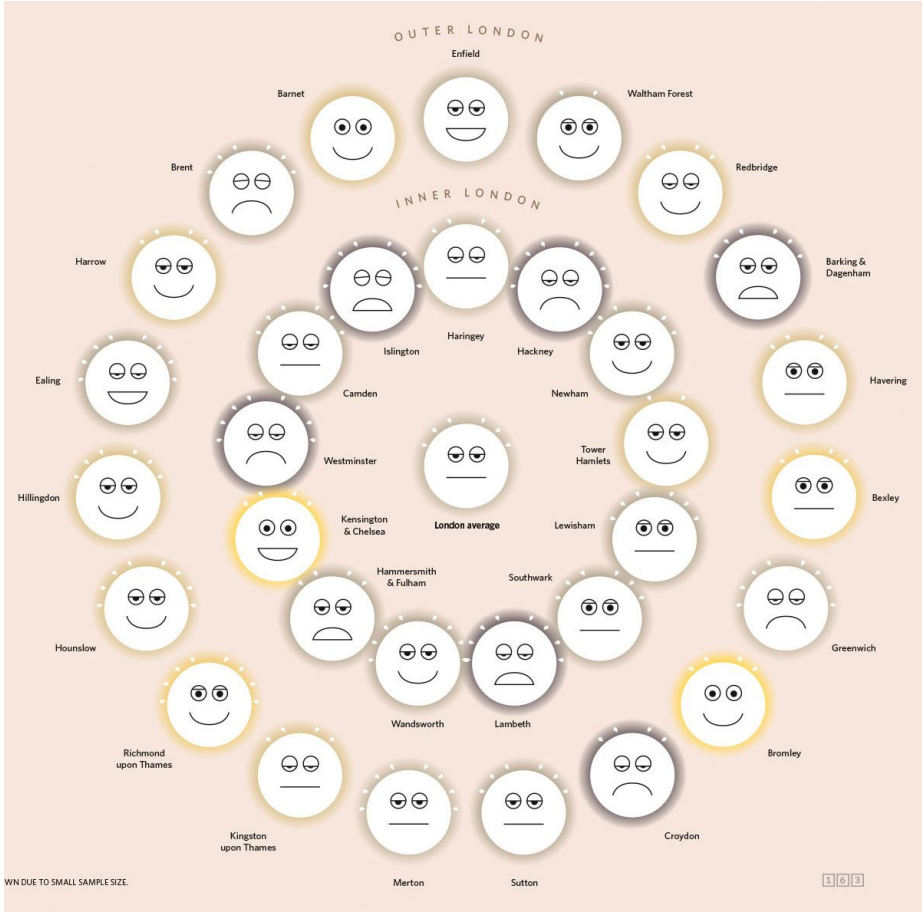


6. Analysis / Modeling


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

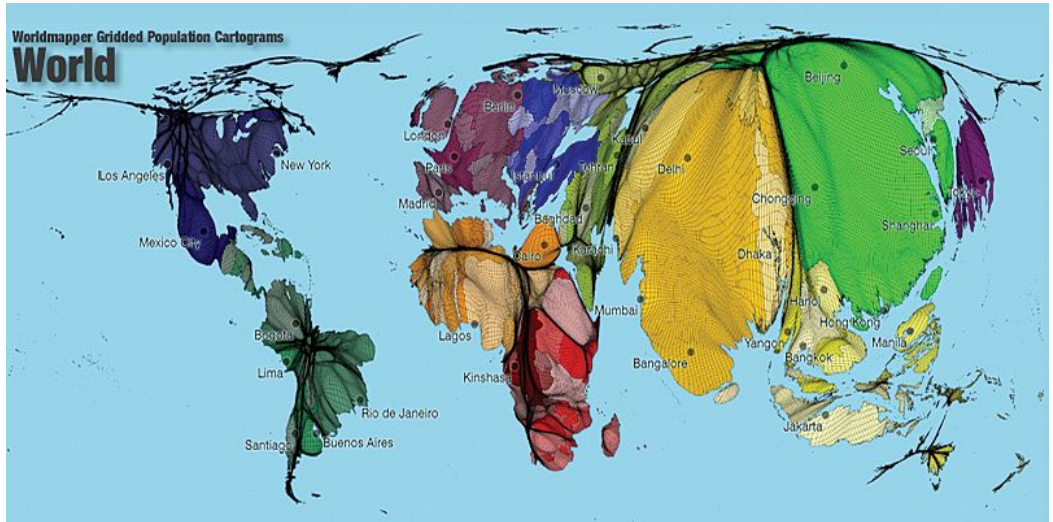
7. Visualisation

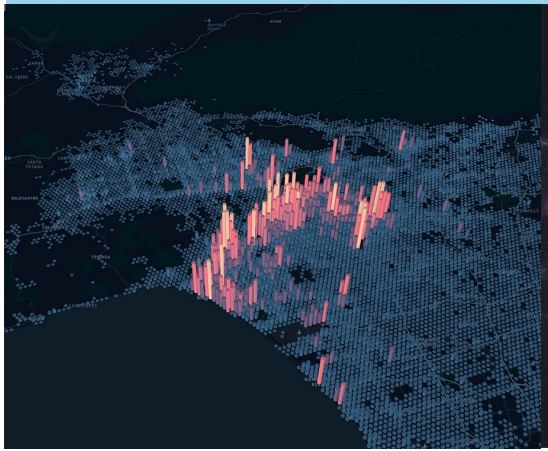


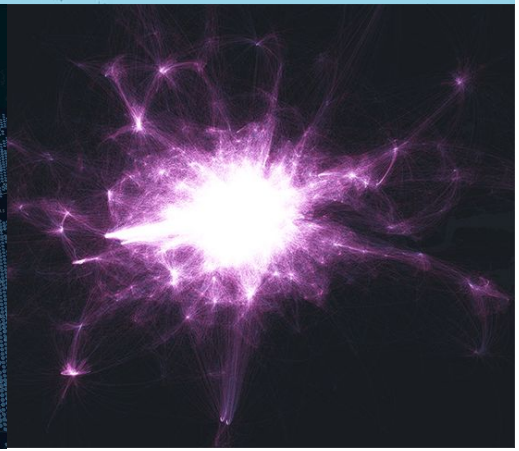



Liberty Canyon Corridor
SANTA MONICA MOUNTAINS NAT. REC. AREA
LOS ANGELES
Malibu
Santa Monica
San Bernardino
SAN BERNARDINO NATIONAL FOREST
Plans are underway for a vegetated overpass across Highway 101. This will give mountain lions in the Santa Monica Mountains an escape route to the Simi Hills.
CHINO HILLS STATE PARK
Coal Canyon Corridor
Riverside
Anaheim
Long Beach
Santa Ana
Irvine
Mission Viejo
SANTA ANA MTS.
UNITED STATES
Palm Springs
SAN JACINTO MTS.
PACIFIC OCEAN
Laguna Beach
CLEVELAND NAT. FOREST
Murrieta
Temecula
Palomar Linkage
SANTA ROSA AND SAN JACINTO MTS. NAT. MONUMENT
SALTON SEA
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.
San Clemente
CAMP PENDLETON MARINE CORPS BASE
Fallbrook
PALOMAR MT
PENINSULAR RANGES
Oceanside
Escondido
Julian
Ramona
ANZA-BORREGO DESERT STATE PARK
LAGUNA MTS.
SAN DIEGO
TIJUANA
Tecate
Jacumba
Parque-to-Park Linkage
La Rumorosa
MEXICO
Laguna Salada
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 north-western foothills.
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.
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.
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
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 - QGIS download and walk-through

Question Time