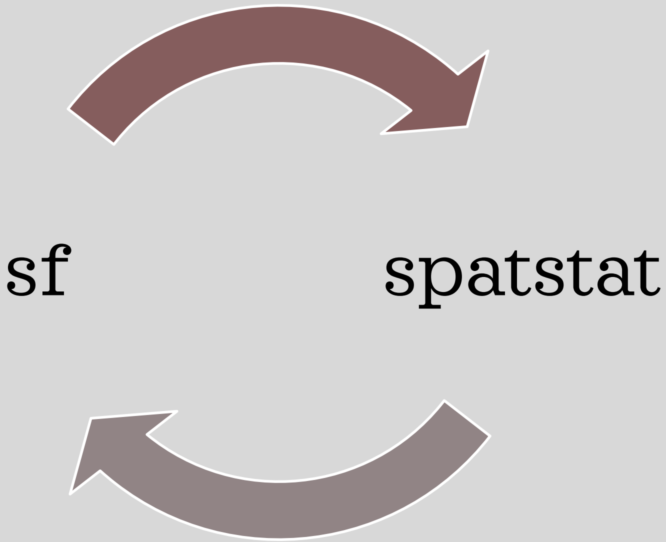
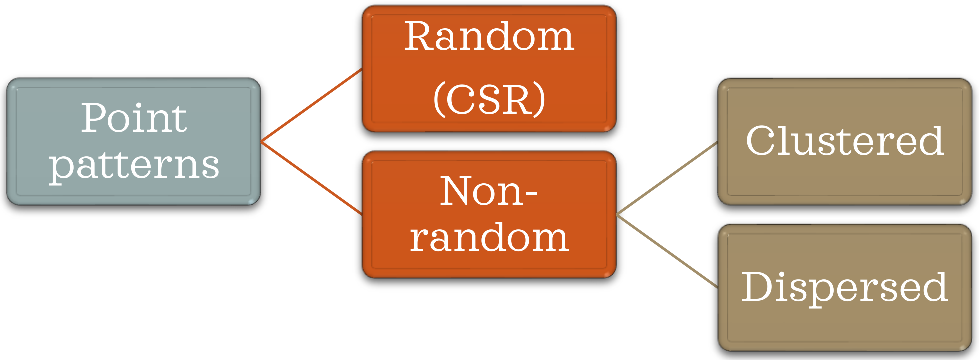
ECON6027 5A

Point Pattern Analysis: Theory

Packages you need

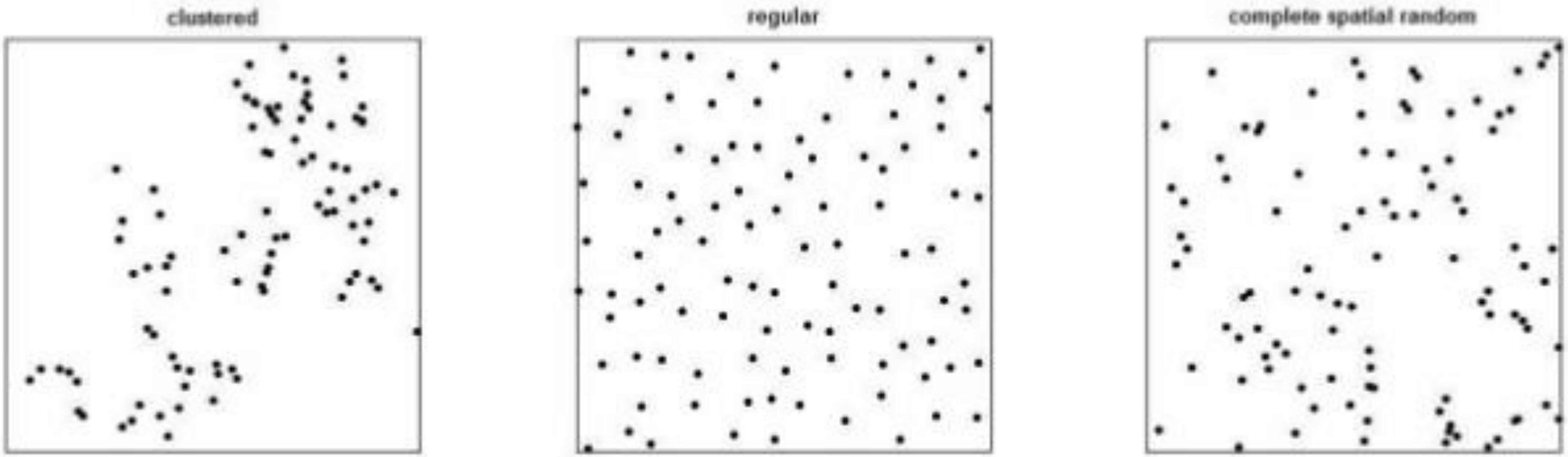


Point pattern analysis: The what?



- The motivation to work with spatial data is partly driven by the need to gain a deep understanding of the spatial structure of a range of phenomena such as **crime incidents, injuries, diseases or bird nesting sites** that are represented by **point features**.
- The focus of point pattern analysis is the examination of any **"static evidence of spacing"**.
  - Thus, the purpose of this lesson is to **explore** the range of approaches that are used to analyse the **distribution patterns of point features**.
- Lack of evidence of spacing is known in the literature as **"complete spatial randomness"** (CSR).

Types of point patterns



$$clustering \xleftarrow{attraction} CSR \xrightarrow{repulsion} regularity$$

Poisson Distribution: revision

- You can use the Poisson distribution when you are interested in counting the number of times an event occurs in an area of opportunity.
  - The probability that an event occurs in one area of opportunity is the same for all areas of opportunity.
  - The number of events that occur in one area of opportunity is independent of the number of events that occur in the other areas of opportunity.
  - The probability that two or more events occur in an area of opportunity approaches zero as the area of opportunity becomes smaller.
  - The average number of events per unit is λ (lambda).
- Probability mass function: If $X \sim Po(\lambda)$,

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

- Properties,
  - $E(X) = \lambda$
  - $Var(X) = \lambda$

Point pattern analysis: The how?

- The **statistical test** for studying point distributions rely on the comparison between an observed spatial pattern and a random theoretical pattern (usually the Poisson distribution).
- The tests are used to determine the probability of the observed pattern under the null that the distribution of points in the study region follows a **homogenous Poisson process (HPP)**.
- For this reason, a CSR process is often referred to as an HPP. However, do bear in mind that HPP implies CSR, but the converse need not be true.

Complete Spatial Randomness (CSR)

Core principles:

First order property: Each event has an equal probability of occurring at any position in the study region, and,

Second order property: The position of any event of occurring is independent of the position of any other.

Goal of a statistical test: test the null hypothesis that the observed pattern is random and is produced by the CSR process either focused on first order or second order property.

How does a *violation of CSR* present itself in the data

1. In the absolute location of a spatial phenomenon, we are bound to encounter a **first-order effect** (no equal probability).

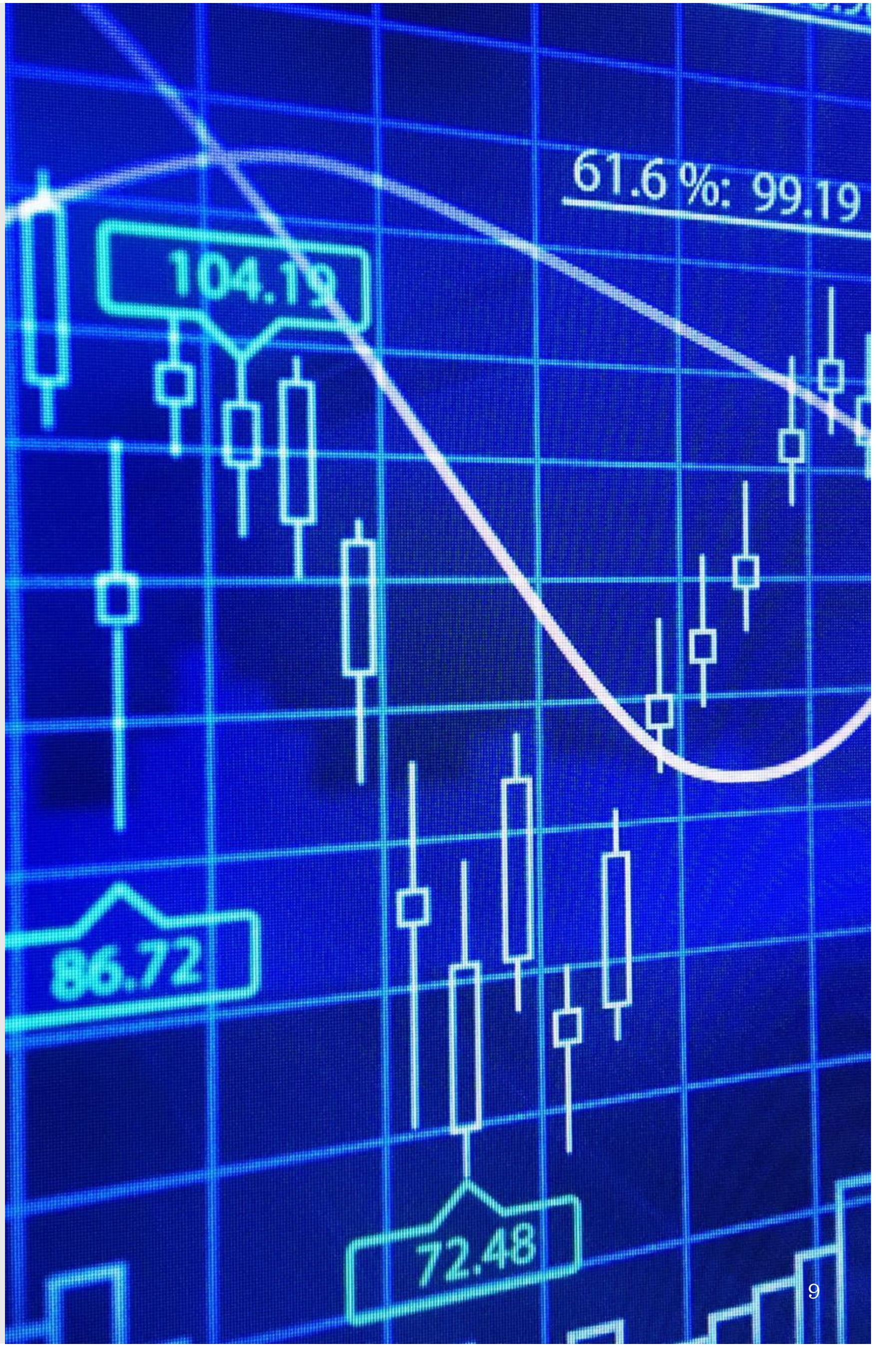
2. If we were to explore the interactions between locations, we are bound to encounter a **second-order effect** (no independence)

Techniques to be discussed:

1. First-order analyses
   - 7b: Quadrat count analysis
   - 7c: Kernel density estimation

2. Second-order analysis
   - 7d: Nearest neighbour analysis
   - 7e: Ripley's K-function analysis

These measures are designed to determine (i) the density of events or (ii) interaction between the locations that develop over space.



Locations of High Dengue Incidence, May 2020, Singapore

Point pattern analysis: The pitfalls

- Impact of scale: magnitude of study and extent
- Modifiable areal unit problem (MAUP)

One may need a apply an appropriate **“edge correction”** in order to reduce the impact of these problems.

Marked vs. Unmarked point patterns

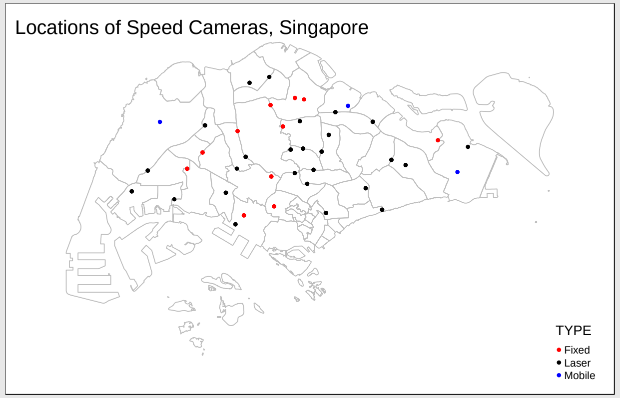
If the dataset we have is only a collection of points, i.e., a set of locations (coordinates) indicating the appearance of a certain event, this is called an **unmarked point pattern**.

If some additional attribute is also observable at the event location? Then we have a **marked point pattern**. Examples:

- Trees in a forest stand. Tree attributes such as diameter, age, height, species, etc are also important as well as understanding whether these attributes are related to the location.
- Type of speed cameras in Singapore.

it is also possible to examine the relationships between the point patterns of different mark types.

- Example: in the 'newhaven' dataset (next up), there are point data for several types of crime. It may be of interest to determine whether forced entry burglaries occur closer to non-forced than one might expect if the two sets of patterns occurred independently.

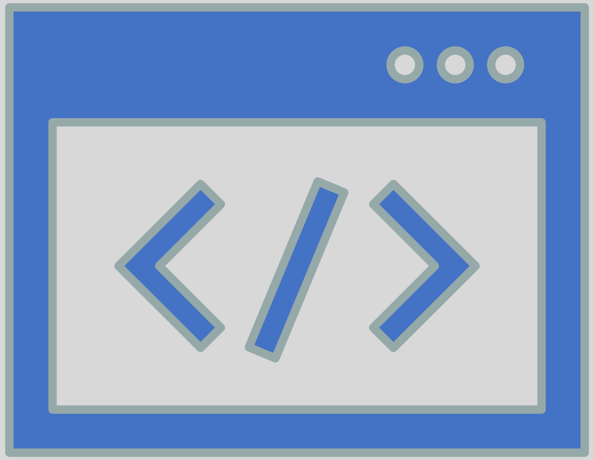


Point pattern analysis using Spatstat

For point pattern analysis we will mainly use the package **“spatstat”**.

In spatstat data is stored in what is called a ppp (planar point pattern) object that incudes,

- Point coordinates
- Boundary of the region



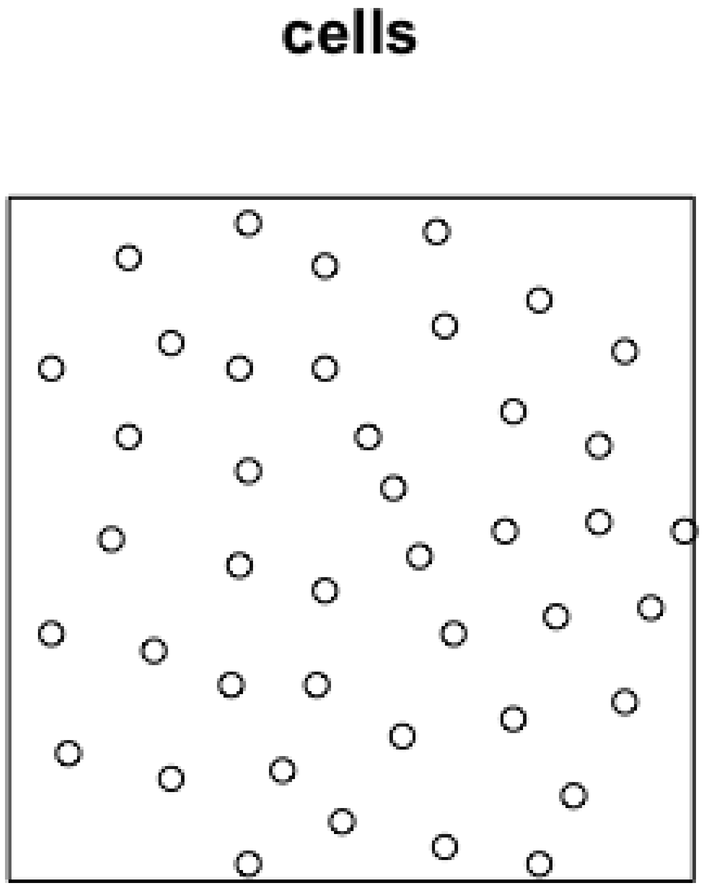
ppp

- Locations of the centres of 42 biological cells observed under optical microscopy in a histological section.

```
> data(cells)

> plot(cells)
```

- Planar point pattern: 42 points
- window: rectangle = [0, 1] x [0, 1] units

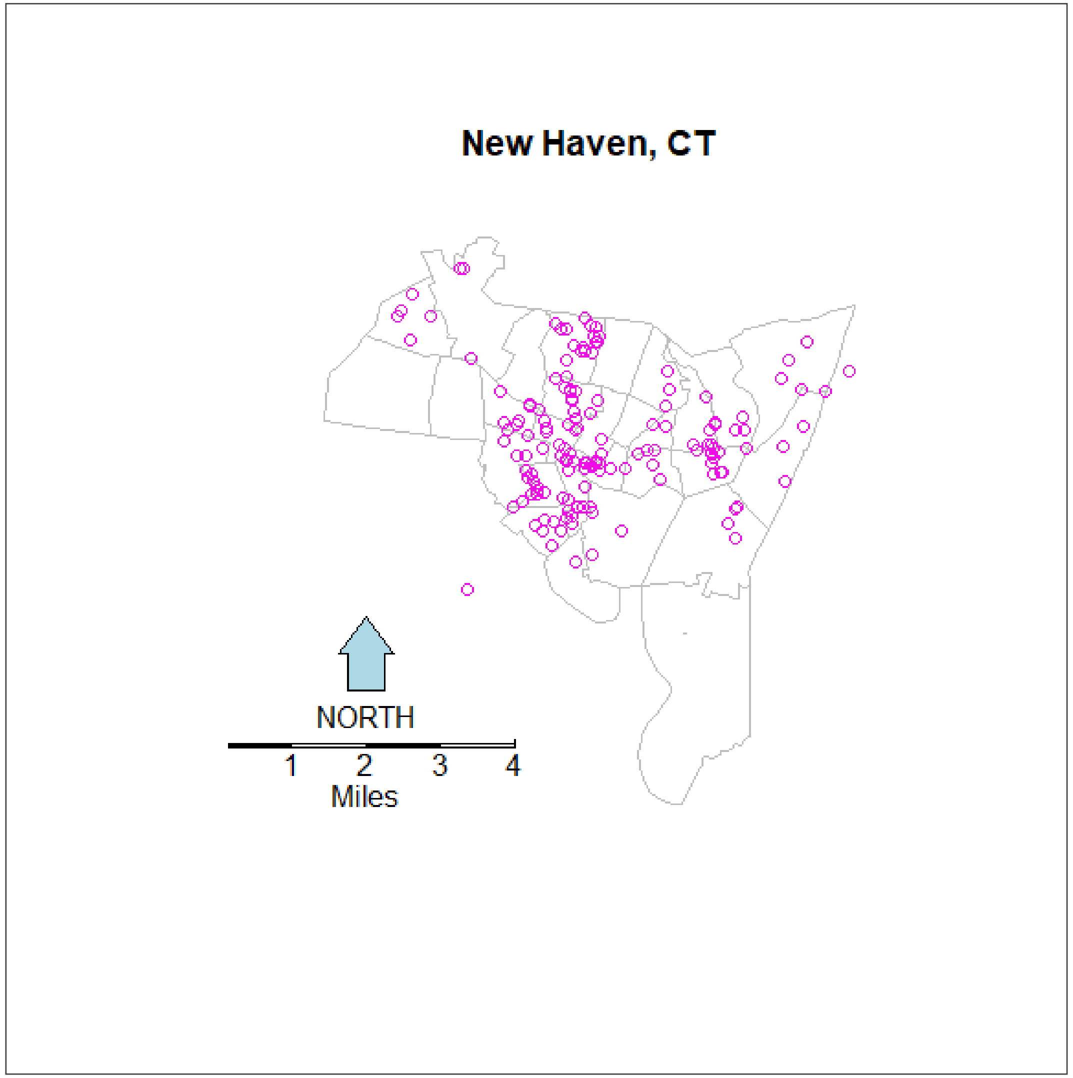


PREP YOUR DATA

Let us see how to prepare a dataset for point pattern analysis using spatstat.

Point geometry to ppp

- As the name suggest, ppp objects **must have a "projected CRS"** as opposed to a geometric CRS.
- The standard ppp format includes,
    i. a set of points and
    ii. a polygon describing the study area *A*.
- When converting a sf object to ppp, there are two aspects to consider:
    i. Point object
    ii. "owin" object created from a Polygons object. The default option is the bounding box of the point data.



Dataset

Incidents of breaches of peace in New Haven:

```
> library(sf)

> breach = st_read("breach.shp") # points

> blocks = st_read("blocks.shp") # polygons

> tm_shape(blocks) + tm_polygons() +
tm_shape(breach) + tm_dots() +
tm_layout(title = "Incidents of breaches of
peace") # plot
```

Point data

- ppp stands for “planar point patterns” which is the class of data that spatstat can process.
- This means spatstat is only able to work with coordinates that are projected.

```
> st_crs(breach)$proj
[1] "lcc"
```

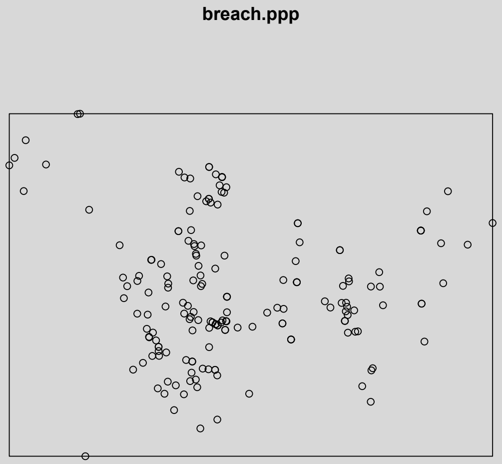
- No problem we can continue. However, if it is FALSE then you must transform your data first using `st_transform()`.

Make ppp object

```
> (breach.ppp = as.ppp(breach))
> plot(breach.ppp)
```

Planar point pattern: 180 points

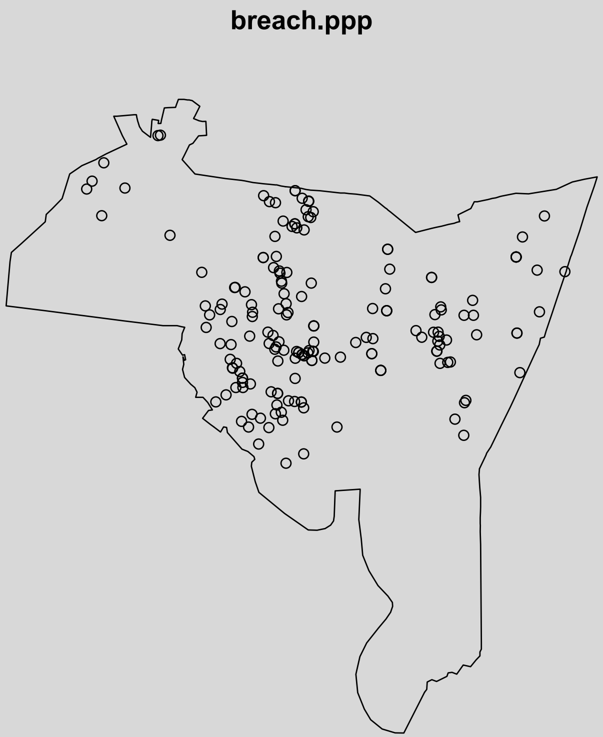
window: rectangle = [536895.7, 569227.2] x [163290.97, 186172.3] units



- However, the boundary is not quite right.

```
> st_bbox(breach)
```

xmin	ymin	xmax	ymax
536895.7	163291.0	569227.2	186172.3



Window

We will use the boundary of New Haven to replace the current rectangular window.

Collapse internal boundaries of New Haven first.

```
> nh = st_union(blocks) # Newhaven outline

> nh.owin = as.owin(nh)

> Window(breach.ppp) = nh.owin

> breach.ppp; plot(breach.ppp)
```

Planar point pattern: 178 points

window: polygonal boundary

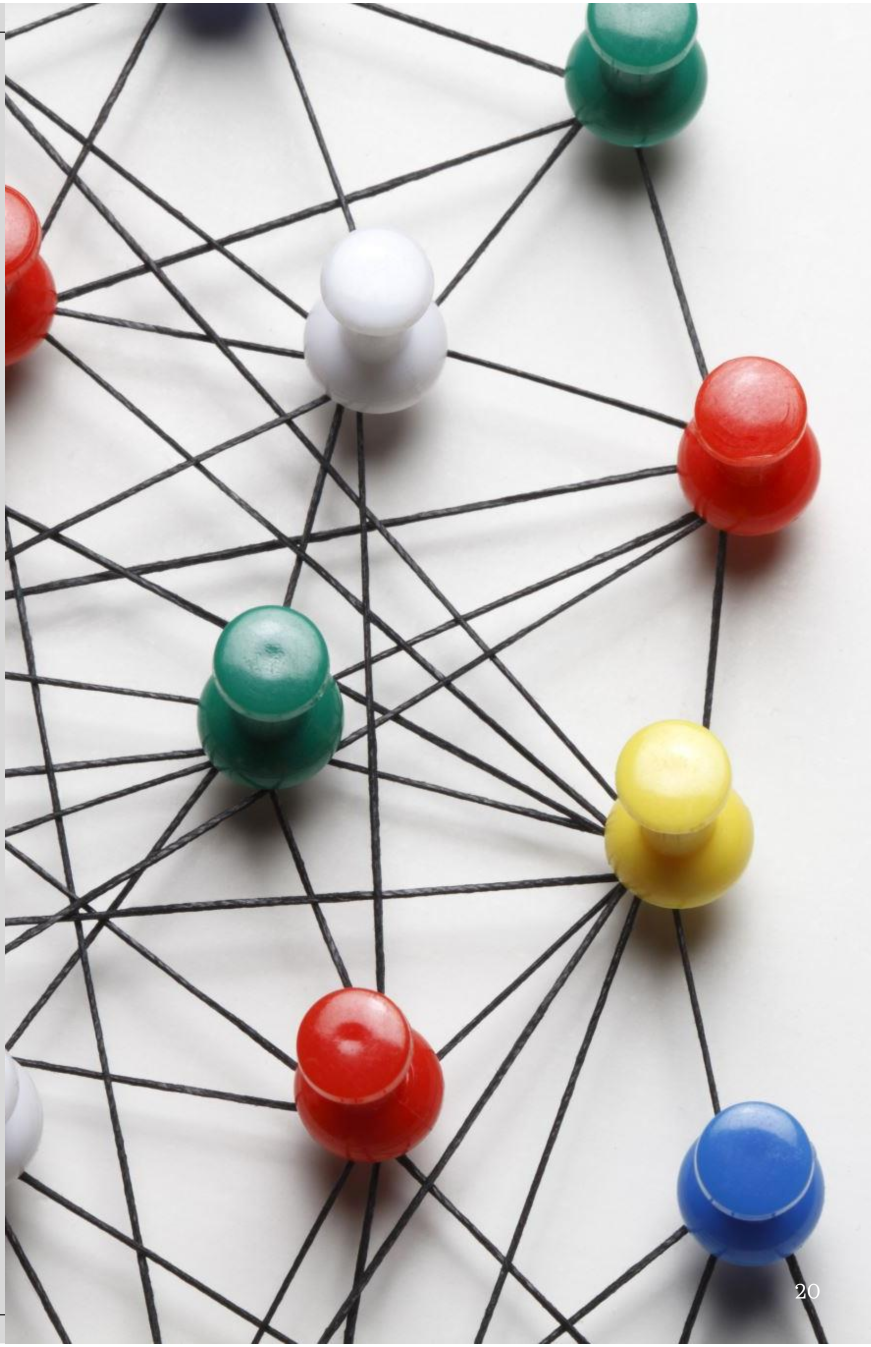
enclosing rectangle: [531731.9, 569625.3] x [147854, 188464.6] units

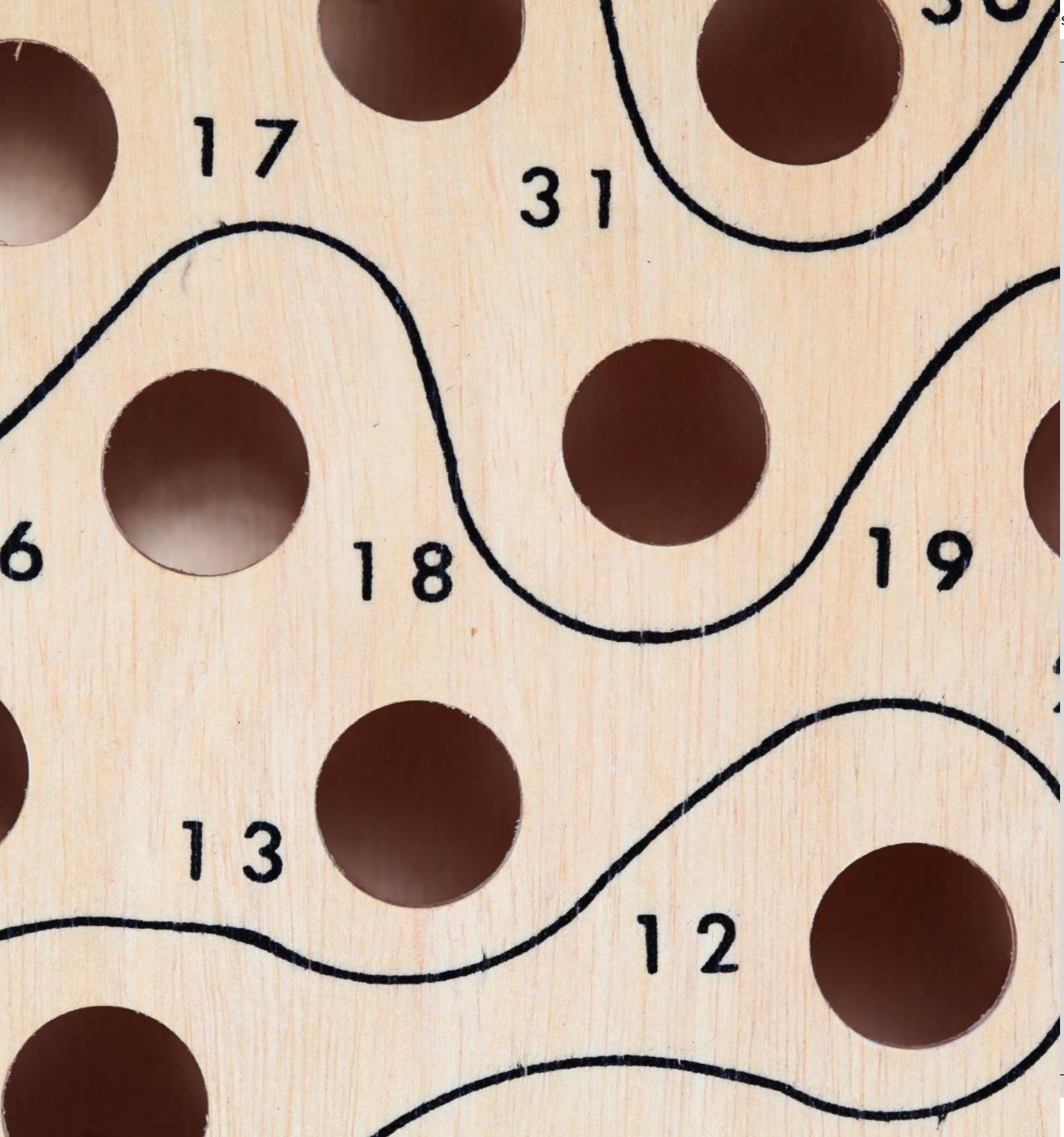
Duplicated points

- If two entries in a dataset are identical, this may or may not be the result of an error.
- Duplication of point coordinates may happen for a variety of reasons and is surprisingly common.
- Check for duplication of points:

```
> duplicated(breach.ppp)
```

- The result is a logical vector, with one entry for each point.
  - TRUE if the current point is identical to an earlier point in the sequence.





What do do about duplicated points

What to do about duplicated points is often unclear; it depends on the context and on the objectives of the analysis. You may,

- Discard duplicated points
- Perturb coordinates slightly
- Make the points of the pattern unique and attach the multiplicities of the points to the pattern as marks.

Since is it possible to have more than one incident of breach of peace in the same location we should use either 2 or 3.

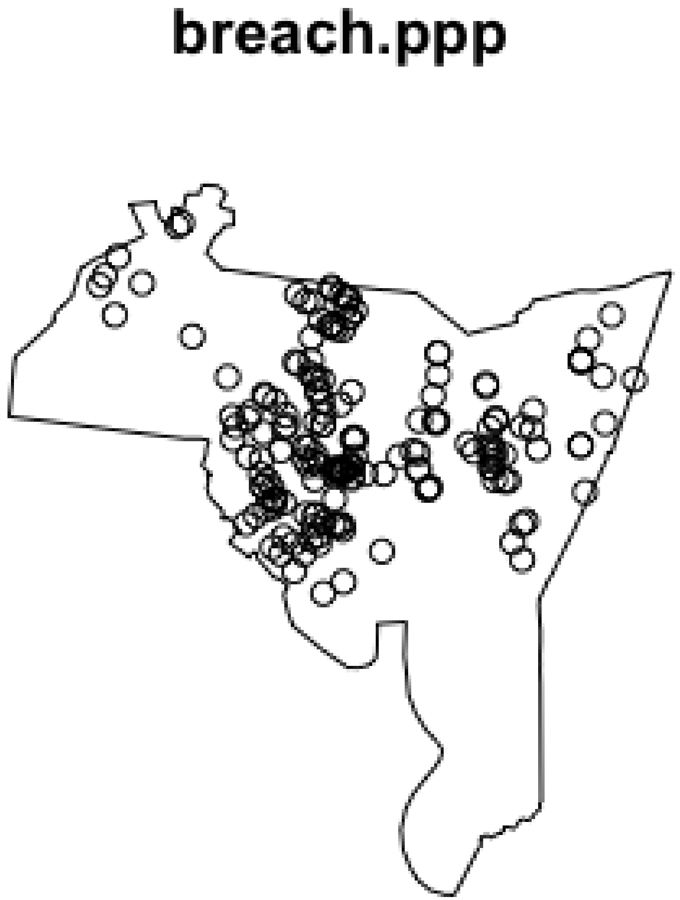
Perturb coordinates

```
> (breach.ppp = rjitter(breach.ppp, retry=T))
```

“retry” argument specifies what to do when a perturbed point lies outside the window of the original point pattern.

If retry=FALSE, the point will be lost; if retry=TRUE, the algorithm will try again.

Since we will not discuss marked point patters in this chapter, I will not take approach 3. (I will show all three methods in a different dataset)



Exercise A

1. Transform the speed camera locations (all types) dataset into a ppp object. (use Singapore boundary to create your window)

Take home points

- What is point pattern analysis
- How do we conduct point pattern analysis in practice
- What is a CSR process and how does it compare to a Poisson process
- Violation of CSR
  - First order effect
  - Second order effect
- Tests for CSR
- What are the potential pitfalls of point pattern analysis
- Basic R packages used for point pattern analysis
- Prep your dataset to ppp to be used in the spatstat package.

References

- ***Spatial Analysis*** by Tonny Oyana, 2nd edition, Chapter 6.
- ***Statistical Methods for Spatial Data Analysis*** by Oliver Schabenberger and Caro A. Gotway, 1st edition, (2005), Chapter 3.
- ***Applied Spatial Data Analysis with R*** by Roger S. Bivand, Edzer Pebesma, and Virgilio Gómez-Rubio, 2nd edition, (2013), Chapter 7.
- https://bookdown.org/lexcomber/brunsdoncomber2e/Ch6.html