Problem 1

Display your output of plot Kmeans(). Does your plot match your expectations?

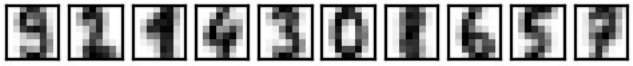

Figure 1: plotKmeans

The plot was easier to read than I thought it would be! I am quite impressed. I thought that I might have to squint a bit more.

Problem 2

In this assignment, you implemented k-means through a Euclidean distance metric. Describe other distance metrics that can be used and how they cluster inputs.

We could have also used distance metrics such as "Mahalanobis Distance." I found this metric online. It also calculates the distance between two points in a multivariate space, but it does it by constructing a probability distribution from the sample and measuring how many standard deviations away a point is from the mean. This would allow the model to take covariance into account–allowing more information about the strength and/or similarity between two data points!

Problem 3

What would you expect the clusters centers (centroids) to look like if use $K < 10$? $K > 10$?

If $K < 10$ then I would expect the clusters to group similar numbers together. Possibly with a mix between 9 and 0 and another mix between 1 and 8 (that might be the response if K = 8). With $K > 10$ I would expect there to be even more separations–likely within a number that has a lot of variation, possibly 4? That might be helpful in that it would capture more nuance where edge cases are being classified incorrectly.