Lab6

Andreas

29 Οκτωβρίου 2018

Assignment 1

This assignment is all about network visualsation of the terrorist connections.

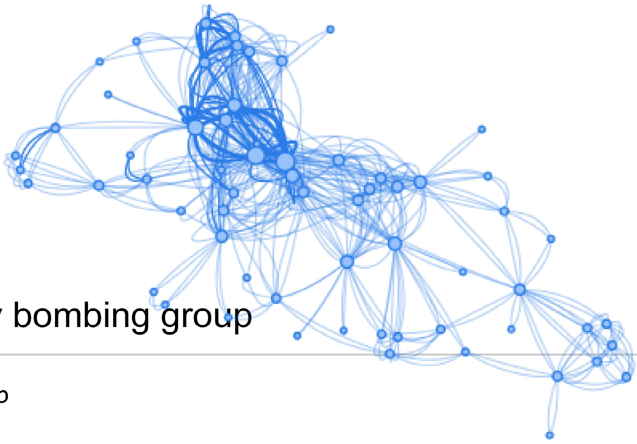
The files given are *trainData.dat* and *trainMeta.dat*.The files have the data about thenetwork of people involved in Madrid bombing.

We plot a graph using the *visNetwork* package

The basic graph is as below

```
library(ggraph)
library(igraph)
library(visNetwork)
library(seriation)
library(plotly)

#setwd("C:/Users/quartermaine/Documents/Visualization/lab_6")
edges <- read.delim("trainData.dat", header = FALSE, sep  = " ")
nodes <- read.delim("trainMeta.dat", header = FALSE, sep = " ")
set.seed(12345)

nodes$id <- rownames(nodes)
colnames(nodes) <- c("bombers", "b_group", "id")
colnames(edges) <- c("temp", "from", "to", "value")
edges$temp <- NULL
graph <- graph.data.frame(edges, directed = T)
degree_value <- degree(graph)
nodes$value <- degree_value[match(nodes$id, names(degree_value))]
nodes <- na.omit(nodes)
nodes$label<-nodes$bombers
#### a.Basic
visNetwork(nodes=nodes,edges=edges,main = "Madrid bombing people network")
```

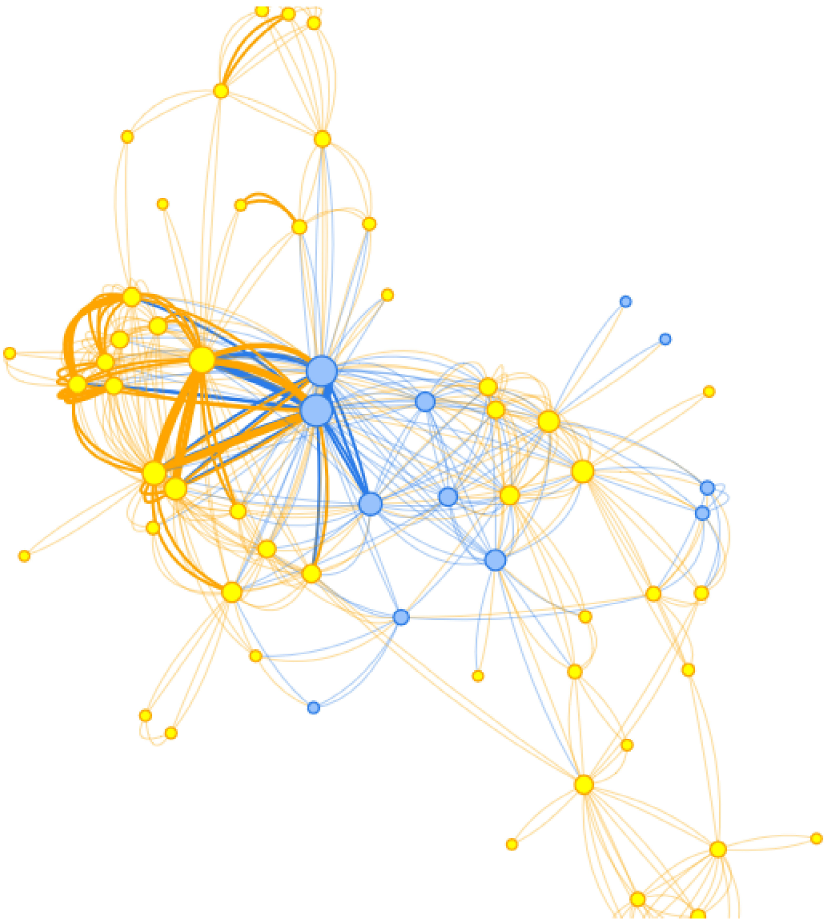
Madrid bombing people network



b.nodes are coloured by bombing group

```
nodes$group<-nodes$b_group

visNetwork(nodes=nodes,edges=edges,main = "Madrid bombing people network")
```

Madrid bombing people network



c. Strength

Strength can be also showed by degree

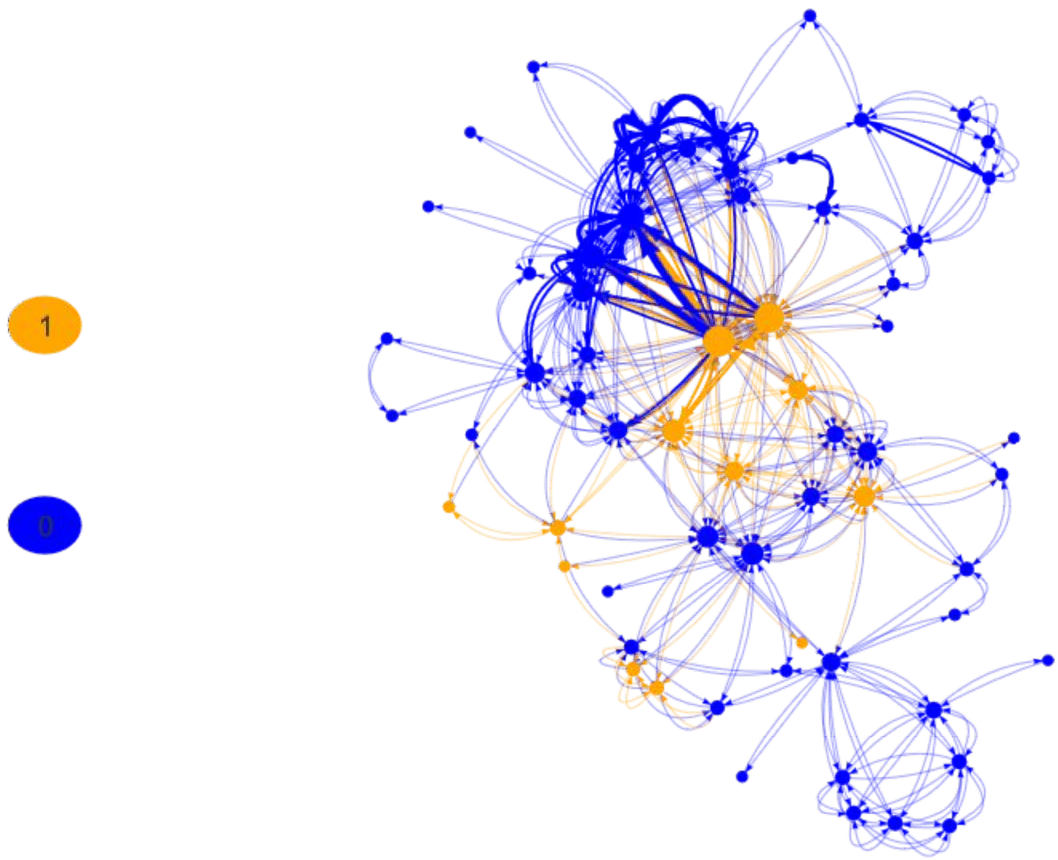
```
graph <- graph.data.frame(edges, directed = T)
degree_value <- degree(graph)
```

d.replusion

```
visNetwork(nodes = nodes, edges = edges, main = "Madrid bombing people network") %
>%
  visGroups(groupname = "0", color = "blue") %>%
  visGroups(groupname = "1", color = "orange") %>%
  visEdges(arrows = "from") %>%
  visOptions(collapse = TRUE,
             selectedBy = "group") %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome()
```

Madrid bombing people network

Select by group

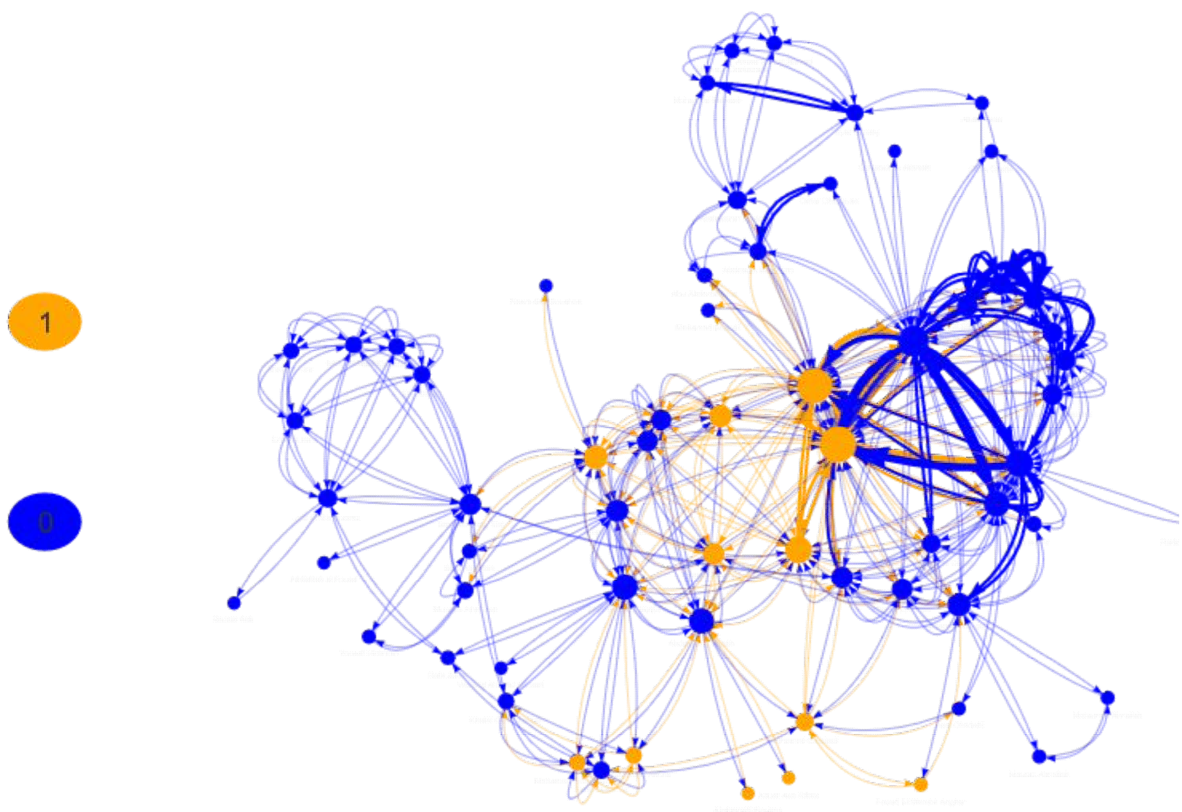


e.Highlighted nodes ans the end graph

```
visNetwork(nodes = nodes, edges = edges,main = "Madrid bombing people network") %
>%
  visGroups(groupname = "0", color = "blue") %>%
  visGroups(groupname = "1", color = "orange") %>%
  visEdges(arrows = "to") %>%
  visOptions(highlightNearest = list(enabled =TRUE,
                                     degree = 1),
             collapse = TRUE,
             selectedBy = "group") %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome()
```

Madrid bombing people network

Select by group

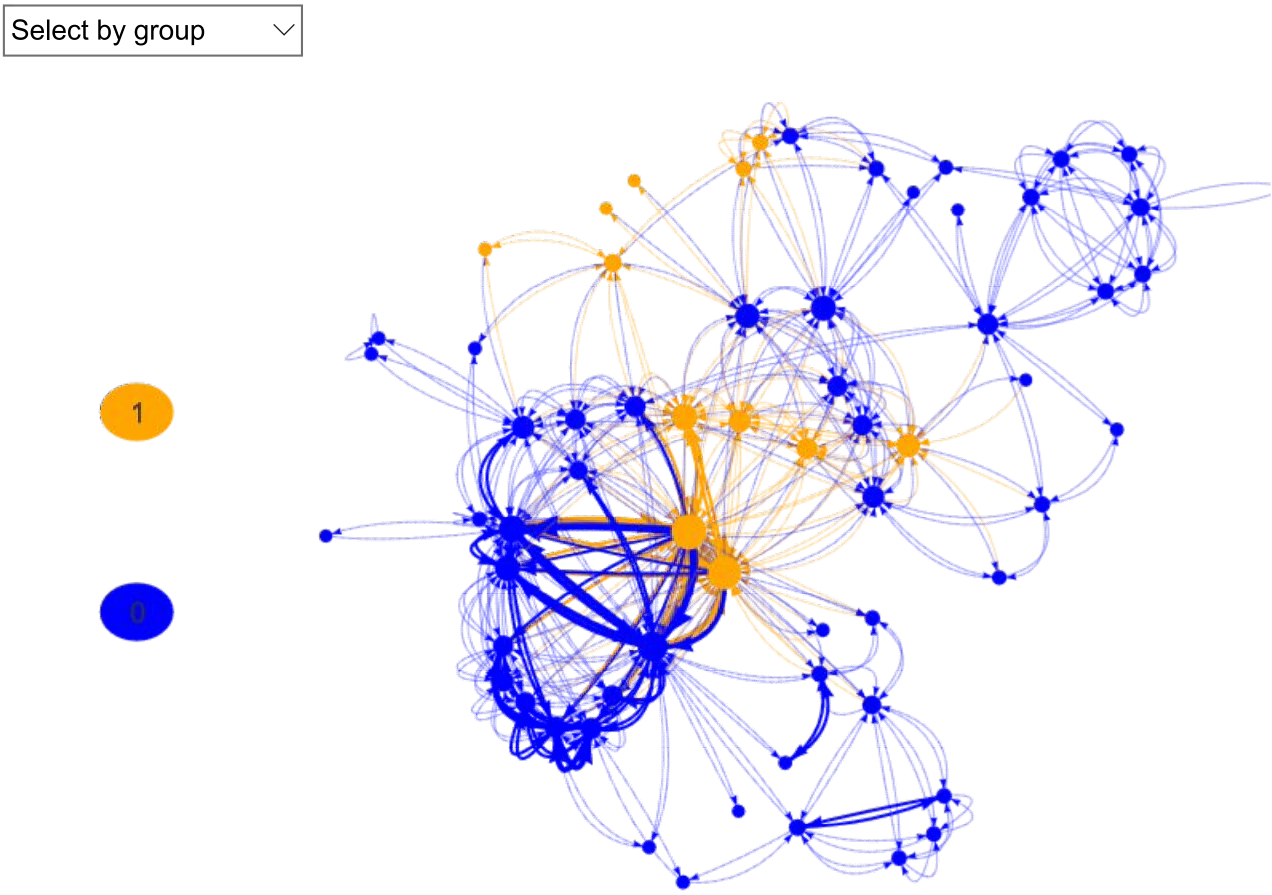


Group 1 is for those involved in blasting and 0 for others.We could see one main cluster centered around Mohamed chaoui,Jamal Zougam,Basel chayoun ,SB abdelmajid Fakher.They are the people actievely involved in bombing.Also there are three other small clusters of family members one with abdel karim,Tayser,mohamed bekkali etc.another with seeman gaby eid,el gitano etc and last one with mohamed chedadi,mohamed oulat akcha etc.

2.

degree = list(from = 1, to = 2)) gives the highlight required

```
visNetwork(nodes = nodes, edges = edges) %>%
  visGroups(groupname = "0", color = "blue") %>%
  visGroups(groupname = "1", color = "orange") %>%
  visEdges(arrows = "from") %>%
  visOptions(highlightNearest = list(enabled =TRUE,
                                    degree = list(from = 1, to = 2)),
           collapse = TRUE,
           selectedBy = "group") %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome()
```

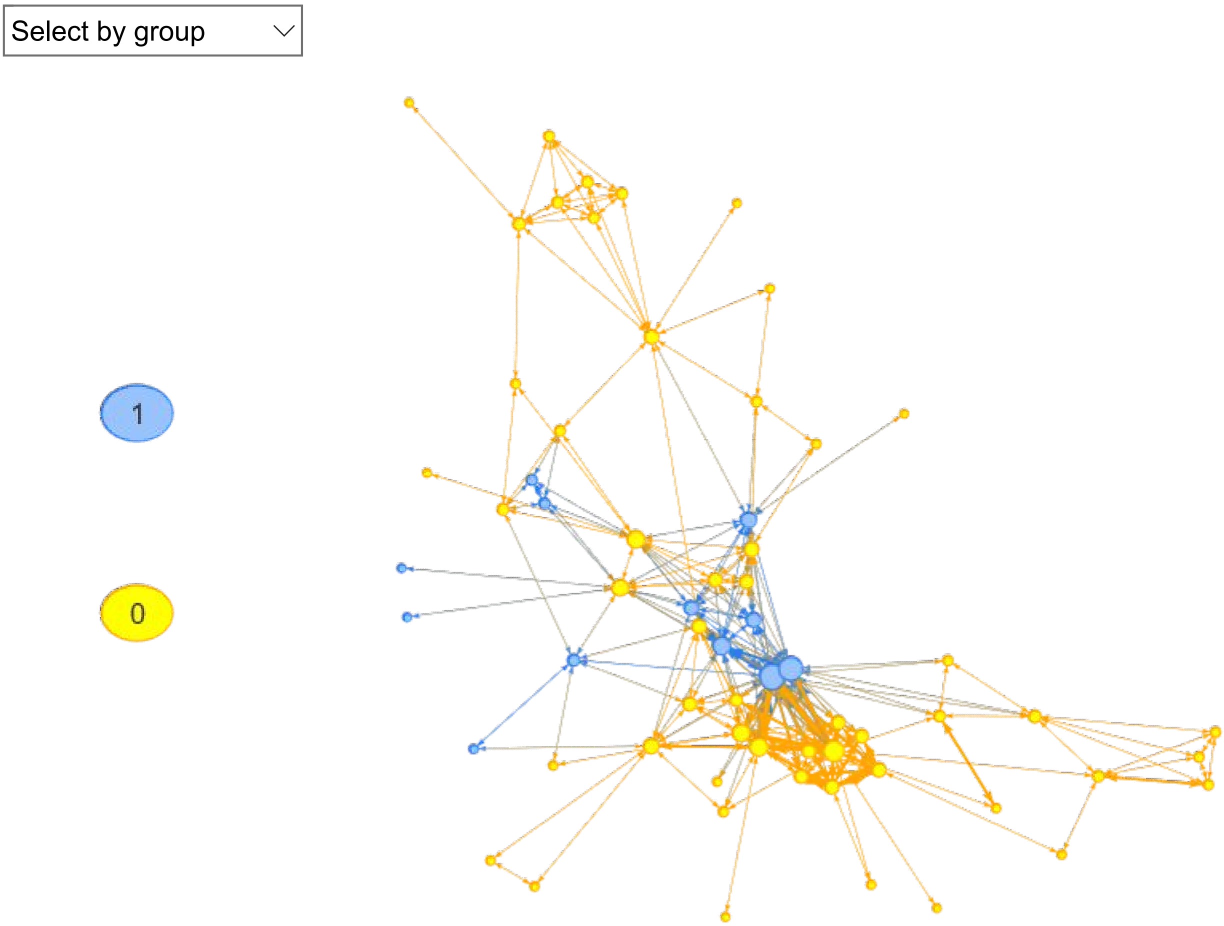


We couls see jamal Zougam and mohamed chaoui were best in spreading news.The BBC 18-4-2014 (http://news.bbc.co.uk/2/hi/europe/3515790.stm) shows him as an early suspect of the attack and he oowned a mobile shop so he could be the point of contact to get all the fake sim cards that connected the people.

3.

```
graph_data_frame <- graph.data.frame(edges, directed = FALSE)
clusters <- cluster_edge_betweenness(graph_data_frame, directed = T)
nodes$clusters <- clusters$membership

visNetwork(nodes = nodes, edges = edges, main = "Madrid Bombing") %>%
  visEdges(arrows = "from") %>%
  visOptions(highlightNearest = list(enabled =TRUE,
                                     degree = list(from = 1, to = 2)),
             collapse = TRUE,
             selectedBy = "group")  %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome() %>%visIgraphLayout()
```

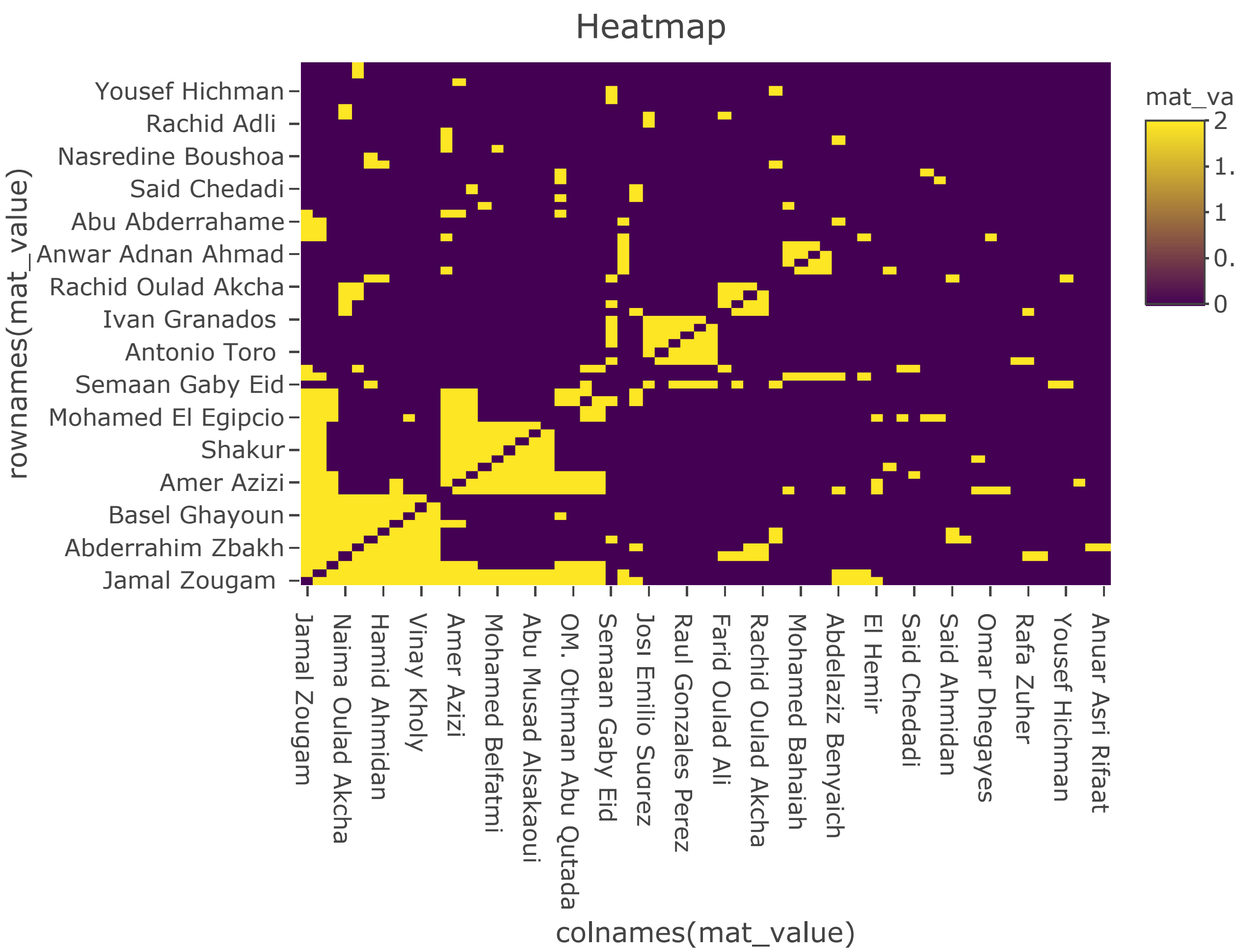
Madrid Bombing



The main cluster with Jamal Zougam is clearly evident in both the graphs.The two other clusters can be also seen one with seeman gaby eid and other with mohamed chedadi,mohamed oulat akcha the third small cluster is not so visible.In conclusion the promintent cluster of Jamal Zougm can be defenetly seen in both the clusters.

4.

```
adjecency <- get.adjacency(graph_data_frame, sparse=F)
colnames(adjecency) <- nodes$label
rownames(adjecency) <- nodes$label
rowdist<-dist(adjecency)


reord<-get_order(seriate(rowdist, "HC"))
mat_value<-adjecency[reord,reord]

##SInce using ubuntu system using the below line of code.



plot_ly(z=~mat_value, x=~colnames(mat_value),
        y=~rownames(mat_value), type="heatmap") %>% layout(title = "Heatmap")
```



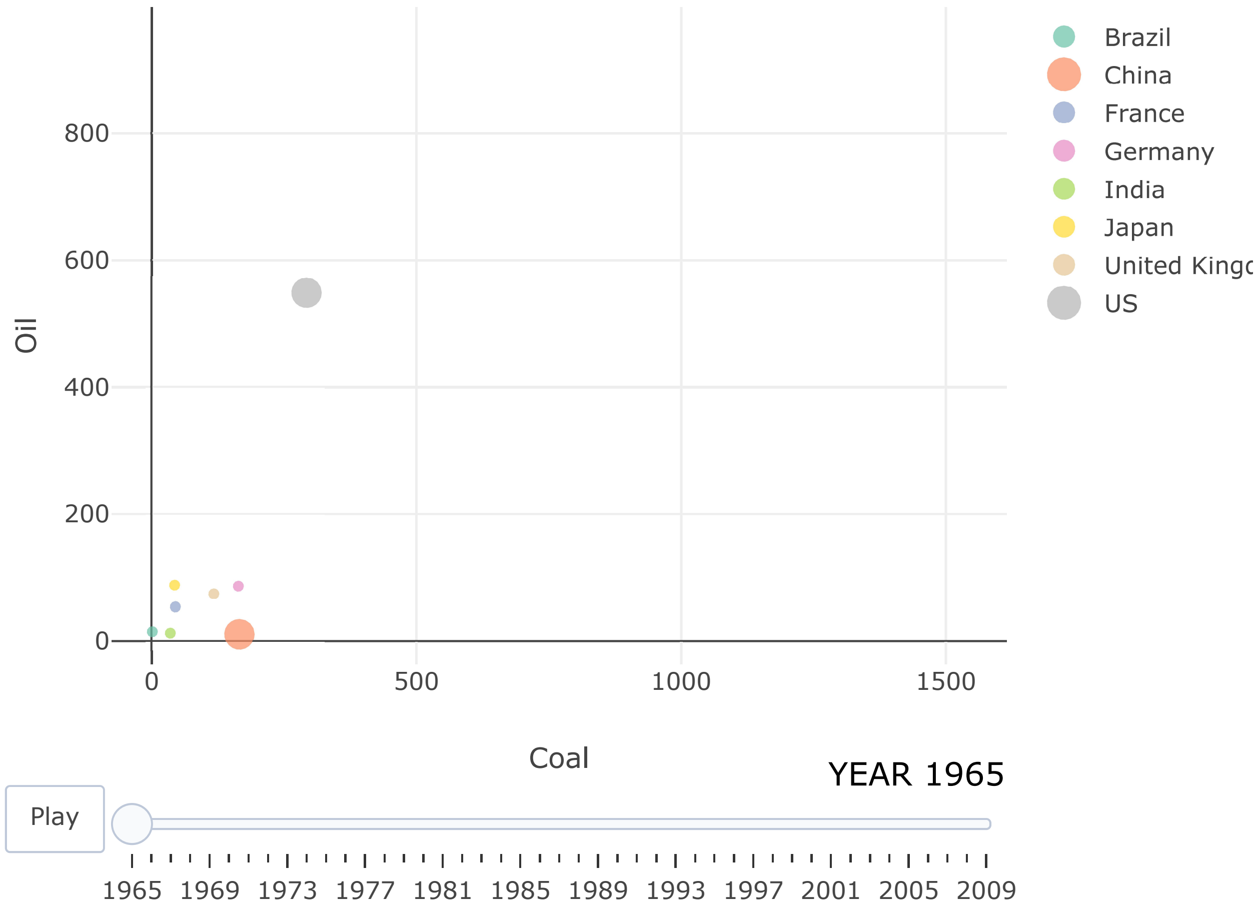
The cluster that is most prominent identified here is same identified in step 1 and step 3.That is the one containing Jamal zougham,Mohamed chaoui etc.

Assignment 2

The consumption of oil and coal in many countries within a time period is given in the oilcoal.csv.

1.plotly visualisation

```
library(tidyr)
library(tourr)
oilcoal<-read.csv("Oilcoal.csv",header=T,sep=";",stringsAsFactors = F)
oilcoal<-oilcoal[,c(1:5)]
oilcoal$Coal<-as.numeric(gsub(",",".",oilcoal$Coal))
oilcoal$Oil<-as.numeric(gsub(",",".",oilcoal$Oil))
oilcoal$Marker.size<-as.numeric(gsub(",",".",oilcoal$Marker.size))
#Visualize data in Plotly as an animated bubble chart of Coal versus Oil
#in which the bubble size corresponds to the country size.
#List several noteworthy features of the investigated animation.

###1

base<-oilcoal%>%plot_ly(x=~Coal,y=~Oil,size=~Marker.size,text=~Country,hoverinfo="t
ext")%>%
  add_markers(color=~Country,frame=~Year,ids=~Country)%>%
  animation_opts(8,redraw = F)%>%animation_slider(
    currentvalue = list(prefix = "YEAR ", font = list(color="black")))
base
```

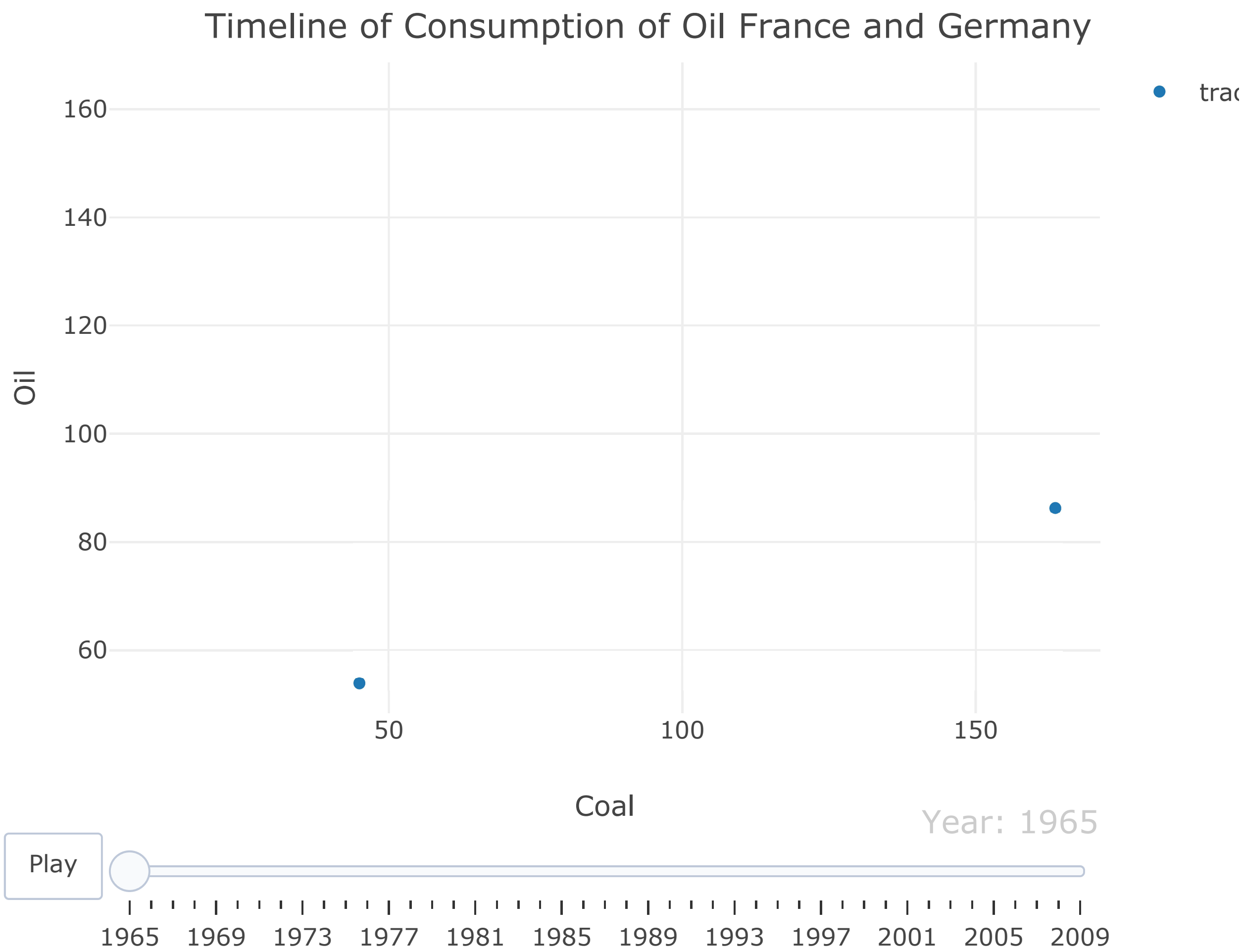


We can say like three countries grab our attention in the animation mostly.US,China and India.We could see that in the start of the animation,that is year 1965,the US had the highest oil and coal consumption ,then germany,uk,britain and china.Coal is a dwindling resource.May be the europen countries and US have started coal mining earlier and over a period of time they would have found it is a bad souce of energy becuase of the bad working condition of miners,price increased as the coal resource decreases and the pollution.SO they might have opted other cleaner source of energy, Us has a high consumtion of oil may be because the population has tripled or so by 2007 and hence domestic consumption increases.

We think the increased consumption of oil and coal and oil in India and china are related to the decreased consumption in other countries.As time progressed India,china,Japan etc became industrialised and other countries have outsourced everyhting to these countries.We can say that producton is very less in the europe and US now.They import a lot from India and China.As Industries increases the consumption becomes high.Ofcourse the domestic need in India and china has increased is another thing because their polpulation is way up the ladder when compared to 1968.

2.Motion chart

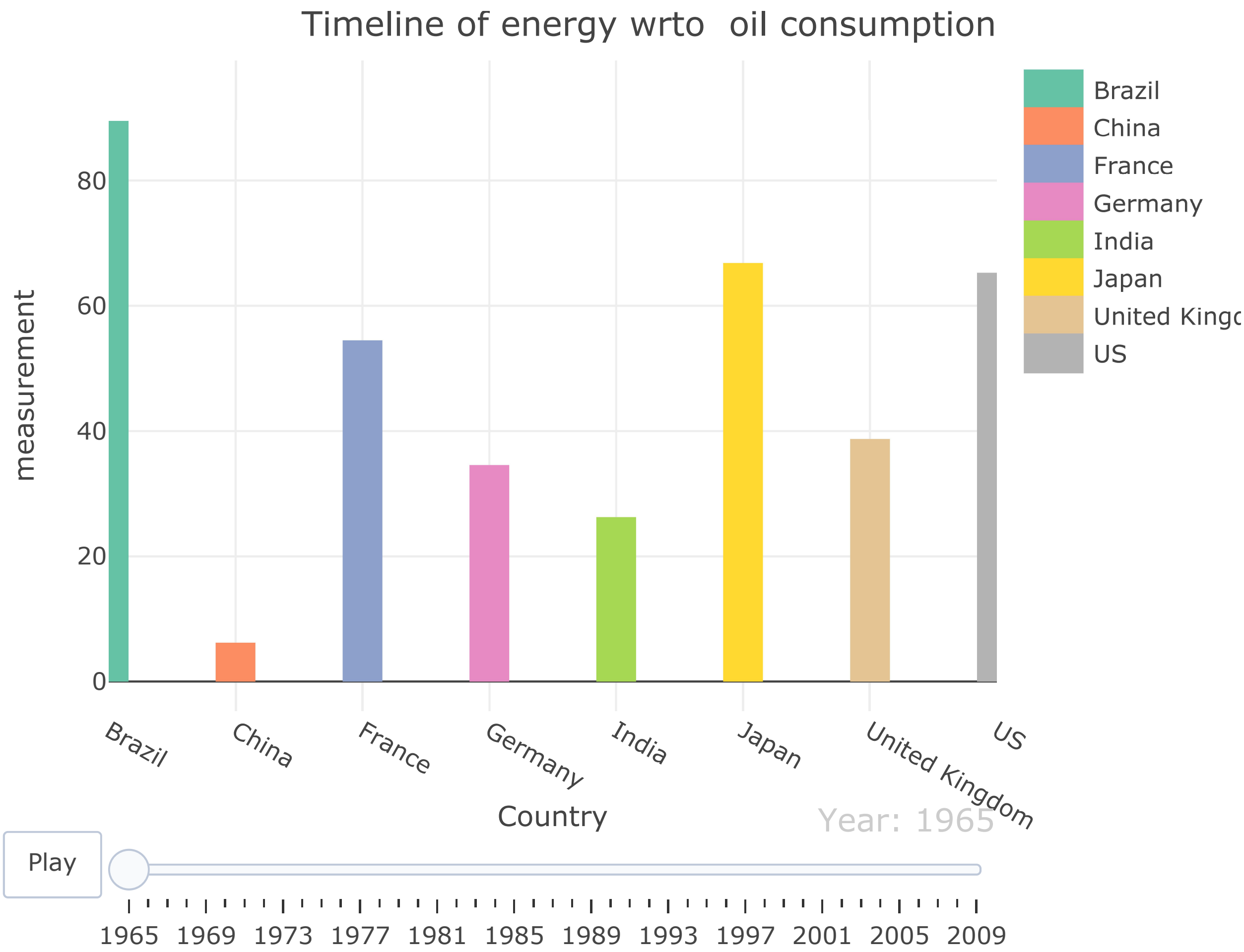
```
oilcoal %>% filter(Country %in% c("France", "Germany")) %>% plot_ly(x=~Coal, y=~Oi
l, frame =~Year, type = 'scatter', text = ~Country, mode = 'markers') %>%
  animation_opts(100, easing = "cubic", redraw = F) %>% layout(title="Timeline of C
onsumption of Oil France and Germany")
```



ONe reason could be that they understood they need to focus on renewable sources of energy and another would be the industires were shifted to the developing countries like china and India.

3.

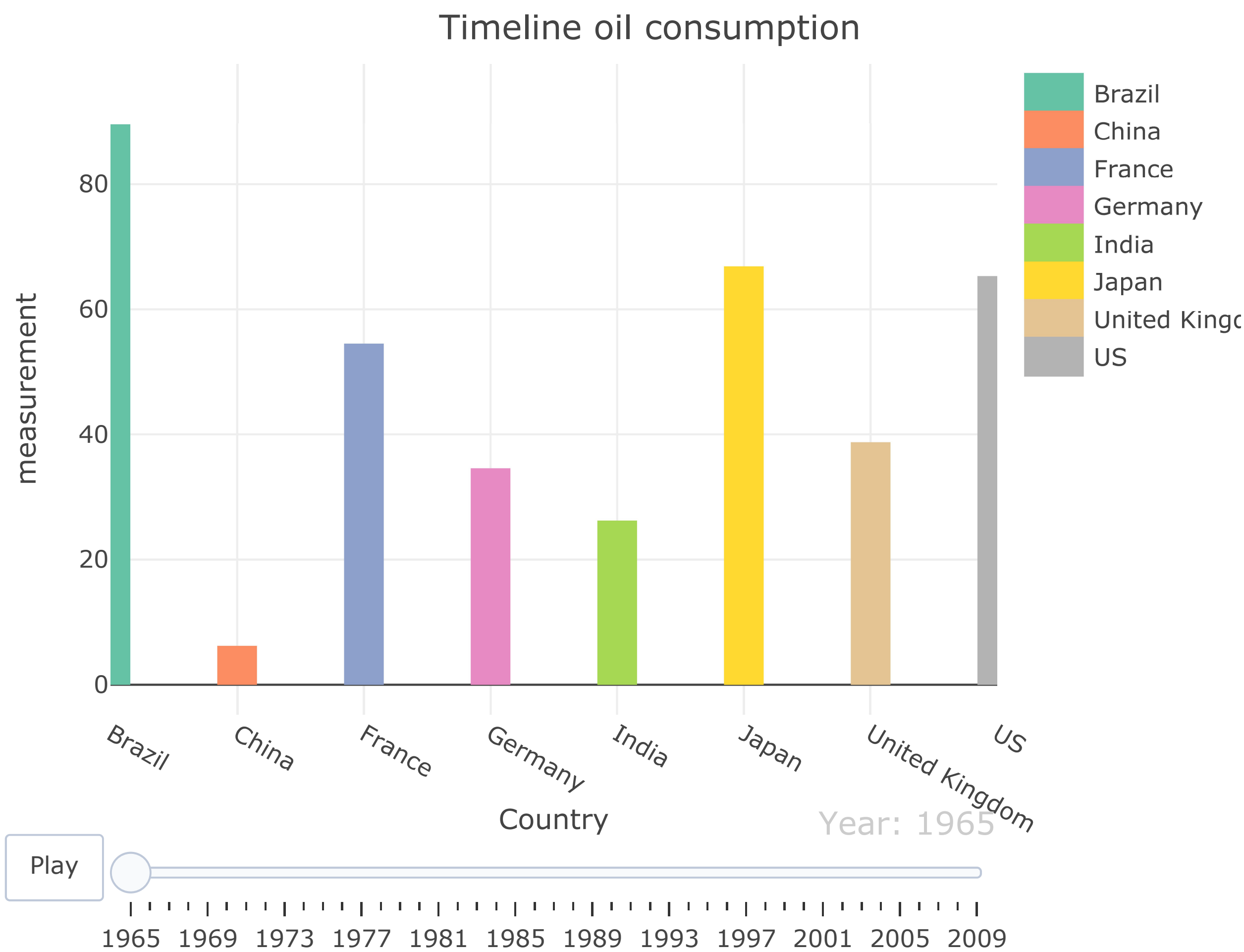
```
oilcoal$oilprop<-100*oilcoal$Oil/(oilcoal$Oil+oilcoal$Coal)

oilcoal$oilprop_0<-0

melt_oil<-gather(oilcoal, condition, measurement,oilprop,oilprop_0)


melt_oil %>%  plot_ly(x=~Country, y=~measurement, frame =~Year, type = 'scatter', m
ode='lines', text = ~Country,  color = ~Country, line = list(width = 20)) %>%
  animation_opts(300, easing = "cubic", redraw = F) %>% layout(title="Timeline of e
nergy wrto  oil consumption")
```



When we use the line plot it is easier to comapre because the length will help as to visualize easier.So the variations can also be visualised better using this plot.But it does not take into account the oilconsumption and the coalconsumption seperately.Or in bubble chart the x and y axis had oil and coal consumption respectively.But in the line plot together of the oil and coal is given as measurement in the y axis.

4.

```
melt_oil %>%  plot_ly(x=~Country, y=~measurement, frame =~Year, type = 'scatter', m
ode='lines', text = ~Country,  color = ~Country, line = list(width = 20)) %>%
  animation_opts(300, easing = "elastic", redraw = F) %>% layout(title="Timeline oi
l consumption")
```

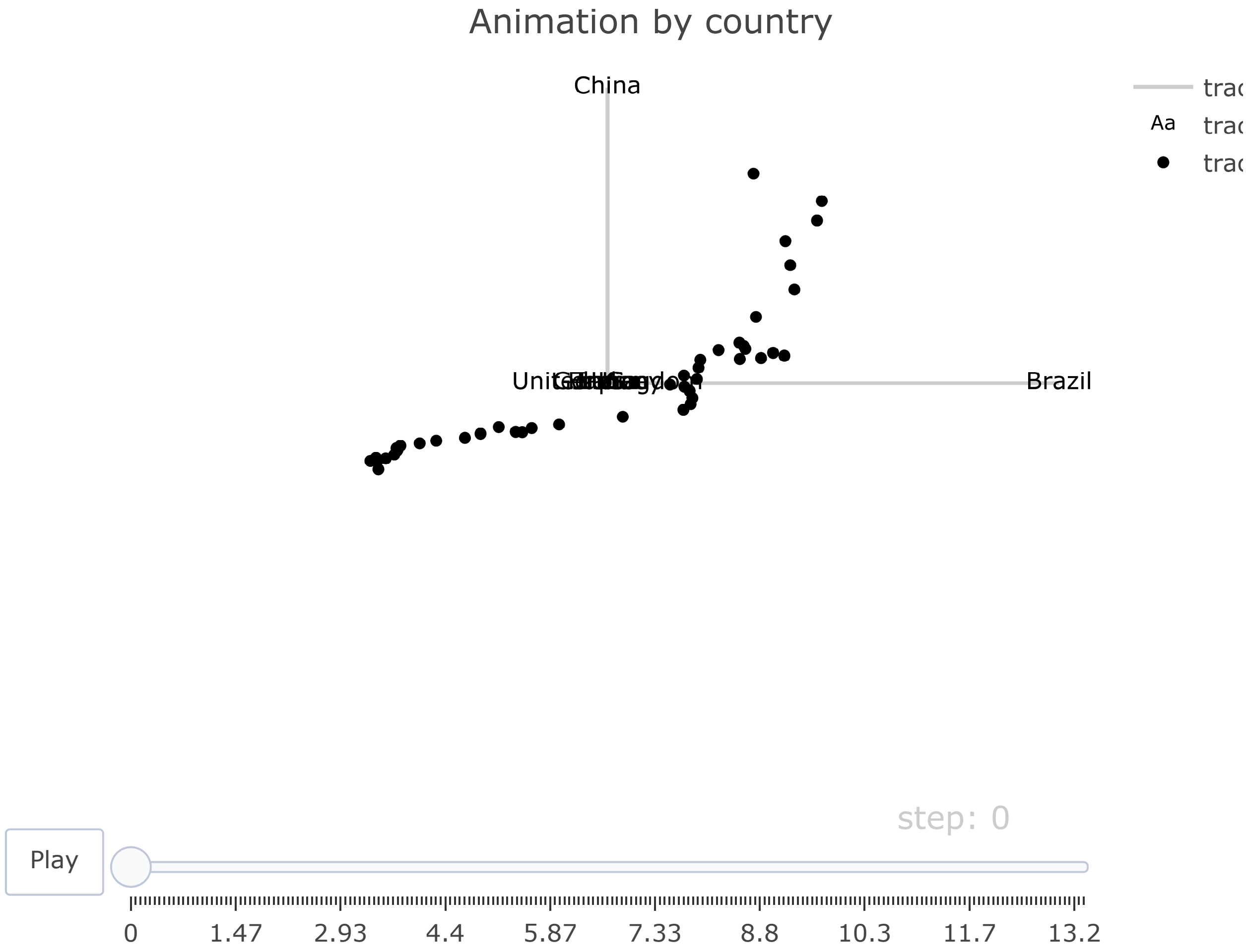


The transitions are visible more clerly here we think.

5.

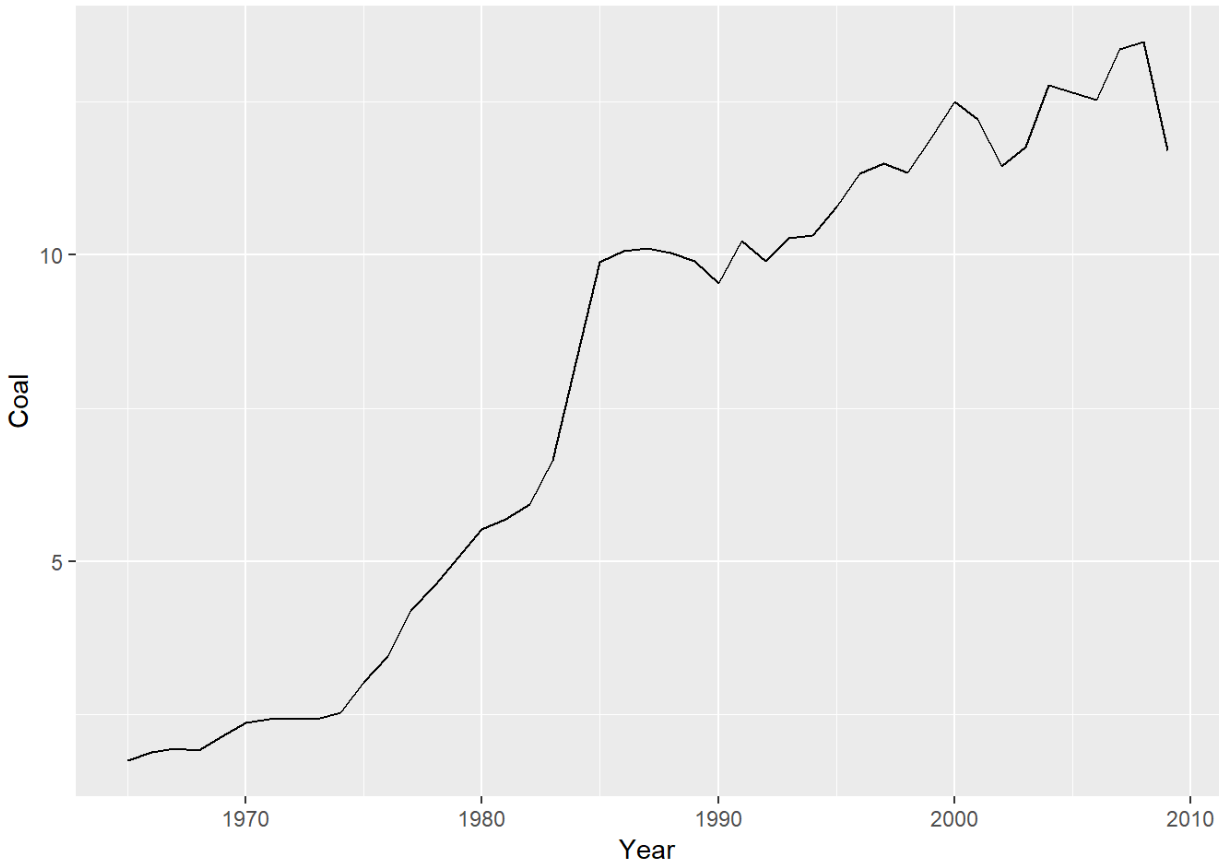
```
oilcoal_c <- oilcoal[, c("Country", "Year", "Coal")]
oilcoal_tour <- oilcoal_c %>%spread(Country, Coal)
oilcoal_scale <- rescale(oilcoal_tour[, 2:9])

rownames(oilcoal_scale) <- oilcoal_tour$Year
colnames(oilcoal_scale) <- names(oilcoal_tour)[-1]

set.seed(12345)
tour <- new_tour(oilcoal_scale, grand_tour(), NULL)
#tour<- new_tour(mat, guided_tour(cmass), NULL)

steps <- c(0, rep(1/15, 200))


Projs<-lapply(steps, function(step_size){
  step <- tour(step_size)
  if(is.null(step)) {
    .GlobalEnv$tour<- new_tour(oilcoal_scale, guided_tour(cmass), NULL)
    step <- tour(step_size)
  }
  step
}
)

# projection of each observation
tour_dat <- function(i) {
  step <- Projs[[i]]
  proj <- center(oilcoal_scale %*% step$proj)
  data.frame(x = proj[,1], y = proj[,2], state = rownames(oilcoal_scale))
}
# projection of each variable's axis
proj_dat <- function(i) {
  step <- Projs[[i]]
  data.frame(
    x = step$proj[,1], y = step$proj[,2], variable = colnames(oilcoal_scale)
  )
}
stepz <- cumsum(steps)
# tidy version of tour data
tour_dats <- lapply(1:length(steps), tour_dat)
tour_datz <- Map(function(x, y) cbind(x, step = y), tour_dats, stepz)
tour_dat <- dplyr::bind_rows(tour_datz)

# tidy version of tour projection data
proj_dats <- lapply(1:length(steps), proj_dat)
proj_datz <- Map(function(x, y) cbind(x, step = y), proj_dats, stepz)
proj_dat <- dplyr::bind_rows(proj_datz)

ax <- list(
  title = "", showticklabels = FALSE,
  zeroline = FALSE, showgrid = FALSE,
  range = c(-1.1, 1.1)
```

```
)
# for nicely formatted slider labels
options(digits = 3)
tour_dat <- highlight_key(tour_dat, ~state, group = "A")
tour <- proj_dat %>%
  plot_ly(x = ~x, y = ~y, frame = ~step, color = I("black")) %>%
  add_segments(xend = 0, yend = 0, color = I("gray80")) %>%
  add_text(text = ~variable) %>%
  add_markers(data = tour_dat, text = ~state, ids = ~state, hoverinfo = "text") %>%
  layout(xaxis = ax, yaxis = ax, title = "Animation by country")
tour
```



Yes the clusters correspond to different year ranges.We think brazil has the largest contribution to this projection

```
oilcoal%>%filter(Country=="Brazil")%>%ggplot(aes(y=Coal,x=Year))+geom_line()
```



Coal
10
5
1970
1980
1990
2000
2010
Year

Appendix

```
library(ggraph)
library(igraph)
library(visNetwork)
library(seriation)
library(plotly)

#setwd("C:/Users/quartermaine/Documents/Visualization/lab_6")
edges <- read.delim("trainData.dat", header = FALSE, sep  = " ")
nodes <- read.delim("trainMeta.dat", header = FALSE, sep = " ")
set.seed(12345)

nodes$id <- rownames(nodes)
colnames(nodes) <- c("bombers", "b_group", "id")
colnames(edges) <- c("temp", "from", "to", "value")
edges$temp <- NULL
graph <- graph.data.frame(edges, directed = T)
degree_value <- degree(graph)
nodes$value <- degree_value[match(nodes$id, names(degree_value))]
nodes <- na.omit(nodes)
nodes$label<-nodes$bombers
#### a.Basic
visNetwork(nodes=nodes,edges=edges,main = "Madrid bombing people network")
nodes$group<-nodes$b_group

visNetwork(nodes=nodes,edges=edges,main = "Madrid bombing people network")
visNetwork(nodes = nodes, edges = edges, main = "Madrid bombing people network") %
>%
  visGroups(groupname = "0", color = "blue") %>%
  visGroups(groupname = "1", color = "orange") %>%
  visEdges(arrows = "from") %>%
  visOptions(collapse = TRUE,
             selectedBy = "group") %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome()
visNetwork(nodes = nodes, edges = edges,main = "Madrid bombing people network") %
>%
  visGroups(groupname = "0", color = "blue") %>%
  visGroups(groupname = "1", color = "orange") %>%
  visEdges(arrows = "to") %>%
  visOptions(highlightNearest = list(enabled =TRUE,
                                     degree = 1),
             collapse = TRUE,
             selectedBy = "group") %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome()


visNetwork(nodes = nodes, edges = edges) %>%
  visGroups(groupname = "0", color = "blue") %>%
  visGroups(groupname = "1", color = "orange") %>%
  visEdges(arrows = "from") %>%
  visOptions(highlightNearest = list(enabled =TRUE,
```

```
                                    degree = list(from = 1, to = 2)),
             collapse = TRUE,
             selectedBy = "group") %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome()

graph_data_frame <- graph.data.frame(edges, directed = FALSE)
clusters <- cluster_edge_betweenness(graph_data_frame, directed = T)
nodes$clusters <- clusters$membership

visNetwork(nodes = nodes, edges = edges, main = "Madrid Bombing") %>%
  visEdges(arrows = "from") %>%
  visOptions(highlightNearest = list(enabled =TRUE,
                                     degree = list(from = 1, to = 2)),
             collapse = TRUE,
             selectedBy = "group")  %>%
  visPhysics(solver= "repulsion") %>%
  visLegend() %>% addFontAwesome() %>%visIgraphLayout()

adjecency <- get.adjacency(graph_data_frame, sparse=F)
colnames(adjecency) <- nodes$label
rownames(adjecency) <- nodes$label
rowdist<-dist(adjecency)


reord<-get_order(seriate(rowdist, "HC"))
mat_value<-adjecency[reord,reord]

##SInce using ubuntu system using the below line of code.




plot_ly(z=~mat_value, x=~colnames(mat_value),
        y=~rownames(mat_value), type="heatmap") %>% layout(title = "Heatmap")
library(tidyr)
library(tourr)
oilcoal<-read.csv("Oilcoal.csv",header=T,sep=";",stringsAsFactors = F)
oilcoal<-oilcoal[,c(1:5)]
oilcoal$Coal<-as.numeric(gsub(",",".",oilcoal$Coal))
oilcoal$Oil<-as.numeric(gsub(",",".",oilcoal$Oil))
oilcoal$Marker.size<-as.numeric(gsub(",",".",oilcoal$Marker.size))
#Visualize data in Plotly as an animated bubble chart of Coal versus Oil
#in which the bubble size corresponds to the country size.
#List several noteworthy features of the investigated animation.

###1

base<-oilcoal%>%plot_ly(x=~Coal,y=~Oil,size=~Marker.size,text=~Country,hoverinfo="t
ext")%>%
  add_markers(color=~Country,frame=~Year,ids=~Country)%>%
  animation_opts(8,redraw = F)%>%animation_slider(
    currentvalue = list(prefix = "YEAR ", font = list(color="black")))
```

```
base

oilcoal %>% filter(Country %in% c("France", "Germany")) %>% plot_ly(x=~Coal, y=~Oi
l, frame =~Year, type = 'scatter', text = ~Country, mode = 'markers') %>%
  animation_opts(100, easing = "cubic", redraw = F) %>% layout(title="Timeline of C
onsumption of Oil France and Germany")

oilcoal$oilprop<-100*oilcoal$Oil/(oilcoal$Oil+oilcoal$Coal)

oilcoal$oilprop_0<-0

melt_oil<-gather(oilcoal, condition, measurement,oilprop,oilprop_0)


melt_oil %>%  plot_ly(x=~Country, y=~measurement, frame =~Year, type = 'scatter', m
ode='lines', text = ~Country,  color = ~Country, line = list(width = 20)) %>%
  animation_opts(300, easing = "cubic", redraw = F) %>% layout(title="Timeline of e
nergy wrto  oil consumption")

melt_oil %>%  plot_ly(x=~Country, y=~measurement, frame =~Year, type = 'scatter', m
ode='lines', text = ~Country,  color = ~Country, line = list(width = 20)) %>%
  animation_opts(300, easing = "elastic", redraw = F) %>% layout(title="Timeline o
f wrto and oil consumption")
oilcoal_c <- oilcoal[, c("Country", "Year", "Coal")]
oilcoal_tour <- oilcoal_c %>%spread(Country, Coal)
oilcoal_scale <- rescale(oilcoal_tour[, 2:9])

rownames(oilcoal_scale) <- oilcoal_tour$Year
colnames(oilcoal_scale) <- names(oilcoal_tour)[-1]

set.seed(12345)
tour <- new_tour(oilcoal_scale, grand_tour(), NULL)
#tour<- new_tour(mat, guided_tour(cmass), NULL)

steps <- c(0, rep(1/15, 200))


Projs<-lapply(steps, function(step_size){
  step <- tour(step_size)
  if(is.null(step)) {
    .GlobalEnv$tour<- new_tour(oilcoal_scale, guided_tour(cmass), NULL)
    step <- tour(step_size)
  }
  step
}
)

# projection of each observation
tour_dat <- function(i) {
  step <- Projs[[i]]
  proj <- center(oilcoal_scale %*% step$proj)
  data.frame(x = proj[,1], y = proj[,2], state = rownames(oilcoal_scale))
```

```r
}
# projection of each variable's axis
proj_dat <- function(i) {
  step <- Projs[[i]]
  data.frame(
    x = step$proj[,1], y = step$proj[,2], variable = colnames(oilcoal_scale)
  )
}
stepz <- cumsum(steps)
# tidy version of tour data
tour_dats <- lapply(1:length(steps), tour_dat)
tour_datz <- Map(function(x, y) cbind(x, step = y), tour_dats, stepz)
tour_dat <- dplyr::bind_rows(tour_datz)

# tidy version of tour projection data
proj_dats <- lapply(1:length(steps), proj_dat)
proj_datz <- Map(function(x, y) cbind(x, step = y), proj_dats, stepz)
proj_dat <- dplyr::bind_rows(proj_datz)

ax <- list(
  title = "", showticklabels = FALSE,
  zeroline = FALSE, showgrid = FALSE,
  range = c(-1.1, 1.1)
)
# for nicely formatted slider labels
options(digits = 3)
tour_dat <- highlight_key(tour_dat, ~state, group = "A")
tour <- proj_dat %>%
  plot_ly(x = ~x, y = ~y, frame = ~step, color = I("black")) %>%
  add_segments(xend = 0, yend = 0, color = I("gray80")) %>%
  add_text(text = ~variable) %>%
  add_markers(data = tour_dat, text = ~state, ids = ~state, hoverinfo = "text") %>%
  layout(xaxis = ax, yaxis = ax, title = "Animation by country")
tour


 oilcoal%>%filter(Country=="Brazil")%>%ggplot(aes(y=Coal,x=Year))+geom_line()
```