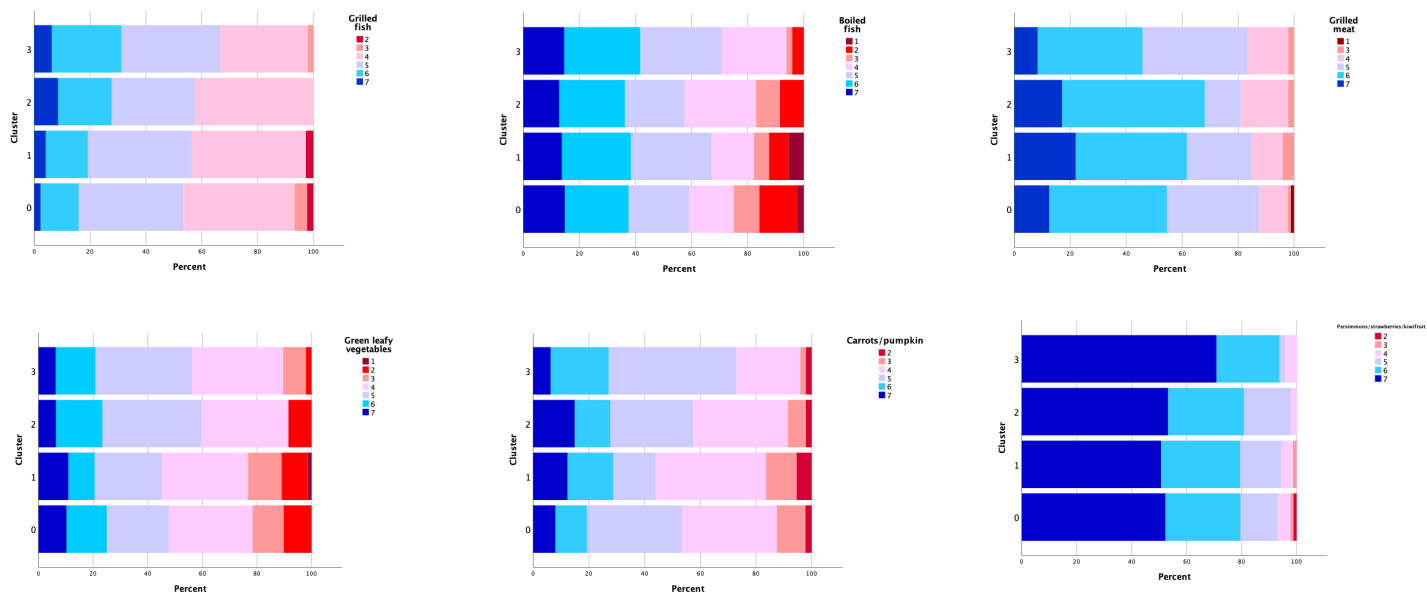
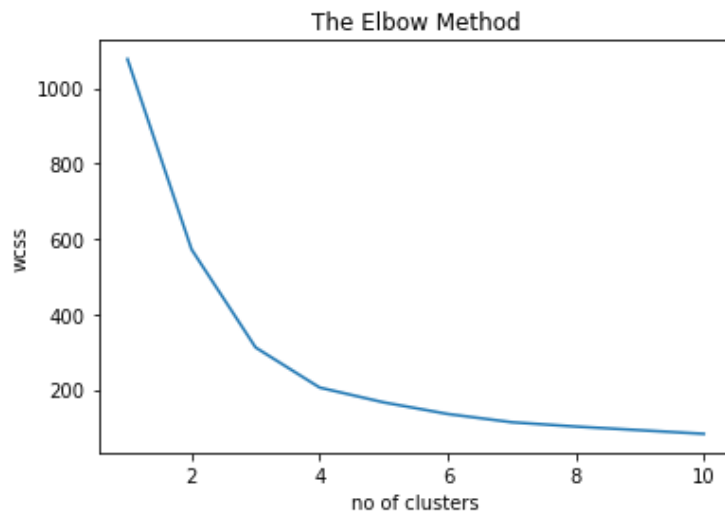




Supplemental figure 1. The histograms compare the % cluster for all samples from the BDHQ dataset. The histogram shows the proportion of each variable of the BDHQ dataset in the clusters. Bars represent a stacked histogram with colors corresponding to different factors of features. The histograms represent overlapping cluster sequencing reads (numbered above stack) centered at each position. The number of layers in a stack varies significantly (from high-frequency intake to low-frequency intake; 1 for the upper stack, dark red, and 9 for the lower stack, dark blue).



Supplemental Figure 2 shows the Elbow method, a data analysis and machine learning technique used to determine the optimal number of clusters in a dataset. The horizontal axis represents the number of clusters, and the vertical axis represents the sum of the squared distances between each data point and its assigned cluster centroid. The Elbow method determines the optimal number of clusters by identifying the point on the plot where the rate of decrease in within-cluster sum of squares (WCSS) slows down significantly, forming an "elbow." In this case, the optimal number of clusters determined by the Elbow method is 4, as indicated by the plot.