AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

UNDERSTANDING MOVEMENT DATA

Seminar@SystemX

Anita Graser
@underdarkGIS

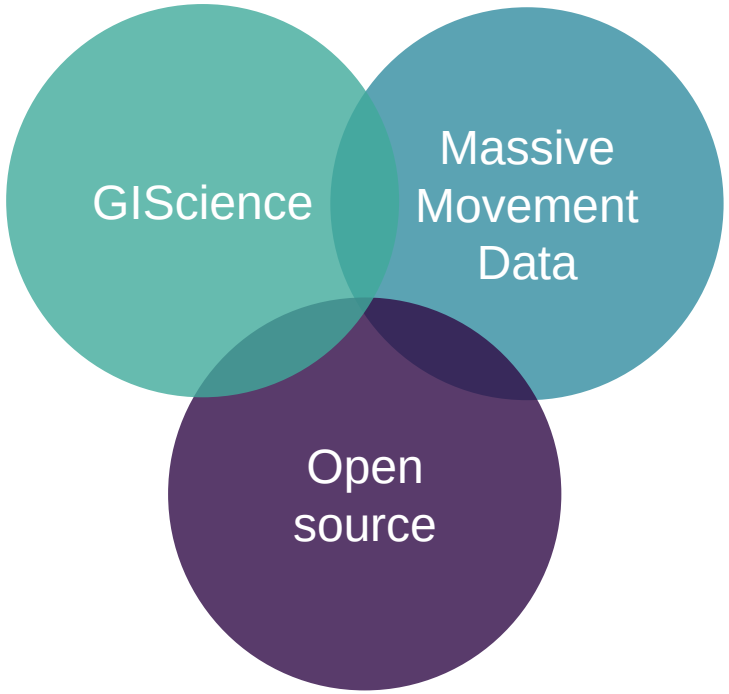
ABOUT

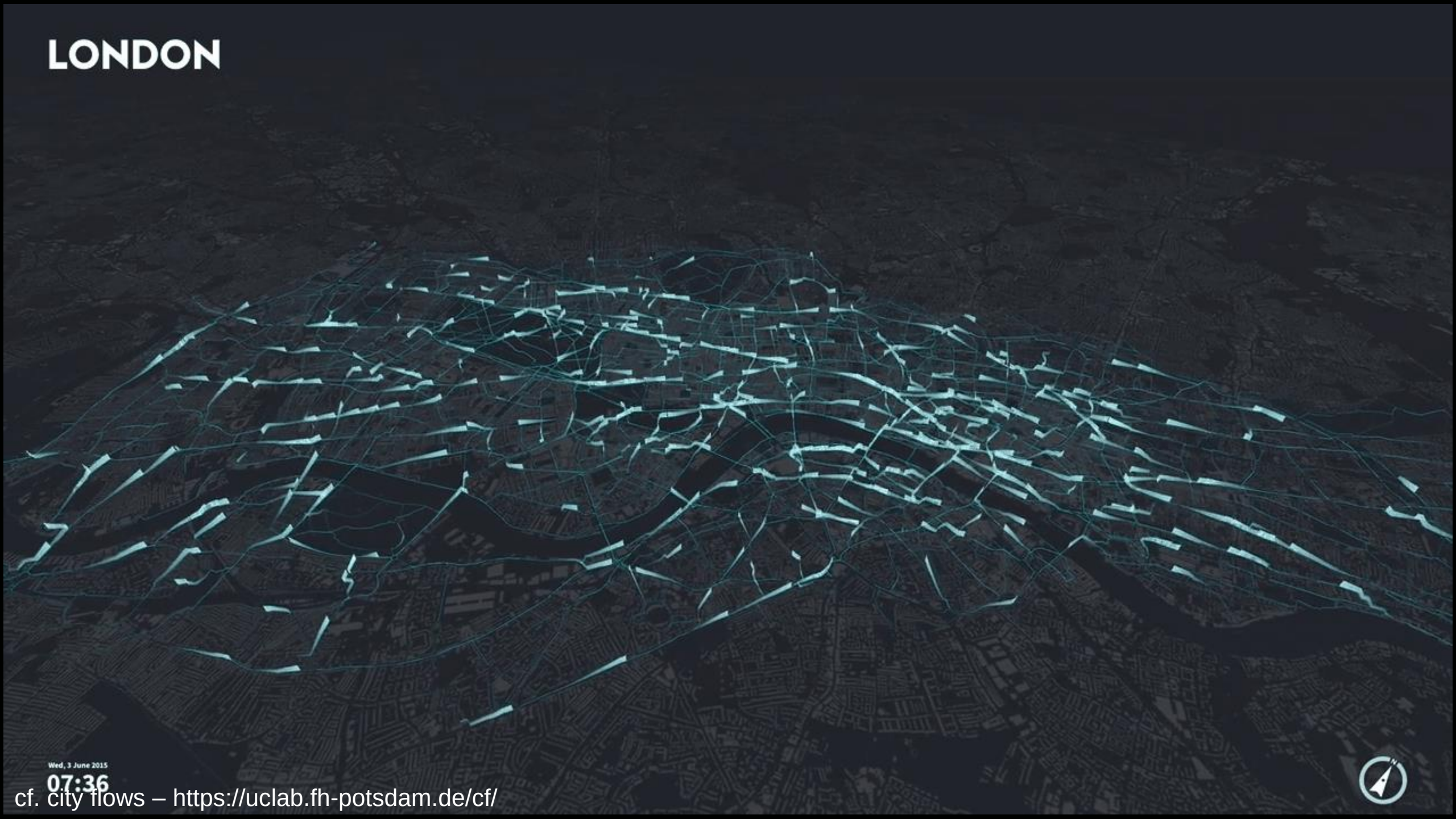
Anita Graser

Scientist @ AIT Austrian Institute of Technology

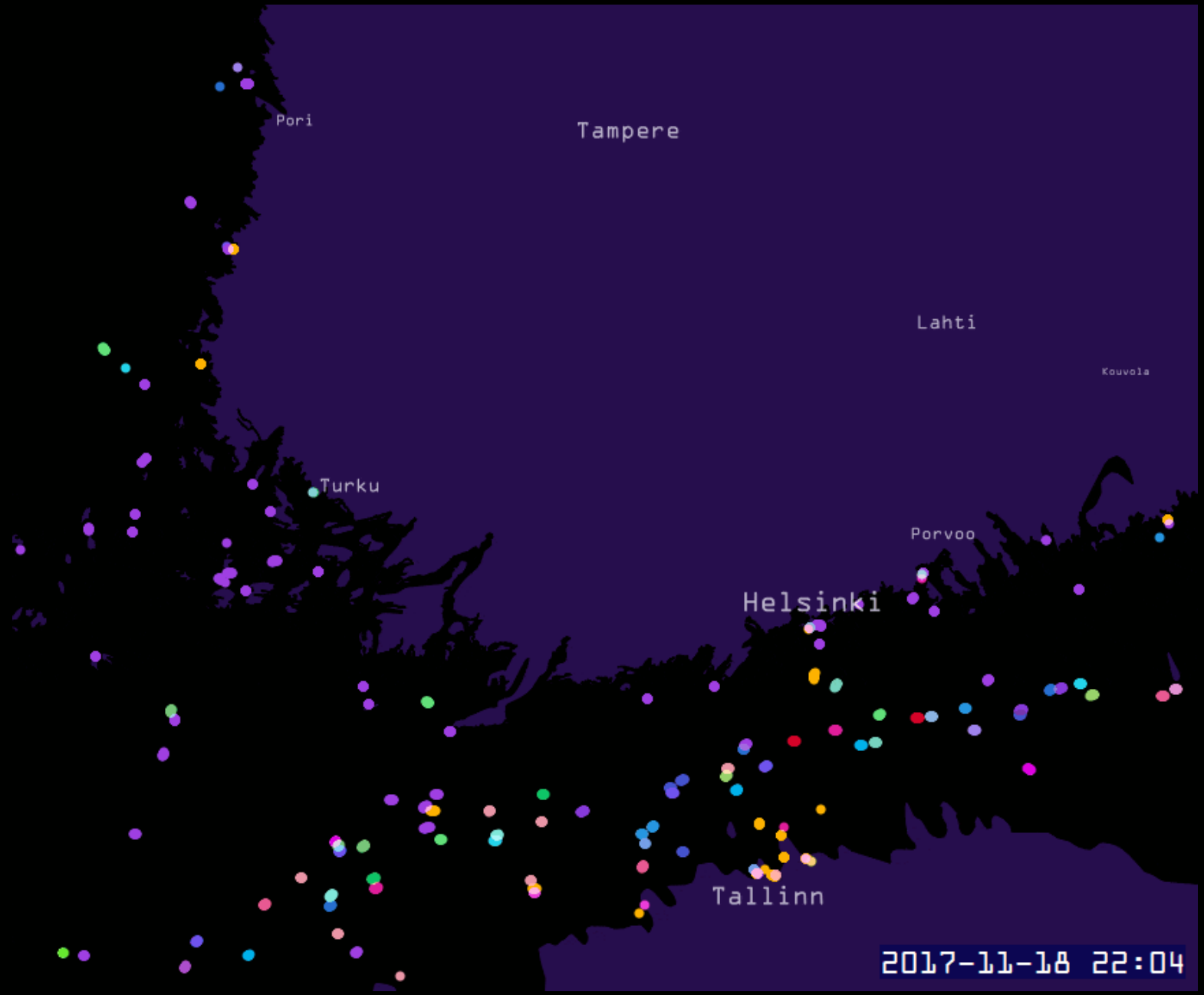
- PhD in Applied Geoinformatics
- Lead developer of MovingPandas since 2018
- MobilityDB Project Steering Committee since 2020
- QGIS Project Steering Committee since 2013
- OSGeo Director 2015-2017
- Moderator on GIS.StackExchange.com
- Author of “Learning QGIS”, “QGIS Map Design” & more

@underdarkGIS





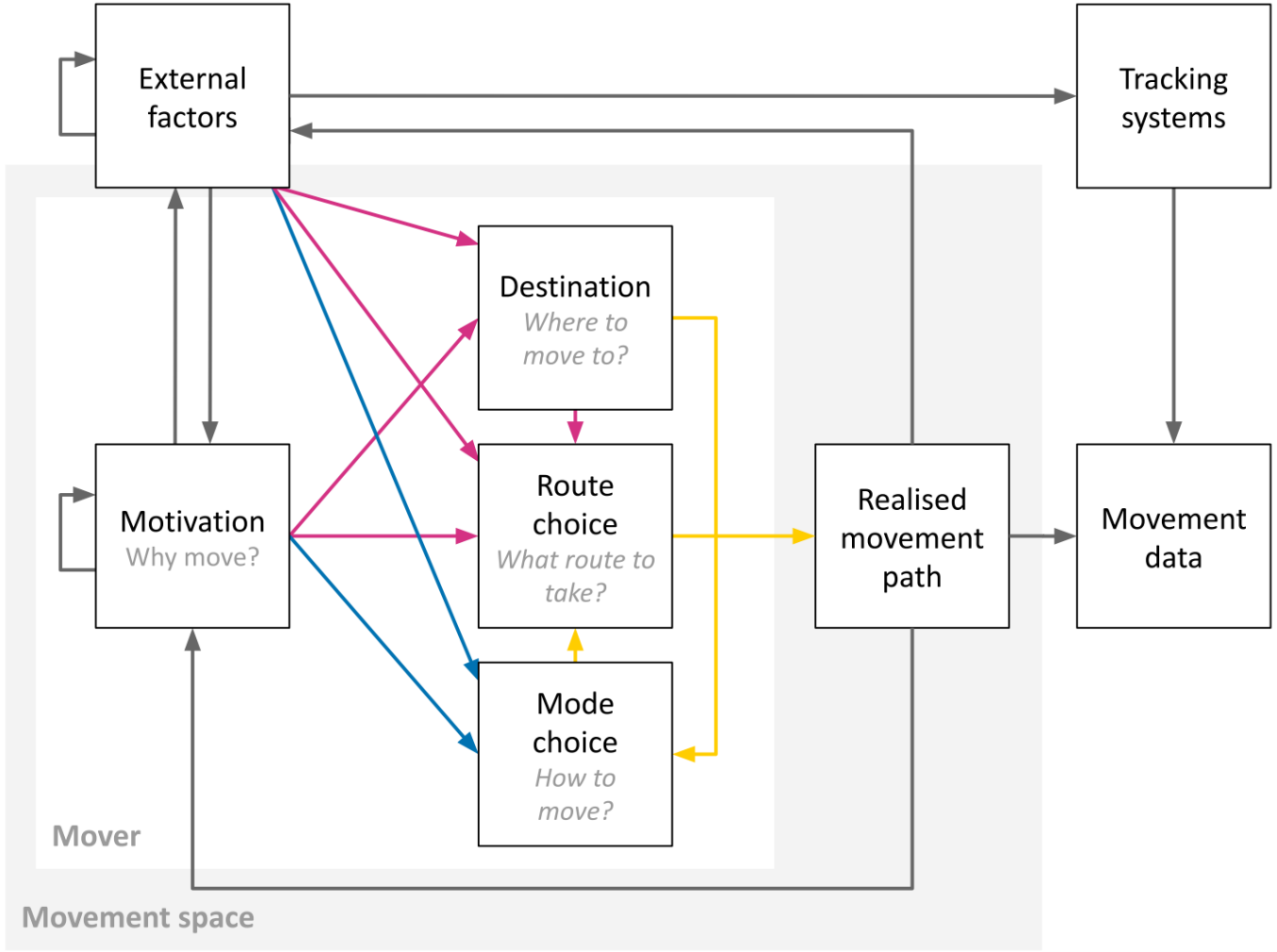
cf. city flows – https://uclab.fh-potsdam.de/cf/



Gulf of Finland – https://tjukanov.org/gulfoffinland

What does a day of bad weather look like? – https://www.youtube.com/watch?v=brX_VhOU3qQ&ab_channel=NATS

MOVEMENT RESEARCH



AUSTRIAN INSTITUTE OF TECHNOLOGY

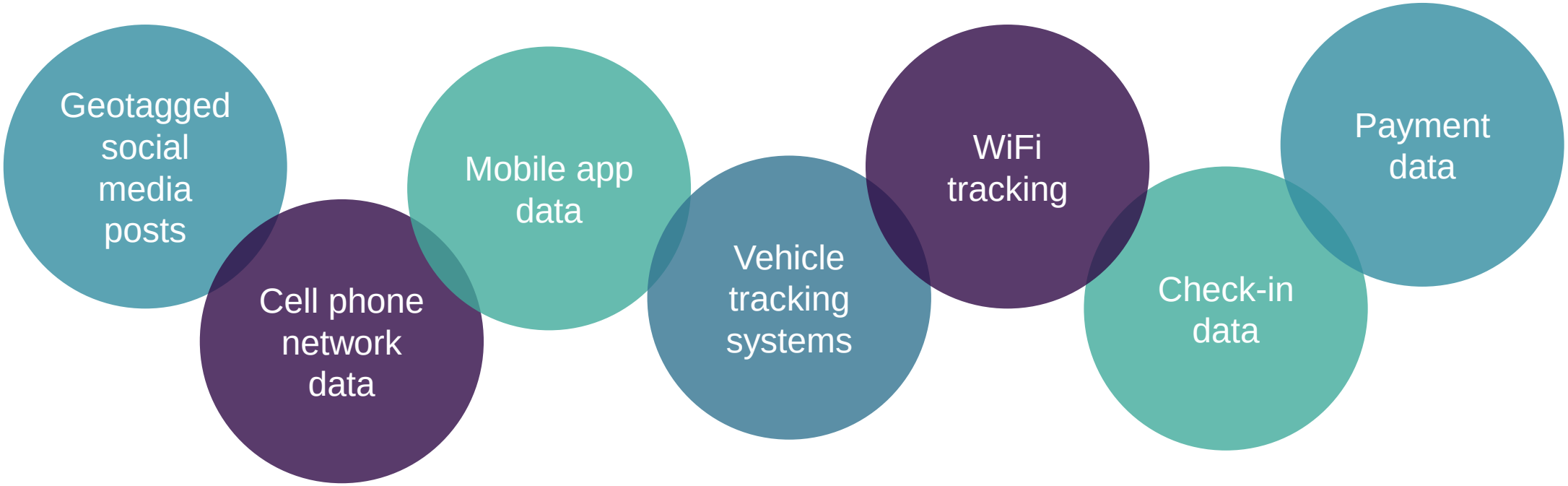
MOVEMENT DATA SCIENCE

- ✓ **Opportunistic reuse** of data

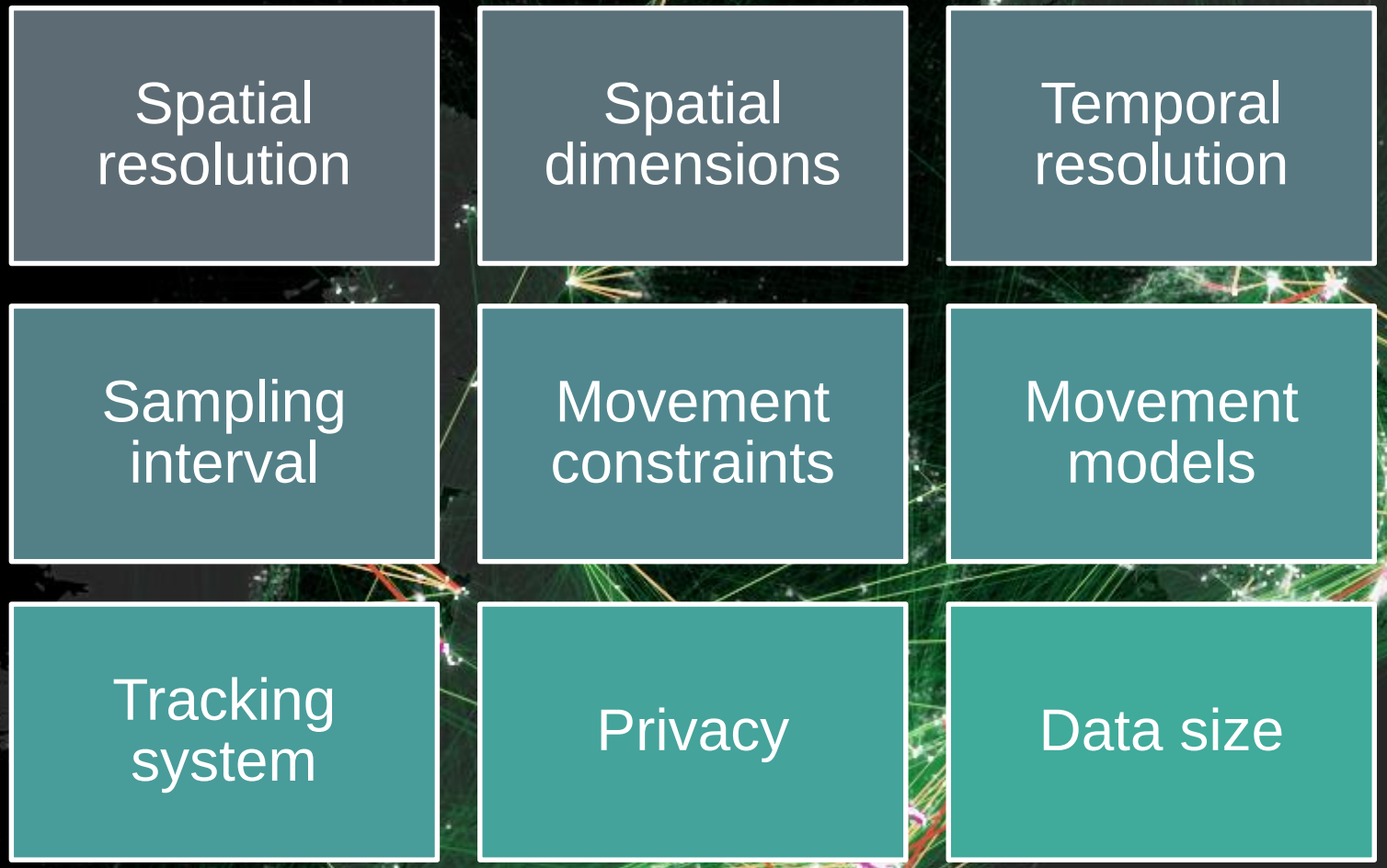
  → Black box / undocumented data collection

  → Usually biased & messy data

"All metadata records are incomplete as it is impossible to foresee future uses" Janowicz et al. (2020) GeoAI



DIMENSIONS OF MOVEMENT DATA



REALITY CHECK

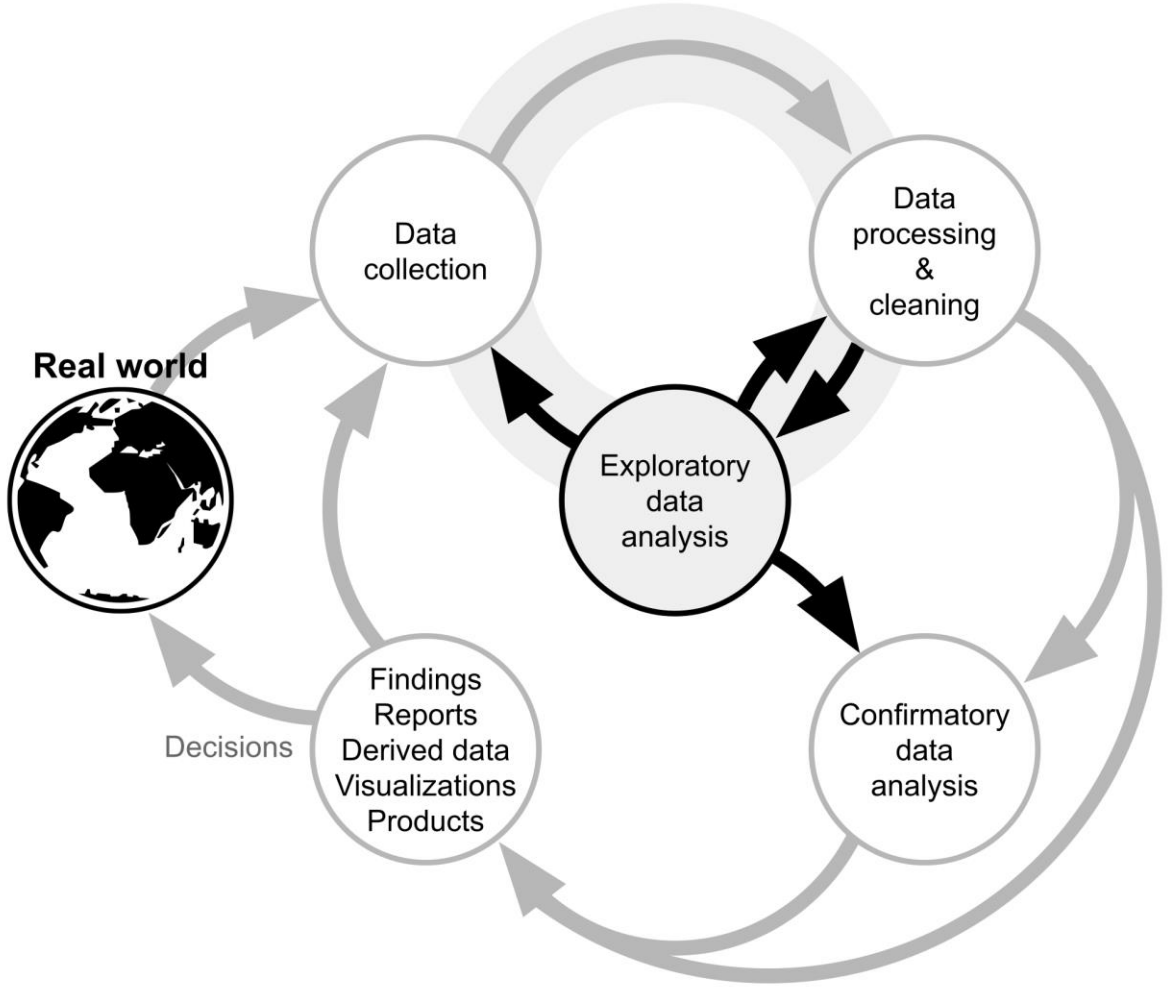
"Data preparation is typically the most time-consuming step in the analysis (Pyle 1999).“

Andrienko et al. 2016

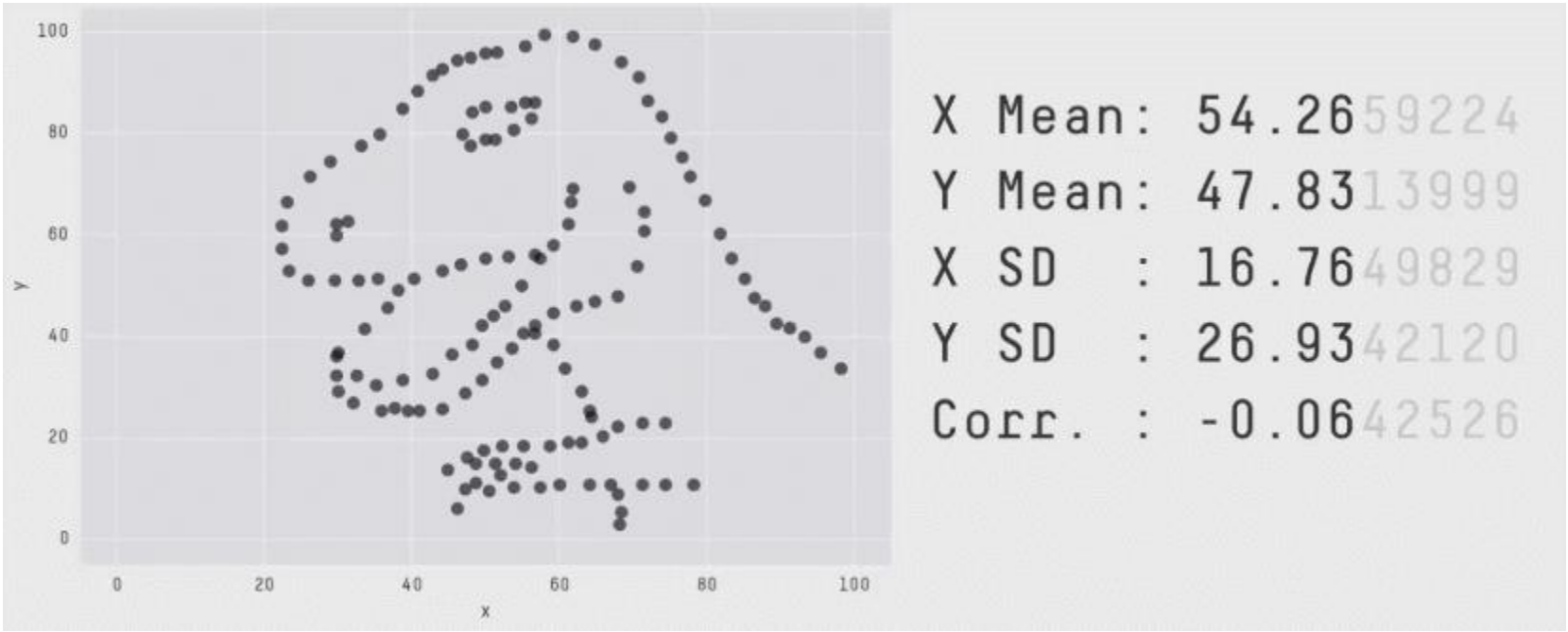
“Data exploration can take up to 50% of the time spent on analysis"

Zuur et al. 2010

EXPLORATORY DATA ANALYSIS



VISUAL TOOLS!



https://www.autodeskresearch.com/publications/samestats

REALITY CHECK

- Rarely collected in lab settings
- Usually created for different purposes

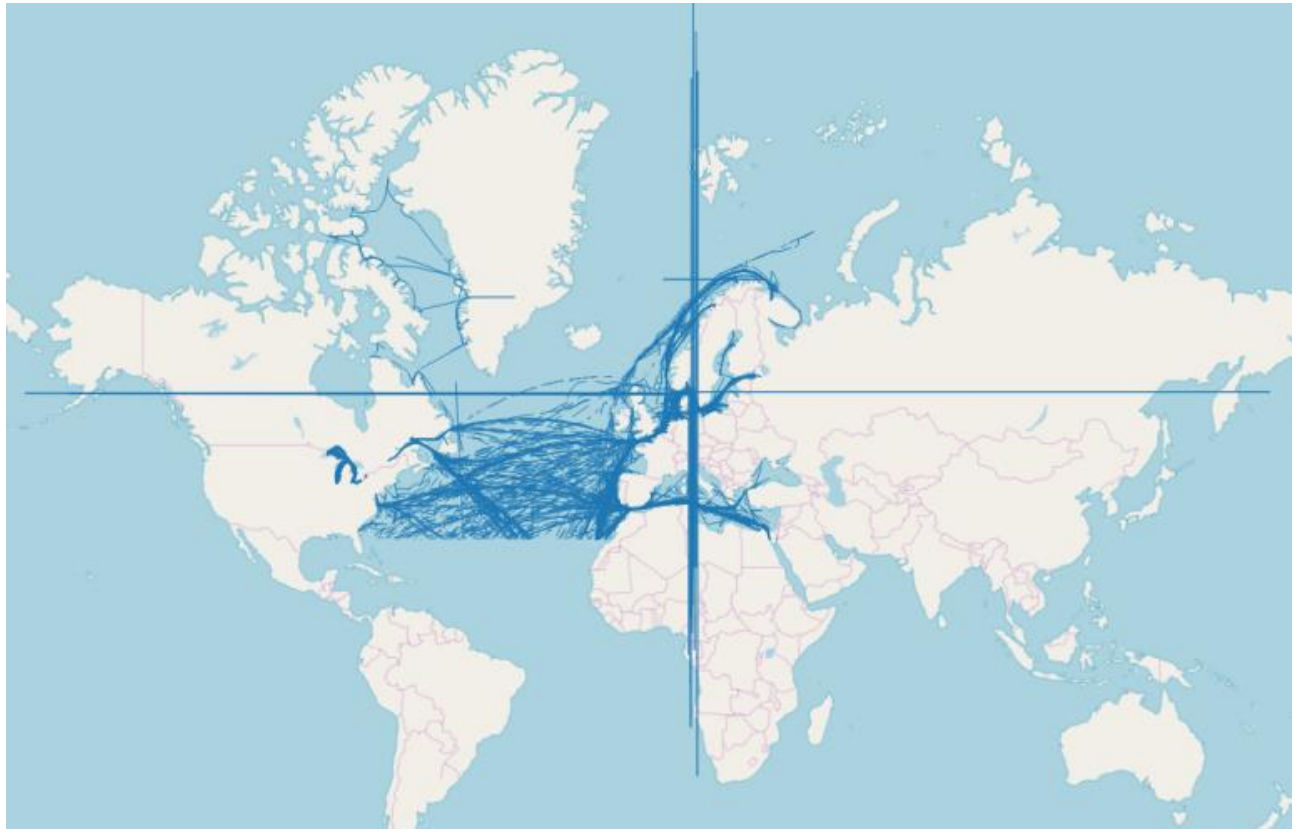
➔ Data quality is usually NOT ideal

REALITY CHECK

Figure 1. Segments are missing in many trajectories due to cutting of the data by a bounding rectangle.

Andrienko et al. (2016) Understanding movement data quality. Journal of Location Based Services.

REALITY CHECK



Graser et al. (2020) Exploratory Trajectory Analysis for Massive Historical AIS Datasets

REALITY CHECK

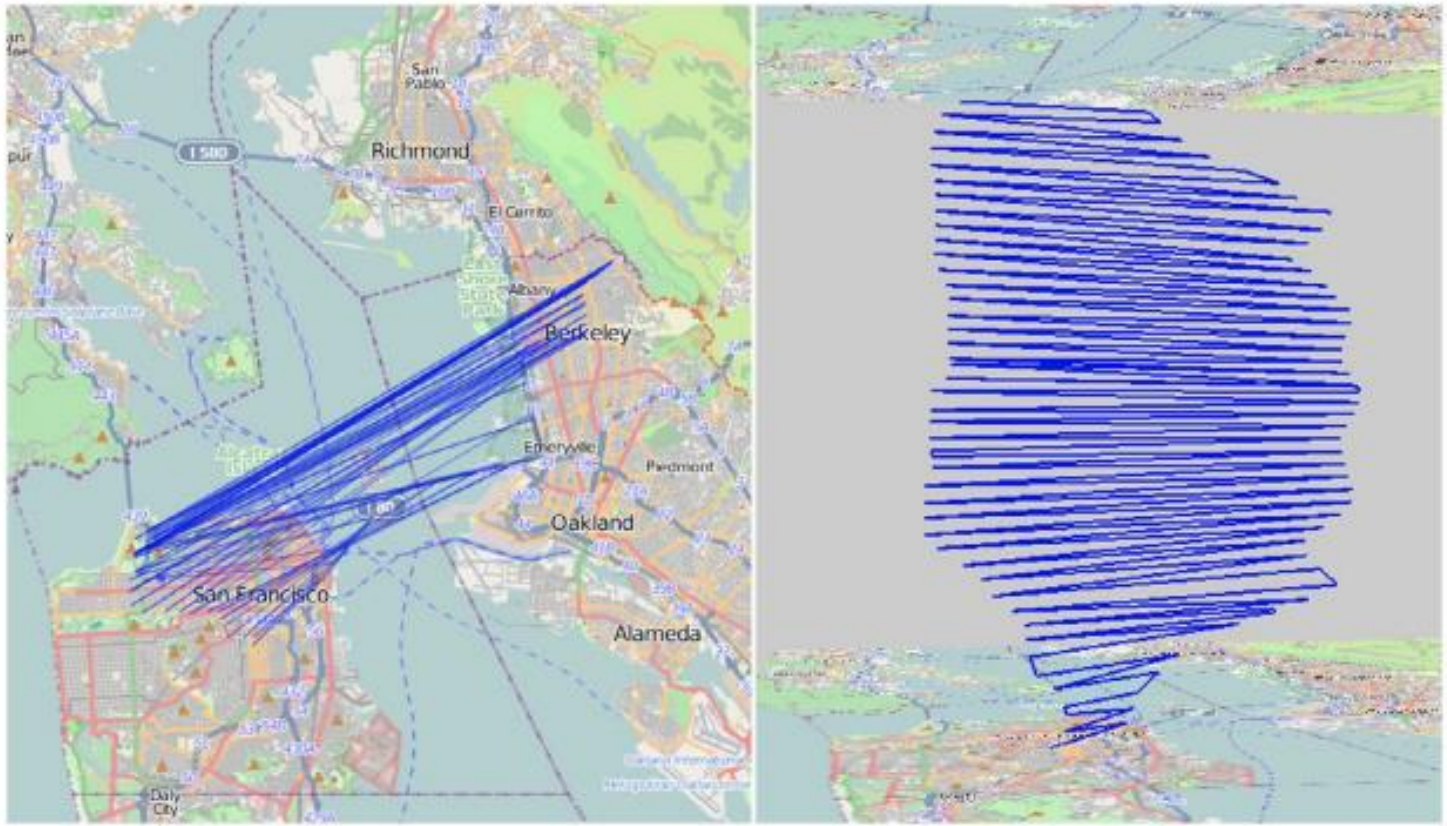


Figure 5. Positions of two distinct movers with the same identifier have been mixed in one trajectory, which therefore has a zigzag shape.

Andrienko et al. (2016) Understanding movement data quality. Journal of Location Based Services.

REALITY CHECK

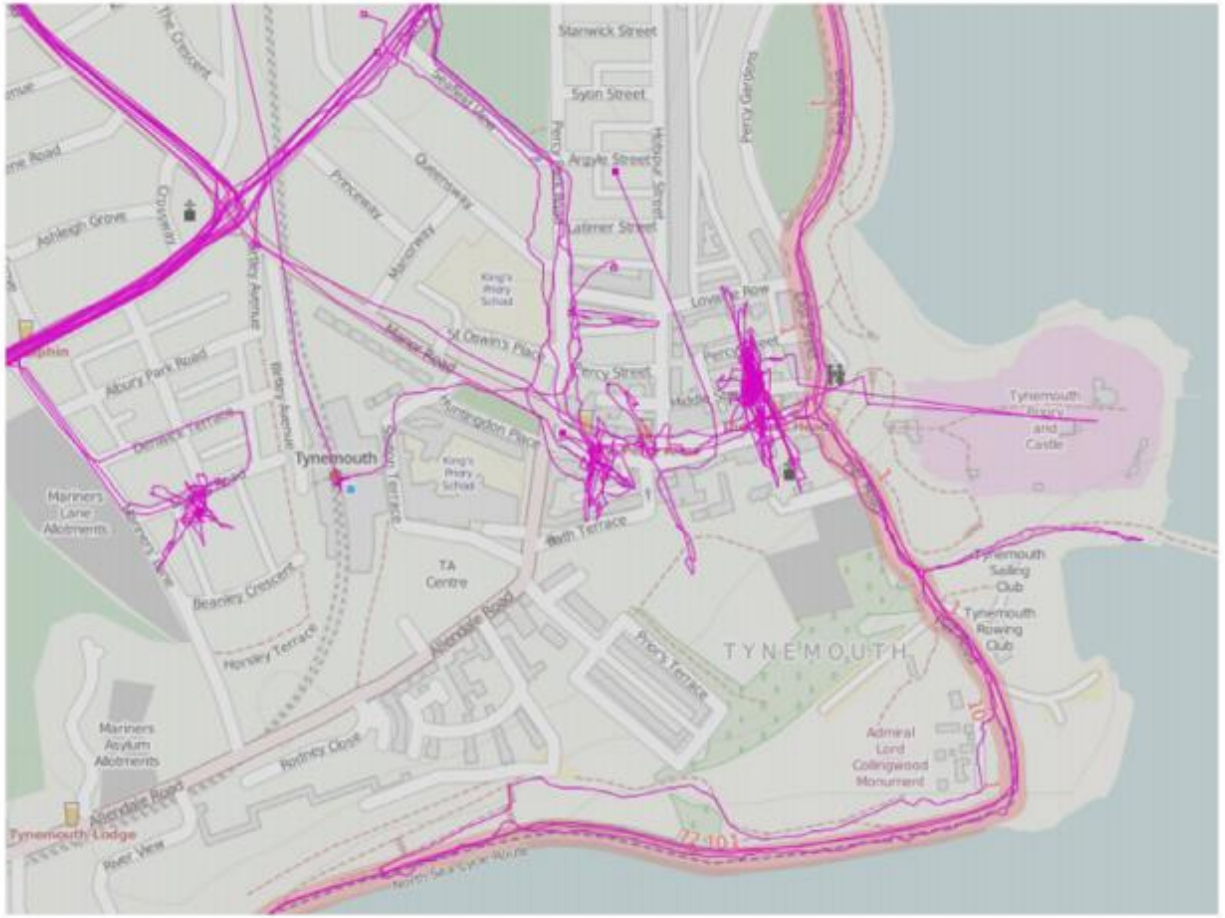


Figure 9. Stop positions appearing as jittered movement.

Andrienko et al. (2016) Understanding movement data quality. Journal of Location Based Services.

REALITY CHECK

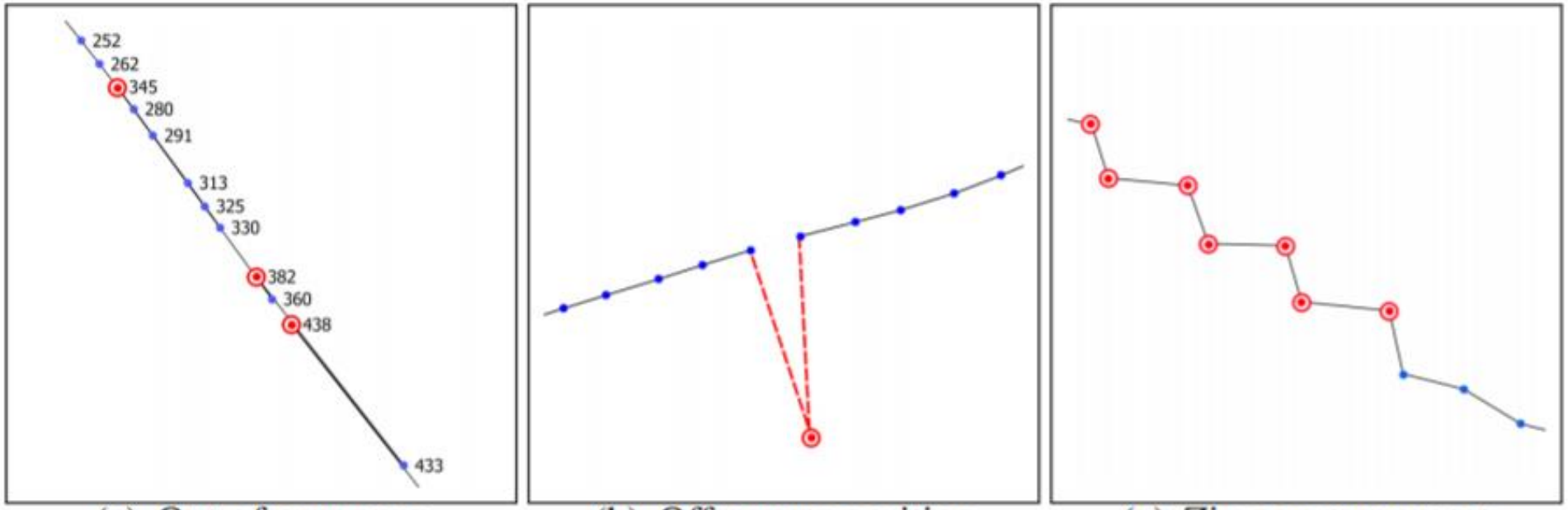
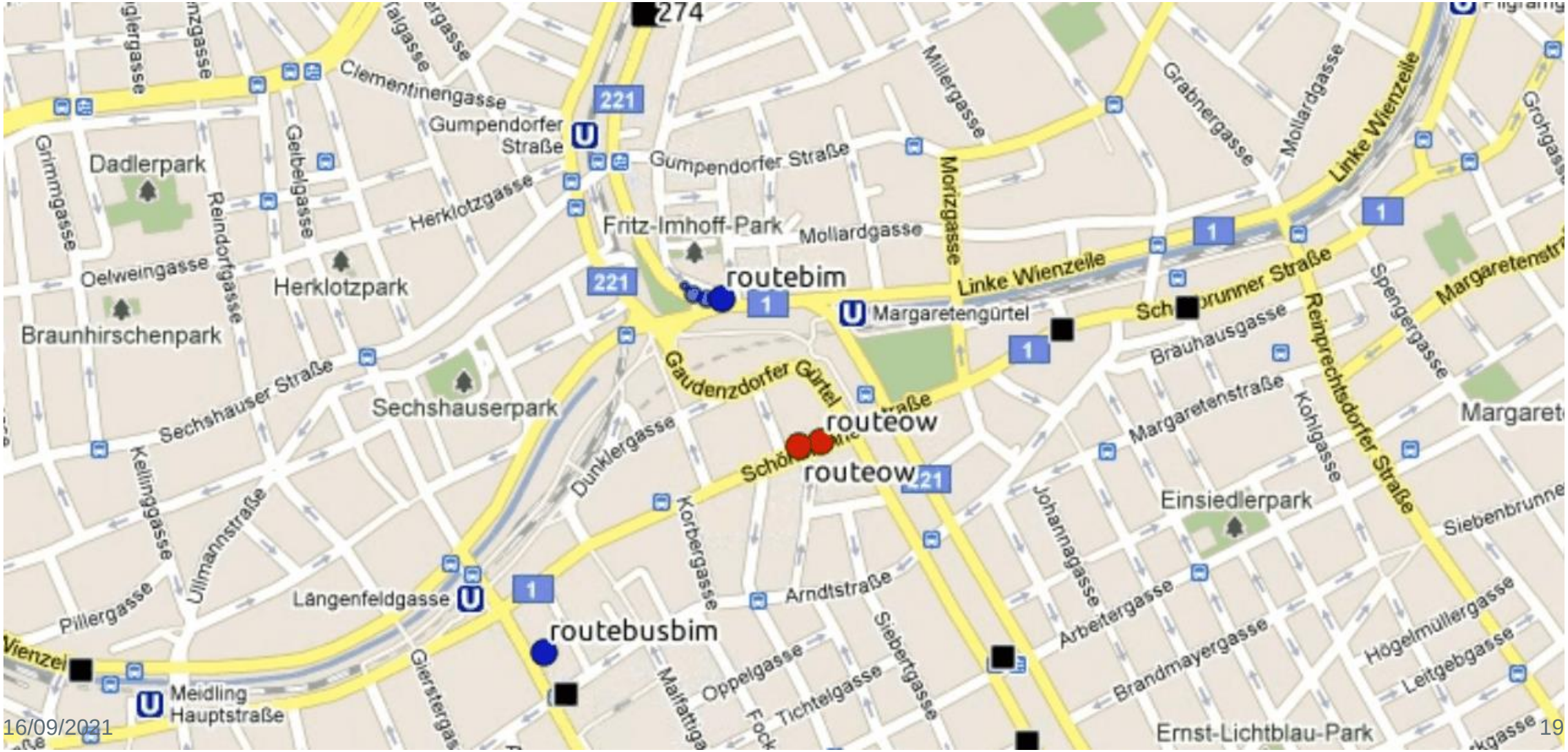


Fig. 2 Noise-related situations along a vessel's course.

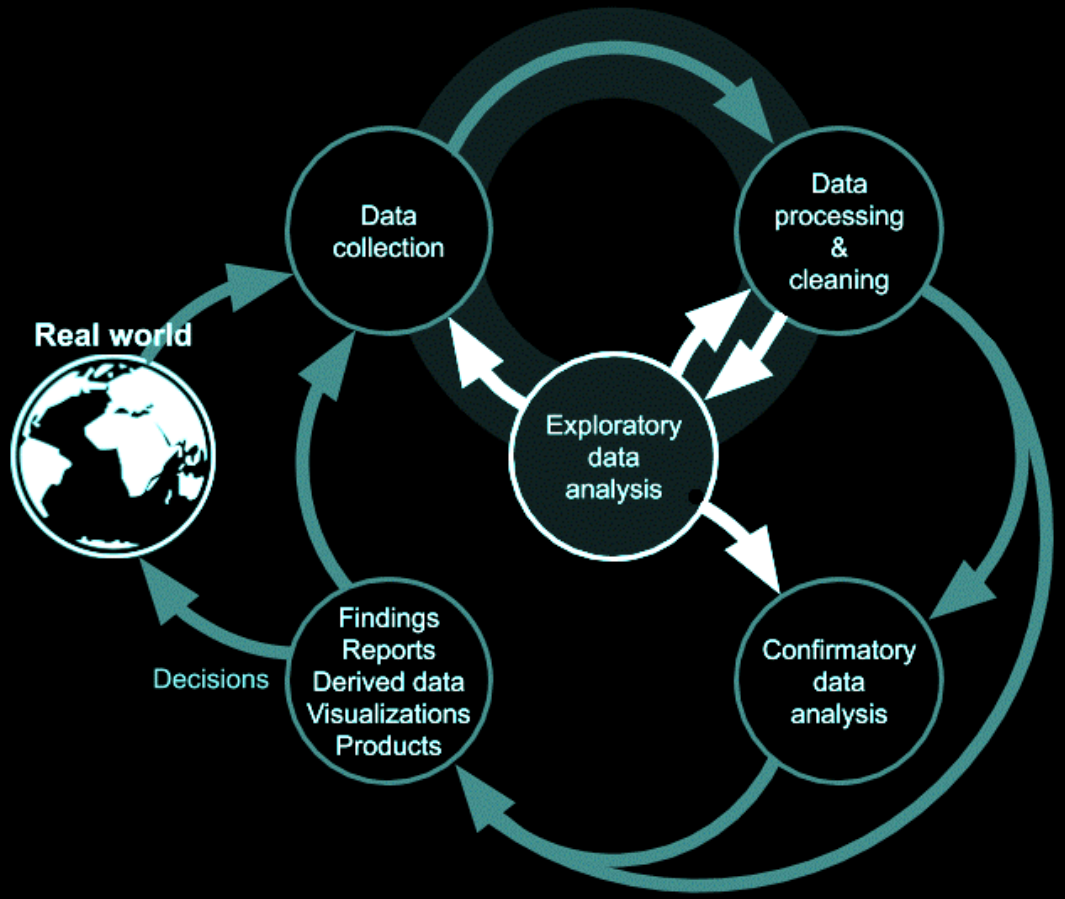
Patroumpas et al. (2017) Online event recognition from moving vessel trajectories. GeoInformatica.

REALITY CHECK



EXPLORATORY DATA ANALYSIS (EDA)

- ✓ Suggest hypotheses
- ✓ Assess assumptions
- ✓ Select statistical tools & techniques
- ✓ Inform further data collection



SYSTEMATIC EDA PROTOCOL

Guide for identifying problems in quasi-continuous movement data

→ Technology agnostic conceptual protocol

Table 1: Protocol steps by type of data quality problem and required processing level.

	Elementary records	Intermediate segments	Overall trajectories
A) Missing data	A-1 Spatial gaps & outliers A-2 Temporal gaps & outliers A-3 Spatiotemporal gaps A-4 Attribute gaps	A-5 Gaps in tracks	
B) Precision problems		B-1 Coordinate imprecision B-2 Timestamp imprecision	
C) Consistency problems		C-1 Sampling heterogeneity	C-2 Mover heterogeneity C-3 Tracker heterogeneity
D) Accuracy problems		D-2 Spatial inaccuracy (jumps) D-3 Temporal inaccuracy	D-1 Object identity issues D-2 Spatial inaccuracy (noise)

Graser (2021) An exploratory data analysis protocol for identifying problems in continuous movement data. Journal of Location Based Services.

SYSTEMATIC EDA PROTOCOL

Guide for identifying problems in quasi-continuous movement data

→ Technology agnostic conceptual protocol

→ Open source reference implementation

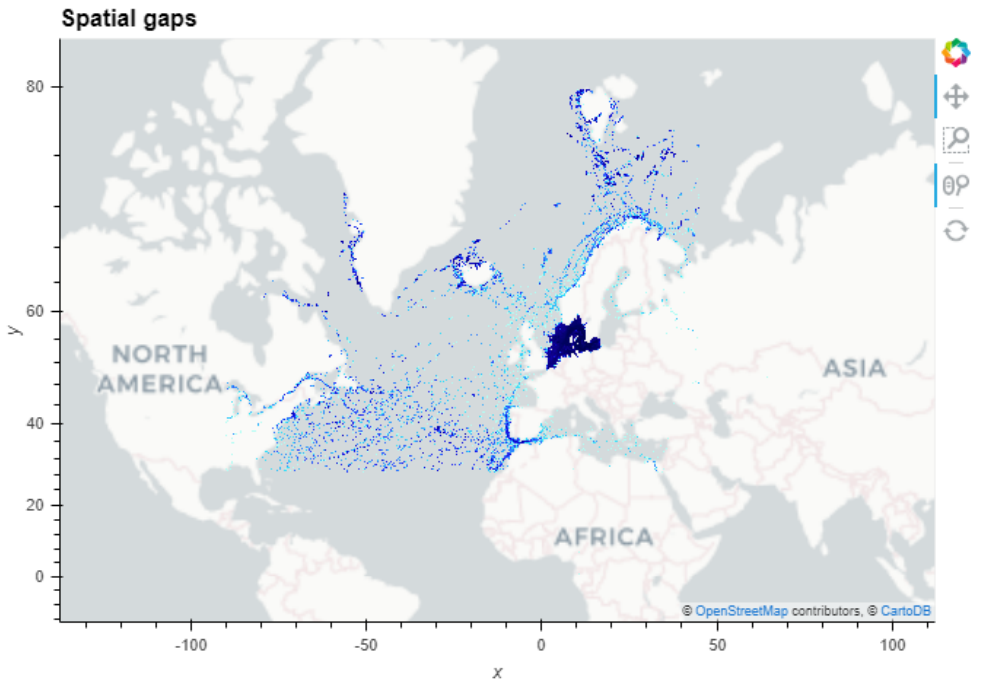
https://zenodo.org/record/4599278

Spatial gaps (selected areas / all movers / whole time span)

In [18]:
```
def plot_point_density(df, title='', width=FIGSIZE[0], height=FIGSIZE[1]):
    opts.defaults(opts.Overlay(active_tools=['wheel_zoom']))
    pts = df.hvplot.scatter(x='x', y='y', title=str(title), datashade=True, frame_width=width, frame_height=h
    return BG_TILES * pts
```

In [19]:
```
plot_point_density(df, title='Spatial gaps')
```

Out[19]:

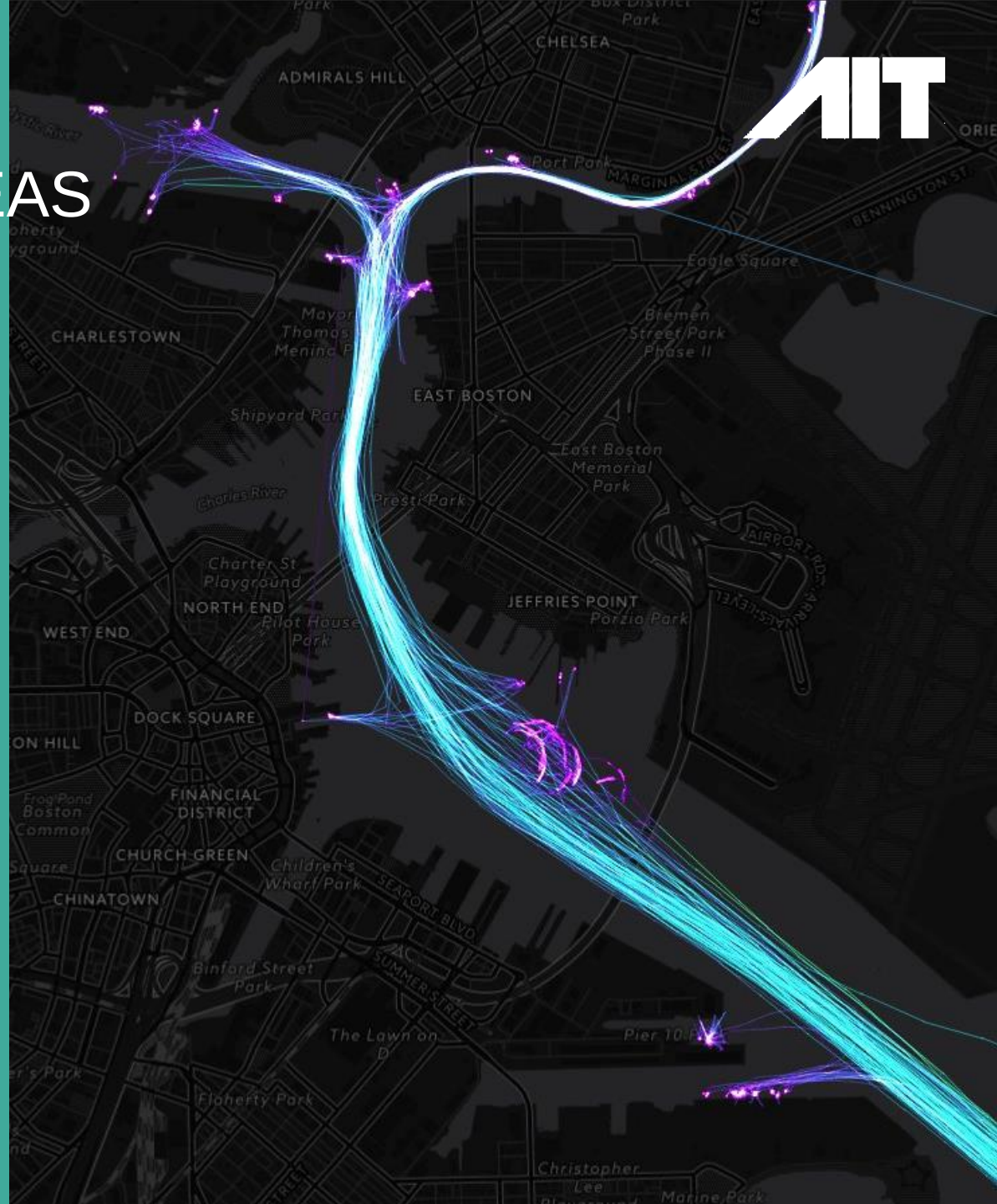


AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

LIVE DEMO

CHALLENGES & RESEARCH AREAS

1. Standardization of movement data models
2. EDA methods & tools for other massive movement data types
3. Bias in massive movement data
4. Data privacy versus EDA
5. Accessibility of EDA tools for massive movement data



AUSTRIAN INSTITUTE OF TECHNOLOGY

ANITA GRASER

 anita.graser@ait.ac.at

 @underdarkGIS

 anitagraser.com