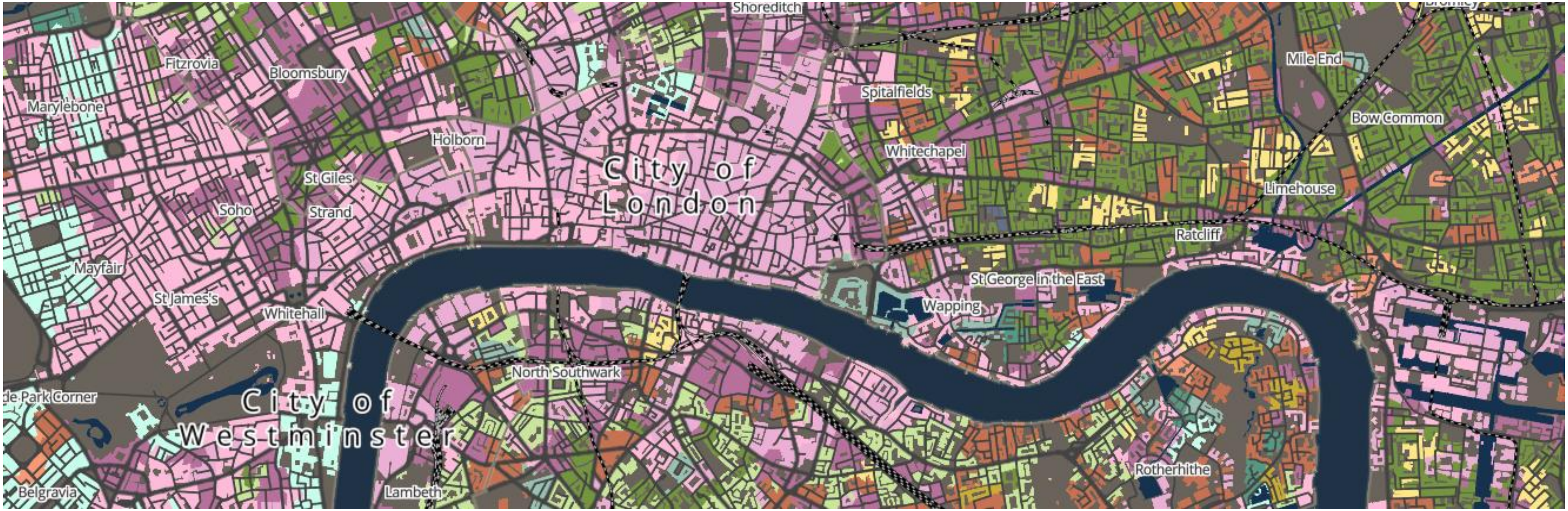
Geocomputation

Point Pattern Analysis



Module outline

W1 Reproducible Spatial Analysis
W2 Spatial Queries and Geometric Operations
W3 Point Pattern Analysis
W4 Spatial Autocorrelation
W5 Spatial Models
W6 Raster Data Analysis
W7 Geodemographic Classification
W8 Accessibility Analysis
W9 Beyond the Choropleth
W10 Complex Visualisations

Core Spatial Analysis

Applied Spatial Analysis

Data Visualisation

This week

- Focus analysis directly on point events rather than aggregating to administrative geography.
- Three main ways to describe or characterise point processes:
  - Descriptive statistics.
  - Distance-based methods: e.g. average nearest neighbour, Ripley’s K.
  - Density-based methods: e.g. DBSCAN, kernel density estimation.
- Some practical examples.

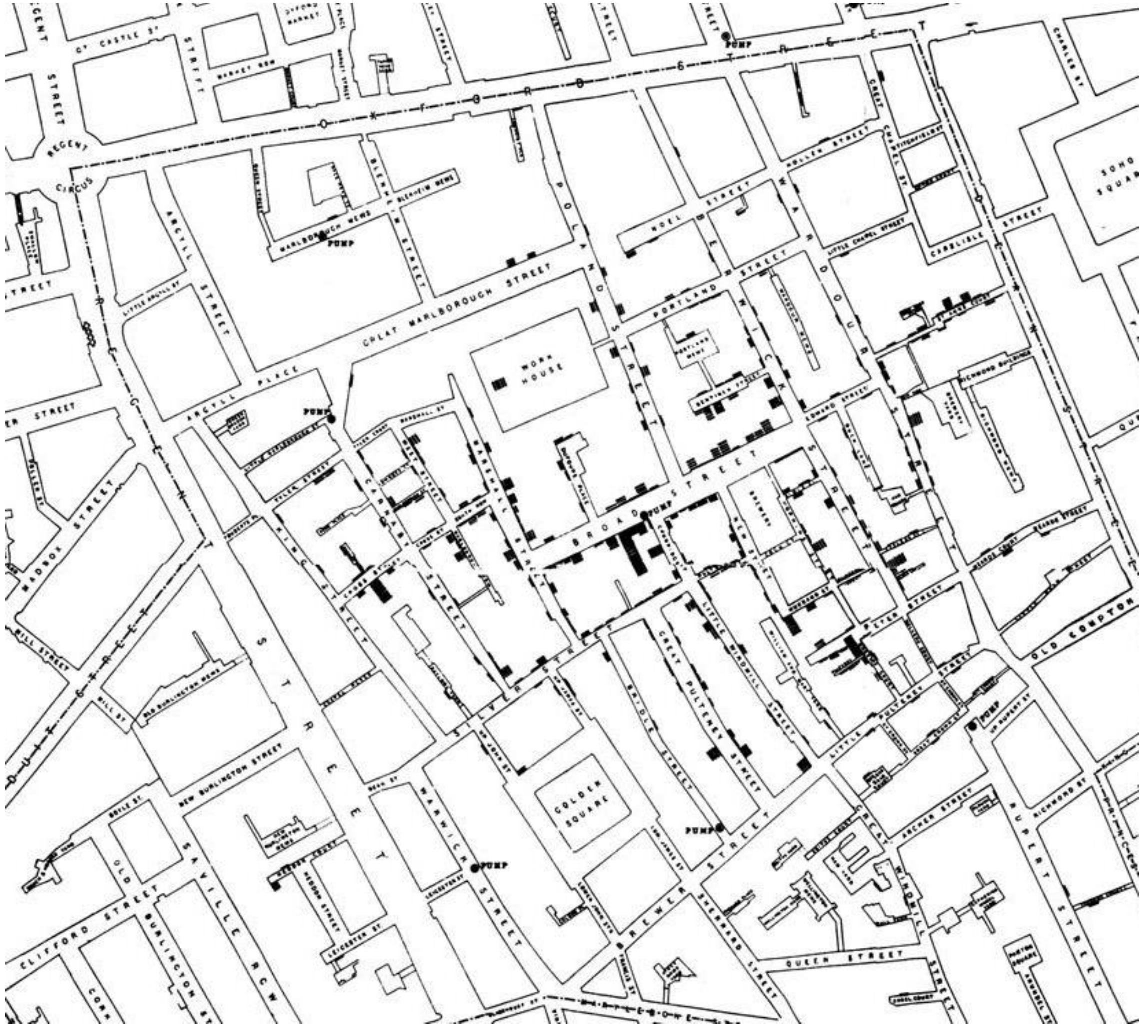
Before we start

- Go to www.menti.com
- Use code: 5997 0800

Why do we want to analyse points?

- Points represent the location of individuals, features, or events, providing a spatial reference for analysis.
- Spatial patterns within the distribution of these points can reveal underlying relationships and trends.
- Analysing points allows us to study the occurrence of phenomena or events in relation to similar attribute values (non-spatial information).

Spatial analysis

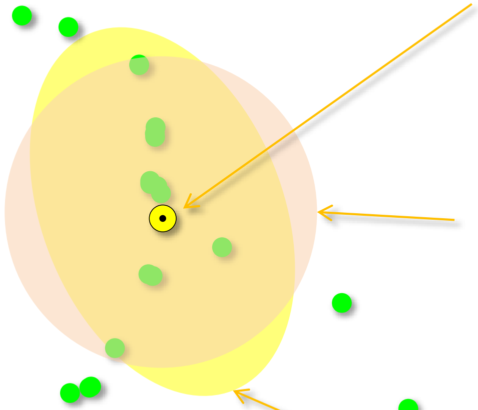


How do we analyse points?

- Descriptive statistics.
- Density-based methods.
- Distance-based methods.

Descriptive statistics

Descriptive statistics



Mean center
computed average X and Y coordinate values.

$$\bar{s} = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right)$$

Standard distance
measure of the variance between the average distance of the features to the mean center.

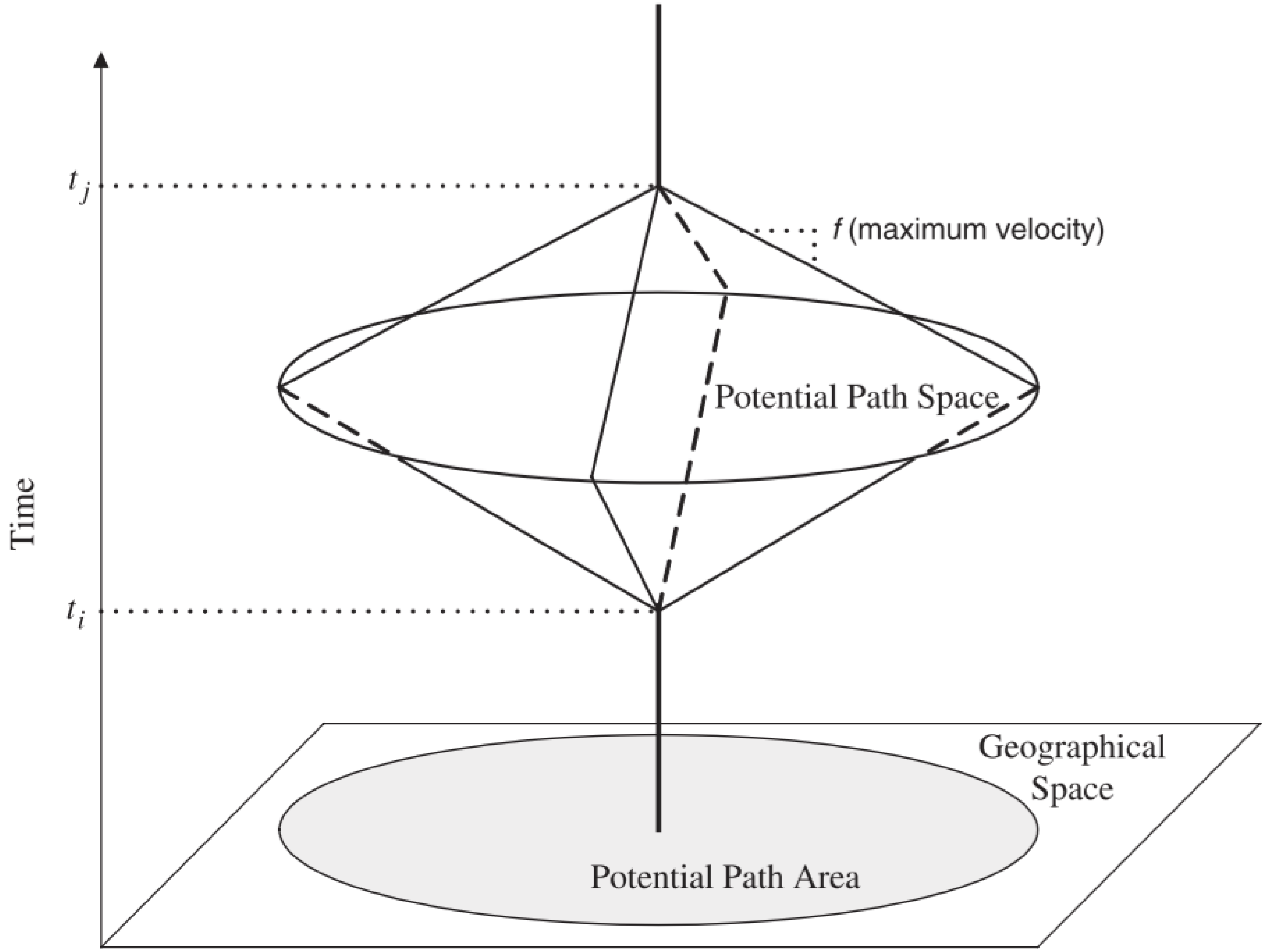
$$d = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_x)^2 + (y_i - \mu_y)^2}{n}}$$

Standard deviational ellipse
separate standard distances for each axis.

$$\left\{ \begin{aligned} d_x &= \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_x)^2}{n}} \\ d_y &= \sqrt{\frac{\sum_{i=1}^{n} (y_i - \mu_y)^2}{n}} \end{aligned} \right.$$

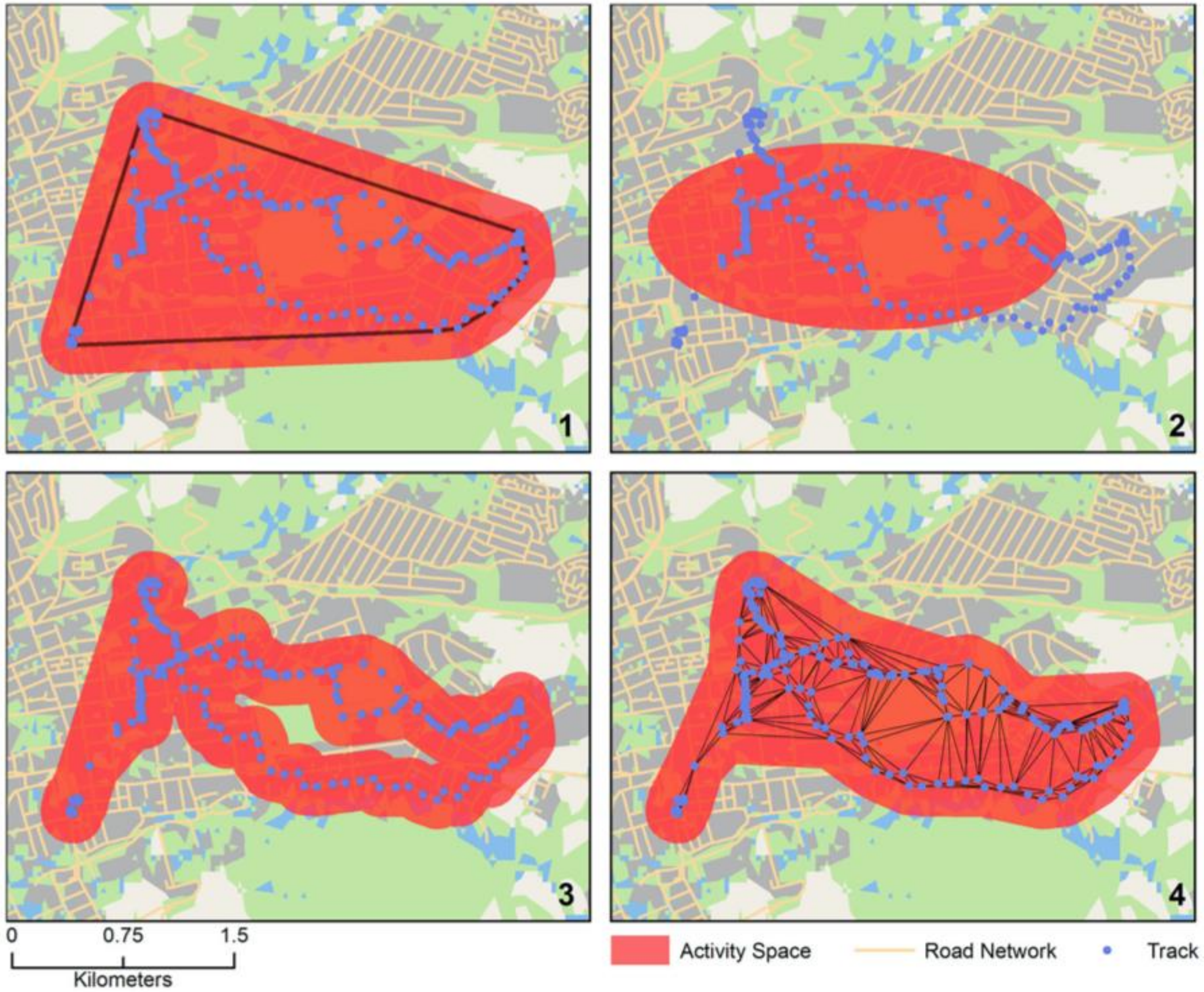
Gimdond, M. 2025. Intro to GIS and Spatial Analysis. [online]
https://mgimond.github.io/Spatial/introGIS.html

Descriptive statistics



Worobey *et al*. 2022

Descriptive statistics



Distance-based measures

Distance-based measures

- Using distance as a summary measure to summarise a point process.
- Measure of spread of the distribution.

Average Nearest Neighbour

- ANN measures the distance between each feature and its nearest neighbour, then averages all these nearest-neighbour distances.
- If the average distance is smaller than the average for a hypothetical random distribution, the features are considered clustered. If the average distance is larger, the features are considered dispersed.

Ripley’s K

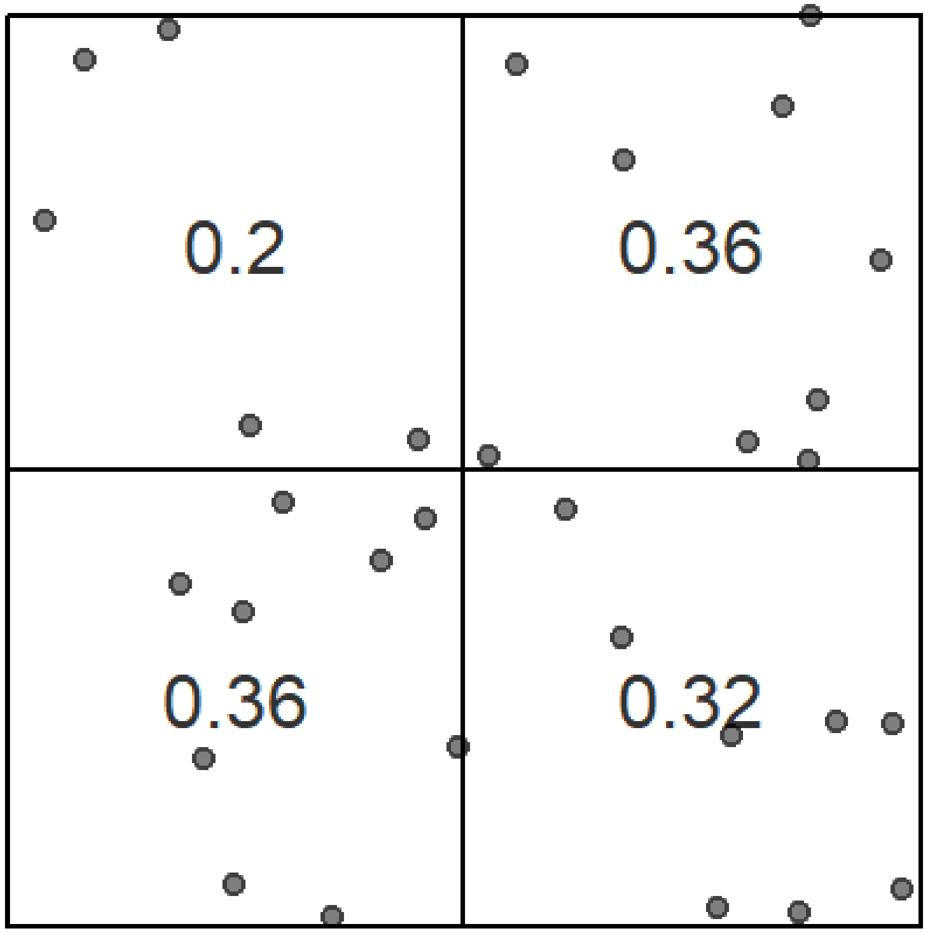
- Ripley’s K counts the number of observations within a user defined set of distances and compares this to a hypothetical (random) pattern of observations.
- Ripley’s K function is generally calculated at multiple distances allowing you to see how point pattern distributions can change with scale. For example, at near distances, the points could cluster, while at farther distances, points could be dispersed.
- Distance-based measure of dispersion across scales.

Density-based measures

Density-based measures

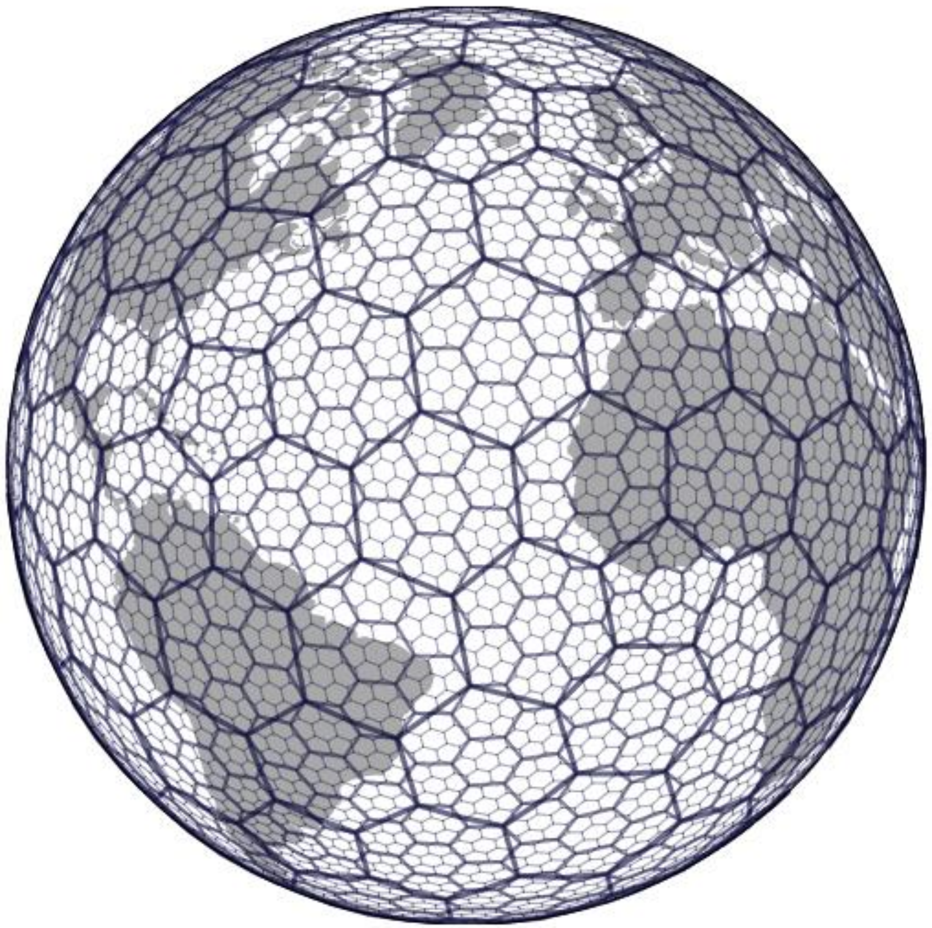
- Measure of the intensity of a point process.
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to find high-density clusters and distinguish noise points.
- Kernel Density Estimation (KDE) smooths point data to estimate the density of points across a given area.

Density-based measures



Gimdond, M. 2025. Intro to GIS and Spatial Analysis. [online]
https://mgimond.github.io/Spatial/introGIS.html

Density-based measures

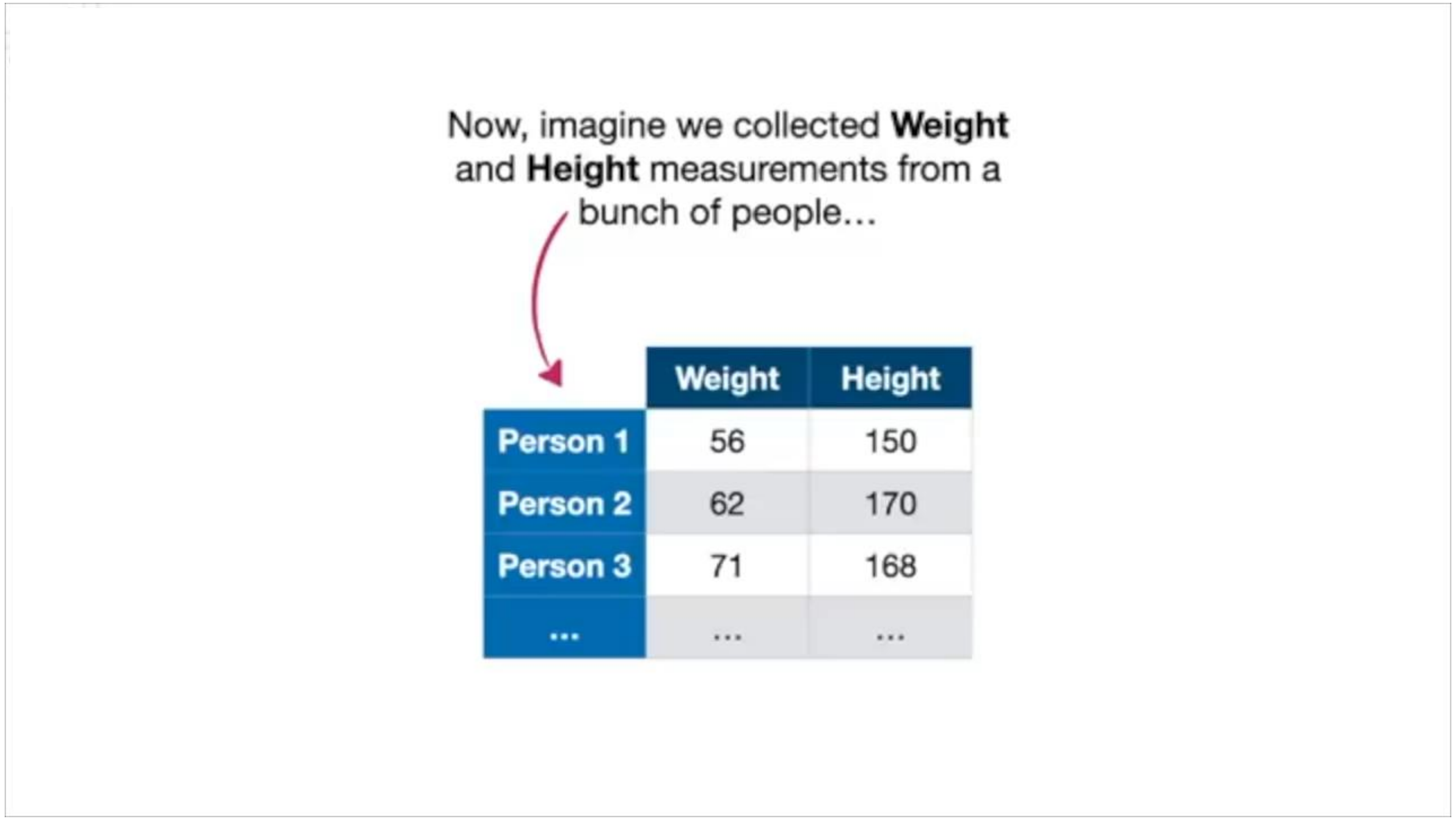


Uber. 2018. *H3: Uber's Hexagonal Hierarchical Spatial Index*. [online]
https://eng.uber.com/h3/

DBSCAN

- Density-Based Spatial Clustering of Applications with Noise (DBSCAN).
- Functionally, DBSCAN detects clusters of points and noise by evaluating density.
- It relies on two key parameters: `minpts` (the minimum number of points required to form a cluster) and `epsilon` (the maximum distance, or search radius, between points to be considered part of the same cluster).
- Points within a cluster must have at least `minpts` points within a radius of `epsilon` to be assigned to that cluster; points that do not meet this condition are classified as noise.

DBSCAN



	Weight	Height
Person 1	56	150
Person 2	62	170
Person 3	71	168
...	...	...

StatQuest. *Clustering with DBSCAN*. [online]
https://www.youtube.com/watch?v=RDZUdRSDOok

DBSCAN

Pros:

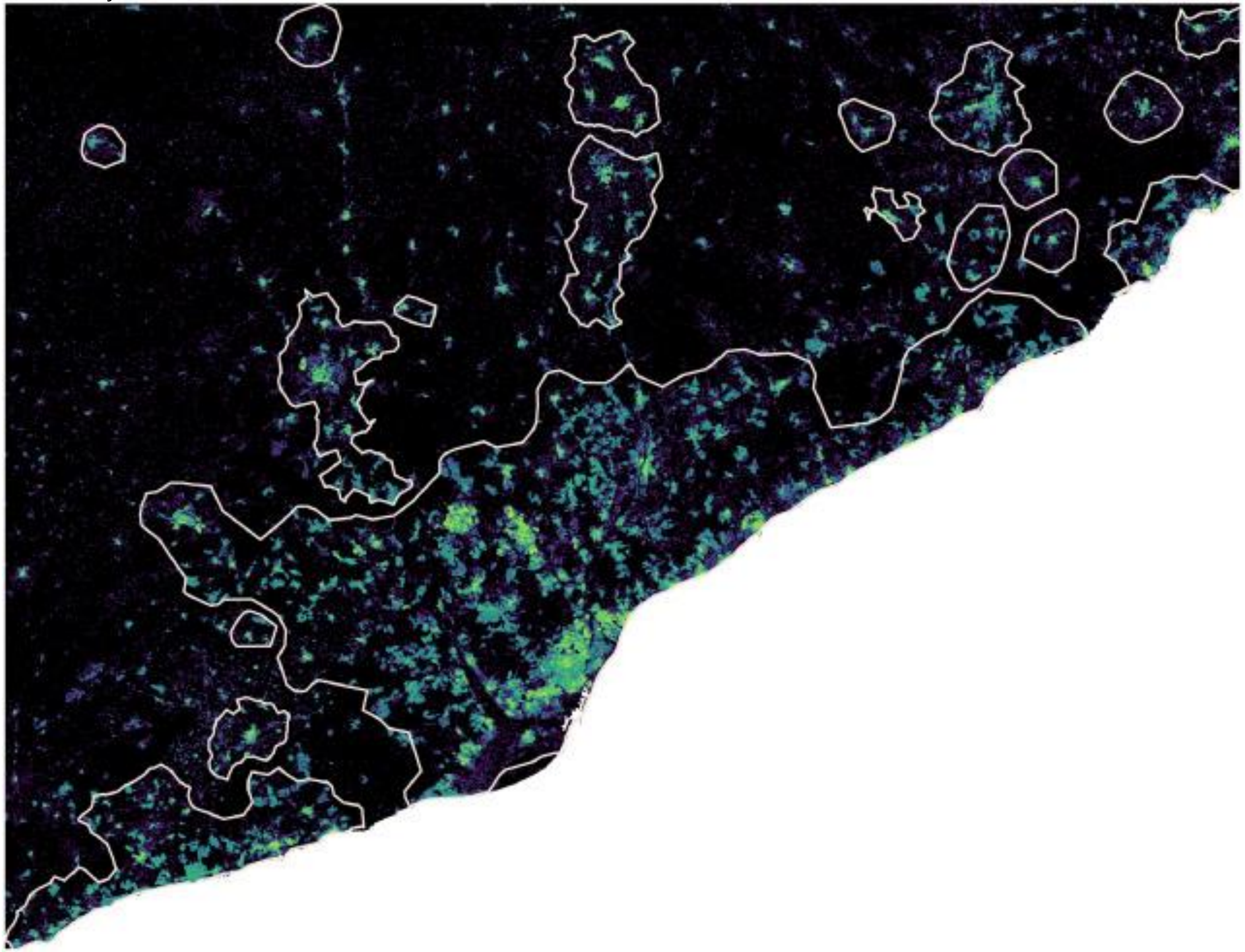
- Does not require specifying the number of clusters beforehand.
- Capable of identifying non-linearly separable clusters.

Cons:

- Results are highly dependent on the choice of distance measure.
- Struggles with datasets containing clusters of varying densities, as a single set of parameters may not accommodate all situations effectively.

DBSCAN

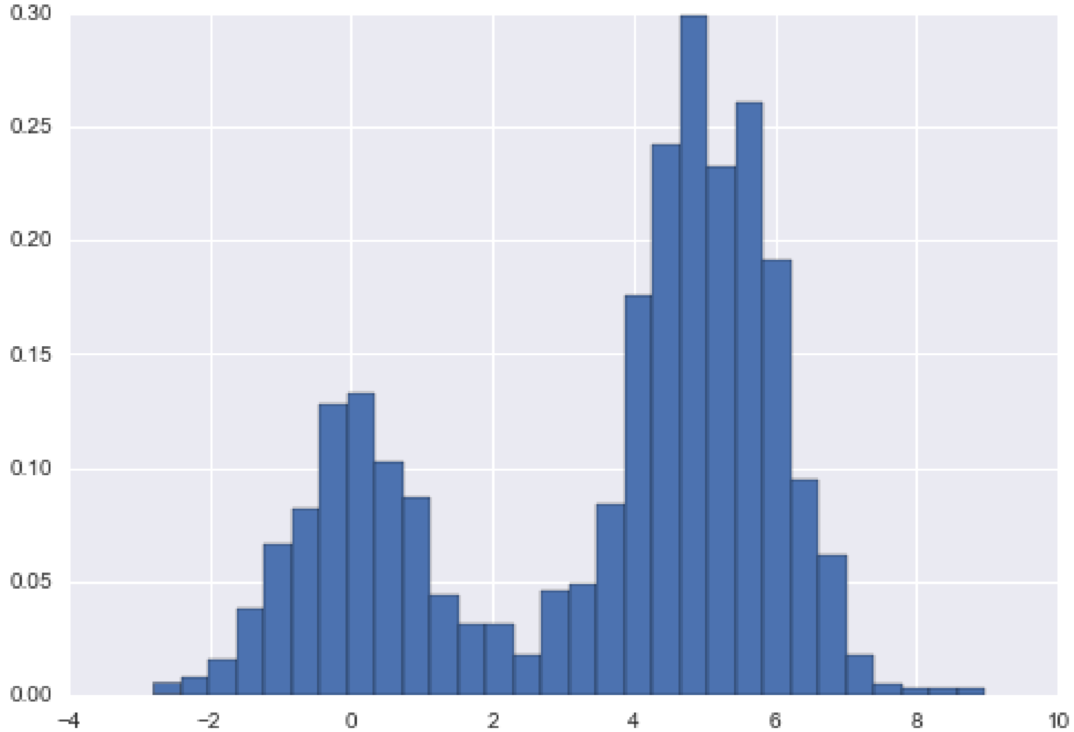
Arribas-Bel *et al*. 2021. Building(s and) cities: Delineating urban areas with a machine learning algorithm. *Journal of Urban Economics* 125: 103217



Kernel density estimation

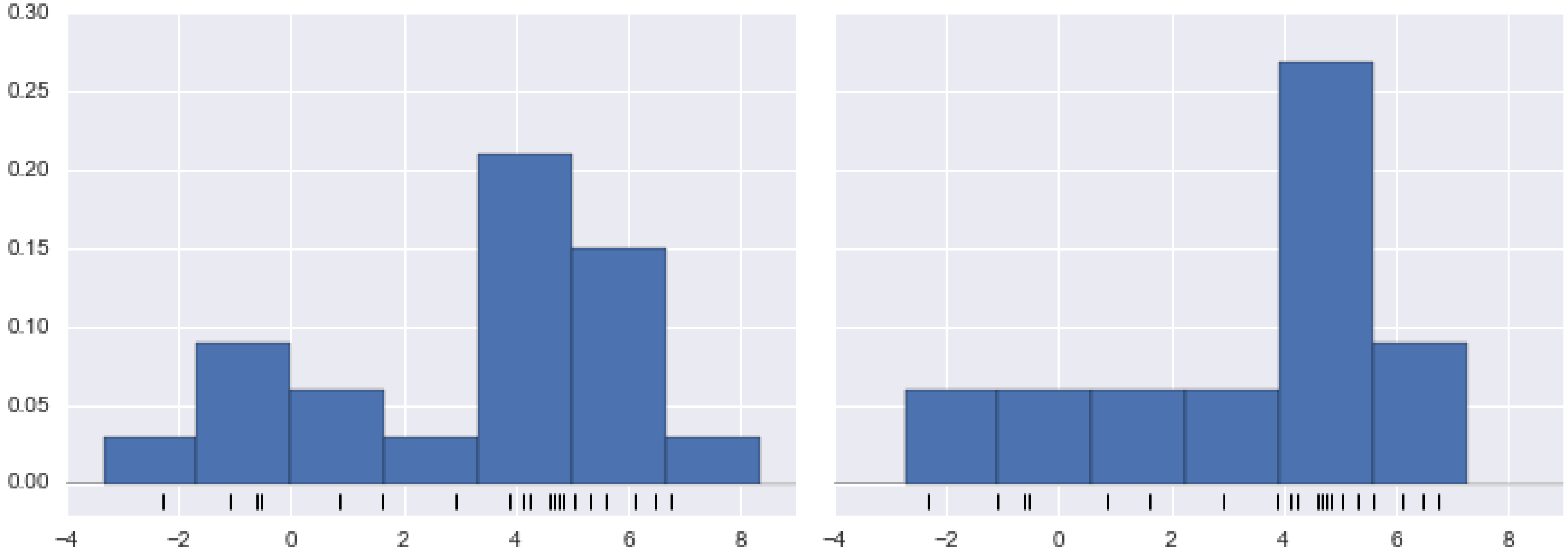
- Heatmap: Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable.
- It uses local information defined by windows (called kernels) to estimate densities of specified features at given locations.
- In essence it is a smoothing function where a continuous curve is created, based on a finite data sample.

Kernel density estimation



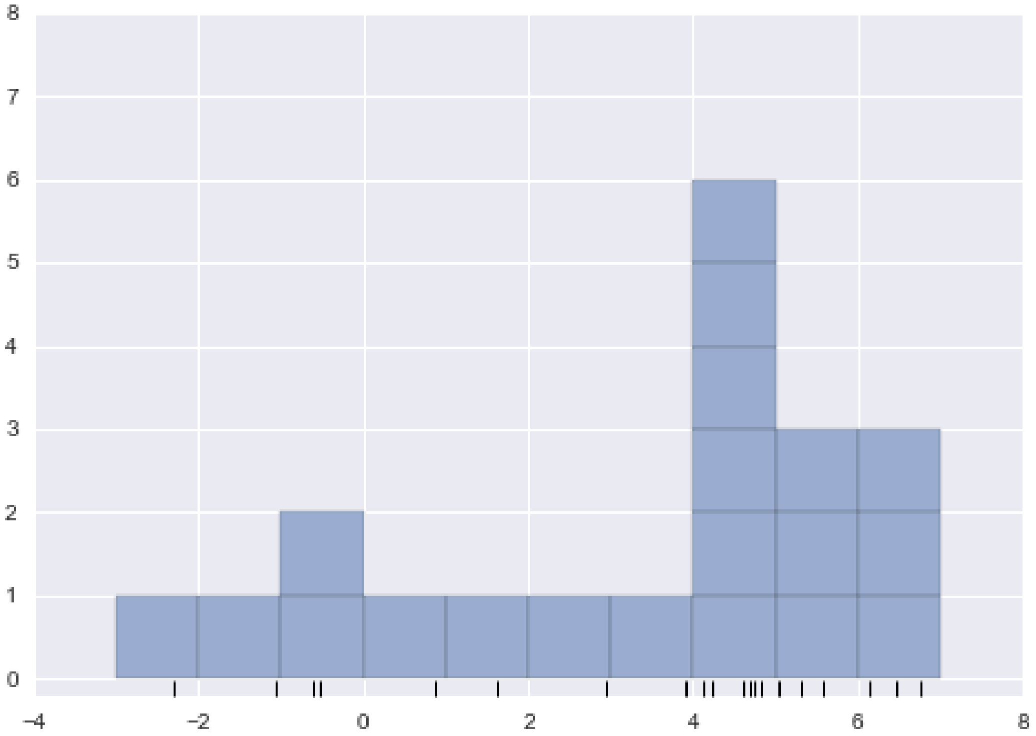
VanderPlas, J. 2016. *Python data science handbook: essential tools for working with data*. O'Reilly Media, Inc.

Kernel density estimation



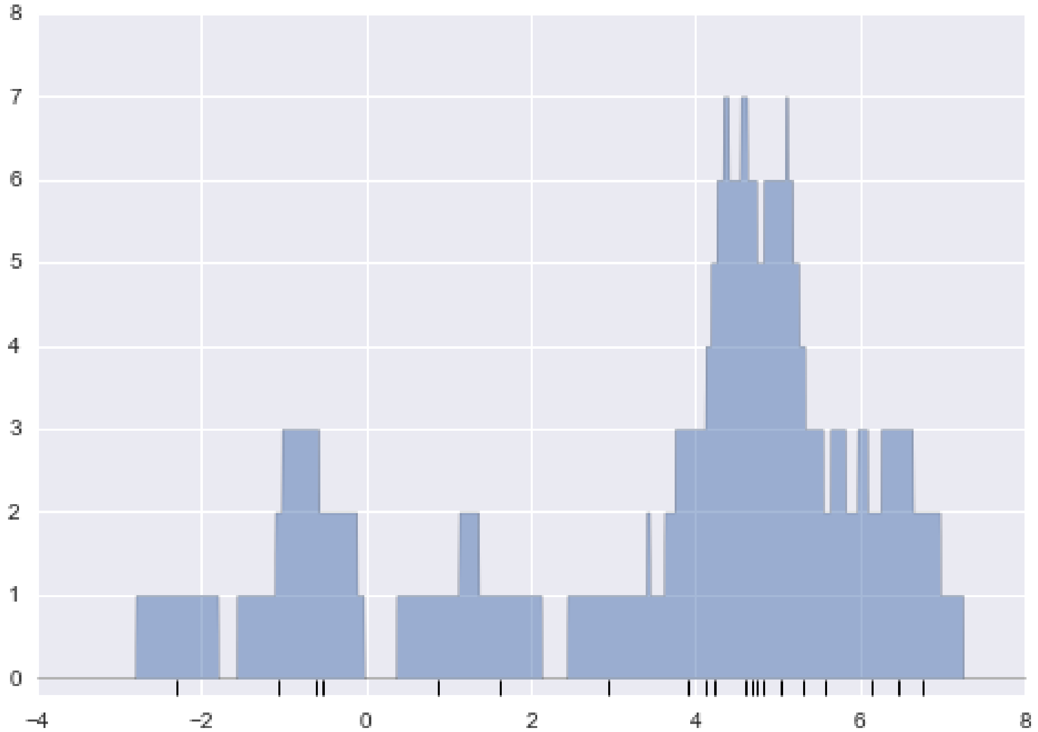
VanderPlas, J. 2016. *Python data science handbook: essential tools for working with data*. O'Reilly Media, Inc.

Kernel density estimation



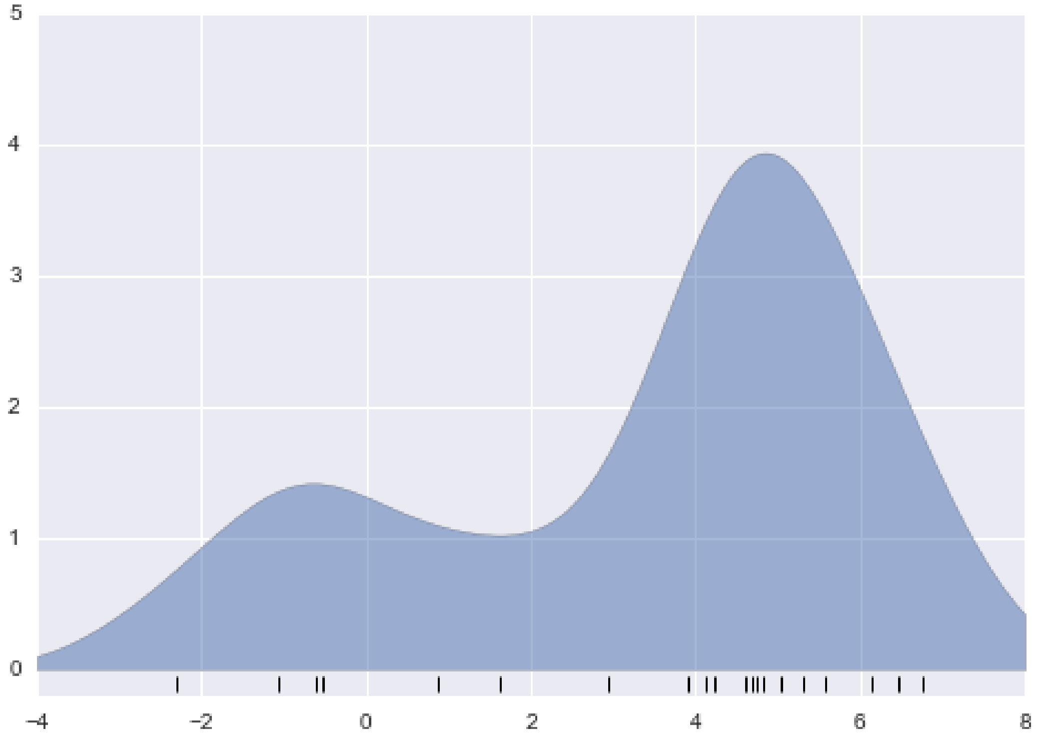
VanderPlas, J. 2016. *Python data science handbook: essential tools for working with data*. O'Reilly Media, Inc.

Kernel density estimation



VanderPlas, J. 2016. *Python data science handbook: essential tools for working with data*. O'Reilly Media, Inc.

Kernel density estimation



VanderPlas, J. 2016. *Python data science handbook: essential tools for working with data.* O'Reilly Media, Inc.

Kernel density estimation

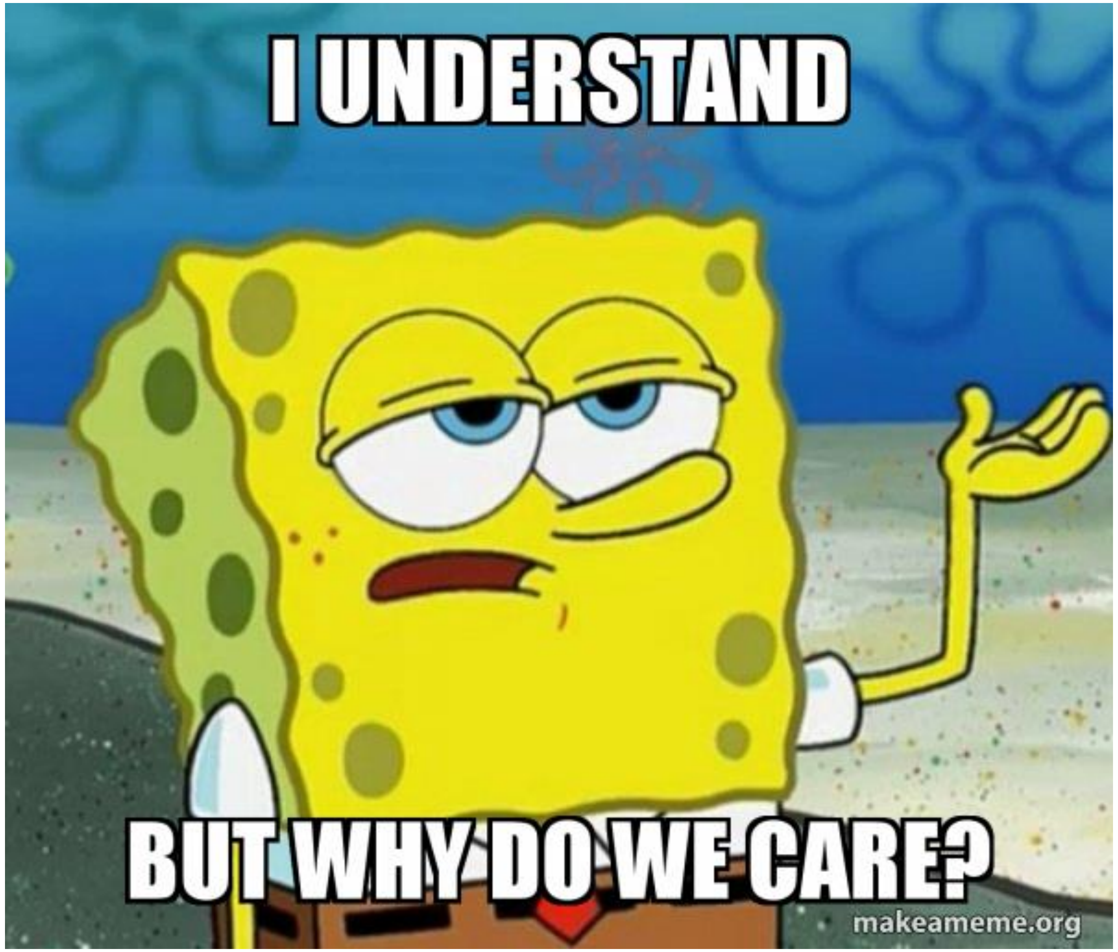
Pros:

- Produces a smooth surface that highlights hotspots without needing to define clusters beforehand.
- Simple to apply and widely used for visualising spatial concentration patterns.

Cons:

- Highly sensitive to the choice of bandwidth.
- Cannot handle areas with very different local densities well, because one global bandwidth may over-smooth or under-smooth (but adaptative bandwidths).

Kernel density estimation



Practical examples

Two examples

Point pattern analysis ‘in action’ in some actual research:

- Not all point data sets can be fully analysed with summary measures, e.g. analysis of GPS trajectory data (‘trajectory data’).
- Some problems do not per se look like a ‘point pattern analysis’ problem, but actually can be conceived as such, e.g. surname profiling.

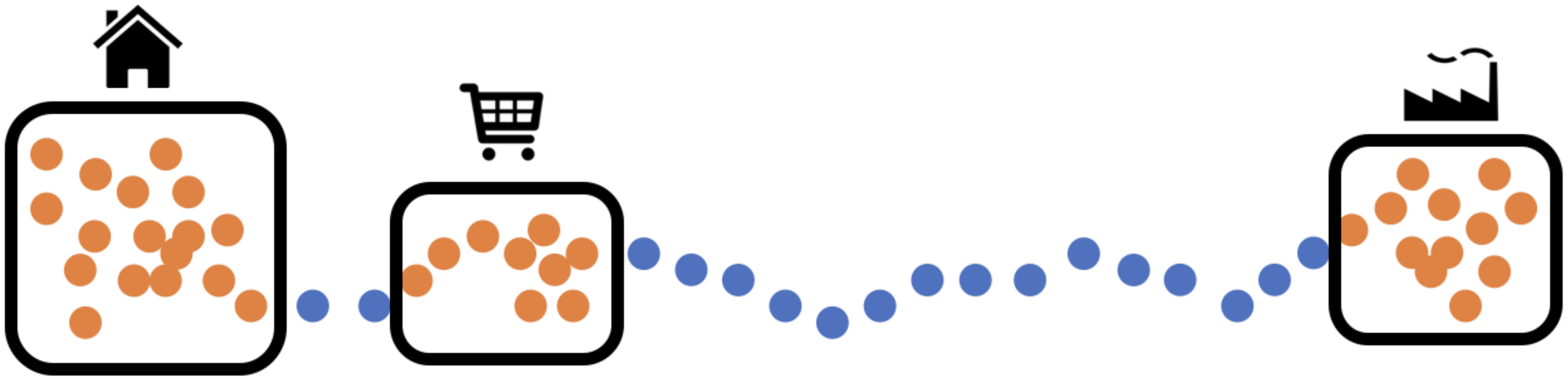
Trajectory data

Van Dijk, J. T. 2018. Identifying activity-travel points from GPS-data with multiple moving windows. *Computers, Environment and Urban System* 70: 84-101

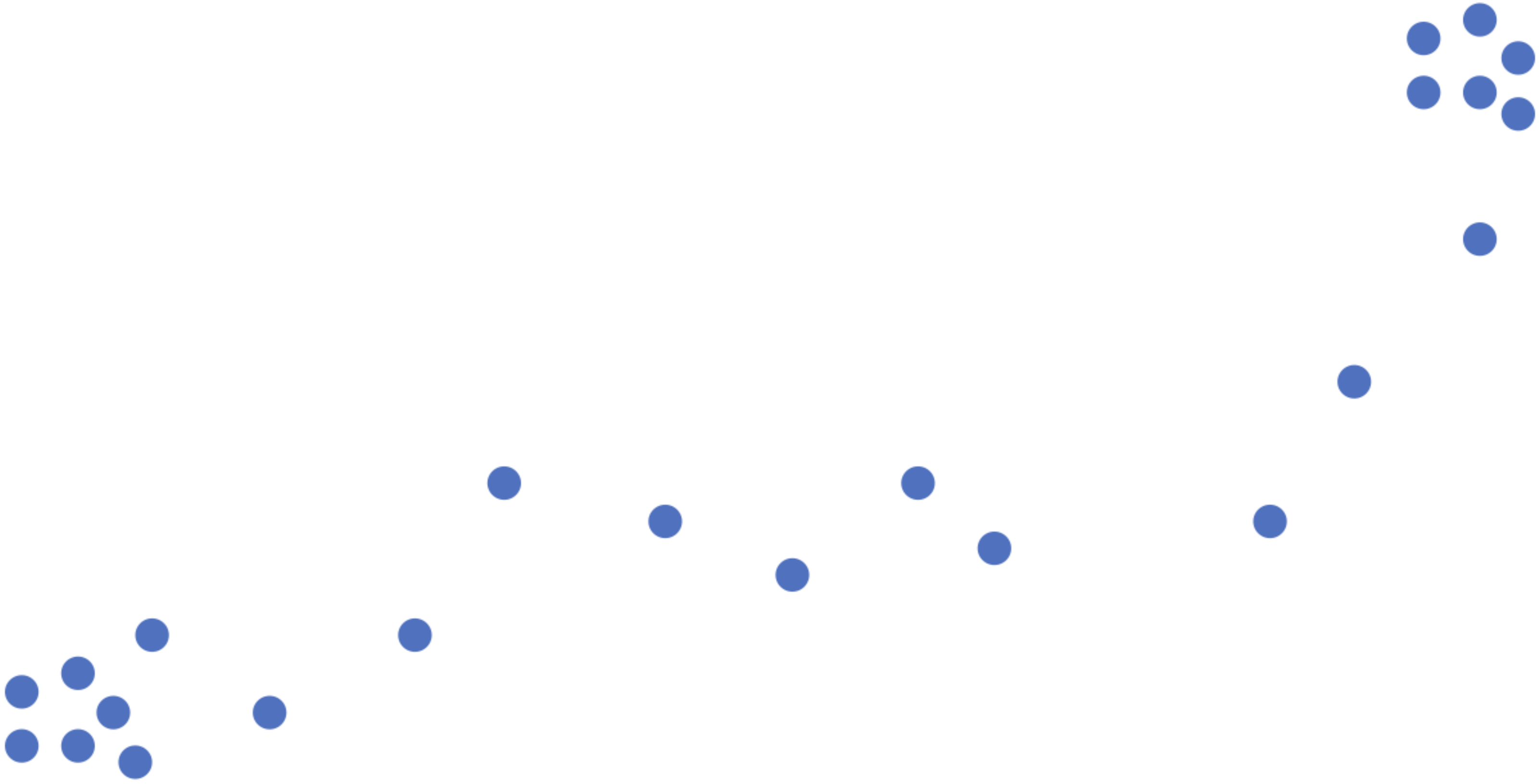
Trajectory data

Van Dijk, J. T. 2018. Identifying activity-travel points from GPS-data with multiple moving windows.
Computers, Environment and Urban System 70: 84-101

Trajectory data

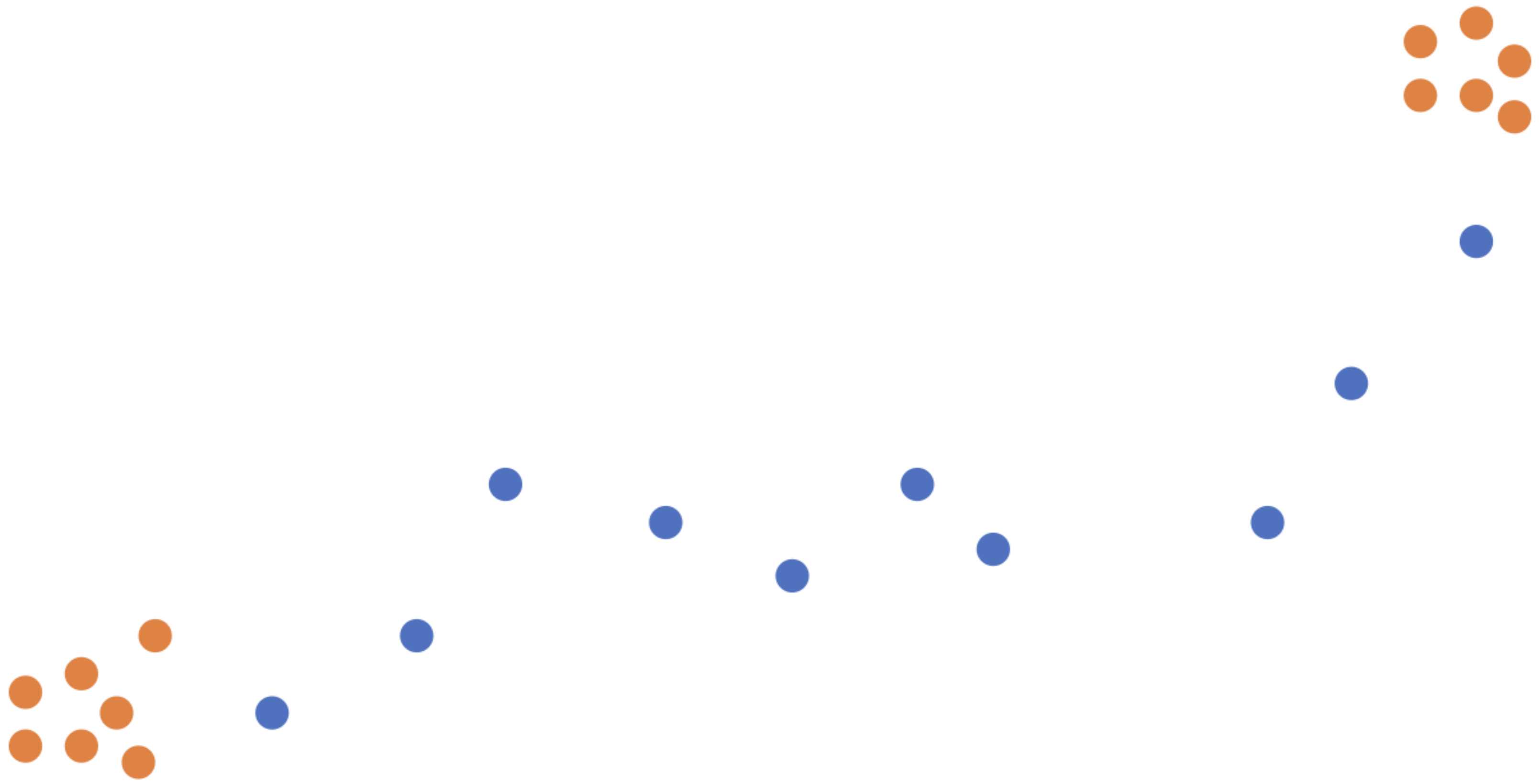


Trajectory data



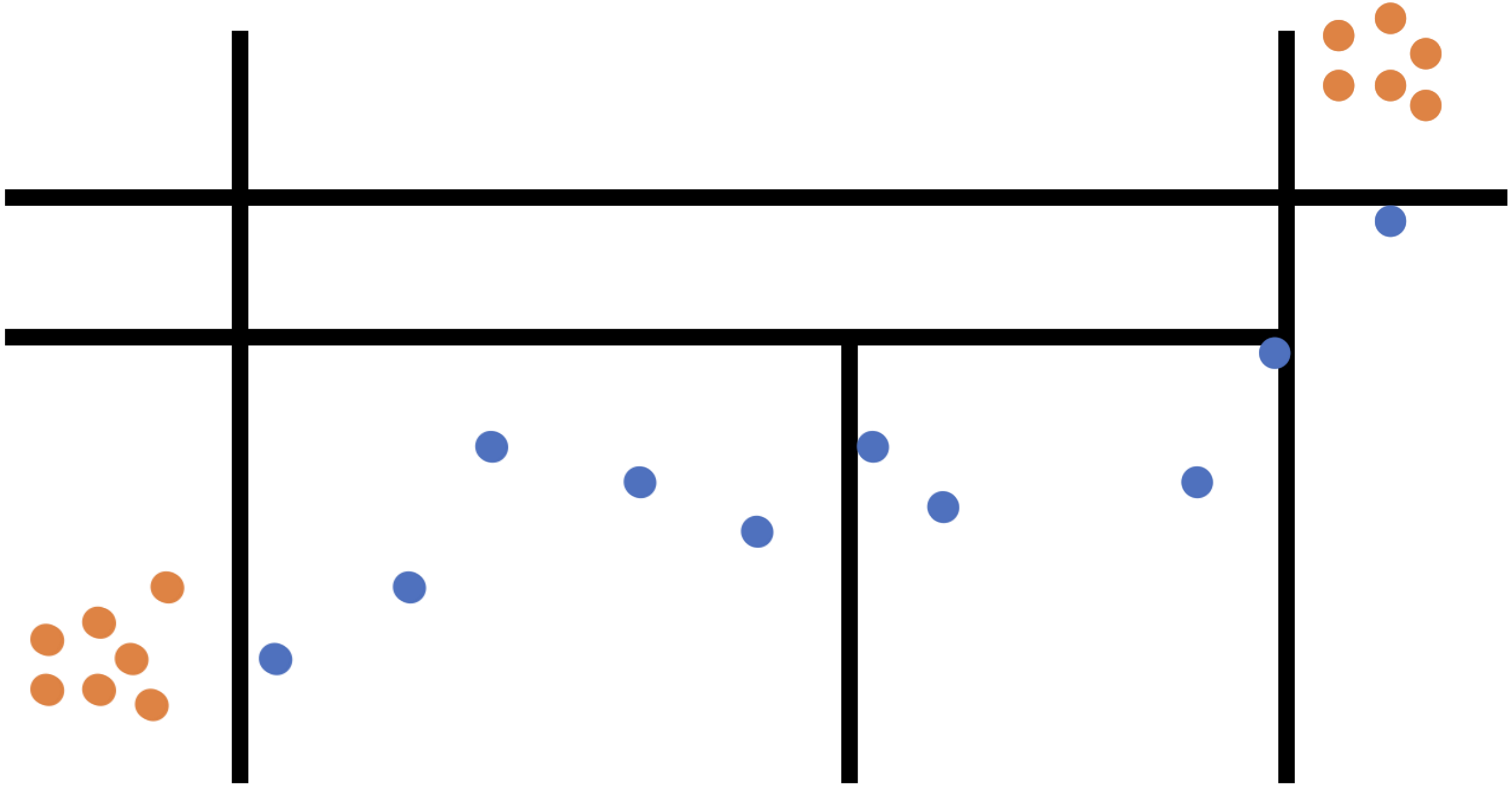
Van Dijk, J. T. & T. de Jong. 2017. Post-processing GPS tracks in reconstructing travelled routes in a GIS-environment: Network subset selection and attribute adjustment. *Annals of GIS* 23(3): 203-217

Trajectory data



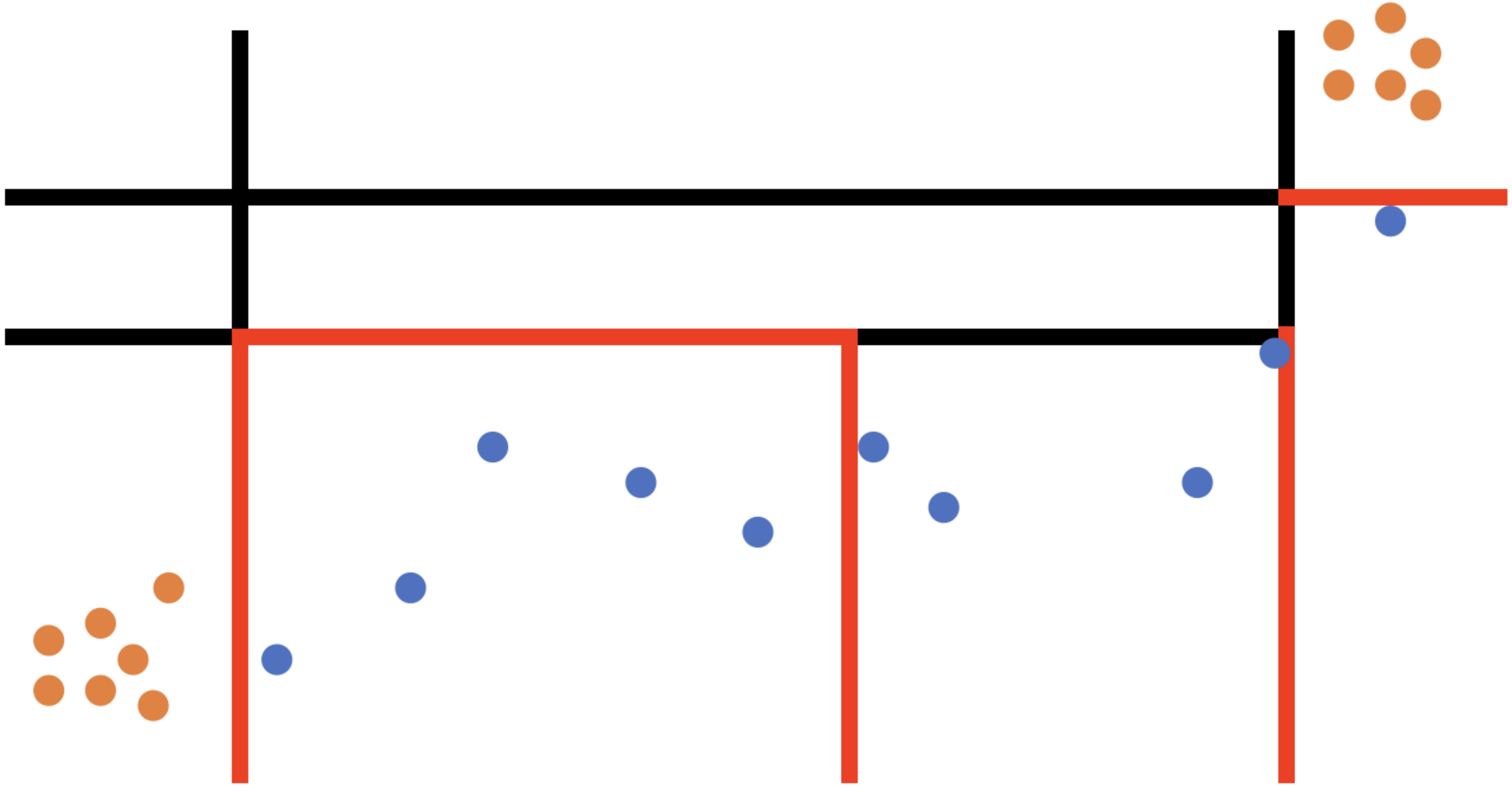
Van Dijk, J. T. & T. de Jong. 2017. Post-processing GPS tracks in reconstructing travelled routes in a GIS-environment: Network subset selection and attribute adjustment. *Annals of GIS* 23(3): 203-217

Trajectory data



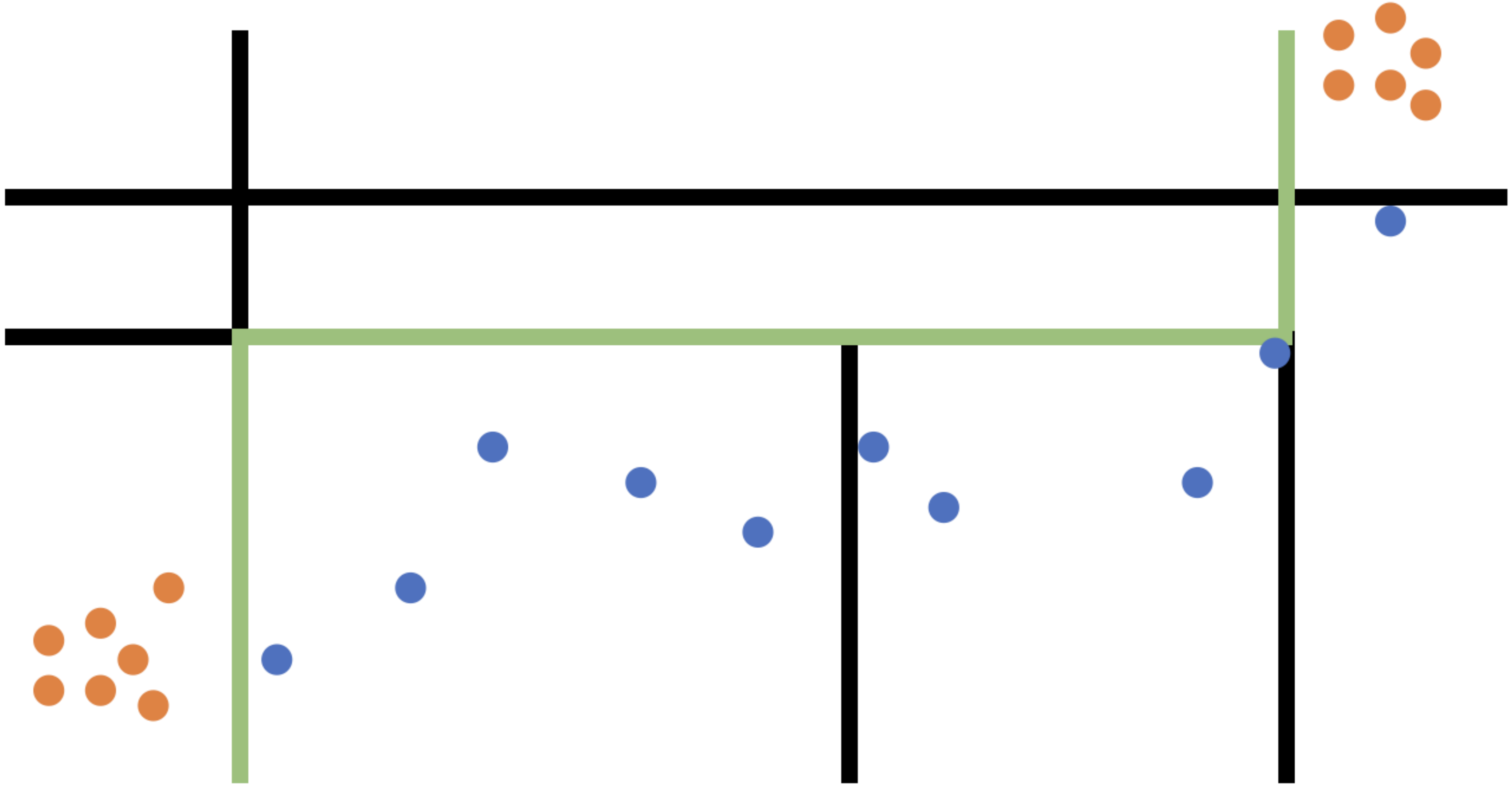
Van Dijk, J. T. & T. de Jong. 2017. Post-processing GPS tracks in reconstructing travelled routes in a GIS-environment: Network subset selection and attribute adjustment. *Annals of GIS* 23(3): 203-217

Trajectory data



Van Dijk, J. T. & T. de Jong. 2017. Post-processing GPS tracks in reconstructing travelled routes in a GIS-environment: Network subset selection and attribute adjustment. *Annals of GIS* 23(3): 203-217

Trajectory data



Van Dijk, J. T. & T. de Jong. 2017. Post-processing GPS tracks in reconstructing travelled routes in a GIS-environment: Network subset selection and attribute adjustment. *Annals of GIS* 23(3): 203-217

Surname profiling

- Personal names contain socio-demographic information.
- But also: many surnames contain spatial information at a variety of scales relating to the origins of many of their bearers.
- Data: Historic Census of Population 1851-1911, Consumer Registers 1997– 2016.
- Total of 1.2 million surnames with locations (Historic Parishes, unit postcodes and geo-coded addresses for several years of data).

“Van Dijk”

“Lansley”

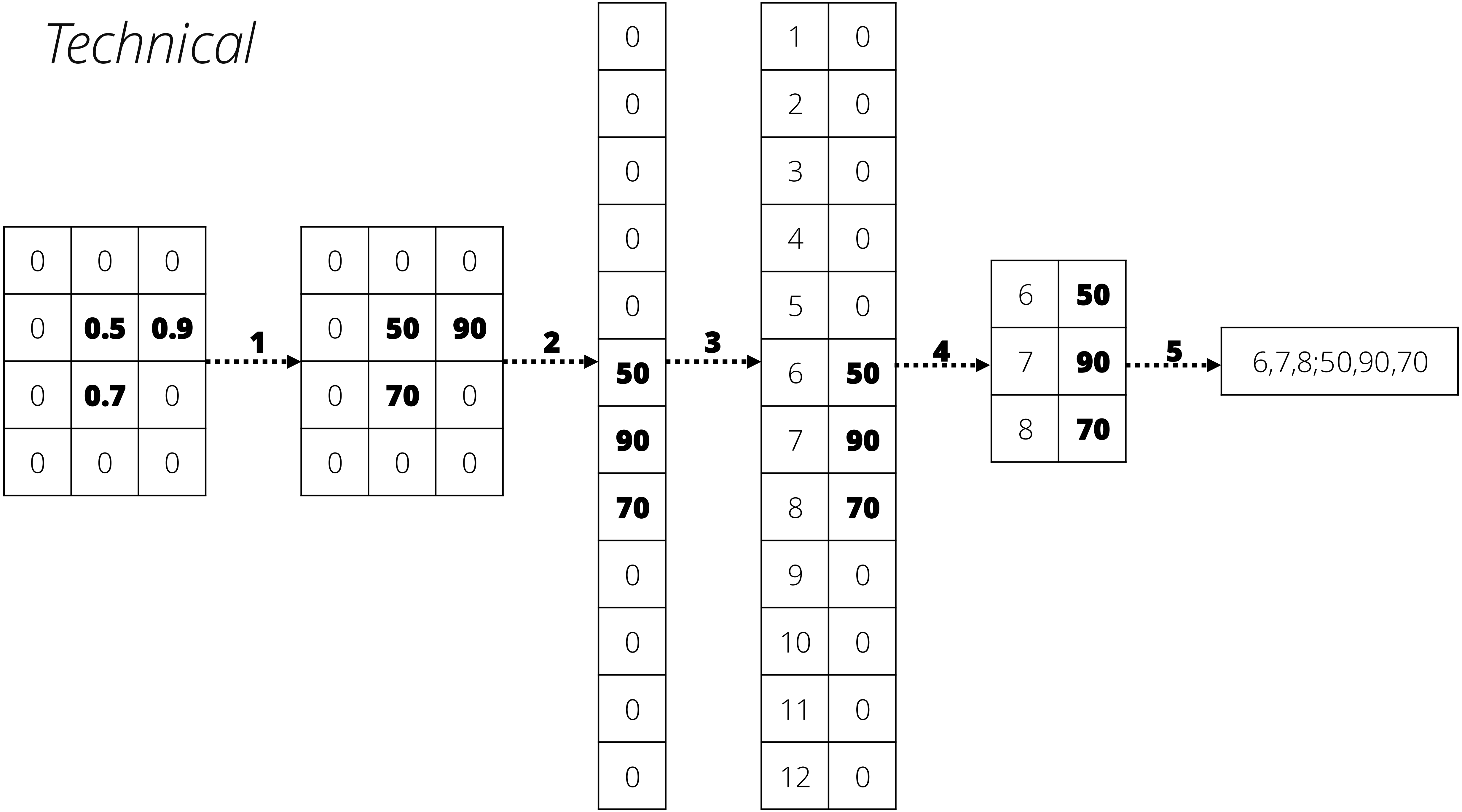
“Rossall”

Surname profiling

Combination on various point pattern analysis techniques:

- kernel densities to map the surname concentrations of names found in Great Britain executed over a 1000m x 1000m grid.
- deconstruction of grids as sparse matrices to optimise storage and database retrieval (storing 1.2 million KDEs is challenging), followed by DBSCAN to create contours of highest relative density (vectorisation).

Technical



0 0 0
0 0.5 0.9
0 0.7 0
0 0 0
1
0 0 0
0 50 90
0 70 0
0 0 0
2
0 0 0 0 0 50 90 70 0 0 0 0
3
1 0
2 0
3 0
4 0
5 0
6 50
7 90
8 70
9 0
10 0
11 0
12 0
4
6 50
7 90
8 70
5
6,7,8;50,90,70

Technical

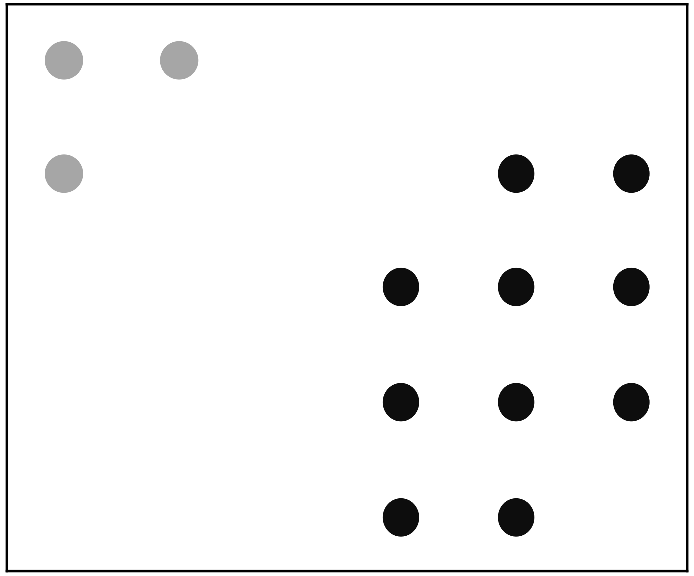
1

50	**90**	20	0	10	10
80	10	0	20	**40**	**50**
30	0	10	**40**	**60**	**80**
0	0	20	**50**	**70**	**50**
0	0	30	**50**	**90**	30

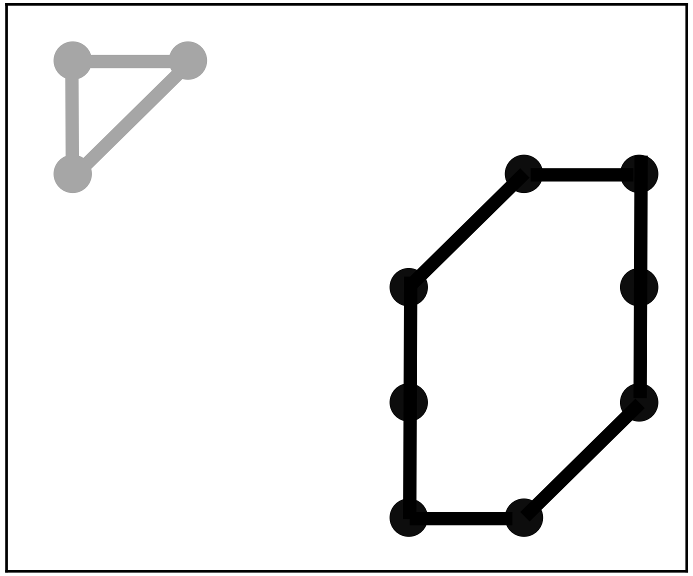
2

50	**90**				
80				**40**	**50**
			40	**60**	**80**
			50	**70**	**50**
			50	**90**	

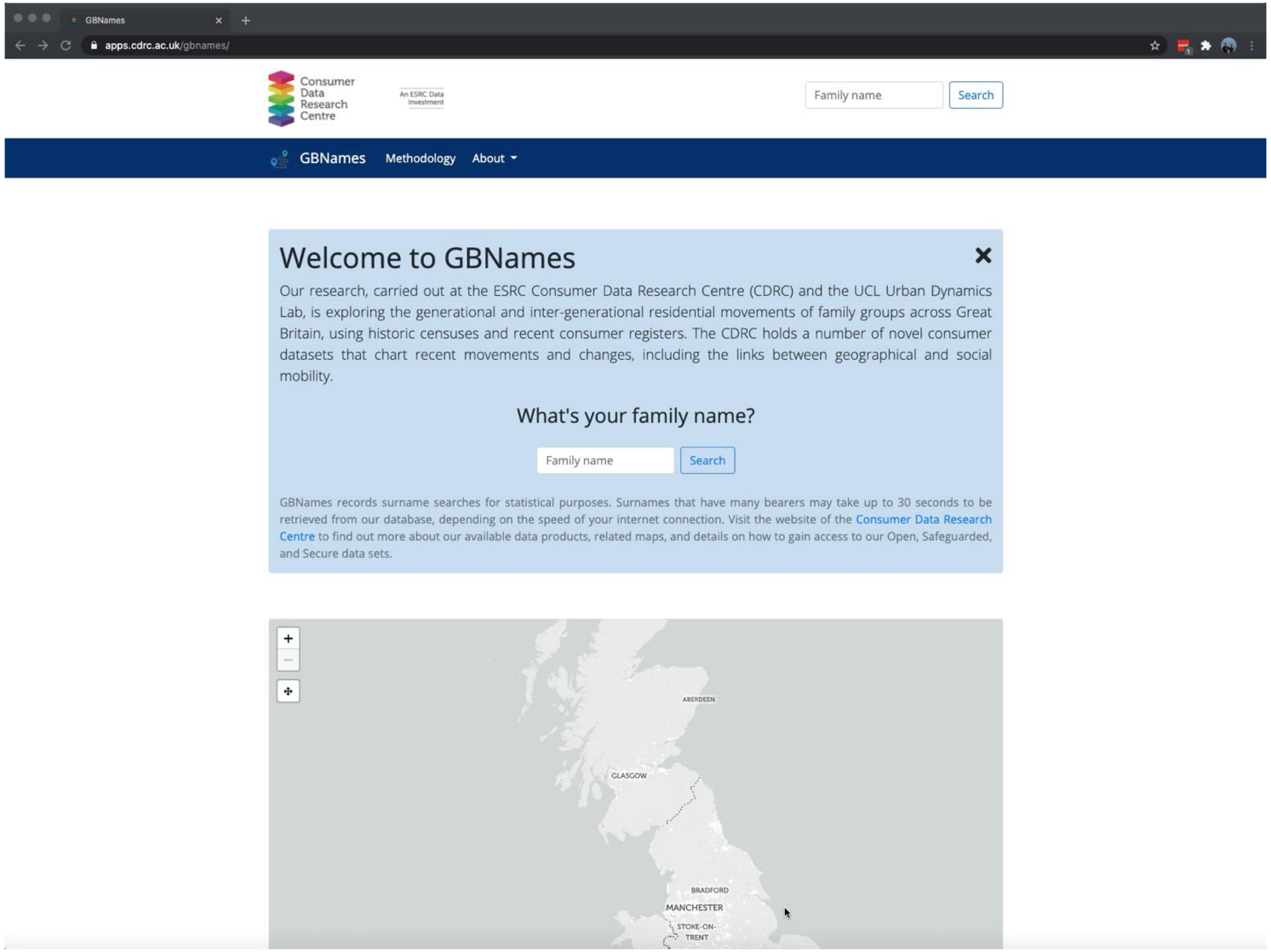
3



4



Surname profiling



Van Dijk, J. T. & P. A. Longley. 2020. Interactive display of surname distributions in historic and contemporary Great Britain. *Journal of Maps* 16(1): 68-76

Conclusion

- Point pattern analysis focuses on individual events in space, rather than aggregated counts in administrative units.
- Different methods can help characterise point processes and identify spatial clusters, including established methods such as DBSCAN and Kernel Density Estimation.
- Some datasets may not immediately look suited to point pattern analysis.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk