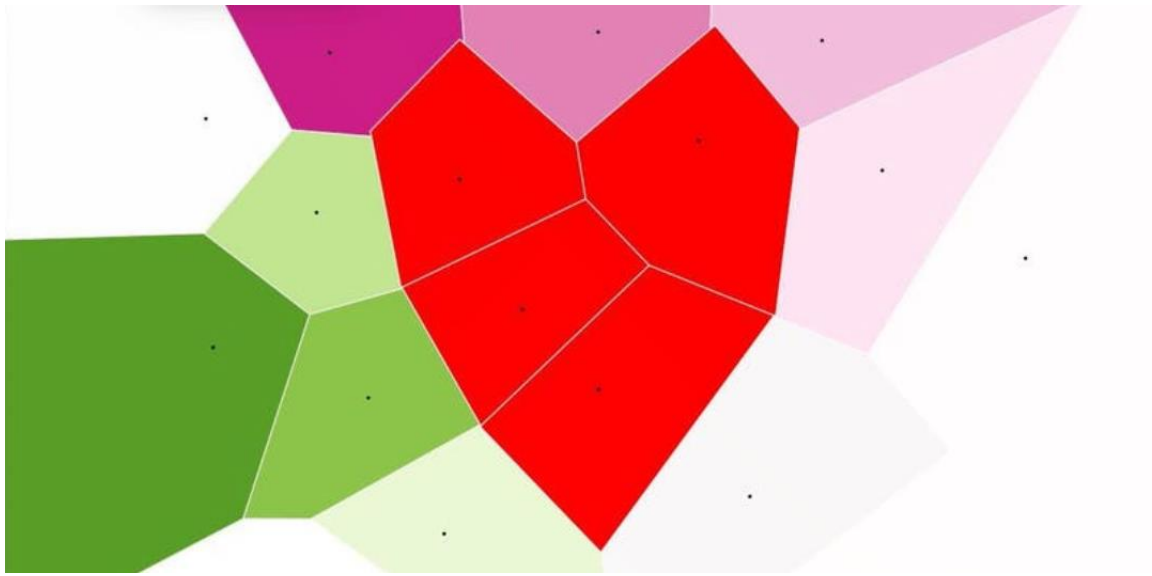
data4good Hackathon



GRÜN STATT GRAU

GOAL

Based on a Natural Language Processing (NPL) data science project, for this Hackathon, we tried to answer one of the business needs, such as:

Survey of areas associated with green roofs.

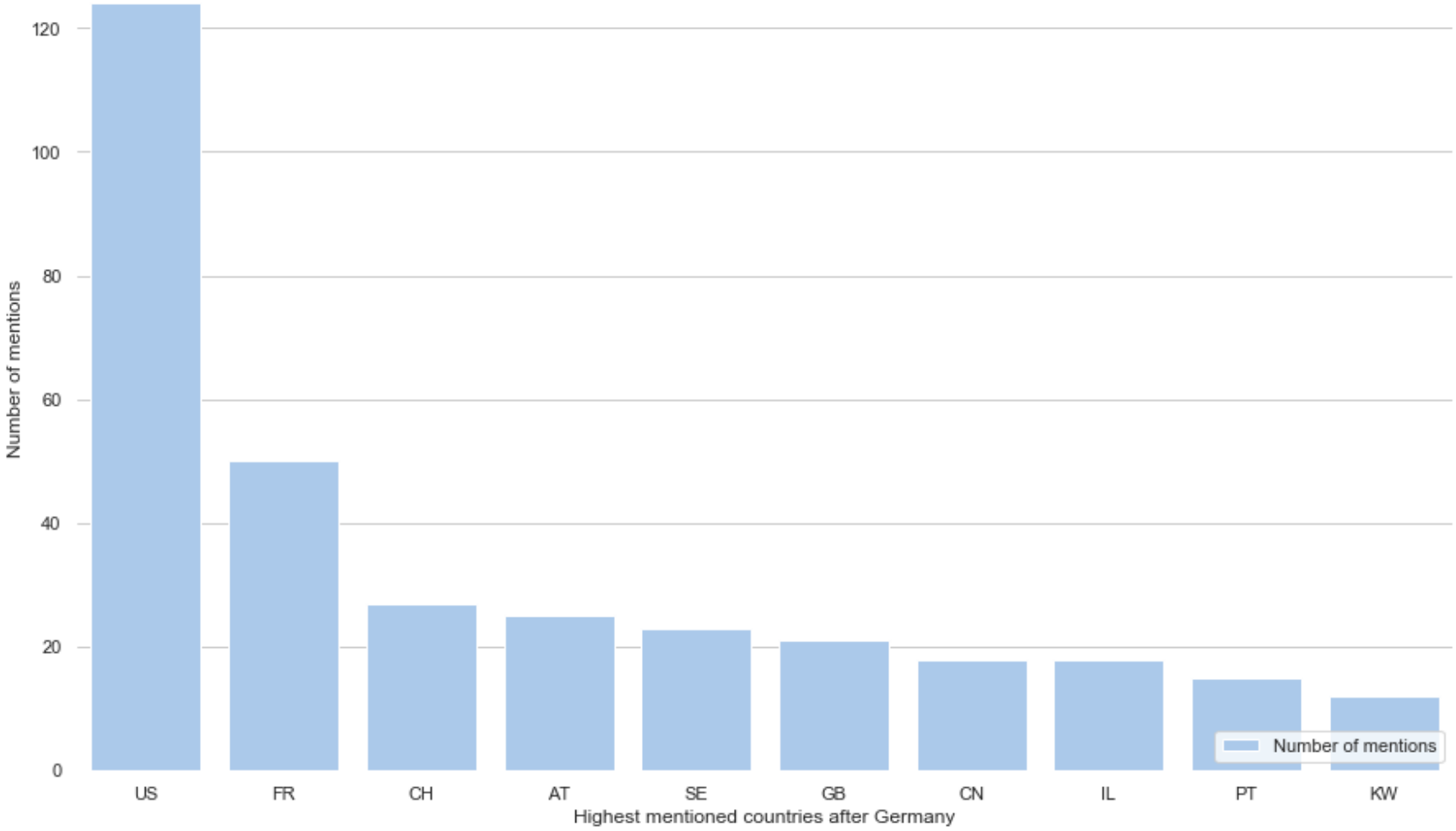
Data comes from German text in magazines about urban green roof tops.

OBJECTIVES:

- Identify in the text references/mentions to location (LOC) (NER with Spacy)

- Counts of the references/mentions for countries and cities (GeoText extracts country and city mentions from text)

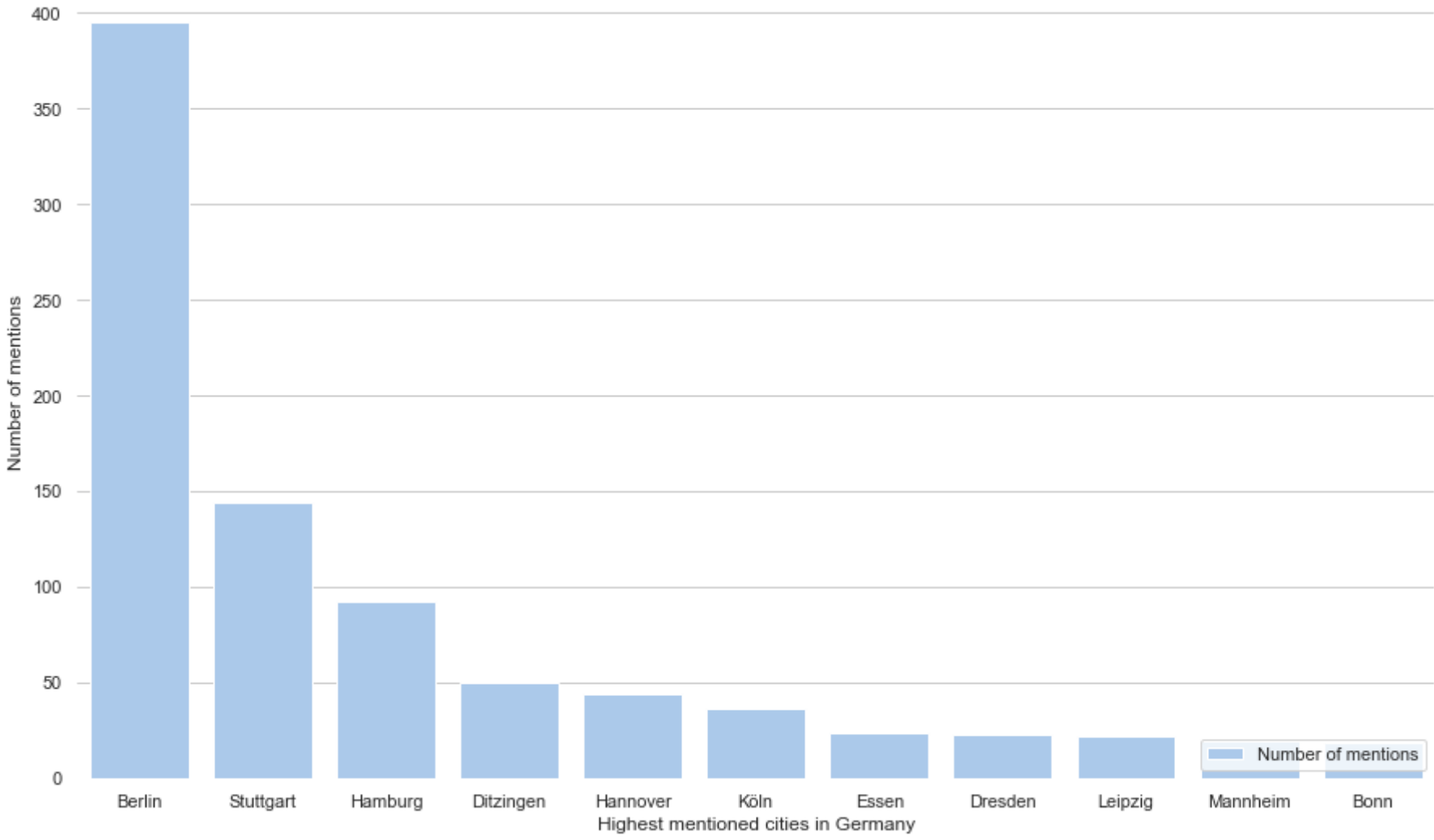
FIRST INSIGHTS FOR COUNTRIES



Note: Germany is mentioned 1290 times and we excluded it from the bar chart to ease the reading. There is also a map (click the green bottom) ☺

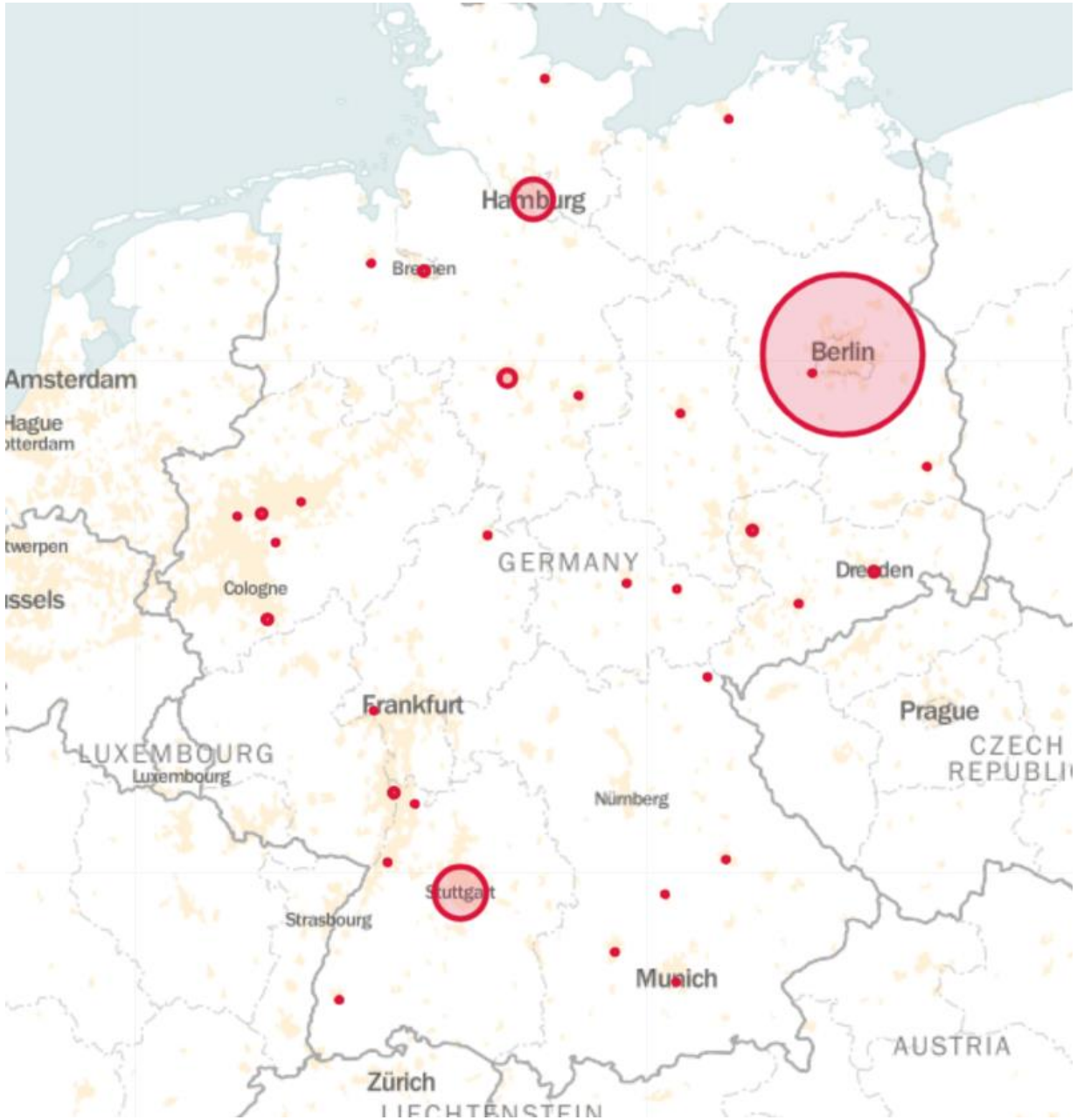
FIRST INSIGHTS FOR COUNTRIES

FIRST INSIGHTS FOR CITIES



Note: There is also a map (click the green bottom) ☺

FIRST INSIGHTS FOR CITIES



IMPORTANT NOTES

1. The locations mentioned are not exclusively linked with green roof top buildings. Such distinction will have to be better developed.

2. Because GeoText has incorporated city names in English, some of our German cities are not directly connected with the English ‘translation’: Koln => Cologne.

3. Classifying locations with topics modelled from text with LDA was unsuccessful.

IMPORTANT NOTES

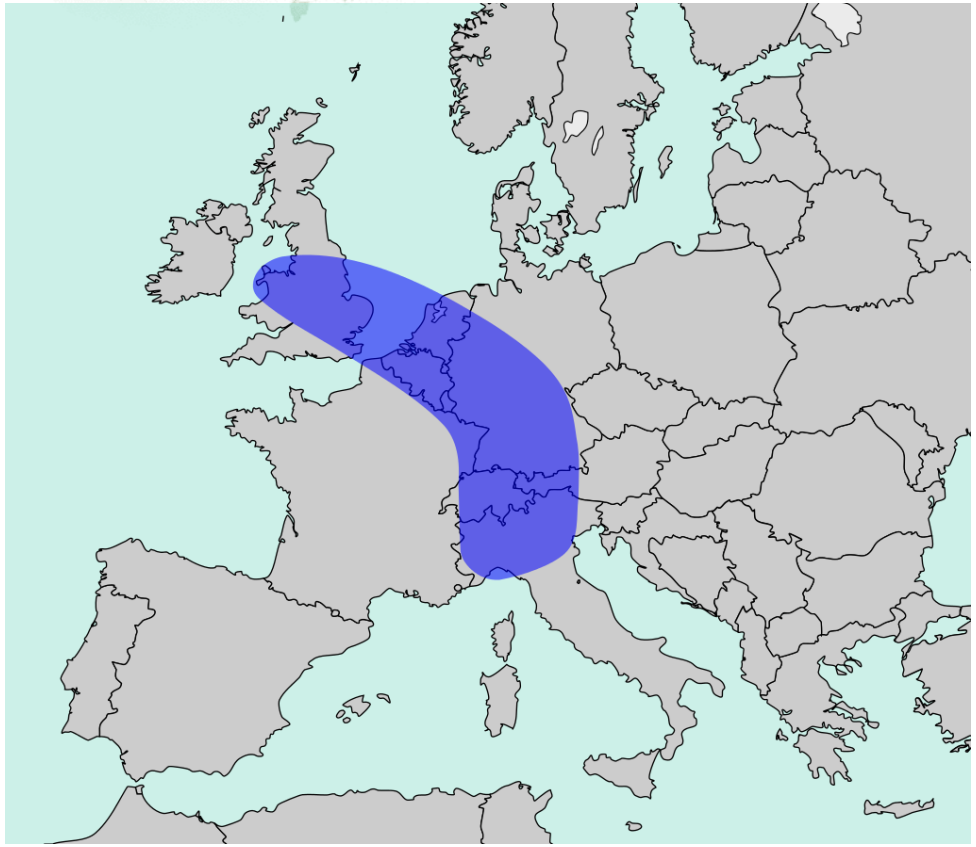
4. Filtering manually allow the identification of cases where, location names where not relevant at all, such as:

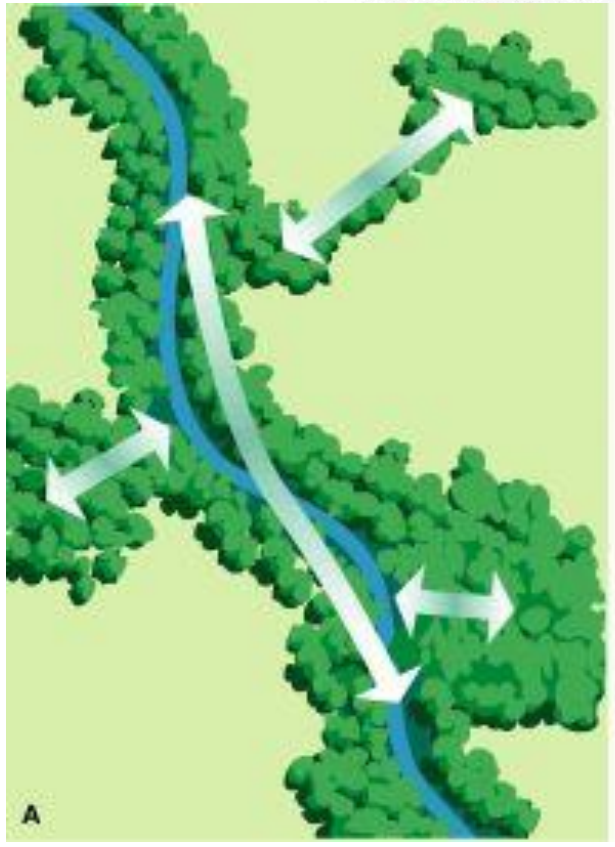
- Nuremberg next to *Gmbh*, was regarding the enterprise location, not the building;

- Locations next to words such as Dachflächen or m2, m3, are more likely be connected with correct locations of green rooftops buildings.

ANALYSIS PROPOSALS FOR POTENTIAL LOCATIONS

THE CORRIDOR CONCEPT IN GEOGRAPHY

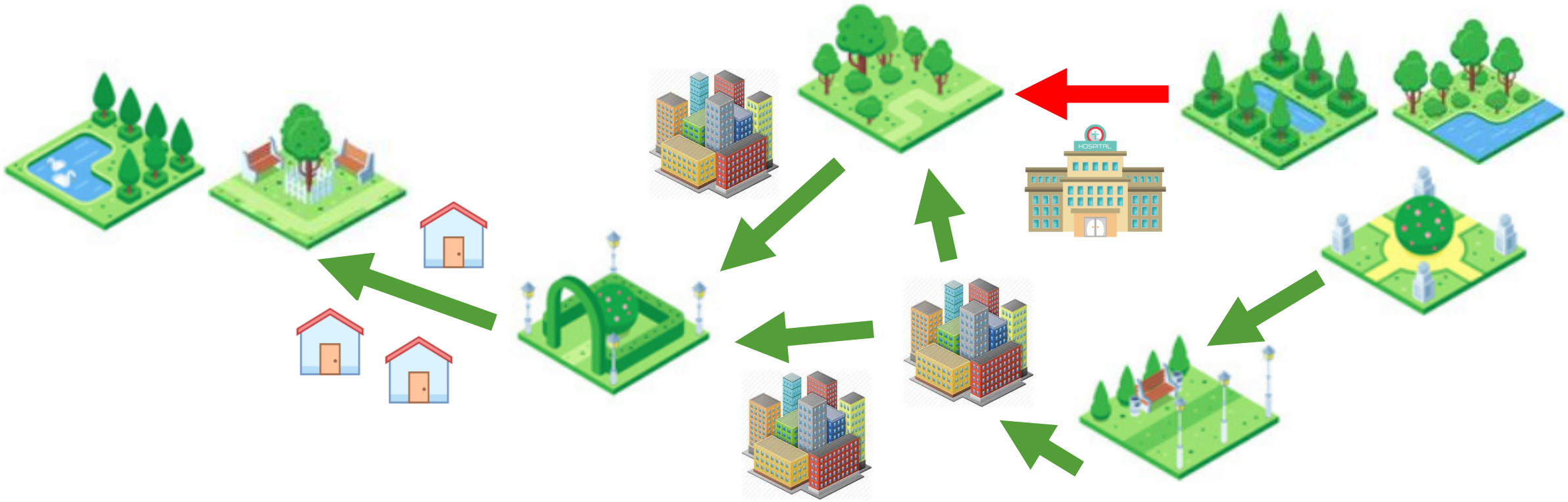




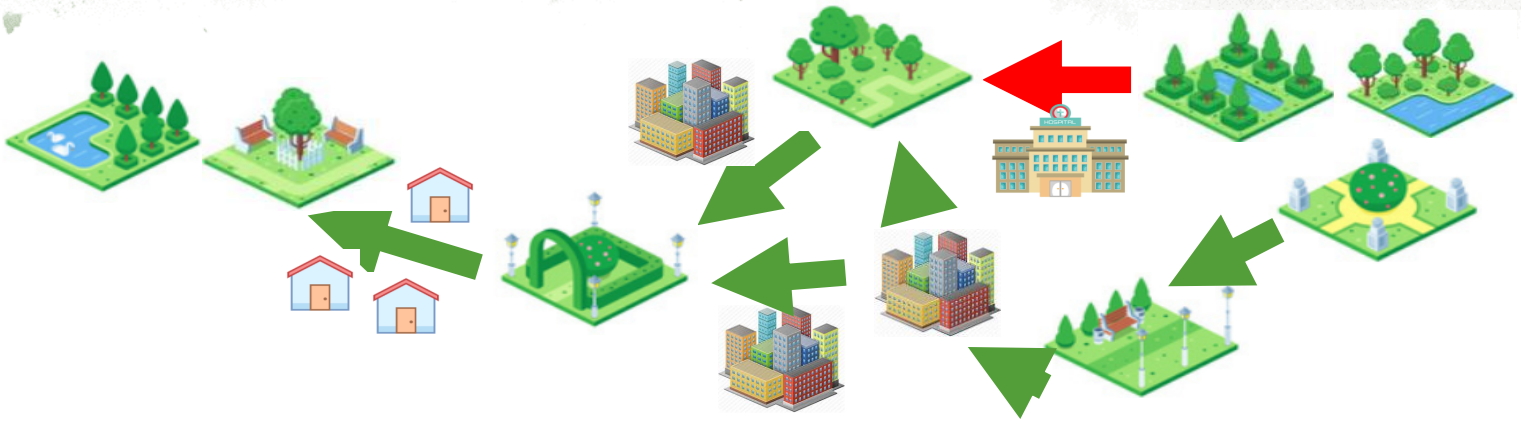
Data collection from other sources:

- Cadastre;
- Buildings structure/architecture/history;
- Urban and transport networks;
- Amenities location or prospection of such;
- Infrastructures;
- Green areas location from a city;
- Demographics (…)

GREEN BUILDINGS CORRIDORS



GREEN CORRIDOR MODEL



- City units (blocks, buildings?) modelled on undir. graph G=(V, E)
- Cost of greening modelled by c:V->[0,∞] weight function
- Already green units is an R subset of V
- Goal: find connected set of vertices A which contains R and minimizes $\sum_{v \in A} C(v)$

References:

- Conrad, Jon M., et al. "Wildlife corridors as a connected subgraph problem." Journal of Environmental Economics and Management 63.1 (2012): 1-18.

- El-Kebir, Mohammed, and Gunnar W. Klau. "Solving the maximum-weight connected subgraph problem to optimality." arXiv preprint arXiv:1409.5308 (2014).

Thank you!

Grünstattgrau Hackathon Team