



## STEP 2c CODING THE INTERVIEWS – CLUSTERING

The coding procedure resulted in many different labels for the causes and effects, many of which overlap in meaning. Even the general concepts (e.g. “economic stress”) were quite varied. The procedure for clustering these labels (including both the general and specific parts of the label) into common groups with their labels was a three-step process based on assigning to each of the original labels an embedding. An embedding is a numerical encoding of the meaning of each label (Chen et al., 2023) in the form of a point in a space, such that two labels with similar meaning are close in this space. For any two such vectors, a measure cosine similarity can be calculated representing the approximate similarity in meaning between the labels which they encode:

1. **Inductive clustering.** First, we grouped the labels into clusters of similar labels using the `hclust()` function from the `stats` package of base R (Team, 2015).
2. **Labelling.** We then asked an AI to find distinct labels for each cluster. We also manually inspected these labels with regard to the original labels within each cluster and adjusted some of them.
3. **Deductive clustering.** We then discarded the original clustering, created embeddings for the new labels, and formed a new set of clusters, one for each of the new labels, assigning each original label to one of the new labels, the one to which it was most similar, providing the similarity was at least higher than a given threshold. This additional deductive step ensures that each member of each new cluster is sufficiently close in meaning to the new cluster label, rather than just to the other members of the cluster.

After each sub-step, we checked the AI’s results to ensure that the instructions were being followed correctly and, if they weren’t, the instructions were tweaked or rewritten and tested again to ensure quality and consistency.

### References

Team (2015). *R: A Language and Environment for Statistical Computing*,.