UCLA Department of Statistics
Edward R. Roybal Learning Center

Mathematical Association of America
Dolciani Mathematics Enrichment Grant

Instructors: Nicolas Christou
Robert Montgomery

Spatial data

- What are spatial data?

- Why spatial data?
  "A major pleasure in working with spatial data is their visualization. Maps are amongst the most compelling graphics, because the space they map is the space we think we live in, and maps can show things we cannot see otherwise". Roger Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, Applied Spatial Data Analysis with R, Use R!, Springer 2008."

- Types of spatial data:

  1. Geostatistical data.

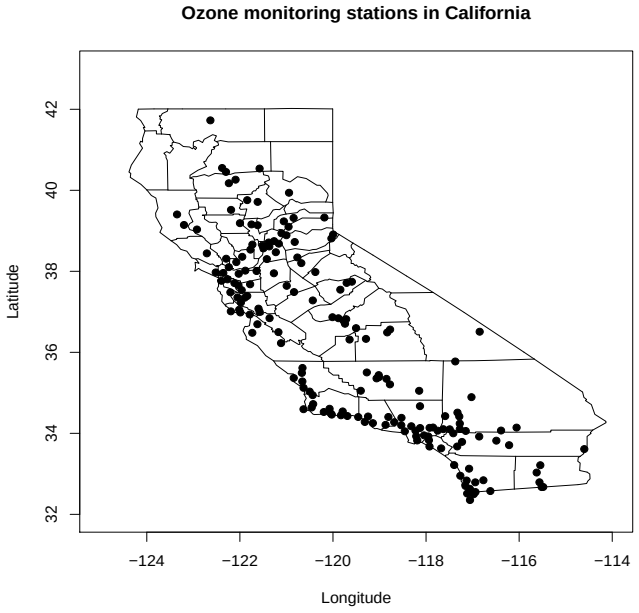
  2. Lattice data.

  3. Point pattern data.

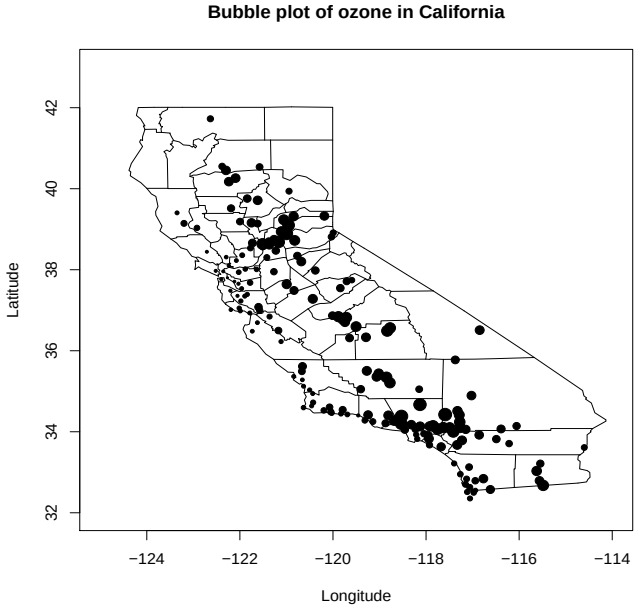
- Geostatistical data.
  1. Identify the variable(s) of interest (e.g. concentration of ozone, amount of rainfall, temperature, elevation, etc.)
  2. Know the coordinates (x, y) of the data points.
  3. Plot the data points on a map.
  4. Construct a bubble plot.
  5. Plot the variable(s) of interest against x and against y. Why?
  6. Predict the variable(s) of interest at unsampled locations.

- Example 1: California ozone data.

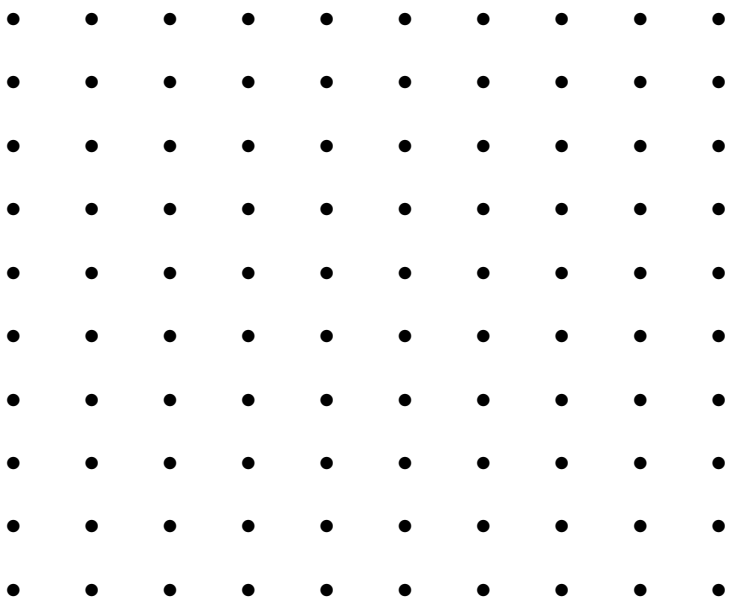
```
a <- read.table("http://www.stat.ucla.edu/~nchristo/maadmeg/
ozone.txt", header=TRUE)

#Load the package "maps":
library(maps)

#We can display the data points and the map using the following commands:
plot(a$x, a$y, xlim=c(-125,-114),ylim=c(32,43), xlab="Longitude",
ylab="Latitude", main="Ozone locations in California")

#Add the map of California:
map("county", "ca",add=TRUE)
```



```
#We can also plot the points relative to their value
#(larger values will be displayed with larger circles)
plot(a$x, a$y, xlim=c(-125,-114),ylim=c(32,43), xlab="Longitude",
ylab="Latitude", main="Ozone locations in California", "n")

map("county", "ca",add=TRUE)

#Use cex argument to construct the bubble plot:
points(a$x, a$y, cex=a$o3/mean(a$o3))
```



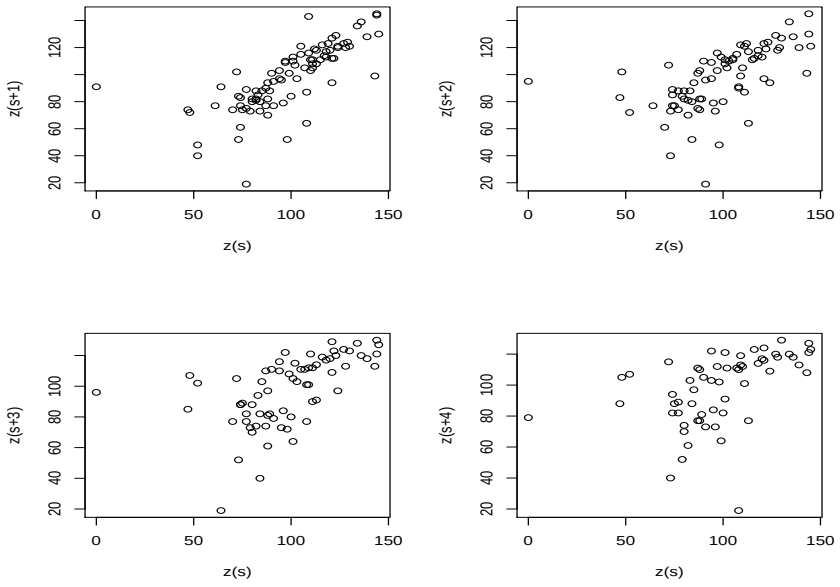
- Example 2: Exploring spatial correlation.
  An h-scatterplot shows all possible pairs of data values whose locations are separated by a certain distance in a particular direction. For example, the figure below shows a 10×10 grid (100 data points - with distance between two consecutive data points 1 m in north-south and east-west direction).



  And below we see the value at each location.

81	77	103	112	123	19	40	111	114	120
82	61	110	121	119	77	52	111	117	124
82	74	97	105	112	91	73	115	118	129
88	70	103	111	122	64	84	105	113	123
89	88	94	110	116	108	73	107	118	127
77	82	86	101	109	113	79	102	120	121
74	80	85	90	96	101	96	72	128	130
75	80	83	87	94	99	95	48	139	145
77	84	74	108	121	143	91	52	136	144
87	100	47	111	124	109	0	98	134	144

In this first example we will plot each value (we place it on the x-axis) against a value that is 1 m apart on the south-north direction (we place it on the y-axis). How many pairs are there? The scatterplot shows a cloud of points distributed around the 45-degree line. We observe that there is an association between nearby data points. If we increase the distance from 1 m to 2 m, 3 m, and 4 m we see that the cloud becomes "fatter" indicating that the values separated by longer distance are not as correlated as with the 1 m case. These plots are shown below:



Similarly, if we move on southwest-northeast direction we will pair all values that are $\sqrt{2}$ m apart. How many pairs are there?

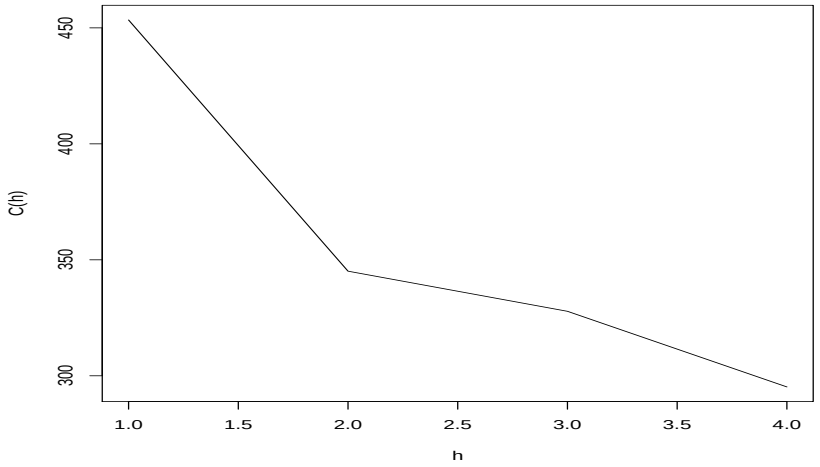
- As with many graphical displays, we need a quantitative summary of the information contained on an h-scatterplot. We want to know the strength of the association between points separated by distance h. We therefore compute the covariance and the correlation. The table below show the values of the covariance and correlation for separation distances $h = 1, 2, 3, 4$ on the south-north direction

h	$C(h)$	$\rho(h)$
1	453.4	0.742
2	345.1	0.590
3	327.8	0.560
4	295.2	0.476

  The plot of the correlation coefficient ($\rho(h)$) against h (the separation distance) it is called *correlogram*. We can also plot the covariance ($C(h)$) against h to get the so called *covariogram*.

- The two figures below show the covariogram and correlogram for these data.

  Covariogram:



  Correlogram: