Análise_de_sentimento

Obejetivo: Analisar o sentimento das pessoas sobre um determinado tema, baseado em tweets.

```
library(twitteR)
```

```
## Warning: package 'twitteR' was built under R version 3.6.1
```

```
library(plyr)
```

```
##
## Attaching package: 'plyr'
```

```
## The following object is masked from 'package:twitteR':
##
##     id
```

```
library(stringr)
library(tm)
```

```
## Warning: package 'tm' was built under R version 3.6.1
```

```
## Loading required package: NLP
```

```
library(sentimentr)
```

```
## Warning: package 'sentimentr' was built under R version 3.6.1
```

```
library(Rstem)
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.6.1
```

```
library(httr)
```

```
##
## Attaching package: 'httr'
```

```
## The following object is masked from 'package:NLP':
##
##     content
```

```
library(rmarkdown)
```

```
## Warning: package 'rmarkdown' was built under R version 3.6.1
```

```
source("utils.R")
```

Criando autenticação no Twitter.

```
key<-"X3tLabKseosqNfnfZIaU2B829"
secret<-"5VsdANG35oBKmrnDd6FTEhGgaklGKX8ujOf0umLwaRatCC8RFg"
token<-"890217190577897473-BIjMvFNPIL6IUQH4TurpfrZ8ZC0b40n"
tokensecret<-"nRrfW2ACMQ2qo7aRgrvonXHI0A2EFeUerpwNnm8Resthy"
setup_twitter_oauth(key, secret, token, tokensecret)
```

```
## [1] "Using direct authentication"
```

Definindo as características da busca: palavra a ser encontrada, idioma, número de tweets.

```
tema<-"Bolsonaro"
lingua<-"pt"
qtd_tweets<-1000
tweetdata=searchTwitter(tema,n=qtd_tweets,lang=lingua)
head(tweetdata)
```

```
## [[1]]
## [1] "JustBira: RT @BolsonaroSP: Bom artigo em que se vê um paralelo econômico entre Brasil e Argenti
##
## [[2]]
## [1] "SalgadoTemplar: RT @AlanaPassosRJ: Se Bolsonaro não fez nada até agora, quem isso aqui<U+0001F9
##
## [[3]]
## [1] "DoniseteSilvei1: RT @cpvpmmg_valente: <U+0001F4CC> Parte 1\nDiscurso de @jairbolsonaro em Parna
##
## [[4]]
## [1] "JorgeCa20557834: RT @BolsonaroSP: Bom artigo em que se vê um paralelo econômico entre Brasil e
##
## [[5]]
## [1] "MarciaSAlmeida: RT @gomes_wilmar: O jogo sujo da Globo e o real motivo de toda a artilharia da
##
## [[6]]
## [1] "kkkkkkkkkknyr: RT @MongeHan: Glr falando que o Bolsonaro ta certo em tirar o hrario de verão pq
```

Tratamento dos dados coletados: limpeza,organização e transformação.

```
library(SnowballC)
```

```
##
## Attaching package: 'SnowballC'

## The following objects are masked from 'package:Rstem':
##
##     getStemLanguages, wordStem
```

```
tweetlist<-sapply(tweetdata, function(x) x$getText())
tweetlist<-iconv(tweetlist, to="utf-8", sub="")
tweetlist<-limpaTweets(tweetlist)
tweetlist<-gsub("afa","",tweetlist)
tweetcorpus<-Corpus(VectorSource(tweetlist))
tweetcorpus<-tm_map(tweetcorpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(tweetcorpus, removePunctuation):
## transformation drops documents
```

```
tweetcorpus<-tm_map(tweetcorpus, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(tweetcorpus, content_transformer(tolower)):
## transformation drops documents
```

```
tweetcorpus<-tm_map(tweetcorpus, function(x)removeWords(x, stopwords()))
```

```
## Warning in tm_map.SimpleCorpus(tweetcorpus, function(x) removeWords(x,
## stopwords())): transformation drops documents
```

Convertendo o objeto Corpus para texto plano.

```
termo_por_documento= as.matrix(TermDocumentMatrix(tweetcorpus), control=list(stopwords=c(stopwords("por
```

Criando a nuvem de palavras.

```
library(RColorBrewer)
library(wordcloud)
```

```
## Warning: package 'wordcloud' was built under R version 3.6.1
```

```
pal2<-brewer.pal(8, "Dark2")
wordcloud(tweetcorpus, min.freq = 2, scale=c(5,1), random.color = F, max.words = 60, random.order = F,
```

Convertendo o objeto texto para o formato de matriz.

```
tweettdm<-TermDocumentMatrix(tweetcorpus)
tweettdm
```

```
## <<TermDocumentMatrix (terms: 1981, documents: 1000)>>
## Non-/sparse entries: 9791/1971209
## Sparsity           : 100%
## Maximal term length: 18
## Weighting          : term frequency (tf)
```

Analisando os dados.

```
findFreqTerms(tweettdm, lowfreq = 20) #20 palavras mais frequentes
```

```
##   [1] "argentina"  "artigo"     "bem"        "bolsonaro"  "bom"
##   [6] "brasil"     "como"       "econ"       "entre"      "fala"
##  [11] "mico"       "motivo"     "paralelo"   "que"        "sucesso"
##  [16] "globo"      "para"       "mas"        "faz"        "presidente"
##  [21] "sobre"      "com"        "mica"       "nordeste"   "pol"
##  [26] "tem"        "das"        "lia"        "uma"        "por"
##  [31] "primeiro"   "quem"       "ser"        "acabar"     "coc"
##  [36] "jair"       "vamos"      "diplomacia" "eduardo"    "fila"
##  [41] "foi"        "insiste"    "negociador" "prio"       "reprovado"
##  [46] "seu"        "teste"      "novo"       "governo"    "ela"
##  [51] "est"        "aprovar"    "diz"        "ncia"       "pode"
```

```
##  [56] "reforma"    "vai"        "comunistas" "corruptos"  "essa"
##  [61] "turma"      "varrer"     "vermelha"   "aos"        "domingos"
##  [66] "mais"       "trabalhar"  "voc"        "ladr"       "parna"
##  [71] "blica"      "elei"       "dia"        "eoea"       "piau"
##  [76] "recep"      "anistia"    "belo"       "constru"    "construir"
##  [81] "dinheiro"   "dios"       "horizonte"  "memorial"   "mila"
##  [86] "pres"       "tiraram"    "agora"      "aprovado"   "empresa"
##  [91] "feriados"   "obrigar"    "ontem"      "pagar"      "amado"
##  [96] "olha"       "pai"        "trabalho"   "deltan"     "pra"
## [101] "acabam"     "feriaa"     "mara"       "principal"  "regra"
## [106] "texto"      "torna"
```

```
findAssocs(tweettdm, "reforma", 0.6) #Buscando associação
```

```
## $reforma
##   anistia   constru construir  dinheiro      dios  memorial      mila
##      0.93      0.93      0.93      0.93      0.93      0.93      0.93
##   tiraram      belo horizonte      pres
##      0.93      0.92      0.92      0.90
```

Classificando os dados em sentimentos.

```
library(Rstem)
library(sentiment)
library(ggplot2)
```

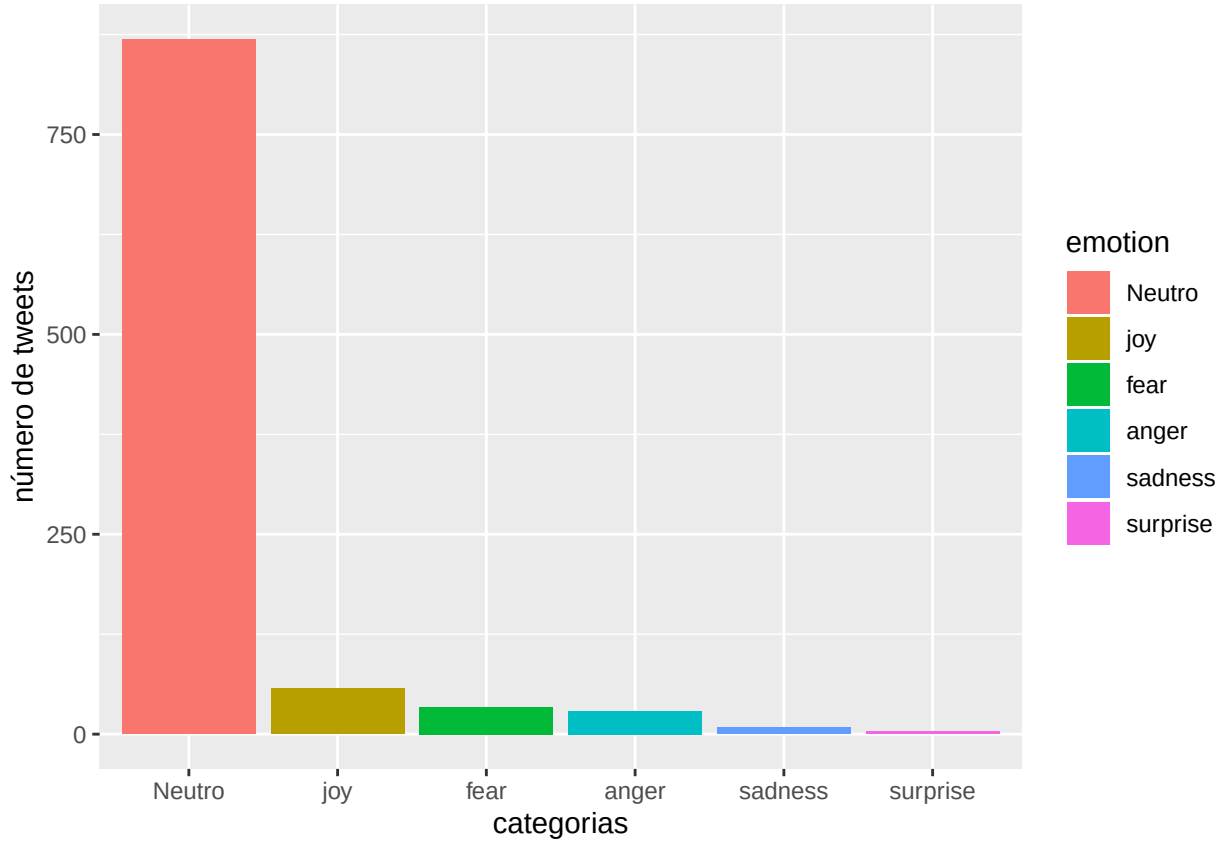
```
##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:NLP':
##
##     annotate
```

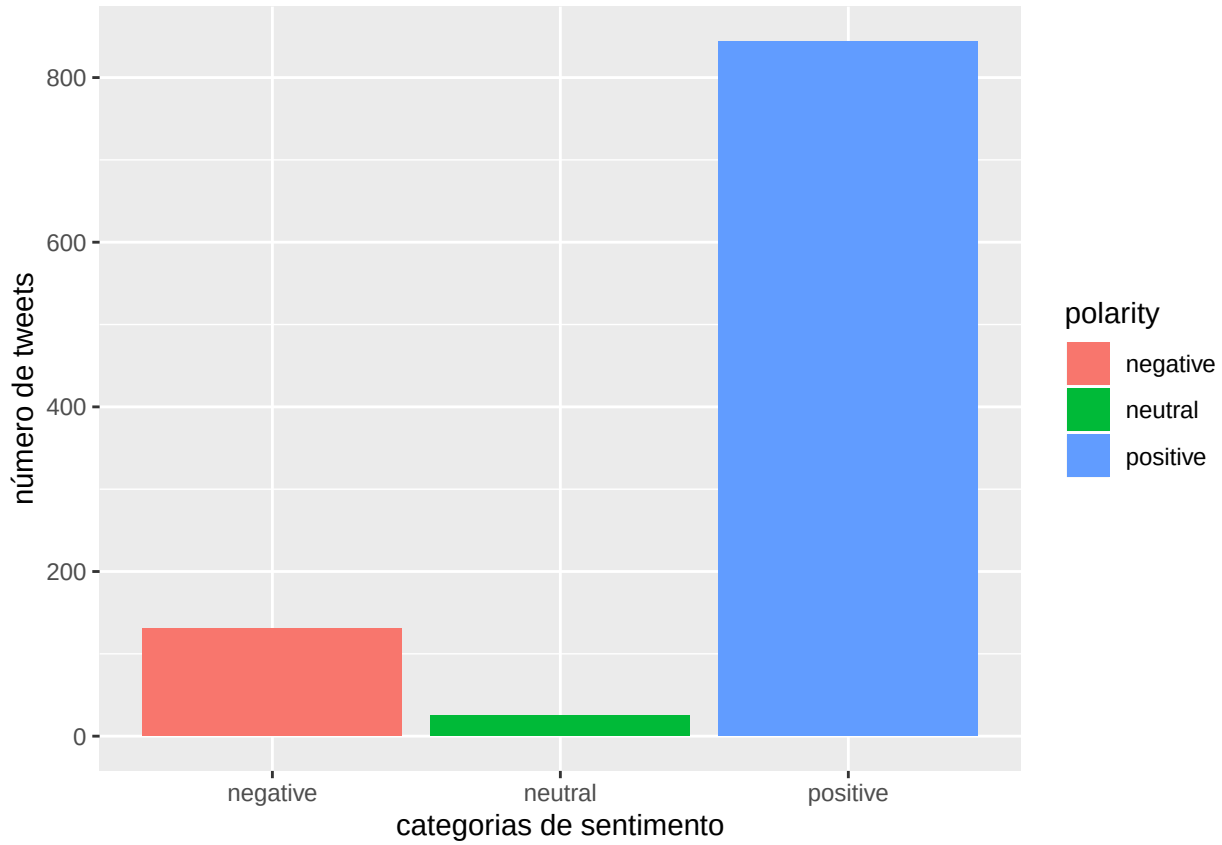
```
class_emo=classify_emotion(tweetlist, algorithm = "bayes", prior=1.0)
emotion=class_emo[,7]
emotion[is.na(emotion)]="Neutro"
class_pol=classify_polarity(tweetlist, algorithm = "bayes")
polarity=class_pol[,4]
sent_df=data.frame(text=tweetlist, emotion=emotion, polarity=polarity, stringsAsFactors = F)
sent_df=within(sent_df, emotion<-factor(emotion, levels=names(sort(table(emotion), decreasing = T))))
```

Visualizando por meio de gráficos

```
#Emoções encontradas
ggplot(sent_df, aes(x=emotion))+geom_bar(aes(y=..count..,fill=emotion))+labs(x="categorias", y="número
```




```
#Polarização
ggplot(sent_df, aes(x=polarity))+geom_bar(aes(y=..count..,fill=polarity))+labs(x="categorias de sentimer
```


800
600
400
200
0
número de tweets
negative
neutral
positive
categorias de sentimento
polarity
negative
neutral
positive