**Ghement Statistical Consulting Company Ltd. © 2013**

**Dealing with large numbers on the vertical axis of a scatterplot**

Often, when we create scatterplots in R, we have to accommodate situations where the numbers displayed on the vertical axis are quite large.

**Example:**

Let us generate some x and y data in R and use these data to create a scatterplot of y versus x. The x values will be small, covering the range 1 to 10. In contrast, the y values will be large, covering the range ...

set.seed(1)

x <- seq(1,10,length=100)

y <- x\*100000 + rnorm(n=length(x),mean=0, sd=50000)

If we the R function plot() to create the scatterplot of y versus x, we will get a default plot which is not very appetizing.

plot(y ~ x)



Two features which make this scatterplot difficult to interpret are:

1. The placement of the y-axis values (parallel to the y axis instead of vertical on the y axis);
2. The use of the scientific notation to display the large numbers appearing on the y axis (e.g., 2e+05 stands for 200,000).

The first of these undesirable features can be addressed using the option las=1 of the plot() function in R.

plot(y ~ x, las=1)



However, now we encounter a new problem – the y-axis label is located too close to the numbers displayed on the y-axis. The formatting of these numbers using scientific notation still needs to be addressed.

Here is the R code that we need to use to obtain a scatterplot where the y-axis label is located far enough from the numbers displayed on this axis and the formatting of the number is no longer using the scientific notation:

par(mar=c(4,6,2,2))

plot(y ~ x, xaxt="n", yaxt="n", xlab="", ylab="")

axis(1,at=pretty(x),labels=pretty(x),las=1)

axis(2,at=pretty(y),labels=format(pretty(y),big.mark=",", scientific=FALSE),las=1)

mtext(text="x", side=1, line=2)

mtext(text="y", side=2, line=5)

title("Scatterplot of y versus x")

The resulting scatterplot is shown below.



The previous R code uses some coding tricks, as follows:

1) Set the left margin of the scatterplot so that it is wider in order to accomodate the large numbers on the y-axis of the scatterplot.

 par(mar=c(4,6,2,2))

2) Construct a scatterplot of y versus x using the function plot() so that it has no x-axis (xaxt="n"), no y-axis (yaxt="n"), no x-axis label (xlab="") and no y-axis label(ylab="").

plot(y ~ x, xaxt="n", yaxt="n", xlab="", ylab="")

3) Add an x-axis to the scatterplot and use the R function pretty() to determine the values to be displayed on the x-axis.

axis(1,at=pretty(x),labels=pretty(x),las=1)

4) Add an y-axis to the scatterplot. Use the R function pretty() to determine the values to be displayed on the y-axis. Use the function format() with the options big.mark="," and scientific=FALSE to format the values to be displayed on the y-axis.

axis(2,at=pretty(y),

 labels=format(pretty(y),big.mark=",", scientific=FALSE),

 las=1)

5) Use the function mtext() to place a label on the x-axis. Because text() was invoked with the options side=1 and line=2, the label will be placed on the second line below the x-axis.

mtext(text="x", side=1, line=2)

5) Use the function mtext() to place a label on the y-axis. Because text() was invoked with the options side=2 and line=5, the label will be placed on the fifth line to the left of the y-axis.

mtext(text="y", side=2, line=5)

6) Add a title to the scatterplot.

title("Scatterplot of y versus x")

~.~

To understand how the mtext() function works, let’s look at the R code below and its corresponding output:

set.seed(1)

x <- seq(1,10,length=100)

y <- x\*100000 + rnorm(n=length(x),mean=0, sd=50000)

par(mar=c(4,6,2,2))

plot(y ~ x, type="n",col.axis= "grey",ylab="",xlab="")

mtext(text="Line 0; Side 2", side=2, line=0, cex=0.8)

mtext(text="Line 1; Side 2", side=2, line=1, cex=0.8)

mtext(text="Line 2; Side 2", side=2, line=2, cex=0.8)

mtext(text="Line 3; Side 2", side=2, line=3, cex=0.8)

mtext(text="Line 4; Side 2", side=2, line=4, cex=0.8)

mtext(text="Line 5; Side 2", side=2, line=5, cex=0.8)

****

Line 1 corresponds to where R would place the y-axis labels.

Line 0 is located immediately below Line 1.

Line 2 is located immediately above Line 1.

Line 3 is located immediately above Line 2.

Line 4 is located immediately above Line 3.

Line 5 is located immediately above Line 4.

Because we invoked the command par(mar=c(4,**6**,2,2)) right before we constructed the scatterplot, we can only have a total of 6 lines to the left of the y-axis, counting from 0 to 5.