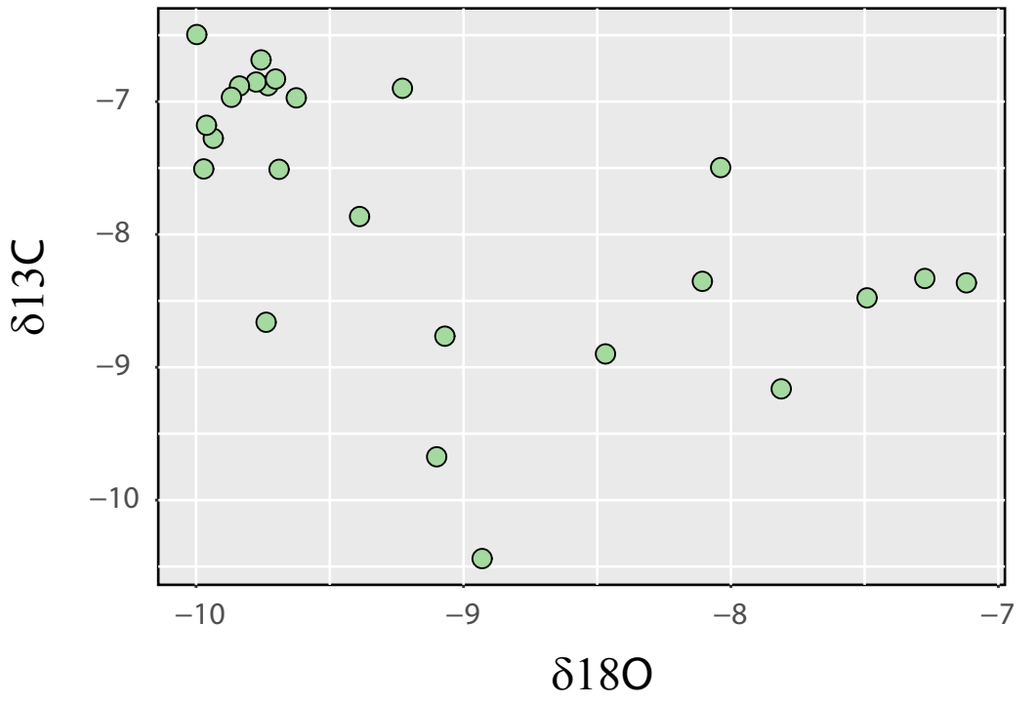
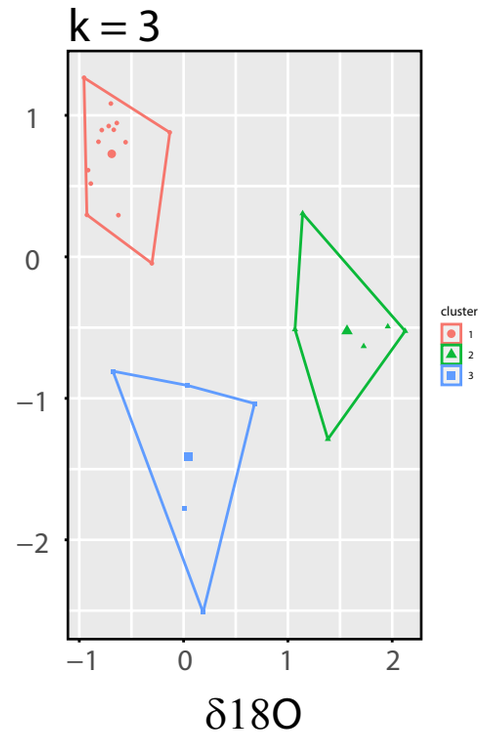
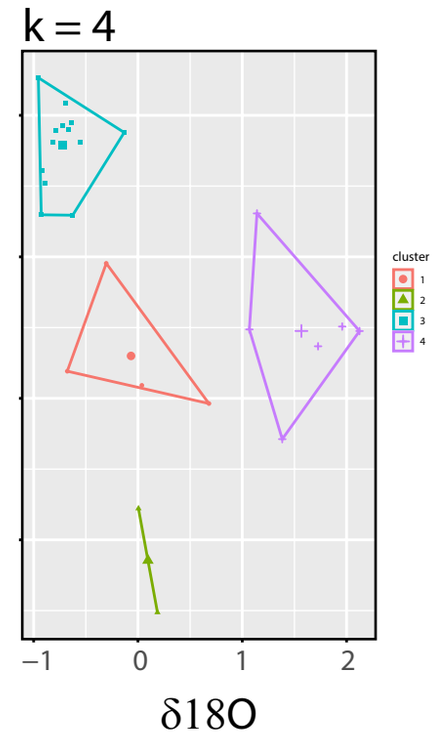
Calcite Isotope K-Means Clustering: Fayetteville Green Lake

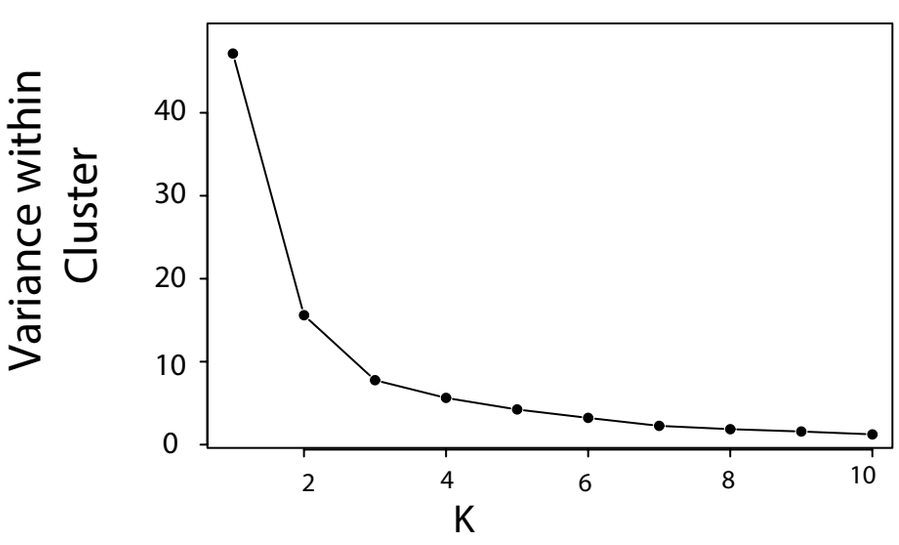
Micah Wiesner

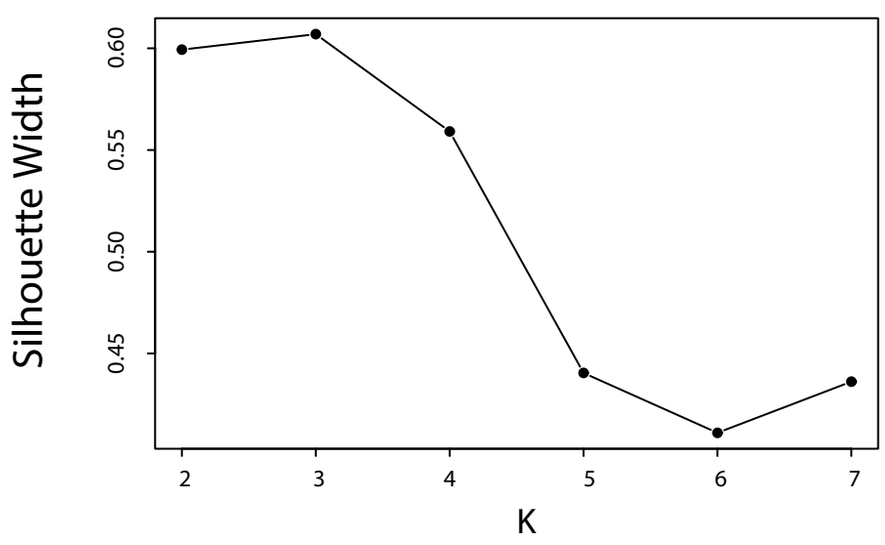


Prior to using the algorithm our eye-naturally intuits a top-left cluster though it is unclear if there is also a central and bottom cluster or more









Analyzing the variance within each cluster (left) and the 'silhouette width' allows the optimal number of clusters to be chosen. 'Silhouette width' takes into account the cohesion between each point of a given cluster and the separation between points of different clusters. Although the variance within cluster doesn't distinguish between the third and fourth k-value very well, the silhouette width shows that k = 3 or k =2 is an optimal number of clusters. Geologically it still unclear how this can be interpreted.

**It should also be noted that a third technique gap statistic could also clarify which k value is best.

In conclusion: Statistically, it is possible that calcite sourced from Fayetteville Green Lake has two or three distinct 'regimes' of formation