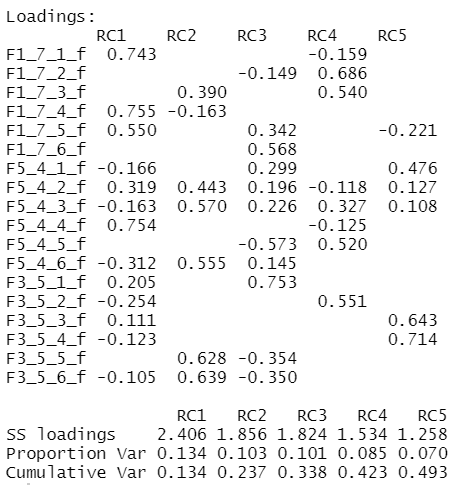
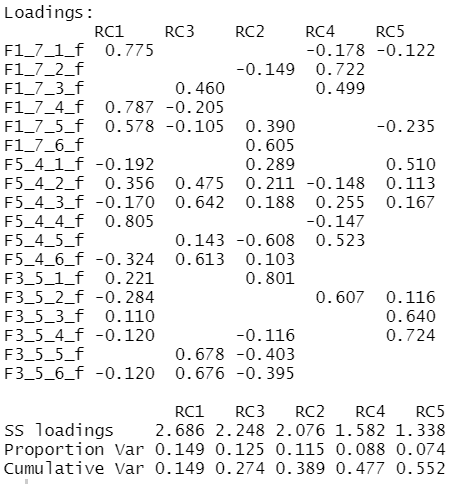
1. **PCA ordered mit 5 factors**

Unterschied zu normaler PCA ist, dass eine polychoric correlation matix verwendet wird

Ergibt folgende factor loadings

Links mit ordered PCA, rechts ‚normale‘ PCA (Ergebnis wie in SPSS)



* Variablen mit höchsten Ladungen sind:

# RC 1: F1\_7\_4\_f , F5\_4\_4\_f

# RC 2: F3\_5\_6\_f, F3\_5\_5\_f

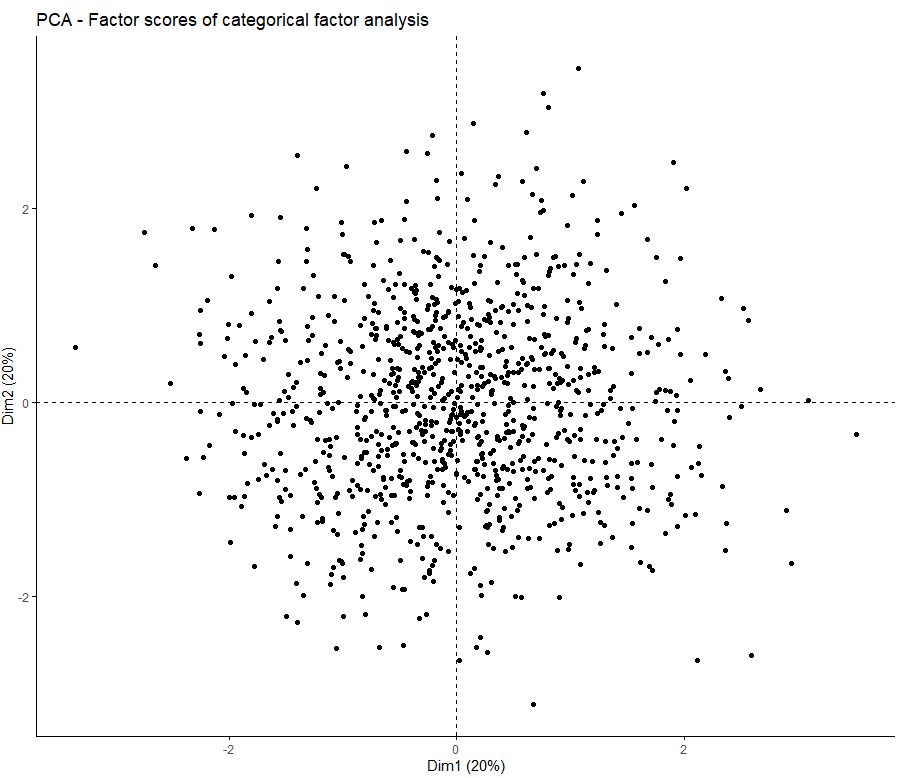
# RC 3: F3\_5\_1\_f, F5\_4\_5\_f

# RC 4: F1\_7\_2\_f, F3\_5\_2\_f

# RC 5: F3\_5\_4\_f, F3\_5\_3\_f

1. **Visual inspection of PCA results**

Plot der 5 factor scores (Erklärung: If there are more than two dimensions (wir haben 5) fviz\_cluster will perform principal component analysis (PCA) and plot the data points according to the first two principal components that explain the majority of the variance.



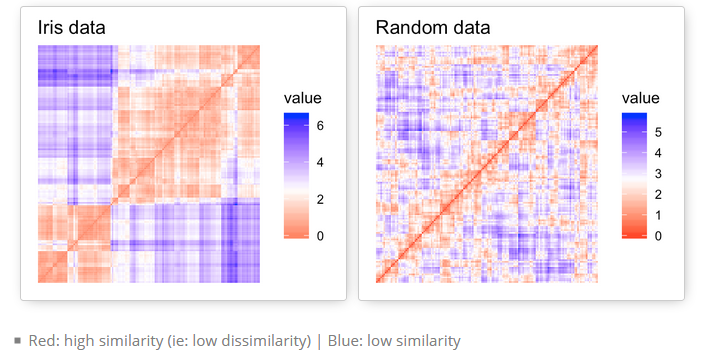
Man sieht hier schon, dass sich keine wirklichen Cluster der Faktoren zeigen.

1. **Hopkins statistic = 0.66 und dissimilary matrix**

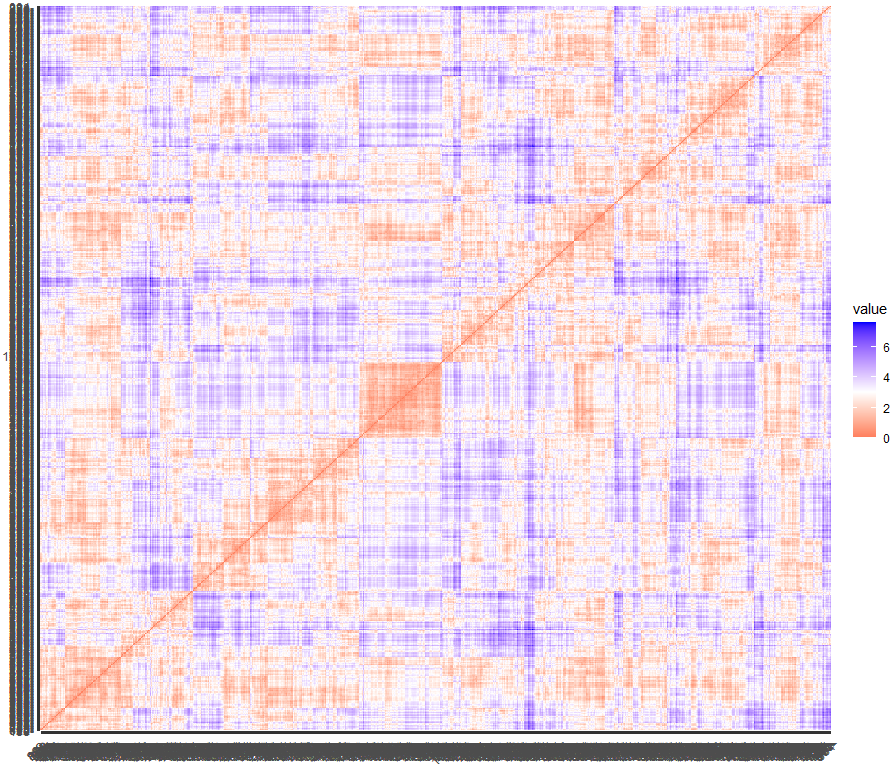
**Hopkins statistic**: If the value of Hopkins statistic is close to 1 (far above 0.5), then we can conclude that the dataset is significantly clusterable (random data has 0.5).

**Visualize the dissimilarity matrix**

Beispiel: Links clusterd, rechts random data

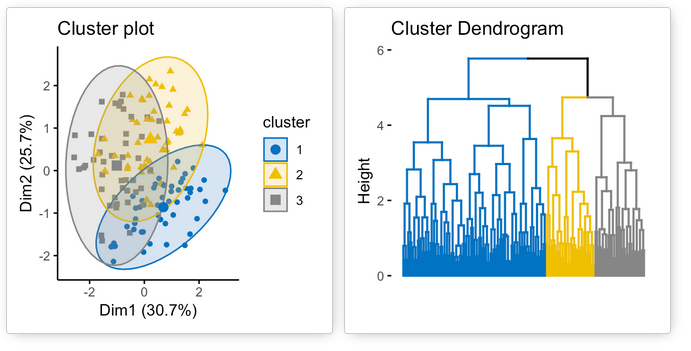


Unser Datensatz (factor scores):



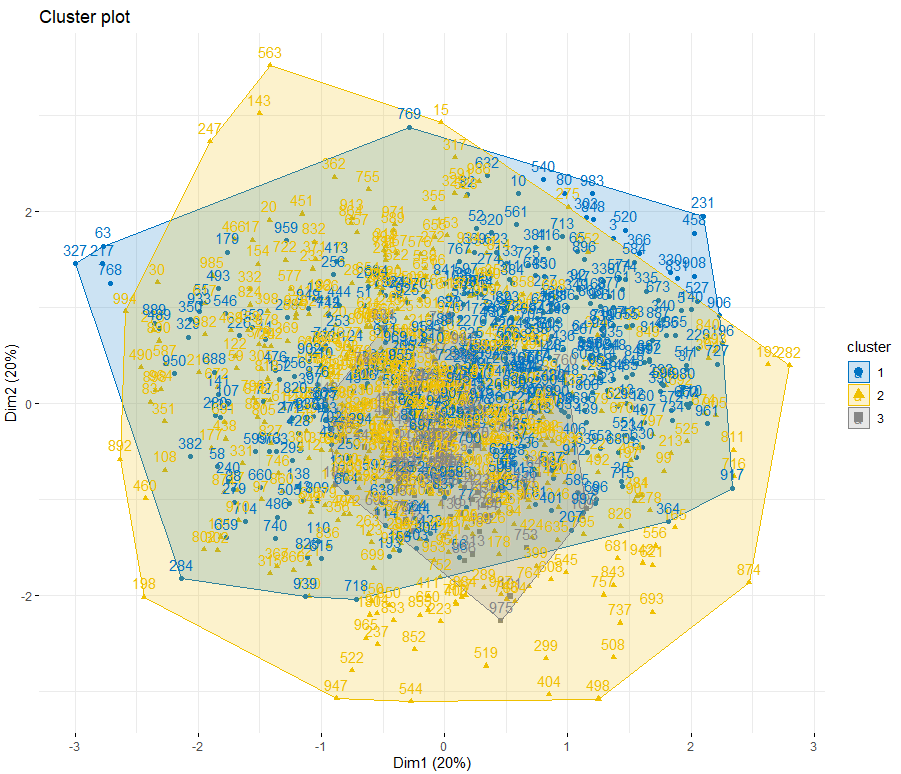
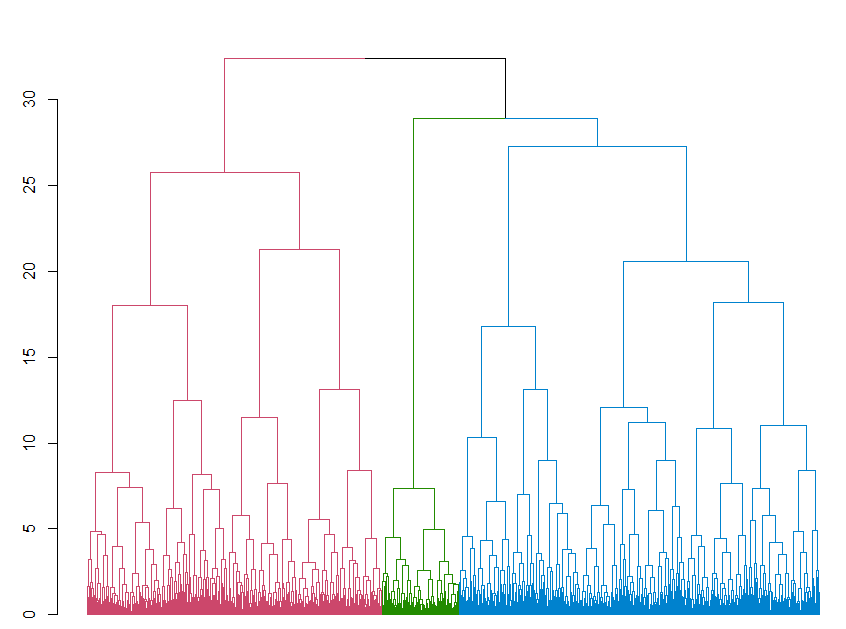
1. **Hierachische Clusteranlyse (mit Euclidean, Ward)**

Beispiel: random data



Mit unseren Daten (als Input factor scores): Mit 3 clustern (auch mit 2 ist es nicht besser)

Cluster Plot Cluster Denogramm



1. **Validity measures**

Score Method Clusters

Connectivity 69.5433 hierarchical 3

Dunn 0.1034 hierarchical 3

Silhouette 0.1165 hierarchical 3

connectivity score ∈ [0, infinity) 🡪 minimize

Dunn ∈ [0,infinity]) 🡪 maximize

Silhouette ∈ [-1,1] 🡪 maximize

**Noch versucht:**

* **Es wird nicht mehr geclustert wenn man mehr als 5 factor scores verwendet**
* **k-means statt hieracical clustering**
* **die am wenigsten korrelierten Variablen rausgeben**
* **keine Faktorenanalyse sondern jene beiden Variablen, die pro Faktor am höchsten laden verwenden**

References

<https://www.datanovia.com/en/lessons/assessing-clustering-tendency/>

Inspect correlations

