

# How Similar are Neighbourhoods in London, Sydney and New York?

Aaron Armour

# Problem Background

- Moving to a new city is a big change
- Finding a similar\* neighbourhood (nbhd) in the new city may be an important consideration
- Probably also care about similarity of the new city as a whole

\* similar in terms of amenities available

- I analyse London, Sydney and New York

# Three Questions

1. How similar are nbhds across the three cities?

- Are there similar nbhds in other cities?
- Are nbhds most similar to others within the same city?

2. For each nbhd find the most similar nbhds from the other two cities.

3. How similar are each pair of cities?

# Data Sources

- Wikipedia for neighbourhood lists

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich <sup>[7]</sup>	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham <sup>[8]</sup>	LONDON	W3, W4	020	TQ205805
Addington	Croydon <sup>[8]</sup>	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon <sup>[8]</sup>	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
Aldborough Hatch	Redbridge <sup>[9]</sup>	ILFORD	IG2	020	TQ455895
Aldgate	City <sup>[10]</sup>	LONDON	EC3	020	TQ334813
Aldwych	Westminster <sup>[10]</sup>	LONDON	WC2	020	TQ307810
Alperton	Brent <sup>[11]</sup>	WEMBLEY	HA0	020	TQ185835
Anerley	Bromley <sup>[11]</sup>	LONDON	SE20	020	TQ345695
Angel	Islington <sup>[8]</sup>	LONDON	EC1, N1	020	TQ345665
Aperfield	Bromley <sup>[11]</sup>	WESTERHAM	TN16	01959	TQ425585
Archway	Islington <sup>[12]</sup>	LONDON	N19	020	TQ285875
Ardleigh Green	Havering <sup>[12]</sup>	HORNCHURCH	RM11	01708	TQ535895
Arkley	Barnet <sup>[12]</sup>	BARNET, LONDON	EN5, NW7	020	TQ225955
Arnos Grove	Enfield <sup>[12]</sup>	LONDON	N11, N14	020	TQ295925
Balham	Wandsworth <sup>[13]</sup>	LONDON	SW12	020	TQ285735
Bankside	Southwark <sup>[14]</sup>	LONDON	SE1	020	TQ325795

# Data Sources

- Foursquare for venue information

```
{'reasons': {'count': 0,
  'items': [{'summary': 'This spot is popular',
    'type': 'general',
    'reasonName': 'globalInteractionReason'}]},
'venue': {'id': '4c4991b19f2ad13a0fd77653',
  'name': 'Canary Riverside',
  'location': {'address': 'Canary Riverside',
    ...
    'formattedAddress': ['Canary Riverside',
      'Canary Wharf',
      'Greater London',
      'R M14',
      'United Kingdom']},
  'categories': [{'id': '4bf58dd8d48988d164941735',
    'name': 'Plaza',
    ...
  }],
'referralId': 'e-0-4c4991b19f2ad13a0fd77653-0'}
```

# Data

- London (Inner Boroughs): 187 nbhds
- Sydney (City of Sydney): 48 nbhds
- New York (Manhattan): 85 nbhds
- Drop nbhds with  $< 50$  venues for London and New York,  $< 30$  venues for Sydney
- London: 61 nbhds
- Sydney: 30 nbhds
- New York: 62 nbhds

# Approach

- One-hot encoding of venue category
- Group by nbhd and average
- So feature vector for each nbhd is observed proportions of each type of venue
- Similarity between nbhds is measured by distance between their feature vectors



# Results – Question 1

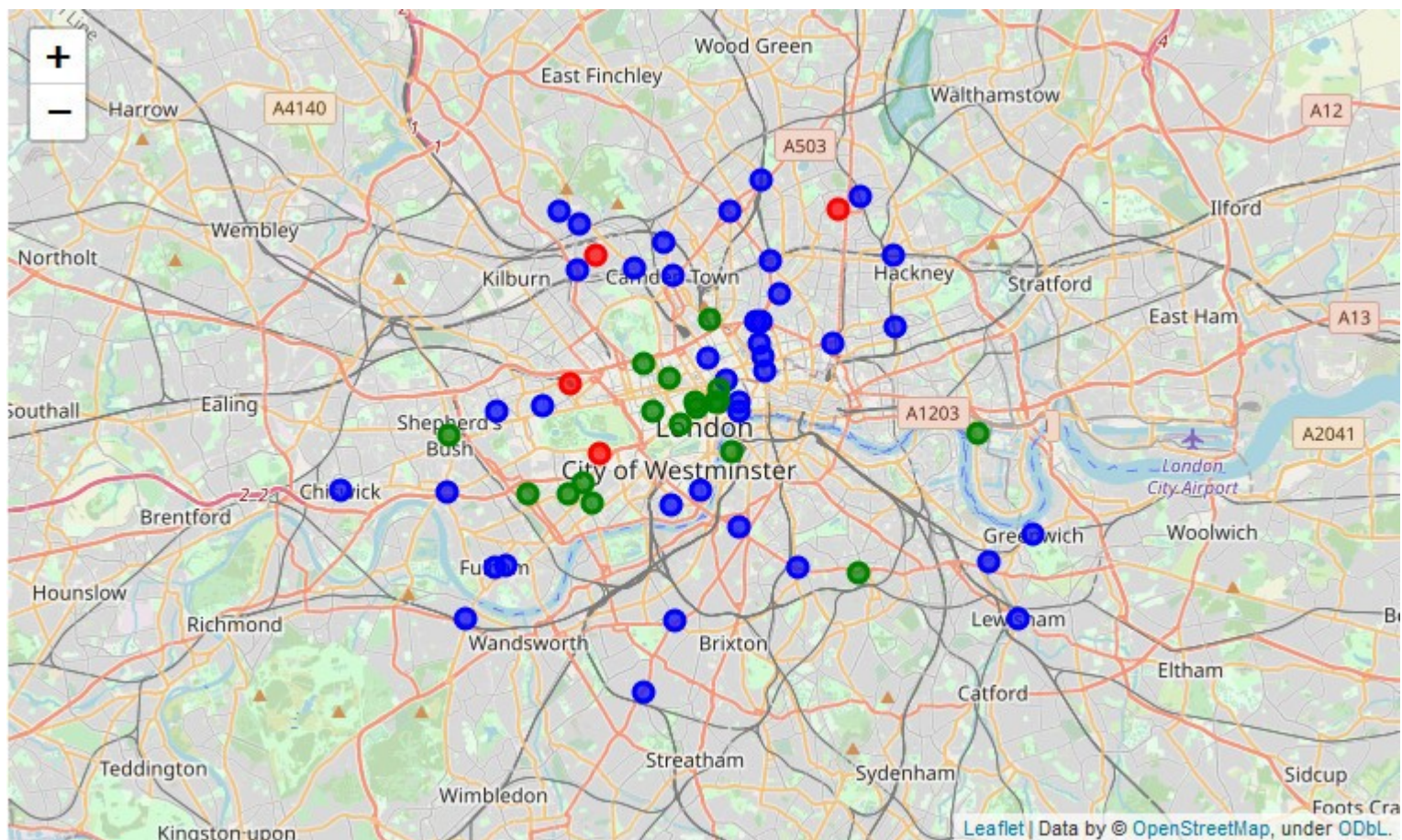
- Clustering nbhds
  - Use K-Means with  $k=3$

Cluster label	London	Sydney	New York
0	4	28	0
1	39	0	0
2	18	2	62



# Results – Question 1

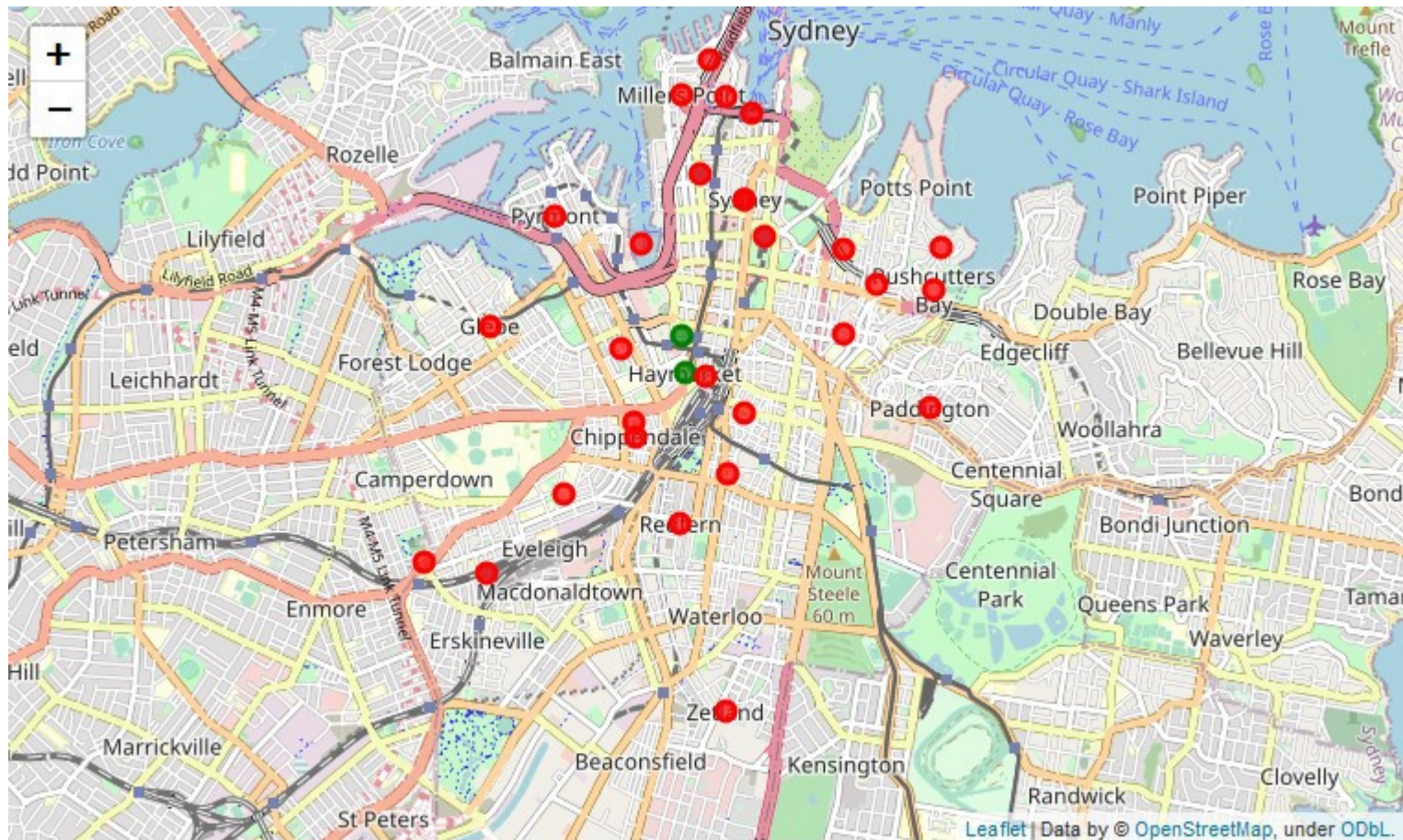
- London clusters:





# Results – Question 1

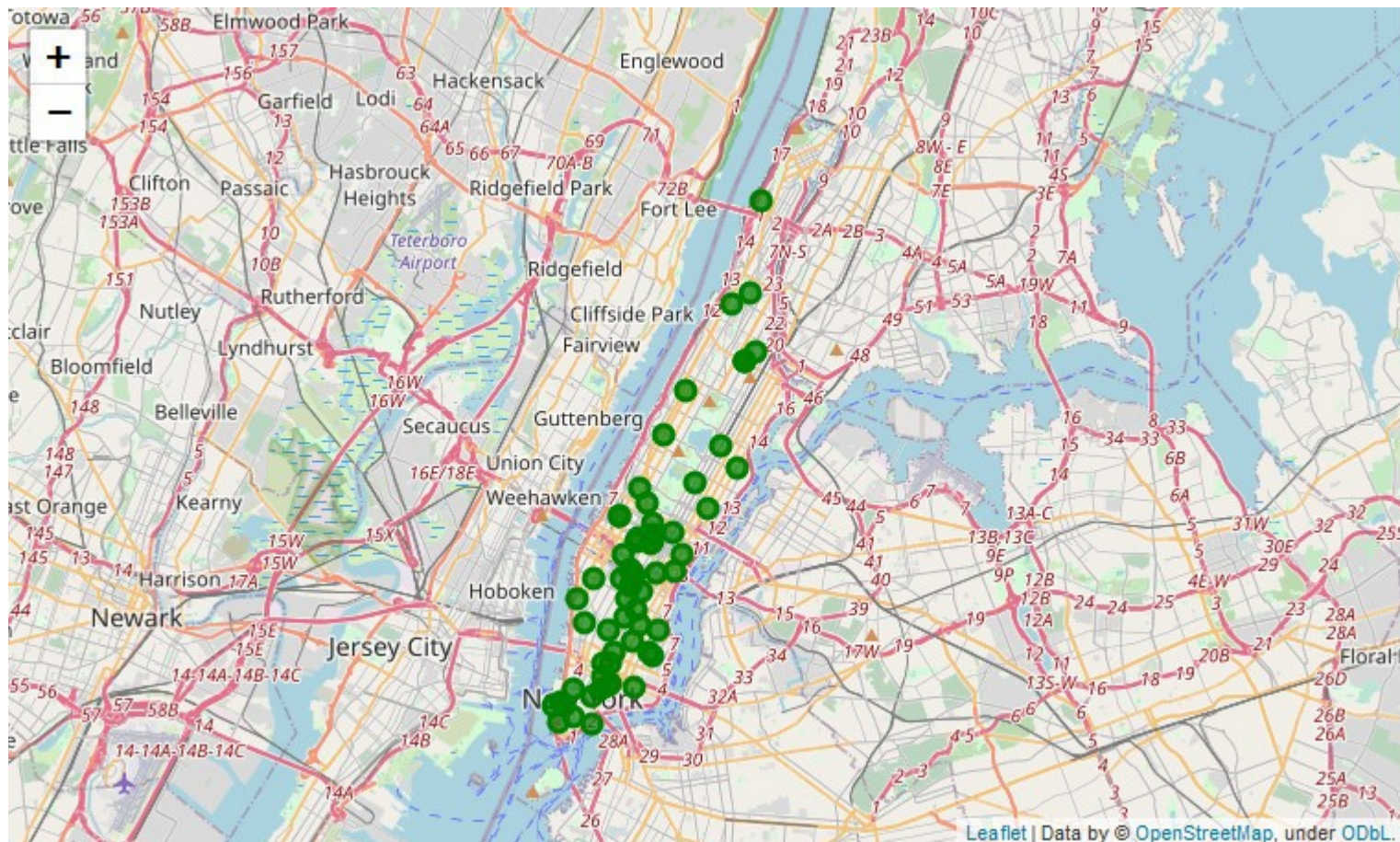
- Sydney clusters:





# Results – Question 1

- New York cluster:



## Results – Question 2

- Average distance between (feature vectors of) nbhds in a given city:

	Intra city average neighbourhood feature distance
London	0.2116
Sydney	0.2011
New York	0.2000

- Given a nbhd, find a nbhd in other cities minimising (feature vector) distance to given one



# Results – Question 2

- Example of results obtained
  - Complete tables in appendix of my report

London neighbourhood	Cluster labels	Sydney		New York	
		Neighbourhood	Feature distance	Neighbourhood	Feature distance
Aldwych	1	The Rocks	0.1951	Midtown	0.1939
Angel	1	Woolloomooloo	0.1866	Tudor City	0.1761
Balham	1	Wynyard	0.2026	Yorkville	0.1989
Bayswater	1	Woolloomooloo	0.1638	Tudor City	0.1700
Belsize Park	0	Woolloomooloo	0.1670	Tudor City	0.1782
Bethnal Green	1	Wynyard	0.1833	Financial District	0.1859
Bloomsbury	1	Wynyard	0.1760	Financial District	0.2000
Brompton	2	The Rocks	0.1681	Murray Hill	0.1573
Camberwell	1	Woolloomooloo	0.2029	Tudor City	0.1874
Camden Town	1	Woolloomooloo	0.1872	Lower East Side	0.1940
Canary Wharf	2	East Sydney	0.1802	Tribeca	0.1518
Chalk Farm	1	East Sydney	0.1425	Lower East Side	0.1615
Chinatown (London)	2	The Rocks	0.1949	Times Square	0.1442
Chiswick	1	Woolloomooloo	0.1942	Tudor City	0.1836
Clapham	1	Woolloomooloo	0.1756	Yorkville	0.1699

# Results – Question 2

- Following tables provide a summary

<b>Sydney neighbourhood</b>	<b>Number of London neighbourhoods</b>	<b>London neighbourhood</b>	<b>Number of Sydney neighbourhoods</b>
<b>Chinatown (Sydney)</b>	1	<b>Bayswater</b>	2
<b>Darling Harbour</b>	1	<b>Brompton</b>	1
<b>Darlinghurst</b>	3	<b>Chalk Farm</b>	14
<b>East Sydney</b>	7	<b>Hackney Central</b>	2
<b>Glebe</b>	2	<b>Hammersmith</b>	2
<b>The Rocks</b>	17	<b>Highbury</b>	1
<b>Woolloomooloo</b>	24	<b>Knightsbridge</b>	1
<b>Wynyard</b>	6	<b>Shoreditch</b>	3
		<b>Stoke Newington</b>	2
		<b>West Hackney</b>	2

# Results – Question 2

New York neighbourhood	Number of London neighbourhoods	London neighbourhood	Number of New York neighbourhoods
Central Harlem	4	Brompton	2
Columbus Circle	1	Canary Wharf	4
Financial District	6	Chalk Farm	2
Flower District	3	Chinatown (London)	3
Garment District	2	Hammersmith	1
Little Australia	3	Marylebone	5
Lower East Side	6	Nag's Head	2
Meatpacking District	1	Putney	4
Midtown	1	Shepherd's Bush	10
Midtown East	1	Shoreditch	20
Murray Hill	1	South Kensington	2
NoHo	1	St Giles	7
SoHo	4		
Times Square	5		
Tribeca	3		
Tudor City	12		
Turtle Bay	1		
Upper East Side	1		
Yorkville	5		



# Results – Question 2

<b>New York neighbourhood</b>	<b>Number of Sydney neighbourhoods</b>	<b>Sydney neighbourhood</b>	<b>Number of New York neighbourhoods</b>
<b>Bowery</b>	2	<b>Chinatown (Sydney)</b>	13
<b>Carnegie Hill</b>	2	<b>East Sydney</b>	30
<b>Financial District</b>	2	<b>Newtown</b>	2
<b>Flower District</b>	3	<b>The Rocks</b>	13
<b>Hudson Heights</b>	10	<b>Woolloomooloo</b>	3
<b>Koreatown</b>	1	<b>Wynyard</b>	1
<b>Little Australia</b>	1		
<b>Lower East Side</b>	4		
<b>Tudor City</b>	5		

# Results – Question 3

- Two measure of similarity at city level
  - Average nbhd feature distance

	London	Sydney	New York
London	0.2116	0.2426	0.2246
Sydney		0.2011	0.2577
New York			0.2000

- Average nbhd matching distance

	London	Sydney	New York
London		0.1959	0.1919
Sydney			0.2236
New York			

- London and New York are most similar

# Conclusions

- There are nbhds more similar to nbhds in other cities than other nbhds in the same city
- London appears to be the most diverse
- For a nbhd in a given city we can find a nbhd in a different city which is fairly similar
- London and New York are the most similar cities of the three analysed