



CentroGeo
19°17'30"N 99°13'17"O 2489m

Análisis geoespacial con R

Introducción a la toma de decisiones basada en datos

Ana J. Alegre, Cristian Silva. Noviembre 25, 2021.

Agenda

Taller - Día 2

1. Preguntas y respuestas
2. Análisis exploratorio de datos (EDA)
3. Sesión de práctica
4. Descanso
5. Análisis de Patrones Puntuales
6. Sesión de práctica
7. Preguntas y respuestas



CentroGeo

19°17'30"N 99°13'17"O 2489m

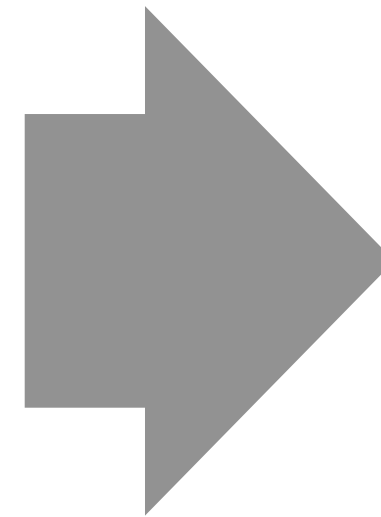
Análisis exploratorio de datos (EDA)

Análisis exploratorio de datos (EDA)

Usando medidas estadísticas para analizar distribuciones de datos

Estadísticas descriptivas:

- Medidas de tendencia central:
 - Moda
 - Mediana
 - Media
- Medidas de dispersión:
 - Rango
 - Desviación estándar
 - Varianza



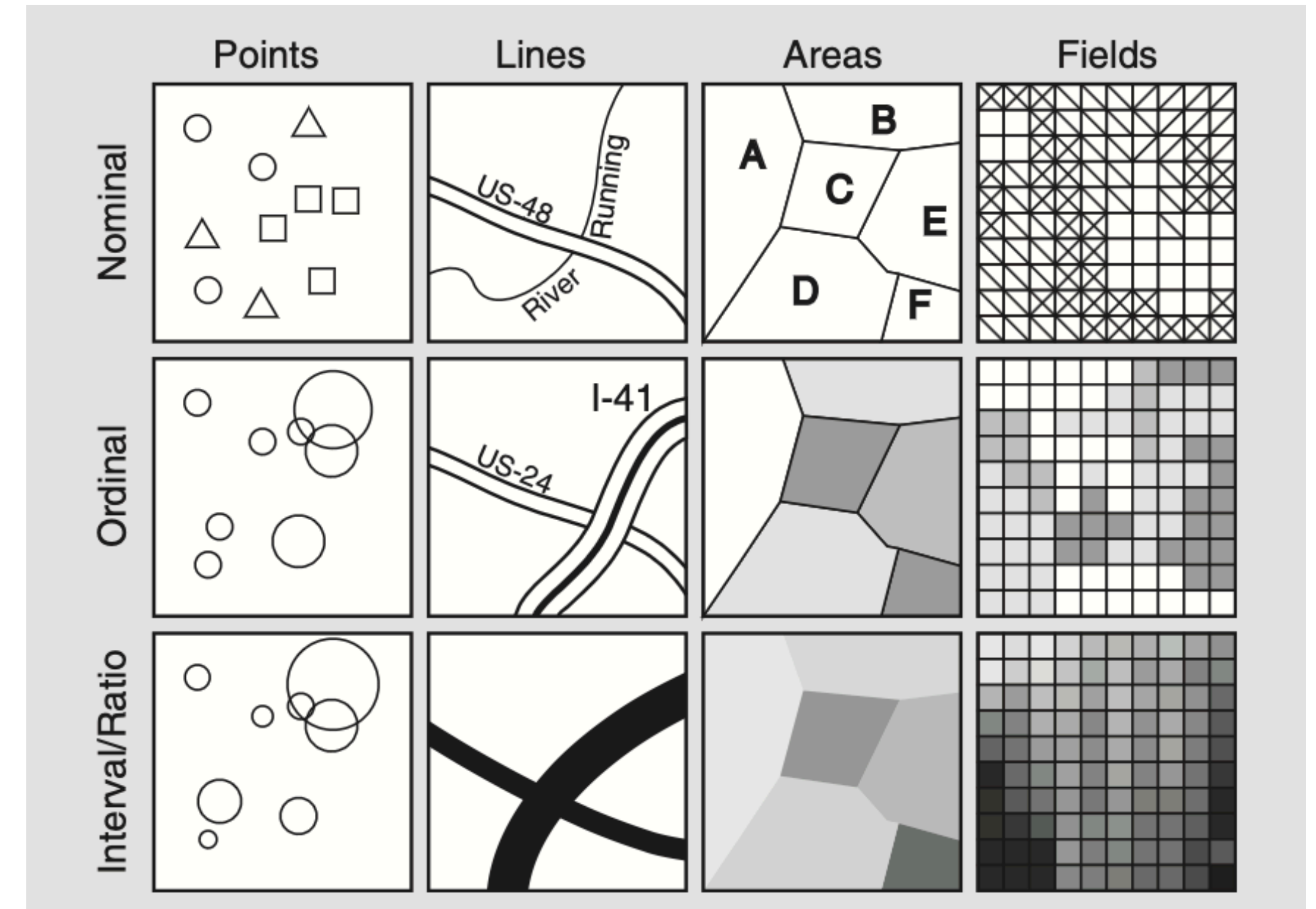
Estadísticas espaciales:

- Medidas espaciales de tendencia central:
 - Centro medio/mediano espacial
 - Centro medio/mediano espacial ponderado
- Medidas espaciales de dispersión:
 - Distancia estándar
 - Elipse de desviación estándar
 - Varianza

Tipos de datos espaciales

Geometrías usadas en los SIG

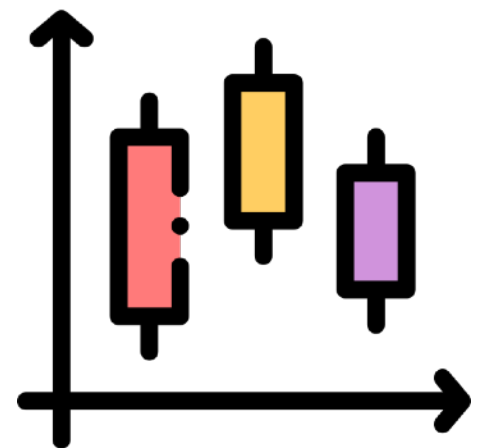
1. **Vista de vector.** Registra las coordenadas (x, y) de ubicación de las características que conforman un mapa.
 - Punto
 - Línea
 - Área
2. **Sistemas *ráster*.** En lugar de empezar con objetos en el suelo, se define una cuadrícula de pequeñas unidades llamadas píxeles en la superficie de la Tierra.



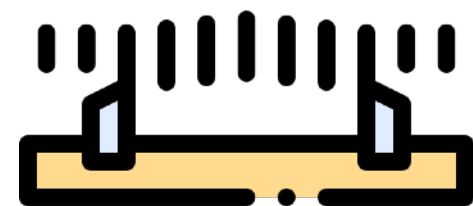
Una representación esquemática de los tipos de datos espaciales entidad-atributo.

Análisis exploratorio de datos (EDA)

Usando medidas estadísticas para analizar distribuciones de datos



1. Identificar los valores más grandes y pequeños.



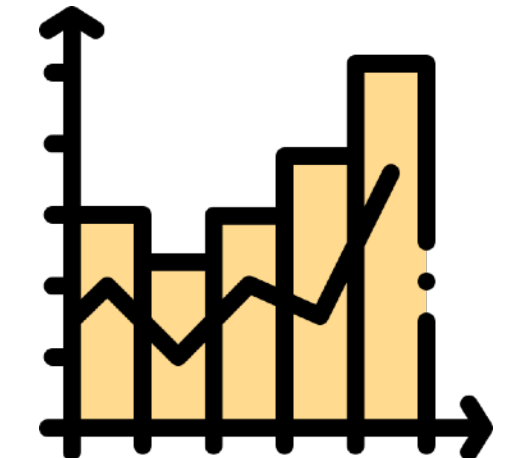
2. Identificar el rango.



3. Determinar el número de clases.

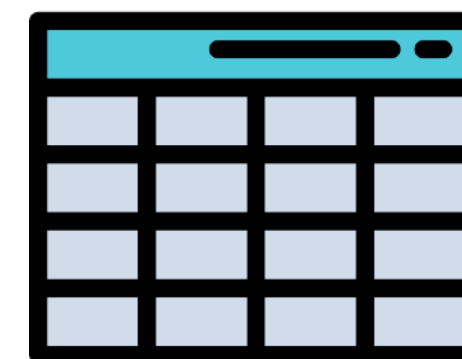


4. Definir el intervalo de clase.



5. Determinar la frecuencia de cada clase

6. Compilar la información en una tabla



Sesión de práctica

Descanso



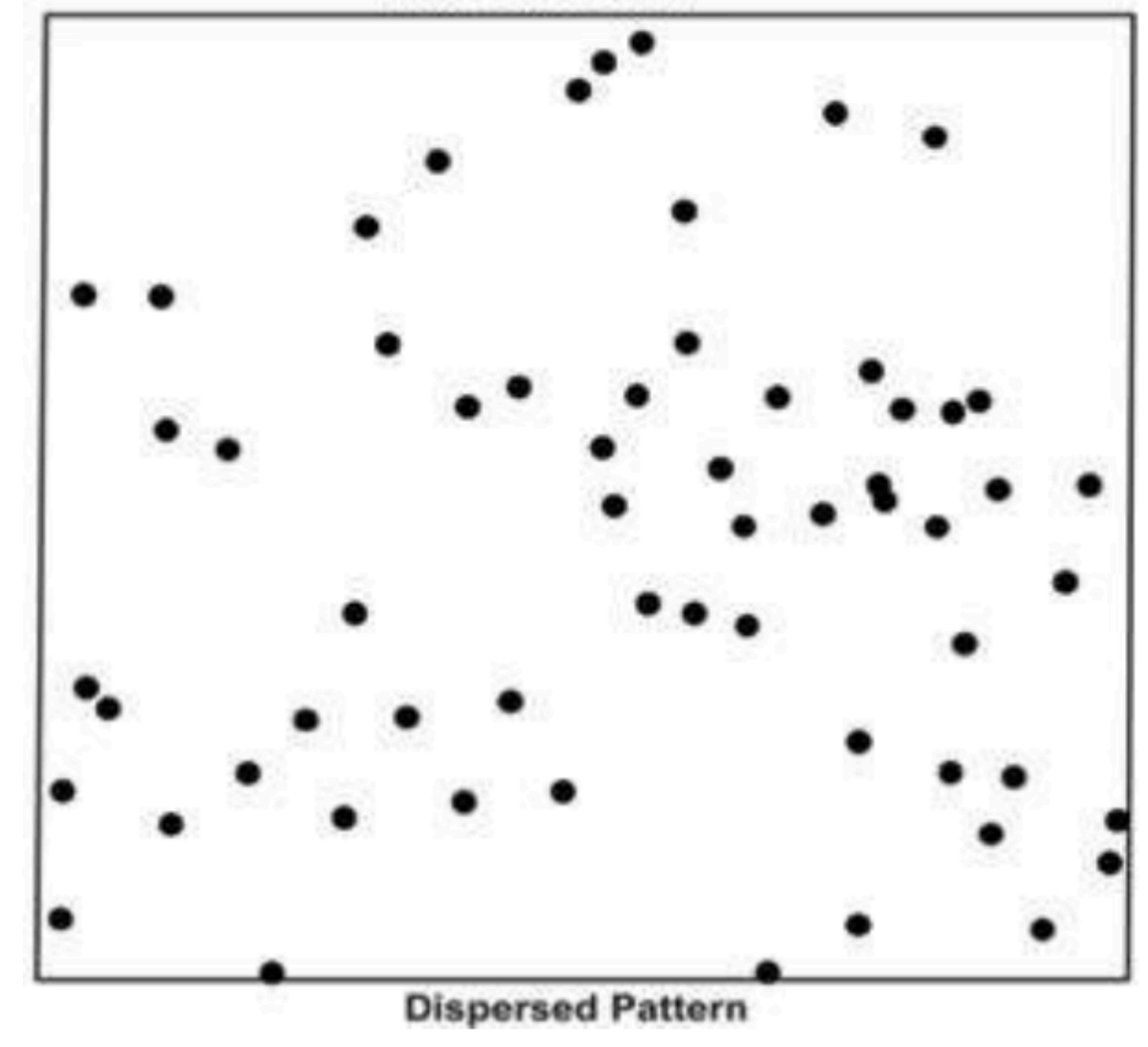
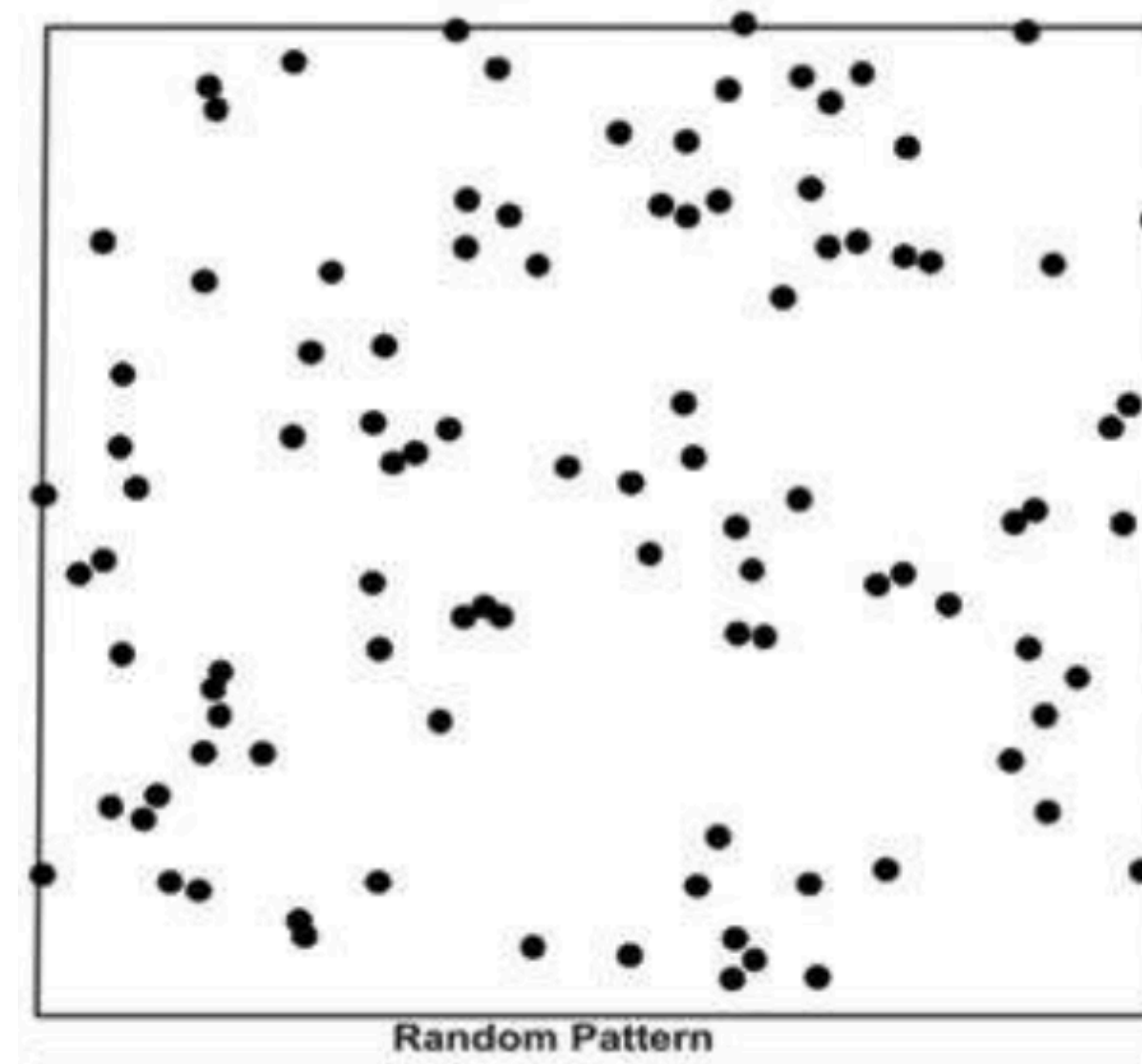
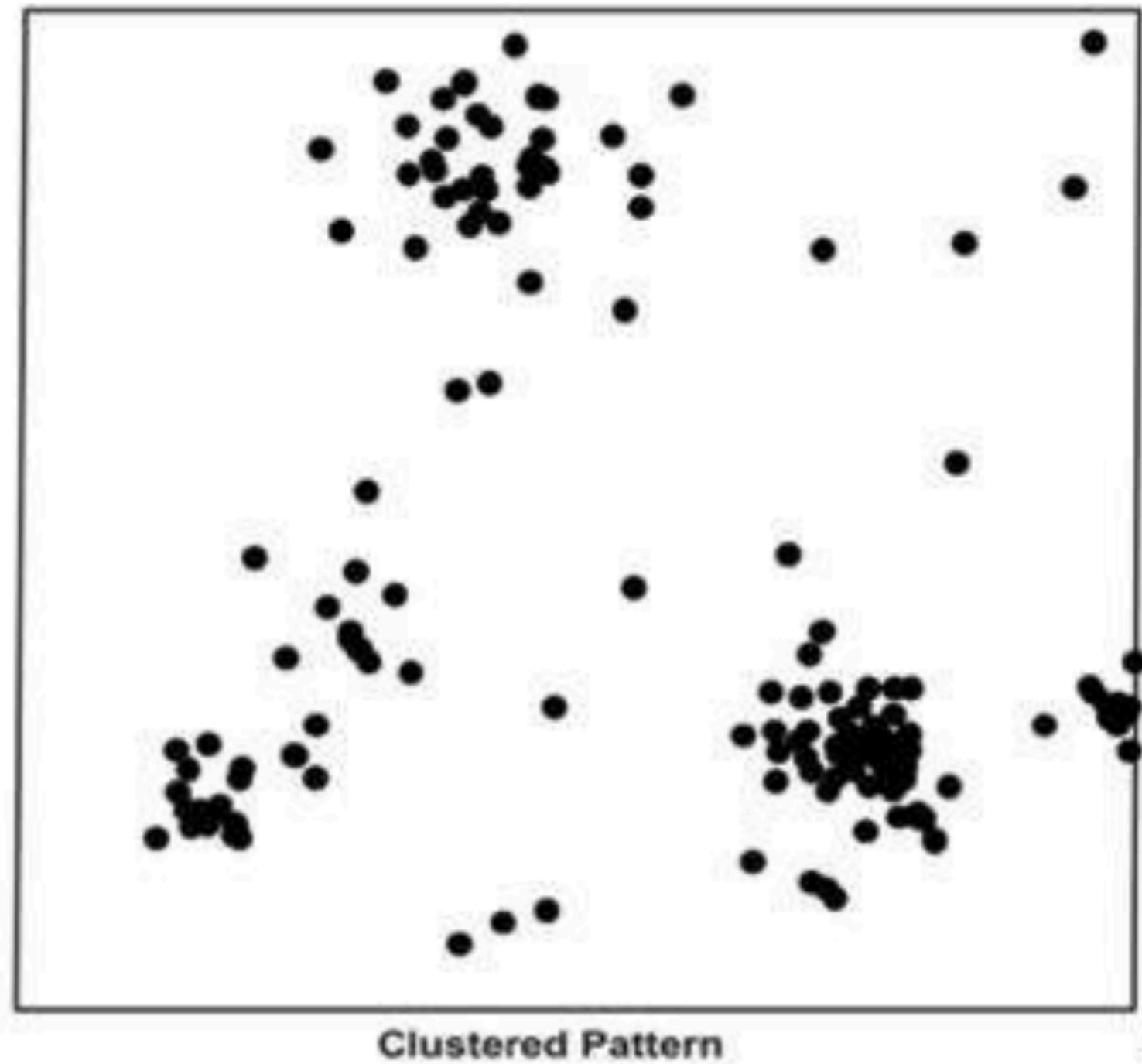
CentroGeo

19°17'30"N 99°13'17"O 2489m

Análisis de Patrones Puntuales

Análisis de patrones puntuales

Las tres estructuras básicas que existen



Análisis de patrones puntuales

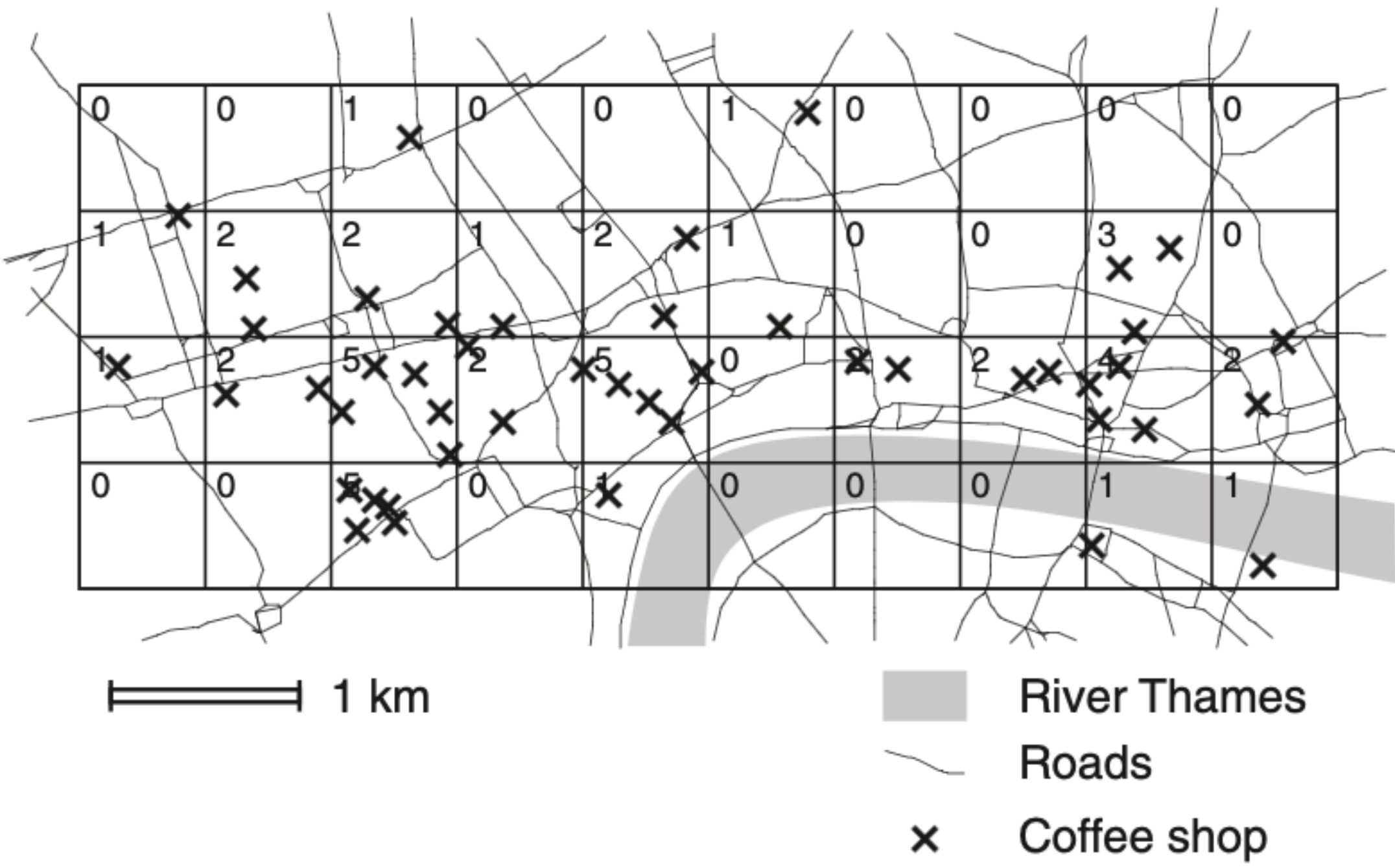
Explorando patrones, distribuciones y tendencias

1. Quadrat Count Conteo por cuadrantes
2. Aproximación por vecinos más cercanos
3. Aproximación por función K
4. Aproximación por estimación de *Kernel*

Análisis de patrones puntuales

Conteo por cuadrante

Ejemplo: Cafeterías en el centro de Londres



Conteos por cuadrante y cálculos de la varianza para el patrón de cafeterías

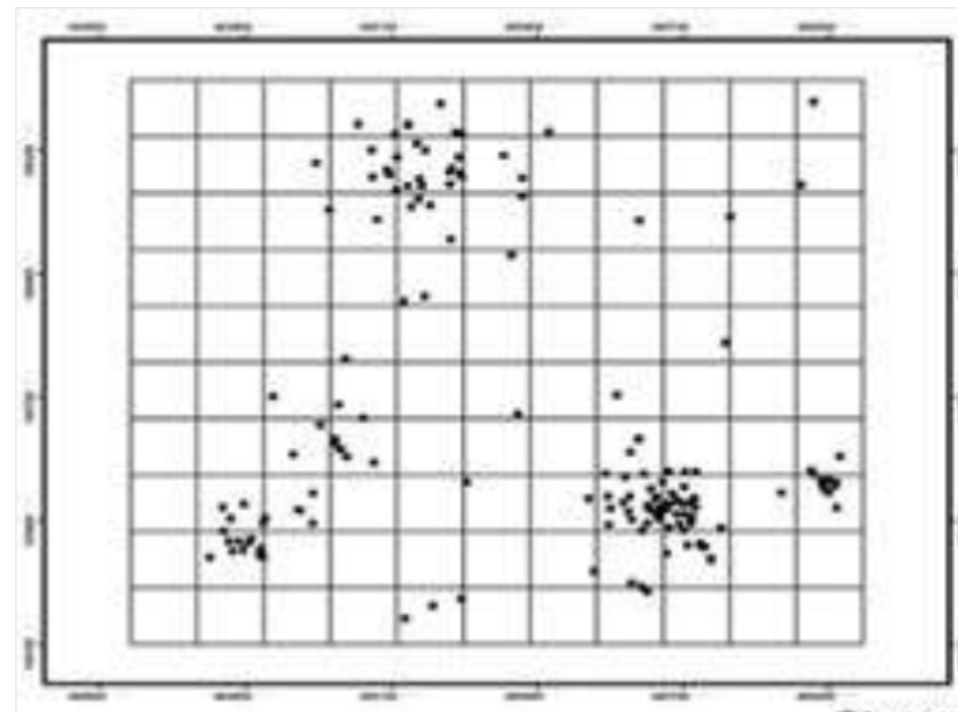
No. of events, K	No. of quadrats, X	$K - \mu$	$(K - \mu)^2$	$X(K - \mu)^2$
0	18	-1.175	1.380625	24.851250
1	9	-0.175	0.030625	0.275625
2	8	0.825	0.680625	5.445000
3	1	1.825	3.330625	3.330625
4	1	2.825	7.980625	7.980625
5	3	3.825	14.630625	43.891875
Totals	40			85.775000

Referencia tomada de O'sullivan, D., & Unwin, D. (2003). *Geographic information analysis*. John Wiley & Sons and, Oyana, T. J. (2020). *Spatial Analysis with R: Statistics, Visualization, and Computational Methods*. CRC press.

Análisis de patrones puntuales

Conteo por cuadrantes

Patrón de cluster

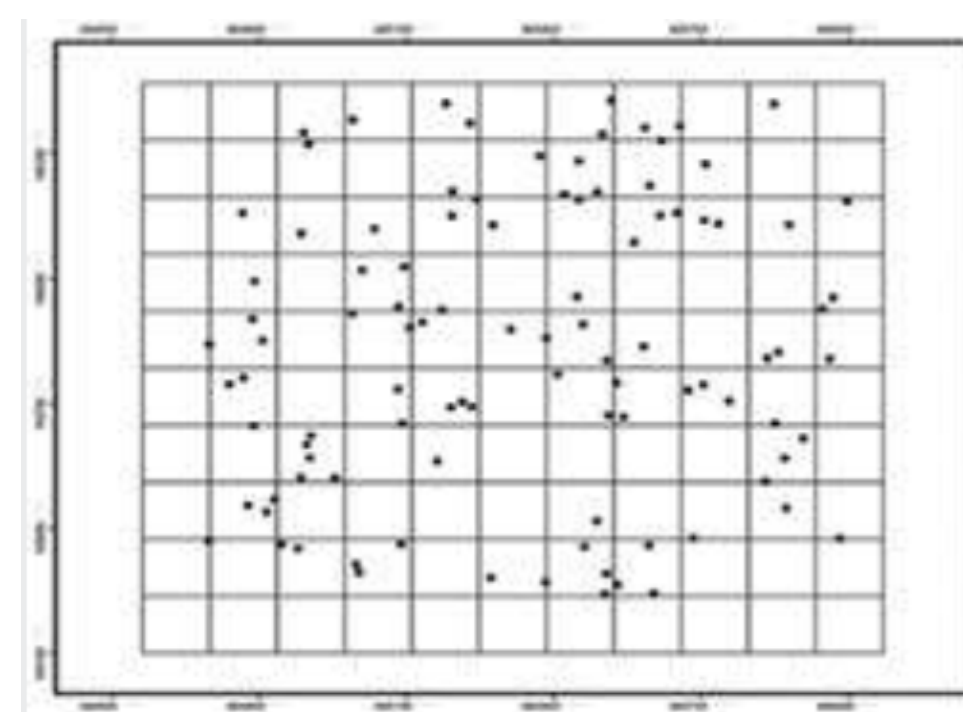


$$\bar{X} = \frac{\sum X_j f_j}{\sum f_j} = \frac{232.2}{110} = 2.1109$$

$$\text{Chi-square statistic } \chi^2 = 2046.77/2.1109 = 969.61$$

$$P\text{-value} = 4.8248\text{E-}138$$

Patrón aleatorio

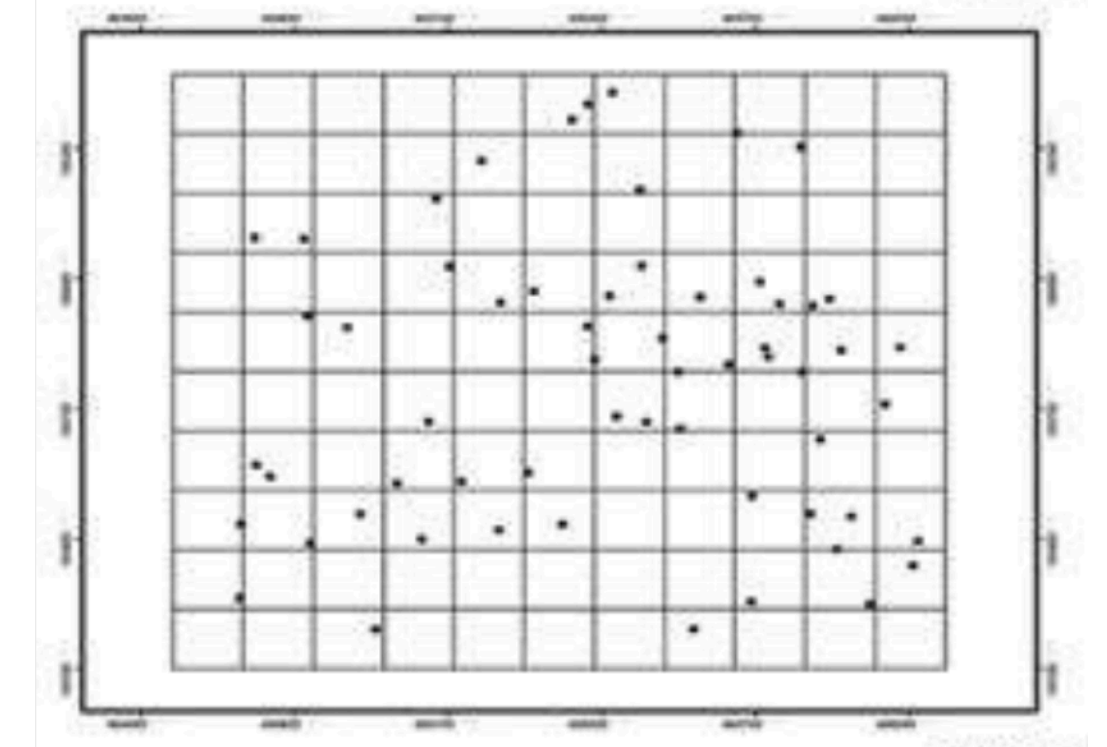


$$\bar{X} = \frac{\sum X_j f_j}{\sum f_j} = \frac{99}{110} = 0.90$$

$$\text{Chi-square statistic } \chi^2 = 119.90/0.9 = 133.22$$

$$P\text{-value} = 0.057389389$$

Patrón disperso



$$\bar{X} = \frac{\sum X_j f_j}{\sum f_j} = \frac{68.2}{110} = 0.62$$

$$\text{Chi-square statistic } \chi^2 = 62.94/0.62 = 101.51$$

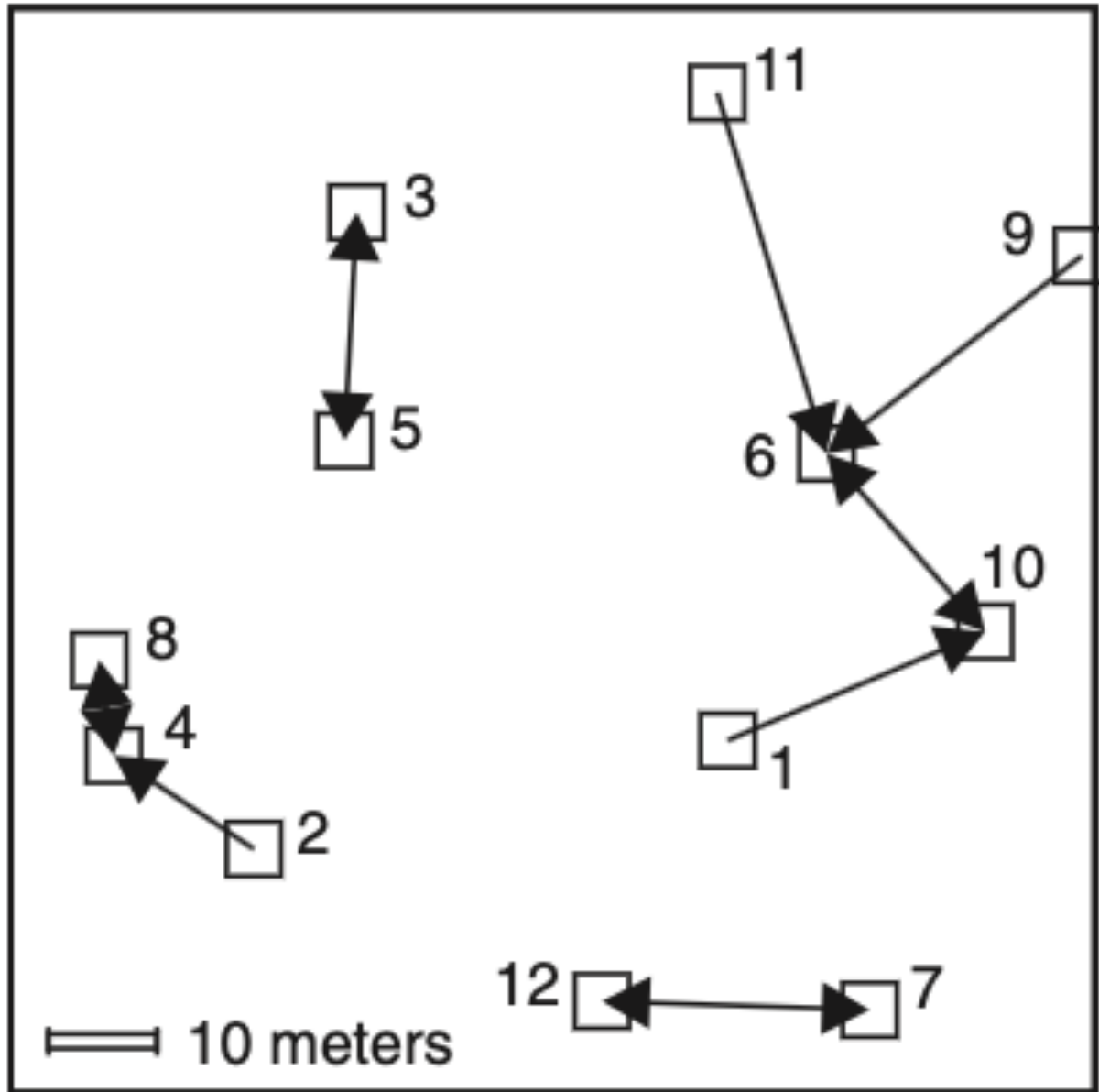
$$P\text{-value} = 0.682243941$$

Referencia tomada de O'sullivan, D., & Unwin, D. (2003). *Geographic information analysis*. John Wiley & Sons and, Oyana, T. J. (2020). *Spatial Analysis with R: Statistics, Visualization, and Computational Methods*. CRC press.

Análisis de patrones puntuales

Aproximación por vecinos más cercanos

Ejemplo: Distancia a los vecinos más cercanos para un pequeño patrón puntual



Cálculos para la distancia de vecinos más cercanos para el patrón puntual

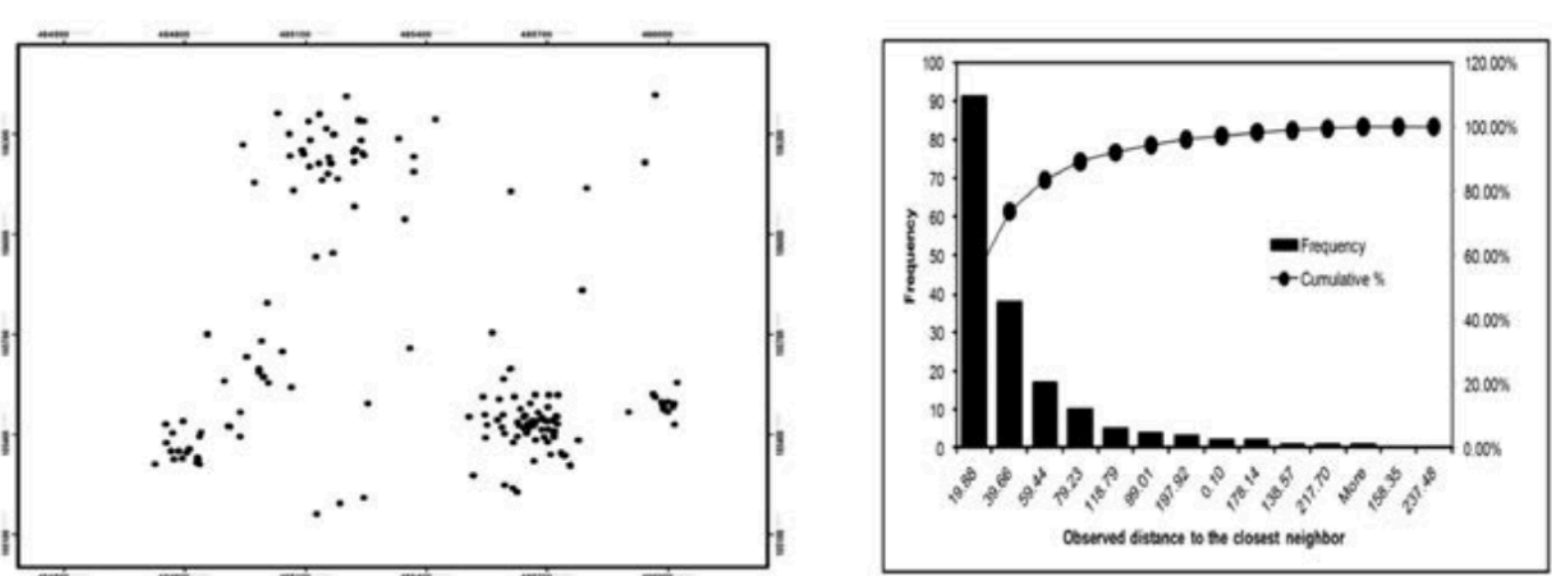
Point	X	Y	Nearest neighbor	D_{\min}
1	66.22	32.54	10	25.59
2	22.52	22.39	4	15.64
3	31.01	81.21	5	21.11
4	9.47	31.02	8	9.00
5	30.78	60.10	3	21.14
6	75.21	58.93	10	21.94
7	79.26	7.68	12	24.81
8	8.23	39.93	4	9.00
9	98.73	77.17	6	29.76
10	89.78	42.53	6	21.94
11	65.19	92.08	6	34.63
12	54.46	8.48	7	24.81

Referencia tomada de O'sullivan, D., & Unwin, D. (2003). *Geographic information analysis*. John Wiley & Sons and, Oyana, T. J. (2020). *Spatial Analysis with R: Statistics, Visualization, and Computational Methods*. CRC press.

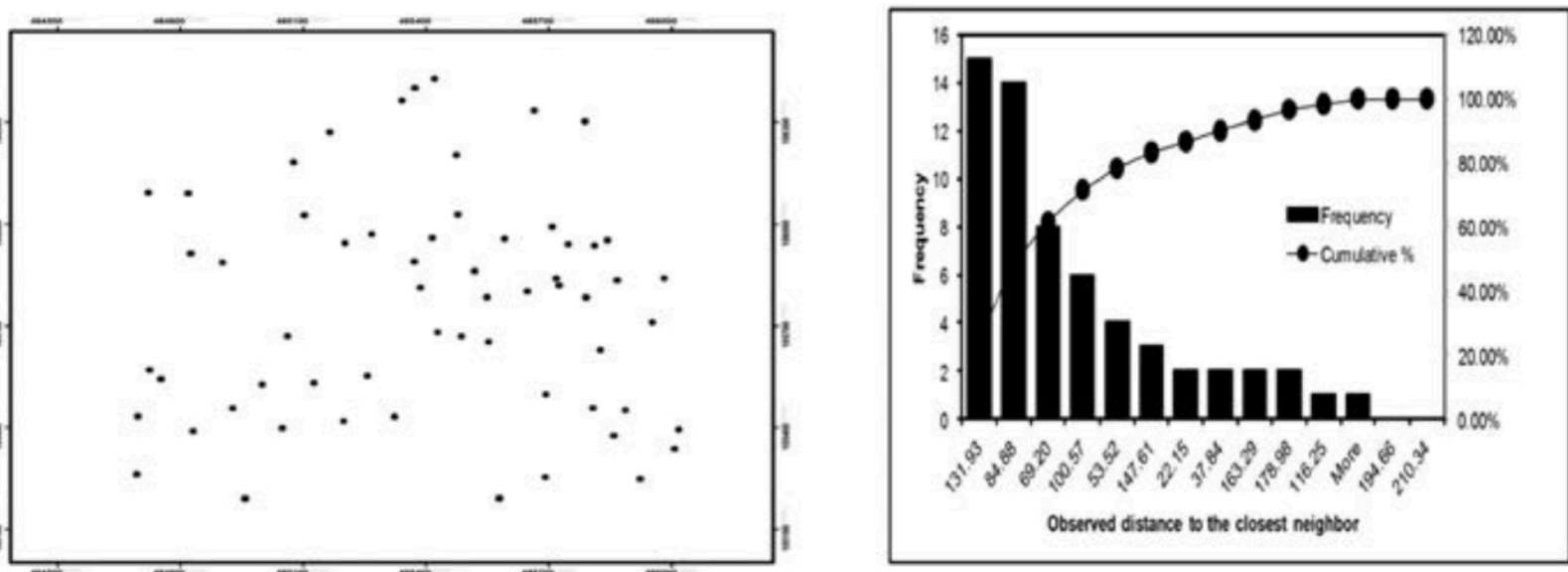
Análisis de patrones puntuales

Aproximación por vecinos más cercanos

Patrón de clusters



Patrón aleatorio



Patrón dispersado

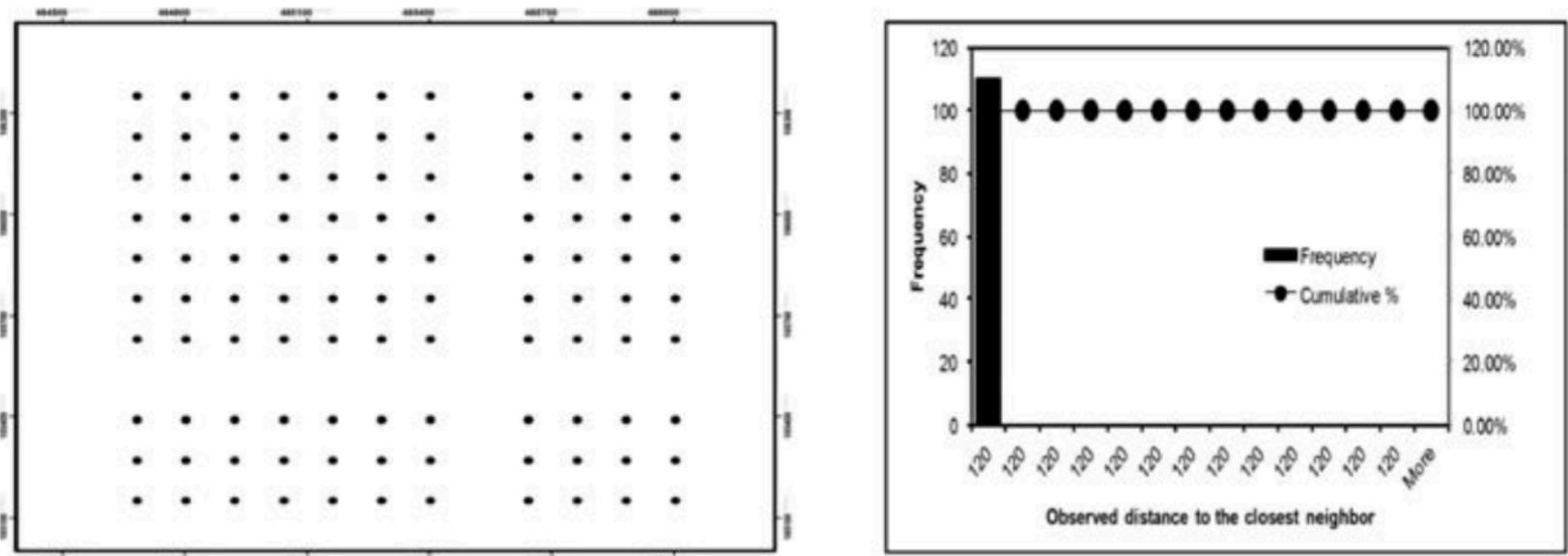


Tabla de análisis de vecinos más cercanos para sitios de anidamiento potencial para las tres distribuciones básicas.

	Observed Mean Distance	Expected Mean Distance	Nearest Neighbor Ratio (R)	z-score	p-value
Clustered	31.671	54.30851	0.58316	−10.549206	0.00000
Dispersed	120	68.5	1.751825	15.08495	0.00000
Random	95.231	92.749438	1.026753	0.396444	0.691778

Referencia tomada de O'sullivan, D., & Unwin, D. (2003). *Geographic information analysis*. John Wiley & Sons and, Oyana, T. J. (2020). *Spatial Analysis with R: Statistics, Visualization, and Computational Methods*. CRC press.

Análisis de patrones puntuales

Aproximación por función K

Hay seis pasos principales para calcular la función K :

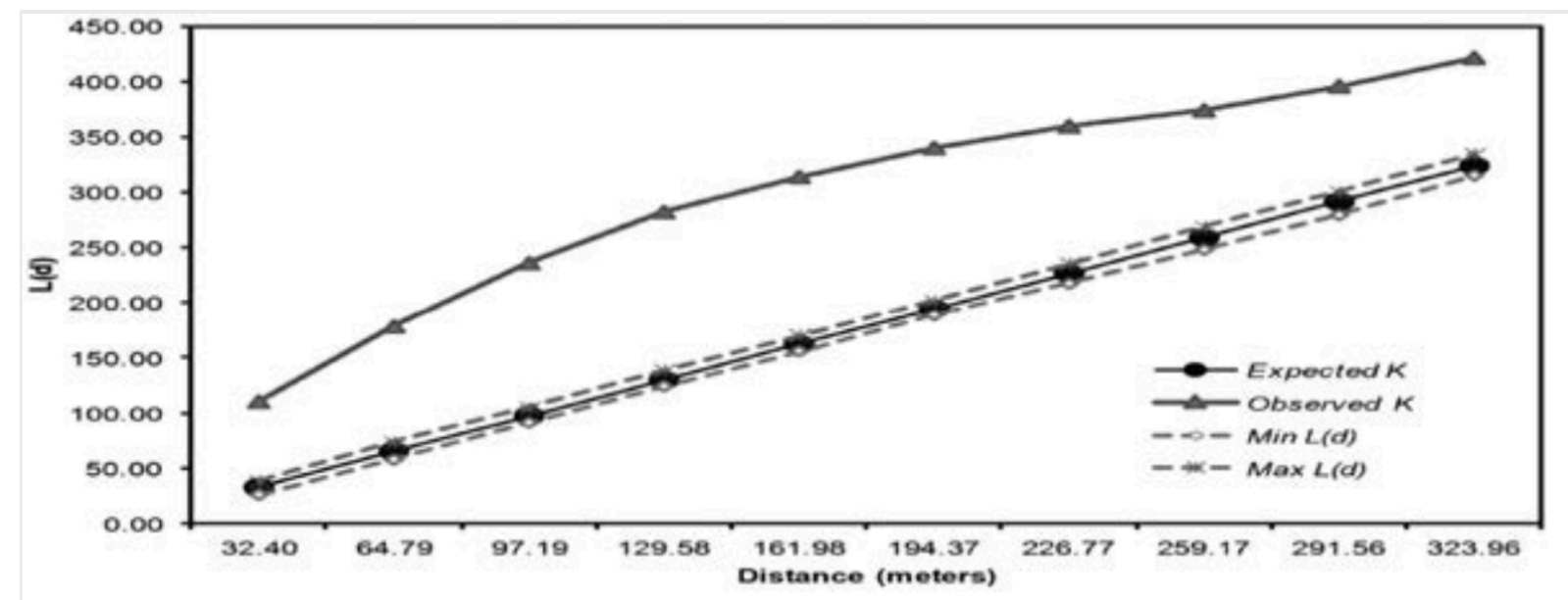
1. Determinar/comparar la K observada y esperada. La K observada se obtiene mediante la construcción de un círculo alrededor de cada evento puntual (i), contando el número de otros eventos (j) dentro del radio (h) del círculo y repitiendo el mismo proceso para todos los demás eventos (i).
2. A continuación, determinar el número promedio de eventos dentro de las bandas de distancia sucesivas. Encontrar la densidad de todos los puntos para el área de estudio. La K observada es la tasa del numerador entre la densidad de los eventos. Entonces se puede comparar con la K esperada, que es un patrón aleatorio, $K(h) = \pi h^2$.
3. Transformar las estimaciones de $K(h)$ en una función de raíz cuadrada para hacerla lineal $L(d)$.
4. Determinar el intervalo de confianza estimando los valores mínimos y máximos de $L(d)$ de varias simulaciones con $\alpha = 0.05$ bajo la hipótesis nula de la distribución aleatoria.
5. Graficar los estimados $L(d)$ en una gráfica para revelar si ocurre algún agrupamiento a ciertas distancias.
6. Interpretar los resultados.

Referencia tomada de O'sullivan, D., & Unwin, D. (2003). *Geographic information analysis*. John Wiley & Sons and, Oyana, T. J. (2020). *Spatial Analysis with R: Statistics, Visualization, and Computational Methods*. CRC press.

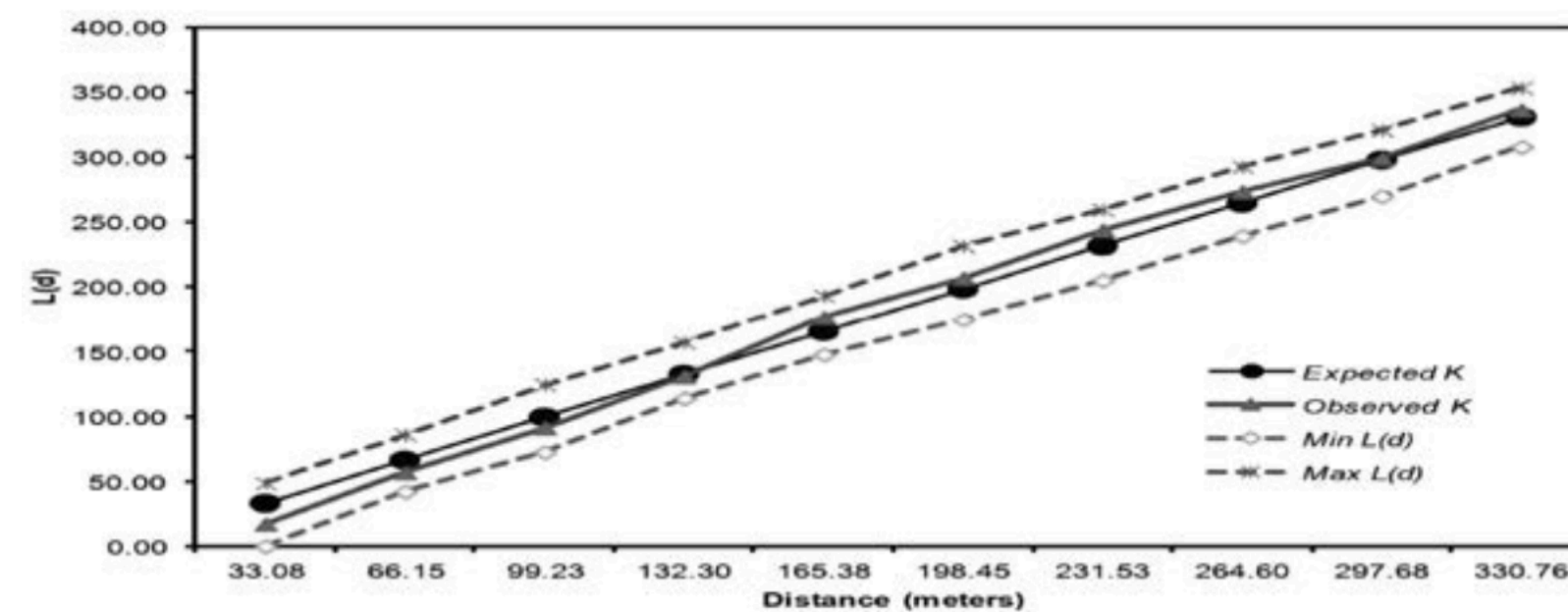
Análisis de patrones puntuales

Aproximación por función K

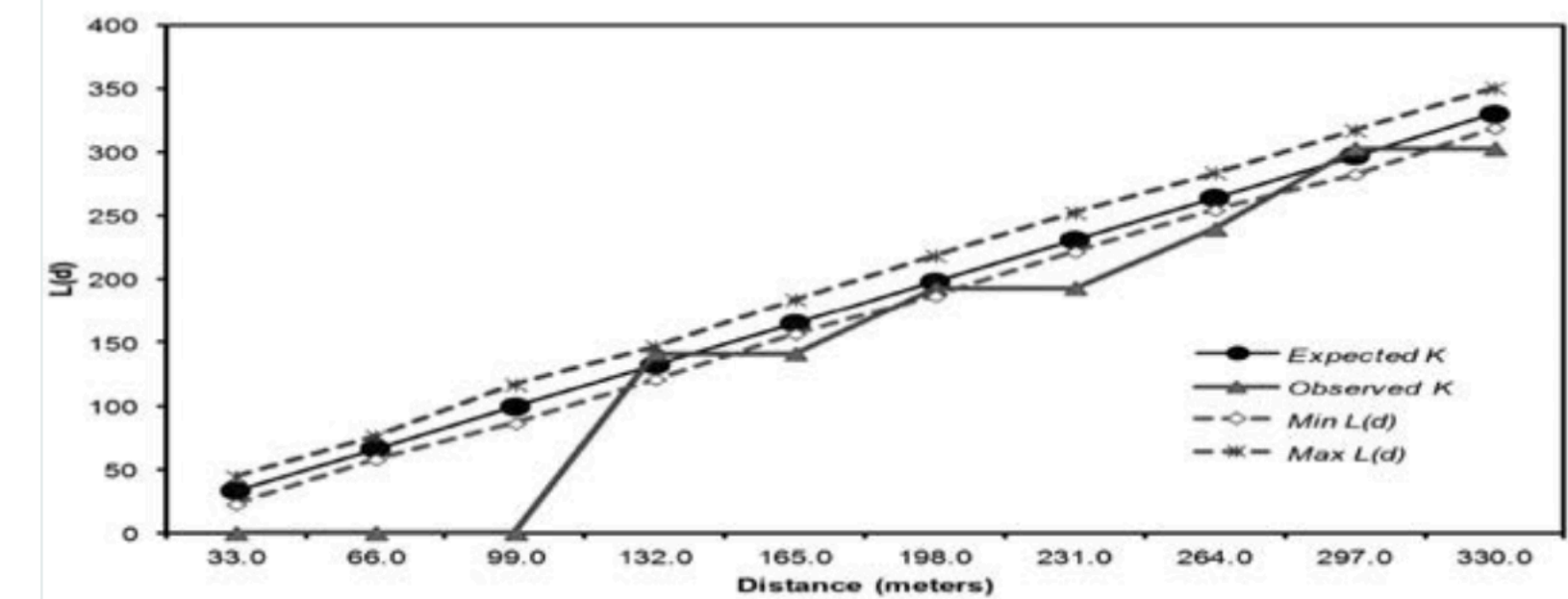
Distribución de clusters



Distribución aleatoria



Distribución dispersa

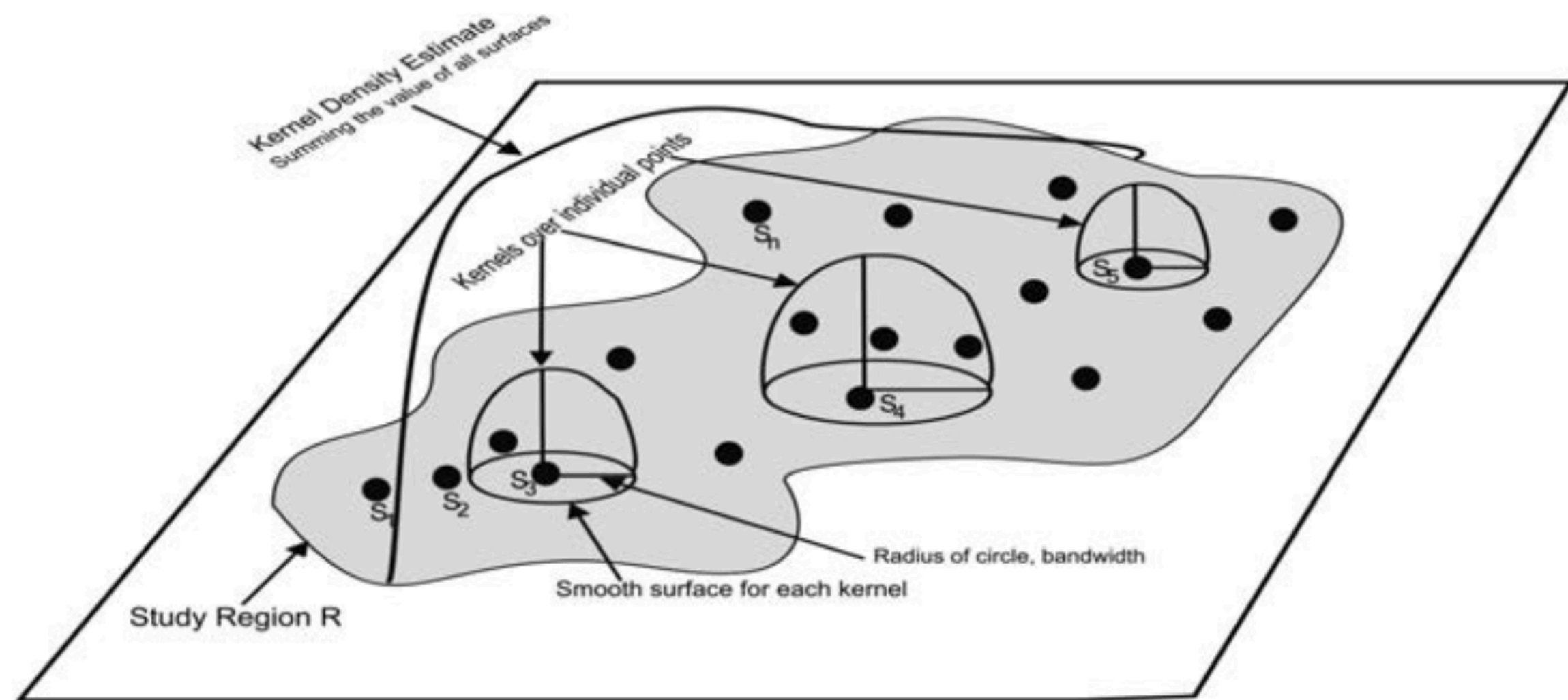


Las gráficas de los valores $L(d)$ para los tres patrones de dispersión de un estudio ecológico obtenido del análisis de la función K . Los hallazgos fueron generados con base en 99 simulaciones bajo la hipótesis nula de la distribución aleatoria.

Análisis de patrones puntuales

Aproximación por estimación de *kernel*

Método de aproximación por estimación de kernel
aplicado para estudiar la región R



$$\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right)$$

τ es el ancho de banda (un parámetro de suavizado, por ejemplo, el radio del círculo)

$k(\cdot)$ es el *kernel*

$s - s_i$ es la distancia entre dos eventos (puntos s y s_i)

Sesión de práctica

Nos vemos mañana. Gracias!