

UrbanHealthOpt

Optimizing urban healthcare placement

Aniruddha Karajgi — ajk8815

Varad Naik - vvn7114

Saumya Parag Phadkar - sp7778

[Github: Polaris000/UrbanHealthOpt](#)

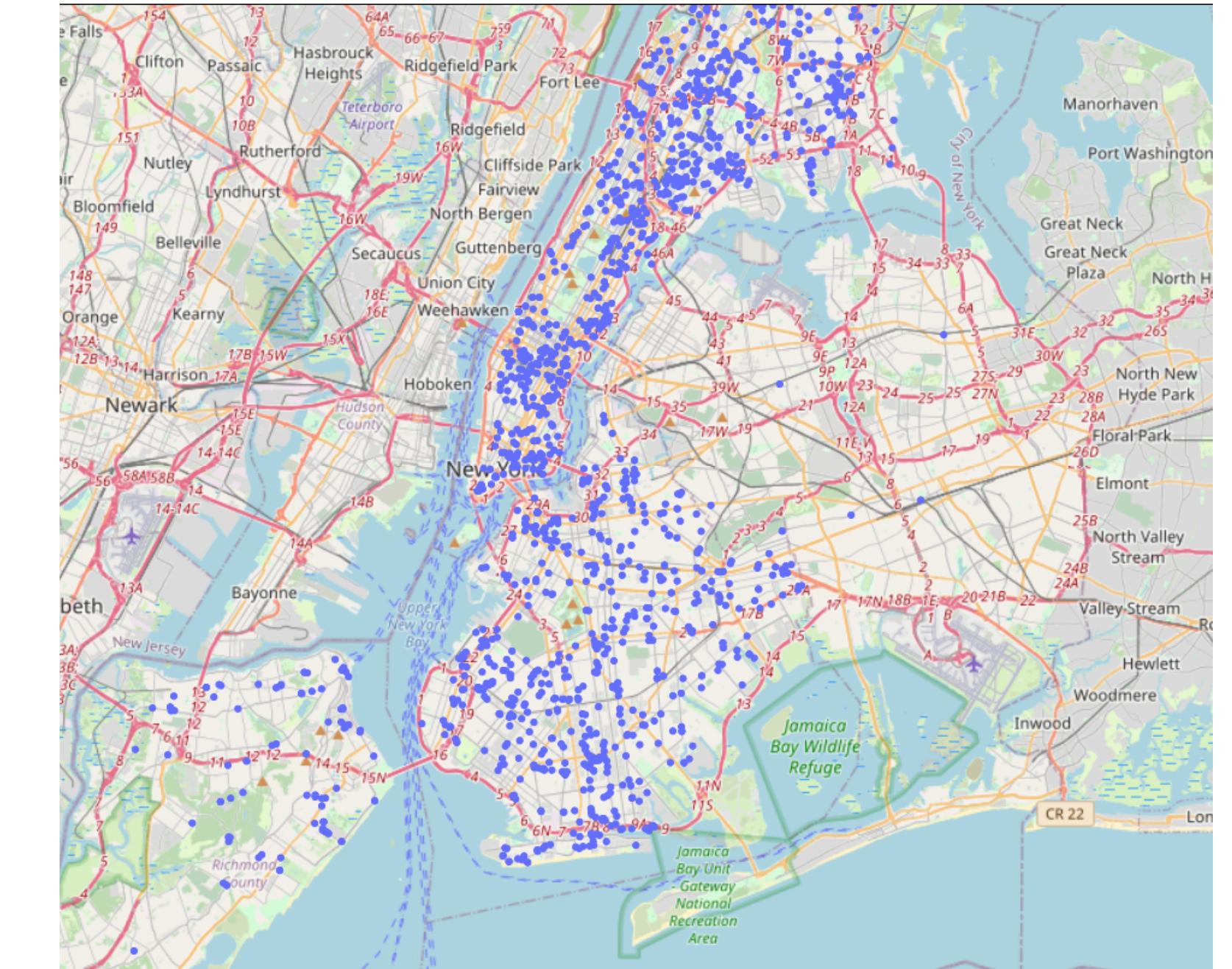
Problem Statement

Optimize the placement of urban healthcare centers across New York based on demographic factors and health indicators.

Datasets

Hospitals Dataset

- Health Facility General Information dataset
 - Source - Health Facilities Information System
 - Granularity level: Coordinates
 - Years: 1900 - 2023
 - Update Frequency: Weekly



Terminology

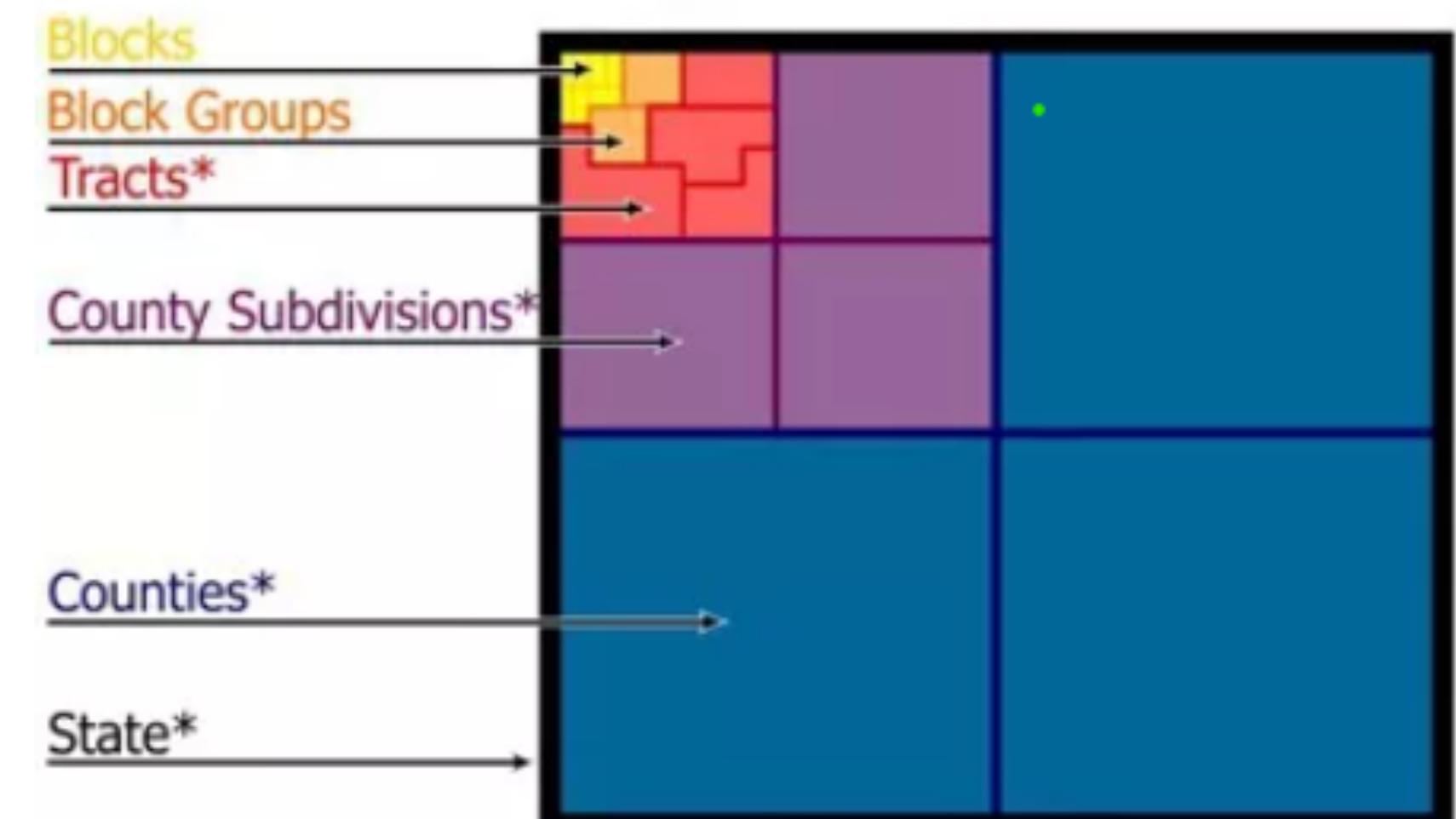
Block - smallest geographic area in a census dataset.

Block groups - A cluster of blocks in the same census tract

Census tract - subdivision with population > 1200 and <8000

County - Primary legal division within a state.

Census Summary Levels:



*Included with Maptitude. Blocks and block groups available separately.

Source

Census Demographic Data

- **Demographics gathered:** Age, income, population.
- **Source** - American Community Survey 5 year estimate datasets.
- **Granularity level:** block group
- **Years:** 2013-21.
- **Update Frequency:** Annual

Census Demographic Data (contd.)

- Merged with census TIGER datasets.
- Calculated weighted age and income.
- Combined datasets for all years and boroughs.

Health Indicators

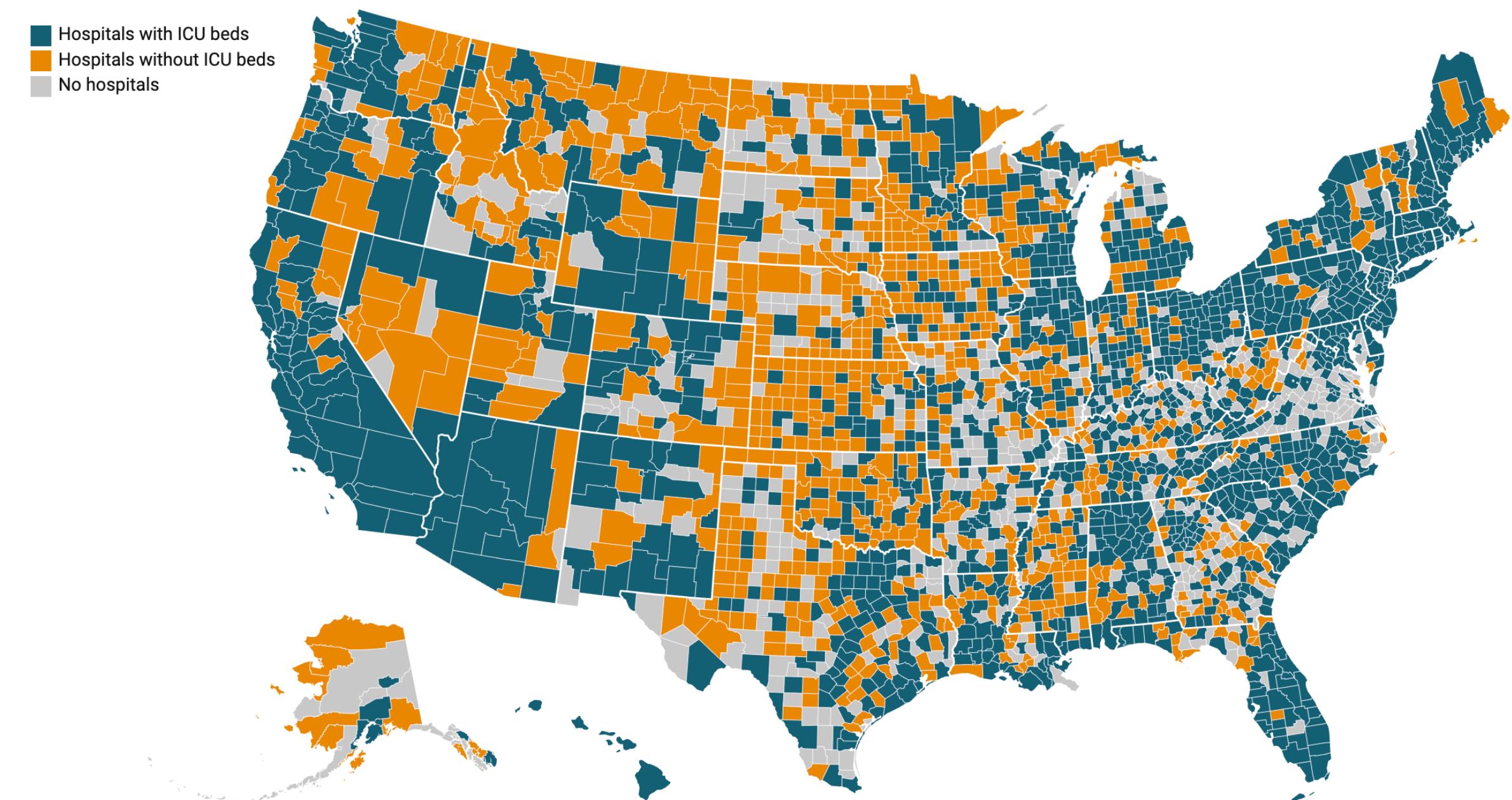
- How do we measure the requirement of healthcare services in an area?
- Examples of Health Indicators:
 - Mortality rates
 - Respiratory Diseases
 - Cardiovascular Diseases
 - Infectious Diseases
- Most relevant factors should be location sensitive

Health Indicators: Data Sources Explored

- **Data World:** Health status Indicators
- **NY Government Health Data:** NYC Vital Statistics
- **Community Health Indicator Reports (CHIRS):** Heart Attack Hospitalization Rate Data
- **nyc.gov:** COVID-19 Hospitalization Rate Data

Number of Beds

- Why are number of beds important?
- Utilization of Bed Count Data in model



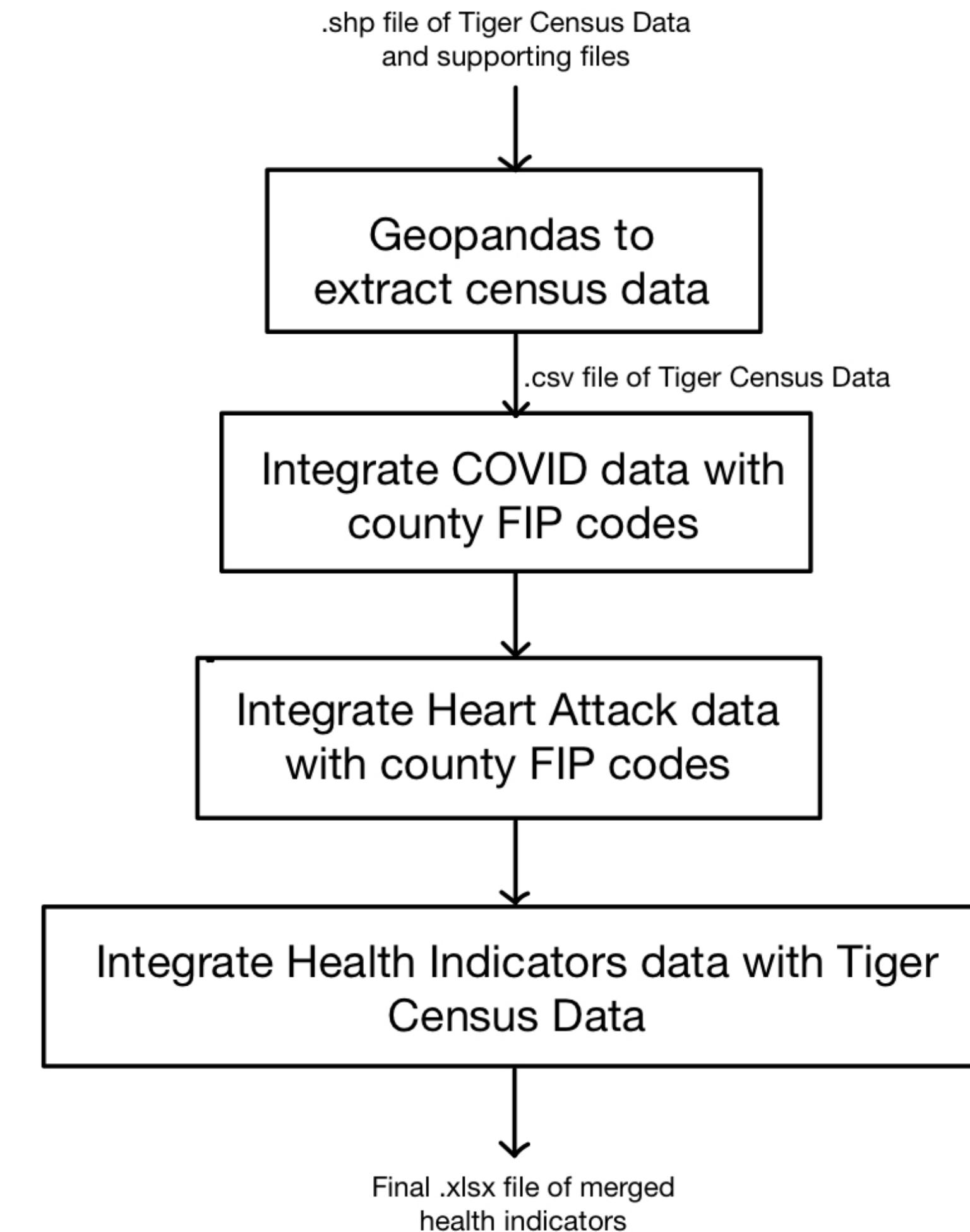
<https://kffhealthnews.org/news/as-coronavirus-spreads-widely-millions-of-older-americans-live-in-counties-with-no-icu-beds/>

Number of Beds: Sources of Data Explored

- **OpenICPSR:** National Neighborhood Data Archive
- **Data World:** Hospital Capacity Data from Harvard Global Health Institute (HGHI)
- **Homeland Infrastructure Foundation - Level Data:** Hospital Data
- **ArcGIS REST Services Directory**
- **Trading Economics:** United States Hospital Beds

Integration of Data

- Location Data Extraction
- COVID Data Integration
- Heart Attack Data Integration
- Final Integration to model



Modeling

The Model

$$\begin{aligned}\mathcal{M}(\text{block_id}) = & l_1 \cdot \frac{1}{\min(x^2, r^2)} \\& + l_2 \cdot f(\overline{\text{age}}) \\& + l_3 \cdot i(\overline{\text{income}}) \\& + l_4 \cdot g(\overline{\text{hosp_rate}}) \\& + l_5 \cdot h(\overline{\text{cardio_cases}}) \\& - \lambda (\#\text{hospitals inside } \min(x^2, r^2))\end{aligned}$$

Parameter Functions

$$f(\text{age}) = (-2.31 \cdot \text{age}) + 0.025 \cdot (\text{age}^2) + 95.608$$

$$g(\text{covid_cases}) = \text{covid_cases}$$

$$h(\text{covid_hosprate}) = \text{covid_hosprate}$$

Placement Selection

$$\text{block_id} \sim \text{Uniform}\left(\text{blocks_ids} \mid \mathcal{M}(\text{block_id}) > 0.8\right)$$

Loss Functions

Loss Functions — Decay

$$\mathcal{L}_{\text{decay}}(y_{\text{true}}, y_{\text{pred}}, \Delta t) = \frac{1}{N} \sum_{i=1}^N \min_{\substack{j=1 \\ t_j \in [t_i - \Delta t, t_i + \Delta t]}}^M d(\text{block}_i, \text{hospital}_j)$$

Loss Functions — Minimum

$$\mathcal{L}_{\text{no_time}}(y_{\text{true}}, y_{\text{pred}}) = \frac{1}{N} \sum_{i=1}^N \min_{j=1}^M d(\text{block}_i, \text{hospital}_j)$$

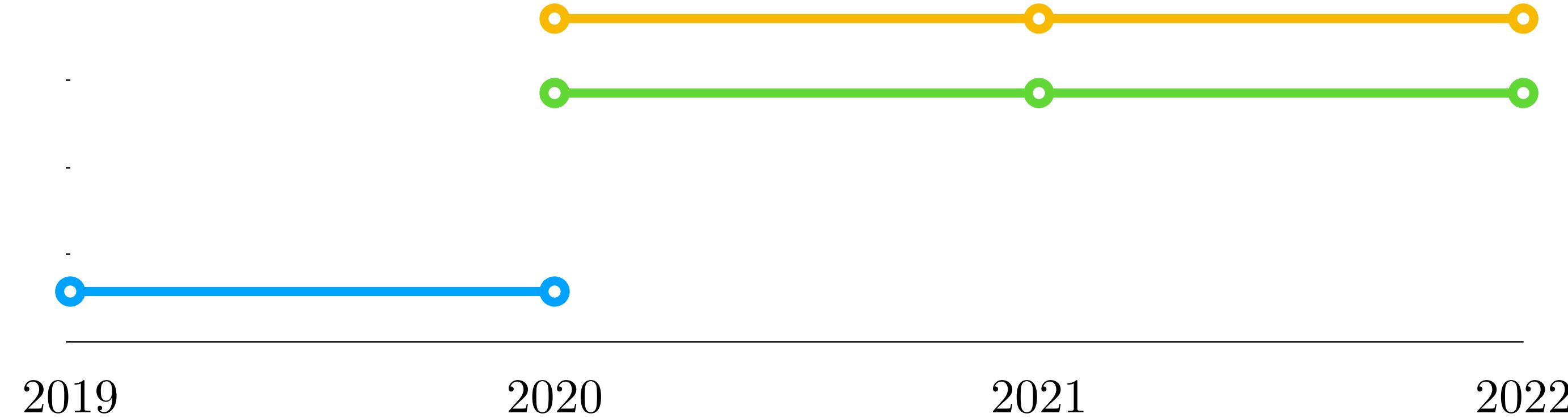
Loss Functions — corresponding

$$\mathcal{L}_{\text{corresponding}}(y_{\text{true}}, y_{\text{pred}}) = \frac{1}{N} \sum_{i=1}^N d(\text{block}_i, \text{hospital}_{\pi(i)})$$

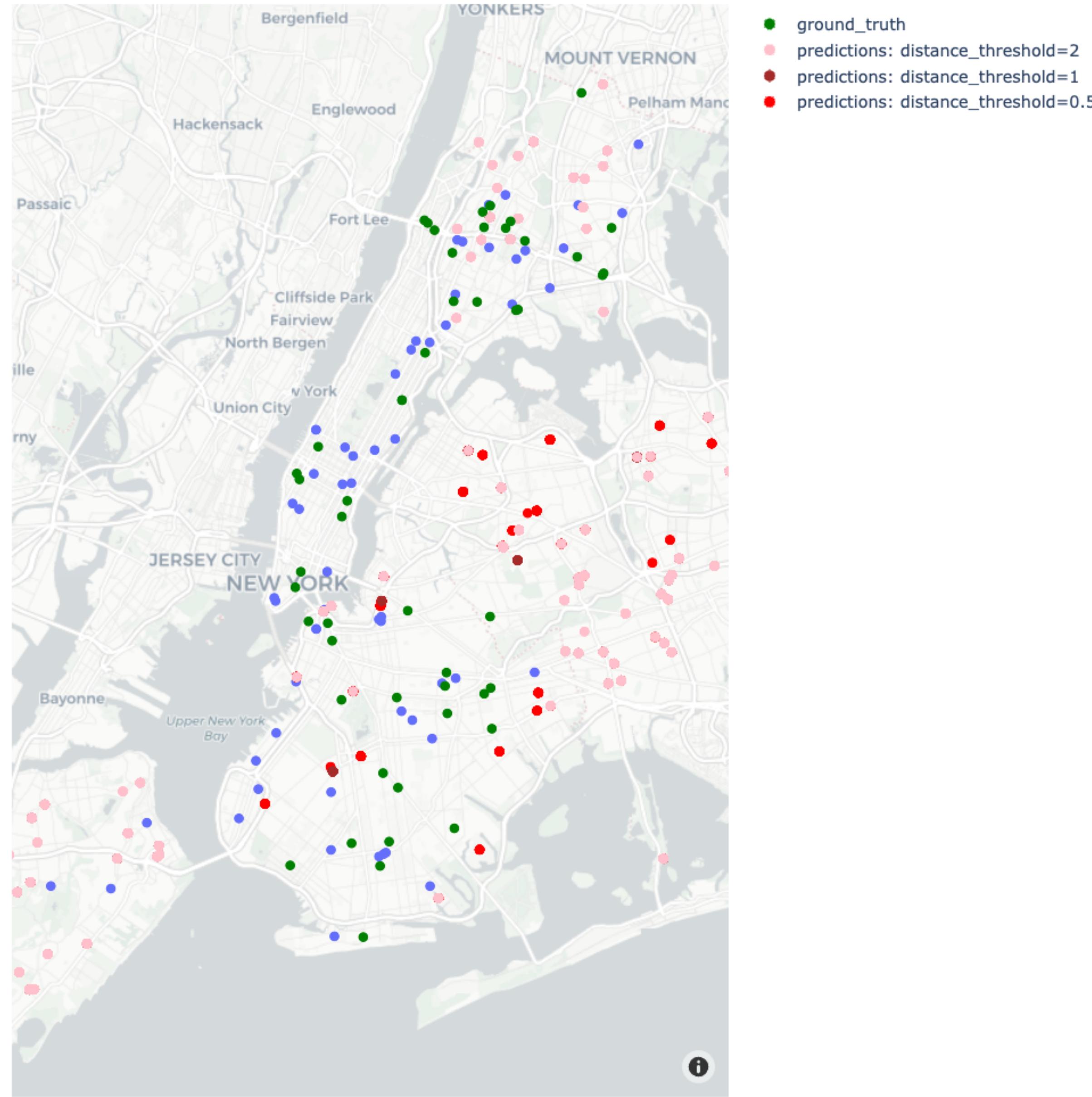
Results

The Simulation

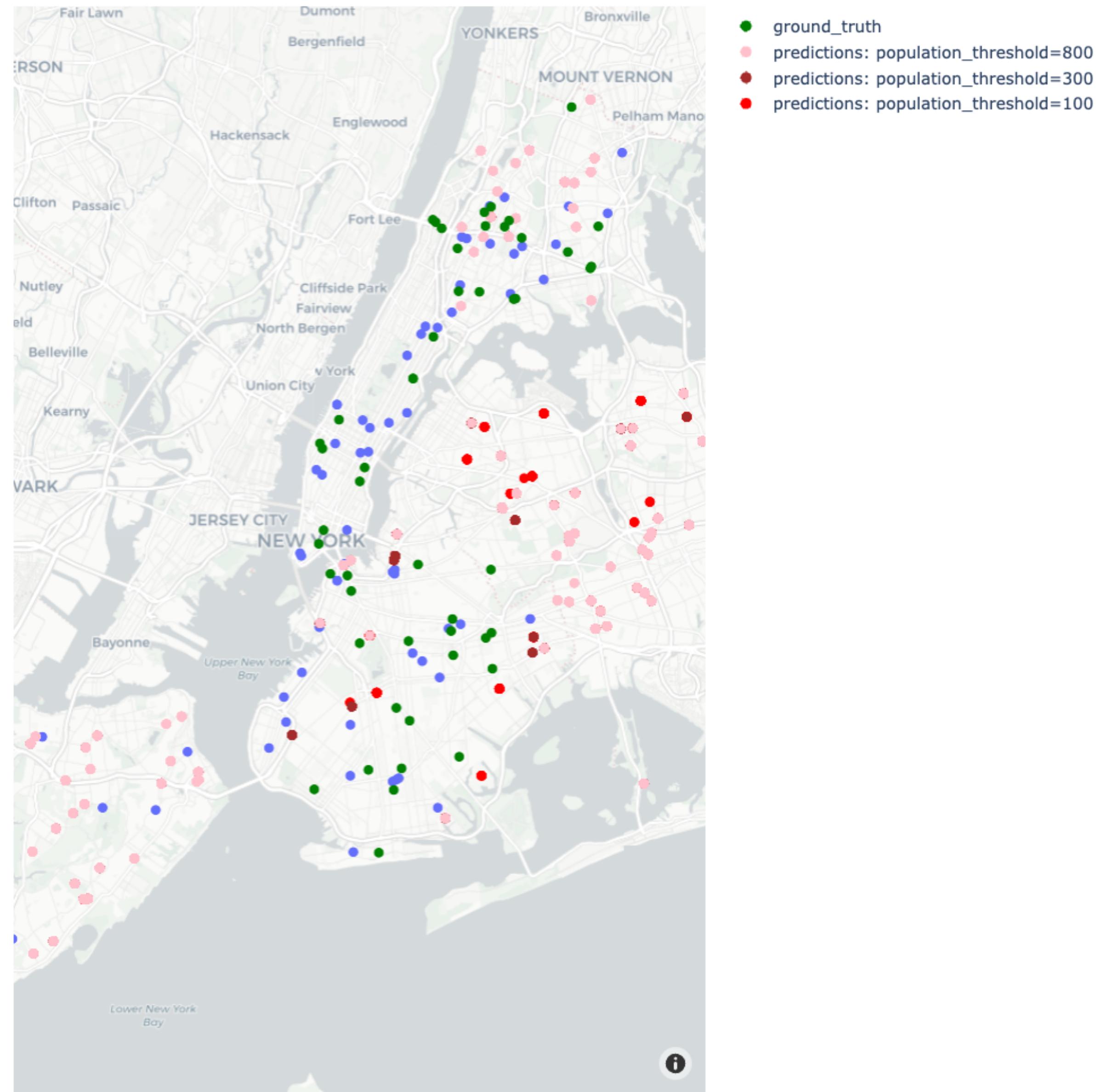
- Pre-existing
- Simulation: ground truth
- Simulation: predictions



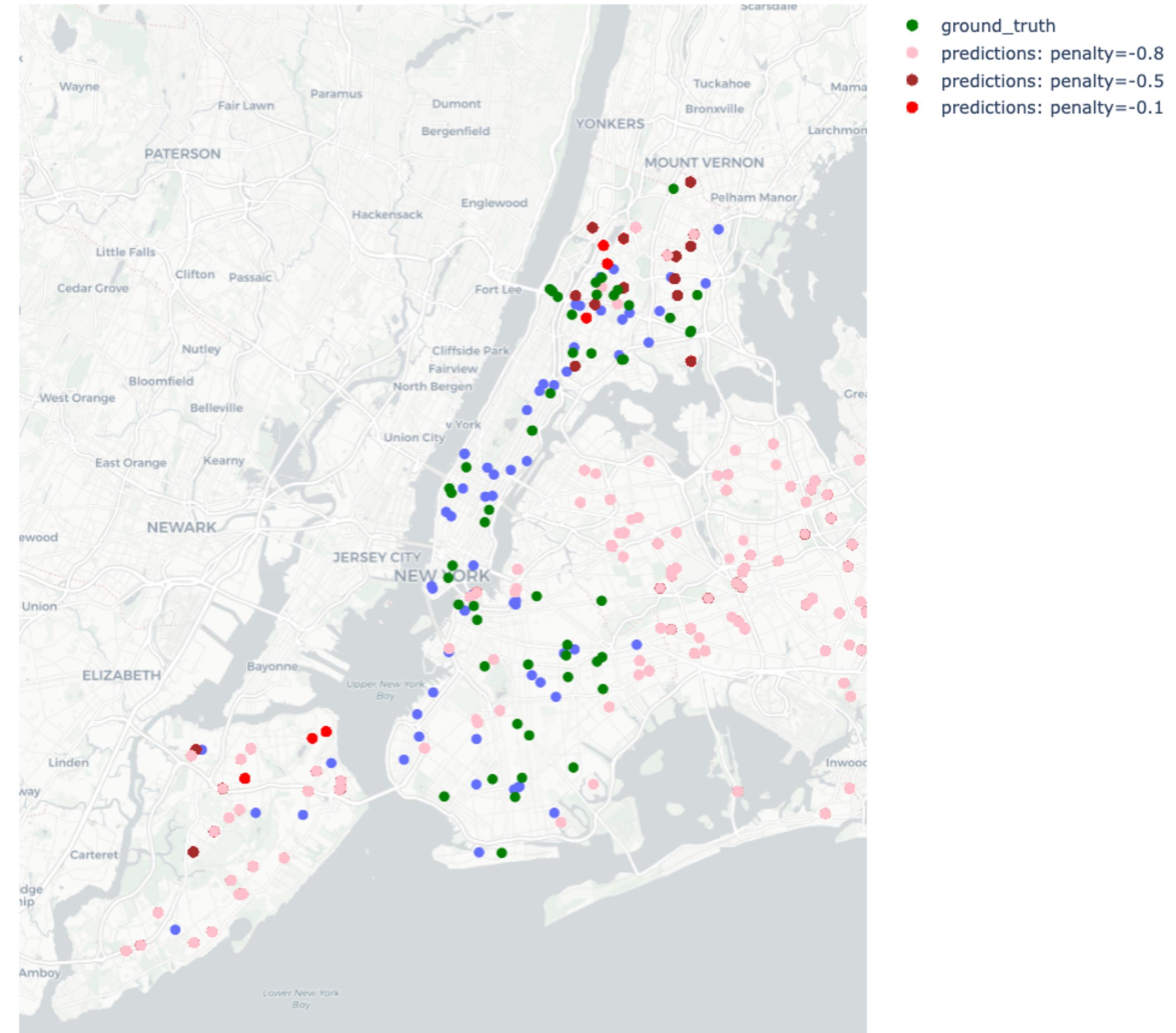
Varying *distance_threshold*



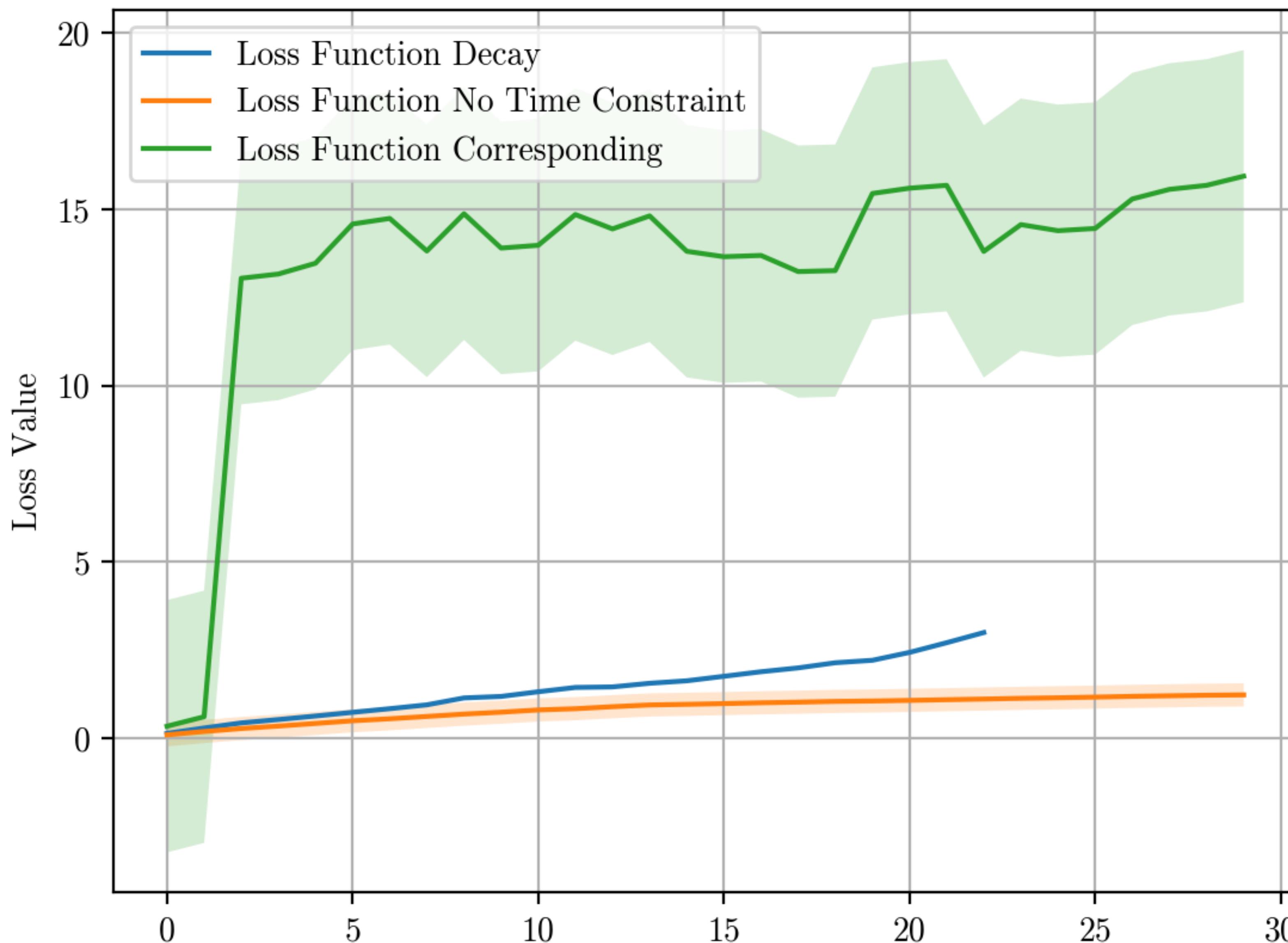
Varying *population_threshold*



Varying penalty (λ)



Loss Curves



Challenges

- Finding datasets at the right geographic level
- Datasets had missing years
- Data Quality and Consistency Issues
- Age and income data was an estimate - getting actual figures is challenging.

Future Work

- Add additional parameters
- Diversify: multiple kinds of hospitals
- Build a UI tool
- Integrate with the census data API.

References

- Number of Beds Data Sources:
 - OpenICPSR : [URL](#)
 - Data World: [URL](#)
 - Homeland Infrastructure Foundation: [URL](#)
 - ArcGIS Rest: [URL](#)
 - Trading Economics: [URL](#)
- Hospital Location Data: [URL](#)
- Health Indicator Data Sources:
 - Data World: [URL](#)
 - NY Government Data: [URL](#)
 - Heart Attack Hospitalization Rate Data: [URL](#)
 - COVID Hospitalization Rate Data: [URL](#)
- Demographic Data Sources:
 - ACS Sex By Age Datasets: [URL](#)
 - ACS Median Income Datasets: [URL](#)

Github: Polaris000/UrbanHealthOpt

