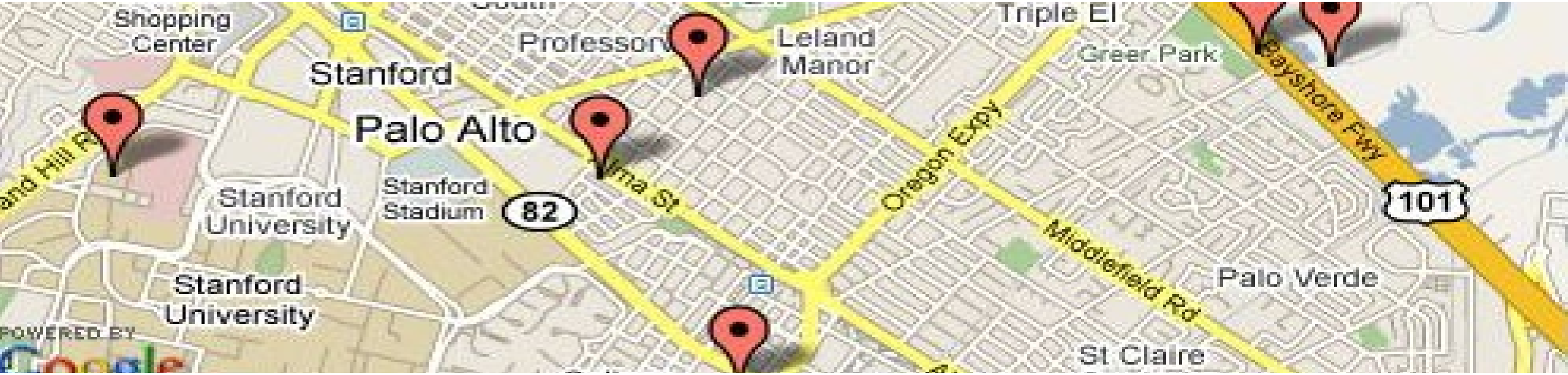
Using Google Maps to Generate Organizational Sampling Frames



Brad R. Fulton and David P. King
Indiana University

SSRC Workshop in Methods

2023.1.20

fulton@iu.edu
@BradRFulton

Sampling Frames

Sampling frame: a list of units forming a population from which a sample is taken

fulton@iu.edu
@BradRFulton

Organizational Research

fulton@iu.edu
@BradRFulton

Organizational Sampling Frames

Government Records

- U.S. Census Bureau, Bureau of Labor Statistics, Internal Revenue Service, etc.
  - Difficult to obtain
  - Not organized by industry
  - Limited to organizations…
    - with employees
    - required to file annual tax returns
    - required to register with the government

fulton@iu.edu
@BradRFulton

Organizational Sampling Frames

- Other Public Records
  - Telephone Directories, U.S. Postal Service Databases, etc.
    - Omits transient organizations
    - Compiled at state and local levels in non-standardized formats
- Proprietary Directories and Databases
  - Expensive
  - Limited to members
  - Infrequently updated

fulton@iu.edu
@BradRFulton

Organizational Sampling Frames

Hypernetwork Sampling

Generate a representative sample of organizations
via a representative sample of individuals

Business organizations, voluntary associations, and religious congregations

Time-consuming

Produces only one sample

Requires complex weighting protocols

fulton@iu.edu
@BradRFulton

Organizational Sampling Frames

Location-based/Point-of-Interest Data Platforms

- Google Maps, SafeGraph, Esri, etc.
  - Greater coverage
  - One consistent format
  - Less expensive
  - Broad range of organizations
  - More efficient
  - Can generate multiple samples

fulton@iu.edu
@BradRFulton

Google Maps

Launched in 2005

> 1 billion users every month

Google Maps Places

> 200 million places

> 25 million updates every day

Pulls data from organizations' websites, online postings, public records, and *Google My Business* registrations

Google does not provide a list of all the places on its platform

Google Maps API

fulton@iu.edu
@BradRFulton

Google Maps API Parameters

place ID: Each place (i.e. organization) is assigned a unique ID

place type: Broadly describes the organization's type (97 place types)

school, hospital, supermarket, etc.

business category: Provides more specific details (4,050 business categories)

charter school, animal hospital, Asian grocery store, etc.

fulton@iu.edu
@BradRFulton

Google Maps API Queries

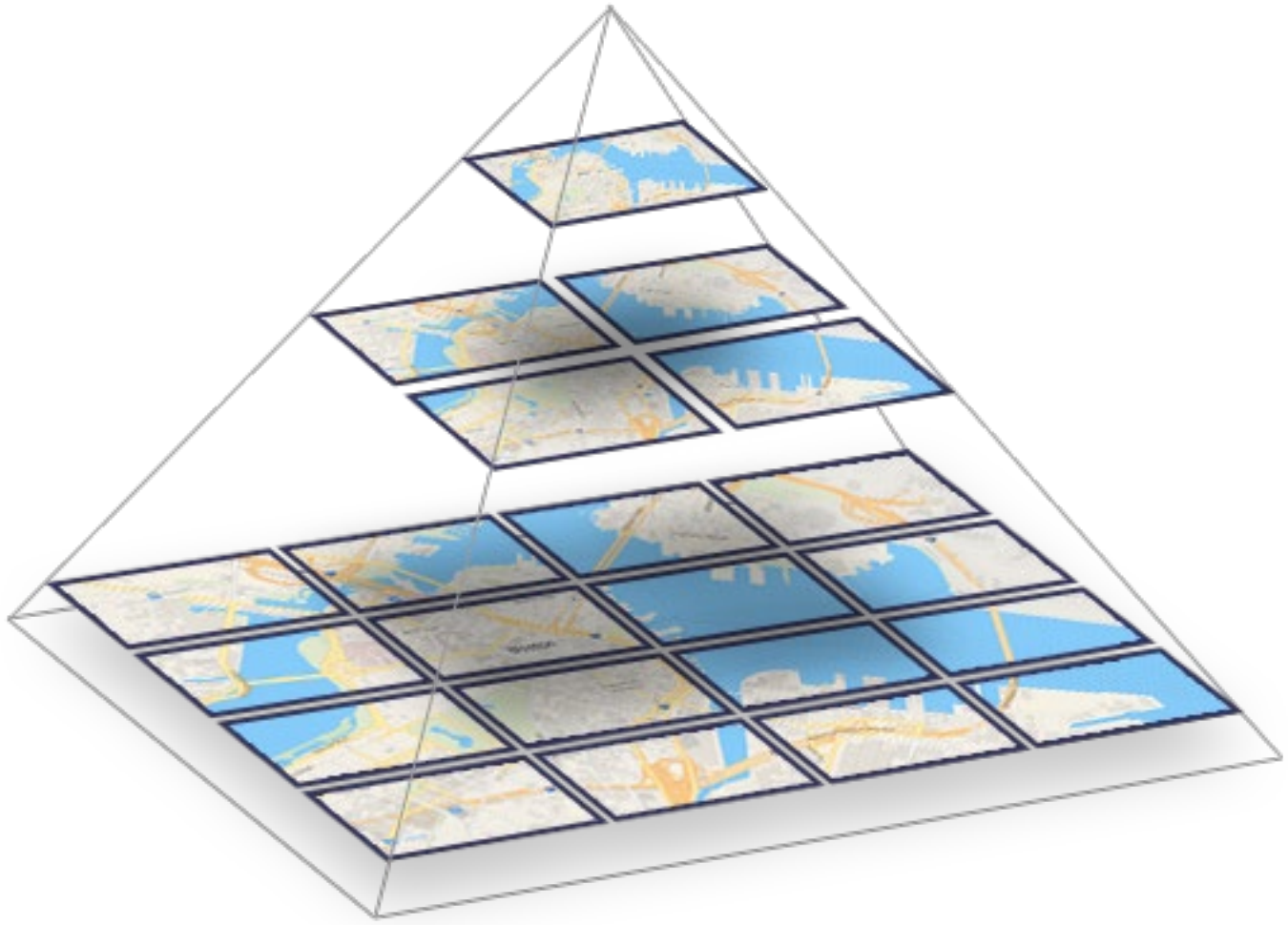
Place Search request: Returns a list of places (specifically *place ID*s) based on a specific search request and corresponding search criteria

Nearby Search request: Allows searches for a particular type of place within a specified geographic area

fulton@iu.edu
@BradRFulton

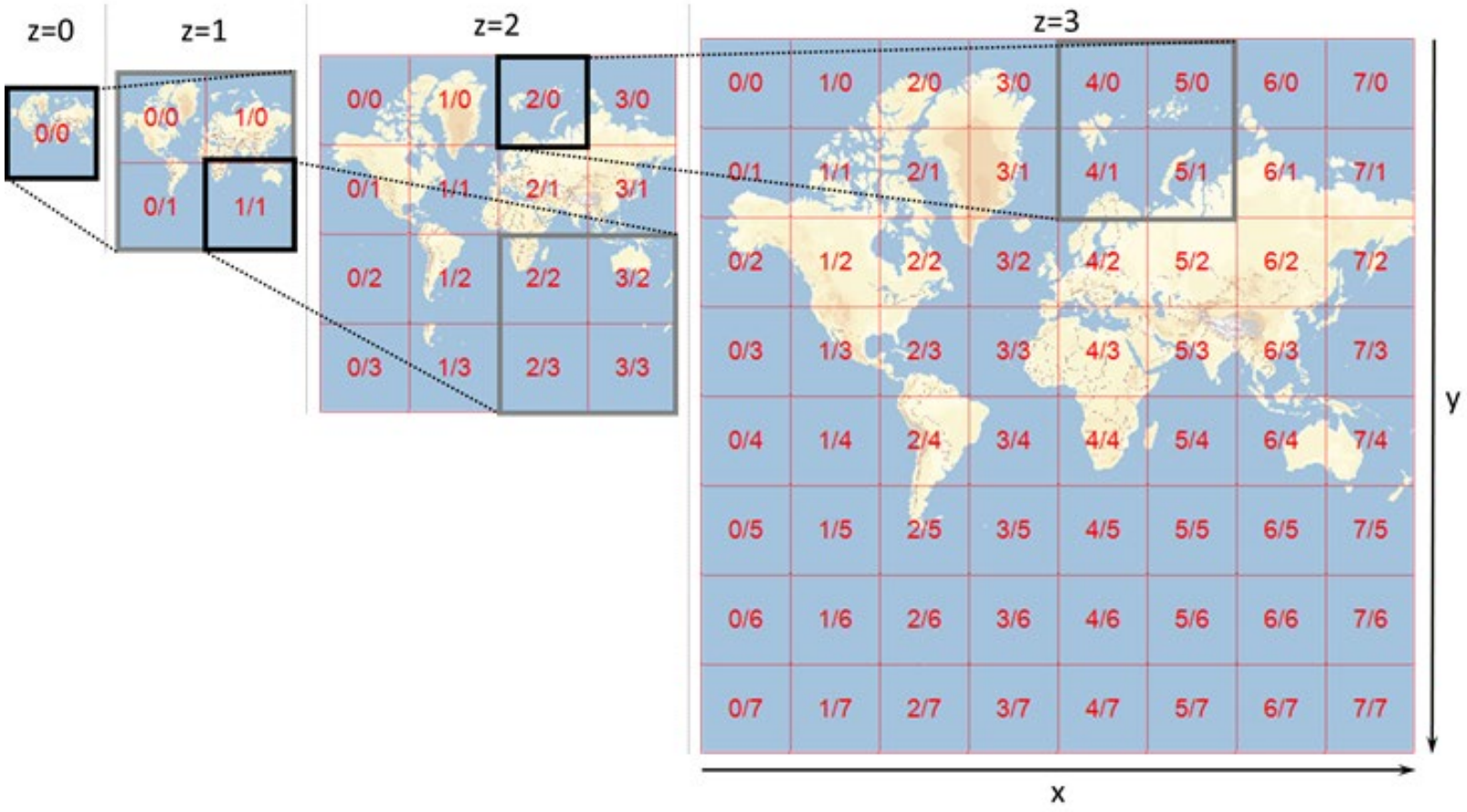
Specifying the Geographic Search Area

Raster Tile Mapping Schema



fulton@iu.edu
@BradRFulton

Raster Tile Mapping Schema



Zoom Level	# of Tiles	Tile Size (m)	Area Represented
0	1	~40M	Whole World
1	4	~20M	1/4 Sphere
2	16	~10M	Continent
3	64	~5M	Lg. Country
⋮	⋮	⋮	⋮
15	~1B	1,223	Square Mile
18	~68B	153	City Block
20	~1T	38	Building

Zoom Level 0

Whole World

fulton@iu.edu
@BradRFulton

Zoom Level 3

United States

fulton@iu.edu
@BradRFulton

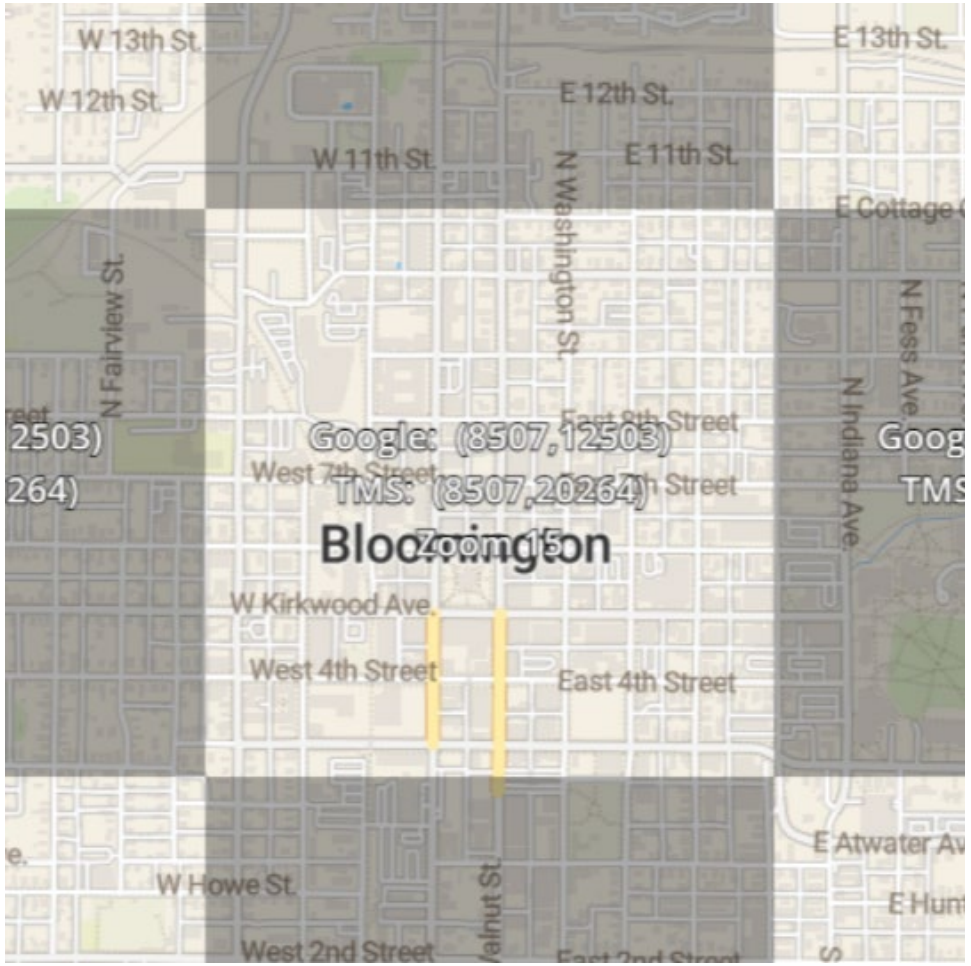
Zoom Level 6

Indiana

fulton@iu.edu
@BradRFulton

Zoom Level 15

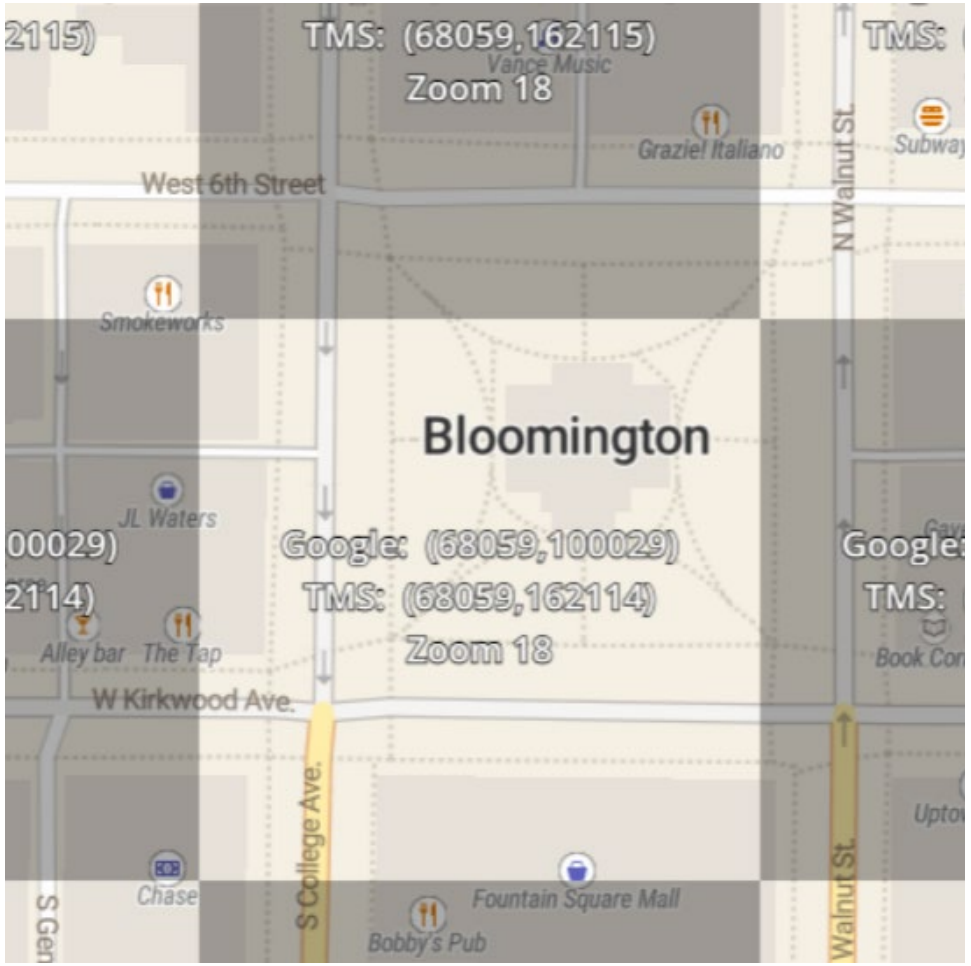
1 square mile



fulton@iu.edu
@BradRFulton

Zoom Level 18

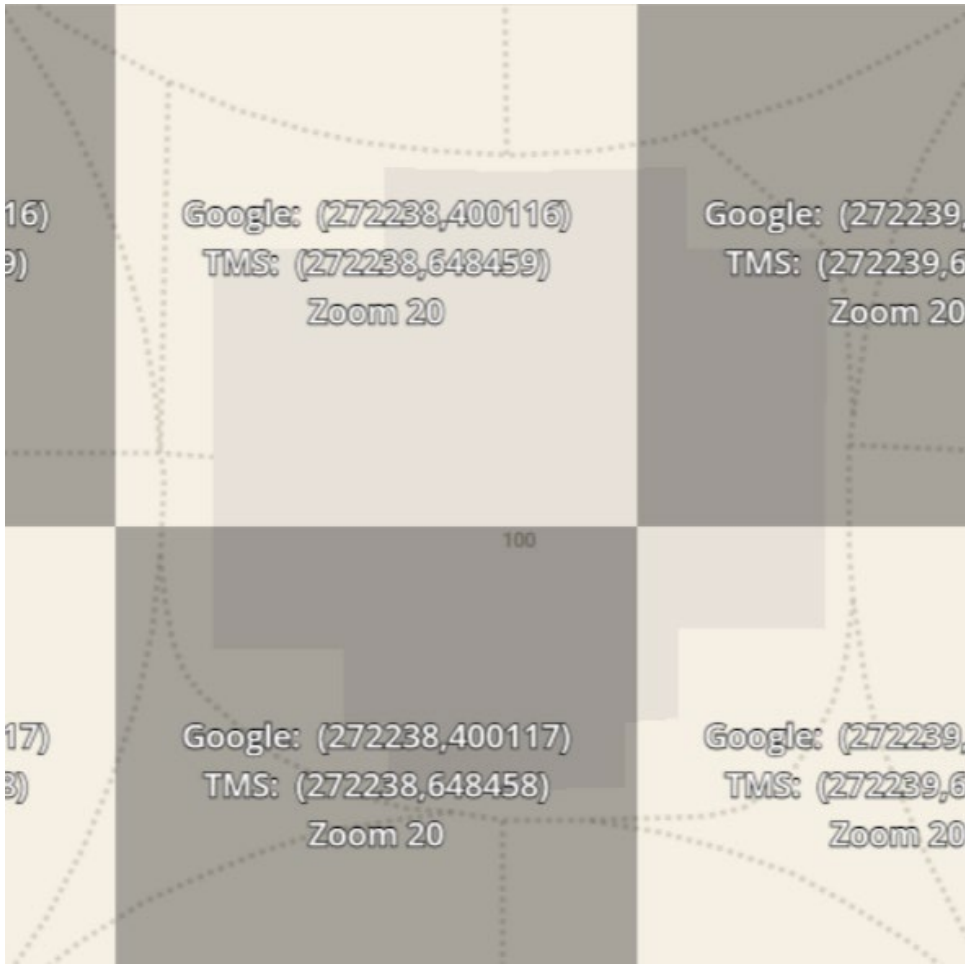
1 city block



fulton@iu.edu
@BradRFulton

Zoom Level 20

1 building



fulton@iu.edu
@BradRFulton

Searching for Organizations

Major Constraint: Search requests will return no more than 60 organizations with each request

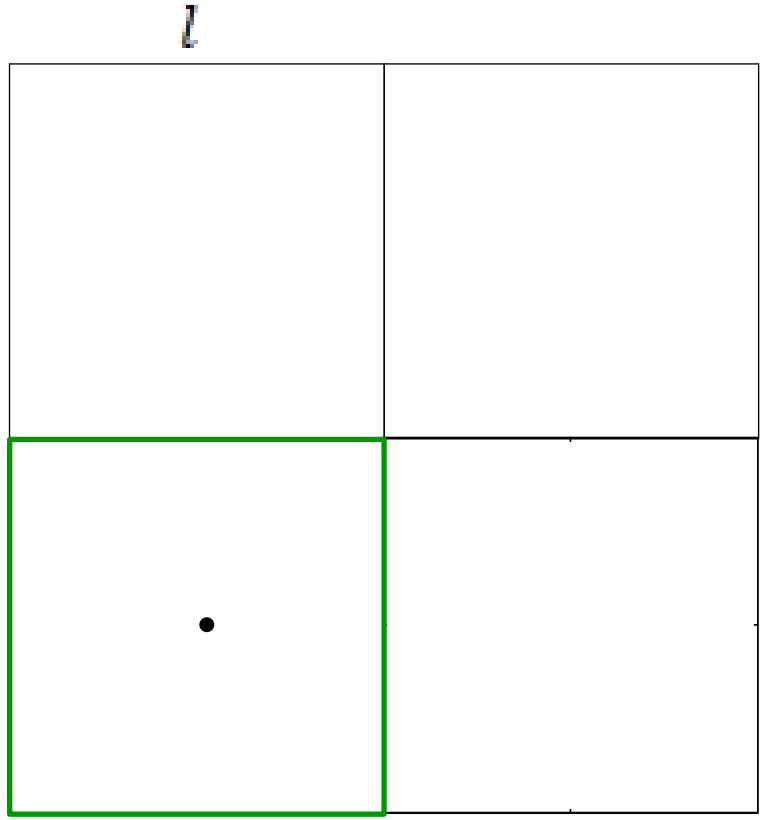
Specify the *place type*

Set the boundary of the search area

Start at zoom level 15 (~1 square mile)
[4 square mile area = 4 tiles]

Select the first tile

Set the query point at the center of the tile



fulton@iu.edu
@BradRFulton

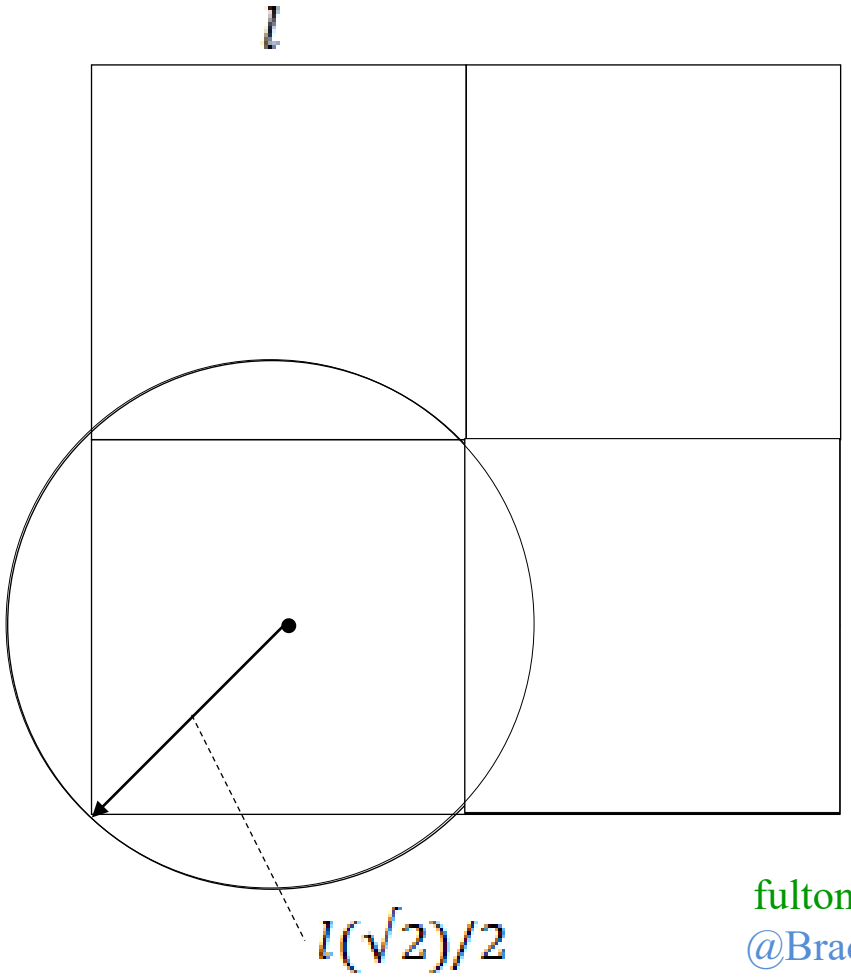
Searching for Organizations

The API's search query area is a circular
(with a user-defined radius)

Set the radius to $l(\sqrt{2})/2$
where l is the length of the tile

If the search returns 60 or more
organizations of interest…

Delete the results

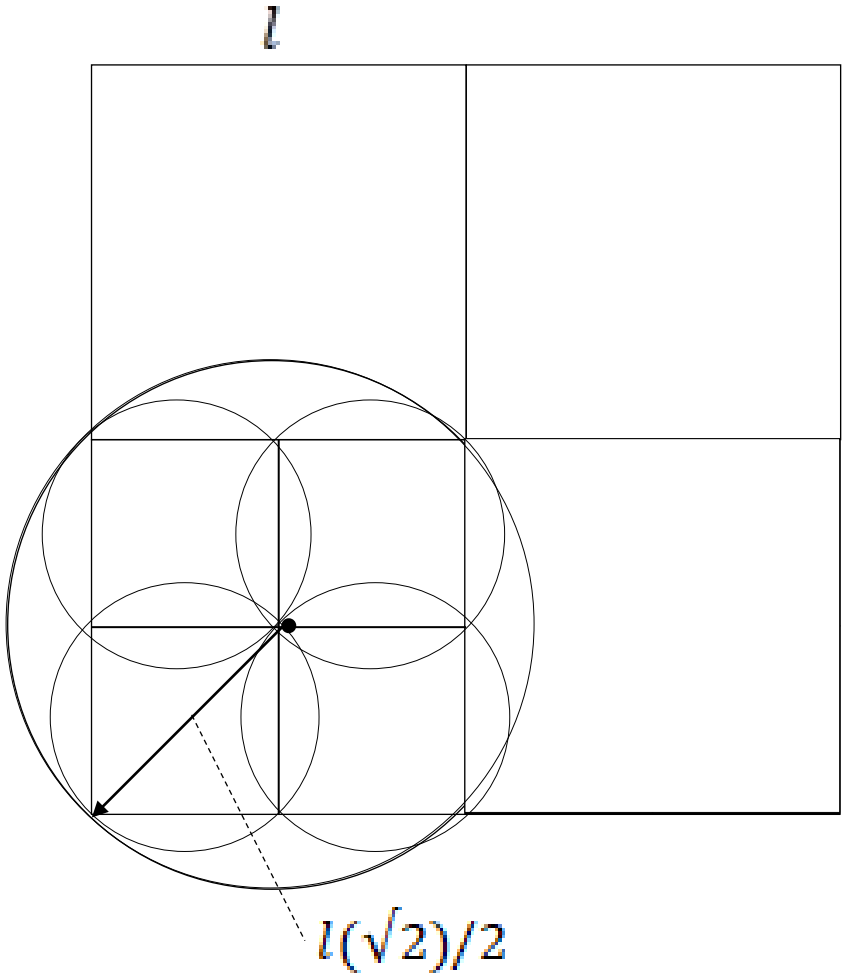


fulton@iu.edu
@BradRFulton

Searching for Organizations

Divide the tile into 4 equally-sized square tiles
(which is zoom level 16)

Repeat until the search returns
fewer than 60 organizations



Demonstration Case

Generate a sampling frame of religious congregations in the U.S.

Congregations are…

- ubiquitous: urban, suburban, and rural
  densely and sparsely populated regions
- diverse: size, age, religious tradition, and online presence
- nomadic: rented space, multi-tenant facilities, and homes
- emergent: high birth and death rates

fulton@iu.edu
@BradRFulton

Running the Program

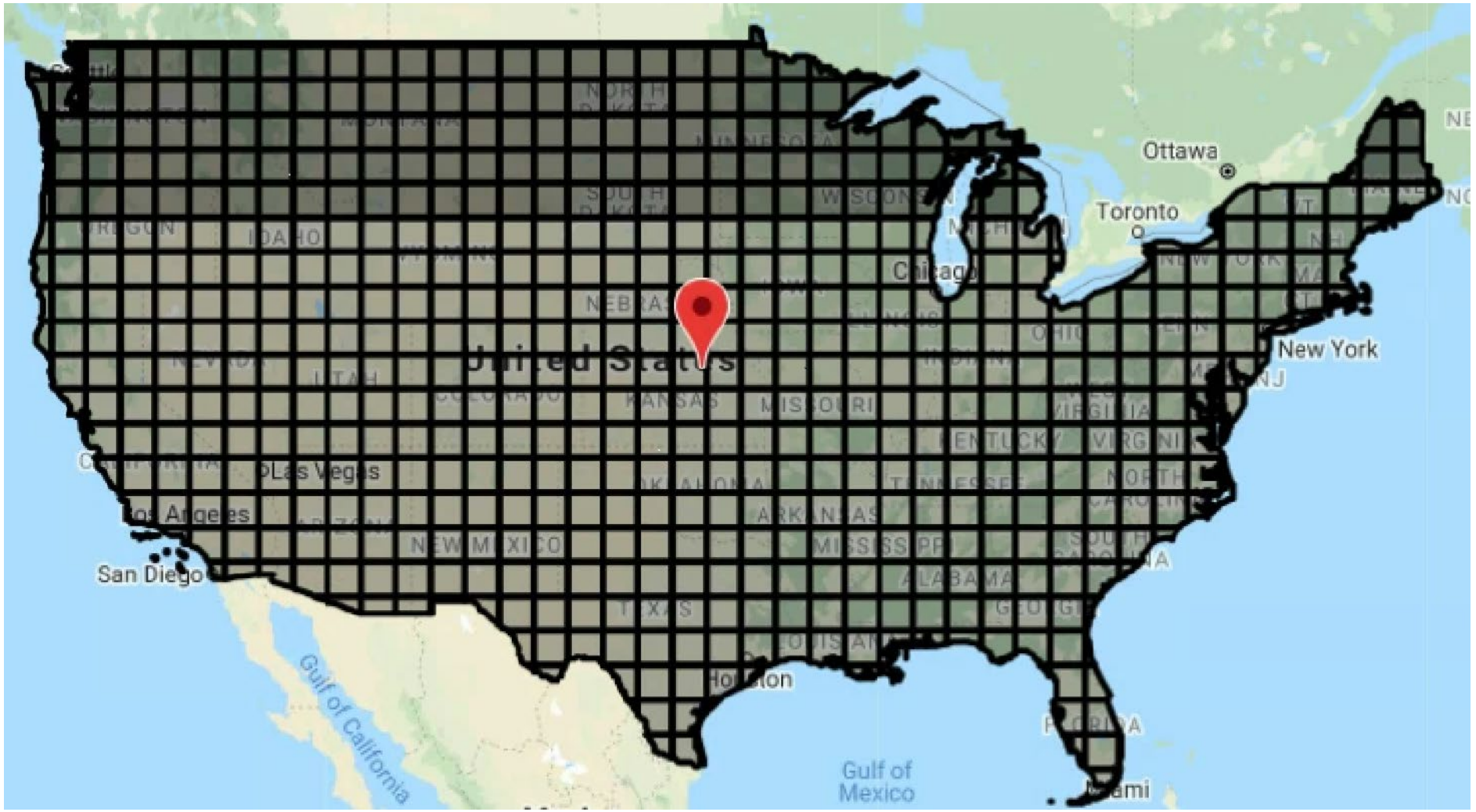
place type: church, synagogue, mosque and Hindu temple

Geographic scope: Continental U.S., Alaska, and Hawaii

Starting zoom level: 15

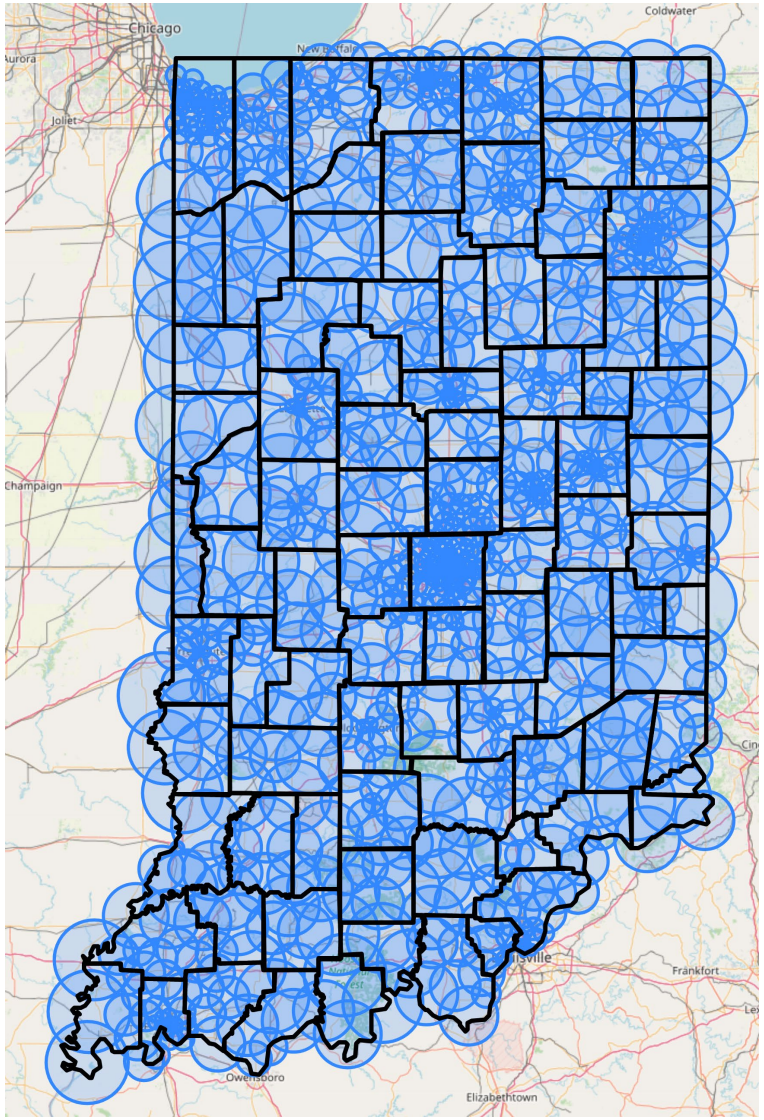
fulton@iu.edu
@BradRFulton

Running the Program



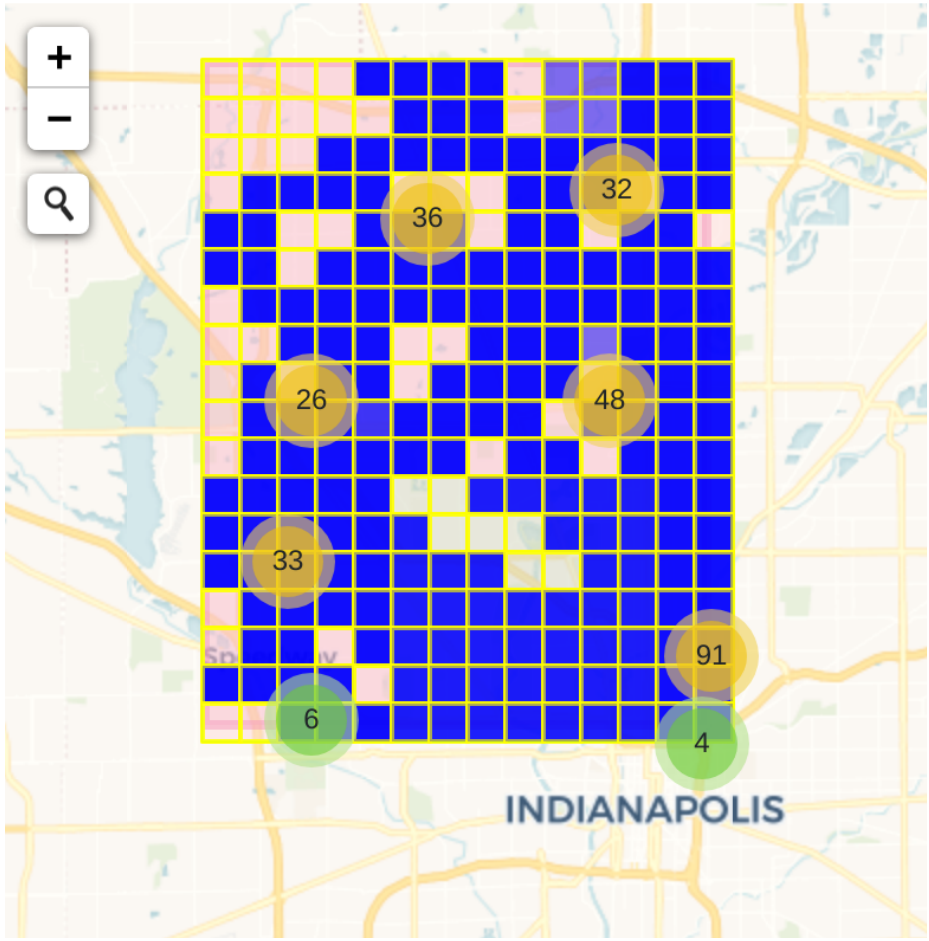
fulton@iu.edu
@BradRFulton

Running the Program



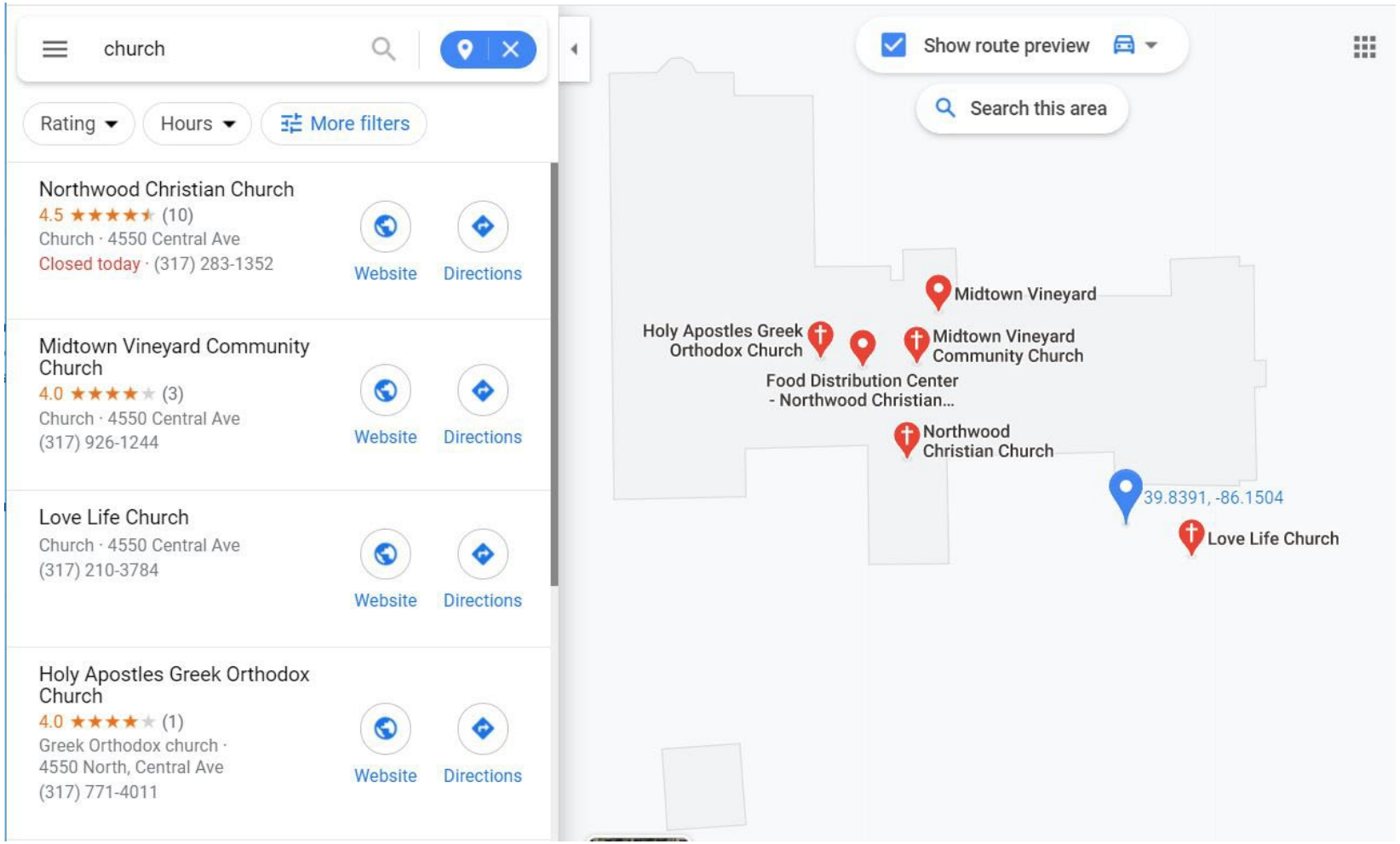
fulton@iu.edu
@BradRFulton

Running the Program



fulton@iu.edu
@BradRFulton

Running the Program



fulton@iu.edu
@BradRFulton

Google Maps' Comprehensiveness

Coverage: % of eligible units in the target population included in the sampling frame

Undercoverage error: Eligible units not included in the sampling frame

Not on Google Maps or Miscategorized by Google Maps

mortuary coded as *place type* cemetery instead of funeral home

Overcoverage error: Ineligible units included in the sampling frame

On Google Maps, but no longer operating or Miscategorized by Google Map

pet store coded as *place type* zoo instead of pet store

Listed more than once on Google Maps

Representativeness: Differences between eligible units included and not included in the sampling frame

fulton@iu.edu
@BradRFulton

Assessing Coverage

Two nationally representative samples of congregations (generated via hypernetwork sampling)

National Study of Congregations' Economic Practices (NSCEP) [N = 1,233]

National Congregations Study (NCS) [N = 1,262]

Percentage of NSCEP congregations on Google Maps: 98%

Percentage of NCS congregations on Google Maps: 97%

Census of congregations in Indianapolis [Center for Congregations]

Of the 937 congregations in Indianapolis, 920 (98%) are on Google Maps

fulton@iu.edu
@BradRFulton

Analysis of Undercoverage

Undercoverage due primarily to omission rather than miscategorization

Among the 27 congregations not in the sampling frame, 2 were miscategorized

- Health center
- Food bank

fulton@iu.edu
@BradRFulton

Assessing Overcoverage

Among the 948 congregations in the Indy sampling frame (generated from Google Maps)

- 13 (1.4%) duplicates
- 10 (1.1%) no longer operating
- 5 (0.5%) not congregations

Overcoverage error: 3%

fulton@iu.edu
@BradRFulton

Assessing Representativeness

Differences between congregations on and *not* on Google Maps

Too rare for rare events logistic regression

Characteristics of congregations *not* on Google Maps

Age of congregation:	new	Distribution similar to all congregations
Religious tradition:	non-Christian	Distribution similar to all congregations
Online presence:	minimal	44% have a website
Meeting facility:	non-traditional	63% own their building
Community context:	rural setting	40% located in an urban setting
Size of congregation:	small	Most have < 100 members (median = 50)

fulton@iu.edu
@BradRFulton

Google Maps' Comprehensiveness

Near-complete coverage

Near-perfect degree of representativeness

fulton@iu.edu
@BradRFulton

Limitations

1. Overcoverage error: including organizations outside the scope / target population
    - Inflates the population count
        - overcoverage can be estimated (1.6%)
2. Duplication error: multiple cases in the frame linked to the same organization
    - Inflates the population count
        - overcoverage can be estimated (1.4%)
    - Duplicate cases are twice as likely to be selected in a random sample
        - a handful of mitigation strategies (see paper)
3. Undercoverage error: not including organizations inside the target population
    - Deflates the population count
        - more difficult to estimate (2%)

fulton@iu.edu
@BradRFulton

Technical Expertise

The Polis Center

Neil Devadasan

fulton@iu.edu
@BradRFulton

Research Assistants















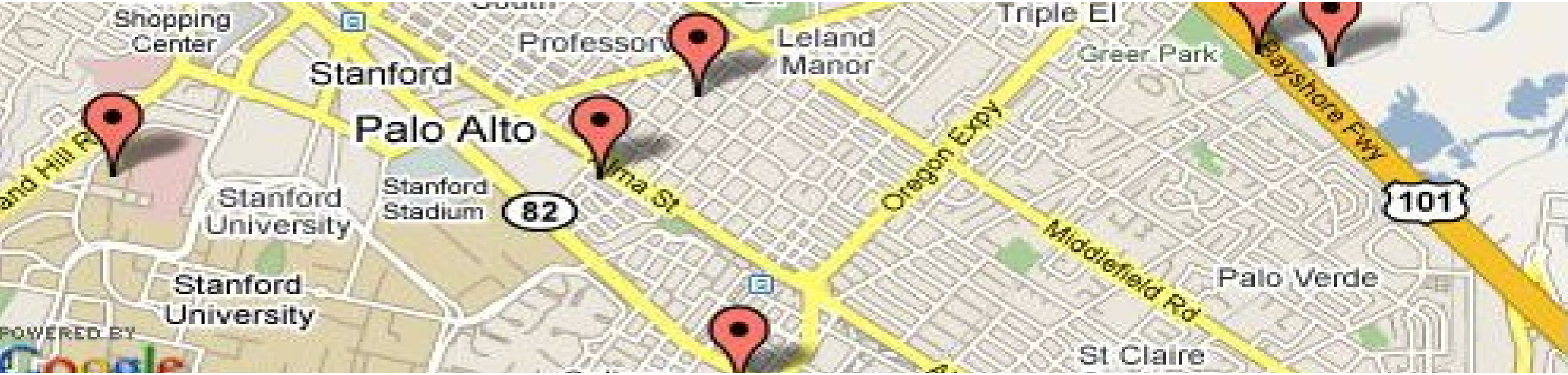



fulton@iu.edu
@BradRFulton

Funding Source

fulton@iu.edu
@BradRFulton

Using Google Maps to Generate Organizational Sampling Frames



Brad R. Fulton and David P. King
Indiana University

SSRC Workshop in Methods

2023.1.20

fulton@iu.edu
@BradRFulton