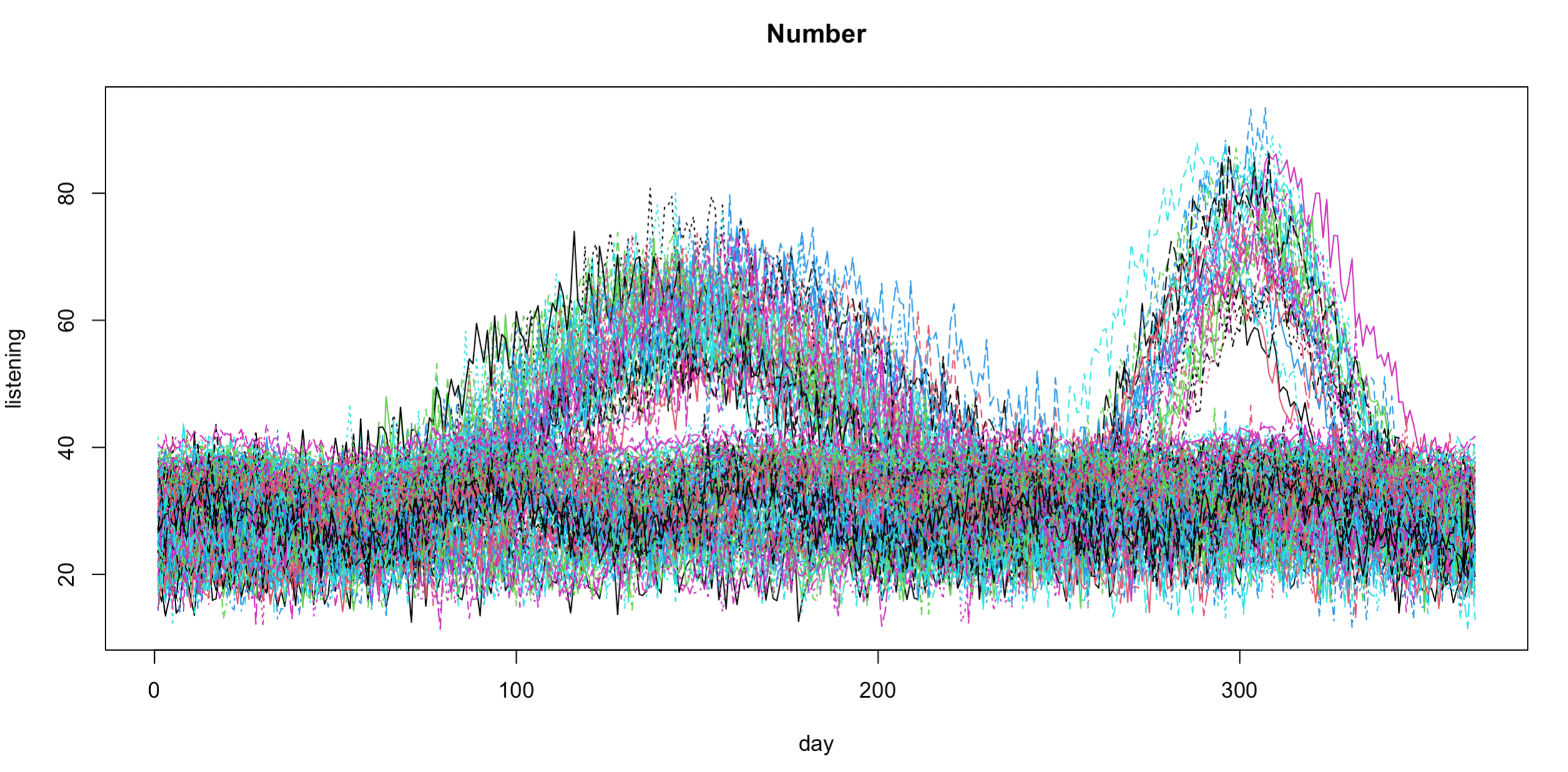
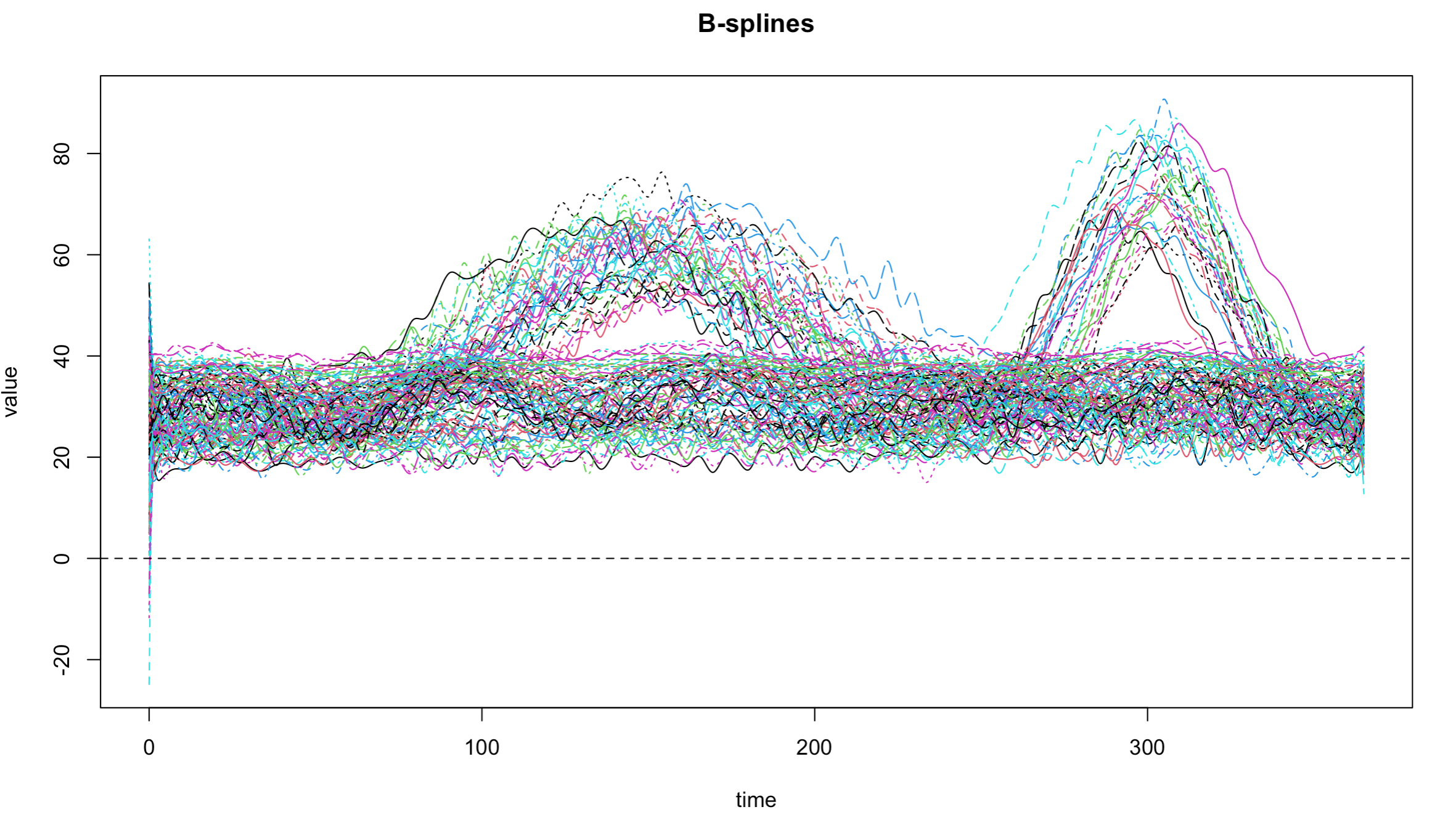
**Problem 4**

We have a dataset about number of listening by day, this are our functions.



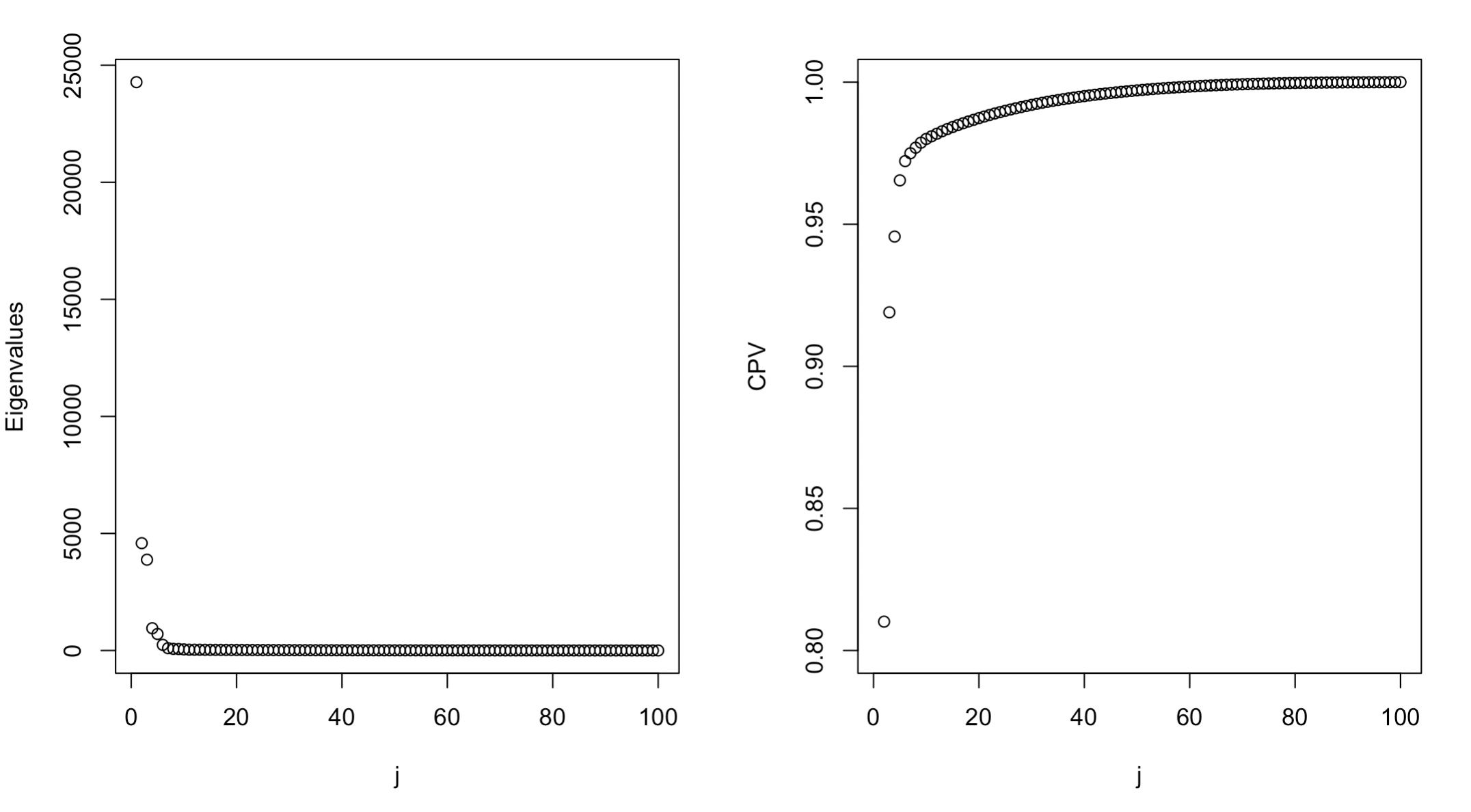
To smooth the data at first we create a set of [NUMBER] cubic splines basis. We project the original functions on them and obtain the following result:



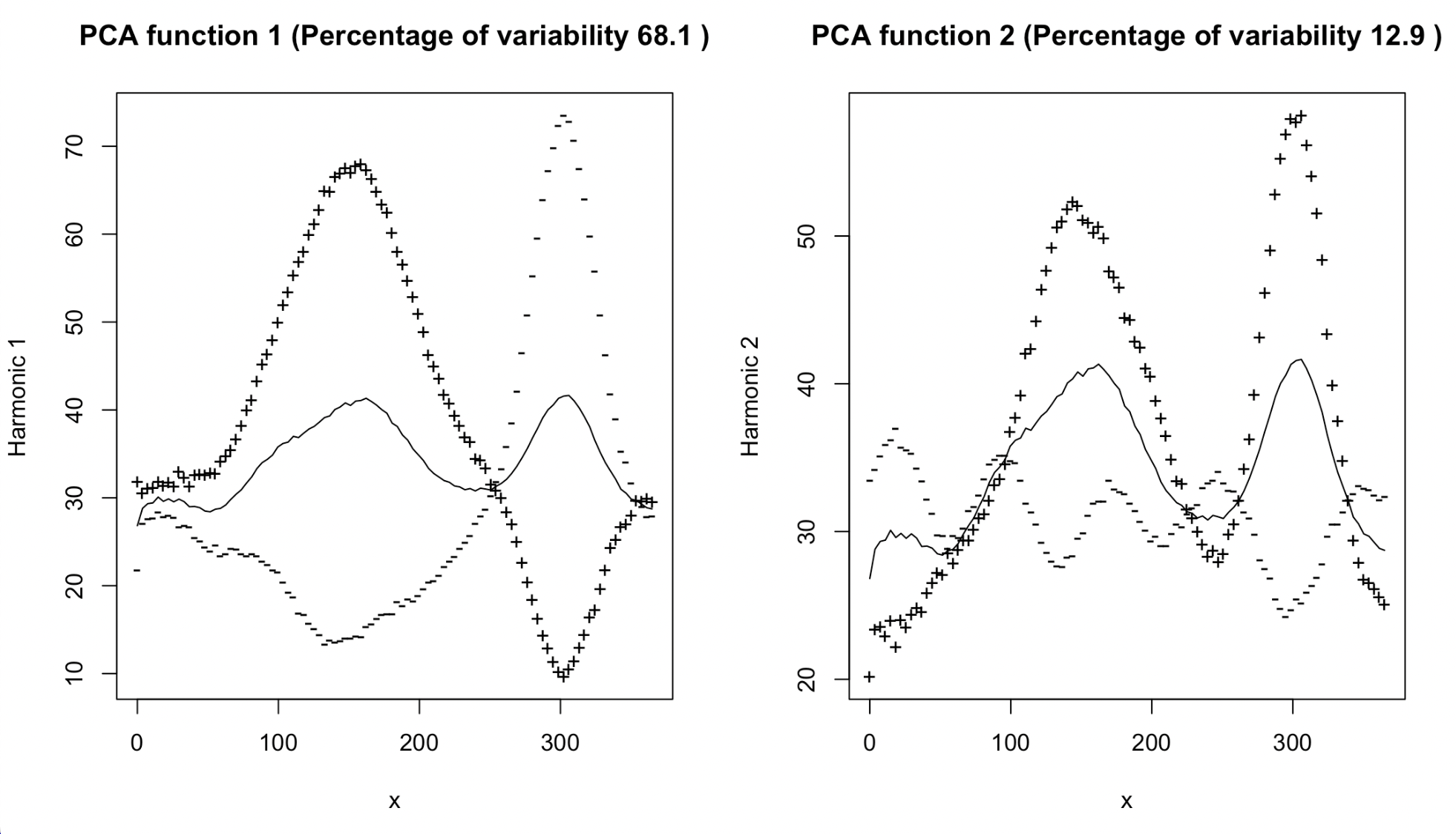
The first three coefficients of the first basis are: 24.76659 23.66195 21.46086

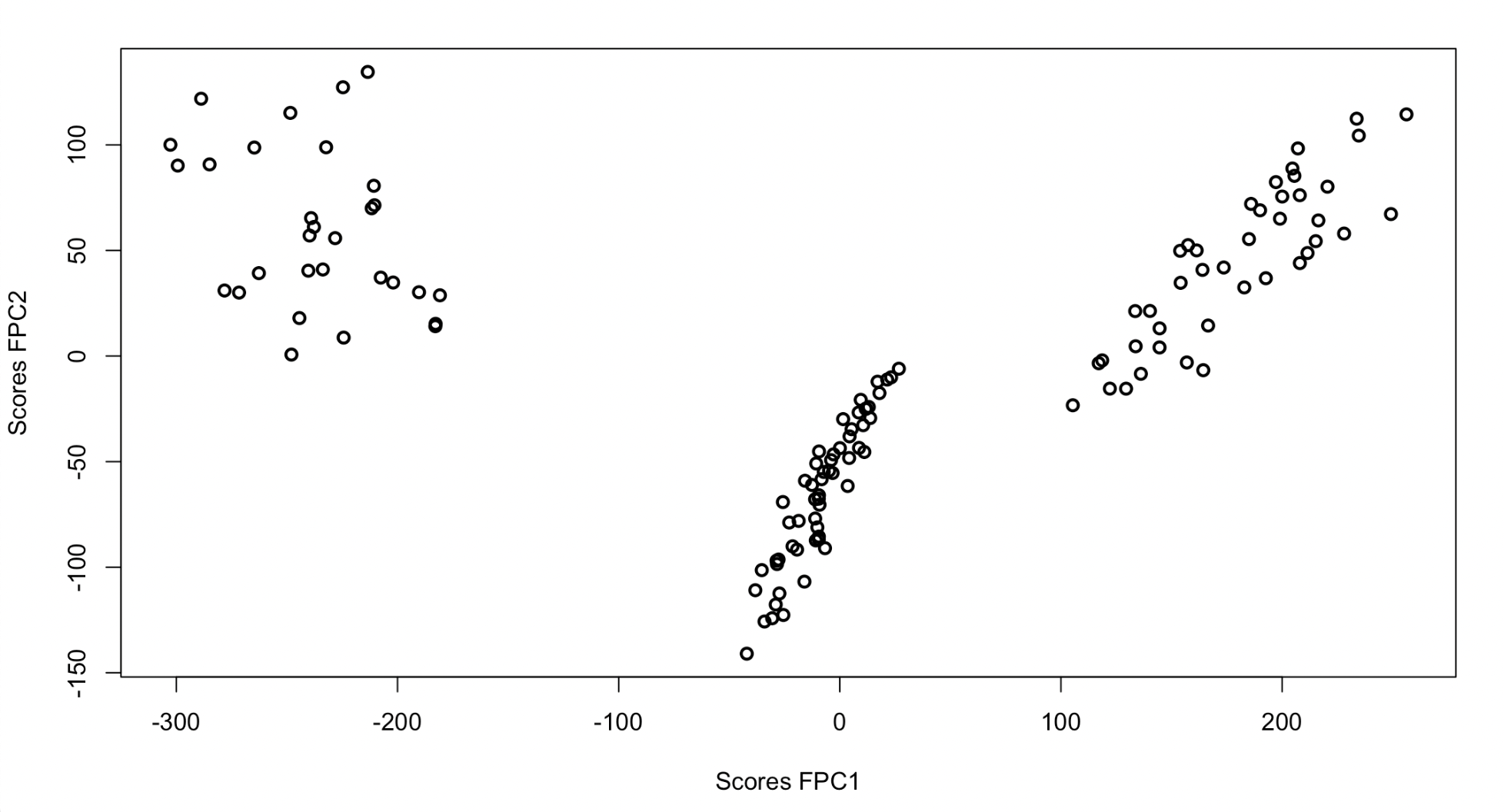
We can now perform Principal component analysis on the data to find the function around which the maximum variability is spread. The first 5 PCS explain respectively 0.68147817 0.12866794 0.10885970 0.02665162 0.01977973 so summed: 0.9654372% of the variance.

Here we can see the screeplot resulting:



We decide to keep just the first two PCS which explain more than 90% of the variance.





We can see groups in the plot.