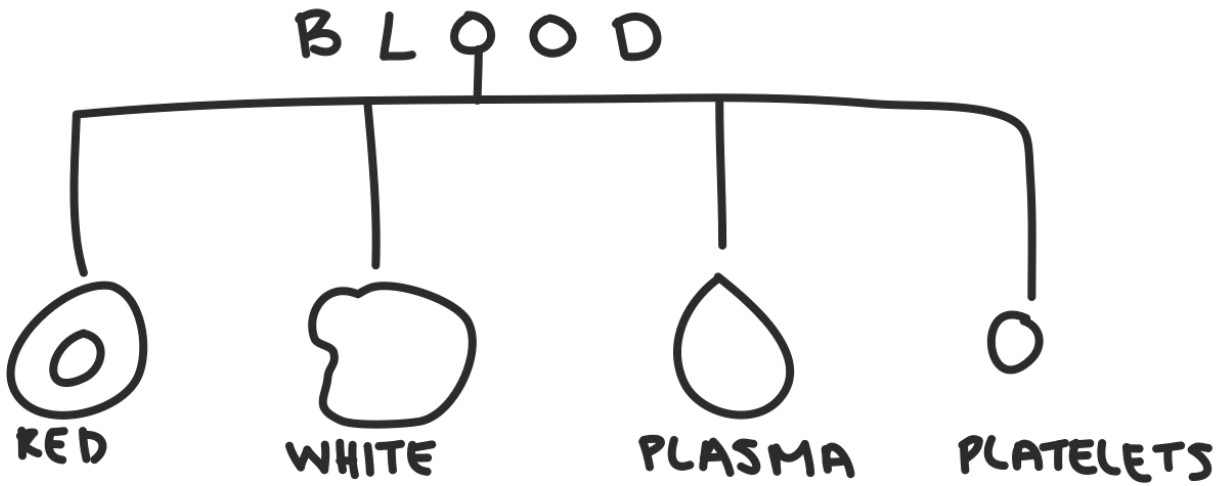
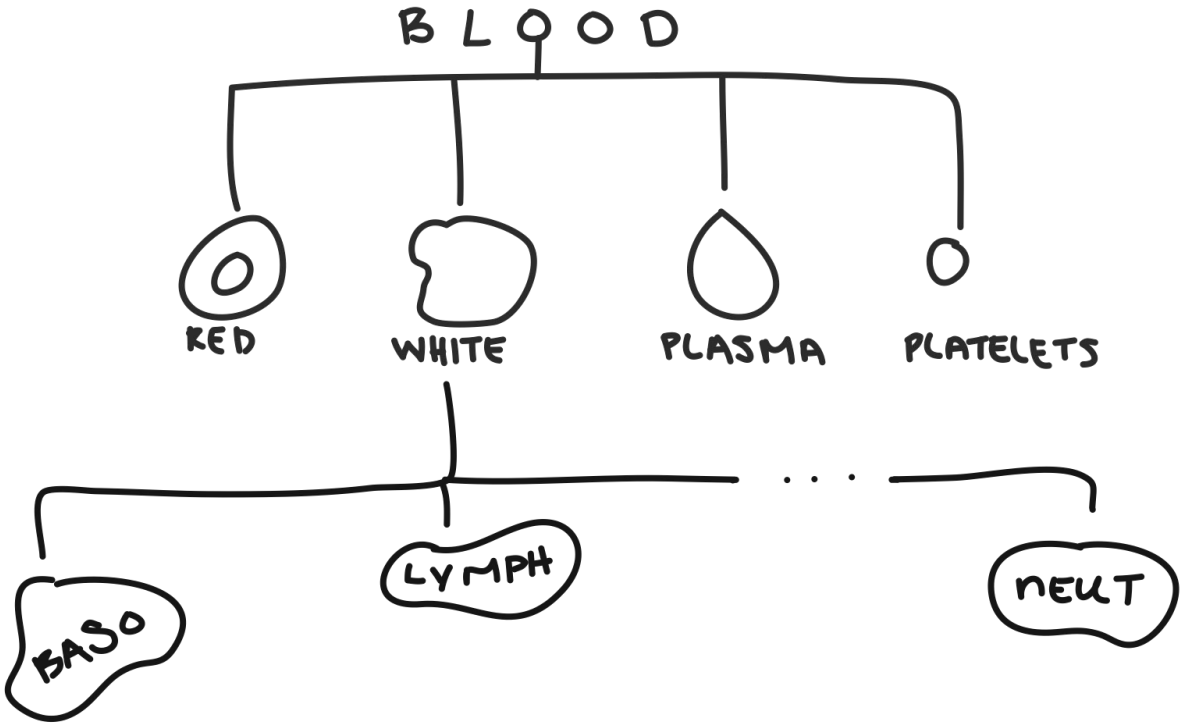
Hierarchical (Agglomerative) Clustering

Chelsea Parlett-Pelleriti

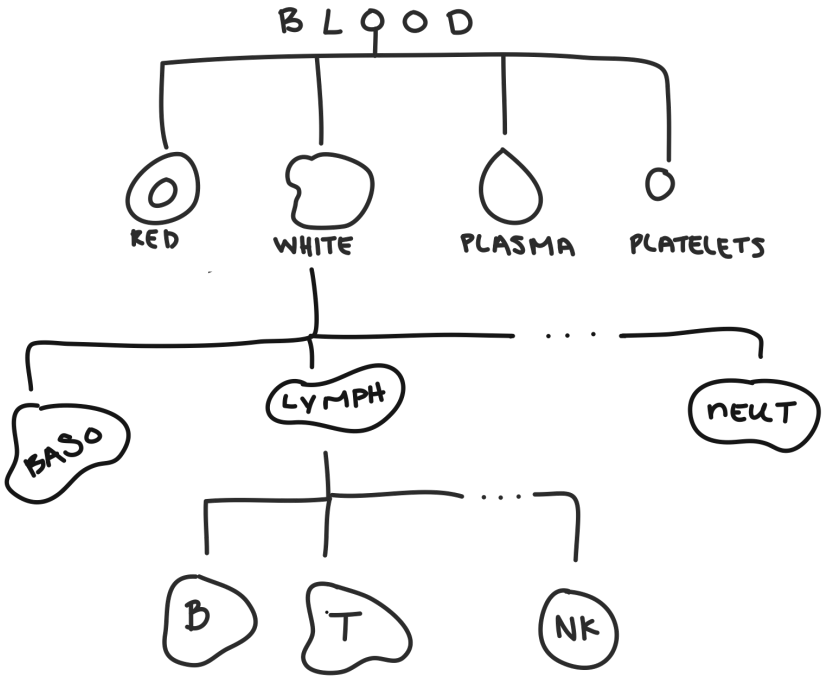
HAC



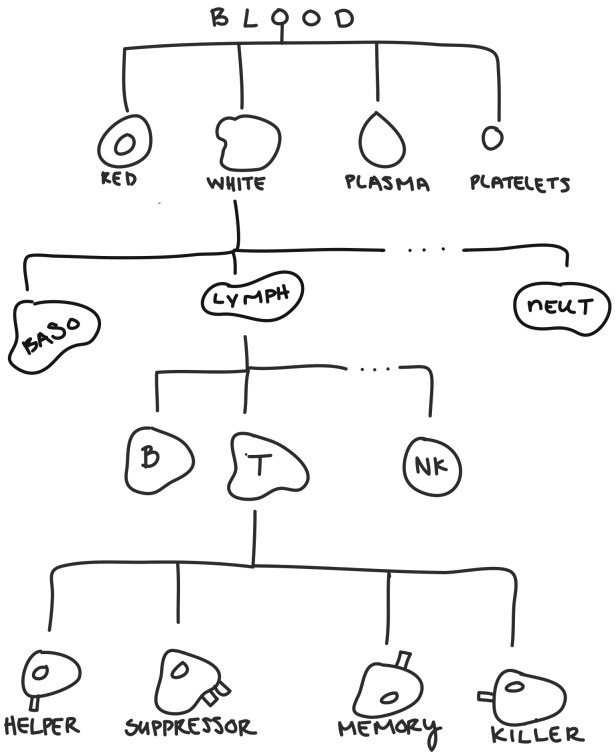
HAC



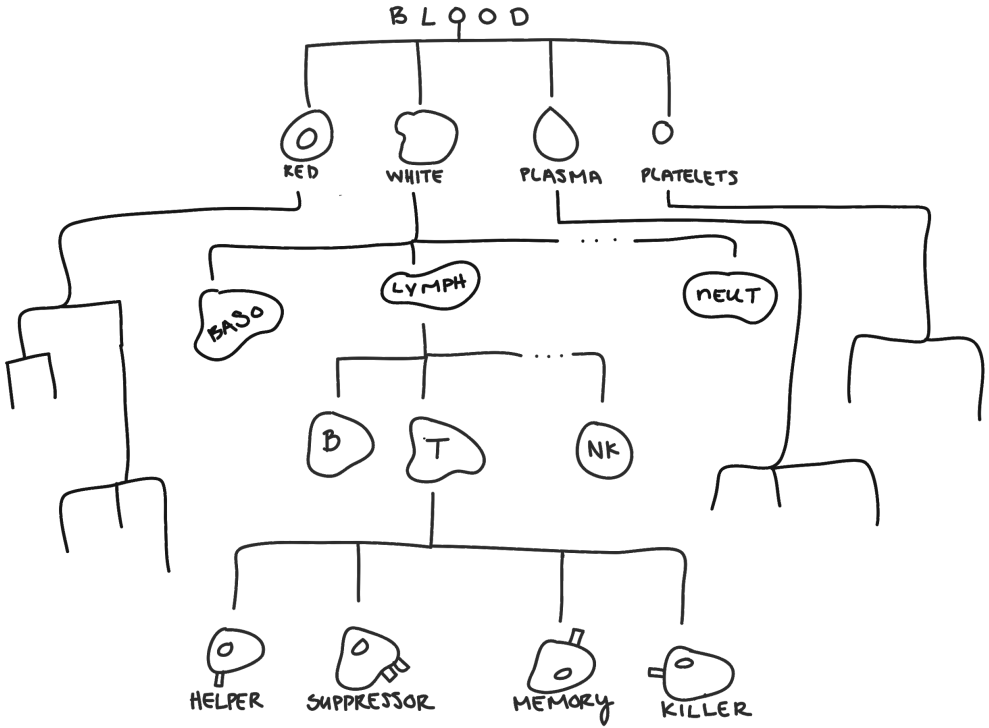
HAC



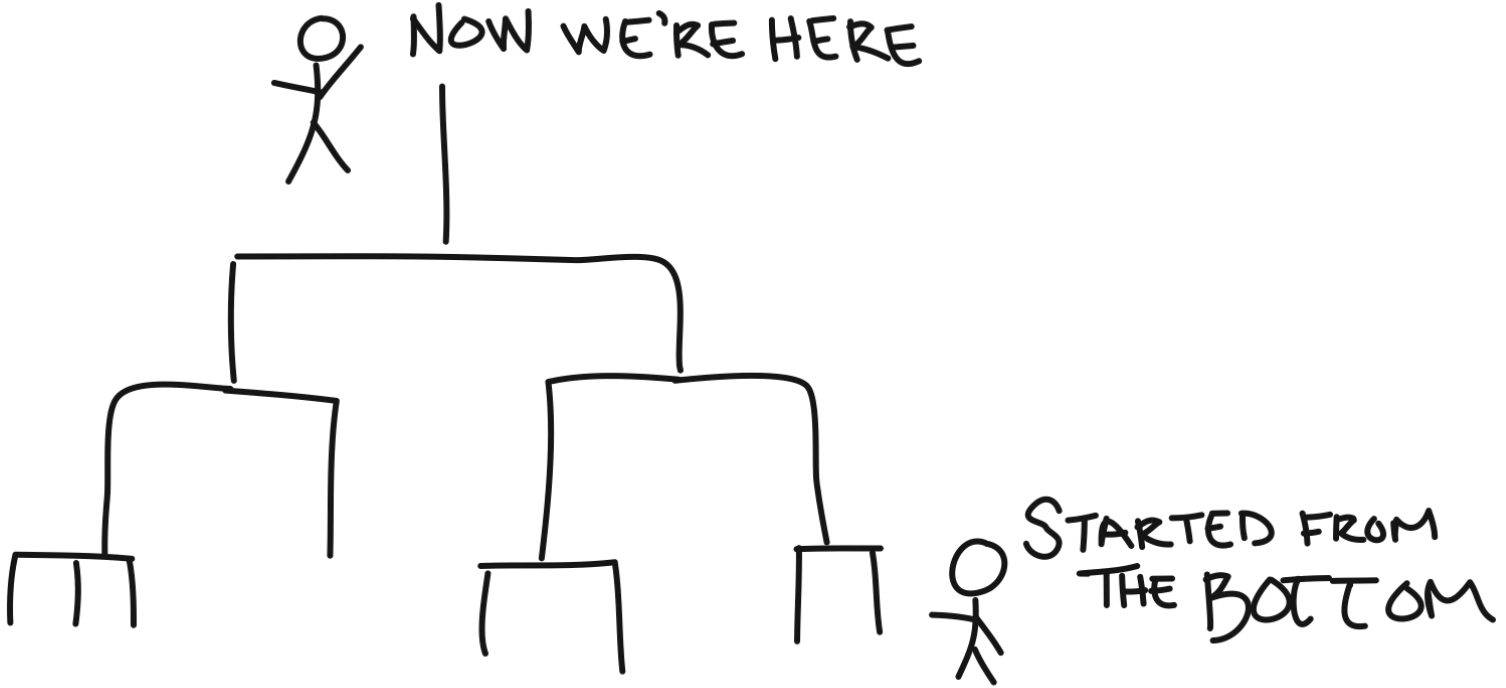
HAC


BLOOD
RED
WHITE
PLASMA
PLATELETS
BASO
LYMPH
NEUT
B
T
NK
HELPER
SUPPRESSOR
MEMORY
KILLER

HAC


BLOOD
RED
WHITE
PLASMA
PLATELETS
BASO
LYMPH
NEUT
B
T
NK
HELPER
SUPPRESSOR
MEMORY
KILLER

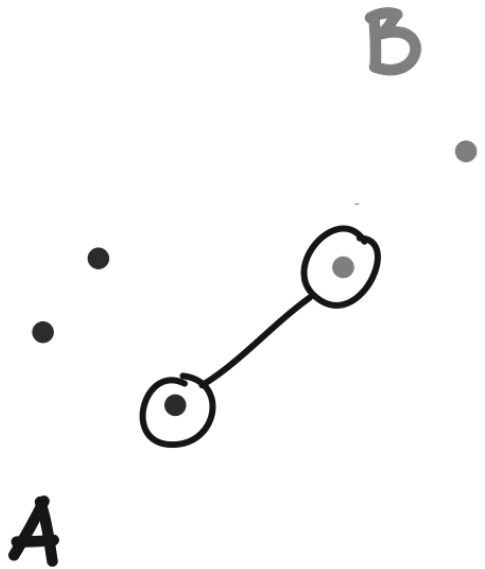
HAC



Distance Metrics

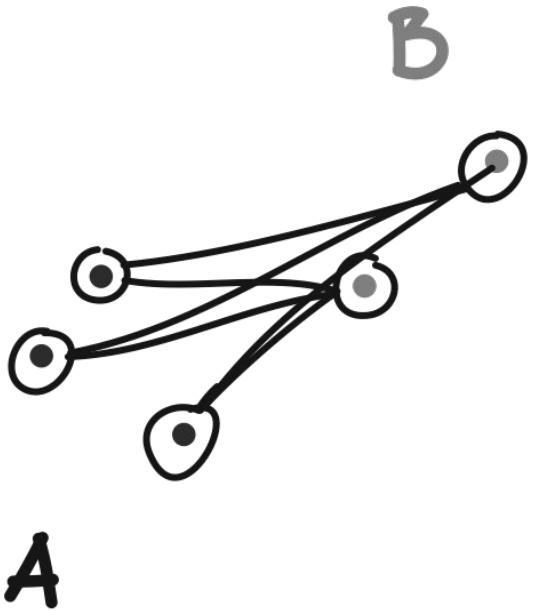
- **Euclidean**: Continuous Data
- **Manhattan**: High Dimensions
- **Hamming**: Categories
- **Cosine**: Word Counts

Linkage Criteria



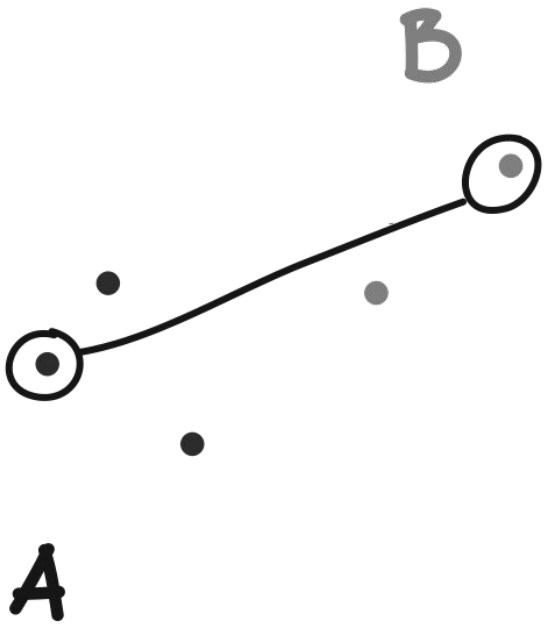
Single

Linkage Criteria



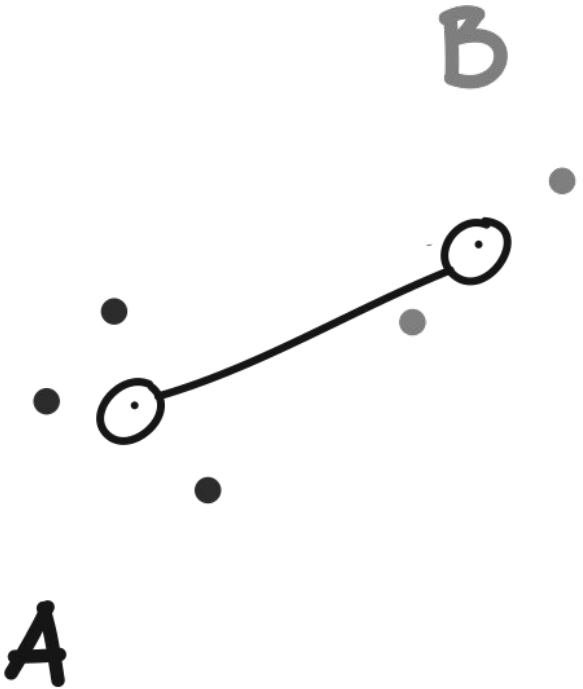
Average

Linkage Criteria



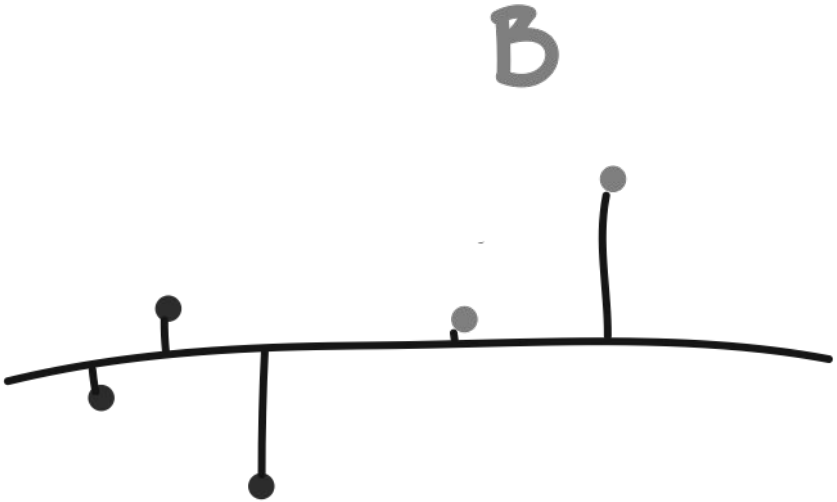
Complete

Linkage Criteria



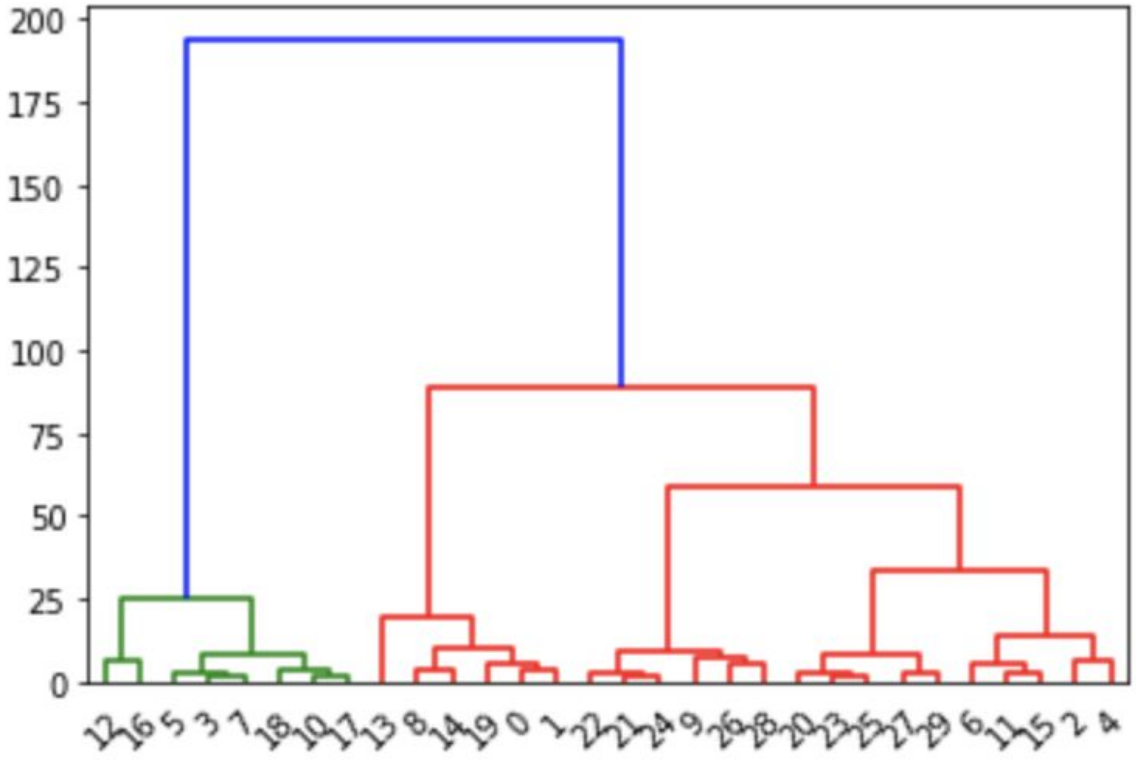
Centroid

Linkage Criteria

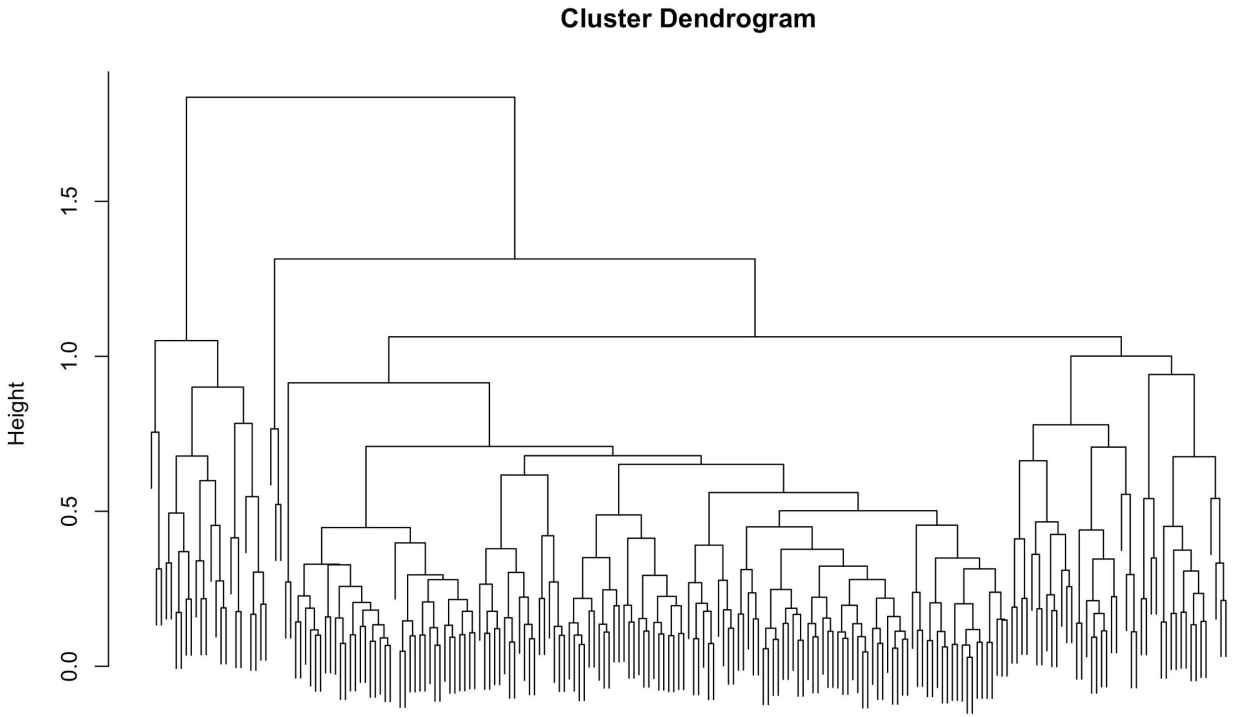


Ward's

Reading a Dendrogram



Reading a Dendrogram



HAC

Pros

- Flexible # of clusters
- Model relationship between clusters (hierarchy)
- Flexibility with linkage

Cons

- Very Slow O(n^3)
- Cannot un-merge clusters

My Master's Thesis