

Análise Espacial de Dados

Aula teórica 2

Análise espacial de um padrão pontual Spatial analysis of a point pattern

24 Fevereiro de 2021

1

Sumário / Summary

- ❑ **Caracterização** (O que é um padrão pontual? Exemplos)
- ❑ **Objetivos da Análise da Distribuição de um padrão Pontual**
- ❑ **Componentes da ADP**
 - ❑ **Visualização** do padrão pontual (estatísticas básicas)
 - ❑ **Análise exploratória:** Técnicas exploratórias para análise das propriedades de 1ª e 2ª ordem de um padrão pontual;
 - ❑ Método de contagem por quadrantes;
 - ❑ Estimador de densidade ("*Kernel Estimation*")
 - ❑ Métodos baseados em distâncias entre eventos (Funções G, F e K)
 - ❑ Problemas de fronteira;

1. Characterization (What is a point pattern? Examples).

2. Objectives of the Point Pattern Analysis (PPA).

3. Components of PPA.

Visualization of the point pattern (spatial statistical parameters).

Exploratory analysis (Exploratory techniques for the analysis of 1st and 2nd order properties of a point pattern);

Quadrant counting method;

Kernel Estimation

Methods based on distances (Functions G, F, and K)

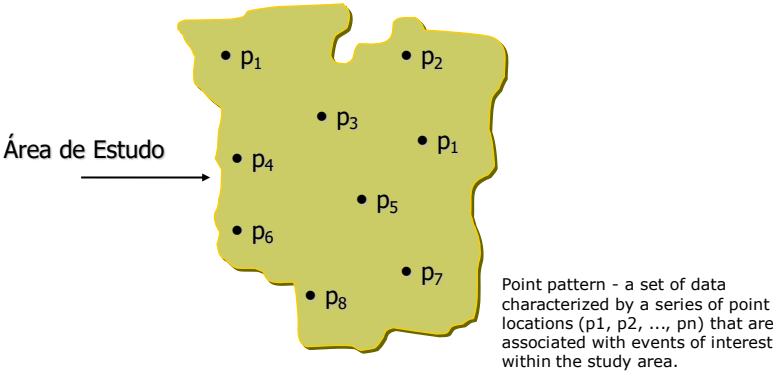
Border issues;

2

Padrão pontual

Point Pattern

- ❑ **Padrão pontual** - conjunto de dados caracterizados por uma série de localizações pontuais (p_1, p_2, \dots, p_n) que estão associados a eventos de interesse dentro da área de estudo.



3

Características de um Padrão Pontual

Characteristics of a Point Pattern

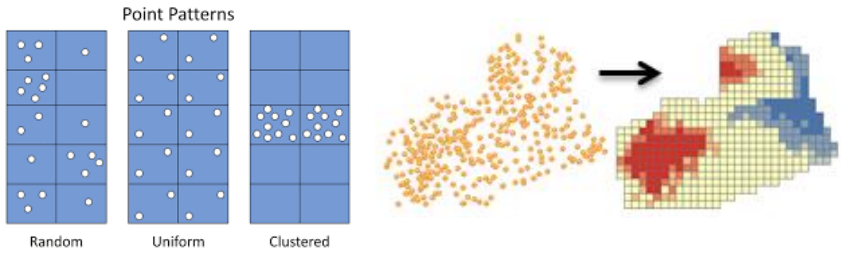
- ❑ Atributos de interesse: **localização geográfica** do evento e/ou **localização geográfica + atributo** (processo pontual marcado) ;
 - ❑ Outra medida de interesse na análise da distribuição de pontos é o nº de ocorrências (**frequência**) na área de estudo;
 - ❑ Todas as entidades representadas por pontos na área de estudo **têm a mesma qualidade**;
 - ❑ Tem de existir uma **correspondência de um-para-um** entre os objectos na área de estudo e os eventos do padrão em análise;
-
- ❑ *Attributes of interest: geographical location of the event and / or geographical location + attribute (marked point process);*
 - ❑ *Another measure of interest in the analysis of point distribution is the number of occurrences (frequency) in the study area;*
 - ❑ *The events are punctual and therefore have no associated area although in many cases these occupy space (e.g. Cities).*
 - ❑ *All entities represented by points in the study area have the same quality;*
 - ❑ *There must be a one-to-one correspondence between the objects in the study area and the events of the pattern under analysis;*

4

Análise Espacial de Padrão Pontual

Spatial analysis of a point pattern

- O objecto de interesse é a distribuição da **localização espacial** dos eventos em estudo.
- Pretende-se estudar se a distribuição espacial destes eventos apresentam um padrão específico (**aleatório**, **aglomerado (cluster)**, ou **regular (uniforme)**) tentando assim testando hipóteses de distribuição.



The object of interest is the **spatial location** of the events being studied. We intend to study the spatial distribution of these points, testing hypotheses about the pattern observed: if it is **random**, **cluster**, or is **regularly distributed**.

5

Objectivos da análise de PP

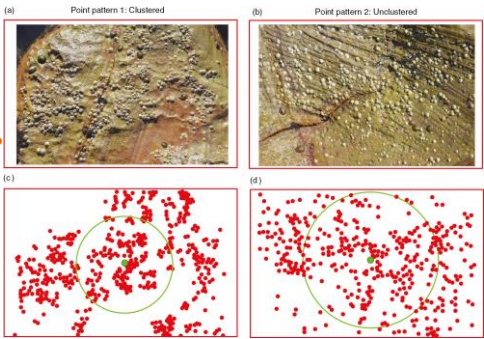
Objectives of the PPA

- Pretende-se **identificar os factores que controlam espacialmente (e temporalmente) os padrões pontuais**;

Que factores determinam este padrão espacial?

- Pretende-se **compreender o processo geográfico que controla o padrão observado**;

Como/Porquê se distribui espacialmente este padrão desta forma?



Source: <https://www.cambridge.org/core/books/spatial-analysis-of-coastal-environments/basic-geographical-analysis-with-spatial-information-in-coastal-environments/4F593538E9F21DB93038EA45E16900CA>

Because we want to identify the factors that spatially (and temporally) control the point patterns;

What factors determine this spatial pattern?

Because we want to understand the geographical process that controls the observed pattern;

How / Why is this spatially distributed this way?

6

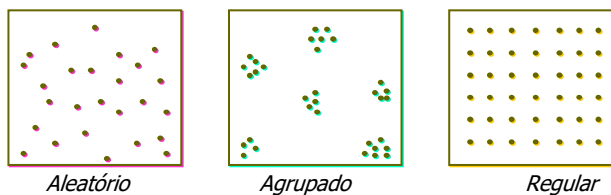
Objectivos da análise de PP

Objectives of the PPA

- ❑ Quantificar a dispersão dos eventos numa área geográfica confinada.
Quantify the dispersion of events in a confined geographic area.
- ❑ Estimar a variação de densidade dos eventos por toda a região de estudo
Estimate the variation of density of events throughout the study region
- ❑ Procurar modelos teóricos que expliquem essa distribuição.
Find theoretical models that represent reality.

Resumindo:

O estudo do PPA pretende verificar se os eventos observados apresentam algum tipo de padrão sistemático (existência de tendências), ao invés de estarem distribuídos aleatoriamente).



Summary: Verify if the observed events present some kind of systematic pattern (existence of tendencies), instead of being distributed randomly).

7

Clusters

- ❑ "Cluster": qualquer agregado de eventos.
 - resultado de classificação onde se procura definir um agrupamento de "semelhantes".
- ❑ Cluster espacial
 - Conjunto de eventos no espaço com proximidade especial grande;
 - Ou ocorrência de "taxas semelhantes" em áreas próximas.
- ❑ O que causa um "cluster"?
 - Agentes infecciosos, contaminação ambiental localizada, efeitos colaterais de tratamentos, etc.

"Cluster": any aggregate of events.
result of classification where we try to define a grouping of "similar".

Space Cluster
Event aggregate in space;
Or occurrence of "similar rates" in nearby areas.

What causes a "cluster"?
Infectious agents, localized environmental contamination, side effects of treatments, etc.

8

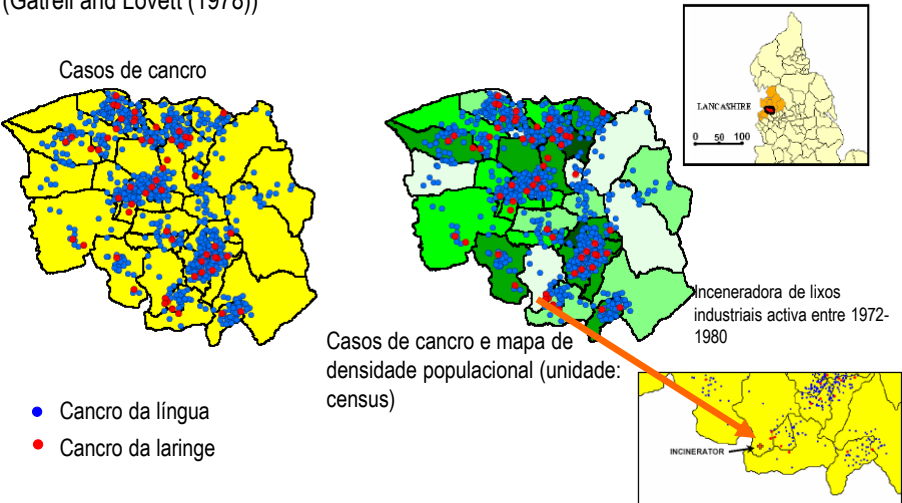
EXEMPLOS / EXAMPLES

- ❑ Epidemiologia / Epidemiology
 - ❖ A distribuição dos casos de uma doença formam um padrão no espaço? Existe associação com alguma fonte de poluição? Há evidência de contágio?
 - ❖ Does the distribution of cases of a disease form a pattern in space? Is there any association with any source of pollution? Is there any evidence of contagion?
- ❑ Sociologia /Sociology
 - ❖ Distribuição de pessoas desempregadas numa determinada região.
 - ❖ Distribution of unemployed persons in a given region.
- ❑ Demografia /Demography
 - ❖ Localização de diferentes etnias num país;
 - ❖ Location of different ethnicities in one country;
- ❑ Criminalogia /Criminology
 - ❖ Roubos que ocorrem em determinadas áreas estão correlacionados com características sócio-económicas ?
 - ❖ Robberies occurring in certain areas are correlated with socio-economic characteristics?
- ❑ Geologia/Paleontologia / Geology / Paleology
 - ❖ Dado um conjunto de amostras, qual a extensão de um depósito mineral ?
 - ❖ Given a set of samples, what is the extent of a mineral deposit?
- ❑ Biologia /Biology
 - ❖ Localização de espécies vegetais/animais de interesse.
 - ❖ Location of plant / animal species of interest.

9

Epidemiologia / Epidemiology

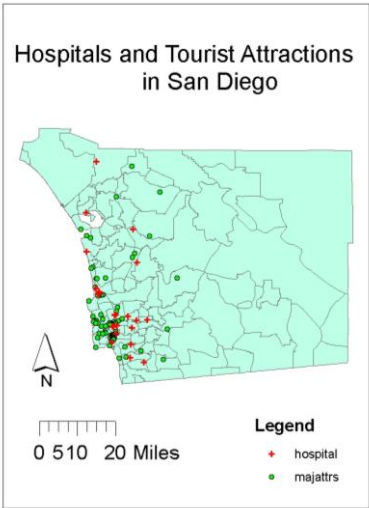
Pessoas com cancro da língua e laringe, em Lancashire (Inglaterra), entre 1974-1983 (Gatrell and Lovett (1978))



10

Social/Cultural/Turístico

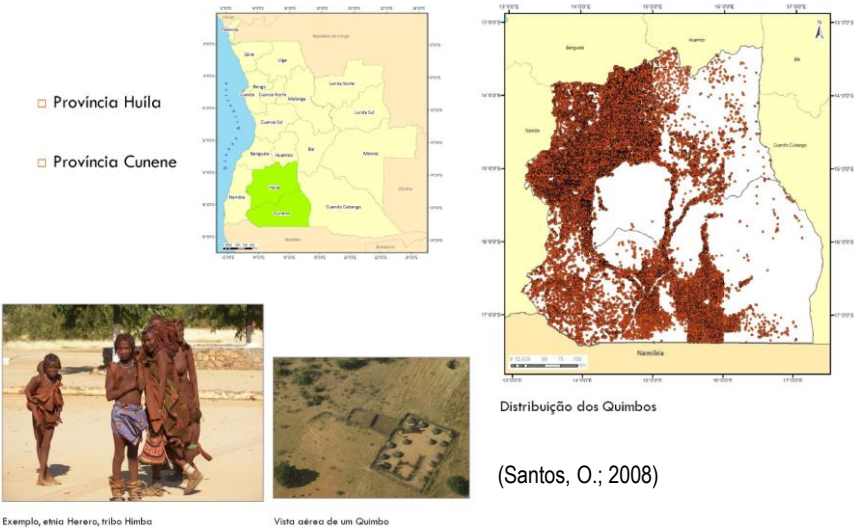
Social / Cultural / Tourist



- O mapa mostra as localizações de hospitais e as atrações turísticas em San Diego
- Questões:
 - Estão os hospitais aleatoriamente distribuídos? Onde localizar um novo hospital?
 - Estão as atrações turísticas aglomeradas espacialmente? Onde localizar um novo parque de diversões?

11

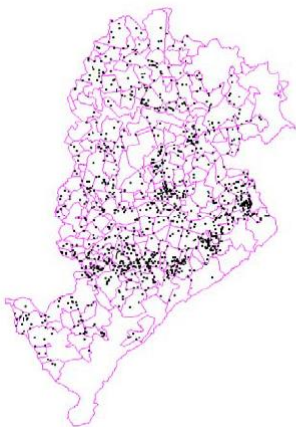
Demografia /Demography



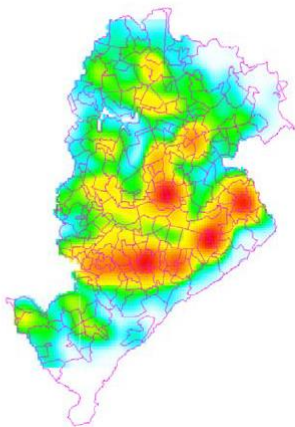
12

Criminalologia / Criminology

Roubos de taxis (Belo Horizonte, Brasil: 1995-2000)



Localização dos eventos

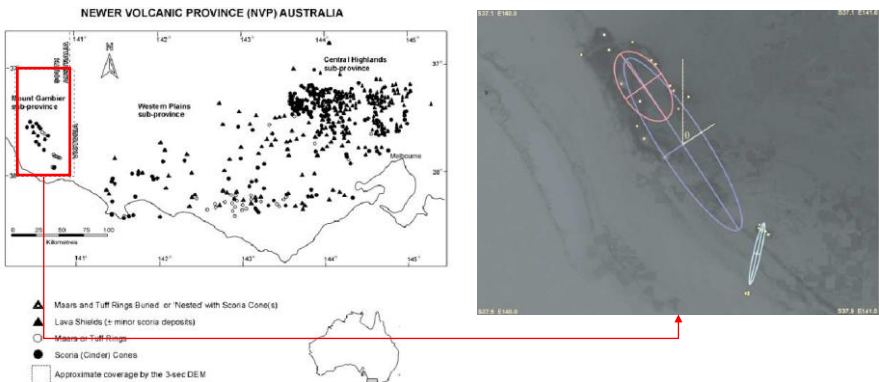


Estimação da densidade de eventos (intensidade por unidade de área)

13

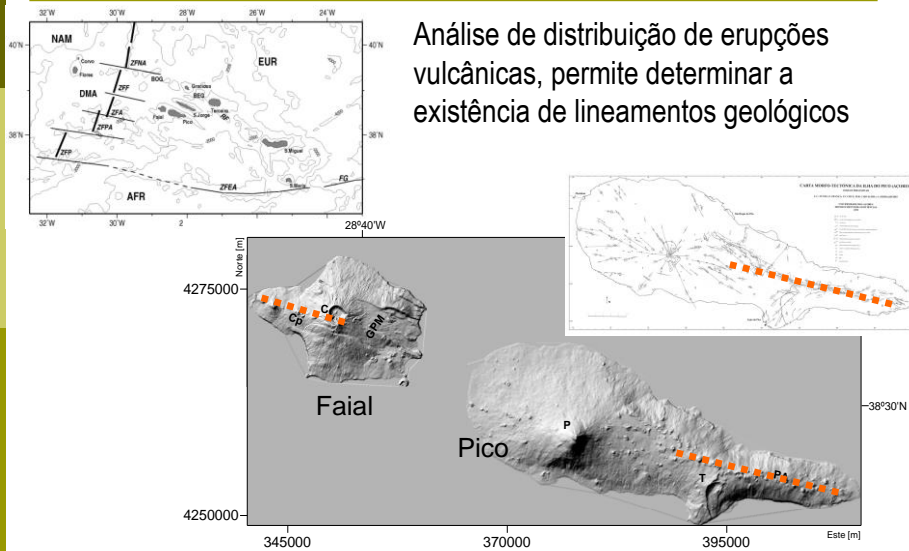
Geologia / Geology

Análise de distribuição de erupções vulcânicas, permite determinar a existência de lineamentos geológicos



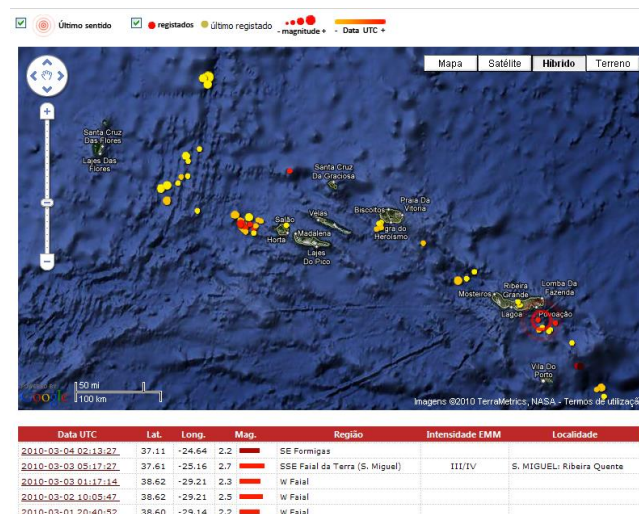
14

Geologia / Geology



15

Geologia / Geology



<http://www.cvarg.azores.gov.pt/Cvarg/CentroVulcanologia/actividadesismo/vulcanica/>

16

Componentes da Análise do Padrões Pontuais de Pontos
Components of the PPA

- 1. Visualização dos dados espaciais
- 2. Análise exploratória dos dados espaciais
- 3. Modelação dos dados espaciais

- 1. Visualization of spatial data
- 2. Exploratory analysis of spatial data
- 3. Spatial data modeling

19

Componentes da Análise do Padrões Pontuais de Pontos
Components of the PPA

- 1º Visualização de dados espaciais

□ **Descritores /Descriptors of PP**

Percepção rápida de como é que os eventos se distribuem no espaço, (i.e. variação e orientação espacial);

 - Frequência / Frequency
 - Densidade / Density
 - Centro geométrico do padrão de pontos / Geometrical Center
 - Dispersão espacial (orientação) / Spatial orientation

Quick perception of how points are distributed in space, (i.e. variation and spatial orientation);
- 2º Análise exploratória de dados espaciais

□ **Detectores /Detectors of PP**

Quantificação, através de indicadores estatísticos, do grau de associação espacial;

 - Contagem por quadrantes / Quadrat count analysis
 - Estimador de intensidade / Kernel density
 - Análise de vizinhos próximos / Nearest neighbor analysis

Recognition and measurement of the pattern of a point distribution;

20

Visualização de dados espaciais

Descritores de DP: Frequência e Densidade

A

Conjunto de localizações
 S com n eventos

Dado um conjunto de S localizações com n eventos:

$$S = \{s_1, s_2, \dots, s_i, \dots, s_n\}$$

A região de estudo A tem uma área a .

□ **Frequência** (n) : N° eventos que ocorrem na região A

□ **Densidade** (λ) : Frequência (n) / Área (a)

$$\lambda = \frac{n}{a} = \frac{no.(S \in A)}{a}$$

A **densidade** depende da área da região de estudo.

21

Visualização de dados espaciais

Descritores de DP: Frequência e Densidade

(12, 15)

A

B

C

D

Frequência = 12 em A, B, C, D

Densidade = 12/180 em A, B, C, D

Igual Frequência
Igual Densidade

mas,
Padrões de distribuição distintos

Quatro padrões de pontos (A, B, C e D).

22

Visualização de dados espaciais

Descritores de DP: Centrografia

É usada para determinar onde é que eventos, descritos pelas suas posições geográficas, estão concentrados (**centro geométrico**) e para medir o grau de dispersão do conjunto de eventos (**distância padrão**).

- **Centro Geométrico (ou centro médio, ou média espacial)** = **identifica a localização de maior concentração de pontos** e calcula-se pela média das coordenadas de localização X e Y

$$\bar{s} = (\mu_x, \mu_y) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right)$$

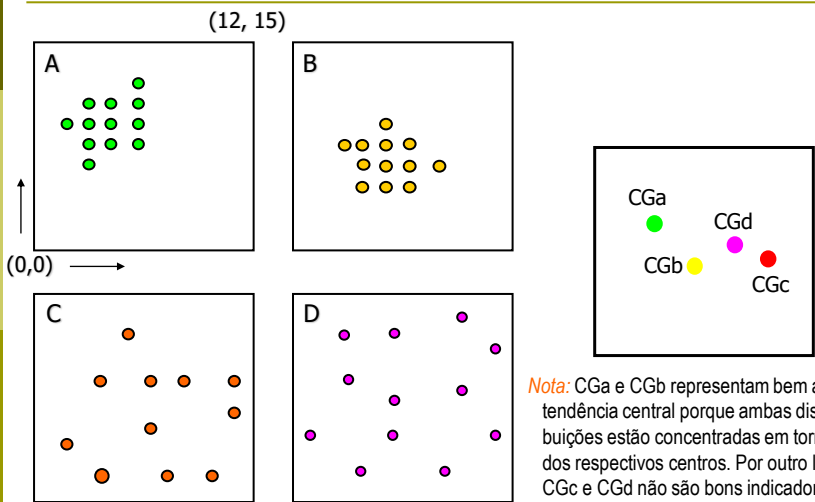
- **Distância padrão (ou dispersão espacial)** = **mede a dispersão dos eventos em torno do seu centro geométrico**. Conceptualmente, é similar aos desvio padrão na estatística clássica, mas a distância padrão expressa-se em unidades de distância.

$$d = \sqrt{\frac{\sum_{i=1}^n [(x_i - \mu_x)^2 + (y_i - \mu_y)^2]}{n}}$$

23

Visualização de dados espaciais

Descritores de DP : Centro Geométrico



Nota: CGa e CGb representam bem a tendência central porque ambas distribuições estão concentradas em torno dos respectivos centros. Por outro lado, CGc e CGd não são bons indicadores para suas respectivas distribuições.

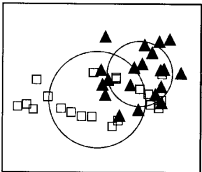
24

Visualização de dados espaciais

Descritores de DP : Círculo Padrão

Círculo (ou as elipses) padrão: Usa a distância padrão como **raio** de um **círculo centrado no centro geométrico** do conjunto dos eventos. Se a distância padrão é calculada separadamente para cada eixo (x e y), representam-se elipses em vez de círculos.

* Delimitam o espaço no qual tendem a estar concentrados 68% das observações



Círculo padrão



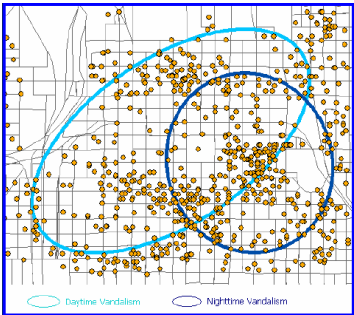
Elipses padrão

* Nas distribuições normais o intervalo entre a média e o desvio padrão positivo e negativo tende a abranger 68% das observações.

25

Visualização de dados espaciais

Descritores de DP : Elipse Padrão



Standard Distance

Input Feature Class: [dropdown]

Output Standard Distance Feature Class: [dropdown]

Circle Size: [dropdown]

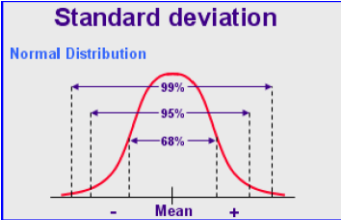
1 Standard Deviation

2 Standard Deviations

3 Standard Deviations

Case Field (optional): [dropdown]

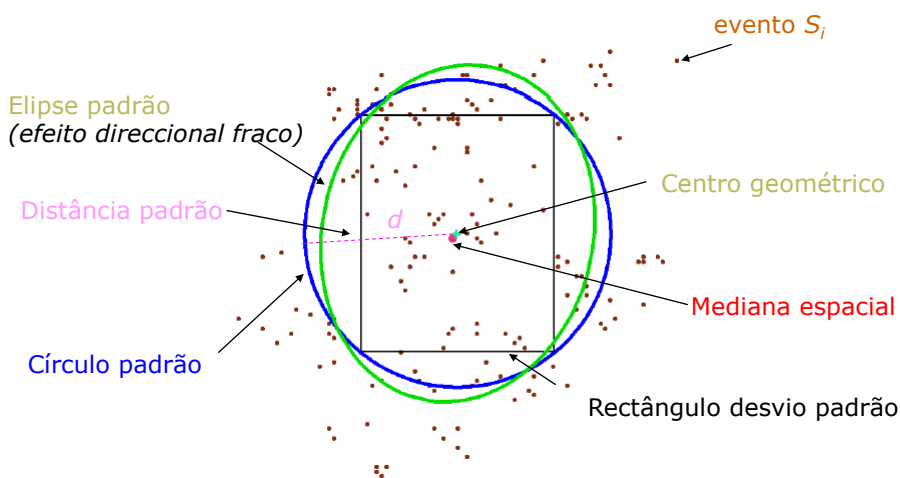
- 1 standard deviation = 68% of features
- 2 standard deviations = 95% of features
- 3 standard deviations = 99% of features



26

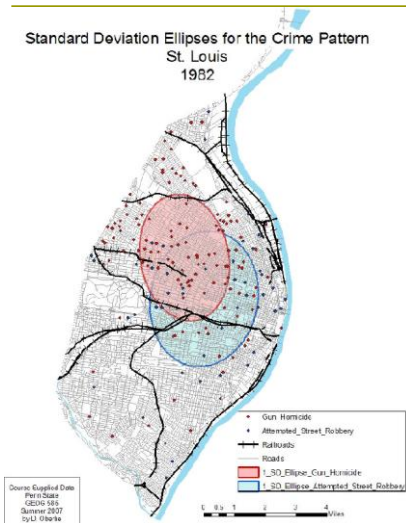
Visualização de dados espaciais

Descriptores de DP: resumo



27

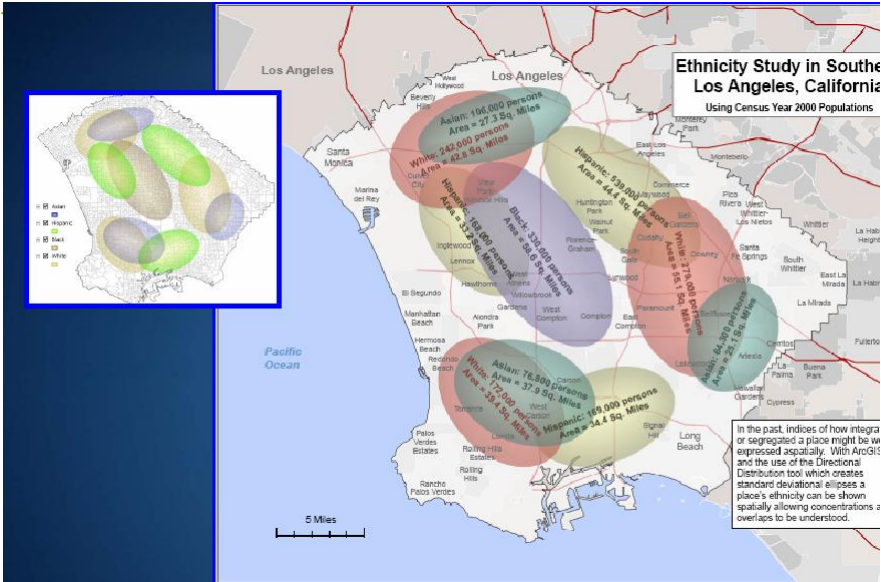
Exemplo 2 (elipses padrão): Distribuição de padrões de crimes



Comparação visual da propagação espacial de eventos distintos

28

Exemplo 3 (elipses padrão): Distribuição etnias na Califórnia



29

Análise Exploratória de dados espaciais
Detectores de DP

✓ Visualização dos dados espaciais

□ Análise Exploratória de dados espaciais

- Identificar se o Padrão é aleatório, disperso ou aglomerado?
- Quantificar o grau de associação espacial

□ Modelação dos dados espaciais

30

Análise Exploratória de dados espaciais

Efeitos de 1ª ordem e Efeitos de 2ª ordem

- ❑ **Processo estocástico:** processo que gera aleatoriamente pontos no espaço. Um processo estocástico pode ser caracterizado por efeitos de 1ª e 2ª ordem;
 - ⇒ **Efeitos de Primeira Ordem**
 - ❑ considerados globais ou de grande escala.
 - ❑ correspondem a variações no valor médio do processo.
 - ❑ Neste caso estamos interessados na **densidade** do processo (N° Eventos / Área).
 - ⇒ **Efeitos de Segunda Ordem**
 - ❑ denominados locais ou de pequena escala.
 - ❑ representam a **dependência espacial** no processo ou seja do grau de associação espacial dos eventos

31

Análise Exploratória de dados espaciais

Técnicas da Análise Exploratórias de 1ª e 2ª ordem

- | | | |
|--------------------------------------|---|--|
| Exploram as propriedades de 1ª ordem | { | 1. Método da contagem por quadrantes |
| | | 2. Estimador de densidade (<i>Kernel density estimation</i>) |
| Exploram as propriedades de 2ª ordem | { | 3. Métodos baseados na medição de distâncias entre pontos |
| | | (vizinho mais próximo, Função G, Função F, e a Função K) |

Para informação mais detalhada de outras técnicas consultar e.g. Statistical Analysis of Spatial Point Patterns by Peter J. Diggle (2001)

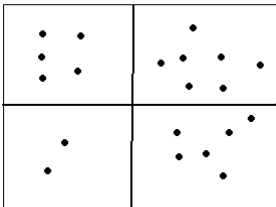
32

Análise Exploratória de dados espaciais

Método da contagem por quadrantes

Este método envolve a simples contagem de eventos em cada quadrante. Quadrantes grandes produzem uma fraca descrição do padrão, mas quadrantes demasiado pequenos podem não conter eventos.

n = nº de quadrantes



$$Mean = \frac{No. \text{ of pts. in the region}}{No. \text{ of quadrants}} = \frac{20}{4} = 5$$

$$Variance = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{2^2 + 5^2 + 6^2 + 7^2 - \frac{(20)^2}{4}}{4-1} = 4.5$$

x_i – frequência de pontos em cada quadrante

Variance to Mean Ratio $\Rightarrow VTMR = \frac{Variance}{Mean} = \frac{4.5}{5} = 0.9$

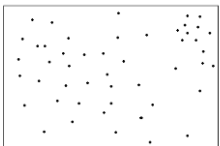
33

Análise Exploratória de dados espaciais

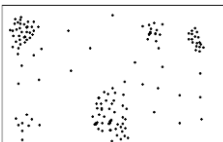
Variance To Mean Ratio (VTMR)



Se VTMR>1, o padrão é de aglomerados (*clusters*). Isto implica que existem conjuntos de dados agrupados e existem áreas extensas sem pontos (*variabilidade espacial grande*).



Se VTMR<1, o padrão é regularmente disperso, implicando que os eventos estão distribuídos mais ou menos regularmente por toda a região (*variabilidade espacial pequena*).



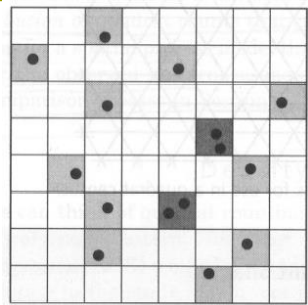
Se VTMR=1, o padrão é aleatório. Isto implica que o padrão de dados não tende nem para o aglomerado nem para a dispersão.

34

Análise Exploratória de dados espaciais

Método da contagem por quadrantes: *census*

- 1. Contagem exaustiva de "células da grelha" (quadrantes) que cobrem toda a área de estudo
- ✓ Escolha da origem, orientação dos quadrantes, e como é que a dimensão dos quadrantes podem afectar a distribuição da frequência
- ✓ Qual a melhor dimensão da grelha ?
- ✓ A dimensão óptima da grelha pode ser calculada por $2A/n$ (onde A=área da região de estudo e n=nº de pontos da distribuição).



Number of events in quadrat	Census, n = 64		Sampling, n = 38	
	Count	Proportion	Count	Proportion
0	51	0.797	29	0.763
1	11	0.172	8	0.211
2	2	0.031	1	0.026
3	0	0.000	0	0.000

Padrão "Clustered": Todos ou quase todos os pontos estão inseridos dentro do mesmo quadrante;

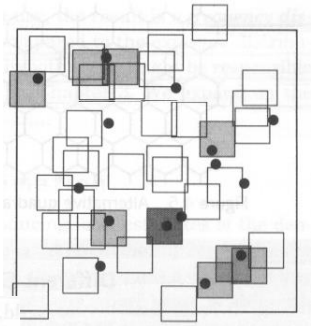
Padrão disperso: todos os quadrantes contêm um número similar de pontos

35

Análise Exploratória de dados espaciais

Método da contagem por quadrantes: *Random sampling*

- O método *Random sampling* é mais utilizado na prática;
- ✓ É possível aumentar a dimensão da amostra adicionando mais quadrantes (para padrões dispersos)
 - ✓ Pode descrever-se o padrão de pontos sem considerar a totalidade dos dados



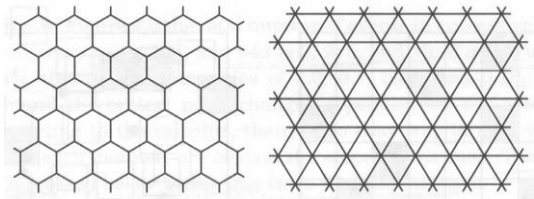
Number of events in quadrat	Census, n = 64		Sampling, n = 38	
	Count	Proportion	Count	Proportion
0	51	0.797	29	0.763
1	11	0.172	8	0.211
2	2	0.031	1	0.026
3	0	0.000	0	0.000

36

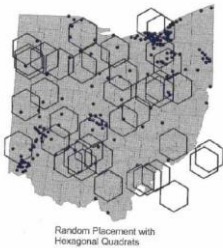
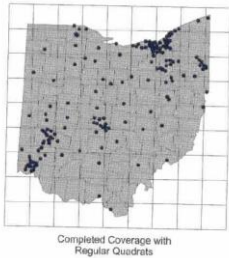
Análise Exploratória de dados espaciais

Método da contagem por quadrantes

Outras formas de quadrantes



Distribuição de
cidades no estado de
Ohio (EUA)



37

Análise Exploratória de dados espaciais

Limitações do Método da contagem por quadrantes

Visualmente os dois padrões são diferentes, mas o método da contagem por quadrantes apresenta o mesmo resultado!

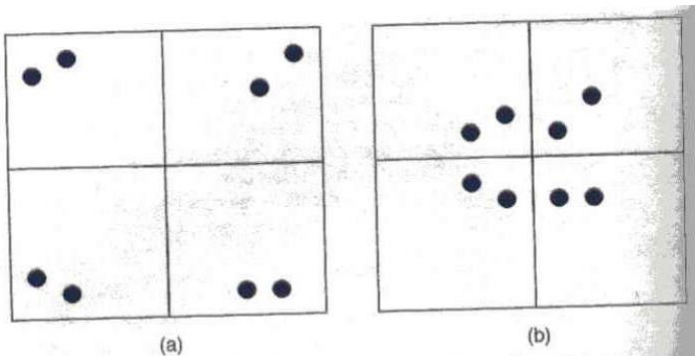
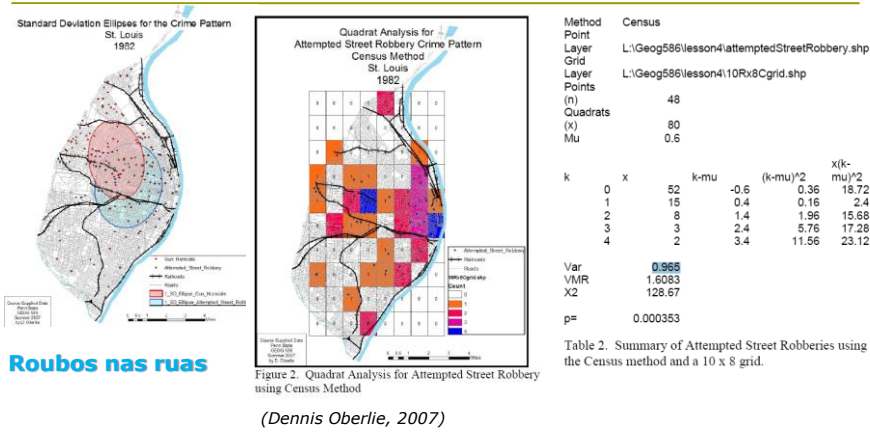


Figure 6.4 Hypothetical patterns illustrating the need for ordered neighbor statistics

38

Análise Exploratória de dados espaciais

Método da contagem por quadrantes



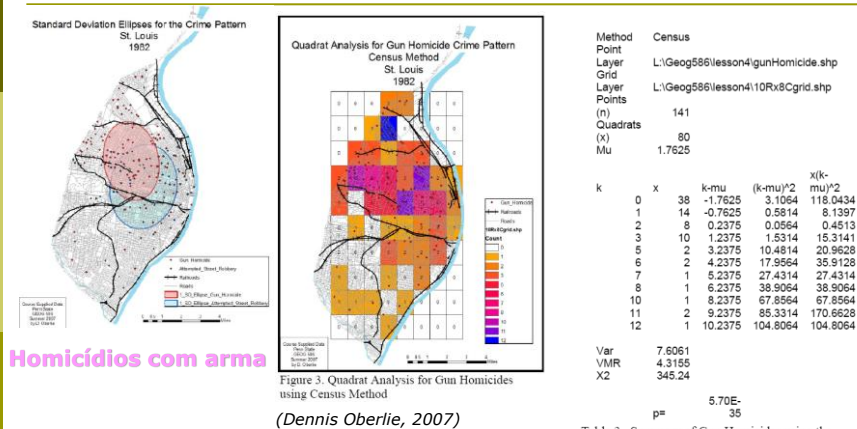
Roubos nas ruas

Média = N° pontos/N° Quad. = 48/80=0.6
Variância = 0.965
VTMR = 1.6 ⇒ Aleatório?

39

Análise Exploratória de dados espaciais

Método da contagem por quadrantes



Homicídios com arma

Média = N° pontos/N° Quad. = 141/80=1.7625
Variância = 7.6061
VTMR = 4.3155 ⇒ Aglomerado?

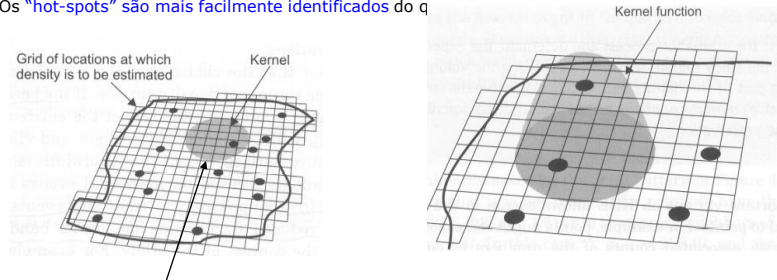
(Os roubos nas ruas apresentam um padrão menos aglomerado do que os homicídios com arma)

40

Análise Exploratória de dados espaciais

Estimador de Densidade (“Kernel Estimation”)

- ❑ O método consiste em converter o conjunto dos eventos pontuais numa superfície contínua que represente a densidade, para tal ajusta-se uma função uni(bi)-dimensional sobre os eventos considerados, cujo valor será proporcional à densidade das amostras por unidade de área. Esta função realiza uma contagem de todos os pontos dentro de uma região de influência (círculo ou esfera), ponderando-os pela distância de cada um à localização de interesse.
- ❑ O método é influenciado pela escolha da “largura da faixa” (raio de um círculo ou esfera centrado nos eventos usados para criar a superfície de densidade). A escolha da “largura da faixa” requer várias tentativas até encontrar a que produz um resultado satisfatório.
- ❑ Os “hot-spots” são mais facilmente identificados do q

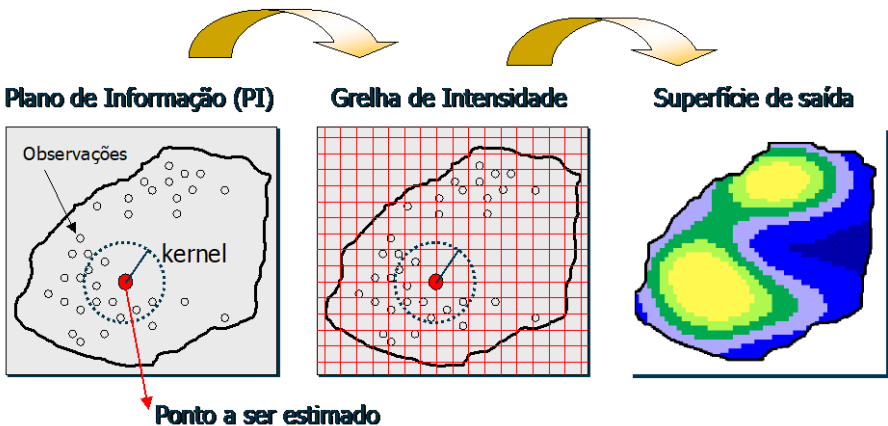


Densidade de amostras por unidade de área

41

Análise Exploratória de dados espaciais

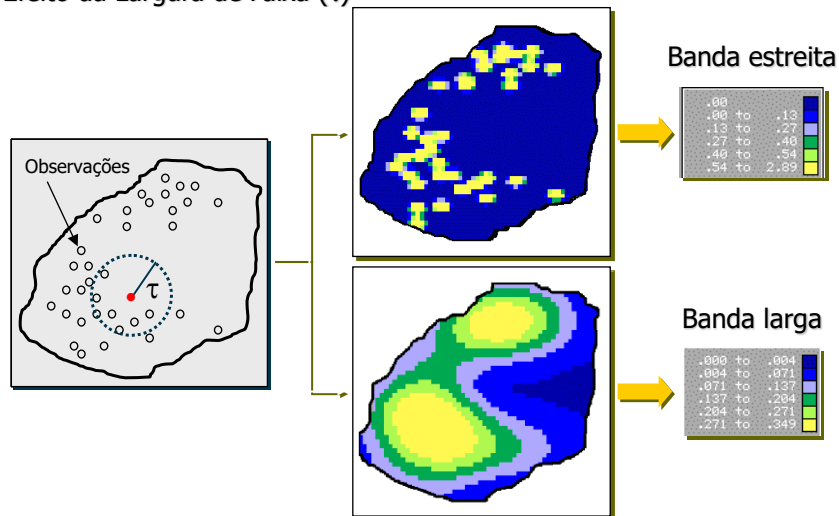
Estimador de Intensidade (“Kernel Estimation”)



42

Estimador de Intensidade (“Kernel Estimation”)

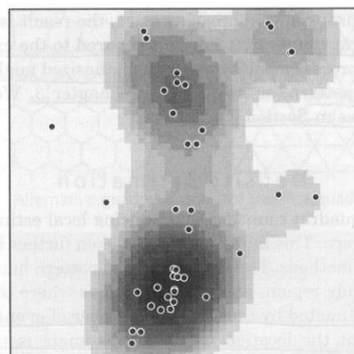
- Efeito da Largura de Faixa (τ)



43

Análise Exploratória de dados espaciais Estimador de Densidade (“Kernel Estimation”)

- 1) Visualização do padrão de pontos permite detectar *hot-spots*.
- 2) Verificar se o processo é ou não estacionário de 1ª ordem (densidade constante) a partir das variações locais de intensidade.
- 3) Ligação dos objectos pontuais a outros dados geográficos (e.g. focos de doenças ou focos de poluição).



44

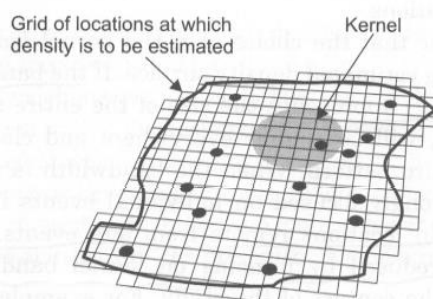
Análise Exploratória de dados espaciais

Exemplo de função uni-dimensional do Estimador de Intensidade

É determinada a densidade em cada local da região de estudo, e não apenas nos locais onde existem eventos.

A densidade é estimada contando o número de eventos (**no**) na região ou *kernel*, centrada no local onde se pretende fazer a estimativa (p)

$$\hat{\lambda}_p = \frac{no.[S \in C(p, r)]}{\pi r^2}$$



Estimação simples de densidade

$C(p, r)$ é o círculo de raio r centrado na localização de interesse p

45

Análise Exploratória de dados espaciais

Exemplo de função bi-dimensional do Estimador de Intensidade

Segundo (Bailey e Gatrell, 1995):

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

Onde:

- A função $I(\cdot)$ -> FDP, escolhida de forma adequada para construir uma superfície contínua sobre os dados.
 - O parâmetro τ denominado "largura de faixa", controla a suavização da superfície gerada.
- ✓ s representa uma localização qualquer na área de estudo e s_i são as localizações dos eventos observados.
- ✓ n representa o número de eventos.

46

Análise Exploratória de dados espaciais

Estimador de Intensidade ("Kernel Estimation"): Funções mais complexas (funções bi-dimensionais)

- Uma função muito utilizada para $I()$ é:

$$I(h) = \frac{3}{\pi} \left(1 - h^2\right)^2$$

onde:

- h representa a distância entre a localização em que desejamos calcular a função e os eventos observados.

- Assim o estimador de intensidade pode ser expresso como:

$$\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{3}{\pi \tau^2} I\left(1 - \frac{h_i^2}{\tau^2}\right)^2$$

onde:

- h_i é a distância entre o ponto a calcular S e o valor observado S_i .

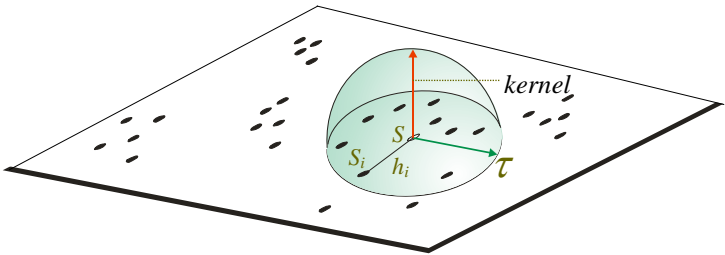
47

Análise Exploratória de dados espaciais

Estimador de Intensidade ("Kernel Estimation"): Funções mais complexas (funções bi-dimensionais)

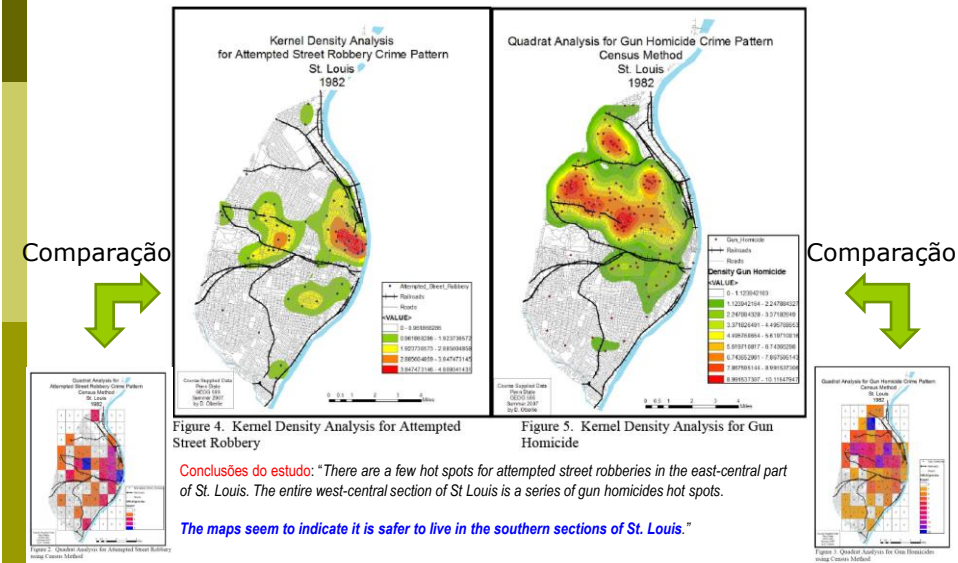
Visão gráfica do estimador de intensidade

$$\hat{\lambda}_{\tau}(s) = \sum_{i=1}^n \frac{3}{\pi \tau^2} I\left(1 - \frac{h_i^2}{\tau^2}\right)^2$$



48

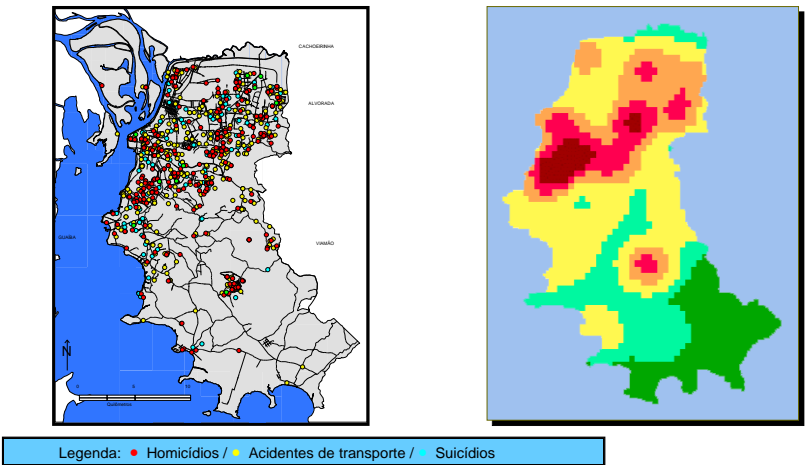
Análise Exploratória de dados espaciais
EXEMPLO: Estimador de Intensidade (“Kernel Estimation”)



49

EXEMPLO: Estimador de Intensidade (“Kernel Estimation”)

Exemplo: Mapeando a violência na cidade de Porto Alegre - RS.

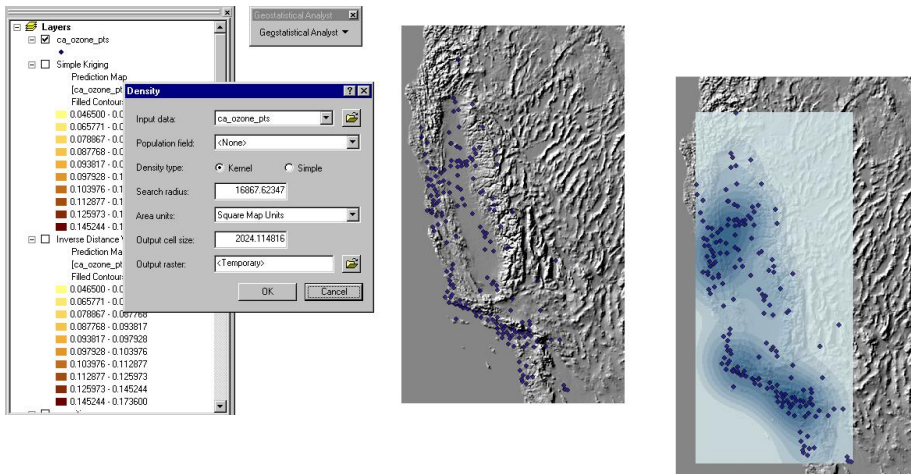


50

Análise Exploratória de dados espaciais

EXEMPLO: Estimador de Intensidade (“Kernel Estimation”)

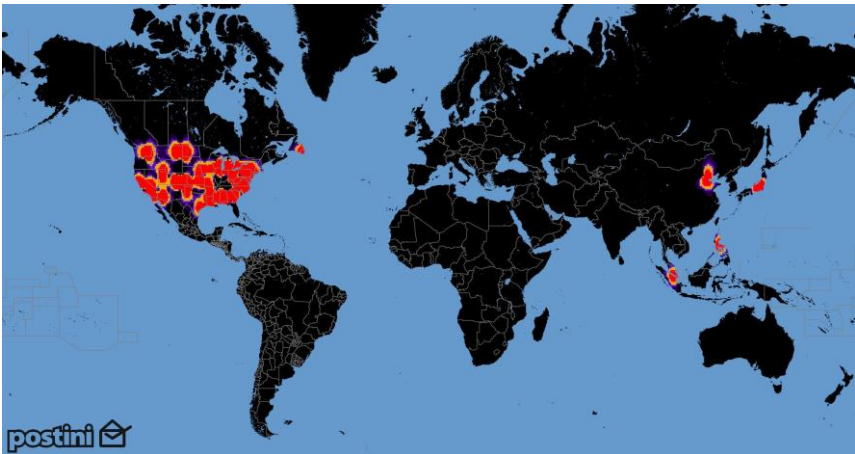
- Valor de ozono medido em estações terrestres



51

EXEMPLO: Estimador de Intensidade (“Kernel Estimation”)

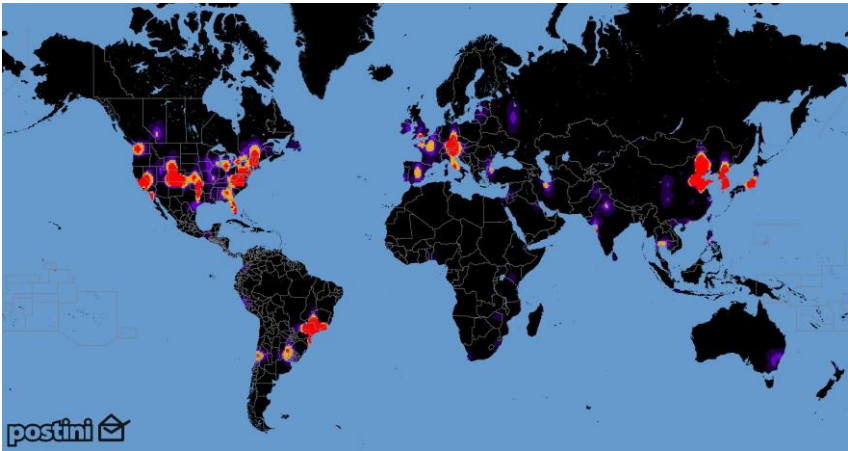
- Exemplo: Origem de vírus computacionais



52

EXEMPLO: Estimador de Intensidade (“Kernel Estimation”)

- Exemplo: Origem de e-mails Spam



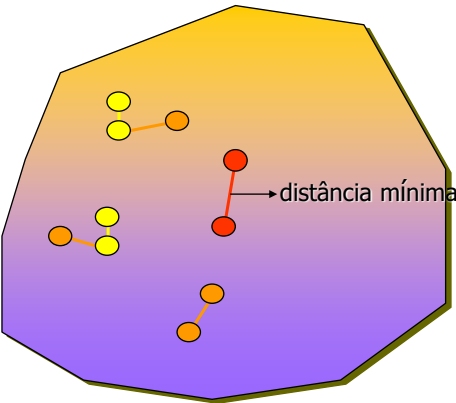
53

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias:

Método do vizinho mais próximo

- Procura as distâncias entre eventos do padrão de pontos;
- É o método mais directo para descrever as propriedades de 2ª ordem de um processo pontual;



Baseado na distância mínima entre os pontos

- 1 distância
- 2 distância
- 3 distância

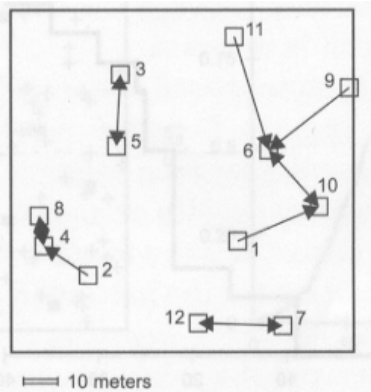
54

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias:

Vizinho mais próximo

O método do vizinho mais próximo estima a função de distribuição cumulativa ($\hat{G}(w)$) a partir das distâncias entre eventos na região de estudo.



Distância euclideana

$$d(s_i, s_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

$$d_{\min}(s_i) = \min_{j \in (1, n) \& j \neq i} (d_{ij})$$

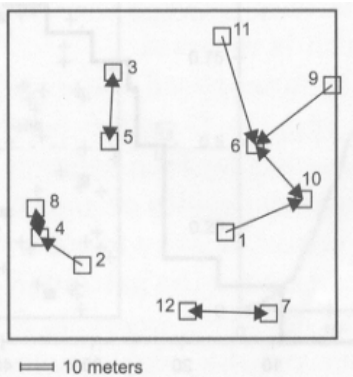
$$\hat{G}(w) = \bar{d}_{\min} = \frac{\sum_{i=1}^n d_{\min}(s_i)}{n}$$

55

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias:

Vizinho mais próximo: Função G



Event	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	21.14
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	42.53	6	21.94
10	89.78	42.53	6	21.94
11	65.19	92.08	6	34.63
12	54.46	8.48	7	24.81

Se o padrão for de cluster, tem um valor alto ou baixo?

$$\bar{d}_{\min}$$

$$\bar{d}_{\min} = 21.62$$

PRECISAMOS DE FUNÇÕES!!!

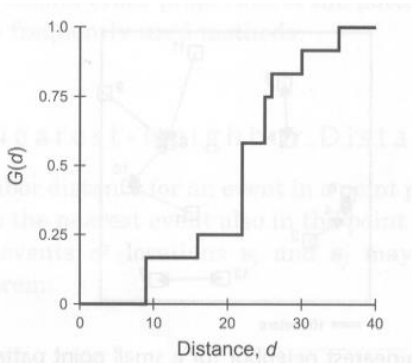
56

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias:

Vizinho mais próximo: Função G

Event	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	21.14
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	42.53	6	21.94
10	89.78	42.53	6	21.94
11	65.19	92.08	6	34.63
12	54.46	8.48	7	24.81



$$G(d) = \frac{no.[d_{min}(s_i) < d]}{n}$$

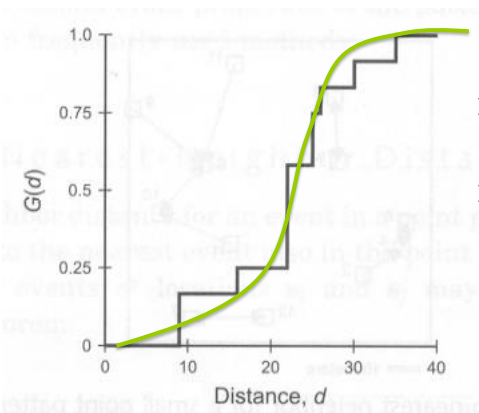
Função cumulativa

57

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias:

Vizinho mais próximo: Função G



O gráfico da função-G indica a forma como os eventos estão espaçados na região de estudo.

- **Clusters:** Se os eventos estão agrupados em clusters, G aumenta rapidamente em pequenas distâncias
- **Dispersos:** Se os eventos tendem para a dispersão, G aumenta lentamente até aproximadamente à distância entre eventos e depois aumenta rapidamente.

58

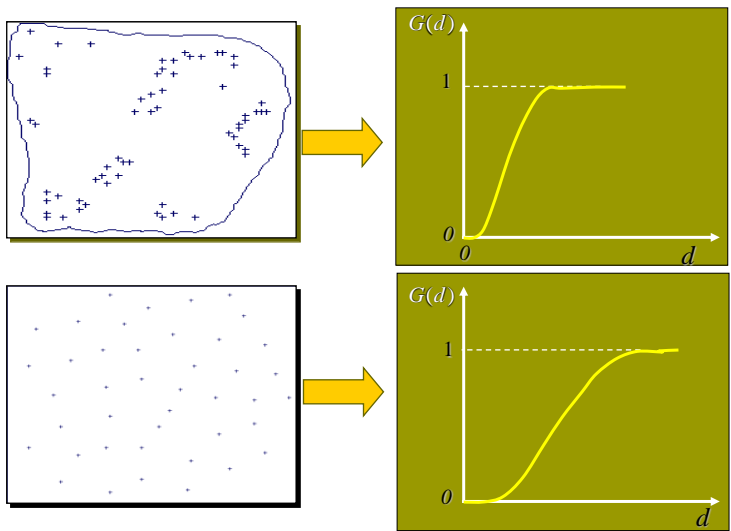
Análise Exploratória de dados espaciais

Métodos Baseados em distâncias:

Vizinho mais próximo: Função G

- Análise exploratória de padrões de distribuição de pontos utilizando

$G(d)$



59

Análise Exploratória de dados espaciais

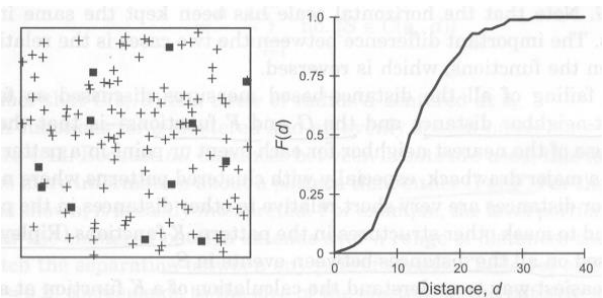
Métodos Baseados em distâncias:

Vizinho mais próximo: Função F

Igual a Função G mas usa apenas um subconjunto de pontos:

- 1) Seleccionam-se aleatoriamente m localizações $\{p_1, p_2, \dots, p_m\}$
- 2) Calcula-se $d_{\min}(p_i, S)$ a distância mínima entre a localização p_i a qualquer evento do padrão de pontos S
- 3) Calcula-se $F(d)$

$$F(d) = \frac{\text{no.}[d_{\min}(p_i, S) < d]}{m}$$

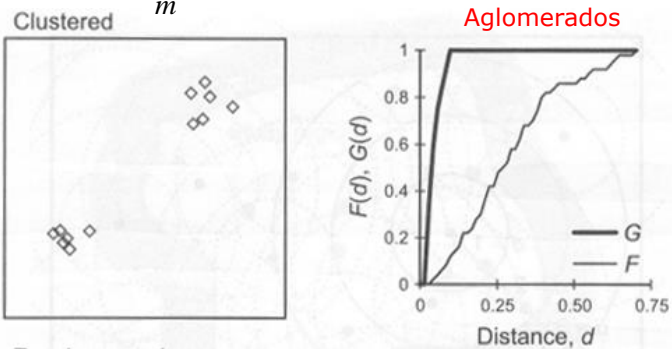


60

Comparação entre a Função G e a Função F

$$G(d) = \frac{no.[d_{\min}(s_i) < d]}{n}$$

$$F(d) = \frac{no.[d_{\min}(p_i, S) < d]}{m}$$

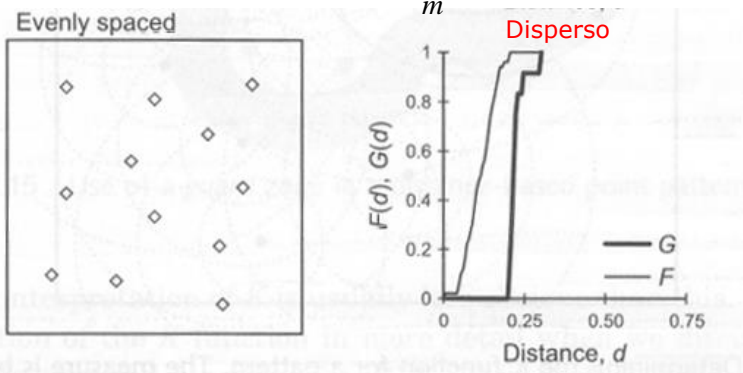


61

Comparação entre a Função G e a Função F

$$G(d) = \frac{no.[d_{\min}(s_i) < d]}{n}$$

$$F(d) = \frac{no.[d_{\min}(p_i, S) < d]}{m}$$



62

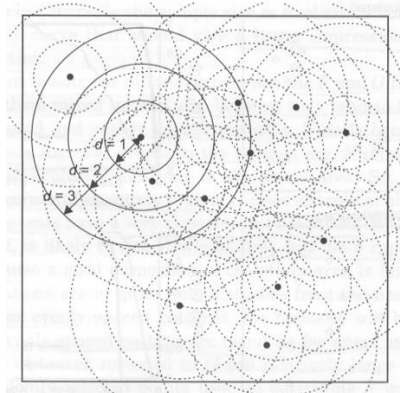
Métodos Baseados em distâncias: Função K - Ripley

1. As funções G e F, utilizam apenas a **distância mínima entre 2 eventos**;
2. Este facto pode constituir uma desvantagem, especialmente em padrões muito aglomerados nos quais as distâncias entre eventos podem ser muito pequenas relativamente a outras distâncias do padrão de pontos.
3. As funções K-Ripley (Ripley, 1976) são baseadas nas distâncias entre todos os eventos da região **S**.

63

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias: Função K-Ripley



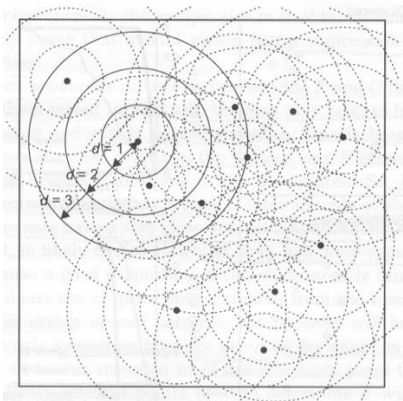
Quatro passos:

1. Para um evento em particular, define-se um círculo centrado no evento (s_i) e com um raio d_i ;
2. Conta-se o número de eventos no interior do círculo;
3. Calcula-se o valor médio das várias contagens em 2);
4. O valor médio é então dividido pela densidade de eventos de toda a área de estudo;

64

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias: Função K-Ripley



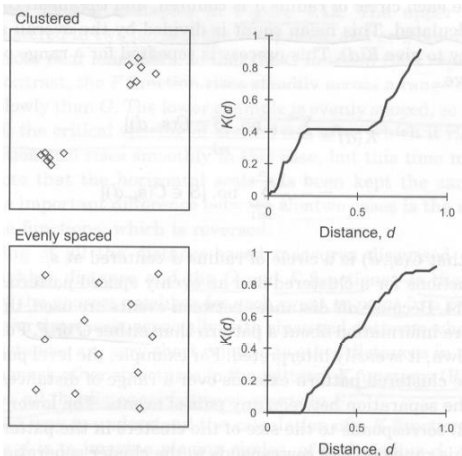
$$K(d) = \frac{\sum_{i=1}^n no.[S \in C(s_i, d)]}{n\lambda}$$
$$= \frac{a}{n} \bullet \frac{1}{n} \sum_{i=1}^n no.[S \in C(s_i, d)]$$

$\lambda = \frac{n}{a}$ É a densidade de eventos na área de estudo

65

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias: Função K-Ripley



Clustered

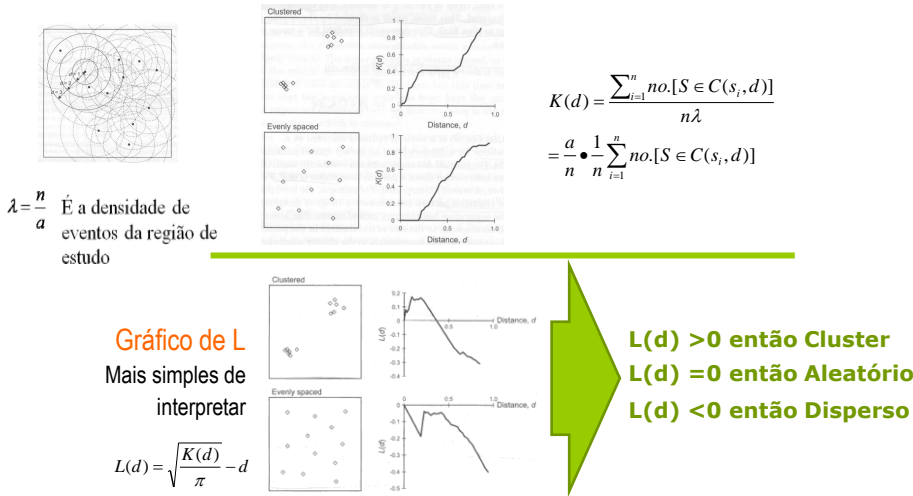
Aleatoriamente distribuídos

66

Análise Exploratória de dados espaciais

Métodos Baseados em distâncias: Função L

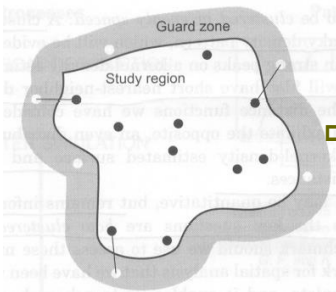
(versão linearizada da Função K-Ripley – tem um gráfico mais compreensível)



67

Método do vizinho mais próximo:

Efeitos de fronteira

- 

Guard zone

Study region
- ❑ O que é?
 - Pontos próximos da fronteira têm menos vizinhos do que pontos no interior da área de estudo;
 - Os resultados relativamente aos pontos próximos da fronteira requerem interpretação mais cuidada;
 - ❑ O que fazer?
 - Um buffer na área de estudo (outward ou inward), e incluir mais ou menos pontos;
 - Utilizar pesos menores para pontos próximos da fronteira;

68

Problemas associados à definição da fronteira

Área demasiado extensa \Rightarrow faz parecer os dados muito concentrados



Área demasiado pequena \Rightarrow faz parecer os dados muito dispersos

