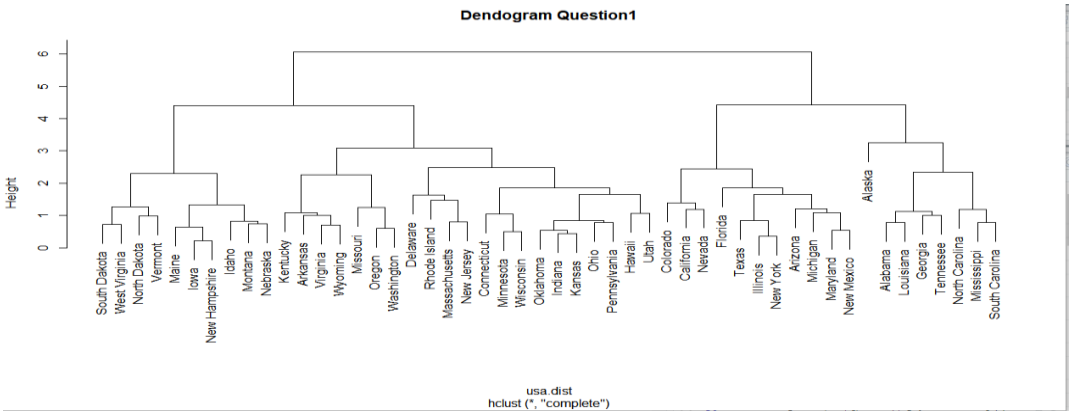
Question 2

- Dendrogram of Hierarchical clustering with complete linkage.



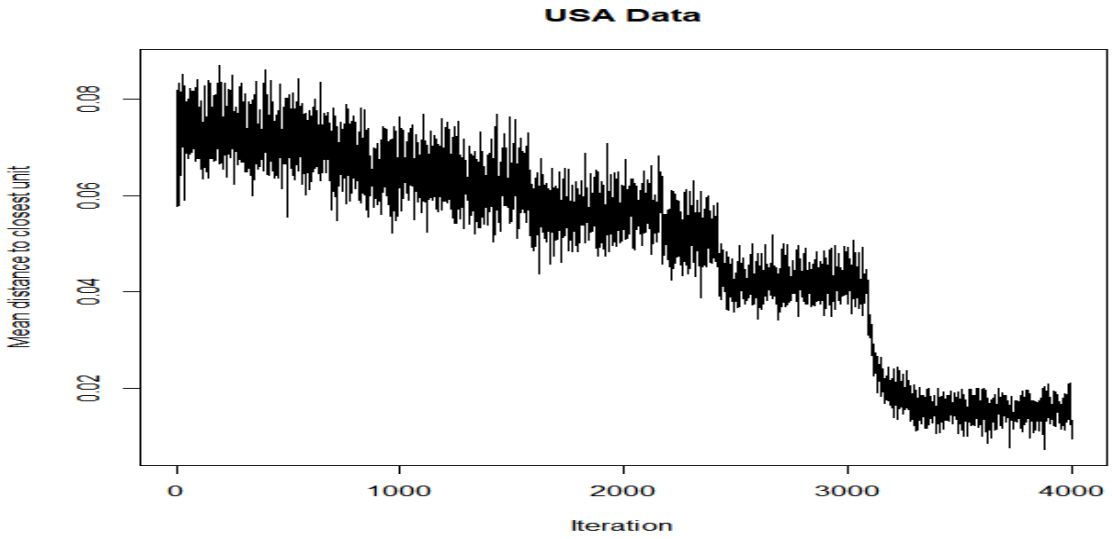
- State belonging to group for USA Arrest Data.

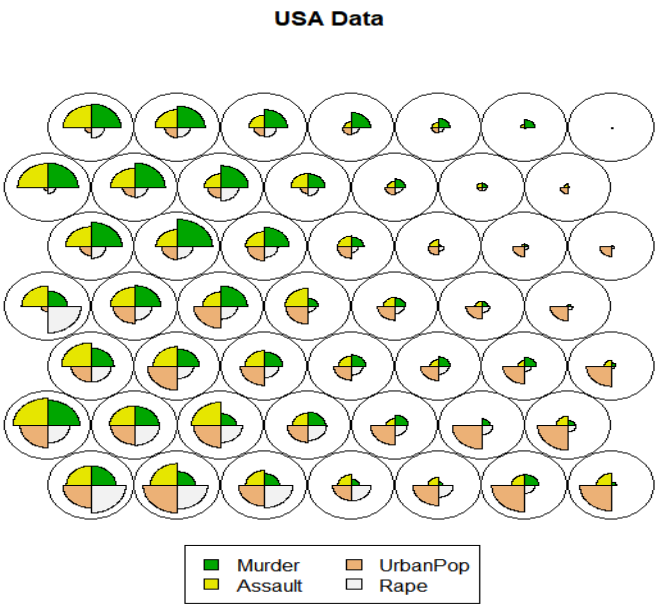
```
> groups.3=cutree(usa.hclust,3)
>
> groups.3
       Alabama         Alaska        Arizona       Arkansas     California       Colorado
             1              1              2              3              2              2
   Connecticut       Delaware        Florida        Georgia         Hawaii          Idaho
             3              3              2              1              3              3
      Illinois        Indiana           Iowa         Kansas       Kentucky      Louisiana
             2              3              3              3              3              1
         Maine       Maryland  Massachusetts       Michigan      Minnesota    Mississippi
             3              2              3              2              3              1
      Missouri        Montana       Nebraska         Nevada  New Hampshire     New Jersey
             3              3              3              2              3              3
    New Mexico       New York North Carolina   North Dakota           Ohio       Oklahoma
             2              2              1              3              3              3
        Oregon   Pennsylvania   Rhode Island South Carolina   South Dakota      Tennessee
             3              3              3              1              3              1
         Texas           Utah        Vermont       Virginia     Washington  West Virginia
             2              3              3              3              3              3
     Wisconsin        Wyoming
             3              3
```

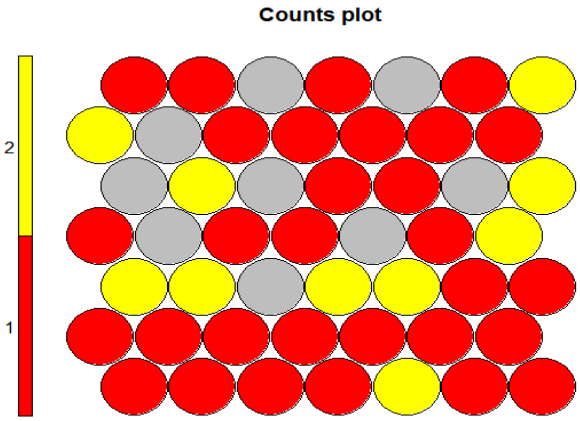
- No of elements in a cluster for hierarchical clustering.

```
> table(groups.3)
groups.3
 1  2  3
 8 11 31
```

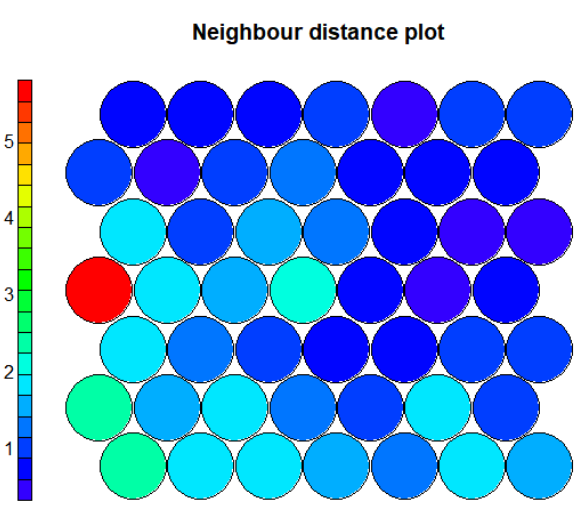
- SOM algorithm on USA arrest Data for rlen =4000 and grid size 7 * 7



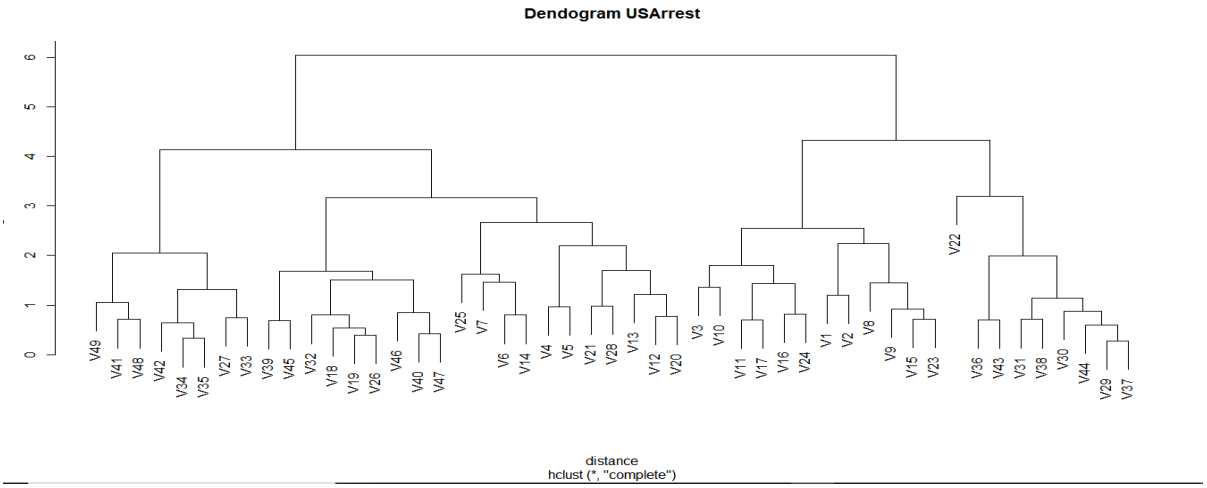




- U- Matrix for the Data.



- Dendrogram of complete linkage for SOM of USArrest data

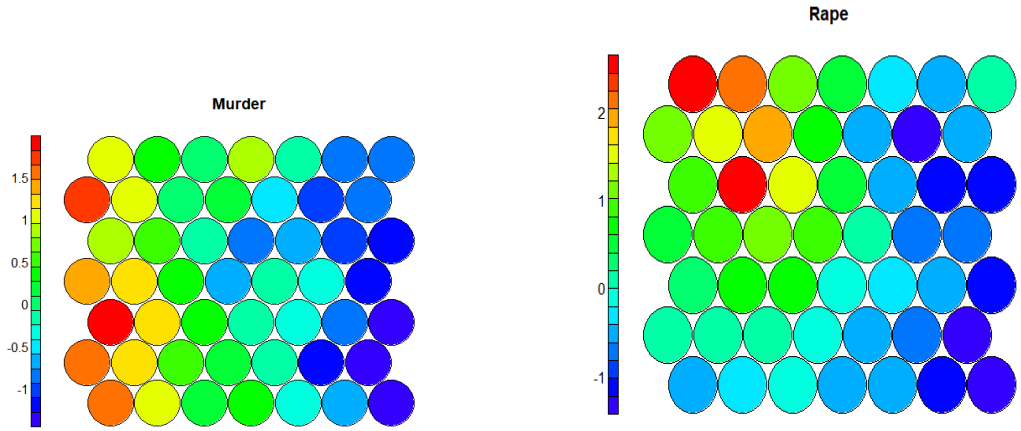


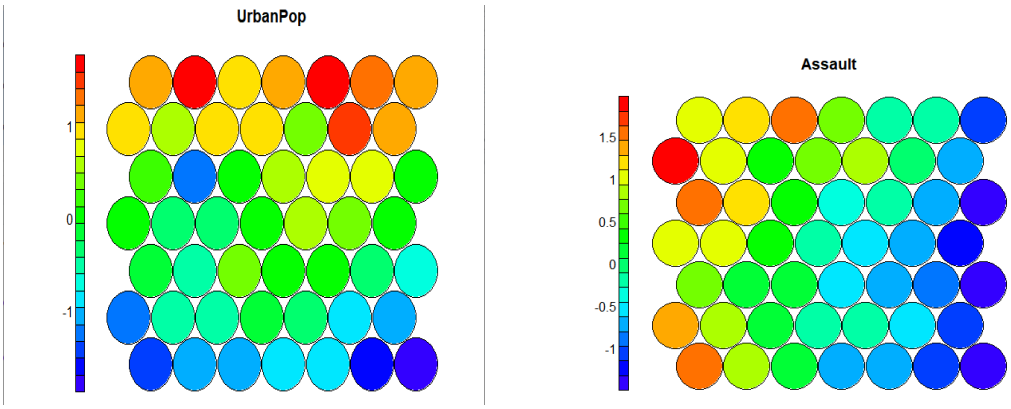
- Cutting the dendrogram to 3 clusters, we get,

```
> groupsom.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   2   2   2   2   1   1   1   1   2   2   2   1   1   1   2   2   2   2   3   1   1   2
V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43 V44 V45 V46 V47 V48 V49
  2   2   2   3   3   3   2   2   2   2   3   3   3   2   2   2   2   3   3   2   2   2   2   2
> table(groupsom.3)
groupsom.3
 1  2  3
12 28  9
>
```

```
> rownames(usa)[groupsom.3==1]
 [1] "Alabama"      "Alaska"       "Arizona"      "Delaware"     "Florida"      "Georgia"
 [7] "Hawaii"       "Iowa"         "Kansas"       "Kentucky"     "Minnesota"    "Mississippi"
[13] "Wyoming"
> rownames(usa)[groupsom.3==2]
 [1] "Arkansas"       "California"     "Colorado"       "Connecticut"    "Idaho"
 [6] "Illinois"       "Indiana"        "Louisiana"      "Maine"          "Maryland"
[11] "Massachusetts"  "Missouri"       "Montana"        "Nebraska"       "Nevada"
[16] "New York"       "North Carolina" "North Dakota"   "Ohio"           "Rhode Island"
[21] "South Carolina" "South Dakota"   "Tennessee"      "Vermont"        "Virginia"
[26] "Washington"     "West Virginia"  "Wisconsin"
> rownames(usa)[groupsom.3==3]
[1] "Michigan"      "New Hampshire" "New Jersey"    "New Mexico"    "Oklahoma"      "Oregon"
[7] "Pennsylvania"  "Texas"         "Utah"
>
```

- Maps for Crimes Like Rape, Assault, Urban Pop, Murder





- We can conclude from the plot that group 1 with states like Alabama, Alaska etc as shown in above figure have highest murder, Rape and Assault cases in Case of SOM maps
- Group 2 has intermediate murder, Rape and Assault cases in case of SOM maps
- Group 3 has lowest murder, Rape and Assault cases in case of SOM maps.

c) We can conclude that SOM is better than hierarchical clustering for high dimensional data. We