

# A UTILIZAÇÃO DE REDES SOCIAIS DA INTERNET PARA OBTENÇÃO DE DADOS.

---

Anderson Castro Soares de Oliveira

VI Workshop de Métodos Probabilísticos e Estatísticos

---

**Twitter**

---

# TWITTER

---

- Twitter é uma rede social e servidor para microblogging, que permite aos usuários enviar e receber atualizações pessoais de outros contatos, em textos de até 140 caracteres.



## Estrutura

	perfis
Privacidade	Público Privado
Conexão	Seguidores
Interação	Tweet Retweet Mensagem privada (DM)
Relação	Interesse de conteúdo
Conteúdo	Texto (140 caracteres) Link Foto Video GIF #hashtags

- O Twitter tem em torno de 500 milhões de usuários

# TWITTER

---

- O Twitter tem em torno de 500 milhões de usuários
- No Brasil tem em torno de 41 milhões de usuários

# TWITTER

---

- O Twitter tem em torno de 500 milhões de usuários
- No Brasil tem em torno de 41 milhões de usuários
- Embora existam muitas plataformas de redes sociais, a grande maioria das pesquisas são realizadas a partir do Twitter



# TWITTER

---

- O Twitter tem em torno de 500 milhões de usuários
- No Brasil tem em torno de 41 milhões de usuários
- Embora existam muitas plataformas de redes sociais, a grande maioria das pesquisas são realizadas a partir do Twitter
- Esta preponderância dos estudos do Twitter deve-se principalmente à disponibilidade de dados, ferramentas e facilidade de análise

## Monitoramento no Twitter

- Para realizar o monitoramento do Twitter é necessário:

## Monitoramento no Twitter

- Para realizar o monitoramento do Twitter é necessário:
  - Ter um software ou plataforma para monitoramento

## Monitoramento no Twitter

- Para realizar o monitoramento do Twitter é necessário:
  - Ter um software ou plataforma para monitoramento
  - Ter um aplicativo dentro do Twitter para comunicar com o software ou plataforma.

## Criando um App no Twitter

- Acessar sua conta no Twitter;

## Criando um App no Twitter

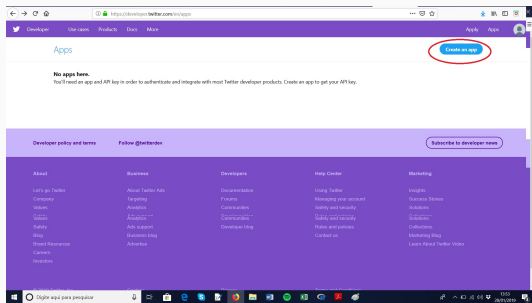
- Acessar sua conta no Twitter;
  - Desde 2018 é necessário ter uma conta de desenvolvedor.

## Criando um App no Twitter

- Acessar sua conta no Twitter;
  - Desde 2018 é necessário ter uma conta de desenvolvedor.
- Acessar <https://apps.twitter.com/>

## Criando um App no Twitter

- Acessar sua conta no Twitter;
  - Desde 2018 é necessário ter uma conta de desenvolvedor.
- Acessar <https://apps.twitter.com/>
  - Clique no botão "Create an app".





## Criando um App no Twitter

- Completar os campos

## Criando um App no Twitter

- Completar os campos
  - App name - Nome para o aplicativo, um que ninguém tenha utilizado Twitter).

## Criando um App no Twitter

- Completar os campos
  - App name - Nome para o aplicativo, um que ninguém tenha utilizado Twitter).
  - Application description - Breve descrição da ferramenta, exemplo: "Ferramenta de captura de dados do Twitter".

## Criando um App no Twitter

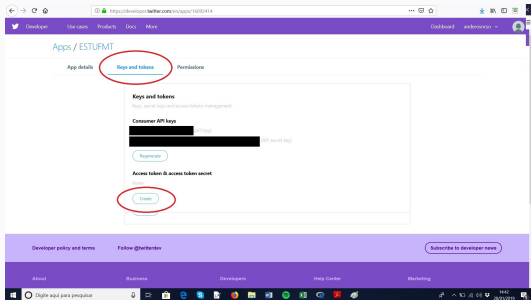
- Completar os campos
  - App name - Nome para o aplicativo, um que ninguém tenha utilizado Twitter).
  - Application description - Breve descrição da ferramenta, exemplo: "Ferramenta de captura de dados do Twitter".
  - Website URL - pagina publica do aplicativo  
<https://twitter.com/username>

## Criando um App no Twitter

- Completar os campos
  - App name - Nome para o aplicativo, um que ninguém tenha utilizado Twitter).
  - Application description - Breve descrição da ferramenta, exemplo: "Ferramenta de captura de dados do Twitter".
  - Website URL - página pública do aplicativo  
<https://twitter.com/username>
  - Tell us how this app will be used - descrição da finalidade do aplicativo

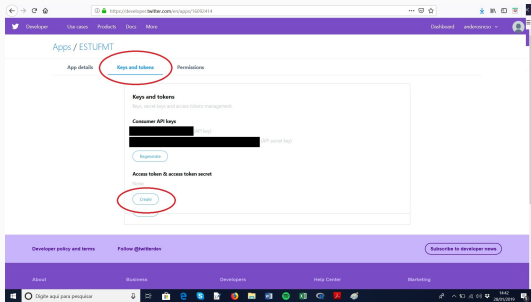
## Criando um App no Twitter

- Após criar o App deve-se criar o chave de acesso



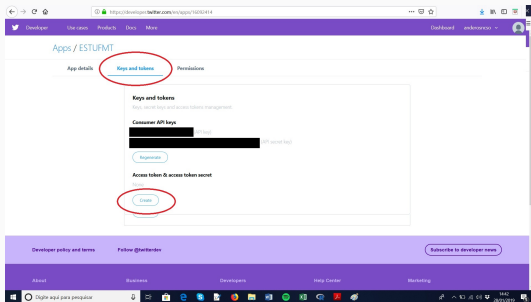
## Criando um App no Twitter

- Após criar o App deve-se criar o chave de acesso



## Criando um App no Twitter

- Após criar o App deve-se criar o chave de acesso

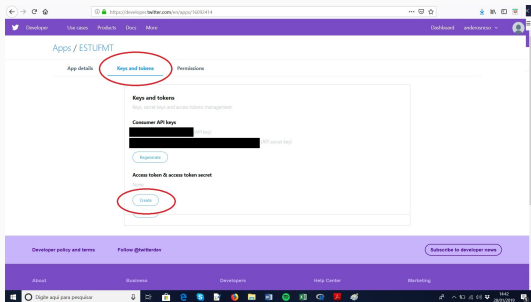


- Obter os codigos:



## Criando um App no Twitter

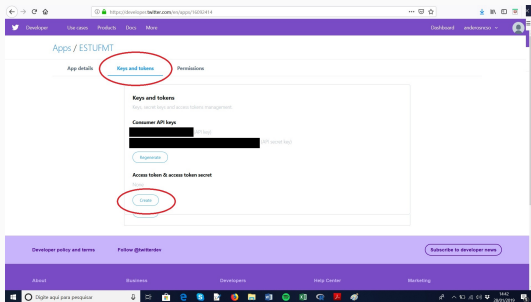
- Após criar o App deve-se criar o chave de acesso



- Obter os códigos:
  - API key

## Criando um App no Twitter

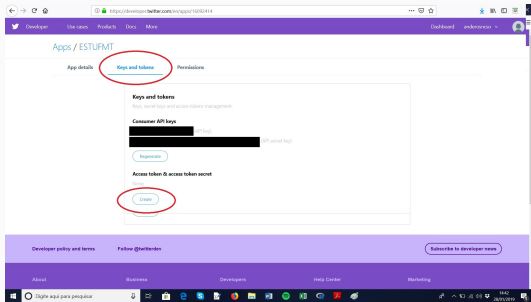
- Após criar o App deve-se criar o chave de acesso



- Obter os códigos:
  - API key
  - API secret key

## Criando um App no Twitter

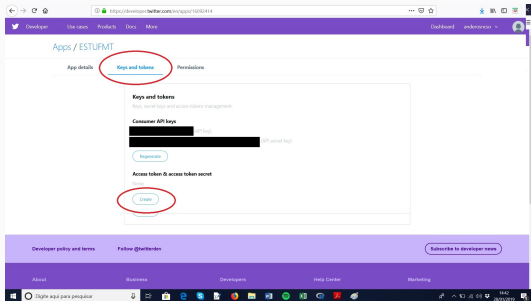
- Após criar o App deve-se criar o chave de acesso



- Obter os códigos:
  - API key
  - API secret key
  - Access token

## Criando um App no Twitter

- Após criar o App deve-se criar o chave de acesso



- Obter os códigos:
  - API key
  - API secret key
  - Access token
  - Access token secret

## Utilizando o R

- Monitoramento do Twitter utilizando o R

## Utilizando o R

- Monitoramento do Twitter utilizando o R
- O pacote `twitteR` destina-se a fornecer acesso à um API do Twitter dentro do R, permitindo aos usuários capturar subconjuntos de dados do Twitter para suas análises.

## Utilizando o twitterR

- Estabelecendo a comunicação entre o R e witter

```
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

## Utilizando o twitterR

- Estabelecendo a comunicação entre o R e witter

```
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

- `consumer_key` - API key do aplicativo criado no Twitter



## Utilizando o twitterR

- Estabelecendo a comunicação entre o R e witter

```
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

- `consumer_key` - API key do aplicativo criado no Twitter
- `consumer_secret` - API secret key do aplicativo criado no Twitter

## Utilizando o twitterR

- Estabelecendo a comunicação entre o R e witter

```
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

- `consumer_key` - API key do aplicativo criado no Twitter
- `consumer_secret` - API secret key do aplicativo criado no Twitter
- `access_token` - Access token do aplicativo criado no Twitter

## Utilizando o twitterR

- Estabelecendo a comunicação entre o R e witter

```
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

- `consumer_key` - API key do aplicativo criado no Twitter
- `consumer_secret` - API secret key do aplicativo criado no Twitter
- `access_token` - Access token do aplicativo criado no Twitter
- `access_secret` - Access token secret do aplicativo criado no Twitter

## Utilizando o twitterR

- Busca por palavras chave

```
searchTwitter(searchString)
```

## Utilizando o twitterR

- Busca por palavras chave

```
searchTwitter(searchString)
```

- `searchStrin` - Palavra ou palavras

```
rt=searchTwitter("IBGE")
```

```
rt=searchTwitter("Censo")
```

```
rt=searchTwitter("IBGE + Censo")
```

## Utilizando o twitterR

- Busca por palavras chave

```
searchTwitter(searchString)
```

- searchString - Palavra ou palavras

```
rt=searchTwitter("IBGE")
```

```
rt=searchTwitter("Censo")
```

```
rt=searchTwitter("IBGE + Censo")
```

- A função `searchTwitter` retorna 25 tweets, para obter mais tweets deve-se especificar o argumento `n`

```
rt=searchTwitter("IBGE + Censo",n=100)
```

## Utilizando o twitterR

- Idioma

```
rt=searchTwitter(searchTwitter,lang)
```

## Utilizando o twitterR

- Idioma

```
rt=searchTwitter(searchTwitter,lang)
```

- `lang` - código do idioma dado pela ISO 639-1.

```
rt=searchTwitter("IBGE + Censo",lang="pt")
```



## Utilizando o twitterR

- Idioma

```
rt=searchTwitter(searchTwitter,lang)
```

- lang - código do idioma dado pela ISO 639-1.

```
rt=searchTwitter("IBGE + Censo",lang="pt")
```

- Data

```
rt=searchTwitter(searchTwitter,since,until)
```

## Utilizando o twitterR

- Idioma

```
rt=searchTwitter(searchTwitter,lang)
```

- lang - código do idioma dado pela ISO 639-1.

```
rt=searchTwitter("IBGE + Censo",lang="pt")
```

- Data

```
rt=searchTwitter(searchTwitter,since,until)
```

- since - data inicial no formato YYYY-MM-DD

## Utilizando o twitterR

- Idioma

```
rt=searchTwitter(searchTwitter, lang)
```

- `lang` - código do idioma dado pela ISO 639-1.

```
rt=searchTwitter("IBGE + Censo", lang="pt")
```

- Data

```
rt=searchTwitter(searchTwitter, since, until)
```

- `since` - data inicial no formato YYYY-MM-DD

- `until` - data final no formato YYYY-MM-DD

```
rt=searchTwitter("IBGE", since="2017-01-01")
```

```
rt=searchTwitter("IBGE", until="2018-12-31")
```

```
rt=searchTwitter("IBGE", since="2017-01-01", until="2018-12-31")
```

## Utilizando o twitterR

- Localização

```
rt=searchTwitter(searchTwitter, geocode)
```

## Utilizando o twitterR

- Localização

```
rt=searchTwitter(searchTwitter, geocode)
```

- geocod - latitude, longitude e raio

## Utilizando o twitterR

- Localização

```
rt=searchTwitter(searchTwitter, geocode)
```

- geocod - latitude, longitude e raio
- Raio pode ter como unidade mi (milhas) ou km (quilômetros)

```
rt=searchTwitter("IBGE", geocode="-22.9196591,-43.3326692,100km")
```

## Utilizando o twitterR

- Obter os assuntos de maior relevância no Twitter  
`getTrends(woeid)`

## Utilizando o twitterR

- Obter os assuntos de maior relevância no Twitter

`getTrends(woeid)`

- `woeid` - O parâmetro WOEID (where on earth id) é o código que identifica um país, cidade ou local



## Utilizando o twitterR

- Obter os assuntos de maior relevância no Twitter

`getTrends(woeid)`

- `woeid` - O parâmetro WOEID (where on earth id) é o código que identifica um país, cidade ou local
- O WOEID do Brasil é 23424768

`trends=getTrends(woeid=23424768)`

## Utilizando o twitterR

- Obter um usuário específico  
`userTimeline(user)`

## Utilizando o twitterR

- Obter um usuário específico

```
userTimeline(user)
```

- `user` - nome de usuário

```
ut=userTimeline("galvaobueno")
```

## Utilizando o twitterR

- Obter um usuário específico

```
userTimeline (user)
```

- `user` - nome de usuário

```
ut=userTimeline ("galvaobueno")
```

- Especificar o numero de tweets

```
ut=userTimeline ("galvaobueno", n=100)
```

## Utilizando o twitterR

- Obter um usuário específico

```
userTimeline(user)
```

- `user` - nome de usuário

```
ut=userTimeline("galvaobueno")
```

- Especificar o numero de tweets

```
ut=userTimeline("galvaobueno",n=100)
```

- Incluir retweets

```
ut=userTimeline("galvaobueno",includeRts=TRUE)
```

## Utilizando o twitterR

- Os dados obtidos do Twitter são em forma de uma lista não estruturada

## Utilizando o twitterR

- Os dados obtidos do Twitter são em forma de uma lista não estruturada
- Para converter esta lista em banco de dados  
`twListToDF (twList)`

## Utilizando o twitterR

- Os dados obtidos do Twitter são em forma de uma lista não estruturada
- Para converter esta lista em banco de dados

```
twListToDF (twList)
```

- `twList` - nome da lista
- ```
tweets=twListToDF (ut)
```



## Utilizando o twitterR

- O banco de dados conterà as seguintes informações

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `text` - texto do tweet

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `text` - texto do tweet
  - `favorited` - Se esse tweet foi favorecido

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `text` - texto do tweet
  - `favorited` - Se esse tweet foi favorecido
  - `favoriteCount` - numero de vezes que foi favorecido

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `text` - texto do tweet
  - `favorited` - Se esse tweet foi favorecido
  - `favoriteCount` - numero de vezes que foi favorecido
  - `replyToSN` - Nome do usuário ao qual está respondendo

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `text` - texto do tweet
  - `favorited` - Se esse tweet foi favorecido
  - `favoriteCount` - numero de vezes que foi favorecido
  - `replyToSN` - Nome do usuário ao qual está respondendo
  - `created` - Quando o tweet foi criado

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `text` - texto do tweet
  - `favorited` - Se esse tweet foi favorecido
  - `favoriteCount` - numero de vezes que foi favorecido
  - `replyToSN` - Nome do usuário ao qual está respondendo
  - `created` - Quando o tweet foi criado
  - `truncated` - Se esse tweet foi truncado

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `text` - texto do tweet
  - `favorited` - Se esse tweet foi favorecido
  - `favoriteCount` - numero de vezes que foi favorecido
  - `replyToSN` - Nome do usuário ao qual está respondendo
  - `created` - Quando o tweet foi criado
  - `truncated` - Se esse tweet foi truncado
  - `replyToSID` - identificação da respota



## Utilizando o twitterR

- O banco de dados conterà as seguintes informações

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `id` - identificação do tweet

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `id` - identificação do tweet
  - `replyToUID` - identificação a quem está respondendo

## Utilizando o twitterR

- O banco de dados conterá as seguintes informações
  - `id` - identificação do tweet
  - `replyToUID` - identificação a quem está respondendo
  - `statusSource` - Origem tweet - smartponhe, computador.

## Utilizando o twitterR

- O banco de dados conterá as seguintes informações
  - `id` - identificação do tweet
  - `replyToUID` - identificação a quem está respondendo
  - `statusSource` - Origem tweet - smartponhe, computador.
  - `screenName` - nome do usuário

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `id` - identificação do tweet
  - `replyToUID` - identificação a quem está respondendo
  - `statusSource` - Origem tweet - smartponhe, computador.
  - `screenName` - nome do usuário
  - `retweetCount` - numero de vezes que o tweet foi retweetado

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `id` - identificação do tweet
  - `replyToUID` - identificação a quem está respondendo
  - `statusSource` - Origem tweet - smartponhe, computador.
  - `screenName` - nome do usuário
  - `retweetCount` - numero de vezes que o tweet foi retweetado
  - `isRetweet` - Se é um retweet

## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `id` - identificação do tweet
  - `replyToUID` - identificação a quem está respondendo
  - `statusSource` - Origem tweet - smartponhe, computador.
  - `screenName` - nome do usuário
  - `retweetCount` - numero de vezes que o tweet foi retweetado
  - `isRetweet` - Se é um retweet
  - `retweeted` - Se foi retweetado



## Utilizando o twitterR

- O banco de dados conterà as seguintes informações
  - `id` - identificação do tweet
  - `replyToUID` - identificação a quem está respondendo
  - `statusSource` - Origem tweet - smartponhe, computador.
  - `screenName` - nome do usuário
  - `retweetCount` - numero de vezes que o tweet foi retweetado
  - `isRetweet` - Se é um retweet
  - `retweeted` - Se foi retweetado
  - `longitude` e `latitude`

## **Análise dos tweets**

- A partir do banco de dados poderá ser feitos diversas análises

## Análise dos tweets

- A partir do banco de dados poderá ser feitas diversas análises
- O interesse em geral é analisar os textos dos tweets

## Análise dos tweets

- A partir do banco de dados poderá ser feitas diversas análises
- O interesse em geral é analisar os textos dos tweets
- O pacote `tm` (Text Mining) fornece algumas funções para analisar os textos

## Análise dos tweets

- A partir do banco de dados poderá ser feitas diversas análises
- O interesse em geral é analisar os textos dos tweets
- O pacote `tm` (Text Mining) fornece algumas funções para analisar os textos
- Obter os tweets de com a palavras "Dengue" a partir de 01/01/2018

```
rt=searchTwitter("dengue",since="2018-01-01",n=10000,lang="pt")
```

## Análise dos tweets

- A partir do banco de dados poderá ser feitas diversas análises
- O interesse em geral é analisar os textos dos tweets
- O pacote `tm` (Text Mining) fornece algumas funções para analisar os textos
- Obter os tweets de com a palavras "Dengue" a partir de 01/01/2018

```
rt=searchTwitter("dengue",since="2018-01-01",n=10000,lang="pt")
```

- Converter o tweets em banco de dados

```
tweets=twListToDF(rt)
```

## Análise dos tweets

- Corrigir acentuação no texto

```
texto=sapply(tweets$text,function(row)  
  iconv(row, "UTF-8", "ASCII//TRANSLIT", sub = ""))
```

## Análise dos tweets

- Corrigir acentuação no texto

```
texto=sapply(tweets$text,function(row)  
  iconv(row, "UTF-8", "ASCII//TRANSLIT", sub = ""))
```

- Colocar cada observação em um vetor

```
texto1=VectorSource(texto)
```



## Análise dos tweets

- Corrigir acentuação no texto

```
texto=sapply(tweets$text,function(row)  
  iconv(row, "UTF-8", "ASCII//TRANSLIT", sub = ""))
```

- Colocar cada observação em um vetor

```
textol=VectorSource(texto)
```

- Armazena os vetores em documentos

```
txt=Corpus(textol)
```

## Análise dos tweets

- O Corpus pode ser resumido em uma matriz documentos-  
termos

```
dtm=DocumentTermMatrix(txt)
```

## Análise dos tweets

- O `Corpus` pode ser resumido em uma matriz documentos-  
termos

```
dtm=DocumentTermMatrix(txt)
```

- Criar uma matrix com as frequências

```
m=as.matrix(dtm)
```

```
v=sort(rowSums(m),decreasing=TRUE)
```

```
d=data.frame(word = names(v),freq=v)
```

```
View(d)
```

## **Análise dos tweets**

- Limpeza do texto - remover alguns caracteres para criar a nuvem de palavras

## Análise dos tweets

- Limpeza do texto - remover alguns caracteres para criar a nuvem de palavras

- Remover @menção

```
removeMetions=function(x) gsub("@\\S+", "", x)  
txt=tm_map(txt, content_transformer(removeMetions))
```

## Análise dos tweets

- Limpeza do texto - remover alguns caracteres para criar a nuvem de palavras

- Remover @menção

```
removeMetions=function(x) gsub("@\\S+", "", x)  
txt=tm_map(txt, content_transformer(removeMetions))
```

- Remover Hashtags

```
removeHash=function(x) gsub('#\\S+', '', x)  
txt=tm_map(txt, content_transformer(removeHash))
```

## **Análise dos tweets**

- Limpeza do texto

## Análise dos tweets

- Limpeza do texto
  - Remover Palavras de parada - preposição, artigos, etc

```
txt=tm_map(txt, removeWords, iconv(stopwords("portuguese")))
```



## Análise dos tweets

- Limpeza do texto

- Remover Palavras de parada - preposição, artigos, etc

```
txt=tm_map(txt, removeWords, iconv(stopwords("portuguese")))
```

- Remover espaços em branco dobrados

```
txt=tm_map(txt, stripWhitespace)
```

## Análise dos tweets

- Nuvem de palavras pode ser obtida utilizando o pacote `wordcloud`

```
wordcloud(txt, max.words = 200, random.order=FALSE, colors=brewer.pal(8, "Dark2"))
```

## Análise dos tweets

- Nuvem de palavras pode ser obtida utilizando o pacote `wordcloud`

```
wordcloud(txt, max.words = 200, random.order=FALSE, colors=brewer.pal(8, "Dark2"))
```

- `max.words` - máximo de palavras;

## Análise dos tweets

- Nuvem de palavras pode ser obtida utilizando o pacote `wordcloud`

```
wordcloud(txt, max.words = 200, random.order=FALSE, colors=brewer.pal(8, "Dark2"))
```

- `max.words` - máximo de palavras;
- `random.order=FALSE` - plotar em ordem de ocorrência;

## Análise dos tweets

- Nuvem de palavras pode ser obtida utilizando o pacote `wordcloud`

```
wordcloud(txt, max.words = 200, random.order=FALSE, colors=brewer.pal(8, "Dark2"))
```

- `max.words` - máximo de palavras;
- `random.order=FALSE` - plotar em ordem de ocorrência;
- `colors` - cores das palavras

## Análise dos tweets

- Utilizando a matriz documento-termo é possível encontrar associação entre os termos

```
dtm=TermDocumentMatrix(txt)
```

```
findAssocs(dtm, terms = "epidemias", corlimit = 0.3)
```

## Análise dos tweets

- Utilizando a matriz documento-termo é possível encontrar associação entre os termos

```
dtm=TermDocumentMatrix(txt)
```

```
findAssocs(dtm, terms = "epidemias", corlimit = 0.3)
```

- **terms** - palavra que quer verificar as associações

## Análise dos tweets

- Utilizando a matriz documento-termo é possível encontrar associação entre os termos

```
dtm=TermDocumentMatrix(txt)
```

```
findAssocs(dtm, terms = "epidemias", corlimit = 0.3)
```

- `terms` - palavra que quer verificar as associações
- `corlimit` - valor de associação mínima entre as palavras



## Monitoramento da Dengue

- Monitorar da dengue em Cuiabá.

## Monitoramento da Dengue

- Monitorar da dengue em Cuiabá.
  - Investigar a localização Tweets com tema dengue

## Monitoramento da Dengue

- Monitorar da dengue em Cuiabá.
  - Investigar a localização Tweets com tema dengue
  - Analisar a relação dos Tweets com a incidência da dengue

## Monitoramento da Dengue

- Monitorar da dengue em Cuiabá.
  - Investigar a localização Tweets com tema dengue
  - Analisar a relação dos Tweets com a incidência da dengue
  - Identificar área de possíveis transmissão da dengue.

## Monitoramento da Dengue

- Dados de notificação de casos de dengue em Cuiabá

## Monitoramento da Dengue

- Dados de notificação de casos de dengue em Cuiabá
  - Obtidos do SINAN

## Monitoramento da Dengue

- Dados de notificação de casos de dengue em Cuiabá
  - Obtidos do SINAN
  - Categorizados por setor censitário da area urbana de Cuiabá.

## Monitoramento da Dengue

- Objetivos do monitoramento do Twitter utilizando o R



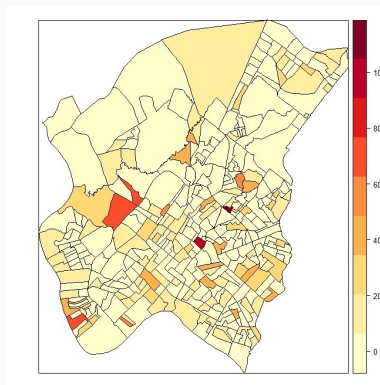
## Monitoramento da Dengue

- Objetivos do monitoramento do Twitter utilizando o R
  - Obter todos os Tweets utilizando a palavra chave "Dengue"

## Monitoramento da Dengue

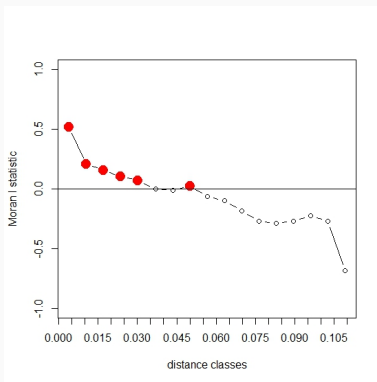
- Objetivos do monitoramento do Twitter utilizando o R
  - Obter todos os Tweets utilizando a palavra chave "Dengue"
  - Assumir que todos os Tweets são relativos a casos de dengue.

## Monitoramento da Dengue



**Figura 1:** Numero de notificações da casos de dengue em Cuiabá por setores censitário no ano de 2017

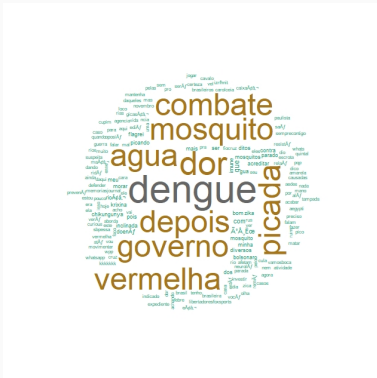
## Monitoramento da Dengue



**Figura 2:** Correlograma de Moran do numero de notificações da casos de dengue em Cuiabá por setores censitário no ano de 2017

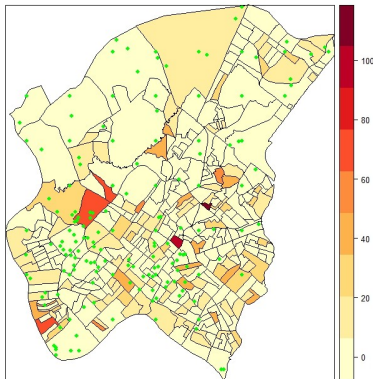
## Monitoramento da Dengue

- Foram obtidos 164 tweets .



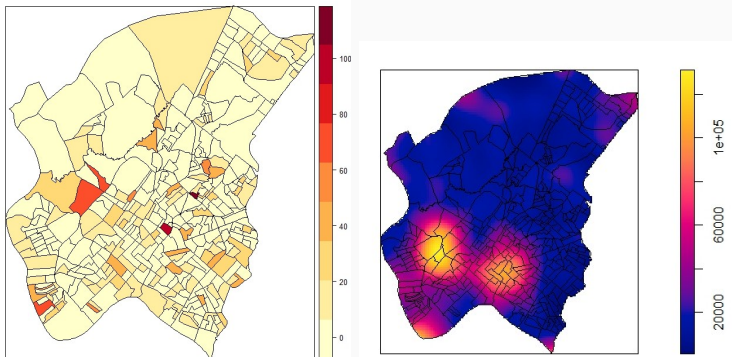
**Figura 3:** Nuvem de palavras dos tweets relativos a dengue em Cuiabá ano de 2017

## Monitoramento da Dengue



**Figura 4:** Tweets relativos a dengue em Cuiabá ano de 2017

## Monitoramento da Dengue



**Figura 5:** Estimador Kernel dos tweets relativos a dengue em Cuiabá por setores censitário no ano de 2017