

Chapter 3

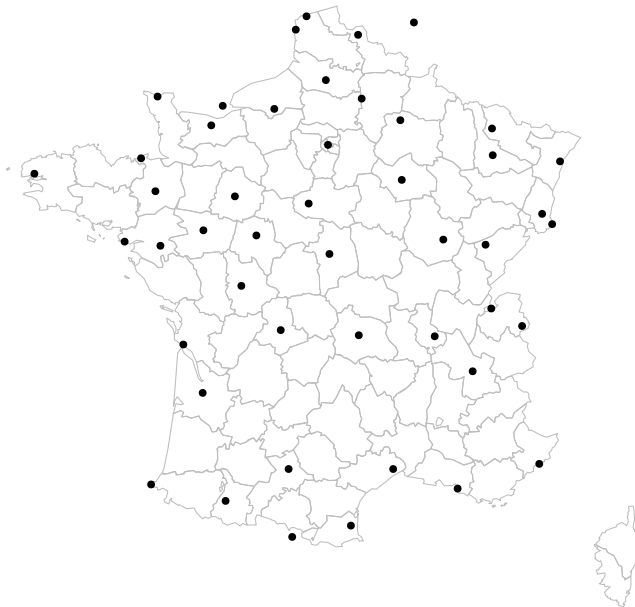
Clustering

Xavier Gendre

M2 SE

Distance data set

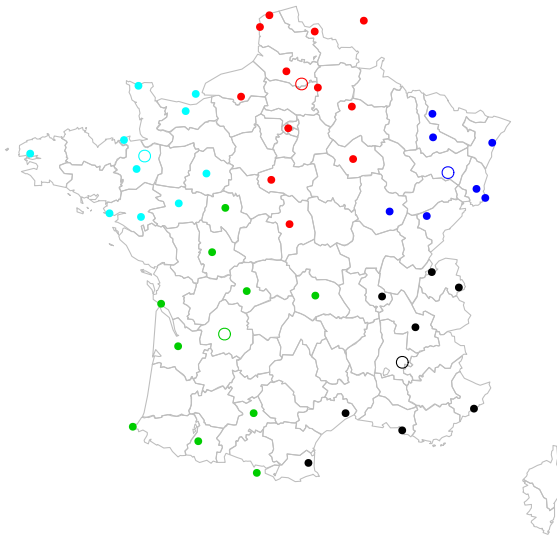
| | Amiens | Andorre | Angers | Bâle | LaBaule | Besançon | Bor |
|----------|--------|---------|--------|------|---------|----------|-----|
| Amiens | 0 | 1020 | 440 | 560 | 590 | 560 | 7 |
| Andorre | 1020 | 0 | 760 | 1130 | 830 | 970 | 4 |
| Angers | 440 | 760 | 0 | 770 | 160 | 620 | 3 |
| Bâle | 560 | 1130 | 770 | 0 | 940 | 160 | 8 |
| LaBaule | 590 | 830 | 160 | 940 | 0 | 770 | 4 |
| Besançon | 560 | 970 | 620 | 160 | 770 | 0 | 7 |
| Bordeaux | 730 | 430 | 340 | 840 | 400 | 700 | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |



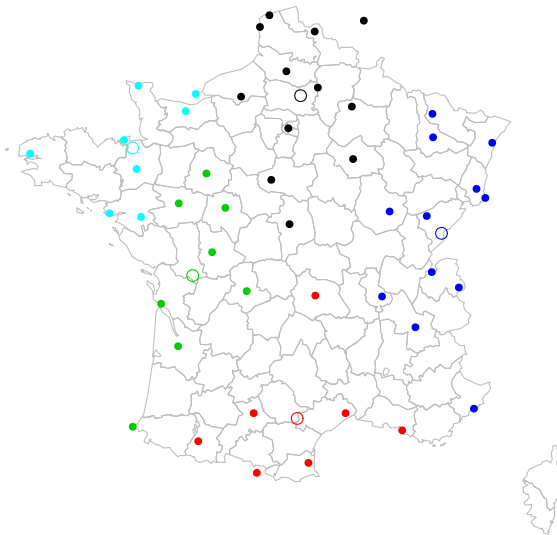
```
K <- 5
start <- sample(1:47, K)
VilleNom[start]

## [1] "Nice"      "Rouen"     "Orléans"  "Reims"    "LeHavre"
```

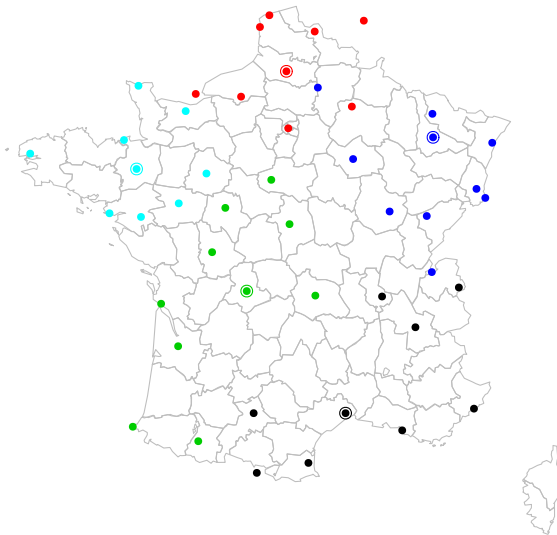
K-means (no variant) : 5 steps



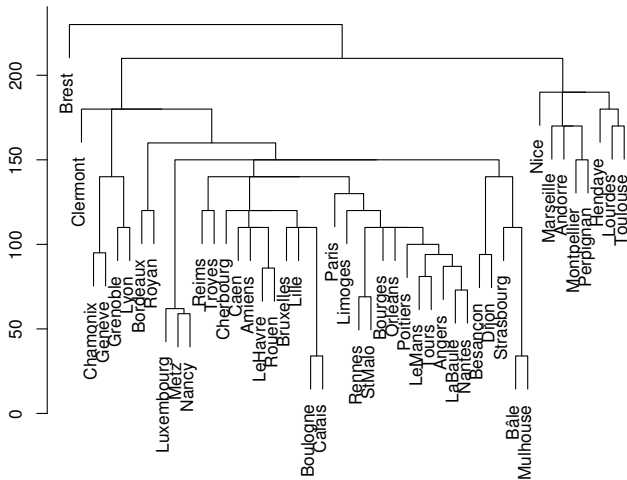
K-means (variant 1) : 3 steps



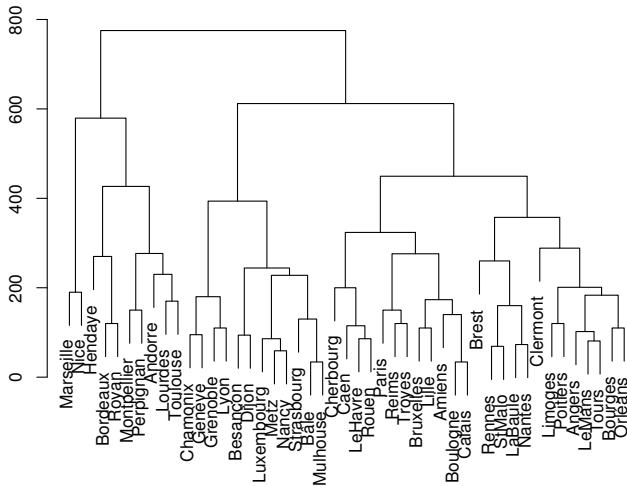
K-means (variant 2) : 4 steps



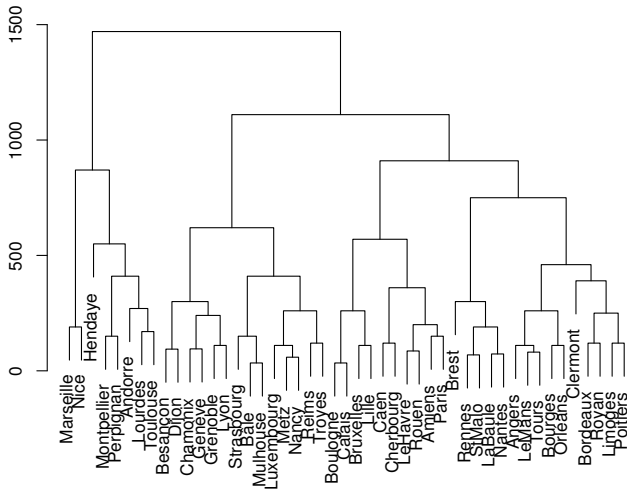
Single linkage



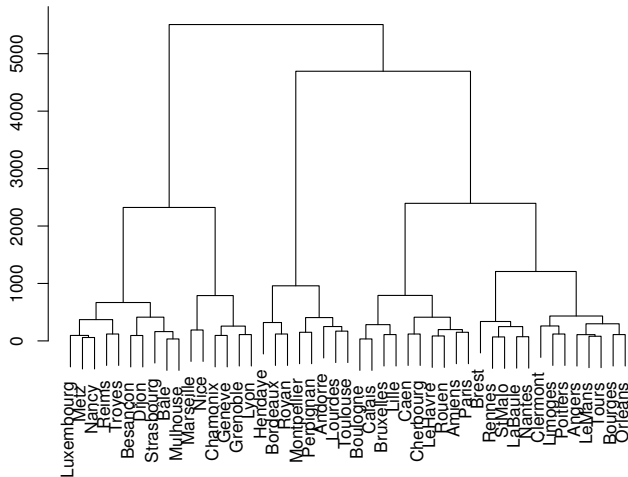
Group average linkage



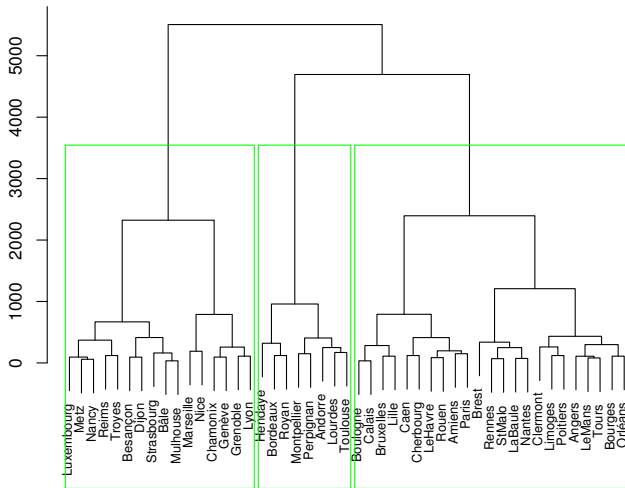
Complete linkage



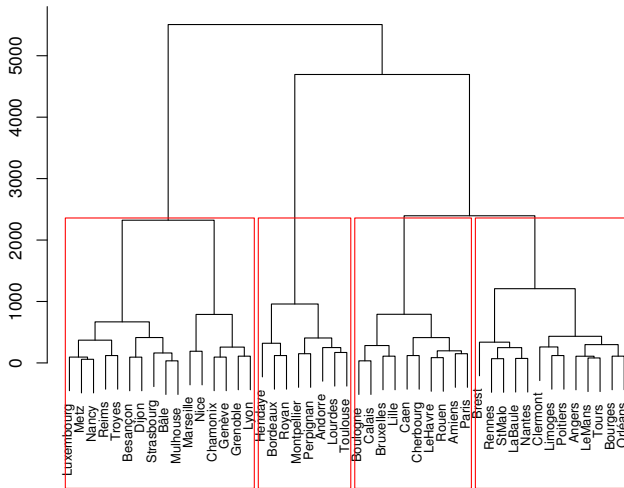
Ward distance



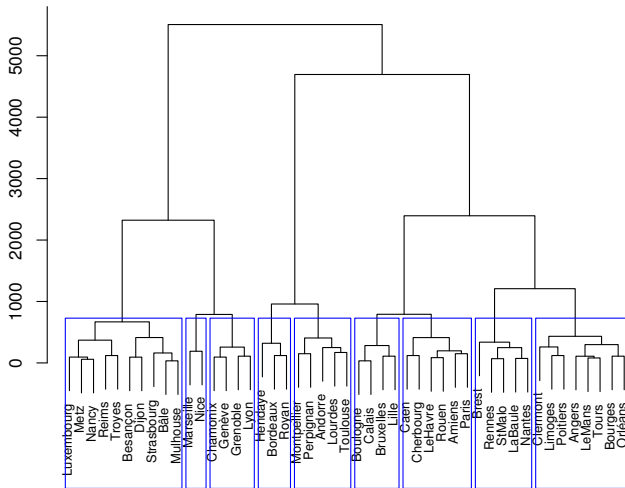
Ward distance - 3 clusters



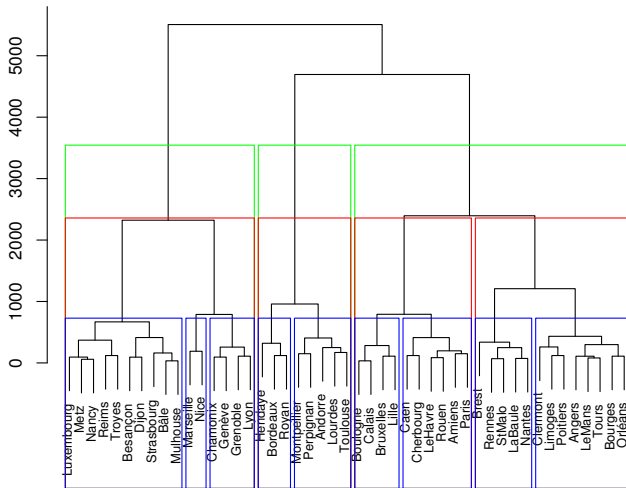
Ward distance - 4 clusters

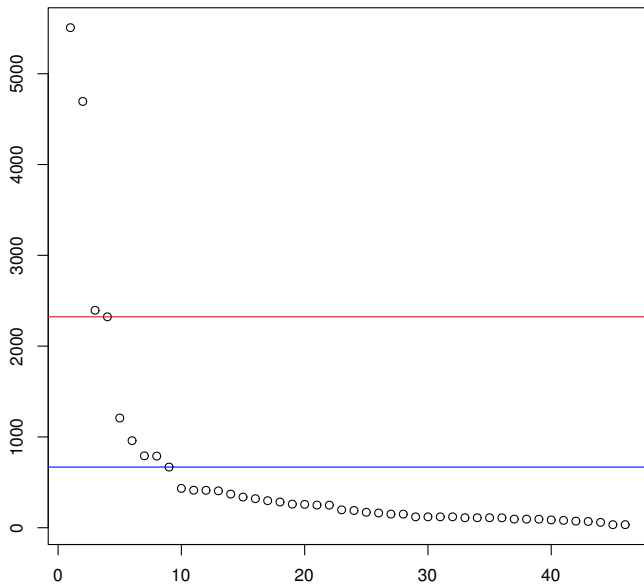


Ward distance - 9 clusters

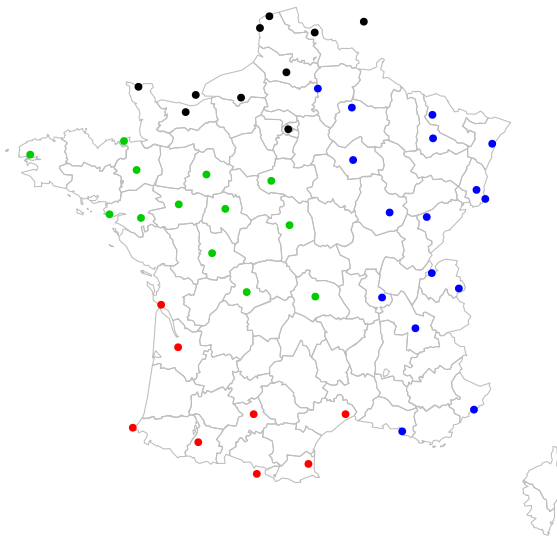


Ward distance - Choice?

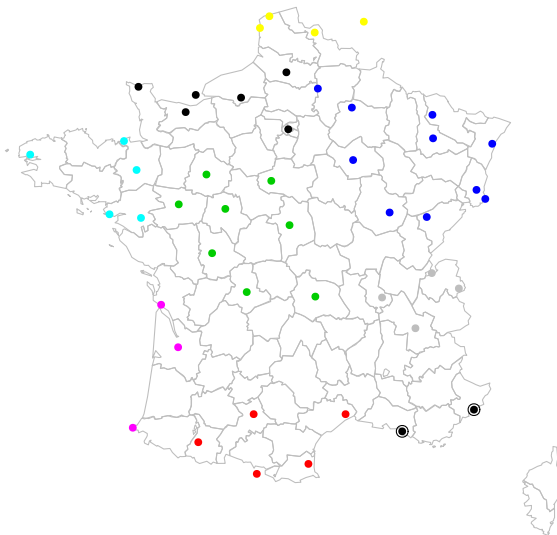




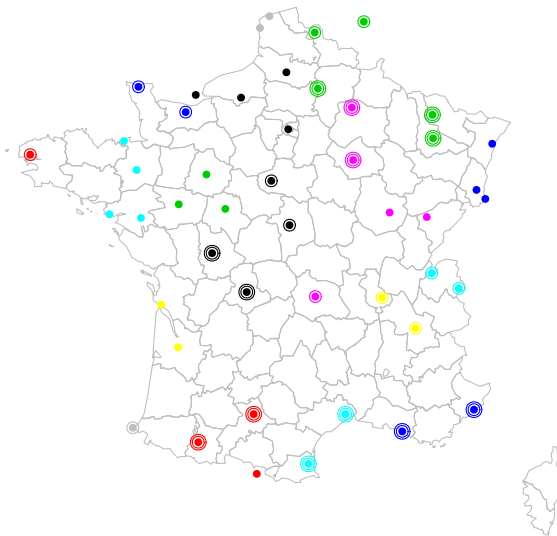
Hierarchical clustering : 4 clusters



Hierarchical clustering : 9 clusters



Hierarchical clustering : 22 clusters



Bibliography

- *The Elements of Statistical Learning : Data Mining, Inference and Prediction*, J. Friedman, T. Hastie and R. Tibshirani (2009)
- *Wiki Stat*, <http://wikistat.fr/>
- *Quick-R*, <http://www.statmethods.net/>