



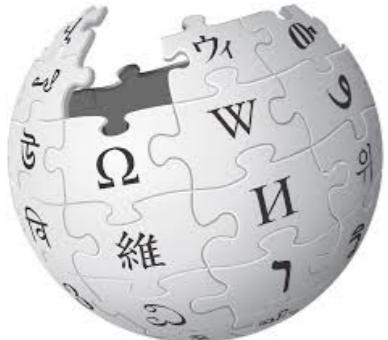
# Final Capstone Project

The Battle of the Cities

# Motivation

---

- Travel
  - Moving
- } to which city?



- Many tools to analyze different aspects, for different cities one by one
- Too many options, time consuming

# Motivation

---

- An automatic tool to:
  - Group similar cities
  - Quantify their differences
- Possible questions to answer:
  - Is a big city with many restaurants more equal to a small city with many restaurants, or a big city with more parks?
  - Which city is more like Dublin: Cologne or Amsterdam?
  - Is Shanghai, Tokyo or Johannesburg more different than São Paulo?

# Data

---

- Latitude and longitude from geopy
- To compare cities, many quantities can be used:
  - Venues
  - Costs
  - Languages
  - ...

# Data

---

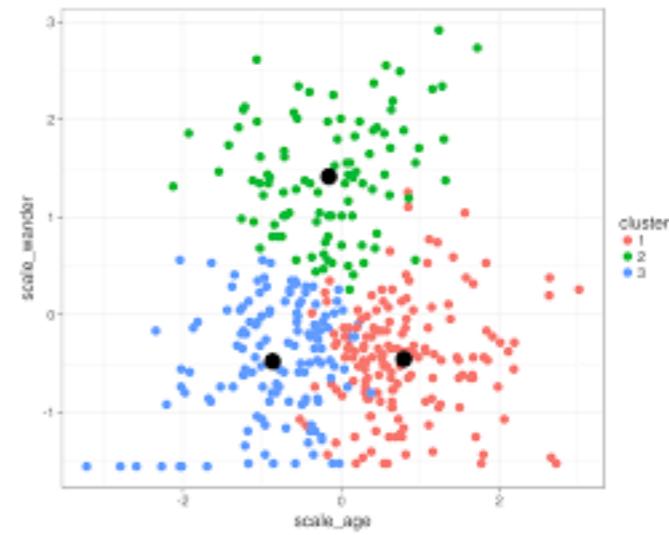
- Latitude and longitude from geopy
- To compare cities, many quantities can be used:



# Methodology

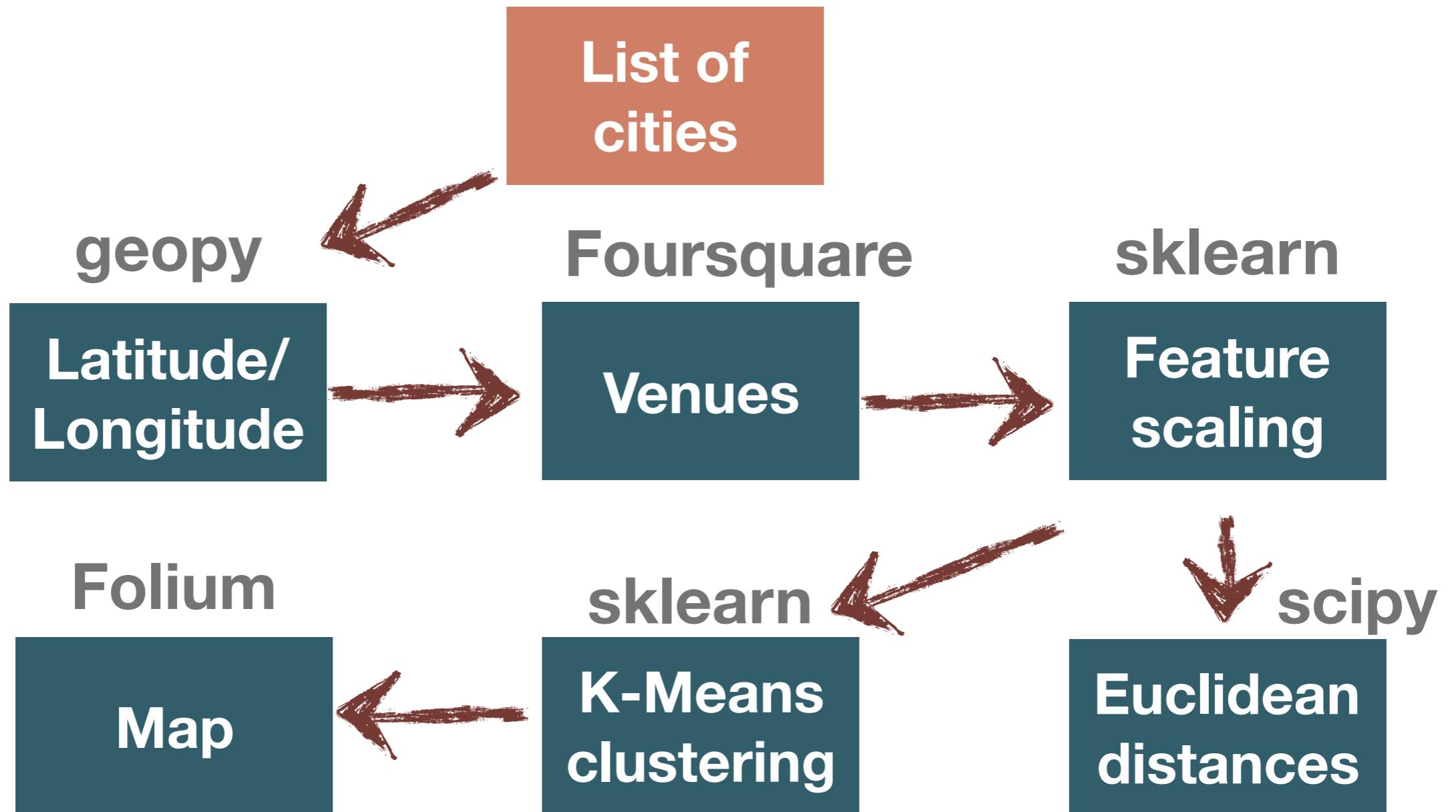
---

- Feature scaling (Standard scaler)
- K-Means clustering algorithm
  - 10 initial configurations
  - 3 clusters
- Euclidian distance
- Folium maps



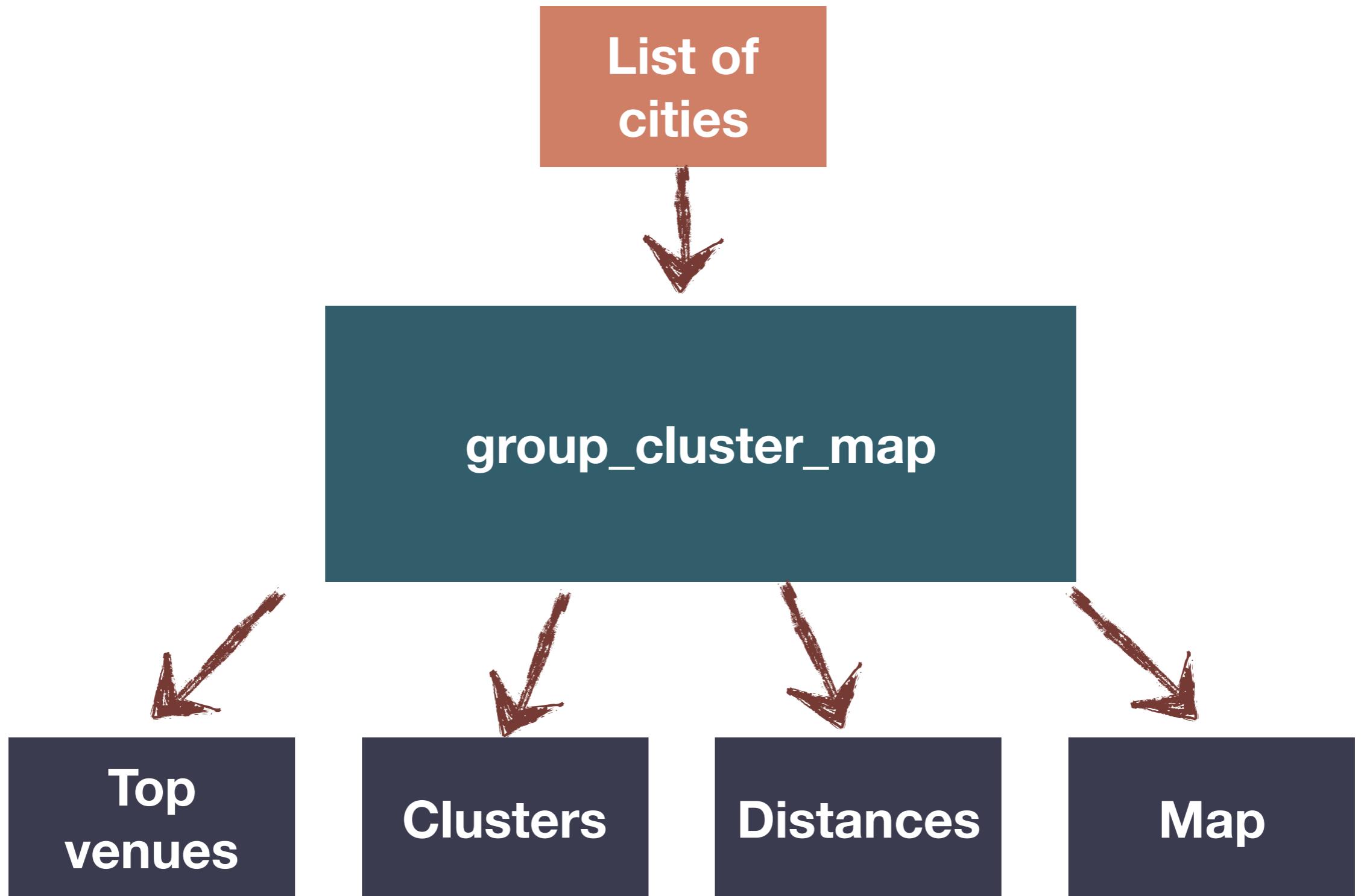
# Methodology

---



# Methodology

---

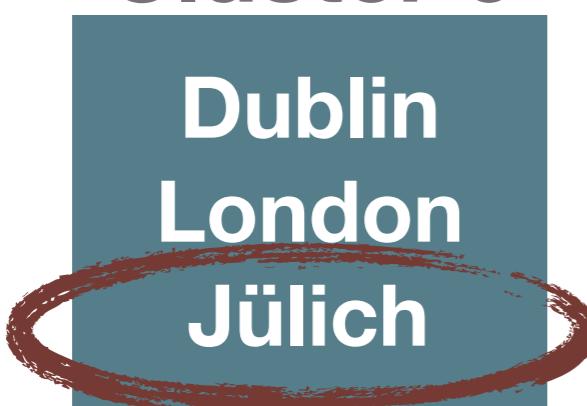


# Results

- Comparing big cities (and a small one)

	Address	Latitude	Longitude	Cluster Labels	Density	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Dublin, Ireland	53.349764	-6.260273	0	0.590	Mobile Phone Shop	Coffee Shop	Salon / Barbershop	Bus Line	Bus Stop	Women's Store	Café	Fast Food Restaurant	Ice Cream Shop	Convenience Store
1	London, England	51.507322	-0.127647	0	0.510	Office	Outdoor Sculpture	Taxi	Government Building	Italian Restaurant	Coworking Space	Embassy / Consulate	Tech Startup	Coffee Shop	Business Service
2	Cologne, Germany	50.938361	6.959974	2	0.610	Hotel	Office	Bar	Italian Restaurant	German Restaurant	Plaza	Gay Bar	Doctor's Office	Residential Building (Apartment / Condo)	Historic Site
3	Juelich, Germany	50.922093	6.361102	0	0.315	Miscellaneous Shop	Pizza Place	Salon / Barbershop	Optical Shop	Café	Bank	Pharmacy	Clothing Store	Shoe Store	Bakery
4	Amsterdam, Netherlands	52.374540	4.897976	2	0.690	Bar	Coffee Shop	Burger Joint	Office	Adult Boutique	Marijuana Dispensary	Bridge	Café	Hostel	Hotel
5	New York, USA	40.712728	-74.006015	2	0.615	Office	Taxi	Food Truck	Government Building	College Classroom	Building	Lawyer	Bar	Bus Station	General College & University
6	Sao Paulo, Brazil	-23.550651	-46.633382	1	0.725	Office	Pharmacy	College Academic Building	Historic Site	Cosmetics Shop	Building	Courthouse	Church	Music Venue	Outdoor Sculpture

Cluster 0



Cluster 1



Cluster 2



Small city was grouped with big cities

# Results

- Comparing big cities (and a small one)

	Dublin, Ireland	London, England	Cologne, Germany	Juelich, Germany	Amsterdam, Netherlands	New York, USA	Sao Paulo, Brazil
Dublin, Ireland	0.000000	24.246767	24.564217	25.478818	24.393808	24.775783	25.935727
London, England	24.246767	0.000000	24.303802	24.877851	24.405267	24.845097	25.162867
Cologne, Germany	24.564217	24.303802	0.000000	22.601027	22.627780	25.166785	25.695264
Juelich, Germany	25.478818	24.877851	22.601027	0.000000	23.770524	24.673967	26.207041
Amsterdam, Netherlands	24.393808	24.405267	22.627780	23.770524	0.000000	22.914864	24.833543
New York, USA	24.775783	24.845097	25.166785	24.673967	22.914864	0.000000	25.365967
Sao Paulo, Brazil	25.935727	25.162867	25.695264	26.207041	24.833543	25.365967	0.000000

São Paulo has the largest distances

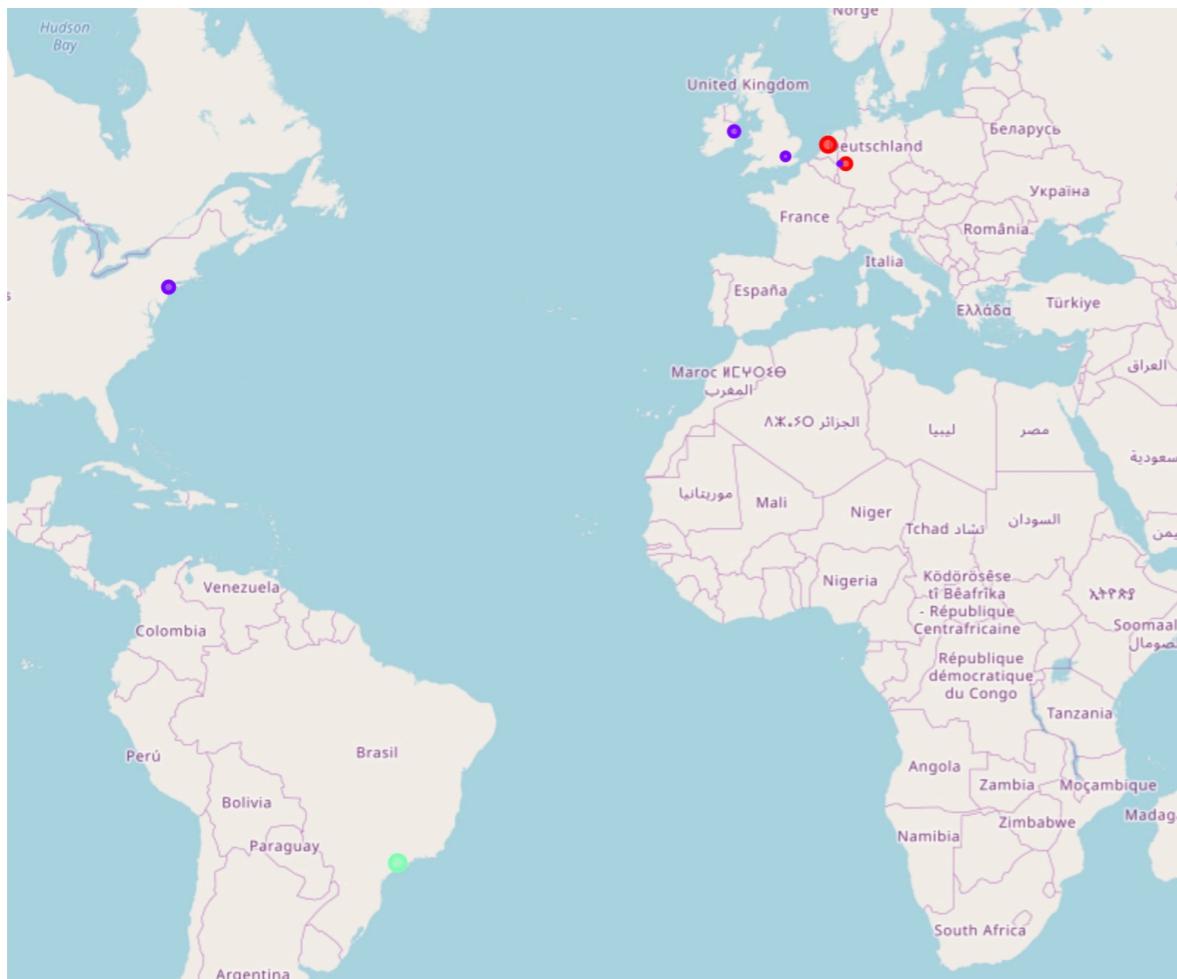


Jülich and Cologne are the closest ones

# Results

---

- Comparing big cities (and a small one)



**Similarity between the cities:**

**Geographical proximity**

**London/Dublin  
Cologne/Jülich**

**Historical connection**

**London/New York**

# Results

---

- Comparing unlike cities  
(with Cologne/Jülich as reference)

	Address	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Shanghai, China	31.232274	121.469175	1	Chinese Restaurant	Coffee Shop	Office	Theater	Art Museum	Japanese Restaurant	Shopping Mall	Tea Room	Convenience Store	Italian Restaurant
1	Tokyo, Japan	35.682839	139.759455	1	Historic Site	Japanese Restaurant	Office	Convenience Store	Italian Restaurant	Coworking Space	Chinese Restaurant	Hotel Bar	Police Station	Lounge
2	Johannesburg, South Africa	-26.205000	28.049722	1	Office	Building	Café	Automotive Shop	Clothing Store	Portuguese Restaurant	Fast Food Restaurant	Bank	Coworking Space	Breakfast Spot
3	Sao Paulo, Brazil	-23.550651	-46.633382	2	Office	Pharmacy	College Academic Building	Historic Site	Building	Cosmetics Shop	Courthouse	Church	Monument / Landmark	Spa
4	Cologne, Germany	50.938361	6.959974	1	Hotel	Office	Bar	Plaza	German Restaurant	Italian Restaurant	Gay Bar	Residential Building (Apartment / Condo)	Road	Brewery
5	Juelich, Germany	50.922093	6.361102	0	Miscellaneous Shop	Pizza Place	Salon / Barbershop	Café	Optical Shop	Pharmacy	Bank	Clothing Store	Doner Restaurant	Bakery

São Paulo is still in an independent cluster

# Results

---

- Comparing unlike cities

	Shanghai, China	Tokyo, Japan	Johannesburg, South Africa	Sao Paulo, Brazil	Cologne, Germany	Juelich, Germany
Shanghai, China	0.000000	23.110470	24.129704	24.264237	23.550396	23.752523
Tokyo, Japan	23.110470	0.000000	23.546244	24.225114	23.327369	23.288283
Johannesburg, South Africa	24.129704	23.546244	0.000000	25.396546	24.318102	23.828016
Sao Paulo, Brazil	24.264237	24.225114	25.396546	0.000000	24.630309	25.496425
Cologne, Germany	23.550396	23.327369	24.318102	24.630309	0.000000	22.966630
Juelich, Germany	23.752523	23.288283	23.828016	25.496425	22.966630	0.000000

**Jülich and Cologne are the closest ones  
(used as a reference)**

**São Paulo has the largest distance from Jülich**

# Results

---

- Comparing unlike cities (without reference)

	Address	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Shanghai, China	31.232274	121.469175	2	Chinese Restaurant	Coffee Shop	Office	Convenience Store	Noodle House	Italian Restaurant	Shopping Mall	Japanese Restaurant	Theater	Art Museum
1	Tokyo, Japan	35.682839	139.759455	1	Historic Site	Japanese Restaurant	Convenience Store	Office	Police Station	Italian Restaurant	Lounge	Coworking Space	Hotel Bar	Chinese Restaurant
2	Johannesburg, South Africa	-26.205000	28.049722	1	Office	Building	Café	Automotive Shop	Clothing Store	Bank	Fast Food Restaurant	Portuguese Restaurant	Coworking Space	Doctor's Office
3	Sao Paulo, Brazil	-23.550651	-46.633382	0	Office	Pharmacy	College Academic Building	Cosmetics Shop	Building	Historic Site	Church	Courthouse	Music Venue	Road

	Shanghai, China	Tokyo, Japan	Johannesburg, South Africa	Sao Paulo, Brazil
Shanghai, China	0.000000	22.822803	22.040519	21.559651
Tokyo, Japan	22.822803	0.000000	23.566859	24.281156
Johannesburg, South Africa	22.040519	23.566859	0.000000	21.715092
Sao Paulo, Brazil	21.559651	24.281156	21.715092	0.000000

São Paulo { similar to Shanghai and Johannesburg  
most different from Tokyo

(effect of the feature scaling when including the reference cities)

# Conclusions

---

- **K-Means clustering and Euclidian distances provide a way to compare cities**
- **Similarities of the cities arise from geographical proximity or historical connections**
- **Size of the city may influence distinction, but is not determinant**
- **Unalike cities may have remarkably similar venue categories**

# Outlook

---

- Improve venues database (e.g., Google Maps)
- Add more features:
  - Cost of living
  - Languages
  - Density or roads/Number of cars
- Decision Tree algorithm