

Rtweet

Análise de Texto - 2022.1

Letícia Lino, Lucas Augusto, Lucas Coelho e Pedro Aguiar

12 de setembro de 2022

Analizando textos do Twitter

O pacote Rtweet

Primeiro acesso

A API do twitter nos disponibilizará os dados que precisamos, mas antes, é necessário obtermos alguns “credenciamentos” para acessá-la:

- Consumer Key
- Consumer Secret
- Access Token
- Access Secret

Para isso, basta acessar esse link <https://developer.twitter.com/en/apps> e seguir os passos a partir de “Create an App”. Após a primeira solicitação de qualquer função do pacote, o R cobrará suas credenciais.

O Pacote

- Com o pacote `rtweet`, podemos localizar os últimos tweets feitos com uma palavra de interesse, por default, a função busca os últimos 100 tweets, mas esses parâmetros podem ser alterados.
- *Atenção!* - O pacote “`rtweet`” *carrega até 18000 tweets a cada 15 minutos*, para cada usuário, entretanto, se for de interesse o parâmetro “`retryonratelimit = TRUE`” pode ser utilizado para que mais tweets sejam carregados. O R contará o tempo para que a próxima consulta possa ser feita!

Ressalva!

- Utilizando rtweet, conseguimos trabalhar apenas com os dados fornecidos nos últimos 6 a 9 dias. Não poderíamos, por exemplo, estudar eleições passadas.

Get_...()

Para usuários

- get_favorites

Get_...()

Para usuários

- get_favorites
- get_followers

Get_...()

Para usuários

- get_favorites
- get_followers
- get_friends

Get_...()

Para usuários

- get_favorites
- get_followers
- get_friends
- get_mentions

Get_...()

Para usuários

- get_favorites
- get_followers
- get_friends
- get_mentions
- get_retweets

Get_...()

Para usuários

- get_favorites
- get_followers
- get_friends
- get_mentions
- get_retweets
- get_timeline

Get_...()

Para usuários

- get_favorites
- get_followers
- get_friends
- get_mentions
- get_retweets
- get_timeline
- get_trends

Lookup!

- `lookup_coords`

Lookup!

- lookup_coords
- lookup_friendships

Lookup!

- lookup_coords
- lookup_friendships
- lookup_users

- ts_data

Atenção! - ts_data converte dados do twitter, em uma espécie de série temporal, essa função pode, por exemplo, juntar as informações temporais (get_timeline) de dois usuários, ou seja, retorna a frequência de tweets em um intervalo de tempo específico.

- ts_plot monta o “gráfico temporal”.

- ts_data
- ts_plot

Atenção! - ts_data converte dados do twitter, em uma espécie de série temporal, essa função pode, por exemplo, juntar as informações temporais (get_timeline) de dois usuários, ou seja, retorna a frequência de tweets em um intervalo de tempo específico.

- ts_plot monta o “gráfico temporal”.

Algumas funções do Rtweet

Procurando termos.

- search_tweets

```
library(rtweet)
library(dplyr)
tweets <- search_tweets("eleições 2022",
                        n = 1800,
                        include_rts = FALSE)
```

Também podemos procurar por hastags!

```
tweetseleicoes <- search_tweets("#eleições2022",
                                type = "popular",
                                n=3)
```

Da onde vieram esses tweets?



Para observações

Coletando, limpando e analisando tweets

```
texto = tweets$full_text
```

Tm

```
library(tm)
texto <- texto %>%
  (removePunctuation) %>%
  (removeNumbers) %>%
  (stripWhitespace)
```

Criando o Corpus

- Corpus = texto + metadados

```
library(tm)
dados <- Corpus(VectorSource(texto))
dados <- dados %>%
  tm_map(content_transformer(tolower))

dados <- dados %>%
  tm_map(removeWords,
          stopwords("pt")) #stopword
dados <- dados %>%
  tm_map(stemDocument) #juntando termos
```

removendo https

```
#removendo "http"  
URL <- function(x) gsub("http[[:alnum:]]*", "", x)  
dados <- tm_map(dados, URL)
```

Criando uma matrix com a frequência de palavras

```
amatrix<- TermDocumentMatrix(dados)
amatrix<-removeSparseTerms(amatrix, 0.98)
matrix <-as.matrix(amatrix)
matrix<-sort(rowSums(matrix), decreasing = TRUE)
```

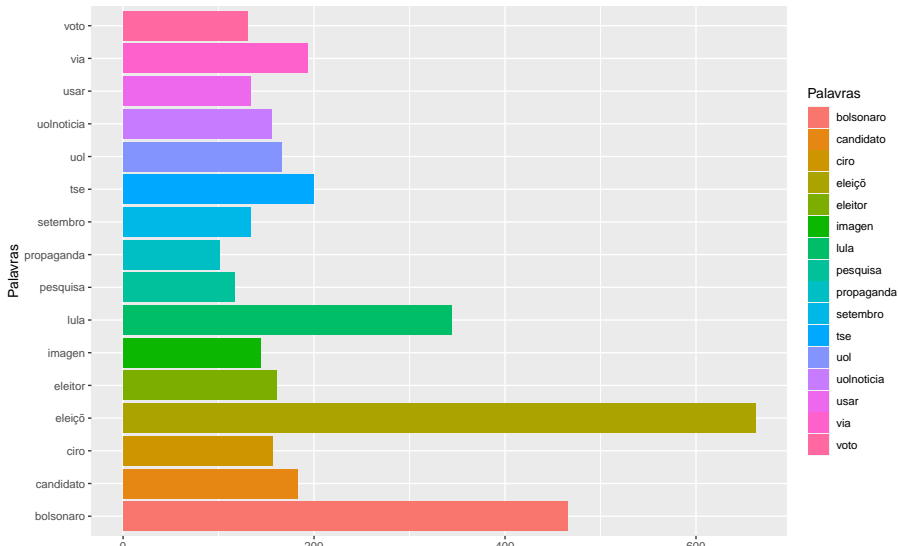
Criando um Data Frame para os gráficos que serão plotados,

```
df_tweets<-data.frame(Palavras = names(matrix),
                      Frequencia = matrix)
```

Frequência de Palavras

Palavras mais citadas nos últimos dias

A partir de tweets que mencionavam eleições 2022



“findFreqTerms” do pacote tm

- Podemos também visualizar as palavras mais utilizando essa nova função.

```
freq <- findFreqTerms(amatrix,  
                      lowfreq = 95)  
freq
```

```
[1] “eleiçõ” “bolsonarista” “vai” “lula” “voto”  
[6] “bolsonaro” “pesquisa” “eleitor” “candidato” “uol”  
[11] “uolnoticia” “via” “campanha” “tse” “ciro”  
[16] “usar” “imagen” “proíb” “setembro” “propaganda”
```

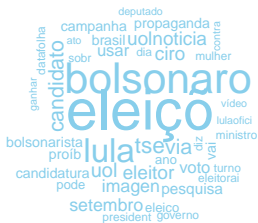
Achando Associação entre termos

```
associacao<-findAssocs(amatrix,  
  "voto",  
  0.1)
```

Termos	Associação
Pode	0.44
Ciro	0.26
Lula	0.23
Eleição	0.23
Primeiro	0.22
Turno	0.22

Nuvem de Palavras

- Wordcloud



- Wordcloud2

O Wordcloud2 é um pacote que nos permitirá plotar uma nuvem de palavras um pouco mais sofisticada.

- interface HTML
- letterCloud
- figPath
- alteração de cor, grafite, vetores de cores, rotação, forma da nuvem e outros.

Funções “get_timeline + ts_plot

- get_timeline(c(“usuario1”, “usuario2”, “usuario3”))
- ts_plot

```
library(dplyr)
jair <- get_timeline("jairbolsonaro",
                     n = 100)
lula <- get_timeline("LulaOficial",
                     n = 100,)
ciro <- get_timeline("cirogomes",
                     n = 100,)
```

```
jair <- cbind(jair, users_data(jair)
              [, c("name", "screen_name")])
lula <- cbind(lula, users_data(lula)
              [, c("name", "screen_name")])
ciro <- cbind(ciro, users_data(ciro)
              [, c("name", "screen_name")])
linhatempo <- rbind(jair, lula, ciro)
linhatempo <- linhatempo %>%
  mutate(Candidato = name)
```

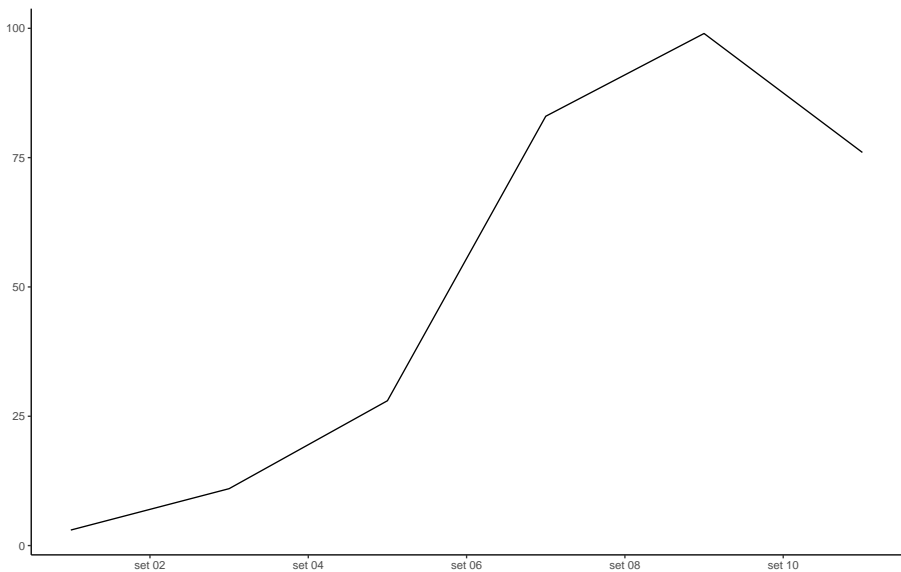
O quanto os candidatos 2022 tweetaram nos últimos dias?

Ts_plot

```
library(ggplot2)
ts_plot(data = linhatepo,
        by = "2 days") +
  theme_classic() +
  labs(
    x = NULL,
    y = NULL,
    title = "Tweets feitos pelos Candidatos",
    subtitle = "Tweets a cada 2 dias.",
    caption = "Fonte: Dados coletados pela API do Twitter")
```

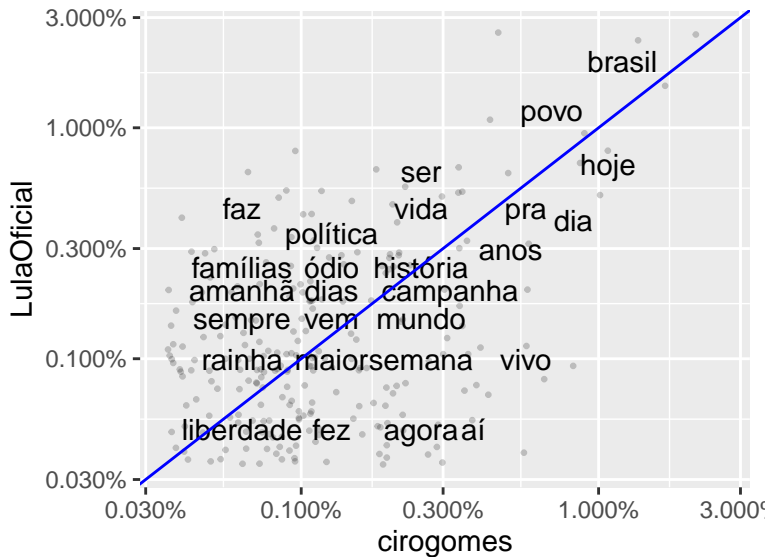
Tweets feitos pelos Candidatos 2022 nos últimos dias

Tweets a cada 2 dias.

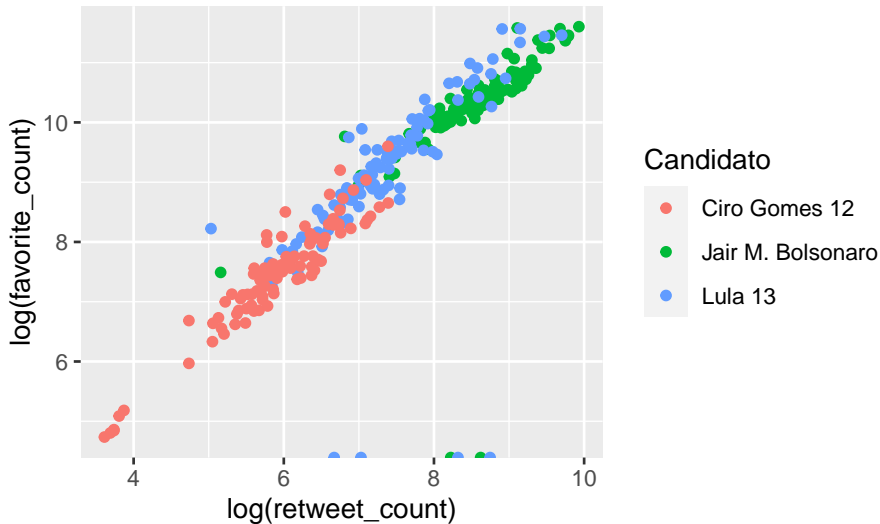


Fonte: Dados coletados pela API do Twitter

Palavras Citadas



Likes e RTs!



Outra Análise:

Média de RTs

```
MediaRT <- linhatepo %>%
  group_by(Candidato) %>%
  summarise(mean(retweet count))
```

MediaRT

```
## # A tibble: 3 x 2
```

```
## Candidato      'mean(retweet_count)'  
## <chr>          <dbl>  
## 1 Ciro Gomes 12      476.  
## 2 Jair M. Bolsonaro 6412.  
## 3 Lula 13        2433.
```

Média de favoritos

```
MediaFav<-linhatempo %>%  
  group_by(Candidato) %>%  
  summarise(mean(favorite_count))  
MediaFav
```

```
## # A tibble: 3 x 2  
##   Candidato      'mean(favorite_count)'  
##   <chr>          <dbl>  
## 1 1 2018-09-12 2447.  
## 2 2018-09-12 37278.  
## 3 2018-09-12 18446.
```

API GOOGLE MAPS

- Permite a utilização gratuita de funções do Google Maps
- Chave de autenticação.
- 25.000 requisições gratuitas.

Tweets que saíram do Brasil.

```
library("rtweet")
nobrasil <- search_tweets("#rockinrio",
                          geocode = lookup_coords("brazil"),
                          type = "popular")
```

Tweets em português, enviados dos Estados Unidos

```
usa<-search_tweets("#rockinrio",
                   lang="pt-BR",
                   type = "recent",
                   geocode = lookup_coords("usa"),
                   n = 3)
```

E as trends?

```
viral <- get_trends("salvador")  
momentos<-viral$trend  
head(momentos,  
      n=5)
```

```
[1] "Enderson" "pardo" "IBGE" "Ludmila" "Anitta"
```

Obrigado!

Todos os códigos, arquivos, imagens e outras informações úteis apresentadas hoje estão no nosso git!

- <https://github.com/pedroaguiar15/C.E.2-AnaliseTexto-2022.1>

Letícia Lino, Lucas Augusto, Lucas Coelho e Pedro Aguiar

Bibliografia

- <https://rdocumentation.org/packages/rtweet/versions/0.7.0>
- <https://www.lume.ufrgs.br/bitstream/handle/10183/149102/001004730.pdf?sequence=1>
- <https://cran.r-project.org/web/packages/rtweet/rtweet.pdf> -
- https://www.gbailey.uk/twitter_workshop/pt2_collection.html