Assessing Spatial Patterns Using Statistics

A session on spatial statistics

Dennis P. Dizon

2018 PSAI Annual Conference | September 2018 | Tagbilaran, Bohol

In this session…

- What is spatial analysis
- Adding spatial property to statistical data

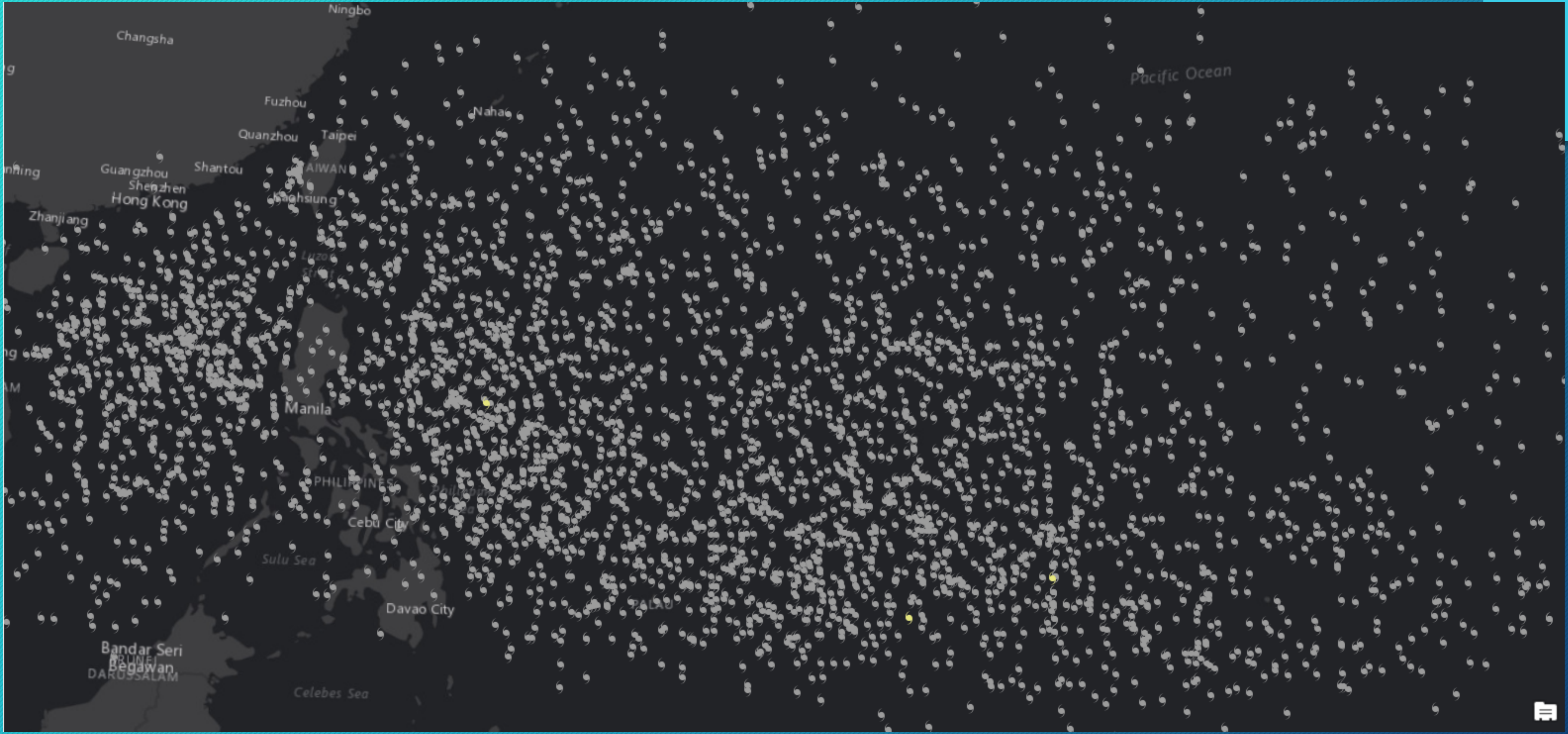
- Measuring geographic distributions
- Identifying patterns
- Identifying clusters

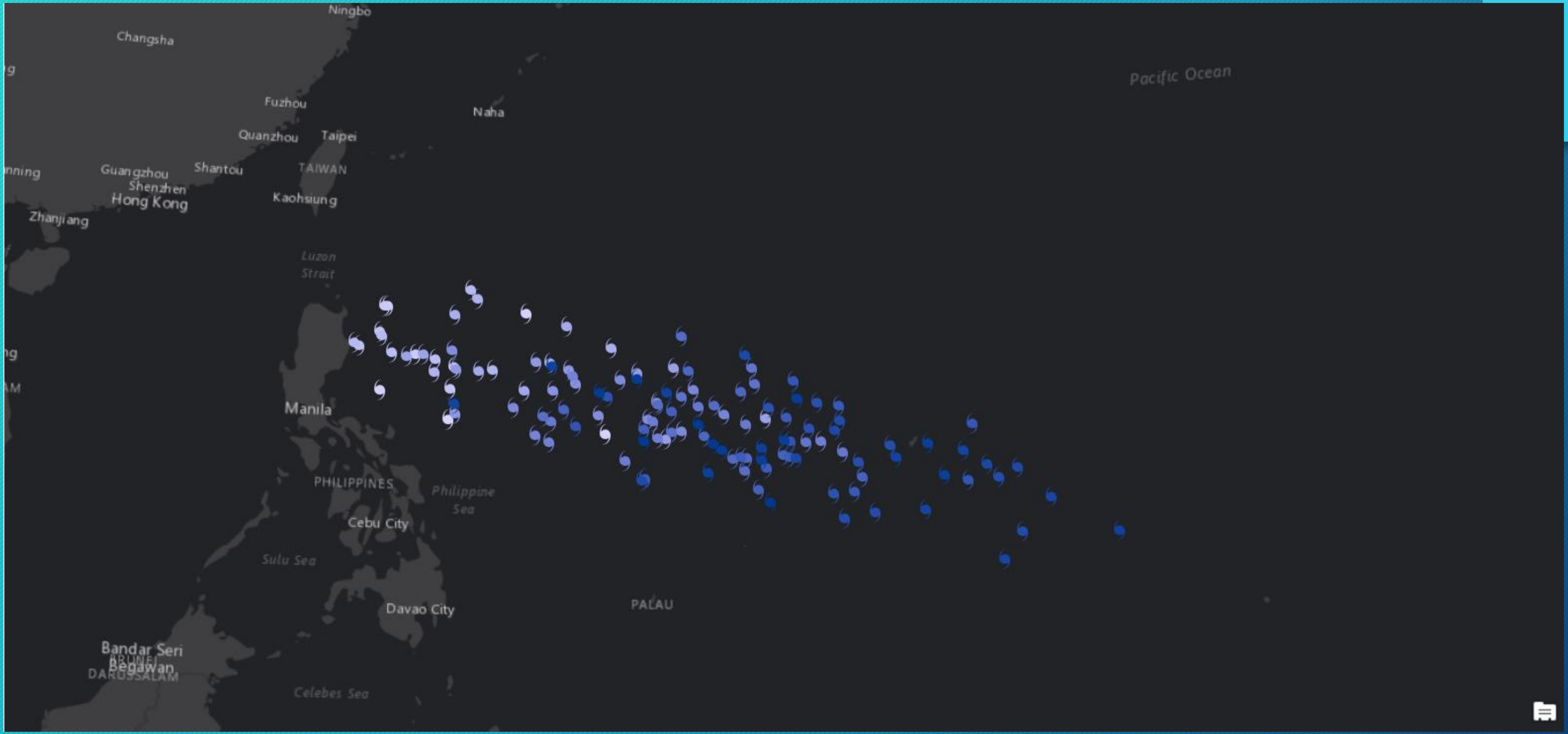
What is spatial analysis?

Serial_Num	Season	Num	Basin	Sub_basin	Name	ISO_time	Nature	wmo_pres	Center	wmo_pres_%
1951050N20139	1951	1	WP	MM	NOT NAMED	1951-02-19 06:00:00	TS	1010	tokyo	0.86
1951077N06158	1951	2	WP	MM	01W:GEORGIA	1951-03-18 06:00:00	TS	1002	tokyo	15.85
1951105N08152	1951	3	WP	MM	02W:HOPE	1951-04-15 00:00:00	TS	1002	tokyo	15.85
1951119N06145	1951	4	WP	MM	03W:IRIS	1951-04-28 12:00:00	TS	1008	tokyo	1.51
1951203N22144	1951	12	WP	MM	NOT NAMED	1951-07-22 00:00:00	TS	1008	tokyo	1.51
1951221N10136	1951	17	WP	MM	NOT NAMED	1951-08-09 00:00:00	TS	1006	tokyo	4.44
1951223N11148	1951	18	WP	MM	07W:MARGE	1951-08-10 12:00:00	TS	987	tokyo	55.59
1951239N11144	1951	20	WP	MM	08W:NORA	1951-08-27 00:00:00	TS	1008	tokyo	1.51
1951255N08148	1951	23	WP	MM	09W:ORA	1951-09-11 18:00:00	TS	1006	tokyo	4.44
1951263N08137	1951	24	WP	MM	10W:PAT	1951-09-20 00:00:00	TS	1008	tokyo	1.51
1951295N21152	1951	28	WP	MM	12W:SARAH	1951-10-22 00:00:00	TS	1010	tokyo	0.86
1951301N16128	1951	30	WP	MM	14W:VERA	1951-10-28 06:00:00	TS	1006	tokyo	4.44
1951321N09143	1951	31	WP	MM	15W:WANDA	1951-11-16 12:00:00	TS	1006	tokyo	4.44
1951337N09150	1951	33	WP	MM	16W:AMY	1951-12-03 00:00:00	NR	1000	tokyo	21.05
1951344N09143	1951	34	WP	MM	17W:BABS	1951-12-10 00:00:00	TS	1006	tokyo	4.44
1952162N15118	1952	2	WP	MM	01W:CHARLOTT:CHARLOTTE	1952-06-09 12:00:00	TS	992	tokyo	43.9
1952171N12130	1952	3	WP	MM	02W:DINAH	1952-06-19 00:00:00	TS	1006	tokyo	4.44
1952180N05144	1952	5	WP	MM	03W:EMMA	1952-06-28 00:00:00	TS	1002	tokyo	15.85
1952236N10135	1952	16	WP	MM	09W:LOIS	1952-08-22 12:00:00	TS	1008	tokyo	1.51
1952241N08137	1952	17	WP	MM	MARY	1952-08-28 00:00:00	TS	1004	tokyo	8.41
1952245N09139	1952	19	WP	MM	11W:NONA	1952-09-01 00:00:00	TS	1006	tokyo	4.44
1952269N11157	1952	25	WP	MM	15W:POLLY	1952-09-24 12:00:00	TS	1004	tokyo	8.41
1952278N22140	1952	26	WP	MM	16W:ROSE	1952-10-04 06:00:00	TS	1006	tokyo	4.44
1952289N15128	1952	29	WP	MM	19W:VAE	1952-10-15 06:00:00	TS	1006	tokyo	4.44
1952312N08150	1952	32	WP	MM	22W:BESS	1952-11-06 18:00:00	TS	1008	tokyo	1.51

3 of 3853 selected

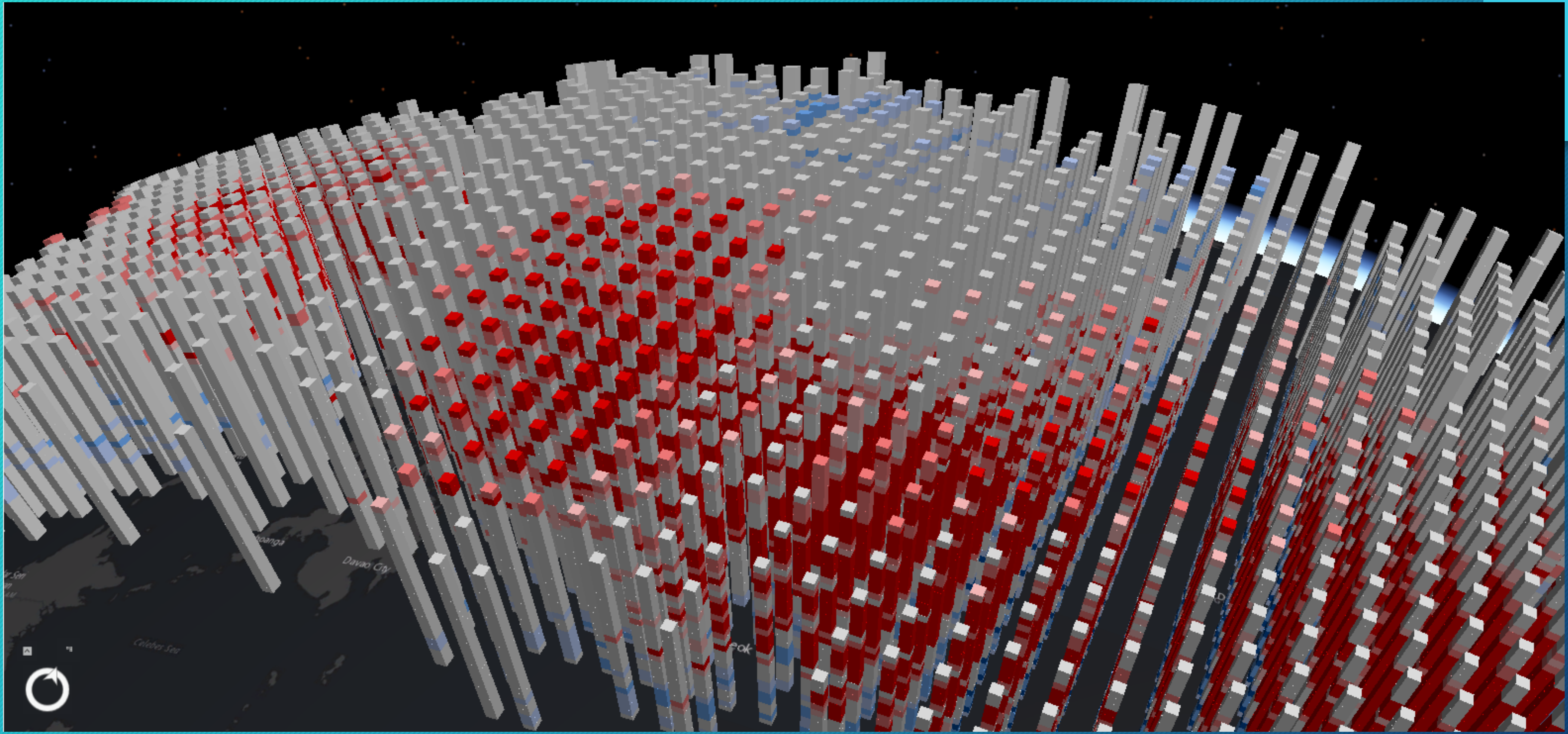
Filters: 100%


Ningbo
Changsha
Fuzhou
Naha
Quanzhou
Taipei
Guangzhou
Shantou
Shenzhen
Hong Kong
Kaohsiung
Zhanjiang
Pacific Ocean
Manila
PHILIPPINES
Cebu City
Sulu Sea
Davao City
PALAU
Bandar Seri Begawan
BRUNEI DARUSSALAM
Celebes Sea



Ningbo
Changsha
Fuzhou
Naha
Pacific Ocean
Quanzhou
Taipei
TAIWAN
Guangzhou
Shantou
Shenzhen
Hong Kong
Kaohsiung
Zhanjiang
Luzon Strait
Manila
PHILIPPINES
Philippine Sea
Cebu City
Sulu Sea
Davao City
PALAU
Bandar Seri Begawan
BRUNEI DARUSSALAM
Celebes Sea

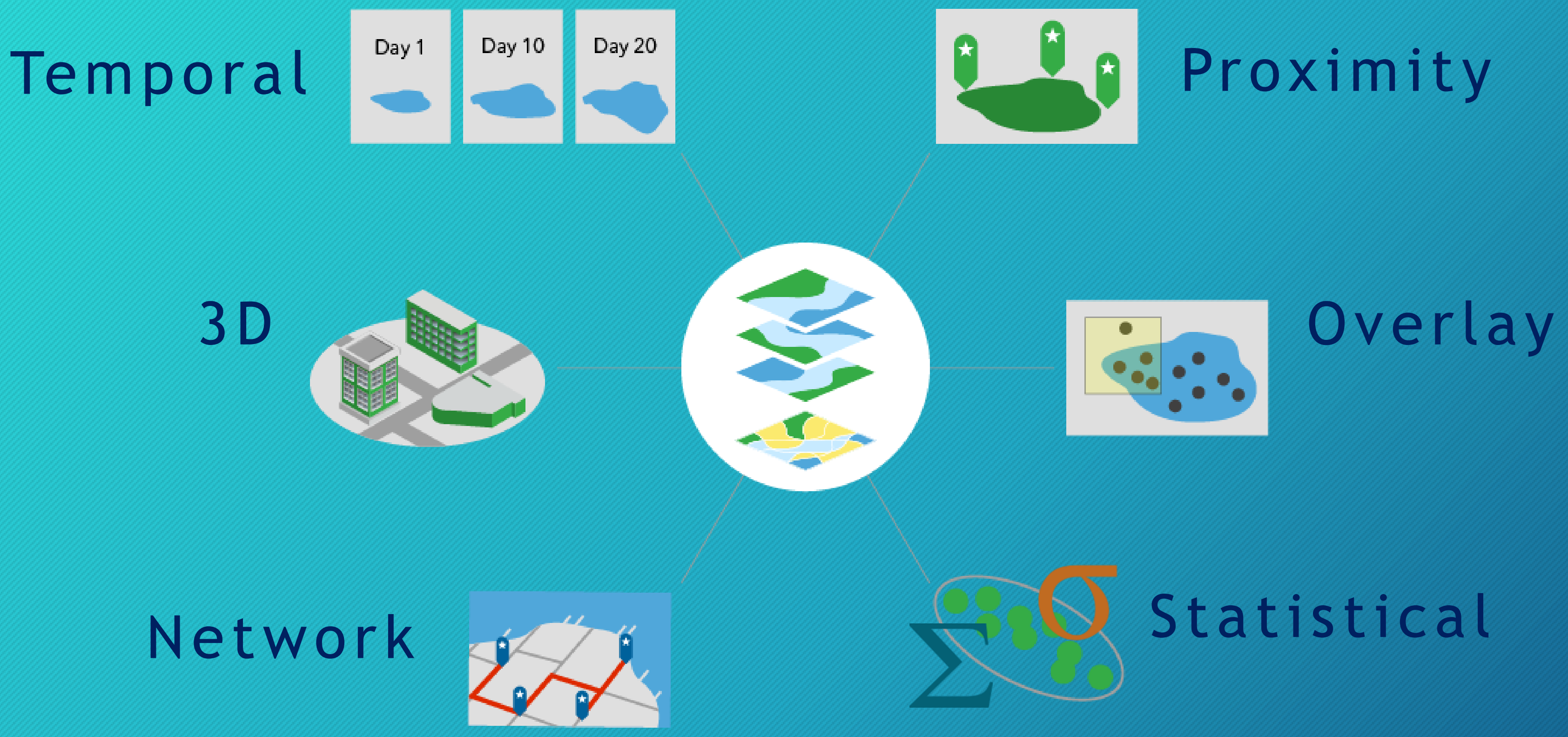

Ningbo
Changsha
Fuzhou
Naha
Quanzhou
Taipei
Guangzhou
Shantou
Shenzhen
Hong Kong
Kaohsiung
Zhanjiang
Manila
Cebu City
Sulu Sea
Davao City
Bandar Seri Begawan
Celebes Sea


Davao City
Celebes Sea

"The process of examining locations, attributes, and their relationships to help answer questions and solve geographical problems."

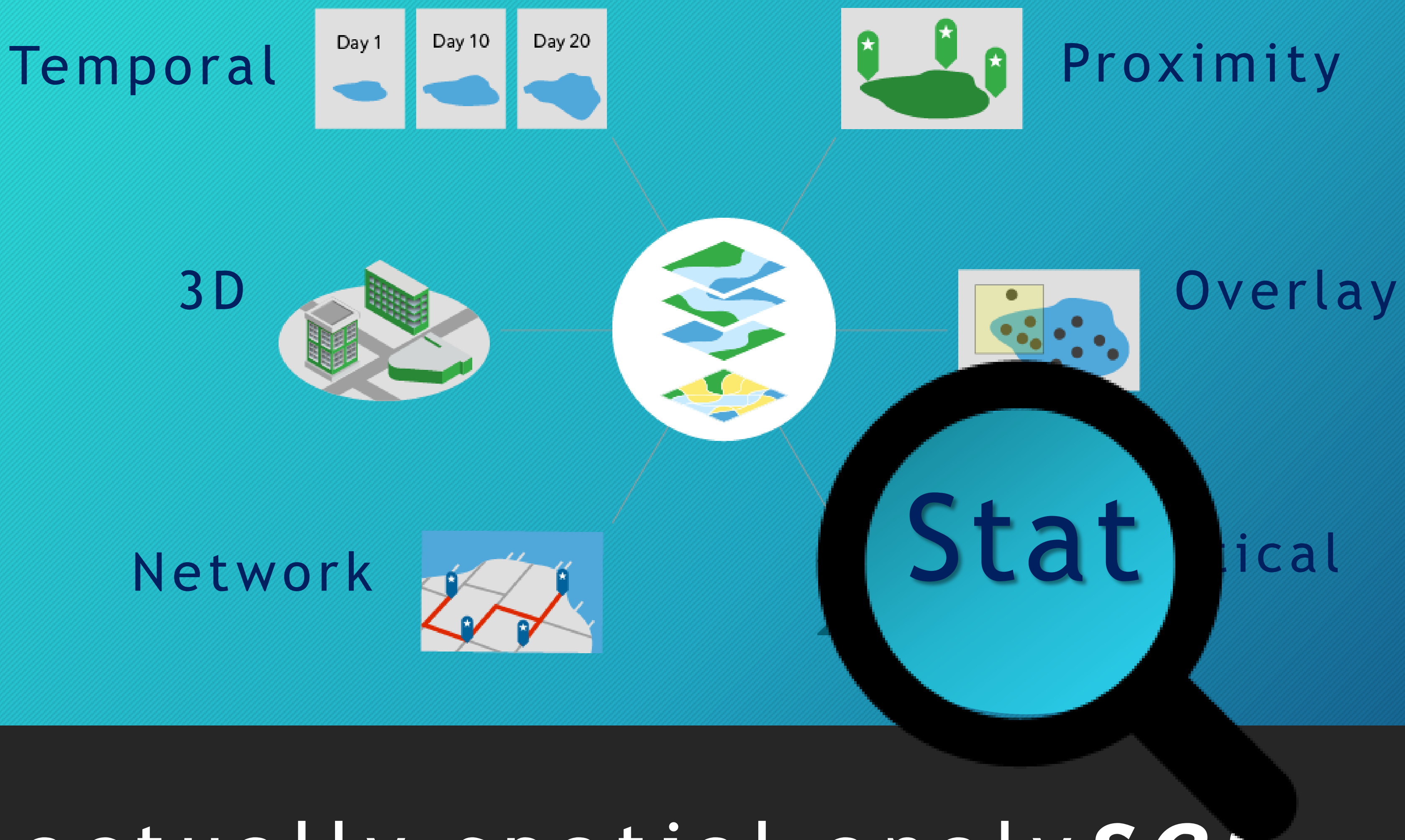
Esri Inc., 2018

Definition: spatial analysis



It's actually spatial analy***ses***

Image credits: Esri Inc., 2018



It’s actually spatial analy*ses*

Image credits: Esri Inc., 2018

“

The application of statistical tools and methods that use **space** and **spatial relationships** directly into their mathematical computation.

”

Esri Inc., 2018

Definition: spatial statistics

“*Space*” aspects, such as…

location coordinates

neighborhood

distances

orientation

address

area

height

boundary contiguity

volume

terrain

centrality

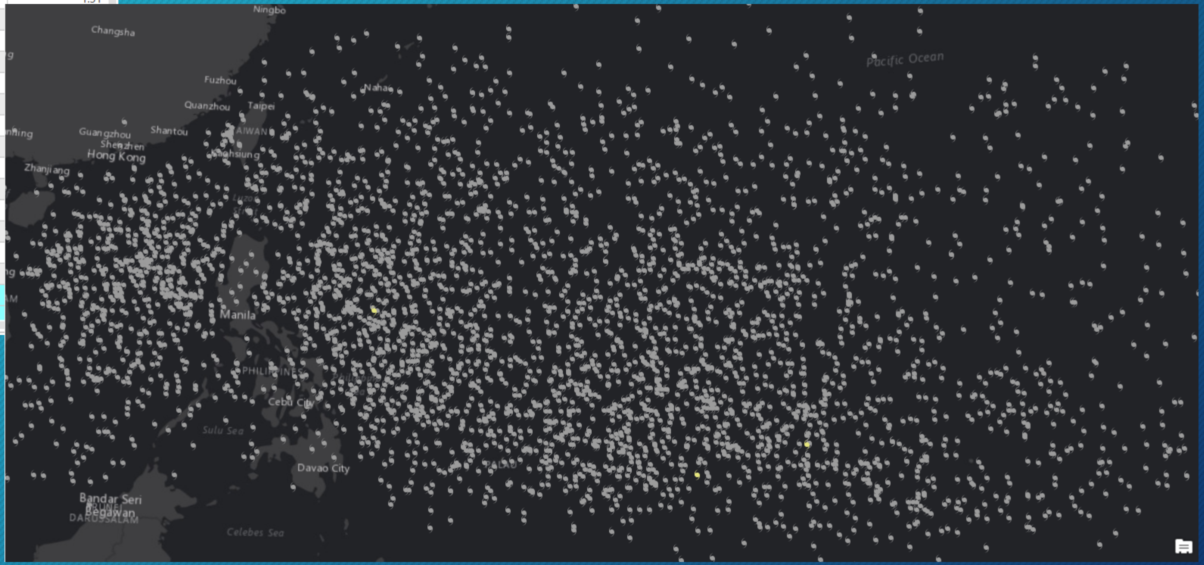
and so on…

“Spatializing” your data

How to turn data…

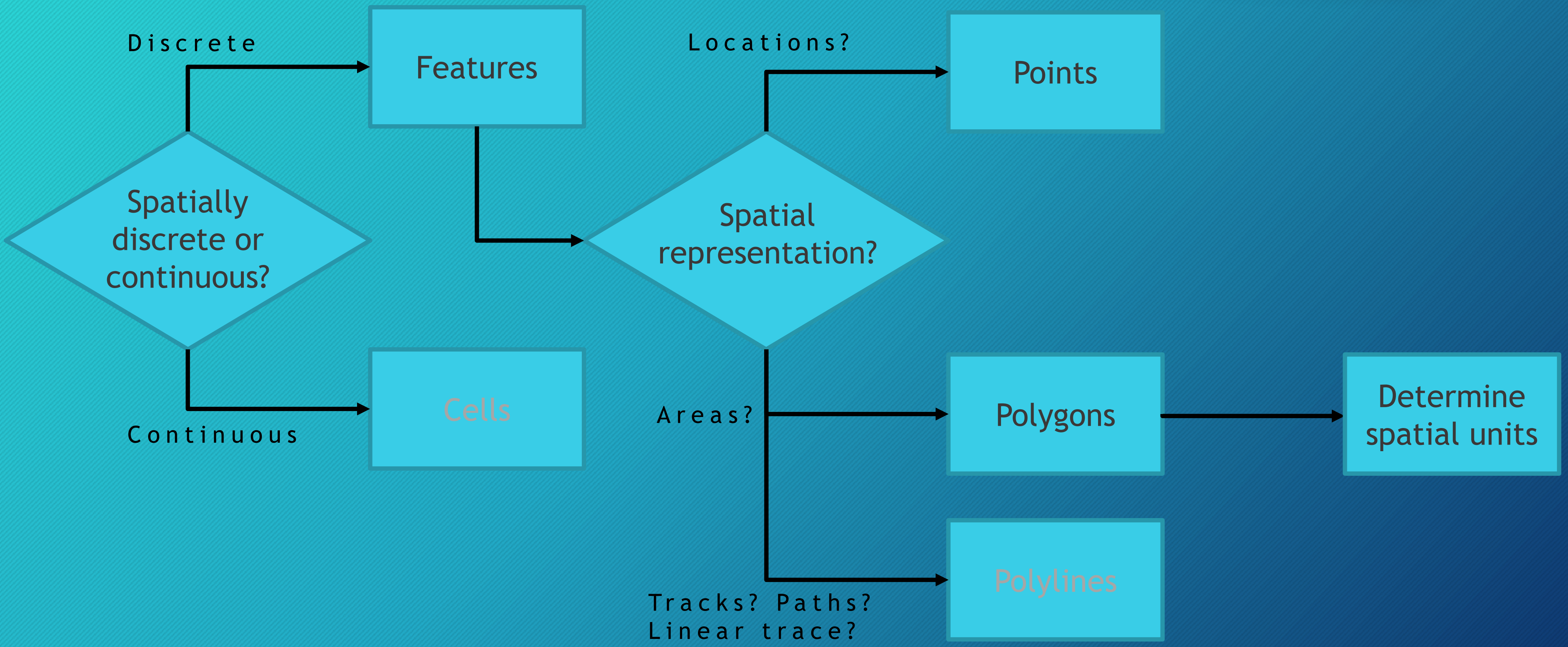
Serial_Num	Season	Num	Basin	Sub_basin	Name	ISO_time	Nature	wmo_pres	Center	wmo_pres_%
1951050N20139	1951	1	WP	MM	NOT NAMED	1951-02-19 06:00:00	TS	1010	tokyo	0.86
1951077N06158	1951	2	WP	MM	01W:GEORGIA	1951-03-18 06:00:00	TS	1002	tokyo	15.85
1951105N08152	1951	3	WP	MM	02W:HOPE	1951-04-15 00:00:00	TS	1002	tokyo	15.85
1951119N06145	1951	4	WP	MM	03W:IRIS	1951-04-28 12:00:00	TS	1008	tokyo	1.51
1951203N22144	1951	12	WP	MM	NOT NAMED	1951-07-22 00:00:00	TS	1008	tokyo	1.51
1951221N10136	1951	17	WP	MM	NOT NAMED	1951-08-09 00:00:00	TS	1006	tokyo	4.44
1951223N11148	1951	18	WP	MM	07W:MARGE	1951-08-10 12:00:00	TS	987	tokyo	55.59
1951239N11144	1951	20	WP	MM	08W:NORA	1951-08-27 00:00:00	TS	1008	tokyo	1.51
1951255N08148	1951	23	WP	MM	09W:ORA	1951-09-11 18:00:00	TS	1006	tokyo	4.44
1951263N08137	1951	24	WP	MM	10W:PAT	1951-09-20 00:00:00	TS	1008	tokyo	1.51
1951295N21152	1951	28	WP	MM	12W:SARAH	1951-10-22 00:00:00	TS	1010	tokyo	
1951301N16128	1951	30	WP	MM	14W:VERA	1951-10-28 06:00:00	TS	1006	tokyo	
1951321N09143	1951	31	WP	MM	15W:WANDA	1951-11-16 12:00:00	TS	1006	tokyo	
1951337N09150	1951	33	WP	MM	16W:AMY	1951-12-03 00:00:00	NR	1000	tokyo	
1951344N09143	1951	34	WP	MM	17W:BABS	1951-12-10 00:00:00	TS	1006	tokyo	
1952162N15118	1952	2	WP	MM	01W:CHARLOTT:CHARLOTTE	1952-06-09 12:00:00	TS	992	tokyo	
1952171N12130	1952	3	WP	MM	02W:DINAH	1952-06-19 00:00:00	TS	1006	tokyo	
1952180N05144	1952	5	WP	MM	03W:EMMA	1952-06-28 00:00:00	TS	1002	tokyo	
1952236N10135	1952	16	WP	MM	09W:LOIS	1952-08-22 12:00:00	TS	1008	tokyo	
1952241N08137	1952	17	WP	MM	MARY	1952-08-28 00:00:00	TS	1004	tokyo	
1952245N09139	1952	19	WP	MM	11W:NONA	1952-09-01 00:00:00	TS	1006	tokyo	
1952269N11157	1952	25	WP	MM	15W:POLLY	1952-09-24 12:00:00	TS	1004	tokyo	
1952278N22140	1952	26	WP	MM	16W:ROSE	1952-10-04 06:00:00	TS	1006	tokyo	
1952289N15128	1952	29	WP	MM	19W:VAE	1952-10-15 06:00:00	TS	1006	tokyo	
1952312N08150	1952	32	WP	MM	22W:BESS	1952-11-06 18:00:00	TS	1008	tokyo	

from *tabular…*

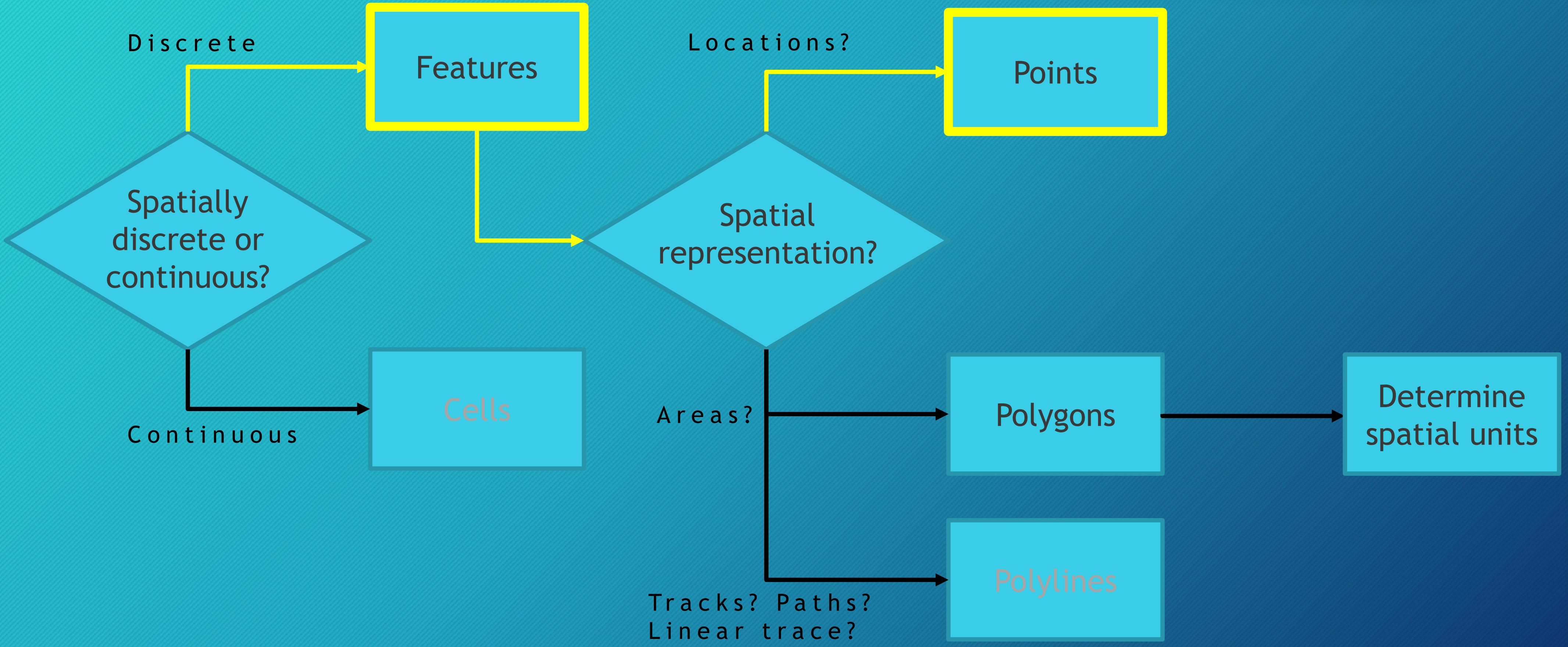


to *spatial.*

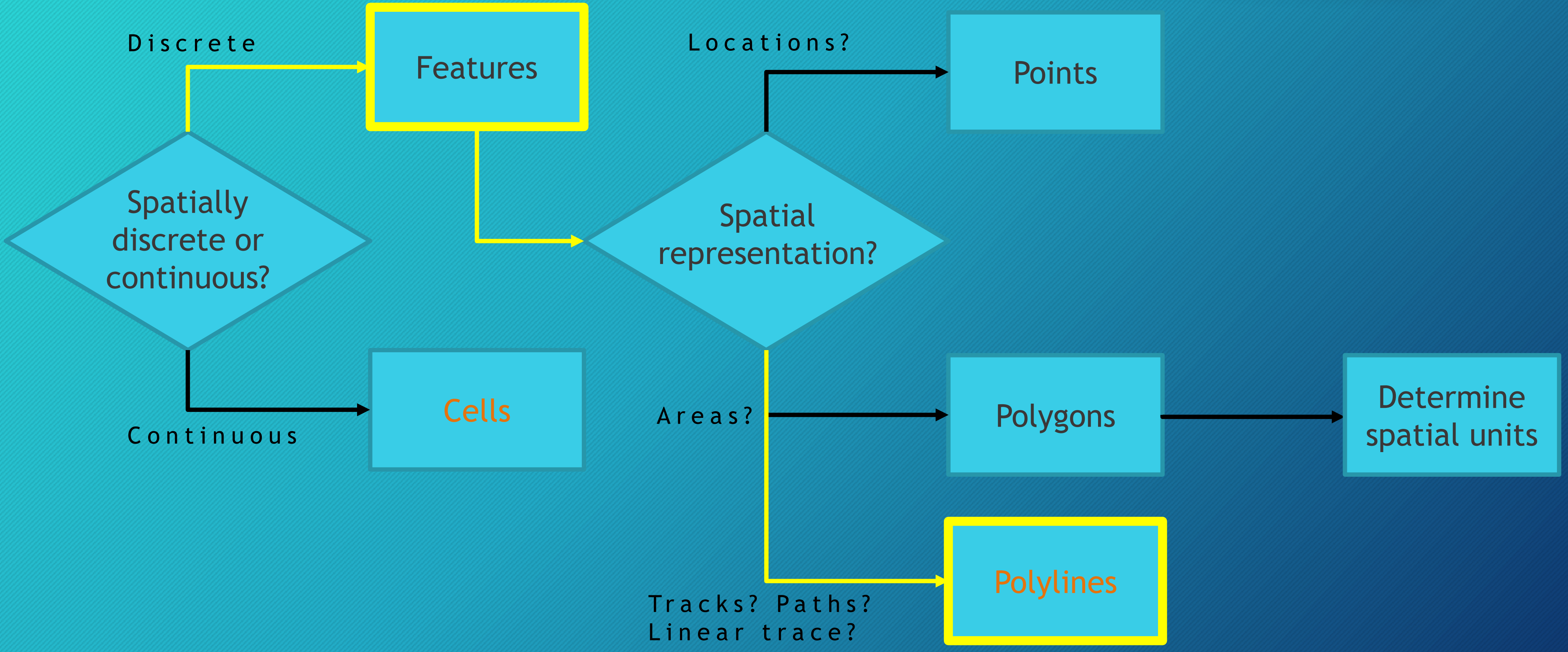
How to introduce space into data



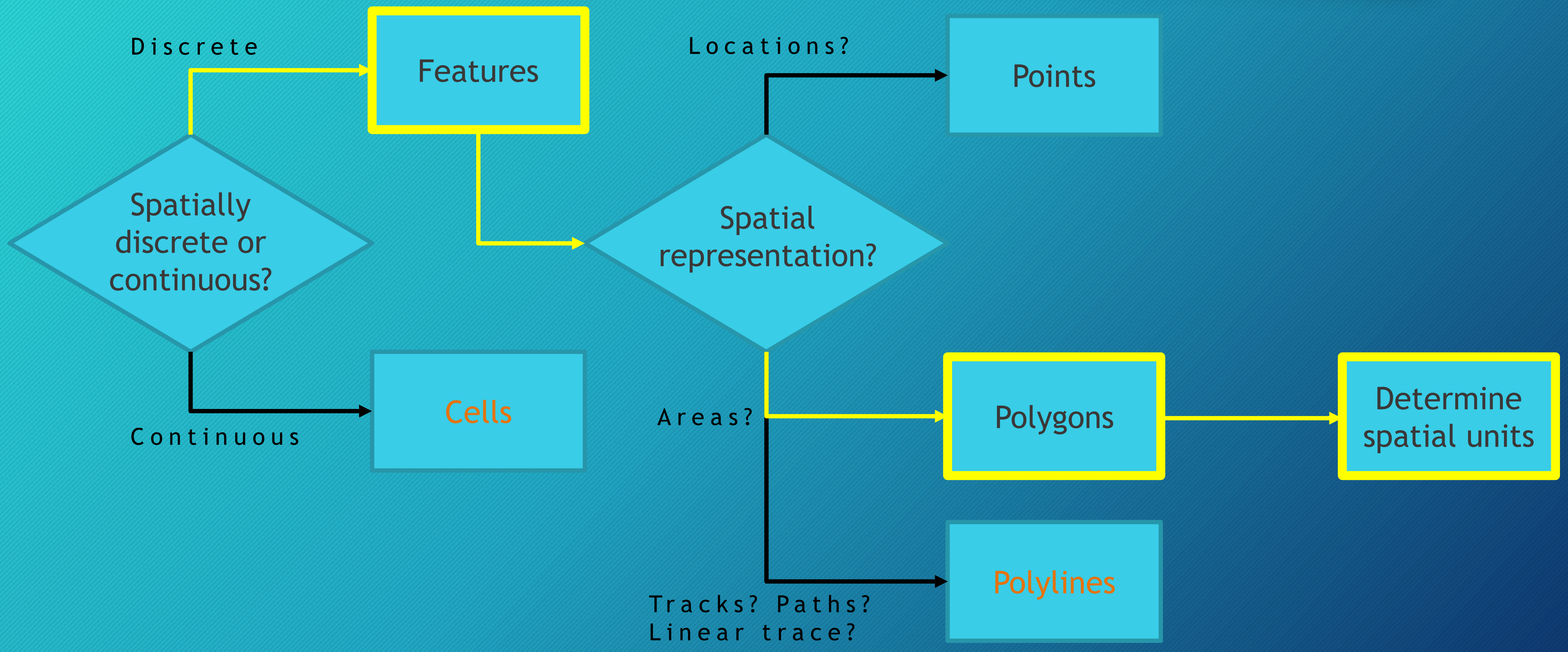
Example: Typhoon origins



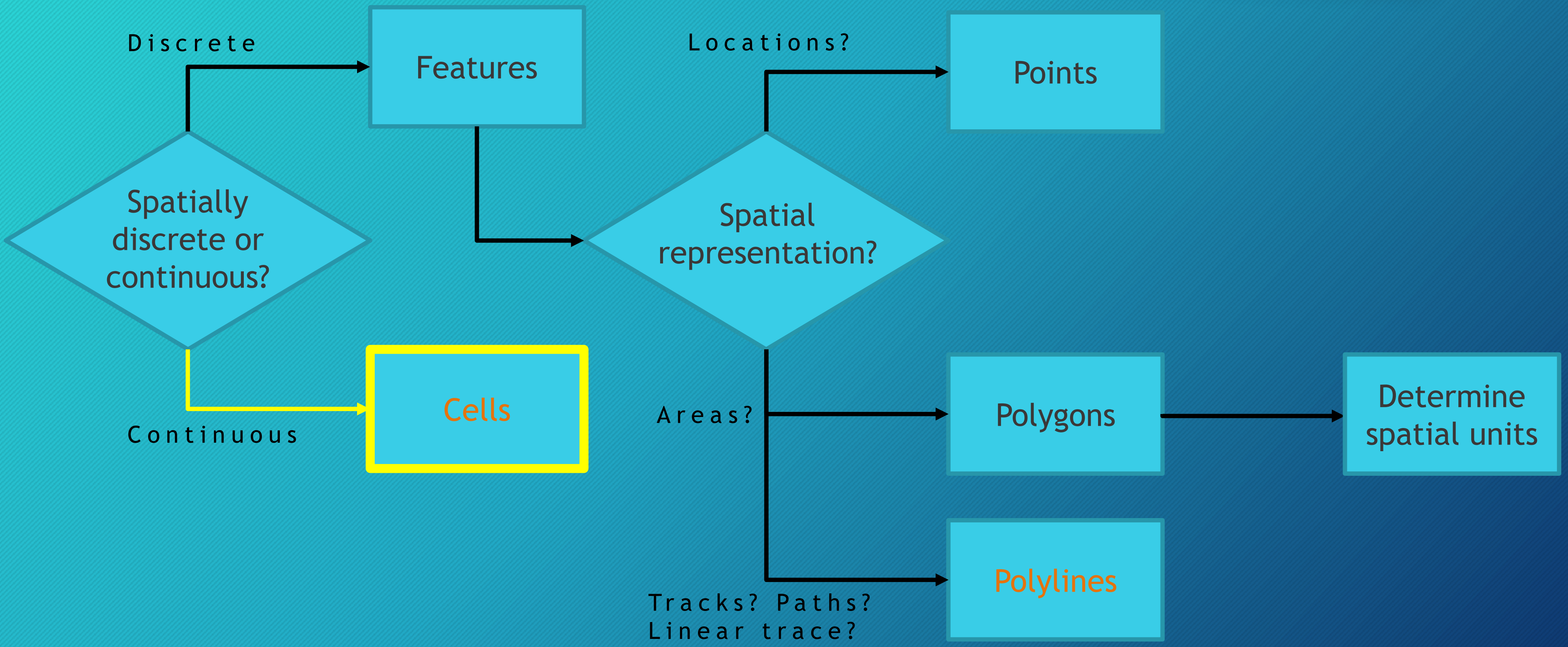
Example: Road inventory



Example: Election results



Example: Soil pH



Same data, many possibilities

Data	Sp. discrete / continuous?	Sp. representation
Crime locations	Discrete	Points
Crime statistics by province	Discrete	Polygons
Annual crime stats by police district	Discrete	Polygons
Criminal record per person	-	-
Crime stats by crime type	-	-

Turning location data to points

Serial_Num	Season	Num	Basin	Sub_basin	Name	ISO_time	Nature	wmo_pres	Center	wmo_pres_%
1951050N20139	1951	1	WP	MM	NOT NAMED	1951-02-19 06:00:00	TS	1010	tokyo	0.86
1951077N06158	1951	2	WP	MM	01W:GEORGIA	1951-03-18 06:00:00	TS	1002	tokyo	15.85
1951105N08152	1951	3	WP	MM	02W:HOPE	1951-04-15 00:00:00	TS	1002	tokyo	15.85
1951119N06145	1951	4	WP	MM	03W:IRIS	1951-04-28 12:00:00	TS	1008	tokyo	1.51
1951203N22144	1951	12	WP	MM	NOT NAMED	1951-07-22 00:00:00	TS	1008	tokyo	1.51
1951221N10136	1951	17	WP	MM	NOT NAMED	1951-08-09 00:00:00	TS	1006	tokyo	4.44
1951223N11148	1951	18	WP	MM	07W:MARGE	1951-08-10 12:00:00	TS	987	tokyo	55.59
1951239N11144	1951	20	WP	MM	08W:NORA	1951-08-27 00:00:00	TS	1008	tokyo	1.51
1951255N08148	1951	23	WP	MM	09W:ORA	1951-09-11 18:00:00	TS	1006	tokyo	4.44
1951263N08137	1951	24	WP	MM	10W:PAT	1951-09-20 00:00:00	TS	1008	tokyo	1.51
1951295N21152	1951	28	WP	MM	12W:SARAH	1951-10-22 00:00:00	TS	1010	tokyo	0.86
1951301N16128	1951	30	WP	MM	14W:VERA	1951-10-28 06:00:00	TS	1006	tokyo	4.44
1951321N09143	1951	31	WP	MM	15W:WANDA	1951-11-16 12:00:00	TS	1006	tokyo	4.44
1951337N09150	1951	33	WP	MM	16W:AMY	1951-12-03 00:00:00	NR	1000	tokyo	21.05
1951344N09143	1951	34	WP	MM	17W:BABS	1951-12-10 00:00:00	TS	1006	tokyo	4.44
1952162N15118	1952	2	WP	MM	01W:CHARLOTT:CHARLOTTE	1952-06-09 12:00:00	TS	992	tokyo	43.9
1952171N12130	1952	3	WP	MM	02W:DINAH	1952-06-19 00:00:00	TS	1006	tokyo	4.44
1952180N05144	1952	5	WP	MM	03W:EMMA	1952-06-28 00:00:00	TS	1002	tokyo	15.85
1952236N10135	1952	16	WP	MM	09W:LOIS	1952-08-22 12:00:00	TS	1008	tokyo	1.51
1952241N08137	1952	17	WP	MM	MARY	1952-08-28 00:00:00	TS	1004	tokyo	8.41
1952245N09139	1952	19	WP	MM	11W:NONA	1952-09-01 00:00:00	TS	1006	tokyo	4.44
1952269N11157	1952	25	WP	MM	15W:POLLY	1952-09-24 12:00:00	TS	1004	tokyo	8.41
1952278N22140	1952	26	WP	MM	16W:ROSE	1952-10-04 06:00:00	TS	1006	tokyo	4.44
1952289N15128	1952	29	WP	MM	19W:VAE	1952-10-15 06:00:00	TS	1006	tokyo	4.44
1952312N08150	1952	32	WP	MM	22W:BESS	1952-11-06 18:00:00	TS	1008	tokyo	1.51

Turning location data to points

Serial_Num	Season	Num	Basin	Sub_basin	Name	ISO_time	Nature	wmo_pres	Center	wmo_pres_%	LONG	LAT
1951050N20139	1951	1	WP	MM	NOT NAMED	1951-02-19 06:00:00	TS	1010	tokyo	0.86	138.5	20
1951077N06158	1951	2	WP	MM	01W:GEORGIA	1951-03-18 06:00:00	TS	1002	tokyo	15.85	158.3	5.7
1951105N08152	1951	3	WP	MM	02W:HOPE	1951-04-15 00:00:00	TS	1002	tokyo	15.85	151.5	8
1951119N06145	1951	4	WP	MM	03W:IRIS	1951-04-28 12:00:00	TS	1008	tokyo	1.51	144.9	6.3
1951203N22144	1951	12	WP	MM	NOT NAMED	1951-07-22 00:00:00	TS	1008	tokyo	1.51	166.5	4
1951221N10136	1951	17	WP	MM	NOT NAMED	1951-08-09 00:00:00	TS	1006	tokyo	4.44	110.65	17.95
1951223N11148	1951	18	WP	MM	07W:MARGE	1951-08-10 12:00:00	TS	987	tokyo	55.59	116.5	17.5
1951239N11144	1951	20	WP	MM	08W:NORA	1951-08-27 00:00:00	TS	1008	tokyo	1.51	136.3	11.5
1951255N08148	1951	23	WP	MM	09W:ORA	1951-09-11 18:00:00	TS	1006	tokyo	4.44	121.5	21
1951263N08137	1951	24	WP	MM	10W:PAT	1951-09-20 00:00:00	TS	1008	tokyo	1.51	103.8	12
1951295N21152	1951	28	WP	MM	12W:SARAH	1951-10-22 00:00:00	TS	1010	tokyo	0.86	136	8.1
1951301N16128	1951	30	WP	MM	14W:VERA	1951-10-28 06:00:00	TS	1006	tokyo	4.44	144.3	21.5
1951321N09143	1951	31	WP	MM	15W:WANDA	1951-11-16 12:00:00	TS	1006	tokyo	4.44	153.1	25.8
1951337N09150	1951	33	WP	MM	16W:AMY	1951-12-03 00:00:00	NR	1000	tokyo	21.05	144.5	10.5
1951344N09143	1951	34	WP	MM	17W:BABS	1951-12-10 00:00:00	TS	1006	tokyo	4.44	135	10
1952162N15118	1952	2	WP	MM	01W:CHARLOTT:CHARLOTTE	1952-06-09 12:00:00	TS	992	tokyo	43.9	113.8	13.9
1952171N12130	1952	3	WP	MM	02W:DINAH	1952-06-19 00:00:00	TS	1006	tokyo	4.44	136	10
1952180N05144	1952	5	WP	MM	03W:EMMA	1952-06-28 00:00:00	TS	1002	tokyo	15.85	147.62	10.92
1952236N10135	1952	16	WP	MM	09W:LOIS	1952-08-22 12:00:00	TS	1008	tokyo	1.51	151	17.7
1952241N08137	1952	17	WP	MM	MARY	1952-08-28 00:00:00	TS	1004	tokyo	8.41	144.45	11.4
1952245N09139	1952	19	WP	MM	11W:NONA	1952-09-01 00:00:00	TS	1006	tokyo	4.44	119.7	13.6
1952269N11157	1952	25	WP	MM	15W:POLLY	1952-09-24 12:00:00	TS	1004	tokyo	8.41	148	21
1952278N22140	1952	26	WP	MM	16W:ROSE	1952-10-04 06:00:00	TS	1006	tokyo	4.44	147.55	8
1952289N15128	1952	29	WP	MM	19W:VAE	1952-10-15 06:00:00	TS	1006	tokyo	4.44	137	8
1952312N08150	1952	32	WP	MM	22W:BESS	1952-11-06 18:00:00	TS	1008	tokyo	1.51	113.3	10.9

Turning location data to points

Serial_Num	Season	Num	Basin	Sub_basin	Name	ISO_time	Nature	wmo_pres	Center	wmo_pres_%
1951050N20139	1951	1	WP	MM	NOT NAMED	1951-02-19 06:00:00	TS	1010	tokyo	0.86
1951077N06158	1951	2	WP	MM	01W:GEORGIA	1951-03-18 06:00:00	TS	1002	tokyo	15.85
1951105N08152	1951	3	WP	MM	02W:HOPE	1951-04-15 00:00:00	TS	1002	tokyo	15.85
1951119N06145	1951	4	WP	MM	03W:IRIS	1951-04-28 12:00:00	TS	1008	tokyo	1.51
1951203N22144	1951	12	WP	MM	NOT NAMED	1951-07-22 00:00:00	TS	1008	tokyo	1.51
1951221N10136	1951	17	WP	MM	NOT NAMED	1951-08-09 00:00:00	TS	1006	tokyo	4.44
1951223N11148	1951	18	WP	MM	07W:MARGE	1951-08-10 12:00:00	TS	987	tokyo	55.59
1951239N11144	1951	20	WP	MM	08W:NORA	1951-08-27 00:00:00	TS	1008	tokyo	1.51
1951255N08148	1951	23	WP	MM	09W:ORA	1951-09-11 18:00:00	TS	1006	tokyo	4.44
1951263N08137	1951	24	WP	MM	10W:PAT	1951-09-20 00:00:00	TS	1008	tokyo	1.51
1951295N21152	1951	28	WP	MM	12W:SARAH	1951-10-22 00:00:00	TS	1010	tokyo	0.86
1951301N16128	1951	30	WP	MM	14W:VERA	1951-10-28 06:00:00	TS	1006	tokyo	4.44
1951321N09143	1951	31	WP	MM	15W:WANDA	1951-11-16 12:00:00	TS	1006	tokyo	4.44
1951337N09150	1951	33	WP	MM	16W:AMY	1951-12-03 00:00:00	NR	1000	tokyo	21.05
1951344N09143	1951	34	WP	MM	17W:BABS	1951-12-10 00:00:00	TS	1006	tokyo	4.44
1952162N15118	1952	2	WP	MM	01W:CHARLOTT:CHARLOTTE	1952-06-09 12:00:00	TS	992	tokyo	43.9
1952171N12130	1952	3	WP	MM	02W:DINAH	1952-06-19 00:00:00	TS	1006	tokyo	4.44
1952180N05144	1952	5	WP	MM	03W:EMMA	1952-06-28 00:00:00	TS	1002	tokyo	15.85
1952236N10135	1952	16	WP	MM	09W:LOIS	1952-08-22 12:00:00	TS	1008	tokyo	1.51
1952241N08137	1952	17	WP	MM	MARY	1952-08-28 00:00:00	TS	1004	tokyo	8.41
1952245N09139	1952	19	WP	MM	11W:NONA	1952-09-01 00:00:00	TS	1006	tokyo	4.44
1952269N11157	1952	25	WP	MM	15W:POLLY	1952-09-24 12:00:00	TS	1004	tokyo	8.41
1952278N22140	1952	26	WP	MM	16W:ROSE	1952-10-04 06:00:00	TS	1006	tokyo	4.44
1952289N15128	1952	29	WP	MM	19W:VAE	1952-10-15 06:00:00	TS	1006	tokyo	4.44
1952312N08150	1952	32	WP	MM	22W:BESS	1952-11-06 18:00:00	TS	1008	tokyo	1.51

?

Turning location data to points

Serial_Num	Season	Num	Basin	Sub_basin	Name	ISO_time	Nature	wmo_pres	Center	wmo_pres_%
1951050N20139	1951	1	WP	MM	NOT NAMED	1951-02-19 06:00:00	TS	1010	tokyo	0.86
1951077N06158	1951	2	WP	MM	01W:GEORGIA	1951-03-18 06:00:00	TS	1002	tokyo	15.85
1951105N08152	1951	3	WP	MM	02W:HOPE	1951-04-15 00:00:00	TS	1002	tokyo	15.85
1951119N06145	1951	4	WP	MM	03W:IRIS	1951-04-28 12:00:00	TS	1008	tokyo	1.51
1951203N22144	1951	12	WP	MM	NOT NAMED	1951-07-22 00:00:00	TS	1008	tokyo	1.51
1951221N10136	1951	17	WP	MM	NOT NAMED	1951-08-09 00:00:00	TS	1006	tokyo	4.44
1951223N11148	1951	18	WP	MM	07W:MARGE	1951-08-10 12:00:00	TS	987	tokyo	55.59
1951239N11144	1951	20	WP	MM	08W:NORA	1951-08-27 00:00:00	TS	1008	tokyo	1.51
1951255N08148	1951	23	WP	MM	09W:ORA	1951-09-11 18:00:00	TS	1006	tokyo	4.44
1951263N08137	1951	24	WP	MM	10W:PAT	1951-09-20 00:00:00	TS	1008	tokyo	1.51
1951295N21152	1951	28	WP	MM	12W:SARAH	1951-10-22 00:00:00	TS	1010	tokyo	0.86
1951301N16128	1951	30	WP	MM	14W:VERA	1951-10-28 06:00:00	TS	1006	tokyo	4.44
1951321N09143	1951	31	WP	MM	15W:WANDA	1951-11-16 12:00:00	TS	1006	tokyo	4.44
1951337N09150	1951	33	WP	MM	16W:AMY	1951-12-03 00:00:00	NR	1000	tokyo	21.05
1951344N09143	1951	34	WP	MM	17W:BABS	1951-12-10 00:00:00	TS	1006	tokyo	4.44
1952162N15118	1952	2	WP	MM	01W:CHARLOTT:CHARLOTTE	1952-06-09 12:00:00	TS	992	tokyo	43.9
1952171N12130	1952	3	WP	MM	02W:DINAH	1952-06-19 00:00:00	TS	1006	tokyo	4.44
1952180N05144	1952	5	WP	MM	03W:EMMA	1952-06-28 00:00:00	TS	1002	tokyo	15.85
1952236N10135	1952	16	WP	MM	09W:LOIS	1952-08-22 12:00:00	TS	1008	tokyo	1.51
1952241N08137	1952	17	WP	MM	MARY	1952-08-28 00:00:00	TS	1004	tokyo	8.41
1952245N09139	1952	19	WP	MM	11W:NONA	1952-09-01 00:00:00	TS	1006	tokyo	4.44
1952269N11157	1952	25	WP	MM	15W:POLLY	1952-09-24 12:00:00	TS	1004	tokyo	8.41
1952278N22140	1952	26	WP	MM	16W:ROSE	1952-10-04 06:00:00	TS	1006	tokyo	4.44
1952289N15128	1952	29	WP	MM	19W:VAE	1952-10-15 06:00:00	TS	1006	tokyo	4.44
1952312N08150	1952	32	WP	MM	22W:BESS	1952-11-06 18:00:00	TS	1008	tokyo	1.51

Plot
(using GIS
software)

“Spatializing” your data

Exercise 1: Turning location data into points

Exercise 1: Turning location data into points

- Step 1: In your web browser, go to the following link:
  https://goo.gl/forms/KcA4hBbEXk0nfutb2

- Step 2: Fill out the form

Exercise 1: Turning location data into points

- Step 3: In your web browser, go to the following link:

  https://arcg.is/19PKyy

- Step 4: Fill out the form

Turning area data to polygons

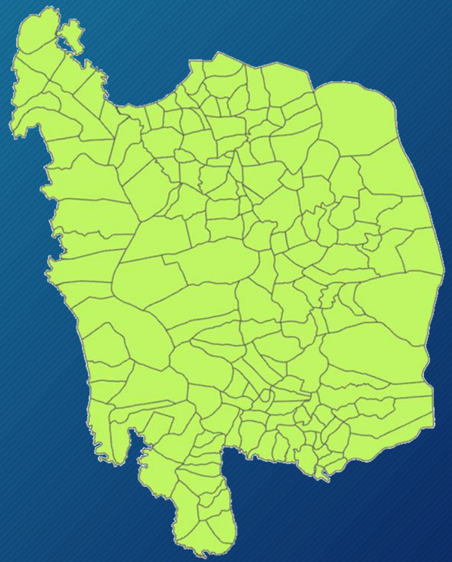
MUNICIPALI	PROVINCE	REGION	PSGC	ZIPCODE	MUN_INCOME	TOTPOP2007	POVERTY	RUR_URB	OLD_NAME
SCIENCE CITY OF MU...	NUEVA ECIJA	III	034917	3119	5th Class	71669	17.91	Partially Urban	MUNOZ
ANAO	TARLAC	III	036901	2310	5th Class	10806	12.16	Partially Urban	ANAO
SAN CLEMENTE	TARLAC	III	036913	2305	5th Class	12458	17.77	Partially Urban	SAN CLEMENTE
CAMILING	TARLAC	III	036903	2306	1st Class	79941	13.88	Partially Urban	CAMILING
NAMPICUAN	NUEVA ECIJA	III	034918	3116	5th Class	11786	31.22	Partially Urban	NAMPICUAN
GUIMBA	NUEVA ECIJA	III	034911	3115	1st Class	96116	28.27	Partially Urban	GUIMBA
QUEZON	NUEVA ECIJA	III	034922	3113	4th Class	33988	32.08	Partially Urban	QUEZON
MAYANTOC	TARLAC	III	036908	2304	3rd Class	27274	20.15	Partially Urban	MAYANTOC
VICTORIA	TARLAC	III	036917	2313	3rd Class	57085	17.77	Partially Urban	VICTORIA
GENERAL MAMERTO...	NUEVA ECIJA	III	034909	3125	4th Class	33354	24.11	Partially Urban	GEN. MAMERTO NAT...
SAN FABIAN	PANGASINAN	I	015533	2433	2nd Class	74005	28.86	Partially Urban	SAN FABIAN
SAN NICOLAS	PANGASINAN	I	015536	2447	3rd Class	33419	31.93	Partially Urban	SAN NICOLAS
POZZORUBIO	PANGASINAN	I	015530	2435	2nd Class	63689	24.7	Partially Urban	POZORRUBIO
SAN MANUEL	PANGASINAN	I	015535	2438	2nd Class	46769	27.13	Partially Urban	SAN MANUEL
MABINI	PANGASINAN	I	015523	2409	4th Class	23338	36.6	Partially Urban	MABINI
CARRANGLAN	NUEVA ECIJA	III	034905	3123	2nd Class	33233	33.81	Partially Urban	CARRANGLAN
SAN JACINTO	PANGASINAN	I	015534	2431	3rd Class	35591	27.62	Partially Urban	SAN JACINTO
BINALONAN	PANGASINAN	I	015512	2436	2nd Class	52722	16.75	Urban	BINALONAN
MANGALDAN	PANGASINAN	I	015526	2432	1st Class	90391	21.64	Urban	MANGALDAN
NATIVIDAD	PANGASINAN	I	015529	2446	4th Class	21560	28.18	Partially Urban	NATIVIDAD
GENERAL TINIO (PAP...	NUEVA ECIJA	III	034910	3104	2nd Class	38640	12.48	Partially Urban	GEN. TINIO (PAPAYA)
JAEN	NUEVA ECIJA	III	034912	3109	3rd Class	63474	20.36	Partially Urban	JAEN
CONCEPCION	TARLAC	III	036905	2316	1st Class	135213	19.77	Partially Urban	CONCEPCION
SAN ANTONIO	NUEVA ECIJA	III	034924	3108	2nd Class	67446	23.73	Partially Urban	SAN ANTONIO
CAPAS	TARLAC	III	036904	2315	1st Class	122084	21.71	Partially Urban	CAPAS

Turning area data to polygons

1

MUNICIPALI	PROVINCE	REGION	PSGC	ZIPCODE	MUN_INCOME	TOTPOP2007	POVERTY	RUR_URB	OLD_NAME
SCIENCE CITY OF MU...	NUEVA ECIJA	III	034917	3119	5th Class	71669	17.91	Partially Urban	MUNOZ
ANAO	TARLAC	III	036901	2310	5th Class	10806	12.16	Partially Urban	ANAO
SAN CLEMENTE	TARLAC	III	036913	2305	5th Class	12458	17.77	Partially Urban	SAN CLEMENTE
CAMILING	TARLAC	III	036903	2306	1st Class	79941	13.88	Partially Urban	CAMILING
NAMPICUAN	NUEVA ECIJA	III	034918	3116	5th Class	11786	31.22	Partially Urban	NAMPICUAN
GUIMBA	NUEVA ECIJA	III	034911	3115	1st Class	96116	28.27	Partially Urban	GUIMBA
QUEZON	NUEVA ECIJA	III	034922	3113	4th Class	33988	32.08	Partially Urban	QUEZON
MAYANTOC	TARLAC	III	36908	2304	3rd Class	27274	20.15	Partially Urban	MAYANTOC
VICTORIA	TARLAC	III	6917	2313	3rd Class	57085	17.77	Partially Urban	VICTORIA
GENERAL MAMERTO...	NUEVA ECIJA	III	034909	3125	4th Class	33354	24.11	Partially Urban	GEN. MAMERTO NAT...
SAN FABIAN	PANGASINAN	I	015533	2433	2nd Class	74005	28.86	Partially Urban	SAN FABIAN
SAN NICOLAS	PANGASINAN	I	015536	2447	3rd Class	33419	31.93	Partially Urban	SAN NICOLAS
POZZORUBIO	PANGASINAN	I	015530	2435	2nd Class	63689	24.7	Partially Urban	POZORRUBIO
SAN MANUEL	PANGASINAN	I	015535	2438	2nd Class	46769	27.13	Partially Urban	SAN MANUEL
MABINI	PANGASINAN	I	015523	2409	4th Class	23338	36.6	Partially Urban	MABINI
CARRANGLAN	NUEVA ECIJA	III	034905	3123	2nd Class	33233	33.81	Partially Urban	CARRANGLAN
SAN JACINTO	PANGASINAN	I	015534	2431	3rd Class	35591	27.62	Partially Urban	SAN JACINTO
BINALONAN	PANGASINAN	I	015512	2436	2nd Class	52722	16.75	Urban	BINALONAN
MANGALDAN	PANGASINAN	I	015526	2432	1st Class	90391	21.64	Urban	MANGALDAN
NATIVIDAD	PANGASINAN	I	015529	2446	4th Class	21560	28.18	Partially Urban	NATIVIDAD
GENERAL TINIO (PAP...	NUEVA ECIJA	III	034910	3104	2nd Class	38640	12.48	Partially Urban	GEN. TINIO (PAPAYA)
JAEN	NUEVA ECIJA	III	034912	3109	3rd Class	63474	20.36	Partially Urban	JAEN
CONCEPCION	TARLAC	III	036905	2316	1st Class	135213	19.77	Partially Urban	CONCEPCION
SAN ANTONIO	NUEVA ECIJA	III	034924	3108	2nd Class	67446	23.73	Partially Urban	SAN ANTONIO
CAPAS	TARLAC	III	036904	2315	1st Class	122084	21.71	Partially Urban	CAPAS

2 boundary data (with equivalent identifiers)



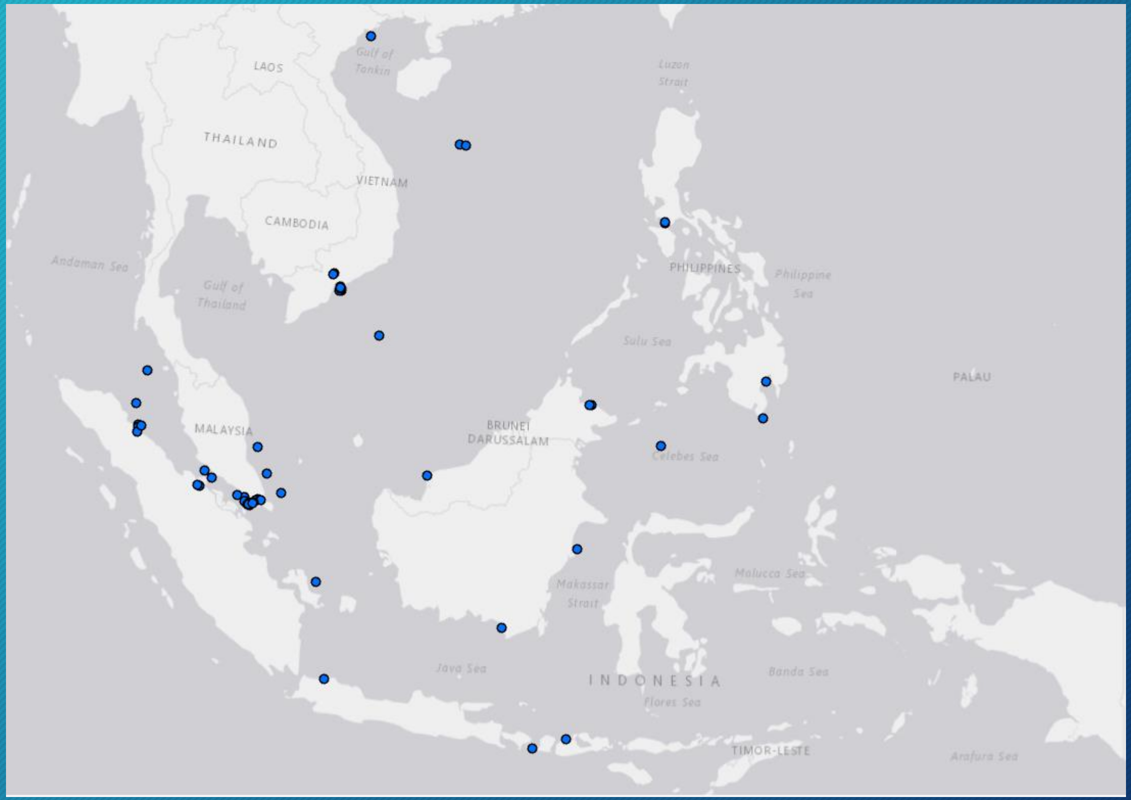
“Spatializing” your data

Demo 1: Turning area data into polygons

Measuring geographic distributions

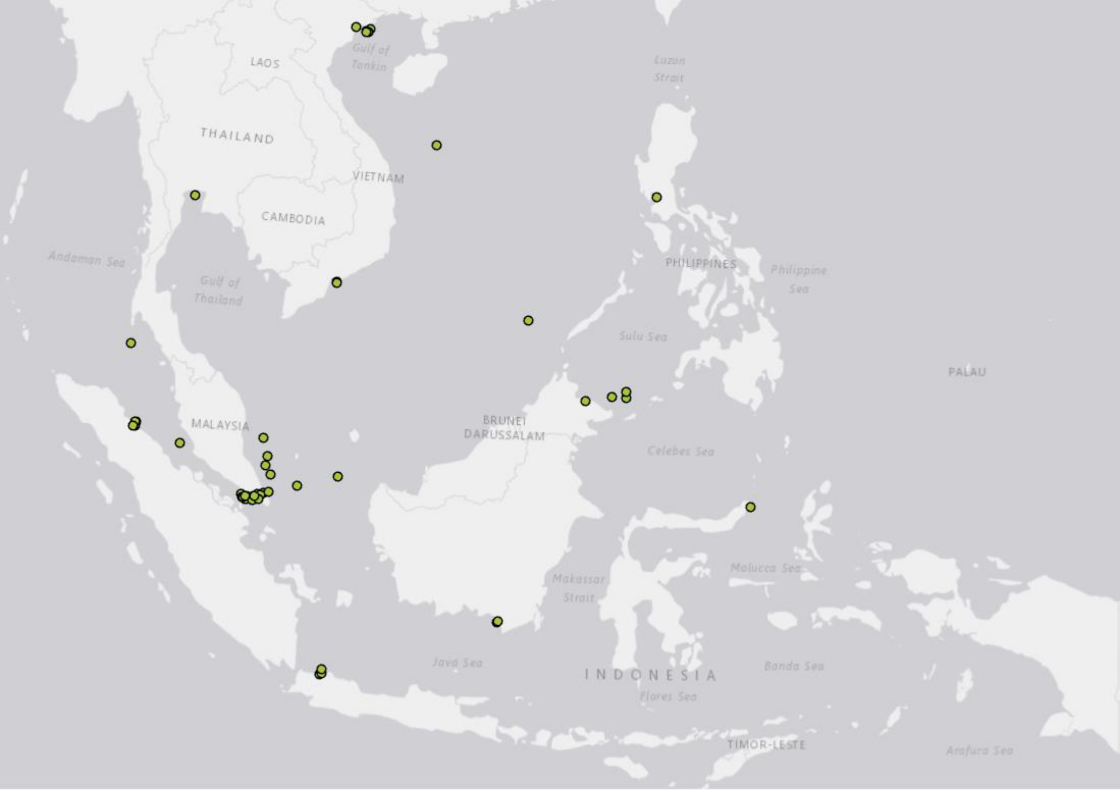
Why measure geographic distribution?

- Comparing the distribution of different features
  - Example: Piracy during monsoon and non-monsoon months in 2015



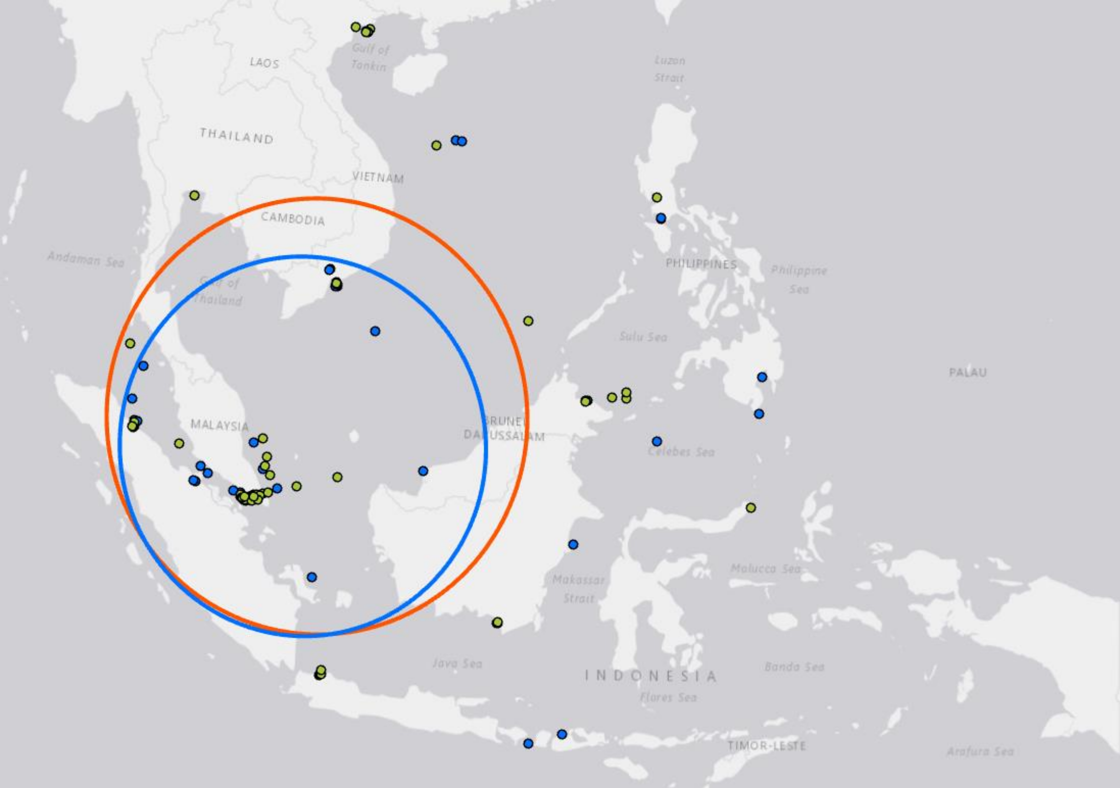
Why measure geographic distribution?

- Comparing the distribution of different features
  - Example: Piracy during monsoon and non-monsoon months in 2015



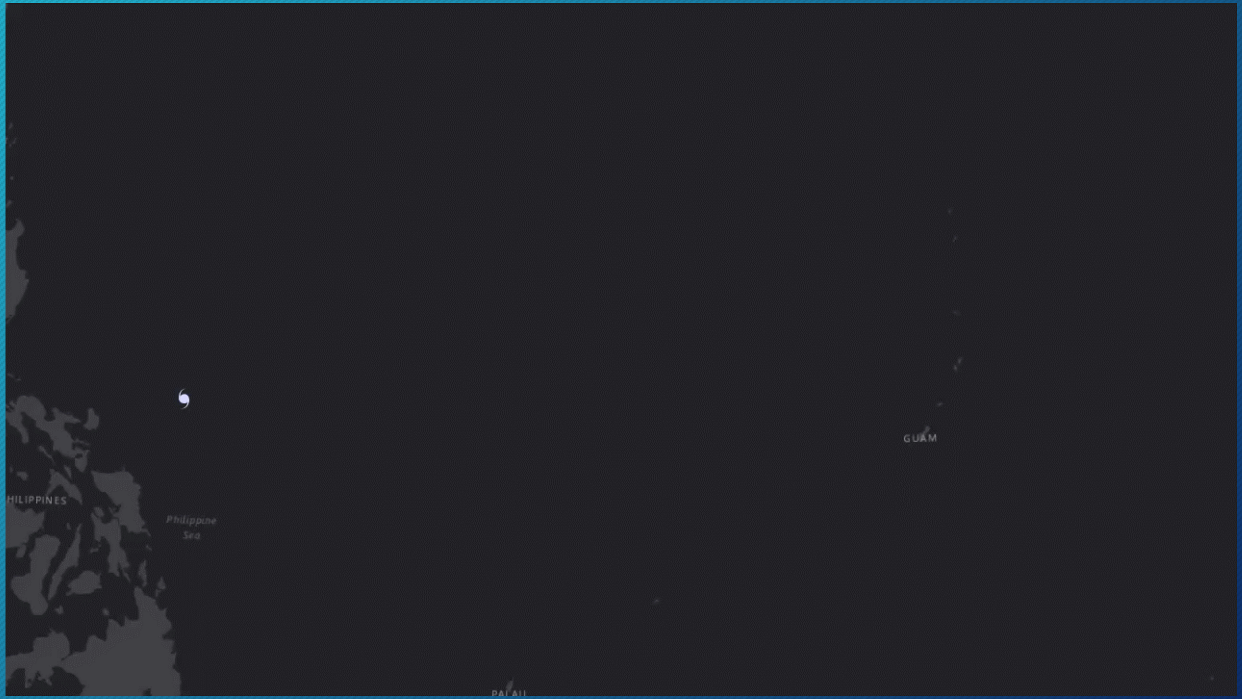
Why measure geographic distribution?

- Comparing the distribution of different features
  - Example: Piracy during monsoon and non-monsoon months in 2015



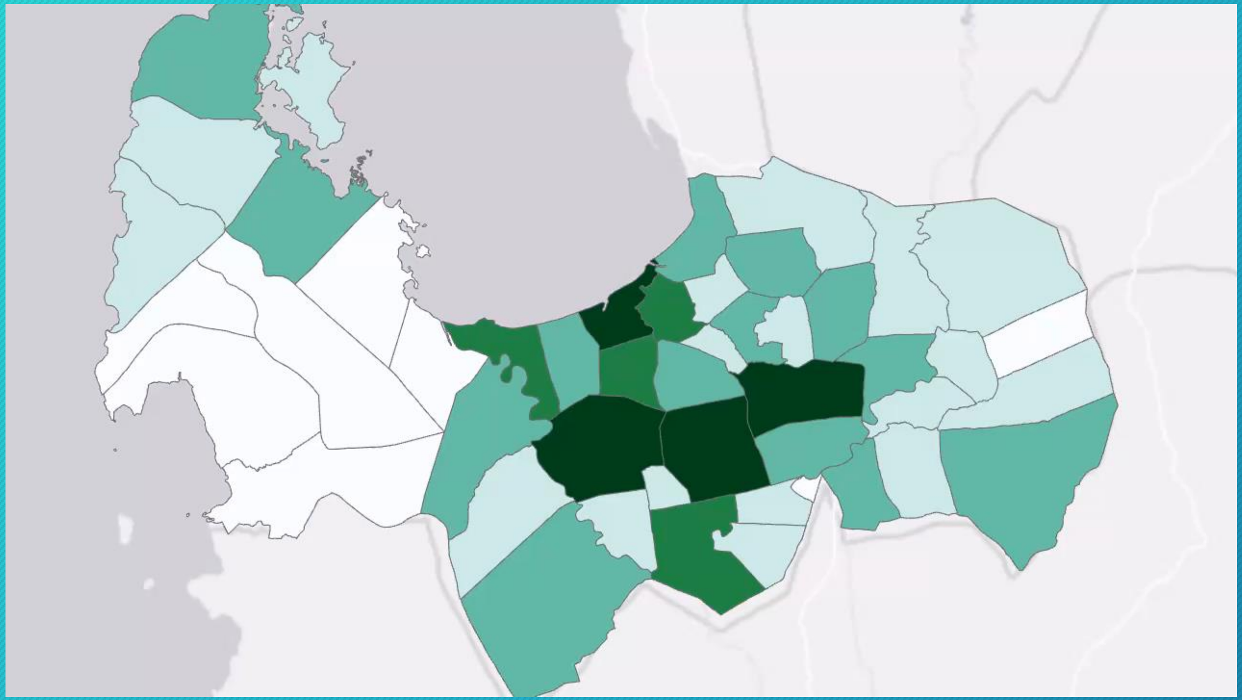
Why measure geographic distribution?

- Tracking change in distribution of features
  - Example: Typhoon origins in the W. Pacific region from 1883 to 2014



Why measure geographic distribution?

- Monitoring change in distribution of values across space
  - Example: Distribution in Pangasinan's population from 1990 to 2015



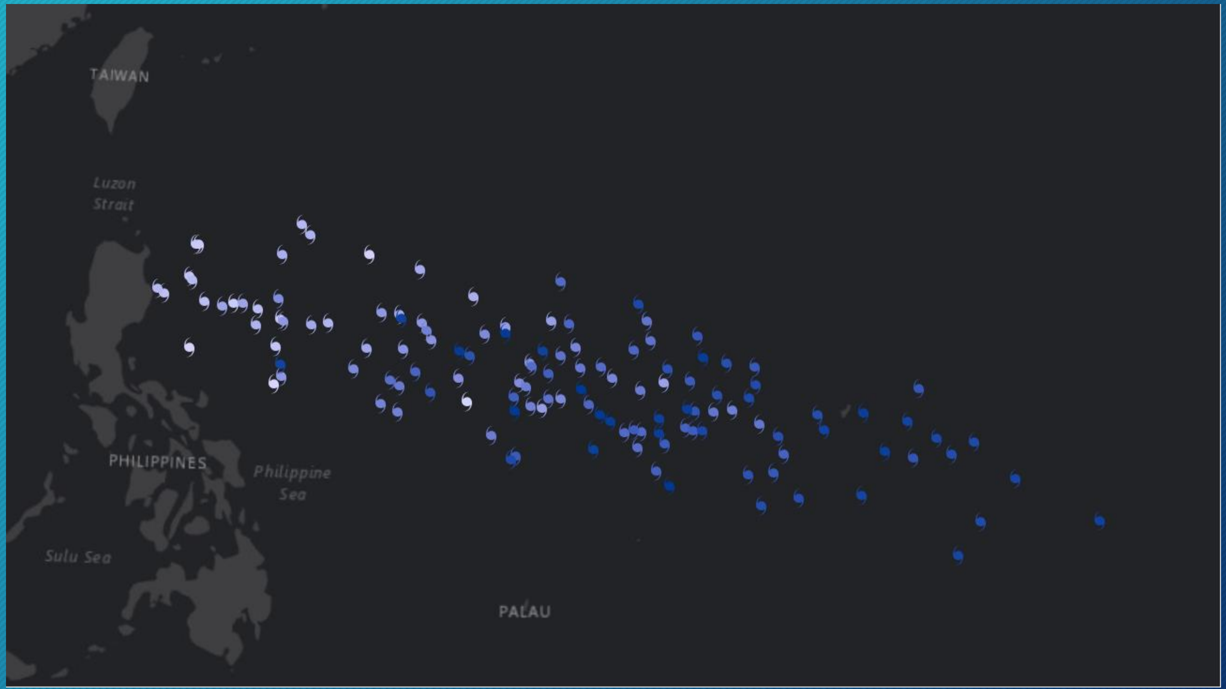
Finding the center

- Mean center
  - Mean of X and Y coordinates, respectively

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

  - Can be weighted by a quantitative field

$$\bar{X}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad , \quad \bar{Y}_w = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

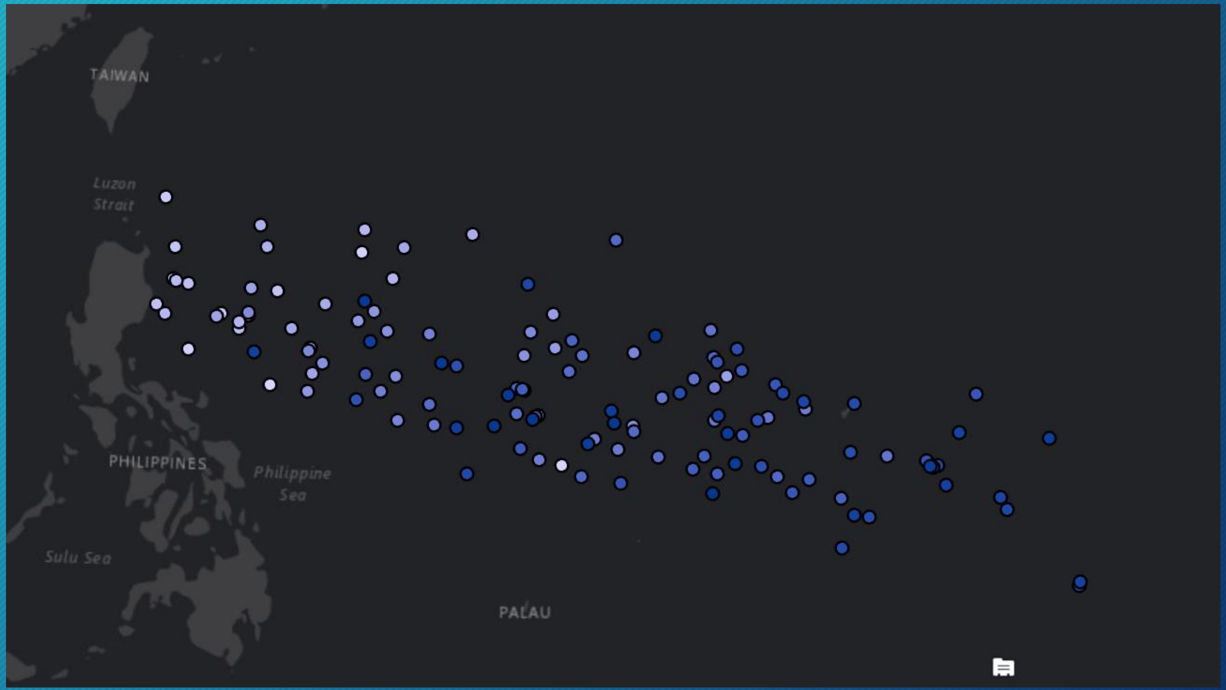


Finding the center

- Median center
  - Location (X,Y) having shortest distance to all features

$$d_i^t = \sqrt{(X_i - X^t)^2 + (Y_i - Y^t)^2 + (Z_i - Z^t)^2}$$

  - May also be weighted by a quantitative field



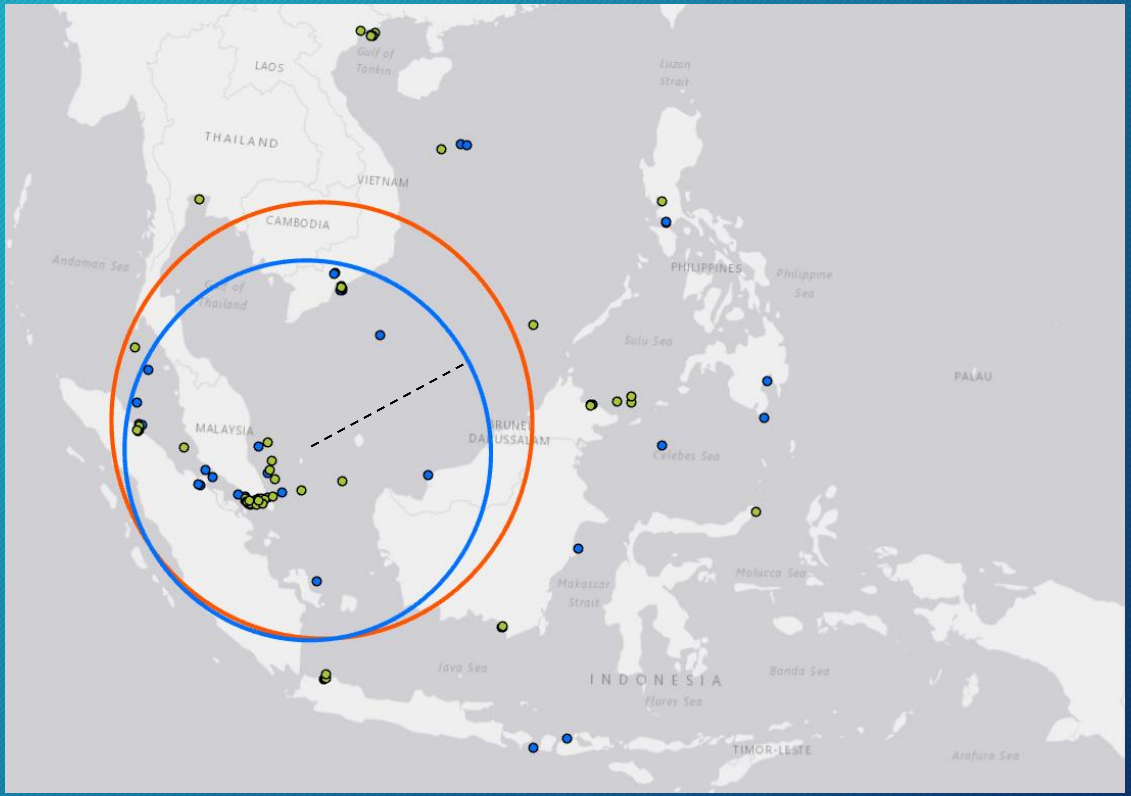
Measuring compactness

- Standard distance deviation
  - Extent of how features deviate from the mean center

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}}$$

  - Can be weighted by a quantitative field

$$SD_w = \sqrt{\frac{\sum_{i=1}^{n} w_i(x_i - \bar{X}_w)^2}{\sum_{i=1}^{n} w_i} + \frac{\sum_{i=1}^{n} w_i(y_i - \bar{Y}_w)^2}{\sum_{i=1}^{n} w_i}}$$



Measuring orientation

- Standard deviational ellipse
  - Alternative to standard distance when features are *anistotropic*
  - Can be weighted by a quantitative field

a

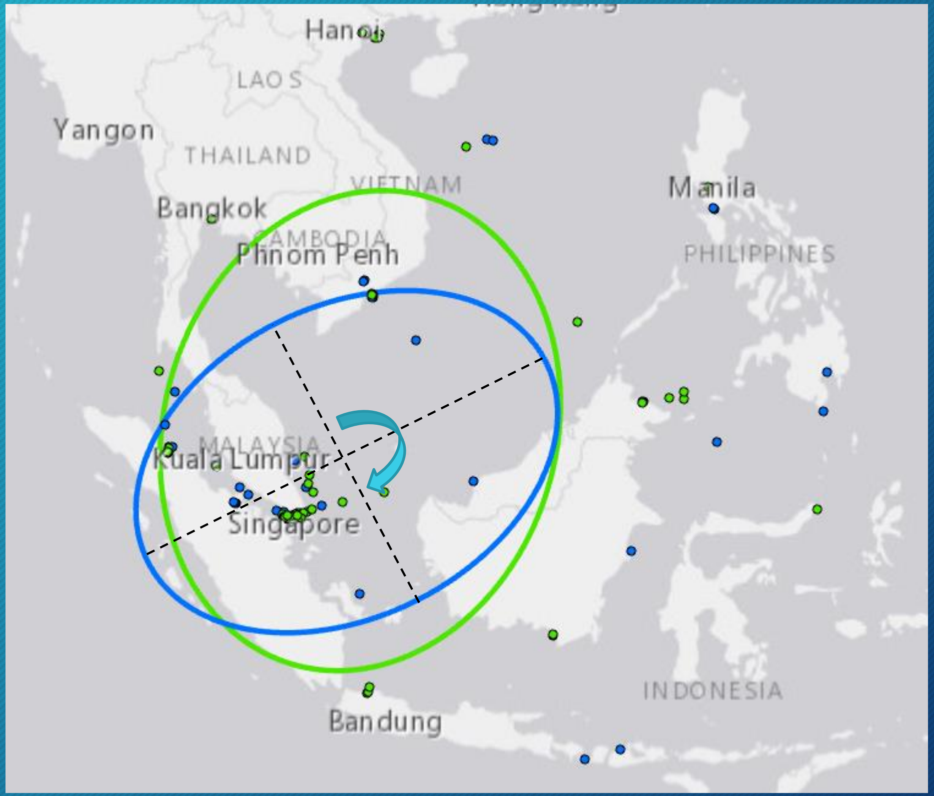
$$SDE_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}}$$

b

$$SDE_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}}$$

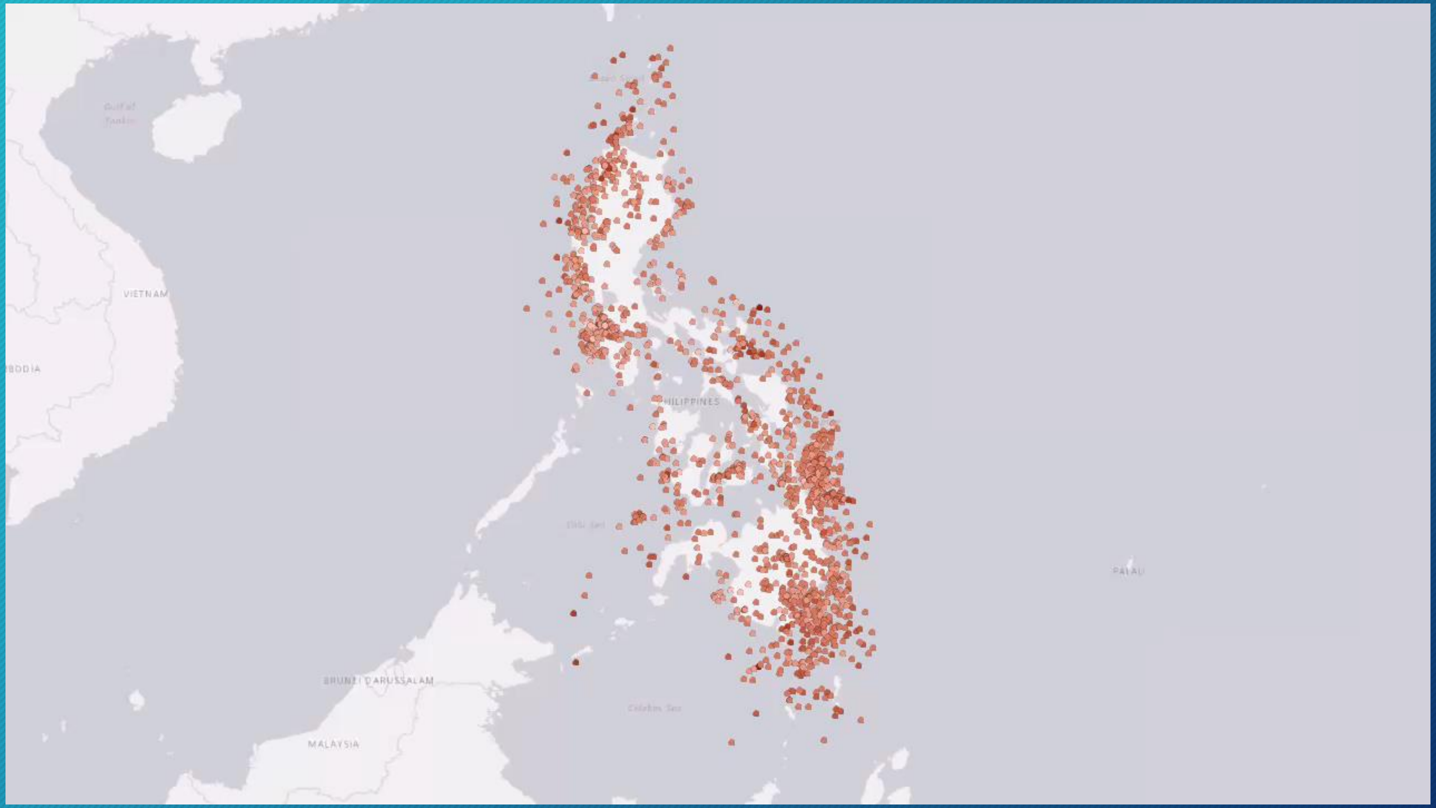
c

$$\begin{aligned}
\tan\ \theta &= \frac{A+B}{C} \\
A &= \left(\sum_{i=1}^{n}\tilde{x}_i^2 - \sum_{i=1}^{n}\tilde{y}_i^2\right) \\
B &= \sqrt{\left(\sum_{i=1}^{n}\tilde{x}_i^2 - \sum_{i=1}^{n}\tilde{y}_i^2\right)^2 + 4\left(\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i\right)^2} \\
C &= 2\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i
\end{aligned}$$



Some tips when measuring distributions

- Use weights when dealing with stationary features
- When there are spatial outliers, use median centers instead of mean center
- May also be used when dealing with 3D features



Measuring geographic distributions

Challenge 1: Calculating mean center, and standard distance deviation in MS Excel

Challenge 1: Measuring geographic distribution in MS Excel

- Step 1: In your web browser, go to the following link:

  https://tinyurl.com/ybata5km

- Step 2: Download a copy of the spreadsheet
  - *File > Download as > Microsoft Excel*

Challenge 1: Measuring geographic distribution in MS Excel

- Step 3: Using MS Excel, calculate the following:
  - Mean center of piracy incidents during *monsoon months*
  - Mean center of piracy incidents during *non-monsoon months*
  - Standard distance of piracy incidents during *monsoon months*
  - Standard distance of piracy incidents during *non-monsoon months*

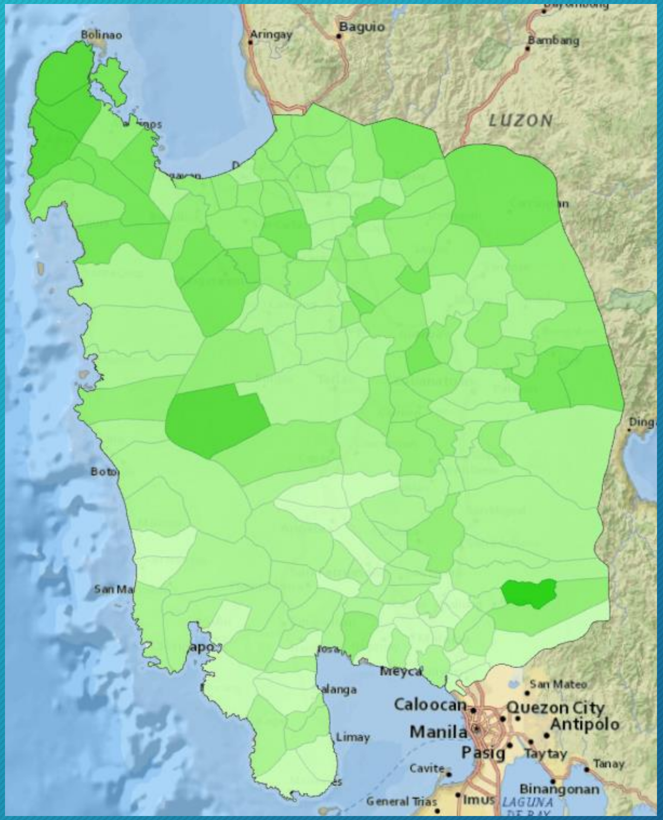
$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

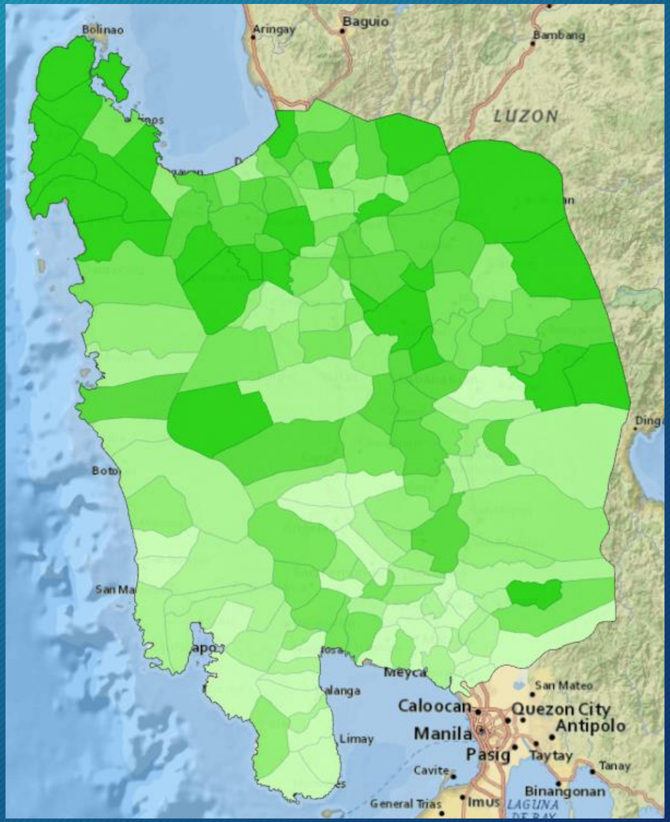
$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}}$$

Identifying patterns

Why identify patterns?

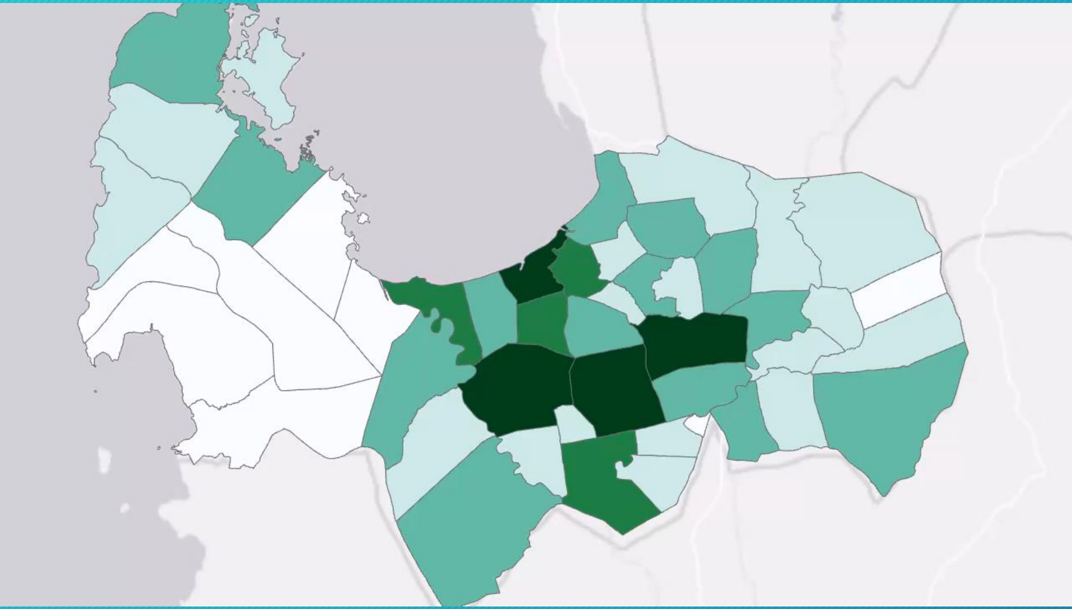
- Better understanding of geographic phenomena
  - Example: Poverty in Central Luzon

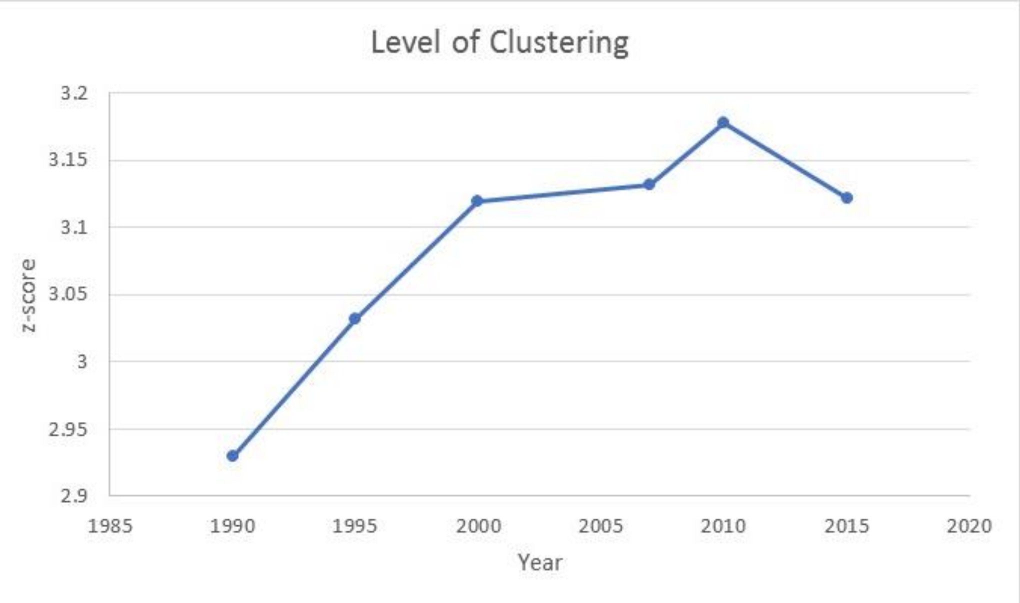




Why identify patterns?

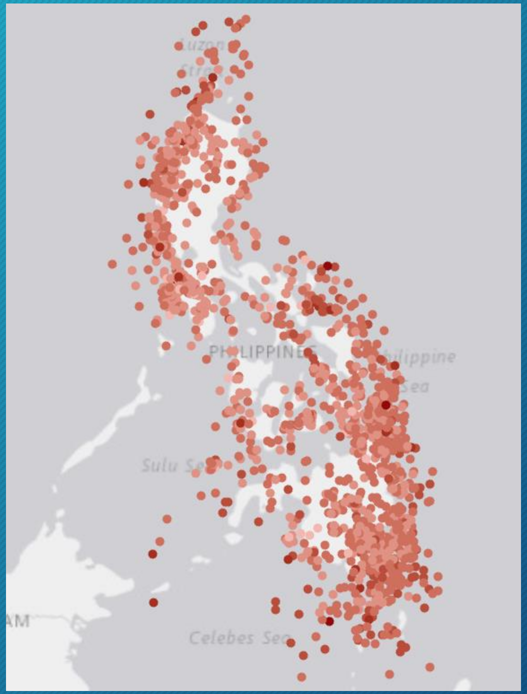
- Monitoring conditions and tracking changes
  - Example: Population of Pangasinan from 1990 to 2015





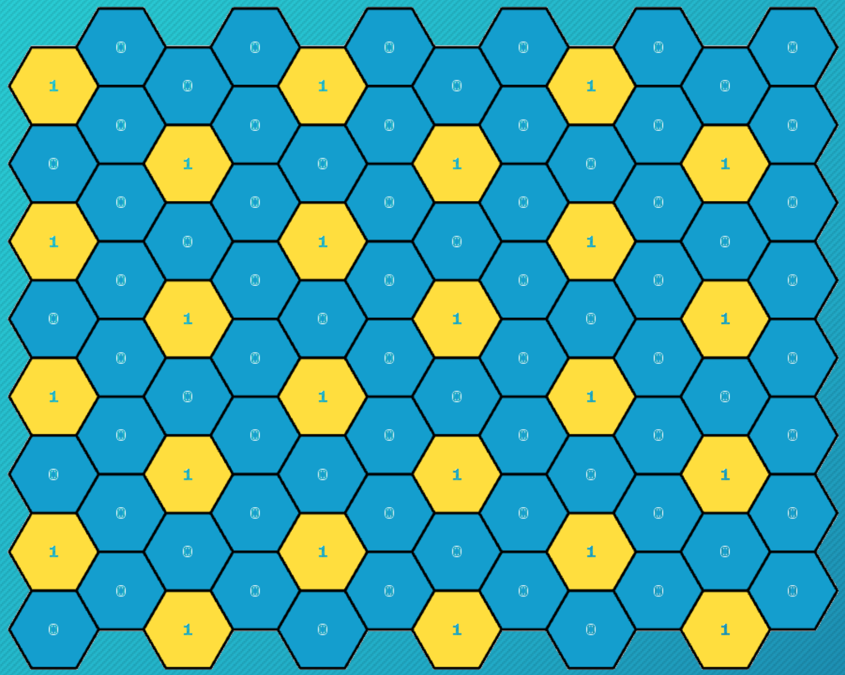
Why identify patterns?

- Comparing the distribution of set of features or values
  - Example: Cluster analysis of earthquakes vs. typhoons

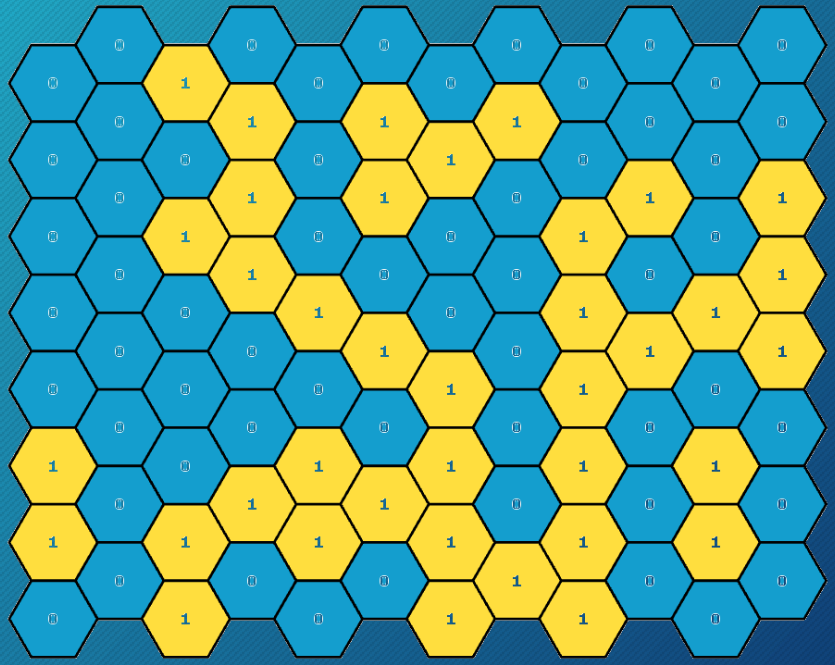


How can statistics identify patterns?

- Comparing the *observed* vs. the *hypothetical* distribution

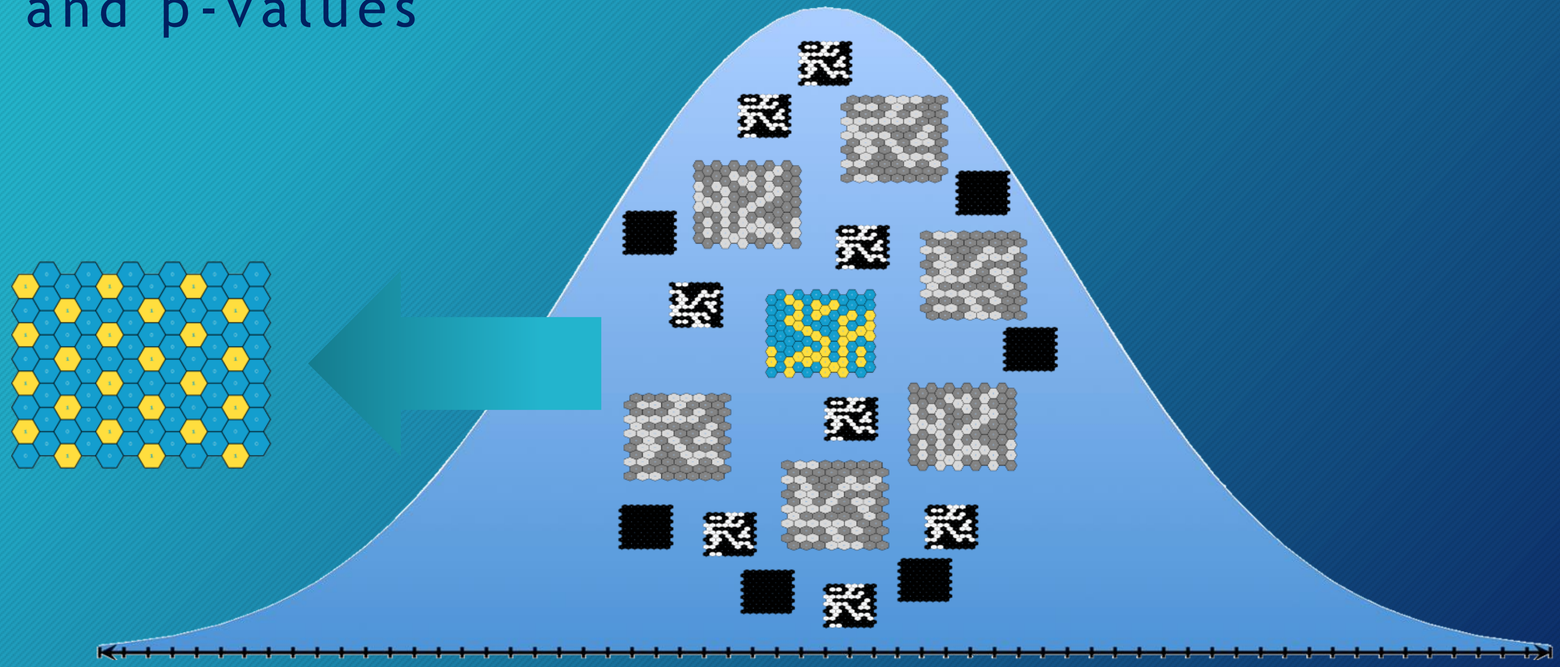


vs.



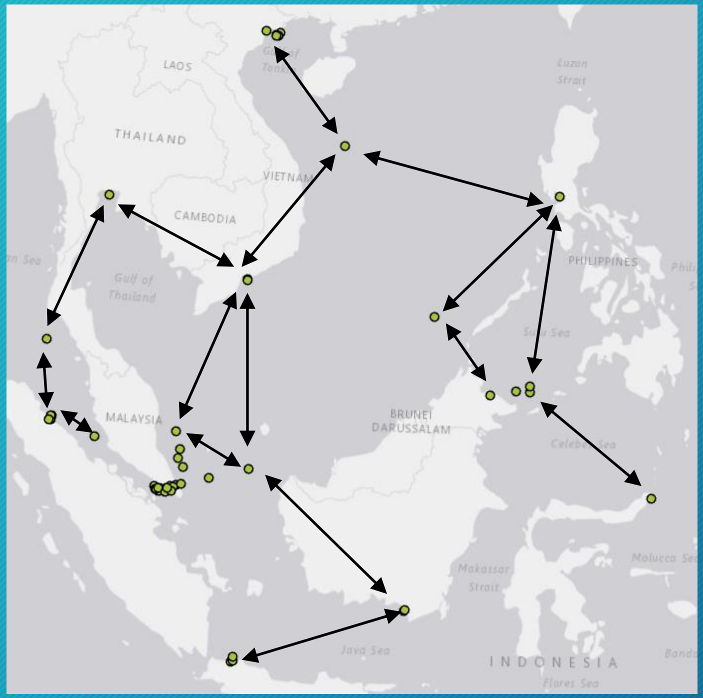
How can statistics identify spatial patterns?

- How different is the *observed* pattern from the *hypothetical* (random) pattern
  - In terms of z-scores and p-values

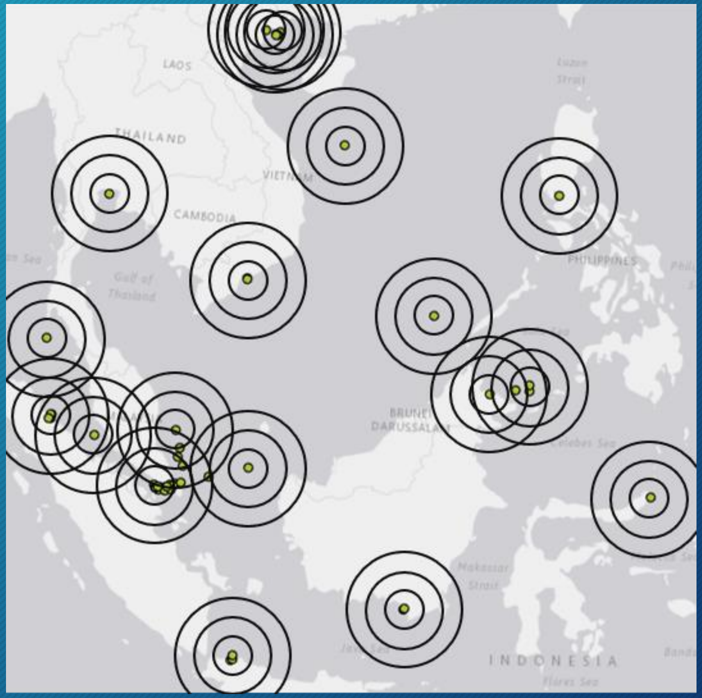


Approaches to identifying patterns
(for feature locations)

Quadrat analysis

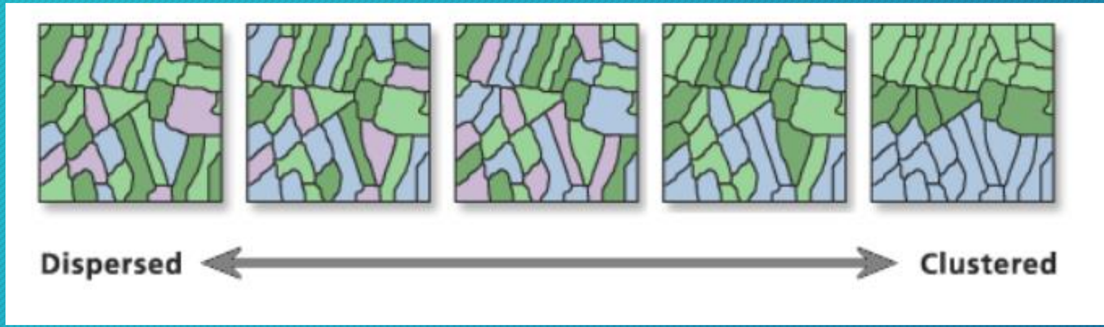


Neighbor distances



Distance bands

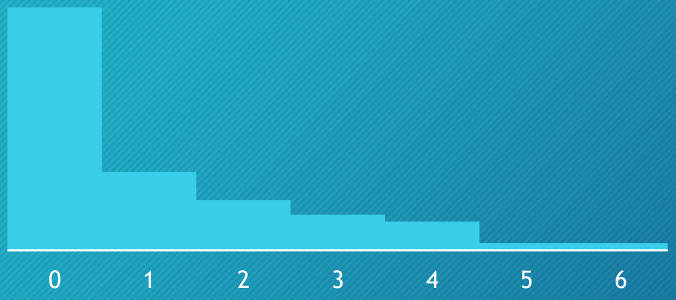
Approaches to identifying patterns
(for area-based values)



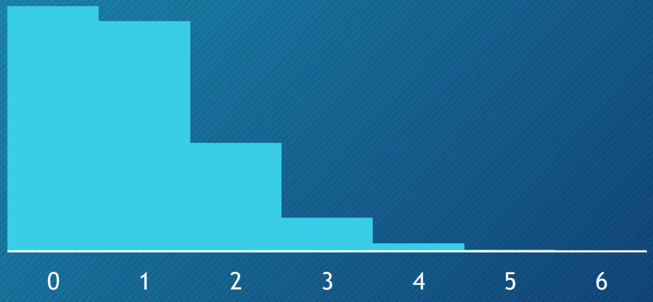
Similarity among neighbors

Approach 1: Quadrat analysis
(for feature locations)

- Actual distribution vs. expected Poissonian distribution of points across quadrats

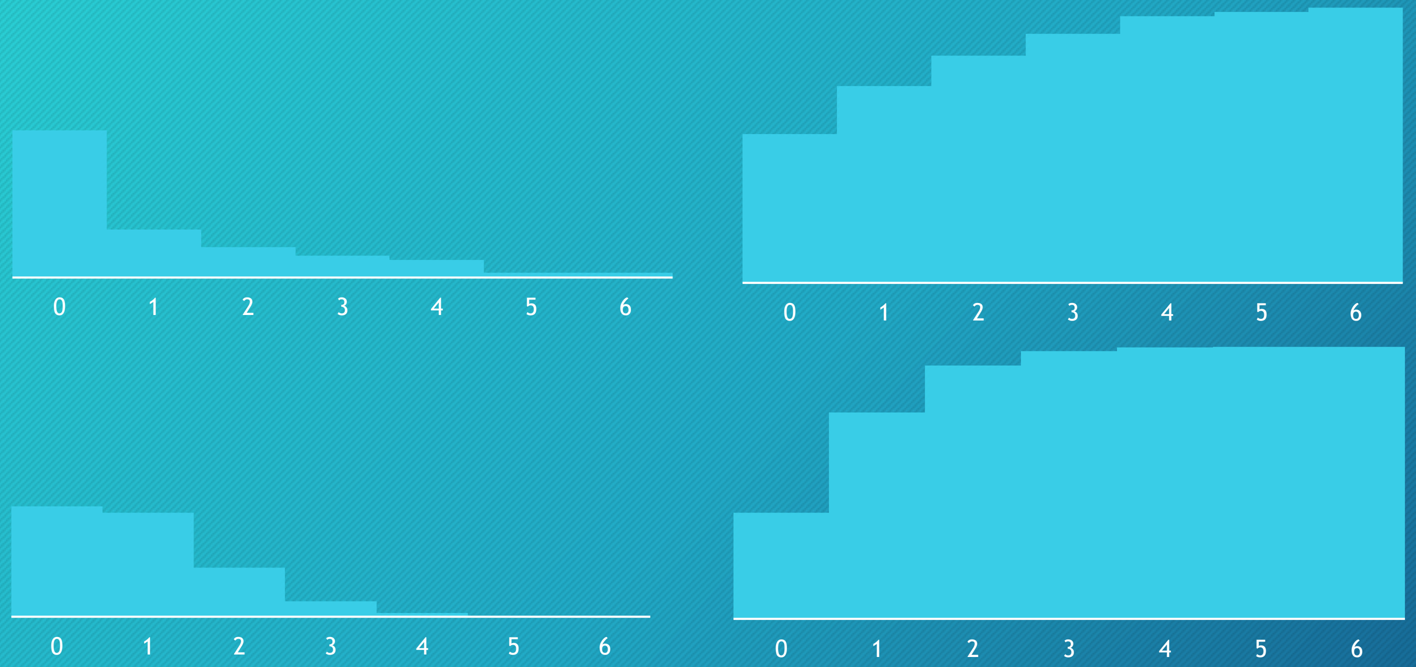


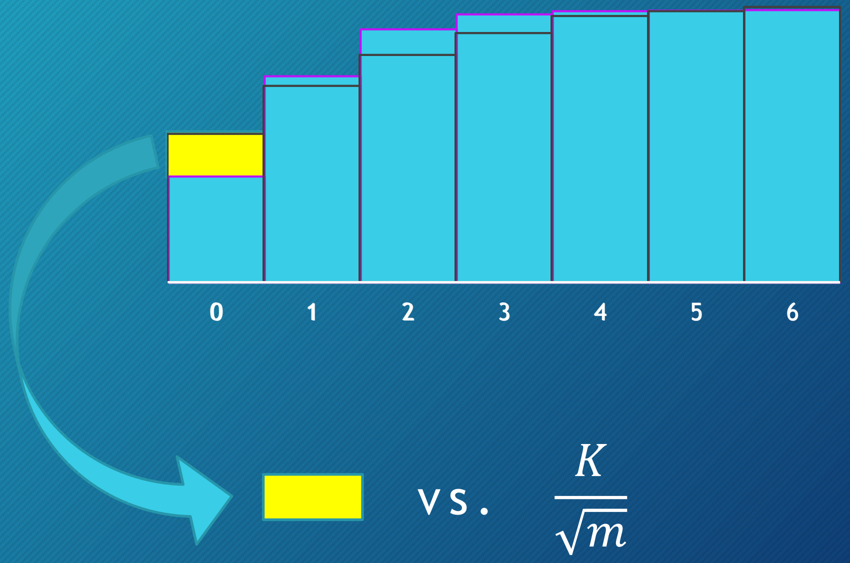
vs.



Approach 1: Quadrat analysis
(for feature locations)

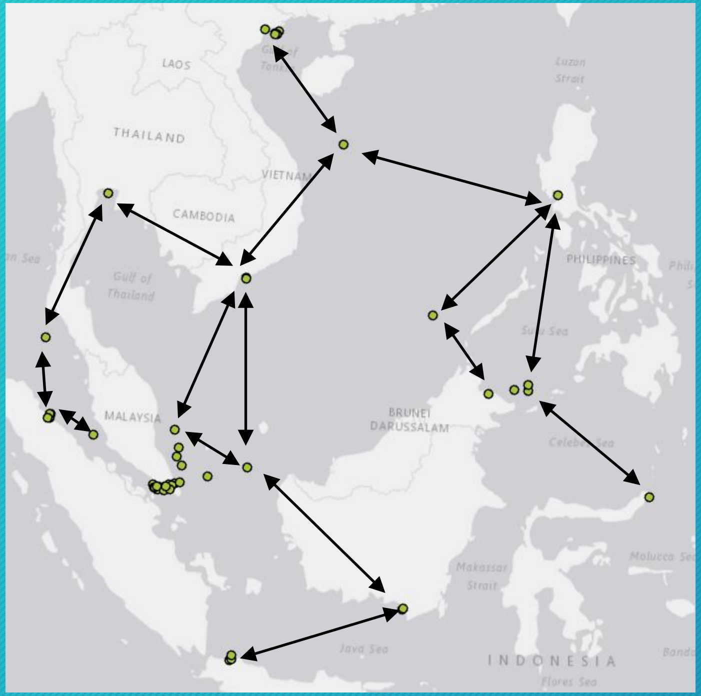
- Comparison performed using:
  - Kolmogorov-Smirnov test
  - Chi-square test





Approach 2: Neighbor distances
(for feature locations)

- Actual vs. expected mean nearest neighbor distance among points



	A	B	C	D	E	F	G
A	0	**988**	2117	2494	3538	3858	4267
B	**988**	0	1725	3348	4308	4004	3601
C	2117	**1725**	0	2804	4034	2567	2309
D	2494	3348	2804	0	**1196**	2510	4897
E	3538	4308	4034	**1196**	0	3277	6034
F	3858	4004	2567	**2510**	3277	0	3433
G	4267	3601	**2309**	4897	6034	3433	0

$$d_o = \frac{\sum c_i}{n}$$

vs.

$$d_e = 0.5\sqrt{A/n}$$

Approach 2: Quadrat analysis (for feature locations)

- Comparison performed using:
  - Nearest neighbor index/ratio
  - Z-test

$$NNI = \frac{d_o}{d_e}$$

$$z = \frac{d_o - d_e}{SE}$$

where

$$SE = \frac{0.2636}{\sqrt{n^2/A}}$$

Identifying point patterns

Challenge 2: Identifying patterns using nearest neighbor approach in MS Excel

Challenge 2: Identifying patterns using nearest neighbor approach

- Step 1: In your web browser, go to the following link:

  https://tinyurl.com/y7ydtfxa

- Step 2: Download a copy of the spreadsheet
  - *File > Download as > Microsoft Excel*

Challenge 2: Identifying patterns using nearest neighbor approach

- Step 3: Using MS Excel, calculate the following:
  - Observed mean distance of drugstores $\bar{D}_O = \frac{\sum_{i=1}^{n} d_i}{n}$
  - Expected mean distance of drugstores $\bar{D}_E = \frac{0.5}{\sqrt{n/A}}$
  - Nearest neighbor index $ANN = \frac{\bar{D}_O}{\bar{D}_E}$
  - Z-score $z = \frac{\bar{D}_O - \bar{D}_E}{SE}$

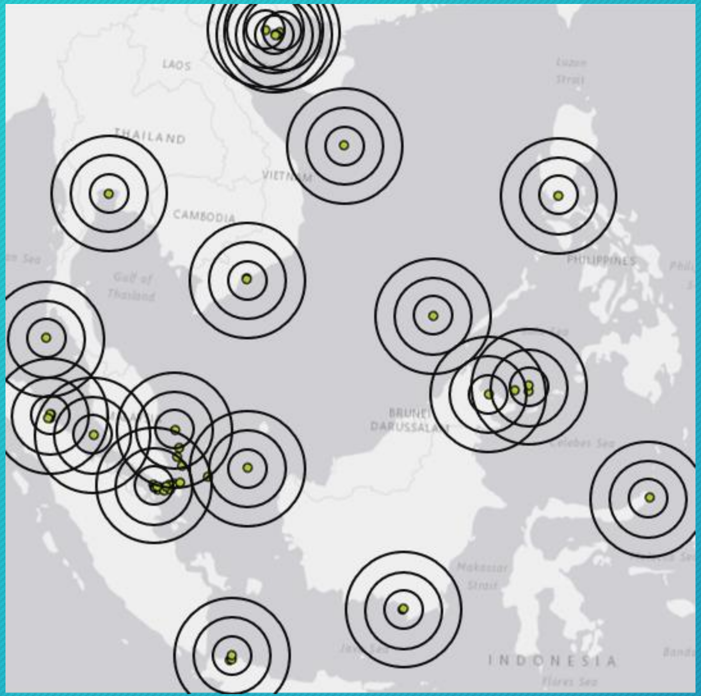
    $$SE = \frac{0.26136}{\sqrt{n^2/A}}$$
  - (optional) p-value

Identifying point patterns

Demo 2: Identifying patterns using nearest neighbor approach in ArcGIS Pro

Approach 3: Distance bands
(for feature locations)

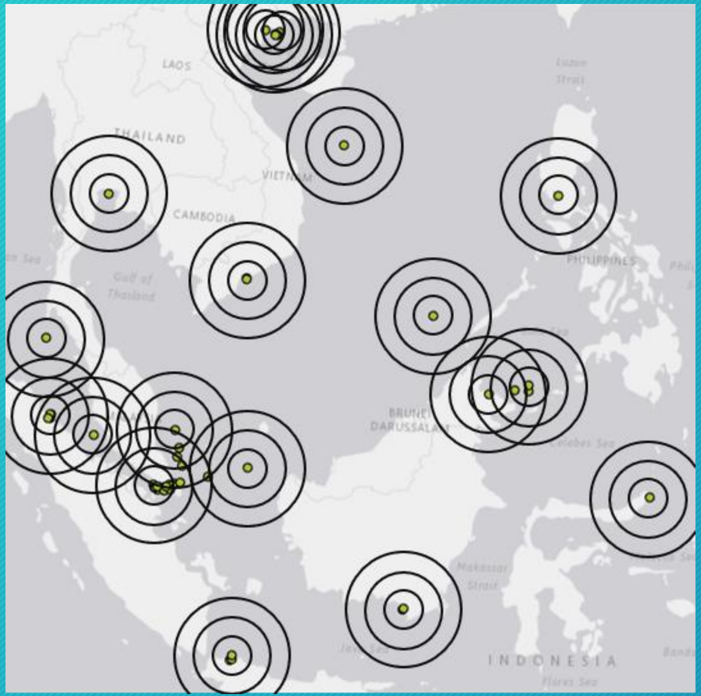
- Actual vs. expected number of points at every distance band *k*



$$K(d) = \frac{A}{n^2} \sum_{i \neq j} \sum_{i \neq j} I_{ij} d_{ij}$$

Approach 3: Distance bands
(for feature locations)

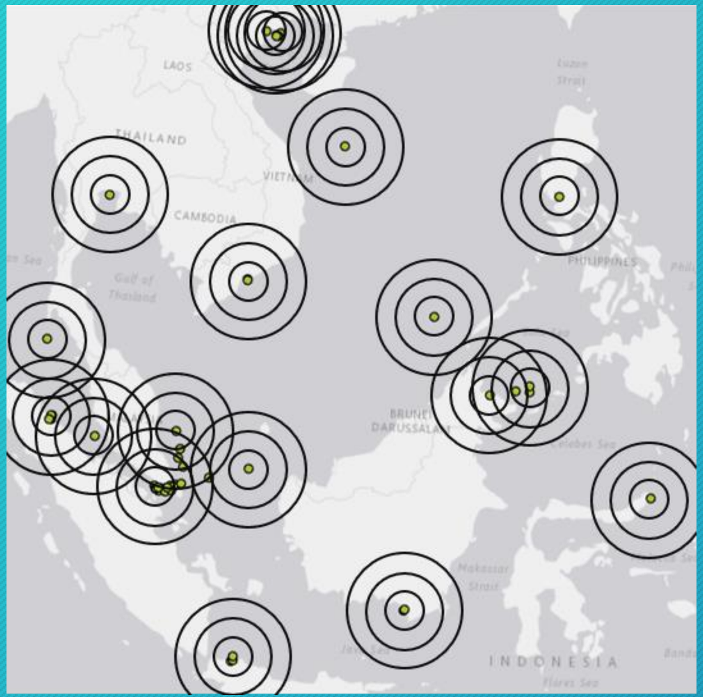
- Actual vs. expected distance of all neighboring points

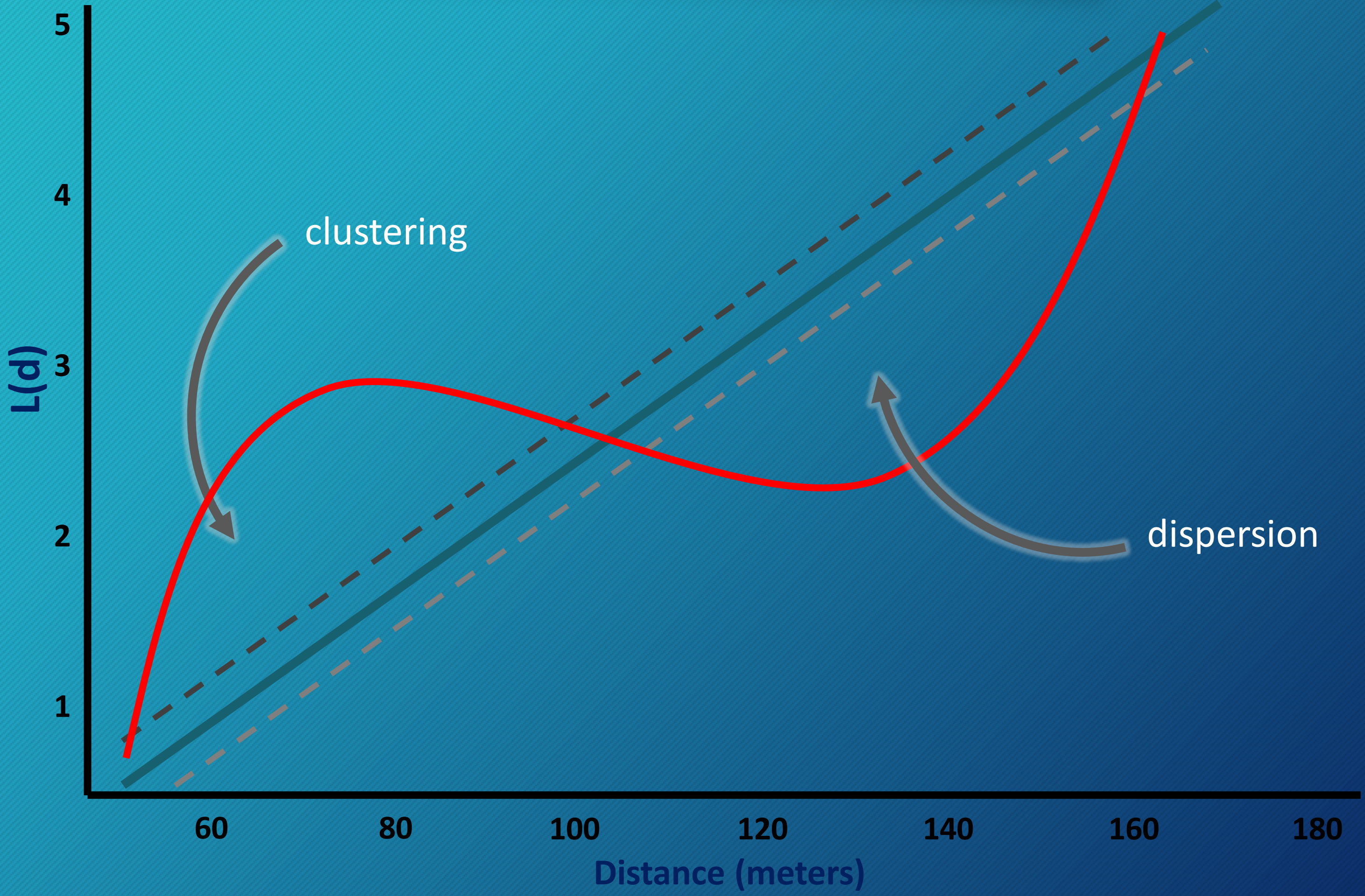


$$K(d) = \frac{A}{n^2} \sum_{i \neq j} \sum_{i \neq j} I_{ij} d_{ij}$$

$$L(d) = \sqrt{\frac{A \sum_{i \neq j} \sum_{i \neq j} I_{ij} d_{ij}}{\pi n(n-1)}} \quad \text{vs.} \quad d$$

Approach 3: Distance bands
(for feature locations)



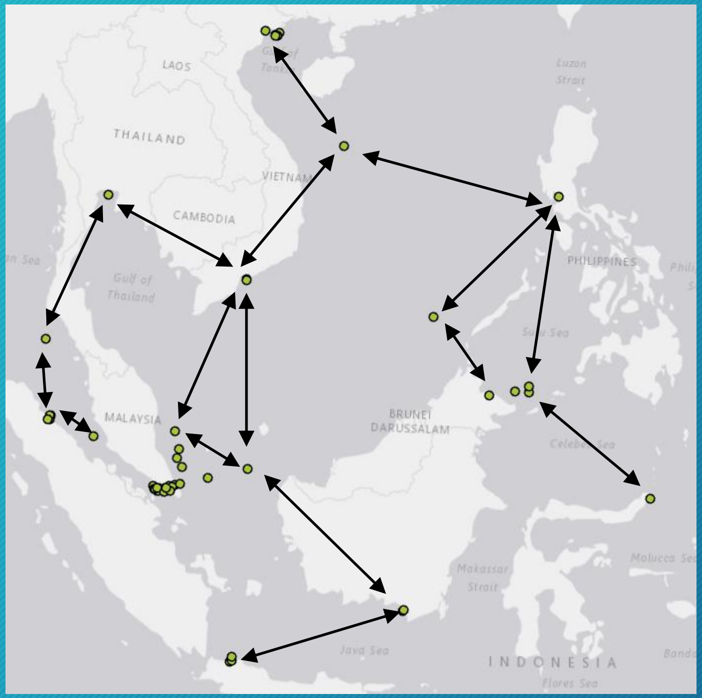


Identifying point patterns

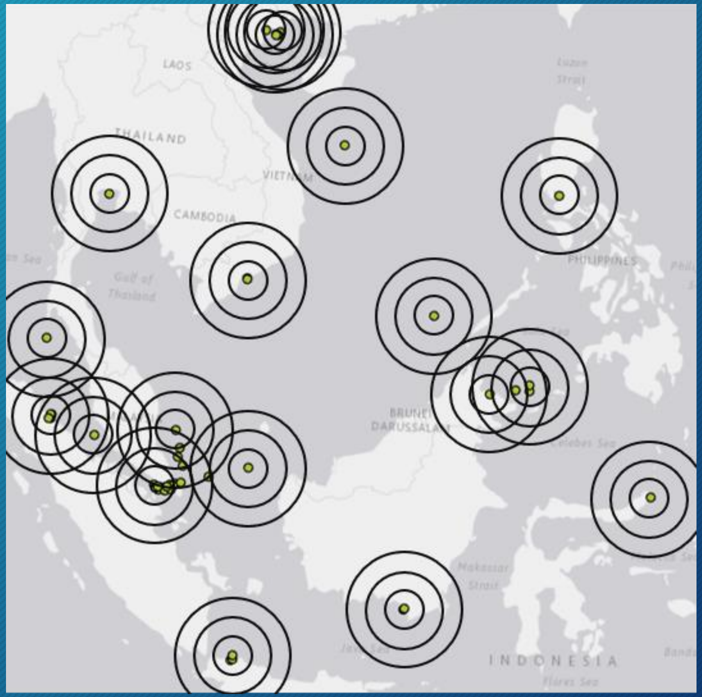
Demo 3: Identifying patterns using distance bands approach in ArcGIS Pro

Recap: Identifying patterns
(for feature locations)

Quadrat analysis



Neighbor distances

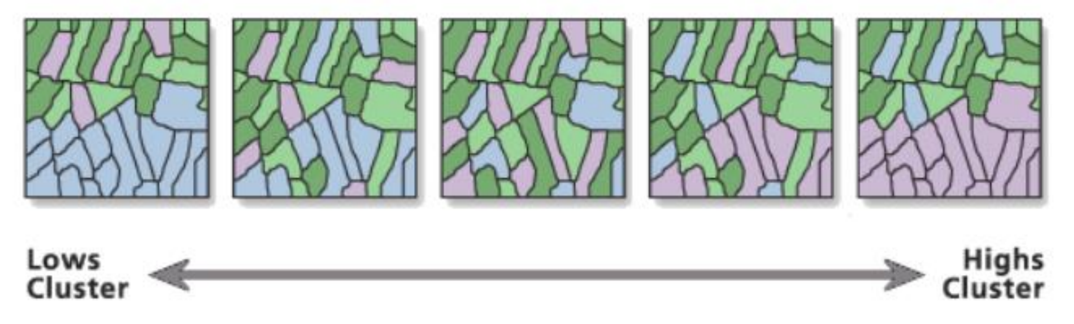


Distance bands

Approaches to identifying patterns
(for area-based values)

- Operates on the concepts of
  - Spatial autocorrelation
  - Tobler's First Law of Geography

Similarity among neighbors

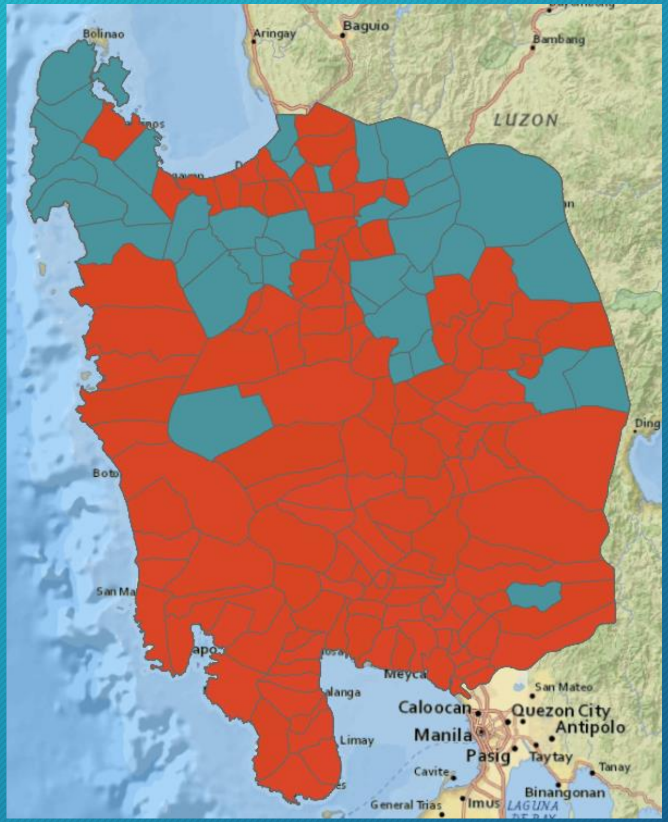


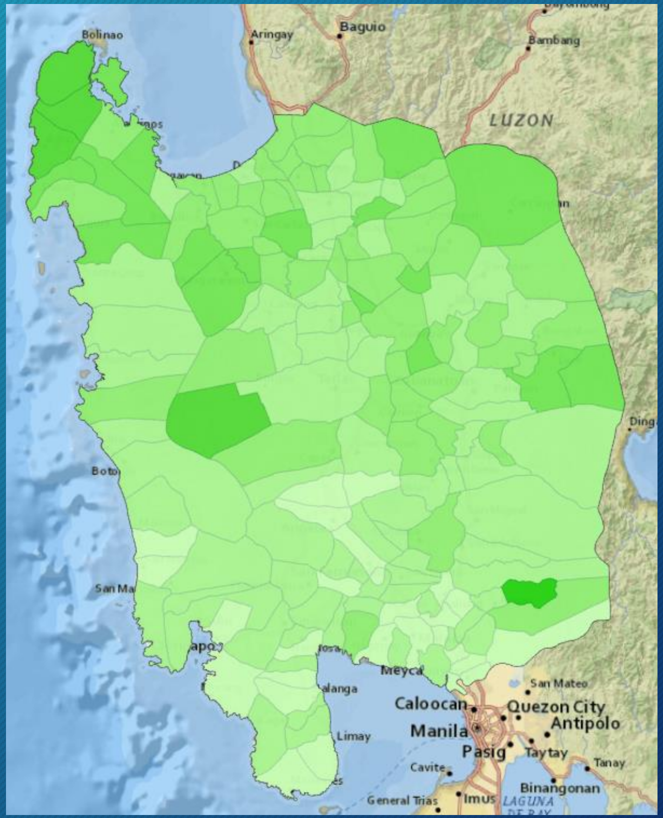
Cluster characterization

"Everything is related to everything else, but nearer things are more related than distant things." – Waldo Tobler

Approach: Similarity of neighbors
(for area-based values)

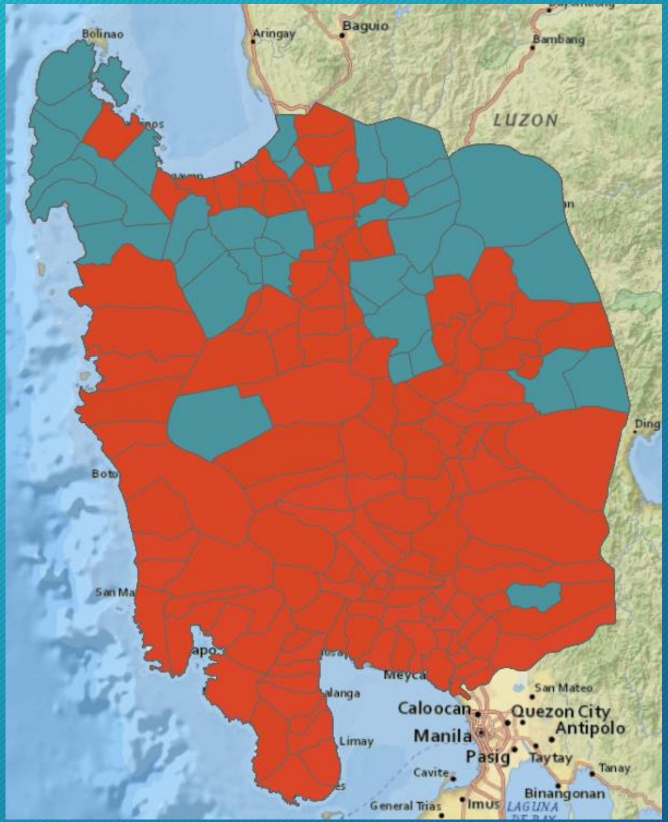
- How nearby values differ or conform with one another
- Possible for categorical, or quantitative values

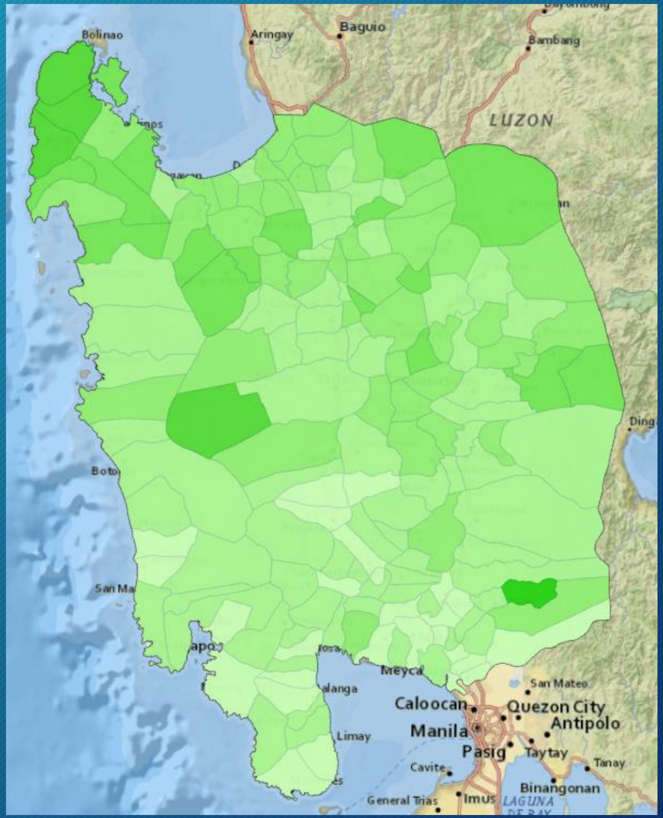




Approach: Similarity of neighbors
(for area-based values)

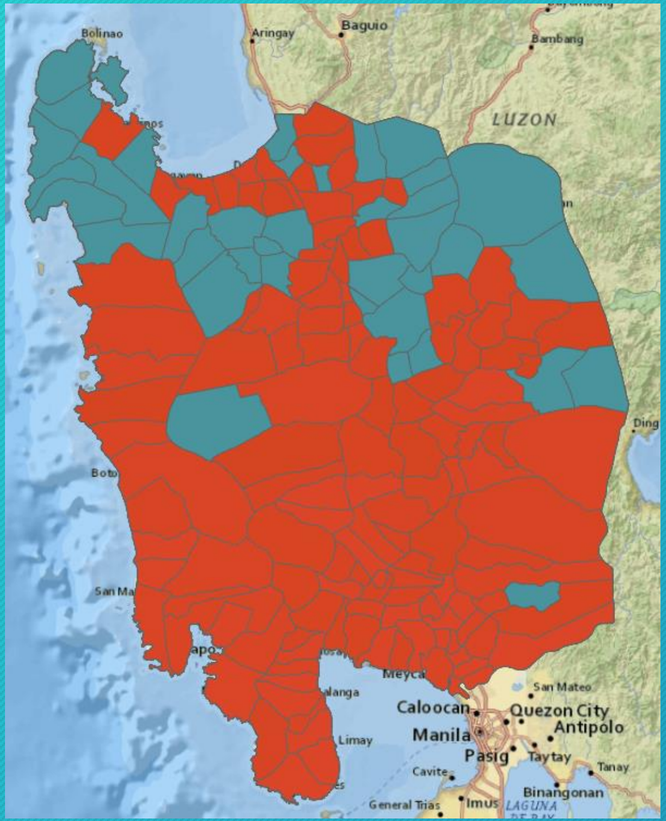
- For categorical data:
  - Join count statistic
- For quantitative data:
  - Geary's *c*
  - Global Moran's *I*





Approach: Similarity of neighbors
(for area-based values)

- Actual vs. expected distribution of adjacent similar categories



$$O_{11} = \sum_i \sum_j [w_{ij}(x_i x_j)] \qquad \text{vs.} \qquad E_{11} = p_1^2 L$$

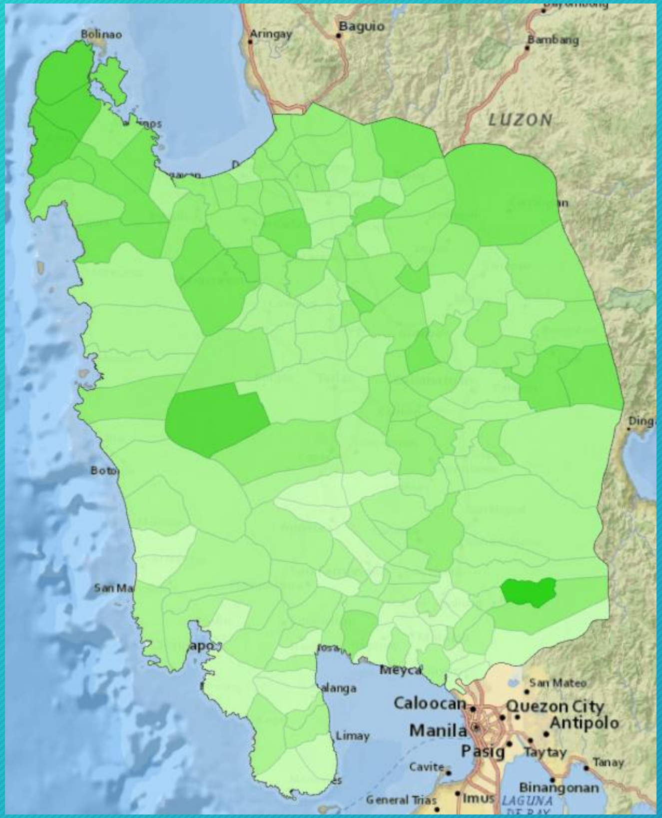
$$z_{01} = \frac{O_{11} - E_{11}}{\sigma_{11}}$$

Identifying area patterns

Demo 4: Identifying patterns using join count test function in *R*

Approach: Similarity of neighbors
(for area-based values)

- Actual vs. expected distribution of adjacent similar values



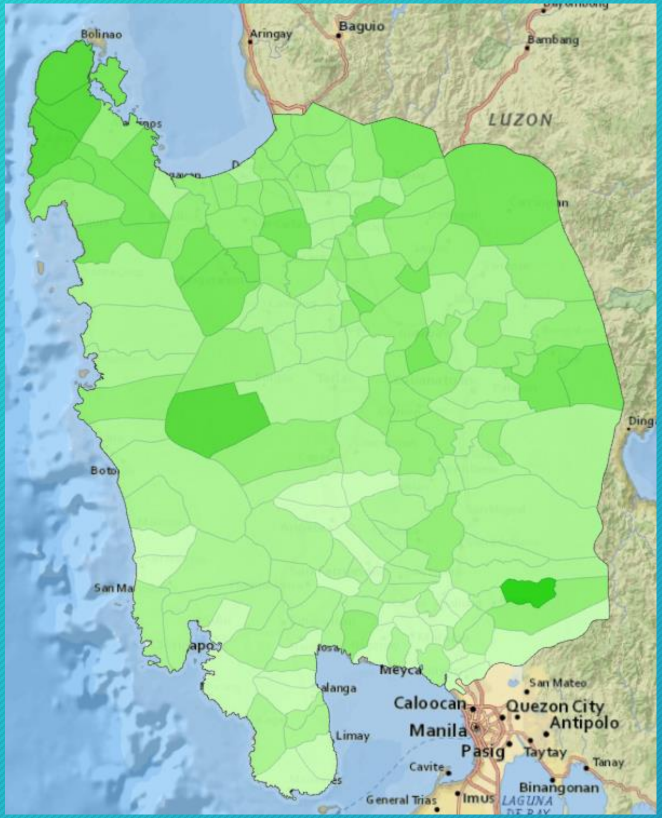
Using Geary's c

$$c_o = \frac{n\sum_i \sum_j w_{ij}(x_i - x_j)^2}{2\sum_i \sum_j w_{ij}(x_i - \bar{X})^2} \quad \text{vs.} \quad c_e = 1$$

$$z_c = \frac{c_O - c_e}{\sigma_e}$$

Approach: Similarity of neighbors (for area-based values)

- Actual vs. expected distribution of adjacent similar values



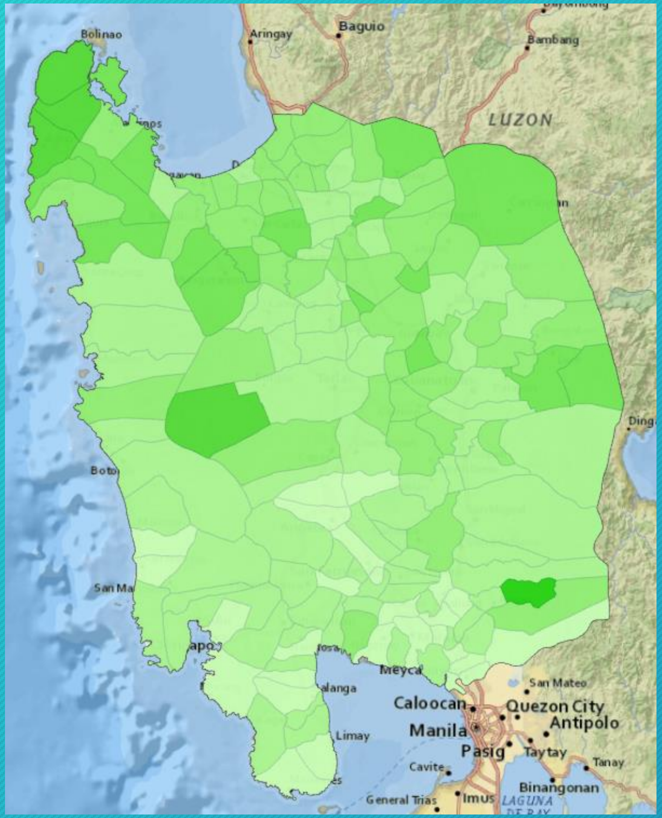
Using Moran's *I*

$$I_o = \frac{n \sum_i \sum_j w_{ij}(x_i - \bar{X})(x_j - \bar{X})}{2 \sum_i \sum_j w_{ij}(x_i - \bar{X})^2} \quad \text{vs.} \quad I_e = \frac{-1}{n-1}$$

$$z_I = \frac{I_o - I_e}{\sigma_e}$$

Approach: Similarity of neighbors
(for area-based values)

- Actual vs. expected distribution of adjacent similar values



Geary's *c*	Moran's *I*	D
$c < 1$	$I > 0$	Clustered
$c = 1$	$I = 0$	Random
$c > 1$	$I < 0$	Dispersed

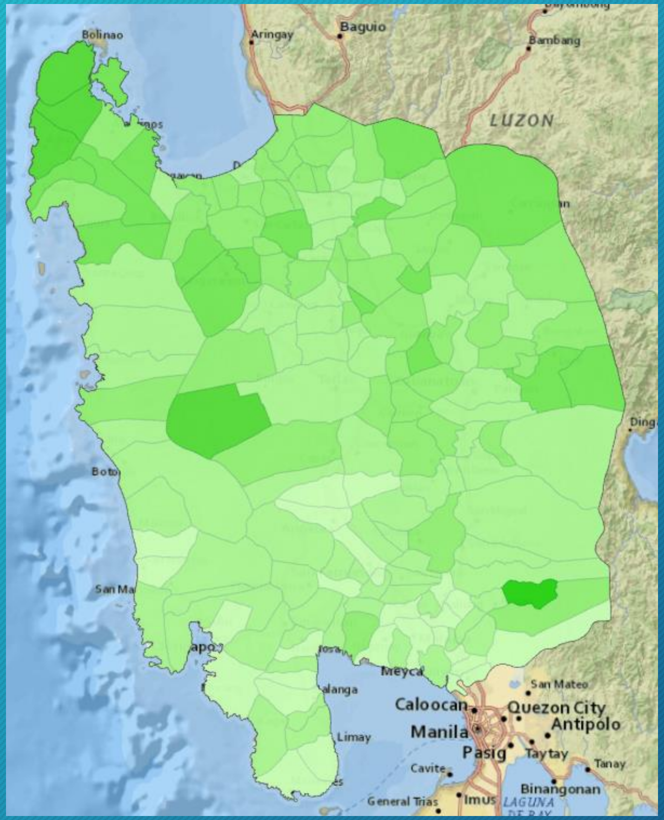
Identifying area patterns

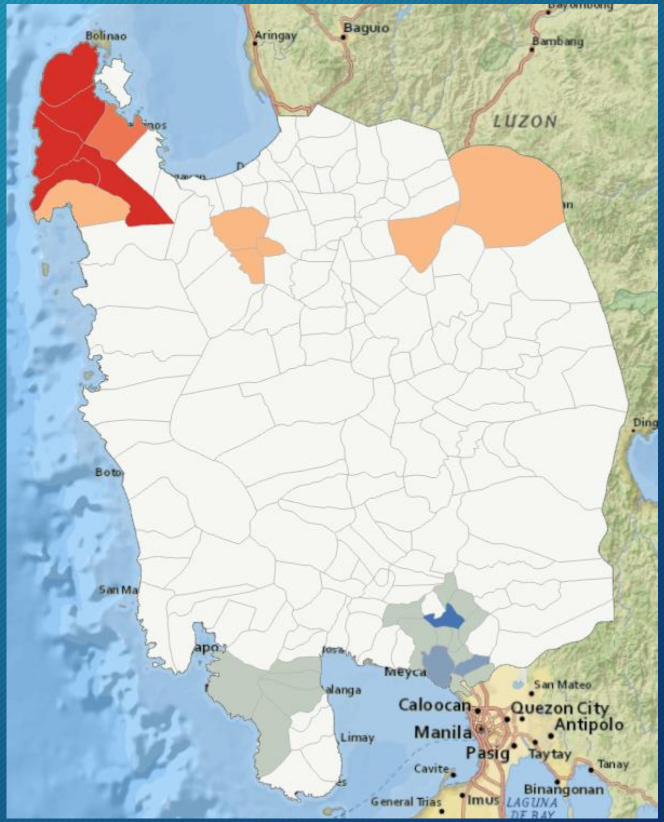
Demo 5: Identifying patterns using Geary’s, and Moran’s test functions in *R*

Identifying clusters

Why identify clusters?

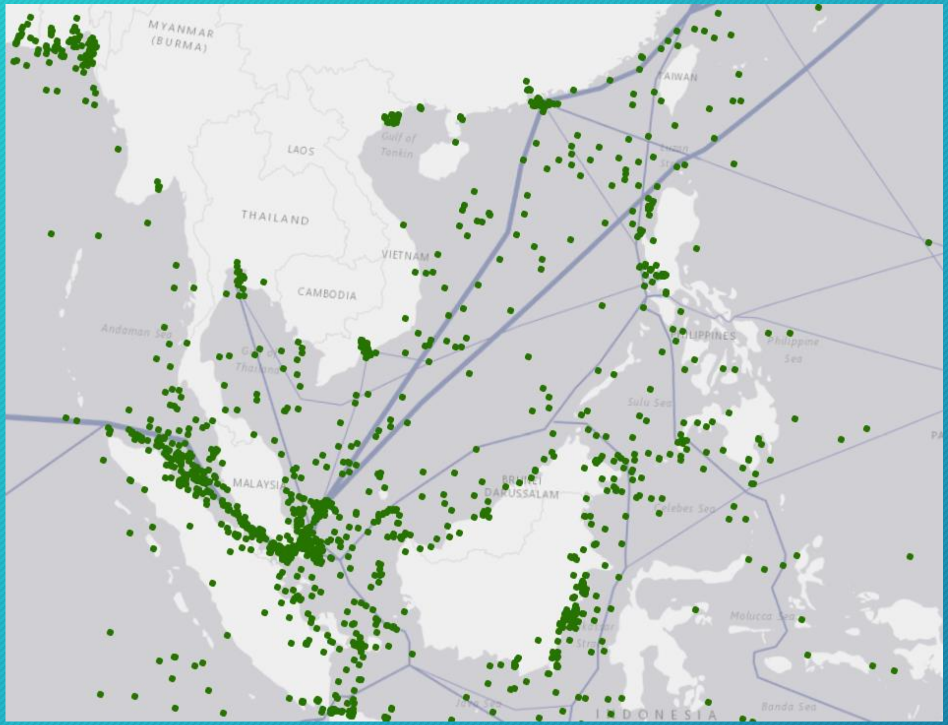
- Delineate specific focus of interest or priority
  - Example: Hotspot and cold-spot of poverty

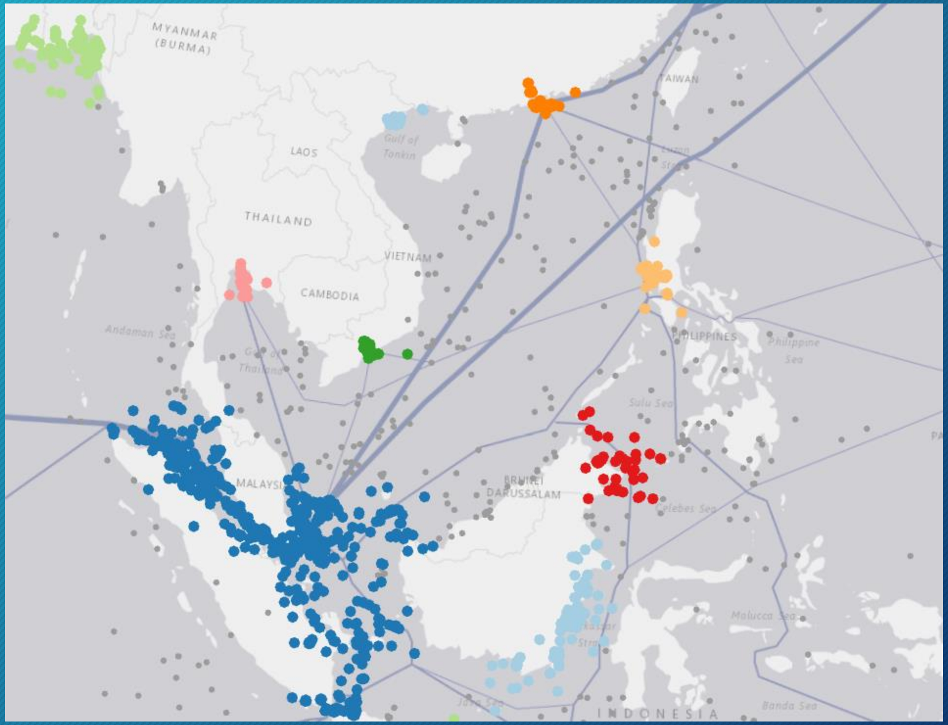




Why identify patterns?

- Explore and examine the possible cause of clustering
  - Example: Piracy and maritime routes





Identifying clusters among feature locations

- Clusters are defined through a *threshold distance*
- May be determined by finding the mean random distance's *confidence interval lower limit*

$$0.26136\sqrt{A/n^2}$$

$$d_c \pm (z * SE)$$

$$0.5\sqrt{A/n}$$

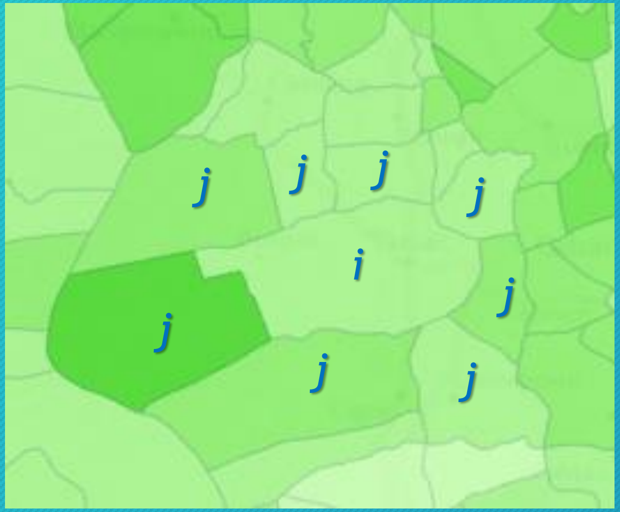
Z-score, depending on confidence level
e.g. 1.96 for 95% c.i.

Identifying point clusters

Demo 6: Detecting point clusters in ArcGIS Pro

Identifying clusters among area-based values

- Features are individually assessed for similarity with its neighboring values
  - Geary's c_i
  - Moran's I_i



$$I_i = \frac{(x_i - \bar{X})}{s^2} \cdot \sum_j w_{ij}(x_j - \bar{X})$$

vs.

$$I_e = \frac{-\sum_j w_{ij}}{n-1}$$

$$z_I = \frac{I_i - I_e}{\sigma_{I_i}}$$

Identifying area clusters

Demo 6: Detecting cluster of area-based values using Moran's I_i in ArcGIS Pro

In summary

"Spatialize" data

As point locations, or as area data

Quadrat analysis

Nearest neighbor index

Geary's *c*

K- and L-function

Moran's *I*

Standard distance deviation

Threshold distance

Standard deviational ellipse

Mean center

Moran's I_i

Assessing Spatial Patterns Using Statistics

Thank you for attending.