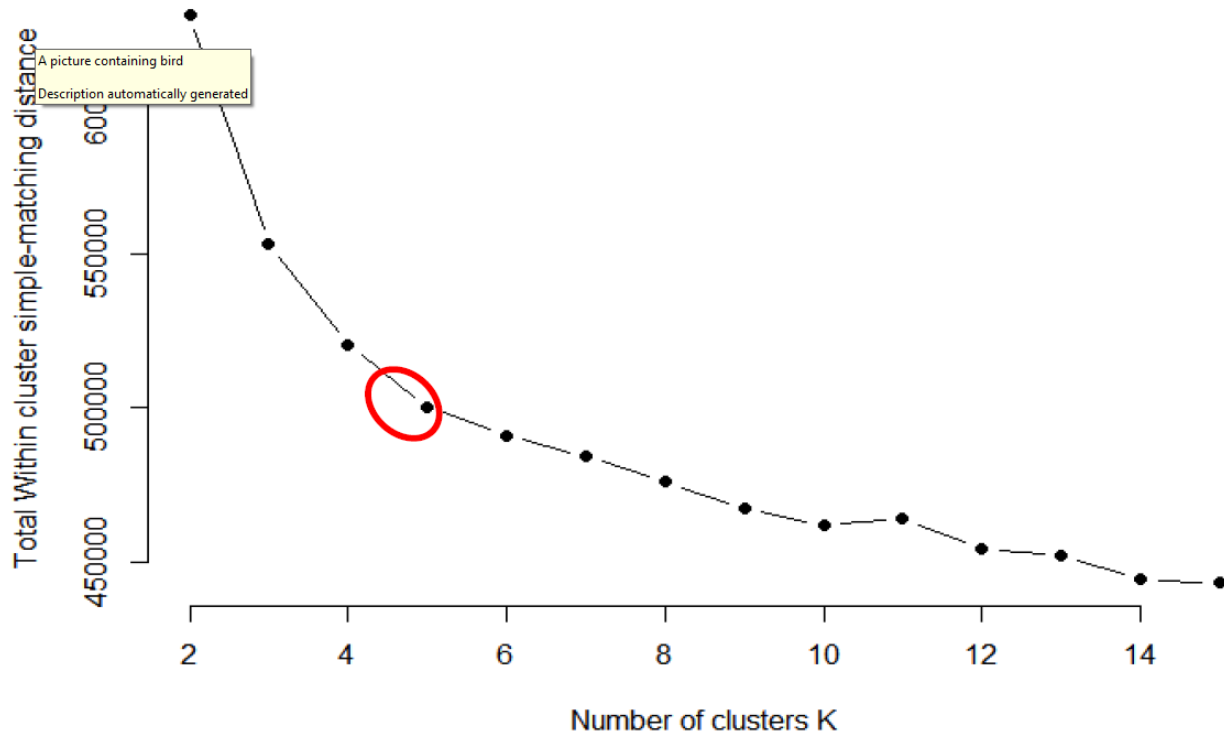# SUPPLEMENTARY FIGURE 1

Figure S1. Selecting the number of clusters (k) for further analysis.

The elbow method was used to identify the best number of suggested clusters. It is a heuristic method of interpreting consistency within cluster analysis and enables finding the appropriate number of clusters for the data set. If the line graph bears a resemblance to an arm, then the elbow (the point of curve inflection) is the "best" value for k. That is an indication that the clustering model fits best at that point. In this case, it is the encircled point (k=5). The y-axis shows the function used to obtain the within-cluster simple-matching distance for each cluster; i.e., by the calculation of the sum of errors for the difference values of k to select the optimal number of clusters, based on the largest sum of error difference between potential k-values. Additionally, k=11 was worth observing, given that its value k $\geq$10, which is unusual for k in an elbow diagram. However, cluster 11 does yield a lower simple matching distance (466000) than cluster 5 (500000) on the y-axis. Lower simple matching distance leads to lower variability and a higher distribution of participants with periodontitis in all the clusters.