Splines en R

Carguemos los datos y la librería necesarias

```r
# install.packages("ISLR") #Instalar para cargar el dataset
library(ISLR)
attach(Wage)
library(splines)
```

Wage dataset

```r
head(Wage)
```

```
year age   maritl   race   education   region
231655 2006 18 1. Never Married 1. White   1. < HS Grad 2. Middle Atlantic
86582 2004 24 1. Never Married 1. White 4. College Grad 2. Middle Atlantic
161300 2003 45   2. Married 1. White 3. Some College 2. Middle Atlantic
155159 2003 43   2. Married 3. Asian 4. College Grad 2. Middle Atlantic
11443 2005 50   4. Divorced 1. White   2. HS Grad 2. Middle Atlantic
376662 2008 54   2. Married 1. White 4. College Grad 2. Middle Atlantic
jobclass   health health_ins logwage   wage
231655 1. Industrial   1. <=Good   2. No 4.318063 75.04315
86582 2. Information 2. >=Very Good   2. No 4.255273 70.47602
161300 1. Industrial   1. <=Good   1. Yes 4.875061 130.98218
155159 2. Information 2. >=Very Good   1. Yes 5.041393 154.68529
11443 2. Information   1. <=Good   1. Yes 4.318063 75.04315
376662 2. Information 2. >=Very Good   1. Yes 4.845098 127.11574
```

Creando conjunto de prueba

```r
agelims = range ( age )
age.grid = seq ( from = agelims [1] , to = agelims [2])
```

Spline cúbico (B-spline)

```r
fit = lm ( wage~bs ( age , knots = c (25 ,40 ,60) ) , data = Wage )
pred = predict ( fit , newdata = list ( age = age.grid ) , se = T )
plot ( age , wage , col =" gray ")
lines ( age.grid , pred$fit , lwd =2)
lines ( age.grid , pred$fit +2* pred$se , lty ="dashed")
lines ( age.grid , pred$fit -2* pred$se , lty ="dashed")
```
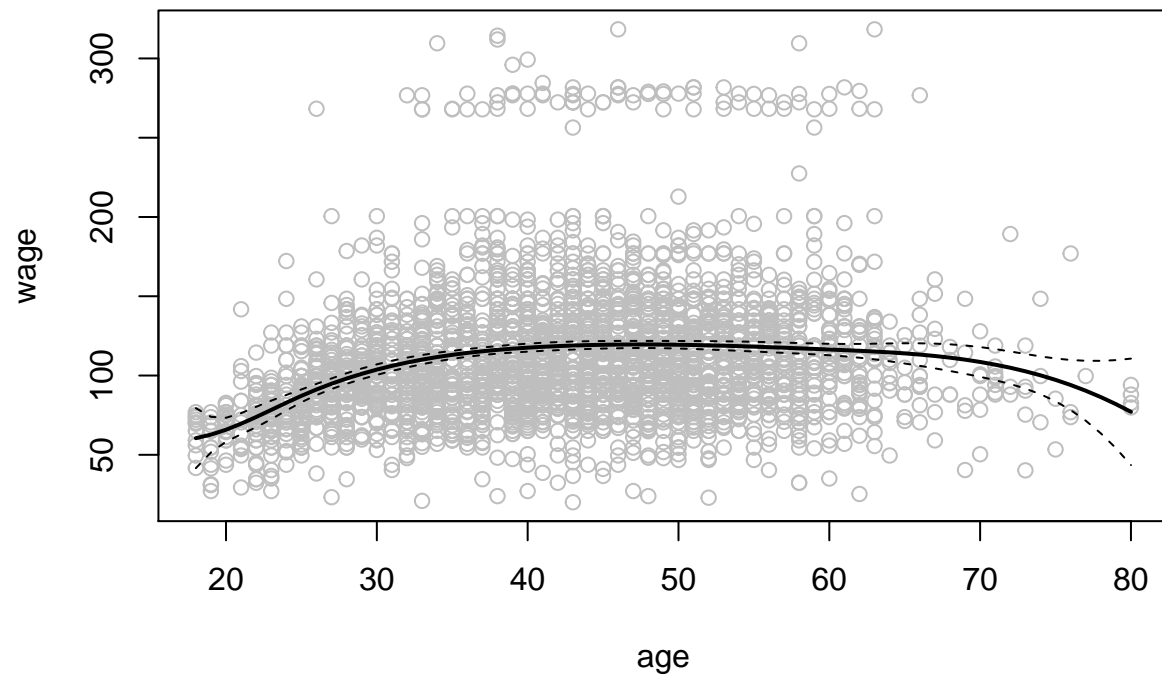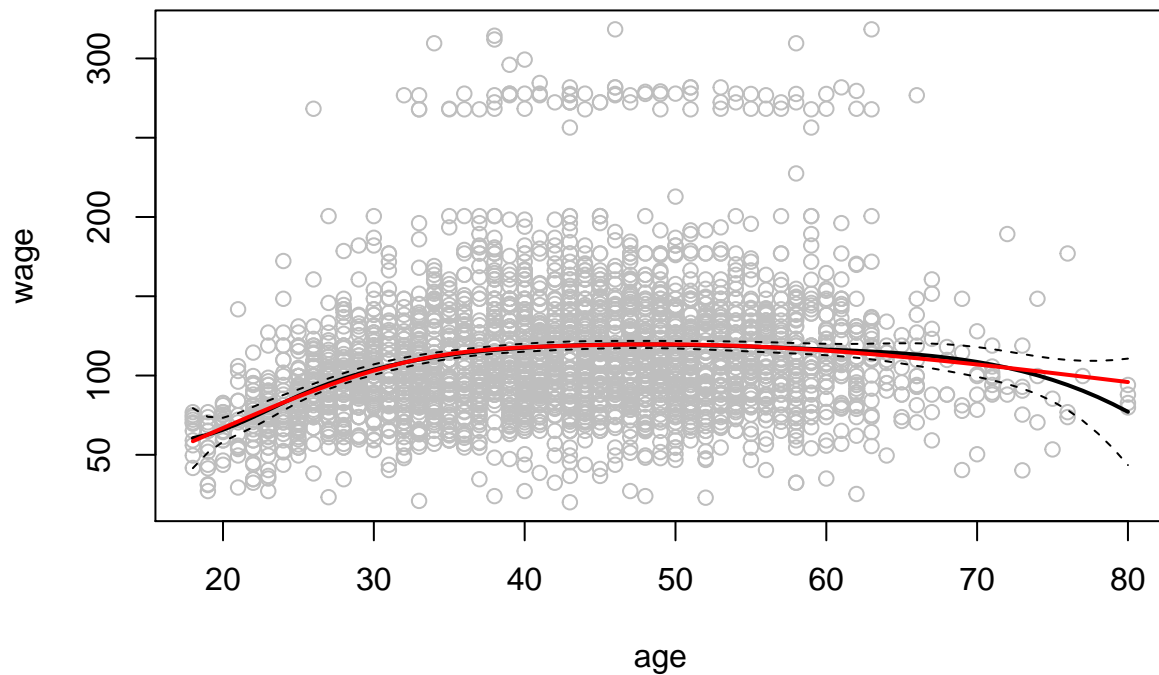
Spline Natural

```r
fit2 = lm ( wage~ns ( age , df =4) , data = Wage )

pred2 = predict ( fit2 , newdata = list ( age = age.grid ) , se = T )

# Grafica pasada
plot ( age , wage , col =" gray ")
lines ( age.grid , pred$fit , lwd =2)
lines ( age.grid , pred$fit +2* pred$se , lty ="dashed")
lines ( age.grid , pred$fit -2* pred$se , lty ="dashed")

# Grafica del spline natural
lines ( age.grid , pred2$fit , col = "red" , lwd =2)
```

Spline suave (Smooth Spline)

```r
fit = smooth.spline ( age , wage , df =16)
fit2 = smooth.spline ( age , wage , cv = TRUE )
```

```
## Warning in smooth.spline(age, wage, cv = TRUE): cross-validation with non-unique
## 'x' values seems doubtful
```

```r
fit2$df
```

```
## [1] 6.794596
```

```r
plot( age , wage , xlim = agelims , cex =.5 , col =" darkgrey ")
title(" Smoothing Spline ")
lines( fit , col =" red " , lwd =2)
lines ( fit2 , col =" blue " , lwd =2)
legend ("topright" , legend = c("16 DF" ,"6.8 DF") , col = c("red","blue") , lty =1 , lwd =2 , cex =.8)
```

Smoothing Spline