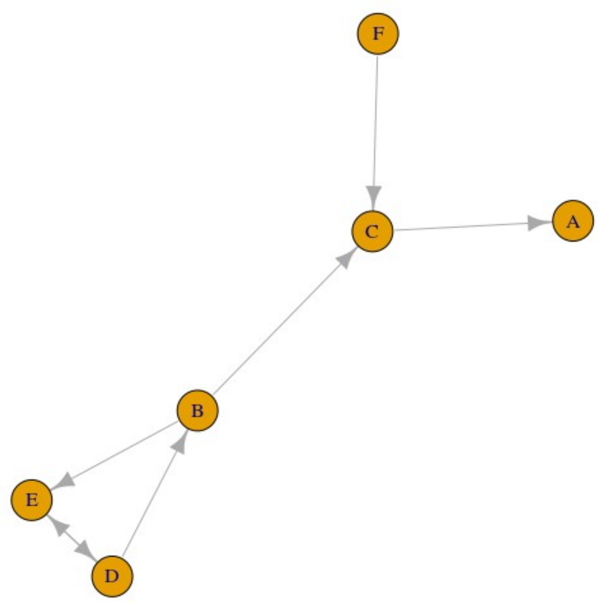
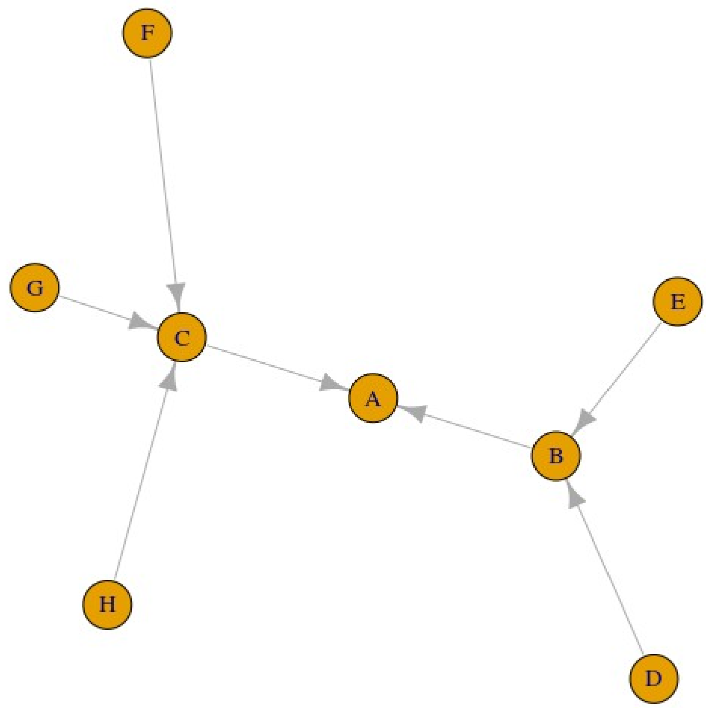
Homework-5

Q.3)




```
> pg1<-page.rank(g,damping = 0.05)
> pg1$vector
        A         B         C         D         E         F 
0.1683271 0.1639395 0.1718214 0.1681380 0.1680380 0.1597361 
> 
> pg2<-page.rank(g,damping = 0.25)
> pg2$vector
        A         B         C         D         E         F 
0.1786588 0.1544288 0.1848587 0.1758772 0.1737324 0.1324441 
> 
> pg3<-page.rank(g,damping = 0.5)
> pg3$vector
         A          B          C          D          E          F 
0.19227231 0.14719411 0.18583257 0.19135235 0.18399264 0.09935603 
> 
> pg4<-page.rank(g,damping = 0.75)
> pg4$vector
         A          B          C          D          E          F 
0.19399617 0.14778661 0.17077331 0.21832113 0.20320659 0.06591619 
> 
> pg5<-page.rank(g,damping = 0.95)
> pg5$vector
         A          B          C          D          E          F 
0.17305017 0.15761096 0.14454445 0.25658531 0.23247617 0.03573294 
```

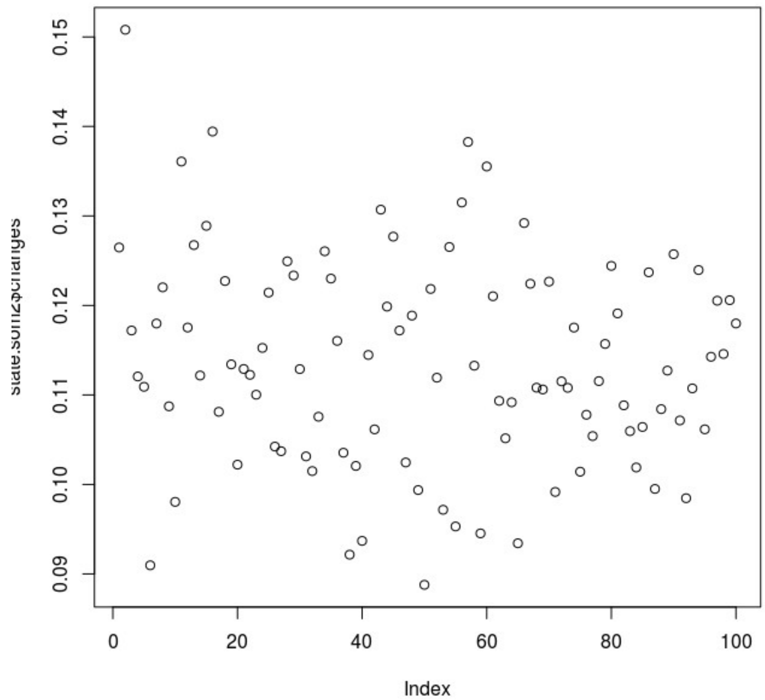
As damping factor increases, page rank value is more sensative to no of incoming links. For F point, we can see rank value goes on decreasing as no of incoming link is zero.
It holds more importance to number of outgoing links than incoming links. C has 2 incoming and 1 outgoing, still rank value is going on decreasing. Surprisingly, for A rank value decreases for very high damping factor. As expected, its value increases with damping factor except that case.




```
  pg.b<-page.rank(g2,damping = 0.15)
> pg.b$vector
        A         B         C         D         E         F         G         H 
0.1541610 0.1418827 0.1582538 0.1091405 0.1091405 0.1091405 0.1091405 0.1091405
```

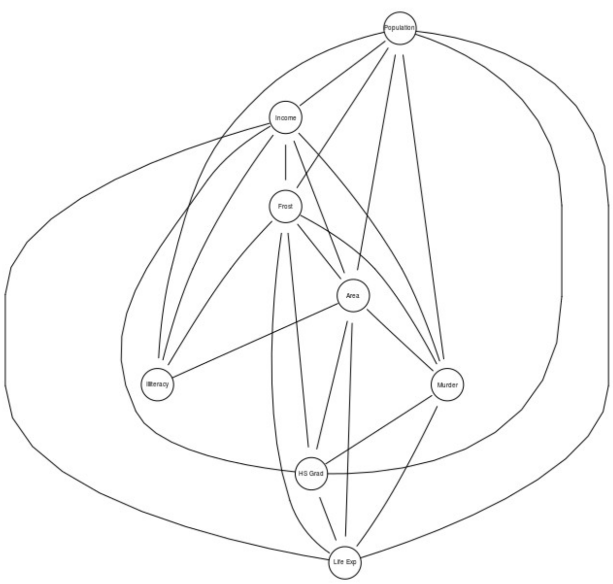
As we can see from this graph, C has most number of incoming nodes(3). So as expected, it has max rank value. Even though it is pointing to A, A has only 2 incoming nodes. So number of incoming nodes dominate over number of outgoing nodes.

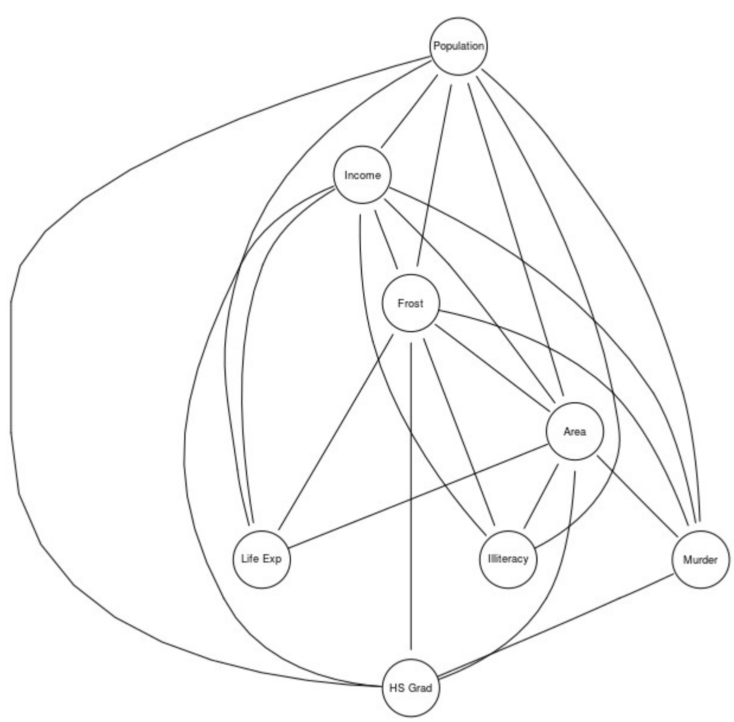
Q.4)

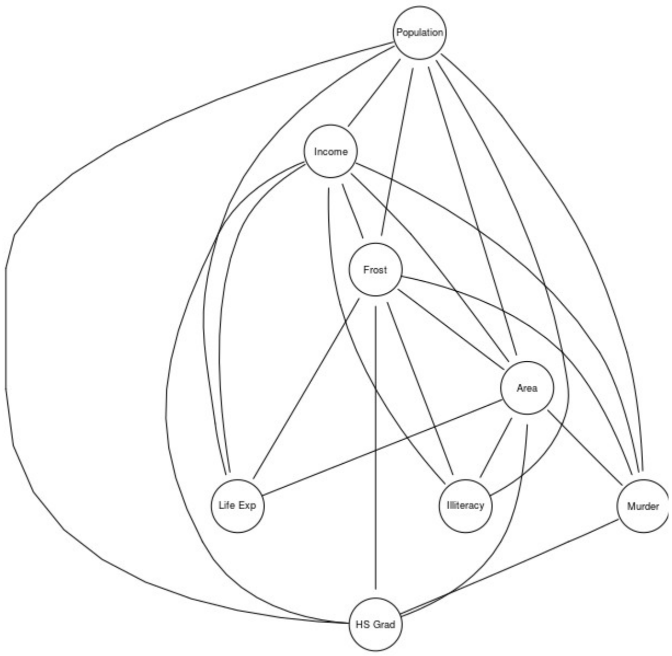


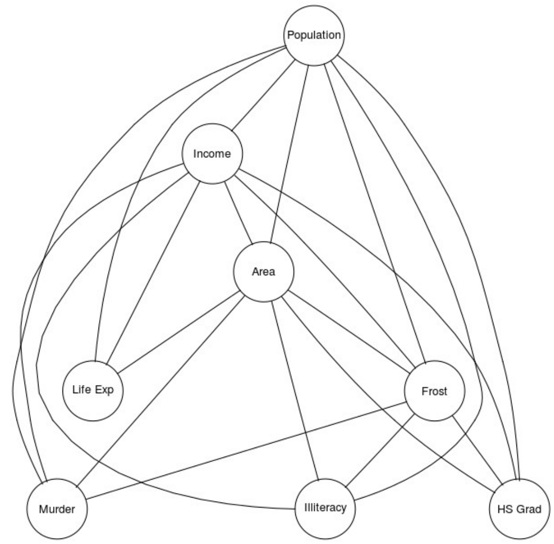
For two clusters, self organising maps are like below. The change in variable is shown in the plot. We can see, reduction is done till 100th iterations.

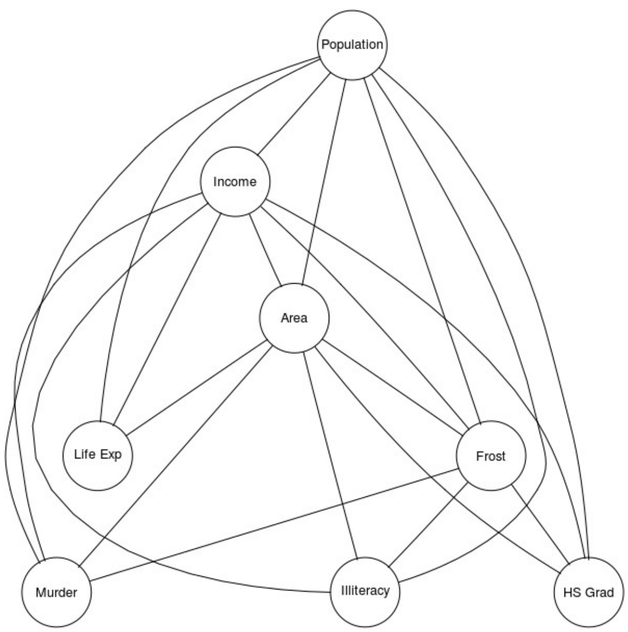
The Self-Organizing Map was created first. The “kohonen” package was used. Using a different dimension of grid, the results are plotted to the right. These results will be used later, as the Graphical Lasso model will be generated next. The “glasso” package was used next to generate graphical lasso models of the data. For this problem, the rho values were set to 2, 4, 6, 10, and 15. Out of the five variables, only three unique plots were generated. These three unique plots are shown below



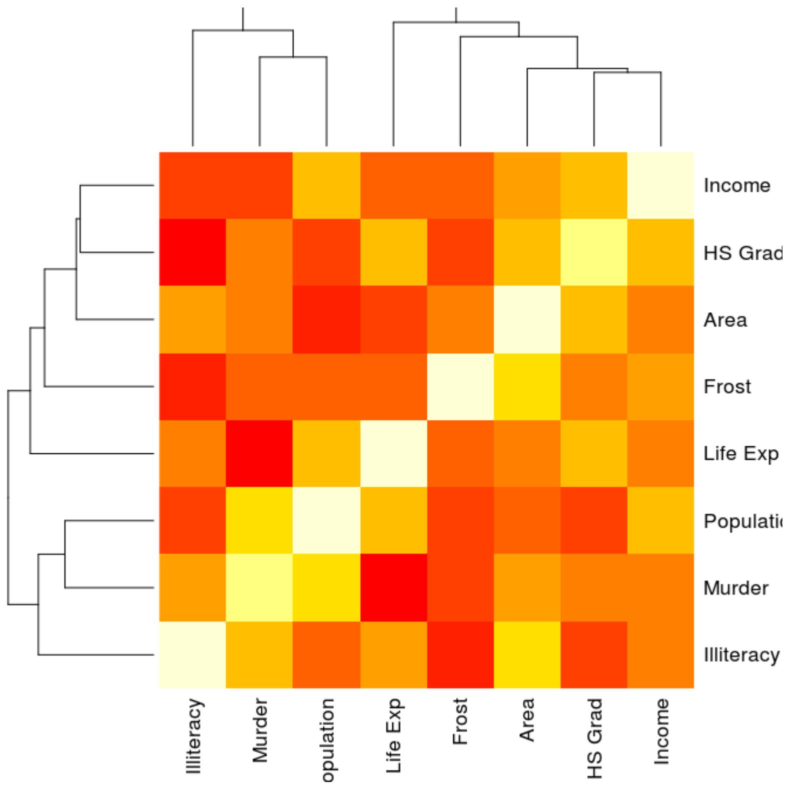




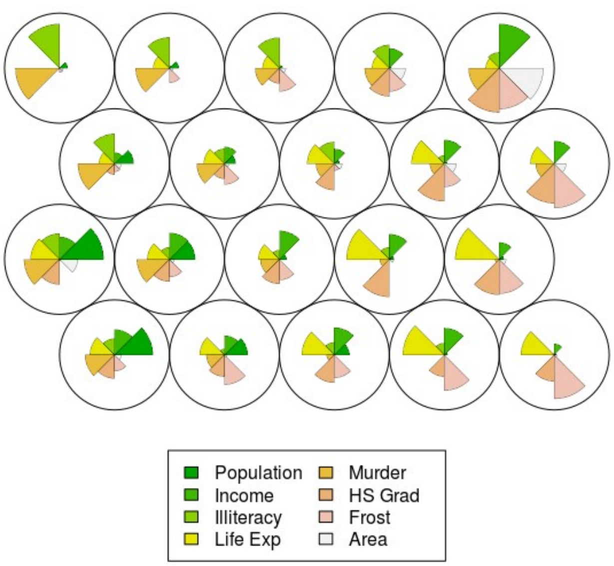




We can see correlation goes on decreasing.But some relations remain the same for all tvalues of rho.



Correlation heatmap generated using glasso model shoes equivalence with the graphical models generated.



For grid size 20. we see same corelation as that of Glasso model. So results of som compliments results of glasso

Q.2)
Cluster 2 had a higher percentage of females. This also backs up the idea that women had a higher chance of surviving. Unusually, from the histograms generated, children had a lower chance of surviving vs. the analysis from Hierarchical Clustering. Further analysis on this would be required. Finally, as expected, It can be seen from the tabulation of Cluster 2 that Class 1 passengers were much

more likely to survive. 125 out of 136 members of Cluster 2 were from Class 1! While I do not have enough evidence to support age being a factor in survivability, it is clear passengers from Class 1 and women were much more likely to survive the disaster, and fared better than men and Class 3 passengers.