

Introduction to Geospatial Data Science

DataKind - Social Data Society - 04/05/2020

Raphaëlle Roffo

[Twitter](#) - [LinkedIn](#)

Today's plan

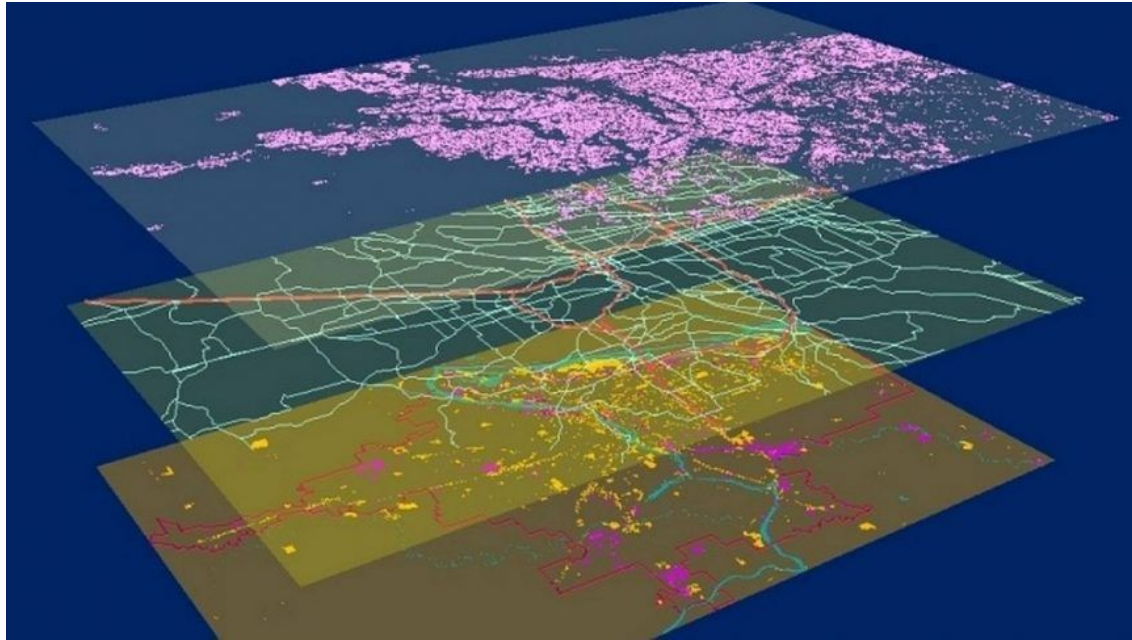


Obviously this blue part here is the land.

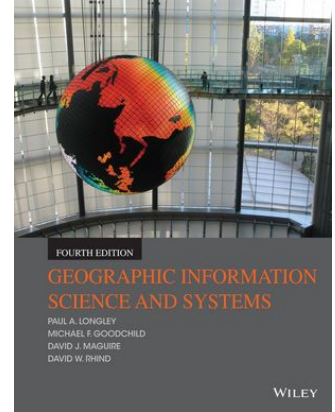


What's GIS?

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

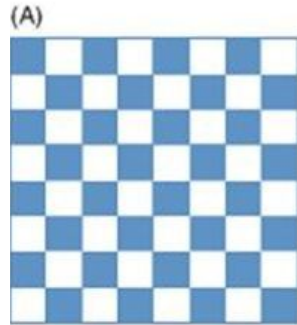


Tobler's (1979) First Law of Geography

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

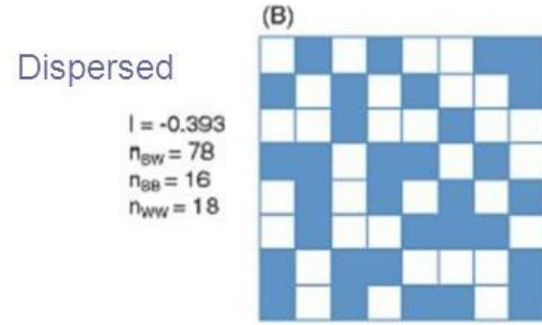


Spatial autocorrelation



Negative

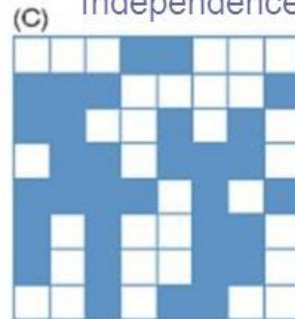
$I = -1.000$
 $n_{BW} = 112$
 $n_{BB} = 0$
 $n_{WW} = 0$



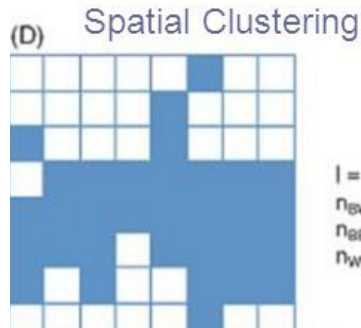
Dispersed

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

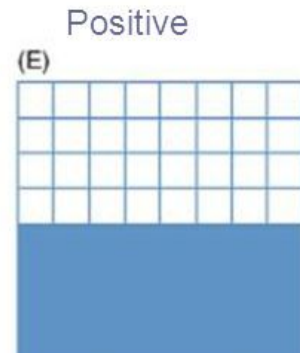
Spatial Independence



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



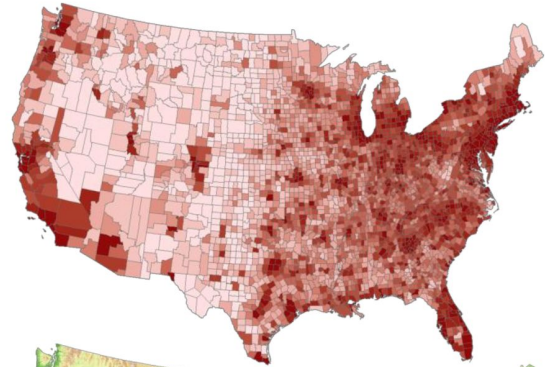
$I = +0.393$
 $n_{BW} = 34$
 $n_{BB} = 42$
 $n_{WW} = 36$



Positive

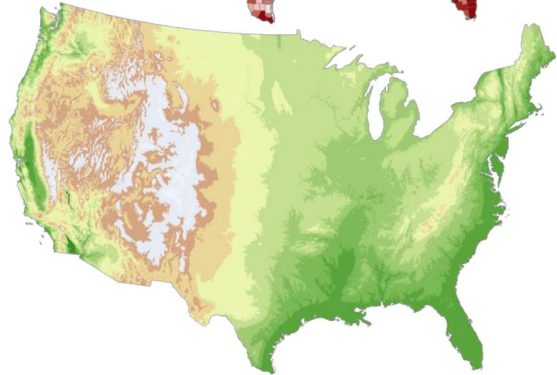
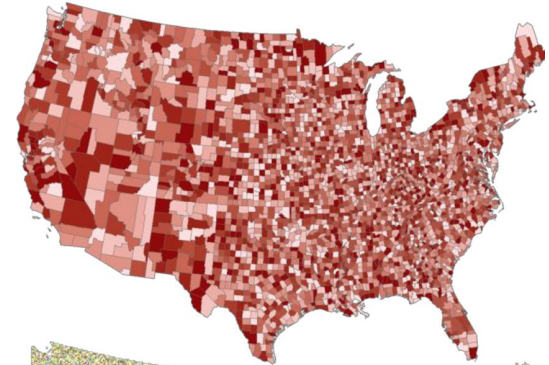
$I = +0.857$
 $n_{BW} = 8$
 $n_{BB} = 52$
 $n_{WW} = 52$

Spatial autocorrelation



If features were randomly distributed ...

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

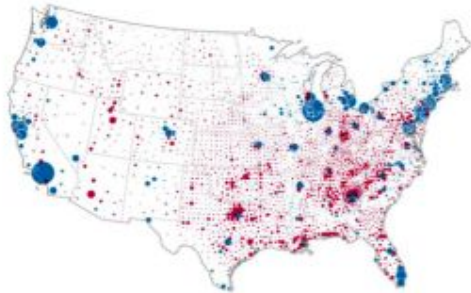
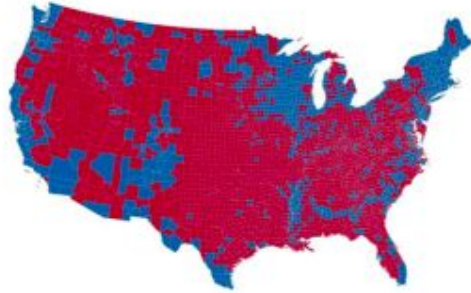


... elevation map of the US would look like this

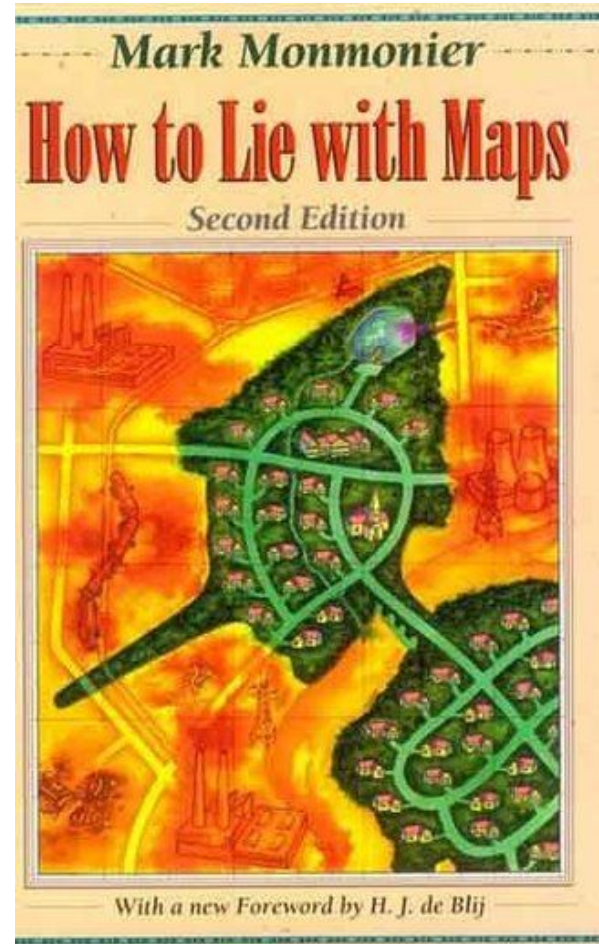


How to lie with maps

The same data,
displayed
two ways.
**Which is
the right
approach?**



HBR.ORG



Spatial is also special...

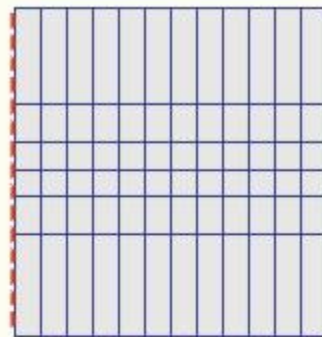
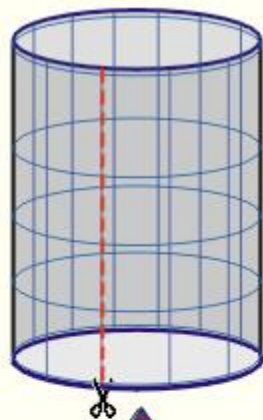
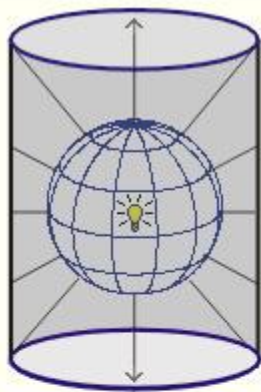
... **because the Earth isn't flat**

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.

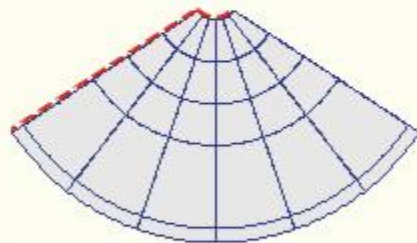
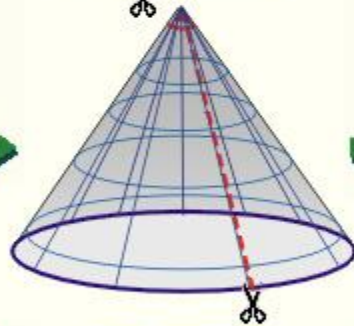
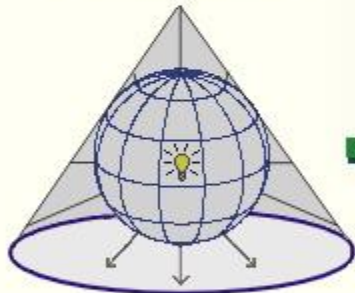
Cylindrical

Mercator



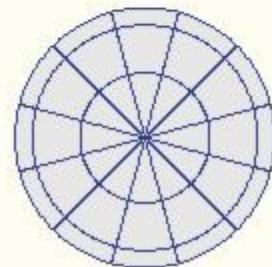
Conical

Perspective Conic



Planar

Orthographic



Projection Concepts
Perspective Examples

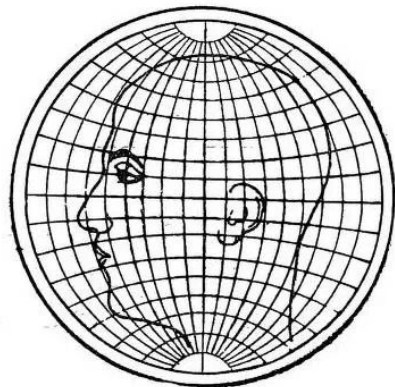


FIG. 42.—Man's head drawn on globular projection.

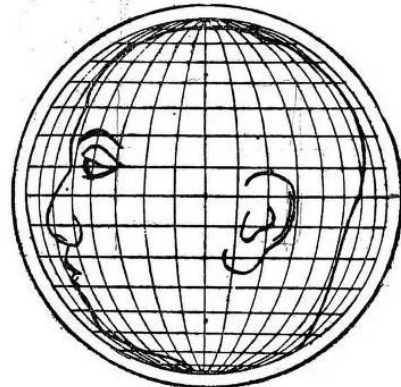


FIG. 43.—Man's head plotted on orthographic projection.

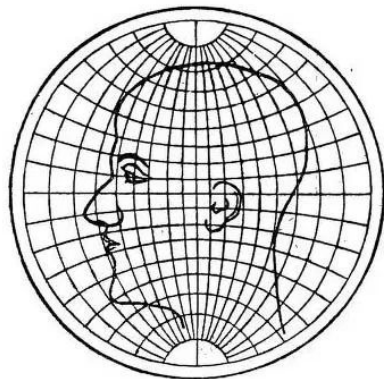


FIG. 44.—Man's head plotted on stereographic projection.

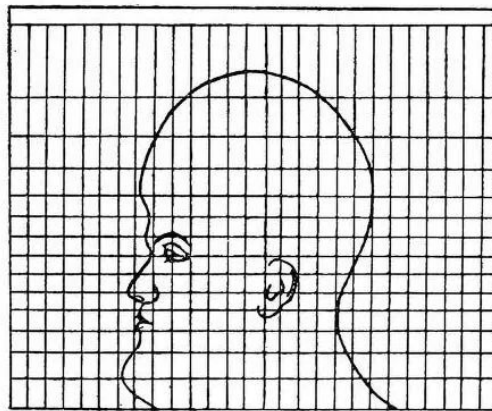
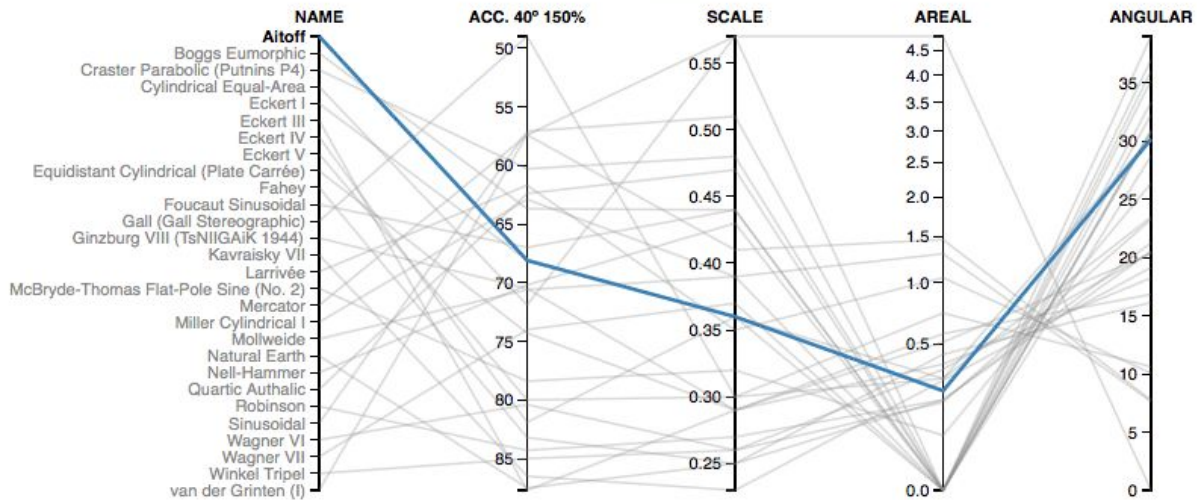
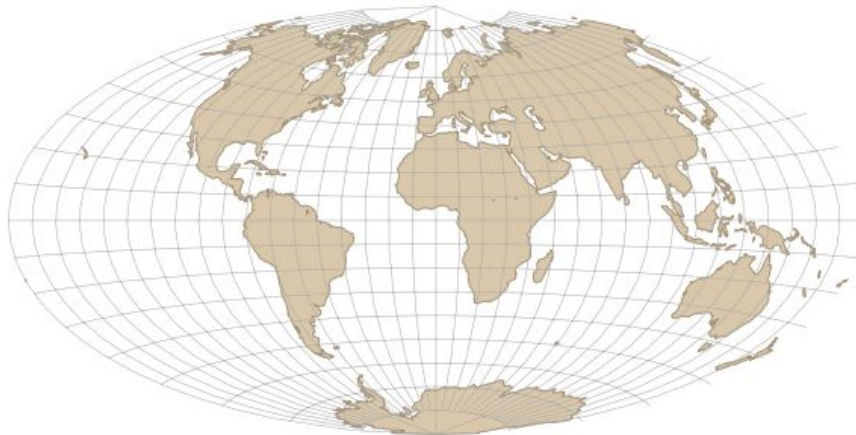


FIG. 45.—Man's head plotted on Mercator projection.





ACTUAL SIZE



Spatial is also special...

Which coordinate reference system (CRS) should you use?

Rule of thumb:

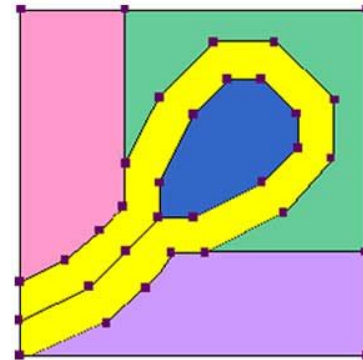
- If you are working on a **global** map: use WGS84 (or Web Mercator)
- In the **UK** the system is OSGB 1936 (British National Grid)
- France: Lambert 1993, US: NAD1927, etc.
- Each coordinate system can also be referenced by an **EPSG code**. You can find the conversion on <https://epsg.io/> (WGS84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

Spatial is special...

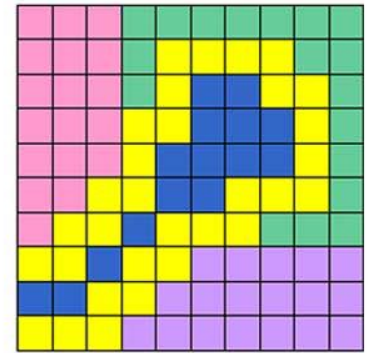


.... and can be represented in two ways:

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



Vector



Raster

Our Tools: open-source

- **QGIS** <3
- **R** (sp, rgdal, rgeos, tmap, raster, dplyr, RColorBrewer, classInt, leaflet...)
- **Python** (shapely, geopandas, folium, rasterio, GDAL, scikit-image, rasterstats, PySAL, etc. <https://medium.com/@chrieke/essential-geospatial-python-libraries-5d82fcc38731>)
- **Kepler.gl** for visualisation only (*/!\ with the CRS!*)

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**

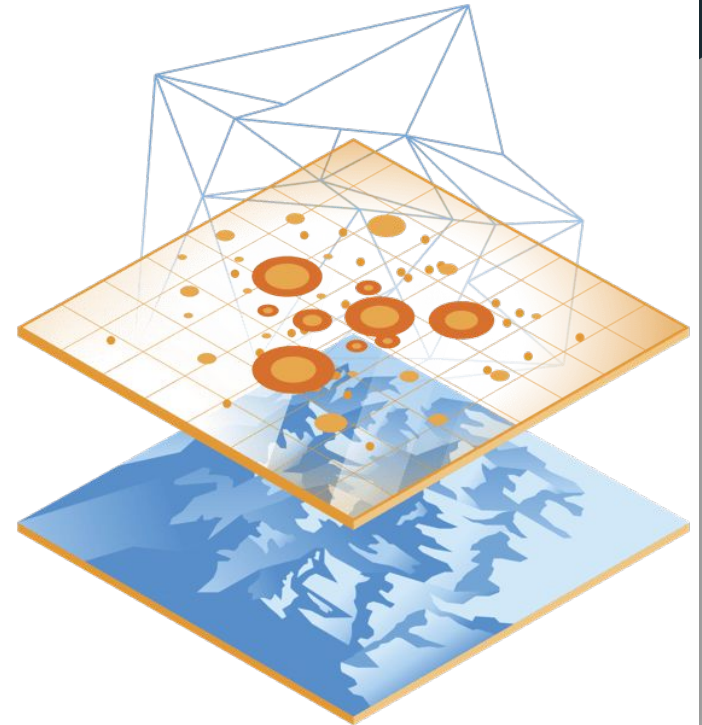


Use Cases

Spatial data analysis for decision making

- Location identification
- Network analysis
- Trends
- Patterns
- Modeling

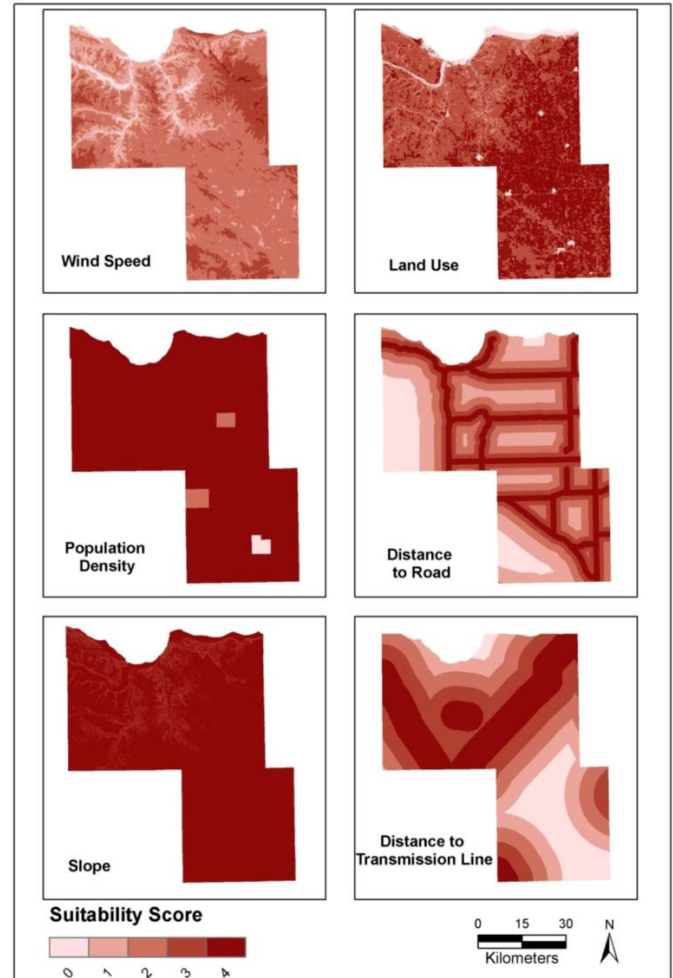
→ **usually to inform decision-making**



Site selection

Wind farms: suitability analysis

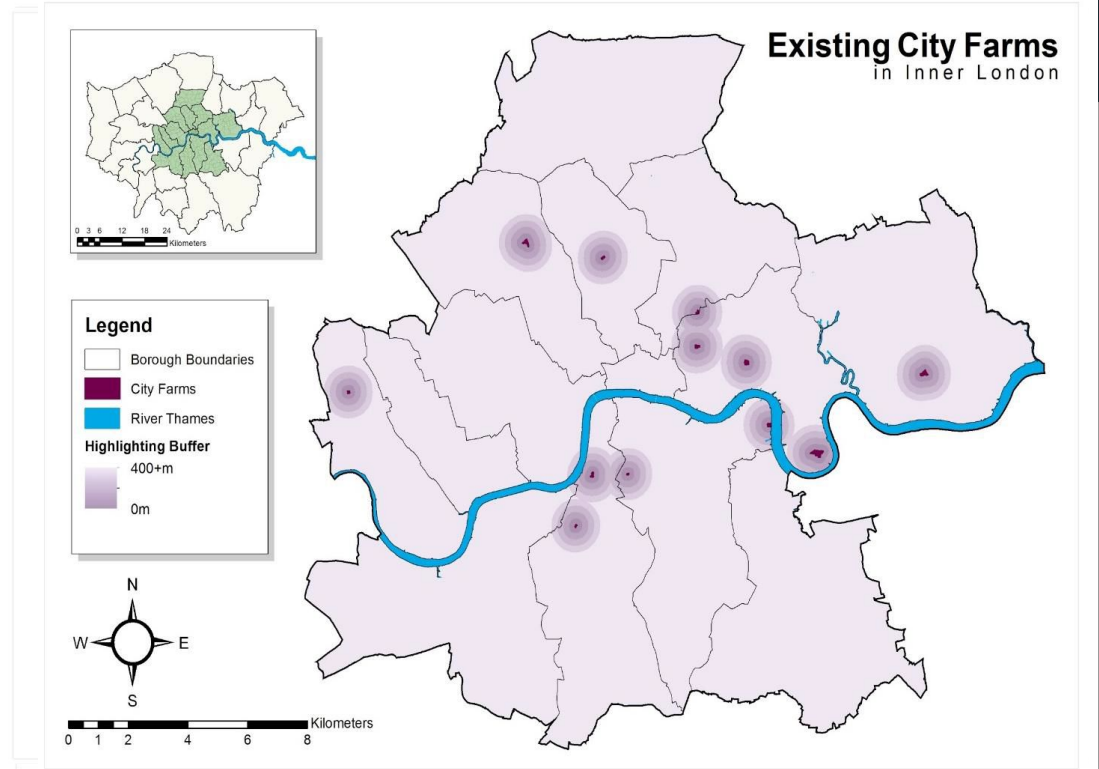
Best spot for a new city farm :)



Site selection

Wind farms

Best spot for a new city farm



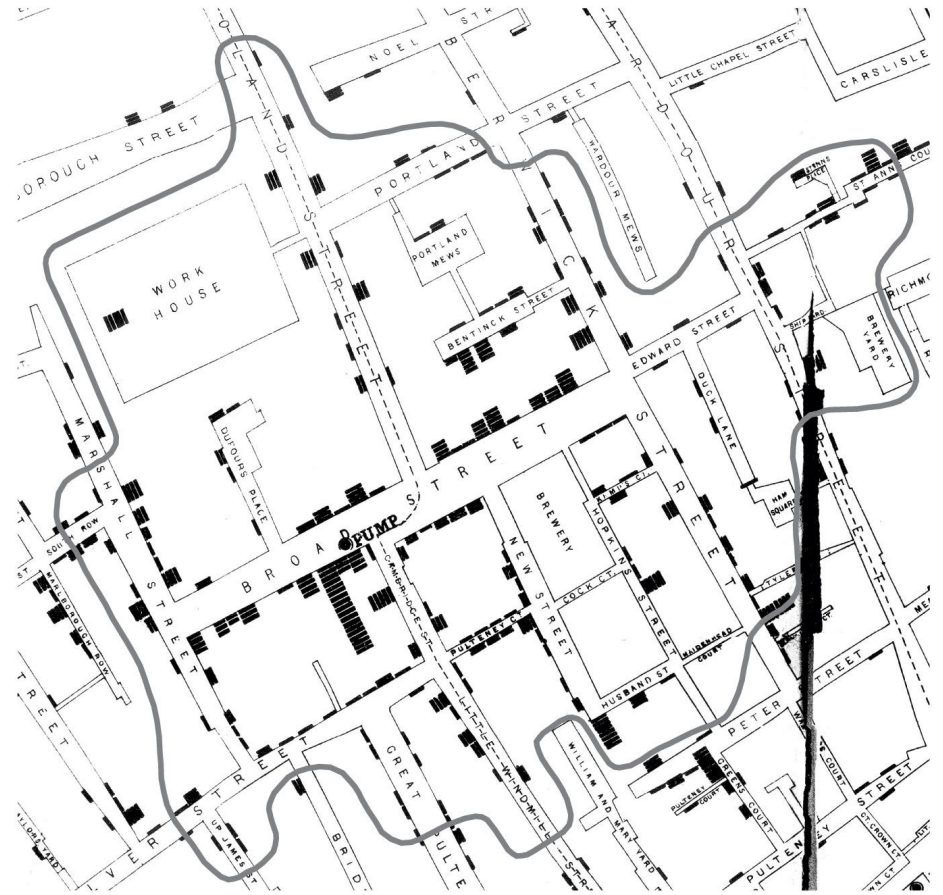
Network Analysis

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



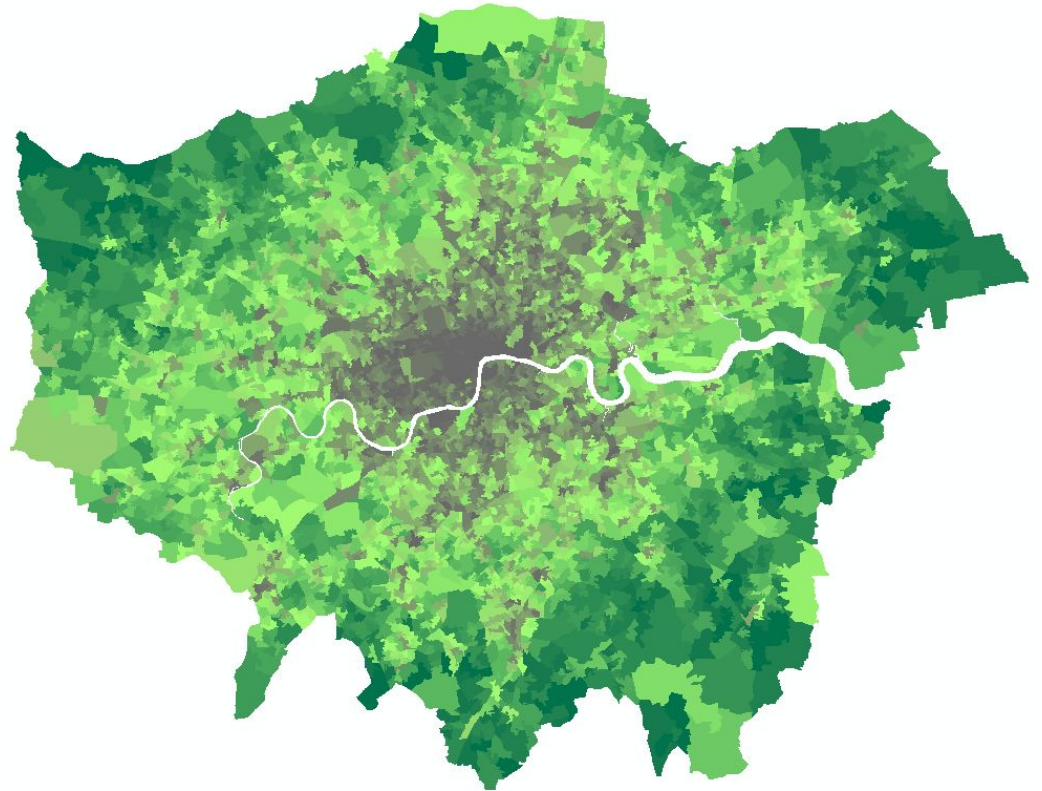
Pattern Detection

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)

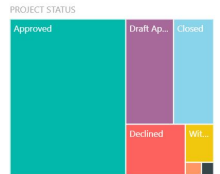
Invest in Arup - Community Engagement

Fund
Tout

Activity Category
Tout

Activity
Tout

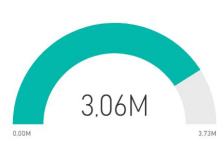
Project Name
Tout



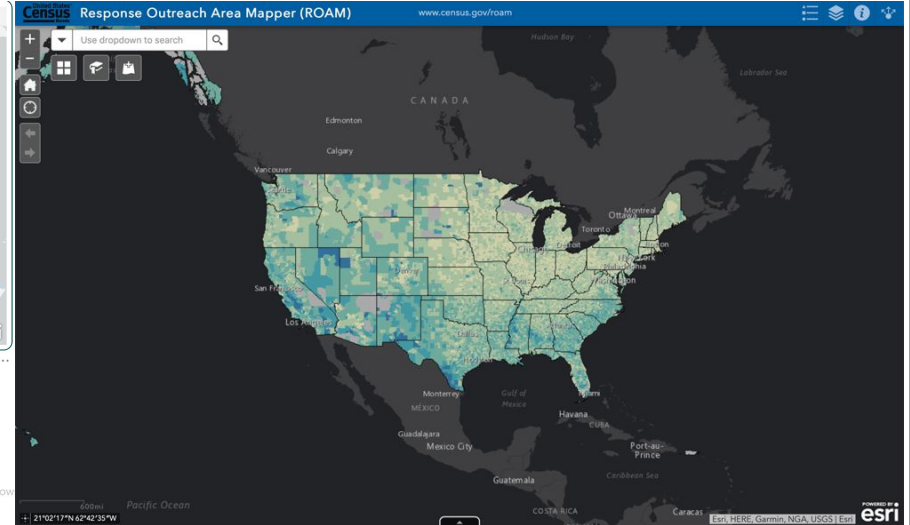
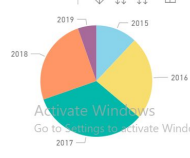
AVERAGE COST BY FUND



TOTAL COST AGAINST FORECAST



START DATE





A typical GIS Workflow

1. Requirements gathering



2. Finding the DATA



2. Finding the DATA: which formats?

Most common formats:

- Shapefiles (careful, it comes in 4 parts)
- Geopackages (QGIS, open-source) / Geodatabases (ESRI)
- PostgreSQL/PostGIS databases : can contain geometries
- Regular databases or tabular data with some kind of lat/long or Eastings/Northings coordinates
- GeoJSON (QGIS, Kepler)

2. Finding the DATA: where do I go?

Check out Robin Wilson's magic list: <https://freegisdata.rtwilson.com/> (probably the most extensive list of free GIS sources)

I frequently use:

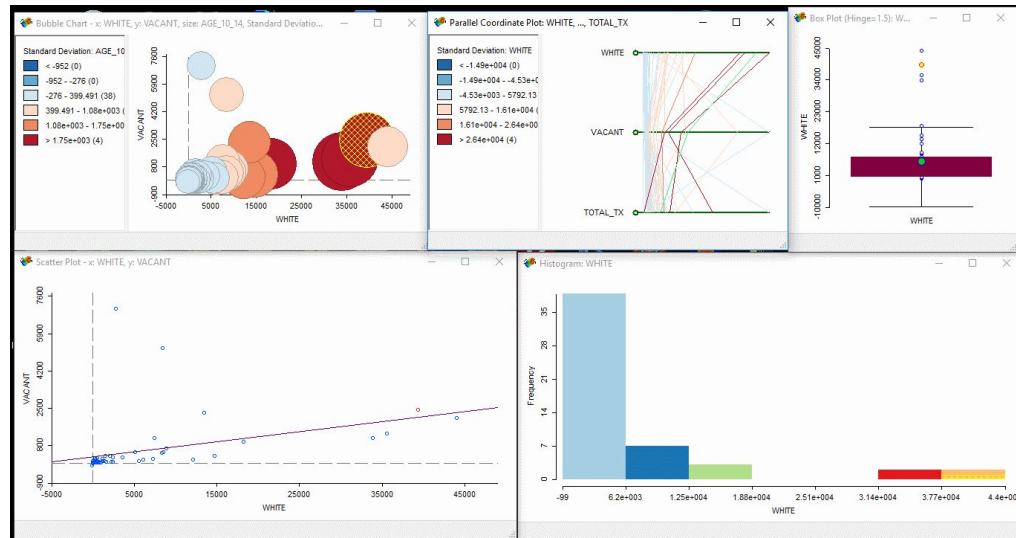
- ONS Geoportal: official England & Wales boundaries <http://geoportal.statistics.gov.uk/>
- National Surveying agency: the Ordnance Survey <https://www.ordnancesurvey.co.uk/business-and-government/products/opendata.html>
- City-level data stores (London DataStore <https://data.london.gov.uk/>)
- National data store UK Open Data <https://data.gov.uk/>
- Environmental Agency / DEFRA <https://environment.data.gov.uk/>
- Administrative boundaries by country: Diva GIS <http://www.diva-gis.org/datadown> or GADM <https://gadm.org/>

3. Exploratory Spatial Data Analysis (ESDA)



3. Exploratory Spatial Data Analysis (ESDA)

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

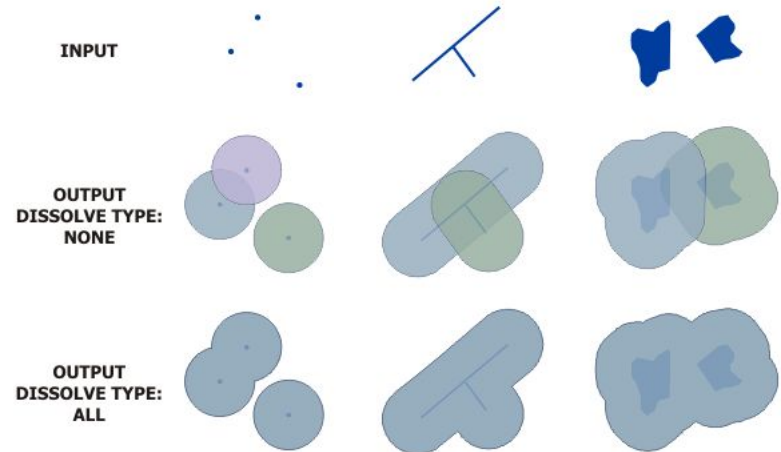
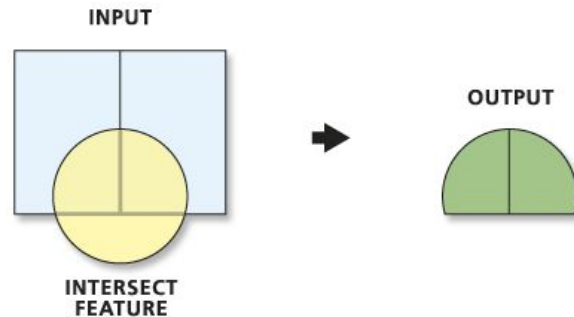


4. Refining the research objective and goals



5. Reshaping your data / feature generation

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

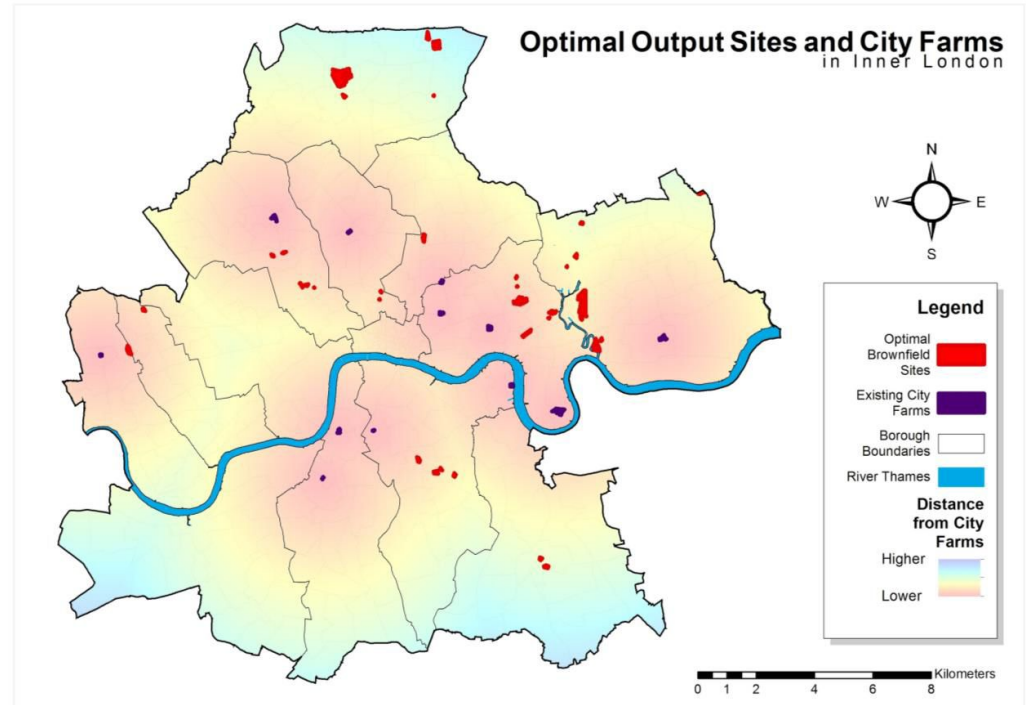


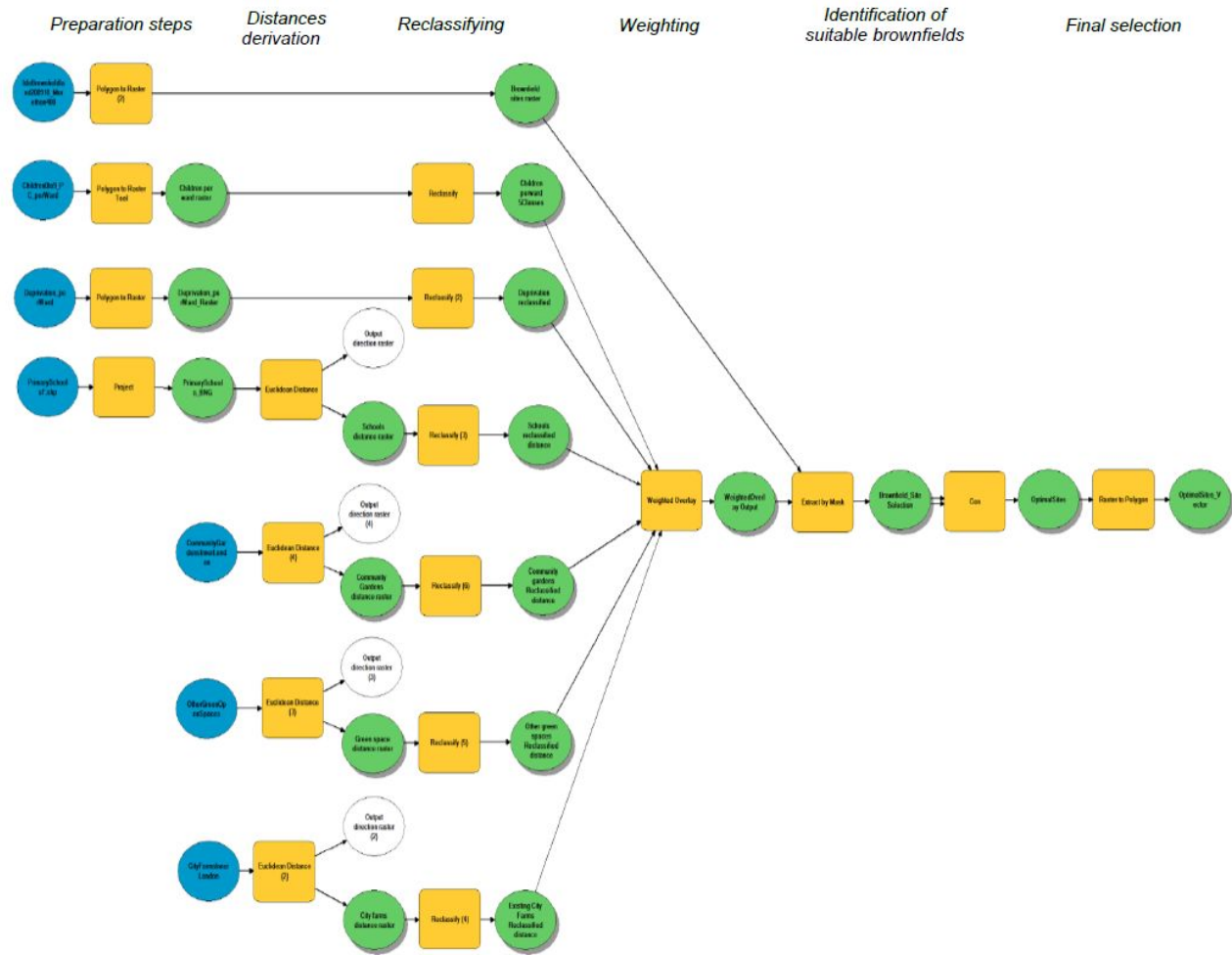
6. Analysis / Modeling



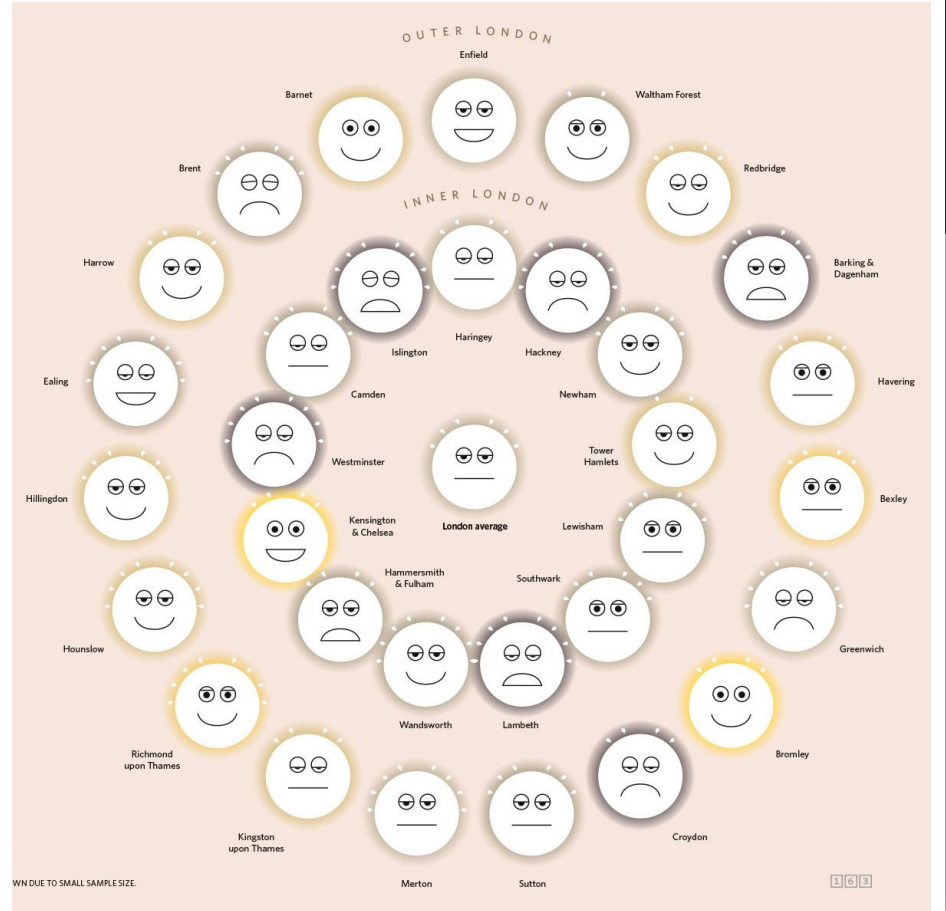
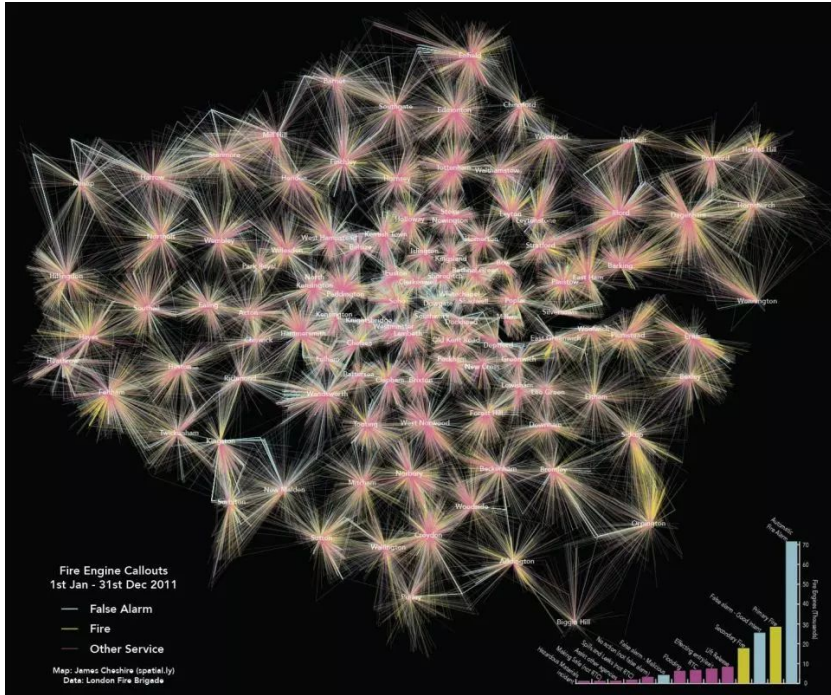
Site selection

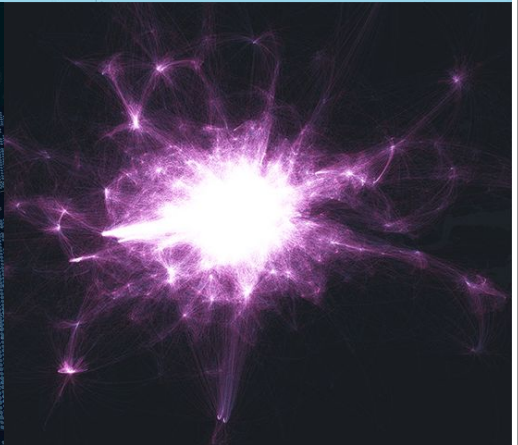
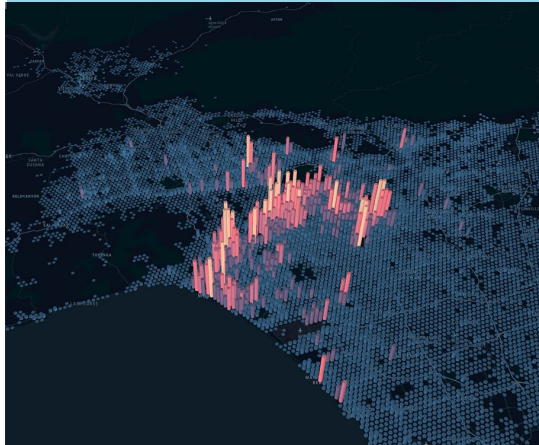
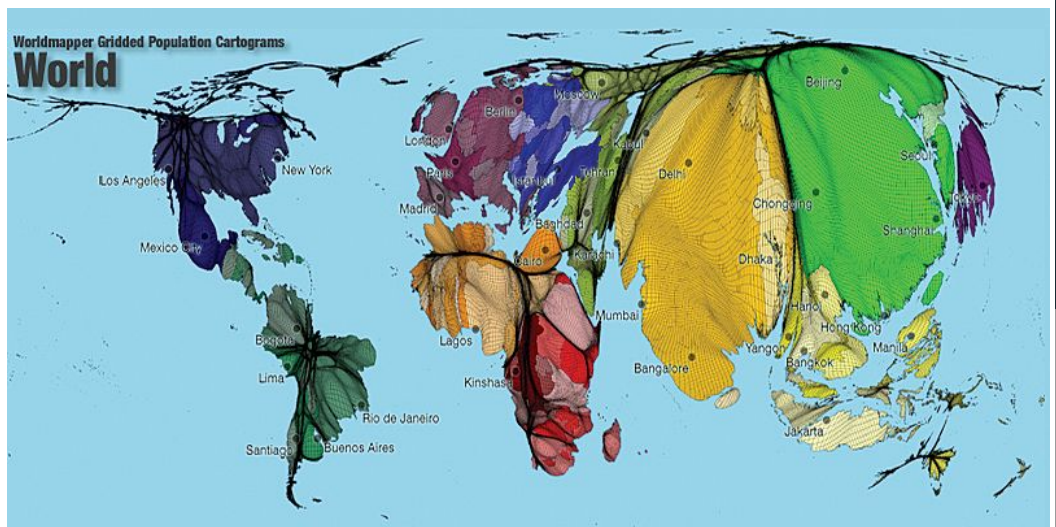
Best spot for a new city farm :)





7. Visualisation







Resources / Tutorials

Mapping / Cartographic Design

Color Brewer:

<http://colorbrewer2.org>

Esri's tips on map design:

<https://www.esri.com/arcgis-blog/products/mapping/mapping/6-easy-ways-to-improve-your-maps/>

Those general tips also apply to maps:

<https://www.columnfivemedia.com/25-tips-to-upgrade-your-data-visualization-design>

QGIS

QGIS Official Training Guide:

<https://www.qgistutorials.com/en/>

Nick Bearman's introduction to QGIS:

<https://github.com/nickbearman/intro-qgis-spatial-data>

Anita Graser's blog:

<https://anitagraser.com/>

R (rgdal, rgeos, tmap, sp, leaflet...)

CDRC tutorial:

<https://data.cdrc.ac.uk/tutorial/an-introduction-to-spatial-data-analysis-and-visualisation-in-r> #

Datacamp!

<https://www.datacamp.com/courses/working-with-geospatial-data-in-r>

Python (geopandas, folium, shapely, gdal, rasterio...)

DataCamp! <https://www.datacamp.com/community/tutorials/geospatial-data-python>
<https://www.datacamp.com/courses/visualizing-geospatial-data-in-python>

Libraries overview:

<https://medium.com/@chrieke/essential-geospatial-python-libraries-5d82fcc38731>

Folium tutorials:

<https://medium.com/datadriveninvestor/visualising-geospatial-data-with-python-d3b1c519f31>

<https://towardsdatascience.com/data-101s-spatial-visualizations-and-analysis-in-python-with-folium-39730da2adf>

Alberto Nieto - Wrangling Data and Visualizing Patterns with Python and GIS

<https://www.youtube.com/watch?v=RNwIE7zp8qE&feature=youtu.be>

All the pretty things!

Beautiful Maps <https://mapsdesign.tumblr.com/>

James Cheshire <http://spatial.ly>

Kenneth Field <https://twitter.com/kennethfield> (Author of Cartography.)

John Nelson <https://adventuresinmapping.com/>

Information is Beautiful

<https://www.informationisbeautifulawards.com/news/323-information-is-beautiful-awards-2018-the-winners>

List of cartography blogs https://blog.feedspot.com/cartography_blogs/

General Support

Women in Geospatial <https://www.womeningeospatial.org/contact>

Twitter: #gischat #gis

How to do Map Stuff webinar:

https://docs.google.com/spreadsheets/d/1TYCFBE5dnIW127Uu_aMVjWGJ_0vBB8RX-4U_TqZDoric/edit#gid=0

(or reach out to me on the DataKind Slack!)



Thanks for listening!