Supplement to "Interactive Exploration of Large Dendrograms with Prototypes"

Andee Kaplan

Department of Statistics, Colorado State University

and

Jacob Bien

Department of Data Sciences and Operations, University of Southern California

May 21, 2022

We provide a comparison of dendrograms resulting from a static plot, `idendr0` (Sieger et al. 2017), and `collapsibleTree` (Khan 2018).

A   Movies Case Study Comparison

We continue to explore a hierarchical clustering of 13,816 movies that was detailed in Section 3.1 of the main text.

Figures 2 – 4 show a sequence of working with this cluster object in `idendr0` (Sieger et al. 2017). The default view does not show the full leaf labels, but this can be adjusted with the margin specification. However, even though the entire dendrogram is showing, now only a small subset of the labels are shown. One can use the interactive zoom feature to have a look at a branch. It is not clear what the branch represents until zoomed in. This is because there are no prototype labels to summarize the contents of a branch and no search function to find the location of a particular label in the tree.

Figure 5 shows the result of loading the movies data into the `collapsibleTree` (Khan 2018) package. It should be noted that this package does not claim to be appropriate for dendrograms, and thus we do not see the typical features we would expect of a dendrogram (specifically, height corresponding to similarity between branches). Additionally, `collapsibleTree` has trouble with a tree of this size, taking over 18 hours to render.
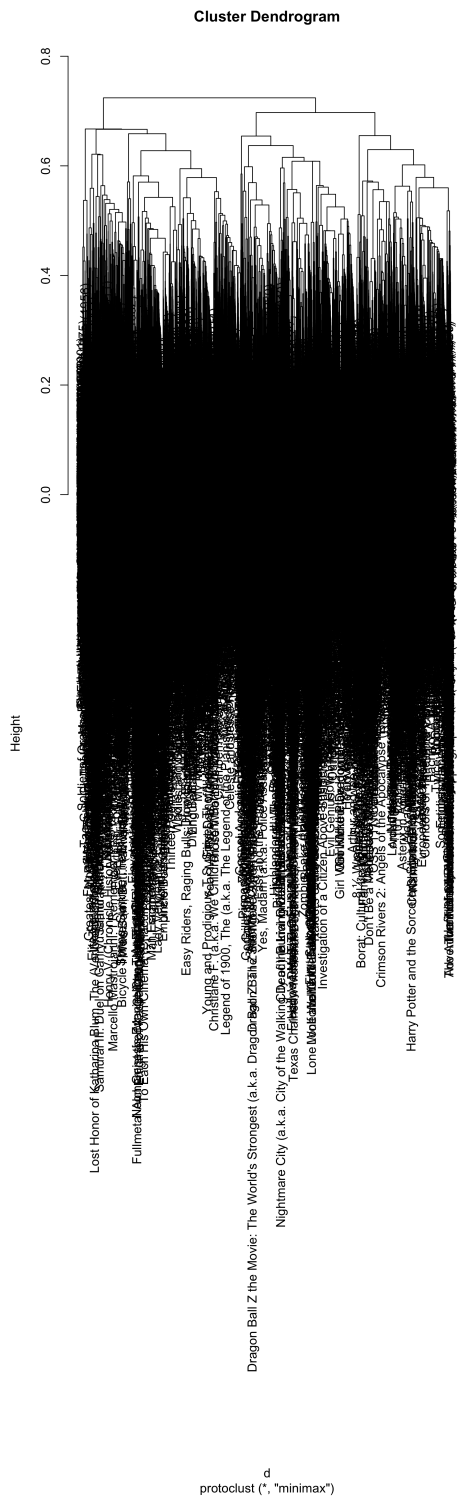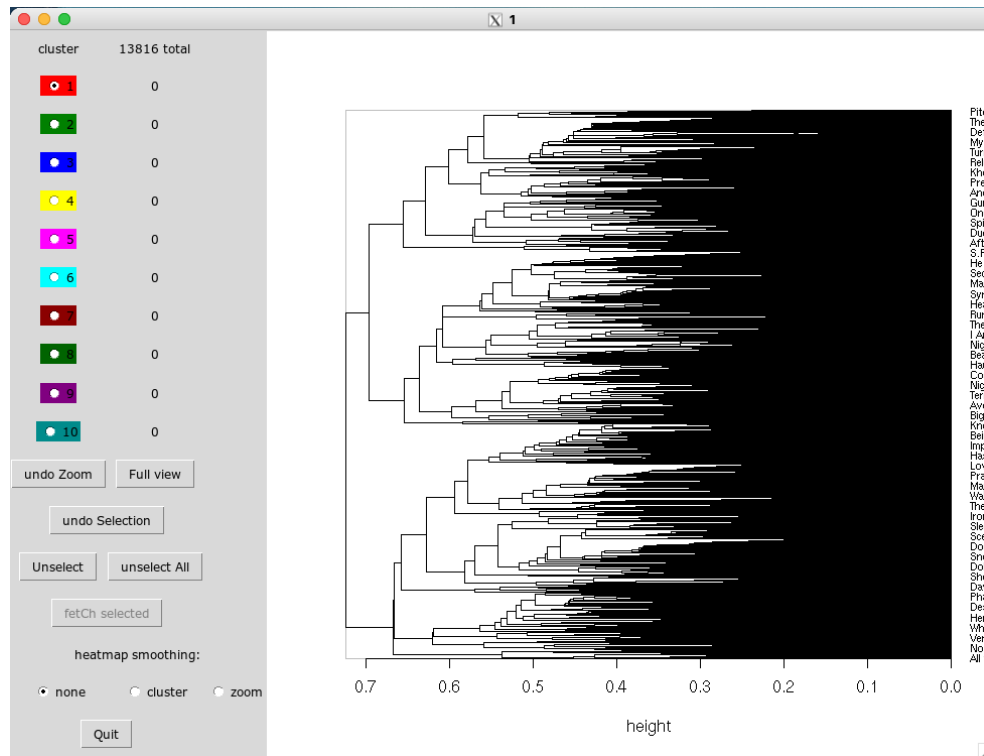
Figure 1: Static dendrogram for Movies case study.

Figure 2: Default view in `idendr0` of the movies data.

References

Khan, A. (2018), *collapsibleTree: Interactive Collapsible Tree Diagrams using 'D3.js'*. R
   package version 0.1.7.
   URL: *https://CRAN.R-project.org/package=collapsibleTree*

Sieger, T., Hurley, C. B., Fišer, K. & Beleites, C. (2017), 'Interactive dendrograms: The r
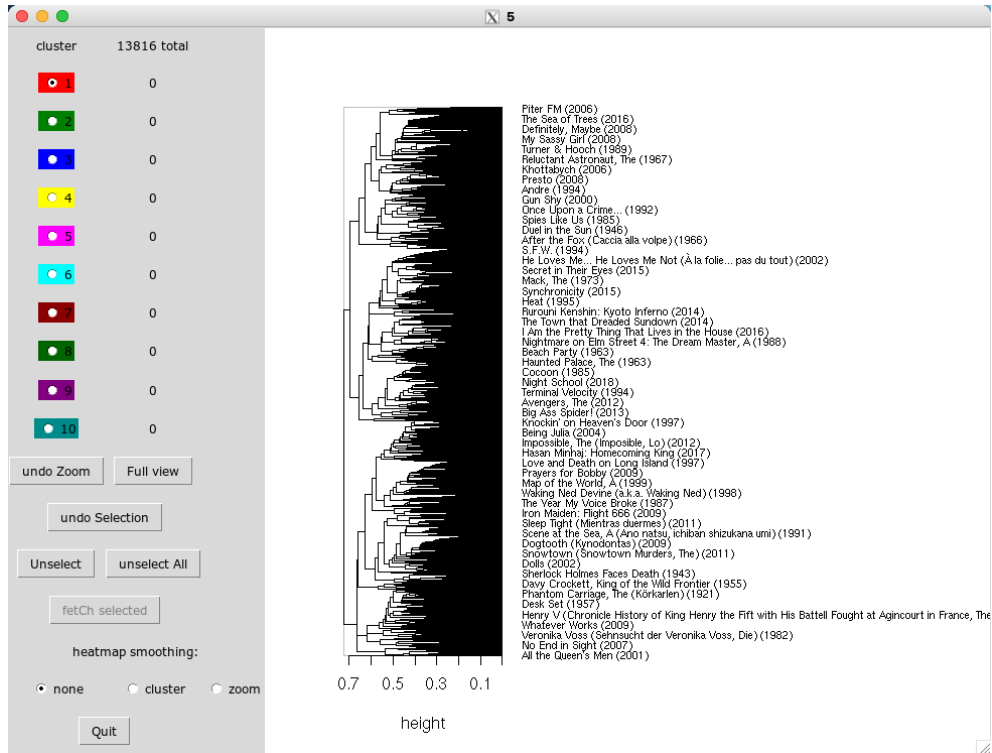   packages idendro and idendr0', *Journal of Statistical Software, Articles* **76**(10), 1–22.

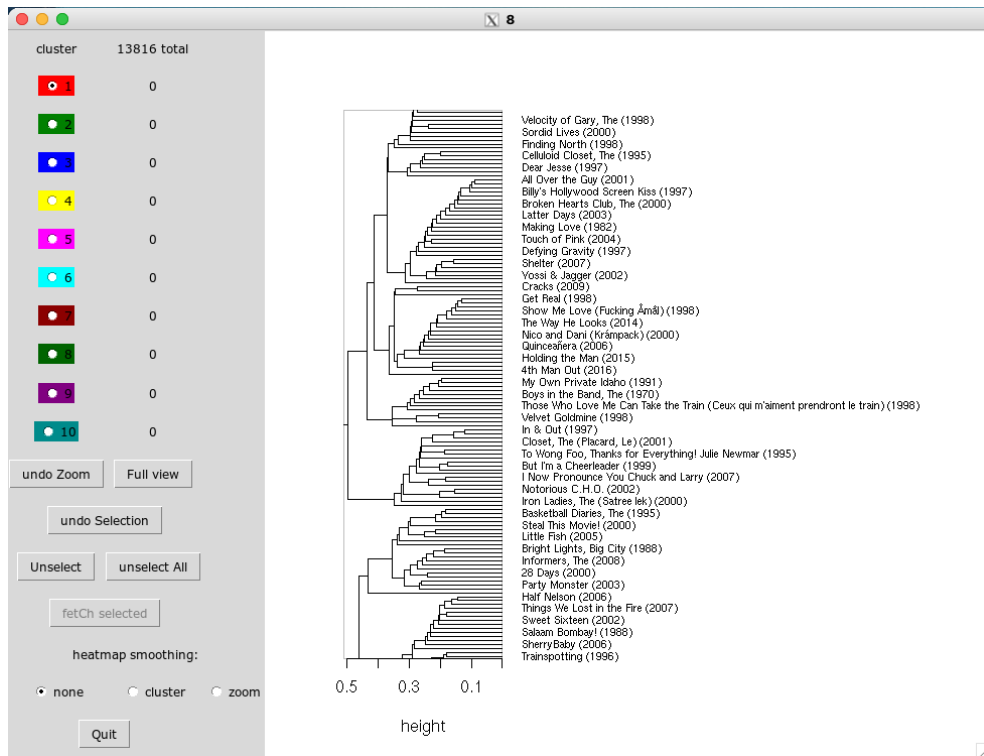Figure 3: Default view in `idendr0` of the movies data with adjusted margins.

Figure 4: Zoomed view in `idendr0` of the movies data.

4

Figure 5: Default view in `collapsibleTree` of the movies data.