```
#        Cours de Stat Mass 05: regression linéaire / résidus
# initialisation des variables "Tailles" et "x", etc.
library(MASS)  # MASS = Modern Applied Statistics with S
data(survey)   #lecture des données de  survey
survey.cc<-survey[complete.cases(survey),] #nettoyage de  survey
x<-survey.cc$Wr.Hnd;x  #x  prends la valeur de la première colonne de  Tailles.
y<-survey.cc$NW.Hnd  #y  prends la valeur de la deuxième colonne de  Tailles.cc
z<-survey.cc$Height  #z  prends la valeur de la troisième colonne de  Tailles.cc

# Nuages
plot(x,y,pch=3) # pch=3 -> "+"  ; pch=10 ajouterait un rond au tour
points(mean(x),mean(y),col="red",pch=19) # (mean(x),mean(y)) est le Centre de Gravité

# Basic Scatterplot Matrix
pairs(~Wr.Hnd+NW.Hnd+Height,data=survey,
   main="Simple Scatterplot Matrix")

# On recommence avec des noms plus simples: x, y, z
MonTableau=data.frame(x,y,z)
head(MonTableau)
# Scatter-plot
pairs(~x+y+z,data=MonTableau,
   main="Matrice des nuages deux à deux")

# un nuage peu rectiligne
plot(x,z,type="p",pch=8)

# un nuage assez rectiligne
plot(x,y,type="p",pch=3)   # pch=1->ronds; 2->triangles; 3->"+";
# et sa regression de y sur x
DteReg_y_sur_x=lm(y~x,data=MonTableau)
abline(coef(DteReg_y_sur_x),col="blue")#ordonnée en 0 (b=Intercept) et pente (a) de la droite y=a
coef(DteReg_y_sur_x)
points(mean(x),mean(y),col="white",pch=19)
# summary(DteReg_y_sur_x) étudierons plus tard ce que fait cette commande
# les "prédites" y de x
y2x=fitted(DteReg_y_sur_x)
points(x,y2x,col="green",pch=1)
points(mean(x),mean(y),col="red",pch=19)

#retour au nuage peu rectiligne des (x,z)
plot(x,z,,pch=3) # pch=3 donnes des "+"
DteReg_z_sur_x=lm(z~x,data=MonTableau)
abline(coef(DteReg_z_sur_x),col="blue")
points(mean(x),mean(z),col="red",pch=19)
coef(DteReg_z_sur_x)#ordonnée à l'origine (b) et pente (a) de la droite y=ax+b (sic!)
# les "prédites" z de x
z2x=fitted(DteReg_z_sur_x)
points(x,z2x,col="green",pch=1)

#étude des "résidus"
windows();dev.cur()   # on ouvre une fenêtre et affiche le nom
Res_z2x=z - z2x
plot(x,Res_z2x)
abline(coef(lm(Res_z2x~x,data=MonTableau)))
hist(Res_z2x)
#hist(Res_z2x,add=T)

# on revient aux "bien alignés" (x,y)
dev.set(2);plot(x,y)
abline(coef(DteReg_y_sur_x),col="blue") # ordonnée à l'origine (a) et pente (b) de la droite y=bx
Res_y2x=y - y2x
dev.set(3)
plot(x,Res_y2x)
abline(coef(lm(Res_z2x~x,data=MonTableau)))
# hist(Res_y2x,add=TRUE)
hist(Res_y2x)
boxplot(Res_z2x,Res_y2x)
hist(Res_z2x)
hist(Res_y2x,add=TRUE)
```

```
# summary(DteReg_y_sur_x)
# summary(DteReg_z_sur_x)

### Call:
### lm(formula = y ~ x, data = MonTableau)
###
### Residuals:
###      Min       1Q   Median       3Q      Max
### -1.92341 -0.42235  0.07489  0.37473  1.67616
###
### Coefficients:
###             Estimate Std. Error t value Pr(>|t|)
### (Intercept) -0.11481    0.36856  -0.311    0.756
### x            1.00212    0.01956  51.236   <2e-16 ***
### ---
### Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
###
### Residual standard error: 0.5357 on 206 degrees of freedom
### Multiple R-squared:  0.9272,    Adjusted R-squared:  0.9269
### F-statistic:  2625 on 1 and 206 DF,  p-value: < 2.2e-16



##### summary(DteReg_z_sur_x)
##### Call:
##### lm(formula = z ~ x, data = MonTableau)
##### Residuals:
#####      Min       1Q   Median       3Q      Max
##### -19.7276  -5.0706  -0.8269   4.9473  25.8704
#####
##### Coefficients:
#####             Estimate Std. Error t value Pr(>|t|)
##### (Intercept) 113.9536     5.4416   20.94   <2e-16 ***
##### x             3.1166     0.2888   10.79   <2e-16 ***
##### ---
##### Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#####
##### Residual standard error: 7.909 on 206 degrees of freedom
##### Multiple R-squared:  0.3612,    Adjusted R-squared:  0.3581
##### F-statistic: 116.5 on 1 and 206 DF,  p-value: < 2.2e-16
```