Transformations

```
# Transforming the X variable
# Example (page 129)
x=c(.5,.5,1,1,1.5,1.5,2,2,2.5,2.5)
y=c(42.5,50.6,68.5,80.7,89,99.6,105.3,111.8,112.3,125.7)
results=lm(y~x)
summary(results) # R^2=0.9256
plot(x,results$res) # note the curve pattern
qqnorm(results$res) # no evidence of non-normality of residuals
bf(x,y) # no evidence of deviation

xp=sqrt(x)
results=lm(y~xp)
summary(results) # R^2=0.9545 - note the improvement
plot(xp,results$res)
qqnorm(results$res)

# Transforming the Y variable
# Example: Plasma (page 132)
data=read.csv("data_plasma.csv",header=T)
attach(data)
plot(Age,Plasma)
results=lm(Plasma~Age)
plot(Age,results$res)
qqnorm(results$res)

log10plasma=log(Plasma,base=10)
plot(Age,log10plasma)
results2=lm(log10plasma~Age)
plot(Age,results2$res)
qqnorm(results2$res)

# Box-Cox Transformations
y=Plasma;x=Age

lambda=seq(-2,2,by=.1)
lambda[21]=0.01 # Just to avoid the special case, lambda=0
iter=length(lambda)
SSEs=array(0,iter) # Storage for the SSEs
k2=(prod(y))^(1/length(y))

for(i in 1:iter){
k1=1/(lambda[i]*k2^(lambda[i]-1))
w=k1*(y^lambda[i]-1)
results=lm(w~x)
SSEs[i]=anova(results)$Sum[2]}

w=k2*(log(y)) # This is for the special case when lambda=0
results=lm(w~x)
SSEs[21]=anova(results)$Sum[2]

lambda[21]=0
cbind(lambda,SSEs) # Choose lambda that minimizes the SSE
plot(lambda,SSEs)

# load package 'MASS'
library(help="MASS")
boxcox(y~x) # Check the resulting plot
```