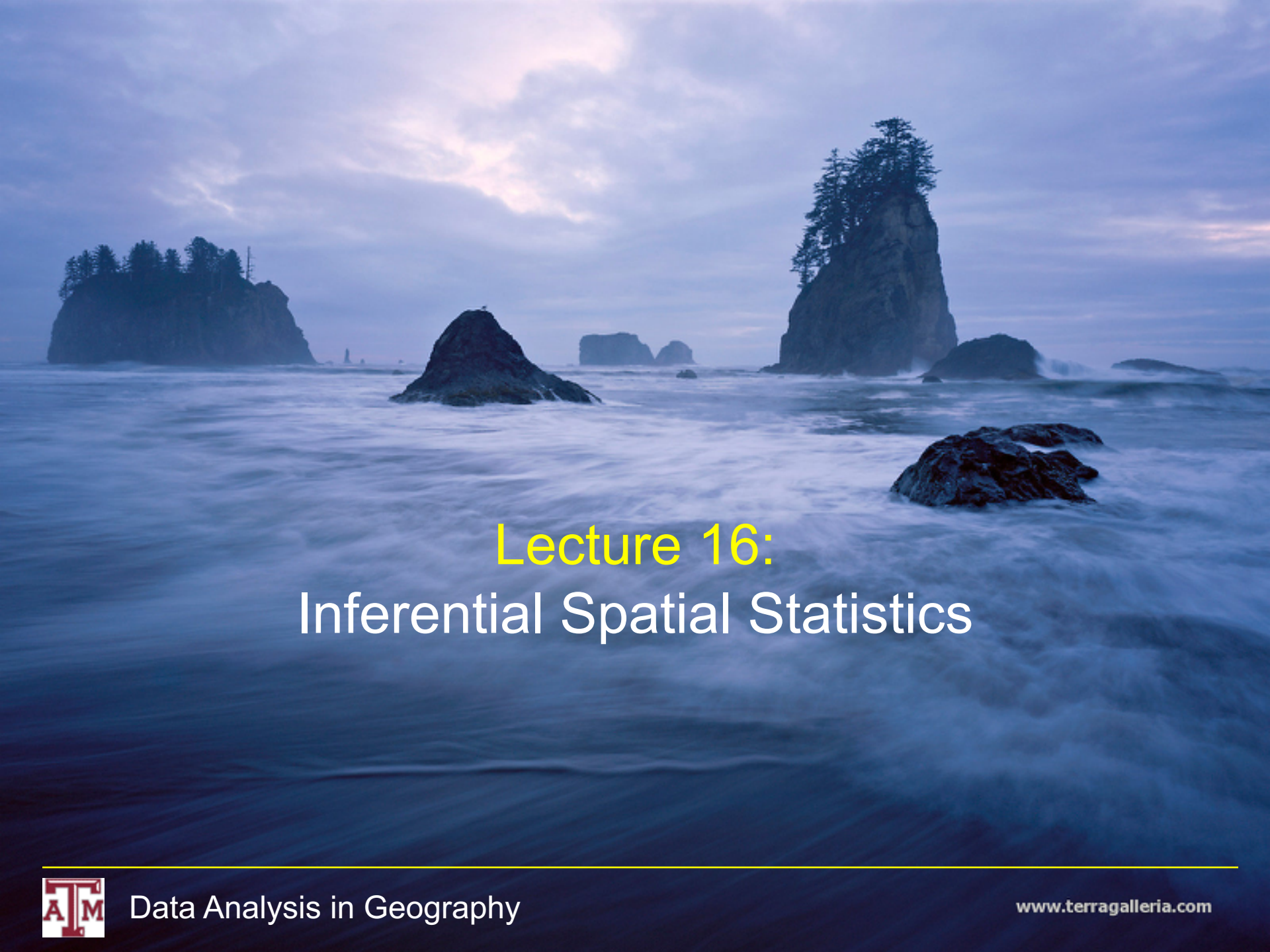


Lecture 16:
Inferential Spatial Statistics

Reminders

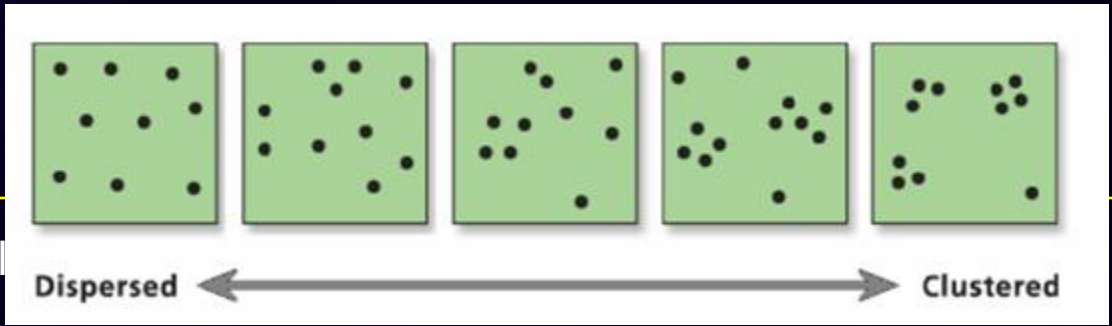
- Assignment #5 – data is due Today (3/21)
- Final Project Topic Proposals are due Tuesday (3/26)
- Assignment #5 – analysis is due next Thursday (3/28)

Inferential Spatial Statistics

- **Inferential**: making predictions or conclusions about a population based on a sample of data
- **Spatial**: relating to space
- Making inferences:
  - Sample(s) about population(s)
  - Observed patterns about underlying processes
- Statistical inference is based on the assumption that you have a randomly spatially distributed sample

Inferential Spatial Statistics

- **Random**: a point is equally likely to occur at any location, and the position of a point is not affected by the position of any other point.
- **Dispersed**: every point is as far from other points as possible: “likely to be distant”
- **Clustered**: every point is close to other points: “likely to be close”



Inferential Spatial Statistics

- Fact: Two times as many people sit “on the corners” rather than opposite at tables in a restaurant
  - Conclusion: Psychological preference for nearness

Inferential Spatial Statistics

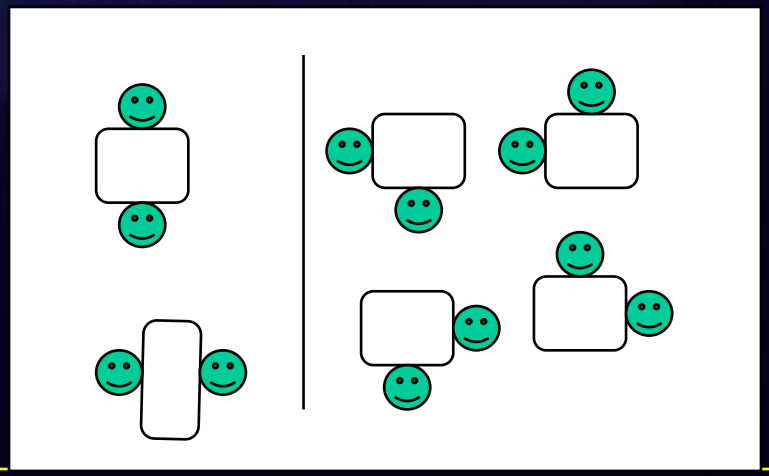
- Fact: Two times as many people sit “on the corners” rather than opposite at tables in a restaurant
  - Conclusion: Psychological preference for nearness

- Actually a random spatial occurrence:
  - Two ways to sit opposite
  - Four ways to sit on the corners

Inferential Spatial Statistics

- Fact: Two times as many people sit “on the corners” rather than opposite at tables in a restaurant
  - Conclusion: Psychological preference for nearness

- Actually a random spatial occurrence:

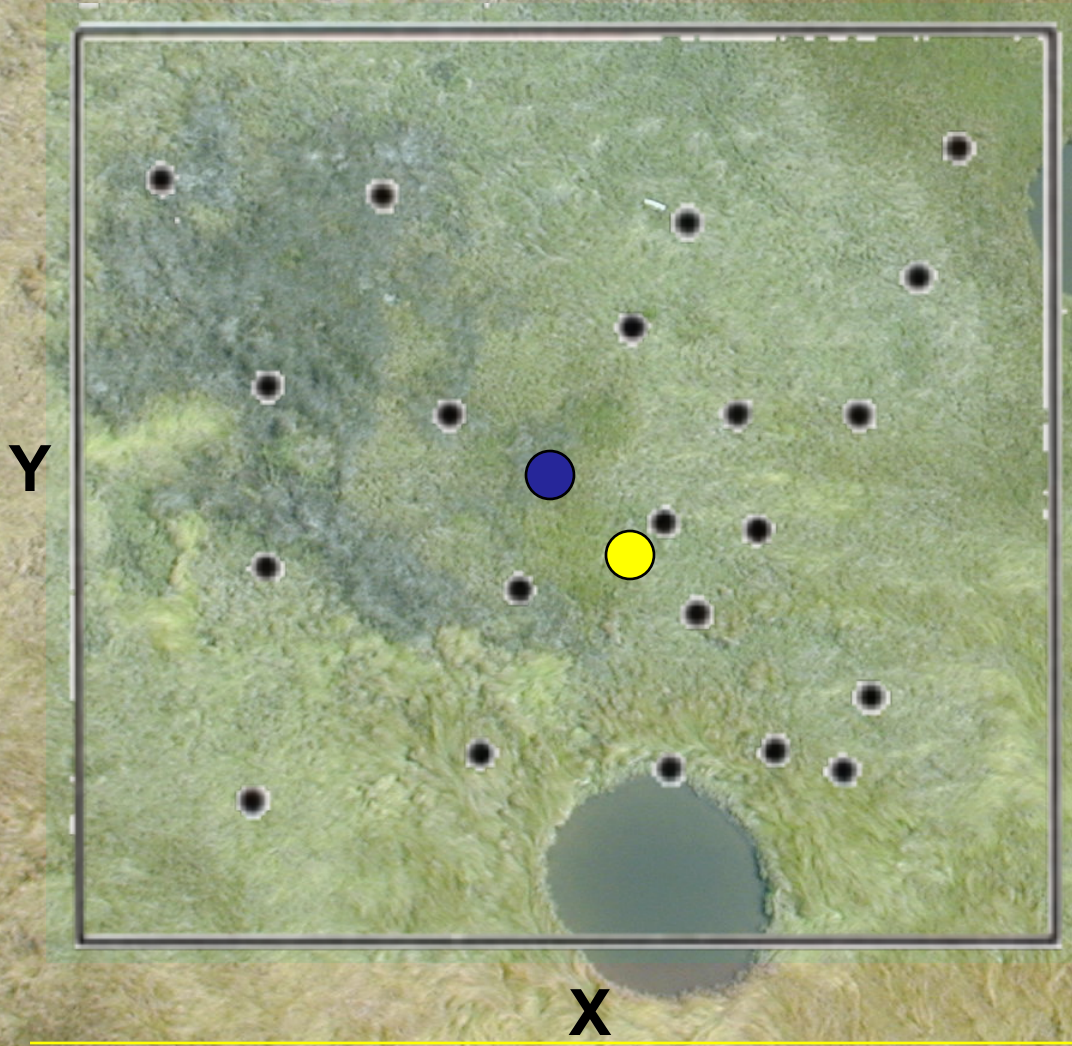


Inferential Spatial Statistics

- Null Hypothesis: Spatial pattern is **random**
  - Independent random process
  - Complete spatial randomness

- Alternative Hypothesis: Spatial pattern is **not random**
  - Clustered or Dispersed

Spatial Statistics



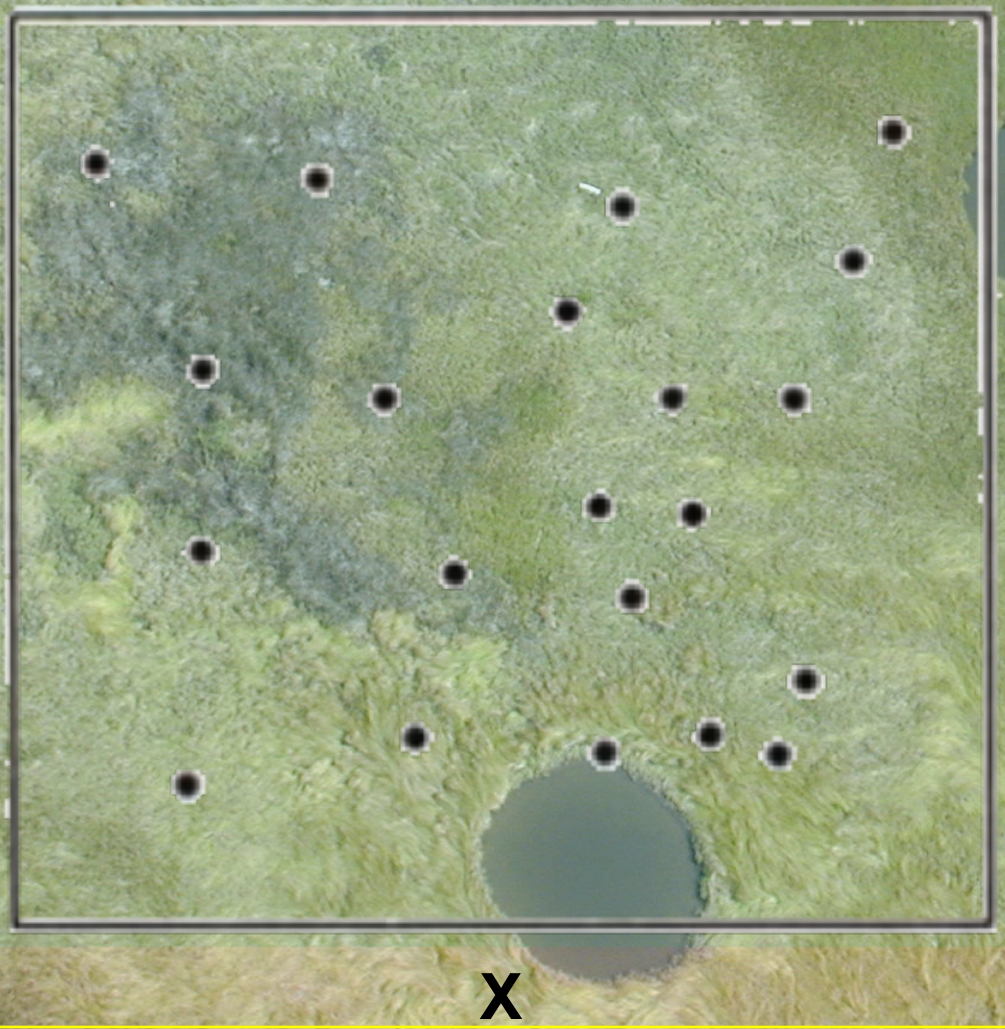
Spatial center (or more specifically the arithmetic center/mean center) is calculated as the average of x and the average of y

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n}$$

$$\overline{Y} = \frac{y_1 + y_2 + ... + y_n}{n}$$

Arithmetic center is different from the geometric center (or centroid)

Weighted Center



Weighted center is based on the sample value

$$\bar{X} = \frac{v_1 x_1 + v_2 x_2 + ... + v_3 x_n}{v_1 + v_2 + ... + v_3}$$

$$\bar{Y} = \frac{v_1 y_1 + v_2 y_2 + ... + v_3 y_n}{v_1 + v_2 + ... + v_3}$$

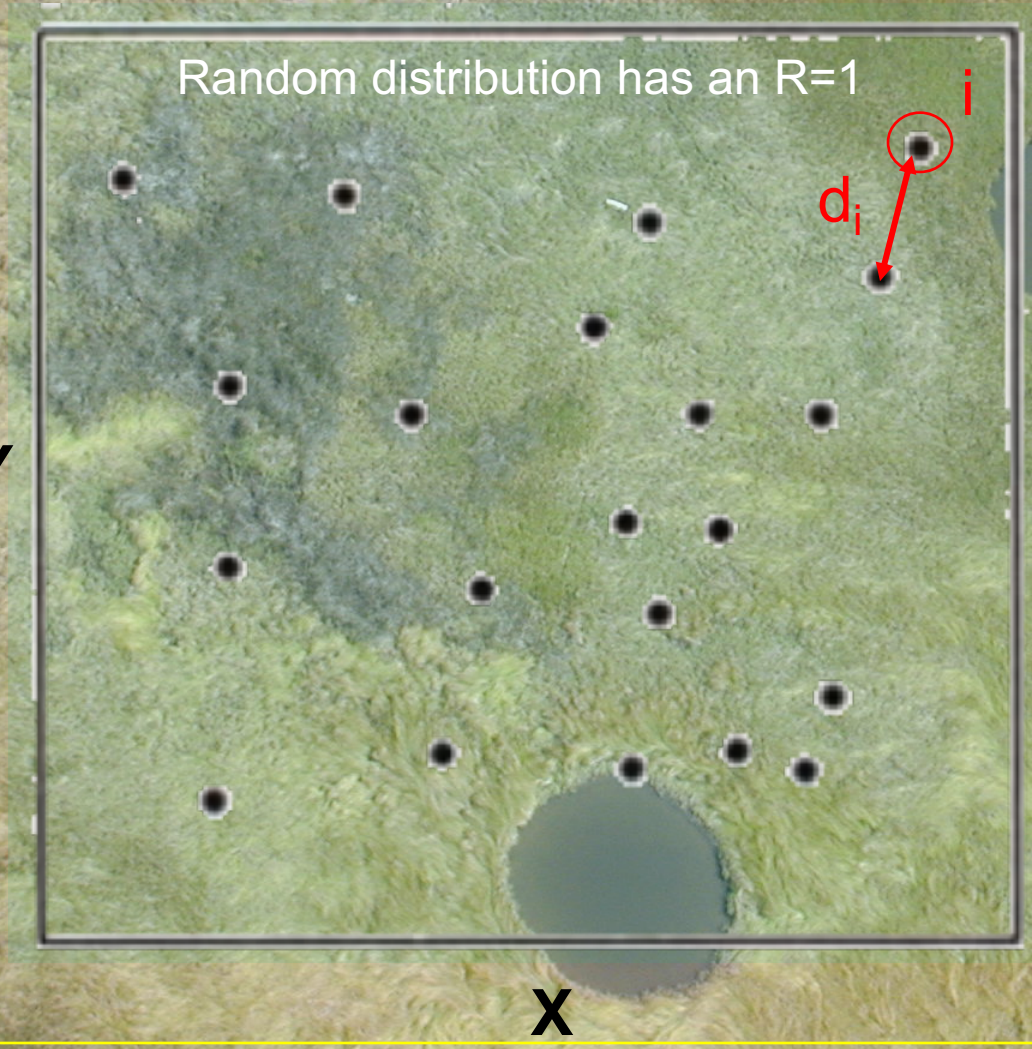
Value dependent center

Standard Distance

Average distance of each sample point to the spatial center

$$S_D = \sqrt{\frac{\Sigma(X_i - \bar{X}_c)^2 + \Sigma(Y_i - \bar{Y}_c)^2}{n}}$$

Nearest Neighbor Distance



What is the spatial arrangement of a variable or your sampling in the study area?

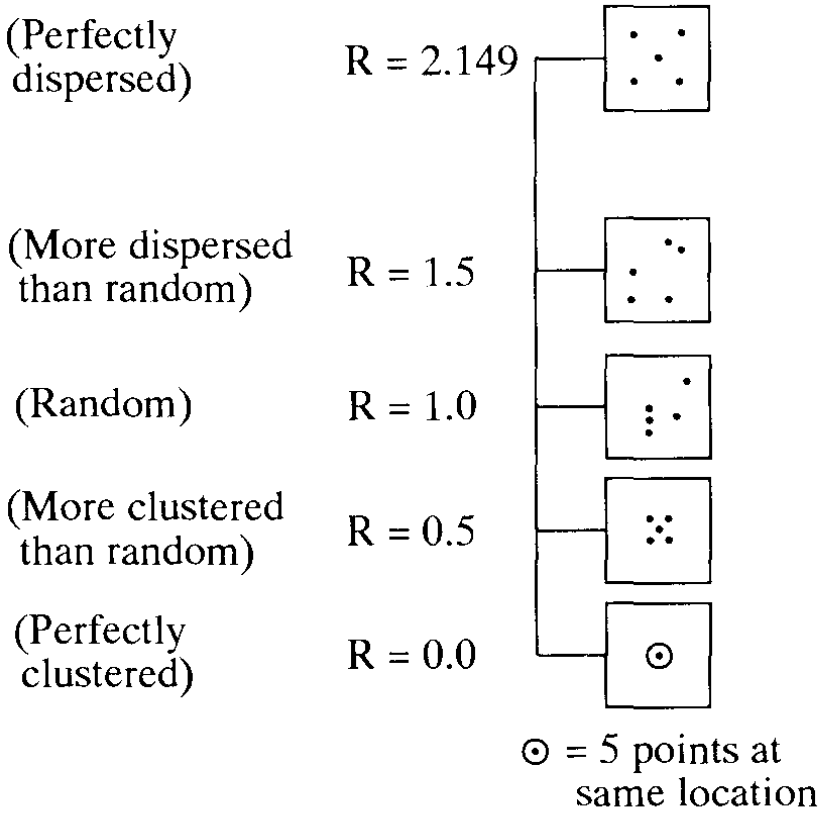
$$NND = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$NND_R = 0.5 \times \frac{1}{\sqrt{\frac{n}{A}}}$$

$$R = \frac{NND}{NND_R}$$

The larger the R the more dispersed the samples, while a smaller R reflects clustering

Figure 11.3 Continuum of *R* Values in Nearest Neighbor Analysis

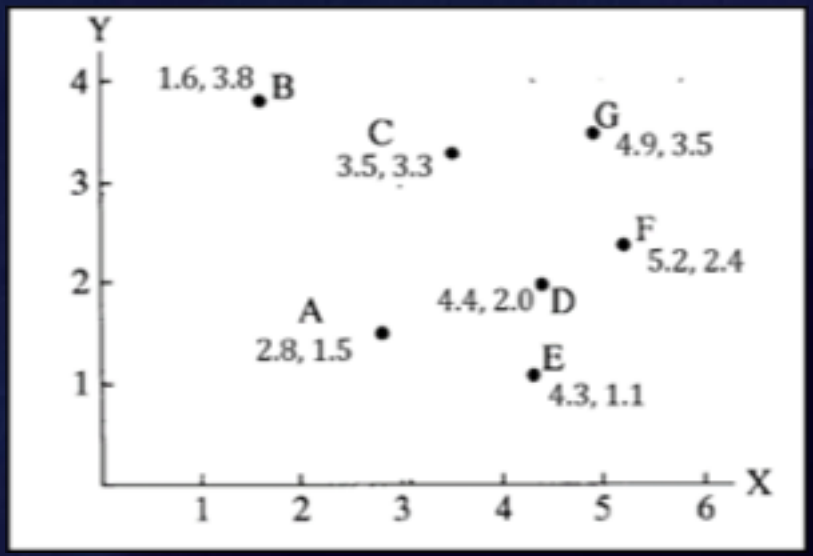


Source: Modified from Taylor, P. J. 1977. *Quantitative Methods in Geography*. Boston: Houghton Mifflin

Nearest Neighbor

- Step 1: Calculate the distance from each point to its nearest neighbor, by calculating the hypotenuse of the triangle

$$NND_{AB} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

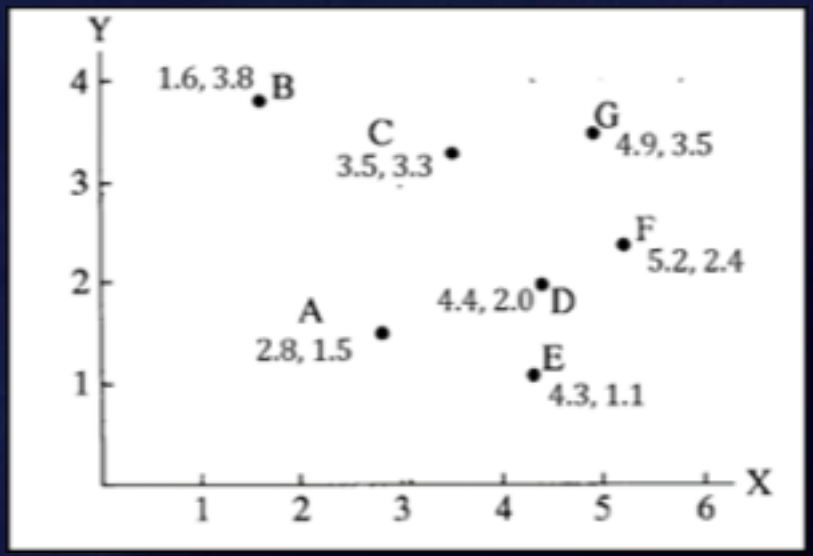


Site	X	Y	NN	NND
A				
B				
C				
D				
E				
F				
G				

Nearest Neighbor

- Step 1: Calculate the distance from each point to its nearest neighbor, by calculating the hypotenuse of the triangle

$$NND_{AB} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

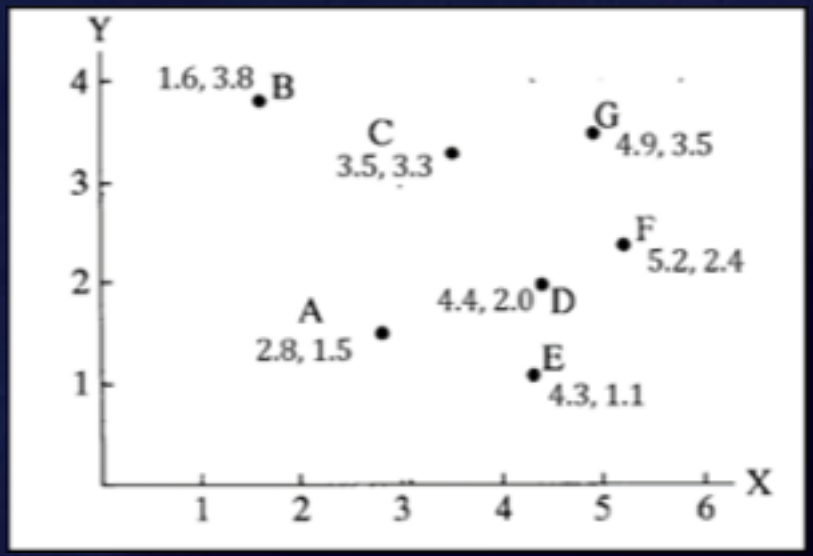


Site	X	Y	NN	NND
A	2.8	1.5	E	1.5524
B	1.6	3.8	C	1.9647
C	3.5	3.3	G	1.4142
D	4.4	2.0	F	0.8944
E	4.3	1.1	D	0.9055
F	5.2	2.4	D	0.8944
G	4.9	3.5	F	1.1402

Nearest Neighbor

- Step 1: Calculate the distance from each point to its nearest neighbor, by calculating the hypotenuse of the triangle

$$NND = \frac{1}{n}\sum_{i=1}^{n} d_i$$

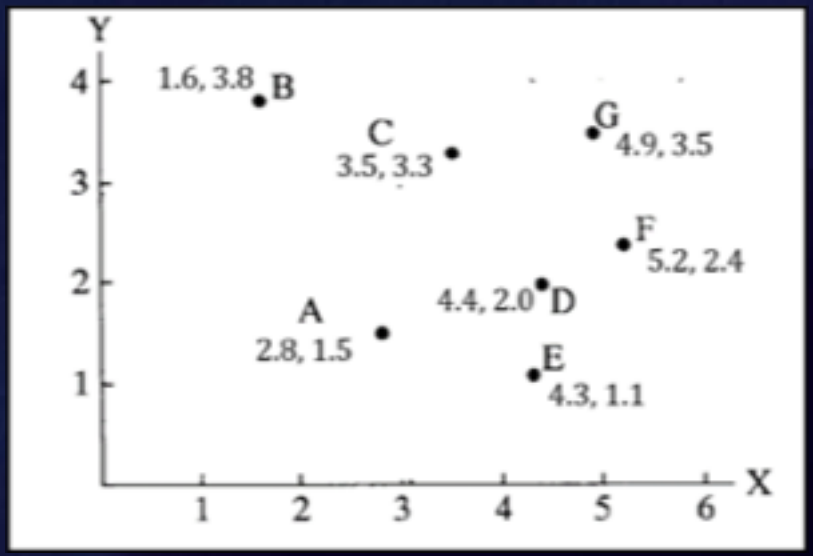


$$NND = \frac{1}{n}\sum_{i=1}^{n} NND_i = \frac{1}{7} \times 8.77 \sim 1.2523$$

Site	X	Y	NN	NND
A	2.8	1.5	E	1.5524
B	1.6	3.8	C	1.9647
C	3.5	3.3	G	1.4142
D	4.4	2.0	F	0.8944
E	4.3	1.1	D	0.9055
F	5.2	2.4	D	0.8944
G	4.9	3.5	F	1.1402

Nearest Neighbor

- Step 2: Calculate NND_R
- The average distance if the pattern were random?

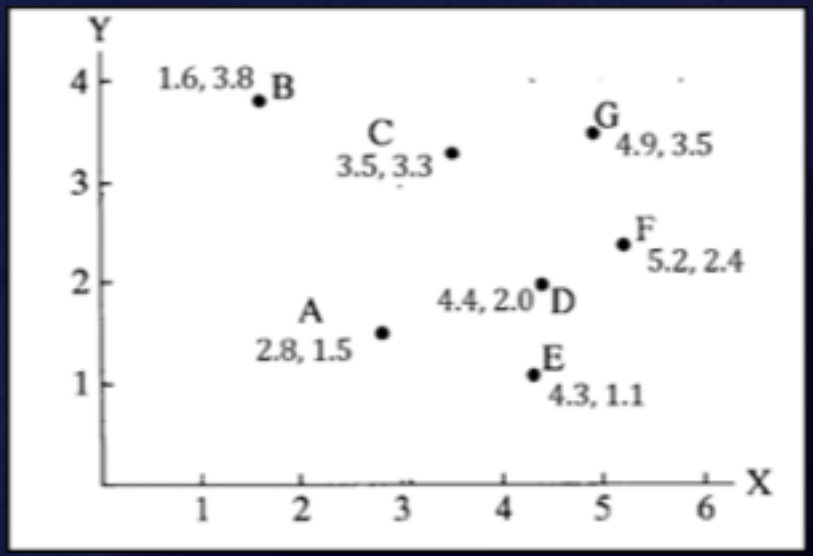


$$NND_R = 0.5 \times \frac{1}{\sqrt{\frac{n}{A}}}$$

$$NND_R = \frac{0.5}{\sqrt{\frac{n}{A}}} = \frac{0.5}{\sqrt{\frac{7}{4 \times 3}}} \sim 0.6547$$

Nearest Neighbor

- Step 3: Compute the standardized nearest neighbor index (R):



$$NND = \frac{1}{n}\sum_{i=1}^{n} NND_i = \frac{1}{7} \times 8.77 \sim 1.2523$$

$$NND_R = \frac{0.5}{\sqrt{\frac{n}{A}}} = \frac{0.5}{\sqrt{\frac{7}{12}}} \sim 0.6547$$

$$R = \frac{NND}{NND_R} \sim 1.9129$$

Nearest Neighbor

$R = 1.9129$

More dispersed than random

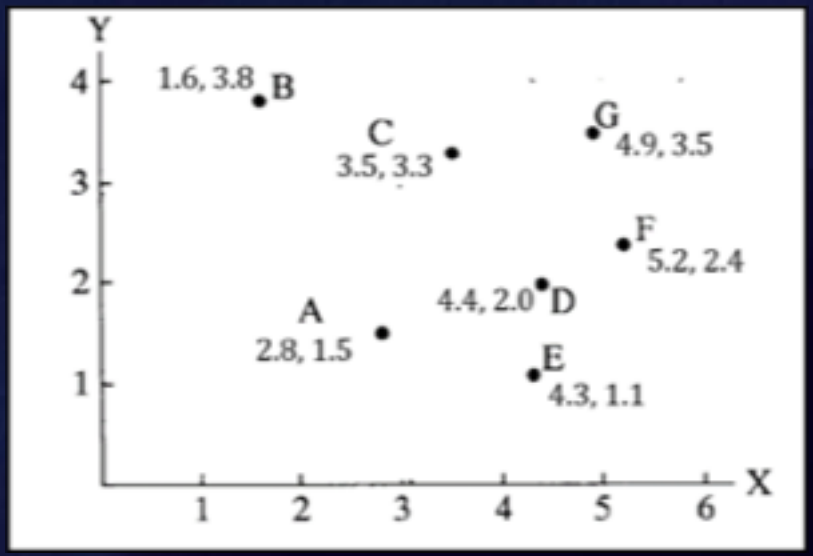
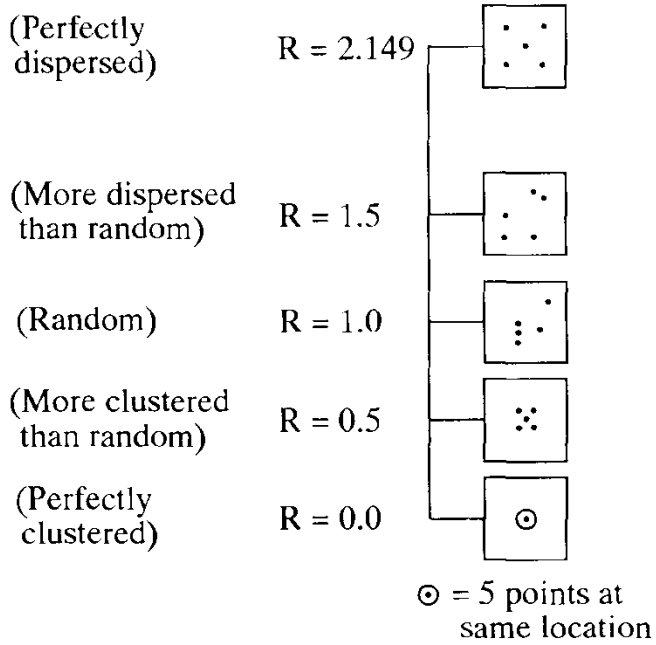


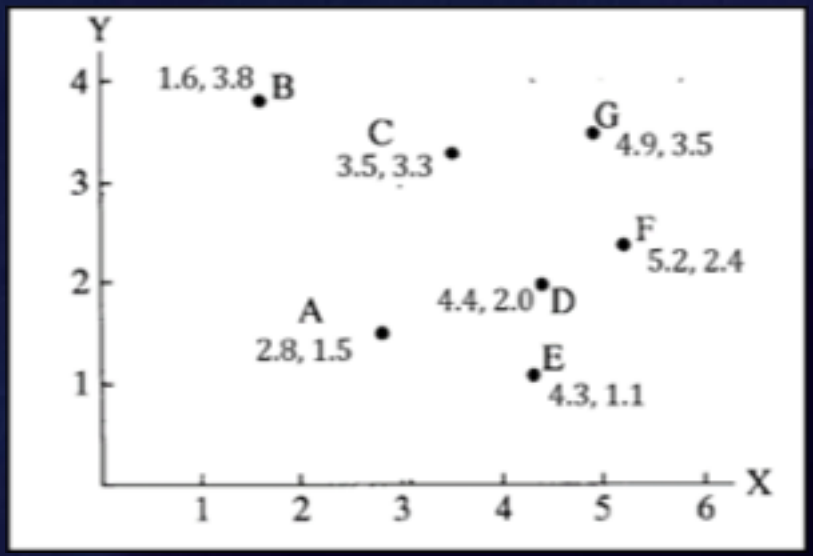
Figure 11.3 Continuum of *R* Values in Nearest Neighbor Analysis



Source: Modified from Taylor, P. J. 1977. *Quantitative Methods in Geography*. Boston: Houghton Mifflin

Nearest Neighbor

- Significance of dispersion/clustering?
  → we can compute Z!



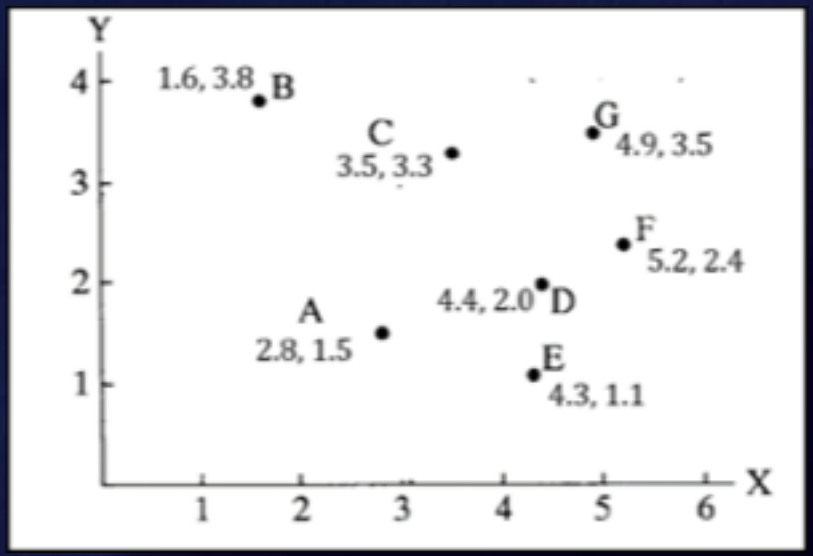
- Null hypothesis:
  Random distribution of points → H0: R = 1

$$Z = \frac{NND - NND_R}{\sigma_{NND}}$$

$$\sigma_{NND} = \frac{0.26136}{\sqrt{\frac{n^2}{A}}}$$

Nearest Neighbor

- Significance of dispersion/clustering?
  → we can compute Z!
- Null hypothesis:
  Random distribution of points → H0: R = 1



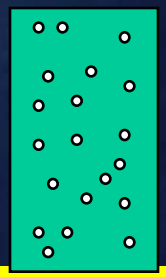
$$NND \sim 1.2523 \qquad NND_R \sim 0.6547$$

$$\sigma_{NND} = \frac{0.26136}{\sqrt{\frac{n^2}{A}}} = \frac{0.26136}{\sqrt{\frac{49}{12}}} \sim 0.1293$$

$$Z = \frac{NND - NND_R}{\sigma_{NND}} \sim \frac{0.5976}{0.1293} \sim 4.62$$

	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Adapted with rounding from Table II of Fisher and Yates 1974.

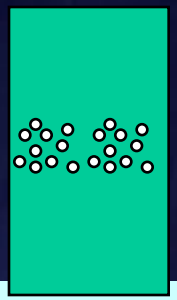


RANDOM

Z = -0.1515

Point	Nearest Neighbor	Distance
1	2	1
2	3	0.1
3	2	0.1
4	5	1
5	4	1
6	5	2
7	6	2.7
8	10	1
9	10	1
10	9	1
		10.9

Mean distance	1.09
Area of Region	50
Density	0.2
Expected Mean	1.118034
R	0.974926

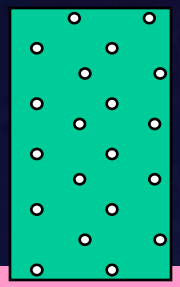


CLUSTERED

Z = 5.508

Point	Nearest Neighbor	Distance
1	2	0.1
2	3	0.1
3	2	0.1
4	5	0.1
5	4	0.1
6	5	0.1
7	6	0.1
8	9	0.1
9	10	0.1
10	9	0.1
		1

Mean distance	0.1
Area of Region	50
Density	0.2
Expected Mean	1.118034
R	0.089443



UNIFORM

Z = 5.855

Point	Nearest Neighbor	Distance
1	3	2.2
2	4	2.2
3	4	2.2
4	5	2.2
5	7	2.2
6	7	2.2
7	8	2.2
8	9	2.2
9	10	2.2
10	9	2.2
		22

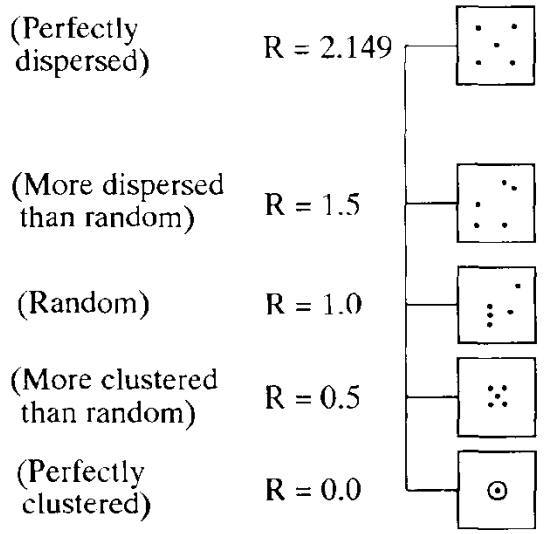
Mean distance	2.2
Area of Region	50
Density	0.2
Expected Mean	1.118034
R	1.96774

$$Z = \frac{NND - NND_R}{\sigma_{NND}}$$

Assumptions and Limitations

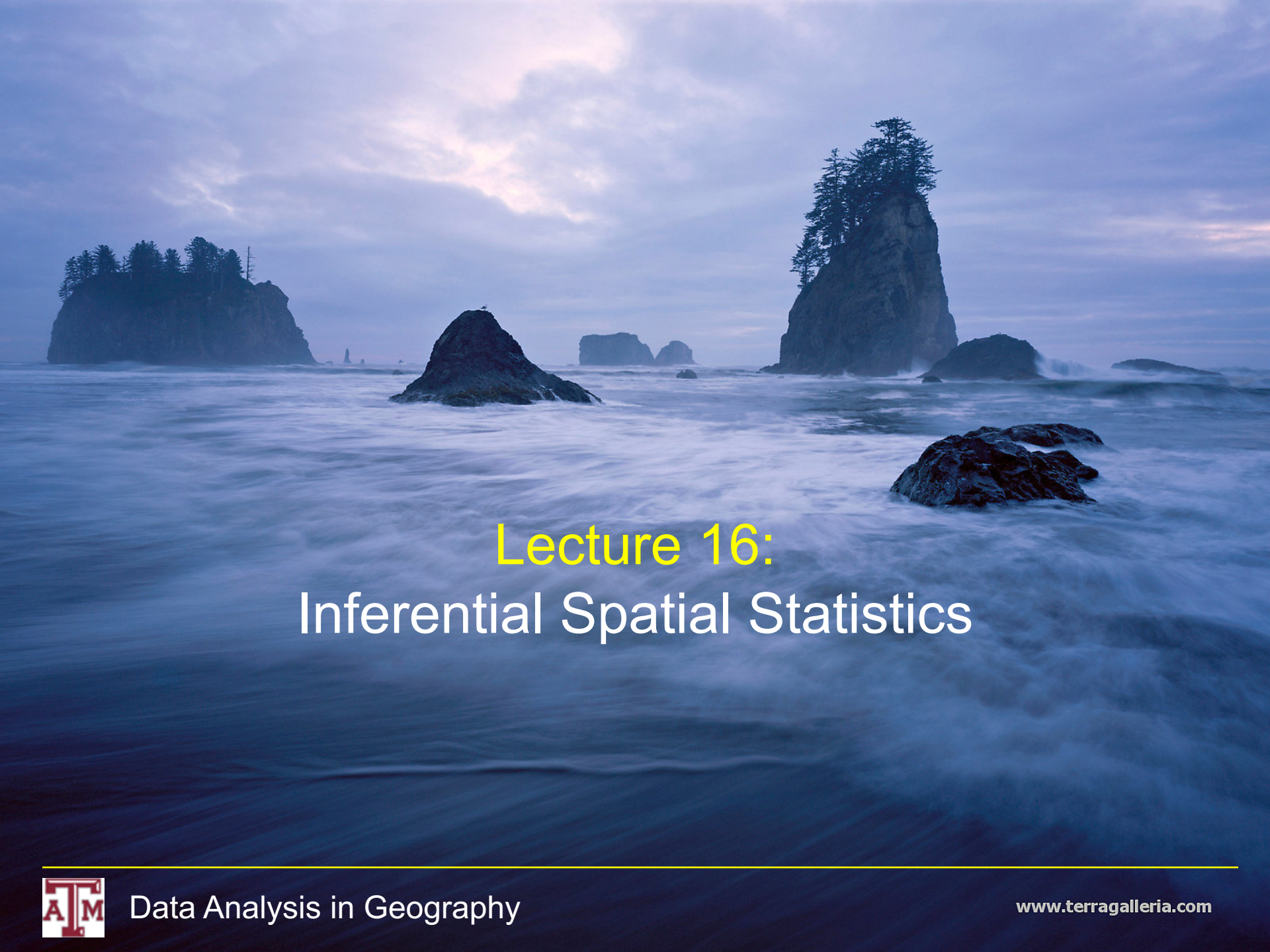
- Random Assumption: Random sample of points from a population
- Independence Assumption: Sample points are independently selected

Figure 11.3 Continuum of *R* Values in Nearest Neighbor Analysis



Source: Modified from Taylor, P. J. 1977. *Quantitative Methods in Geography*. Boston: Houghton Mifflin

Let's Open Up R…



Lecture 16:
Inferential Spatial Statistics