

lab5

Andreas and Priya

11 October 2018

##Assignment 1

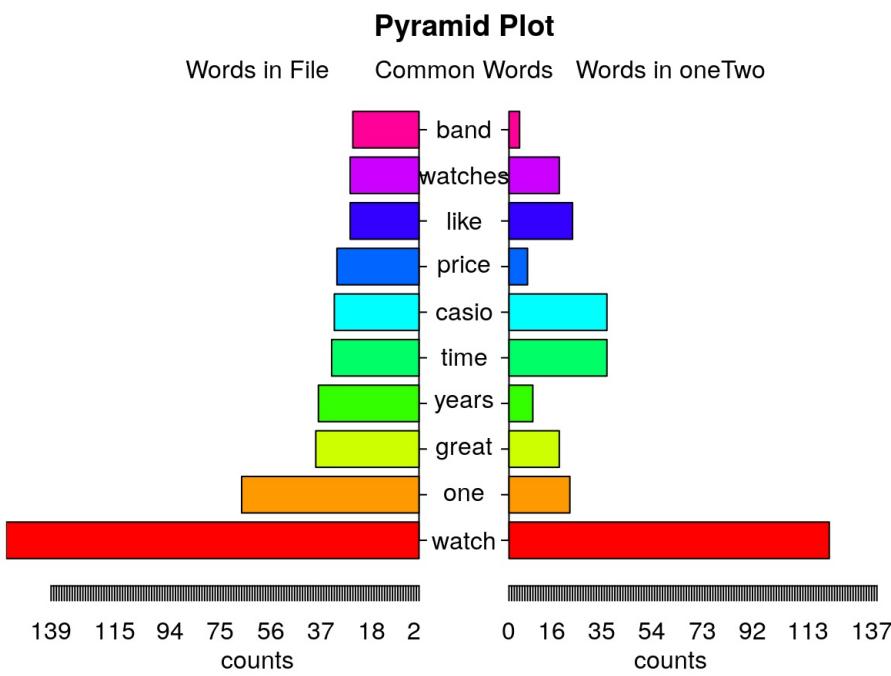
###Word Cloud Plot for five



###Word Cloud Plot for One Two



###Pyramid Plot



```
## [1] 5.1 4.1 4.1 2.1
```

The word cloud for five.txt which is the good feedback has more frequent words like watch,great,one,price,casio,years,battery etc.This makes sense as words watch and casio will be often coming in the words as this is the product.As it is a positive feedback,people are happy with the price.Great is a posisive word and may be telling it last for long with word years.

The oneTwo.txt is the negative feedback and the frequent words are amazon,one,back,watch,casio,battery etc.casio and watch comes frequent like before as this is the product.Amazon ,may be because they are not happy with the amazon more than product(may be the deal,return policy,shipping etc).They would like to return back the product and hence **back** has come multiple times.

###Phrasenets

Phrase nets of five.text

Figure 1. positive.

phrase nets of negative words

Figure 2. negative

word trees

word tree of battery with the positive feedback

battery and **replacement** changed for the period. Oh, I think the rubber wristband dried up and broke a few years. Just battery replacement for the period. Casio makes this same watch with the atomic movement, solar power. I wanted to make sure that my son would like this style and size. A couple of times. The watch repair guy said that once he opens up the watch to change the battery, it won't stay waterproof. So when he goes to swim, he just takes it off. Then the crystal broke for so 3 or 4 times. The last time the battery was changed, the jeweler told me the push button won't stay waterproof. Maybe I'm prejudiced, but for more than 25 years I've depended on my Casios and they have never failed me. The battery lasted 11 years. Replaced it 3-4 years ago still working. Its slim and love it also. Dropped it face down on concrete. One of the bands came loose. Dont drop this watch face about that many straps. I've traveled the planet, dived in three oceans with it, it has gone through change, no service, no nothing. My wife's Cartier keeps breaking and needs a 700\$ service every 5 years. goes. I just got a new one cause its rare when a watch shop makes it waterproof again after replacing the battery how many watches go thru more than one generation of batteries? I think my black one (about 15 years) every few years are minor issues. Pretty good for the money. In Australia, it cost about 3 times more. I was changed, the jeweler told me the pushers had become clogged. He cleaned them, and I wore it another 1 life around 2 years. Casio makes great watches. Also have a casio digital that came w/lithium battery. The battery lasted 11 years. Replaced it 3-4 years ago still working. Its slim and love it also. If were to spend more money, I dropped and broke it. So, I finally got a new one, the black face version this time for a change of pa

still works fine and with great accuracy after quite a few years. Just battery replacement for the period. Oh, I think the rubber wristband dried up and broke at one point after years of wearing it. I had to replace it. It is beginning to show some wear and tear - mostly scratches on the crystal. The rest of the case seems to be fine. But nice complements made about the old watch recently - never had that happen before. So that's why I got another one with the effect of free shipping. I like the price, it's a solid watch. Sturdy and durable. You never have to worry about it , like you would an expensive watch. This product has 5 stars is very nice, excellent product. i recommend it to every bahu

Figure 4. wordtree battery pos feedback

word tree with the word battery in negative feedback

battery and after that it start to be slower I dont know what the hell..... this did not fix the problem, so I returned the watch to the store and exchanged it for an id all seems well. You do need special tool to get back off of watch. I will update if it quits working. The analog portion does not keep time. I replaced the battery and this did not fix the problem, so I If you just have someone replace the battery without pressure testing it will fail. This watch was given At the least I would recommend trying this monitor on before purchasing it to make sure you are changed and it worked well for a few months, but then stopped again. Sad, because I love the look. shortly after PURCHASING this watch. After replacing the battery, it did not too long after that so I'm done. It did not too long after that so I'm done fooling with it. It's a shame because it's a really nice looking watch quit in less than 12 months and after it was replaced, the watch died completely. Looks and price are my top priority. The watch is still under warranty, but I'm just going to return it and get my money back from Casio. I had the battery replaced and the watch wouldn't keep time. I returned the watch to Casio for repair and replaced the watch wouldn't keep time. I returned the watch to Casio for repair and was told it would be \$ only last a year. It will cost more than the watch to properly replace the battery. If you just have someone replace, work fine for another couple of months, then the same problem happened again - hence the issue because it will run fine for days keeping time well and then I will look at it a day later and then it will be already. The back is designed to be opened with special tools so you cannot do it yourself making it a bit more in the first watch. Had this watch for one and a half years when the push buttons started sticking while depre

Figure 4. wordtree battery neg feedback

Analysis

Looking at the above graphs we come to conclusions like

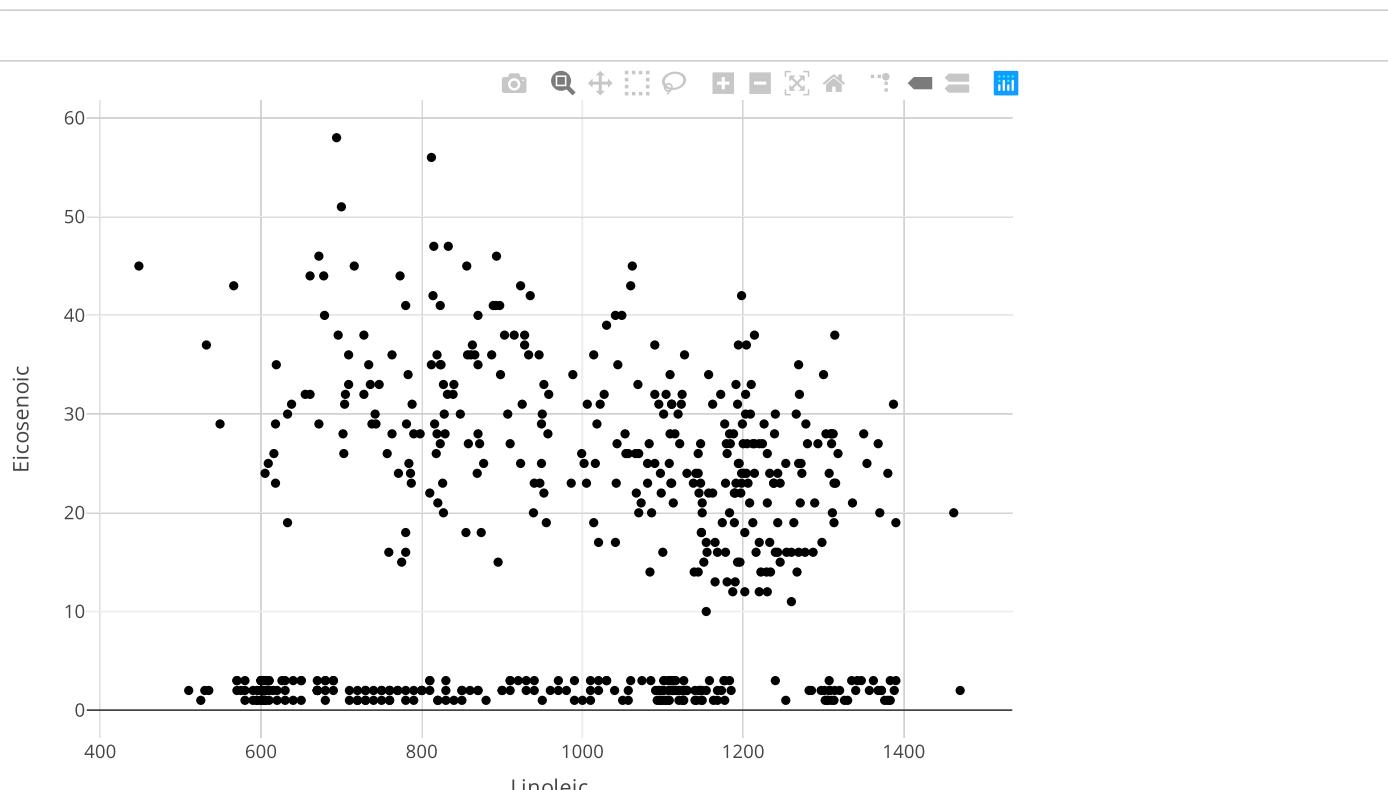
The watch has analog and digital display and it appears in black,gold and white. It has sporty look,rubber band and waterprrof and good value for money. The satisfied customers are talking about modern look,water resistance,some of them say the battery last long and it is good priced. Disatisfied customers talk about the defective alrams,battery once changed the watch does not work well,The analog portions becomes defective etc. Also Replacing battery is costly .

Seems like the customers are satisfied with the look and feel and price of the watch .With some changes in the battery and all it can be a good product.

Assignment 2

1.Scatter Plot eicosenoic vs linoleic

Choose region



The value of eicosenoic is between one and three. # ## 2.Linked Scatterplot and Bar chart

values of stearic



Brush color

rgba(228, 1

152 375
248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

152 248 344

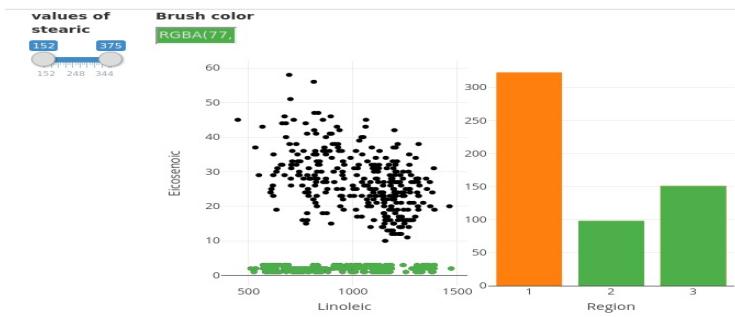
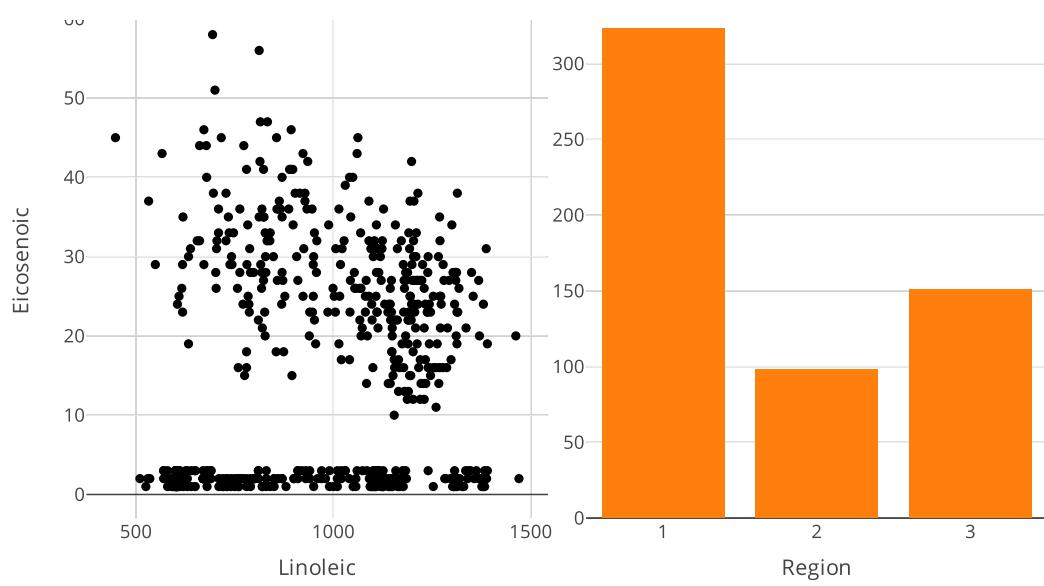


Figure 1. coloured

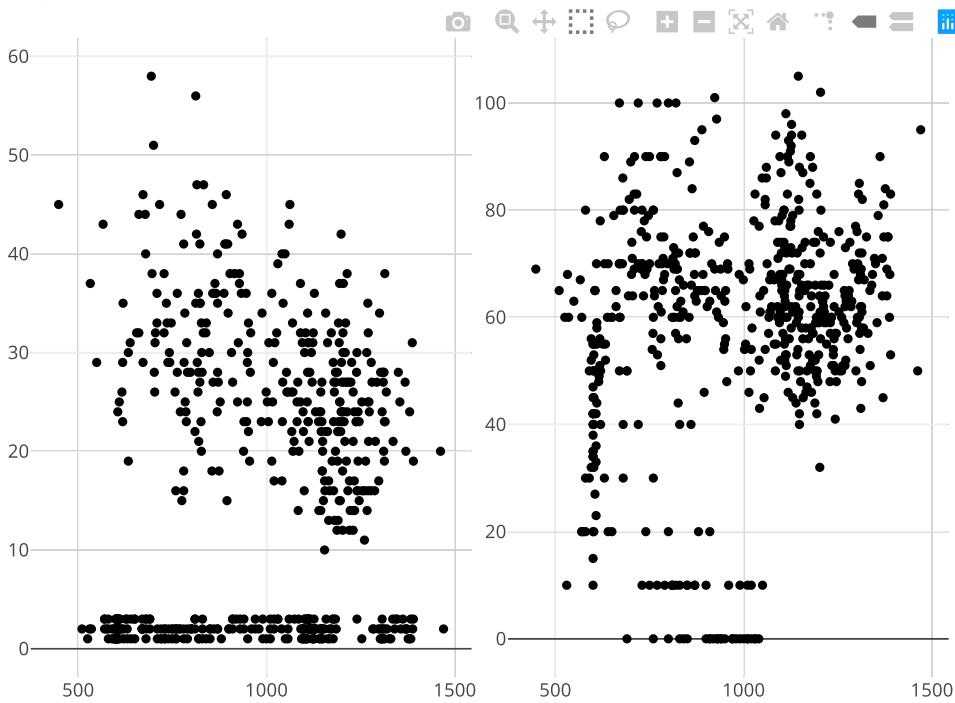
Using brushing when we take out the lower values of eicosenoic(between 1 and 3),we can see they correspond to region south and sardinia Island.In south region the amount of stearic acid is between 190 and 270.

The operators used here are selection operators for brushing,Connection operator for connecting the plots and Filtering operator for filtering stearic values

##3.Linked Scatter Plots

Brush color

`rgba(228,2`



Brush color
rgba(228,2)

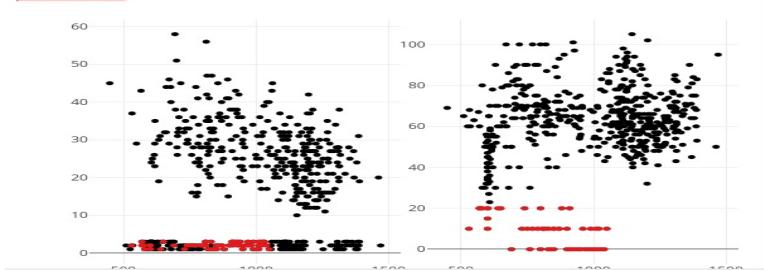


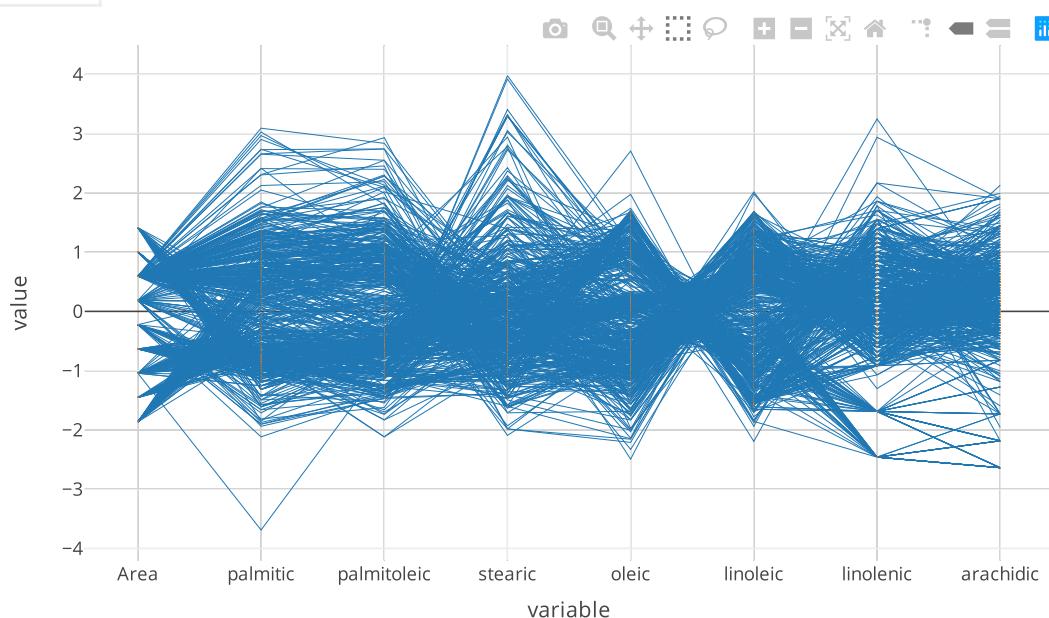
Figure 1. linked scatter

The values of arachidic below 40 are outliers in the plot2 are also outliers in plot 1. IN plot one their eicosenoic values are between 1 and 3.

####4.Parallel Coordinate Plot linked Barplot and Scatterplot

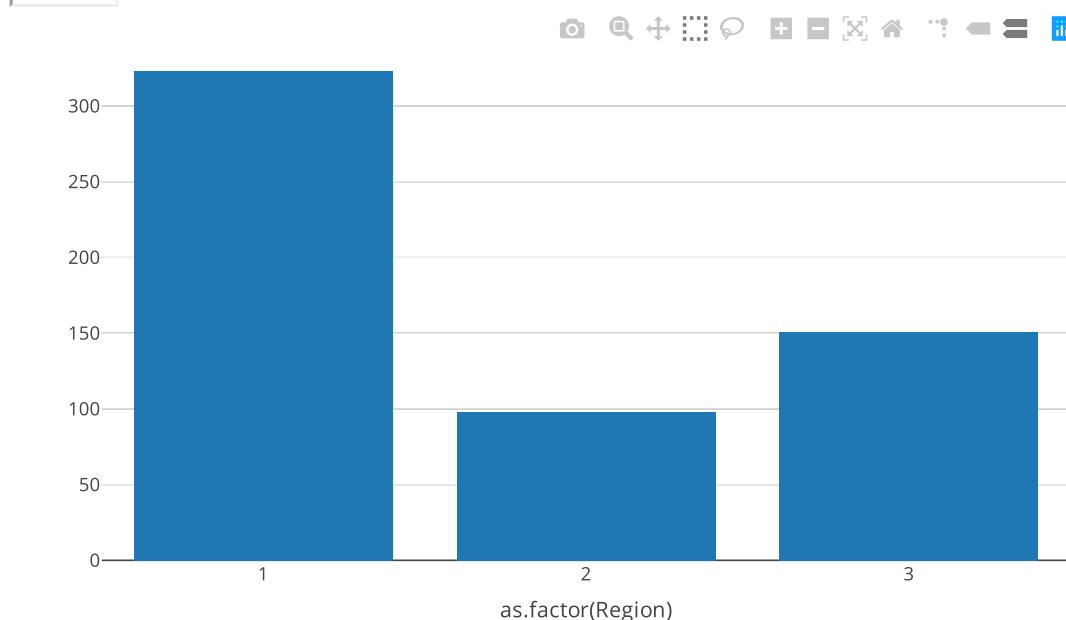
Brush color

rgba(228,2)



Brush color

rgba(228,2)



Brush color

rgba(228,2)

palmitic ▼

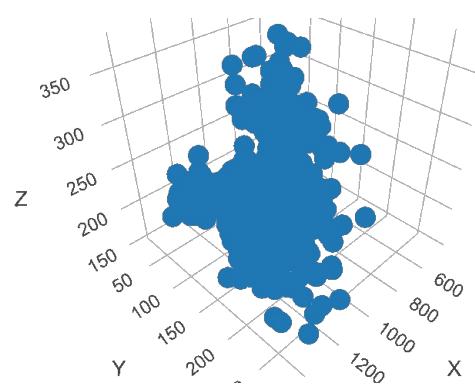
Distribution of Oil



palmitic ▼



palmitic ▼



When we analysed the parallel coordinate plot, linolenic, oleic and palmitic seemed to have more clusters hence they are influential variables to differentiate regions.

In South region we can observe multiple clusters. linolenic, oleic and stearic forms cluster, steraric, oleic, linoleic forms another cluster.

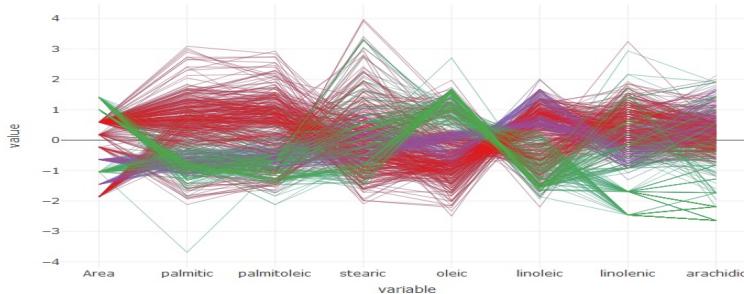


Figure 1. plot_123

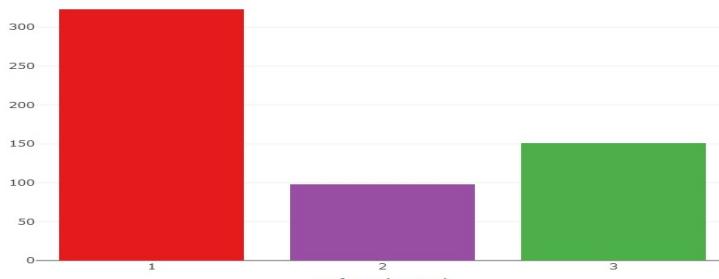


Figure 1. plot_123

oleic ▼

Distribution of Oil

linolenic ▼

palmitic ▼

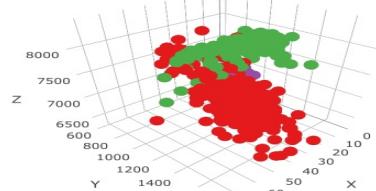


Figure 1. plot_123

As seen in the above screenshots, when we select the 3 variables and brush the regions with 3 different colours, the plot is almost clustered in 3 separate groups.

##5.Analysis

In step 4 we have used:

selection operator while brushing the plots ,connection operator to connect plots, Reconfiguring operator dynamically selecting axis variables in 3D plot on data value space operand

An additional interaction operator which can be used is Filtering operator.This operator can be used to filter based on regions and analyse the concentration of acids in oil from each region. This data can be used for analysing the region of origin of an oil given the concentration of acids in that oil.s

##Appendix

```
## ----message=FALSE,echo=FALSE-----  
library(tm)  
library(wordcloud)
```

```

library(RColorBrewer)
#library(viridisLite)
library(plotrix)
library(tidyverse)

## ----message=FALSE,echo=F-----
five<-read.table("Five.txt",header=F, sep='\n',encoding="latin1")#Read file
five$doc_id=1:nrow(five)
colnames(five)[1]<-"text"
mycorpus <- Corpus(DataframeSource(five)) #Creating corpus (collection of text data)

clean_corpus <- function(corpus){
  corpus <- tm_map(corpus, removePunctuation)
  corpus <- tm_map(corpus, content_transformer(tolower))
  corpus <- tm_map(corpus, stripWhitespace)
  corpus <- tm_map(corpus,function(x) removeWords(x,stopwords("english")))
  return(corpus)
}

mycorpus <- clean_corpus(mycorpus)
tdm <- TermDocumentMatrix(mycorpus) #Creating term-document matrix
m <- as.matrix(tdm)

#here we merge all rows
v <- sort(rowSums(m),decreasing=TRUE) #Sum up the frequencies of each word
d <- data.frame(word = names(v),freq=v) #Create one column=names, second=frequencies

pal<-brewer.pal(6,"PRGn") #Create palette of colors
#color_pal <- cividis(n = 8)

wordcloud(d$word,d$freq, scale=c(8,.3),min.freq=2,max.words=100, random.order=F, rot.per=.15, colors=col_pal, vfont=c("sans serif","plain"))

wordcloud(d$word,d$freq, scale=c(8,.3),min.freq=2,max.words=100, random.order=F, rot.per=.15, colors=col_pal, vfont=c("sans serif","plain"))

## ----message=F,echo=F-----
oneTwo<-read.table("OneTwo.txt",header=F,sep="\n")
oneTwo$doc_id=1:nrow(oneTwo)
colnames(oneTwo)[1]<-"text"

mycor <- Corpus(DataframeSource(oneTwo))

mycor <- clean_corpus(mycor)
tdm1<- TermDocumentMatrix(mycor) #Creating term-document matrix
m1 <- as.matrix(tdm1)

#here we merge all rows
v1 <- sort(rowSums(m1),decreasing=TRUE) #Sum up the frequencies of each word
d1<- data.frame(word = names(v1),freq=v1)
pal<-brewer.pal(8,"Dark2")

wordcloud(d1$word,d1$freq, scale=c(8,.3),min.freq=2,max.words=100, random.order=F, rot.per=.15, colors=col_pal, vfont=c("sans serif","plain"))

## ----message=FALSE,warning=FALSE,echo=F-----
j<-left_join(d,d1,by="word")

pyramid.plot(
  j$freq.x[1:10],j$freq.y[1:10],
  # Words
  labels = j$word[1:10],
  top.labels = c("Words in File", "Common Words", "Words in oneTwo"),
  main = "Pyramid Plot",gap=17,unit="counts")

## ----message=FALSE,echo=FALSE-----
library(tidyverse)
library(plotly)
library(plyr)
library(crosstalk)
library(GGally)

```

```

olive<-read.csv("olive.csv")
olive$Region<-as.factor(olive$Region)
levels(olive$Region)<-c("1","2","3")
o<-SharedData$new(olive)
oo<-SharedData$new(olive,~Region,group = "Choose region")

## ----set-options,message=F,echo=F-----
scatterolive <- plot_ly(oo, y = ~eicosenoic, x = ~linoleic)%>%group_by(Region)%>%
  add_markers(color = I("black"),name="hollow")%>%highlight(on="plotly_hover",persistent = F,selectize = T) %>% l
ayout(xaxis = list(title = "Linoleic"),
      yaxis = list(title = "Eicosenoic"))

scatterolive

## ----message=FALSE,warning=FALSE,echo=F-----

barolive <- plot_ly(oo, x=~as.factor(Region))%>%add_histogram()%>%layout(barmode="overlay",xaxis=list(title="Regio
n"))

bscols(widths=c(2, NA),filter_slider("stearic", "values of stearic", oo, ~stearic),subplot(scatterolive,barolive,t
itleY = TRUE, titleX = TRUE)%>%highlight(on="plotly_select", dynamic=T, persistent = T, opacityDim = I(1))%>%hide
_legend())

## ----message=FALSE,echo=F-----
scatter1 <- plot_ly(o, x = ~linoleic, y = ~eicosenoic) %>%
  add_markers(color = I("black")) %>% layout(xaxis = list(title = "linoleic"),
                                                yaxis = list(title = "Eicosenoic"))
scatter2 <- plot_ly(o, x = ~linoleic, y = ~arachidic) %>%
  add_markers(color = I("black"))%>% layout(xaxis = list(title = "linoleic"),
                                                yaxis = list(title = "arachidic"))

subplot(scatter1,scatter2)%>%
  highlight(on="plotly_select", dynamic=T, persistent=T, opacityDim = I(1))%>%hide_legend()

## ----message=FALSE,fig.width=20, fig.height=18,echo=F-----
oliveparallelcord<-ggparcoord(olive, columns = c(3:10))

d<-plotly_data(ggplotly(oliveparallelcord))%>%group_by(.ID)
d1<-SharedData$new(d, ~.ID, group="olive")

p1<-plot_ly(d1, x=~variable, y=~value)%>%
  add_lines(line=list(width=0.3))%>%
  add_markers(marker=list(size=0.3),
              text=~.ID, hoverinfo="text")

oildata <- olive
oildata$.ID=1:nrow(oildata)
d2<-SharedData$new(oildata, ~.ID, group="olive")
p2 <- plot_ly(d2, x=~as.factor(Region))%>%add_histogram()%>%layout(barmode="overlay")

ButtonsX=list()
for (i in 4:10){
  ButtonsX[[i-3]]= list(method = "restyle",
                        args = list( "x", list(olive[[i]])),
                        label = colnames(olive)[i])
}

ButtonsY=list()
for (i in 4:10){
  ButtonsY[[i-3]]= list(method = "restyle",

```

```

        args = list( "y", list(olive[[i]])),
        label = colnames(olive)[i])
    }

ButtonsZ=list()
for (i in 4:10){
  ButtonsZ[[i-3]]= list(method = "restyle",
                        args = list( "z", list(olive[[i]])),
                        label = colnames(olive)[i])
}

axx <- list(
  title = "X"
)

axy <- list(
  title = "Y"
)

axz <- list(
  title = "Z"
)
p3 <- plot_ly(d2,x= ~linoleic,y = ~palmitoleic, z = ~stearic)%>%add_markers()%>%
  layout(xaxis=list(title=""), yaxis=list(title=""),
         title = "Distribution of Oil",
         updatemenus = list(
           list(y=0.9, buttons = ButtonsX),
           list(y=0.6, buttons = ButtonsY),
           list(y=1.2, buttons = ButtonsZ)
         )%>%layout(scene = list(xaxis=axx,yaxis=axy,zaxis=axz))

ps<-htmltools::tagList(p1%>%
                          highlight(on="plotly_select", dynamic=T, persistent = T, opacityDim = I(1))%>%
                          hide_legend(),
                          p2%>%
                          highlight(on="plotly_select", dynamic=T, persistent = T, opacityDim = I(1))%>%
                          hide_legend(),
                          p3%>%
                          highlight(on="plotly_click", dynamic=T, persistent = T, opacityDim = I(1))%>%
                          hide_legend())
)
htmltools::browsable(ps)

```