Tutorial to reproduce cluster analysis on the Papuan Malay spontaneous data.

Version 03-2021

This tutorial is meant as a practical guide to reproduce the contour clustering example in the accompanying article (Kaland, 2021, Section 3). For an in-depth description of the clustering procedures, see the manual that accompanies the article.

Files needed (from https://constantijnkaland.github.io/contourclustering/):
    - contour_clustering_GUI.R
    - pmy_spontaneous_phrases.csv

Software needed:
    - R (https://cran.r-project.org/)
    - RStudio (https://rstudio.com/)

R-packages needed:
    - ggplot2
    - reshape
    - reshape2
    - dplyr
    - shiny
Package installation instructions:
   - https://cran.r-project.org/doc/manuals/r-release/R-admin.html#Installing-packages

Download all of the above files, and install all the software and packages before continuing this tutorial.

Reference:

Kaland, C.C.L. (2021). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*.

1. Loading the contour clustering graphical user interface (GUI)

- Run R-studio

- From the R-studio main screen: File > Open file > Select 'contour_clustering_GUI.R'

- Run app:
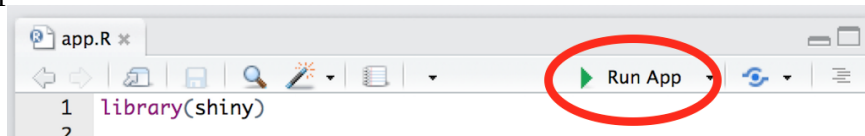
2. Load the datafile and prepare the data for clustering

- Load 'pmy_spontaneous_phrases.csv':

- Select speaker correction method 'Standardise':

- Apply selected options to data:

3. Start the cluster analysis.

- Obtain the dendrogram:

- Inspect table and plot:
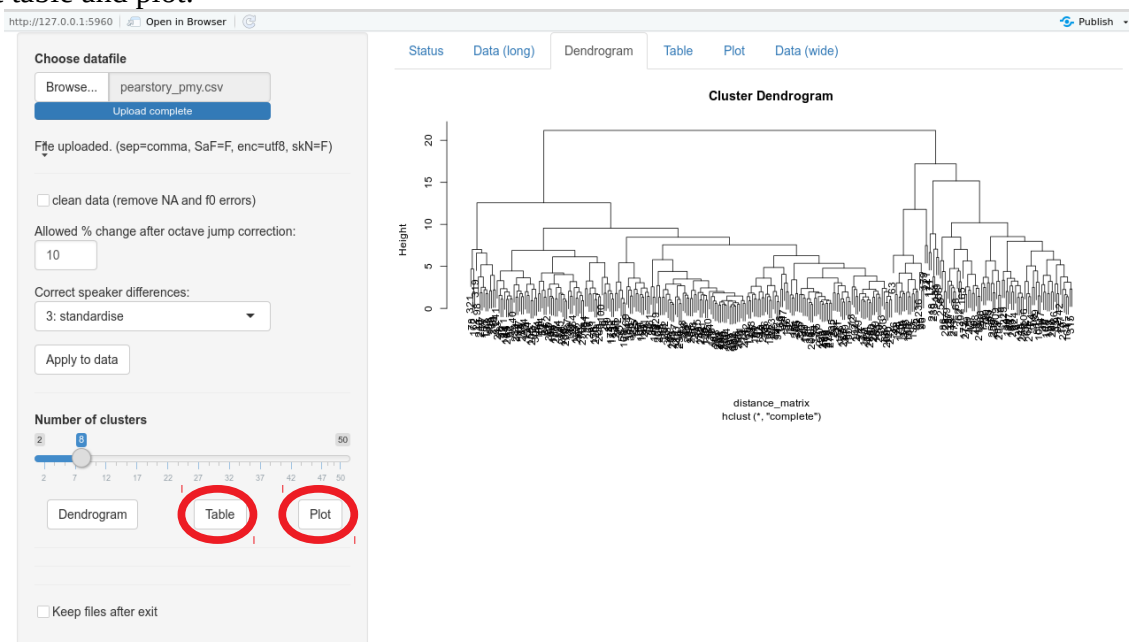
- Set number of clusters to 25 and apply subsetting until no clusters are flagged (2x):

- Set number of clusters back to 9 and obtain plot (identical to Figure 3d in Kaland 2021).