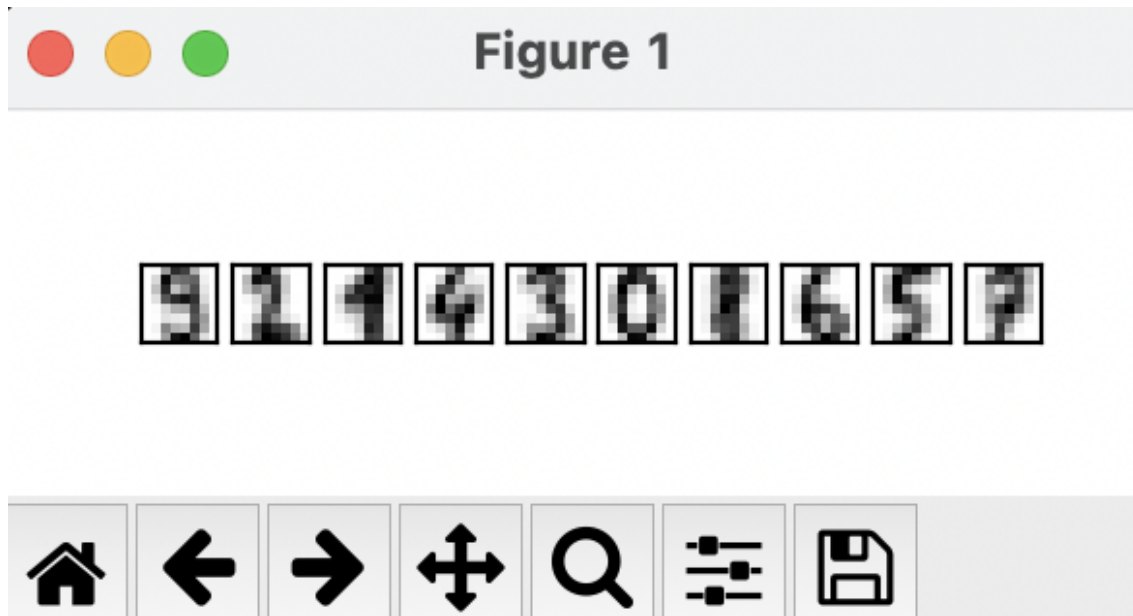


Problem 1

Below is my plot:



I expected to see digits plotted, similarly to the image above. Moreover, I also expected there to be some degree of indistinguishability between digits— for instance, as seen above, the numbers 8 and 0 appear to be very similar. Moreover, the 1 and 4 also look very similar. However, these digits are all different enough that it is clear to see which is which.

Problem 2

Another distance metric that can be used when calculating distances between points is Manhattan distance. Essentially, Manhattan distance is the sum of absolute differences between points across all possible dimensions. Thus, in a two-dimensional space, the sum of the absolute distance in x and y directions would need to be taken. Therefore, in clustering, the Manhattan distance aggregates pairwise absolute differences between each different variable to calculate distances between two points.

Problem 3

As stated in the handout, K is the number of clusters that we have. Thus, in this homework assignment, $K = 10$ since we needed 10 clusters for the 10 different numbers we were categorizing. If $K > 10$, due to distances, data would not always necessarily go to the same

cluster they were originally in because of distances. Thus, when the value of K changes, each cluster will then have less data or more data that would be spread between the numbers. Thus, if $K > 10$, each cluster has less data, and if $K < 10$, each cluster has more data. When $K < 10$, this would produce blurrier images (centroids) since there would be fewer clusters. This would cause digits such as 8 and 0, for instance, to appear very similar. However, when $K > 10$, there would be too little data for digits to be distinguishable, which would cause the centroids to also appear blurry.