Geospatial Data Science, Spring 2022

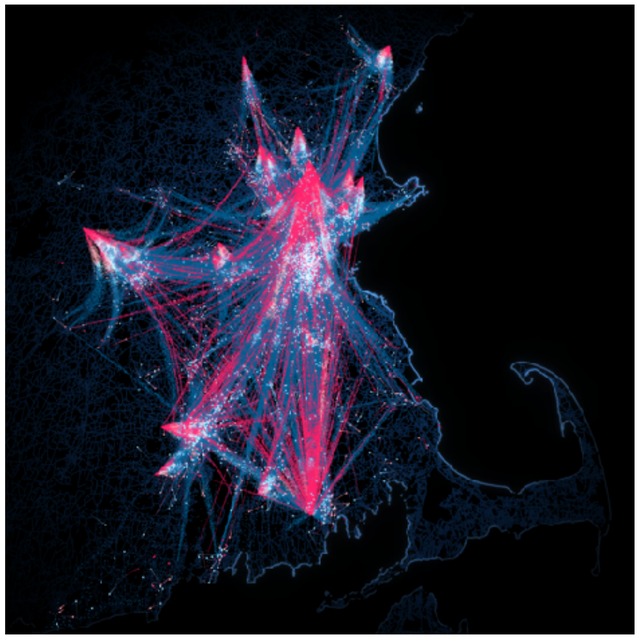
Lecture 1: Introduction & Geometric objects

Instructor: Michael Szell
Feb 3, 2022

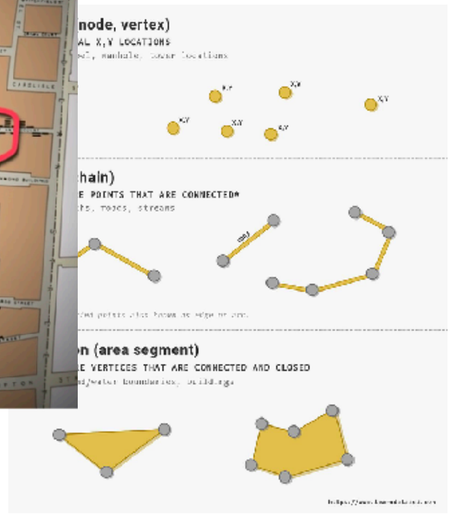
IT UNIVERSITY OF COPENHAGEN

Today you will learn about GDS, John Snow, and shapely

What is Geospatial Data Science?



John Snow & shapely



Course Overview & Logistics

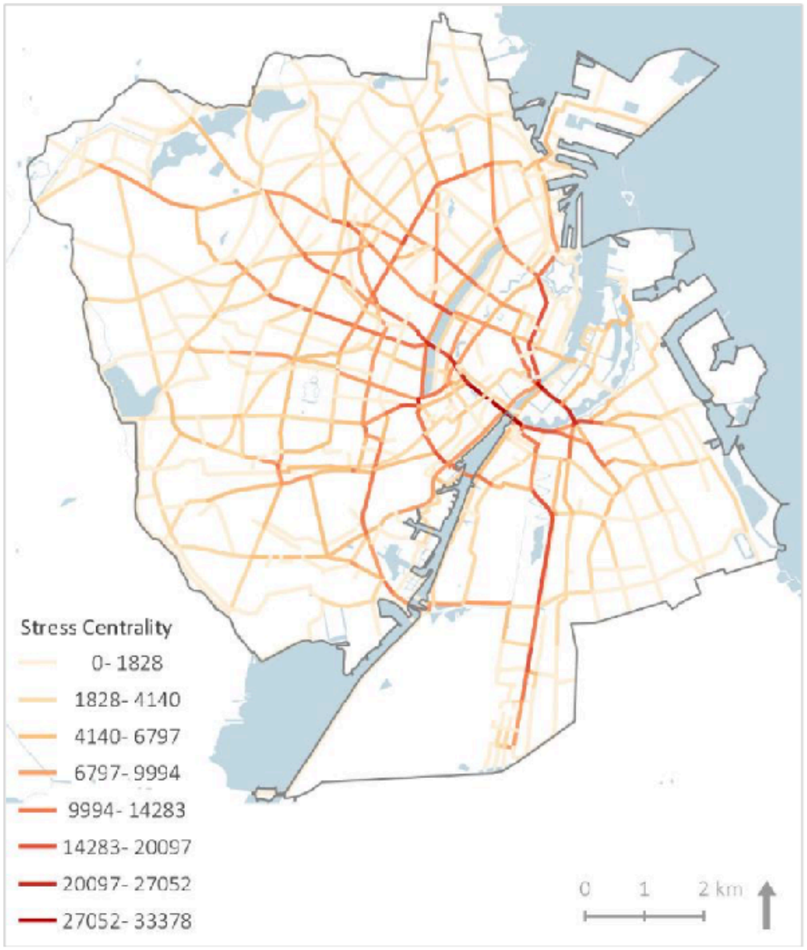
Schedule GDS2022 — Last update: 2022-01-18

Room 4A66, 10:00–14:00

Lecture	Instructor	CalW	Date	Topic	Event
1	M	5	Feb 3	Introduction	Invitations
2	M	6	Feb 10	Geospatial data in Python	
3	M	7	Feb 17	Choropleth mapping	
4	M	8	Feb 24	Spatial weights	
5	M	9	Mar 3	Spatial autocorrelation	
6	M	10	Mar 10	Spatial clustering	
7	M	11	Mar 17	Point pattern analysis	
8	M	12	Mar 24	Spatial networks and OSM	
9	M	13	Mar 31	OSMnx	Projects proposed
10	A+A	14	Apr 7	Bicycle network analysis	Projects approved
Holiday week		15			
11	M/A/A	16	Apr 21	Applications	Open class exercise
12	M/A/A	17	Apr 28	Applications	Open class exercise
13	M/M/A	18	May 5	Applications	Open class exercise
14	M/M/M?	19	May 12	Applications	Open class exercise
		20	May 20		Project submission

Introduction: Anastassia Vybornova + Ane Rahbek Vierø

Data Scientists researching:

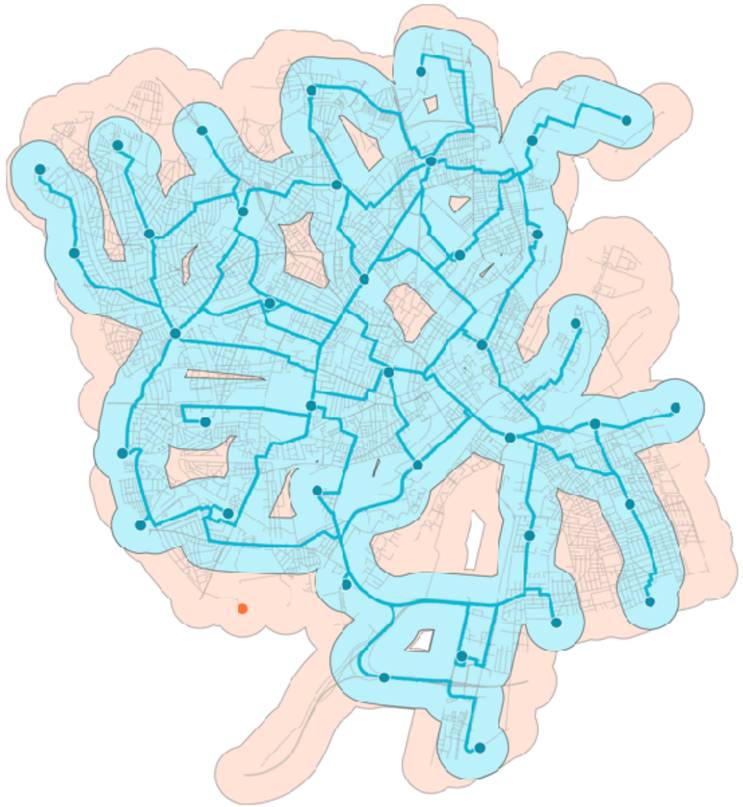


Bicycle networks

Python

Introduction: Michael Szell

Data Scientist researching:



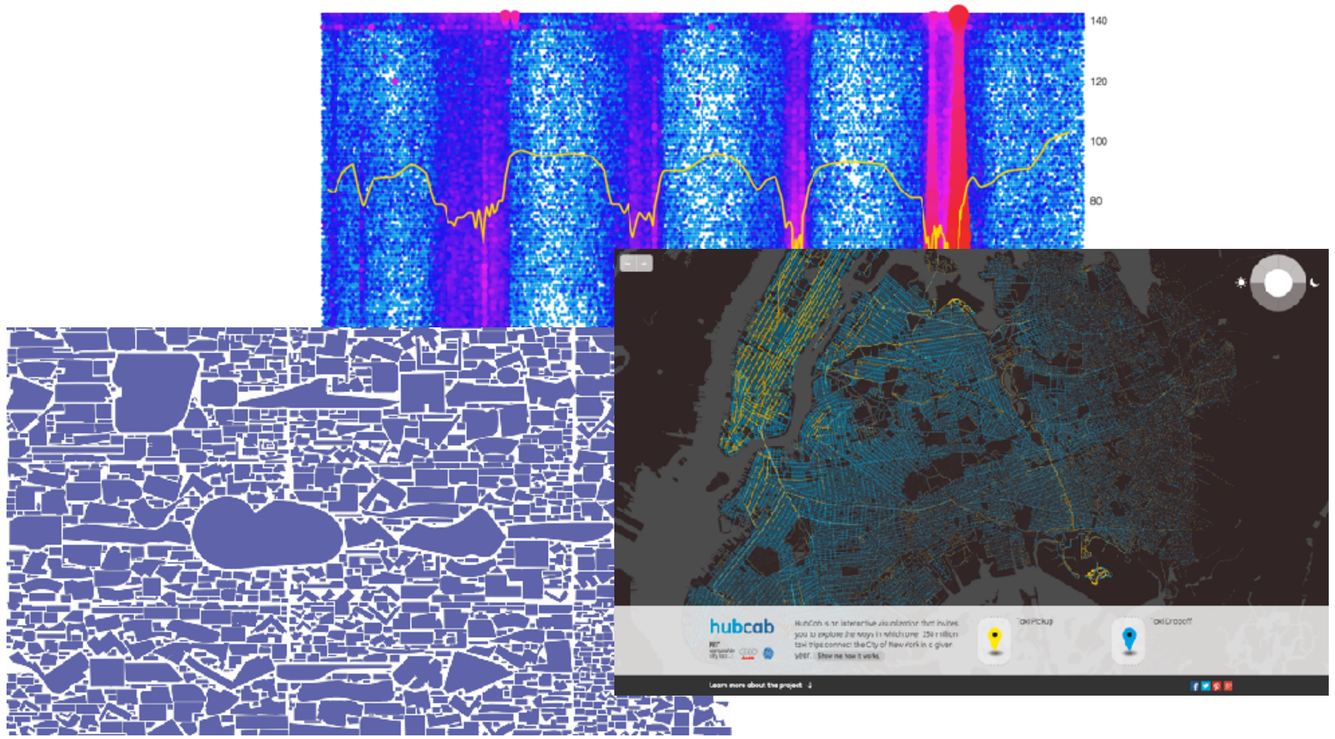
Bicycle networks

Human mobility

Python

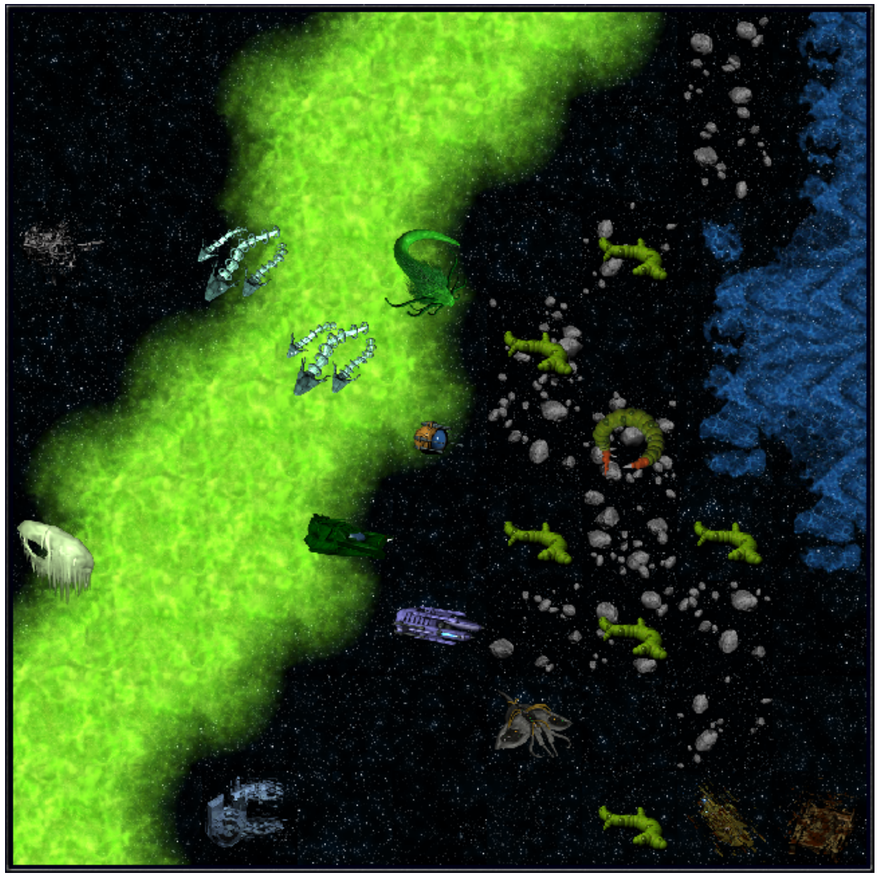
Introduction: Michael Szell

Data Scientist creating:



Data Visualizations

Python, Javascript, MongoDB



MMOG

PHP, C/C++, MySQL

Introduction: Michael Szell

Data Scientist having background in / working with:

Physicists

Architects, Urban planners

Industrial designers

Computer scientists

Game developers

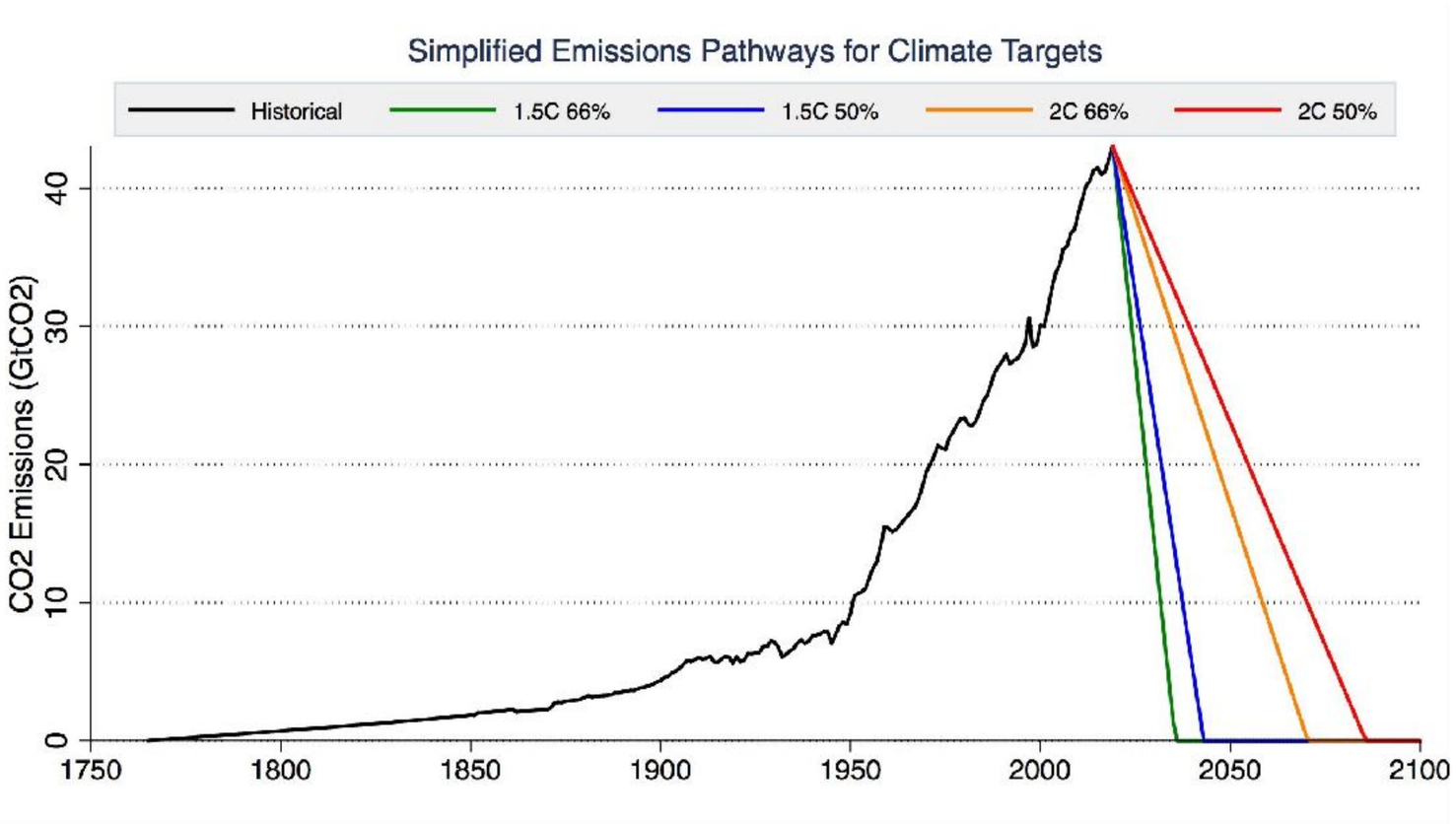
Mathematicians

Economists

Medical doctors

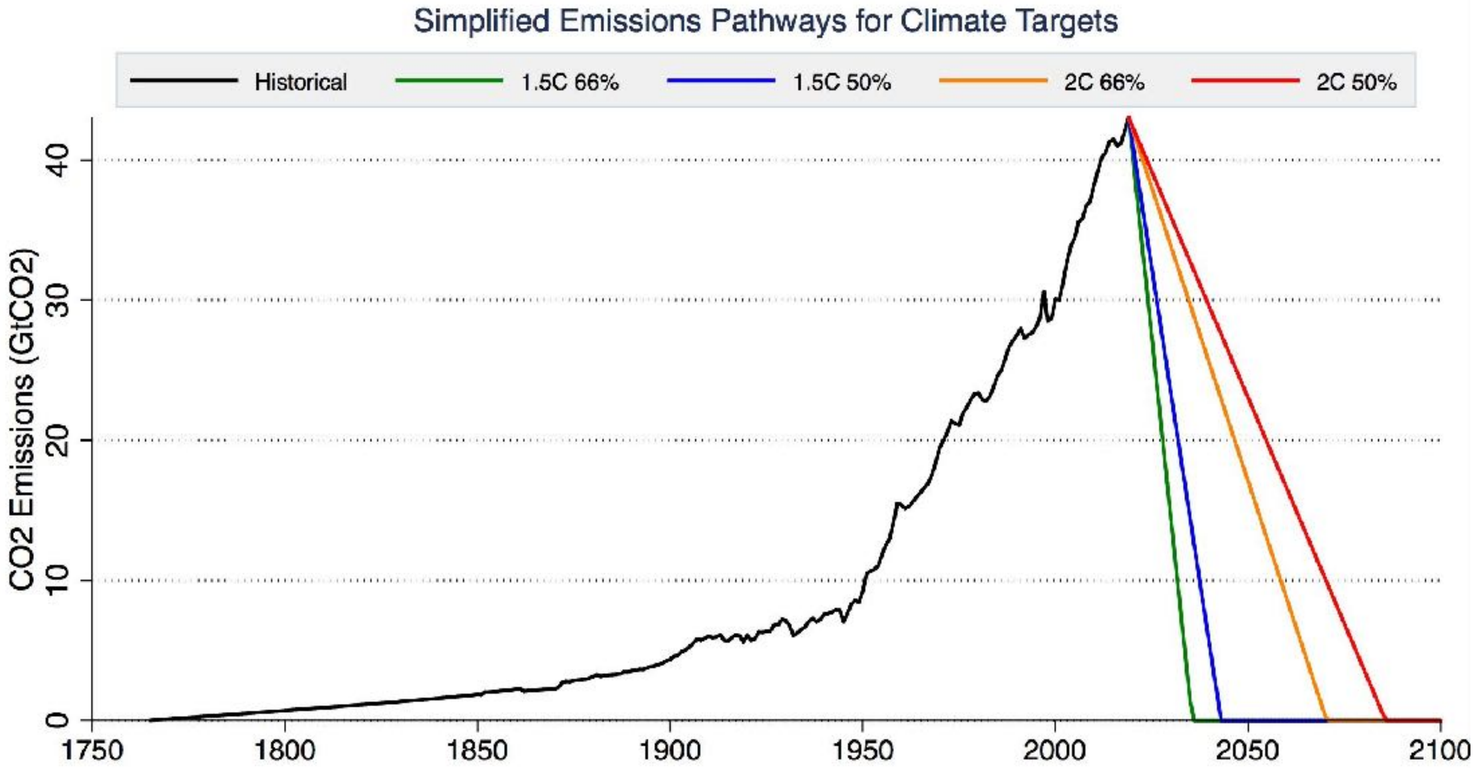
Michael's story: How should we grow bike networks?

Climate change is a big problem



Michael's story: How should we grow bike networks?

Climate change is a big problem - Cycling has huge potential to help!



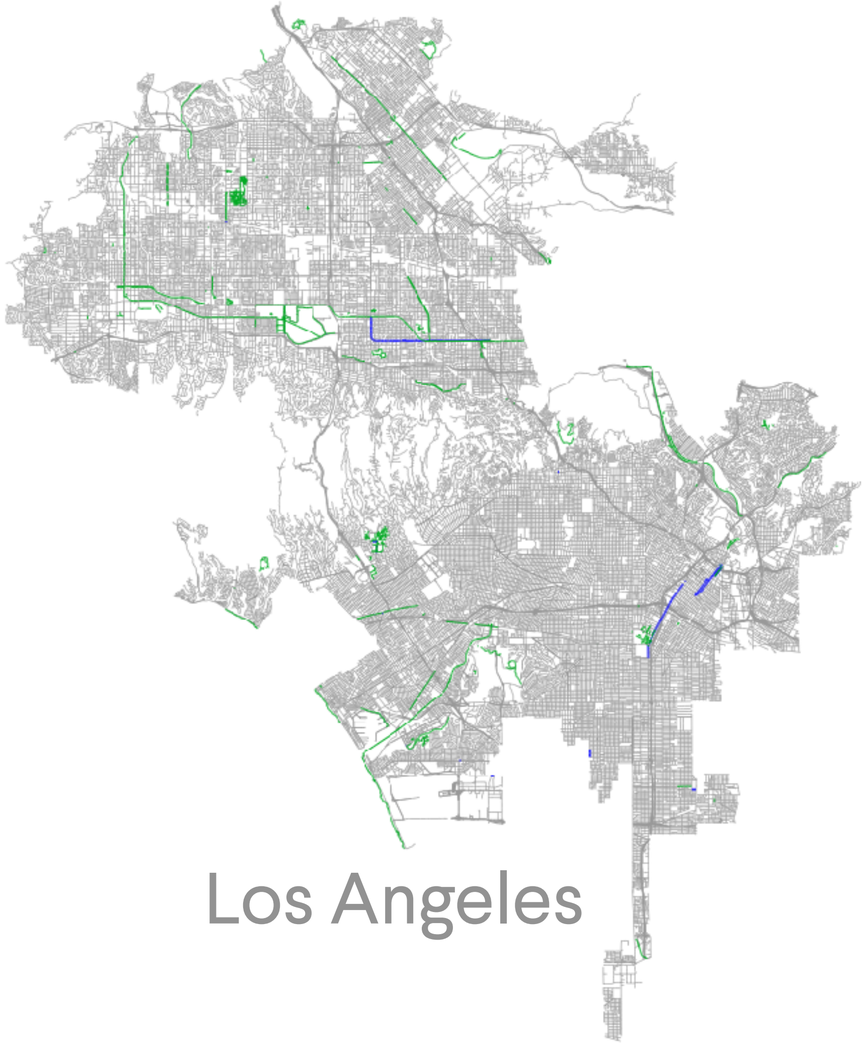
Home 〉 News 〉 Obsessing over electric cars is impeding the race to net zero: More active travel is essential

Obsessing over electric cars is impeding the race to net zero: More active travel is essential

https://www.ox.ac.uk/news/2021-06-14-obsessing-over-electric-cars-impeding-race-net-zero-more-active-travel-essential

Michael's story: How should we grow bike networks?

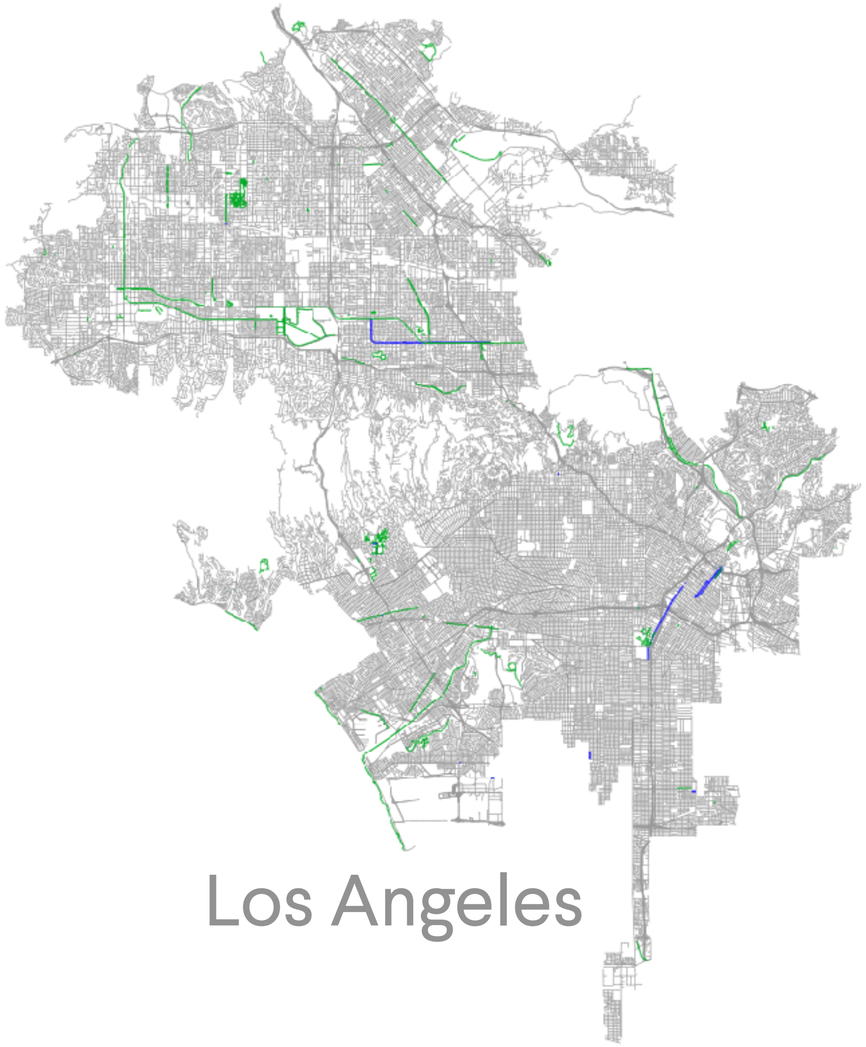
Most cities have no bike infrastructure



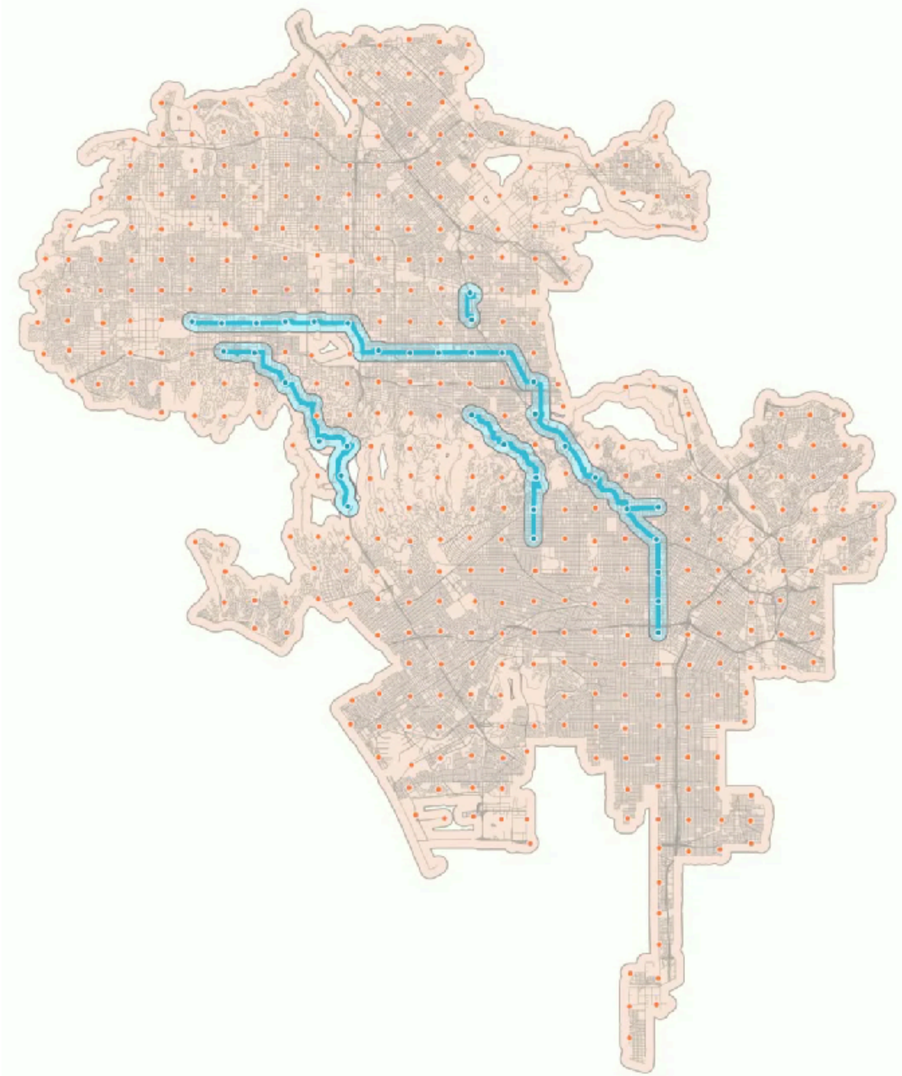
Los Angeles

Michael's story: How should we grow bike networks?

Most cities have no bike infrastructure

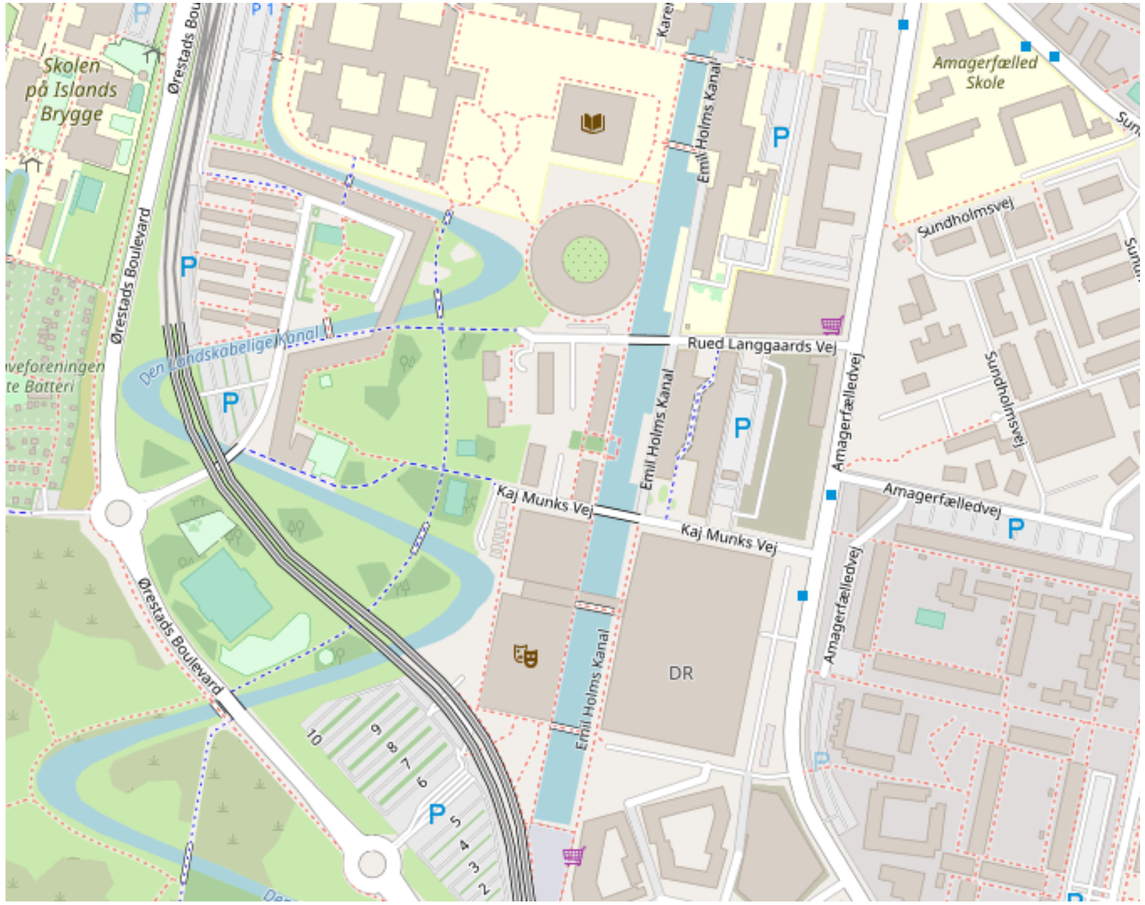


Los Angeles



Michael's story: How should we grow bike networks?

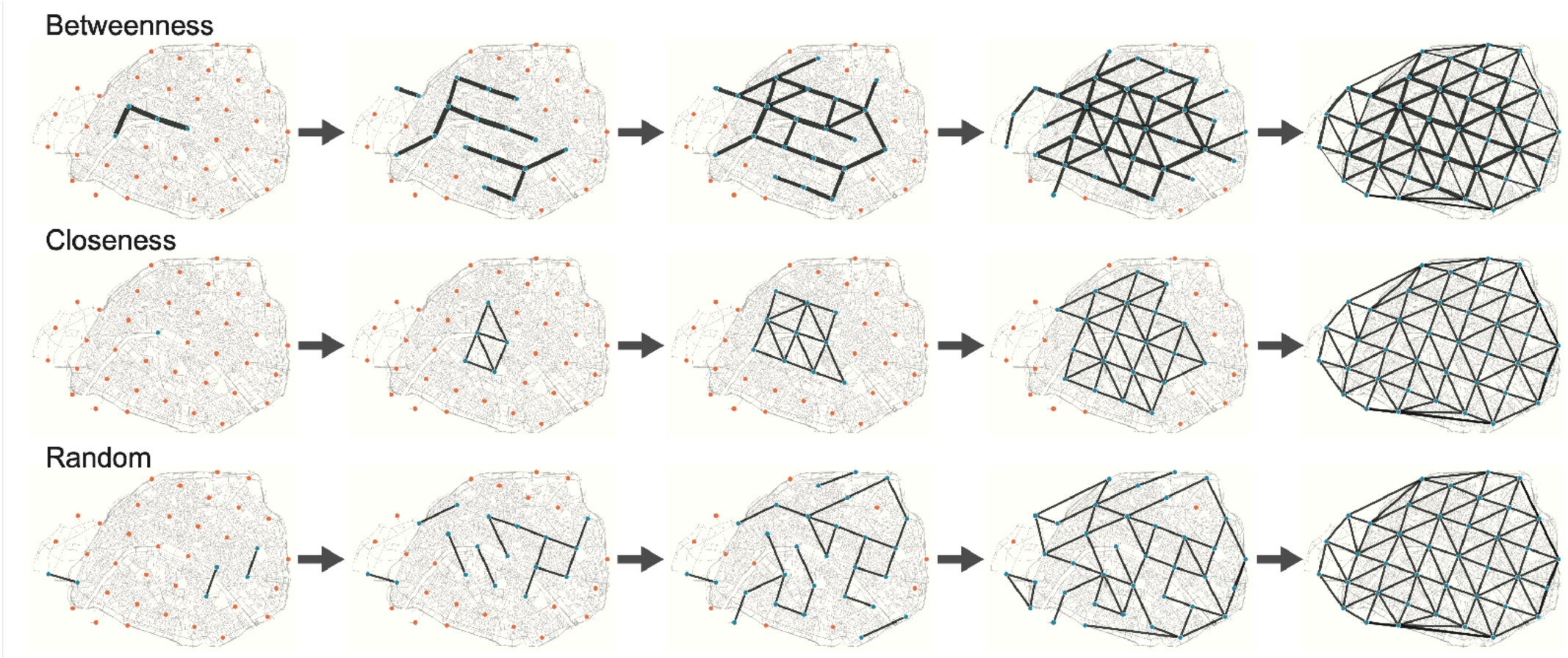
We downloaded data from OpenStreetMap




```
20811    <nd ref="661727421"/>
20812    <nd ref="298719840"/>
20813    <nd ref="298719836"/>
20814    <nd ref="298719835"/>
20815    <tag k="amenity" v="university"/>
20816    <tag k="architect" v="Henning Larsen"/>
20817    <tag k="building" v="university"/>
20818    <tag k="name" v="IT-Universitetet"/>
20819    <tag k="name:en" v="IT University"/>
20820    <tag k="name:ko" v="코펜하겐 IT 대학교"/>
20821    <tag k="official_name" v="IT-Universitetet i København"/>
20822    <tag k="official_name:en" v="IT University of Copenhagen"/>
20823    <tag k="phone" v="+45 72185000"/>
20824    <tag k="short_name" v="ITU"/>
20825    <tag k="toilets:wheelchair" v="yes"/>
20826    <tag k="website" v="https://www.itu.dk/"/>
20827    <tag k="wheelchair" v="yes"/>
20828    <tag k="wikidata" v="Q124882"/>
20829    <tag k="wikipedia" v="da:IT-Universitetet i København"/>
20830  </way>
20831  <way id="27220159" visible="true" version="4" changeset="31300934" timestamp="2015-05-
       19T22:57:15Z" user="Niels Elgaard Larsen" uid="1288">
20832    <nd ref="298719841"/>
20833    <nd ref="298719842"/>
20834    <nd ref="298719843"/>
20835    <nd ref="298719864"/>
20836    <nd ref="298719845"/>
20837    <nd ref="298719841"/>
20838    <tag k="building" v="yes"/>
20839  </way>
20840  <way id="27220160" visible="true" version="13" changeset="75317252" timestamp="2019-10-
       05T16:23:56Z" user="wheelmap_visitor" uid="290680">
20841    <nd ref="298719849"/>
20842    <nd ref="298719858"/>
20843    <nd ref="298719861"/>
20844    <nd ref="298719852"/>
```

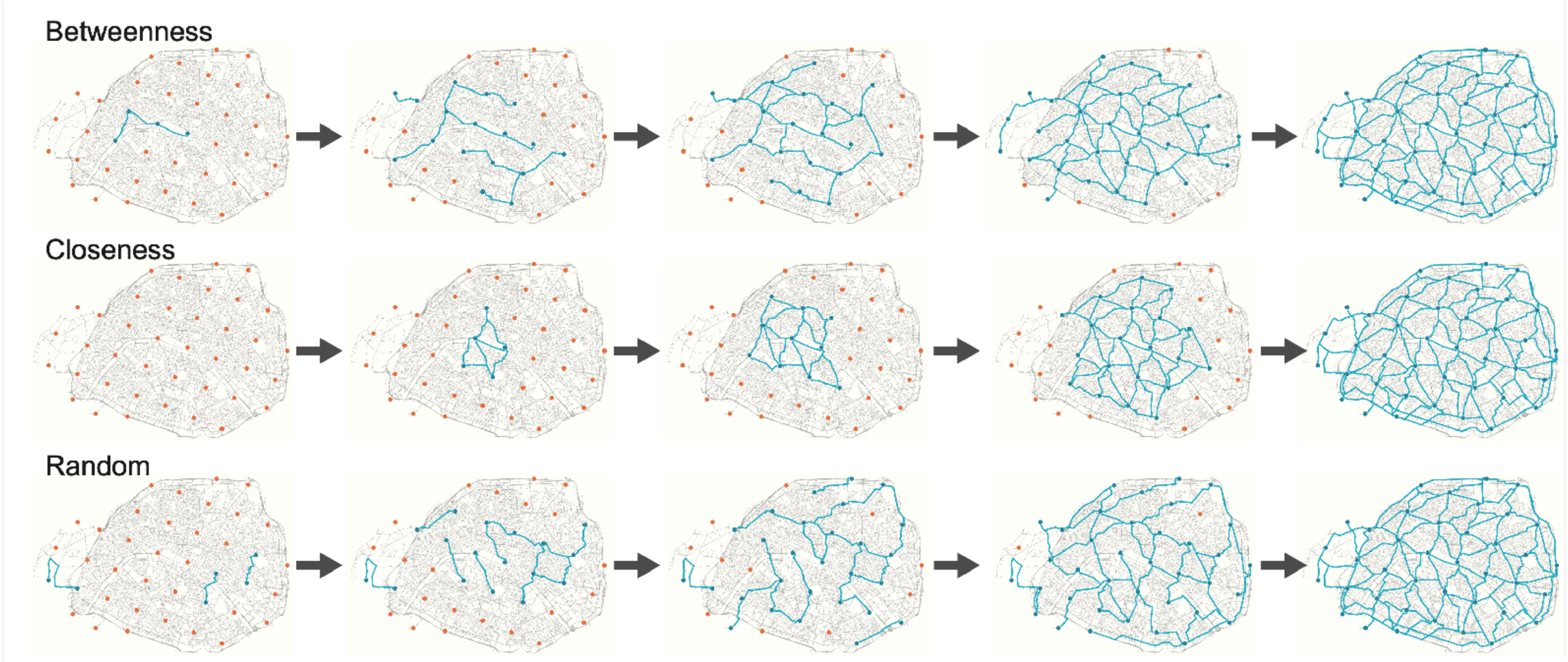
Michael's story: How should we grow bike networks?

We built a triangulation algorithm



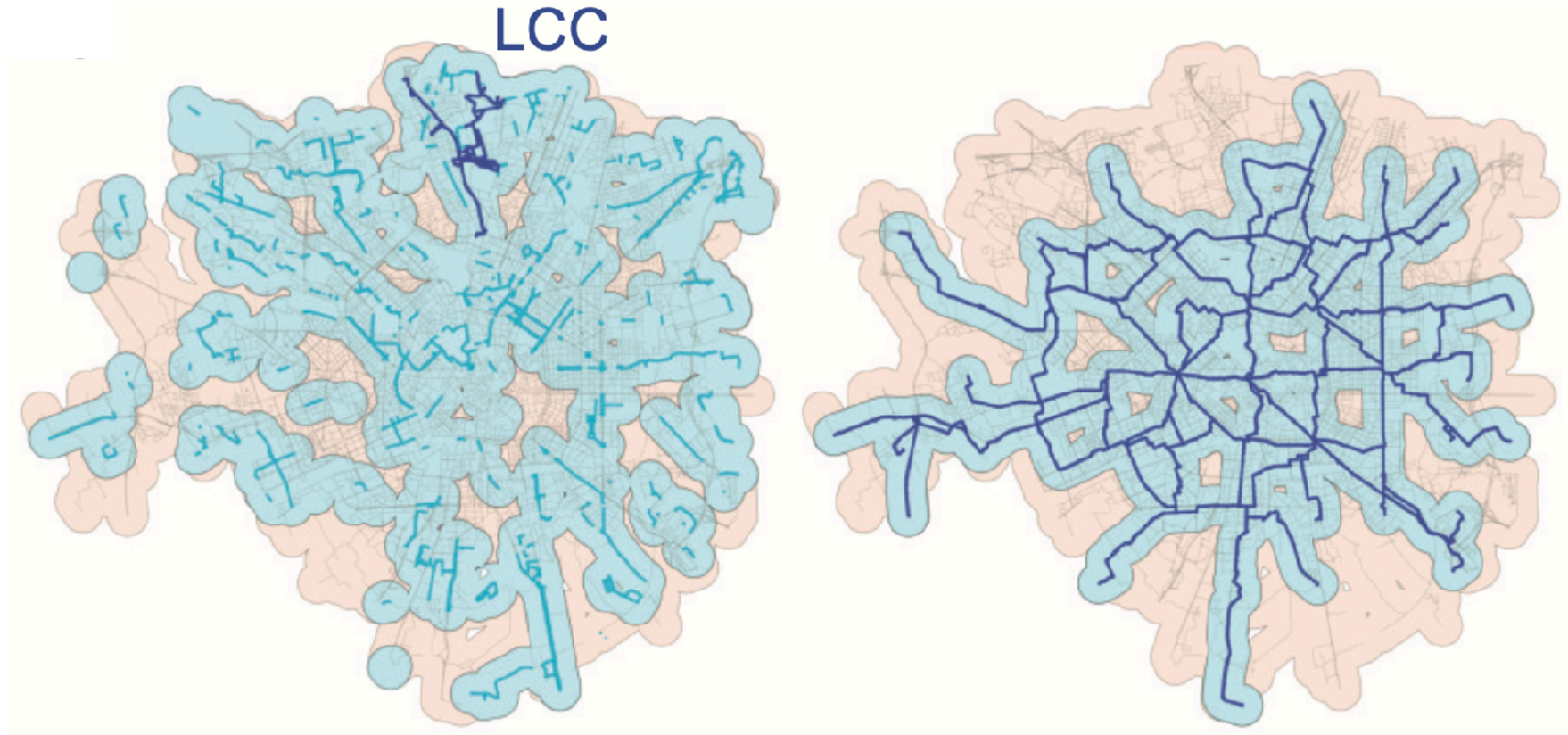
Michael's story: How should we grow bike networks?

We built a triangulation algorithm



Michael's story: How should we grow bike networks?

Real bike networks are scattered -
We could do much better!



Real Milan

Synthetic Milan

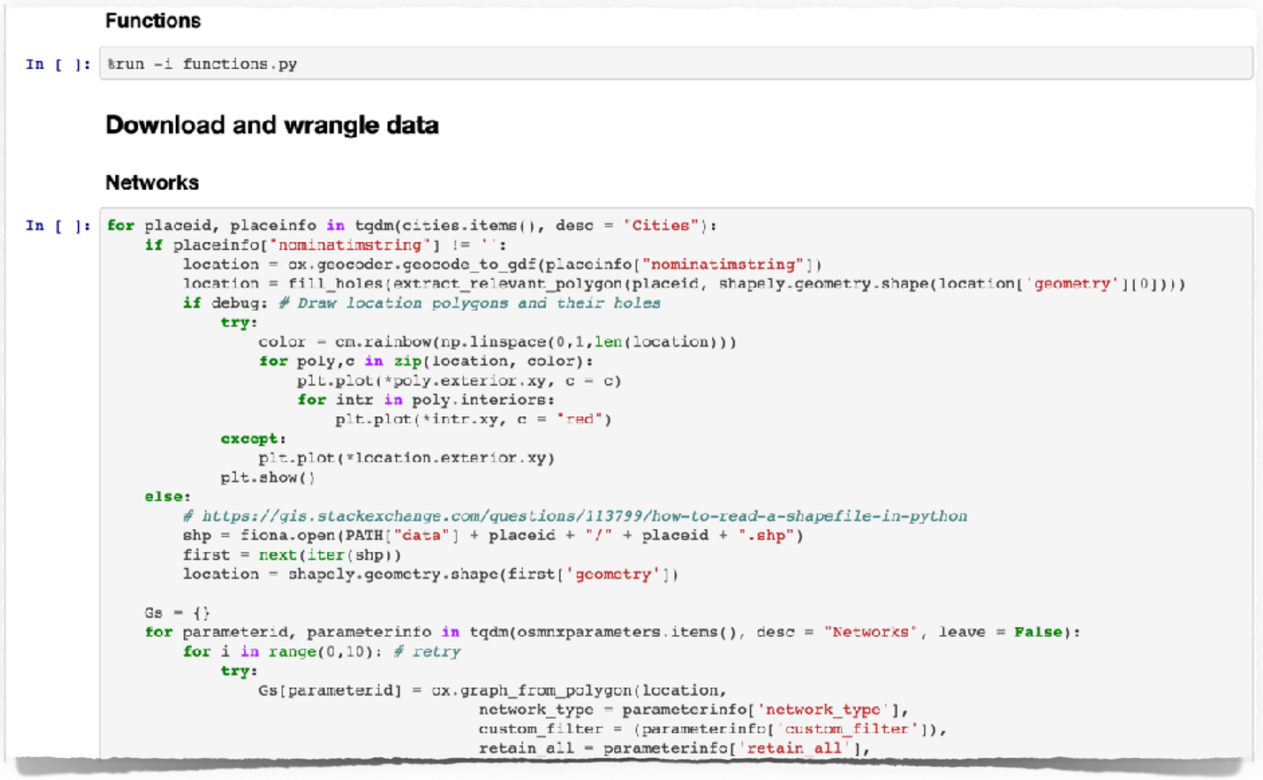
Michael's story: How should we grow bike networks?

My ITU Master students built growbike.net

Michael's story: How should we grow bike networks?

We open-sourced all code here:

https://github.com/mszell/bikenwgrowth



Functions

In [ ]:
```
%run -i functions.py
```

Download and wrangle data

Networks

In [ ]:
```
for placeid, placeinfo in tqdm(cities.items(), desc = 'Cities'):
    if placeinfo['nominatimstring'] != '':
        location = ox.geocoder.geocode_to_gdf(placeinfo["nominatimstring"])
        location = fill_holes(extract_relevant_polygon(placeid, shapely.geometry.shape(location['geometry'][0])))
        if debug: # Draw location polygons and their holes
            try:
                color = cm.rainbow(np.linspace(0,1,len(location)))
                for poly,c in zip(location, color):
                    plt.plot(*poly.exterior.xy, c = c)
                    for intr in poly.interiors:
                        plt.plot(*intr.xy, c = "red")
            except:
                plt.plot(*location.exterior.xy)
            plt.show()
    else:
        # https://gis.stackexchange.com/questions/113799/how-to-read-a-shapefile-in-python
        shp = fiona.open(PATH["data"] + placeid + "/" + placeid + ".shp")
        first = next(iter(shp))
        location = shapely.geometry.shape(first['geometry'])

    Gs = {}
    for parameterid, parameterinfo in tqdm(osmnxparameters.items(), desc = "Networks", leave = False):
        for i in range(0,10): # retry
            try:
                Gs[parameterid] = ox.graph_from_polygon(location,
                                    network_type = parameterinfo['network_type'],
                                    custom_filter = (parameterinfo['custom_filter']),
                                    retain_all = parameterinfo['retain_all'],
```

What is Geospatial Data Science?

mentimeter link

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.

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

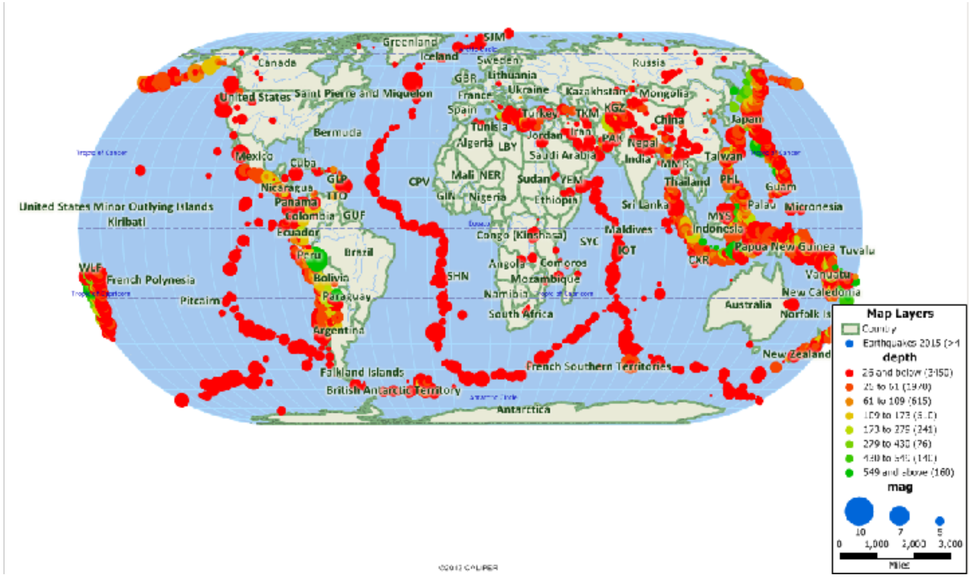
What is Geospatial Data?

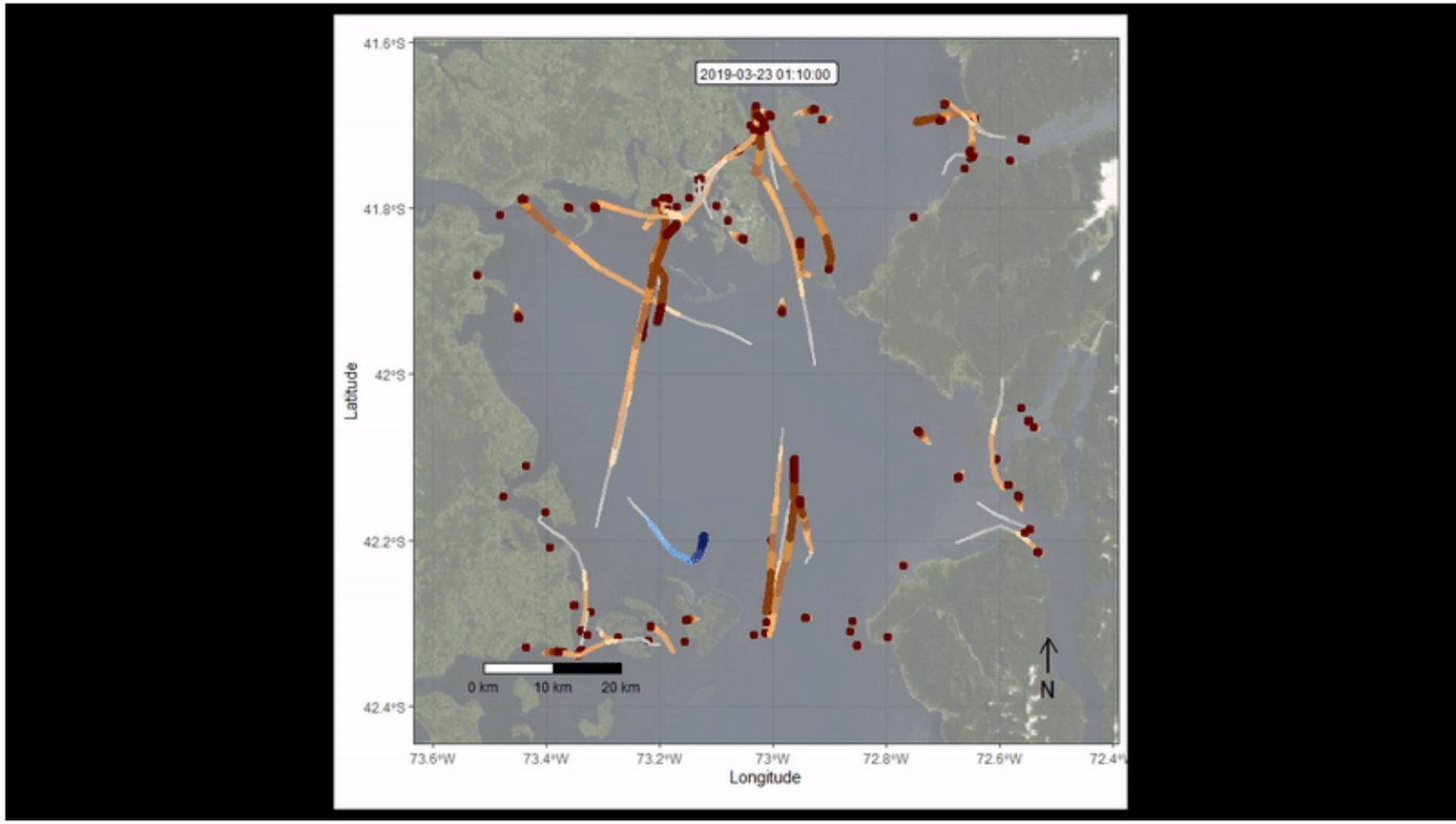
Geospatial data is information that describes objects, events or other features with a location on or near the surface of the earth.

- Coordinates
- Attributes
- Temporal information

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

There is static and dynamic data





There is vector and raster data

Vector: Geometric objects

.shp, .svg

Raster: Grid of pixels

.tif, .jpg, .png, .bmp

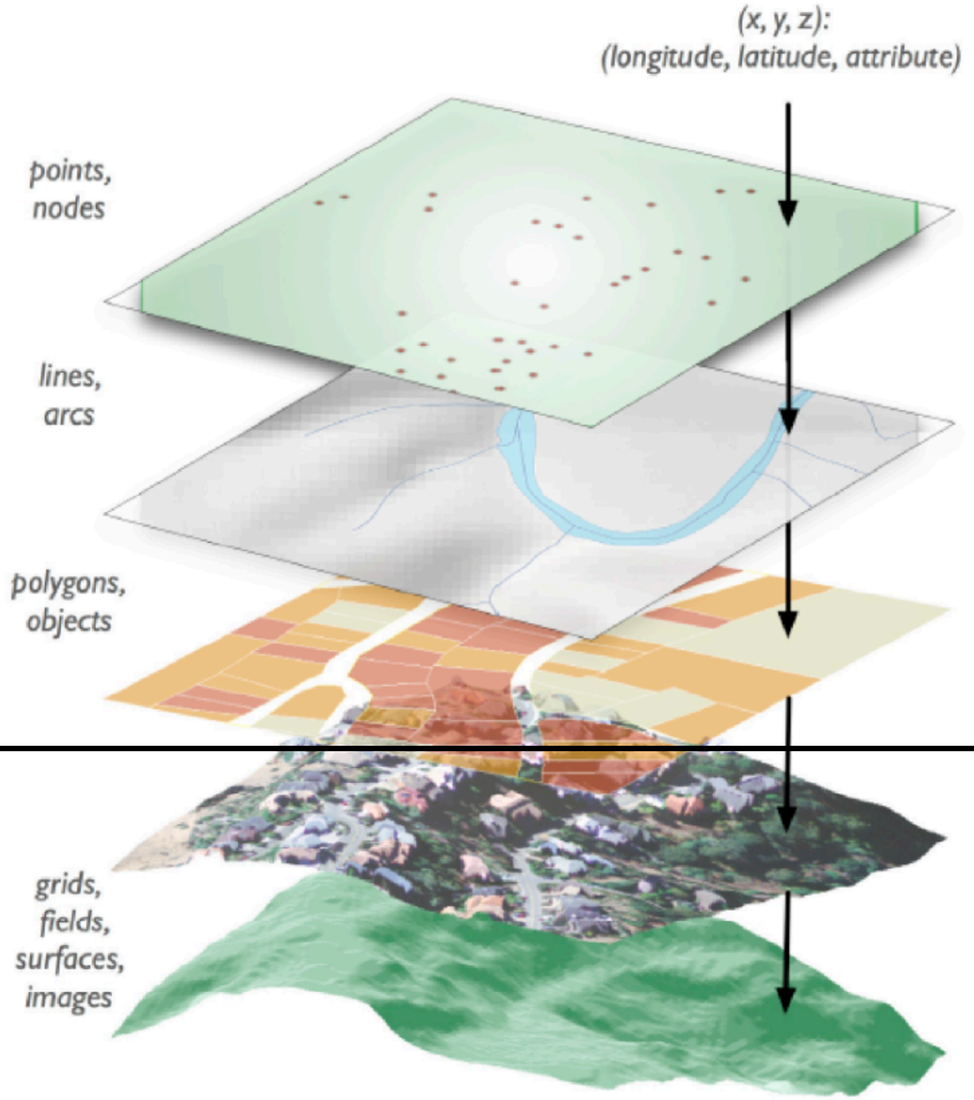
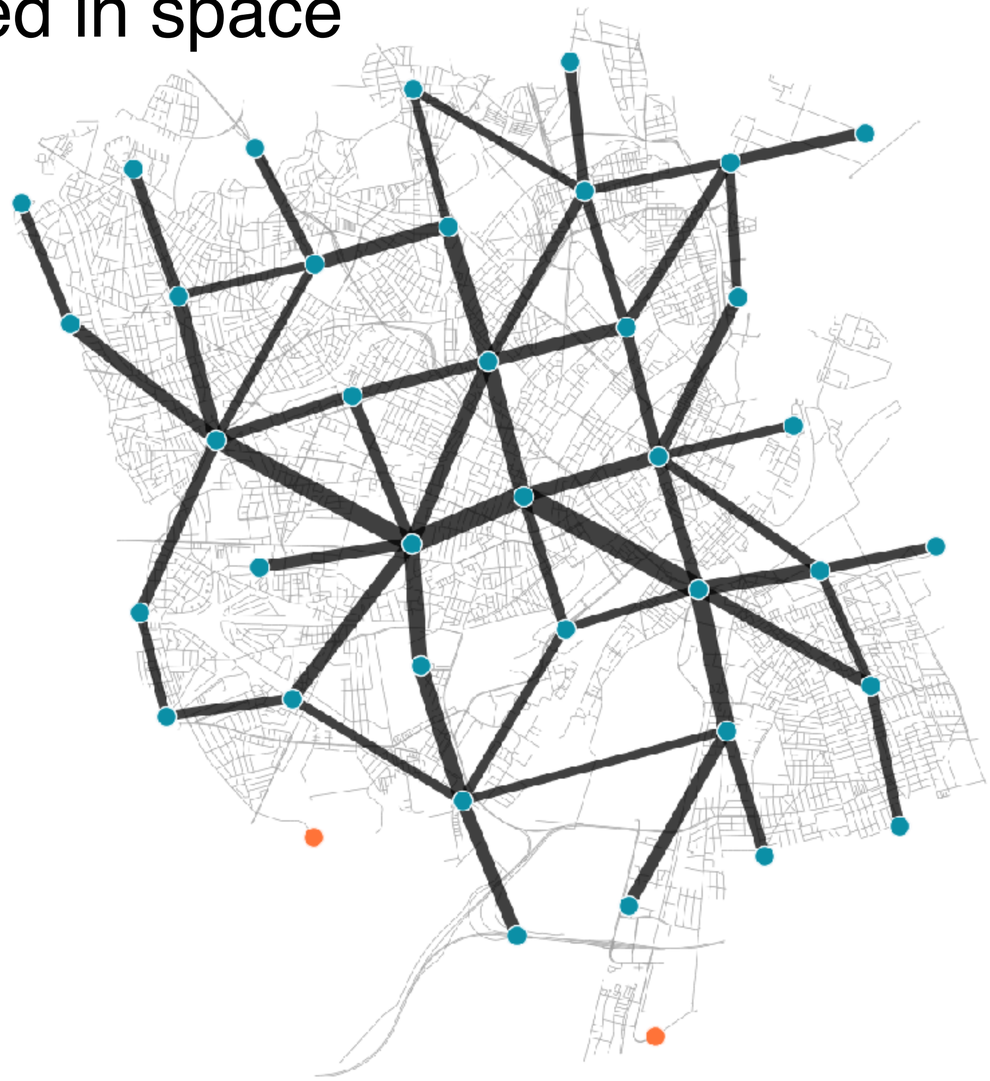
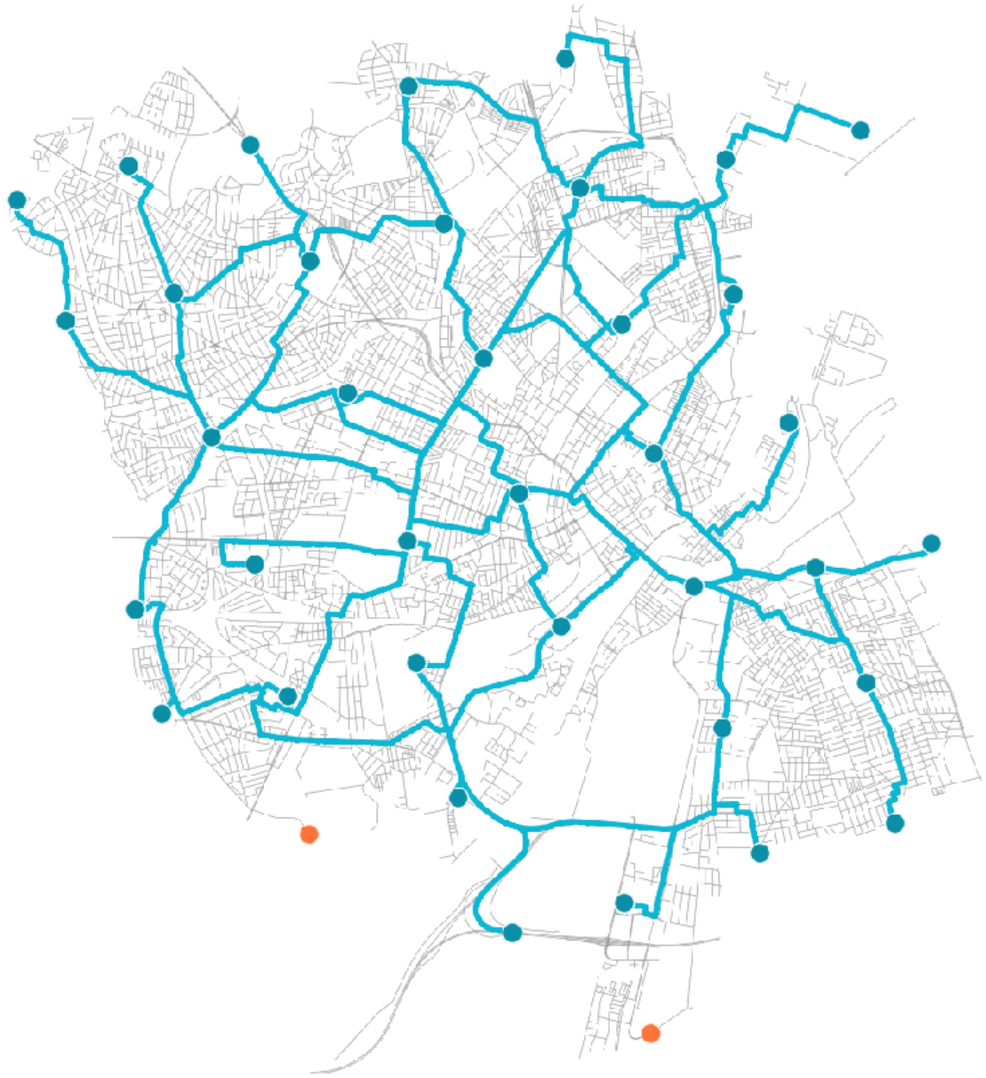


Image: Maggi Kelly, UC Berkeley

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

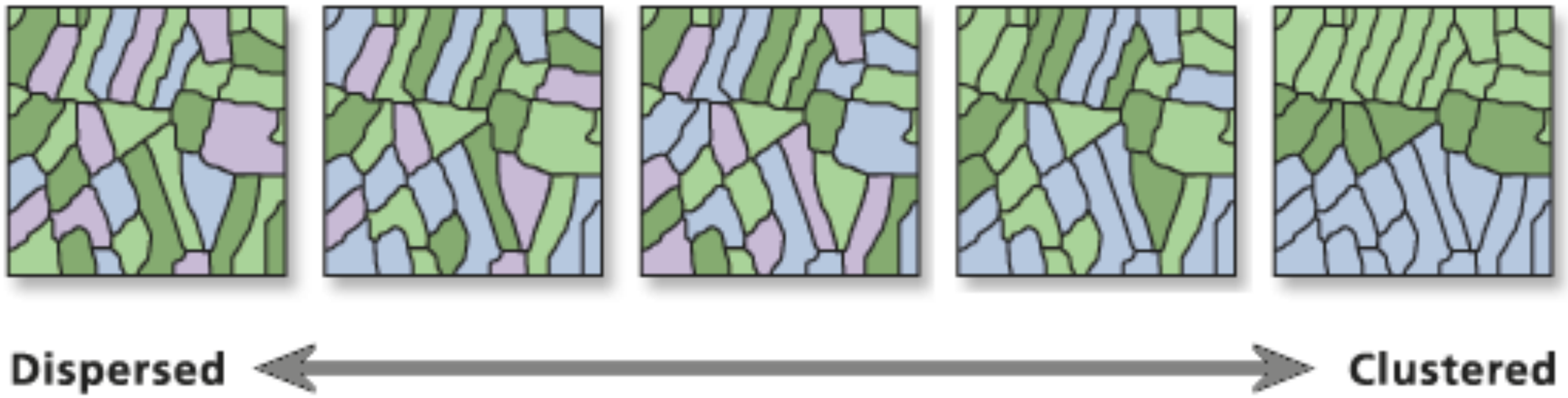
GDS asks: How do things relate in space?

How can we formalize this question?

How can we operationalize this question?

GDS is mathematical, visual, and computational

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



Geospatial Data Science stands on the shoulder of giants

	Most proponents come from	Since	Goal
Urban/regional planning	Urban planning, Architecture	3000 BC	City/Regional development
Human geography	Social sciences	18th century	Socio-cultural observation/ critique
Geoinformatics	Geography, Computer Science	1980s	Spatial analysis
Urban computing	Computer science	2000s	Urban Big Data and technology analysis
Complex spatial systems	Physics	2000s	Parsimonious models
Geospatial data science	Geography, Computer science, Geoinformatics, Physics	2010s	Quantitative understanding

This course will teach you:

- Data structures and principles of GDS
- Gathering and preprocessing large-scale geospatial data
- State-of-the-art computational tools for GDS
- Spatial network analysis
- Real world applications of these techniques in an applied context

https://learnit.itu.dk/local/coursebase/view.php?ciid=940

We use the Geographic Data Science platform `gds_env`

https://darribas.org/gds_env/

Course topics

Unit	Topic	New library	Main source
1	Geometric objects	shapely	Tenkanen, Heikinheimo, Aagesen
2	Geospatial data in Python	geopandas	Tenkanen, Heikinheimo, Aagesen, Arribas-Bel
3	Choropleth mapping	PySAL, contextily	Arribas-Bel
4	Spatial weights		Arribas-Bel
5	Spatial autocorrelation	esda	Arribas-Bel
6	Spatial clustering	sklearn	Arribas-Bel
7	Point pattern analysis	pointpats	Arribas-Bel
8	OpenStreetMap	OSMnx, pyrosm	Boeing, Tenkanen
9	Spatial networks	spaghetti	Boeing, Gaboardo, Rey, Lumnitz
10-14	Mobility / Applications	scikit-mobility	Pappalardo

https://learnit.itu.dk/local/coursebase/view.php?ciid=940

Course overview

Geographic Data Science with Python

Schedule GDS2022 — Last update: 2022-04-25

Room 4A56, 10:00-14:00

Lecture	Instructor	CalW	Date	Topic	Event
1	M	5	Feb 3	**Introduction**	Installations
2	M	6	Feb 10	**Geospatial data in Python**	
3	M	7	Feb 17	**Choropleth mapping**	
4	M	8	Feb 24	**Spatial weights**	
5	M	9	Mar 3	**Spatial autocorrelation**	
6	M	10	Mar 10	**Spatial clustering**	
7	M	11	Mar 17	**Point pattern analysis**	
8	M	12	Mar 24	**OSM+OSMnx**	
9	M	13	Mar 31	**Individual mobility**	Projects proposed
10	M	14	Apr 7	**Spatial networks**	Projects approved
Holiday week		15			
11	M	16	Apr 21	**Mobility patterns**	Tutorial: MovingPandas
12	A+A	17	Apr 28	**Bicycle network analysis**	Tutorial: OSMnx filters
13	M	18	May 5	**Aggregate mobility**	Tutorial: Folium maps
14	M	19	May 12	**Sustainable mobility**	Tutorial: Report writing
		20	May 20		Project submission

Course overview

Lecture 10:00-11:45
Lunch 11:45-12:15
Exercise 12:15-13:50

- Group project deadlines
- group ready
- project idea submitted
- approved
- final submission

Schedule GDS2022 — Last update: 2022-04-25

Room 4A56, 10:00-14:00

Lecture	Instructor	CalW	Date	Topic	Event
1	M	5	Feb 3	**Introduction**	Installations
2	M	6	Feb 10	**Geospatial data in Python**	
3	M	7	Feb 17	**Choropleth mapping**	
4	M	8	Feb 24	**Spatial weights**	
5	M	9	Mar 3	**Spatial autocorrelation**	
6	M	10	Mar 10	**Spatial clustering**	
7	M	11	Mar 17	**Point pattern analysis**	
8	M	12	Mar 24	**OSM+OSMnx**	
9	M	13	Mar 31	**Individual mobility**	Projects proposed
10	M	14	Apr 7	**Spatial networks**	Projects approved
Holiday week		15			
11	M	16	Apr 21	**Mobility patterns**	Tutorial: MovingPandas
12	A+A	17	Apr 28	**Bicycle network analysis**	Tutorial: OSMnx filters
13	M	18	May 5	**Aggregate mobility**	Tutorial: Folium maps
14	M	19	May 12	**Sustainable mobility**	Tutorial: Report writing
		20	May 20		Project submission

After this course you will be able to:

- Use Python libraries programmatically for geospatial data analysis.
- Reflect on the motivation and inner workings of the main methodological approaches of GDS, both analytical and visual.
- Critically evaluate the suitability of a specific GDS technique, what it can offer and how it can help answer questions of interest.
- Explain how to interpret results, in a process of turning data into insights.

https://learnit.itu.dk/local/coursebase/view.php?ciid=940

In case of issues, contact:

Student Affairs and Programmes (SAP)

sap@itu.dk

Michael's office hours

Tuesdays 11:00-12:00, office 3F11 (DR building)

misz@itu.dk

We can only cover a small part of all tools

The typical GDS workflow:

Set Up Your Environment

- Virtual machine environments (Linux-based)
- Spatial databases (PostgreSQL/PostGIS) with multi-user editing and versioning (GeoGig)

Wrangle Data

- APIs (Google Maps, OpenStreetMap)
- Modern data formats and tools (GeoJSON, GDAL)

Analyze Data

- Vector-based analysis using ArcPy and PySAL (within IPython)
- Raster-based analysis using Rasterio (within IPython) and R

Visualize and Publish Data

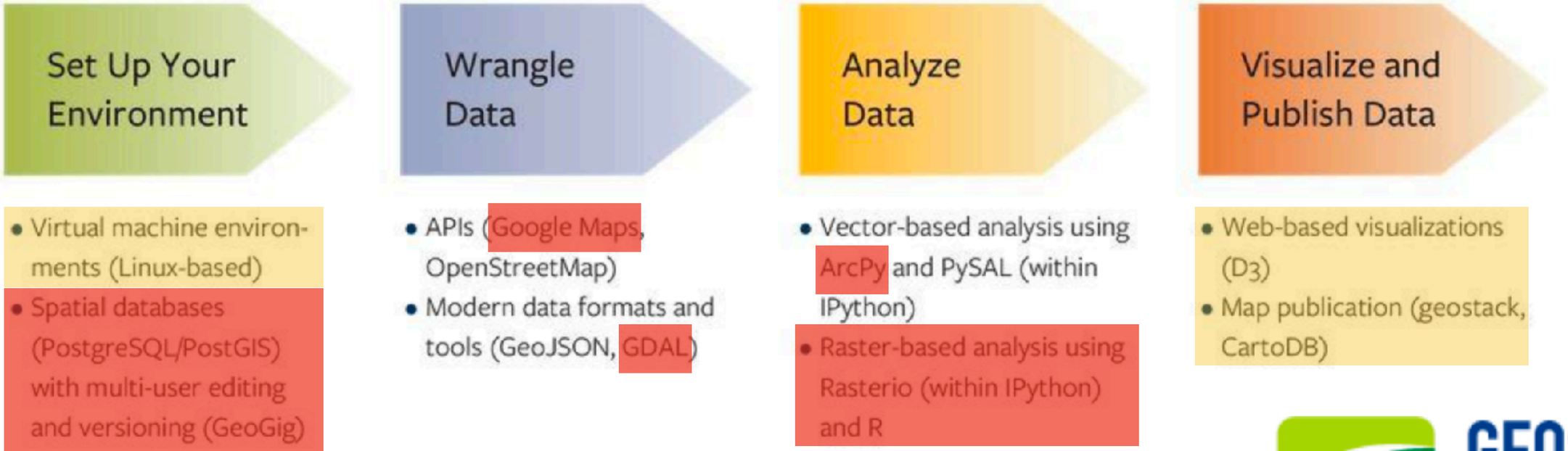
- Web-based visualizations (D3)
- Map publication (geostack, CartoDB)

GEOSPATIAL INNOVATION FACILITY

Cutting-Edge Mapping Technology at UC Berkeley

We can only cover a small part of all tools

The typical GDS workflow:



GEOSPATIAL INNOVATION FACILITY

Cutting-Edge Mapping Technology at UC Berkeley

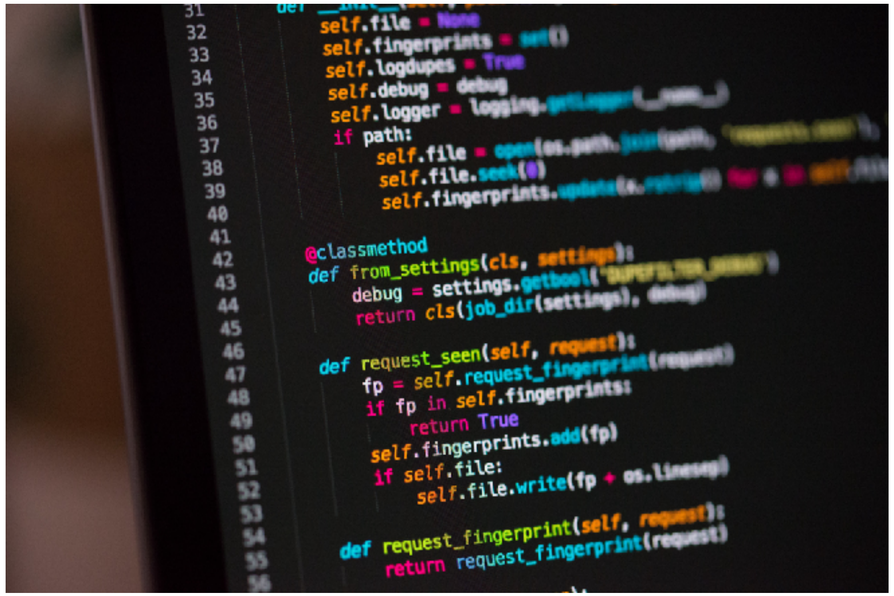
We will only cover the procedural part of GDS (in Python)

DB handling

PostgreSQL

mongoDB

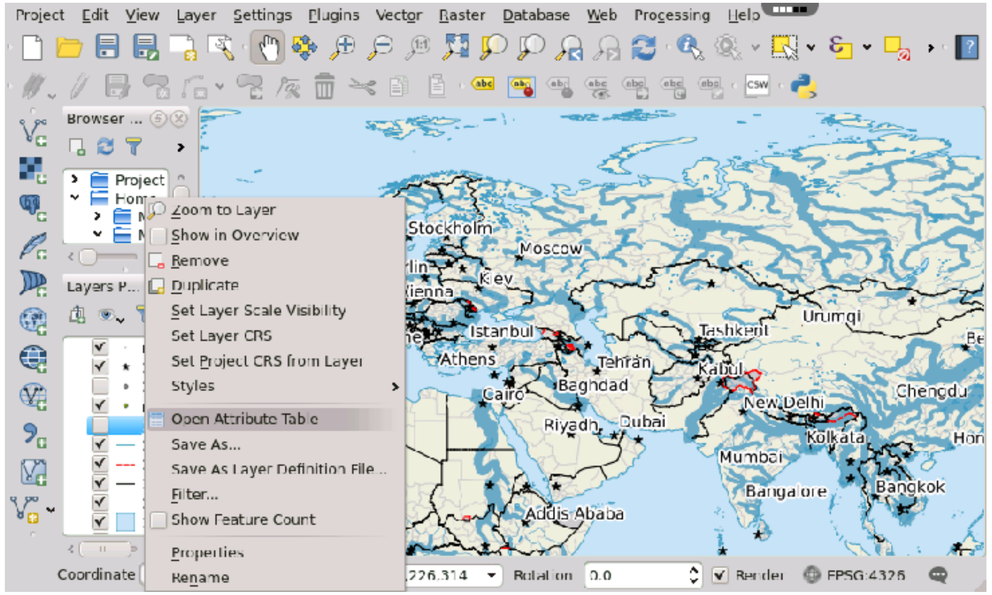
Procedural



python™

R

Point & Click GIS

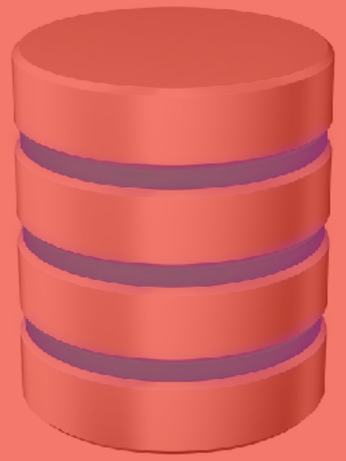


QGIS

ArcGIS

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

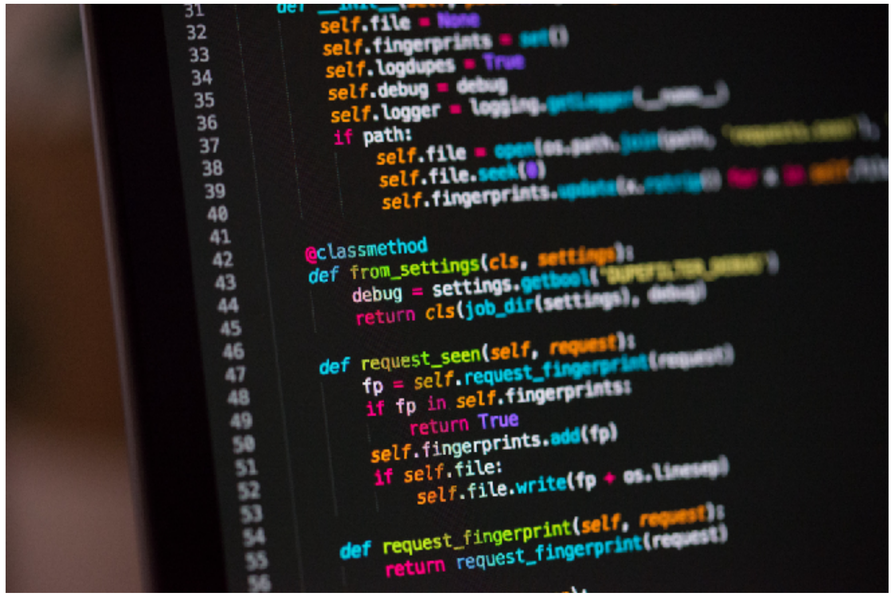
DB handling



PostgreSQL

mongoDB

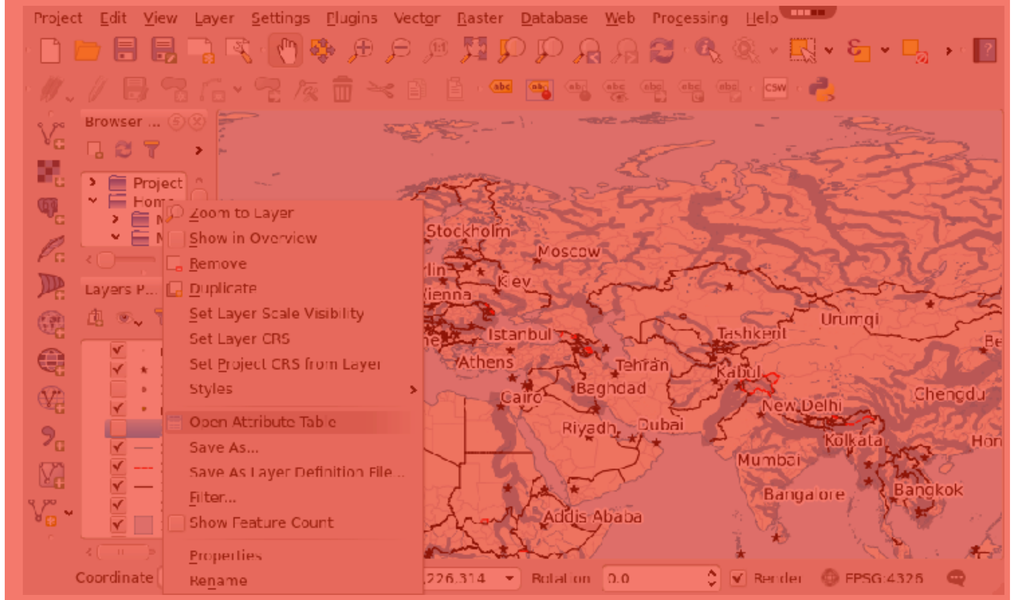
Procedural



python™

R

Point & Click GIS



QGIS

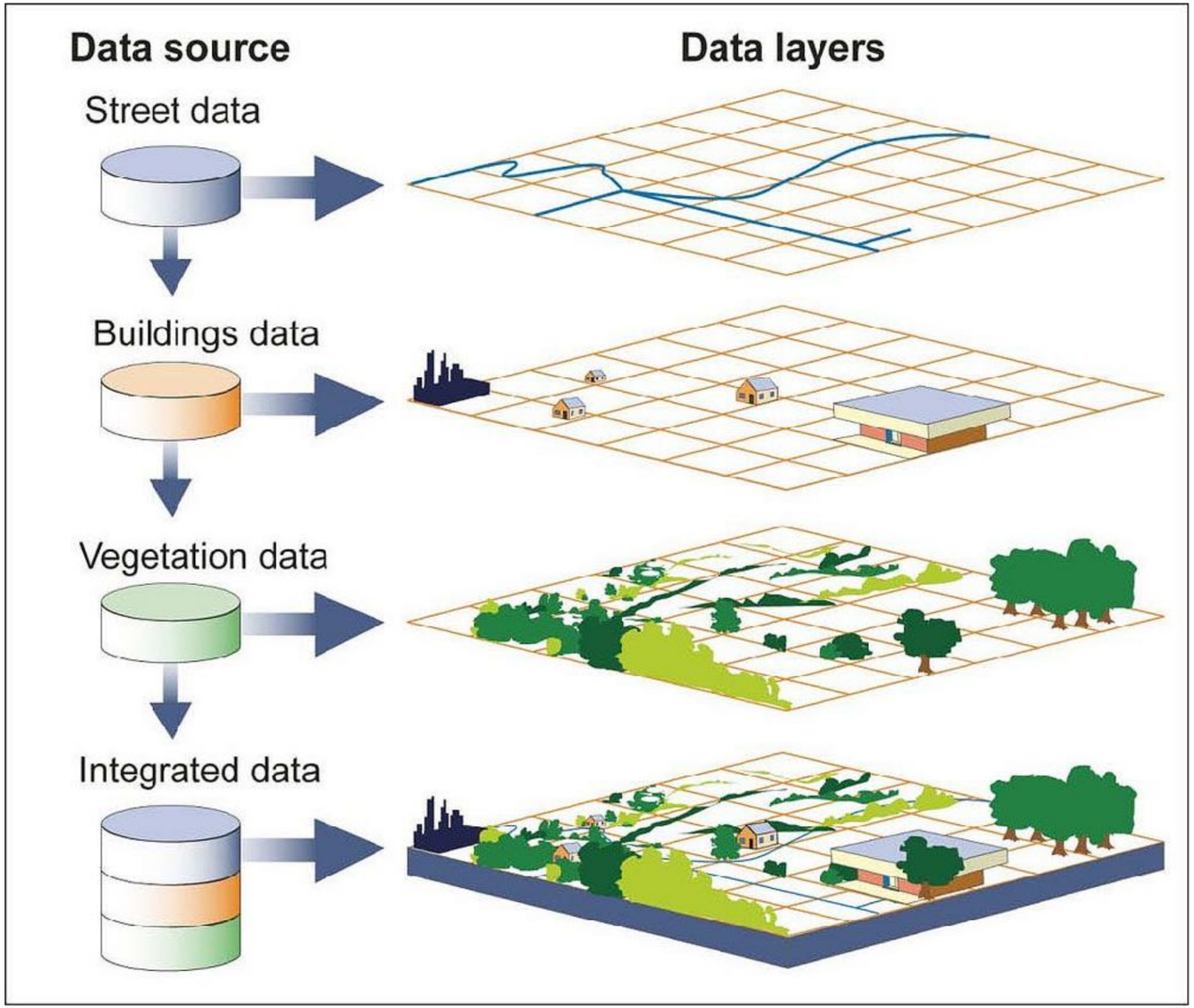
ArcGIS

GIS

A Geographic information system (GIS) is a system for capturing, storing, checking, and displaying geospatial data.

Can refer to:

- concrete software
- abstract concept



Source: GAO.

https://www.nationalgeographic.org/encyclopedia/geographic-information-system-gis/

Who is John Snow?

https://www.youtube.com/watch?v=lNjrAXGRda4

Jupyter