MATH 437/537 ASSIGNMENT IX Due April 8, 2020

The dataset `www.mines.edu/~wnavidi/math437537/cereal.csv` contains data on 43 brands of cereal from three manufacturers: General Mills (G), Kellogg's (K), and Quaker (Q). For each brand, the number of calories, and amounts of protein, fat, sodium, fiber, carbohydrates, sugar, and potassium per serving are given.

1. Use logistic regression to classify each cereal as to manufacturer. Construct the confusion matrix. Note: The default maximum number of iterations of the fitting procedure is 100. You may need to set the `maxit` option to increase this number.

2. Use cross-validation (leave one out method) to estimate the misclassification rate of the logistic regression method. Construct the confusion matrix.

3. Classify each cereal as to manufacturer, using the k-nearest neighbors method, using $k = 1$. Construct the confusion matrix. Explain why every item is classified correctly.

4. Use cross-validation (leave one out method) to estimate the misclassification rate of the k-nearest neighbors method for $k = 1$, $k = 3$, and $k = 7$.
   (a) Construct the confusion matrix for each value of k.
   (b) Explain why the number of items classified as "Q" decreases as k increases.