Exploring vernacular perceptions of spatial entities: Using Twitter data and R for delimiting vague, informal neighbourhoods in Inner London, UK.

Luke Clasper
UNIGIS UK
ltclasper@aol.com
ltclasper.net

Talk Structure

- What is vernacular geography?
- How to capture vernacular geography?
- What is Ambient Geographic Information (AGI)?
- R statistical software.
- Data collection from Twitter.
- Qualitative coding.
- Results and vernacular neighbourhood delimitation.
- Elements forming basis of vernacular neighbourhoods.
- Source platforms of Tweets.
- Limitations of Study and Further Work.

What is Vernacular Geography?

- Folksonomy and the informal, bottom-up, unofficial, colloquial nature of how people discuss and mentally conceive geographic place and structure.
- Individual's awareness of fuzzy, abstract geographic regions (official and unofficial) in relation to their own location.
- Important applications:
  - Emergency Services
  - Deliveries
  - In-vehicle Navigation
  - Allocations of Services and Census

How to Capture Vernacular Geography?

- Difficult: personal, casual, vague, qualitative.
- Traditional methods like questionnaires and sketch maps.
- Web scraping.
- Flickr.
- AIG (Ambient Geographic Information) from social media like Twitter can collect many responses from unconscious participants.

What is AGI? (See *et al*., 2016)

- Passively volunteered data.
- An example of crowdsourced geographic information.
- Participants are unconsciously involved in the study.
- Contributors of the data are often the focus of the study.
- Contributors of the data also act as sensors for observable phenomena.
- AGI can help us study behaviour and patterns in social systems.

R Statistical Software

- Open Source and free of charge.
- Language and software environment.
- Spatial capabilities.
- Research reproducibility and self documentation.
- Efficient with large datasets and repeating tasks.
- OSM (and other) base maps.
- Thousands of code libraries.
- Active community.
- Steep learning curve but rewarding.

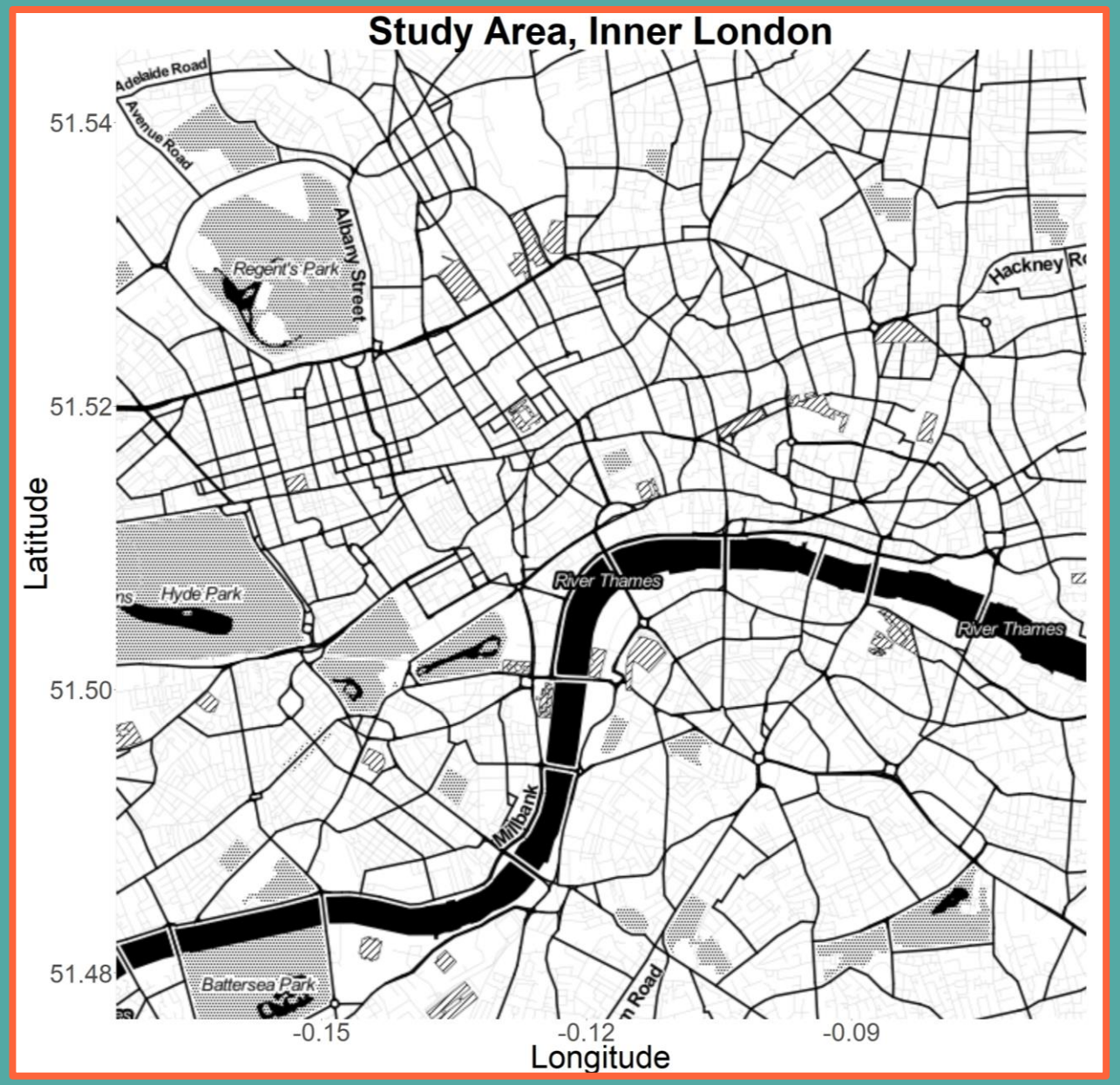
R Libraries Used

- rgdal (GDAL)
- ggmap (basemaps)
- GISTools (spatial analysis)
- sp (spatial methods)
- spatstat (point pattern analysis)
- hexbin (hexagonal binning)
- aspace (spatial point patterns)
- ggplot2 (graphics, visualisations)
- ggthemes (style themes)
- twitteR (connects to Twitter API)
- Httr (HTTP web data)
- RColorBrewer (colour schemes)
- geosphere (spherical trigonometry)
- reshape (data aggregation)
- gganimate (ggplot animation)

Neighbourhoods for Study

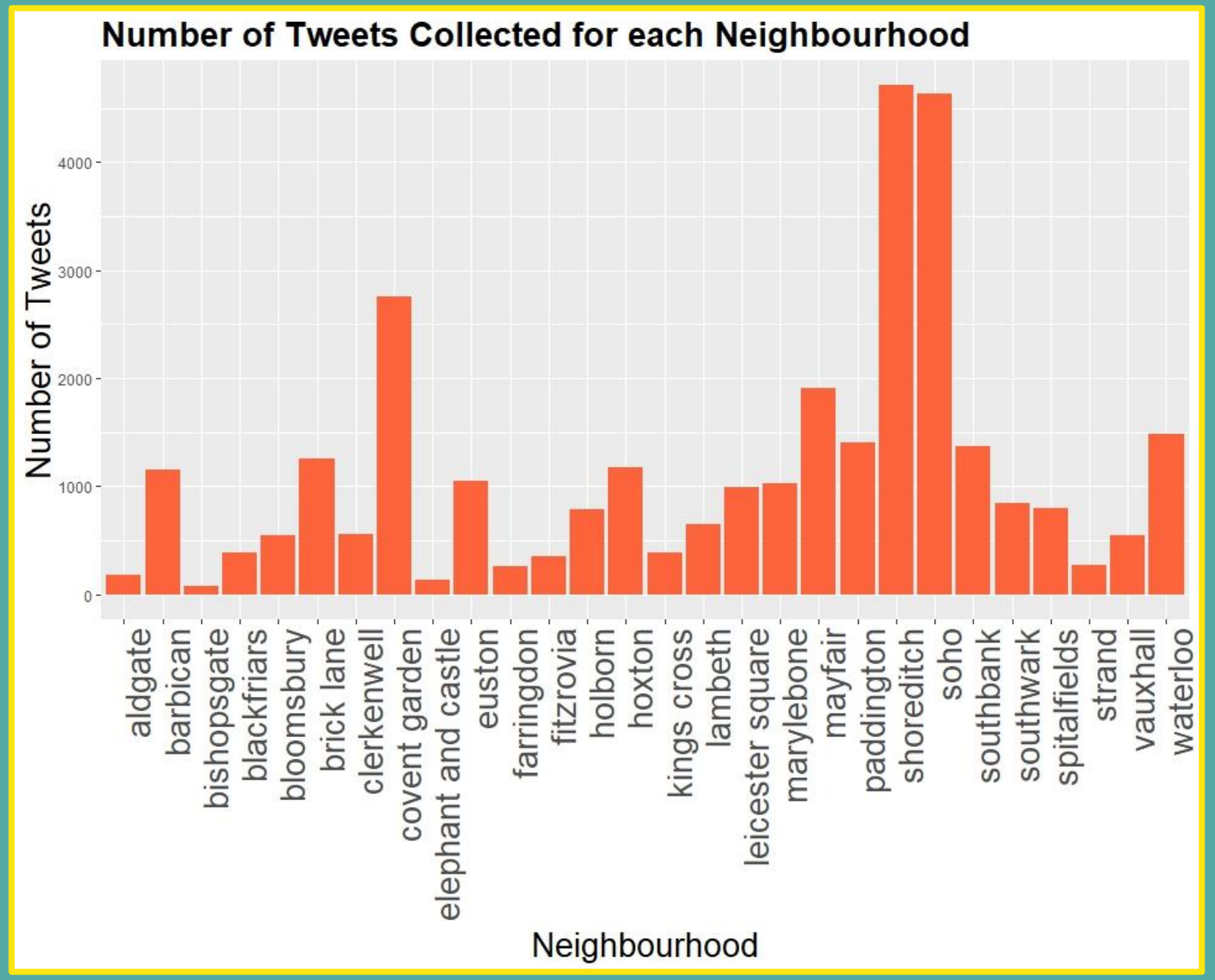
Neighbourhood has current official boundary? (Y/N).

- Aldgate (Y)
- Barbican (N)
- Bishopsgate (Y)
- Blackfriars (N)
- Bloomsbury (Y)
- Brick Lane (N)
- Clerkenwell (Y)
- Covent Garden (Y)
- Elephant and Castle (N)
- Euston (N)
- Farringdon (Y)
- Fitzrovia (N)
- Holborn (Y)
- Hoxton (Y)
- Kings Cross (Y)
- Lambeth (Y)
- Leicester Square (N)
- Marylebone
- Mayfair (N)
- Paddington (N)
- Shoreditch (N)
- Soho (N)
- Southbank (N)
- Southwark (Y)
- Spitalfields (Y)
- Strand (N)
- Vauxhall (N)
- Waterloo (N)

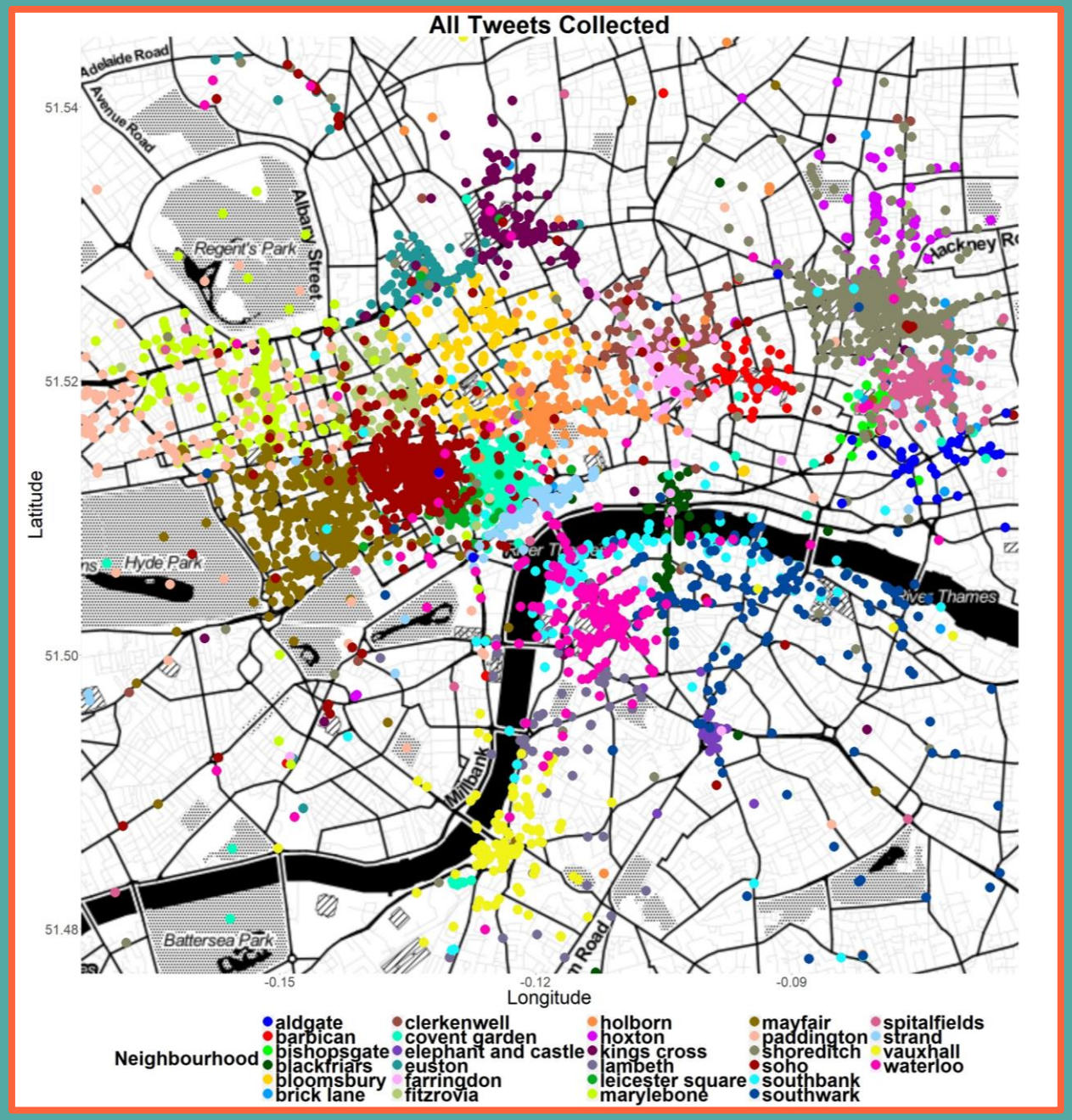


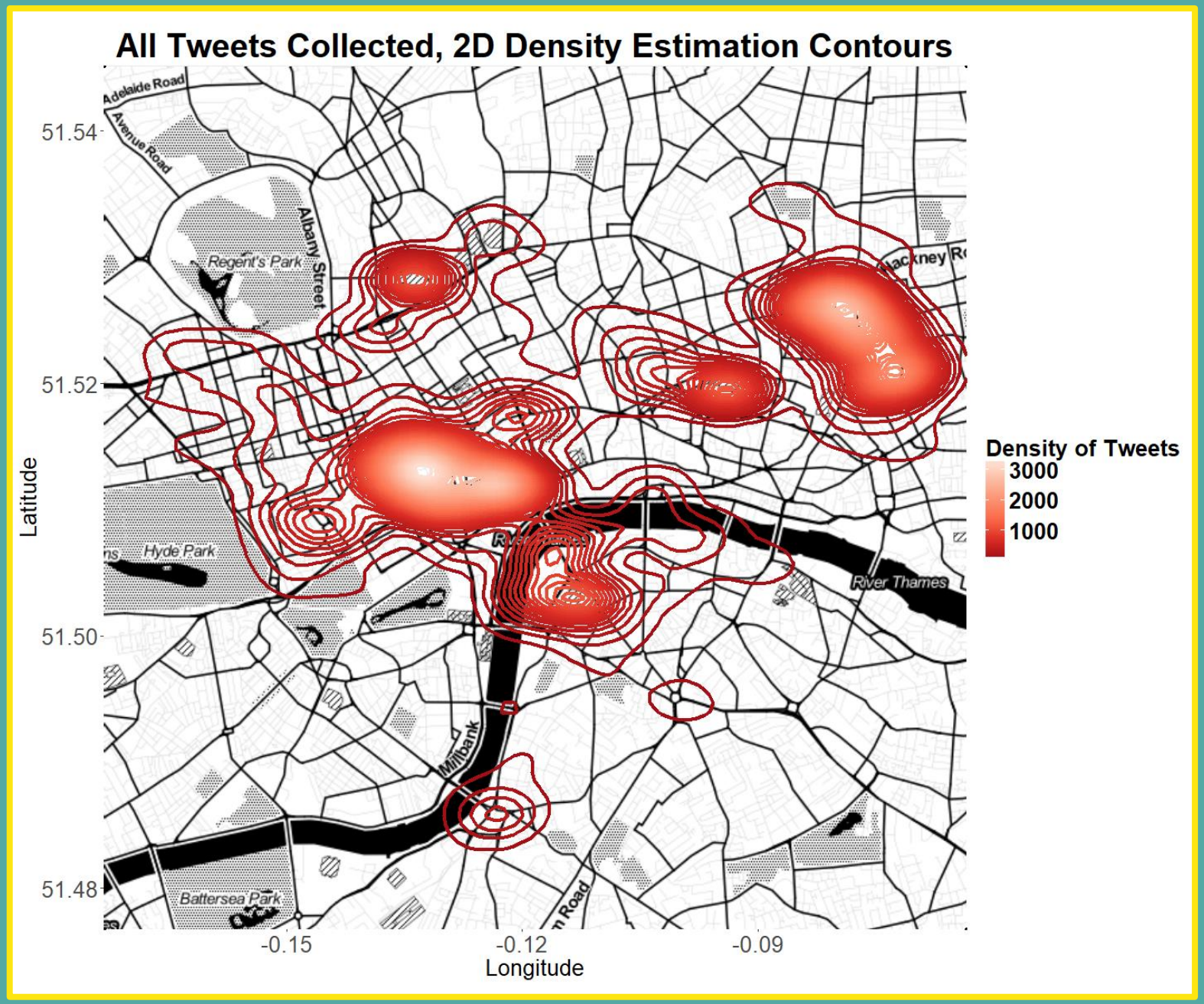
Data Collection from the Twitter API

- twitteR library.
- setup_twitter_oauth() make authorised connection.
- API key, API token, API secret from Twitter.
- searchTwitter() for:
  - Looped keywords, Hastags e.g. shoreditch and #Shoreditch
  - Geocode Radius (5 miles)
  - Geocode Centroid (51.508107, -0.126449)


Number of Tweets Collected for each Neighbourhood
Number of Tweets
4000
3000
2000
1000
0
aldgate
barbican
bishopsgate
blackfriars
bloomsbury
brick lane
clerkenwell
covent garden
elephant and castle
euston
farringdon
fitzrovia
holborn
hoxton
kings cross
lambeth
leicester square
marylebone
mayfair
paddington
shoreditch
soho
southbank
southwark
spitalfields
strand
vauxhall
waterloo
Neighbourhood

- Tweets mirroring social activity.
- Dense Tweets in West End and East End.
- City of London, Westminster and residential areas have sparse Tweets.
- Tweets replicating underlying urban structure.
- Neighbourhood clusters forming from keyword searches.




All Tweets Collected, 2D Density Estimation Contours
Adelaide Road
Avenue Road
Albany Street
Regent's Park
Hyde Park
River Thames
Millbank
Battersea Park
51.54
51.52
51.50
51.48
Latitude
-0.15
-0.12
-0.09
Longitude
Density of Tweets
3000
2000
1000

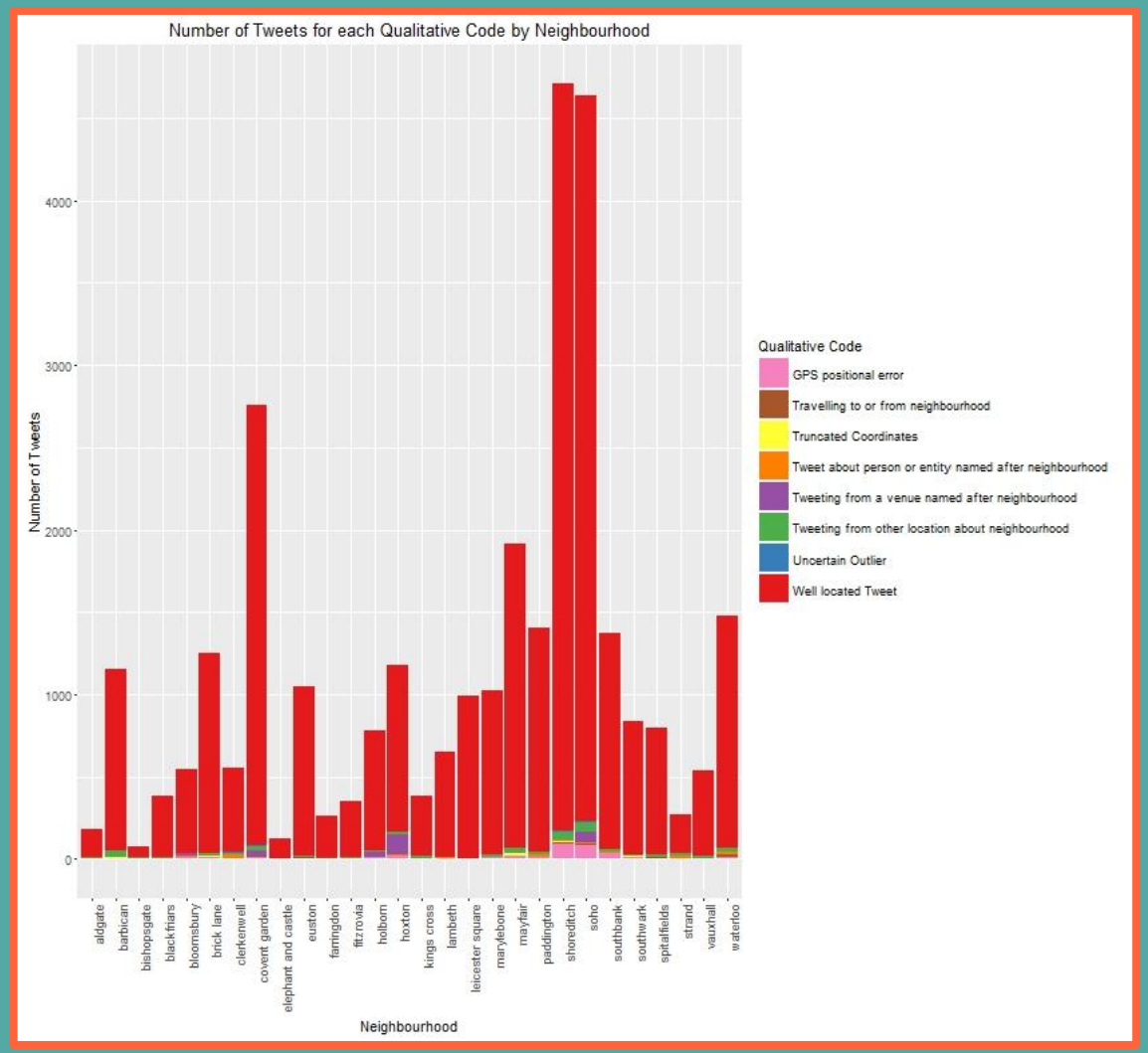
Qualitative Coding

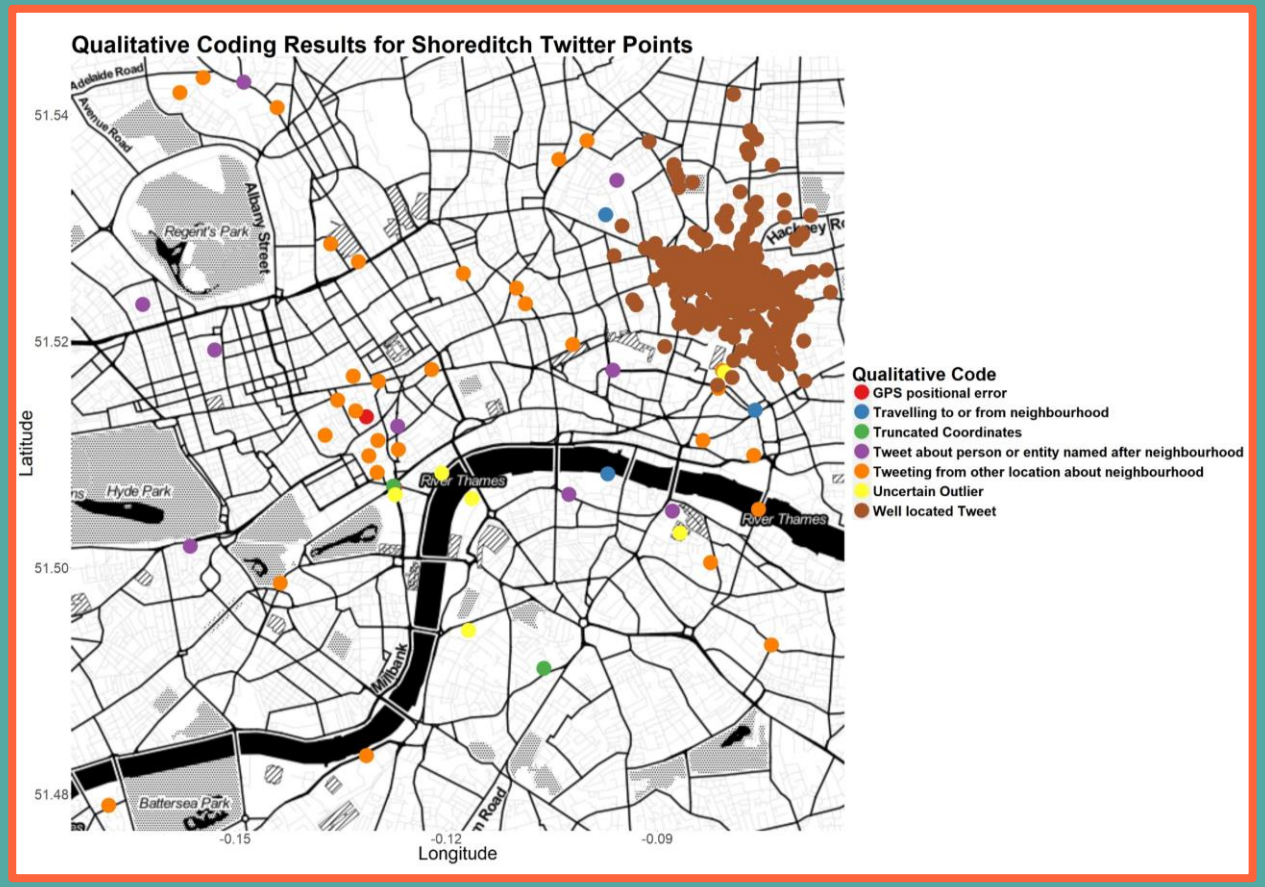
- Social media data is unreliable (Lovelace *et al*, 2016).
- AGI from social media is qualitative.
- Qualitative coding used to improve data quality.
  - Manual examination of text topic
  - Visual analysis for precision
- A way of interpreting and filtering data in order to categorise or classify it into themes (Cope, 2003).

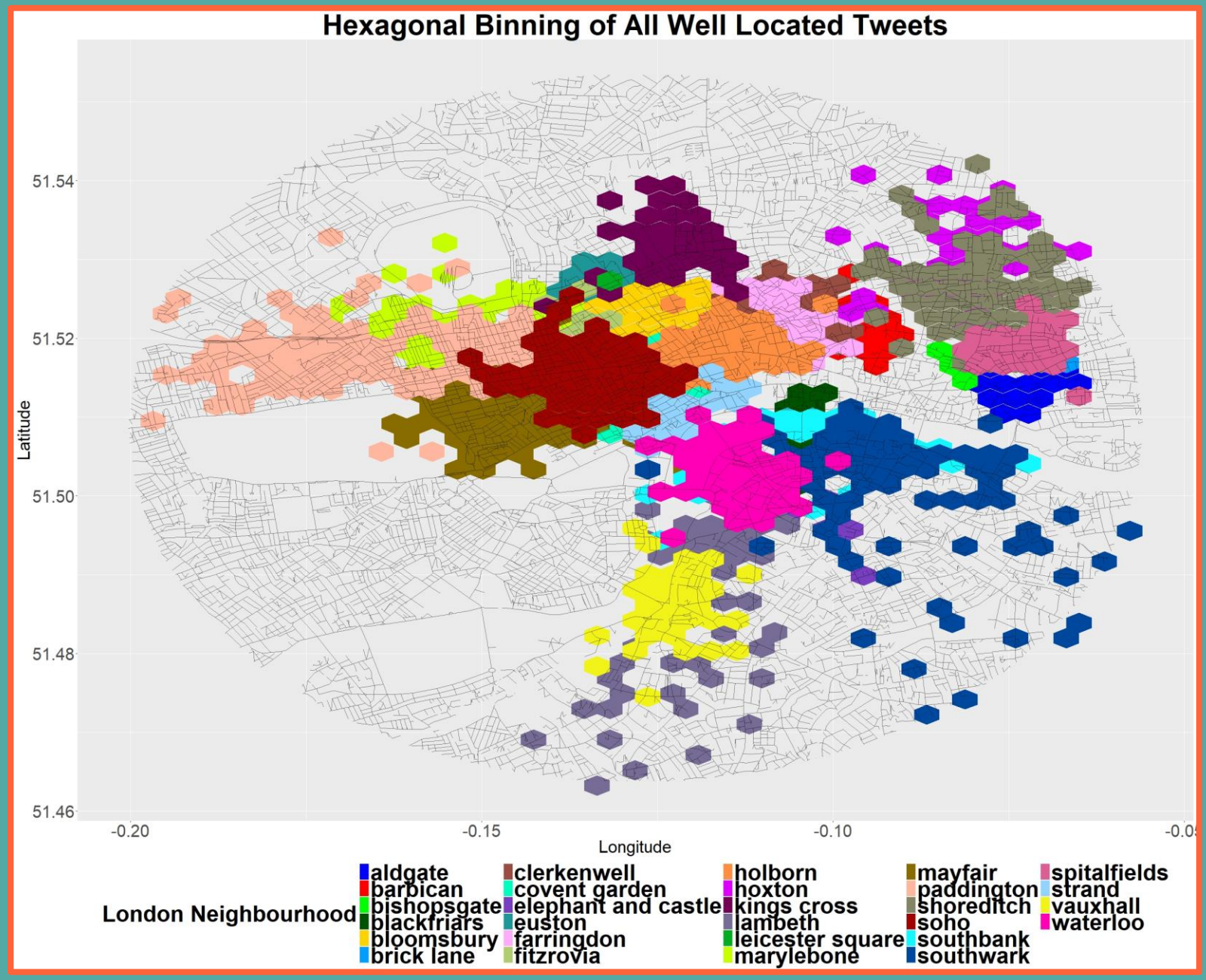
Qualitative Coding Categories

- Travelling to or from neighbourhood.
  - "Made it as far as Covent Garden en-route to Soho, gotta experience the gay night life here in London… "
- Tweeting from other locational about neighbourhood.
  - "Still feeling stuffed after yesterday's meal at @BodeansBBQ in soho. huge massive portions, can't wait to go again."
- Tweeting from a venue named after neighbourhood.
  - "Gym on a Sunday, that's how committed I am to undoing the holiday damage! (@ Soho Gyms Farringdon in London, UK)"
- Well located Tweet.
  - "I'm at Gail's Artisan Bakery in Soho"
- GPS Positional Error.
- Tweet about person or entity named after neighbourhood.
  - "Paul Strand 1890- 1976 arguably one of the greatest documentary photographers of 20th century"
- Truncated Coordinates.
- Uncertain Outlier.

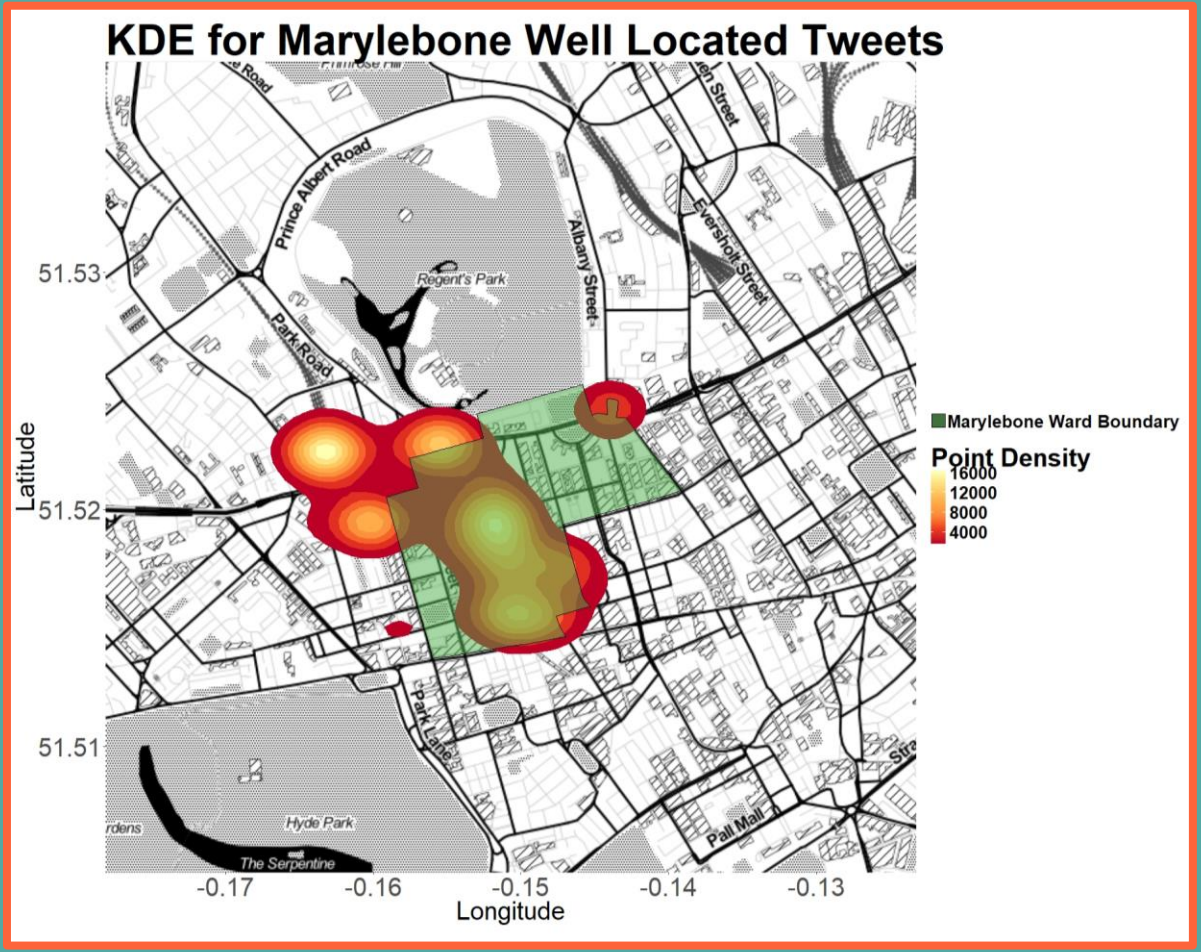
Qualitative Coding Results

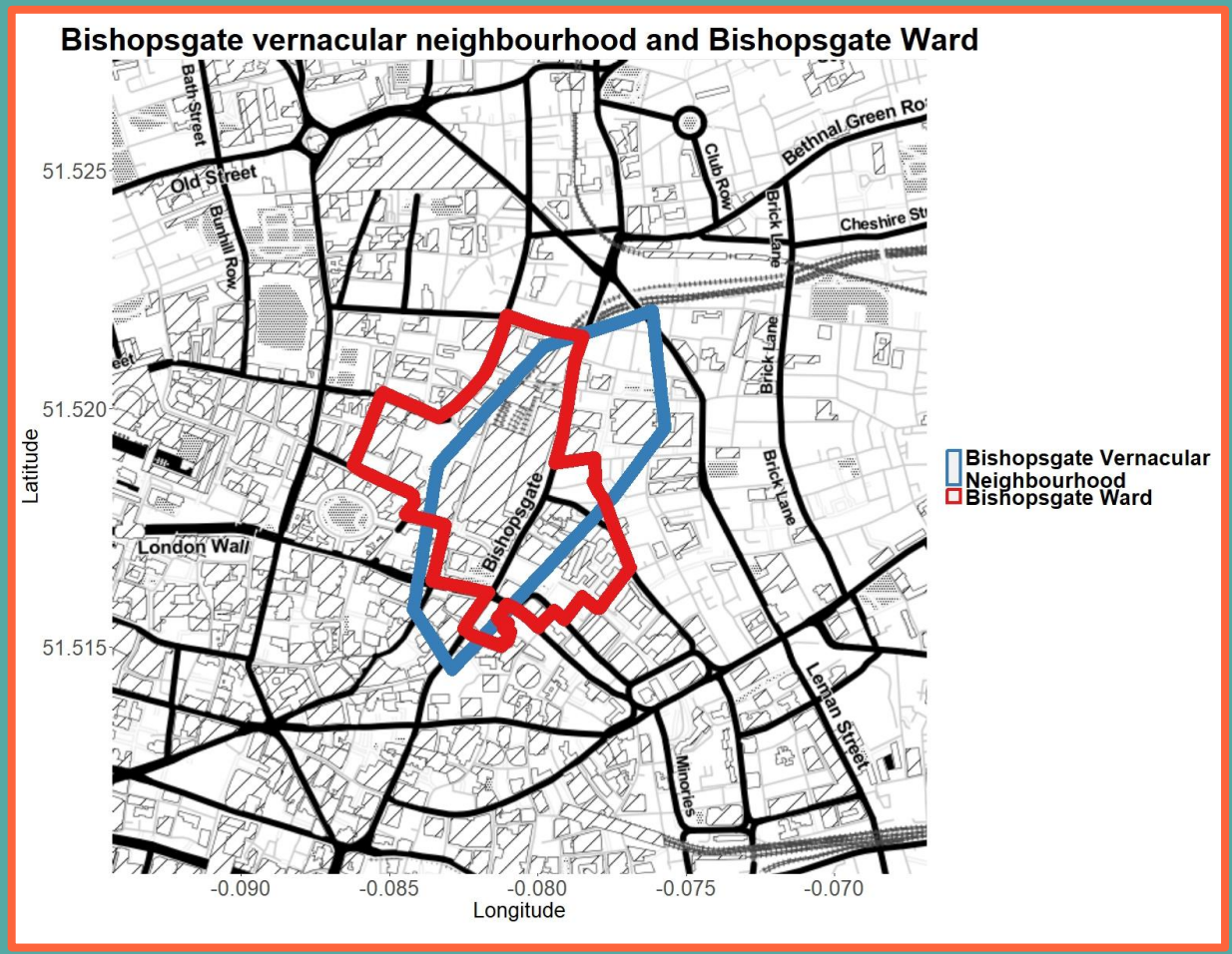


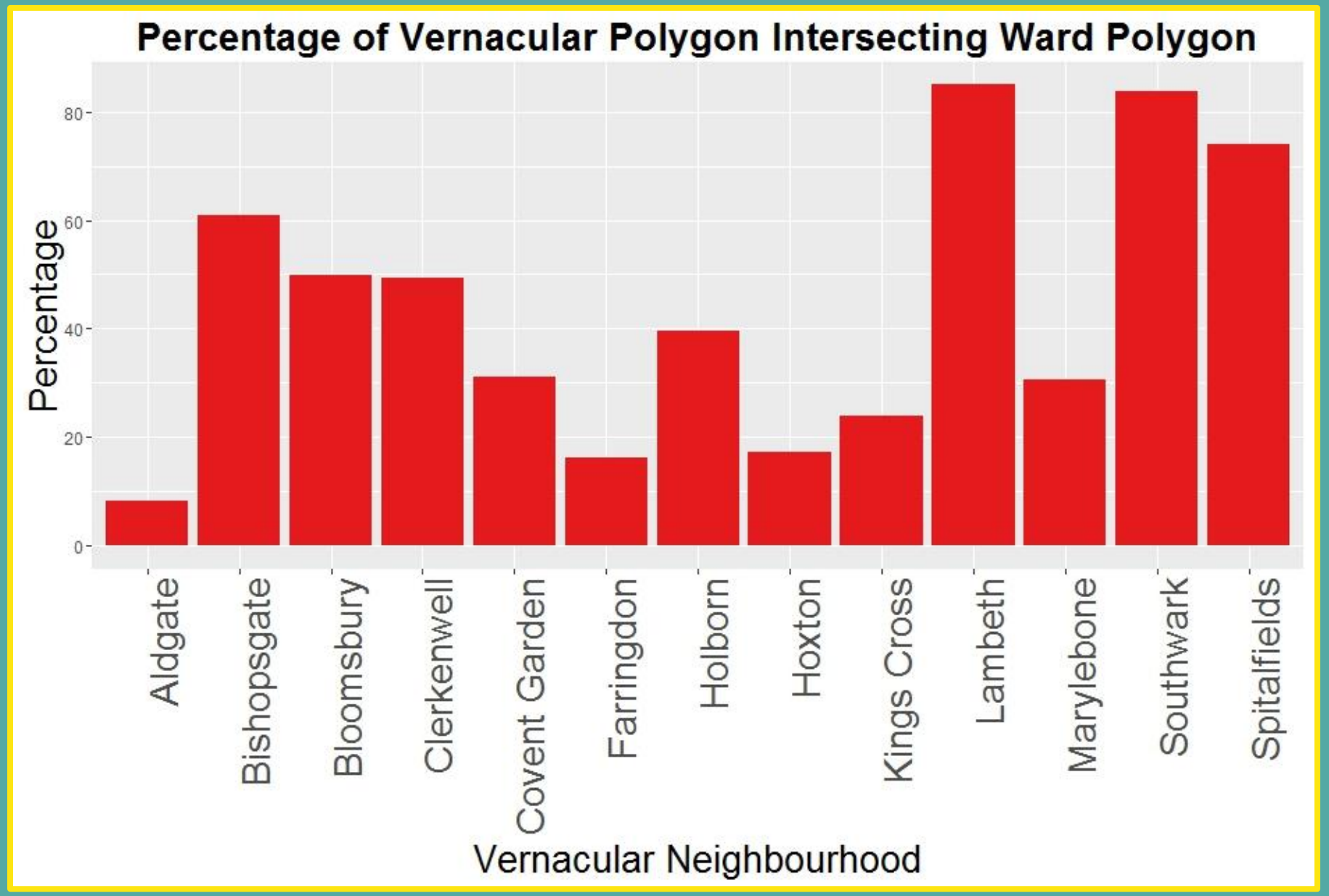



Hexagonal Binning of All Well Located Tweets
51.54
51.52
51.50
51.48
51.46
Latitude
-0.20
-0.15
-0.10
Longitude
London Neighbourhood
aldgate
barbican
bishopsgate
blackfriars
bloomsbury
brick lane
clerkenwell
covent garden
elephant and castle
euston
farringdon
fitzrovia
holborn
hoxton
kings cross
lambeth
leicester square
marylebone
mayfair
paddington
shoreditch
soho
southbank
southwark
spitalfields
strand
vauxhall
waterloo

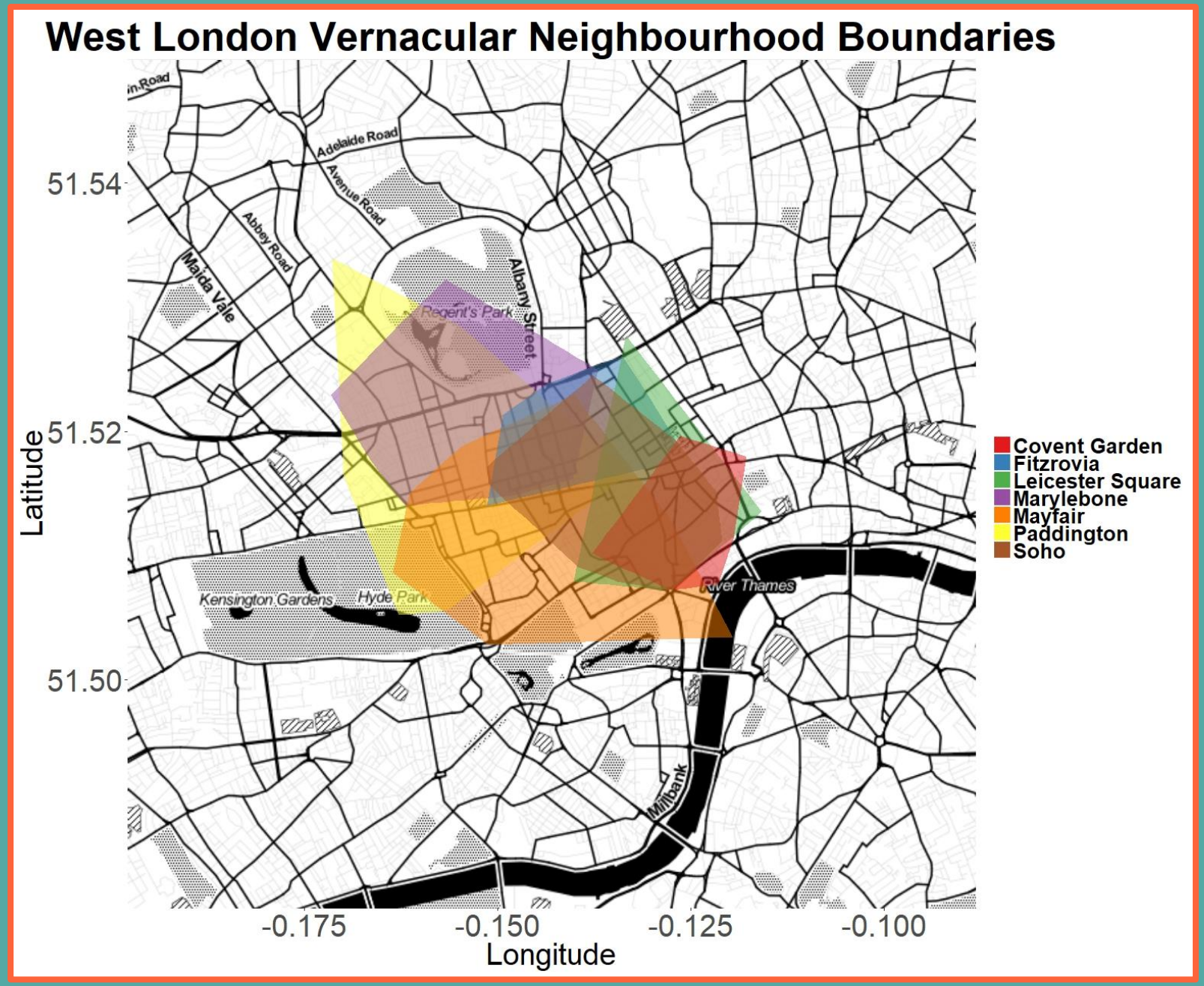
Neighbourhood Comparison to Official Administrative Boundaries



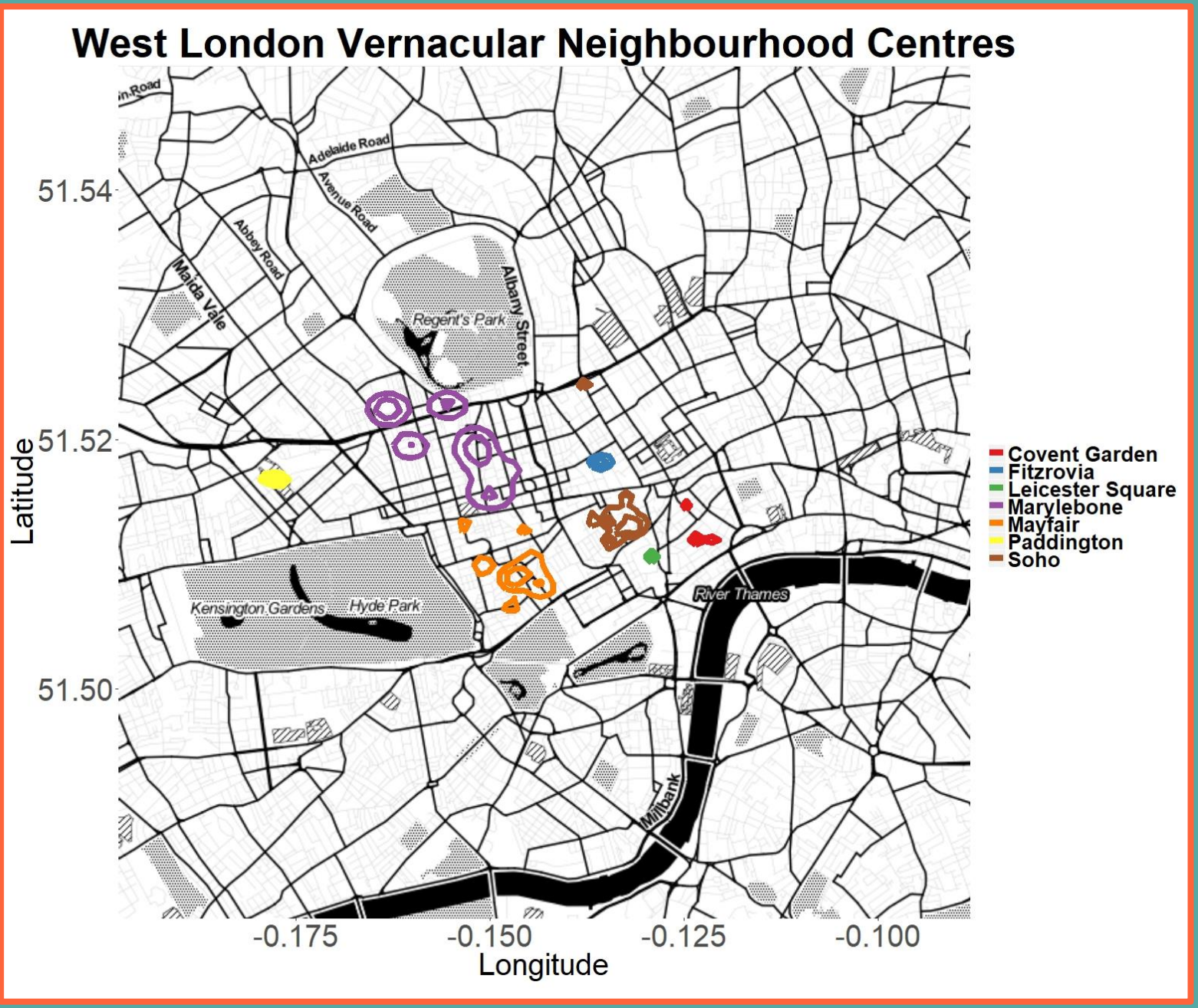



Percentage of Vernacular Polygon Intersecting Ward Polygon
Percentage
0
20
40
60
80
Aldgate
Bishopsgate
Bloomsbury
Clerkenwell
Covent Garden
Farringdon
Holborn
Hoxton
Kings Cross
Lambeth
Marylebone
Southwark
Spitalfields
Vernacular Neighbourhood

Vernacular Neighbourhood Delimitation by Convex Hull

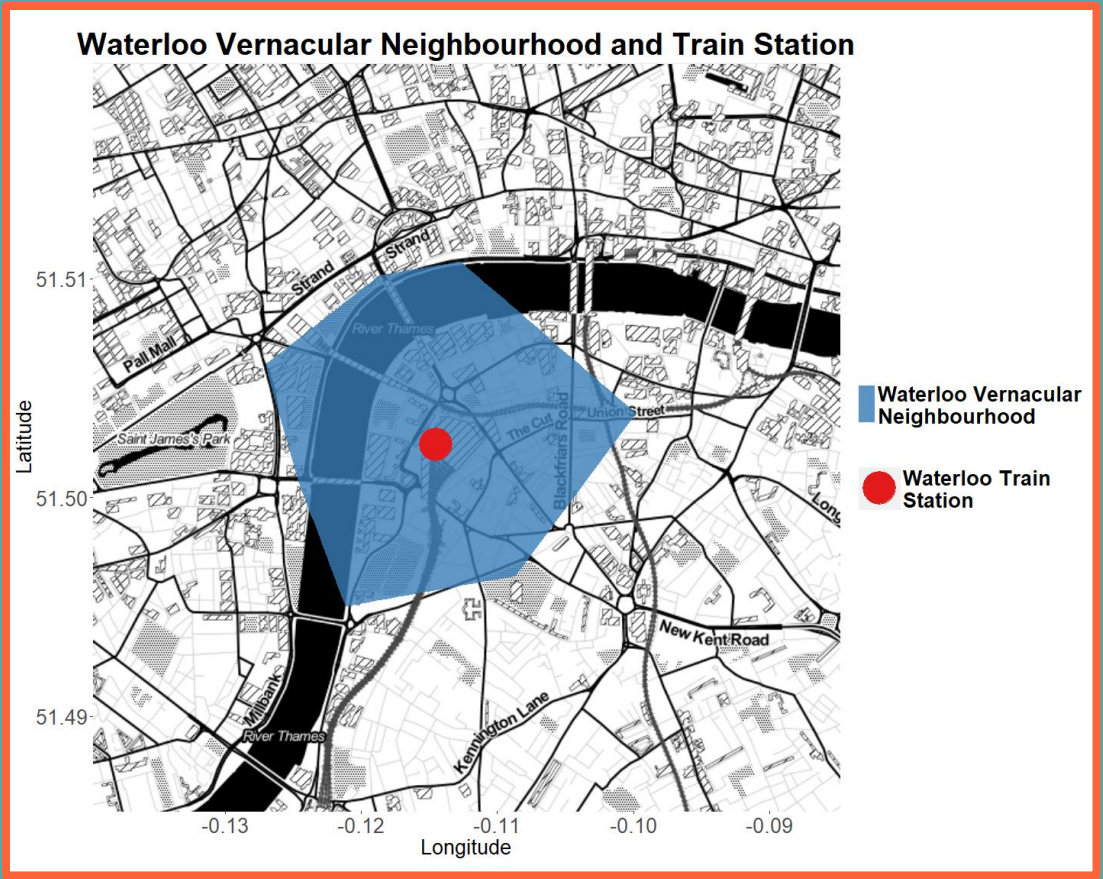


Vernacular Neighbourhood Centres

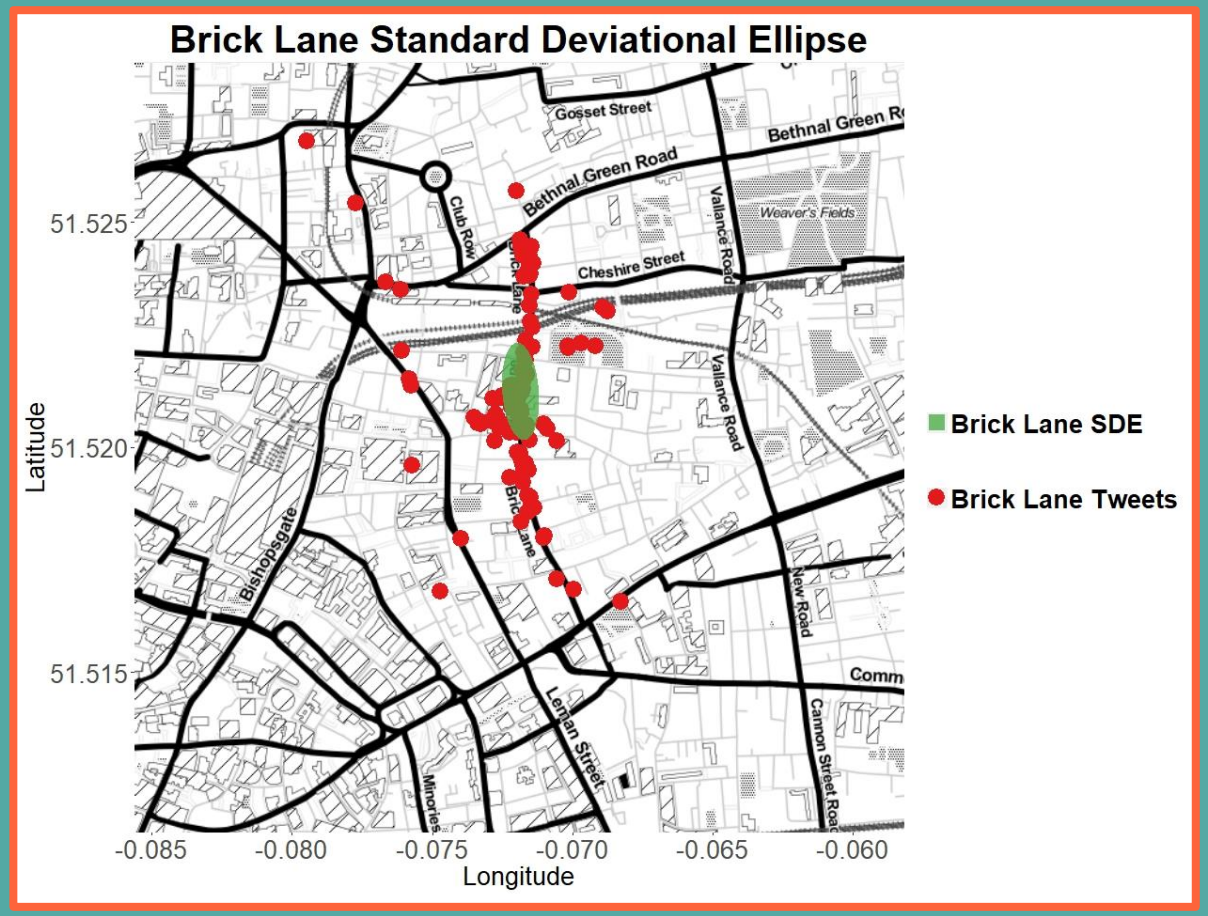


Lynch's (1960) Elements Form Basis of Vernacular Neighbourhoods

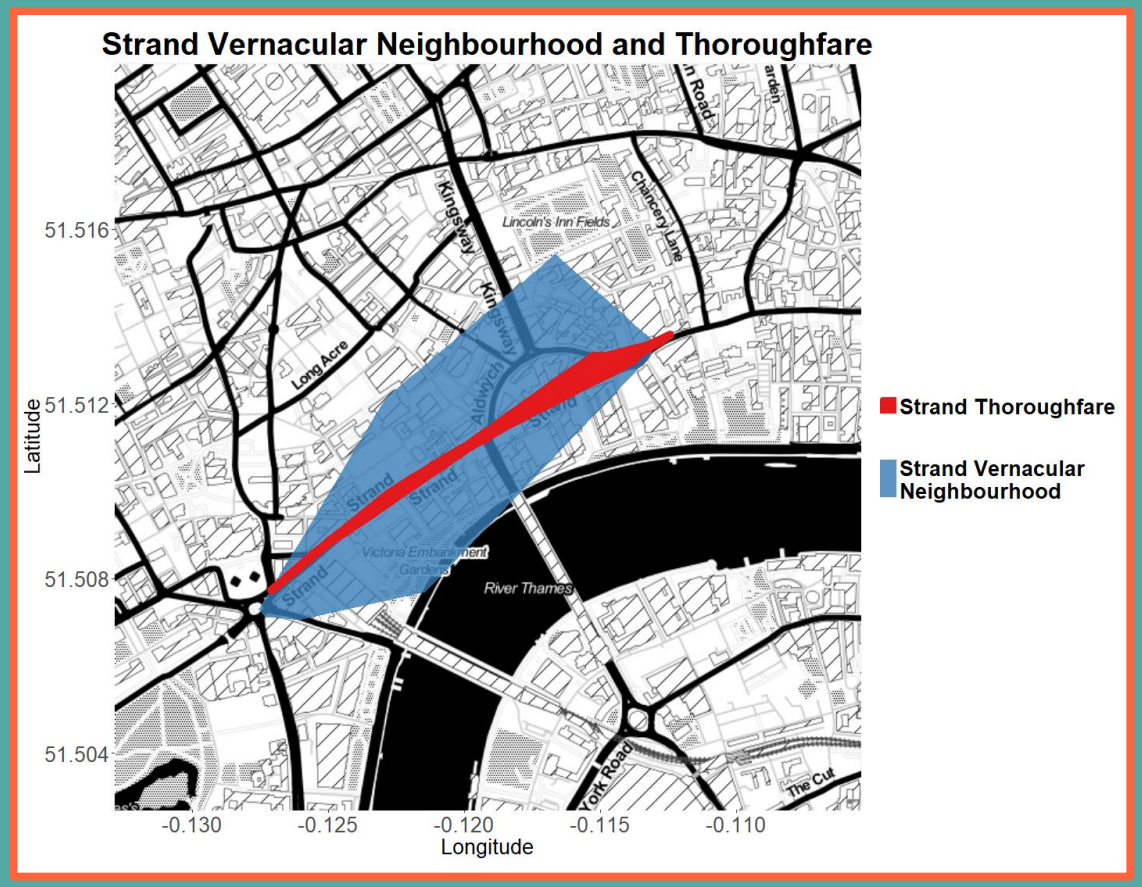
Nodes and Landmarks



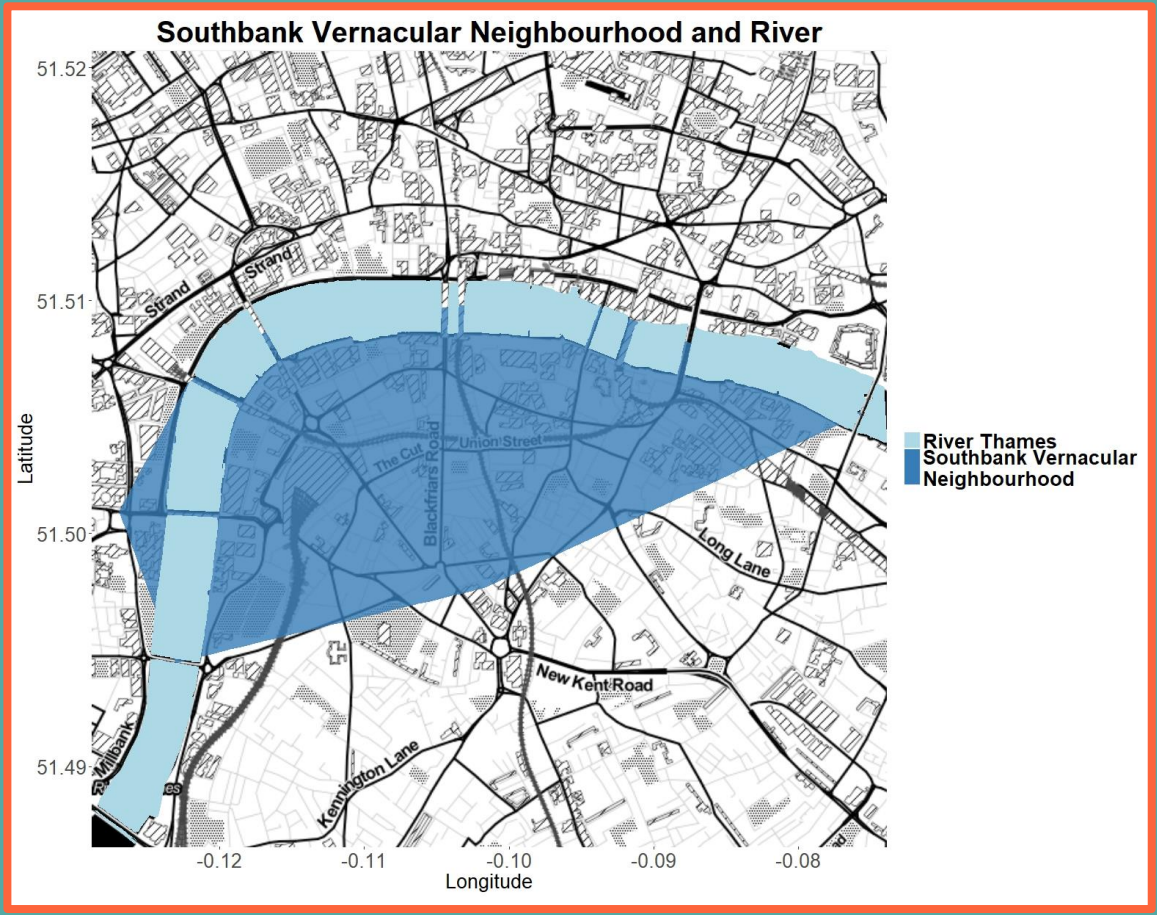
Path Elements, Lynch (1960)

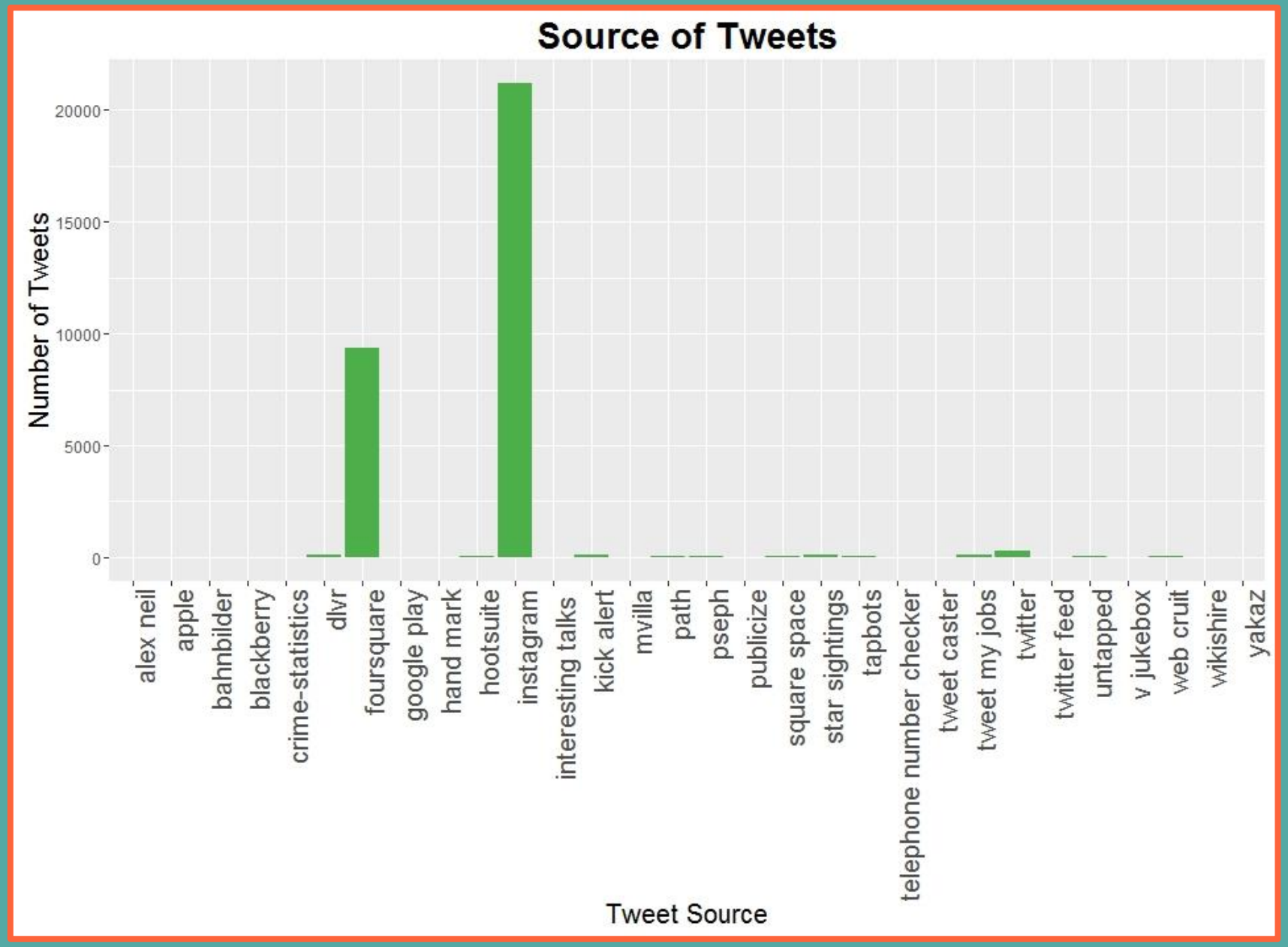


More Path Elements, Lynch (1960)



Edge Elements, Lynch (1960)



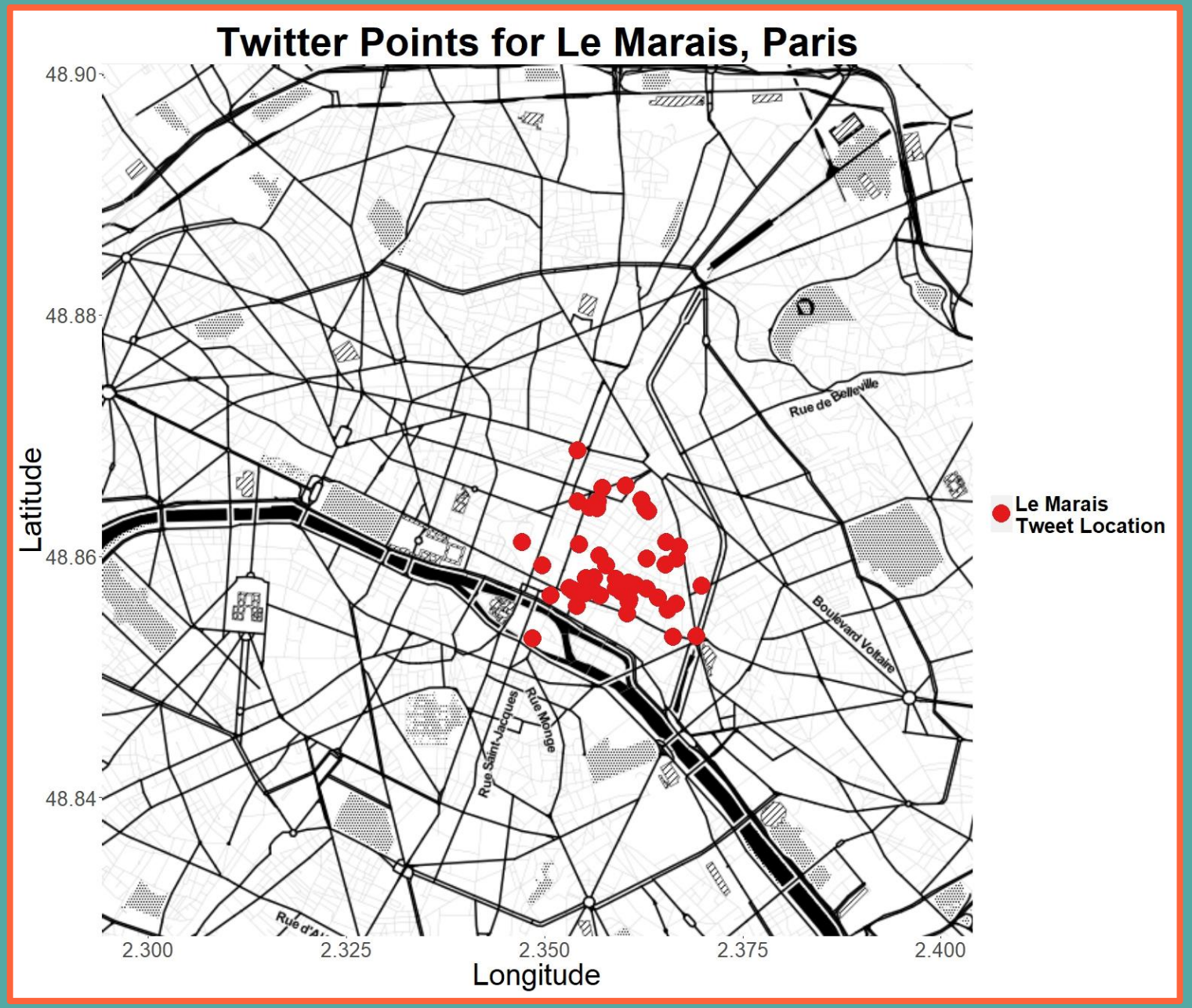

Source of Tweets
Number of Tweets
20000
15000
10000
5000
0
alex neil
apple
bahnbilder
blackberry
crime-statistics
dlvr
foursquare
google play
hand mark
hootsuite
instagram
interesting talks
kick alert
mvilla
path
pseph
publicize
square space
star sightings
tapbots
telephone number checker
tweet caster
tweet my jobs
twitter
twitter feed
untapped
v jukebox
web cruit
wikishire
yakaz
Tweet Source

Limitations of Study and Further Work.

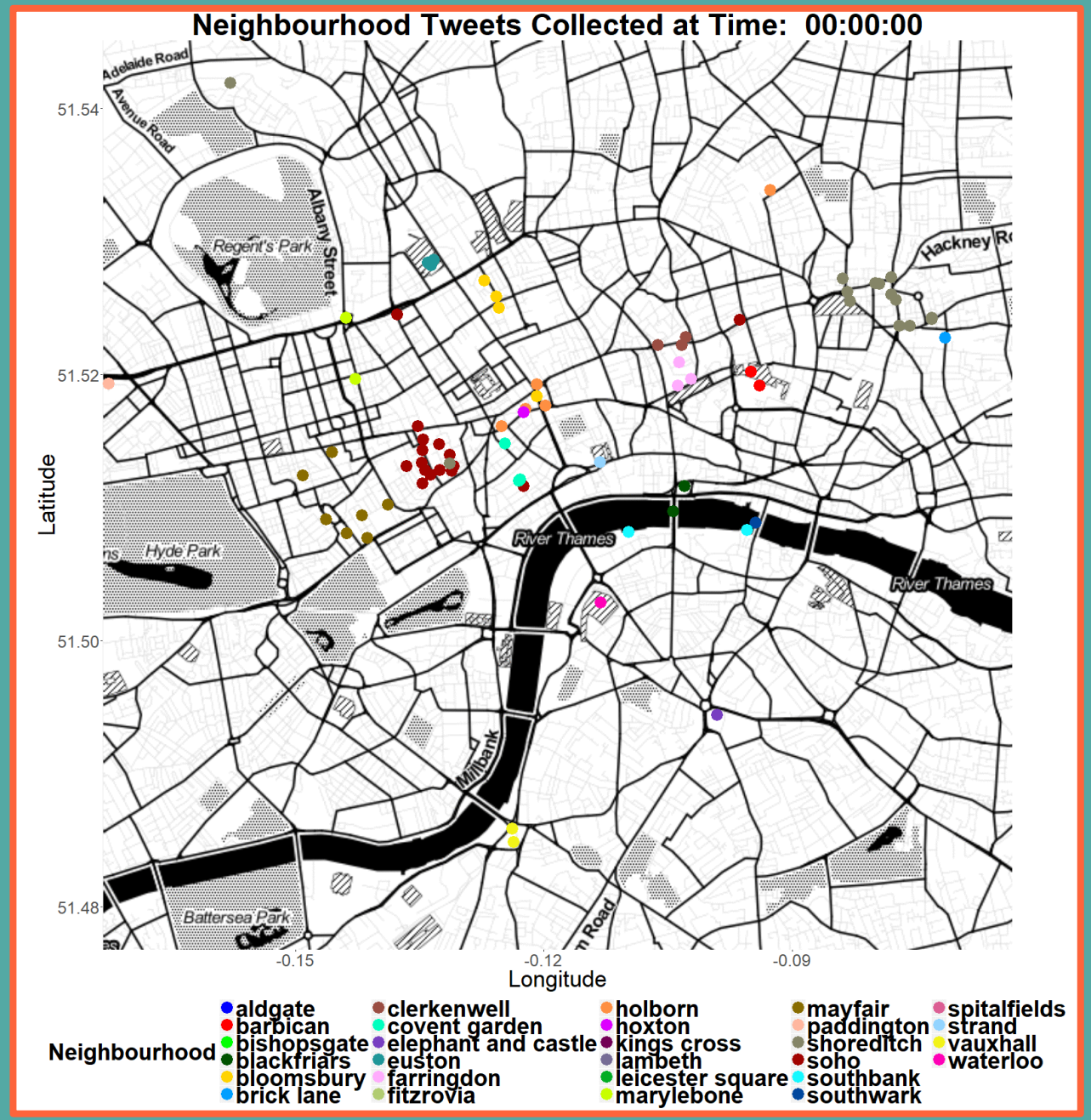
- Qualitative Coding process subjective and time consuming.
- Twitter API Limits number of Tweets harvested.
- 31,692 Tweets collected but only 14,832 users.
- Socio/economic/cultural demographics of Twitter users.
- Machine learning algorithm to sort Tweets (in R).
- Test data quality by comparing to results from Flickr.
- Apply to other cities, e.g. Paris.
- Look at spatio-temporal aspect of Twitter data.

Further Work

Apply to Other Cities, Paris



Spatio-temporal Characteristic of Tweets.



References and Acknowledgements

- Cope, M. (2003) Coding Transcripts and Diaries. In Clifford, N., French, S. & Valentine, G. (Editors) Key Methods in Geography. Chapter 27: 440-452. London: Sage.
- Lovelace, R., Birkin, M., Cross, P. & Clarke, M. (2016) From Big Noise to Big Data: Towards the verification of Large Data sets for Understanding Region Retails Flows. Geographical Analysis 48: 59-81.
- Lynch, K. (1960) The Image of the City. Cambridge (Massachusetts):The MIT Press.
- See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H., Milcinski, G., Niksic, M., Painho, M., Podor, A., Olteanu-Raimond, A. & Rutzinger, M. (2016) Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. International Journal of Geo-Information, 5 (55): 1-23.