Crowdsourcing Pedestrian and Bicyclist Activity Data

TRB UTC Spotlight Conference

December 1, 2016

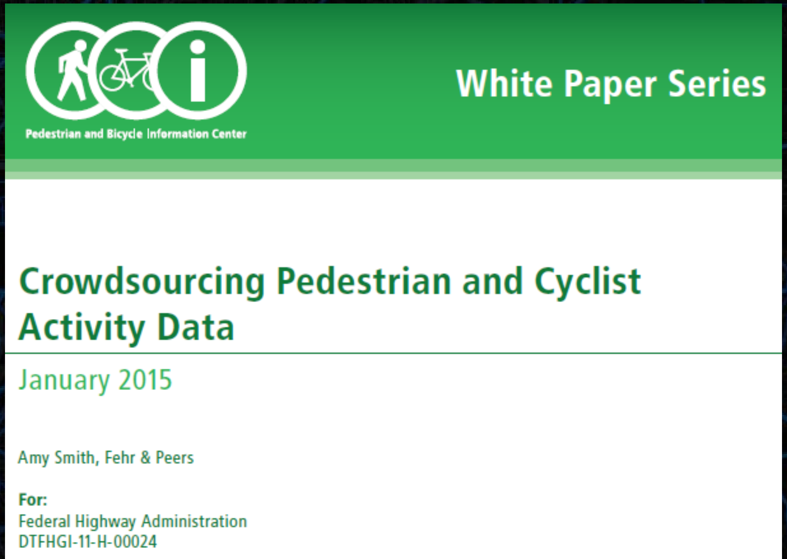
FEHR PEERS DC

Alex Rixey

@alexrixey

Overview

1 Crowdsourced Data Typology
2 Data Source Examples
3 Data Challenges
4 Applications
5 New Crowdsourced Data Opportunities
6 Better Data → Better Safety Analysis?


Pedestrian and Bicycle Information Center

White Paper Series

Crowdsourcing Pedestrian and Cyclist Activity Data

January 2015

Amy Smith, Fehr & Peers

For:
Federal Highway Administration
DTFHGI-11-H-00024

1 | Crowdsourced Data Typology

In-Situ

Thematic

Thumbtack

Spatial Inventory

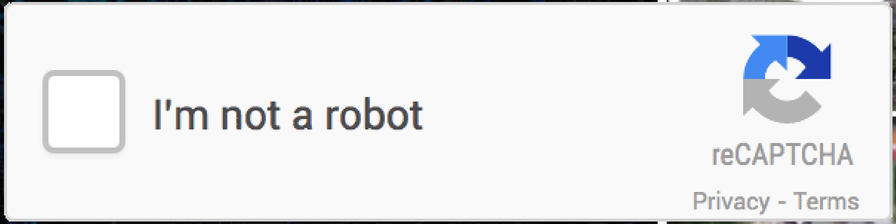
1 | Additional Crowdsourced Data Characteristics

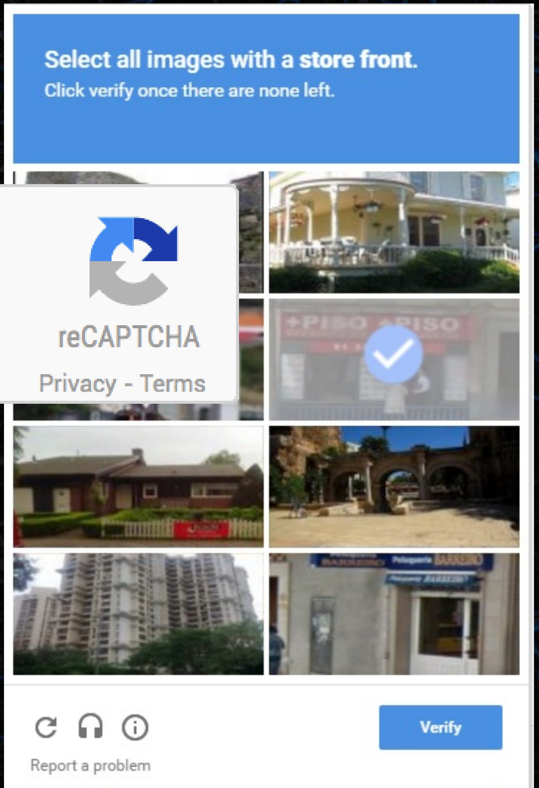
Explicit

Defined problem communicated directly to participants

Implicit

Participants may be unaware of secondary use of their data





Misra, Gooze, Watkins, Asad and Le Dantec (2014)

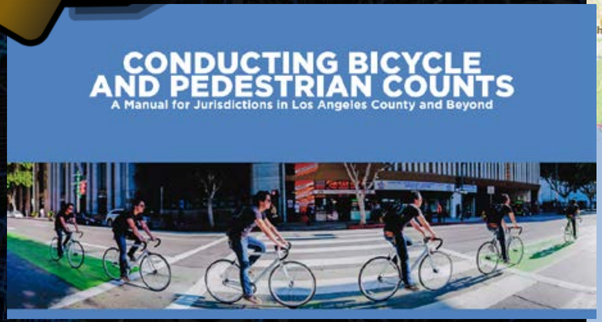
1 | Additional Crowdsourced Data Characteristics

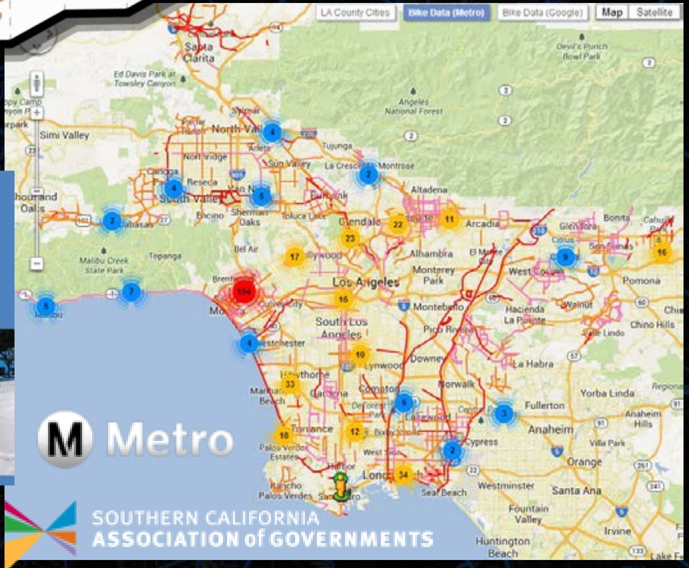
General Purpose

Does not require specialized knowledge from participants

Domain-Specific

Data collected from participants with existing expertise





Misra, Gooze, Watkins, Asad and Le Dantec (2014)

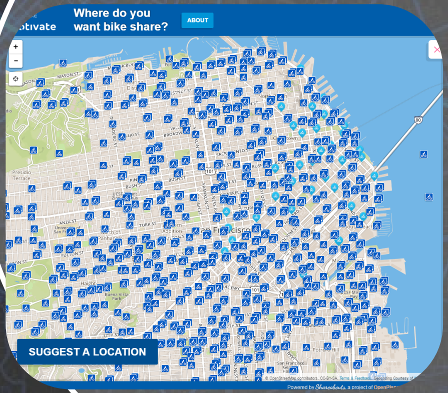
1 | Additional Crowdsourced Data Characteristics

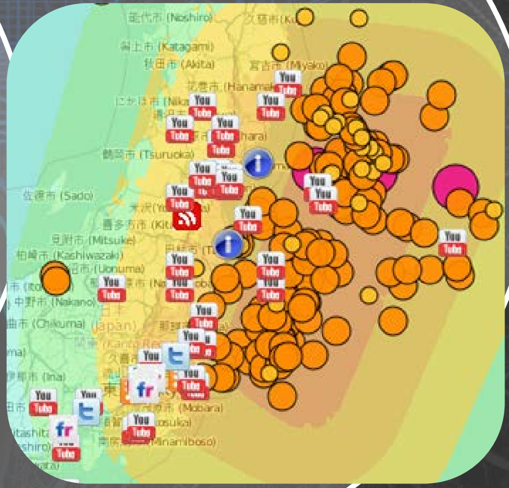
Geocentric

Audience-Centric

Event-Centric

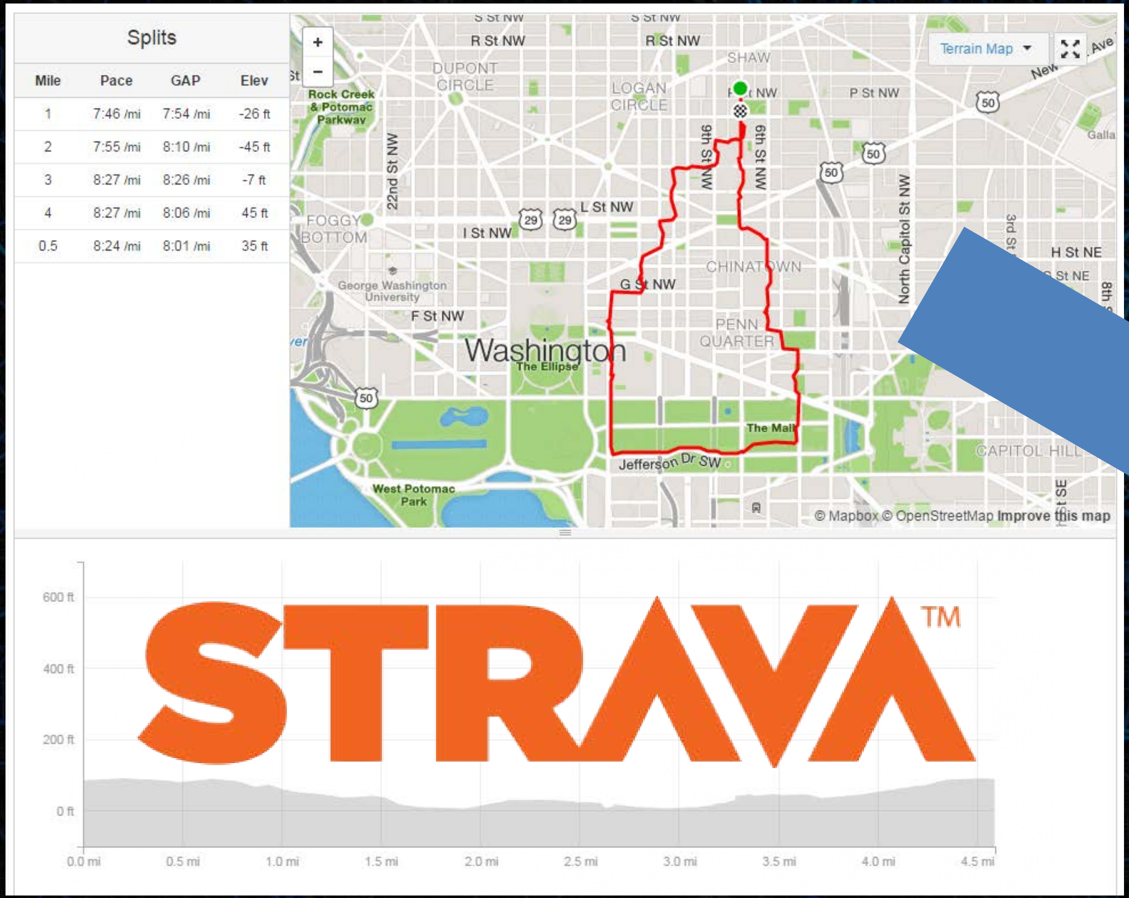
Global





Misra, Gooze, Watkins, Asad and Le Dantec (2014)

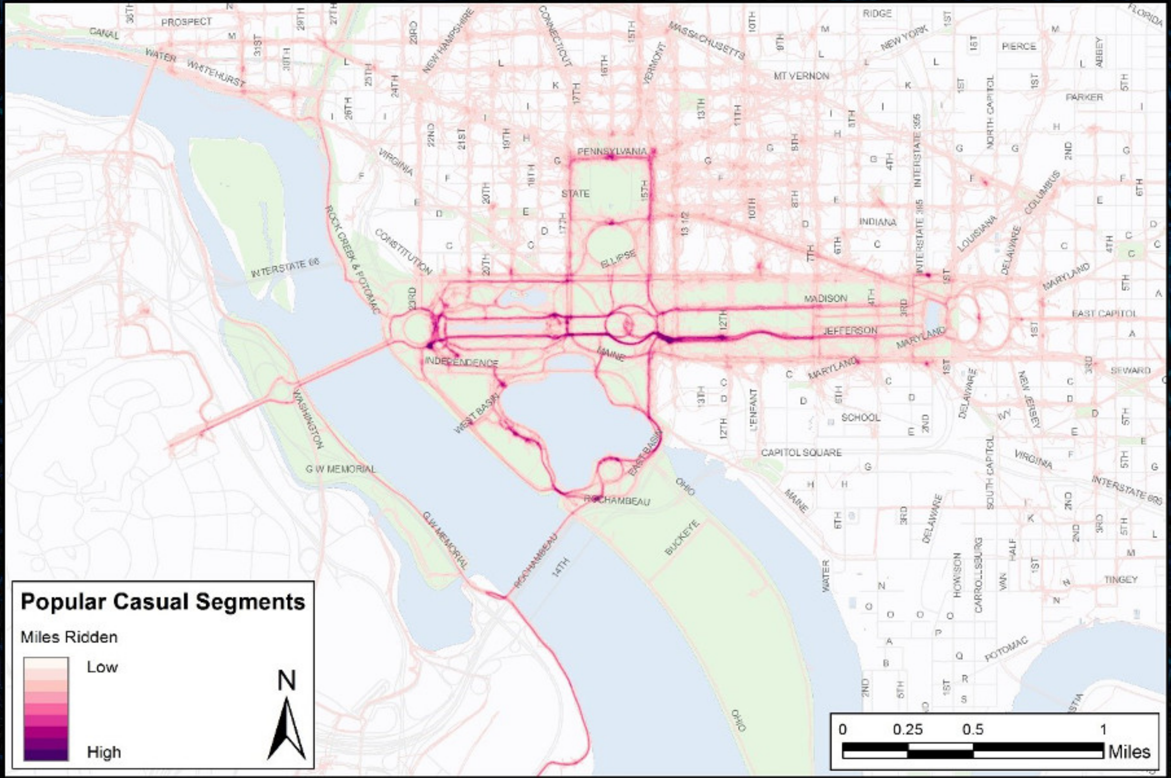
2 | In-Situ Data (Explicit)

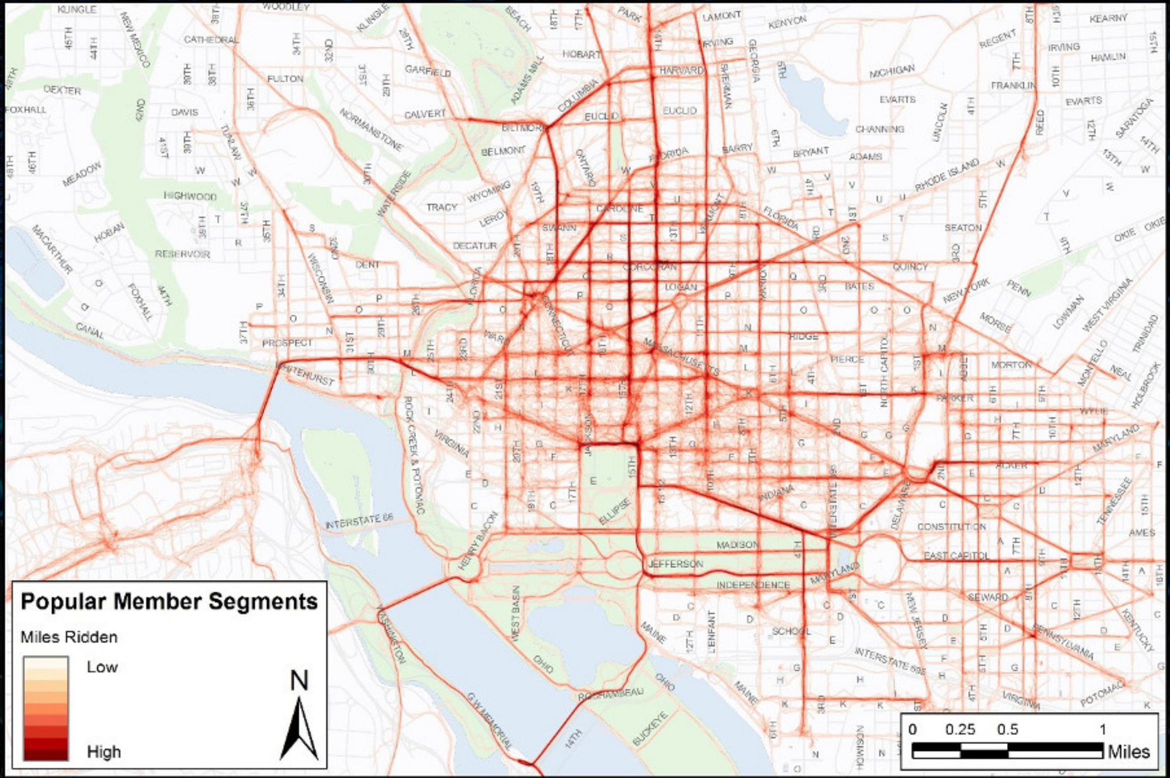


STRAVA | METRO

2 | In-Situ Data (Implicit)

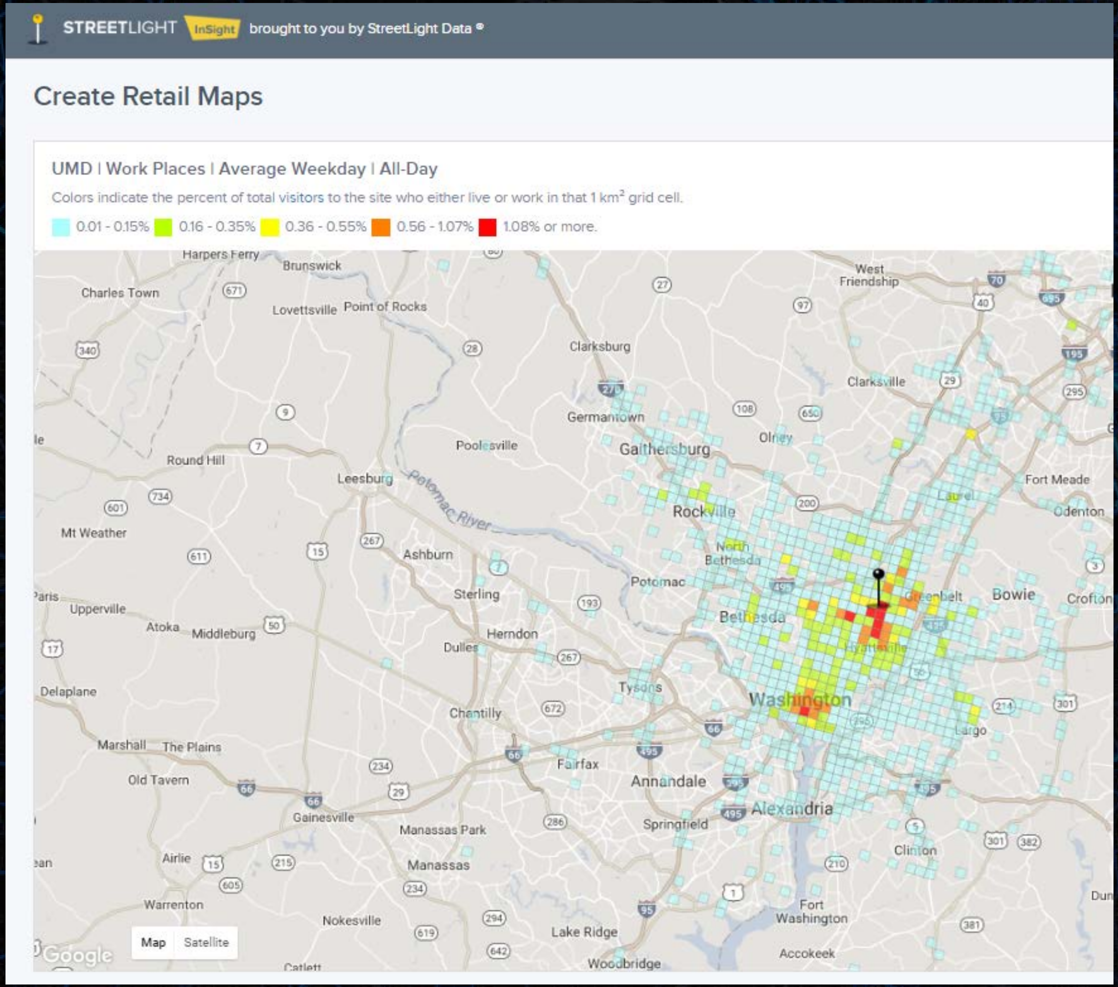
capital bikeshare™





Source: Joe Wergin

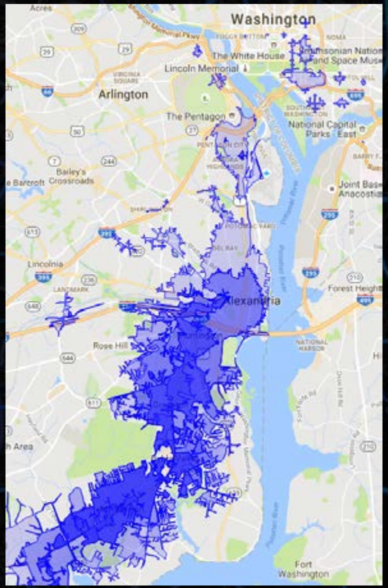
2 | In-Situ Data (Implicit)



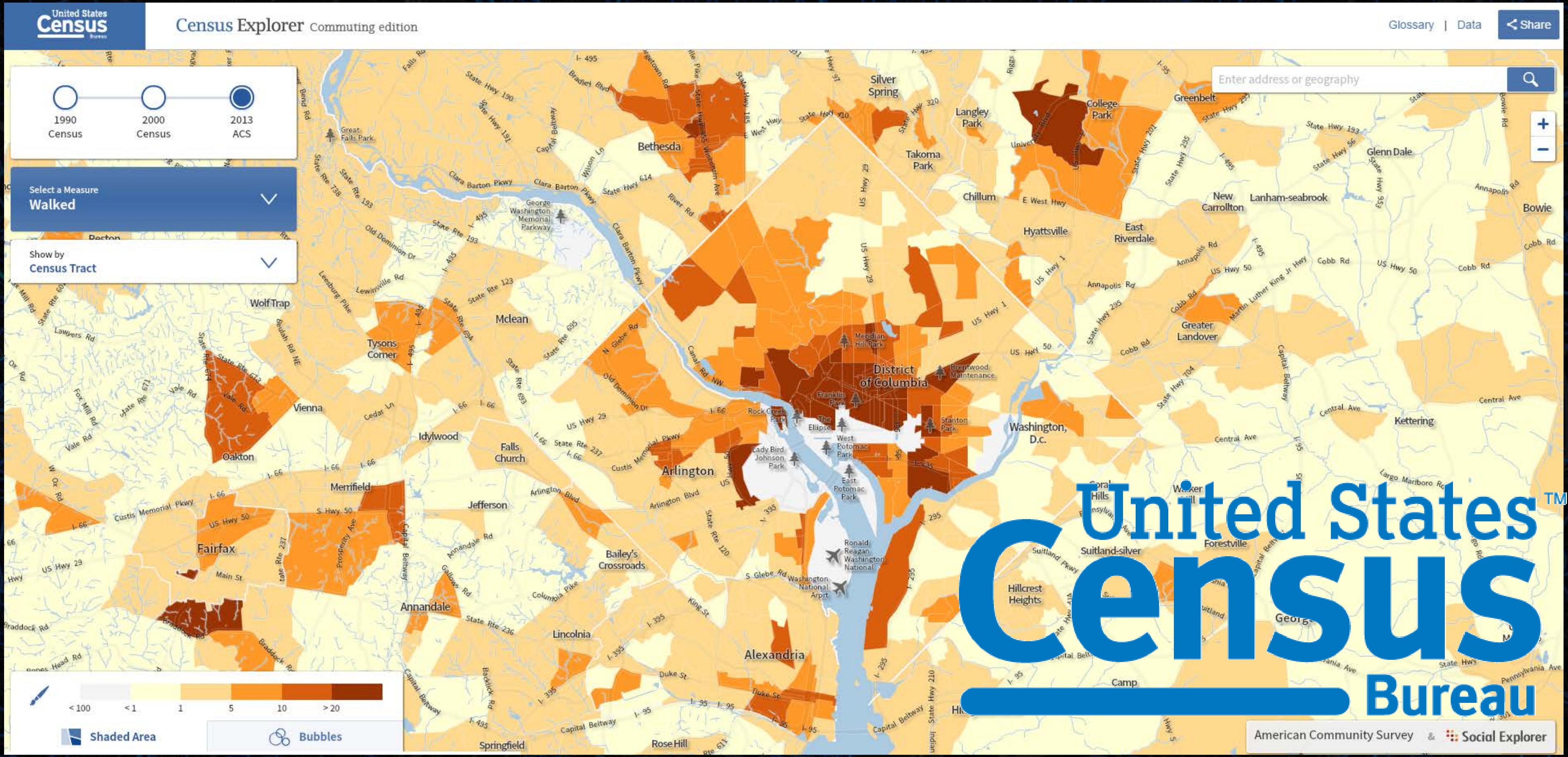
airsage
The power of where and when

INRIX

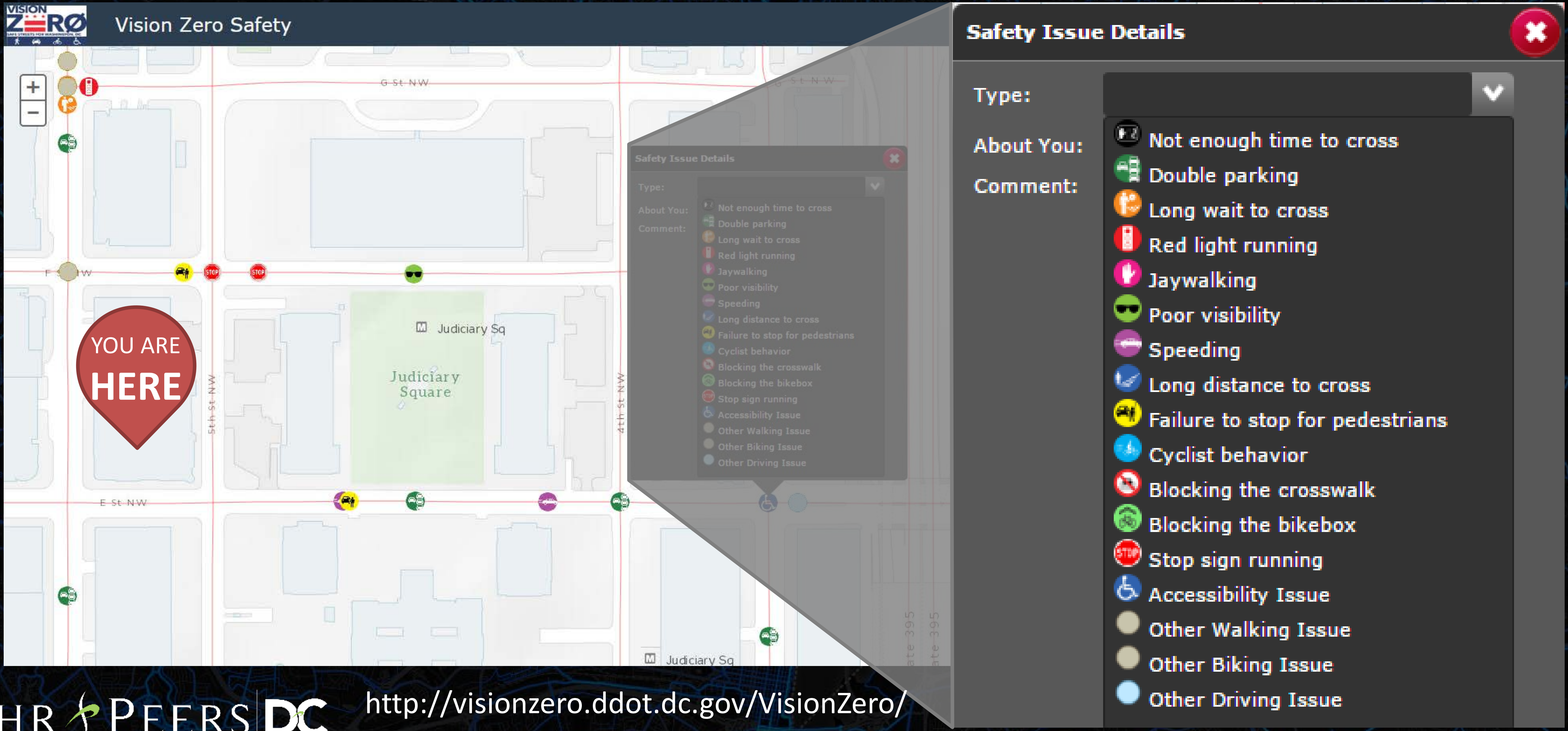
STREETLIGHTDATA



2 | Thematic Data

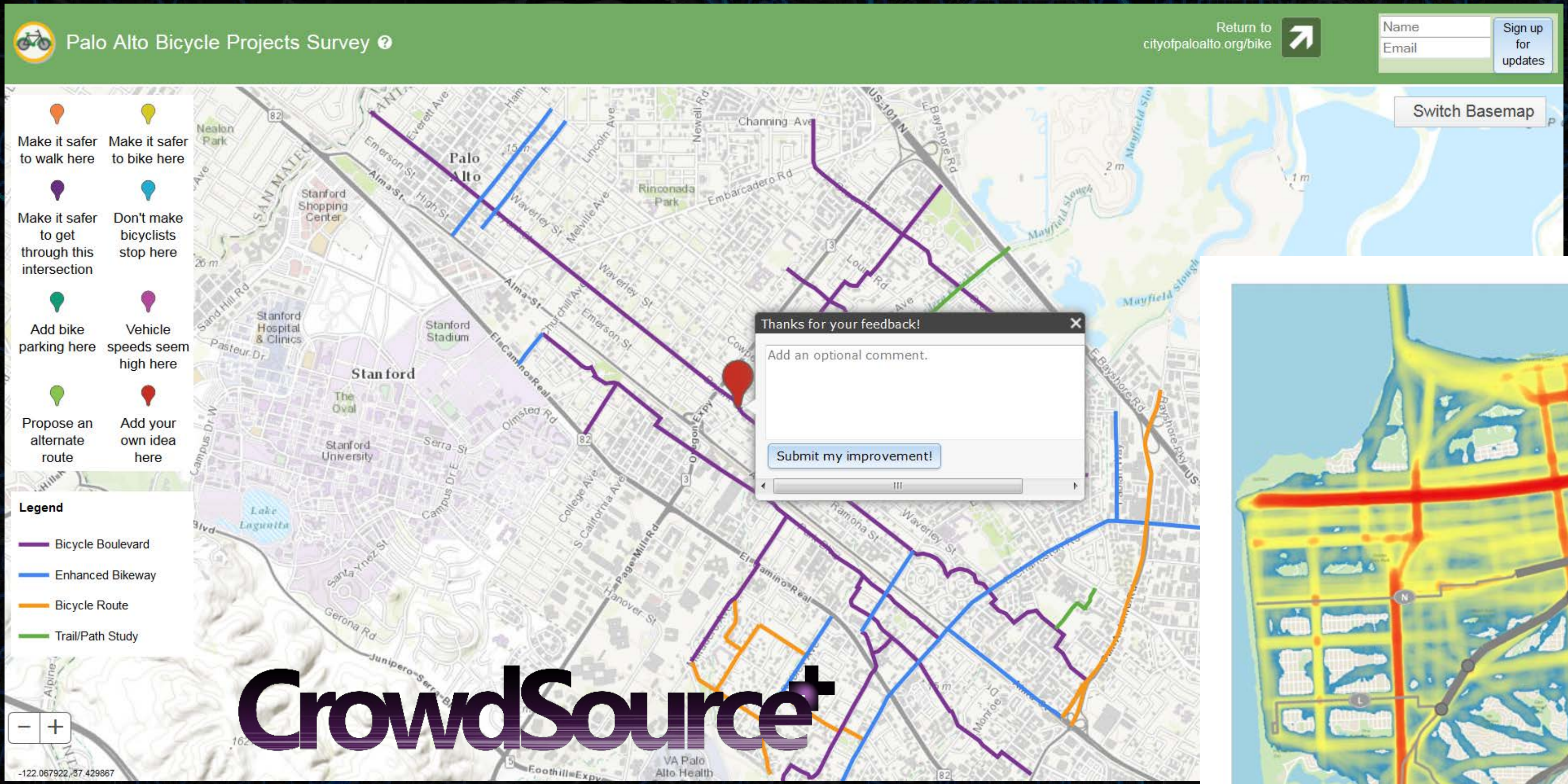


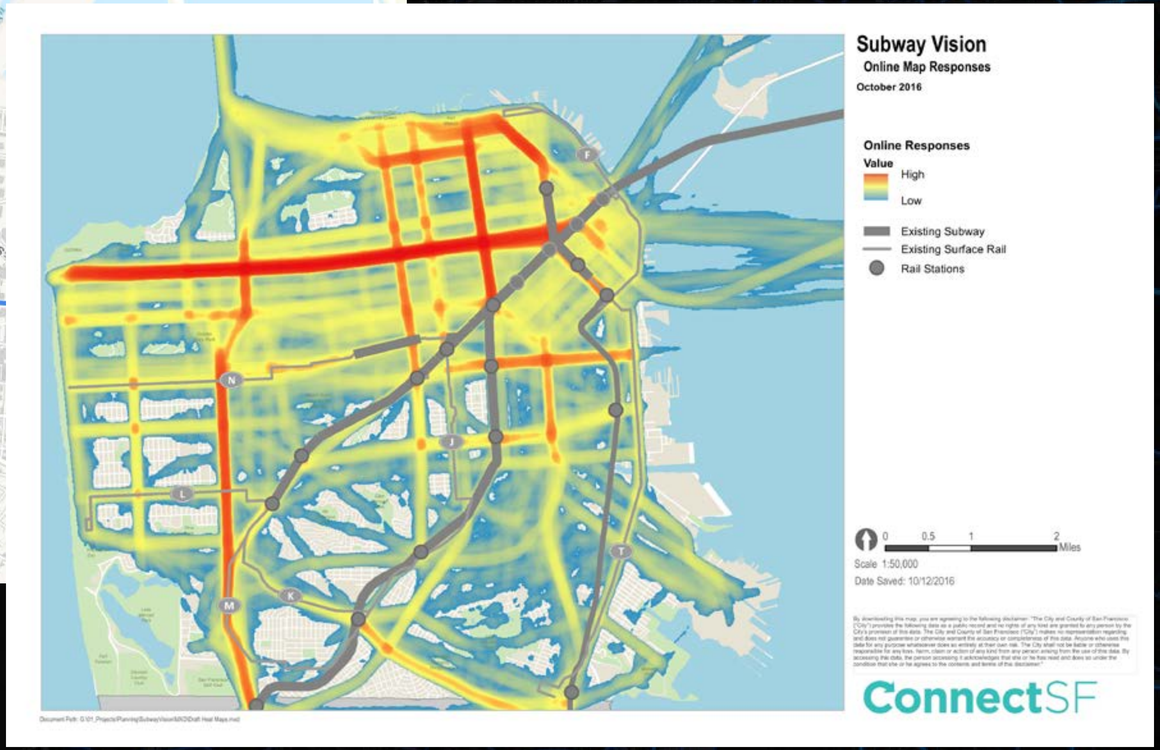
2 | Thumbtack Data



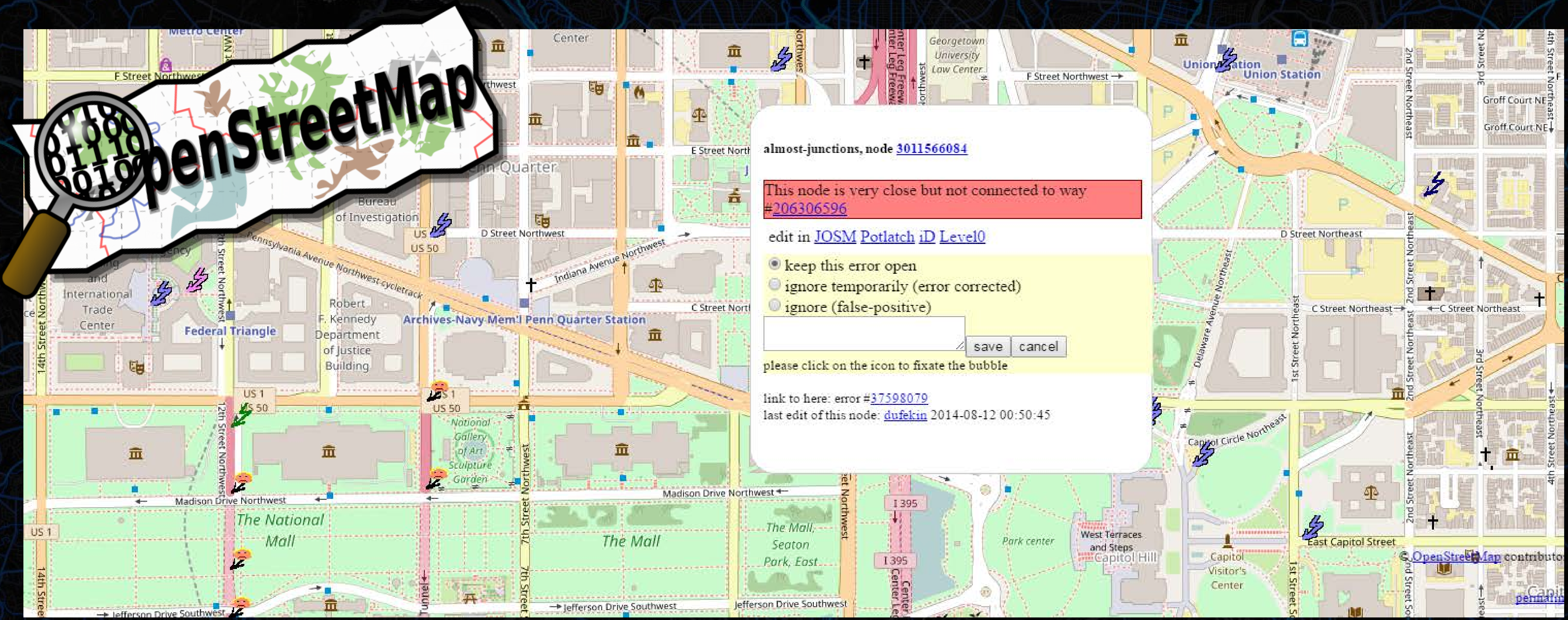
http://visionzero.ddot.dc.gov/VisionZero/

2 | Thumbtack Data

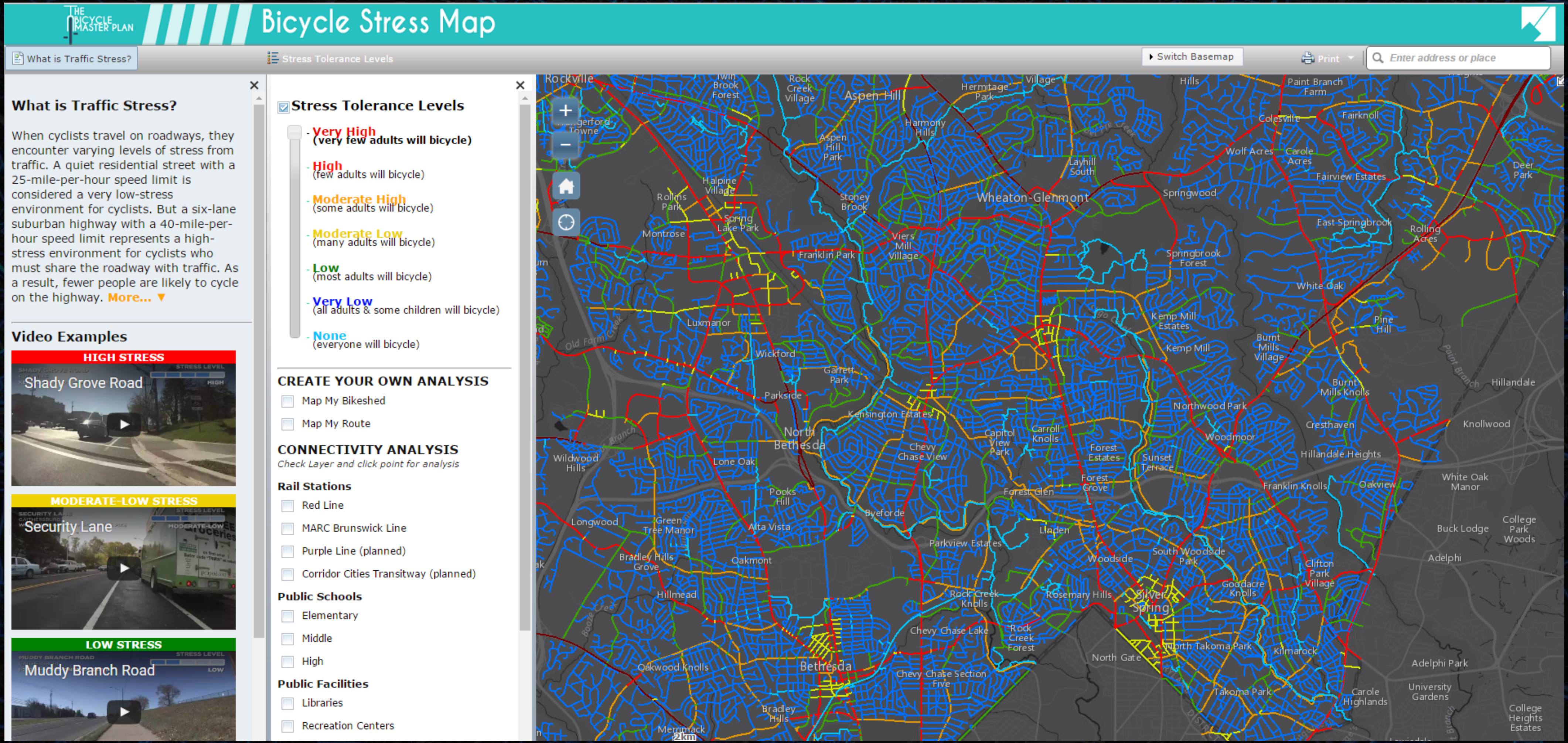




2 | Spatial Inventory Data



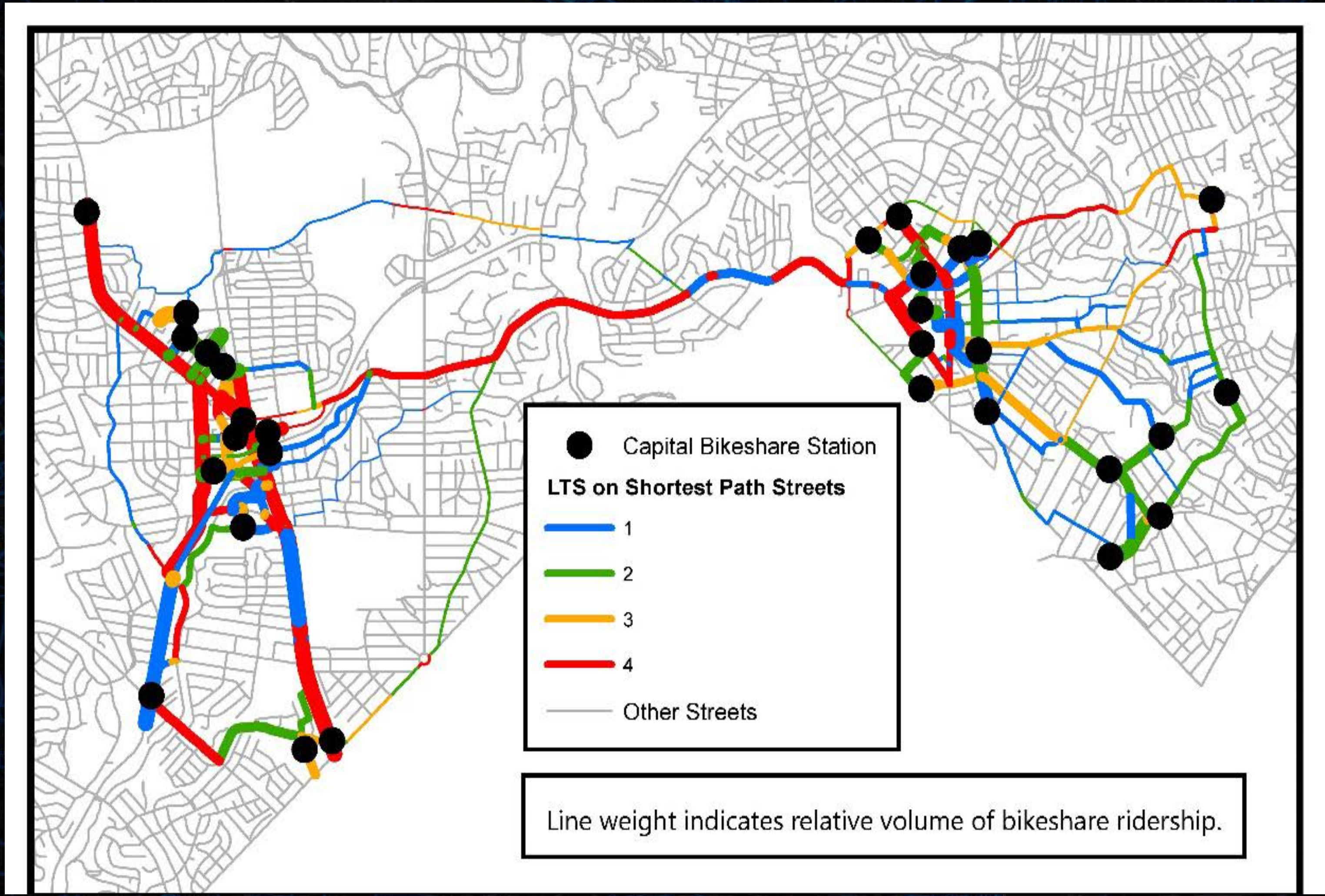
2 | Spatial Inventory Data



3 | (Some of the) Current Data Challenges

- Privacy concerns
- Proprietary vs. Open data
- Biases, e.g:
  - Underreporting of collision data
  - Self-selection
  - Demographic
  - Honesty and validity
- Interpretation

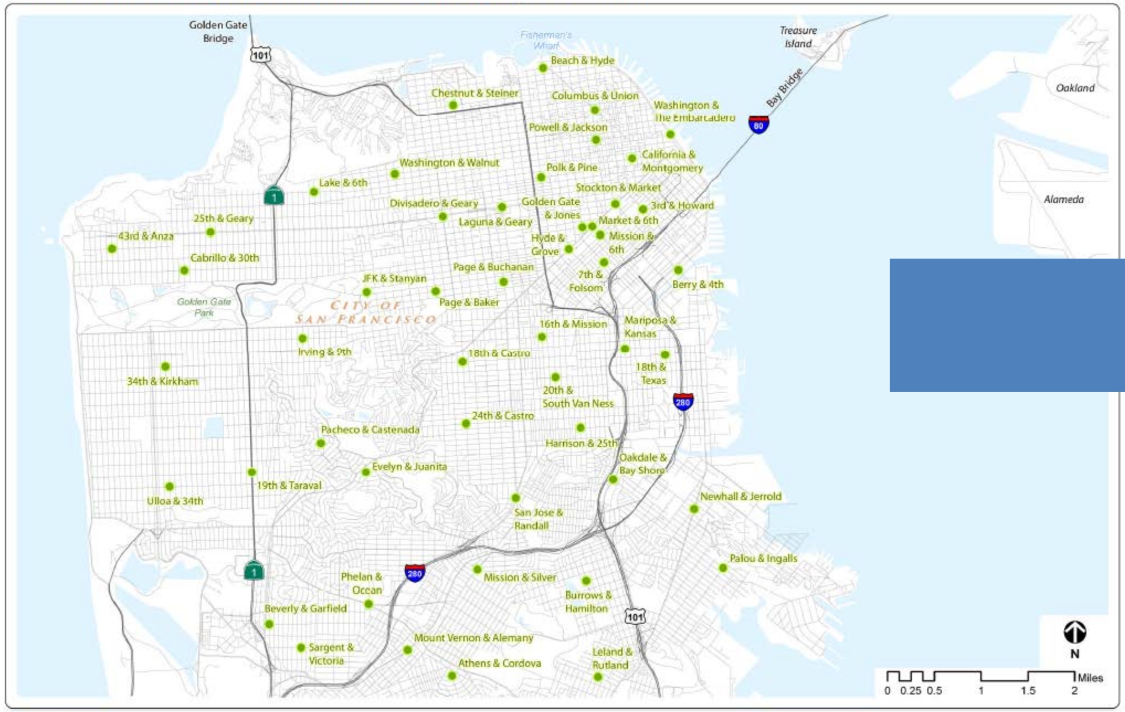
4 | Modeling CaBi Ridership



capital bikeshare™

4 | Estimating Crossing Risk

FIGURE 1 San Francisco Pedestrian Volume Model Study Intersections



(1 mile = 1.61 kilometers)

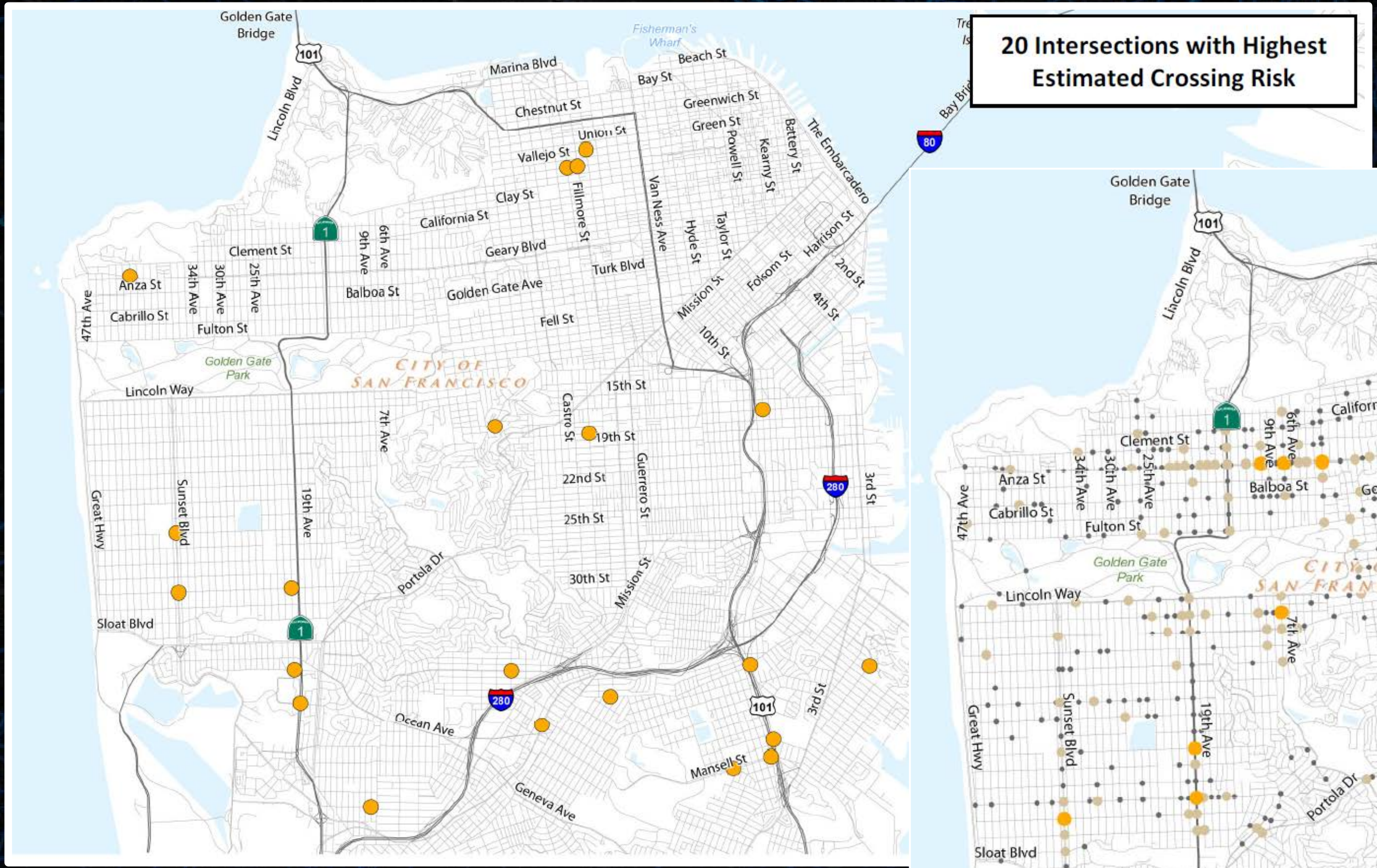
TABLE 4 Preferred San Francisco Pedestrian Volume Model

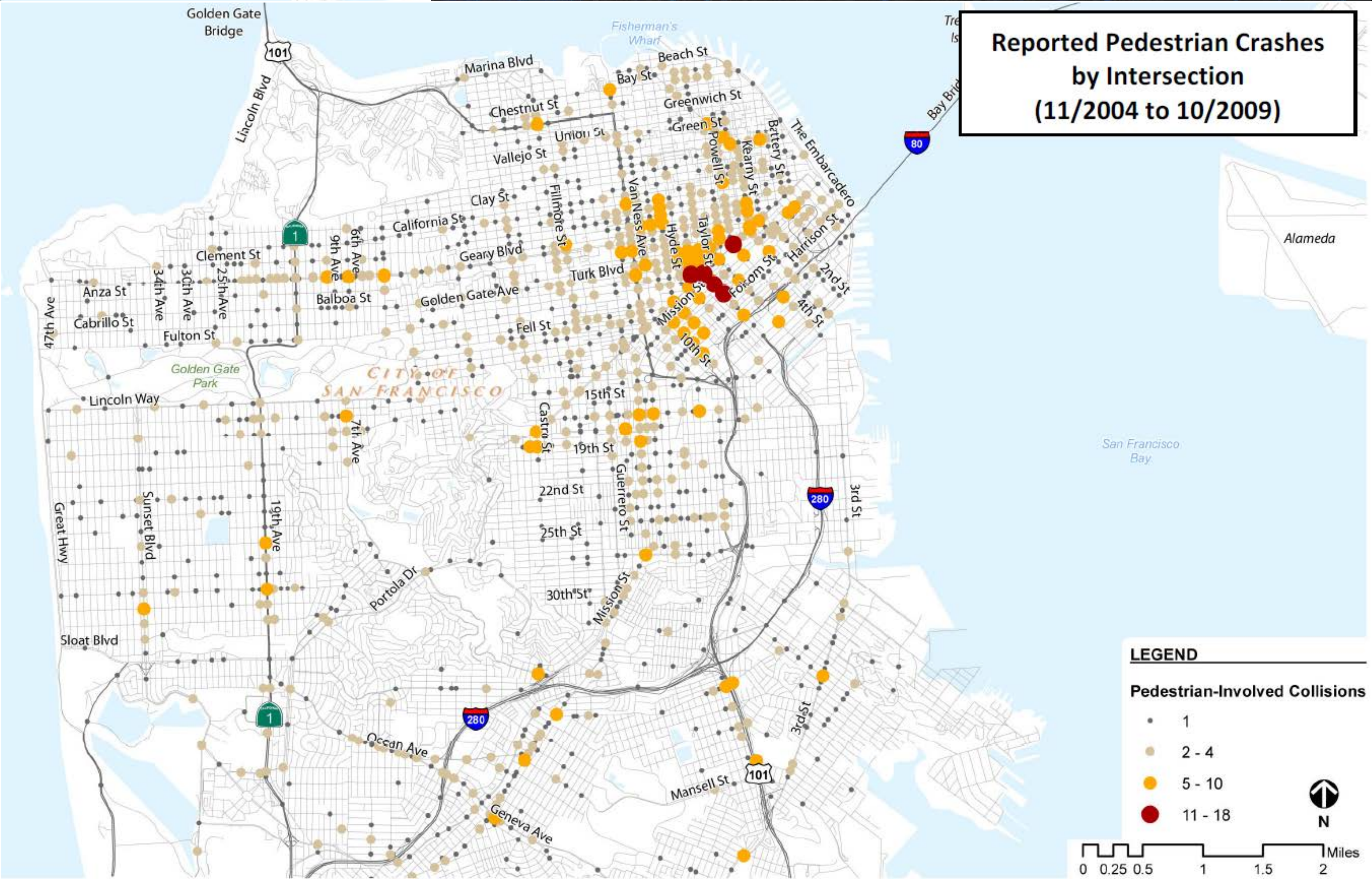
Dependent Variable = Natural Logarithm of Total Annual Pedestrian Intersection Crossings[1]			
	Recommended Model		
Model Variables[2]	**Coefficient**	**t-value**	**p-value**
Total households within 1/4 mile (10,000s)	1.81	2.12	0.040
Total employment within 1/4 mile (100,000s)	2.43	2.22	0.032
Intersection is in a high-activity zone	1.27	3.79	0.000
Maximum slope on any intersection approach leg (100s)	-9.40	-3.07	0.004
Intersection is within 1/4 mile of a university campus	0.635	1.45	0.154
Intersection is controlled by a traffic signal	1.16	4.03	0.000
Constant	12.9	33.29	0.000
Overall Model			
Sample Size (N)	50		
Adjusted R^2-Value	0.804		
F-Value (Test value)	34.4 ($p < 0.001$)		

1) The dependent variable is the natural logarithm of the annual pedestrian intersection crossing volume at each of the 50 study intersections. This represents the sum of all crossings on each approach leg within 50 feet of intersection. The annual volume estimate is extrapolated from a two-hour manual count taken in September 2009 or July-August 2010. The extrapolation method accounts for variations in pedestrian activity by time of day, day of week, weather, and land use.
2) All distances used to calculate the model variables are straight-line distances rather than roadway network distances.

Schneider, Henry, Mitman, Stonehill and Koehler

4 | Estimating Crossing Risk





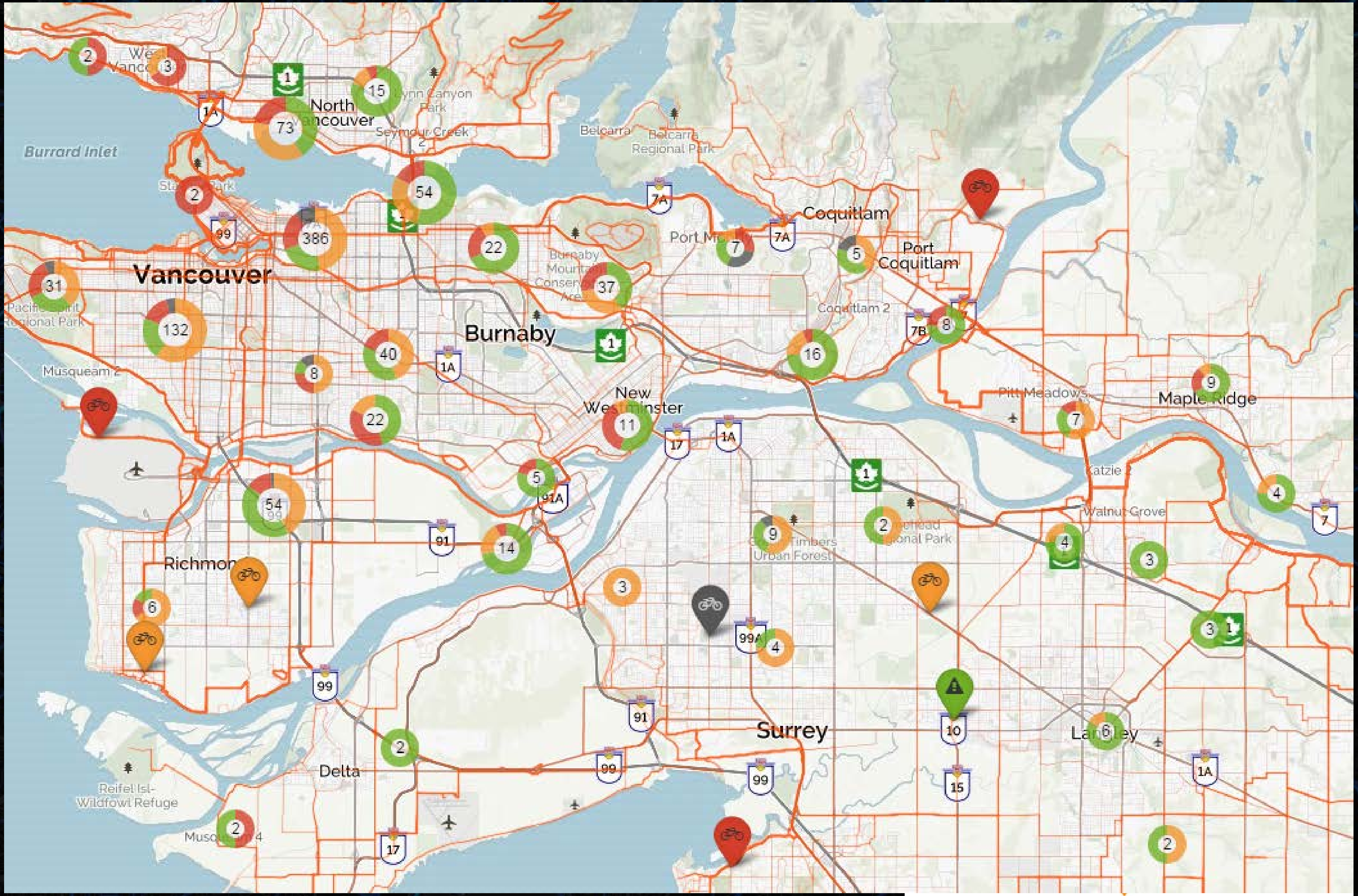
Schneider, Henry, Mitman, Stonehill and Koehler

5 | More Collision and Near-Miss Data?

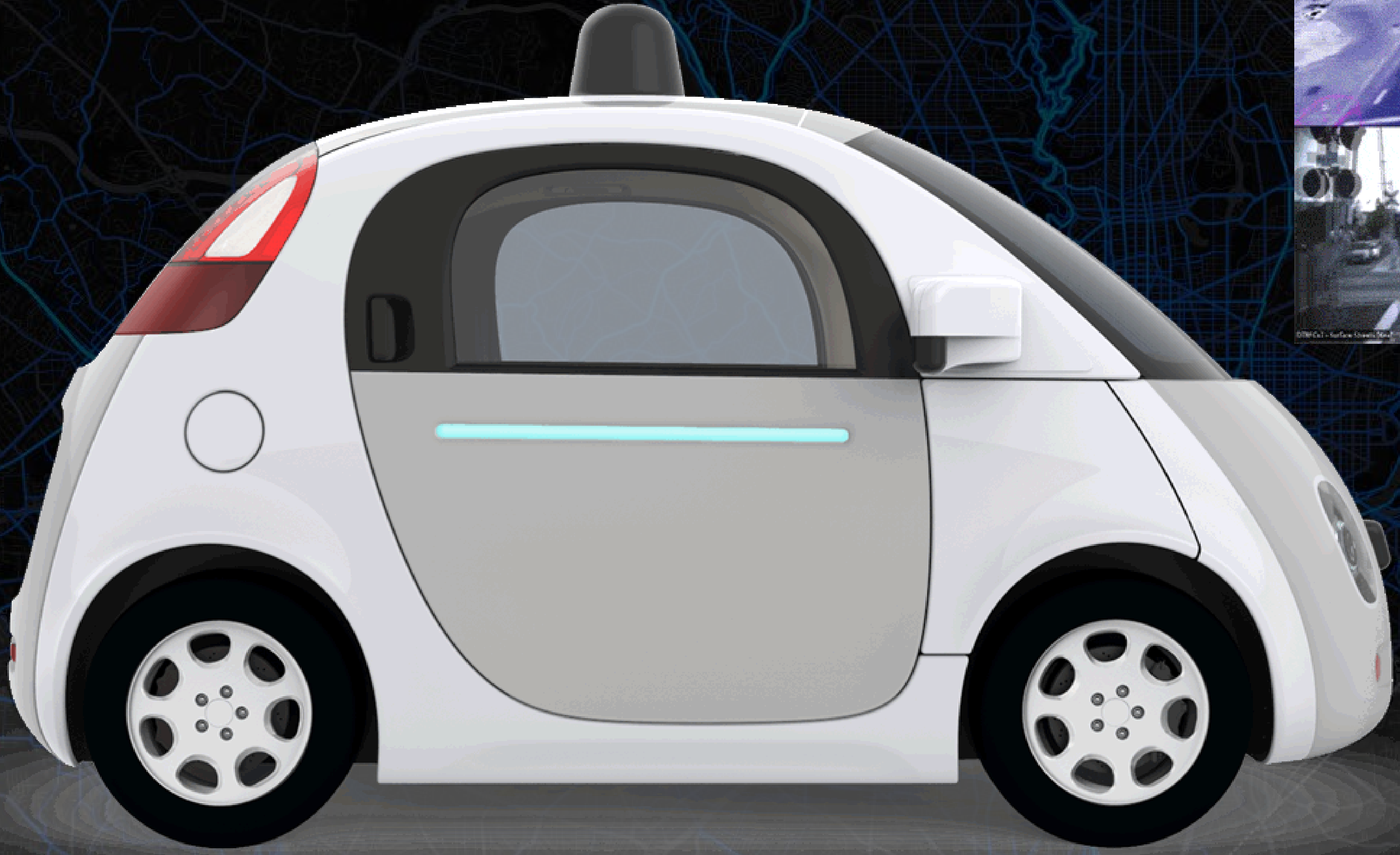
CRASH APP

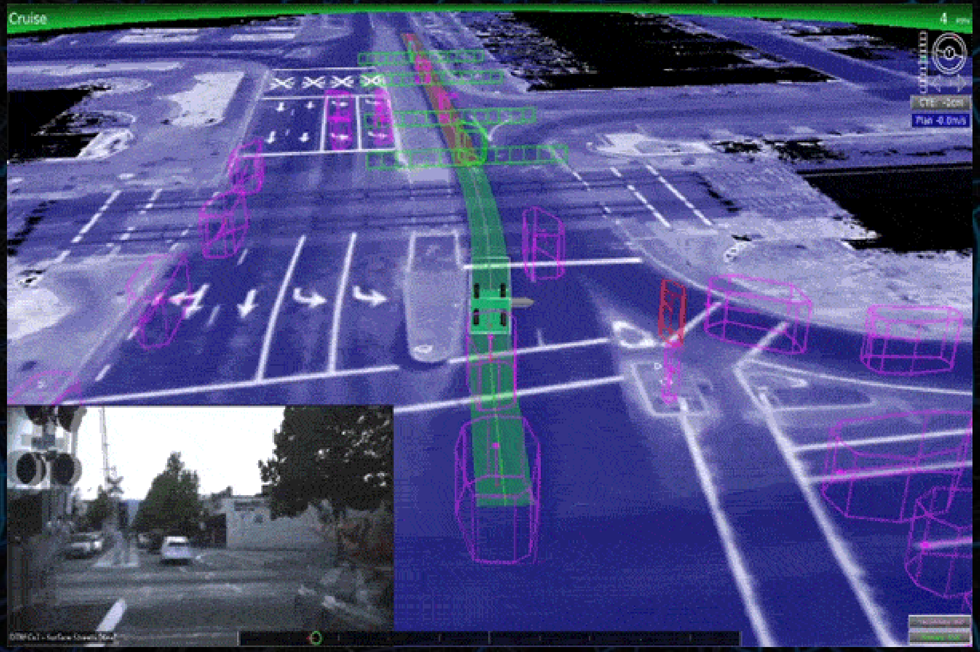
5 | More Collision and Near-Miss Data?

BikeMaps.org



5 | More Infrastructure Data





http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles

Source: Google

5 | More Trip Data

Google

UBER

lyft

5 | StreetLight Data + Cuebiq

- New mobile device data
- Refining "tripifying" algorithm
- New metrics (Spring '17?):
  - Zonal O-D data for active modes
  - All-modes data with likely mode split

6 | So what?

6 | So what? Imagine we have…

- Better collision and near miss data
- Better trip data
- Better spatial data

6 | How can we improve safety (analysis)?

- Improve **reactive** analysis
- Apply more **proactive** analysis
- Develop a **new** approach?

6 | Current Practice – Reactive Analysis

Collisions (Injuries, Fatalities)

6 | Current Practice – Reactive Analysis

$$\frac{\#\text{ Collisions (Injuries, Fatalities)}}{\text{Exposure}} = \text{Rate!}$$

6 | Current Practice – Reactive Analysis

$$\frac{\text{\# Collisions (Injuries, Fatalities)}}{\text{Exposure}} = \text{Rate!}$$

Assess Collision Profile → Identify Countermeasure

6 | Improving Reactive Analysis

Easier self-reporting

More comprehensive professional reporting

$$\frac{\text{\# Collisions (Injuries, Fatalities)}}{\text{Exposure}} = \text{Rate!}$$

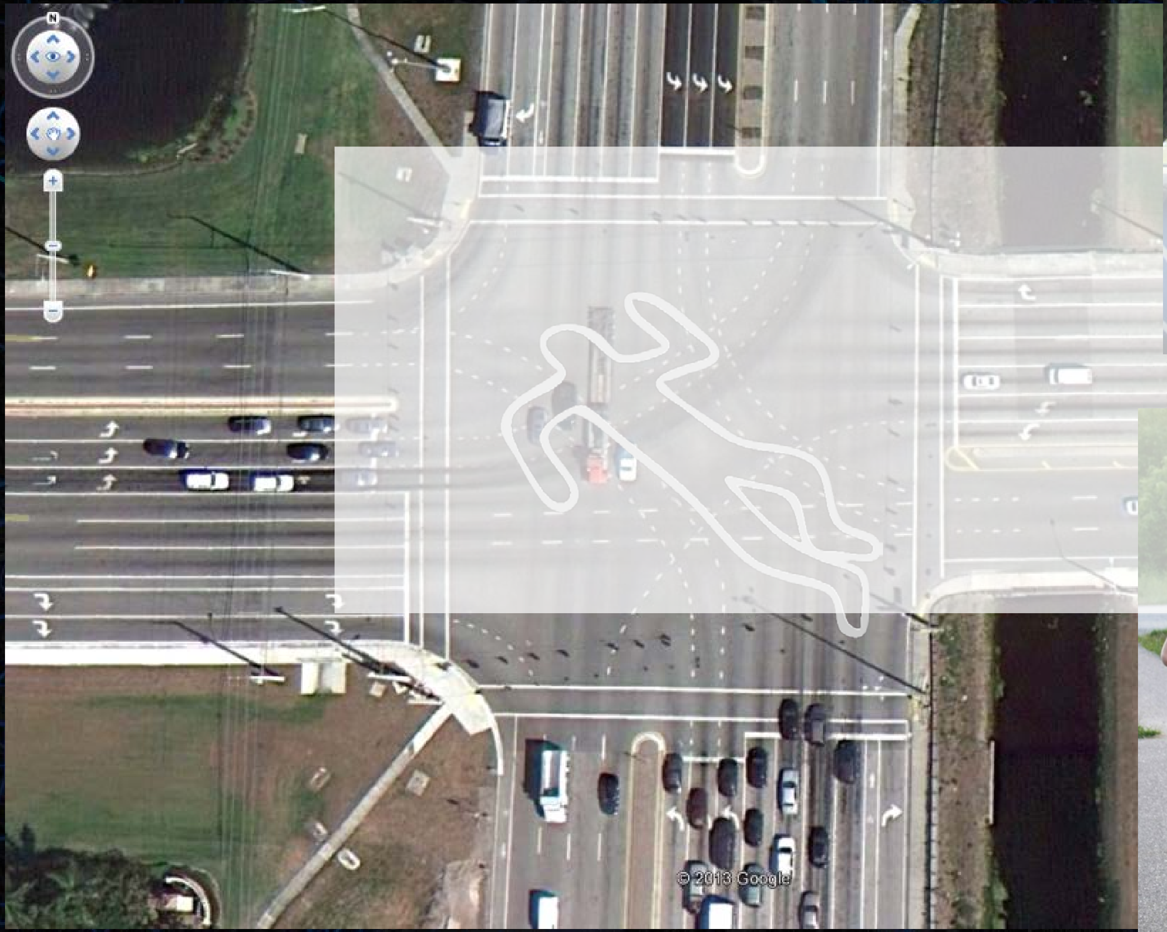
*Better!

Better trip data

6 | Conduct Proactive Analysis

- Where is ped&bike activity suppressed by unsafe conditions? Can short trip data help ID?
- Can we attempt better predictive analysis based on characteristics of known unsafe locations?
- Can we design to prevent injuries and fatalities before they become “hotspots?”

6 | Flipping the script?



6 | Measuring Safety – "Units of Bad per Good"

Crashes/Injuries/Fatalities

per 100 Million VMT

per Million
Entering Vehicles

per Resident

per Employee

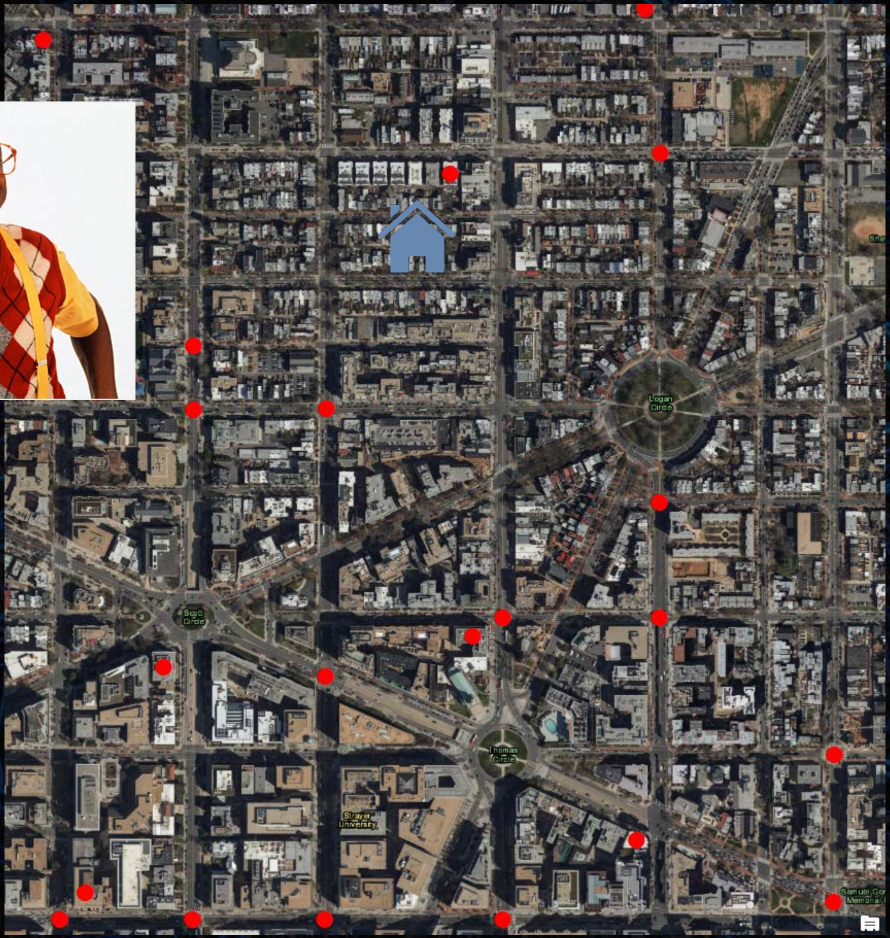
per Grocery Trip

per $ of Goods Delivered

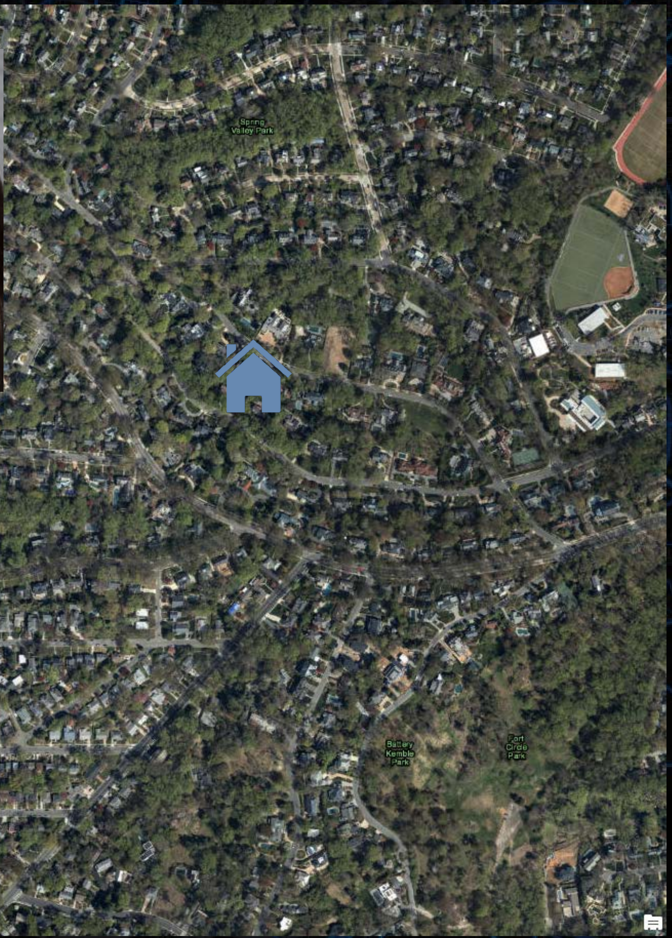
A multimodal full accounting of the safety costs ("risks") of travel

6 | Safer for whom? Measuring an externality…

Logan Circle



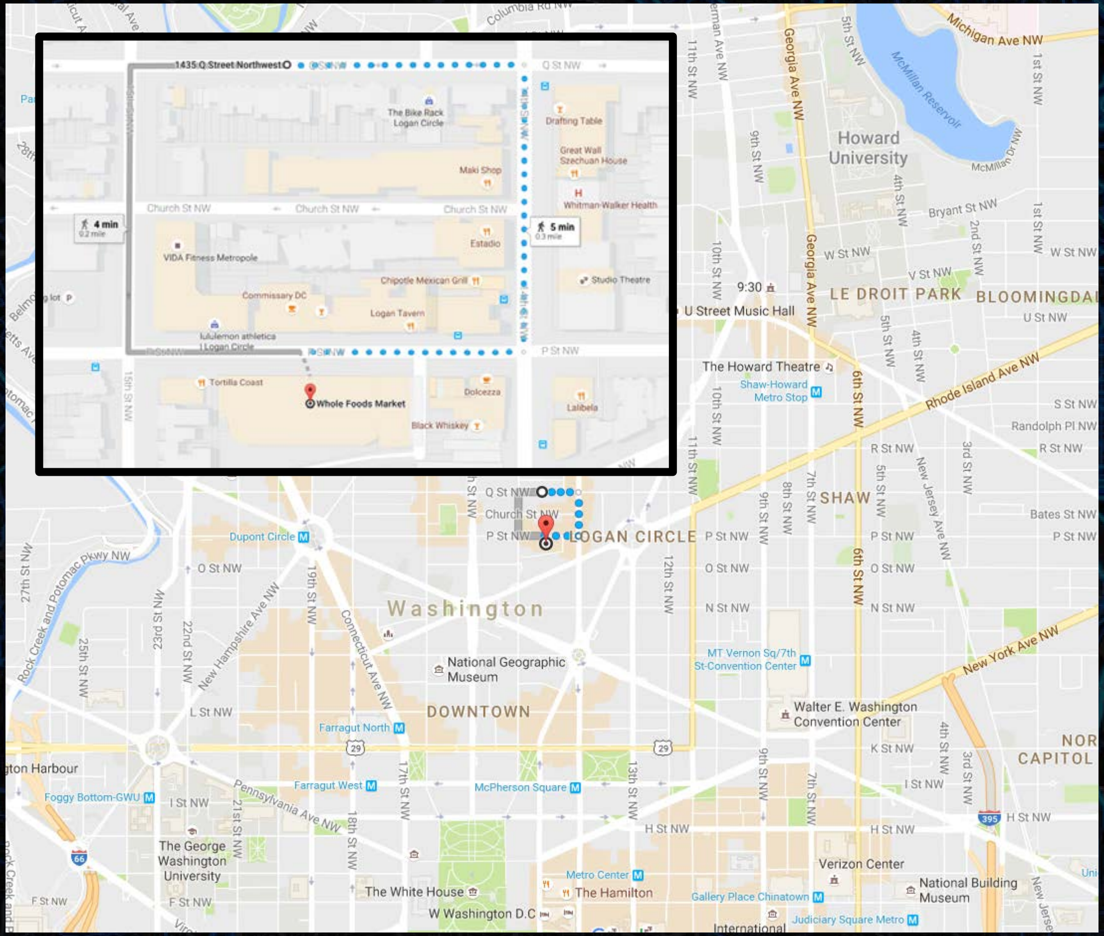
Spring Valley



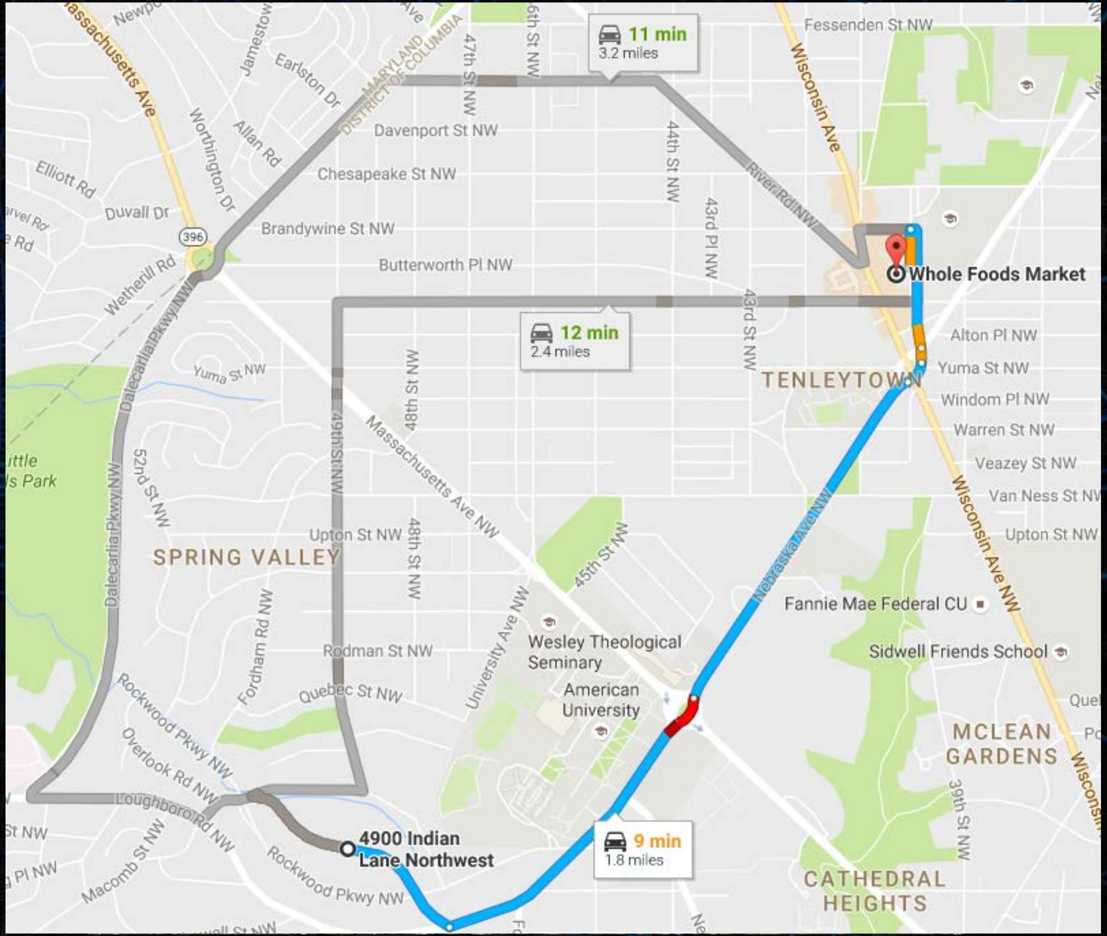
● Injury or Fatality Collision Involving Pedestrian

6 | Grocery trip

Logan Circle



Spring Valley



6 | Grocery trip

Logan Circle

0.2 miles

2 intersections

0 crash history locations

Spring Valley

1.8 miles

20 intersections

1 U-turn

1 crash history location

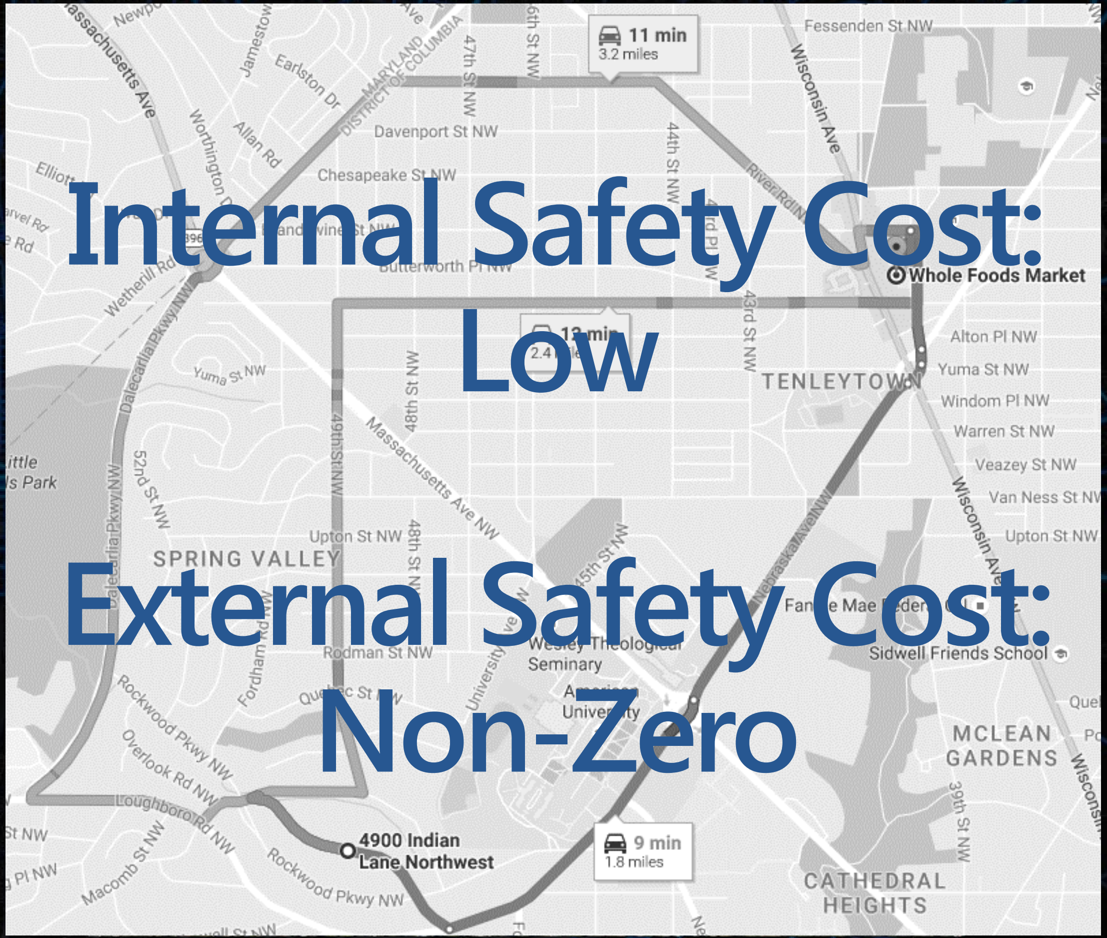
6 | Grocery trip safety costs (risks)

Logan Circle

Internal Safety Cost: Low

External Safety Cost: Zero?

Spring Valley



Internal Safety Cost: Low

External Safety Cost: Non-Zero

FEHR PEERS DC
Crowdsourcing Ped+Bike Data

6 | Grocery trip + commute trip + lunch trip + …

Logan Circle

Total Safety Cost: ?

Spring Valley

Total Safety Cost: ?

6 | Non-Intersection Countermeasures?

- Can we improve safety by…
  - Designing safer land use + transportation systems?
  - Reducing VMT and vehicle speeds on local roads while helping travelers accomplish their trip purposes?

How else can we use data to improve bike+ped safety?

Thank you!

FEHR PEERS DC

Alex Rixey
@alexrixey