Measurement, Scaling, & Dimensional Analysis
Prof. Adam Enders
ICPSR, 2019

Assignment 5: Multidimensional Scaling

I start my analysis by using the "dist" function to calculate Euclidean distances from the original $n \times k$ matrix of n survey respondents' affective reactions to m salient political figures and parties. This results in a $k \times k$ square, symmetric matrix $\mathbf{\Delta}$ of distances that we can conceive of as dissimilarities. Since We are not employing direct (dis)similarity judgements (and, probably even if we were), it is likely not theoretically appropriate to conceptualize the dissimilarities data as interval (or ratio) level. Rather, we'll treat these distance measures as ordinal level, where the output distances will be weakly monotonic functions of the input dissimilarities data.

The Stress_1 values for two dimensional solutions (optimal for graphically presenting results) for each year (presented below) all suggest excellent fit of the thermometer distance data to a two-dimensional nonmetric MDS model. Indeed, the largest Stress_1 value of 0.010, which is associated with the 2016 data, is still substantially smaller than the 0.05 rule of thumb for interpreting the model fit as "excellent." So far, so good.

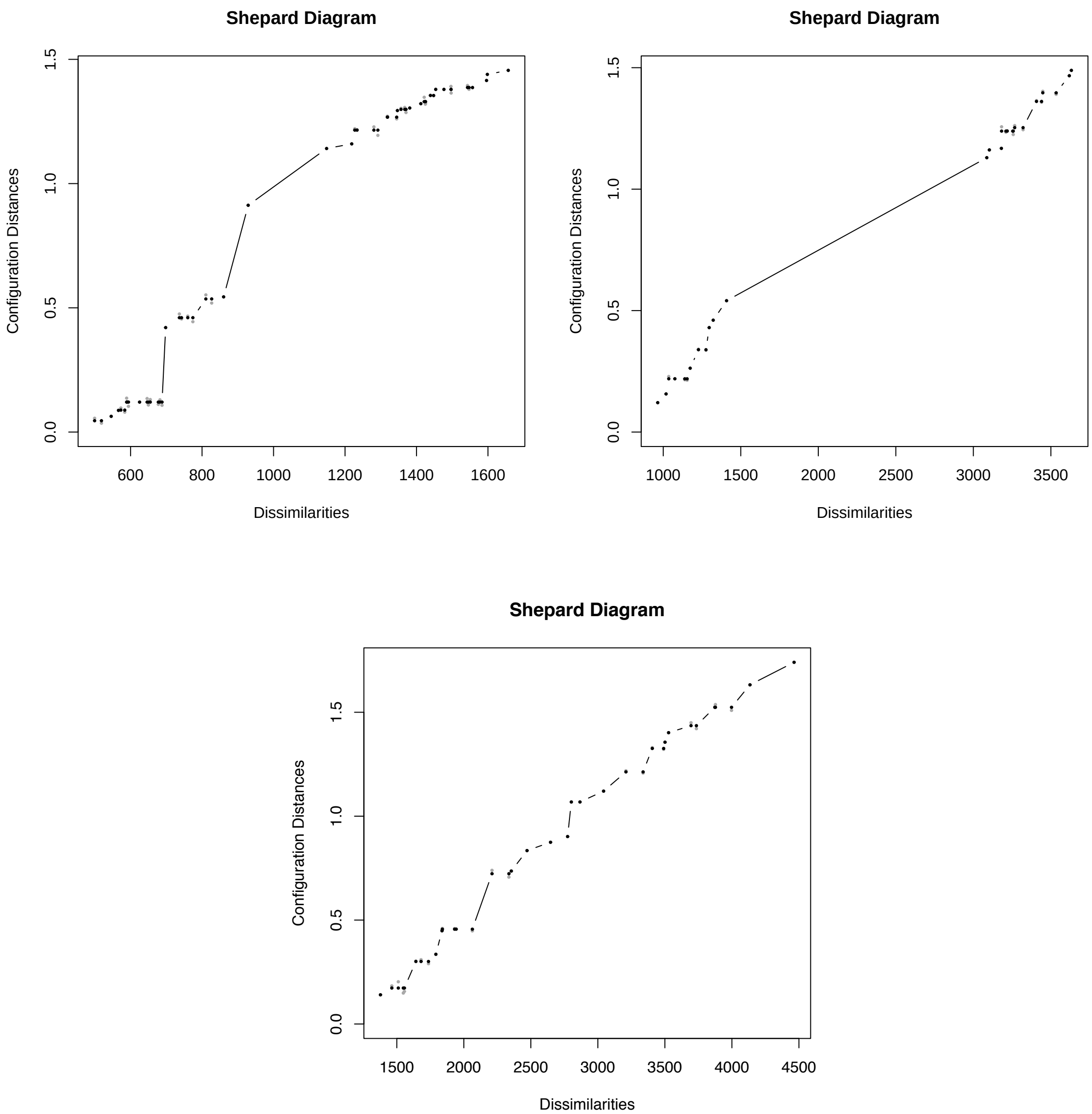
```
> nonmetric08$stress
[1] 0.009219182
> nonmetric12$stress
[1] 0.005787854
> nonmetric16$stress
[1] 0.009985961
```

Next, I examine Shepard plots of the scaled (or, output) distances against the input dissimilarities. With nonmetric MDS analyses, we're hoping to find (weakly) monotonic patterns in the plot. In each case, there are only very few barely perceptible deviations from monotonicity. This makes sense given the low Stress_1 values we observed above. Note, however, the relatively low range of scaled distances associated with a relatively large range of input dissimilarities in the first (2008) and second (2012) plots. This is disconcerting. Though the relationships are monotonic, the patterns we observe suggest that in both plots there are two clusters of objects with very little variability within clusters and substantial dissimilarity between clusters. This is the type of pattern that frequently accompanies degenerate solutions.

The Spearman rank order correlations between the input dissimilarities and output distances are extremely high in all cases. High correlations will, generally, go hand in hand with low Stress_1 values. However, as we just saw with the Shepard plots, Stress_1 and these correlations alone cannot determine whether the resultant MDS configuration will really be *useful*.

Figure 1: Shepard plots of output (scaled) distances from NONMETRIC MDS analyses against input dissimilarities with a monotonic curve.

,



```
> cor(nonmetric08$delta, nonmetric08$confdist, method = "spearman")
[1] 0.987518
> cor(nonmetric12$delta, nonmetric12$confdist, method = "spearman")
[1] 0.9797482
> cor(nonmetric16$delta, nonmetric16$confdist, method = "spearman")
[1] 0.990991
```

Finally, I plot the nonmetric MDS configurations for each year in an attempt to both lend some substantive interpretation to the geometric structure produced by the analyses, and to more fully consider the potential for degenerate solutions. It is pretty clear from each of the plots that something like a liberal-conservative ideological dimension exists within the data, and in each case it is nearly parallel to the first MDS axis (which we should not substantively interpret!). Vertical variability in the stimulus points might be interpreted as something like average "popularity." For instance, Condoleezza Rice was more popular among both Democrats/liberals and Republican/conservatives than any of the elected officials or parties in 2008, and certainly more popular than Rush Limbaugh who is low on that dimension (this can be confirmed by checking the averages for the original thermometers). Similarly in 2016, Gary Johnson was despised by neither the average Democrat, nor the average Republican. In some sense, then, his average popularity was higher than any of the more salient parties or associated political figures. The dendrograms produced using the results of hierarchical cluster analyses tell similar stories.

Despite being able to some information from the MDS configurations, the 2008 and 2012 nonmetric MDS results would probably be classified as degenerate solutions by most MDS experts. In 2008, the only variability in the vertical direction is due to Condoleezza Rice and Ruch Limbaugh. In fact, if you remove them, the clusters of Republican and Democratic Party stimulus points become even tighter. The plot really doesn't tell us all that much about the psychological proximity between the stimuli other than the fact that Democrats and Republicans are different. More or less the same can be said of the 2012 solution. All of the Republican stimuli are very close to each other on one side of the configuration, and all of the Democratic stimuli are very close to each other on the other side. The 2016 solution seems to be fine.

Even though the 2008 and 2012 MDS solutions are technically problematic, we might still be able to make a useful substantive inference about the psychology of the American electorate in those years. The major reason why we wouldn't observe within-party variation in inter-point distances is, of course, because people didn't perceive stimuli associated with a given party to be all that different. Furthermore, we would observe a substantial distance between clusters of party stimuli because people who had positive feelings toward one party and very negative feelings toward the other party. Most of us would probably call this set of conditions "polarization." In other words, the degenerate solutions are telling us something potentially useful about Americans' perceptions of and affective attachments toward major political stimuli: we love our ingroup, and despise our outgroup.

Treating the data metric and re-estimating the models doesn't seem to change any of the inferences we've already made under assumptions of nonmetric data. Stress$_1$ values have

Figure 2: MDS configurations for 2008, 2012, and 2016 ANES data.

,

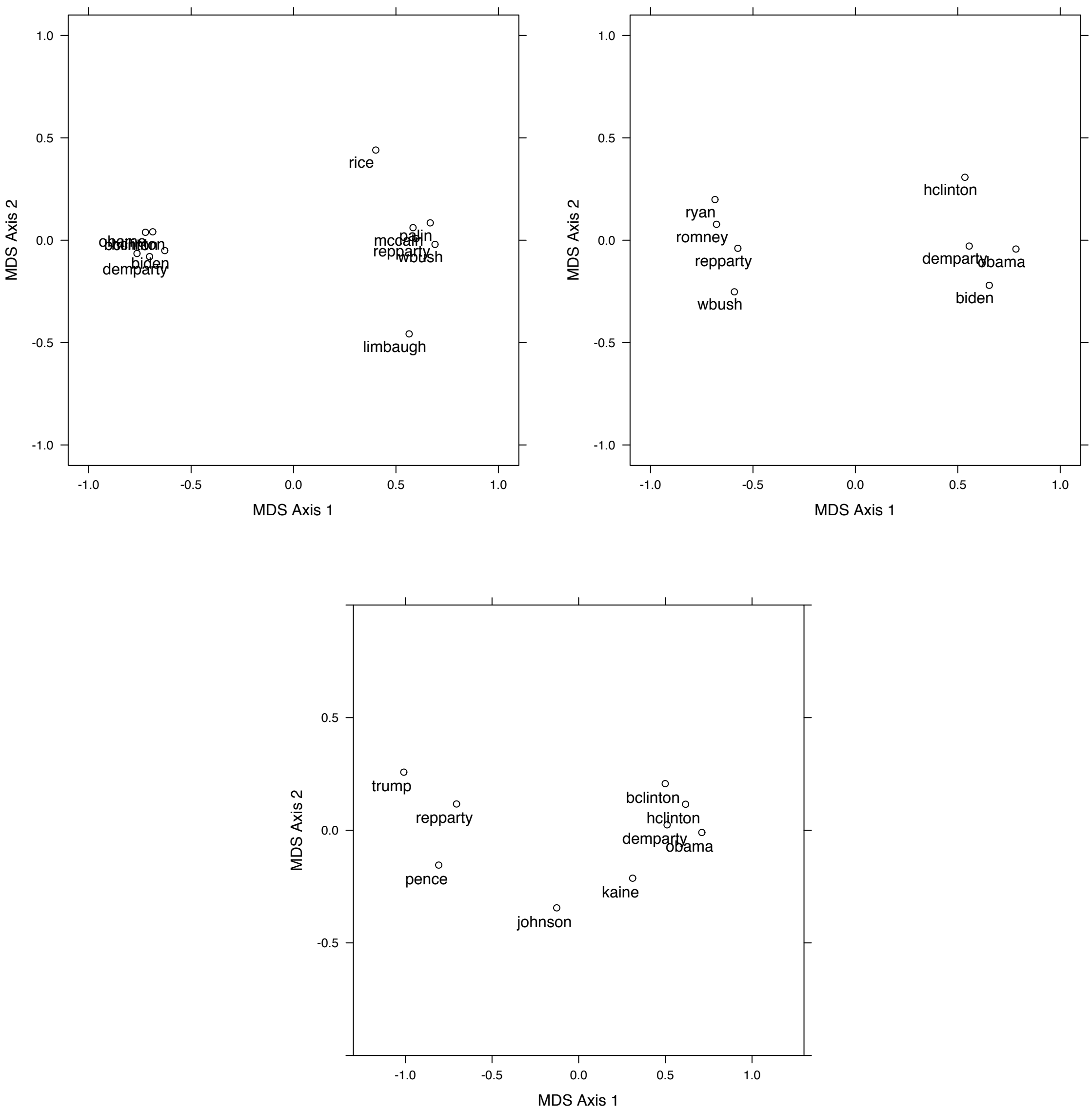
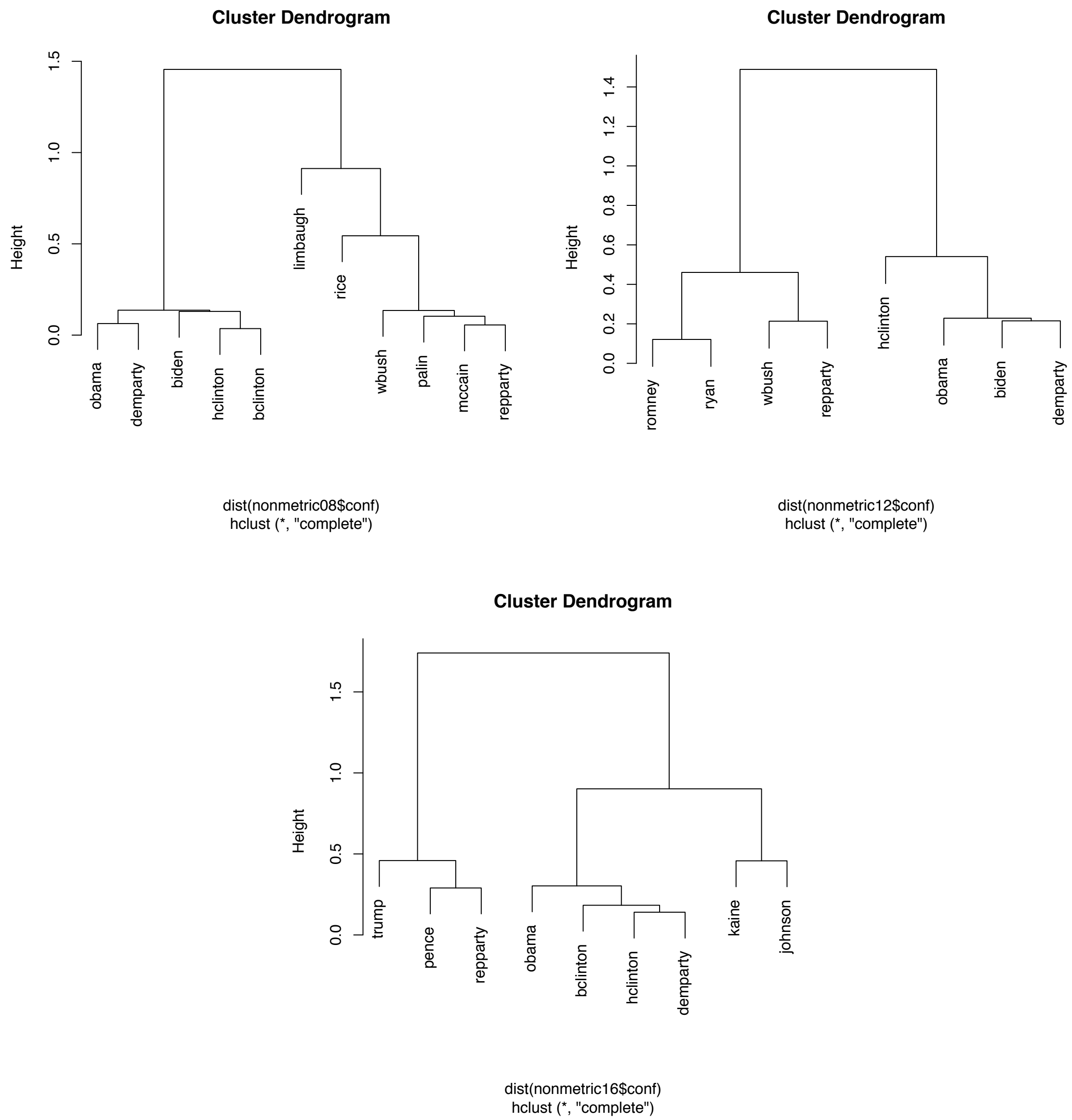


Figure 3: Dendrograms from hierarchical cluster analyses of Euclidean distances computed political stimuli via the nonmetric MDS coordinates.

,



increased, as they almost always will, because we've imposed more constraints on the procedure. However, they still suggest "excellent" model fit by conventional standards. The Shepard plots for 2008 and 2012 still suggest that we might have degenerate solutions for those years. And, the plotted configurations (not pictured) look substantively identical to the nonmetric configurations. In this case, we could probably consider the feeling thermometer dissimilarities measures to be interval-level, though we really don't acquire any more information by doing so.

```
> metric08$stress
[1] 0.03298808
> metric12$stress
[1] 0.01894932
> metric16$stress
[1] 0.02689223


> cor(metric08$delta, metric08$confdist, method = "pearson")
[1] 0.998049
> cor(metric12$delta, metric12$confdist, method = "pearson")
[1] 0.9994234
> cor(metric16$delta, metric16$confdist, method = "pearson")
[1] 0.9986696

> cor(nonmetric08$confdist, metric08$confdist)
[1] 0.983916
>
> cor(nonmetric12$confdist, metric12$confdist)
[1] 0.9956645
>
> cor(nonmetric16$confdist, metric16$confdist)
[1] 0.9946522
```

Figure 4: Shepard plots of output (scaled) distances from METRIC MDS analyses against input dissimilarities with a monotonic curve.

,