Getting started with `spatstat`

Adrian Baddeley, Rolf Turner and Ege Rubak

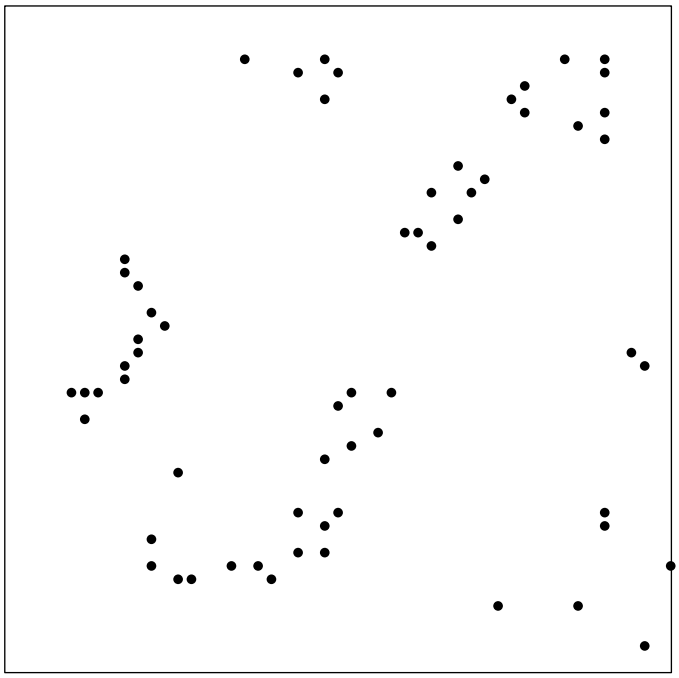
For `spatstat` version `3.6-4`

Welcome to `spatstat`, a package in the R language for analysing spatial point patterns.

This document will help you to get started with `spatstat`. It gives you a quick overview of `spatstat`, and some cookbook recipes for doing basic calculations.

What kind of data does `spatstat` handle?

`Spatstat` is mainly designed for analysing *spatial point patterns*. For example, suppose you are an ecologist studying plant seedlings. You have pegged out a 10×10 metre rectangle for your survey. Inside the rectangle you identify all the seedlings of the species you want, and record their (x, y) locations. You can plot the (x, y) locations:



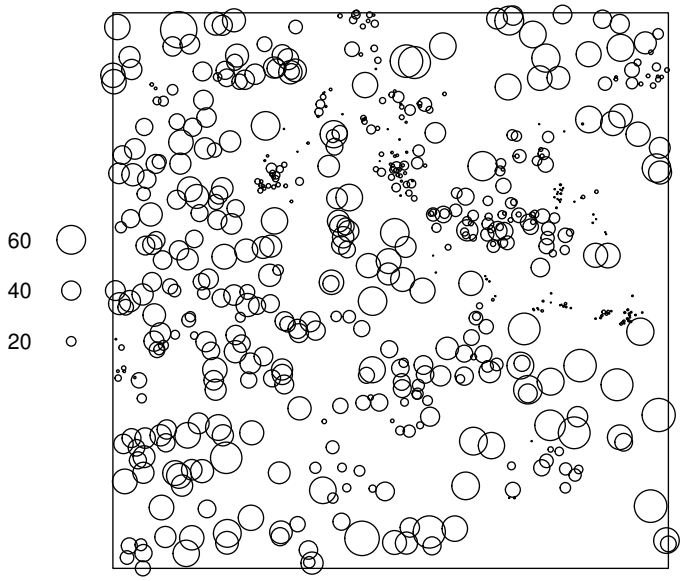
This is a *spatial point pattern* dataset.

Methods for analysing this kind of data are summarised in the highly recommended book by Diggle [4], or our own book [1], or other references in the bibliography below.

Alternatively the points could be locations in one dimension (such as road accidents recorded on a road network) or in three dimensions (such as cells observed in 3D microscopy).

You might also have recorded additional information about each seedling, such as its height, or the number of fronds. Such information, attached to each point in the point pattern, is called a *mark*

variable. For example, here is a stand of pine trees, with each tree marked by its diameter at breast height (dbh). The circle radii represent the dbh values (not to scale).



You might also have recorded supplementary data, such as the terrain elevation, which might serve as explanatory variables. These data can be in any format. `Spatstat` does not usually provide capabilities for analysing such data in their own right, but `spatstat` does allow such explanatory data to be taken into account in the analysis of a spatial point pattern.

`Spatstat` is **not** designed to handle point data where the (x, y) locations are fixed (e.g. temperature records from the state capital cities in Australia) or where the different (x, y) points represent the same object at different times (e.g. hourly locations of a tiger shark with a GPS tag). These are different statistical problems, for which you need different methodology.

What can `spatstat` do?

`Spatstat` supports a very wide range of popular techniques for statistical analysis for spatial point patterns, for example

- kernel estimation of density/intensity
- quadrat counting and clustering indices
- detection of clustering using Ripley's K-function
- spatial logistic regression
- model-fitting
- Monte Carlo tests

as well as some advanced statistical techniques.

`Spatstat` is one of the largest packages available for R, containing over 1000 commands. It is the product of 25 years of software development by leading researchers in spatial statistics.

How do I start using `spatstat`?

1. Install R on your computer

   Go to `r-project.org` and follow the installation instructions.

2. Install the `spatstat` package in your R system

   Start R and type `install.packages("spatstat")`. If that doesn't work, go to `r-project.org` to learn how to install Contributed Packages.

3. Start R

4. Type `library(spatstat)` to load the package.

5. Type `help(spatstat)` for information.

How do I get my data into `spatstat`?

Here is a cookbook example. Suppose you've recorded the (x, y) locations of seedlings, in an Excel spreadsheet. You should also have recorded the dimensions of the survey area in which the seedlings were mapped.

1. In Excel, save the spreadsheet into a comma-separated values (CSV) file.

2. Start R

3. Read your data into R using `read.csv`.

   If your CSV file is called `myfile.csv` then you could type something like

   ```
   > mydata <- read.csv("myfile.csv")
   ```

   to read the data from the file and save them in an object called `mydata` (or whatever you want to call it). You may need to set various options inside the `read.csv()` command to get this to work for your file format: type `help(read.csv)` for information.

4. Check that `mydata` contains the data you expect.

   For example, to see the first few rows of data from the spreadsheet, type

   ```
   > head(mydata)
   ```

   ```
     x    y diameter height
   1 -1.99 0.93        1    1.7
   2 -1.02 0.41        1    1.7
   3 -4.91 1.99        1    1.6
   4 -4.47 1.45        5    4.1
   5 -4.30 0.91        3    3.1
   6 -3.81 0.81        4    4.3
   ```

   To select a particular column of data, you can type `mydata[,3]` to extract the third column, or `mydata$x` to extract the column labelled `x`.

5. Type `library(spatstat)` to load the `spatstat` package

6. Now convert the data to a point pattern object using the `spatstat` command `ppp`.

   Suppose that the `x` and `y` coordinates were stored in columns 3 and 7 of the spreadsheet. Suppose that the sampling plot was a rectangle, with the x coordinates ranging from 100 to 200, and the y coordinates ranging from 10 to 90. Then you would type

   ```
   >   mypattern <- ppp(mydata[,3], mydata[,7], c(100,200), c(10,90))
   ```

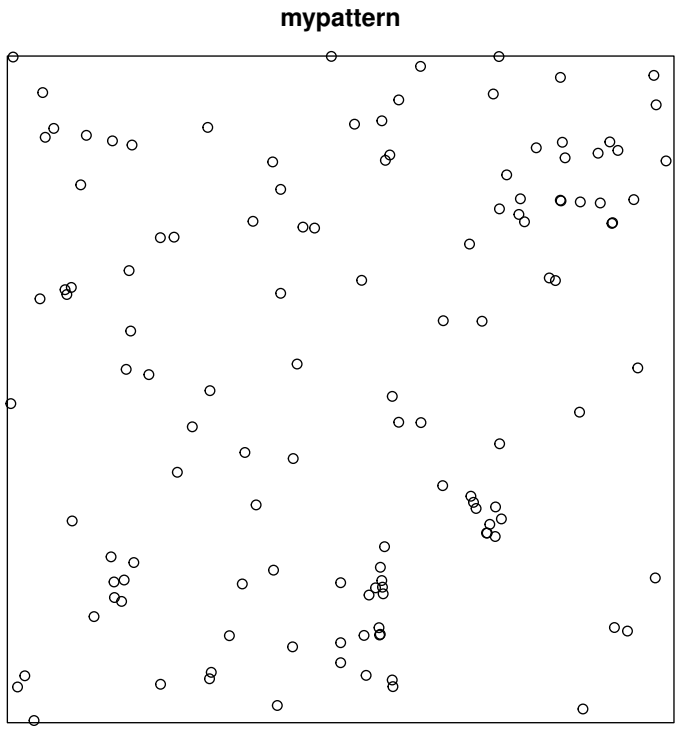
   The general form is

   ```
   > ppp(x.coordinates, y.coordinates, x.range, y.range)
   ```

   Note that this only stores the seedling locations. If you have additional columns of data (such as seedling height, seedling sex, etc) these can be added as *marks*, later.

7. Check that the point pattern looks right by plotting it:

   ```
   > plot(mypattern)
   ```



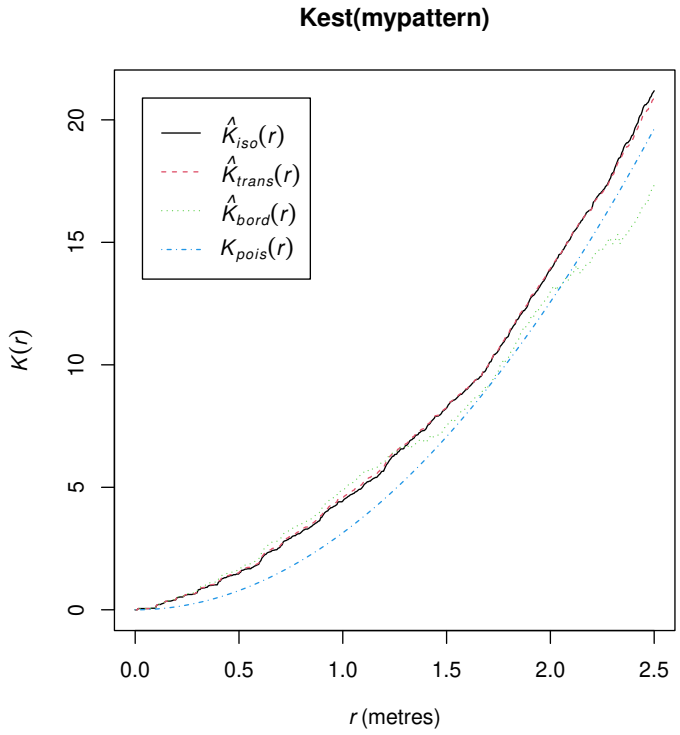
8. Now you are ready to do some statistical analysis. Try the following:

   - Basic summary of data: type

     ```
     > summary(mypattern)
     ```

   - Ripley's K-function:

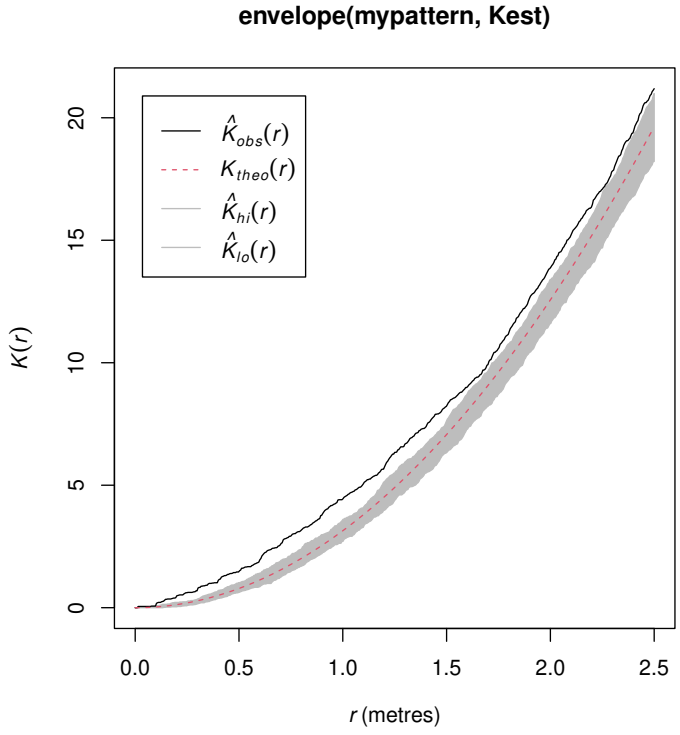
     ```
     > plot(Kest(mypattern))
     ```



For more information, type `help(Kest)`

- Envelopes of K-function:

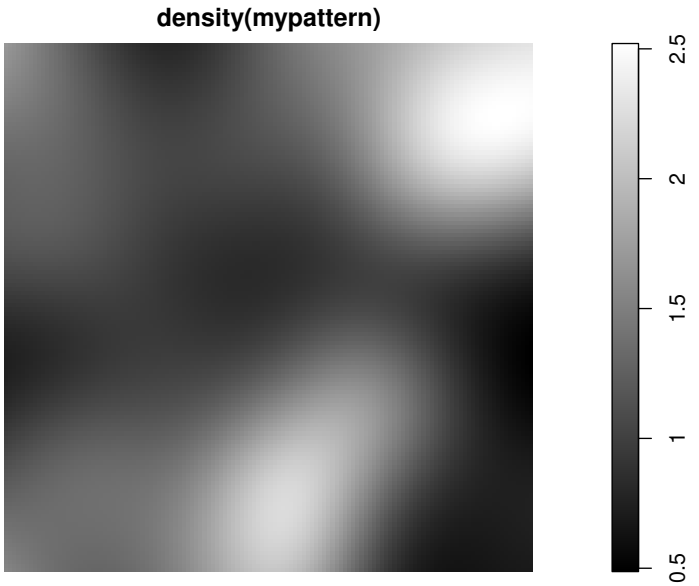
  `> plot(envelope(mypattern,Kest))`



For more information, type `help(envelope)`

- kernel smoother of point density:

```
> plot(density(mypattern))
```



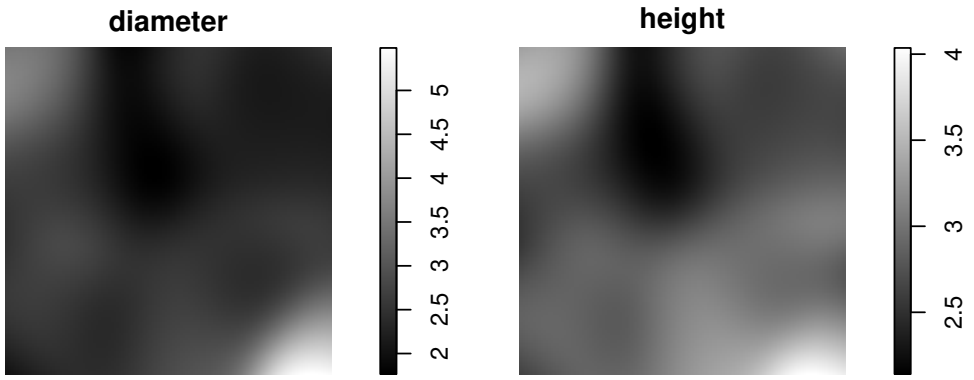
For more information, type `help(density.ppp)`

9. Next if you have additional columns of data recording (for example) the seedling height and seedling sex, you can add these data as *marks*. Suppose that columns 5 and 9 of the spreadsheet contained such values. Then do something like

   ```
   > marks(mypattern) <- mydata[, c(5,9)]
   ```

   Now you can try things like the kernel smoother of mark values:

   ```
   > plot(Smooth(mypattern))
   ```



10. You are airborne! Now look at the book [1] for more hints.

How do I find out which command to use?

Information sources for `spatstat` include:

- the Quick Reference guide: a list of the most useful commands.

  To view the quick reference guide, start R, then type `library(spatstat)` and then `help(spatstat)`. Alternatively you can download a pdf of the Quick Reference guide from the website `www.spatstat.org`

- online help:

  The online help files are useful — they give detailed information and advice about each command. They are available when you are running `spatstat`. To get help about a particular command `blah`, type `help(blah)`. There is a graphical help interface, which you can start by typing `help.start()`. Alternatively you can download a pdf of the entire manual (1000 pages!) from the website `www.spatstat.org`.

- vignettes:

`Spatstat` comes installed with several 'vignettes' (introductory documents with examples) which can be accessed using the graphical help interface. They include a document about `Handling shapefiles`.

- book:

  Our book [1] contains a complete course on `spatstat`.

- website:

  Visit the `spatstat` package website `www.spatstat.org`

- forums:

  Join the forum `R-sig-geo` by visiting `r-project.org`. Then email your questions to the forum. Alternatively you can ask the authors of the `spatstat` package (their email addresses are given in the package documentation).

References

[1] A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R.* Chapman & Hall/CRC Press, 2015.

[2] R. Bivand, E.J. Pebesma, and V. Gómez-Rubio. *Applied spatial data analysis with R.* Springer, 2008.

[3] N.A.C. Cressie. *Statistics for Spatial Data.* John Wiley and Sons, New York, second edition, 1993.

[4] P.J. Diggle. *Statistical Analysis of Spatial Point Patterns.* Hodder Arnold, London, second edition, 2003.

[5] M.J. Fortin and M.R.T. Dale. *Spatial analysis: a guide for ecologists.* Cambridge University Press, Cambridge, UK, 2005.

[6] A.S. Fotheringham and P.A. Rogers, editors. *The SAGE Handbook on Spatial Analysis.* SAGE Publications, London, 2009.

[7] C. Gaetan and X. Guyon. *Spatial statistics and modeling.* Springer, 2009. Translated by Kevin Bleakley.

[8] A.E. Gelfand, P.J. Diggle, M. Fuentes, and P. Guttorp, editors. *Handbook of Spatial Statistics.* CRC Press, 2010.

[9] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns.* John Wiley and Sons, Chichester, 2008.

[10] J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton, 2004.

[11] D.U. Pfeiffer, T. Robinson, M. Stevenson, K. Stevens, D. Rogers, and A. Clements. *Spatial analysis in epidemiology.* Oxford University Press, Oxford, UK, 2008.

[12] L.A. Waller and C.A. Gotway. *Applied spatial statistics for public health data.* Wiley, 2004.