esri | THE SCIENCE OF WHERE

Spatial Data Mining: Essentials of Cluster Analysis

Lauren Bennett, PhD

Jenora D'Acosta

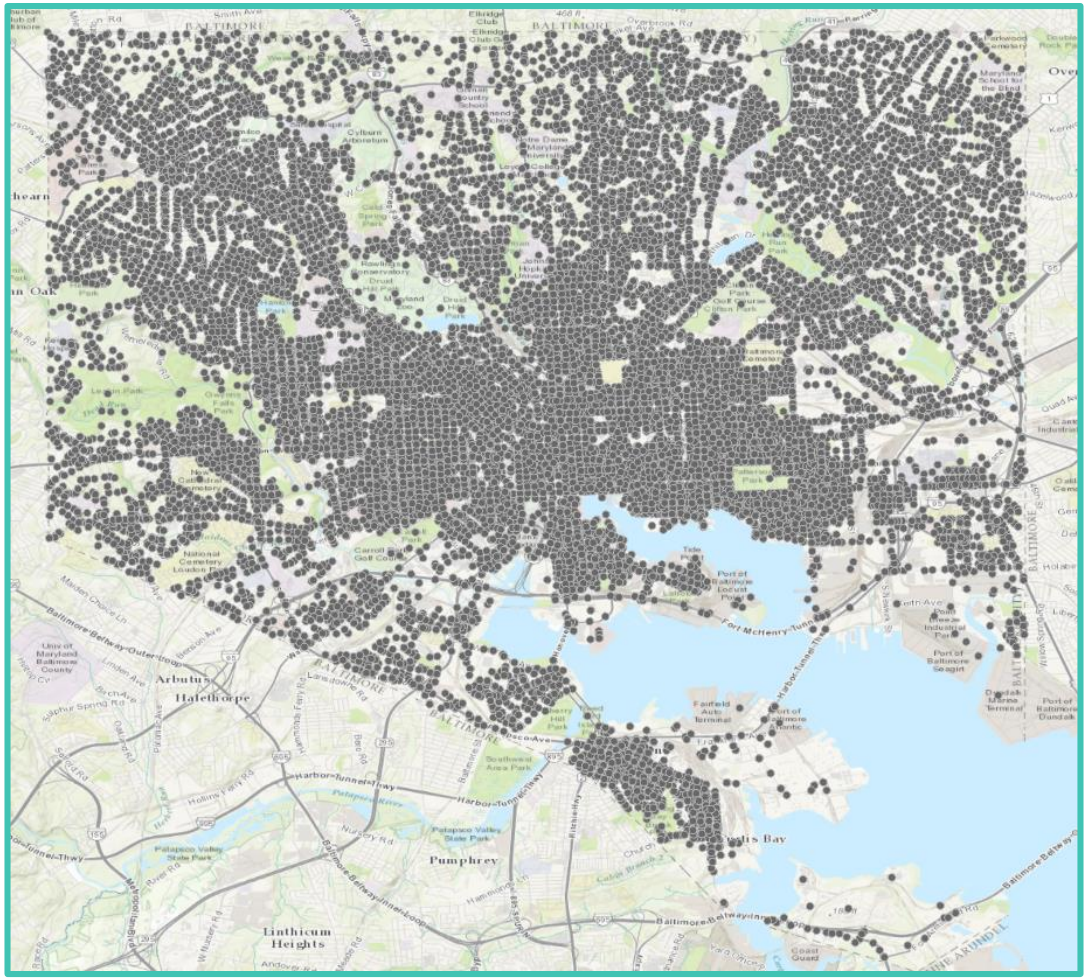
Flora Vale

esriurl.com/spatialstats

Subjectivity of Maps

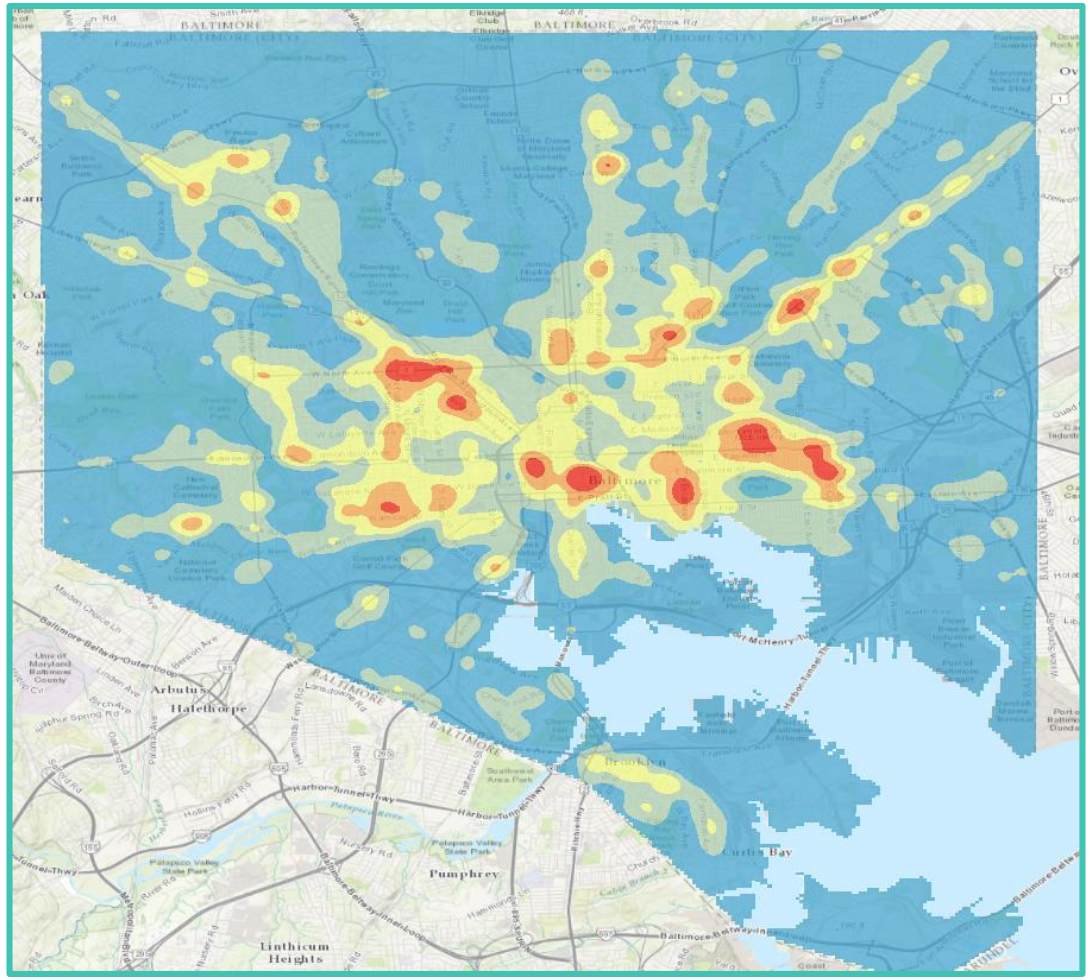
The map as data

High Priority 911 Calls in Baltimore



The map as data

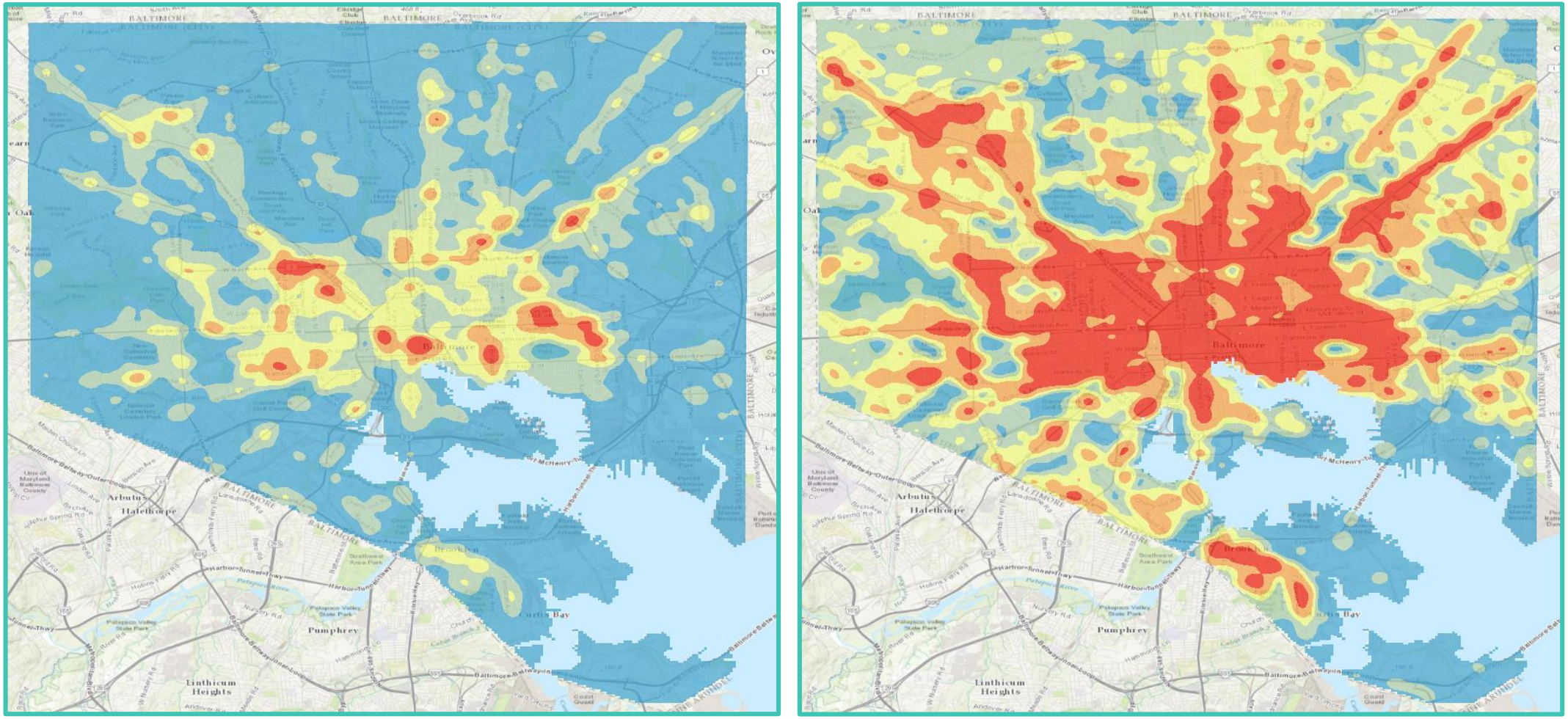
High Priority 911 Calls in Baltimore



Where are the hot spots? Where is the variation greater?

The map as data

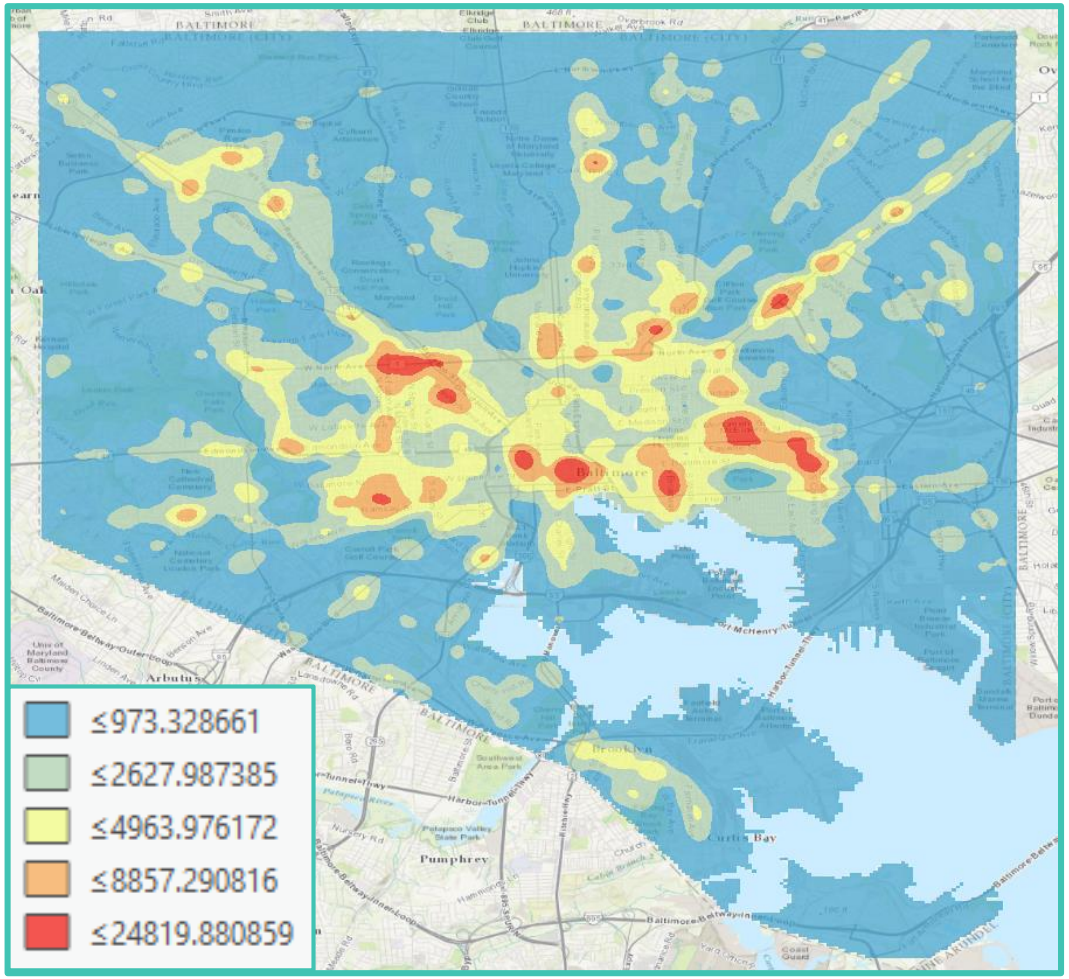
High Priority 911 Calls in Baltimore



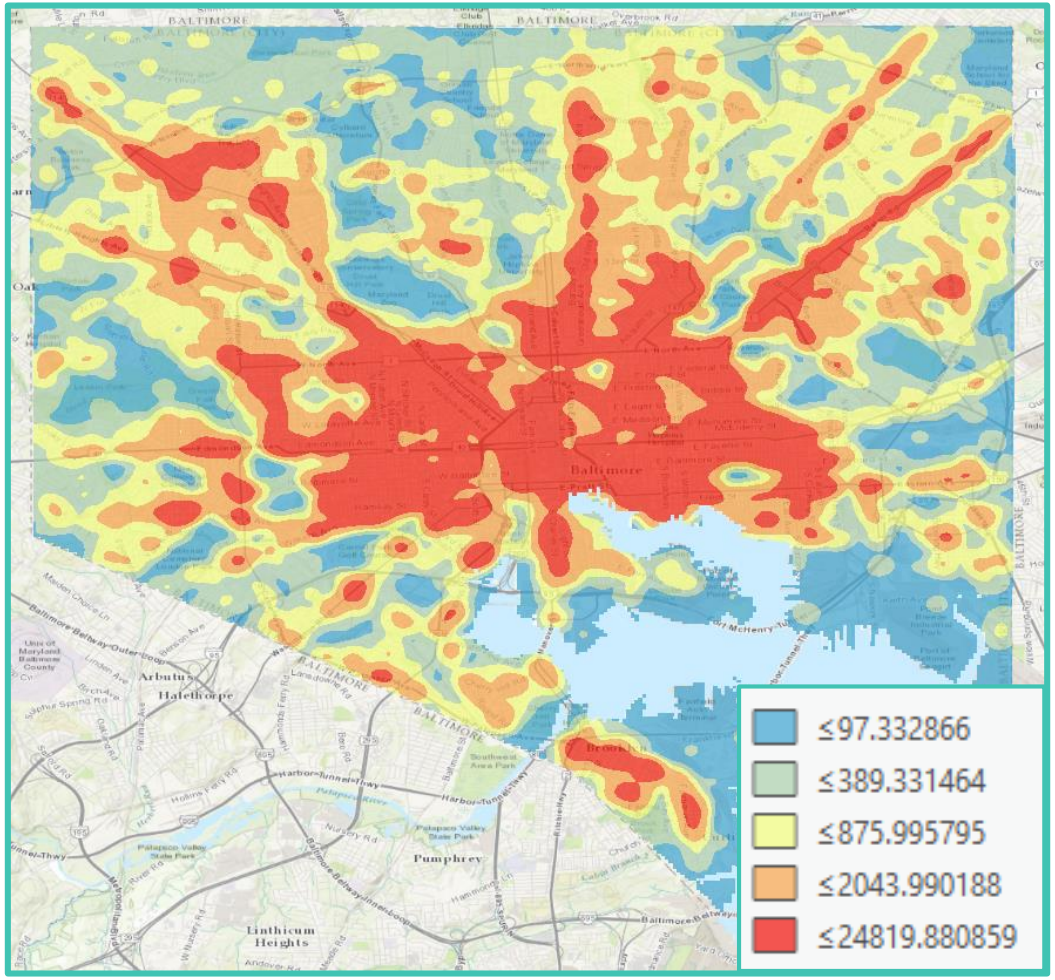
Where are the hot spots? Where is the variation greater?

The subjectivity of visual pattern analysis

Natural Breaks



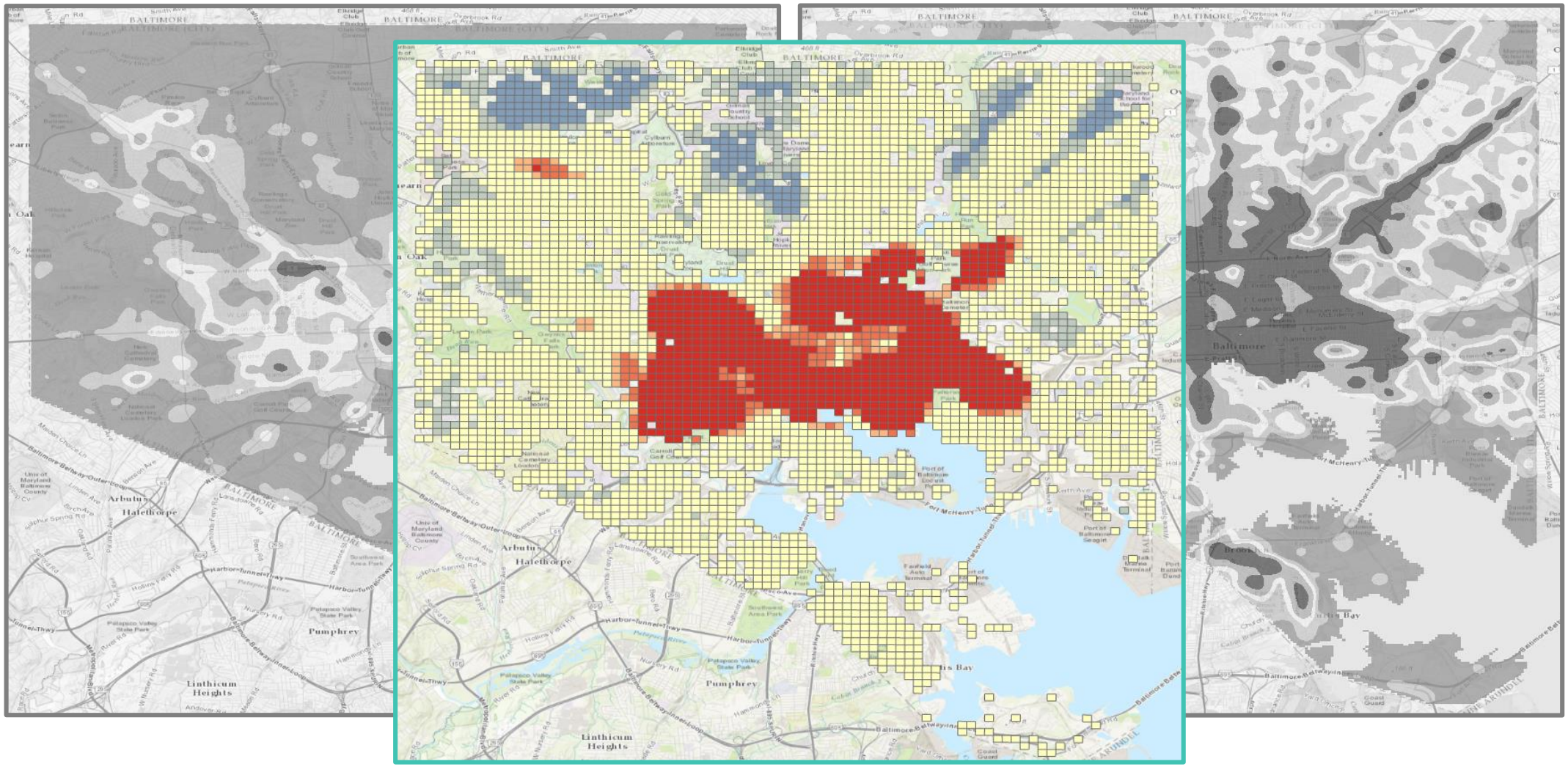
Quantile



Where are the hot spots? Where is the variation greater?

Minimizing the subjectivity

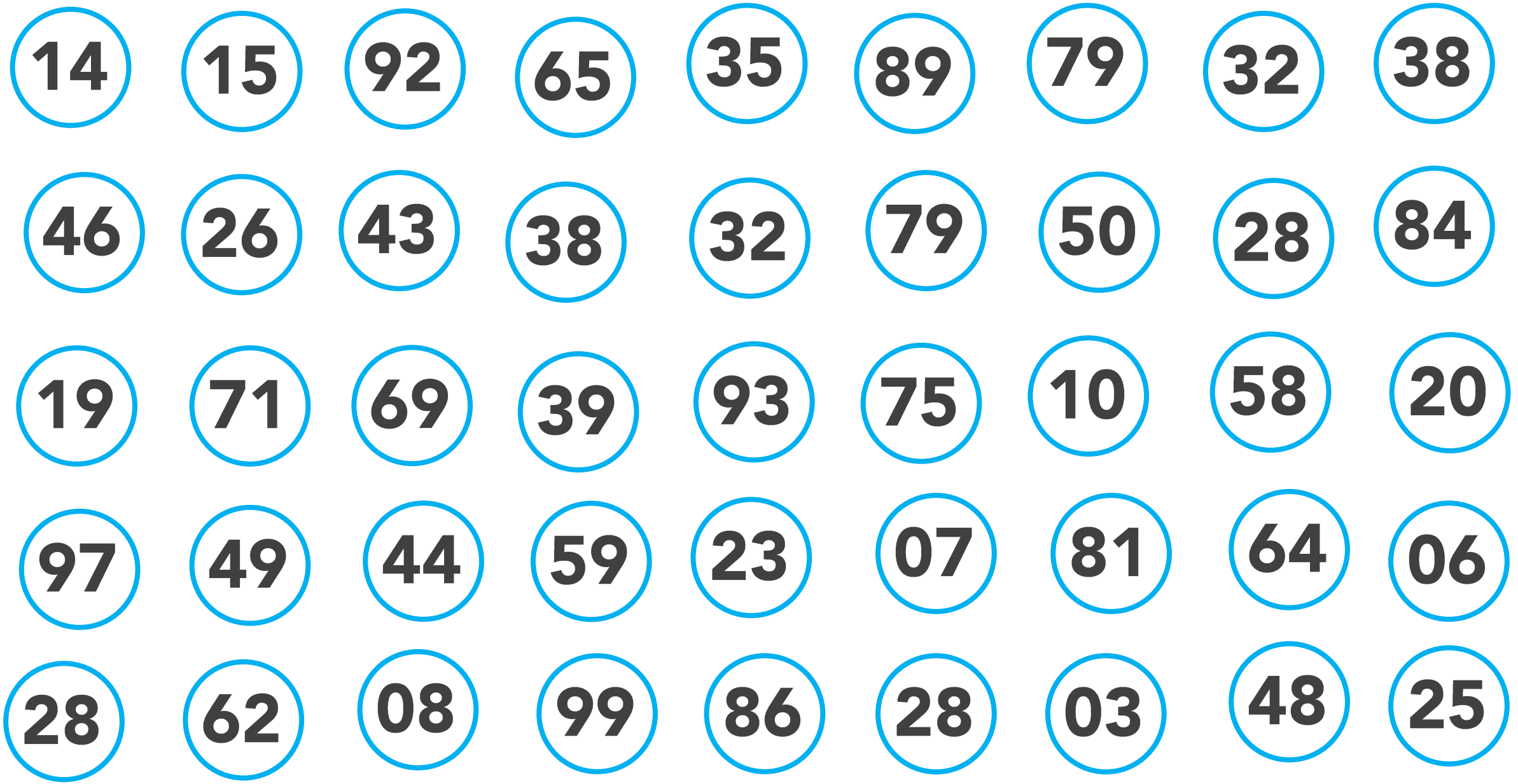
Turning the map into information

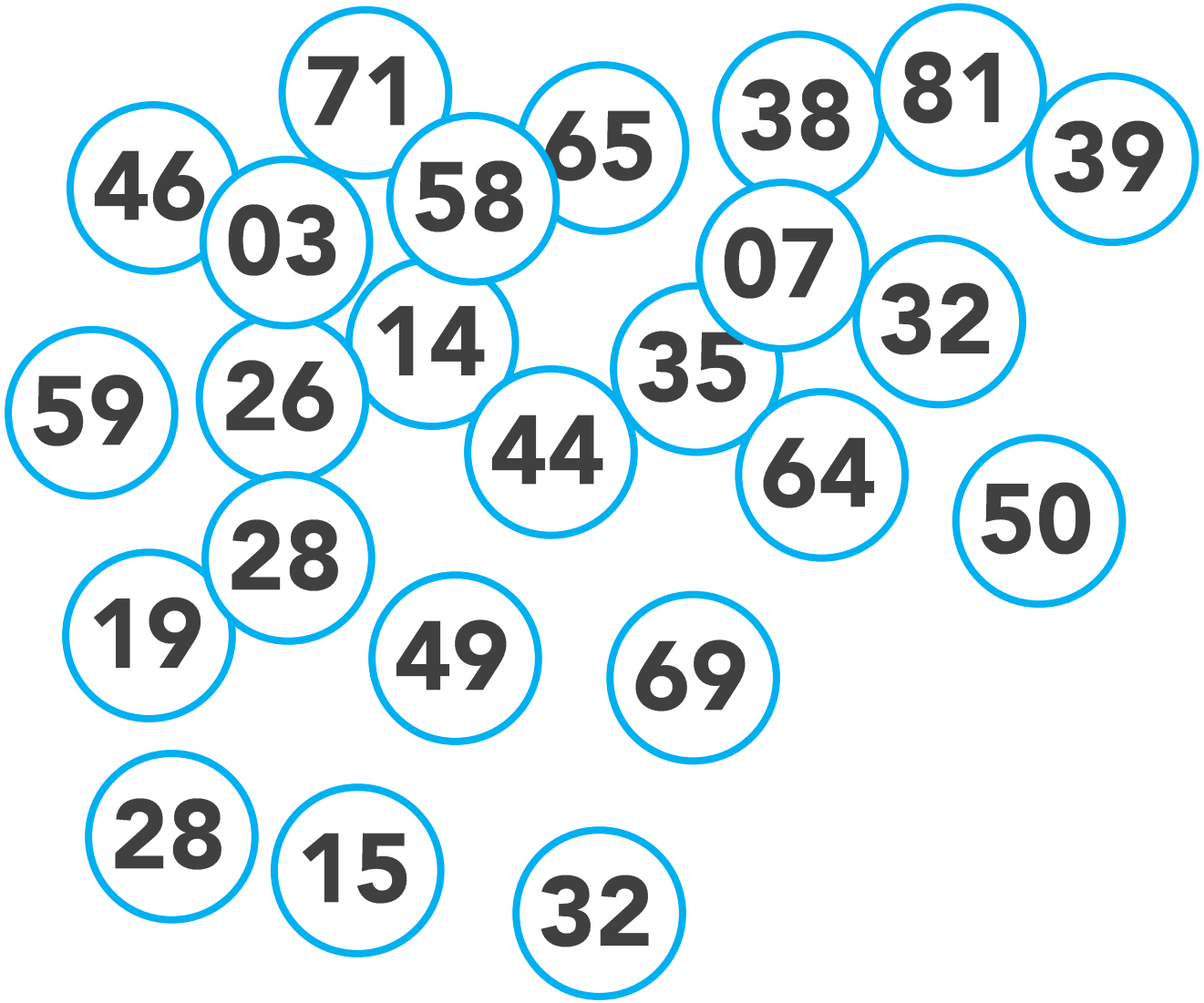


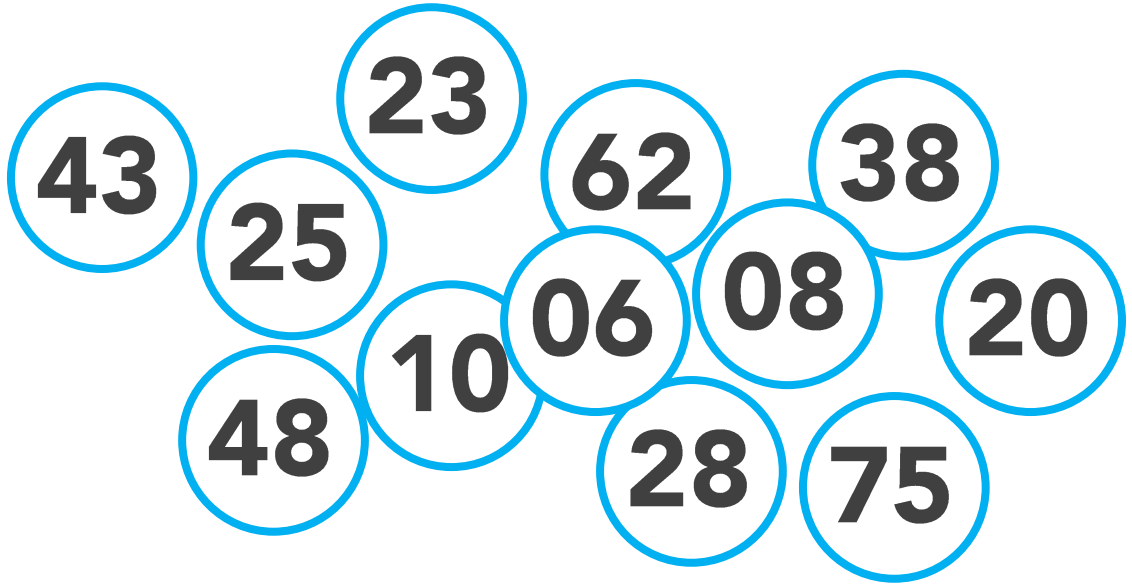
Inferential Statistics

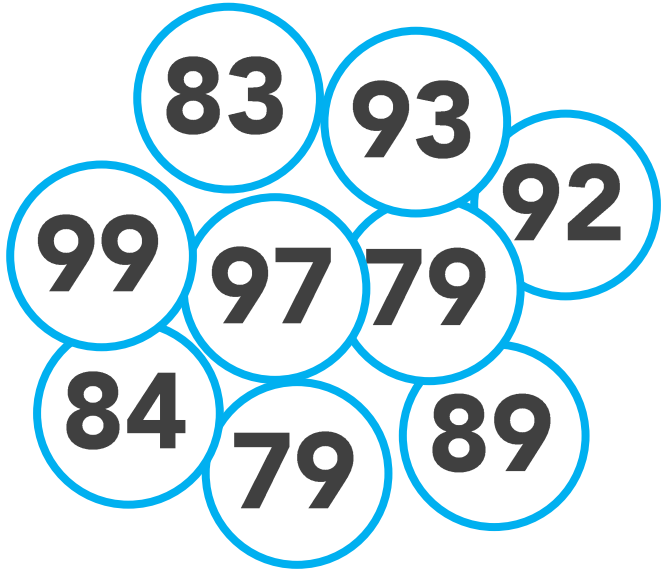

Complete Spatial
RANDOMNESS

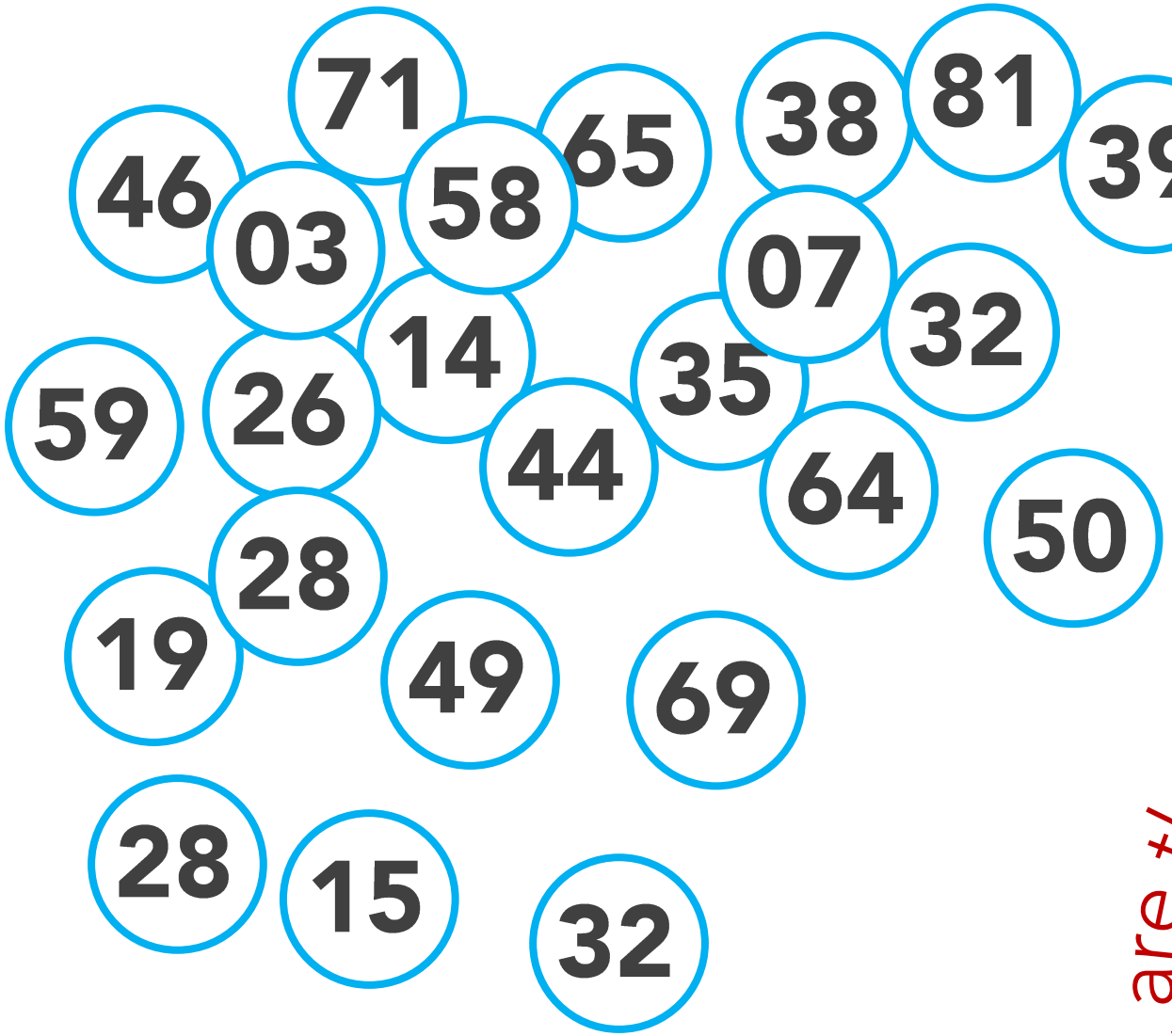
Is there a PATTERN?

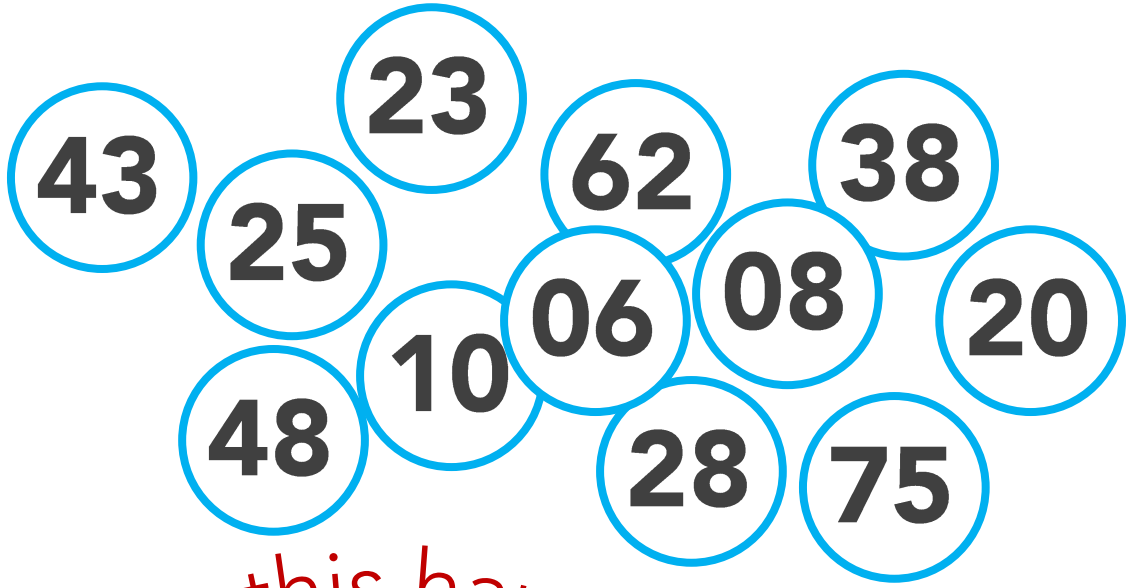

14 15 92 65 35 89 79 32 38
46 26 43 38 32 79 50 28 84
19 71 69 39 93 75 10 58 20
97 49 44 59 23 07 81 64 06
28 62 08 99 86 28 03 48 25

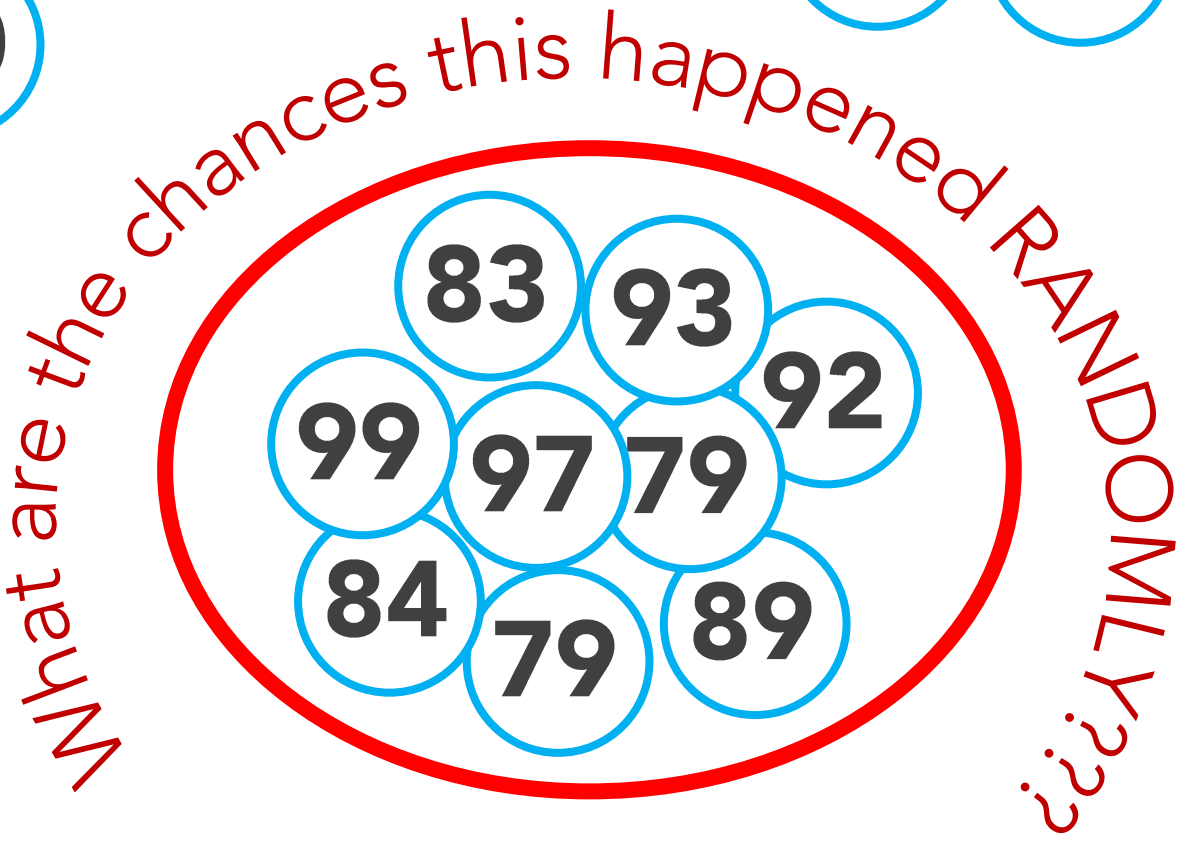

71
65
38
81
39
46
58
03
07
32
14
35
59
26
44
64
50
28
19
49
69
28
15
32


23
43
62
38
25
06
08
20
10
48
28
75


83
93
92
99
97
79
84
79
89

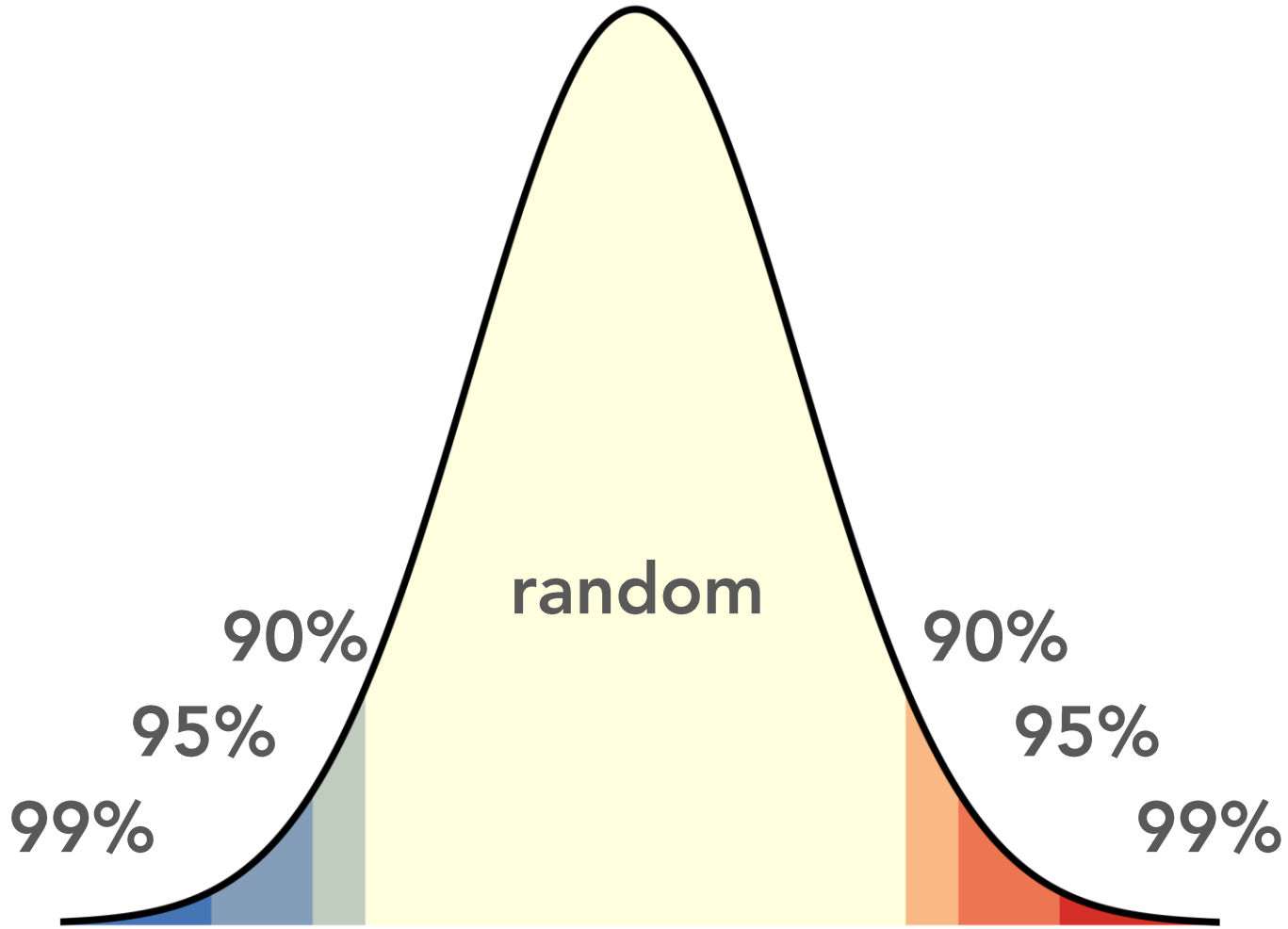

71
46
03
58
65
38
81
39
07
32
14
35
59
26
44
64
50
28
19
49
69
28
15
32

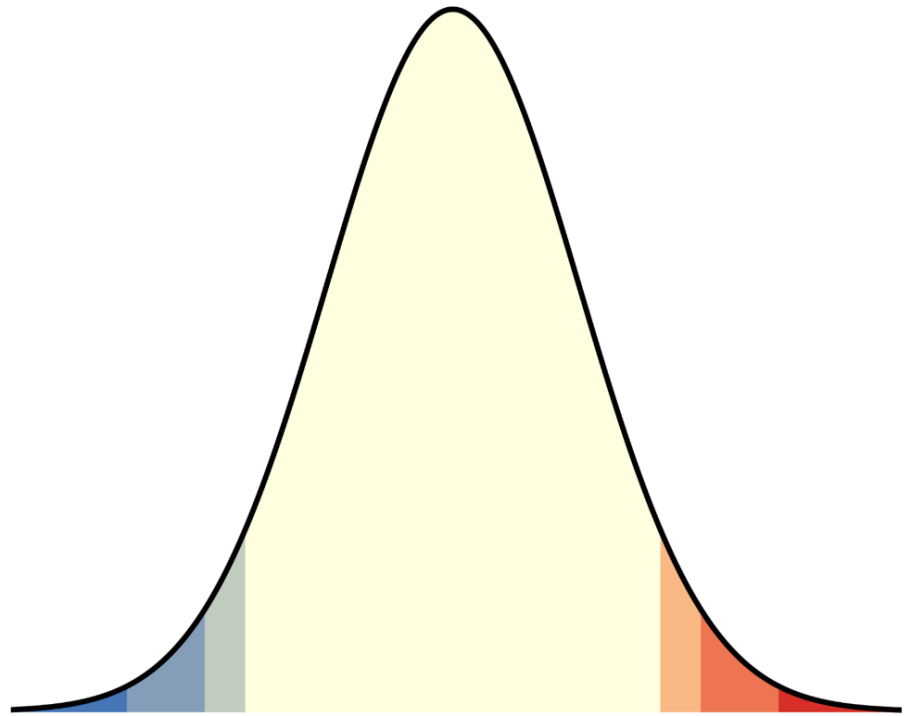

23
43
25
62
38
06
08
20
10
48
28
75


What are the chances this happened RANDOMLY???
83
93
92
99
97
79
84
79
89

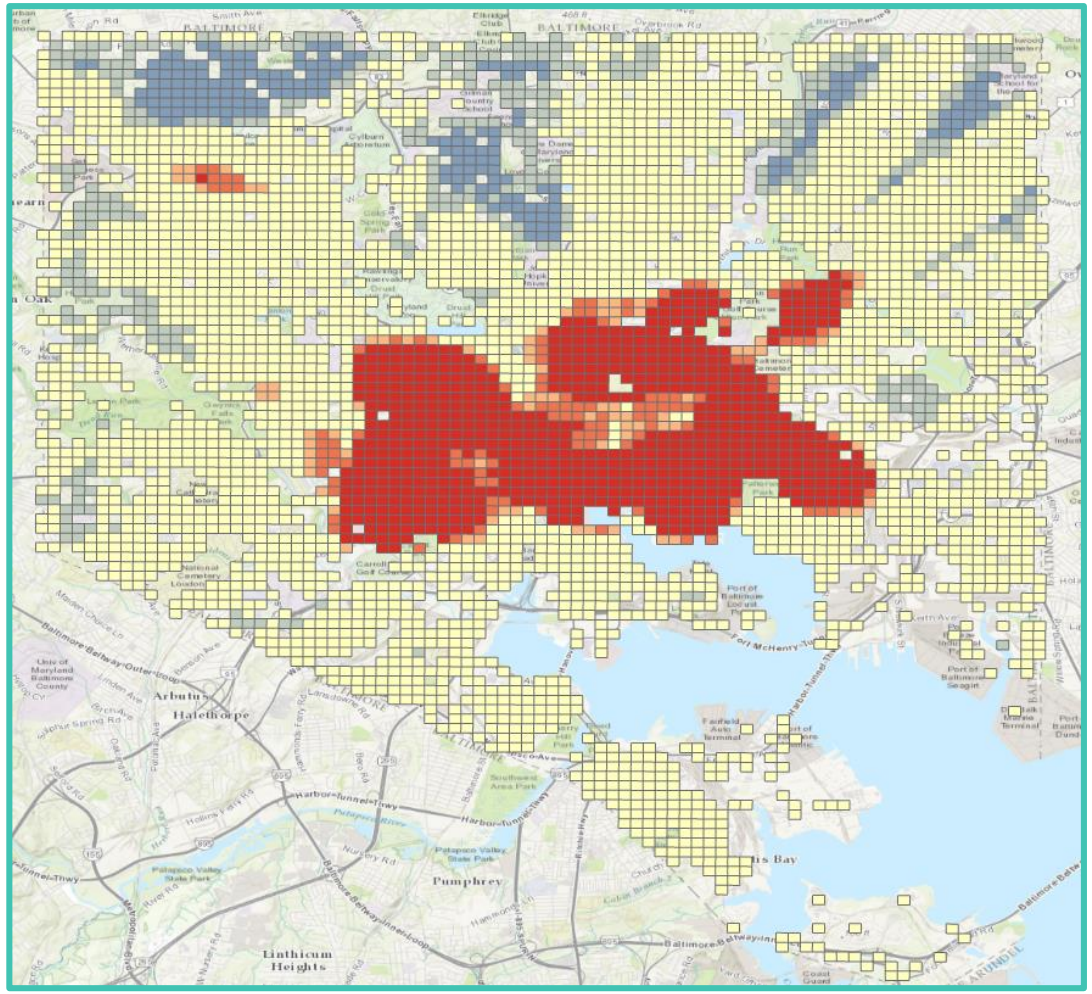
z-scores
p-values

z-scores and p-values



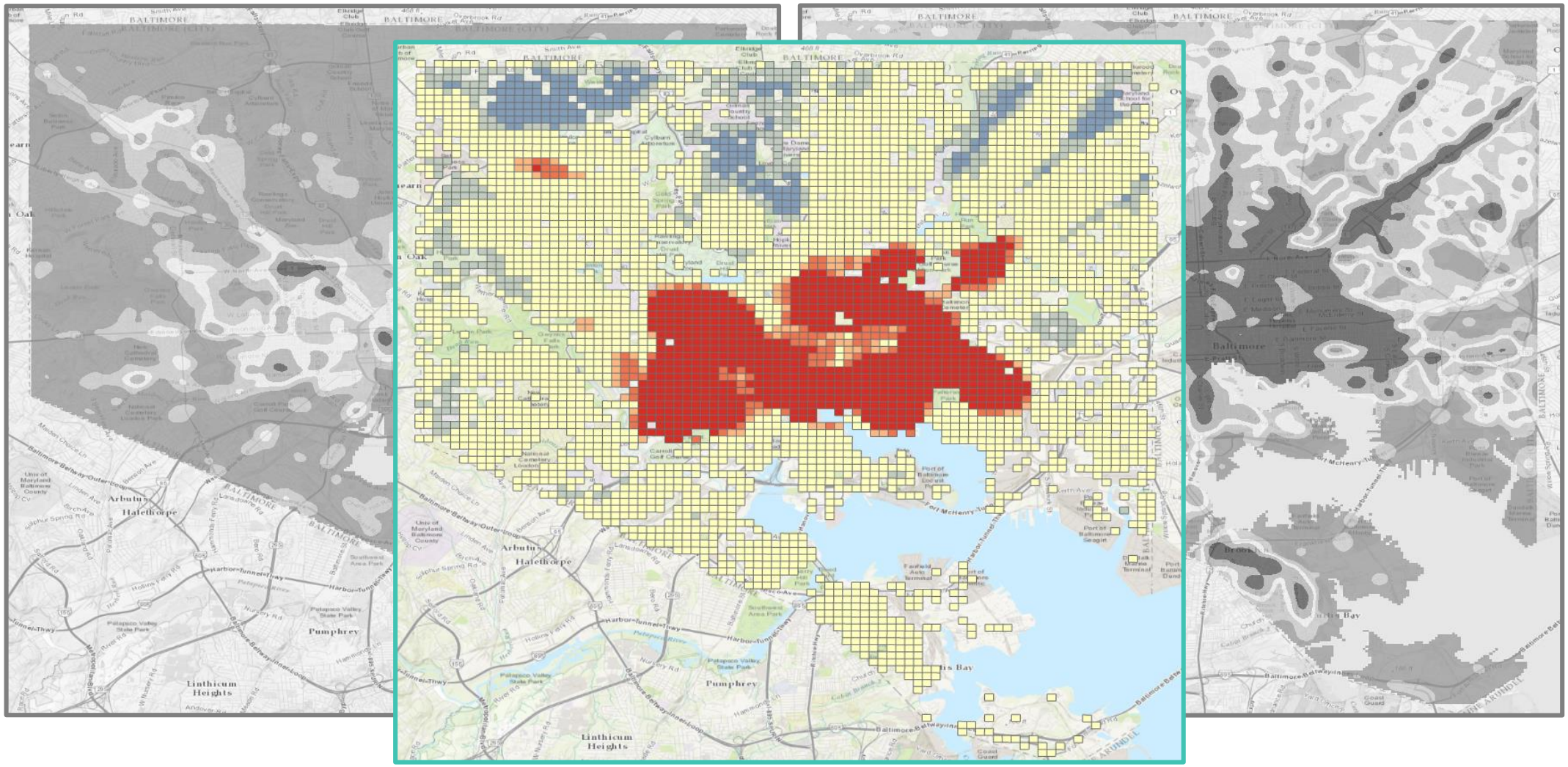


z-scores	...-2.58	-1.96	-1.65	0	1.65	1.96	2.58...
p-values	...0.01	0.05	0.10	0	0.10	0.05	0.01...



Minimizing the subjectivity

Turning the map into information

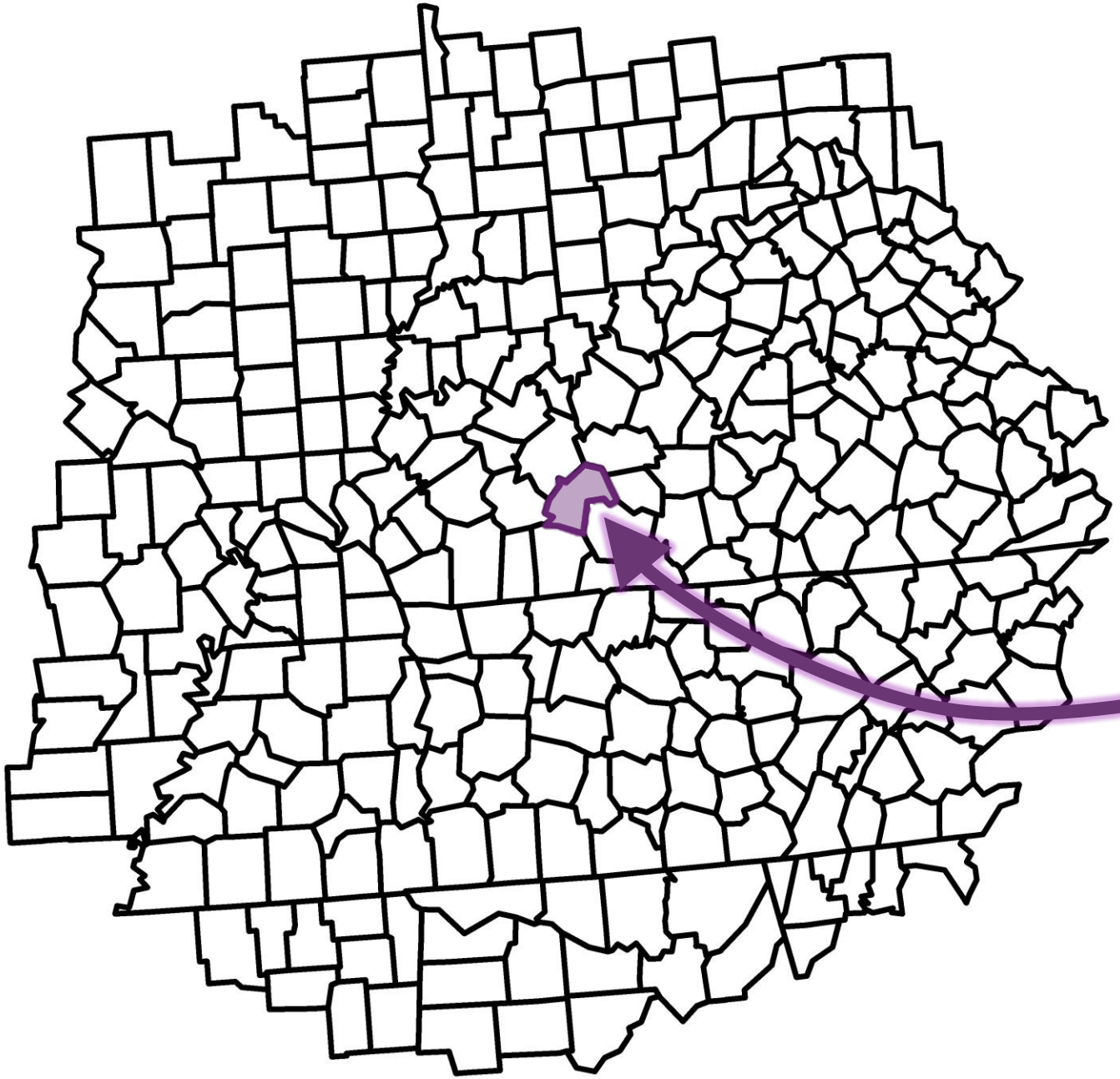


"...everything is related to everything else, but near things are more related than distant things."

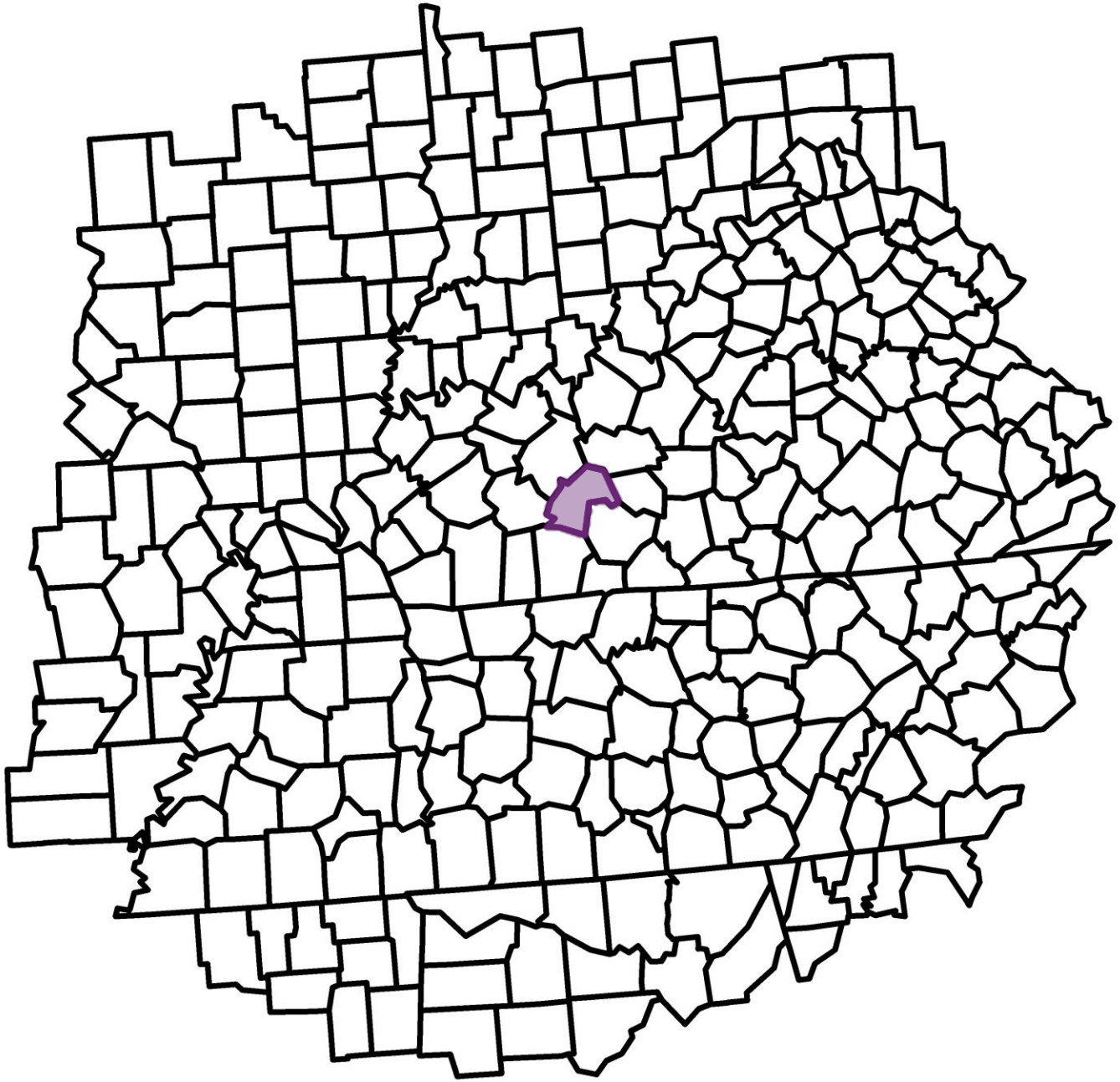
Hot Spot Analysis

given a set of weighted features, identifies statistically significant hot spots and cold spots using the Getis-Ord Gi* statistic

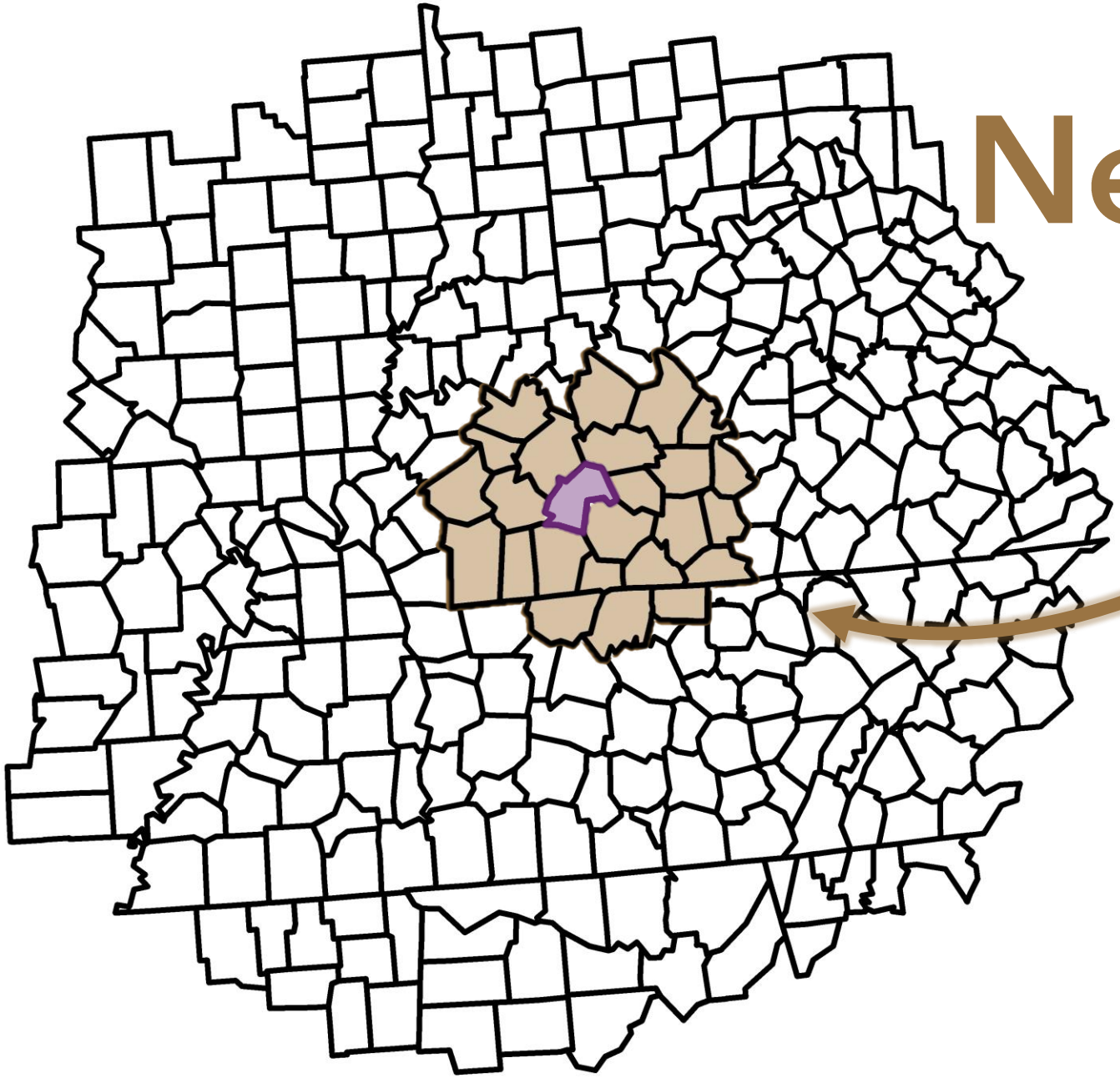
Polygons



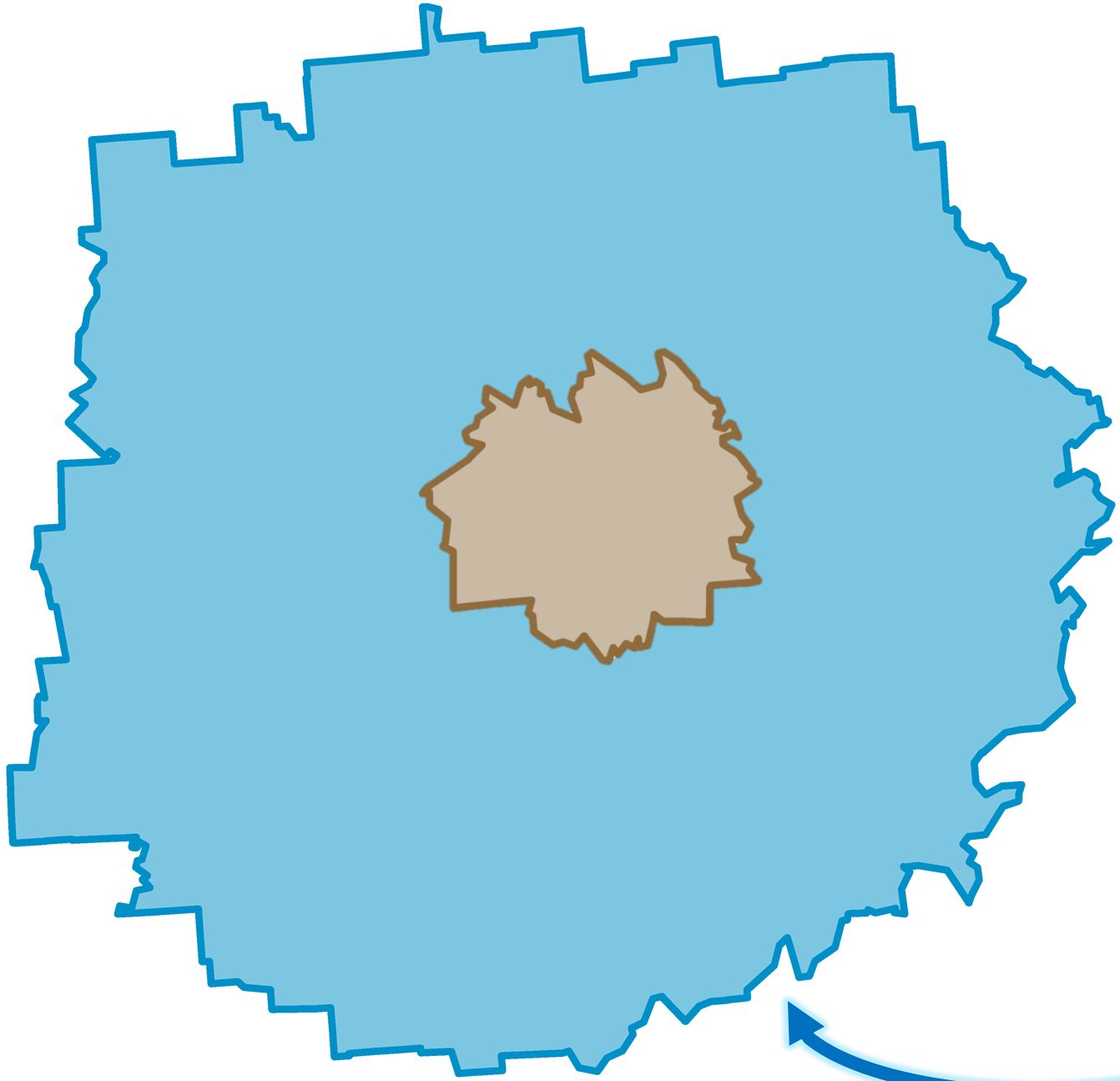
feature



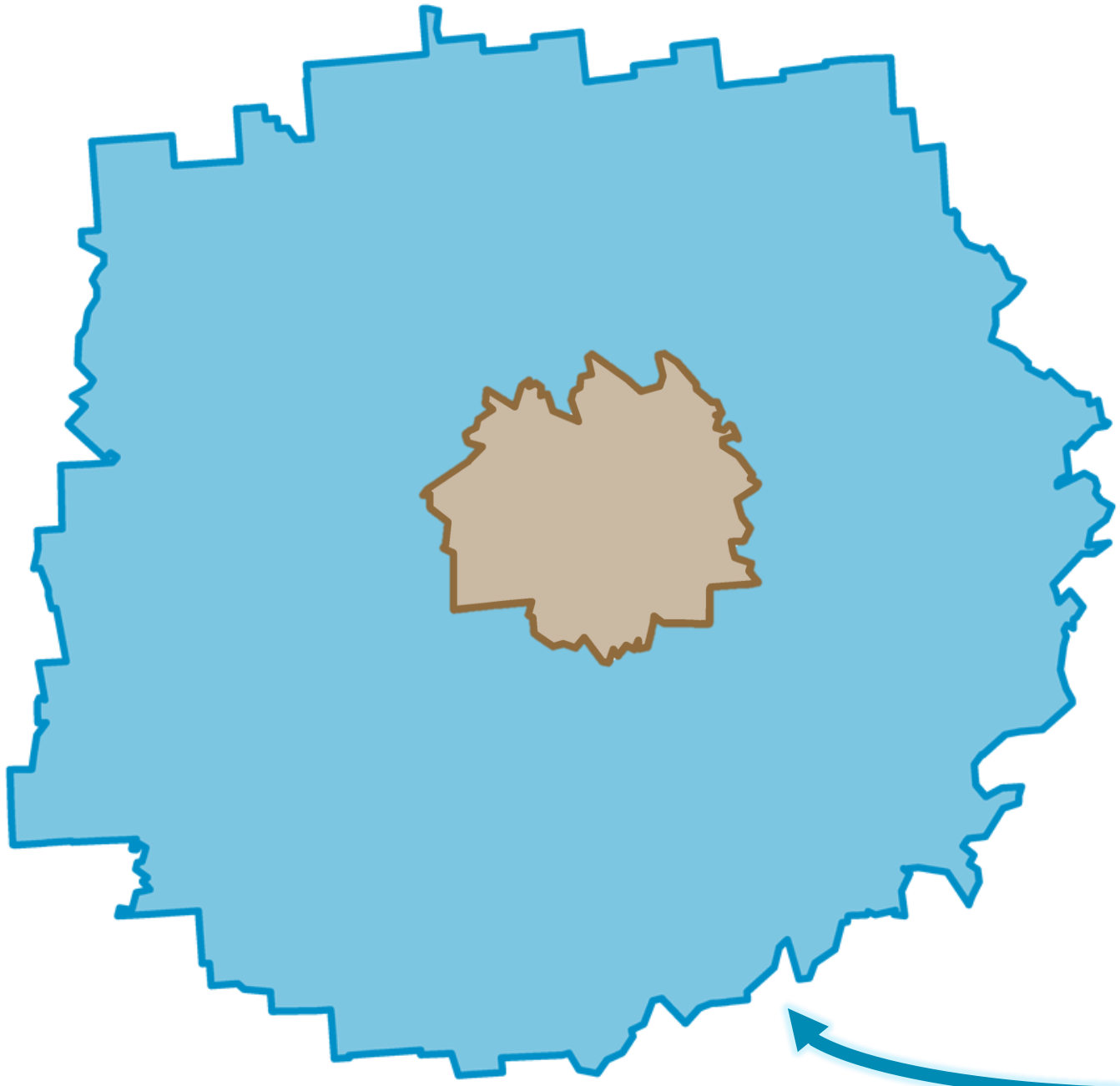
each feature has a **value**



Neighborhood

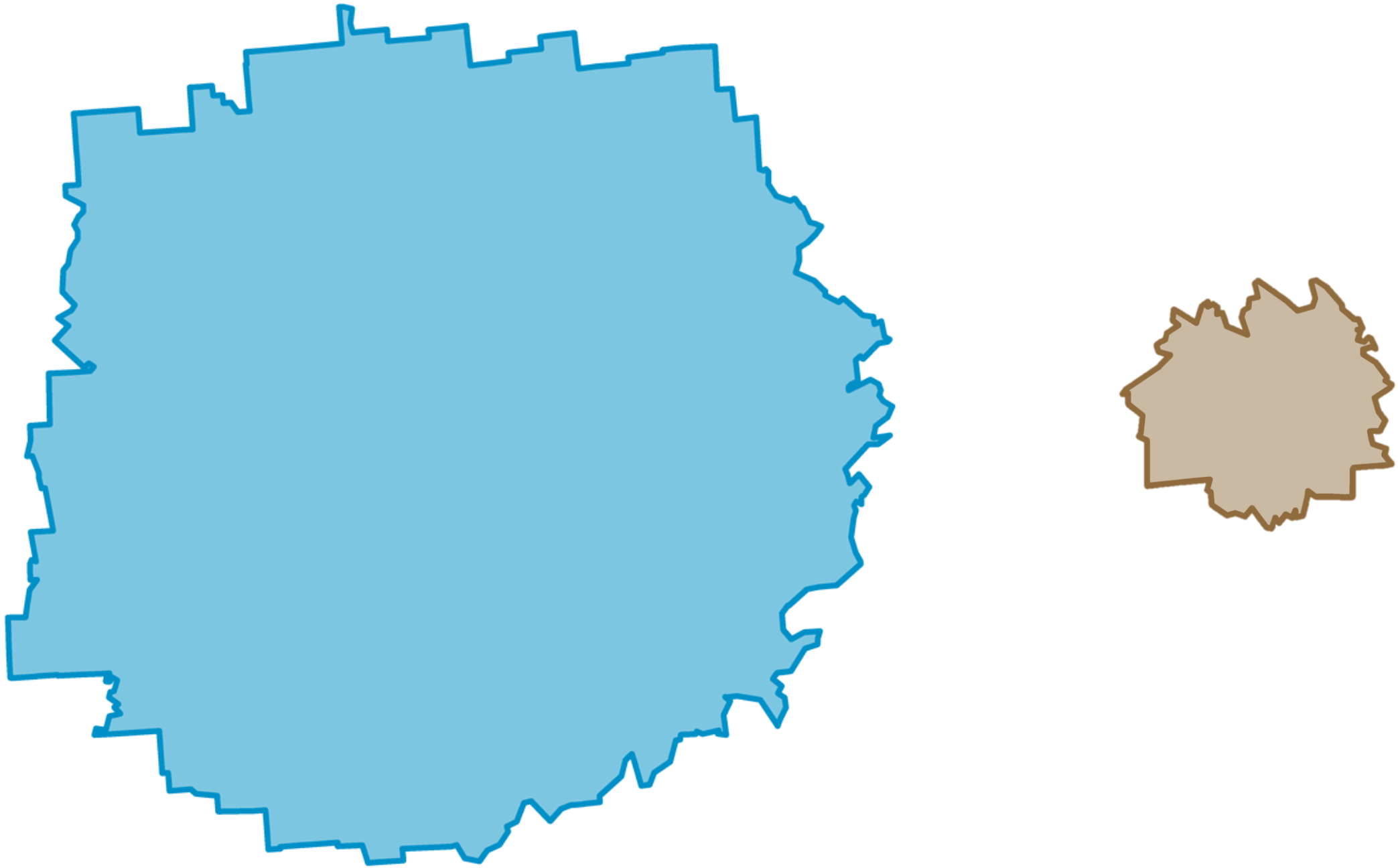


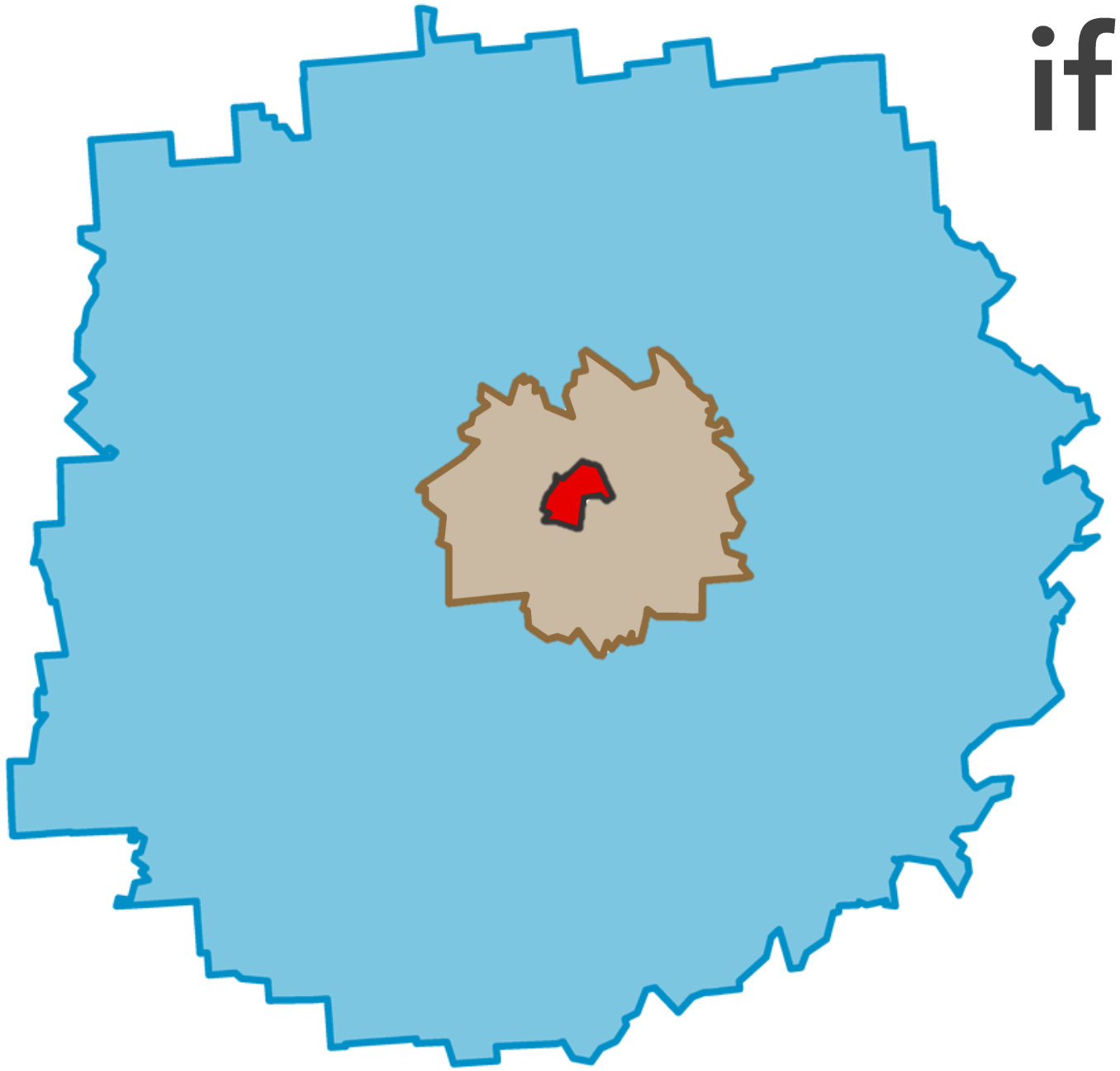
Study Area



is this

significantly different from this?





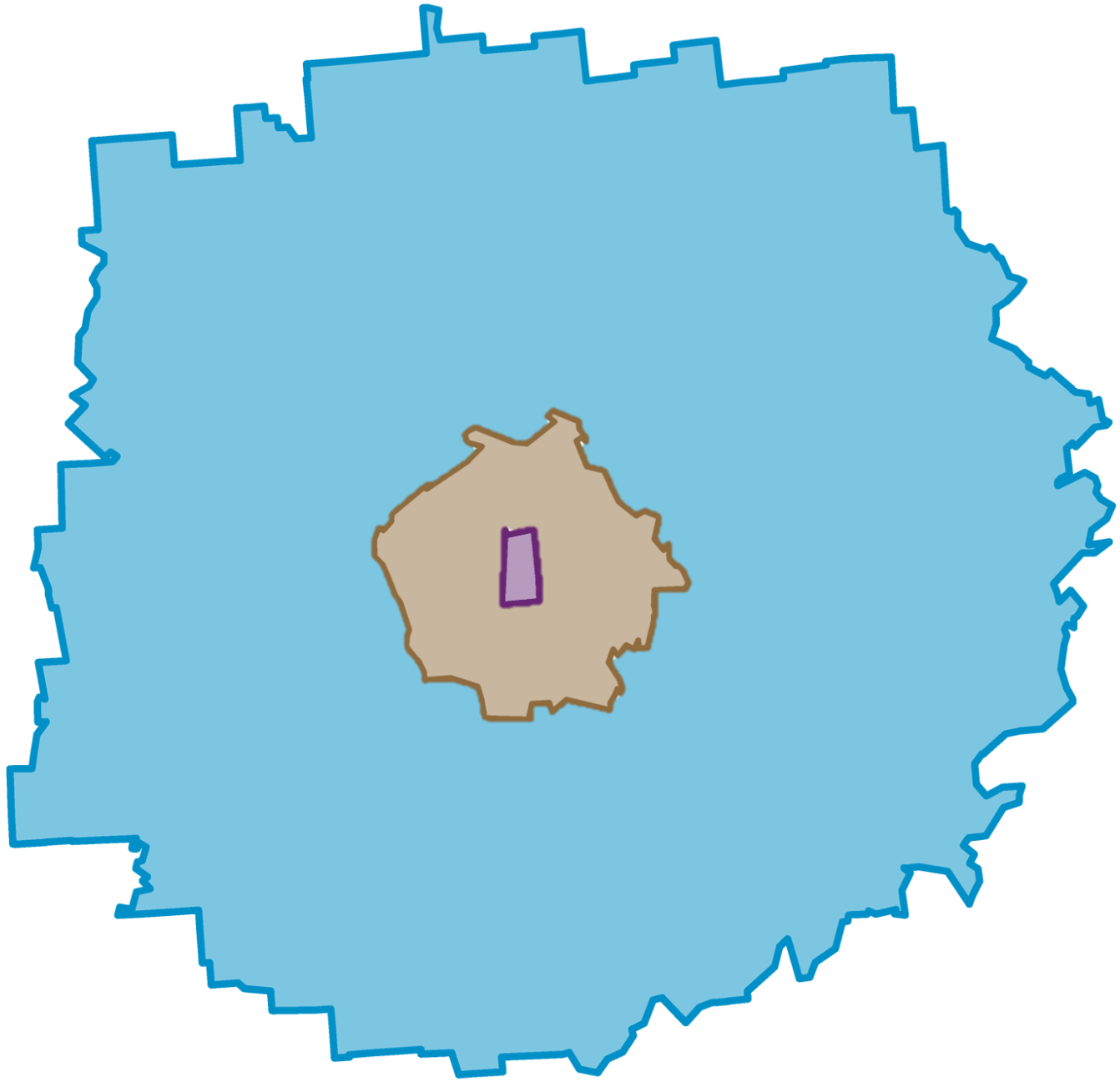
if significantly higher...

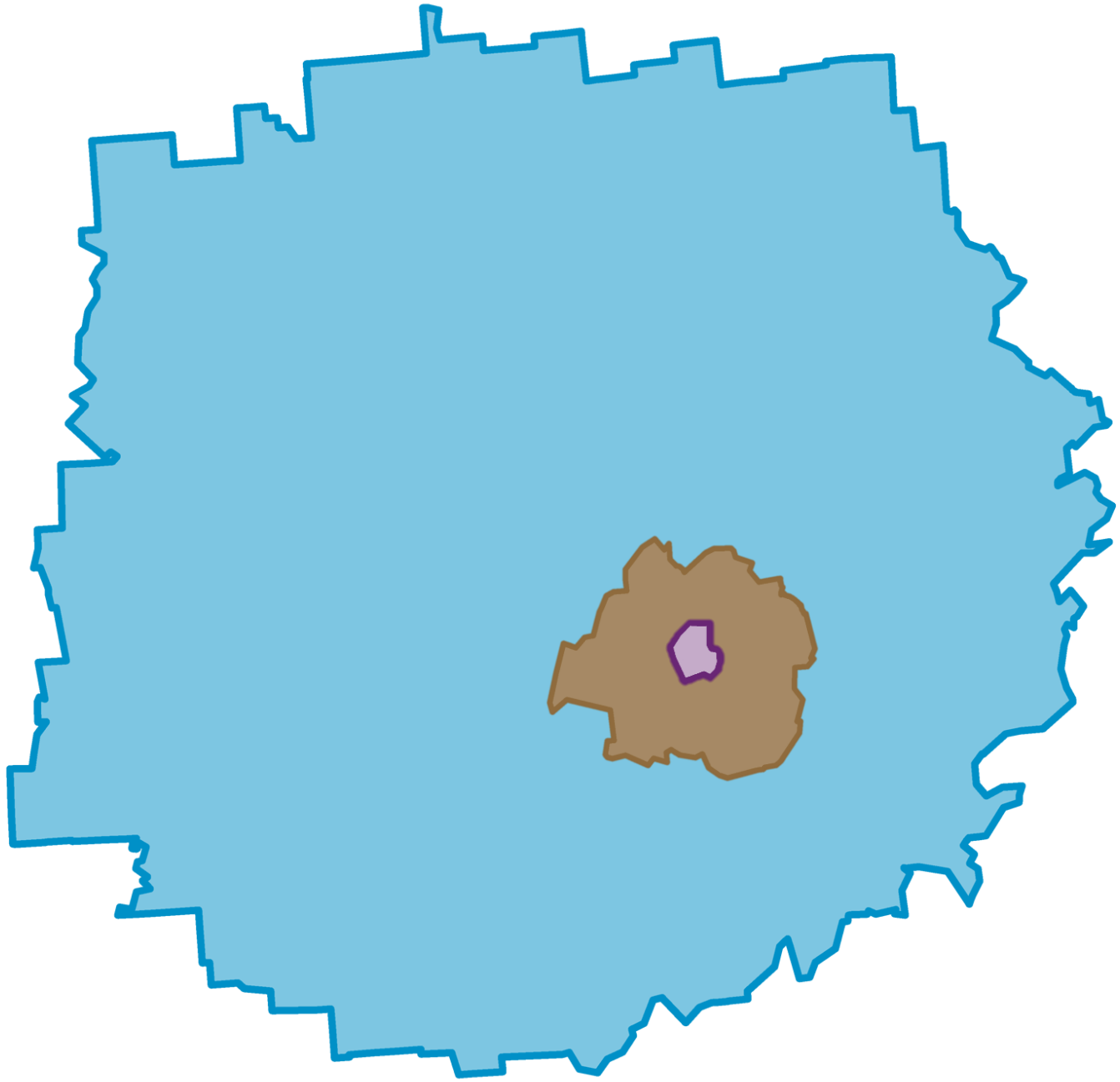
feature is marked as a **hot spot**!

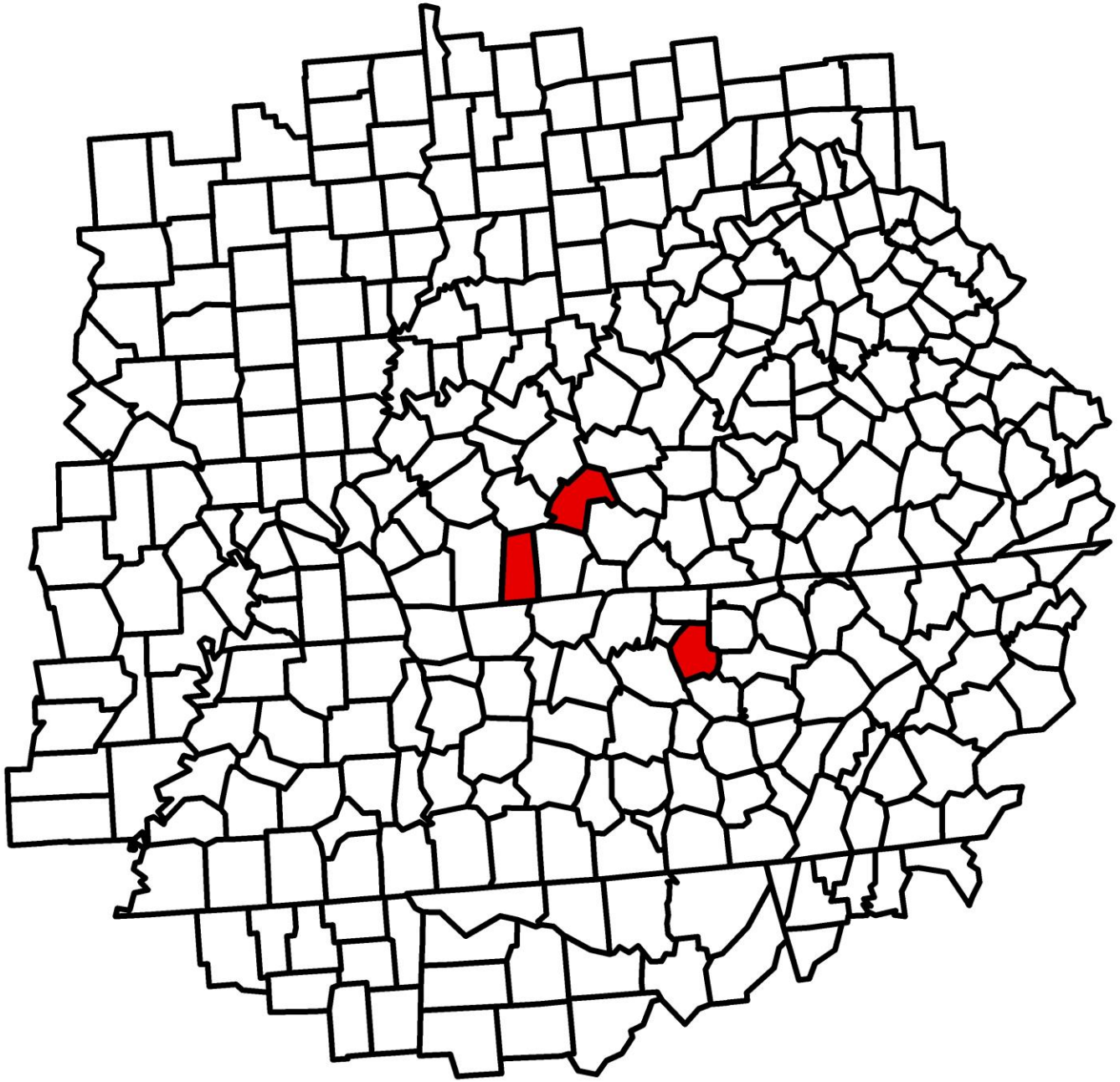
Hot Spot - 90% Confidence

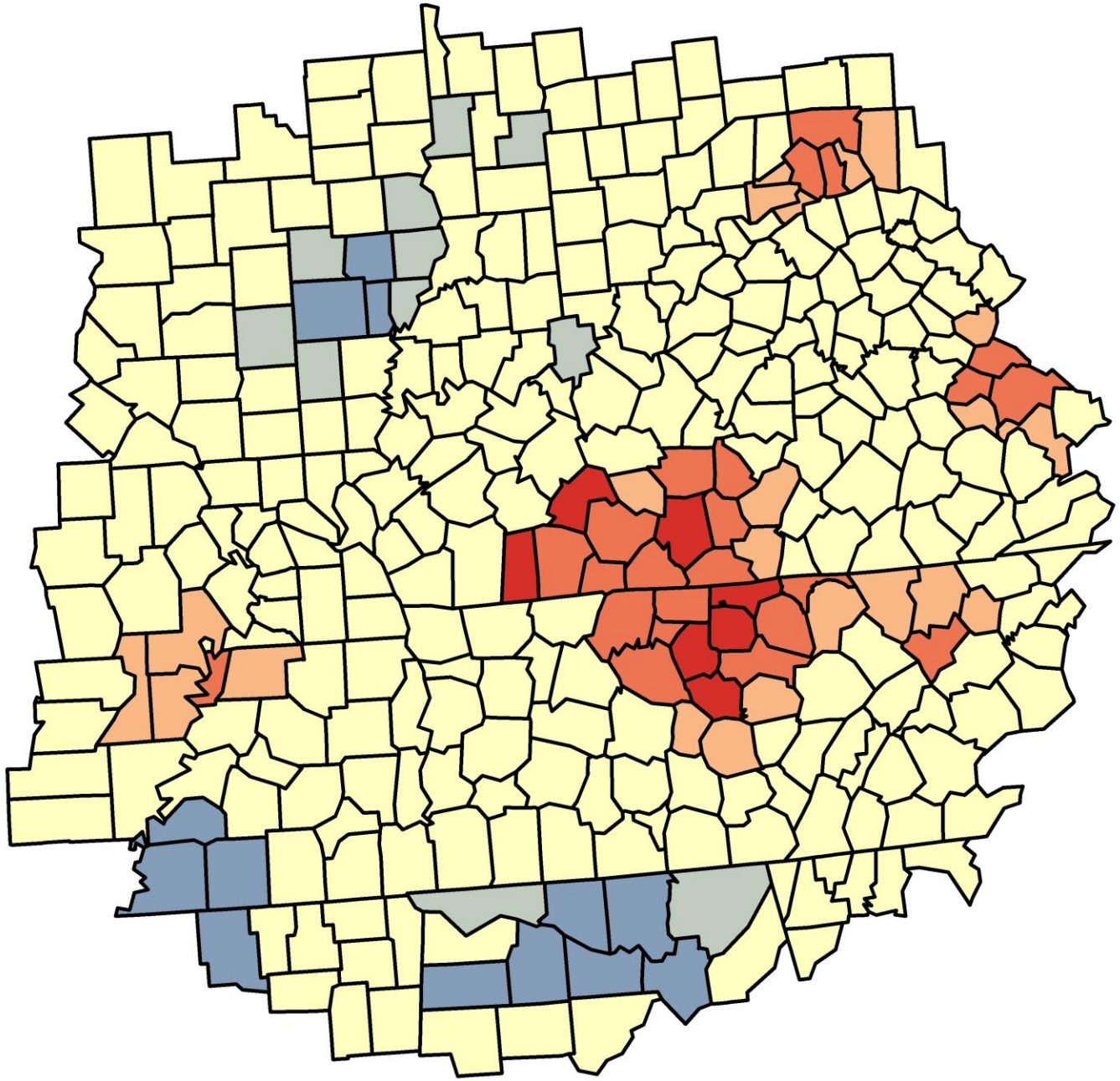
Hot Spot - 95% Confidence

Hot Spot - 99% Confidence









Cold Spot - 99% Confidence

Cold Spot - 95% Confidence

Cold Spot - 90% Confidence

Not Significant

Hot Spot - 90% Confidence

Hot Spot - 95% Confidence

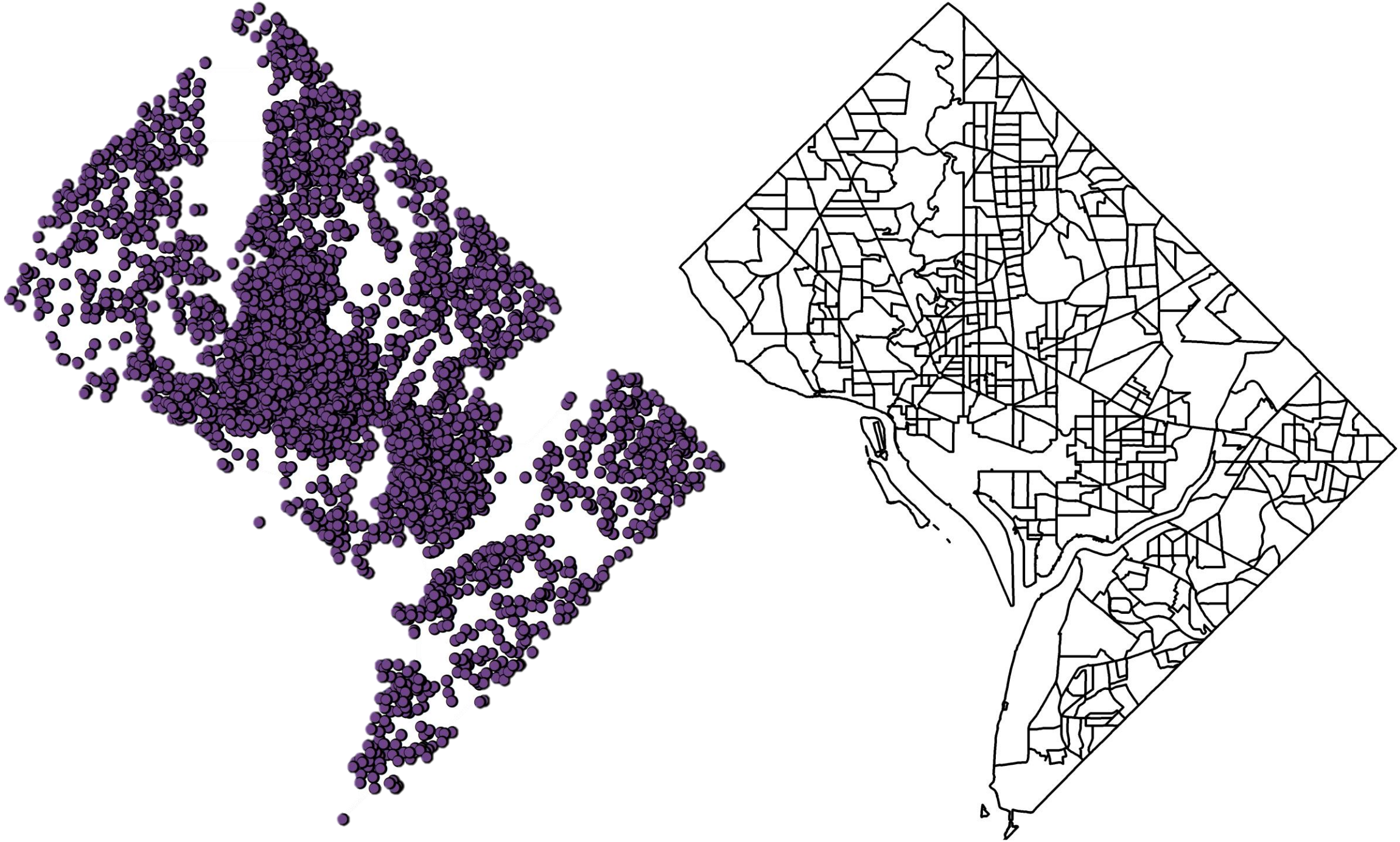
Hot Spot - 99% Confidence

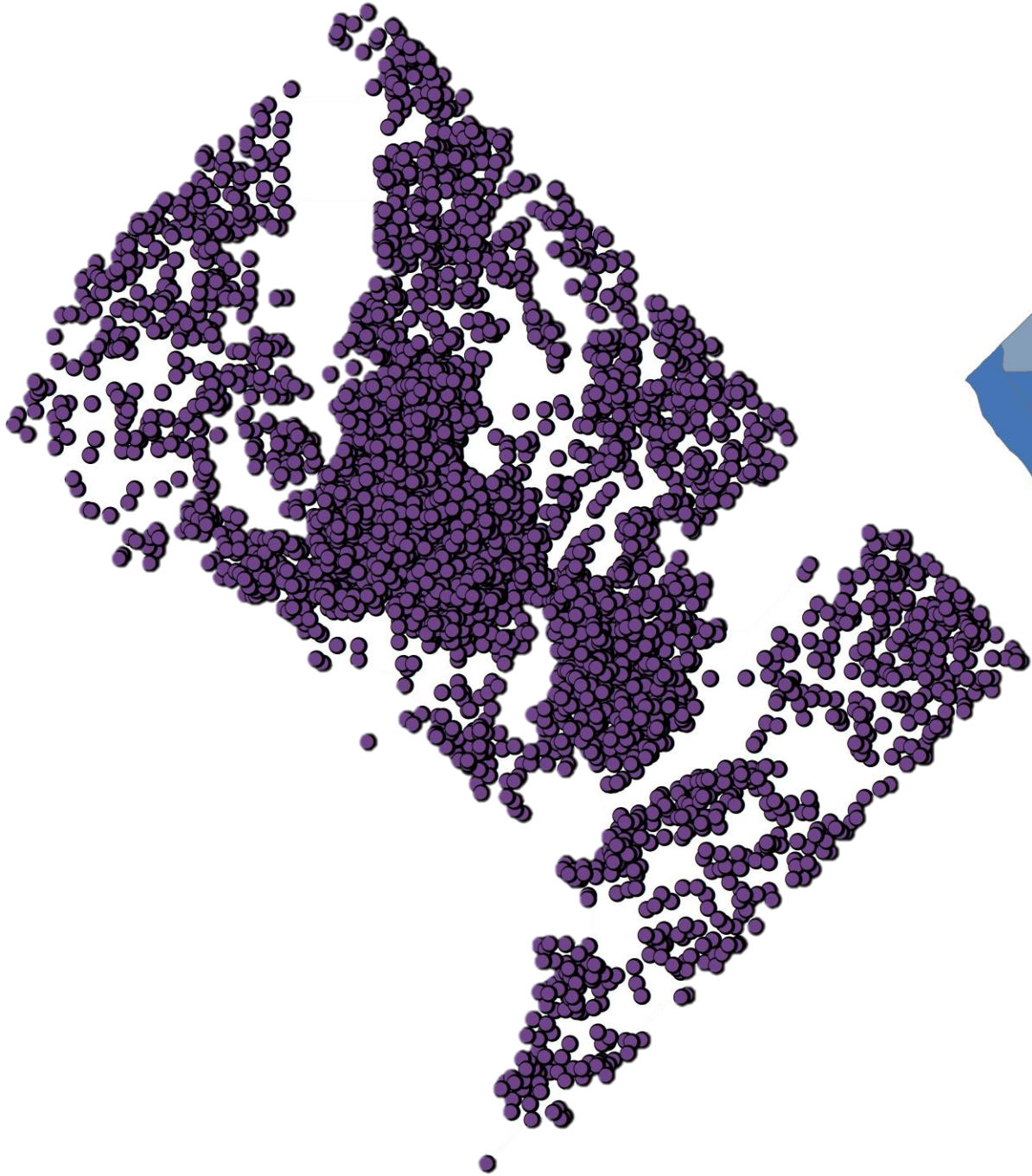
…how do we know if it's SIGNIFICANTLY different???

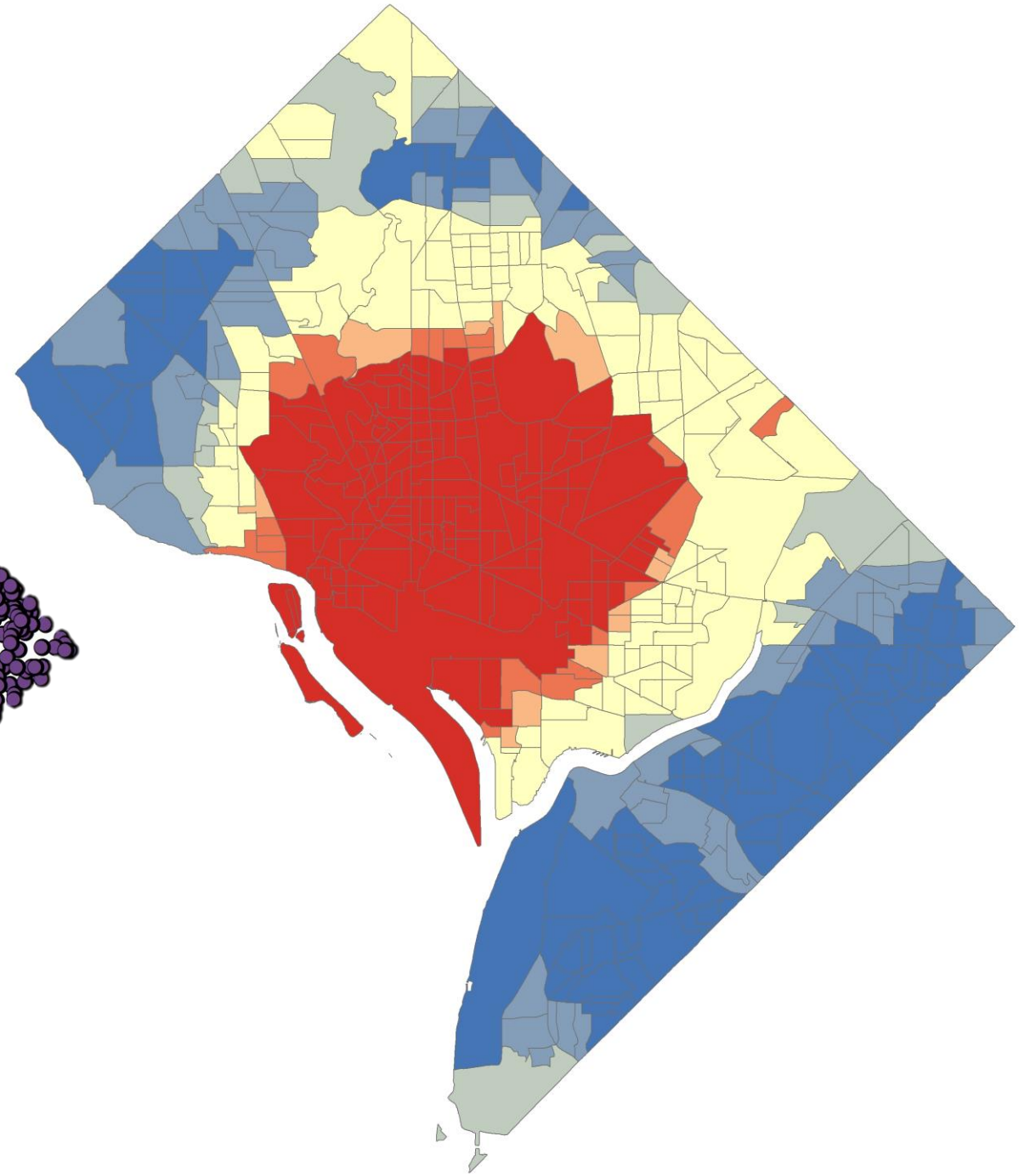
MATH!

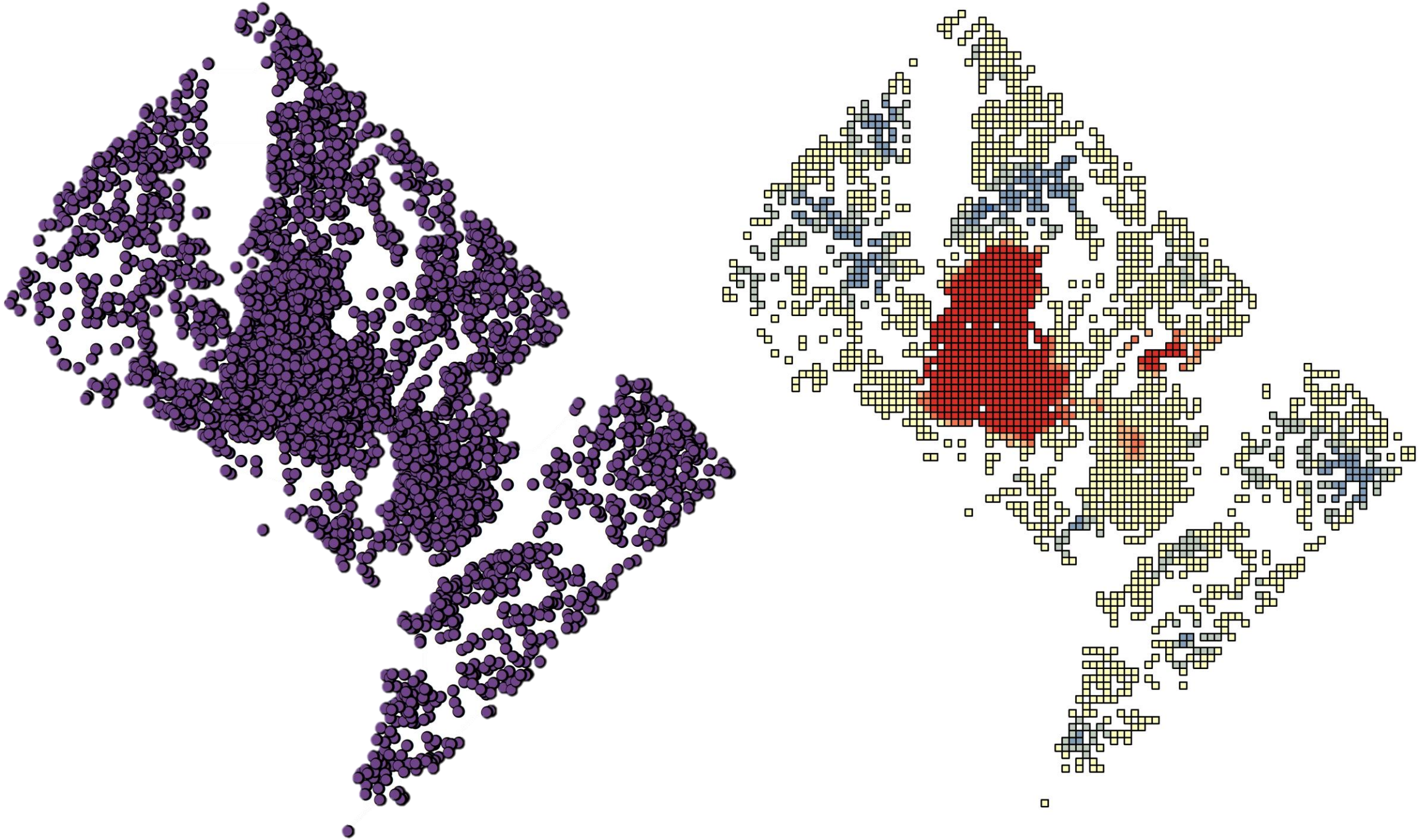
Getis-Ord Gi* Statistic

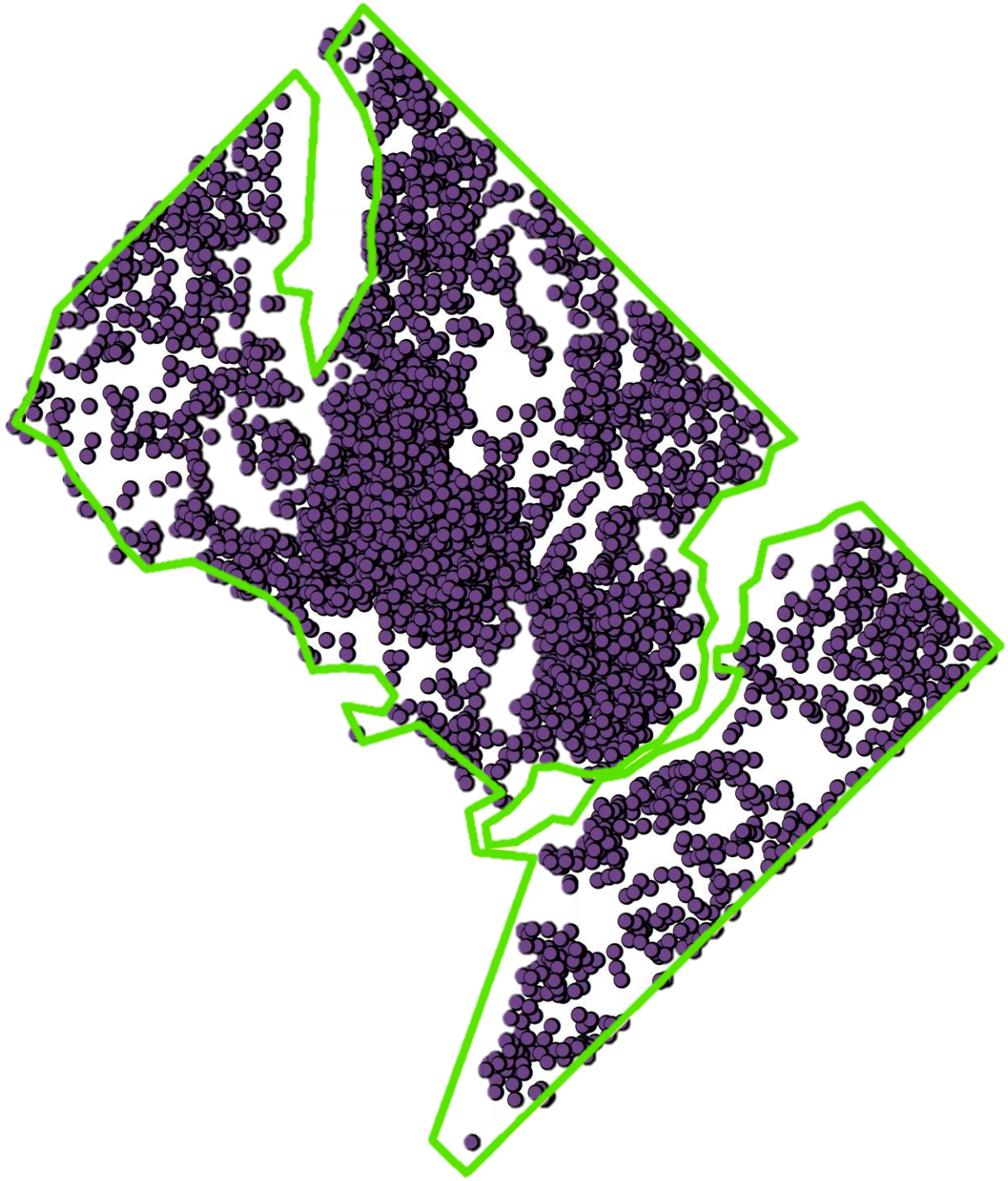
Points

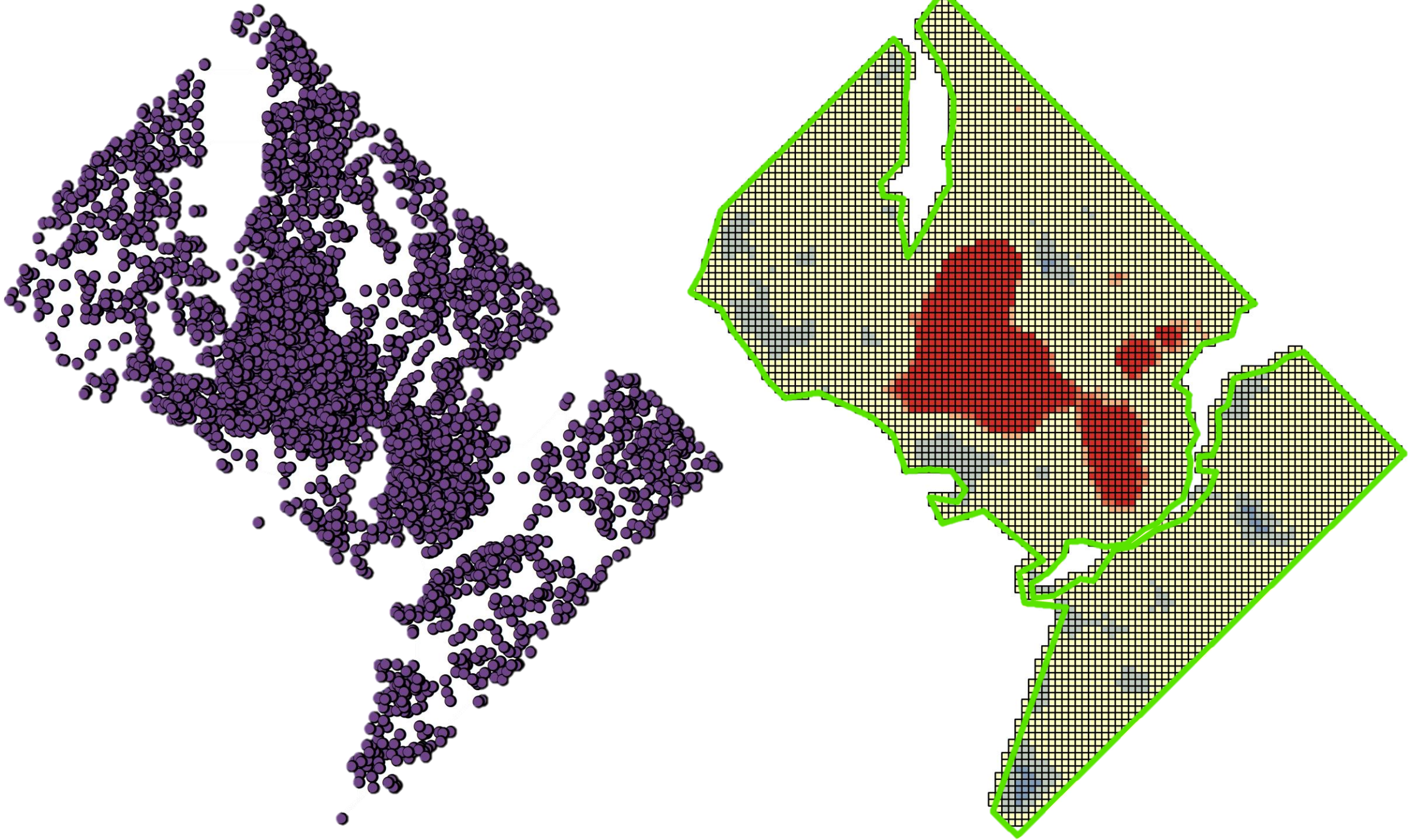


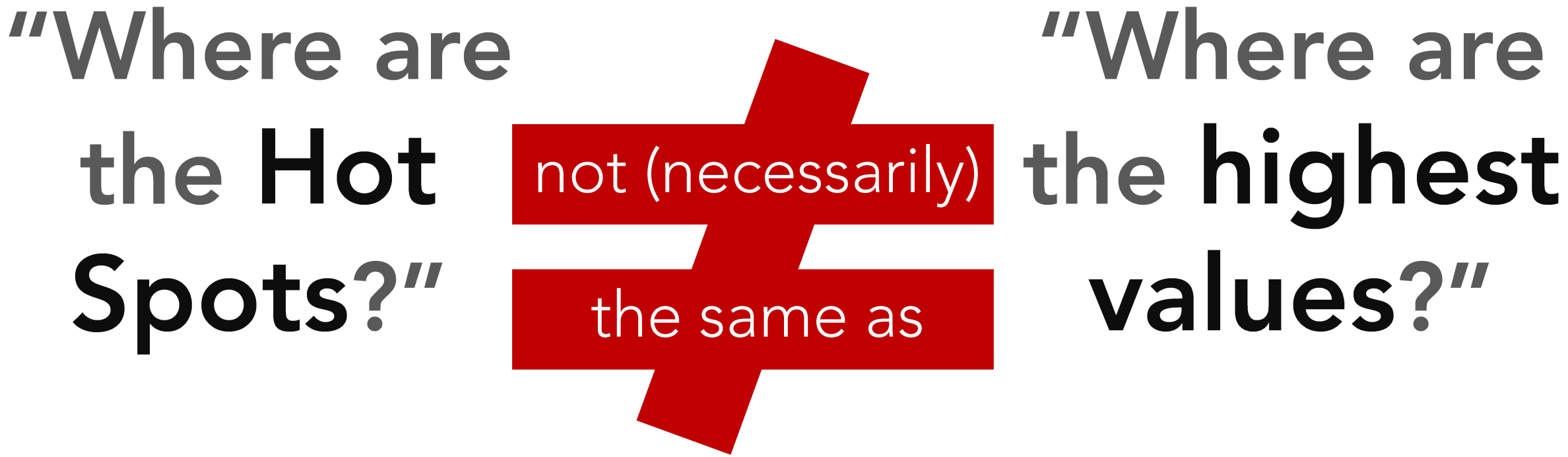


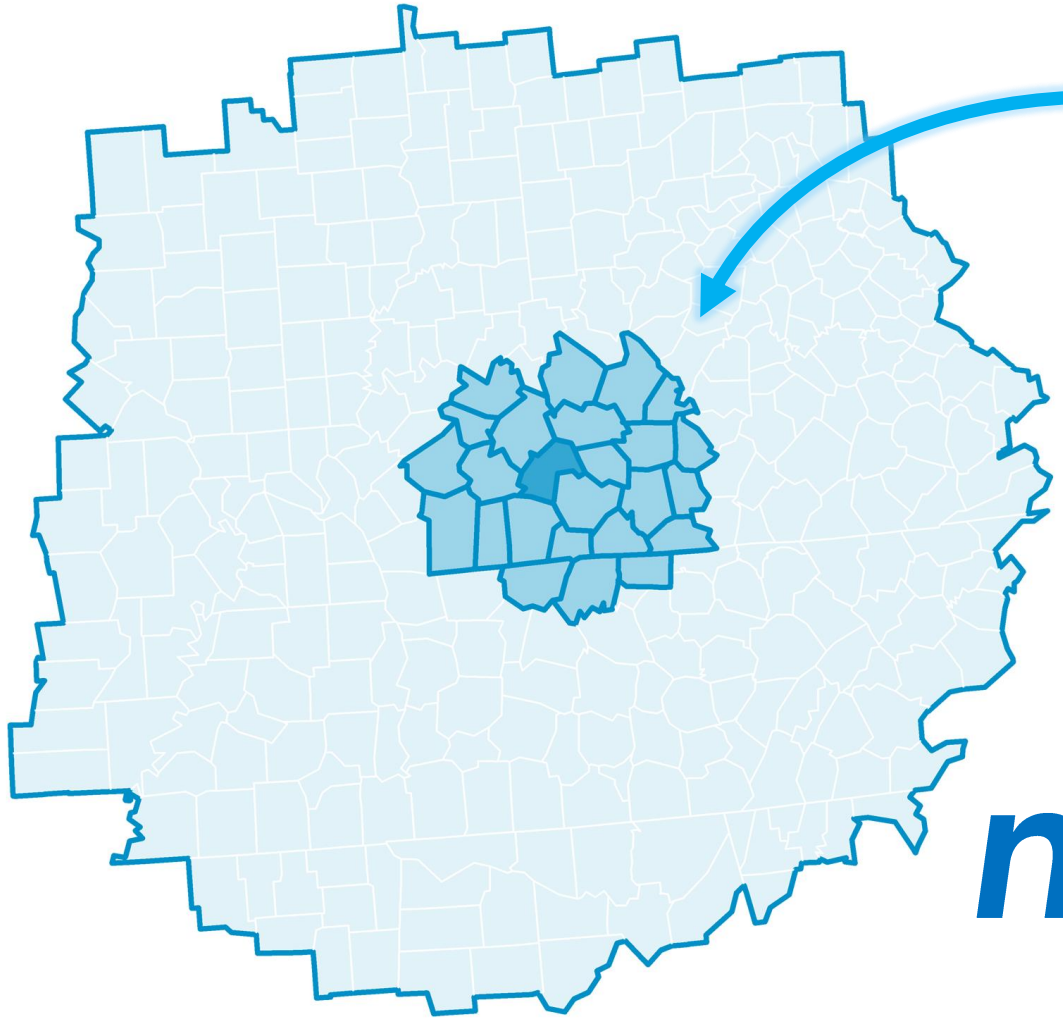






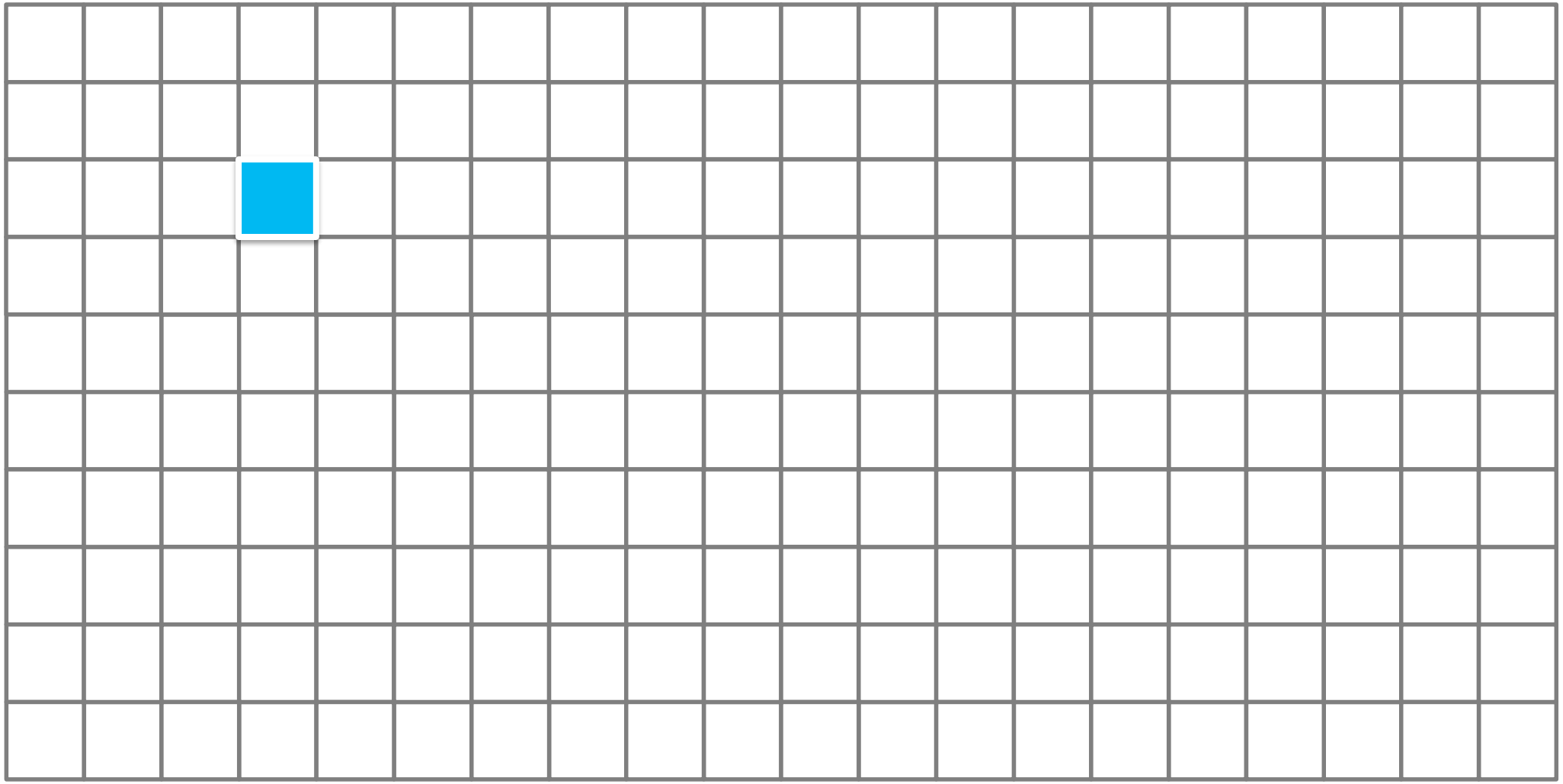



"Where are the Hot Spots?"
not (necessarily)
the same as
"Where are the highest values?"

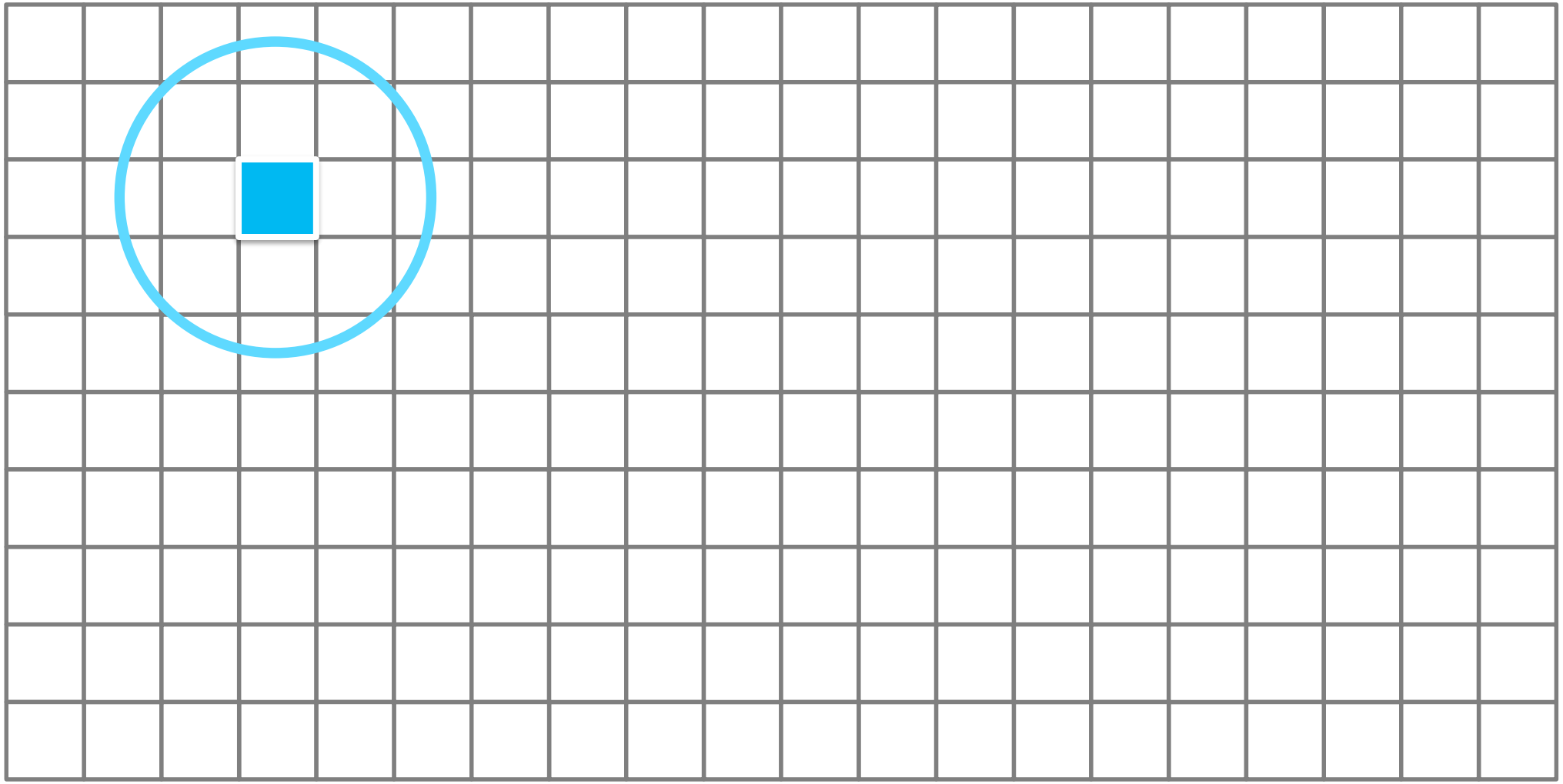


How are *neighborhood* sizes determined?

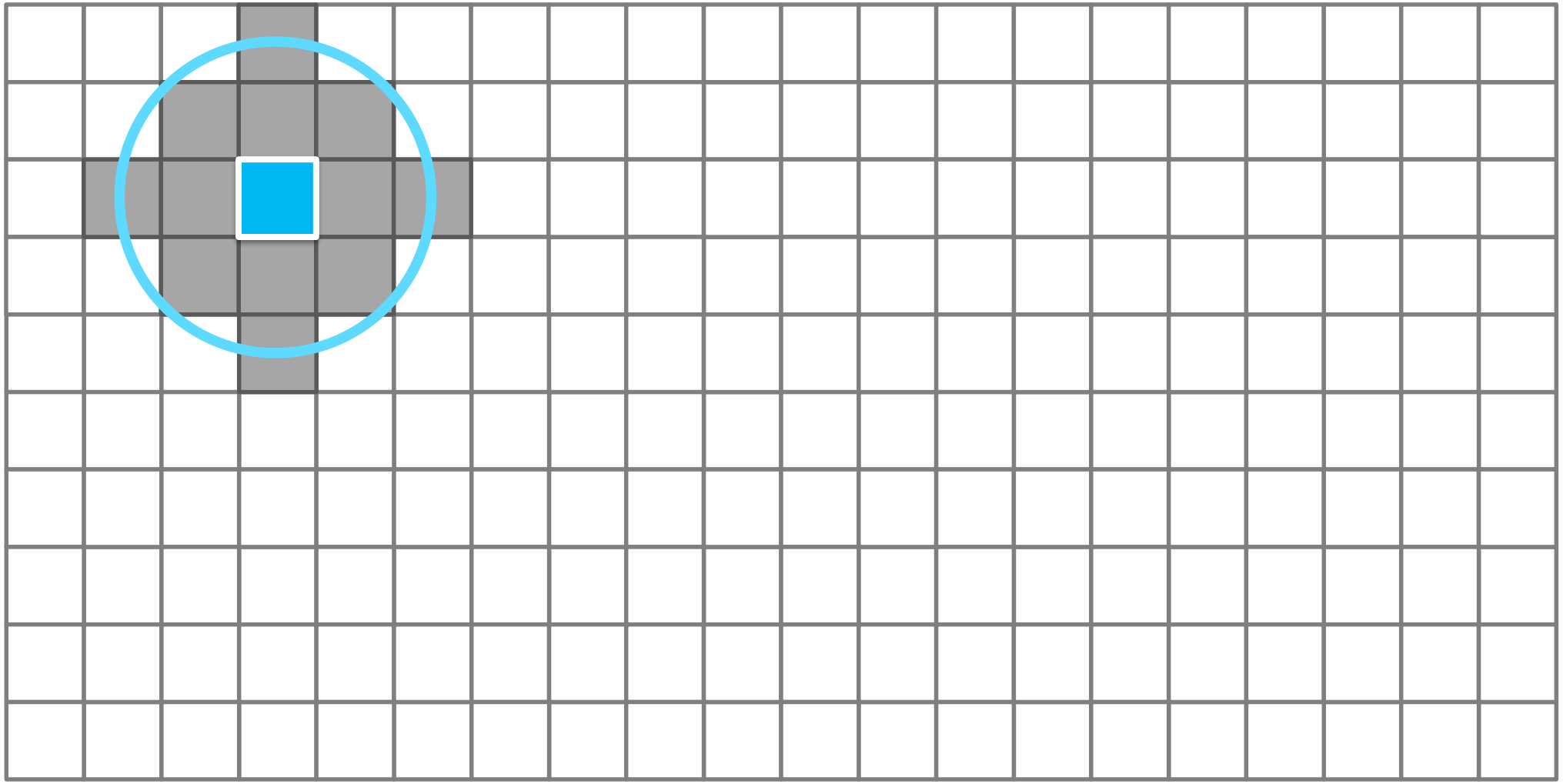
Fixed Distance Band



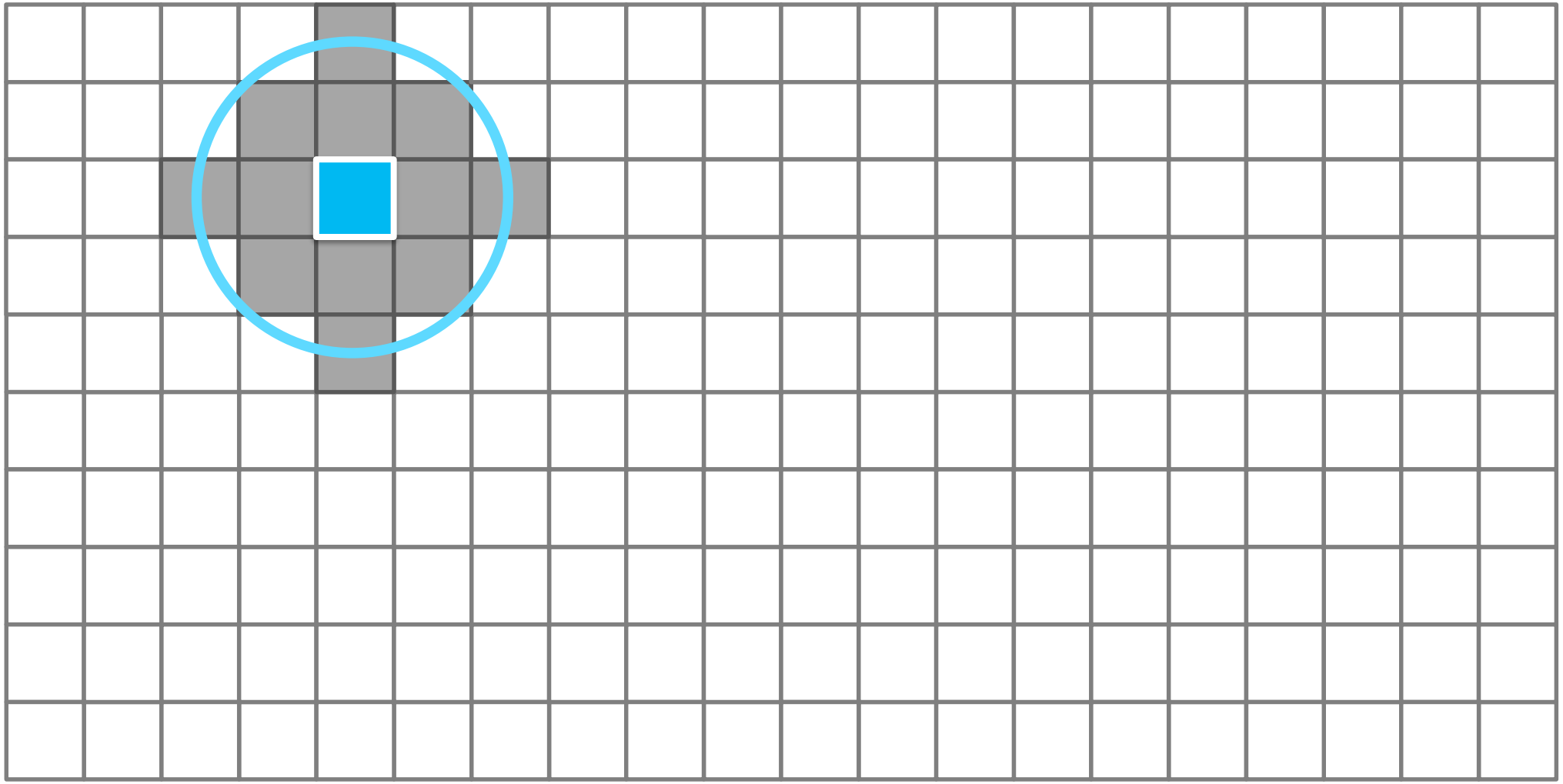
Fixed Distance Band



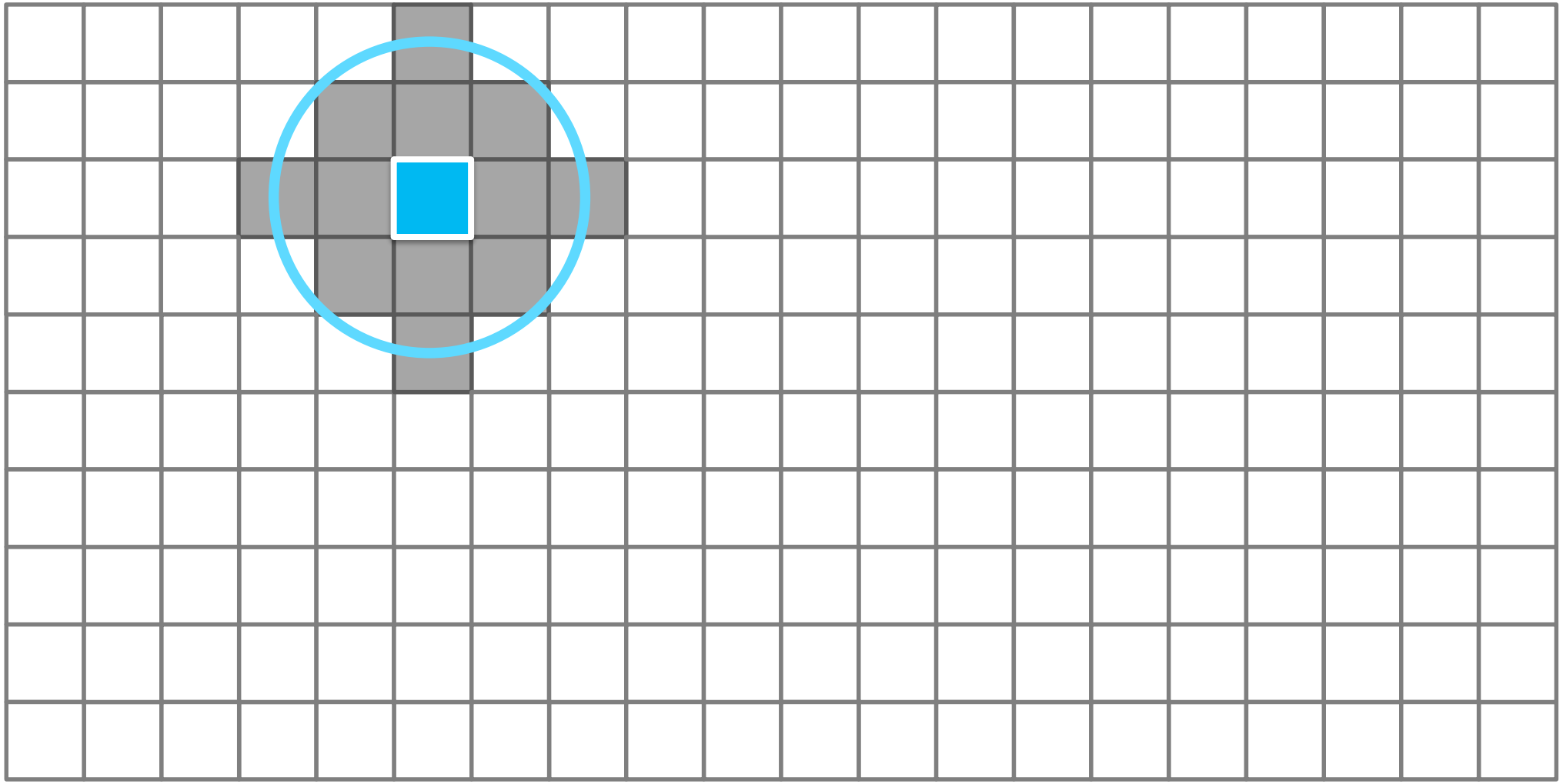
Fixed Distance Band



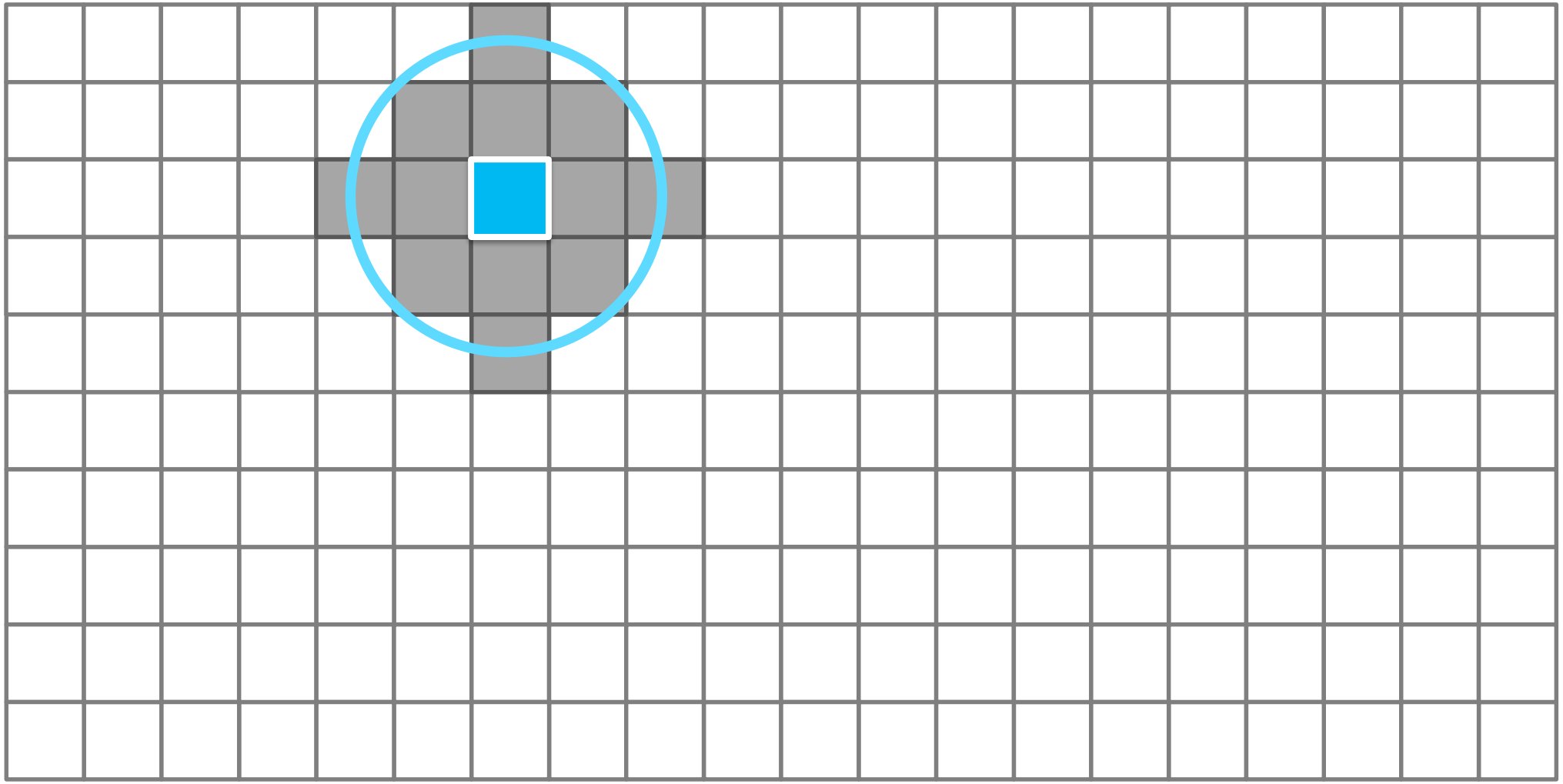
Fixed Distance Band



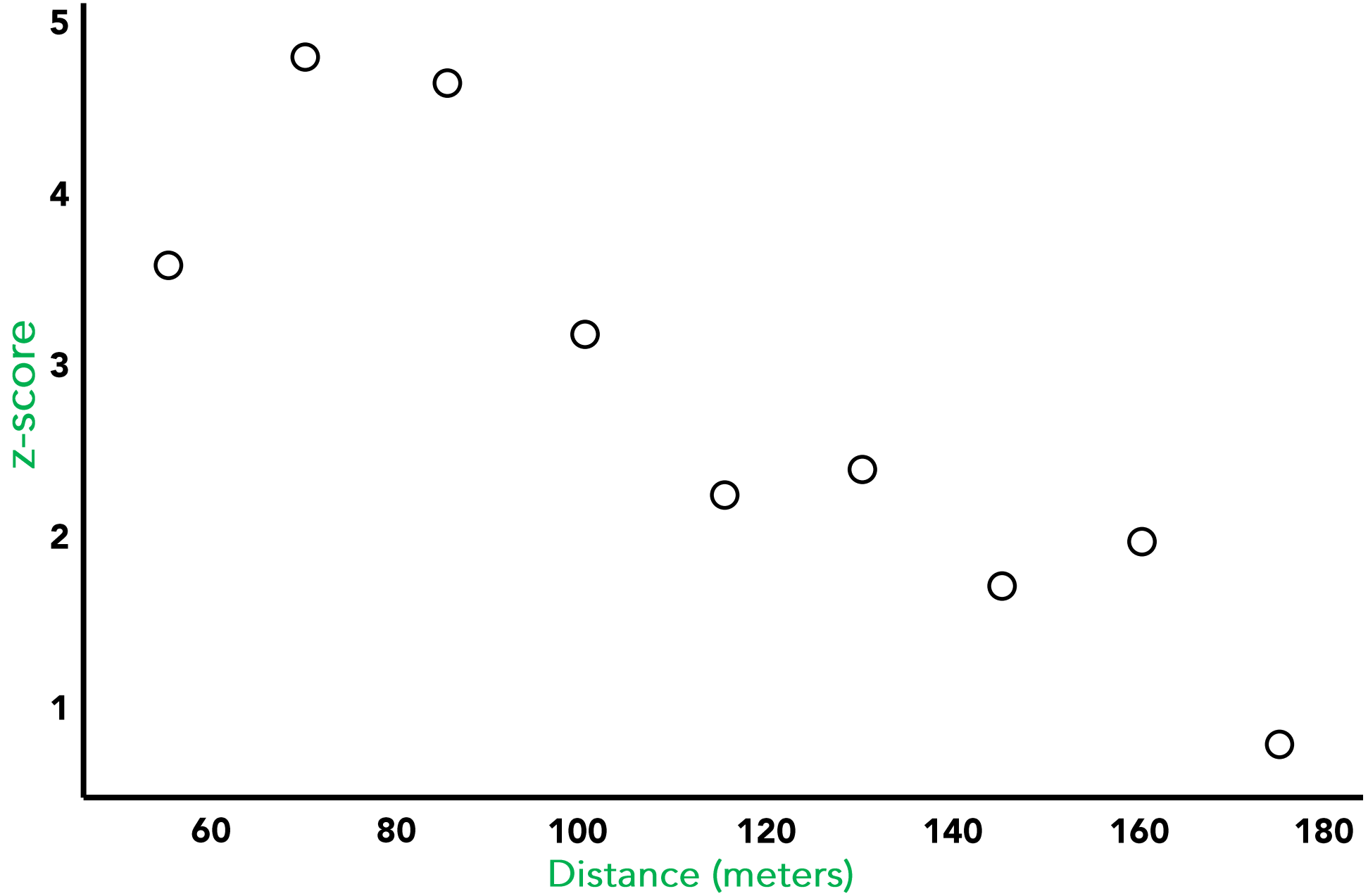
Fixed Distance Band



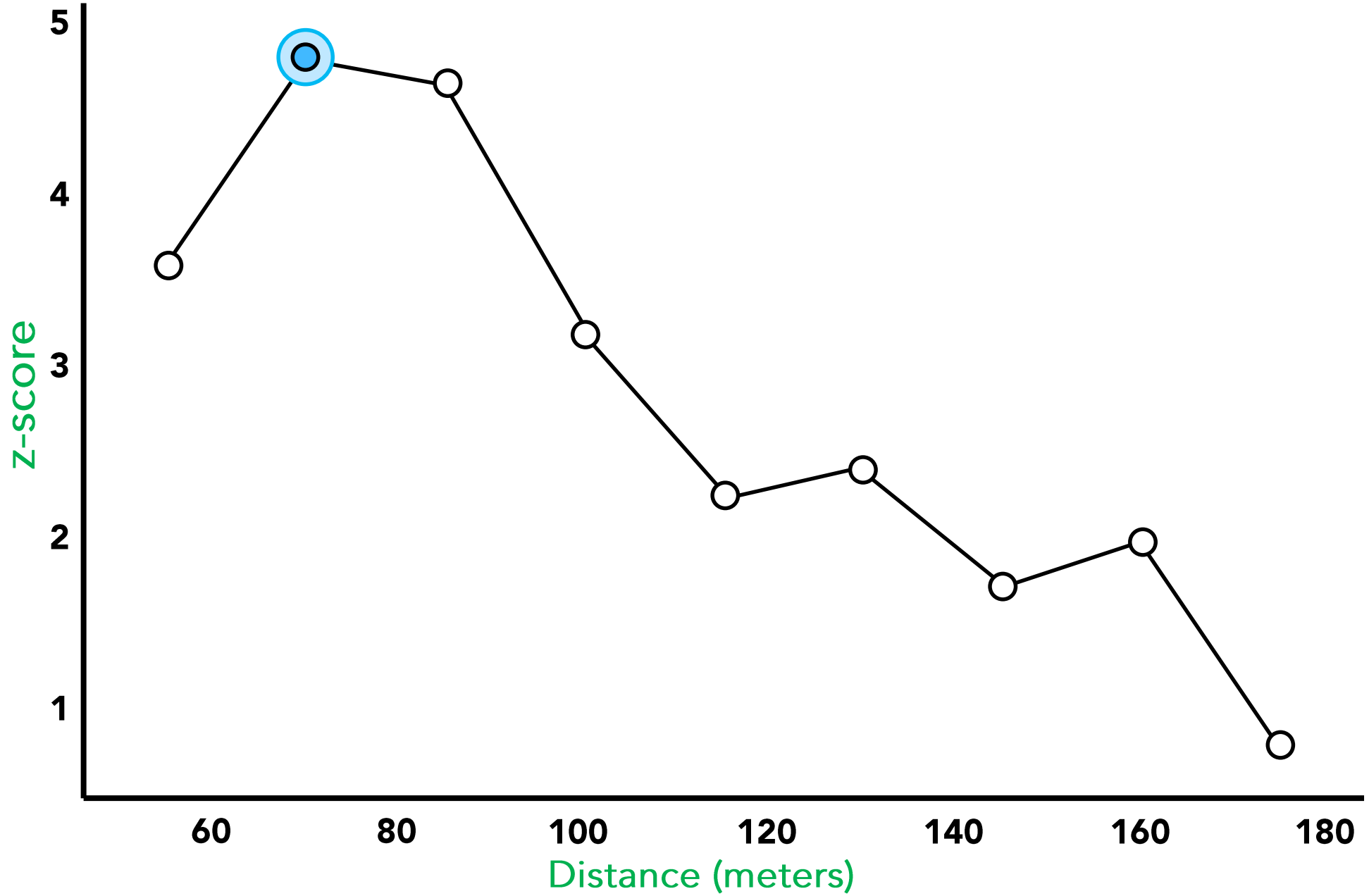
Fixed Distance Band



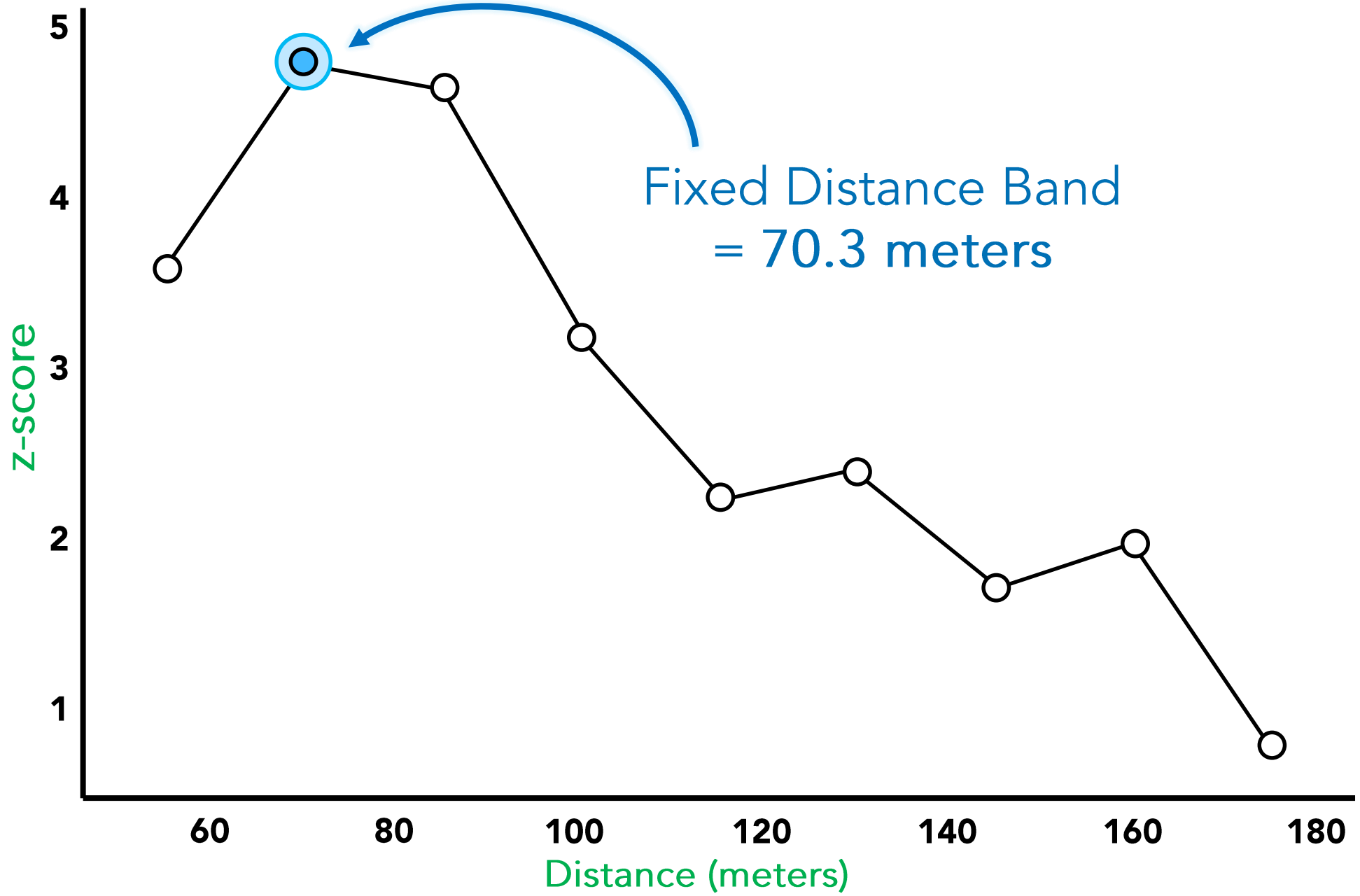
Spatial Autocorrelation by Distance

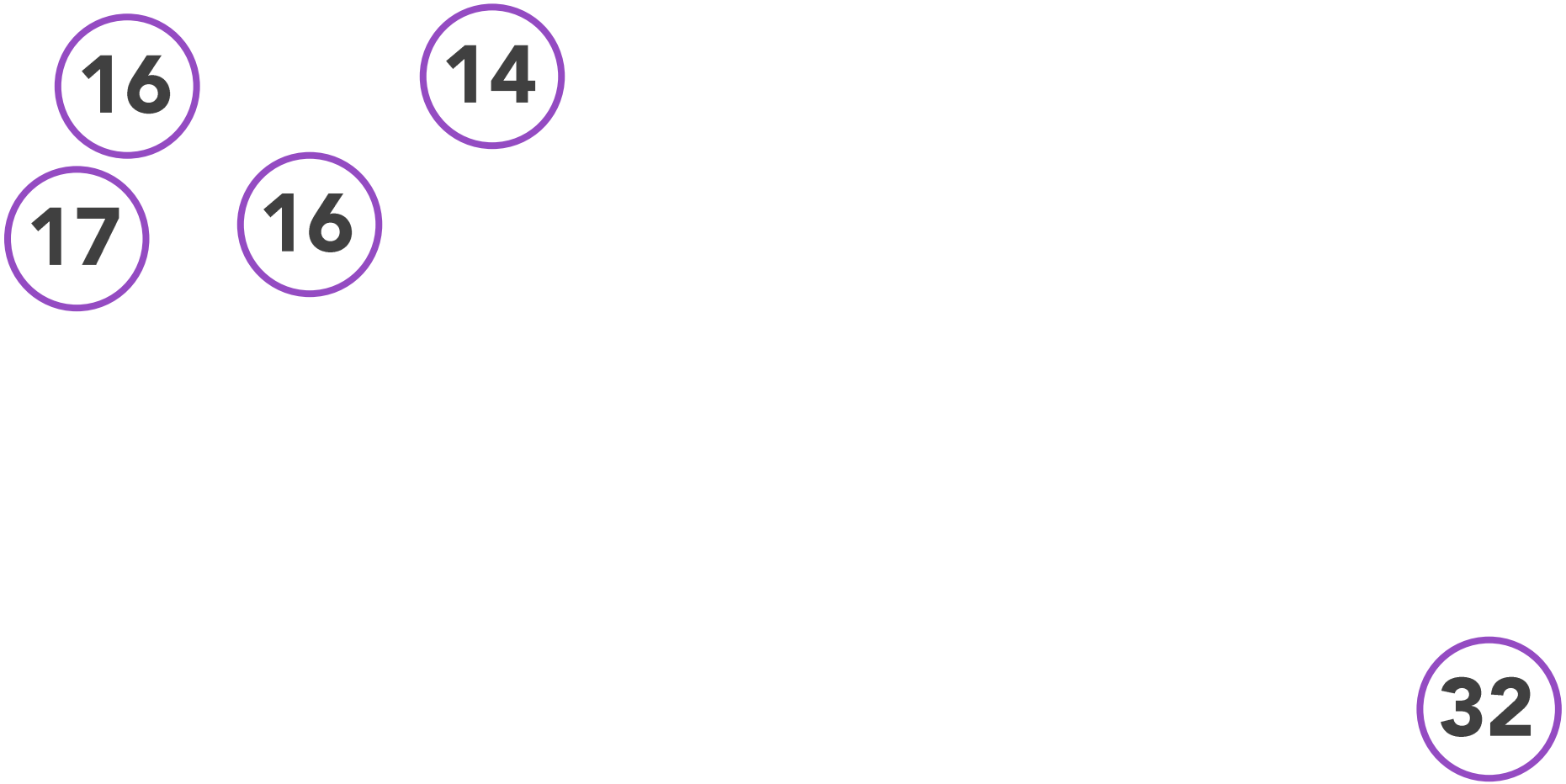


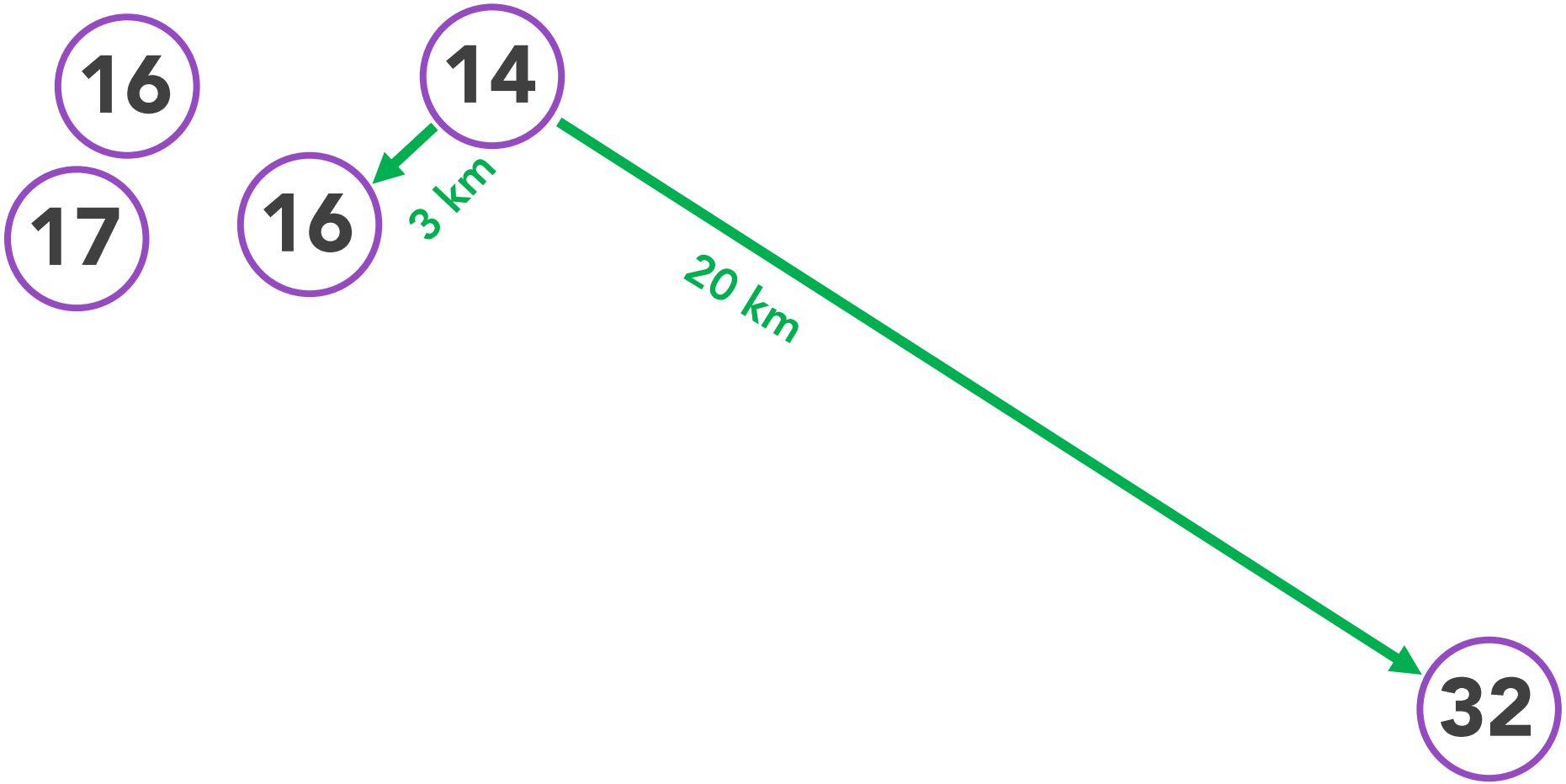
Spatial Autocorrelation by Distance

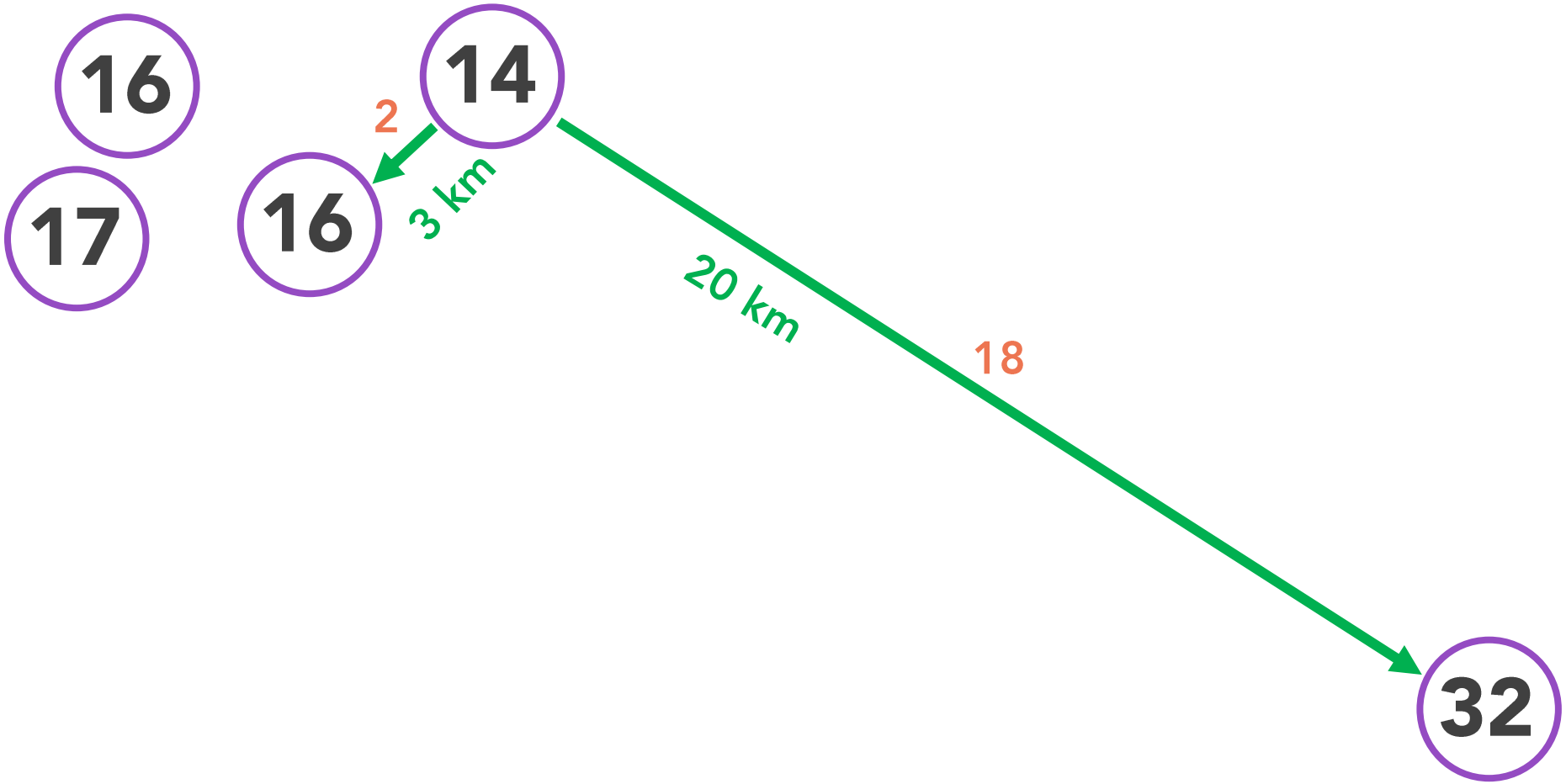


Spatial Autocorrelation by Distance

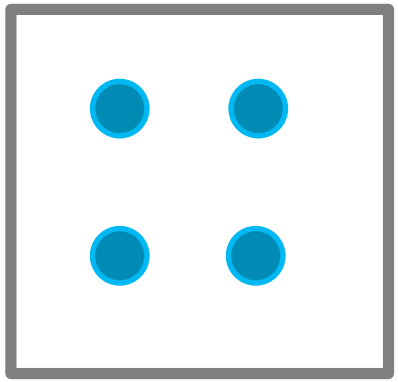



16
14
17
16
32

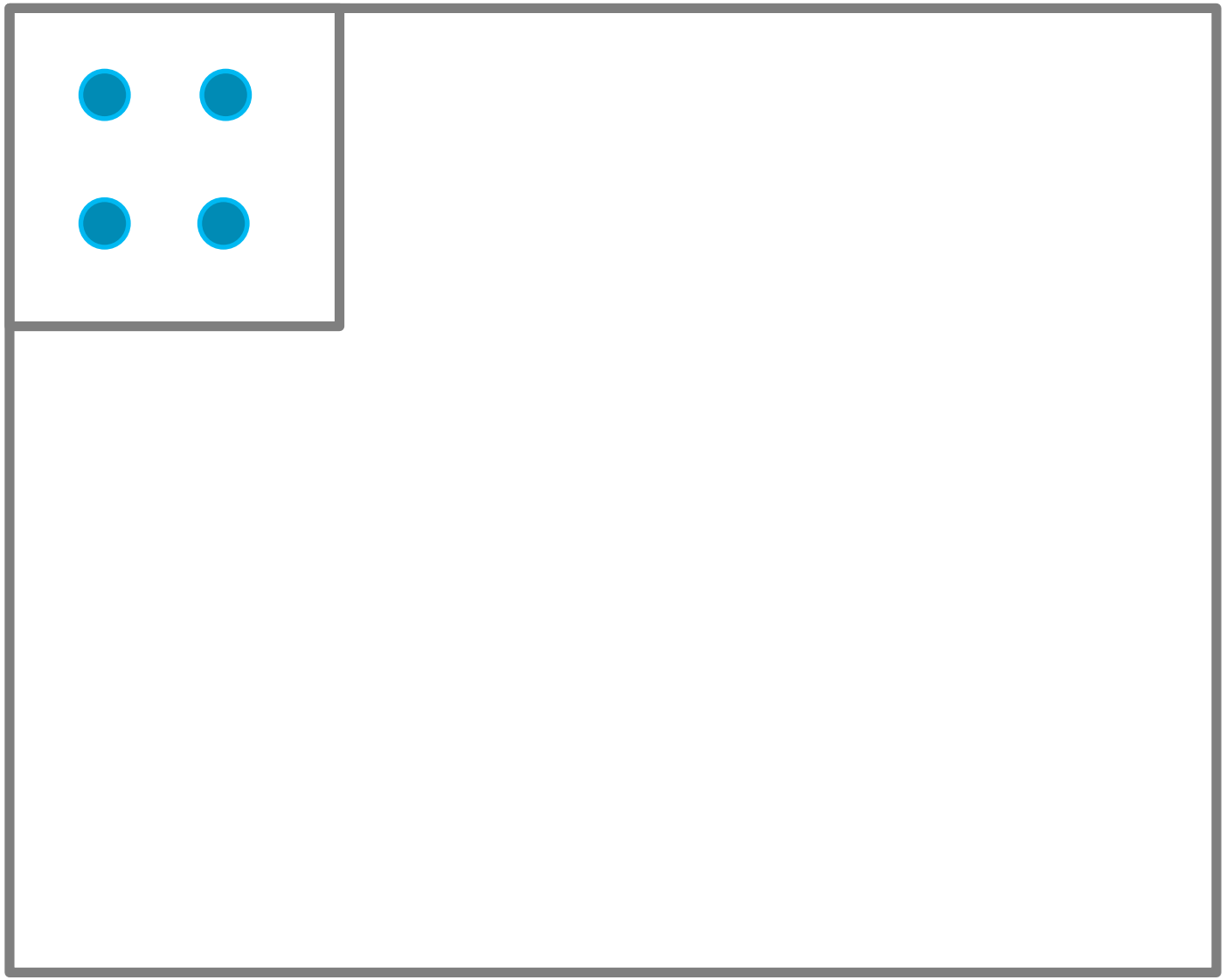

16
17
14
16
3 km
20 km
32


16
17
14
2
16
3 km
20 km
18
32

dispersed

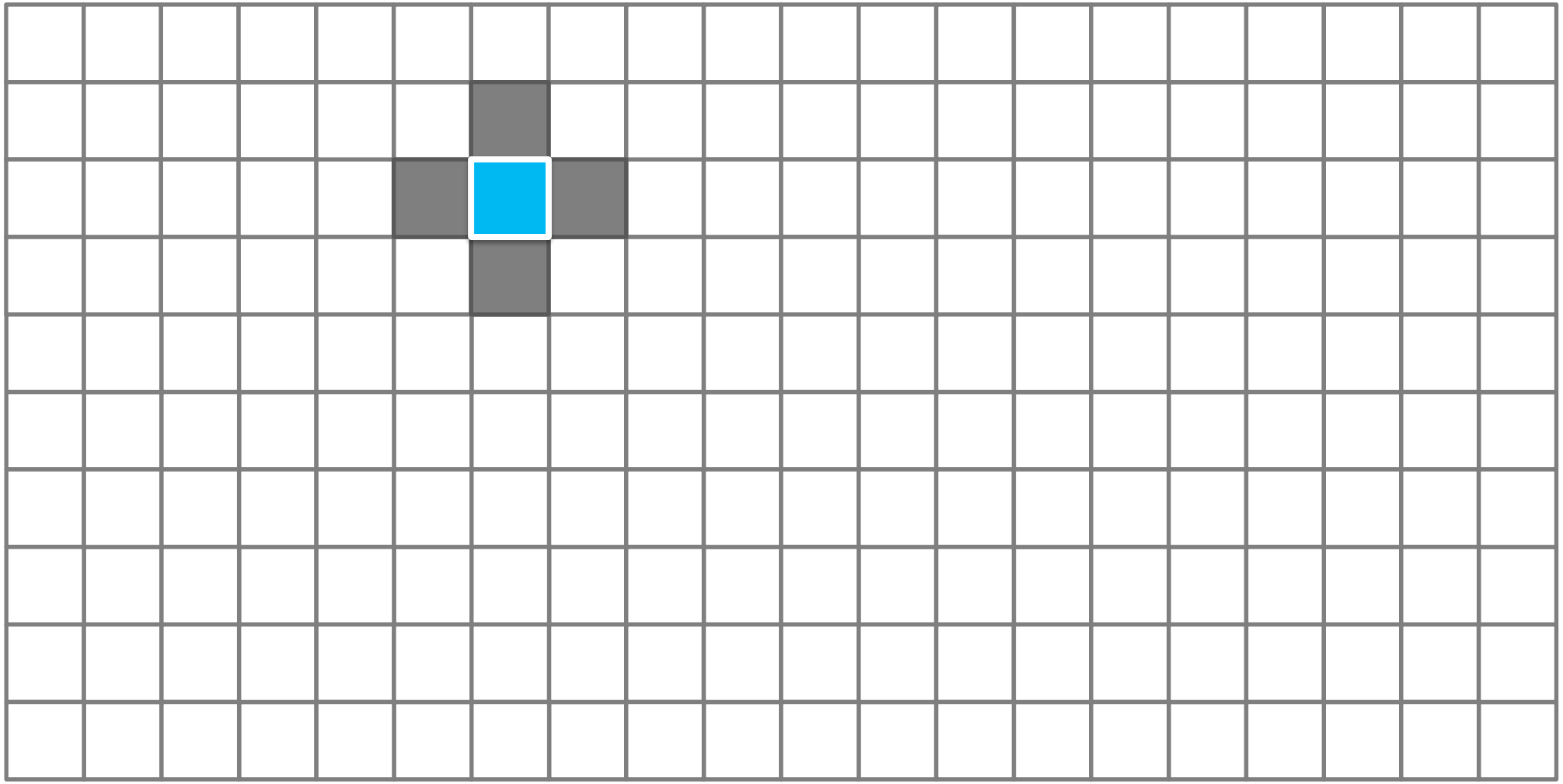


clustered



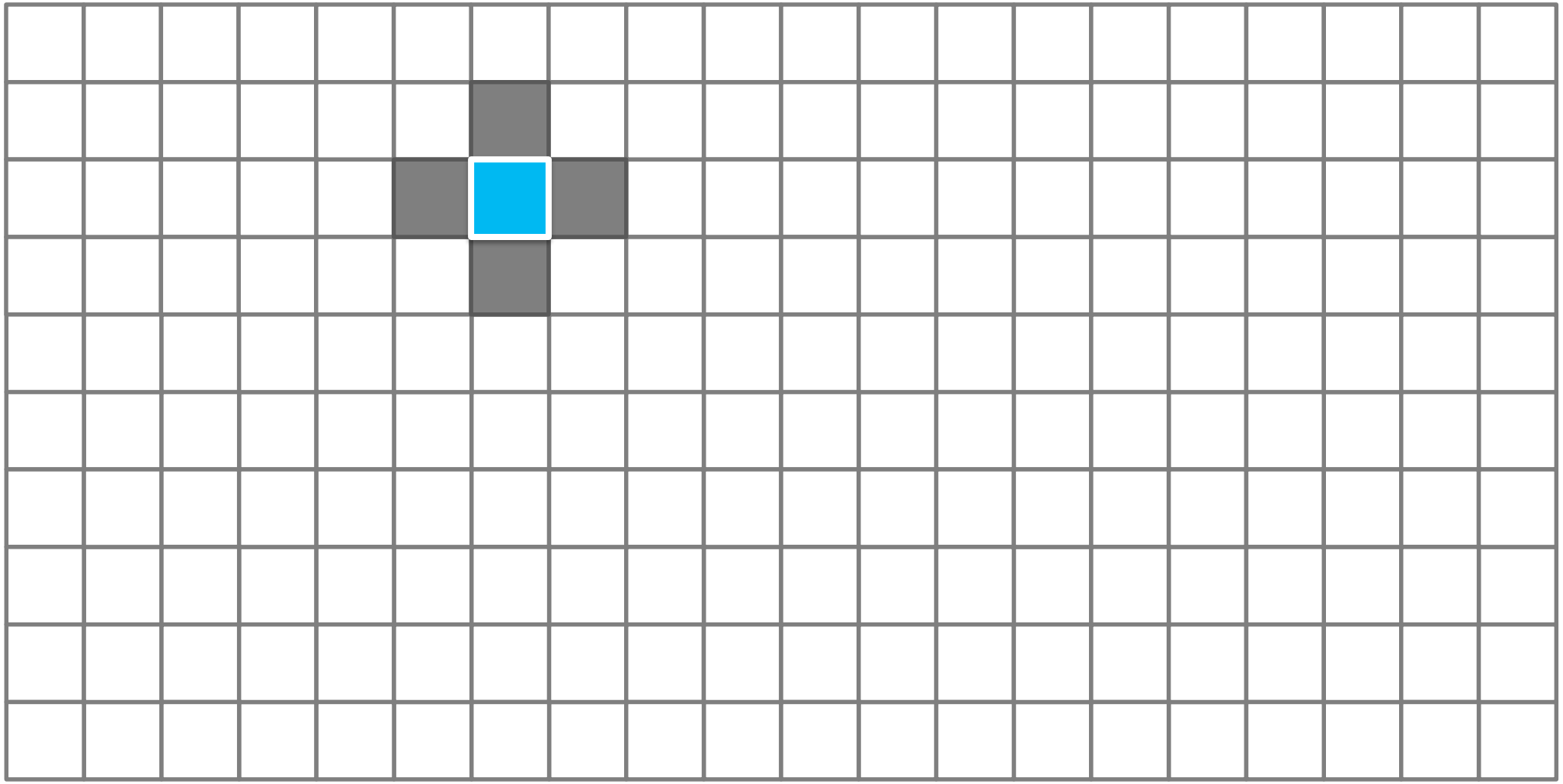
Contiguity

Edges



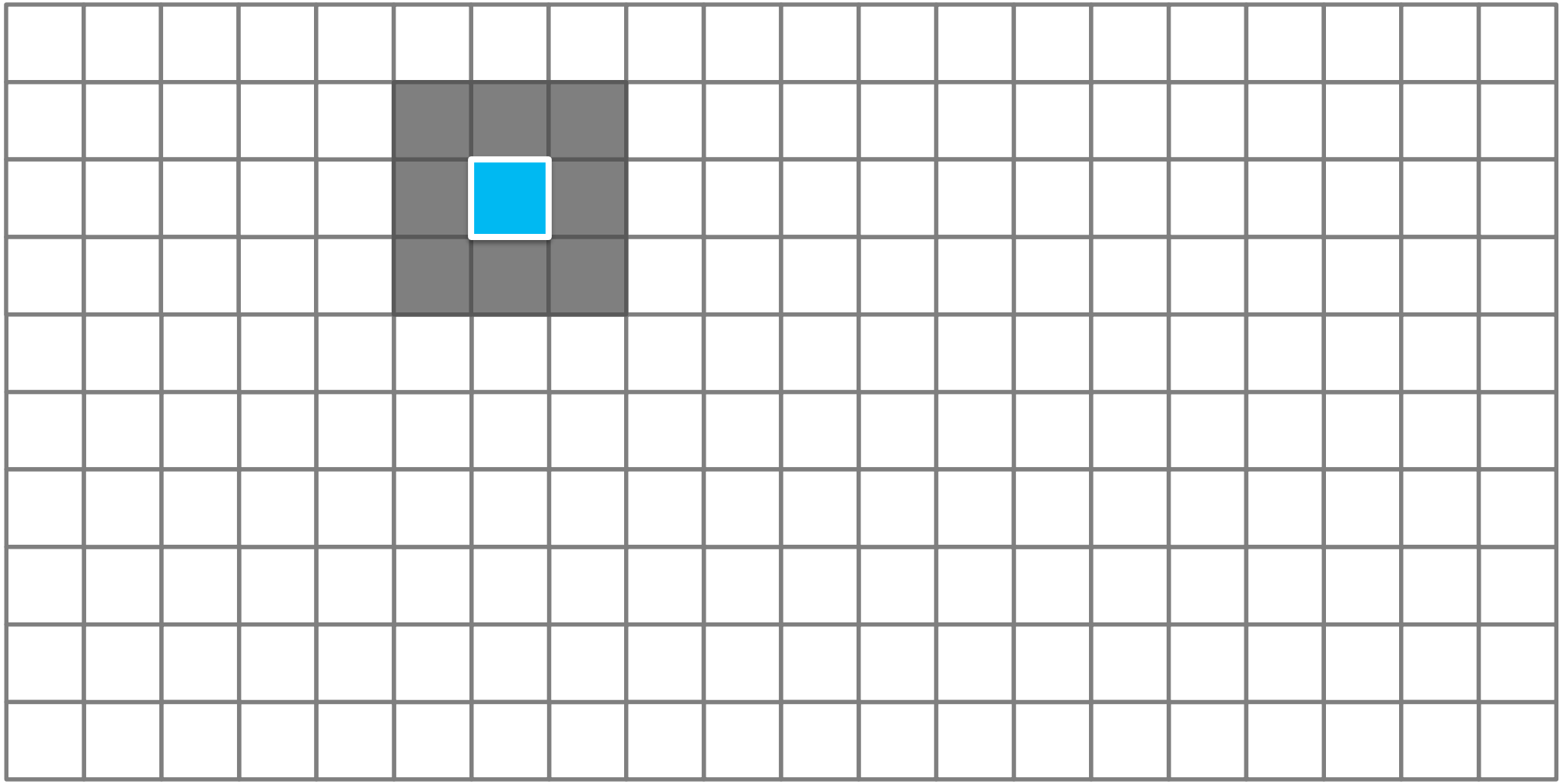
Contiguity

Rook's Case



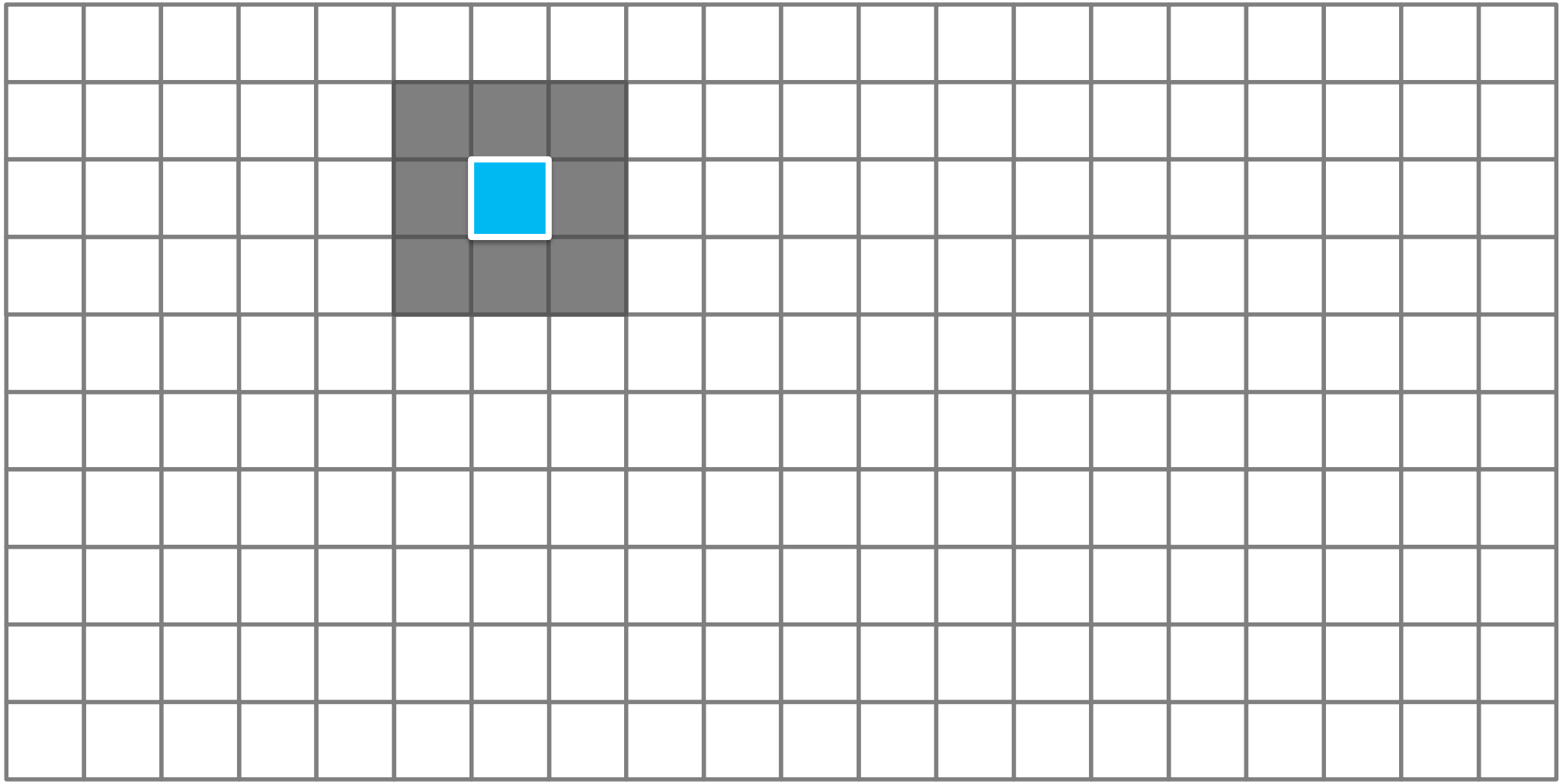
Contiguity

Edges/Corners



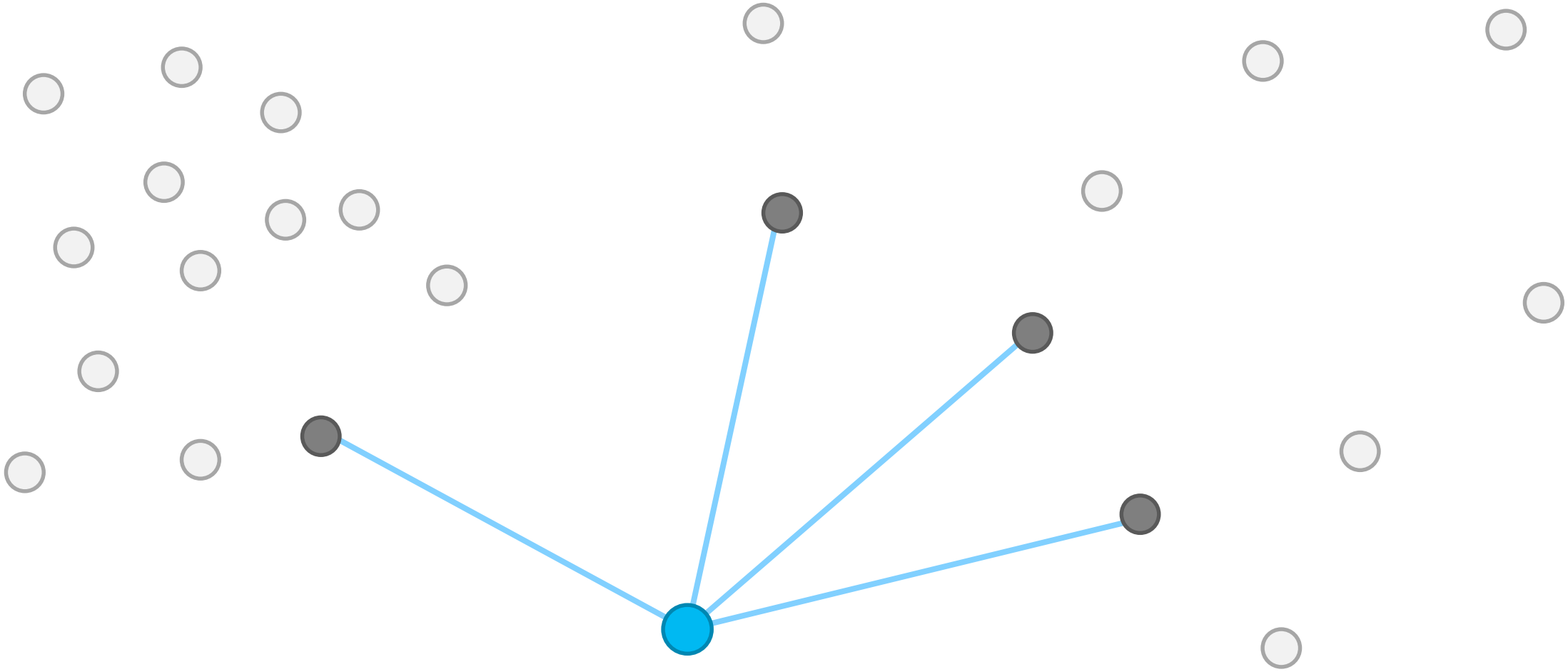
Contiguity

Queen's Case

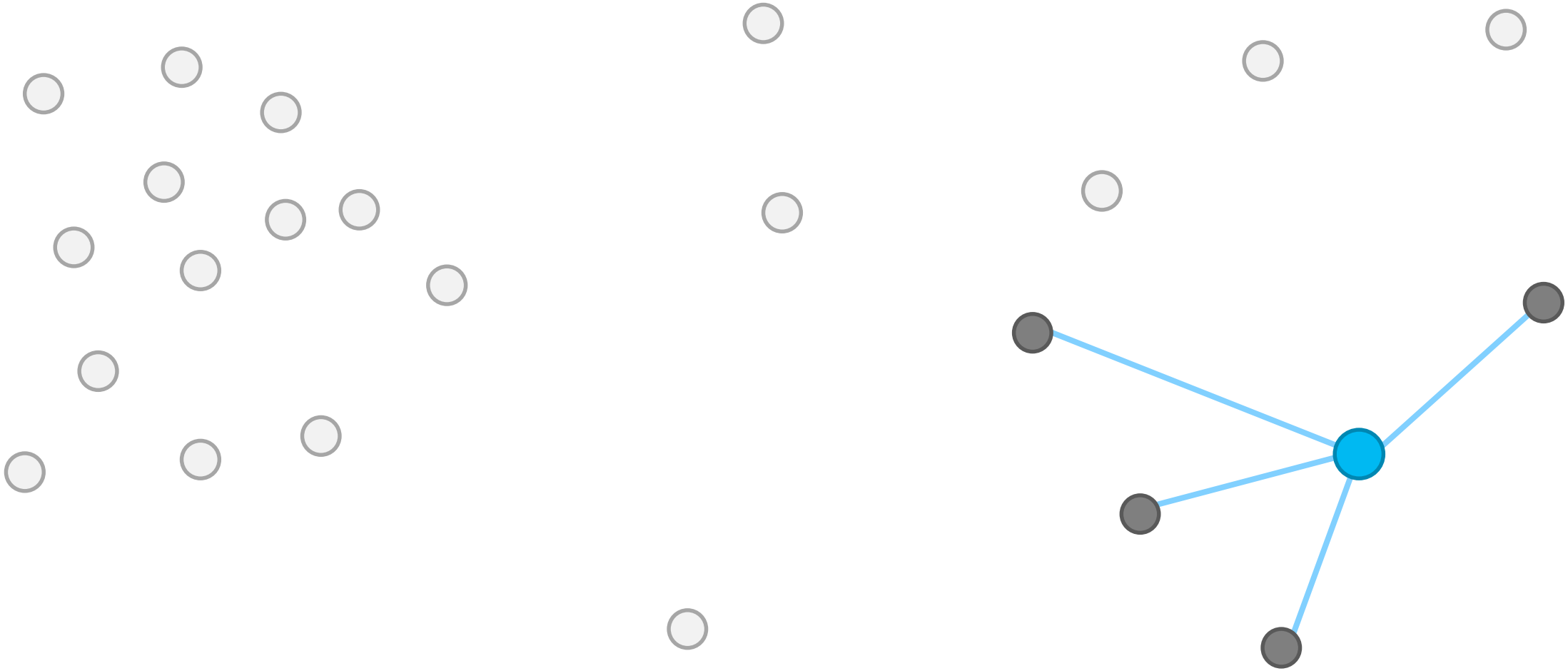


K Nearest Neighbors

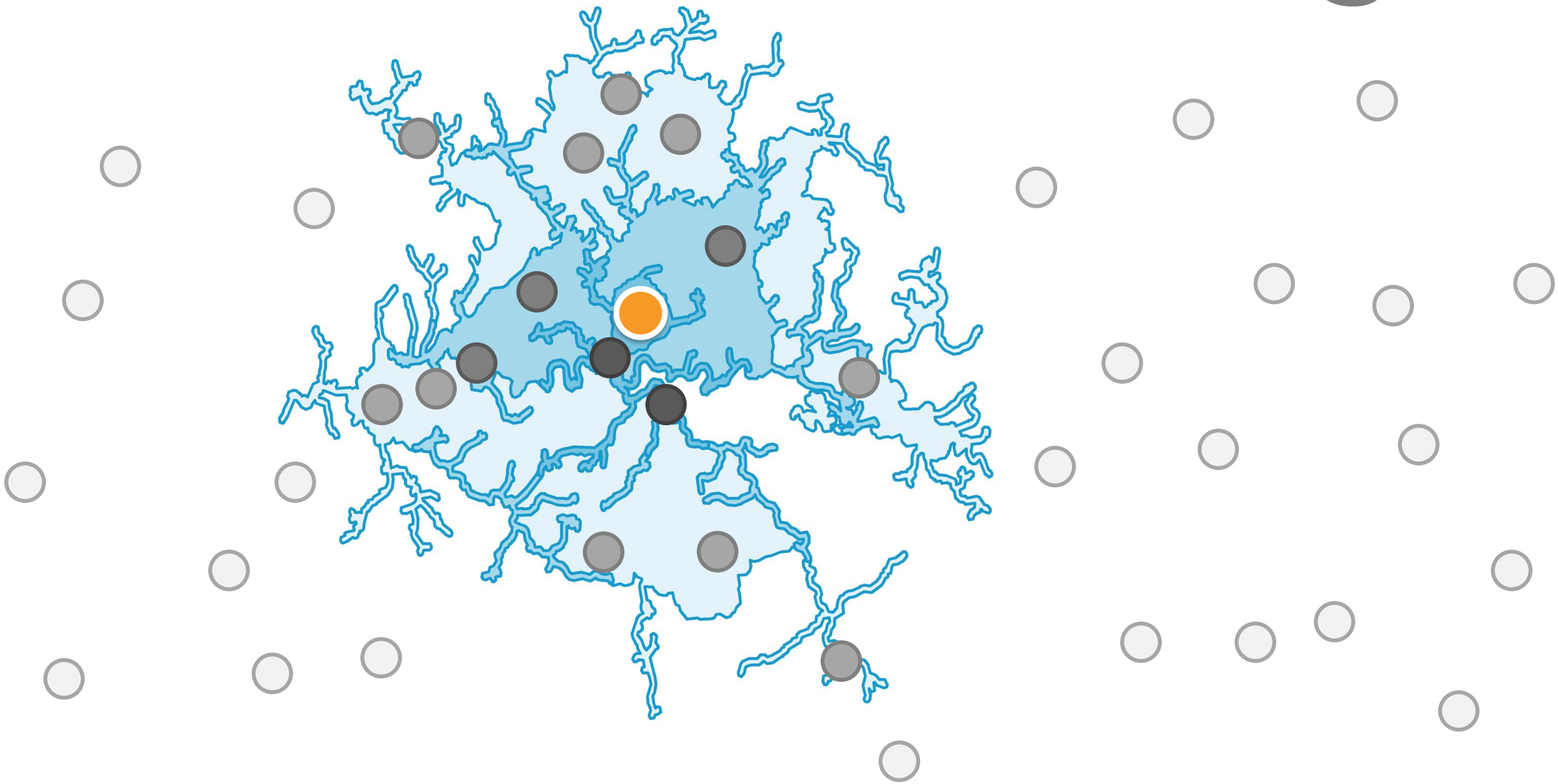
K Nearest Neighbors



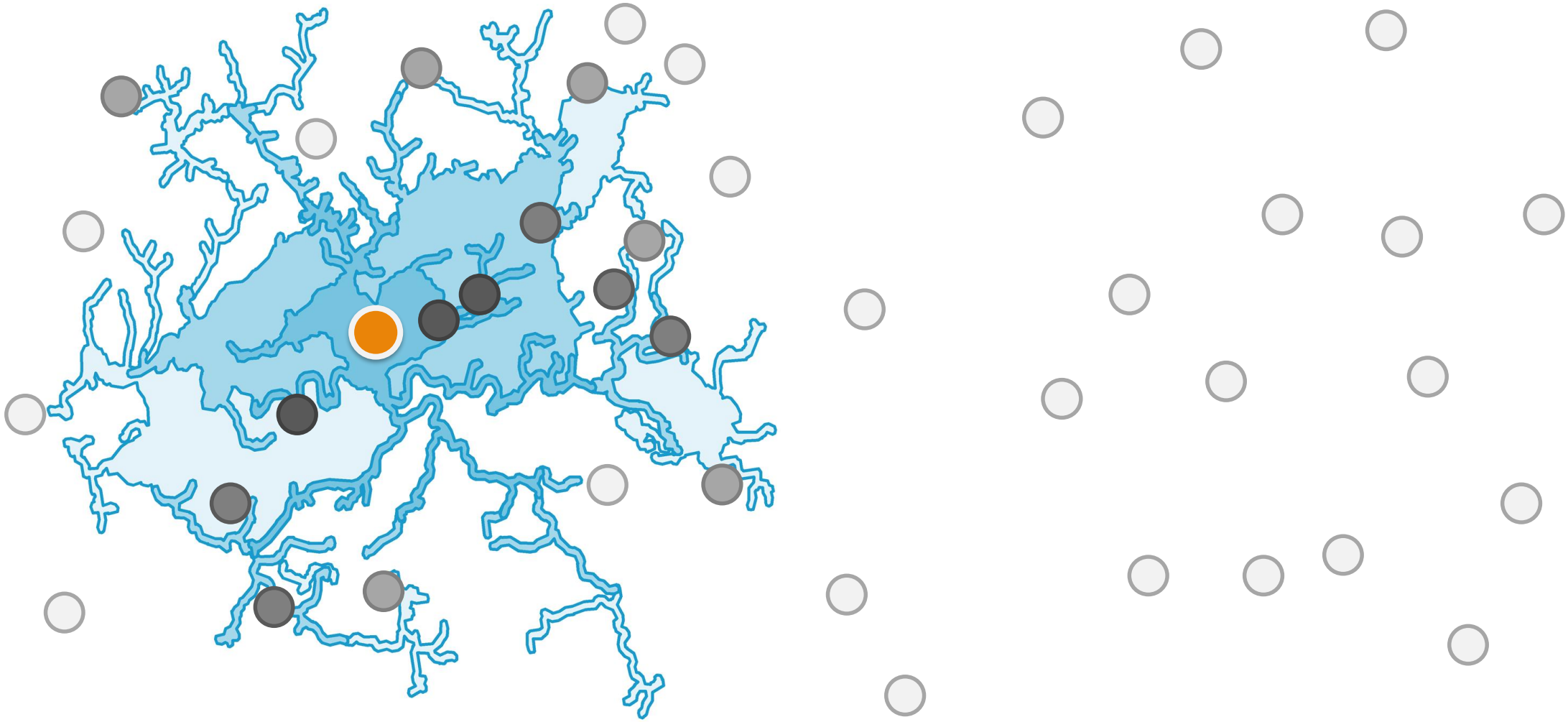
K Nearest Neighbors



Network Spatial Weights



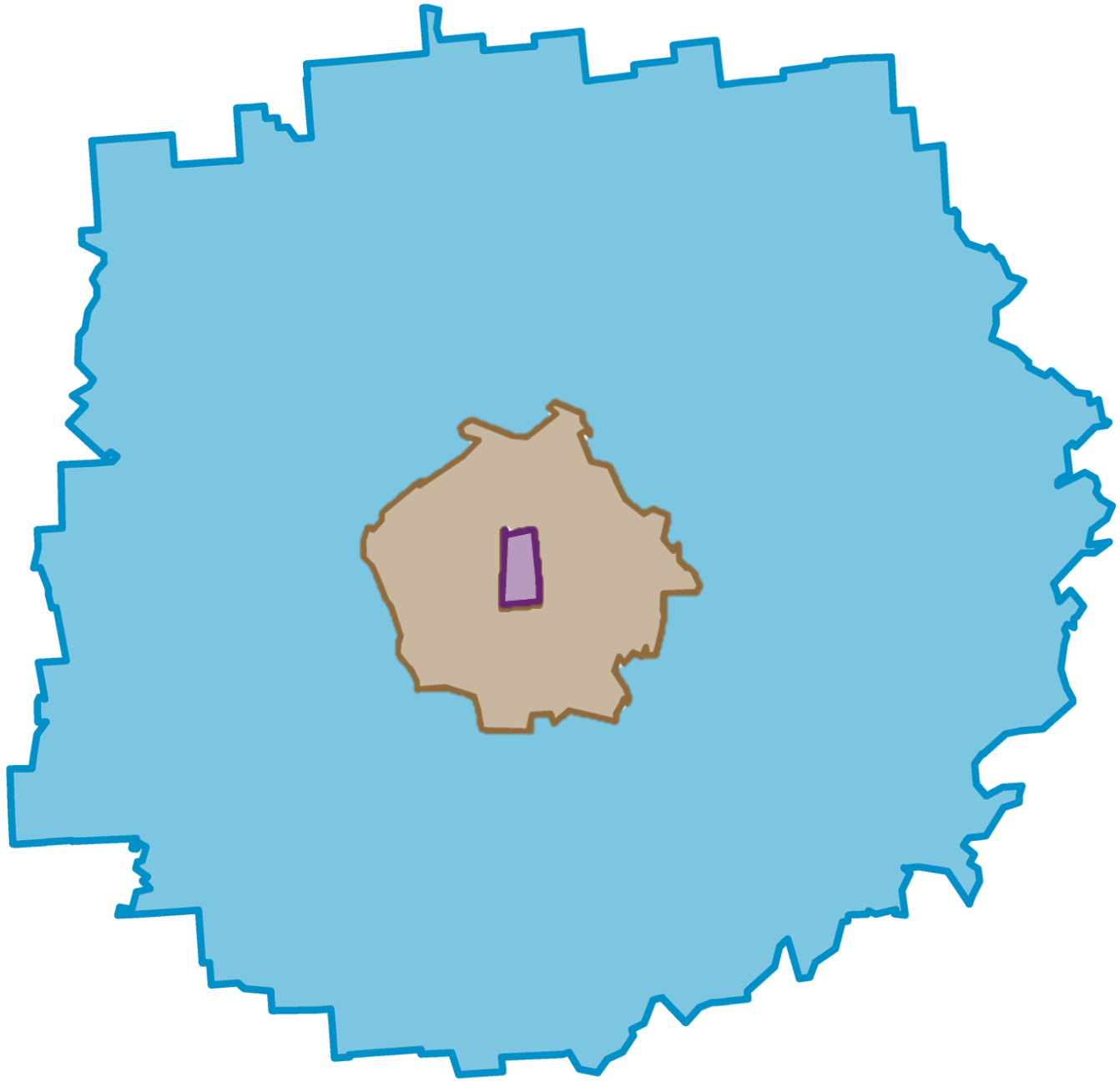
Network Spatial Weights

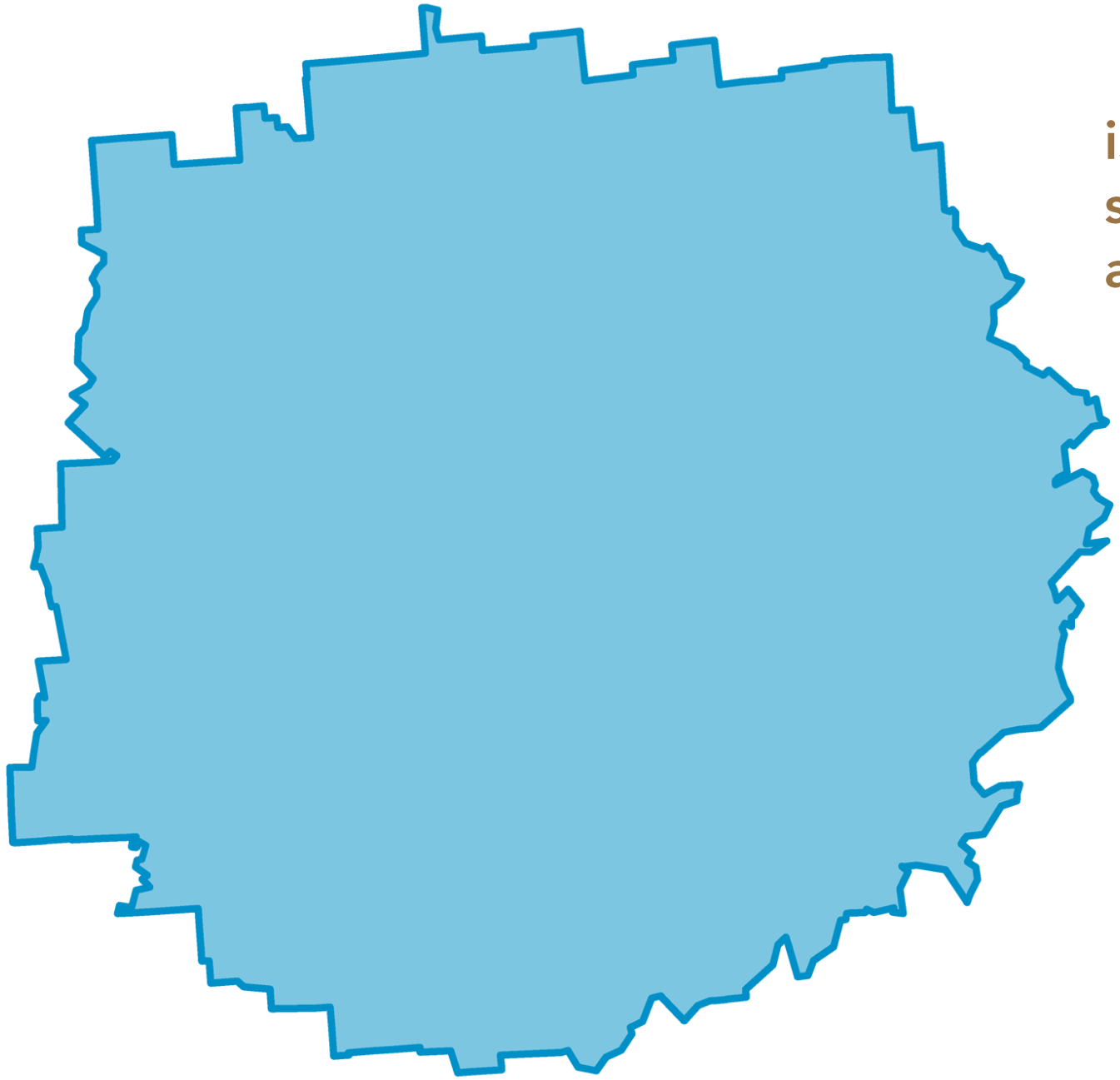


demo

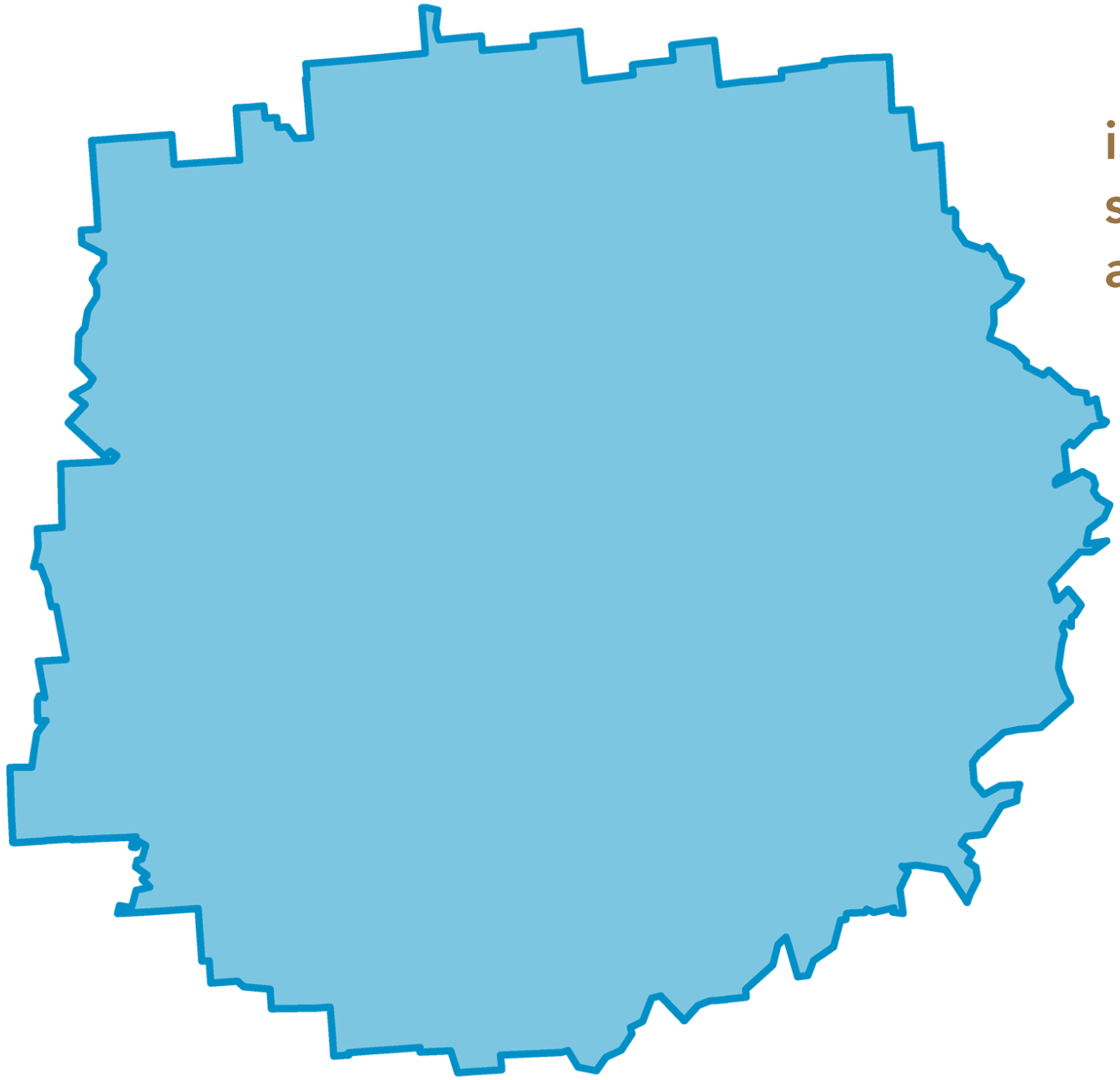
Cluster and Outlier Analysis

given a set of weighted features, identifies statistically significant hot spots, cold spots, and spatial outliers using the Anselin Local Moran's I statistic

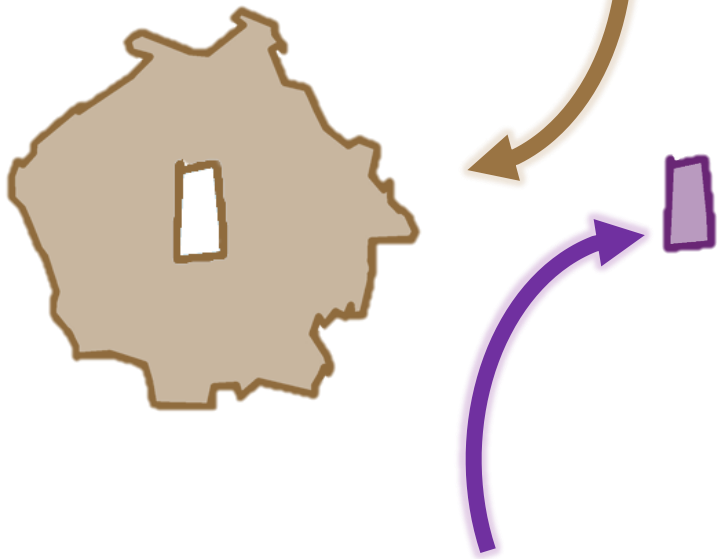




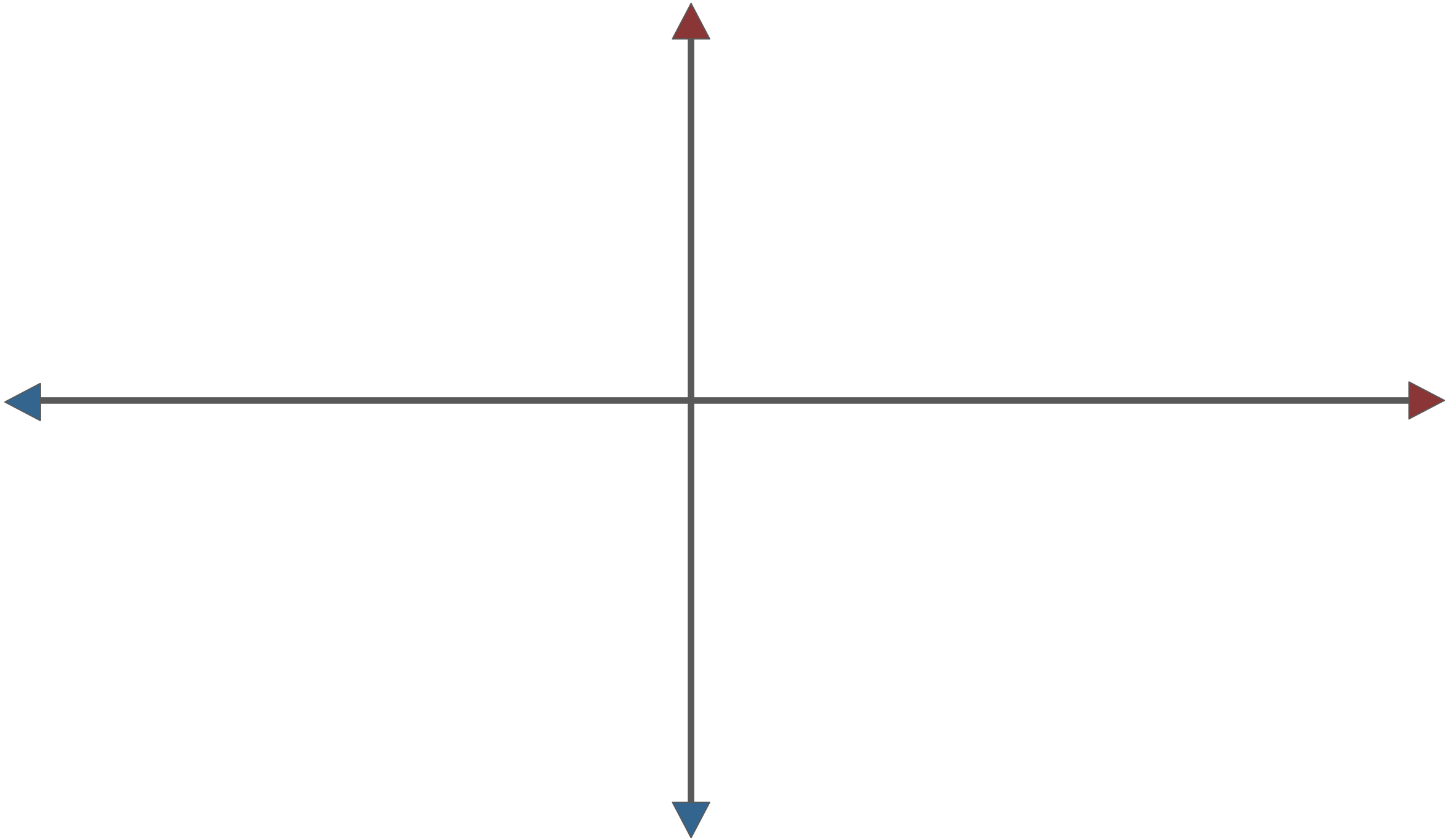
is the neighborhood
significantly different from
all other neighborhoods?

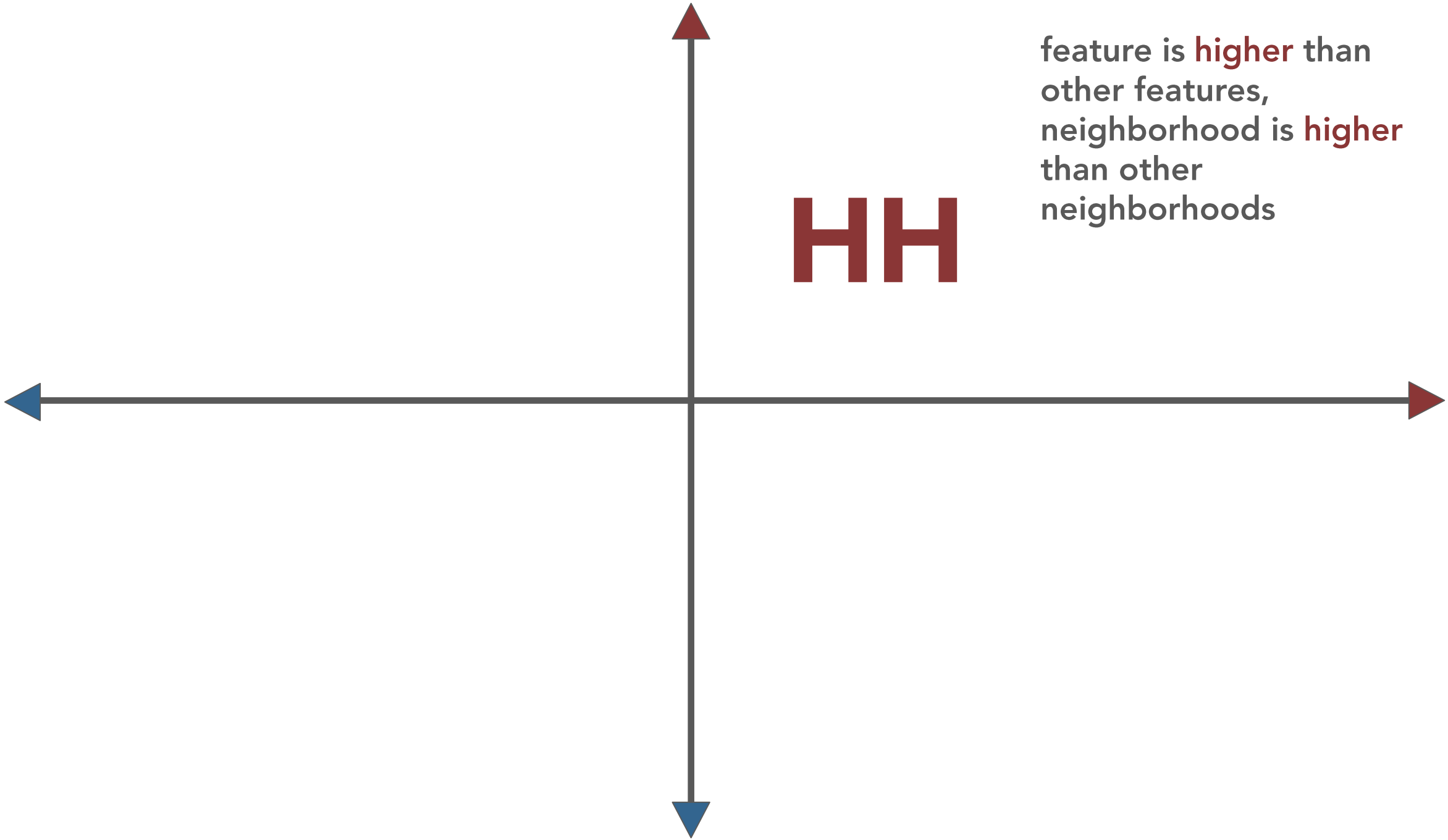


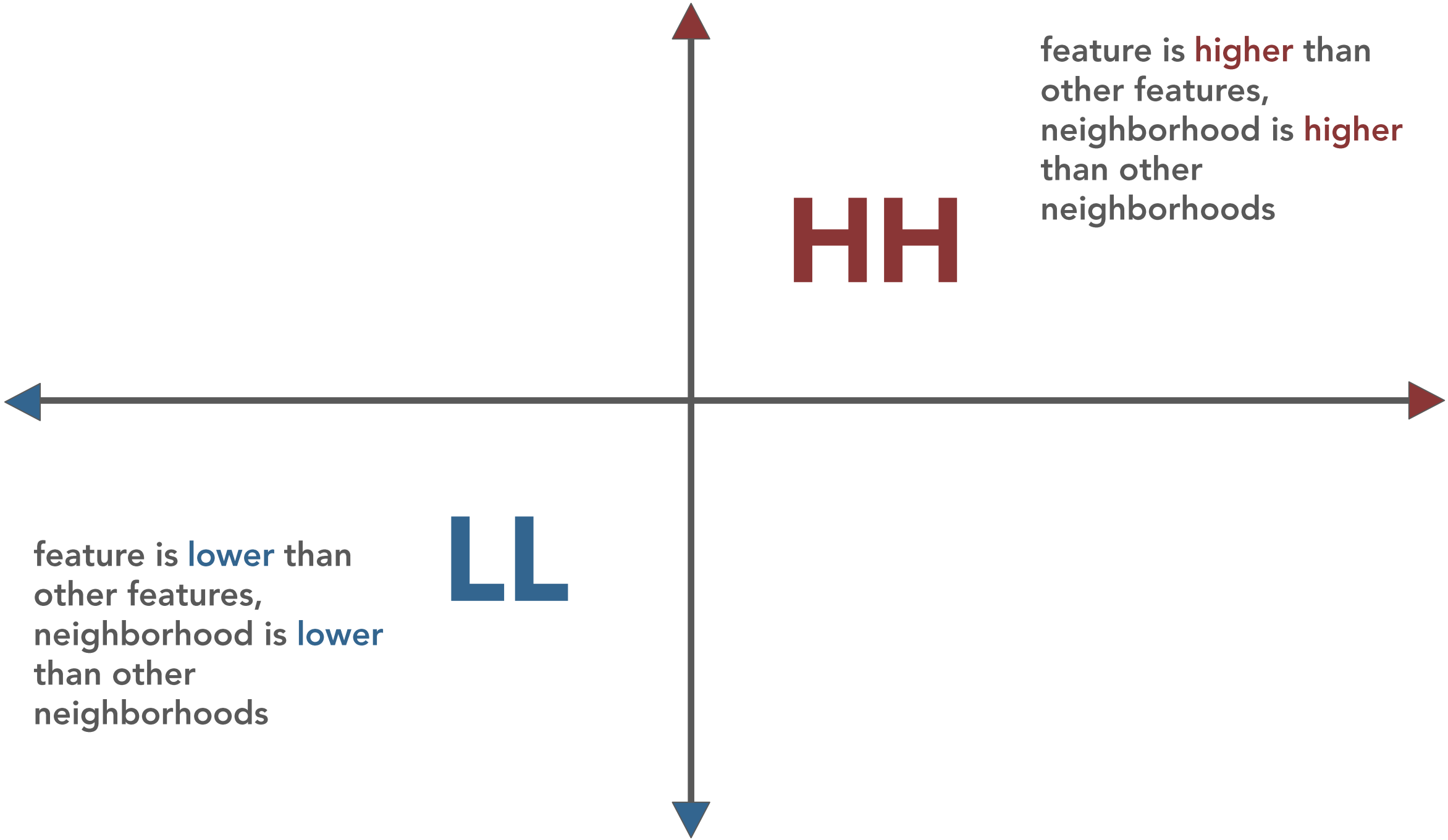
is the neighborhood significantly different from all other neighborhoods?

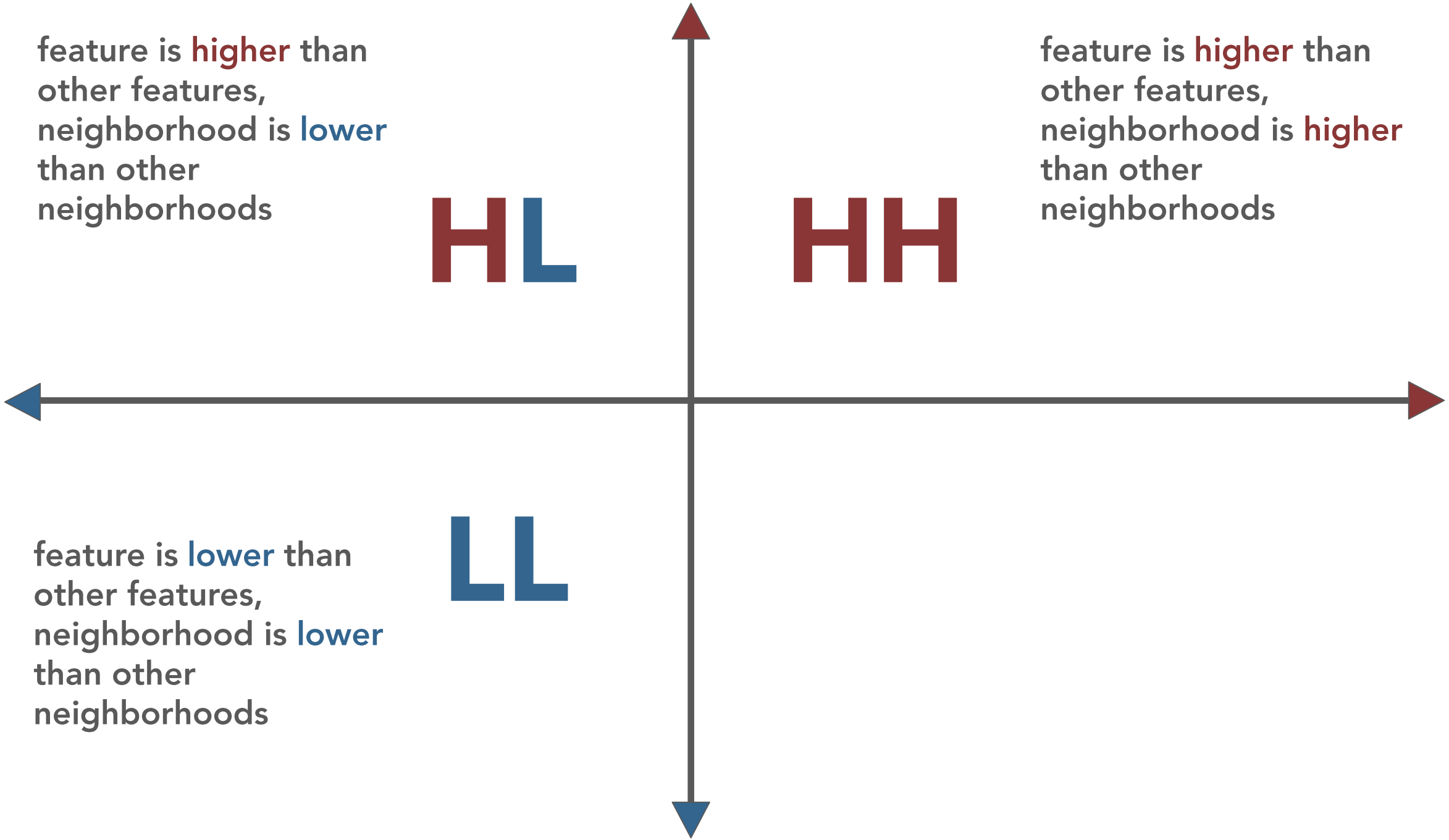


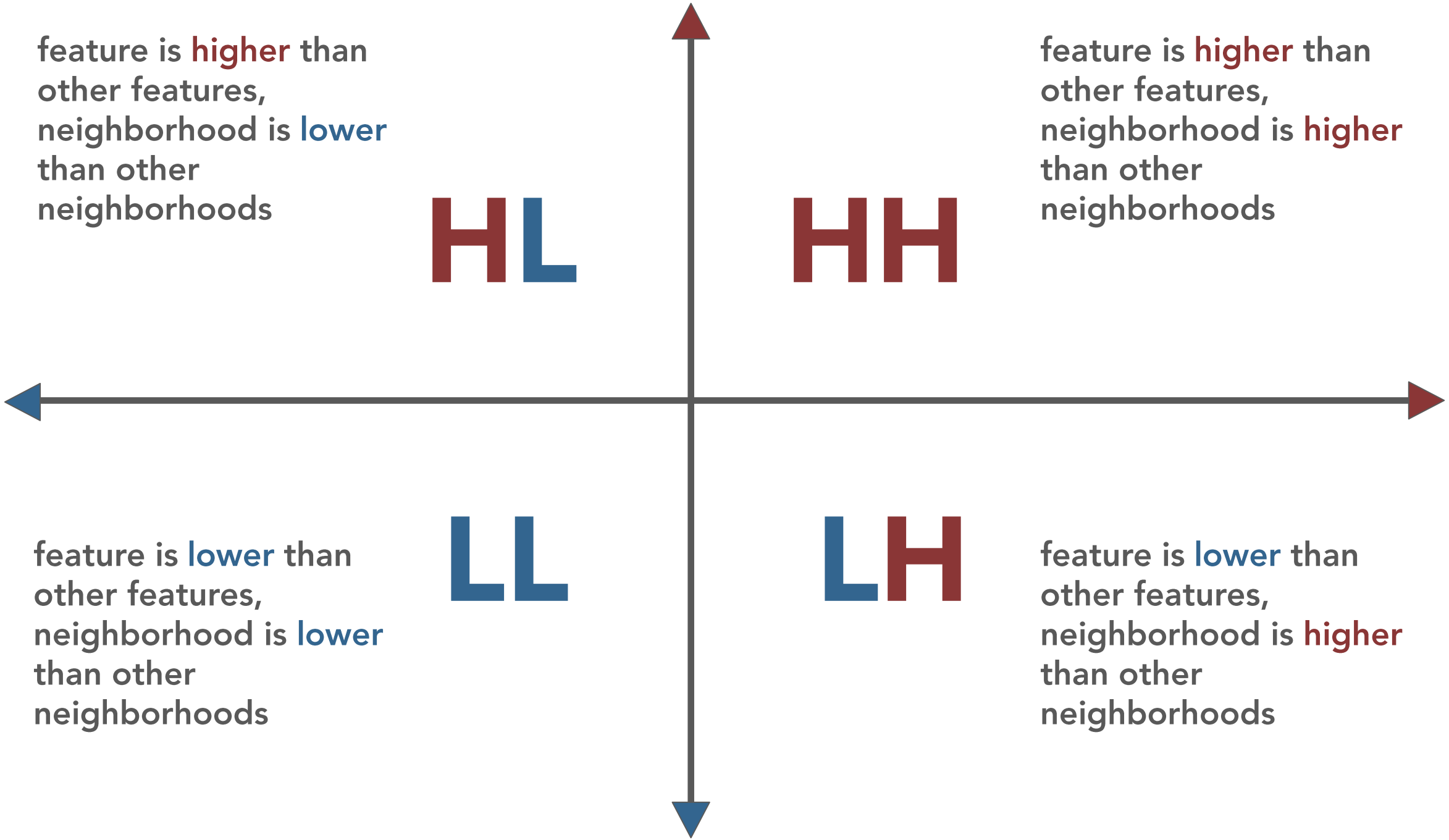
is the feature significantly different from all other features?



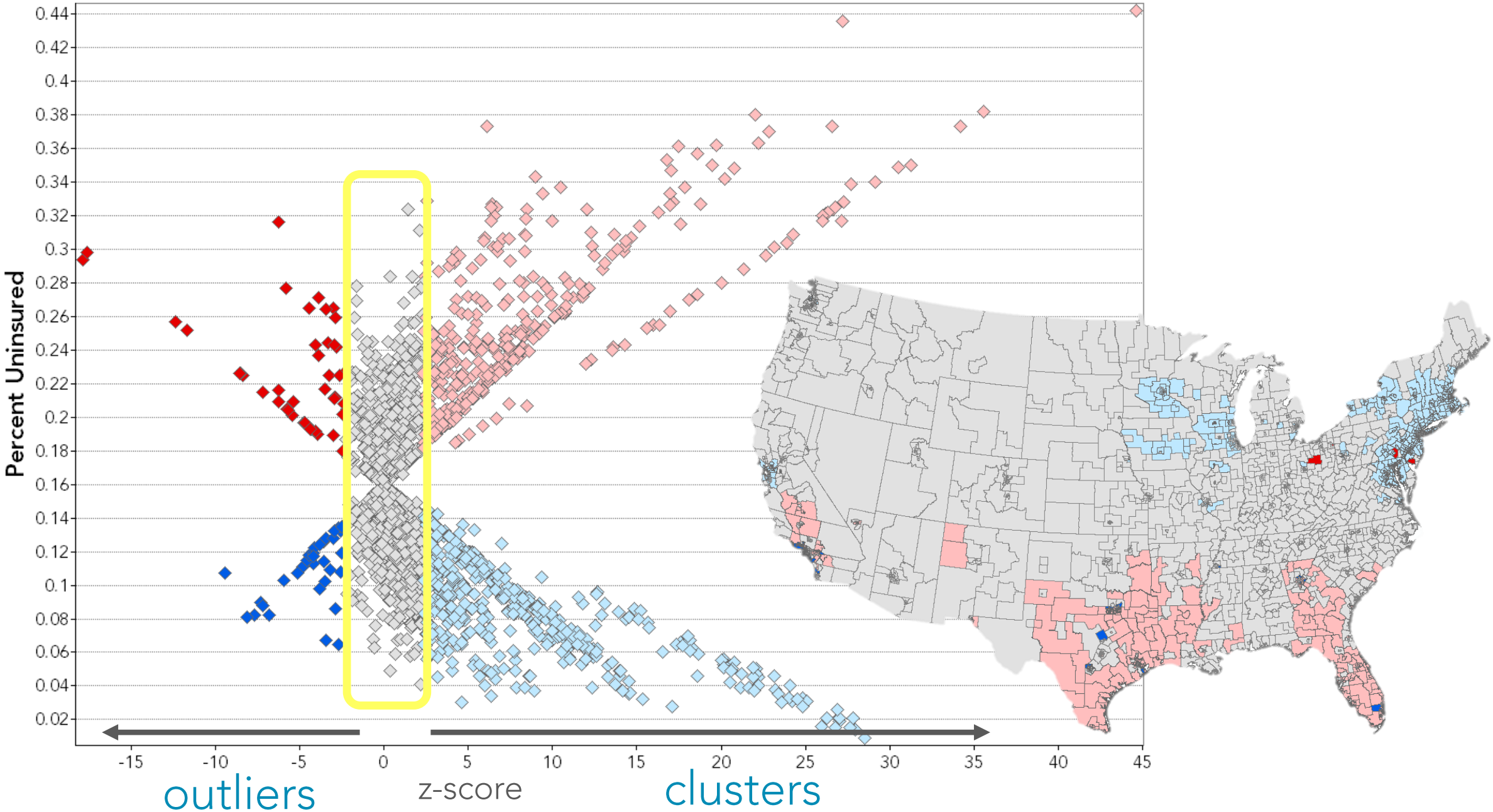

HH
feature is higher than other features, neighborhood is higher than other neighborhoods


HH
feature is higher than other features, neighborhood is higher than other neighborhoods
LL
feature is lower than other features, neighborhood is lower than other neighborhoods


feature is higher than other features, neighborhood is lower than other neighborhoods
HL
HH
feature is higher than other features, neighborhood is higher than other neighborhoods
feature is lower than other features, neighborhood is lower than other neighborhoods
LL


feature is higher than other features, neighborhood is lower than other neighborhoods
HL
HH
feature is higher than other features, neighborhood is higher than other neighborhoods
feature is lower than other features, neighborhood is lower than other neighborhoods
LL
LH
feature is lower than other features, neighborhood is higher than other neighborhoods

Cluster and Outlier Analysis

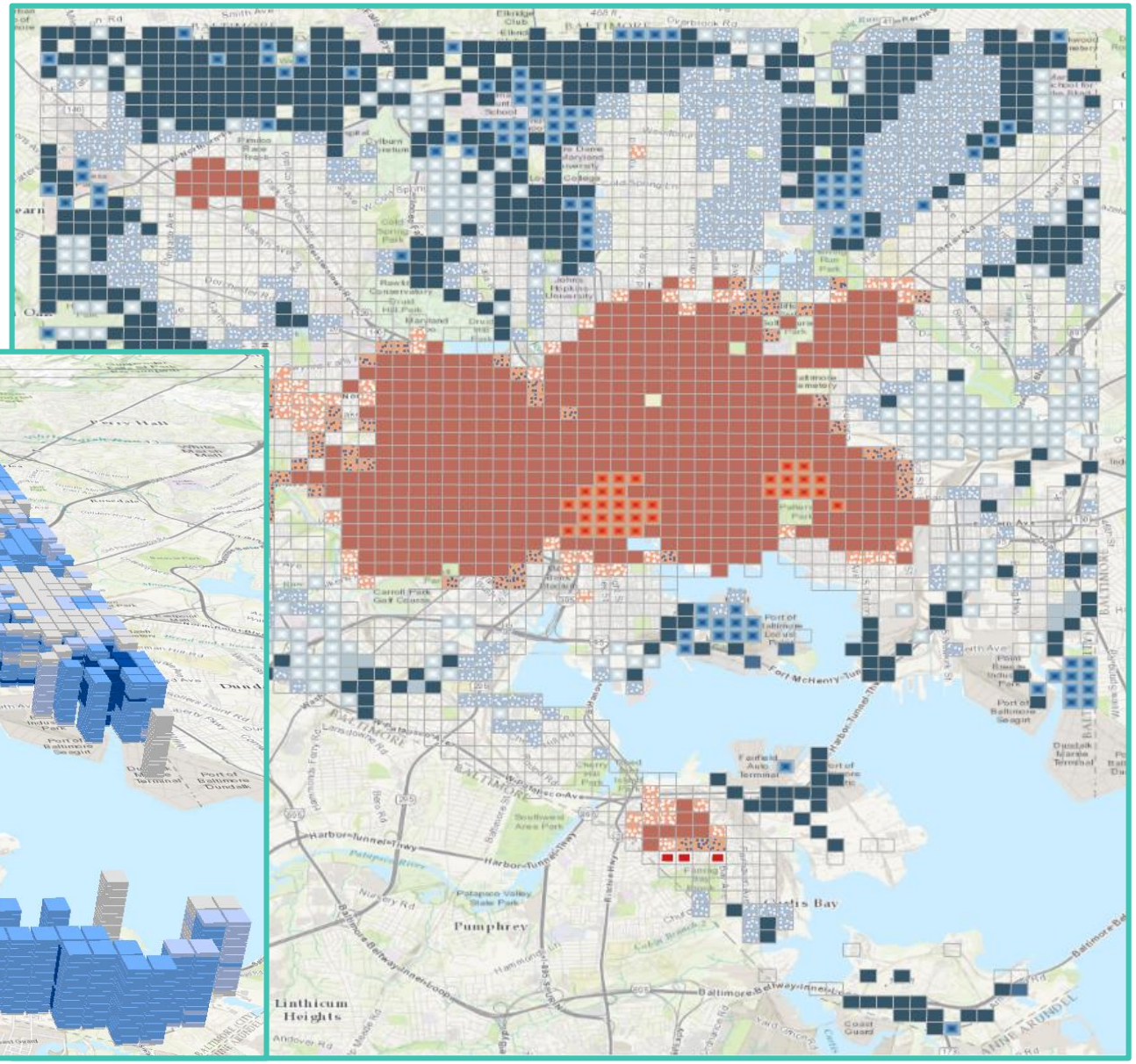


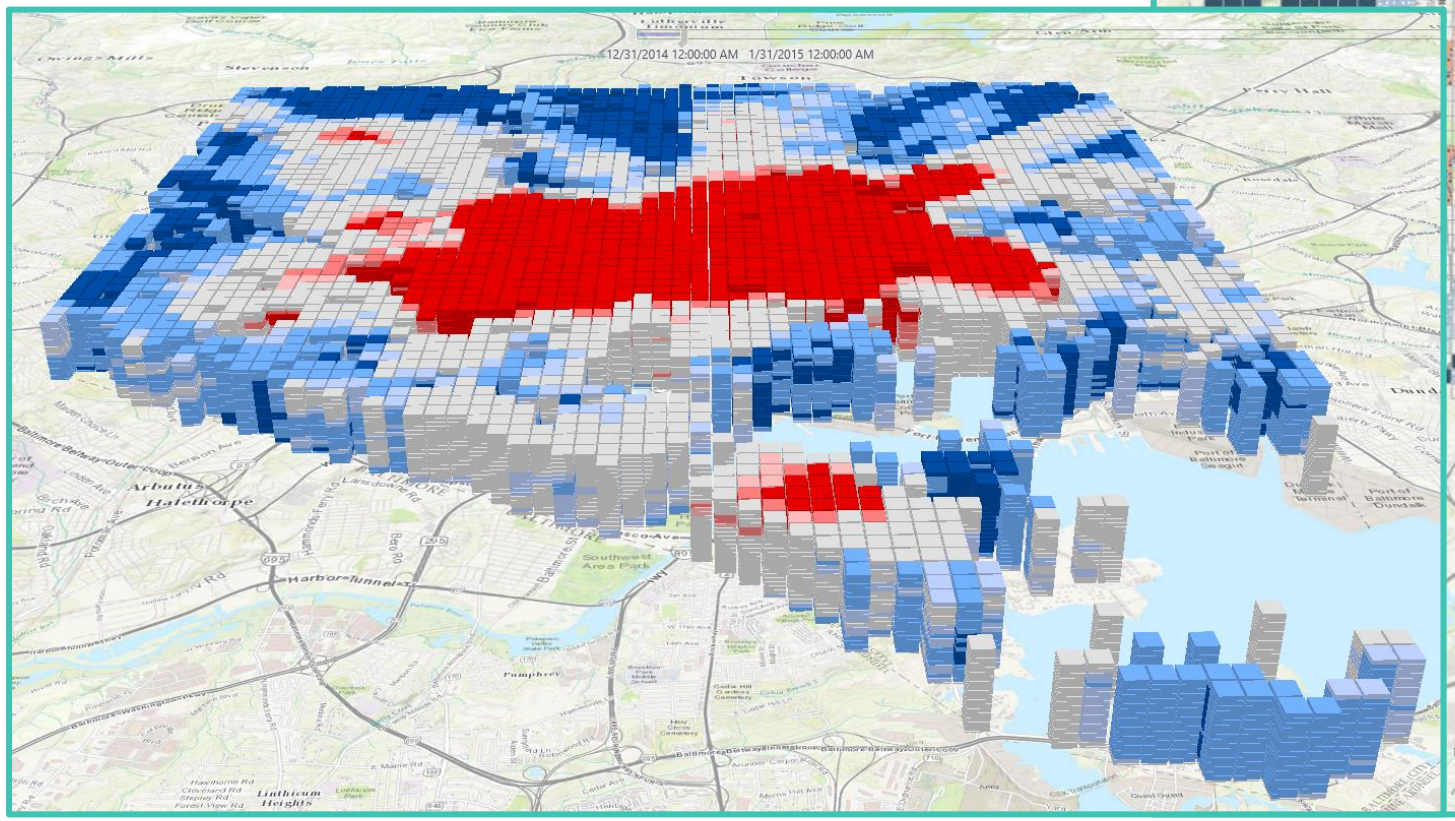
demo

Space Time Pattern Mining

Defining neighbors in space and time

911 Calls





Want to learn more???

esriurl.com/spatialstats

TUESDAY

~~8:30a From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 15B~~

~~10a Data Visualization for Spatial Analysis 10~~

~~2:30p Spatial Data Mining I: Essentials of Cluster Analysis 15B~~

~~4p From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 14B~~

WEDNESDAY

~~10a Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 29C~~

~~2:30p Spatial Data Mining I: Essentials of Cluster Analysis Room 15A~~

4p Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 31B

THURSDAY

10a Data Visualization for Spatial Analysis 07A/B

1p Beyond Where: Modeling Spatial Relationships and Making Predictions 17B

4p Beyond Where: Modeling Spatial Relationships and Making Predictions 17A

lbennett@esri.com
jdacosta@esri.com
fvale@esri.com

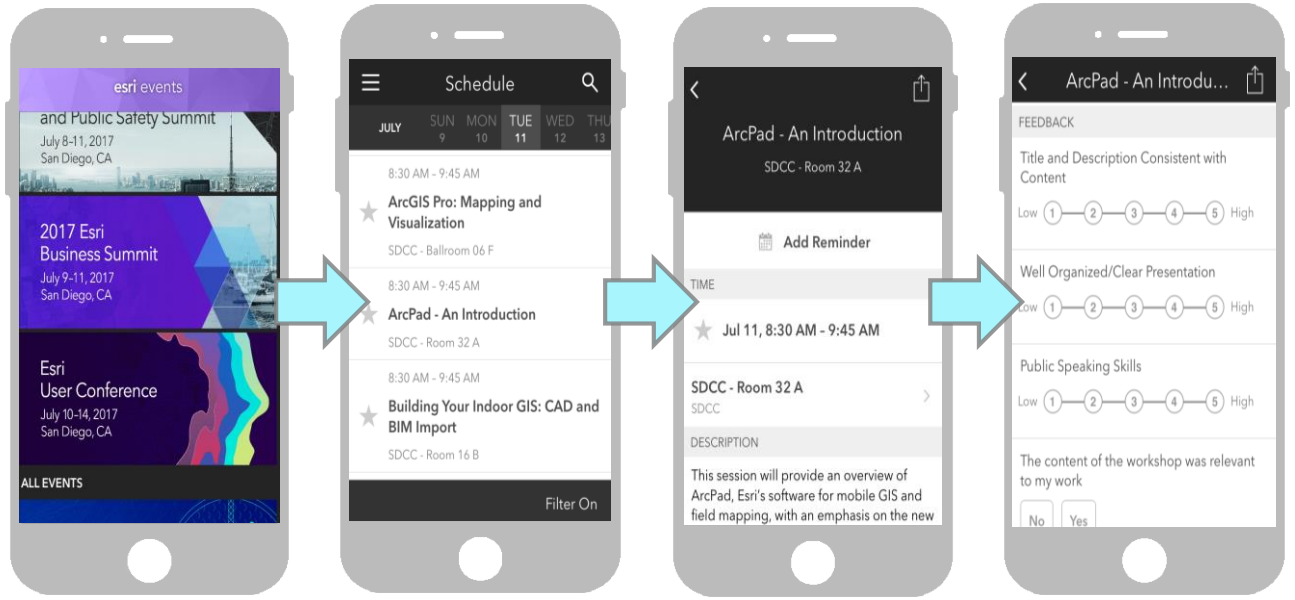
Please fill out a course survey!!

Download the Esri Events app and find your event

Select the session you attended

Scroll down to find the survey

Complete Answers and Select "Submit"



Want to learn more???

esriurl.com/spatialstats

TUESDAY

~~8:30a From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 15B~~

~~10a Data Visualization for Spatial Analysis 10~~

~~2:30p Spatial Data Mining I: Essentials of Cluster Analysis 15B~~

4p From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 14B

WEDNESDAY

10a Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 29C

2:30p Spatial Data Mining I: Essentials of Cluster Analysis Room 15A

4p Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 31B

THURSDAY

10a Data Visualization for Spatial Analysis 07A/B

1p Beyond Where: Modeling Spatial Relationships and Making Predictions 17B

4p Beyond Where: Modeling Spatial Relationships and Making Predictions 17A

esri | THE SCIENCE OF WHERE

lbennett@esri.com
jdacosta@esri.com
fvale@esri.com

Please fill out a course survey!!

Download the Esri Events app and find your event

Select the session you attended

Scroll down to find the survey

Complete Answers and Select "Submit"