

# Geographic Data Science - Lecture VII

## Grouping Data over Space

Dani Arribas-Bel

# Today

- The need to group data
- Geodemographic analysis
- Non-spatial clustering
- Regionalization
- Examples "in the wild"

The need to group data

*Everything should be made as simple as possible, but  
not simpler*

Albert Einstein

# The need to group data

- The world (and its problems) are **complex** and **multidimensional**
- **Univariate** analysis involves focusing **only one** way of measure the world

# The need to group data

- The world (and its problems) are **complex** and **multidimensional**
- **Univariate** analysis involves focusing **only one** way of measure the world
- Sometimes, world issues are best understood as **multivariate**:
  - Percentage of foreign-born Vs. *What is a neighborhood?*
  - Years of schooling Vs. *Human development*
  - Monthly income Vs. *Deprivation*

## *Grouping as simplifying*

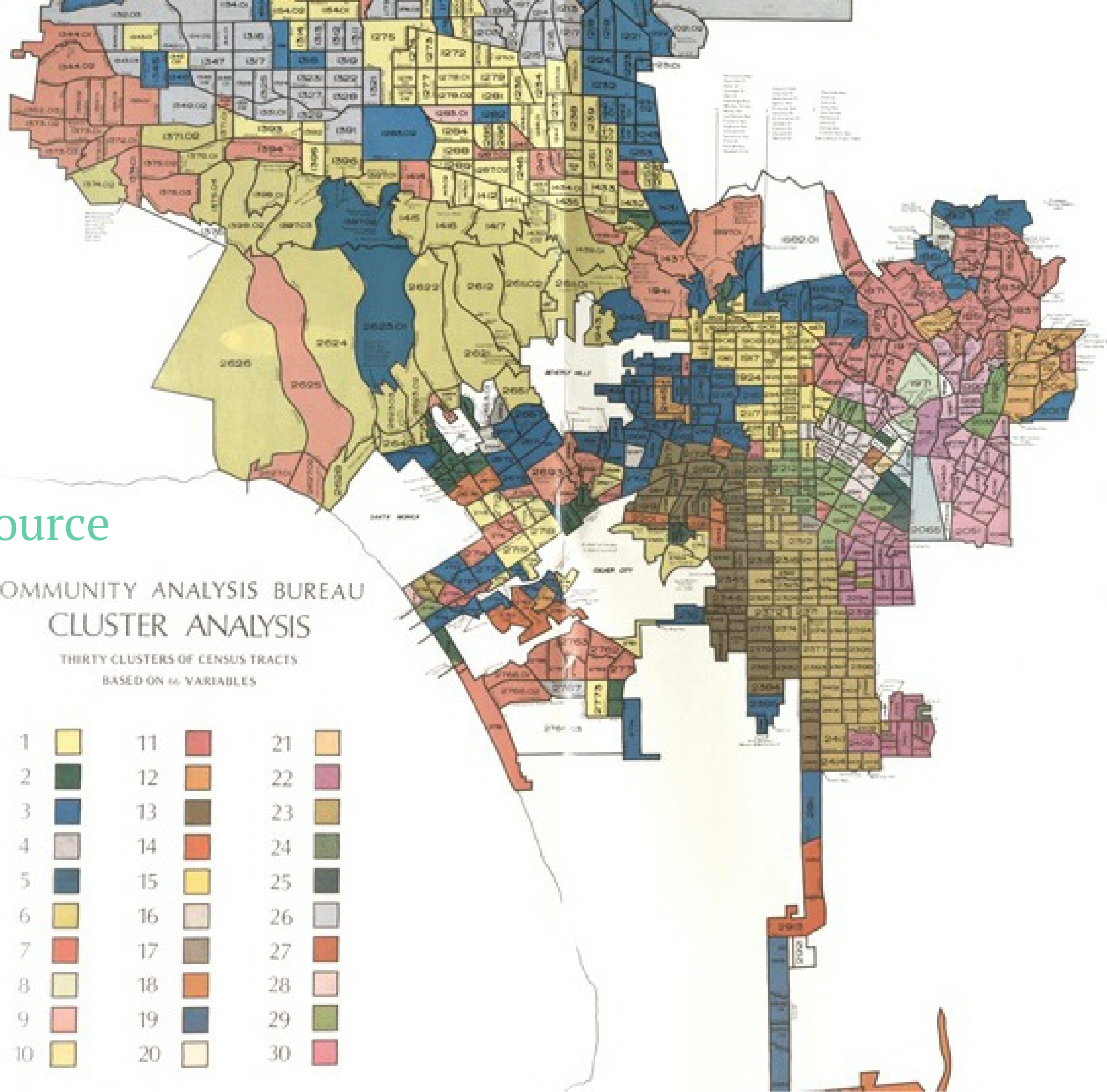
- Define a given number of categories based on **many characteristics** (multi-dimensional)
- Find the **category** where each observation *fits best*
- **Reduce complexity**, keep all the **relevant information**
- Produce easier-to-understand outputs

# Geodemographic analysis



# Geodemographic analysis

- Technique developed in 1970's attributed to Richard Webber
- **Identify similar neighborhoods** → Target urban deprivation funding
- Originated in the **Public** Sector (policy) and spread to the **Private** sector (marketing and business intelligence)



Source

COMMUNITY ANALYSIS BUREAU  
 CLUSTER ANALYSIS  
 THIRTY CLUSTERS OF CENSUS TRACTS  
 BASED ON 46 VARIABLES

# CDRC Maps

**DATA CHOOSER**

Classifications: Retail

Select a map:  
2011 Area Classif/n of OAs

[Download this data](#)

**MAP OPTIONS**

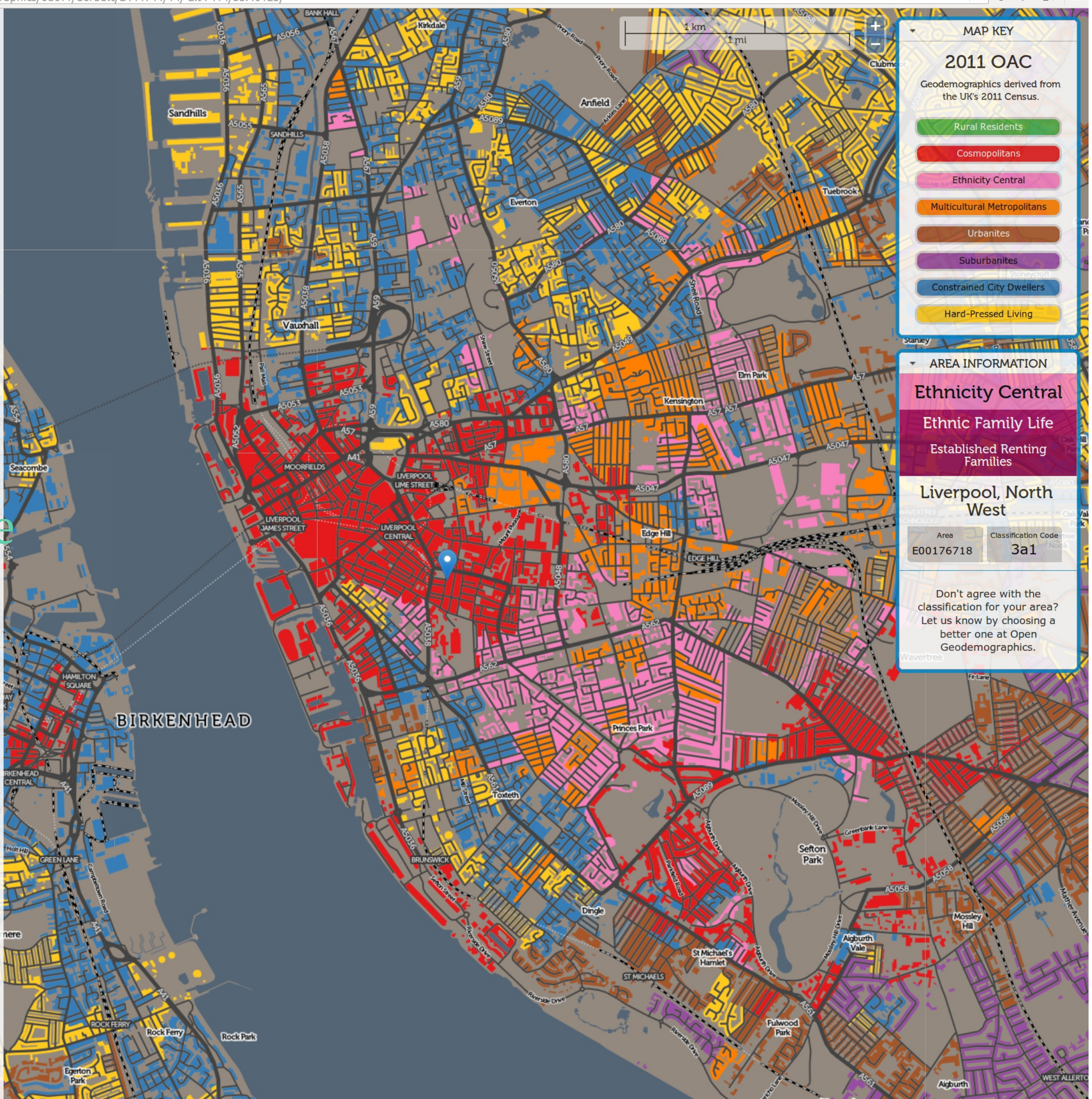
Layers: Land Labels

Toggle: Retail Centres

Download: retail centre locations

Postcode:

Source



**MAP KEY**

**2011 OAC**  
Geodemographics derived from the UK's 2011 Census.

- Rural Residents
- Cosmopolitans
- Ethnicity Central
- Multicultural Metropolitans
- Urbanites
- Suburbanites
- Constrained City Dwellers
- Hard-Pressed Living

**AREA INFORMATION**

**Ethnicity Central**

Ethnic Family Life  
Established Renting Families

**Liverpool, North West**

Area	Classification Code
E00176718	3a1

Don't agree with the classification for your area? Let us know by choosing a better one at Open Geodemographics.

Like Share Tweet

Important note: Classifications are an average across the local area, rather than for individual houses, therefore the colour coding on a building is not necessarily indicative of that building.

*How do you segment/cluster observations over space?*

- Statistical clustering
- Explicitly spatial clustering (regionalization)

# Non-spatial clustering

**Split** a dataset into **groups** of observations that are **similar within** the group and **dissimilar between** groups, based on a series of **attributes**

**Machine learning**

**Unsupervised**

## **Machine learning**

- The computer *learns* some of the properties of the dataset without the human specifying them

## **Unsupervised**



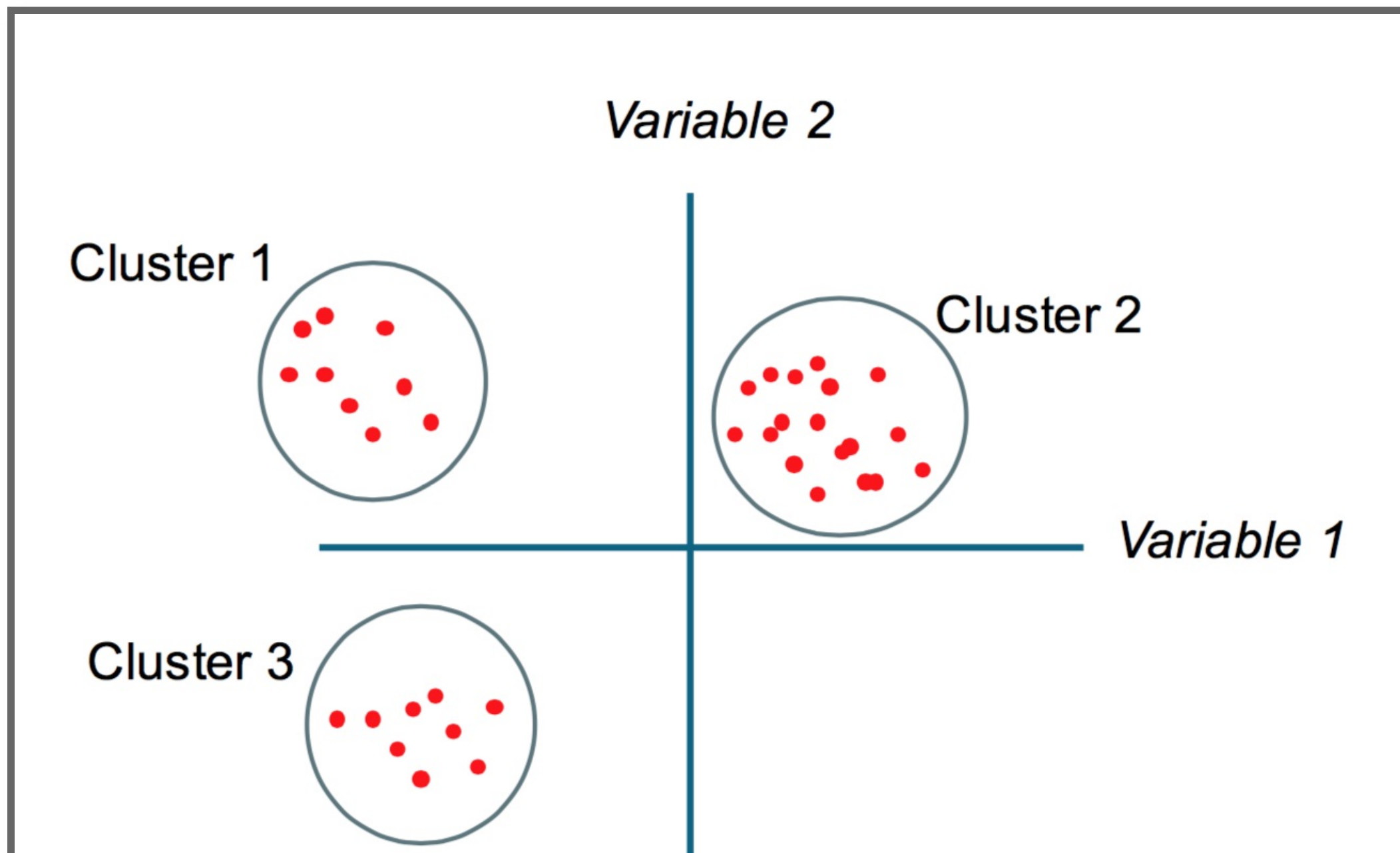
## **Machine learning**

- The computer *learns* some of the properties of the dataset without the human specifying them

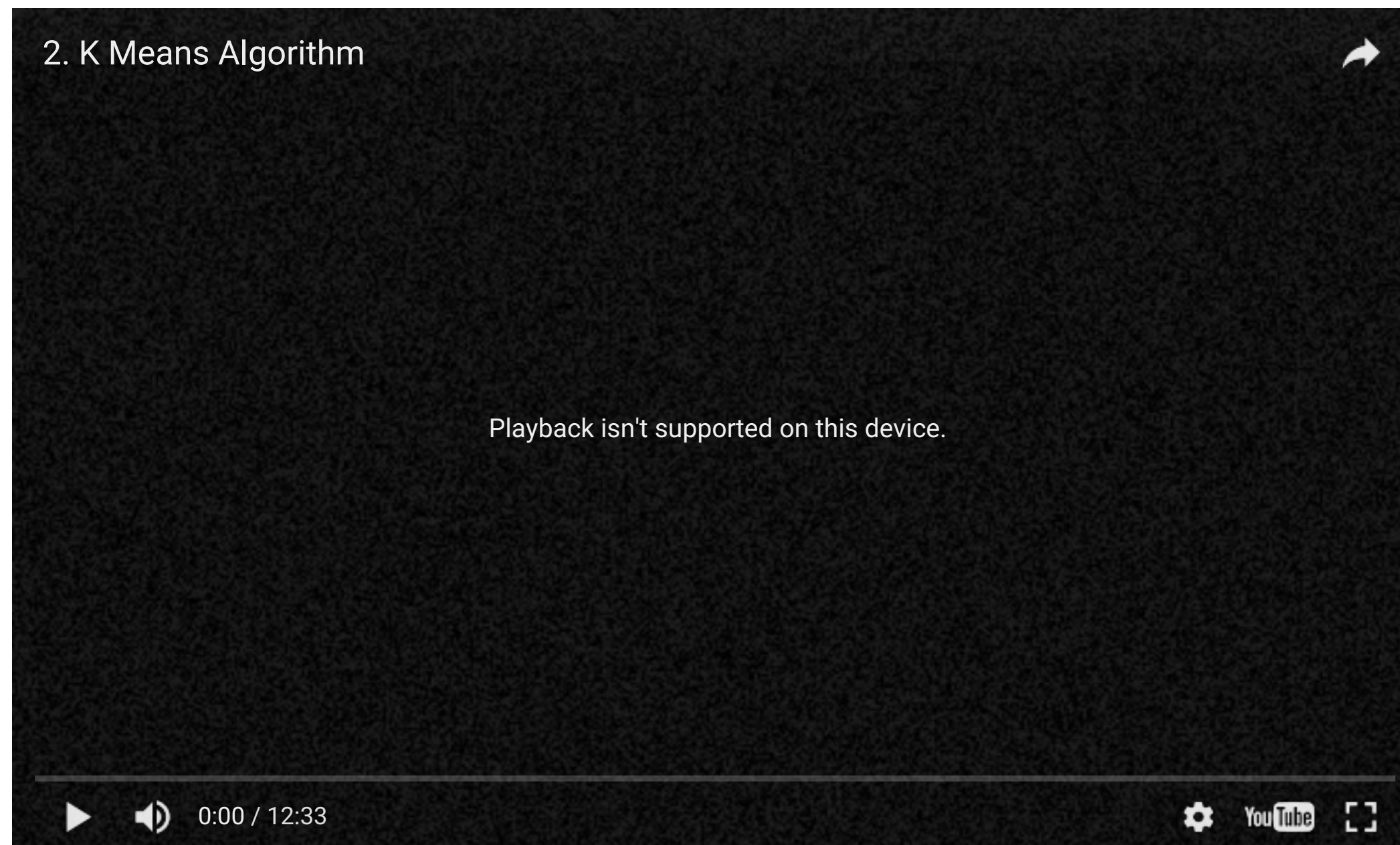
## **Unsupervised**

- There is no a-priori structure imposed on the classification → before the analysis, no observations is in a category

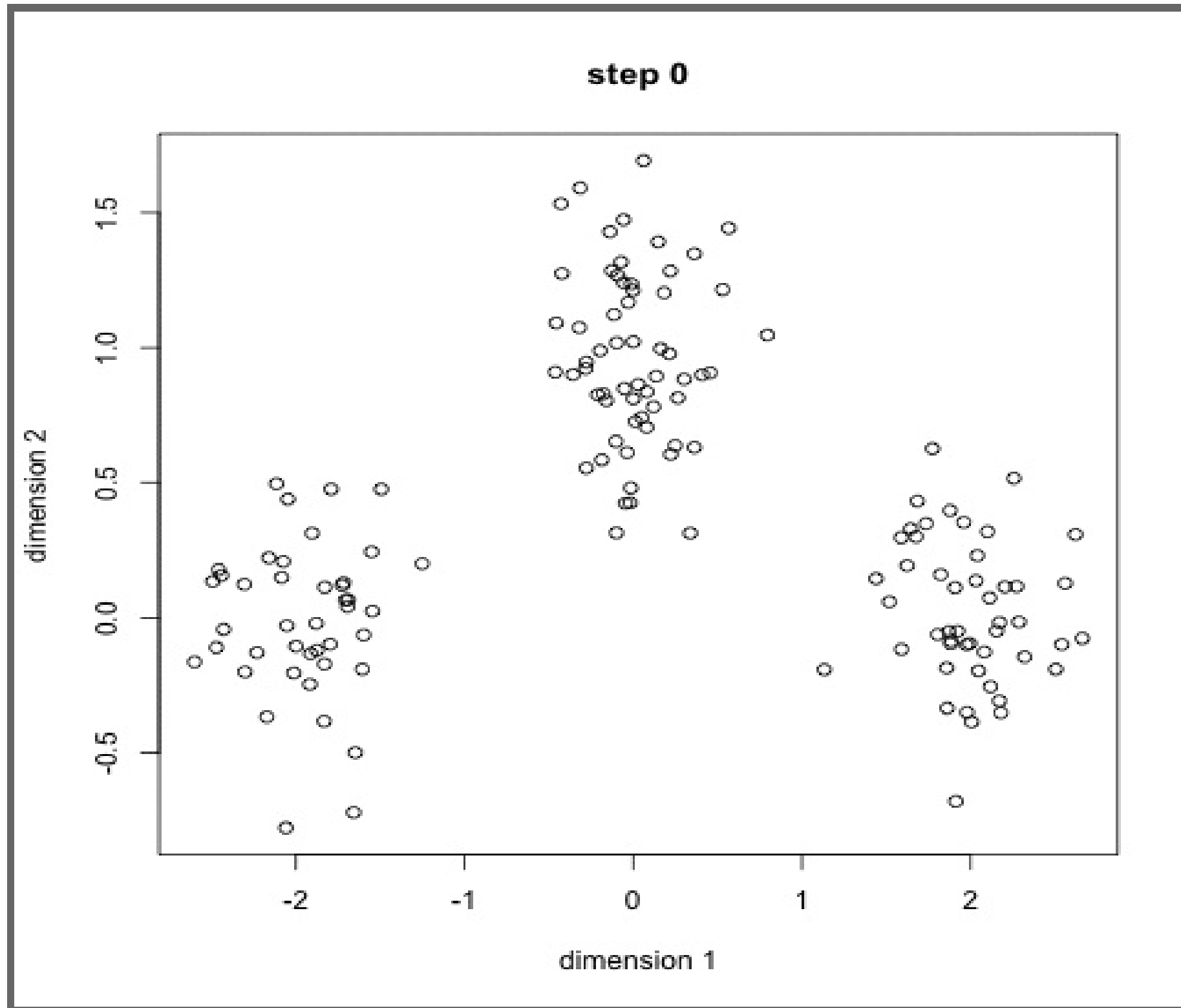
# Intuition



# K-means [Source]



# K-means [Source]



# More clustering...

- Hierarchical clustering
- Agglomerative clustering
- Spectral clustering
- Neural networks (e.g. Self-Organizing Maps)
- DBScan
- ...

Different properties, different best usecases

See [interesting comparison](#) table

# Regionalization

# Machine Learning

# **Spatial** Machine Learning



## **Spatial** Machine Learning

*Aggregating basic spatial units (**areas**) into larger units (**regions**)*

# Regionalization

**Split** a dataset into **groups** of observations that are **similar within** the group and **dissimilar between** groups, based on a series of **attributes...**

# Regionalization

**Split** a dataset into **groups** of observations that are **similar within** the group and **dissimilar between** groups, based on a series of **attributes...**

...with the additional constraint observations need to be **spatial neighbors**

# Regionalization

Duque et al. (2007)

# Regionalization

- All the methods aggregate geographical areas into a predefined number of regions, while optimizing a particular aggregation criterion;

Duque et al. (2007)

# Regionalization

- The areas within a region must be geographically connected (the spatial contiguity constraint);

Duque et al. (2007)

# Regionalization

- The number of regions must be smaller than or equal to the number of areas;

Duque et al. (2007)

# Regionalization

- Each area must be assigned to one and only one region;

Duque et al. (2007)



# Regionalization

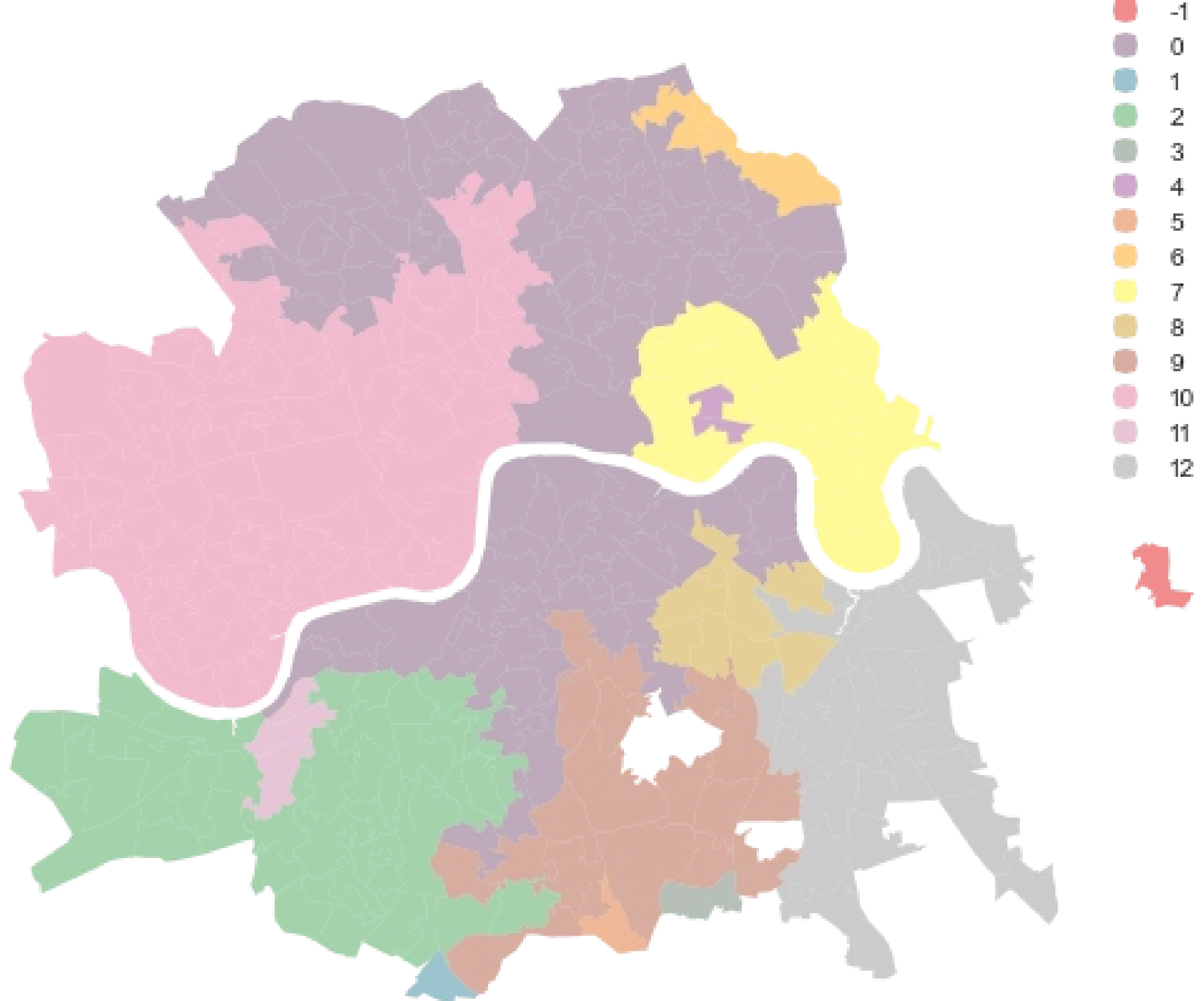
- Each region must contain at least one area.

Duque et al. (2007)

# Regionalization

- All the methods aggregate geographical areas into a predefined number of regions, while optimizing a particular aggregation criterion;
- The areas within a region must be geographically connected (the spatial contiguity constraint);
- The number of regions must be smaller than or equal to the number of areas;
- Each area must be assigned to one and only one region;
- Each region must contain at least one area.

Duque et al. (2007)



# Algorithms

- Automated Zoning Procedure (AZP)
- Arisel
- Max-P
- ...

See [Duque et al. \(2007\)](#) for an excellent, though advanced, overview

# Examples

# Census geographies

*Environment and Planning A* 1995, volume 27, pages 425-446

---

## Algorithms for reengineering 1991 Census geography

---

S Openshaw, L Rao†

School of Geography, University of Leeds, Leeds LS2 9JT, England

Received 22 April 1994; in revised form 6 October 1994

---

# Find a Neighbourhood in London

What kind of neighbourhood are you looking for?

- Dining ×
- Artsy 7
- Nightlife 9
- Shopping 15
- Loved by Londoners 11
- Touristy 14
- Great Transit 13
- Museums & the Arts 10
- Peace & Quiet 7
- Along the River Thames 5
- West End 7

23 neighbourhoods match Dining. [See all listings](#)

## AirBnb neighborhoods



### Soho

Giving the green light to its red light reputation.

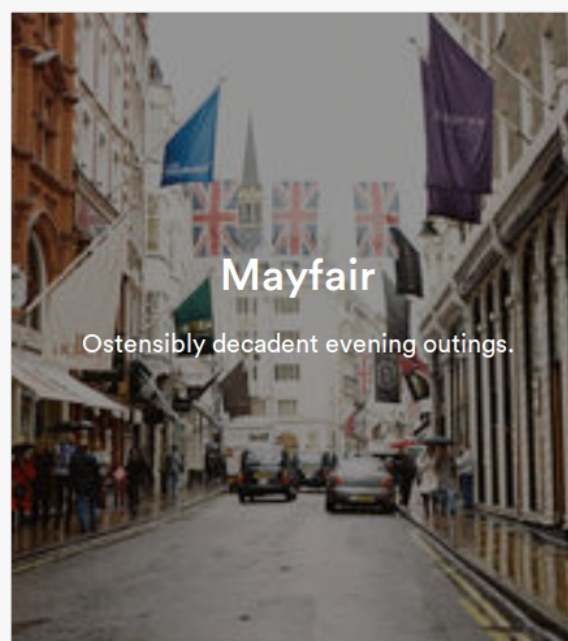
Eclectic • Open-minded • China Town • Lgbt Friendly



### Whitechapel/Brick Lane

Gritty outside with gallant insides.

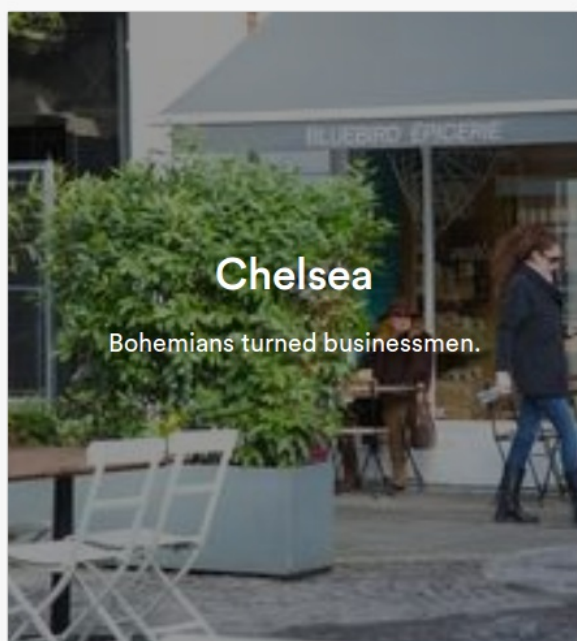
Buzzing • Curry • Arty • Graffiti



### Mayfair

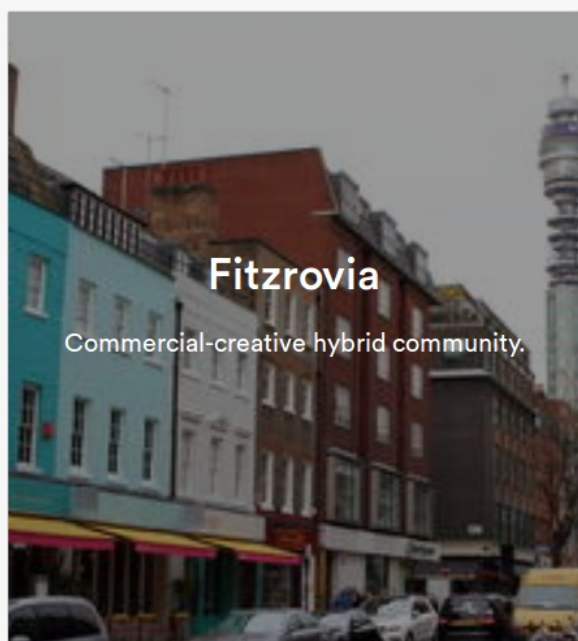
Ostensibly decadent evening outings.

Luxury Shopping • Ballrooms • Heart Of London • Fashionable



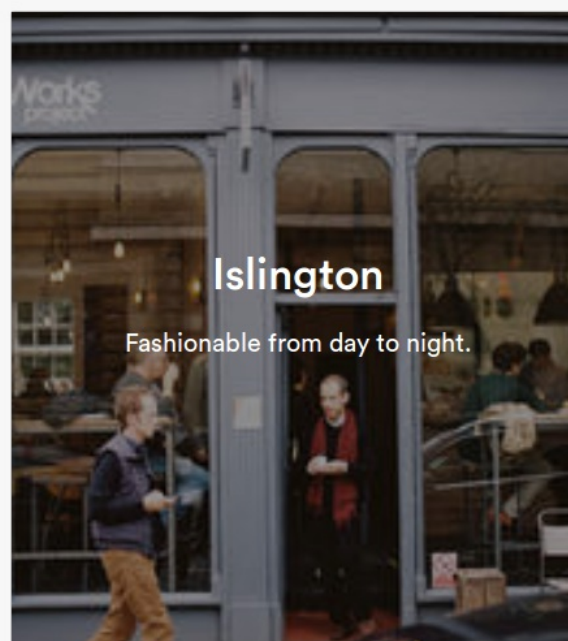
### Chelsea

Bohemians turned businessmen.



### Fitzrovia

Commercial-creative hybrid community.



### Islington

Fashionable from day to night.



Livehoods are a new way to understand the city, using social media data.

### Re-Imagining the City in the Age of Social Media

Livehoods offer a new way to conceptualize the dynamics, structure, and character of a city by analyzing the social media its residents generate. By looking at people's check-in patterns at places across the city, we create a mapping of the different dynamic areas that comprise it. Each Livehood tells a different story of the people and places that shape it.

> MORE

### Using Machine-Learning to Study Cities

Our research hypothesis is that the character of an urban area is defined not just by the types of places found there, but also by the people that make it part of their daily life. To explore this idea, we use data from approximately 18 million check-ins collected from the location-based social network foursquare, and apply clustering algorithms to discover the different areas of the city.

> MORE

## Livehoods

### Current Maps



> New York City



> San Francisco



> Pittsburgh



> More Maps

### News and Press

#### Livehood at ICWSM

Our work with Livehoods won the best paper award at ICWSM in Dublin this June! **Watch the video from our presentation.**

#### Livehoods on CBC Radio

Justin was on the CBC Radio program Spark talking with host Nora Young about the Livehoods Project. **Listen to the full interview.**

#### Livehoods in the Atlantic

Livehoods appeared as the Map of the Day on the Atlantic's Cities blog. **See their post about us.**

#### Wired Insider

Wired's Insider blog says Livehoods is "taking a big swing" at mining insights into "cultural habits and how societies flow." **Read the full post.**

> MORE

### Recent Tweets

#### @tiffehr

Best map/location mashup I've seen in quite some time: <http://livehoods.org/maps/nyc#> (Via <http://roomthily.tumblr.com>)

#### @Werner

Livehoods is a cool CMU research project to visualize cities through the use of social media (@foursquare in this case) <http://ww.ly/IJZ3We>

#### @tomcoates

The 'Related' tab on <http://livehoods.org> is the best. See which neighboring places people travel too. Algorithmic divination of commuting!

#### @brainpicker

Forget neighborhoods, it's about Livehoods — Carnegie Mellon maps the dynamic character of cities through social media <http://j.mp/HzmkoN>

#### @kellan

clearly i live on the wrong side of the bqe - <http://livehoods.org/maps/nyc>

> MORE

### Subscribe to our newsletter

Find out more about Livehoods and get updates on future developments by subscribing to our mailing list.

EMAIL\*

NAME

SUBSCRIBE



# Recapitulation

- Some problems are truly **highly dimensional** and univariate representations are not appropriate
- **Clustering** can help reduce complexity by creating **categories** that retain statistical information but are easier to understand
- Two main types of clustering in this context:
  - Geo-demographic analysis
  - Regionalization



Geographic Data Science'16 - Lecture 7 by **Dani Arribas-Bel** is licensed under a **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License**.