Introduction to Geoinformatics for Epidemiology

Cole Brokamp

23 September 2018

GRAPPH

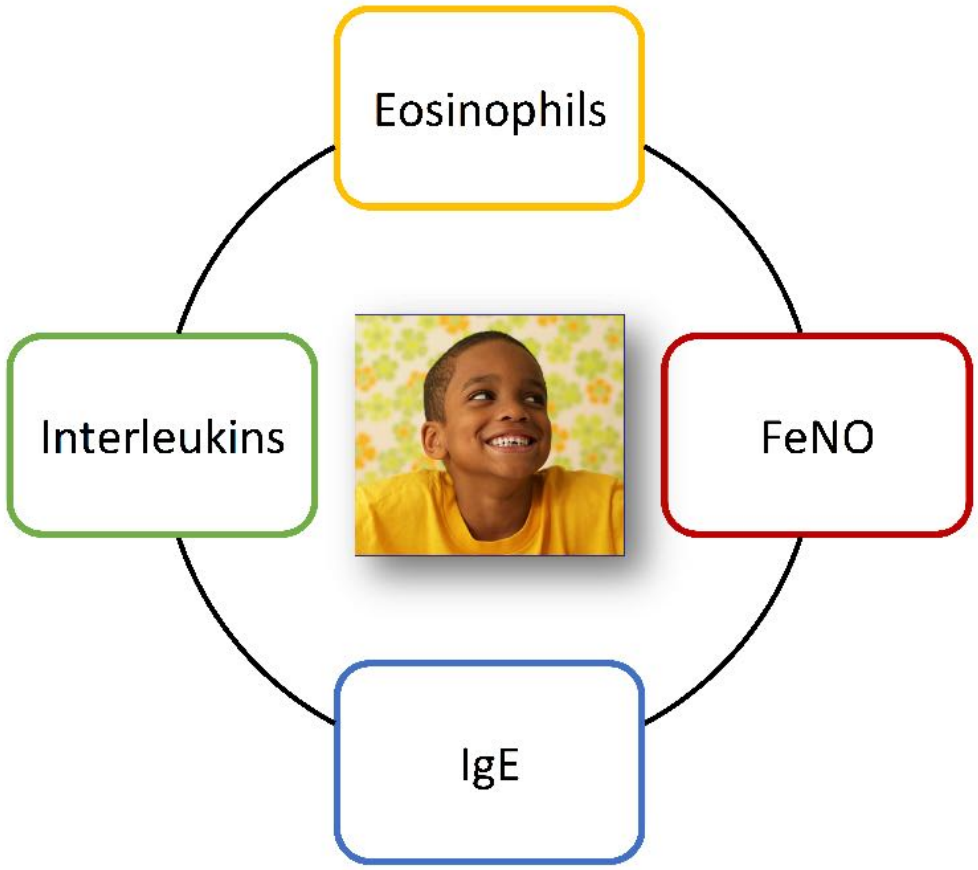
Geospatial **R**esearch **A**ccelerator for **P**recision **P**opulation **H**ealth

- An interdisciplinary team of faculty, staff, and clinicians at Cincinnati Children's Hospital Medical Center with expertise in:
  - programming and informatics
  - (bio)statistics
  - epidemiology
  - population health
- Collaboration across divisions of:
  - Biostatistics and Epidemiology
  - Biomedical Informatics
  - Information Services
  - General & Community Pediatrics
  - Hospital Medicine
  - James M. Anderson Center

Table of contents

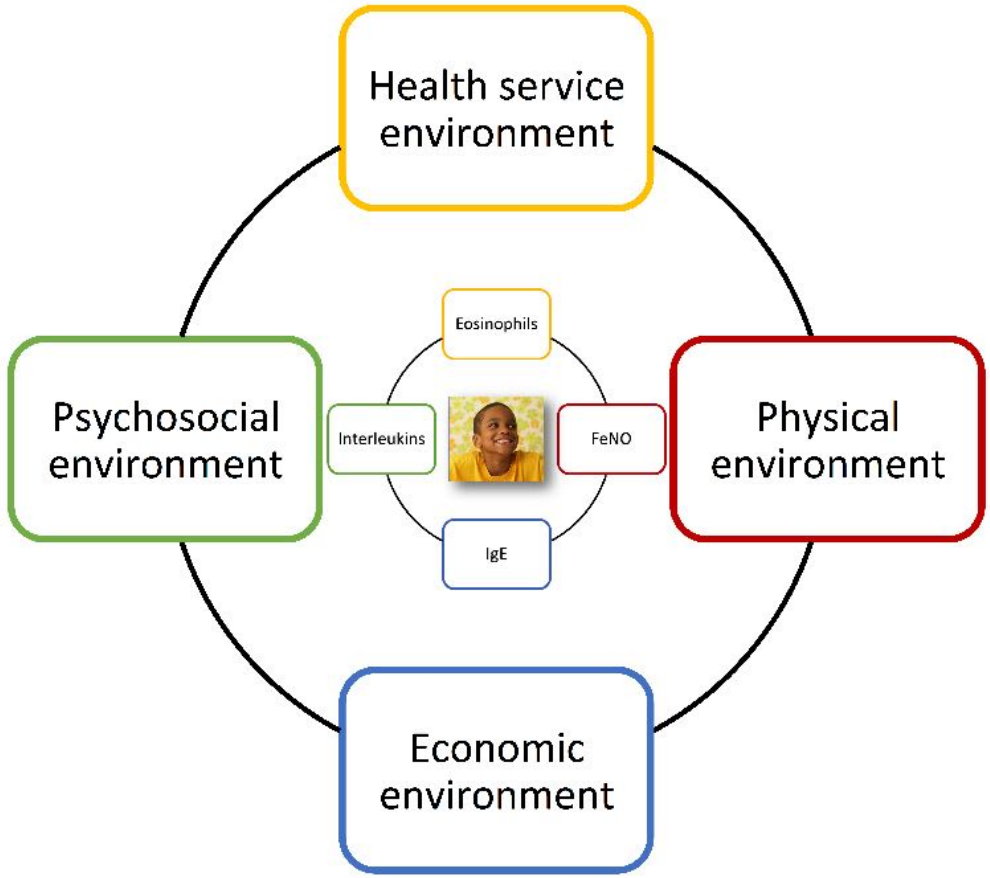
From precision medicine . . .

Biomarker: any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease



. . . to precision public health

Geomarker: any objective, contextual or geographic measure that influences or predicts the incidence of outcome or disease



Geomarkers and Health

Geomarkers are a powerful predictor of disease, disorder, injury, and mortality.

The role of environment on health has been known for decades, but the data and tools needed to characterize the relationship at a population level were not available until recently.

Geoinformatics

Geoinformatics defined as:

> *"The science and technology dealing with the structure and character of spatial information, its capture, its classification and qualification, its storage, processing, portrayal and dissemination, including the infrastructure necessary to secure optimal use of this information"* [1]

[1] P.L.N. Raju, Fundamentals of Geographic Information Systems

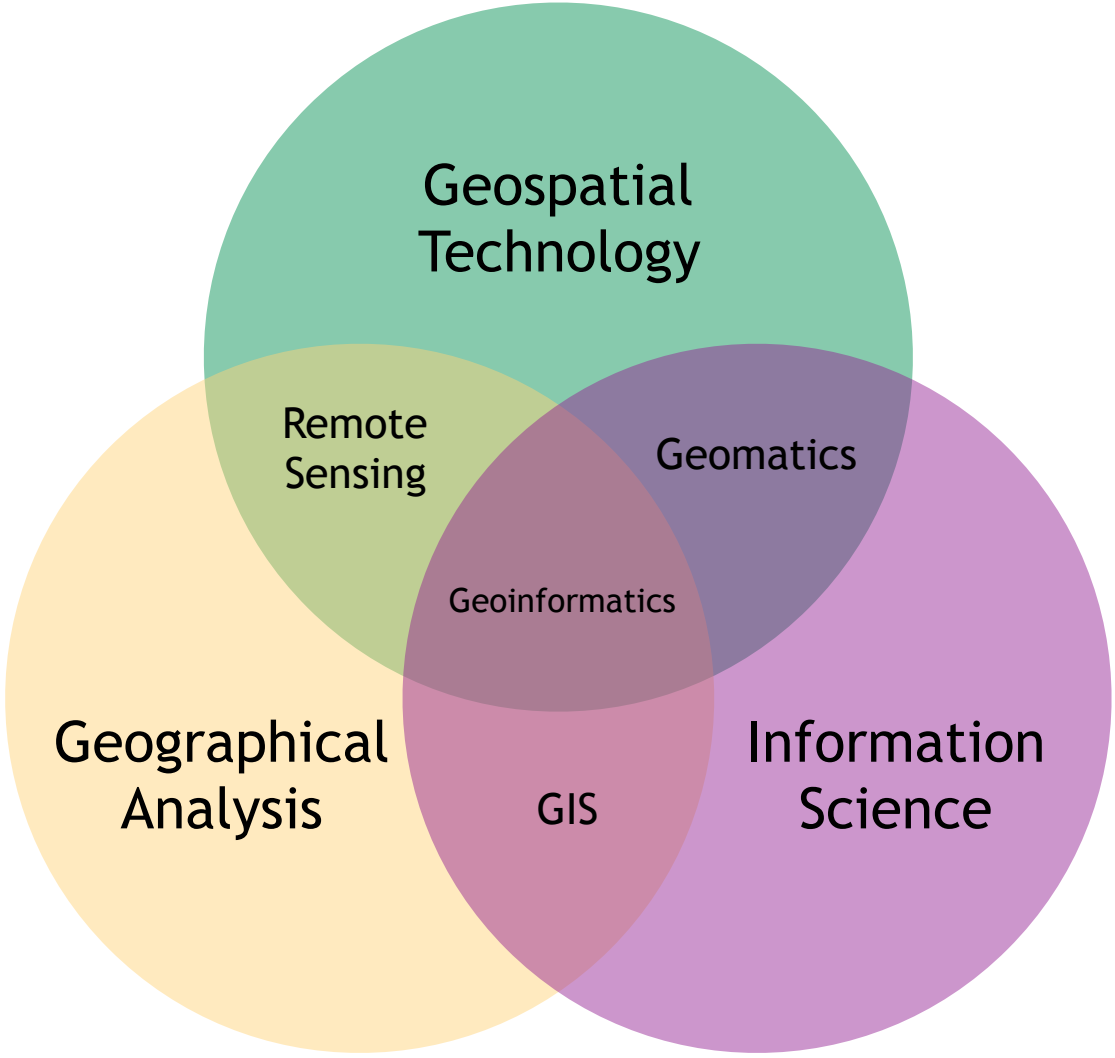
Geoinformatics

You know it as:

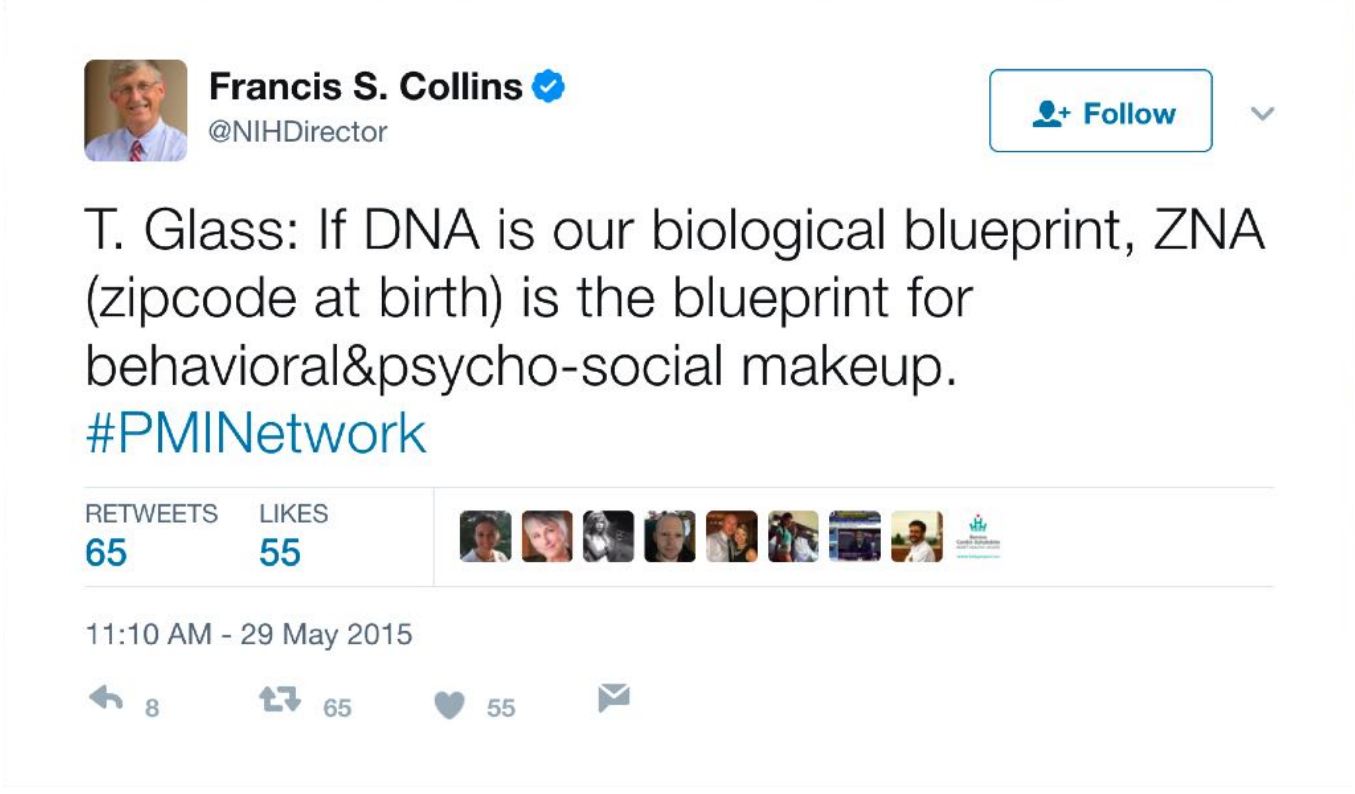
- Google maps
- GPS navigation
- Hamilton County Auditor's Office
- Weather forecasting
- US Postal Service
- Election maps

Geoinformatics

Transforming spatial data into actionable knowledge



Geoinformatics and Health



Geoinformatics and Health

Uses:

- Identifying study population
- Identifying potential sources and routes of exposures
- Estimating environmental levels of pollutants
- Measuring community characteristics
- Estimating personal socioeconomic characteristics
- Statistical models with spatial correlations
- Temporal exposure estimates

Geoinformatics and Health

Advantages:

- Direct measurement of exposure or personal characteristics often not feasible in cohort- or EHR- based studies
  - Often use retrospective records, only containing home addresses
  - Geomarkers are highly variable with respect to time and space
- Links outcome, exposure, and confounding data by location
- Data usually available publicly

GRAPPH Application Arms

1. Innovative Research
   - Area-based SES confounders
   - Place-based predictors of health outcomes
2. Place-based Patient Care
   - Add information to clinical decision making
   - Deeper risk assessment and/or intervention deployment
3. Community-centered Quality Improvement
   - Illustrate health disparities
   - Identify disparity-reducing interventions

GRAPPH Application Arms

Cincinnati Children's

Geospatial **R**esearch **A**ccelerator for **P**recision **P**ublic **H**ealth

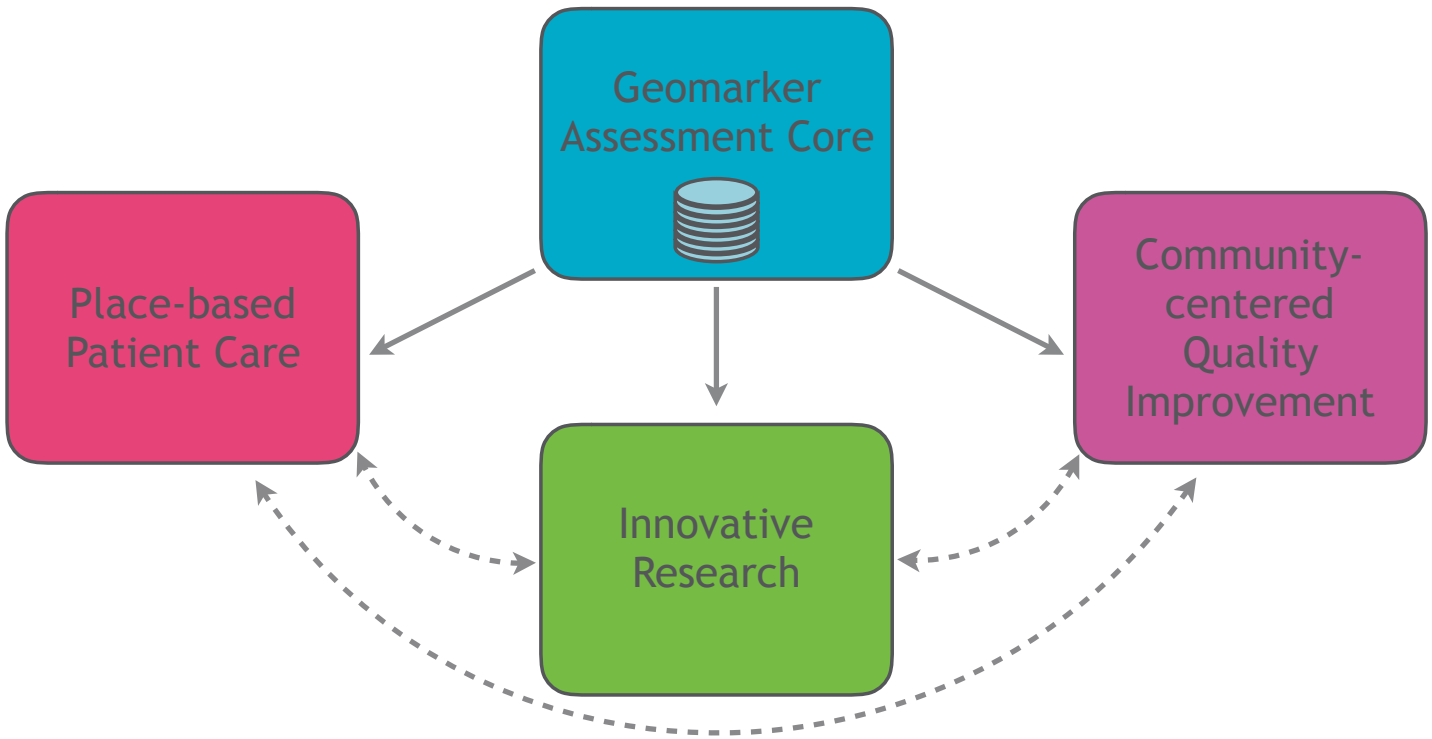


Table of contents

Challenges

Aggregating different spatial and temporal resolutions from disparate sources like government agencies, academic departments, hospitals, pharmacies, and others

Available Geomarker Data

Available Geomarker Data

Cole Brokamp

(cole.brokamp@cchmc.org)

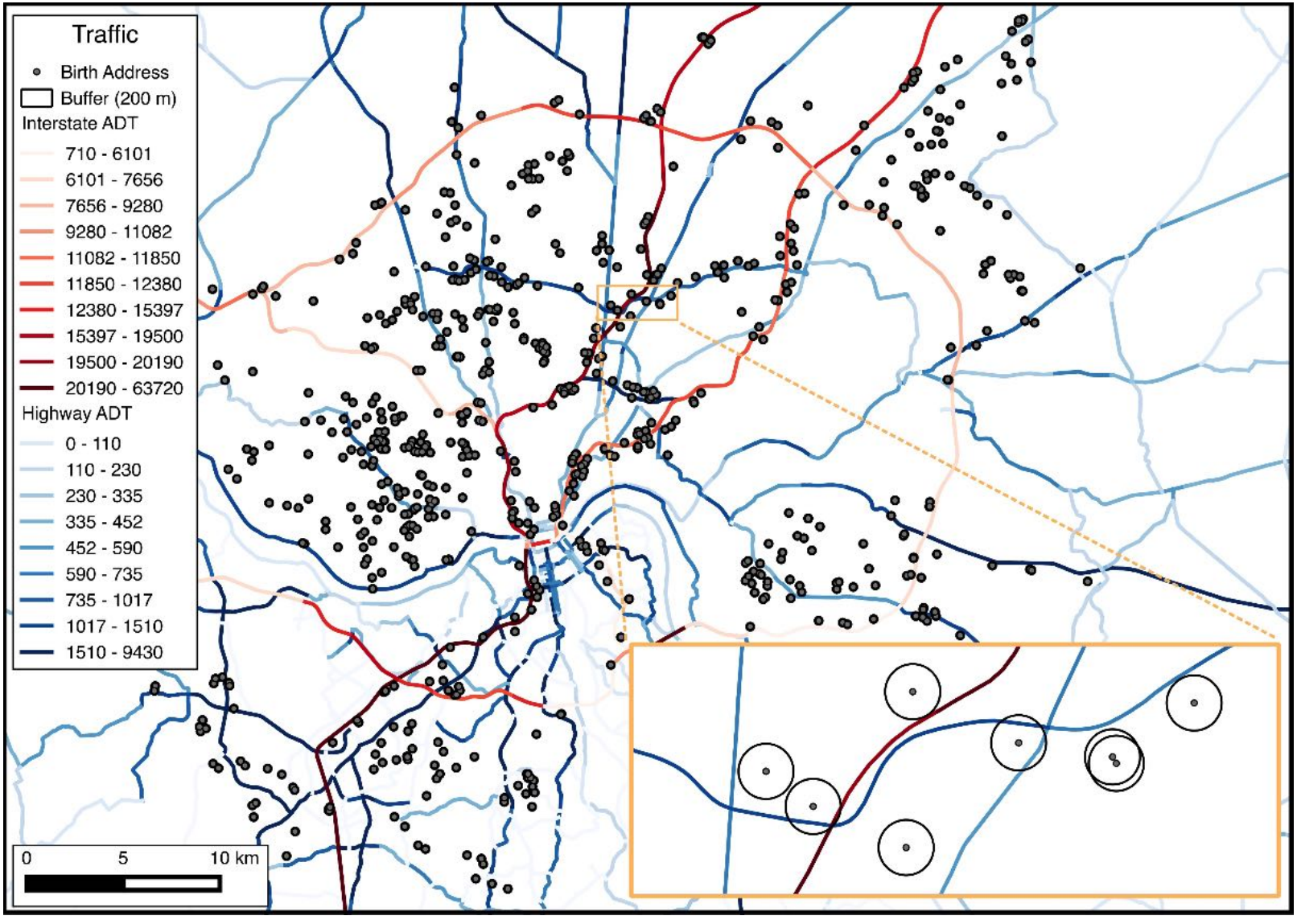
Level 1

Database	Example Data	Space	Time
Topologically Integrated Geographic Encoding and Referencing (TIGER)	Proximity to roadway by type, locations of airports, railroads, bodies of water, elevation	Exact location (nationwide)	Yearly
Open Street Map (OSM)	Public transportation routes	Exact location	Current
Global Historical Climatology Network (GHCN)	Temperature, relative humidity, precipitation, wind speed and direction	Interpreted from > 5000 stations (nationwide)	Daily
National Emissions Inventory (NEI)	Location and amount of emissions	Exact location	Yearly
EPA Air Quality System (AQS)	Fixed site monitoring of ambient pollutant levels, air quality alerts	Interpreted from over 2500 stations (nationwide)	Daily
US Department of Transportation (DOT)	Highway traffic intensity	Exact location (nationwide)	Yearly

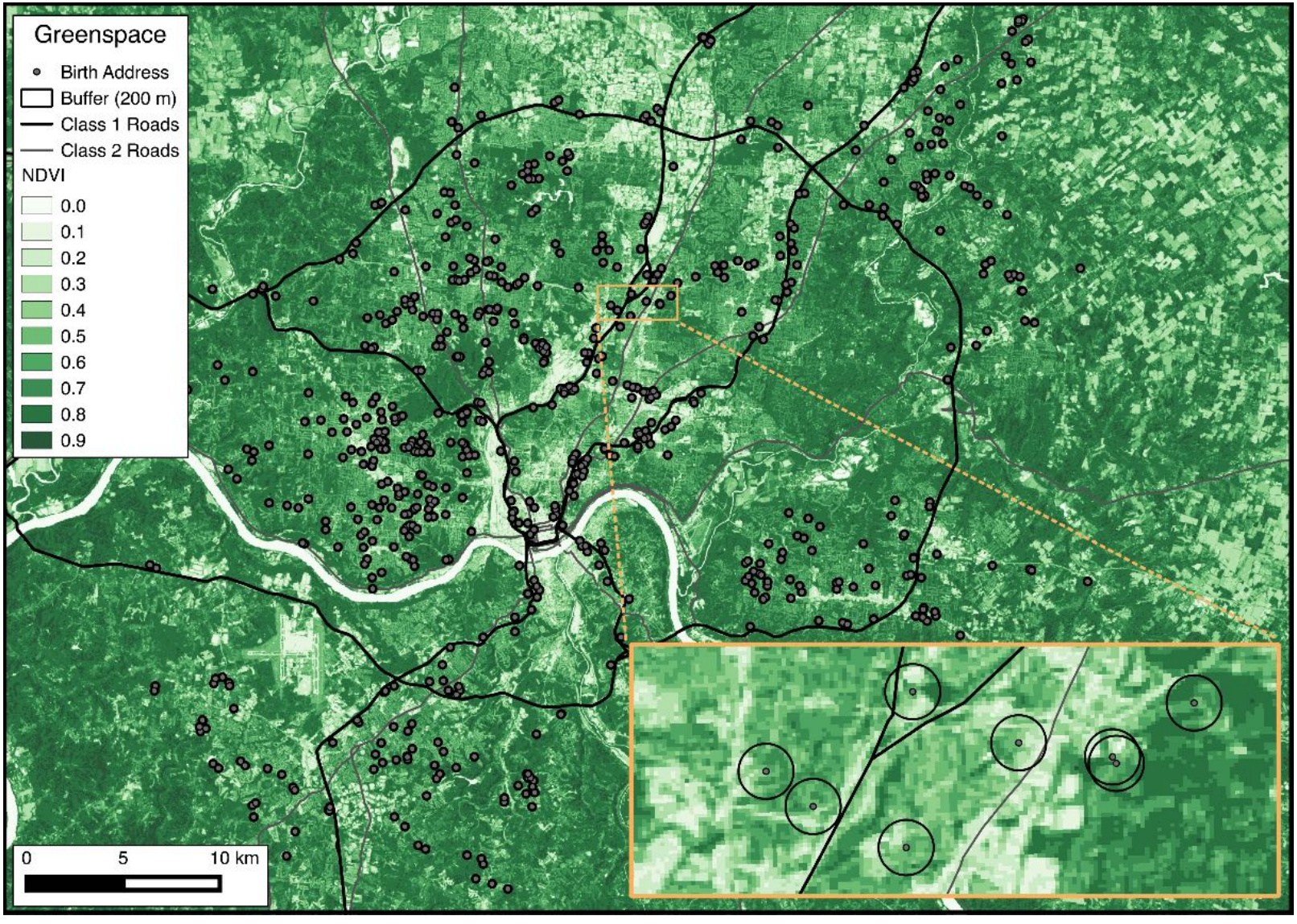
Level 2

Database	Example Data	Space	Time
US Census American Community Survey (ACS)	Population, socioeconomic indicators	Census tract or block (nationwide)	Yearly
North American Regional Reanalysis (NARR)	Air temperature, planetary boundary height, relative humidity, precipitation, wind	North America (0.3°×0.3°)	8-times (2001 – 2016); daily means (1979 – 2000)
Moderate Resolution Imaging Spectroradiometer (MODIS)	Greenspace, land cover, aerosol optical thickness, surface reflectance	3km square grids (nationwide)	Daily (2000 – 2015)
National Land Cover Database (NLCD)	21 classes of land cover classifications	30 x 30 m grid	Annual (1992, 2001, 2006, 2011)
Smart Location Database (SLD)	Location efficiency, characteristics of the built environment	Census tract	2010
Modified Retail Food Environment Index (mRFEI)	Fraction of food retailers that are "healthy"	Census Tract	2011

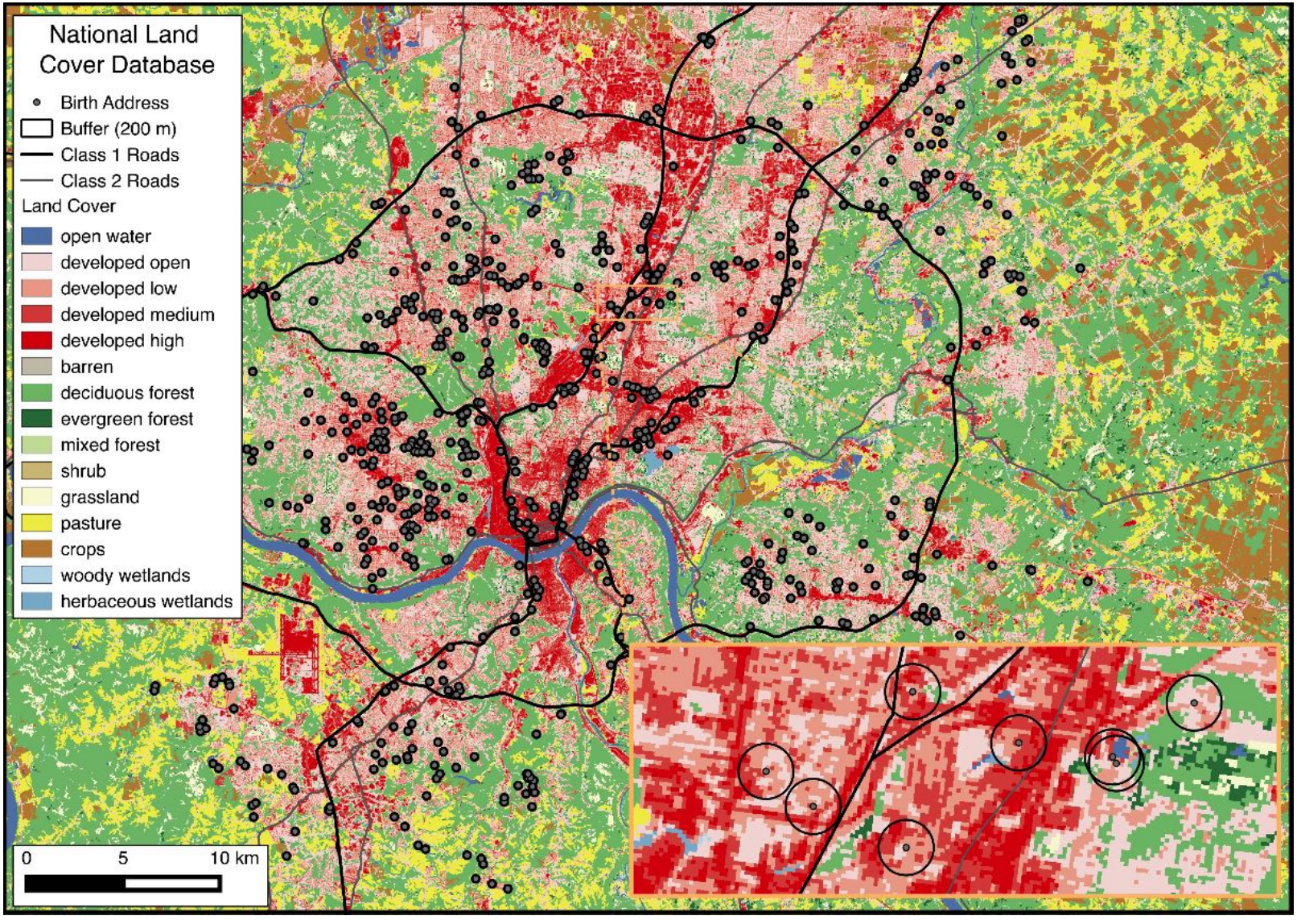
Truck Traffic



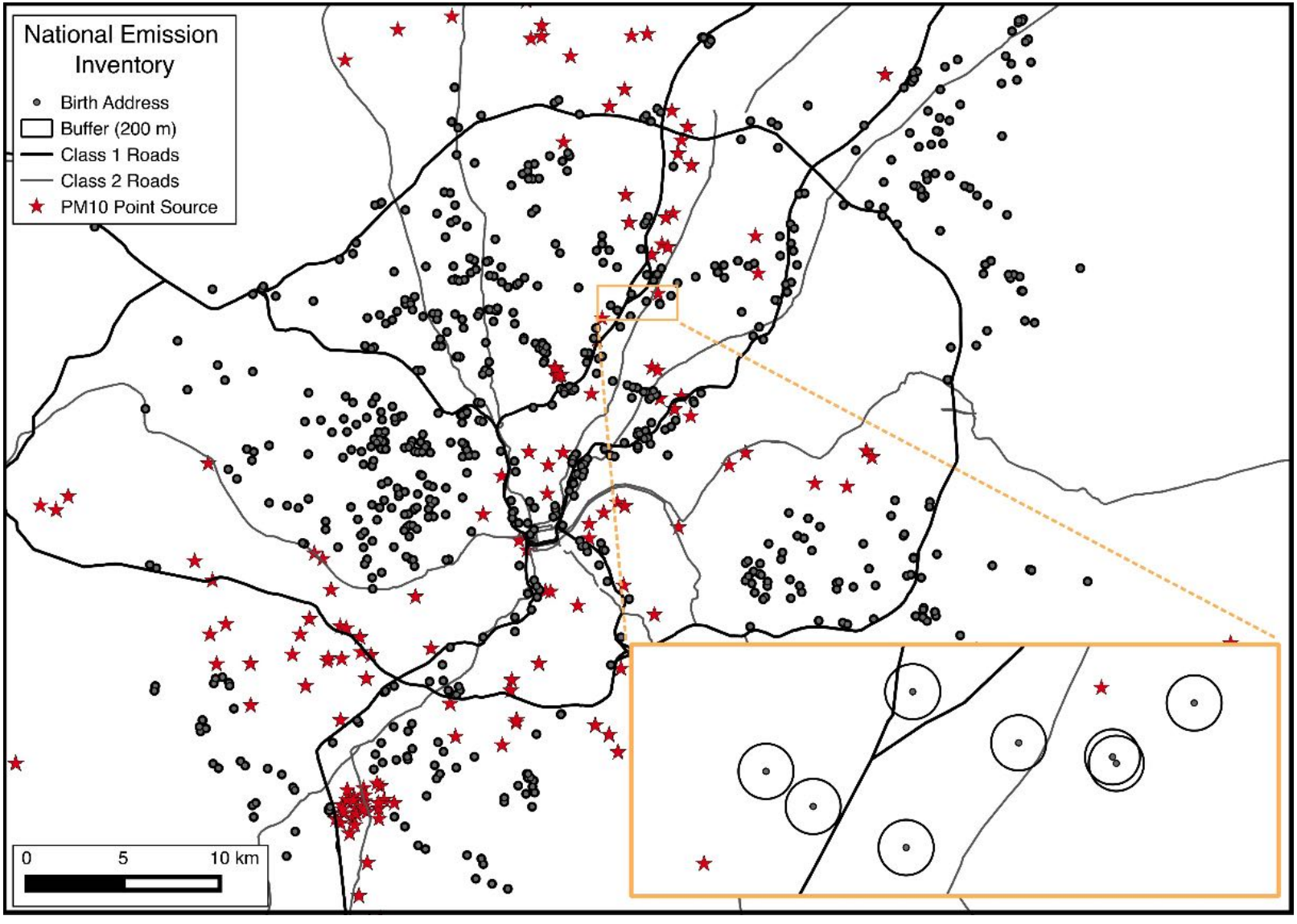
Greenspace


Greenspace
Birth Address
Buffer (200 m)
Class 1 Roads
Class 2 Roads
NDVI
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
5
10 km

Land Cover


National Land Cover Database
Birth Address
Buffer (200 m)
Class 1 Roads
Class 2 Roads
Land Cover
open water
developed open
developed low
developed medium
developed high
barren
deciduous forest
evergreen forest
mixed forest
shrub
grassland
pasture
crops
woody wetlands
herbaceous wetlands
0
5
10 km

NEI Point Sources



PM2.5 Across Space and Time

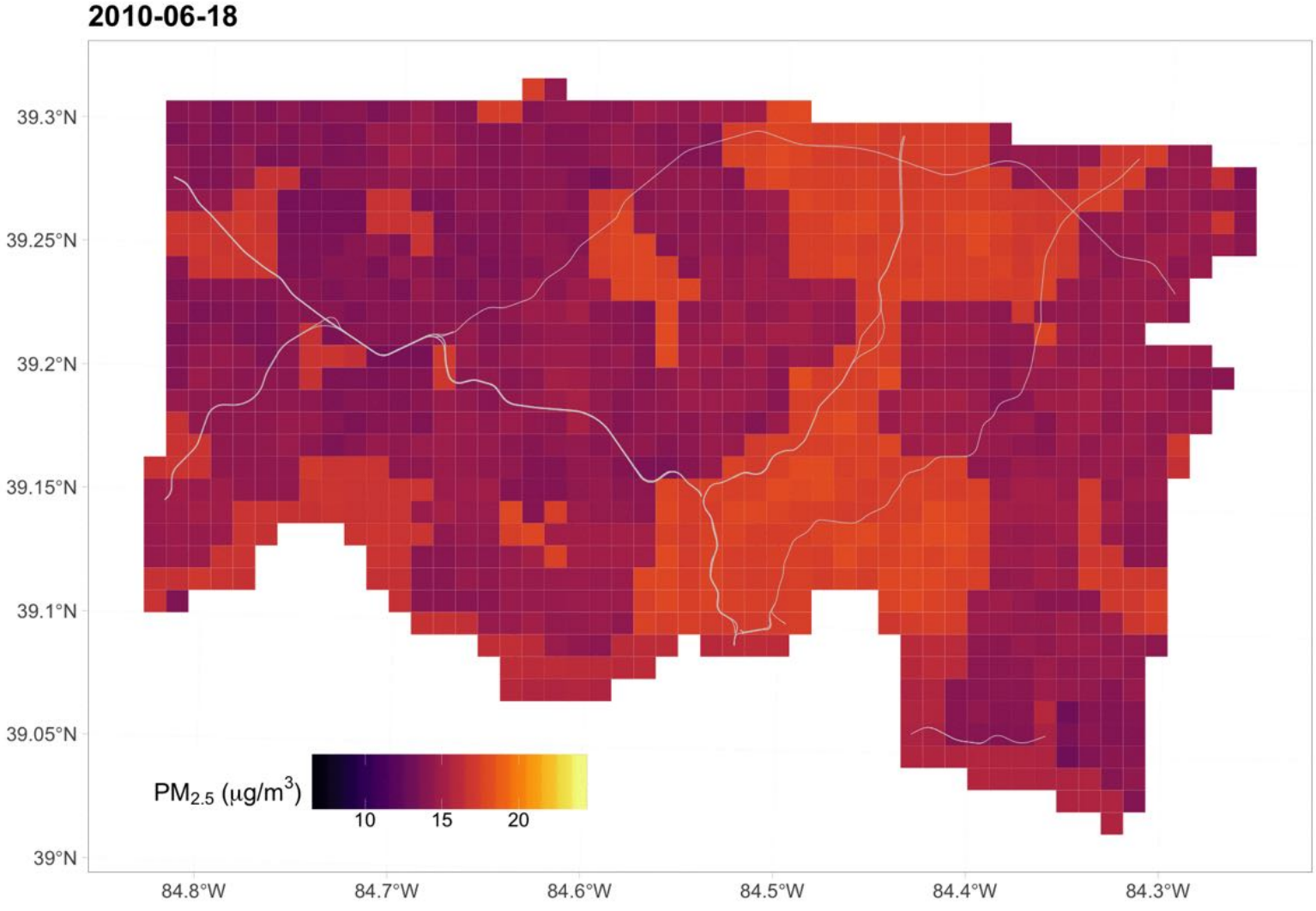


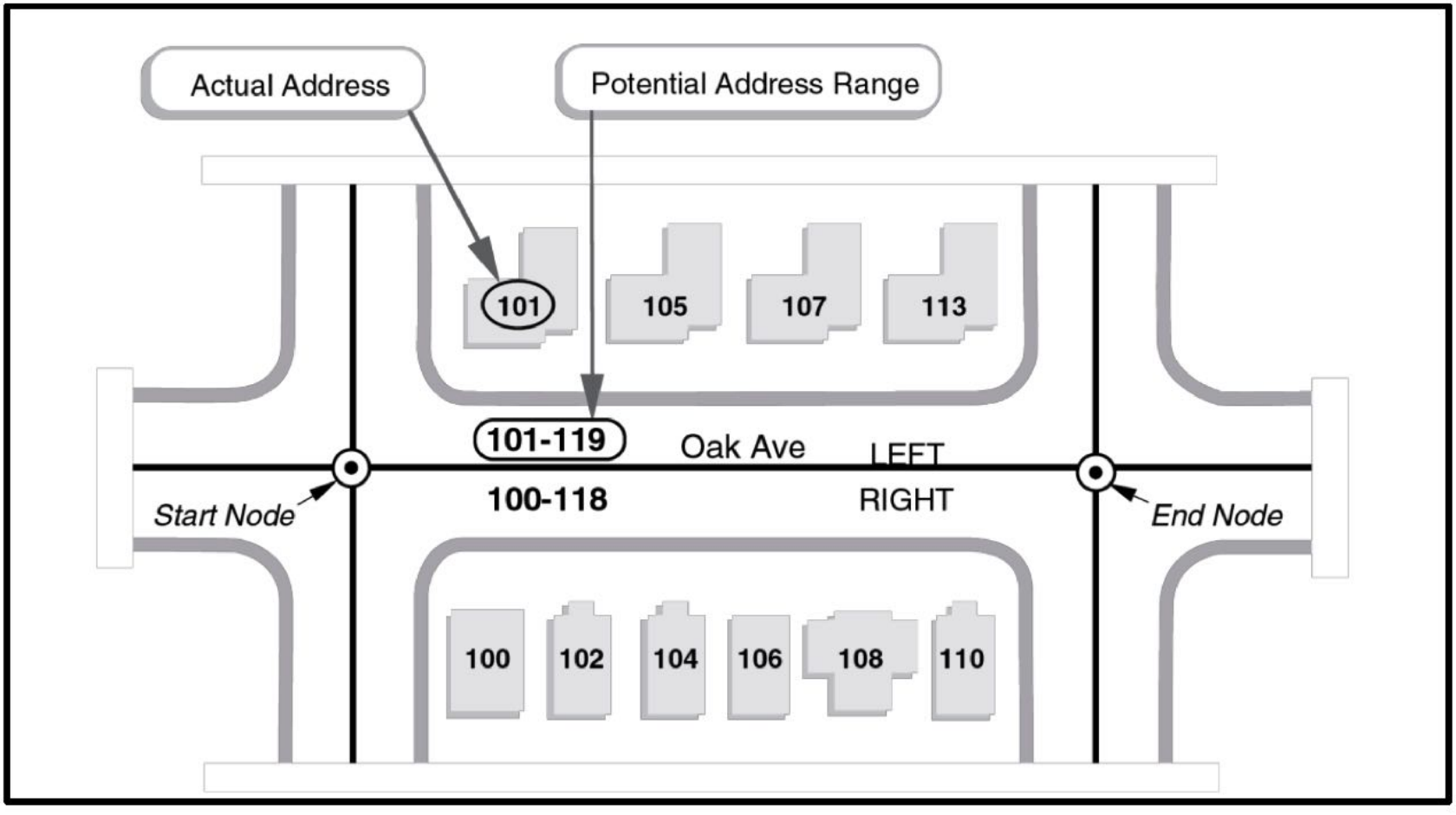
Table of contents

Geocoding

Geocoding

- converting location information text into coordinates
- most often a postal address into latitude and longitude
- most consider it a “magic black box”, but is error prone

Street range address



Exact location

Exposure Assessment

1. Containing Geography:
   - Census tract linkage to survey data
   - Census block linked to population density
   - Neighborhood linked to policies or characteristics

Exposure Assessment

1. Containing Geography:
   - Census tract linkage to survey data
   - Census block linked to population density
   - Neighborhood linked to policies or characteristics
2. Radial Measures:
   - Buffer designated around location with a radius
   - Length / density of predicted sources
   - Calculate mean, total length, or fraction within buffer

Exposure Assessment

1. Containing Geography:
   - Census tract linkage to survey data
   - Census block linked to population density
   - Neighborhood linked to policies or characteristics
2. Radial Measures:
   - Buffer designated around location with a radius
   - Length / density of predicted sources
   - Calculate mean, total length, or fraction within buffer
3. Exact Location:
   - Proximity to predicted source
   - Nearest neighbor weighting and krigging
   - Prediction models (land use models, etc)

Geoinformatics is more than mapping

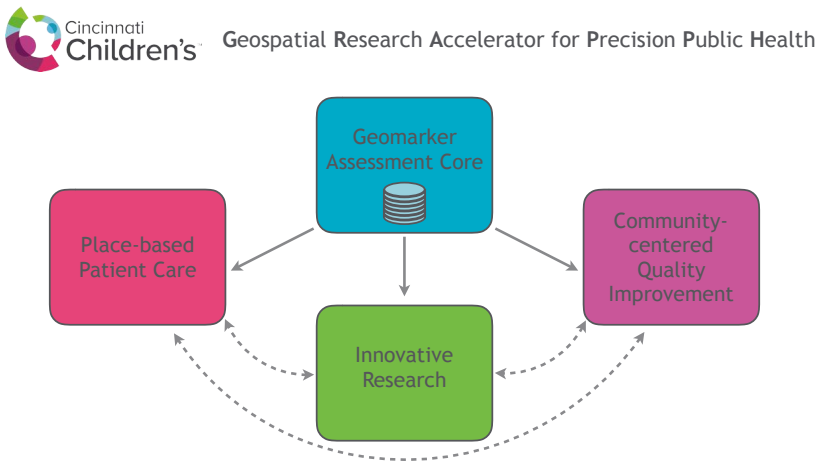
- Interactive and linked visualizations
- Statistics and distributions to put a place on a relative scale
- Predictive modeling and inference

Table of contents

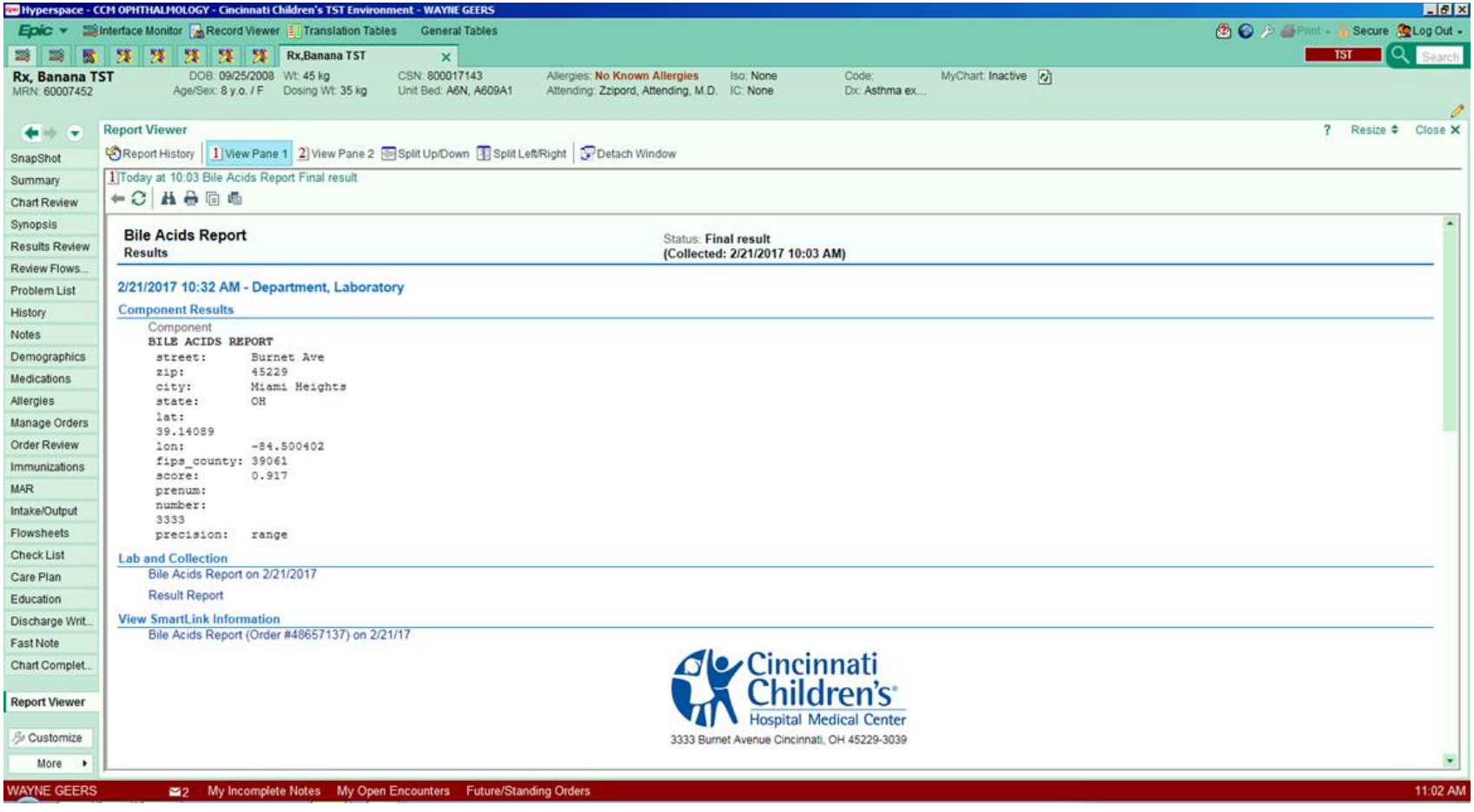
App aims

Who is your audience? What is the purpose of your application?

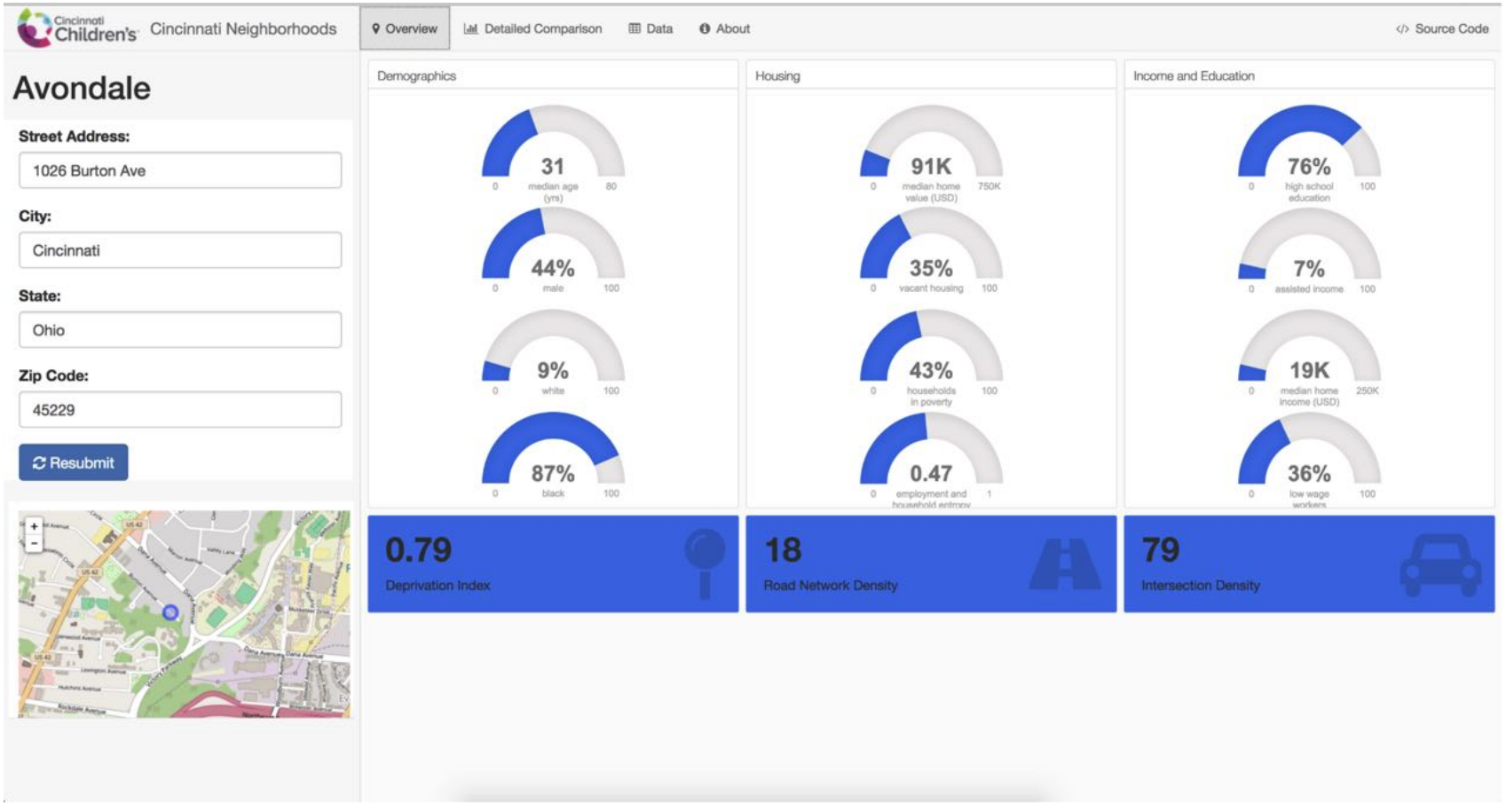
- hypothesis generation
- community engagement
- place-based care
- research database supplementation
- quality improvement
- public education



Place-based patient care

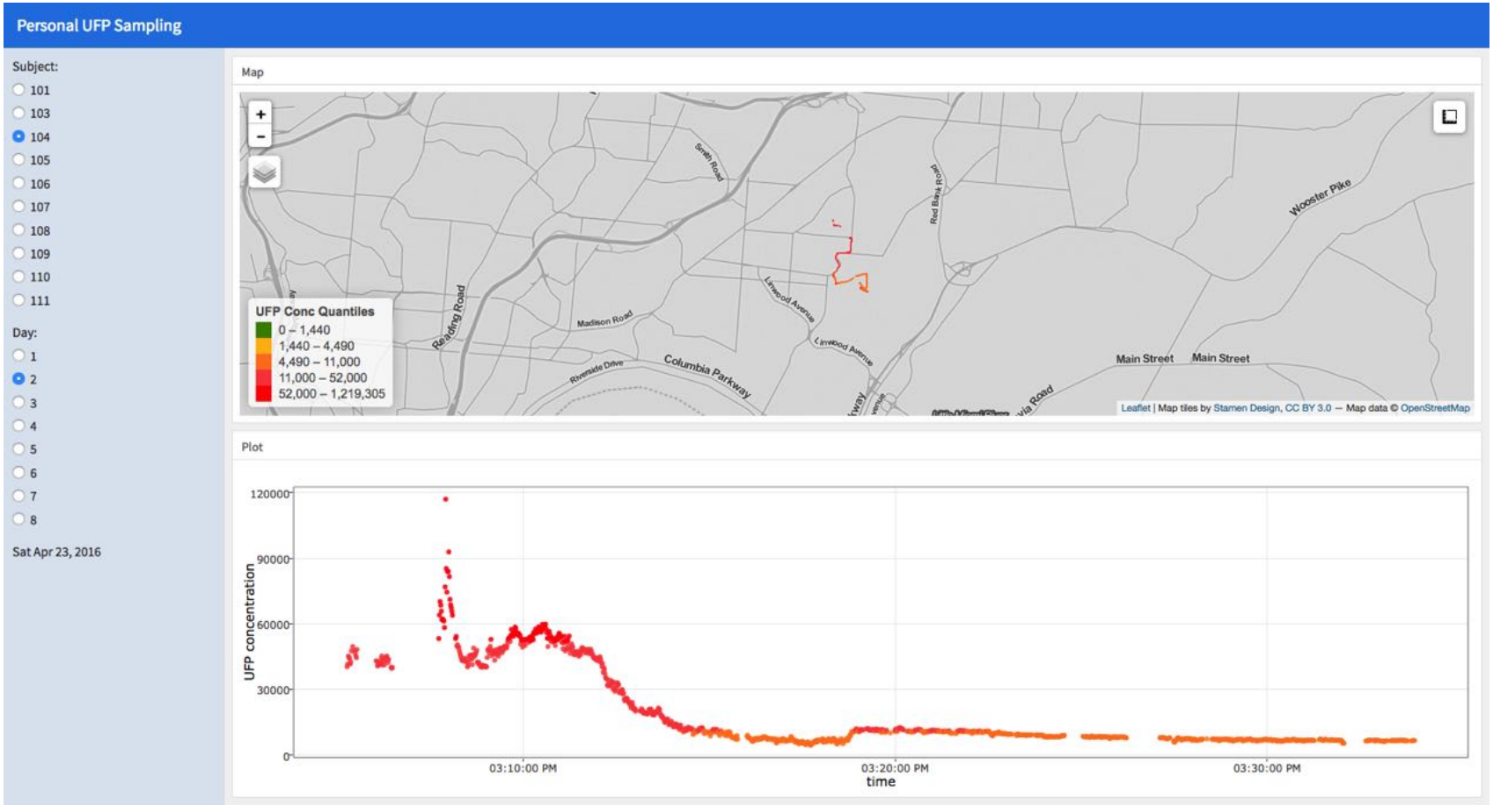


Place-based patient care

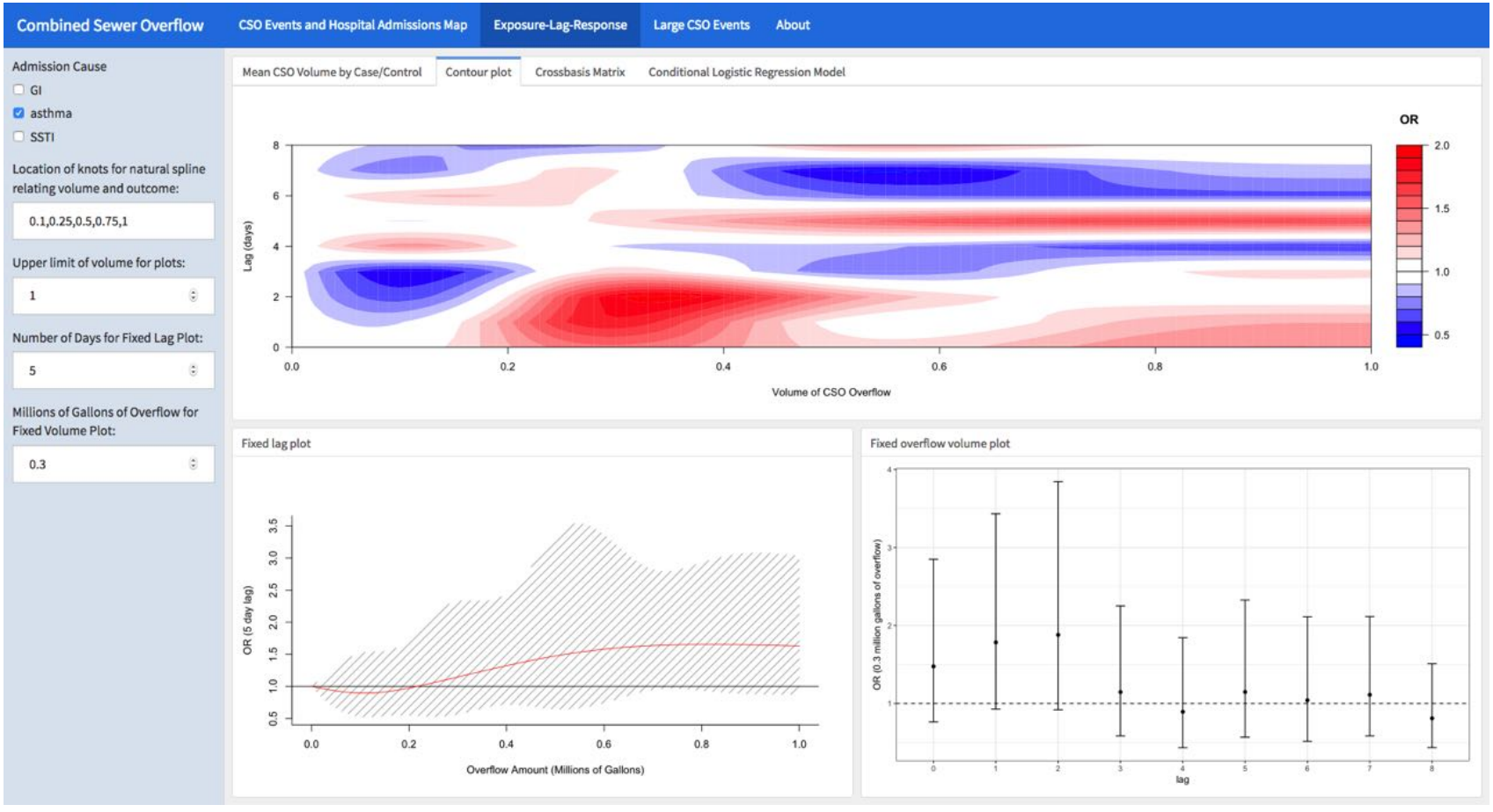


Place-based patient care

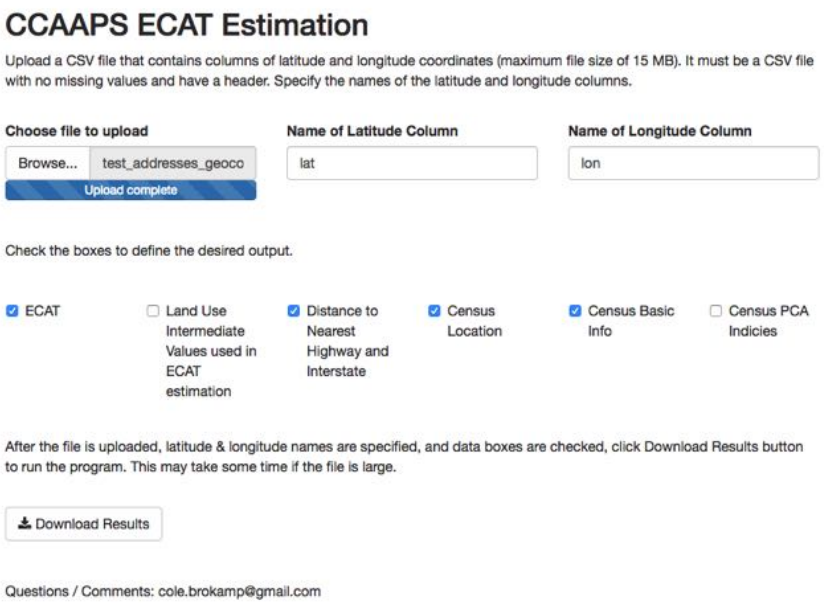
Innovative research



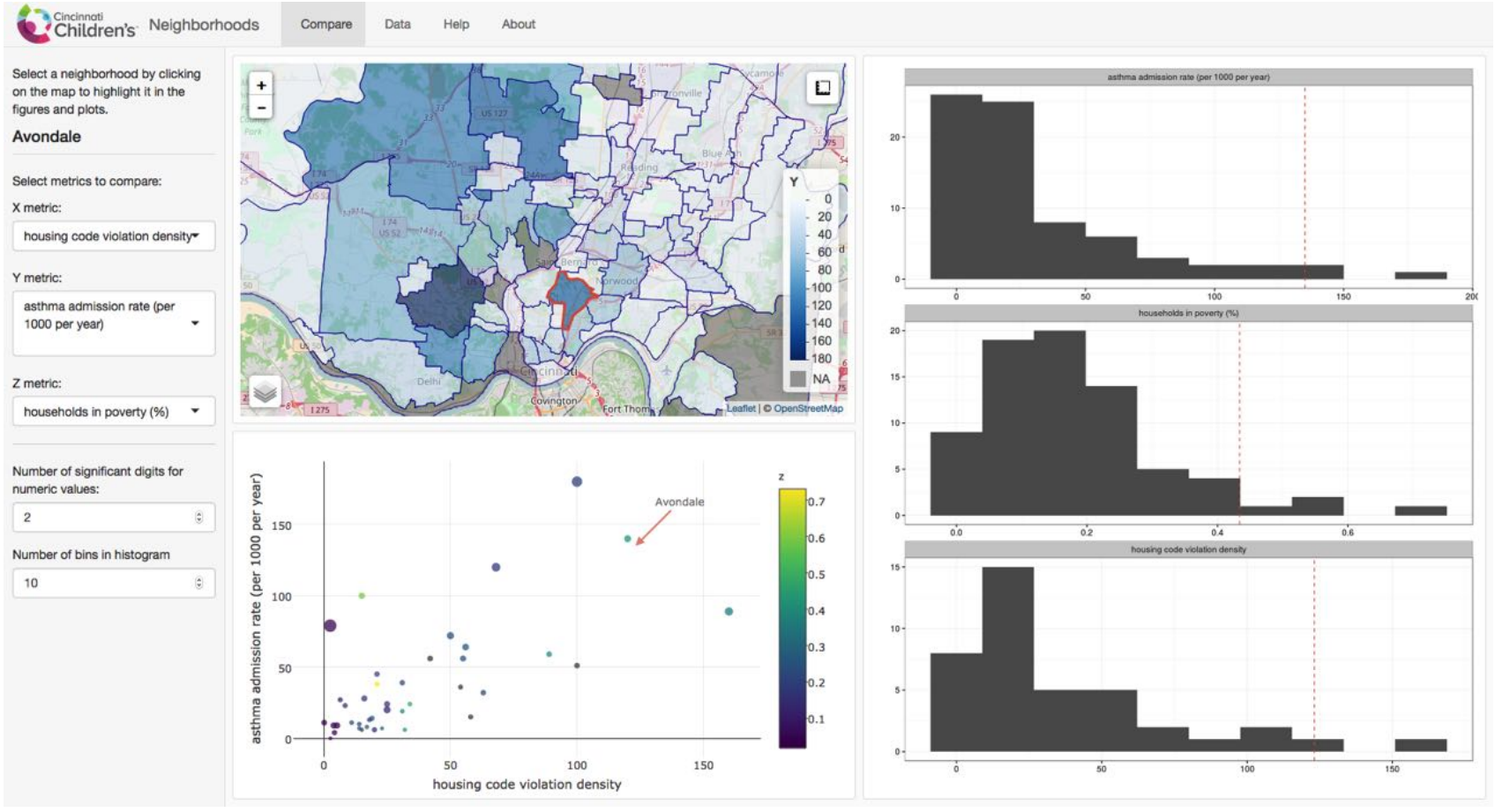
Innovative research



Innovative research



Community-centered quality improvement



http://geomarker.io/neighborhood

Thank you

Cole Brokamp

cole.brokamp@cchmc.org
@cole_brokamp
https://colebrokamp.com