Tracking Citi Bike reliability & measuring service inequity using real-time data

Dan Levine

Office of NYC Comptroller Brad Lander

https://github.com/NYCComptroller/citi-bike-gbfs

nycsodata24.sched.com #nycSOdata #opendataweek

NYC SCHOOL OF DATA

NEW YORK CITY COMPTROLLER
BRAD LANDER

Tracking Citi Bike reliability & measuring service inequity using real-time data

School of Data Presentation

MARCH 2024

https://github.com/NYCComptroller/citi-bike-gbfs

Agenda

- Key findings and recommendations
- Data
- Analysis methods
- Hands-on how-to

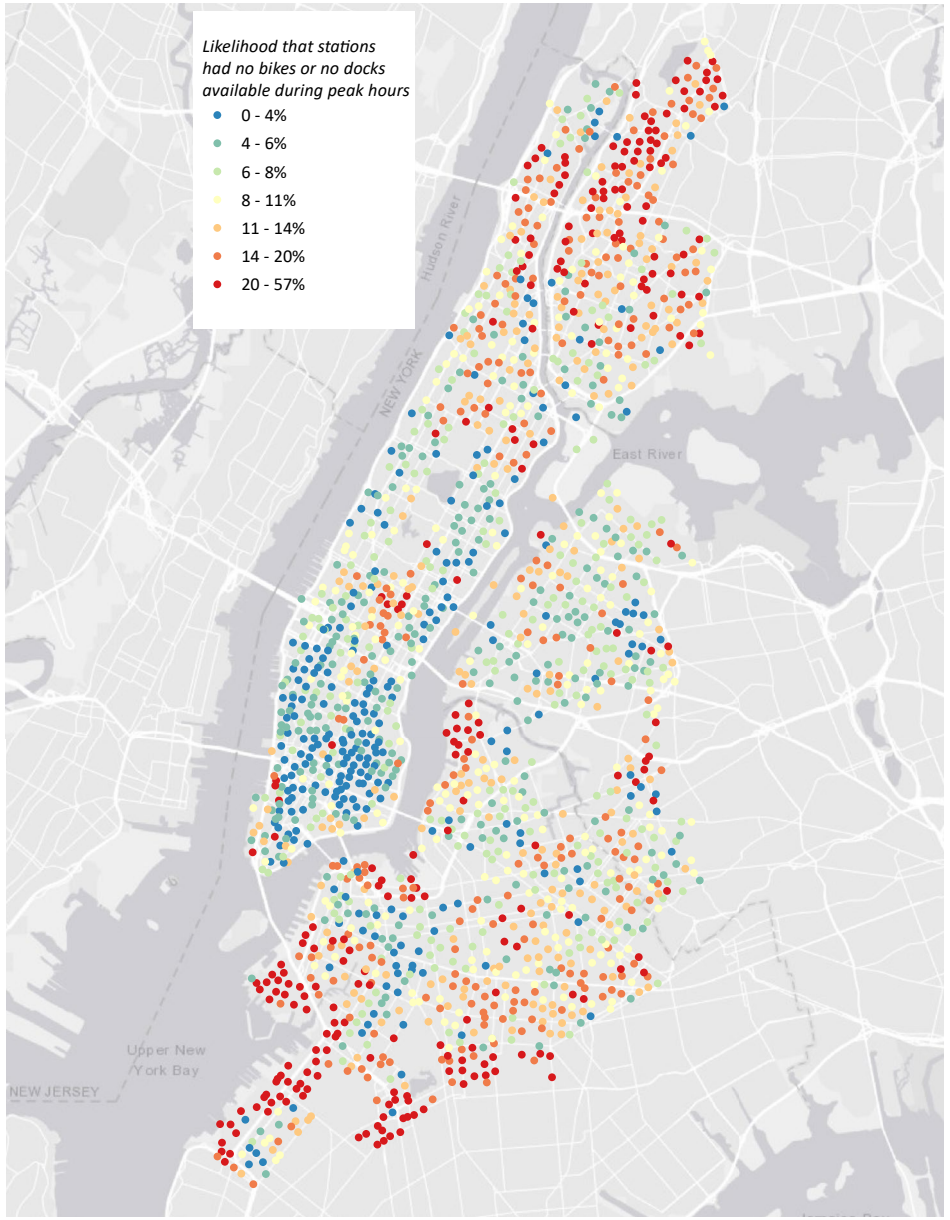
Problem: unreliable Citi Bike

Key findings & policy recommendations

Photo: Elzbieta Sekowska / Shutterstock.com

Key findings

- Citi Bike is **not consistently available** in all neighborhoods
- Least reliable service in areas home to **more Hispanic, Black, and low-income residents**
- While Citi Bike has **cut back** on rebalancing



Policy recommendations

- Create **neighborhood-level** performance standards
- Strengthen **enforcement** of basic performance standards
- Incentivize better performance by **offering payments or subsidies** when the operator consistently **exceeds** service standards
- **Improve transparency** through enhanced public reporting

HOW WE DID IT:

Data

Photo: Elzbieta Sekowska / Shutterstock.com

Needed more granular data

citi bike®
provided by lyft

April 2022 Monthly Report

SLA 6 – Station Deactivation, De-Installation, Re-Installation, and Adjustment

Description: As directed by DOT, Citi Bike must perform: (i) Station Deactivation(s); (ii) Station De-Installation(s); (iii) Station Re-Installation(s); (iv) Station Adjustment(s).
DOT will provide a minimum of 48 hours advance notice prior to any of the above, except in instances where the continued presence/activity of the Station has been determined to pose a threat to public safety. Deactivated Stations must be reactivated within 24 hours of direction from DOT. De-Installed or Adjusted Stations must be reinstalled or Readjusted to their original configurations within 72 hours of direction from DOT.

Performance Rate: 100%

SLA 7 – Snow Removal

Description: Following snow events, Citi Bike must remove snow within 12 hours so as to maintain:
(i) Parallel pedestrian clear path adjacent to Stations located on sidewalks and in plazas; and
(ii) Perpendicular pedestrian paths through Stations where gaps in Docks provide pedestrian access.

Performance Rate: 100%

SLA 8 – Program Functionality

Description: The Program is completely unavailable, such that no Program user can dock, undock, and Wrench Bicycles as intended, exclusive of planned Program outages for upgrades and maintenance as agreed upon by Citi Bike and DOT and Program outages caused by an Event of Force Majeure.
Performance Rate: 100%

SLA 9 – Bicycle Availability

Description: This Bicycle Availability requirement is met if the monthly average Bicycle Fleet Level, recorded once each Day of the month between the hours of 11:00 AM and 3:00 PM, meets or exceeds the required Bicycle Fleet Level.

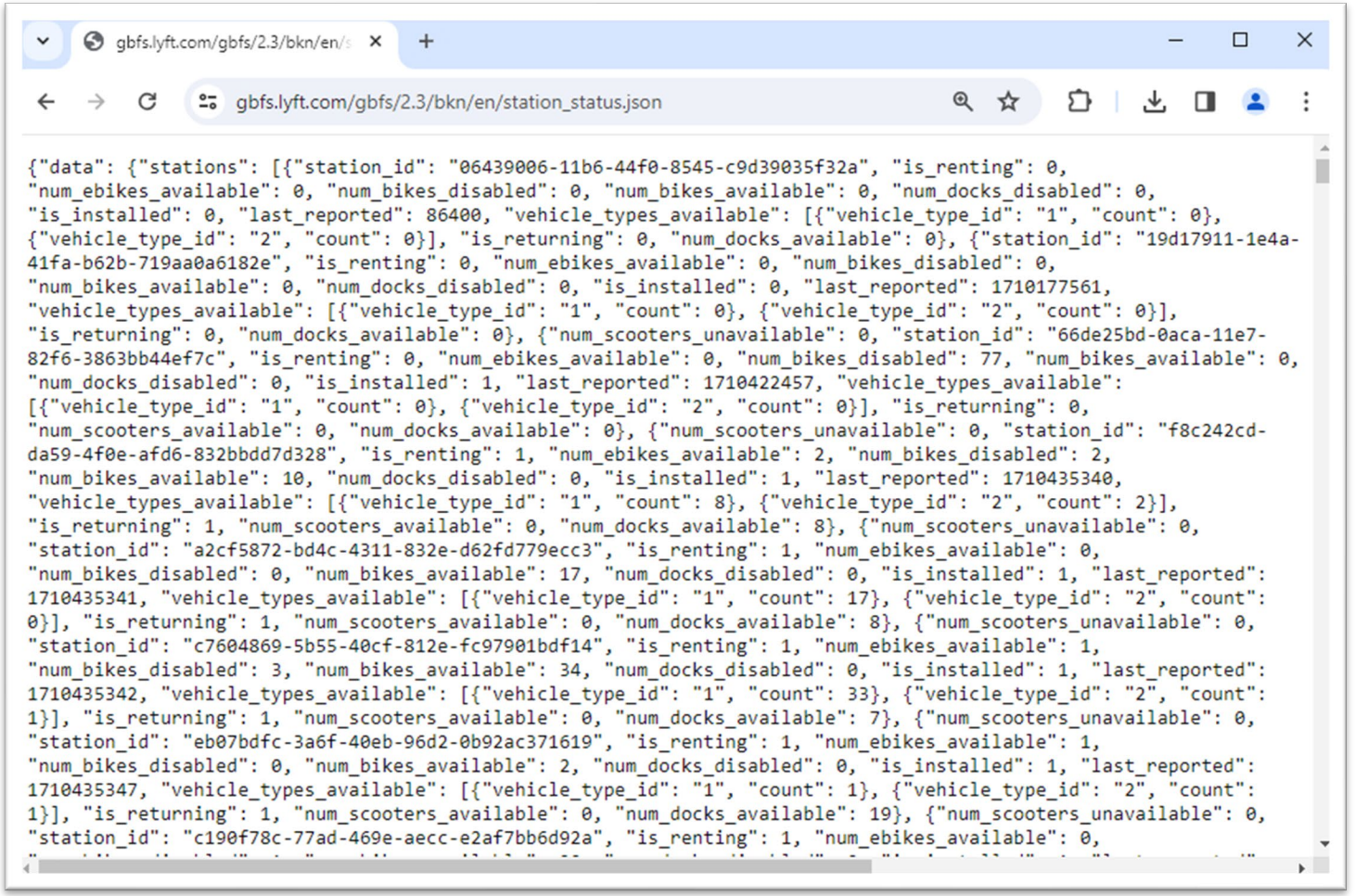
Performance Rate: 100%

The Citi Bike program is operated by NYC Bike Share, LLC, a subsidiary of Lyft, Inc.

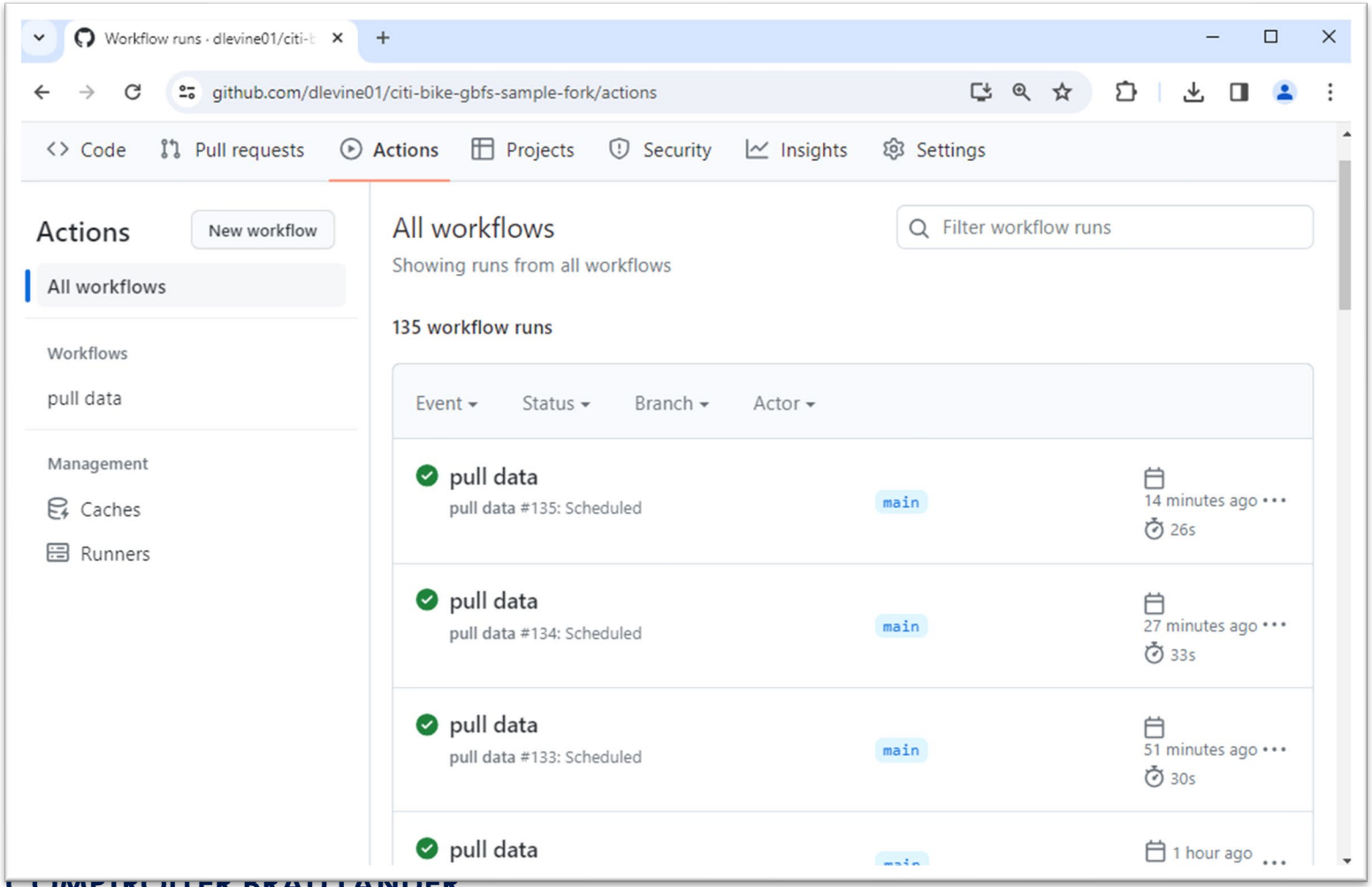
Real-time detail

Citi Bike GBFS

(General Bikeshare Feed Specification)



Recording live data



Build detailed data

last_updated	station_id	capacity	is_renting	is_returning	num_docks_available	num_bikes_available	num_ebikes_available	num_bikes_disabled
2024-03-12 17:40:49	66dc2995-0aca-11e7-82f6-3863bb44ef7c	51	0	0	0	0	0	0
	06439006-11b6-44f0-8545-c9d39035f32a	48	0	0	0	0	0	0
	19d17911-1e4a-41fa-b62b-719aa0a6182e	39	0	0	0	0	0	0
	1861678548643203686	25	1	1	8	15	8	2
	cd2d9dab-7708-4685-a56f-9412c738de7e	23	1	1	1	22	1	0
	66db2a71-0aca-11e7-82f6-3863bb44ef7c	45	1	1	0	41	16	4
	901ad0c4-383e-490a-8b54-0656ce2358d6	19	1	1	3	11	2	5
	66dc292c-0aca-11e7-82f6-3863bb44ef7c	55	1	1	50	3	1	2
	4c03fa2d-89da-4f0b-8f19-c7de5edbe256	25	1	1	3	21	3	0
	9af90faf-0b9b-451b-9cb0-20ff421ca1d9	27	1	1	18	9	1	0

HOW WE DID IT:

Analysis

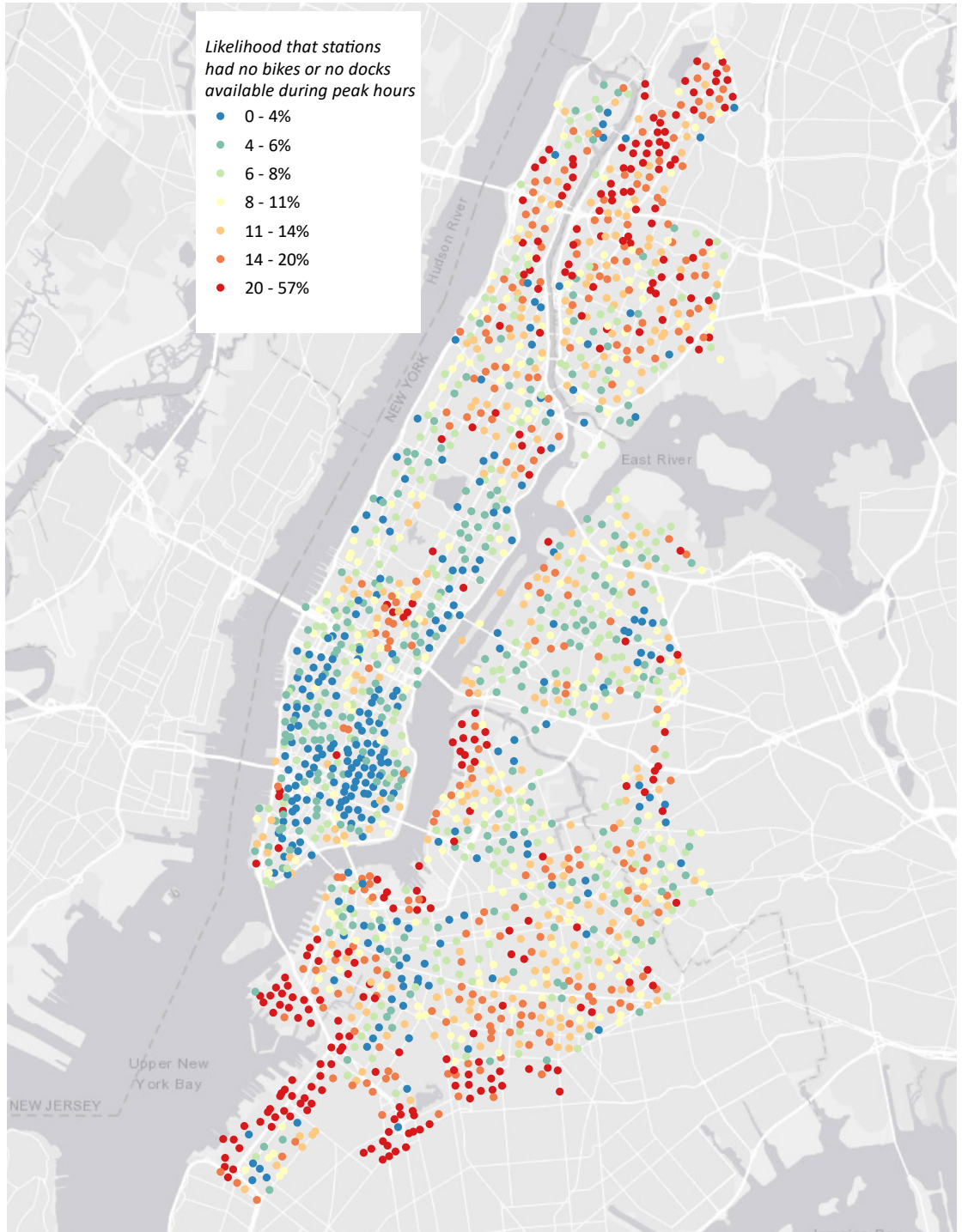
Photo: Elzbieta Sekowska / Shutterstock.com

Compute service measures

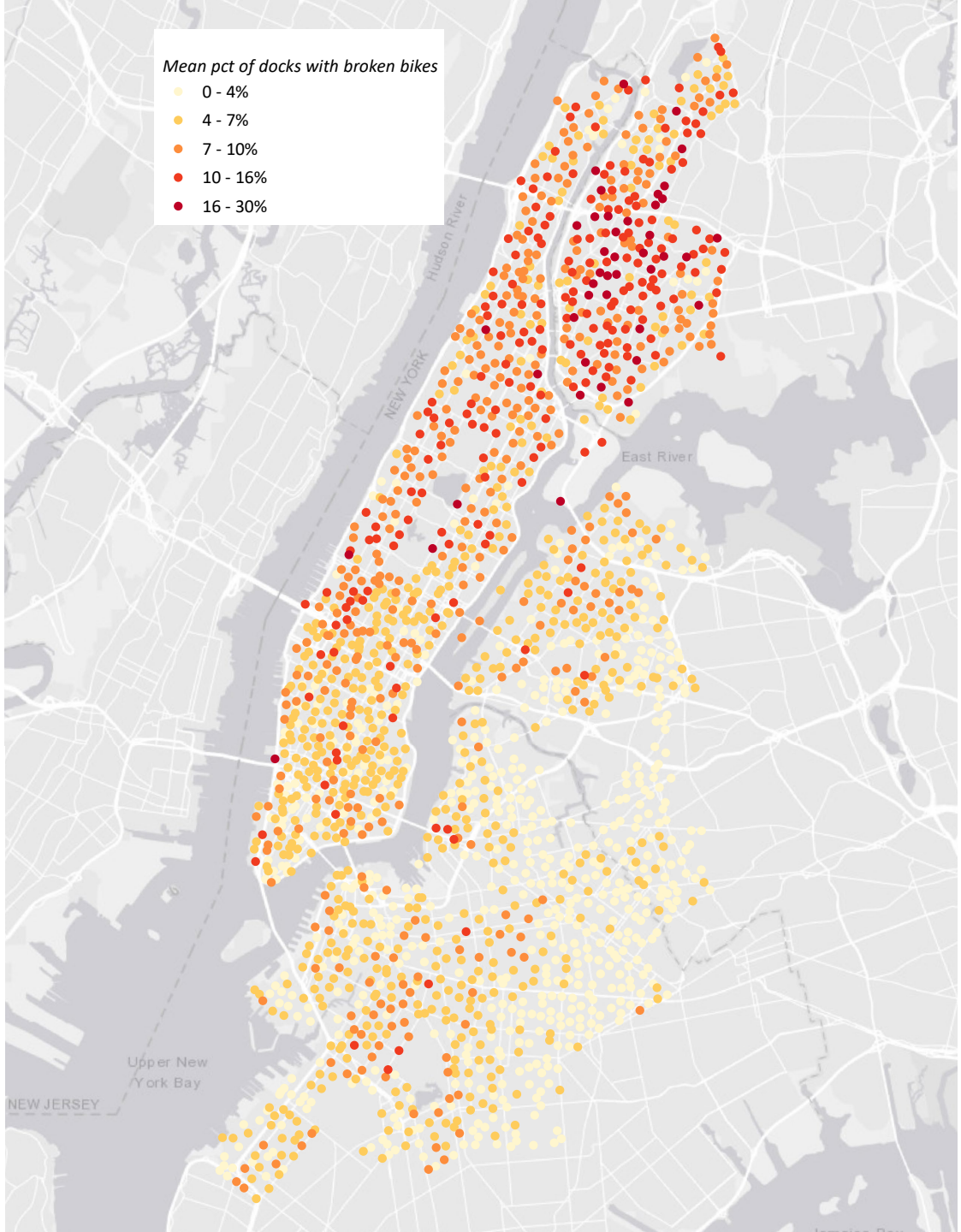
- Frequency no docks or no bikes available
- Duration no docks or no bikes available
- Portion of docks with broken bikes

Explore the data:

Unequal service between neighborhoods



Bronx riders burdened with stations loaded with broken bikes

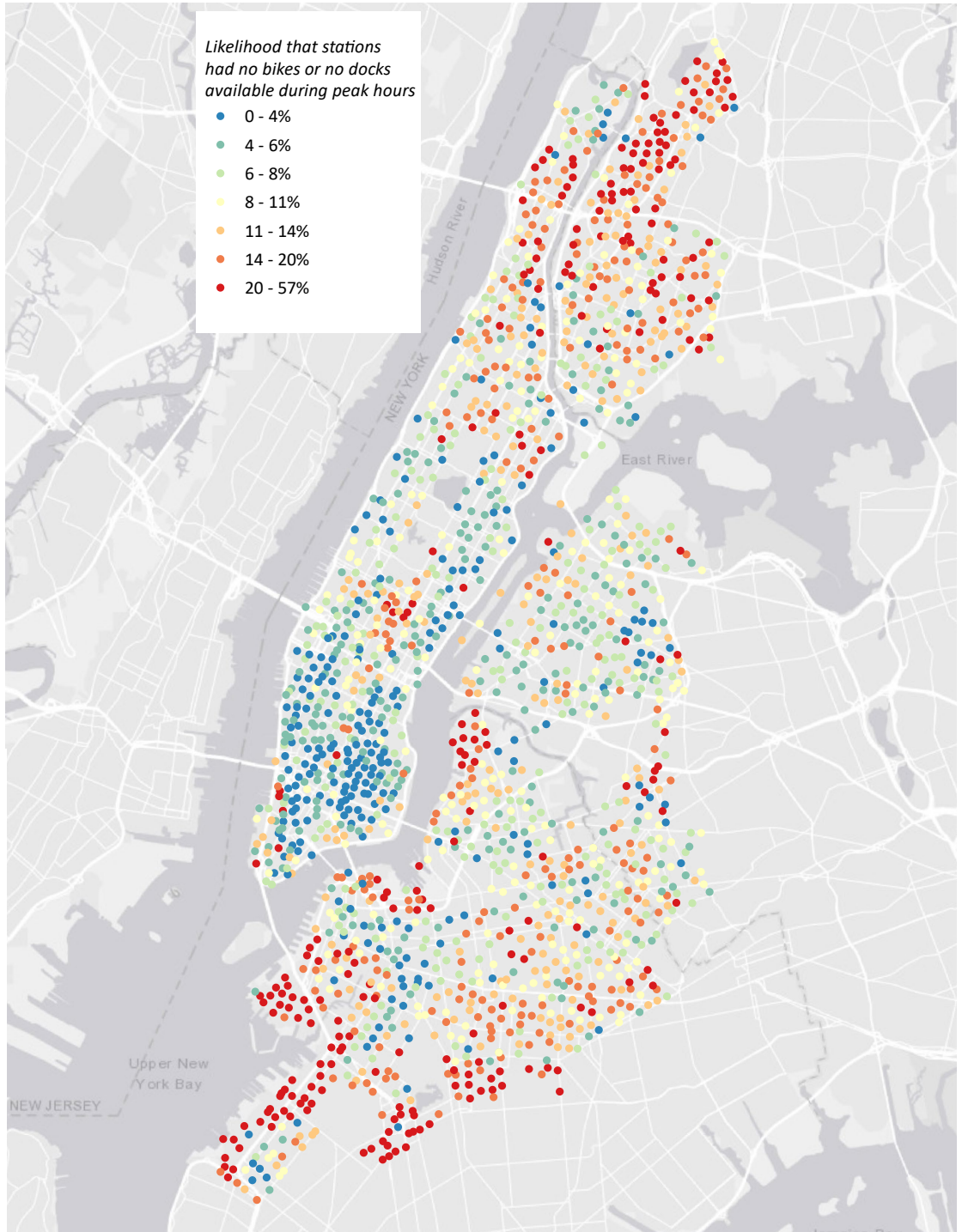


Failing to meet service standards

11,000
violations

$812,000
in potential penalties

Unequal service between neighborhoods



HOW WE DID IT:

Compute clusters of poor service

Photo: Elzbieta Sekowska / Shutterstock.com

Find clusters

Local indicators of spatial association

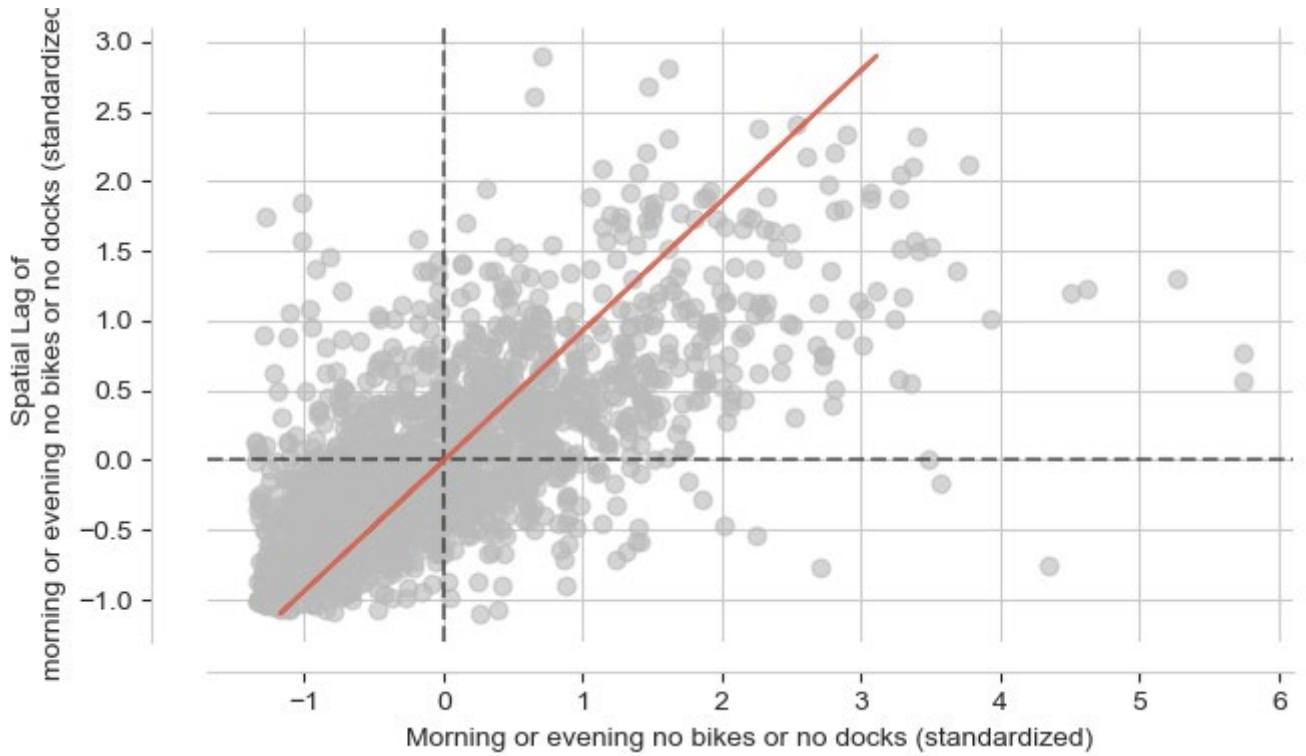
a spatial method with statistical power

Neighbors

1. Compute weighted average of nearby stations

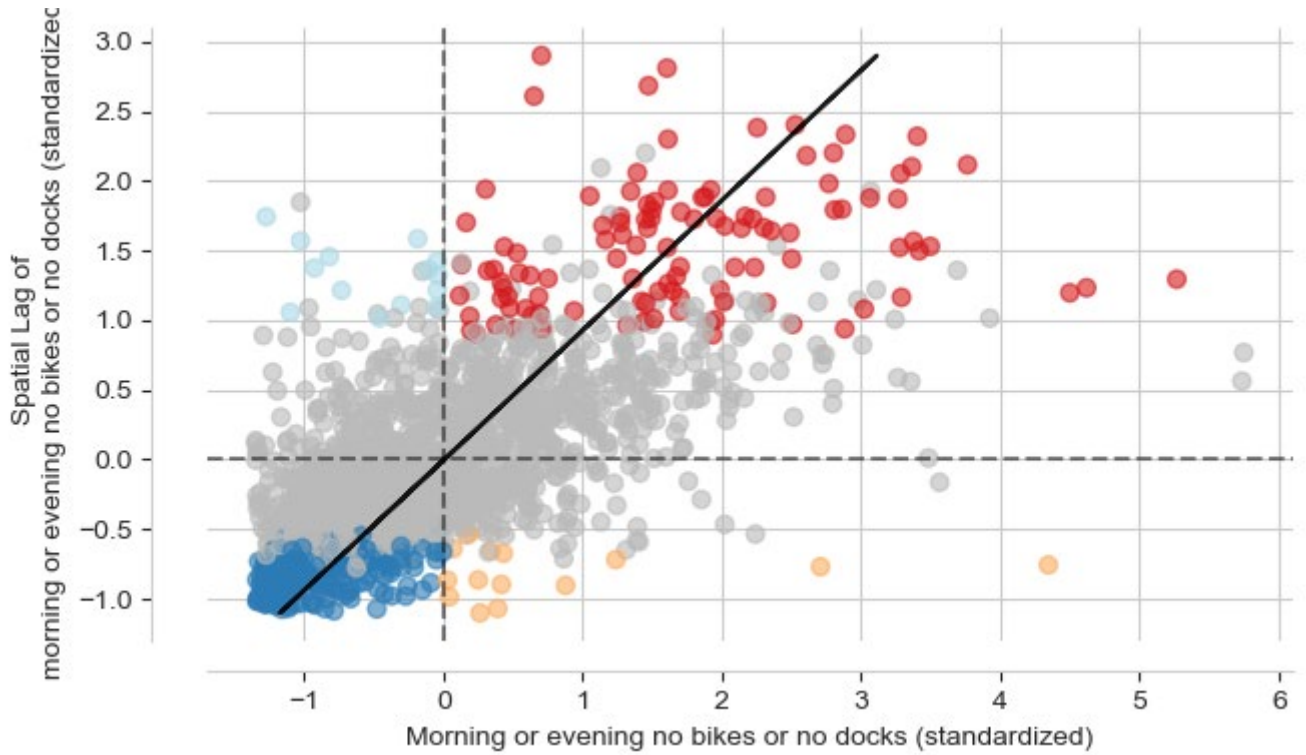
Local spatial autocorrelation

2. Compare station to its neighbors

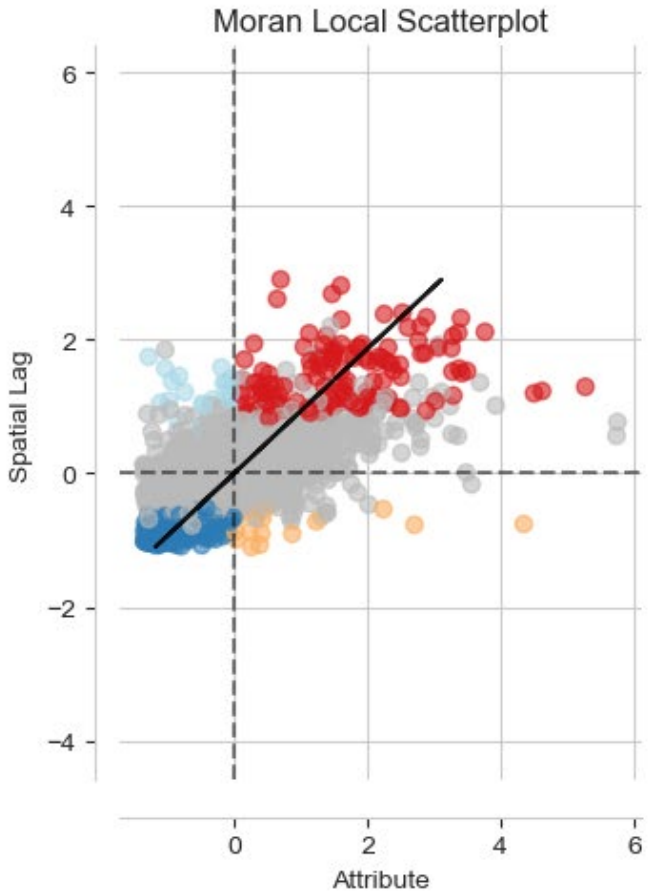


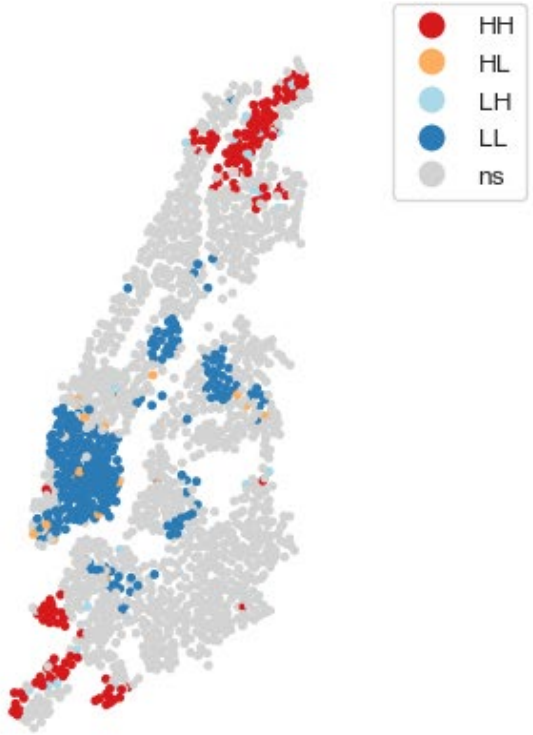
Local spatial autocorrelation

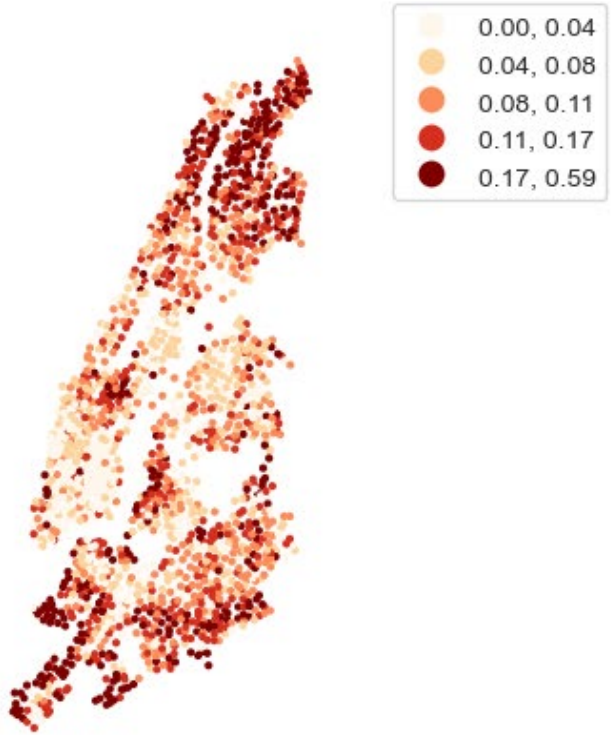
3. Shuffle arrangement to test unlikelihood *(pseudo-significance)*



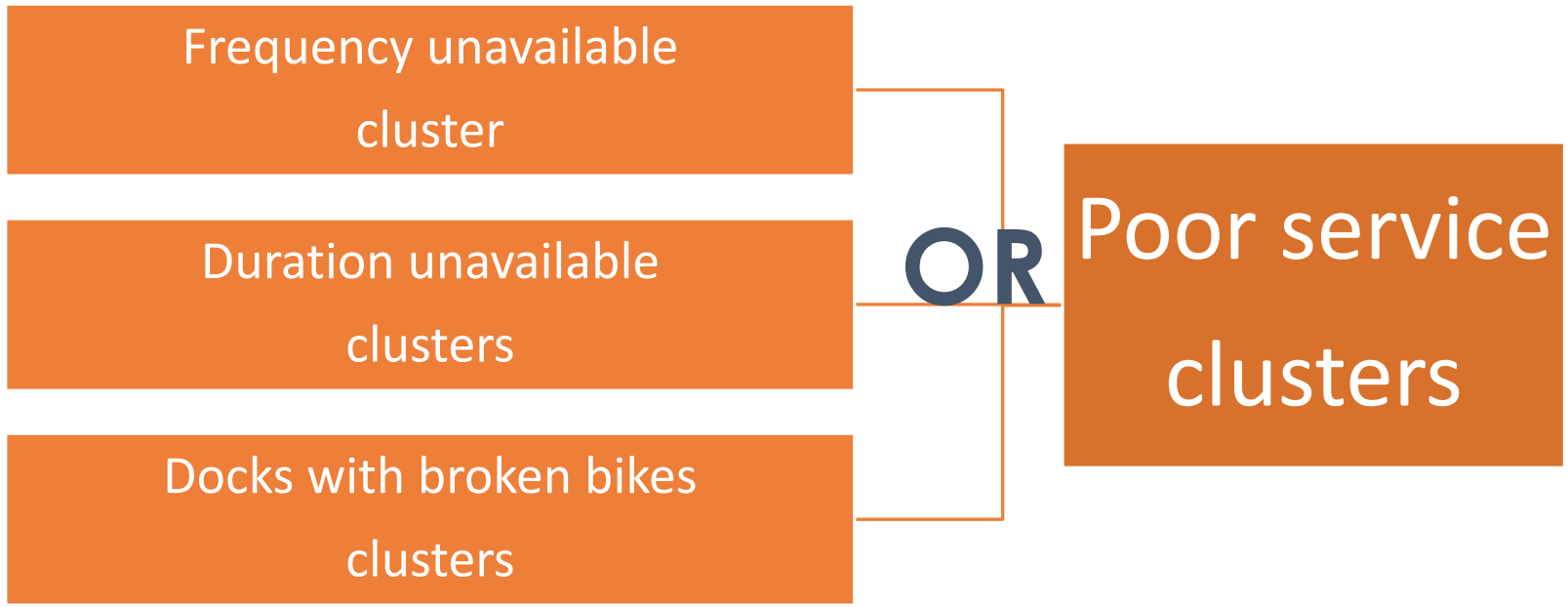
Get a measure of significant hotspots





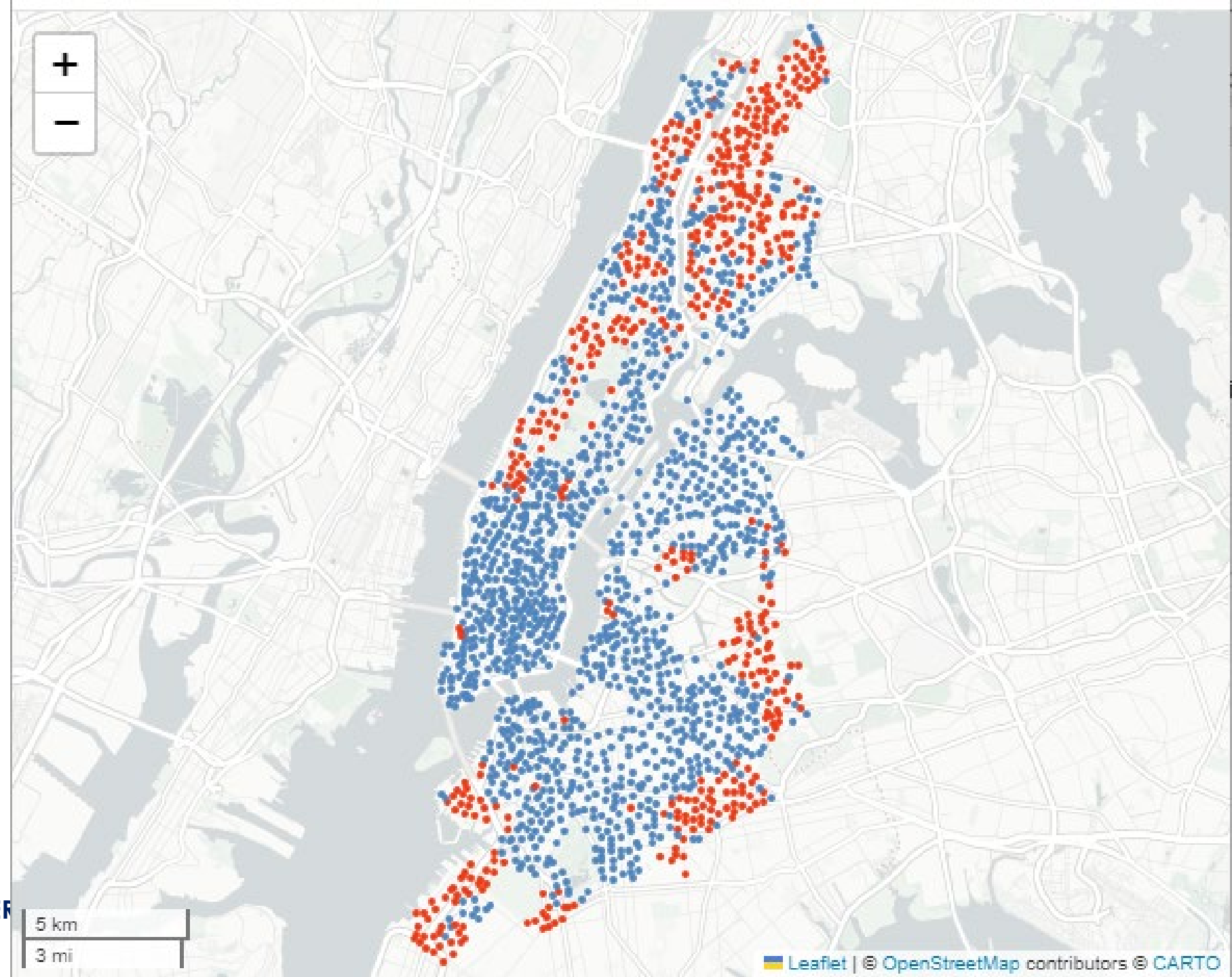


Compute hotspots across each measure



Combine

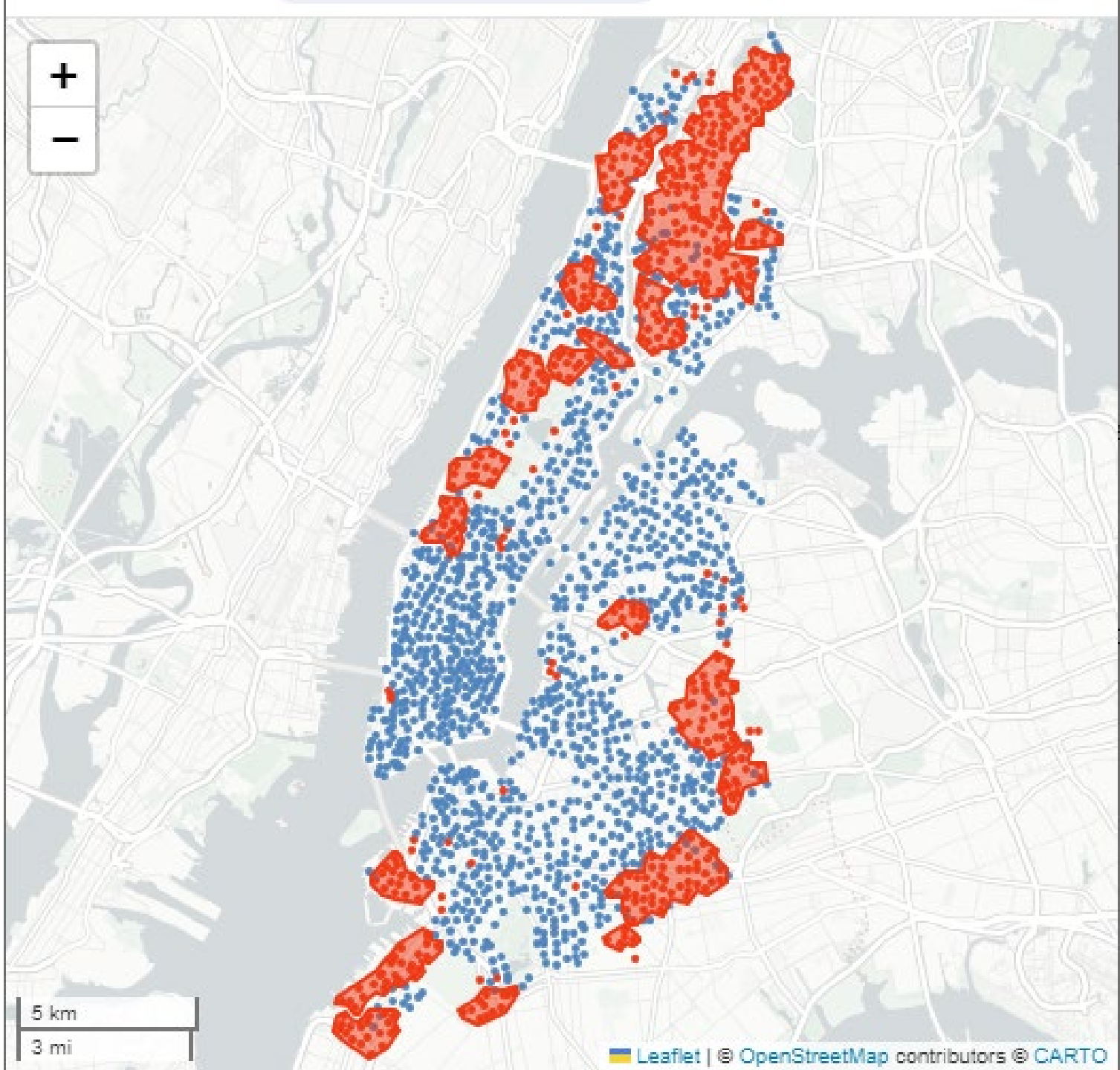
Poor performance stations



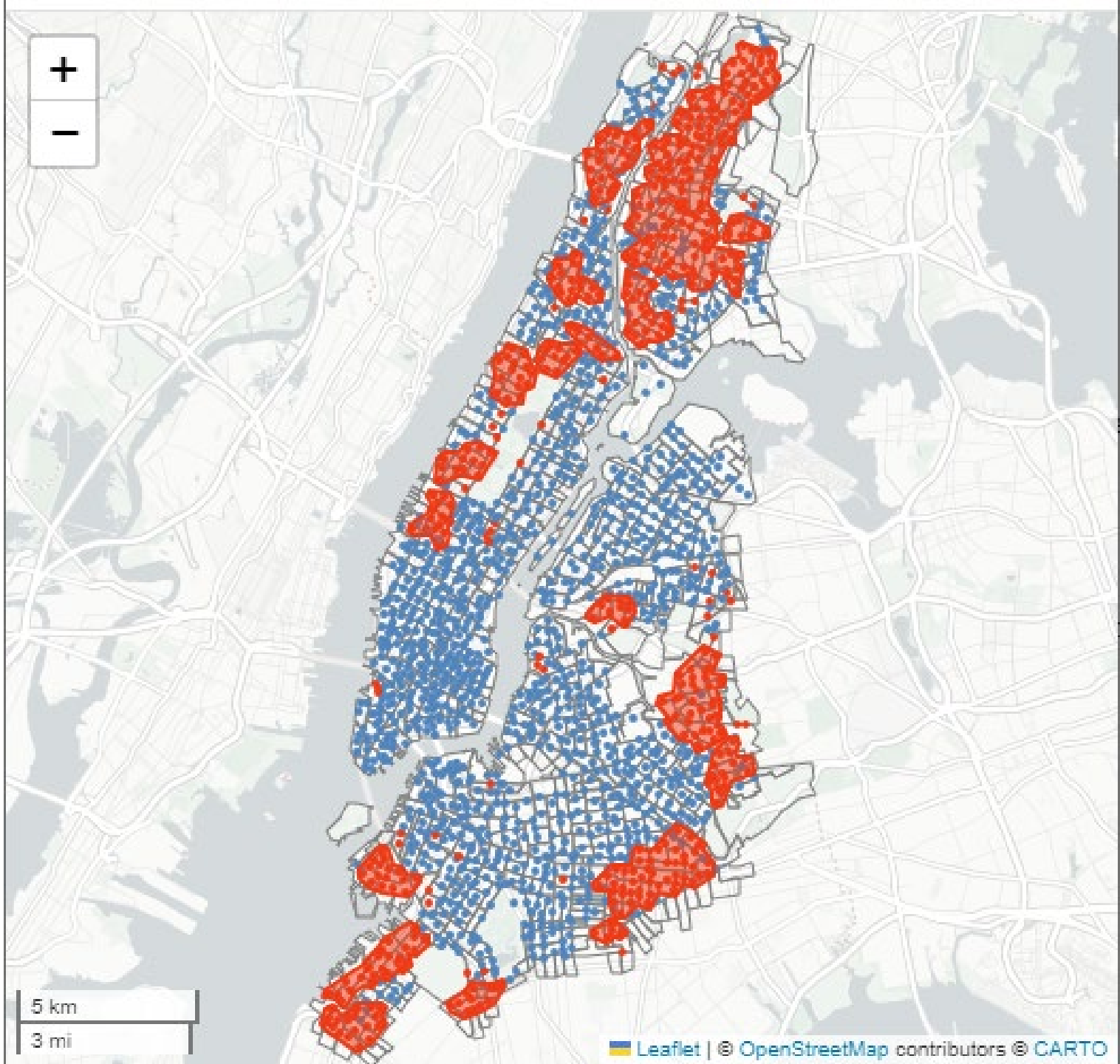
Find clusters

Keep groups with at least 5 stations within ¼ mile

Build concave hulls as boundaries

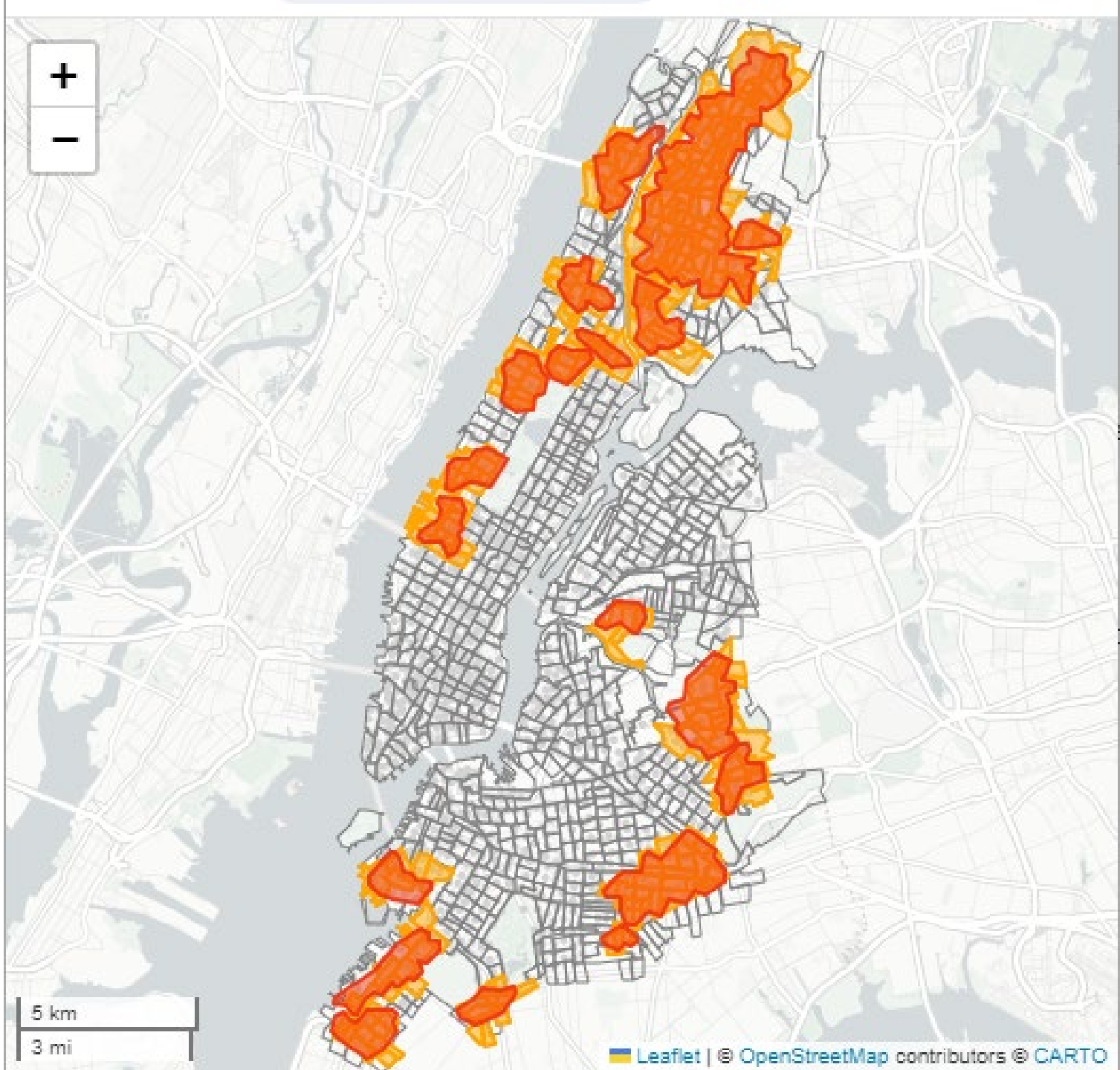


Join Tracts

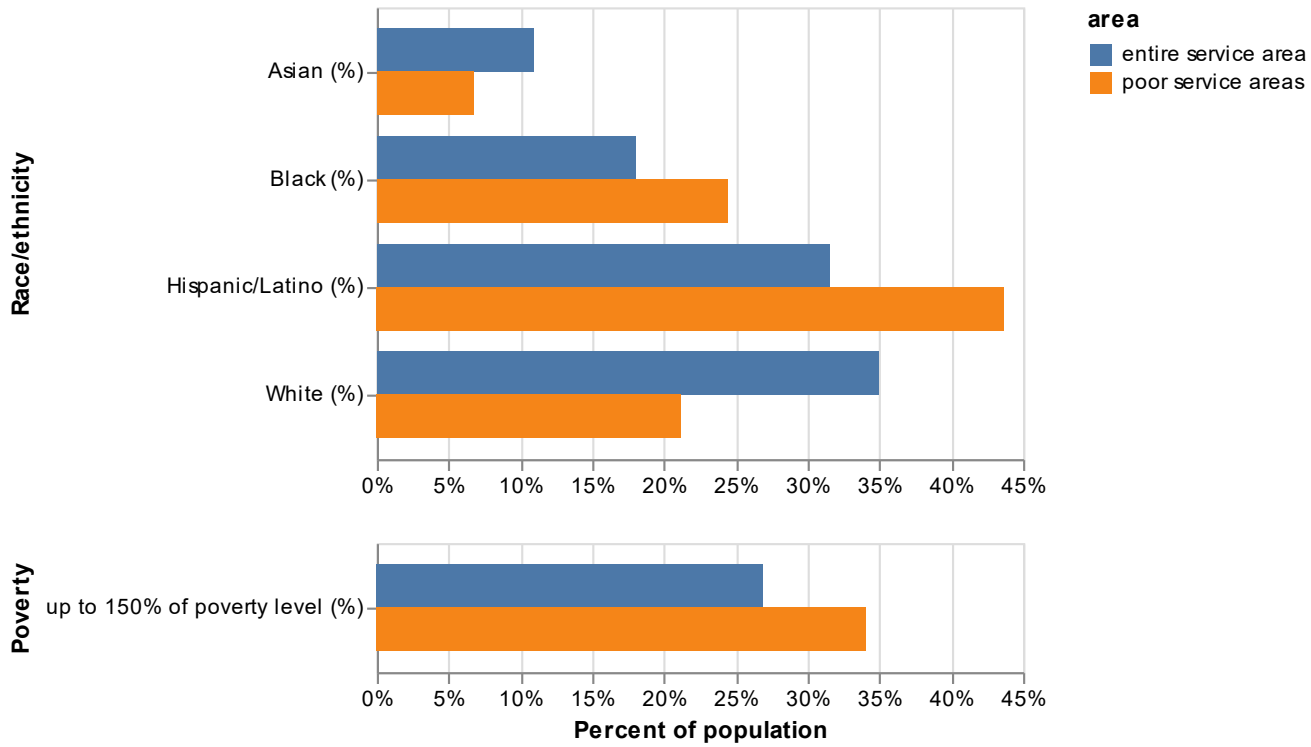


Join Tracts

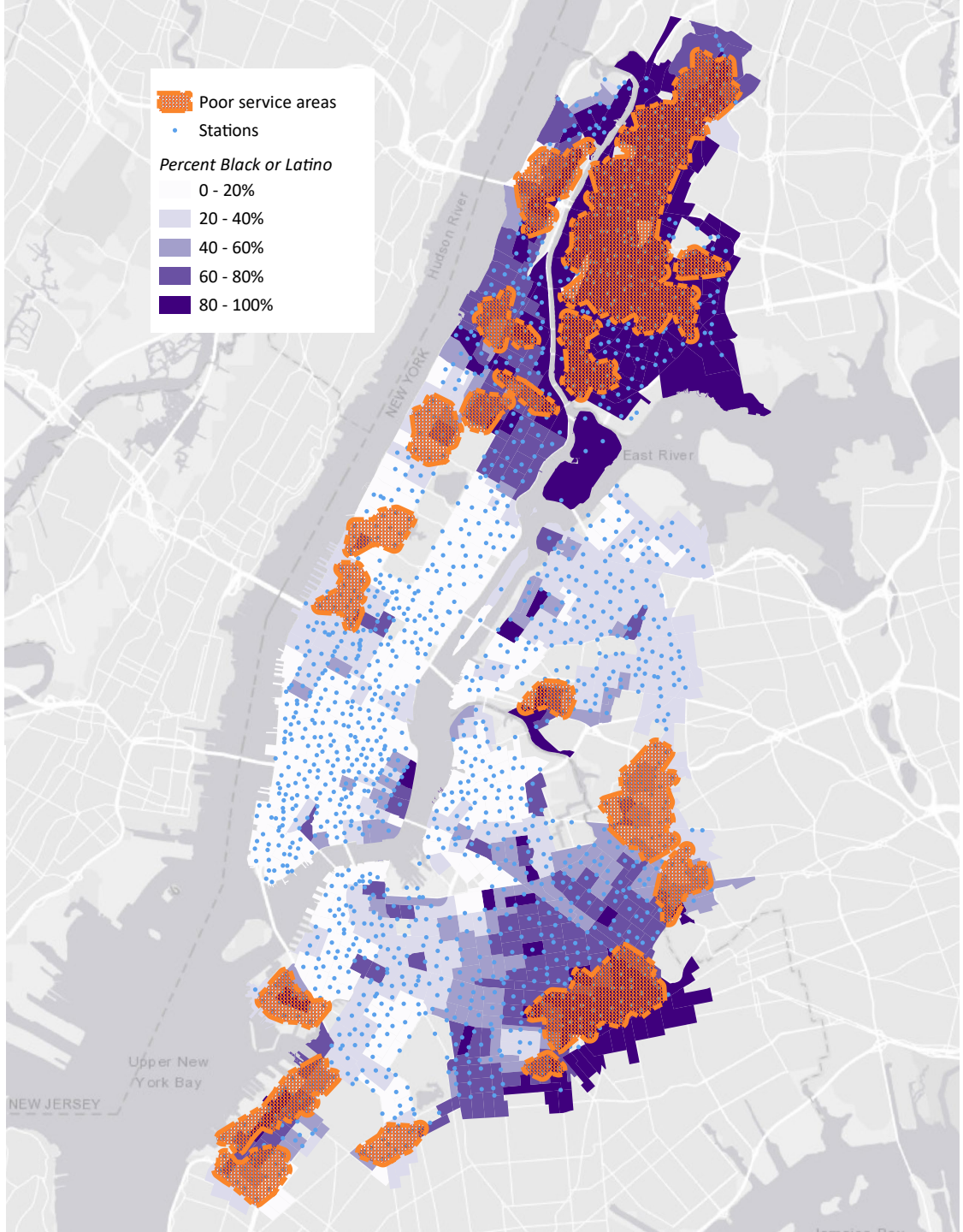
Tracts intersecting poor service clusters and entire service area



Least reliable service in areas home to more Hispanic, Black, and low-income residents



Least reliable service in areas home to more Hispanic & Black residents



Compute clusters of poor service.ipynb README.md

emplate repo for school of data > citi-bike-gbfs > Examples > Compute clusters of poor service.ipynb > compute

- Code + Markdown | Run All | Restart | Clear All Outputs | Variables | Outline ...

```
    for measure in measures:
```

But how actually?

```
            w=w_idw,
            transformation='r',
            permutations=1000,
            n_jobs=-1,
            seed=1
        )

        measure_result = (
            stations_service_measures
            [[measure]]
            .assign(
                q = measure_local_moran.q,
                p_z_sim = measure_local_moran.p_z_sim,
                p_sim = measure_local_moran.p_sim,
                significant_cluster = lambda row: (
                    row['q'].map(QUADRANT_LABELS)
                    .where(
                        (row['p_z_sim'] < alpha) &
                        (row['q'].isin([1,3]))
                    )
                ),
            )
            .rename(columns={
                'q':f'{measure}_q',
                'p_z_sim':f'{measure}_p_z_sim',
                'p_sim':f'{measure}_p_sim',
                'significant_cluster':f'{measure} significant_cluster',
```

PySAL

Python Spatial Analysis Library

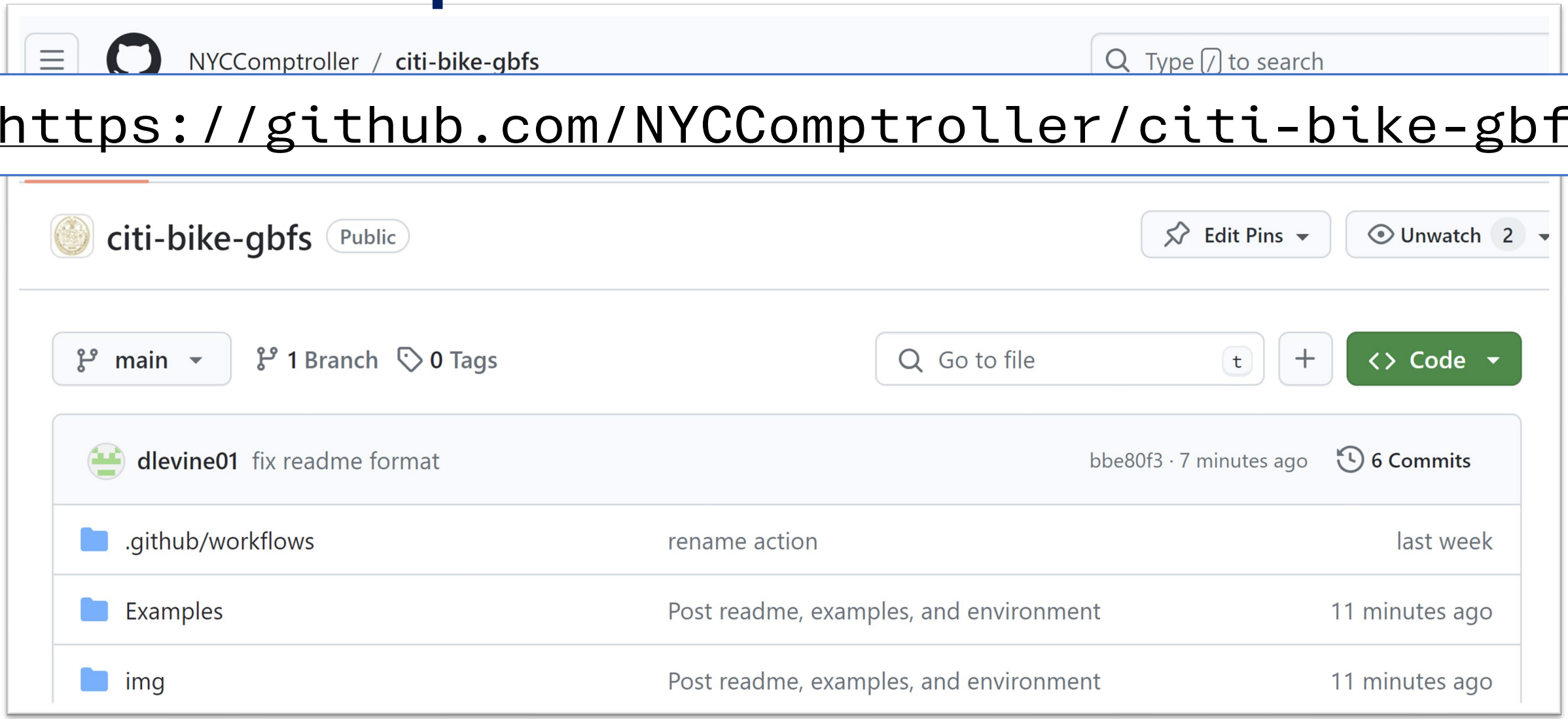
or:

Q&A

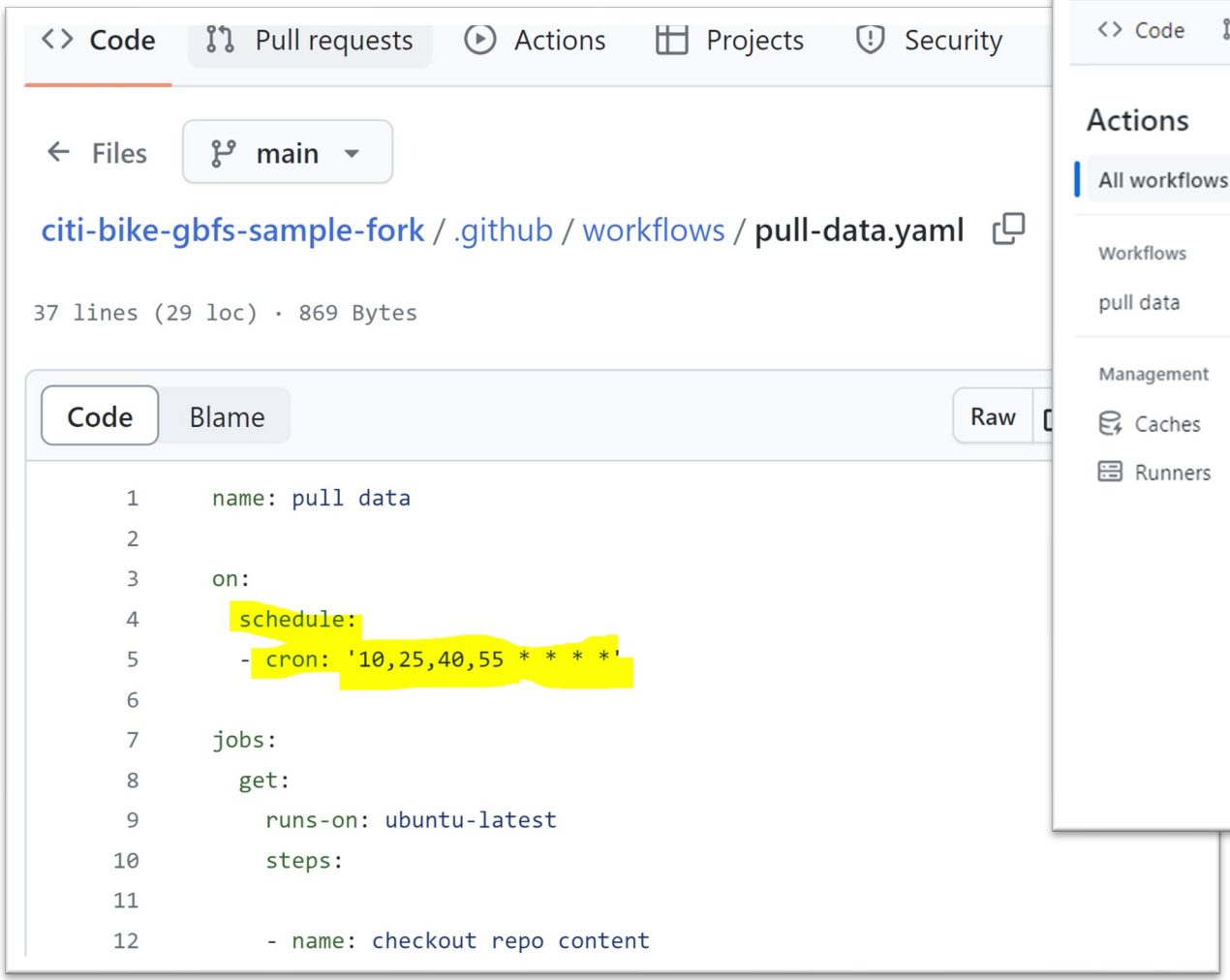
Photo: Elzbieta Sekowska / Shutterstock.com

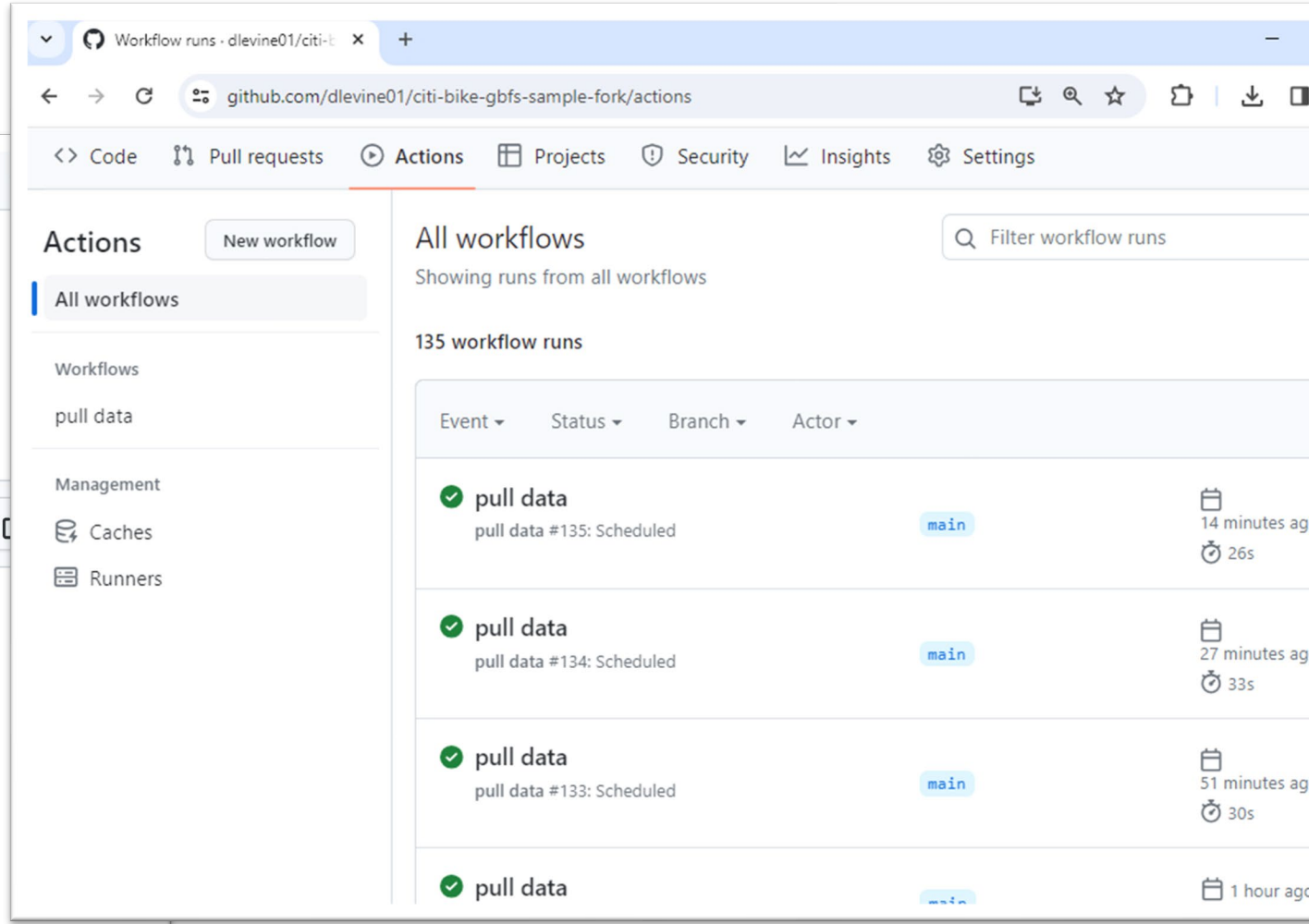
Fork the repo

https://github.com/NYCComptroller/citi-bike-gbfs



Start recording

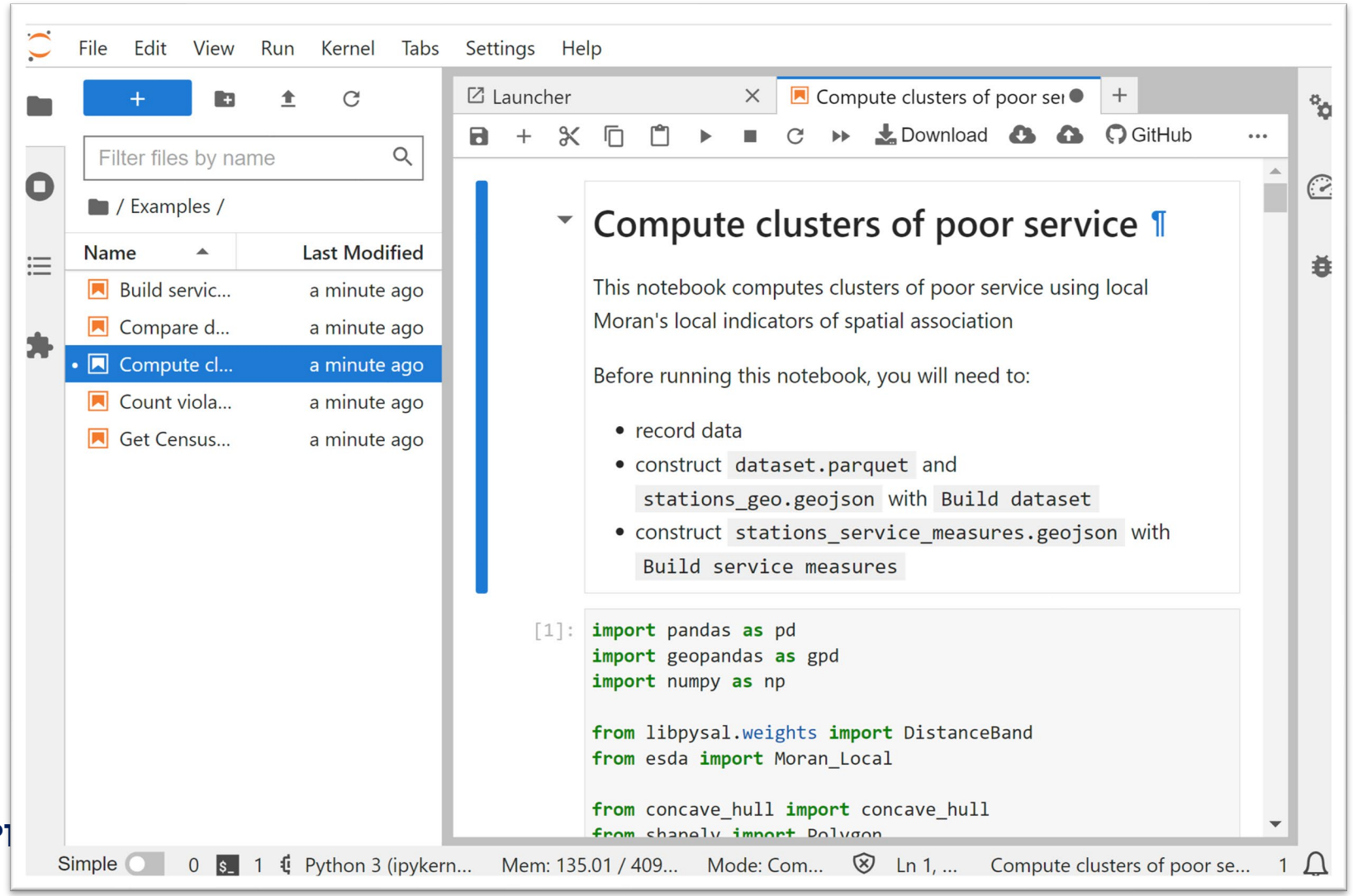




Build dataset

last_updated	station_id	capacity	is_renting	is_returning	num_docks_available	num_bikes_available	num_ebikes_available	num_bikes_disabled
2024-03-12 17:40:49	66dc2995-0aca-11e7-82f6-3863bb44ef7c	51	0	0	0	0	0	0
	06439006-11b6-44f0-8545-c9d39035f32a	48	0	0	0	0	0	0
	19d17911-1e4a-41fa-b62b-719aa0a6182e	39	0	0	0	0	0	0
	1861678548643203686	25	1	1	8	15	8	2
	cd2d9dab-7708-4685-a56f-9412c738de7e	23	1	1	1	22	1	0
	66db2a71-0aca-11e7-82f6-3863bb44ef7c	45	1	1	0	41	16	4
	901ad0c4-383e-490a-8b54-0656ce2358d6	19	1	1	3	11	2	5
	66dc292c-0aca-11e7-82f6-3863bb44ef7c	55	1	1	50	3	1	2
	4c03fa2d-89da-4f0b-8f19-c7de5edbe256	25	1	1	3	21	3	0
	9af90faf-0b9b-451b-9cb0-20ff421ca1d9	27	1	1	18	9	1	0

View on Binder



Your turn!

- Possible extensions:
  - Check whether service has improved or patterns have changed
  - Use station stats to measure reliability or service in other ways
  - Check reliability at your favorite station
  - Check on availability in real time, or closer to real-time
- Other applications:
  - Recording any other online data feed to build a longitudinal dataset
  - Clustering analysis and demographic comparison

Thank you!

Stay in touch
at beta.nyc/links

βetaNYC

NYC SCHOOL OF DATA

nycsodata24.sched.com #nycSOdata #opendataweek


NYC SCHOOL OF DATA
OPEN DATA WEEK
Powered by NYC OpenData