

Sentimental Airbnb

Ajaya Adhikari
Ramy Al Sharif
Lorenzo Gasparini
Vasileios Milius
Ioannis Papadopoulos
(Group 16)

Pipeline

Data Collection



Geolocated Tweets



Extraction



Sentiment



WordCloud

Data Collection



Geolocated Prices



Spatial Clustering



1. Data Collection & Processing



Streaming &
Storing



- English
- Geolocated

Sentiment per
neighborhood?



TextBlob



Sentiment

Word-of-mouth per
neighborhood?



Text Processing

- Removing stopwords
- Removing words \leq 2 letters
- Aggregation



WordCloud



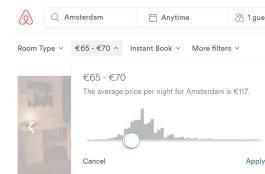
Need for geolocated accommodations



Normalized Geolocated Prices



8000



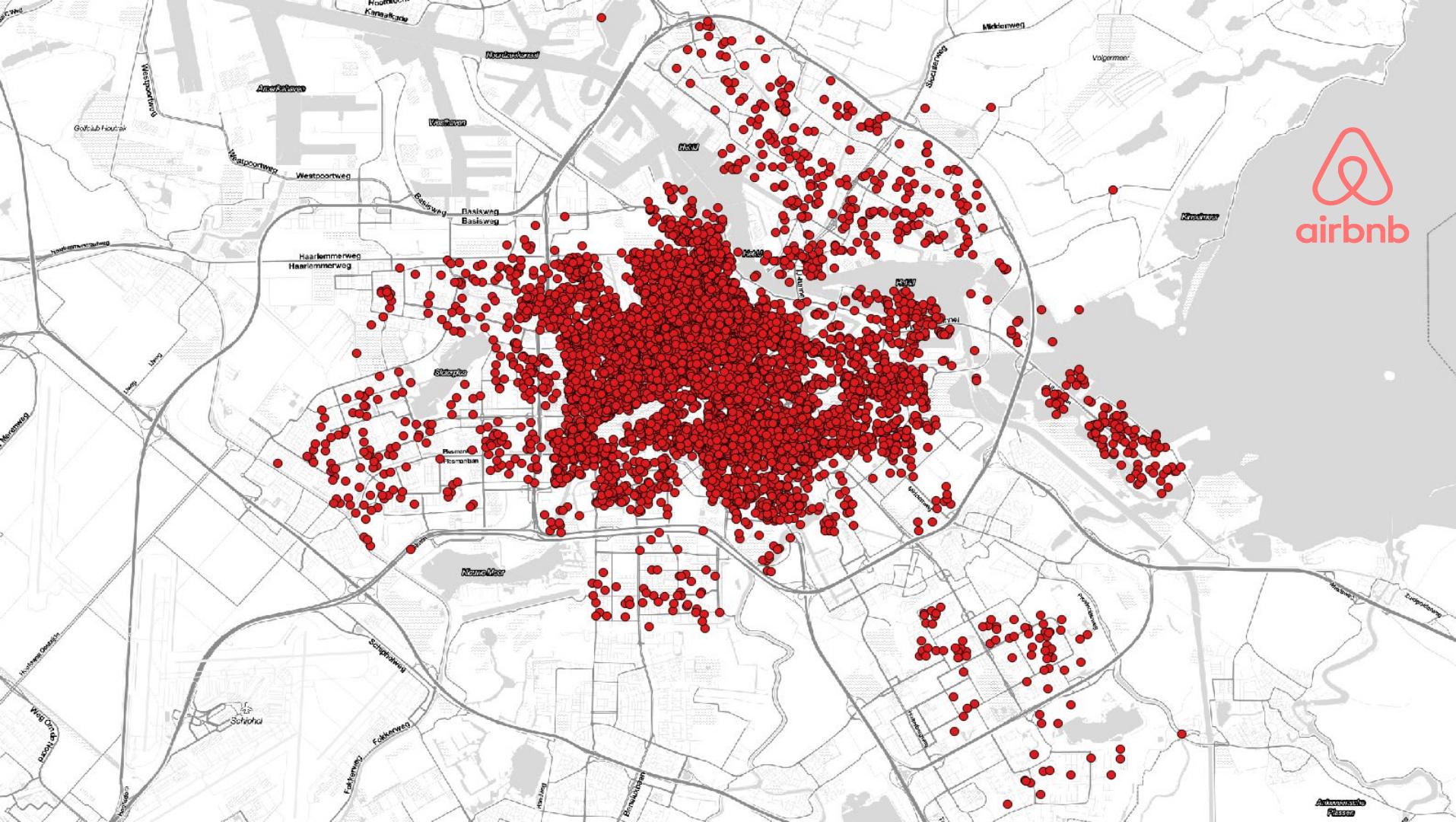
1 2 3 ... 17 >

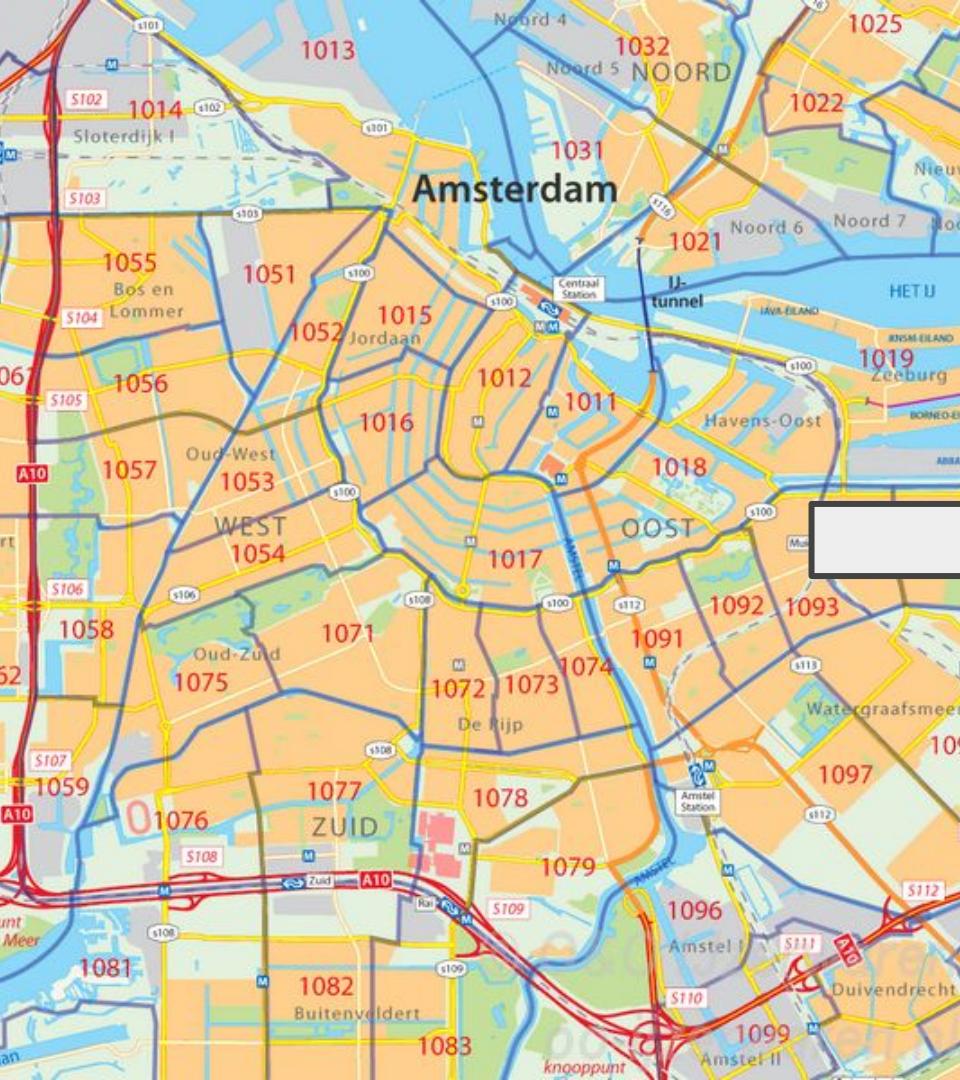
1 – 18 of 300+ Rentals



Limitations of the number of results

2. Clustering

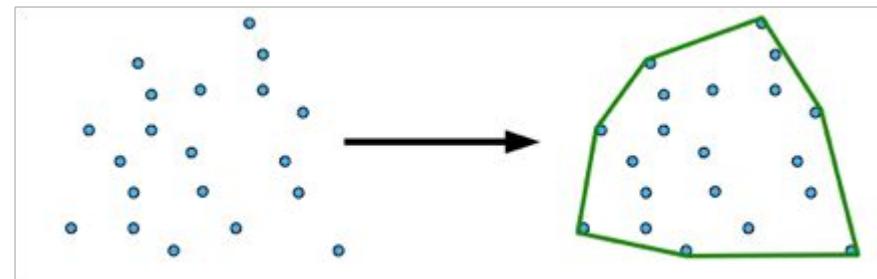


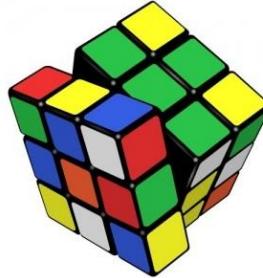
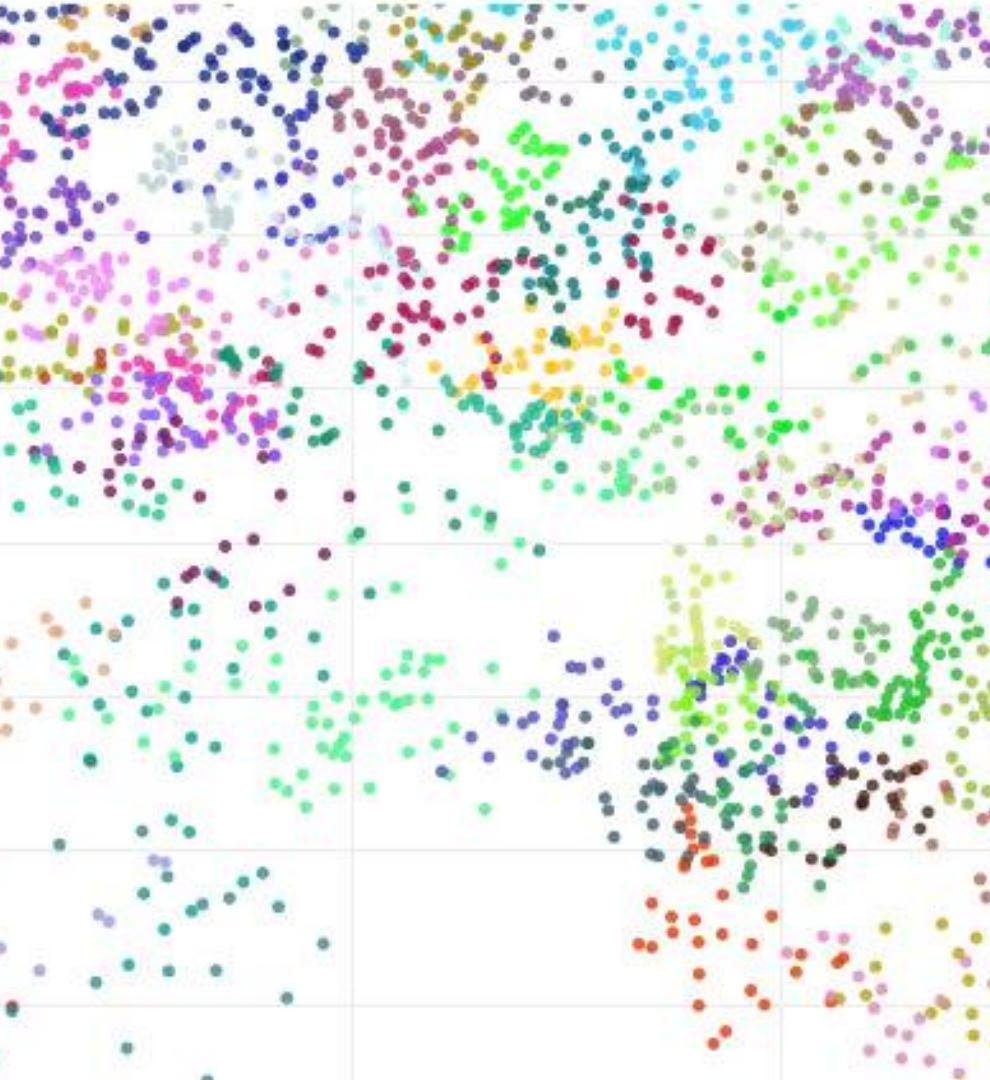


K Nearest Neighbors

+

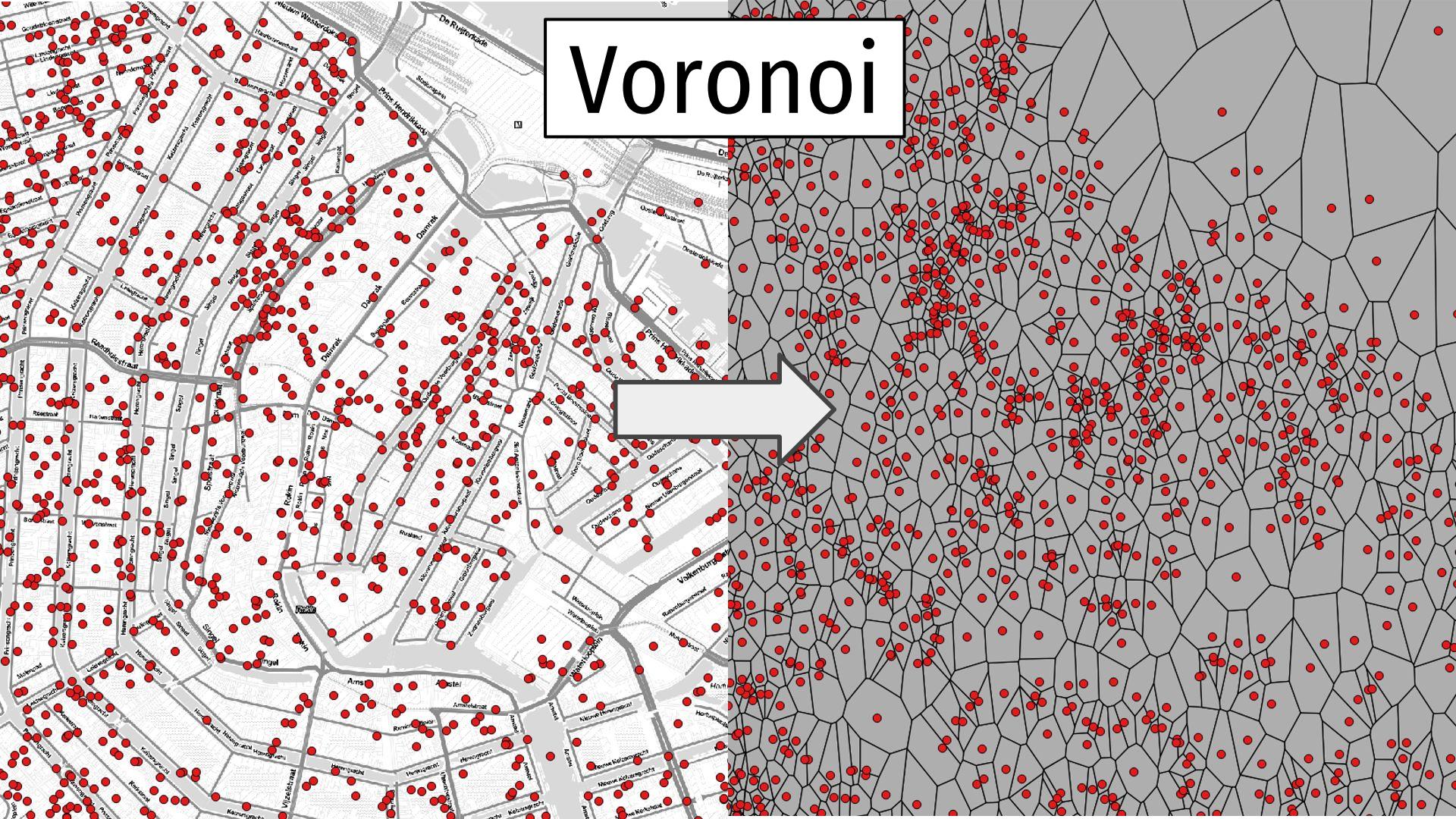
Convex hull





Overlapping
Convex- Hull
Regions

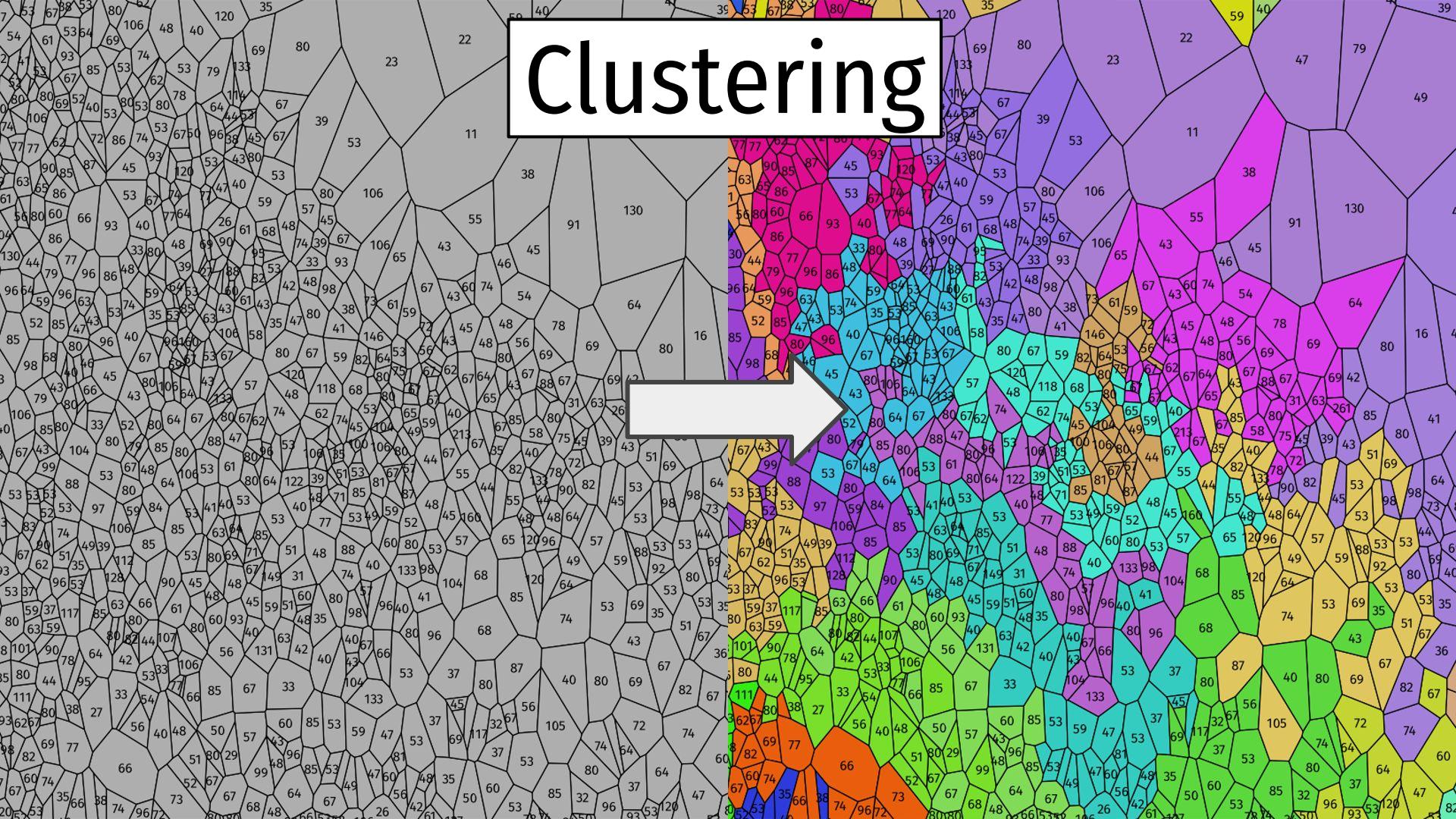
Voronoi



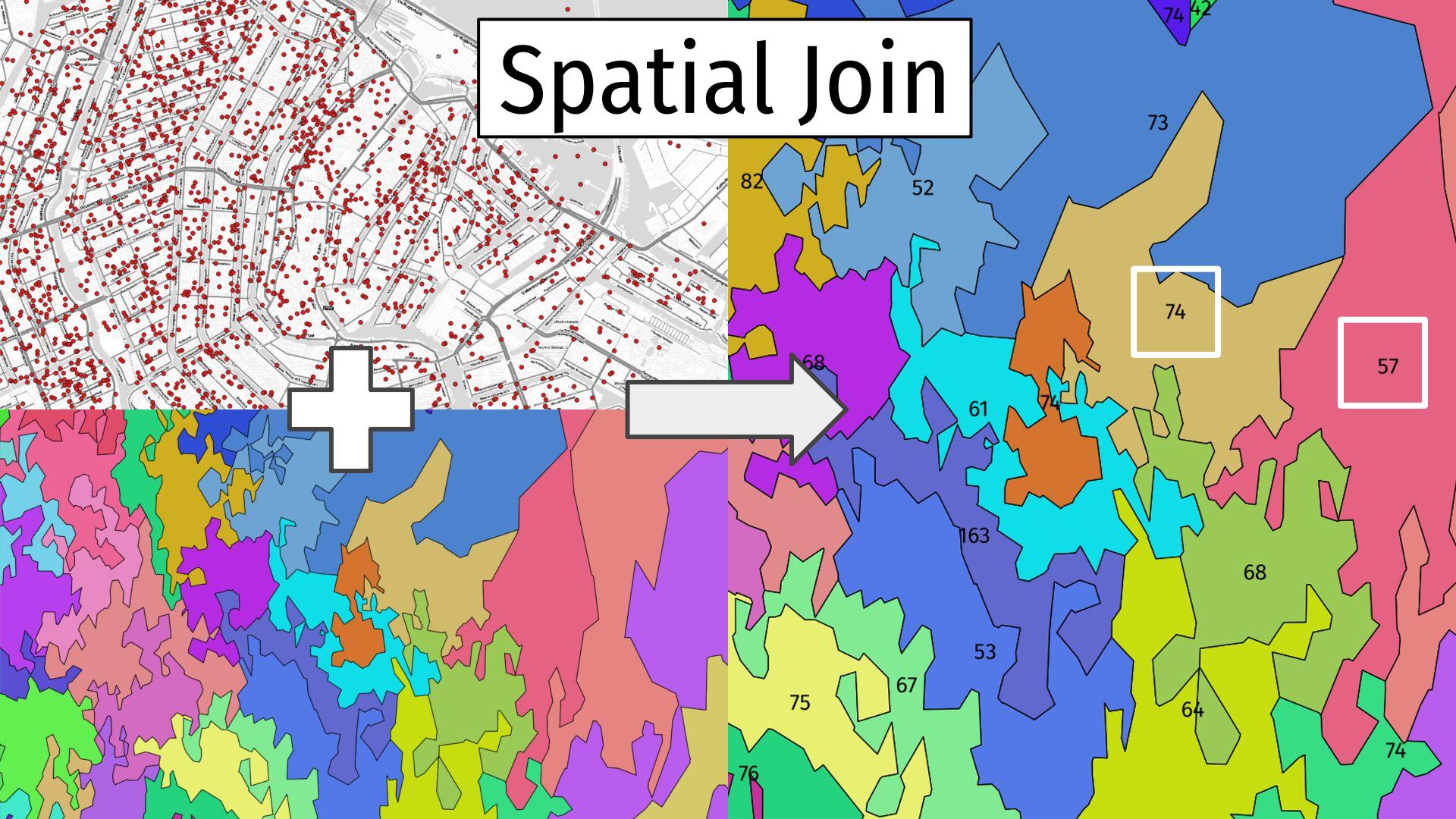
Max-p regions

- Finds maximum number of heterogeneous regions
 - Airbnb clusters
- The data determines the number of regions and the shape
 - Normalized airbnb prices
- Constraint per region
 - Sum of capacity greater than 150

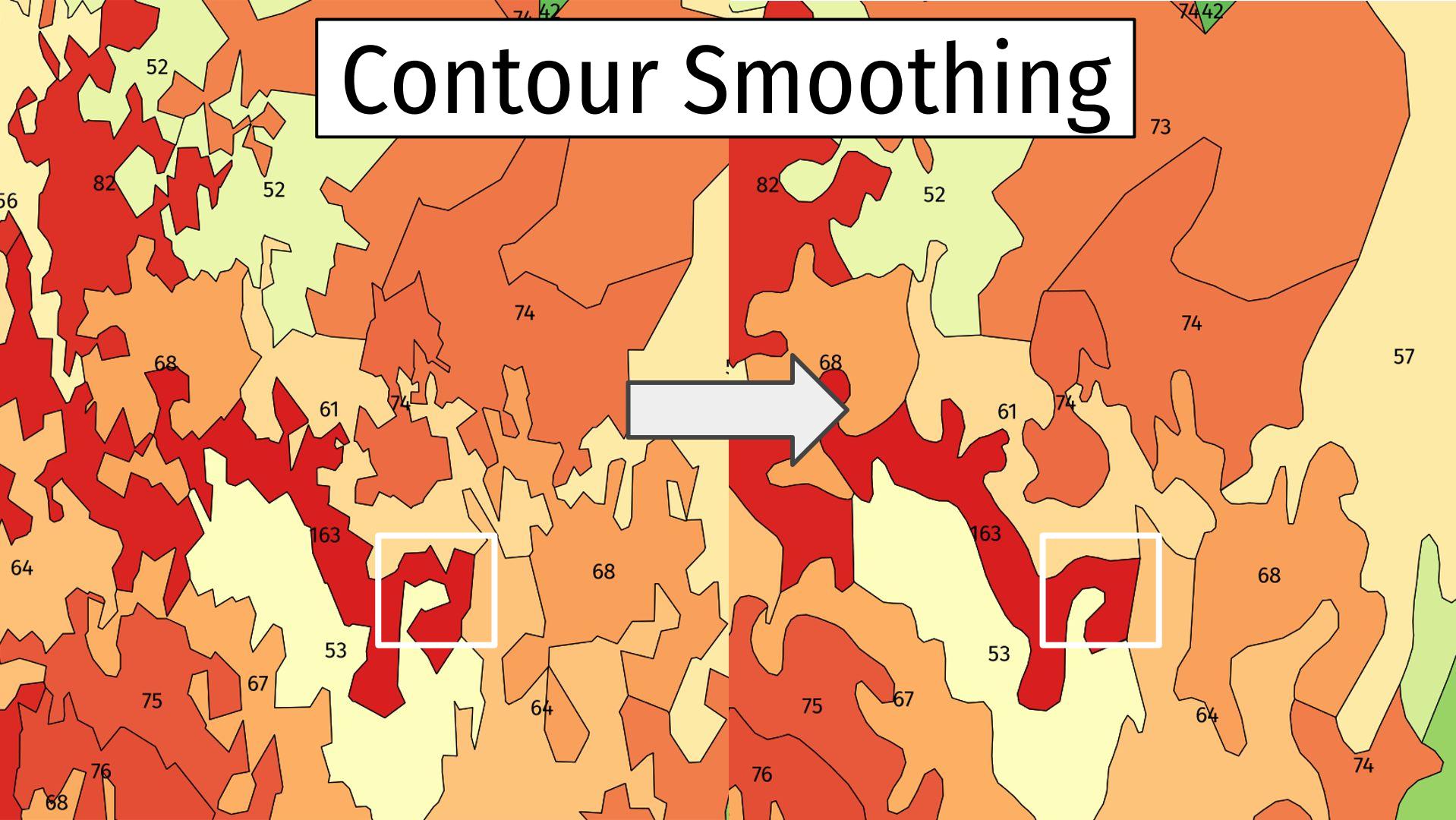
Clustering



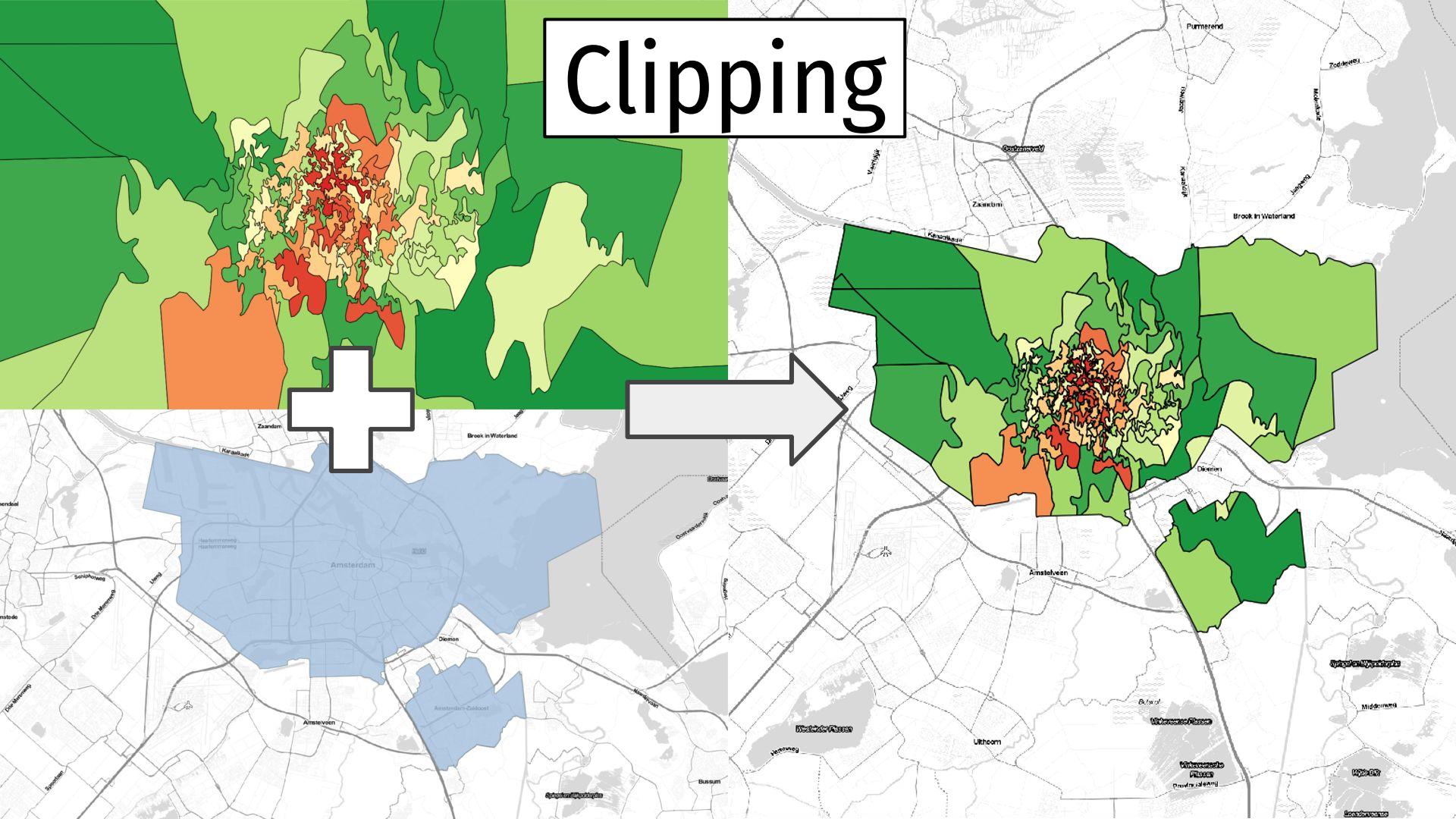
Spatial Join

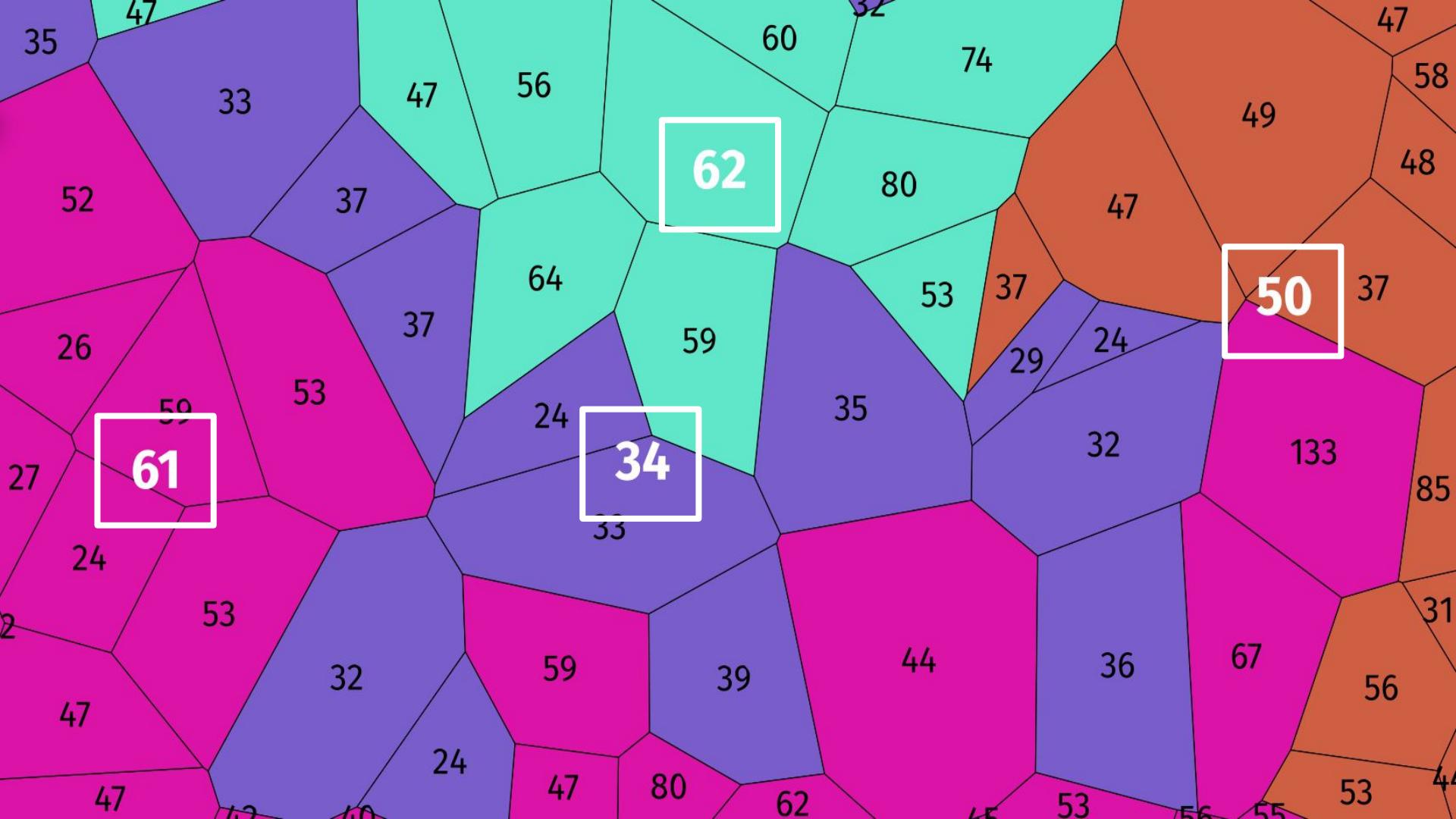


Contour Smoothing

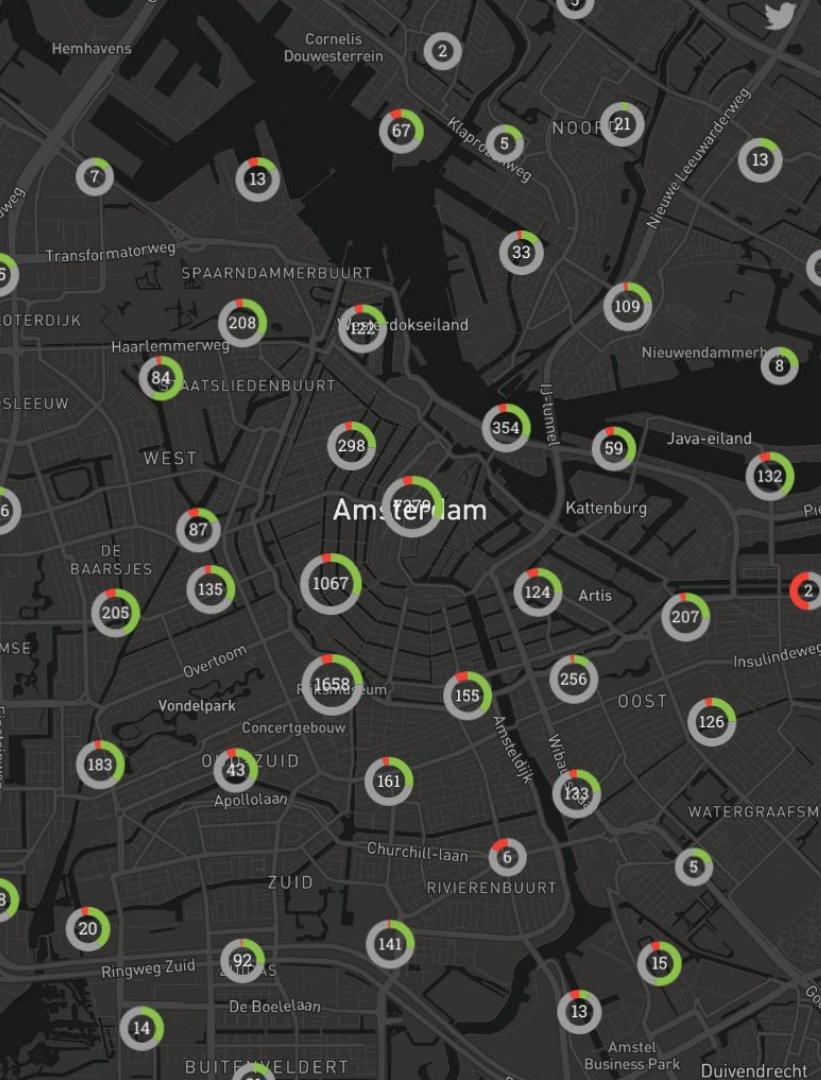


Clipping

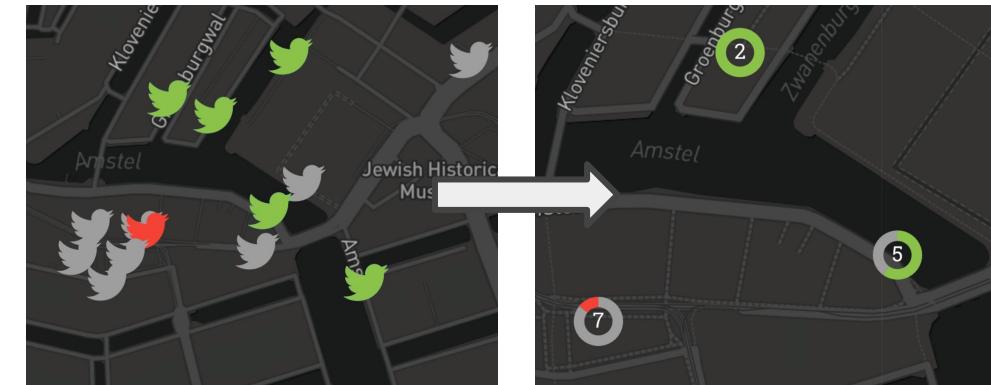




3. Visualization

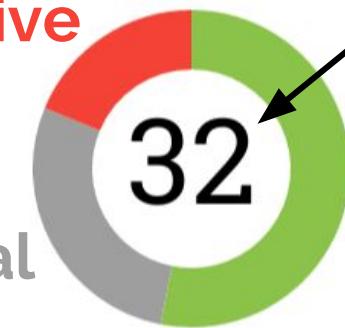


Tweets Clusters



% Negative

% Neutral



Number
of Tweets

% Positive

Word Cloud

- Pops up when the user clicks on a point
 - Displays the most used words in the tweets contained in the blue radius
 - More frequent words are shown with a bigger font size

*Colours are purely aesthetic

Live demo

4. Conclusion



Our Vision

The Near Future



Instagram

pararius.nl



FOURSQUARE

Our Team





Thank you.