

Erik Rosenstrom
erosenstrom3
Nancy Deaton
ndeaton3

Report

For the k – means; k = 4 seems to be the optimal number of clusters. The output is

```
Initial centroid 0: 3.600000 79.000000
Initial centroid 1: 1.800000 54.000000
Initial centroid 2: 3.333000 74.000000
Initial centroid 3: 2.283000 62.000000
Final centroid 0: 4.371818 84.948052
Final centroid 1: 2.006362 51.155172
Final centroid 2: 4.238557 76.012658
Final centroid 3: 2.281861 61.472222
```

Each centroid is placed just about equal distance from the others. Versus if k = 5 the output is

```
Initial centroid 0: 3.600000 79.000000
Initial centroid 1: 1.800000 54.000000
Initial centroid 2: 3.333000 74.000000
Initial centroid 3: 2.283000 62.000000
Initial centroid 4: 4.533000 85.000000
Final centroid 0: 4.342621 79.439394
Final centroid 1: 2.006362 51.155172
Final centroid 2: 4.100368 73.315789
Final centroid 3: 2.251829 61.285714
Final centroid 4: 4.384566 86.509434
```

The centroids aren't as evenly spread out. Due to the fairly small ranges in the data to begin with, smaller of number of clusters will be more meaningful.

One can look at the data and see a pattern of values, these values correlate to final cluster values nicely.

KNN produces values the appear to fit into the k-means output.