

Desafio Técnico - A3Data

Caio Gabriel Barreto Balieiro

14/04/2022

Desafio Técnico - A3Data

O teste técnico para a vaga de Cientista de Dados da A3Data consiste na exploração da base dados “Ocorrências Aeronáuticas na Aviação Civil Brasileira” dos dados abertos do governo (<https://dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira>).

```
# Pacotes utilizados
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(dplyr)
```

```
# Conjunto de dados carregados para o teste
```

```
ocorrencia <- read.csv('ocorrencia.csv', sep = ';')
aeronave <- read.csv('aeronave.csv', sep = ';')
ocorrencia_tipo <- read.csv('ocorrencia_tipo.csv', sep = ';')
fator_contribuinte <- read.csv('fator_contribuinte.csv', sep = ';')
recomendacao <- read.csv('recomendacao.csv', sep = ';')
```

```
# Juntando em uma única tabela (Tabelão de dados!)
```

```
ocor_aren <- left_join(ocorrencia, aeronave,
                      by = c('codigo_ocorrencia2' = 'codigo_ocorrencia2'))

ocor_tipo <- left_join(ocor_aren, ocorrencia_tipo,
                      by = c('codigo_ocorrencia1' = 'codigo_ocorrencia1'))
```

```

ocor_contribuinte <- left_join(ocor_tipo, fator_contribuinte,
                              by = c('codigo_ocorrencia3' = 'codigo_ocorrencia3') )

dados <- left_join(ocor_contribuinte, recomendacao,
                  by = c('codigo_ocorrencia4' = 'codigo_ocorrencia4') )

# Carregando as primeiras 5 linhas e 5 colunas do conjunto de dados criado através
# dos 5 conjunto de dados (csv) disponíveis para o problema

dados[1:5, 1:5]

```

```

##      codigo_ocorrencia codigo_ocorrencia1 codigo_ocorrencia2 codigo_ocorrencia3
## 1          52242          52242          52242          52242
## 2          45331          45331          45331          45331
## 3          45331          45331          45331          45331
## 4          45331          45331          45331          45331
## 5          45331          45331          45331          45331
##      codigo_ocorrencia4
## 1          52242
## 2          45331
## 3          45331
## 4          45331
## 5          45331

```

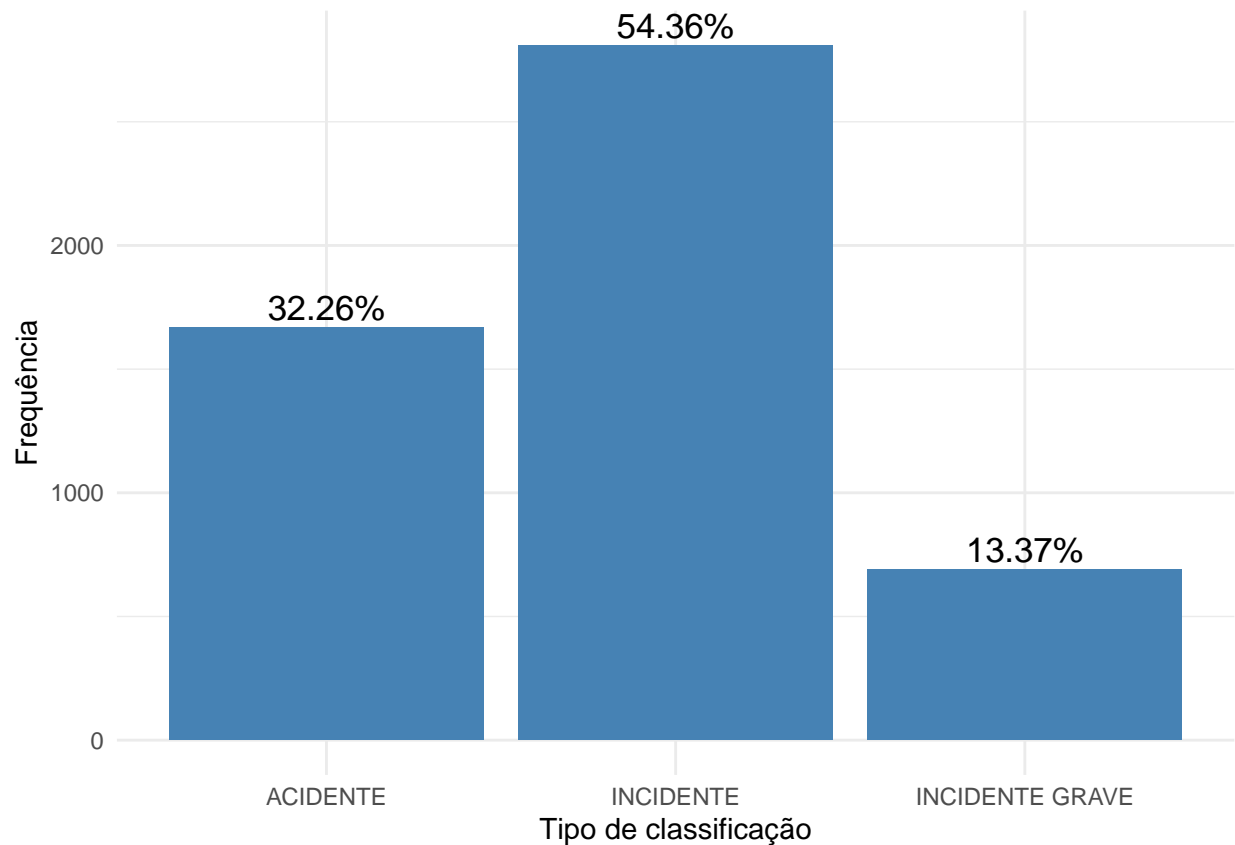
Para o conjunto de dados acima, foi considerado uma junção utilizando a função `left_join` do pacote `dplyr` em R. Esta função permite que dado dois `data.frame` (Objeto em R utilizado para armazenar tabelas) checa-se em quais linhas são comuns nos dois `data.frame`, caso a segunda tabela não possua linhas com mesma chave primária que a primeira tabela, preenche-se por NA. Essa formulação de construção de tabela, foi empregada em todos os 5 csv que estão disponíveis no site apresentado no início deste relatório. Dessa forma, pode-se extrair informações completas para as análises seguintes. Por fim, é apresentado as cinco primeiras linhas e cinco primeiras colunas do conjunto de dados que será considerado nas próximas análises.

Análise Descritiva para o conjunto de dados

```

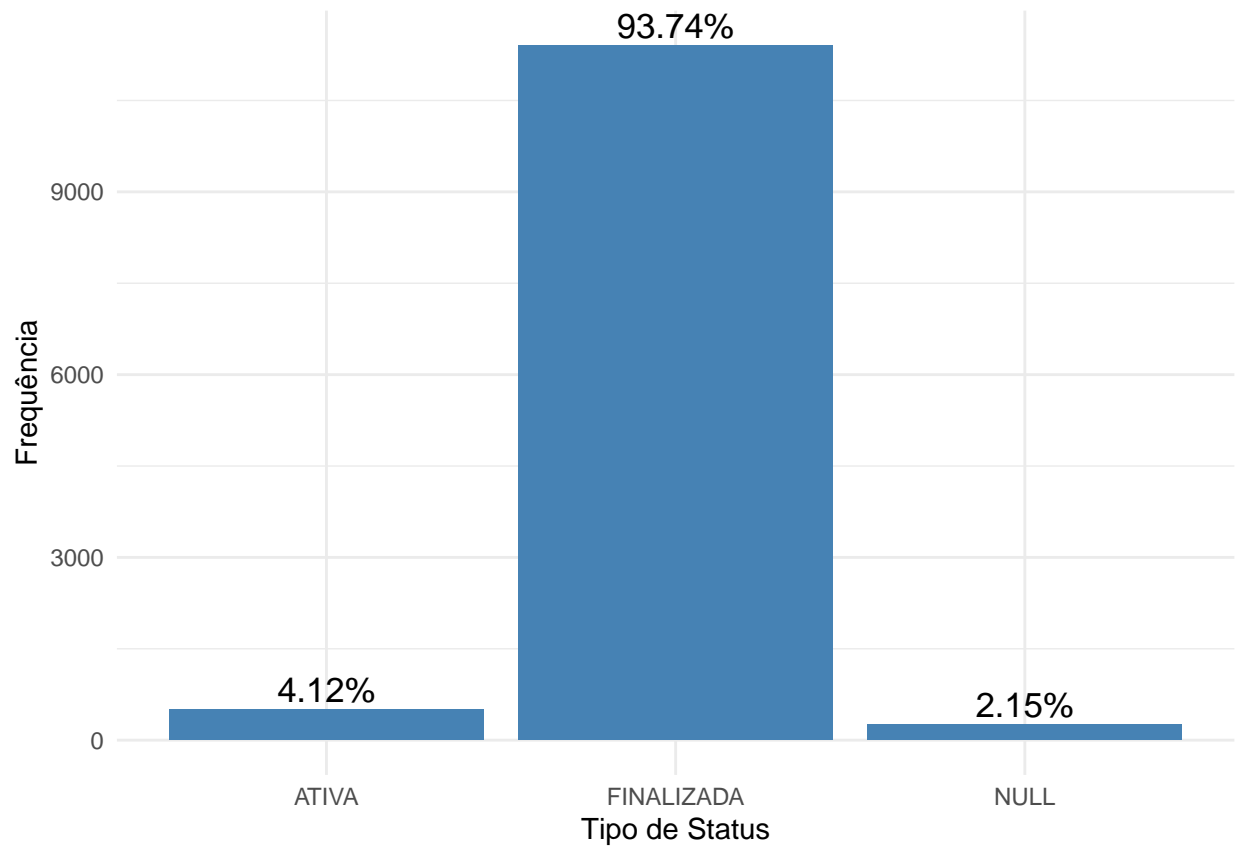
ggplot(data=ocorrencia, aes(x=ocorrencia_classificacao)) +
  geom_bar(stat="count", fill="steelblue")+
  geom_text(aes(label= paste0(round(prop.table(..count..)*100,2), "%")),
            stat= "count", vjust=-0.3, size=4.5)+ylab("Frequência") +
  xlab("Tipo de classificação") +
  theme_minimal()

```

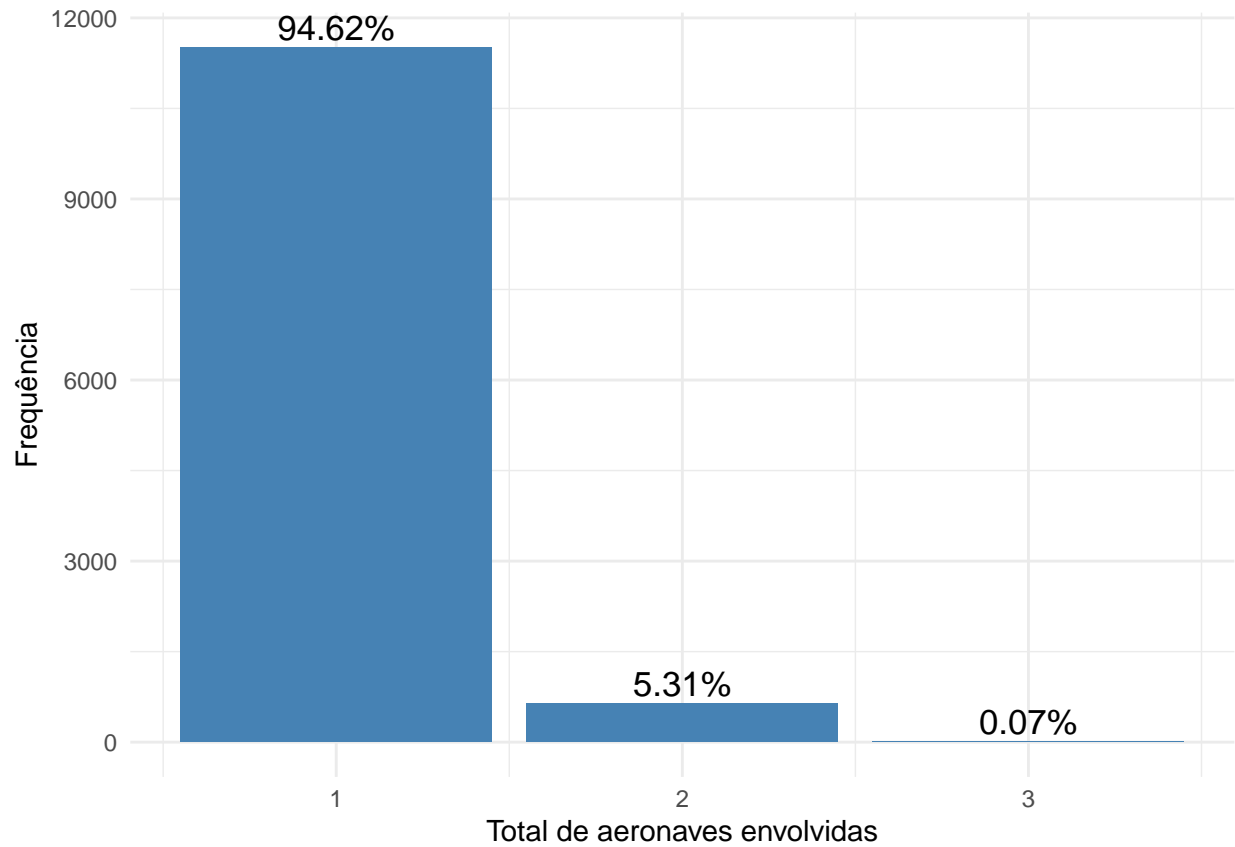


No que se refere a classificação de ocorrência de acidentes aéreos, tem-se que 56.36% das ocorrências são de fato acidentes, 25.49% são incidentes e apenas 18.19% são incidentes graves.

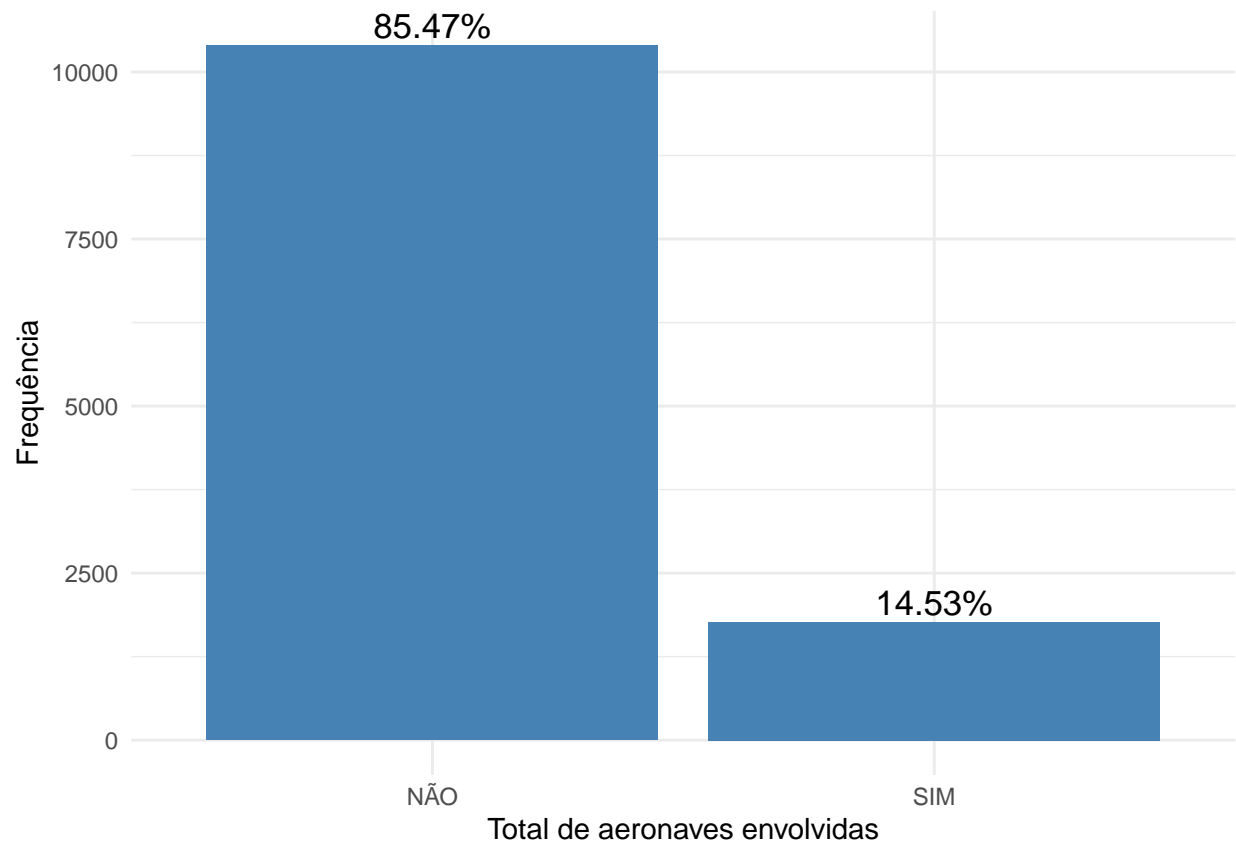
```
ggplot(data=dados, aes(x=investigacao_status)) +
  geom_bar(stat="count", fill="steelblue")+
  geom_text(aes(label= paste0(round(prop.table(..count..)*100,2), "%"), stat= "count", vjust=-0.3, size=12),
    ylab("Frequência") +
    xlab("Tipo de Status") +
    theme_minimal()
```



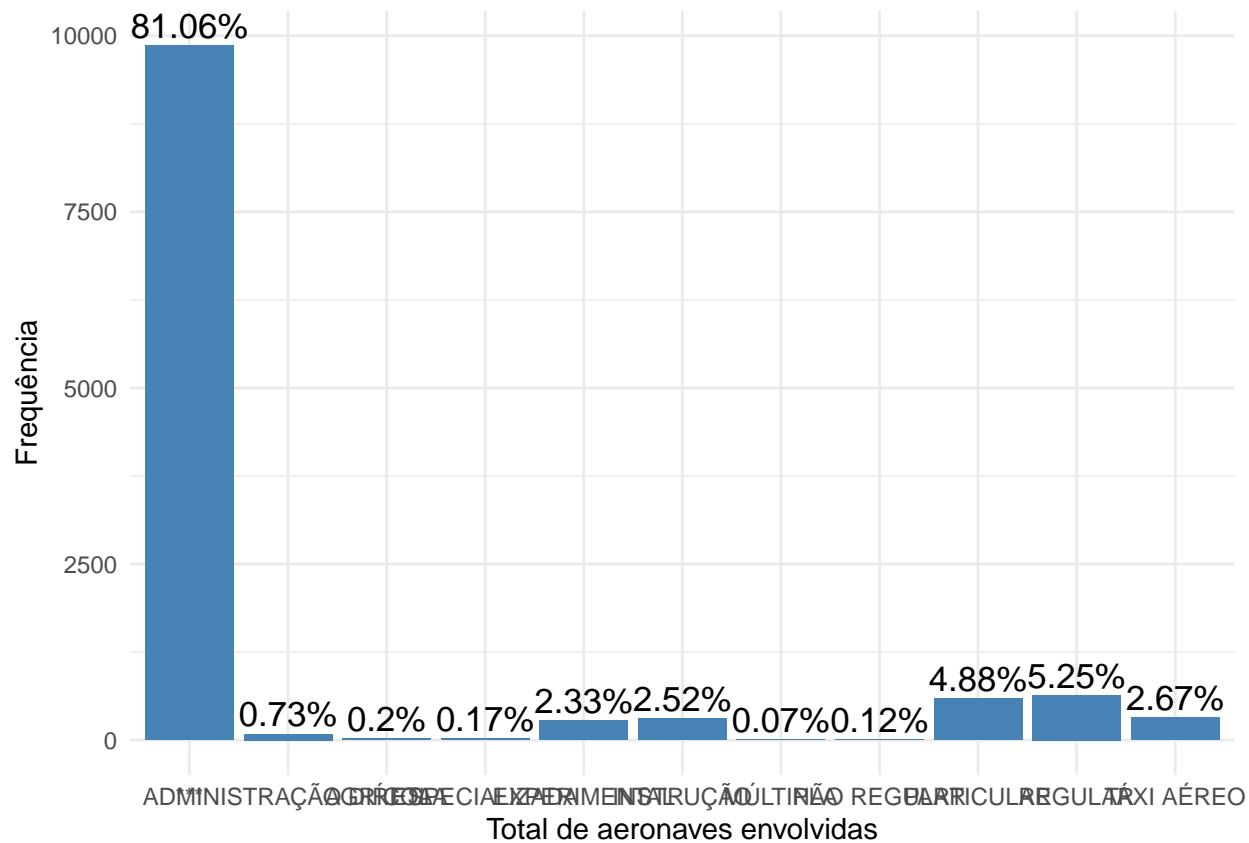
```
ggplot(data=dados, aes(x=total_aeronaves_envolvidas)) +  
  geom_bar(stat="count", fill="steelblue")+  
  geom_text(aes(label= paste0(round(prop.table(..count..)*100,2), "%")),stat= "count", vjust=-0.3, size=12)+  
  ylab("Frequência") +  
  xlab("Total de aeronaves envolvidas") +  
  theme_minimal()
```



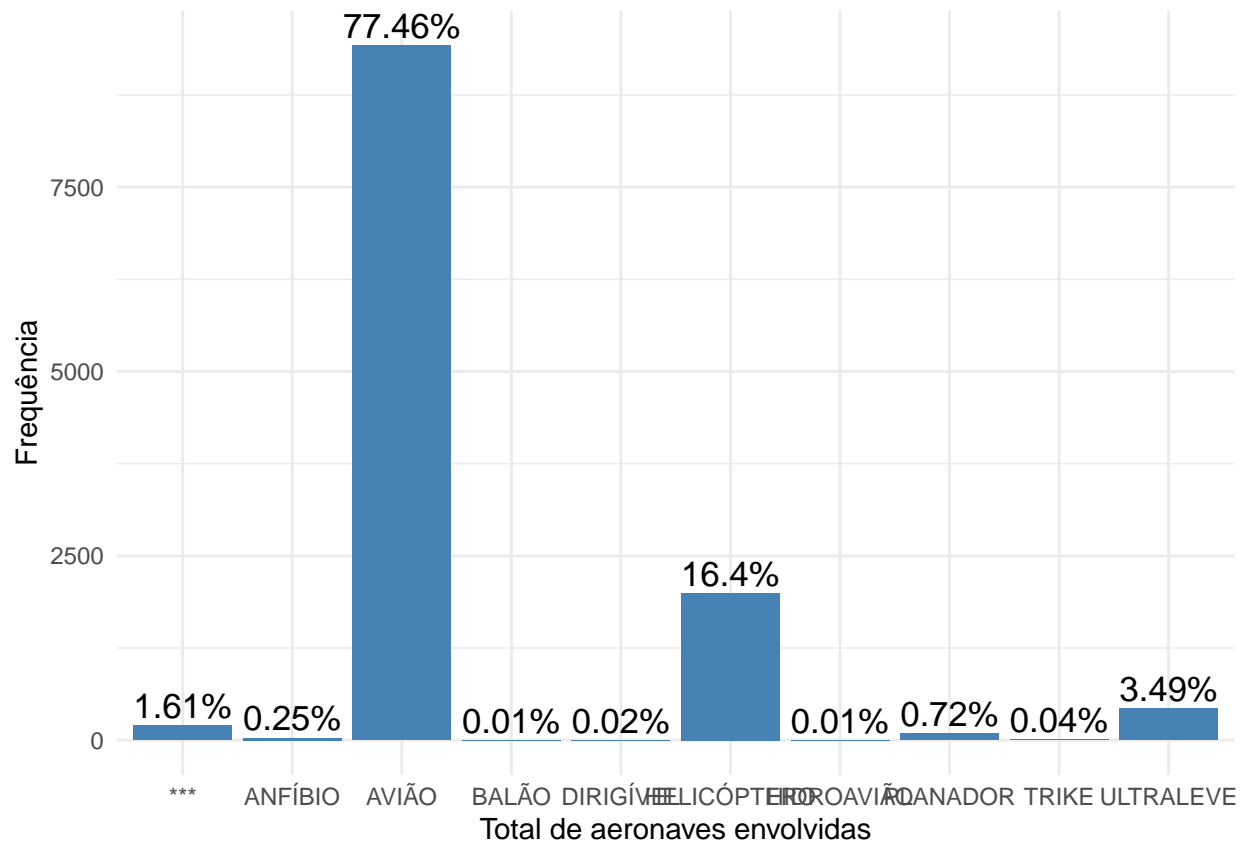
```
ggplot(data=dados, aes(x=ocorrendia_saida_pista)) +  
  geom_bar(stat="count", fill="steelblue")+  
  geom_text(aes(label= paste0(round(prop.table(..count..)*100,2), "%")),stat= "count", vjust=-0.3, size=12)+  
  ylab("Frequência") +  
  xlab("Total de aeronaves envolvidas") +  
  theme_minimal()
```



```
ggplot(data=dados, aes(x=aeronave_operador_categoria)) +  
  geom_bar(stat="count", fill="steelblue")+  
  geom_text(aes(label= paste0(round(prop.table(..count..)*100,2), "%")),stat= "count", vjust=-0.3, size=12, color="black")+  
  ylab("Frequência") +  
  xlab("Total de aeronaves envolvidas") +  
  theme_minimal()
```



```
ggplot(data=dados, aes(x=aeronave_tipo_veiculo)) +
  geom_bar(stat="count", fill="steelblue")+
  geom_text(aes(label= paste0(round(prop.table(..count..)*100,2), "%")),stat= "count", vjust=-0.3, size=10) +
  ylab("Frequência") +
  xlab("Total de aeronaves envolvidas") +
  theme_minimal()
```



```
ggplot(data=dados, aes(x=aeronave_motor_tipo)) +
  geom_bar(stat="count", fill="steelblue")+
  geom_text(aes(label= paste0(round(prop.table(..count..)*100,2), "%")),stat= "count", vjust=-0.3, size=10) +
  ylab("Frequência") +
  xlab("Total de aeronaves envolvidas") +
  theme_minimal()
```