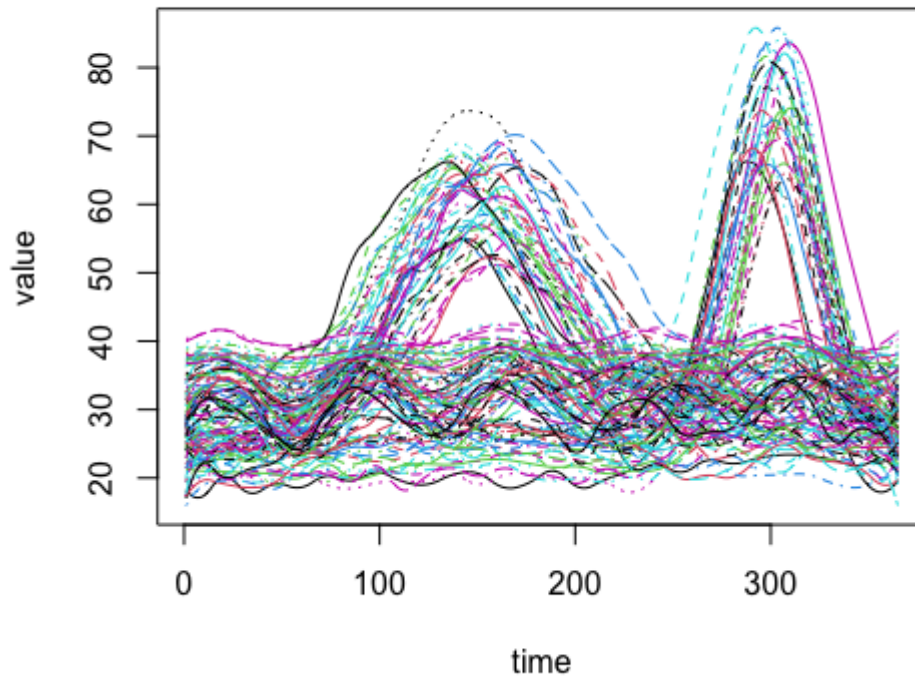


Exercise 4

The choice of breaks is evenly spaced.

I perform the smoothing with the required parameters and get the following result:



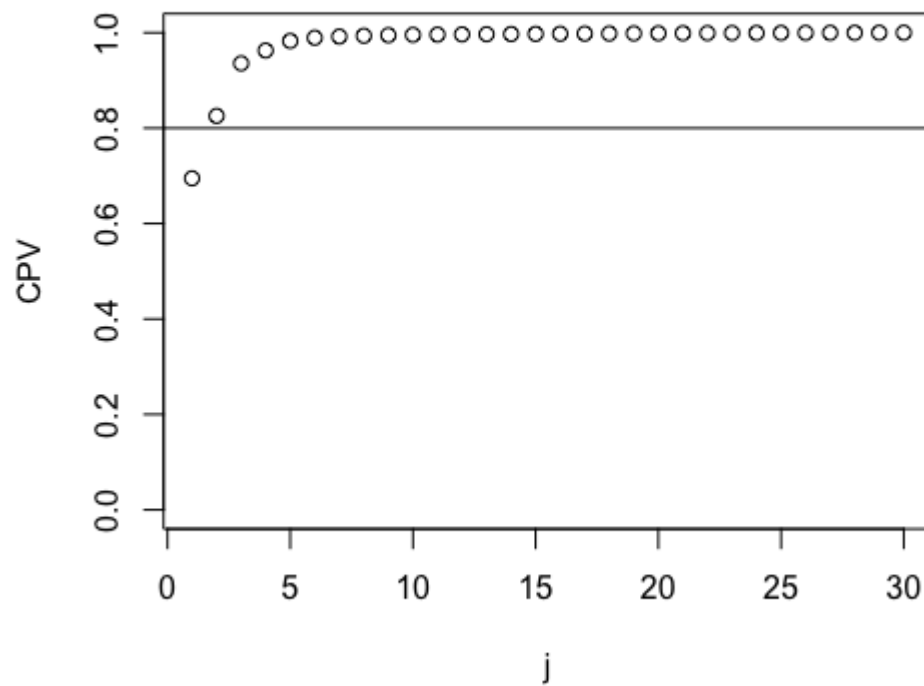
The first 3 coefficients for song 1 are

```
bsp13.1  bsp13.2  bsp13.3  
18.21189 15.85323 21.34703
```

Then I perform a Functional Principal Components Analysis (FPCA). The first 5 components explain respectively

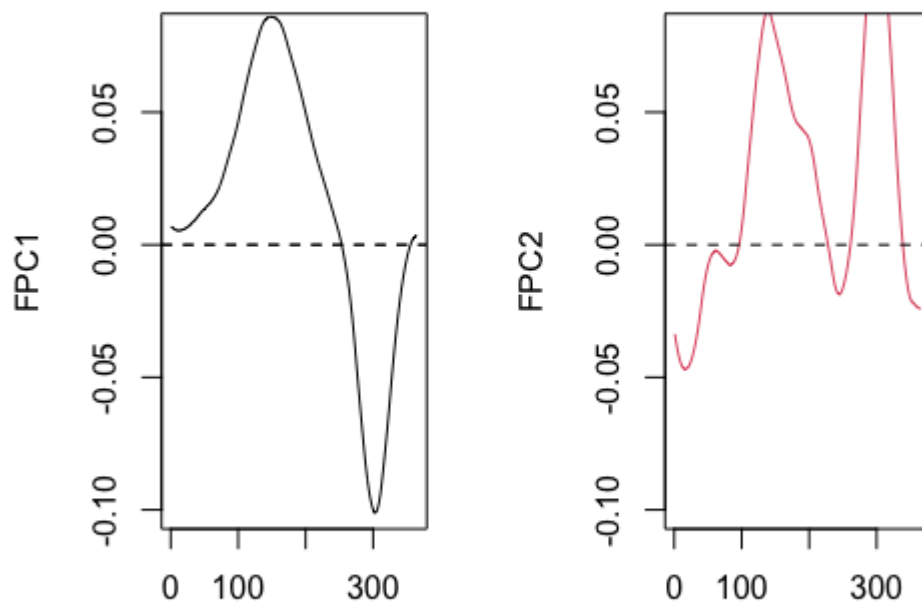
```
0.69479647 0.13081329 0.11012629 0.02697289 0.02000239
```

of the total variance, indeed here is the screeplot.



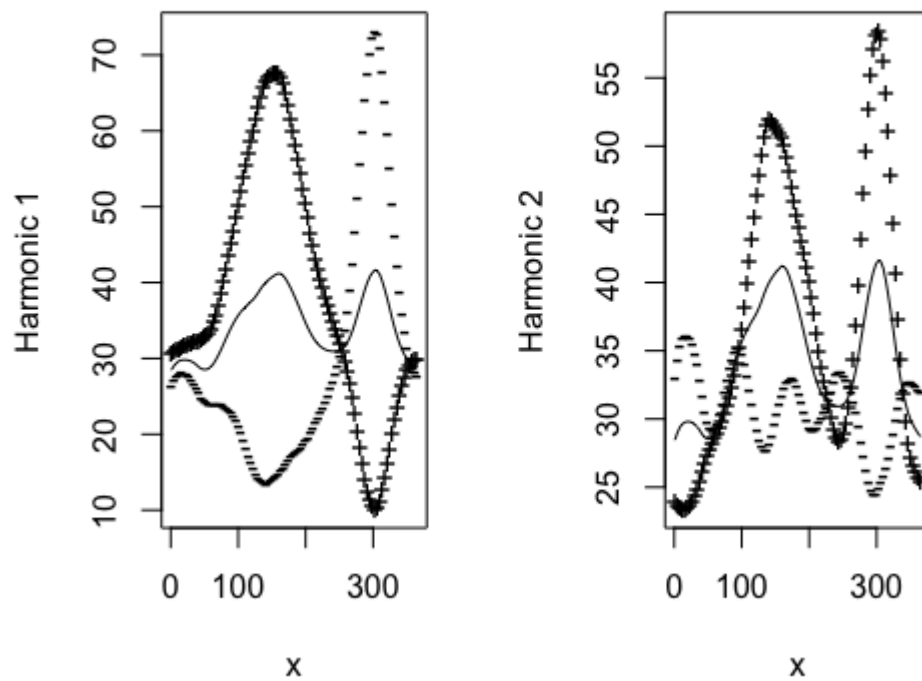
I would propose as possible dimensional reduction to project the original data on the PC space of the first 2 components (0.8256098 of variance explained).

The first 2 PCs are the following



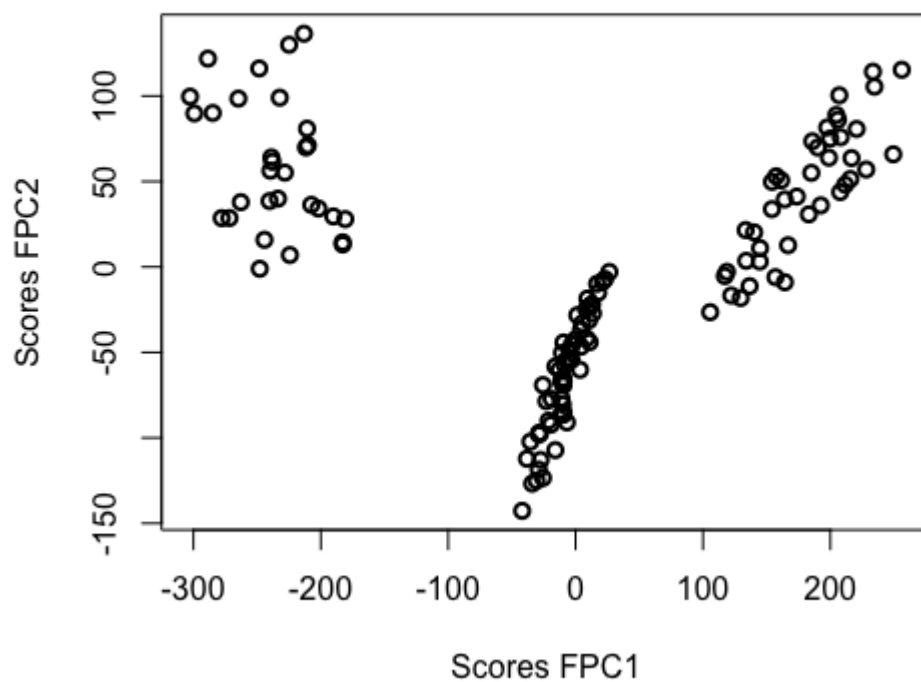
The retained PCs as perturbation of the mean are as follows

action 1 (Percentage of varin **action 2 (Percentage of vari**



An observation with a positive PC1 is above the mean in spring/summer and below in winter, this could be referred to known “summer hits”. The inverse hold for when PC1 is negative. Positive PC2 is associated to songs which are well liked both in spring/summer and winter, and negative PC2 is associated to songs which are oscillating below the mean all the year long.

The plot of the scores is



I can clearly see 3 clusters, mostly discriminated by the PC1. Cluster on the right has positive PC1 thus is most likely a cluster of summer hits, while cluster on the left is most likely the opposite. PC2 doesn't seem to discriminate very well.