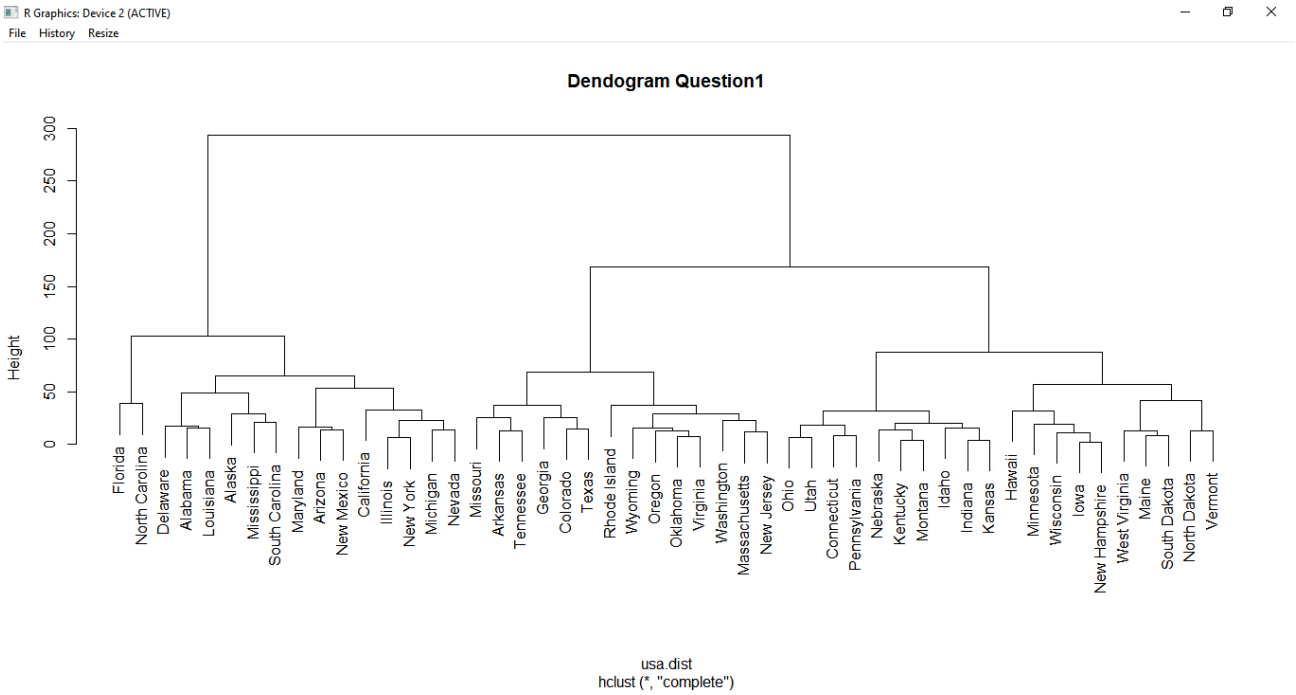
Question 1:

a)



Dendrogram For hierarchical clustering with complete linkage and Eculidean distance.

b) Cutting dendrogram at height with 3 clusters

groups.3=cutree(usa.hclust,3)

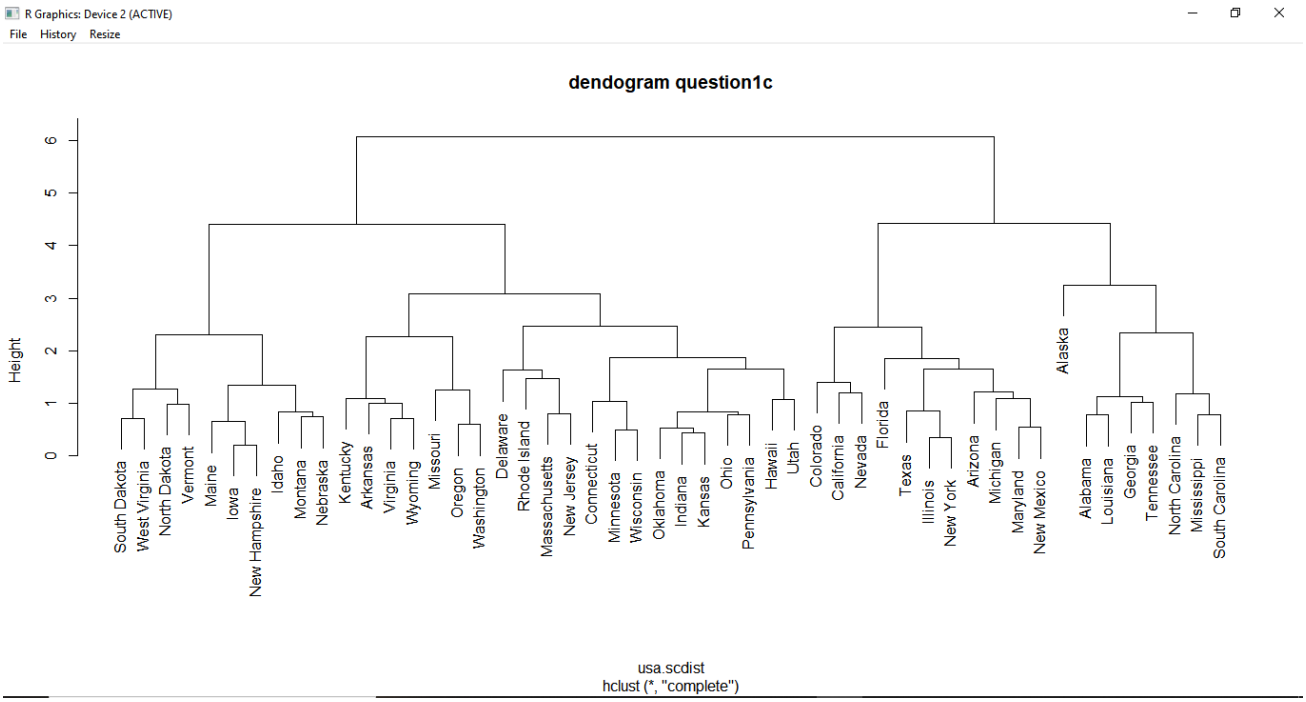
groups.3

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

```
> table(groups.3)
groups.3
 1  2  3
16 14 20
```

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

c) variables to be scaled to standard deviation one.



d) Effect of scaling on variables of hierarchical clustering:

```
> scgroups.3=cutree(usa.schclust,3)
> scgroups.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
>
```

```
> table(scgroups.3)
scgroups.3
 1  2  3
 8 11 31
> table(groups.3,scgroups.3)
        scgroups.3
groups.3  1  2  3
       1  6  9  1
       2  2  2 10
       3  0  0 20
>
```

Conclusion drawn from difference between scaled and unscaled variables:

- Clusters obtained from cutting the dendrogram into 3 cluster is different for scaled data.
- Maximum height of dendrogram is changed after scaling the variables.
- Variable scaling is necessary for standardization of data set. As observations in the dataset are measured on different units therefore they can have dissimilarities. Therefore scaling of variable is necessary.