Geographic Data Science

The need to group data

Carmen Cabrera and Elisabetta Pietrostefani

The need to group data

Simplicity and complexity

“Everything should be made as simple as possible, but not simpler”

Attributed to Albert Einstein

The need to group data

- The world is complex and multidimensional.
- Univariate analysis considers one dimension.
- Many issues need a multivariate perspective.

Single measures and broader concepts

Percentage foreign-born	**Neighbourhood**
Years of schooling	**Human development**
Monthly income	**Deprivation**

Grouping as simplifying

- Define categories using many characteristics.
- Assign each observation to the category it fits best.
- Reduce complexity while retaining relevant information.
- Produce outputs that are easier to understand.

Types of grouping

- Non-spatial clustering
- Regionalisation

Geographic Data Science

Non-spatial clustering

Non-spatial clustering

Grouping by attributes

Split a dataset into groups using a set of attributes.

Within groups

Observations are similar.

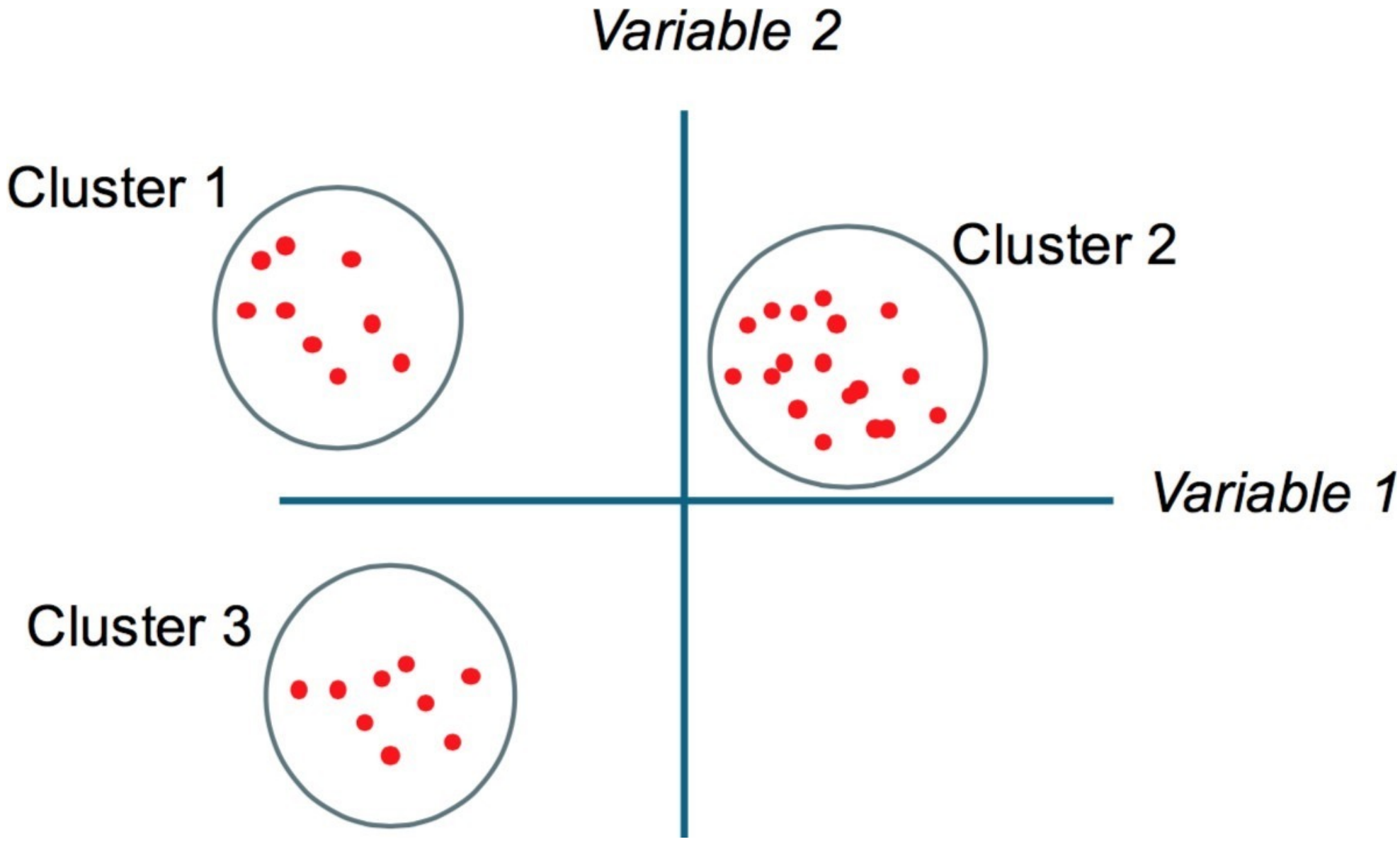
Between groups

Observations are dissimilar.

Machine learning

Unsupervised

Intuition



k-means

- The most popular clustering algorithm
- Useful, with limitations

More clustering methods

- Hierarchical clustering
- Agglomerative clustering
- Spectral clustering
- Neural networks, e.g. self-organising maps
- DBSCAN

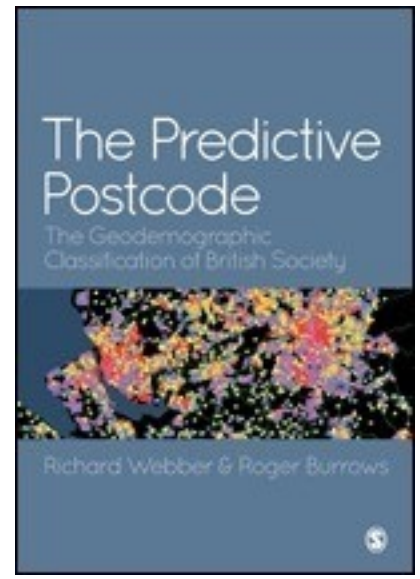
Different properties suit different use cases.

Geodemographic analysis

Geodemographic analysis

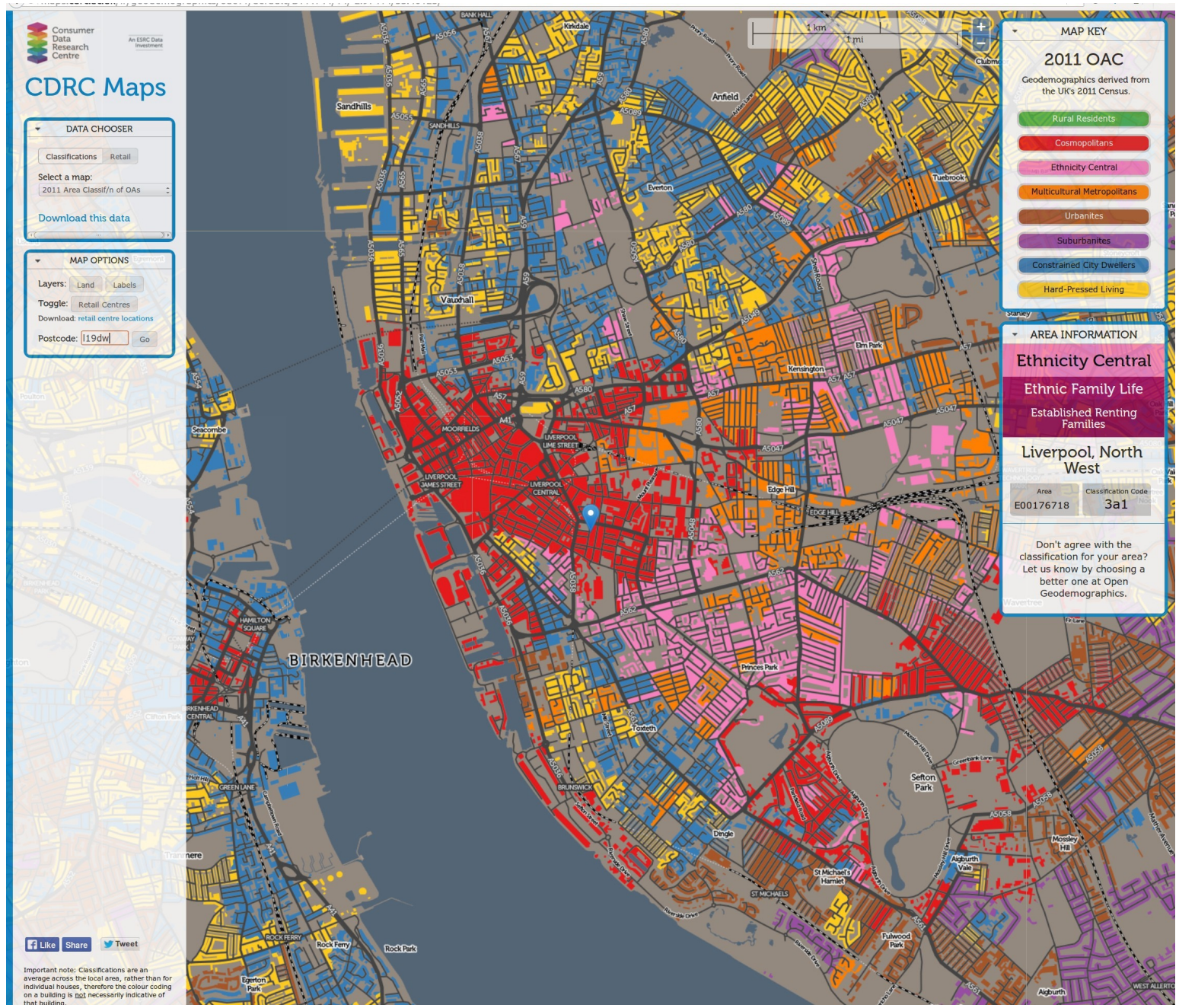
1970s: Richard Webber

- Identify similar neighbourhoods.
- Target funding for urban deprivation.



Applications expanded from public sector policy to private sector marketing and business intelligence.

Internet user classification



CDRC

Regionalisation

Unsupervised spatial machine learning

Aggregate basic spatial units (areas) into larger units (regions).

Regionalisation

Group observations using a set of attributes, with similarity within groups and differences between groups.

Spatial constraint

Observations in the same region must be spatially connected neighbours.

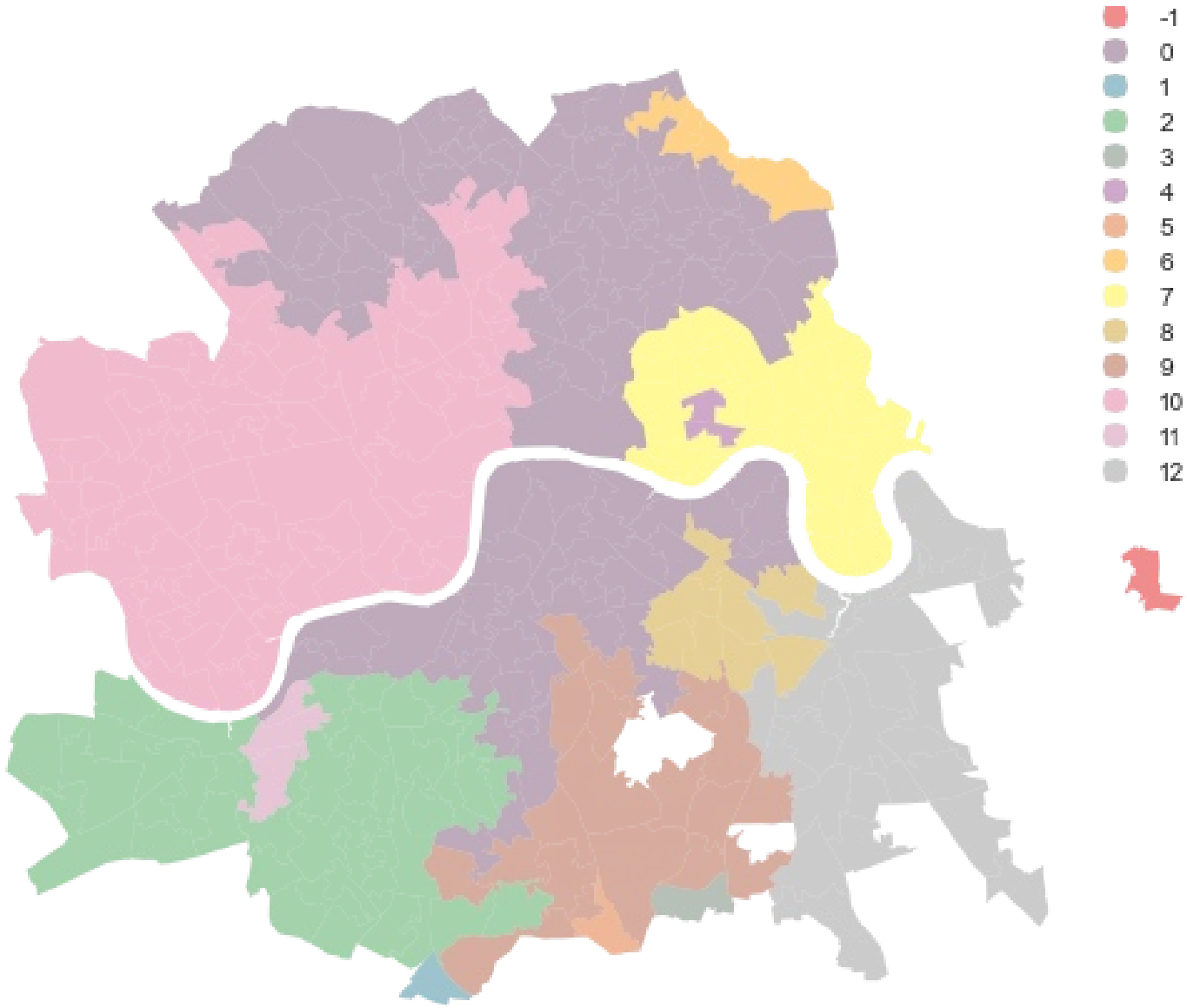
Regionalisation

Regionalisation: defining constraints

- Aggregate areas into a predefined number of regions, optimising an aggregation criterion.
- Areas within each region must be geographically connected.
- The number of regions cannot exceed the number of areas.
- Assign each area to exactly one region.
- Each region must contain at least one area.

Duque et al. (2007)

Regionalisation: map example



Regionalisation algorithms

- Automated Zoning Procedure (AZP)
- Max-P
- Other regionalisation methods

Duque et al. (2007) provide an advanced overview.

Examples

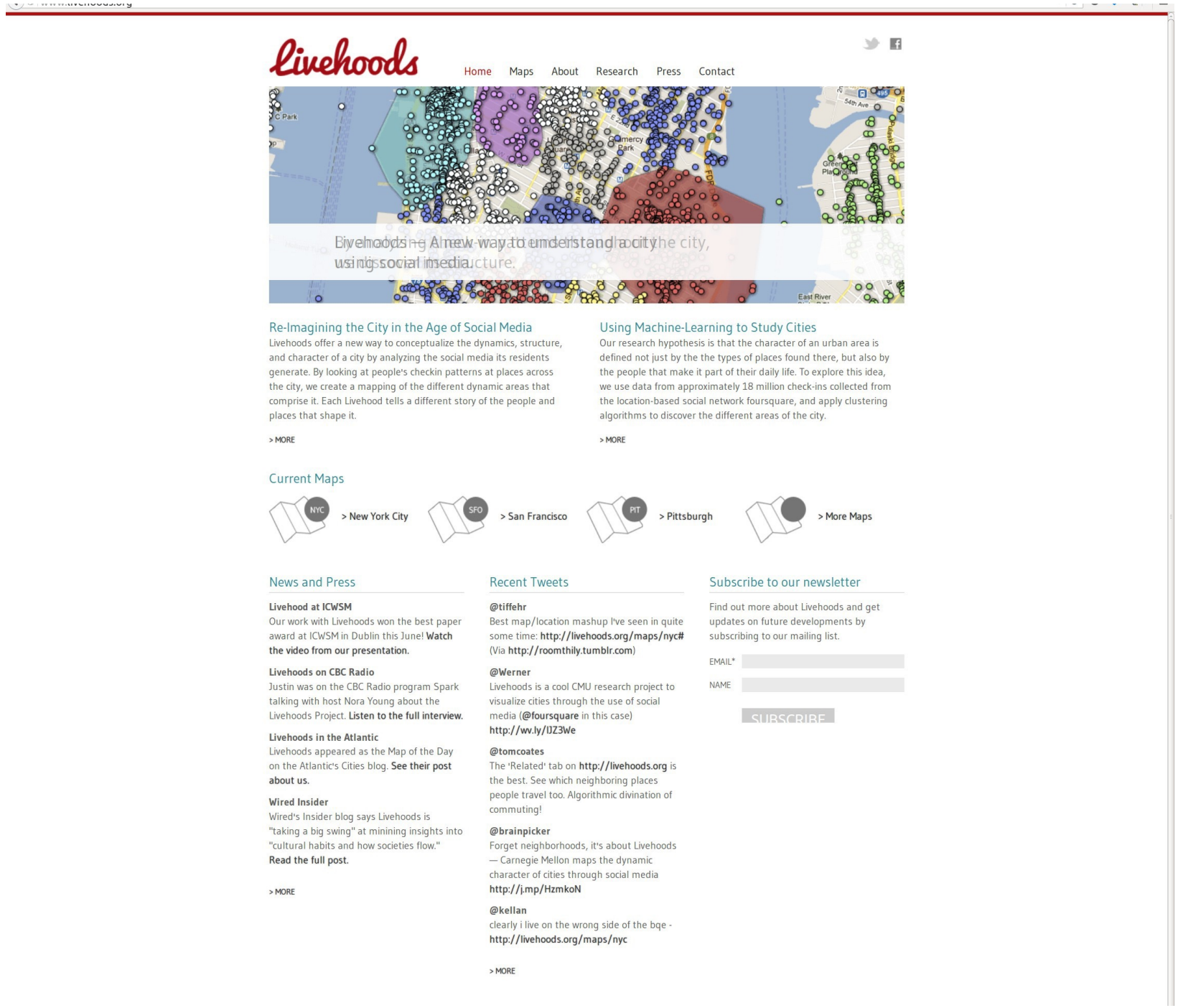
Census geographies

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

Livehoods


Livehoods
Home
Maps
About
Research
Press
Contact
Livehoods – A new way to understand a 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 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
Current Maps
NYC
> New York City
SFO
> San Francisco
PIT
> 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 minining 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

Attribution and licence

These slides were originally created by
Dani Arribas-Bel for a course in Geographic Data Science.

Creative Commons Attribution-ShareAlike 4.0 International Licence

