

# Geographic Data Science - Lecture VIII

## 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 is **complex** and multidimensional
- **Univariate** analysis focuses on **only one dimension**
- Sometimes, world issues are best understood as **multivariate**. E.g.
  - Percentage of foreign-born Vs. *What is a neighborhood?*
  - Years of schooling Vs. *Human development*
  - Monthly income Vs. *Deprivation*

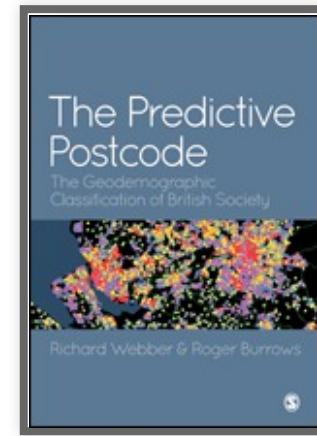
# *Grouping as simplifying*

- 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

- 1970's, Richard Webber
- Identify similar neighborhoods
  - Target urban deprivation funding
- Public Sector (policy) →  
Private sector (marketing and business intelligence)



# CDRC Maps

**DATA CHOOSEN**

Classifications Retail

Select a map: 2011 Area Classif/n of OAs

Download this data

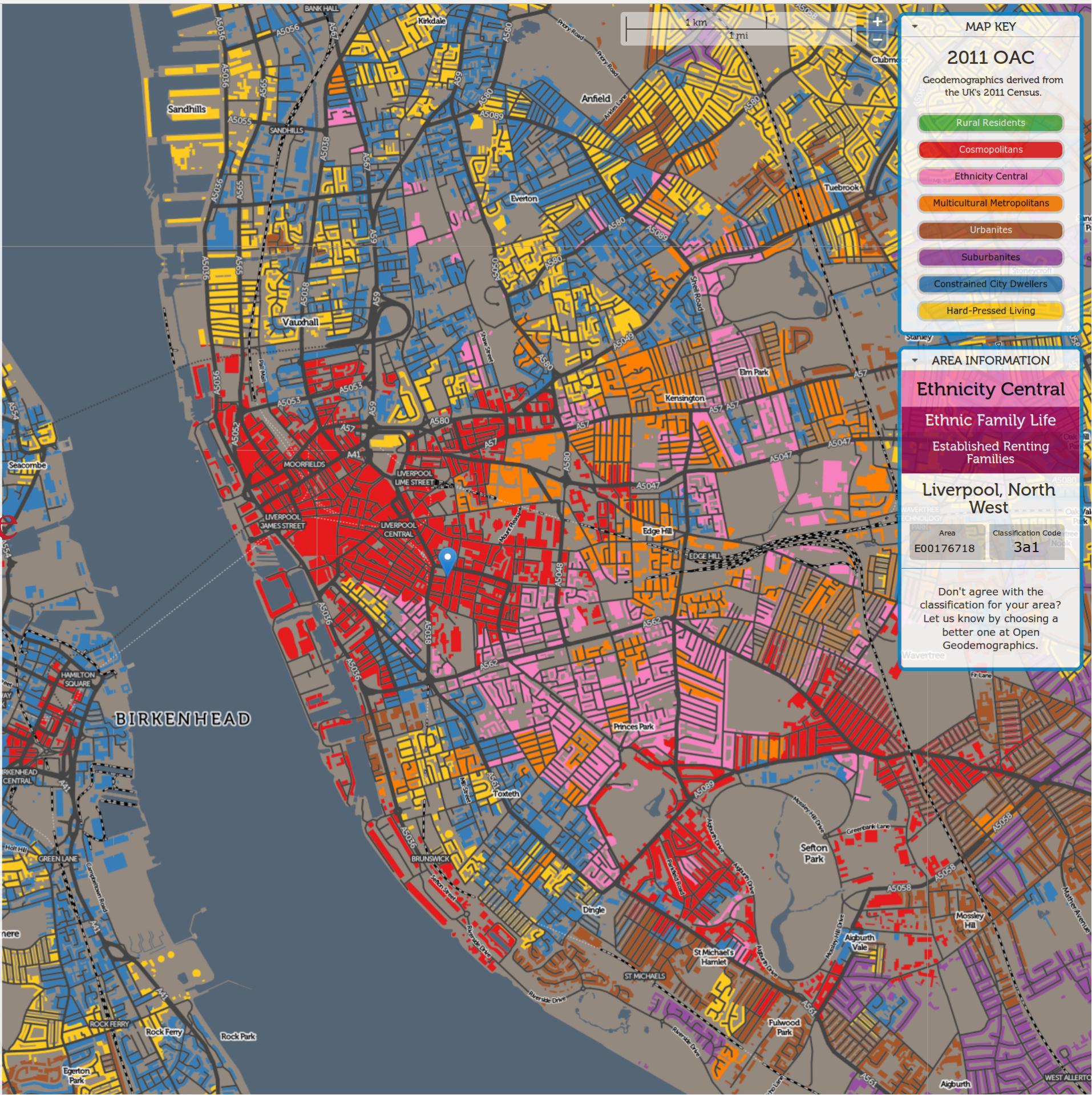
**MAP OPTIONS**

Layers: Land Labels

Toggle: Retail Centres

Download: retail centre locations

Postcode: L19dw | Go



Source

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

- Statistical clustering
- Explicitly spatial clustering (regionalization)

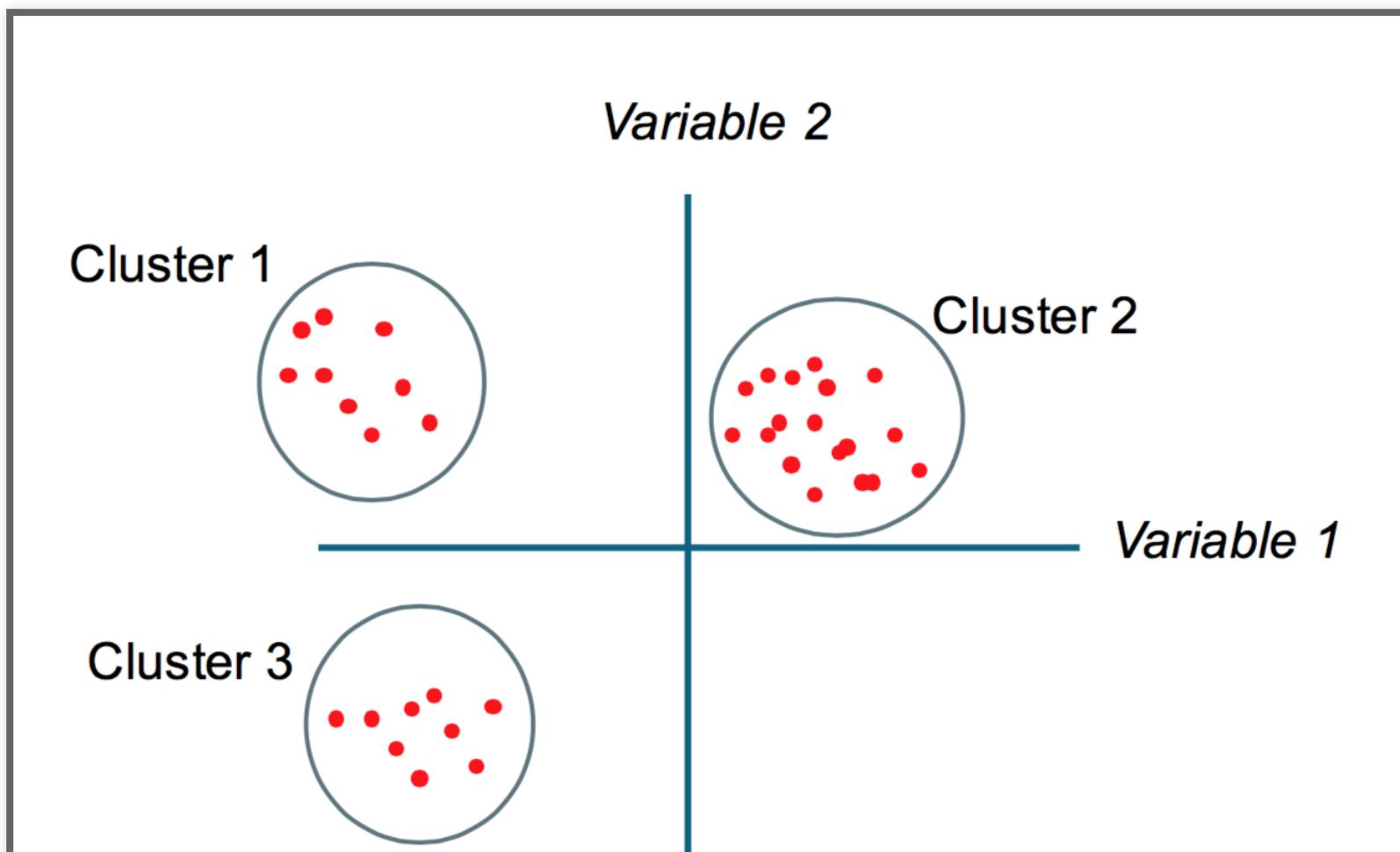
# Non-spatial clustering

**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

# Intuition



# 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

# Unsupervised 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**...

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

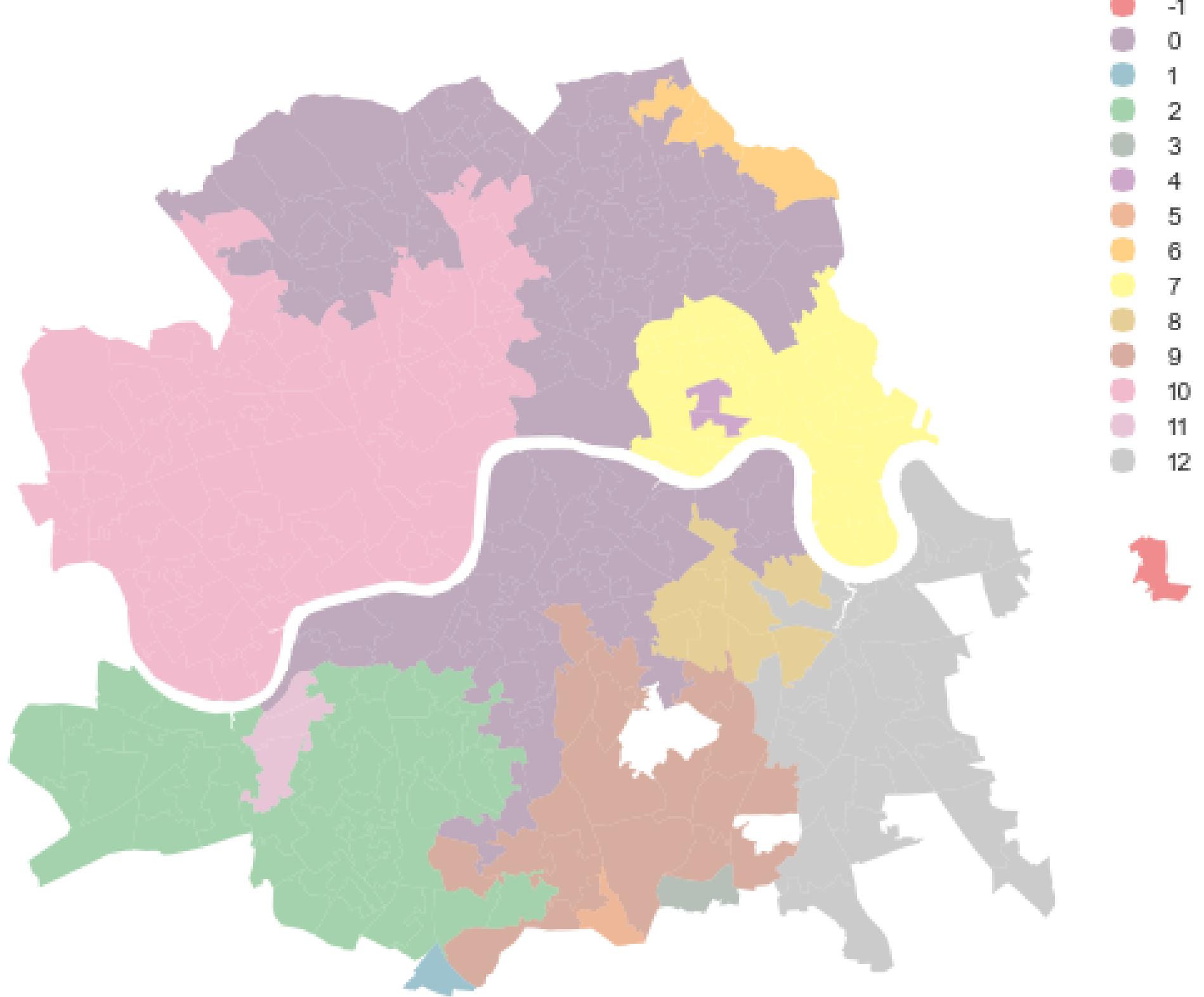
# Regionalization

Duque et al. (2007)

# Regionalization

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

# Non-spatial clustering

Spotify

Browse Radio

YOUR LIBRARY

Your Daily Mix

Recently Played

Songs

Albums

Artists

Stations

Local Files

Videos

Podcasts

PLAYLISTS

Discover Weekly

Release Radar

Classical

Electronic

HipHop

Jazz

Playlists by others

Tras la cena

After dinner

Cumple

Mi patria

party - summer street

Deftones walk

TX-trip

Liked from Radio

Krugman Friday Night Music

Dinner

Cosmopolitan

Personal-recommendations

Book bands

New Playlist

< > Search

dreamessence

Friend activity

townhallsymphon... Step Out José González The Secret Life Of...

1236163056 Divine Hammer The Breeders Last Splash

\_maxi Sensationz of the M... Mental Nomad Subterranean Hit...

FIND FRIENDS

# Your Daily Mixes

Play the music you love, without the effort. Packed with your favorites and new discoveries.

**Your Daily Mix 1**  
Daily Mix 1  
Gata Cattana, DELLA FUENTE, ToteKing and more  
MADE FOR DREAMESSENCE

**Your Daily Mix 2**  
Daily Mix 2  
Aphex Twin, George FitzGerald, Nosaj Thing and more  
MADE FOR DREAMESSENCE

**Your Daily Mix 3**  
Daily Mix 3  
Dedekind Cut, Helena Hauff, Taylor Deupree and more  
MADE FOR DREAMESSENCE

**Your Daily Mix 4**  
Daily Mix 4  
Berliner Philharmoniker, Alexandre Tharaud, Sir Colin Davis and more  
MADE FOR DREAMESSENCE

**Your Daily Mix 5**  
Daily Mix 5  
GoGo Penguin, Yussef Kamaal, Blue Lab Beats and more  
MADE FOR DREAMESSENCE

**Your Daily Mix 6**  
Daily Mix 6  
Fleet Foxes, Andrew Bird, Iron & Wine and more  
MADE FOR DREAMESSENCE

X ◀ ▶ ⏪ ⏩ ⏴ ⏵

CA Homepage | CA Political

https://ca-political.com/?\_hstc=163013475.b1001e9ef2d9e98e22fa0fd952f21a5c.1528388939083.1528388939083.1528388939083.1&\_hssc=163013

Read CA's latest press releases

CA  
Political

# Data-driven campaigns

**Cambridge Analytica (UK) Limited, SCL Group Limited, SCL Analytics Limited, SCL Commercial Limited, SCL Social Limited and SCL Elections Limited (together "the Companies")**

On 3 May 2018, Vincent John Green and Mark Newman, insolvency practitioners at Crowe Clark Whitehill LLP were appointed independent Joint Administrators of the Companies under order of the High Court.

Please click on this link to contact the Joint Administrators.

Contact address:  
Crowe Clark Whitehill LLP, 4 Mount Ephraim Road, Tunbridge Wells, Kent TN1 1EE  
E-mail:  
recoverysolutions@crowecw.co.uk

Let's talk



We find your voters and move them to action.

CA Political has redefined the relationship between data and campaigns. By knowing your electorate better, you can achieve greater influence while lowering overall costs.

**TRUMP**  
MAKE AMERICA GREAT AGAIN!

TedCruz<sup>2016</sup>

BEN CARSON  
FOR PRESIDENT 2016

JOHN  
**BOLTON**  
PAC

MAKE AMERICA  
NUMBER 1

THOM  
**TILLIS**  
U.S. Senate

# Regionalisation

# Census geographies

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

---

## Algorithms for reengineering 1991 Census geography

---

S Openshaw, L Rao<sup>†</sup>

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

Received 22 April 1994; in revised form 6 October 1994

---



By livehoods A new way to understand the city, using social media.

## 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 checkin 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://wv.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'18 - Lecture 8 by **Dani Arribas-Bel** is licensed under a **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License**.