

TPC1

Grupo4

2023-09-23

Nota:

Deve efetuar todos os Save com "Save with encoding UTF-8" de modo a manter palavras acentuadas e caracteres especiais"

Base de dados:condutores.csv

```
# Remover tudo!

# Incluir as libraries de que necessita
library(VIM)

## Warning: package 'VIM' was built under R version 4.2.3

## Loading required package: colorspace

## Loading required package: grid

## VIM is ready to use.

## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues

##
## Attaching package: 'VIM'

### The following object is masked from 'package:datasets':
###   sleep

library(plyr)

## Warning: package 'plyr' was built under R version 4.2.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

library(dplyr)

##
## Attaching package: 'dplyr'

### The following objects are masked from 'package:plyr':
###   arrange, count, desc, failwith, id, mutate, rename, summarise,
###   summarize

## The following objects are masked from 'package:stats':
###   filter, lag

## The following objects are masked from 'package:base':
###   intersect, setdiff, setequal, union

library(lubridate)

## Warning: package 'lubridate' was built under R version 4.2.3

##
## Attaching package: 'lubridate'

### The following objects are masked from 'package:base':
###   date, intersect, setdiff, union

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## Warning: package 'readr' was built under R version 4.2.3

## Warning: package 'purrr' was built under R version 4.2.3

## Warning: package 'stringr' was built under R version 4.2.3

## Warning: package 'forcats' was built under R version 4.2.3

## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ forcats 1.0.0   ✓ stringr 1.5.0
## ✓ purrr  1.0.1   ✓ tibble  3.1.8
## ✓ readr  2.1.4   ✓ tidyr  1.3.0

## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::arrange() masks plyr::arrange()
## ✖ purrr::compact() masks plyr::compact()
## ✖ dplyr::count() masks plyr::count()
## ✖ dplyr::desc() masks plyr::desc()
## ✖ dplyr::failwith() masks plyr::failwith()
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::id() masks plyr::id()
## ✖ dplyr::lag() masks stats::lag()
## ✖ dplyr::mutate() masks plyr::mutate()
## ✖ dplyr::rename() masks plyr::rename()
## ✖ dplyr::summarise() masks plyr::summarise()
## ✖ dplyr::summarize() masks plyr::summarize()
## ! Use the [ ]8;http://conflicted.r-lib.org/ to force all conflicts to become errors

library(psych)

## Warning: package 'psych' was built under R version 4.2.3

##
## Attaching package: 'psych'
##
## The following objects are masked from 'package:ggplot2':
##   %%, alpha

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.2.3

## corrplot 0.92 loaded

library(ggplot2)
library(ggfortify)

## Warning: package 'ggfortify' was built under R version 4.2.3

library(lsr)# Cramer's V e Eta
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.2.3

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3W8a
```

Questão 1 [5 valores]

Leitura dos dados condutores.csv.

```
Data <-read.csv("condutores.csv", header=TRUE, stringsAsFactors =T,sep=";",dec=".",check.names=F,na.strings=c("N
A","NÃO DEFINIDO"),fileEncoding = "UTF-8")
```

Averigue a existência de valores omissos e, caso existam, identifique as respetivas variáveis. Realize a imputação dos valores omissos da variável "Tempo.Condução.Continuada" considerando as variáveis Tipo.Veiculo, Tipo.Serviço e Distrito. Faça upload do sumário de Tempo.Condução.Continuada após a imputação.

```
summary(is.na(Data))

##   Id. Acidente      Datahora      Sexo      Lesões a 30 dias
##   Mode :logical      Mode :logical      Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209      FALSE:39514      FALSE:48209
##   TRUE :695
##   Licença Condução Teste Alcool      Ações Condutores
##   Mode :logical      Mode :logical      Mode :logical
##   FALSE:39285      FALSE:48172      FALSE:39764
##   TRUE :924      TRUE :37      TRUE :445
##   Inf. Comp. a Ações e Manobras Nomeoutrosfactores Tempo Condução Continuada
##   Mode :logical      Mode :logical      Mode :logical      Mode :logical
##   TRUE :48699      FALSE:48209      FALSE:39872
##   TRUE :110      TRUE :1137
##   Acessórios Condutores Categoria Veiculos Tipo Veiculo      Tipo Serviço
##   Mode :logical      Mode :logical      Mode :logical      Mode :logical
##   FALSE:39699      FALSE:48155      FALSE:29756      FALSE:48186
##   TRUE :119      TRUE :54      TRUE :18453      TRUE :183
##   Veiculo Especial Ano matricula      Inspeção Periódica Certificado Adr
##   Mode :logical      Mode :logical      Mode :logical      Mode :logical
##   FALSE:232      FALSE:38865      FALSE:48014      FALSE:19
##   TRUE :39977      TRUE :2144      TRUE :195      TRUE :48198
##   Carga Lotação      Pneus      Seguros      Distrito
##   Mode :logical      Mode :logical      Mode :logical      Mode :logical
##   FALSE:39719      FALSE:39773      FALSE:48116      FALSE:48209
##   TRUE :498      TRUE :436      TRUE :93
##   Concelho      Conductor Gr.Etario(<=5) SUM Conductor Gr.Etario(6-9) SUM
##   Mode :logical      Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(10-14) SUM Conductor Gr.Etario(15-17) SUM
##   Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(18-20) SUM Conductor Gr.Etario(21-24) SUM
##   Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(25-29) SUM Conductor Gr.Etario(30-34) SUM
##   Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(35-39) SUM Conductor Gr.Etario(40-44) SUM
##   Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(45-49) SUM Conductor Gr.Etario(50-54) SUM
##   Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(55-59) SUM Conductor Gr.Etario(65-69) SUM
##   Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(70-74) SUM Conductor Gr.Etario(>=75) SUM
##   Mode :logical      Mode :logical
##   FALSE:48209      FALSE:48209
##
##   Conductor Gr.Etario(Não Def.) SUM
##   Mode :logical
##   FALSE:48209
##

Data1<-Data[,c("Tempo Condução Continuada", "Tipo Veiculo", "Tipo Serviço", "Distrito")]

Data1<-KNM(Data1, variable = "Tempo Condução Continuada", k = 200)

summary(is.na(Data1))

##   Tempo Condução Continuada Tipo Veiculo      Tipo Serviço      Distrito
##   Mode :logical      Mode :logical      Mode :logical      Mode :logical
##   FALSE:48209      FALSE:29756      FALSE:48186      FALSE:48209
##   TRUE :18453      TRUE :183
##   Tempo Condução Continuada_imp
##   Mode :logical
##   FALSE:48209
##

Data1<- Data1[, c("Tempo Condução Continuada", "Tipo Veiculo", "Tipo Serviço", "Distrito")]

Data$`Tempo Condução Continuada`<-Data1$`Tempo Condução Continuada`

summary(Data$`Tempo Condução Continuada` )

##   De 1 a 3 horas      De 3 a 5 horas      Ignorada Mais de 5 horas      Menos de 1 hora
##   2338      182      16224      123      21422
```

Questão 2 [5 valores] Efetue a análise descritiva da variável Ano.matricula, de modo a completar:

#Efetue a análise descritiva da variável Ano.matricula, de modo a completar:

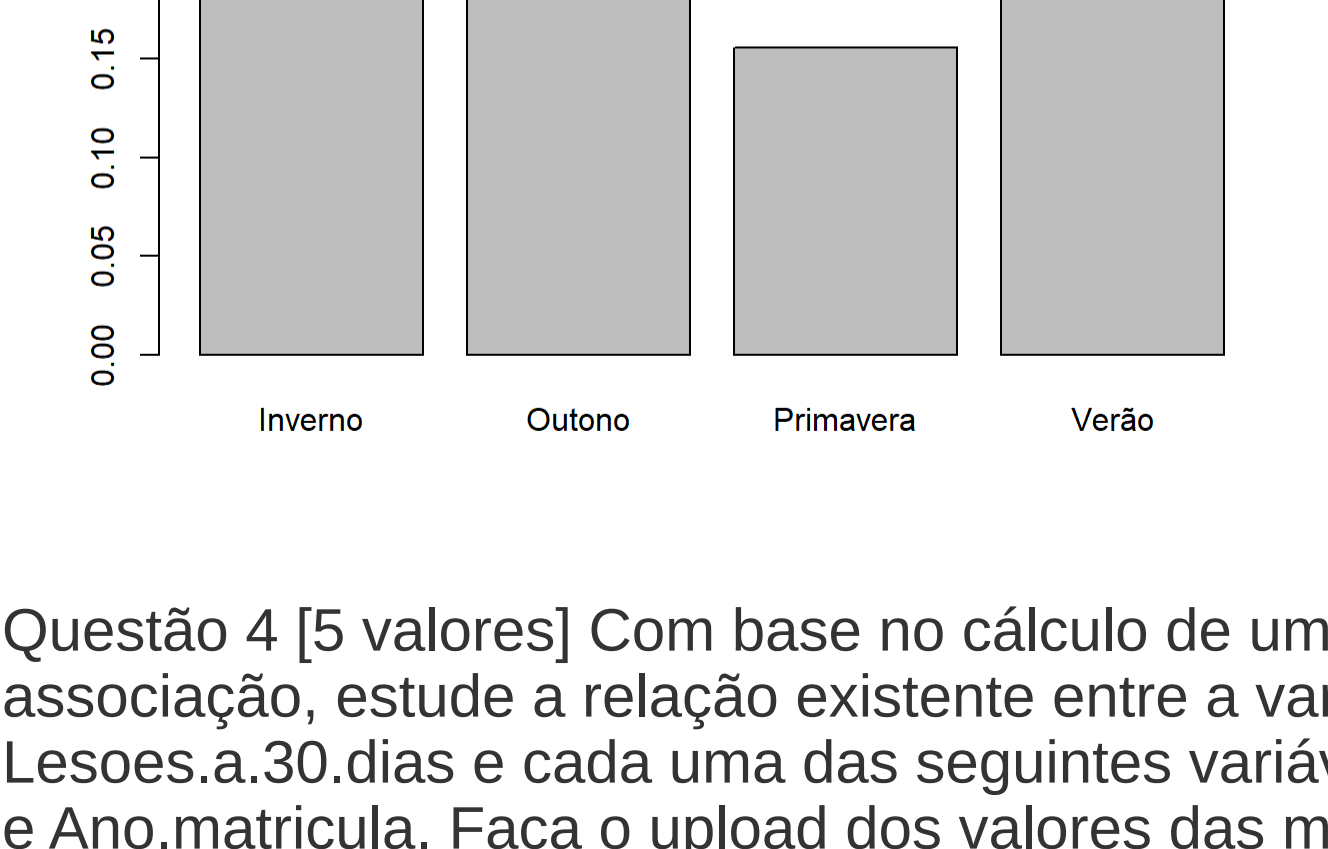
#O veiculo mais antigo é do ano 1914 e a percentagem (cumulativa) de veiculos com ano de matricula inferior ou ig ual a 2087 é 52,5%(percentagem com uma casa decimal); os veiculos com ano de matricula inferior a 1977 (indique o ano) podem ser considerados outliers. Considerando o ano de 2020 para o cálculo da idade dos veiculos, observa-s e que 25% dos veiculos têm menos de 7 anos.

Questão 3 [5 valores] Crie a variável nominal Estação_Ano com as classes “Inverno”, “Primavera”, “Verão” e “Outono” e faça upload do correspondente gráfico de barras apresentando as frequências relativas.

```
Data$Estacao_ano<- rep(NA,48209)
Data$Datahora<- ymd_hms(Data$Datahora)
Data$Estacao_ano[which(Data$Datahora > "2020-3-22" & Data$Datahora <= "2020-6-20" )]<- "Primavera"
Data$Estacao_ano[which(Data$Datahora > "2020-6-20" & Data$Datahora <= "2020-9-22" )]<- "Verão"
Data$Estacao_ano[which(Data$Datahora > "2020-9-22" & Data$Datahora <= "2020-12-21" )]<- "Outono"
Data$Estacao_ano[which(Data$Datahora > "2020-12-21" & Data$Datahora <= "2021-1-1" )]<- "Inverno"
Data$Estacao_ano[which(Data$Datahora >= "2020-1-1" & Data$Datahora <= "2020-3-22" )]<- "Inverno"

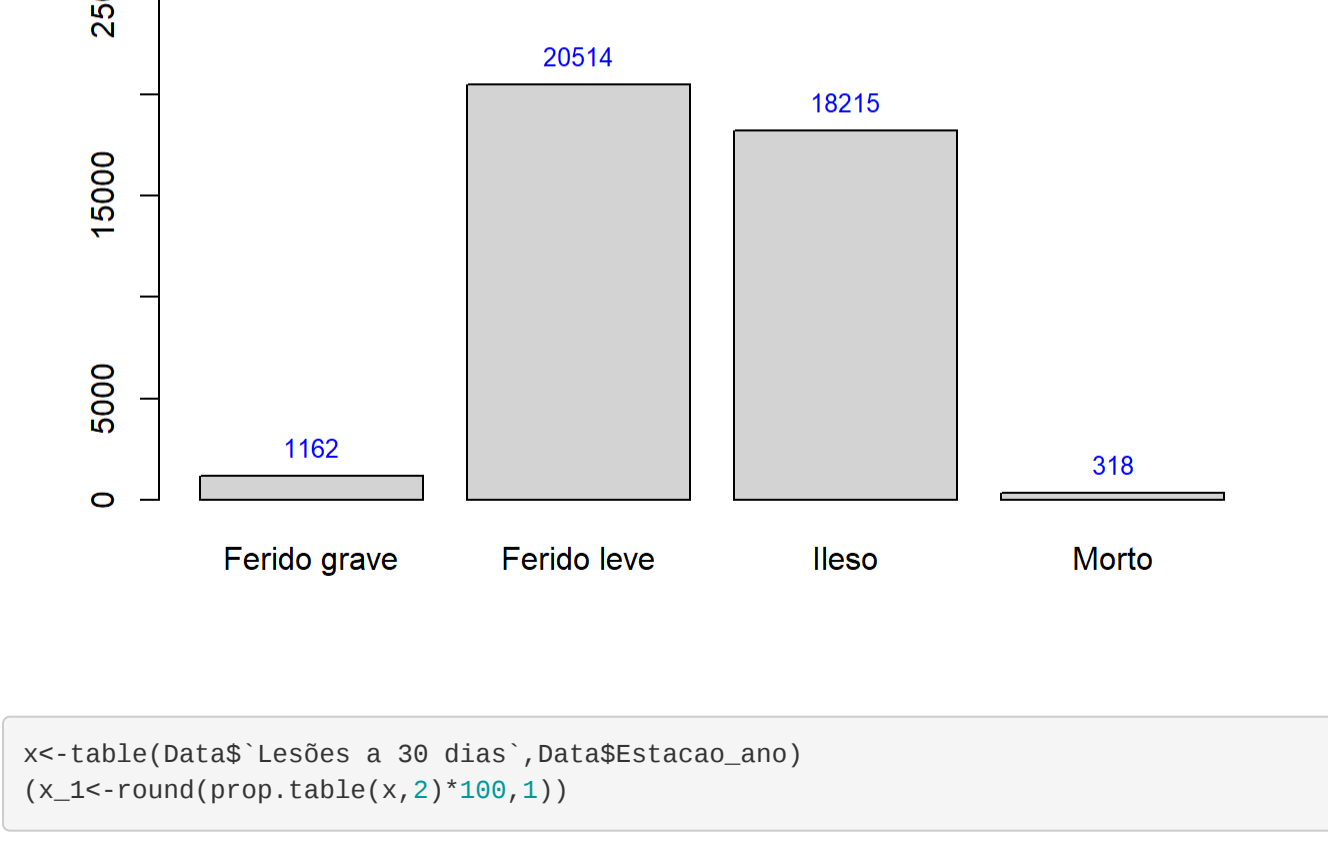
freq.rel<-table(Data$Estacao_ano)/length(Data$Estacao_ano)

barplot(freq.rel)
```



Questão 4 [5 valores] Com base no cálculo de uma medida de associação, estude a relação existente entre a variável Lesões.a.30.dias e cada uma das seguintes variáveis: Estacao_ano e Ano.matricula. Faça o upload dos valores das medidas de associação e dos gráficos que ilustram as mesmas associações.

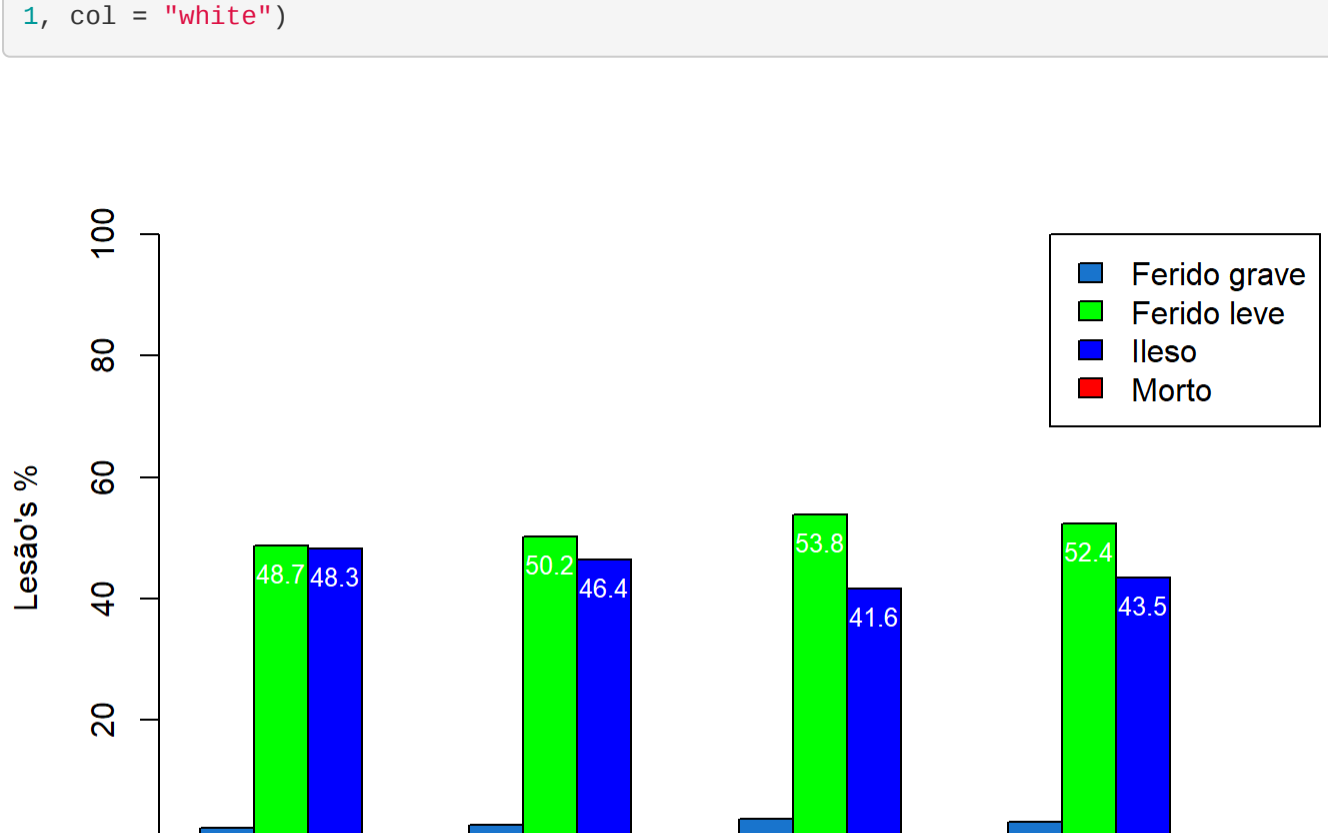
```
coords <- table(Data$`Lesões a 30 dias`,Data$Estacao_ano)
p<-barplot(coords, ylim=c(0,3000),main = "Frequencia",col="lightgrey")
text(x = p, y = coords, label = coords, cex = 0.8,pos = 3, col = "blue")
```



```
x<-table(Data$`Lesões a 30 dias`,Data$Estacao_ano)
(x_1<-round(prop.table(x,2)*100,1))

##
##
##   Ferido grave      Inverno      Outono      Primavera      Verão
##   48.7      2.3      50.2      53.8      41.6
##   Ileso      48.3      46.4      41.6      43.5
##   Morto      0.6      0.8      0.9      0.9
```

```
coords <- x_1
p<-barplot(x_1,beside=TRUE,legend.text = TRUE, ylim=c(0,100), xlab="Estação",ylab="Lesão's %", col= c("dodgerblue3","green", "blue", "red"),args.legend = list(x = "to
pright",inset =
c(- 0.05, 0)))
text(x = p, y = coords, label = coords,cex = 0.8, pos =
1, col = "white")
```



```
cramersV(Data$`Lesões a 30 dias`,Data$Estacao_ano)

## [1] 0.8317989

ggplot(Data, aes(x=Data$`Lesões a 30 dias`, y=Data$`Ano matricula`)) +
  geom_violin(width=1,trim=FALSE)+
  geom_boxplot(width=0.1, color="grey",
  alpha=0.2)

## Warning: Use of ``Data$`Lesões a 30 dias` `` is discouraged.
## i Use `Lesões a 30 dias` instead.

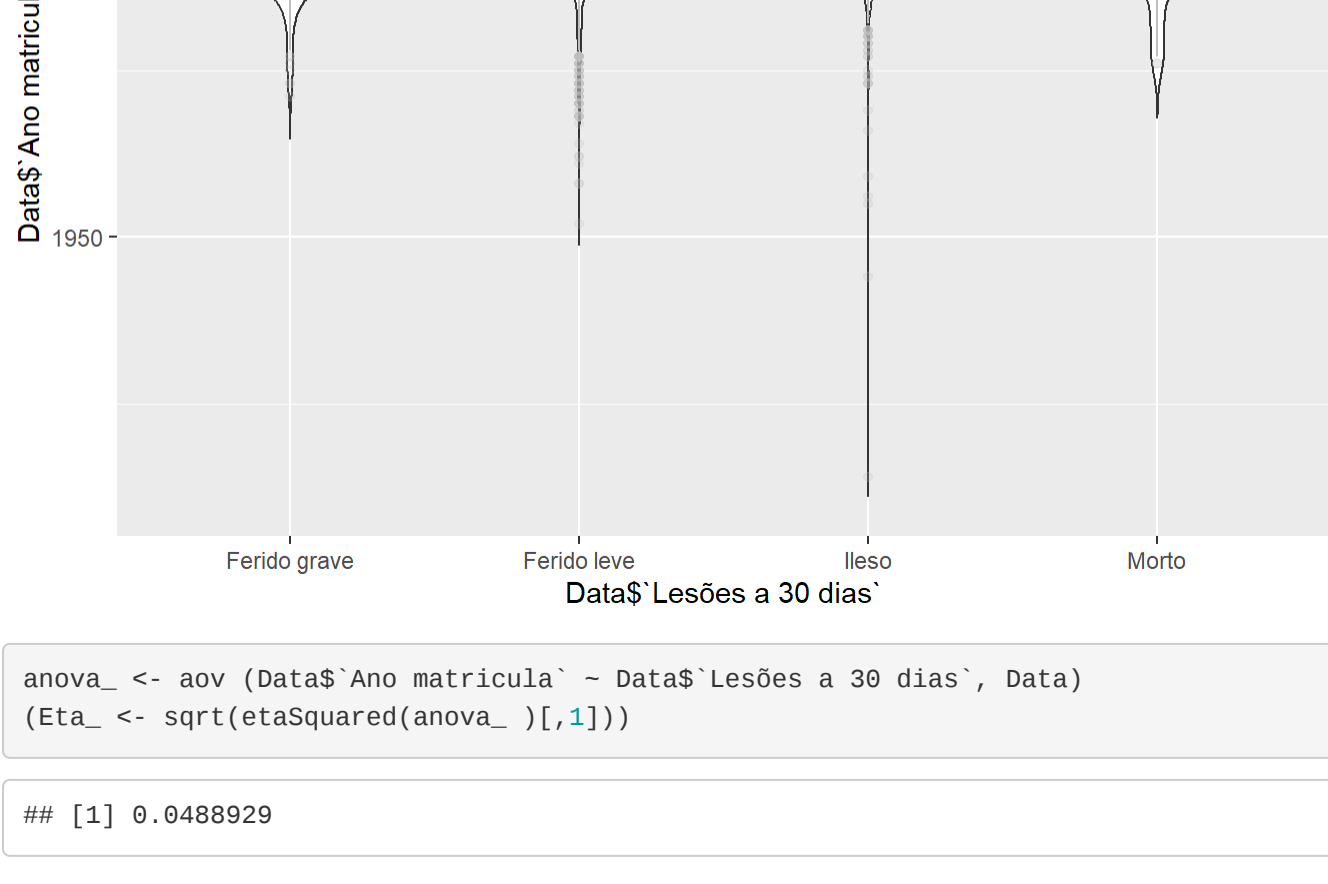
## Warning: Use of ``Data$`Ano matricula` `` is discouraged.
## i Use `Ano matricula` instead.

## Warning: Use of ``Data$`Lesões a 30 dias` `` is discouraged.
## i Use `Lesões a 30 dias` instead.

## Warning: Use of ``Data$`Ano matricula` `` is discouraged.
## i Use `Ano matricula` instead.

## Warning: Removed 2144 rows containing non-finite values ( `stat_ydensity()` ).

## Warning: Removed 2144 rows containing non-finite values ( `stat_boxplot()` ).
```



```
anova_ <- aov (Data$`Ano matricula` ~ Data$`Lesões a 30 dias`, Data)
(Eta_ <- sqrt(etasquared(anova_)[,1]))

## [1] 0.8488929
```

Tarefa final: Submeta, no Moddle, um ficheiro pdf resultado da compilação do TEMPLATE_QUIZ1.

Caso os resultados apresentados não sejam coerentes com as respostas dadas, a classificação será penalizada.