The Livehoods Project

Utilizing Social Media to Understand the Dynamics of a City

Justin Cranshaw
jcransh@cs.cmu.edu

Raz Schwartz
razs@andrew.cmu.edu

Jason I. Hong
jasonh@cs.cmu.edu

Norman Sadeh
sadeh@cs.cmu.edu

School of Computer Science
Carnegie Mellon University

@livehoods | livehoods.org

Neighborhoods

Neighborhoods

- Neighborhoods provide **order to the chaos** of the city
- Help us determine: *where* to **live** / **work** / **play**
- Provide **haven** / **safety** / **territory**
- Municipal government: **organizational unit** in resource allocation
- Centers of **commerce** and economic **development**. **Brands**.
- A sense of **cultural identity** to residents

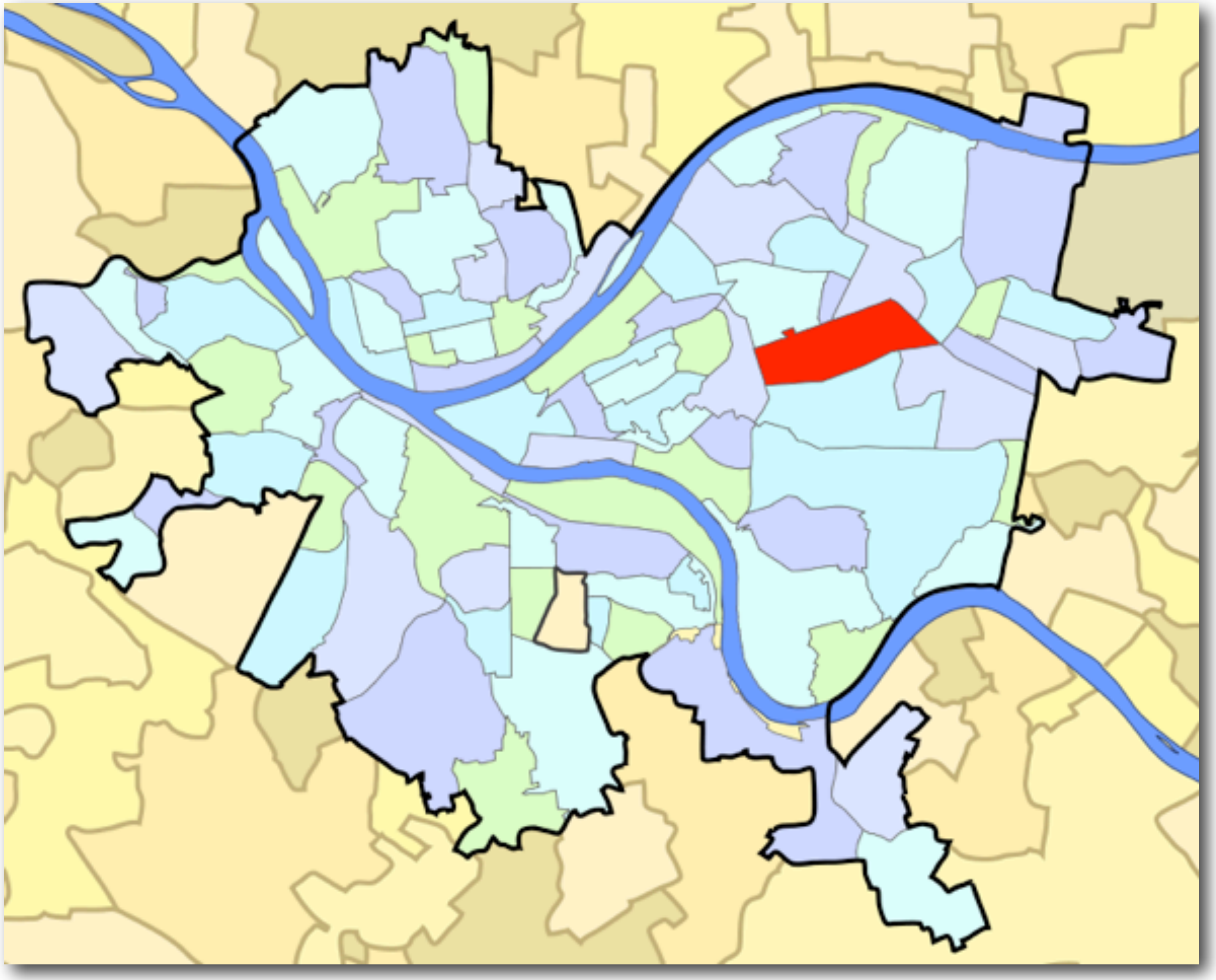
What comes to mind when you picture **your neighborhood**?

The Image of a Neighborhood

You're probably **not** imagining this.

The Image of a Neighborhood

You're probably **not** imagining this.



The Image of a Neighborhood

What you're imagining most likely
looks a lot more like this.

The Image of a Neighborhood

What you're imagining most likely looks a lot more like this.

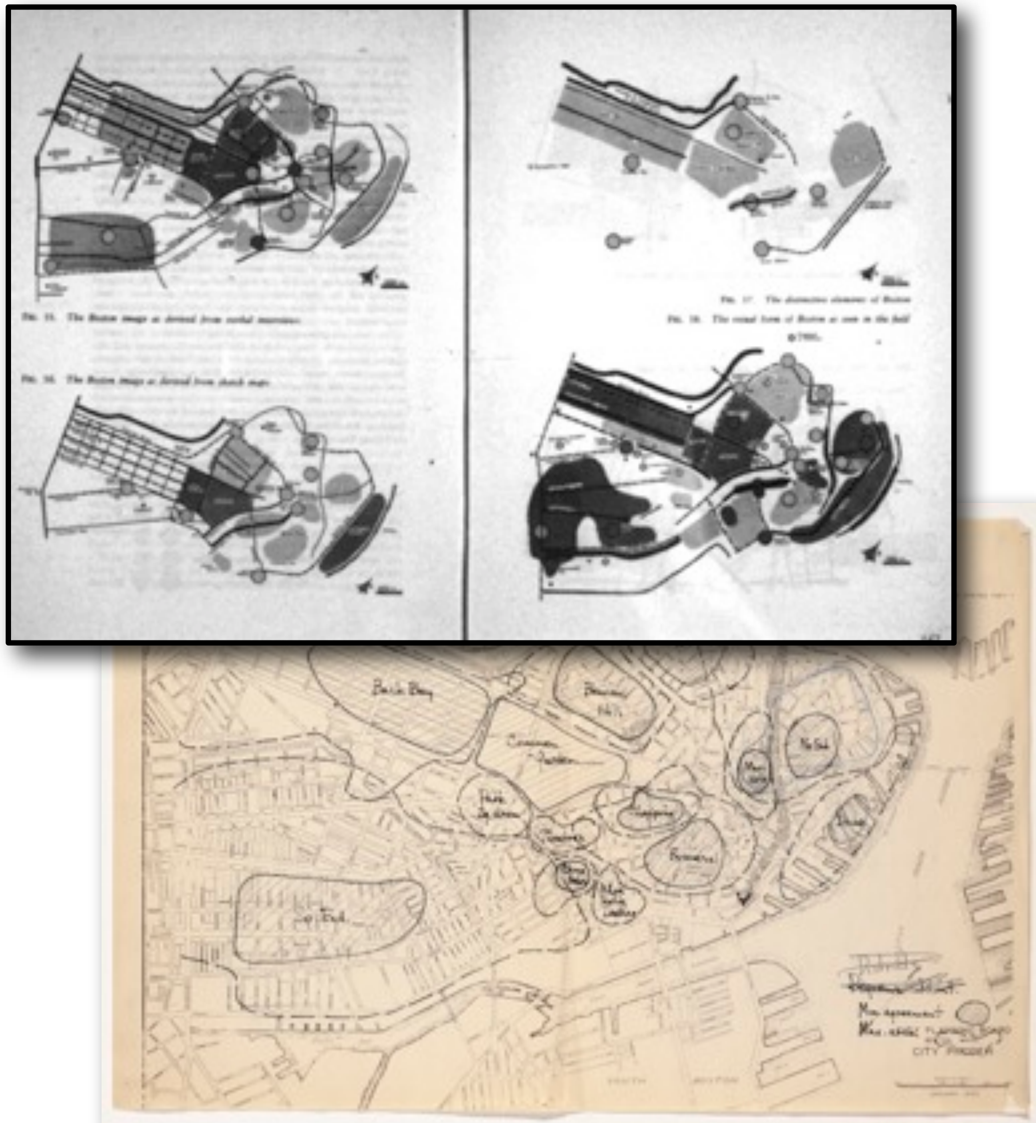
The Image of a Neighborhood

What you're imagining most likely looks a lot more like this.

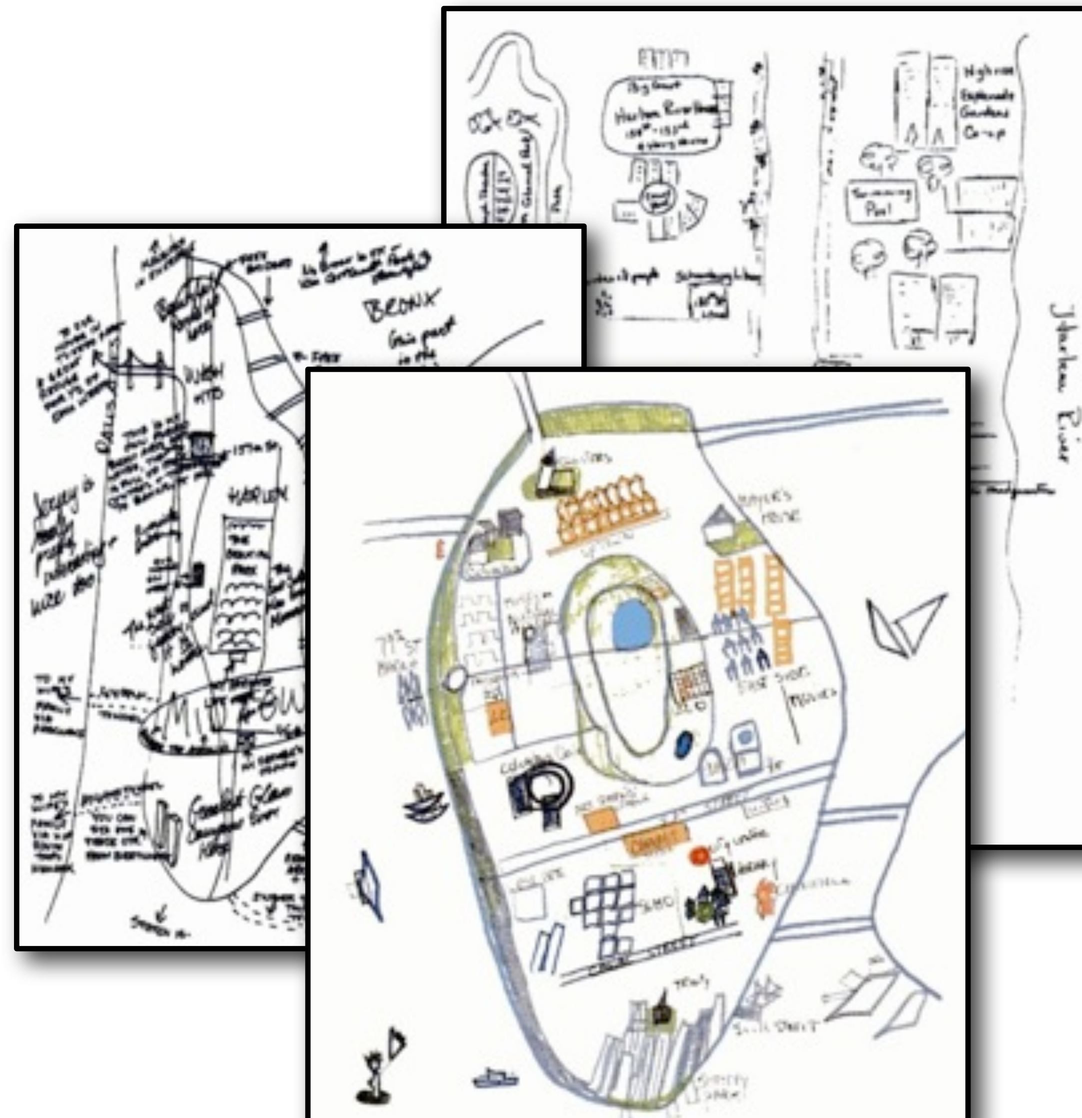
Every citizen has had long associations with some part of his city, and his image is soaked in memories and meanings.

---*Kevin Lynch, The Image of a City*

Studying Perceptions: Cognitive Maps



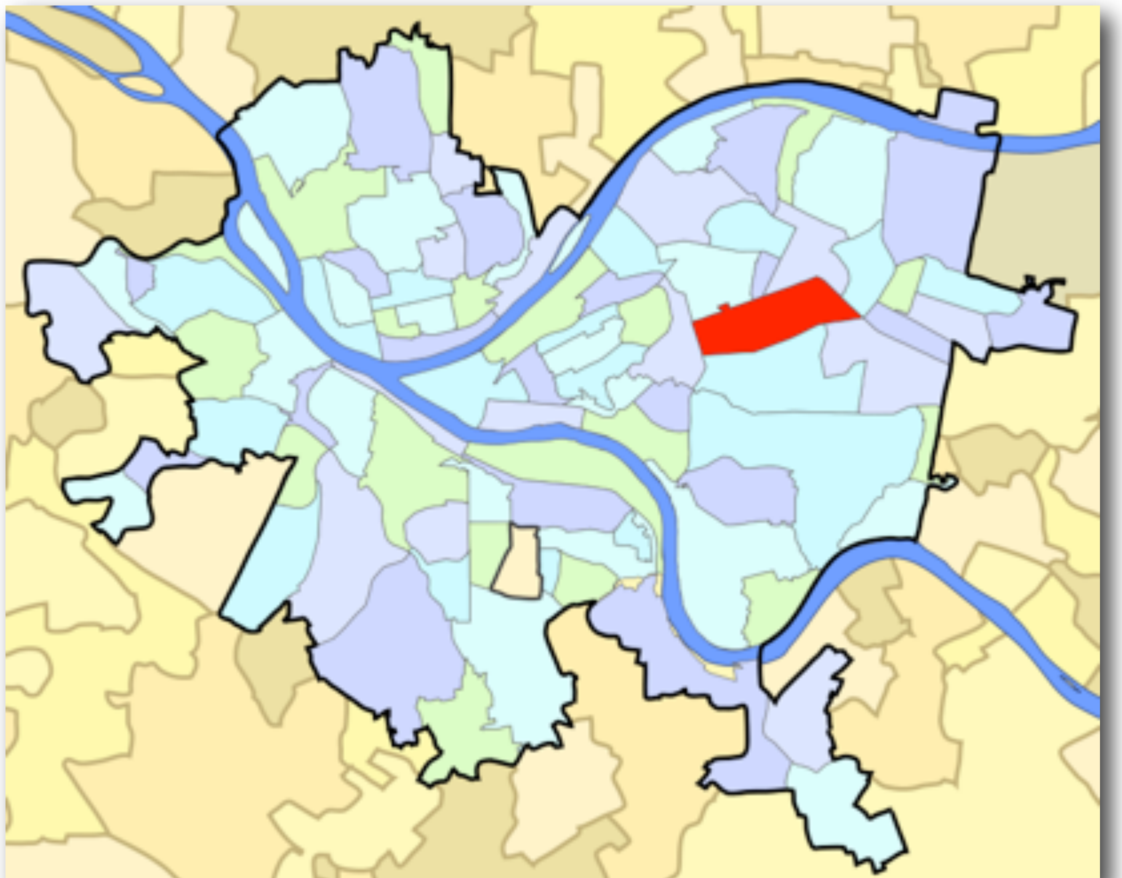
Kevin Lynch, 1960



Stanley Milgram, 1977

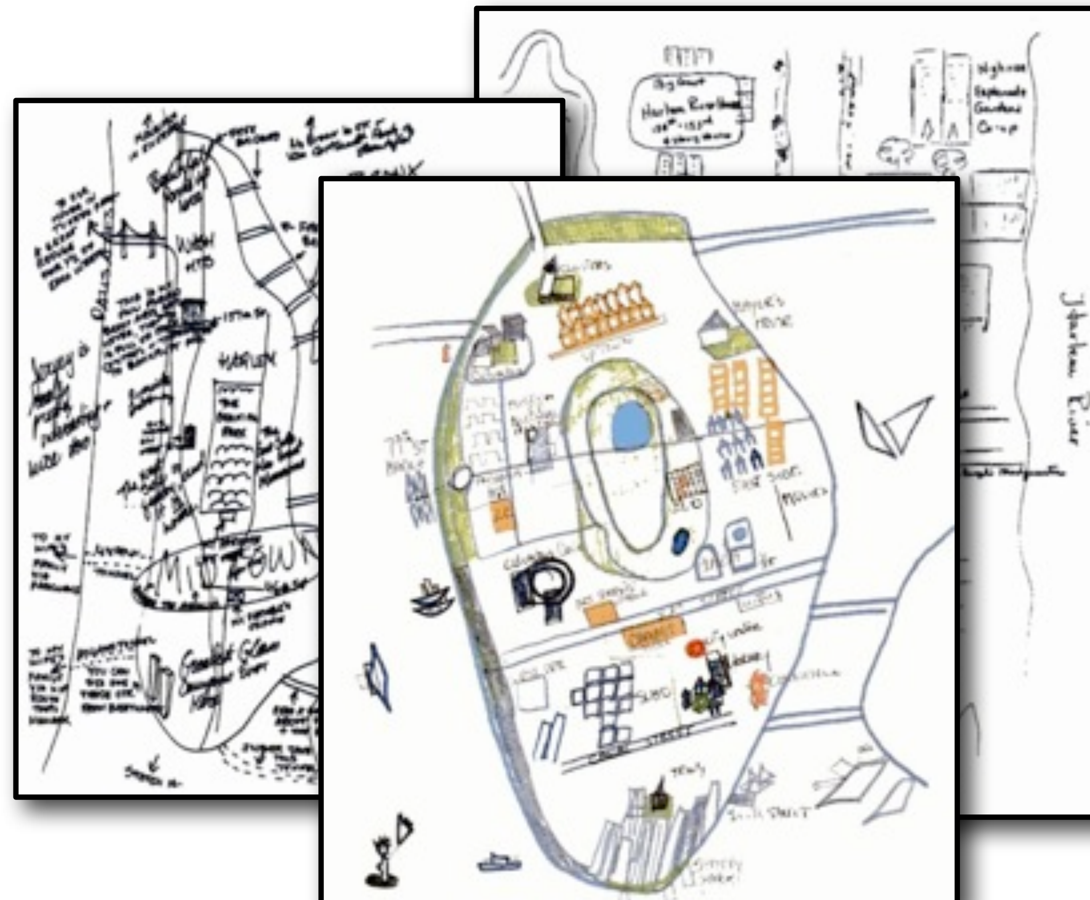
Two Perspectives

"Politically constructed"



Neighborhoods have fixed borders defined by the city government.

"Socially constructed"



Neighborhoods are organic, cultural artifacts. Borders are blurry, imprecise, and may be different to different people.

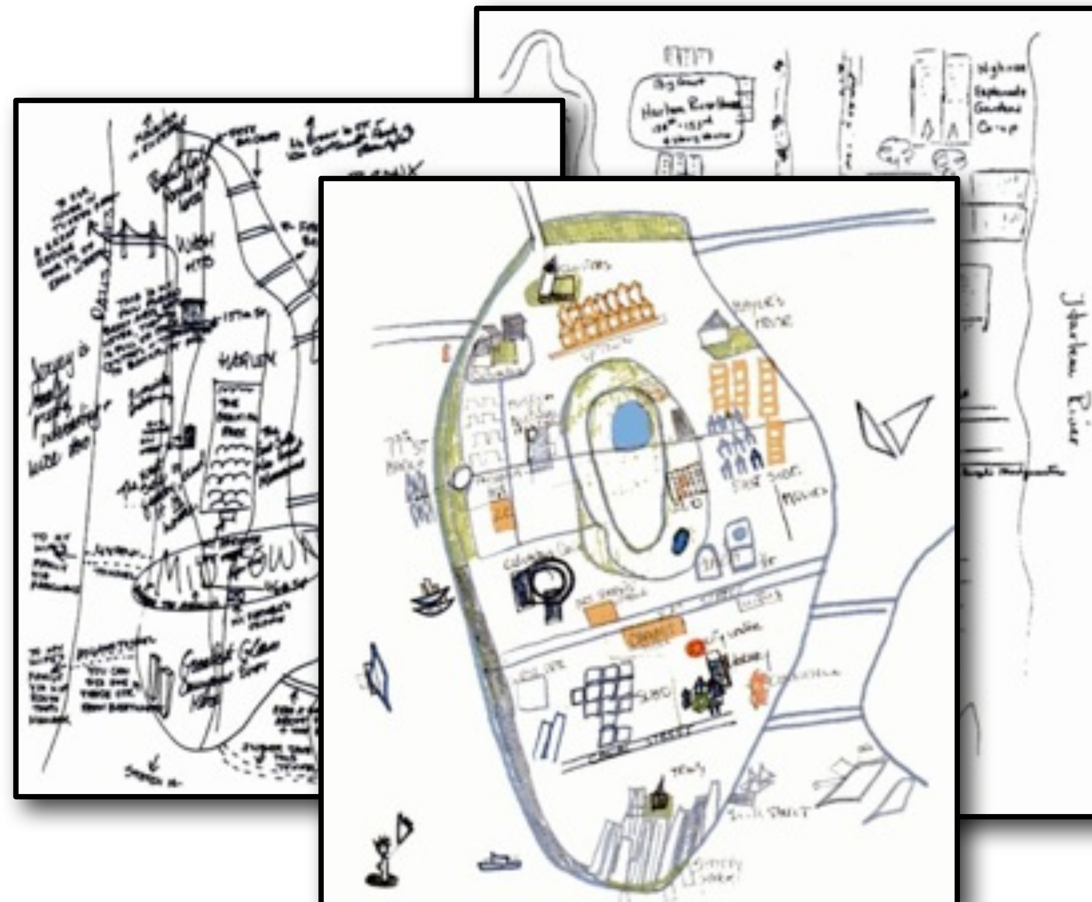
Collective Cognitive Maps

Can we discover automated ways of identifying the “organic” boundaries of the city?

Can we extract local cultural knowledge from social media?

Can we build a collective cognitive map from data?

“Socially constructed”



Neighborhoods are organic, cultural artifacts. Borders are blurry, imprecise, and may be different to different people.

Observing the City:
SmartPhones

We seek to leverage location-based mobile social networks such as foursquare, which let users broadcast the places they visit to their friends via check-ins.

foursquare

Check-in


Round Corner Cantina
+ ADD TO MY TO-DO LIST
3723 Butler St
at 38th St
Pittsburgh, PA 15201
(412) 904-2279
Twitter
Like 5
Map
Arsenal Park
Lower Lawrenceville
Categories: Bar, Tacos, Mexican
Tags: fancy cocktails, pimm's cup, jukebox, vegetarian, seviche, margaritas, friendly bartenders, hipsters, bar, outdoor seating, happy hour, tacos
edit

Hypothesis

- The **character of an urban** area is **defined** not just by the the types of places found there, but also **by the people who make the area part of their daily routine.**
- Thus we can characterize a place by observing the people that visit it.
- To discover areas of unique character, we should look for clusters of nearby venues that are visited the same people

The moving elements of a city, and in particular the people and their activities, are as important as the stationary physical parts.

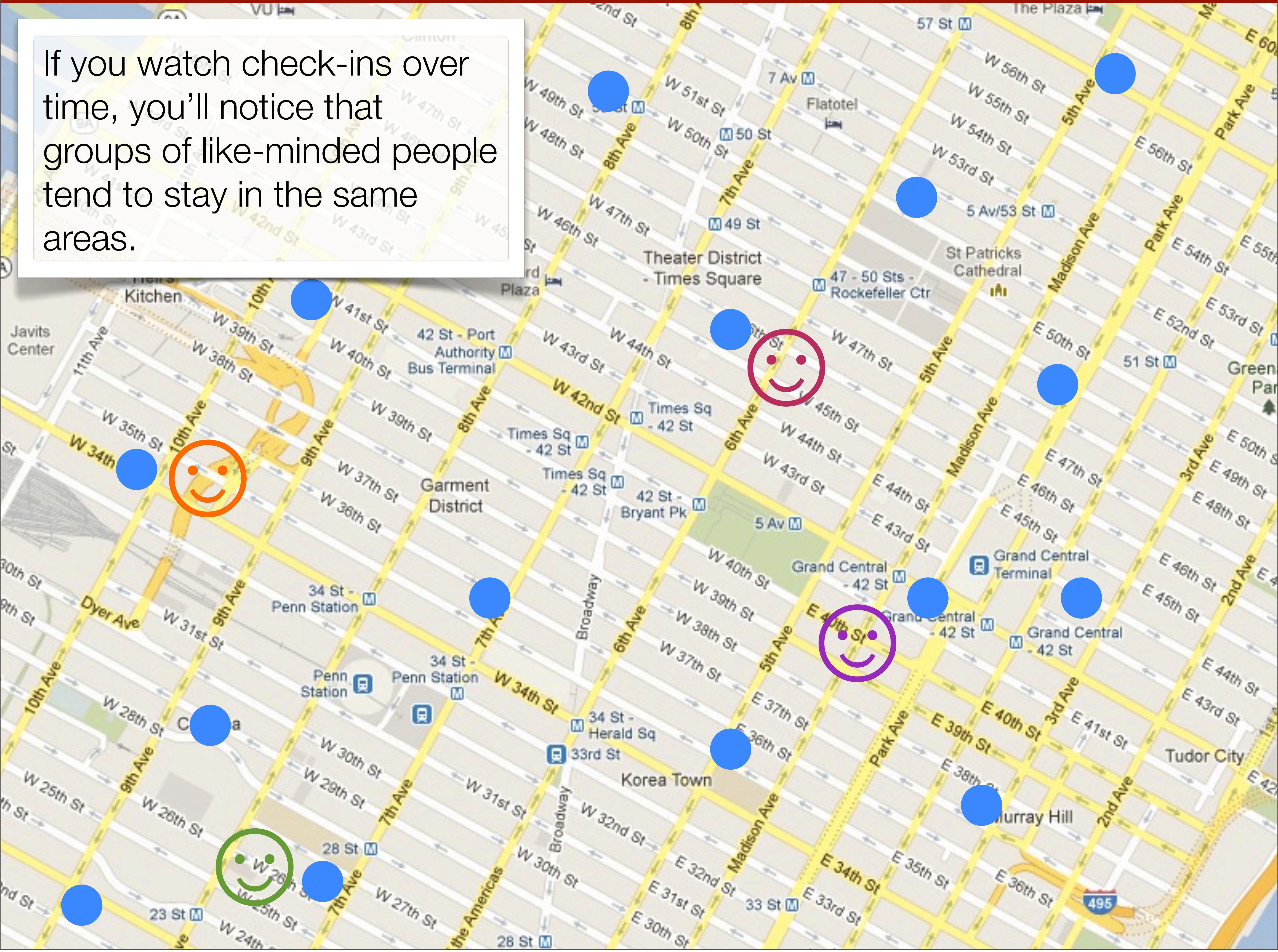
---Kevin Lynch, The Image of a City

Clustering The City

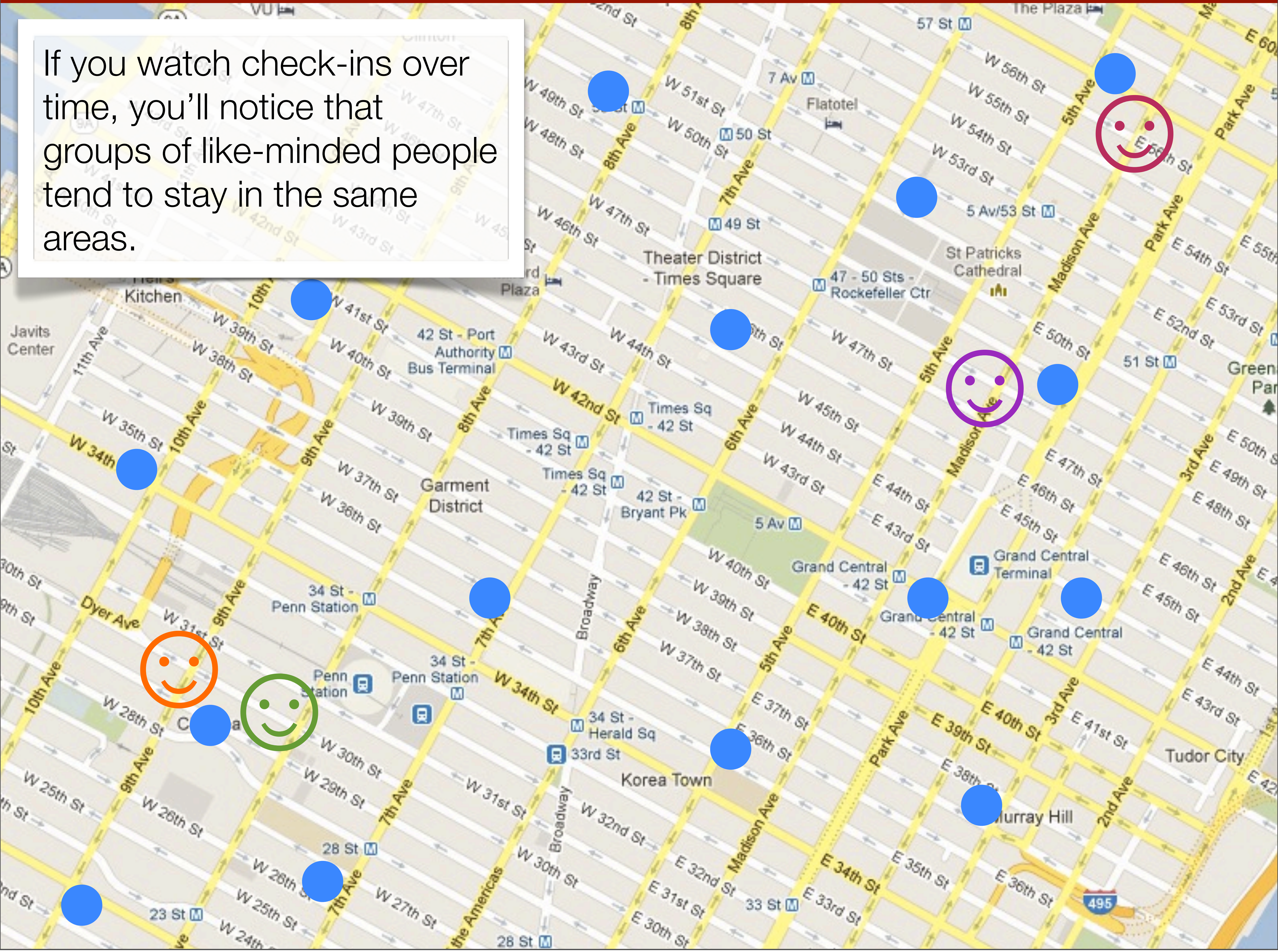
- Clusters should be of geographically contiguous foursquare venues.
- Clusters should have a distinct character from one another, perceivable by city residents.
- Clusters should be such that venues within a cluster are more likely to be visited by the same users than venues in different clusters.

Clustering Intuition

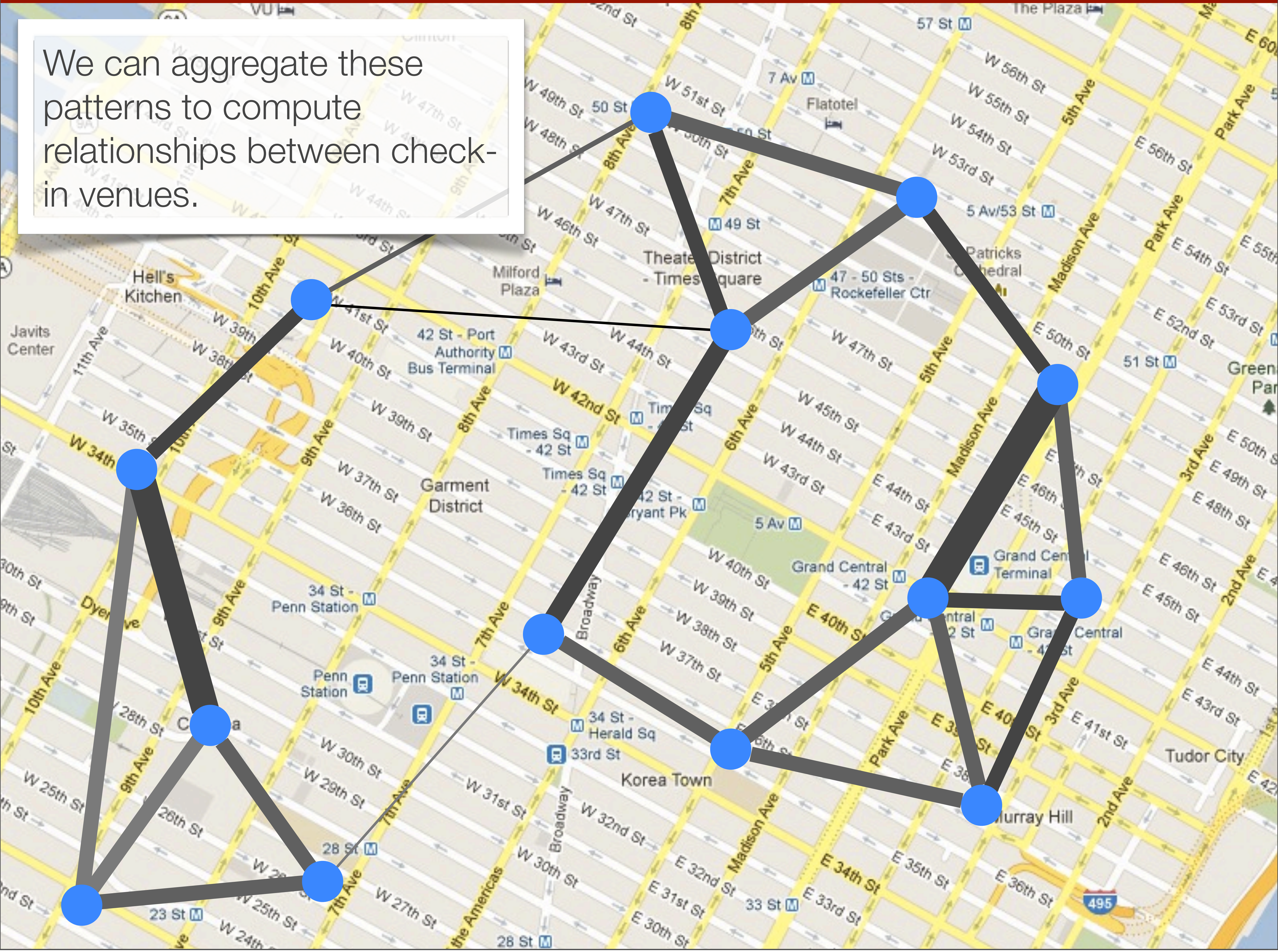
If you watch check-ins over time, you'll notice that groups of like-minded people tend to stay in the same areas.



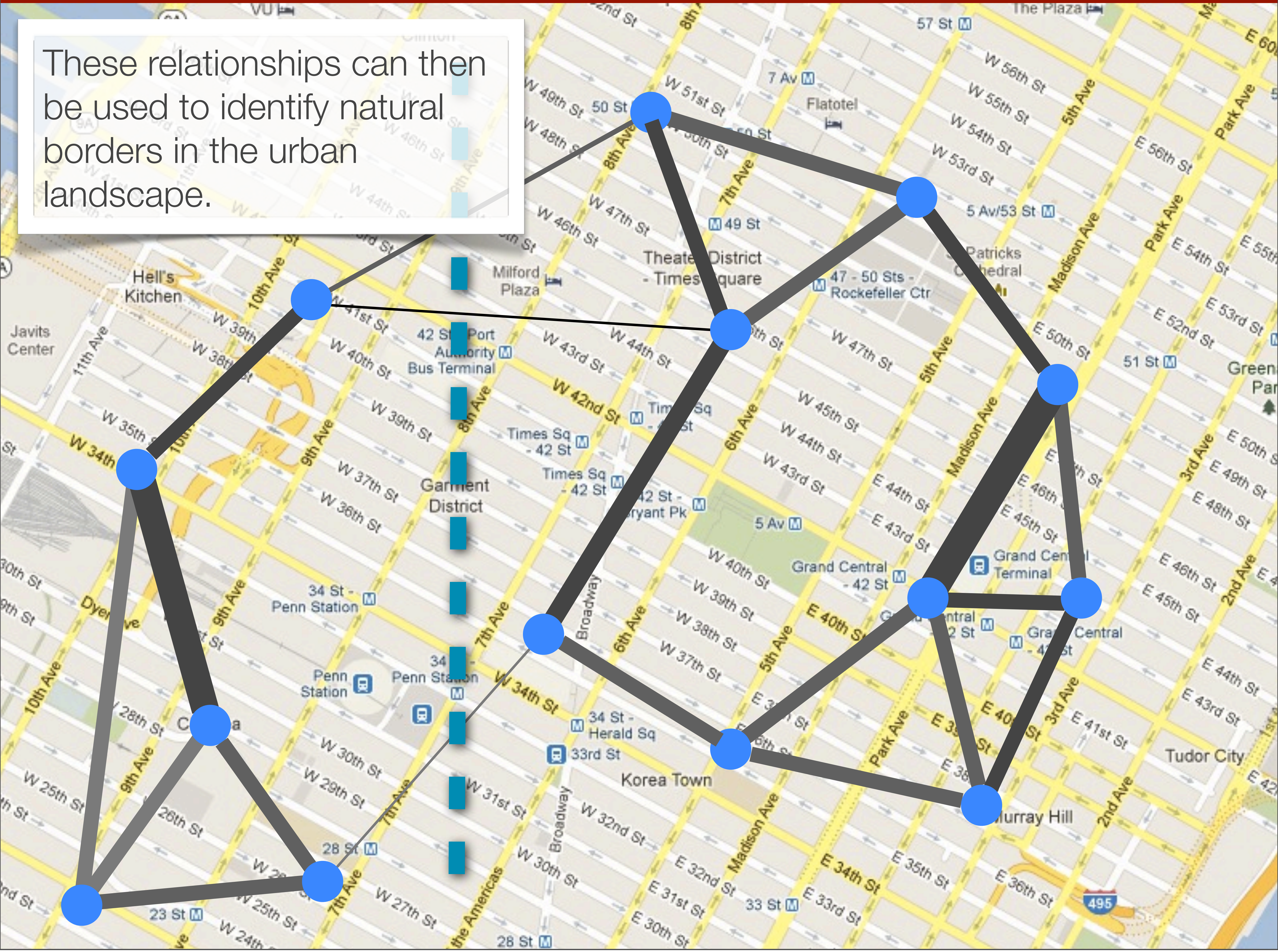
If you watch check-ins over time, you'll notice that groups of like-minded people tend to stay in the same areas.



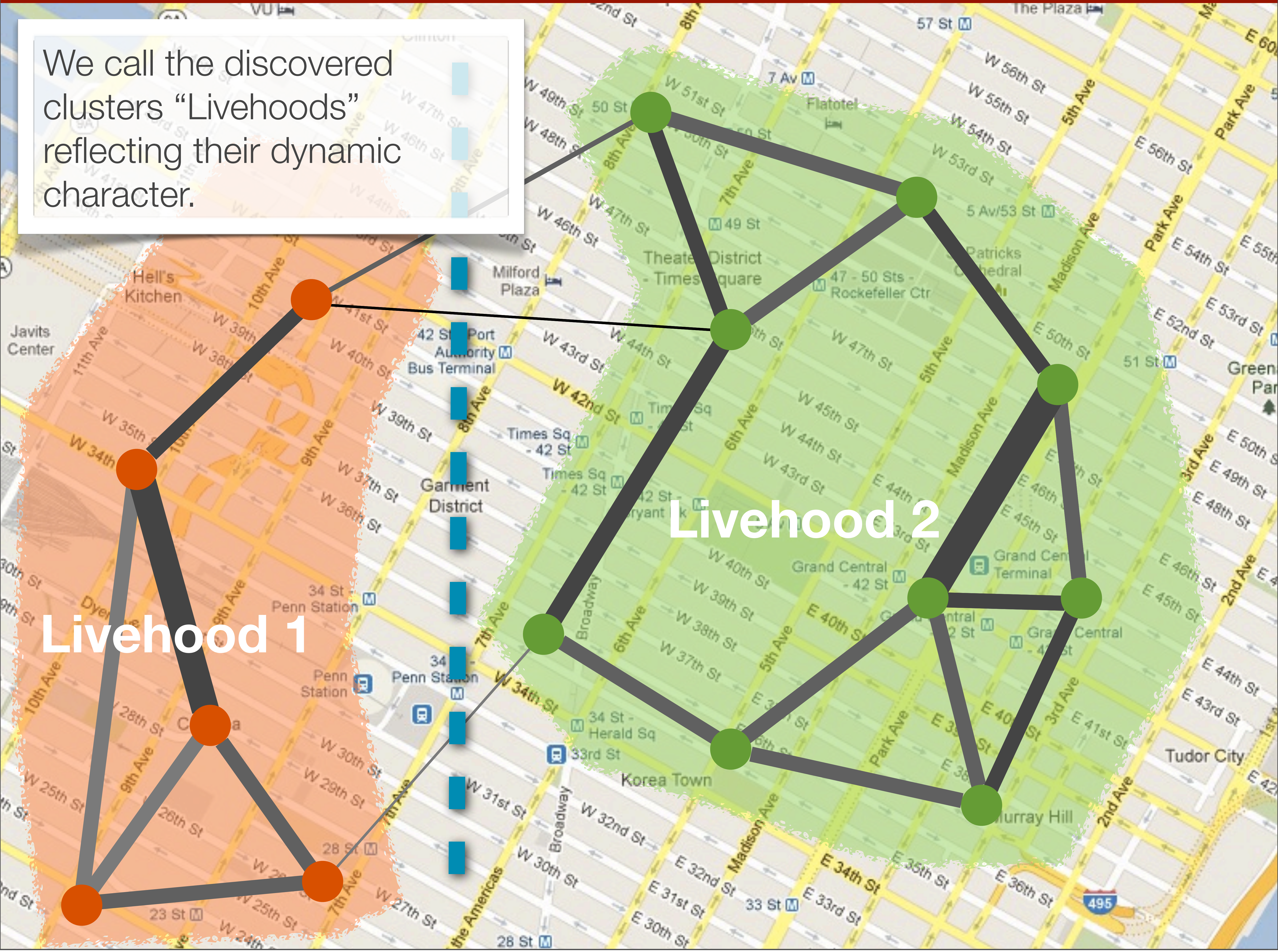
We can aggregate these patterns to compute relationships between check-in venues.



These relationships can then be used to identify natural borders in the urban landscape.



We call the discovered clusters “Livehoods” reflecting their dynamic character.



Clustering Methodology

Social Venue Similarity

We can get a notion of how similar two places are by looking at who has checked into them.

c_v is a vector where the u^{th} component counts the number of times user u checks in to venue v.

We can think of c_v as the **bag of checkins** to venue v.

We can then look at the cosine similarity between venues.

$$s(i,j) = \frac{c_i \cdot c_j}{||c_i|| \; ||c_j||}$$

$$c_v = \begin{bmatrix} \#\text{ of } u_1 \text{ checkins } v \\ \#\text{ of } u_2 \text{ checkins to } v \\ \vdots \\ \#\text{ of } u_{n_U} \text{ checkins to } v \end{bmatrix}$$

Problems With This Approach

(1) The resulting similarity graph is **quite sparse**

(2) Similarity tends to be dominated by **"hub" venues**

Venue Affinity Matrix

- In addition to capturing the **social similarity** between places, we want clusters to be **geographically contiguous**
- We also need to overcome the limitations of social similarity (sparsity and hub biases)

We derive an affinity matrix A that blends social affinity with spatial proximity:

$$A = (a_{i,j})_{i,j=1,\dots,n_V}$$

$$a_{i,j} = \begin{cases} s(i,j) + \alpha & \text{if } j \in N_m(i) \text{ or } i \in N_m(j) \\ 0 & \text{otherwise} \end{cases}$$

$N_m(i)$ are the *m* nearest geographic neighbors to venue *i*.

Restricting to nearest neighbors **overcomes bias by "hub" venues** such as airports, and it **adds geographic contiguity**.

α is a small positive constant that **overcomes sparsity** in pairwise co-occurrence data.

Spectral Clustering

- Given the affinity matrix *A*, we segment check-in venues using well studied spectral clustering techniques
- We use the variation of spectral clustering introduced by **Ng, Jordan, and Weiss** [NIPS 2001]
- We select *k* (number of clusters) as is common by looking for **large gaps in consecutive eigenvalues** between and upper and lower allowable k.

Ng, Jordon, and Weiss

(1) *D* is the diagonal degree matrix

(2) $L = D - A$

(3) $L_{norm} = D^{-1/2} L D^{1/2}$

(4) Find $e_1, \ldots, e_k$, the *k* smallest evects's of L_{norm}

(5) $E = [e_1, \ldots, e_k]$ and let $y_1, \ldots, y_{n_V}$ be the rows of *E*

(6) Clustering $y_1, \ldots, y_{n_V}$ with KMeans induces a clustering $A_1, \ldots, A_k$ of the original data.

Post Processing

- We introduce a post processing step to clean up any degenerate clusters
- We separate the subgraph induced by each A_i into connected components, creating new clusters for each.
- We delete any clusters that span too large a geographic area (“background noise”) and reapportion the venues to the closest non-degenerate cluster by (single linkage) geographic distance

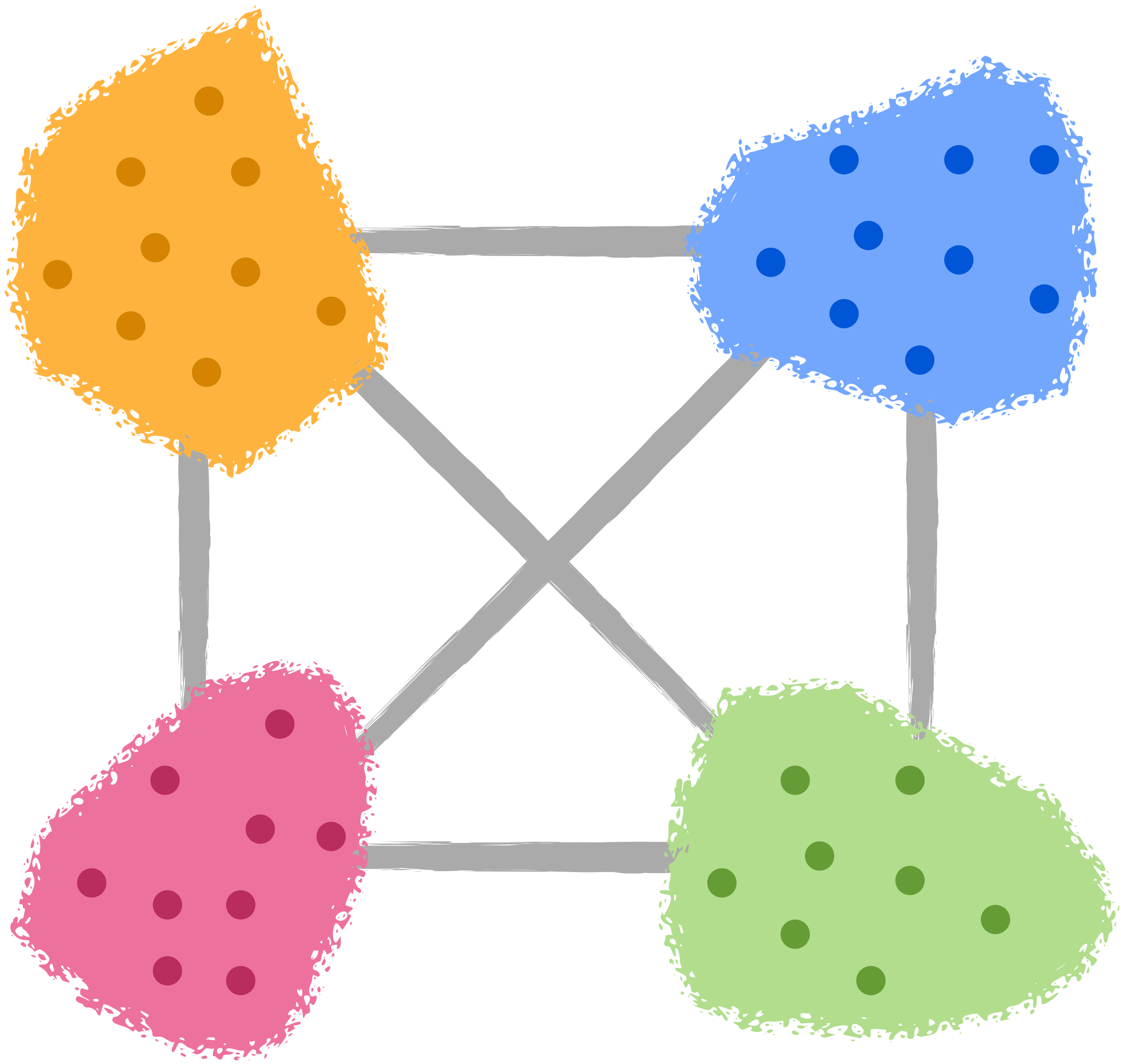
Related Livehoods

To examine how the Livehoods are related to one another, for each pair of Livehoods, we compute a similarity score.

For Livehood A_i the vector c_{A_i} is the **bag of check-ins** to A_i.

Each component measures for a given user u the number of times u has checked in to *any* venue in A_i.

$$s(A_i, A_j) = \frac{c_{A_i} \cdot c_{A_j}}{||c_{A_i}||\ ||c_{A_j}||}$$

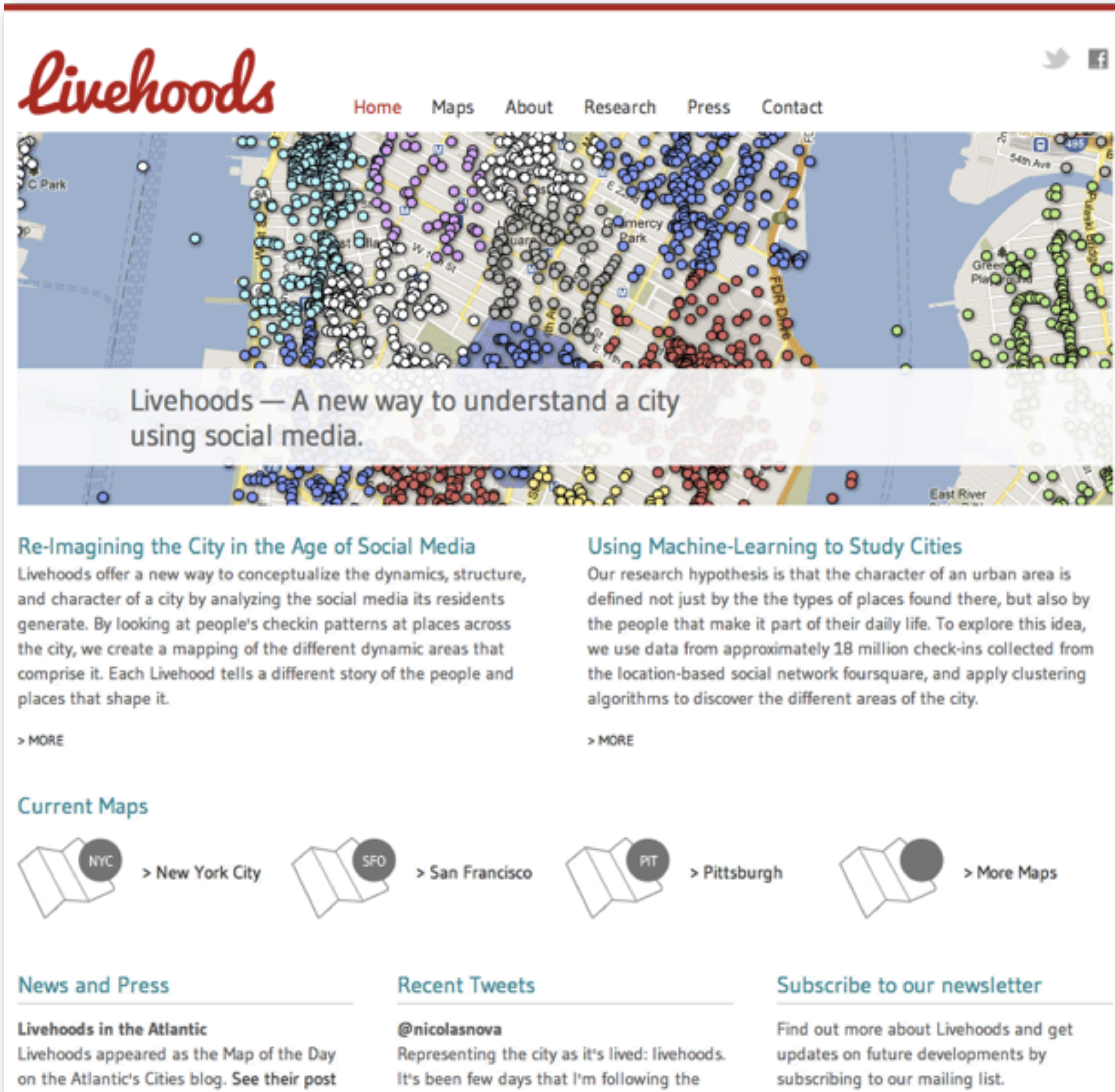


Data

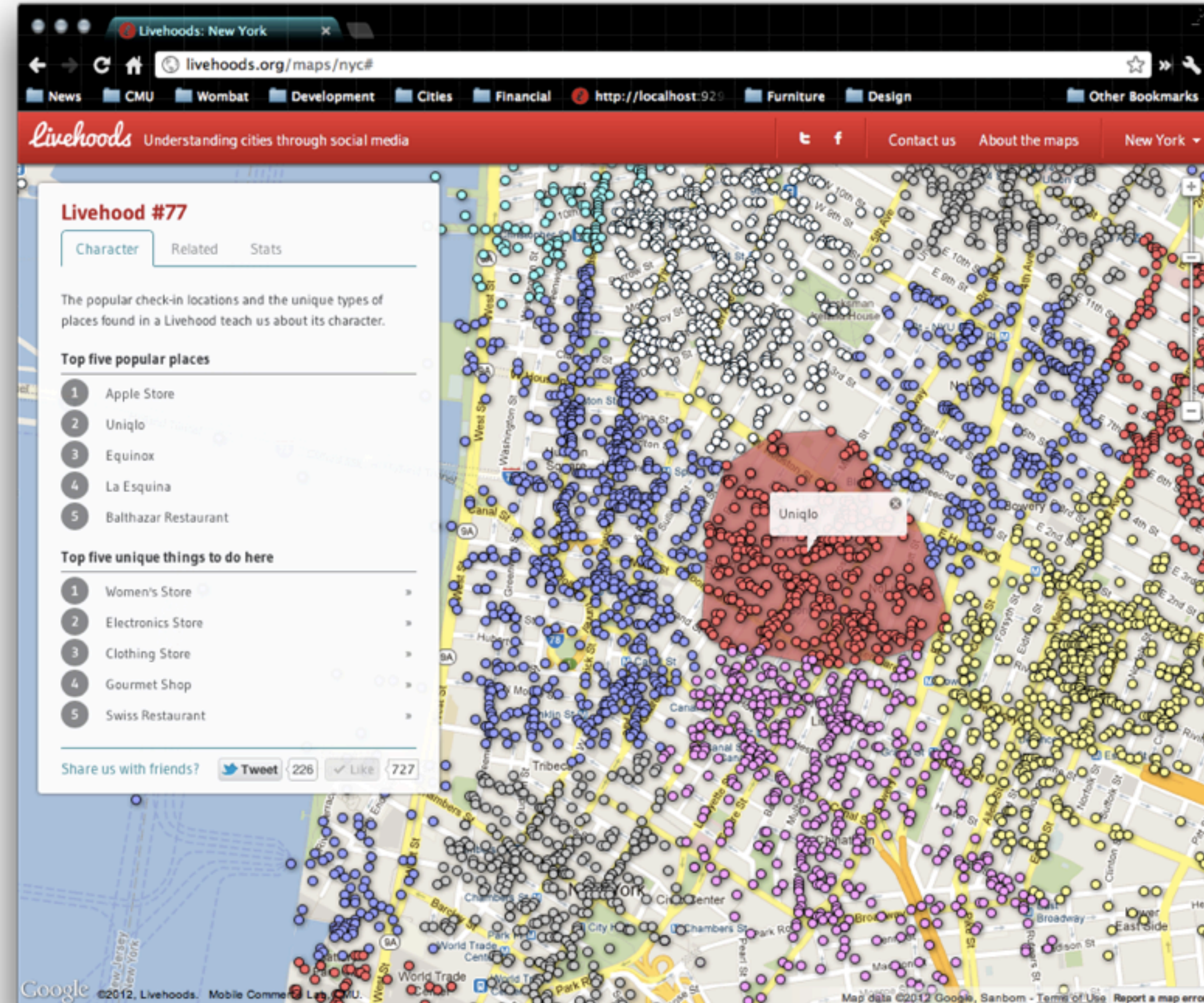
The Data

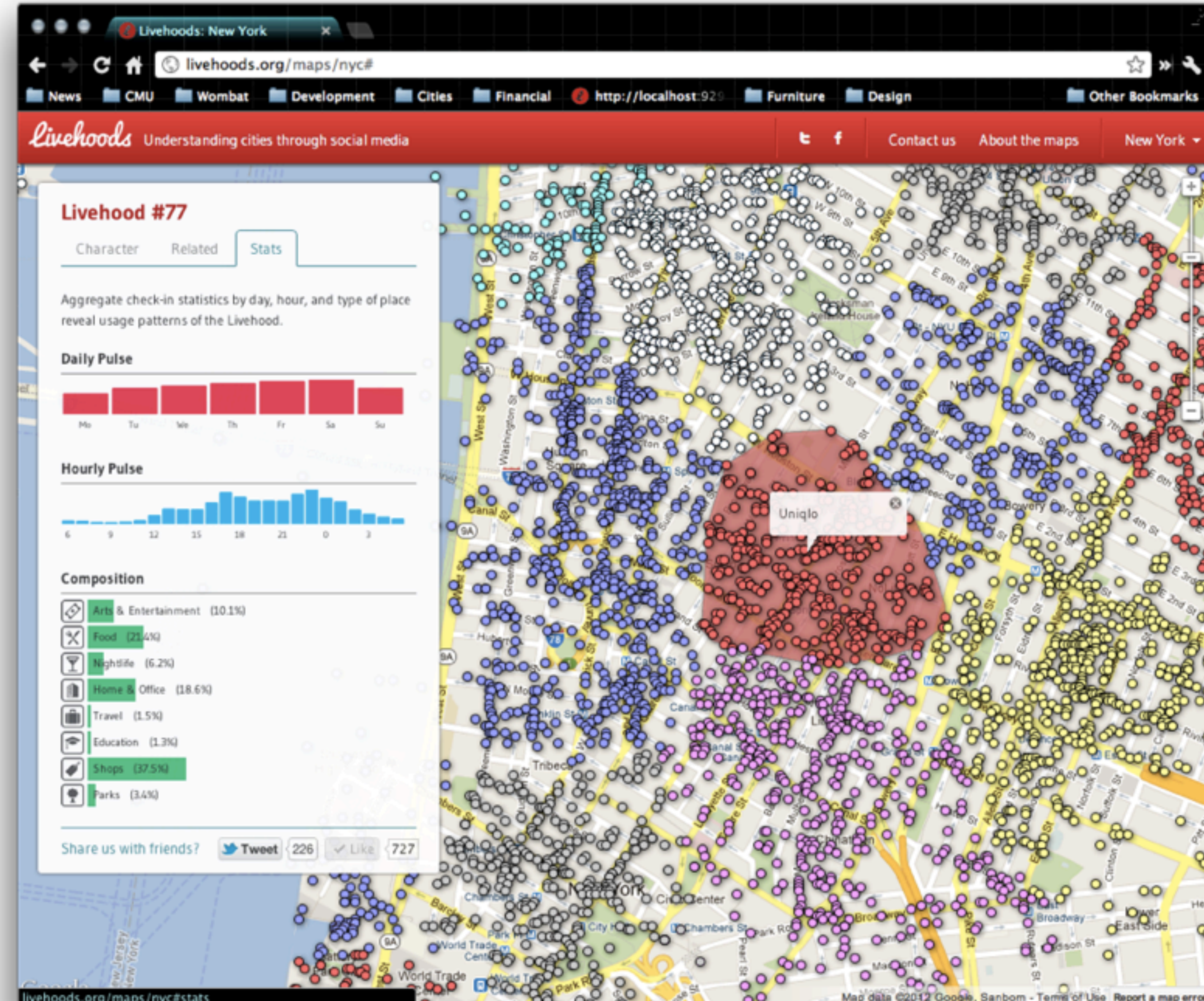
- Foursquare check-ins are by default private
- We can gather check-ins that have been shared publicly on Twitter.
- Combine the 11 million foursquare check-ins from the dataset Chen et al. dataset [ICWSM 2011] with our own dataset of 7 million checkins gathered between June and December of 2011.
- Aligned these Tweets with the underlying foursquare venue data (venue ID and venue category)

livehoods.org

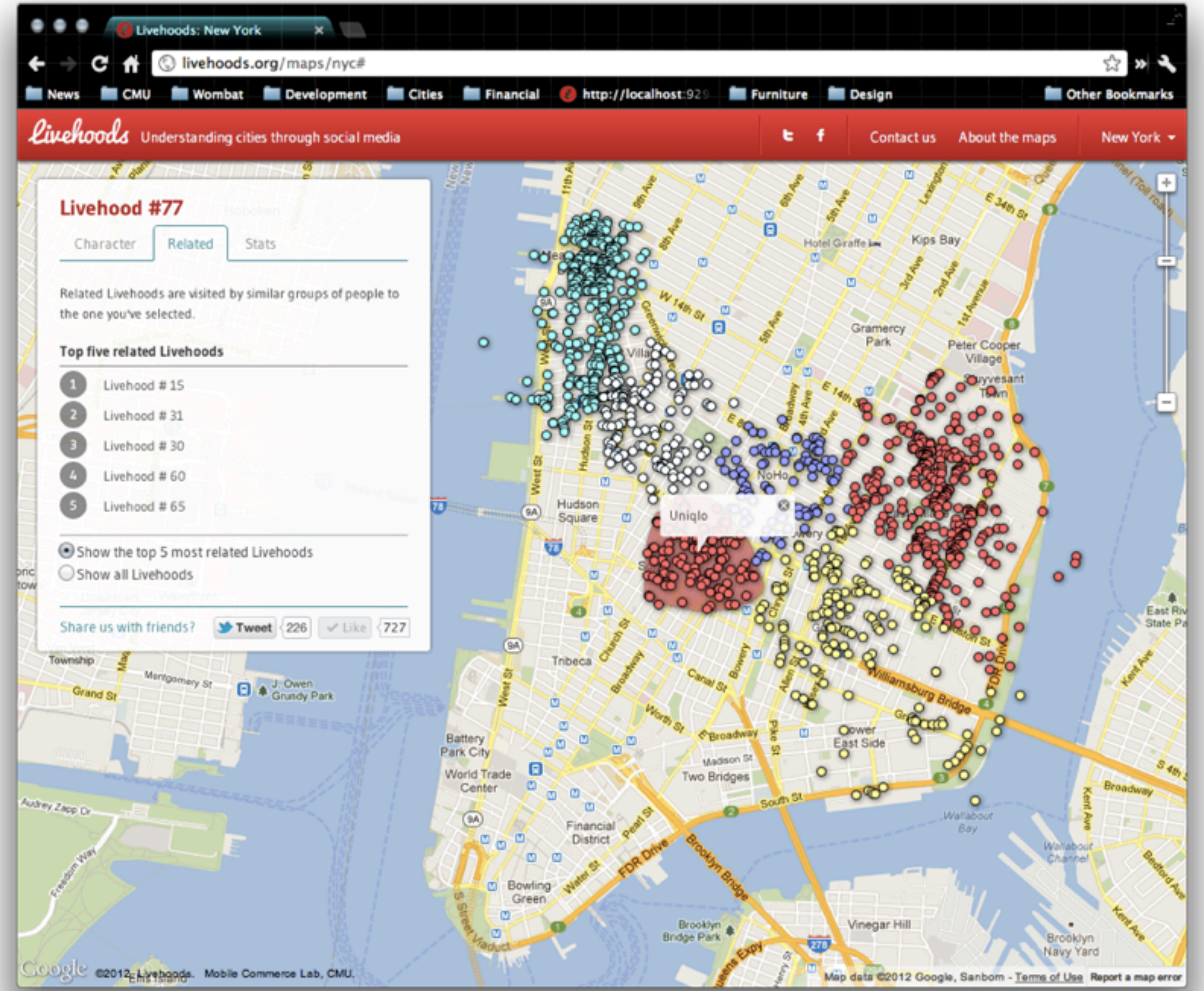


Currently maps of **New York**, **San Francisco Bay**, **Pittsburgh**, and **Montreal**


Livehoods: New York
livehoods.org/maps/nyc#
News
CMU
Wombat
Development
Cities
Financial
Furniture
Design
Other Bookmarks
Livehoods
Understanding cities through social media
Contact us
About the maps
New York
Livehood #77
Character
Related
Stats
The popular check-in locations and the unique types of places found in a Livehood teach us about its character.
Top five popular places
1 Apple Store
2 Uniqlo
3 Equinox
4 La Esquina
5 Balthazar Restaurant
Top five unique things to do here
1 Women's Store
2 Electronics Store
3 Clothing Store
4 Gourmet Shop
5 Swiss Restaurant
Share us with friends?
Tweet 226
Like 727
Uniqlo
©2012, Livehoods. Mobile Commerce Lab, CMU.
Map data ©2012 Google, Sanborn - Terms of Use
Report a map error



Livehoods: New York
livehoods.org/maps/nyc#
News
CMU
Wombat
Development
Cities
Financial
Furniture
Design
Other Bookmarks
Livehoods
Understanding cities through social media
Contact us
About the maps
New York
Livehood #77
Character
Related
Stats
Aggregate check-in statistics by day, hour, and type of place reveal usage patterns of the Livehood.
Daily Pulse
Mo Tu We Th Fr Sa Su
Hourly Pulse
6 9 12 15 18 21 0 3
Composition
Arts & Entertainment (10.1%)
Food (21.4%)
Nightlife (6.2%)
Home & Office (18.6%)
Travel (1.5%)
Education (1.3%)
Shops (37.5%)
Parks (3.4%)
Share us with friends?
Tweet 226
Like 727
Uniqlo
livehoods.org/maps/nyc#stats


Livehoods: New York
livehoods.org/maps/nyc#
News
CMU
Wombat
Development
Cities
Financial
Furniture
Design
Other Bookmarks
Livehoods Understanding cities through social media
Contact us
About the maps
New York
Livehood #77
Character
Related
Stats
Related Livehoods are visited by similar groups of people to the one you've selected.
Top five related Livehoods
1 Livehood # 15
2 Livehood # 31
3 Livehood # 30
4 Livehood # 60
5 Livehood # 65
Show the top 5 most related Livehoods
Show all Livehoods
Share us with friends?
Tweet 226
Like 727
Uniqlo
©2012 Livehoods. Mobile Commerce Lab, CMU.
Map data ©2012 Google, Sanborn - Terms of Use Report a map error

Evaluation

Pittsburgh Livehoods

How do we evaluate this?

- Livehoods are different from neighborhoods, *but how*? In our evaluation, we want to **characterize *what* Livehoods are.**
  - Do residents **derive social meaning** from the Livehoods mapping?
  - Can Livehoods help **elucidate the various forces** that shape and define the city?
- Quantitative (algorithmic) evaluation methods fall far short of capturing such concepts.

Evaluation

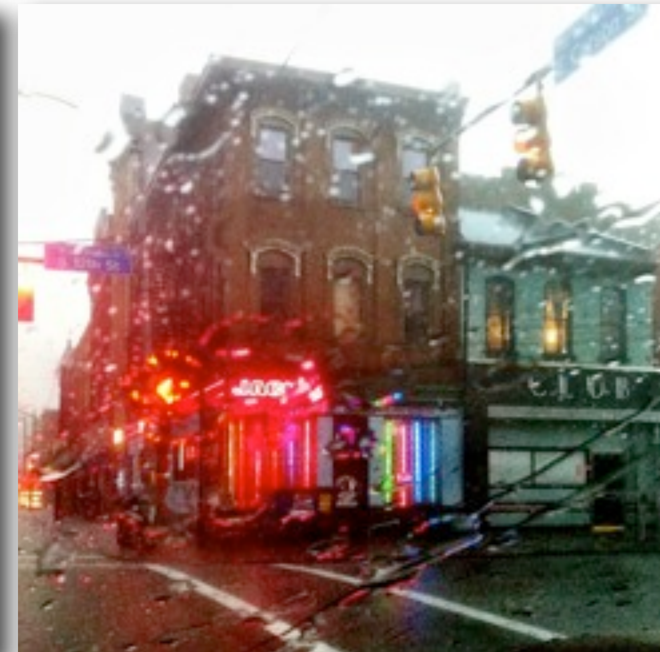
- To see how well our algorithm performed, we **interviewed 27 residents of Pittsburgh**
- Residents recruited through a social media campaign, with various neighborhood groups and entities as seeds
- Semi-structured Interview protocol **explored the relationship among Livehoods**, **municipal borders**, and the participants own **perceptions of the city**.
- Participants must have lived in their neighborhood for at least 1 year

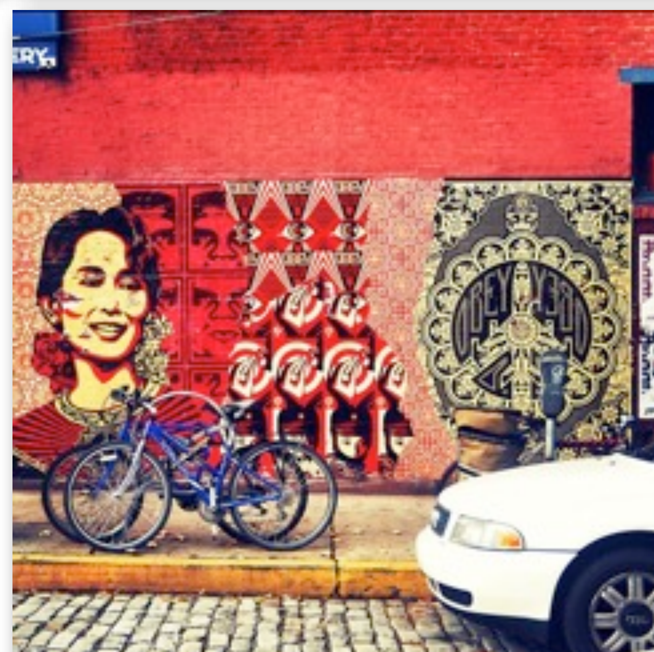
Interview Protocol

- Each interview lasted approximately one hour
- Began with a discussion of their backgrounds in relation their neighborhood.
- Asked them to draw the boundaries of their neighborhood over it (their cognitive map).
- Is there a “shift in feel” of the neighborhood?
- Municipal borders changing?
- Show Livehoods clusters (ask for feedback)
- Show related Livehoods (ask for feedback)

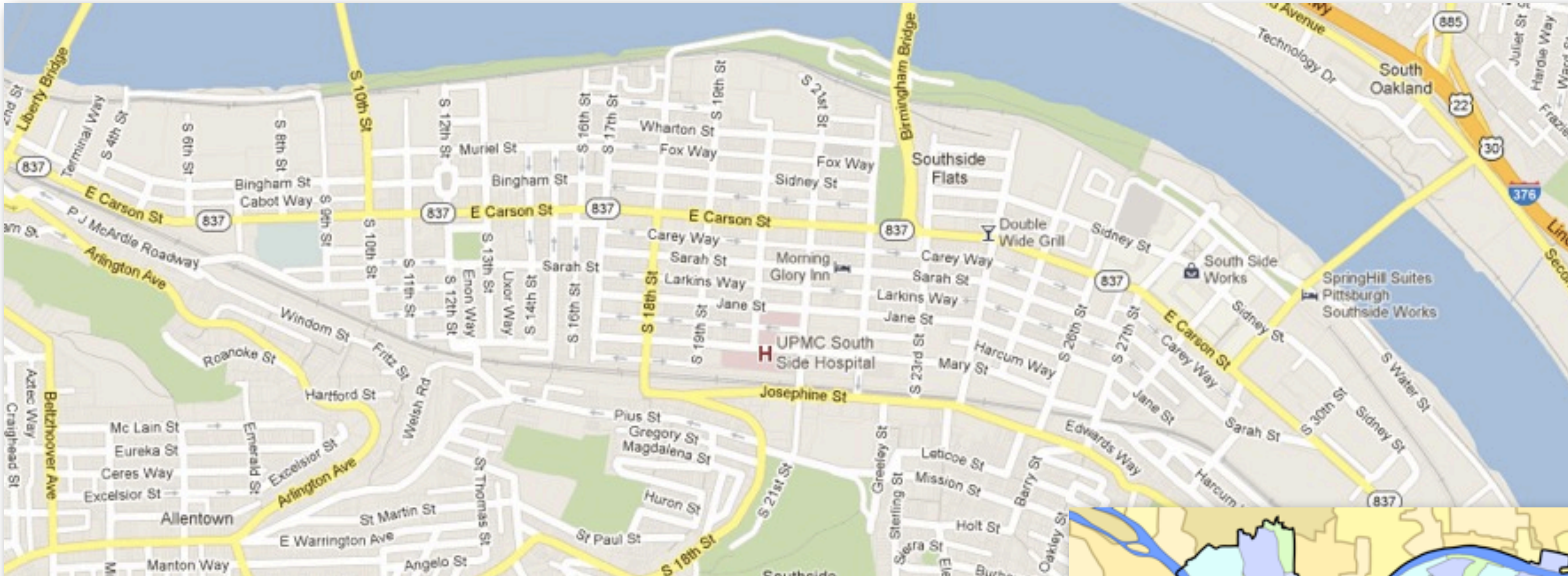
Interview Results

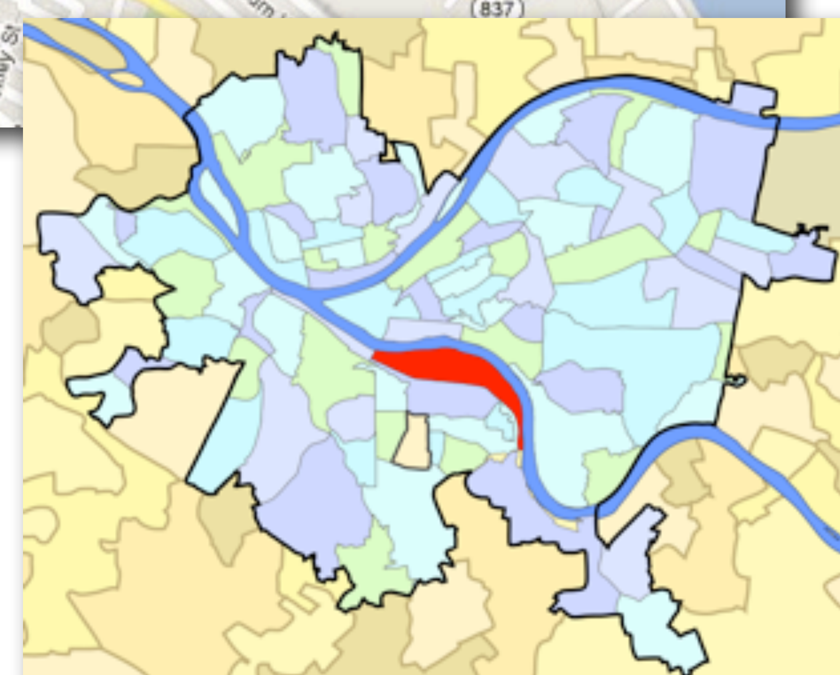
South Side Pittsburgh



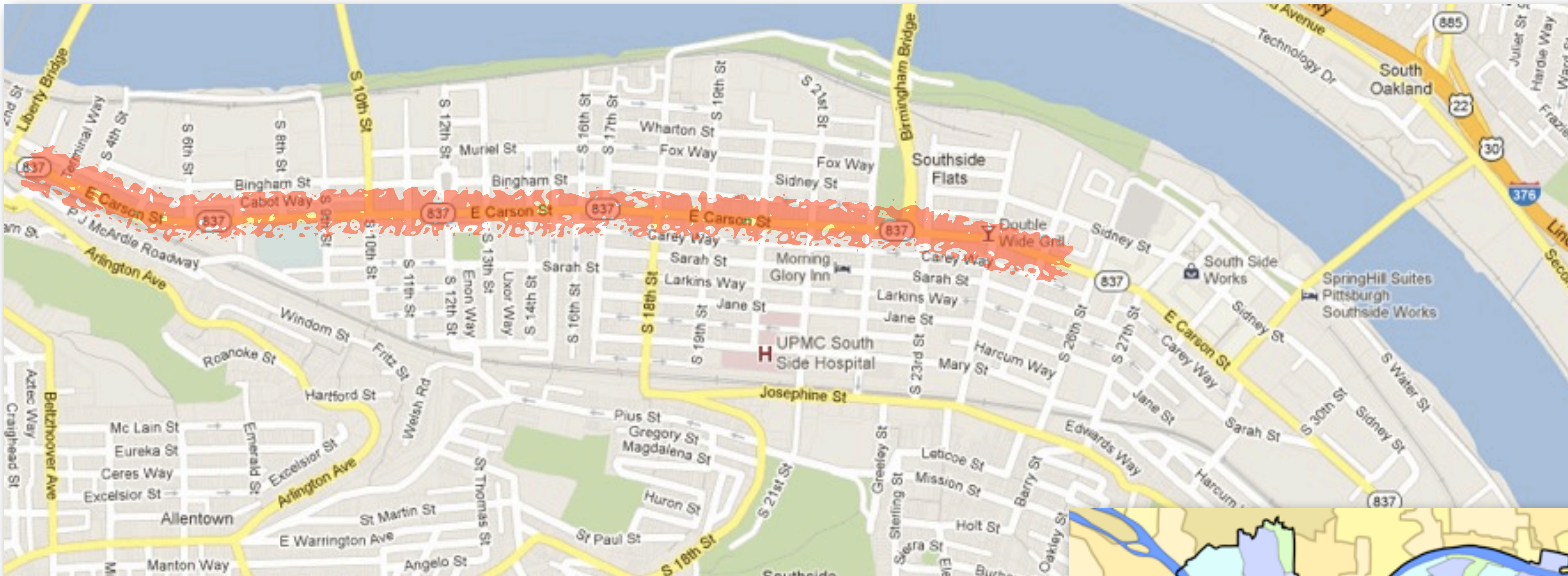


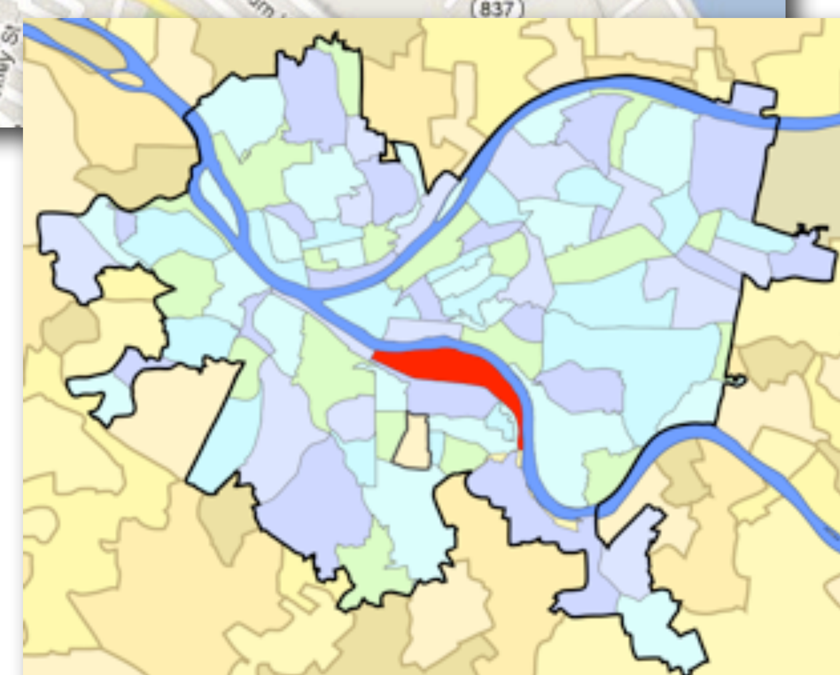
South Side Pittsburgh





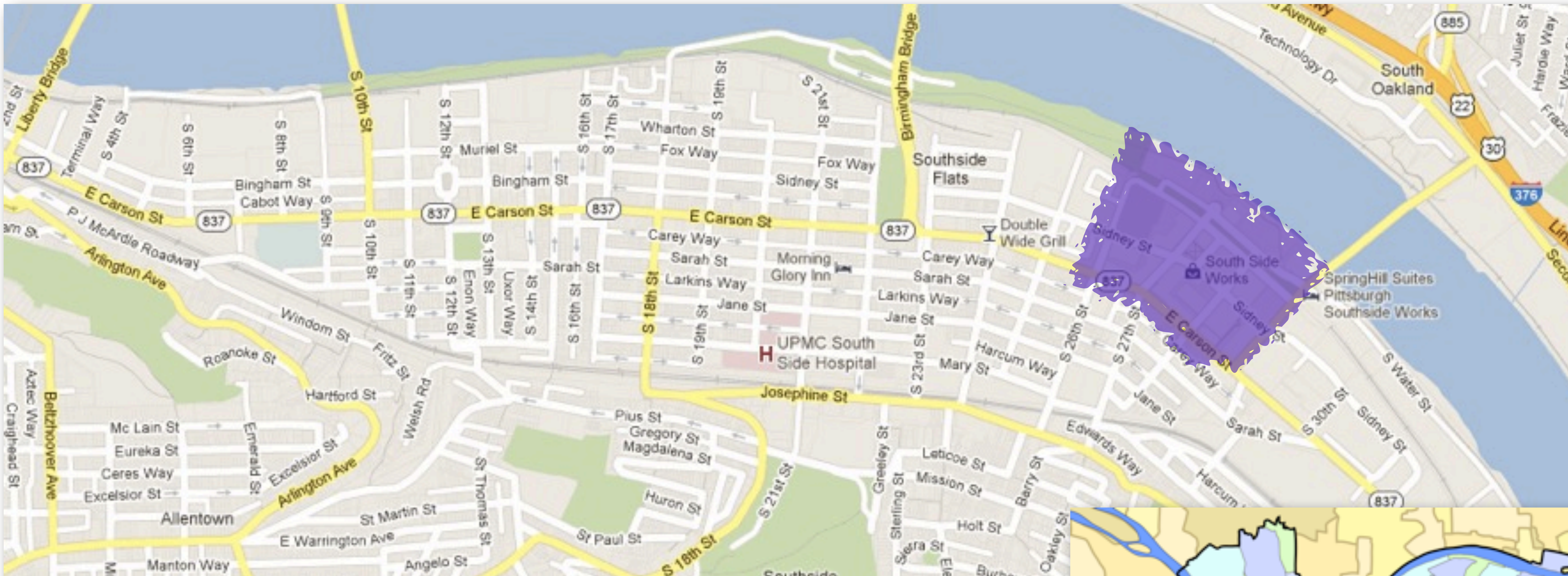
South Side Pittsburgh

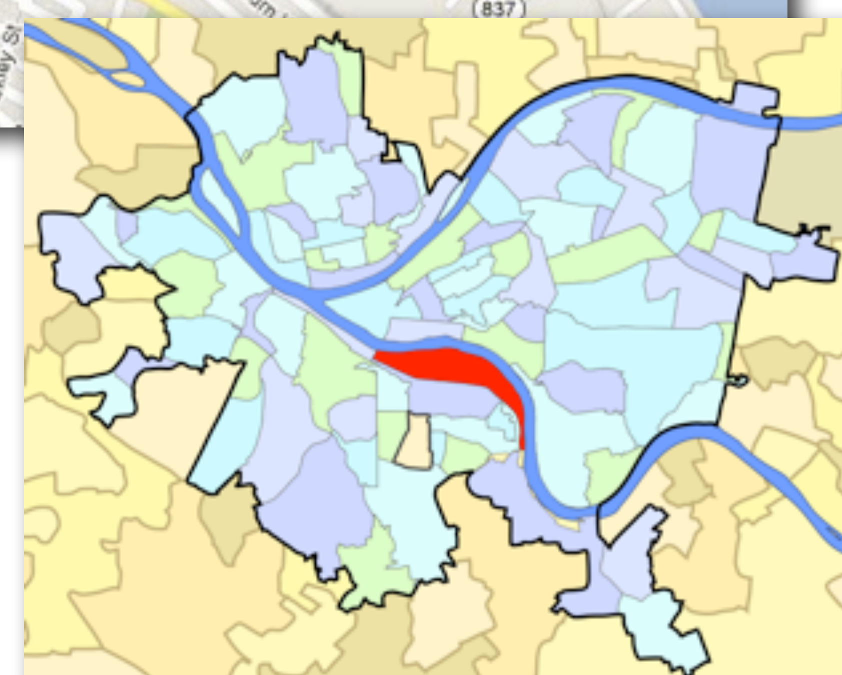




Carson Street runs along the length of South Side, and is densely packed with **bars, restaurants, tattoo parlors,** and **clothing** and **furniture shops.** It is the most popular destination for nightlife.

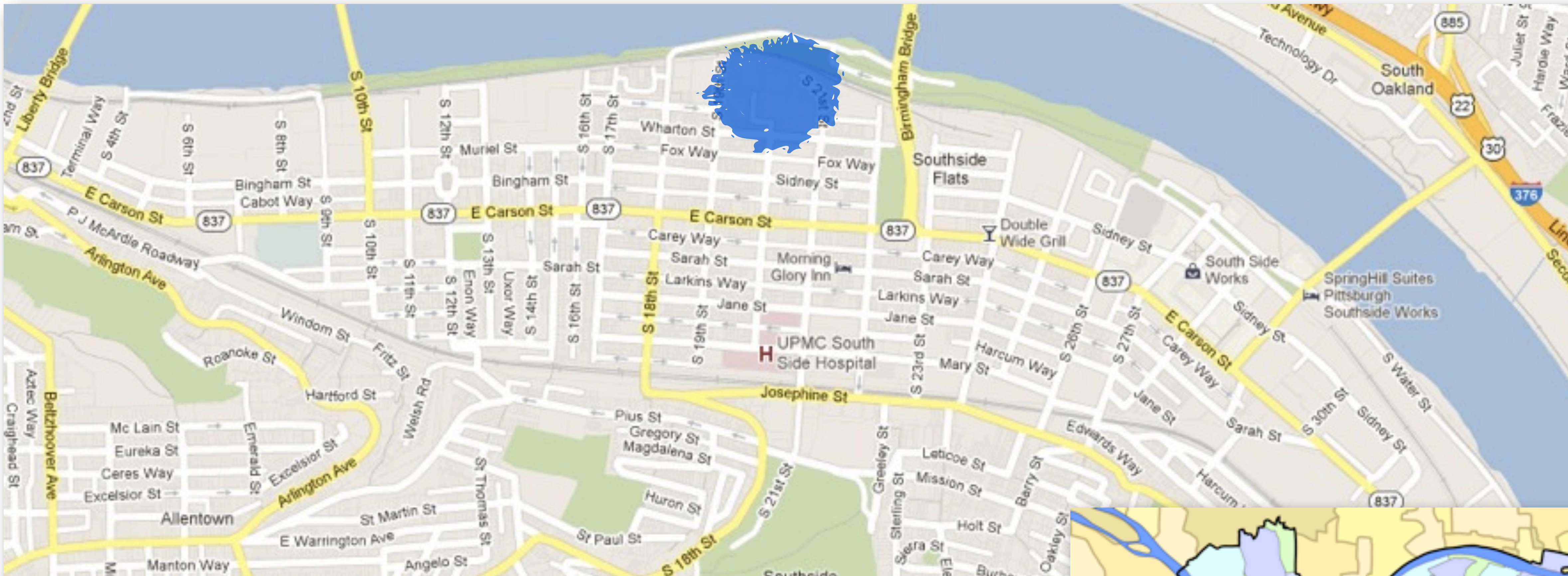
South Side Pittsburgh

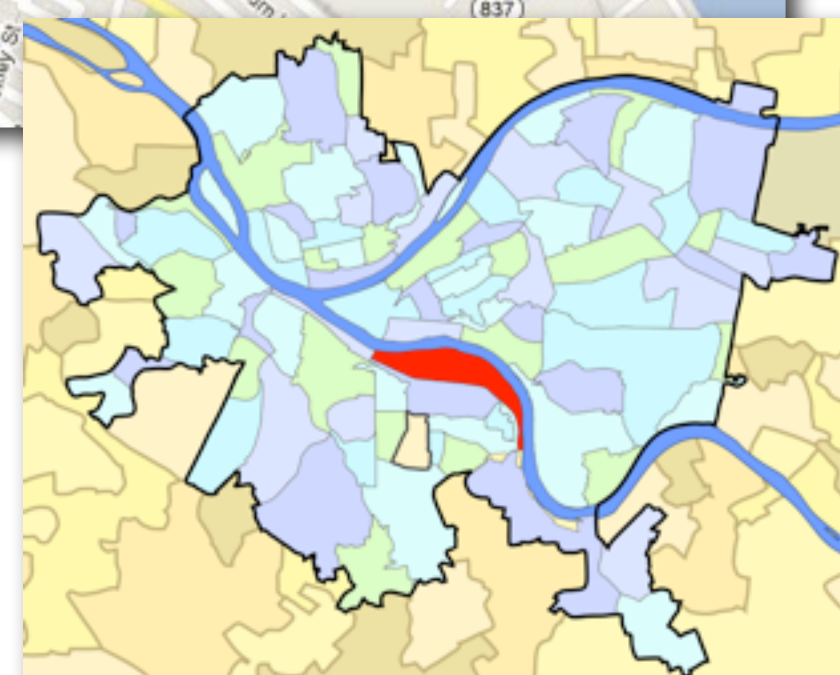




South Side Works is a recently built, mixed-use *outdoor shopping mall*, containing **nationally branded apparel** stores and **restaurants**, **upscale condominiums**, and **corporate offices**.

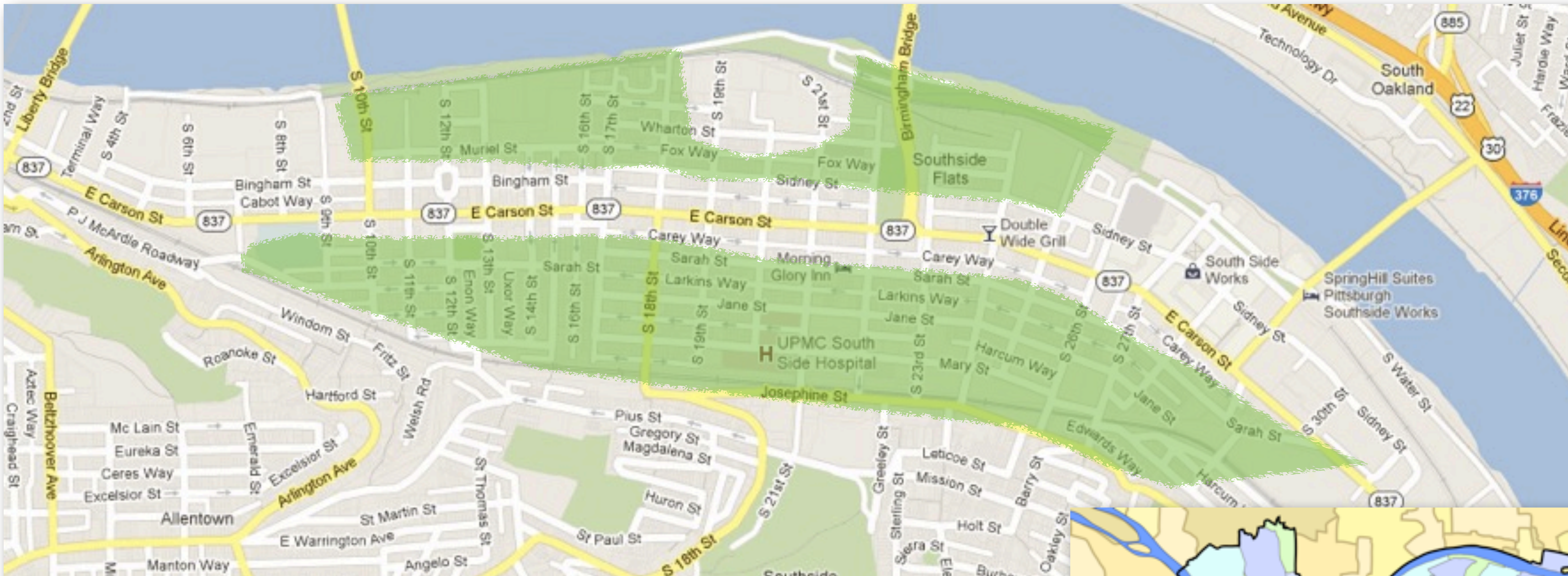
South Side Pittsburgh

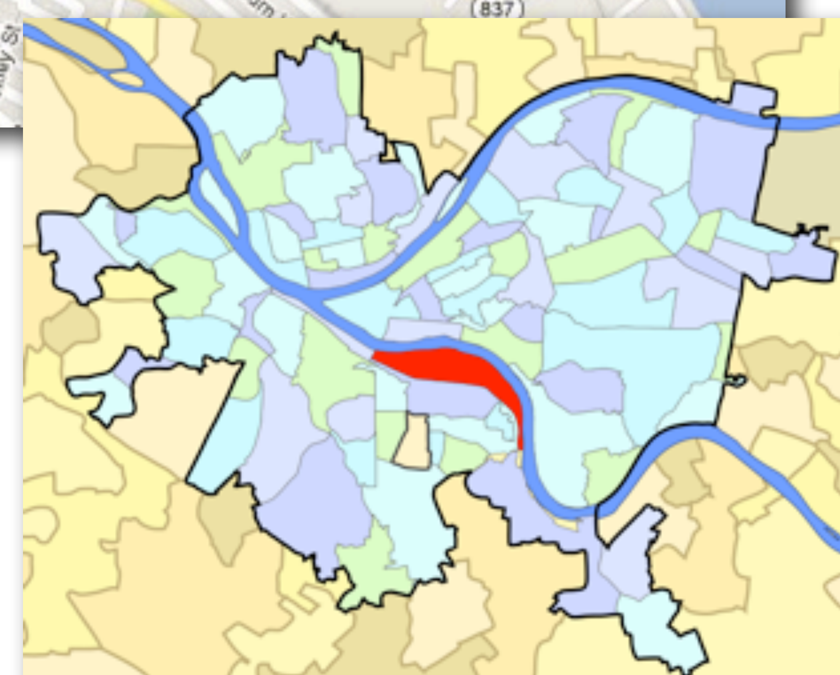




There is an small, somewhat older strip-mall that contains the only **super market** (grocery) in South Side. It also has a **liquor store**, an **auto-parts** store, a **furniture rental** store and other small chain stores.

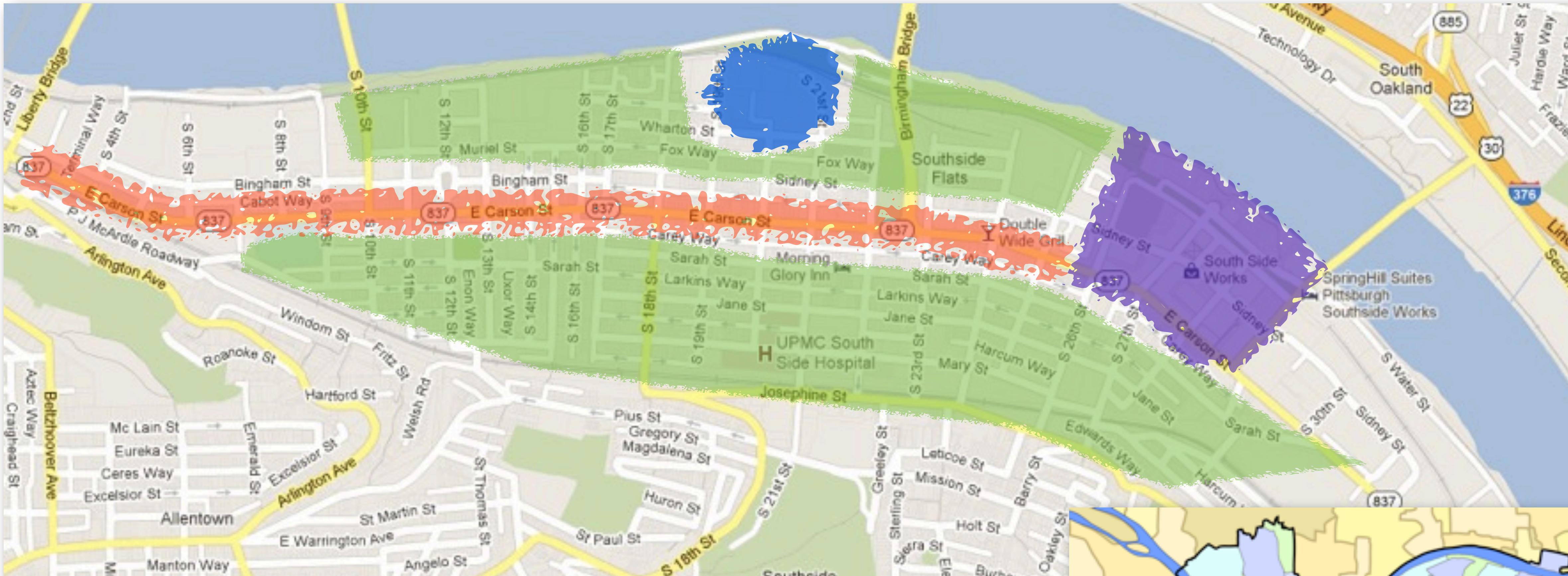
South Side Pittsburgh

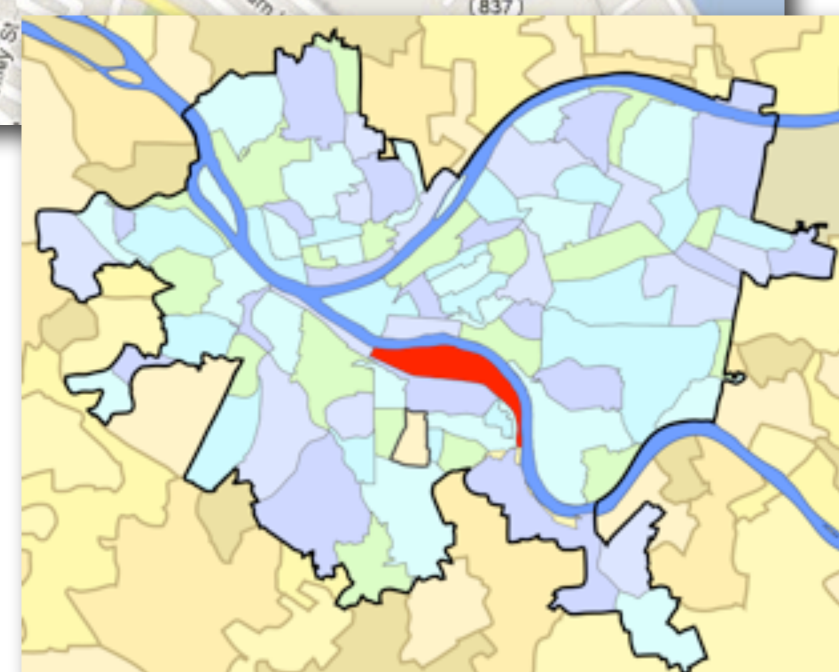




The rest of South Side is predominantly residential, consisting of mostly smaller row houses.

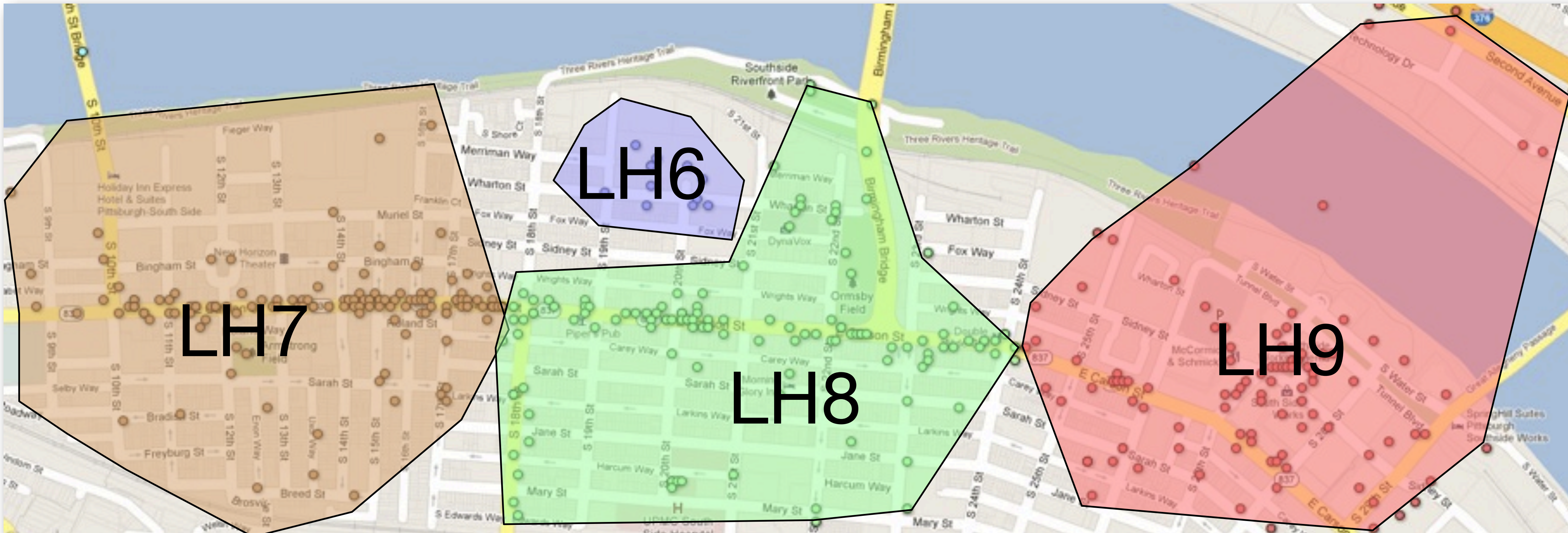
South Side Pittsburgh





My Cognitive Map of South Side

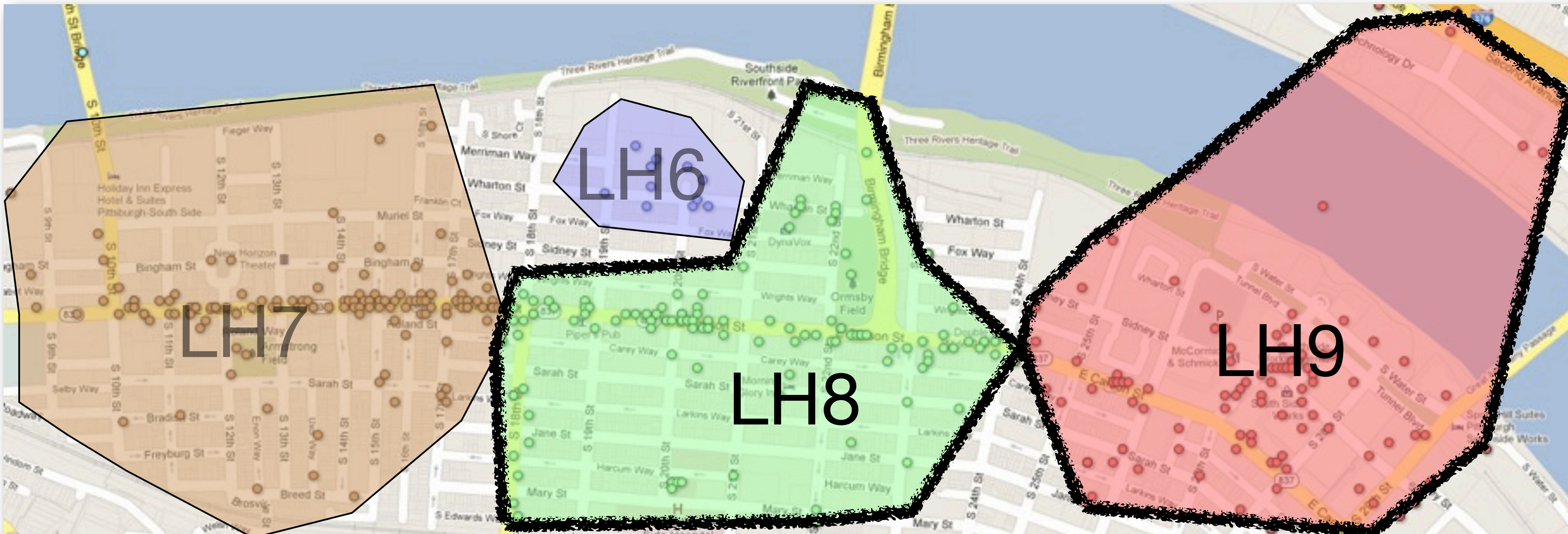
South Side Pittsburgh



The Livehoods of South Side

I'll show the evidence in support of the Livehoods clusters in South Side, and will describe the forces that shape the city that the Livehoods highlight.

South Side Pittsburgh

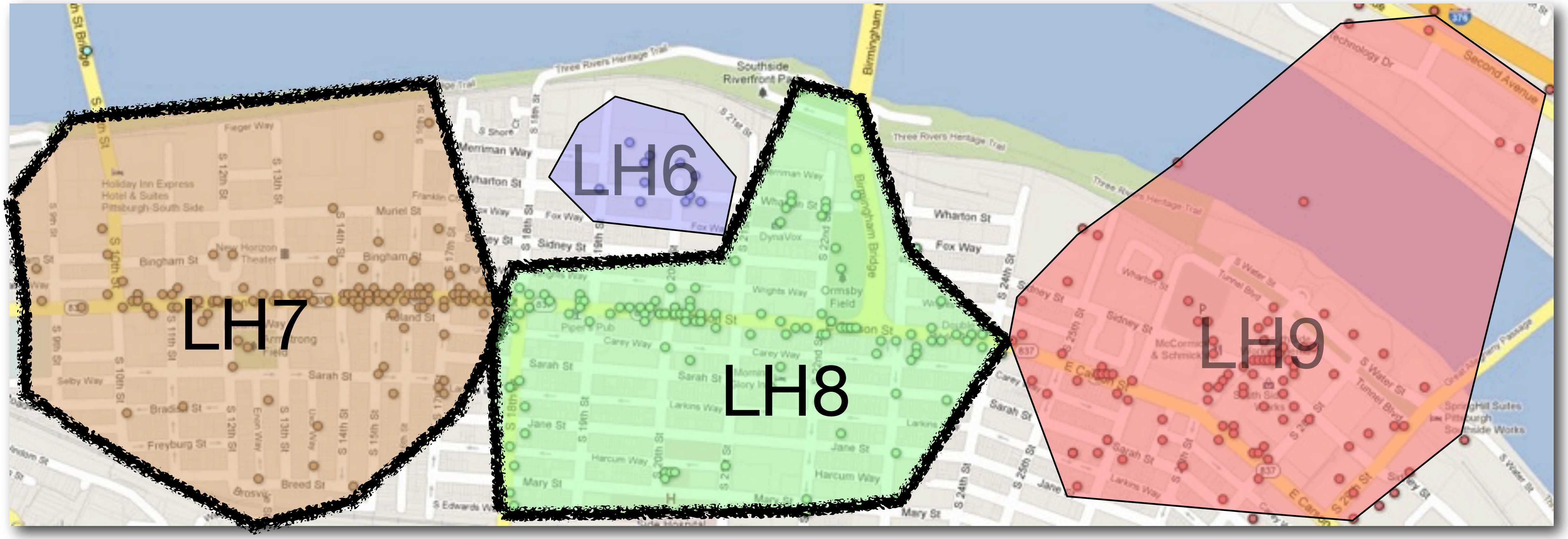


LH8 vs LH9

Demographic Differences

“Ha! Yes! See, here is my division! Yay! Thank you algorithm! ... I definitely feel where the South Side Works, and all of that is, is a very different feel.”

South Side Pittsburgh

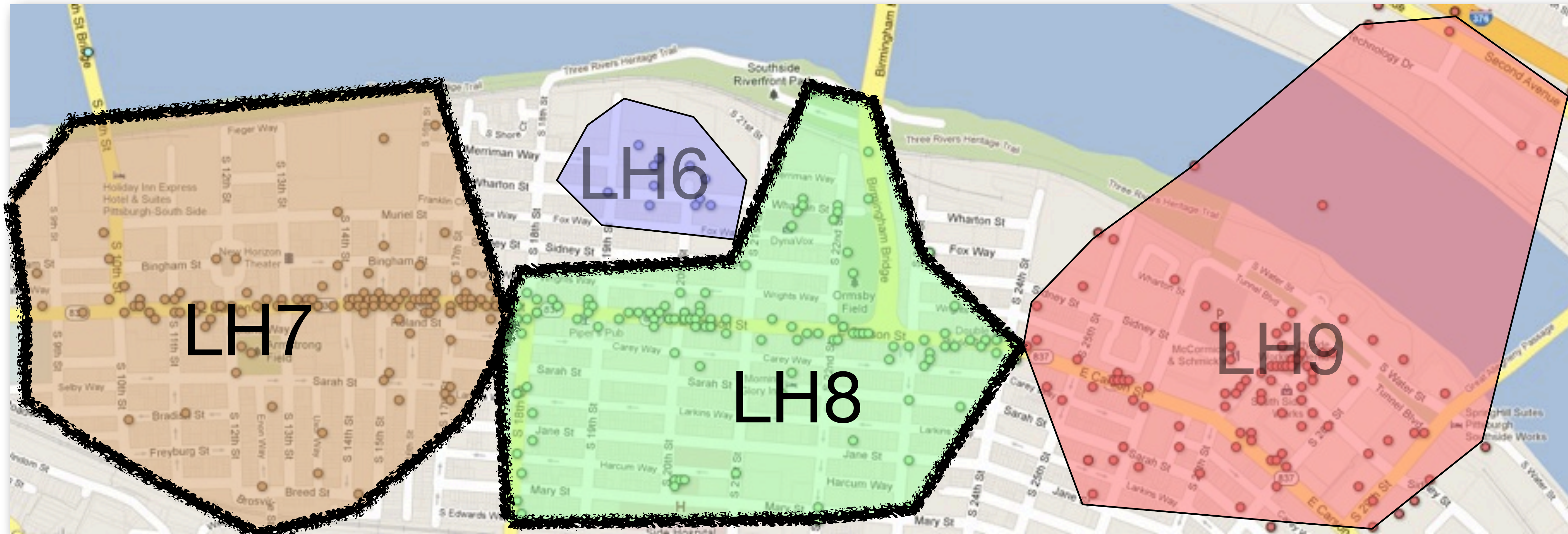


LH7 vs LH8

Architecture & Urban Design

“from an urban standpoint it is a lot tighter on the western part once you get west of 17th or 18th [LH7].”

South Side Pittsburgh

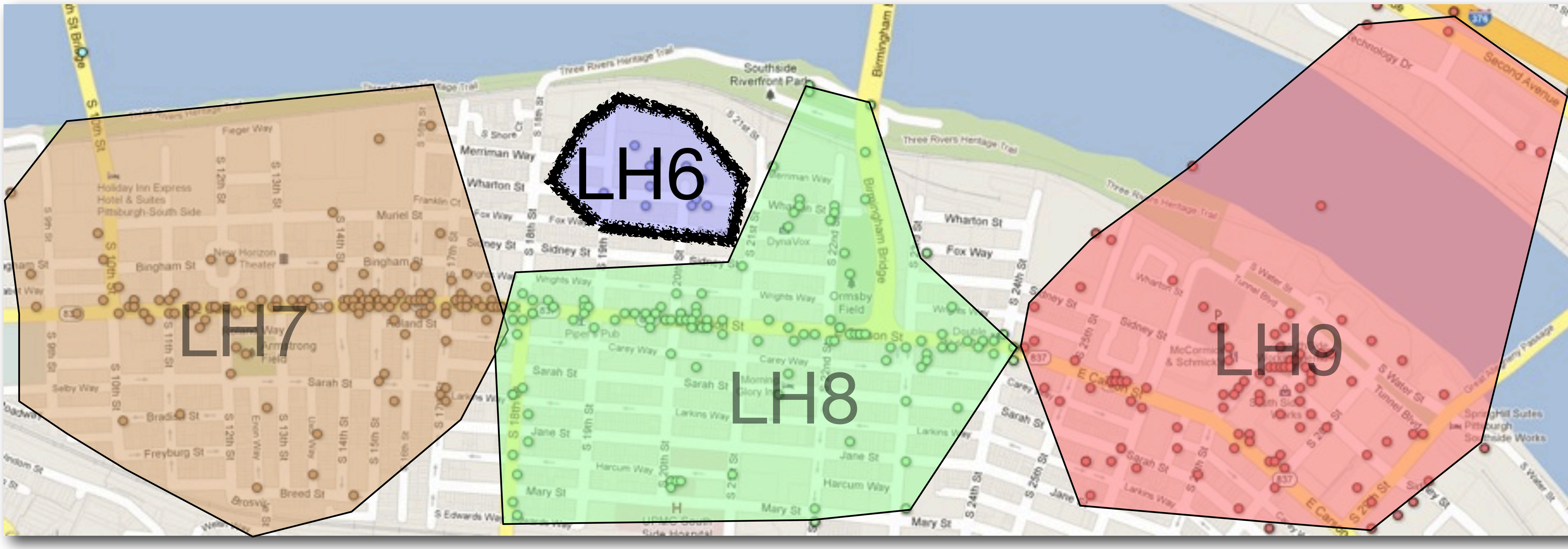


LH7 vs LH8

Safety

“Whenever I was living down on 15th Street [LH7] I had to worry about drunk people following me home, but on 23rd [LH8] I need to worry about people trying to mug you... so it’s different. It’s not something I had anticipated, but there is a distinct difference between the two areas of the South Side.”

South Side Pittsburgh



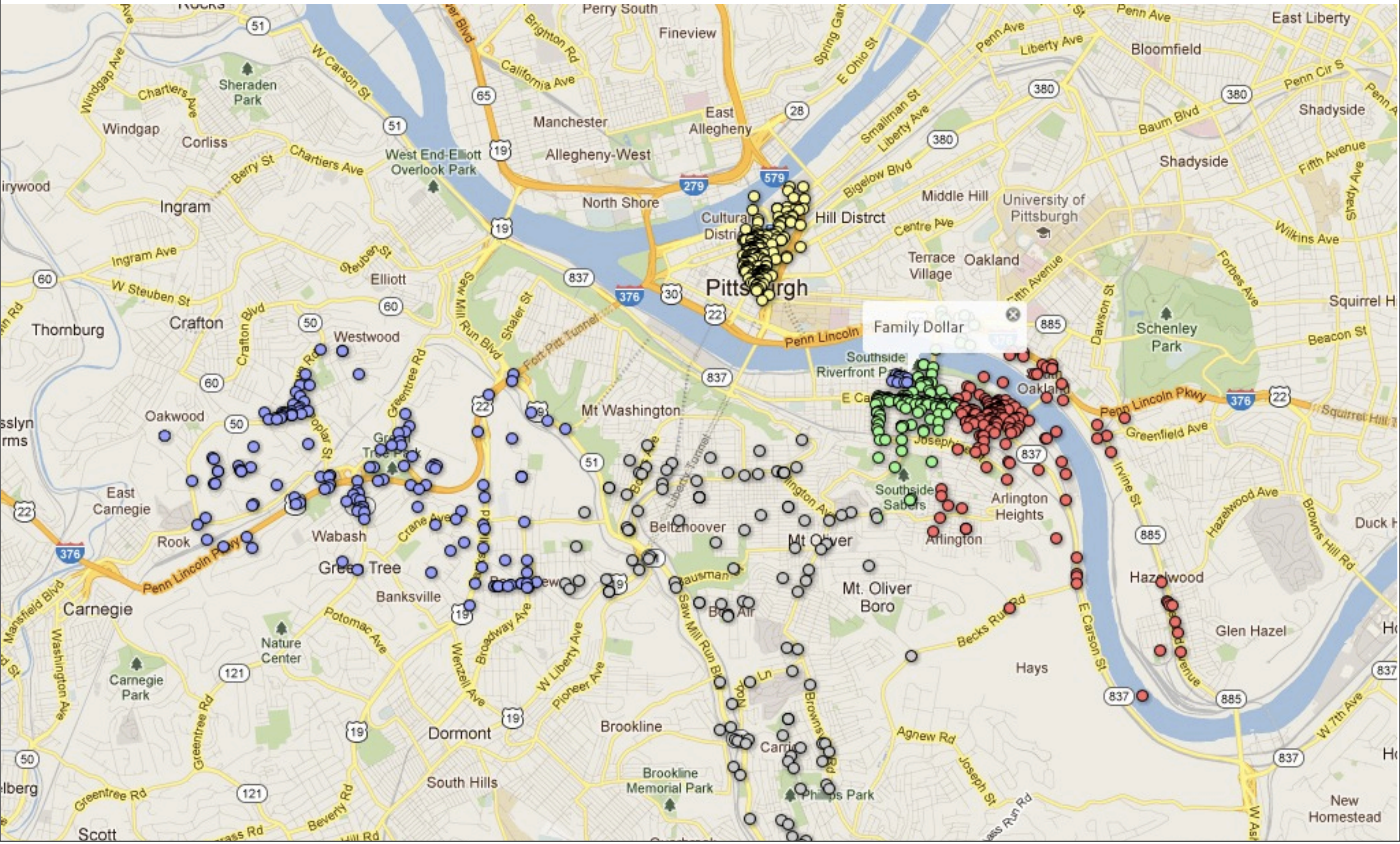
LH6

Demographic Differences

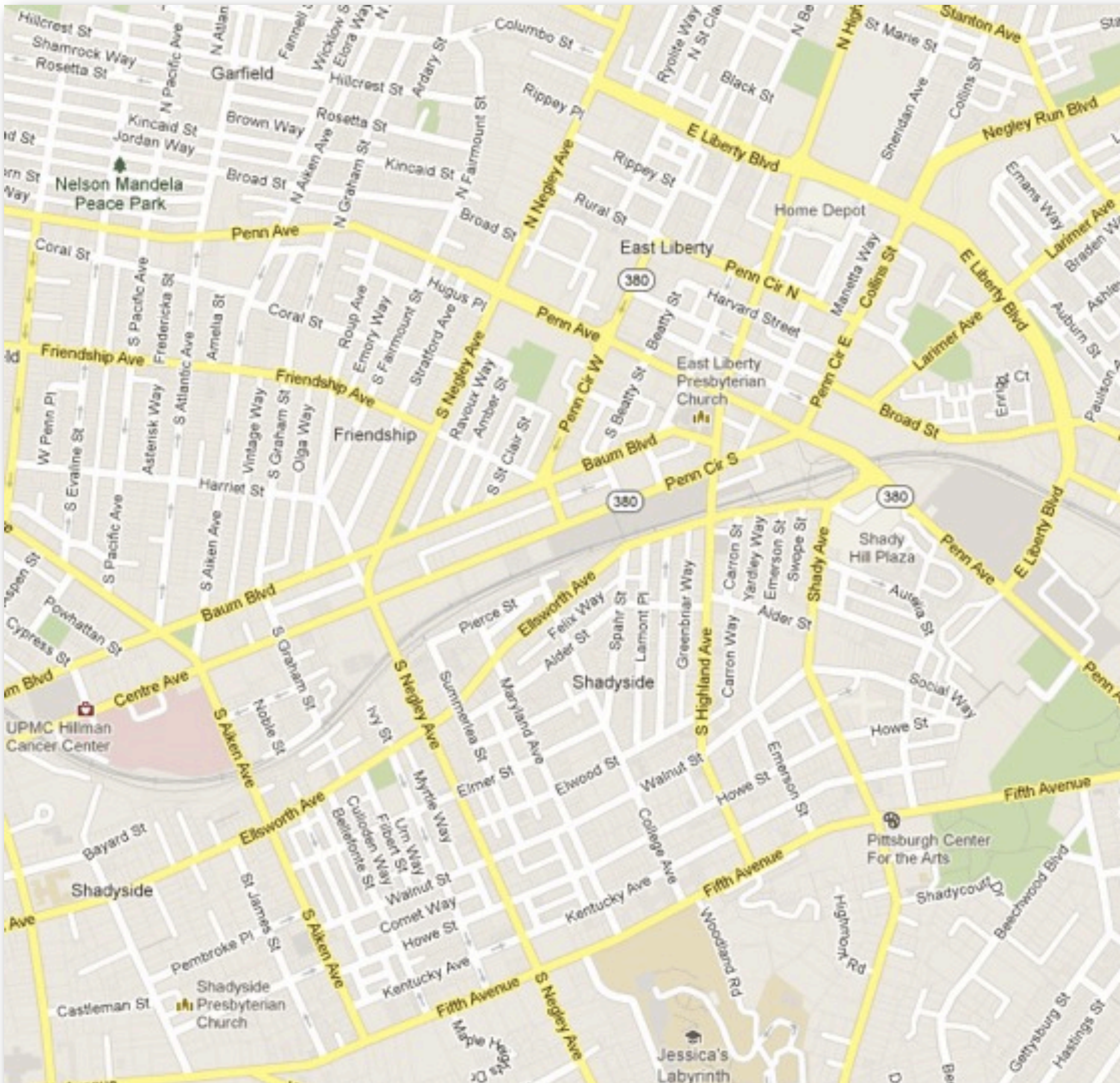
"There is this interesting mix of people there I don't see walking around the neighborhood. I think they are coming to the Giant Eagle from lower income neighborhoods...I always assumed they came from up the hill."

South Side Pittsburgh

"I always assumed they came from up the hill."



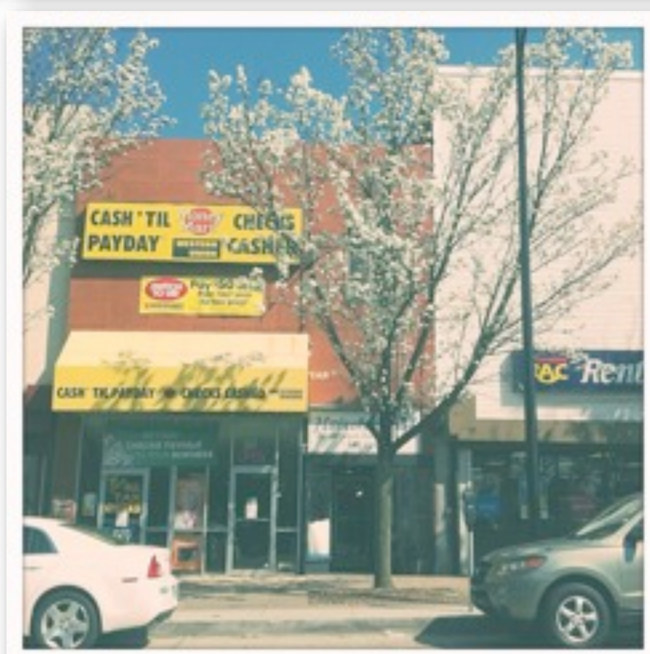
Shadyside and East Liberty



A Teaser...

Shadyside

East Liberty



The Train Tracks

The Whole Foods

The Pedestrian Bridge

Conclusions

- Throughout our interviews we found very strong evidence in support of the clustering
- Interviews showed that residence found strong social meaning behind the Livehood clusters.
- We also found that Livehoods can help shed light on the various forces that shape people's behavior in the city, the city including demographics, economic factors, cultural perceptions and architecture.

Limitations

- Most Livehoods had real social meaning to participants, but no algorithm is perfect. There are certainly Livehoods that don't make sense.
- There are obvious biases to using foursquare data. However this a limitation to the data, not our methodology.
- Some populations are left out (the digital divide)
- We don't want to overemphasize sharp divisions between Livehoods. In reality neighborhoods blend into one another.
- This is not comparative work. We're not making the claim that ours model is the best model for capturing the areas of a city, only that its a good model.

Future Work

- Comparative Work
  - Our model works well, but it's just one way to accomplish the end goal.
  - It would be fascinating to compare how different model variations segment the city differently.

Thanks!

Please explore our maps at livehoods.org
You can also find us on on Facebook and Twitter **@livehoods**.

Justin Cranshaw
jcransh@cs.cmu.edu

Raz Schwartz
razs@andrew.cmu.edu

Jason I. Hong
jasonh@cs.cmu.edu

Norman Sadeh
sadeh@cs.cmu.edu

School of Computer Science
Carnegie Mellon University

@livehoods | livehoods.org

http://distilleryimage7.s3.amazonaws.com/88a6ca749b8b11e180c9123138016265_7.jpg

http://distilleryimage7.s3.amazonaws.com/0a7e8f2c6ad711e180d51231380fcd7e_7.jpg

http://distilleryimage0.s3.amazonaws.com/b57fcd849d6b11e1b10e123138105d6b_7.jpg

http://distilleryimage9.s3.amazonaws.com/94a7a046453711e19e4a12313813ffc0_7.jpg

http://distilleryimage3.s3.amazonaws.com/1733a942876411e1be6a12313820455d_7.jpg

http://distilleryimage1.s3.amazonaws.com/22c1c74e6af911e180d51231380fcd7e_7.jpg

http://distilleryimage2.s3.amazonaws.com/70cb3dca9ebc11e18cf91231380fd29b_7.jpg

http://instagr.am/p/LHByf3CtNl/

http://distilleryimage9.s3.amazonaws.com/d7b9e2249bca11e1a92a1231381b6f02_7.jpg

http://distilleryimage6.s3.amazonaws.com/4de101bc414411e1abb01231381b65e3_7.jpg

http://distilleryimage7.s3.amazonaws.com/93bcdec8a04311e18bb812313804a181_7.jpg

http://distilleryimage4.instagram.com/3d3ef156a9eb11e181bd12313817987b_7.jpg

http://distilleryimage0.s3.amazonaws.com/db4691625bdd11e19896123138142014_7.jpg

http://distilleryimage0.s3.amazonaws.com/65ba3af242d811e19896123138142014_7.jpg

http://instagr.am/p/K_ZE9roeq6/

http://www.flickr.com/photos/award33/4730513546

http://distilleryimage8.s3.amazonaws.com/9cc5f582919211e1be6a12313820455d_7.jpg

http://distillery.s3.amazonaws.com/media/2011/10/07/1d2d0bc47db44acf9cf15017103af06a_7.jpg

http://distilleryimage5.s3.amazonaws.com/fe99d252a69f11e1abd612313810100a_7.jpg

http://instagr.am/p/LN_tt3MzHl/

http://distilleryimage8.s3.amazonaws.com/88d3de6639a111e1a87612313804ec91_7.jpg

http://distilleryimage2.s3.amazonaws.com/d57f456c534011e1abb01231381b65e3_7.jpg

http://instagr.am/p/K8O4JLoY-4/

http://distilleryimage8.s3.amazonaws.com/dfb52f1268b011e19e4a12313813ffc0_7.jpg

http://distilleryimage1.s3.amazonaws.com/fe0235ce7a0c11e181bd12313817987b_7.jpg

http://distilleryimage0.s3.amazonaws.com/2dc3f470923911e181bd12313817987b_7.jpg

http://distilleryimage0.s3.amazonaws.com/0c6a6cac9af711e19dc71231380fe523_7.jpg

http://distilleryimage5.s3.amazonaws.com/fe99d252a69f11e1abd612313810100a_7.jpg

http://distilleryimage4.s3.amazonaws.com/d3db2b5ca36a11e18bb812313804a181_7.jpg

http://distilleryimage7.s3.amazonaws.com/ca6d892a13f011e180c9123138016265_7.jpg