

Chapter 10 in Everitt and Hothorn (2010)

Scatterplot Smoothers and Generalised Additive Models

Load the HSAUR2 and Rcmdr either
using the commands: `library(HSAUR2);library(Rcmdr)`
or from the R Console using menus **Packages > Load package ... > select HSAUR2 and Rcmdr > Ok**

| [LOWESS](#) | [Generalized Additive Model](#)

Subtopics:

[Olympic 1500m Times](#)

[Air Pollution in US Cities](#)

[Risk Factors for Kyphosis](#)

Olympic 1500m Times

From the R Commander menus select **Data > Data in packages > Read data set from an attached package... >**

Double click on **HSAUR2** and select **men1500m**, then click **OK**.

To see a description, from the R commander menu select **Data > Active data set > Help on active data set (if available)**

Click **View data set** to view it.

From the R Commander menu, select **Graphs > Scatterplot... > x:year, y: time**, and click **ok**.

76 Scatterplot

x-variable (pick one) y-variable (pick one)

time year time year

Options

Identify points ☐

Jitter x-variable ☐

Jitter y-variable ☐

Log x-axis ☐

Log y-axis ☐

Marginal boxplots ☒

Least-squares line ☒

Smooth line ☒

Show spread ☒

Span for smooth 50

Plotting Parameters

Plotting characters <auto>

Point size 1.0

Axis text size 1.0

Axis-labels text size 1.0

x-axis label y-axis label

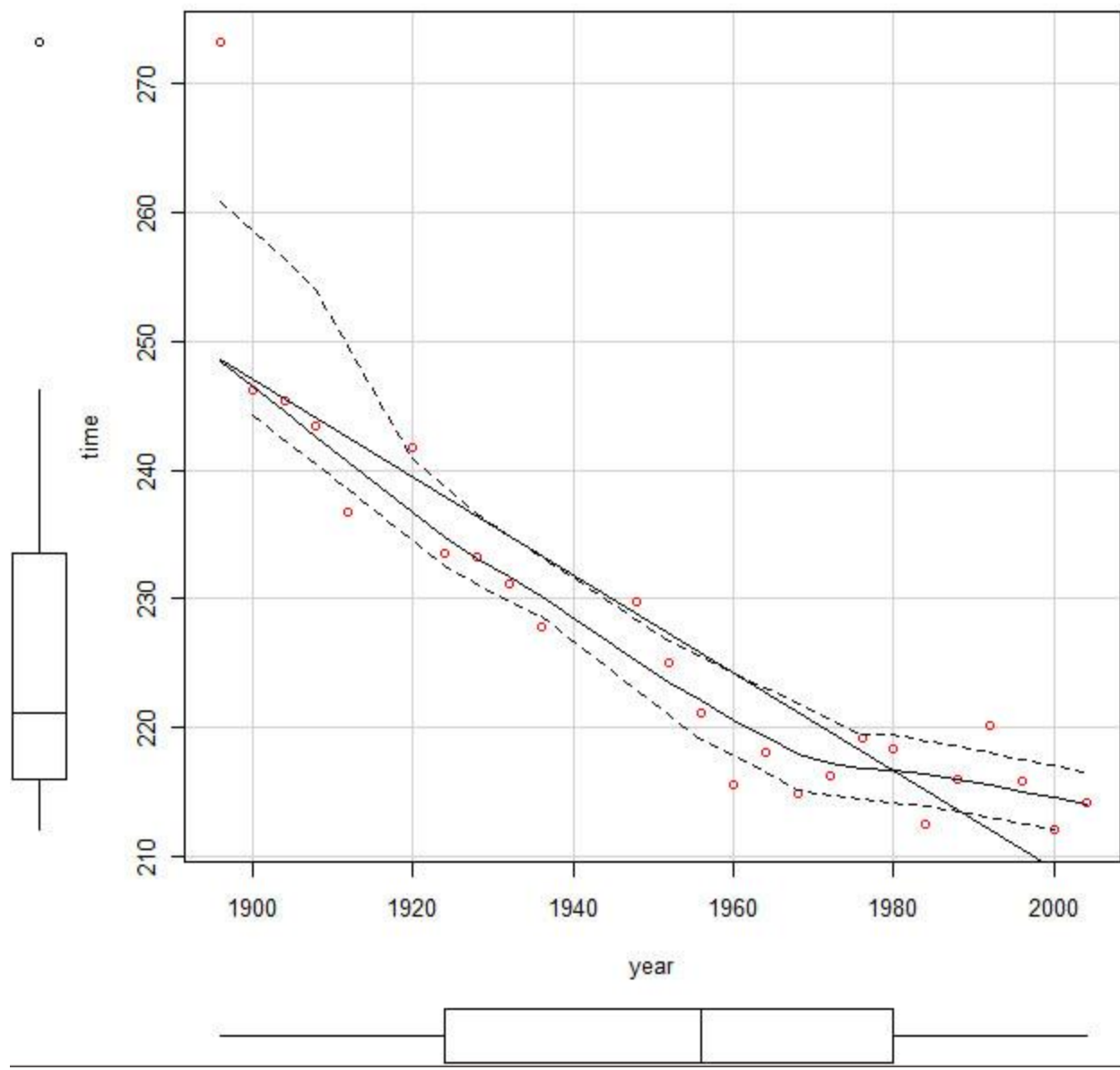
<auto> <auto>

Subset expression

<all valid cases>

Plot by groups...

OK Cancel Help

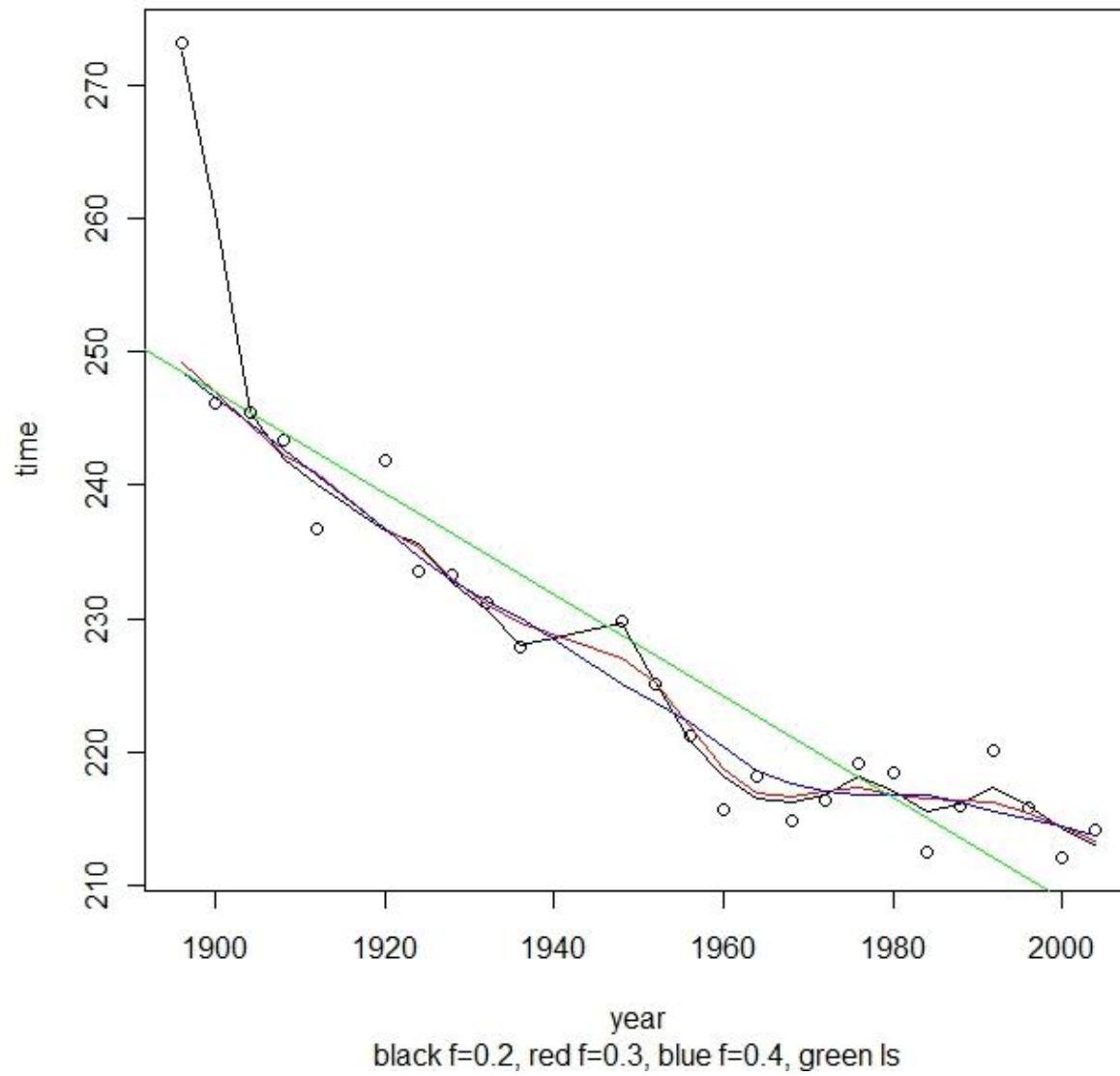


The straight solid line is the linear regression, while the solid curved line is the lowess smooth. The dotted lines show the confidence bound around the lowess smooth.

To see the effect of changing the proportion of points that influence the smooth (**f** in command or **Span for smooth** in menu) enter the following commands into the Script Window and submit them.

```
f1=lowess(men1500m$year,men1500m$time,f=0.2)
f2=lowess(men1500m$year,men1500m$time,f=0.3)
f3=lowess(men1500m$year,men1500m$time,f=0.4)
plot(men1500m$year,men1500m$time,xlab="year",ylab="time",
main="Men's 1000m Times",sub="black f=0.2, red f=0.3, blue f=0.4, green ls")
lines(f1$x,f1$y,type="l")
lines(f2$x,f2$y,type="l",col="red")
lines(f3$x,f3$y,type="l",col="blue")
abline(lm(men1500m$time~men1500m$year),col="green")
```

Men's 1000m Times



To use generalized additive models, from the R Commander menus select **Tools > Load packages... > mgcv** and click **Ok**.

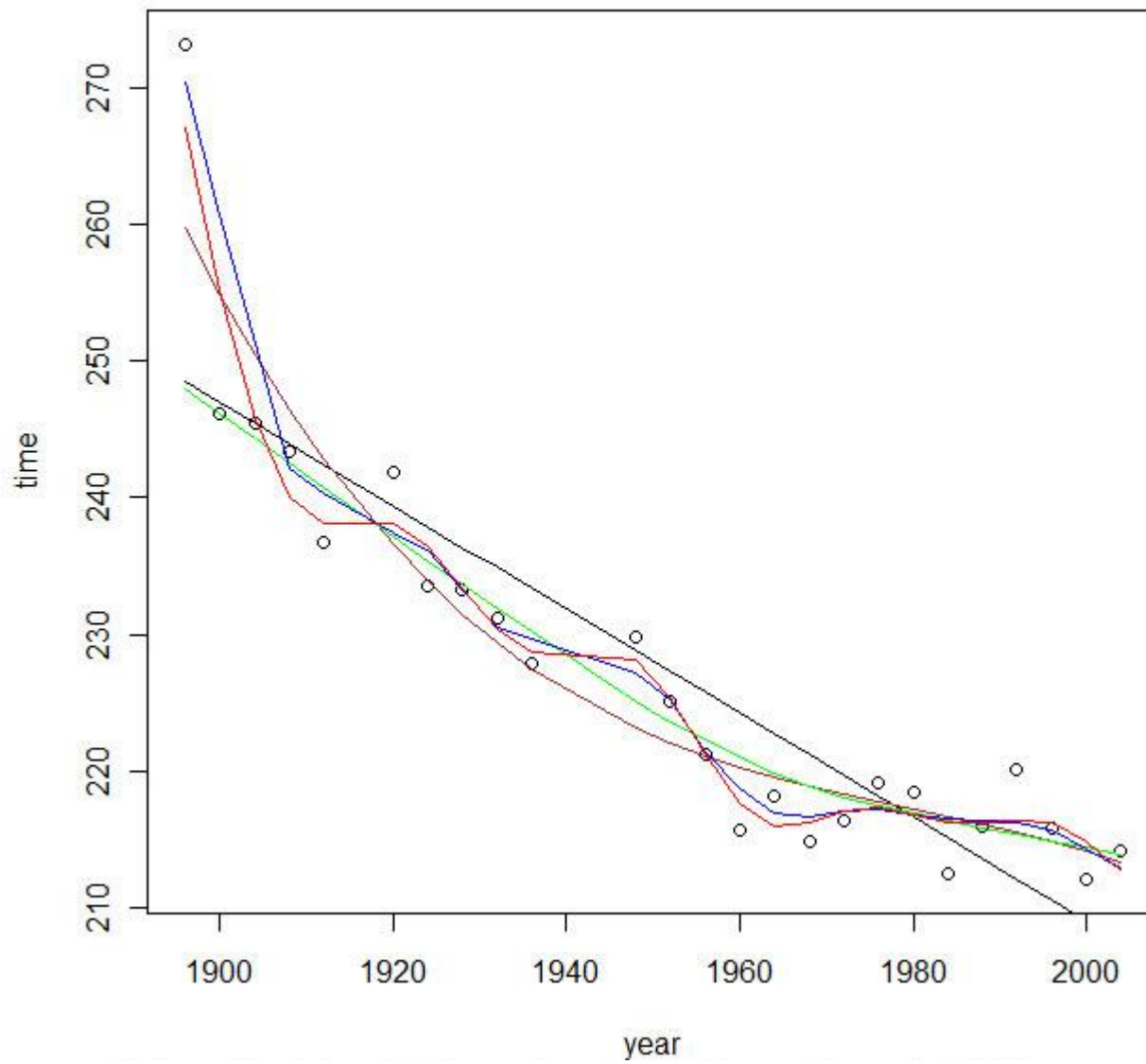
If you have not installed this package, you can do so from the R Console.

Please Note: When loading mboost, you may need to answer 1 in the R Console, or R will freeze waiting for an answer.

To compare fits, enter the following commands into the Script Window and submit them.

```
x=men1500m$year
f1=predict(lm(time~year,data=men1500m))
f2=predict(lm(time~poly(year,3),data=men1500m))
f3=lowess(men1500m$year,men1500m$time)
f3=f3$y
f4=lowess(men1500m$year,men1500m$time,f=0.25)
f4=f4$y
f5=predict(gam(time~s(year),data=men1500m))
plot(men1500m$year,men1500m$time,xlab="year",ylab="time",main="Men's 1000m Times",
sub="lm linear black, lm cubic brown, lowess default green, lowess f=0.25 blue, gam red")
lines(x,f1,type="l")
lines(x,f2,type="l",col="brown")
lines(x,f3,type="l",col="green")
lines(x,f4,type="l",col="blue")
lines(x,f5,type="l",col="red")
```

Men's 1000m Times



Air Pollution in US Cities

From the R Commander menus select **Data > Data in packages > Read data set from an attached package... >**

Double click on **HSAUR2** and select **USairpollution**, then click **OK**.

To see a description, from the R commander menu select **Data > Active data set > Help on active data set (if available)**

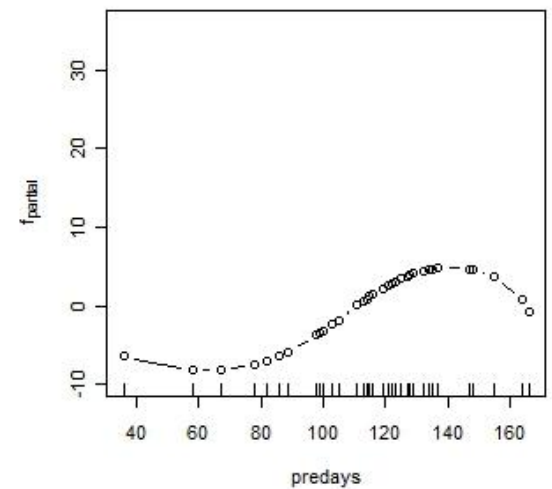
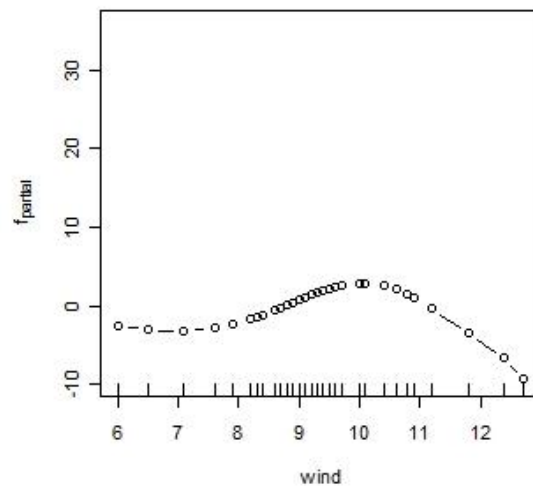
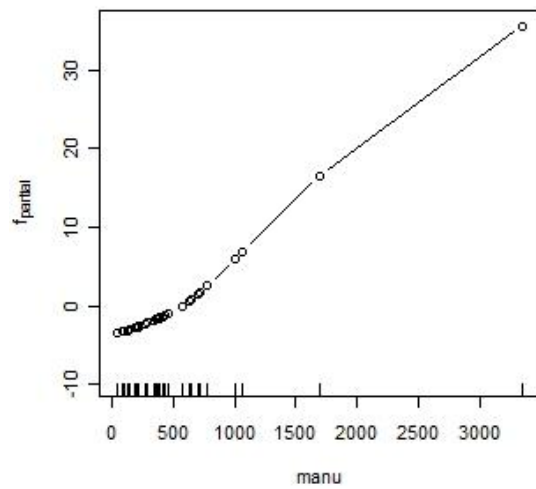
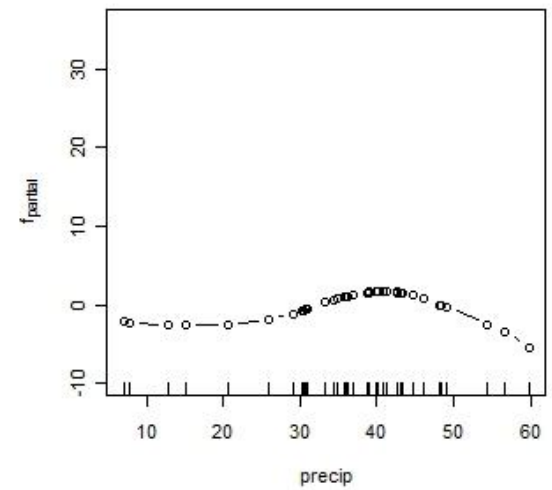
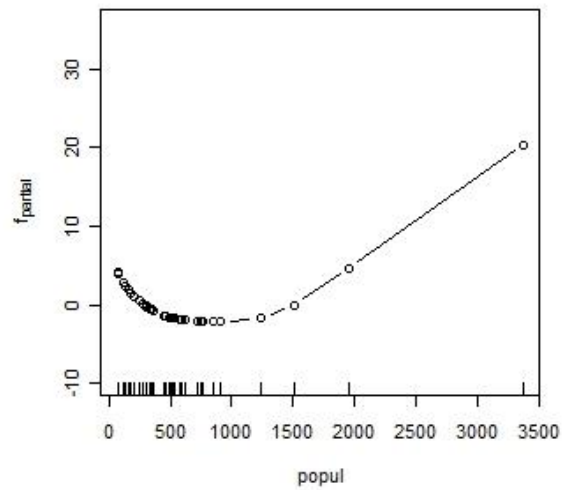
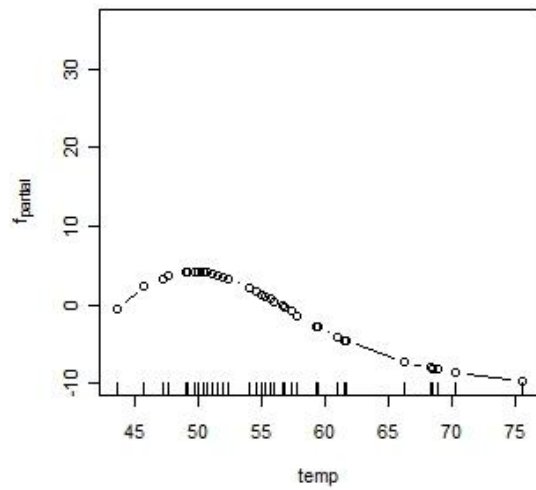
Click **View data set** to view it.

From the R Commander menus select **Tools > Load packages... >** select **mboost** and **mgcv** (if you did not load it above) and click **Ok**.

If you have not installed these packages, you can do so from the R Console.

Enter the following commands into the Script Window and submit them.

```
m1=gamboost(SO2 ~ ., data=USairpollution)  
aic1=AIC(m1); aic1  
p1=m1[mstop(aic1)] # = m1[40]  
layout(matrix(1:6, ncol=3))  
plot(p1, ask=FALSE)
```

Risk Factors for Kyphosis

If you have not already installed rpart from the R console, do so now.

From the R Commander menus select **Tools > Load packages... > rpart** and click **Ok**.

From the R Commander menus select **Data > Data in packages > Read data set from an attached package... >**

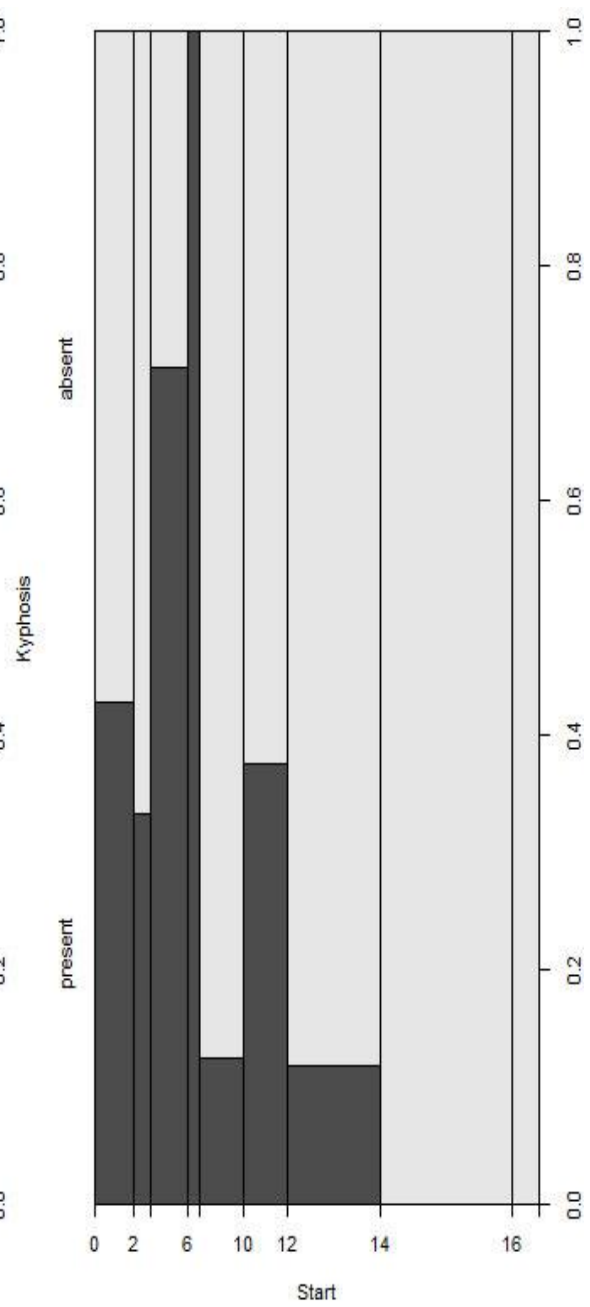
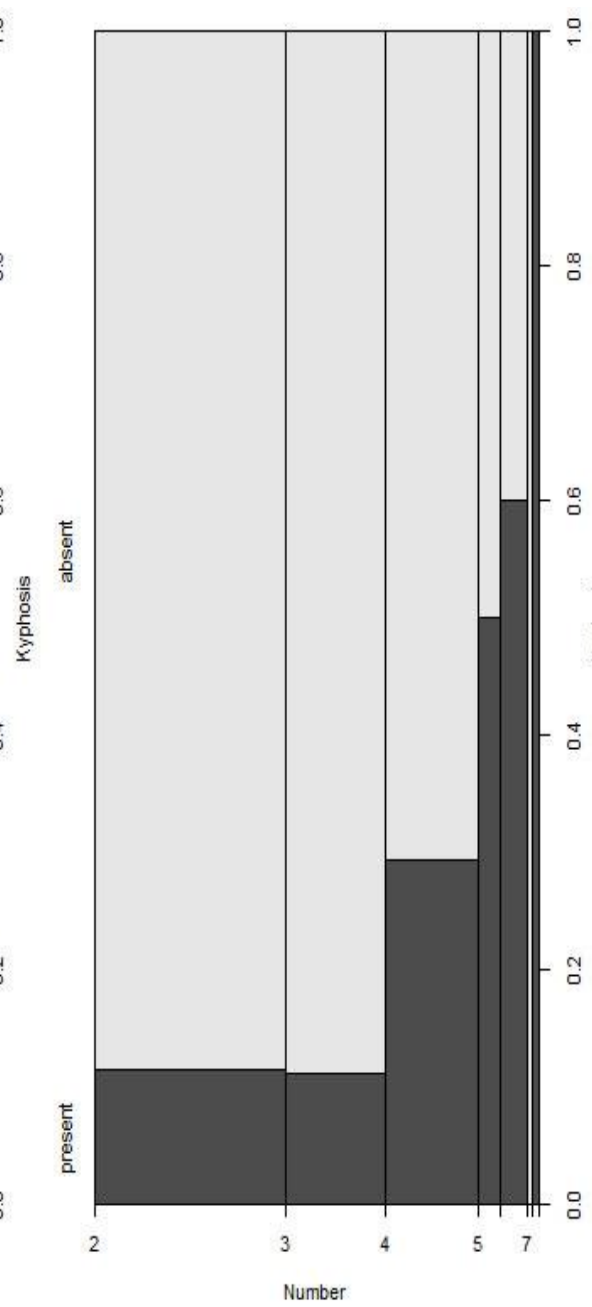
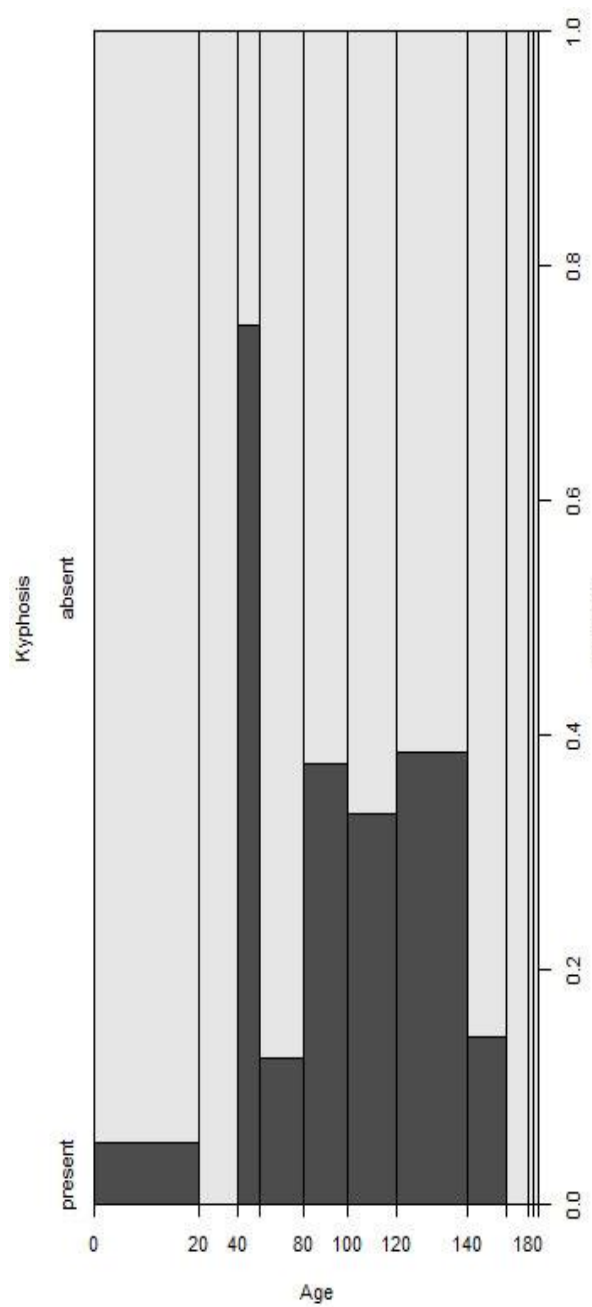
Double click on **rpart** and select **kyphosis**, then click **OK**.

To see a description, from the R commander menu select **Data > Active data set > Help on active data set (if available)**

Click **View data set** to view it.

Enter the following commands into the Script Window and submit them.

```
layout(matrix(1:3, nrow=1))  
spineplot(Kyphosis~Age, data=kyphosis, ylevels=c("present","absent"))  
spineplot(Kyphosis~Number, data=kyphosis, ylevels=c("present","absent"))  
spineplot(Kyphosis~Start, data=kyphosis, ylevels=c("present","absent"))
```



Copy and paste the following commands into the Script Window and submit them.
bs="cr" selects a cubic spline basis, k selects the dimension of the basis (degree of smoothing).

```
m1=gam(Kyphosis ~ s(Age,bs="cr")+ s(Number,bs="cr",k=3)+ s(Start,bs="cr",k=3),  
  family=binomial, data=kyphosis); m1  
trans=function(x) binomial())$linkinv(x)  
layout(matrix(1:3,nrow=1))  
plot(m1, select=1, shade=TRUE, trans=trans)  
plot(m1, select=2, shade=TRUE, trans=trans)  
plot(m1, select=3, shade=TRUE, trans=trans)
```

```
Family: binomial  
Link function: logit  
Formula:  
Kyphosis ~ s(Age, bs = "cr") + s(Number, bs = "cr", k = 3) +  
  s(Start, bs = "cr", k = 3)  
Estimated degrees of freedom:  
2.2267 1.2190 1.8421 total = 6.28774  
UBRE score: -0.2335850
```

