

# Statistical Learning Methods Exercise 11

Tobias Famos

## Preliminaries

Loading the Votes

```
votes <- read.table("/home/tobias/unibe/statistical methods in R/Exercise 11/Votes.txt", header=T)
sum(is.na(votes))
```

```
## [1] 0
```

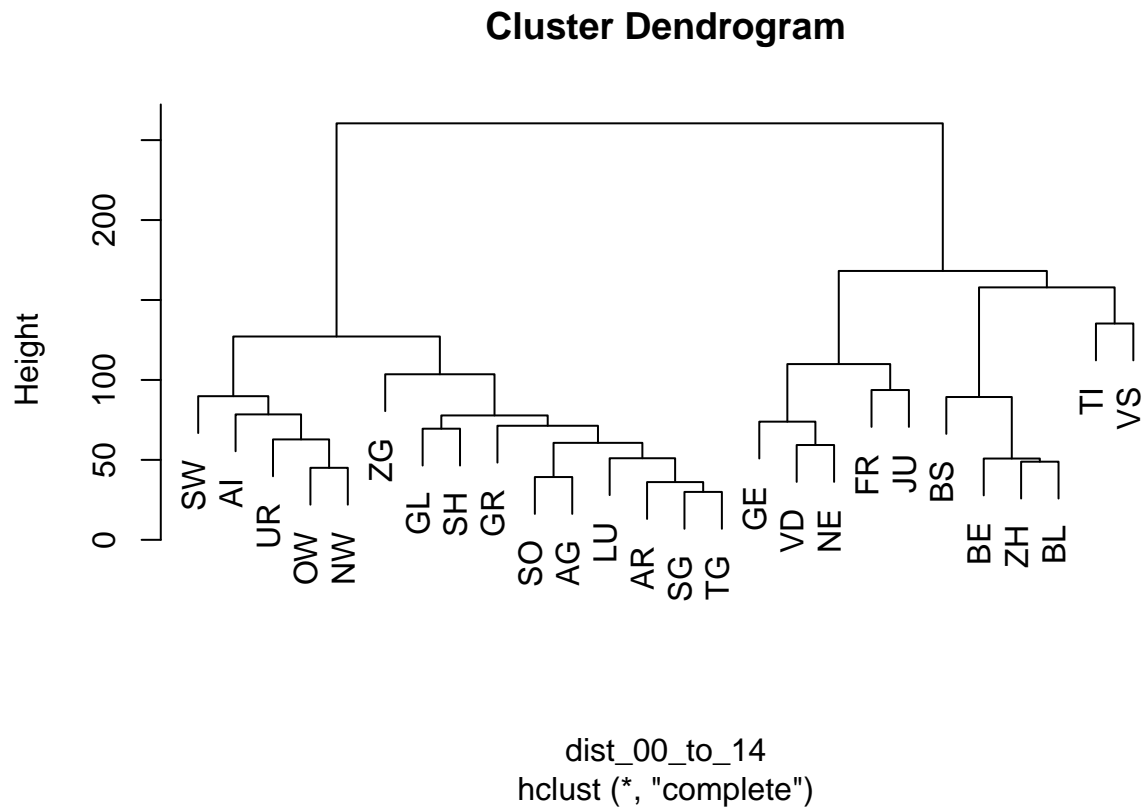
Add the row names to prevent objects to only show up as numbers in plots

```
rownames(votes) <- votes$Vote
```

## Task 1 Cluster the votes from 2000 until 20145

Select the votes for the clustering

```
votes_00_to_14 <- votes[28:161]
dist_00_to_14 <- dist(votes_00_to_14)
cluster_00_to_14 <- hclust(dist_00_to_14, method="complete")
plot(cluster_00_to_14)
```

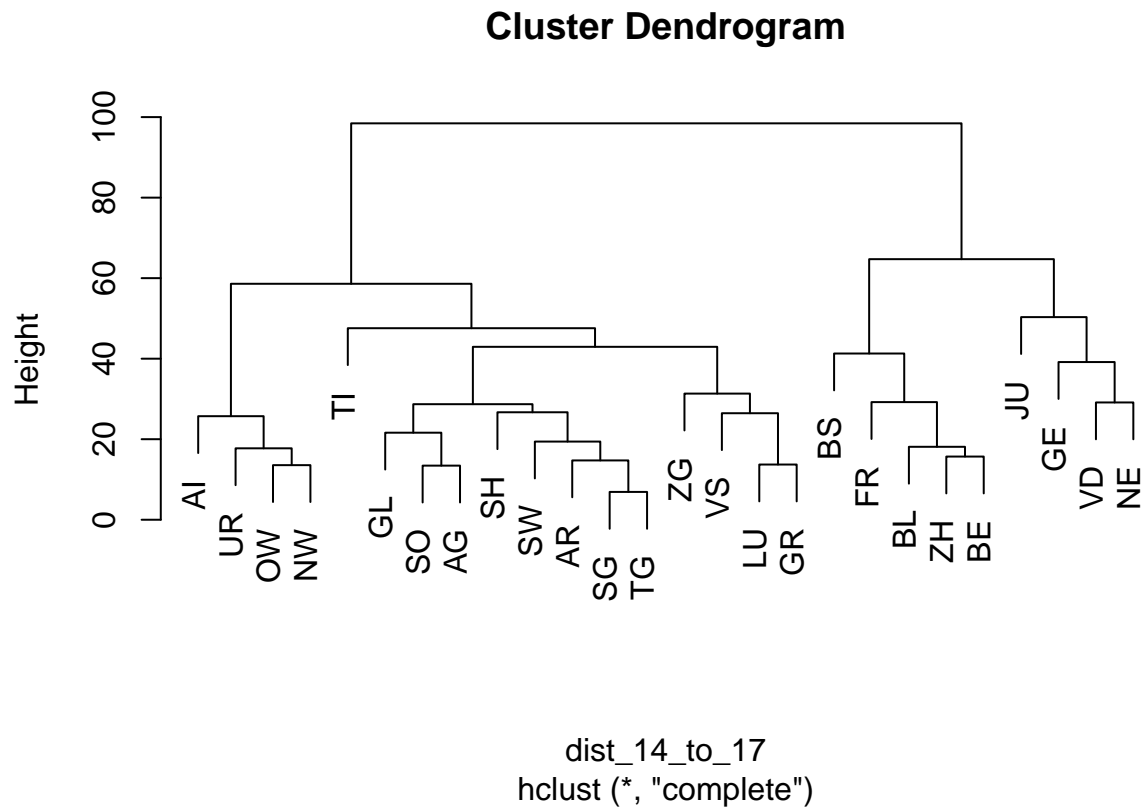


## Task 2 Cluster votes from 2014 to 2017

```
votes_14_to_17 <- votes[0:27]
dist_14_to_17 <- dist(votes_14_to_17)

## Warning in dist(votes_14_to_17): NAs introduced by coercion

labels <- votes$Vote
cluster_14_to_17 <- hclust(dist_14_to_17, method="complete")
plot(cluster_14_to_17)
```



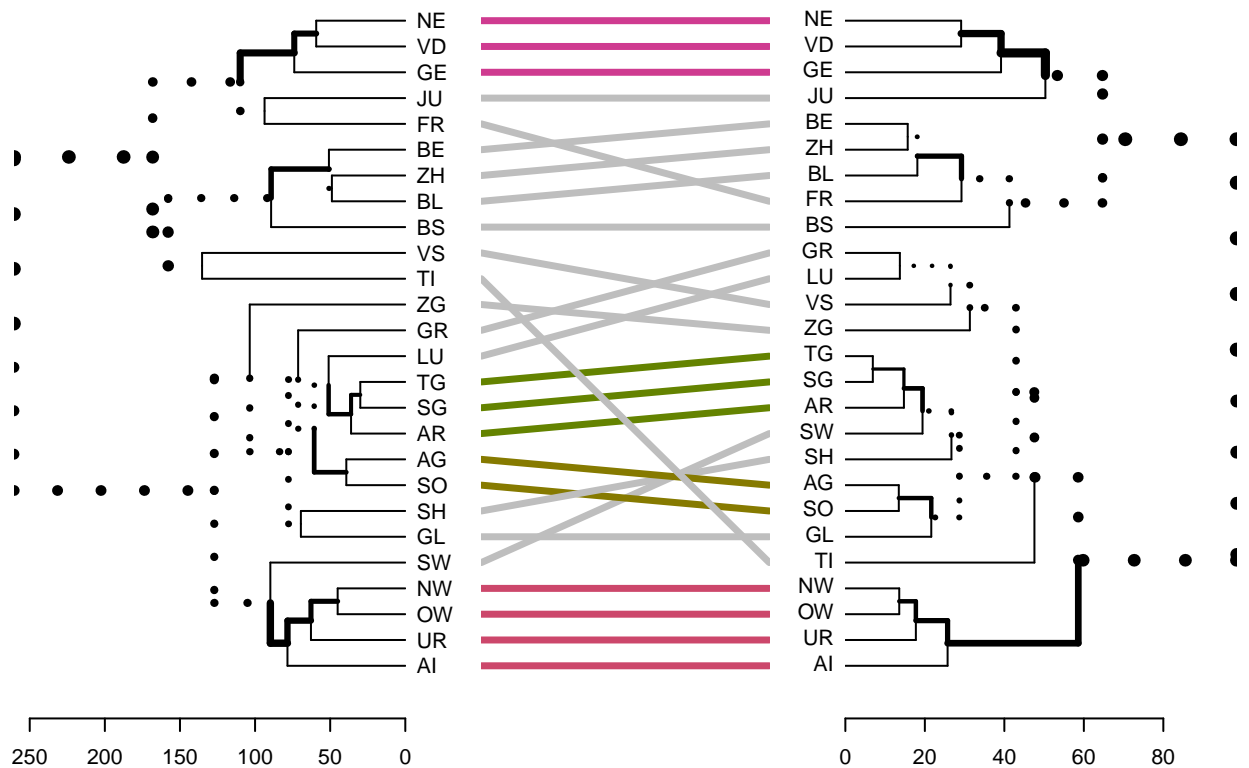
## Task 3 Compare

```
library(dendextend)
```

```
##
## -----
## Welcome to dendextend version 1.15.2
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## You may ask questions at stackoverflow, use the r and dendextend tags:
##   https://stackoverflow.com/questions/tagged/dendextend
##
## To suppress this message use: suppressPackageStartupMessages(library(dendextend))
## -----
##
## Attaching package: 'dendextend'
##
## The following object is masked from 'package:stats':
##
##   cutree
dend00 <- as.dendrogram (cluster_00_to_14)
dend14 <- as.dendrogram (cluster_14_to_17)
```

```
dend_list <- dendlist(dend00, dend14)
```

```
dendlist(dend00, dend14) %>%  
  untangle(method = "step1side") %>% # Find the best alignment layout  
  tanglegram()
```



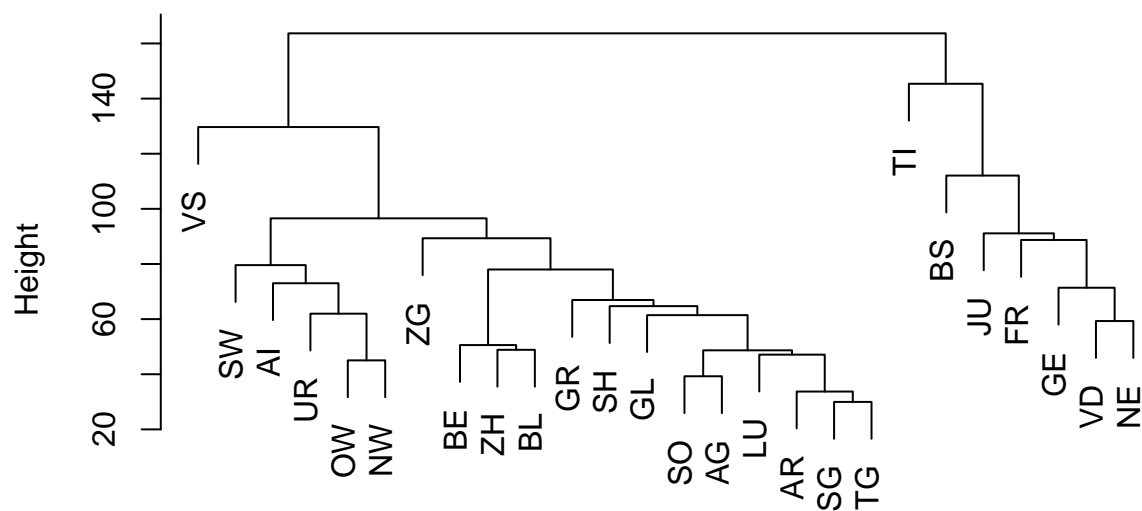
In both clustering approaches there are two pretty distinct clusters emerging. When comparing the two pretty distinct clusters, the only different canton in them is the TI.

## Task 4 Average Link method

For 2000 to 2014

```
cluster_00_to_14_average <- hclust(dist_00_to_14, method="average")  
plot(cluster_00_to_14_average)
```

## Cluster Dendrogram

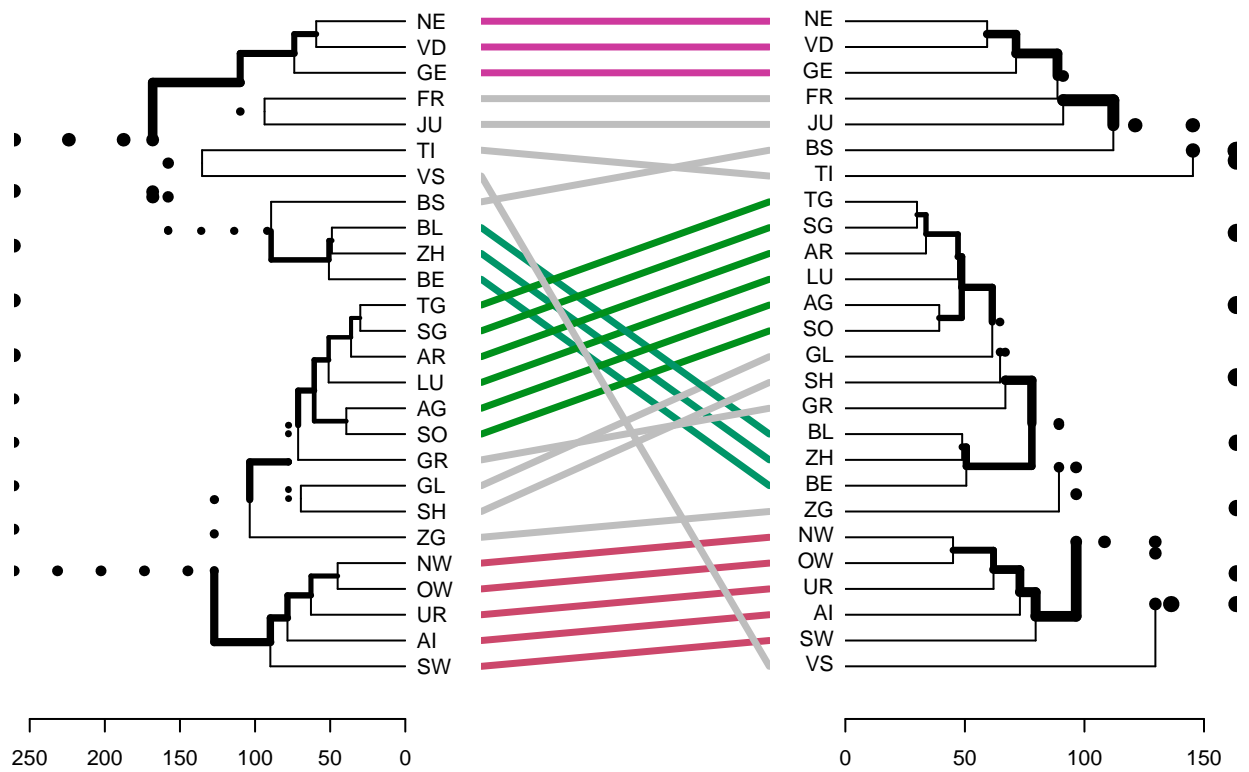


```
dist_00_to_14
hclust (*, "average")
```

```
dend <- as.dendrogram (cluster_00_to_14)
dendAverage <- as.dendrogram (cluster_00_to_14_average)

dend_list <- dendlist(dend, dendAverage)

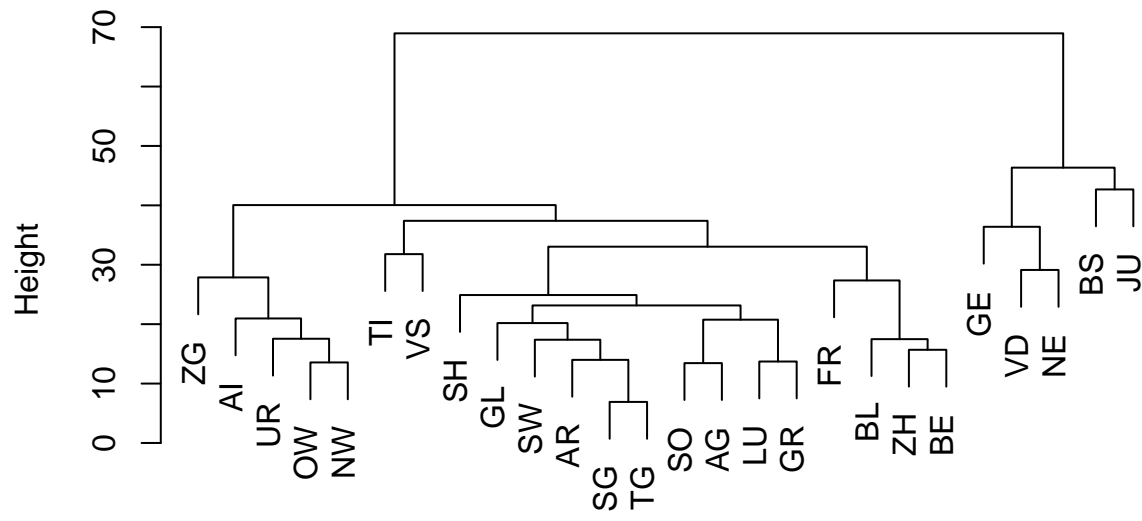
dendlist(dend, dendAverage) %>%
  untangle(method = "step1side") %>% # Find the best alignment layout
  tanglegram()
```



When clustering the same data with the average method, we get again two pretty distinct clusters, but the members of the clusters have changed quite a lot

```
cluster_14_to_17_average <- hclust(dist_14_to_17, method="average")
plot(cluster_14_to_17_average)
```

## Cluster Dendrogram

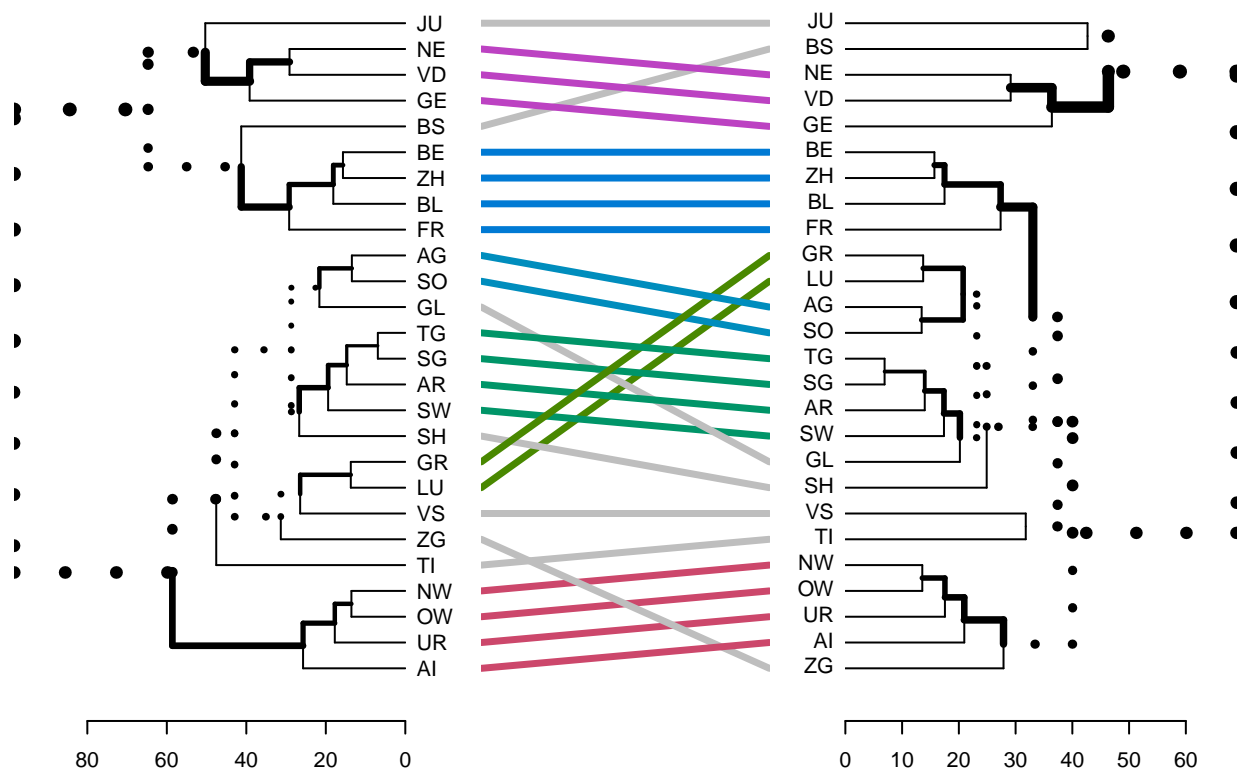


dist\_14\_to\_17  
hclust (\*, "average")

```
dend <- as.dendrogram (cluster_14_to_17)
dendAverage <- as.dendrogram (cluster_14_to_17_average)

dend_list <- dendlist(dend, dendAverage)

dendlist(dend, dendAverage) %>%
  untangle(method = "step1side") %>% # Find the best alignment layout
  tanglegram()
```



The same thing holds for the 2014 to 2017 Data