Liz Roten

Using location-based services to locate high activity areas within Twin Cities Regional Parks

2019-10-08

Women in Machine Learning and Data Science

METROPOLITAN COUNCIL

Outline

- Why regional parks?
- Location-based services and StreetLight data
- Case study: Long Lake Regional Park

Why regional parks?

- The Metropolitan Council is required by law to estimate how many people visit the regional parks system annually
- Regional parks are owned and managed by implementing agencies

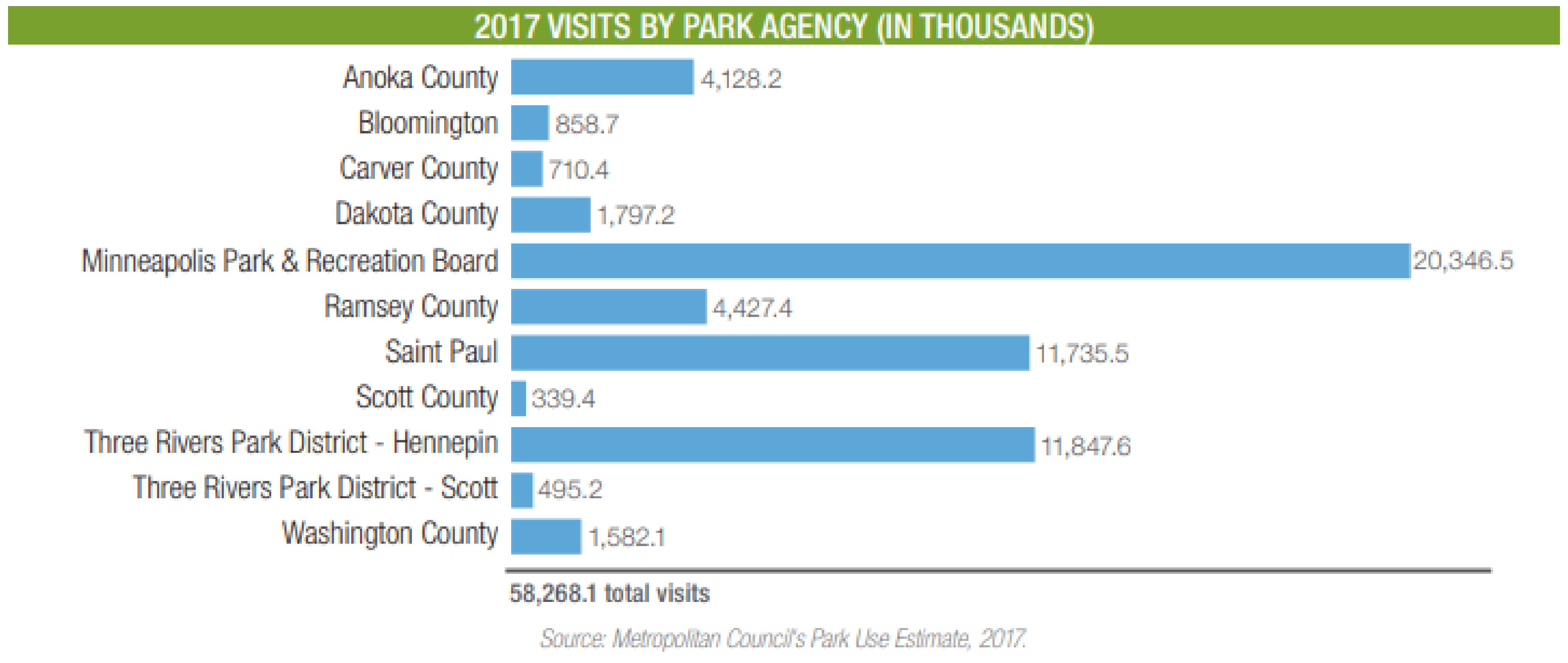
68 Regional parks, park reserves, & special features

78 Regional trails

METROPOLITAN COUNCIL

How we usually research visitors

- Annual Use Estimates
  - Representative at the agency level
- Visitor Studies
  - In-person sampled surveys
  - Representative at the agency level
- Limitations
  - Every park has its unique features and geography
  - Self-selection survey bias
  - Funding restrictions



Image, chart: Metropolitan Council

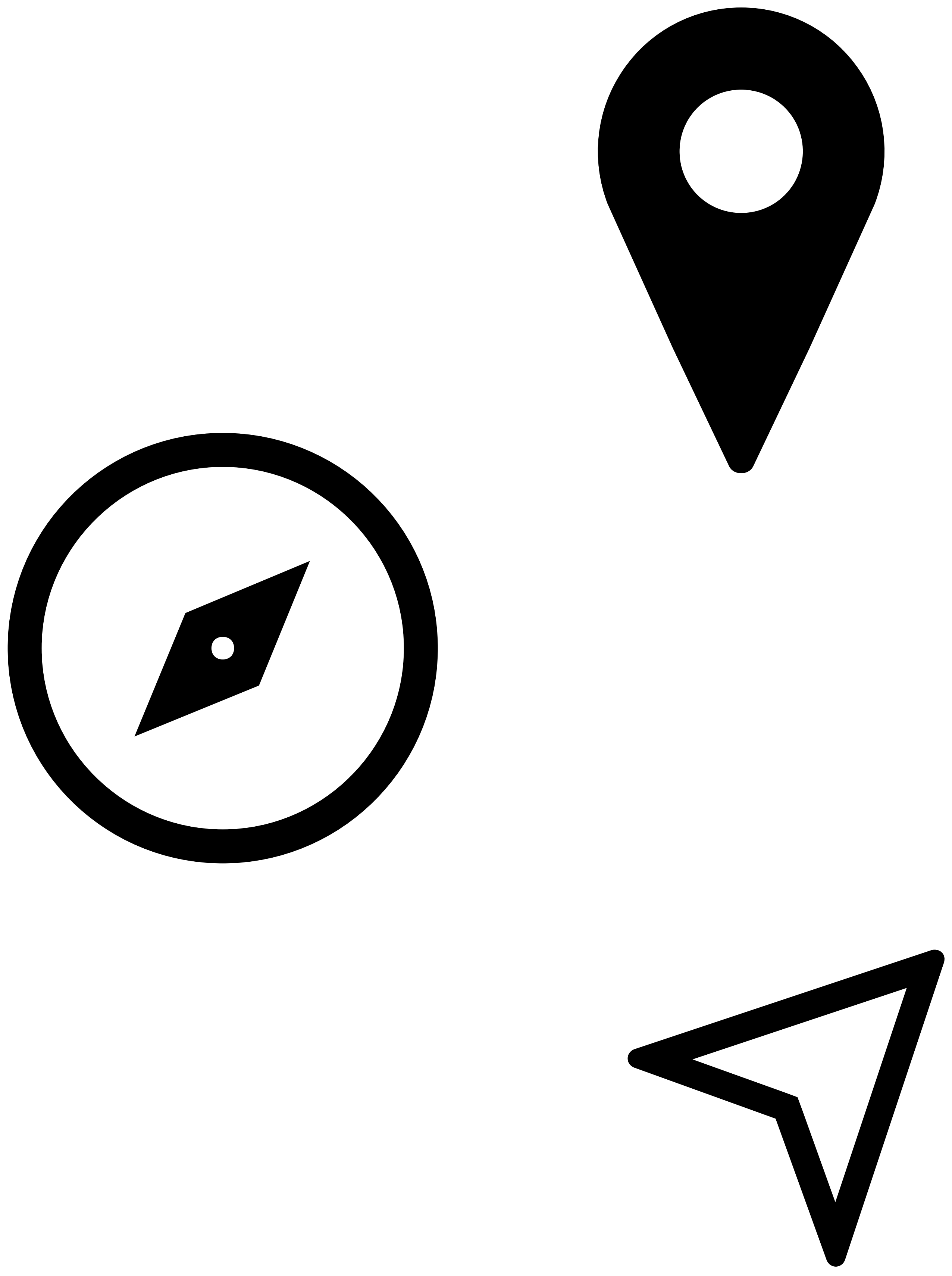
Outline

- Why regional parks?
- **Location-based services and StreetLight data**
- Case study: Long Lake Regional Park

METROPOLITAN COUNCIL

Take out your phone!

- Go to your settings
  - Or try swiping down from the top of your screen, and look for an icon like these
- If your location is turned on, then your data is being collected by *someone*
- Apps and services use your location to provide services



StreetLight Data

- Provides platform for running individual analyses, such as zone activity and origin-destination
- Uses LBS data gathered from apps, such as shopping, dating, weather, productivity
- Accurate to 20 meters (65 feet)

"StreetLight Data is the first company to make using real-world transportation data easy, efficient, and affordable."

Image, quote: "About Us" page

Data Privacy

Personal identifiable information is removed by the LBS data provider

StreetLight does not process, use, or distribute personally identifiable information

Users contractually commit to not using StreetLight metrics in combination with other materials to try to identify individuals.

StreetLight data characteristics

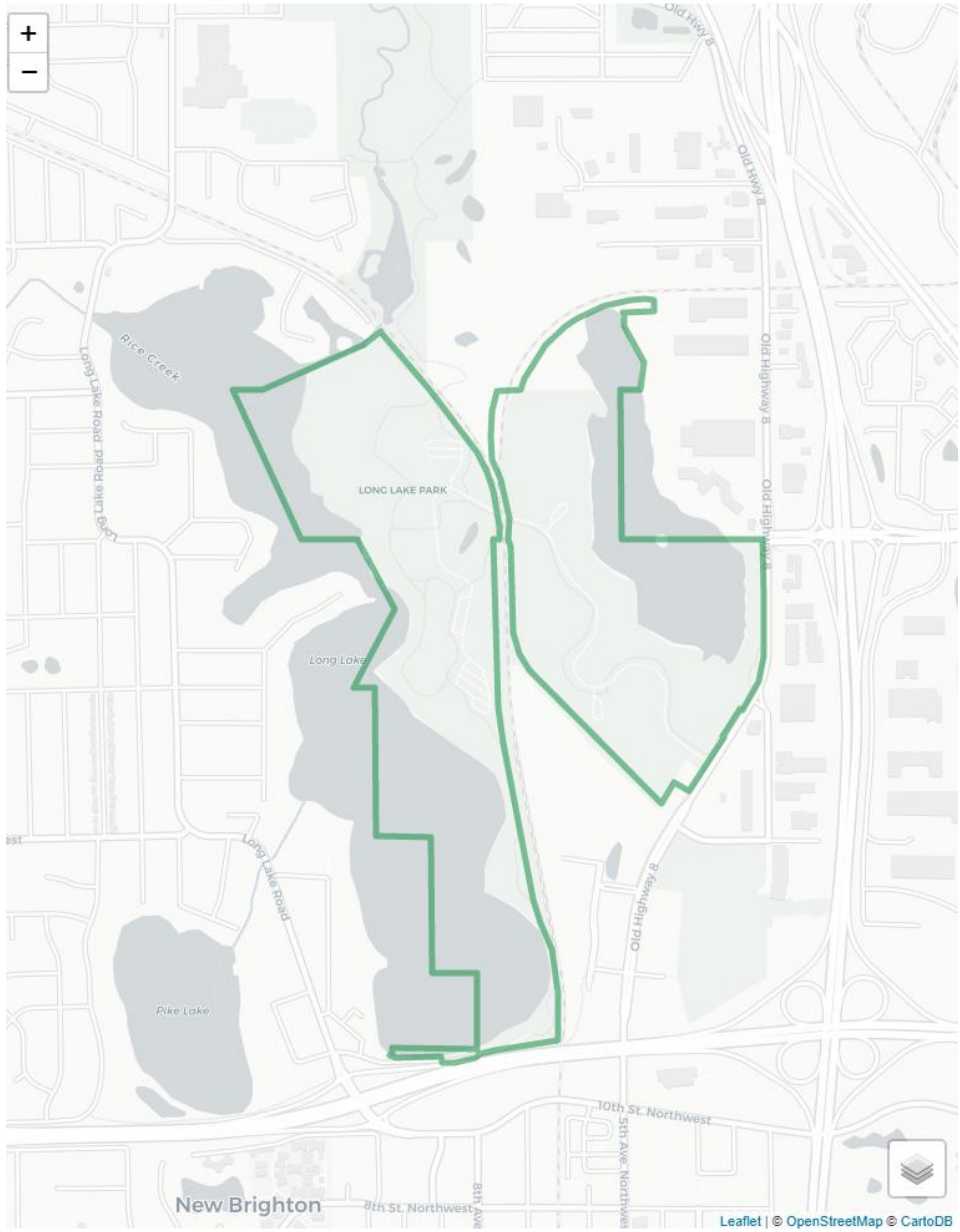
- **StreetLight Traffic Index** a normalized measure of relative traffic in each area
  - The Index is NOT a count of devices or vehicles
- Calculated using LBS data, with contextual verification
- Large sample size
- Bike and pedestrian (limited)

Outline

- Why regional parks?
- Location-based services and StreetLight data
- **Case study: Long Lake Regional Park**

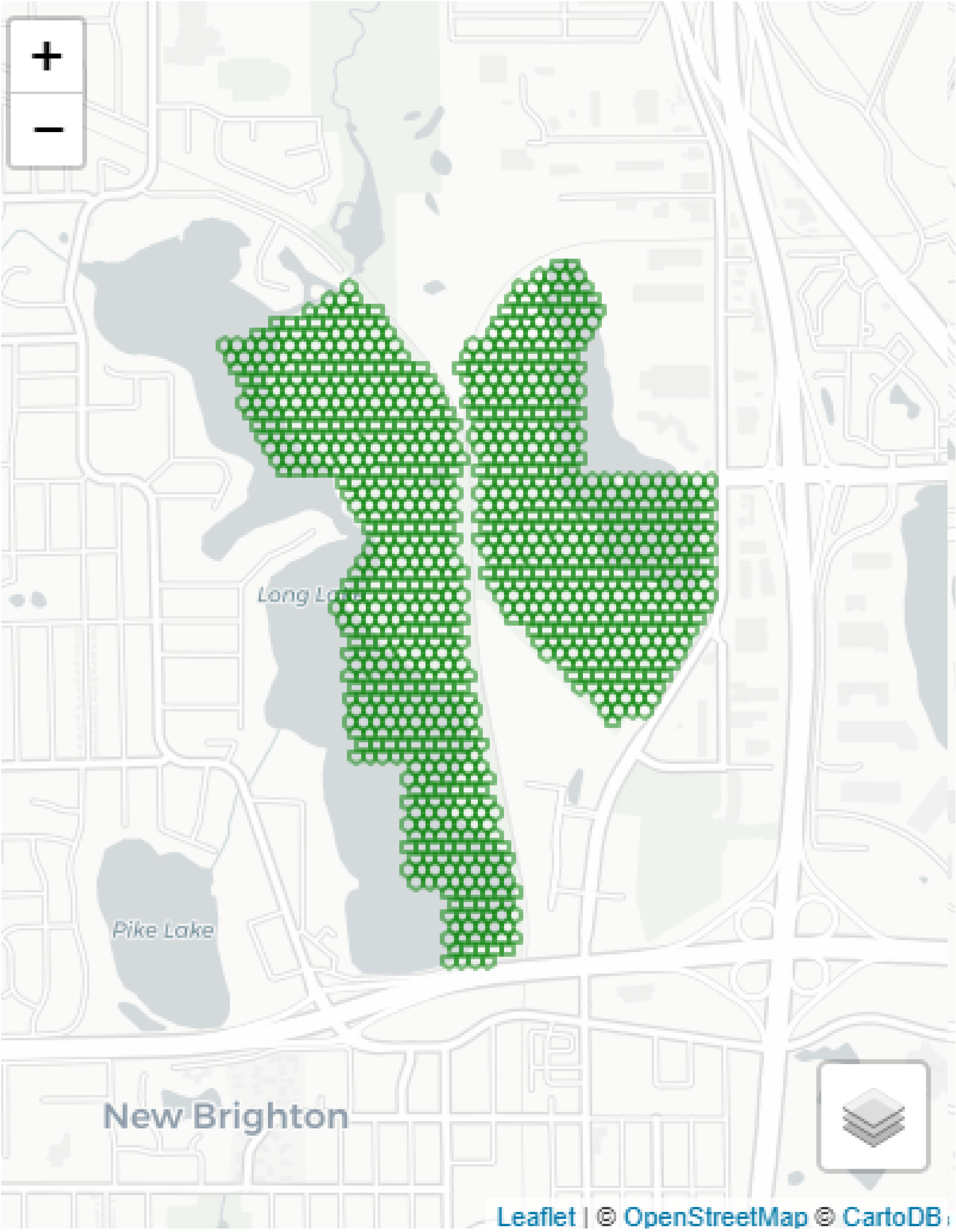
Long Lake Regional Park

- Regional park in Ramsey county
- Amenities include
  - Guarded swimming beach
  - Playgrounds
  - Boat launch
  - Pavilion
  - Trails



Long Lake Regional Park

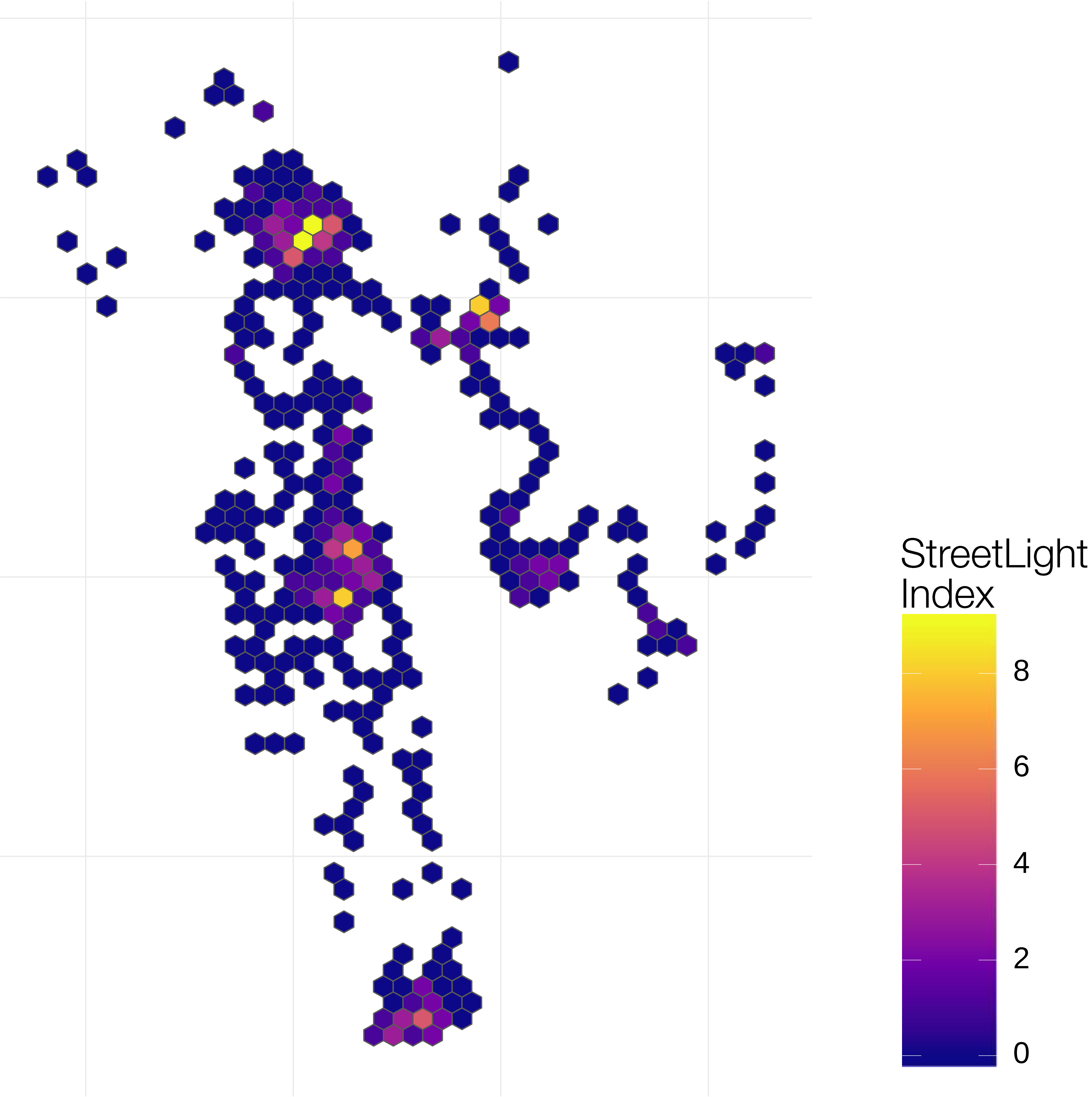
- We want to know where there is the most activity within the park
- Now broken into 996 hexagons, each 1.2 square kilometers
- Careful to avoid private residences!



Run the StreetLight Analysis

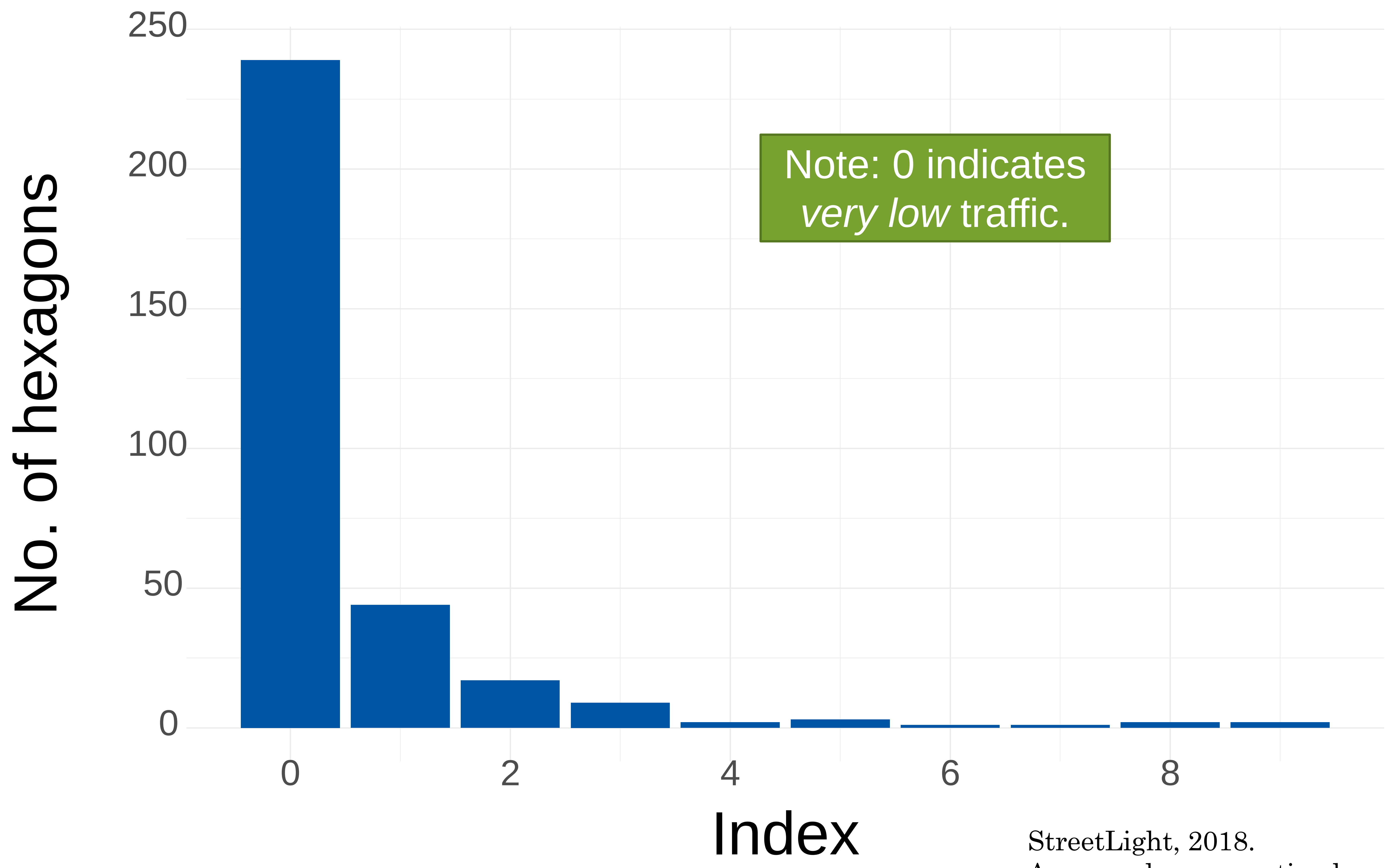
- Zone Activity Analysis with…
  - All months in 2018
  - Only trips that *end* in the park
  - On any day of the week (M-Su)
  - During recreation hours (8am-8pm)
- Results
  - Sample size of approx. 2,000 devices and 4,000 trips
  - StreetLight Index range from 0-9

StreetLight Index



Mississippi St. Mississippi St.
Rice Creek Road Rice Creek Road
Rice Creek
Long Lake Road
LONG LAKE PARK
Lake Diana
Silver Lake Rd. Northwest
Long Lake
14th St. Northwest
Long Lake Road
Innsbruck Drive
Palmer Drive
Pike Lake
Old Highway 8
County Highway 96 County Highway 96
Old Highway 10
10th St. Northwest
Index
0
1
2
3
4
5
6
7
8
9
Leaflet | © OpenStreetMap © CartoDB

Traffic Index Distribution, Long Lake

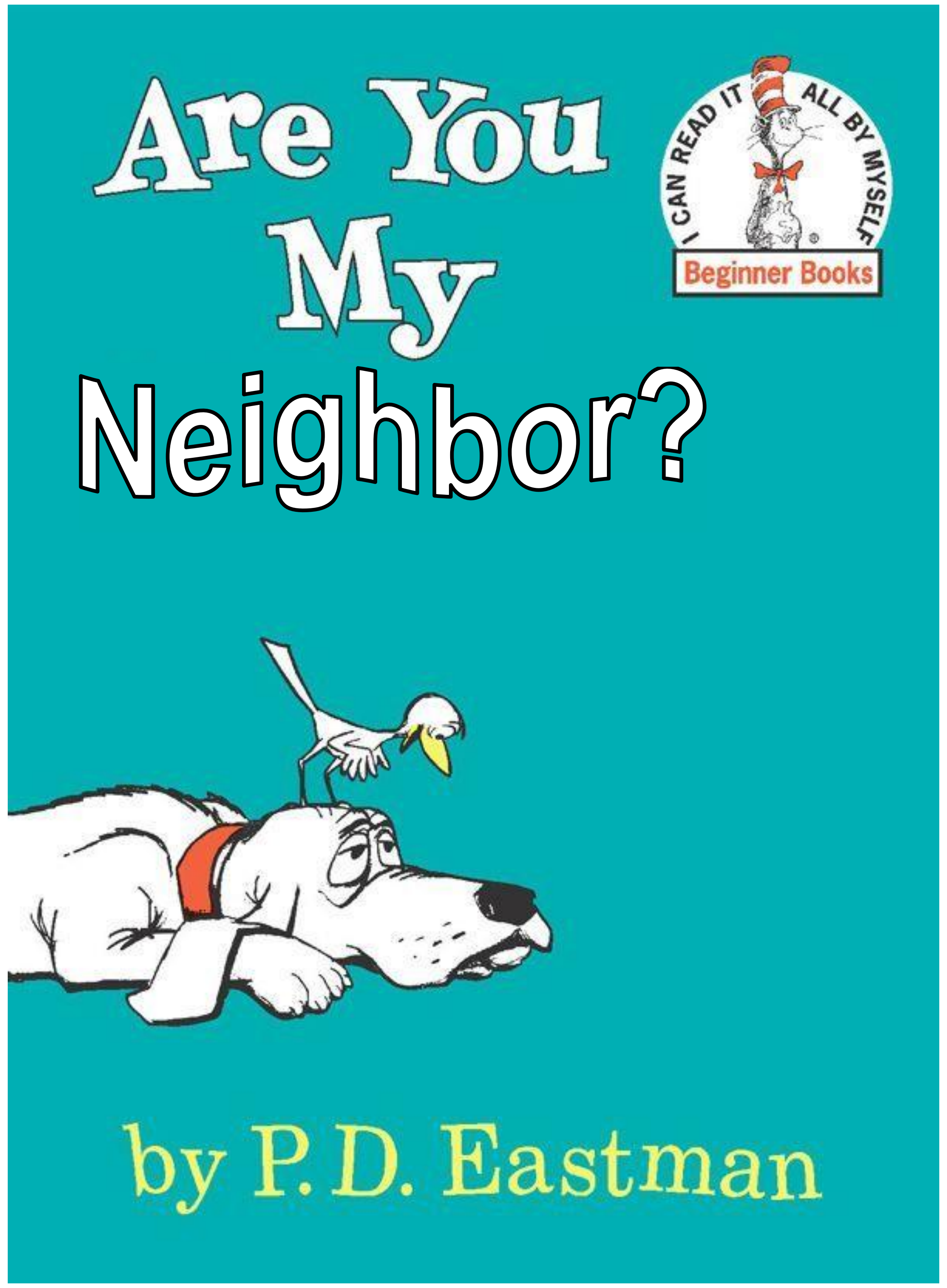


StreetLight, 2018.
Average day, recreation hours

METROPOLITAN COUNCIL

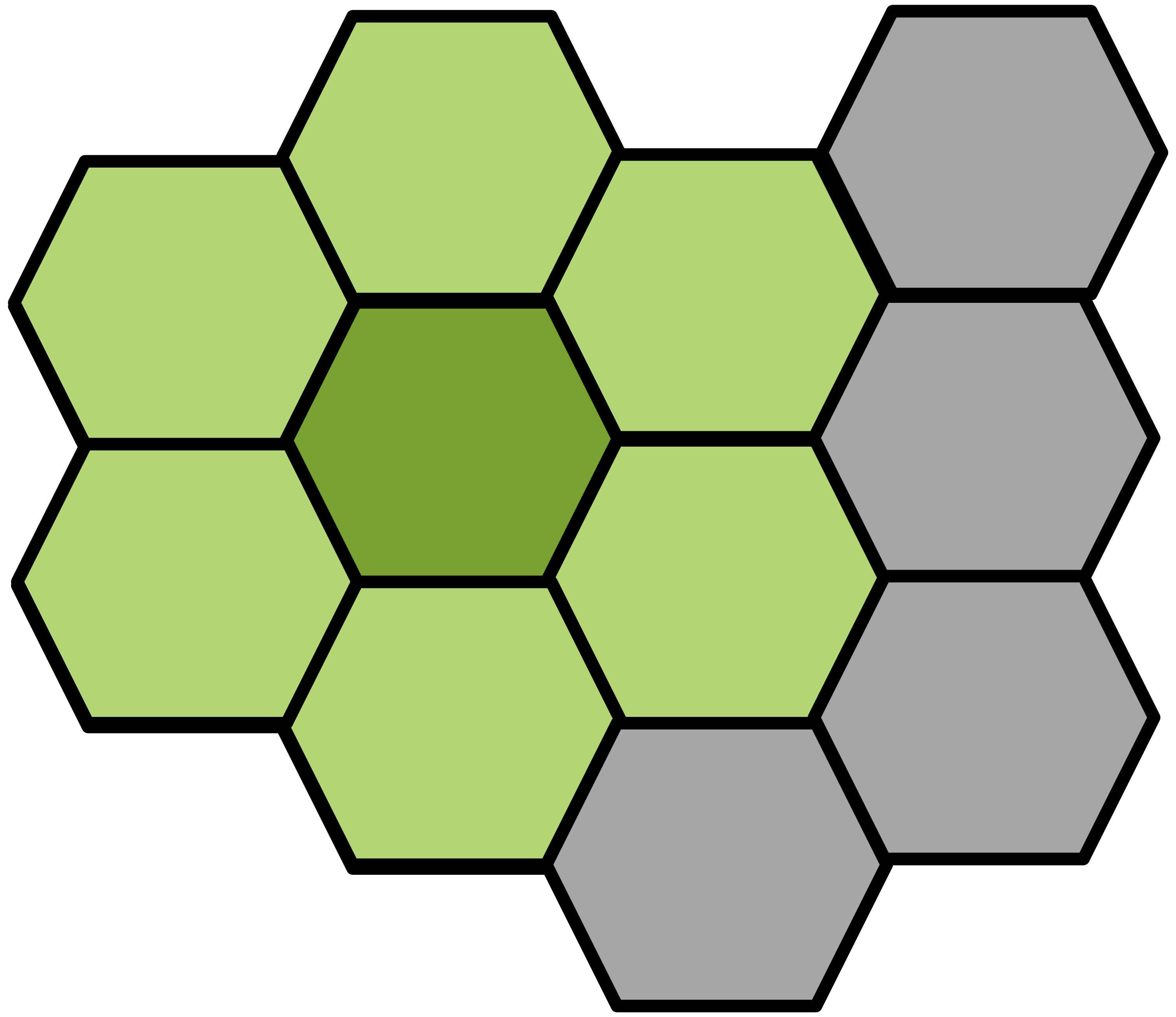
Testing for spatial autocorrelation

Identifying neighbors

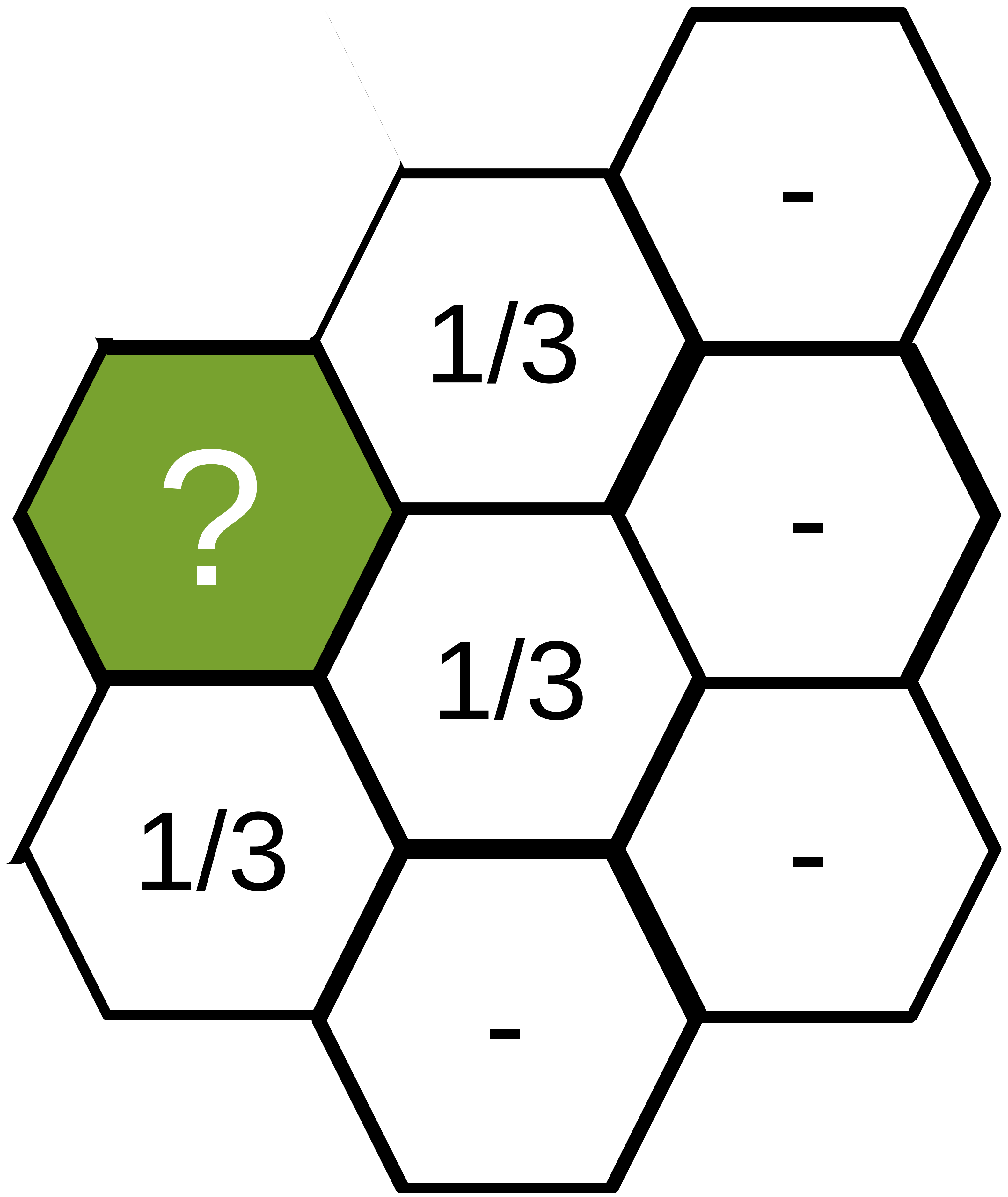


Image, Amazon

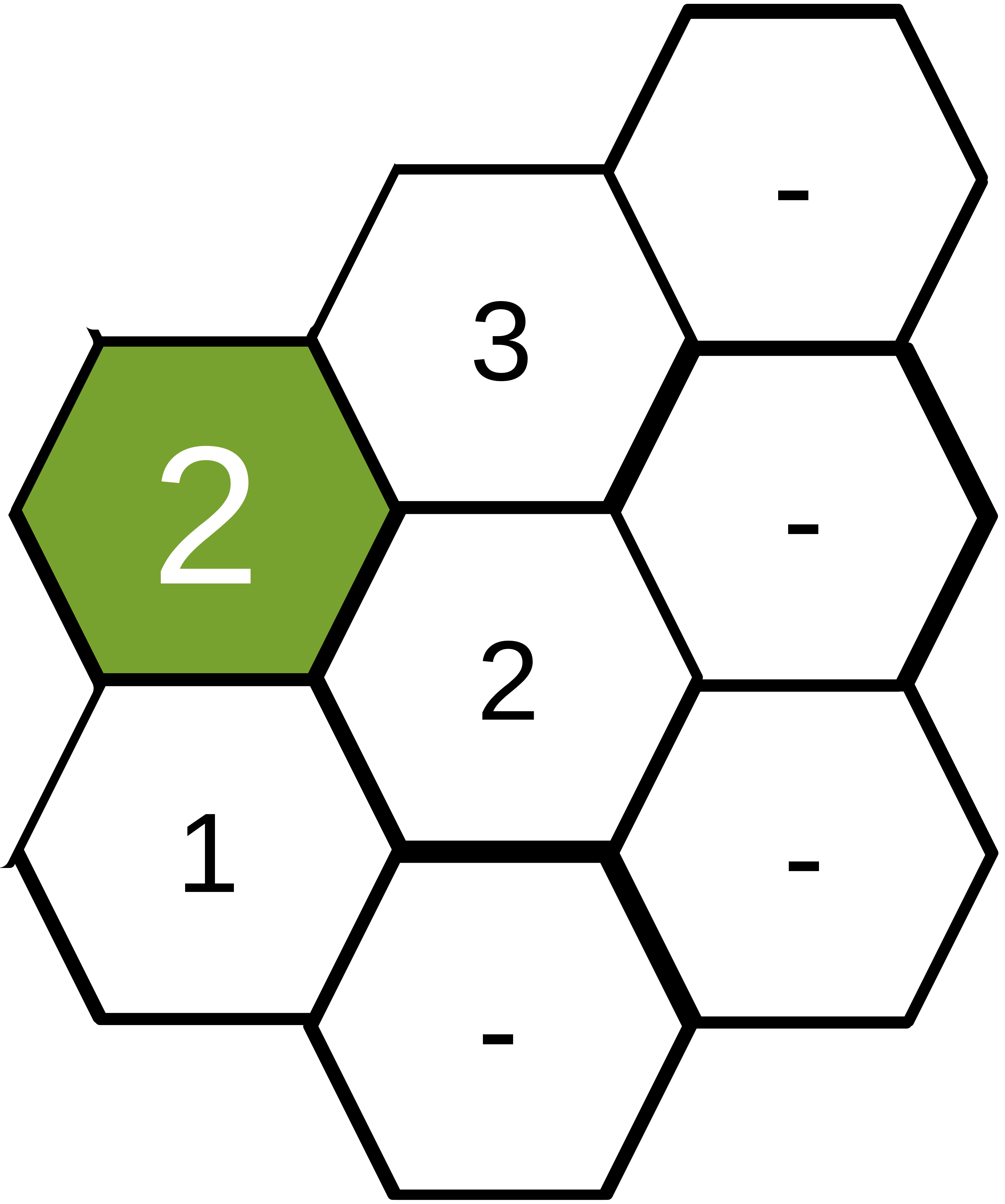
- In this case, neighbors are hexagons that share a line segment (rook's scheme)



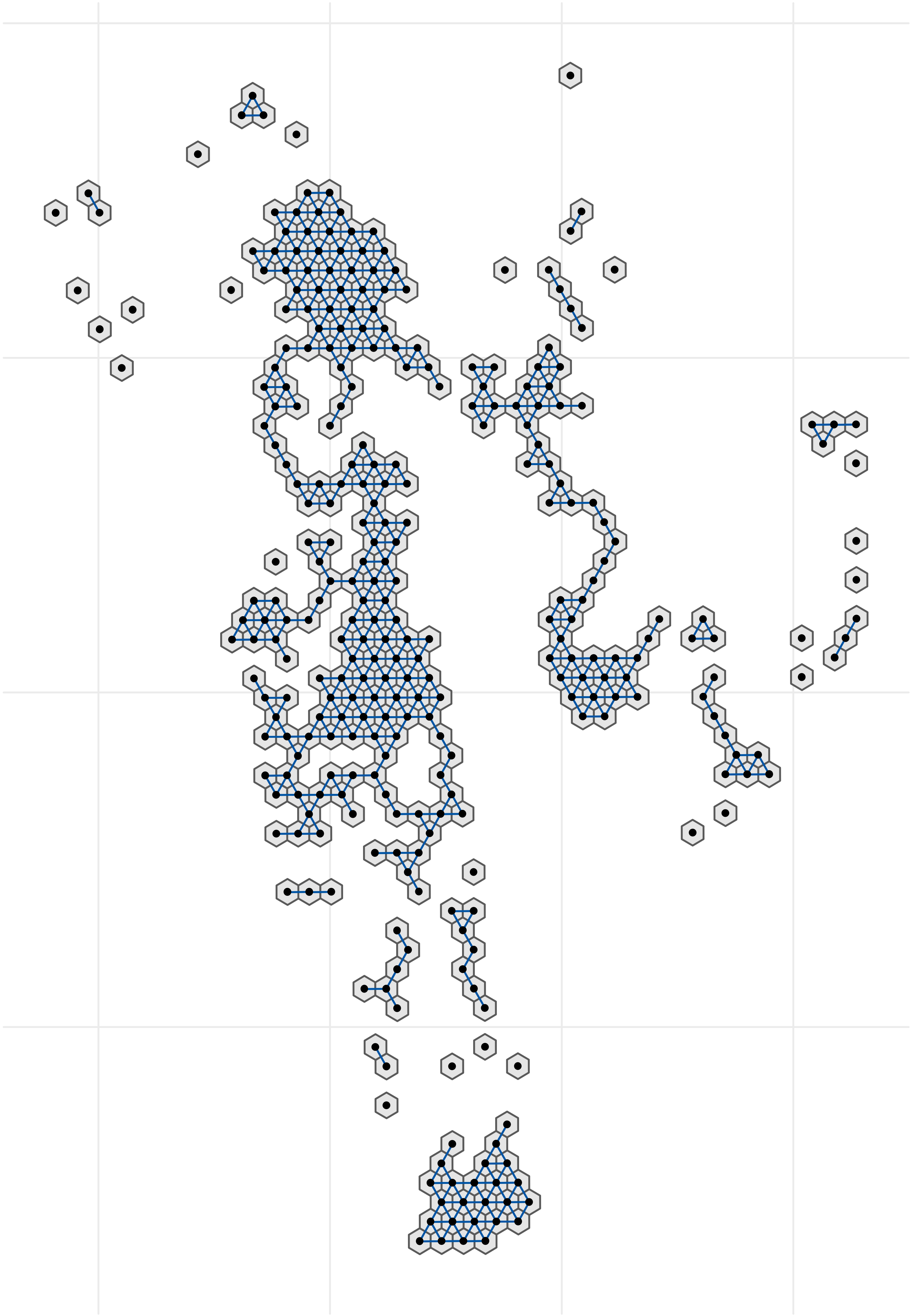
Spatial weights matrix



Create lagged variable



Create neighbors and weights



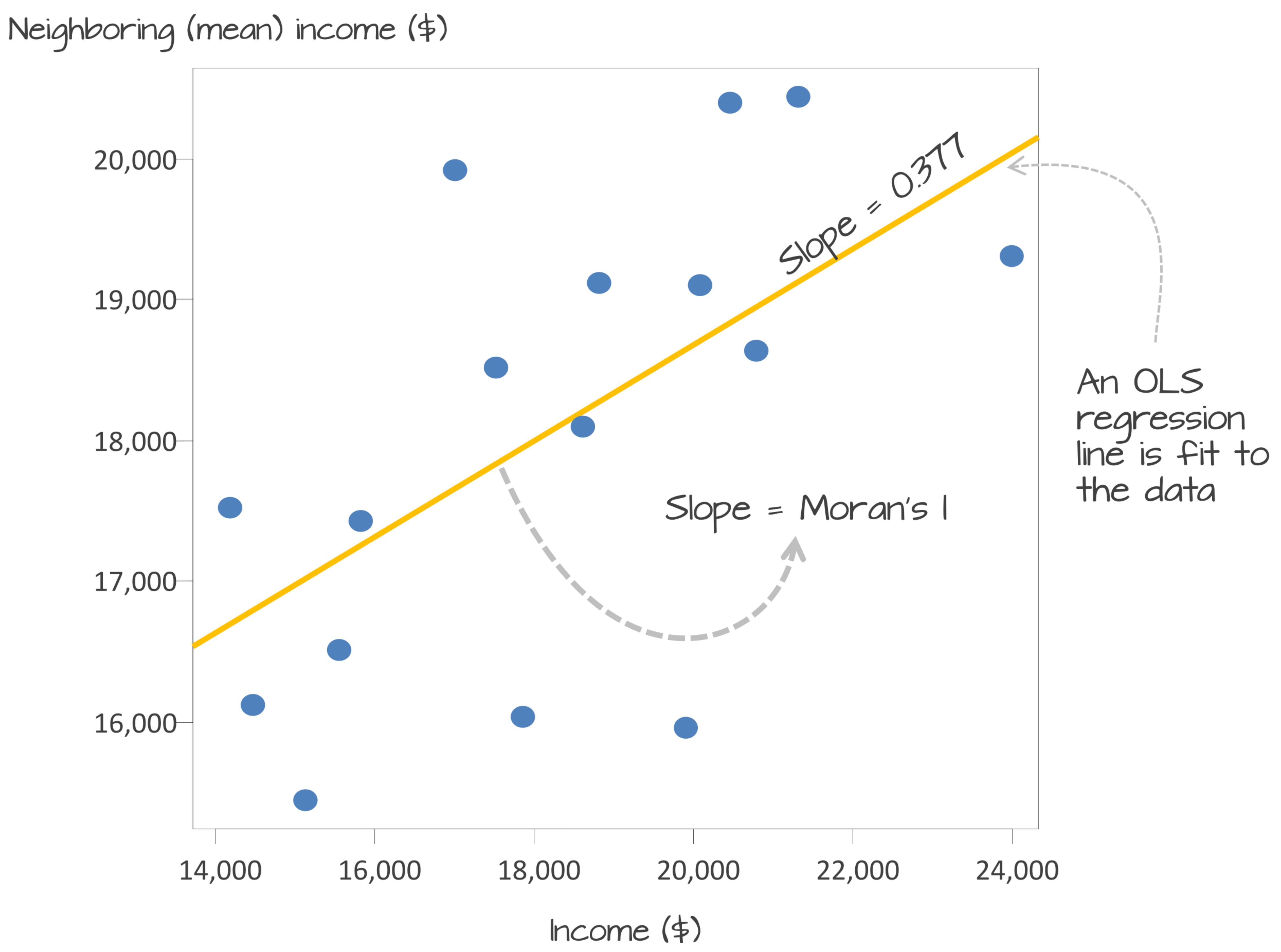
```
neighbors <-
poly2nb(pl = long_lake_spatial,
        queen = FALSE)

neighbors_weights_list <-
nb2listw(neighbors,
        style = "S",
        zero.policy = TRUE)
# allows NA values
```

“S” specifies a variance stabilizing coding scheme (Tiefelsdorf et al. 1999).

Measure clustering with Moran’s I

“Moran’s I is the correlation coefficient for the relationship between a variable and its surrounding values.”

Gimond, 2019

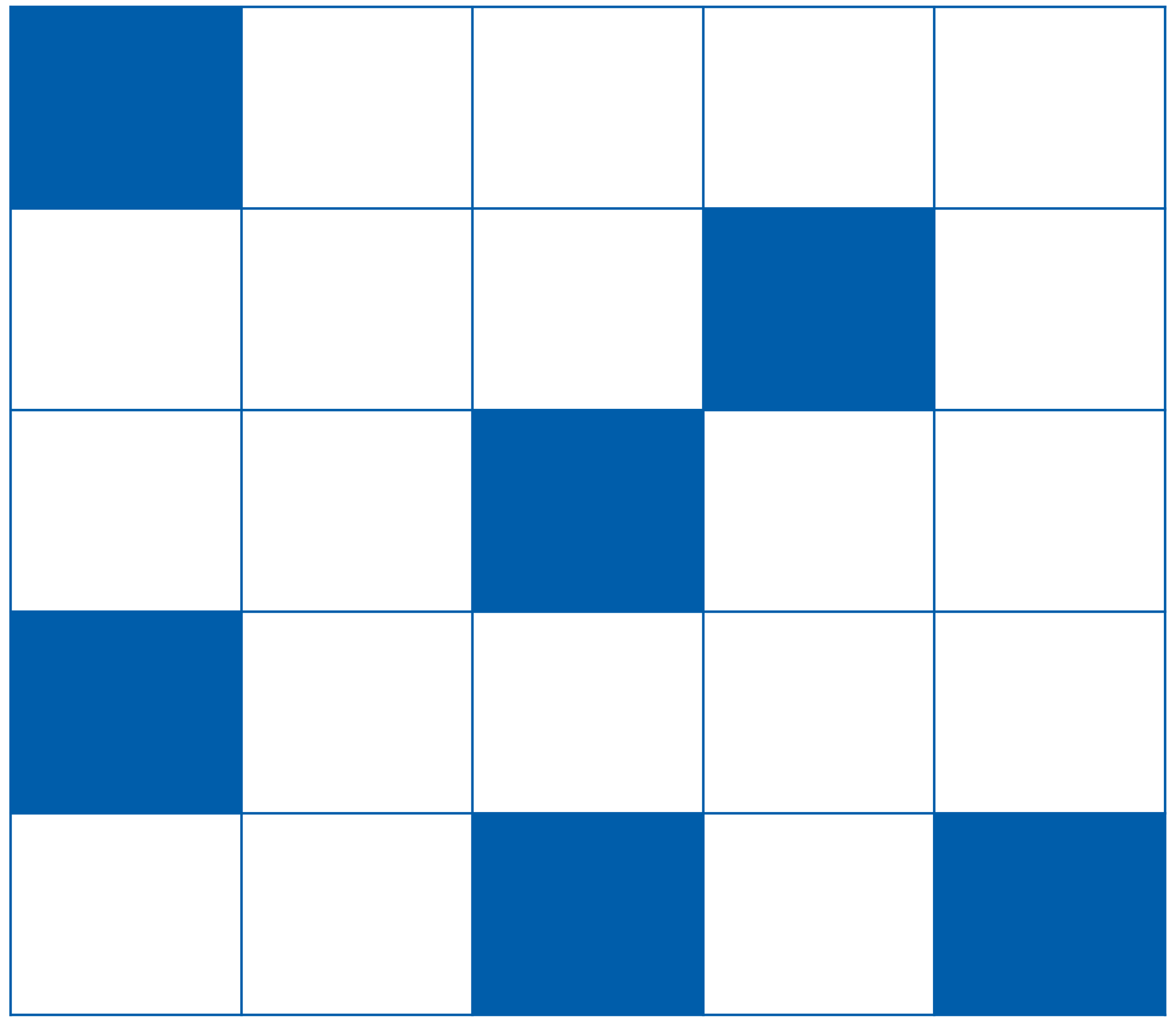


Gimond, Manuel. 2019. *Intro to GIS and Spatial Analysis.* Colby College

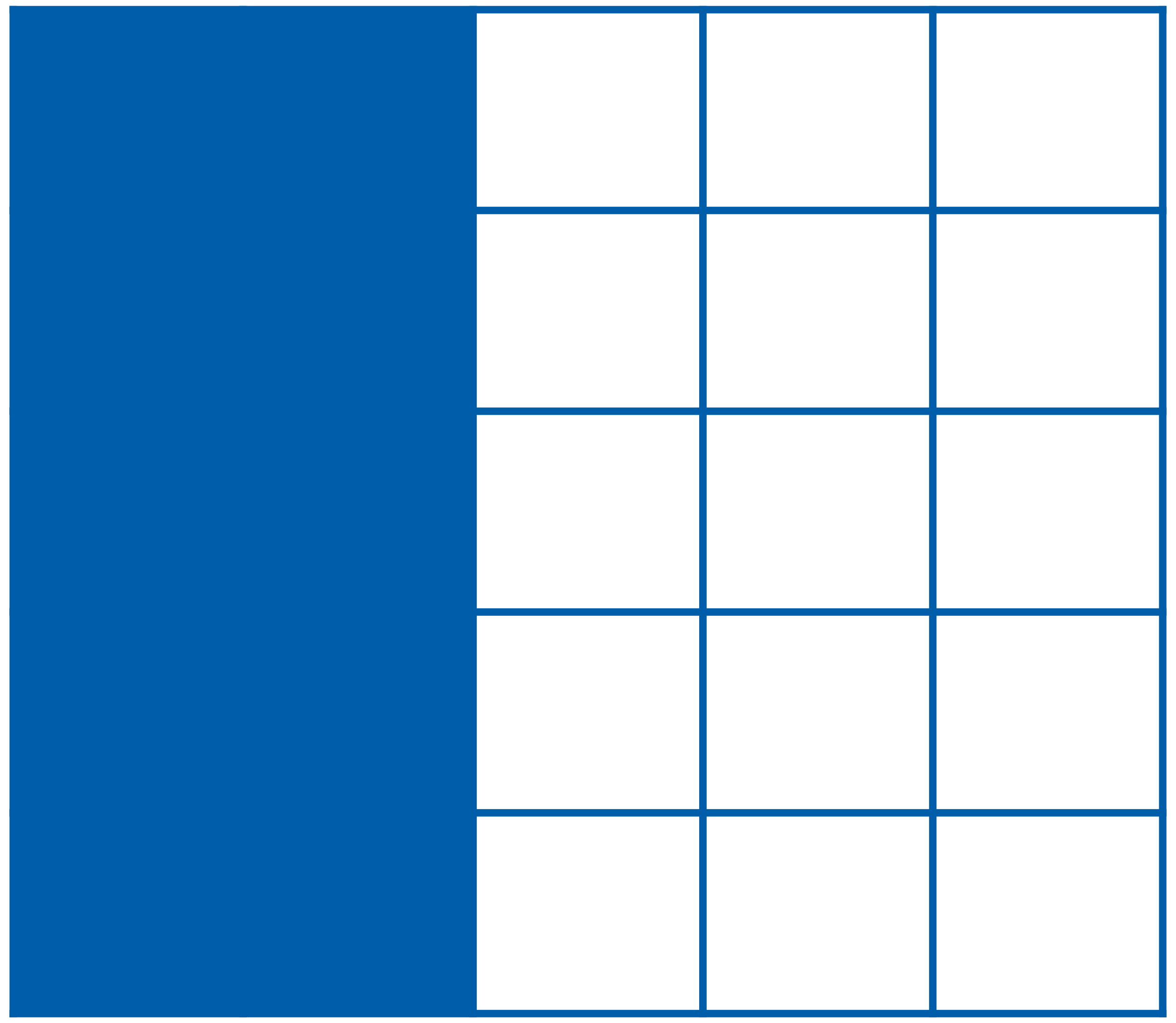
METROPOLITAN COUNCIL

Moran’s I value

Moran’s I = -1,
Perfect dispersion



Moran’s I = 0,
Random



Moran’s I = 1,
Clustered

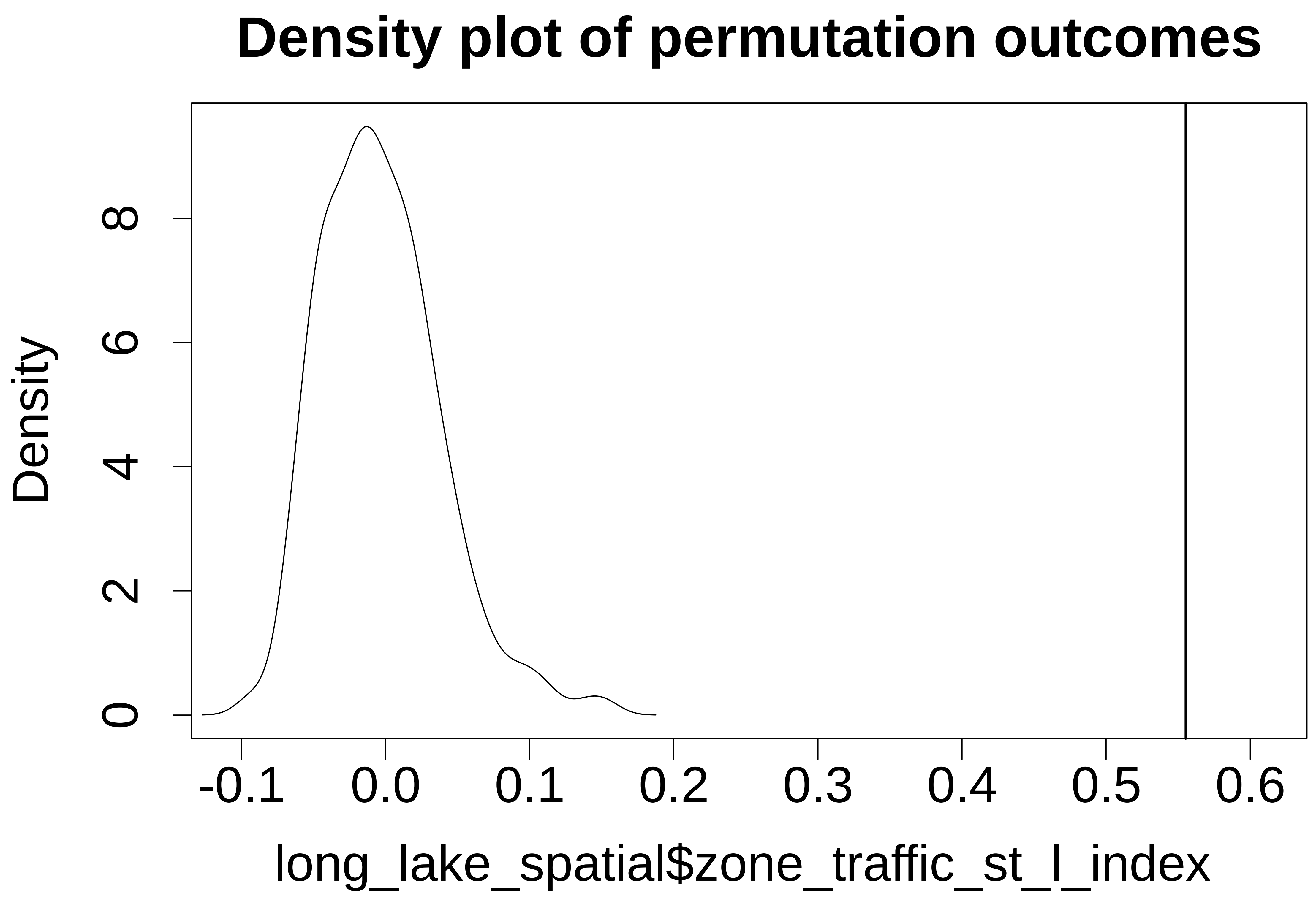
Global Moran's I

```
Moran I test under randomisation

data:  long_lake_spatial$zone_traffic_st_l_index
weights: neighbors_weights_list  n reduced by no-neighbour
observations

Moran I statistic standard deviate = 12.78, p-value <
0.0000000000000022
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.555167330      -0.003389831       0.001910092
```

Monte Carlo simulation

Assigns a random value to every hexagon and runs a Moran's I test. Repeat 600 times

```
statistic = 0.55517, observed
rank = 600, p-value = 0.001667
alternative hypothesis: greater
```



Find “hotspots” with LISA

Quickmeme

- LISA (Local Indicators of Spatial Association)[1]
  - Used to find clusters within the entire dataset
  - Breaks down the Global Moran’s I by calculating Moran’s I for each hexagon
  - Finds significant areas with a *pseudo* p-value

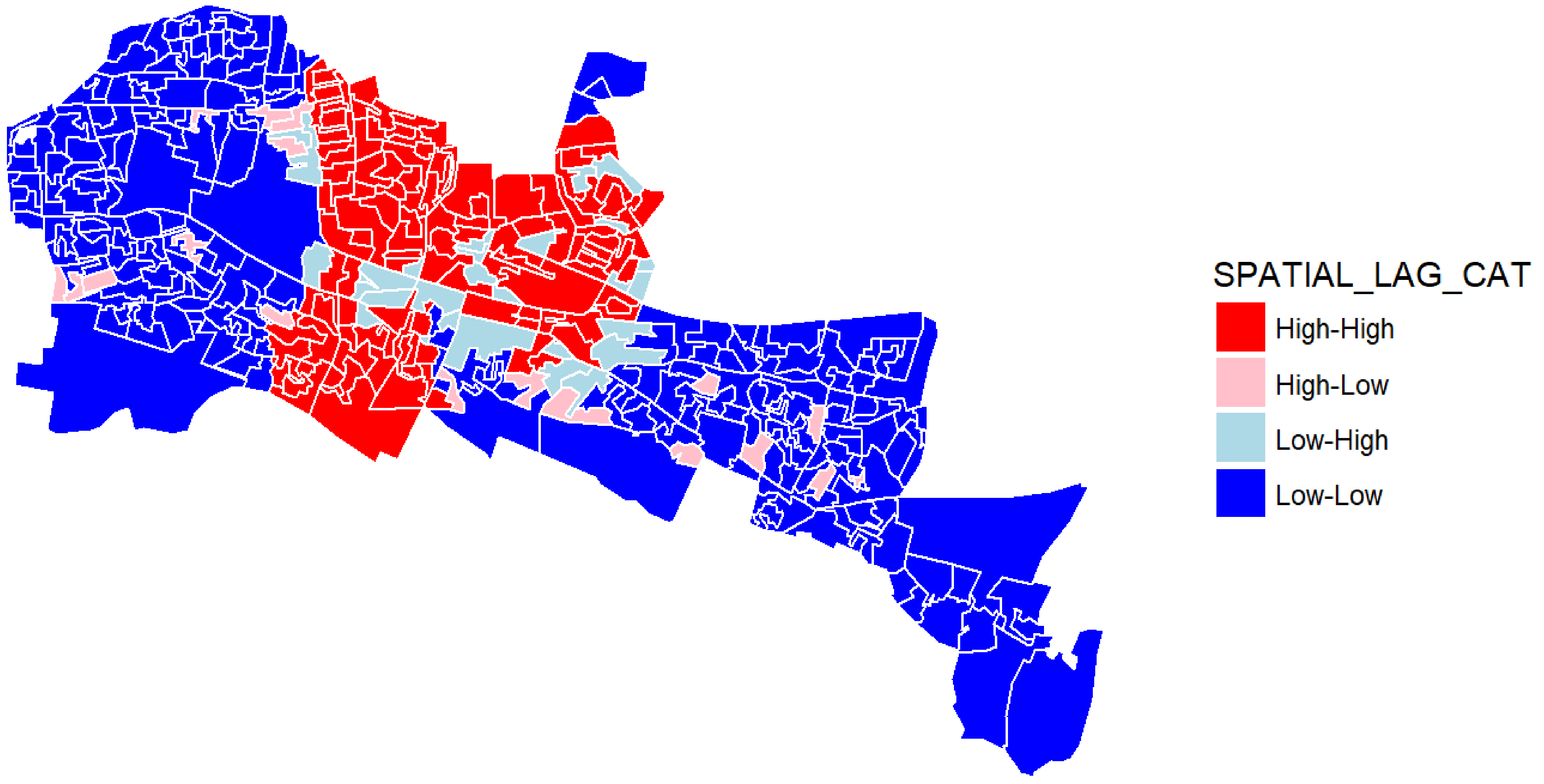
```
LM_Results <- localmoran(
    long_lake_spatial$zone_traffic_st_l_index,
    neighbors_weights_list,
    p.adjust.method = "fdr”,
    na.action = na.exclude,
    zero.policy = TRUE)
```

METROPOLITAN COUNCIL

1: Anselin et al. 2006 GeoDa

LISA categories

- The Moran's I for each hexagon isn't that useful
- A common practice is to assign categories based on z-scores of the original value and the lagged value
- These identify spatial clusters and spatial outliers



[Clustering of Census Recorded Ethnic Background]
Simon Hailstone, 2017

LISA categories, explained

[Observed z-score] - [Lagged z-score]

LISA categories, explained

[Observed z-score] - [Lagged z-score]

If > 0, “High”
If < 0, “Low”

METROPOLITAN COUNCIL

LISA categories, explained

[Observed z-score] - [Lagged z-score]

If > 0, “High”
If < 0, “Low”

METROPOLITAN COUNCIL

LISA categories, explained

High - High

Hexagon is relatively **high** and surrounded by **high** values.
This is a **high** traffic spatial **cluster**!

METROPOLITAN COUNCIL

LISA categories, explained

High - Low

Hexagon is relatively **high** and surrounded by **low** values.
This is a spatial **outlier**!

METROPOLITAN COUNCIL

LISA categories, explained

Low - Low

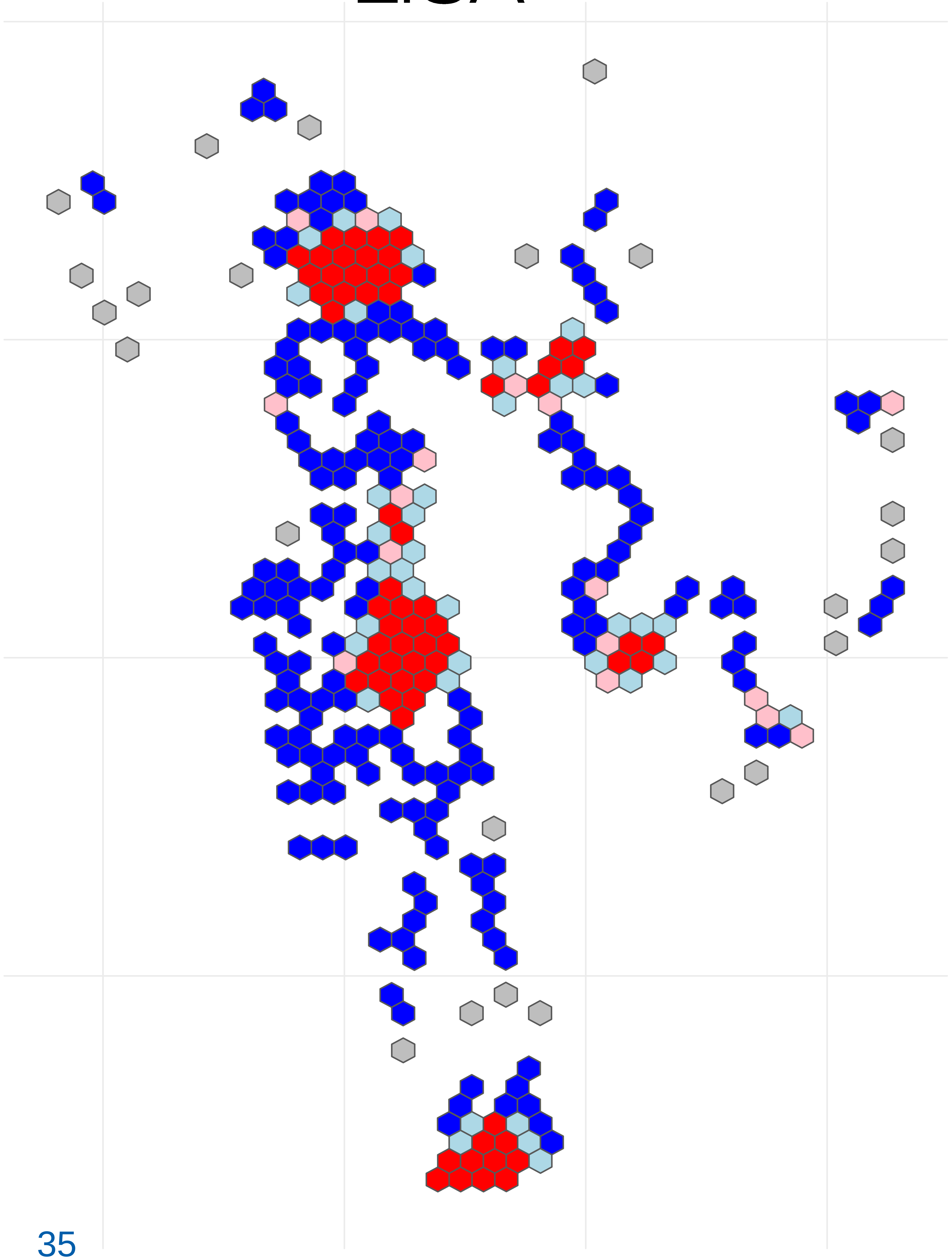
Hexagon is relatively **low** and surrounded by **low** values.
This is a **low** traffic spatial **cluster**!

METROPOLITAN COUNCIL

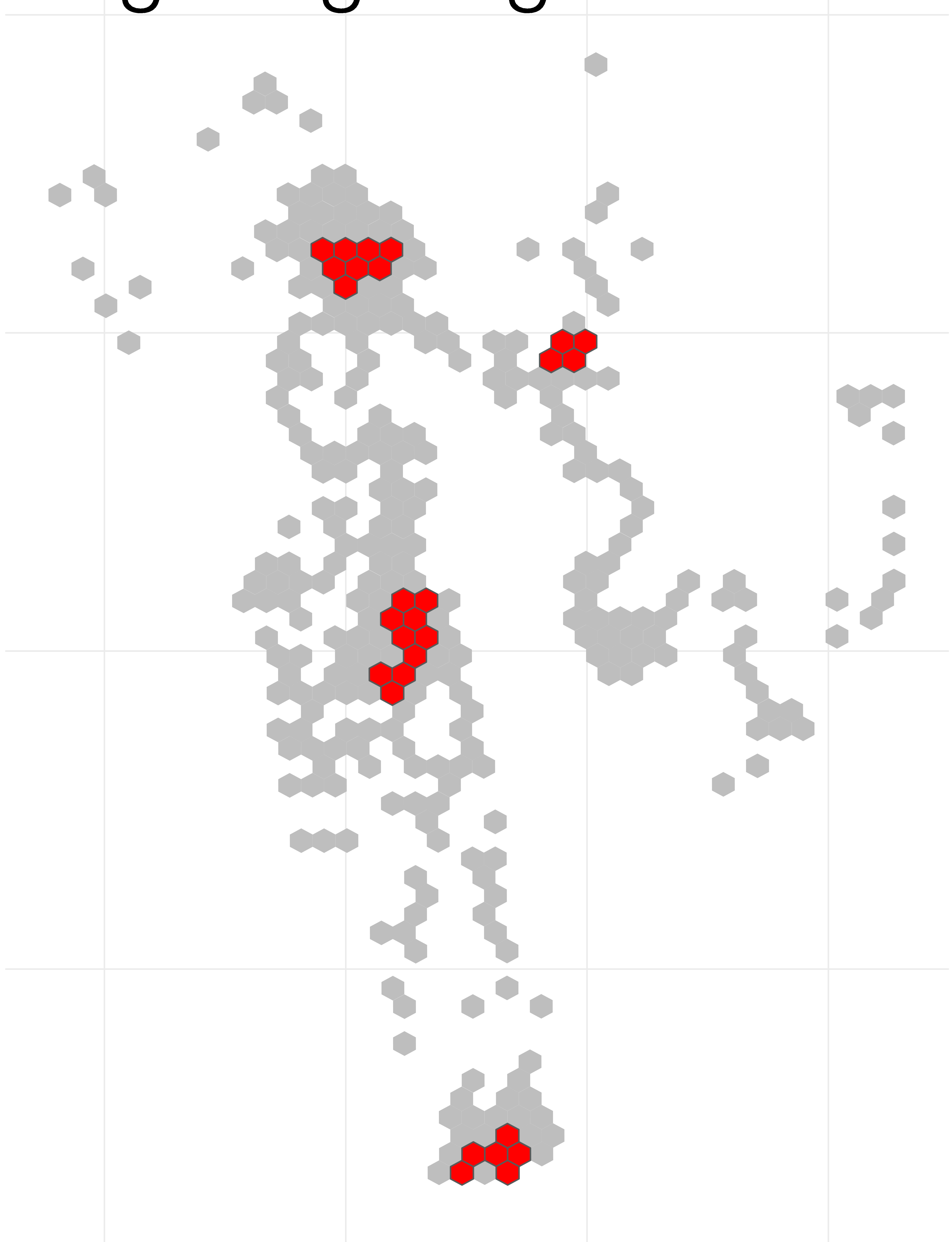
Observed and lagged Index z-scores

METROPOLITAN COUNCIL

LISA



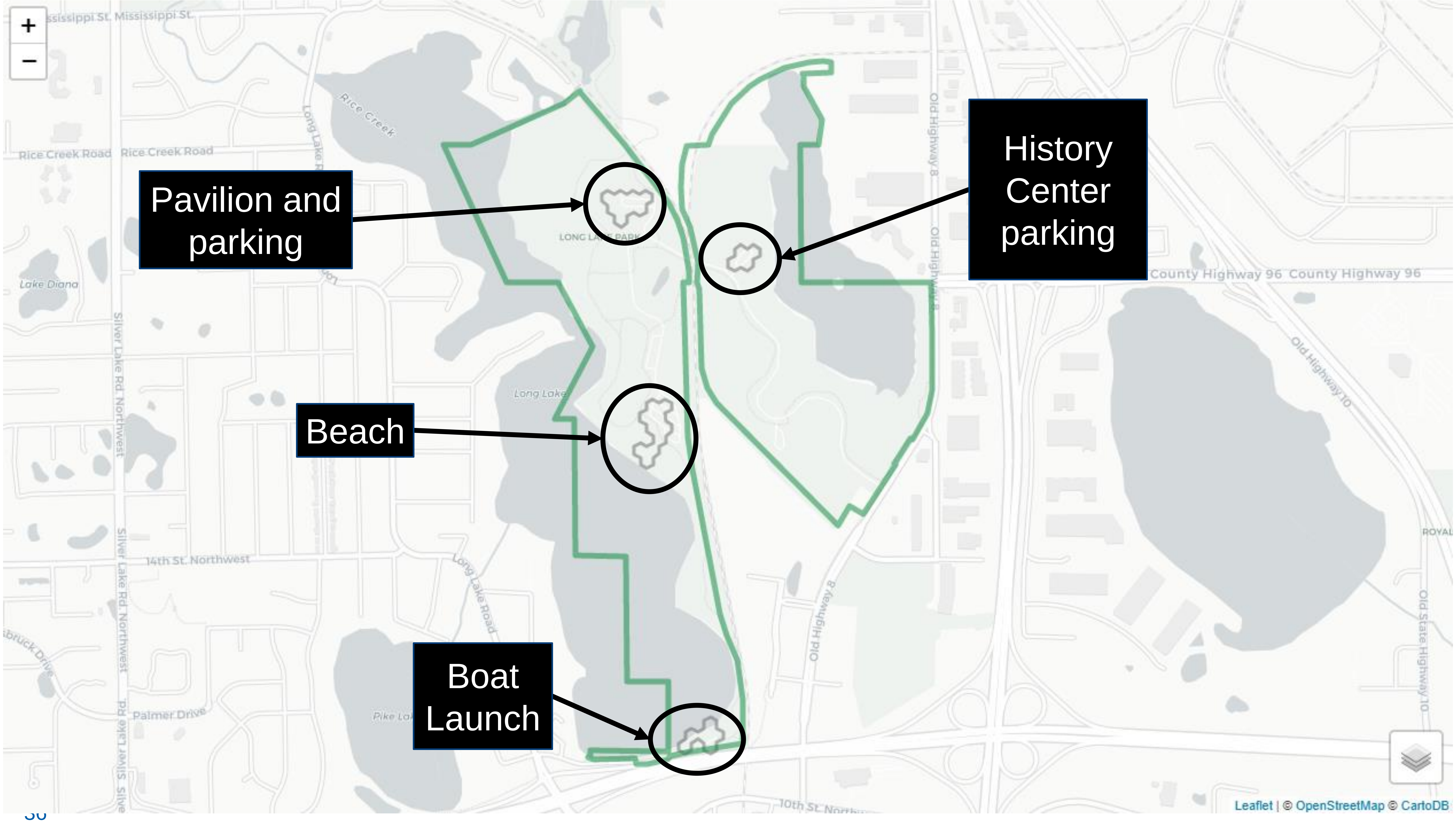
High-High Significant



LISA

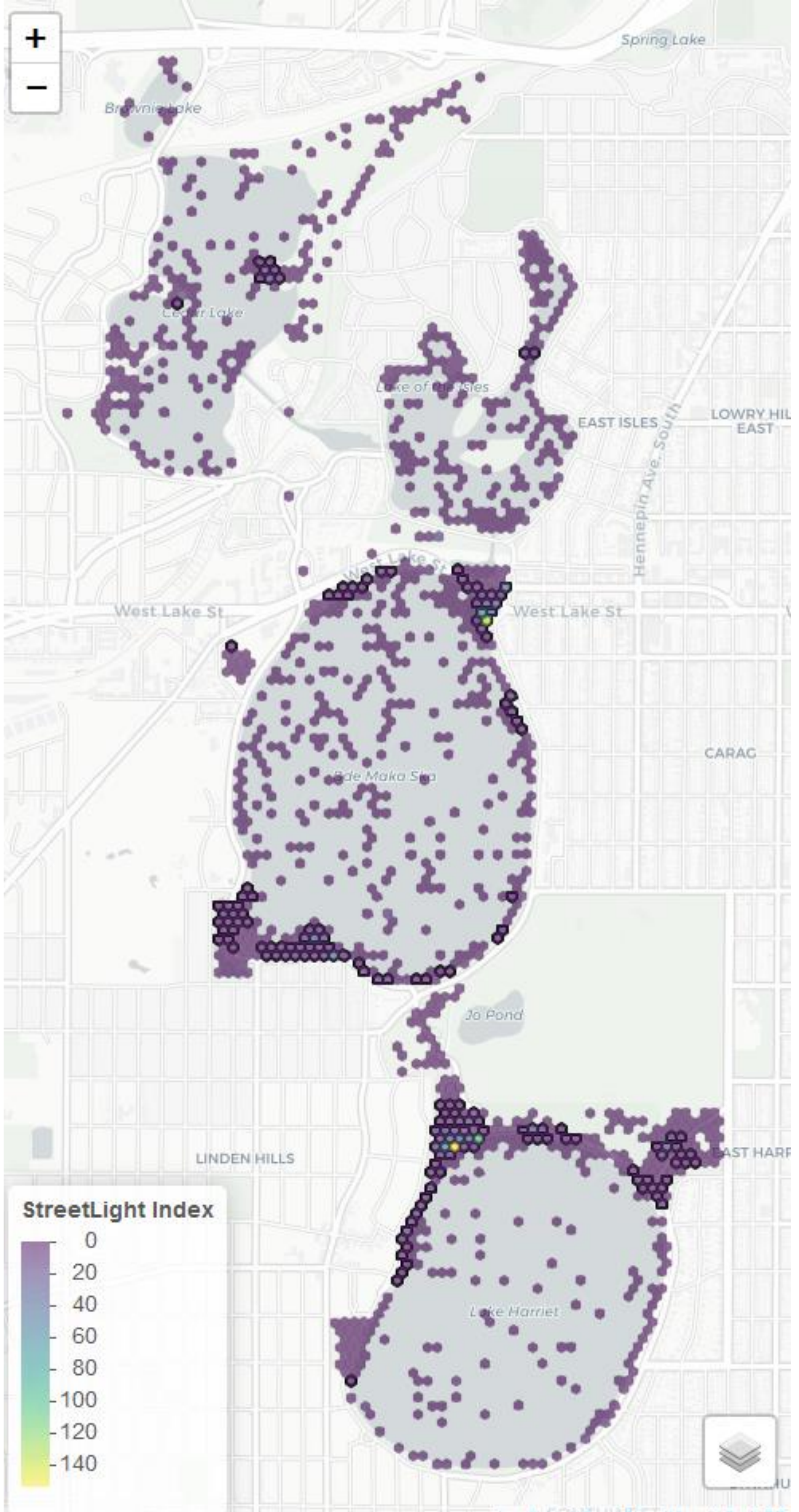
- Equivalent
- High-High
- High-Low
- Low-High
- Low-Low

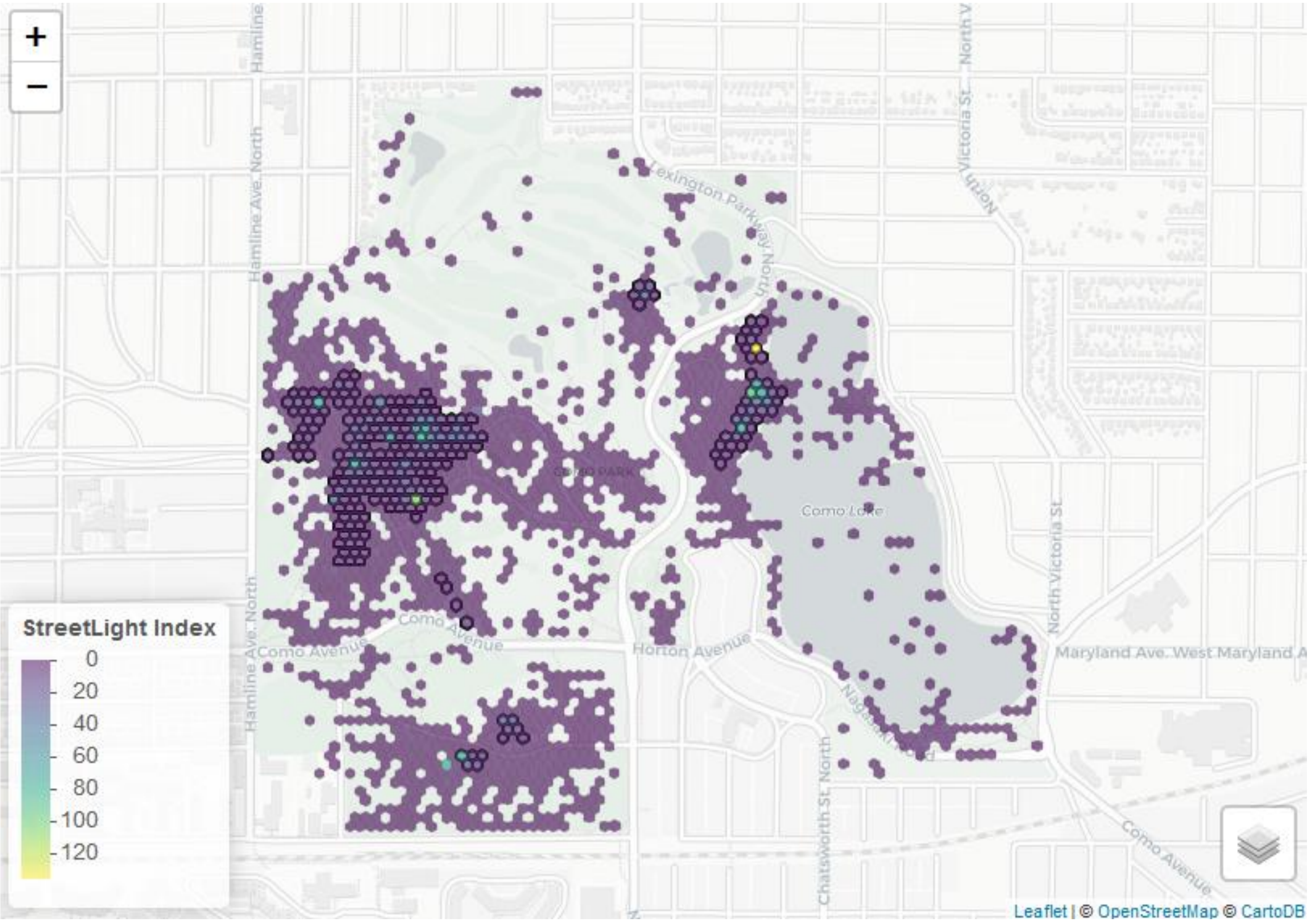
METROPOLITAN COUNCIL

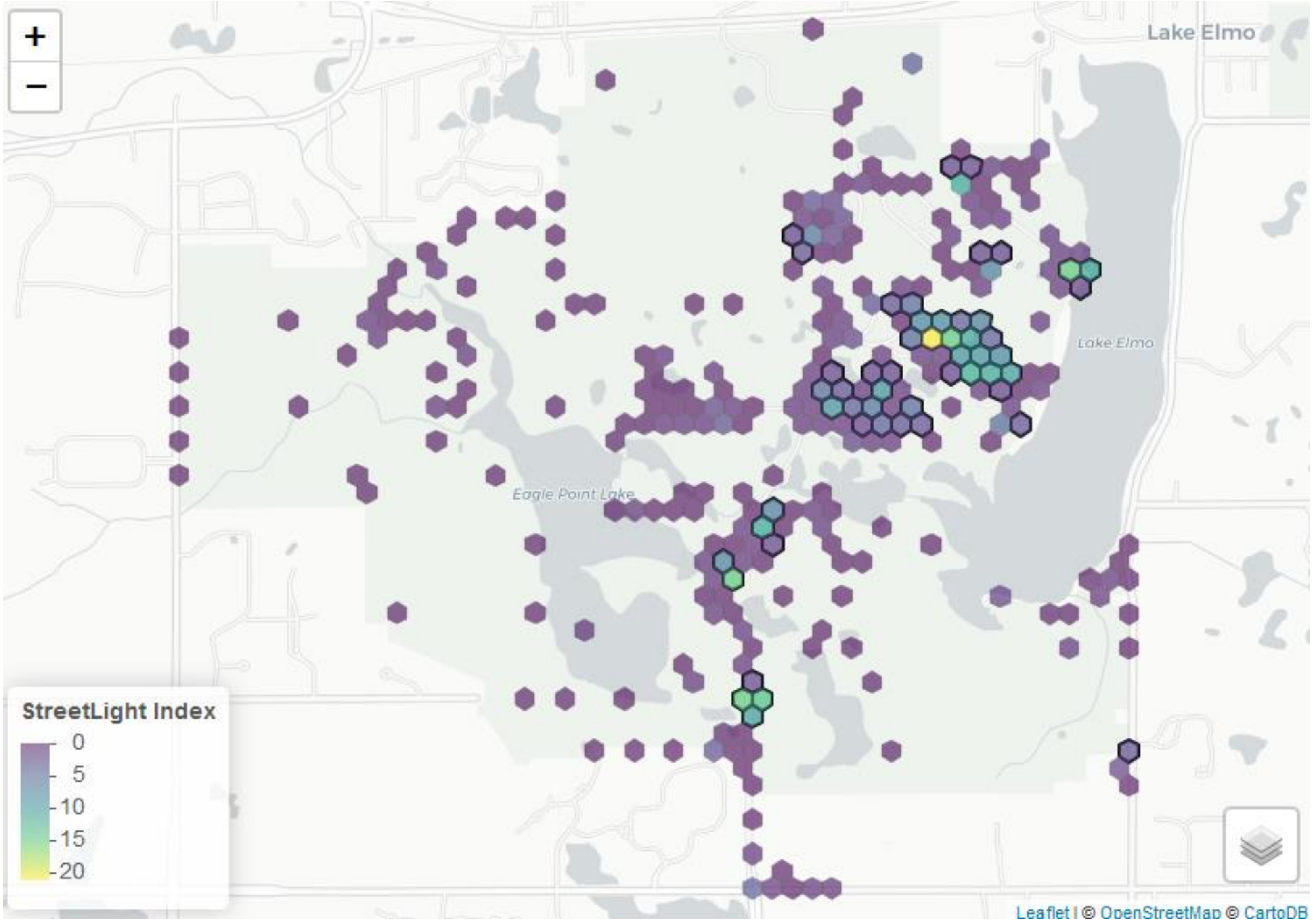


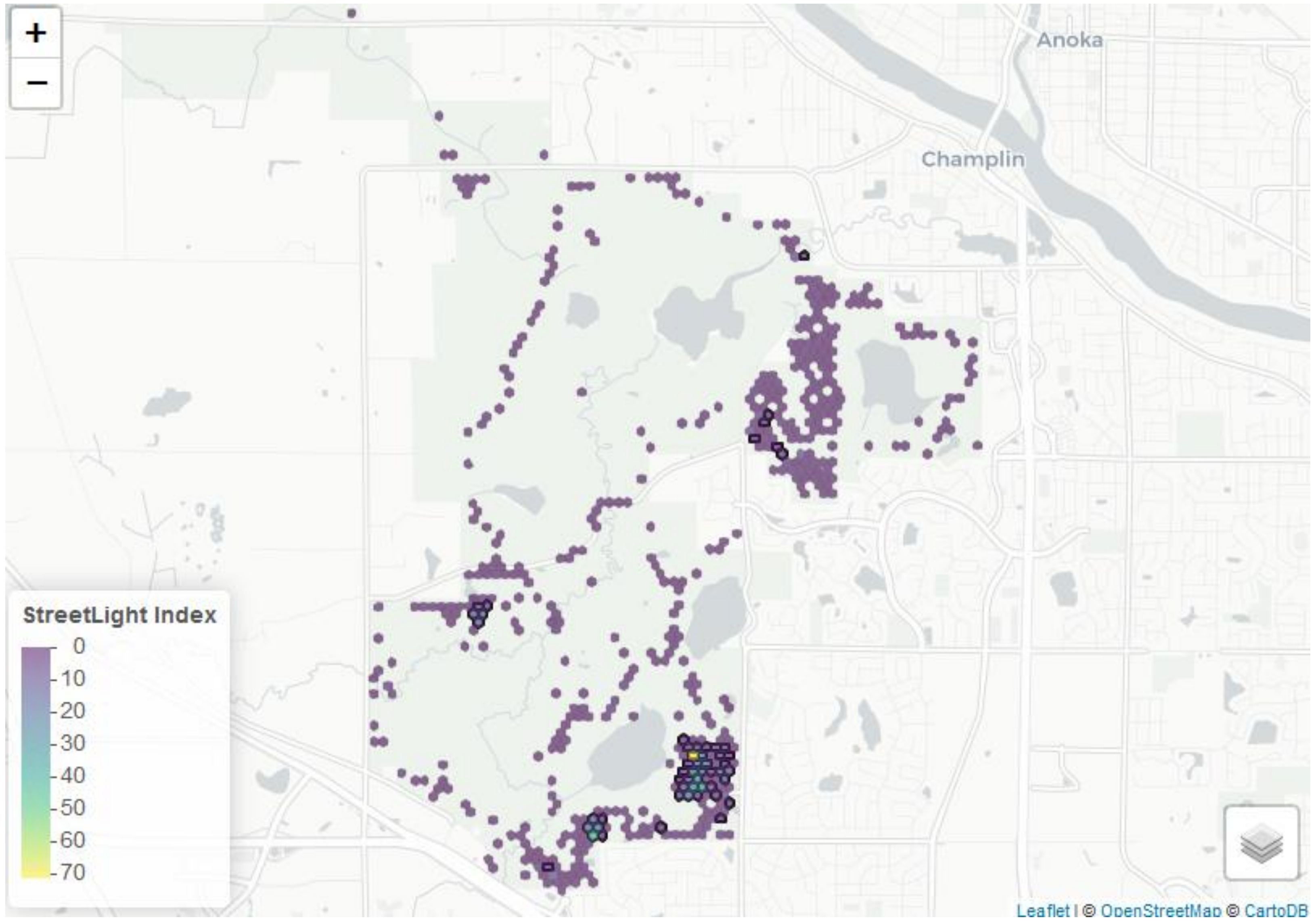
Pavilion and parking
History Center parking
Beach
Boat Launch
Mississippi St.
Rice Creek Road
Rice Creek
Lake Diana
Silver Lake Rd. Northwest
14th St. Northwest
Long Lake
Long Lake Road
Old Highway 8
County Highway 96
Old Highway 10
Palmer Drive
Old State Highway 10
Leaflet | © OpenStreetMap © CartoDB

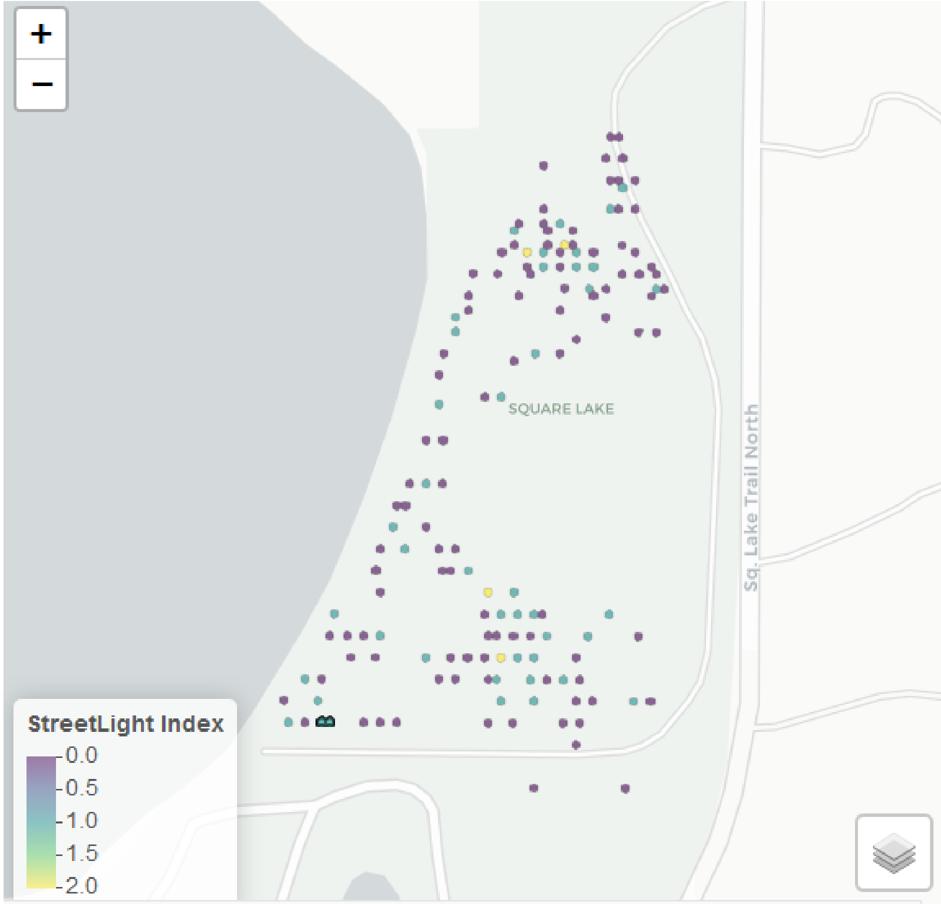
Now repeat!











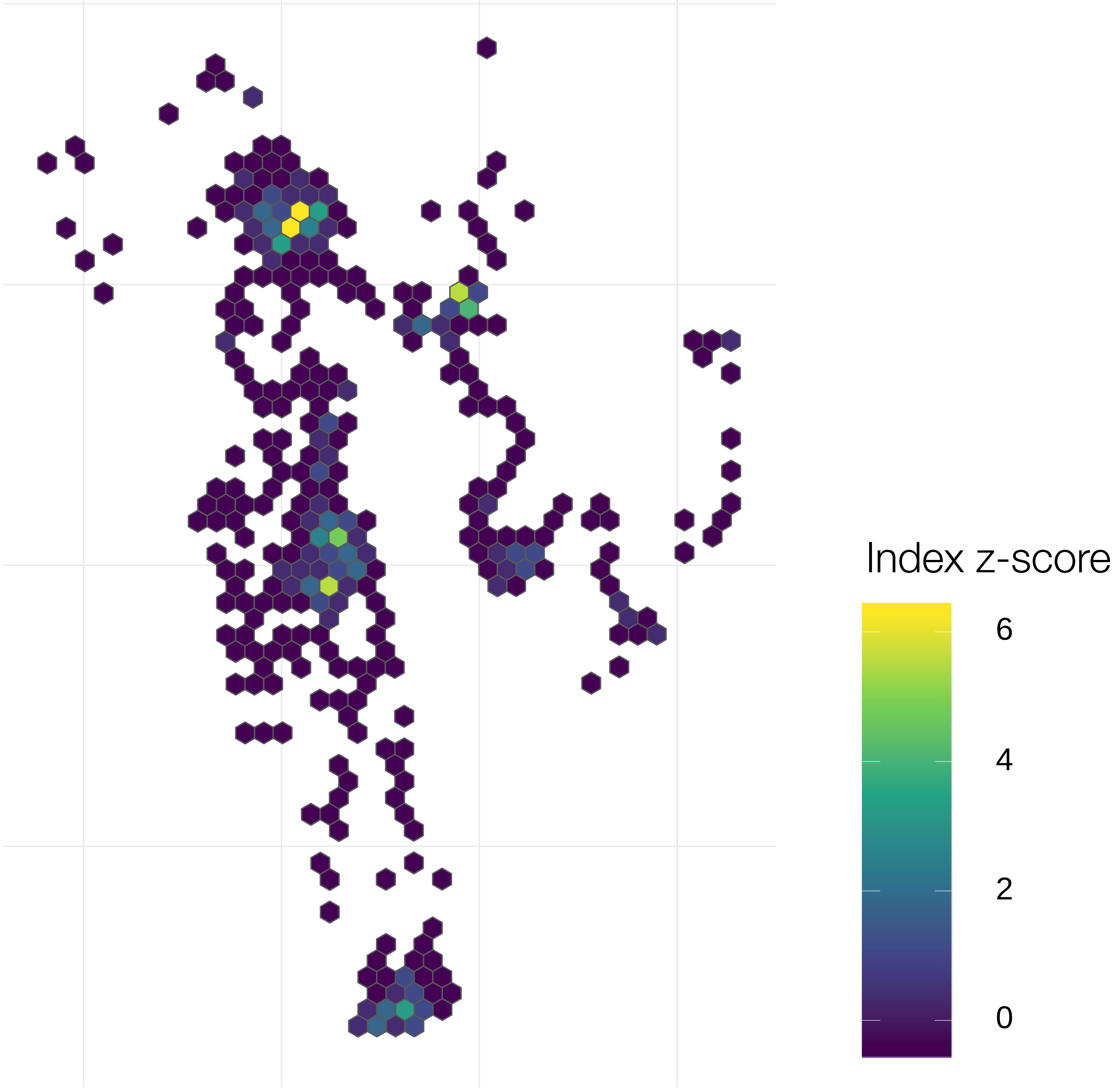
METROPOLITAN
COUNCIL

METROPOLITAN COUNCIL

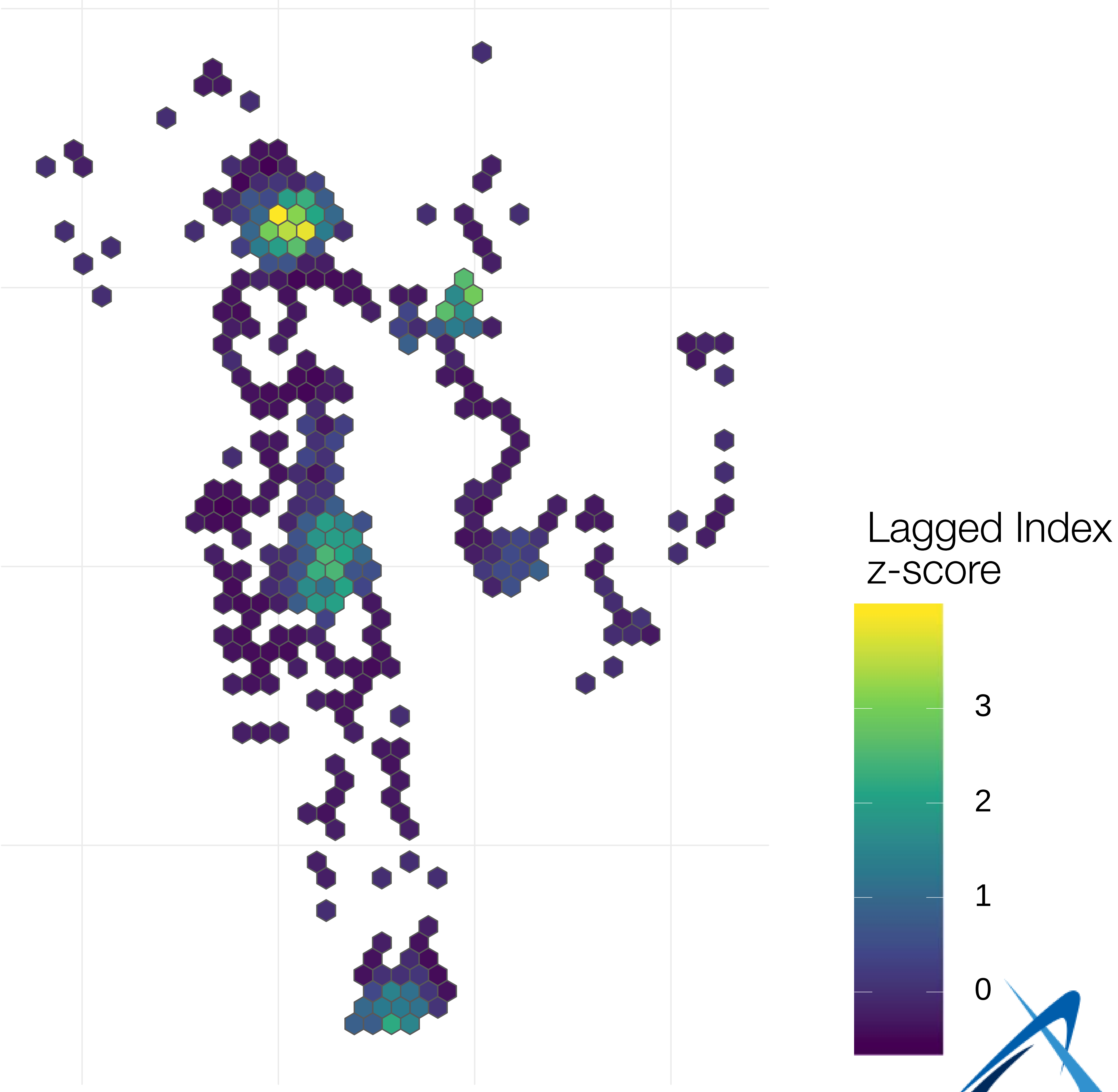
Liz Roten, Associate Data Scientist
elizabeth.roten@metc.state.mn.us
https://lizroten.info

Index and Lagged Index z-score

Observed StreetLight Index z-score



Lagged StreetLight Index z-score



METROPOLITAN COUNCIL

References and R packages

metrocouncil.org/data

Anselin, Luke. 2018. “Applications of Spatial Weights.” In *GeoDa: An Introduction to Spatial Data Analysis*, 1.12 ed. The Center for Spatial Data Science at The University of Chicago. https://geodacenter.github.io/workbook/4d_weights_applications/lab4d.html.

Arnold, Jeffrey B. 2019. *Ggthemes: Extra Themes, Scales and Geoms for ’Ggplot2’*. https://CRAN.R-project.org/package=ggthemes.

Birch, Colin P. D., Sander P. Oom, and Jonathan A. Beecham. 2007. “Rectangular and Hexagonal Grids Used for Observation, Experiment and Simulation in Ecology.” *Ecological Modelling* 206 (3): 347–59. https://doi.org/10.1016/j.ecolmodel.2007.03.041.

Bivand, Roger. 2019. *Spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. https://CRAN.R-project.org/package=spdep.

Bivand, Roger, Tim Keitt, and Barry Rowlingson. 2019. *Rgdal: Bindings for the ’Geospatial’ Data Abstraction Library*. https://CRAN.R-project.org/package=rgdal.

Bivand, Roger, Jakub Nowosad, and Robin Lovelace. 2019. *SpData: Datasets for Spatial Analysis*. https://CRAN.R-project.org/package=spData.

Caldas de Castro, Marcia, and Burton H Singer. 2006. “Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association.” *Geographical Analysis* 38 (2): 180–208.

Chapman McGrew, J., Jr, Arthur Lembo Jr, and Charles B. Monroe. 2014. *An Introduction to Statistical Problem Solving in Geography*. 3rd ed. Waveland Press, Inc.

Cheng, Joe, Bhaskar Karambelkar, and Yihui Xie. 2018. *Leaflet: Create Interactive Web Maps with the Javascript ’Leaflet’ Library*. https://CRAN.R-project.org/package=leaflet.

Dowle, Matt, and Arun Srinivasan. 2019. *Data.table: Extension of ‘Data.frame‘*. https://CRAN.R-project.org/package=data.table.

Firke, Sam. 2019. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gandrud, Christopher. 2016. *Repmis: Miscellaneous Tools for Reproducible Research*. https://CRAN.R-project.org/package=repmis.

Gimond, Manuel. 2019. *Intro to GIS and Spatial Analysis*. Colby College. https://mgimond.github.io/Spatial/index.html.

Grossenbacher, Timo. 2019. “Bivariate Maps with Ggplot2 and Sf Timo Grossenbacher.” Timo Grossenbacher. *Bivariate Maps with Ggplot2 and Sf*. https://timogrossenbacher.ch/2019/04/bivariate-maps-with-ggplot2-and-sf/.

Kyle Walker. 2016. “Spatial Neighbors in R - an Interactive Illustration.” *KYLE WALKER DATA*. /2016/07/spatial-neighbors-in-r---an-interactive-illustration/.

Maxwell B. Joseph. 2015. “Plotting Spatial Neighbors in Ggmap.” *R-Bloggers*. https://www.r-bloggers.com/plotting-spatial-neighbors-in-ggmap/.

Pebesma, Edzer. 2019. *Sf: Simple Features for R*. https://CRAN.R-project.org/package=sf.

Pebesma, Edzer, and Roger Bivand. 2018. *Sp: Classes and Methods for Spatial Data*. https://CRAN.R-project.org/package=sp.

Qiu, Yixuan, and authors/contributors of the included fonts. See file AUTHORS for details. 2017. *Showtextdb: Font Files for the ’Showtext’ Package*. https://CRAN.R-project.org/package=showtextdb.

———. 2018. *Sysfonts: Loading Fonts into R*. https://CRAN.R-project.org/package=sysfonts.

Qiu, Yixuan, and authors/contributors of the included software. See file AUTHORS for details. 2019. *Showtext: Using Fonts More Easily in R Graphs*. https://CRAN.R-project.org/package=showtext.

R Core Team. 2019a. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

———. 2019b. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Roger Bivand. 2017. “Creating Neighbours.” Vignette. *Spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. https://rdrr.io/rforge/spdep/f/inst/doc/nb.pdf.

Roten, Liz. 2019. *CdResearchTemplates: A Package for the Metropolitan Council Research Team*.

Roten, Liz, and Metropolitan Council. 2019. *StreetlightR: Functions for Processing, Plotting, and Mapping Streetlight and Related Data*.

Simon Hailstone. 2017. “Spatial Clustering of Census Data.” *RPubs*. https://rpubs.com/Hailstone/346625.

Tiefelsdorf, Michael, Daniel A Griffith, and Barry Boots. 1999. “A Variance-Stabilizing Coding Scheme for Spatial Link Matrices.” *Environment and Planning A* 31 (1): 165–80.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2019. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Lionel Henry. 2019. *Tidyr: Easily Tidy Data with ’Spread()’ and ’Gather()’ Functions*. https://CRAN.R-project.org/package=tidyr.

Zhu, Hao. 2019. *KableExtra: Construct Complex Table with ’Kable’ and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.

METROPOLITAN COUNCIL