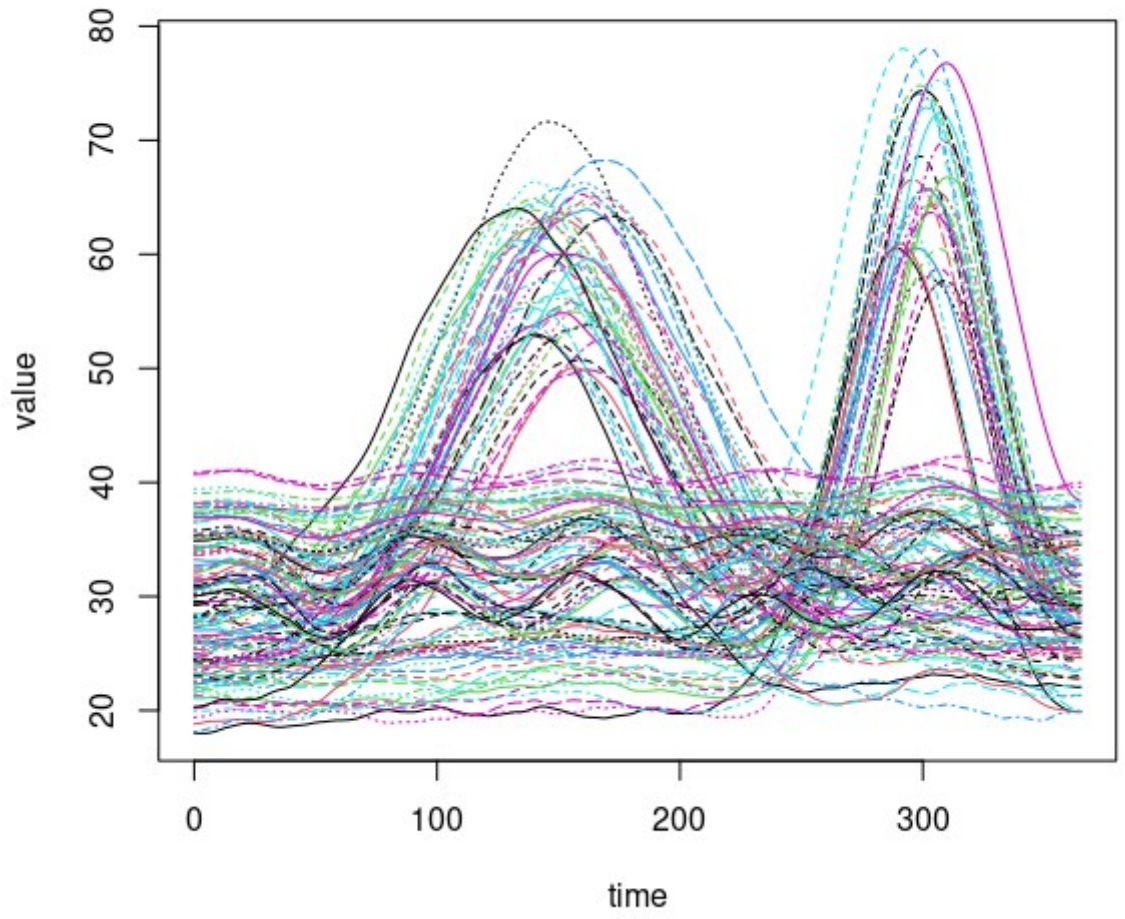
Ex4
a)

10/10

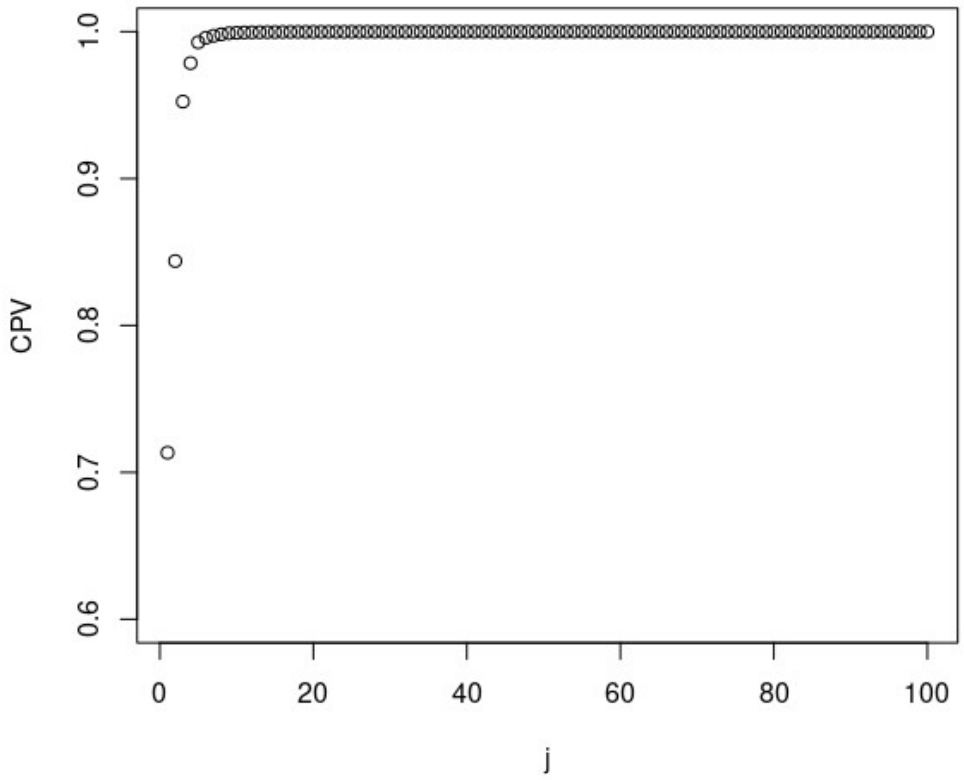


We choose 100 basis, the estimate for the first 3 parameters for the first song are:
18.03991,17.99415 and 18.00049

ok

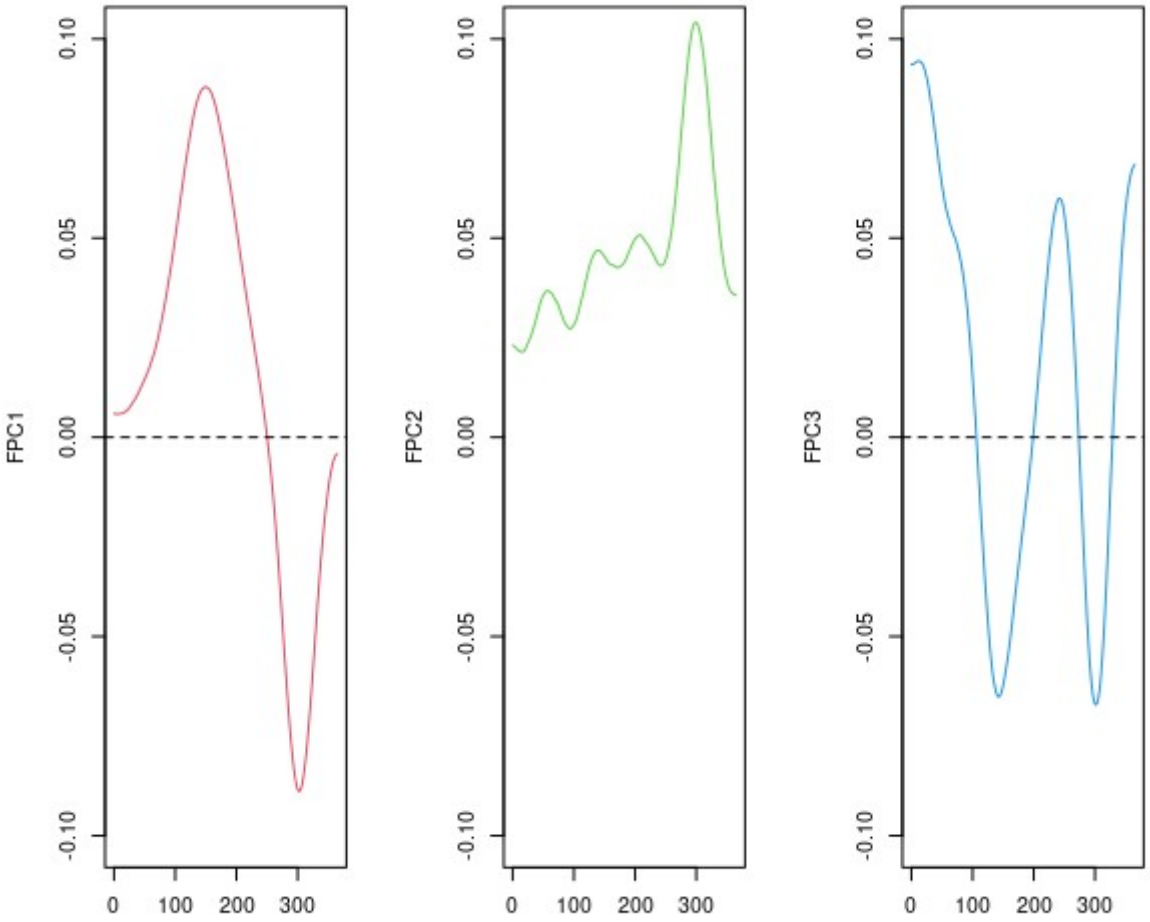
b)
Variances along first 5 components: 21990.1361 4022.7950 3346.0051 807.2823 437.7569

Screeplot

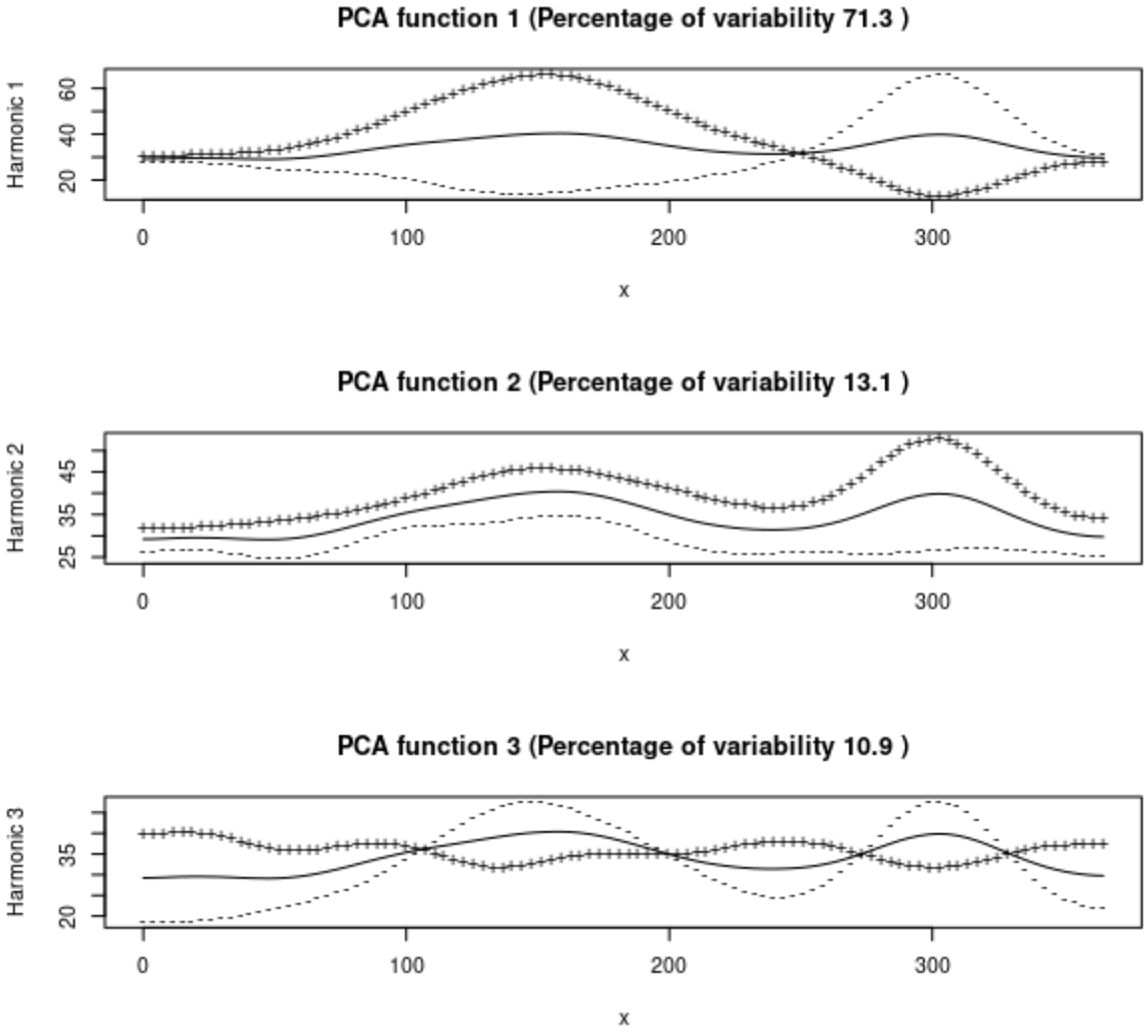


c)
Cumulative proportion of variance explained by first 3 PC: 0.7133807 0.8438840 0.9524315,
So we choose 3 principal components

Plot of the the retained principal components:



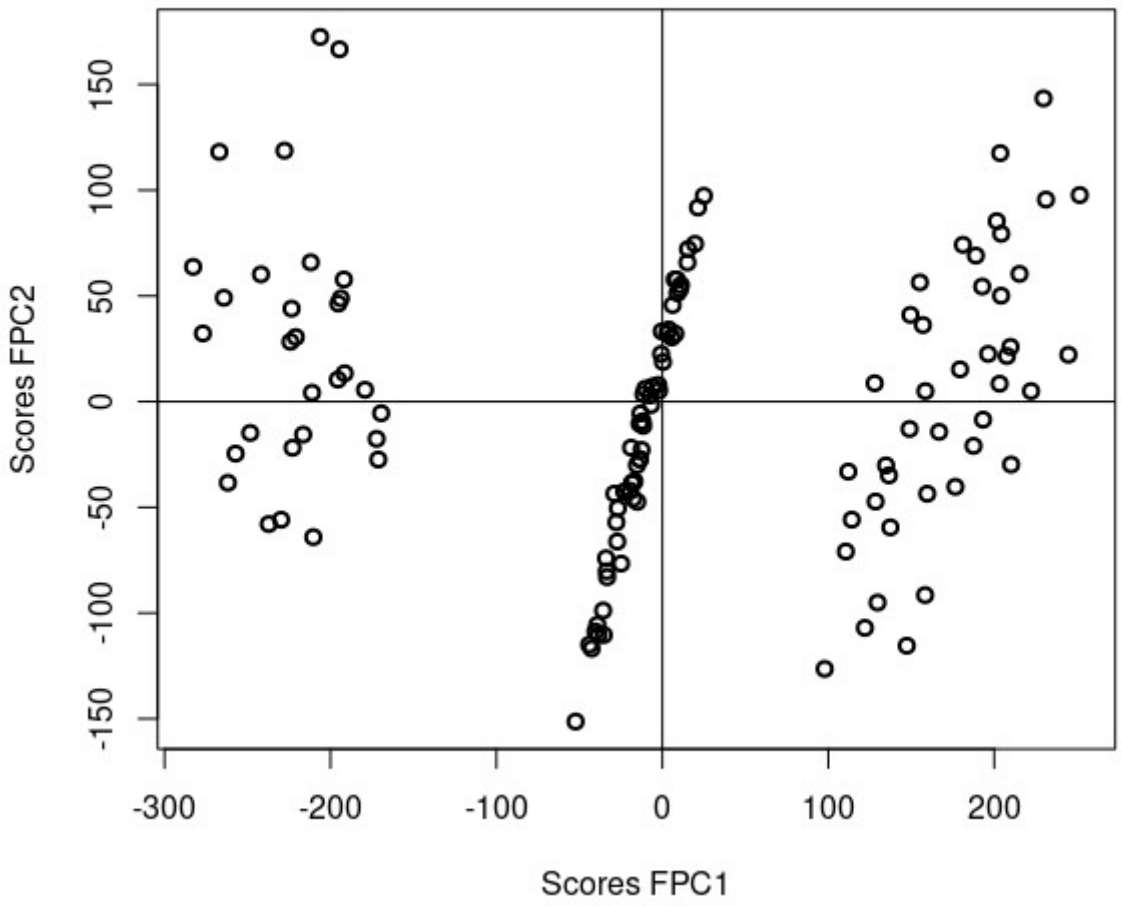
d)



Interpretation:

1. Songs with high PC1 are more listened during days 100-200 of the year, on the other hand, songs with low PC1 are more listened at the end of the year
2. Variation in amplitude that is more evident at the end of the year; songs with high PC2 are more listened, especially at the end of the year
3. Songs with high PC3 have less variation of the listening during the year, songs with low PC3 have instead an high variation.

e)



Interpretation:
We can see a clear clustering between the songs:

1. The first group of songs have low PC1: they could be Christmas song, since they are listened more during the end of the year
2. The second group has no particular levels of PC1 or PC2, they could be songs listened during all the year
3. The third group is like the first, but wit high PC1: they could be summer songs
4. Notice that along the PC2 we can't see a specific behavior: only that group1 is more listened at the end of the year, while group2 not (coherent with the previous interpretation)