

# Portal da Transparência: Análise de Dados de Viagens

Gladson Rabelo | gladsonrf1@gmail.com

2025-02-08

## Carregando os pacotes

```
pacman::p_load(caret, e1071, dplyr, ggplot2, caTools)
```

## Carregando os dados

```
# Importando os dados do arquivo CSV  
# Arquivo disponível no site do governo, mas pode ser encontrado nesse link também  
# https://www.mediafire.com/file/z40nyaqm3hqfmoy/2024_viagem.csv/file  
viagens <- read.csv("2024_viagem.csv", sep = ';', dec = ',')
```

## Visualizando distribuição os dados

```
# Exibir as primeiras linhas do dataset  
head(viagens)
```

## Resumo estatístico do dataset

```
summary(viagens)
```

## Resumo estatístico da coluna 'Valor.passagens'

```
summary(viagens$Valor.passagens)  
dim(viagens)
```

## Estrutura dos dados

```
library(dplyr)  
# Visualizar estrutura do dataset  
glimpse(viagens)
```

## Transformação dos dados

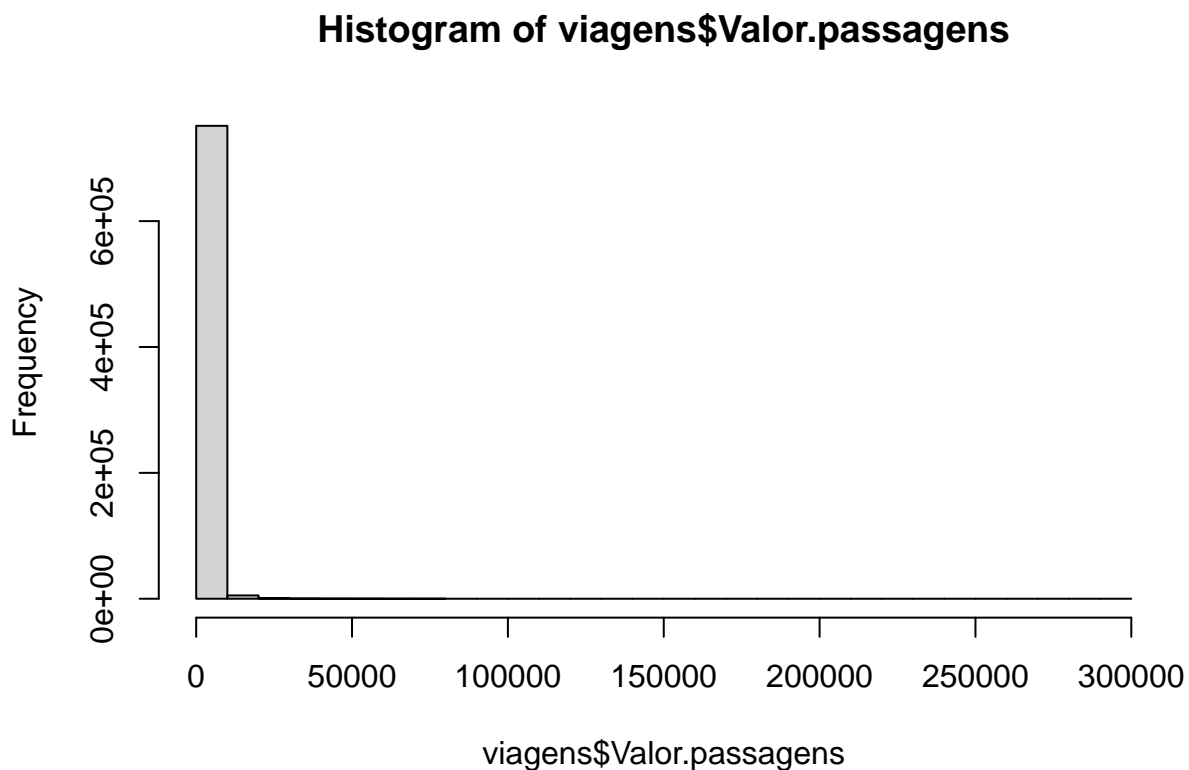
```
# Convertendo a coluna de datas para o formato Date
viagens$data.inicio <- as.Date(viagens$Período...Data.de.início, "%d/%m/%Y")

# Verificando a estrutura novamente após a conversão
glimpse(viagens)

# Criando uma nova coluna apenas com Ano/Mês
viagens$data.inicio.formatada <- format(viagens$data.inicio, "%Y-%m")
```

## Exploração dos dados

```
# Criando um histograma da coluna 'Valor.passagens'
hist(viagens$Valor.passagens)
```

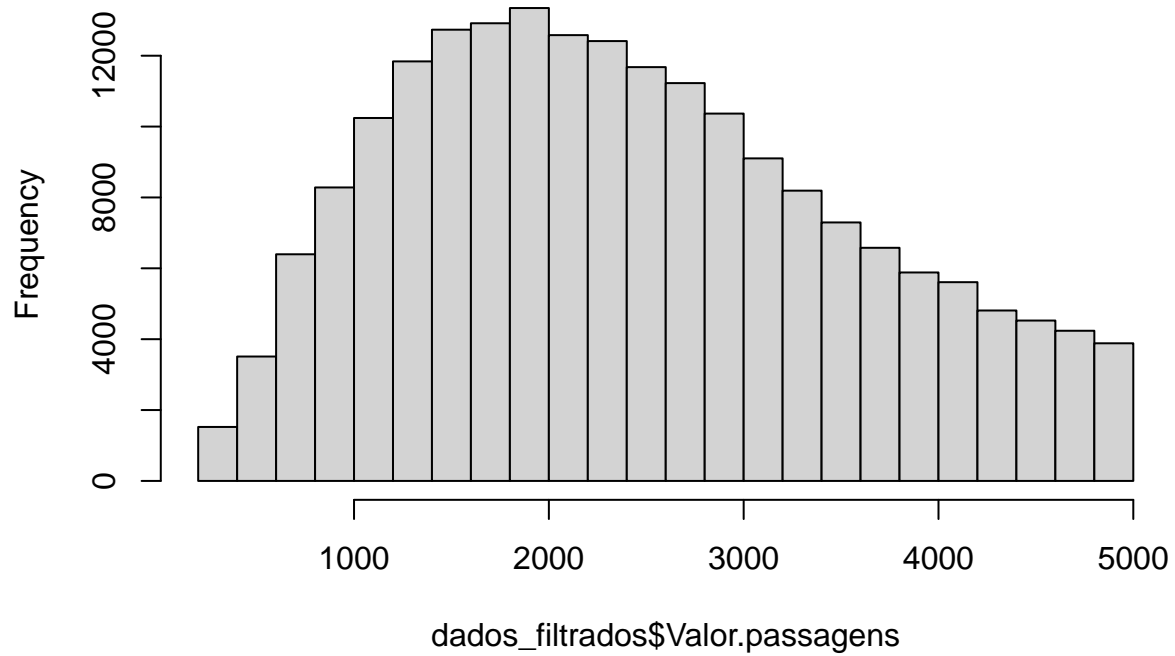


## Filtragem de dados

```
# Filtrando valores entre 200 e 5000
dados_filtrados <- viagens %>%
  select(Valor.passagens) %>%
  filter(Valor.passagens >= 200 & Valor.passagens <= 5000)

# Visualizando os dados filtrados
hist(dados_filtrados$Valor.passagens)
```

## Histogram of dados\_filtrados\$Valor.passagens

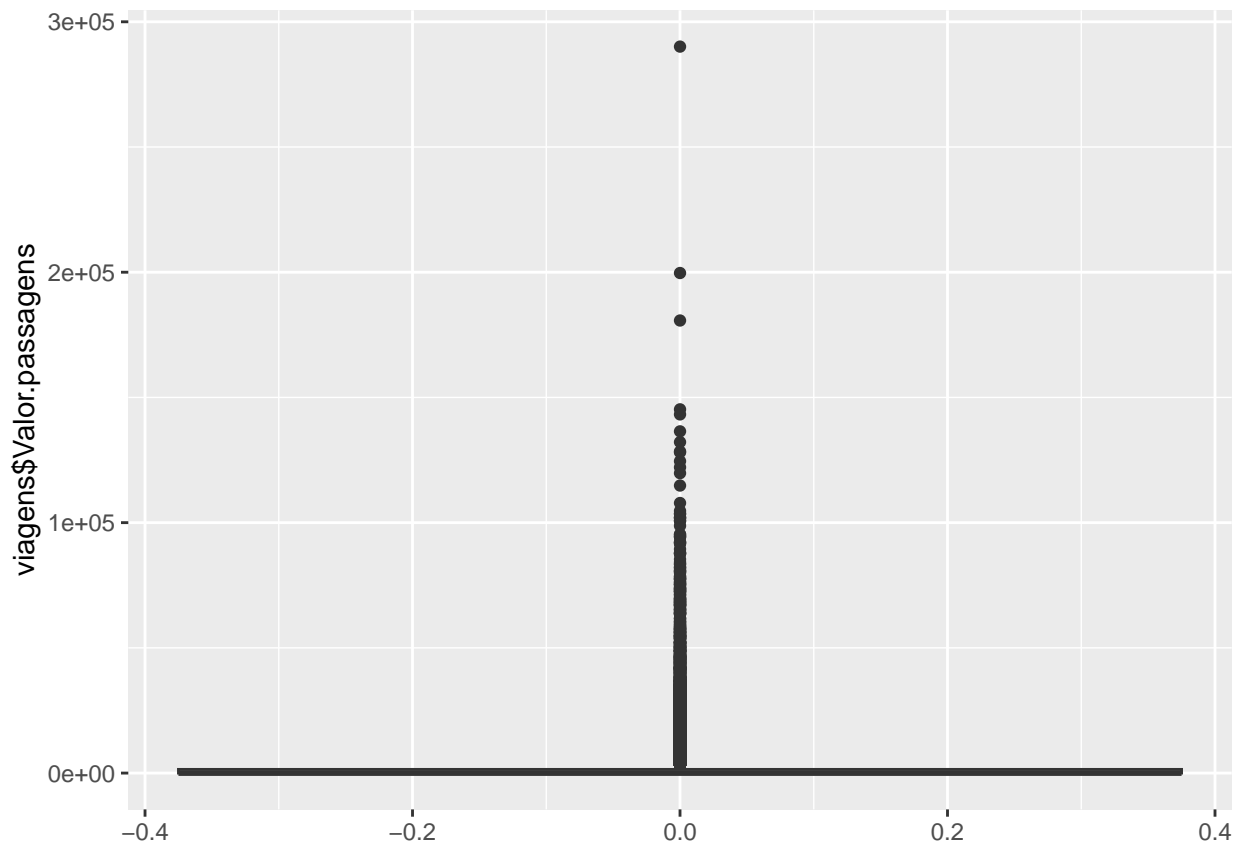


### Estatísticas descritivas

```
# Resumo estatístico da coluna 'Valor.passagens'
summary(viagens$Valor.passagens)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##       0       0       0    1045    1468 290096 

# Criando boxplot para visualizar distribuição
# Basic box plot
ggplot(viagens, aes(y=viagens$Valor.passagens)) +
  geom_boxplot()
```



### Analisando categorias

```
# Convertendo a coluna 'Situação' para um fator
viagens$Situação <- factor(viagens$Situação)

# Estrutura da coluna 'Situação'
str(viagens$Situação)

# Contagem de registros por categoria
table(viagens$Situação)

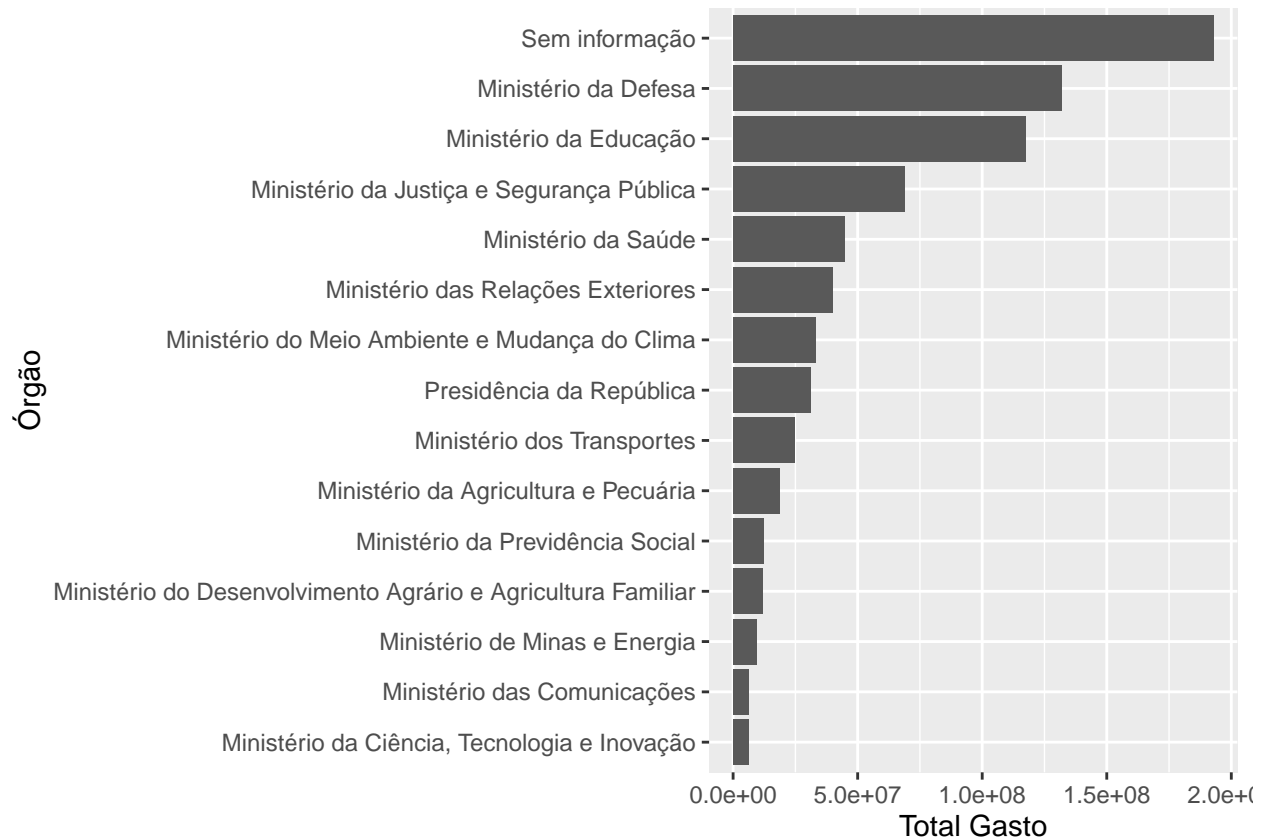
# Percentual de cada categoria
prop.table(table(viagens$Situação)) * 100
```

### Visualização dos resultados

#### Gastos por órgão

```
p1 <- viagens %>%
  group_by(Nome.do.órgão.superior) %>%
  summarise(total_gasto = sum(Valor.passagens)) %>%
  arrange(desc(total_gasto)) %>%
  top_n(15)
```

```
# Gráfico de barras dos 15 órgãos que mais gastam
ggplot2::ggplot(p1, aes(x = reorder(Nome.do.órgão.superior, total_gasto), y = total_gasto)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(x = "Órgão", y = "Total Gasto")
```



## Gastos por cidade

```
p2 <- viagens %>%
  group_by(Destinos) %>%
  summarise(total_gasto = sum(Valor.passagens)) %>%
  arrange(desc(total_gasto)) %>%
  top_n(15)

# Gráfico de barras das 15 cidades com maior gasto
ggplot2::ggplot(p2, aes(x = reorder(Destinos, total_gasto), y = total_gasto)) +
  geom_bar(stat = "identity", fill = "#0ba791") +
  geom_text(aes(label = total_gasto), vjust = 0.3, size = 3) +
  coord_flip() +
  labs(x = "Destino", y = "Total Gasto")
```

