Lab-6

Group22 - balra340, tejma768

November 23, 2018

Assignment1

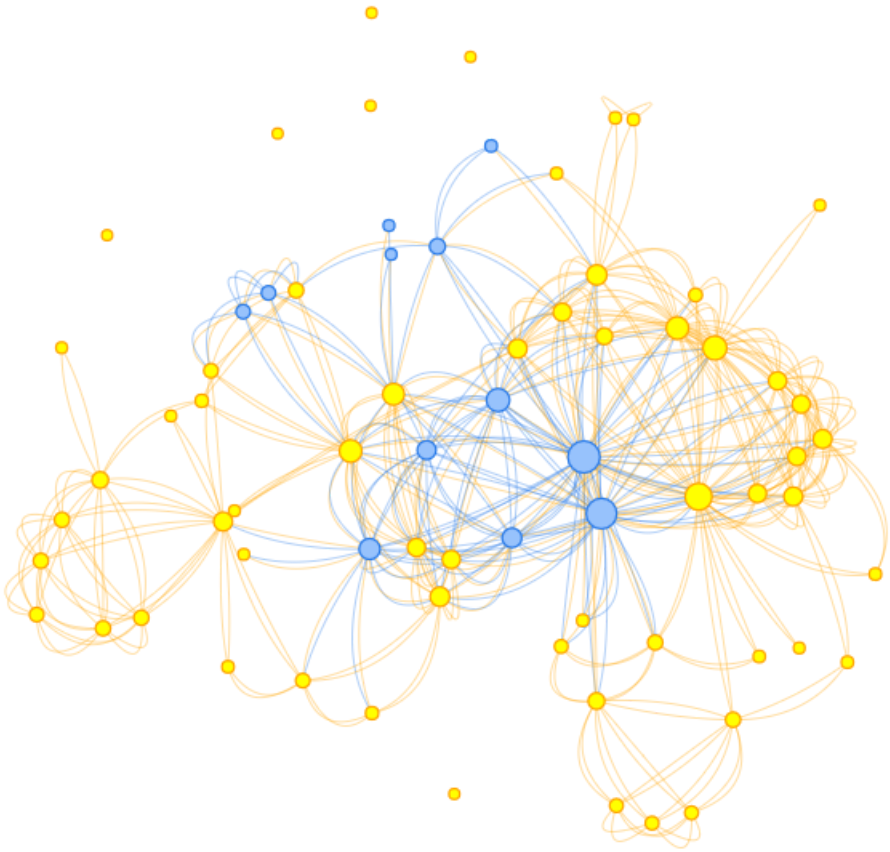
Question 1

step 1

we are supposed to make a graph using visNetwork

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

#1.
#making nodes

nodes1 <- read.table("trainMeta.dat" , header = F)

nodes1$id <- 1:nrow(nodes1)

colnames(nodes1) <- c("label","group","id")

nodes1 <- nodes1[c("id","group","label")]


#making links

links1 <- read.table("trainData.dat", header=F, as.is=F)

colnames(links1) <- c("from","to","strength")


g <- graph_from_data_frame(d=links1, vertices=nodes1, directed=T)

nodes1$value <- strength(g)


result <- visNetwork(nodes1,links1) %>% visPhysics(solver='repulsion') %>%
          visOptions(highlightNearest = list(enabled = TRUE, degree = 1,
          labelOnly = FALSE, hover = TRUE), nodesIdSelection = TRUE)

result
```

Select by id ▼



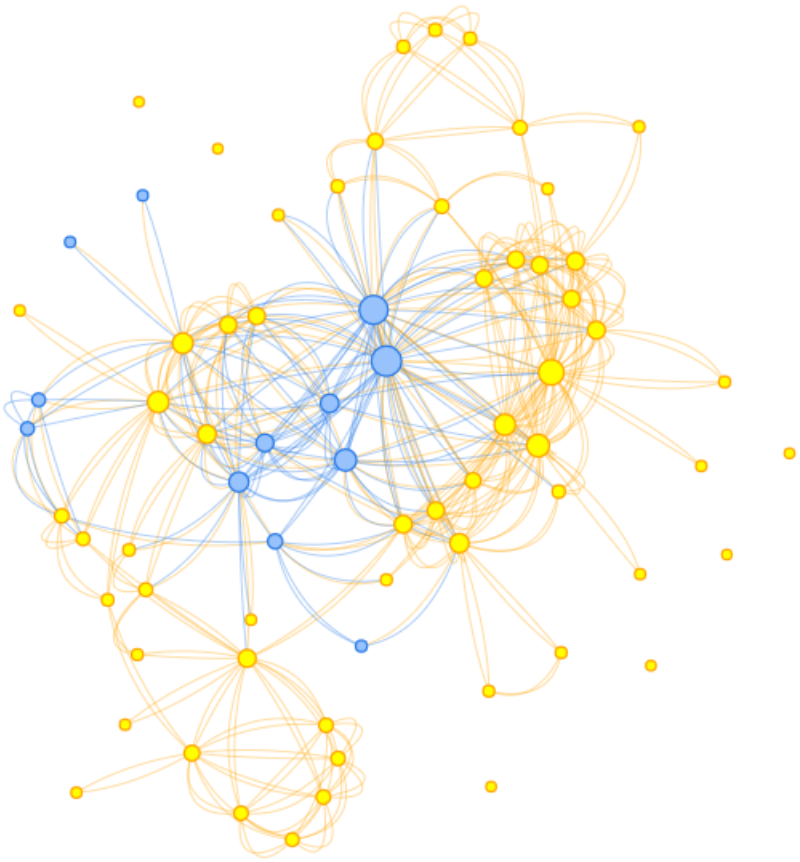
We are able to find 3 clusters. Contacts of Said berrack, Amer Azizi and Mohamed Chaoui seems to form a big cluster, Second cluster at the left is formed by Hamid Ahmidan and Naima Oulad Akcha. Third cluster formed at the right is of connections of Jose Emilio Suarez and Tayisir Alouny.

step 2

Now we are supposed to add the functionality to the previous plot. That highlight all the nodes that are connected to the selected node by length 1 or 2.

```
result2 <- visNetwork(nodes1,links1) %>% visPhysics(solver='repulsion') %>%
          visOptions(highlightNearest = list(enabled = TRUE, degree = 1,labelOnly = FALSE, hover
= TRUE), nodesIdSelection = TRUE)

result2
```

Select by id ▼

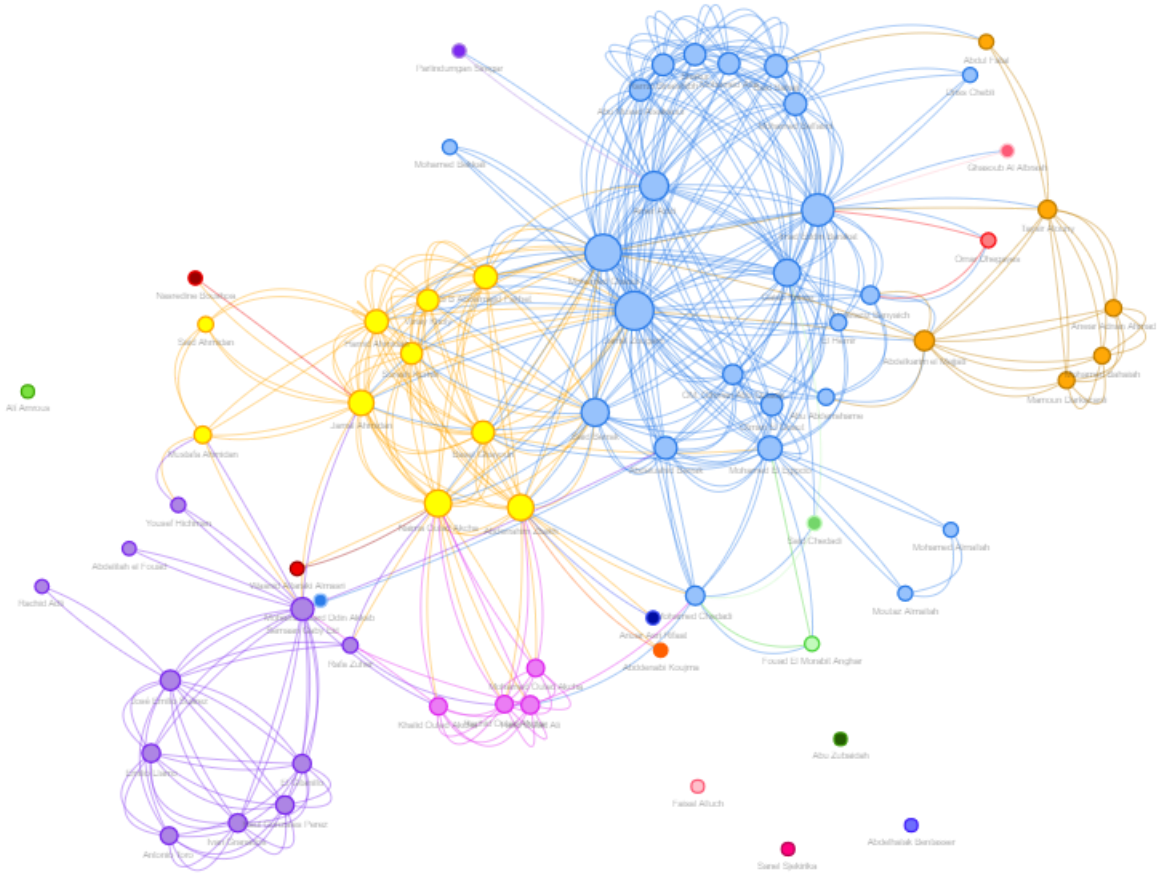


We feel that Said Berrak and Jamal Zougam can actually spread the information faster as their connections seem to form larger nodes. Jamal Zougam had a mobile store near Madrid, hence was suspected to be the person who sold telephones that were used to blow up bombs in the Madrid Attack. As all his connections are large nodes. He has a huge role in the madrid attack.

step 3

Now we compute clusters by optimising edge betweenness.

```
new_nodes1 <- nodes1
net <- graph_from_data_frame(d=links1, vertices=nodes1, directed=F)
ceb <- cluster_edge_betweenness(net)
new_nodes1$group=ceb$membership

result3 <- visNetwork(new_nodes1,links1) %>% visPhysics(solver='repulsion') %>%
  visOptions(highlightNearest = list(enabled = TRUE, degree = 1,labelOnly = FALSE, hover = TRUE
))

result3
```

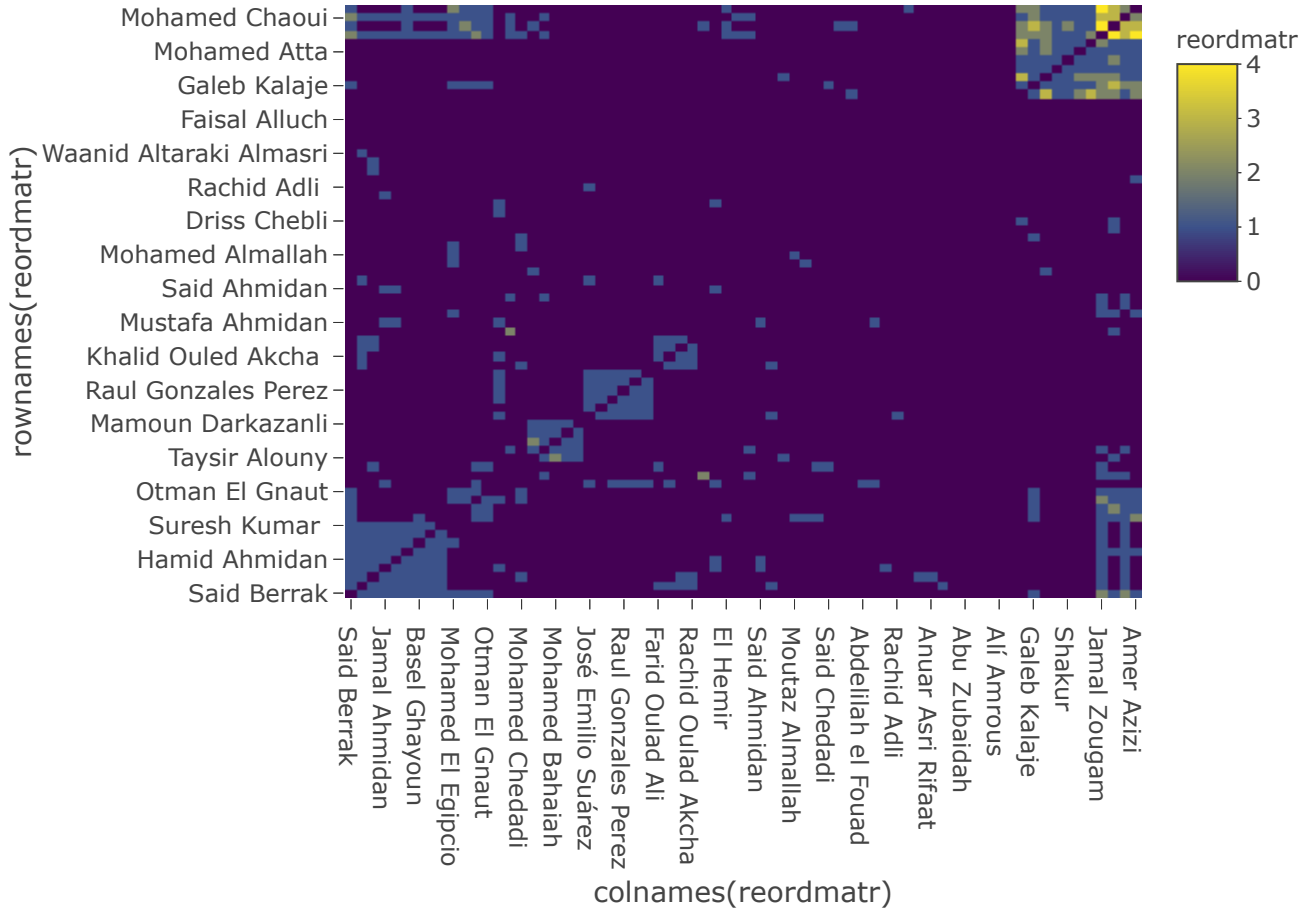


More clusters are visible in this plot compared to part 1. Yes, we are able to find the same clusters as the ones that were found in 1st part. Clusters formed by Said Berrack, Jamal Zougam, Mohamed Chaoui, Amer Azizi and Imad Eddin Barakat were also present in part 1.

step 4

Using adjacency matrix representation we make a heat map in this part of question.

```
netm <- get.adjacency(net, attr="strength", sparse=F)



colnames(netm) <- V(net)$label
rownames(netm) <- V(net)$label

rowdist<-dist(netm)

library(seriation)
order1<-seriate(rowdist, "HC")
ord1<-get_order(order1)

reordmatr<-netm[ord1,ord1]

library(plotly)

plot_ly(z=~reordmatr, x=~colnames(reordmatr),
        y=~rownames(reordmatr), type="heatmap")
```

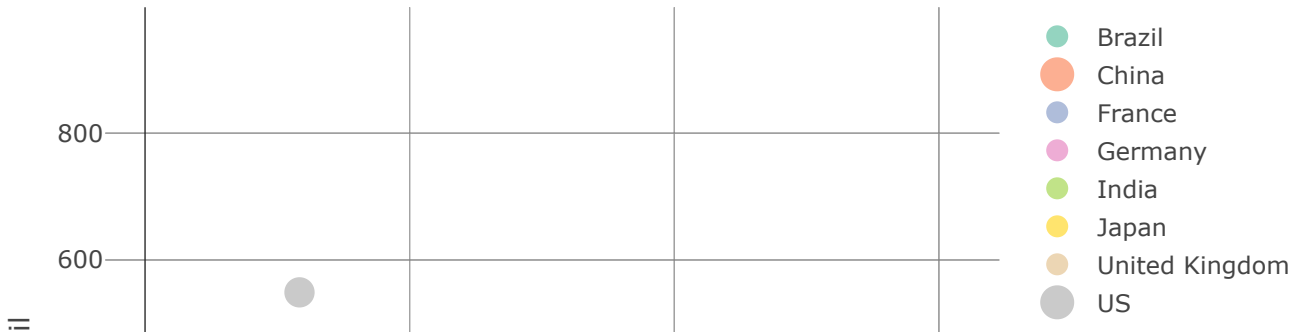


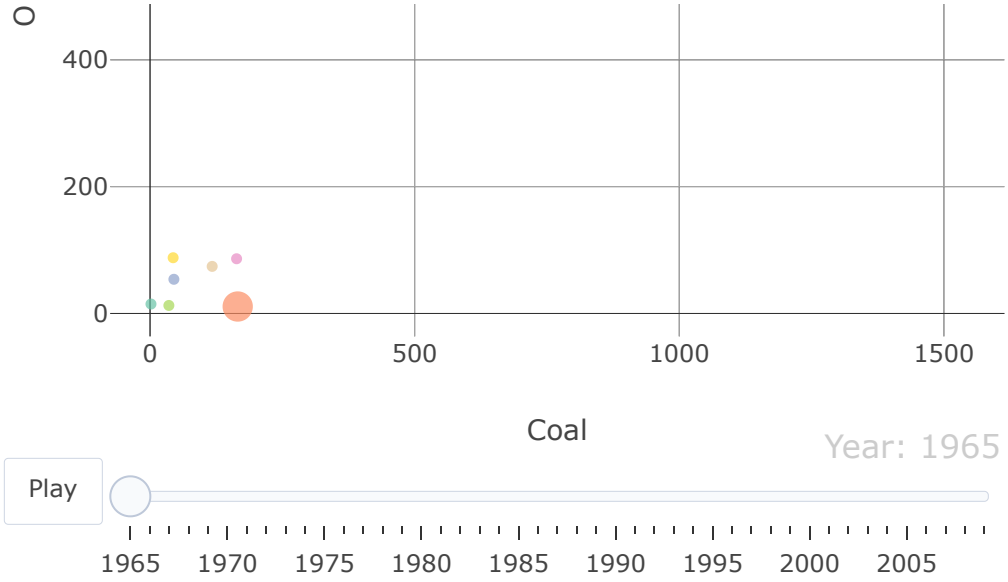
There are two main clusters. one in the top right and one in the bottom left. Both clusters were found in part3 but not in 1

Assignment 2

1. 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.

```
my_data <- read.csv2("Oilcoal.csv")

p <- plot_ly(my_data, x = ~Coal, y = ~Oil, size = ~Marker.size, color = ~Country, frame = ~Year
  ) %>%add_markers%>%animation_opts(
  100, easing = "elastic", redraw = F
)
p
```

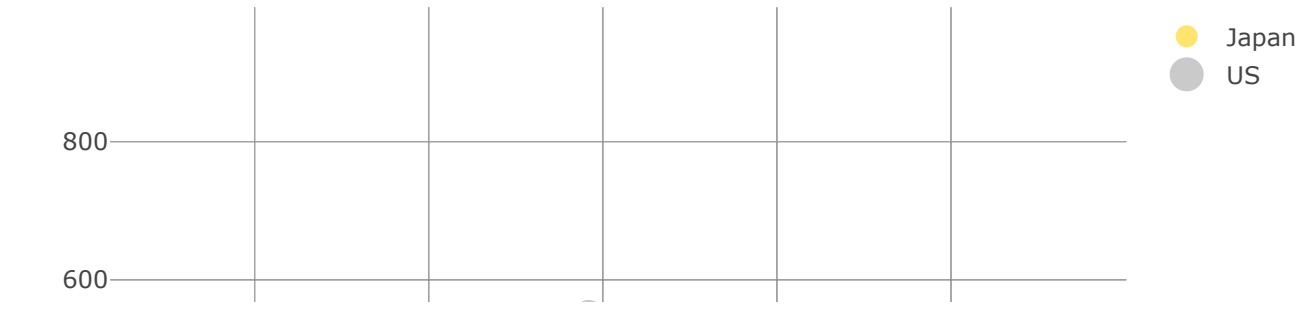


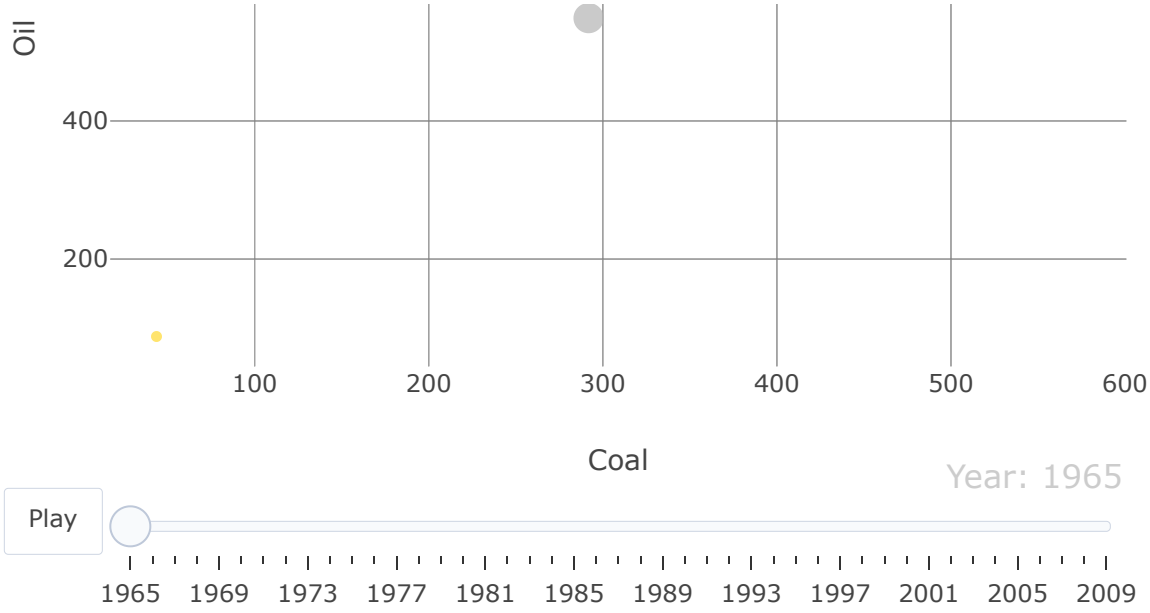


When we want to see graph for many years is when we use Animated Graphs, By looking at the bubble plot it is clear that Coal and Oil consumption has increased over the years. US seems to have increased its oil consumption from the year 1947, and the consumption level has increased from 200 to 500. And China has increased Coal consumption from the year 1986, the consumtion rate has increased from 700 to 1600

2. Find two countries that had similar motion patterns and create a motion chart including these countries only. Try to find historical facts that could explain some of the sudden changes in the animation behavior.

```
US <- filter(my_data , Country == "US")

Japan <- filter(my_data , Country == "Japan")

my_data1 <- rbind(US, Japan)

p1 <- plot_ly(my_data1,
  x = ~Coal,
  y = ~Oil,
  size = ~Marker.size,
  color = ~Country,
  frame = ~Year,
  text = ~Country,
  hoverinfo = "text"
  ) %>%add_markers%>%animation_opts(
  100, easing = "elastic", redraw = F
)
p1
```





Two countries that have similar motion patterns are Japan and US. Both Japan and US have increased consumption of oil in the previous years. With the need for use of enery resources, consumption of oil is increasing, US being the largest among the oil consumers.

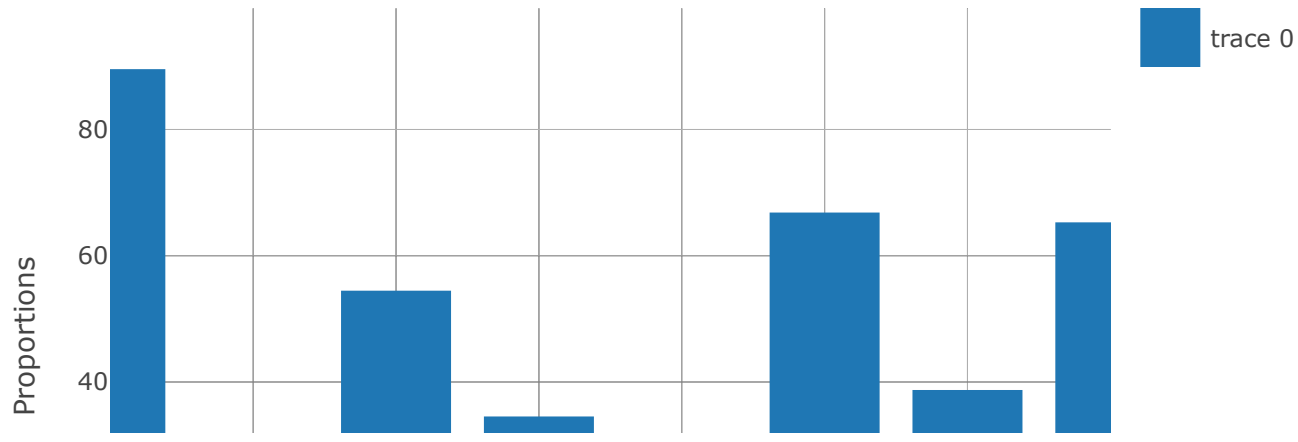
3

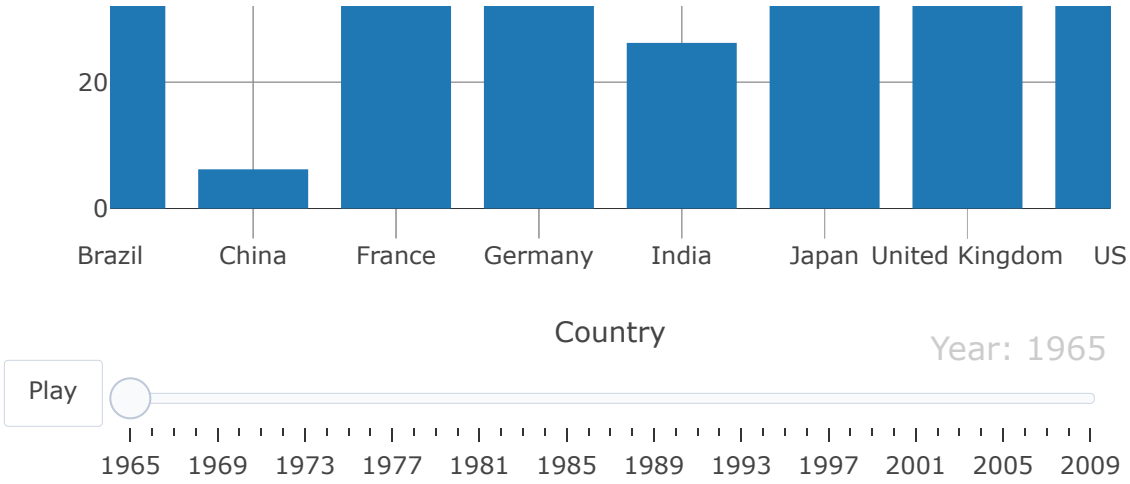
3.1 Create a new data frame that for each year and country contains two rows: one that shows oil_p and another row containing 0 in oil_p column.

```
Pdata2 <- my_data%>%group_by(Year,Country)%>%mutate(Proportions=0)
Pdata3 <- my_data%>%group_by(Year,Country)%>%mutate(Proportions=(Oil/(Oil+Coal))*100)
bar_graph <- rbind(Pdata2, Pdata3)
```

3.2 Make an animated line plot of Oilp versus Country where you group lines by Country and make them thicker.

```
p2 <- bar_graph%>% plot_ly(x =  ~ Country, y =  ~ Proportions, frame =  ~ Year) %>%
add_lines(size=1)%>%animation_opts(100, easing="cubic",redraw=F)

p2
```

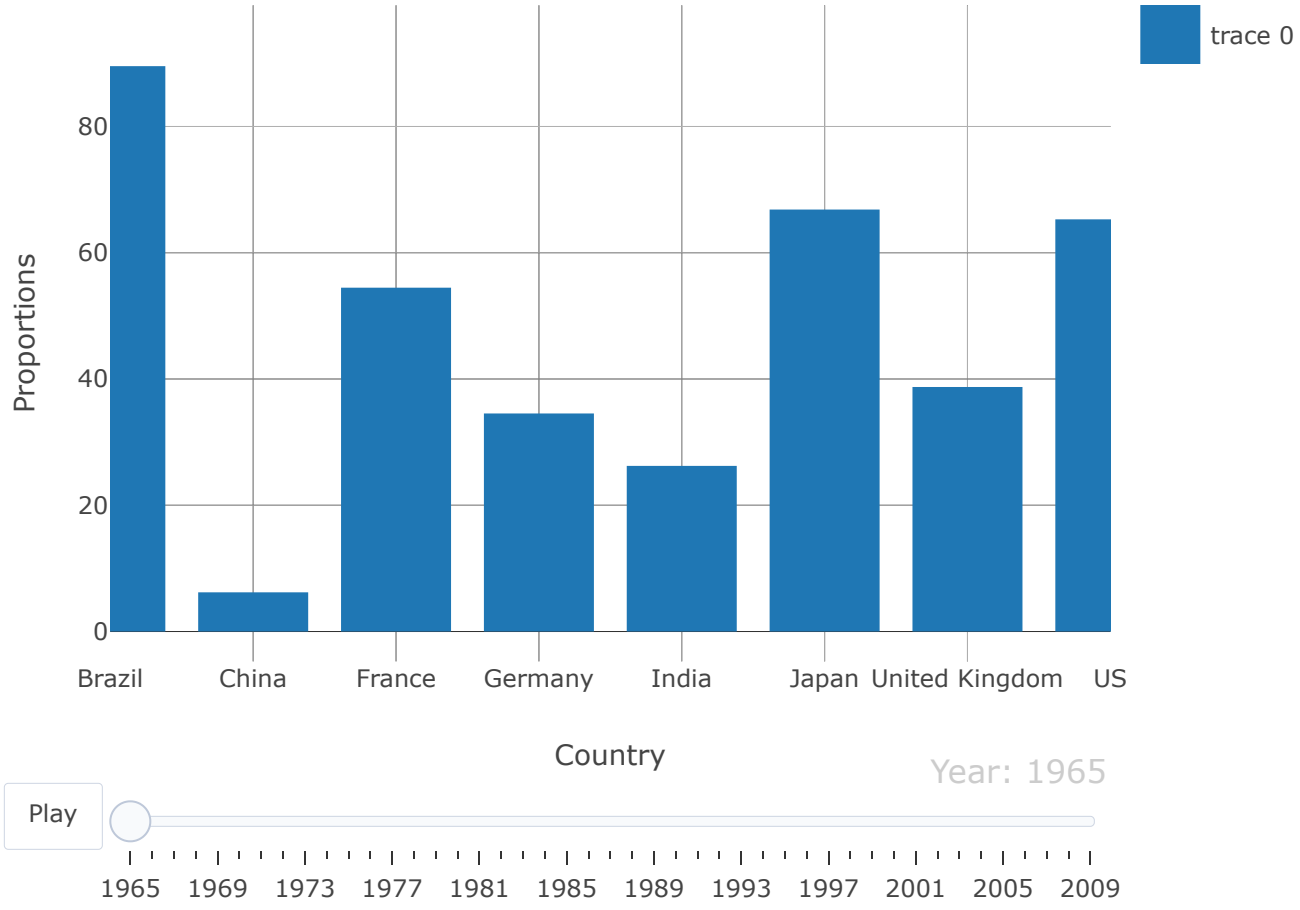




The advantages of using a Animated bubble chart is that it provides clear information regarding the Coal and Oil consumption of various contries over the years. The only disadvantage is that we can not recognize the increase by looking at the plot for once. From the Animated line plot we are easily able to identify the increase in consumption of resources.

4. Repeat the previous step but use elastic transition (easing). Which advantages and disadvantages can you see with this animation?

```
p3 <- bar_graph%>% plot_ly(x =  ~ Country, y =  ~ Proportions, frame =  ~ Year) %>%
add_lines(size=1)%>%animation_opts(300, easing="elastic",redraw=F)

p3
```

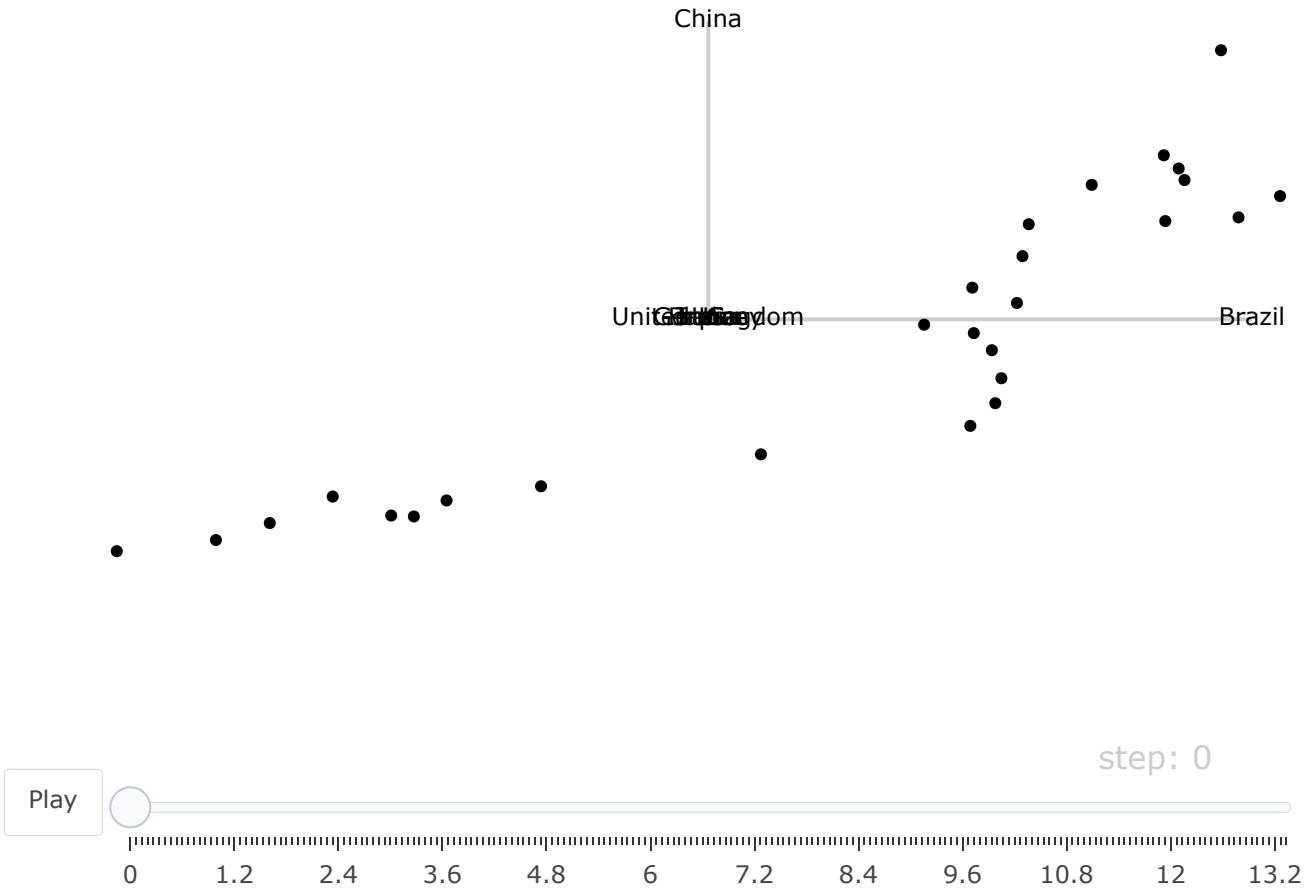


By changing the easing from cubic to elastic, thought the plot remains the same, we are able to notice stepwise transition, which makes it much easier for us to understand the increase in consumption of resources of respective Countries. The plot looks similar to easeInOutBounce, as even in easeInOutBounce we are able to notice stepwise transition. The disadvantage is that it is difficult to analyse consumption of resources per year using bar chart as there are drastic changes.

5. Use Plotly to create a guided 2D-tour visualizing Coal consumption.

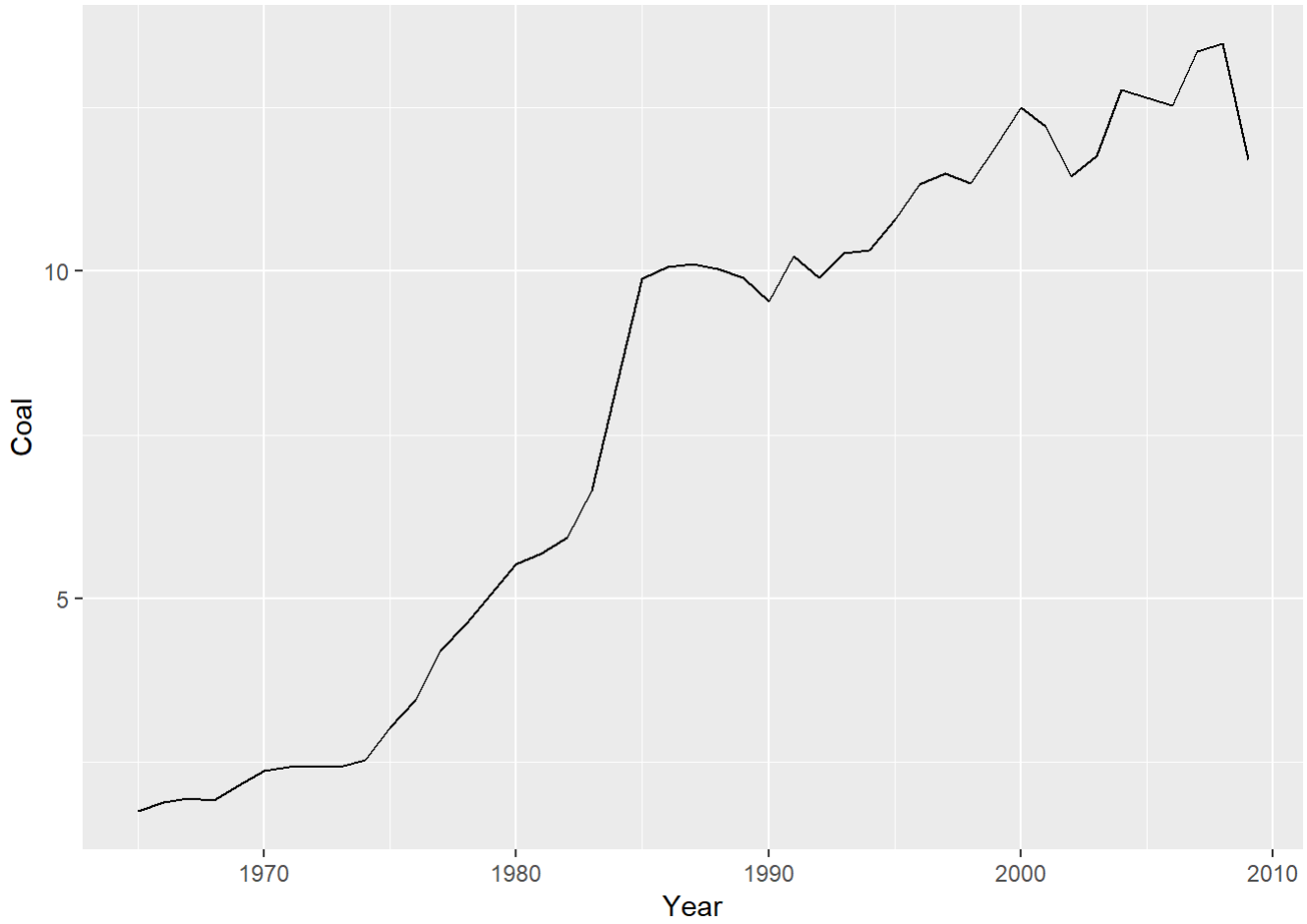
```
library(tourr)
library(plotly)
library(reshape2)

data4 <- as.matrix(my_data %>% dcast(Year~Country, value.var = "Coal"))

rownames(data4) <- data4[,"Year"]
mat <- data4[,-1] %>% scale()

set.seed(12345)
tour <- new_tour(mat, grand_tour(), 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(mat, guided_tour(cmass), NULL)
    step <- tour(step_size)
  }
  step
})

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

# 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(mat)
  )
}

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, showlegend = F)

tour
```



Yes, Clusters do correspond to different year ranges. We can see two clusters by looking at the guided tour, one at 0 and the other one immediately after 7.47. Variable Brazil has the largest contribution to this projection. The increase in coal consumption of Brazil can be interpreted by a Time Series plot of Coal consumption.

```
p5 <- my_data %>% filter(Country=="Brazil")%>%ggplot(aes(y=Coal,x=Year))+geom_line()
p5
```



Brazil data seems to be constantly rising. Brazils coal consumption has drastically increased from 1975 to 1985, from 1985 onwards there is a non uniform increase and thus having a largest contribution to the projection.