

# Análise de Dados dos Gastos Parlamentares

11 de Julho

```
library(dplyr, warn.conflicts = FALSE)
library(ggplot2)
theme_set(theme_bw())
```

## Resumo

Neste relatório analisei informações referentes aos gastos das Cotas para Exercício dos parlamentares Paraibanos durante o corrente ano, 2016. Estes dados estão disponíveis na transparência da Câmara Federal disponível em <http://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/dados-abertos-cota-parlamentar> (<http://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/dados-abertos-cota-parlamentar>). A análise, feita na linguagem de programação R, tem o objetivo de investigar como os 13 Deputados Federais Paraibanos gastam suas cotas Parlamentares.

## Introdução

Os deputados federais são eleitos com o intuito de representar os interesses da população a nível legislativo e assim contribuir para o engrandecimento do Estado. Para isso, cada deputado tem uma cota mensal para cobrir as despesas referentes ao exercício parlamentar, como: Passagens, telefonia, contratação de pessoal, divulgação da atividade parlamentar, entre outros.

O valor da cota parlamentar é diferente para cada estado da União, porque leva em consideração o custo referente a passagens aéreas para os diferentes estados da federação. Para o estado da Paraíba temos a seguinte cota:

PB		R\$ 42.032,56
----	--	---------------

Logo,

```
valorDaCota = 42032.56
```

Diante de uma das maiores crises da política brasileira, aonde deputados são acusados constantemente de corrupção, analisar como gastam o dinheiro da União é uma atividade fundamental de cidadania.

Neste relatório tenho o objetivo de responder duas questões a cerca dos gastos parlamentares:

1. Em que tipo de despesas nossos parlamentares gastam mais recursos de sua cota?
2. Quais tipos de despesas têm despesas que mais variam, que têm valores mais desiguais?

Para respondê-las este artigo foi dividido em algumas seções:

- Os Dados

- Processamento dos Dados
- Respostas as questões
- Algumas Palavras
- Referências

## Os Dados

Os dados disponíveis no site da Transparência da Câmara Federal são em formato XML. A conversão para csv (comma-separated value) foi feita pelo professor Nazareno e disponibilizado no link <https://github.com/nazareno/ciencia-de-dados-1/blob/master/dados/ano-atual.csv.tgz> (<https://github.com/nazareno/ciencia-de-dados-1/blob/master/dados/ano-atual.csv.tgz>). O arquivo conta com as descrições dos dados parlamentares distribuídos em vinte e nove (29) variáveis, incluindo quando e onde ocorreu os gastos, o valor do documento e nome do deputado.

## Processamento dos Dados

Após clonar o repositório e extrair o banco de dados o mesmo foi lido e setado a uma variável `gastosDeputadosBr`.

A data da análise é mostrada a seguir:

```
dataDaAnalise = date()  
dataDaAnalise
```

```
## [1] "Mon Jul 11 18:14:16 2016"
```

Leitura dos Arquivos:

```
gastosDeputadosBr = read.csv("~/Documentos/sampaio/AnaliseDeDados/DataAnalysis  
I/ano-atual.csv",  
                             comment.char = "#", encoding = "UTF-8")
```

A estrutura de entrada é visualizada

```
str(gastosDeputadosBr)
```

```
## 'data.frame':    149485 obs. of  29 variables:
## $ sgPartido      : Factor w/ 28 levels "DEM","PCdoB",...: 1 1 1 1
1 1 1 1 1 1 ...
## $ numMes         : int  1 3 2 3 2 3 2 4 5 2 ...
## $ txtCNPJCPF     : num  5.94e+12 5.94e+12 5.94e+12 2.34e+12 2.34e
+12 ...
## $ vlrGlosa       : num  6 1.6 0 9.85 0 0 0 0 0 0 ...
## $ txNomeParlamentar : Factor w/ 572 levels "ABEL MESQUITA JR.",...: 1
1 1 1 1 1 1 1 1 1 ...
## $ sgUF           : Factor w/ 27 levels "AC","AL","AM",...: 22 22 2
2 22 22 22 22 22 22 ...
## $ nuCarteiraParlamentar : int  1 1 1 1 1 1 1 1 1 1 ...
## $ vlrDocumento   : num  37.4 38.8 33.3 464.3 112.1 ...
## $ nuDeputadoId    : int  3074 3074 3074 3074 3074 3074 3074 3074 3
074 3074 ...
## $ numSubCota      : int  1 1 1 1 1 3 3 3 3 3 ...
## $ txtDescricaoEspecificacao: Factor w/ 4 levels "Aeronaves","Embarçaõe
s",...: NA NA NA NA NA 4 4 4 4 4 ...
## $ datEmissao      : Factor w/ 13405 levels "2015-11-04T00:00:0
0",...: 2088 4820 2088 6841 3267 6274 3398 8557 10874 3132 ...
## $ codLegislatura   : int  55 55 55 55 55 55 55 55 55 55 ...
## $ numParcela       : int  0 0 0 0 0 0 0 0 0 0 ...
## $ txtDescricao     : Factor w/ 18 levels "ASSINATURA DE PUBLICAÇÃO
S",...: 11 11 11 11 11 2 2 2 2 2 ...
## $ txtNumero        : Factor w/ 117141 levels "0","00","000",...: 323
55 32355 117131 1773 4706 1678 53795 54457 54638 13253 ...
## $ numEspecificacaoSubCota : int  0 0 0 0 0 1 1 1 1 1 ...
## $ txtFornecedor    : Factor w/ 15074 levels "03SINPETAXI",...: 4227
4227 4227 5336 5336 2661 2661 2661 2661 3309 ...
## $ numLote          : int  1268870 1282185 1268878 1289623 1282183 1
278448 1268791 1289609 1295137 1268790 ...
## $ indTipoDocumento : int  0 0 0 0 0 0 0 0 0 0 ...
## $ numAno           : int  2016 2016 2016 2016 2016 2016 2016 2016 2
016 2016 ...
## $ numRessarcimento : int  5369 5417 5369 5442 5417 5413 5369 5441 5
466 5367 ...
## $ vlrLiquido       : num  31.4 37.2 33.3 454.5 112.1 ...
## $ ideDocumento     : int  5928744 5970849 5928732 5993742 5970856 5
959453 5928619 5986310 6011374 5928626 ...
## $ ideCadastro      : int  178957 178957 178957 178957 178957 178957
178957 178957 178957 178957 ...
## $ nuLegislatura     : int  2015 2015 2015 2015 2015 2015 2015 2015 2
015 2015 ...
## $ txtPassageiro     : Factor w/ 2877 levels "A BARROS DE OLVEIRA DOS
SANTO",...: NA NA NA NA NA NA NA NA NA NA ...
## $ txtTrecho         : Factor w/ 1613 levels "AFL/CGB","AFL/CGB/AF
L",...: NA NA NA NA NA NA NA NA NA NA ...
## $ vlrRestituicao     : num  NA NA NA NA NA NA NA NA NA NA ...
```

As variáveis consideradas na análise foram:

- txNomeParlamentar
- sgPartido

- sgUF
- vlrLiquido (O seu conteúdo é valor que será realmente descontado ou acrescentado à cota)
- txtDescricao (O seu conteúdo é a descrição do Tipo de Despesa relativo à despesa em questão.))

```
#Seleciona as variáveis que serão utilizadas
gastosDeputadosPB <- select(gastosDeputadosBr, sgUF ,sgPartido , txNomeParlamentar, vlrLiquido, txtDescricao, numMes)

#Filtra apenas os deputados paraibanos
gastosDeputadosPB = gastosDeputadosPB %>% filter(sgUF == "PB")
```

Para facilitar a análise o nome das variáveis foram renomeadas para melhor analisá-las e em seguida é impresso as primeiras cinco linhas do dataframe.

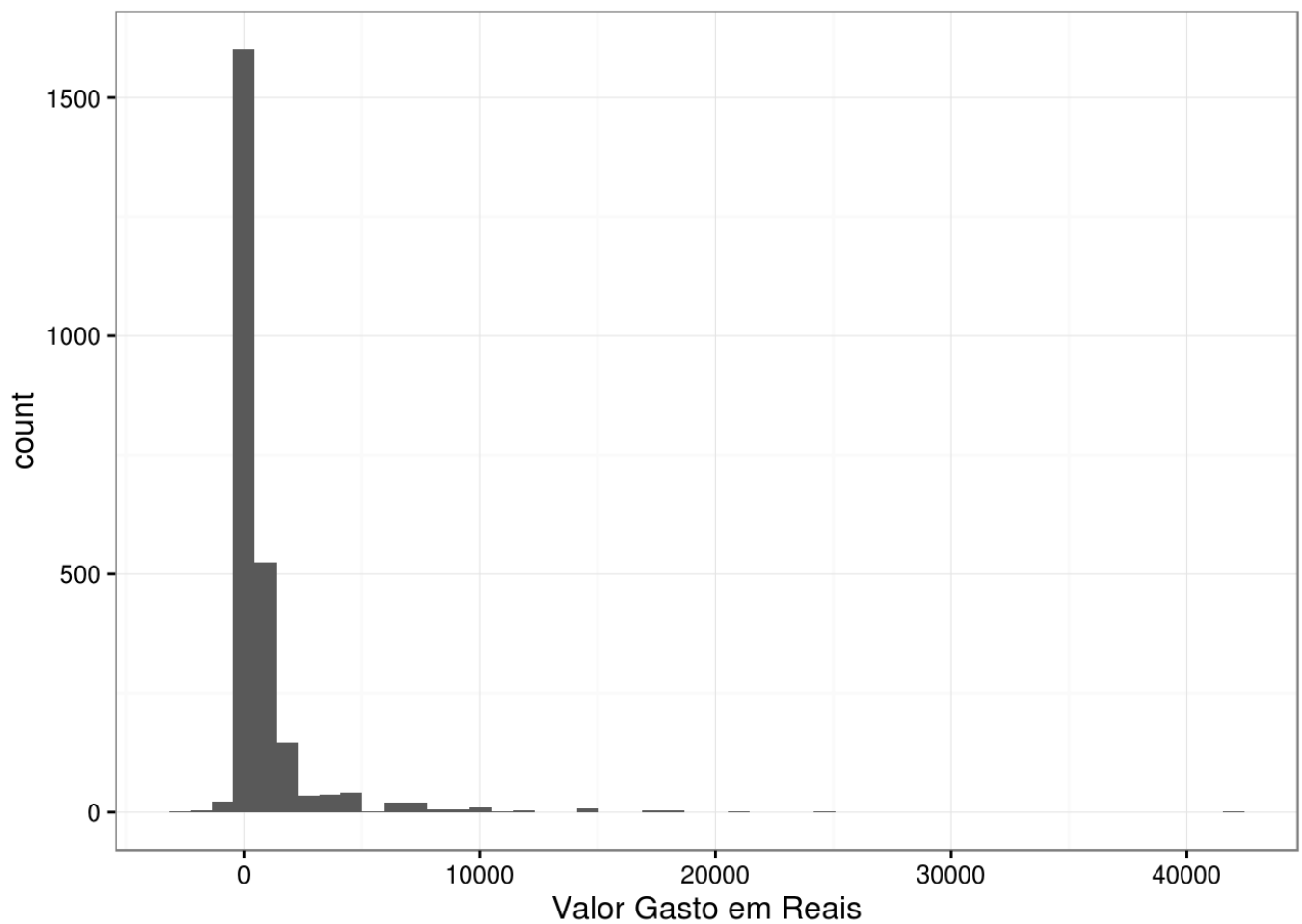
```
names(gastosDeputadosPB) <- c("UF", "Partido", "Nome", "ValorGasto", "Descricao", "Mes")
#Imprime as dez primeiras linhas
head(gastosDeputadosPB,5)
```

```
##   UF Partido      Nome ValorGasto
## 1 PB      PP AGUINALDO RIBEIRO      583
## 2 PB      PP AGUINALDO RIBEIRO      583
## 3 PB      PP AGUINALDO RIBEIRO      583
## 4 PB      PP AGUINALDO RIBEIRO      583
## 5 PB      PP AGUINALDO RIBEIRO      583
##                                     Descricao Mes
## 1 MANUTENÇÃO DE ESCRITÓRIO DE APOIO À ATIVIDADE PARLAMENTAR 1
## 2 MANUTENÇÃO DE ESCRITÓRIO DE APOIO À ATIVIDADE PARLAMENTAR 2
## 3 MANUTENÇÃO DE ESCRITÓRIO DE APOIO À ATIVIDADE PARLAMENTAR 3
## 4 MANUTENÇÃO DE ESCRITÓRIO DE APOIO À ATIVIDADE PARLAMENTAR 4
## 5 MANUTENÇÃO DE ESCRITÓRIO DE APOIO À ATIVIDADE PARLAMENTAR 5
```

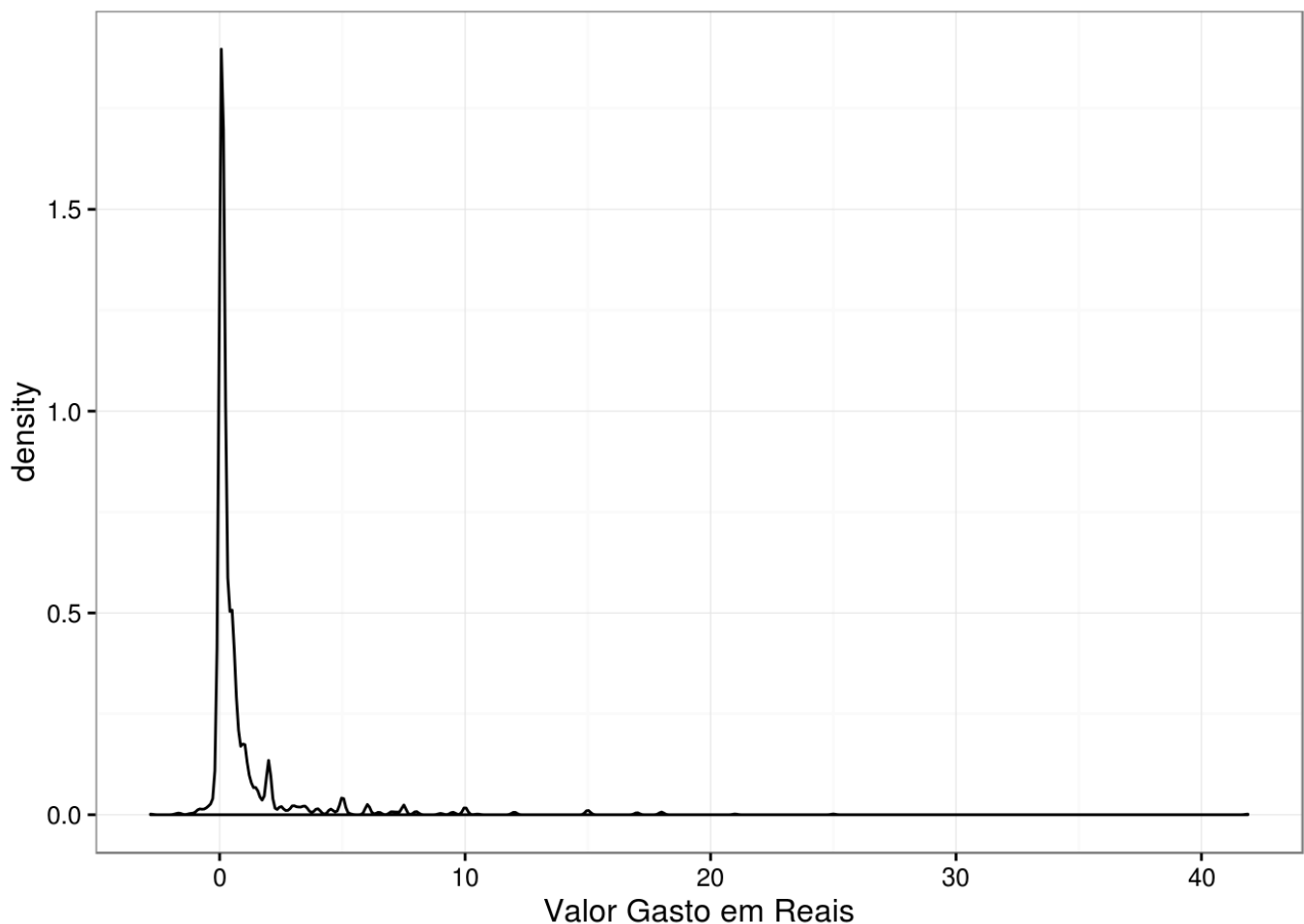
## Como varia os gastos dos deputados?

Para isto é gerado um gráfico para melhor visualizar a variação dos gastos dos parlamentares.

```
ggplot(gastosDeputadosPB, mapping = aes(x = ValorGasto)) +
  geom_histogram(bins = 50) + xlab("Valor Gasto em Reais")
```



```
ggplot(gastosDeputadosPB) +  
  geom_density(kernel = "gaussian", mapping = aes(x = ValorGasto/1e3)) + xlab("Valor Gasto em Reais")
```



Os valores estão muito concentrados a esquerda do gráfico, assimétricos, além disto os valores crescem exponencialmente. Para facilitar a visualização é plotado um gráfico monolog. É importante frisar que ao plotar um gráfico monolog os valores negativos, referentes a compensação de débitos não serão mostrados.

```
summary(gastosDeputadosPB$ValorGasto)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2820.0   70.0   190.0   847.9   639.0 41890.0
```

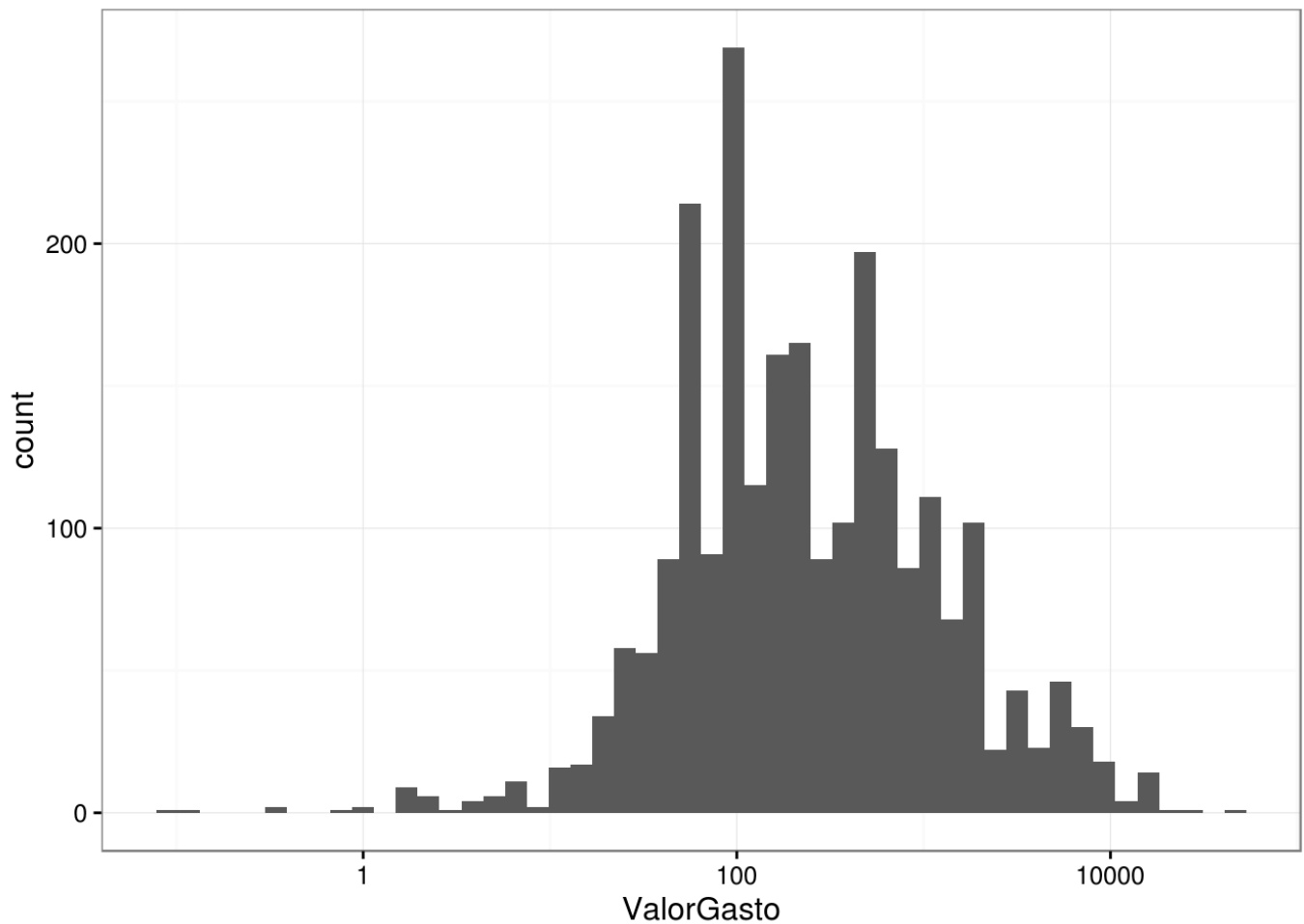
```
DesvioPadrao <- summarise(gastosDeputadosPB, sd(ValorGasto))
DesvioPadrao
```

```
##      sd(ValorGasto)
## 1          2184.164
```

```
ggplot(gastosDeputadosPB, mapping = aes(x = ValorGasto)) +
  geom_histogram(bins = 50) +
  scale_x_log10()
```

```
## Warning in self$trans$transform(x): NaNs produzidos
```

```
## Warning: Removed 80 rows containing non-finite values (stat_bin).
```



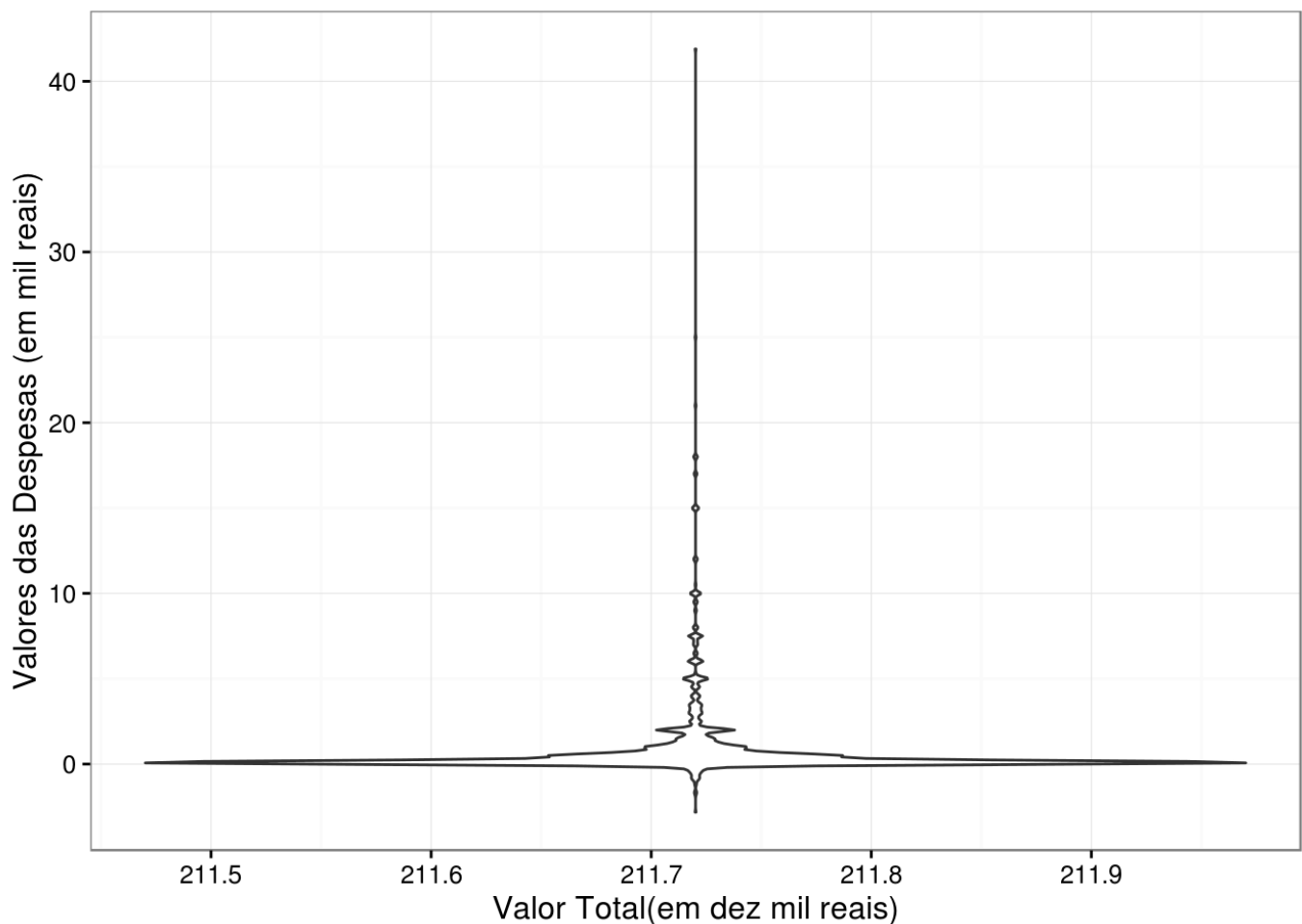
Os valores são mais concentrados próximos à mediana (R\$190,00) e 75% dos gastos são inferiores há R\$650,00. Contudo, o desvio padrão apresenta um valor de R\$2184,16 evidenciando que os valores são muito dispersos.

```
ValorTotal <- gastosDeputadosPB %>% summarise(sum(ValorGasto))
```

```
ggplot(gastosDeputadosPB, mapping = aes(x = ValorTotal/1e4, y = ValorGasto/
1e3, width = .5)) +
  geom_violin() + xlab("Valor Total(em dez mil reais)") + ylab("Valores das D
espesas (em mil reais)")
```

```
## Don't know how to automatically pick scale for object of type data.frame. De
faulting to continuous.
```

```
## Warning in density.default(x, weights = w, bw = bw, adjust = adjust, kernel
## = kernel, : sum(weights) != 1 -- will not get true density
```



O valor negativo significa que a despesa é um bilhete de compensação de viagem aérea, significando que o referido valor compensa um outro bilhete emitido e não utilizado pelo deputado. Já o outro extremo, onde a despesa é de mais de 41 mil reais, refere-se ao gasto do *Deputado Benjamin Maranhão* com divulgação parlamentar no mês de Maio.

O deputado Benjamin Maranhão gastou quase toda sua cota do mês de maio com divulgação parlamentar, tal fato levanta a seguinte questão: Será que os deputados estão gastando mais do que deviam?

Para isto precisamos recorrer à legislação da Câmara Federal, precisamente no **Artigo 13 do ATO DA MESA de número 23**, promulgada no dia 23 de Maio de 2009, que diz:

Art. 13. O saldo da Cota não utilizado acumula-se ao longo do exercício financeiro, vedada a acumulação de saldo de um exercício para o seguinte.

Parágrafo 1º - A Cota somente poderá ser utilizada para despesas de competência do respectivo exercício financeiro.

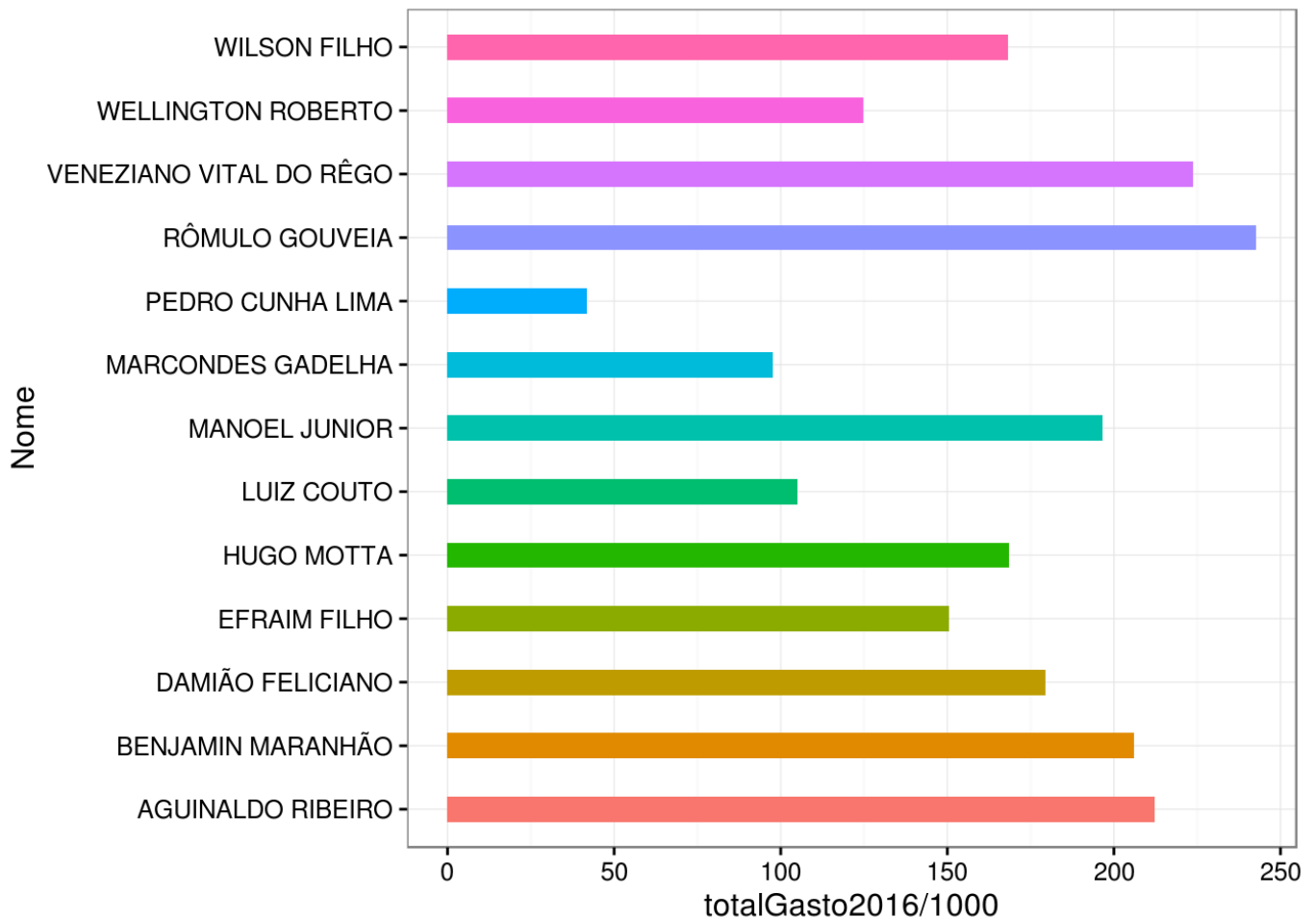
Parágrafo 2º - A importância que exceder, no exercício financeiro, o saldo de Cota disponível será deduzida automática e integralmente da remuneração do parlamentar ou do saldo de acerto de contas de que ele seja credor, revertendo-se à conta orçamentária própria da Câmara dos Deputados.

Diante disto, para responder a dúvida acima, foi calculado o valor total gasto - até o mês de junho - e verificado se algum deputado excedeu a cota total (Cota \* Quantidade de meses passado)

```
totalGastoDeputado = gastosDeputadosPB %>% group_by(Nome) %>% summarise(tot
alGasto2016 = sum(ValorGasto))
```



```
ggplot(totalGastoDeputado, mapping = aes(x = Nome,
                                          y = totalGasto2016/1e3,
                                          fill = Nome, width = .4)) +
geom_bar(stat = "identity") +
theme(legend.position="none")+
coord_flip()
```



Assim, nenhum deputado excedeu o total da Cota Parlamentar para os seis primeiros meses do ano.

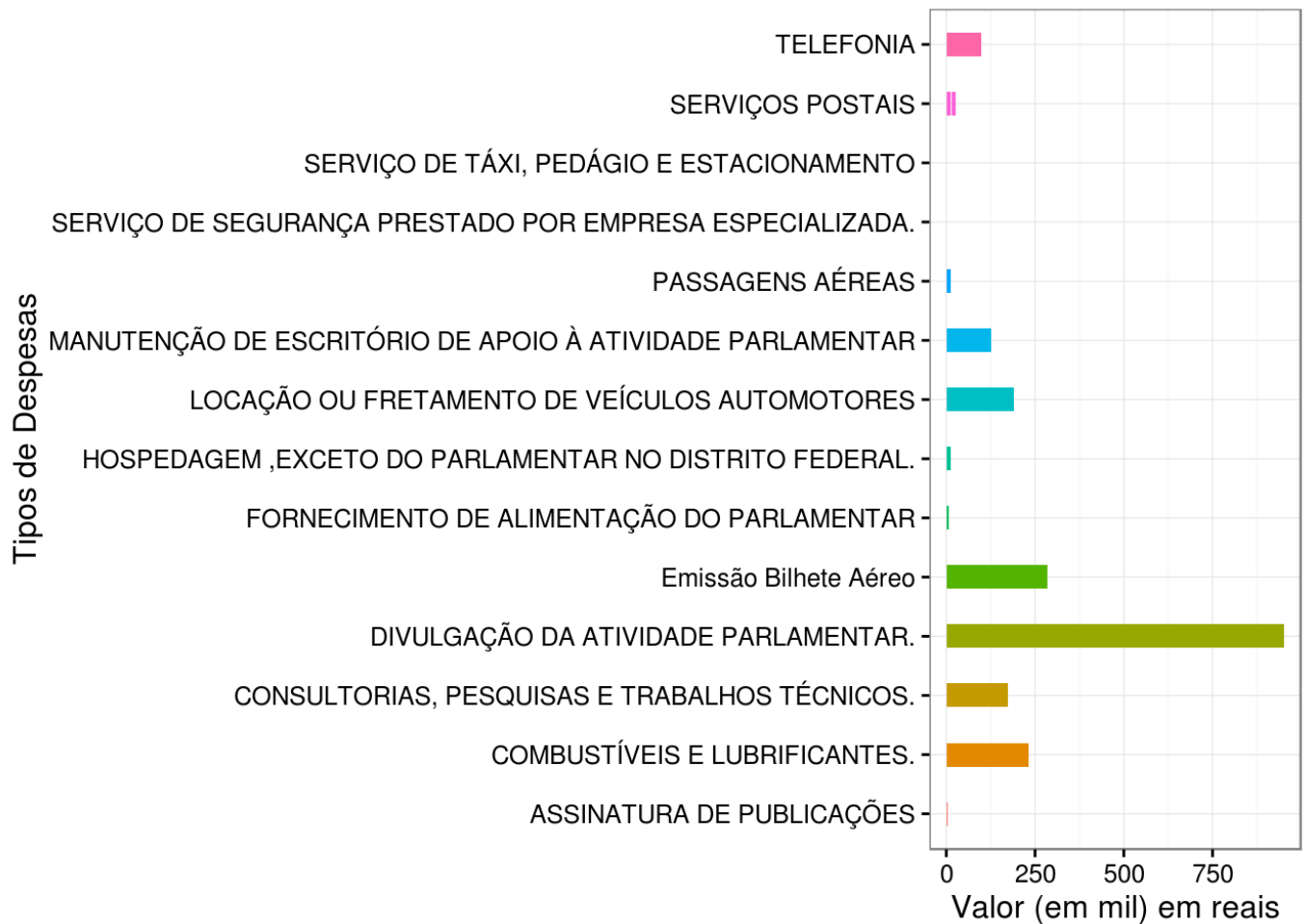
## Respostas da Questões

### Em que tipo de despesas nossos parlamentares gastam mais recursos de sua cota?

Para responder tal questão, foi preciso plotar o gráfico onde a Descrição dos Gastos e seus respectivos valores estejam contidos.

```
ggplot(gastosDeputadosPB, mapping = aes(x = Descricao,
                                          y = ValorGasto/1e3,
                                          fill = Descricao, width = .4)) +
geom_bar(stat = "identity") +
theme(legend.position="none")+
xlab("Tipos de Despesas") + ylab("Valor (em mil) em reais") +
coord_flip()
```

```
## Warning: Stacking not well defined when ymin != 0
```



No gráfico acima é possível perceber que os maiores gastos da Cota Parlamentar são para divulgação parlamentar e passagens aéreas. Assim, analisei cada uma dessas despesas.

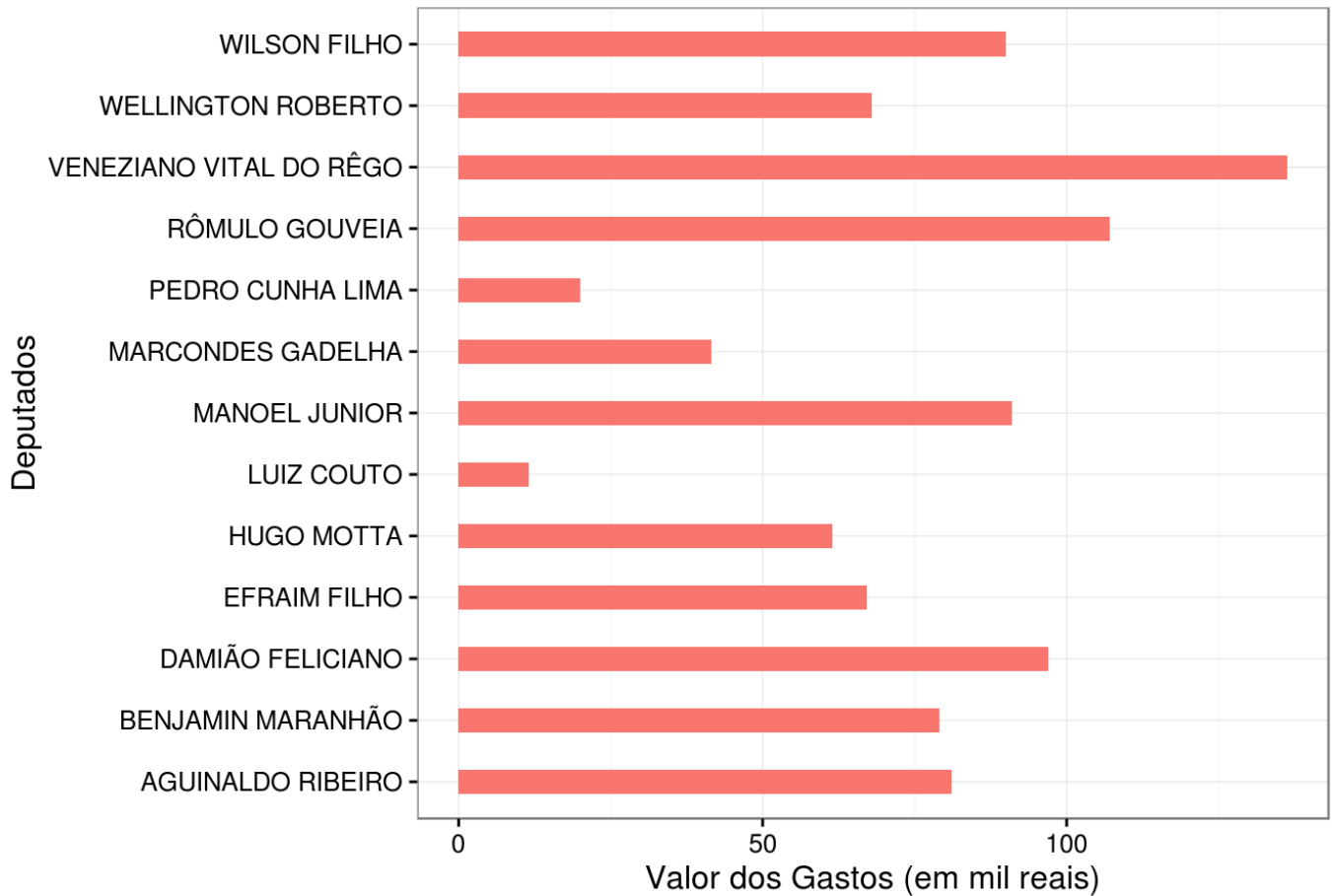
Separei os dois grupos que serão analisados: 1) os gastos com divulgação parlamentar; e 2) os gastos com despesas aéreas.

```
despesaDivulgacao = gastosDeputadosPB %>% filter(Descricao == "DIVULGAÇÃO
DA ATIVIDADE PARLAMENTAR.")
despesaAerea = gastosDeputadosPB %>% filter(Descricao == "Emissão Bilhete
Aéreo")
```

Imprimindo o gráfico referente as Despesas com Divulgação Parlamentar, temos:

```
ggplot(despesaDivulgacao, mapping = aes(x = Nome, y = ValorGasto/1e3, fill
= Descricao, width = 0.4, alpha = 0.4 )) +
  theme(legend.position = "none") +
  xlab("Deputados") + ylab("Valor dos Gastos (em mil reais)") +
  geom_bar(stat = "identity") + ggtitle("Despesas com Divulgação Parlama
r por Deputado") +
  coord_flip()
```

## Despesas com Divulgação Parlamentar por Deputado



O Deputado que mais gastou da sua cota com divulgação parlamentar foi o *Deputado Veneziano Vital do Rêgo (PMDB)* com a quantia total de R\$ 136.300,00.

```
despesaDivulgacao %>% filter(Nome == "VENEZIANO VITAL DO RÊGO") %>% summarise(
  despesaComDivulgacao = sum(ValorGasto))
```

```
## despesaComDivulgacao
## 1 136300
```

Em contra partida, o *Deputado Luiz Couto (PT)* foi o que menos dedicou dos seus recursos para divulgação parlamentar. Todo o seu gasto foi de R\$ 11.500,00

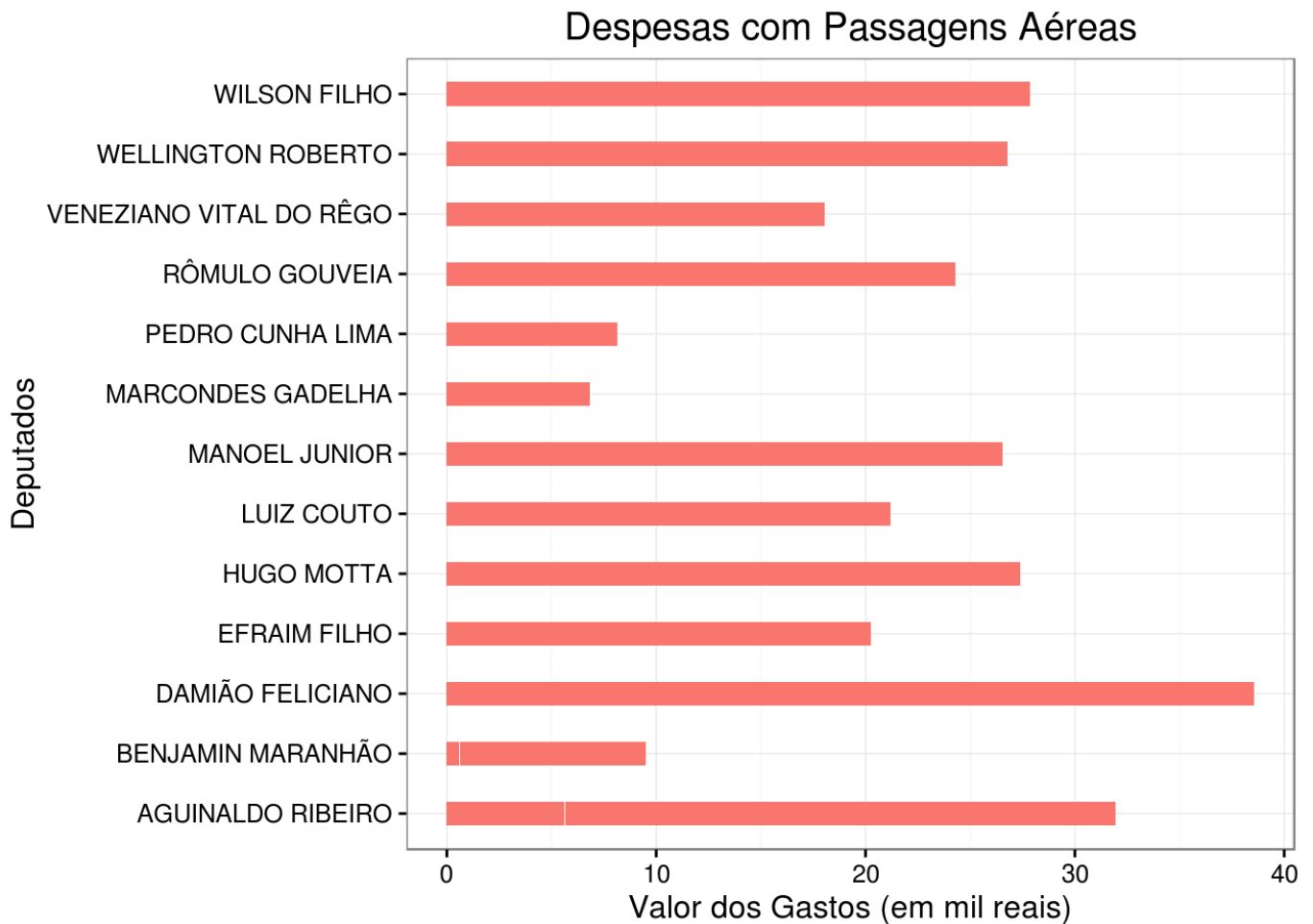
```
despesaDivulgacao %>% filter(Nome == "LUIZ COUTO") %>% summarise(despesaComDi
  lvgacao = sum(ValorGasto))
```

```
## despesaComDivulgacao
## 1 11500
```

Para as despesas áreas, temos:

```
ggplot(despesaAerea, mapping = aes(x = Nome, y = ValorGasto/1e3, fill = Descricao, width = 0.4, alpha = 0.4 )) +
  theme(legend.position = "none") +
  xlab("Deputados") + ylab("Valor dos Gastos (em mil reais)") +
  geom_bar(stat = "identity") + ggtitle("Despesas com Passagens Aéreas") +
  coord_flip()
```

```
## Warning: Stacking not well defined when ymin != 0
```



O deputado que mais gastou com despesas aéreas foi o *Deputado Damião Feliciano (PDT)* com um valor total de R\$ 38.558,75. Já o que poupou mais o bolso do brasileiro foi o *Deputado Marcondes Gadelha (PSC)* com um gasto total de R\$ 6.824,09.

```
despesaAerea %>% filter(Nome == "DAMIÃO FELICIANO") %>% summarise(despesaComPassagens = sum(ValorGasto))
```

```
## despesaComPassagens
## 1 38558.75
```

```
despesaAerea %>% filter(Nome == "MARCONDES GADELHA") %>% summarise(despesaComPassagens = sum(ValorGasto))
```

```
##   despesaComPassagens
## 1                6824.09
```

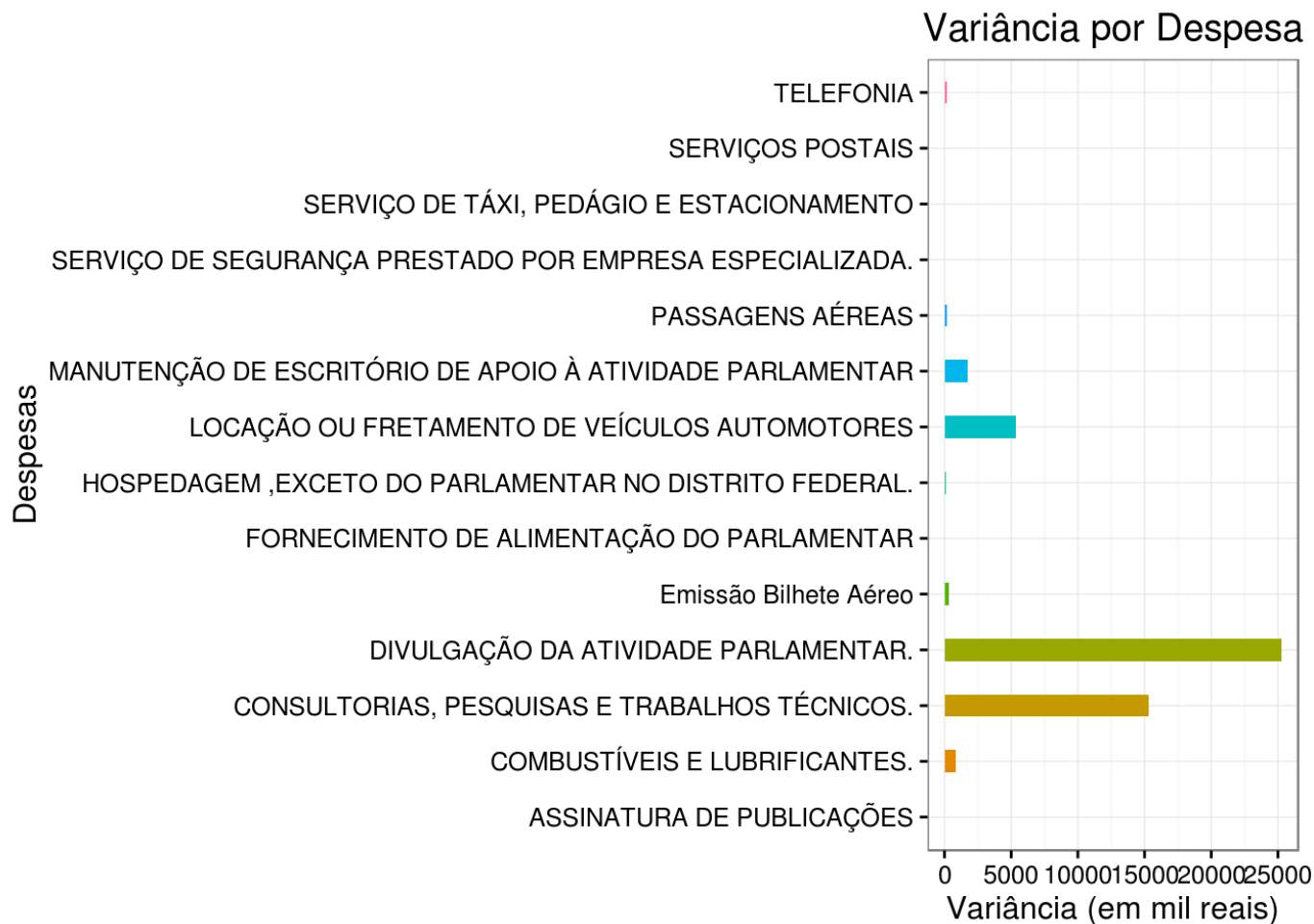
## Quais tipos de despesas têm despesas que mais variam, que têm valores mais desiguais?

O variância é uma medida utilizada por nós, analistas, para determinar a dispersão dos valores, ou seja, quão longe os valores estão do valor esperado. Assim, foi calculado a variância de cada despesa e plotado um gráfico para ajudar na determinação das variáveis que mais variam, ou seja, que tem o maior valor de variância.

```
varianciaTotal = gastosDeputadosPB %>% group_by(Descricao) %>% summarise(vari
ancia = var(ValorGasto))
varianciaTotal
```

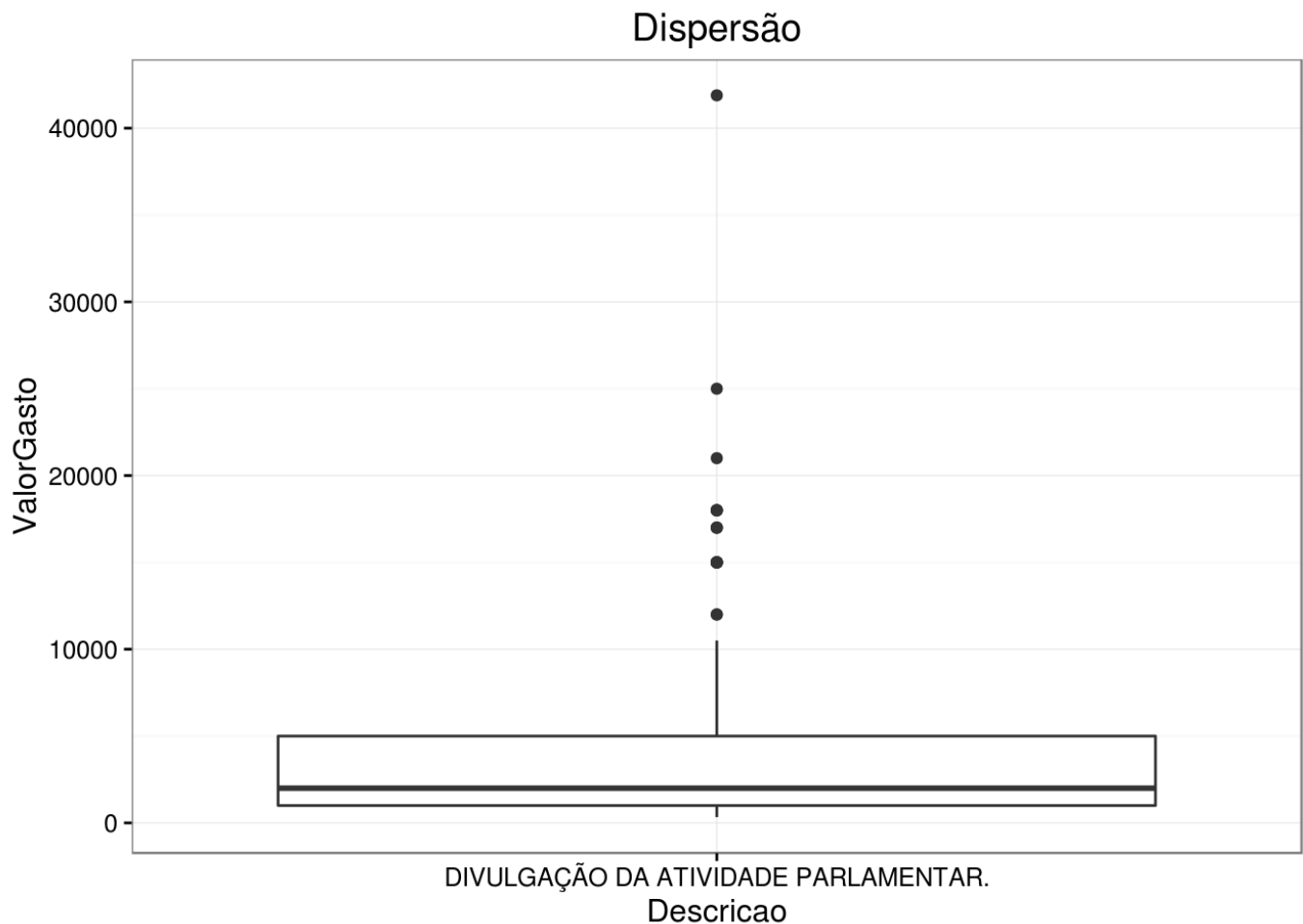
```
## # A tibble: 14 x 2
##                               Descricao      variancia
##                               <fctr>         <dbl>
## 1                ASSINATURA DE PUBLICAÇÕES 4.702895e+04
## 2          COMBUSTÍVEIS E LUBRIFICANTES. 8.776184e+05
## 3    CONSULTORIAS, PESQUISAS E TRABALHOS TÉCNICOS. 1.534398e+07
## 4    DIVULGAÇÃO DA ATIVIDADE PARLAMENTAR. 2.526388e+07
## 5              Emissão Bilhete Aéreo 3.202918e+05
## 6    FORNECIMENTO DE ALIMENTAÇÃO DO PARLAMENTAR 3.157532e+03
## 7    HOSPEDAGEM ,EXCETO DO PARLAMENTAR NO DISTRITO FEDERAL. 1.317135e+05
## 8    LOCAÇÃO OU FRETAMENTO DE VEÍCULOS AUTOMOTORES 5.365456e+06
## 9    MANUTENÇÃO DE ESCRITÓRIO DE APOIO À ATIVIDADE PARLAMENTAR 1.719872e+06
## 10          PASSAGENS AÉREAS 1.803634e+05
## 11  SERVIÇO DE SEGURANÇA PRESTADO POR EMPRESA ESPECIALIZADA. 2.856100e+00
## 12    SERVIÇO DE TÁXI, PEDÁGIO E ESTACIONAMENTO 1.227474e+04
## 13          SERVIÇOS POSTAIS 2.638242e+04
## 14          TELEFONIA 1.592099e+05
```

```
ggplot(varianciaTotal, mapping = aes(x = Descricao, y = variancia/1e3, fill = D
escricao, width = 0.4, alpha = 0.4 )) +
  theme(legend.position = "none") +
  xlab("Despesas") + ylab("Variância (em mil reais)") +
  geom_bar(stat = "identity") + ggtitle("Variância por Despesa") +
  coord_flip()
```



Pelo gráfico plotado acima é possível visualizar que as despesas que apresentam a maior variação são de Divulgação Parlamentar e Consultorias, Pesquisas e Trabalhos Técnicos.

```
divulParlamentar = gastosDeputadosPB %>% group_by(Descricao) %>% filter(Descricao == "DIVULGAÇÃO DA ATIVIDADE PARLAMENTAR.")
ggplot(divulParlamentar, mapping = aes(x = Descricao,
                                         y = ValorGasto)) + ggtitle("Dispersão") +
geom_boxplot()
```

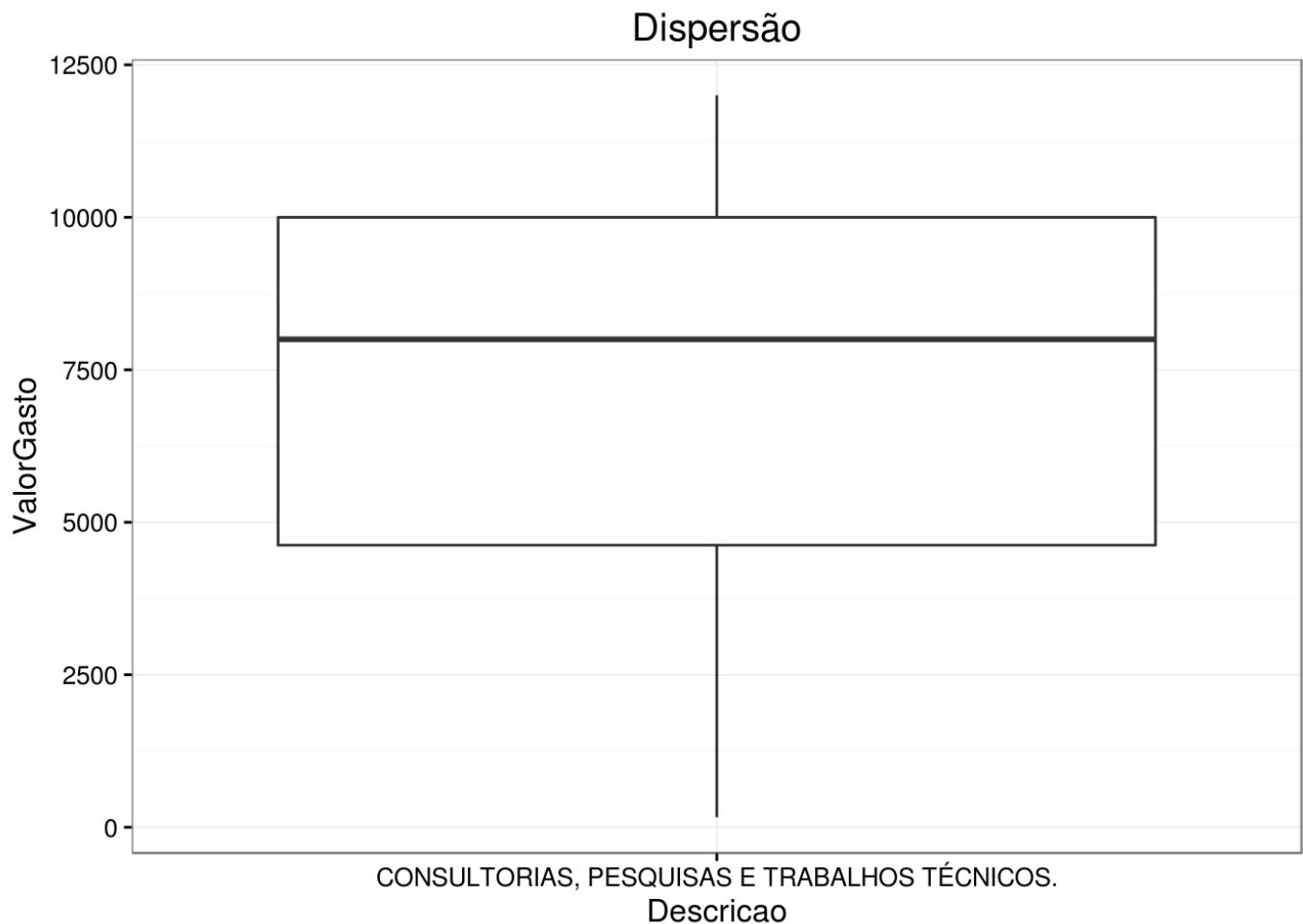


```
summary(divulParlamentar$ValorGasto)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 334.3  1000.0   2000.0   3980.0  5000.0 41890.0
```

Os valores variam vertiginosamente para a divulgação parlamentar de valores de R\$334,30 até R\$41.890,00. Além disto, 75% dos valores estão abaixo de cinco mil reais. Com o mesmo pensamento para Consultorias Técnicas, temos:

```
consulTecnica = gastosDeputadosPB %>% group_by(Descricao) %>% filter(Descricao
== "CONSULTORIAS, PESQUISAS E TRABALHOS TÉCNICOS.")
ggplot(consulTecnica, mapping = aes(x = Descricao,
                                   y = ValorGasto)) + ggtitle("Dispersão") +
geom_boxplot()
```



```
summary(consulTecnica$ValorGasto)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	166.2	4625.0	8000.0	6641.0	10000.0	12000.0

Para os gastos com Consultoria Técnica temos que os valores variam de R\$166,00 até R\$12.000,00.

## Algumas Palavras

Durante a elaboração do relatório enfrentei inúmeras dificuldades no que tange tanto à linguagem de programação quanto relacionar valores de medidas a situações que acontecem no nosso mundo.

Além disto, foi possível perceber o quanto os gastos com as despesas variam para os Deputados de uma mesma unidade federativa, o que levanta a questão: Será mesmo necessário uma cota de mais de 42 mil reais para os nossos representantes custearem seus trabalhos?

## Referências

ATO DA MESA, Numero 43 (<http://www2.camara.leg.br/login/int/atomes/2009/atodamesa-43-21-maio-2009-588364-publicacaooriginal-112820-cd-mesa.html>)

R for Data Science (<http://r4ds.had.co.nz/introduction.html#prerequisites>)