16/7

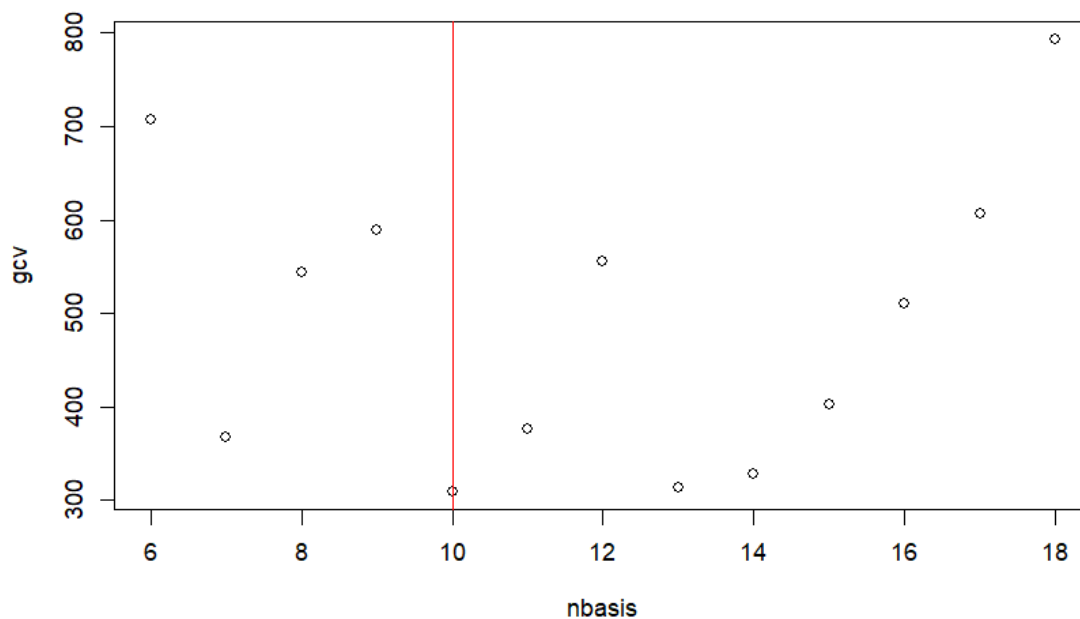## KMA per trovare clusters in all day (e poi in mean day)

Firstly I load fda and fdakma library and the data.

I choose the b-spline basis because the data don't present any periodicity.

Then I search for the optimal number of basis with generalized cross validation considering only the first observation and then using that basis for all the other 123 observation:

Finding 10 as optimal number of basis.

Then I smooth all the data:
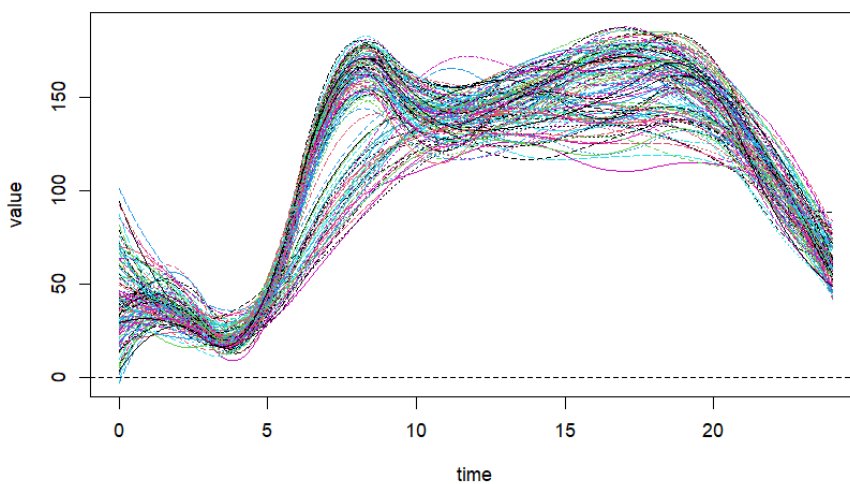
Now I proceeded doing KMA to see if there are some clusters.

Firstly I tried allowing for alignment, but then, after tried with NOAlignment, the result was really similar and moreover it appears to me (and Luca) that makes no sense aligning the functions because in this way we lose the variability for interpreting our data.

So here I report the result given by the kma function with two cluster because from the plot of the smoothed data it seems to be like that, NOalignment, k-mean as center method and similarity computed according to d0.pearson i.e. this similarity measure is the cosine of the angle between the two functions f and g (R help of kma.similarity → d0.pearson)
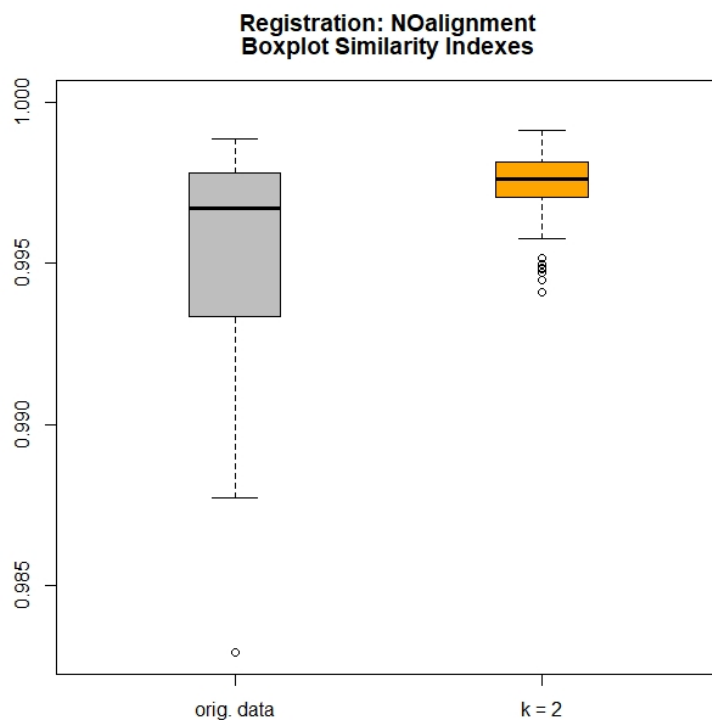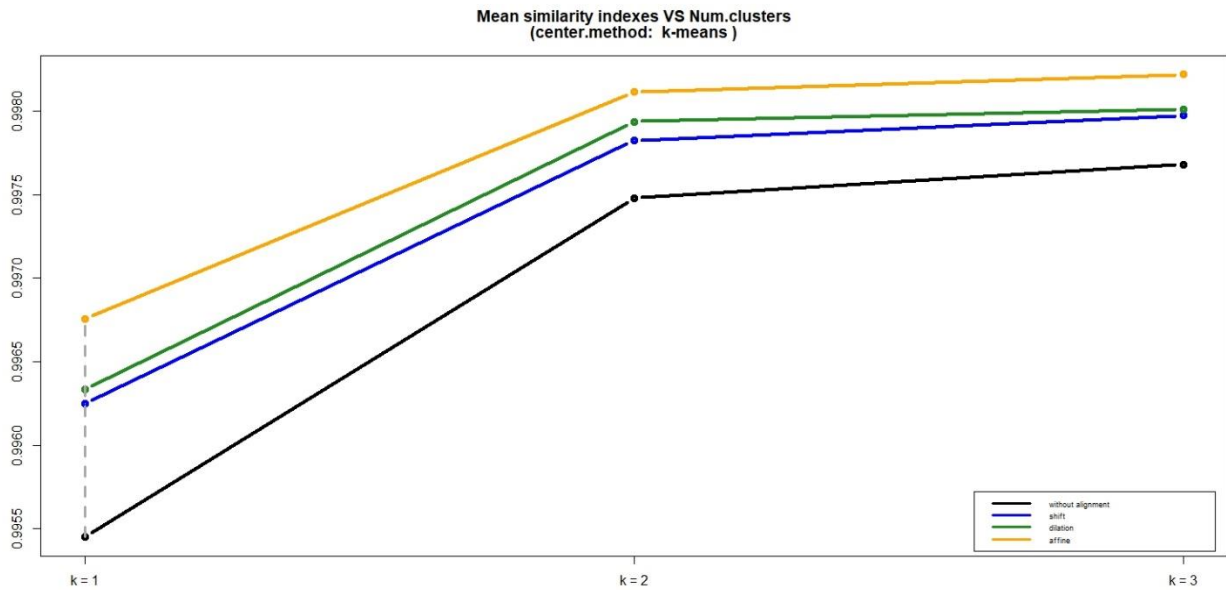

[does this makes sense to be consider a good result? Orig data has 0.996 and k = 2 0.997, so the result doesn't change]

Two clusters:
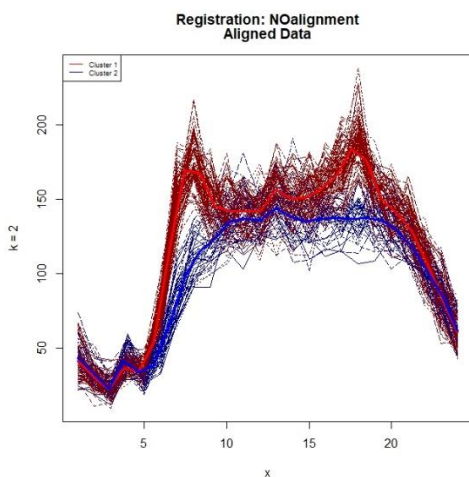
I do the following analysis to understand which number of cluster is optimal and which method I have to choose:

Mean similarity indexes VS Num.clusters
(center.method: k-means)

We choose k = 2 NOAlignment.

[does this makes sense to be consider a good result? Orig data has 0.996 and k = 2 0.997, so the result doesn't change (see scale of th axes)]