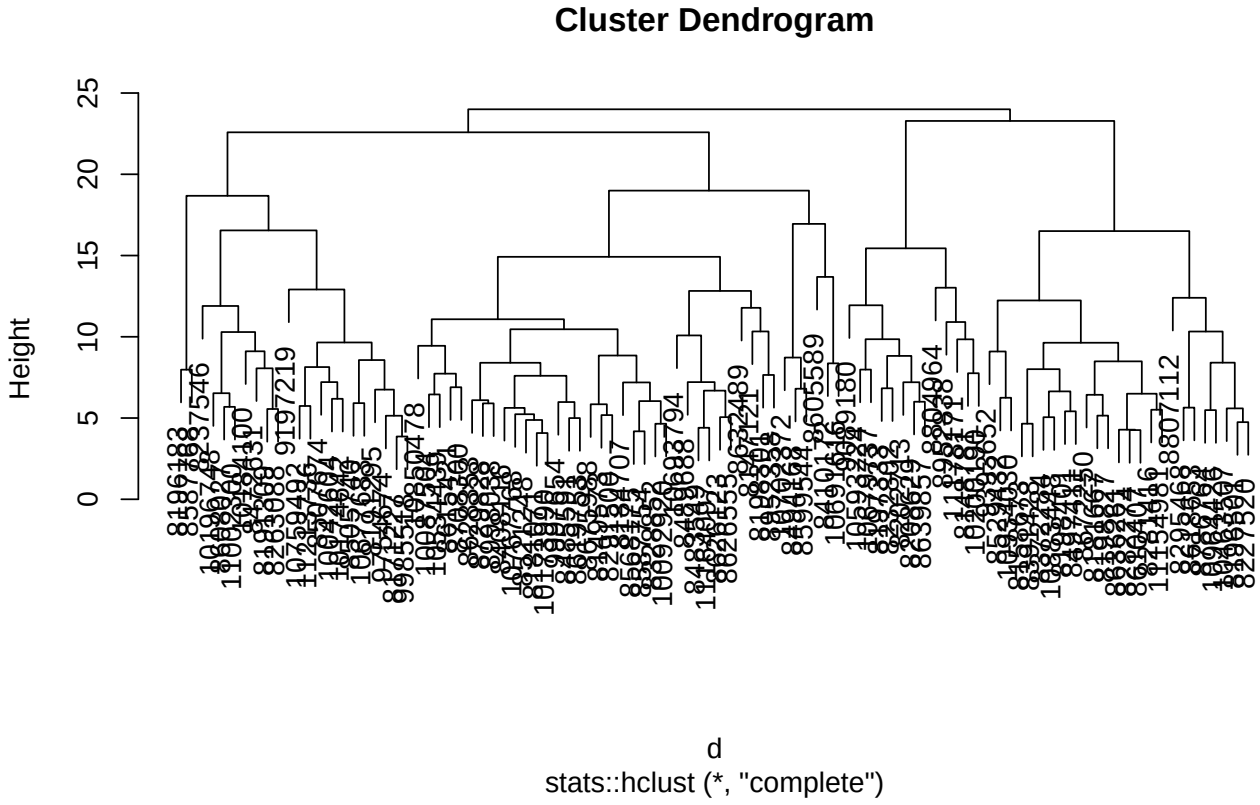
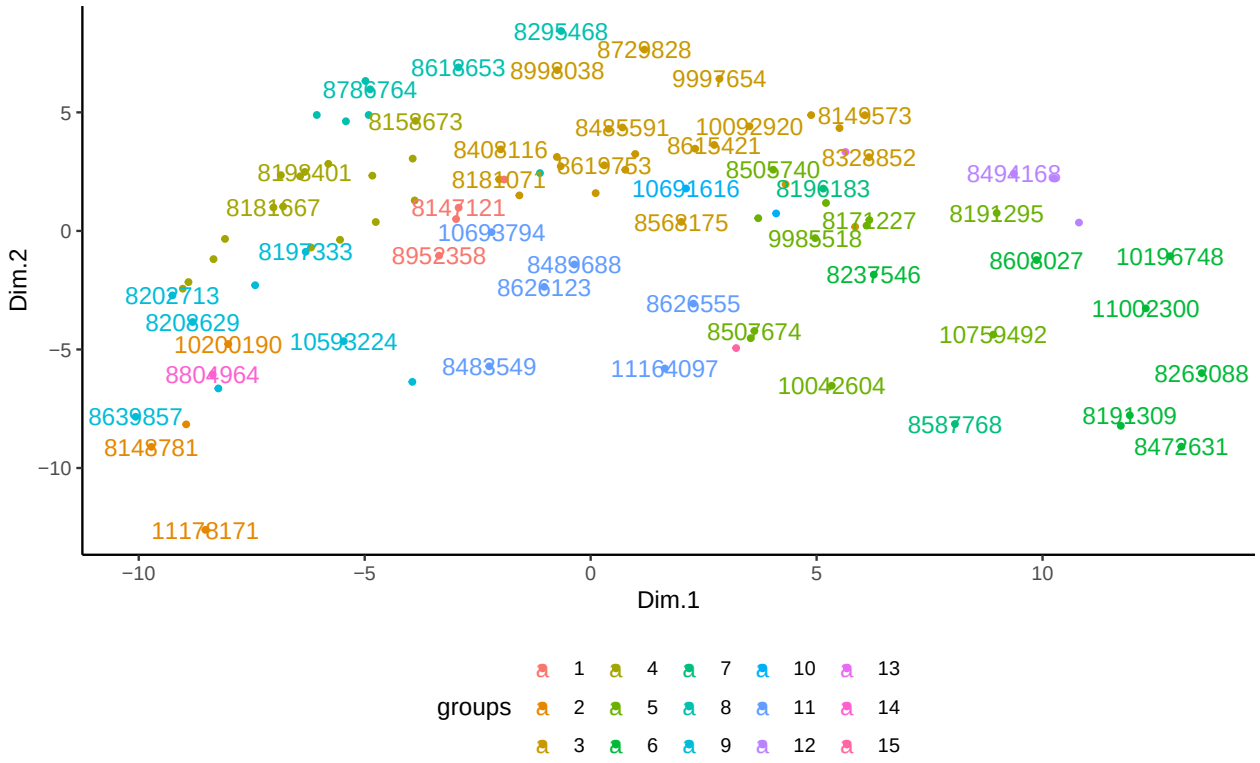
Clustering real data with 100 customers (no missing value)

1 Clutering approach

1. *Compute quantiles of distributions across each hour of day*
2. *Compute JS distance between households for each hour of day*
3. *Total distance between households computed as sum of JS distances for all hours*
4. *Cluster using this distance with hierarchical clustering algorithm (method "complete")*

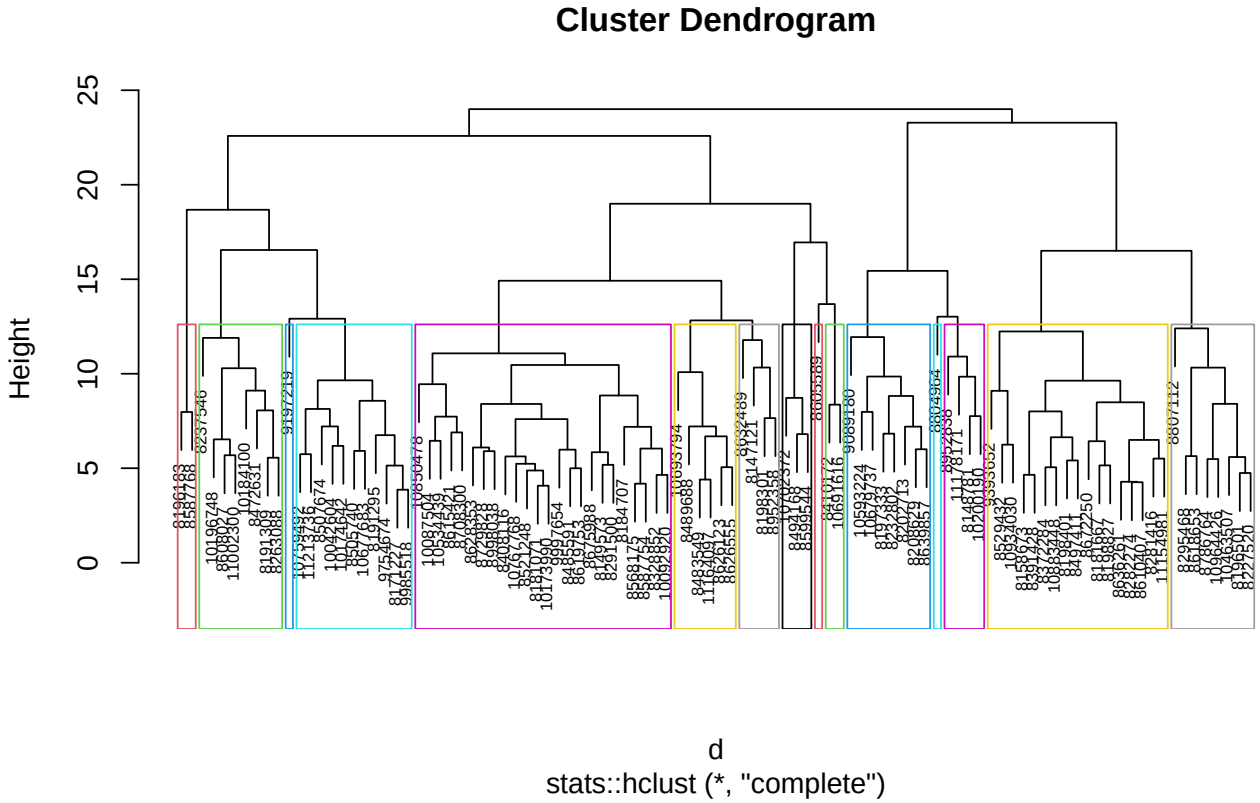
2 Clustering results





Optimal number of clusters as defined by the *fpc::nselectboot* is 15.

$k = 15$



3 Cluster characterization

The quartile deviation of different clusters are drawn. The shape of daily load curve is different for all of them. Across each cluster, customers should show different shape and within each cluster, customers should differ in size.