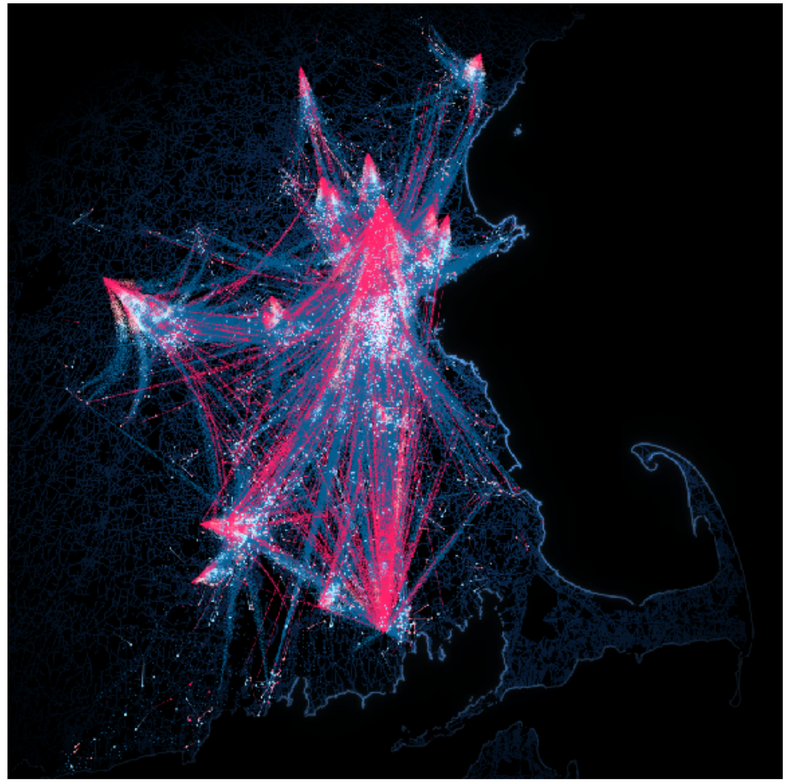
Part 0: Introduction

Michael Szell
Ane Rahbek Vierø
Anastassia Vybornova

Jul 17, 2023



We are NERDS, ITU (nerds.itu.dk)

Michael Szell
Assoc Prof
Network & Data Science

misz@itu.dk
michael.szell.net

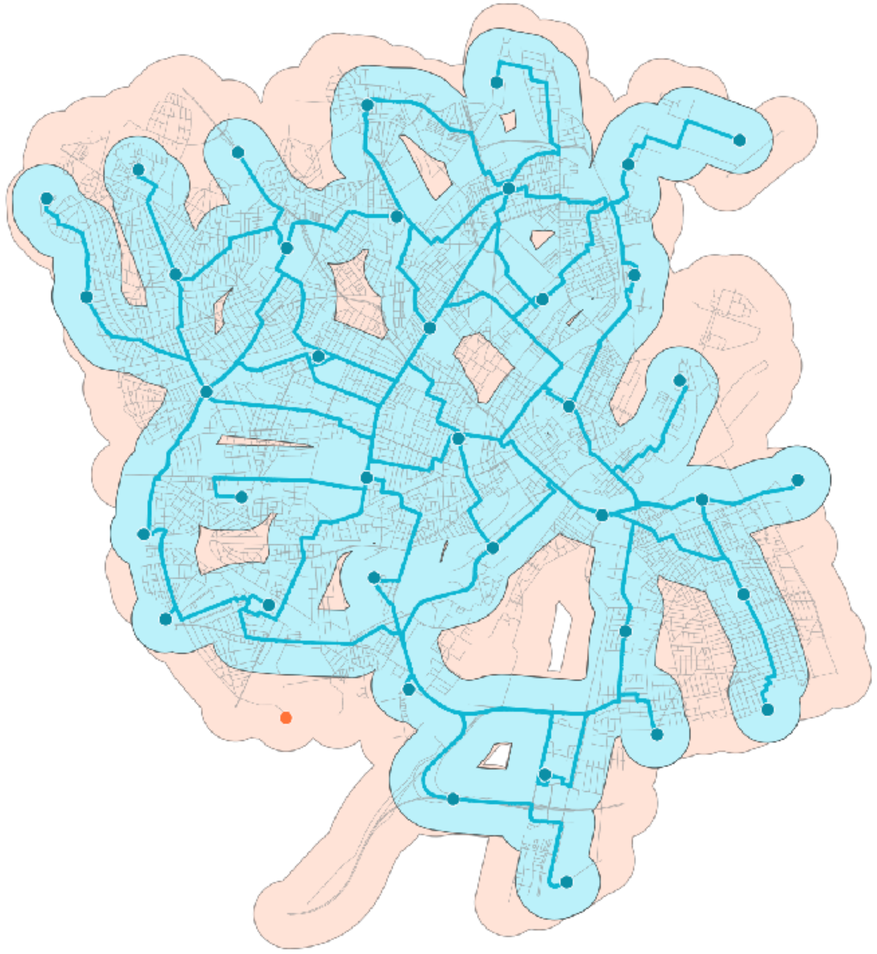
Ane Rahbek Vierø
PhD Student
GIS, Urban Planning

anev@itu.dk
anerv.github.io

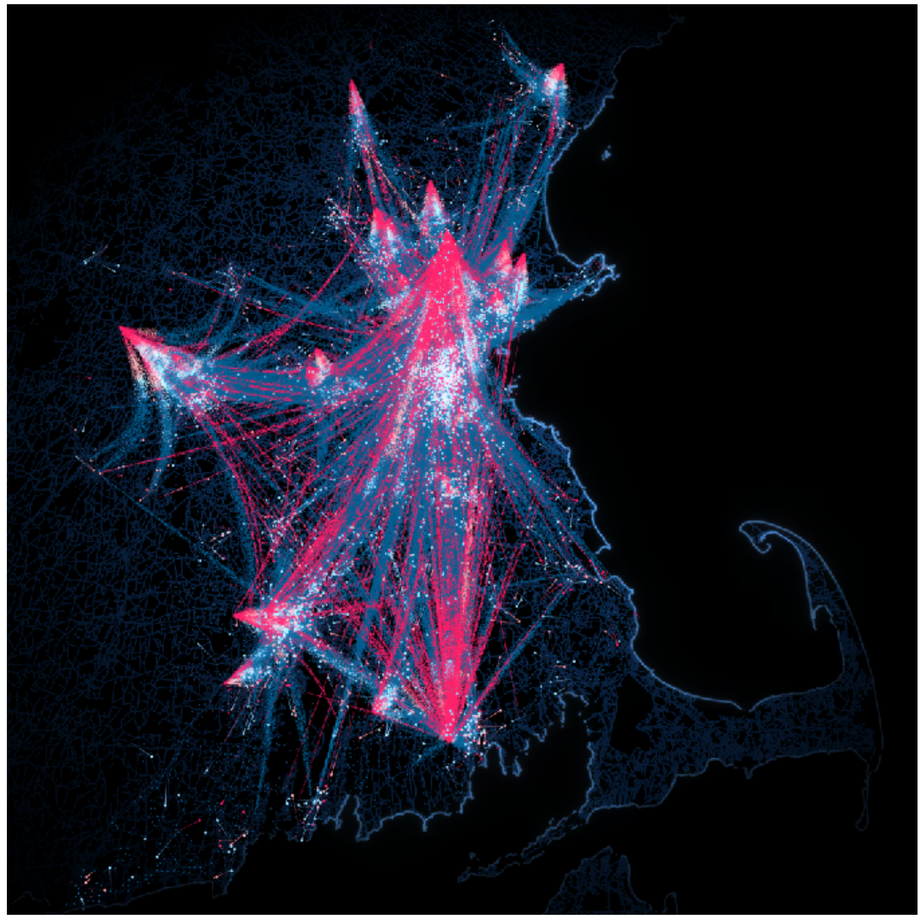
Anastassia Vybornova
PhD Student
Physics, Env Sci, Comm

anvy@itu.dk
github.com/anastassiavybornova

Our current research:



Bicycle networks

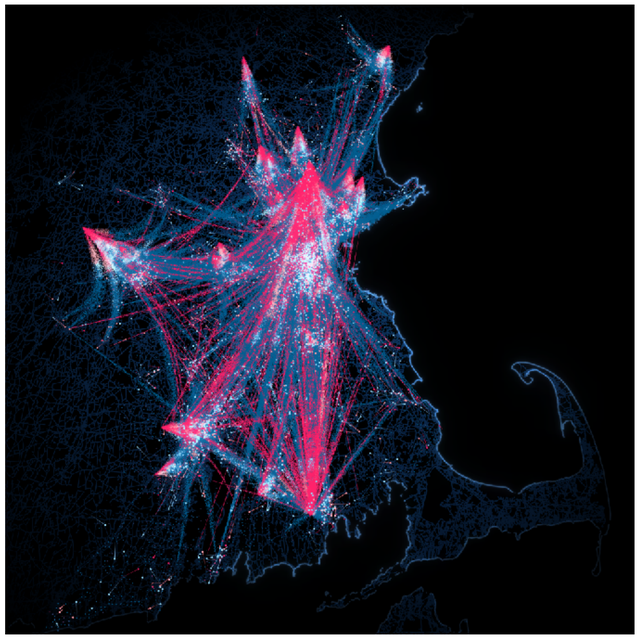


Human mobility

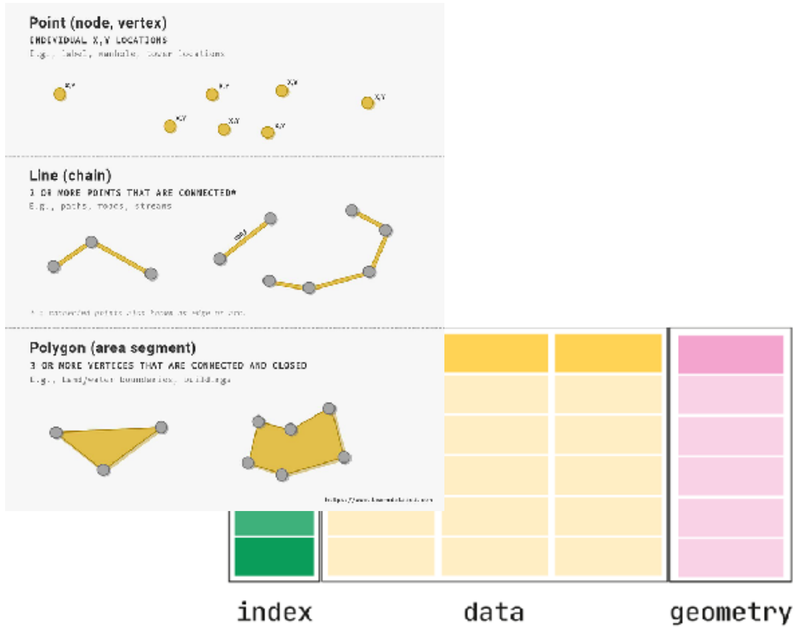
Python

Today you will learn about Geospatial Data Science (GDS)

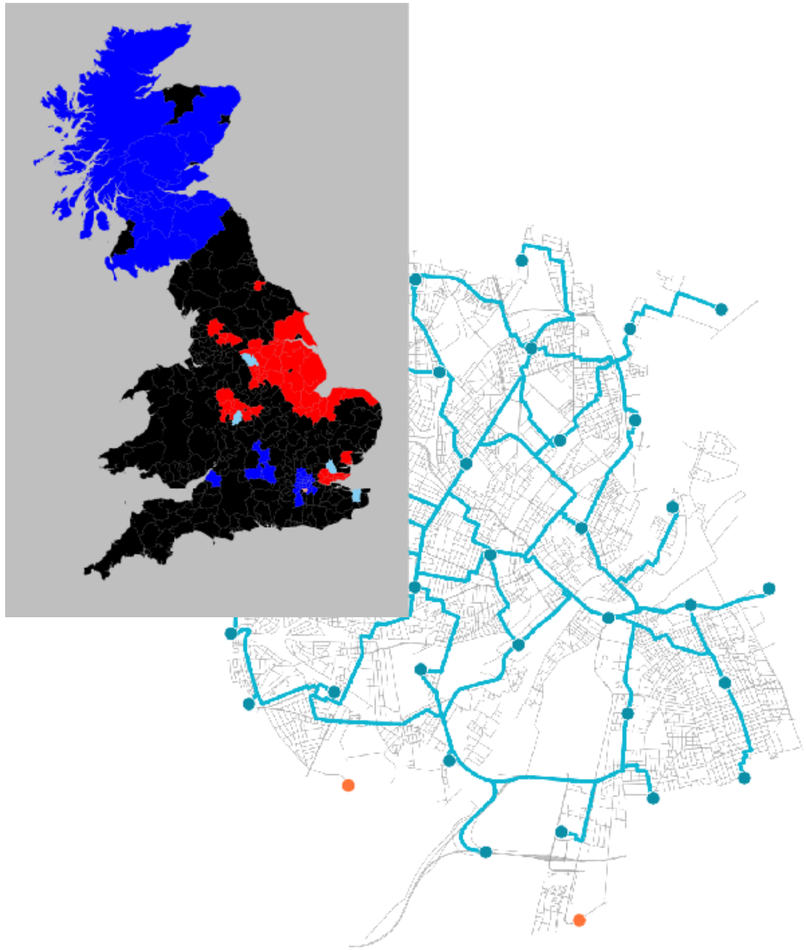
What is GDS?
Why learn about it?



Data & Python libraries



Statistical tools & applications



Geospatial Data Science is all the things that exist in ‘regular’ data science - but with a focus on space and location

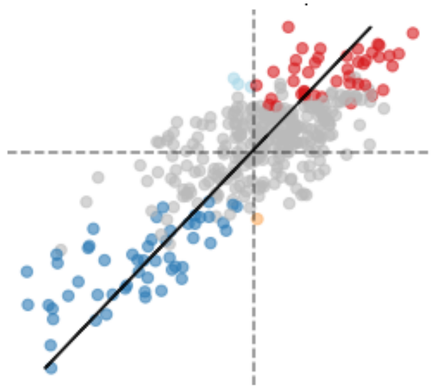
Why is this relevant to computational social scientists?

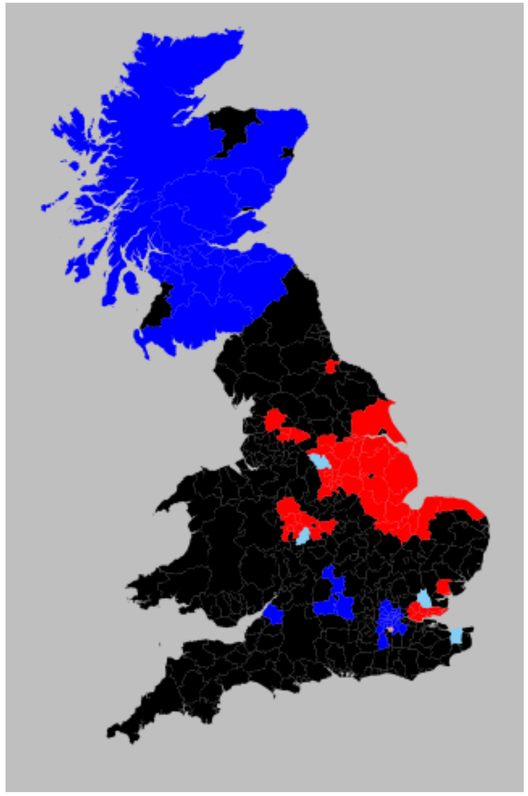
More and more social data are spatial

We need the statistical/computational tools to correctly handle space

$$\mathbf{y}_{\text{lag}} = \left(\sum_{j=1}^{n} w_{ij} y_j \right)_i = W\mathbf{y}$$

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}\, z_i\, z_j}{\sum_i z_i^2}$$





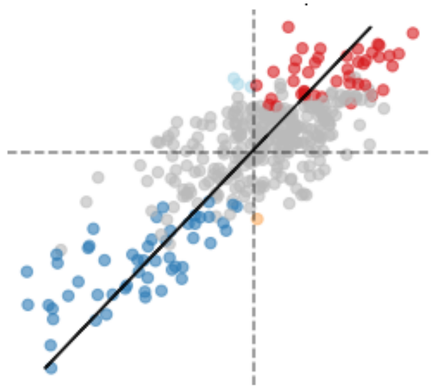
Why is this relevant to computational social scientists?

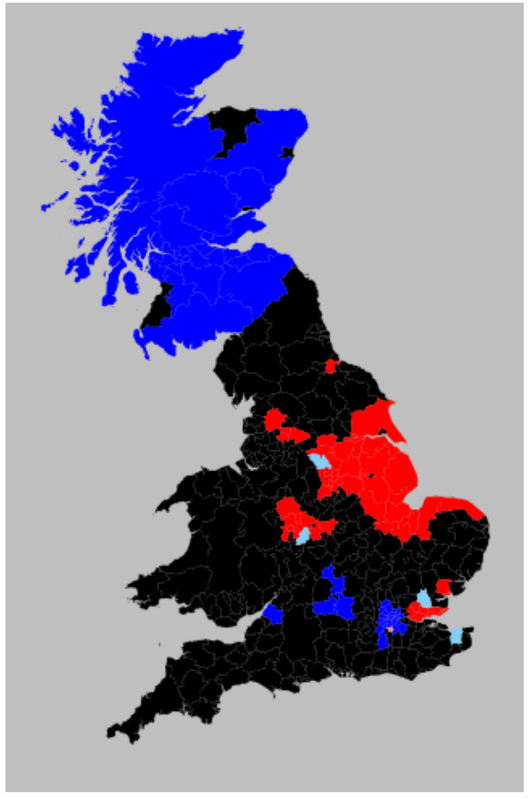
More and more social data are spatial

We need the statistical/computational tools to correctly handle space

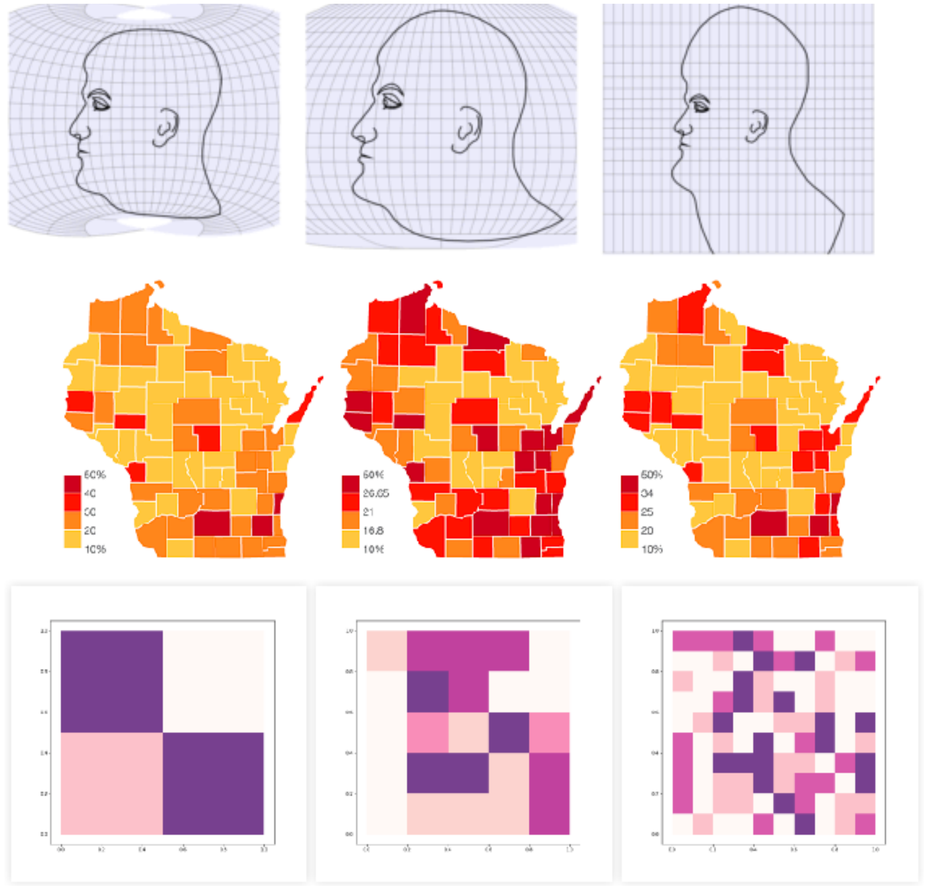
$$\mathbf{y}_{\text{lag}} = \left(\sum_{j=1}^{n} w_{ij} y_j \right)_i = W\mathbf{y}$$

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}\, z_i\, z_j}{\sum_i z_i^2}$$





We need to be aware of pitfalls



What is Geospatial Data?

Geospatial data is information that describes objects, events or other features with a location on or near the surface of the earth.

- Coordinates 
- Attributes 
- Temporal information 

https://www.ibm.com/topics/geospatial-data

There is vector and raster data

Vector: Geometric objects

.gpkg, .shp, .svg, geojson

Raster: Grid of pixels

.tif, .jpg, .png, .bmp

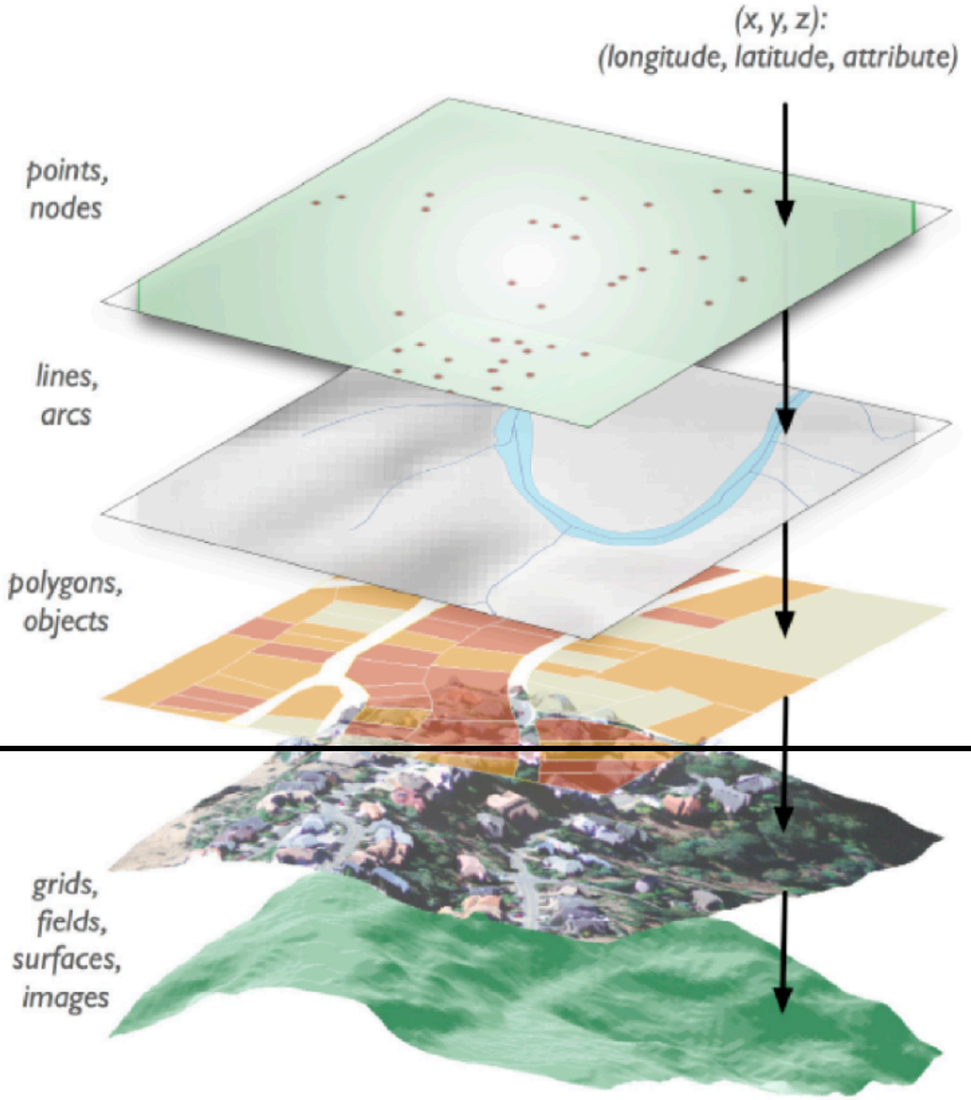
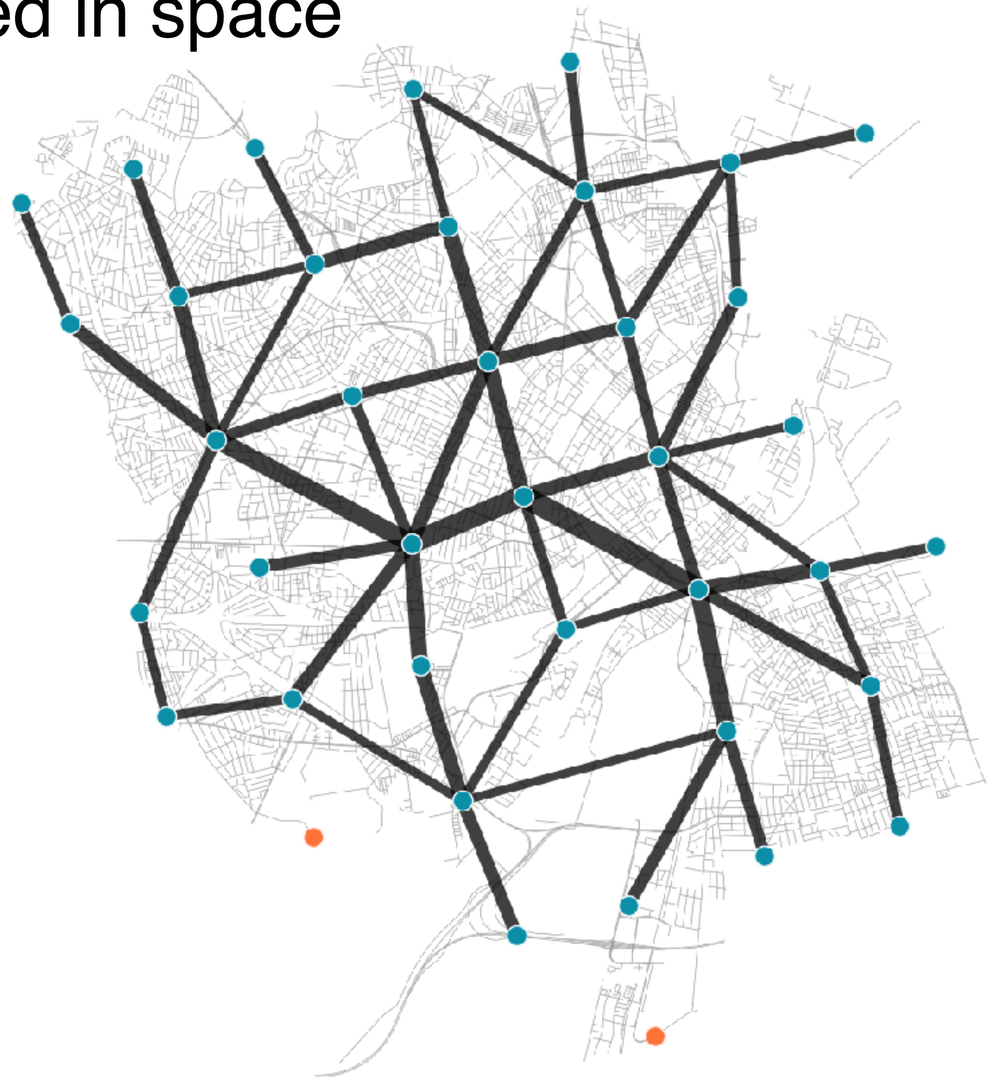
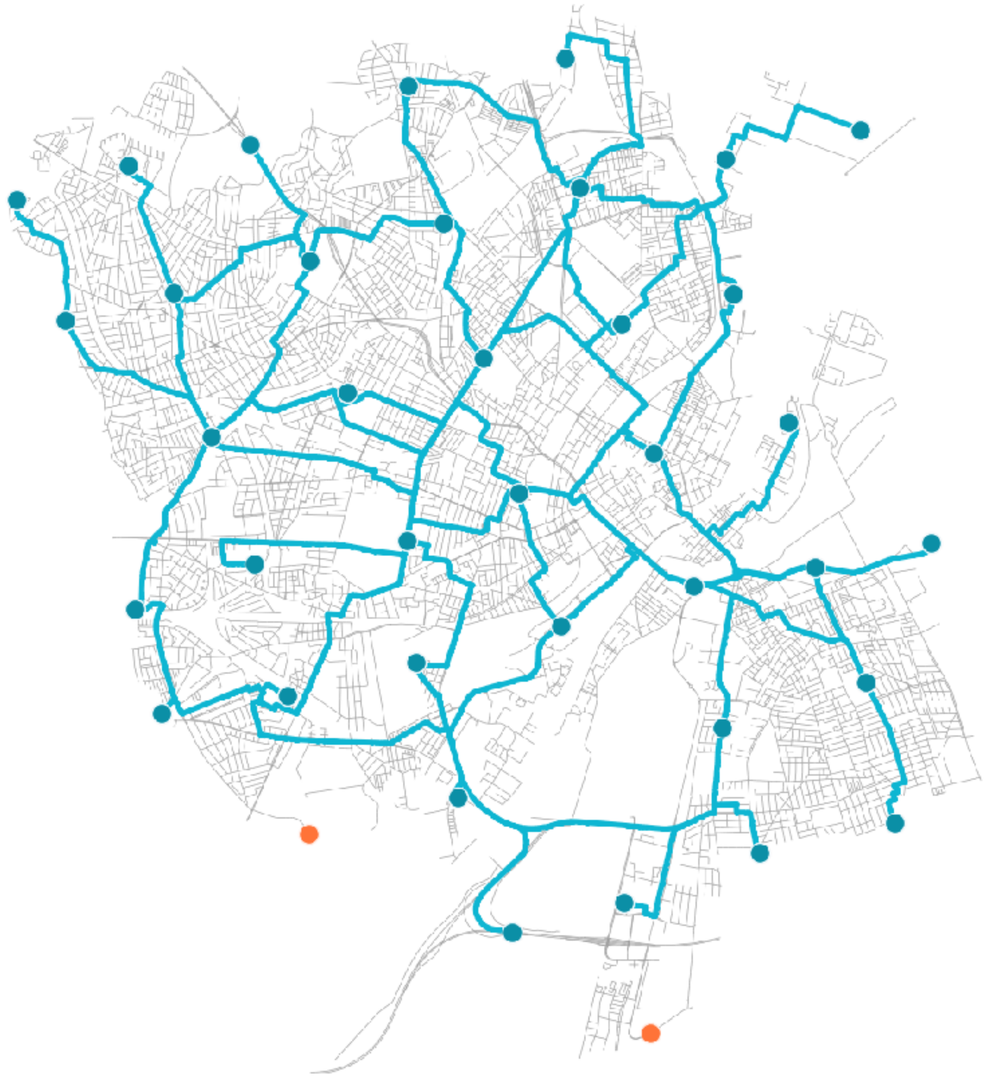


Image: Maggi Kelly, UC Berkeley

There is network data

Structure (topology),
embedded in space





GDS asks: How do things relate in space?

> Everything is related to everything else, but near things are more related than distant things.
>
> Tobler's 1st law of geography

GDS asks: How do things relate in space?

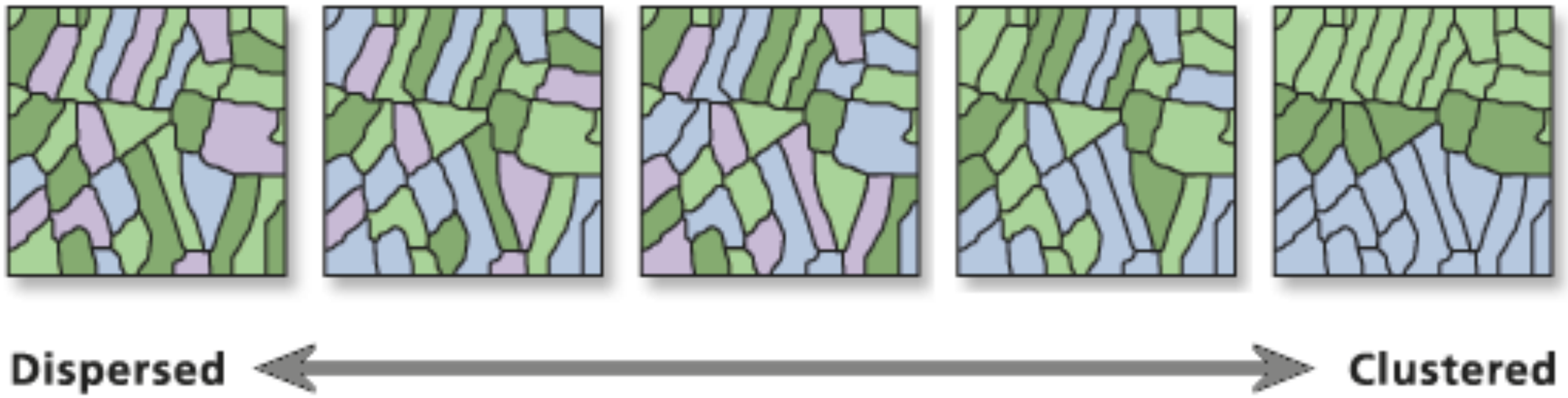
How can we formalize this question?

How can we operationalize this question?

How can we visualize this question?

GDS is mathematical, visual, and computational

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$



We use the Geographic Data Science platform `gds_env`

💻 Preparations

Before the tutorial, install `gds_py` (version 9.0) from `gds_env`:
https://darribas.org/gds_env/stacks/gds_py/#install

https://github.com/NERDSITU/gdstutorial

https://darribas.org/gds_env/

Tutorial schedule

July 17, 2023	Part	Topic
09:00 - 09:05	0. Introduction	Introduction
09:05 - 09:40	1. Data Handling	Data & Geometry, CRS, Libraries
09:40 - 09:50		*10 min break*
09:50 - 10:20	2. Spatial Statistics	Choropleth Maps, Spatial Autocorrelation
10:30 - 11:00		*30 min break*
11:00 - 11:35	3. OpenStreetMap	Introduction to OSM & OSMnx
11:35 - 11:45		*10 min break*
11:45 - 12:30	4. Spatial Networks	Spatial Networks with Geopandas

https://github.com/NERDSITU/gdstutorial

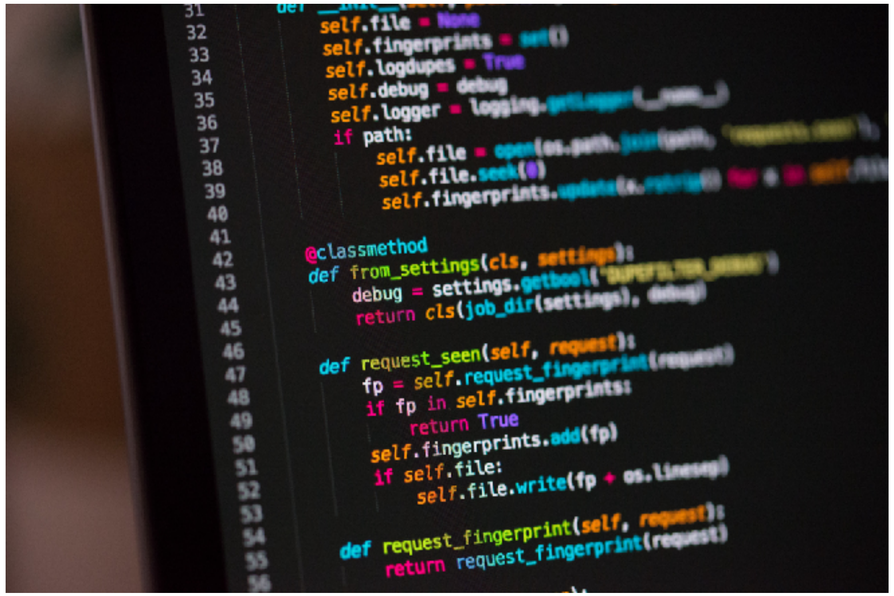
We will only cover the procedural part of GDS (in Python)

DB handling

PostgreSQL

mongoDB

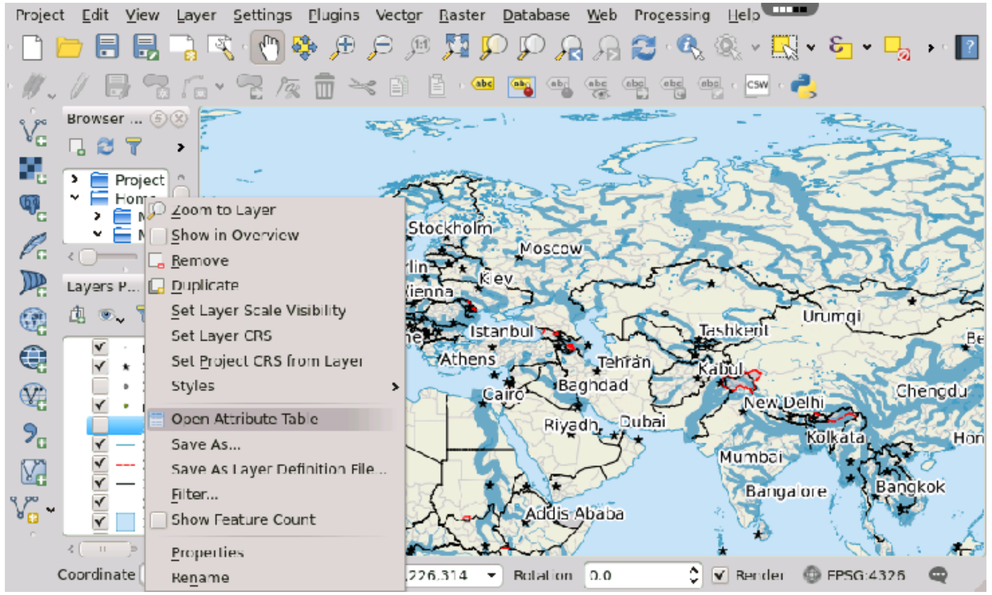
Procedural



python™

R

Point & Click GIS

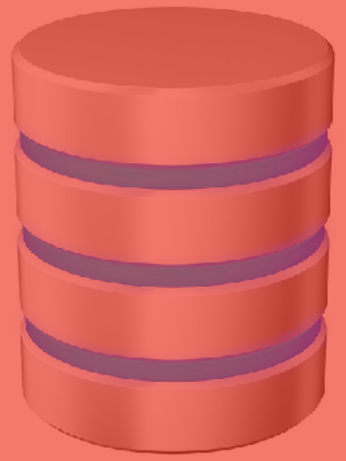


QGIS

ArcGIS

We will only cover the procedural part of GDS (in Python)

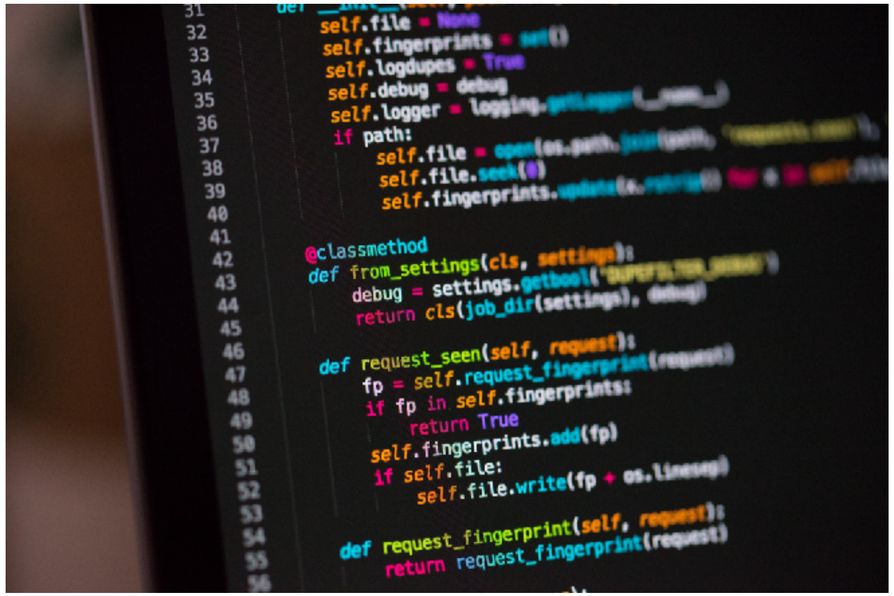
DB handling



PostgreSQL

mongoDB

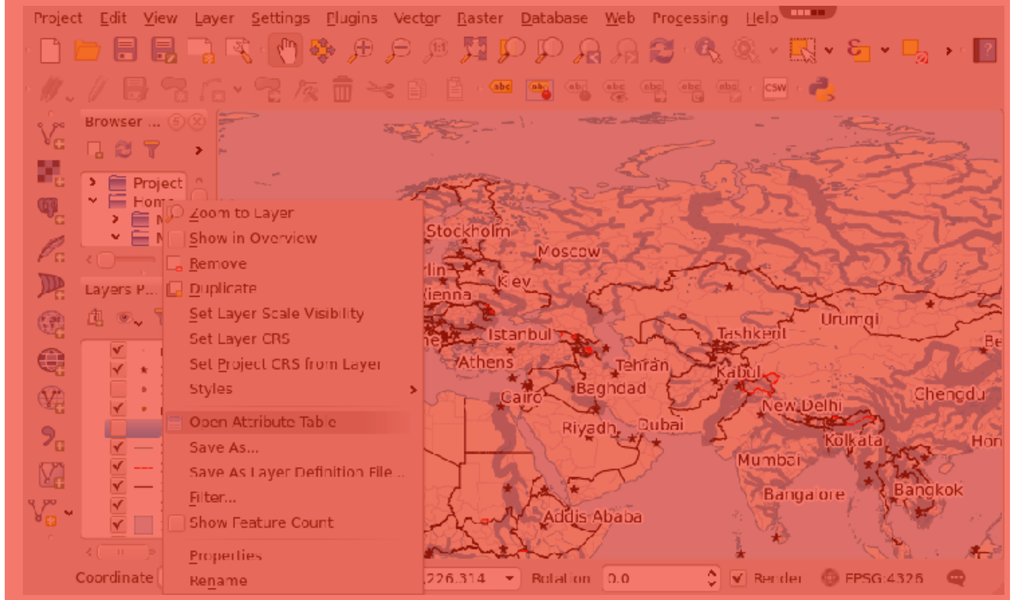
Procedural



python™

R

Point & Click GIS



QGIS

ArcGIS