Geographic Data Science - Lecture VIII

Points

Dani Arribas-Bel

Today

- The point of points
- Point patterns
- Visualization of point patterns

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…

CRIMEMAPPING.com
Building Safer Communities

Enter an address: SEARCH

Try: 131 N. Isabel Street, Glendale, CA

RECEIVE CRIME ALERTS

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

[Source]

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.

[Source]

Dot size is roughly proportional to tree trunk diameter.

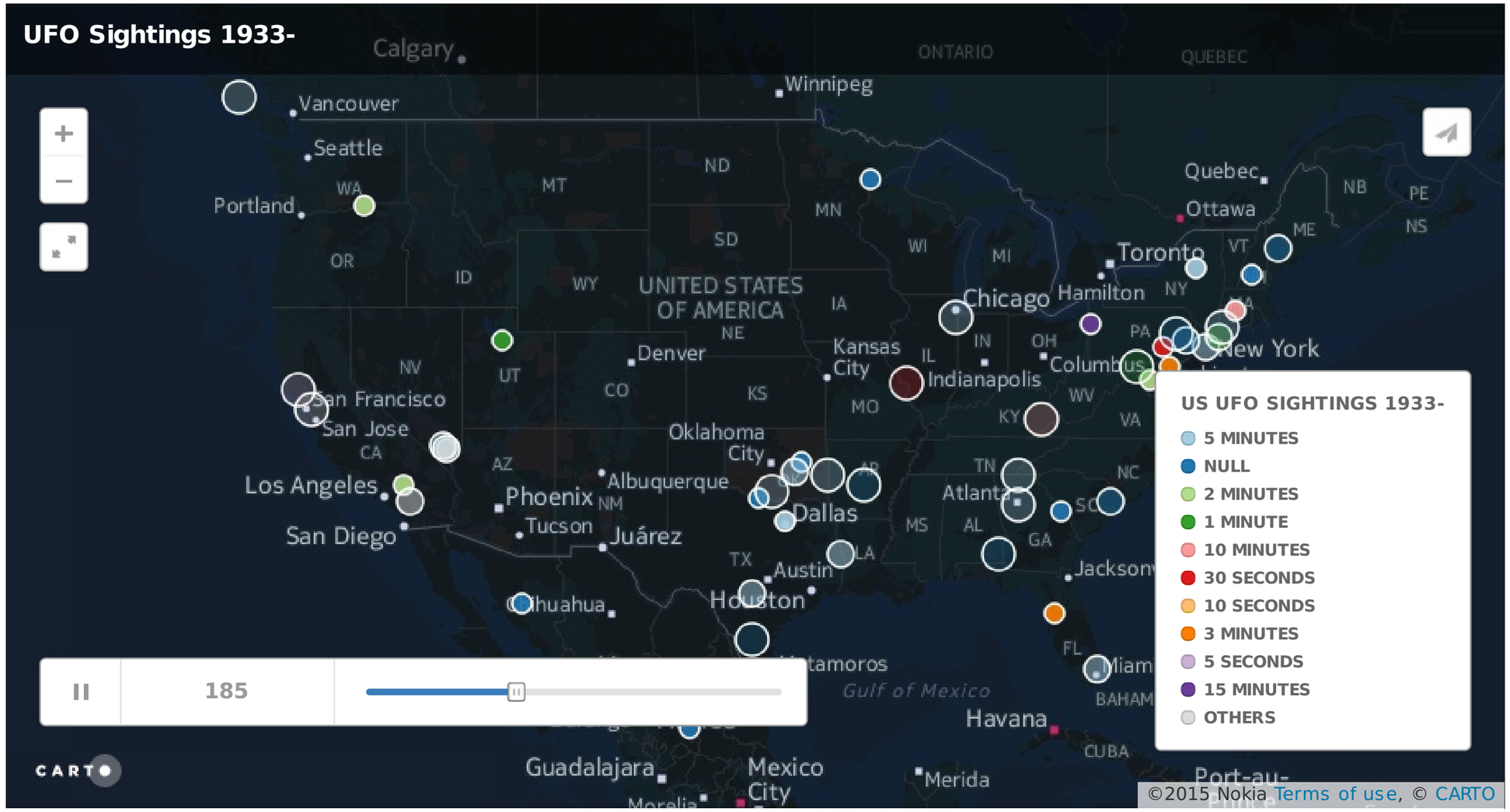
Created by Jill Hubley | Leaflet | Mapbox Terms & Feedback, CartoDB attribution

FILTER BY SPECIES

BASE MAP ON OFF

+ −

UFO Sightings (1933-)



Map created by lcpearso

8

Geo-tagged tweets

Point patterns

Point 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 PPs

Visualization of PPs

Two routes (today):

- Aggregate ↔ "Histogram"
- Smooth ↔ KDE

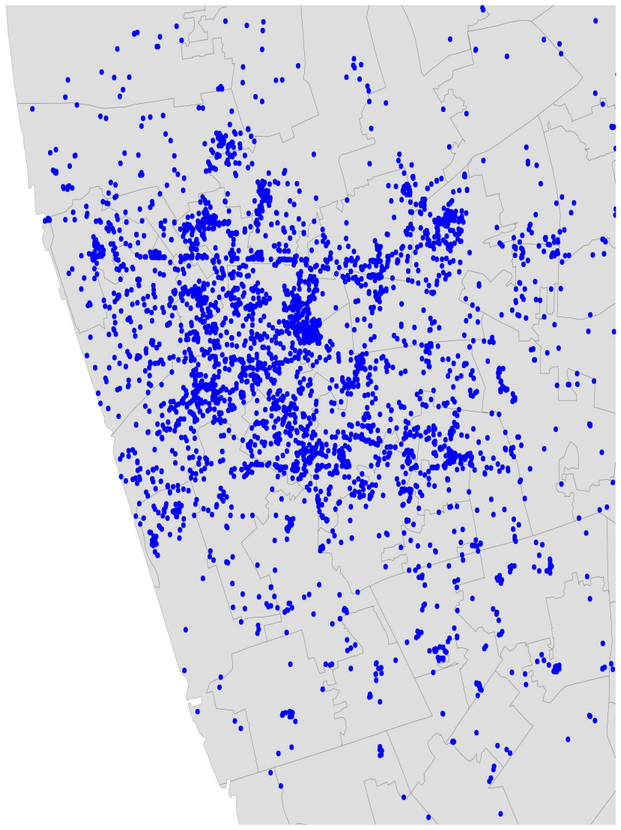
Aggregation

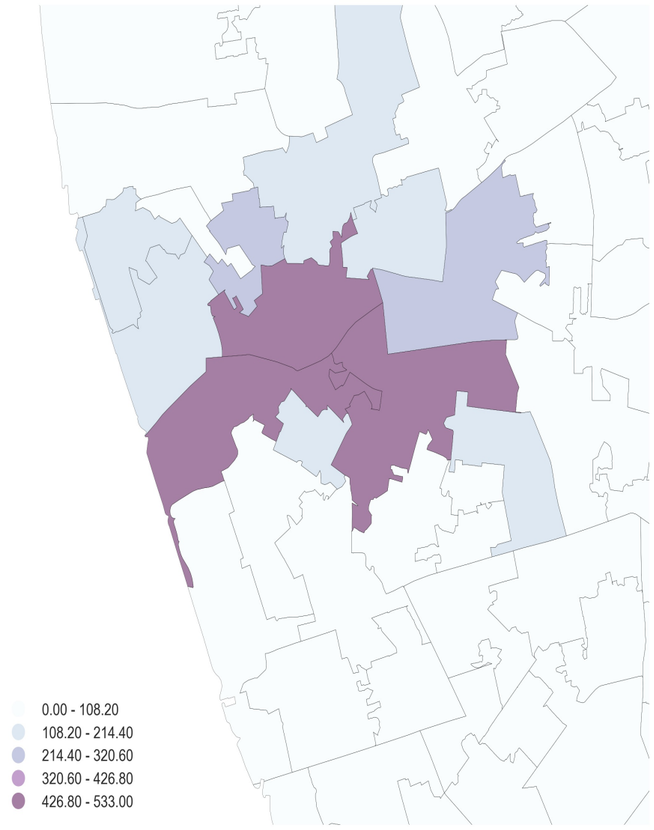
Points meet polygons

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

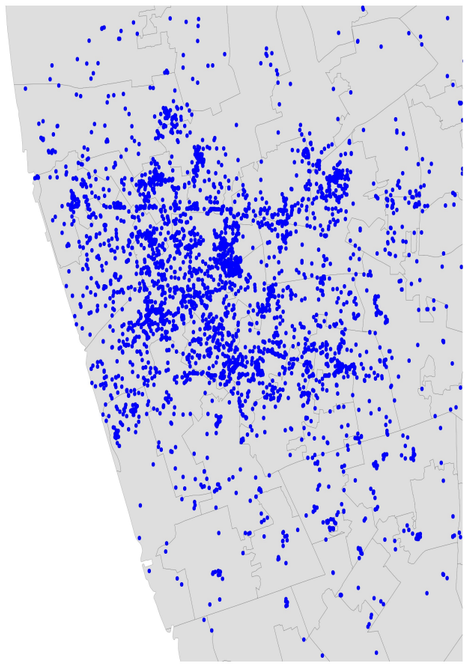
Hex-binning

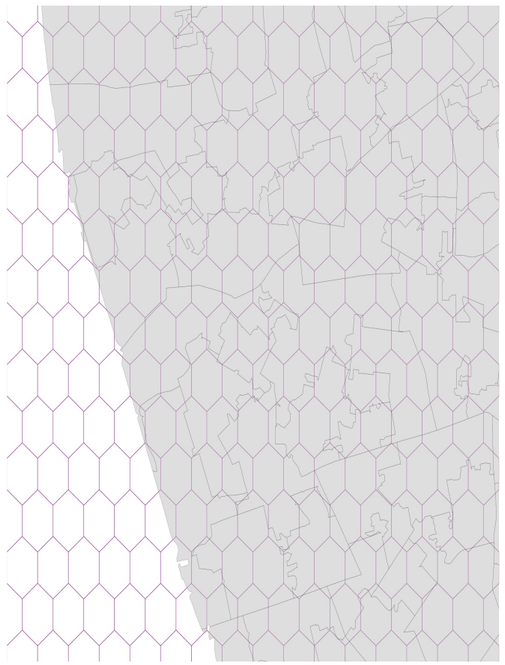
If no polygon boundary seems like a good candidate for aggregation…

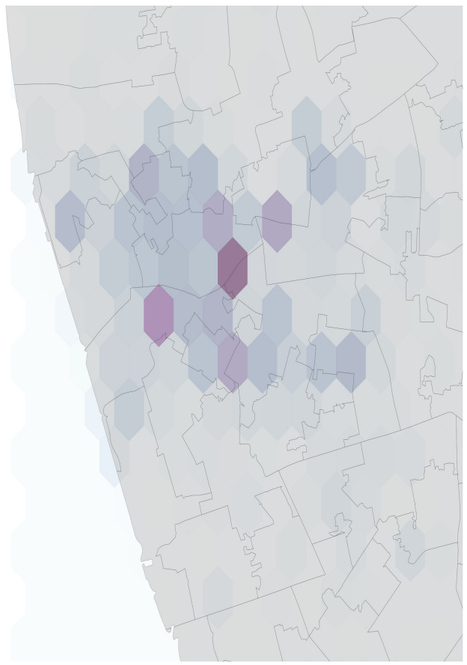
…draw a hexagonal (or squared) tesselation!!!

Hexagons…

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







But…

(Arbitrary) aggregation may induce MAUP (see Lecture 4)

+

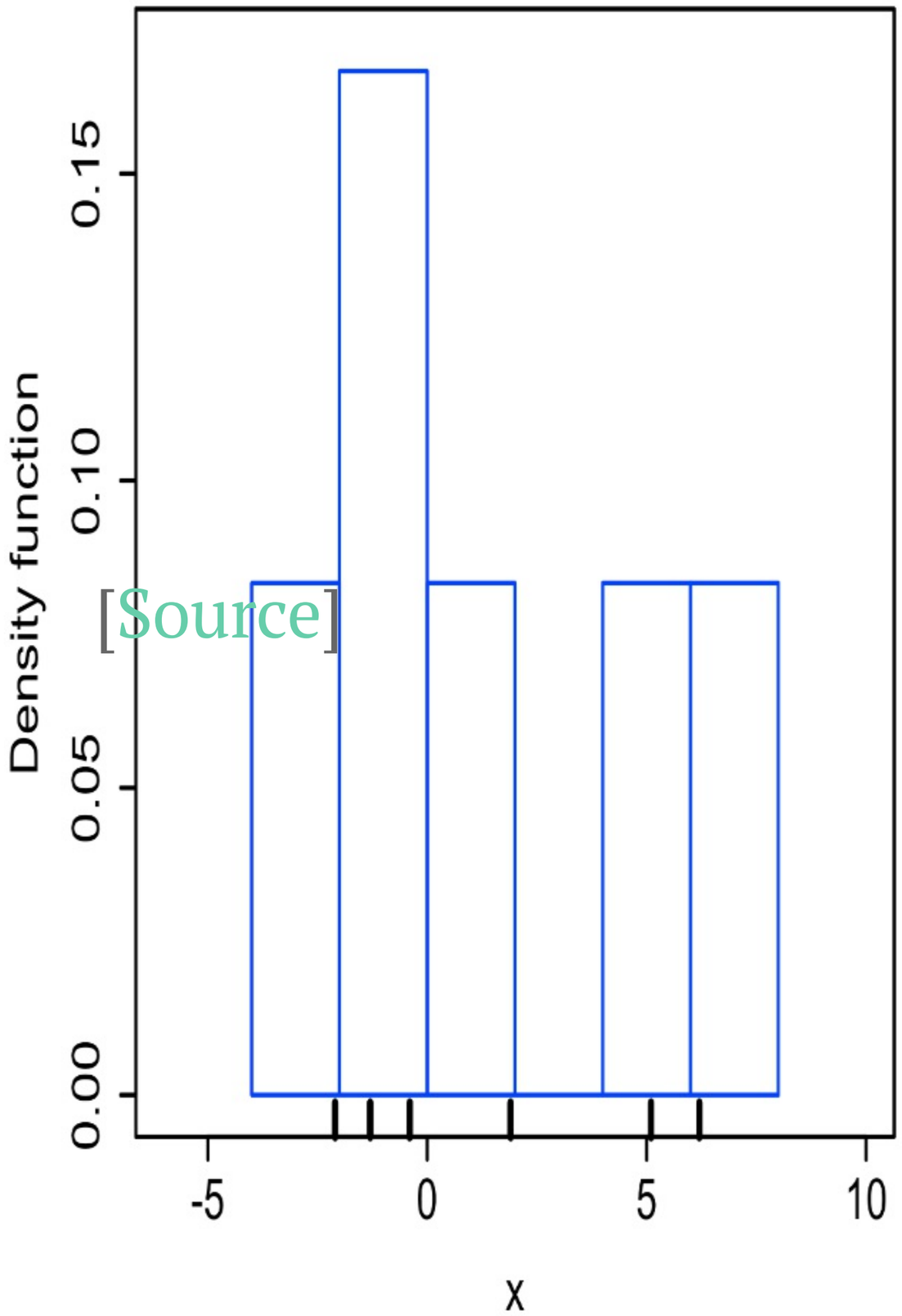
Points usually represent events that affect only part of the population and hence are best considered as rates (see Lecture 4)

Kernel Density Estimation

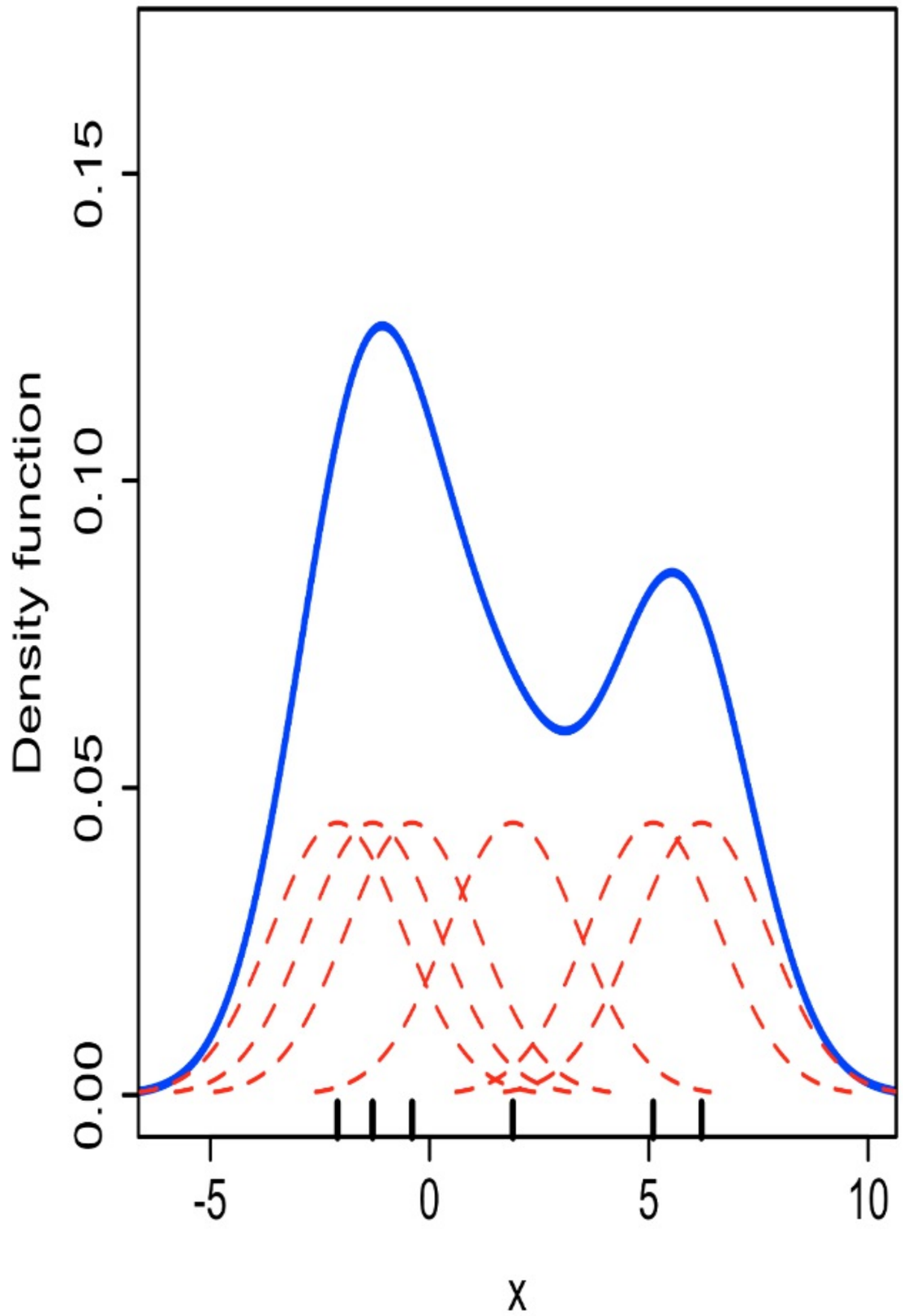
Kernel Density Estimation

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


Density function
0.00
0.05
0.10
0.15
-5
0
5
10
x

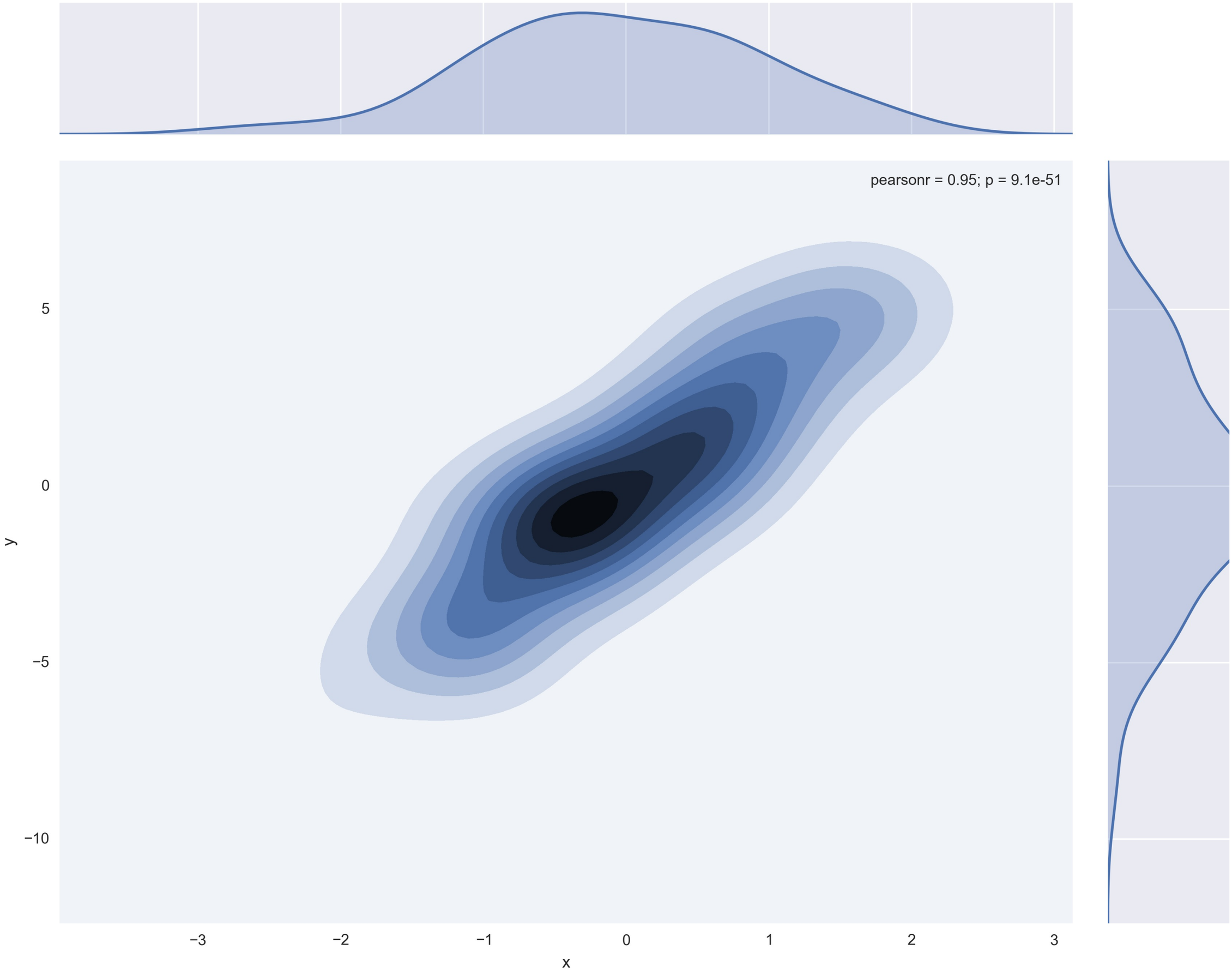
[Source]

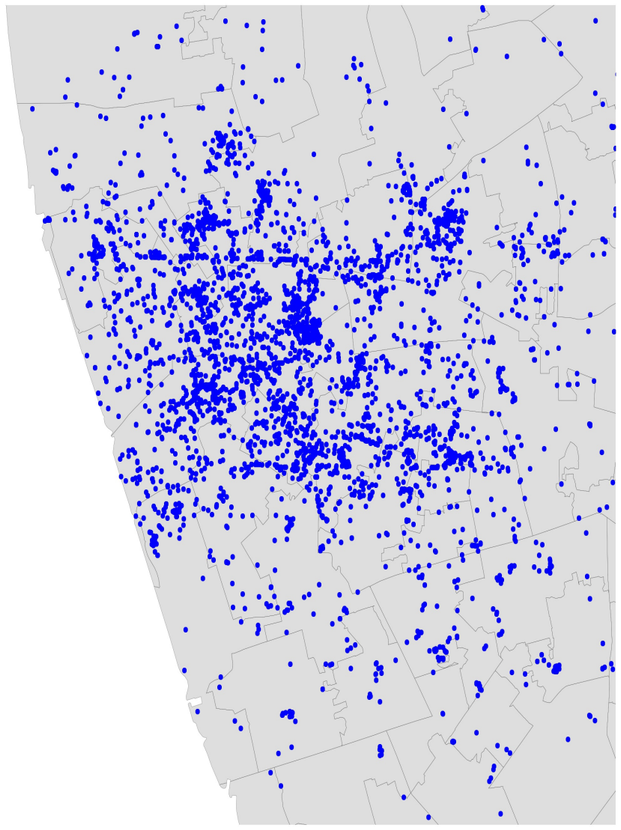

Density function
0.00
0.05
0.10
0.15
-5
0
5
10
x

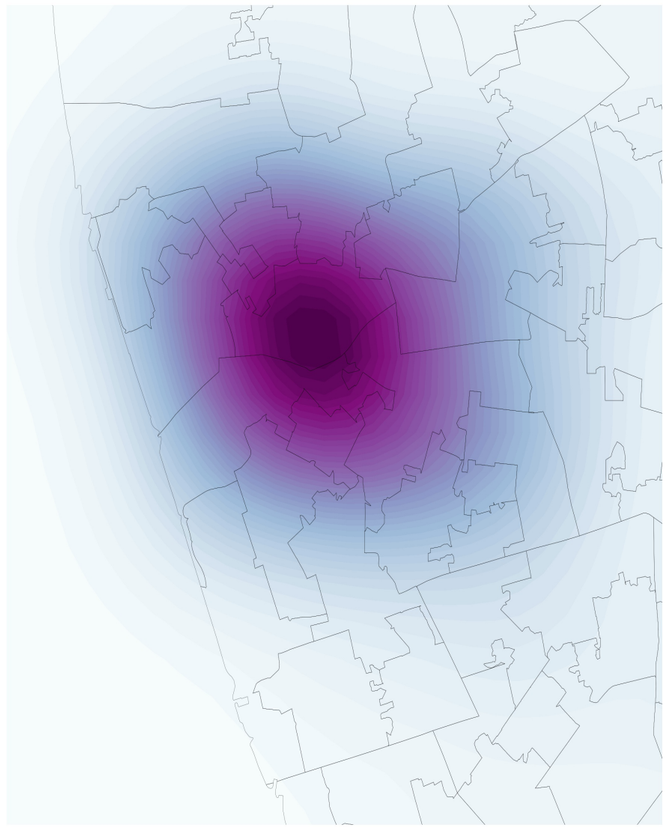
Bivariate (spatial) KDE

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





Finding clusters of PPs

Concentrations/agglomerations of points over space, significantly more so than in the rest of the space considered

Huge literature spanning spatial analysis, statistics and computer science. Today, we'll look at…

Density

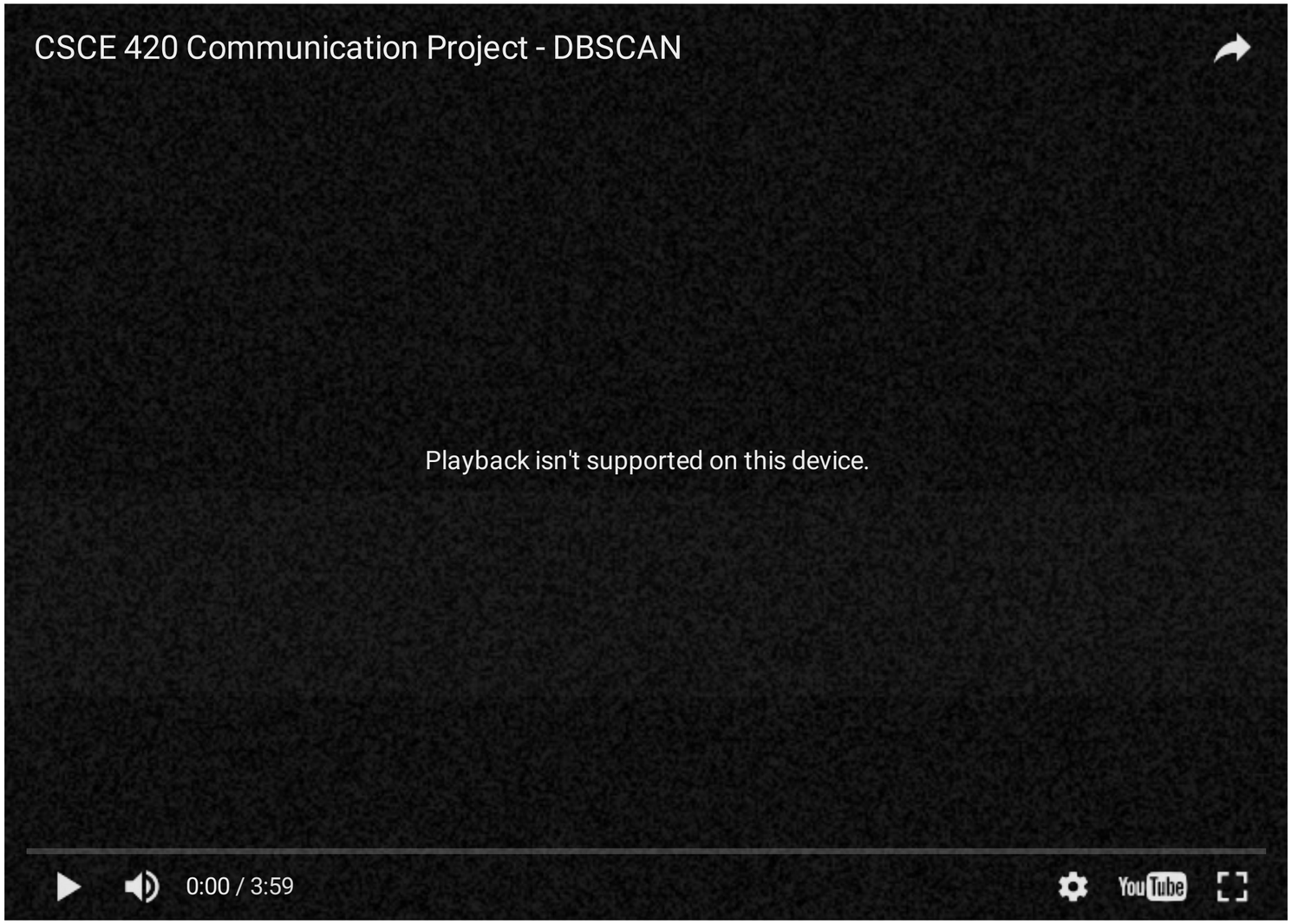
Based

Spatial

Clustering of

Applications with

Noise


CSCE 420 Communication Project - DBSCAN
Playback isn't supported on this device.
0:00 / 3:59

DBSCAN

(Additional) Pros:

- Not necessarily spatial
- Very fast to run so → scales relatively well → applicable to large datasets

(Additional) Cons:

- Not based on any probabilistic model (no inference)
- Hard to learn about the underlying process