Comparing Neighborhoods from Coast to Coast

By Yiwei Wang

Introduction

Compare cities:

- How do we do it?
  Access to venue categories in neighborhoods

- Who can benefit from the analysis

  - Individual relocation decisions

  - Small businesses

Data

- Neighborhood names

  Source: Wikipedia

- Location of neighborhood centers

  Google geocoding service (API)

- Venues in the neighborhood

  Foursquare.com (API)

Methodology

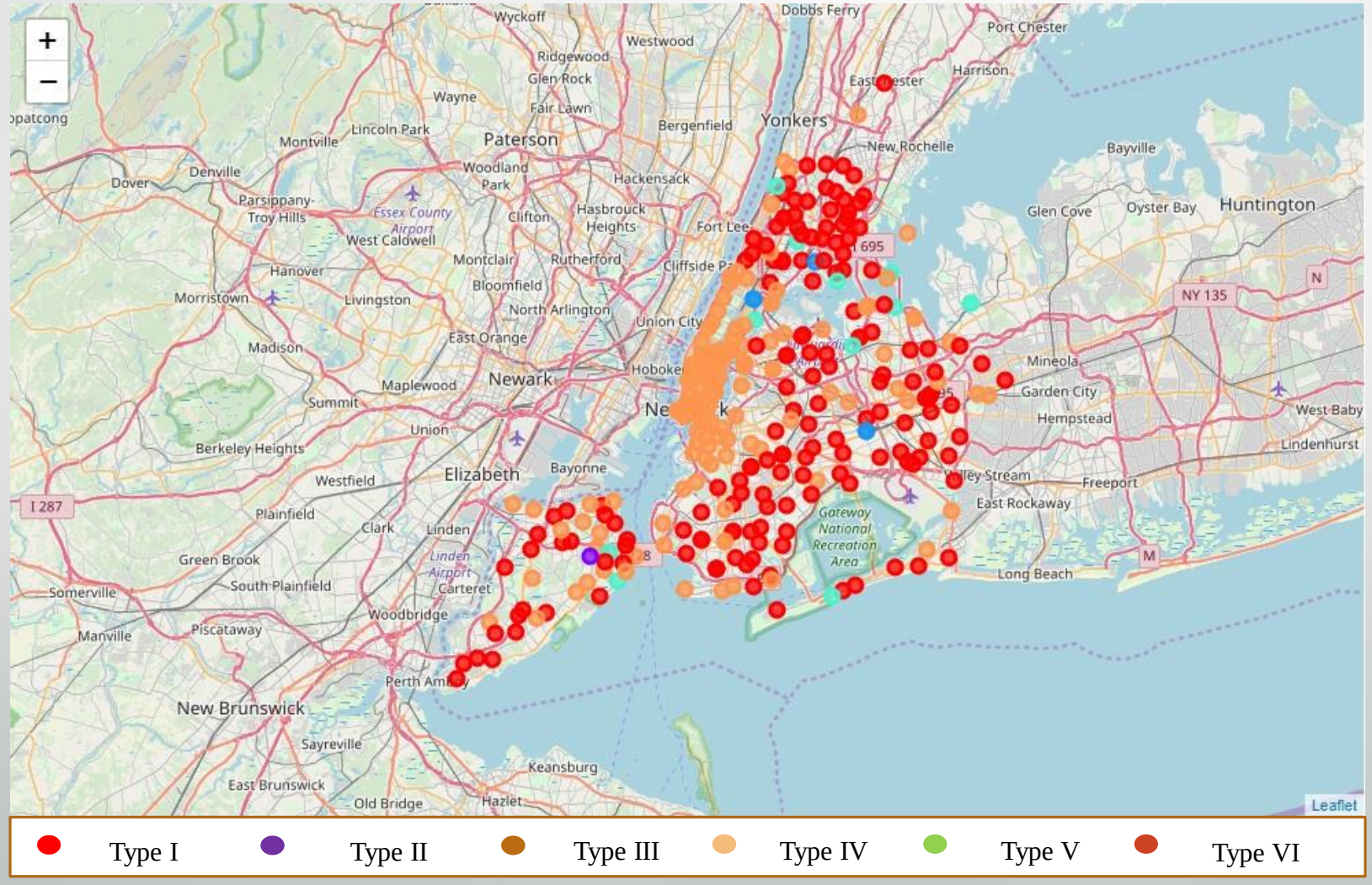
Machine learning technique:

K-means clustering

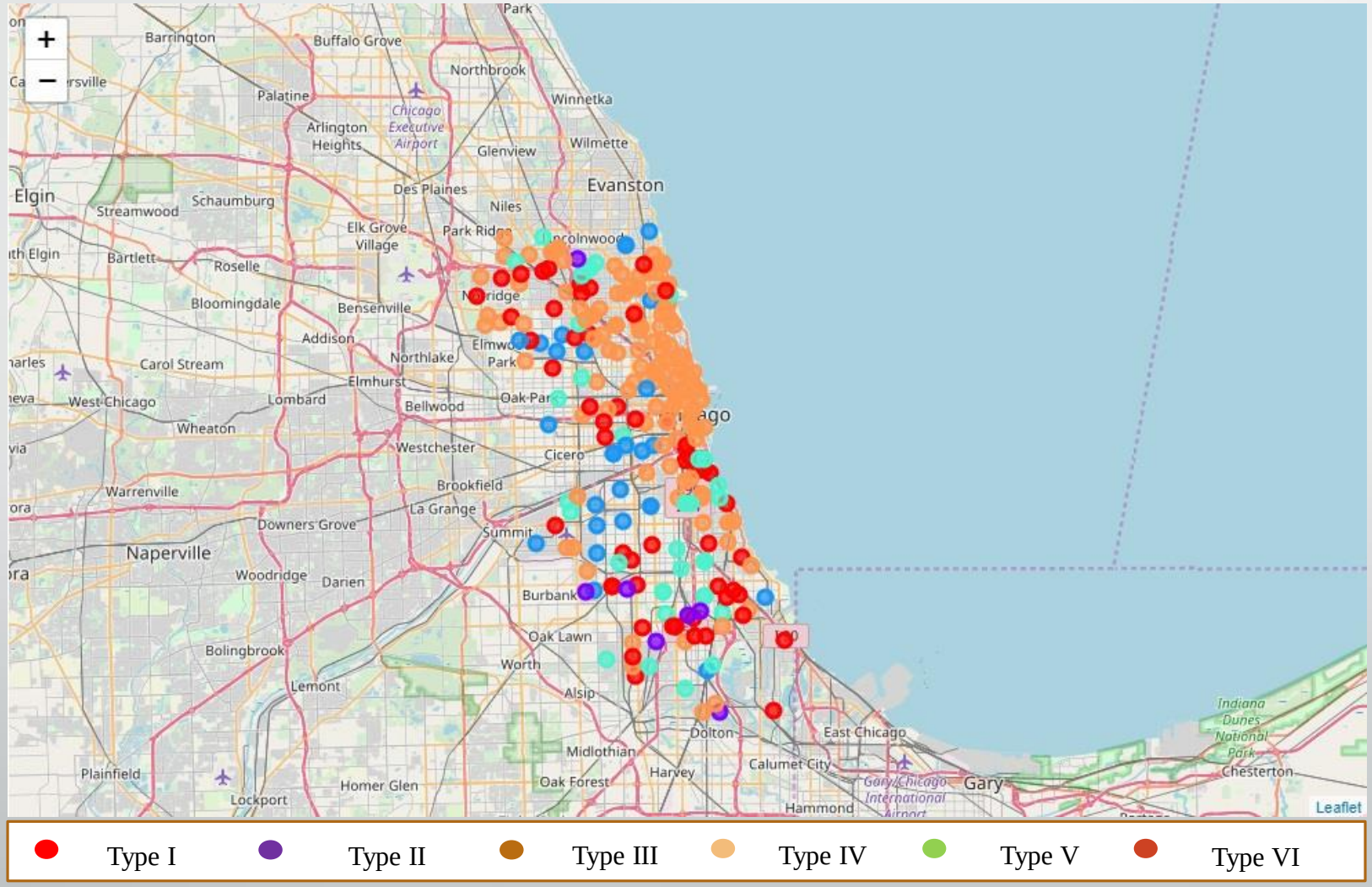
Minimize within-cluster variance

$$V_m = \sum_{i=1}^{n} \sum_{j=1}^{S} (x_{ij} - C_{mj})^2$$

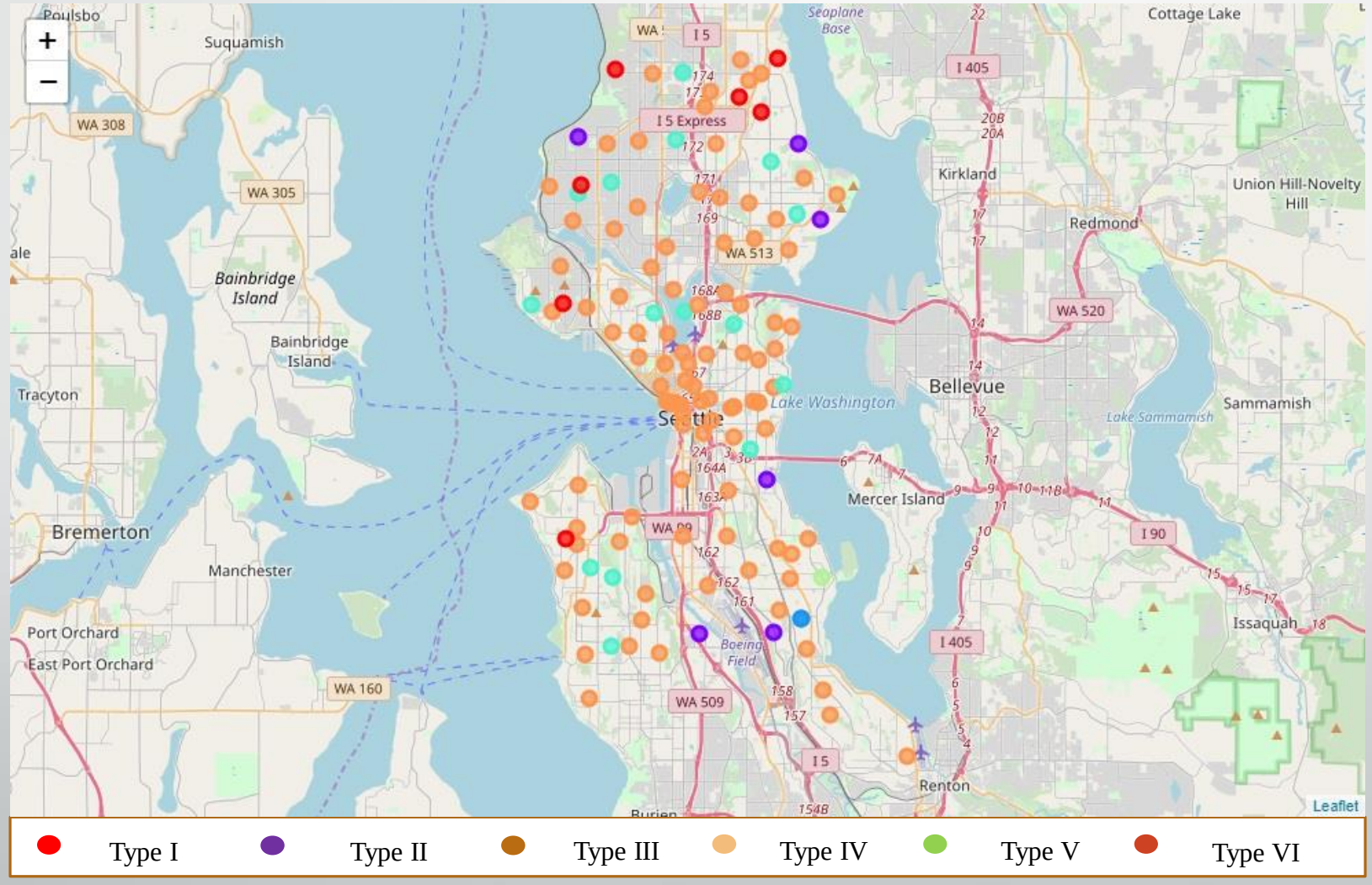
Results – New York



Results – Chicago



Results – Seattle



Discussion

Composition of neighborhood types in three cities

	Type I	Type II	Type III	Type IV	Type V	Type VI
Chicago	22.54%	2.87%	11.07%	13.93%	0.00%	49.59%
New York	51.88%	0.31%	1.56%	4.06%	0.31%	41.88%
Seattle	5.56%	4.76%	1.59%	11.90%	0.79%	75.40%

Variance between cities:

New York - Chicago: 1115
New York - Seattle: 3351
Seattle – Chicago: 1053

Seattle and Chicago are more similar!

Discussion (cont'd)

Top venues in different types of clusters.

	Pizza Place	Fast Food Restaurant	Pharmacy	Sandwich Place	Chinese Restaurant	Donut Shop	Deli / Bodega	Factory	Fabric Shop	Bank
Type I	4.71%	3.04%	2.75%	2.63%	2.43%	2.40%	2.19%	2.05%	2.02%	2.02%

	Coffee Shop	Bar	Pizza Place	Bakery	Italian Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Park	Gym
Type VI	4.10%	2.78%	2.69%	2.04%	2.04%	1.92%	1.85%	1.77%	1.77%	1.56%

Conclusion

- 690 neighborhoods in New York, Chicago, and Seattle are clustered into 6 groups based on their access to different venue categories using k-means clustering.
- Chicago and Seattle appear to have more in common in terms of their composition of various neighborhood types.
- The methods applied in this project can be further refined and extended to solve business problems as well as help individuals make relocation decisions.