WiMLDS
Women in Machine Learning & Data Science

# Sneak peek at Geodata

## WiMLDS Zurich / Litix / SIT Academy

2022.05.18, Zurich

SIT Academy

# Before we start!

Give your names to Kate -> we need to create the user accounts

# Who are we?

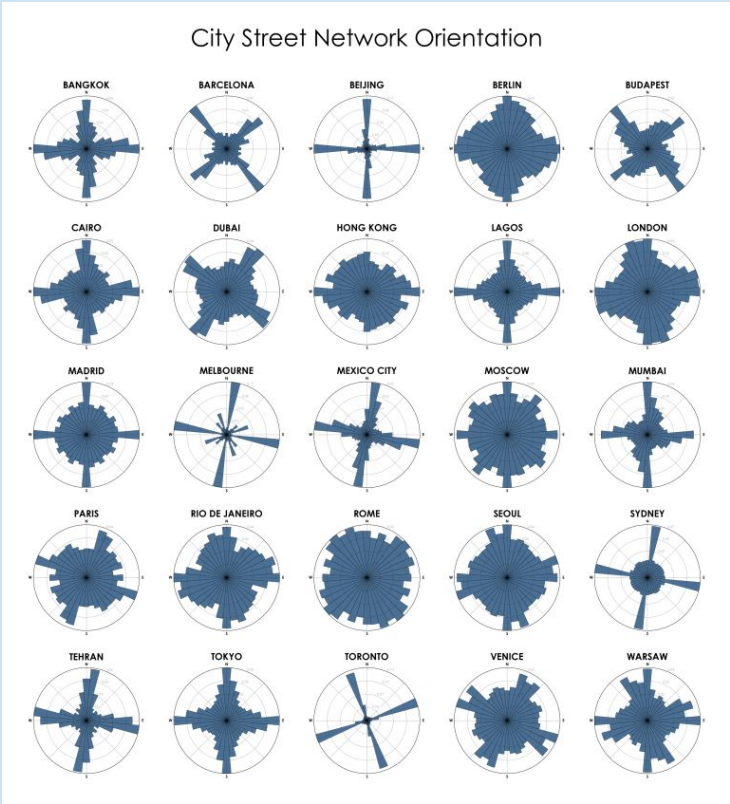
| | |
|---|---|
| WiMLDS Zurich: | Our mission is to support and promote women and gender minorities who are practicing, studying or are interested in the fields of machine learning and data science. |
| Litix: | Consortium of freelancers in the area of data analytics, geodata/GIS and applied statistics. |
| SIT Academy: | SIT Academy is a coding and data science academy offering a variety of programs to motivated learners. |

Presenter today: Agnes Gubicza from WiMLDS

# What is going to happen today?

- Brief introduction to the theory
- 2+1 exercises
  - Visualize different types of geodata / geoservices
  - Spatial analytics
- Have fun!

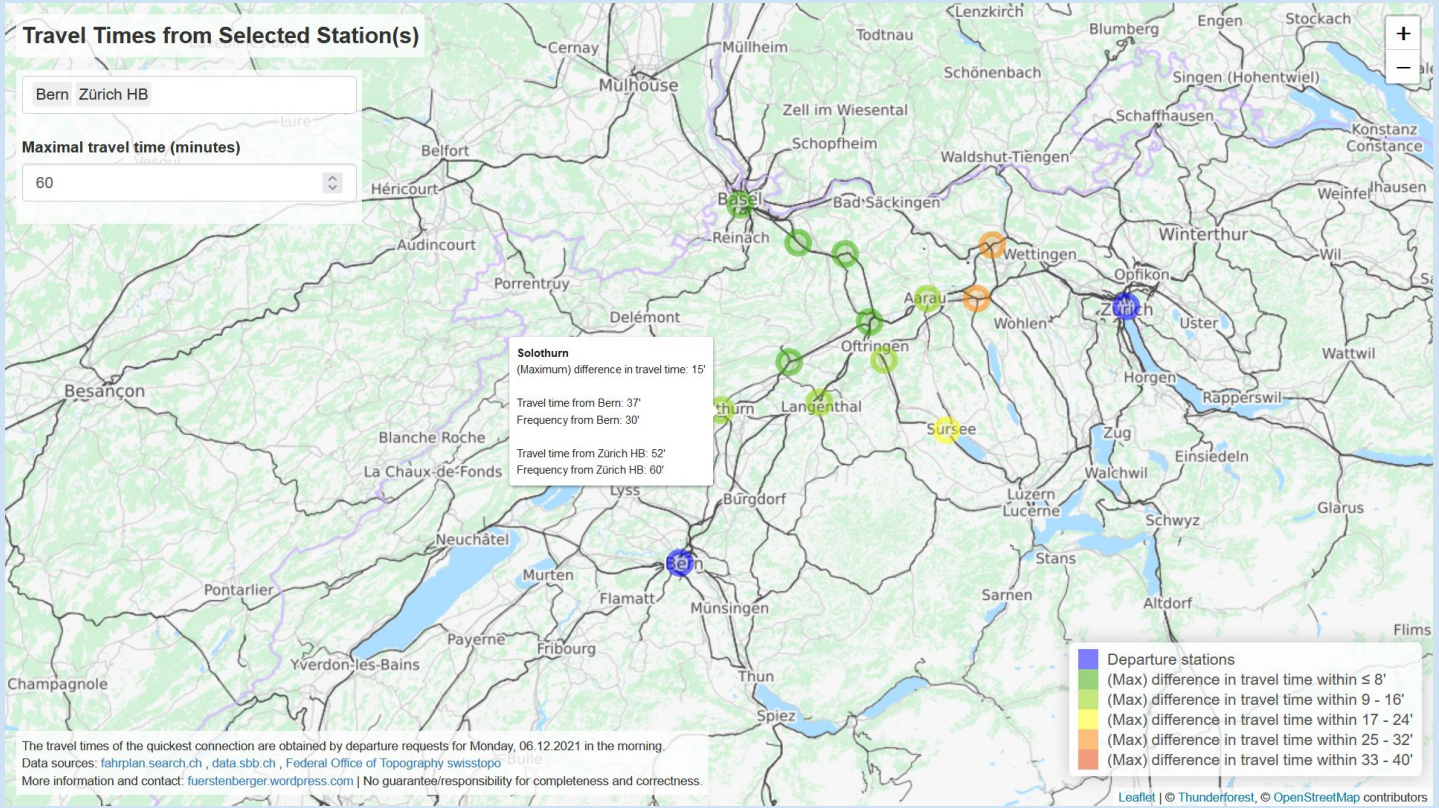
# Cool examples: Urban data visualization



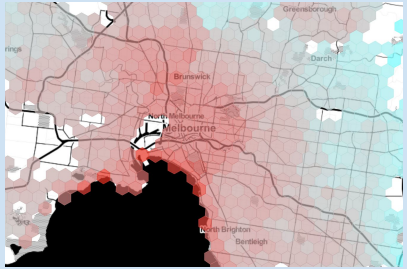

*https://geoffboeing.com/2018/07/city-street-orientations-world/*

https://fuerstenberger.shinyapps.io/minimal_travel_times/

# Where should we meet to have equal commute?

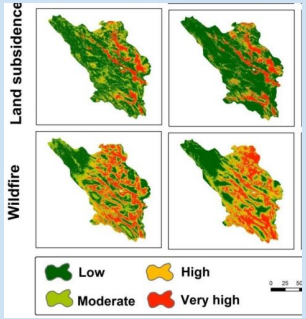


# Machine Learning with Geodata



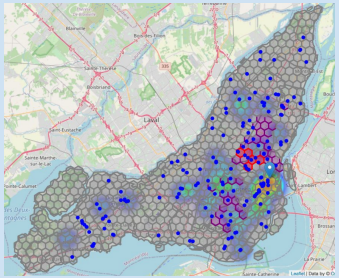

Housing prices prediction

*https://towardsdatascience.com/the-impact-of-geospatial-features-on-machine-learning-3a71c99f080a*



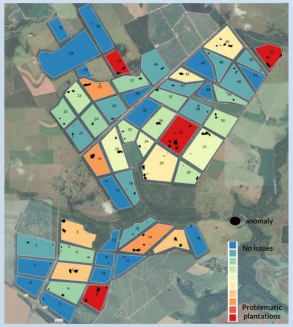

Natural hazards prediction

*https://www.nature.com/articles/s41598-020-69233-2*



Finding optimal placement

*https://towardsdatascience.com/the-battle-of-the-neighborhoods-open-a-movie-theater-in-montreal-355cf5c679b8*




Annotating satellite imagery

*https://www.gamaya.com/solutions/projects-partnerships*

Mining mobility data

*https://www.mdpi.com/2220-9964/8/10/434*

# What is geodata?

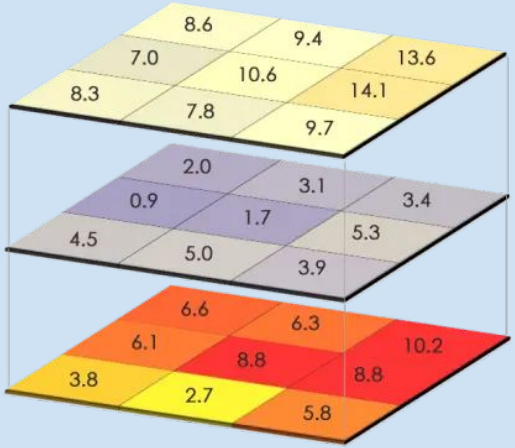
Geodata:

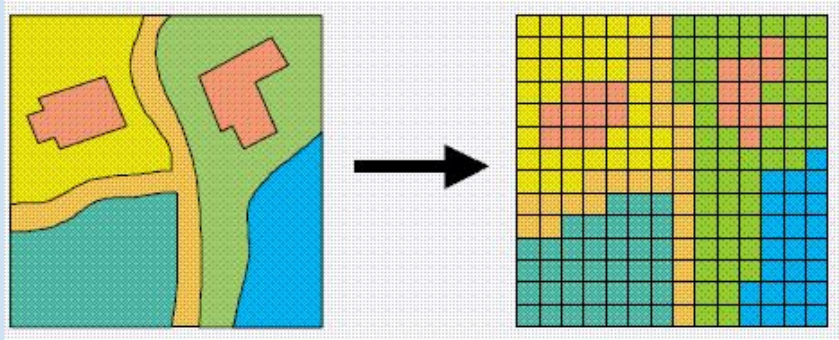
- *spatial data*
- geographical reference
- Focus on location in the real world

GIS:

- Geographical Information System
- generic term for geodata related *software*

# Raster data



*https://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/raster-basics.htm*

Array, grid of numbers that correspond to particular geolocation

Height model

*https://gisgeography.com/spatial-data-types-vector-raster/*

# Vector GIS objects

**points**

Object w/o size/area

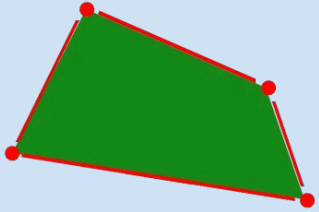
Landmarks, cities, people, points of interest

**lines**

Path: Set of points connected in a particular order

Rivers, roads

**polygons**

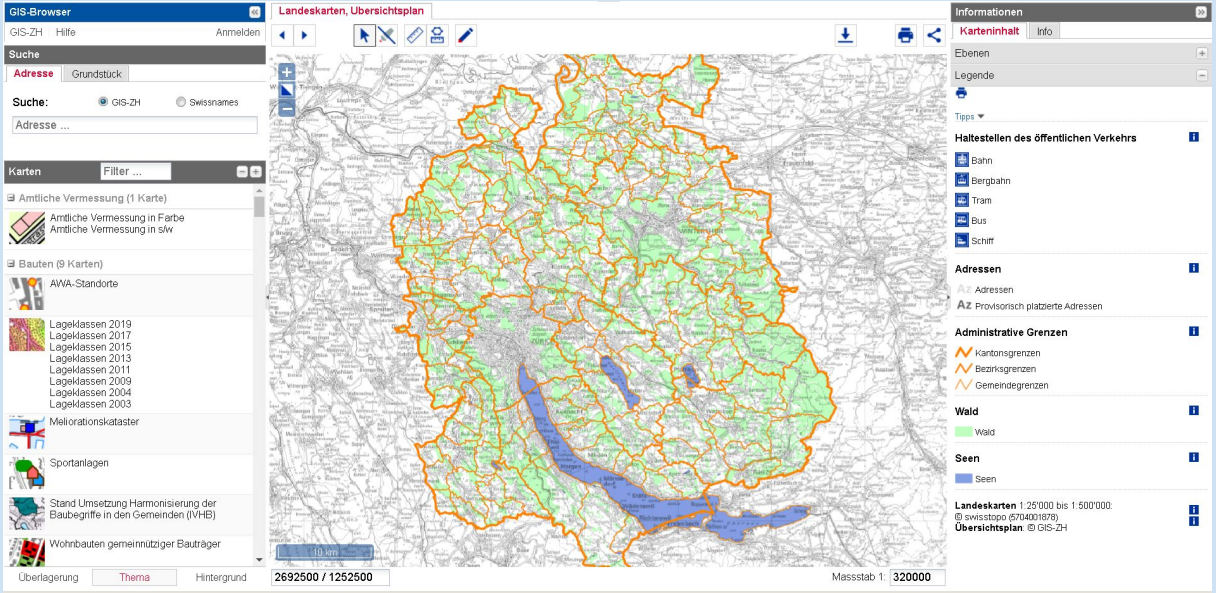


Area with boundaries: Set of points connected in a particular order, closed

Lakes, building footprint, country borders

Vector objects may come with **attribute tables**

# Geodata, maps, layers




geoportal of Kt Zurich

- geodata is stored in layers
- map is a set of layers

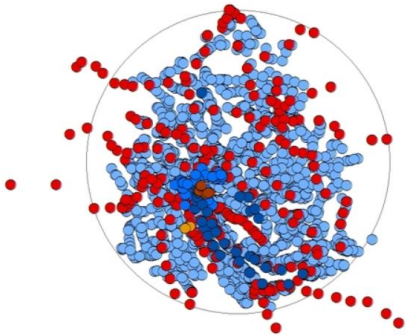
To display a layer, a symbology / style (e.g. color code) is used:

- height model with color ramp
- road network layer with and without symbology

# Demo data access

Geodatensatz

Haltestellen des öffentlichen Verkehrs



Kontakt +

Inhalt / Identifikation -

**GIS-ZH Nr.**
140

**Bezeichnung**
Haltestellen des öffentlichen Verkehrs

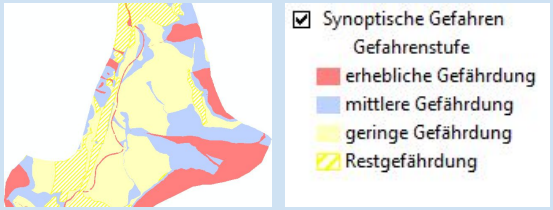
http://www.geolion.zh.ch/geodatensatz/883

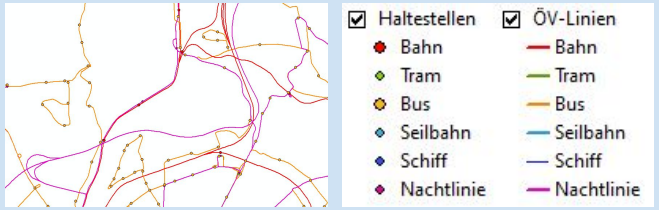
**Abgabeformat**
DXF (.dxf); ESRI Shapefile (.shp); Comma separated text (.csv); ESRI File Geodatabase (.gdb); GeoPackage (.gpkg);

| | |
|---|---|
| **WMS** | WMS ÖV-Güteklassen |
| | WMS Haltestellen des öffentlichen Verkehrs |
| | WMS Administrative Einteilungen (OGD) |
| **WFS** | WFS Haltestellen des öffentlichen Verkehrs |
| | Allgemeiner WFS für Open-Government-Data des Kantons Zürich (OGD ZH WFS) |

# Geoservices

**WMS: web map service** - raster data or maps (incl. symbology and legend information)



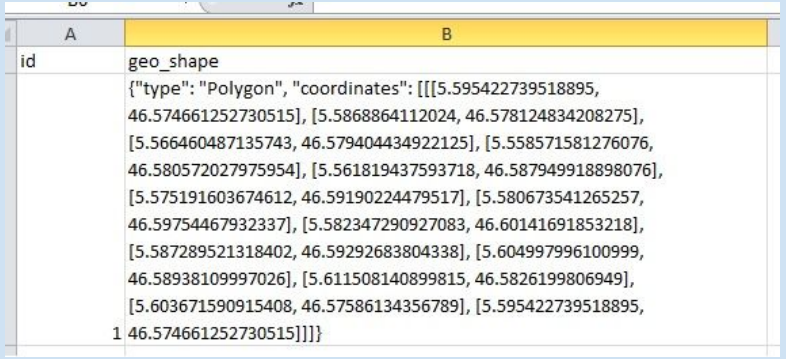

**WFS: web feature service -** vector data (with or without symbology, with attribute table)

| | geodb_oid | gefstufe_hw | label_hw | |
|---|---|---|---|---|
| 0 | 256486 | 2 | hf4 | POLYGON ((2684271.590 12 |
| 1 | 256479 | 2 | hs4 | POLYGON ((2684899.757 12 |
| 2 | 256482 | 2 | hf4 | POLYGON ((2684428.206 12 |

| | geodb_oid | objectid | nhstid | diva_nr | chstname | cna |
|---|---|---|---|---|---|---|
| 0 | 413985 | 83665 | 8576245 | 2986 | Regensdorf, Zentrum | Regensc |
| 1 | 413986 | 83666 | 8576246 | 1850 | Regensdorf, Ostring | Regensc |
| 2 | 413987 | 83667 | 8576248 | 943 | Watt, Geerenweg | W |

# Common static geodata formats

- GeoJSON
  - extension of JSON
- ESRI / Shapefile
  - very messy
- Tabular data
  - CSV
- ...many others



| | A | B |
|---|---|---|
| | id | geo_shape |
| 1 | | {"type": "Polygon", "coordinates": [[[5.595422739518895, 46.574661252730515], [5.5868864112024, 46.578124834208275], [5.566460487135743, 46.579404434922125], [5.558571581276076, 46.580572027975954], [5.561819437593718, 46.587949918898076], [5.575191603674612, 46.59190224479517], [5.580673541265257, 46.59754467932337], [5.582347290927083, 46.60141691853218], [5.587289521318402, 46.59292683804338], [5.604997996100999, 46.58938109997026], [5.611508140899815, 46.5826199806949], [5.603671590915408, 46.57586134356789], [5.595422739518895, 46.574661252730515]]]} |

# Geopandas

Familiar Pandas routines

Additional geo-methods

- _.area, _.bounds
- _.distance, _.centroid
- _.contains, _.intersects

Dtype: geometry

Powerful, but cumbersome to install on Windows :(

```
In [12]: world['centroid_column'].head()
Out[12]:
0    POINT (163.85316 -17.31631)
1      POINT (34.75299 -6.25773)
2     POINT (-12.13783 24.29117)
3     POINT (-98.14238 61.46908)
4    POINT (-112.59944 45.70563)
Name: centroid_column, dtype: geometry
```

# Plan for today

- Use Python Geopandas for analysis and Leaflet for visualization
- 2 in-class exercises + 1 extra exercise
- You will work on the server:
    - Create a user account:

**34.65.177.87**

    - Pull the repo (THIS WILL OVERWRITE!!!):

**https://tinyurl.com/GeoDataWS**

- The server will stay up until **Monday**, so you can play with the exercises also after the workshop.

# Outro / discussion

# Thank you!

https://www.meetup.com/Zurich-Women-in-Machine-Learning-and-Data-Science/