Question 2

Yufei Yin

Question 2

- Further modify your code so that instead of initializing the cluster assignments of the data items randomly, instead assign data item i to cluster ((i-1)%%K)+1. (i.e., if K = 3 and there are 7 data items, the initialization assigns the data items in order to 1, 2, 3, 1, 2, 3, 1).

*Run this further modified version of kmedians on the parkinsons dataset with K = 3, and iters = 1000. Report the locations of the 3 clusters.

```
kmedians = function(x, k, iters){
  N = dim(x)[1]
  D = dim(x)[2]

  centres = matrix(NA, k, D)
  clusters = rep(NA, N)

  for (i in 1:N){
    clusters[i] = ((i-1) %% k)+1
  }

  for (iter in 1:iters){
    for (k in 1:k){
      for (d in 1:D){
        centres[k, d] = median(x[clusters == k, d])
      }
    }
    distanceMatrix <- matrix(NA, nrow=N, ncol=k)
    for(i in 1:k) {
      distanceMatrix[,i] <- rowSums(t(abs(t(x)-centres[i,])))
    }
    clusters <- apply(distanceMatrix, 1, which.min)
    centres <- apply(x, 2, tapply, clusters, median)
  }
  return(list(locations=centres, assignment=clusters))
}
```

```
data = read.table(file = 'parkinsons.data', sep = ',', header = TRUE)
data = data[,-1]
res = kmedians(x = data, k = 3, iters = 1000)
res$locations
```

```
##   MDVP.Fo.Hz. MDVP.Fhi.Hz. MDVP.Flo.Hz. MDVP.Jitter... MDVP.Jitter.Abs.
## 1     158.219     208.7010      79.5430        0.00564            4e-05
## 2     202.908     228.6375     184.0220        0.00297            1e-05
## 3     117.572     131.9825     106.1855        0.00494            4e-05
```

```
##   MDVP.RAP MDVP.PPQ Jitter.DDP MDVP.Shimmer MDVP.Shimmer.dB. Shimmer.APQ3
## 1  0.00287 0.003170   0.008620      0.02574           0.2550      0.01441
## 2  0.00167 0.001705   0.005010      0.01756           0.1545      0.00947
## 3  0.00252 0.002835   0.007565      0.02470           0.2255      0.01350
##   Shimmer.APQ5 MDVP.APQ Shimmer.DDA      NHR     HNR status      RPDE       DFA
## 1     0.015820 0.019560     0.04322 0.018400 20.4220      1 0.5091270 0.7040870
## 2     0.010915 0.012780     0.02841 0.005595 24.5745      0 0.4318620 0.7221695
## 3     0.014020 0.019485     0.04049 0.010290 22.3375      1 0.5393685 0.7552775
##     spread1   spread2       D2      PPE
## 1 -5.571843 0.2217110 2.555477 0.215558
## 2 -6.981070 0.1769335 2.375411 0.109477
## 3 -5.514255 0.2343545 2.280550 0.215035
```