Deterministic interpolation

GIS 5923 Spatial Statistics

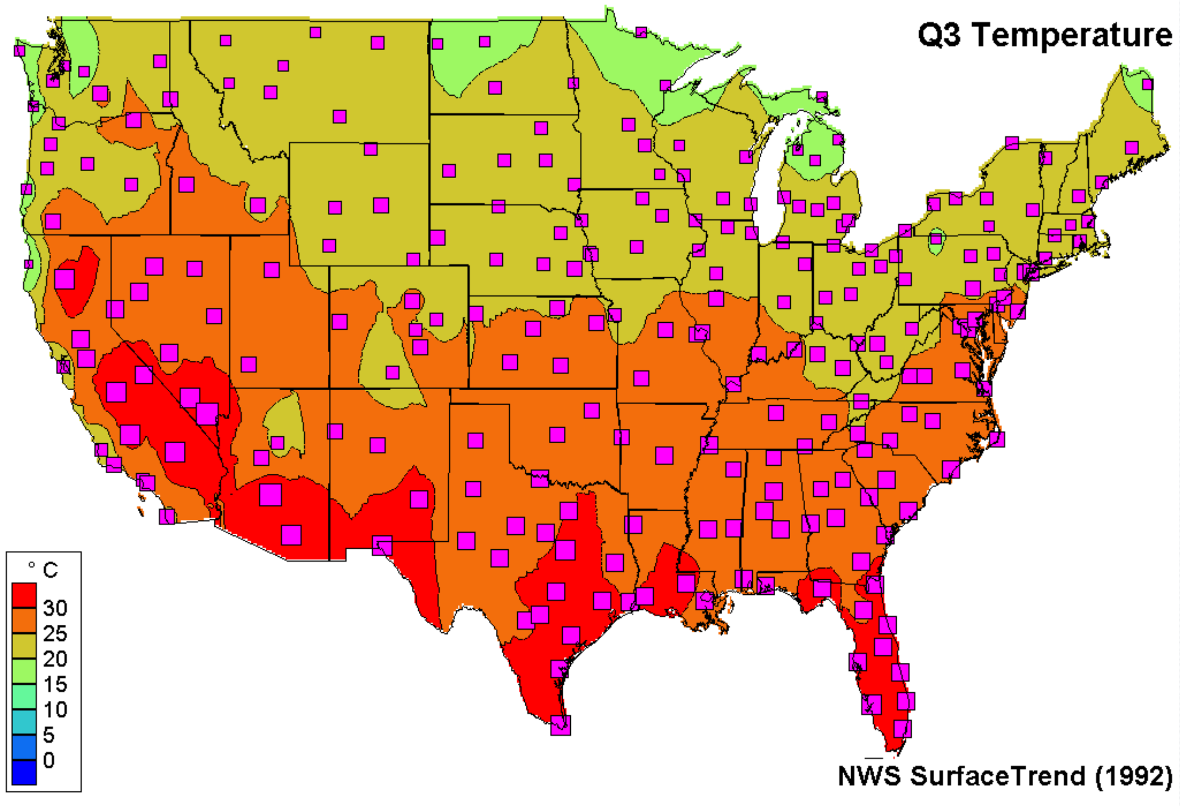
COLLEGE OF ATMOSPHERIC AND GEOGRAPHIC SCIENCES
DEPARTMENT OF GEOGRAPHY
AND ENVIRONMENTAL SUSTAINABILITY
The UNIVERSITY of OKLAHOMA

Interpolation and Geostatistics

- A field view is that the world consists of attributes that are continuously variable and measurable across space. Some authors call these geostatistical data
- Geostatistical data often come as measurements at a limited number of locations
- The pattern of observed locations is usually not of primary interest, since it often results from economical and physical constraints of the sample study design

Geostatistical data: example

- For example, we might have temperature observed at a series of weather stations.
- How can we interpolate a continuous field of temperature from these discrete observations?



COLLEGE OF ATMOSPHERIC AND GEOGRAPHIC SCIENCES
DEPARTMENT OF GEOGRAPHY
AND ENVIRONMENTAL SUSTAINABILITY
The UNIVERSITY of OKLAHOMA

Scalar vs. vector fields

- A scalar field is any quantity characterized only by its magnitude or amount, independent of any spatial coordinate system in which it is measured. Example: land elevation
- We represent scalar fields by the very general equation $z_i = f(\mathbf{s}_i)$, where z is called the height of the response variable
- A vector field is any mapped quantity with both magnitude and direction. It is not independent of the spatial coordinate system used. Example: land surface slope

Assumptions of field data

- Generally we make two assumptions about field data:

- Continuity: for every location s_i in the study area, the height variable z_i measurable at the same place. (for some applications it is common to assume that z is differentiable everywhere)

- Single-valued: for every location there is only one value of z

COLLEGE OF ATMOSPHERIC AND GEOGRAPHIC SCIENCES
DEPARTMENT OF GEOGRAPHY
AND ENVIRONMENTAL SUSTAINABILITY
The UNIVERSITY of OKLAHOMA

How do we model and store field data?

- To model field data, we generally take two steps:
- 1) Sampling the real surface
- 2) Employing some form of interpolation or modeling to predict the continuous surface between the sampled points

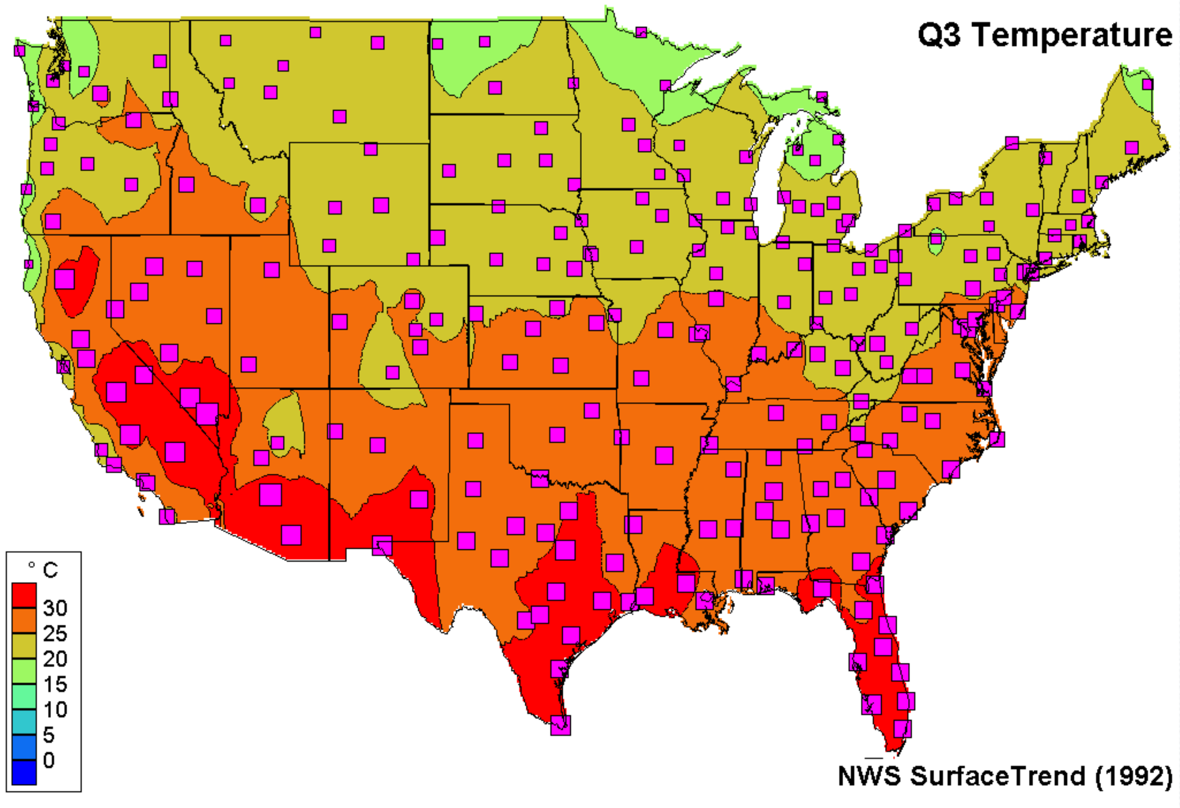
Step 1: Sampling the real surface

- Usually, we have measured values at a set of known locations: rain gauges, temperature stations, bore holes, etc. These locations are known as control points.

- Increasingly, field data are acquired from aerial and remote sensing surveys, including LiDAR. In this case, z-values are on a regular grid that can be of high resolution.

Interpolation to create a continuous surface

- To create a continuous field, we need to interpolate based on the values observed at the control points



COLLEGE OF ATMOSPHERIC AND GEOGRAPHIC SCIENCES
DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY
The UNIVERSITY of OKLAHOMA

Spatial interpolation

- Spatial interpolation is the prediction of exact values of attributes at unsampled locations from measurements made at control points. How?

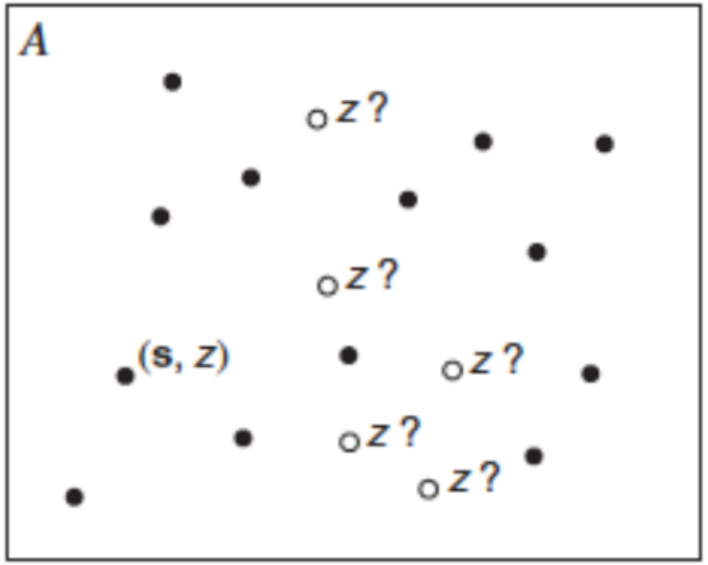


Figure 9.3 The interpolation problem. Control points are black circles, where we know the location, s, and the height, z, but we require the field height, z, anywhere in the region A—say, at the unfilled circle locations.

Source: O'Sullivan and Unwin

Non-spatial approach to interpolation: take the mean of the control points

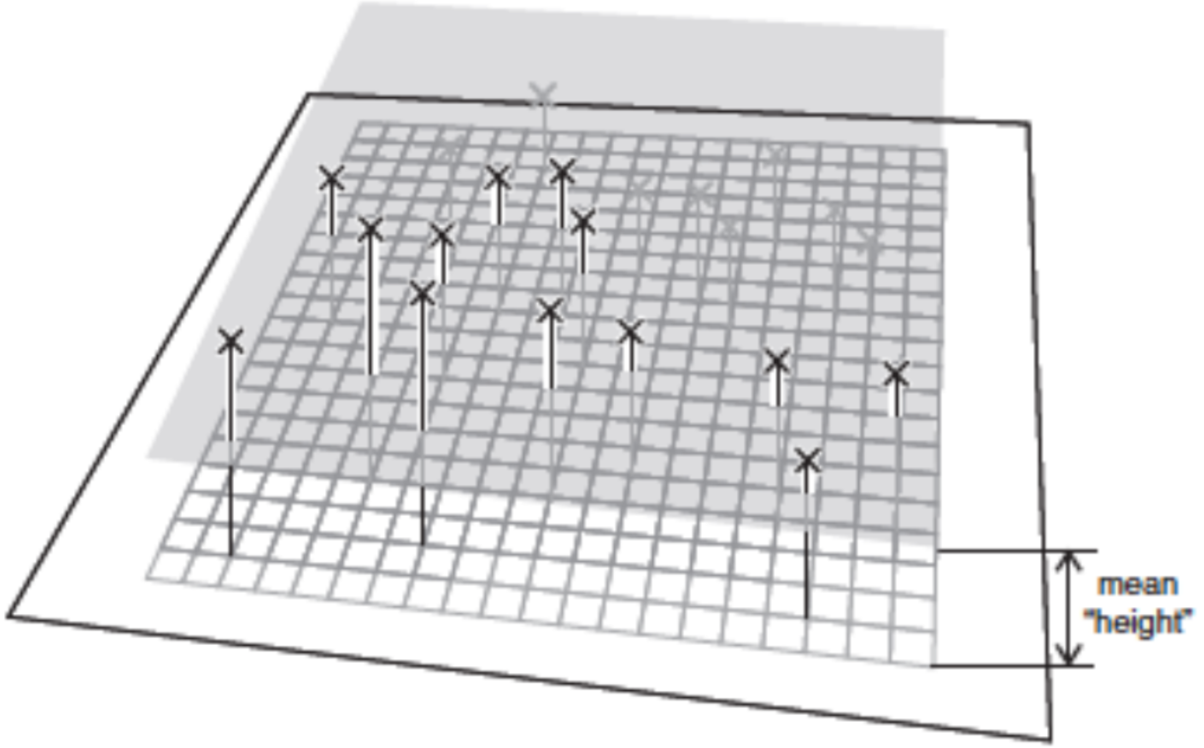


Figure 9.5 Not taking space into account in prediction.

Source: O'Sullivan and Unwin

Spatial approach 1: proximity polygons

- A simple improvement on using the mean is to assign every unsampled point the value at its nearest control point using proximity polygons

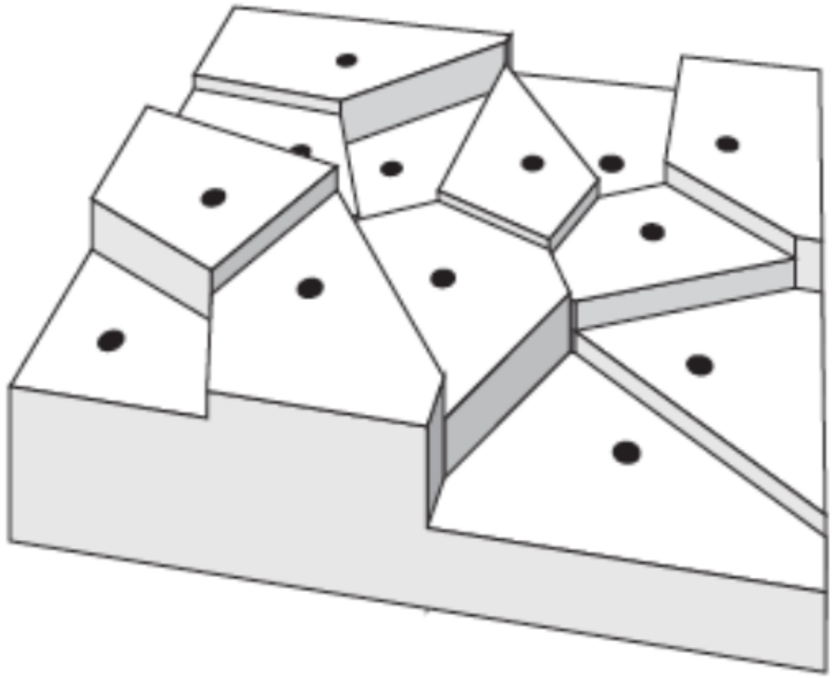


Figure 9.6 The "blocky," discontinuous results of an interpolation using proximal polygons.

Source: O'Sullivan and Unwin

Spatial approach 2: The local average

- Perhaps better is to calculate the local spatial means of the sample points. For example, we might take the mean of the 10 nearest control points, or the mean of all control points within 100m.

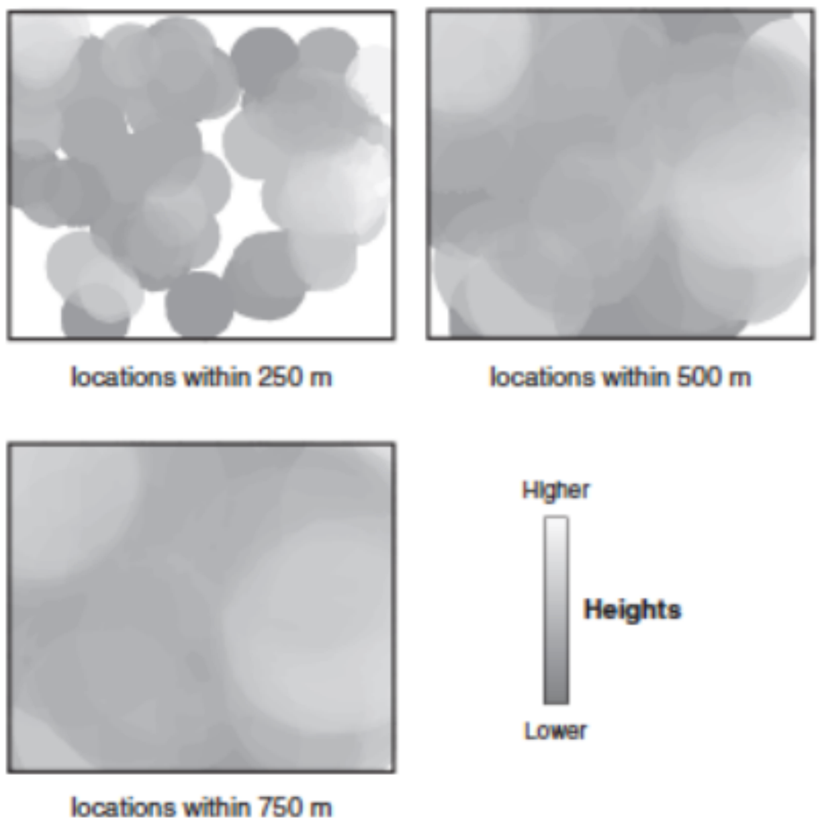


Figure 9.7 Interpolation using the mean of control points within 250, 500, and 750 m of locations to be estimated. Areas where no estimates are made are shown in white.

Source: O'Sullivan and Unwin

Spatial approach 3: Inverse distance weighted spatial average

- A further refinement to interpolating unknown values is to use inverse distance weighting when determining the mean.

- The non-weighted local mean calculation at *m* neighboring control points is:

$$z_j = \frac{1}{m}\sum_{i=1}^{m} z_i$$

- The weighted spatial average can be calculated as in other spatial analysis settings:

$$z_j = \sum_{i=1}^{m} w_{ij} z_i$$

Inverse distance weighting (cont.)

- The basic approach is to set the weights as inversely proportional to the distance between the control point *i* and location *j*

$$w_{ij} \propto \frac{1}{d_{ij}}$$

- More generally, we can adjust the weight using an exponent *k* to arrive at the formula for the weights:

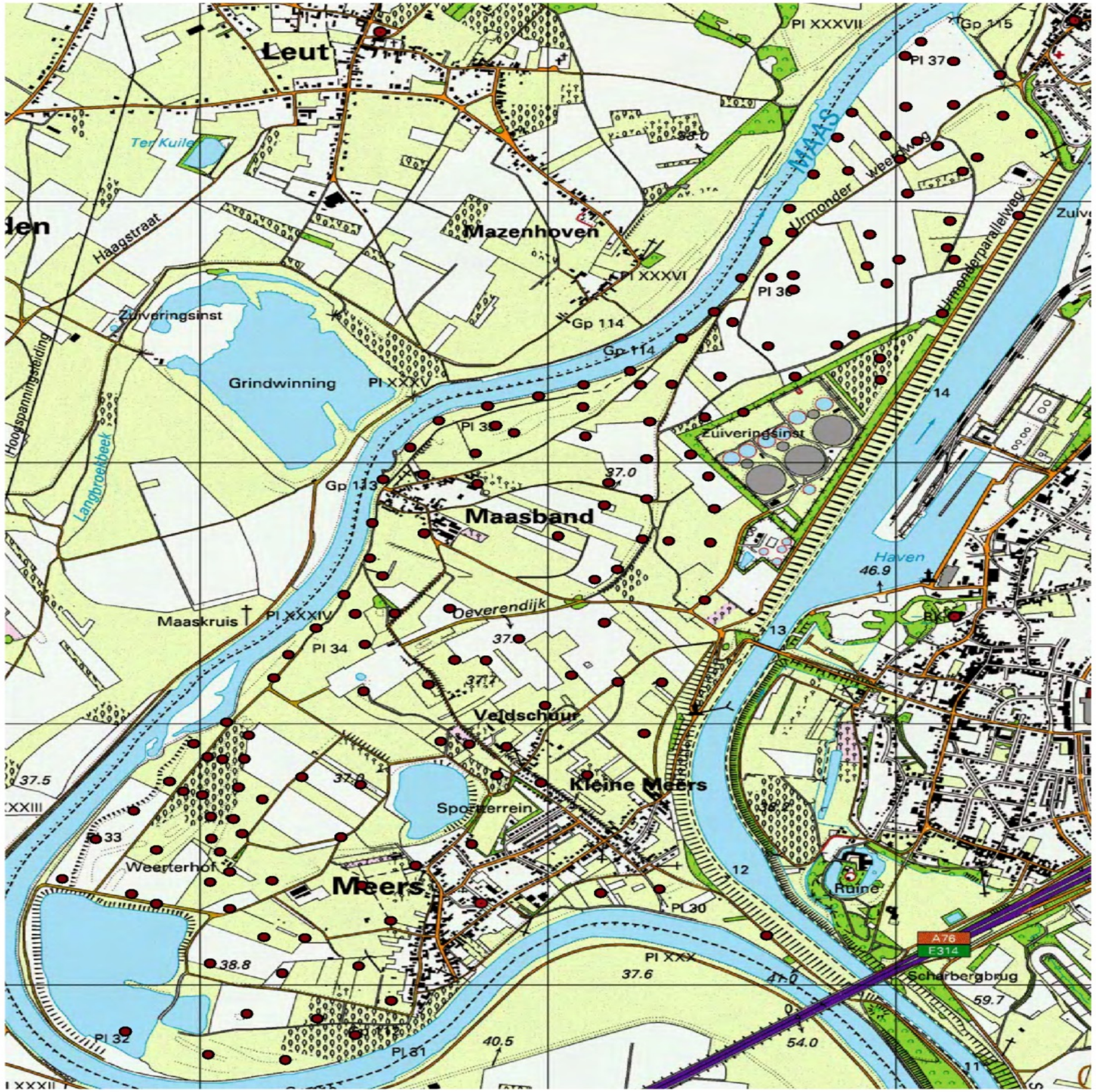
$$w_{ij} \propto \frac{1}{d_{ij}^{k}}$$

- Higher values of k decrease the effect of more distance control points and create a "peakier" map. Values of k less than 1 increase the effect of distant points and create a smoother map.

Meuse data

- We will work through some examples using a data set from the Meuse River
- Historic metal mining has caused the widespread dispersal of lead, zinc, copper and cadmium in the alluvial soil.
- The pollutants may constrain the land use in these areas, so detailed maps are required that identify zones with high concentrations.
- Our specific objective will be to generate a map of a heavy metal (zinc) in soil using point observations, and a range of auxiliary maps.

Meuse river, the Netherlands



Variables in the point data

- xa numeric vector;
- ya numeric vector;
- Cadmium: topsoil cadmium concentration, mg kg-1 soil ("ppm");
- Copper: topsoil copper concentration, mg kg-1 soil ("ppm")
- Lead: topsoil lead concentration, mg kg-1 soil ("ppm")
- Zinc: topsoil zinc concentration, mg kg-1 soil ("ppm")
- Elev: relative elevation above local river bed, m
- Dist: distance to the Meuse; obtained from the nearest cell in meuse.grid,
- Om: organic matter, kg (100 kg)-1 soil (percent)
- Ffreq: flooding frequency class: 1 = once in two years; 2 = once in ten years; 3 = one in 50 years
- Soil: soil type according to the 1:50 000 soil map of the Netherlands
- Lime: lime class: 0 = absent, 1 = present by field test with 5% HCl
- Landuse:
- dist.m: distance to river Meuse in meters

R

- Let’s look at some examples in R…