



Clustering the biggest tech hub cities

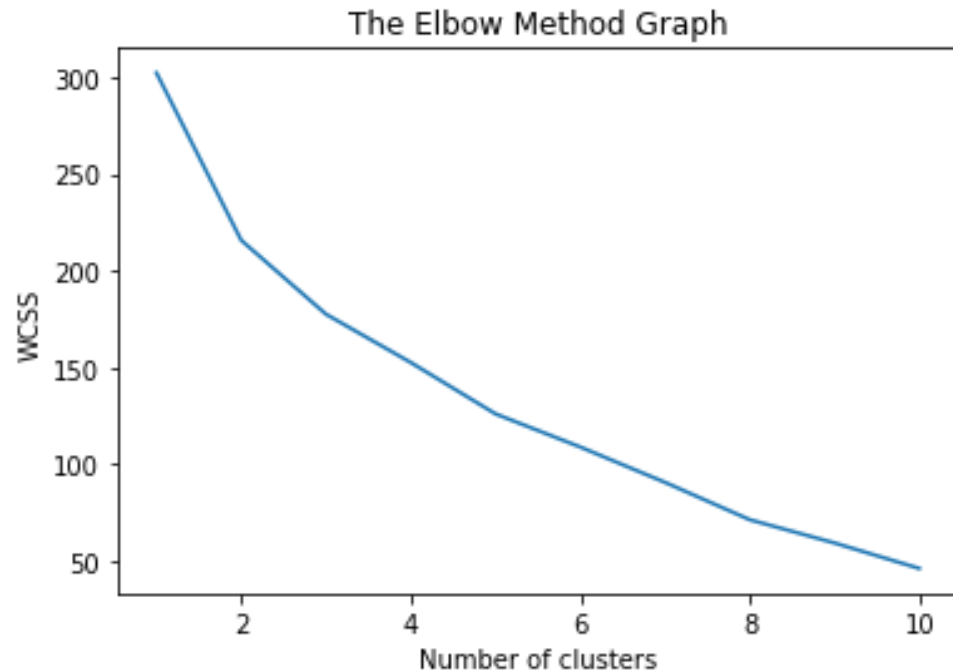
MAY 11, 2021

Data acquisition and cleaning

Three different sets of data:

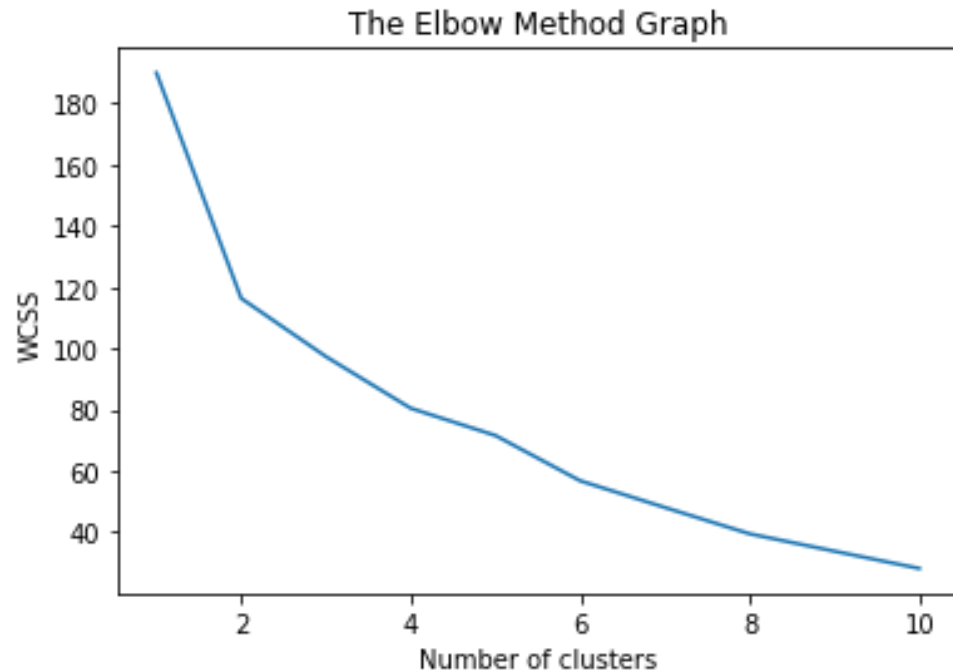
- General population, economy, environment data
- Innovation and startups ecosystem indicators
- Foursquare API data

1st set evaluation



LABELS CITY		SINGAPORE	
0	DUBLIN	1	1
	BARCELONA	1	1
	BERLIN	1	1
	GENEVA	1	1
	HONG KONG	1	1
1	LONDON	1	1
	LOS ANGELES	1	1
	NEW YORK	1	1
	PARIS	1	1
2	STOCKHOLM	1	1
	SYDNEY	1	1
	TORONTO	1	1
	VANCOUVER	1	1
3	BEIJING	1	1
	SHANGHAI	1	1
	AMSTERDAM	1	1
4	BOSTON	1	1
	CHICAGO	1	1

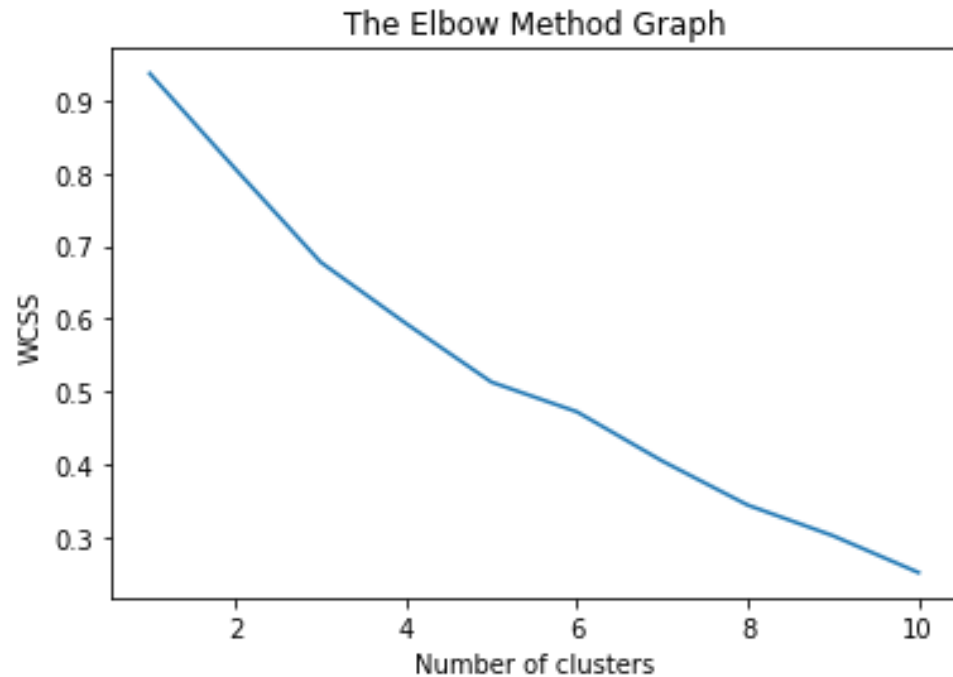
2nd set evaluation



	LABELS_2	CITY
0	BARCELONA	1
	CHICAGO	1
	DUBLIN	1
1	BEIJING	1
	BOSTON	1
	LONDON	1
	LOS ANGELES	1
	NEW YORK	1
	SHANGHAI	1
2	AMSTERDAM	1

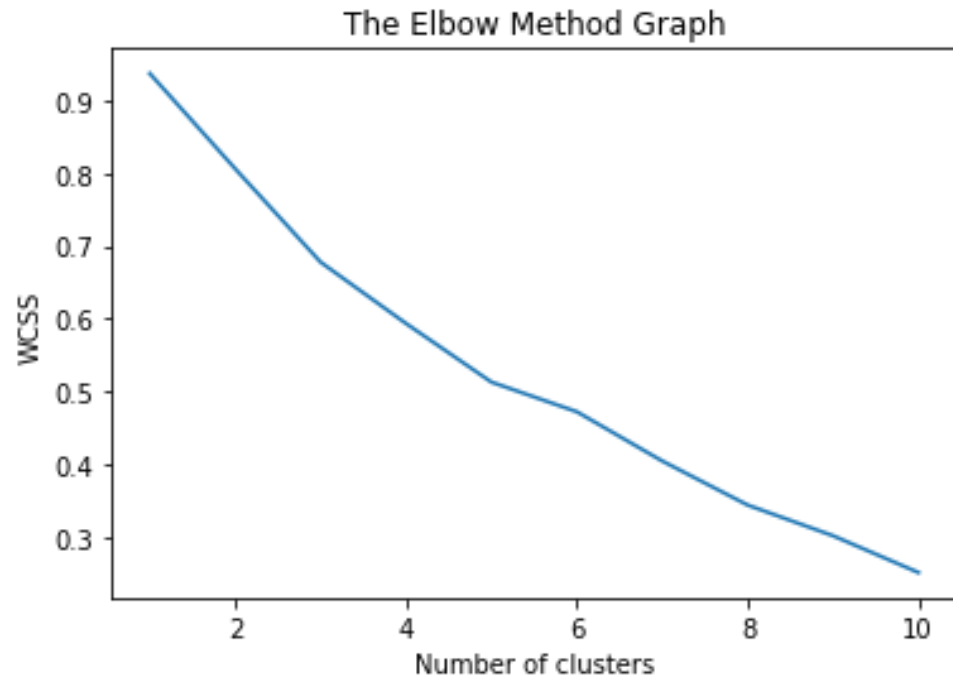
BERLIN	1	
GENEVA	1	
PARIS	1	
SINGAPORE	1	
STOCKHOLM	1	
TORONTO	1	
3	HONG KONG	1
	SYDNEY	1
	VANCOUVER	1

3rd set evaluation



LABELS_3 CITY			VANCOUVER 1		
0	BOSTON	1	2	BEIJING	1
	LONDON	1		SHANGHAI	1
	LOS ANGELES	1		AMSTERDAM	1
	NEW YORK	1	3	BARCELONA	1
	PARIS	1		CHICAGO	1
	STOCKHOLM	1		DUBLIN	1
	TORONTO	1		GENEVA	1
1	BERLIN	1	SYDNEY 1		
	HONG KONG	1			
	SINGAPORE	1			

4th set evaluation



LABELS_4 CITY

0 GENEVA 1

SYDNEY 1

1 BARCELONA 1

BERLIN 1

BOSTON 1

CHICAGO 1

DUBLIN 1

LONDON 1

LOS ANGELES 1

NEW YORK 1

PARIS 1

SINGAPORE 1

STOCKHOLM 1

TORONTO 1

VANCOUVER 1

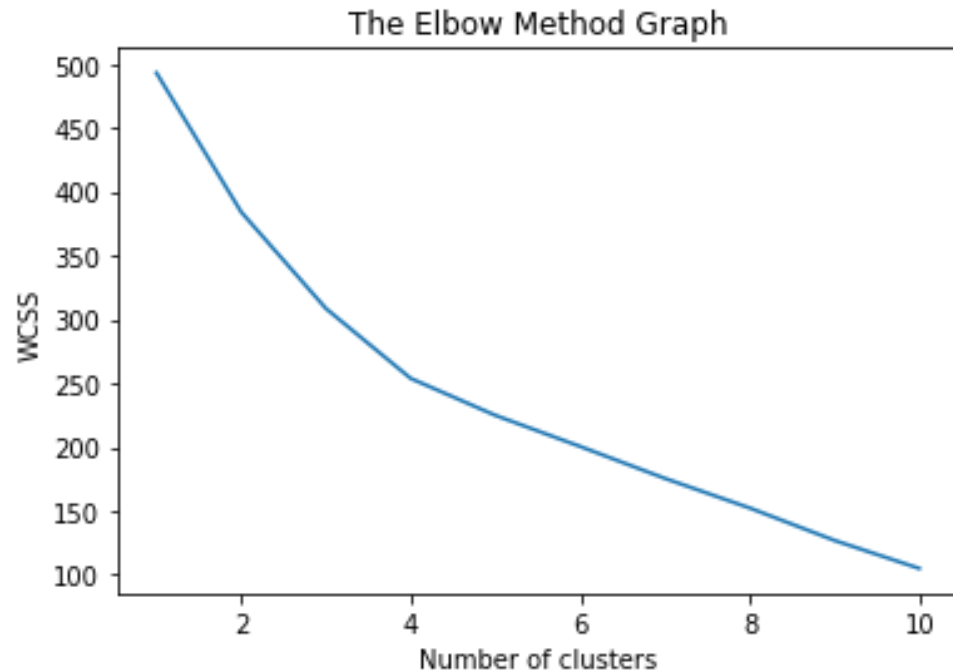
2 SHANGHAI 1

3 AMSTERDAM 1

BEIJING 1

4 HONG KONG 1

6th set evaluation

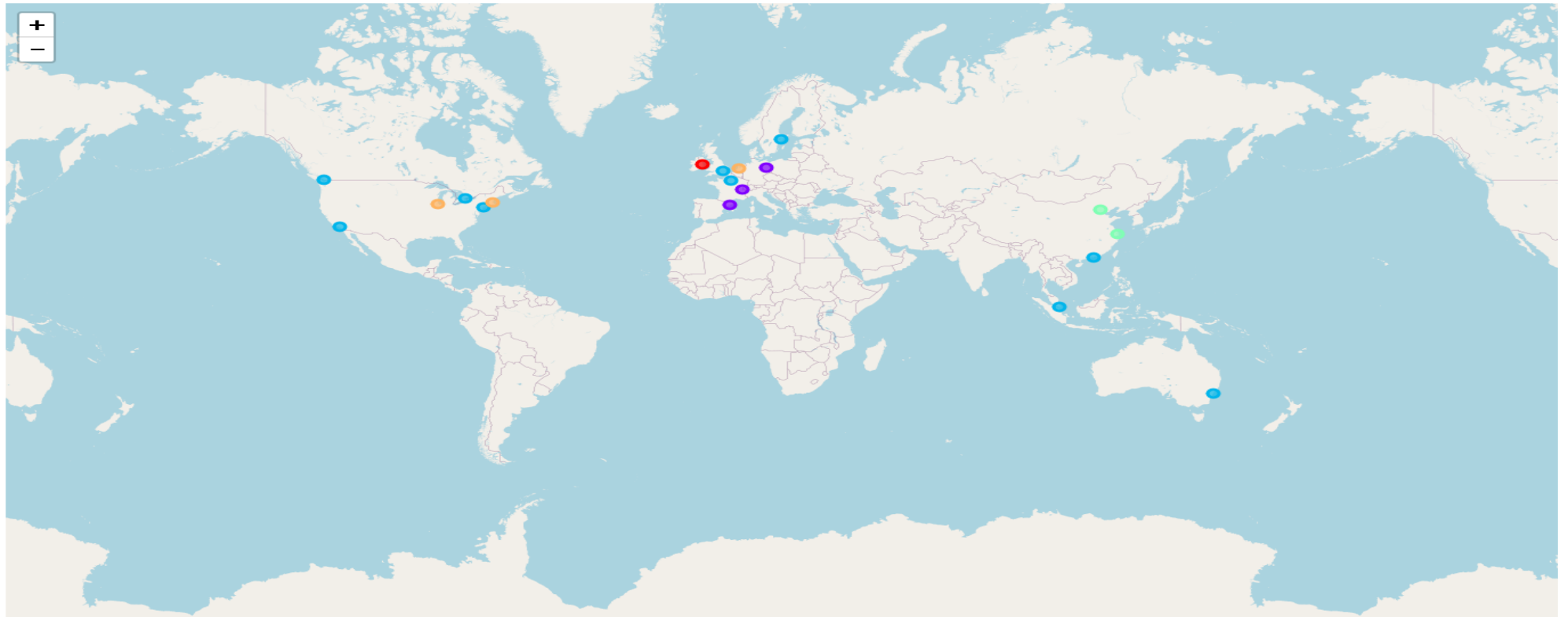


LABELS_6 CITY			SINGAPORE		
0	BOSTON	1	2	BEIJING	1
	LONDON	1		SHANGHAI	1
	LOS ANGELES	1	3	AMSTERDAM	1
	NEW YORK	1		BARCELONA	1
	PARIS	1		CHICAGO	1
	STOCKHOLM	1			
	TORONTO	1			
1	BERLIN	1			
	HONG KONG	1			

Combined labels

	city	try_1	try_2	try_3	try_4	try_6
0	New York	2	1	0	1	0
1	London	2	1	0	1	0
2	Beijing	3	1	2	3	2
3	Boston	4	1	0	1	0
4	Los Angeles	2	1	0	1	0
5	Shanghai	3	1	2	2	2
6	Paris	2	2	0	1	0
7	Berlin	1	2	1	1	1
8	Stockholm	2	2	0	1	0
9	Toronto	2	2	0	1	0
10	Singapore	2	2	1	1	1
11	Amsterdam	4	2	3	3	3
12	Chicago	4	0	3	1	3
13	Geneva	1	2	3	0	3
14	Sydney	2	3	3	0	3
15	Vancouver	2	3	1	1	1
16	Hong Kong	2	3	1	4	1
17	Barcelona	1	0	3	1	3
18	Dublin	0	0	3	1	3

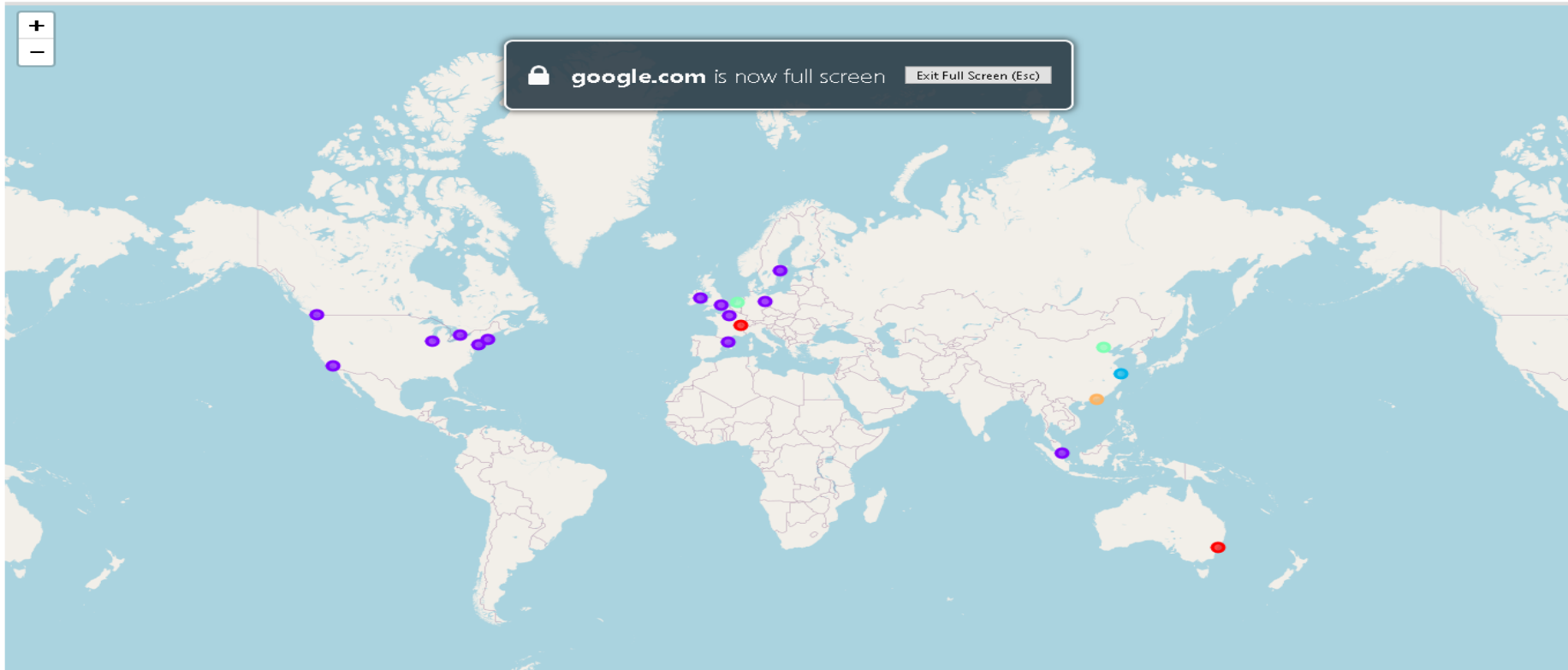
Map – 1st set



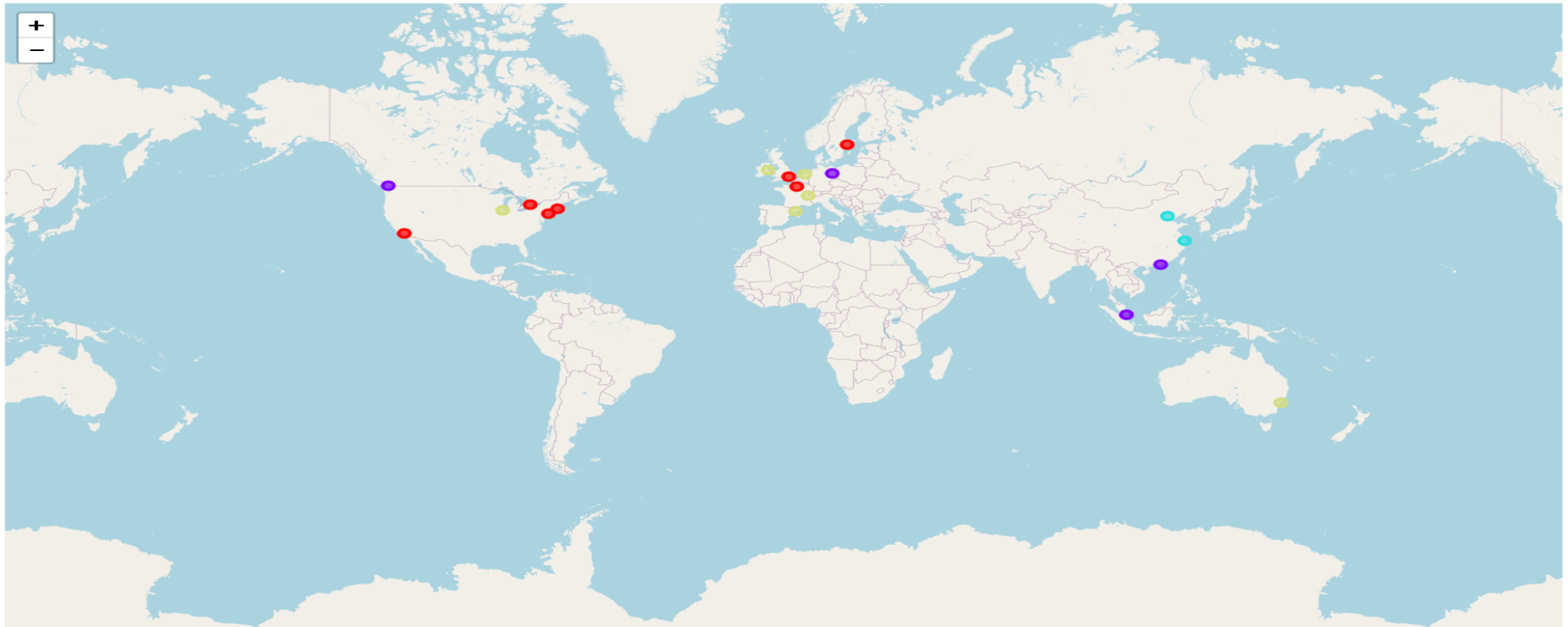
Map – 2nd set

Map – 3rd set

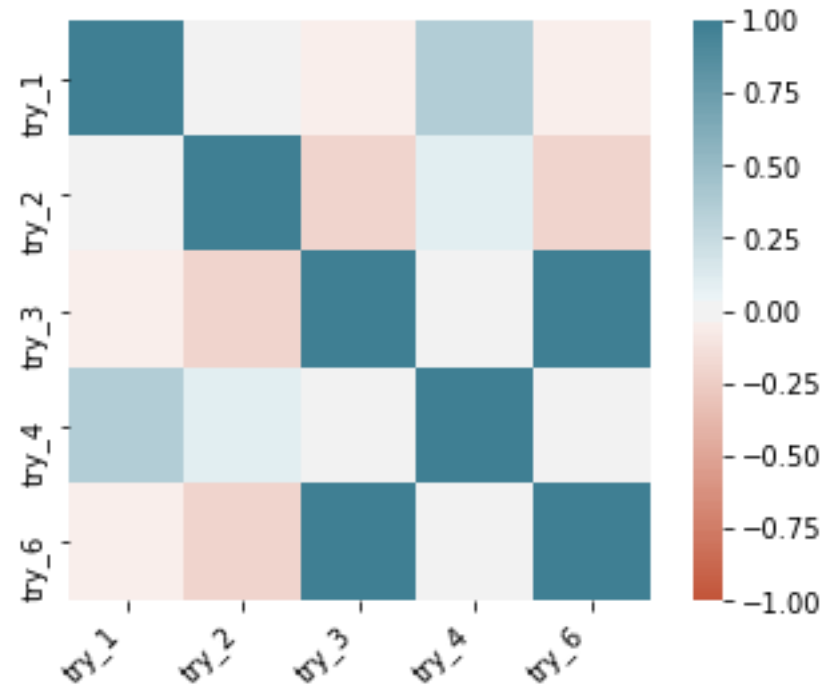
Map – 4th set



Map – 6th set



Heat-map of clusters correlations



Conclusions

- An approach to evaluate clusters obtained with different datasets was proposed
- Need to further investigate quality of application of clustering algorithm and improve results precision
- Data quality should be improved
- East-Asian/Australian clusters rarely overlap with other clusters
- Other relations and patterns should be investigated explain what causes different clustering configurations

Your best quote that reflects your approach... “It’s one small step for man, one giant leap for mankind.”

- NEIL ARMSTRONG