

1.

K value	SSE	AIC
2	173.33	49.33
3	265.42	739.42
4	399.27	1031.27
5	478.83	1268.83
6	604.34	1552.34

2. The best k value we found is k=3. Looking at the external metrics (clusters) we found that k=3 finds is able to distinctly separate protein synthesis and protein degradation. Although we found that k=3 still cannot separate TCA cycle and oxidative phosphorylation too well. We thought that k=3 is still the best k because k=4 has the same issue, but has a higher SSE and AIC and thus is ultimately not as good as k=3.

3. They used hierarchical clustering. They justified by noting that genes most closely clustered together were actually just different sequences representing the same expression. More generally, genes in the same bigger clusters were genes that had similar function. They talk about hierarchical clustering representing a natural way of looking at complex data, first noticing the large scale features and then looking at the details.