spatial data analysis 🌐 lab

Using **R** to work with spatial data; using **Kriging** to interpolate density values in a spatial data frame.

Using **R** in a Jupyter Notebook:

- Heavy metals pollution data
- Sampled from the Meuse River in Belgium
- The spatial sampling is not performed on a close-spaced regular grid. Therefore, spatial interpolation or Kriging is used to create a map of the expected concentration of metal pollution.
  - Load and examine the properties of these data.
  - Explore the spatial dependency using the Variogram of the metal pollution data.
  - Use Kriging to interpolate and create a map of metal pollution.

The Meuse data is loaded into **R** and the **data.frame** is converted to a **SpatitalPointsDataFrame**.

A spatial grid is required to interpolate the Pollution Data; the **meuse.grid** data frame is loaded.

The **meuse.grid** data frame is converted to a **SpatialPixelsDataFrame** using the function as before:

In [7]:
```
summary(meuse.grid)
```

```
Object of class SpatialPixelsDataFrame
Coordinates:
     min    max
x 178440 181560
y 329600 333760
Is projected: TRUE
proj4string :
[+proj=stere +lat_0=52.15616055555555 +lon_0=5.38763888888889
+k=0.999908 +x_0=155000 +y_0=463000 +ellps=bessel +units=m +no_defs
+towgs84=565.2369,50.0087,465.658,
-0.406857330322398,0.350732676542563,-1.8703473836068, 4.0812]
Number of points: 3103
Grid attributes:
  cellcentre.offset cellsize cells.dim
x            178460       40        78
y            329620       40       104
Data attributes:
     part.a           part.b            dist          soil     ffreq
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   1:1665   1: 779
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.1193   2:1084   2:1335
 Median :0.0000   Median :1.0000   Median :0.2715   3: 354   3: 989
 Mean   :0.3986   Mean   :0.6014   Mean   :0.2971
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.4402
 Max.   :1.0000   Max.   :1.0000   Max.   :0.9926
```

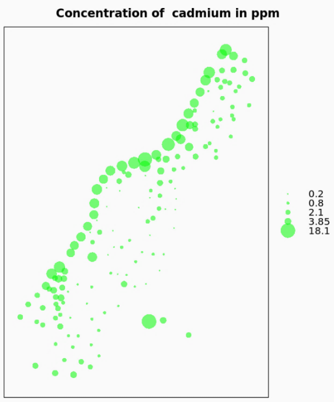
In [4]:
```
summary(meuse)
```

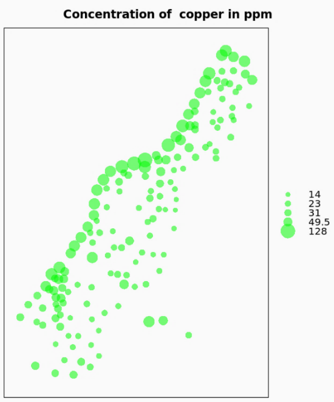
```
Object of class SpatialPointsDataFrame
Coordinates:
     min    max
x 178605 181390
y 329714 333611
Is projected: TRUE
proj4string :
[+proj=stere +lat_0=52.15616055555555 +lon_0=5.38763888888889
+k=0.999908 +x_0=155000 +y_0=463000 +ellps=bessel +units=m +no_defs
+towgs84=565.2369,50.0087,465.658,
-0.406857330322398,0.350732676542563,-1.8703473836068, 4.0812]
Number of points: 155
Data attributes:
    cadmium          copper           lead            zinc
 Min.   : 0.200   Min.   : 14.00   Min.   : 37.0   Min.   : 113.0
 1st Qu.: 0.800   1st Qu.: 23.00   1st Qu.: 72.5   1st Qu.: 198.0
 Median : 2.100   Median : 31.00   Median :123.0   Median : 326.0
 Mean   : 3.246   Mean   : 40.32   Mean   :153.4   Mean   : 469.7
 3rd Qu.: 3.850   3rd Qu.: 49.50   3rd Qu.:207.0   3rd Qu.: 674.5
 Max.   :18.100   Max.   :128.00   Max.   :654.0   Max.   :1839.0

      elev             dist               om          ffreq soil   lime
 Min.   : 5.180   Min.   :0.00000   Min.   : 1.000   1:84   1:97   0:111
 1st Qu.: 7.546   1st Qu.:0.07569   1st Qu.: 5.300   2:48   2:46   1: 44
 Median : 8.180   Median :0.21184   Median : 6.900   3:23   3:12
 Mean   : 8.165   Mean   :0.24002   Mean   : 7.478
 3rd Qu.: 8.955   3rd Qu.:0.36407   3rd Qu.: 9.000
 Max.   :10.520   Max.   :0.88039   Max.   :17.000
                                    NA's   :2
    landuse       dist.m
 W      :50   Min.   :  10.0
 Ah     :39   1st Qu.:  80.0
 Am     :22   Median : 270.0
 Fw     :10   Mean   : 290.3
 Ab     : 8   3rd Qu.: 450.0
 (Other):25   Max.   :1000.0
 NA's   : 1
```

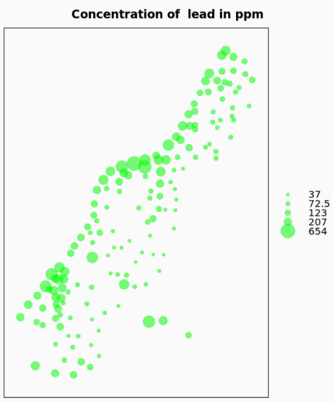
The plots demonstrate the following:

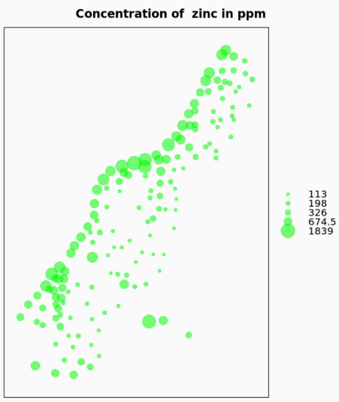
- The distribution of sample points is not uniform in space. The densest sampling occurs along the Meuse River, from the upper right to the lower left of the diagram.
- The distribution of metal concentration is not uniform in space. The highest concentration occurs along the path of the Meuse River. The metals are likely carried by the

Spatial Maps of the sampled metal coordinates where the bubble size represents a numeric value:



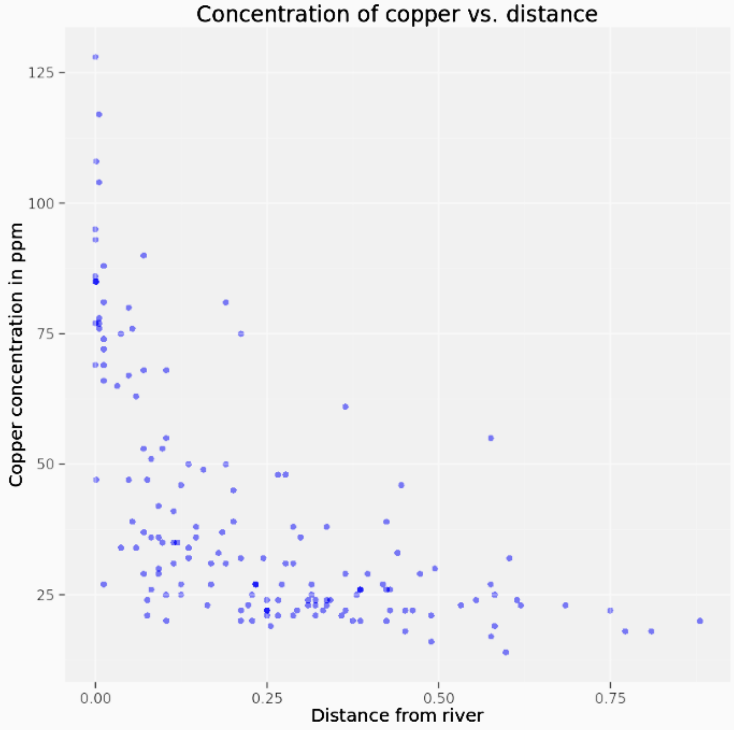


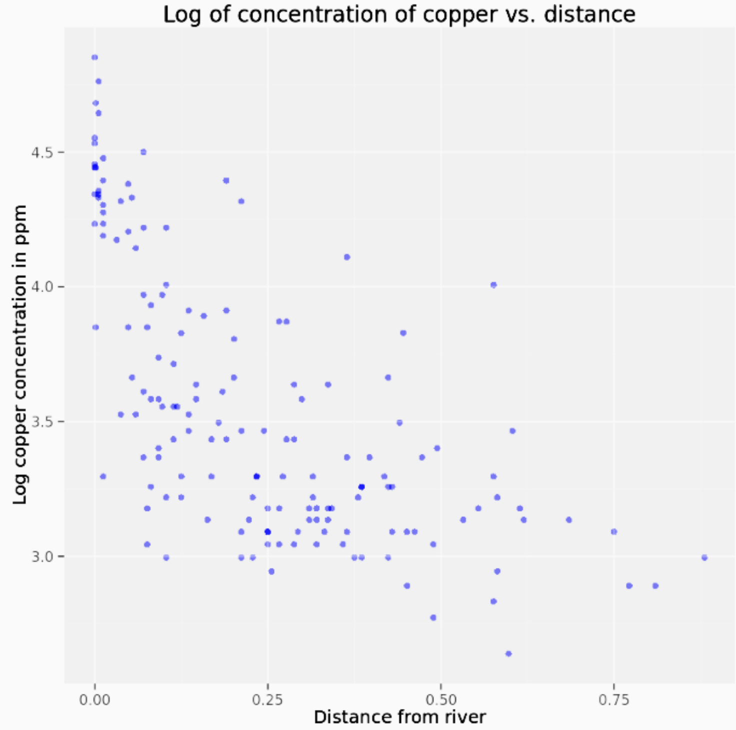




The spatial dependency structure of the metal concentration using **Variogram** methods.

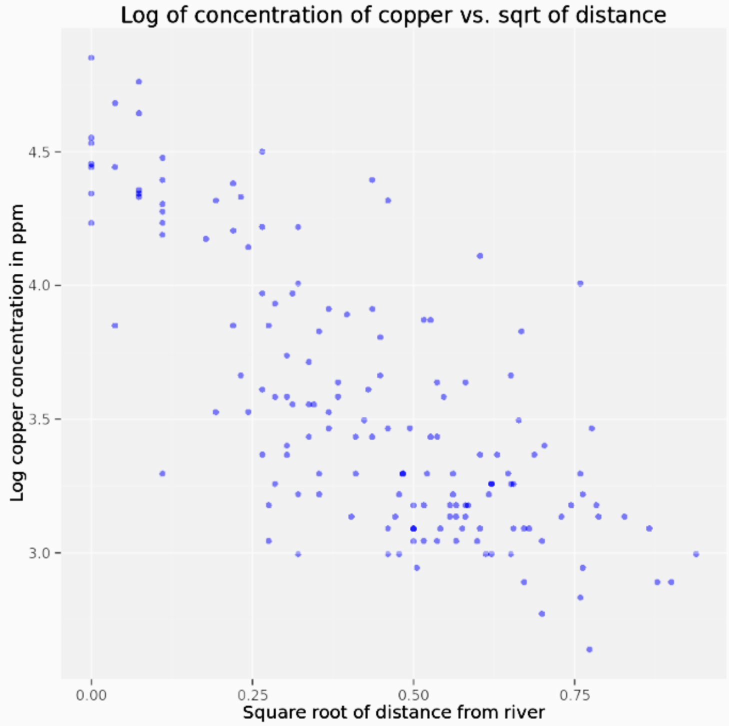
The **dist** column measuring distance from the river is used to compute spatial dependency:

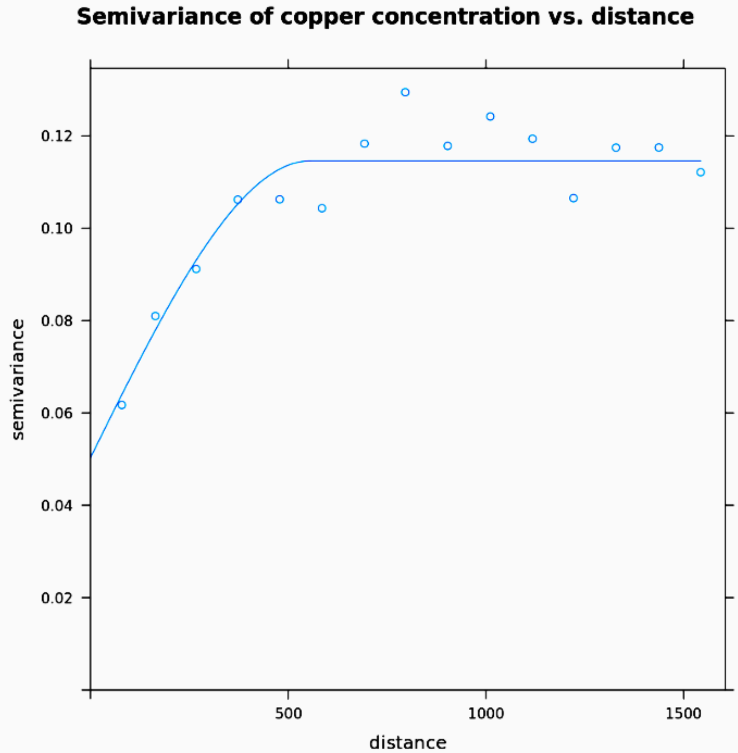




The lack of linearity between copper and distance is more linear after applying logarithmic scale; although still not linear. The scale is further transformed through the log of copper concentration plotted against the squared root of distance from the river; the data is now close to Linear

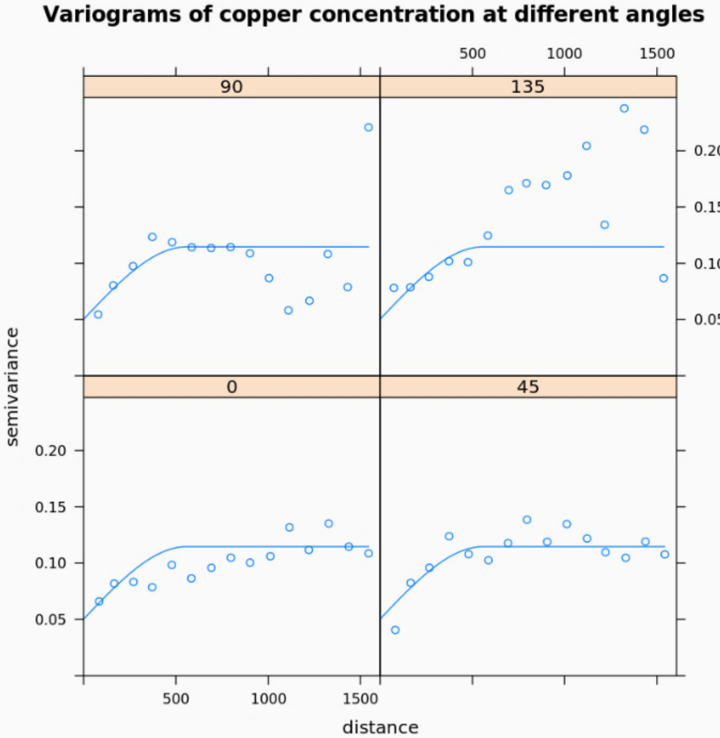
A Variogram can now be properly fitted to the data in the linear diagram

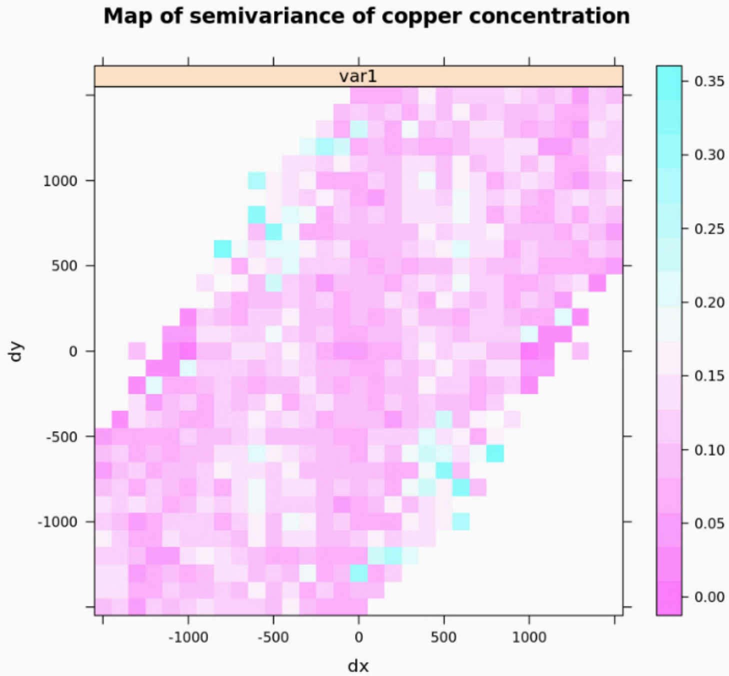




The **semivariance** versus **distance** relationships shows a reasonably good fit.
The **Semivariogram** was computed using the **orthogonal distances** from the river.
However, there can be dependency structures in a spatial diagram in any direction.

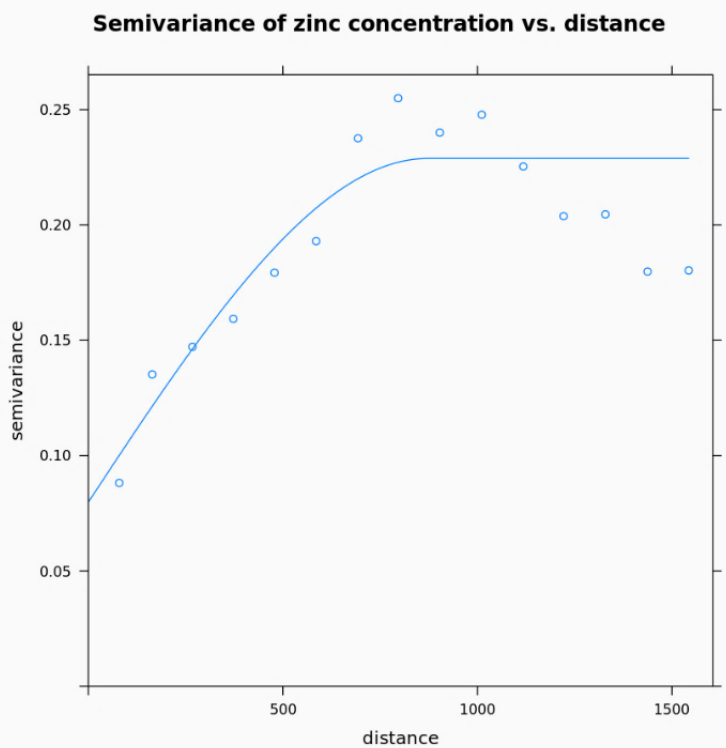
It is noted that the fit remains constant at **0** and **45°** but degrades at **90** and **135°**. The semivariance will proceed to be mapped using the `map = TRUE` argument:





The map illustrates a majority of low semivariance areas with small amounts of high semivariance.

The Variogram analysis is applied equally to the Zinc Concentration of the Meuse River (ex):

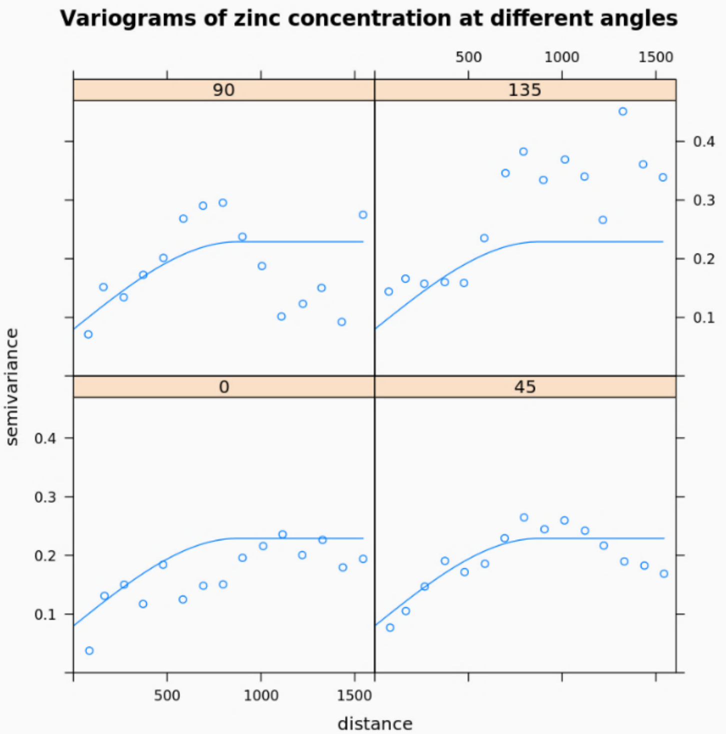


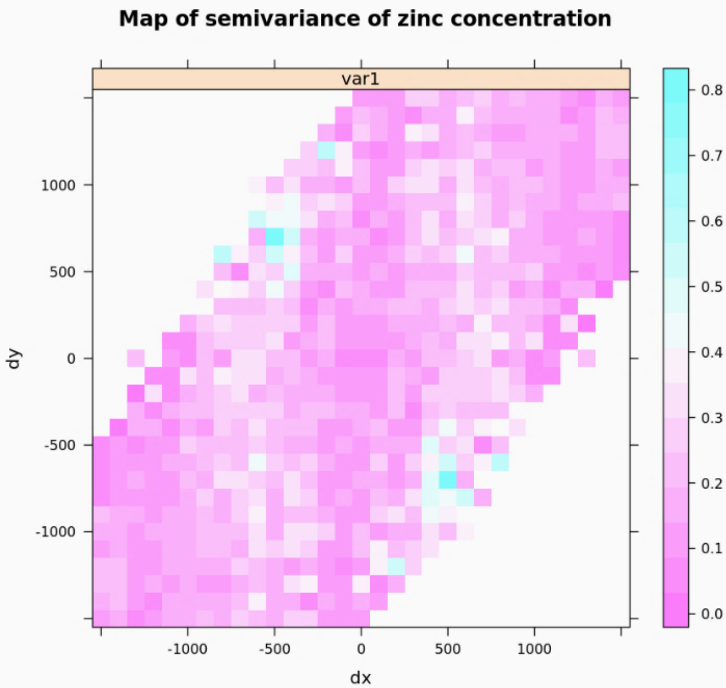
The **semivariance** versus **distance** relationships shows a slightly poorer fit than that of copper.

The **Semivariogram** was computed using the **orthogonal distances** from the river.

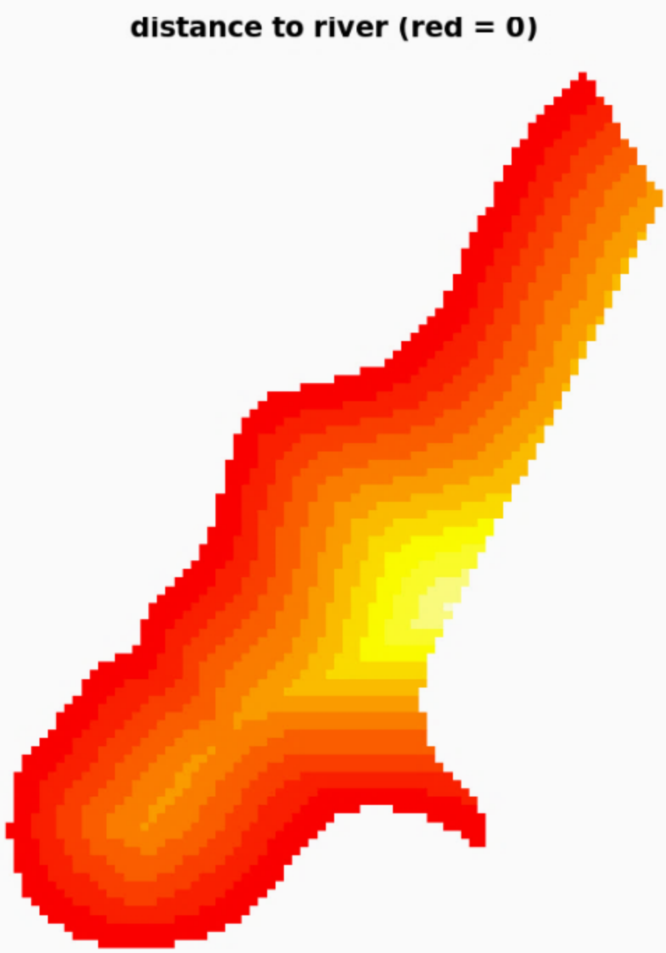
However, there can be dependency structures in a spatial diagram in any direction.

It is noted that the fit remains constant at **0** and **45°** but degrades at **90** and **135°**. The semivariance will proceed to be mapped using the `map = TRUE` argument:





Interpolation of the Copper Concentration



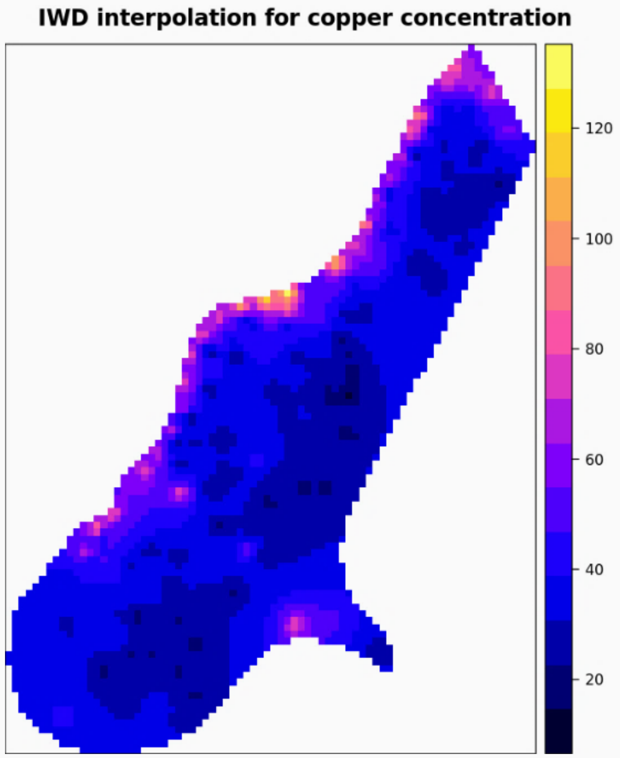
Subsequent to obtaining a valid Variogram, the values can be interpolated. Using the Kriging method, irregularly sampled cooper concentrations can be interpolated to a regular grid.

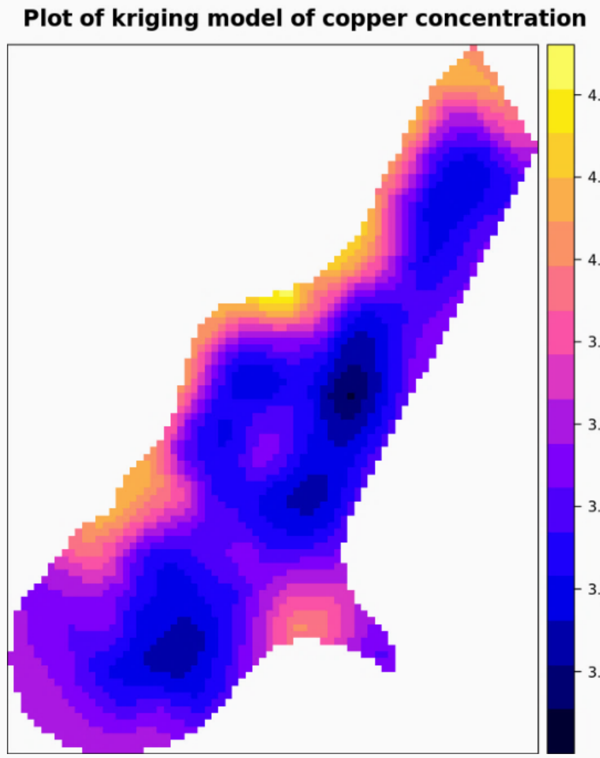
The initial step would be to create a map of the grid used for interpolation (distance from river):

The Meuse River runs from upper right to the lower left of the diagram, curves toward the upper right, and finally toward the lower right. As expected, the lowest distance is near the curve of the river and the greatest distances are near the center (yellow).

Using the distance on the grid the **iwd** function computes the inverse weighted distance of copper concentration. Two new columns are created: **var1.pred** and **var1.var** representing the inverse weighted distance and the variance of computed **IWD**.

The **IWD** is mapped on the following page.





The Inverse Weighted Distance of Copper Concentration contains “hotspots” or high values along the river path; with lower concentrations being found further away from the river. Regardless of valuable insights, the graph does not represent an interpolation.

Kriging uses the **Spatial Dependency Structure** of the copper Variogram computed earlier.

Examining the map of kriged copper concentration. As expected, the highest concentrations of copper pollution are near the Meuse River. Areas of low concentration are further from the river.

Interpolation of the Zinc Concentration