7. Groups

Hierarchical Dendrogram

*# install.packages("ggdendro")*

**library**(ggplot2)

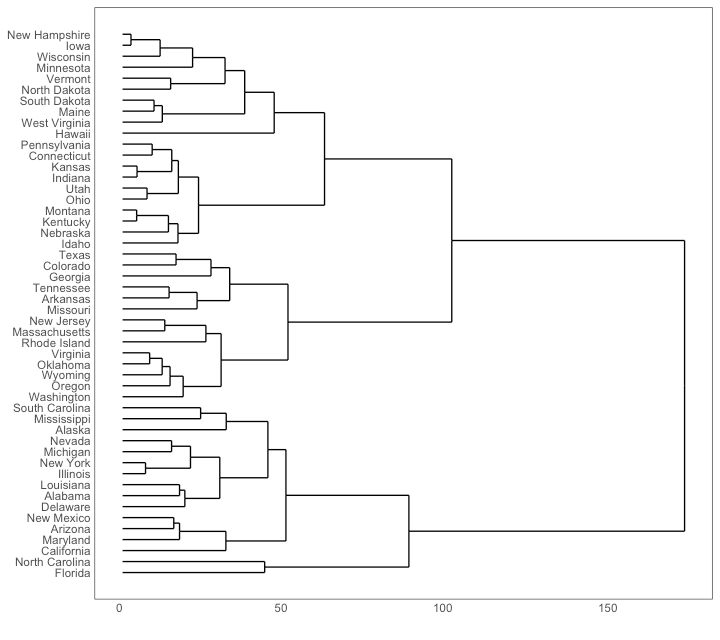
**library**(ggdendro)

**theme\_set**(**theme\_bw**())

hc <- **hclust**(**dist**(USArrests), "ave") *# hierarchical clustering*

*# plot*

**ggdendrogram**(hc, rotate = TRUE, size = 2)



Clusters

It is possible to show the distinct clusters or groups using geom\_encircle(). If the dataset has multiple weak features, you can compute the principal components and draw a scatterplot using PC1 and PC2 as X and Y axis.

The geom\_encircle() can be used to encircle the desired groups. The only thing to note is the dataargument to geom\_circle(). You need to provide a subsetted dataframe that contains only the observations (rows) that belong to the group as the data argument.

*# devtools::install\_github("hrbrmstr/ggalt")*

**library**(ggplot2)

**library**(ggalt)

**library**(ggfortify)

**theme\_set**(**theme\_classic**())

*# Compute data with principal components ------------------*

df <- iris[**c**(1, 2, 3, 4)]

pca\_mod <- **prcomp**(df) *# compute principal components*

*# Data frame of principal components ----------------------*

df\_pc <- **data.frame**(pca\_mod$x, Species=iris$Species) *# dataframe of principal components*

df\_pc\_vir <- df\_pc[df\_pc$Species == "virginica", ] *# df for 'virginica'*

df\_pc\_set <- df\_pc[df\_pc$Species == "setosa", ] *# df for 'setosa'*

df\_pc\_ver <- df\_pc[df\_pc$Species == "versicolor", ] *# df for 'versicolor'*

*# Plot ----------------------------------------------------*

**ggplot**(df\_pc, **aes**(PC1, PC2, col=Species)) +

**geom\_point**(**aes**(shape=Species), size=2) + *# draw points*

**labs**(title="Iris Clustering",

subtitle="With principal components PC1 and PC2 as X and Y axis",

caption="Source: Iris") +

**coord\_cartesian**(xlim = 1.2 \* **c**(**min**(df\_pc$PC1), **max**(df\_pc$PC1)),

ylim = 1.2 \* **c**(**min**(df\_pc$PC2), **max**(df\_pc$PC2))) + *# change axis limits*

**geom\_encircle**(data = df\_pc\_vir, **aes**(x=PC1, y=PC2)) + *# draw circles*

**geom\_encircle**(data = df\_pc\_set, **aes**(x=PC1, y=PC2)) +

**geom\_encircle**(data = df\_pc\_ver, **aes**(x=PC1, y=PC2))

