## Analise de Sentimentos - Rio2016

Guilherme Righetto
12 de fevereiro de 2017

#### Análise de Sentimentos

Este é um projeto sobre a análise de sentimentos a partir dos tweets das olimpíadas do Rio 2016. O objetivo é captutar dados do Twitter e realizar a criação de uma nuvem das palavras e de um análise de sentimentos (Positivo/Negativo/Neutro). Todo o projeto será descrito nas próximas seções.

#### **Pacotes**

```
library(twitteR)
library(tm)
library(wordcloud)
library(SnowballC)
library(Rstem)
library(sentiment)
library(plotly)
library(plyr)
```

## Autenticação do Twitter

O código apresentado realiza a autenticação realizada pelo pacote 'twitteR', é válido ressaltar que é necessário criar um aplicação na área de desenvolvedor do twitter.

```
api_key <- 'IpbVirlaHHLkTP8MfkL5Yuavm'
api_secret <- '7f0U01fQj5QBLNxRxo85k9k1Hze7HDl0Mta0Bj0KTwynnSCulA'
token <- '830527057423691776-hVh1j10t8yNsX6f98NVwyEAfwFvYts0'
token_secret <- 'vAwDfU9buhjd7yaFrdyv5zcCo2HWefTd6amN7L5K79LVY'
setup_twitter_oauth(api_key, api_secret, token, token_secret)</pre>
```

## [1] "Using direct authentication"

## Carregar Tweets

As próximas funções executam uma busca de 500 tweets com a string de busca '#Rio2016' em português e depois transformam os dados em um data frame.

```
tweets <- searchTwitter("#Rio2016",n=500, lang = 'pt')
tweetpt <- twListToDF(tweets)</pre>
```

## Limpeza e Pré-processamento dos Dados

Nesta etapa, será realizado a limpeza de dados como também o agrupamento de todos os tweets utilizando o pacote de text mining(tm). A limpeza realizada é referente aos caracteres especiais, acentuação, números e stopwords.

```
tweetpt <- iconv(tweetpt$text, 'UTF-8', 'latin1', 'byte')
tweetpt = gsub("https://.+","",tweetpt)

tweetpt = gsub("<.+>","",tweetpt)

tweetpt <- paste(tweetpt, collapse = ' ')

s_words <- c("sobre","após","httpstcopeidluqlc","seis","xedxaxbdxedxbxa","well","superb")
tweetpt = Corpus(VectorSource(tweetpt))
tweetpt = gsub("(RT|via)((?:\b\\W*@\\w+)+)", "", tweetpt)
tweetpt = Corpus(VectorSource(tweetpt))
tweetpt <- tm_map(tweetpt, content_transformer(tolower))
tweetpt <- tm_map(tweetpt, removeNumbers)
tweetpt <- tm_map(tweetpt, removePunctuation)
tweetpt = tm_map(tweetpt, removeWords, stopwords('pt'))
tweetpt = tm_map(tweetpt, removeWords, s_words)
tweetpt <- tm_map(tweetpt, PlainTextDocument)</pre>
```

#### Nuvem de Palavras

Construção de uma nuvem de palavras utilizando os dados.

```
set.seed(111)
wordcloud(tweetpt, min.freq = 3,max.words=200, random.order=FALSE, rot.per=0.35, colors=brewer.pal(8, "
                                                                                                                                                                                                      Conquistando prataginda gumenta di sides tempo conquistando prataginda gumenta conquistando en la culturalnrio país de conquistante en la culturalnrio país de conquis
                                                                                                                                                                                                                                                                                                                                                                                                                                                     castelo esportivo delegação favor
                                                                      articipantes and comentar articipantes and ingresso of today and ingresso of today and ingresso of escritorio of injudgosparalimpicos of injudgosparalimpicos of injudgosparalimpicos of injudgosparalimpicos of injudgosparalimpicos
                                                                                                                                                                                                                          zada
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              odebrecht ataca
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              ltimo outro Öneymarjr
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   foda
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        etas etas
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  elipe paragraphical sortes sor
                                                                                                                              tudo ver brasilbra
                                                                                                                                                                                                                                                                                                              matéria 🛫
                                                                                 nosken prasilbra diado go nosken olímpico diagona olímpico famoso
                                                                                                                                   Cerimônia deu entraria gaucho motivo fossemos lindo, elegado character mudou felicidade bolt mobilidade mobilidade modifica de la construitu o exatos exatos
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               abandono
                                                                                                                                                                                                                                                                                                                                                                                                                                            jogosolimpicos de o
                                                                                                                                                                                                                                                               mobilidade
                                                                                                                                                                                                                                                      norteamericana
saudadessobraram
```

## Análise de Sentimentos

Nesta seção será realizado uma análise de sentimentos utilizando os pacotes 'Rstem' e 'sentiment'. Assim, os dados serão carregados e pré-processados novamente.

```
tweetst <- twListToDF(tweets)
tweetst <- iconv(tweetst$text, 'UTF-8', 'latin1', 'byte')</pre>
```

## Classificação

O código a seguir utiliza a função 'classify\_polarity' que classifica os tweets em três classes (Positive, Negative e Neutral) utilizando o algoritmo Naive Bayes.

```
class_pol = classify_polarity(tweetst, algorithm = "bayes")
polarity = class_pol[,4]
```

## Adequação

Nesta etapa será realizado uma adequação do retorno do classificador para construir um gráfico de pizza com melhor entendimento.

```
sent_df = data.frame(text = tweetst, polarity = polarity, stringsAsFactors = FALSE)

sent_df$polarity = gsub("positive","POSITIVO",sent_df$polarity)
sent_df$polarity = gsub("negative","NEGATIVO",sent_df$polarity)
sent_df$polarity = gsub("neutral","NEUTRO",sent_df$polarity)

sent_df <- count(sent_df,~polarity)
sent_df$polarity <- paste(sent_df$polarity,sent_df$freq)</pre>
```

## Gráfico de Pizza

Construção do gráfico de pizza utilizando o pacote 'plotly'.

# Análise de sentimentos sobre as Olimpiadas Rio2016

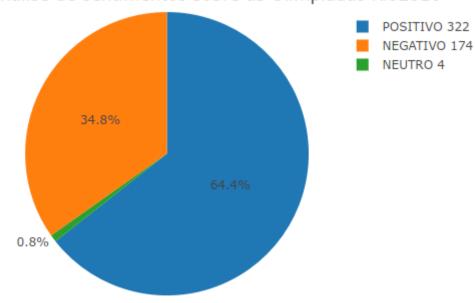


Figure 1: Análise de Sentimentos.