

Unsupervised Learning

Dani Arribas-Bel

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*

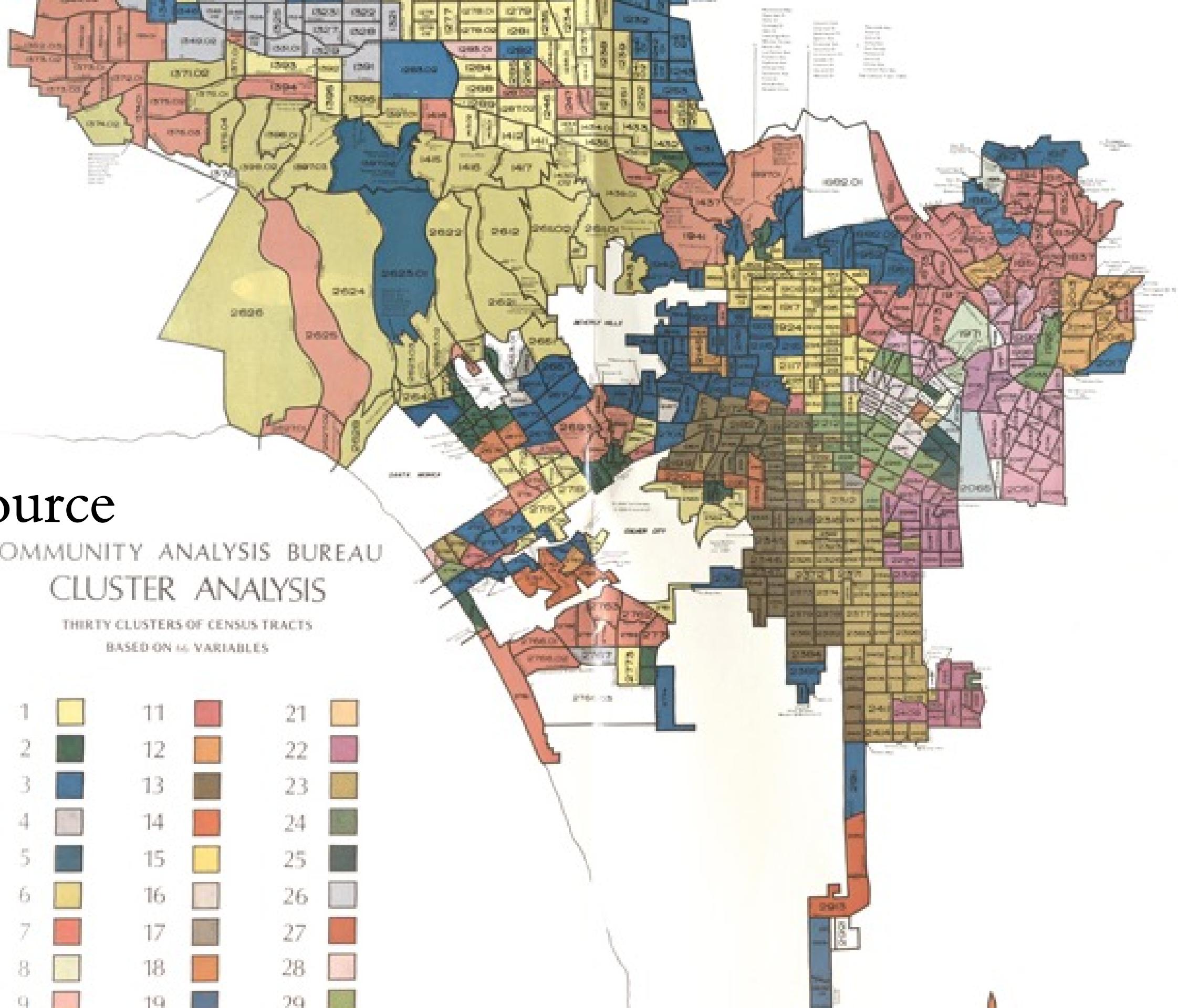
Grouping as simplifying

- 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

- 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)



CDRC Maps

Geodemographics derived from the UK's 2011 Census.

DATA CHOOSEN

Classifications Retail

Select a map:
2011 Area Classif/n of OAs

[Download this data](#)

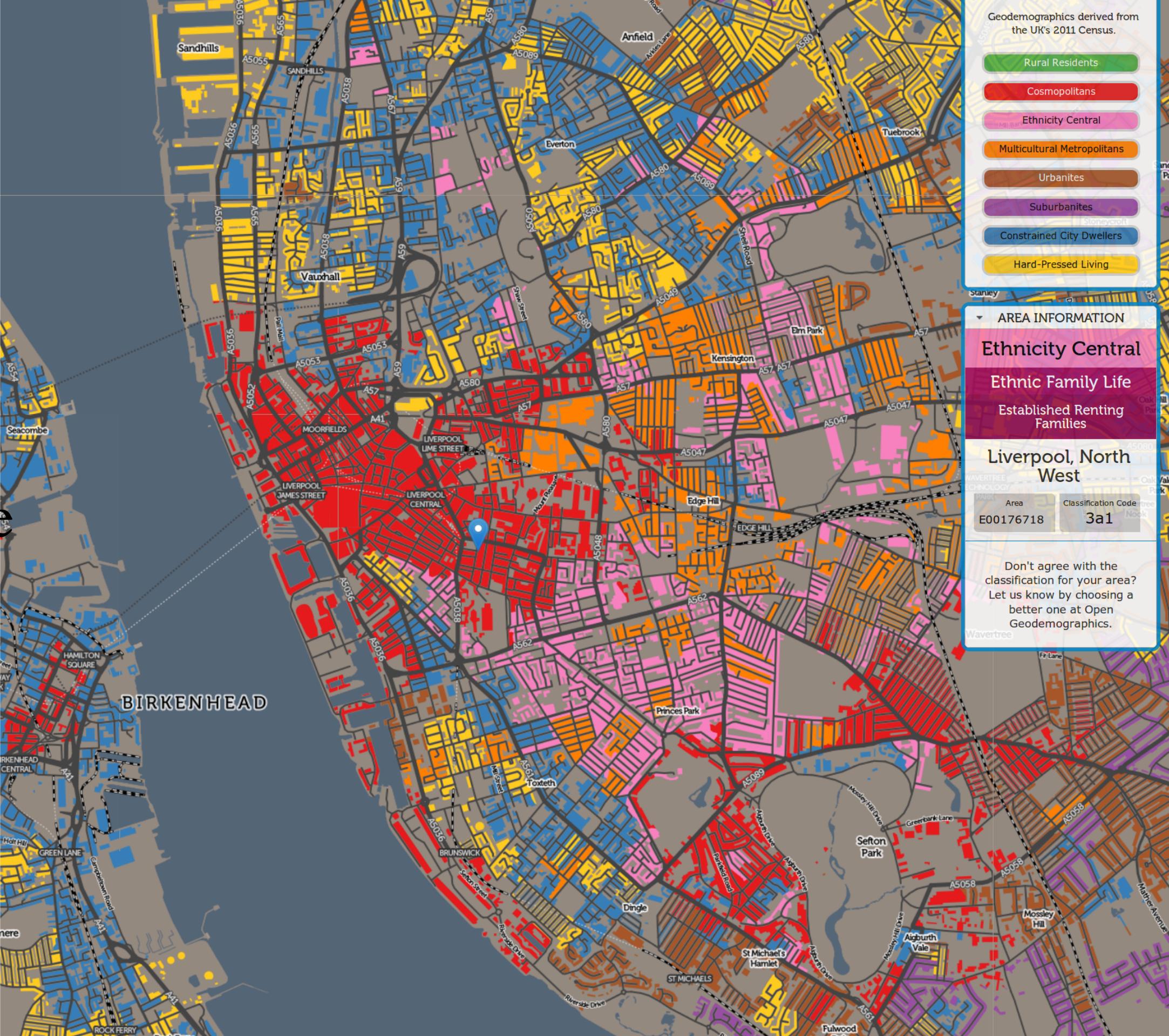
MAP OPTIONS

Land Labels

Toggle: Retail Centres

Download: retail centre locations

Postcode: L19dwl



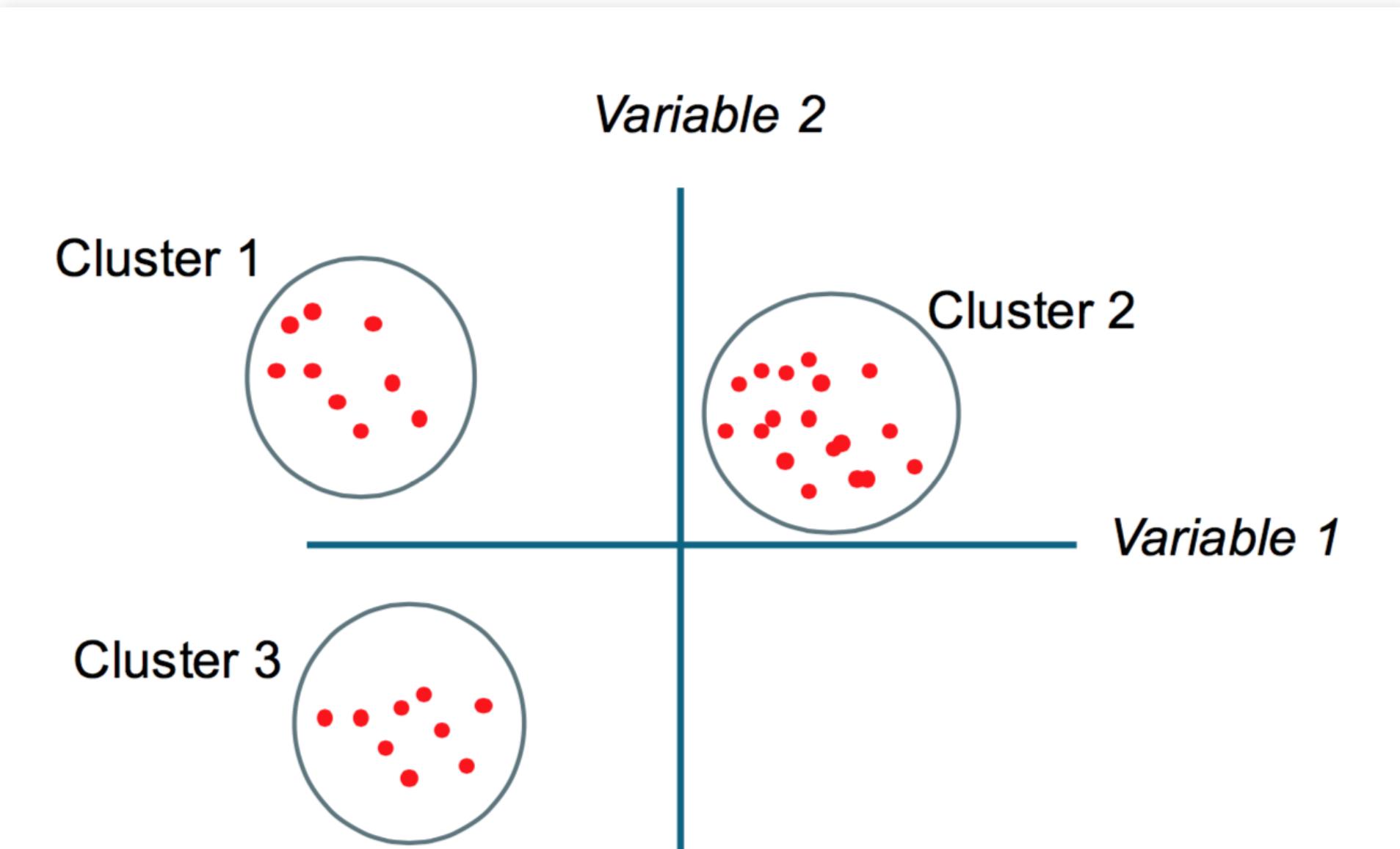
Clustering

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
- Topological Data Analysis
- ...

Different properties, different best usecases

See interesting comparison table

Examples

Frequently Bought Together



Total price: £32.97

Add all three to Basket

i These items are dispatched from and sold by different sellers. [Show details](#)

- This item:** Green and Black's Organic Dark Chocolate 85 Percent Cocoa 100 g (Pack of 5) £11.62 (£2.32 / 100 g)
- [Green and Black's Organic Ginger Dark 100 g \(Pack of 5\) £10.40](#) (£2.08 / 100 g)
- [Green and Black's Organic Dark Chocolate Maya Gold 100 g \(Pack of 5\) £10.95](#) (£2.19 / 100 g)

Customers Who Bought This Item Also Bought



Green and Black's Organic
Ginger Dark 100 g (Pack of
5)
 15
£10.40



Green and Black's Organic
Dark Chocolate Maya Gold
100 g (Pack of 5)
 5
£10.95



Green and Black's Organic
Dark Chocolate 100 g
(Pack of 5)
 22
£8.20



Vivani Organic Dark
Chocolate with 85% Coco
100 g (Pack of 5)
 25
£11.95

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

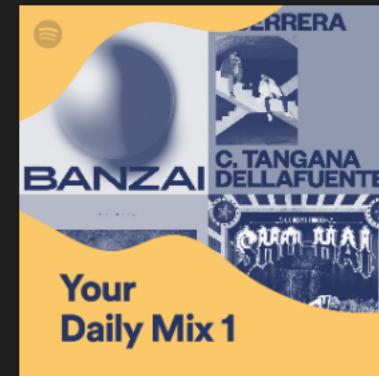
Personal-recommendations

Bookends

New Playlist

Your Daily Mixes

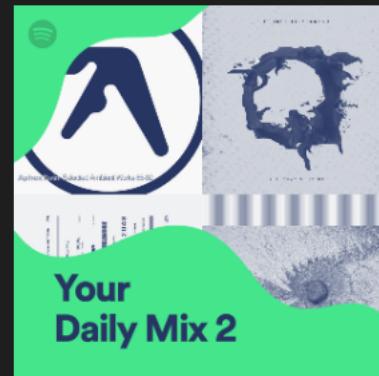
Play the music you love, without the effort. Packed with your favorites and new discoveries.



Daily Mix 1

Gata Cattana, DELLA FUENTE, ToteKing and more

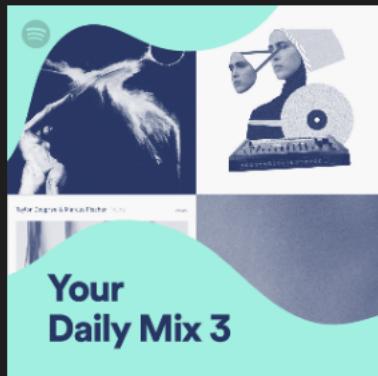
MADE FOR DREAMESSENCE



Daily Mix 2

Aphex Twin, George FitzGerald, Nosaj Thing and more

MADE FOR DREAMESSENCE



Daily Mix 3

Dedekind Cut, Helena Hauff, Taylor Deupree and more

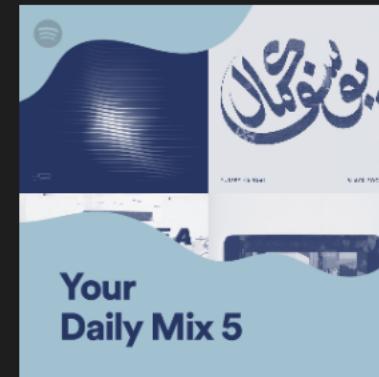
MADE FOR DREAMESSENCE



Daily Mix 4

Berliner Philharmoniker, Alexandre Tharaud, Sir Colin Davis and more

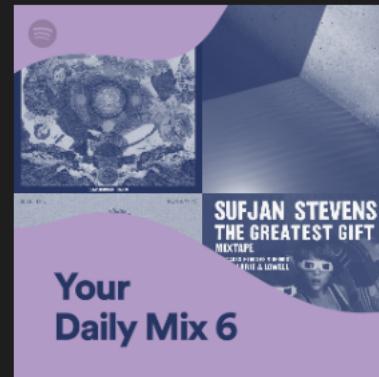
MADE FOR DREAMESSENCE



Daily Mix 5

GoGo Penguin, Yussef Kamaal, Blue Lab Beats and more

MADE FOR DREAMESSENCE



Daily Mix 6

Fleet Foxes, Andrew Bird, Iron & Wine and more

MADE FOR DREAMESSENCE

dreamessence

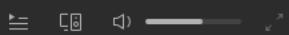
Friend activity

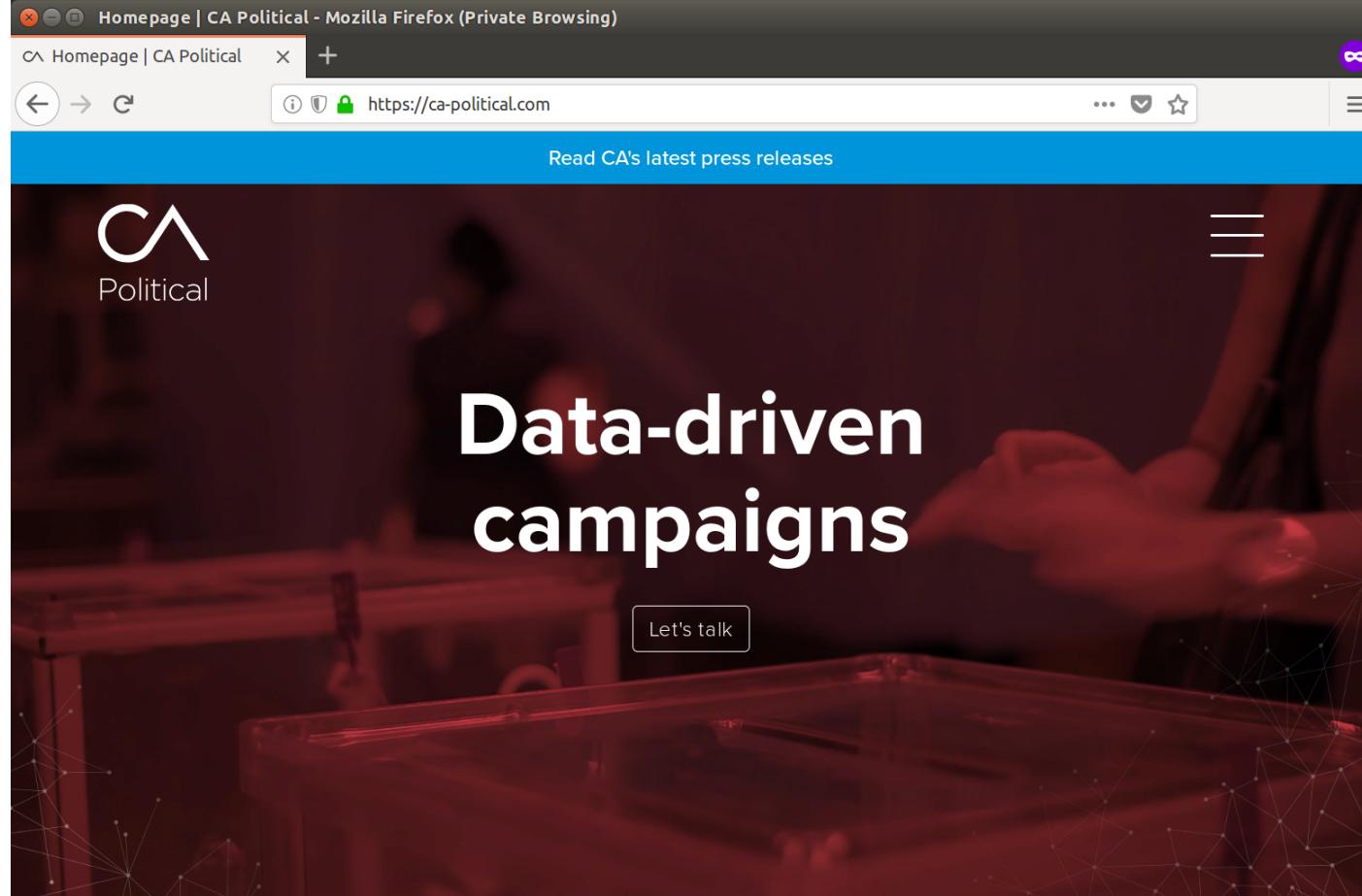
townhallsymphon... ↵
Step Out
José González
⌚ The Secret Life O...

1236163056 6 h
Divine Hammer
The Breeders
⌚ Last Splash

_maxi 10 h
Sensationz of the M...
Mentol Nomad
⌚ Subterranean Hit...

FIND FRIENDS





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.

"There are no longer any experts except Cambridge Analytica."

- Frank Luntz, Political Pollster



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



Unsupervised Learning by Dani Arribas-Bel is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.