Solutions to Exercise Series 3

1. Boxplots.

a) *Create a boxplot of the lead content. Add a title and label the axis with the variable name and the unit of the measurements.*

```
> boxplot( d.jura[,"pb"], ylab="Pb content [mg/kg]",
+     main="Metals in soils around La Chaux-de-Fonds"
+ )
```

b) *As **a)**, but use a logarithmic scale for the axis and display the data as a horizontal boxplot. Add lines to mark the guide and the trigger values. How many observations exceed the guide and the trigger value?*

```
> boxplot( d.jura[,"pb"], horizontal=TRUE, log="x",
+     xlab="Pb content [mg/kg]",
+     main="Metals in soils around La Chaux-de-Fonds"
+ )
> abline( v=c(50,200), col=c("orange","red"), lwd=2, lty="dotted" )
> table( cut( d.jura[,"pb"], breaks=c(0,50,200,1000) ) )
    (0,50]    (50,200] (200,1e+03]
       150         107           2
```

107 (2) `pb` measurements exceed the guide (trigger) value.

c) *Create boxplots of the lead content for all categories of* `rock`*. Add notches to display approximate 95 % confidence intervals of the group medians. Does the lead content differ between the rock types? Why is the confidence interval for category* portlandian *much wider than the intervals of the other categories?*

```
> boxplot( pb~rock, data=d.jura, notch=T, log="y",
+     names=c("arg","kim","port","quart","seq" ),
+     xlab="Rock type", ylab="Pb content [mg/kg]",
+     main="Pb content per rock type"
+ )
> table( d.jura[ "rock"] )
   argovian kimmeridgian  portlandian   quaternary    sequanian
         53           85            3           55           63
```

There are only 3 observations in the category `portlandian`.

d) *Create boxplots of all 7 metals in a single plot.*

```
> boxplot( d.jura[,c("cd","co","cr","cu","ni","pb","zn")],
+     log="y", ylab="metal content [mg/kg]"
+ )
```

Scatterplots.

2. **a)** *Create a plot of the positions of the sampling locations. Add a title to the plot and label the axis.*

```
> plot( y~x, data=d.jura, asp=1, xlab="easting [km]", ylab="northing [km]",
+     main="Positions of sampling locations"
+ )
```

b) *Create a plot that displays the spatial distribution of the rock type by different symbols and add a legend.*

```
> plot( d.jura[,c("x","y")], asp=1, pch=as.numeric( d.jura[,"rock"] ),
+     xlab="easting [km]", ylab="northing [km]",
+     main="Positions of sampling locations"
+ )
> legend( "topright", pch=1:nlevels( d.jura[,"rock"] ),
+     legend=substring( levels(d.jura[,"rock"]), 1, 3 ), bty="n"
+ )
```

c) *As **b)**, but add another set symbols to show the landus. Add legends for both factors.*

```
> plot( d.jura[,c("x", "y")], asp=1, pch=as.numeric( d.jura[,"rock"] ),
+     xlab="easting [km]", ylab="northing [km]",
+     main="Positions of sampling locations"
+ )
> points( d.jura[,c("x","y")], pch=1, cex=2,
+     col=as.numeric( d.jura[,"landuse"] )
+ )
> legend( "topright", pch=c( NA, 1:nlevels(d.jura[,"rock"]) ),
+     legend=c("Rock type", substring( levels(d.jura[,"rock"]), 1, 3 ) ), bty="n"
+ )
> legend( "topleft", col=c( NA, 1:nlevels( d.jura[,"landuse"] ) ), pch=1,
+     legend=c( "Landuse", levels( d.jura[,"landuse"] ) ), bty="n"
+ )
```

d) *Create a "bubble plot" for the lead content, color the circles by rock type and add legends. Can you detect any patterns in the spatial distribution of* `pb`*?*

```
> plot( d.jura[,c("x","y")], asp=1, col=as.numeric( d.jura[,"rock"] ),
+     cex=sqrt( d.jura[,"pb"] )/5
+ )
> t.lead <- c( 20, 50, 100, 200 )
> legend( "topleft", x.intersp=1.5, y.intersp=1.7,
+     legend=c( "lead [mg/kg]", as.character(t.lead) ),
+     pch=1, pt.cex=c( NA, sqrt(t.lead)/5 ),  bty="n"
+ )
> legend( "topright", pch=1, col=c( NA, 1:nlevels( d.jura[,"rock"] ) ),
+     legend=c( "Rock type", substring( levels( d.jura[,"rock"] ), 1, 3 ) ),
+     bty="n"
+ )
```

Pb measurement at sites on Portlandian rock (black symbols) are on average a bit smaller (cf. Problem 1c). Furthermore, measurements at adjacent points appear to be more similar to each other than measurements at points spaced farther apart.

e) *Read the lead content for a couple of points from the plot generated in **d)** by using* `identify()`*. What do you notice about the spatial variation of* `pb` *at different spatial scales?*

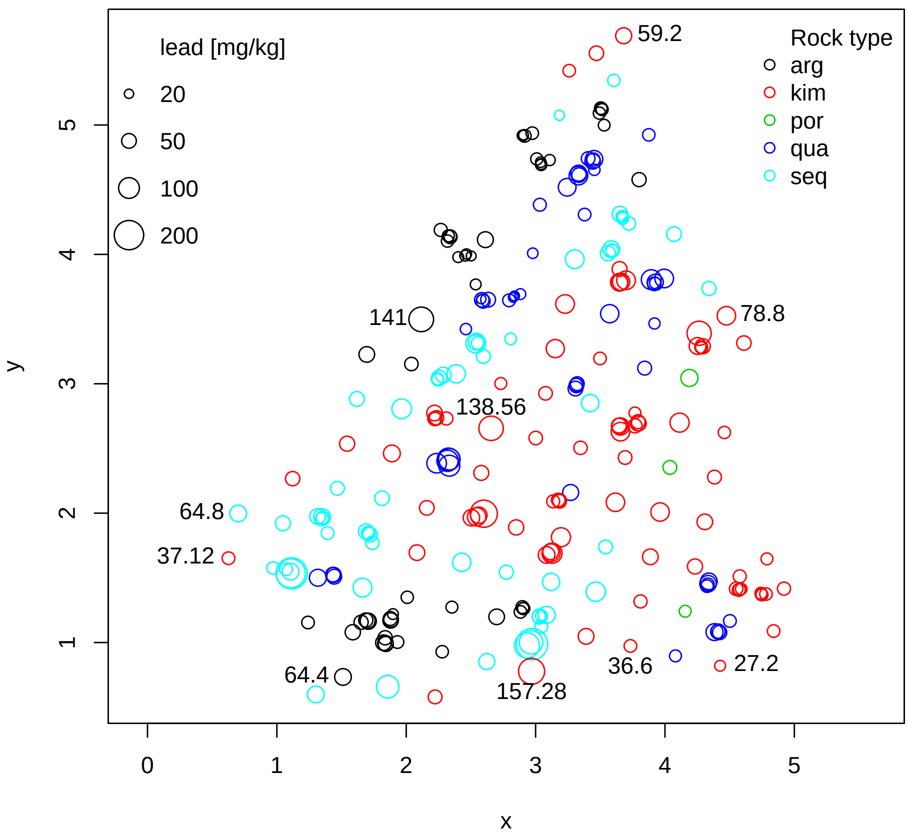

Figure 1: Bubble plot of lead (Problem 2d & 2e).

```
> identify( d.jura[,c("x","y")], labels=d.jura[,"pb"] )
[1]    8   90   98 107 108 140 142 147 185 216
> plot( d.jura[, ("x", y")], asp=1, xlim=c(1,3), ylim=c(1,3),
+      cex=sqrt(d.jura[ "pb"])/5
+ )
> identify( d.jura[,c("x","y")], labels=d.jura[,"pb"] )
[1]  60 112 196 197
```

`identify` returns for the selected points the row numbers in the matrix or dataframe. As stated above, points close to each other have similar `pb` content.

f) *Re-create a plot with the sampling locations and use `locator()` and `polygon()` to draw a bounding polygon around the survey area.*

```
> plot( d.jura[,c("x","y")], asp=1 )
> t.coords <- locator( type="p", pch=3, col="red" )
> polygon( t.coords, border="red", lty="dashed" )
```

g) *Find out what happens if you type `plot(y~x, data=d.jura, type="l")`. What rule is used to connect the points?*

`plot(..., type="l"|"o"|"b"|"c")` and `lines(...)` connects points corresponding to subsequent elements (rows) in a vector, matrix, or in a dataframe by line segments.

h) *Generate again a plot of the sampling locations and annotate them by the row numbers of the observations in the dataframe.*

```
> plot( y ~ x, data=d.jura, asp=1)
> text( d.jura[,c("x","y")], labels= 1:259, pos=4, cex=0.7)
```