Geographic Data Science

Point Patterns

Elisabetta Pietrostefani & Carmen Cabrera-Arnau

The *point* of points

Points like polygons

- Points *can* represent “fixed” entities
- In this case, points are qualitatively similar to polygons/lines
- The goal here is, taking location fixed, to model other aspects of the data

Points like polygons

Examples: - Cities (in most cases) - Buildings - Polygons represented as their centroid - …

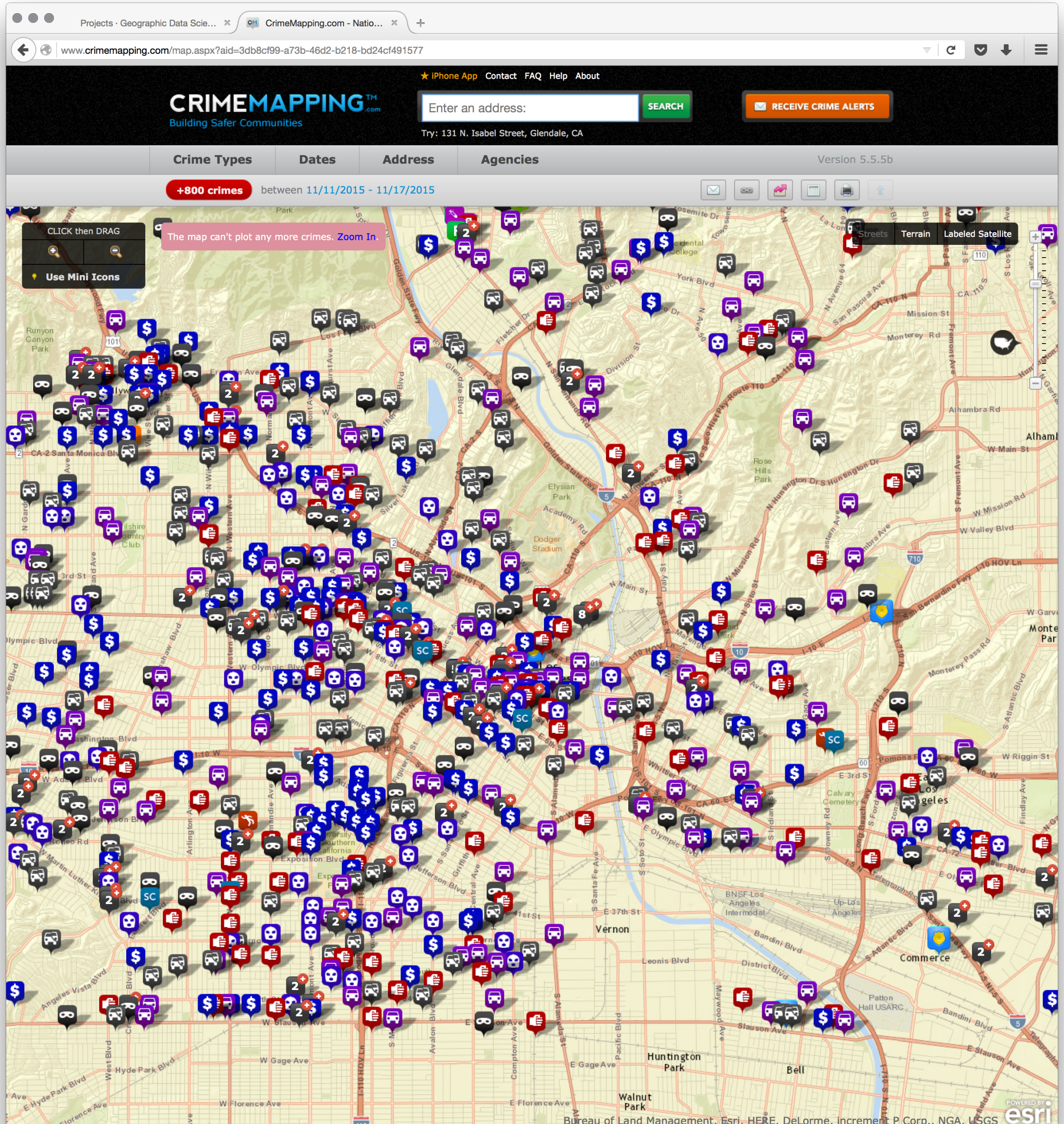
When points are not polygons

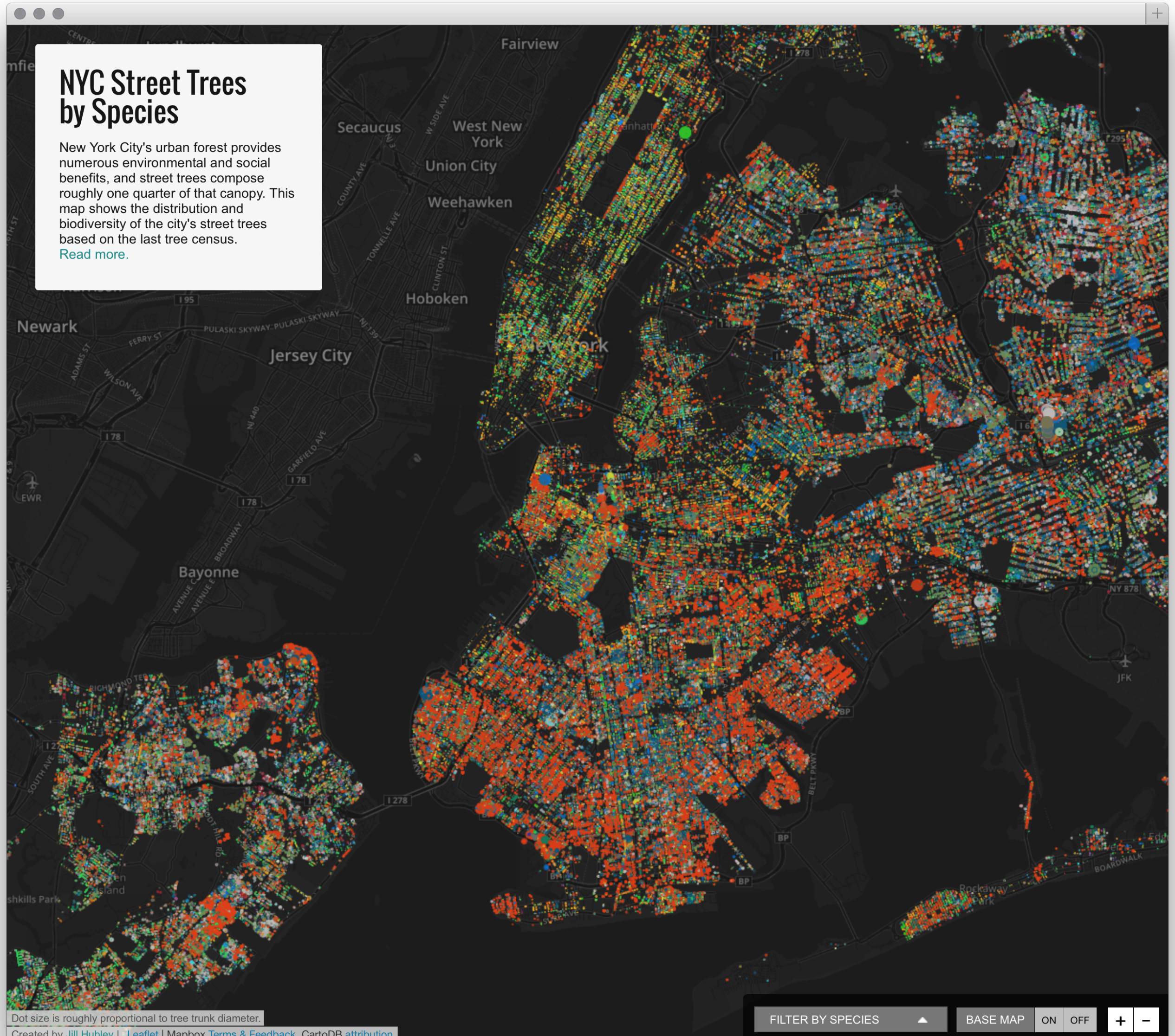
Point data are not only a different geometry than polygons or lines…

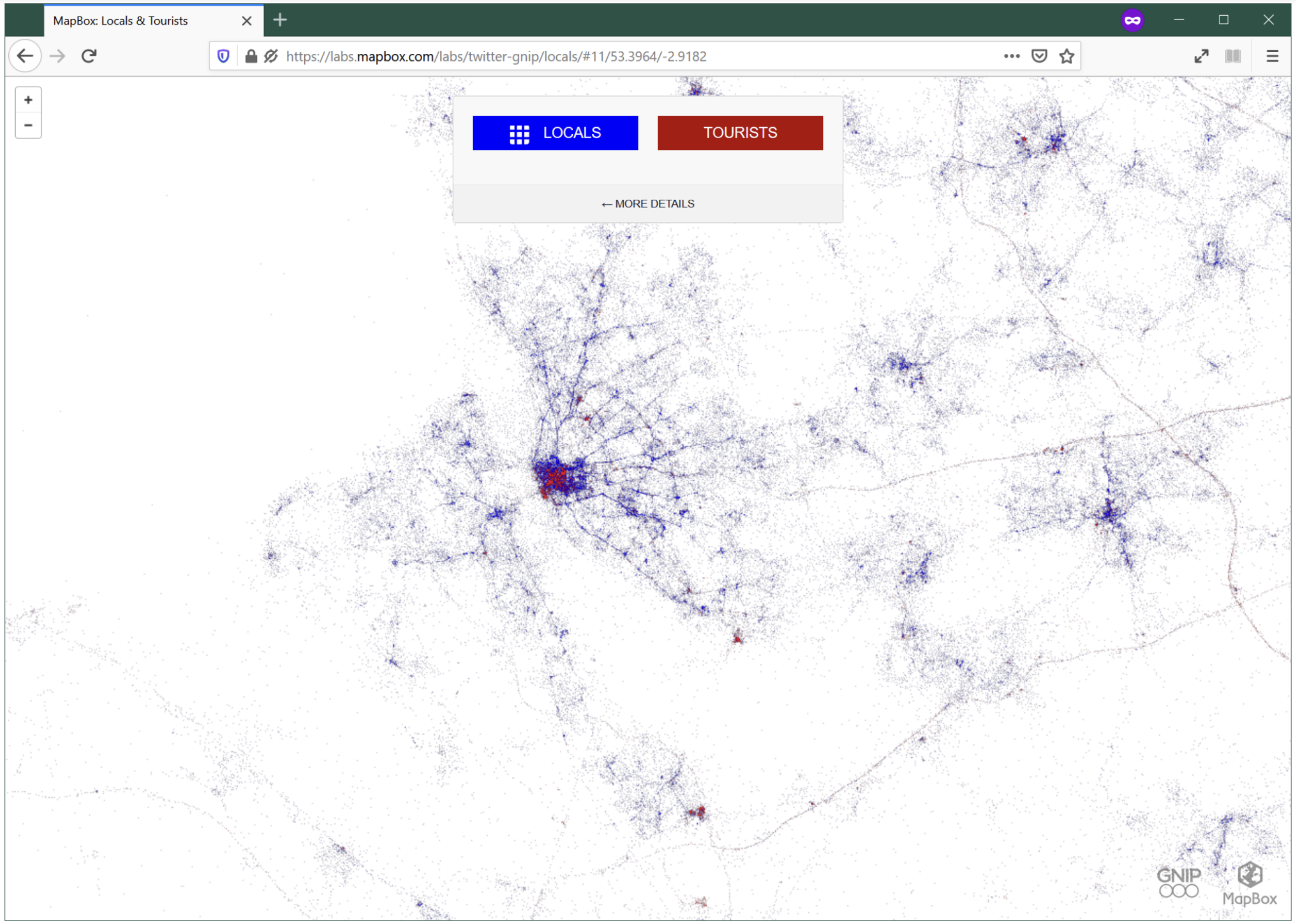
… Points can also represent a fundamentally different way to approach spatial analysis

Points unlike polygons

A few examples


Projects · Geographic Data Scie...
CrimeMapping.com - Natio...
www.crimemapping.com/map.aspx?aid=3db8cf99-a73b-46d2-b218-bd24cf491577
iPhone App
Contact
FAQ
Help
About
CRIMEMAPPING
Building Safer Communities
Enter an address:
SEARCH
RECEIVE CRIME ALERTS
Try: 131 N. Isabel Street, Glendale, CA
Crime Types
Dates
Address
Agencies
Version 5.5.5b
+800 crimes
between 11/11/2015 - 11/17/2015
CLICK then DRAG
Use Mini Icons
The map can't plot any more crimes. Zoom In
Streets
Terrain
Labeled Satellite
Vernon
Commerce
Huntington Park
Bell
Walnut Park
Bureau of Land Management, Esri, HERE, DeLorme, increment P Corp., NGA, USGS
esri


NYC Street Trees by Species
New York City's urban forest provides numerous environmental and social benefits, and street trees compose roughly one quarter of that canopy. This map shows the distribution and biodiversity of the city's street trees based on the last tree census.
Read more.
Fairview
Secaucus
West New York
Union City
Weehawken
Hoboken
Newark
Jersey City
New York
Bayonne
EWR
JFK
Dot size is roughly proportional to tree trunk diameter.
Created by Jill Hubley | Leaflet | Mapbox Terms & Feedback, CartoDB attribution
FILTER BY SPECIES
BASE MAP
ON
OFF
+
−


MapBox: Locals & Tourists
https://labs.mapbox.com/labs/twitter-gnip/locals/#11/53.3964/-2.9182
LOCALS
TOURISTS
← MORE DETAILS
GNIP
MapBox

Points patterns

Points patterns

Distribution of **points** over a portion of **space** Assumption is a point can happen anywhere on that space, but only happens in specific locations

- **Unmarked**: locations only
- **Marked**: values attached to each point

Point Pattern Analysis

Describe, characterize, and explain point patterns, focusing on their **generating process**

- Visual exploration
- Clustering properties and clusters
- Statistical modeling of the underlying processes

Visualization of Point Patterns

Visualization of PPs

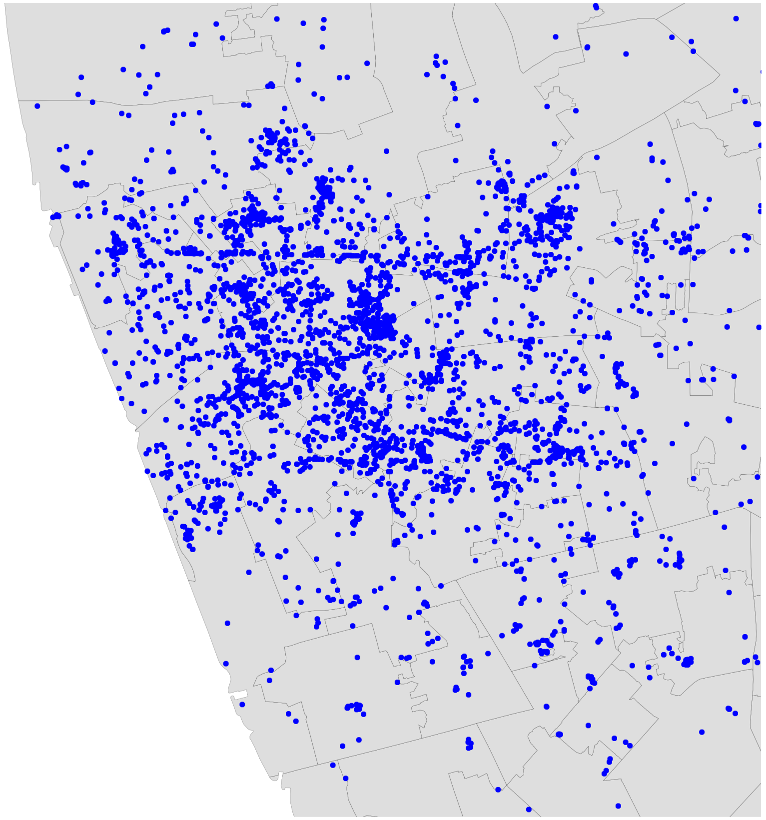
Four routes (today):

- One-to-one mapping – “Scatter plot”
- Aggregate – “Histogram”
- Smooth – KDE
- Smooth – Interpolation

One-to-one

- Intuitive
- Effective in small datasets
- Limited as size increases until useless

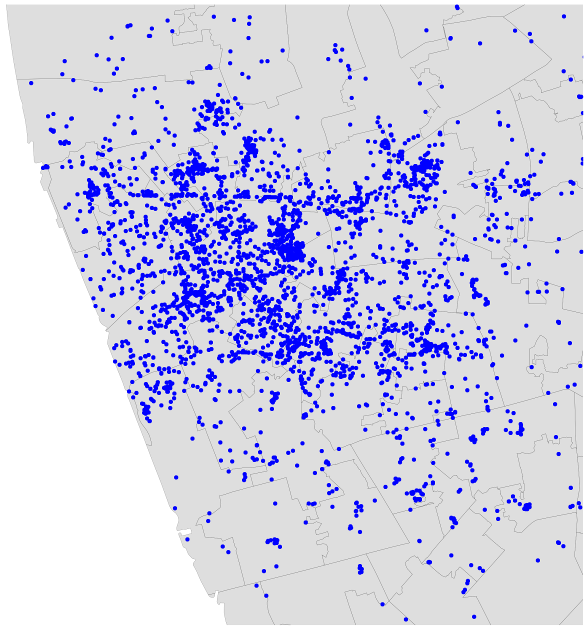
One-to-one

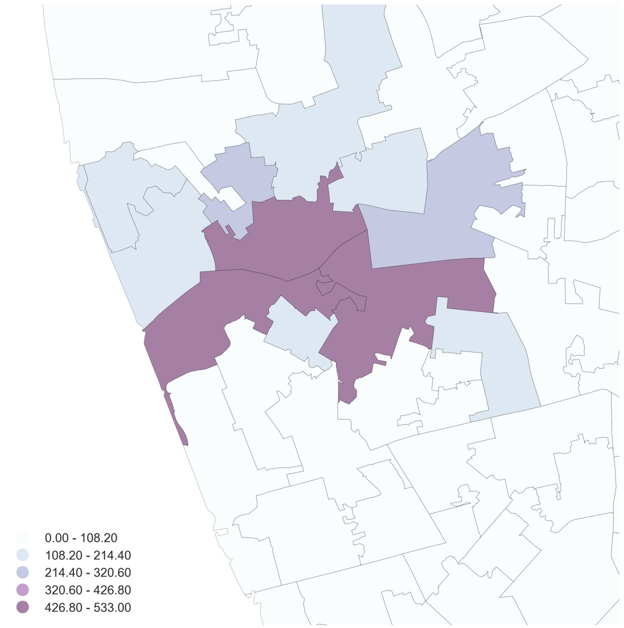


Aggregation

Points meet polygons

- Use polygon boundaries and count points per area [Insert your skills for choropleth mapping here!!!]
- But, the polygons need to *“make sense”* (their delineation needs to relate to the point generating process)



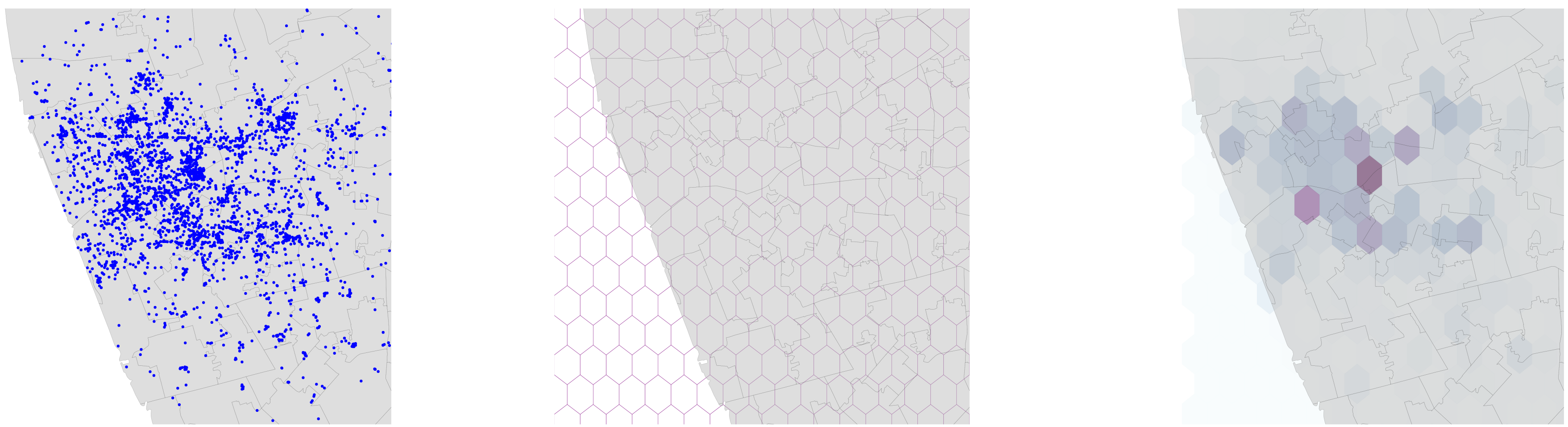

0.00 - 108.20
108.20 - 214.40
214.40 - 320.60
320.60 - 426.80
426.80 - 533.00

Hex-binning

If no polygon boundary seems like a good candidate for aggregation… …draw a hexagonal (or squared) tesselation!!!

Hexagons…

- Are regular
- Exhaust the space (Unlike circles)
- Have many sides (minimize boundary problems)



But

- (Arbitrary) aggregation may induce MAUP
- Points usually represent events that affect only part of the population and hence are best considered as rates

Kernet Density Estimation (KDE)

KDE

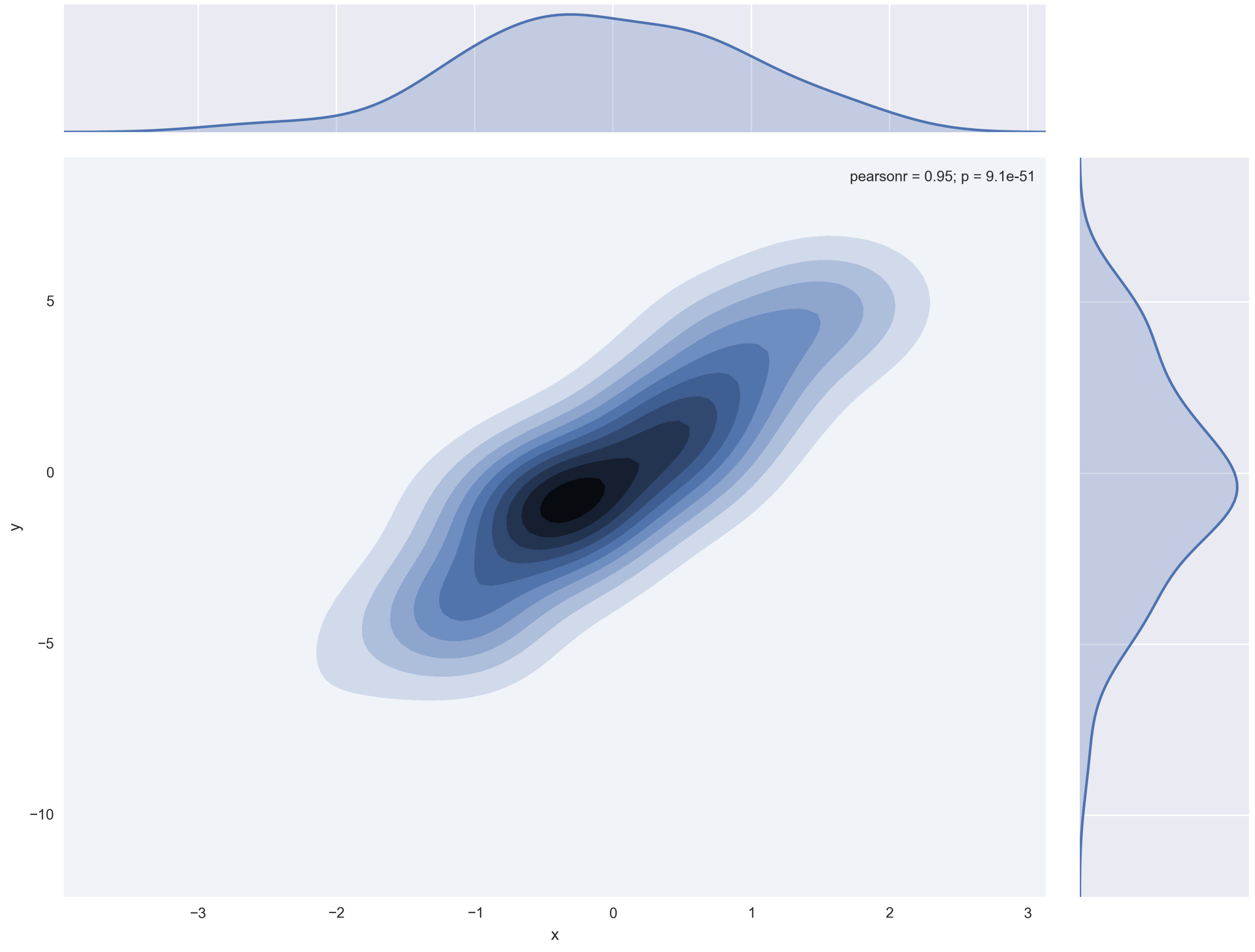
Estimate the (**continuous**) observed distribution of a variable

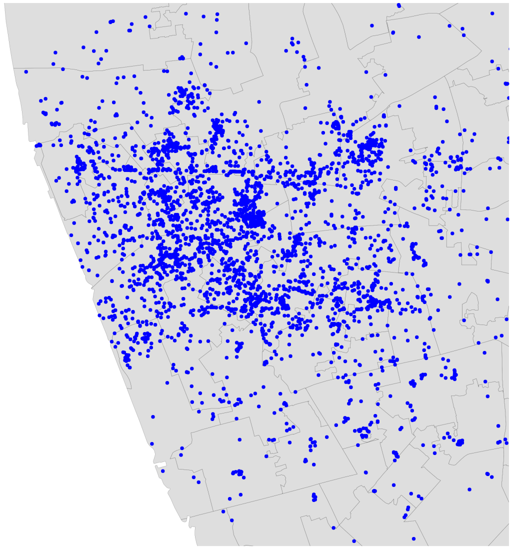
- Probability of finding an observation at a given point
- “Continuous histogram”
- Solves (much of) the MAUP problem, but not the underlying population issue

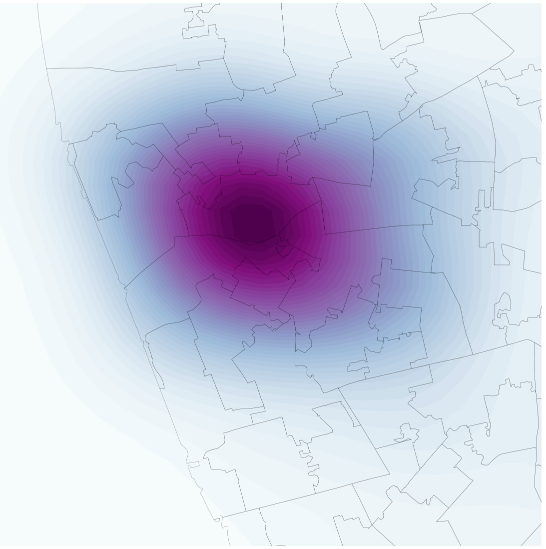
Bivariate (spatial) KDE

Probability of finding observations at a given point in space

- **Bivariate** version: distribution of pairs of values
- In **space**: values are coordinates (XY), locations
- Continuous “version” of a choropleth


pearsonr = 0.95; p = 9.1e-51
5
0
−5
−10
y
−3
−2
−1
0
1
2
3
x





Interpolation

- Estimating values spatially continuous variables for spatial locations where they **have not** been observed, based on observations.
- **Geostatistics**, is concerned with the modelling, prediction and simulation of spatially continuous phenomena.

Inverse Distance Weighting (IDW)

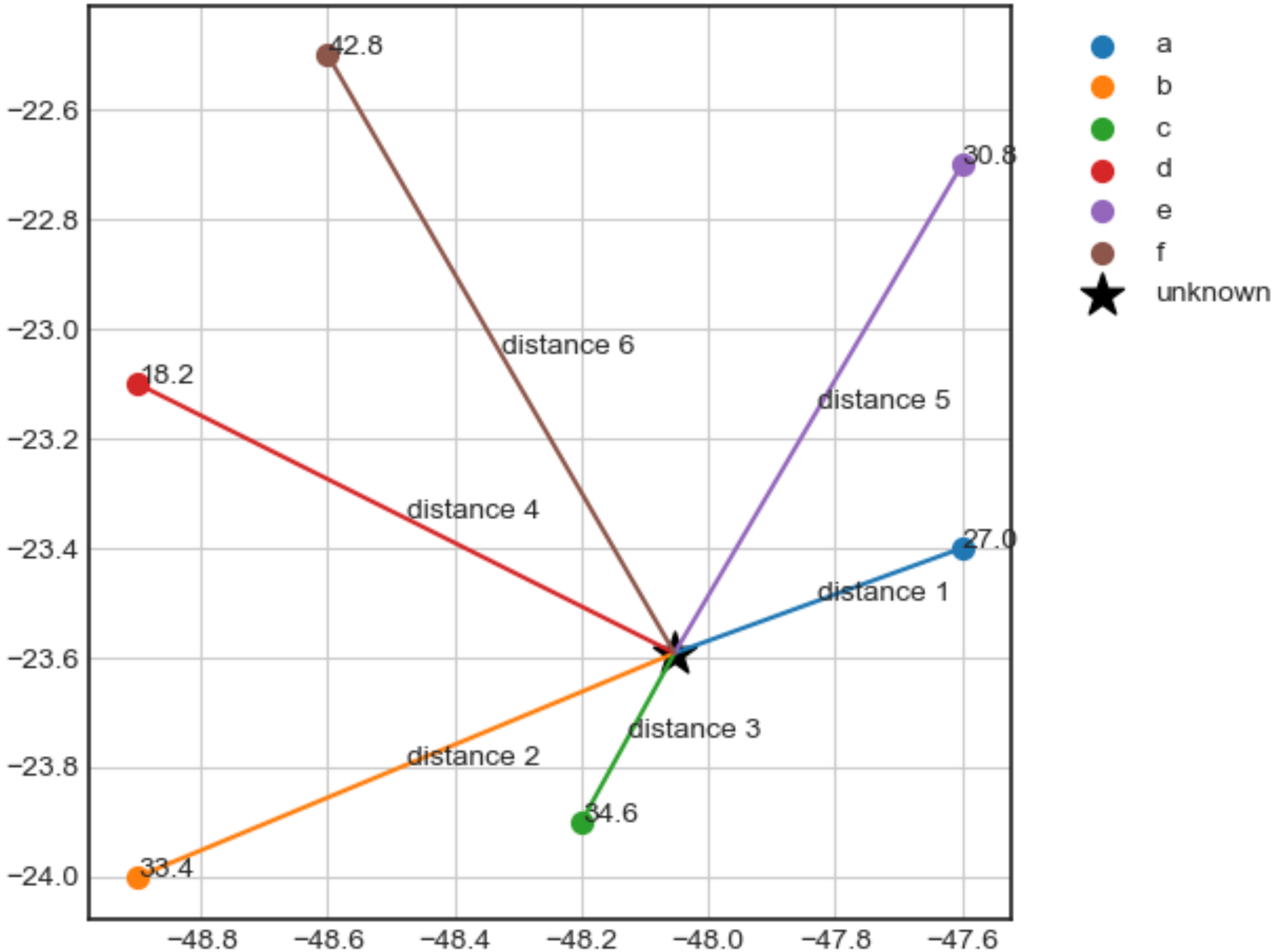
- We observe a property of a phenomenon $Z(s)$ at a **limited** number of sample locations, and are interested in the property value at **all** locations.
- Have to predict it for unobserved locations.

Kriging

If we were predicting prices

$$\text{Price}_i = \sum_{j=1}^{N} w_j * \text{Price}_j + \epsilon_i$$

- with $w_j = (\frac{1}{d_{ij}})^2$ for all i and $j \neq i$
- d the distance between i and j.

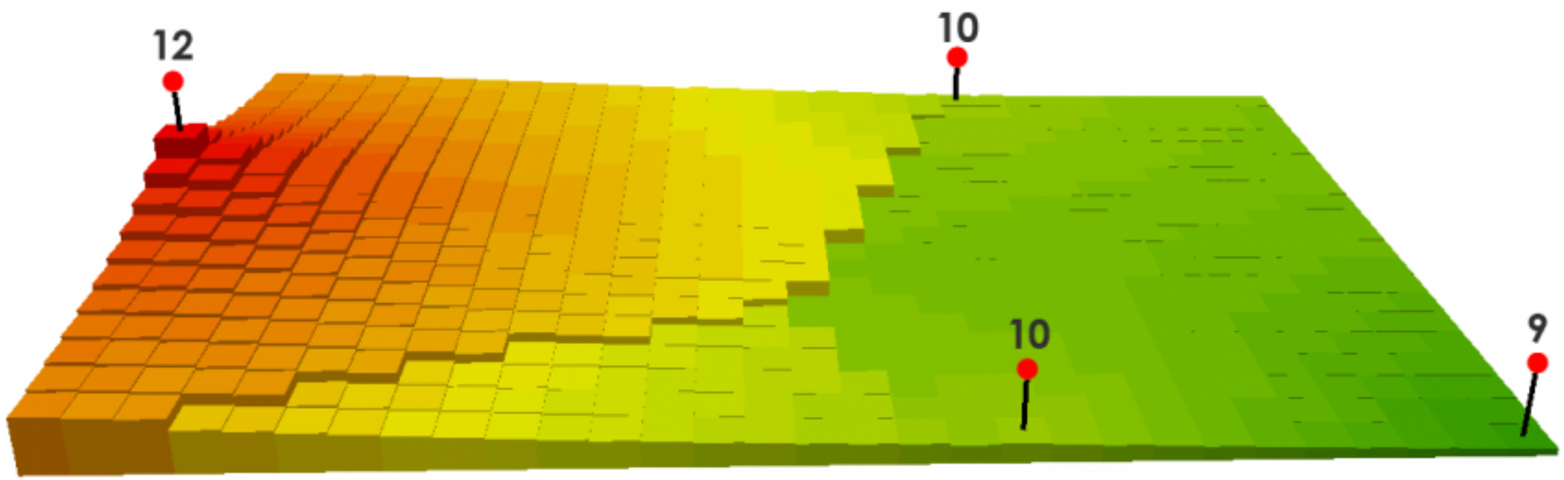

42.8
30.8
18.2
27.0
34.6
33.4
distance 1
distance 2
distance 3
distance 4
distance 5
distance 6
−22.6
−22.8
−23.0
−23.2
−23.4
−23.6
−23.8
−24.0
−48.8
−48.6
−48.4
−48.2
−48.0
−47.8
−47.6
a
b
c
d
e
f
unknown

Parametres

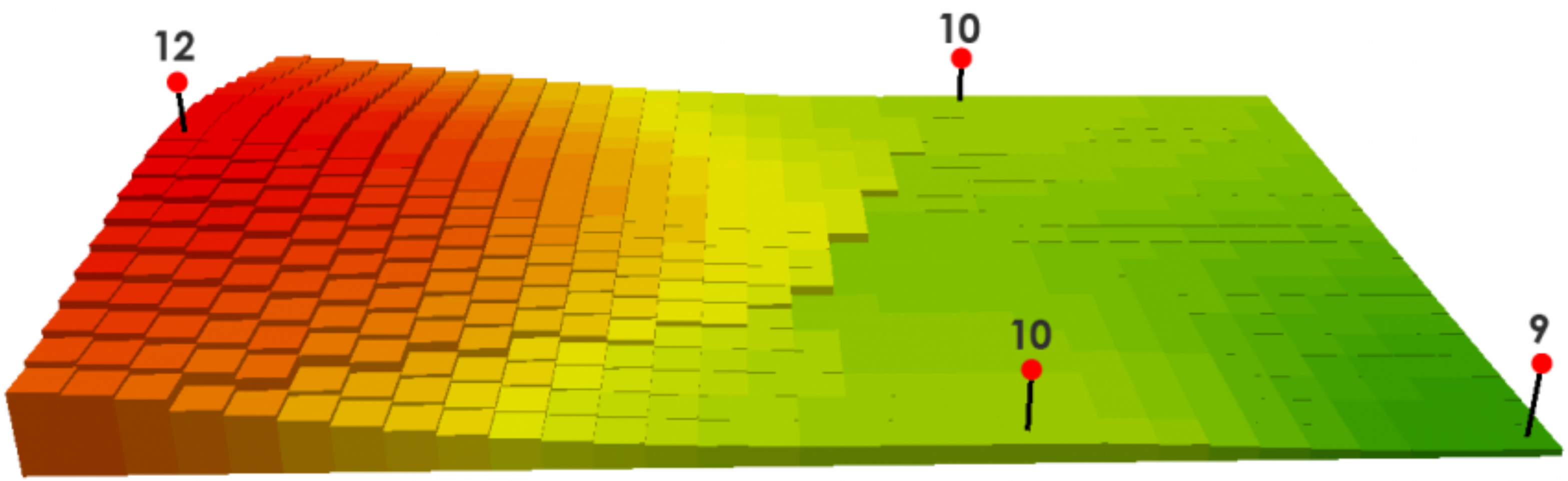
- **Variable**: for example price
- **Nearest Neighbours** : the number of nearest observations that should be used
- **idp** : set inverse distance power to 2

A super useful link here

Parametres



idp = 1



idp = 2

Density-Based Spatial Clustering of Applications with Noise, or DBSCAN

Questions