



TWITTADAS PRESIDENCIAIS: ANÁLISE DOS TWEETS DO ATUAL E ALGUNS EX PRESIDENTES DO BRASIL

Luiz Fernando Coelho Passos¹, Daniel dos Santos², Lyncoln Sousa de Oliveira³ e

Gabriel Perez Claro Santana Mizuno⁴

Introdução

Com o surgimento e avanço da globalização a acessibilidade ao acesso à internet vem crescendo continuamente e alcançado todas as classes sociais. Com isso é bem comum as redes sociais ganharem mais usuários e consequentemente visualização. Além dessa expansão observou-se que muitos governantes, de diversos países (desenvolvidos ou não), vem usando redes sociais para comunicar-se com internautas de seus países.

Dentre estas plataformas se destaca o Twitter, uma rede social focada em comunicação por mensagens textuais curtas e diretas (tweets), onde um governante pode se expressar e ter uma interação rápida com seus seguidores, assim tendo um feedback dos internautas em relação a sua mensagem.

Esses tweets são dados e sua análise gera informações relevantes sobre o usuário, podendo ser utilizadas para observar como se comporta a frequência das palavras que os governantes usam para se comunicar, ou até mesmo usar mecanismos de análise de sentimentos para ranquear essas mensagens em números positivos ou negativos.

Objetivos

Objetivo Geral: Analisar o twitter do presidente e ex presidentes utilizando as linguagens computacionais R e Python.

Objetivo Específico: Extrair os tweets dos presidentes; descobrir as palavras mais frequentes; analisar o sentimento dos tweets e visualizar o mais positivo e mais negativo.

¹ Universidade Federal Fluminense/UFF – Dep. Estatística, lfcpassos@id.uff.br

² Universidade Federal Fluminense/UFF – Dep. Estatística, dd_santos@id.uff.br

³ Universidade Federal Fluminense/UFF – Dep. Estatística, lyncolnsousa@id.uff.br

⁴ Universidade Federal Fluminense/UFF – Dep. Estatística, gabrielmizuno@gmail.com



Material e Método

Os conceitos e etapas foram baseados em (SILGE e ROBINSON, 2016). O método segue 2 etapas: para obter os tweets do presidente e ex presidentes foi necessário utilizar a API do Twitter, cujo acesso foi feito através da biblioteca *tweepy* da linguagem computacional python e as análises foram feitas utilizando as bibliotecas *tidytext* para manipulação dos tweets, *ggplot2* para visualização em gráfico de barras das 20 palavras mais frequentes, *ggwordcloud* para confecção da nuvem de palavras, *ggraph* para confecção de grafos e *lexiconPT* para realizar a análise de sentimento.

Foi criado um repositório no diretório *github*, onde pode ser acessado pela URL <https://github.com/lyncoln/Twittadas-Presidenciais> contendo a estrutura do estudo, com as seguintes pastas:

- code: scripts contendo os códigos em R e Python
- data: arquivos de dados
- man: arquivos de imagens

Resultados e Discussão

Foram analisados os últimos 16.090 tweets, coletados no dia 03 de abril de 2019, divididos entre o governante e ex-governantes: Jair Bolsonaro, Dilma Rousseff, Fernando Collor de Melo, Michel Temer e Luiz Inácio Lula da Silva.

Para obtenção dos resultados foram usados as seguintes análises gráficas: Nuvem de Palavras, Gráfico de Barras e Grafos, a partir delas pode-se observar:

- Na Nuvem de Palavras as 150 palavras mais frequentes de cada presidente sendo diferenciadas pelo tamanho da própria palavra.
- No Gráfico de Barras as 20 palavras mais frequentes de cada presidente e a quantidade das mesmas.
- A relação entre as palavras na visualização de Grafos Orientados.

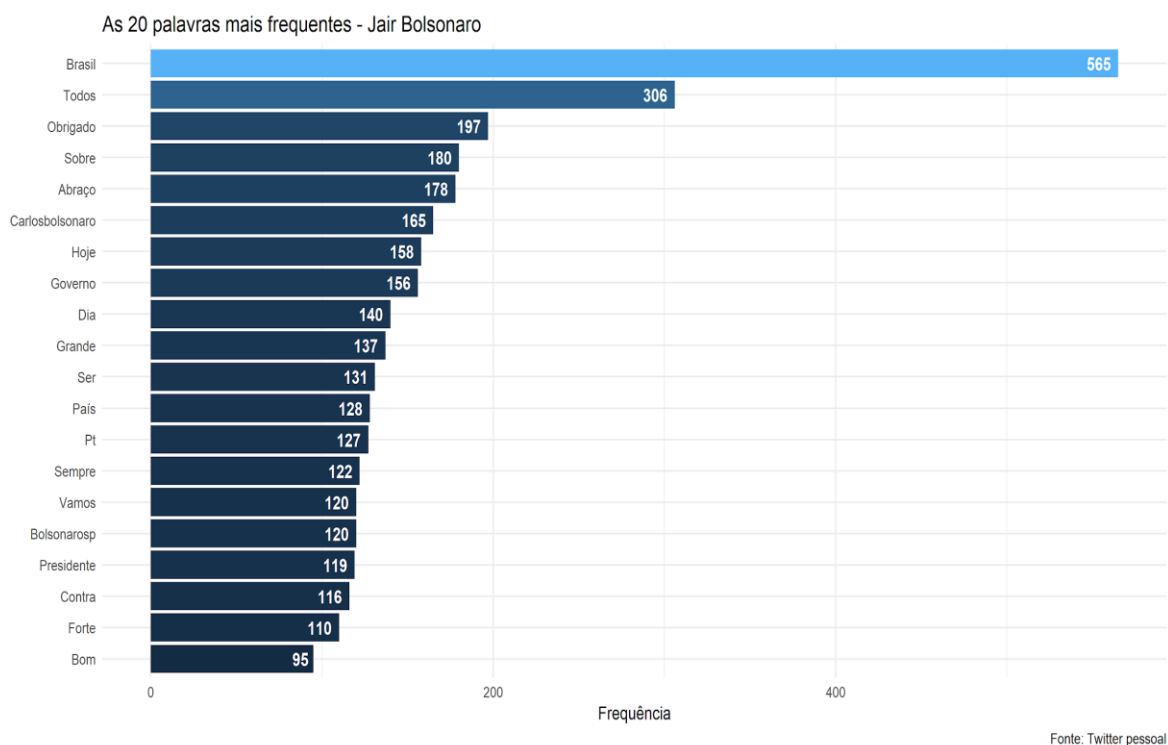
Já para a criação do ranque de sentimentos foi usado a técnica de Análise de Sentimento que foi baseada em todos os tweets coletados e a partir deles pode-se observar os tweets com maior(es) e menor(es) classificação(ões). Baseado na classificação recuperou-se o tweet para ser visualizado.



IV SEMINÁRIO INTERNACIONAL DE ESTATÍSTICA COM R & PYTHON E AS TENDÊNCIAS DE COLABORAÇÃO NITERÓI, 21 A 23 DE MAIO DE 2019



Devido ao grande número de imagens geradas com o estudo, usaremos apenas algumas das visualizações referente ao atual presidente, Jair Bolsonaro, e ao ex presidente



Lula.

Figura 1 – As 20 palavras mais frequentes - Jair Bolsonaro

Fonte: Elaborado pelos autores, 2019.

Word Cloud - Luiz Inacio Lula da Silva



Figura 2 – Nuvem de palavras - Luiz Inácio Lula da Silva

Fonte: Elaborado pelos autores, 2019.



Conclusão

A escolha da linguagem de programação Python se mostrou bastante eficiente, pois coletou todos os 16.090 tweets dos presidentes estudados em apenas alguns segundos. Já o R e suas bibliotecas facilitaram a criação das visualizações, sendo fácil de ser executado e as deixando atrativas para os leitores.

Com base nas visualizações geradas pelo estudo pode-se concluir que existe uma predominância de palavras relacionadas a patriotismo e identidade nacional.

Além disso, o atual presidente possui um grande número de citações ao seu partido rival, no caso “PT”. Já o ex-presidente Lula cita várias vezes o nome de Haddad demonstrando seu apoio durante as eleições de 2018.

Referências

BACHE, S e WICKHAM, H. **magrittr: A Forward-Pipe Operator for R**. R package version 1.5. URL <https://CRAN.R-project.org/package=magrittr>, 2014.

BENOIT.K, MUHR.D e WATANABE.K. **stopwords: Multilingual Stopword Lists**. R package version 0.9.0. URL <https://CRAN.R-project.org/package=stopwords>, 2017.

CSARDI,G e NEPUSZ,T: **The igraph software package for complex network research**, *InterJournal, Complex Systems* 1695. URL <http://igraph.org>, 2016.

GONZAGA, S. **lexiconPT: Lexicons for Portuguese Text Analysis**. R package version 0.1.0. URL <https://CRAN.R-project.org/package=lexiconPT>, 2017.

HENRY, L e WICKHAM, H. **purrr: Functional Programming Tools**. R package version 0.3.2. URL <https://CRAN.R-project.org/package=purrr>, 2019.

MCKINNEY, W. **Powerful data structures for data analysis, time series, and statistics**. Python library version 0.24.2. URL <https://pypi.org/project/pandas/>, 2019.

PEDERSEN,T. **ggraph: An Implementation of Grammar of Graphics for Graphs and Networks**. R package version 1.0.2. <https://CRAN.R-project.org/package=ggraph>, 2018.

PENNEC,E e SLOWIKOWSKI,K. **ggwordcloud: A Word Cloud Geom for 'ggplot2'**. R package version 0.3.0. URL <https://cran.r-project.org/package=ggwordcloud>, 2018.

PYTHON Software Foundation. **Python Language Reference**, version 3.7.1. URL <http://www.python.org>, 2019.

R Core Team. **R: A language and environment for statistical computing**. Version 3.5.0. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>, 2018.



ROESSLEIN, J. **Twitter library for python**. Python library version 3.7.0. URL <https://pypi.org/project/tweepy/>, 2018.

SANTOS, Bárbara Ferreira. Apesar de expansão, acesso à internet no Brasil ainda é baixo. **EXAME**, São Paulo, 29 de janeiro de 2018. Disponível em: <https://exame.abril.com.br/brasil/apesar-de-expansao-acesso-a-internet-no-brasil-ainda-e-baixo/>. Acesso em: 07/04/2019.

SILGE, Julia e ROBINSON, David. **Text Mining with R A Tidy Approach**. Estados Unidos: O'Reilly, 2017.

SILGE, J e ROBINSON, D. **Tidyttext: Text Mining and Analysis Using Tidy Data Principles in R**. R package version 0.2.0. URL <http://dx.doi.org/10.21105/joss.00037>, 2016.

WICKHAM, H. **stringr: Simple, Consistent Wrappers for Common String Operations**. R package version 1.4.0. URL <https://CRAN.R-project.org/package=stringr>, 2019.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016.

WICKHAM, H e HENRY, L. **tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions**. R package version 0.8.3. URL <https://CRAN.R-project.org/package=tidyr>, 2019.

WICKHAM, H, FRANÇOIS, R, HENRY, L e MÜLLER, KJ. **dplyr: A Grammar of Data Manipulation**. R package version 0.8.0.1. URL <https://CRAN.R-project.org/package=dplyr>, 2019.