Mineração de texto aplicada à Lei de Acesso à informação - LAI

true

Rio de Janeiro, 30 de outubro de 2019

Packages for this routine

BASE DE DADOS E ANÁLISE EXPLORATÓRIA

Importação dos dados

Caminho do projeto

```
PATH = "..;/proj_eSIC_v10/textmining_pt/DATA/"
```

Importação ee estrutura dos dados

Tabela1: Pedidos e-SIC

- Pedidos e-SIC

Estrutura dos dados

```
glimpse(Pedidos_eSIC)
```

```
## Observations: 625
## Variables: 9
## $ Protocolo                                <chr> "16853006234201716"...
## $ `Órgão Superior`                         <chr> "EPE - Empresa de P...
## $ `Data de Abertura`                       <dttm> 2017-08-19 20:26:4...
## $ `Prazo de Atendimento`                   <dttm> 2017-09-11 23:59:5...
## $ Situação                                 <chr> "Respondido", "Resp...
## $ `Descrição do Pedido`                    <chr> "A Empresa de Pesqu...
## $ `Descrição da Forma de Resposta do Pedido` <chr> "Pelo sistema (com ...
## $ `Resumo da Solicitação`                  <chr> "Empresa de Pesquis...
## $ `Data da Resposta`                       <dttm> 2017-08-30 21:19:1...
```

Tabela2: Respostas Diretorias da EPE

- Respostas e-SIC (DIRETORIAS EPE)

Estrutura dos dados

```
glimpse(Respostas_EPE)
```

```
## Observations: 705
## Variables: 3
## $ ProtocoloPedido                    <chr> "99938000045201565", "9993...
## $ DataRegistro                       <dttm> 2015-07-24, 2015-07-28, 2...
## $ DiretoriaEPE_ResponsavelPelaDemanda <chr> "DGC", "DEA", "DEA", "DEE"...
```

Tabela3: Stopwords

- Stopwords

```
FILE2 = "DATA/stopwords_PT_FINAL.csv"
stopwords_pt = read.csv(paste0(PATH,FILE2), sep = ';', header = F, encoding = "UTF-8")
stopwords_pt = stopwords_pt[,-2];
cat(paste0("O nosso vetor de stopwords contém ",length(stopwords_pt), " palavras únicas"))
```

```
## O nosso vetor de stopwords contém 734 palavras únicas
```

```
## dim(stopwords_pt); class(stopwords_pt)
stopwords_pt = as.character(stopwords_pt)
stopwords_pt[1:14]
```

```
##  [1] "a"        "à"        "acerca"   "acesso"   "adeus"    "agora"
##  [7] "agradeço" "agradeco" "aí"       "ai"       "ainda"    "alem"
## [13] "além"     "algmas"
```

Tabelas4,5,6: Dicionários de variáveis e-SIC

- Dicionário > BASE DE DADOS - REAL PRO TEXTO DO TCC

Dicionário de variáveis - PEDIDOS

```
dicionario = "DATA/Dicionario-Dados-Exportacao.txt"
dic_pedidos = read.delim(paste0(PATH,dicionario), sep = "-", skip = 3, header = FALSE, nrows = 21) %>%
  select(-V1)
colnames(dic_pedidos) = c("Nome das variáveis", "Tipo e descrição da variável")
#dimnames(dic_pedidos); View(dic_pedidos)
```

Dicionário de variáveis - RECURSOS

```
dic_recursos = read.delim(paste0(PATH,dicionario), sep = "-", skip = 30, header = FALSE, nrows = 17) %>%
  select(-V1)
colnames(dic_recursos) = c("Nome das variáveis", "Tipo e descrição da variável")
#dimnames(dic_recursos); View(dic_recursos)
```

Dicionário de variáveis - SOLICITANTES

```
dicionario = "DATA/Dicionario-Dados-Exportacao.txt"
dic_solicitantes = read.delim(file = paste0(PATH,dicionario), sep = "-", skip = 53, header = FALSE, nrou
  select(-V1)
colnames(dic_solicitantes) = c("Nome das variáveis", "Tipo e descrição da variável")
#dimnames(dic_solicitantes); View(dic_solicitantes)
```

Transformação e pré-processamento dos dados

Filtra, Transforma e Unifica bases

Filtro1: tabela consulta de pedidos

Filtrando apenas as variáveis de interesse do estudo na tabela de consulta de pedidos

```
LAI = Pedidos_eSIC
LAI = LAI %>% select(Protocolo, `Data de Abertura`, `Prazo de Atendimento`, `Descrição do Pedido`, `Resu
```

Transformação1: renomeando colunas

Reescrevendo o nome das variáveis de ambas tablelas

```
colnames(LAI) = c("Protocolo", "DATA_REGISTRO", "DATA_PRAZOATEND", "DESCRI_PEDIDO",
                      "RESUMO_PEDIDO", "DATA_RESPOSTA")
LAI1 = Respostas_EPE
colnames(LAI1) = c("Protocolo", "DATA_REGISTRO", "DIRETORIAS")
# glimpse(LAI1)
```

Transformação2: transforma as.factor( ) variável DIRETORIAS

character em factor

Transformação3: cria a variável PEDIDO = RESUMO + DESCRIÇÃO

```
LAI$PEDIDO = paste(LAI$RESUMO_PEDIDO, LAI$DESCRI_PEDIDO)
```

Análise1: Quantitativo de pedidos por diretoria

Transformação3: substitui NA por OUTROS (coluna DIRETORIAS)

Primeiro conta o número de pedidos por diretoria (observações por categorias)

```
LAI1 %>%
    group_by(DIRETORIAS) %>%  count(sort = TRUE) %>%
  kable("latex", caption = "Quantitativo de solicitações por Diretoria/EPE via e-SIC - substituição NA e
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

```
## Warning: Factor `DIRETORIAS` contains implicit NA, consider using
## `forcats::fct_explicit_na`

## Warning: Factor `DIRETORIAS` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

Table 1: Quantitativo de solicitações por Diretoria/EPE via e-SIC - substituição NA em OUTROS

DIRETORIAS	n
DEE	244
DEA	240
DGC	121
OUTROS	66
DPG	33
NA	1

Existe um valor NA, vamos subistitui-lo como OUTROS

```
LAI1 = LAI1 %>%
  replace_na(list(DIRETORIAS = "OUTROS"))
```

Tabela1: Quantitativo de pedidos por diretoria - sem reclassificação

- Tabela 01 número de solcitações/pedidos de informação (sem NA)

```
pedidos_diretoria = LAI1 %>%
 count(DIRETORIAS, sort = TRUE, name = "total_pedidos")
pedidos_diretoria %>%
  kable("latex", caption = "Quantitativo de solicitações por Diretoria/EPE via e-SIC - sem reclassifica
        booktabs = T) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 2: Quantitativo de solicitações por Diretoria/EPE via e-SIC - sem reclassificação

DIRETORIAS	total_pedidos
DEE	244
DEA	240
DGC	121
OUTROS	67
DPG	33

Verificamos a existência de 4 diretorias, sendo elas: *DEA*, *DEE*, *DGC*, *DPG* e *OUTROS*. Essa última é devido a existência de informações solicitadas que não são de competência de nenhuma das cinco diretorias, daí a necessidade de uma última categoria *OUTROS* para atender essas demandas.

Fica nítida o desbalanceamento do número de pedidos por categoria. Enquanto as diretorias *DEE* e *DEA* possuem, respectivamente, 244 e 240 pedidos verifica-se uma diferença grande do número de pedido das diretorias *DGC* e *DPG* e também da categoria *OUTRAS*, onde se forem somadas possuem um total de 221 pedidos conjuntamente.

A seguir, um passo importante de reclassificação será executado devido ao número pequeno de solicitações para as diretorias DGC e DPG Apenas uma solcitação existente no nosso banco de dados para essa diretoria. Iremos, portanto, unificar essa demanda à categoria *OUTROS*. A seguir, verificamos nas tabela 01 e 02 a distribuição de pedidos por diretoria antes e após reclassificação das mesmas.

Tabela1: Quantitativo de pedidos por diretoria - sem reclassificação

Vamos criar uma nova variável: DIRETORIA que é basicamente uma reclassificação da variável DIRETORIAS

Vamos, primeiro, armazenar um vetor com o nome das categorias de DIRETORIAS originais.

```
(diretorias = levels(as.factor(LAI1$DIRETORIAS)))
```

```
## [1] "DEA"    "DEE"    "DGC"    "DPG"    "OUTROS"
```

Transformação4 - Reclassificação das Diretorias

Respostas e-SIC - Reclassificação Diretorias

```
LAI1$DIRETORIAS = as.character(LAI1$DIRETORIAS) # glimpse(LAI1)
LAI1 = LAI1 %>%
  mutate(DIRETORIA = ifelse(DIRETORIAS == "DGC", "OUTROS",
                            ifelse(DIRETORIAS == "DPG", "OUTROS",DIRETORIAS)))
(diretorias1 = levels(as.factor(LAI1$DIRETORIA)))
```

```
## [1] "DEA"    "DEE"    "OUTROS"
```

Tabela2: Quantitativo de pedidos por diretoria - após reclassificação

- Tabela 02 número de solcitações/pedidos de informação - após reclassificação

```r
pedidos_diretoria1 = LAI1 %>%
 count(DIRETORIA, sort = TRUE, name = "total_pedidos")
pedidos_diretoria1 %>%
  kable("latex", caption = "Quantitativo de solicitações por Diretoria/EPE via e-SIC - após reclassifica
        booktabs = T) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 3: Quantitativo de solicitações por Diretoria/EPE via e-SIC - após reclassificação

DIRETORIA	total_pedidos
DEE	244
DEA	240
OUTROS	221

Temos, finalmente um maior balanceamento nas categorias da nossa variável resposta com 244, 240 e 221 pedidos que foram destinios à *DEE*, *DEA* e *OUTROS*, respectivamente. Onde *OUTROS* é a categoria formada com a união dos pedidos das diretorias *DGC*, *DPG* e *OUTROS*.

A reclassificação foi, também, uma decisão suportada por análises préveias do presente estudo. Foi avaliada a viabilidade de aplicar o estudo com as categorias originais, entretanto na fase de modelagem preditiva o desempenho do modelo do Random Forest foi muito inferior comparado ao modelo após reclassificação. Um motivo plausível para a melhoria de performance pode ser por conta do maior balanceamento entre as categorias da variável resposta **Diretoria**, em questão.

- Unificando as Bases

É necessário, agora, unificar as bases de dados pertinentes às solicitações e respostas.

Join1: União das bases em questão

```r
LAI1 = LAI1 %>% select(-DATA_REGISTRO); #dim(LAI1)
DB = left_join(x = LAI, y = LAI1, by = "Protocolo") %>%
  drop_na()
#View(head(DB))
```

```r
glimpse(DB)
```

```
## Observations: 624
## Variables: 9
## $ Protocolo       <chr> "16853006234201716", "18600000523201890", "234...
## $ DATA_REGISTRO   <dttm> 2017-08-19 20:26:47, 2018-03-07 18:29:43, 201...
## $ DATA_PRAZOATEND <dttm> 2017-09-11 23:59:59, 2018-03-28 23:59:59, 201...
## $ DESCRI_PEDIDO   <chr> "A Empresa de Pesquisa Energética (vinculada a...
## $ RESUMO_PEDIDO   <chr> "Empresa de Pesquisa Energética", "Demanda ou ...
## $ DATA_RESPOSTA   <dttm> 2017-08-30 21:19:17, 2018-03-08 16:50:43, 201...
## $ PEDIDO          <chr> "Empresa de Pesquisa Energética A Empresa de P...
## $ DIRETORIAS      <chr> "DGC", "OUTROS", "DEA", "DEE", "DEA", "DPG", "...
## $ DIRETORIA       <chr> "OUTROS", "OUTROS", "DEA", "DEE", "DEA", "OUTR...
```

```r
cat(paste0("Existem ", dim(DB)[1]," observações/pedidos na base de dados."))
```

```
## Existem 624 observações/pedidos na base de dados.
```

```r
cat(paste0("Com registros de pedidos datados entre ", format(min(DB$DATA_REGISTRO), '%d de %B de %Y'),"
```

```
## Com registros de pedidos datados entre 07 de Julho de 2015 a 25 de Março de 2019.
```

Ver Anexo 01 c/ amostra dos dados da tabela que serpá utilizada para manipulação daqui pra frente.

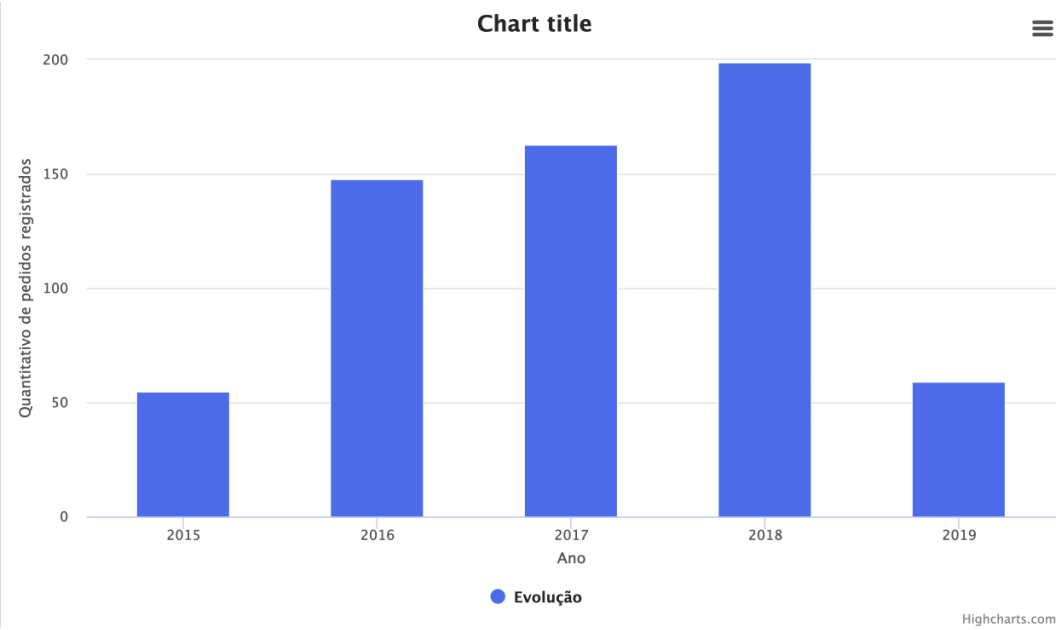
- Evolução de pedidos:

Data de registro do pedido

```
db_evolPedido = DB %>% select(Protocolo, DIRETORIAS, DIRETORIA, DATA_REGISTRO) %>%
  mutate(DIASEMANA_REGISTRO = weekdays(DB$DATA_REGISTRO),
         HORA_REGISTRO = hour(DB$DATA_REGISTRO),
         MES_REGISTRO = base::months.Date(DB$DATA_REGISTRO),
         ANO_REGISTRO = year(DB$DATA_REGISTRO))

ano_evolution = db_evolPedido %>%
  group_by(ANO_REGISTRO) %>% count()

hc2_1 <- highchart() %>%
  hc_add_series(data = ano_evolution$n,
                type = "column",
                name = "Evolução",
                showInLegend = TRUE,
                tooltip = list(valueDecimals = 2, valuePrefix = "",
                               valueSuffix = " pedidos registrados"),
                               color = "#5F83EE", fillOpacity = 0.1) %>%
  hc_yAxis(title = list(text = "Quantitativo de pedidos registrados"),
           allowDecimals = FALSE, max = max(ano_evolution$n),
           labels = list(format = "{value}")) %>%
  hc_xAxis(title = list(text = "Ano"),
           categories = ano_evolution$ANO_REGISTRO,
           tickmarkPlacement = "on",
           opposite = FALSE) %>%
  hc_title(#text = "Evolução de pedidos registrados via LAI (EPE)",
           style = list(fontWeight = "bold")) %>%
  hc_subtitle(text = paste("")) %>%
      hc_tooltip(valueDecimals = 2,
                 pointFormat = "Importância: {point.y}")%>%
                 #pointFormat = "Variável: {point.x} <br> Importância: {point.y}")
      hc_credits(enabled = TRUE,
                 #text = "Fonte: CGU, e-SIC. Elaboração: Leal, Alize; Pimenta, Ewerson.",
                 style = list(fontSize = "10px")) %>%
  hc_exporting(enabled = TRUE, filename = "F6_1-importance-Pimenta")
#hc <- hc %>%
#  hc_add_theme(hc_theme_darkunica())
hc2_1

ano_evolution_DIR = db_evolPedido %>%
  group_by(DIRETORIAS, ANO_REGISTRO) %>% count()


DEE = ano_evolution_DIR %>% filter(DIRETORIAS == "DEE") %>% arrange(desc(ANO_REGISTRO), .by_group = TRU
DEA = ano_evolution_DIR %>% filter(DIRETORIAS == "DEA") %>% arrange(desc(ANO_REGISTRO), .by_group = TRU
DGC = ano_evolution_DIR %>% filter(DIRETORIAS == "DGC") %>% arrange(desc(ANO_REGISTRO), .by_group = TRU
DPG = ano_evolution_DIR %>% filter(DIRETORIAS == "DPG") %>% arrange(desc(ANO_REGISTRO), .by_group = TRU
OUTROS = ano_evolution_DIR %>% filter(DIRETORIAS == "OUTROS") %>% arrange(desc(ANO_REGISTRO), .by_group
```

```r
hc2_2 <- highchart() %>%
  hc_add_series(data = DEE$n,
                type = "line",
                name = "DEE",
                showInLegend = TRUE,
                tooltip = list(valueDecimals = 0, valuePrefix = "",
                               valueSuffix = " pedidos registrados"),
                               color = "#5F83EE", fillOpacity = 0.1) %>%
  hc_add_series(data = DEA$n,
                type = "line",
                name = "DEA",
                showInLegend = TRUE,
                tooltip = list(valueDecimals = 0, valuePrefix = "",
                               valueSuffix = " pedidos registrados"),
                               color = "skyblue", fillOpacity = 0.1) %>%
  hc_add_series(data = DGC$n,
                type = "line",
                name = "DGC",
                showInLegend = TRUE,
                tooltip = list(valueDecimals = 0, valuePrefix = "",
                               valueSuffix = " pedidos registrados"),
                               color = "green", fillOpacity = 0.1) %>%
  hc_add_series(data = DPG$n,
                type = "line",
                name = "DPG",
                showInLegend = TRUE,
                tooltip = list(valueDecimals = 0, valuePrefix = "",
                               valueSuffix = " pedidos registrados"),
                               color = "black", fillOpacity = 0.1) %>%
  hc_add_series(data = OUTROS$n,
                type = "area",
                name = "OUTROS",
                showInLegend = TRUE,
                tooltip = list(valueDecimals = 0, valuePrefix = "",
                               valueSuffix = " pedidos registrados"),
                               color = "pink", fillOpacity = 0.5) %>%

  hc_yAxis(title = list(text = "Quantitativo de pedidos registrados"),
           allowDecimals = FALSE, max = max(DEE$n, DEA$n, DGC$n, DPG$n, OUTROS$n),
           labels = list(format = "{value}"), #minorTickInterval = "auto",
           #minorGridLineDashStyle = "LongDashDotDot",
           showFirstLabel = TRUE,
           showLastLabel = TRUE) %>%
  hc_xAxis(title = list(text = "Ano"),
           categories = ano_evolution$ANO_REGISTRO,
           tickmarkPlacement = "on",
           opposite = FALSE) %>%
  hc_title(#text = "Evolução de pedidos registrados via LAI (EPE)",
           style = list(fontWeight = "bold")) %>%
  hc_subtitle(text = paste("")) %>%
      hc_tooltip(valueDecimals = 2,
                 pointFormat = "Número de {point.y}")%>%
                 #pointFormat = "Variável: {point.x} <br> Importância: {point.y}")
```

```
      hc_credits(enabled = TRUE,
                 #text = "Fonte: CGU, e-SIC. Elaboração: Leal, Alize; Pimenta, Ewerson.",
                 style = list(fontSize = "10px")) %>%
  hc_exporting(enabled = TRUE, filename = "F6_1-importance-Pimenta")
#hc <- hc %>%
#  hc_add_theme(hc_theme_darkunica())
hc2_2
```




```
summary(DB$DATA_REGISTRO)
class()
inic = as.Date(min(DB$DATA_REGISTRO), format = "%m/%d/%y %H:%M:%S", tz = "UTC")
fim = max(DB$DATA_REGISTRO), date_format())
cat(paste0("Período de pedidos registrados que serão utilizados nessa análise ", inic, " até ", fim))

time_index_h <- seq(from = as.POSIXct(inic),
                    to = as.POSIXct(fim), by = "hour")
time_index_w <- weekdays(time_index_h)
# or
#time_index_w <- lubridate::wday(time_index_h)
library(lubridate)
date<-ymd_hms("2016-06-06 09:45:12")
wday(date)
```

Mineração de texto

Palavras por pedido

Análise2: distribuição de frequência de palavras por diretoria e algumas estatísticas descritivas

Ferramentas

Iniciamos as manipulações utilizando recursos da função `unnest_tokens( )` do pacote `library(tidytext)`

que nos permite trabalhar com textos em um formato `tidy`, ou seja que coloca uma palavra por linha em uma única coluna, formando, assim, *termos/palavras* por linha. Utilizamos, também, ainda os recursos do pacote `library(diplyr)` para, posteriormente, agrupar esses termos por diretoria e calcular a frequência dos *termos*.

Verificamos que as 10 palavras mais frequentes em todos os pedidos realizados são palavras sem acréscimo contextual, pois essas não acrescentam nenhum sentido semântico como, por exemplo: preposições (de, da, do, para, em, no), conjunção (e) e artigos(o,a).

Citar o que é preoposição.

Tabela3: Palavras mais frequentes

- Tabela 03 Palavras mais frequentes em todo o conjunto de solicitações

```
library(tidytext)
palavras <- DB %>%
  unnest_tokens(palavra, PEDIDO) %>%
  count(palavra, sort = TRUE) %>%
  ungroup()

palavras[0:10,] %>%
  kable("latex", caption = "Principais palavras com stopwords",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 4: Principais palavras com stopwords

palavra	n
de	4.071
a	1.311
e	1.162
do	900
o	891
da	821
para	712
no	584
em	575
energia	553

Tabelas4,5,6: Palavras mais frequentes por diretoria

Cria o objeto de palavras por diretoria

```
palavras_diretoria <- DB %>%
  unnest_tokens(palavra, PEDIDO) %>%
  count(DIRETORIA,palavra, sort = TRUE) %>%
  ungroup() %>%  droplevels() %>% drop_na()

palavras_diretoria$DIRETORIA = as.factor(palavras_diretoria$DIRETORIA)
```

Tabelas4: Palavras mais frequentes DEA

- Tabela 04 Palavras mais frequentes no conjunto de solicitações por diretoria

```r
DEA_termo =
palavras_diretoria %>%
  filter(DIRETORIA == "DEA") %>% droplevels()

DEA_termo %>%
    top_n(n = 10) %>%
  kable("latex", caption = "Principais palavras com stopwords (DEA)",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

```
## Selecting by n
```

Table 5: Principais palavras com stopwords (DEA)

DIRETORIA	palavra	n
DEA	de	1.467
DEA	a	405
DEA	e	389
DEA	do	319
DEA	energia	303
DEA	o	292
DEA	dados	273
DEA	da	259
DEA	por	256
DEA	no	252

Tabelas5: Palavras mais frequentes DEE

- Tabela 05 Palavras mais frequentes no conjunto de solicitações por diretoria

```r
DEE_termo =
palavras_diretoria %>%
  filter(DIRETORIA == "DEE") %>% droplevels()

DEE_termo %>%
    top_n(n = 10) %>%
  kable("latex", caption = "Principais palavras com stopwords (DEA)",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

```
## Selecting by n
```

Tabelas6: Palavras mais frequentes OUTROS

- Tabela 06 Palavras mais frequentes no conjunto de solicitações por diretoria

```r
OUTROS =
palavras_diretoria %>%
  filter(DIRETORIA == "OUTROS") %>% droplevels()

OUTROS %>%
    top_n(n = 10) %>%
  kable("latex", caption = "Principais palavras com stopwords (OUTROS)",
```

Table 6: Principais palavras com stopwords (DEA)

DIRETORIA	palavra	n
DEE	de	1.241
DEE	a	415
DEE	e	318
DEE	do	317
DEE	o	268
DEE	da	267
DEE	para	251
DEE	energia	202
DEE	no	182
DEE	que	161

```
      booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

```
## Selecting by n
```

Table 7: Principais palavras com stopwords (OUTROS)

DIRETORIA	palavra	n
OUTROS	de	1.363
OUTROS	a	491
OUTROS	e	455
OUTROS	o	331
OUTROS	da	295
OUTROS	do	264
OUTROS	para	210
OUTROS	em	207
OUTROS	que	199
OUTROS	ou	180

Mesmo assim, abrindo para cada uma das 3 possíveis cateogorias da variável **Diretoria** temos que as principais palavras não agregam nenhum valor semântico, exceto pela palavra energia que apareceu na oitava e nona colocação de maior frequência dos documentos de pedidos enviados à *DEA* e *DEE*, respectivamente. Isso devido ao excesso de uso de **stop words** em textos humanos.

Em passos mais adiante serão removidas essas palavras, **stop words**, e a partir da remoção o trabalho se dará apenas com palavras de sentido semântico relevante aos subjetivos solicitados às diretorias, acrescentando assim maior assertividade do modelo de classificação.

Verificamos, antes disso, o total, freq. e média de palavras por diretoria, bem como comparações 2 a 2 para cada uma das categorias. E avançamos um pouco com gráficos da contagem de frequência e a lei de **Zipf** que dá suporte as conclusões do passo anterior e, a por conseguinte, é definida a estatística de **tf_idf** (**term frequency times inverse document frequency**), uma estatística utilizada para ressaltar termos relevantes para um documento em particular.

Análise2: Comparação de freq. de palavras por diretoria

- Total de palavras por diretoria, total de pedidos por diretoria e número médio de palavras por pedido e diretoria

```
total_palavras = palavras_diretoria %>%
  group_by(DIRETORIA) %>%
  summarize(total_palavras = sum(n))

total_palavras$DIRETORIA = as.character(total_palavras$DIRETORIA)
total_palavras = left_join(x = total_palavras, y = pedidos_diretoria1,
                           by = "DIRETORIA") %>%
mutate(media_palavras_porpedidoEdiretoria = total_palavras/total_pedidos)
```

Tabelas7: Total de palavras por diretoria, total de pedidos por diretoria e número médio de palavras por pedido e diretoria

- Total de palavras por diretoria, total de pedidos por diretoria e número médio de palavras por pedido e diretoria

```
total_palavras %>%
  kable("latex", caption = "Total de palavras, total de pedidos e número médio de palavras
        por pedido e diretoria",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 8: Total de palavras, total de pedidos e número médio de palavras por pedido e diretoria

DIRETORIA	total_palavras	total_pedidos	media_palavras_porpedidoEdiretoria
DEA	17.211	240	71,71250
DEE	15.536	244	63,67213
OUTROS	16.970	221	76,78733

Temos que o número médio de palavras por pedido é parecido entre as diretorias. com médias de 55 palavras por pedido para DEE e 69,7 e 61,7, respectivamente para DEA e OUTROS.

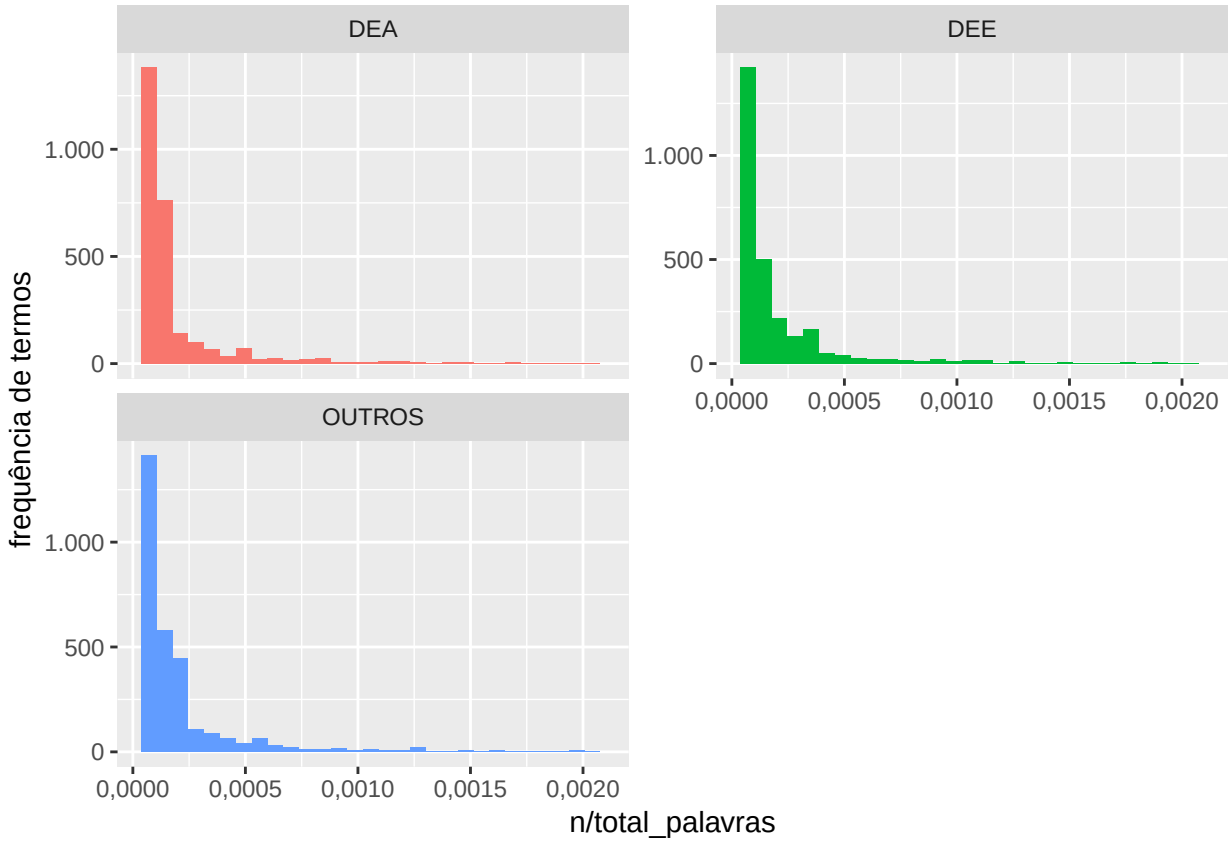
Figura1: Distribuição de frequência de termos por diretoria

- Distribuição da freq. de palavras usadas em solicitações por diretoria (histograma)

```
diretoria_palavras <- DB %>%
  unnest_tokens(palavra, PEDIDO) %>%
  count(DIRETORIA, palavra, sort = TRUE) %>%
  ungroup()

diretoria_palavras = left_join(diretoria_palavras, total_palavras, by = "DIRETORIA")

library(ggplot2)
gcomma <- function(x) format(x, big.mark = ".", decimal.mark = ",", scientific = FALSE)

ggplot(diretoria_palavras, aes(n/total_palavras, fill = DIRETORIA)) +
  geom_histogram(show.legend = FALSE) + xlim(NA, 0.0021) +
facet_wrap(~DIRETORIA, ncol = 2, scales = "free_y") +
  scale_y_continuous(labels=gcomma)  +
```

```
  scale_x_continuous(labels=gcomma, limits = c(NA, 0.0021)) +
  labs(y = "frequência de termos")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which
## will replace the existing scale.

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 182 rows containing non-finite values (stat_bin).

## Warning: Removed 3 rows containing missing values (geom_bar).
```



Pelos histogramas fica claro que as distribuições da frequência de termos por diretoria possuem caudas mais alongadas à direita. Além disso, algumas frequências não foram evidenciadas no gráfico por questões de escala. De fato, as palavras/termos de maior recorrência nos documentos/textos são as de menor relevância em contexto semântica.

Sabemos, portanto, que queremos encontrar valor exatamente nas partes mais longas à direita das distribuições de frequência de termos, uma vez que ali se encontram as palavras de maior relevância contextual.

Logo, a seguir, usamos da definição da lei **Zipf** que afirma que a frequência que uma palavra (ou termo) aparece em um documento é inversamente proporcional ao seu ranque.

> lei de Zipf's

Citar, aqui, "There are very long tails to the right for these novels (those extremely common words!) that we have not shown in these plots. These plots exhibit similar distributions for all the novels, with many words that occur rarely and fewer words that occur frequently." pág. 31 (Silge, Robinson). Que averigua que documentos de texto tendem a ter distribuições de frequência de palavras similar, por conta das stopwords.

Ainda de acordo com os autores, "Distributions like those shown in Figure 3-1 are typical in language. In fact,

those types of long-tailed distributions are so common in any given corpus of natural lan- guage (like a book, or a lot of text from a website, or spoken words) that the relation- ship between the frequency that a word is used and its rank has been the subject of study." e por essa razão e a relação verificada por George Zipf da relação inversa entre freq. de palavra e ranque tiramos valor dos documentos partindo dessas premissas.

- Ranque de palavras pela pela lei de **Zipf**

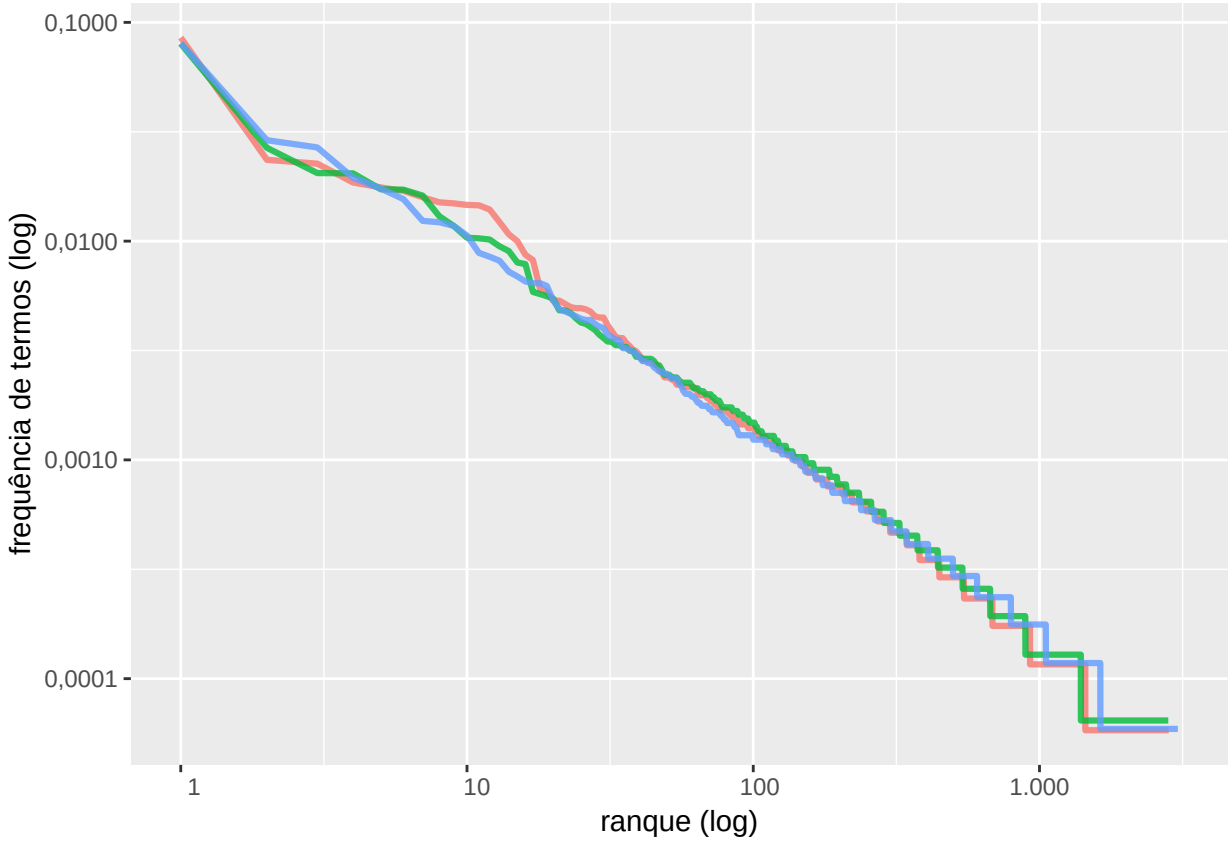
```
freq_by_rank <- diretoria_palavras %>%
group_by(DIRETORIA) %>%
mutate(ranque = row_number(),
`frequência de termos` = n/total_palavras)
```

Figura1: Lei de Zipf

- Zipf's law

```
#plot1
freq_by_rank %>%
ggplot(aes(ranque, `frequência de termos`, color = DIRETORIA)) +
  geom_line(size = 1.1, alpha = 0.8, show.legend = FALSE) + scale_x_log10() +
  scale_y_log10(labels=gcomma) +
  scale_x_log10(labels=gcomma) +
  labs(y = "frequência de termos (log)", x = "ranque (log)")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which
## will replace the existing scale.
```



Vemos que exatamente nas extremidades do gráfico tem-se uma não sobreposição de frequências por diretoria. Detalhe que o gráfico, em questão, está na escala logarítmica no eixo x (ranque) e eixo y (freq. de termos).

Plotando desta forma, a relação inversamente proporcional terá uma inclinação constante e negativa.

Tendo em vista, portanto, que o gráfico referido está em cordenadas log-log e dado a semelhança de todos os documentos de texto das diferentes diretorias, afirmamos que para todas as diretorias pela Lei de **Zipf** a relação entre ranque e freq. de termos assumirá, sempre, uma inclinação negativa, ou seja,

Daí, aplicando a escala log-log temos que e podemos aplicar um ajuste a fim de encontrar um intercepto e coef. angular para traçar no gráfico anterior.

$$frequência \propto \frac{1}{ranque} \implies log(frequência) \propto log\left(\frac{1}{ranque}\right)$$

Reescrever e exlicar a seguimentação em 3 partes como uma "lei de potenciacao dividida em 3 partes" e então utilizar do seguimento do meio, onde as freq. de ternos sao mais semelhantes para diferentes ranques das diferentes diretorias. Fica claro pela eq.

"Notice that Figure 3-2 is in log-log coordinates. We see that all six of Jane Austen's novels are similar to each other, and that the relationship between rank and fre- quency does have negative slope. It is not quite constant, though; perhaps we could view this as a broken power law with, say, three sections. Let's see what the exponent of the power law is for the middle section of the rank range."

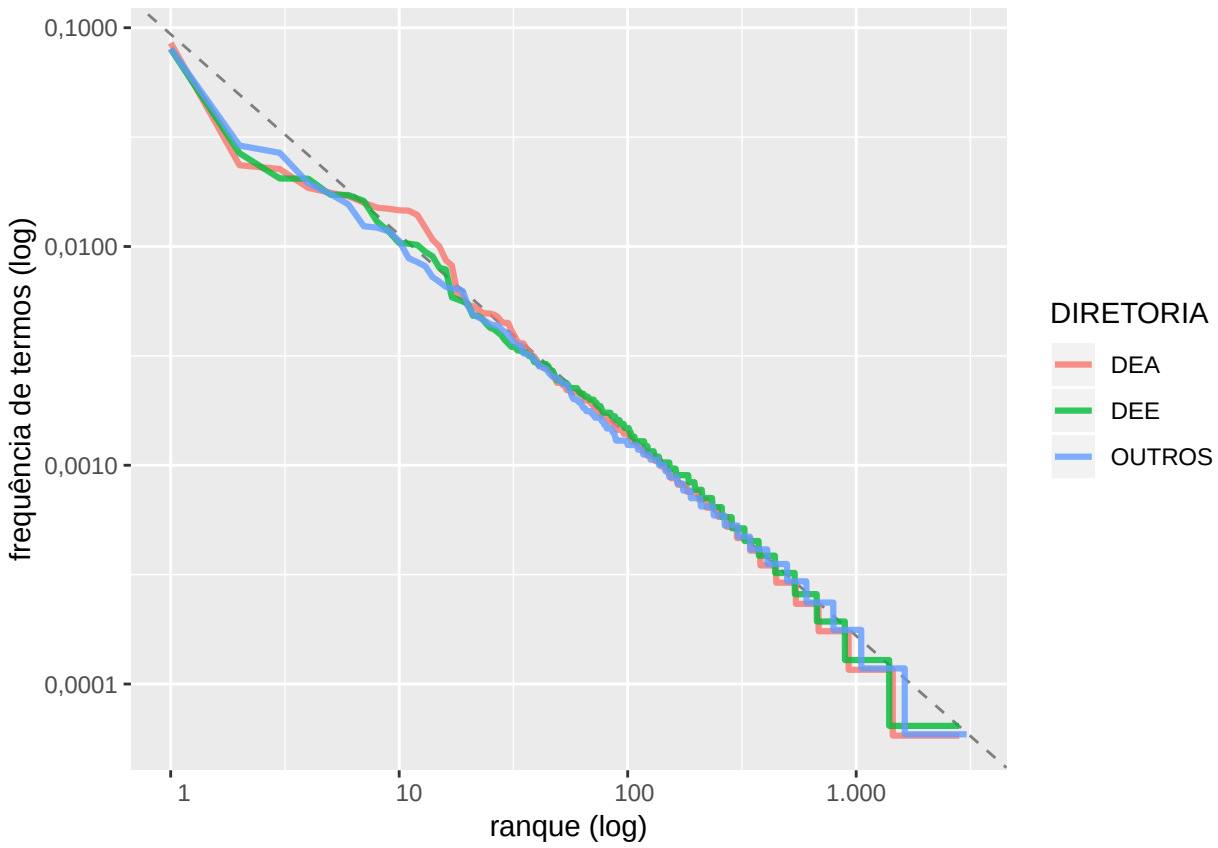
```
rank_subset <- freq_by_rank %>%
      filter(ranque < 500, ranque > 50)

(zipf_ajusteloglog <- lm(log10(`frequência de termos`) ~ log10(ranque),
                         data = rank_subset))
```

```
##
## Call:
## lm(formula = log10(`frequência de termos`) ~ log10(ranque),
##     data = rank_subset)
##
## Coefficients:
##   (Intercept)  log10(ranque)
##       -1.0298        -0.9166
```

Finalmente, traçando e sobrepondo o gráfico anterior com os valores de initercepto e coeficiente angular obtidos no ajuste do passo anterior temos a figura a seguir.

Figura1: Lei de Zipf + ajuste log-log

```
freq_by_rank %>%
ggplot(aes(ranque, `frequência de termos`, color = DIRETORIA)) +
geom_abline(intercept = coefficients(zipf_ajusteloglog)[1], slope = coefficients(zipf_ajusteloglog)[2],
geom_line(size = 1.1, alpha = 0.8, show.legend = TRUE) +
  scale_y_log10(labels=gcomma) +
  scale_x_log10(labels=gcomma) +
  labs(y = "frequência de termos (log)", x = "ranque (log)")
```



The Bind tf_idf

Fundamentar o uso da estatística **tf_idf**, bem como descrever a definição.

> The idea of tf-idf is to find the important words for the content of each document by decreasing the weight for commonly used words and increasing the weight for words that are not used very much in a collection or corpus of documents, in this case, the group of Jane Austen's novels as a whole. Calculating tf-idf attempts to find the words that are important (i.e., common) in a text, but not too common. Let's do that now. The bind_tf_idf function in the tidytext package takes a tidy text dataset as input with one row per token (term), per document. One column (word here) contains the terms/tokens, one column contains the documents (book in this case), and the last necessary column contains the counts, or how many times each document contains each term (n in this example). We calculated a total for each book for our explora- tions in previous sections, but it is not necessary for the bind_tf_idf function; the table only needs to contain all the words in each document.

```
round_df <- function(x, digits) {
    # round all numeric variables
    # x: data frame
    # digits: number of digits to round
    numeric_columns <- sapply(x, mode) == 'numeric'
    x[numeric_columns] <-  round(x[numeric_columns], digits)
    x
}
```

- Palavras mais relevantes de acordo com a estatística **tf_idf**

```r
diretoria_palavras_tfidf <- diretoria_palavras %>%
  bind_tf_idf(palavra, DIRETORIA, n) %>%
  select(-total_palavras, -total_pedidos, -media_palavras_porpedidoEdiretoria) %>%
  arrange(desc(tf_idf))

#options(digits=4)
set.seed(7456)
amostra1 = sample(seq(1:dim(diretoria_palavras_tfidf)[1]), 10, replace = FALSE)
round_df(diretoria_palavras_tfidf[amostra1,],6)  %>%
  kable("latex", caption = "Total de palavras, total de pedidos e número médio de palavras
        por pedido e diretoria",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 9: Total de palavras, total de pedidos e número médio de palavras por pedido e diretoria

DIRETORIA	palavra	n	tf	idf	tf_idf
DEE	potencial	5	0,000322	0,405465	0,000130
DEA	consumidores	34	0,001975	0,000000	0,000000
DEE	características	1	0,000064	0,000000	0,000000
DEA	deste	5	0,000291	0,000000	0,000000
DEA	contar	2	0,000116	1,098612	0,000128
DEA	conjuntura	1	0,000058	0,405465	0,000024
DEE	creio	1	0,000064	1,098612	0,000071
DEA	d'apote	1	0,000058	1,098612	0,000064
OUTROS	órgãos	11	0,000648	0,000000	0,000000
DEE	orientador	1	0,000064	0,405465	0,000026

A estatística faz um trabalho brilhante ao ressaltar as palavras mais relevantes dentro de cada conjunto de documentos (diretorias). As tabelas a seguir mostram as 10 palavras mais relevantes de acordo com a estatística tf_idf por diretoria

Tabela8: top 12 termos ordenados pela estatística tf_idf (DEE)

```r
round_df(diretoria_palavras_tfidf,5) %>%
  filter(DIRETORIA == "DEE") %>%
  top_n(12,tf_idf) %>%
  kable("latex", caption = "Top 10 termos (DEE)",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Tabela9: top 12 termos ordenados pela estatística tf_idf (DEA)

```r
round_df(diretoria_palavras_tfidf,5) %>%
  filter(DIRETORIA == "DEA") %>%
  top_n(12,tf_idf) %>%
  kable("latex", caption = "Top 10 termos (DEA)",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Tabela10: top 10 termos ordenados pela estatística tf_idf (OUTROS)

Table 10: Top 10 termos (DEE)

DIRETORIA	palavra	n	tf	idf	tf_idf
DEE	leilão	75	0,00483	1,09861	0,00530
DEE	cadastrados	19	0,00122	1,09861	0,00134
DEE	eólica	32	0,00206	0,40547	0,00084
DEE	r1	10	0,00064	1,09861	0,00071
DEE	rev0	10	0,00064	1,09861	0,00071
DEE	habilitados	9	0,00058	1,09861	0,00064
DEE	porto	9	0,00058	1,09861	0,00064
DEE	ventos	9	0,00058	1,09861	0,00064
DEE	cálculos	8	0,00051	1,09861	0,00057
DEE	módulos	8	0,00051	1,09861	0,00057
DEE	parâmetros	8	0,00051	1,09861	0,00057
DEE	puc	8	0,00051	1,09861	0,00057

Table 11: Top 10 termos (DEA)

DIRETORIA	palavra	n	tf	idf	tf_idf
DEA	municípios	13	0,00076	1,09861	0,00083
DEA	faixa	11	0,00064	1,09861	0,00070
DEA	nuclear	10	0,00058	1,09861	0,00064
DEA	riachão	10	0,00058	1,09861	0,00064
DEA	eia	8	0,00046	1,09861	0,00051
DEA	kwh	8	0,00046	1,09861	0,00051
DEA	rima	8	0,00046	1,09861	0,00051
DEA	balanço	20	0,00116	0,40547	0,00047
DEA	distribuição	20	0,00116	0,40547	0,00047
DEA	solar	20	0,00116	0,40547	0,00047
DEA	ambiental	19	0,00110	0,40547	0,00045
DEA	condicionado	7	0,00041	1,09861	0,00045
DEA	porcentagem	7	0,00041	1,09861	0,00045

```
round_df(diretoria_palavras_tfidf,5) %>%
  filter(DIRETORIA == "OUTROS") %>%
  top_n(12,tf_idf) %>%
  kable("latex", caption = "Top 10 termos (DEA)",
        booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Ou simplesmente verificamos através de um gráfico

Figura2: Termos mais relevantes por diretoria pela estatística tf_idf

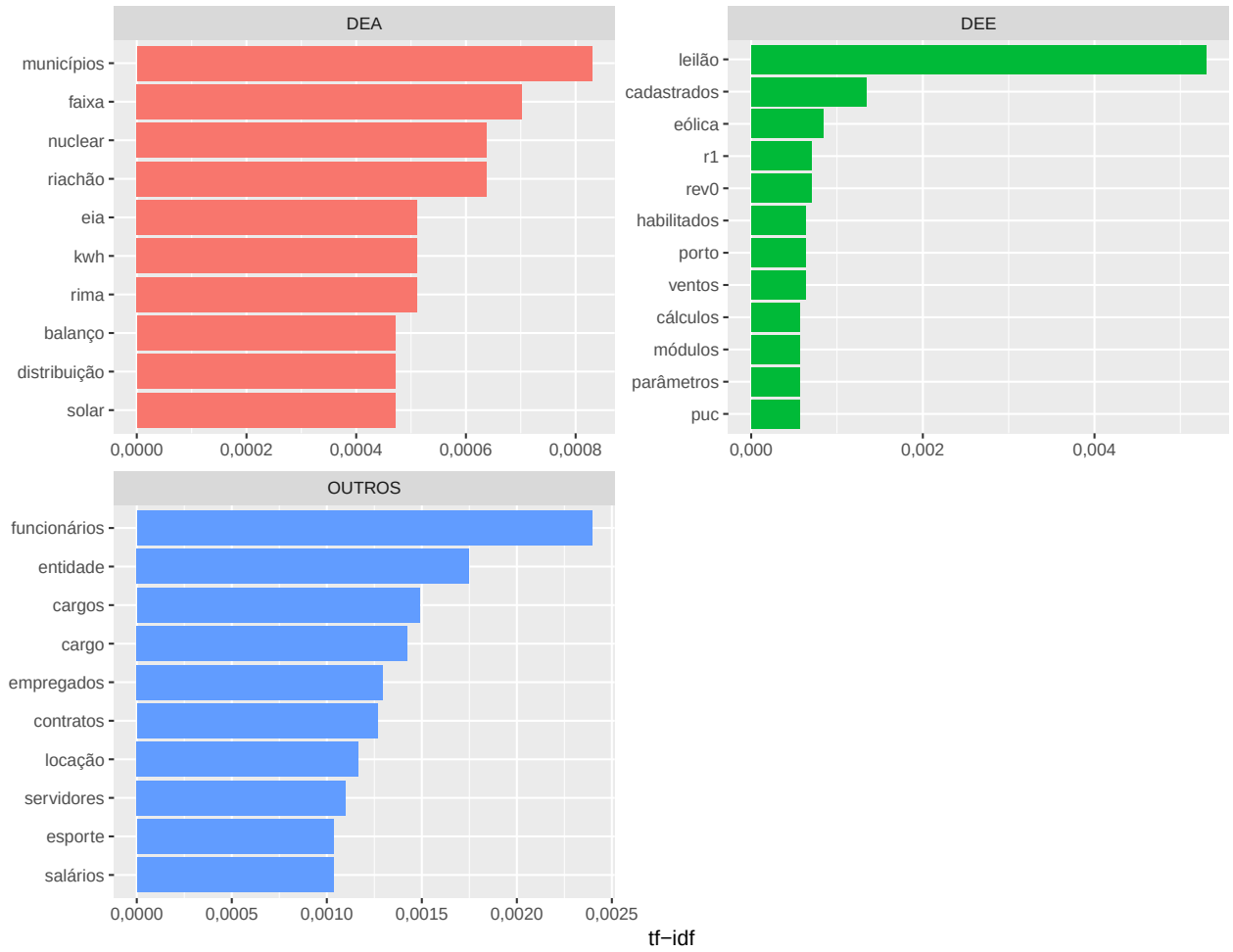
```
diretoria_palavras <- DB %>%
  unnest_tokens(palavra, PEDIDO) %>%
  count(DIRETORIA, palavra, sort = TRUE) %>%
  ungroup()
diretoria_palavras = left_join(diretoria_palavras, total_palavras, by = "DIRETORIA")
```

Table 12: Top 10 termos (DEA)

DIRETORIA	palavra	n	tf	idf	tf_idf
OUTROS	funcionários	37	0,00218	1,09861	0,00240
OUTROS	entidade	27	0,00159	1,09861	0,00175
OUTROS	cargos	23	0,00136	1,09861	0,00149
OUTROS	cargo	22	0,00130	1,09861	0,00142
OUTROS	empregados	20	0,00118	1,09861	0,00129
OUTROS	contratos	53	0,00312	0,40547	0,00127
OUTROS	locação	18	0,00106	1,09861	0,00117
OUTROS	servidores	17	0,00100	1,09861	0,00110
OUTROS	esporte	16	0,00094	1,09861	0,00104
OUTROS	salários	16	0,00094	1,09861	0,00104
OUTROS	concurso	43	0,00253	0,40547	0,00103
OUTROS	patrocínio	14	0,00082	1,09861	0,00091

Figura3: Top 10 termos por diretoria (ordenados pela estatística tf_idf e com stop words e sem stemming

```
plot_diretoria_palavras <- diretoria_palavras %>%
  bind_tf_idf(palavra, DIRETORIA, n) %>%
  arrange(desc(tf_idf)) %>%
  mutate(palavra = factor(palavra, levels = rev(unique(palavra)))) %>%
  mutate(DIRETORIA = factor(DIRETORIA, levels = c("DEA","DEE","OUTROS")))
#View(head(plot_diretoria_palavras))
#jpeg("02_freq_palavras_dir.jpeg")
plot_diretoria_palavras %>%
  group_by(DIRETORIA) %>%
  top_n(10, tf_idf) %>%
  ungroup() %>%
  mutate(palavra = reorder(palavra, tf_idf)) %>%
  ggplot(aes(palavra, tf_idf, fill = DIRETORIA)) +
  geom_col(show.legend = FALSE) +
  labs(x = NULL, y = "tf-idf") +
  facet_wrap(~DIRETORIA, ncol = 2, scales = "free") +
  coord_flip() +
  scale_y_continuous(labels=gcomma)
```




```
#dev.off()
```

Filtrando um pedaço de texto

```
DB %>%
  filter(str_detect(PEDIDO, "in")) %>%
  select(PEDIDO) %>%
  head()
```

```
## # A tibble: 6 x 1
##   PEDIDO
##   <chr>
## 1 "Empresa de Pesquisa Energética A Empresa de Pesquisa Energética (vincul~
## 2 Demanda ou carga Energética total (comercial, industrial, residencial e ~
## 3 "para EPE empresa de pesquisa energética - dados sobre custos de energia~
## 4 "Dados sobre Consumo de Energia por Unidades da Federação Boa tarde,\n\n~
## 5 Manifestação no processo de licenciamento da UHE Castanheira (processo n~
## 6 "Dados sobre quantidade de energia consumida e dinheiro pago pelo consum~
```

Uma limpeza removendo palavras sem significado semântico (**stop words**) pode auxiliar o algoritmo a retornar palavras ainda mais acertivas, bem como o tratamento de **stemming**, abordados a seguir.

Colocar tudo em minúsculo

```
DB$PEDIDO1 = tolower(DB$PEDIDO)
```

Stopwords

Com o arquivo de **stop words** , vamos remover as palavras sem sentido semântico

```
mystopwords <- data_frame(palavra = stopwords_pt)
for (j in 1:dim(DB)[1]) {
  for(i in 1:dim(mystopwords)[1]){
  stopw = as.character(mystopwords[i,1])
  DB$PEDIDO1[j] = gsub(paste0("\\ ",stopw," "), " ", as.character(DB$PEDIDO1[j]))
}
}
```

Ou simplesmente

```
mystopwords <- data_frame(palavra = stopwords_pt)
```

```
## Warning: `data_frame()` is deprecated, use `tibble()`.
## This warning is displayed once per session.
```

```
DB$PEDIDO1 <- removeWords(DB$PEDIDO1, mystopwords$palavra)
#View(head(DB))
```

Stemming

Podemos diminuir redundâncias por parte do algoritmo ensinando-o a compreender palavras que podem estar escritas de forma diferente mas que em significado semântico são semelhantes. Para isso, analisamos o radical de palavras com um mesmo prefixo mas com sufixos diferentes seja por quisistos como gênero ou plural.

Exemplos:

leilão $\propto$ leilões estado $\propto$ estados região $\propto$ regiões

Usando o pacote `ptstem`

```
library(ptstem)
temp_stem1 = proc.time()
stemming1 = ptstem(DB$PEDIDO1)
tempo_stem1 = proc.time() - temp_stem1
```

- Frequência de palavras por diretoria do stemming 1

```
diretoria_palavras_stem1 <- DB %>%
  mutate(PEDIDO1 = stemming1) %>%
  unnest_tokens(palavra, PEDIDO1) %>%
  count(palavra, sort = TRUE) %>%
  ungroup()
```

```
cat(paste0("Utilizando o algoritmo de stemming do pacote 'ptstem' o número de palavras chaves sem stemm
```

```
## Utilizando o algoritmo de stemming do pacote 'ptstem' o número de palavras chaves sem stemming reduz
```

Usando o pacote `rslp`

```
library(rslp)
temp_stem2 = proc.time()
```

```r
stemming2 = rslp(DB$PEDIDO1)
tempo_stem2 = proc.time() - temp_stem2
```

- Frequência de palavras por diretoria do stemming 2

```r
diretoria_palavras_stem2 <- DB %>%
  mutate(PEDIDO1 = stemming2) %>%
  unnest_tokens(palavra, PEDIDO1) %>%
  count(palavra, sort = TRUE) %>%
  ungroup()
```

```r
cat(paste0("Utilizando o algoritmo de stemming do pacote 'rslp' o número de palavras chaves sem stemmin
```

```
## Utilizando o algoritmo de stemming do pacote 'rslp' o número de palavras chaves sem stemming reduziu
```

Uma redução considerável no número de termos ocorreu ao usar o algoritmo `ptstem`, cerca de 61% de redução de termos versus 36% utilziando o algoritmo `rslp`, ou seja, o algoritmo ptstem foi mais eficiente na tarefa de agrupar os semelhantes (termos únicos).

Vale ressaltar, tabmbém, o tempo de processamento que ambos os algoritmos requerem.

```r
cat(paste0("O tempo de processamento do stemming rslp( ) foi de ",round(tempo_stem1[3],2), ' segundos d
```

```
## O tempo de processamento do stemming rslp( ) foi de 11.24 segundos decorridos.
```

```r
remove(tempo_stem1)
```

```r
cat(paste0("O tempo de processamento do stemming rslp( ) foi de ",round(tempo_stem2[3],2), ' segundos d
```

```
## O tempo de processamento do stemming rslp( ) foi de 0.82 segundos decorridos.
```

```r
remove(tempo_stem2)
```

O tempo decorrido para processamento do algoritmo do `ptstem` foi de aproximadamente 12 segundos versus 1 segundo decorrido para o processamento do algoritmo do `rslp`. Logo, o `rslp` é quase 12 vezes mais eficiente em termos de tempo de processamento. Além disso, o `rslp` remove acentuações e caracteres como "ç". Isso irá nos ajudar mais a frente quando utilizarmos os principais termos como variáveis binárias e preditoras do modelo de classificação.

Entretanto, o algoritmo mais lento, `ptstem`, foi mais interessante em termos de redução do número de termos únicos, cerca de 25% menos termos únicos em relação ao outro algoritmo. Além disso, por se tratar de uma base de dados relativamente pequena, 625 pedidos, e pouco mais de 4 mil termos únicos em todo o conjunto de texto, além disso vamos utilizar de um alto poder de processamento da máquina no referido estudo. Optamos, portanto, por utilizar ambos algoritmos. Vamos, primeiro, aplicar o removedor de sufixos da lingua portuguesa `rslp` seguido do `ptstem`.

Comparação do texto original c/ os 2 algoritmos e o final implementados após diferentes **stemmings**

```r
DB$PEDIDO[227]
```

```
## [1] "Destino dos honorários sucumbências Prezados, boa tarde. Desejo obter informações acerca da des
```

```r
#stemming1[227]
#stemming2[227]
DB$PEDIDO1[227]
```

```
## [1] "destino  honorários sucumbências ,  . desejo obter    destinação dada  honorários  sucumbência
```

```r
DB$PEDIDO[350]
```

```
## [1] "SOLICITAÇÃO DE NOTAS TÉCNICAS Prezados(as) Senhores(as),\n\nsou o Engº Eletricista Lidinei Serg
```

```r
#stemming1[350]
#stemming2[350]
DB$PEDIDO1[350]
```

```
## [1] "  notas técnicas () senhores(),\n\n  eng° eletricista lidinei sergio mesquita neri (ex-colaborad
```

```r
DB$PEDIDO[615]
```

```
## [1] "Gás do pré-sal Com fundamento na Lei 12.527/2011 (Lei de Acesso a\nInformações Públicas) venho a
```

```r
#stemming1[615]
#stemming2[615]
DB$PEDIDO1[615]
```

```
## [1] "gás  pré-sal  fundamento  lei 12.527/2011 (lei   \n públicas)   requerer  ,   20 dias corridos
```

```r
DB$PEDIDO[617]
```

```
## [1] "Questionamento sobre dados do PIB apresentados em Planos Decenais de Expansão Energética (PDEs)
```

```r
#stemming1[617]
#stemming2[617]
DB$PEDIDO1[617]
```

```
## [1] "questionamento    pib apresentados  planos decenais  expansão energética (pdes) publicados  .
```

```r
DB$PEDIDO[619]
```

```
## [1] "Dados distribuição de energia- UF Amapá Verifiquei que no Plano Decenal de Expansão de Energia
```

```r
#stemming1[619]
#stemming2[619]
DB$PEDIDO1[619]
```

```
## [1] " distribuição  - uf amapá verifiquei   plano decenal  expansão   (2006/ 2015)   tabela -25 - \"
```

Cria, antes, uma variáveil PEDIDO1 que repete os passos feitos aos termos quanto ao stemming so que no texto fonte.

```r
DB$PEDIDO1 = tolower(DB$PEDIDO)
mystopwords <- data_frame(palavra = stopwords_pt)
DB$PEDIDO1 <- removeWords(DB$PEDIDO1, mystopwords$palavra) # Remove Stop Words
DB$PEDIDO1 <- removePunctuation(DB$PEDIDO1) # Remove Punctuation

rm_accent <- function(str,pattern="all") {
   if(!is.character(str))
    str <- as.character(str)

  pattern <- unique(pattern)

  if(any(pattern=="Ç"))
    pattern[pattern=="Ç"] <- "ç"

  symbols <- c(
    acute = "áéíóúÁÉÍÓÚýÝ",
    grave = "àèìòùÀÈÌÒÙ",
    circunflex = "âêîôûÂÊÎÔÛ",
    tilde = "ãõÃÕñÑ",
    umlaut = "äëïöüÄËÏÖÜÿ",
    cedil = "çÇ"
```

```r
  )

  nudeSymbols <- c(
    acute = "aeiouAEIOUyY",
    grave = "aeiouAEIOU",
    circunflex = "aeiouAEIOU",
    tilde = "aoAOnN",
    umlaut = "aeiouAEIOUy",
    cedil = "cC"
  )

  accentTypes <- c("´","`","^","~","¨","ç")

  if(any(c("all","al","a","todos","t","to","tod","todo")%in%pattern)) # opcao retirar todos
    return(chartr(paste(symbols, collapse=""), paste(nudeSymbols, collapse=""), str))

  for(i in which(accentTypes%in%pattern))
    str <- chartr(symbols[i],nudeSymbols[i], str)

  return(str)
}


DB$PEDIDO1 <- rm_accent(DB$PEDIDO1) # Remove accent patterns
#View(head(DB))
#View(DB$PEDIDO1)
```

```r
#View(head(DB))
### CARACTERES
DB$PEDIDO1 = gsub("-"," ",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:.:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:,:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:':]"," ",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:!:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:?:]","",DB$PEDIDO1)
#DB$PEDIDO1 = gsub("[:-:]","_",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:_:]"," ",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:__:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:;:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:&:]"," ",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:/:]"," ",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:(:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:):]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:%:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:º:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:°:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:ª:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("\\d+","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[0-9]"," ",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:\n\t:]"," ",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:\t:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:\n:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("[:§:]","",DB$PEDIDO1)
DB$PEDIDO1 = gsub("\\s+"," ",DB$PEDIDO1)
```

```r
DB$PEDIDO1 = gsub("\"","  ",DB$PEDIDO1)


### STEMMINGS
#DB$PEDIDO1[143]
#DB$PEDIDOz = rslp(DB$PEDIDO1)
#DB$PEDIDOz = ptstem(DB$PEDIDO1, complete = FALSE)
DB$PEDIDO1 = ptstem(rslp(DB$PEDIDO1), complete = FALSE)
DB$PEDIDO1 = gsub("\\s+","  ",DB$PEDIDO1)
#teste1 = ptstem(rslp(DB$PEDIDO1), complete = FALSE)
#DB$PEDIDOz[143]
#DB$PEDIDOz[537]

## REMOVE STOP WORDS novamente
mystopwords <- data_frame(palavra = stopwords_pt)
DB$PEDIDO1 <- removeWords(DB$PEDIDO1, mystopwords$palavra)
```

```r
### PALAVRAS
#DB$PEDIDO1 =gsub("\\b(Leiloes)", "leilao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Leiloar)", "leilao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(leiloes)", "leilao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(leiloar)", "leilao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(leiloes)", "leilao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Energetica)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energetica)", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Eletricas)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(eletricas)", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Eletricos)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(eletricos)", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Eletrico)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(eletrico)", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Eletricidade)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(eletricidade)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energetica)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energeticas)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energetico)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energeticos)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energia)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energias)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energy)", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(energies)", "eletrica", DB$PEDIDO1)
x#DB$PEDIDO1 =gsub("\\b(Termoeletricas)", "termoeletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(termoeletricas)", "termoeletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Termeletrica)", "termoeletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(termeletrica)", "termoeletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Termeletricas)", "termoeletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(termeletricas)", "termoeletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Hidreletricas)", "hidreletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(hidreletricas)", "hidreletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Hidroeletricas)", "hidreletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(hidroeletricas)", "hidreletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Hidroeletricos)", "hidreletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(hidroeletricos)", "hidreletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Hidroeletrica)", "hidreletrica", DB$PEDIDO1)
```

```
DB$PEDIDO1 =gsub("\\b(hidroeletrica)", "hidreletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Administracao)", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(administracao)", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Administrativo)", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(administrativo)", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Administrativos)", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(administrativos)", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Administrativa)", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(administrativa)", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Administrativas)", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(administrativas)", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Consumo)", "consumo", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Consumidores)", "consumo", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(consumidores)", "consumo", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Consumidor)", "consumo", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(consumidor)", "consumo", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(Consumir)", "consumo", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\b(consumir)", "consumo", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\b(http)>", "", DB$PEDIDO1)
#View(DB$PEDIDO1)


DB$PEDIDO1 =gsub("\\ leiloes ", "leilao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ leiloar ", "leilao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ leiloes ", "leilao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Energetica ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energetica ", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Eletricas ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ eletricas ", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Eletricos ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ eletricos ", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Eletrico ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ eletrico ", "eletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Eletricidade ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ eletricidade ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energetica ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energeticas ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energetico ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energeticos ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energia ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energias ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energy ", "eletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ energies ", "eletrica", DB$PEDIDO1)
x#DB$PEDIDO1 =gsub("\\ Termoeletricas ", "termoeletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ termoeletricas ", "termoeletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Termeletrica ", "termoeletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ termeletrica ", "termoeletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Termeletricas ", "termoeletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ termeletricas ", "termoeletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Hidreletricas ", "hidreletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ hidreletricas ", "hidreletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Hidroeletricas ", "hidreletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ hidroeletricas ", "hidreletrica", DB$PEDIDO1)
```

```
#DB$PEDIDO1 =gsub("\\ Hidroeletricos ", "hidreletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ hidroeletricos ", "hidreletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Hidroeletrica ", "hidreletrica", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ hidroeletrica ", "hidreletrica", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Administracao ", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ administracao ", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Administrativo ", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ administrativo ", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Administrativos ", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ administrativos ", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Administrativa ", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ administrativa ", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Administrativas ", "administracao", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ administrativas ", "administracao", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Consumo ", "consumo", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Consumidores ", "consumo", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ consumidores ", "consumo", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Consumidor ", "consumo", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ consumidor ", "consumo", DB$PEDIDO1)
#DB$PEDIDO1 =gsub("\\ Consumir ", "consumo", DB$PEDIDO1)
DB$PEDIDO1 =gsub("\\ consumir ", "consumo", DB$PEDIDO1)
```

Frequência de palavras por diretoria

```
diretoria_palavras_stem3 <- DB %>%
  unnest_tokens(palavra, PEDIDO1) %>%
  count(DIRETORIA, palavra, sort = TRUE) %>%
  ungroup()
```

Stopwords

Com o arquivo de **stop words** previamente inserido vamos, primeiramente, transforma-lo em um data_frame a fim de futuramente utilizá-lo para extrair do texto palavras em comum.

Freq. de palavras sem stopwords por diretoria

```
mystopwords <- data_frame(palavra = stopwords_pt)
diretoria_palavras_noSTOP <- anti_join(diretoria_palavras_stem3, mystopwords,
                                       by = "palavra")
```

Filtrando um pedaço de texto

```
DB %>%
  filter(str_detect(PEDIDO1, "leiloes")) %>%
  select(PEDIDO1) %>%
  head()
```

Comparação do texto original c/ os 2 algoritmos e o final implementados após diferentes **stemmings**

```
DB$PEDIDO[227]
```

```
## [1] "Destino dos honorários sucumbências Prezados, boa tarde. Desejo obter informações acerca da des
```

```r
ptstem(DB$PEDIDO[227])
```

```
## [1] "Destino dos honorários sucumbências Prezados, boa tarde. Desejo obter informações acerca da des
```

```r
rslp(DB$PEDIDO[227])
```

```
## [1] "Destino dos honorarios sucumbencias Prezados, boa tarde. Desejo obter informacoes acerca da des
```

```r
DB$PEDIDO1[227]
```

```
## [1] "destin honora sucumbenc desej obt destinaca dad honora sucumbenc ambit empr repart advog carr i
```

```r
DB$PEDIDO[350]
```

```
## [1] "SOLICITAÇÃO DE NOTAS TÉCNICAS Prezados(as) Senhores(as),\n\nsou o Engº Eletricista Lidinei Serg
```

```r
ptstem(DB$PEDIDO[350])
```

```
## [1] "SOLICITAÇÃO DE NOTAS TÉCNICAS Prezados(as) Senhores(as),\n\nsou o Engº Eletricista Lidinei Serg
```

```r
rslp(DB$PEDIDO[350])
```

```
## [1] "SOLICITACAO DE NOTAS TECNICAS Prezados(as) Senhores(as),\n\nsou o Engº Eletricista Lidinei Serg
```

```r
DB$PEDIDO1[350]
```

```
## [1] " not tecn senh eng eletric lidin sergi mesquit ner excolabor eletrobr eletronucl gentil inform
```

```r
DB$PEDIDO[617]
```

```
## [1] "Questionamento sobre dados do PIB apresentados em Planos Decenais de Expansão Energética (PDEs)
```

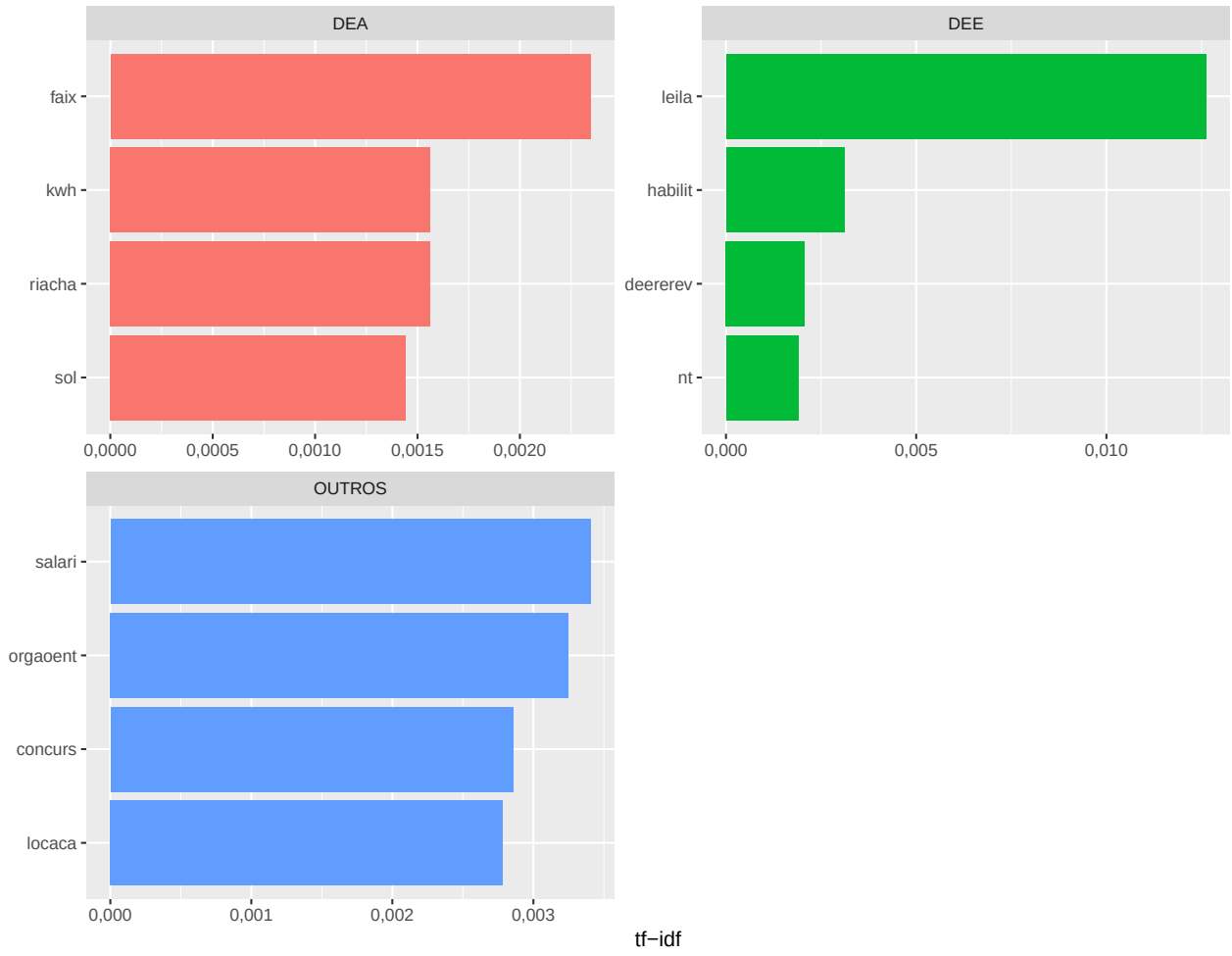
```r
DB$PEDIDO1[617]
```

```
## [1] "question pib apresent plan decen expansa energ pde public explicaco val tax cresc pib nacion ap
```

Figura4: Termos mais relevantes por diretoria pela estatística tf_idf, após stemming porém ainda com stop words

Vamos, agora, plotar as quinze palavras mais relevantes de acordo com a estatística **tf_idf**, por diretoria

```r
plot_diretoria_palavras_stem <- diretoria_palavras_stem3 %>%
  bind_tf_idf(palavra, DIRETORIA, n) %>%
  arrange(desc(tf_idf)) %>%
  mutate(palavra = factor(palavra, levels = rev(unique(palavra)))) %>%
  mutate(DIRETORIA = factor(DIRETORIA, levels = c("DEA","DEE","OUTROS")))
#View(head(plot_diretoria_palavras))
#jpeg("02_freq_palavras_dir.jpeg")
plot_diretoria_palavras_stem %>%
  group_by(DIRETORIA) %>%
  top_n(4, tf_idf) %>%
  ungroup() %>%
  mutate(palavra = reorder(palavra, tf_idf)) %>%
  ggplot(aes(palavra, tf_idf, fill = DIRETORIA)) +
  geom_col(show.legend = FALSE) +
  labs(x = NULL, y = "tf-idf") +
  facet_wrap(~DIRETORIA, ncol = 2, scales = "free") +
  coord_flip() +
  scale_y_continuous(labels=gcomma)
```



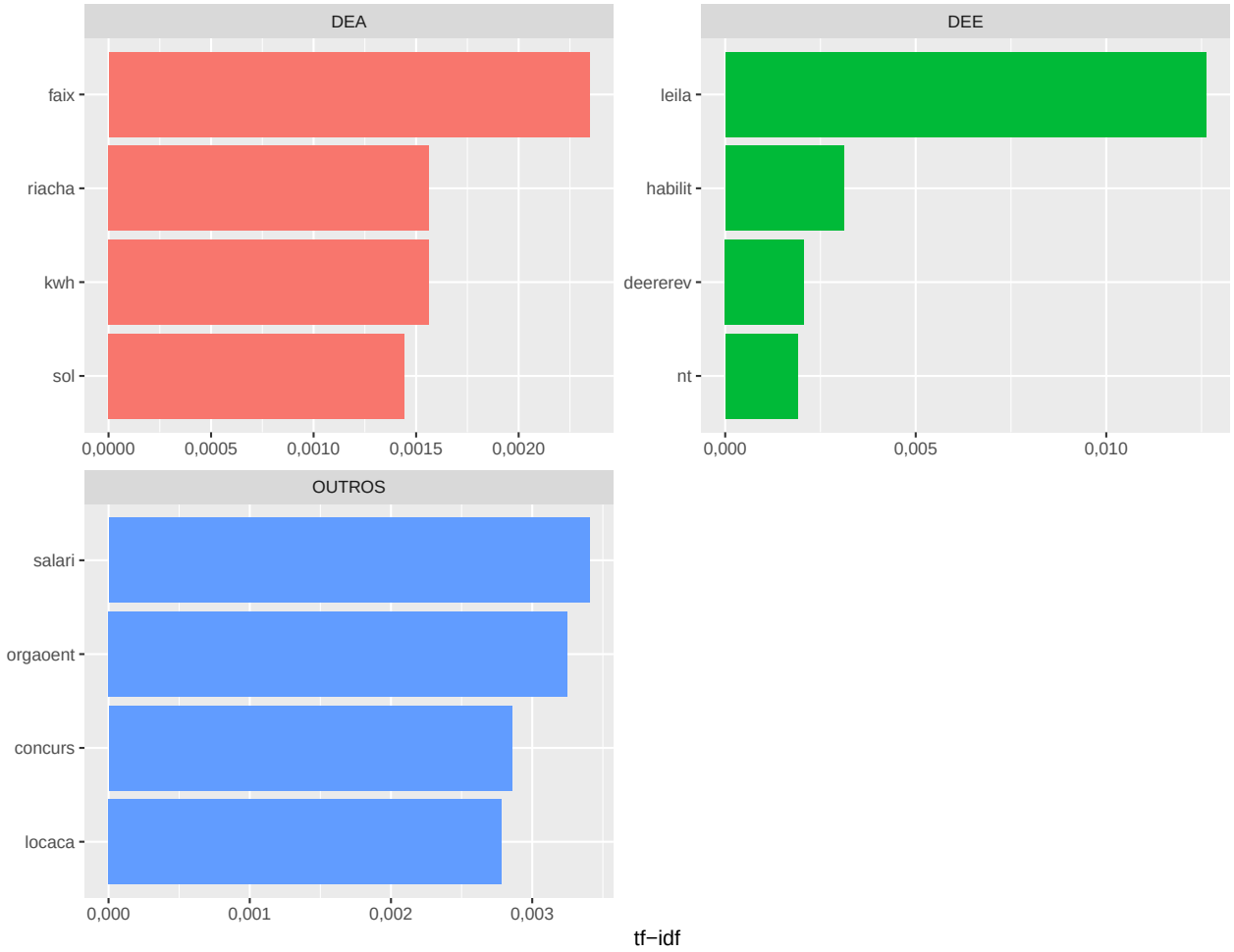

```
#dev.off()
```

Figura5: Termos mais relevantes por diretoria pela estatística tf_idf, após stemming e sem stop words

Sim, a remoção de **stop words** não alterou em nada a ordem das 4 palavras mais relevantes de acordo com a estatística.

```
#diretoria_palavras_noSTOP_noSTOP
plot_diretoria_palavras_noSTOP <- diretoria_palavras_noSTOP %>%
  bind_tf_idf(palavra, DIRETORIA, n) %>%
  arrange(desc(tf_idf)) %>%
  mutate(word = factor(palavra, levels = rev(unique(palavra)))) %>%
  mutate(DIRETORIA = factor(DIRETORIA, levels = c("DEA","DEE","OUTROS")))
#plot_diretoria_palavras_noSTOP
#windows.options(width=10, height=10)
#jpeg("03_freq_palavras_dir_nostop.jpeg")
plot_diretoria_palavras_noSTOP %>%
  group_by(DIRETORIA) %>%
  top_n(4, tf_idf) %>%
  ungroup() %>%
  mutate(palavra = reorder(palavra, tf_idf)) %>%
  ggplot(aes(palavra, tf_idf, fill = DIRETORIA)) +
```

```
  geom_col(show.legend = FALSE) +
  labs(x = NULL, y = "tf-idf") +
  facet_wrap(~DIRETORIA, ncol = 2, scales = "free") +
  coord_flip() +
  scale_y_continuous(labels=gcomma)
```



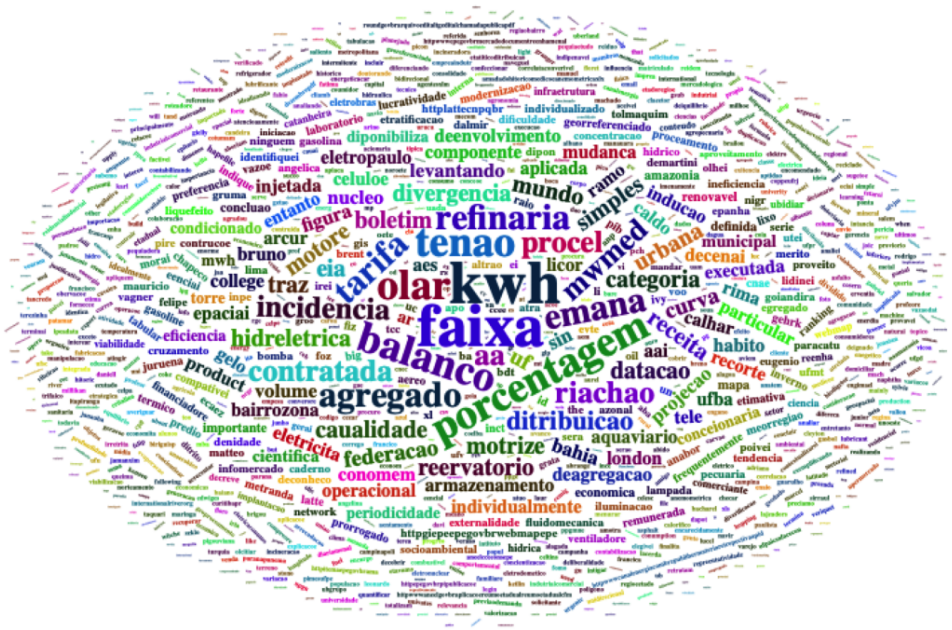

```
#dev.off()
```

Wordcloud2 - DEE

```
set.seed(6423)
plot_diretoria_palavras_noSTOP %>%
  filter(DIRETORIA == "DEE") %>%
  select(word = palavra, freq = tf_idf) %>%
  mutate(word = as.factor(word)) %>%
  #top_n(150, freq) %>%
  as.data.frame() %>%
  wordcloud2(shuffle = TRUE, color = "random-dark", shape = "circle", size = 1.10)
```

Wordcloud2 - DEA

```
set.seed(6423)
plot_diretoria_palavras_noSTOP %>%
  filter(DIRETORIA == "DEA") %>%
  select(word = palavra, freq = tf_idf) %>%
  mutate(word = as.factor(word)) %>%
  #top_n(150, freq) %>%
  as.data.frame() %>%
  wordcloud2(shuffle = TRUE, color = "random-dark", shape = "circle", size = .25)
```



Wordcloud2 - OUTROS

```
set.seed(6423)
plot_diretoria_palavras_noSTOP %>%
  filter(DIRETORIA == "OUTROS") %>%
  select(word = palavra, freq = tf_idf) %>%
  mutate(word = as.factor(word)) %>%
  #top_n(150, freq) %>%
  as.data.frame() %>%
  wordcloud2(shuffle = TRUE, color = "random-dark", shape = "circle", size = 0.35)
```

Comparação de frequências dois a dois (sem stopwords e com stemming)

Vamos agora comparar a frequência de palavras entre diretorias. Antes disso, vamos criar documentos de texto no formato tidy separadamente para cada uma das 3 cateorias: *DEA*, *DEE* e *OUTROS*.

```
{r, child="03_1_compara2a2.Rmd"}
```

```
{r child = '032_textminingpart2.Rmd'}
```

MODELAGEM - APLICAÇÃO E RESULTADOS

Preparação e partição de dados

Recapitulando, chegamos portanto, a uma base de dados donde foram aplicadas 2 diferentes técnicas de **stemming**, também a remoção de **stopwords** e fazendo uso da estatística **tf_idf** a fim de ressaltar os termos mais relevantes de cada documento de texto.

Vamos, portanto, contar o número de termos únicos dentro de cada diretoria.

Tabela11: Número de termos únicos por diretoria

```
key_DIR = plot_diretoria_palavras_noSTOP %>%
  group_by(DIRETORIA) %>%
  count(DIRETORIA)
 key_DIR %>%
   kable("latex", caption = "Número de termos por diretoria",
         booktabs = T, format.args = list(decimal.mark = ',', big.mark = ".")) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 13: Número de termos por diretoria

DIRETORIA	n
DEA	1.556
DEE	1.551
OUTROS	1.622

Vamos, agora, selecionar as $n = 250$ palavras mais importantes de cada uma das 4 diretorias e da categoria 'OUTROS'. Para isso vamos, primeiro, separar os documentos em 3 documentos distintos, um para cada diretoria.

```
#n=1500
#n=500
#n=500

termos_dir_DEE =
plot_diretoria_palavras_noSTOP %>%
filter(DIRETORIA == "DEE")
termos_DEE = termos_dir_DEE #%>%top_n(n, tf_idf)

termos_dir_DEA =
plot_diretoria_palavras_noSTOP %>%
filter(DIRETORIA == "DEA")
termos_DEA = termos_dir_DEA #%>%top_n(n, tf_idf)

termos_dir_OUTROS =
plot_diretoria_palavras_noSTOP %>%
filter(DIRETORIA == "OUTROS")
termos_OUTROS = termos_dir_OUTROS #%>%top_n(n*3, tf_idf)

termos_dir = bind_rows(mutate(termos_DEE, DIRETORIA = "DEE"),
                        mutate(termos_DEA, DIRETORIA = "DEA"),
                        mutate(termos_OUTROS, DIRETORIA = "OUTROS")) %>%
  select(palavra) %>%
  unique()
```

```
gg <- termos_dir$palavra
gg <- unique(gg)
fe <- matrix(data = 0, nrow = length(DB$PEDIDO1), ncol = length(gg))
fe <- data.frame(fe); colnames(fe) <- gg
i=j=0
for(i in 1:length(DB$Protocolo)){
        for(j in 1:length(gg)){
                g <- grepl(gg[j], DB$PEDIDO1[i])
                if(g == TRUE){
                        fe[i, j] <- 1
```

```
                }
        }
}

#sum(rowSums(fe))
dim(fe)
```

```
## [1]  624 3032
```

```
#colSums(fe)
cat(paste0("Existem ", dim(fe)[2], " termos/palavras-chaves únicas na matriz em questão."))
```

```
## Existem 3032 termos/palavras-chaves únicas na matriz em questão.
```

```
NumTermos = as_tibble(rbind(apply(fe,2,sum)))
NumTermos = gather(NumTermos, key = "termo", value = "Num_Pedidos")
NumTermos = NumTermos[order(NumTermos$Num_Pedidos, decreasing = TRUE), ]
#View(colSums(fe))
#View(colnames(fe))
```

Vamos excluir alguns termos com pouca frequência, abaixo ou iguais a 2 (mediana)

```
mediana1 = summary(colSums(fe))[3]
 #removing unecessary terms
    exclui_termos <- as.character(c())
     cbind(Termos = colnames(fe), Freq_Termos = colSums(fe))
```

```
##
## leila
## habilit
## deererev
## nt
## parametr
## ama
## modul
## cvu
## termoeletr
## sol
## hidreletr
## ahe
## deficit
## limit
## habilitaca
## anemometr
## rs
## vent
## ccpectet
## conex
## eol
## gt
## mau
## mon
## monitor
## pucri
## rap
## sin
```

```
## subestaco
## medico
## reserv
## barr
## bifac
## climatolog
## confi
## deedearerev
## deentrev
## furn
## neg
## nodal
## piau
## precostet
## prt
## pv
## reproduz
## suap
## tust
## tv
## unita
## art
## ler
## memor
## particip
## venc
## iii
## ofici
## torr
## antoni
## cec
## ceg
## cgh
## cmo
## cms
## consequenc
## dinh
## disposi
## etap
## fever
## gna
## hibr
## hom
## imetam
## instruco
## itaipu
## luiz
## maranh
## ns
## offshor
## pne
## presum
## reforc
## rod
```

```
## supracit
## tet
## vian
## vnr
## cme
## interlig
## kv
## medica
## outorg
## reposica
## veloc
## acr
## anared
## import
## instalaca
## pch
## pic
## supr
## altern
## bel
## ce
## conclu
## distribuica
## fich
## mwh
## regim
## acrerondon
## acu
## aeroger
## amap
## aneelleila
## anemograf
## ansi
## antecip
## aport
## atlan
## ave
## bat
## brizon
## cedr
## chapeub
## compromiss
## comprovaca
## contribuint
## correspondenc
## csp
## cteep
## cvp
## dec
## deck
## deeptr
## delr
## deslig
## determinaco
```

```
## disput
## distint
## dro
## elenc
## encontral
## especificaca
## exig
## exigenc
## fabbr
## fac
## ganh
## guacu
## hidraulic
## hidroeletr
## hidrogeraca
## httpwwwgovbrtransmissaopaginasdadosparaestudosdeplanejamentodatransmisscaopdeaspx
## igap
## incentiv
## infom
## ingl
## interligaca
## intermitenc
## inventa
## irma
## jandairarn
## lavr
## lelio
## mgw
## minuci
## mitig
## morr
## mossor
## nac
## ndpespemm
## obtev
## onshor
## onsr
## paragu
## participaco
## pd
## permanent
## peruib
## petpelp
## prerequisit
## propriedad
## pucmin
## recebel
## recuperaca
## reep
## refrigeraca
## regaseificaca
## relicit
## renat
## replic
```

```
## revitalizaca
## rondon
## sav
## sgtaneel
## sisorh
## solarimetr
## subestaca
## subterrane
## sudo
## tapaj
## tecnicoeconom
## thiag
## transmis
## trombet
## tronc
## uirapuru
## unesp
## uniform
## utegnsp
## utilz
## volg
## vr
## cab
## digit
## divulgaca
## documentaca
## extensa
## intermedi
## matema
## mdi
## nenhum
## palm
## parec
## precotet
## rora
## semestr
## sergip
## siti
## tes
## verd
## viabil
## abastec
## acredit
## agost
## celg
## cheg
## cientif
## co
## cost
## defin
## descrit
## estaca
## estrut
## federac
```

```
## it
## joa
## nom
## paragraf
## posi
## pr
## propost
## proveni
## softw
## ufv
## abril
## acresc
## alteraca
## aproveit
## cce
## cepel
## chesf
## compreensa
## coqu
## decret
## deentr
## desloc
## destac
## dispost
## emit
## escal
## estaco
## estagi
## evt
## explicit
## gros
## gw
## importanc
## instalaco
## intercambi
## lot
## lt
## ltd
## palestr
## parcel
## pg
## po
## pratic
## questo
## regulament
## respons
## shap
## simulaco
## soluca
## subsistem
## sud
## tarif
## tocantim
## abengo
```

```
## abrac
## absorca
## abun
## acces
## aceita
## adem
## aeg
## aep
## aliquotabas
## alteraco
## alumin
## amorf
## ampliaca
## aneelanatel
## anteced
## aparent
## applied
## aprendiz
## apresentaco
## aproximaca
## apt
## aracat
## aracatub
## aripu
## arranj
## assum
## atra
## atrat
## atribuica
## attl
## austr
## autotransform
## bag
## barcel
## basei
## bauru
## biolog
## boj
## borb
## cajuru
## camarg
## camocimc
## cano
## capivaricacho
## caracterizaca
## carbonit
## carolin
## carvaa
## castr
## cavac
## cee
## cel
## cemiggt
## chav
```

```
## cho
## chu
## cient
## clic
## clovil
## cnu
## colisa
## colomb
## coment
## comig
## comparec
## competi
## comunita
## comwh
## concluso
## concret
## condiz
## condu
## conec
## conexo
## confeccion
## consorci
## contingenc
## copcec
## copelgt
## copp
## coppead
## coronel
## correca
## corrig
## coxip
## crei
## ctapr
## cuiab
## curitib
## dal
## declar
## deeit
## deerer
## defas
## deliber
## deliberaca
## descont
## descri
## desss
## deten
## devera
## dezen
## diferenc
## dificil
## dificult
## dimensa
## diog
## diogosignorgmail
```

```
## diox
## dissertacaa
## dist
## distribuicaotransmissa
## doc
## docu
## domici
## donwload
## doument
## ebrasil
## ebtt
## edrr
## educ
## efetu
## elaboracaa
## elasticidaderend
## eletromagne
## eletronort
## eletrosul
## empreed
## en
## ena
## enacel
## enfrent
## engi
## enrroc
## enterr
## enumer
## enunci
## eolicofotovolta
## ere
## erval
## espirit
## estend
## estiv
## estreit
## esutd
## exc
## exceca
## exp
## expanca
## explor
## facult
## fal
## falenc
## far
## ferr
## fh
## fici
## figueired
## fixaca
## florianopolil
## formul
## franc
```

```
## francil
## ftp
## gafanhot
## gave
## gcp
## gd
## ge
## gen
## genpow
## georeferenc
## geron
## getn
## gfil
## gg
## go
## googl
## gov
## graduand
## grav
## graziell
## grid
## gusm
## hidreledr
## hidrolog
## historioc
## httpgovbrsitesptpublicacoesabertospublicacoespublicacoesarquivospublicacaodeeitaverscaofinalpdf
## httponsorgbrptpaginasresultadosoperacaohistoricooperaca
## httpwwwcceeorgbrportalfacesoquefazemosmenulateralleiloesadfctrlstatecpigtknafrloopfafrloopdadfctrlsta
## httpwwwgovbrleiloesdocumentsleilcbesdeenergiadereservacbalercadastradospdf
## idiom
## ifsul
## igpd
## igual
## imediat
## imperatriz
## imprens
## imprensalightbr
## inat
## incentivoobrigaca
## incompati
## informac
## inglesfiqu
## iniic
## inmet
## insum
## intens
## interromp
## invaria
## irradiaca
## irrigaca
## itapipoc
## itaqu
## ituting
## jeann
```

```
## joanneum
## joas
## jup
## kay
## kelman
## kwmw
## leia
## len
## lest
## liberal
## licenci
## lilian
## log
## londrin
## mad
## manipulal
## manobra
## mantovilil
## marin
## marmel
## marqu
## martim
## mascarenh
## massayosh
## mctic
## mecanc
## meteorolog
## metereolog
## migraca
## mim
## mistur
## modulosinver
## monocristalin
## mora
## moura
## msctech
## multipl
## mv
## mva
## mwm
## mwp
## narr
## neblin
## obrig
## observanc
## obteca
## obtel
## ocultaca
## oner
## onsdt
## out
## padra
## palavr
## pali
```

```
## paraben
## parad
## parigot
## paril
## patric
## paulin
## pbc
## pedil
## pelot
## perc
## perfeit
## permanenc
## perme
## peron
## pet
## petr
## pibit
## pinh
## plant
## play
## pleite
## policristalin
## pouc
## pow
## prai
## preceit
## preez
## prem
## pressa
## print
## produ
## programaca
## propic
## prospecco
## prospect
## prosper
## proteca
## pusch
## pwf
## qtd
## qualific
## qued
## quix
## ra
## rafael
## rastre
## rd
## rea
## realocaca
## recomendaco
## reconstru
## redaca
## ree
## refent
```

```
## regulamentaco
## relacional
## relocaca
## repet
## repr
## reproduzindoatualiz
## requis
## rest
## retific
## reversi
## reviso
## rit
## rosan
## rosangel
## rot
## rr
## rua
## russ
## sagr
## saik
## salin
## salt
## sazonal
## scienc
## screen
## sctaneel
## seccion
## secunda
## serr
## servira
## sign
## silici
## simul
## sincer
## situaco
## sofr
## soliciti
## solt
## somato
## stat
## superaca
## superfici
## supramencion
## sustenta
## taref
## tarj
## tecnic
## tecnog
## ted
## telecomunicac
## telfax
## termomaranha
## traf
## trair
```

```
## trajet
## transcrit
## transform
## transformaca
## transito
## transmit
## tributaca
## trimestr
## tubara
## tvpuc
## uberab
## uff
## umid
## umidad
## unid
## unisin
## upload
## urani
## usmw
## utv
## vaness
## velh
## verificaca
## versu
## vicent
## vicereit
## vistainteri
## vontad
## zao
## adot
## adquir
## afirm
## altur
## ana
## andr
## aplica
## apont
## associ
## ata
## atl
## atualizaca
## automa
## bah
## base
## ben
## capitul
## carv
## castanh
## cemig
## cesp
## chuv
## codig
## comercializ
## comunic
```

```
## concessa
## conclusa
## concurs
## condicion
## condico
## conect
## confirm
## cons
## conservaca
## consolid
## constituica
## curv
## deede
## denomin
## direca
## divisa
## eletrobr
## energis
## entend
## err
## estatut
## gastrading
## goi
## imped
## kw
## leit
## licitaca
## luc
## manutenca
## met
## metod
## minim
## mwmed
## ne
## negoci
## newav
## obtiv
## oliv
## pa
## par
## pedr
## perfil
## portug
## posteri
## produt
## profes
## qualificaca
## raza
## recomendaca
## region
## relev
## renova
## resum
## rural
```

```
## signific
## sirv
## subgrup
## suport
## telefon
## temperat
## th
## titul
## uf
## universita
## veicul
## verifiq
## vigent
## vii
## visa
## volum
## workshop
## xxxii
## ab
## aba
## abord
## acat
## ader
## adoca
## agrup
## alex
## alinh
## aloc
## an
## antig
## aparelh
## apres
## apresentaca
## aprofund
## apuraca
## ar
## arin
## assegur
## assist
## atmosfer
## audienc
## automaca
## baliz
## barb
## benchmarking
## boletim
## bols
## californ
## cam
## campu
## candidat
## carbon
## carol
## cart
```

```
## ced
## celp
## centroo
## certam
## cienc
## clim
## cnpq
## coelb
## coelc
## combin
## comissa
## compo
## configuraca
## consequ
## constat
## contabil
## convoc
## copel
## corr
## cosern
## ct
## cult
## dad
## dea
## decisa
## deent
## def
## delim
## desempenh
## dispo
## dispos
## doutorand
## edica
## edita
## efetiv
## eficienc
## eng
## ent
## equlibr
## escolh
## essenc
## estatal
## estoqu
## eventual
## exclusiv
## express
## extr
## extra
## extraca
## fatur
## financ
## firm
## fix
## fomal
```

```
## fun
## fundament
## gwm
## hidr
## ibam
## ilh
## implement
## indenizaca
## indicaca
## iniciaca
## inscrica
## int
## integraca
## isol
## isoladomw
## ivan
## jardim
## km
## laud
## lembr
## lenh
## lev
## liber
## licit
## liv
## lixiv
## lucen
## luz
## mach
## manau
## mand
## mater
## mesquit
## minut
## mov
## ms
## nasc
## nel
## net
## noro
## nov
## nucle
## nul
## ob
## obrigato
## obrigatoriedad
## ocorrenc
## oest
## of
## opca
## opinia
## ora
## orientaca
## otim
```

```
## our
## ouv
## pacienc
## pais
## part
## patamar
## percent
## permanec
## pertin
## pes
## pleit
## poli
## pont
## posgraduaca
## pragma
## prefer
## premiss
## prime
## problem
## prop
## prud
## publicaco
## quinzen
## ramal
## rapid
## reduca
## reduz
## referenci
## reiter
## rel
## ren
## requisit
## respectiv
## restrico
## reunio
## rev
## ric
## rj
## rn
## sec
## seca
## secret
## selec
## semina
## sergi
## setor
## setorcomerc
## sic
## silv
## simplific
## simultane
## situ
## submercadonort
## subsidi
```

```
## sudestec
## sufici
## suger
## sulgip
## tom
## tr
## trech
## triangul
## tribut
## txt
## ufrj
## uftm
## ultrassecret
## university
## usuari
## vaza
## verif
## vi
## vigenc
## vitor
## eletr
## estud
## tecn
## eolic
## document
## relat
## empr
## empreend
## projet
## not
## usin
## leilo
## cop
## refer
## utiliz
## transmissa
## calcul
## gerac
## realiz
## potenc
## font
## pesquil
## cadastr
## expansa
## pde
## energ
## demand
## ba
## consum
## regia
## cust
## apresent
## analis
## ga
```

```
## garant
## inform
## parqu
## encontr
## atend
## exist
## quant
## sit
## anex
## consider
## seri
## ute
## disponivel
## fisic
## ped
## precis
## rio
## fotovolta
## metodolog
## obt
## referenc
## regio
## capac
## detalh
## lei
## algum
## atenci
## gentil
## mens
## nacion
## nord
## plan
## planej
## process
## seguint
## set
## desagreg
## natur
## progr
## arqu
## biomass
## desenvolv
## dispon
## disponibilizaca
## etc
## linh
## necessit
## period
## prec
## pres
## sul
## trat
## val
## combusti
```

```
## ger
## ii
## list
## mme
## model
## mont
## om
## port
## previs
## senh
## tabel
## term
## uhe
## basic
## envi
## esco
## instal
## operac
## produca
## atual
## const
## consult
## est
## fornec
## horari
## inclu
## merc
## municipi
## public
## red
## ult
## aneel
## atenca
## att
## carg
## despach
## histor
## invest
## med
## min
## respeit
## sant
## anual
## certificaca
## contat
## gwh
## planilh
## result
## revisa
## total
## abaix
## academ
## artig
## cenari
```

```
## distribu
## engenh
## excel
## flux
## graf
## industr
## link
## livr
## long
## marg
## nort
## receb
## relacion
## renov
## univ
## vist
## acim
## acord
## agent
## aguard
## anteri
## cicl
## crite
## elabor
## eletric
## encaminh
## entr
## espec
## fat
## geograf
## horizont
## marc
## medi
## mw
## necess
## novembr
## obr
## participaca
## pec
## pergunt
## recent
## registr
## regul
## respost
## retorn
## sid
## sistem
## solicit
## tecnolog
## unidad
## uso
## ajud
## avaliaca
## camp
```

```
## cit
## colet
## consegu
## construca
## cont
## contrat
## contratac
## cur
## decen
## di
## diss
## econom
## emissa
## especific
## fed
## feit
## impact
## jan
## localizaca
## mat
## matriz
## mostr
## obje
## ole
## org
## pass
## paul
## pod
## proced
## publ
## realizaca
## ref
## seg
## termeletr
## acompanh
## agradec
## alun
## ambit
## and
## assunt
## auxili
## canal
## compr
## consid
## coorden
## definica
## dev
## diesel
## duvid
## eletron
## equip
## esclarec
## graduaca
## indic
```

```
## inici
## licenc
## lucr
## mai
## ministe
## necessa
## ola
## pag
## pal
## poss
## produz
## real
## requ
## sc
## sigil
## utilizaca
## vend
## acess
## agu
## ambient
## antecipad
## atenc
## bac
## car
## classific
## correspond
## diari
## elaboraca
## especif
## estabelec
## inter
## interess
## iv
## man
## melhor
## mencion
## ocorr
## orig
## possu
## pra
## praz
## prev
## prez
## princip
## publicaca
## pud
## recomend
## report
## respec
## setembr
## unic
## vers
## adic
## ambi
```

```
## aprov
## ativ
## baix
## banc
## brut
## catarin
## centr
## ch
## cidad
## compar
## compreend
## conced
## contribu
## convers
## cord
## daniel
## dat
## dentr
## descrev
## desej
## dezembr
## diretriz
## edit
## envolv
## escrev
## esper
## estad
## estatil
## estim
## event
## fic
## funcion
## junt
## legal
## legislaca
## mail
## maquin
## observ
## oferec
## ofert
## outubr
## pagin
## previst
## resid
## respond
## responsa
## sent
## serv
## shapefil
## soc
## trabalh
## us
## var
## vis
```

```
## abert
## alt
## amazon
## anuari
## aspect
## branc
## cl
## comp
## compartilh
## composica
## conhec
## conjunt
## consig
## difer
## diret
## embas
## enderec
## evoluca
## exportaca
## fabric
## func
## funca
## hav
## identific
## identificaca
## implementaca
## impost
## incis
## integr
## julh
## justific
## localiz
## mape
## mod
## modal
## nega
## oper
## orc
## pertenc
## pesso
## petrole
## plataform
## post
## preench
## presid
## prest
## pret
## procur
## rend
## reunia
## segment
## segu
## text
## tir
```

```
## transparenc
## turbin
## varia
## verific
## via
## abr
## aceit
## adequaca
## administr
## alcanc
## aplicaca
## aquisica
## are
## atuaca
## aument
## autor
## autorizaca
## avali
## bagac
## benefici
## biog
## busc
## can
## capit
## caracteris
## carat
## cativ
## cerc
## cert
## cham
## classificaca
## cnpj
## colaborac
## comec
## comerc
## companh
## computac
## concession
## contempl
## context
## contribuico
## convenc
## corret
## cresc
## cri
## depend
## deriv
## determin
## discrimin
## disp
## disponibiliz
## divulg
## entreg
## escrit
```

```
## especi
## estrutur
## exempl
## expost
## facil
## falt
## feir
## foc
## form
## futur
## gast
## govern
## grau
## hipotes
## implantaca
## inflexibil
## institu
## internac
## internet
## interval
## intuit
## kg
## levant
## liqu
## manifestaca
## mant
## mar
## max
## mestr
## monograf
## mund
## norm
## notic
## numer
## obtenca
## ofic
## onlin
## ord
## orient
## otimizaca
## papel
## paraib
## pdf
## perceb
## perd
## permit
## pretend
## projeco
## protocol
## quebr
## question
## questiona
## recurs
## represent
```

```
## ressalt
## restrica
## seguranc
## separ
## simpl
## sintes
## superi
## tax
## tend
## torn
## transport
## faix
## kwh
## riacha
## balanc
## porcent
## agreg
## incidenc
## procel
## refin
## seman
## celulos
## eiar
## reservato
## projeca
## aquavia
## bairr
## causal
## comerci
## compon
## consom
## dataca
## desagregaca
## divergenc
## espac
## format
## hor
## iluminaca
## injet
## lic
## motriz
## traz
## figur
## municip
## francisc
## aai
## ae
## arcur
## armazen
## cadern
## cald
## calh
## cna
## colleg
```

```
## contruco
## dens
## eia
## eletropaul
## entant
## execut
## external
## felip
## gel
## georreferenci
## habit
## individ
## induca
## inferi
## juruen
## london
## micr
## mudanc
## particul
## pib
## pir
## rai
## ram
## recort
## rim
## som
## ufb
## vazo
## vera
## xl
## map
## periodic
## webmap
## categor
## distrit
## infraestrut
## resenh
## tel
## adri
## aflu
## alago
## alban
## altra
## anab
## anat
## angel
## angelin
## aplicaco
## arcondicion
## bar
## bdt
## big
## biodiges
## boc
```

```
## bomb
## calorif
## campanh
## chapec
## compati
## concentraca
## concesso
## cris
## cruz
## dalm
## demartin
## desconhec
## dificuldad
## domicili
## enerd
## entretant
## escass
## espanh
## estratificaca
## eugeni
## financi
## fluidomecan
## fornecel
## foz
## frequ
## gehrk
## geod
## georeferenci
## gil
## grum
## grup
## horosazon
## httplattescnpqbr
## httpsgisepegovbrwebmapep
## identifiq
## inct
## indiqu
## individu
## ineficienc
## infomerc
## inp
## inst
## invern
## ire
## ivy
## laborato
## lal
## lamp
## latt
## li
## lidin
## lim
## liquefeit
## lix
```

```
## lucrat
## manuel
## matogross
## matte
## merit
## mesorregia
## mestrand
## metropolit
## miner
## modernizaca
## moral
## mt
## ner
## network
## nigr
## ning
## oil
## olh
## paracatu
## paranapanem
## pecu
## pernambuc
## populac
## possi
## preferenc
## prorrog
## proveit
## ranking
## regin
## remuner
## ribeira
## segreg
## selecion
## socioambient
## sucess
## sugesto
## supply
## tabul
## tcc
## temp
## tendenc
## test
## the
## to
## tolmasquim
## uc
## ufmt
## uno
## urgenc
## urucucoarimanau
## utel
## vagn
## ventil
## aplic
```

```
## hidraul
## intern
## receit
## unia
## anp
## brun
## constru
## conteud
## csv
## divid
## gasolin
## ideal
## junh
## motor
## negr
## product
## sociedad
## urban
## acion
## aco
## aere
## cas
## client
## cobr
## destin
## efeit
## existenc
## gasodut
## gesta
## glp
## imens
## institut
## operaco
## pemat
## predi
## proc
## residenc
## sab
## territo
## tip
## tipic
## util
## about
## abrang
## acessi
## acrescent
## adicion
## agentesxlsm
## agrad
## agregaca
## agricol
## agronom
## agropecu
## alag
```

```
## amadadoshistoricosmedicoesanemometricasxlsx
## amand
## amig
## amostr
## anaerob
## aneelcceeom
## antecedenc
## apiac
## apliqu
## apot
## aptida
## arrecadaca
## asa
## asfalt
## asphalt
## assent
## at
## atencioas
## ating
## atrel
## aurel
## availabl
## avanc
## averigu
## azul
## bacharel
## baian
## barril
## bell
## bidirec
## biriguisp
## bren
## brent
## brentinternationalriversorg
## brig
## build
## but
## cal
## campin
## campinapolil
## canalenerg
## cand
## capt
## caus
## cav
## cc
## ceb
## celesc
## celtim
## cez
## chec
## classese
## cleyton
## cliamb
```

```
## cln
## cnec
## cok
## colun
## comportament
## compres
## comunidad
## conceitu
## confecc
## conscien
## conscientizaca
## consideraca
## consumidoresse
## consumption
## contabiliz
## contabilizaca
## contrast
## conver
## coppeufrj
## corre
## correlaca
## correlataconversi
## correntin
## costum
## cosum
## cpflpaul
## cristalin
## curitibapr
## custs
## daniell
## dav
## decid
## deliberal
## descobr
## descrevel
## desiquilib
## diariamens
## diariasseman
## diciona
## diferec
## diferenci
## dimenso
## dispus
## diversificaca
## doutor
## dta
## duplicaca
## dutovi
## educaca
## electr
## elegi
## elektr
## eletrodomes
## eletronucl
```

```
## email
## emisso
## empes
## empresaindustr
## encaix
## encarecid
## encarg
## encomend
## eneerg
## enege
## energe
## energeticac
## energu
## energy
## engan
## engismael
## enorm
## ensai
## epitaci
## equilib
## escut
## estadosregio
## estatisi
## estatisticosdistribuica
## eten
## evaporaca
## excet
## excolabor
## execuca
## exemplific
## explicaco
## export
## fabi
## fabricaca
## facti
## faculdad
## fai
## famili
## fil
## firewall
## fl
## florest
## following
## fornecimentox
## fort
## fot
## francasp
## freez
## fuel
## gas
## gasbol
## generaca
## gent
## geoespac
```

```
## geoprocess
## geraco
## giselly
## goiand
## goiandirag
## grang
## gsub
## guarulh
## historic
## httpgovbrmercadopaginasdefaultaspx
## httpgovbrptpublicacoesabertospublicacoesconsumoeletricaconsumomensaleletricaclasseregioessubsistem
## httproundsgovbrarquivoseditaltgeditalchamadapublicapdf
## httpsistemasgovbram
## httpwwwaneelgovbraplicacoesresumoestadualresumoestadualcfm
## httpwwwcanalenergiabrzpublishermateriasretrospectivaasp
## httpwwwgovbrmercadodocumentsresenhamensaldezembropdf
## ide
## idealiz
## iden
## importaca
## inciner
## incineraca
## industriaiscomerc
## industrialcomerc
## iner
## inf
## influenc
## instanc
## instig
## intermit
## internat
## ipeadat
## iri
## irrestrit
## ita
## itapirang
## ivinhem
## jaa
## jalil
## jamanxim
## janaub
## jen
## joic
## jonatan
## juni
## just
## karl
## ketlin
## konzen
## lacerd
## laje
## latitud
## laur
## learning
```

```
## legisl
## leonard
## light
## lik
## login
## longitud
## lpg
## lubricant
## lubrific
## lucel
## lug
## machin
## madr
## mainly
## mamanu
## manausam
## manipul
## manipulaca
## marcel
## maring
## materiaspr
## matricul
## mba
## mensur
## mercadolog
## metal
## meter
## mid
## milho
## millikan
## modernizaco
## mp
## naamazon
## naca
## naft
## naphth
## nav
## naveg
## navi
## need
## nerg
## nev
## norte
## observaco
## ocean
## od
## onquant
## oth
## pacot
## padro
## paran
## parecil
## passi
## peg
```

```
## pesquis
## pesquisaestud
## petroleum
## pi
## picon
## pigouvi
## pimesufp
## poligon
## populacaoveicul
## ppeaufop
## ppg
## ppgmn
## pretomg
## previsaodemand
## prezadx
## production
## prof
## profund
## progress
## propi
## proporca
## prorrogaca
## proviso
## prudentesp
## quantific
## quas
## queim
## rec
## recicl
## recuper
## refined
## refriger
## regiaobairr
## regioesest
## relac
## relaco
## relevanc
## representat
## requisica
## residencialindustr
## residu
## restaurant
## retrat
## retrofit
## rg
## riv
## rodrig
## rote
## rubr
## sa
## salient
## sanit
## santanasalvadorb
## servent
```

```
## shopping
## sinc
## sinte
## sirraul
## solciti
## solicital
## sous
## stand
## ston
## subbac
## sugesta
## sup
## supermerc
## switch
## sylvi
## szkl
## tabulaca
## tak
## tancred
## taquar
## tdr
## tema
## tenc
## tep
## teres
## terren
## that
## todav
## too
## topic
## totaliz
## trifas
## turqu
## uberland
## ufccaen
## ufgd
## ufpaicsafacecon
## ufpr
## ufsc
## ufu
## ufvjm
## unidirec
## unifacef
## univat
## universit
## upgn
## used
## usprp
## vali
## valorizaca
## varej
## variaca
## variaco
## vazaodat
```

```
## veget
## veraa
## veriq
## viabilizaca
## vic
## visu
## vo
## voip
## wat
## when
## will
## ach
## agenc
## alexandr
## atu
## biocombusti
## compreeend
## constitu
## correg
## decorrenc
## demonstr
## departament
## df
## direcion
## disposica
## dut
## edwig
## empreg
## enquadr
## ex
## execu
## fabr
## fer
## fernand
## fiz
## formaca
## frequenc
## fund
## gerenc
## implant
## indigen
## indispensa
## inic
## malh
## mauric
## membr
## mg
## microgeraca
## mor
## motiv
## parc
## sal
## saudaco
## shp
```

```
## simil
## sp
## tempor
## ufsm
## adequ
## administraca
## aeronav
## antema
## assemble
## august
## autoridad
## av
## avis
## bast
## bibliotec
## bioenerg
## bloc
## celul
## cep
## clar
## compet
## complet
## comunicaco
## concorrenc
## confidencial
## contextualiz
## continuaca
## corp
## correi
## deix
## descrica
## design
## dire
## dirig
## discre
## discriminaca
## distanc
## doming
## edifici
## entidad
## equival
## estrang
## estrateg
## ferrament
## fgv
## forc
## fundaca
## gabriel
## gerenci
## gerent
## getuli
## gnl
## gradu
## gui
```

```
## impress
## indenizaco
## inteligenc
## is
## lanc
## manifest
## mecan
## melac
## melh
## metr
## objet
## organ
## pe
## peric
## permission
## petrobr
## priv
## pront
## prov
## quadr
## quent
## redig
## refefer
## remuneraca
## restrit
## risc
## saud
## situaca
## sr
## tcu
## termin
## tribun
## turc
## urgent
## varg
## vind
## virtud
## visit
## volt
## web
## xlsx
## salari
## orgaoent
## locaca
## esport
## imovel
## patrocini
## pessoal
## vag
## carr
## incen
## prestaca
## fiscal
## profiss
```

```
## acuc
## comunicaca
## reajust
## conselh
## control
## gratificaca
## seleca
## anal
## audi
## contrataco
## direit
## jurid
## licitaco
## loc
## ocup
## organiz
## ouvid
## patrocin
## sed
## segur
## vincul
## adit
## advog
## cronolog
## demissa
## desp
## exclus
## fiscalizaca
## instruca
## organizaca
## pil
## possibilit
## quantit
## regr
## retir
## terceir
## terceirizaca
## aluguel
## conarq
## confianc
## defer
## estagia
## etanol
## funco
## labex
## ministr
## pad
## quaisqu
## republ
## visual
## apostil
## ato
## aven
## cole
```

```
## complement
## conform
## constituc
## continu
## convenco
## cooperaca
## corporat
## curricul
## desinfecca
## disciplin
## empresar
## ergonom
## espor
## exploraca
## farmac
## folh
## instaur
## instrument
## logis
## mao
## natalin
## penal
## reg
## repartica
## salar
## sediment
## seleco
## temporal
## tesour
## vig
## word
## almoxarif
## alug
## ape
## autoriz
## banh
## bqa
## caix
## certific
## cgu
## comprov
## comput
## contenci
## deciso
## decorr
## dl
## drog
## ed
## efet
## embrap
## exercici
## fas
## gar
## glob
```

```
## gom
## governament
## independ
## instituc
## iso
## lo
## noutr
## pda
## perspec
## plr
## preced
## prega
## prego
## principi
## refeica
## relatoriosintes
## renovaca
## resoluca
## semelh
## siasgweb
## sidec
## sobreprec
## soluco
## tonel
## validaca
## venh
## vid
## aca
## acid
## acolh
## acontec
## acorda
## administrativad
## advocac
## ago
## ajust
## alimentaca
## aline
## ambienteecolog
## anticorrupca
## api
## arbit
## arquitet
## arquivis
## ataqu
## audit
## aut
## aviaca
## barboz
## be
## beatriz
## bilho
## brasiliadf
## brazili
```

```
## caf
## carl
## cesgran
## chef
## cidada
## coaching
## coloc
## comit
## competenc
## concursadosterceirizadoscomission
## condica
## condomini
## condut
## confecca
## confirmaca
## contenh
## convocaca
## convocaco
## coordenad
## corporaca
## credenci
## cronogram
## csll
## curt
## defes
## deleg
## demisso
## deput
## descart
## descobert
## desenh
## dest
## destinaca
## devolv
## dieg
## discorr
## doenc
## dou
## down
## dpgr
## duraca
## ebserh
## eric
## estipul
## experienc
## exteri
## extern
## feri
## fiq
## flutu
## fotograf
## gisel
## gratuit
## hardw
```

```
## hierarqu
## honora
## hospital
## httpgovbrsitesptservidoresconcursospublicosdocumentscbaconcursopcbablicodaepesituacacaodasconvocacacl
## httpwwwwebserhgovbr
## igovt
## inclusiv
## inexistenc
## instauraca
## intranet
## invert
## investig
## irpj
## isabell
## jornal
## julg
## jus
## justifiqu
## lai
## ldo
## load
## locata
## medic
## mei
## ment
## nacionalinternac
## nil
## nomin
## nun
## ocupaca
## ozoni
## patrimon
## pav
## planocodig
## planoprogram
## pol
## premium
## prepar
## preservaca
## preven
## prevenca
## prim
## publicita
## qvt
## razo
## reda
## reflet
## reubl
## revi
## rodovia
## salvaguard
## secretarioassist
## sen
## sucumbenc
```

```
## sug
## sumul
## teletrabalh
## titulaca
## tst
## ufrg
## uspol
## validad
## vari
## vet
## viag
## vide
## vim
## vou
## wik
## abrangenc
## acompanhal
## act
## adequal
## adi
## adjunt
## admit
## adverbi
## advertenc
## afer
## afet
## afo
## afrobrasil
## agend
## agiliz
## alc
## alessi
## alienaca
## aliment
## alimentacaorefeica
## alter
## ampar
## analistasconsul
## ancor
## andaresconjunt
## anon
## antepost
## apf
## aposent
## archivisttoolkit
## arit
## arm
## arq
## arquvi
## artific
## artur
## assess
## assessor
## assin
```

```
## atest
## atom
## atualil
## atualiz
## atualizaco
## aul
## ausent
## autentic
## autom
## backup
## bande
## bdi
## beb
## beneficia
## bimestr
## biomet
## blackenergy
## bm
## bnd
## bovesp
## brind
## busqu
## cade
## cala
## cancel
## carenc
## cargofunca
## carto
## cdfunca
## cedric
## celebr
## celet
## cesam
## check
## chilen
## ciberne
## cipacissp
## cipacomissa
## cissp
## civil
## classif
## cobert
## colabor
## colocaca
## comand
## comenta
## comission
## company
## comparaco
## compil
## compre
## compressa
## conarqctd
## conf
```

```
## confiancacarg
## confus
## congen
## conjugaca
## conso
## constanci
## consul
## contatal
## contavel
## contemplaca
## contentdm
## contratual
## convit
## coordenaca
## correco
## correlat
## cpf
## cuid
## cumpr
## custe
## danill
## david
## dedic
## defen
## defici
## definitivadermatos
## delivery
## demandara
## denunc
## departamentoslaborato
## dependenc
## descumpr
## desembarg
## desembols
## desestimul
## desfaz
## desvi
## deu
## difusa
## discretizaca
## discussa
## dispensadosexoner
## dispensaexoneraca
## dispers
## div
## divers
## domes
## dsic
## ecologic
## econometr
## efici
## element
## eletronic
## elev
```

```
## emanuel
## empa
## empenh
## empregadodirig
## empresaentidadeorga
## encerr
## enfim
## enforc
## enriquec
## ensin
## entrev
## epoc
## escop
## estatuta
## estimul
## estrit
## estu
## exam
## exat
## exclu
## expect
## expl
## falh
## farmaciaval
## fech
## ferna
## fidalg
## flexi
## fme
## formula
## formulaca
## forncec
## fracking
## fraud
## frot
## fur
## fut
## gavet
## ged
## geolocalizaca
## geolog
## governanc
## guerr
## havel
## hist
## homossex
## httpgovbrptimprensanoticiasdecidereformularconcurs
## httphdlhandlenet
## httpsbengovbrdownloadscadntesedorelatcbriofinalwebpdf
## httpsgisepeprdgovbrwebmapep
## httpswwwcomprasgovernamentaisgovbrindexphplegislacaoinstrucoesnormativasinstrucaonormativadezembr
## httpswwwnovacanaindustriainvestimentobiocombustiveisdemandamprojet
## httpwwwgovbrpdeeformsepeestudoaspx
## httpwwwgovbrptabcdenergiaplanejamentoenerg
```

```
## httpwwwgovbrptpublicacoesabertospublicacoesconsumomensaleletricaclasseregioessubsistem
## human
## icaatom
## ido
## incid
## inicop
## injustific
## inscrev
## inser
## instrumentalizaca
## intenca
## interi
## investigaca
## iptu
## janain
## jeton
## jorn
## jos
## judic
## juiz
## jun
## kim
## lauraramosifmsedubr
## layout
## lerdortantracosebissinosesurd
## letr
## lin
## liquidaca
## locaco
## locat
## louc
## lum
## malefici
## malw
## marcu
## marluc
## metropol
## microond
## microsot
## milen
## milh
## minigeraca
## minor
## missa
## mist
## mix
## monet
## moss
## mpdg
## mpog
## muit
## nalgum
## natal
## nomeaco
```

```
## nomenclat
## notpety
## oci
## ocupacionalsiderosecataratadoenc
## omit
## oportun
## opt
## orcament
## orcamenta
## orcamentariofinanc
## orden
## ordin
## organogram
## orgaosent
## orientaco
## origin
## pac
## palh
## panoram
## parnaib
## pbt
## pcdccd
## pdfdoc
## pend
## pequen
## percorr
## perit
## pety
## planinvest
## plas
## plena
## poc
## poduca
## pontual
## posico
## pre
## precav
## preestabelec
## prelanc
## premenc
## premiaca
## preparaca
## presidentevicepresidentesdiretoressuperintendentesger
## prevenco
## previdenc
## prezaod
## privatizaca
## probabil
## profer
## profissa
## proprieta
## proteg
## providenc
## providenci
```

```
## psicosoc
## publicidadepublicaco
## qua
## qualificacaoaperfeico
## quarent
## radial
## ransomw
## rb
## rdcarq
## reconcav
## reembols
## reformul
## regulaco
## remuneraco
## renom
## repart
## reposito
## representaco
## requer
## resciso
## resolucaotcu
## respe
## respeitos
## restabelec
## retribuica
## retro
## revog
## ricard
## robert
## rogeri
## ruid
## salariobas
## sane
## secretosigil
## segred
## sensi
## servidoresempreg
## shapfil
## sian
## sib
## sigad
## significativ
## situac
## slid
## socialcpf
## startup
## submet
## subordin
## subprocuradoresger
## substituica
## suficienteseficaz
## sujeit
## sumar
## supeir
```

```
## susan
## suspensa
## tampouc
## tempora
## terrestr
## tessarin
## tid
## toe
## tpa
## trac
## trag
## transferenc
## transgener
## trf
## tripulaca
## tsv
## ttdttdd
## tubulaca
## uasg
## unb
## unidadegerenciadiretoriasuperintendenc
## vacanc
## ved
## verb
## vicepresid
## vidr
## violaca
## violent
## vot
## wannacry
## whatsapp
## willian
## xlsxlsx
## zone
##
## leila
## habilit
## deererev
## nt
## parametr
## ama
## modul
## cvu
## termoeletr
## sol
## hidreletr
## ahe
## deficit
## limit
## habilitaca
## anemometr
## rs
## vent
## ccpectet
```

```
## conex
## eol
## gt
## mau
## mon
## monitor
## pucri
## rap
## sin
## subestaco
## medico
## reserv
## barr
## bifac
## climatolog
## confi
## deedearerev
## deentrev
## furn
## neg
## nodal
## piau
## precostet
## prt
## pv
## reproduz
## suap
## tust
## tv
## unita
## art
## ler
## memor
## particip
## venc
## iii
## ofici
## torr
## antoni
## cec
## ceg
## cgh
## cmo
## cms
## consequenc
## dinh
## disposi
## etap
## fever
## gna
## hibr
## hom
## imetam
## instruco
```

```
## itaipu
## luiz
## maranh
## ns
## offshor
## pne
## presum
## reforc
## rod
## supracit
## tet
## vian
## vnr
## cme
## interlig
## kv
## medica
## outorg
## reposica
## veloc
## acr
## anared
## import
## instalaca
## pch
## pic
## supr
## altern
## bel
## ce
## conclu
## distribuica
## fich
## mwh
## regim
## acrerondon
## acu
## aeroger
## amap
## aneelleila
## anemograf
## ansi
## antecip
## aport
## atlan
## ave
## bat
## brizon
## cedr
## chapeub
## compromiss
## comprovaca
## contribuint
## correspondenc
```

```
## csp
## cteep
## cvp
## dec
## deck
## deeptr
## delr
## deslig
## determinaco
## disput
## distint
## dro
## elenc
## encontral
## especificaca
## exig
## exigenc
## fabbr
## fac
## ganh
## guacu
## hidraulic
## hidroeletr
## hidrogeraca
## httpwwwgovbrtransmissaopaginasdadosparaestudosdeplanejamentodatransmisscaopdeaspx
## igap
## incentiv
## infom
## ingl
## interligaca
## intermitenc
## inventa
## irma
## jandairarn
## lavr
## lelio
## mgw
## minuci
## mitig
## morr
## mossor
## nac
## ndpespemm
## obtev
## onshor
## onsr
## paragu
## participaco
## pd
## permanent
## peruib
## petpelp
## prerequisit
## propriedad
```

```
## pucmin
## recebel
## recuperaca
## reep
## refrigeraca
## regaseificaca
## relicit
## renat
## replic
## revitalizaca
## rondon
## sav
## sgtaneel
## sisorh
## solarimetr
## subestaca
## subterrane
## sudo
## tapaj
## tecnicoeconom
## thiag
## transmis
## trombet
## tronc
## uirapuru
## unesp
## uniform
## utegnsp
## utilz
## volg
## vr
## cab
## digit
## divulgaca
## documentaca
## extensa
## intermedi
## matema
## mdi
## nenhum
## palm
## parec
## precotet
## rora
## semestr
## sergip
## siti
## tes
## verd
## viabil
## abastec
## acredit
## agost
## celg
```

```
## cheg
## cientif
## co
## cost
## defin
## descrit
## estaca
## estrut
## federac
## it
## joa
## nom
## paragraf
## posi
## pr
## propost
## proveni
## softw
## ufv
## abril
## acresc
## alteraca
## aproveit
## cce
## cepel
## chesf
## compreensa
## coqu
## decret
## deentr
## desloc
## destac
## dispost
## emit
## escal
## estaco
## estagi
## evt
## explicit
## gros
## gw
## importanc
## instalaco
## intercambi
## lot
## lt
## ltd
## palestr
## parcel
## pg
## po
## pratic
## questo
## regulament
```

```
## respons
## shap
## simulaco
## soluca
## subsistem
## sud
## tarif
## tocantim
## abengo
## abrac
## absorca
## abun
## acces
## aceita
## adem
## aeg
## aep
## aliquotabas
## alteraco
## alumin
## amorf
## ampliaca
## aneelanatel
## anteced
## aparent
## applied
## aprendiz
## apresentaco
## aproximaca
## apt
## aracat
## aracatub
## aripu
## arranj
## assum
## atra
## atrat
## atribuica
## attl
## austr
## autotransform
## bag
## barcel
## basei
## bauru
## biolog
## boj
## borb
## cajuru
## camarg
## camocimc
## cano
## capivaricacho
## caracterizaca
```

```
## carbonit
## carolin
## carvaa
## castr
## cavac
## cee
## cel
## cemiggt
## chav
## cho
## chu
## cient
## clic
## clovil
## cnu
## colisa
## colomb
## coment
## comig
## comparec
## competi
## comunita
## comwh
## concluso
## concret
## condiz
## condu
## conec
## conexo
## confeccion
## consorci
## contingenc
## copcec
## copelgt
## copp
## coppead
## coronel
## correca
## corrig
## coxip
## crei
## ctapr
## cuiab
## curitib
## dal
## declar
## deeit
## deerer
## defas
## deliber
## deliberaca
## descont
## descri
## desss
```

```
## deten
## devera
## dezen
## diferenc
## dificil
## dificult
## dimensa
## diog
## diogosignorgmail
## diox
## dissertacaa
## dist
## distribuicaotransmissa
## doc
## docu
## domici
## donwload
## doument
## ebrasil
## ebtt
## edrr
## educ
## efetu
## elaboracaa
## elasticidaderend
## eletromagne
## eletronort
## eletrosul
## empreed
## en
## ena
## enacel
## enfrent
## engi
## enrroc
## enterr
## enumer
## enunci
## eolicofotovolta
## ere
## erval
## espirit
## estend
## estiv
## estreit
## esutd
## exc
## exceca
## exp
## expanca
## explor
## facult
## fal
## falenc
```

```
## far
## ferr
## fh
## fici
## figueired
## fixaca
## florianopolil
## formul
## franc
## francil
## ftp
## gafanhot
## gave
## gcp
## gd
## ge
## gen
## genpow
## georeferenc
## geron
## getn
## gfil
## gg
## go
## googl
## gov
## graduand
## grav
## graziell
## grid
## gusm
## hidreledr
## hidrolog
## historioc
## httpgovbrsitesptpublicacoesabertospublicacoespublicacoesarquivospublicacaodeeitaverscaofinalpdf
## httponsorgbrptpaginasresultadosoperacaohistoricooperaca
## httpwwwcceeorgbrportalfacesoquefazemosmenulateralleiloesadfctrlstatecpigtknafrloopfafrloopdadfctrlsta
## httpwwwgovbrleiloesdocumentsleilcbesdeenergiadereservacbalercadastradospdf
## idiom
## ifsul
## igpd
## igual
## imediat
## imperatriz
## imprens
## imprensalightbr
## inat
## incentivoobrigaca
## incompati
## informac
## inglesfiqu
## iniic
## inmet
## insum
```

```
## intens
## interromp
## invaria
## irradiaca
## irrigaca
## itapipoc
## itaqu
## ituting
## jeann
## joanneum
## joas
## jup
## kay
## kelman
## kwmw
## leia
## len
## lest
## liberal
## licenci
## lilian
## log
## londrin
## mad
## manipulal
## manobra
## mantovilil
## marin
## marmel
## marqu
## martim
## mascarenh
## massayosh
## mctic
## mecanc
## meteorolog
## metereolog
## migraca
## mim
## mistur
## modulosinver
## monocristalin
## mora
## moura
## msctech
## multipl
## mv
## mva
## mwm
## mwp
## narr
## neblin
## obrig
## observanc
```

```
## obteca
## obtel
## ocultaca
## oner
## onsdt
## out
## padra
## palavr
## pali
## paraben
## parad
## parigot
## paril
## patric
## paulin
## pbc
## pedil
## pelot
## perc
## perfeit
## permanenc
## perme
## peron
## pet
## petr
## pibit
## pinh
## plant
## play
## pleite
## policristalin
## pouc
## pow
## prai
## preceit
## preez
## prem
## pressa
## print
## produ
## programaca
## propic
## prospecco
## prospect
## prosper
## proteca
## pusch
## pwf
## qtd
## qualific
## qued
## quix
## ra
## rafael
```

```
## rastre
## rd
## rea
## realocaca
## recomendaco
## reconstru
## redaca
## ree
## refent
## regulamentaco
## relacional
## relocaca
## repet
## repr
## reproduzindoatualiz
## requis
## rest
## retific
## reversi
## reviso
## rit
## rosan
## rosangel
## rot
## rr
## rua
## russ
## sagr
## saik
## salin
## salt
## sazonal
## scienc
## screen
## sctaneel
## seccion
## secunda
## serr
## servira
## sign
## silici
## simul
## sincer
## situaco
## sofr
## soliciti
## solt
## somato
## stat
## superaca
## superfici
## supramencion
## sustenta
## taref
```

```
## tarj
## tecnic
## tecnog
## ted
## telecomunicac
## telfax
## termomaranha
## traf
## trair
## trajet
## transcrit
## transform
## transformaca
## transito
## transmit
## tributaca
## trimestr
## tubara
## tvpuc
## uberab
## uff
## umid
## umidad
## unid
## unisin
## upload
## urani
## usmw
## utv
## vaness
## velh
## verificaca
## versu
## vicent
## vicereit
## vistainteri
## vontad
## zao
## adot
## adquir
## afirm
## altur
## ana
## andr
## aplica
## apont
## associ
## ata
## atl
## atualizaca
## automa
## bah
## base
## ben
```

```
## capitul
## carv
## castanh
## cemig
## cesp
## chuv
## codig
## comercializ
## comunic
## concessa
## conclusa
## concurs
## condicion
## condico
## conect
## confirm
## cons
## conservaca
## consolid
## constituica
## curv
## deede
## denomin
## direca
## divisa
## eletrobr
## energis
## entend
## err
## estatut
## gastrading
## goi
## imped
## kw
## leit
## licitaca
## luc
## manutenca
## met
## metod
## minim
## mwmed
## ne
## negoci
## newav
## obtiv
## oliv
## pa
## par
## pedr
## perfil
## portug
## posteri
## produt
```

```
## profes
## qualificaca
## raza
## recomendaca
## region
## relev
## renova
## resum
## rural
## signific
## sirv
## subgrup
## suport
## telefon
## temperat
## th
## titul
## uf
## universita
## veicul
## verifiq
## vigent
## vii
## visa
## volum
## workshop
## xxxii
## ab
## aba
## abord
## acat
## ader
## adoca
## agrup
## alex
## alinh
## aloc
## an
## antig
## aparelh
## apres
## apresentaca
## aprofund
## apuraca
## ar
## arin
## assegur
## assist
## atmosfer
## audienc
## automaca
## baliz
## barb
## benchmarking
```

```
## boletim
## bols
## californ
## cam
## campu
## candidat
## carbon
## carol
## cart
## ced
## celp
## centroo
## certam
## cienc
## clim
## cnpq
## coelb
## coelc
## combin
## comissa
## compo
## configuraca
## consequ
## constat
## contabil
## convoc
## copel
## corr
## cosern
## ct
## cult
## dad
## dea
## decisa
## deent
## def
## delim
## desempenh
## dispo
## dispos
## doutorand
## edica
## edita
## efetiv
## eficienc
## eng
## ent
## equlibr
## escolh
## essenc
## estatal
## estoqu
## eventual
## exclusiv
```

```
## express
## extr
## extra
## extraca
## fatur
## financ
## firm
## fix
## fomal
## fun
## fundament
## gwm
## hidr
## ibam
## ilh
## implement
## indenizaca
## indicaca
## iniciaca
## inscrica
## int
## integraca
## isol
## isoladomw
## ivan
## jardim
## km
## laud
## lembr
## lenh
## lev
## liber
## licit
## liv
## lixiv
## lucen
## luz
## mach
## manau
## mand
## mater
## mesquit
## minut
## mov
## ms
## nasc
## nel
## net
## noro
## nov
## nucle
## nul
## ob
## obrigato
```

```
## obrigatoriedad
## ocorrenc
## oest
## of
## opca
## opinia
## ora
## orientaca
## otim
## our
## ouv
## pacienc
## pais
## part
## patamar
## percent
## permanec
## pertin
## pes
## pleit
## poli
## pont
## posgraduaca
## pragma
## prefer
## premiss
## prime
## problem
## prop
## prud
## publicaco
## quinzen
## ramal
## rapid
## reduca
## reduz
## referenci
## reiter
## rel
## ren
## requisit
## respectiv
## restrico
## reunio
## rev
## ric
## rj
## rn
## sec
## seca
## secret
## selec
## semina
## sergi
```

```
## setor
## setorcomerc
## sic
## silv
## simplific
## simultane
## situ
## submercadonort
## subsidi
## sudestec
## sufici
## suger
## sulgip
## tom
## tr
## trech
## triangul
## tribut
## txt
## ufrj
## uftm
## ultrassecret
## university
## usuari
## vaza
## verif
## vi
## vigenc
## vitor
## eletr
## estud
## tecn
## eolic
## document
## relat
## empr
## empreend
## projet
## not
## usin
## leilo
## cop
## refer
## utiliz
## transmissa
## calcul
## gerac
## realiz
## potenc
## font
## pesquil
## cadastr
## expansa
## pde
```

```
## energ
## demand
## ba
## consum
## regia
## cust
## apresent
## analis
## ga
## garant
## inform
## parqu
## encontr
## atend
## exist
## quant
## sit
## anex
## consider
## seri
## ute
## disponivel
## fisic
## ped
## precis
## rio
## fotovolta
## metodolog
## obt
## referenc
## regio
## capac
## detalh
## lei
## algum
## atenci
## gentil
## mens
## nacion
## nord
## plan
## planej
## process
## seguint
## set
## desagreg
## natur
## progr
## arqu
## biomass
## desenvolv
## dispon
## disponibilizaca
## etc
```

```
## linh
## necessit
## period
## prec
## pres
## sul
## trat
## val
## combusti
## ger
## ii
## list
## mme
## model
## mont
## om
## port
## previs
## senh
## tabel
## term
## uhe
## basic
## envi
## esco
## instal
## operac
## produca
## atual
## const
## consult
## est
## fornec
## horari
## inclu
## merc
## municipi
## public
## red
## ult
## aneel
## atenca
## att
## carg
## despach
## histor
## invest
## med
## min
## respeit
## sant
## anual
## certificaca
## contat
```

```
## gwh
## planilh
## result
## revisa
## total
## abaix
## academ
## artig
## cenari
## distribu
## engenh
## excel
## flux
## graf
## industr
## link
## livr
## long
## marg
## nort
## receb
## relacion
## renov
## univ
## vist
## acim
## acord
## agent
## aguard
## anteri
## cicl
## crite
## elabor
## eletric
## encaminh
## entr
## espec
## fat
## geograf
## horizont
## marc
## medi
## mw
## necess
## novembr
## obr
## participaca
## pec
## pergunt
## recent
## registr
## regul
## respost
## retorn
```

```
## sid
## sistem
## solicit
## tecnolog
## unidad
## uso
## ajud
## avaliaca
## camp
## cit
## colet
## consegu
## construca
## cont
## contrat
## contratac
## cur
## decen
## di
## diss
## econom
## emissa
## especific
## fed
## feit
## impact
## jan
## localizaca
## mat
## matriz
## mostr
## obje
## ole
## org
## pass
## paul
## pod
## proced
## publ
## realizaca
## ref
## seg
## termeletr
## acompanh
## agradec
## alun
## ambit
## and
## assunt
## auxili
## canal
## compr
## consid
## coorden
```

```
## definica
## dev
## diesel
## duvid
## eletron
## equip
## esclarec
## graduaca
## indic
## inici
## licenc
## lucr
## mai
## ministe
## necessa
## ola
## pag
## pal
## poss
## produz
## real
## requ
## sc
## sigil
## utilizaca
## vend
## acess
## agu
## ambient
## antecipad
## atenc
## bac
## car
## classific
## correspond
## diari
## elaboraca
## especif
## estabelec
## inter
## interess
## iv
## man
## melhor
## mencion
## ocorr
## orig
## possu
## pra
## praz
## prev
## prez
## princip
## publicaca
```

```
## pud
## recomend
## report
## respec
## setembr
## unic
## vers
## adic
## ambi
## aprov
## ativ
## baix
## banc
## brut
## catarin
## centr
## ch
## cidad
## compar
## compreend
## conced
## contribu
## convers
## cord
## daniel
## dat
## dentr
## descrev
## desej
## dezembr
## diretriz
## edit
## envolv
## escrev
## esper
## estad
## estatil
## estim
## event
## fic
## funcion
## junt
## legal
## legislaca
## mail
## maquin
## observ
## oferec
## ofert
## outubr
## pagin
## previst
## resid
## respond
```

```
## responsa
## sent
## serv
## shapefil
## soc
## trabalh
## us
## var
## vis
## abert
## alt
## amazon
## anuari
## aspect
## branc
## cl
## comp
## compartilh
## composica
## conhec
## conjunt
## consig
## difer
## diret
## embas
## enderec
## evoluca
## exportaca
## fabric
## func
## funca
## hav
## identific
## identificaca
## implementaca
## impost
## incis
## integr
## julh
## justific
## localiz
## mape
## mod
## modal
## nega
## oper
## orc
## pertenc
## pesso
## petrole
## plataform
## post
## preench
## presid
```

```
## prest
## pret
## procur
## rend
## reunia
## segment
## segu
## text
## tir
## transparenc
## turbin
## varia
## verific
## via
## abr
## aceit
## adequaca
## administr
## alcanc
## aplicaca
## aquisica
## are
## atuaca
## aument
## autor
## autorizaca
## avali
## bagac
## benefici
## biog
## busc
## can
## capit
## caracteris
## carat
## cativ
## cerc
## cert
## cham
## classificaca
## cnpj
## colaborac
## comec
## comerc
## companh
## computac
## concession
## contempl
## context
## contribuico
## convenc
## corret
## cresc
## cri
```

```
## depend
## deriv
## determin
## discrimin
## disp
## disponibiliz
## divulg
## entreg
## escrit
## especi
## estrutur
## exempl
## expost
## facil
## falt
## feir
## foc
## form
## futur
## gast
## govern
## grau
## hipotes
## implantaca
## inflexibil
## institu
## internac
## internet
## interval
## intuit
## kg
## levant
## liqu
## manifestaca
## mant
## mar
## max
## mestr
## monograf
## mund
## norm
## notic
## numer
## obtenca
## ofic
## onlin
## ord
## orient
## otimizaca
## papel
## paraib
## pdf
## perceb
## perd
```

```
## permit
## pretend
## projeco
## protocol
## quebr
## question
## questiona
## recurs
## represent
## ressalt
## restrica
## seguranc
## separ
## simpl
## sintes
## superi
## tax
## tend
## torn
## transport
## faix
## kwh
## riacha
## balanc
## porcent
## agreg
## incidenc
## procel
## refin
## seman
## celulos
## eiar
## reservato
## projeca
## aquavia
## bairr
## causal
## comerci
## compon
## consom
## dataca
## desagregaca
## divergenc
## espac
## format
## hor
## iluminaca
## injet
## lic
## motriz
## traz
## figur
## municip
## francisc
```

```
## aai
## ae
## arcur
## armazen
## cadern
## cald
## calh
## cna
## colleg
## contruco
## dens
## eia
## eletropaul
## entant
## execut
## external
## felip
## gel
## georreferenci
## habit
## individ
## induca
## inferi
## juruen
## london
## micr
## mudanc
## particul
## pib
## pir
## rai
## ram
## recort
## rim
## som
## ufb
## vazo
## vera
## xl
## map
## periodic
## webmap
## categor
## distrit
## infraestrut
## resenh
## tel
## adri
## aflu
## alago
## alban
## altra
## anab
## anat
```

```
## angel
## angelin
## aplicaco
## arcondicion
## bar
## bdt
## big
## biodiges
## boc
## bomb
## calorif
## campanh
## chapec
## compati
## concentraca
## concesso
## cris
## cruz
## dalm
## demartin
## desconhec
## dificuldad
## domicili
## enerd
## entretant
## escass
## espanh
## estratificaca
## eugeni
## financi
## fluidomecan
## fornecel
## foz
## frequ
## gehrk
## geod
## georeferenci
## gil
## grum
## grup
## horosazon
## httplattescnpqbr
## httpsgisepegovbrwebmapep
## identifiq
## inct
## indiqu
## individu
## ineficienc
## infomerc
## inp
## inst
## invern
## ire
## ivy
```

```
## laborato
## lal
## lamp
## latt
## li
## lidin
## lim
## liquefeit
## lix
## lucrat
## manuel
## matogross
## matte
## merit
## mesorregia
## mestrand
## metropolit
## miner
## modernizaca
## moral
## mt
## ner
## network
## nigr
## ning
## oil
## olh
## paracatu
## paranapanem
## pecu
## pernambuc
## populac
## possi
## preferenc
## prorrog
## proveit
## ranking
## regin
## remuner
## ribeira
## segreg
## selecion
## socioambient
## sucess
## sugesto
## supply
## tabul
## tcc
## temp
## tendenc
## test
## the
## to
## tolmasquim
```

```
## uc
## ufmt
## uno
## urgenc
## urucucoarimanau
## utel
## vagn
## ventil
## aplic
## hidraul
## intern
## receit
## unia
## anp
## brun
## constru
## conteud
## csv
## divid
## gasolin
## ideal
## junh
## motor
## negr
## product
## sociedad
## urban
## acion
## aco
## aere
## cas
## client
## cobr
## destin
## efeit
## existenc
## gasodut
## gesta
## glp
## imens
## institut
## operaco
## pemat
## predi
## proc
## residenc
## sab
## territo
## tip
## tipic
## util
## about
## abrang
## acessi
```

```
## acrescent
## adicion
## agentesxlsm
## agrad
## agregaca
## agricol
## agronom
## agropecu
## alag
## amadadoshistoricosmedicoesanemometricasxlsx
## amand
## amig
## amostr
## anaerob
## aneelcceeom
## antecedenc
## apiac
## apliqu
## apot
## aptida
## arrecadaca
## asa
## asfalt
## asphalt
## assent
## at
## atencioas
## ating
## atrel
## aurel
## availabl
## avanc
## averigu
## azul
## bacharel
## baian
## barril
## bell
## bidirec
## biriguisp
## bren
## brent
## brentinternationalriversorg
## brig
## build
## but
## cal
## campin
## campinapolil
## canalenerg
## cand
## capt
## caus
## cav
```

```
## cc
## ceb
## celesc
## celtim
## cez
## chec
## classese
## cleyton
## cliamb
## cln
## cnec
## cok
## colun
## comportament
## compres
## comunidad
## conceitu
## confecc
## conscien
## conscientizaca
## consideraca
## consumidoresse
## consumption
## contabiliz
## contabilizaca
## contrast
## conver
## coppeufrj
## corre
## correlaca
## correlataconversi
## correntin
## costum
## cosum
## cpflpaul
## cristalin
## curitibapr
## custs
## daniell
## dav
## decid
## deliberal
## descobr
## descrevel
## desiquilib
## diariamens
## diariasseman
## diciona
## diferec
## diferenci
## dimenso
## dispus
## diversificaca
## doutor
```

```
## dta
## duplicaca
## dutovi
## educaca
## electr
## elegi
## elektr
## eletrodomes
## eletronucl
## email
## emisso
## empes
## empresaindustr
## encaix
## encarecid
## encarg
## encomend
## eneerg
## enege
## energe
## energeticac
## energu
## energy
## engan
## engismael
## enorm
## ensai
## epitaci
## equilib
## escut
## estadosregio
## estatisi
## estatisticosdistribuica
## eten
## evaporaca
## excet
## excolabor
## execuca
## exemplific
## explicaco
## export
## fabi
## fabricaca
## facti
## faculdad
## fai
## famili
## fil
## firewall
## fl
## florest
## following
## fornecimentox
## fort
```

```
## fot
## francasp
## freez
## fuel
## gas
## gasbol
## generaca
## gent
## geoespac
## geoprocess
## geraco
## giselly
## goiand
## goiandirag
## grang
## gsub
## guarulh
## historic
## httpgovbrmercadopaginasdefaultaspx
## httpgovbrptpublicacoesabertospublicacoesconsumoeletricaconsumomensaleletricaclasseregioessubsistem
## httproundsgovbrarquivoseditaltgeditalchamadapublicapdf
## httpsistemasgovbram
## httpwwwaneelgovbraplicacoesresumoestadualresumoestadualcfm
## httpwwwcanalenergiabrzpublishermateriasretrospectivaasp
## httpwwwgovbrmercadodocumentsresenhamensaldezembropdf
## ide
## idealiz
## iden
## importaca
## inciner
## incineraca
## industriaiscomerc
## industrialcomerc
## iner
## inf
## influenc
## instanc
## instig
## intermit
## internat
## ipeadat
## iri
## irrestrit
## ita
## itapirang
## ivinhem
## jaa
## jalil
## jamanxim
## janaub
## jen
## joic
## jonatan
## juni
```

```
## just
## karl
## ketlin
## konzen
## lacerd
## laje
## latitud
## laur
## learning
## legisl
## leonard
## light
## lik
## login
## longitud
## lpg
## lubricant
## lubrific
## lucel
## lug
## machin
## madr
## mainly
## mamanu
## manausam
## manipul
## manipulaca
## marcel
## maring
## materiaspr
## matricul
## mba
## mensur
## mercadolog
## metal
## meter
## mid
## milho
## millikan
## modernizaco
## mp
## naamazon
## naca
## naft
## naphth
## nav
## naveg
## navi
## need
## nerg
## nev
## norte
## observaco
## ocean
```

```
## od
## onquant
## oth
## pacot
## padro
## paran
## parecil
## passi
## peg
## pesquis
## pesquisaestud
## petroleum
## pi
## picon
## pigouvi
## pimesufp
## poligon
## populacaoveicul
## ppeaufop
## ppg
## ppgmn
## pretomg
## previsaodemand
## prezadx
## production
## prof
## profund
## progress
## propi
## proporca
## prorrogaca
## proviso
## prudentesp
## quantific
## quas
## queim
## rec
## recicl
## recuper
## refined
## refriger
## regiaobairr
## regioesest
## relac
## relaco
## relevanc
## representat
## requisica
## residencialindustr
## residu
## restaurant
## retrat
## retrofit
## rg
```

```
## riv
## rodrig
## rote
## rubr
## sa
## salient
## sanit
## santanasalvadorb
## servent
## shopping
## sinc
## sinte
## sirraul
## solciti
## solicital
## sous
## stand
## ston
## subbac
## sugesta
## sup
## supermerc
## switch
## sylvi
## szkl
## tabulaca
## tak
## tancred
## taquar
## tdr
## tema
## tenc
## tep
## teres
## terren
## that
## todav
## too
## topic
## totaliz
## trifas
## turqu
## uberland
## ufccaen
## ufgd
## ufpaicsafacecon
## ufpr
## ufsc
## ufu
## ufvjm
## unidirec
## unifacef
## univat
## universit
```

```
## upgn
## used
## usprp
## vali
## valorizaca
## varej
## variaca
## variaco
## vazaodat
## veget
## veraa
## veriq
## viabilizaca
## vic
## visu
## vo
## voip
## wat
## when
## will
## ach
## agenc
## alexandr
## atu
## biocombusti
## compreeend
## constitu
## correg
## decorrenc
## demonstr
## departament
## df
## direcion
## disposica
## dut
## edwig
## empreg
## enquadr
## ex
## execu
## fabr
## fer
## fernand
## fiz
## formaca
## frequenc
## fund
## gerenc
## implant
## indigen
## indispensa
## inic
## malh
## mauric
```

```
## membr
## mg
## microgeraca
## mor
## motiv
## parc
## sal
## saudaco
## shp
## simil
## sp
## tempor
## ufsm
## adequ
## administraca
## aeronav
## antema
## assemble
## august
## autoridad
## av
## avis
## bast
## bibliotec
## bioenerg
## bloc
## celul
## cep
## clar
## compet
## complet
## comunicaco
## concorrenc
## confidencial
## contextualiz
## continuaca
## corp
## correi
## deix
## descrica
## design
## dire
## dirig
## discre
## discriminaca
## distanc
## doming
## edifici
## entidad
## equival
## estrang
## estrateg
## ferrament
## fgv
```

```
## forc
## fundaca
## gabriel
## gerenci
## gerent
## getuli
## gnl
## gradu
## gui
## impress
## indenizaco
## inteligenc
## is
## lanc
## manifest
## mecan
## melac
## melh
## metr
## objet
## organ
## pe
## peric
## permission
## petrobr
## priv
## pront
## prov
## quadr
## quent
## redig
## refefer
## remuneraca
## restrit
## risc
## saud
## situaca
## sr
## tcu
## termin
## tribun
## turc
## urgent
## varg
## vind
## virtud
## visit
## volt
## web
## xlsx
## salari
## orgaoent
## locaca
## esport
```

```
## imovel
## patrocini
## pessoal
## vag
## carr
## incen
## prestaca
## fiscal
## profiss
## acuc
## comunicaca
## reajust
## conselh
## control
## gratificaca
## seleca
## anal
## audi
## contrataco
## direit
## jurid
## licitaco
## loc
## ocup
## organiz
## ouvid
## patrocin
## sed
## segur
## vincul
## adit
## advog
## cronolog
## demissa
## desp
## exclus
## fiscalizaca
## instruca
## organizaca
## pil
## possibilit
## quantit
## regr
## retir
## terceir
## terceirizaca
## aluguel
## conarq
## confianc
## defer
## estagia
## etanol
## funco
## labex
```

```
## ministr
## pad
## quaisqu
## republ
## visual
## apostil
## ato
## aven
## cole
## complement
## conform
## constituc
## continu
## convenco
## cooperaca
## corporat
## curricul
## desinfecca
## disciplin
## empresar
## ergonom
## espor
## exploraca
## farmac
## folh
## instaur
## instrument
## logis
## mao
## natalin
## penal
## reg
## repartica
## salar
## sediment
## seleco
## temporal
## tesour
## vig
## word
## almoxarif
## alug
## ape
## autoriz
## banh
## bqa
## caix
## certific
## cgu
## comprov
## comput
## contenci
## deciso
## decorr
```

```
## dl
## drog
## ed
## efet
## embrap
## exercici
## fas
## gar
## glob
## gom
## governament
## independ
## instituc
## iso
## lo
## noutr
## pda
## perspec
## plr
## preced
## prega
## prego
## principi
## refeica
## relatoriosintes
## renovaca
## resoluca
## semelh
## siasgweb
## sidec
## sobreprec
## soluco
## tonel
## validaca
## venh
## vid
## aca
## acid
## acolh
## acontec
## acorda
## administrativad
## advocac
## ago
## ajust
## alimentaca
## aline
## ambienteecolog
## anticorrupca
## api
## arbit
## arquitet
## arquivis
## ataqu
```

```
## audit
## aut
## aviaca
## barboz
## be
## beatriz
## bilho
## brasiliadf
## brazili
## caf
## carl
## cesgran
## chef
## cidada
## coaching
## coloc
## comit
## competenc
## concursadosterceirizadoscomission
## condica
## condomini
## condut
## confecca
## confirmaca
## contenh
## convocaca
## convocaco
## coordenad
## corporaca
## credenci
## cronogram
## csll
## curt
## defes
## deleg
## demisso
## deput
## descart
## descobert
## desenh
## dest
## destinaca
## devolv
## dieg
## discorr
## doenc
## dou
## down
## dpgr
## duraca
## ebserh
## eric
## estipul
## experienc
```

```
## exteri
## extern
## feri
## fiq
## flutu
## fotograf
## gisel
## gratuit
## hardw
## hierarqu
## honora
## hospital
## httpgovbrsitesptservidoresconcursospublicosdocumentscbaconcursopcbablicodaepesituacacaodasconvocacacl
## httpwwwwebserhgovbr
## igovt
## inclusiv
## inexistenc
## instauraca
## intranet
## invert
## investig
## irpj
## isabell
## jornal
## julg
## jus
## justifiqu
## lai
## ldo
## load
## locata
## medic
## mei
## ment
## nacionalinternac
## nil
## nomin
## nun
## ocupaca
## ozoni
## patrimon
## pav
## planocodig
## planoprogram
## pol
## premium
## prepar
## preservaca
## preven
## prevenca
## prim
## publicita
## qvt
## razo
```

```
## reda
## reflet
## reubl
## revi
## rodovia
## salvaguard
## secretarioassist
## sen
## sucumbenc
## sug
## sumul
## teletrabalh
## titulaca
## tst
## ufrg
## uspol
## validad
## vari
## vet
## viag
## vide
## vim
## vou
## wik
## abrangenc
## acompanhal
## act
## adequal
## adi
## adjunt
## admit
## adverbi
## advertenc
## afer
## afet
## afo
## afrobrasil
## agend
## agiliz
## alc
## alessi
## alienaca
## aliment
## alimentacaorefeica
## alter
## ampar
## analistasconsul
## ancor
## andaresconjunt
## anon
## antepost
## apf
## aposent
## archivisttoolkit
```

```
## arit
## arm
## arq
## arquvi
## artific
## artur
## assess
## assessor
## assin
## atest
## atom
## atualil
## atualiz
## atualizaco
## aul
## ausent
## autentic
## autom
## backup
## bande
## bdi
## beb
## beneficia
## bimestr
## biomet
## blackenergy
## bm
## bnd
## bovesp
## brind
## busqu
## cade
## cala
## cancel
## carenc
## cargofunca
## carto
## cdfunca
## cedric
## celebr
## celet
## cesam
## check
## chilen
## ciberne
## cipacissp
## cipacomissa
## cissp
## civil
## classif
## cobert
## colabor
## colocaca
## comand
```

```
## comenta
## comission
## company
## comparaco
## compil
## compre
## compressa
## conarqctd
## conf
## confiancacarg
## confus
## congen
## conjugaca
## conso
## constanci
## consul
## contatal
## contavel
## contemplaca
## contentdm
## contratual
## convit
## coordenaca
## correco
## correlat
## cpf
## cuid
## cumpr
## custe
## danill
## david
## dedic
## defen
## defici
## definitivadermatos
## delivery
## demandara
## denunc
## departamentoslaborato
## dependenc
## descumpr
## desembarg
## desembols
## desestimul
## desfaz
## desvi
## deu
## difusa
## discretizaca
## discussa
## dispensadosexoner
## dispensaexoneraca
## dispers
## div
```

```
## divers
## domes
## dsic
## ecologic
## econometr
## efici
## element
## eletronic
## elev
## emanuel
## empa
## empenh
## empregadodirig
## empresaentidadeorga
## encerr
## enfim
## enforc
## enriquec
## ensin
## entrev
## epoc
## escop
## estatuta
## estimul
## estrit
## estu
## exam
## exat
## exclu
## expect
## expl
## falh
## farmaciaval
## fech
## ferna
## fidalg
## flexi
## fme
## formula
## formulaca
## forncec
## fracking
## fraud
## frot
## fur
## fut
## gavet
## ged
## geolocalizaca
## geolog
## governanc
## guerr
## havel
## hist
```

```
## homossex
## httpgovbrptimprensanoticiasdecidereformularconcurs
## httphdlhandlenet
## httpsbengovbrdownloadscadntesedorelatcbriofinalwebpdf
## httpsgisepeprdgovbrwebmapep
## httpswwwcomprasgovernamentaisgovbrindexphplegislacaoinstrucoesnormativasinstrucaonormativadezembr
## httpswwwnovacanaindustriainvestimentobiocombustiveisdemandamprojet
## httpwwwgovbrpdeeformsepeestudoaspx
## httpwwwgovbrptabcdenergiaplanejamentoenerg
## httpwwwgovbrptpublicacoesabertospublicacoesconsumomensaleletricaclasseregioessubsistem
## human
## icaatom
## ido
## incid
## inicop
## injustific
## inscrev
## inser
## instrumentalizaca
## intenca
## interi
## investigaca
## iptu
## janain
## jeton
## jorn
## jos
## judic
## juiz
## jun
## kim
## lauraramosifmsedubr
## layout
## lerdortantracosebissinosesurd
## letr
## lin
## liquidaca
## locaco
## locat
## louc
## lum
## malefici
## malw
## marcu
## marluc
## metropol
## microond
## microsot
## milen
## milh
## minigeraca
## minor
## missa
## mist
```

```
## mix
## monet
## moss
## mpdg
## mpog
## muit
## nalgum
## natal
## nomeaco
## nomenclat
## notpety
## oci
## ocupacionalsiderosecataratadoenc
## omit
## oportun
## opt
## orcament
## orcamenta
## orcamentariofinanc
## orden
## ordin
## organogram
## orgaosent
## orientaco
## origin
## pac
## palh
## panoram
## parnaib
## pbt
## pcdccd
## pdfdoc
## pend
## pequen
## percorr
## perit
## pety
## planinvest
## plas
## plena
## poc
## poduca
## pontual
## posico
## pre
## precav
## preestabelec
## prelanc
## premenc
## premiaca
## preparaca
## presidentevicepresidentesdiretoressuperintendentesger
## prevenco
## previdenc
```

```
## prezaod
## privatizaca
## probabil
## profer
## profissa
## proprieta
## proteg
## providenc
## providenci
## psicosoc
## publicidadepublicaco
## qua
## qualificacaoaperfeico
## quarent
## radial
## ransomw
## rb
## rdcarq
## reconcav
## reembols
## reformul
## regulaco
## remuneraco
## renom
## repart
## reposito
## representaco
## requer
## resciso
## resolucaotcu
## respe
## respeitos
## restabelec
## retribuica
## retro
## revog
## ricard
## robert
## rogeri
## ruid
## salariobas
## sane
## secretosigil
## segred
## sensi
## servidoresempreg
## shapfil
## sian
## sib
## sigad
## significativ
## situac
## slid
## socialcpf
```

```
## startup
## submet
## subordin
## subprocuradoresger
## substituica
## suficienteseficaz
## sujeit
## sumar
## supeir
## susan
## suspensa
## tampouc
## tempora
## terrestr
## tessarin
## tid
## toe
## tpa
## trac
## trag
## transferenc
## transgener
## trf
## tripulaca
## tsv
## ttdttdd
## tubulaca
## uasg
## unb
## unidadegerenciadiretoriasuperintendenc
## vacanc
## ved
## verb
## vicepresid
## vidr
## violaca
## violent
## vot
## wannacry
## whatsapp
## willian
## xlsxlsx
## zone
```

```
    z=0
    for (k in 1:dim(fe)[2]) {
      if (colSums(fe)[k] <= mediana1) {
        exclui_termos[z] <- colnames(fe)[k]
        z = z+1
      }
    }

#length(exclui_termos) # [1] 409 [1] 2118
```

```r
cat(paste0("Existem ", length(exclui_termos), " termos com freq. menor ou igual a mediana (", mediana1,
```

```
## Existem 1521 termos com freq. menor ou igual a mediana (2). Logo, se removermos estes o número de va
```

```r
    fe <- fe %>% select(-exclui_termos)
```

```r
cat(paste0("Existem, agora, ", dim(fe)[2], " termos/palavras-chaves únicas. E a nova mediana da frequên
```

```
## Existem, agora, 1511 termos/palavras-chaves únicas. E a nova mediana da frequência de termos restant
```

Critério de escolha dos termos, se a frequência for maior ou igual a 10

IMPLEMENTAR

```r
highchart() %>%
  hc_add_series(data = NumTermos$Num_Pedidos,
                type = "bar",
                name = "# de pedidos",
                showInLegend = FALSE,
                tooltip = list(valueDecimals = 0, valuePrefix = "", valueSuffix = ""), color="blue") %>%
  hc_yAxis(title = list(text = "Quantitativo de pedidos"),
           allowDecimals = TRUE, max = (max(NumTermos$Num_Pedidos)+103),
           labels = list(format = "{value}")) %>%
  hc_xAxis(title = list(text = "Termo"),
           categories = NumTermos$termo,
           tickmarkPlacement = "on",
           opposite = FALSE) %>%
  hc_title(text = "Quantitativo de pedidos por termo (sem exclusividade)",
           style = list(fontWeight = "bold")) %>%
  hc_subtitle(text = paste("")) %>%
      hc_tooltip(valueDecimals = 2,
                 pointFormat = "{point.y} pedidos")%>%
                 #pointFormat = "Variável: {point.x} <br> Missing: {point.y}")
      hc_credits(enabled = TRUE,
                 text = "Fonte: CGU, e-SIC (2019). Elaboração: Ewerson Pimenta.",
                 style = list(fontSize = "10px")) %>%
  hc_exporting(enabled = TRUE, filename = "F3-filmes-genero-Pimenta")
```

```r
db_modelo0 = as_tibble(cbind(select(DB,Protocolo, DATA_REGISTRO, DIRETORIAS, DIRETORIA),fe))
db_modelo = as_tibble(cbind(select(DB,DIRETORIA),fe))
```

```r
# __Porcentagem de ZEROS por variável__

zeros <- (colSums(fe==0)/nrow(fe)*100); var <- names(fe)
db_zero <- data.frame(var,zeros); rownames(db_zero) <- NULL
db_zero <- db_zero[order(db_zero$zeros, decreasing = TRUE), ]

hc4_1 <- highchart() %>%
  hc_add_series(data = db_zero$zeros,
                type = "bar",
                name = "Porcentagem de zeros",
                showInLegend = FALSE,
                tooltip = list(valueDecimals = 2, valuePrefix = "", valueSuffix = " %"), color="pink") %
  hc_yAxis(title = list(text = "Porcentagem de zero"),
           allowDecimals = TRUE, max = 100,
```

```
          labels = list(format = "{value}%")) %>%
  hc_xAxis(categories = db_zero$var,
           tickmarkPlacement = "on",
           opposite = FALSE) %>%
  hc_title(text = "Porcentagem de zeros por variável",
           style = list(fontWeight = "bold")) %>%
  hc_subtitle(text = paste("")) %>%
      hc_tooltip(valueDecimals = 2,
                 pointFormat = "Zeros: {point.y}")%>%
                 #pointFormat = "Variável: {point.x} <br> Missing: {point.y}")
      hc_credits(enabled = TRUE,
                 text = "Fonte: IMDB/KAGGLE. Elaboração: Ewerson Pimenta.",
                 style = list(fontSize = "10px")) %>%
  hc_exporting(enabled = TRUE, filename = "Fig00-Pimenta")
#hc <- hc %>%
#  hc_add_theme(hc_theme_darkunica())
hc4_1; remove(hc4_1, var, zeros)
```

Modelos de classificação

Partição dos dados

Particionando a base de dados em Treino e Teste, esses dois (Treino e Teste) também terão armazenados as diretorias que foram responsaveis por cada pedido via amostragem probabilística dos dados originais separadamente das bases de Treino e Teste.

```
#db_modelo = as_tibble(cbind(select(DB,DIRETORIA),fe))
#getwd()
#setwd("/Users/ewersonpimenta/Desktop/ESIC_TCC/TCC_v2.1/RMARKDOWN/WEB_APP/")
#write.csv(db_modelo0, file = "db_modelo_rf_v21.csv", row.names = FALSE)
#db_modelo = read.csv("db_modelo_rf_v10.csv",header = T); dim(db_modelo)
#db_modelo = db_modelo %>% select(-r,-venc)
#write.csv(db_modelo, file = "db_modelo_rf_v11.csv", row.names = FALSE)
db_modelo$DIRETORIA <- as.factor(db_modelo$DIRETORIA)
#levels(db_modelo$DIRETORIA)
```

Para amostragem aleatória simples

```
set.seed(098798) # 756446 ou 75452 (OOB_erro: 35,63% ACC: 65,44%) # 2967 (OOB_erro: 34,15% ACC: 64,98%)
intrain <- createDataPartition(y = db_modelo$DIRETORIA, p = 0.65, list = FALSE)
training <- db_modelo[intrain,]
testing <- db_modelo[-intrain,]
```

Modelagem 1 - Random Forest (RF)

Random Forest (RF) - Metodologia

Descrição 1. Random Forest foi desenvolvido para agregar árvores de decisão (modelo de classificação);
2. Pode ser usado para modelo de classificação (p/ var. resposta categórica) ou regressão (no caso de haver variável resposta contínua);
3. Evita *overfitting*;
4. Permite trabalhar com um largo número de características de um conjunto de dados;
5. Auxilia na seleção de variáveis baseada em um algoritmo que calcula a importância por variável (assim,

tendo conhecimento de quais variáveis são mais importantes, podemos usar essa informação para outros modelos de classificação);
6. User-friendly: apenas 2 parâmetros livres:

- Trees - ntrees, default 500 (Nº de árvores);
- Variáveis selecionadas via amostragem aleatória candidatas à cada "split" (quebra da árvore) - mtry, default $\sqrt{p}$ p/ classificação e $\frac{p}{3}$ p/ regressão (p: nº de features/variáveis);

Passo-a-Passo

É realizado em 3 passos:

1. Desenha as amostras via bootstrap do número de árvores *ntrees*;

2. Para cada amostra via bootstrap, cresce o número de árvores "un-puned" para a escolha da melhor quebra da árvore baseado na amostra aleatória do valor predito de mtry a cada nó da árvore;

- 3. Faz classificação de novos valores usando a maioria de votos p/ classificação e usa a média p/ regressão baseada nas amostras de ntrees.

Random Forest - Aplicação e Resultados

Inicialmente utilizaremos o pacote `randomForest` que implmenta o algoritmo de Random Forest de Breiman (baseado na clusterização de Breiman, originalmente codificada em Fortran) que tem por finalidade classificar e/ou criar regressão. Além disso, pode ser usado em um modelo não supervisionado para avaliar proximidades entre pontos.

Estamos usando, a partir daqui, a base de treino.

```
#library(randomForest)
#library(rpart)
#library(rpart.plot)
#rf <- randomForest(proximity = T,ntree = 38,do.trace = T,WR~.,data=training)
set.seed(9984512)
# Training with classification tree
rf <- rpart(DIRETORIA ~ ., data=training, method="class", xval = 4, )
print(rf, digits = 3)
```

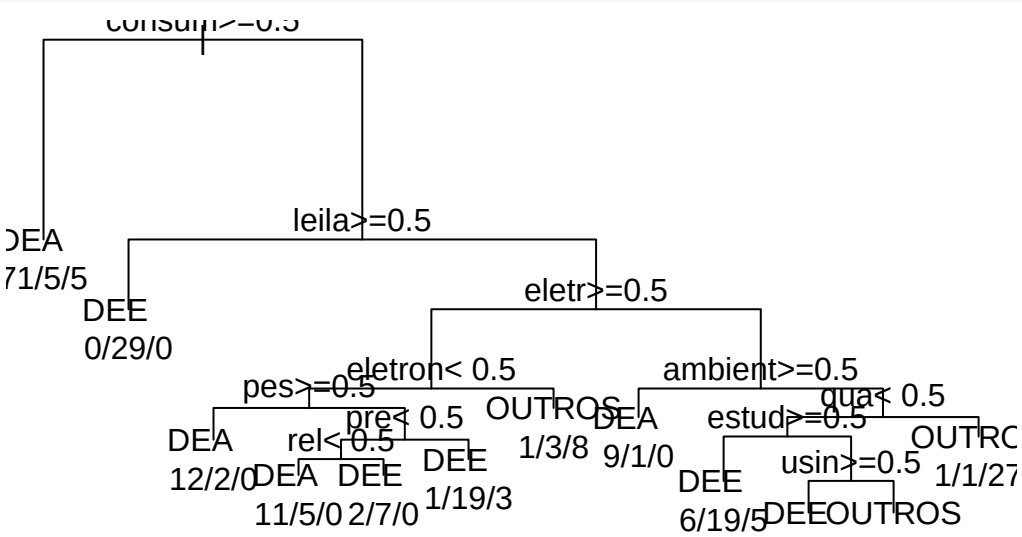
```
## n= 407
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##   1) root 407 267 DEA (0.3440 0.3366 0.3194)
##     2) consum>=0.5 81  10 DEA (0.8765 0.0617 0.0617) *
##     3) consum< 0.5 326 194 DEE (0.2117 0.4049 0.3834)
##       6) leila>=0.5 29   0 DEE (0.0000 1.0000 0.0000) *
##       7) leila< 0.5 297 172 OUTROS (0.2323 0.3468 0.4209)
##        14) eletr>=0.5 74  38 DEE (0.3649 0.4865 0.1486)
##          28) eletron< 0.5 62  29 DEE (0.4194 0.5323 0.0484)
##            56) pes>=0.5 14   2 DEA (0.8571 0.1429 0.0000) *
##            57) pes< 0.5 48  17 DEE (0.2917 0.6458 0.0625)
##             114) pre< 0.5 25  12 DEA (0.5200 0.4800 0.0000)
##               228) rel< 0.5 16   5 DEA (0.6875 0.3125 0.0000) *
##               229) rel>=0.5 9   2 DEE (0.2222 0.7778 0.0000) *
##             115) pre>=0.5 23   4 DEE (0.0435 0.8261 0.1304) *
```

```
##          29) eletron>=0.5 12   4 OUTROS (0.0833 0.2500 0.6667) *
##        15) eletr< 0.5 223 109 OUTROS (0.1883 0.3004 0.5112)
##          30) ambient>=0.5 10   1 DEA (0.9000 0.1000 0.0000) *
##          31) ambient< 0.5 213  99 OUTROS (0.1549 0.3099 0.5352)
##            62) qua< 0.5 184  97 OUTROS (0.1739 0.3533 0.4728)
##             124) estud>=0.5 30  11 DEE (0.2000 0.6333 0.1667) *
##             125) estud< 0.5 154  72 OUTROS (0.1688 0.2987 0.5325)
##               250) usin>=0.5 14   3 DEE (0.1429 0.7857 0.0714) *
##               251) usin< 0.5 140  59 OUTROS (0.1714 0.2500 0.5786) *
##            63) qua>=0.5 29   2 OUTROS (0.0345 0.0345 0.9310) *
```

```r
attributes(rf)
```

```
## $names
##  [1] "frame"               "where"               "call"
##  [4] "terms"               "cptable"             "method"
##  [7] "parms"               "control"             "functions"
## [10] "numresp"             "splits"              "variable.importance"
## [13] "y"                   "ordered"
##
## $xlevels
## named list()
##
## $ylevels
## [1] "DEA"    "DEE"    "OUTROS"
##
## $class
## [1] "rpart"
```

```r
plot(rf)
text(rf, use.n = TRUE)
```




```r
# Predict the testing set with the trained model
predictions <- predict(rf, testing, type = "class")

# Accuracy and other metrics
confusionMatrix(predictions, as.factor(testing$DIRETORIA))
```

```
## Confusion Matrix and Statistics
```

```
##
##           Reference
## Prediction DEA DEE OUTROS
##     DEA     47  14      9
##     DEE      6  32      8
##     OUTROS  21  27     53
##
## Overall Statistics
##
##                Accuracy : 0.6083
##                  95% CI : (0.5399, 0.6737)
##     No Information Rate : 0.341
##     P-Value [Acc > NIR] : 8.746e-16
##
##                   Kappa : 0.4141
##
##  Mcnemar's Test P-Value : 0.0003788
##
## Statistics by Class:
##
##                      Class: DEA Class: DEE Class: OUTROS
## Sensitivity              0.6351     0.4384        0.7571
## Specificity              0.8392     0.9028        0.6735
## Pos Pred Value           0.6714     0.6957        0.5248
## Neg Pred Value           0.8163     0.7602        0.8534
## Prevalence               0.3410     0.3364        0.3226
## Detection Rate           0.2166     0.1475        0.2442
## Detection Prevalence     0.3226     0.2120        0.4654
## Balanced Accuracy        0.7371     0.6706        0.7153
```

Olhando as 6 primeiras observações real X predito

```r
p1 <- predict(rf,training)
head(p1)
```

```
##         DEA       DEE    OUTROS
## 1 0.1714286 0.2500000 0.5785714
## 2 0.8765432 0.0617284 0.0617284
## 3 0.2000000 0.6333333 0.1666667
## 4 0.8765432 0.0617284 0.0617284
## 5 0.1714286 0.2500000 0.5785714
## 6 0.9000000 0.1000000 0.0000000
```

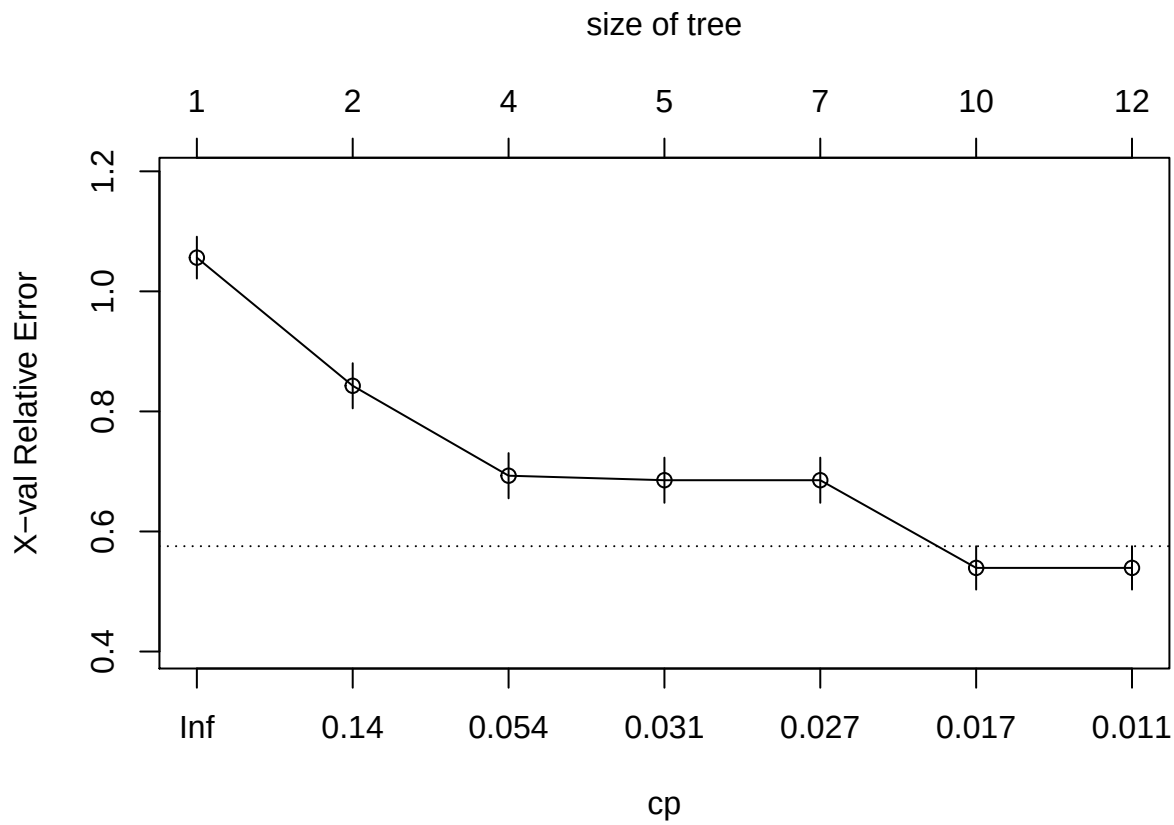
```r
head(training$DIRETORIA)
```

```
## [1] OUTROS DEA    DEE    DEA    OUTROS DEA
## Levels: DEA DEE OUTROS
```

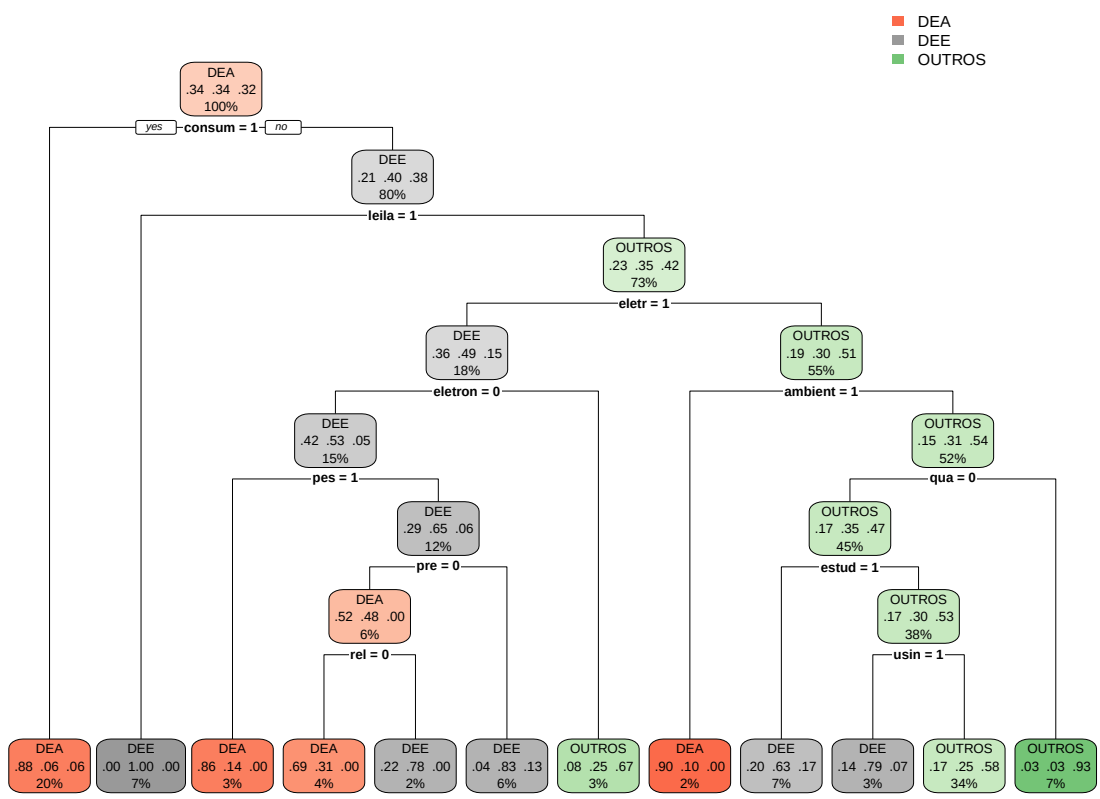
Selecionando uma árvore

```r
rp <- rpart::rpart(formula = DIRETORIA~.,data=training)
```

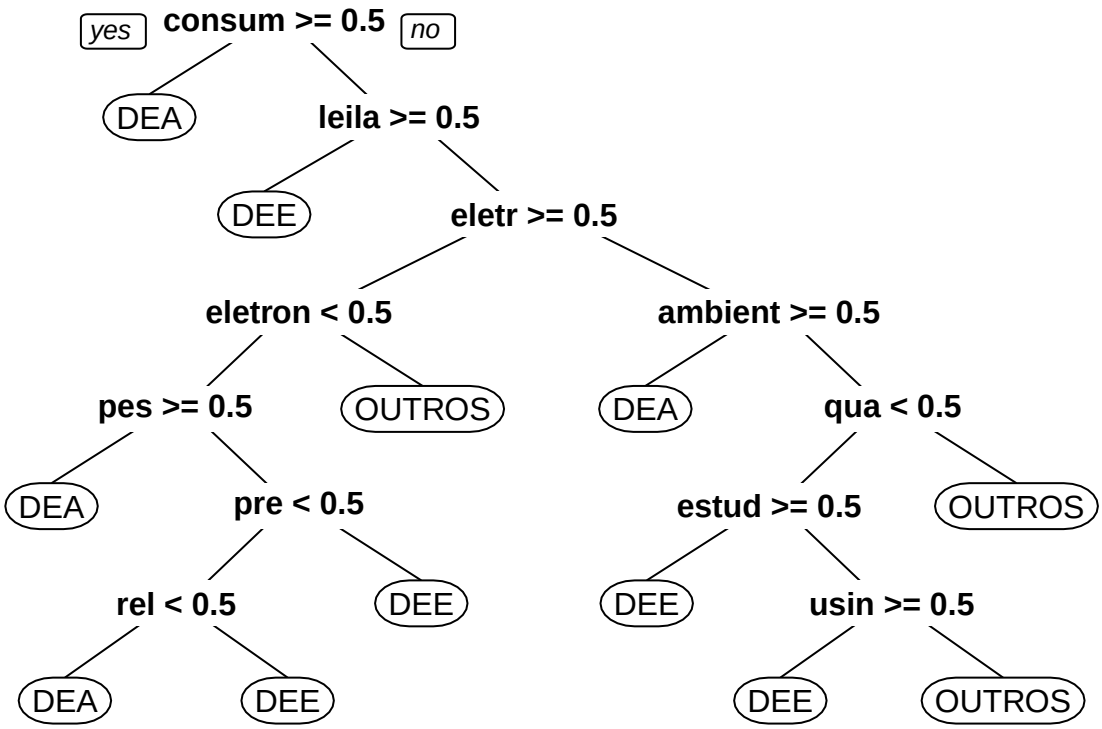
```r
rpart::plotcp(rf)
```




```r
rpart.plot(rf)
```



```r
rpart.plot.version1(rf)
```



Outra forma de escrever o modelo é usando a função `randomForest( )`

```r
set.seed(09986755)
rf1 <- randomForest(as.factor(DIRETORIA) ~ ., data=training,
                    importance = TRUE,
                    proximity = TRUE)
rf1
```

```
##
## Call:
##  randomForest(formula = as.factor(DIRETORIA) ~ ., data = training,      importance = TRUE, proximity
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 38
##
##         OOB estimate of  error rate: 25.31%
## Confusion matrix:
##        DEA DEE OUTROS class.error
## DEA     97  24     19   0.3071429
## DEE     18 101     18   0.2627737
## OUTROS   9  15    106   0.1846154
```

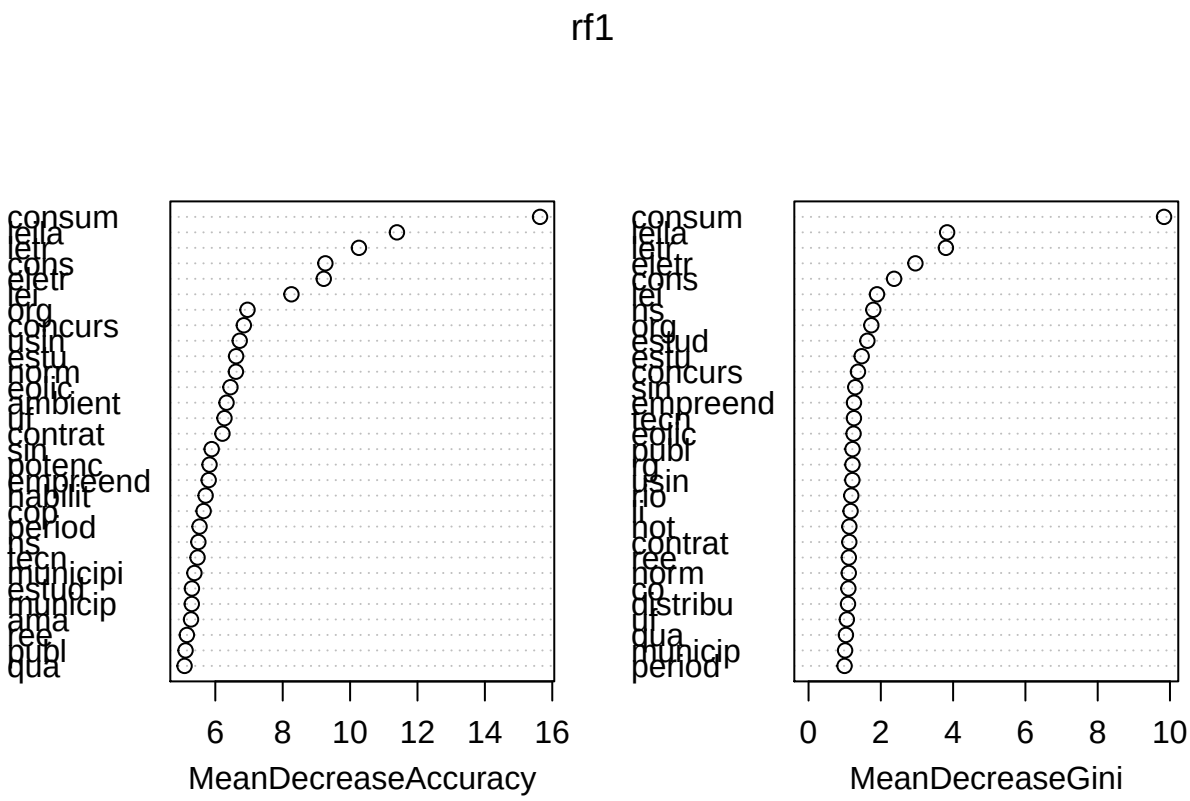
```r
# Predict the testing set with the trained model
predictions1 <- predict(rf1, testing, type = "class")

# Accuracy and other metrics
confusionMatrix(predictions1, as.factor(testing$DIRETORIA))
```

```
## Confusion Matrix and Statistics
##
```

```
##           Reference
## Prediction DEA DEE OUTROS
##     DEA     45  10      7
##     DEE     17  52     10
##     OUTROS  12  11     53
##
## Overall Statistics
##
##                Accuracy : 0.6912
##                  95% CI : (0.6252, 0.752)
##     No Information Rate : 0.341
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.5372
##
##  Mcnemar's Test P-Value : 0.365
##
## Statistics by Class:
##
##                      Class: DEA Class: DEE Class: OUTROS
## Sensitivity              0.6081     0.7123        0.7571
## Specificity              0.8811     0.8125        0.8435
## Pos Pred Value           0.7258     0.6582        0.6974
## Neg Pred Value           0.8129     0.8478        0.8794
## Prevalence               0.3410     0.3364        0.3226
## Detection Rate           0.2074     0.2396        0.2442
## Detection Prevalence     0.2857     0.3641        0.3502
## Balanced Accuracy        0.7446     0.7624        0.8003
```

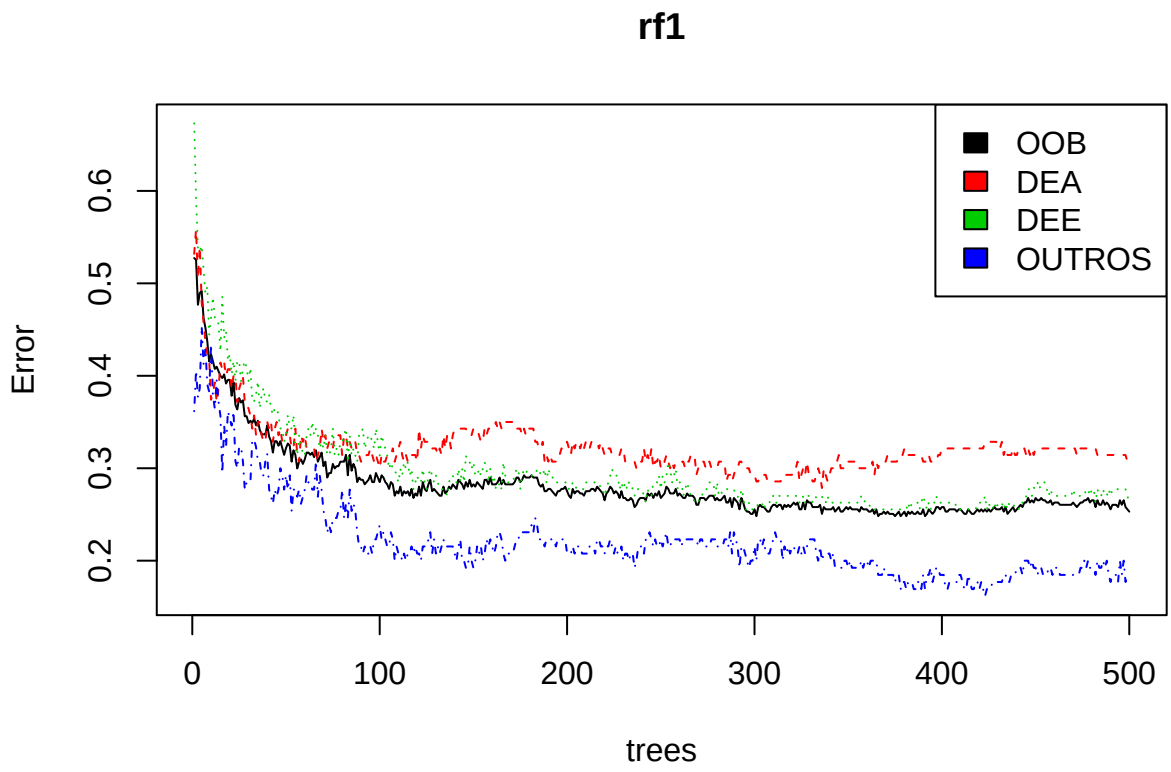
Importância de variáveis

```
RF_importance = randomForest::importance(rf1)[order(randomForest::importance(rf1)[,1], decreasing = TRU
randomForest::varImpPlot(rf1)
```




```r
{r, out.width = "400px",echo = FALSE, eval = TRUE, message=FALSE, include = TRUE}
knitr::include_graphics(paste0(PATH,"IMAGENS/RF_var_importance.png"))
```

```r
plot(rf1)
legend('topright', colnames(rf1$err.rate), col=1:5, fill=1:5)
```



A partir de $n = 420$ árvores a taxa do erro **OOB (Out of Bag)** tende a estabilizar.

Tuning do modelo

Fixando, então, $n = 420$ árvores

Aparentemente $mtry = 26$ parece ser um bom palpite para o segundo parâmetro do random forest, uma vez que esse retornou menor taxa de erro **OOB**, $27,03\%$. Entretanto esse erro ainda é muito alto. Vamos reescrever o modelo com os parâmetros tunados.

Aparentemente $mtry = 28$ parece ser um bom palpite para o segundo parâmetro do random forest, uma vez que esse retornou menor taxa de erro **OOB**, $26,54\%$. Entretanto esse erro ainda é muito alto. Vamos reescrever o modelo com os parâmetros tunados.

Aparentemente $mtry = 38$ parece ser um bom palpite para o segundo parâmetro do random forest, uma vez que esse retornou menor taxa de erro **OOB**, $26,54\%$. Entretanto esse erro ainda é muito alto. Vamos reescrever o modelo com os parâmetros tunados.

```
set.seed(09986755)
rf2 <- randomForest(as.factor(DIRETORIA) ~ ., data=training,
                    ntree = 420,
                    mtry = 38,
                    importance = TRUE,
                    proximity = TRUE)
rf2
```

```
##
## Call:
##  randomForest(formula = as.factor(DIRETORIA) ~ ., data = training,      ntree = 420, mtry = 38, impo
##                Type of random forest: classification
##                      Number of trees: 420
## No. of variables tried at each split: 38
```

```
##
##         OOB estimate of  error rate: 25.31%
## Confusion matrix:
##        DEA DEE OUTROS class.error
## DEA     95  26     19   0.3214286
## DEE     17 101     19   0.2627737
## OUTROS   8  14    108   0.1692308
```

```r
# Predict the testing set with the trained model
predictions2 <- predict(rf2, testing, type = "class")

# Accuracy and other metrics
(rf2_CONFUSIONM = confusionMatrix(predictions2, as.factor(testing$DIRETORIA)))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction DEA DEE OUTROS
##     DEA     44  10      6
##     DEE     18  53     11
##     OUTROS  12  10     53
##
## Overall Statistics
##
##                Accuracy : 0.6912
##                  95% CI : (0.6252, 0.752)
##     No Information Rate : 0.341
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.5372
##
##  Mcnemar's Test P-Value : 0.2276
##
## Statistics by Class:
##
##                      Class: DEA Class: DEE Class: OUTROS
## Sensitivity              0.5946     0.7260        0.7571
## Specificity              0.8881     0.7986        0.8503
## Pos Pred Value           0.7333     0.6463        0.7067
## Neg Pred Value           0.8089     0.8519        0.8803
## Prevalence               0.3410     0.3364        0.3226
## Detection Rate           0.2028     0.2442        0.2442
## Detection Prevalence     0.2765     0.3779        0.3456
## Balanced Accuracy        0.7414     0.7623        0.8037
```

```r
p2 <- predict(rf2,training)
as.character(head(p2))
```

```
## [1] "OUTROS" "DEA"    "DEE"    "DEA"    "OUTROS" "DEA"
```

```r
head(training$DIRETORIA)
```

```
## [1] OUTROS DEA    DEE    DEA    OUTROS DEA
## Levels: DEA DEE OUTROS
```

```r
(DEA_erroCLASS = sum(rf2_CONFUSIONM$table[1,2:3])/ sum(rf2_CONFUSIONM$table[1,]))
```

```
## [1] 0.2666667
```

```
(DEE_erroCLASS = sum(rf2_CONFUSIONM$table[2,c(1,3)])/ sum(rf2_CONFUSIONM$table[2,]))
```

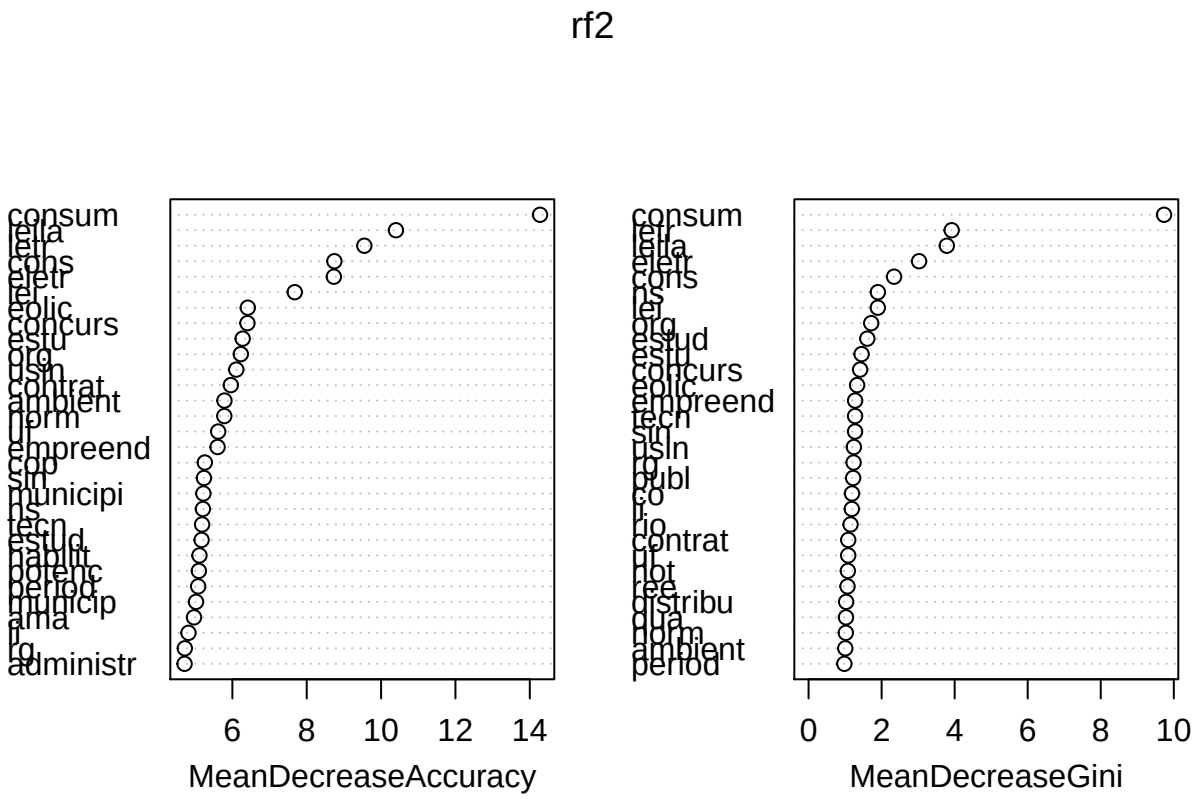
```
## [1] 0.3536585
```

```
(OUTROS_erroCLASS = sum(rf2_CONFUSIONM$table[3,1:2])/ sum(rf2_CONFUSIONM$table[3,]))
```

```
## [1] 0.2933333
```

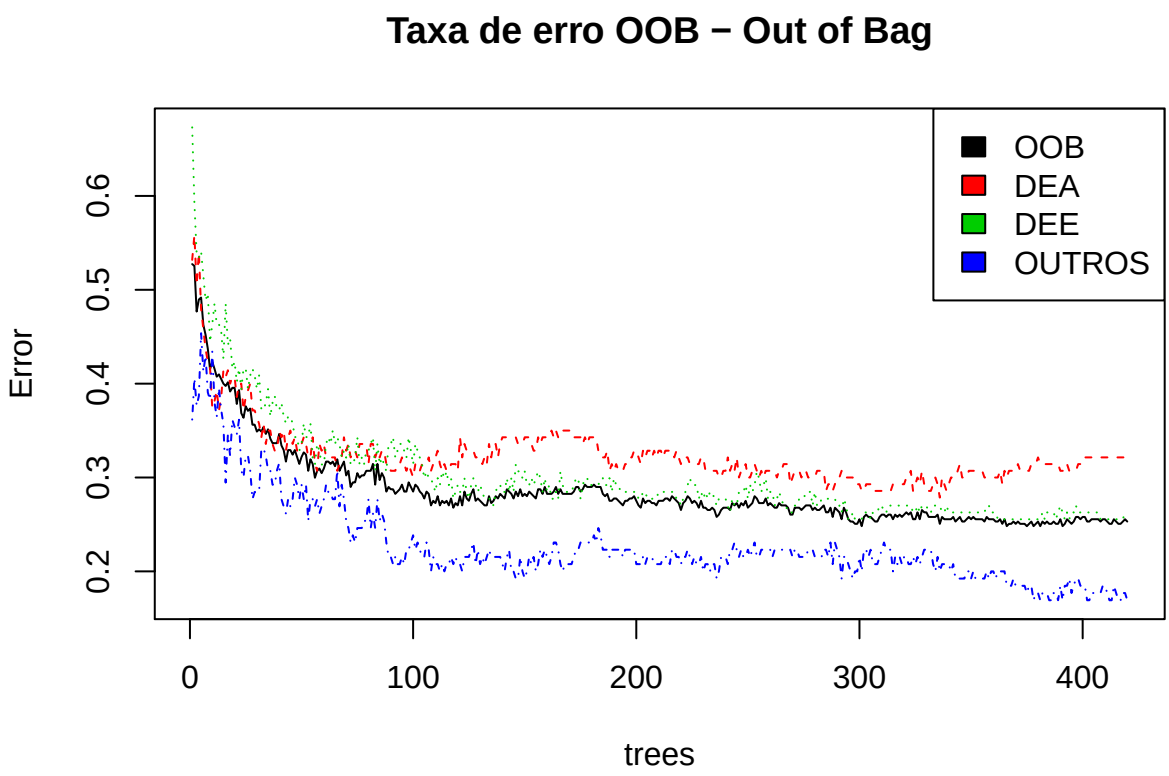
Acurácia de aproximadamente 72% na base de teste. E as taxas de erro de classificação foram 30%, 44% e 45% para *DEA*, *DEE* e *OUTROS*, respectivamente. Houve um melhor desempenho na classificação do modelo para a categoria *DEA*

```
RF_importance = randomForest::importance(rf2)[order(randomForest::importance(rf2)[,1], decreasing = TRU
randomForest::varImpPlot(rf2)
```



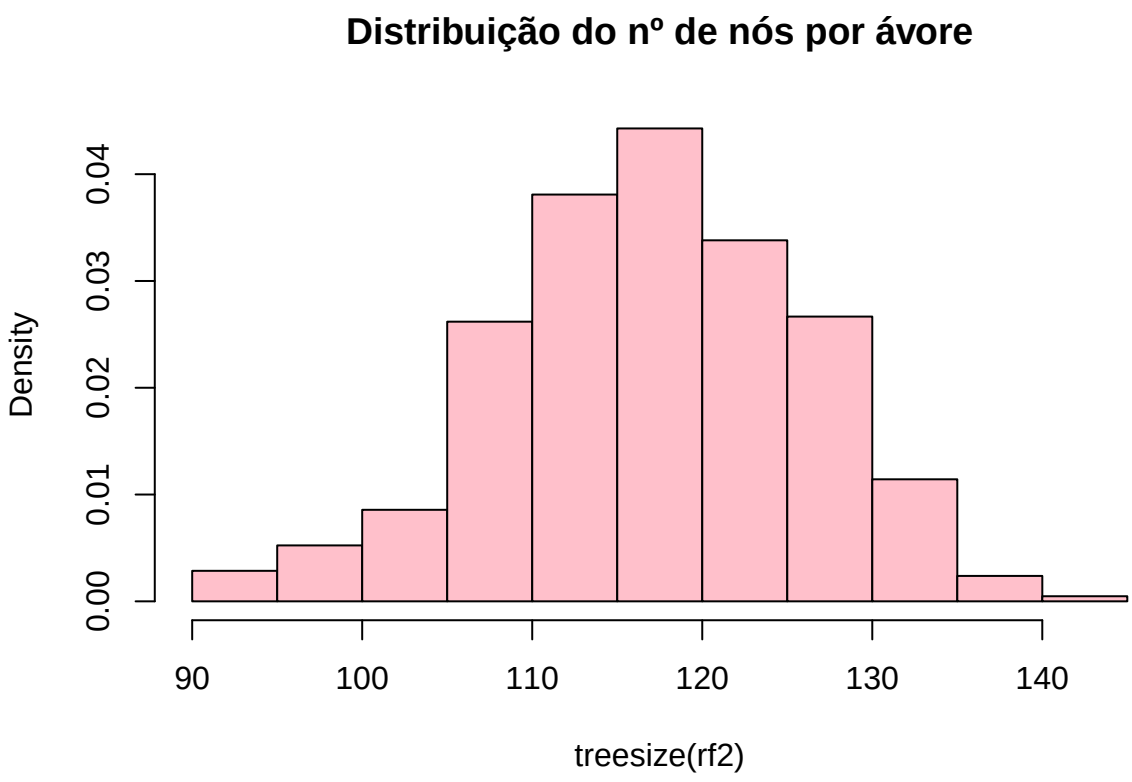
Taxa de Erro Random Forest

```
plot(rf2, main = "Taxa de erro OOB - Out of Bag")
legend('topright', colnames(rf2$err.rate), col=1:5, fill=1:5)
```



Histograma do Número de nós por árvore

```
hist(treesize(rf2), probability = T,
     main = "Distribuição do nº de nós por ávore",
     col = "pink")
```



Vamos excluir as variáveis que não retornaram valor de importância para o algoritmo do random forest.

```
RF_importance = randomForest::importance(rf2)[order(randomForest::importance(rf2)[,1], decreasing = TRU
```

```
RF1 = data.frame(variables = rownames(RF_importance), importance = RF_importance[,4])
RF1 = RF1[order(RF1$importance, decreasing = TRUE),]
rownames(RF1) <- NULL
#summary(RF1)
```

```
RF2 = RF1[1:20,]
#library("highcharter")
hc6_1 <- highchart() %>%
  hc_add_series(data = RF2$importance,
                type = "bar",
                name = "Importância",
                showInLegend = FALSE,
                tooltip = list(valueDecimals = 2, valuePrefix = "", valueSuffix = "")) %>%
  hc_yAxis(title = list(text = "Importância"),
           allowDecimals = TRUE, max = 12,
           labels = list(format = "{value}")) %>%
  hc_xAxis(title = list(text = "Fatores"),
           categories = RF2$variables,
           tickmarkPlacement = "on",
           opposite = FALSE) %>%
  hc_title(text = "Importância por fator - Random Forest",
           style = list(fontWeight = "bold")) %>%
  hc_subtitle(text = paste("")) %>%
      hc_tooltip(valueDecimals = 2,
                 pointFormat = "Importância: {point.y}")%>%
                 #pointFormat = "Variável: {point.x} <br> Importância: {point.y}")
      hc_credits(enabled = TRUE,
                 text = "Fonte: CGU, e-SIC. Elaboração: Leal, Alize; Pimenta, Ewerson.",
                 style = list(fontSize = "10px")) %>%
  hc_exporting(enabled = TRUE, filename = "F6_1-importance-Pimenta")
#hc <- hc %>%
#  hc_add_theme(hc_theme_darkunica())
hc6_1
```

Vamos excluir todas as variáveis que retornaram importância menor ou igual a zero.

```
variaveis_sem_importancia = RF1 %>% filter(as.character(importance) <= 0)
#summary(variaveis_sem_importancia)
variaveis_sem_importancia = as.character(variaveis_sem_importancia$variables)
```

```
training1 = training %>% select(-(variaveis_sem_importancia))
testing1 = testing %>% select(-(variaveis_sem_importancia))
db_modelo1 = db_modelo %>% select(-(variaveis_sem_importancia))
#DB_HISTORICO <- db_modelo0 %>% select(-(variaveis_sem_importancia))
#write.csv(DB_HISTORICO, file = "DB_LAI-EPE_HISTORICO.csv", row.names = FALSE)
```

```
cat(paste0("O número de variáveis da base de históricos (cheia) reduziu de ", dim(training)[2], " para
```

```
## O número de variáveis da base de históricos (cheia) reduziu de 1512 para 790.
```

```
set.seed(09986755) #2967
rf3 <- randomForest(as.factor(DIRETORIA) ~ ., data=training1,
                    ntree = 420,
```

```r
                    mtry = 28,
                    importance = TRUE,
                    proximity = TRUE)
rf3
```

```
##
## Call:
##  randomForest(formula = as.factor(DIRETORIA) ~ ., data = training1,      ntree = 420, mtry = 28, impo
##                Type of random forest: classification
##                      Number of trees: 420
## No. of variables tried at each split: 28
##
##         OOB estimate of  error rate: 26.78%
## Confusion matrix:
##        DEA DEE OUTROS class.error
## DEA     95  29     16   0.3214286
## DEE     19  99     19   0.2773723
## OUTROS   7  19    104   0.2000000
```

```r
# Predict the testing set with the trained model
predictions3 <- predict(rf3, testing1[,-1], type = "class")
#predict(rf3, testing1[,-1], type = "prob")

# Accuracy and other metrics
confusionMatrix(predictions3, as.factor(testing1$DIRETORIA))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction DEA DEE OUTROS
##     DEA     46   9      6
##     DEE     17  53     10
##     OUTROS  11  11     54
##
## Overall Statistics
##
##                Accuracy : 0.7051
##                  95% CI : (0.6396, 0.7649)
##     No Information Rate : 0.341
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.5579
##
##  Mcnemar's Test P-Value : 0.2637
##
## Statistics by Class:
##
##                      Class: DEA Class: DEE Class: OUTROS
## Sensitivity              0.6216     0.7260        0.7714
## Specificity              0.8951     0.8125        0.8503
## Pos Pred Value           0.7541     0.6625        0.7105
## Neg Pred Value           0.8205     0.8540        0.8865
## Prevalence               0.3410     0.3364        0.3226
## Detection Rate           0.2120     0.2442        0.2488
## Detection Prevalence     0.2811     0.3687        0.3502
```

```
## Balanced Accuracy        0.7584     0.7693        0.8109
```

```
p3 <- predict(rf3,training1)
as.character(head(p3))
```

```
## [1] "OUTROS" "DEA"    "DEE"    "DEA"    "OUTROS" "DEA"
```

```
head(training1$DIRETORIA)
```

```
## [1] OUTROS DEA    DEE    DEA    OUTROS DEA
## Levels: DEA DEE OUTROS
```

```
matriz_conf <- confusionMatrix(predictions3, as.factor(testing1$DIRETORIA))
matriz_conf$overall[1]
```

```
##  Accuracy
## 0.7050691
```

```
cat(paste0("A taxa de erro OOB do modelo final é de ", round(mean(colMeans(rf3$err.rate))*100,1), "%. E
```

```
## A taxa de erro OOB do modelo final é de 28.5%. E a acurácia do modelo verificada na base de teste é
```

```
set.seed(09986755)
rf4 <- randomForest(as.factor(DIRETORIA) ~ ., data=db_modelo,
                    ntree = 420,
                    mtry = 45,
                    importance = TRUE,
                    proximity = TRUE)
rf4
```

```
##
## Call:
##  randomForest(formula = as.factor(DIRETORIA) ~ ., data = db_modelo,      ntree = 420, mtry = 45, imp
##                Type of random forest: classification
##                      Number of trees: 420
## No. of variables tried at each split: 45
##
##         OOB estimate of  error rate: 26.92%
## Confusion matrix:
##        DEA DEE OUTROS class.error
## DEA    149  39     26   0.3037383
## DEE     32 154     24   0.2666667
## OUTROS  19  28    153   0.2350000
```

```
# Predict the testing set with the trained model
#predictions4 <- predict(rf4, testing, type = "class")

# Accuracy and other metrics
#confusionMatrix(predictions4, as.factor(testing$DIRETORIA))

#p4 <- predict(rf4,training)
#as.character(head(p4))
#head(training$DIRETORIA)
```

Comparacao do poder de predicao dos modelos treinados e propostos

```
as.character(p1[1:15])
```

```
##  [1] "0.171428571428571"  "0.876543209876543"  "0.2"
```

```
##  [4] "0.876543209876543"  "0.171428571428571"  "0.9"
##  [7] "0.0434782608695652" "0.876543209876543"  "0"
## [10] "0.876543209876543"  "0.876543209876543"  "0.857142857142857"
## [13] "0.857142857142857"  "0.0344827586206897" "0.171428571428571"
```

```
as.character(rf$y[1:15])
```

```
##  [1] "3" "1" "2" "1" "3" "1" "3" "1" "2" "1" "1" "1" "1" "3" "2"
```

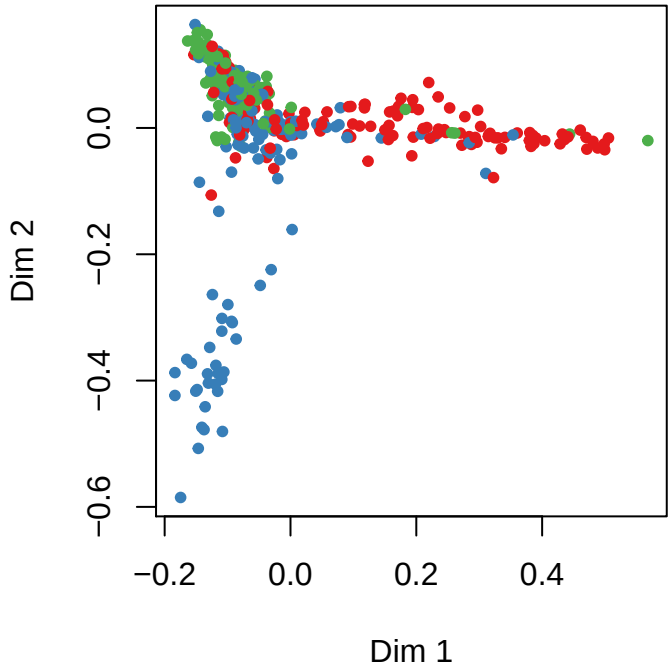
```
as.character(p2[1:15])
```

```
##  [1] "OUTROS" "DEA"    "DEE"    "DEA"    "OUTROS" "DEA"    "OUTROS"
##  [8] "DEA"    "DEE"    "DEA"    "DEA"    "DEA"    "DEA"    "OUTROS"
## [15] "DEE"
```

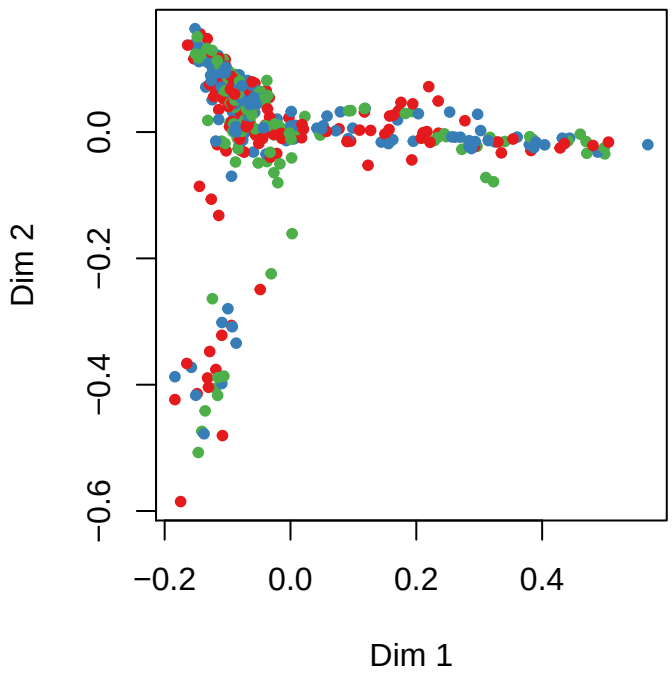
```
#as.character(p4[1:15])
training$DIRETORIA[1:15]
```

```
##  [1] OUTROS DEA    DEE    DEA    OUTROS DEA    OUTROS DEA    DEE    DEA
## [11] DEA    DEA    DEA    OUTROS DEE
## Levels: DEA DEE OUTROS
```

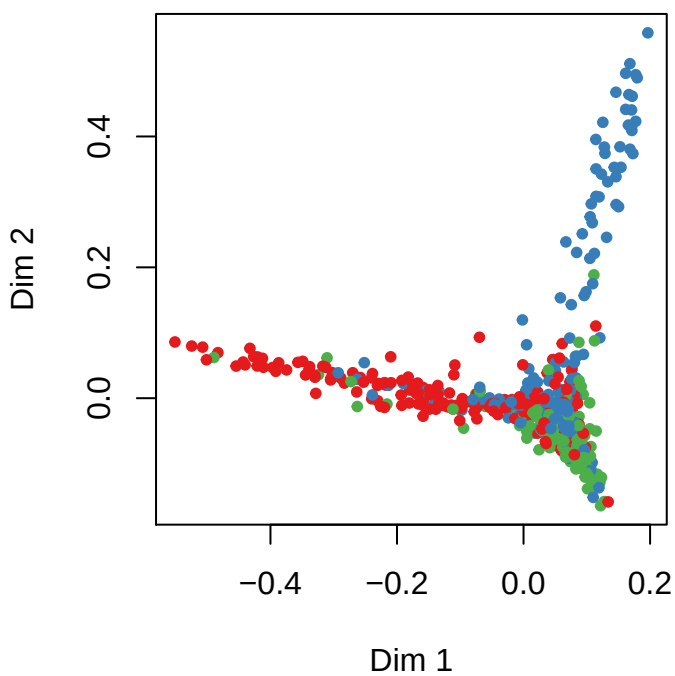
```
#edit(MDSplot)
fig.align="center"
training1$DIRETORIA = as.factor(training1$DIRETORIA)
testing1$DIRETORIA = as.factor(testing1$DIRETORIA)
db_modelo$DIRETORIA = as.factor(db_modelo$DIRETORIA)

MDSplot(rf3, training1$DIRETORIA, pch=20) # MDIM_treino
```



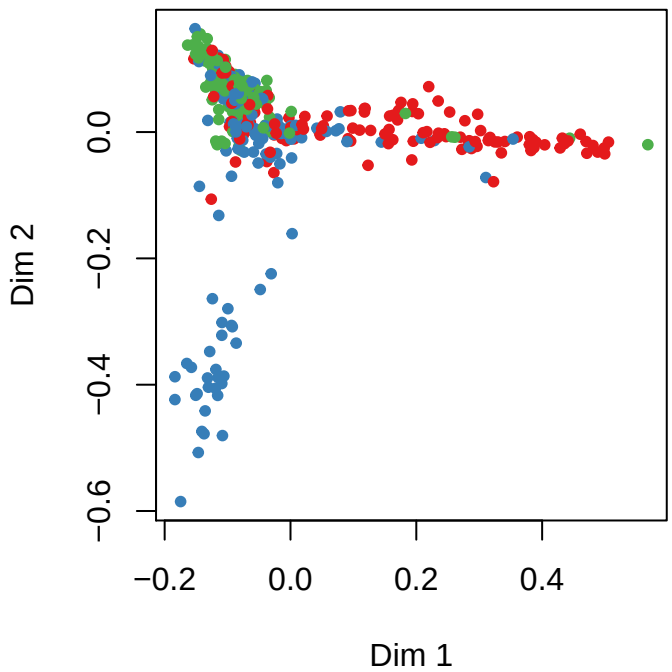

```
MDSplot(rf3, testing1$DIRETORIA, pch=20) # MDIM_teste
```




```
MDSplot(rf4, db_modelo$DIRETORIA, pch=20) # MDIM_BASECHEIA
```

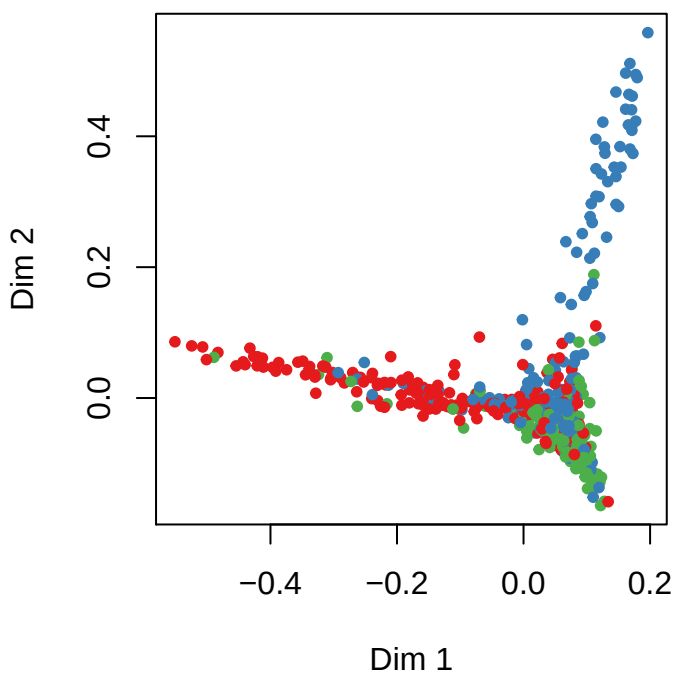



```
sum(MDSplot(rf3, training1$DIRETORIA, pch=20)$eig[1:2]); # sum(MDIM_teste$eig[1:2])  # is the same
```




```
## [1] 16.35569
```

```
sum(MDSplot(rf4, db_modelo$DIRETORIA, pch=20)$eig[1:2]); # sum(MDIM_teste$eig[1:2])  # is the same
```




```
## [1] 19.14914
```

```
DIR = db_modelo$DIRETORIA
x = db_modelo[,-1]; colnames(x) <- c(DIR)
x = x[,1:300]
hchart(princomp(x, cor = FALSE))
```