1. Consider the data set growth in fda library related to growth profiles of sample of boys and girls (note that this is a extended data set from the previously considered):

```
> library(fda)
> data(growth)
> Y1<-growth$hgtf
> Y2<-growth$hgtm
> t<-growth$age
>
> par(mfrow=c(1,2))
> matplot(t, Y1, pch=19, col="red", main="Height - Girls")
> matlines(t, Y1, col="red")
> matplot(t, Y2, pch=19, col="blue", main="Height - Boys")
> matlines(t, Y1, col="blue")
>
> head(Y1[,1:5])
     girl01 girl02 girl03 girl04 girl05
1      76.2   74.6   78.2   77.7     76
1.25   80.4   78.0   81.8   80.5     80
1.5    83.3   82.0   85.4   83.3     83
1.75   85.7   86.9   87.9   87.0     86
2      87.7   90.0   89.6   90.3     89
3      96.0   94.9   97.1   98.6     96
> head(Y2[,1:5])
     boy01 boy02 boy03 boy04 boy05
1     81.3  76.2  76.8  74.1  74.2
1.25  84.2  80.4  79.8  78.4  76.3
1.5   86.4  83.2  82.6  82.6  78.3
1.75  88.9  85.4  84.7  85.4  80.3
2     91.4  87.6  86.7  88.1  82.2
3    101.1  97.0  94.2  98.6  89.4


A list containing the heights of 39 boys and 54 girls from age 1 to 18 and the ages at
which they were collected.

hgtm - a 31 by 39 numeric matrix giving the heights in centimeters of 39 boys at 31 ages.
hgtf - a 31 by 54 numeric matrix giving the heights in centimeters of 54 girls at 31 ages.
age - a numeric vector of length 31 giving the ages at which the heights were measured.

Tuddenham, R. D., and Snyder, M. M. (1954)
"Physical growth of California boys and girls from birth to age 18",
University of California Publications in Child Development, 1, 183-364.
```

(a) Convert the $\mathbf{Y}_1$ data set related growth curves of girls to functional objects $y_{i1}(t) = \boldsymbol{\phi}'\mathbf{c}_{i1}$ by using $K = 15$ B-splines of order 6 and knots with equal intervals, and plot the 54 smoothed curves on one graph. Use the variable `age` as the t ="time" variable, and do the smoothing by using the least squares method to estimate coefficient vectors $\mathbf{c}_{i1}$.

(1 point)

(b) Convert the $\mathbf{Y}_1$ data set related growth curves of girls to functional objects $y_{i1}(t) = \boldsymbol{\phi}'\mathbf{c}_{i1}$ by using $K = 20$ B-splines of order 6 and knots with equal intervals. Do the smoothing, that is, estimation of coefficient vectors $\mathbf{c}_{i1}$, by using the penalized least squares method with roughness penalty being the square of the second derivative $[D^2 x_i(t)]^2$. You may use the value $\lambda = 0.01$ as the weight to the roughness penalty. Plot the 54 smoothed curves on one graph.

(1 point)

(c) Convert also the $\mathbf{Y}_2$ data set related growth curves of boys to functional objects $y_{i2}(t) = \boldsymbol{\phi}'\mathbf{c}_{i2}$ by using $K = 20$ B-splines of order 6 and knots with equal intervals. Do the smoothing, that is, estimation of coefficient vectors $\mathbf{c}_{i2}$, by using the penalized least squares method with roughness penalty being the square of the second derivative $[D^2 x_i(t)]^2$. You may use the value $\lambda = 0.01$ as the weight to the roughness penalty. After obtaining smoothed functionals $y_{i1}(t), y_{i2}(t)$, construct the sample mean functions

$$\bar{y}_1(t) = \frac{\sum_{i=1}^{N} \hat{y}_{i1}(t)}{N}, \qquad \bar{y}_2(t) = \frac{\sum_{i=1}^{N} \hat{y}_{i2}(t)}{N}$$

where N is the amount of sampling units in the data. Plot the mean curves $\bar{y}_1(t), \bar{y}_2(t)$ on one graph. What are the sample means at $t = 15$, i.e., what are $\bar{y}_1(15)$ and $\bar{y}_2(15)$?

(1 point)

(d) Evaluate how the first derivatives of $y_{i1}(t) = \boldsymbol{\phi}'\mathbf{c}_{i1}$ are behaving. Particularly, try to approximate time points t_* when the first derivative is zero for the sampling units $i = 1, 2, \ldots, 54.$, i.e., $\frac{d\hat{y}_{i1}(t)}{dt}|_{t=t_*} = 0$.

(1 point)

(e) Consider the first sampling unit $i = 1$ of the girls, and predict her height from the age of 18 to 19 year old. That is, predict the value of $y_1(t)$, when $t \in (18, 19]$.

(1 point)

(f) Let the be a new girl with observed measured values of height up to age of 4 being

```
   1  1.25   1.5  1.75     2     3     4
77.7  80.5  83.3  87.0  90.3  98.6 106.3
```

Predict her height from the age of 4 to 18 year old.

(1 point)

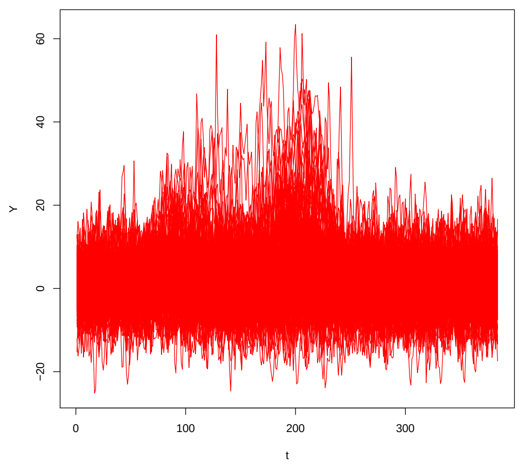
2. Consider the dataset emotion in library FDboost.

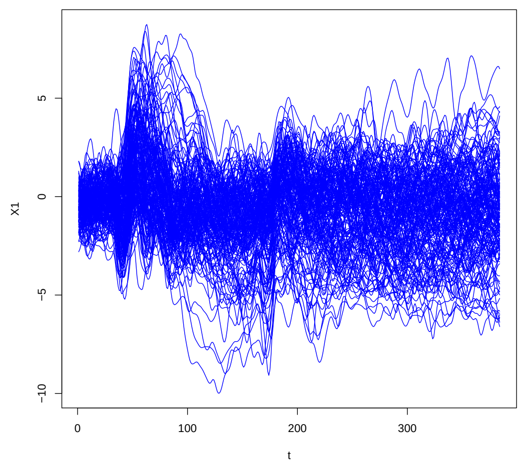
```
> library(fda)
> library(FDboost)
> data(emotion)
> Y<-t(emotion$EMG)
> X1<-t(emotion$EEG)
> t<-emotion$t
> matplot(t, Y, type="n", xlab="t", ylab="Y")
> matlines(t, Y, lty=1, col="red")
> matplot(t, X1, type="n", xlab="t", ylab="X1")
> matlines(t, X1, lty=1, col="blue")
> x2<-emotion$power
> x2
  [1] high high low  low  high high low  low  high high low  low  high high low  low  high high l
Levels: high low
> x3<-emotion$game_outcome
> x3
  [1] gain loss gain loss gain loss gain loss gain loss gain loss gain loss gain loss gain loss g
Levels: gain loss
> x4<-emotion$control
> x4
  [1] high high high high low  low  low  low  high high high high low  low  low  low  high high h
Levels: high low

To analyse the functional relationship between electroencephalography (EEG) and
facial electromyography (EMG), Gentsch et al. (2014) simultaneously recorded EEG
and EMG signals from 24 participants while they were playing a computerised gambling task.

y=EMG signal
x1=EEG signal
x2=power - factor variable with levels high and low
x3=game_outcome - factor variable with levels gain and loss
x4=control - factor variable with levels high and low

The aim is to explain potentials in the EMG signal by study settings as well as the EEG signal.
```





(a) Convert the X_1 and Y data sets to functional objects $x_{i1}(t) = \boldsymbol{\phi}'\mathbf{c}_{i_{X_1}}$ and $y_i(t) = \boldsymbol{\phi}'\mathbf{c}_{i_Y}$. Decide yourself the appropriate basis system $\boldsymbol{\phi}$, and estimate coefficient vectors $\mathbf{c}_{i_{X_1}}$ and $\mathbf{c}_{i_Y}$ either by the ordinary least squares method or by the penalized least squares method.

(1 point)

(b) Consider the linear model

$$y_i(t) = \beta_0(t) + \beta_1(t)x_{i1}(t) + u_i(t),$$

where $u_i(t)$ is the random error functional. Choose the appropriate basis system(s) $\boldsymbol{\phi}_\beta$ for functions $\beta_0(t)$ and $\beta_1(t)$. Plot the estimated functions $\hat{\beta}_0(t)$ and $\hat{\beta}_1(t)$.

(2 points)

(c) Consider the effect of variables x_2, x_3, x_4 to the functional objects $y_i(t)$. Thus consider the functional linear model

$$y_i(t) = \beta_0(t) + \sum_{j=2}^{4} \beta_j(t)x_{ij} + u_i(t),$$

where $y_i(t)$ are the smoothed curves, and the values x_{ij} are either 0 or 1, $j = 2, 3, 4$. Choose basis structures for functions $\beta_0(t), \beta_2(t), \beta_3(t), \beta_4(t)$. Calculate estimate for the functional $y_{i_*}(t)$, when the explanatory variables are set to the levels

$$x_{i_*2} = \mathsf{high}, \qquad x_{i_*3} = \mathsf{gain}, \qquad x_{i_*4} = \mathsf{high}.$$

Also calculate estimate for the functional $y_{i_\#}(t)$, when the explanatory variables are set to the levels

$$x_{i_\#2} = \mathsf{low}, \qquad x_{i_\#3} = \mathsf{loss}, \qquad x_{i_\#4} = \mathsf{low}.$$

Do the estimates of the functionals $y_{i_*}(t)$ and $y_{i_\#}(t)$ differ clearly? Draw the estimates $\hat{y}_{i_*}(t)$ and $\hat{y}_{i_\#}(t)$ into the same graph.

(2 points)

(d) Consider the linear model

$$y_i(t) = \beta_0(t) + \beta_1(t)x_{i1}(t) + \sum_{j=2}^{4} \beta_j(t)x_{ij} + u_i(t), \qquad j = 2, 3, 4.$$

Plot the fitted functions $\hat{y}_i(t)$ in a graph.

(1 point)

3. (a) Let the functional $y_i(t) = \boldsymbol{\phi}(t)'\mathbf{c}_i$ has the periodic basis

$$\boldsymbol{\phi}(t) = \begin{pmatrix} 1 \\ \sin(\omega t) \\ \cos(\omega t) \end{pmatrix}.$$

What is then the derivative $\frac{d\,y_i(t)}{d\,t}$?

(2 points)

(b) Let $\hat{y}_1(t) = \boldsymbol{\phi}(t)'\hat{\mathbf{c}}_1, \hat{y}_2(t) = \boldsymbol{\phi}(t)'\hat{\mathbf{c}}_2, \ldots, \hat{y}_N(t) = \boldsymbol{\phi}(t)'\hat{\mathbf{c}}_N$ be smoothed functionals, where the smoothing is done by the least squares method $\hat{\mathbf{c}}_i = (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{y}_i$. The functional sample mean $\bar{y}(t)$ is defined as

$$\bar{y}(t) = \frac{\hat{y}_1(t) + \hat{y}_2(t) + \cdots + \hat{y}_N(t)}{N}.$$

Simplify the expression for the sample mean $\bar{y}(t)$ as far as you can. That is, what kind of transformation the sample mean $\bar{y}(t)$ is from the observed data $\mathbf{Y} = (\mathbf{y}_1 : \mathbf{y}_2 : \cdots : \mathbf{y}_N)$.

(2 points)

(c) Consider the observed values $\mathbf{y}_i = \boldsymbol{\Phi}\mathbf{c}_i + \boldsymbol{\varepsilon}_i$ of the functional $y_i(t) = \boldsymbol{\phi}(t)'\mathbf{c}_i$. Let us do the smoothing by the penalized least squares method

$$\min_{\mathbf{c}_i} (\mathbf{y}_i - \boldsymbol{\Phi}\mathbf{c}_i)'(\mathbf{y}_i - \boldsymbol{\Phi}\mathbf{c}_i) + \lambda \mathbf{c}_i'\mathbf{R}\mathbf{c}_i,$$

where $\mathbf{R} = \int \left(\frac{d\phi(t)}{dt}\right)\left(\frac{d\phi(t)}{dt}\right)' dt$. Show that the solution for the penalized least squares method smoothing is

$$\hat{\mathbf{c}}_i = (\boldsymbol{\Phi}'\boldsymbol{\Phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\Phi}'\mathbf{y}_i.$$

(2 points)