

Análise de Acidentes com animais selvagens envolvendo aeronaves 1990-2023

Integrantes:

João Carlos Gonçalves, Tomás Mello, Eric, Felipe



12 de dezembro, 2023

Introdução

O relatório visa construir uma análise exploratória de um conjunto de dados chamado STRIKE_REPORTS. Este conjunto de dados inclui uma coleção abrangente de incidentes de colisão entre aviões e animais selvagens envolvendo aeronaves militares, comerciais ou civis de 1990 a 2023. Os incidentes foram obtidos a partir da Base de Dados de Acidentes com Animais Selvagens da Administração Federal da Aviação (FAA). O conjunto abrange uma série de informações relacionadas aos incidentes, incluindo pormenores sobre as aeronaves envolvidas, espécies de animais selvagens, localizações geográficas e vários fatores que contribuem para as ocorrências e análise.

Objetivos

Objetivos

-Pontos para analisar, correlacoes com batida em relacao ao voo, trajetoria, velocidade, altura. Se quanto mais rapido estiver mais dano tem o aviao -Da para tirar conclusoes do que é cada parte do voo, em qual altura qual velocidade, etc -Sugerir mudancas nos trajetos ou horarios. Estacoes dos anos que tem mais passaros migrando -Ver quais aeroportos tem mais indices assim -Tipos de avioes, e quais tem mais danos

Variaveis de Interesse com dados completos ou quse completos:

SPECIES_ID ou SPECIES especie que mais tem impacto

SOURCE

NR_FATALITIES OU NR_INJURIES, Poucos observacoes para o tamanho da amostra, mas é muito interessante

Size tamanho do passaro

Birds_Struck muito bom pare correlacionar com o dano, o tamanho também

REMAINS_COLLECTED para combinar com a especie e ver se ajudou a identificar

EFFECT poucos efeitos na maioria dos casos, mas interessante ver se o impacto teve que interditar voo ou etc

COST_REPAIRS & COST_OTHERS importante para ver os custos de concertar a aeronave, ou os custos com gasolina, cancelar o voo, reembolssar passagens, hotel, etc. Porem so tem 5000 observacoes para cada. É uma variavel quantitativa, boa para fazer correlação/boxplot. Tem os ajustados a inflação para comparar melhor

AOS Tempo que a aeronave teve que ficar sem vooar dps, 14k observações. Outra quantitativa, boa. Se for blank/vazio, é pq nao ficou nenhuma hora sem

Distance quantitativa, mas boa parte é zero

PRECIPITATION - Se tinha chuva, nevoa, etc Latitude e longitude, mas como usar?

E variaveis de tempo/data: INCIDENT_MONTH, INCIDENT_YEAR, INCIDENT_MONTH, TIME_OF_DAY, TIME

E muitas quantitativas falando onde na aeronave ocorreu o impacto e onde teve dano, achar algum jeito de incorporar

Carregar bibliotecas e ler a base de dados

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Traduzir ou transformar valores

Tabelas (Lembrar de fazer com e sem os valores desconhecidos, pq eles sao os maiores), e também adicionar mais informacoes na tabela fora o numero de voos, o que botar?

```
## # A tibble: 607 x 2
##   OPID  Voos
##   <chr> <int>
## 1 UNK   81287
## 2 BUS   24493
## 3 SWA   24200
## 4 AAL   20481
## 5 UAL   17603
## 6 DAL   13109
## 7 FDX    9602
```

```
## 8 SKW      6778
## 9 UPS      6377
## 10 JBU     4983
## # i 597 more rows
```

```
## # A tibble: 602 x 2
##   OPERATOR      Voos
##   <chr>        <int>
## 1 UNKNOWN      81307
## 2 BUSINESS    24493
## 3 SOUTHWEST AIRLINES 24201
## 4 AMERICAN AIRLINES 20481
## 5 UNITED AIRLINES  17603
## 6 DELTA AIR LINES  13110
## 7 FEDEX EXPRESS   9602
## 8 SKYWEST AIRLINES  6778
## 9 UPS AIRLINES    6377
## 10 JETBLUE AIRWAYS  4983
## # i 592 more rows
```

```
## # A tibble: 601 x 2
##   OPERATOR      Voos
##   <chr>        <int>
## 1 BUSINESS    24493
## 2 SOUTHWEST AIRLINES 24201
## 3 AMERICAN AIRLINES 20481
## 4 UNITED AIRLINES  17603
## 5 DELTA AIR LINES  13110
## 6 FEDEX EXPRESS   9602
## 7 SKYWEST AIRLINES  6778
## 8 UPS AIRLINES    6377
## 9 JETBLUE AIRWAYS  4983
## 10 1US AIRWAYS     3961
## # i 591 more rows
```

```
## # A tibble: 2,615 x 2
##   AIRPORT      Voos
##   <chr>        <int>
## 1 UNKNOWN      35418
## 2 DENVER INTL AIRPORT 9620
## 3 DALLAS/FORT WORTH INTL ARPT 7759
## 4 CHICAGO O'HARE INTL ARPT 6239
## 5 JOHN F KENNEDY INTL 5984
## 6 MEMPHIS INTL 4813
## 7 SALT LAKE CITY INTL 3722
## 8 SACRAMENTO INTL 3608
## 9 DETROIT METRO WAYNE COUNTY ARPT 3600
## 10 ORLANDO INTL 3589
## # i 2,605 more rows
```

```
## # A tibble: 2,614 x 2
##   AIRPORT      Voos
##   <chr>        <int>
## 1 DENVER INTL AIRPORT 9620
## 2 DALLAS/FORT WORTH INTL ARPT 7759
## 3 CHICAGO O'HARE INTL ARPT 6239
## 4 JOHN F KENNEDY INTL 5984
```

```
## 5 MEMPHIS INTL 4813
## 6 SALT LAKE CITY INTL 3722
## 7 SACRAMENTO INTL 3608
## 8 DETROIT METRO WAYNE COUNTY ARPT 3600
## 9 ORLANDO INTL 3589
## 10 HARTSFIELD - JACKSON ATLANTA INTL ARPT 3529
## # i 2,604 more rows
```

```
## # A tibble: 603 x 2
##   AIRCRAFT Voos
##   <chr> <int>
## 1 UNKNOWN 81535
## 2 B-737-700 14287
## 3 A-320 13993
## 4 B-737-800 11100
## 5 CRJ100/200 9503
## 6 B-737-300 9337
## 7 A-319 8617
## 8 B-757-200 8219
## 9 EMB-170 7572
## 10 EMB-145 5985
## # i 593 more rows
```

```
## # A tibble: 602 x 2
##   AIRCRAFT Voos
##   <chr> <int>
## 1 B-737-700 14287
## 2 A-320 13993
## 3 B-737-800 11100
## 4 CRJ100/200 9503
## 5 B-737-300 9337
## 6 A-319 8617
## 7 B-757-200 8219
## 8 EMB-170 7572
## 9 EMB-145 5985
## 10 A-321 5348
## # i 592 more rows
```

```
## # A tibble: 68 x 2
##   STATE Voos
##   <chr> <int>
## 1 "" 35501
## 2 "TX" 24794
## 3 "FL" 20616
## 4 "CA" 20431
## 5 "NY" 14371
## 6 "CO" 12420
## 7 "IL" 11942
## 8 "TN" 8503
## 9 "PA" 8190
## 10 "OH" 8133
## # i 58 more rows
```

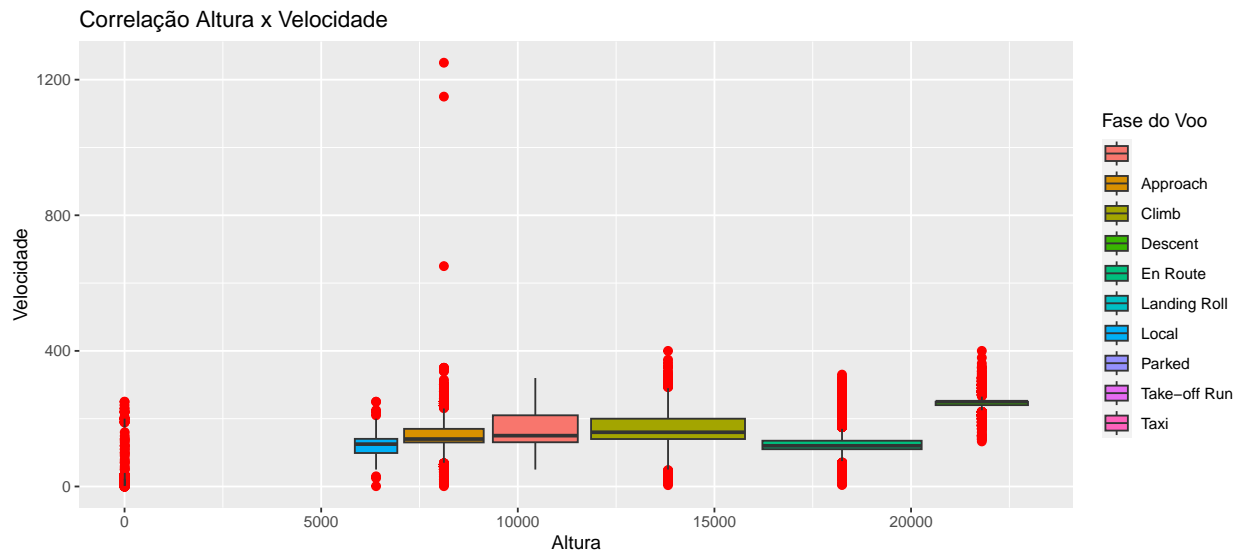
```
## # A tibble: 67 x 2
##   STATE Voos
##   <chr> <int>
## 1 TX 24794
```

```
## 2 FL 20616
## 3 CA 20431
## 4 NY 14371
## 5 CO 12420
## 6 IL 11942
## 7 TN 8503
## 8 PA 8190
## 9 OH 8133
## 10 NJ 7282
## # i 57 more rows
```

```
#Grafico de Correlação entre altura e velocidade, mas nao acho que esse seja bom, pois são duas variaveis quantitativas
def_strikes %>%
  ggplot(aes(x=HEIGHT , y=SPEED, fill=PHASE_OF_FLIGHT)) + geom_boxplot(outlier.color = "red", outlier.size = 2, outlier.fill="red")
```

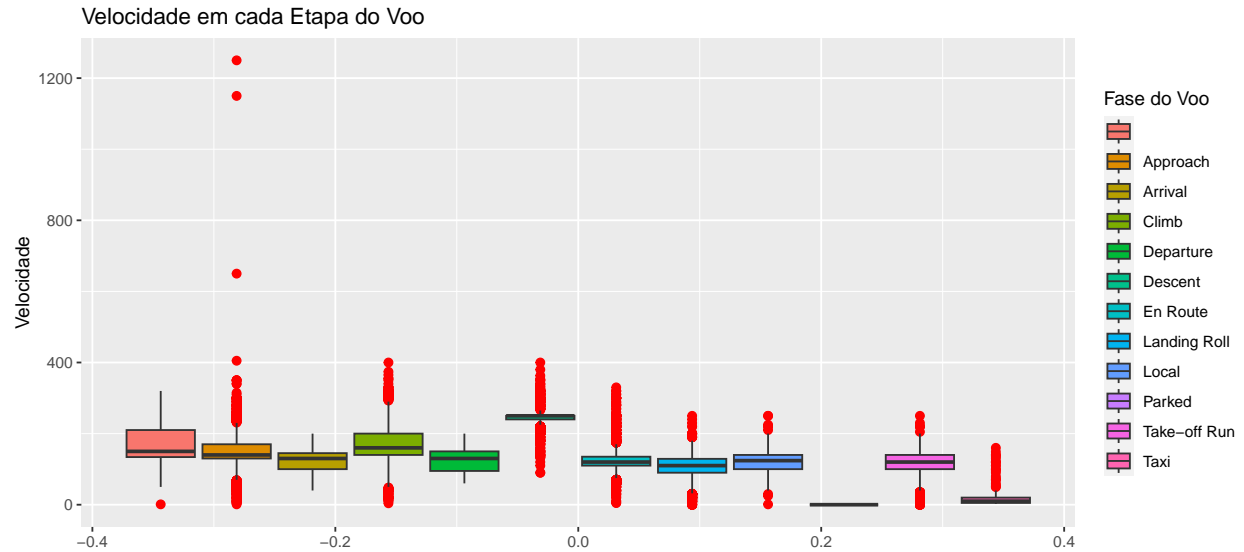
```
## Warning: Removed 140003 rows containing missing values ('stat_boxplot()').
```

```
## Warning: Removed 55954 rows containing non-finite values ('stat_boxplot()').
```



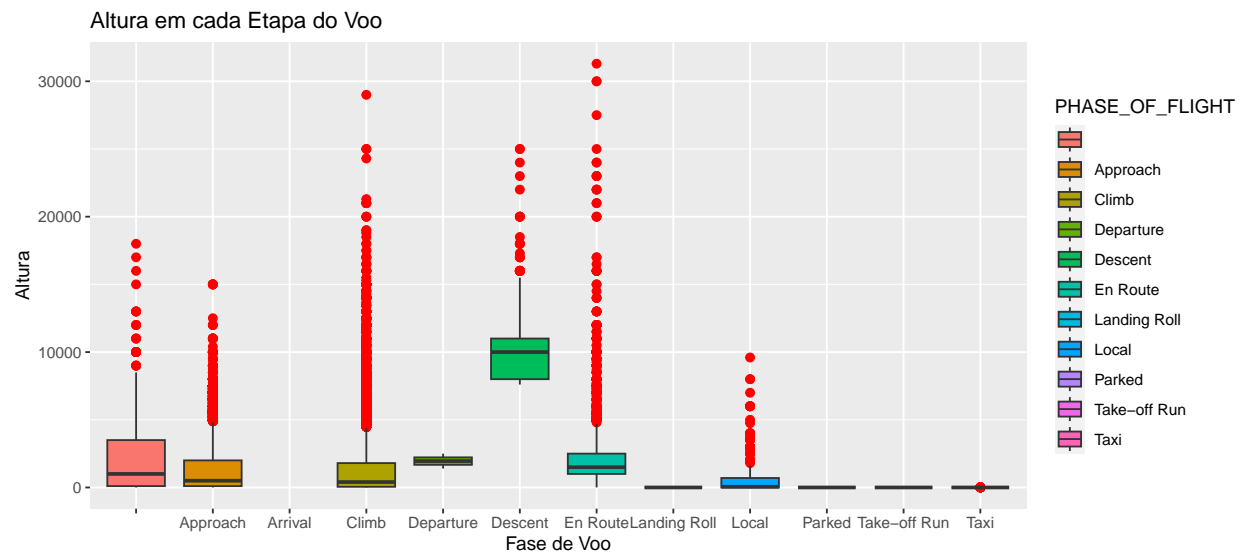
```
#Boxplot da velocidade em cada etapa do Voo
def_strikes %>%
  ggplot(aes(y=SPEED, fill=PHASE_OF_FLIGHT)) + geom_boxplot(outlier.color = "red", outlier.size = 2, outlier.fill="red")
```

```
## Warning: Removed 194573 rows containing non-finite values ('stat_boxplot()').
```



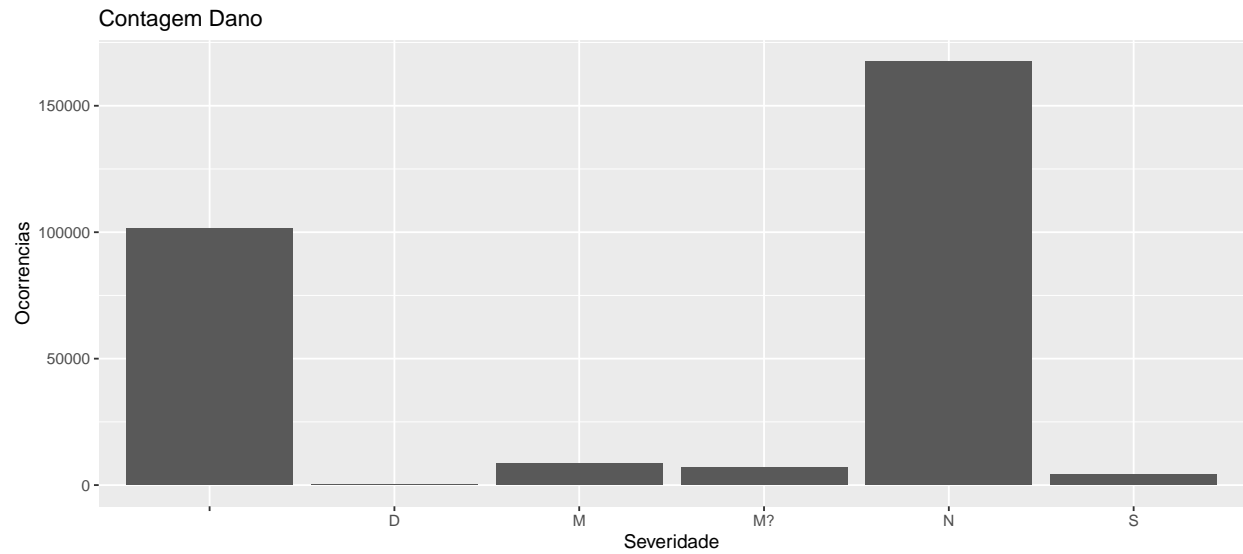
```
#Altura em cada fase de voo"
def_strikes %>%
  ggplot(aes(x=PHASE_OF_FLIGHT, y=HEIGHT, fill=PHASE_OF_FLIGHT)) + geom_boxplot(outlier.color = "red", outlier.size=
```

```
## Warning: Removed 140003 rows containing non-finite values ('stat_boxplot()').
```

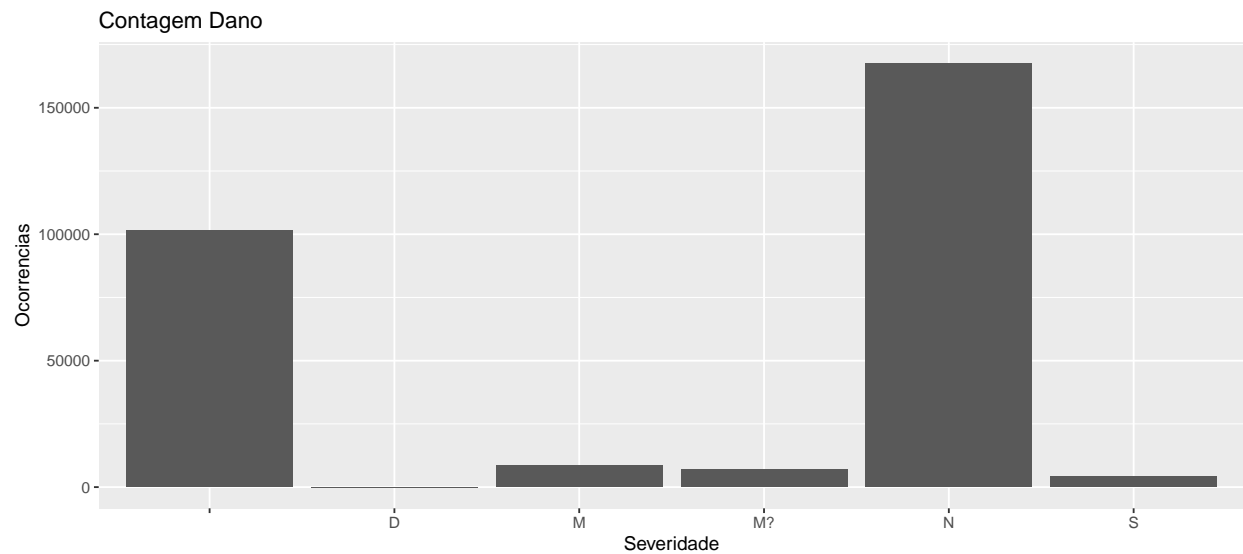


```
#Grafico de Barras do dano, o fill por AC_Mass nao esta funcionando
def_strikes %>%
  ggplot(aes(x= DAMAGE_LEVEL, fill =AC_MASS)) + geom_bar() + labs(title="Contagem Dano", x="Severidade", y="Ocorren
```

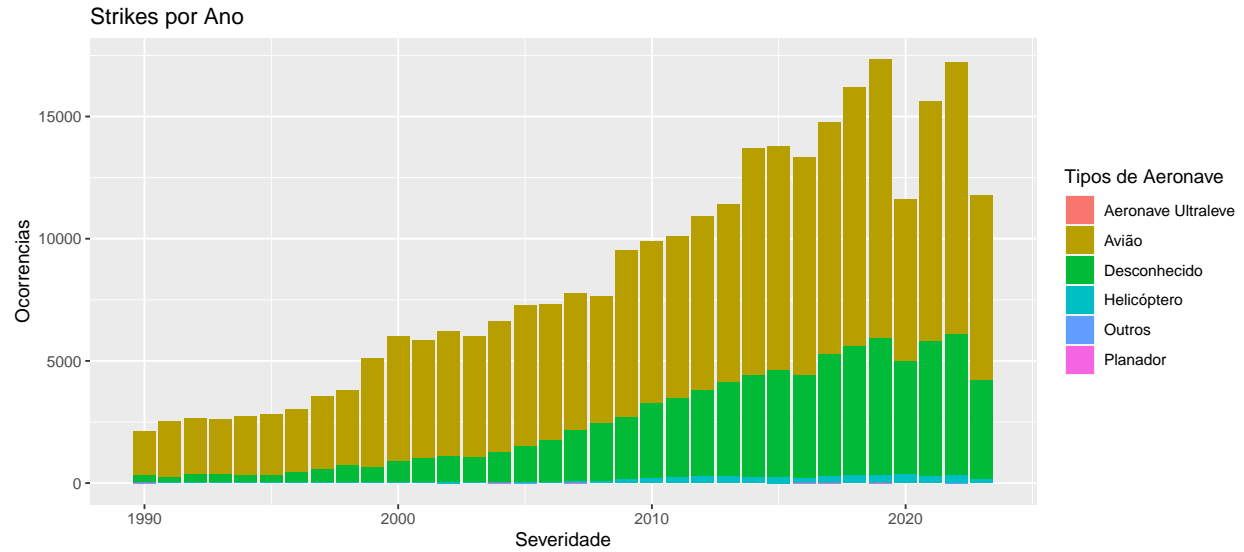
```
## Warning: The following aesthetics were dropped during statistical transformation: fill
## i This can happen when ggplot fails to infer the correct grouping structure in
## the data.
## i Did you forget to specify a 'group' aesthetic or to convert a numerical
## variable into a factor?
```



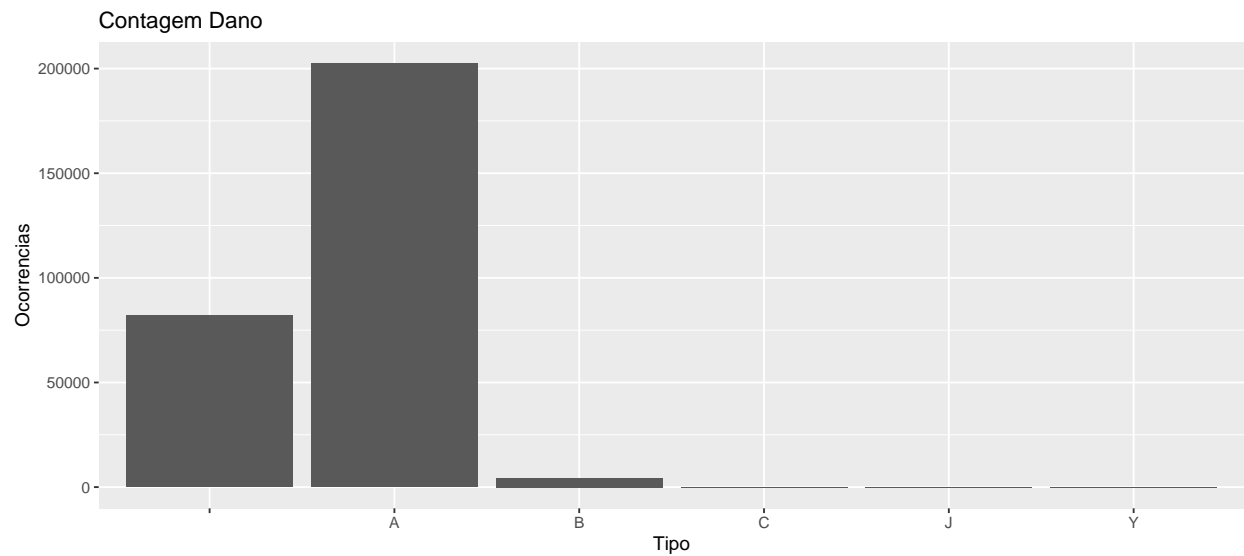
```
#Contagem da severidade dos impactos
def_strikes %>%
  ggplot(aes(x= DAMAGE_LEVEL)) + geom_bar() + labs(title="Contagem Dano", x="Severidade", y="Ocorrencias")
```



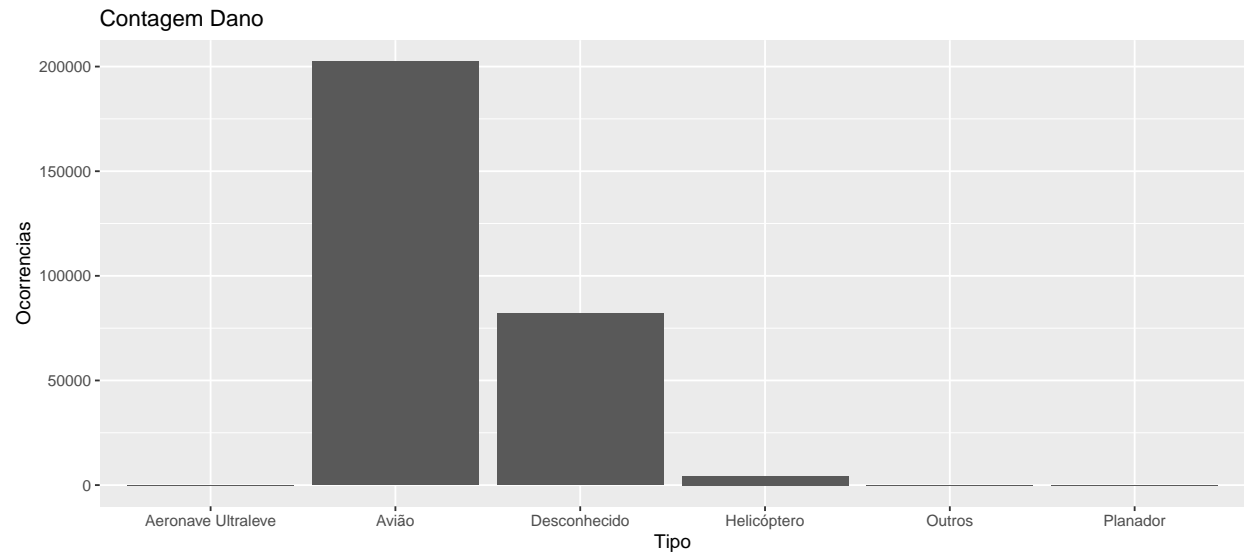
```
#Anos e batidas de aeronave, fazer a sem os unknow/desconhecidos tambem mais para frente
def_strikes %>%
  ggplot(aes(x= INCIDENT_YEAR, fill = Tipo_de_Aeronave)) + geom_bar() + labs(title="Strikes por Ano", x="Severidade")
```

```
#teste para ver se a mudança de nome esta entregando os mesmos resultados para a contagem de ocorrencias com cada t
def_strikes %>%
  ggplot(aes(x= AC_CLASS)) + geom_bar() + labs(title="Contagem Dano", x="Tipo", y="Ocorrencias")
```

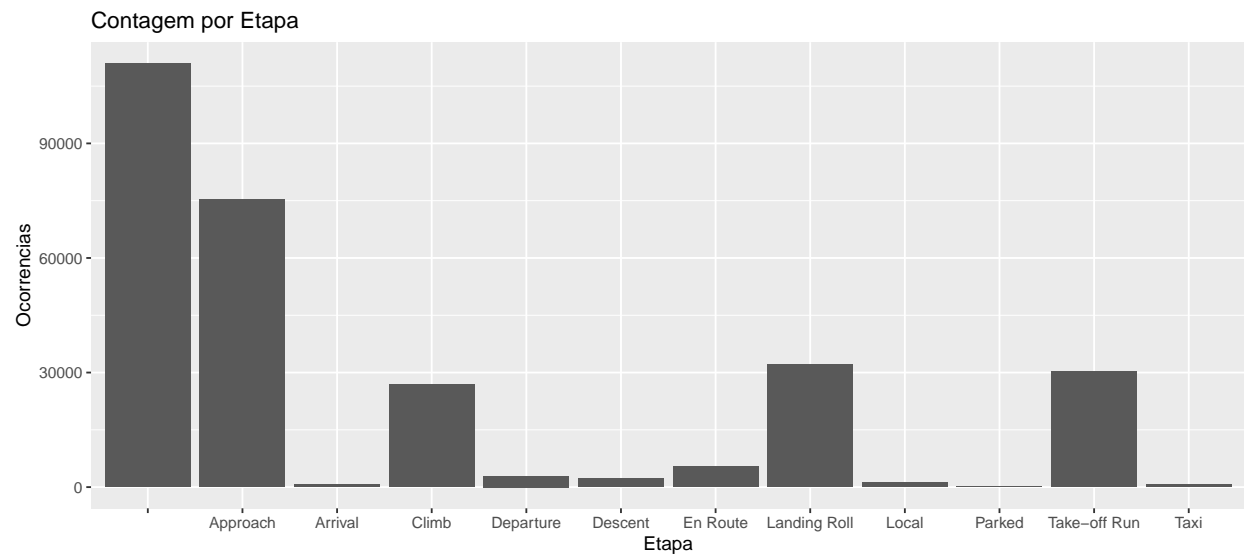


```
def_strikes %>%
  ggplot(aes(x= Tipo_de_Aeronave)) + geom_bar() + labs(title="Contagem Dano", x="Tipo", y="Ocorrencias")
```



#Qual etapa do Voo tem mais batidas, com isso percebemos que quanto mais baixo mais bate, o que é logico, porque as

```
def_strikes %>%
  ggplot(aes(x= PHASE_OF_FLIGHT)) + geom_bar() + labs(title="Contagem por Etapa", x="Etapa", y="Ocorrencias")
```



```
#def_strikes %>%
#  gg(plot)

#def_strikes %>%
```