

# Crowdsourcing Pedestrian and Bicyclist Activity Data

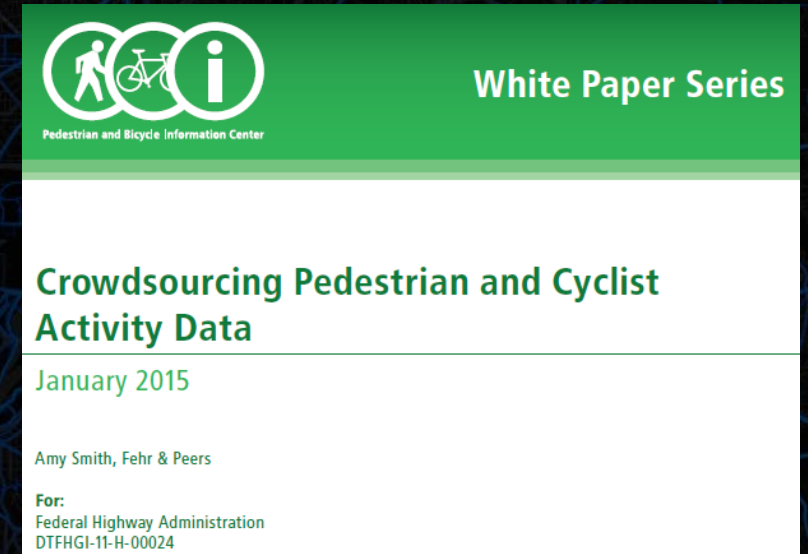
---

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?



# 1 | Crowdsourced Data Typology



**In-Situ**



**Thematic**



**Thumbtack**



**Spatial  
Inventory**

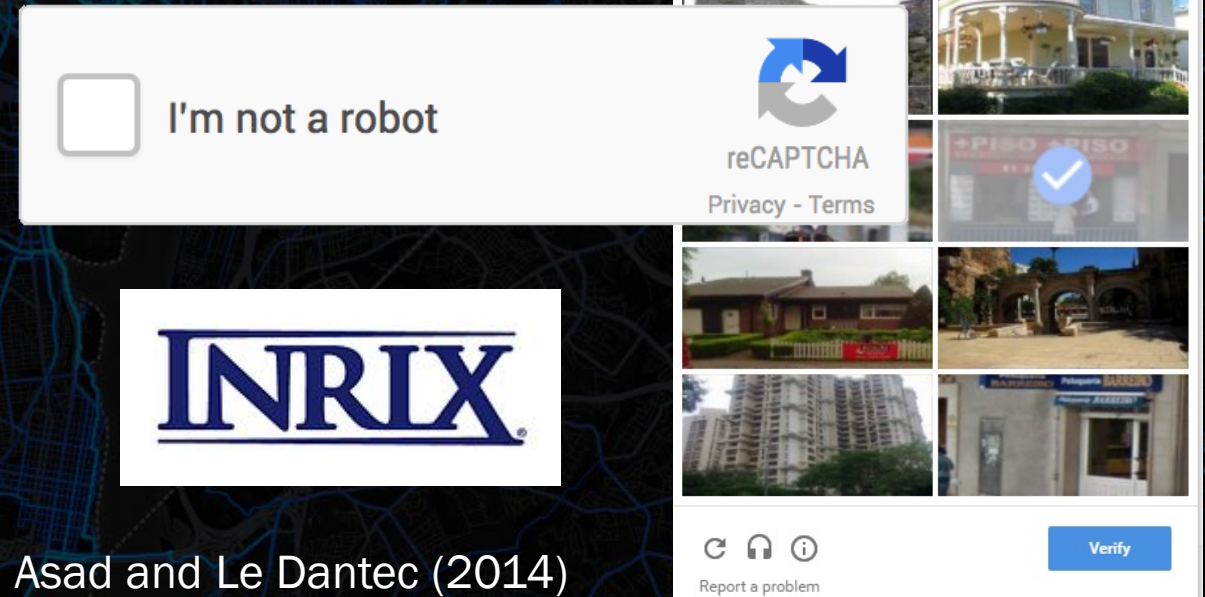
# 1 | Additional Crowdsourced Data Characteristics

## Explicit

Defined problem communicated directly to participants

## Implicit

Participants may be unaware of secondary use of their data



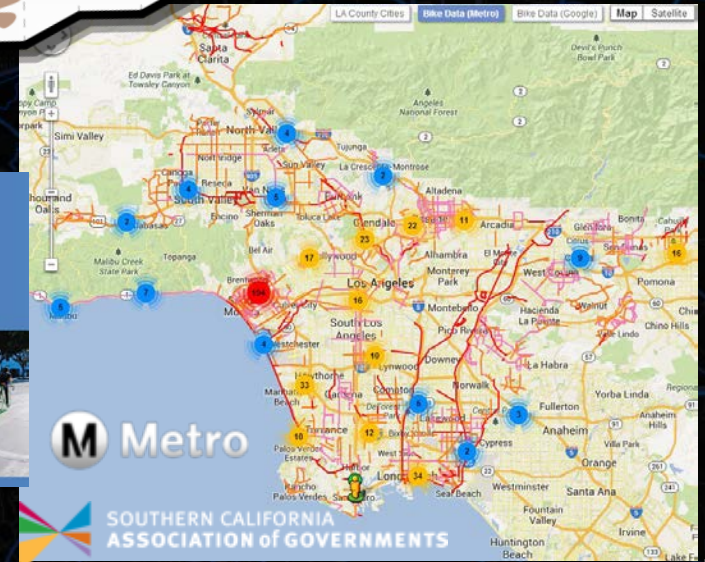
# 1 | Additional Crowdsourced Data Characteristics

## General Purpose

Does not require specialized knowledge from participants

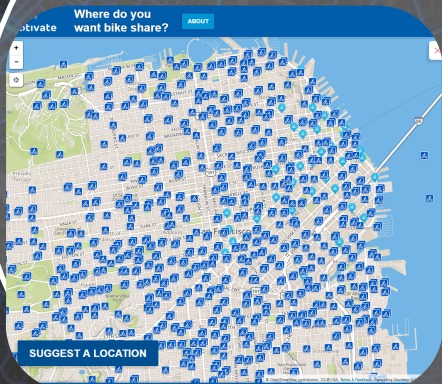
## Domain-Specific

Data collected from participants with existing expertise



# 1 | Additional Crowdsourced Data Characteristics

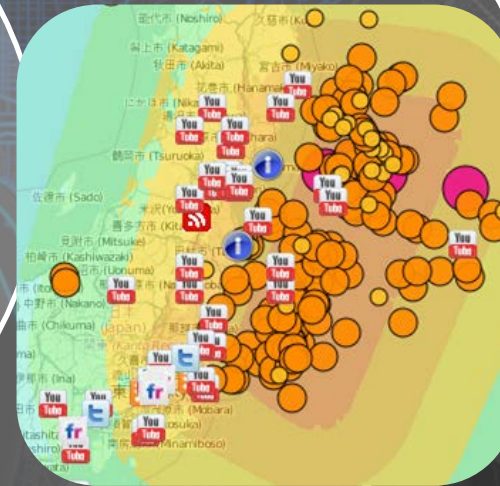
Geocentric



Audience-Centric



Event-Centric

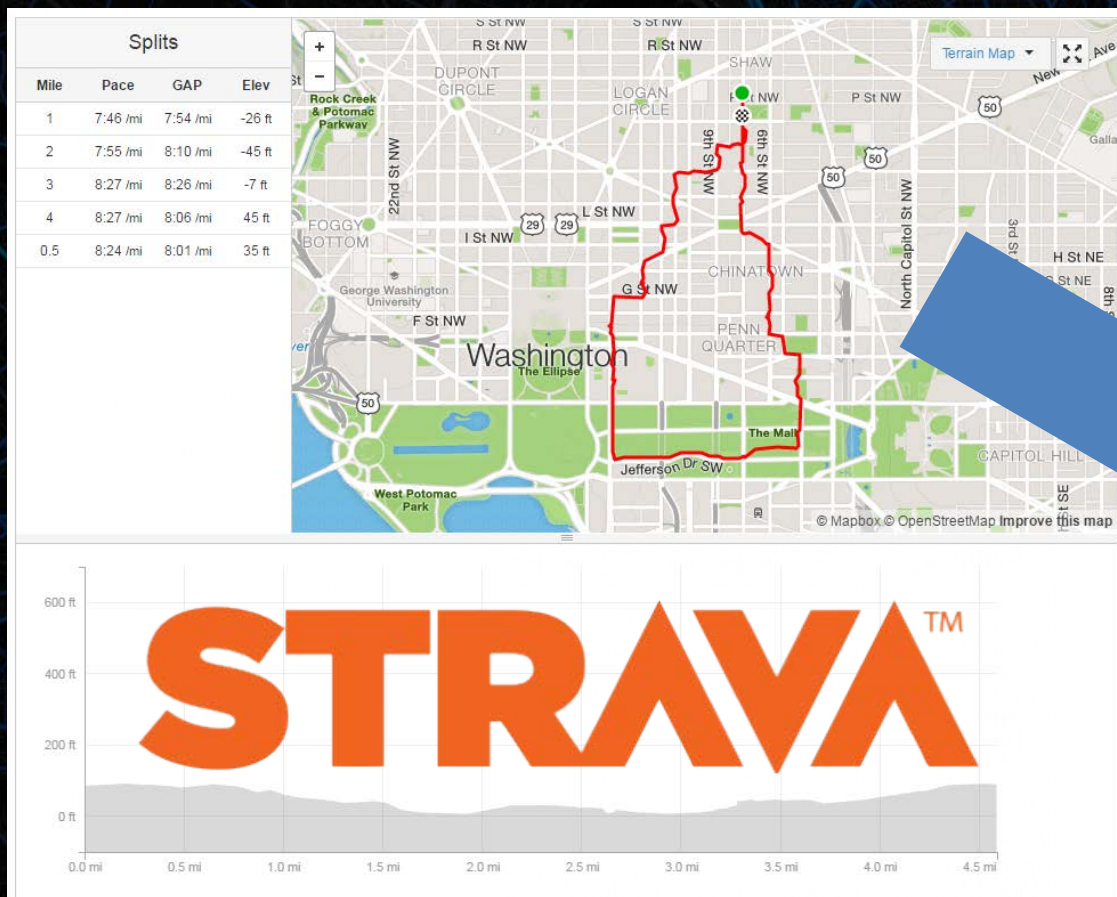


Global

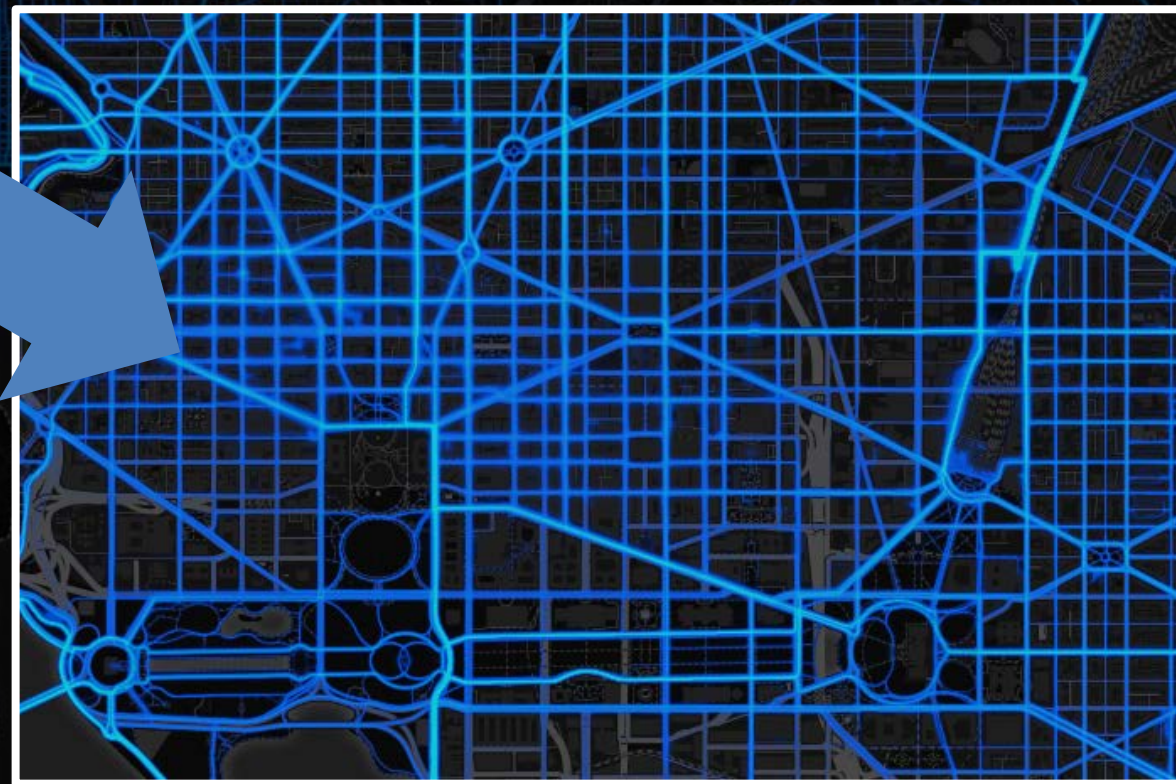




## 2 | In-Situ Data (Explicit)



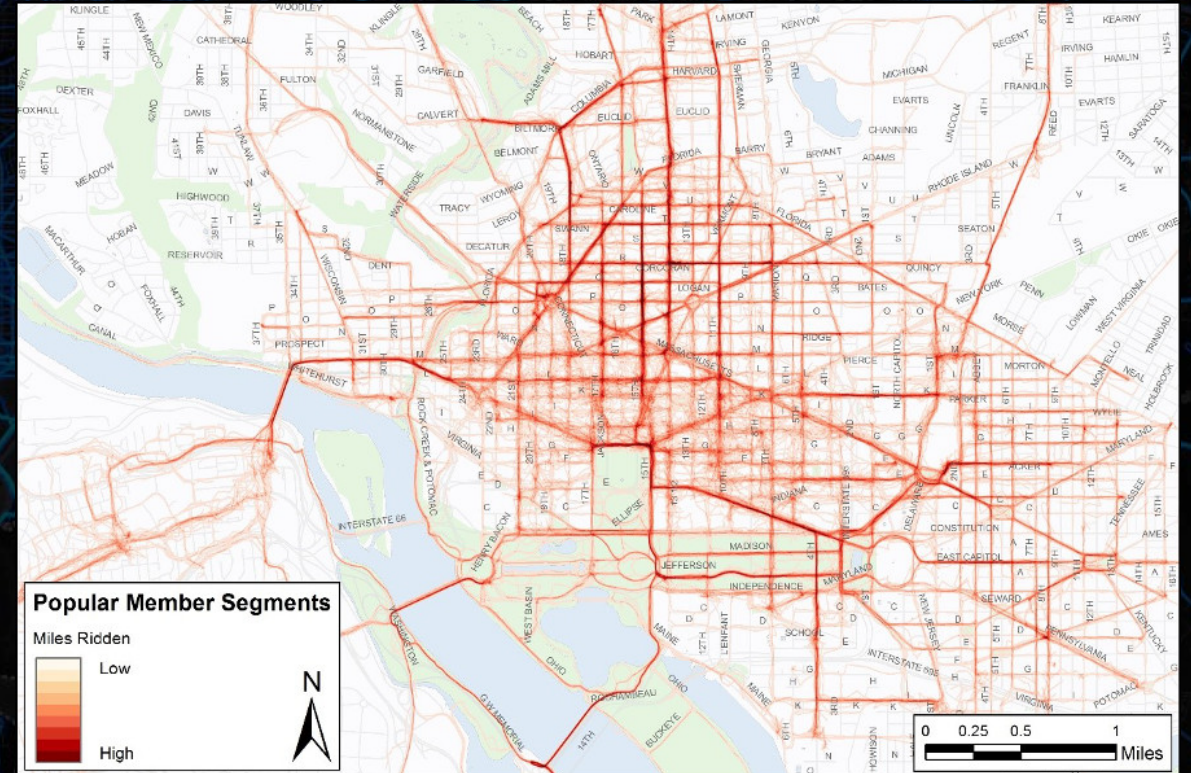
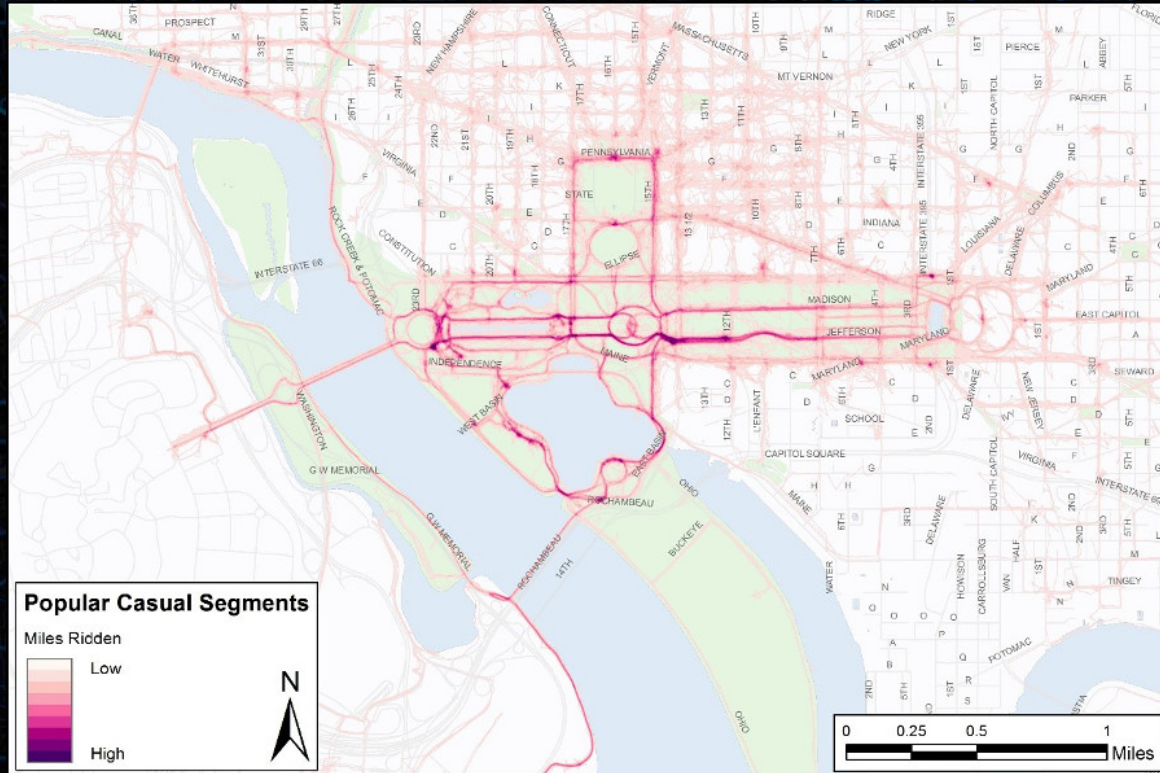
**STRAVA** | METRO



# 2 | In-Situ Data (Implicit)



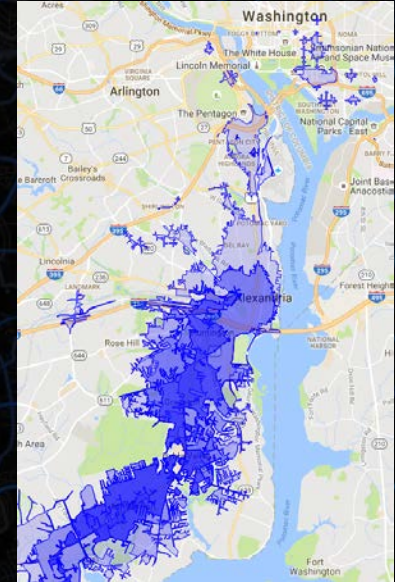
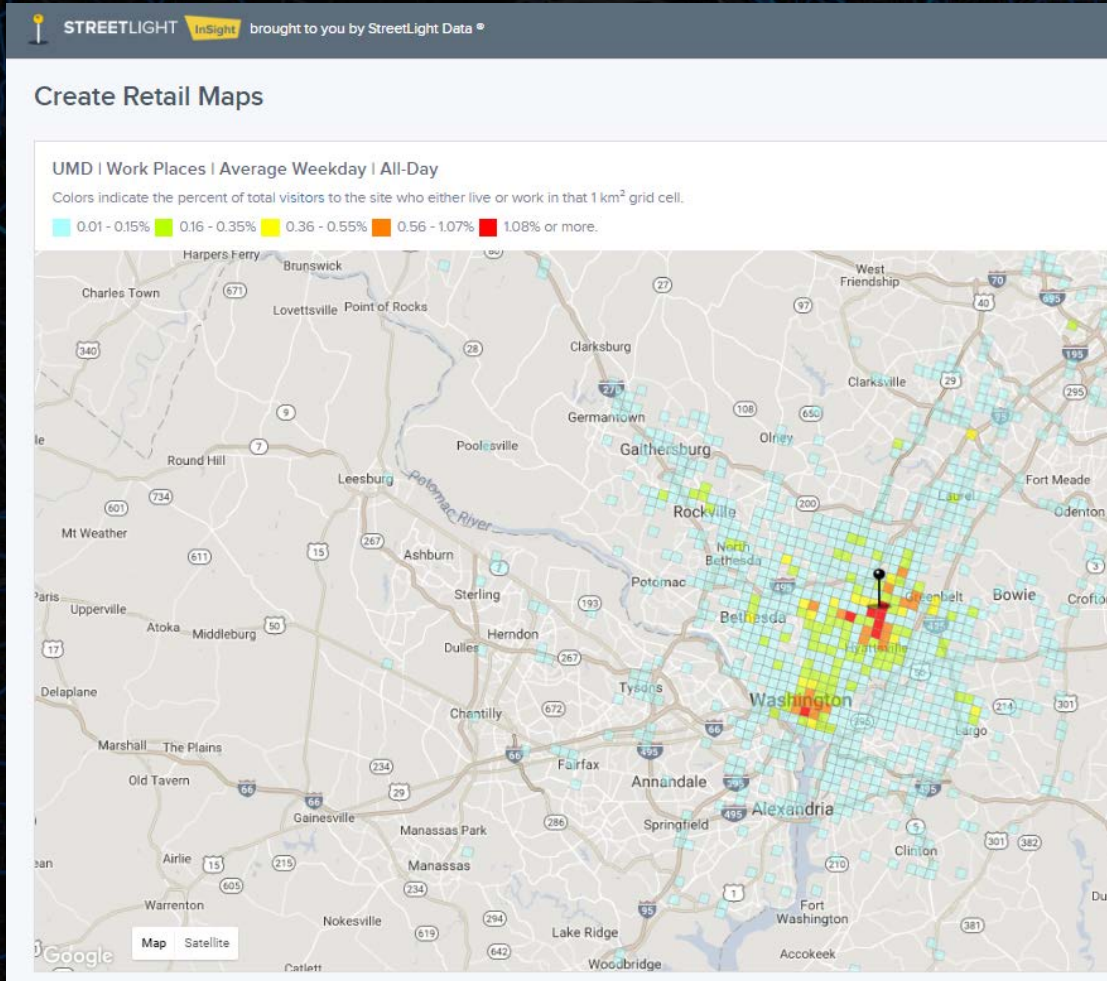
# capital bikeshare™



Source: Joe Wergin

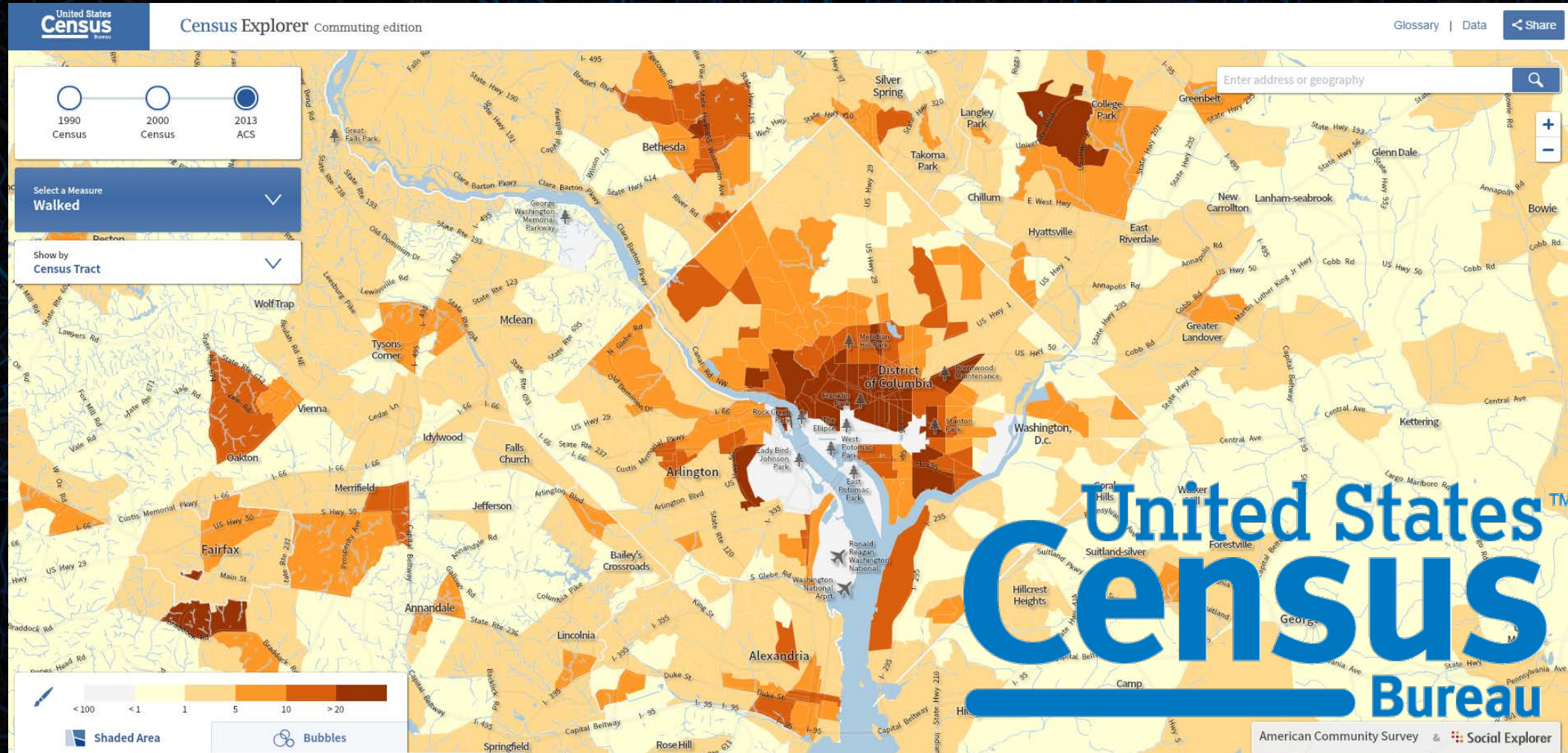


# 2 | In-Situ Data (Implicit)





# 2 | Thematic Data





# 2 | Thumbtack Data

**VISION ZERO**  
SAFE STREETS FOR WASHINGTON DC

### Vision Zero Safety

**YOU ARE HERE**

Map labels: G St-NW, F St-NW, E St-NW, 5th St-NW, 4th St-NW, Judiciary Sq, Judiciary Square.

#### Safety Issue Details

Type: [Dropdown]

About You: [Dropdown]

Comment: [Text Area]

- Not enough time to cross
- Double parking
- Long wait to cross
- Red light running
- Jaywalking
- Poor visibility
- Speeding
- Long distance to cross
- Failure to stop for pedestrians
- Cyclist behavior
- Blocking the crosswalk
- Blocking the bikebox
- Stop sign running
- Accessibility Issue
- Other Walking Issue
- Other Biking Issue
- Other Driving Issue



# 2 | Thumbtack Data

**Palo Alto Bicycle Projects Survey** [Return to cityofpaloalto.org/bike](#)

Switch Basemap

- Make it safer to walk here
- Make it safer to bike here
- Make it safer to get through this intersection
- Don't make bicyclists stop here
- Add bike parking here
- Vehicle speeds seem high here
- Propose an alternate route
- Add your own idea here

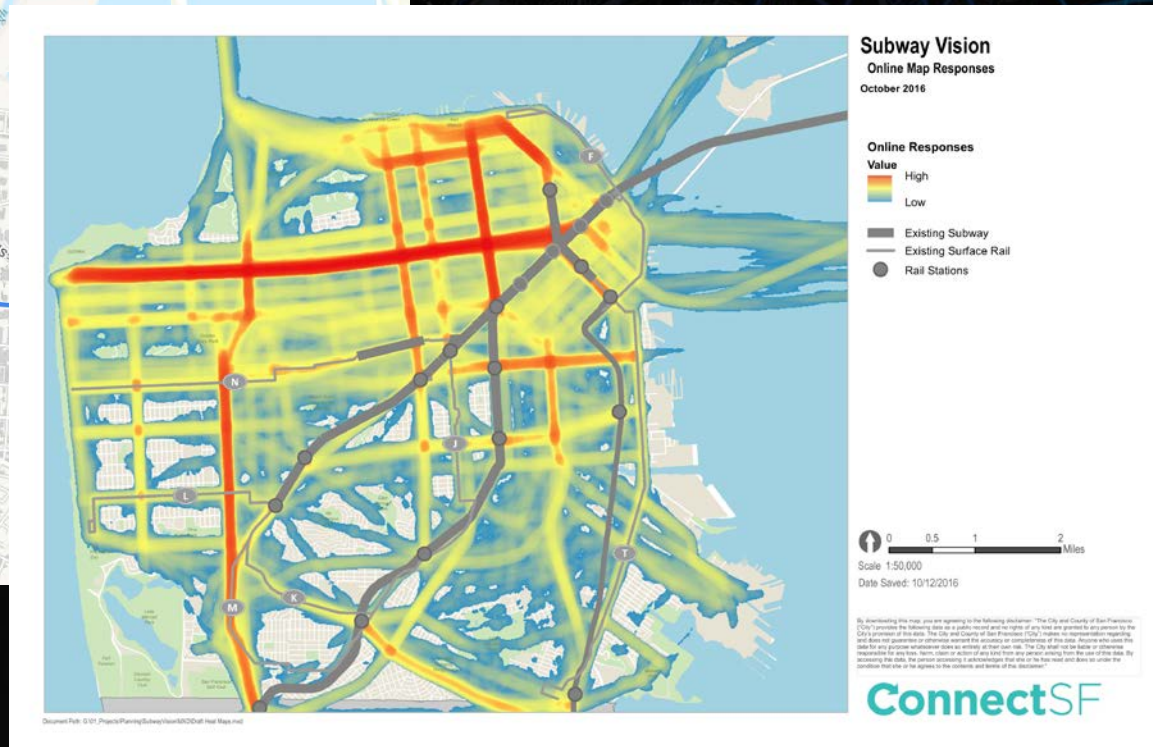
**Legend**

- Bicycle Boulevard
- Enhanced Bikeway
- Bicycle Route
- Trail/Path Study

**Thanks for your feedback!**

Add an optional comment.

**CrowdSource+**





# 2 | Spatial Inventory Data



almost-junctions, node [3011566084](#)

This node is very close but not connected to way [#206306596](#)

edit in [JOSM](#) [Potlatch](#) [iD](#) [Level0](#)

- keep this error open
- ignore temporarily (error corrected)
- ignore (false-positive)

save cancel

please click on the icon to fixate the bubble

link to here: error [#37598079](#)  
last edit of this node: [dufekin](#) 2014-08-12 00:50:45



# 2 | Spatial Inventory Data

### THE BICYCLE MASTER PLAN

## Bicycle Stress Map

What is Traffic Stress? Stress Tolerance Levels

Switch Basemap Print Enter address or place

#### What is Traffic Stress?

When cyclists travel on roadways, they encounter varying levels of stress from traffic. A quiet residential street with a 25-mile-per-hour speed limit is considered a very low-stress environment for cyclists. But a six-lane suburban highway with a 40-mile-per-hour speed limit represents a high-stress environment for cyclists who must share the roadway with traffic. As a result, fewer people are likely to cycle on the highway. [More...](#)

#### Stress Tolerance Levels

- Very High** (very few adults will bicycle)
- High** (few adults will bicycle)
- Moderate High** (some adults will bicycle)
- Moderate Low** (many adults will bicycle)
- Low** (most adults will bicycle)
- Very Low** (all adults & some children will bicycle)
- None** (everyone will bicycle)

#### CREATE YOUR OWN ANALYSIS

- Map My Bikedshed
- Map My Route

#### CONNECTIVITY ANALYSIS

Check Layer and click point for analysis

#### Rail Stations

- Red Line
- MARC Brunswick Line
- Purple Line (planned)
- Corridor Cities Transitway (planned)

#### Public Schools

- Elementary
- Middle
- High

#### Public Facilities

- Libraries
- Recreation Centers

#### Video Examples

- HIGH STRESS**  
Shady Grove Road
- MODERATE-LOW STRESS**  
Security Lane
- LOW STRESS**  
Muddy Branch Road

# 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

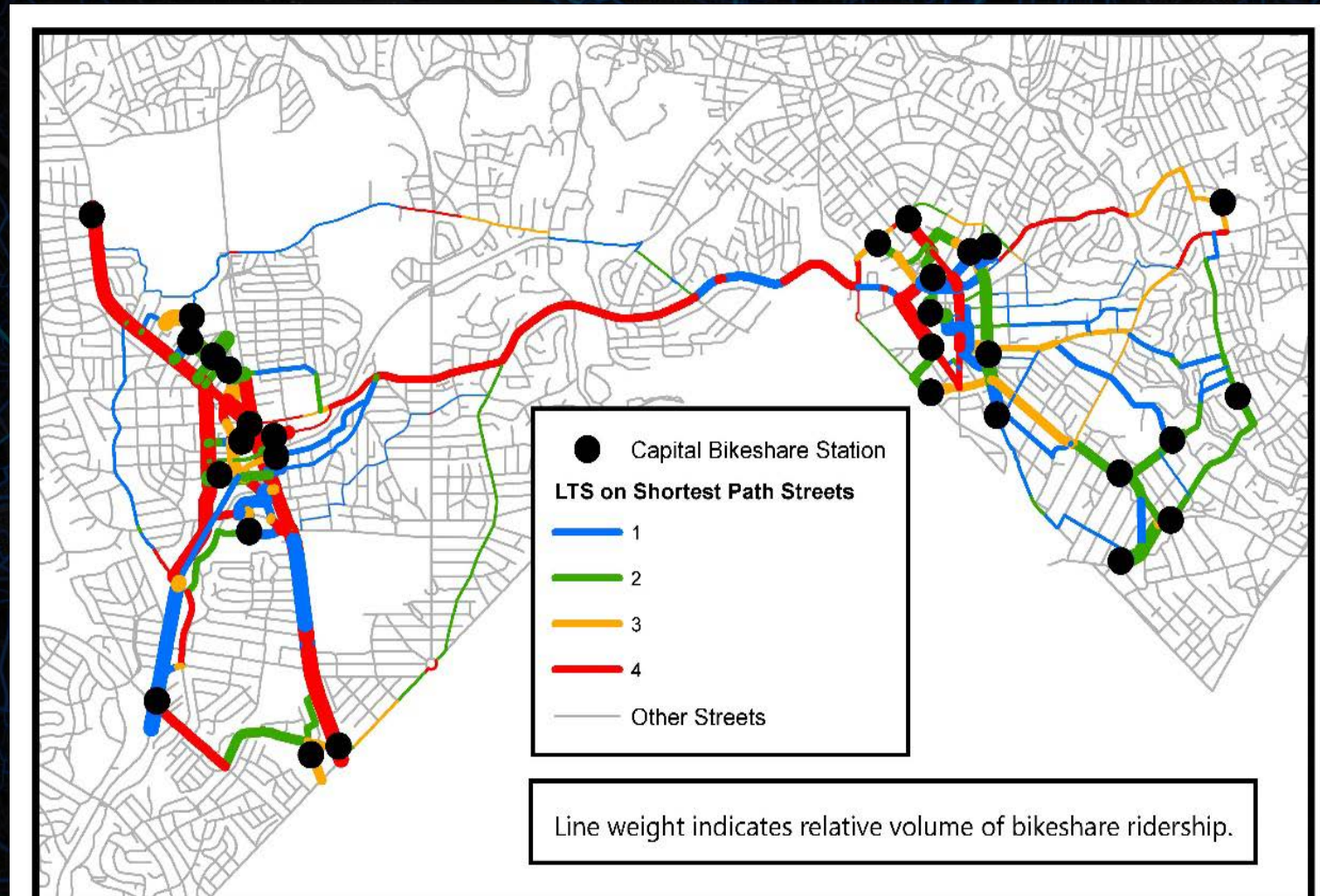
# 4 | Modeling CaBi Ridership



+



+



capital  
bikeshare™



# 4 | Estimating Crossing Risk



FIGURE 1 San Francisco Pedestrian Volume Model Study Intersections

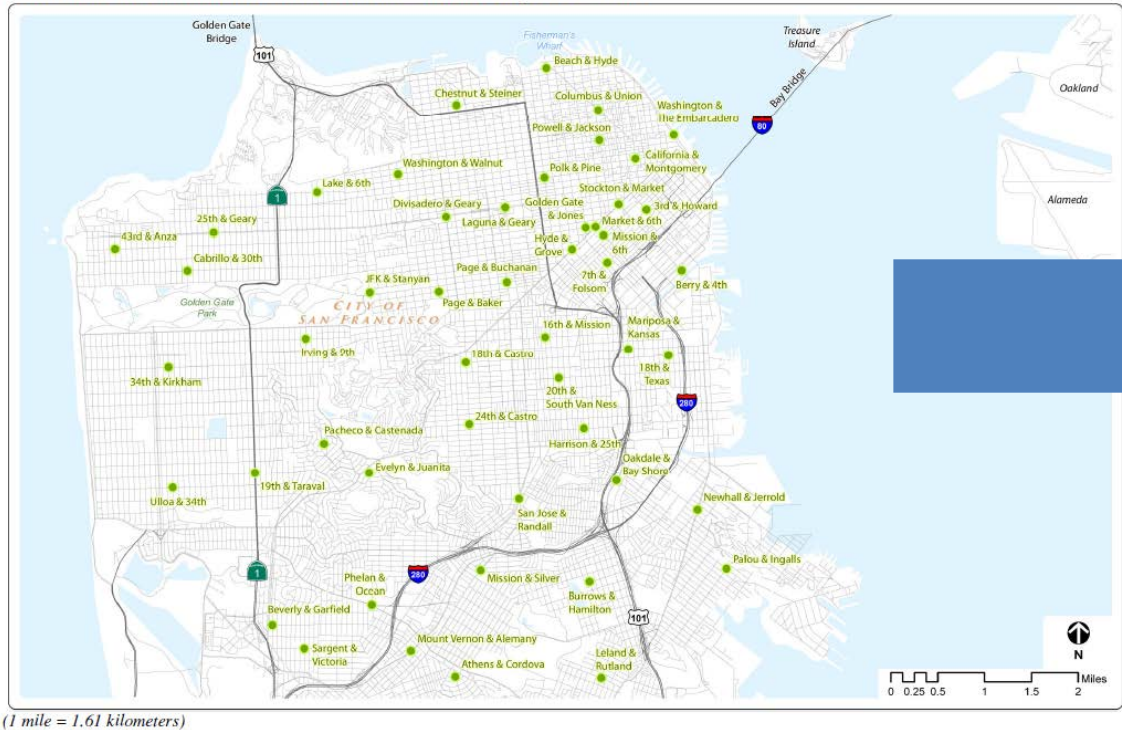


TABLE 4 Preferred San Francisco Pedestrian Volume Model

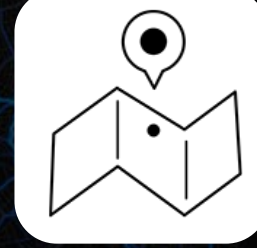
Dependent Variable = Natural Logarithm of Total Annual Pedestrian Intersection Crossings <sup>1</sup>			
	Recommended Model		
Model Variables <sup>2</sup>	Coefficient	t-value	p-value
Total households within 1/4 mile (10,000s)	1.81	2.12	0.040
Total employment within 1/4 mile (100,000s)	2.43	2.22	0.032
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
<b>Overall Model</b>			
Sample Size (N)		50	
Adjusted R <sup>2</sup> -Value		0.804	
F-Value (Test value)		34.4 ( <i>p</i> < 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

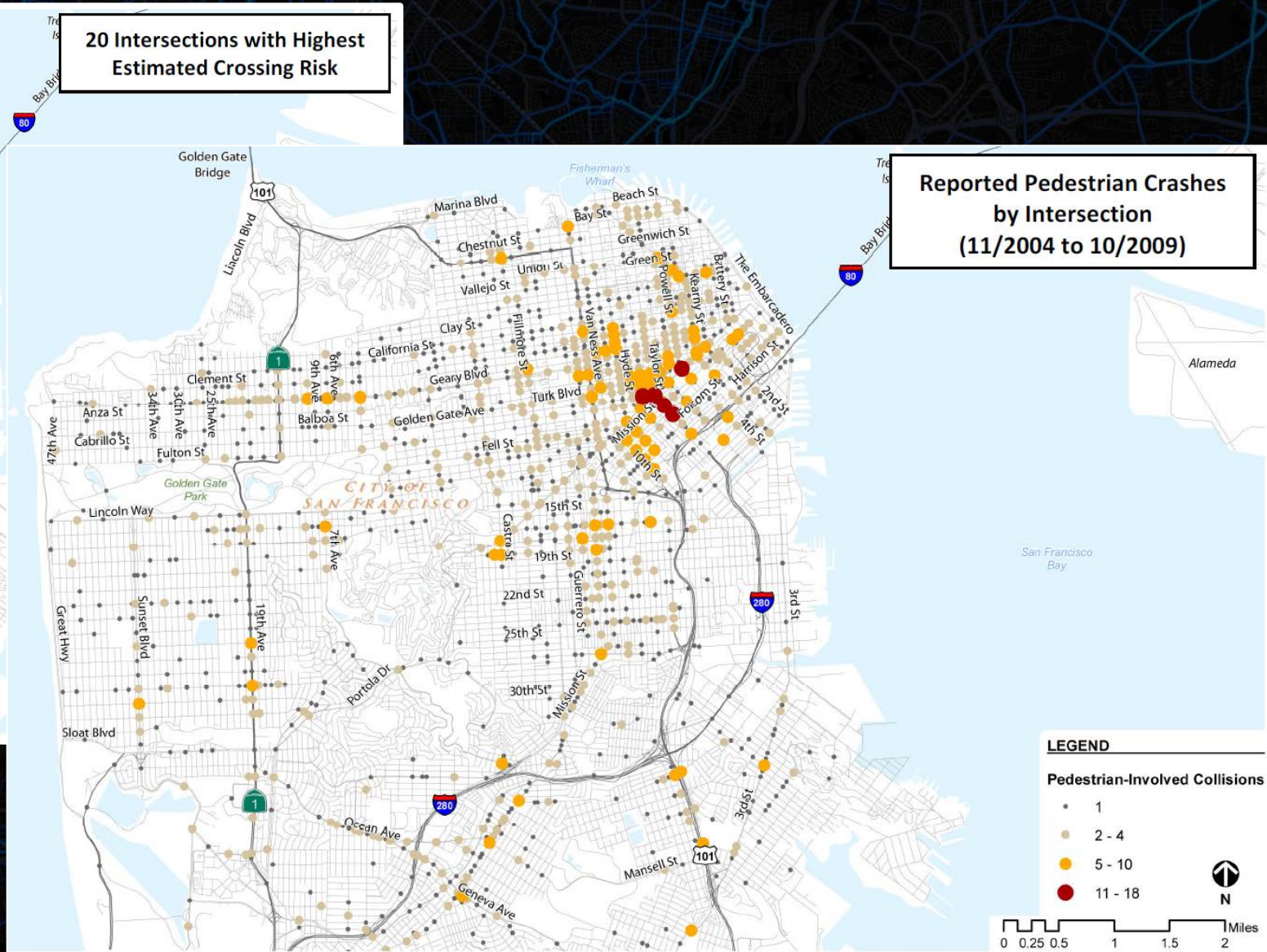
# 4 | Estimating Crossing Risk



+

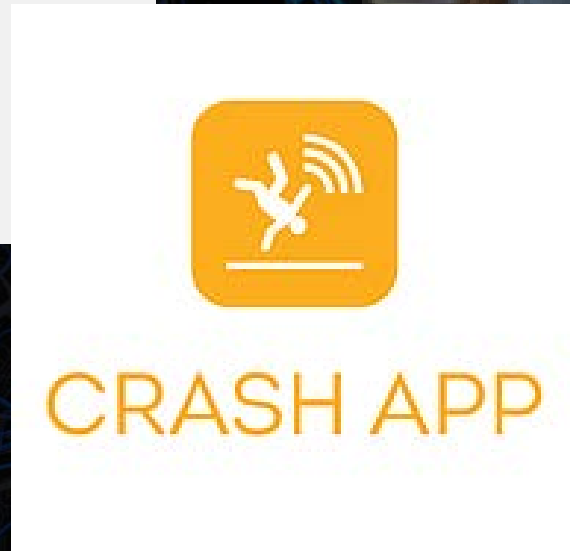


+

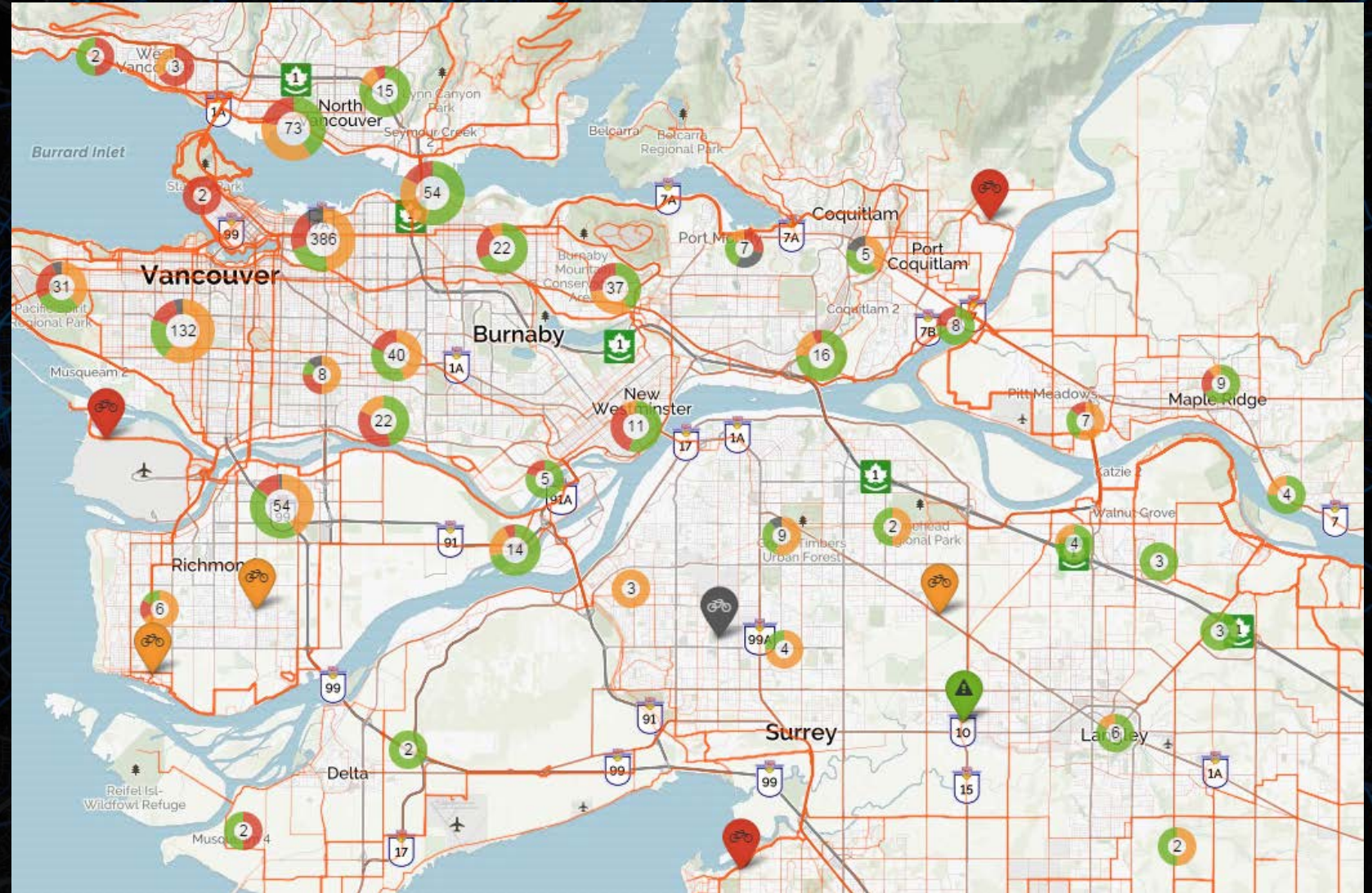


Schneider, Henry, Mitman, Stonehill and Koehler

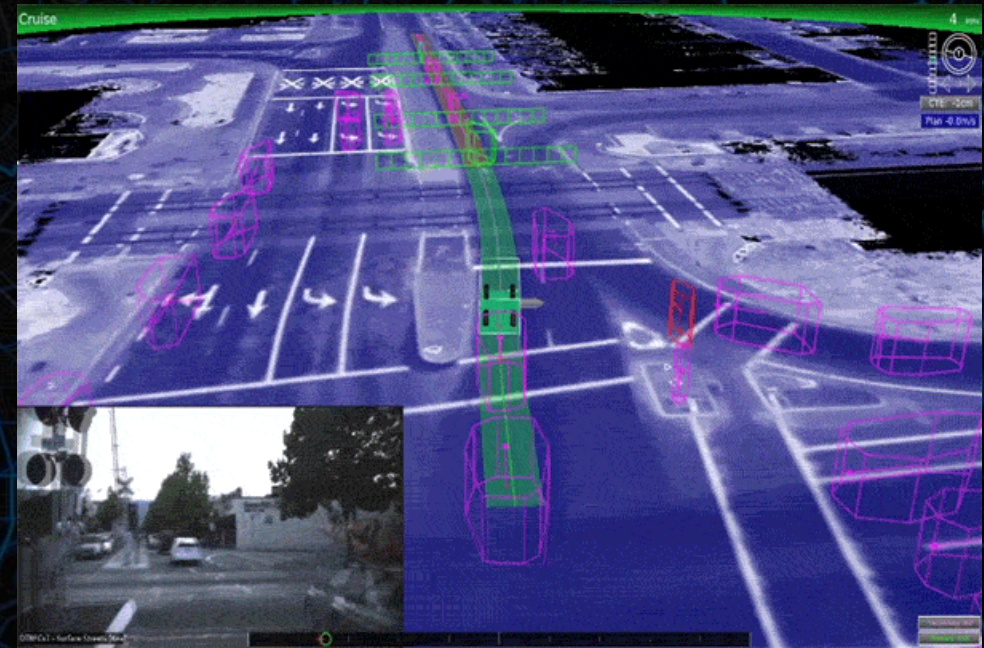
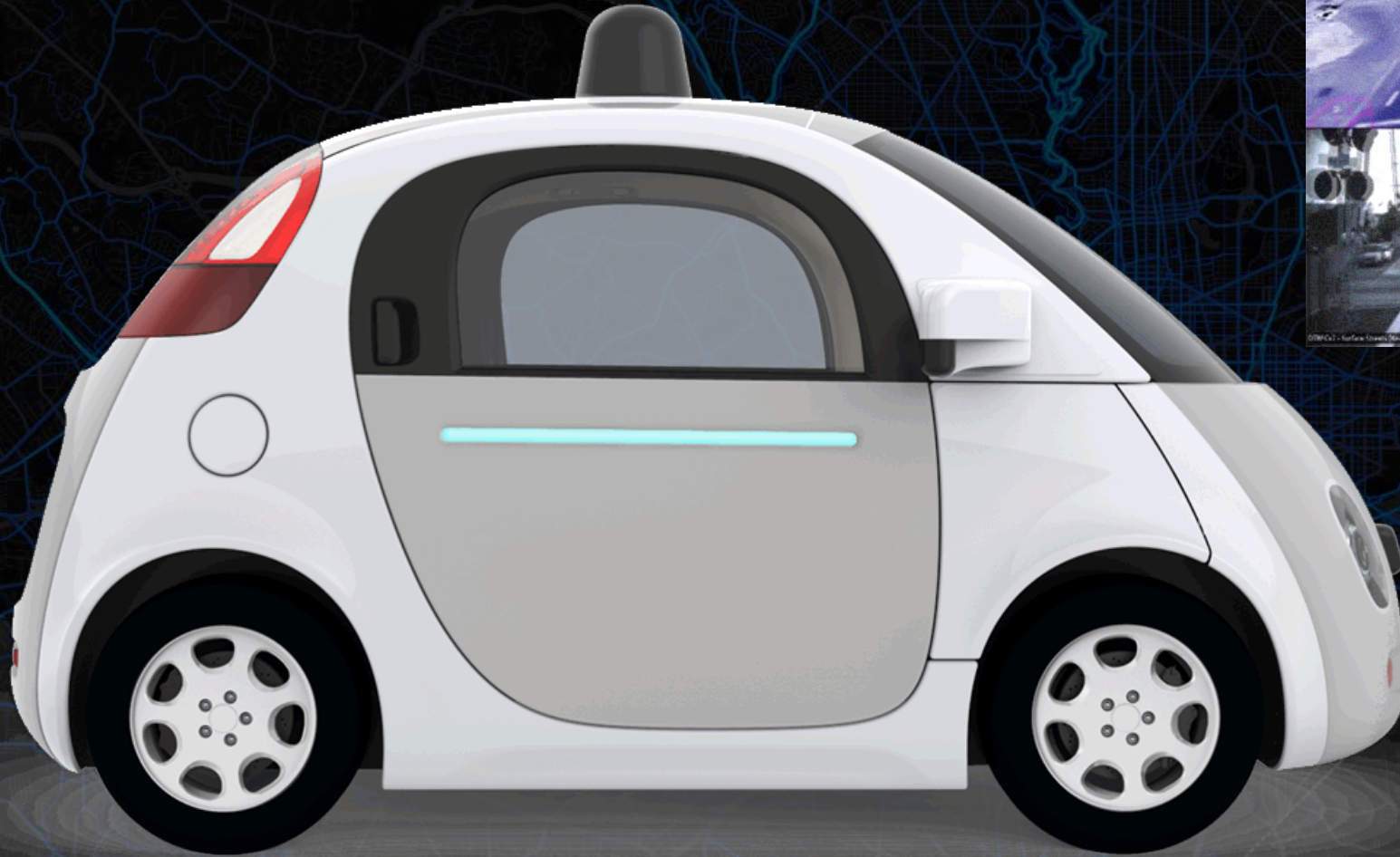
# 5 | More Collision and Near-Miss Data?



# 5 | More Collision and Near-Miss Data?



# 5 | More Infrastructure Data



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

## 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



C U E B I Q



# 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

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

Identify  
Countermeasure

Assess  
Collision  
Profile

# 6 | Improving Reactive Analysis

Easier self-reporting

More comprehensive professional reporting

# Collisions (Injuries, Fatalities)

Exposure

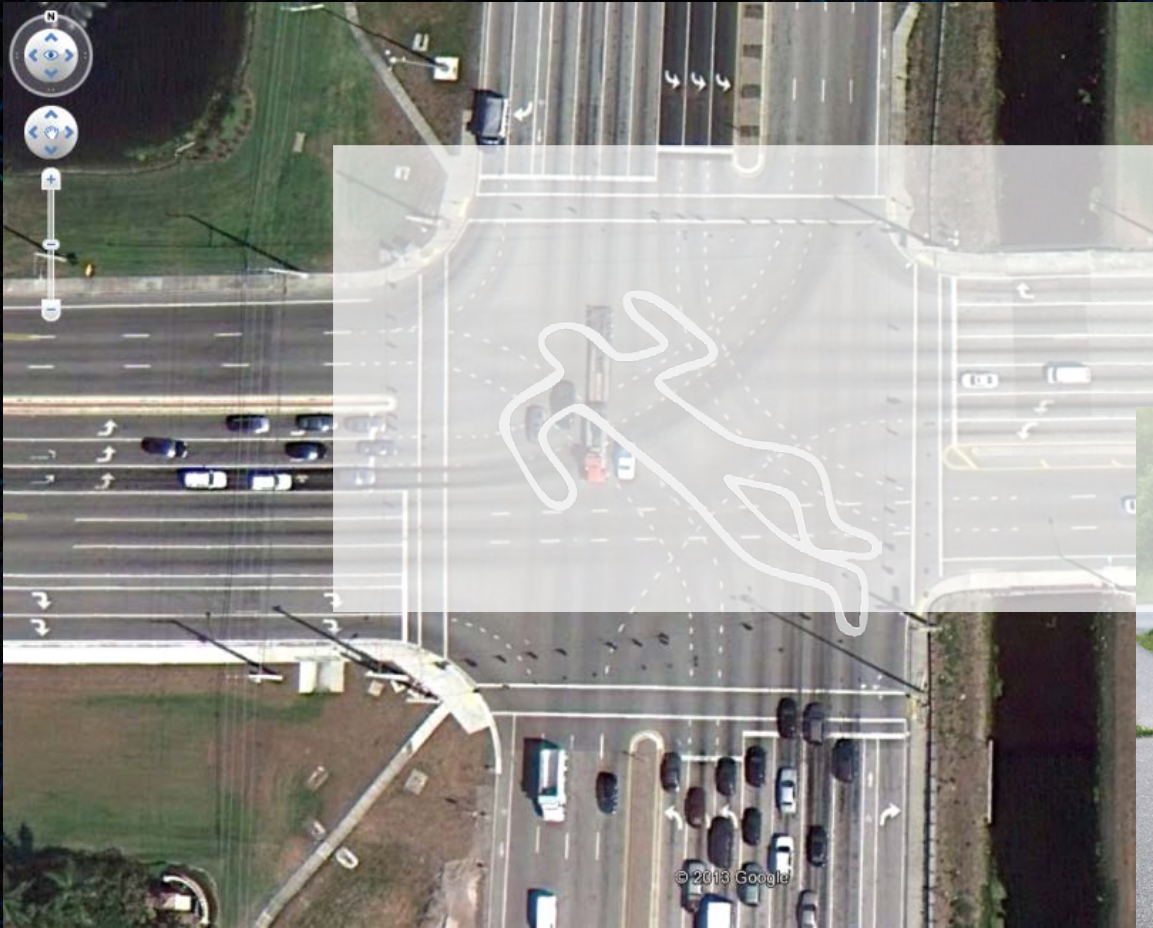
Better trip data

<sup>\*Better!</sup>  
= Rate!

## 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

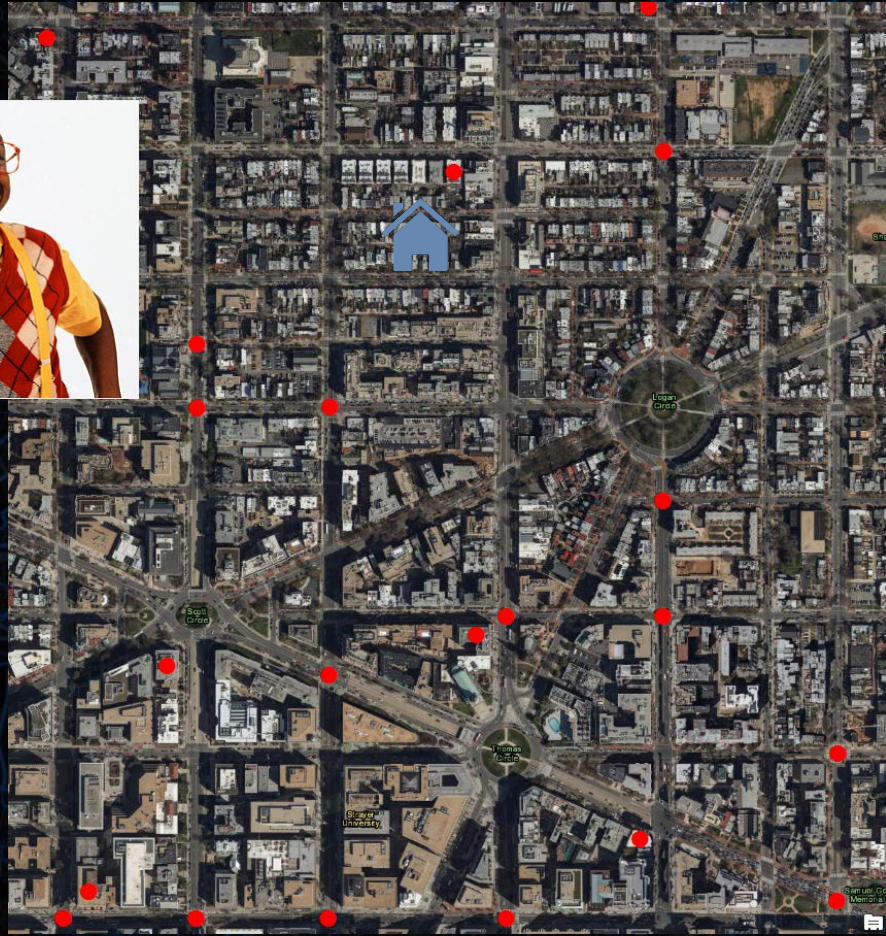
per Employee

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

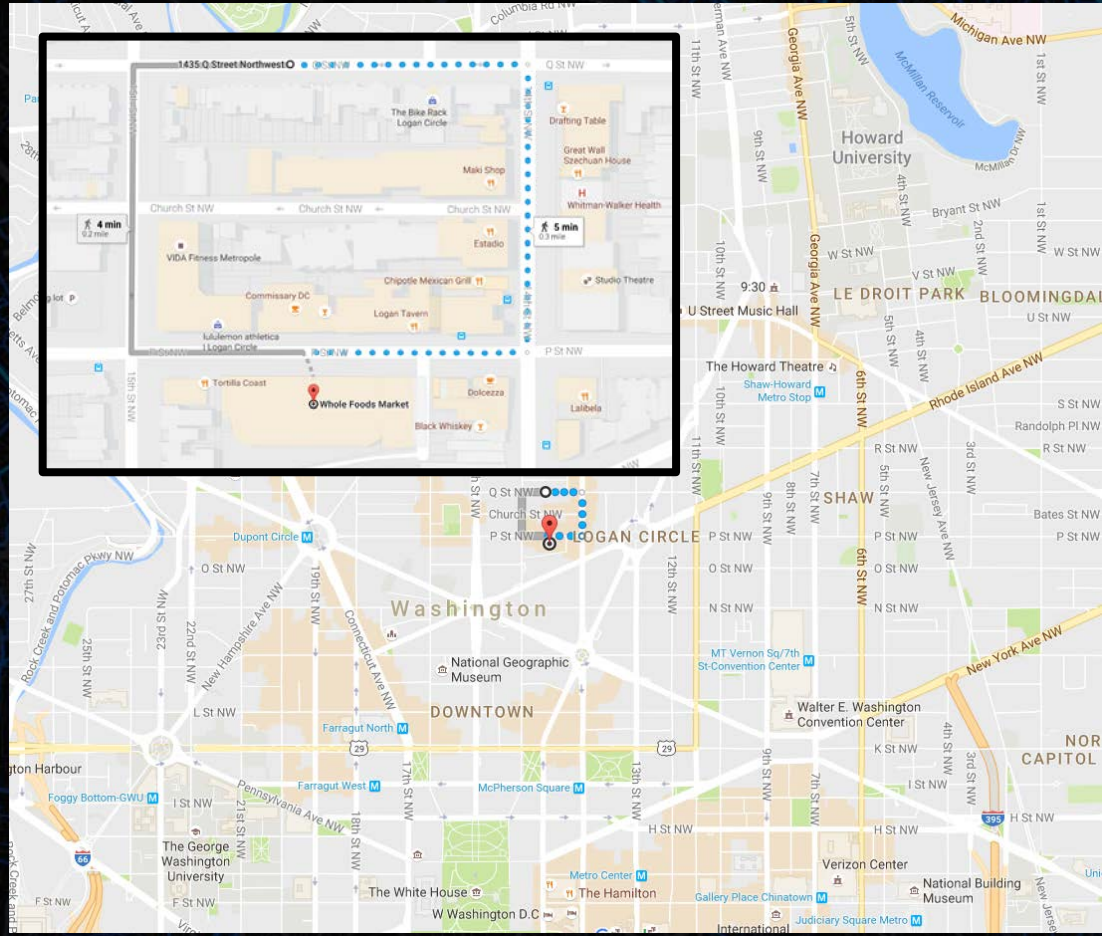


Spring Valley

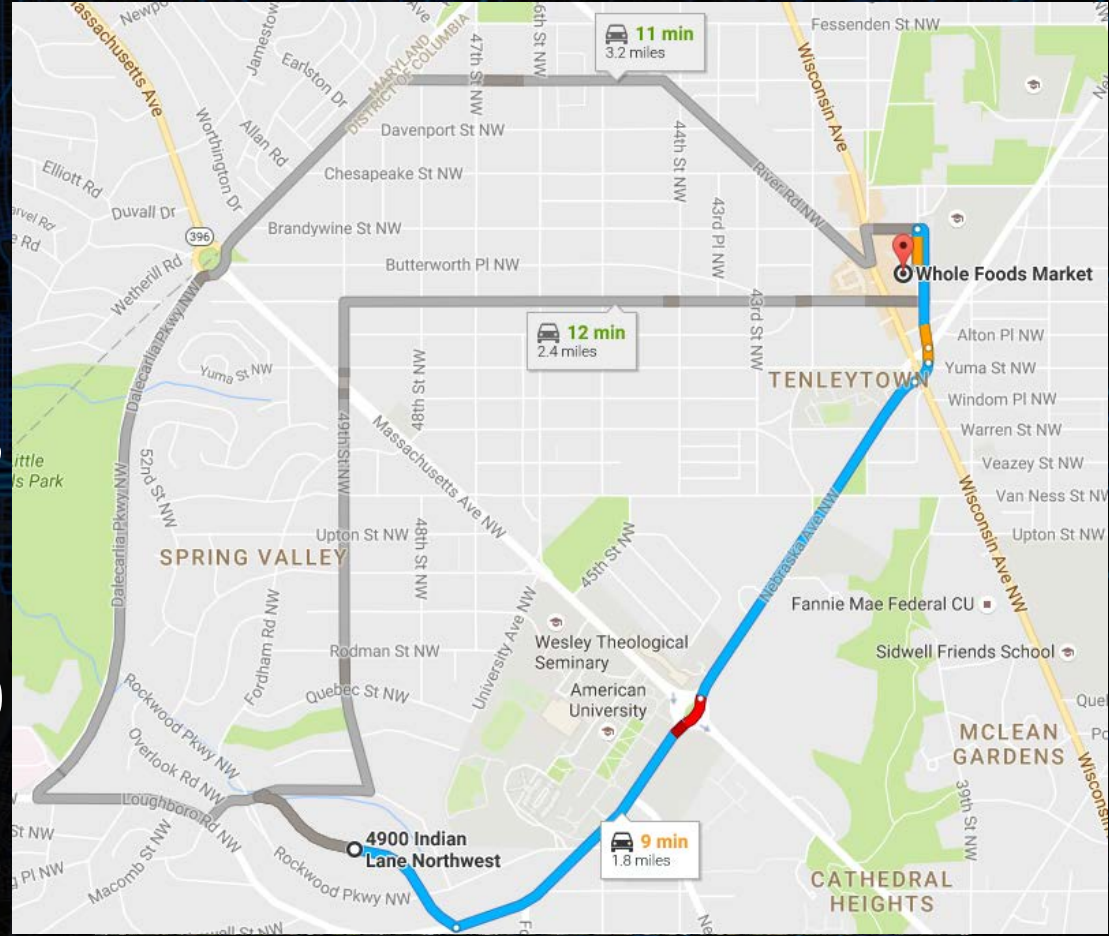
● Injury or Fatality Collision Involving Pedestrian

# 6 | Grocery trip

Logan Circle



Spring Valley



# 6 | Grocery trip

Logan Circle

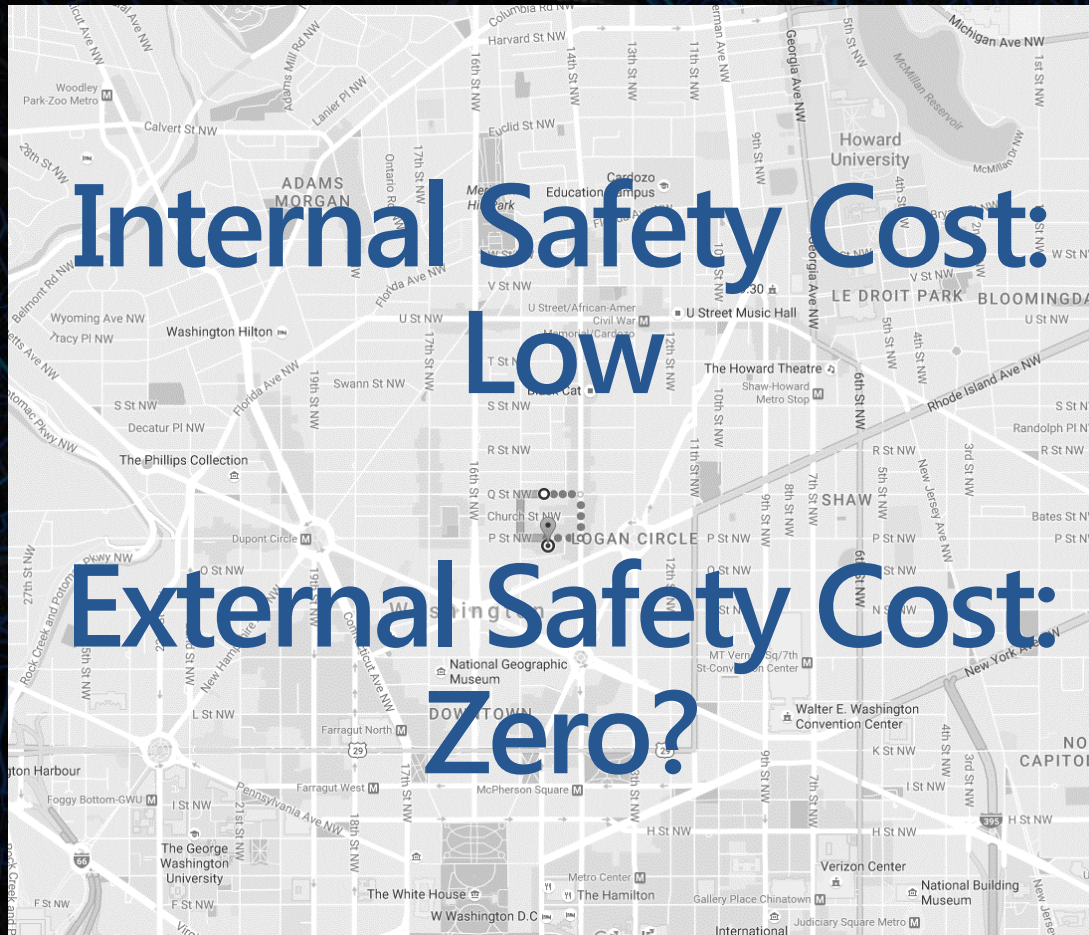


Spring Valley

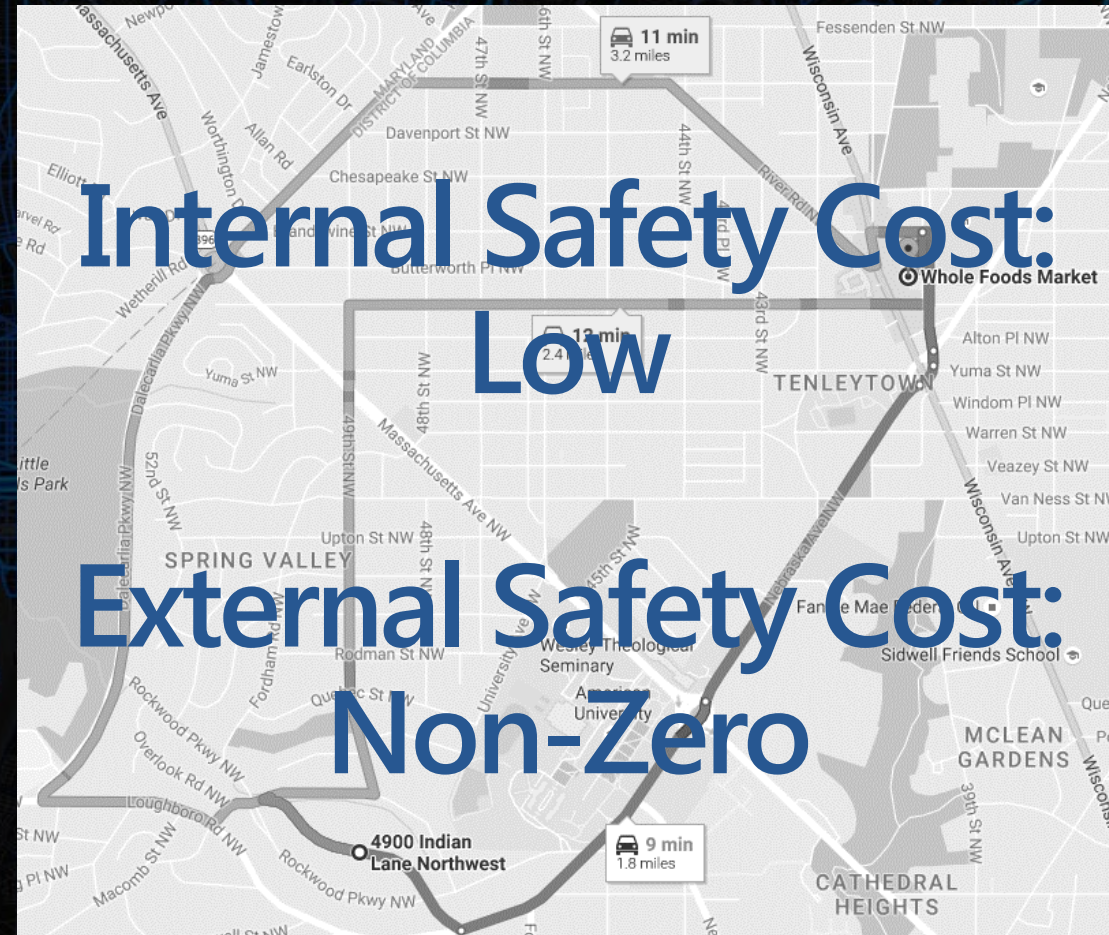


# 6 | Grocery trip safety costs (risks)

Logan Circle



Spring Valley



# 6 | Grocery trip + commute trip + lunch trip + ...



Total Safety Cost: ?

Logan Circle



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