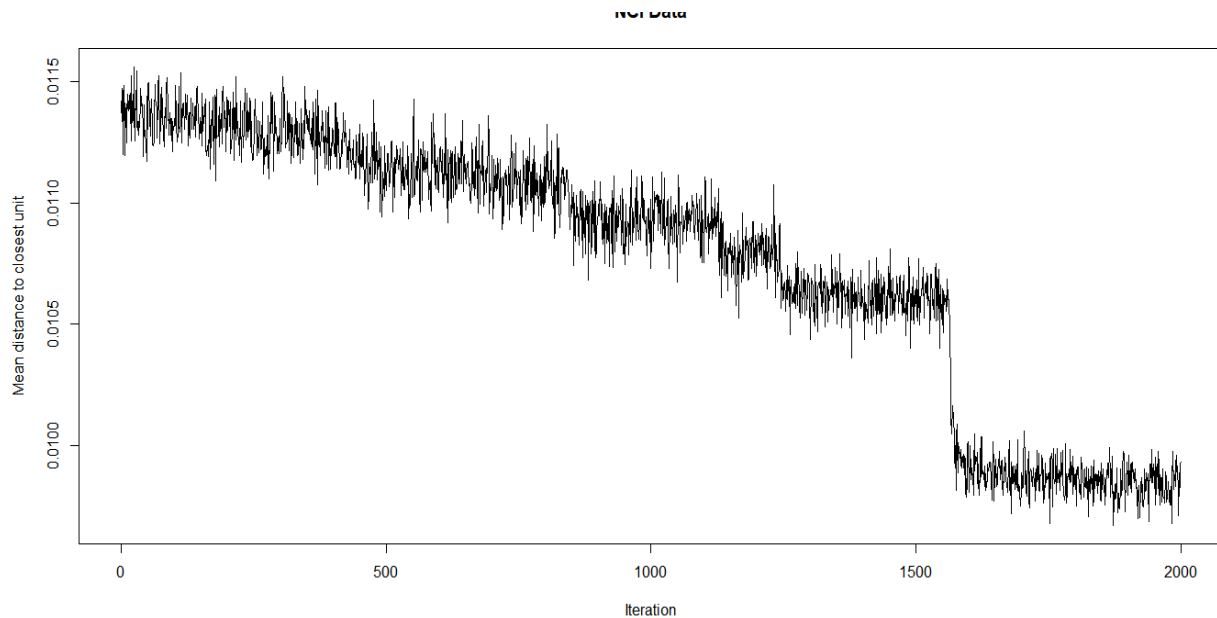


## Question 1

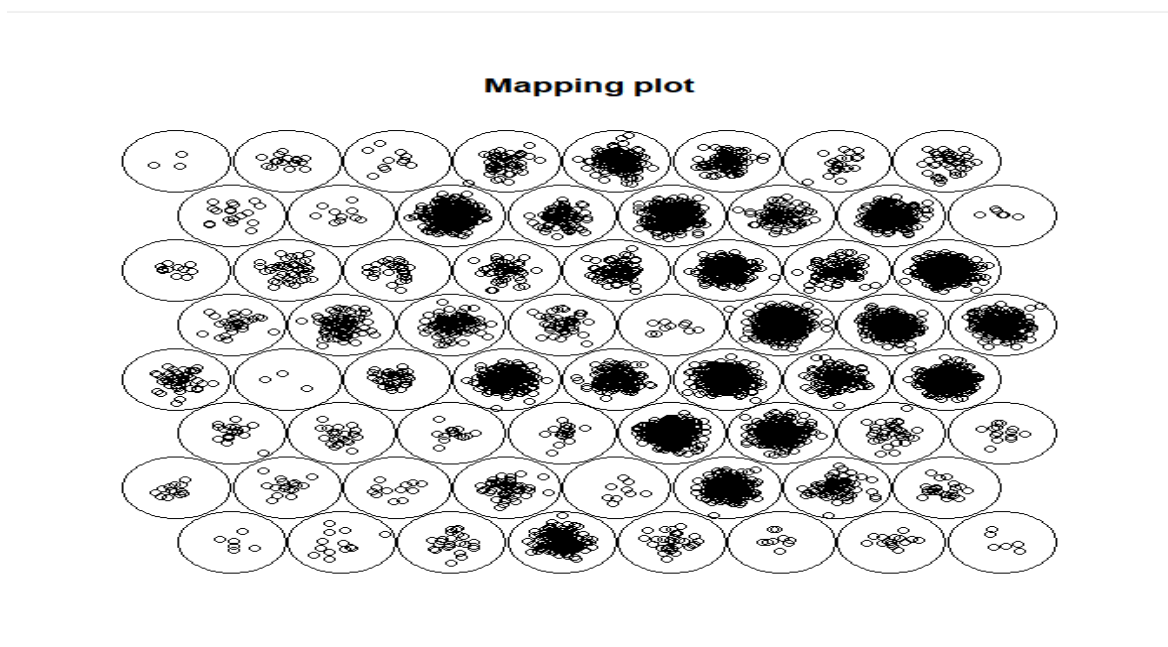
- Calculating the SOM algorithm for the NCI data.

```
som_grid <- somgrid(xdim = 8, ydim = 8, topo = "hexagonal")  
nci.som <- som(nci.scaled, grid = som_grid, rlen = 2000)
```

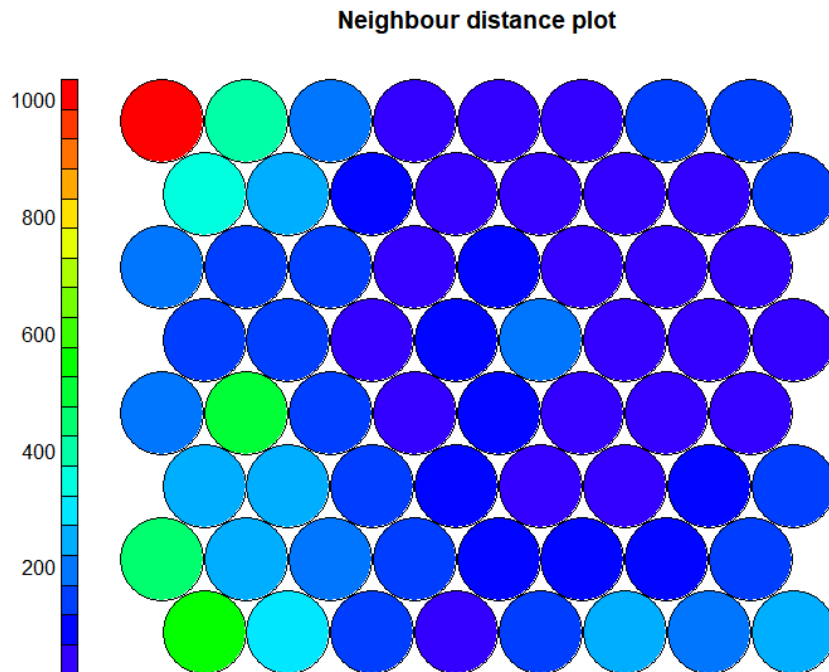
- Mean distance to closest unit vs Iteration Graph Converges at 'rlen' 2000



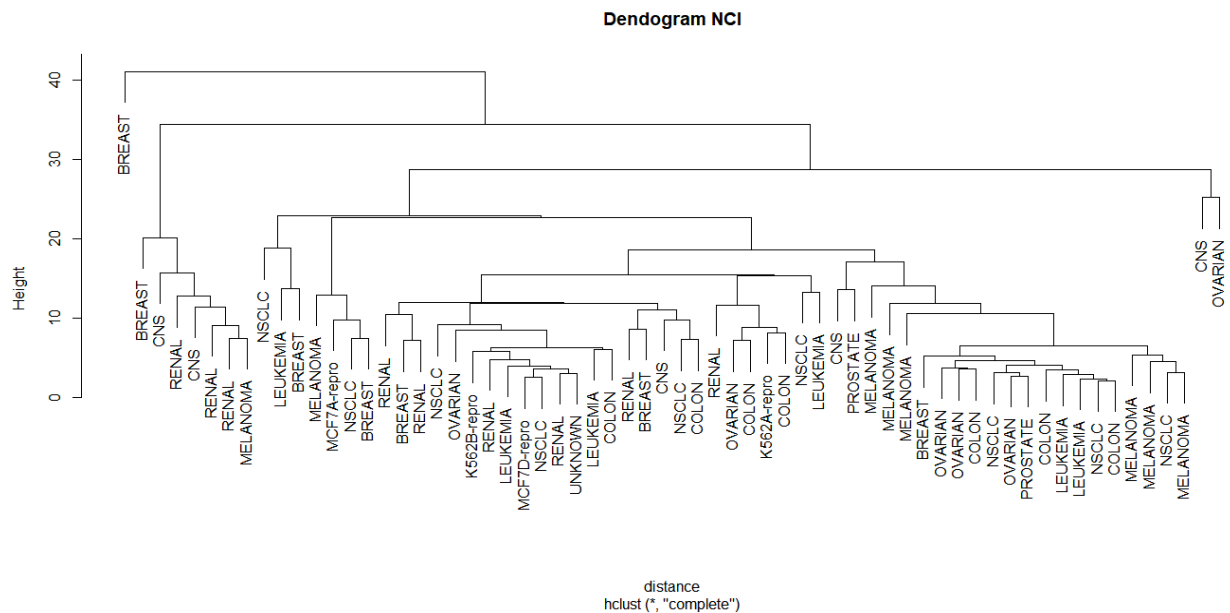
- Mapping Plot for NCI data for 8\*8 grid



- U-Matrix Visualization: From U-matrix visualization plot below we get the idea that there could be 3 groups for SOM algorithm



- Performing Hierarchical cluster we get the following dendrogram



- Cluster size after 3 clusters are formed in hierarchical clustering. We can see that cluster 1 contains most of the tumor cells data.

```

> groups.3<-cutree(nci.hclust,3)
> groups.3
  V1  V2  V3  V4  V5  V6  V7  V8  V9  V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V21 V22 V23 V24 V25
1   1   1   1   1   2   2   2   1   1   2   1   1   1   2   2   1   1   1   1   1   1   2   1   1
V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43 V44 V45 V46 V47 V48 V49 V50
1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
V51 V52 V53 V54 V55 V56 V57 V58 V59 V60 V61 V62 V63 V64
1   1   1   1   1   3   1   1   1   1   1   1   1   1
> table(groups.3)
groups.3
 1  2  3
56  7  1

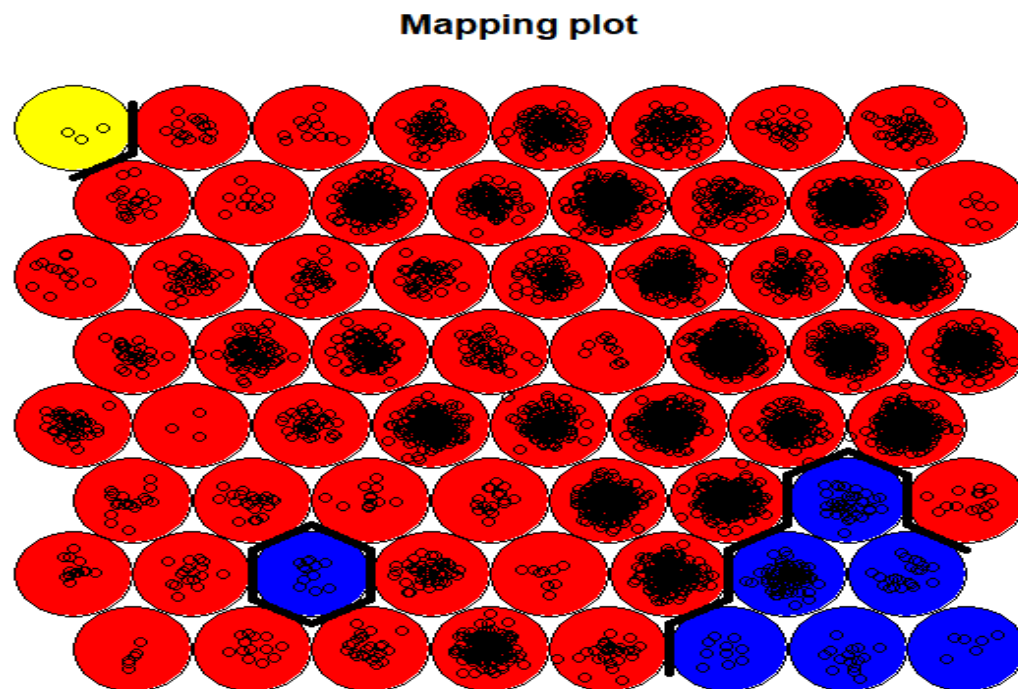
```

- Showing clusters with boundaries on SOM plot.

```

> plot(nci.som, type = "mapping", col = "black", bgcol = bgcolumns)
> add.cluster.boundaries(nci.som, groups.3)

```



- From the figure below we can see that group 3 contains only one BREAST tumor cells, Group 2 contains some CNS, RENAL, MELANOMA and BREAST tumor cells and group 1 contains all of the OVARIAN, PROSTATE, COLON, LEUKEMIA, NSCLC tumor cells including some CNS, BREAST, MELANOMA and RENAL
- We can see that BREAST tumor cells is distributed in all of the three group, CNS, MELANOMA and RENAL are distributed in group 1 and group 2
- Group 1 contains all of the types of tumor cells.

```
> unique(colnames(nci))[groups.3==3]
[1] NA
> unique(colnames(nci))[groups.3==1])
[1] "CNS"      "RENAL"    "BREAST"    "NSCLC"    "UNKNOWN"  "OVARIAN"
[7] "PROSTATE" "LEUKEMIA" "K562B-repro" "K562A-repro" "COLON"    "MCF7A-repro"
[13] "MCF7D-repro" "MELANOMA"
> unique(colnames(nci))[groups.3==2])
[1] "CNS"      "BREAST"    "RENAL"    "MELANOMA"
> unique(colnames(nci))[groups.3==3])
[1] "BREAST"
```