Sentiment Analysis Interview 1

Alvaro Rivera

2023-02-23

```
pkg <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE)
  sapply(pkg, require, character.only = TRUE)
}
packages <- c("tidyverse","raster","sf","ggspatial","cluster","factoextra",
              "NbClust","tidyr","forecast","semTools","corrplot",
              "corrr","haven","psych","dplyr","lavaan","readr","cvms","tm",
              "NLP","SnowballC","RColorBrewer","wordcloud","wordcloud2",
              "RefManageR","bibliometrix","GGally","quanteda","ggplot2",
              "ggpubr","Factoshiny","syuzhet","RColorBrewer","tokenizers",
              "stringr","sentimentr","stringi","stopwords","twitteR",
              "mscstexta4r","plyr","psych","corrr","latticeExtra",
              "semPlot","lavaan","readr","lme4","sjPlot","gvlma","Rcmdr",
              "tidymodels","caret","lmtest","gapminder","png","rtweet","knitr")
pkg(packages)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.4.1     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.4     v forcats 1.0.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading required package: raster
```

```
## Loading required package: sp
```

```
##
## Attaching package: 'raster'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## Loading required package: sf
```

```
## Linking to GEOS 3.10.2, GDAL 3.4.3, PROJ 8.2.0; sf_use_s2() is TRUE
```

```
## Loading required package: ggspatial
```

```
## Loading required package: cluster

## Loading required package: factoextra

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

## Loading required package: NbClust

## Loading required package: forecast

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## Loading required package: semTools

## Loading required package: lavaan

## This is lavaan 0.6-11
## lavaan is FREE software! Please report any bugs.

##

## ################################################################################

## This is semTools 0.5-6

## All users of R (or SEM) are invited to submit functions or ideas for functions.

## ################################################################################

##
## Attaching package: 'semTools'

## The following object is masked from 'package:readr':
##
##     clipboard

## Loading required package: corrplot

## corrplot 0.92 loaded

## Loading required package: corrr

##
## Attaching package: 'corrr'

## The following object is masked from 'package:raster':
##
##     stretch

## Loading required package: haven

## Loading required package: psych

##
## Attaching package: 'psych'

## The following objects are masked from 'package:semTools':
##
##     reliability, skew

## The following object is masked from 'package:lavaan':
##
##     cor2cov
```

```
## The following object is masked from 'package:raster':
##
##     distance

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

## Loading required package: cvms

## Loading required package: tm

## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##     annotate

##
## Attaching package: 'tm'

## The following object is masked from 'package:lavaan':
##
##     inspect

## Loading required package: SnowballC

## Loading required package: RColorBrewer

## Loading required package: wordcloud

## Loading required package: wordcloud2

## Loading required package: RefManageR

## Loading required package: bibliometrix

## To cite bibliometrix in publications, please use:
##
## Aria, M. & Cuccurullo, C. (2017) bibliometrix: An R-tool for comprehensive science mapping analysis,
##                                  Journal of Informetrics, 11(4), pp 959-975, Elsevier.
##
##
## https://www.bibliometrix.org
##
##
## For information and bug reports:
##                          - Send an email to info@bibliometrix.org
##                          - Write a post on https://github.com/massimoaria/bibliometrix/issues
##
## Help us to keep Bibliometrix free to download and use by contributing with a small donation to suppo
##
##
## To start with the shiny web-interface, please digit:
## biblioshiny()

##
## Attaching package: 'bibliometrix'
```

```
## The following object is masked from 'package:raster':
##
##     trim

## Loading required package: GGally

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

## Loading required package: quanteda

## Package version: 3.2.1
## Unicode version: 14.0
## ICU version: 70.1

## Parallel computing: 8 of 8 threads used.

## See https://quanteda.io for tutorials and examples.

##
## Attaching package: 'quanteda'

## The following object is masked from 'package:tm':
##
##     stopwords

## The following objects are masked from 'package:NLP':
##
##     meta, meta<-

## Loading required package: ggpubr

##
## Attaching package: 'ggpubr'

## The following object is masked from 'package:cvms':
##
##     font

## The following object is masked from 'package:forecast':
##
##     gghistogram

## The following object is masked from 'package:raster':
##
##     rotate

## Loading required package: Factoshiny

## Loading required package: FactoMineR

## Loading required package: shiny

##
## Attaching package: 'shiny'

## The following object is masked from 'package:cvms':
##
##     validate

## Loading required package: FactoInvestigate

## Loading required package: syuzhet
```

```
##
## Attaching package: 'syuzhet'

## The following object is masked from 'package:psych':
##
##     rescale

## Loading required package: tokenizers

## Loading required package: sentimentr

##
## Attaching package: 'sentimentr'

## The following object is masked from 'package:syuzhet':
##
##     get_sentences

## Loading required package: stringi

## Loading required package: stopwords

##
## Attaching package: 'stopwords'

## The following object is masked from 'package:tm':
##
##     stopwords

## Loading required package: twitteR

## Error: package or namespace load failed for 'twitteR' in loadNamespace(i, c(lib.loc, .libPaths()), v
##  there is no package called 'rjson'

## Loading required package: mscstexta4r

## Loading required package: plyr

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following object is masked from 'package:ggpubr':
##
##     mutate

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following object is masked from 'package:purrr':
##
##     compact

## Loading required package: latticeExtra

## Loading required package: lattice
```

```
##
## Attaching package: 'latticeExtra'
## The following object is masked from 'package:ggplot2':
##
##     layer
## Loading required package: semPlot
## Loading required package: lme4
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
## Attaching package: 'lme4'
## The following object is masked from 'package:raster':
##
##     getData
## Loading required package: sjPlot
## Loading required package: gvlma
## Loading required package: Rcmdr
## Loading required package: splines
## Loading required package: RcmdrMisc
## Loading required package: car
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:psych':
##
##     logit
## The following object is masked from 'package:dplyr':
##
##     recode
## The following object is masked from 'package:purrr':
##
##     some
## Loading required package: sandwich
##
## Attaching package: 'RcmdrMisc'
## The following object is masked from 'package:psych':
##
##     reliability
```

```
## The following object is masked from 'package:semTools':
##
##     reliability

## Loading required package: effects

## Use the command
##     lattice::trellis.par.set(effectsTheme())
##   to customize lattice options for effects plots.
## See ?effectsTheme for details.

## The Commander GUI is launched only in interactive sessions

##
## Attaching package: 'Rcmdr'

## The following object is masked from 'package:shiny':
##
##     radioButtons

## The following object is masked from 'package:base':
##
##     errorCondition

## Loading required package: tidymodels

## -- Attaching packages --------------------------------------- tidymodels 0.2.0 --

## v broom        1.0.3     v rsample      0.1.1
## v dials        0.1.1     v tune         0.2.0
## v infer        1.0.0     v workflows    0.2.6
## v modeldata    0.1.1     v workflowsets 0.2.1
## v parsnip      0.2.1     v yardstick    0.0.9
## v recipes      0.2.0

## -- Conflicts ------------------------------------------ tidymodels_conflicts() --
## x psych::%+%()          masks ggplot2::%+%()
## x yardstick::accuracy() masks forecast::accuracy()
## x scales::alpha()       masks psych::alpha(), ggplot2::alpha()
## x NLP::annotate()       masks ggplot2::annotate()
## x plyr::arrange()       masks dplyr::arrange()
## x plyr::compact()       masks purrr::compact()
## x plyr::count()         masks dplyr::count()
## x plyr::desc()          masks dplyr::desc()
## x scales::discard()     masks purrr::discard()
## x Matrix::expand()      masks tidyr::expand()
## x raster::extract()     masks tidyr::extract()
## x plyr::failwith()      masks dplyr::failwith()
## x dplyr::filter()       masks stats::filter()
## x recipes::fixed()      masks stringr::fixed()
## x plyr::id()            masks dplyr::id()
## x dplyr::lag()          masks stats::lag()
## x latticeExtra::layer() masks ggplot2::layer()
## x plyr::mutate()        masks ggpubr::mutate(), dplyr::mutate()
## x infer::observe()      masks shiny::observe()
## x Matrix::pack()        masks tidyr::pack()
## x car::recode()         masks dplyr::recode()
## x plyr::rename()        masks dplyr::rename()
## x raster::select()      masks dplyr::select()
```

```
## x car::some()           masks purrr::some()
## x yardstick::spec()     masks readr::spec()
## x recipes::step()       masks stats::step()
## x plyr::summarise()     masks dplyr::summarise()
## x plyr::summarize()     masks dplyr::summarize()
## x Matrix::unpack()      masks tidyr::unpack()
## x recipes::update()     masks Matrix::update(), lavaan::update(), raster::update(), stats::update()
## * Use tidymodels_prefer() to resolve common conflicts.

## Loading required package: caret

##
## Attaching package: 'caret'

## The following objects are masked from 'package:yardstick':
##
##     precision, recall, sensitivity, specificity

## The following object is masked from 'package:purrr':
##
##     lift

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following object is masked from 'package:quanteda':
##
##     index

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: gapminder

## Loading required package: png

## Loading required package: rtweet

##
## Attaching package: 'rtweet'

## The following object is masked from 'package:syuzhet':
##
##     get_tokens

## The following object is masked from 'package:purrr':
##
##     flatten

## Loading required package: knitr

##    tidyverse       raster           sf    ggspatial      cluster   factoextra
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##      NbClust        tidyr     forecast     semTools     corrplot        corrr
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##        haven        psych        dplyr       lavaan        readr         cvms
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
```

```
##           tm          NLP    SnowballC RColorBrewer    wordcloud   wordcloud2
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##   RefManageR bibliometrix       GGally     quanteda      ggplot2       ggpubr
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##   Factoshiny      syuzhet RColorBrewer   tokenizers      stringr   sentimentr
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##      stringi    stopwords      twitteR  mscstexta4r         plyr        psych
##         TRUE         TRUE        FALSE         TRUE         TRUE         TRUE
##        corrr latticeExtra      semPlot       lavaan        readr         lme4
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##       sjPlot        gvlma        Rcmdr   tidymodels        caret       lmtest
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##    gapminder          png       rtweet        knitr
##         TRUE         TRUE         TRUE         TRUE
```

Now I will upload the file

```
setwd("/home/alrier/Descargas/")
participant1=read_file("/home/alrier/Descargas/p1tr.txt")
```

Tokenization process...

```
tokensparticipant1<-tokens(participant1,what = "word",
                           remove_punct = TRUE,
                           remove_symbols =TRUE,
                           remove_numbers =TRUE,
                           remove_url =TRUE,
                           remove_separators =TRUE,
                           split_hyphens =TRUE)
tokensparticipant1
```

```
## Tokens consisting of 1 document.
## text1 :
##  [1] "Interview"   "Transcript"  "Digestive"   "Disorders"   "Brenda"
##  [6] "What"        "are"         "the"         "experiences" "of"
## [11] "people"      "with"
## [ ... and 3,279 more ]
```

This is one of the cleaning proces using regx and gsub to identify paterns into the text and delete everything that could be "contamintant"

```
text <- gsub("@\\w+", "", tokensparticipant1)
text <- gsub("https?://.+", "", text)
text <- gsub("\\d+\\w*\\d*", "", text)
text <- gsub("#\\w+", "", text)
text <- gsub("[^\x01-\x7F]", "", text)
text <- gsub("[[:punct:]]", " ", text)
# Remove spaces and newlines
text <- gsub("\n", " ", text)
text <- gsub("^\\s+", "", text)
text <- gsub("\\s+$", "", text)
text <- gsub("[ |\t]+", " ", text)
#Now i will make a second "cleaning" round, just to be sure.
# remove rt
text = gsub("rt", "", text)
# remove at
text= gsub("@\\w+", "", text)
```

```
# remove punctuation
text = gsub("[[:punct:]]", "", text)
# remove numbers (pilas con esta porque a veces los númers son útiles)
text= gsub("[[:digit:]]", "", text)
# remove links http
text= gsub("http\\w+", "", text)
# remove tabs
text = gsub("[ |\t]{2,}", "", text)
# remove blank spaces at the beginning
text = gsub("^ ", "", text)
# remove blank spaces at the end
text = gsub(" $", "", text)
#more unusual characters
a1<-gsub("[^\x01-\x7F]", "", text)
```

In this part of the code, I am celaning the word matrix.

I delete the stopwords, but I will delet every other words that I consider are not specially usefull for the analysis, convert the terms matrix into a word corpus and organize it into a vector.

```
discurso1 <- tolower(a1)
discurso1 <- removeWords(discurso1, words = stopwords("english"))
discurso1 <- removeWords(discurso1,words=c("really","like","just","can","know","don","yeah","think","wer
discurso1 <- removePunctuation(discurso1)
discurso1 <- removeNumbers(discurso1)
corpus1 <- Corpus(VectorSource(discurso1))
letras1<- TermDocumentMatrix(corpus1)
letrasmatrix1 <- as.matrix(letras1)
vector1 <- rowSums(letrasmatrix1)
Vectorr1<- sort(vector1, decreasing = T)
Vectorr1[1:30]
```

```
##      gluten        feel         eat       blood    symptoms       bowel
##          16          15          15           9           9           8
##     disease  ulcerative     colitis        sick      doctor        diet
##           7           7           7           7           7           6
##      weight      issues  medication        food   disorders intolerance
##           6           5           5           5           4           4
##         cut        best    accutane    diseases   remission        good
##           4           4           4           4           4           4
##      taking      friend        stay       gotta        make        lose
##           4           4           4           4           4           4
```

This is a matrix with the word occurrences in the text ordered by frequency.

```
dataletras1 <- data.frame(word= names(Vectorr1),freq=sort(Vectorr1, decreasing= T))
dataletras1
```

```
##                       word freq
## gluten              gluten   16
## feel                  feel   15
## eat                    eat   15
## blood               blood    9
## symptoms          symptoms    9
## bowel                bowel    8
## disease            disease    7
```

```
## ulcerative       ulcerative    7
## colitis             colitis    7
## sick                   sick    7
## doctor               doctor    7
## diet                   diet    6
## weight               weight    6
## issues               issues    5
## medication       medication    5
## food                   food    5
## disorders         disorders    4
## intolerance     intolerance    4
## cut                    cut    4
## best                   best    4
## accutane           accutane    4
## diseases           diseases    4
## remission         remission    4
## good                   good    4
## taking               taking    4
## friend               friend    4
## stay                   stay    4
## gotta                gotta    4
## make                   make    4
## lose                   lose    4
## meat                   meat    4
## sense                sense    4
## aware                aware    4
## digestive         digestive    3
## says                   says    3
## inflammatory   inflammatory    3
## label                 label    3
## body                   body    3
## protein             protein    3
## five                   five    3
## actually           actually    3
## anything           anything    3
## gotten               gotten    3
## better               better    3
## treat                 treat    3
## skin                   skin    3
## even                   even    3
## completely       completely    3
## healthy             healthy    3
## chron                chron    3
## said                   said    3
## medical             medical    3
## acid                   acid    3
## reflux               reflux    3
## gonna                gonna    3
## scope                scope    3
## look                  look    3
## two                    two    3
## way                    way    3
## middle              middle    3
## labels              labels    3
```

```
## hormones           hormones    3
## things               things    3
## hey                     hey    3
## couldn               couldn    3
## wanted               wanted    3
## week                   week    3
## able                   able    3
## mom                     mom    3
## side                   side    3
## dad                     dad    3
## doesn                doesn    3
## god                     god    3
## pay                     pay    3
## friends             friends    3
## finding             finding    3
## eating               eating    3
## least                 least    3
## kids                   kids    3
## poor                   poor    3
## balance             balance    3
## animal               animal    3
## cancer               cancer    3
## humor                 humor    3
## experiences     experiences    2
## problems           problems    2
## digestion         digestion    2
## half                   half    2
## celiac               celiac    2
## colonoscopy     colonoscopy    2
## test                   test    2
## break                 break    2
## wheat                 wheat    2
## either               either    2
## across               across    2
## board                 board    2
## joint                 joint    2
## four                   four    2
## bites                 bites    2
## pasta                 pasta    2
## else                   else    2
## headaches         headaches    2
## call                   call    2
## tried                 tried    2
## ever                   ever    2
## life                   life    2
## connection       connection    2
## shown                 shown    2
## research           research    2
## happened           happened    2
## medications     medications    2
## seen                   seen    2
## response           response    2
## markers             markers    2
## told                   told    2
```

```
younger             younger    2
three                 three    2
absorbing         absorbing    2
inflammation   inflammation    2
intense             intense    2
hell                   hell    2
never                 never    2
bathroom           bathroom    2
problem             problem    2
saying               saying    2
black                 black    2
white                 white    2
several             several    2
finally             finally    2
pubey                 pubey    2
always               always    2
tell                   tell    2
hospital           hospital    2
cup                     cup    2
send                   send    2
psych                 psych    2
thought             thought    2
figure               figure    2
wanting             wanting    2
years                 years    2
class                 class    2
gerd                   gerd    2
severe               severe    2
pizza                 pizza    2
bread                 bread    2
sugars               sugars    2
especially       especially    2
easier               easier    2
last                   last    2
next                   next    2
gets                   gets    2
specific           specific    2
times                 times    2
great                 great    2
someone             someone    2
nice                   nice    2
miserable         miserable    2
stick                 stick    2
overwhelming   overwhelming    2
grocery             grocery    2
read                   read    2
laughs               laughs    2
want                   want    2
reading             reading    2
lifestyle         lifestyle    2
changed             changed    2
done                   done    2
stomach             stomach    2
book                   book    2
```

```
## foods                foods    2
## hunter              hunter    2
## gatherer          gatherer    2
## tough                tough    2
## heavy                heavy    2
## paying              paying    2
## connected         connected    2
## lead                  lead    2
## everything       everything    2
## power                power    2
## control            control    2
## exercise           exercise    2
## stress               stress    2
## emotion            emotion    2
## need                  need    2
## present            present    2
## advocate          advocate    2
## tune                  tune    2
## watch                watch    2
## animals            animals    2
## soy                    soy    2
## felt                  felt    2
## gained              gained    2
## everyone          everyone    2
## bit                    bit    2
## course              course    2
## transcript       transcript    1
## brenda              brenda    1
## cope                  cope    1
## old                    old    1
## realize            realize    1
## exact                exact    1
## suffer              suffer    1
## define              define    1
## discomfo          discomfo    1
## marker              marker    1
## intolerant       intolerant    1
## record              record    1
## digest              digest    1
## absorb              absorb    1
## handle              handle    1
## processed         processed    1
## personally       personally    1
## pains                pains    1
## exhaustion       exhaustion    1
## incredibly       incredibly    1
## full                  full    1
## anymore             anymore    1
## nauseous           nauseous    1
## beyond              beyond    1
## migraine           migraine    1
## hormonal           hormonal    1
## acne                  acne    1
## creams              creams    1
```

```
## antibiotics     antibiotics    1
## retin                 retin    1
## wheats               wheats    1
## cleared             cleared    1
## months               months    1
## exactly             exactly    1
## appear               appear    1
## entire               entire    1
## studies             studies    1
## possibly           possibly    1
## coming               coming    1
## process             process    1
## tested               tested    1
## internal           internal    1
## medicine           medicine    1
## panel                 panel    1
## according         according    1
## ibd                     ibd    1
## gastrologist   gastrologist    1
## recommended     recommended    1
## wanna                 wanna    1
## made                   made    1
## family               family    1
## history             history    1
## unfounately     unfounately    1
## uncle                 uncle    1
## deceased           deceased    1
## second               second    1
## cousins             cousins    1
## intestines       intestines    1
## proper               proper    1
## nutrients         nutrients    1
## gut                     gut    1
## burn                   burn    1
## esophagus         esophagus    1
## mouth                 mouth    1
## rest                   rest    1
## asthma               asthma    1
## willing             willing    1
## immediately     immediately    1
## instant             instant    1
## diarrhea           diarrhea    1
## sweat                 sweat    1
## unable               unable    1
## transcribe       transcribe    1
## lesion               lesion    1
## talked               talked    1
## dermatologist dermatologist    1
## underlying       underlying    1
## hesitant           hesitant    1
## wasn                   wasn    1
## typical             typical    1
## textbook           textbook    1
## case                   case    1
```

```
## discussed         discussed    1
## interviews       interviews    1
## fear                   fear    1
## using                 using    1
## terrified         terrified    1
## changing           changing    1
## throws               throws    1
## treating           treating    1
## trial                 trial    1
## fire                   fire    1
## youngest           youngest    1
## sixth                 sixth    1
## grade                 grade    1
## come                   come    1
## severely           severely    1
## filled               filled    1
## ache                   ache    1
## rash                   rash    1
## belly                 belly    1
## distend             distend    1
## unit                   unit    1
## mental               mental    1
## filling             filling    1
## reaction           reaction    1
## gold                   gold    1
## standard           standard    1
## diagnosing       diagnosing    1
## seems                 seems    1
## reluctance       reluctance    1
## unwillingness unwillingness    1
## sure                   sure    1
## taught               taught    1
## med                     med    1
## setting             setting    1
## see                     see    1
## honesty             honesty    1
## weren                 weren    1
## page                   page    1
## egos                   egos    1
## agendas             agendas    1
## days                   days    1
## enough               enough    1
## discharge         discharge    1
## bleeding           bleeding    1
## weeks                 weeks    1
## later                 later    1
## anemic               anemic    1
## ended                 ended    1
## chance               chance    1
## practice           practice    1
## beat                   beat    1
## beast                 beast    1
## prescription   prescription    1
## tammy                 tammy    1
```

```
## none                  none    1
## parents             parents    1
## growing             growing    1
## understand       understand    1
## telling             telling    1
## labeled             labeled    1
## put                     put    1
## fact                   fact    1
## take                   take    1
## similar             similar    1
## favorite           favorite    1
## loved                 loved    1
## burgers             burgers    1
## ate                     ate    1
## balanced           balanced    1
## loading             loading    1
## funny                 funny    1
## sounded             sounded    1
## crackers           crackers    1
## basis                 basis    1
## complex             complex    1
## sugar                 sugar    1
## addicting         addicting    1
## longer               longer    1
## difficult         difficult    1
## paier                 paier    1
## weird                 weird    1
## drinking           drinking    1
## underage           underage    1
## turned               turned    1
## enjoy                 enjoy    1
## night                 night    1
## drink                 drink    1
## beers                 beers    1
## water                 water    1
## cheat                 cheat    1
## salad                 salad    1
## restaurants     restaurants    1
## accommodating accommodating    1
## feeling             feeling    1
## ostracized       ostracized    1
## restaurant       restaurant    1
## complicated     complicated    1
## experience       experience    1
## jumps                 jumps    1
## mind                   mind    1
## couple               couple    1
## ask                     ask    1
## menu                   menu    1
## year                   year    1
## ago                     ago    1
## rolls                 rolls    1
## eyes                   eyes    1
## grouped             grouped    1
```

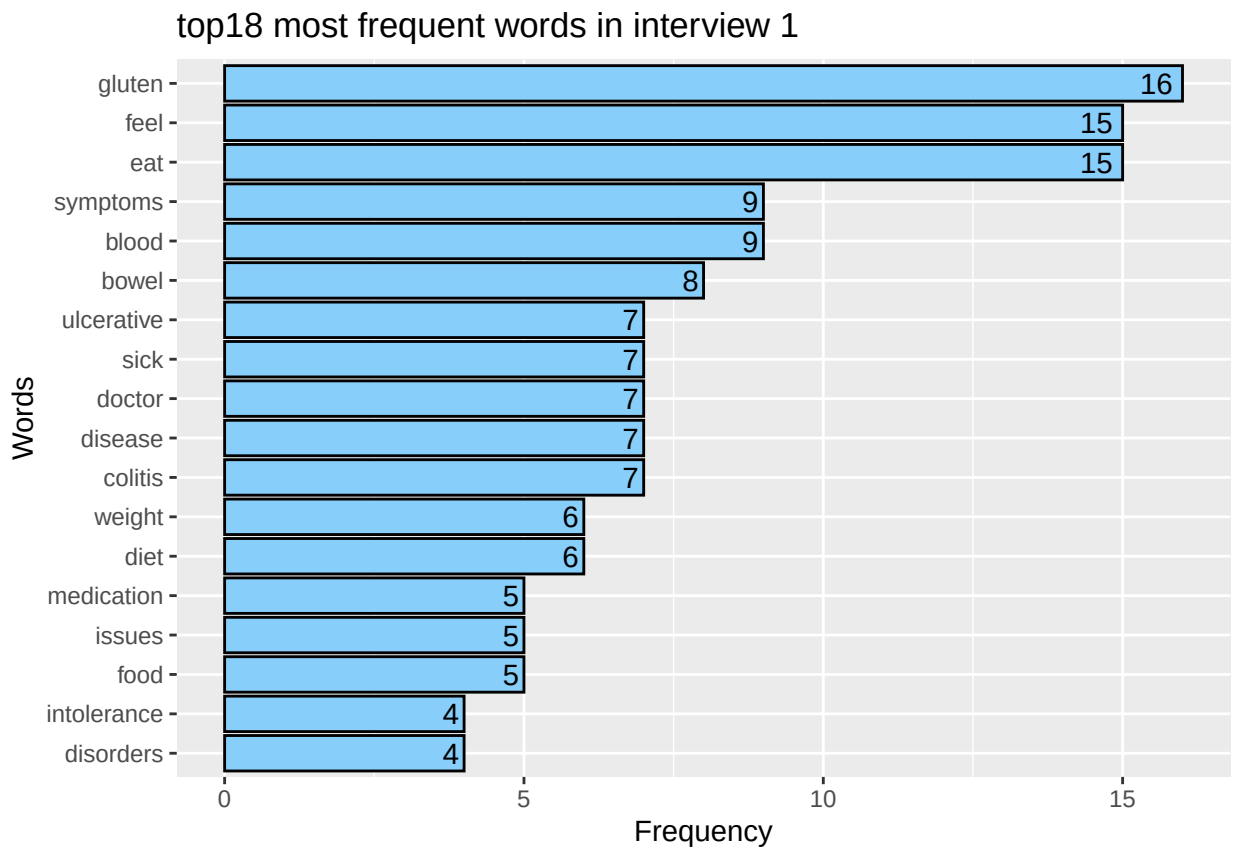
```
## cheaper             cheaper    1
## probably           probably    1
## flow                   flow    1
## cutting             cutting    1
## desperately     desperately    1
## close                 close    1
## strangers         strangers    1
## judgment           judgment    1
## media                 media    1
## hyped                 hyped    1
## won                     won    1
## mass                   mass    1
## market               market    1
## honest               honest    1
## store                 store    1
## takes                 takes    1
## forever             forever    1
## every                 every    1
## long                   long    1
## downfall           downfall    1
## places               places    1
## haven                 haven    1
## anyone               anyone    1
## home                   home    1
## addition           addition    1
## breakdown         breakdown    1
## moment               moment    1
## called               called    1
## describes         describes    1
## based                 based    1
## follow               follow    1
## fruits               fruits    1
## vegetables       vegetables    1
## alone                 alone    1
## grateful           grateful    1
## brought             brought    1
## sounds               sounds    1
## suppo                 suppo    1
## pass                   pass    1
## along                 along    1
## differences     differences    1
## teachers           teachers    1
## daughters         daughters    1
## steroids           steroids    1
## maybe                 maybe    1
## attention         attention    1
## pain                   pain    1
## shit                   shit    1
## pants                 pants    1
## understood       understood    1
## inside               inside    1
## manifest           manifest    1
## outside             outside    1
## concerned         concerned    1
```

```
## ways                  ways    1
## ticking             ticking    1
## bomb                   bomb    1
## flu                     flu    1
## far                     far    1
## adapted             adapted    1
## new                     new    1
## hardest             hardest    1
## committed         committed    1
## conquered         conquered    1
## putting             putting    1
## yoga                   yoga    1
## easily               easily    1
## feed                   feed    1
## making               making    1
## choices             choices    1
## focusing           focusing    1
## emotional         emotional    1
## tied                   tied    1
## reactions         reactions    1
## danger               danger    1
## becoming           becoming    1
## oriented           oriented    1
## battle               battle    1
## fears                 fears    1
## emotions           emotions    1
## bring                 bring    1
## back                   back    1
## trigger             trigger    1
## phrases             phrases    1
## help                   help    1
## keep                   keep    1
## happens             happens    1
## focus                 focus    1
## stresses           stresses    1
## listening         listening    1
## clear                 clear    1
## shouldn             shouldn    1
## lethargic         lethargic    1
## exhausted         exhausted    1
## immediate         immediate    1
## leaning             leaning    1
## towards             towards    1
## organic             organic    1
## documentaries documentaries    1
## treated             treated    1
## given                 given    1
## rate                   rate    1
## vegetarian       vegetarian    1
## upped                 upped    1
## grains               grains    1
## lentils             lentils    1
## supposed           supposed    1
## horrible           horrible    1
```

```
## breaking           breaking    1
## honestly           honestly    1
## unfair               unfair    1
## consumer           consumer    1
## woman                woman    1
## cranky             cranky     1
## grass                grass    1
## fed                    fed    1
## injected         injected     1
## causing           causing     1
## giving             giving     1
## infuriating   infuriating     1
## versus             versus     1
## line                 line     1
## bills               bills     1
## sho                    sho    1
## end                    end    1
## holds               holds     1
## accountable   accountable     1
## question         question     1
## check               check     1
## females           females     1
## sometimes       sometimes     1
## dinner             dinner     1
## conversations conversations  1
## habits             habits     1
## open                 open     1
## communication communication  1
## provides         provides     1
## joke                 joke     1
## accepting       accepting     1
## positive         positive     1
## move                 move     1
## forward           forward     1
## top                   top     1
## choosing         choosing     1
## relation         relation     1
```

Now I will take the top 18 most used and most important words in the document and print it in a bar plot in which I show the number of repetitions, the word and the order of "importance"

```
dataletras1[1:18, ] %>%
   mutate(word=fct_reorder(word,freq)) %>%
   ggplot(aes(word, freq)) +
   geom_bar(stat = "identity", color = "black", fill = "#87CEFA") +
   geom_text(aes(hjust = 1.3, label = freq))+
   coord_flip() +
   labs(title = "top18 most frequent words in interview 1",  x = "Words", y = "Frequency")
```



Now, I will make a beautiful wold cloud plot with the same words I have previously created in the bar plot.

```
wordcloud(words = dataletras1$word, freq = dataletras1$freq, min.freq = 4,
          max.words=30, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(7, "Dark2"), scale=c(3.5,1.25))
```

In this part starts all the sentiment analysis, I will call the syunzhet package and call the get_sentiments function.

Once the press is done, I will check the summary. It is important to mention that sentiment analysis has a variation between 1 and 0 so each emotion presented here, included the sentiments (positive and negative) are presented in the same coefficient.

To understand the summary you have to check this same coefficient in the average row.

```
sentimientos_df <- get_nrc_sentiment(discurso1, lang="english")
```

```
## Warning: `spread_()` was deprecated in tidyr 1.2.0.
## i Please use `spread()` instead.
## i The deprecated feature was likely used in the syuzhet package.
##   Please report the issue to the authors.
```

```
head(sentimientos_df)
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative positive
## 1     0            0       0    0   0       0        0     0        0        0
## 2     0            0       0    0   0       0        0     1        0        0
## 3     0            0       0    0   0       0        0     0        0        0
## 4     0            0       0    0   0       0        0     0        0        0
## 5     0            0       0    0   0       0        0     0        0        0
## 6     0            0       0    0   0       0        0     0        0        0
```

```
summary<-summary(sentimientos_df)
```

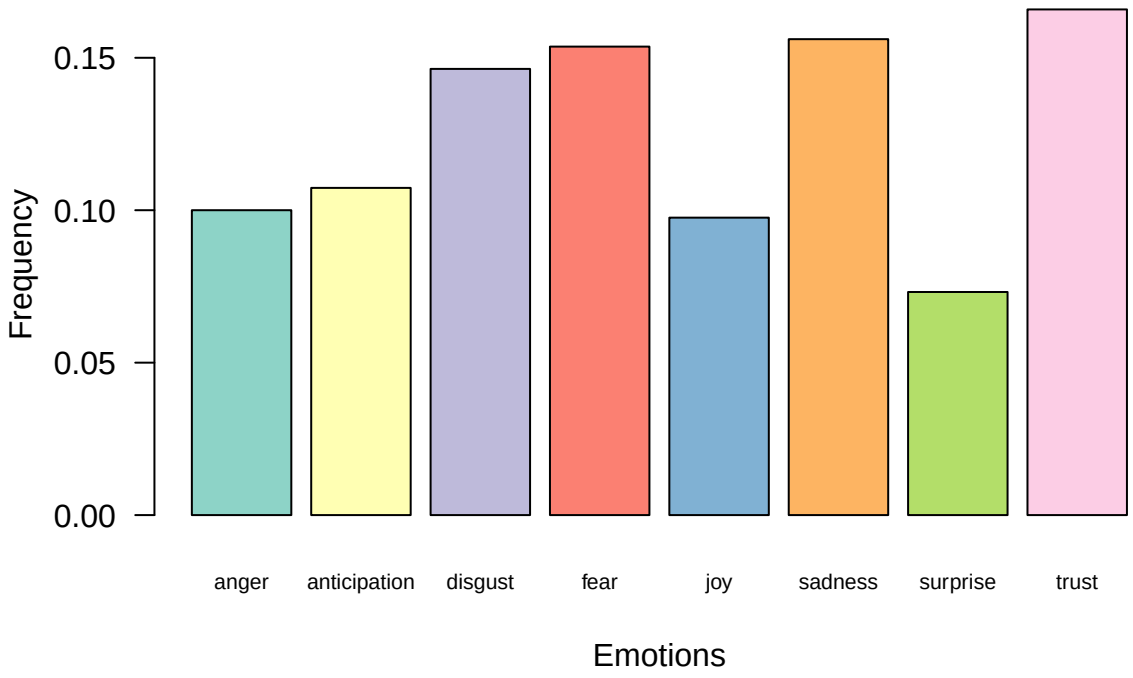
Here is he summary.

```
print(summary<-summary(sentimientos_df))
```

```
##      anger          anticipation       disgust            fear
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01246   Mean   :0.01337   Mean   :0.01823   Mean   :0.01914
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##       joy            sadness          surprise           trust
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.000000   Median :0.00000
##  Mean   :0.01215   Mean   :0.01945   Mean   :0.009116   Mean   :0.02066
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000000   Max.   :1.00000
##     negative          positive
##  Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000
##  Mean   :0.02917   Mean   :0.03464
##  3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000
```

Now I will print a bar plot with the 8 emotions that the interview number 1 express.

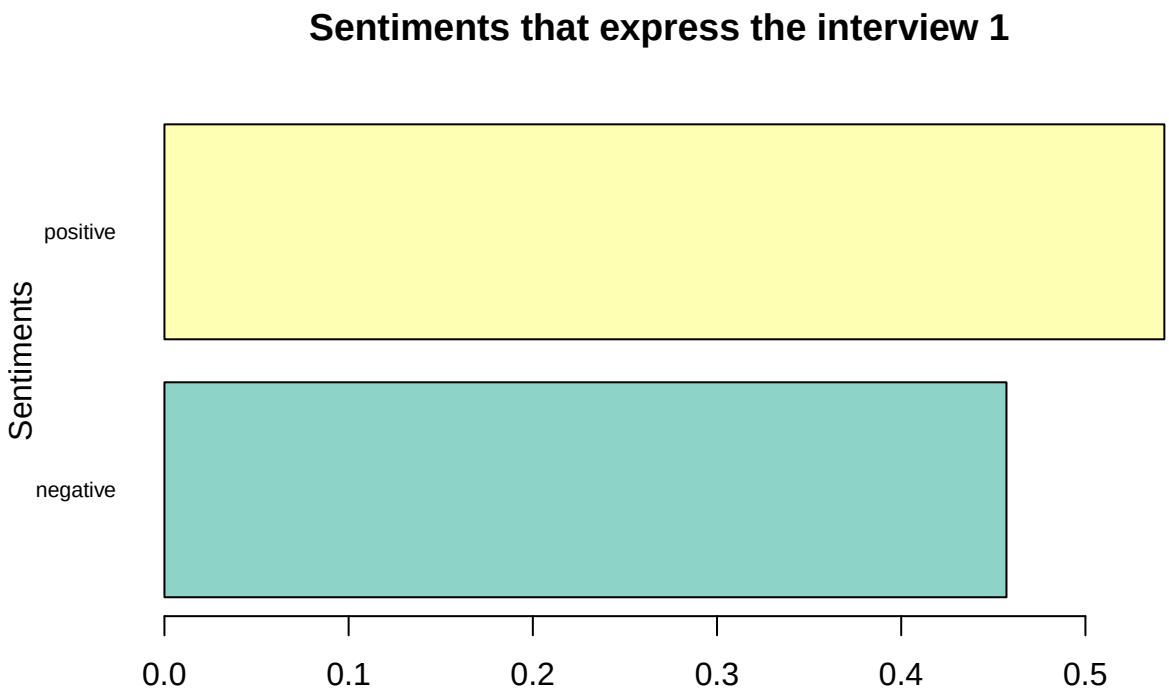
```
#graphic emotions
barplot(
  colSums(prop.table(sentimientos_df[, 1:8])),
  space = 0.2,
  horiz = F,
  las = 1,
  cex.names = 0.7,
  col = brewer.pal(n = 8, name = "Set3"),
  main = "8 different emotions expresed in the interview",
  xlab="Emotions", ylab = "Frequency")
```



Now, this is the a plot that has the two sentiments present in the document.

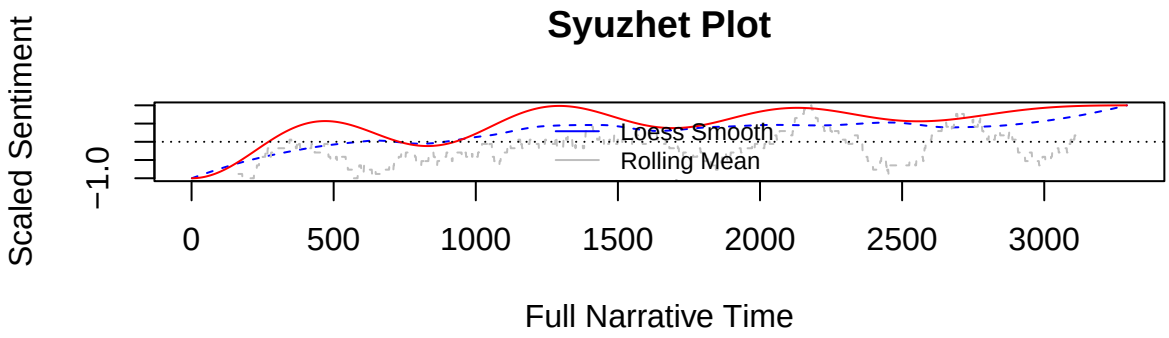
```
#graphic sentiments
barplot(
  colSums(prop.table(sentimientos_df[, 9:10])),
  space = 0.2,
  horiz = T,
  las = 1,
  cex.names = 0.7,
  col = brewer.pal(n = 2, name = "Set3"),
  main = "Sentiments that express the interview 1",
  sub = "There are two different sentiments: Positive and Negative",
  xlab="Frequency", ylab = "Sentiments")
```

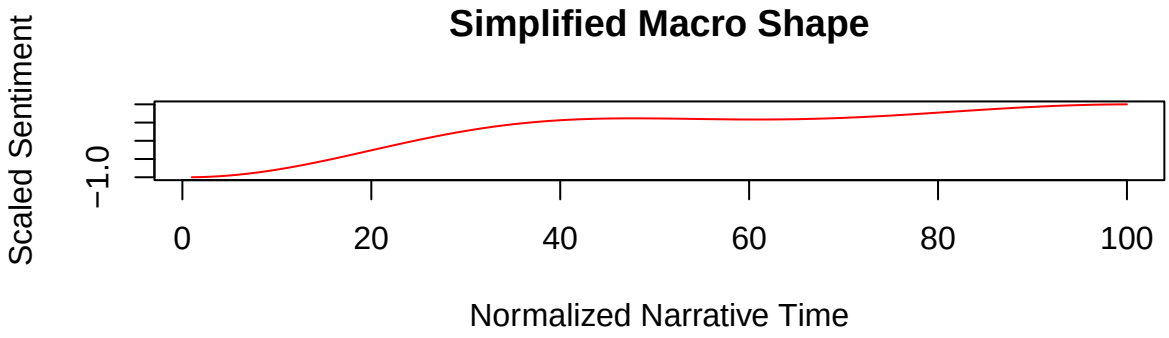
```
## Warning in brewer.pal(n = 2, name = "Set3"): minimal value for n is 3, returning requested palette wi
```



There are two different sentiments: Positive and Negative

```
#Now, in a graphic i will draw the way in which the dialog has changed between
#positive sentiments and negative ones
sentimientos_valencia <- (sentimientos_df$negative *-1)+sentimientos_df$positive
simple_plot(sentimientos_valencia)
```


Syuzhet Plot
Scaled Sentiment
−1.0
Loess Smooth
Rolling Mean
0
500
1000
1500
2000
2500
3000
Full Narrative Time


Simplified Macro Shape
Scaled Sentiment
−1.0
0
20
40
60
80
100
Normalized Narrative Time