Topic 2: Exploratory spatial data analysis

This topic overview some methods of exploratory spatial data analysis with R. In particular, we consider

- Plotting with **geoR.**
- Basic analysis with **sp.**
- Visualising the **Meuse data.**
- **Fitting spatial trends.**

Spatial Exploratory Data Analysis

Spatial exploratory data analysis starts with the plotting maps of measured variables and their descriptive statistics.

geoR

The package GEOR provides functions for geostatistical data analysis.

To load package and its simulated data S100, type:

```
> library(geoR)
> data(s100)
```

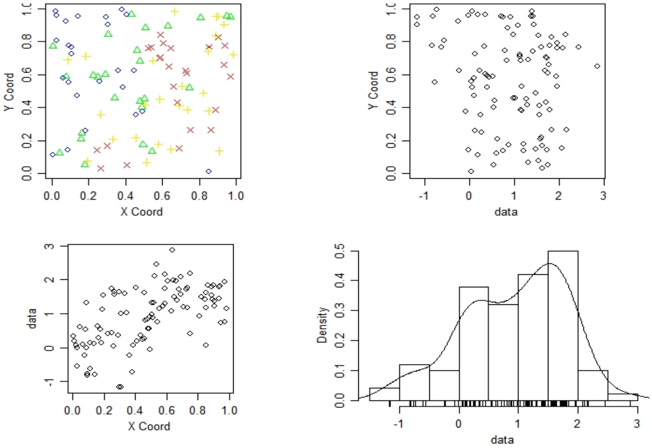
A quick summary of the data can be obtained typing

```
> summary(s100)
```

It will return basic information about the coordinates and data values.

The function **plot.geodata** shows a 2 x 2 display with data locations (top plots) and data versus coordinates (bottom plots):

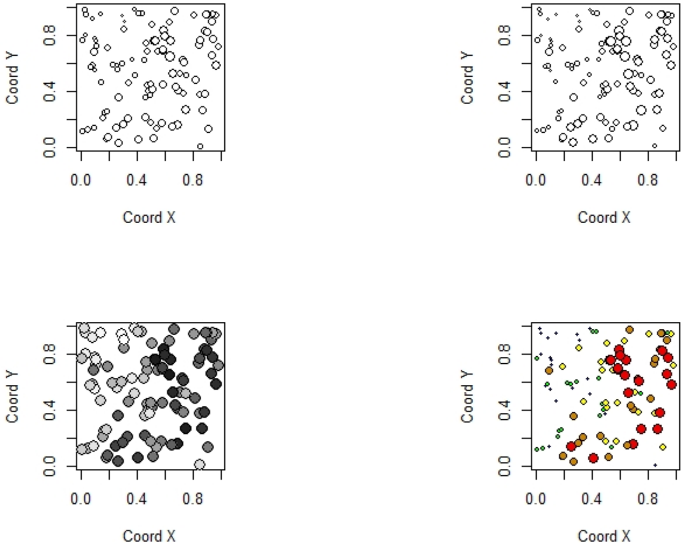
```
> plot(s100)
```



The function **points.geodata** produces a plot showing the data locations.

There are options to specify point sizes, patterns and colors, which can be set to be proportional to the data values or specified quantiles. Some examples of graphical outputs are illustrated by the commands and corresponding plots as shown below:

```
> par(mfrow = c(2,2))
> points(s100, xlab = "Coord X", ylab = "Coord Y")
> points(s100, xlab = "Coord X", ylab = "Coord Y",
+ pt.divide = "rank.prop")
> points(s100, xlab = "Coord X", ylab = "Coord Y",
+ cex.max = 1.7, col = gray(seq(1, 0.1, l=100)),
+ pt.divide = "equal")
> points(s100, pt.divide = "quintile", xlab = "Coord X",
+ ylab = "Coord Y")
```


Coord Y
Coord X
0.0
0.4
0.8

Basic analysis with sp

We use MEUSE data:

```
> library(lattice)
> library(sp)
> data(meuse)
> ?meuse
```

This data set gives locations and top soil heavy metal concentrations (ppm), along with a number of soil and landscape variables, collected in a flood plain of the river Meuse in Belgium.

Heavy metal concentrations are bulk sampled from areas of approximately 15m x 15m.

This data frame contains the following columns in which we are interested:

- **x** a numeric vector; x-coordinate (m) in RDM (Dutch topographical map coordinates)
- **y** a numeric vector; y-coordinate (m) in RDM (Dutch topographical map coordinates)
- **zinc** topsoil zinc concentration
- **dist** distance to river Meuse; obtained from the nearest cell in meuse.grid, which in turn was derived by a spread (spatial distance) GIS operation, therefore it is accurate up to 20 metres; normalized $[0, 1]$

First we transform the data the Spatial class:

```
> coordinates(meuse) <- c("x", "y")
```

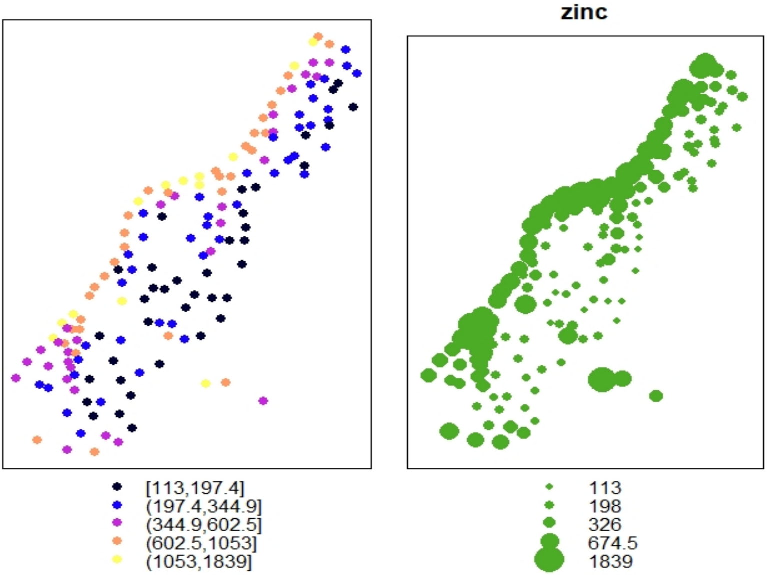
Then, to plot the observed value, we can use colour

```
> spplot(meuse, "zinc", do.log = T)
```

or symbol size:

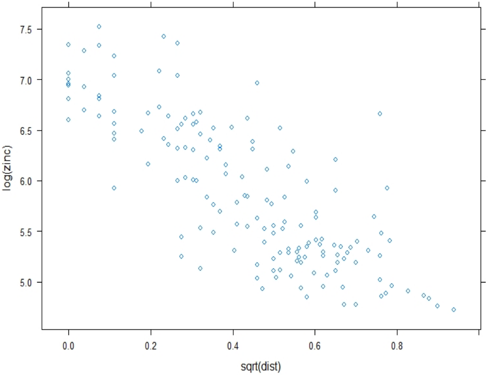
```
> bubble(meuse, "zinc", do.log = T, key.space = "bottom")
```

The evident structure here is that zinc concentration is larger close to the river Meuse banks.


zinc
[113,197.4]
(197.4,344.9]
(344.9,602.5]
(602.5,1053]
(1053,1839]
113
198
326
674.5
1839

There is a spatial trend, such as the relation between top soil zinc concentration and distance to the river. However this trend is non-linear. Applying appropriate transformations of the concentration and distance attributes it can be made approximately linear:

```
> xyplot(log(zinc) ~ sqrt(dist), as.data.frame(meuse))
```



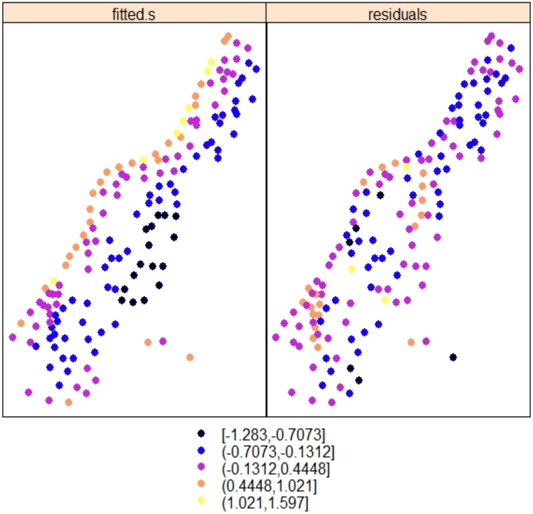
Therefore, we can also plot maps with fitted values and with residuals:

```
> zn.lm <- lm(log(zinc) ~ sqrt(dist), meuse)
> meuse$fitted.s <- predict(zn.lm,meuse)-mean(predict(zn.lm,meuse))
> meuse$residuals <- residuals(zn.lm)
> spplot(meuse, c("fitted.s", "residuals"))
```

where the formula $y \sim x$ indicates dependency of y on x.

The resulting figure reveals that although the trend removes a large part of the variability, the residuals do not appear to behave as spatially unstructured: residuals with a similar value occur regularly close to another.

So, we need some random field model to describe the residuals and improve the prediction.


fitted.s
residuals
[-1.283,-0.7073]
(-0.7073,-0.1312]
(-0.1312,0.4448]
(0.4448,1.021]
(1.021,1.597]

Fitting spatial trends.

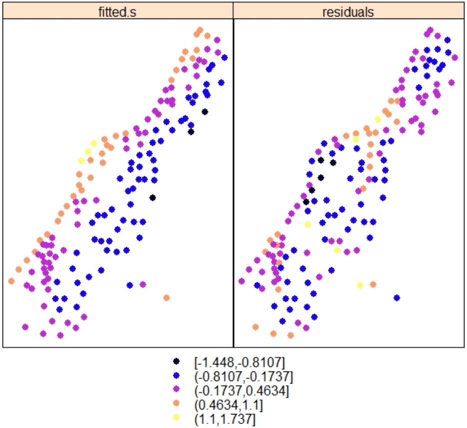
A special form of linear regression is obtained when polynomials of spatial coordinates are used for predictors, for example a second-order polynomial. This form is called trend surface analysis.

To use **lm** for trend surface analysis, for example, for the second order trend apply `I` (to treat powers and products as is):

```
> fit <- lm(log(zinc) ~I(x^2)+I(y^2)+I(x*y)+x+y,meuse)
> fit
Call:
lm(formula = log(zinc) ~ I(x^2) + I(y^2) + I(x * y) + x + y,
data = meuse)
Coefficients:
(Intercept)       I(x^2)       I(y^2)     I(x * y)
2.395e+04    8.575e-07    8.467e-07   -1.623e-06
x            y
2.279e-01   -2.684e-01
```

To plot maps with fitted values and residuals, use the commands:

```
> meuse$fitted.s <- predict(fit,meuse) - mean(predict(fit,meuse))
> meuse$residuals <- residuals(fit)
> spplot(meuse, c("fitted.s", "residuals"))
```



Key R commands

read.csv(x)	*reads a csv file and creates a data frame from it*
cbind(x)	*combines the sequence x by columns*
plot.geodata(x)	*produces a 2 x 2 display of geostatistical data*
points.geodata(x)	*produces a plot with points indicating the data locations*
bubble(x)	*creates a bubble plot of spatial data*
xyplot(x)	*produces a bivariate scatterplot*
lm(x)	*fits a linear model*
I(x)	*indicates that x should be treated "as is"*
predict(x)	*predicts using a fitted model*
residuals(x)	*extracts model's residuals*