Simulating Clusters

Illustrate what happens in randomly clustered data.

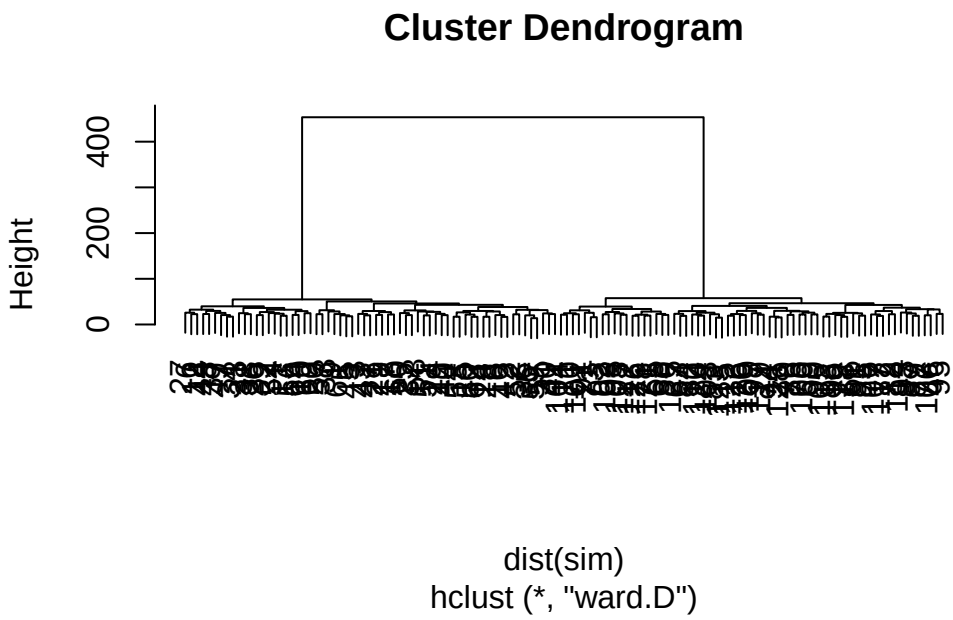
We want to better understand the output we obtained from our actual data. So, see what some 'idealised' scenarios for clustered data would look like.

Simulate our data

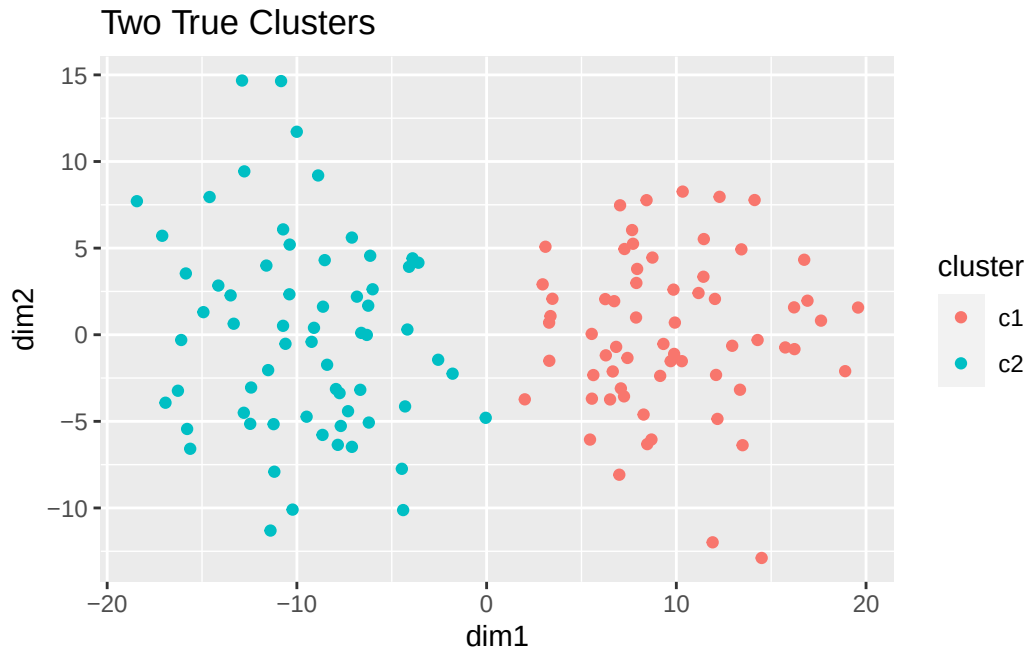
What if there are 2 clusters?

First, we get a noticeable elbow in the elbow plot:

Our dendrogram clearly shows four clusters, too:



Third, when applying MDS to visualise, we get a nicely separated plot:

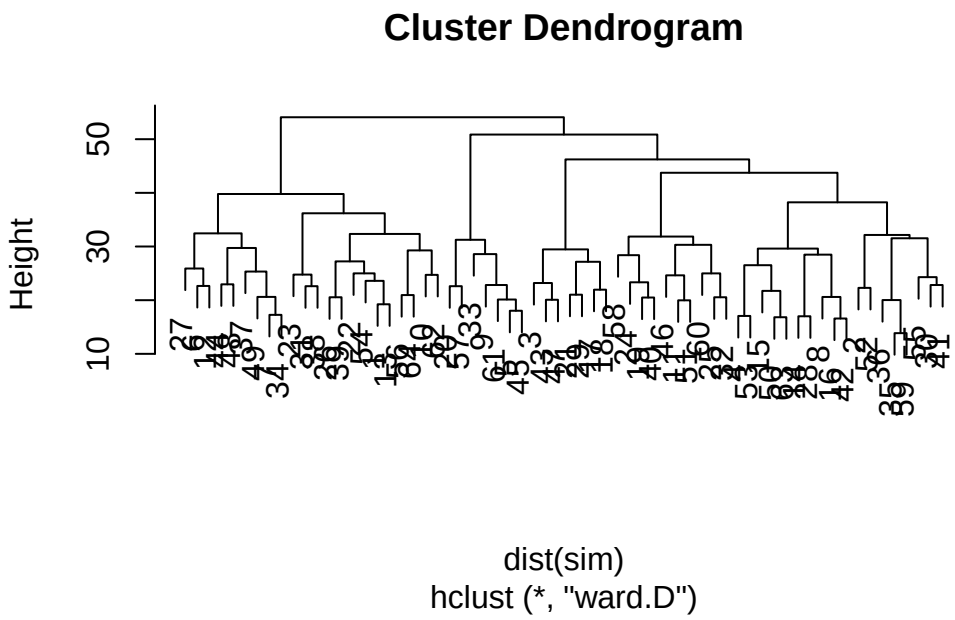


What if there is no cluster structure?

First, we don't get an elbow. Instead the within-SS decreases monotonically, showing slight slowing down the further we go - maybe exponential decay?

```
plot(1:20, within, main = "Elbow Plot One Cluster", xlab = "n Clusters", ylab = "Total Within
```

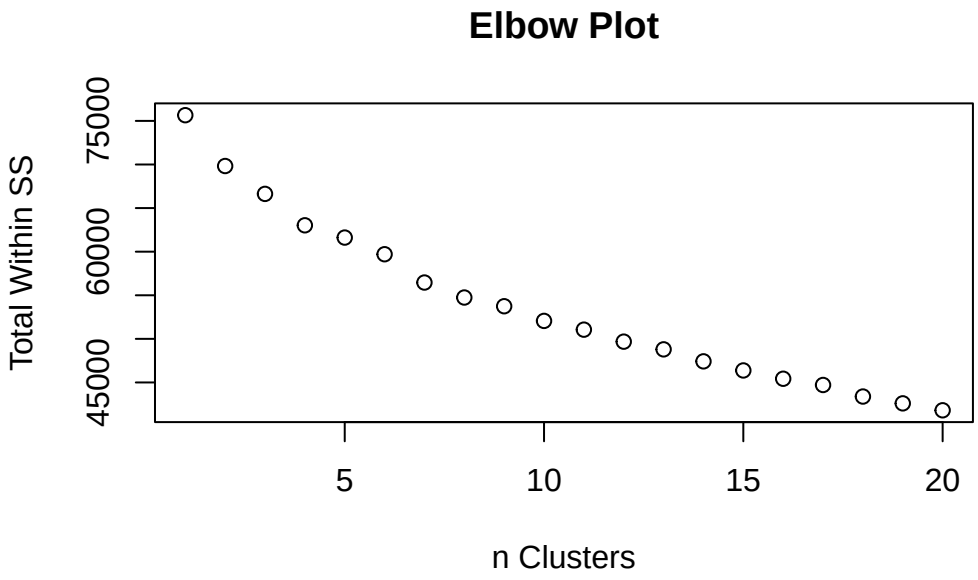
Ward's dendrogram shows a confused multi-layer structure with lots of 'clusters' within clusters with relatively low distance (height) between them:



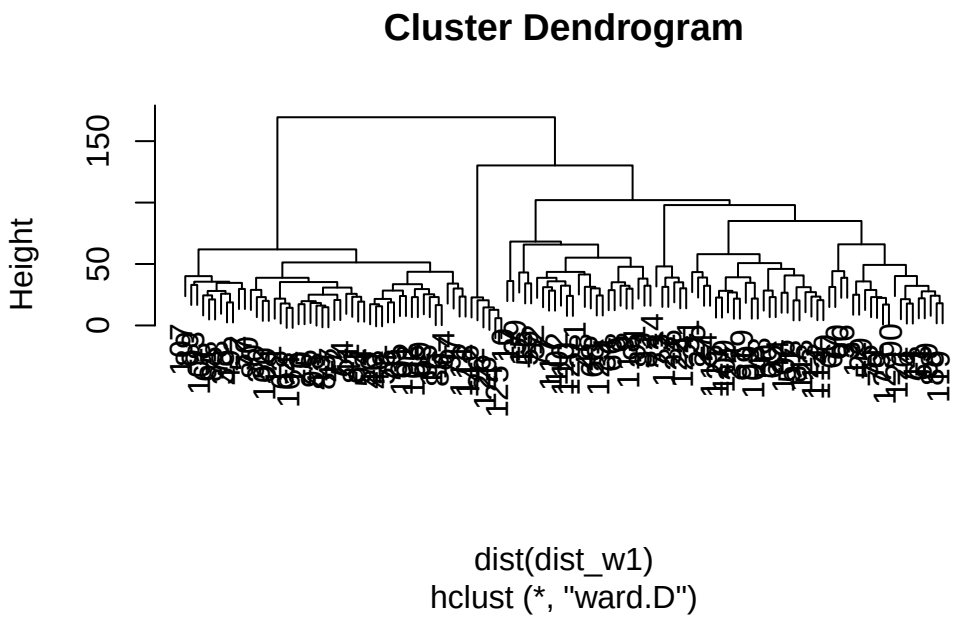
Similarly, when applying MDS our spatial structure is implausible:

How does our real data look?

First, our elbow diagram lacks any clearly identifiable elbow:



Second, our dendrogram looks suspiciously similar to the random case:



Third, the mds-plot looks conspicuously chaotic: