Datasets provided for `spatstat`

Adrian Baddeley, Rolf Turner and Ege Rubak

For `spatstat` version 1.64-1

This document is an overview of the spatial datasets that are provided for the `spatstat` package.

To flick through a nice display of all the data sets that come with `spatstat` type `demo(data)`. To see information about a given data set, type `help(`*`name`*`)` where *name* is the name of the data set. To plot a given data set, type `plot(`*`name`*`)`.

Datasets in `spatstat` are "lazy-loaded", which means that they can be accessed simply by typing their name. Not all packages do this; in some packages you have to type `data(`*`name`*`)` in order to access a data set.

To list all the datasets in `spatstat`, you need to type `data(package="spatstat.data")`. This is because, for efficiency, the datasets are actually installed in a sub-package `spatstat.data`. This is the only time you should ever need to mention `spatstat.data` explicitly. When the `spatstat` package is loaded by the command `library(spatstat)`, the sub-package `spatstat.data` is automatically loaded.

1 List of datasets

1.1 Point patterns in 2D

Here is a list of the standard point pattern data sets that are supplied with the current installation of `spatstat.data`:

name	description	marks	covariates	window
amacrine	rabbit amacrine cells	cell type	·	▭
anemones	sea anemones	diameter	·	▭
ants	ant nests	species	zones	⬡
bdspots	breakdown spots	·	·	◯
bei	rainforest trees	·	topography	▭
betacells	cat retinal ganglia	cell type, area	·	▭
bramblecanes	bramble canes	age	·	▭
bronzefilter	bronze particles	diameter	·	▭
cells	biological cells	·	·	▭
chorley	cancers	case/control	·	▨
clmfires	forest fires	cause, size, date	elevation, orientation, slope, land use	▨
copper	copper deposits	·	fault lines	▭
demopat	artificial data	type	·	▨
finpines	trees	diam, height	·	▭
gordon	people in a park	·	·	▨
gorillas	gorilla nest sites	group, season	terrain, vegetation, heat, water	▨
hamster	hamster tumour cells	cell type	·	▭
humberside	child leukaemia	case/control	·	▨
hyytiala	mixed forest	species	·	▭
japanesepines	Japanese pines	·	·	▭
lansing	mixed forest	species	·	▭
longleaf	trees	diameter	·	▭
mucosa	gastric mucosa cells	cell type	·	▭
murchison	gold deposits	·	faults, rock type	▨
nbfires	wildfires	several	·	▨
nztrees	trees	·	·	▭
paracou	trees	adult/juvenile	·	▭
ponderosa	trees	·	·	▭
redwood	saplings	·	·	▭
redwood3	saplings	·	·	▭
redwoodfull	saplings	·	zones	▭
shapley	galaxies	magnitude, recession, SE	·	⬡
simdat	simulated pattern	·	·	▭
sporophores	fungi	species	·	◯
spruces	trees	diameter	·	▭
swedishpines	trees	·	·	▭
urkiola	mixed forest	species	·	▨
vesicles	synaptic vesicles	·	zones	▨
waka	trees	diameter	·	▭

The shape of the window containing the point pattern is indicated by the symbols ▭ (rectangle), ◯ (disc), ⬡ (convex polygon) and ▨ (irregular polygon).

Additional information about the data set *name* may be stored in a separate list *name*.extra. Currently these are the available options:

Name	Contents
ants.extra	field and scrub subregions;
	additional map elements; plotting function
bei.extra	covariate images
chorley.extra	incinerator location; plotting function
gorillas.extra	covariate images
nbfires.extra	inscribed rectangle; border type labels
ponderosa.extra	data points of interest; plotting function
redwoodfull.extra	subregions; plotting function
shapley.extra	individual survey fields; plotting function
vesicles.extra	anatomical regions

For demonstration and instruction purposes, raw data files are available for the datasets `vesicles`, `gorillas` and `osteo`.

1.2 Other Data Types

There are also the following spatial data sets which are not 2D point patterns:

name	description	format
austates	Australian states	tessellation
cetaceans	marine survey	replicated 2D point patterns
chicago	crimes	point pattern on linear network
demohyper	simulated data	replicated 2D point patterns with covariates
dendrite	dendritic spines	point pattern on linear network
flu	virus proteins	replicated 2D point patterns
heather	heather mosaic	binary image (three versions)
osteo	osteocyte lacunae	replicated 3D point patterns with covariates
pyramidal	pyramidal neurons	replicated 2D point patterns in 3 groups
residualspaper	data & code from Baddeley et al (2005)	2D point patterns, R function
simba	simulated data	replicated 2D point patterns in 2 groups
spiders	spider webs	point pattern on linear network
waterstriders	insects on water	replicated 2D point patterns

Additionally there is a dataset `Kovesi` containing several colour maps with perceptually uniform contrast.

2 Information on each dataset

Here we give basic information about each dataset. For further information, consult the help file for the particular dataset.

`amacrine`: Amacrine cells

Locations of displaced amacrine cells in the retina of a rabbit. There are two types of points, "on" and "off".

```
> plot(amacrine)
```

amacrine
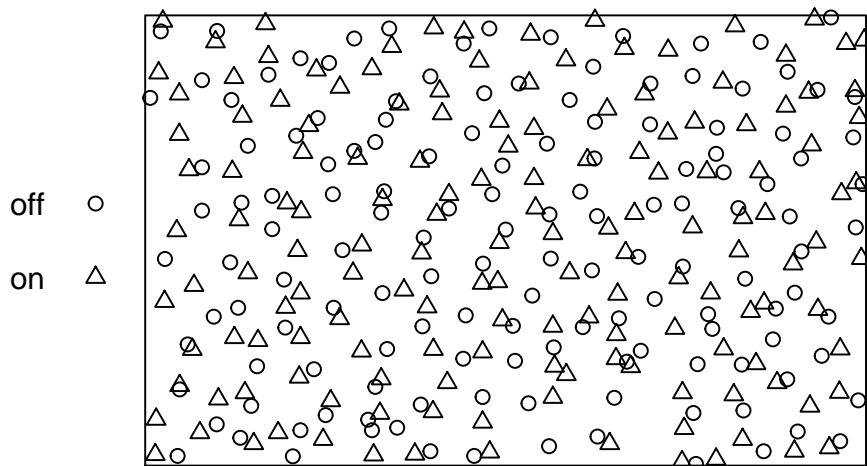
anemones: Sea Anemones

These data give the spatial locations and diameters of sea anemones on a boulder near sea level.

```
> plot(anemones, markscale=1)
```

anemones
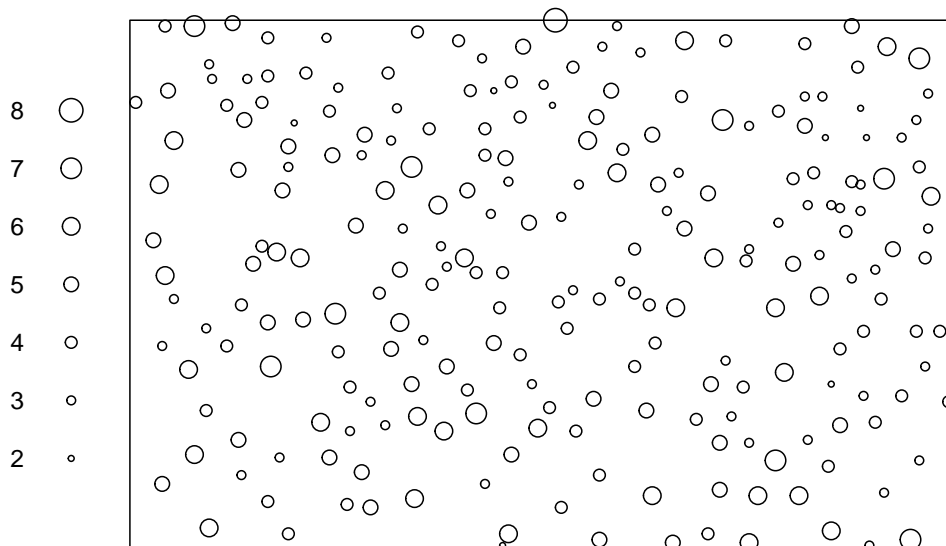
ants: Ants' nests

Spatial locations of nests of two species of ants at a site in Greece. The full dataset (supplied here) has an irregular polygonal boundary, while most analyses have been confined to two rectangular subsets of the pattern (also supplied here).

```
> ants.extra$plotit()
```

ants

Cataglyphis ○

Messor △
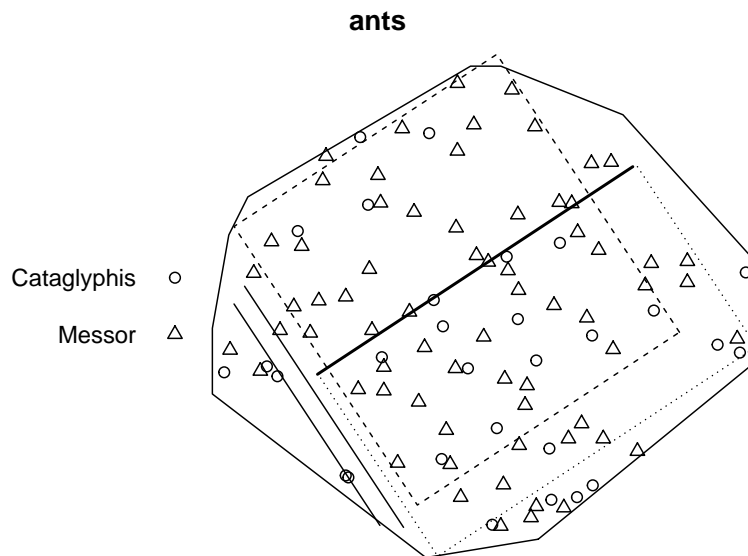
austates: **Australian states**

The states and large mainland territories of Australia are represented as polygonal regions forming a tessellation.

```
> plot(austates)
```


austates

bdspots: **Breakdown spots**

A list of three point patterns, each giving the locations of electrical breakdown spots on a circular electrode in a microelectronic capacitor.

```
> plot(bdspots, equal.scales=TRUE, pch="+",
+       panel.args=function(i)list(cex=c(0.15, 0.2, 0.7)[i]))
```
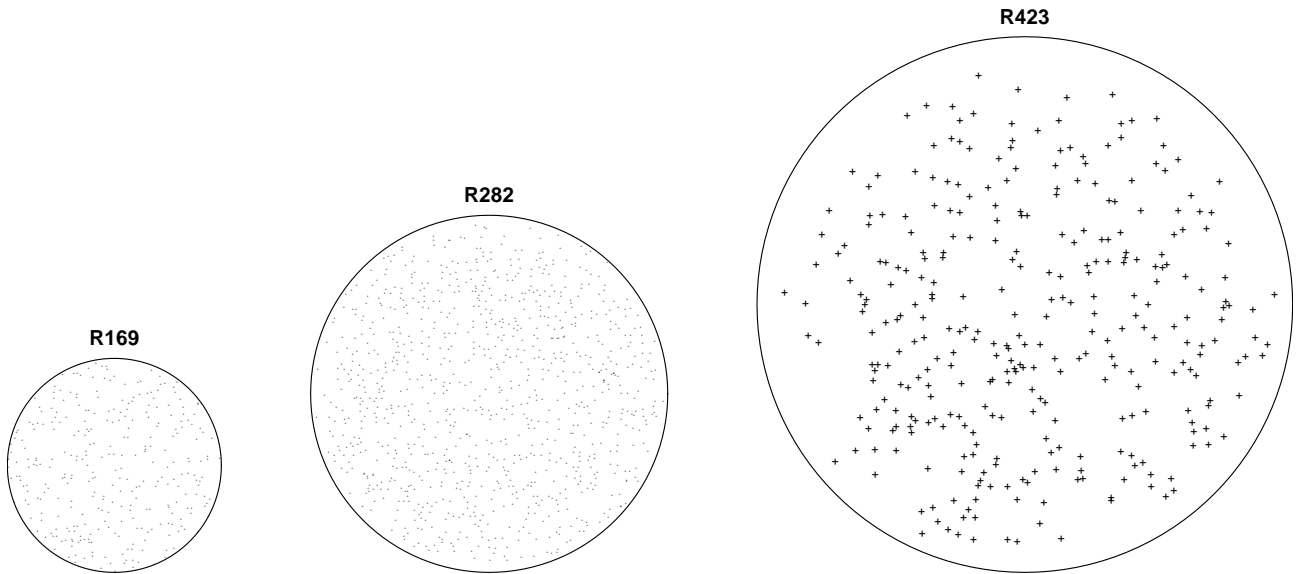
bei: Beilschmiedia data

Locations of 3605 trees in a tropical rain forest. Accompanied by covariate data giving the elevation (altitude) and slope of elevation in the study region.

```
> plot(bei.extra$elev, main="Beilschmiedia")
> plot(bei, add=TRUE, pch=16, cex=0.3)
```

Beilschmiedia
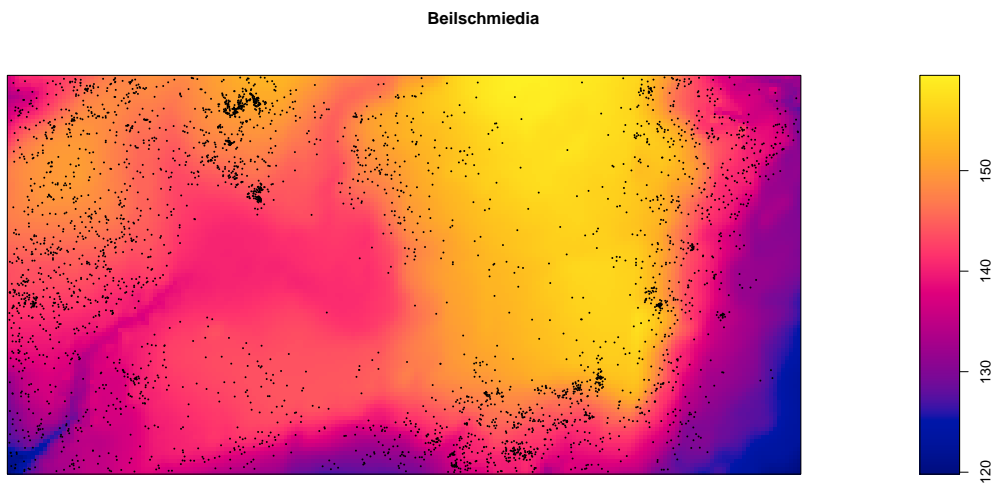

```
> M <- persp(bei.extra$elev,
+            theta=-45, phi=18, expand=7,
+            border=NA, apron=TRUE, shade=0.3,
+            box=FALSE, visible=TRUE,
+            main="")
> perspPoints(bei, Z=bei.extra$elev, M=M, pch=16, cex=0.3)
```

6
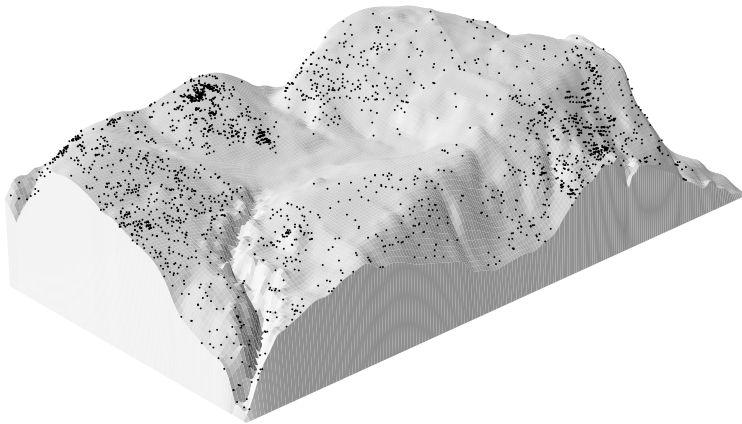
betacells: Beta ganglion cells

Locations of beta ganglion cells in cat retina, each cell classified as 'on' or 'off' and also labelled with the cell profile area.

```
> plot(betacells)
```
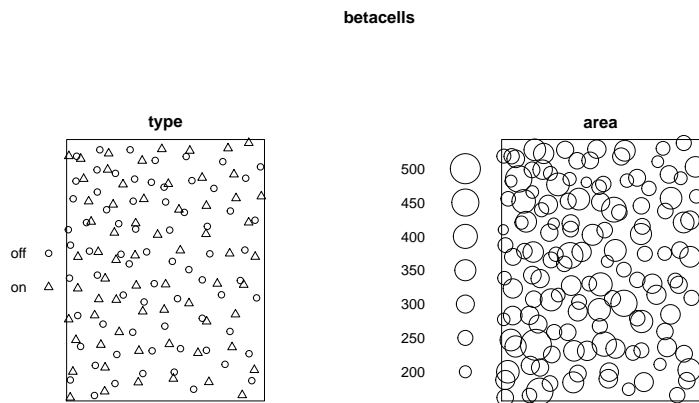

bramblecanes: Bramble canes

```
> plot(bramblecanes, cols=1:3)
```
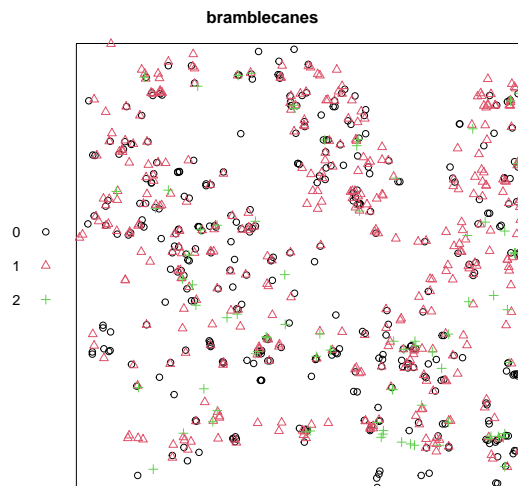
bramblecanes

```
> plot(split(bramblecanes))
```


split(bramblecanes)
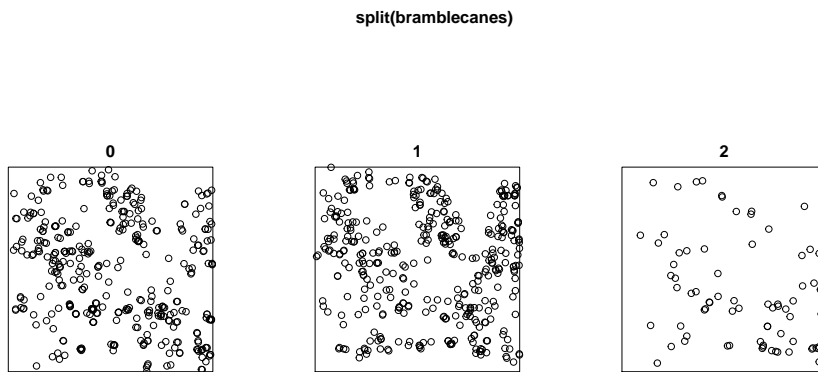
bronzefilter: Bronze filter section profiles

Spatially inhomogeneous pattern of circular section profiles of particles, observed in a longitudinal plane section through a gradient sinter filter made from bronze powder.

```
> plot(bronzefilter,markscale=2)
```

bronzefilter
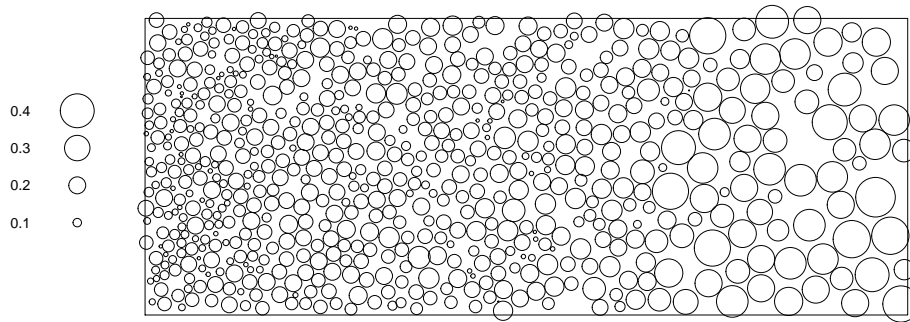
cells: Biological cells

Locations of the centres of 42 biological cells observed under optical microscopy in a histological section. Often used as a demonstration example.

```
> plot(cells)
```

cells
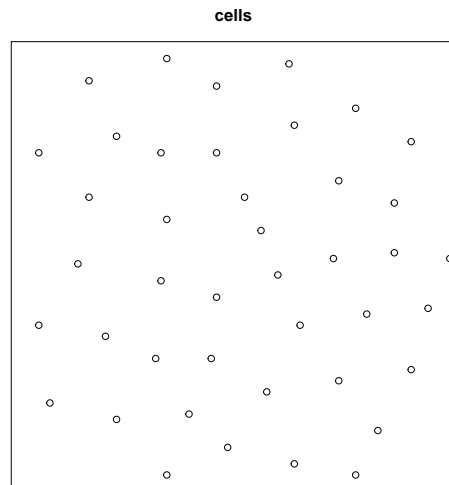
cetaceans: Survey of marine species

Recorded sightings of whales, dolphins and other marine species in a series of surveys. Replicated 2D marked point patterns.

```
> plot(cetaceans.extra$patterns, main="Cetaceans data", cols=1:5, hsep=1)
```

Cetaceans data
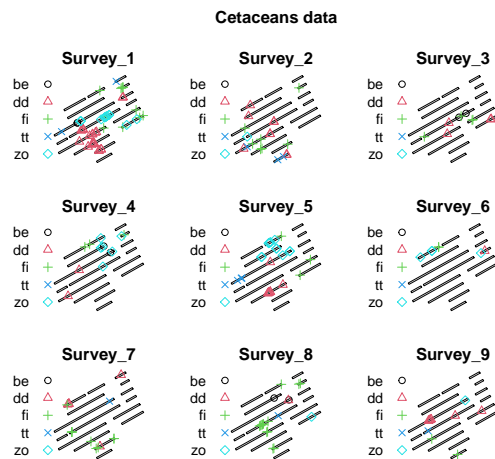
chicago: Chicago crimes

Locations (street addresses) of crimes reported in a two-week period in an area close to the University of Chicago. A multitype point pattern on a linear network.

```
> plot(chicago, main="Chicago Crimes", col="grey",
+       cols=c("red", "blue", "black", "blue", "red", "blue", "blue"),
+       chars=c(16,2,22,17,24,15,6), leg.side="left", show.window=FALSE)
```
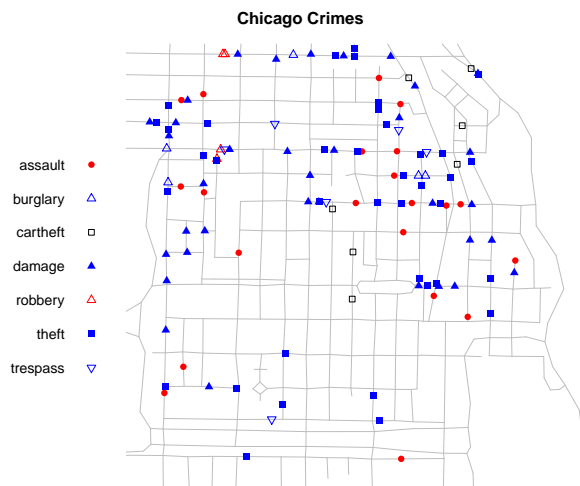

chorley: Chorley-Ribble cancer data

Spatial locations of cases of cancer of the larynx and cancer of the lung, and the location of a disused industrial incinerator. A marked point pattern, with an irregular window and a simple covariate.

```
> chorley.extra$plotit()
```

Chorley–Ribble Data
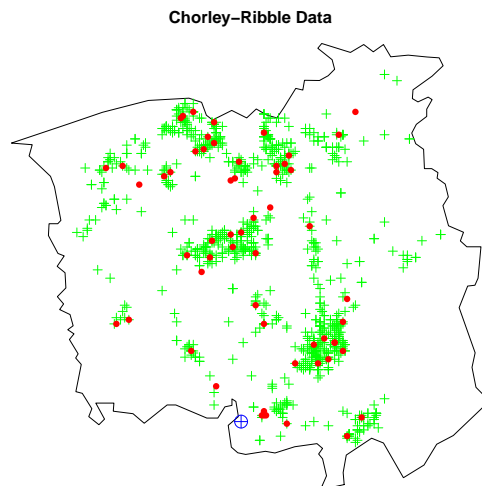
clmfires: Castilla-La Mancha Fires

Forest fires in the Castilla-La Mancha region of Spain between 1998 and 2007. A point pattern with 4 columns of marks:

cause	cause of fire
burnt.area	total area burned, in hectares
date	date of fire
julian.date	date of fire in days since 1.1.1998

```
> plot(clmfires, which.marks="cause", cols=2:5, cex=0.25,
+      main="Castilla-La Mancha forest fires")
```

Castilla–La Mancha forest fires
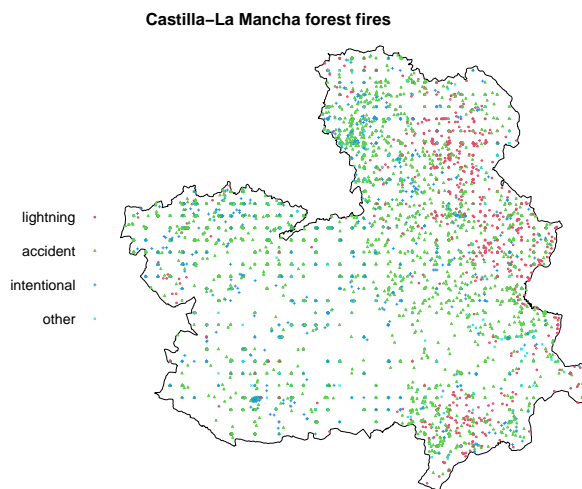
The accompanying dataset `clmfires.extra` is a list of two items `clmcov100` and `clmcov200` containing covariate information for the entire Castilla-La Mancha region. Each of these two elements is a list of four pixel images named `elevation`, `orientation`, `slope` and `landuse`.

```
> plot(clmfires.extra$clmcov200, main="Covariates for forest fires")
```

11

Covariates for forest fires
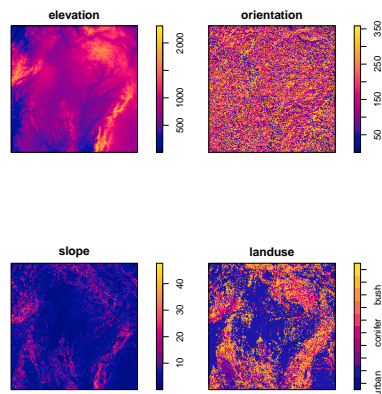
copper: Queensland copper data

These data come from an intensive geological survey in central Queensland, Australia. They consist of 67 points representing copper ore deposits, and 146 line segments representing geological 'lineaments', mostly faults.

```
> plot(copper$Points, main="Copper")
> plot(copper$Lines, add=TRUE)
```


Copper
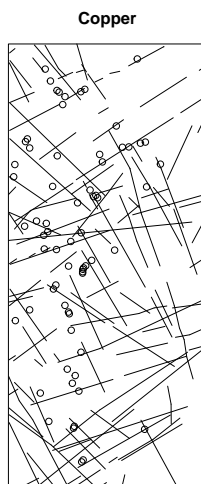
demohyper

A synthetic example of a `hyperframe` for demonstration purposes.

```
> plot(demohyper, quote({ plot(Image, main=""); plot(Points, add=TRUE) }),
+        parargs=list(mar=rep(1,4)))
```

demohyper
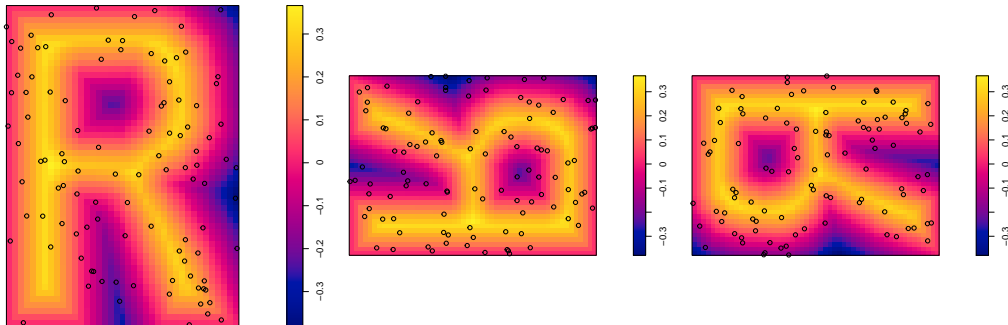
demopat

A synthetic example of a point pattern for demonstration purposes.

```
> plot(demopat)
```


demopat
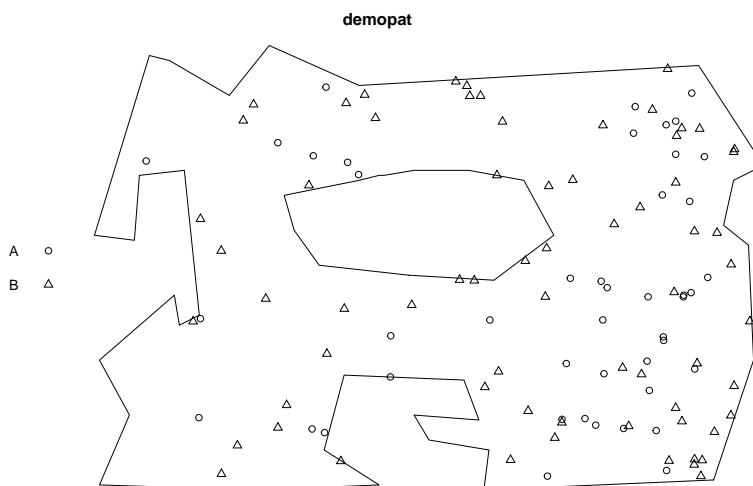
dendrite

Dendrites are branching filaments which extend from the main body of a neuron (nerve cell) to propagate electrochemical signals. Spines are small protrusions on the dendrites.

This dataset gives the locations of 566 spines observed on one branch of the dendritic tree of a rat neuron. The spines are classified according to their shape into three types: mushroom, stubby or thin.

```
> plot(dendrite, leg.side="bottom", main="", cex=0.75, cols=2:4)
```

mushroom   stubby   thin
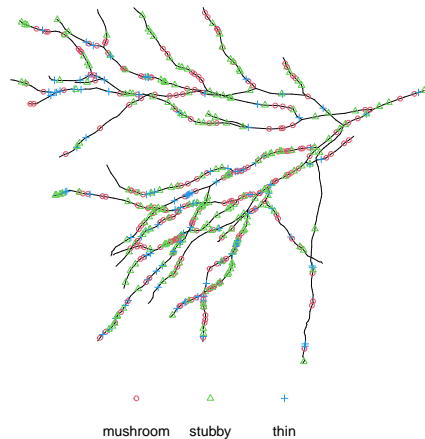
`finpines`: **Finnish pine saplings**

Locations of 126 pine saplings in a Finnish forest, their heights and their diameters.

```
> plot(finpines, main="Finnish pines")
```

Finnish pines
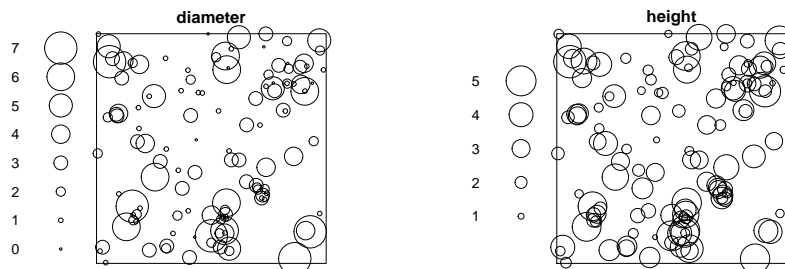
`flu`: **Influenza virus proteins**

The `flu` dataset contains replicated spatial point patterns giving the locations of two different virus proteins on the membranes of cells infected with influenza virus.

It is a `hyperframe` containing point patterns and explanatory variables.

```
> wildM1 <- with(flu, virustype == "wt" & stain == "M2-M1")
> plot(flu[wildM1, 1, drop=TRUE],
+       main=c("flu data", "wild type virus, M2-M1 stain"),
+       chars=c(16,3), cex=0.4, cols=2:3)
```

14

flu data
wild type virus, M2–M1 stain
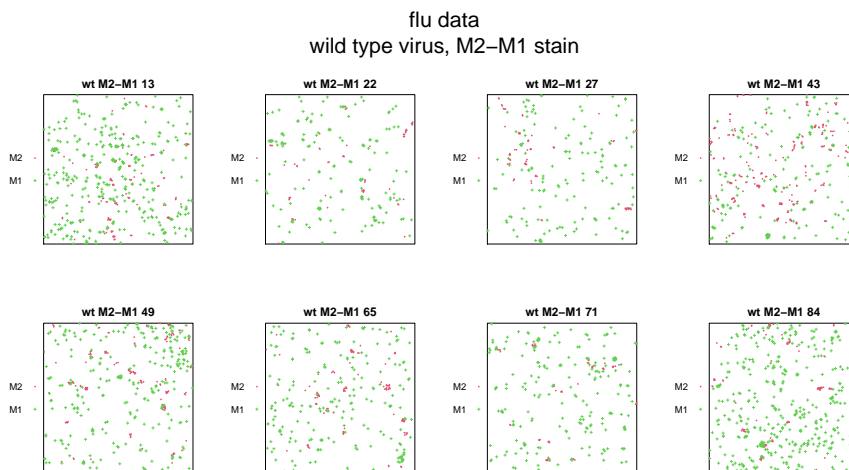
gordon: People in Gordon Square

Locations of people sitting on a grass patch on a sunny afternoon.

```
> plot(gordon, main="People in Gordon Square", pch=16)
```


People in Gordon Square
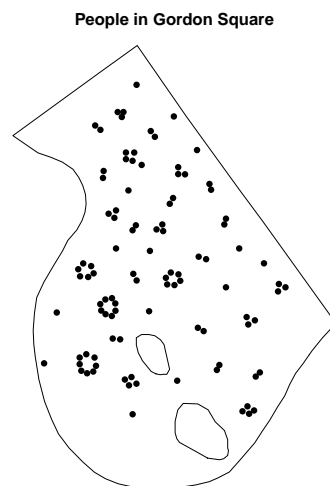
gorillas: Gorilla nesting sites

Locations of nesting sites of gorillas, and associated covariates, in a National Park in Cameroon.

gorillas is a marked point pattern (object of class "ppp") representing nest site locations.

gorillas.extra is a named list of 7 pixel images (objects of class "im") containing spatial covariates. It also belongs to the class "listof".

```
> plot(gorillas, which.marks=1, chars=c(1,3), cols=2:3, main="Gorilla nest sites")
```

Gorilla nest sites
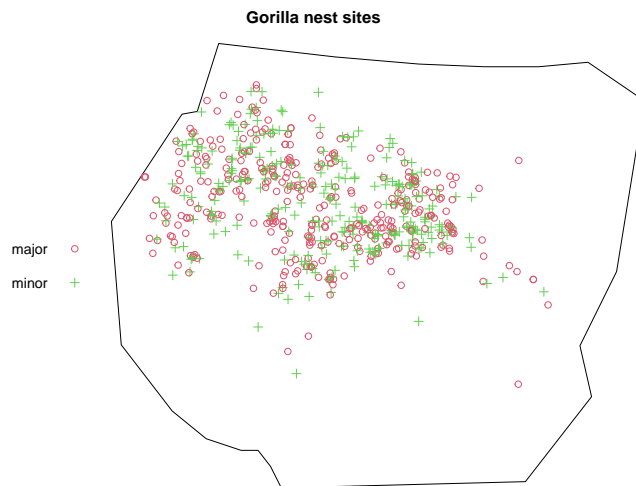
The `vegetation` covariate is also available as a raw ASCII format file,

```
> system.file("rawdata/gorillas/vegetation.asc", package="spatstat")
```

`hamster`: Hamster kidney cells

Cell nuclei in hamster kidney, each nucleus classified as either 'dividing' or 'pyknotic'. A multitype point pattern.

```
> plot(hamster, cols=c(2,4))
```

hamster
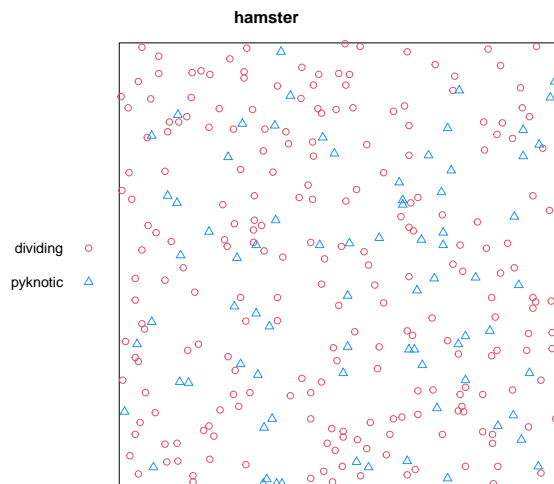
`heather`: Heather mosaic

The spatial mosaic of vegetation of the heather plant, recorded in a 10 by 20 metre sampling plot in Sweden. A list with three entries, representing the same data at different spatial resolutions.

```
> plot(heather)
```

heather

| fine | medium | coarse |
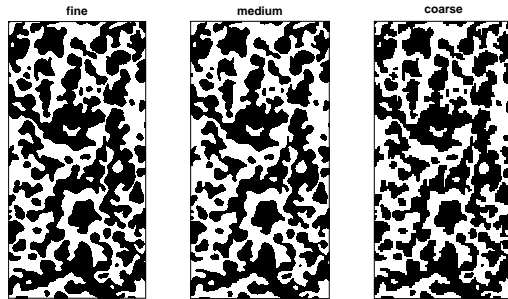
humberside: **Childhood Leukemia and Lymphoma**

Spatial locations of cases of childhood leukaemia and lymphoma, and randomly-selected controls, in North Humberside. A marked point pattern.

```
> plot(humberside)
```
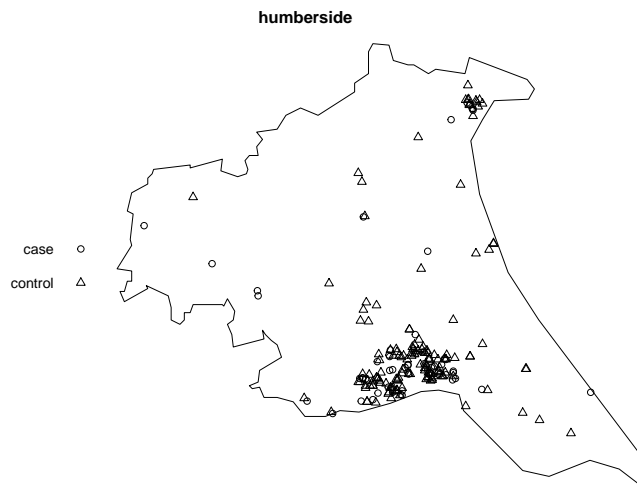

humberside

The dataset `humberside.convex` is an object of the same format, representing the same point pattern data, but contained in a larger, 5-sided convex polygon.

hyytiala: **Mixed forest**

Spatial locations and species classification for trees in a Finnish forest.

```
> plot(hyytiala, cols=2:5)
```

hyytiala
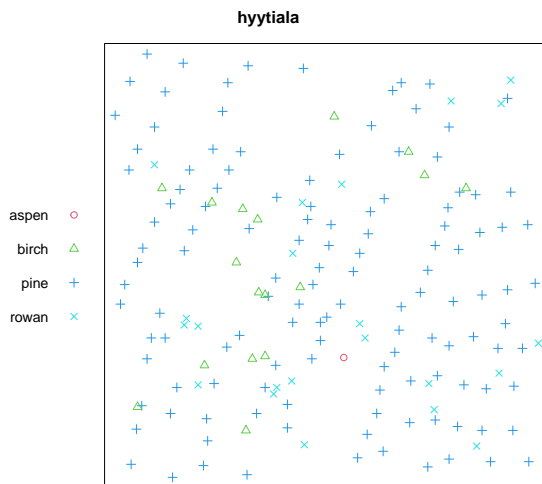
japanesepines: Japanese black pine saplings

Locations of Japanese black pine saplings in a square sampling region in a natural forest. Often used as a standard example.

```
> plot(japanesepines)
```


japanesepines
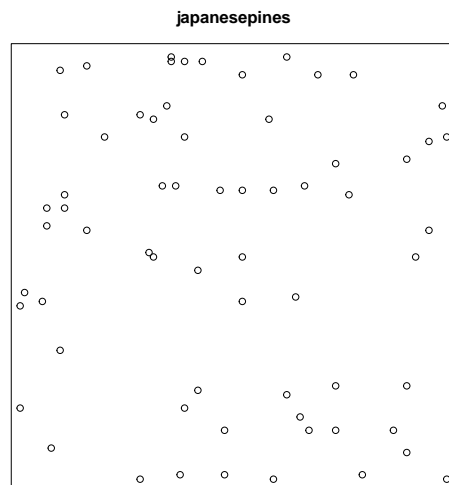
lansing: Lansing Woods

Locations and botanical classification of trees in a forest. A multitype point pattern with 6 different types of points. Includes duplicated points.

```
> plot(lansing)
```

lansing

blackoak ○
hickory △
maple +
misc ×
redoak ◇
whiteoak ▽

```
> plot(split(lansing))
```

split(lansing)

| blackoak | hickory | maple |
| misc | redoak | whiteoak |
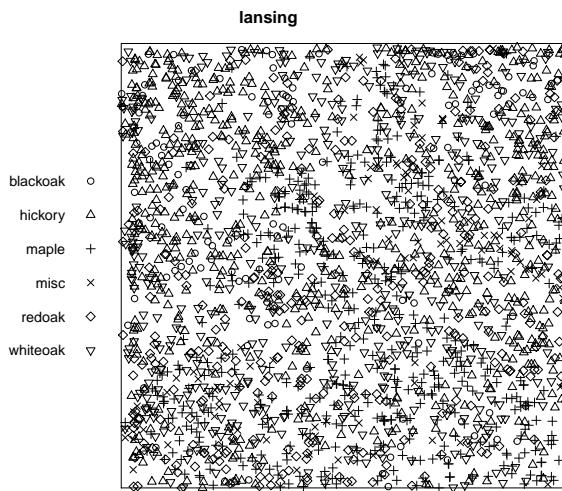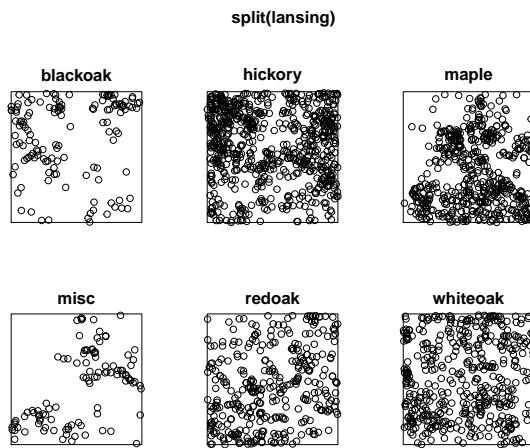
longleaf: Longleaf Pines

Locations and diameters of Longleaf pine trees.

```
> plot(longleaf)
```

longleaf
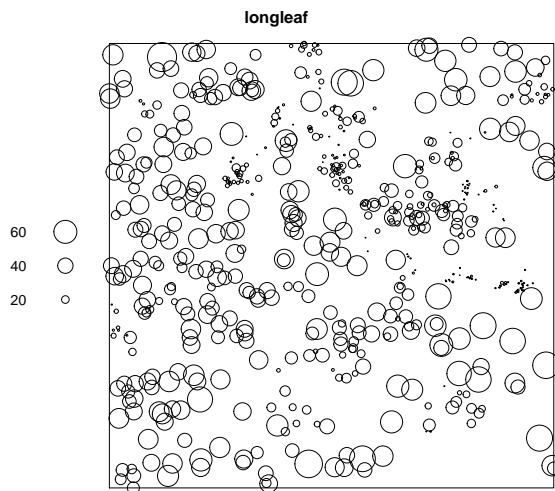
mucosa: Gastric Mucosa Cells

A bivariate inhomogeneous point pattern, giving the locations of the centres of two types of cells in a cross-section of the gastric mucosa of a rat.

```
> plot(mucosa, chars=c(1,3), cols=c("red", "green"))
> plot(mucosa.subwin, add=TRUE, lty=3)
```

mucosa
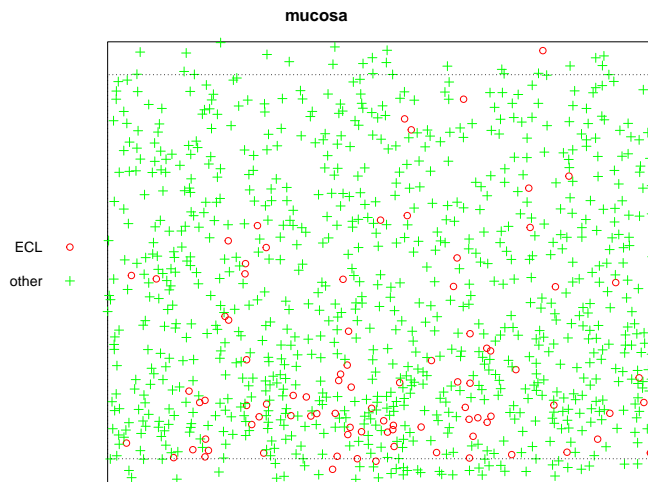
murchison: Murchison Gold Deposits

Spatial locations of gold deposits and associated geological features in the Murchison area of Western Australia. A list of three elements:

- `gold`, the point pattern of gold deposits;
- `faults`, the line segment pattern of geological faults;
- `greenstone`, the subregion of greenstone outcrop.

```
> plot(murchison$greenstone, main="Murchison data", col="lightgreen")
> plot(murchison$gold, add=TRUE, pch=3, col="blue")
> plot(murchison$faults, add=TRUE, col="red")
```

Murchison data
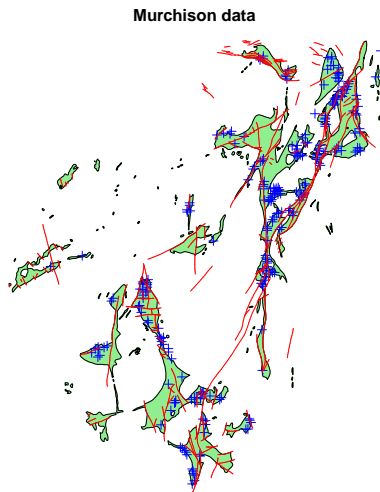
nbfires: New Brunswick Fires

Fires in New Brunswick (Canada) with marks giving information about each fire.

```
> plot(nbfires, use.marks=FALSE, pch=".")
```

nbfires


```
> plot(split(nbfires), use.marks=FALSE, chars=".")
```

split(nbfires)

1987	**1989**	**1990**	**1991**
1992	**1993**	**1994**	**1995**
1996	**1997**	**1998**	**1999**
2000	**2001**	**2002**	**2003**
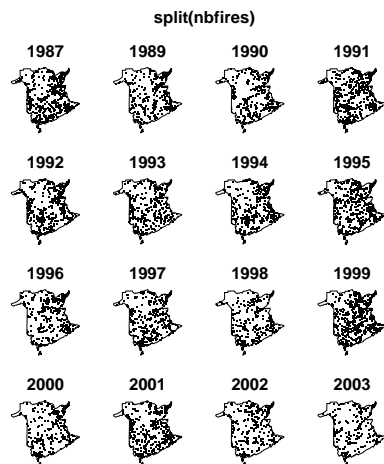

```
> par(mar=c(0,0,2,0))
> plot(split(nbfires)$"2000", which.marks="fire.type",
+       main=c("New Brunswick fires 2000", "by fire type"),
+       cols=c("blue", "green", "red", "cyan"),
+       leg.side="left")
```
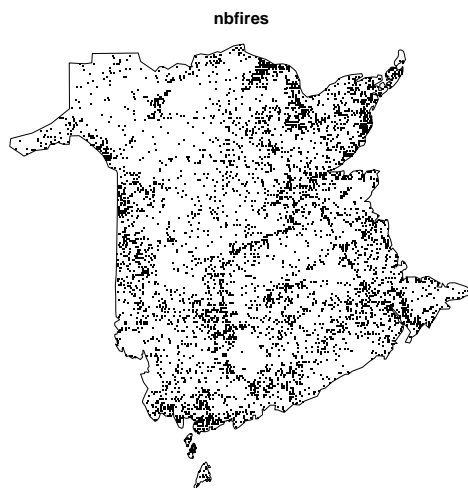

New Brunswick fires 2000
by fire type

forest o
grass △
dump +
other ×
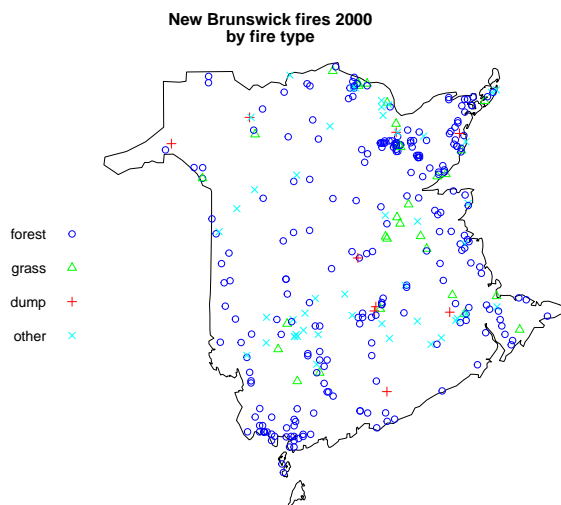
nztrees: New Zealand Trees

Locations of trees in a forest plot in New Zealand. Often used as a demonstration example.

```
> plot(nztrees)
> plot(trim.rectangle(as.owin(nztrees), c(0,5), 0), add=TRUE, lty=3)
```

nztrees
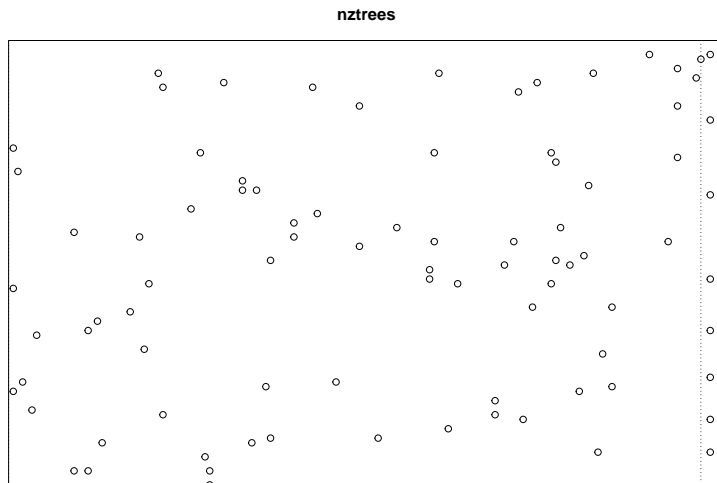
osteo: Osteocyte Lacunae

Replicated three-dimensional point patterns: the three-dimensional locations of osteocyte lacunae observed in rectangular volumes of solid bone using a confocal microscope. A `hyperframe` containing 3D point patterns and explanatory variables.

```
> plot(osteo[1:10,], main.panel="", pch=21, bg='white')
```

osteo[1:10,]
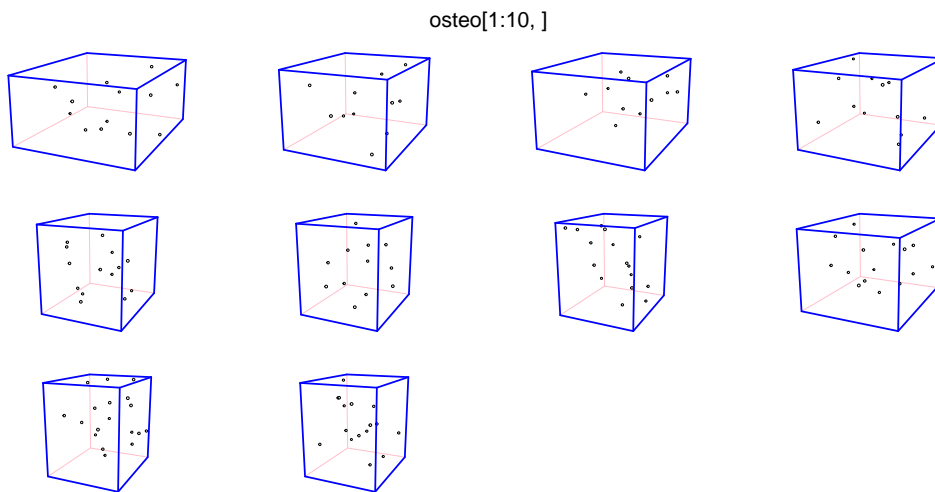
For demonstration and instruction purposes, the raw data from the 36th point pattern are available in a plain ascii file in the `spatstat` installation,

```
> system.file("rawdata/osteo/osteo36.txt", package="spatstat")
```

paracou: Kimboto trees

Point pattern of adult and juvenile Kimboto trees recorded at Paracou in French Guiana. A bivariate point pattern.

```
> plot(paracou, cols=2:3, chars=c(16,3))
```

paracou

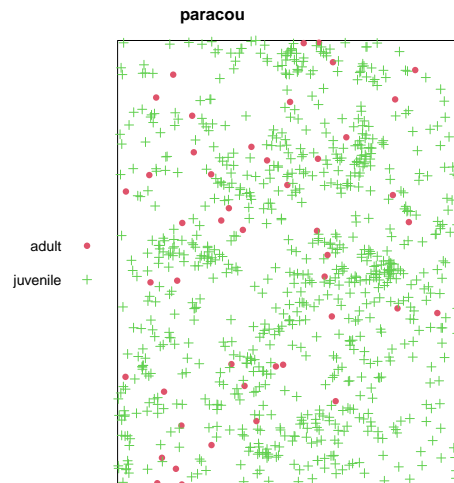adult •
juvenile +

ponderosa: **Ponderosa Pines**

Locations of Ponderosa Pine trees in a forest. Several special points are identified.

```
> ponderosa.extra$plotit()
```


ponderosa
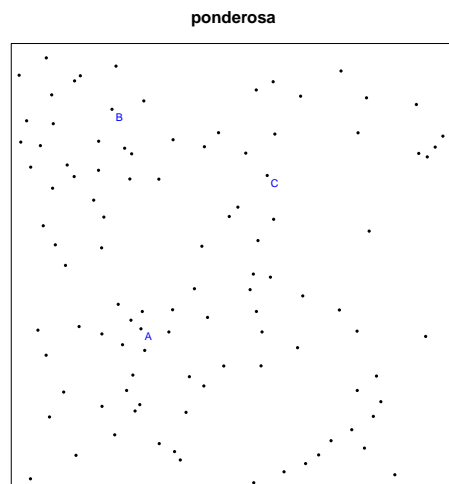
pyramidal: **Pyramidal Neurons in Brain**

Locations of pyramidal neurons in sections of human brain. There is one point pattern from each of 31 human subjects. The subjects are divided into three groups: controls (12 subjects), schizoaffective (9 subjects) and schizophrenic (10 subjects).

```
> pyr <- pyramidal
> pyr$grp <- abbreviate(pyramidal$group, minlength=7)
> plot(pyr, quote(plot(Neurons, pch=16, main=grp)), main="Pyramidal Neurons")
```

Pyramidal Neurons
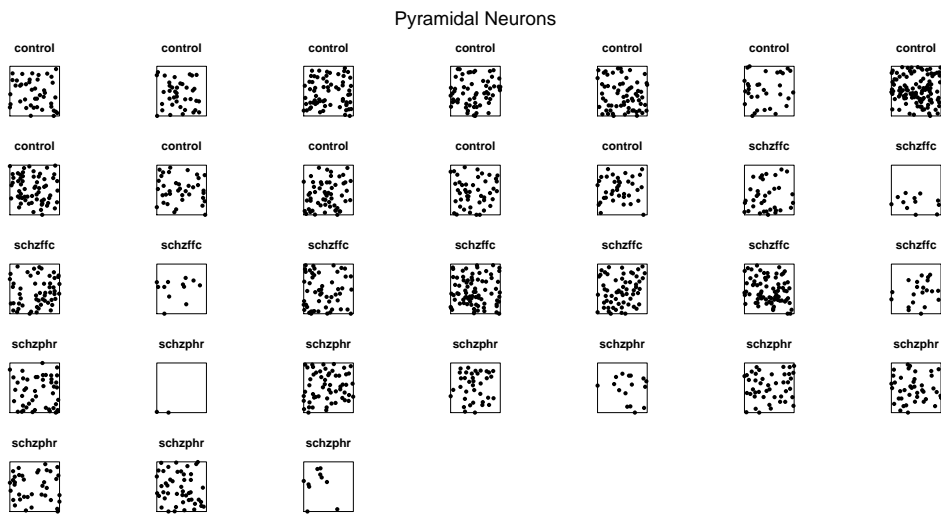
redwood, redwood3, redwoodfull: Redwood seedlings and saplings

California Redwood seedlings and saplings in a forest. There are two versions of this dataset: `redwood` and `redwoodfull`.

The `redwoodfull` dataset is the full data. It is spatially inhomogeneous in density and spacing of points.

The `redwood` dataset is a subset of the full data, selected because it is apparently homogeneous, and has often been used as a demonstration example. This comes in two versions commonly used in the literature: `redwood` (coordinates given to 2 decimal places) and `redwood3` (coordinates given to 3 decimal places).

```
> plot(redwood)
> plot(redwood3, add=TRUE, pch=20)
```
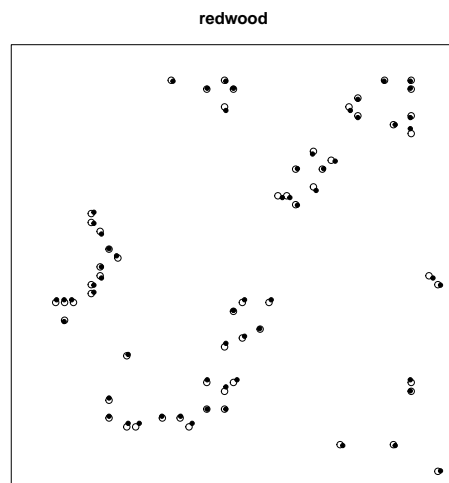

redwood

```
> redwoodfull.extra$plotit()
```

Strauss's redwood data
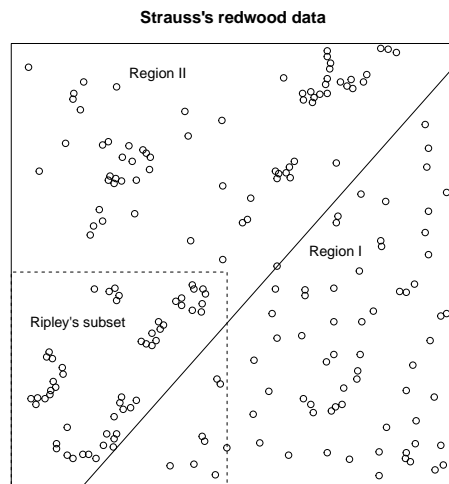
residualspaper: Data from residuals paper

Contains the point patterns used as examples in
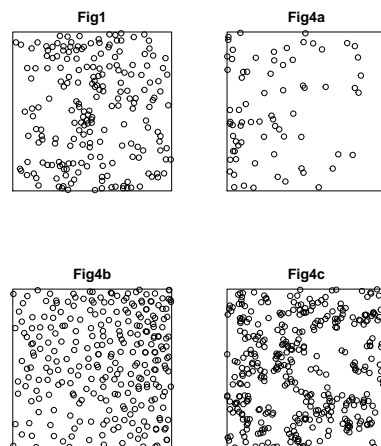
> A. Baddeley, R. Turner, J. Møller and M. Hazelton (2005) Residual analysis for spatial point processes. *Journal of the Royal Statistical Society, Series B* **67**, 617–666
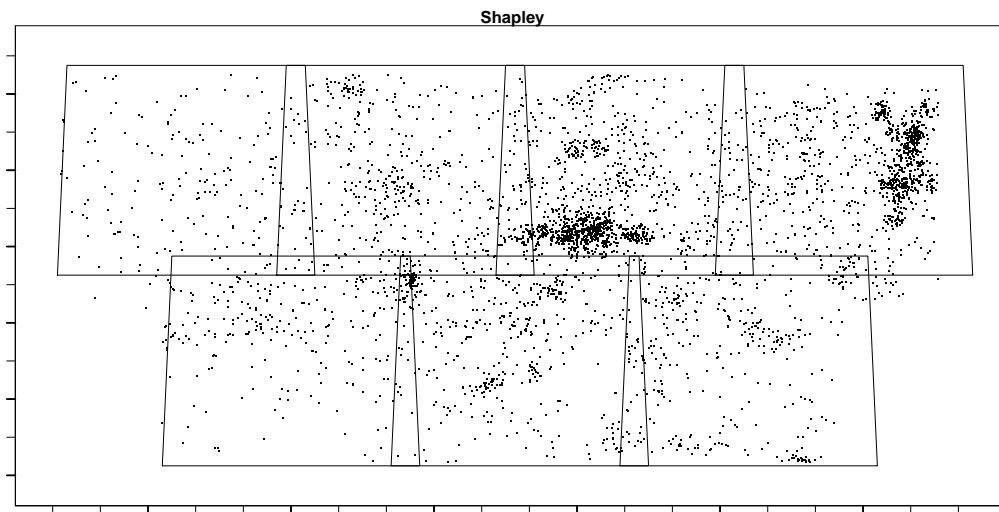
along with R code.

```
> plot(as.solist(residualspaper[c("Fig1", "Fig4a", "Fig4b", "Fig4c")]),
+      main="")
```


shapley: Shapley Galaxy Concentration

Sky positions of 4215 galaxies in the Shapley Supercluster (mapped by radioastronomy).
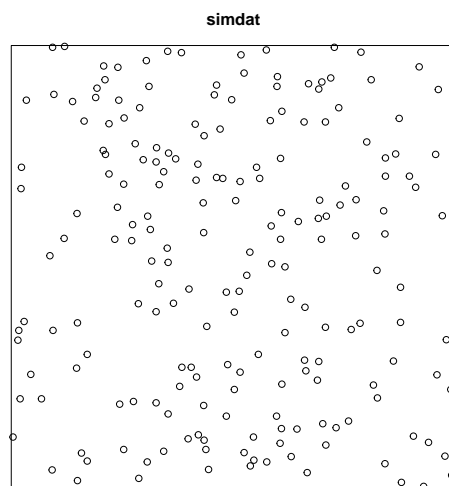
```
> shapley.extra$plotit(main="Shapley")
```

Shapley

`simdat`: **Simulated data**

Another simulated dataset used for demonstration purposes.

```
> plot(simdat)
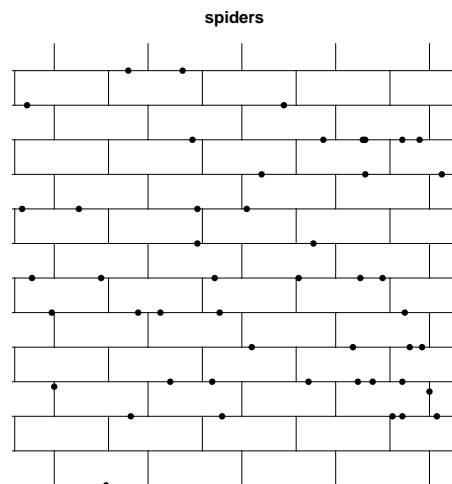```



**simdat**

`spiders`: **Spider webs**

Spider webs across the mortar lines of a brick wall. A point pattern on a linear network.

```
> plot(spiders, pch=16, show.window=FALSE)
```

spiders

sporophores: **Sporophores**

Sporophores of three species of fungi around a tree.

```
> plot(sporophores, chars=c(16,1,2), cex=0.6)
> points(0,0,pch=16, cex=2)
> text(15,8,"Tree", cex=0.75)
```

sporophores
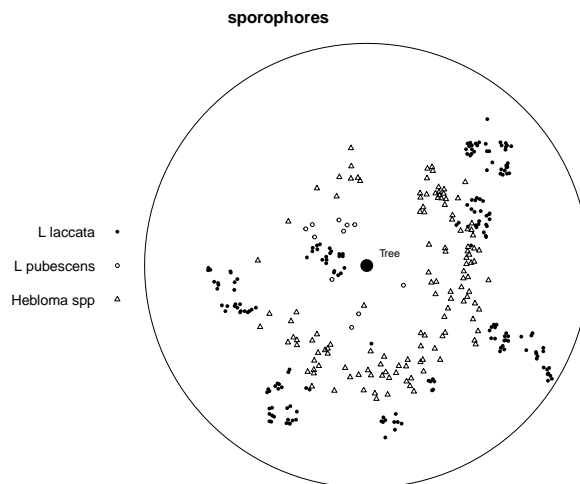
spruces: **Spruces in Saxony**

Locations of Norwegian spruce trees in a natural forest stand in Saxonia, Germany. Each tree is marked with its diameter at breast height.

```
> plot(spruces, maxsize=min(nndist(spruces)))
```

spruces

swedishpines: Swedish Pines
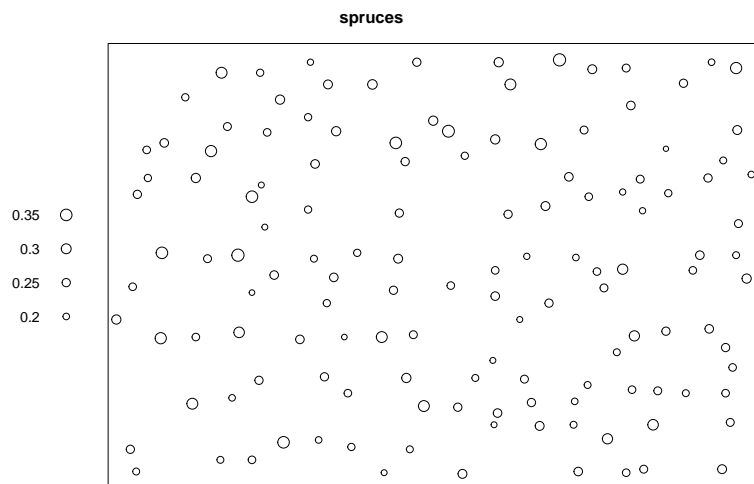
Locations of pine saplings in a Swedish forest. Often used as a demonstration example.
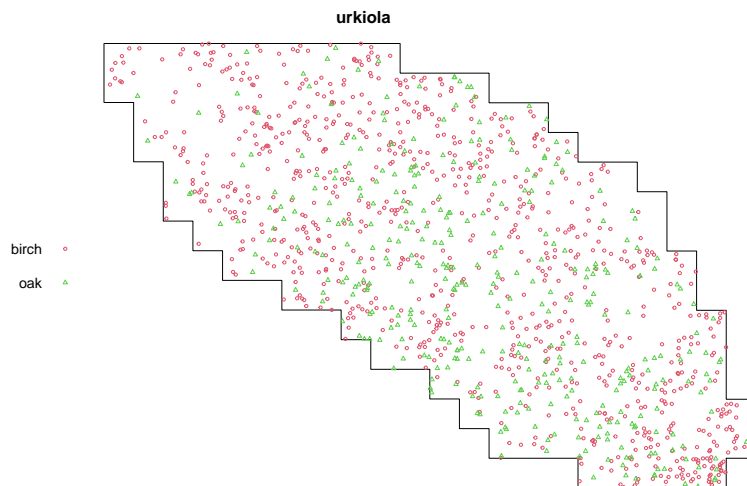
```
> plot(swedishpines)
```

swedishpines
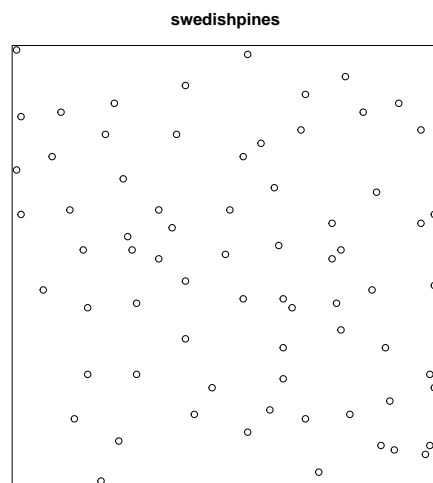
urkiola: trees in a wood

Locations of birch and oak trees in a secondary wood in Urkiola Natural Park (Basque country, northern Spain). Irregular window, bivariate point pattern.

```
> plot(urkiola, cex=0.5, cols=2:3)
```

`waka`: trees in Waka National Park

Spatial coordinates of each tree, marked by the tree diameter at breast height.

```
> par(mar=c(0,0,2,0))
> plot(waka, markscale=0.04, main=c("Waka national park", "tree diameters"))
```
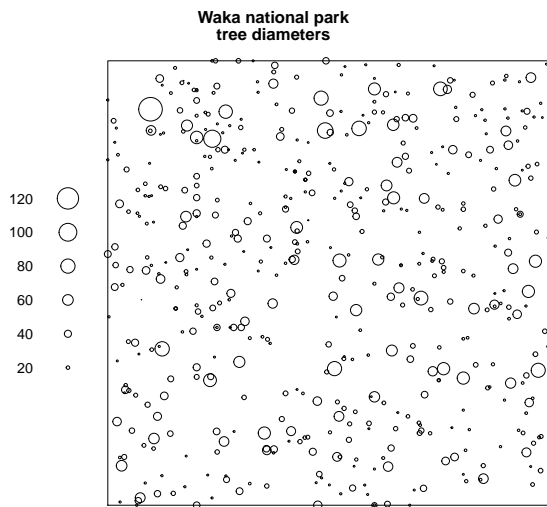

`vesicles`: synaptic vesicles

Point pattern of synaptic vesicles observed in rat brain tissue.

```
> v <- rotate(vesicles, pi/2)
> ve <- lapply(vesicles.extra, rotate, pi/2)
> plot(v, main="Vesicles")
> plot(ve$activezone, add=TRUE, lwd=3)
```

Vesicles

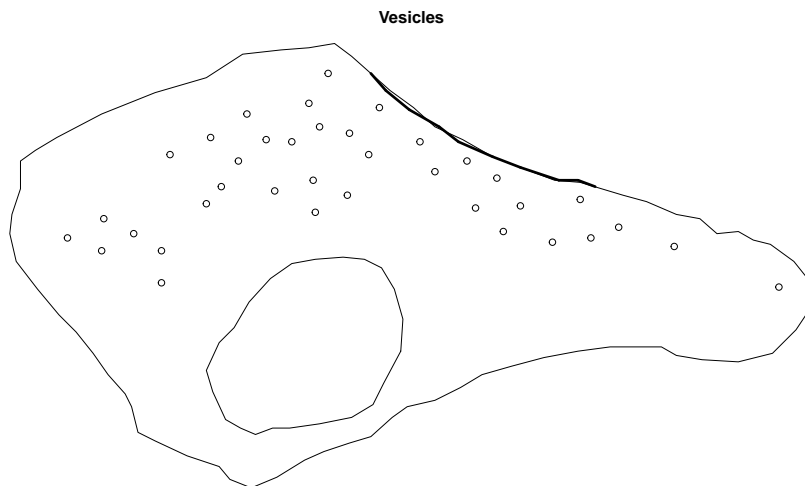The auxiliary dataset `vesicles.extra` is a list with entries

`presynapse`	outer polygonal boundary of presynapse
`mitochondria`	polygonal boundary of mitochondria
`mask`	binary mask representation of vesicles window
`activezone`	line segment pattern representing the active zone.

For demonstration and training purposes, the raw data files for this dataset are also provided in the `spatstat` package installation:

`vesicles.txt`	spatial locations of vesicles	
`presynapse.txt`	vertices of `presynapse`	
`mitochondria.txt`	vertices of `mitochondria`	The files are in the folder `rawdata/vesicles` in
`vesiclesimage.tif`	greyscale microscope image	
`vesiclesmask.tif`	binary image of `mask`	
`activezone.txt`	coordinates of `activezone`	

the `spatstat` installation directory. The precise location of the files can be obtained using `system.file`, for example

```
> system.file("rawdata/vesicles/mitochondria.txt", package="spatstat")
```

`waterstriders`: Insects on a pond

Three independent replications of a point pattern formed by insects on the surface of a pond.

```
> plot(waterstriders)
```

waterstriders

Component_1   Component_2   Component_3