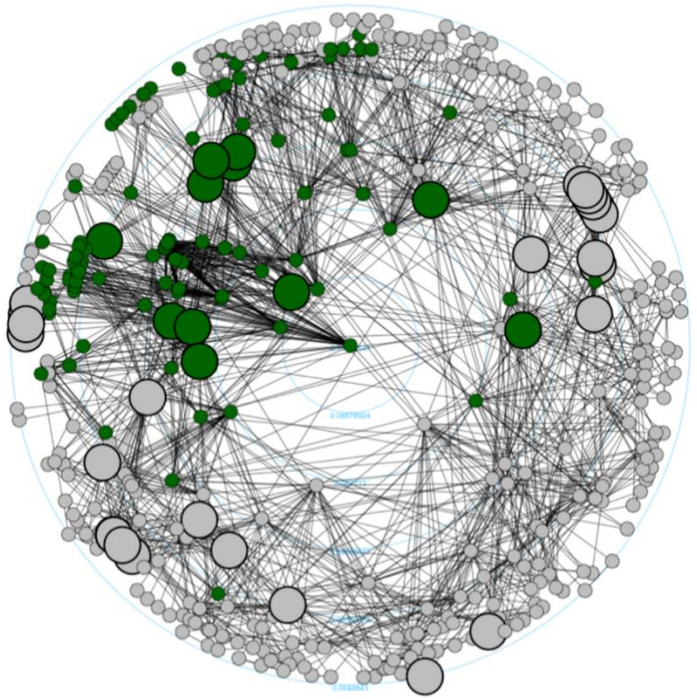
Geographic Data Science – guest lecture

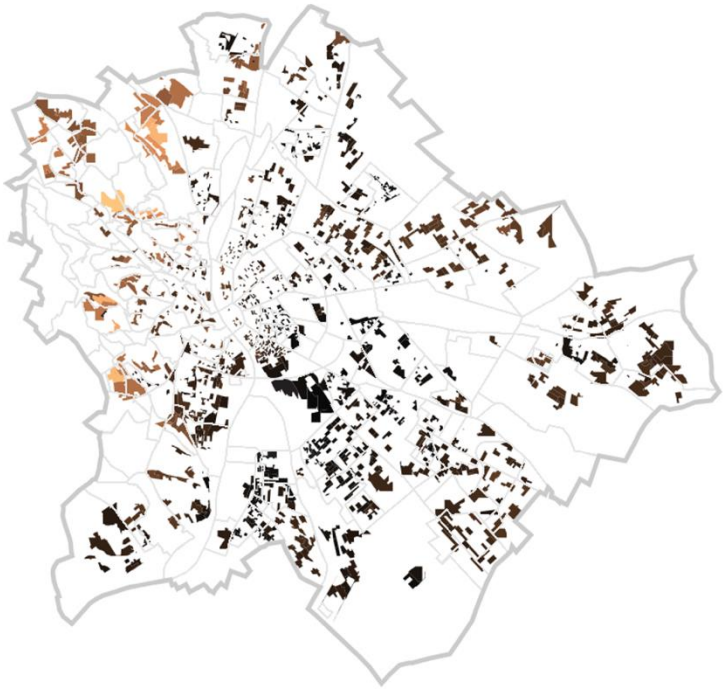
Sándor Juhász

sandorjuhasz.com

sandor.juhasz@uni-corvinus.hu

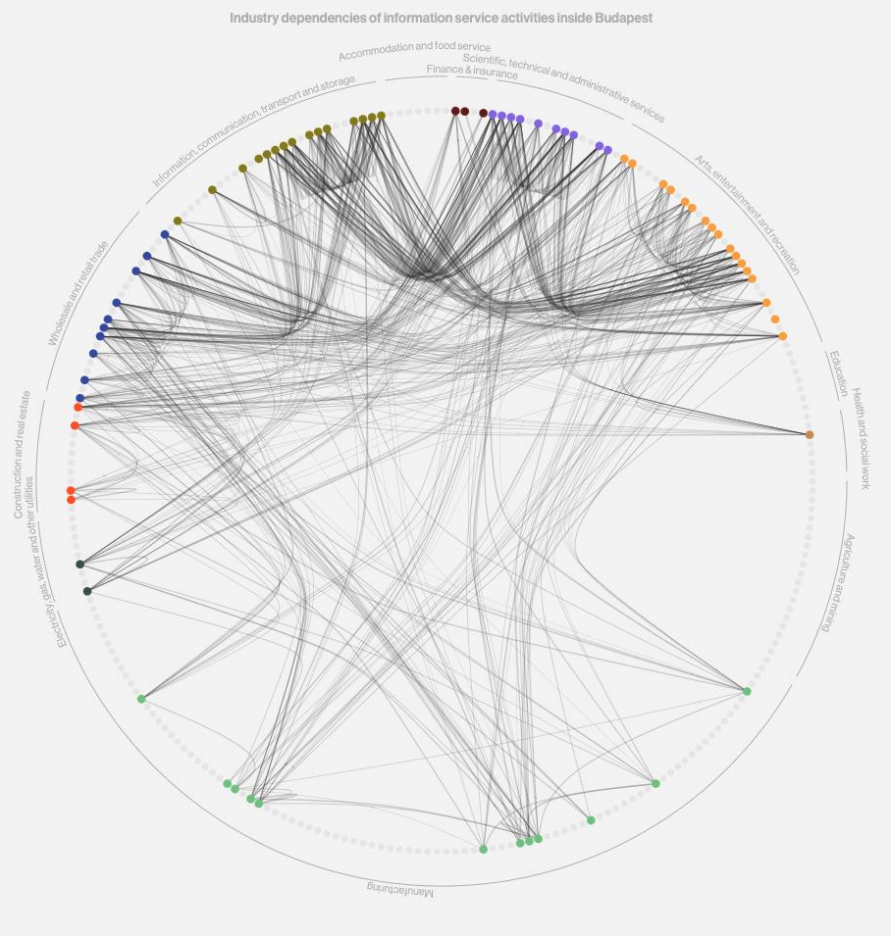


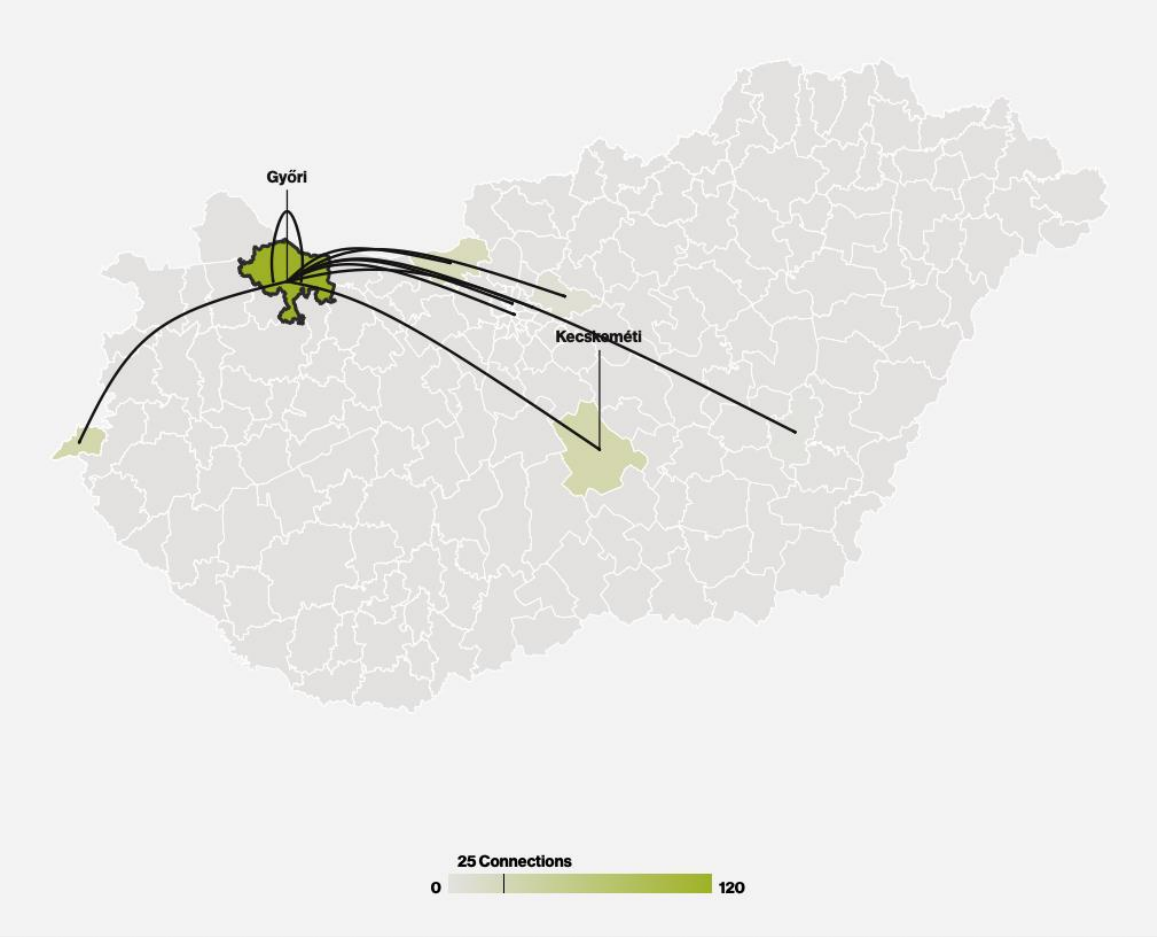
SOCIO-ECONOMIC NETWORKS



URBAN DATA SCIENCE

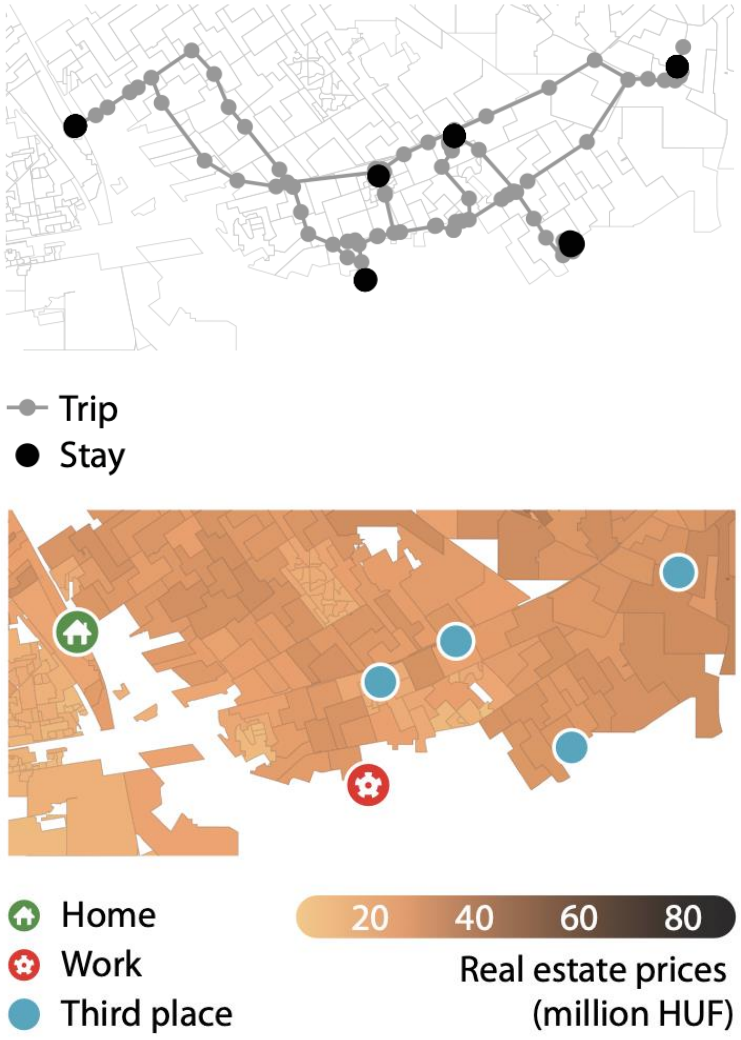
Current research – Supply networks and labor flows

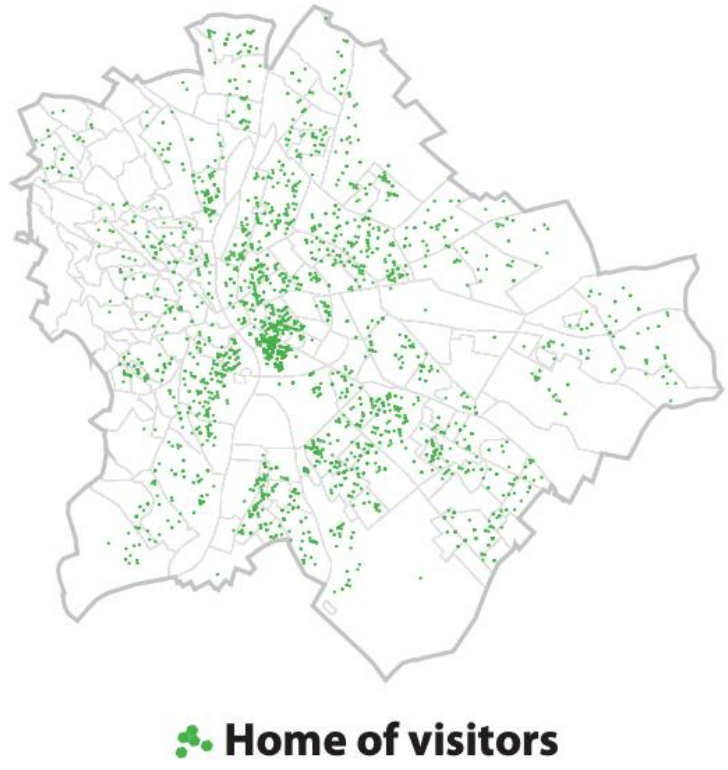




https://vis.csh.ac.at/colocation-suppliers/

Current research – Cities through individual mobility data





Juhász et al. (2023) **EPJ Data Science**

Geographic Data Science

- Full course in **Social Data Science 1-year MA program** (spring semester)
- We follow giants with local twists http://darribas.org/gds18/
- Tutorials by the NERDS (https://nerds.itu.dk)

Michael Szell
Assoc Prof
ITU Copenhagen

Anastassia Vybornova
PhD Student
ITU Copenhagen

Dani Arribas-Bel
Professor in Geographic Data Science, University of Liverpool (UK)

Geographic Data Science – today

- What is GDS (Geographic Data Science)?
- Why learn about it?
- Data and pyhon libraries
- Statistical tools and applications

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

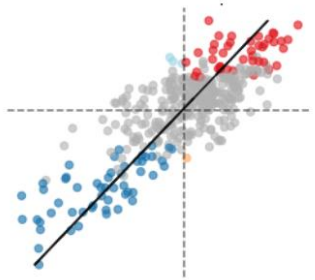
Why is this relevant?

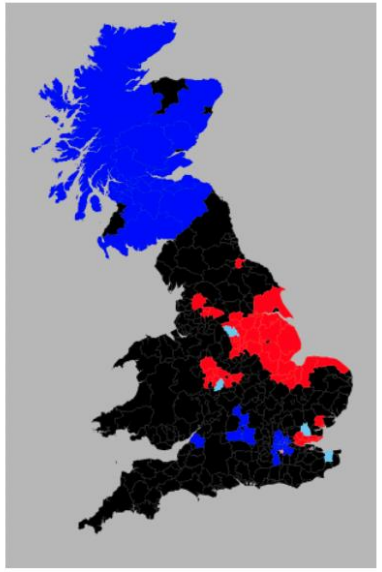
More and more social and economical data are spatial

Need for statistical tools to correctly handle space

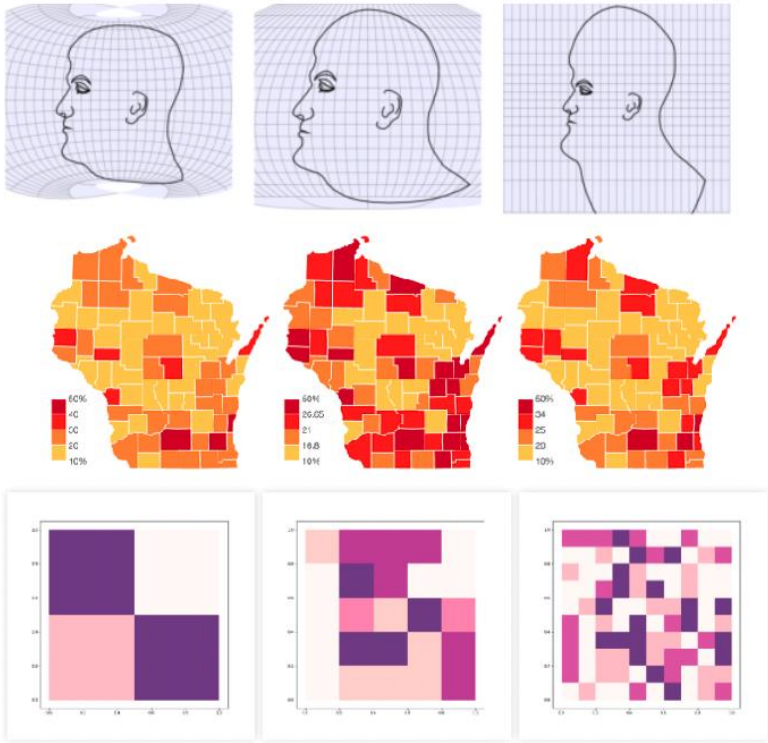
$$\mathbf{y}_{\text{lag}} = \left(\sum_{j=1}^{n} w_{ij} y_j \right)_i = W\mathbf{y}$$

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





We need to be aware of pitfalls



What is geographic data?

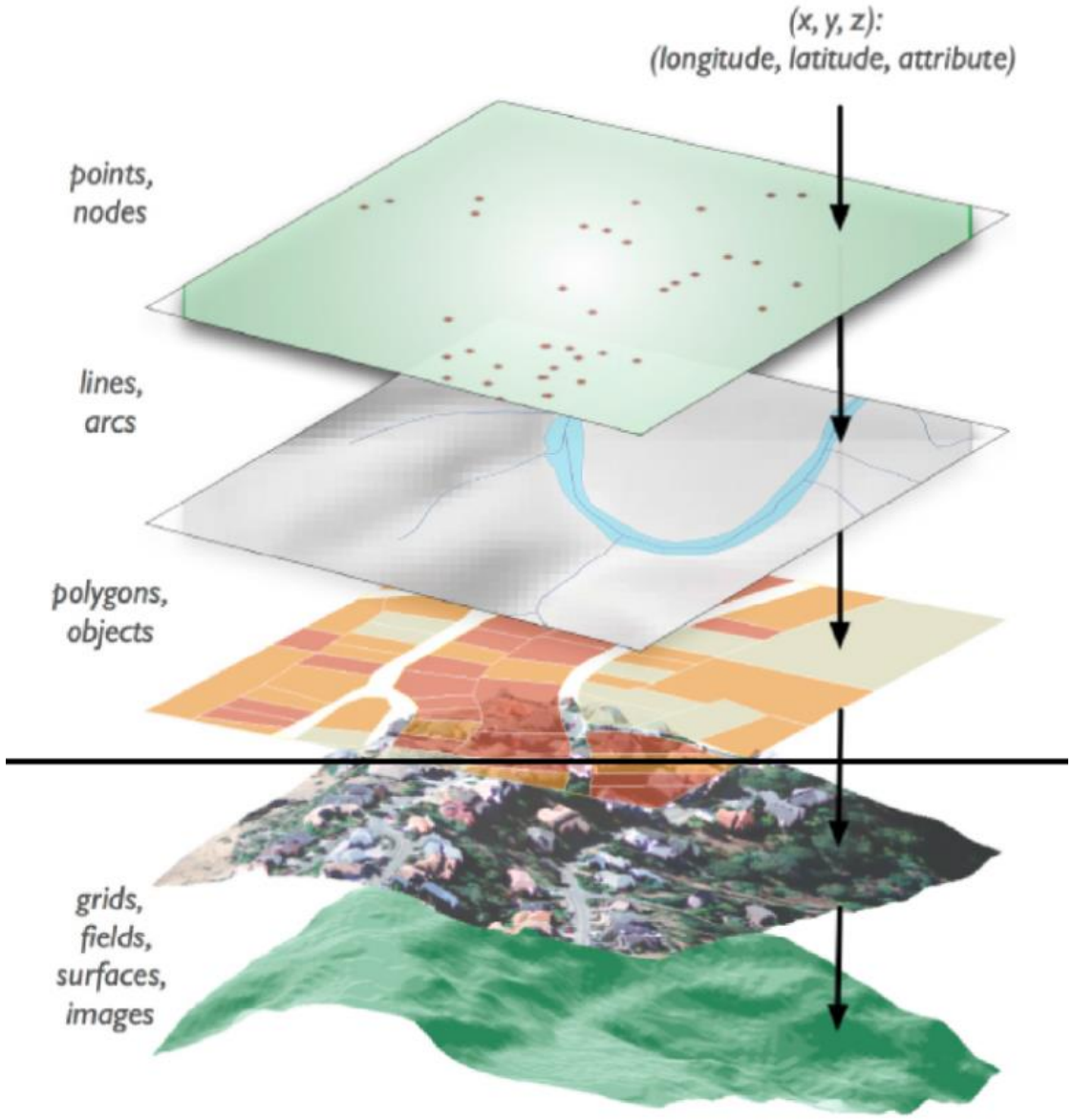
Geographic or geospatial data is information that describes objects, events or other features with a location on or near the surface of the earth

- Coordinates (latitutde, longitude)
- Attibutes (temperature, …)
- Temporal infomration (time stamp, …)

Standard data formats

Vector data: geometric objects
.gpkg, .shp, .svg, .geojson

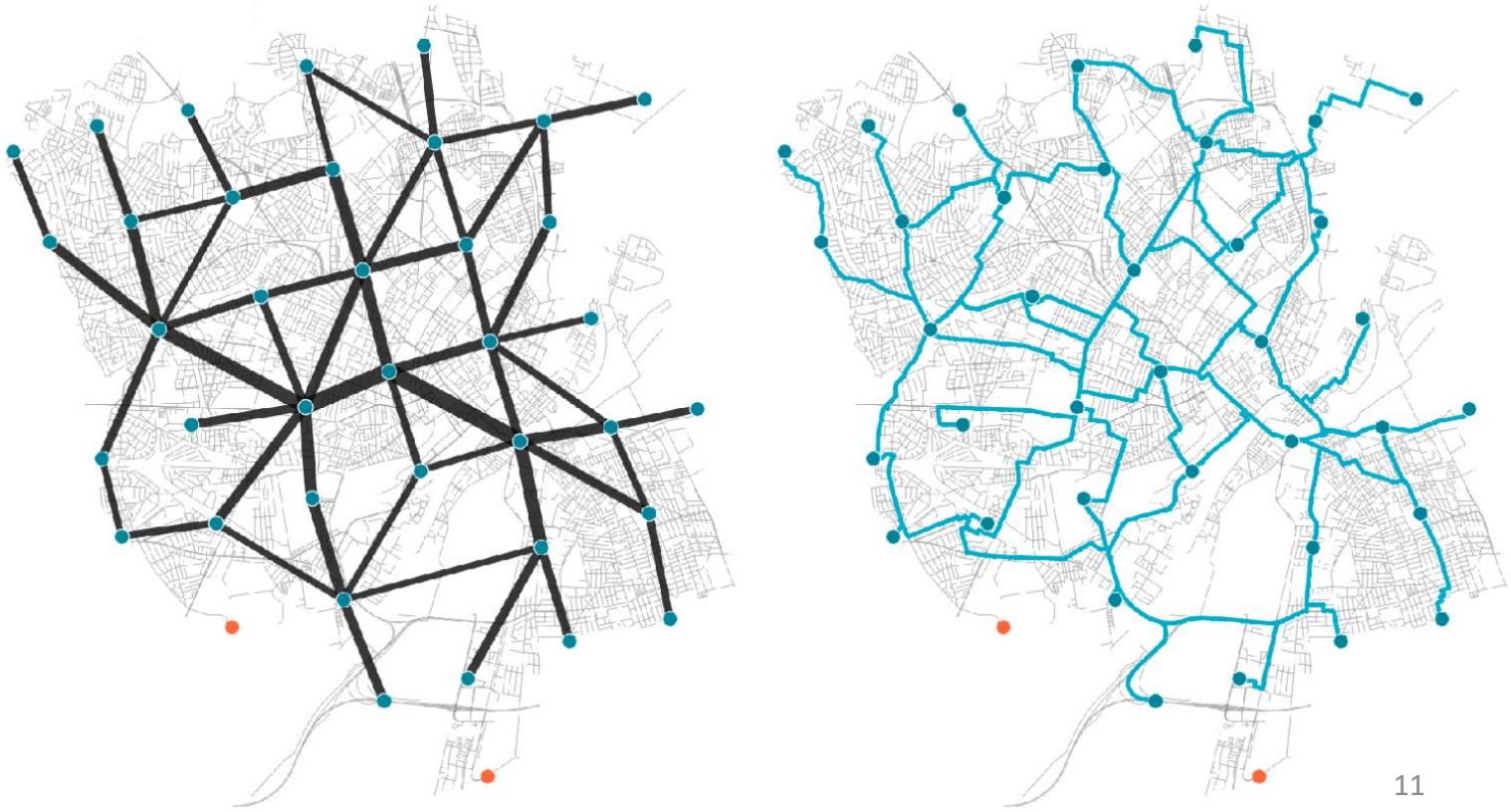
Raster data: grid of pixels
.tif, .jpg, .png, .bmp



Source: Kelly, M. UC Berkeley

Network data

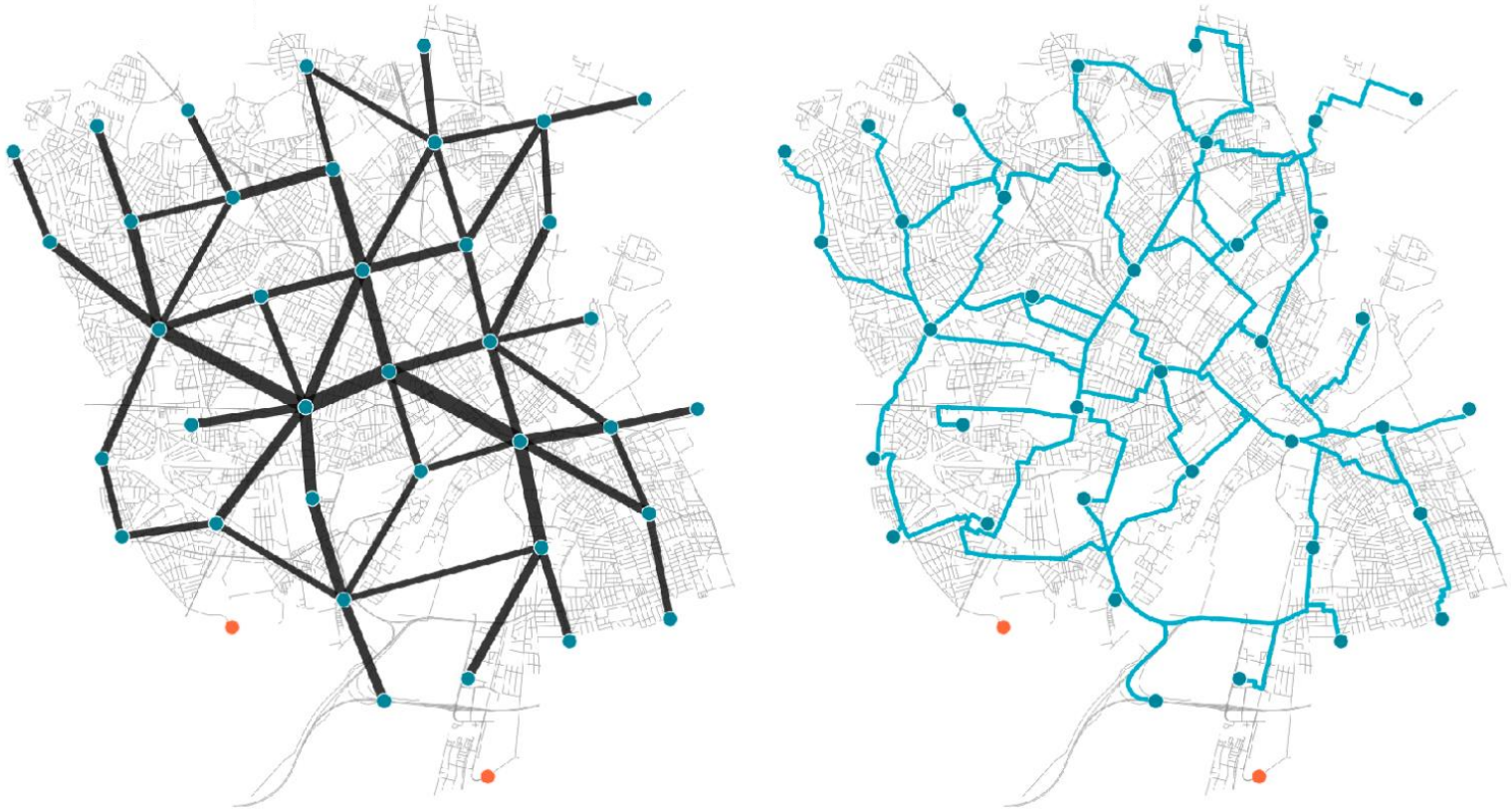
Structure (topology), embedded in space



Szell et al. (2022) **Scientific Reports**

Tobler’s first law of geography

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



Szell et al. (2022) **Scientific Reports**

How do things relate in space?

How to formalize, operationalize and visualize this question?

Practice sessions

Part1 – Introduction to geographic data

Part 2 – OSM and spatial networks

Part 3 – Spatial autocorrelation and gravity models

https://github.com/sandorjuhasz/geoDS-guest-lectures

Procedural part of Geographic Data Science

DB handling

Procedural

Point and click GIS

Procedural part – geographic data basics

- Coordinate Reference Systems (CRS)
- Libraries and data handling

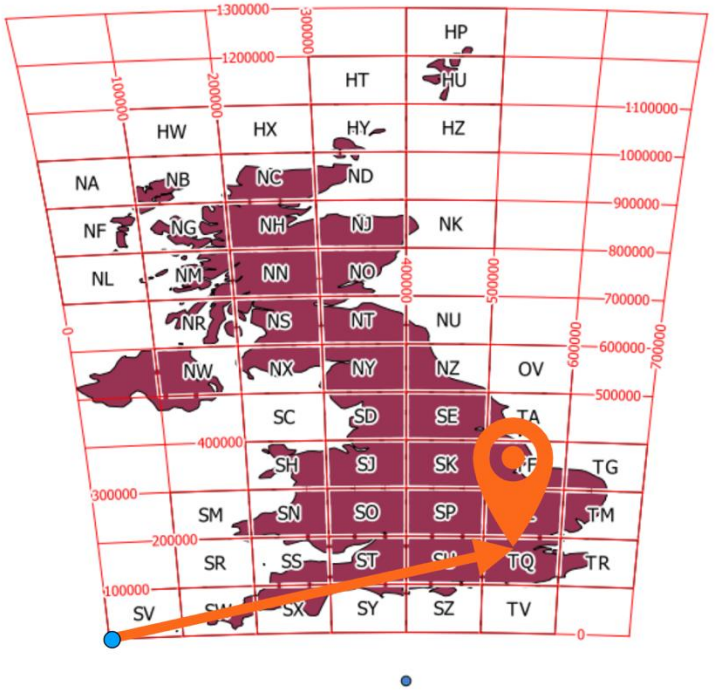
Coordinate Reference Systems (CRS)

Geographic RS

$(\phi, \lambda) = (-0.1, 51.5)$

Longitude	Degrees N/S from equator
Latitude	Degrees W/E from meridian

Projected RS

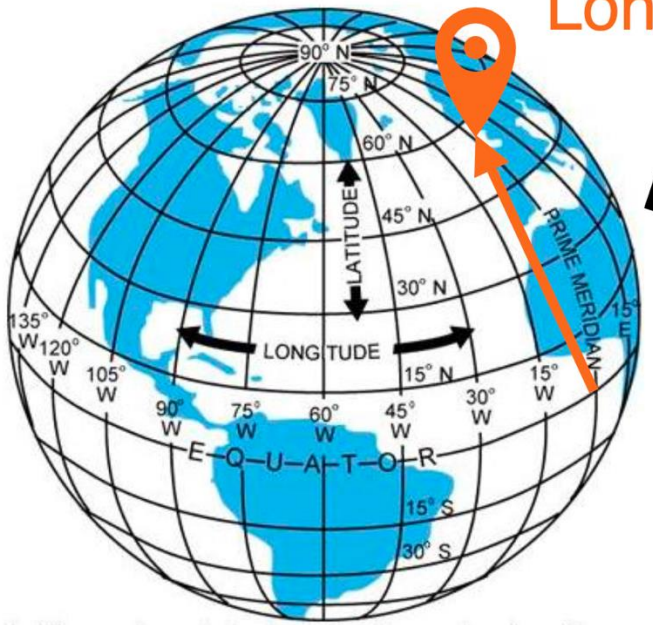


(x,y) = (530000,180000)

Easting	Meters east from origin (bottom left)
Northing	Meters north from origin (bottom left)

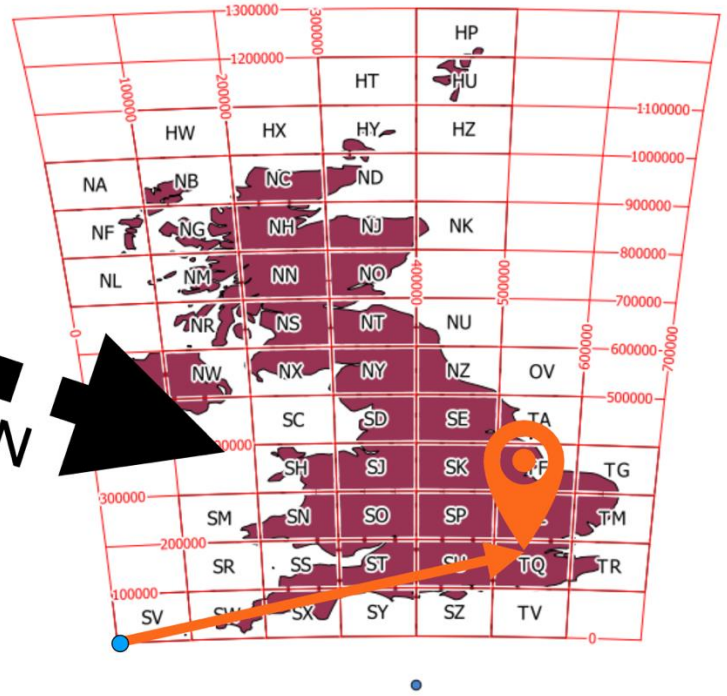
Coordinate Reference Systems (CRS)

Geographic RS



MAP PROJECTION

Projected RS



(ϕ, λ) = (-0.1,51.5)

(x,y) = (530000,180000)

Longitude	Degrees N/S from equator	Easting	Meters east from origin (bottom left)
Latitude	Degrees W/E from meridian	Northing	Meters north from origin (bottom left)

Common world-spanning reference systems

WGS 84 (World Geodetic System 1984) / EPSG:4326

WGS 84 - Web/Pseudo-Mercator / EPSG:3857

Google Maps

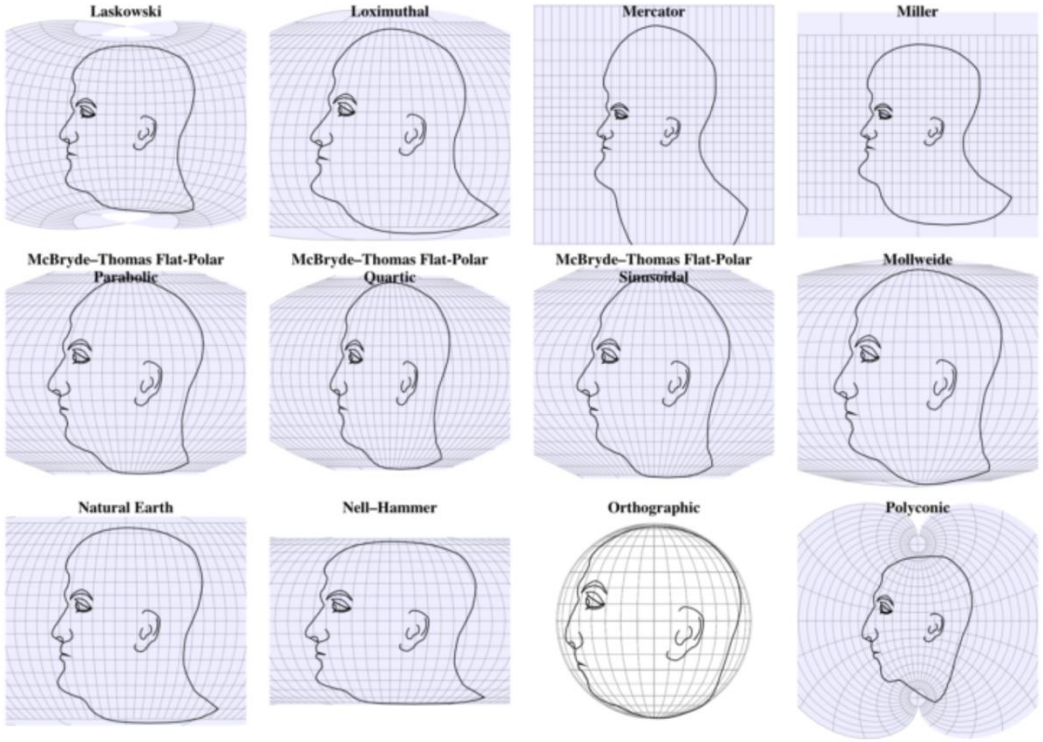
epsg.io

From MapTiler Team

All map projections are wrong

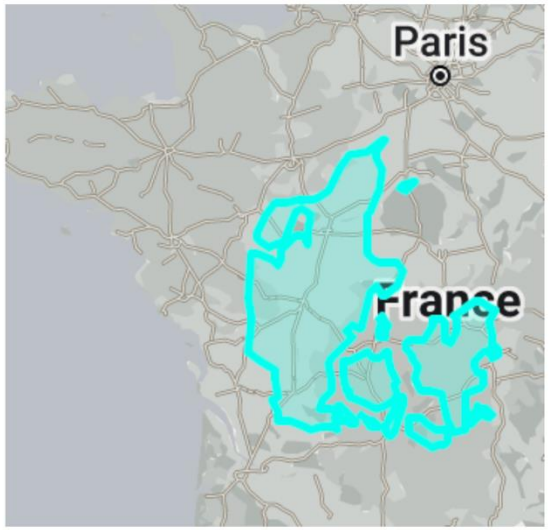
… but some are useful

https://www.thetruesize.com/



Usual pitfalls

Nr. 1 – wrong or missing CRS!



Nr. 2 – (lat,lon) vs. (lon,lat)

lon, lat	lat, lon
formats • GeoJSON ref • KML ref • Shapefile ref • WKT ref • WKB ref • geobuf ref	formats • GeoRSS ref • Encoded Polylines (Google) ref • iCalendar ref
javascript apis • OpenLayers ref • d3 ref • ArcGIS API for JavaScript ref • Mapbox GL JS ref	javascript apis • Leaflet ref • Google Maps API ref

Data handling and libraries

GeoPandas

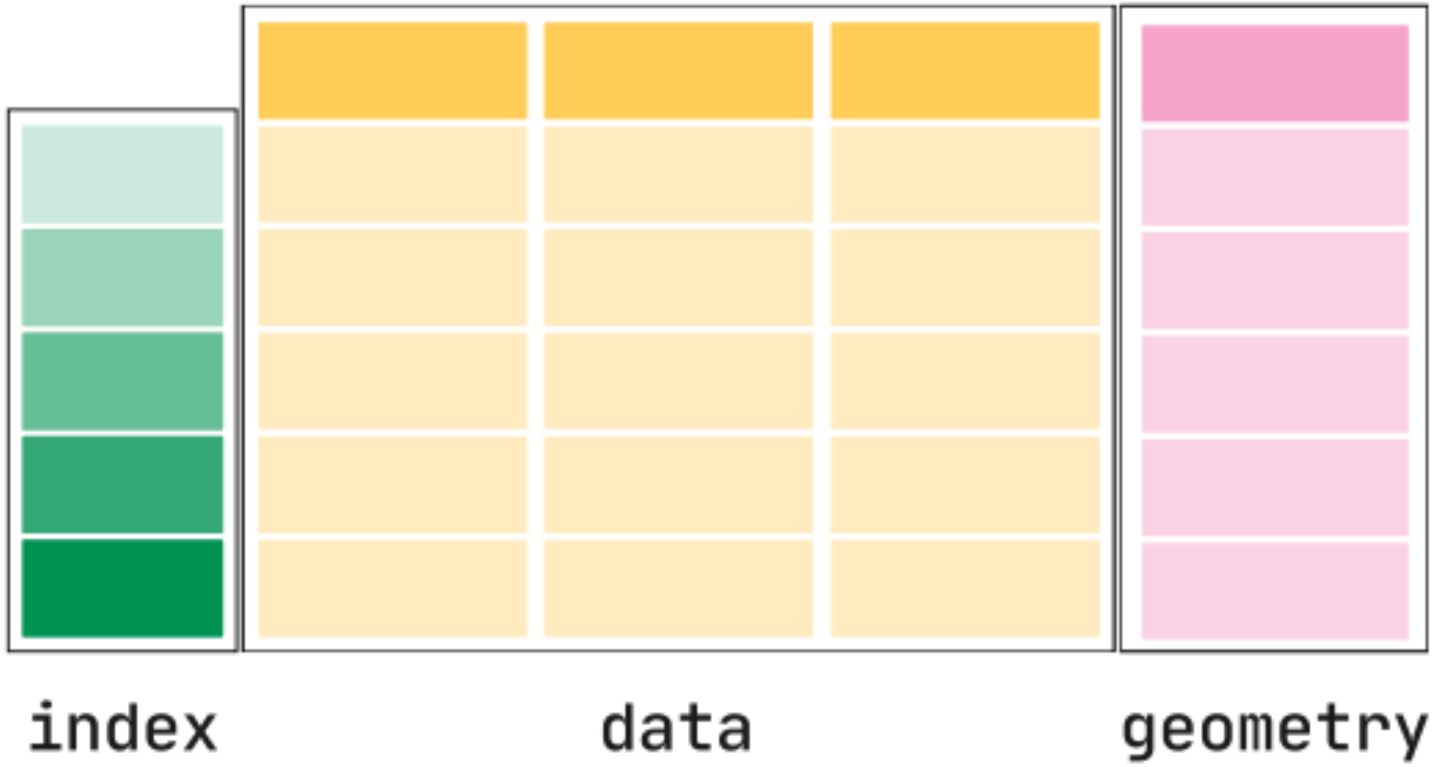
The spatial extension for pandas

Data handling and libraries

https://geopandas.org/

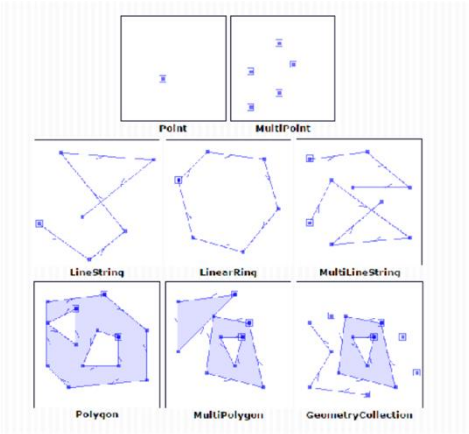
GeoDataFrame

GeoSeries



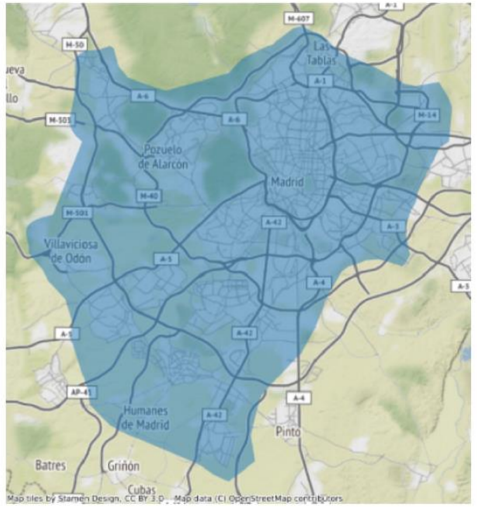
Data handling and libraries

Uses Shapely for geometries

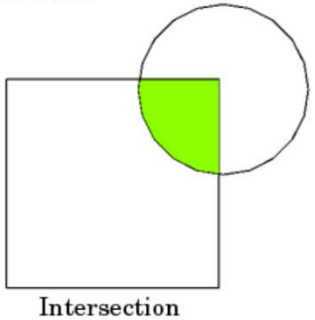


CRS support

Static maps



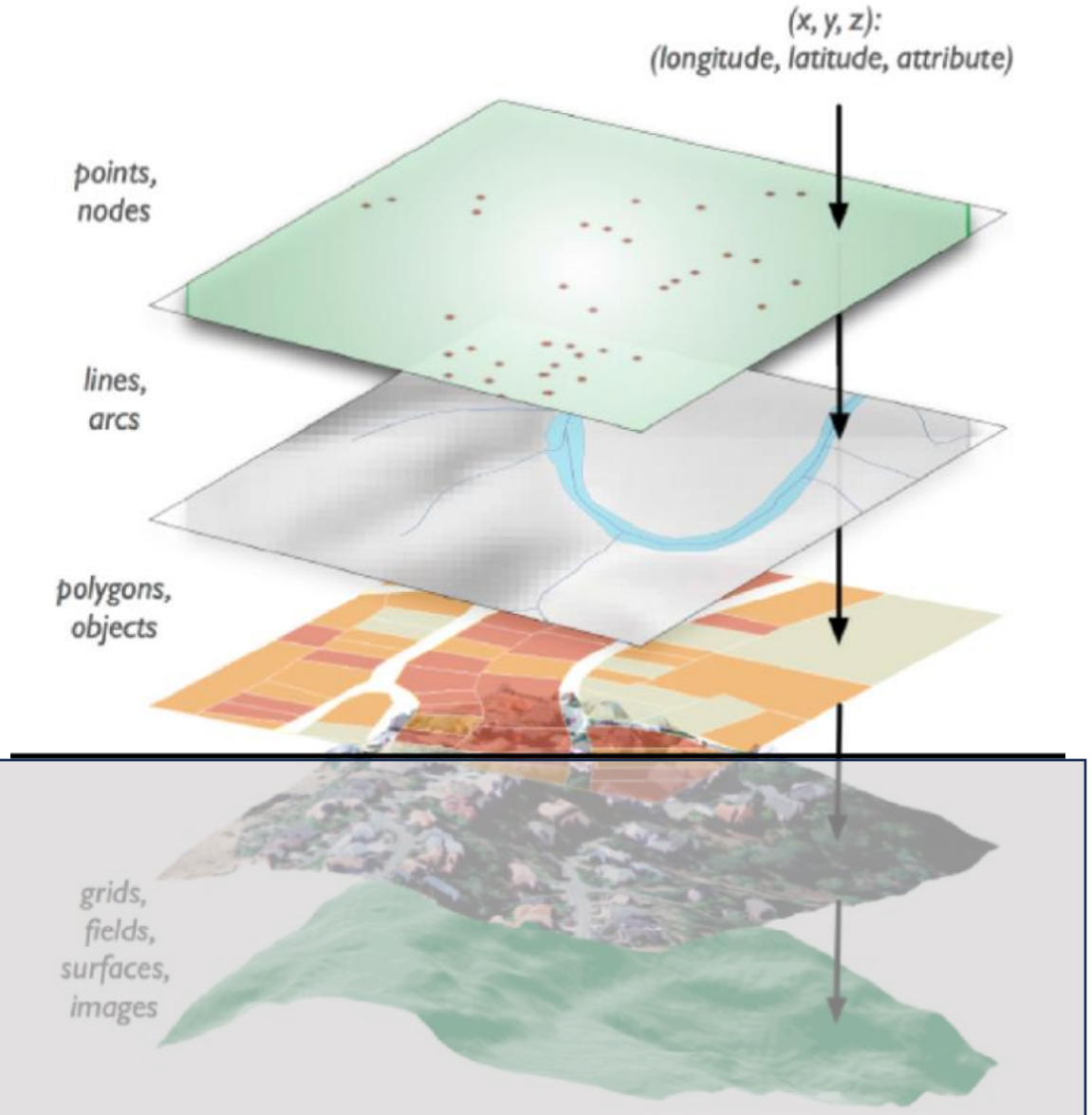
Spatial operations



GeoPandas handles vector data

Vector data: geometric objects
.gpkg, .shp, .svg, .geojson

Raster data: grid of pixels
.tif, .jpg, .png, .bmp



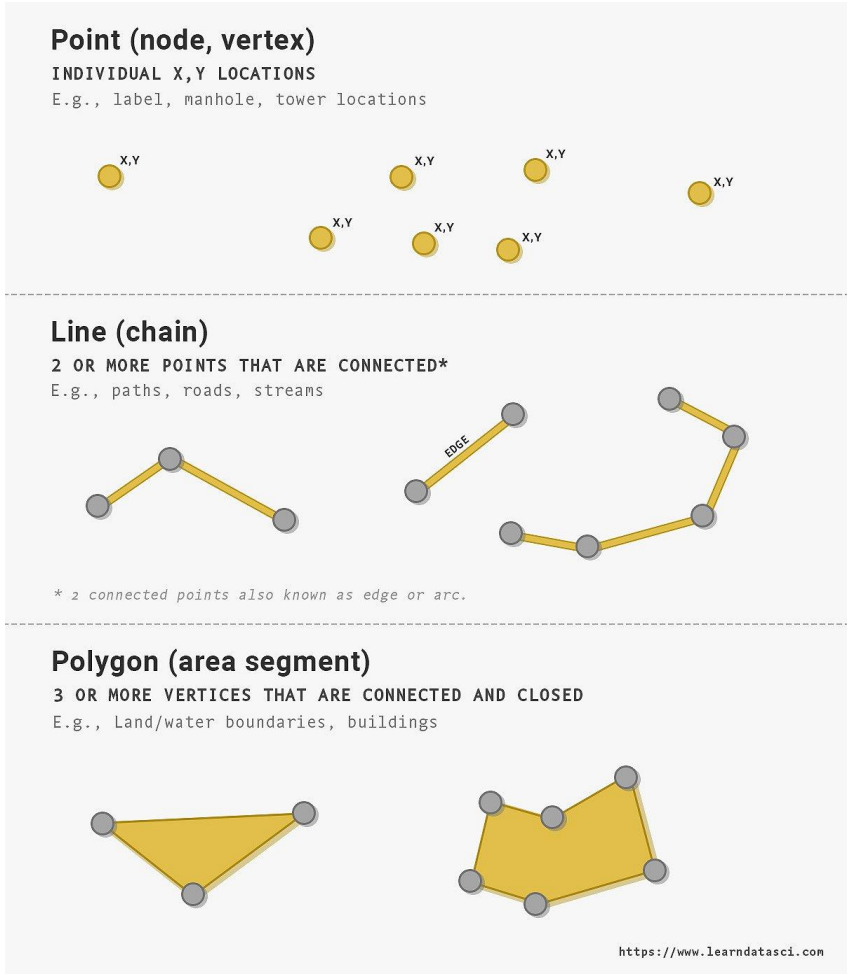
(rasterio – for raster data)

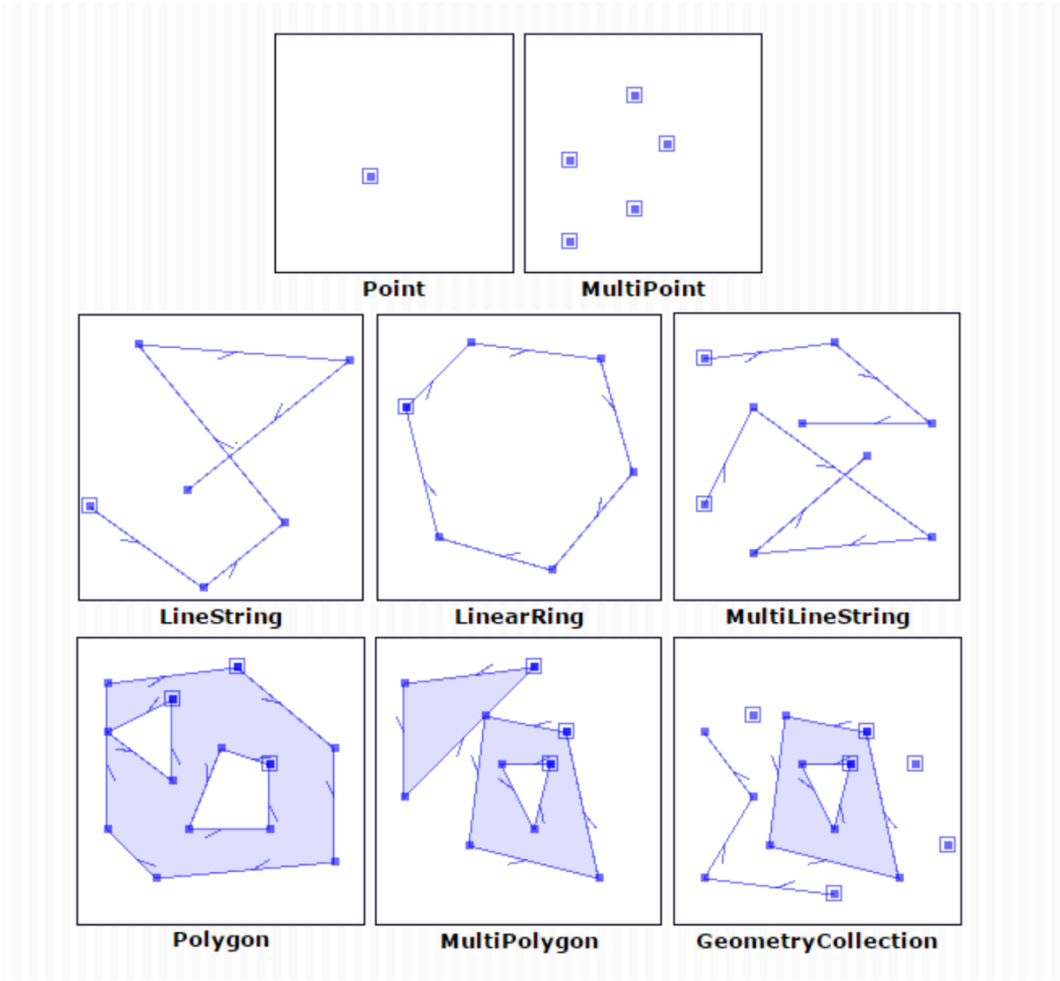
Source: Kelly, M. UC Berkeley

Data handling – file formats

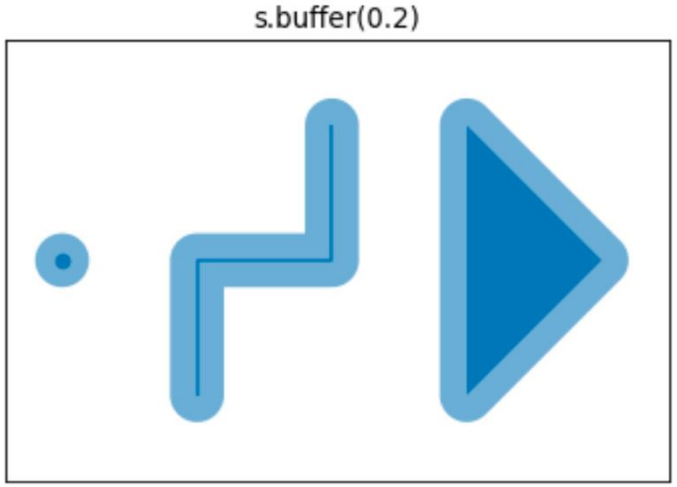
Geopackage – ‘universal’
GeoJSON – web-optimized
CSV
Shapefiles – old school
….

Basic geometric objects are handled by shapely

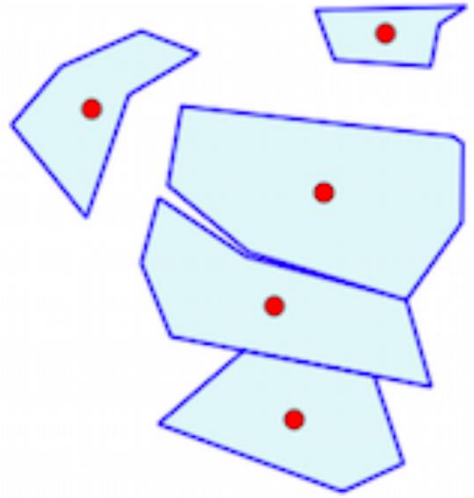




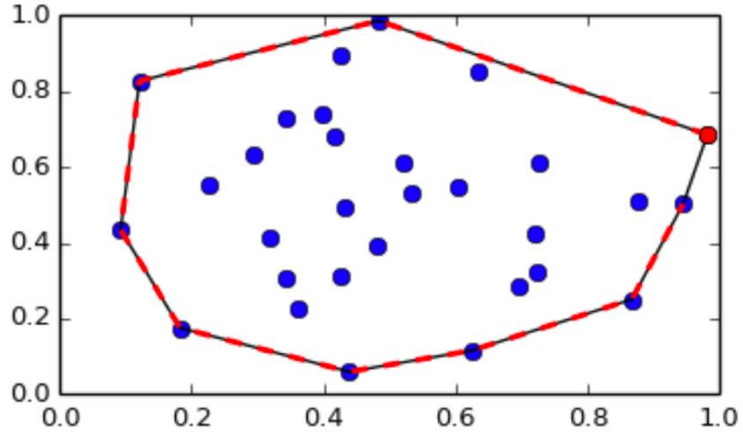
Spatial data operations in GeoPandas/Shapely



`gdf.buffer(10)`

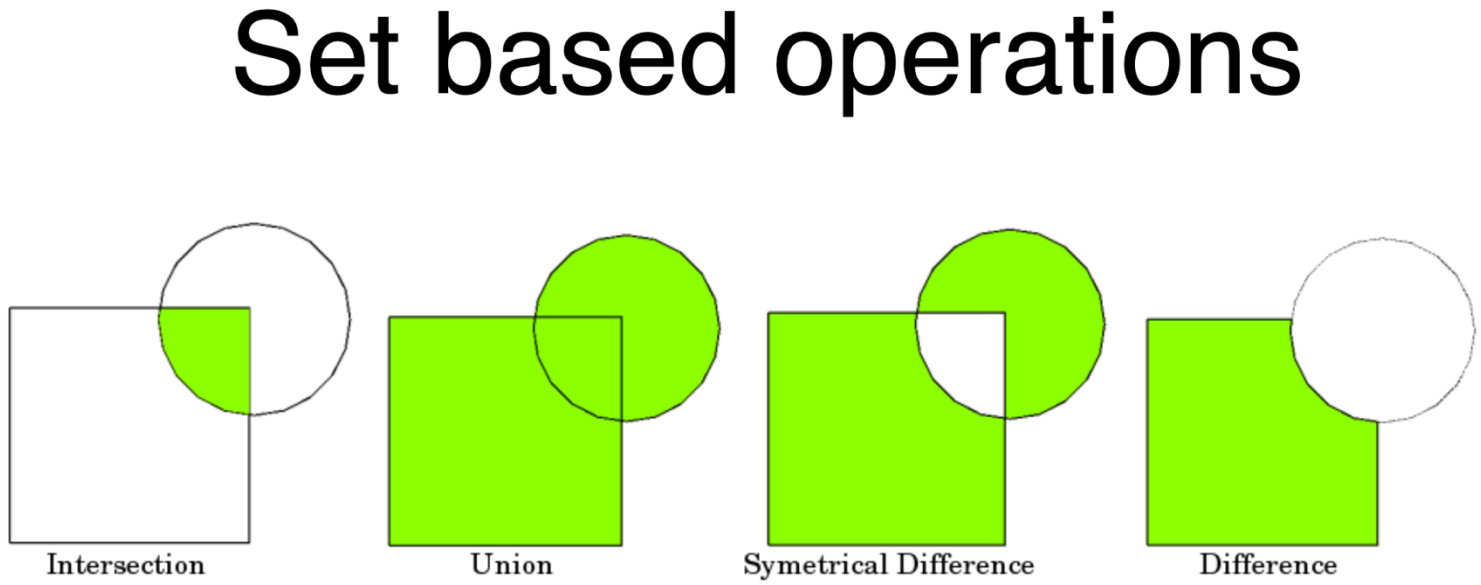


`gdf.centroid`



`gdf.convex_hull`

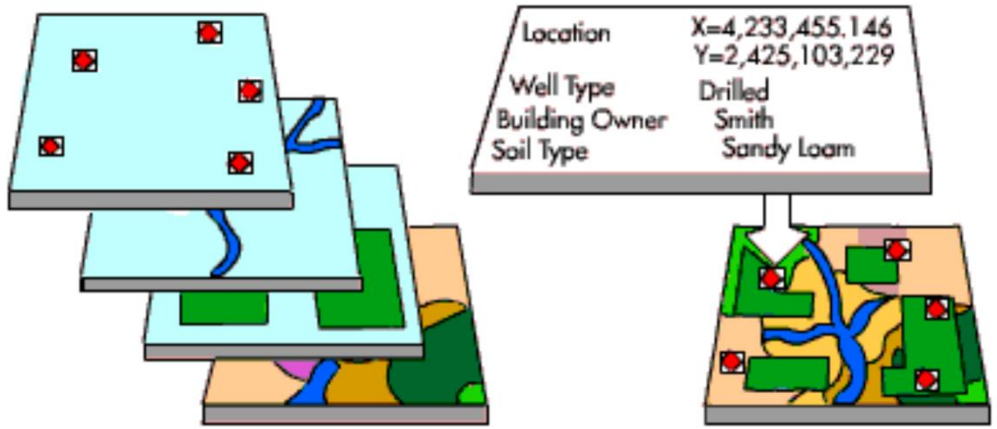
Spatial data operations in GeoPandas/Shapely




```
intersection = gdf1.overlay(gdf2, how='intersection')
```

Spatial data operations in GeoPandas

Spatial join

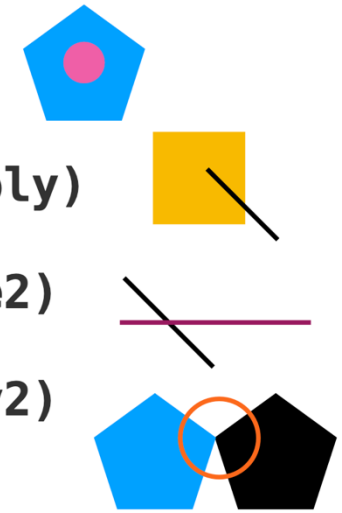



```
join = point_gdf.sjoin(poly_gdf)
```

Spatial data operations in GeoPandas/Shapely

Spatial queries

```
point.within(poly)
line.intersects(poly)
line1.crosses(line2)
poly1.touches(poly2)
```



…and many more!

Thank you! Let’s explore together!

sandorjuhasz.com