Point pattern analysis

경희대학교
KYUNG HEE UNIVERSITY

Overview

- Describing point patterns
  - Point patterns and marked point patterns
  - Questions about point patterns
  - First- and second-order effects
  - Centrographic measures
    - Mean centre, standard distance, standard deviational ellipse
- Density-based analysis
  - Quadrat analysis
  - Variance-to-mean ratio (VMR)
  - Coffee shops in London
  - Weakness of quadrat analysis

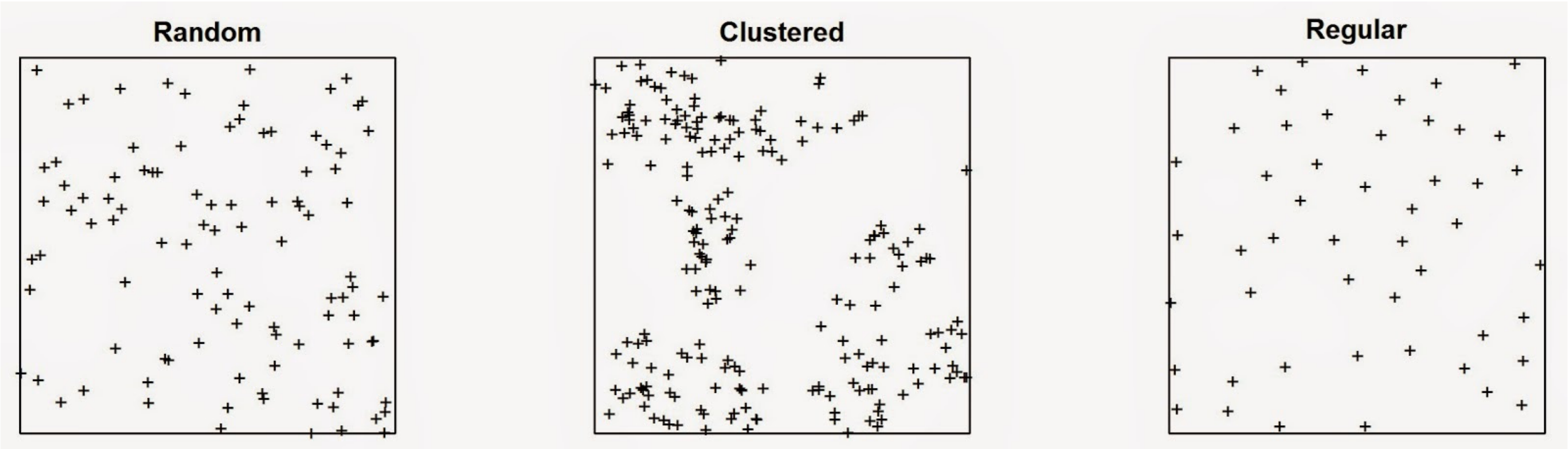
Describing point patterns

Point patterns

- Point patterns
  - Patterns where the only data are the locations of a set of point objects (e.g., crimes, deaths from some disease, plants of various species, and archaeological finds)
  - A point pattern of n events can be more formally defined as a set of locations $S = \{s_1, s_2 \dots s_n\}$, in which each event s_i has locational coordinates (x_i, y_i).
- Marked point patterns
  - Include both location and value (e.g., location and revenue of cafés in Hoegi-dong)

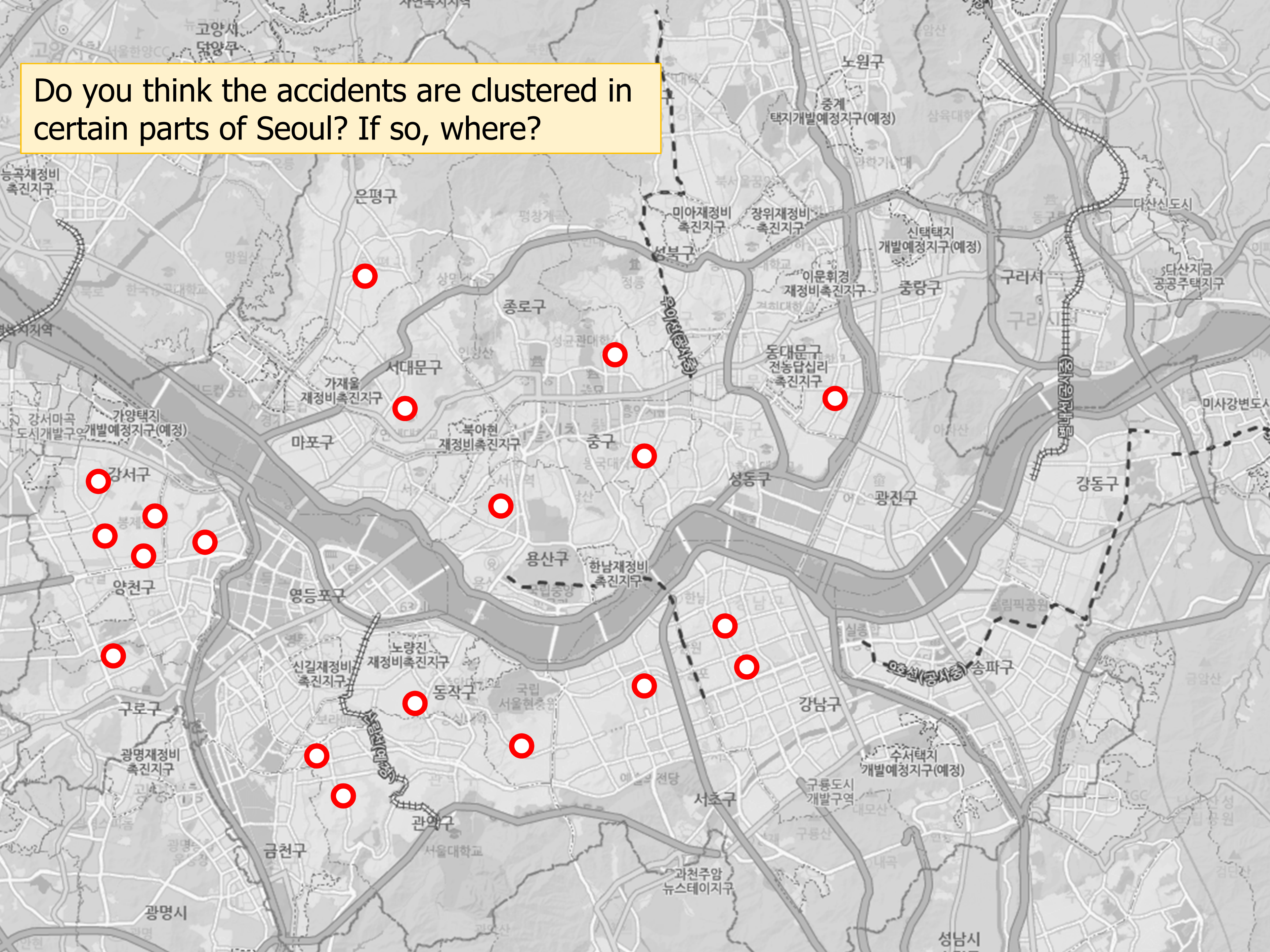
Questions about point patterns

- Is the pattern random, or is it structured in some fashion?
  - Clustered if the points are located closer than random
  - Dispersed if the points are farther than random
- What is the process that might have generated the pattern?
  - Not easy to answer …

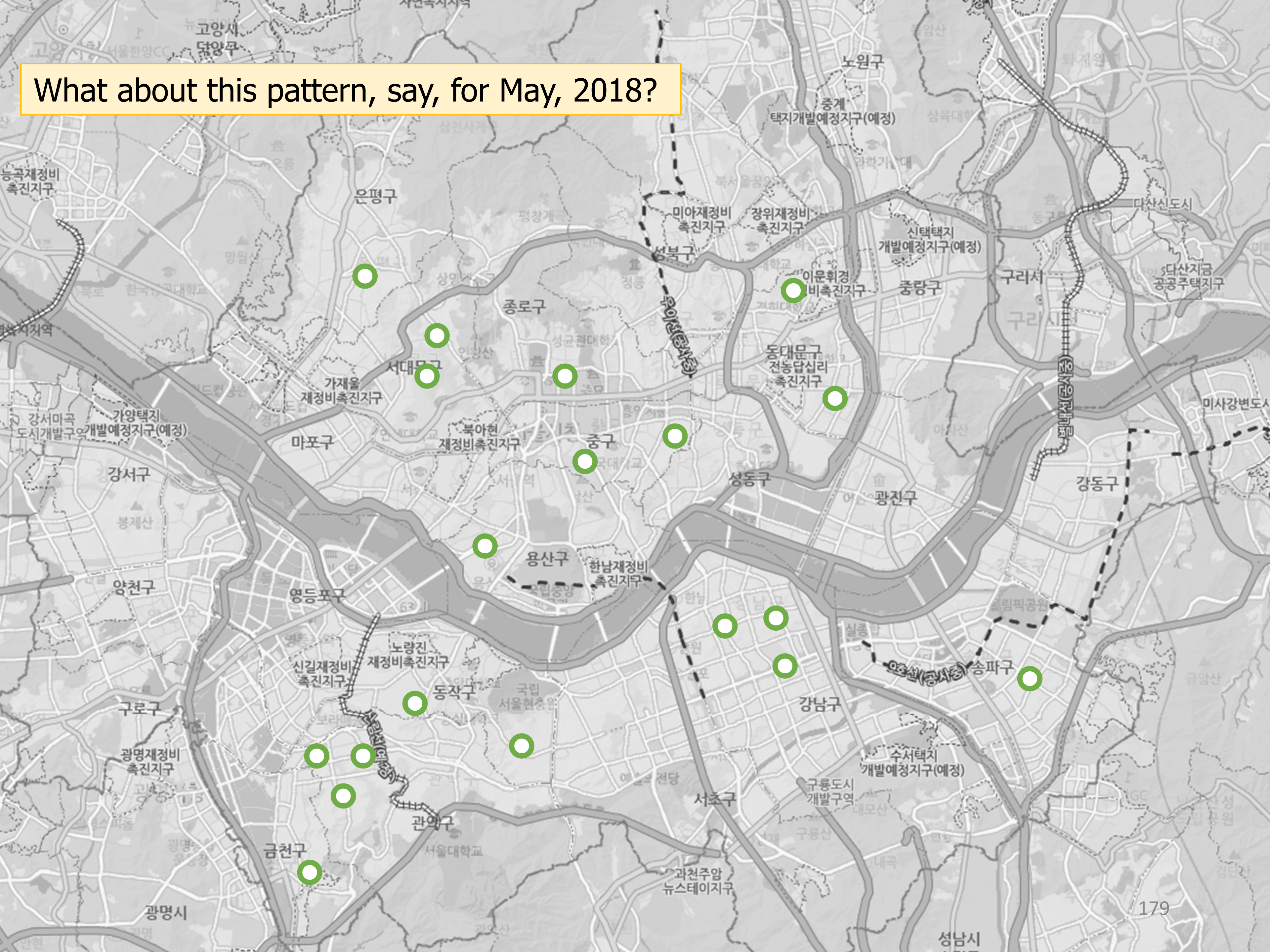


Source: http://r-video-tutorial.blogspot.com/2015/05/introductory-point-pattern-analysis-of.html

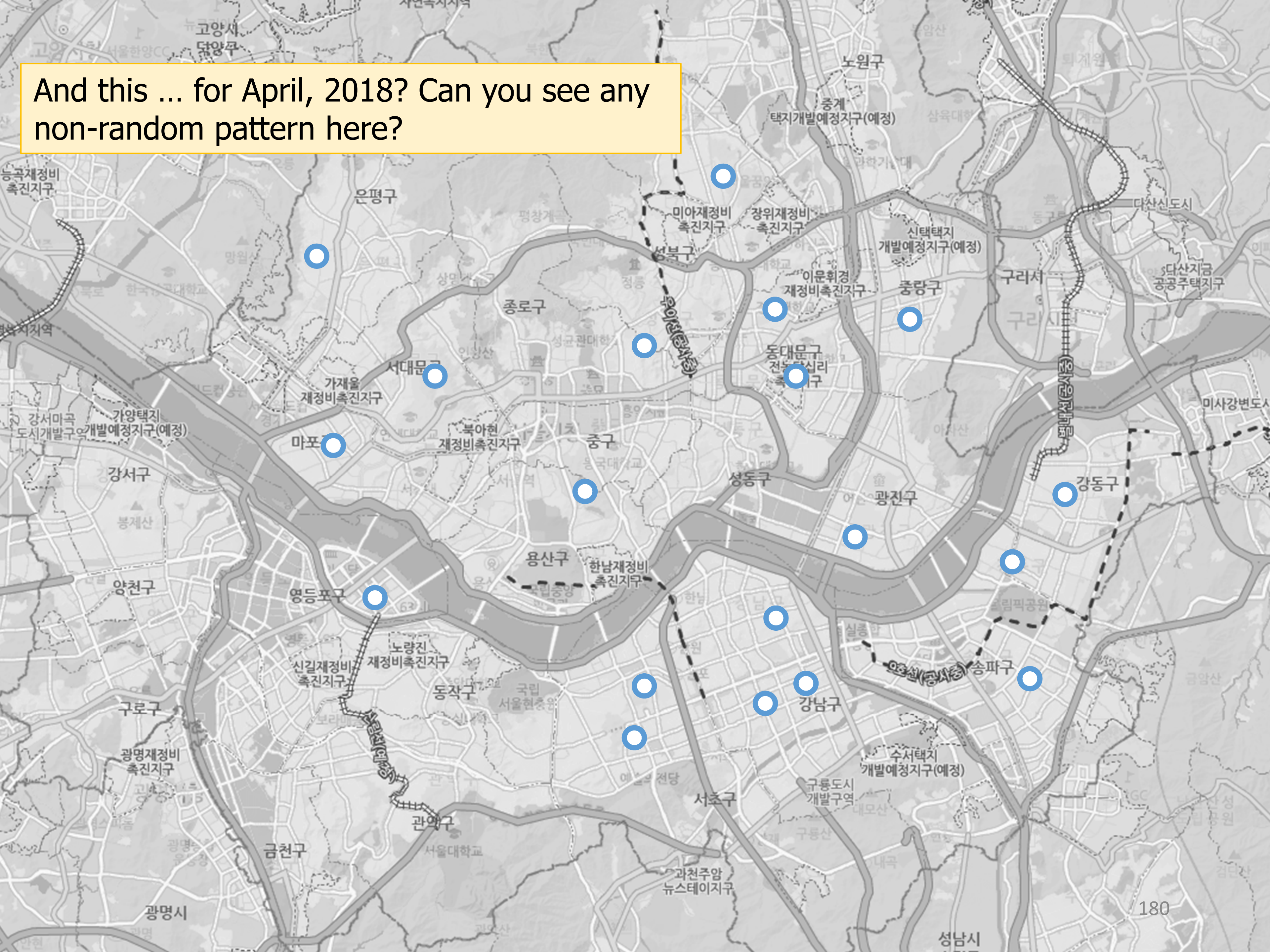
Do you think the accidents are clustered in certain parts of Seoul? If so, where?



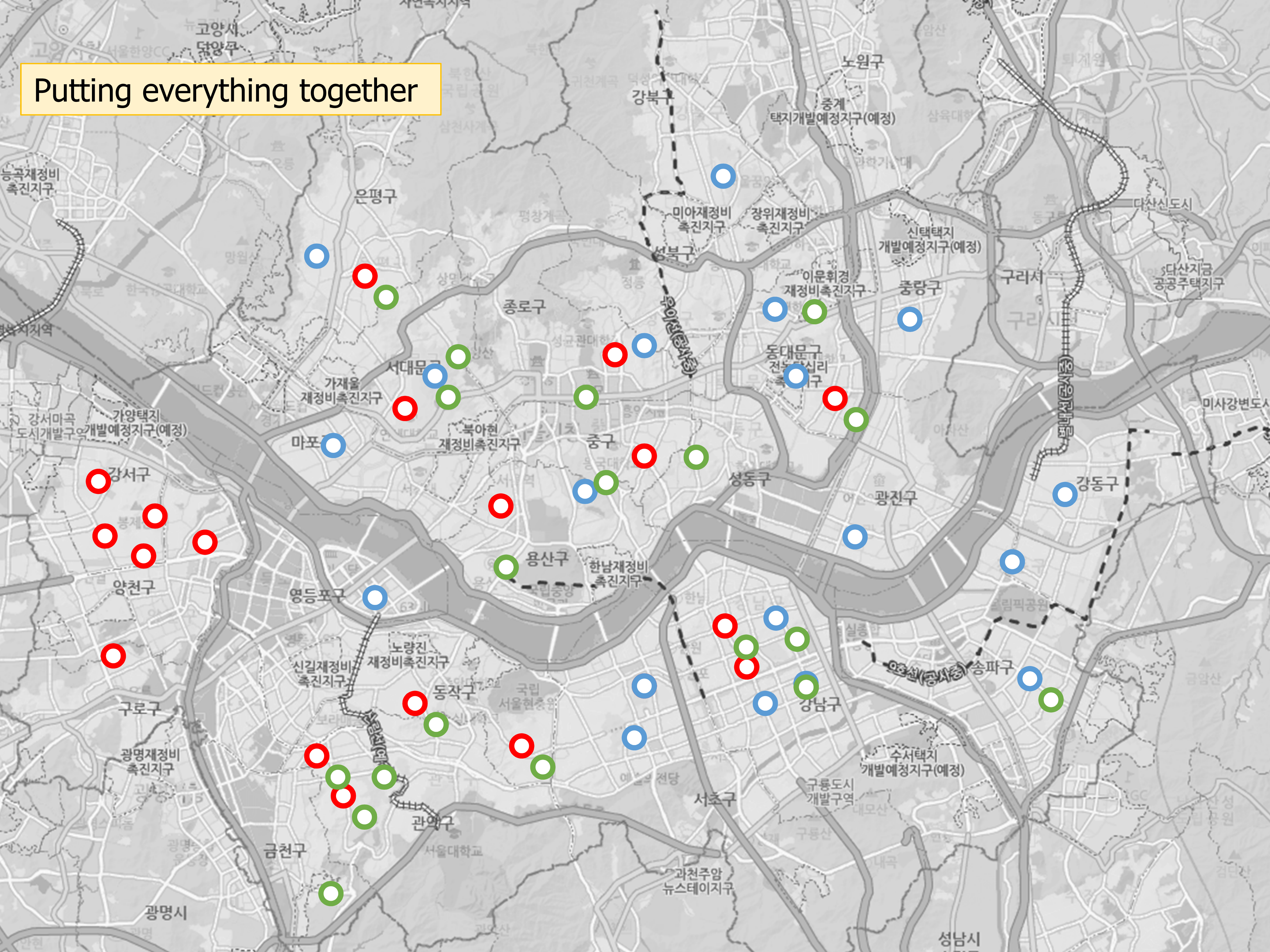
What about this pattern, say, for May, 2018?



And this … for April, 2018? Can you see any non-random pattern here?



Putting everything together



First- and second-order effects

- First-order effects
  - It relates to variation in the mean value of a process over space (e.g., south-to-north decrease in average temperatures).
  - The presence of first-order effects can be problematic for some methods, such as kriging.
- Second-order effects
  - It relates to the spatial correlation or spatial dependence in the process.
  - This is often what is left in the pattern once first-order variations have been removed.

First- and second-order effects

- Any point pattern analysis relates to either first-order or second-order spatial effects.
- In practice, it is difficult to distinguish first- and second-order effects in a spatial pattern by the analysis of spatial data.

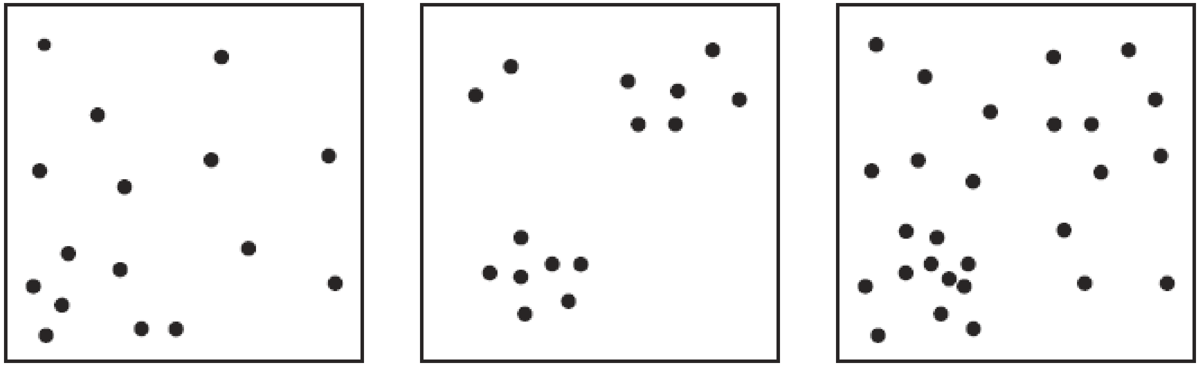


Figure 5.1 The difficulty of distinguishing first- and second-order effects.

Source: p. 124, O'Sullivan and Unwin (2010)

Centrographic measures

- Descriptive statistics to provide summary descriptions
  - Mean centre, standard distance, standard deviational ellipse
- Can be useful for comparing one distribution with another (say, population distribution in 1990 and that in 2000)
- Do not provide much information about the pattern itself

Mean centre

- The point whose coordinates are the average of the corresponding x and y coordinates of all the events in the pattern, S:

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

  - Sometimes referred to as the centre of gravity, or centroid

- Weighted mean centre can be computed by weighting x_i and y_i coordinates with another variable w_i.

Standard distance

- Standard deviation of a point data set that indicates how dispersed the events (i.e., points) are around their mean centre:

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

  – Can be used to plot a summary circle, centred at (μ_x, μ_y) with radius d.

Standard deviational ellipse

- Represent the overall shape of the point pattern
  - Calculate the standard distance in the x and y directions separately, and rotate it by the angle θ:

$$\tan\theta = \frac{A + B}{C}$$

where $A = \{\sum_{i=1}^{n}(x_i - \mu_x)^2 - \sum_{i=1}^{n}(y_i - \mu_y)^2\}$,

$$B = \sqrt{\left\{\sum_{i=1}^{n}(x_i - \mu_x)^2 - \sum_{i=1}^{n}(y_i - \mu_y)^2\right\}^2 + 4\{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)\}^2},$$

$$C = 2\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y).$$

Source: How Directional Distribution (Standard Deviational Ellipse) works, http://help.arcgis.com/en/arcgisdesktop/10.0/help/005p/005p0000001q000000.htm

Standard deviational ellipse in R

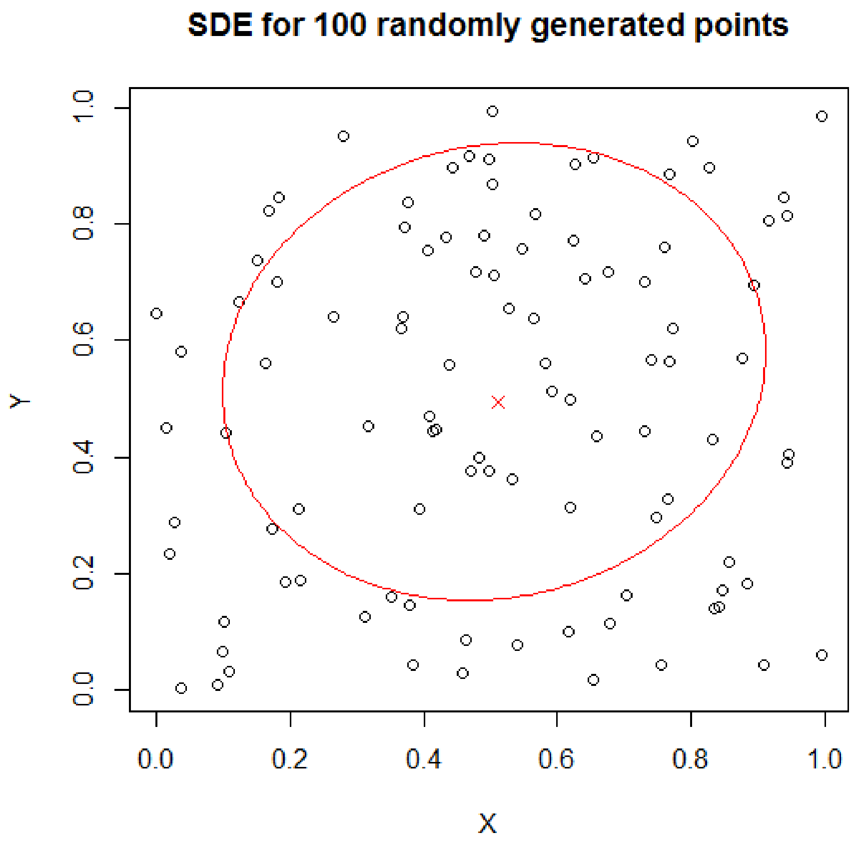
- ArcGIS offers tools for calculating the mean centre point and the standard deviational ellipse.
- CrimeStat III and OpenGeoDa, both of which are freely available, provide easy-to-use functions.
- In R, things are a bit trickier.
  - The `aspace` package can be a good starting point.

Standard deviational ellipse in R

- The following code produces a standard deviational ellipse in R.

```
library(aspace)
coords <- cbind(x = runif(100), y = runif(100))
z <- calc_sde(points = coords)
plot(coords,
  main = "SDE for 100 randomly generated points",
  xlab = "X", ylab = "Y")
lines(z[,2], z[,3], col = "Red")
points(mean(coords[,1]), mean(coords[,2]),
  pch = 4, col = "Red")
```

Standard deviational ellipse in R



Density-based analysis

Quadrat analysis

- Records the number of events in a pattern that occur in a set of cells, or quadrats, of some fixed size.

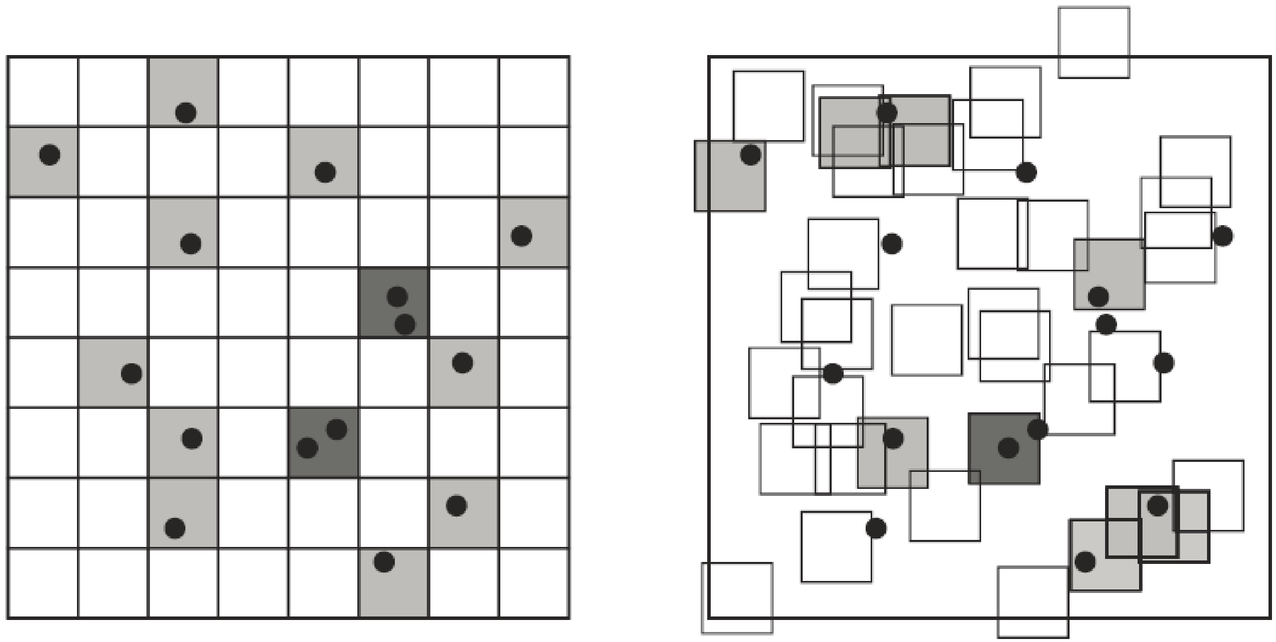


Figure 5.4 The two quadrat count methods: an exhaustive census (left) and random sampling (right). Quadrats containing events are shaded.

Source: p. 128, O’Sullivan and Unwin (2010)

Quadrat analysis

- Random quadrats
  - More frequently applied in field work (e.g., in surveying vegetation in plant ecology)
  - Estimate the likely number of events in a quadrat-shaped region by random sampling
  - Can describe a point pattern without having complete data on the whole pattern, provided that an appropriate sampling approach is used
- Exhaustive census-based quadrats
  - More commonly used in geographic applications such as spatial epidemiology or criminology
  - Can also be hexagonal or triangular

Variance-to-mean ratio

- A frequency distribution indicating how many quadrats contain n events can be constructed by recording the number of events that occur in each quadrat.
- The ratio of variance to the mean, the variance-mean ratio (VMR) can tell whether the pattern is likely to be generated from the Poisson process.

$$\mathrm{VMR} = \frac{\sigma^2}{\mu}$$

  - Only defined when the mean is non-zero, and is generally only used for positive statistics (e.g., count data!)

Variance-to-mean ratio

- If the data were generated from a Poisson distribution:

$$\mathrm{VMR} = 1$$

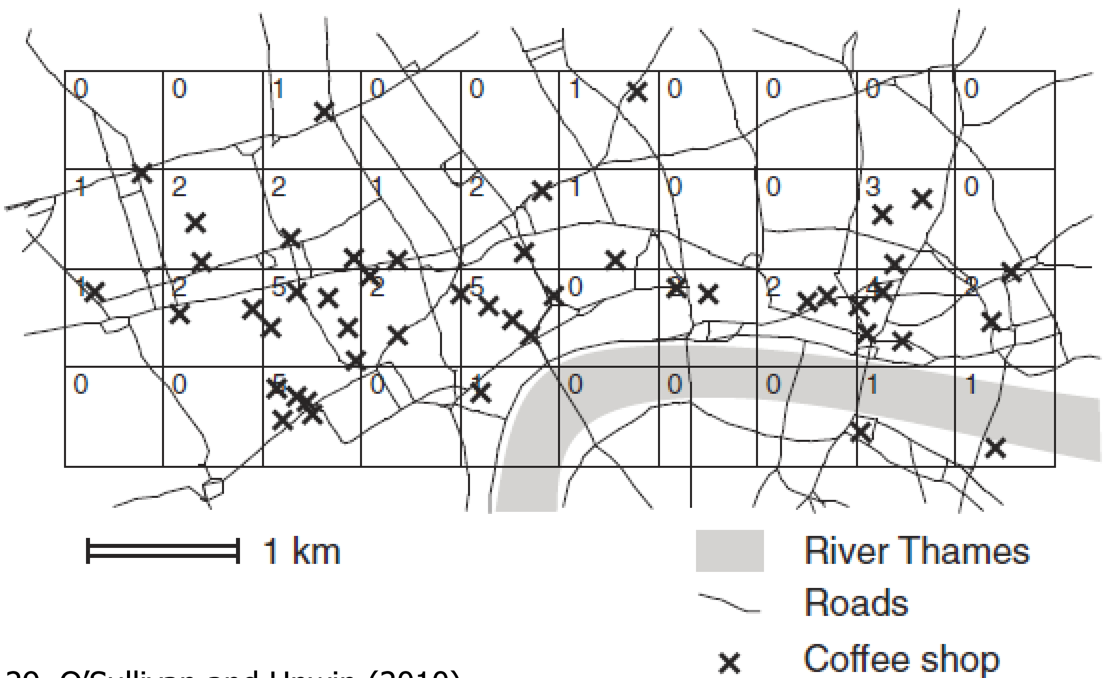
  Because the expected value and variance of a Poisson-distributed random variable are identical.

- In general:
  - $\mathrm{VMR} > 1$ indicates a tendency toward clustering in the pattern.
  - $\mathrm{VMR} < 1$ indicates an evenly spaced arrangement.

Can you explain why?

Coffee shops in London

- The number of coffee shops, $K = 47$, and the number of quadrats, $X = 40$.
  - The mean, $\mu = K/X$, is 1.175.



Source: p. 129, O'Sullivan and Unwin (2010)

Coffee shops in London

- What is VMR for this example?
  - What does it mean?

$$\sigma^2 = \frac{85.775}{40}$$

Table 5.1 Quadrat Counts and Calculation of the Variance for the Coffee Shop Pattern

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**

Source: p. 130, O'Sullivan and Unwin (2010)

Weakness of quadrat analysis

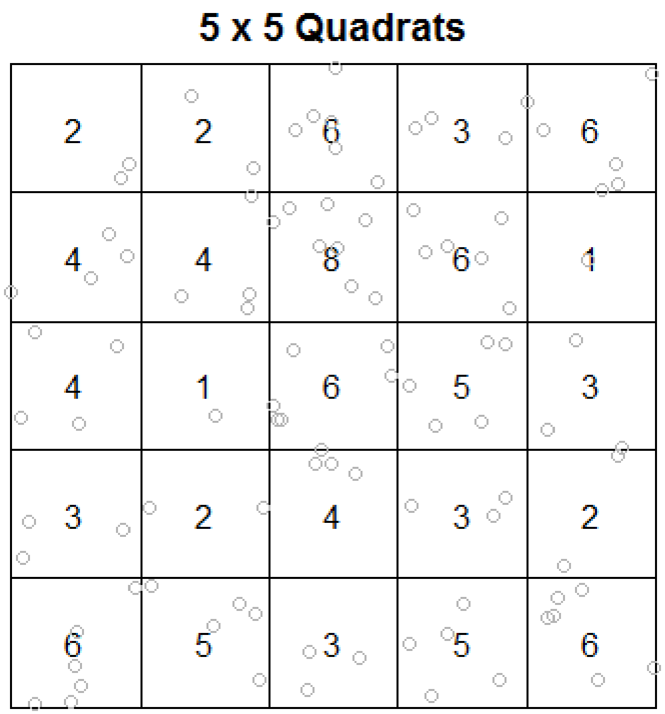
- Sensitive to the choice of quadrat shape and size
  - Large quadrats produce a very coarse description of the pattern.
  - As quadrat size is reduced, many will contain no events and only a few will contain more than one, so the set of counts is not useful as a description of pattern variability.
- A single measure for the entire distribution, so variations within the region are not recognised

Quadrat analysis in R

- You will learn how to perform quadrat analysis in R from the lab:

```
coords.ppp <- ppp(x = coords[,1], y = coords[,2])
build.quad <- quadratcount(coords.ppp, 5, 5)
plot(build.quad, main = "5 x 5 Quadrats")
points(coords.ppp, col = "Grey 70")
quadrat.test(coords.ppp, 5, 5)
#         Chi-squared test of CSR using quadrat counts
# data:  coords.ppp
# X-squared = 20.5, df = 24, p-value = 0.668
# Quadrats: 5 by 5 grid of tiles
# Warning message:
...
```

Quadrat analysis in R



Quadrat analysis in R

- See what happens when we change the number of quadrats:
  - The conclusion changes dramatically!

```
quadrat.test(coords.ppp, 50, 500)
#        Chi-squared test of CSR using quadrat counts
# data:  coords.ppp
# X-squared = 25400, df = 24999, p-value = 0.03698
# Quadrats: 50 by 500 grid of tiles
# Warning message:
# Some expected counts are small; chi^2 approximation
# may be inaccurate
```

References

- Geographic Information Analysis, 2nd Ed. (O'Sullivan & Unwin, 2010, Wiley)