*Spatial Statistics Lab 9*

**Onyedikachi J Okeke**

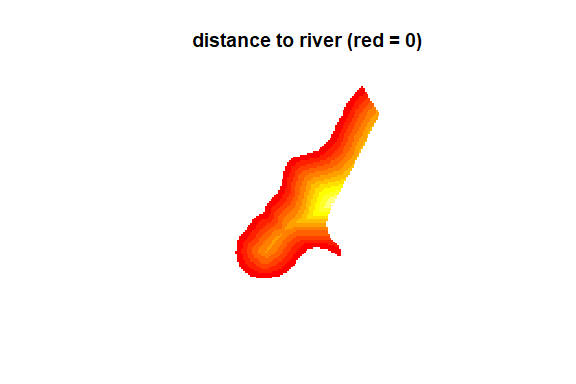
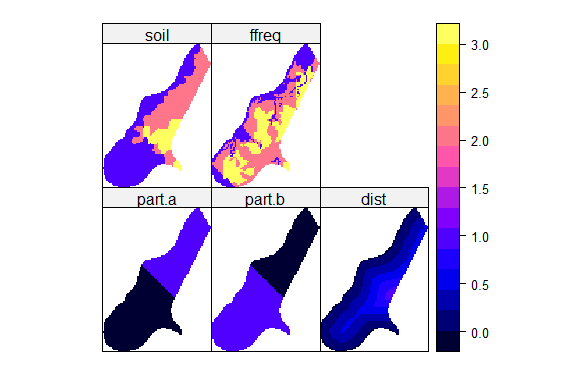
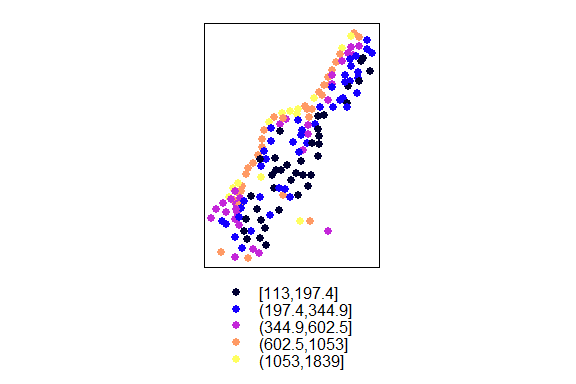
2024-04-24

### 0.0 To Load the library

library(sp)  
library(spatstat)  
library(sf)  
library(spatstat.geom)  
library(ctv)  
library(terra)  
library(spdep)  
library(rgdal)  
library(rgdal)  
library(terra)  
library(RColorBrewer)  
library(classInt)  
library(epitools)  
library(DCluster)  
library(lmtest)  
library(car)  
library(spatialreg)  
library(spdep)  
library(classInt)  
library(spgwr)  
library(ggplot2)  
library(gstat)  
library(lattice)

### 0.1 Loading data set comprises of four heavy metals (e.g. zinc) measured in the top soil.

library(lattice)  
library(sp)  
data(meuse) #use the system file  
coordinates(meuse) <- c("x", "y")  
#map the zinc value  
data(meuse.grid)  
spplot(meuse, "zinc", do.log = T)  
#create a raster surface of distance to river  
coordinates(meuse.grid) = ~x+y  
proj4string(meuse.grid) <- CRS("+init=epsg:28992")  
gridded(meuse.grid) = TRUE  
spplot(meuse.grid)  
image(meuse.grid["dist"])  
title("distance to river (red = 0)")



**Q1. Take a screen shot of the two plots. Do you see any potential correlation between the zinc concentration and distance to river? Explain if any.**

There appears to be a pattern that could indicate a correlation between zinc concentration and distance to the river. In the third image, the area closest to the river is marked in red, denoting a value of 0 (zero distance from the river). As we move further away from the river, the colors shift towards yellow and then orange, indicating increasing distance.

If we compare this pattern with the first image, which seems to show varying concentrations of zinc (assuming each color represents a range of zinc concentration), there is a visible trend where higher concentrations of zinc (depicted by darker colors) tend to occur closer to the river. This spatial distribution suggests that zinc concentration could potentially be higher near the river and decrease with distance.

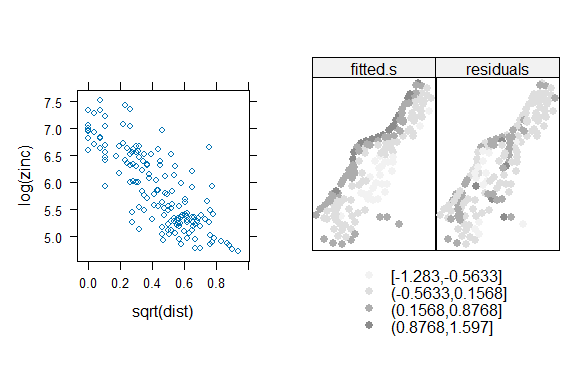
The distance to river chart explicitly shows the distance to the river, with areas closest to the river depicted in red and those farther away in yellow to orange gradients. If we overlay the concept from the first chart onto this heat map, the higher concentrations of zinc seem to be in the areas depicted as closer to the river, which are red.

From a visual standpoint, the charts suggest that there might be a spatial relationship where the concentration of zinc decreases as the distance from the river increases. However, this observation assumes that darker colors represent higher concentrations of zinc, which would need to be verified with the actual data.

## Fit a linear regression to model the relationship between zinc and distance to the river. Plot maps with fitted values and with residuals.

print(xyplot(log(zinc)~sqrt(dist), as.data.frame(meuse), asp = 1), split = c(1,1,2,1), more = TRUE)  
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)  
print(spplot(meuse, c("fitted.s", "residuals"), cuts=4, col.regions=grey.colors(5, 0.95, 0.55, 2.2)), split = c(2,1,2,1))

##### Result



**Q2. Take a screen shot of the plots. Describe the pattern of the residuals.**

The residual plot on the right of the provided chart displays a pattern indicative of heteroscedasticity—a condition where the variability of the residuals is not consistent across all levels of the fitted values. This phenomenon is visually suggested by the funnel-shaped spread of residuals, broadening towards the extremes of fitted values. Such a pattern often points to model mis-specification, meaning that the current model may not be adequately capturing the complexity of the underlying relationship between the predictor variables and the response. Specifically, the residuals are more positive at the low and high ends of the fitted values, while closer to zero in the center range. This could imply that a simple linear model is insufficient, and a non-linear model or a transformation of variables may be necessary to achieve homoscedasticity—constant variance of residuals—and improve the model's predictive accuracy.

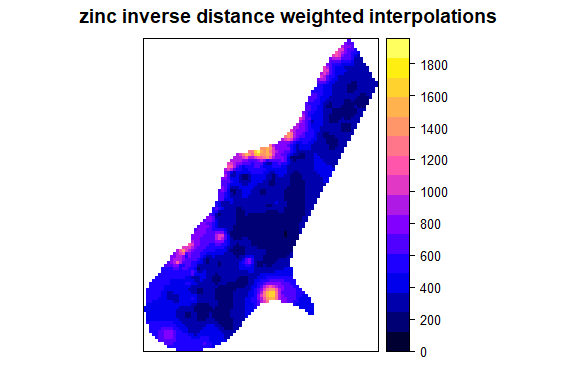
## Converting into a SpatialPixels DataFrame

library(gstat)  
data(meuse.grid)  
coordinates(meuse.grid) <- c("x", "y")  
meuse.grid <- as(meuse.grid, "SpatialPixelsDataFrame")  
###inverse distance weighted interpolation  
  
idw.out <- idw(zinc~1, meuse, meuse.grid, idp = 2.5)  
as.data.frame(idw.out)[1:5,]  
spplot(idw.out["var1.pred"], main = "zinc inverse distance weighted interpolations")

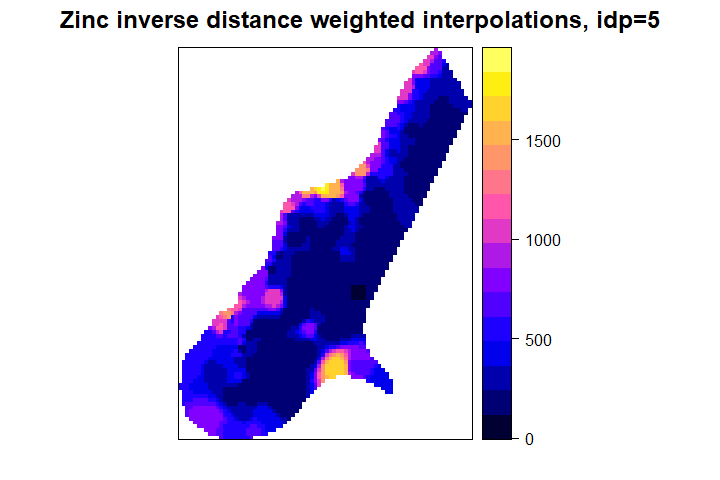
##### Result

## [inverse distance weighted interpolation]

## x y var1.pred var1.var  
## 1 181180 333740 701.9621 NA  
## 2 181140 333700 799.9616 NA  
## 3 181180 333700 723.5780 NA  
## 4 181220 333700 655.3131 NA  
## 5 181100 333660 942.0218 NA

  
**Q3. Take a screen shot of the plot. Modify the idp value to 5 and rerun the above statements. Compare the result against the first one.**

In the first chart, the standard IDW interpolation is applied without specifying the power parameter, which determines how the weight decreases with distance. The color intensity represents higher concentrations of zinc and is most pronounced in specific localized areas, indicating potential hotspots of zinc concentration.



The second chart specifies an IDW power parameter (`idp=5`), which intensifies the influence of nearby points on the interpolation compared to those farther away. The effect of this parameter is visible; the interpolation is smoother, with less variation in zinc concentration across the area. The hotspots are less pronounced, and the overall variation in concentration levels is decreased compared to the first chart.

The two charts provided depict zinc concentration interpolated across a geographical area using an inverse distance weighting (IDW) method, but with a noticeable difference in the parameter settings. The difference between the two interpolations is likely due to the influence of the power parameter on the weighting function. A higher power parameter value like `idp=5` means that closer observations will have a much greater influence on the interpolated values than those farther away, leading to smoother gradients and potentially less pronounced extremes. This suggests that the choice of IDW parameters can significantly affect the interpretation of spatial data, which should be carefully considered in environmental assessments and policy-making.

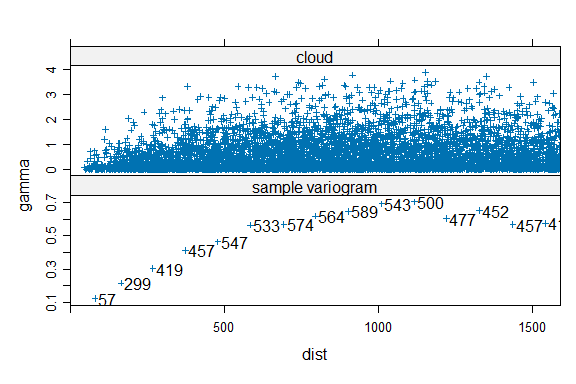
## Variograms calculations

library(lattice)  
  
log(zinc)~1   
#means that we assume a constant trend for the variable log(zinc).  
svgm <- variogram(log(zinc) ~ 1, meuse)  
  
svgm  
  
cld <- variogram(log(zinc) ~ 1, meuse, cloud = TRUE)  
svgm <- variogram(log(zinc) ~ 1, meuse)  
d <- data.frame(gamma = c(cld$gamma, svgm$gamma),  
dist = c(cld$dist, svgm$dist),  
id = c(rep("cloud", nrow(cld)), rep("sample variogram", nrow(svgm)))  
)  
xyplot(gamma ~ dist | id, d,  
scales = list(y = list(relation = "free", ylim = list(NULL, c(-.005,0.7)))),  
layout = c(1, 2), as.table = TRUE,  
panel = function(x,y, ...) {  
if (panel.number() == 2)  
ltext(x+10, y, svgm$np, adj = c(0,0.5)) #$  
panel.xyplot(x,y,...)  
},  
xlim = c(0, 1590),  
cex = .5, pch = 3  
)

##### Result

## log(zinc) ~ 1

## np dist gamma dir.hor dir.ver id  
## 1 57 79.29244 0.1234479 0 0 var1  
## 2 299 163.97367 0.2162185 0 0 var1  
## 3 419 267.36483 0.3027859 0 0 var1  
## 4 457 372.73542 0.4121448 0 0 var1  
## 5 547 478.47670 0.4634128 0 0 var1  
## 6 533 585.34058 0.5646933 0 0 var1  
## 7 574 693.14526 0.5689683 0 0 var1  
## 8 564 796.18365 0.6186769 0 0 var1  
## 9 589 903.14650 0.6471479 0 0 var1  
## 10 543 1011.29177 0.6915705 0 0 var1  
## 11 500 1117.86235 0.7033984 0 0 var1  
## 12 477 1221.32810 0.6038770 0 0 var1  
## 13 452 1329.16407 0.6517158 0 0 var1  
## 14 457 1437.25620 0.5665318 0 0 var1  
## 15 415 1543.20248 0.5748227 0 0 var1



**Q4. Take a screen shot of the plot. What are the differences between a variogram cloud (top one) and empirical variogram (button one).**

The variogram cloud and the empirical variogram are two representations of spatial dependence, but they display data differently:

**Variogram Cloud:**

The top plot, known as a variogram cloud, shows every pair of observations as a point, with the separation distance on the x-axis and the semivariance on the y-axis.

It provides detailed insight into the variability between all individual pairs of data points, making it easier to spot outliers or anomalies in spatial continuity.

This cloud can appear cluttered due to the sheer number of data point comparisons, especially for large datasets.

**Empirical Variogram:**

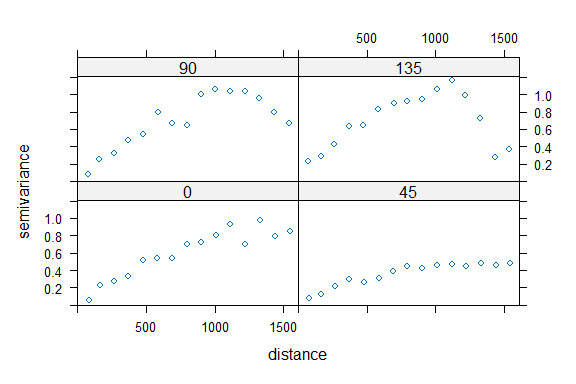
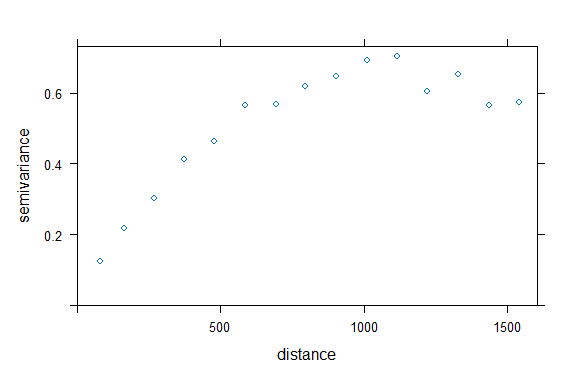
The bottom plot is the empirical variogram, which aggregates the semivariances of data points that fall within specified distance intervals (or bins) and averages them.

This method provides a more generalized and clearer overview of spatial dependence, showing the average degree of variability for pairs of points at different distances.

The empirical variogram is smoother and less detailed than the variogram cloud, as it reduces variability within the bins to a single mean value, which can be used to model the variogram.

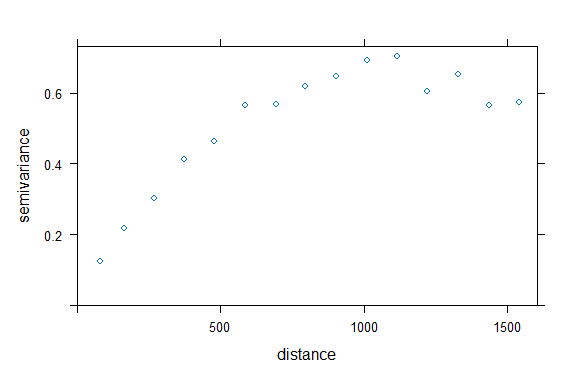
## Cutoff, Lag Width, Direction Dependence

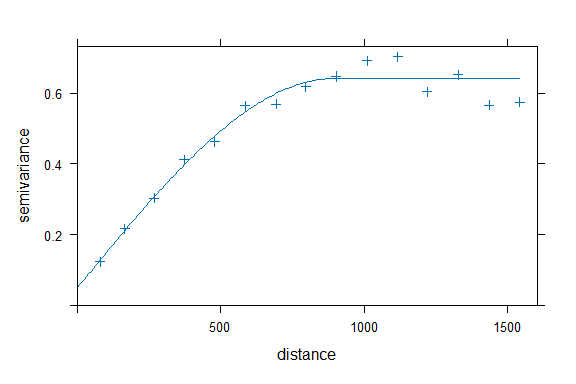
plot(variogram(log(zinc) ~ 1, meuse))  
  
plot(variogram(log(zinc) ~ 1, meuse, alpha = c(0, 45, 90, 135)))



Q5. Take a screen shot of the plot. Describe and compare the patterns in the four subplots.

## Variogram Modelling

v <- variogram(log(zinc) ~ 1, meuse)  
plot(v)  
v <- variogram(log(zinc) ~ 1, meuse)  
v.fit <- fit.variogram(v, vgm(1, "Sph", 800, 1)) ##spherical model  
plot(v, v.fit, pch = 3)

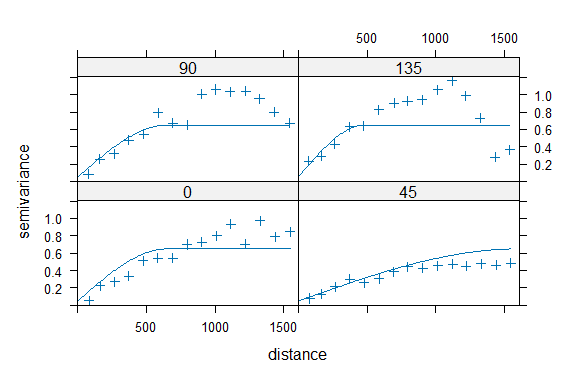


Q6. Take a screen shot of the plot.

## Classify directions into four direction intervals

v.dir <- variogram(log(zinc)~1,meuse,alpha=(0:3)\*45)  
v.anis <- vgm(.6, "Sph", 1600, .05, anis=c(45, 0.3))  
print(plot(v.dir, v.anis, pch=3))

##### Result



Q7. Take a screen shot of the plot.

## Simple, Ordinary, and Universal Kriging

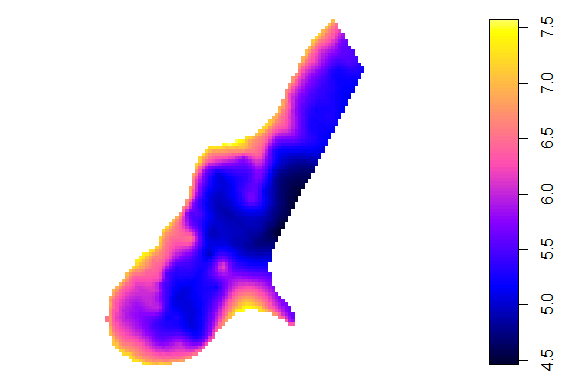
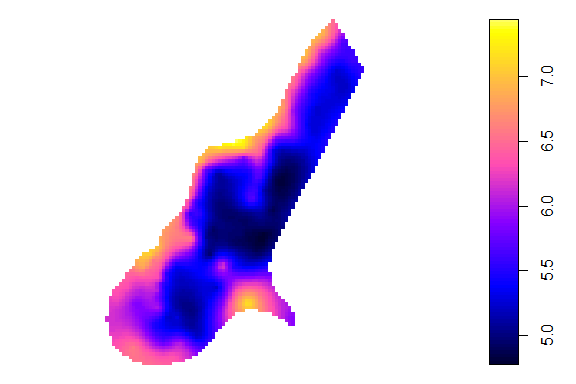
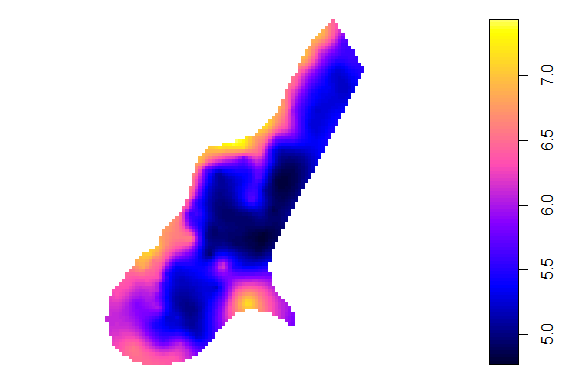
lz.sk <- krige(log(zinc)~1, meuse, meuse.grid, v.fit, beta = 5.9)  
lz.ok <- krige(log(zinc)~1, meuse, meuse.grid, v.fit)  
lz.uk <- krige(log(zinc)~sqrt(dist), meuse, meuse.grid, v.fit)  
plot(lz.sk)  
plot(lz.ok)  
plot(lz.uk)

##### Result

## [using simple kriging]

## [using ordinary kriging]

## [using universal kriging]



**Q8. Take a screen shot of these plots and describe their differences.**

**Simple Kriging:** This method would produce a map with smooth transitions between zinc concentration levels, assuming a constant mean. The resulting interpolation might appear more uniform, with less local variability since it relies on a global mean that doesn't change over space.

**Ordinary Kriging**: The map from ordinary kriging would likely show more variation and potentially more detail in the spatial pattern. Since ordinary kriging does not assume a constant mean, it adapts to local variations, resulting in a map with potentially more defined high and low zinc concentration areas based on the local data points.

**Universal Kriging:** For universal kriging, the map would include a spatial trend or drift in the data. This means you'd see a pattern indicating a gradient or directional trend in zinc concentrations. The interpolation would capture this trend on top of the local variations, potentially creating a more complex map with varying gradients.

## Model Diagnostics (cross validation)

sel100 <- sample(1:155, 100)  
m.model <- meuse[sel100,]  
m.valid <- meuse[-sel100,]  
v100.fit <- fit.variogram(variogram(log(zinc)~1, m.model), vgm(1, "Sph", 800, 1))  
m.valid.pr <- krige(log(zinc)~1, m.model, m.valid, v100.fit) #[using ordinary kriging]  
resid.kr <- log(m.valid$zinc) - m.valid.pr$var1.pred  
summary(resid.kr)  
resid.mean <- log(m.valid$zinc) - mean(log(m.valid$zinc))  
R2 <- 1 - sum(resid.kr^2)/sum(resid.mean^2)  
R2

##### Result

## [using ordinary kriging]

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.98266 -0.30203 -0.08812 -0.10224 0.15703 0.65301

## [1] 0.6671816

The ordinary kriging results indicate a median value of -0.08812 and a mean of -0.10224, suggesting a slight negative bias in the data. The interquartile range, from -0.30203 to 0.15703, shows moderate variability. The value 0.6671816 may represent a measure of spatial autocorrelation or kriging error, indicating a moderate level of spatial dependency or prediction accuracy within the dataset.