

Manipulação e Análise de Dados com R

Ricson Ramos

2025-02-20

Importação de Bibliotecas

```
library(dplyr)
library(ggplot2)
library(tidyr)
```

Carregamento dos Dados

```
dados <- read.csv("datatran2024.csv", sep = ";", fill = TRUE, check.names = FALSE, encoding = "latin1")
```

Exploração Inicial dos Dados

```
##Estrutura das Variáveis
str(dados)
```

```
## 'data.frame':    60365 obs. of  30 variables:
## $ id              : chr  "571789" "571804" "571806" "571818" ...
## $ data_inversa     : chr  "01/01/2024" "01/01/2024" "01/01/2024" "01/01/2024" ...
## $ dia_semana       : chr  "segunda-feira" "segunda-feira" "segunda-feira" "segunda-feira" ...
## $ horario          : chr  "03:56:00" "04:50:00" "04:30:00" "06:30:00" ...
## $ uf               : chr  "ES" "PI" "BA" "SE" ...
## $ br               : int   101 343 116 101 364 251 116 251 116 262 ...
## $ km               : chr  "38" "185" "578" "18" ...
## $ municipio        : chr  "CONCEICAO DA BARRA" "PIRIPIRI" "BREJOES" "MALHADA DOS BOIS" ...
## $ causa_acidente   : chr  "Ultrapassagem Indevida" "Manobra de mudança de faixa" "Ingestão de a
## $ tipo_acidente     : chr  "Colisão lateral sentido oposto" "Colisão frontal" "Colisão frontal"
## $ classificacao_acidente: chr  NA "Com Vítimas Fatais" "Com Vítimas Fatais" "Com Vítimas Feridas" .
## $ fase_dia         : chr  "Plena Noite" "Amanhecer" "Plena Noite" "Amanhecer" ...
## $ sentido_via       : chr  "Crescente" "Decrescente" "Decrescente" "Crescente" ...
## $ condicao_meteorologica: chr  "Céu Claro" "Céu Claro" "Céu Claro" "Céu Claro" ...
## $ tipo_pista        : chr  "Simples" "Simples" "Simples" "Dupla" ...
## $ tracado_via       : chr  "Reta" "Reta" "Curva" "Declive;Reta" ...
## $ uso_solo          : chr  "Não" "Sim" "Não" "Não" ...
## $ pessoas           : int   3 2 3 2 3 3 4 4 2 1 ...
## $ mortos            : int   0 1 1 0 0 0 2 0 0 0 ...
## $ feridos_leves     : int   0 0 0 0 0 1 0 2 1 0 ...
```

```
## $ feridos_graves      : int  1 0 0 1 0 0 0 0 0 0 ...
## $ ilesos              : int  1 1 1 0 2 1 2 1 0 1 ...
## $ ignorados           : int  1 0 2 2 1 1 0 3 1 0 ...
## $ feridos             : int  1 0 0 1 0 1 0 2 1 0 ...
## $ veiculos            : int  3 2 4 3 3 3 2 5 2 1 ...
## $ latitude            : chr  "-18,48261" "-4,29603281" "-13,07158302" "-10,35601949" ...
## $ longitude           : chr  "-39,92379" "-41,76732659" "-39,9611107" "-36,90552235" ...
## $ regional            : chr  "SPRF-ES" "SPRF-PI" "SPRF-BA" "SPRF-SE" ...
## $ delegacia           : chr  "DELO4-ES" "DELO2-PI" "DELO3-BA" "DELO2-SE" ...
## $ uop                 : chr  "UOP02-DELO4-ES" "UOP01-DELO2-PI" "UOP02-DELO3-BA" "UOP02-DELO2-SE"
```

```
##Resumo Estatístico
summary(dados)
```

```
##      id      data_inversa      dia_semana      horario
## Length:60365 Length:60365 Length:60365 Length:60365
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##      uf      br      km      municipio
## Length:60365 Min. : 0.0 Length:60365 Length:60365
## Class :character 1st Qu.:101.0 Class :character Class :character
## Mode :character Median :158.0 Mode :character Mode :character
## Mean :209.2
## 3rd Qu.:319.0
## Max. :495.0
## causa_acidente tipo_acidente classificacao_acidente
## Length:60365 Length:60365 Length:60365
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##      fase_dia      sentido_via      condicao_metereologica
## Length:60365 Length:60365 Length:60365
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##      tipo_pista      tracado_via      uso_solo      pessoas
## Length:60365 Length:60365 Length:60365 Min. : 1.000
## Class :character Class :character Class :character 1st Qu.: 2.000
## Mode :character Mode :character Mode :character Median : 2.000
## Mean : 2.588
## 3rd Qu.: 3.000
## Max. :61.000
##      mortos      feridos_leves      feridos_graves      ilesos
## Min. : 0.0000 Min. : 0.0000 Min. : 0.0000 Min. : 0.000
## 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 0.000
## Median : 0.0000 Median : 1.0000 Median : 0.0000 Median : 1.000
## Mean : 0.0831 Mean : 0.8728 Mean : 0.2776 Mean : 1.043
```

```
## 3rd Qu.: 0.0000 3rd Qu.: 1.0000 3rd Qu.: 0.0000 3rd Qu.: 1.000
## Max. :23.0000 Max. :43.0000 Max. :30.0000 Max. :60.000
## ignorados feridos veiculos latitude
## Min. : 0.0000 Min. : 0.00 Min. : 1.000 Length:60365
## 1st Qu.: 0.0000 1st Qu.: 1.00 1st Qu.: 1.000 Class :character
## Median : 0.0000 Median : 1.00 Median : 2.000 Mode :character
## Mean : 0.4018 Mean : 1.15 Mean : 1.995
## 3rd Qu.: 1.0000 3rd Qu.: 1.00 3rd Qu.: 2.000
## Max. :33.0000 Max. :45.00 Max. :26.000
## longitude regional delegacia uop
## Length:60365 Length:60365 Length:60365 Length:60365
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
```

```
##Visualização das Primeiras Linhas
head(dados)
```

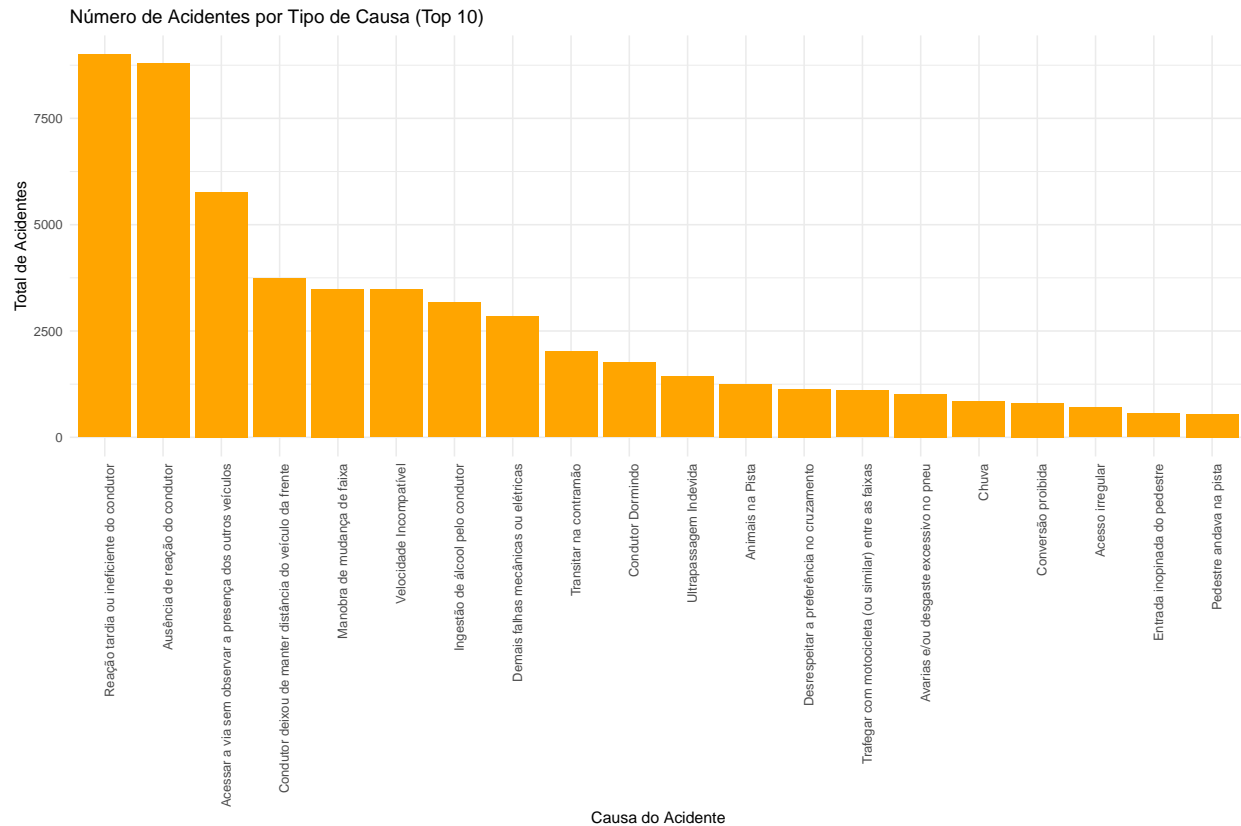
```
##      id data_inversa  dia_semana horario uf  br  km      municipio
## 1 571789 01/01/2024 segunda-feira 03:56:00 ES 101 38 CONCEICAO DA BARRA
## 2 571804 01/01/2024 segunda-feira 04:50:00 PI 343 185 PIRIPIRI
## 3 571806 01/01/2024 segunda-feira 04:30:00 BA 116 578 BREJOES
## 4 571818 01/01/2024 segunda-feira 06:30:00 SE 101 18 MALHADA DOS BOIS
## 5 571838 01/01/2024 segunda-feira 05:00:00 MT 364 240 RONDONOPOLIS
## 6 571855 01/01/2024 segunda-feira 11:50:00 MG 251 447 GRAO MOGOL
##
##      causa_acidente
## 1      Ultrapassagem Indevida
## 2      Manobra de mudança de faixa
## 3      Ingestão de álcool pelo condutor
## 4      Reação tardia ou ineficiente do condutor
## 5 Condutor deixou de manter distância do veículo da frente
## 6      Velocidade Incompatível
##
##      tipo_acidente classificacao_acidente  fase_dia sentido_via
## 1 Colisão lateral sentido oposto <NA> Plena Noite Crescente
## 2      Colisão frontal Com Vítimas Fatais Amanhecer Decrescente
## 3      Colisão frontal Com Vítimas Fatais Plena Noite Decrescente
## 4 Saída de leito carroçável Com Vítimas Feridas Amanhecer Crescente
## 5      Colisão traseira Sem Vítimas Pleno dia Crescente
## 6      Colisão traseira Com Vítimas Feridas Pleno dia Crescente
##
##      condicao_meteorologica tipo_pista tracado_via uso_solo pessoas mortos
## 1 Céu Claro Simples Reta Não 3 0
## 2 Céu Claro Simples Reta Sim 2 1
## 3 Céu Claro Simples Curva Não 3 1
## 4 Céu Claro Dupla Declive;Reta Não 2 0
## 5 Céu Claro Dupla Reta Não 3 0
## 6 Chuva Simples Aclive;Curva Não 3 0
##
##      feridos_leves feridos_graves ilesos ignorados feridos veiculos latitude
## 1 0 1 1 1 1 3 -18,48261
## 2 0 0 1 0 0 2 -4,29603281
## 3 0 0 1 2 0 4 -13,07158302
## 4 0 1 0 2 1 3 -10,35601949
## 5 0 0 2 1 0 3 -16,17914141
```

```
## 6          1          0          1          1          1          3 -16,43408996
##      longitude regional delegacia          uop
## 1      -39,92379  SPRF-ES  DEL04-ES  UOP02-DEL04-ES
## 2 -41,76732659  SPRF-PI  DEL02-PI  UOP01-DEL02-PI
## 3   -39,9611107  SPRF-BA  DEL03-BA  UOP02-DEL03-BA
## 4  -36,90552235  SPRF-SE  DEL02-SE  UOP02-DEL02-SE
## 5  -54,78905337  SPRF-MT  DEL02-MT  UOP01-DEL02-MT
## 6  -43,27992195  SPRF-MG  DEL12-MG  UOP01-DEL12-MG
```

Análises Exploratórias

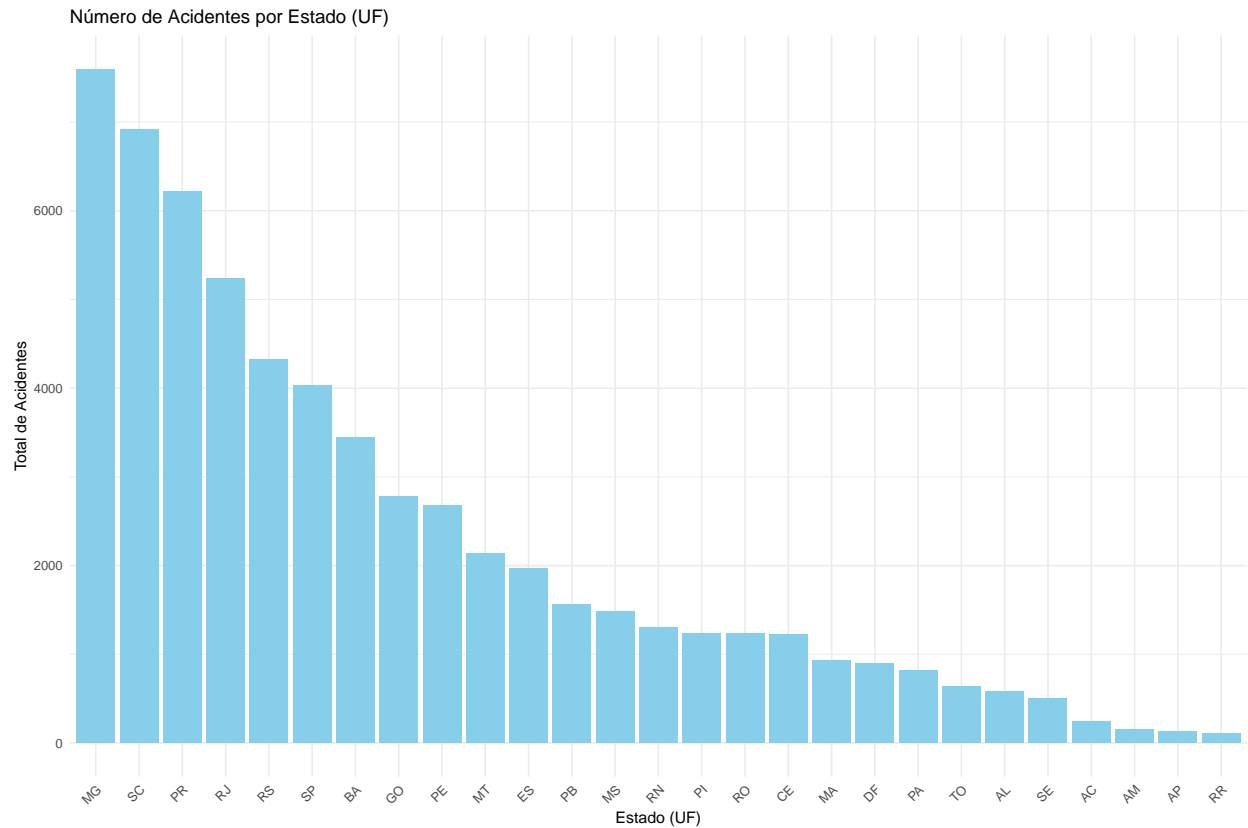
Número de Acidentes por Tipo de Causa (Top 20)

```
ggplot(dados %>%
  group_by(causa_acidente) %>%
  summarise(total = n()) %>%
  arrange(desc(total)) %>%
  head(20), aes(x = reorder(causa_acidente, -total), y = total)) +
geom_bar(stat = "identity", fill = "orange") +
theme_minimal() +
labs(
  title = "Número de Acidentes por Tipo de Causa (Top 10)",
  x = "Causa do Acidente",
  y = "Total de Acidentes"
) +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



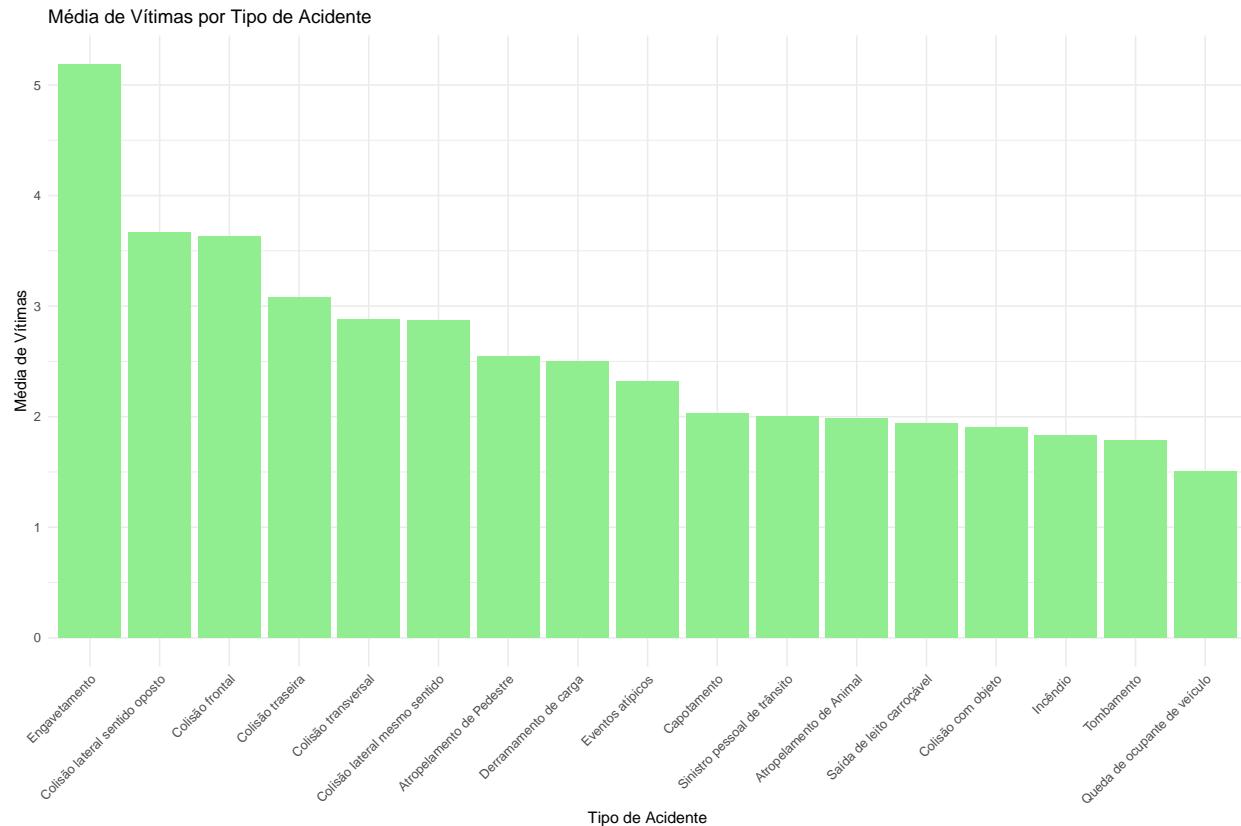
Número de Acidentes por Estado (UF)

```
ggplot(dados %>%
  group_by(uf) %>%
  summarise(total_acidentes = n()) %>%
  arrange(desc(total_acidentes)),
  aes(x = reorder(uf, -total_acidentes), y = total_acidentes)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  theme_minimal() +
  labs(
    title = "Número de Acidentes por Estado (UF)",
    x = "Estado (UF)",
    y = "Total de Acidentes"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



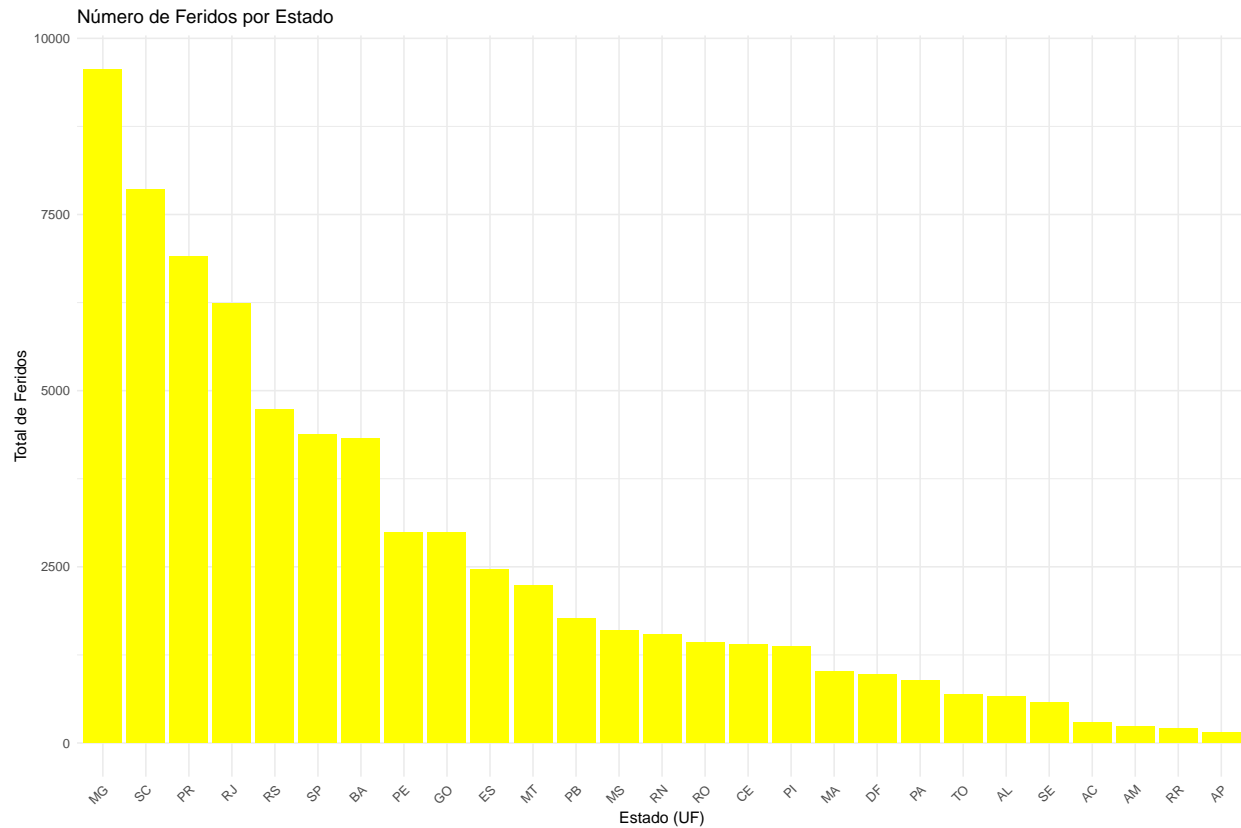
Média de Vítimas por Tipo de Acidente

```
ggplot(dados %>%
  group_by(tipo_acidente) %>%
  summarise(media_vitimas = mean(pessoas, na.rm = TRUE)) %>%
  arrange(desc(media_vitimas)), aes(x = reorder(tipo_acidente, -media_vitimas), y = media_vitimas)
  geom_bar(stat = "identity", fill = "lightgreen") +
  theme_minimal() +
  labs(
    title = "Média de Vítimas por Tipo de Acidente",
    x = "Tipo de Acidente",
    y = "Média de Vítimas"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



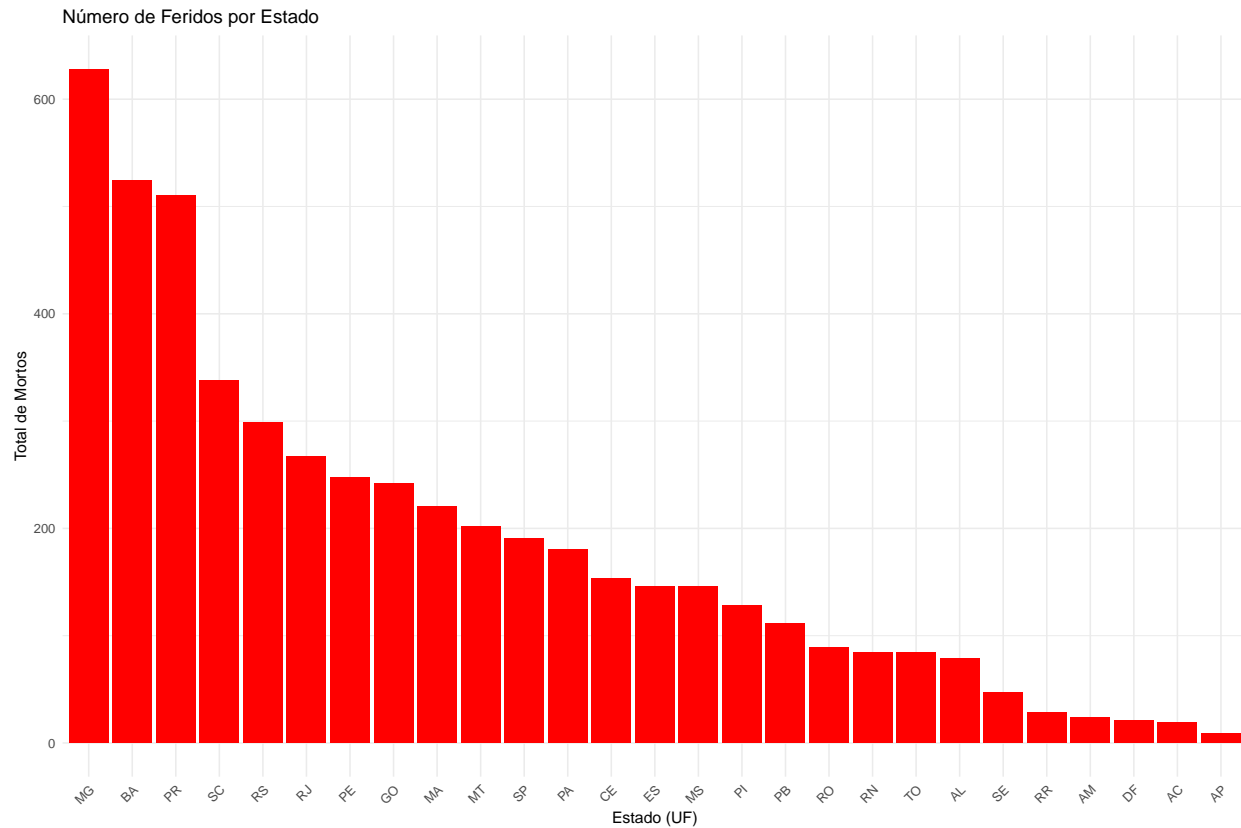
Número de Feridos por Estado

```
dados %>%
  group_by(uf) %>%
  summarise(
    total_feridos = sum(feridos, na.rm = TRUE)
  ) %>%
  arrange(desc(total_feridos)) %>% # Organiza do maior para o menor
  ggplot(aes(x = reorder(uf, -total_feridos), y = total_feridos)) +
  geom_bar(stat = "identity", fill = "yellow") +
  theme_minimal() +
  labs(
    title = "Número de Feridos por Estado",
    x = "Estado (UF)",
    y = "Total de Feridos",
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



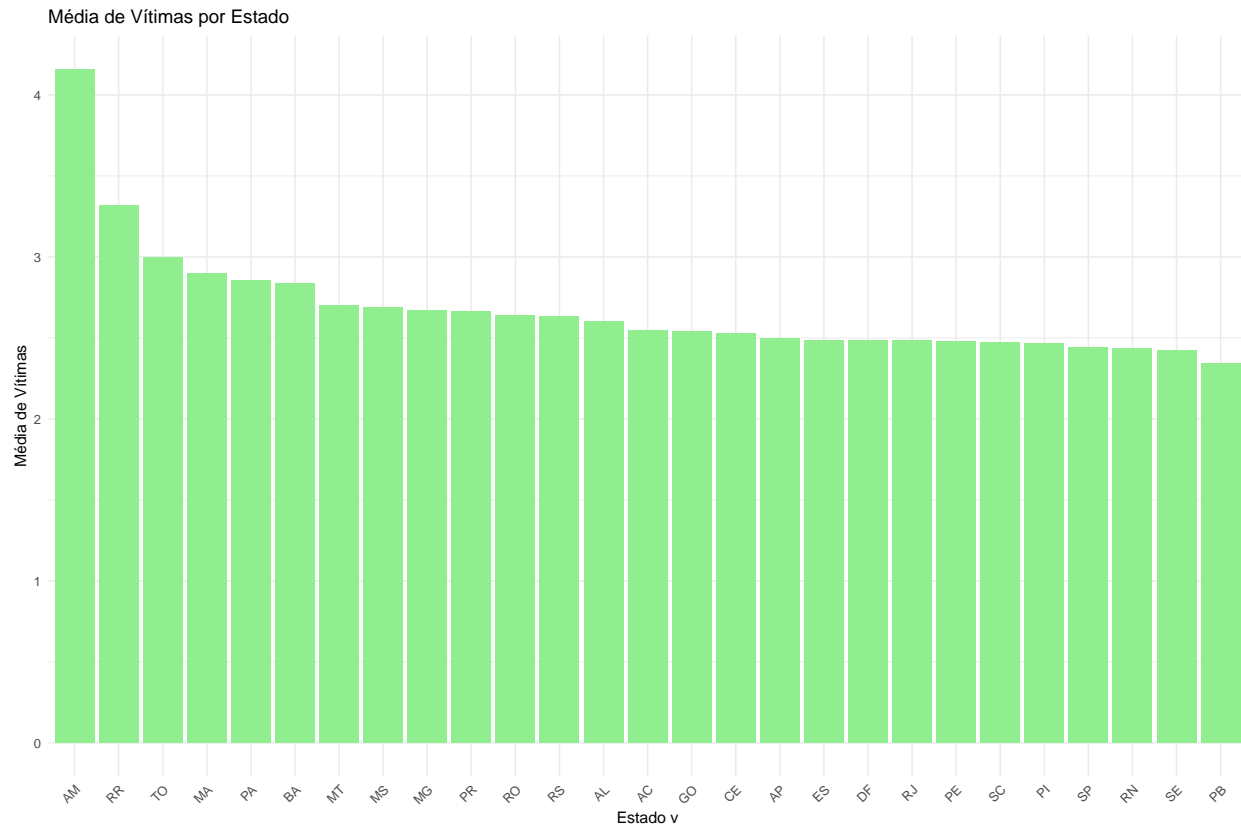
Número de Mortos por Estado

```
dados %>%
  group_by(uf) %>%
  summarise(
    total_mortos = sum(mortos, na.rm = TRUE)
  ) %>%
  arrange(desc(total_mortos)) %>%
  ggplot(aes(x = reorder(uf, -total_mortos), y = total_mortos)) +
  geom_bar(stat = "identity", fill = "red") +
  theme_minimal() +
  labs(
    title = "Número de Feridos por Estado",
    x = "Estado (UF)",
    y = "Total de Mortos",
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Média de Vítimas por Estado (UF)

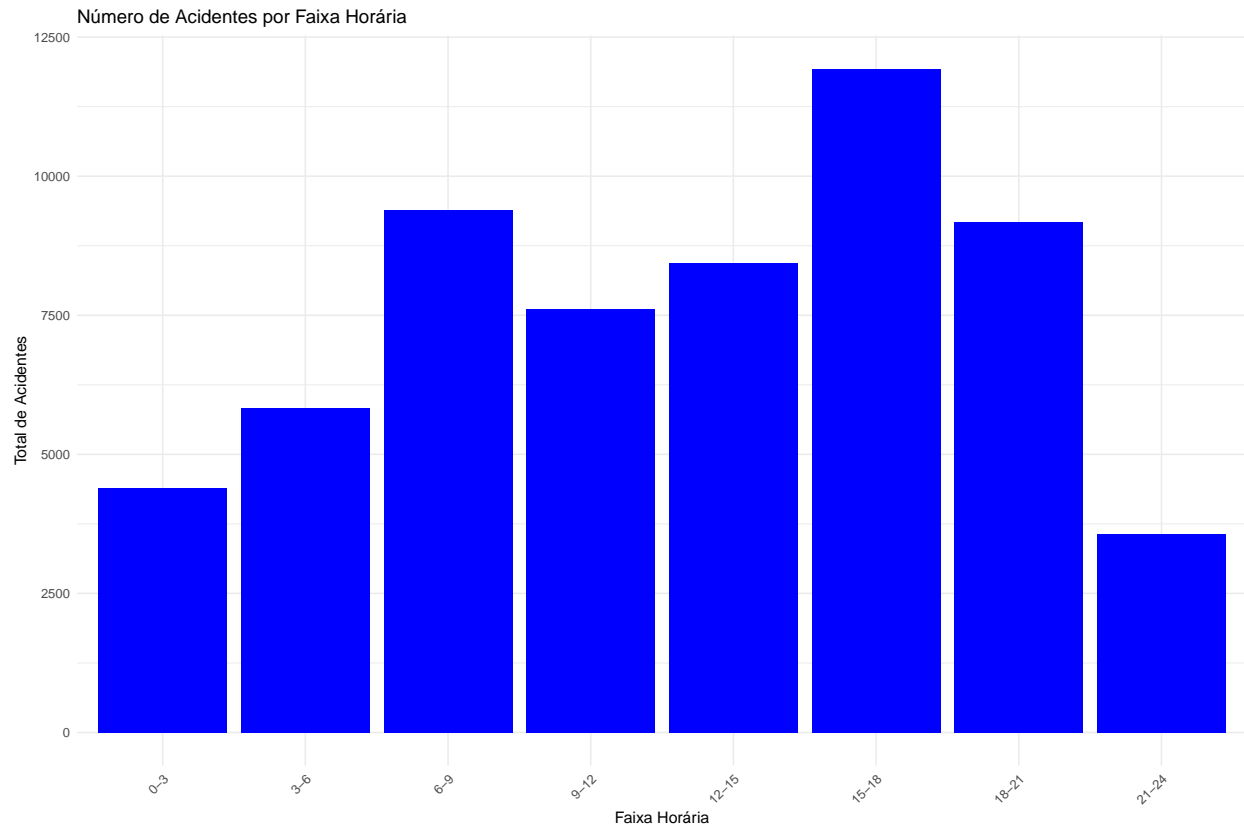
```
ggplot(dados %>%
  group_by(uf) %>%
  summarise(media_vitimas = mean(pessoas, na.rm = TRUE)) %>%
  arrange(desc(media_vitimas), aes(x = reorder(uf, -media_vitimas), y = media_vitimas)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  theme_minimal() +
  labs(
    title = "Média de Vítimas por Estado",
    x = "Estado v",
    y = "Média de Vítimas"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Número de Acidentes por Faixa de Hora (baseado em horário)

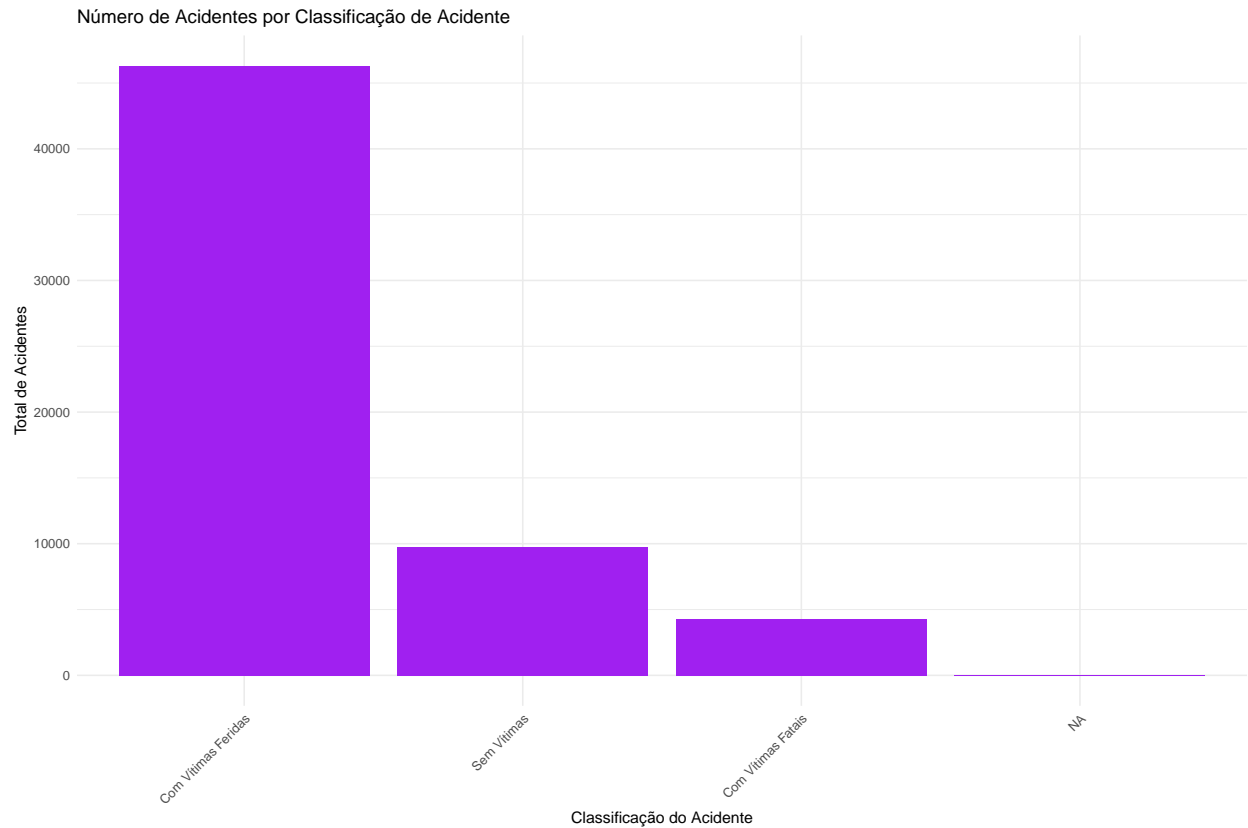
```
dados %>%
  mutate(faixa_horaria = cut(as.numeric(substr(horario, 1, 2)),
                             breaks = seq(0, 24, by = 3),
                             labels = paste0(seq(0, 21, by = 3), "-", seq(3, 24, by = 3)),
                             include.lowest = TRUE)) %>%

  group_by(faixa_horaria) %>%
  summarise(total_acidentes = n()) %>%
  ggplot(aes(x = faixa_horaria, y = total_acidentes)) +
  geom_bar(stat = "identity", fill = "blue") +
  theme_minimal() +
  labs(
    title = "Número de Acidentes por Faixa Horária",
    x = "Faixa Horária",
    y = "Total de Acidentes"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



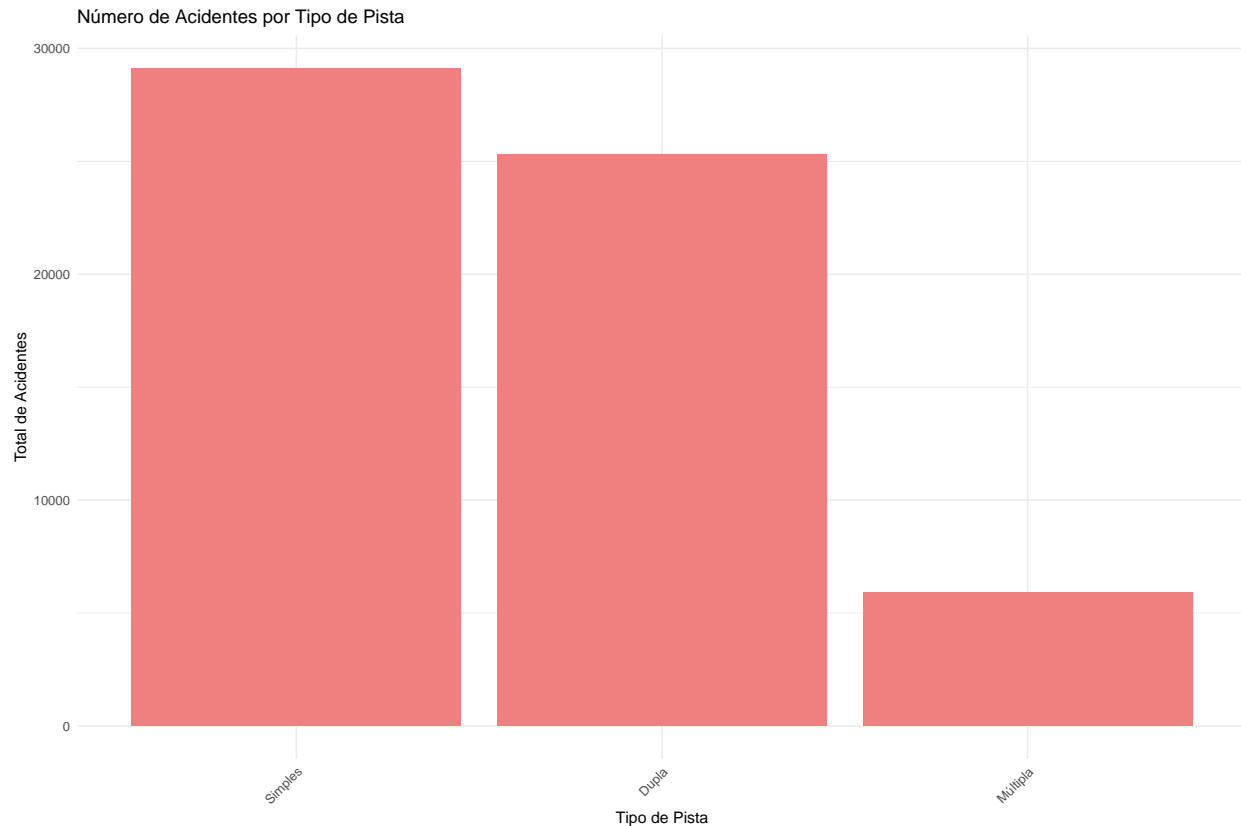
Análise de Acidentes por Classificação de Acidente

```
ggplot(dados %>%
  group_by(classificacao_acidente) %>%
  summarise(total_acidentes = n()) %>%
  arrange(desc(total_acidentes)), aes(x = reorder(classificacao_acidente, -total_acidentes), y =
  geom_bar(stat = "identity", fill = "purple") +
  theme_minimal() +
  labs(
    title = "Número de Acidentes por Classificação de Acidente",
    x = "Classificação do Acidente",
    y = "Total de Acidentes"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



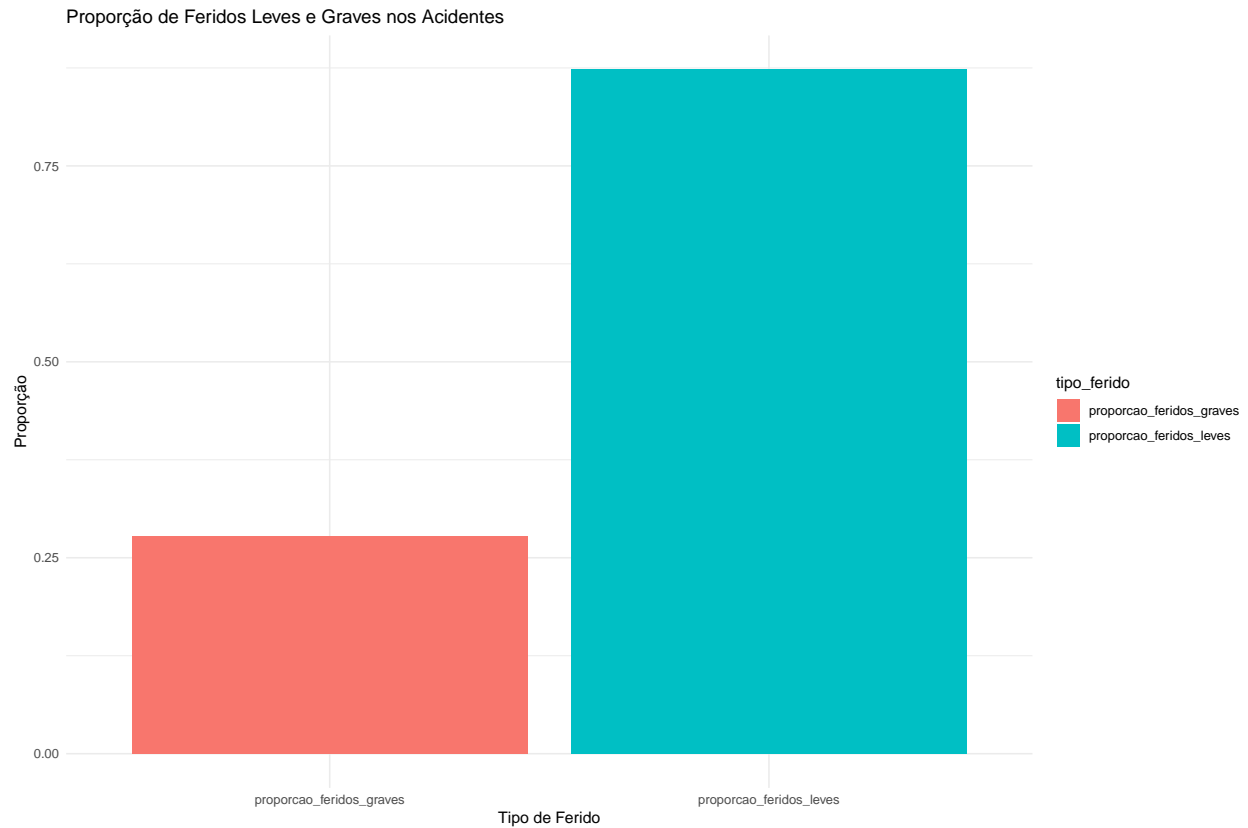
Distribuição de Acidentes por Tipo de Pista

```
ggplot(dados %>%
  group_by(tipo_pista) %>%
  summarise(total_acidentes = n()) %>%
  arrange(desc(total_acidentes)), aes(x = reorder(tipo_pista, -total_acidentes), y = total_acidentes)
) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  theme_minimal() +
  labs(
    title = "Número de Acidentes por Tipo de Pista",
    x = "Tipo de Pista",
    y = "Total de Acidentes"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



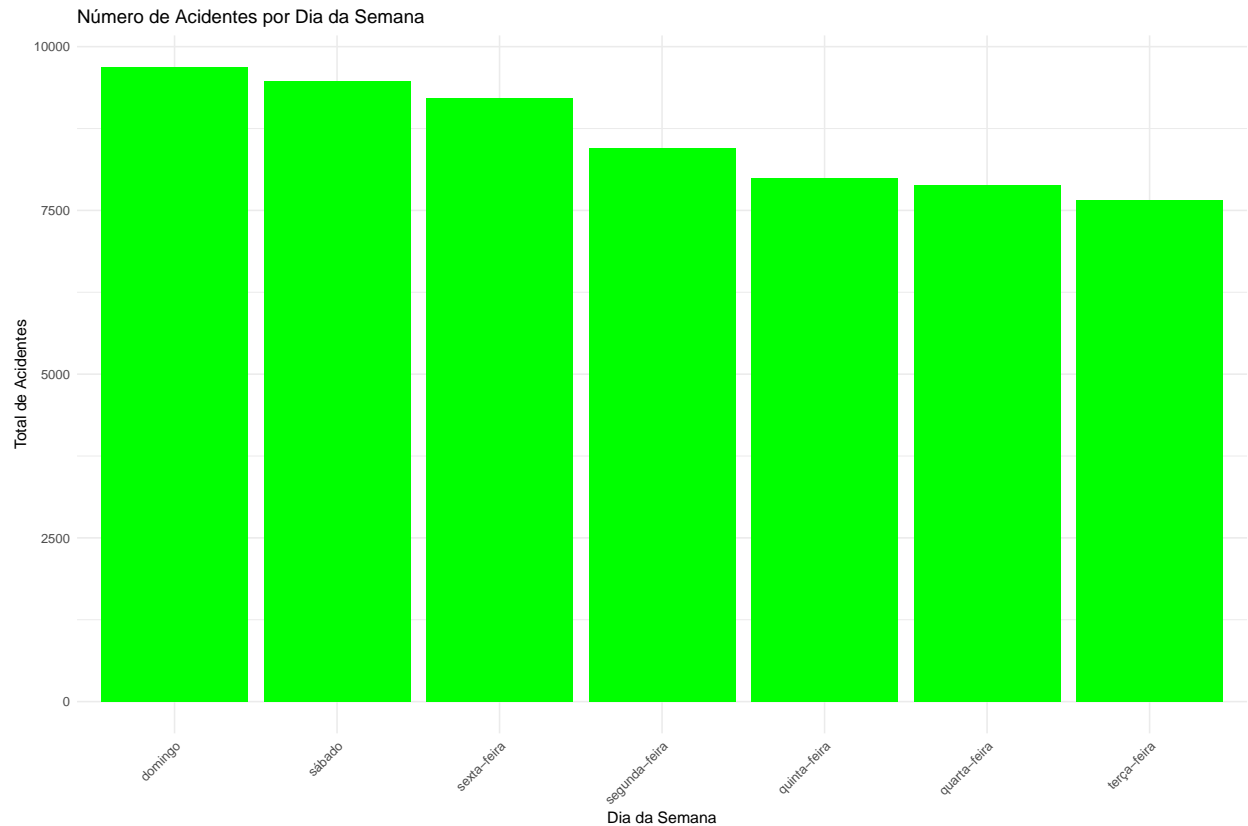
Proporção de Acidentes com Feridos Leves e Graves

```
dados %>%
  summarise(
    proporcao_feridos_leves = sum(feridos_leves, na.rm = TRUE) / n(),
    proporcao_feridos_graves = sum(feridos_graves, na.rm = TRUE) / n()
  ) %>%
  pivot_longer(cols = c(proporcao_feridos_leves, proporcao_feridos_graves),
               names_to = "tipo_ferido",
               values_to = "proporcao") %>%
  ggplot(aes(x = tipo_ferido, y = proporcao, fill = tipo_ferido)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(
    title = "Proporção de Feridos Leves e Graves nos Acidentes",
    x = "Tipo de Ferido",
    y = "Proporção"
  )
```



Número de Acidentes por Dia da Semana

```
ggplot(dados %>%
  group_by(dia_semana) %>%
  summarise(total_acidentes = n()) %>%
  arrange(desc(total_acidentes)), aes(x = reorder(dia_semana, -total_acidentes), y = total_acidentes)
) +
  geom_bar(stat = "identity", fill = "green") +
  theme_minimal() +
  labs(
    title = "Número de Acidentes por Dia da Semana",
    x = "Dia da Semana",
    y = "Total de Acidentes"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Análise de Acidentes por Causa e Condição Meteorológica

```
dados_resumo <- dados %>%
  group_by(condicao_meteorologica) %>%
  summarise(
    total_mortos = sum(mortos, na.rm = TRUE),
    total_feridos = sum(feridos, na.rm = TRUE),
    total_feridos_graves = sum(feridos_graves, na.rm = TRUE),
    total_feridos_leves = sum(feridos_leves, na.rm = TRUE)
  ) %>%
  # Calcular o total de acidentes em cada condição meteorológica
  mutate(total_acidentes = total_mortos + total_feridos + total_feridos_graves + total_feridos_leves)

# Calcular o total geral de acidentes
total_acidentes_geral <- sum(dados_resumo$total_acidentes, na.rm = TRUE)

# Calcular a porcentagem de acidentes por condição meteorológica
dados_resumo <- dados_resumo %>%
  mutate(porcentagem = (total_acidentes / total_acidentes_geral) * 100)

# Criar o gráfico com a porcentagem de acidentes por condição meteorológica
ggplot(dados_resumo, aes(x = reorder(condicao_meteorologica, -porcentagem), y = porcentagem, fill = condicao_meteorologica)) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = scales::percent_format(scale = 1)) + # Formatar o eixo Y para porcentagem
  theme_minimal() +
```

```
labs(
  title = "Porcentagem de Acidentes por Condição Meteorológica",
  x = "Condição Meteorológica",
  y = "Porcentagem de Acidentes",
  fill = "Condição Meteorológica"
) +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

