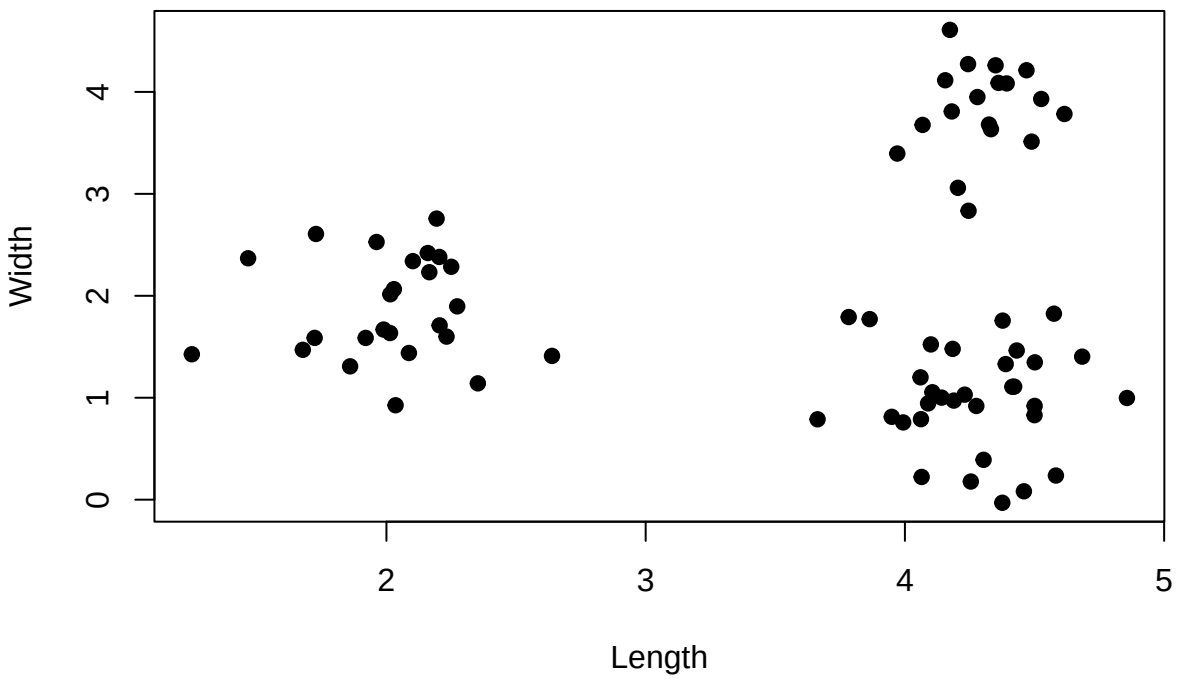
Exam: 2021/06/18

Marco Scarpelli

06 giugno, 2024

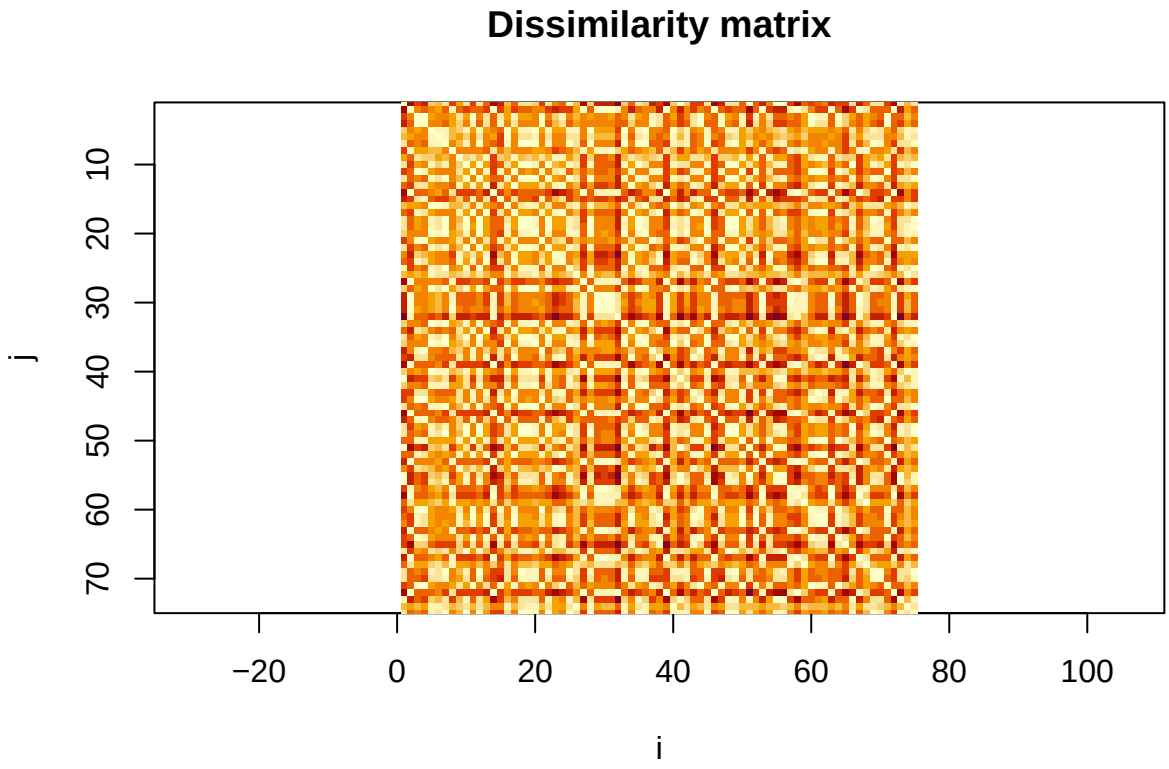
Dataset exploration

```
##        Length  Width
## Site_1 4.4590 0.0823
## Site_2 4.2797 3.9496
## Site_3 2.2500 2.2847
## Site_4 1.9620 2.5277
## Site_5 4.1843 1.4793
## Site_6 4.5748 1.8245
```

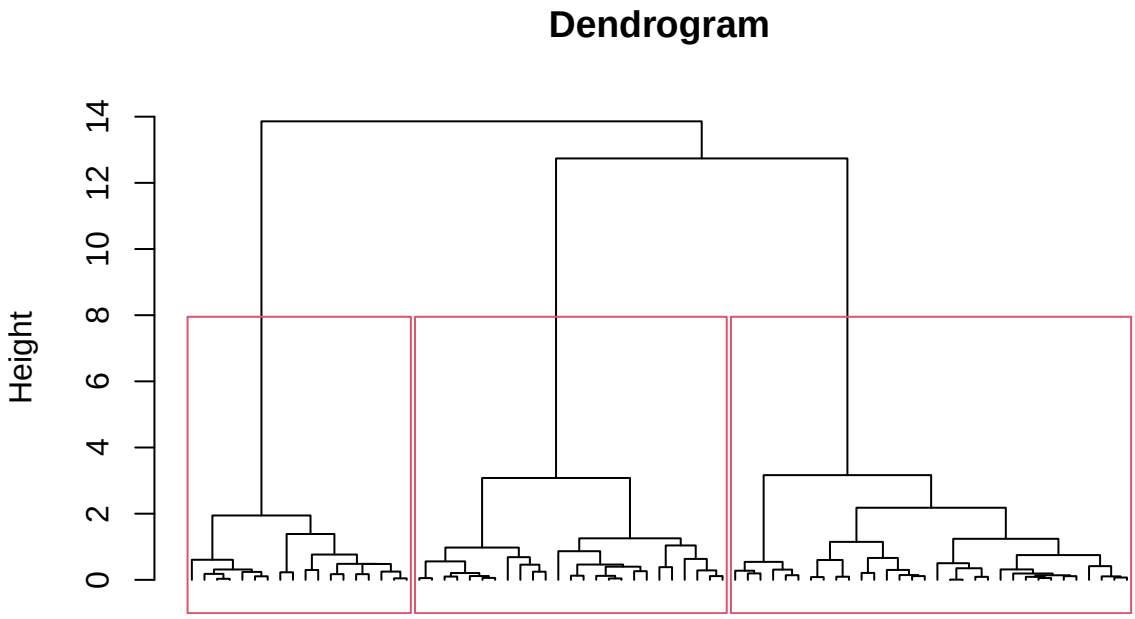


Point A

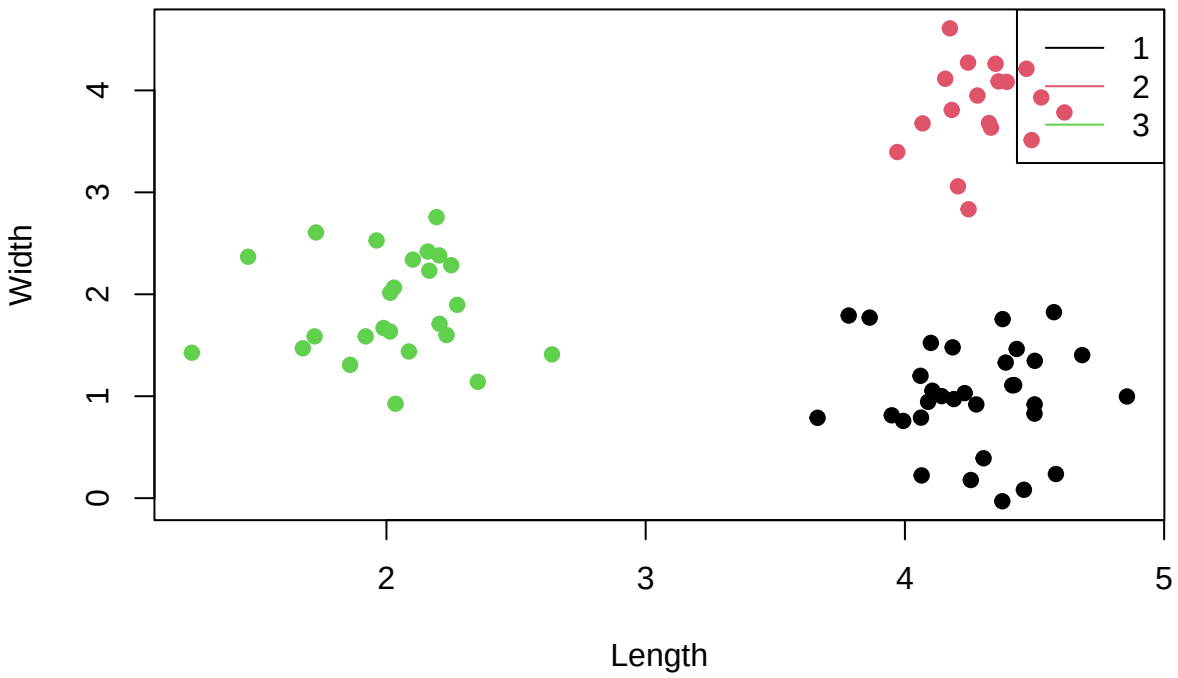
We report the dissimilarity matrix:



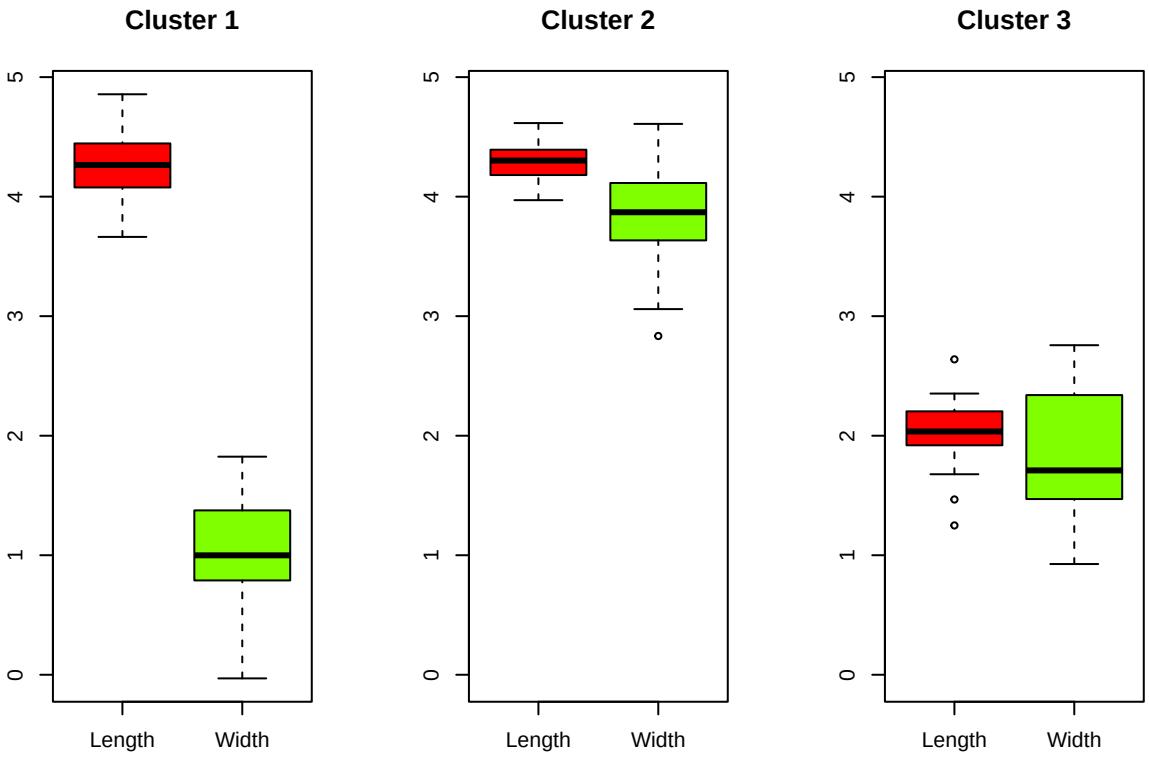
We run the clustering algorithm and provide the dendrogram. From it, it seems that the correct amount of clusters is 3, since there is a great height since the two-to-three split and the next one. We hence highlight the cut at three clusters:



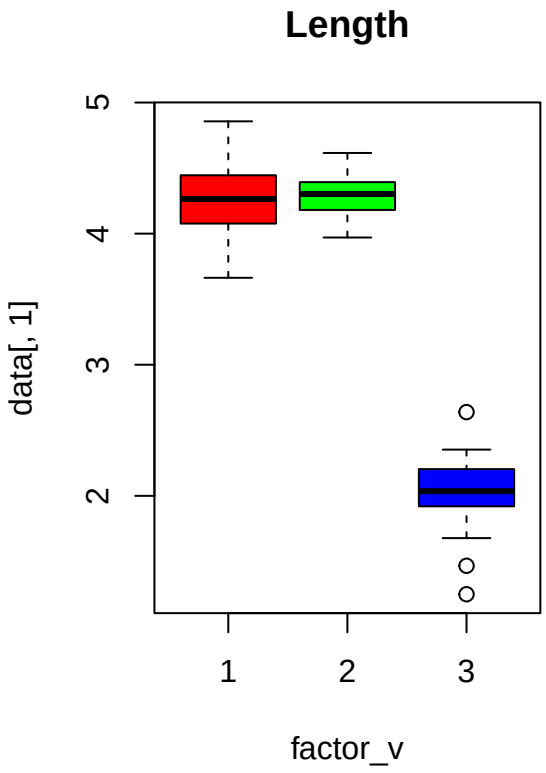
We report the resulting clustering:

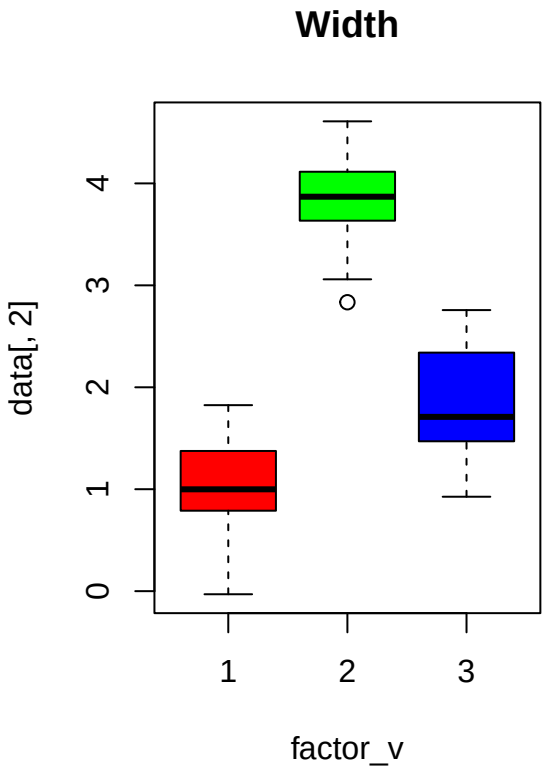


And other plots, by cluster:



and feature:





Point B

Let us check whether the three clusters have the same covariance structure; we will check the residuals after fitting the model. We will check the three covariance matrices and verify that no element on the diagonal is larger than 10 times (10 is specific for MANOVA, usually it is 4) any other element on the diagonals:

```
##              Length        Width
## Length  0.071160980 -0.008768262
## Width  -0.008768262  0.258558836
```

```
##            Length      Width
## Length 0.02705947 0.01211026
## Width  0.01211026 0.19607752
```

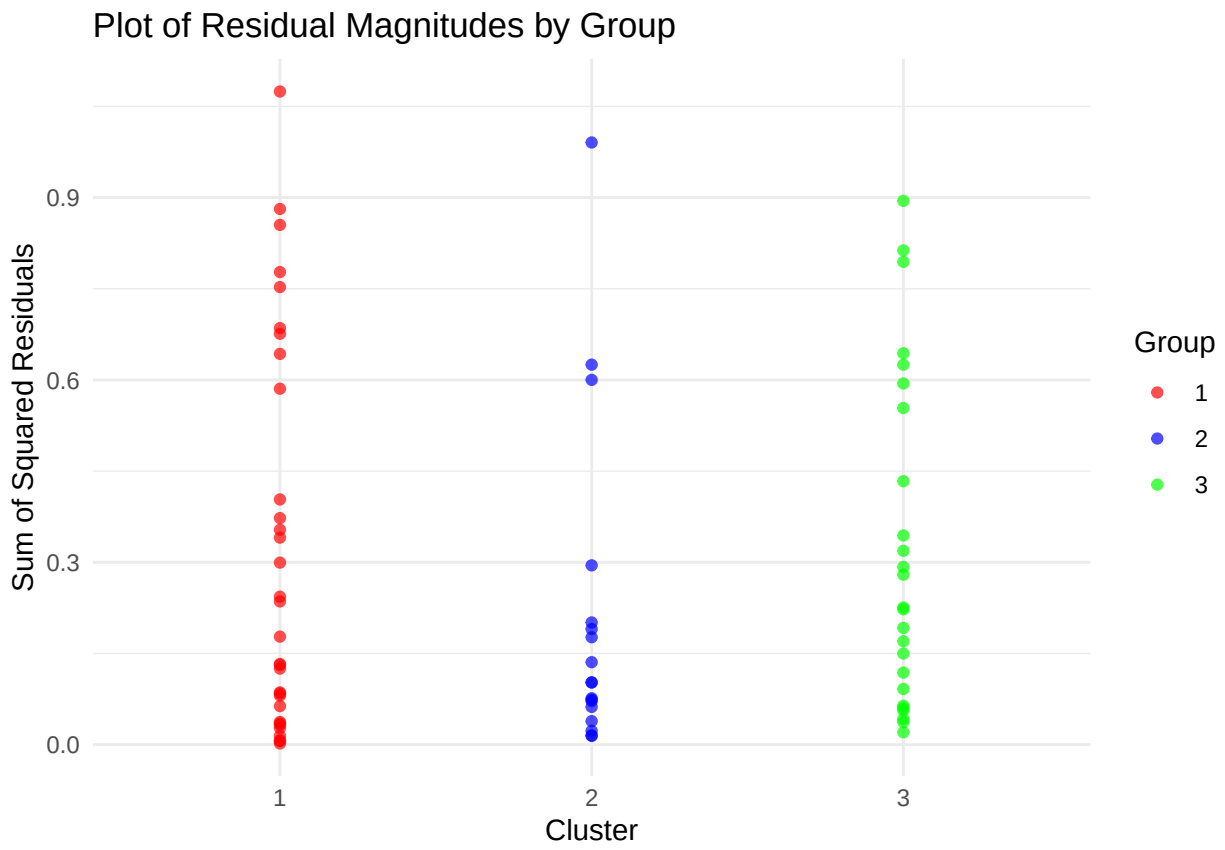
```
##              Length        Width
## Length 0.0848463042 0.0009094381
## Width  0.0009094381 0.2500743862
```

The test has positive results; we may proceed.

```
##           Df     Wilks approx F num Df den Df    Pr(>F)
## factor_v   2 0.0083476   353.05      4    142 < 2.2e-16 ***
## Residuals 72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The summary reports that the membership to a cluster does indeed have an effect on the mean features of a stone flake, as the p-value is practically 0.

Let us check whether the residuals are homoscedastic:




```
##              Length        Width
## Length  0.071160980 -0.008768262
## Width  -0.008768262  0.258558836
```

```
##            Length      Width
## Length 0.02705947 0.01211026
## Width  0.01211026 0.19607752
```

```
##              Length        Width
## Length 0.0848463042 0.0009094381
## Width  0.0009094381 0.2500743862
```

The homoscedasticity does not seem to be respected too well.

Point C

```
## $`diff 1-2`
##             inf12      sup12
## Length -0.2216299  0.1474938
## Width  -3.1823644 -2.4733238
##
## $`diff 2-3`
##           inf23    sup23
## Length 2.084059 2.471340
## Width  1.584009 2.327927
##
```

```
## $`diff 1-3`
##            inf13      sup13
## Length  2.073422  2.4078408
## Width  -1.193064 -0.5506875
```

Only the interval on length between the first and second clusters constains 0, which means that these two clusters could likely have the same mean. All other combinations of clusters have means that differ, some with more confidence than others.