

# Insights Twitter

Cesar Vinicius cesarvinici@gmail.com (mailto:cesarvinici@gmail.com)

24/11/2019

## Analise de mensagens no twitter no dia 08/11/2019

Vamos (tentar) analisar alguns tweets gerados no dia 08/11/2019 e ver qual insights podemos tirar de lá.

Antes que me perguntem, **porque dia 08/11/2019?**

- Bom, imaginei que usuários de do twitter estariam pavorosos pois foi o dia que o ex presidente Lula foi solto da prisão após a decisao sobre prisões em segunda instância do STF.

Isso dito acredito que um *disclaimer* seja necessário. Bem sabemos que política é um assunto um tanto perigoso para se tratar no Brasil nos últimos tempos, pois de modo geral as pessoas estão cada vez mais polarizadas. Com isso minha intenção aqui é ser o mais imparcial possível.

Três pesquisas foram realizadas no twitter neste dia, usei as *keywords* **Lula** e **Bolsonaro** e uma terceira pesquisa sem uma *keyword* definida, de forma a pegar tudo possível.

Como uso uma versão gratuita da API do twitter, a pesquisa foi um tanto limitada, consegui apenas 4/5 mil tweets em cada pesquisa. Mas acredito que já de pra tirar uns *insights* interessantes!

Para realizar este trabalho foi usado Python para coleta e salvamento dos dados em um container docker rodando um MongoDB e R para as análises.

Todo código está disponível no Github (<https://github.com/cesarvinici/analise-twitter>)

## Lula

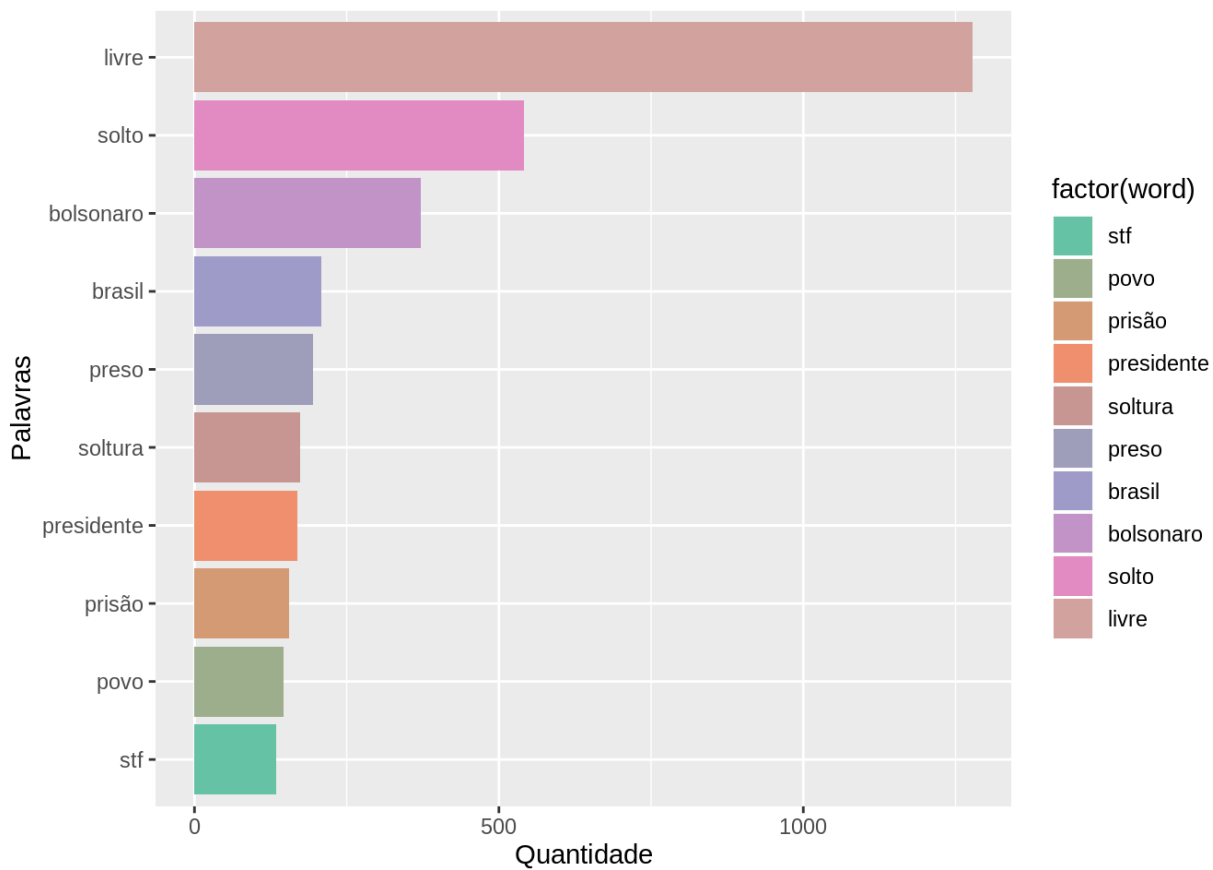
Vamos primeiro para análise utilizando a *keyword* 'Lula':

- Abaixo uma *WordCloud* com as palavras que mais aparecem na pesquisa:



Algumas palavras já imaginávamos que iriam aparecer, como “Livre” de (Lula Livre) e afins. Percebemos que a palavra ‘Bolsonaro’ também aparece com uma boa frequência. É possível verificar que algumas pessoas não ficaram muito feliz com a decisão pois as palavras “ladrão” e “bandido” aparecem. Algumas provocações foram feitas para eleitores do Bolsonaro, já que a palavra comumente usada “Bolsominion” também aparece com certa frequência.

Para confirmar algumas previsões, vamos ver um gráfico de barras mostrando a contagem das palavras:

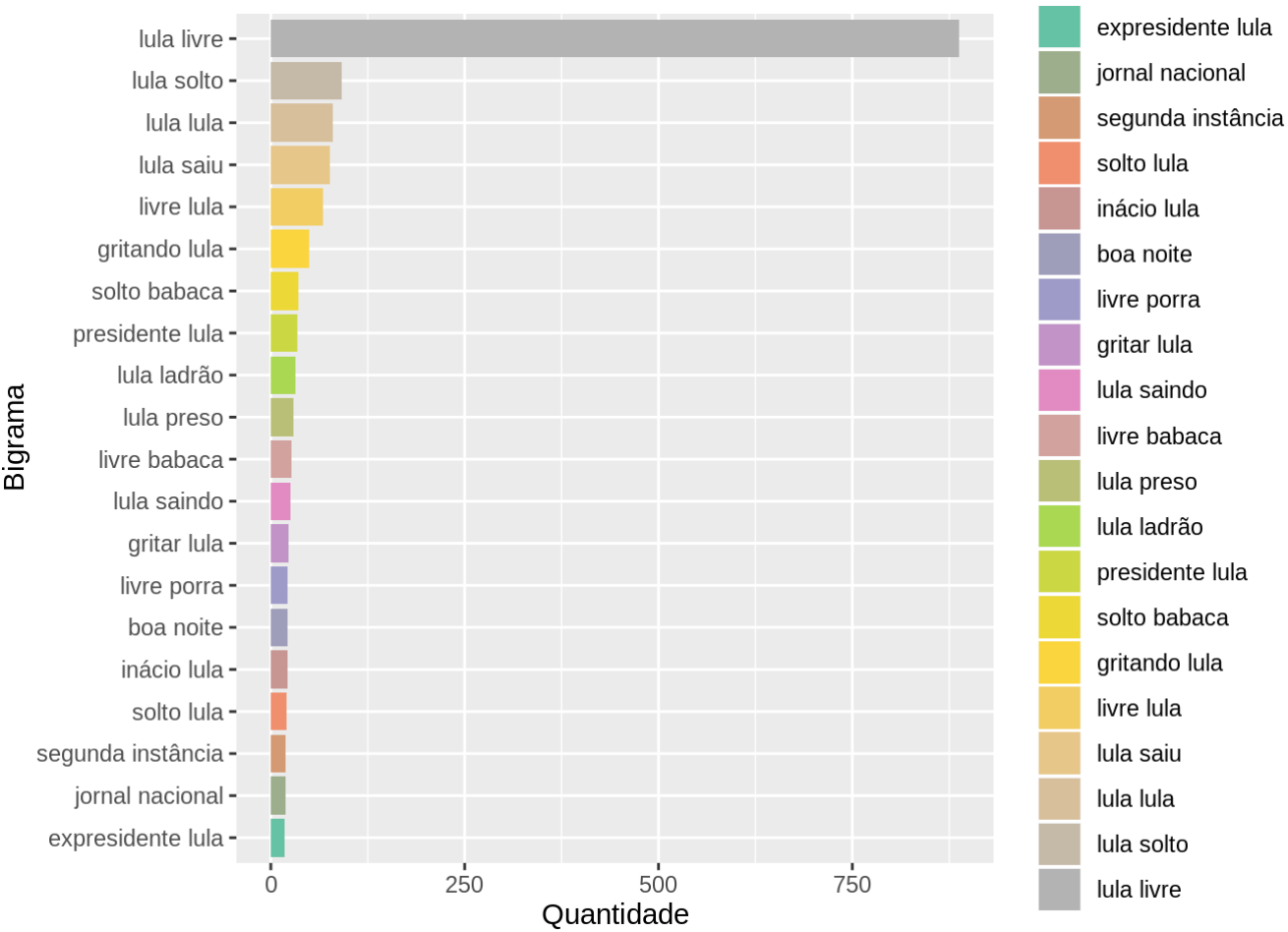


### WordCloud Keyword Lula

Nenhuma surpresa quando as palavras “livre” (Lula livre) ou solto, Bolsonaro está no top 3 o que pode ser surpreendente e ao mesmo tempo levar à indagação se são eleitores do Bolsonaro ou eleitores do Lula que estão preocupados com o candidato do outro lado do polo?

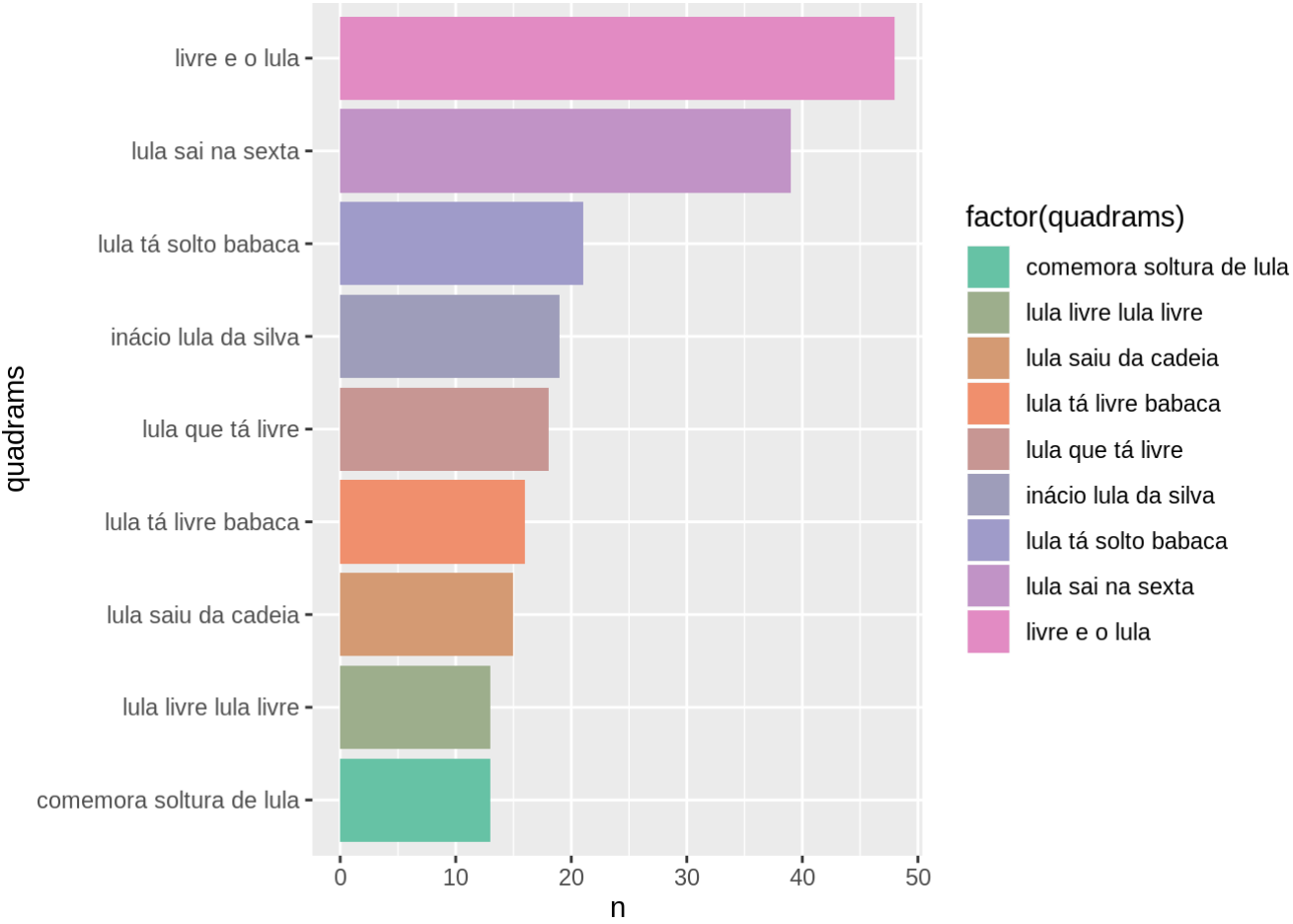
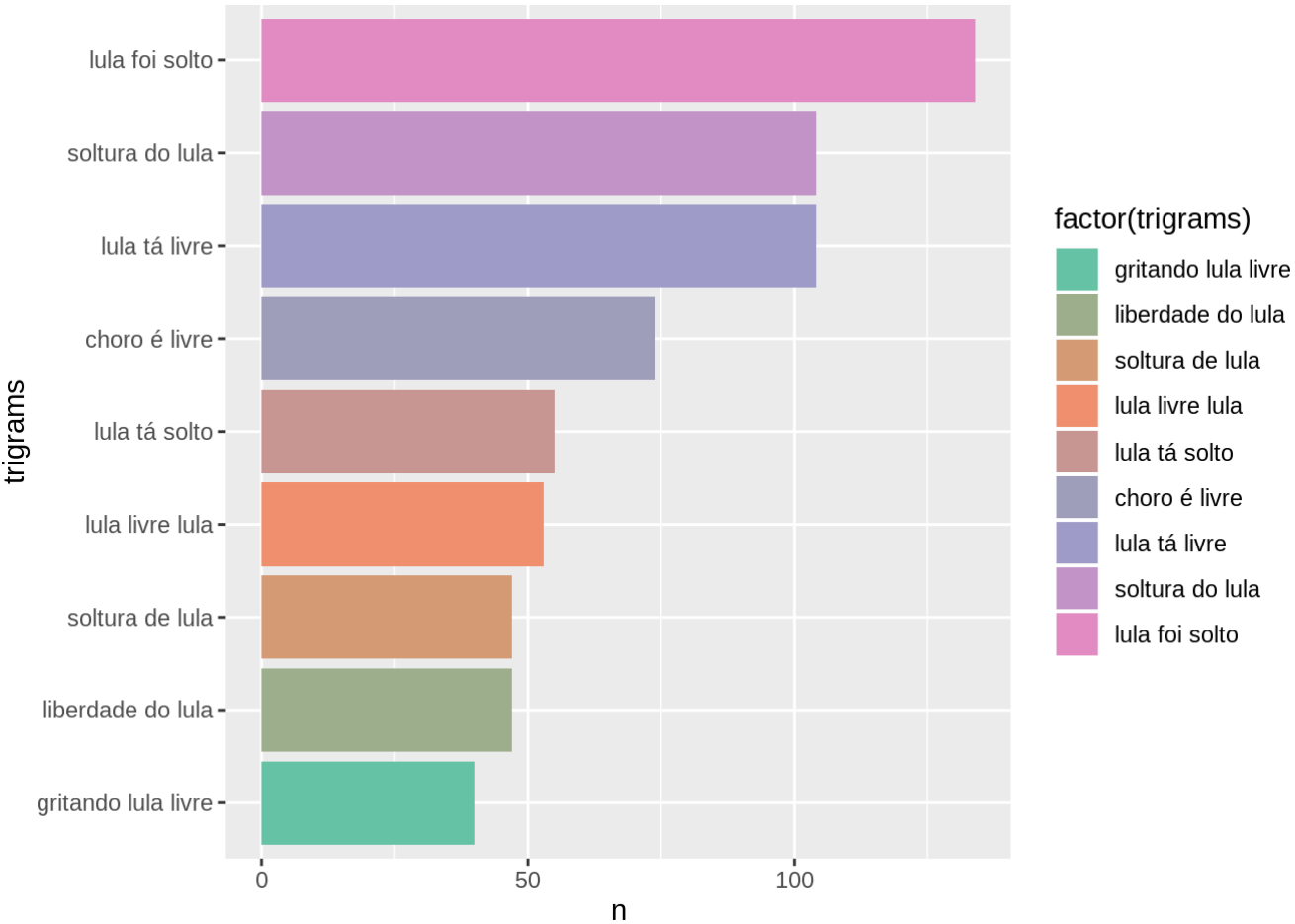
Beleza, já sabemos quais foram as palavras que mais se repetiram, mas e as frases? Bigramas, Trigramas e etc? é possível saber? **yes it is!**

Vamos aos bigramas (se você não sabe o que é um bigrama Clique aqui (<https://en.wikipedia.org/wiki/Bigram>))



Bom, pra surpresa de ninguém o bigrama mais falado foi “Lula livre”, e de modo geral com excessão de Boa noite e jornal nacional todas as outras frases foram em um sentido de comemoração à solturado Ex Presidente.

E frases com 3 ou mais palavras?

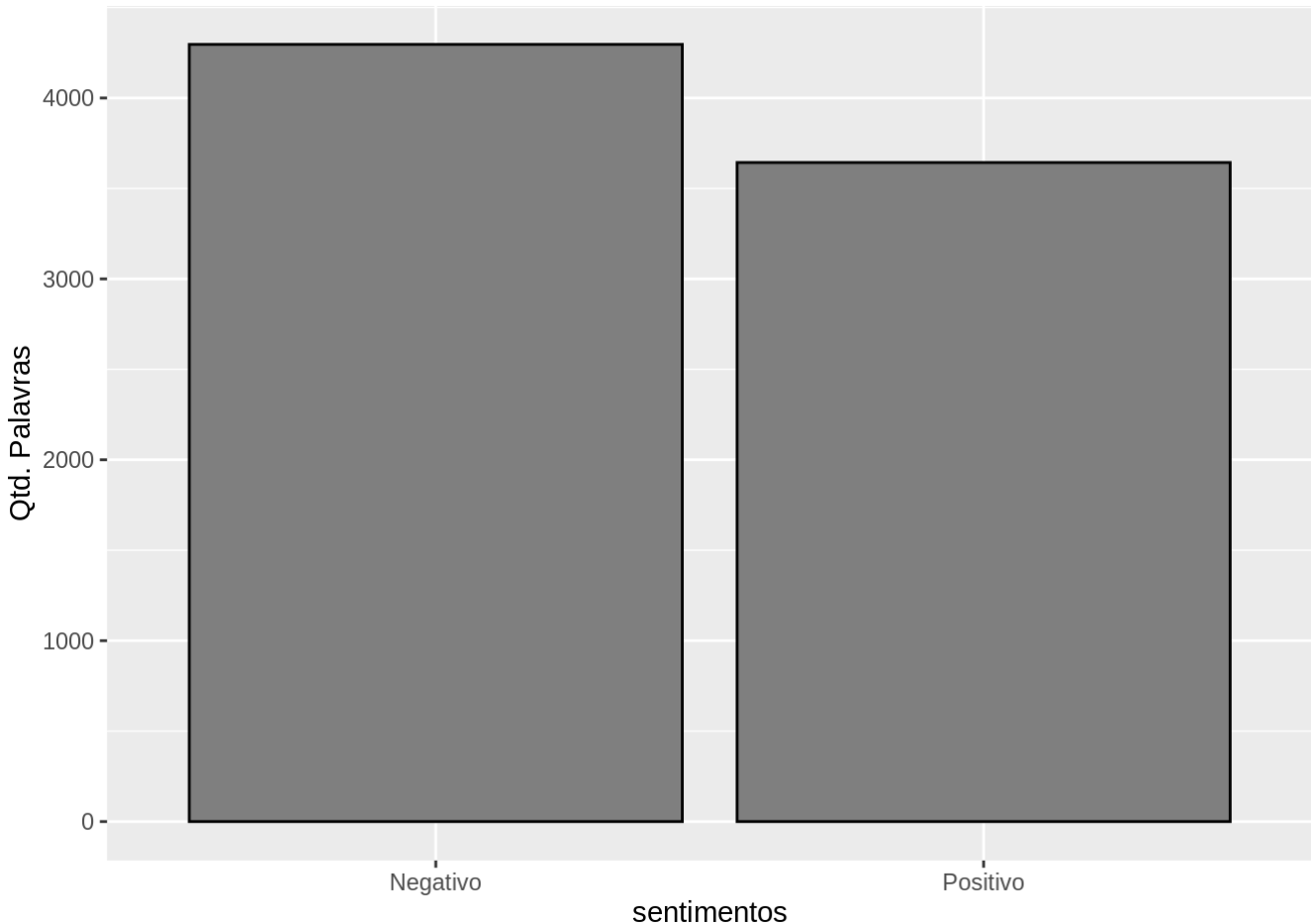


Via de regra são pessoas ou comemorando ou, ao que parece simplesmente comentando a soltura.

Bom, se formos analisar friamente não obtivemos nenhuma grande surpresa nas palavras nem nas frases ditas pelos usuários do Twitter ao que se refere ao Lula, mas e em uma análise de sentimentos, será que eles se portaram de forma mais agressiva ou mais positiva?

Vamos analisar!

```
## Joining, by = "word"
```

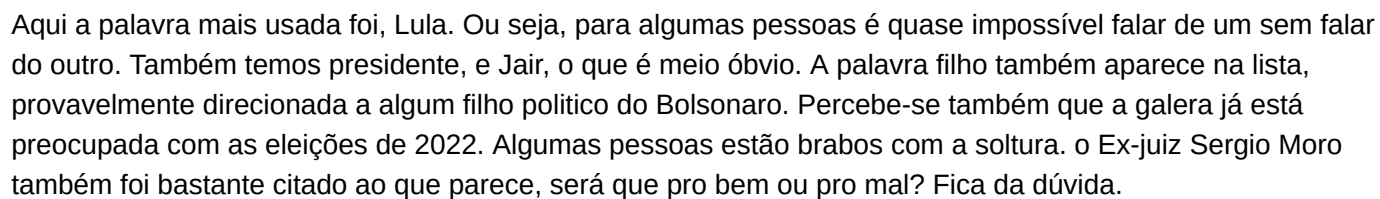


Na comparação palavras com um tom mais negativos foram mais “faladas” que as positivas, porém precisamos levar em consideração o contexto, já que “palavrões” foram usados em um tom de positivo, de comemoração.

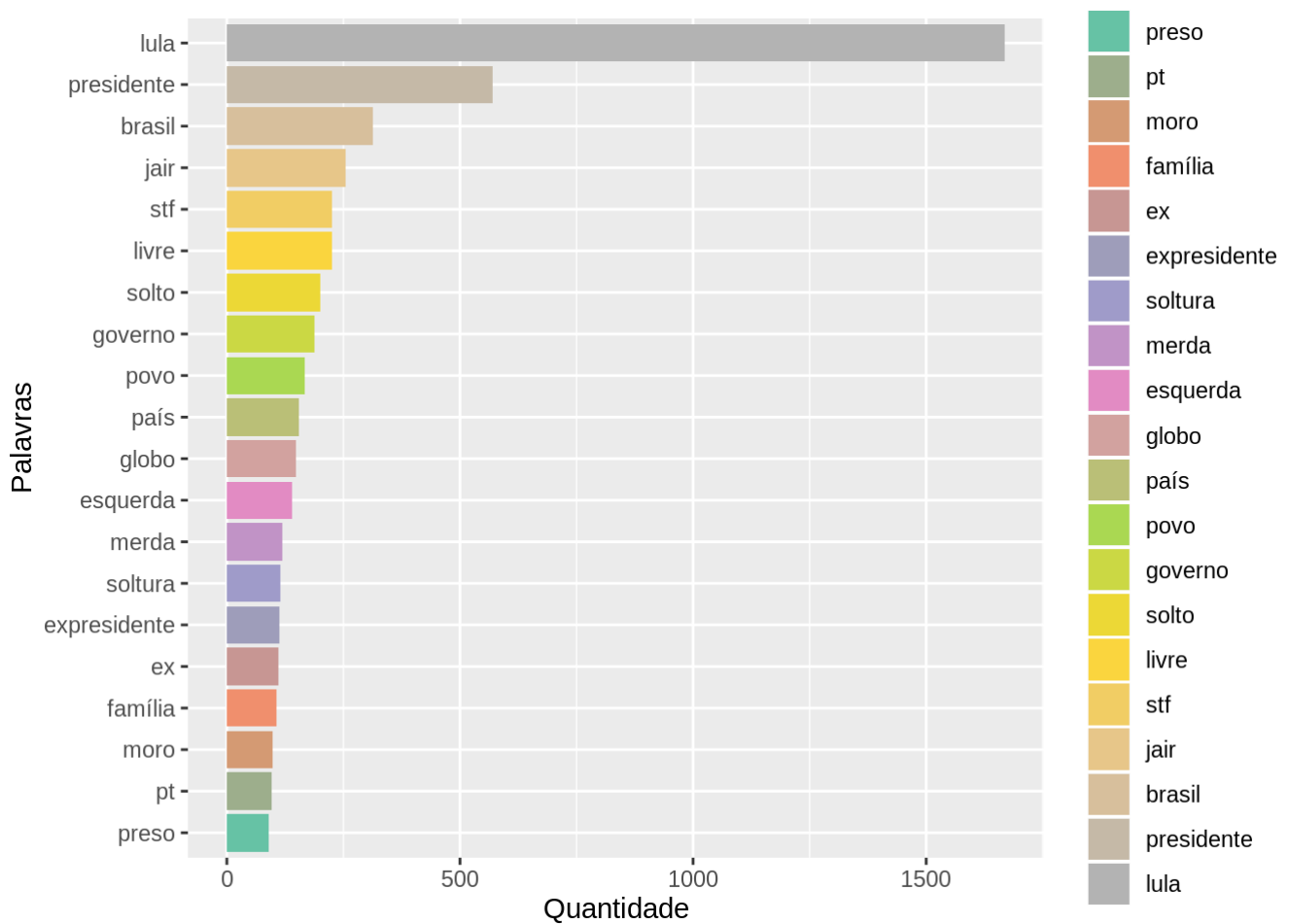
## Bolsonaro

Vamos para a pesquisa utilizando a palavra-chave “Bolsonaro”

ps: Da mesma forma que tiramos a palavra-chave Lula, aqui também iremos tirar Bolsonaro, já que, ela aparece em todos os twites.



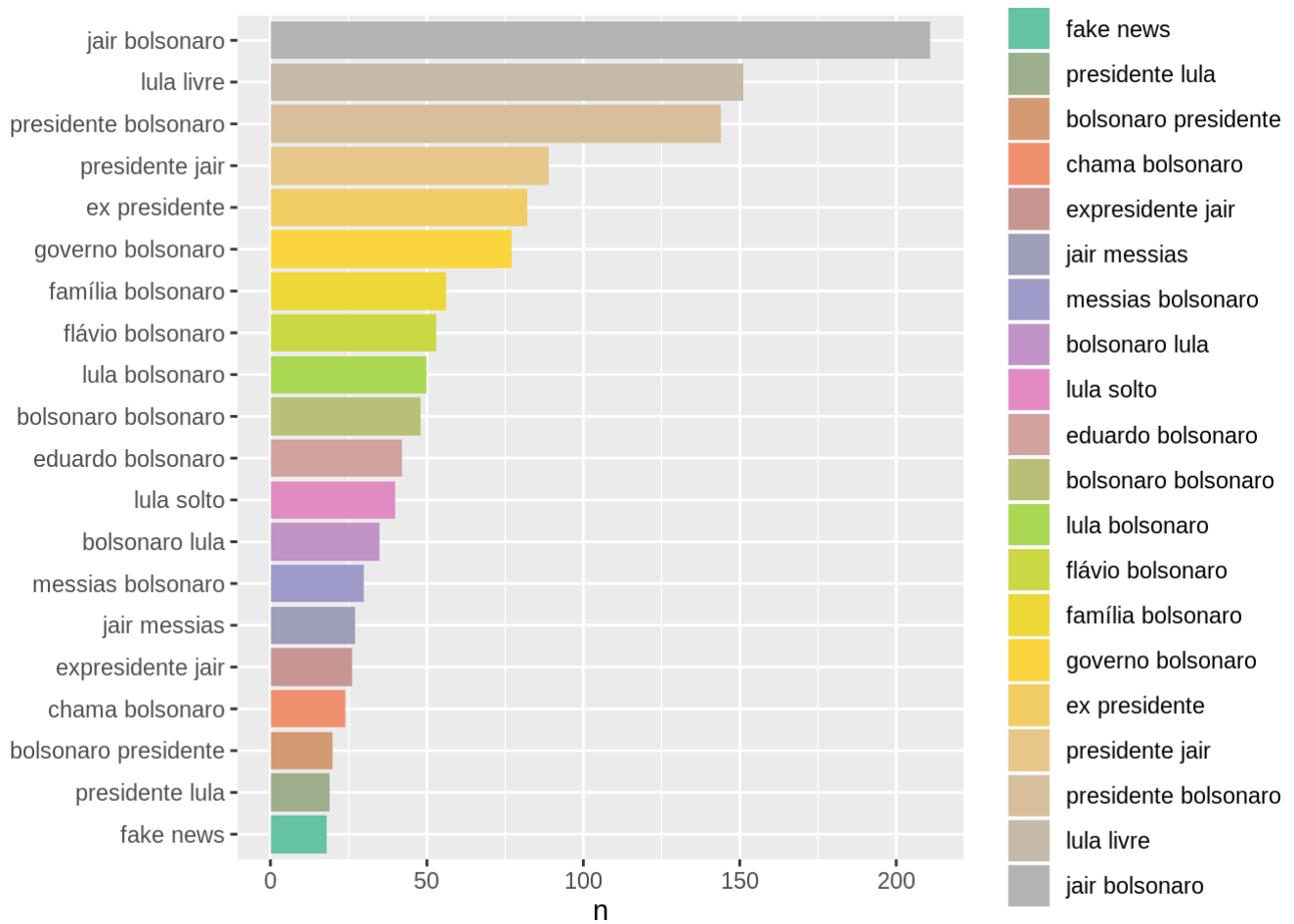
Bom, podemos comprovar algumas teorias olhando a contagem de palavras, vamos lá!



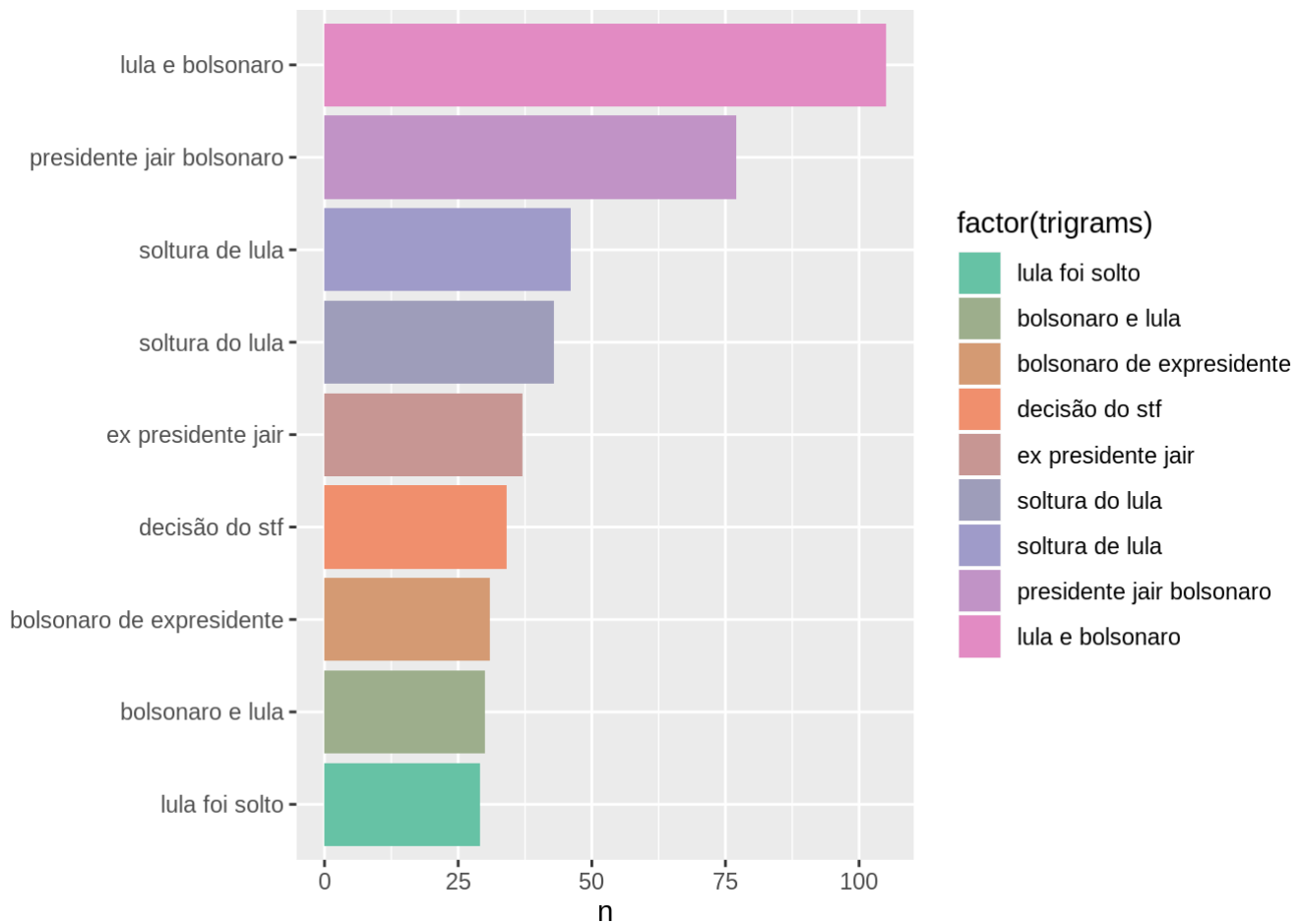
Por mais que olhando para o gráfico, existem muitas referências ao PT ou ao Lula. Foram recuperados 3900 twites com a palavra-chave bolsonaro, e em mais de 1500 também aparece a palavra Lula. Lembrando que quando fizemos utilizando a palavra-chave Lula, o qual retornou mais de 4 mil twites, Bolsonaro não chegou a ser citado 500 vezes.

Bigramas, Trigramas e Quadrigramas.

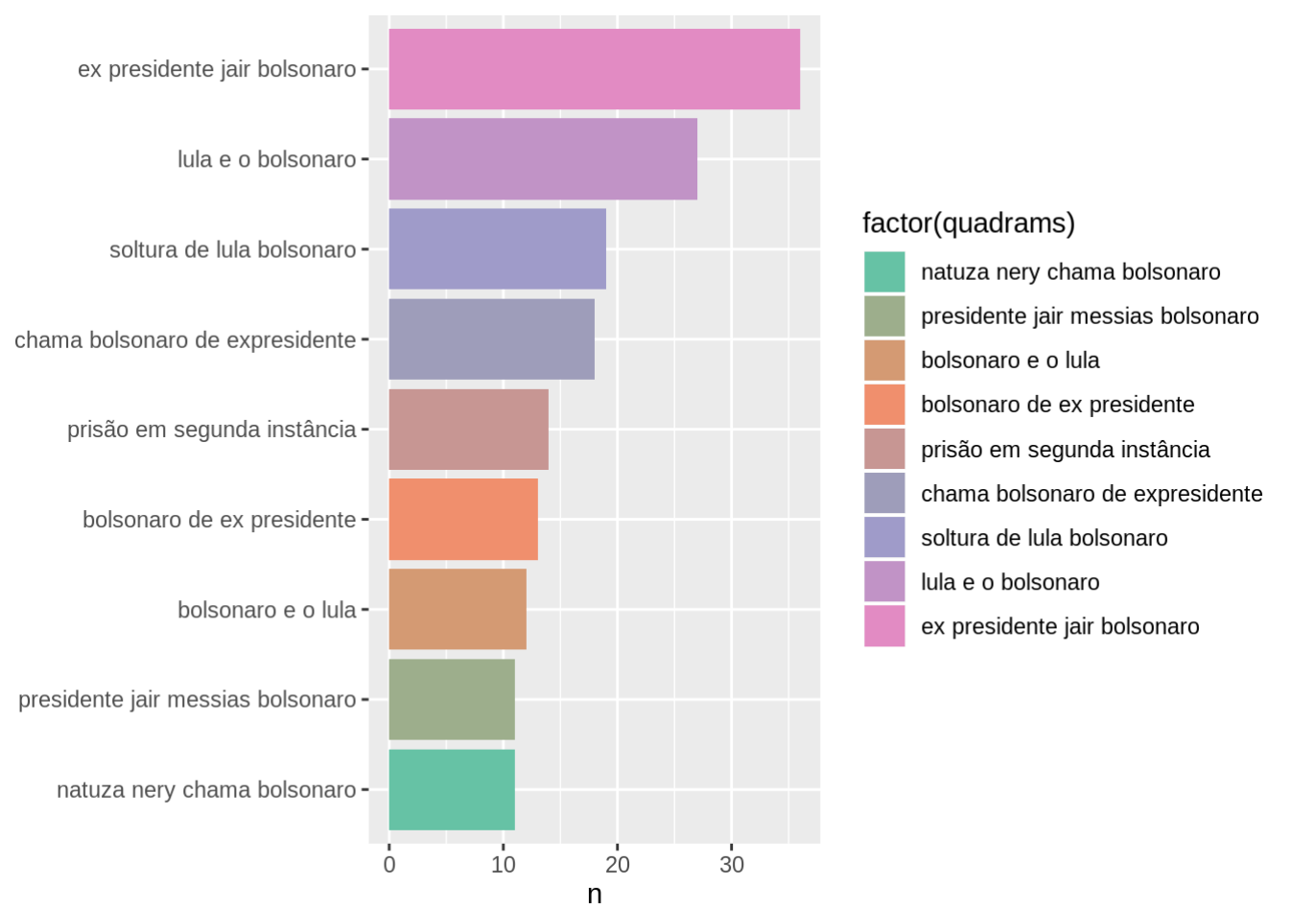




Bigramas basicamente com citações à família Bolsonaro. Interessante ver que mesmo usando a palavra-chave Bolsonaro, o segundo bigrama que mais se repetiu foi “Lula Livre”.

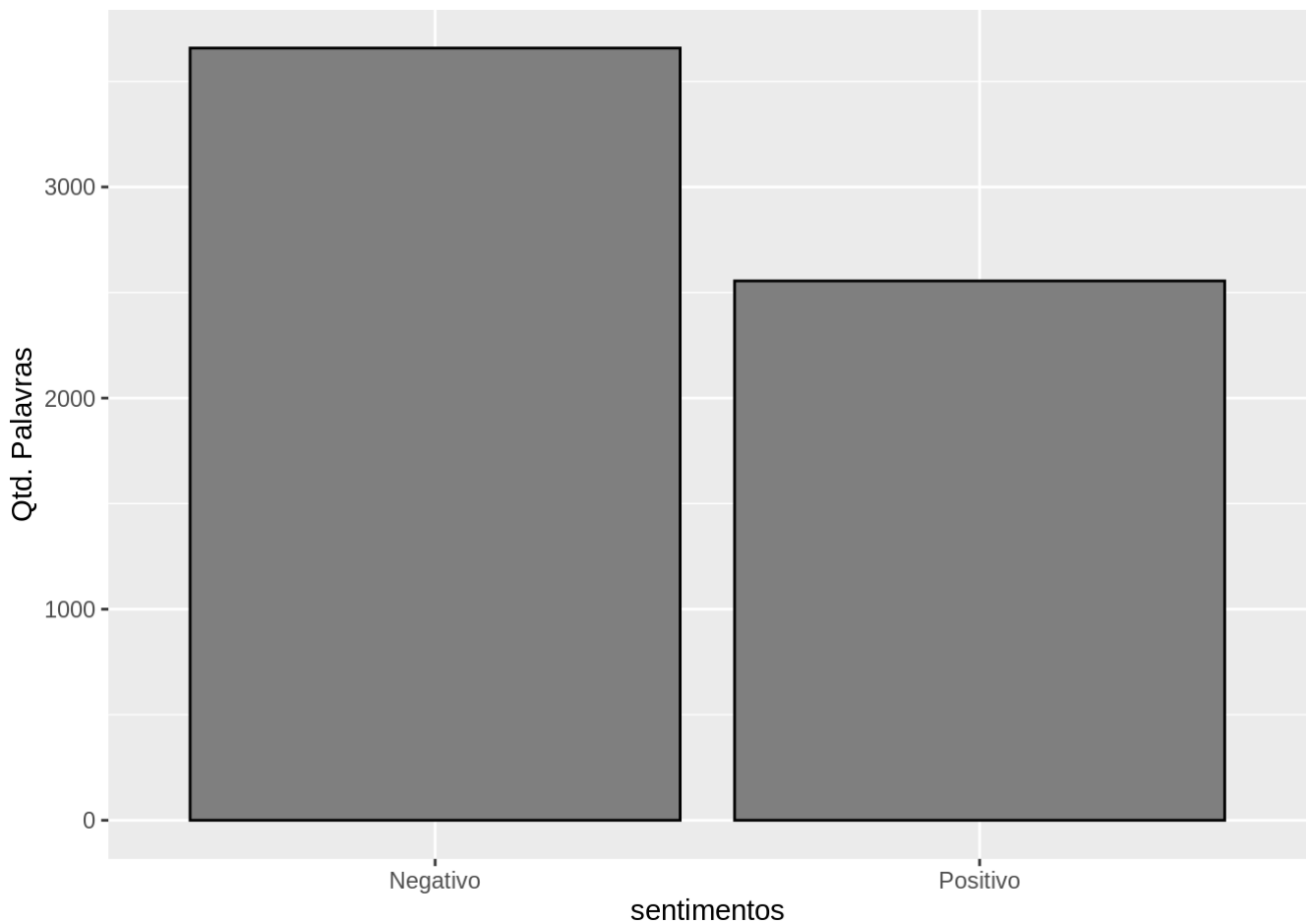


Nos trigramas, mais algumas citações e alguns comentários sobre a decisão do STF



Nos quadrigramas podemos verificar alguns comentários de sobre a reporter que chamou o atual presidente de ex presidente.

```
## Joining, by = "word"
```

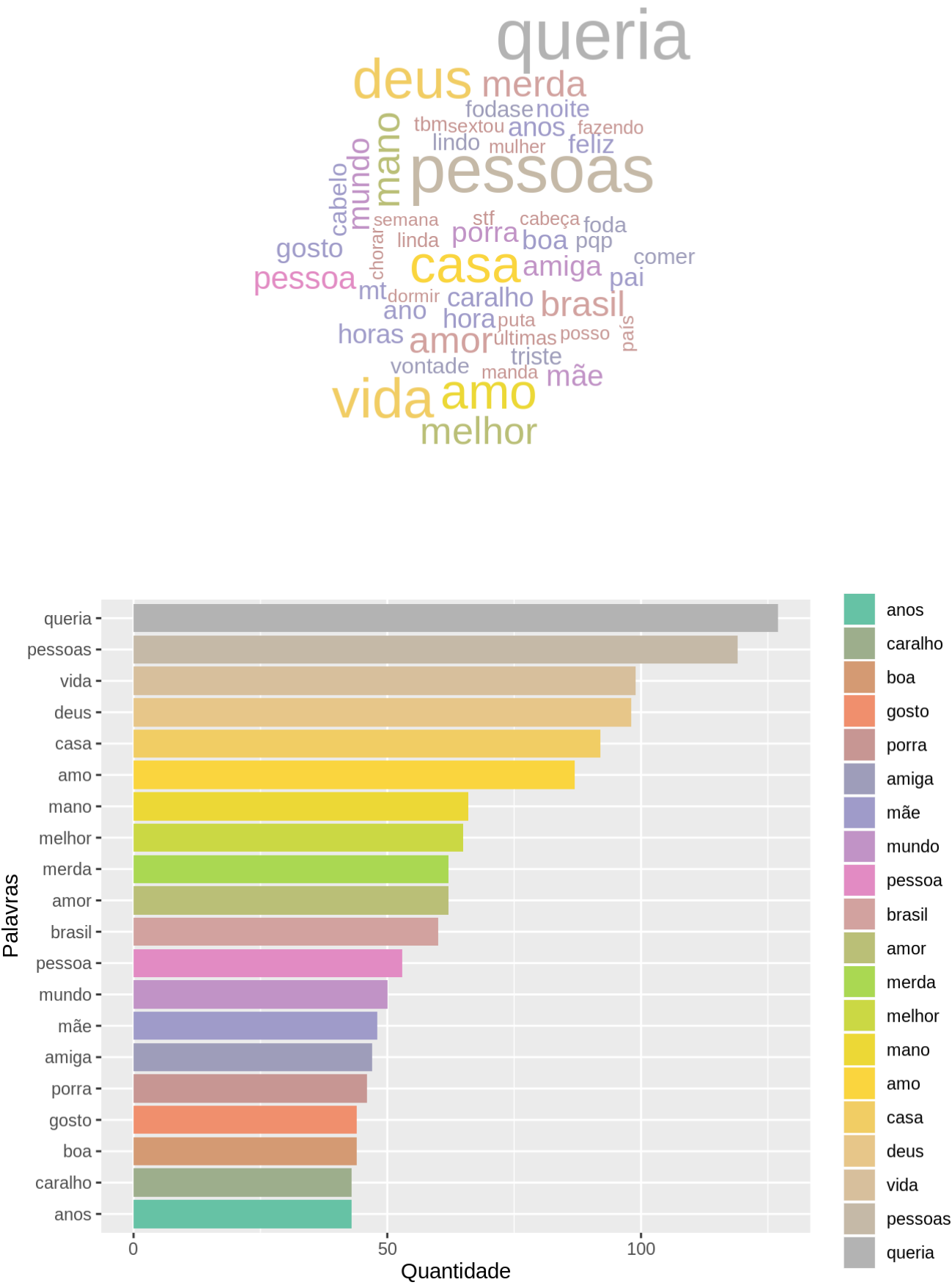


Pesquisa realizada com a palavra-chave bolsonaro retornou dados muito mais negativos, o que acredito já era de se esperar pois já que várias pessoas eram contra a decisão.

## Twites Gerais

E se realizássemos uma pesquisa sem usar nenhuma palavra-chave? Será que a maioria dos usuários estão pensando em política ou não?

Vamos descobrir!



Bom, via de regra ao que parece as pessoas não estavam preocupadas com o ocorrido a não ser uma citação ao STF no worldcloud.

Bom, acredito que nenhuma surpresa nos dados informados, algumas pessoas contentes e outras nem tanto. Muito se falou da Família Bolsonaro, muito Lula Livre.

Se pegarmos uma amostra de dados sem usar uma palavra-chave podemos percebermos que as pessoas falam tanto assim sobre política.

Bom, espero que os dados acima possam trazer algum tipo de conhecimento. Qualquer dúvida estou a disposição.