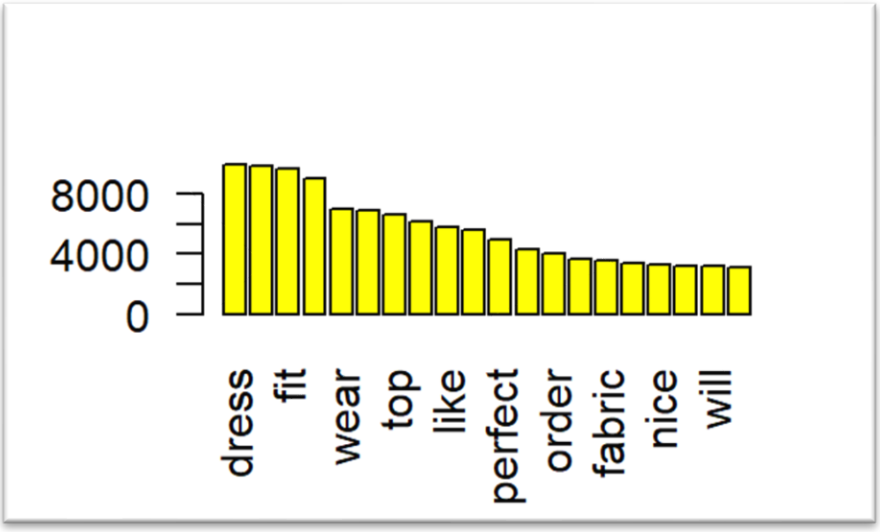
HW5
109306045 資管二黃筠茜

```
#1
install.packages("jiebaR")
library("jiebaR")
install.packages("tmcn")
library(tmcn)
library(tidyverse)
library(tidyr)
library(readr)

#推薦的
data1=data[which(data$`Recommended IND`==1),]
library(tm)
## Make a vector source and a corpus
x=Corpus(VectorSource(data1$`Review Text`))
##Clean text
x=tm_map(x, tolower) #convert to lower case 換成小寫
x<-tm_map(x, content_transformer(tolower))
x=tm_map(x, removePunctuation) #remove punctuation 標點符號

x=tm_map(x, removeWords, stopwords("english"))
x=tm_map(x, stemDocument)

x_tdm <- TermDocumentMatrix(x)
inspect(x_tdm)
review_m <- as.matrix(x_tdm)
freq_df <- rowSums(review_m)
freq_df <- sort(freq_df, decreasing = T)
freq_df[1:10]
barplot(freq_df[1:20], col = "yellow", las = 2)
```



看得出前十多的字有 fit like perfect nice 等樂觀字眼

```
freq_df <- data.frame(word = names(freq_df),
                                    num = freq_df)
library(wordcloud2)
wordcloud2(freq_df,size=0.5)
```

Love fit dress like 都是正面的詞

```
#語意分析 正向負面
install.packages("tidytext")
library(tidytext)
```

```
get_sentiments("bing")
bing_word_counts <- freq_df %>%
    inner_join(get_sentiments("bing")) %>%
    count(word, sentiment, sort = TRUE)

bing_word_counts

table(bing_word_counts$sentiment)
```

```
negative positive
     371      266
```

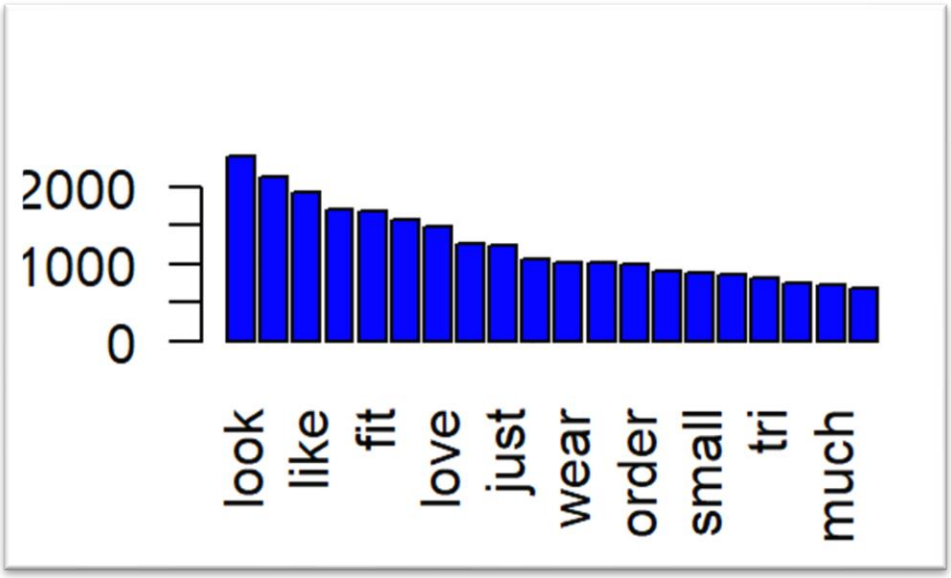
雖然 negative 較多 但兩者並沒有差很多

```
bing_word_counts %>%
    filter(sentiment == "positive") %>%
    select(word,n)%>%
    wordcloud2()
```

推薦者為留言較為正面族群 像是有 angel afford awe
顯現出他們對產品的滿意

#不推薦

```
data2 = data[which(data$`Recommended IND`==0),]
x2=Corpus(VectorSource(data2$`Review Text`))

x2=tm_map(x2, tolower)
x2<-tm_map(x2, content_transformer(tolower))

x2=tm_map(x2, removePunctuation)

x2=tm_map(x2, removeWords, stopwords("english"))
x2=tm_map(x2, stemDocument)

x2_tdm <- TermDocumentMatrix(x2)
inspect(x2_tdm)

review_m2 <- as.matrix(x2_tdm)

freq_df2 <- rowSums(review_m2)

freq_df2 <- sort(freq_df2, decreasing = T)

freq_df2[1:10]
barplot(freq_df2[1:20], col = "blue", las = 2)
```



可看出幾個主要關鍵字,但沒有太多的負面字眼,只是相對的正面字眼更少

```
freq_df2 <- data.frame(word = names(freq_df2),
```

num = freq_df2)

wordcloud2(freq_df2,size=0.5)

文字雲顯示正面字更少

```
get_sentiments("bing")
bing_word_counts2 <- freq_df2 %>%
    inner_join(get_sentiments("bing")) %>%
    count(word, sentiment, sort = TRUE)

bing_word_counts

table(bing_word_counts2$sentiment)
```

```
negative positive
     297      181
```

Negative 字更多了

```
bing_word_counts2 %>%
    filter(sentiment == "negative") %>%
```

```
  select(word,n)%>%
  wordcloud2()
```

出現負面字詞 broke annoy bust,所以較不推薦

```
#2
library(rvest)
library(magrittr)
library(httr)
library(jsonlite)
library(tidyverse)

options(stringsAsFactors = FALSE)#保留 str 型態  不拆開
options(encoding = "UTF-8")#設定編碼
dcardurl <- 'https://www.dcard.tw/_api/forums/'#設定 URL
board<-'relationship'#看板設定
mainurl <- paste0(dcardurl,board,'/posts?popular=false')
resdata <- fromJSON(content(GET(mainurl), "text"))
head(resdata[,c(1,2)])
n <- 1800#抓 1800 篇文章
page <- (1800/30)-1#每幾筆抓一項
end <- resdata$id[length(resdata$id)]
end
for(i in 1:page){
```

```
    url <- paste0(mainurl,"&before=",end)
    print(url)
    tmpres <- fromJSON(content(GET(url), "text"))
    end <- tmpres$id[length(tmpres$id)]
    resdata <- bind_rows(resdata[,c(1:12)],tmpres[,c(1:12)])
}
rm(tmpres)
head(resdata)
count <-table(cc[resdata[,2]])#

newd = data.frame(count)#

head(newd[order(newd$Freq,decreasing = TRUE),],20)
newdd = newd[order(newd$Freq,decreasing = TRUE),]
wordcloud2(newdd)

word <- cc[resdata[,2]]
newd = data.frame(table(word))

newd %>%
    filter(!str_detect(word, "[a-zA-Z0-9]+")) %>%    #去掉 english and number
    filter(nchar(as.character(word)) > 1) %>% #一個字的去掉
    filter( Freq > 2) ->temp    #可留下頻率>某數字

wordcloud2(temp,size=0.4)
```

可看出感情版的幾個主要關鍵字是男友 我 女友 怎麼辦,看來主要大家都在煩惱感情問題