

# lab5

Andreas and Priya

11 Οκτωβρίου 2018

- Assignment 1
  - Word Cloud Plot for five
  - Word Cloud Plot for One Two
  - Pyramid Plot
  - Phrasenets
- Assignment 2
  - 1.Scatter Plot eicoseinoic vs linoleic
  - 3.Linked Scatter Plots
  - 4.Parallel Coordinate Plot linked Barplot and Scatterplot
- Appendix

## Assignment 1

```
library(tm)
library(wordcloud)
library(RColorBrewer)
library(viridisLite)
library(plotrix)
library(tidyverse)
```

### Word Cloud Plot for five

```
five<-read.table("Five.txt", header=F, sep='\n')#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(corporus){
  corporus <- tm_map(corporus, removePunctuation)
  corporus <- tm_map(corporus, content_transformer(tolower))
  corporus <- tm_map(corporus, stripWhitespace)
  corporus <- tm_map(corporus,function(x) removeWords(x,stopwords("english")))
  return(corporus)
}

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"))
```



## Word Cloud Plot for One Two

```

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=pal, vfont=c("sans serif","plain"))

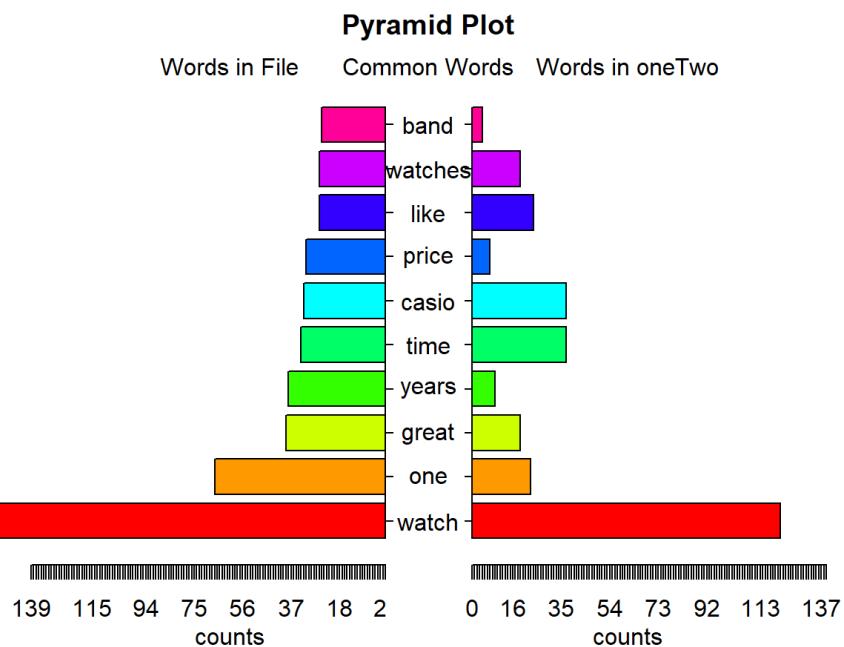
```



## Pyramid Plot

```
j<-left_join(d,d1,by="word")
## Warning: Column `word` joining factors with different levels, coercing to
## character vector

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")
```



```
## [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

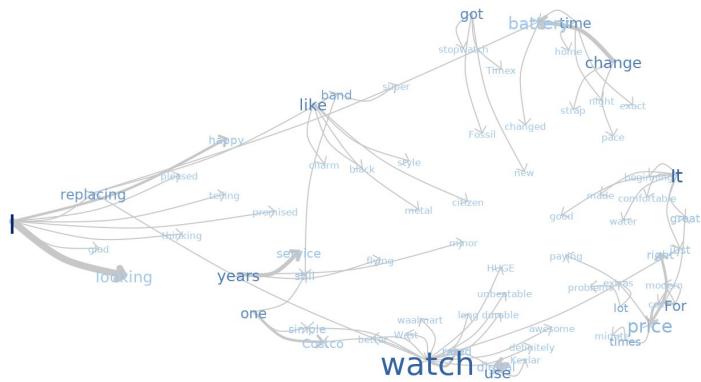


Figure 1. positive.

## phrase nets of negative words

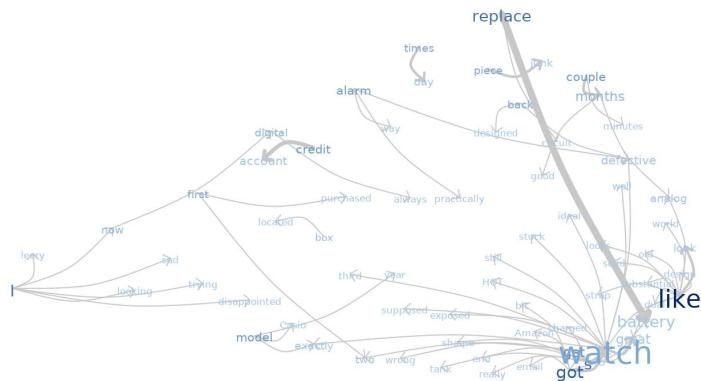


Figure 2. negative

## word trees

word tree of not with the positive feedback

TOO BIG OR SMALL OR TOO HEAVY. QUITE THIN I  
 , comfortable. looks snazzy. Priced right. Ok..  
 and chunky. The dual time feature keeps trac  
 large, easy to read and has both analog and easy to set digital r  
 expensive, and not too big and chunky. The dual time feature k  
 woks great but looks like a \$500 dive watch. It so far seems very  
 does it look sexy but its durable super high quality and doesnt sc  
 looks great (kinda like an Omega) but it is very durable. This wa  
 how long I have had it, but a guess would be easily over 6 years. I  
 I ever want another model. I'm surprised that so many people cri  
 diver but I don't want a watch which is afraid of the water. This one is no  
 a Casio) broke after a basketball hit it, popping off the glass and knocking  
 beat. Looks good, feels robust but not heavy and from the other reviews should  
 heavy and from the other reviews should last quite a while. WATCH IS NOT TOO  
 noticeable in the pictures. IMO the dial should be prominent instead of the bezel.  
 tried to see if it is waterproof but it looks like it will hold up very well This is a great  
 in a good way!! This watch met all my requirements: it's a face watch, it's got a sto  
 chic but the most durable watches you can buy! It fulfilled completely my expecta  
 disappointed. This one has a larger dial, which is easier to see without my glasses,  
 . I haven't put it to a 100-m test but it shrugs off casual water from boating, water  
 great in the luminosity dept. but just viewable at night. Limited functionality, but i  
 find it cheaper elsewhere. It's perfect for surfing, swimming, etc. and has a stylish  
 water resistant)This is a great watch, well worth the money.I am here looking to re  
 go wrong owning this piece. It will serve you well for many, many years. I have see  
 really. I like to know that I'm wearing a watch when I have one on. Bottom line gr  
 partial to strap bands so I have transferred the metal band from the old one so tha  
 strong or long. But for \$50, it is a great watch still. My previous Casio served for te  
 recommend scuba diving unless the watch is rated 200 or 300m, I have been scub  
 offer as much as this Casio. For the cost, the features, and the overall value, I don't  
 lost a second. The digital display is a little small, but I don't use it much anyway ar

Figure 4. wordtree not pos feedback

word tree with the word amazon in negative feedback

Both times, the watch movement stopped working correct  
 The watch battery only last a year. It will cost more than t  
 Amazon sent it again, to the main USPS facility, different  
 If Amazon wants to credit the 49.00, I would be willing to  
 com, certainly not. Thank you all. REVIEW: I hope this re  
 nope, UPS would not deliver to my PO Box. Well, I went to  
 I don't want this watch anymore, and I want it replaced, n  
 to save the situation and keep the customer. And you win :  
 'S warranty so now I'll have to chase down Casio for a repl  
 money or anyone's. I paid for the watch, I got the watch,  
 team even if you don't want to replace it, or give me c  
 was very courteous and friendly. The Casio guys  
 decides to reason and have the courtesy to credi  
 correct this situation, and take back t  
 (THANK YOU AMAZON for not even a e-mail response). My strong  
 for not even a e-mail response). My strong recommendation: SKIP  
 ! The watch is great, but there's more to this case than just a senten  
 sent it again, to the main USPS facility, different place, that's where  
 /UPS delivery policies, it's more like a world away. So I call Amazo  
 back and talk again with another representative in India, very frier  
 did have the goodwill to send a replacement, then a second replace  
 expires, and now I'm on my own. Well, Casio will cover for it. Yeah,  
 courteously to reconsider this issue, and give me a replacement iter  
 sends me an email (one of those that look like spam and people del  
 representative I just spoke to (today is September 9, 2010), has "N  
 does make an effort to satisfy every customer, sometimes there are  
 nor Casio want to deal with it. I would warn every Amazon custome  
 customer to research the watches well. This particular watch retails  
 AGAIN. I would pay the premium somewhere else. Sadly, the once  
 has been eroded with so many irregularities in this case. By the way  
 wants to credit the 49.00, I would be willing to purchase another, n  
 this time. If Amazon decides to reason and have the courtesy to cre

Figure 4. wordtree amazon neg feedback

# Assignment 2

```
library(tidyverse)
library(plotly)
library(plyr)
library(crosstalk)
library(GGally)

olive<-read.csv("olive.csv")
olive$Region<-as.factor(olive$Region)
levels(olive$Region)<-c("North","South","Sardinia island")
oo<-SharedData$new(olive)
oo<-SharedData$new(olive,~Region,group = "Choose region")
```

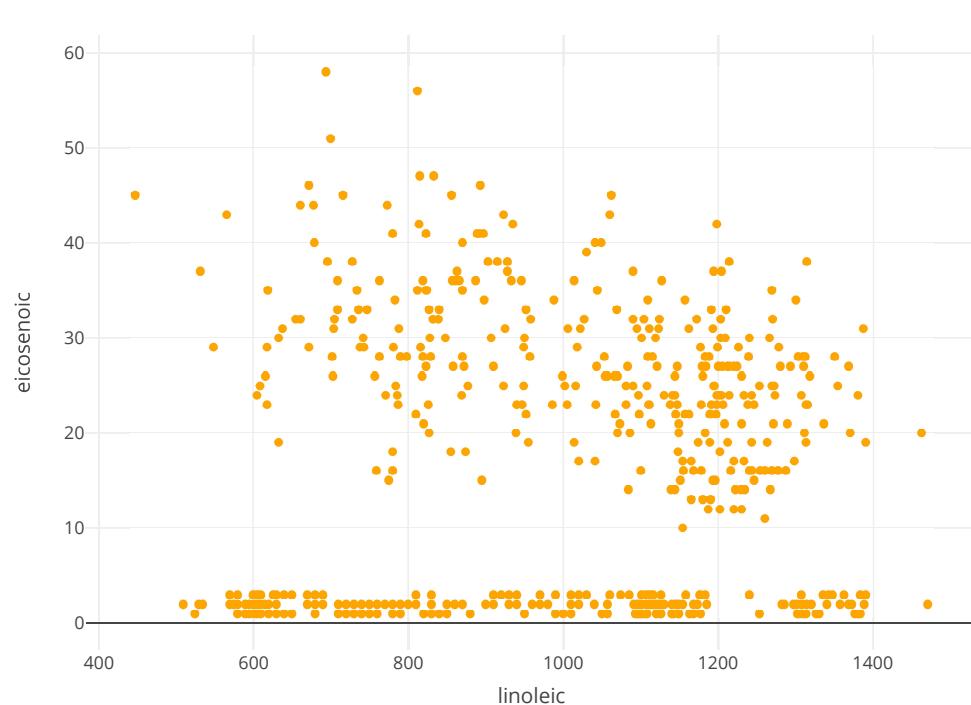
## 1.Scatter Plot eicoseinoic vs linoleic

```
scatterolive <- plot_ly(oo, y = ~eicosenoic, x = ~linoleic) %>%group_by(Region)%>%
  add_markers(color = I("orange"),name="hollow")%>%highlight(on="plotly_hover",persistent = F,selectize = T)

scatterolive
```

## Setting the `off` event (i.e., 'plotly\_doubleclick') to match the `on` event (i.e., 'plotly\_hover'). You can change this default via the `highlight()` function.

Choose region



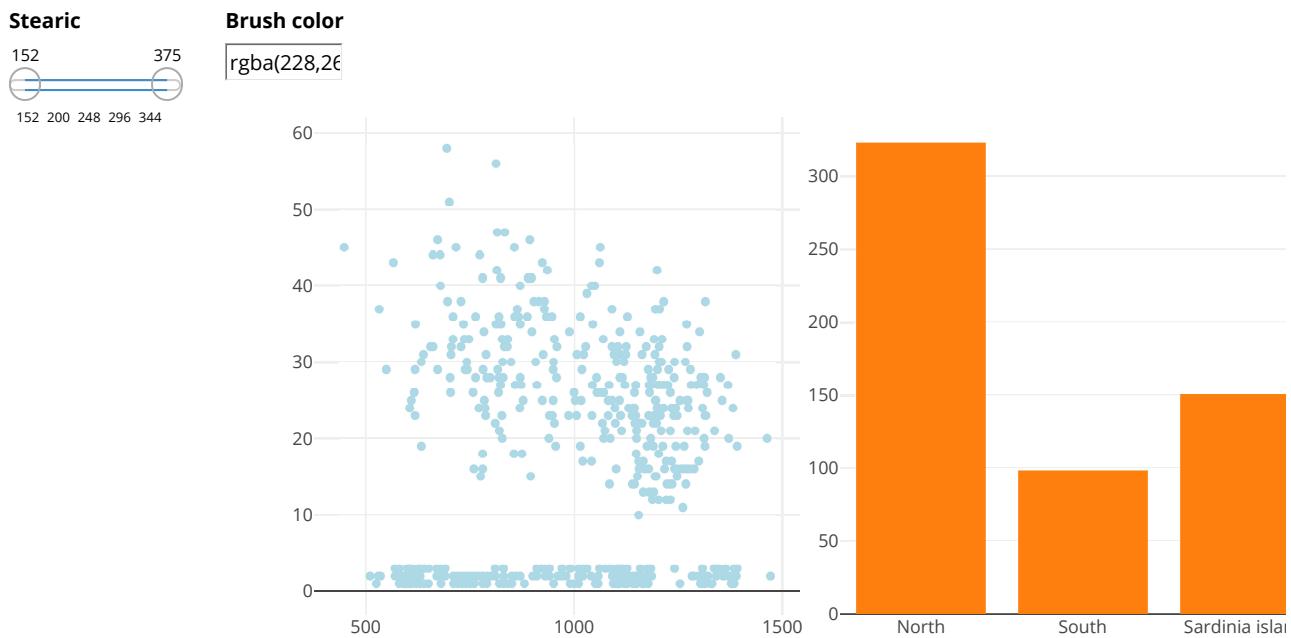
The value of eicosenoic is between one and three. ####2.Linked Scatterplot and Bar chart

```
scatterolive2 <- plot_ly(o, y = ~eicosenoic, x = ~linoleic) %>%
  add_markers(color = I("lightblue"))

barolive <-plot_ly(o, x=~Region,group="Select your region")%>%add_histogram()%>%layout(barmode="overlay")
```

## Warning in plot\_ly(o, x = ~Region, group = "Select your region"): The group argument has been deprecated. Use `group\_by()` or split instead.  
## See `help('plotly\_data')` for examples

```
bscols(widths=c(2, NA),filter_slider("Slider", "Stearic", o, ~stearic)
      ,subplot(scatterolive2,barolive)%>%
        highlight(on="plotly_select", dynamic=T, persistent = T, opacityDim = I(1))%>%hide_legend())
```



The brush colour and the slider/filter were the interaction operators used in this step.

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.Now when we use slider to find the higher values of stearic we can see that they correspond to the region of north and Sardinia Island

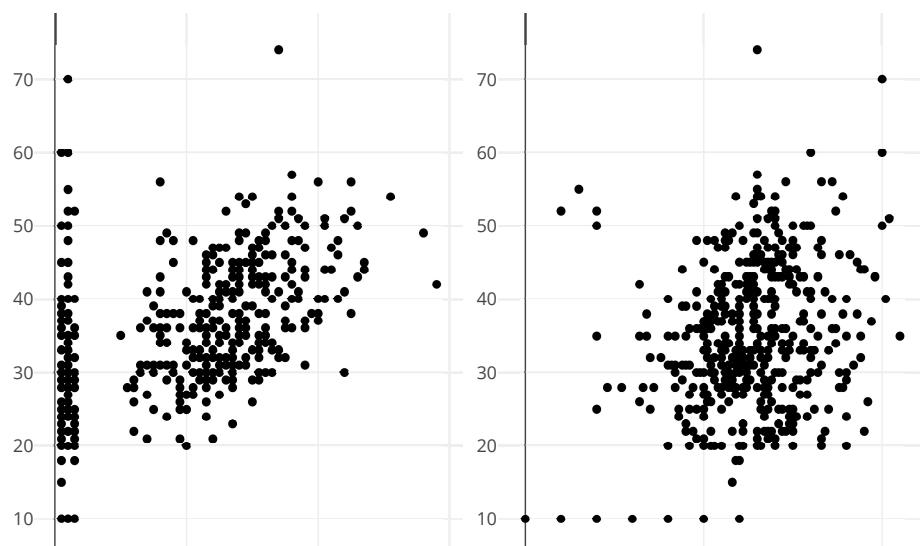
### 3.Linked Scatter Plots

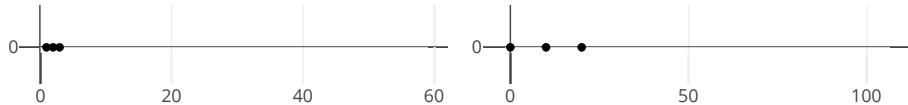
```
scatter1 <- plot_ly(o, x_= ~eicosenoic, y_= ~linolenic) %>%
  add_markers(color = I("black"))
scatter2 <- plot_ly(o, x_= ~arachidic, y_= ~linolenic) %>%
  add_markers(color = I("black"))

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

**Brush color**

rgba(228,26)





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

```

p<-ggparcoord(olive,columns = c(4:11))

d<-plotly_data(ggplotly(p))%>%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")

ButtonsX=list()
for (i in 4:11){
  ButtonsX[[i-3]]= list(method = "restyle",
                        args = list( "x", list(olive[[i]])),
                        label = colnames(olive)[i])
}

ButtonsY=list()
for (i in 4:11){
  ButtonsY[[i-3]]= list(method = "restyle",
                        args = list( "y", list(olive[[i]])),
                        label = colnames(olive)[i])
}

ButtonsZ=list()
for (i in 4:11){
  ButtonsZ[[i-3]]= list(method = "restyle",
                        args = list( "z", list(olive[[i]])),
                        label = colnames(olive)[i])
}

p3<-plot_ly(o,x=~palmitic,y=~stearic,z=~oleic)%>%add_markers() %>%
  layout(xaxis=list(title=""), yaxis=list(title=""),
         title = "Select variable:",
         updatemenus = list(
           list(y=0.9, buttons = ButtonsX),
           list(y=0.7, buttons = ButtonsY),
           list(y=0.5, buttons = ButtonsZ)
         )
       )

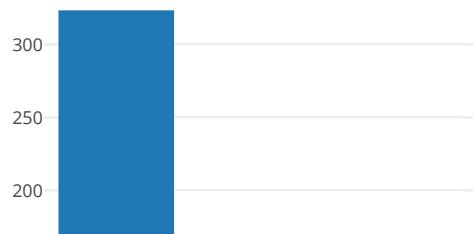
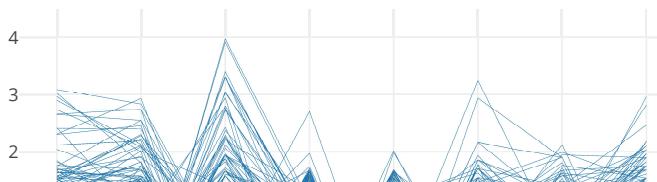
bscols(p1%>%highlight(on="plotly_select", dynamic=T, persistent = T, opacityDim = I(1))%>%
  hide_legend(),barolive,
  p3%>%highlight(on="plotly_click", dynamic=T, persistent = T)%>%hide_legend(),
  widths = c(7,5,9))

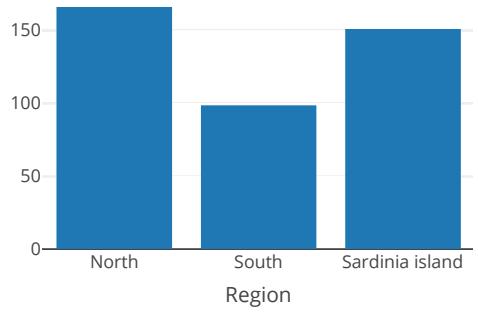
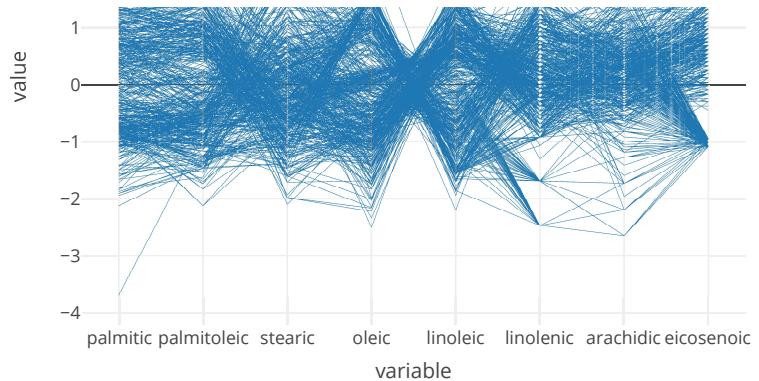
```

## Warning in bscols(p1 %>% highlight(on = "plotly\_select", dynamic = T, persistent = T, : Sum of bscolumn width units is greater than 12

#### Brush color

rgba(228,26





#### Brush color

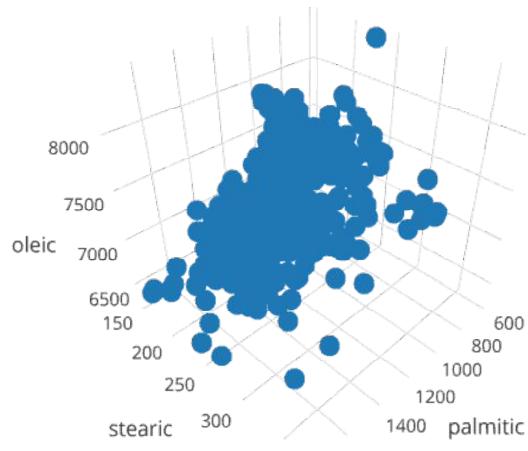
`rgba(228,26`

Select variable:

palmitic ▼

palmitic ▼

palmitic ▼



The parallel coordinate plot demonstrate clusters among observations that belong to the same region. i think oleic,linoleic,Eicosenoic are three influential variables.while using this three variable we think we can differentiate regions.

## Appendix

```

library(tm)
library(wordcloud)
library(RColorBrewer)
library(viridisLite)
library(plotrix)
library(tidyverse)
five<-read.table("Five.txt",header=F, sep='\n')#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"))

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"))

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")

library(tidyverse)
library(plotly)
library(plyr)
library(crosstalk)

```

