# Introduction to Crowdsourcing in Research

DLab Workshop, UC Berkeley

Kate Beck
Program Lead, SafeTREC

Jill Cooper
Co-Director, SafeTREC

# Agenda

- Introduction to crowdsourcing
- Crowdsourcing uses
- Benefits and concerns
- Case study: crowdsourcing in transportation safety
- Discussion and crowdsourcing activities

# Crowdsourcing

- Uses collective knowledge to meet organizational or research-oriented goals
- A bottom-up approach to meet top-down goals
- Involves mutually beneficial outcomes
- Used to gather information, solve problems, generate and prioritize ideas, and complete tasks
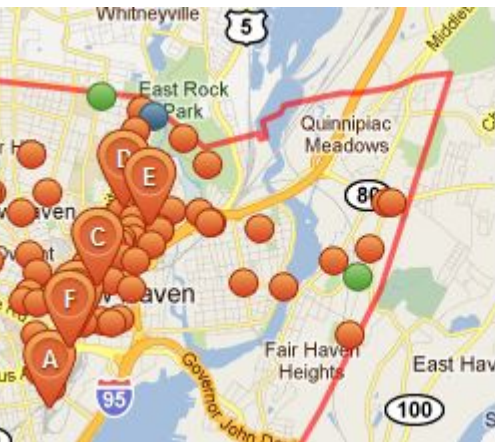
# Crowdsourcing vs. Citizen Science

- Citizen Science: members of the public actively participate in scientific processes, ex. developing research questions, collecting and analyzing data, interpreting results, etc.

# Crowdsourcing Uses

1) Information Gathering
   - SeeClickFix
   - Ushahidi
   - online survey platforms

# Crowdsourcing Uses

2) Empirical Problem Solving
   - Innocentive
   - GitHub

# Crowdsourcing Uses

3) Idea generation, prioritization and decision-making
   - MTC's Transformational Projects
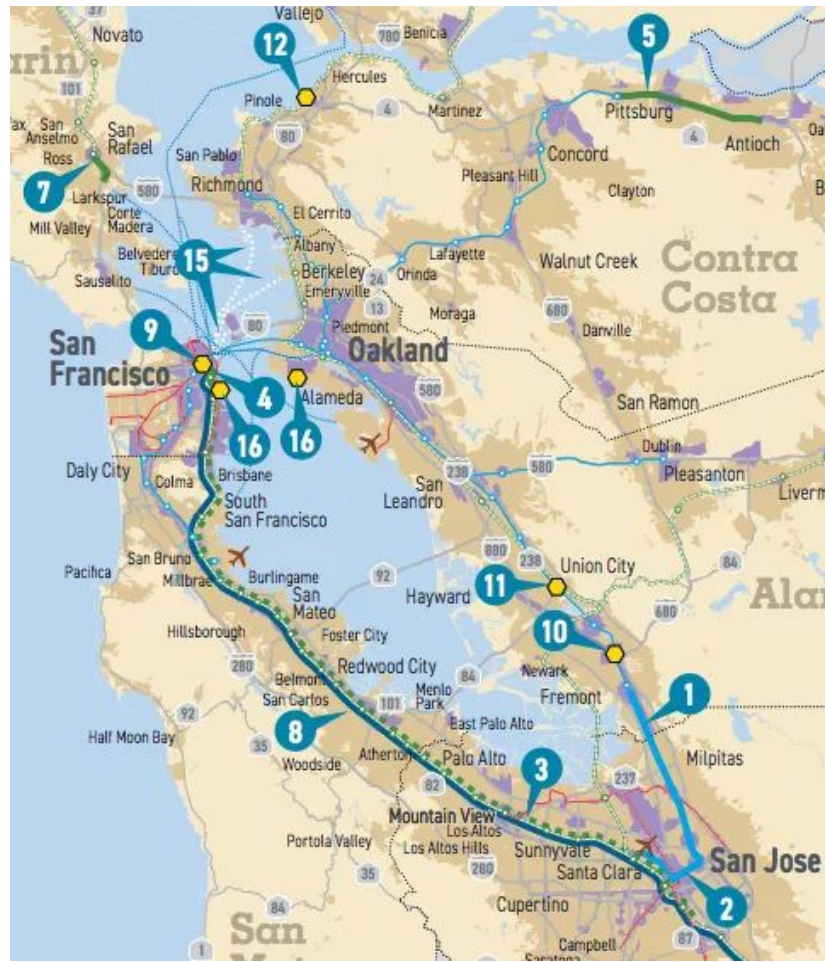
# Crowdsourcing Uses

4) Tasking
- Amazon's Mechanical Turk
- Zooniverse

# Crowdsourcing Issues

- Motivation
- Representativeness of participants
- Privacy and legality
- Misuse of the platform
- Critical mass
- Ethical issues, "crowdsploitation"

# Planning a Crowdsourcing Project

- What issue are you trying to address?
- Who can help you solve this problem? How will participants this group benefit from helping?
- What are the best ways for participants to be involved?
- How can you reach participants?

# Crowdsourcing Tips

- Use a crowdsourced platform or dataset that already exists
- understand potential participants' motivations to (and not to) participate
- Recognize and reduce barriers to participation
- Communicate mutual benefits

# Analysis - Perceptions of Crowdsourcing Technologies

| Internal Factors Impacting Crowdsourcing Success | | External Factors Impacting Crowdsourcing Success | |
|---|---|---|---|
| **Strengths** | **Weaknesses** | **Opportunities** | **Threats** |
| <ul><li>Crowdsourcing can collect information that is comparable to existing data sources</li><li>Crowdsourcing tools and data should be easy for participants to use</li><li>Crowdsourcing should provide information that builds trust between researchers, external partners and participants</li></ul> | <ul><li>The data collected may not be accurate or verifiable</li><li>Crowdsourcing may not protect the participant privacy</li><li>Data ownership may not be clear</li></ul> | <ul><li>Crowdsourcing can encourage accountability</li><li>Crowdsourcing can support existing data collection and engagement efforts</li></ul> | <ul><li>The potential participants may face barriers to accessing crowdsourcing tools</li><li>Researchers or external partners may not acknowledge the crowdsourced information</li></ul> |

# Case Study: Crowdsourcing in Transportation

# Street Story

- Street Story helps community groups and agencies collect and understand information that is important for transportation safety but is difficult to gather
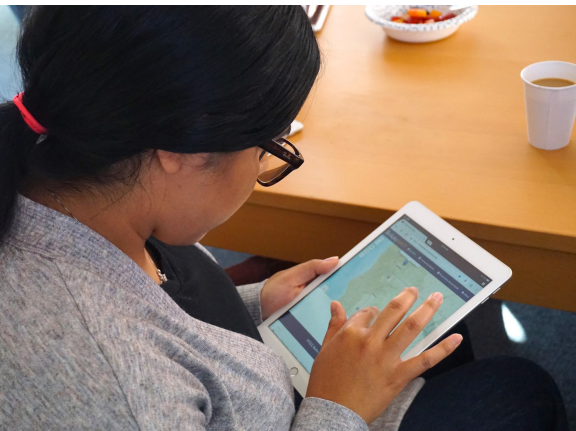
# Police Reported Collisions
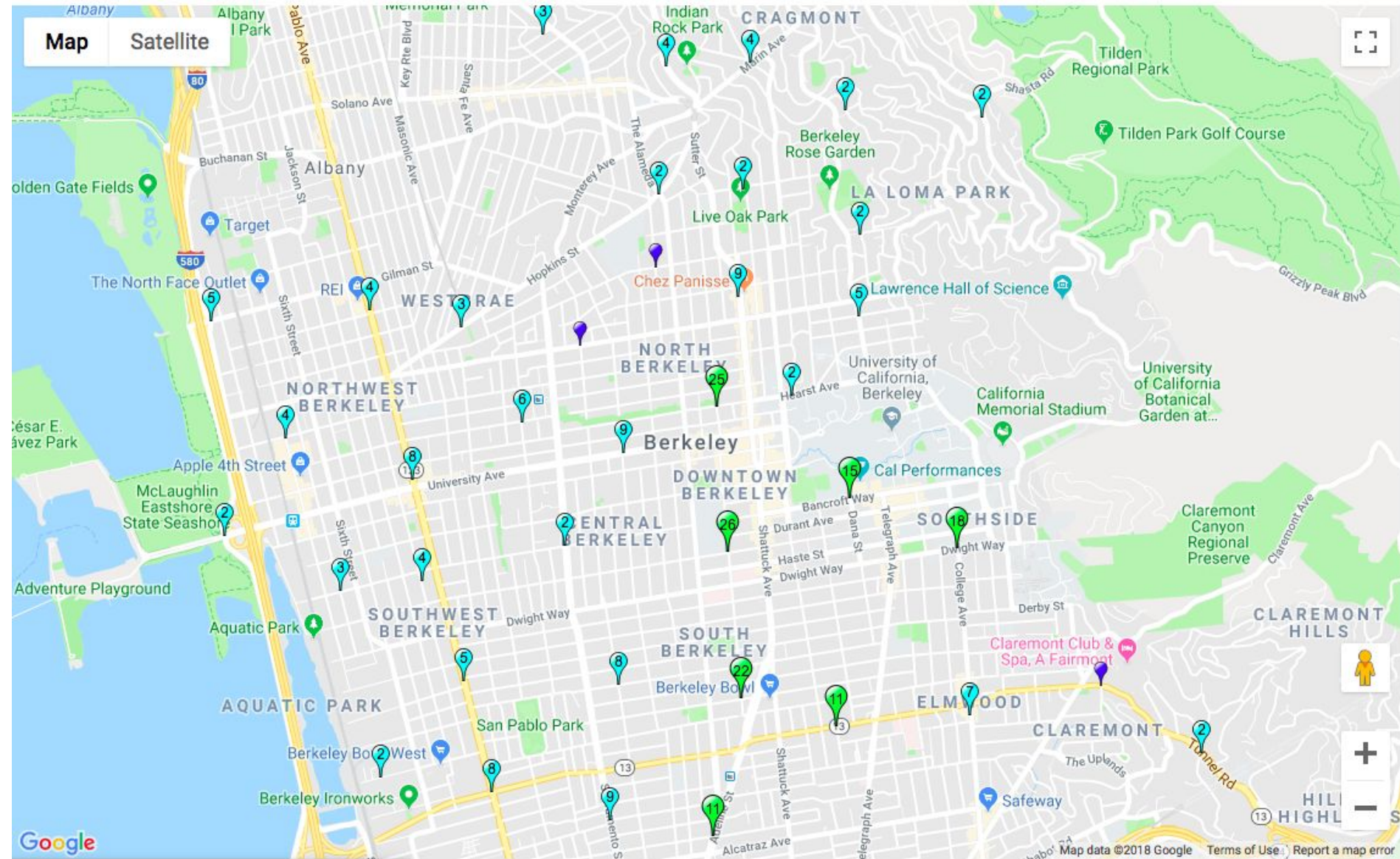
# Local Knowledge

# Designing the Street Story Program

- Interviews, focus groups, pilot testing

| Key Partners | Key Activities | Value to Participants | Participant Relationships | Participant Segments |
|---|---|---|---|---|
| | **Key Resources** | | | |

| Cost | Revenue Streams |
|---|---|
| | |

# Street Story Program Model
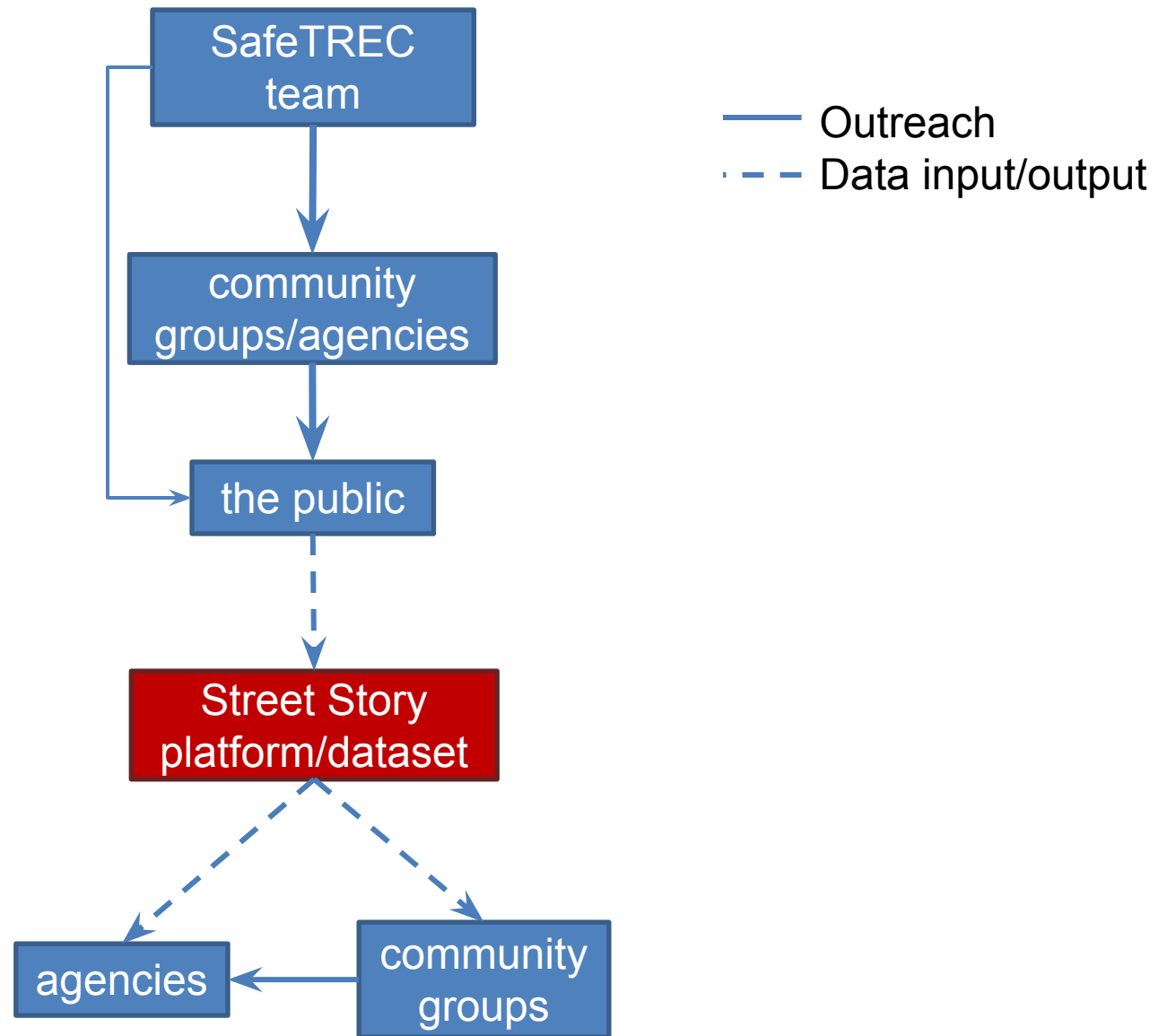
# How to Report

https://streetstory.berkeley.edu/
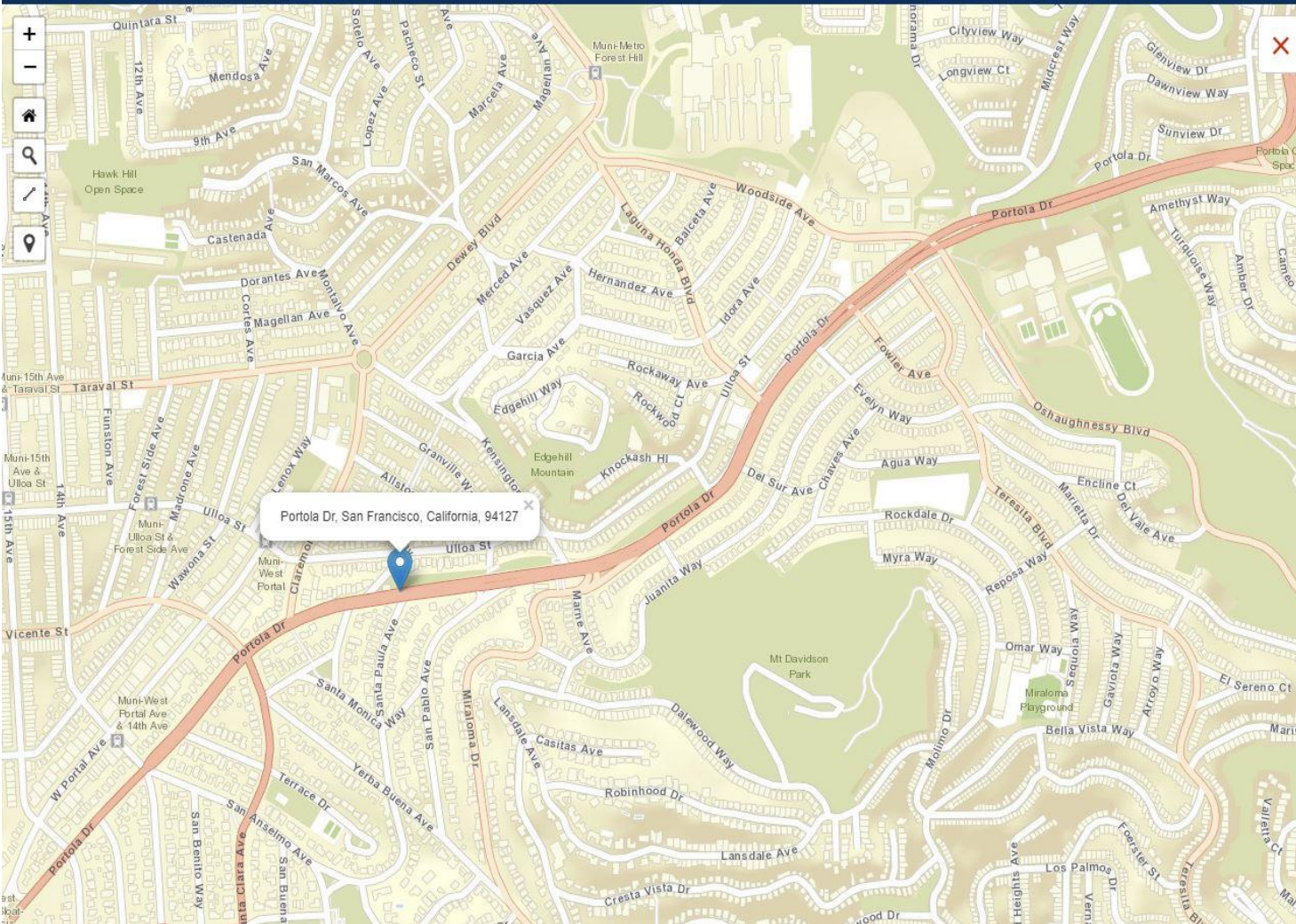
# Street Story Quantitative Data

**Report Map**



**Report Type**
- 🟥 Collision
- 🟧 Near-Miss
- 🟨 Generally Unsafe
- 🟩 Safe

## Travel Modes Summary



Legend:
- 🔵 Biking
- 🟢 Walking
- 🟤 Scootering
- 🟣 Using a mobility device
- 🟠 Riding in a vehicle
- 🔵 Multiple Modes
- ⚪ Other
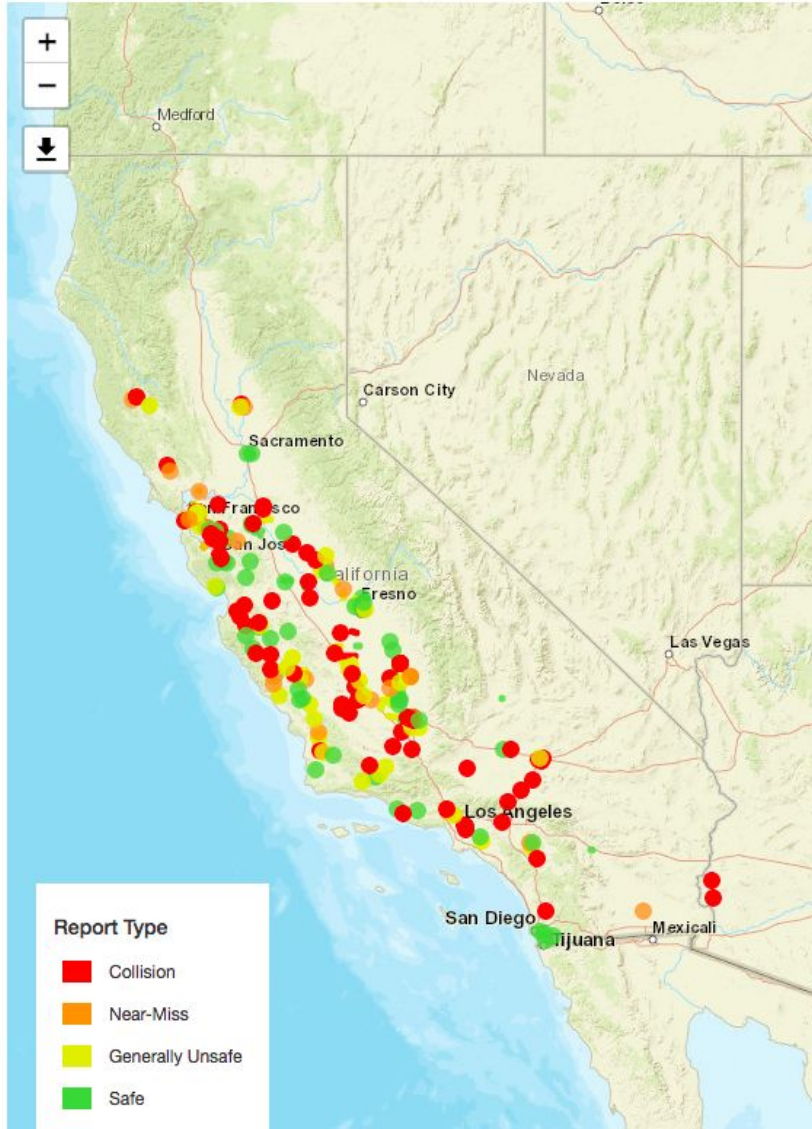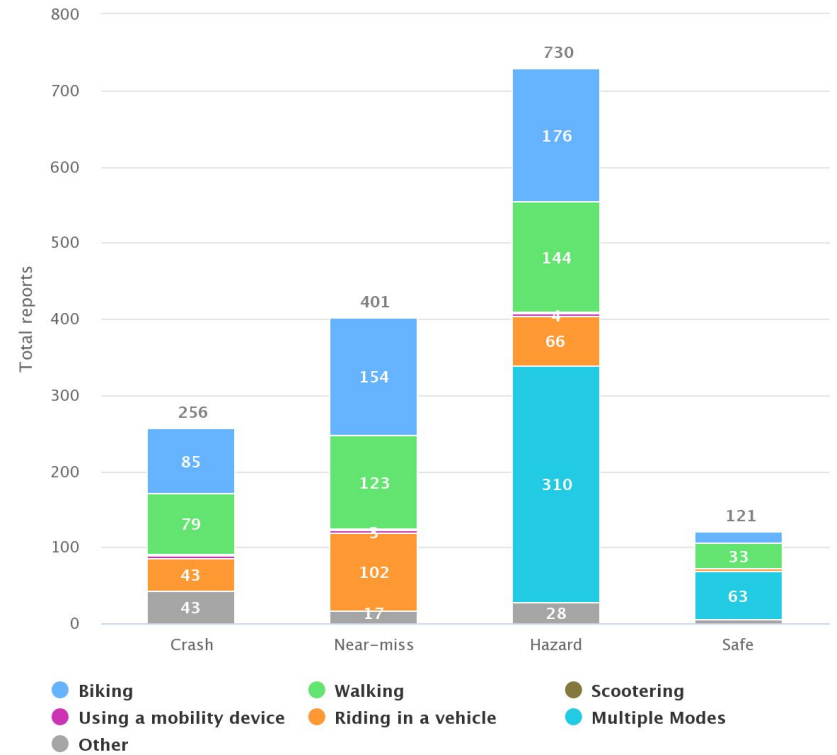
| Report Type | Count | Percent |
|---|---|---|
| Crash | 256 | 17 % |
| Near-miss | 401 | 27 % |
| Hazard | 730 | 48 % |
| Safe | 121 | 8 % |
| Total | 1508 | |

# Street Story Quantitative Data

## Demographic Information

| Gender | Count | Percent |
|---|---|---|
| Female | 424 | 39 % |
| Male | 449 | 41 % |
| Non Binary | 10 | 1 % |
| No Response | 214 | 20 % |

| Age | Count | Percent |
|---|---|---|
| 18-25 | 62 | 6 % |
| 26-64 | 653 | 60 % |
| 65 years or older | 84 | 8 % |
| No Response | 298 | 27 % |

| Resident in area | Count | Percent |
|---|---|---|
| Yes | 725 | 66 % |
| No | 134 | 12 % |
| No Response | 238 | 22 % |

| Participation in community meeting | Count | Percent |
|---|---|---|
| Often (more than 5 times in the last 5 years) | 181 | 16 % |
| Sometimes (1-5 times in the last 5 years) | 309 | 28 % |
| Never | 401 | 37 % |
| No Response | 206 | 19 % |

| Have disability | Count | Percent |
|---|---|---|
| Yes | 62 | 6 % |
| No | 689 | 63 % |
| No Response | 346 | 32 % |

| First time respondent | Count | Percent |
|---|---|---|
| Yes | 858 | 78 % |
| No | 239 | 22 % |
| No Response | 0 | 0 % |

| Race/Ethnicity | Count | Percent |
|---|---|---|
| American Indian/Alaska Native | 16 | 1 % |
| Asian | 59 | 5 % |
| Black/African American | 32 | 3 % |
| Latino/Hispanic | 192 | 16 % |
| Native Hawaiian/Pacific Islander | 15 | 1 % |
| White | 594 | 51 % |
| Other | 65 | 6 % |
| No Response | 196 | 17 % |

* Race/Ethinity will not add up to 100% because of multiple choices.

Date Generated: 9/11/2019

# Street Story Qualitative Data

"This intersection experiences regular collisions. On several occasions a car has run up onto the sidewalk. There have been numerous close calls with pedestrians. Cars frequently run the red light at this intersection."

**Codes**
Persistent Issue
Collision Type
Driver Behavior

# Addressing Crowdsourcing Issues

- Motivation and mutual benefits
- Participant representativeness
- Privacy issues
- Misuse of the platform

# Lessons Learned

1. Recognize when to collect data, and when to use existing sources
2. Understand participants' motivation
3. Speak in a language that participants understand
4. Recognize the time it takes to build and maintain relationships

# Resources

- UC Berkeley Coalition for Education & Outreach
- FieldScope - platform for hosting citizen science project data http://www.fieldscope.org
- SciStarter - Platform to find citizen science projects and audiences https://scistarter.org
- Citizen science: crowdsourcing for research, Catherine Lichten et al., University of Cambridge (2018)
- A Methodological Framework for Crowdsourcing in Research, Michael Keatinga and Robert D. Furberg, RTI International (2013)

# Sources

1) [Crowdsourcing, by Daren C. Brabham (2013)](#)
2) [Crashes on and Near College Campuses: A Comparative Analysis of Pedestrian and Bicyclist Safety](#). Loukaitou-Sideris, Medury, et al. Journal of the American Planning Association Vol. 80, Iss. 3, 2014.
3) [Investigating the underreporting of pedestrian and bicycle safety crashes in and around university campuses-a crowdsourcing approach](#). Medury, Grembek, et al. Accident Analysis and Prevention, 2017.
4) [Challenges, Crowdsourcing, Citizen Science: What's the Dif?](#) Digital.gov, 2015.

Berkeley SafeTREC
SAFE TRANSPORTATION RESEARCH AND EDUCATION CENTER

# Contact Information

Kate Beck
[katembeck@berkeley.edu](mailto:katembeck@berkeley.edu)

Jill Cooper
[cooperj@berkeley.edu](mailto:cooperj@berkeley.edu)

Berkeley SafeTREC
SAFE TRANSPORTATION RESEARCH AND EDUCATION CENTER