

## PACDII: Quiz III

ID Grupo e nome do representante

2023-10-05

```
# Remover tudo!  
  
# Incluir as libraries de que necessita  
library(tidyverse)  
  
library(conflicted)  
conflicts_prefer(dplyr::filter)  
  
library(here)  
  
library(magrittr)  
library(tidymodels)  
  
library(dplyr)
```

### III.1 [5 valores]

Leia “peões.csv” e retenha apenas as variáveis “Tipo.Natureza”, “Natureza”, “Ações.Peão” e “Lesões.a.30.dias”.

```
read <- read.csv(here("peões.csv"), stringsAsFactors = T, check.names =  
T)  
df <- read %>% select(Tipo.Natureza, Natureza, Ações.Peão,  
Lesões.a.30.dias)  
df %>% Hmisc::describe()  
  
## .  
##  
## 4 Variables      5968 Observations  
## -----  
## -----  
## Tipo.Natureza  
##      n missing distinct  
##  5968      0         3  
##  
## Value      Atropelamento      Colisão      Despiste  
## Frequency      5759          150         59  
## Proportion      0.965      0.025      0.010  
## -----  
## -----  
## Natureza  
##      n missing distinct
```

```
##      5968      0      15
##
## lowest : Atropelamento com fuga
Atropelamento de animais                                Atropelamento de
peões                                                    Colisão choque em cadeia
Colisão com fuga
## highest: Despiste com capotamento
Despiste com colisão com veículo imobil. ou obstáculo Despiste com
dispositivo de retenção                                Despiste sem dispositivo de
retenção                                                Despiste simples
## -----
-----
## Acções.Peão
##      n missing distinct
##      5968      0      13
##
## lowest : A sair ou entrar num veículo
Atravessando em passagem sinalizada
Atravessando em passagem sinalizada com desrespeito da sinalização
semafórica                                              Atravessando fora da passagem de peões a
mais de 50m de uma passagem ou quando não exista passagem Atravessando
fora da passagem de peões, a menos de 50m de uma passagem
## highest: NÃO DEFINIDO
Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo
Transitando pela berma ou passeio
Transitando pela direita da faixa de rodagem
Transitando pela esquerda da faixa de rodagem
## -----
-----
## Lesões.a.30.dias
##      n missing distinct
##      5968      0      4
##
## Value      Ferido grave  Ferido leve      Ileso      Morto
## Frequency          428      5341          4      195
## Proportion      0.072      0.895      0.001      0.033
## -----
-----
```

Descarte todos os casos com dados omissos.

```
dfN <- df %>% filter(Acções.Peão != "NÃO DEFINIDO") %>% drop_na()
dfN$Acções.Peão %<>% fct_drop()
dfN %>% Hmisc::describe()

## .
##
## 4 Variables      5760 Observations
## -----
-----
## Tipo.Natureza
```

```

##          n missing distinct
##      5760          0          3
##
## Value      Atropelamento      Colisão      Despiste
## Frequency          5562          141          57
## Proportion          0.966          0.024          0.010
## -----
##
## Natureza
##          n missing distinct
##      5760          0          15
##
## lowest : Atropelamento com fuga
Atropelamento de animais                      Atropelamento de
peões                      Colisão choque em cadeia
Colisão com fuga
## highest: Despiste com capotamento
Despiste com colisão com veículo imobil. ou obstáculo Despiste com
dispositivo de retenção                      Despiste sem dispositivo de
retenção                      Despiste simples
## -----
##
## Acções.Peão
##          n missing distinct
##      5760          0          12
##
## lowest : A sair ou entrar num veículo
Atravessando em passagem sinalizada
Atravessando em passagem sinalizada com desrespeito da sinalização
semafórica                      Atravessando fora da passagem de peões a
mais de 50m de uma passagem ou quando não exista passagem Atravessando
fora da passagem de peões, a menos de 50m de uma passagem
## highest: Em trabalhos na via
Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo
Transitando pela berma ou passeio
Transitando pela direita da faixa de rodagem
Transitando pela esquerda da faixa de rodagem
## -----
##
## Lesões.a.30.dias
##          n missing distinct
##      5760          0          4
##
## Value      Ferido grave  Ferido leve      Ileso      Morto
## Frequency          414          5153          4          189
## Proportion          0.072          0.895          0.001          0.033
## -----
##

```

Adote dummy coding para as variáveis “Tipo.Natureza” e “Lesões.a.30.dias” (mantendo estas variáveis originais no seu conjunto de dados). No Quiz faça upload de um sumário dos dados “peoes\_d” assim obtidos.

```
dfND <- dfN %>% fastDummies::dummy_cols(select_columns =
c("Tipo.Natureza", "Lesões.a.30.dias"), remove_most_frequent_dummy = T)
peoes_d <- dfND
peoes_d %>% summary()
```

```
##          Tipo.Natureza
## Atropelamento:5562
## Colisão      : 141
## Despiste     :  57
##
##
##
##
##                                     Natureza
## Atropelamento de peões                :5266
## Atropelamento com fuga                : 295
## Colisão com outras situações          :  60
## Despiste com colisão com veículo imobil. ou obstáculo: 34
## Colisão com veículo ou obstáculo na faixa de rodagem : 31
## Colisão lateral com outro veículo em movimento      : 26
## (Other)                                         : 48
##
##
## Acções.Peão
## Atravessando em passagem sinalizada                :2193
## Em plena faixa de rodagem                          : 736
## Atravessando fora da passagem de peões, a menos de 50m de uma passagem : 688
## Atravessando fora da passagem de peões a mais de 50m de uma passagem ou quando não exista passagem: 540
## Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo   : 482
## Transitando pela berma ou passeio                  : 473
## (Other)                                             : 648
##
##
## Lesões.a.30.dias Tipo.Natureza_Colisão Tipo.Natureza_Despiste
## Ferido grave: 414   Min.    :0.000000   Min.    :0.000000
## Ferido leve :5153   1st Qu.:0.000000   1st Qu.:0.000000
## Ileso       :  4    Median :0.000000   Median :0.000000
## Morto       : 189   Mean    :0.02448   Mean    :0.009896
##                                     3rd Qu.:0.000000   3rd Qu.:0.000000
##                                     Max.    :1.000000   Max.    :1.000000
##
##
## Lesões.a.30.dias_Ferido grave Lesões.a.30.dias_Ileso
## Lesões.a.30.dias_Morto
## Min.    :0.000000   Min.    :0.000000   Min.    :0.000000
## 1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000
## Median :0.000000   Median :0.000000   Median :0.000000
## Mean    :0.07187   Mean    :0.0006944   Mean    :0.03281
## 3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000
```

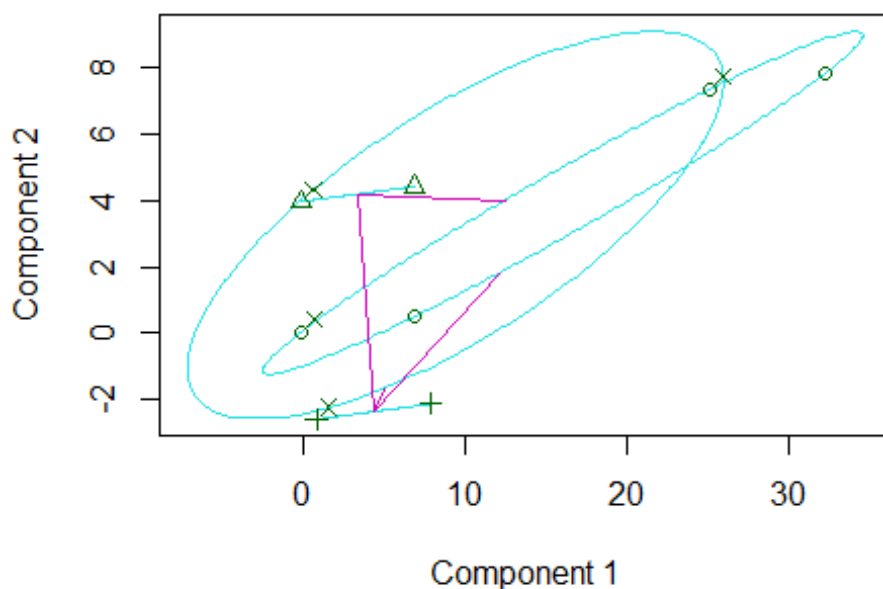
```
## Max. :1.00000 Max. :1.0000000 Max. :1.00000
##
```

III.2) [5 valores] Efectue o agrupamento de “peoes\_d” com o algoritmo PAM, usando a distância Manhattan com base nas variáveis dummy (recodificação das categorias de “Tipo.Natureza” e “Lesões.a.30.dias”). Obtenha 4 clusters.

```
set.seed(1)
peoes_d.pam <- peoes_d %>%
  select(-Tipo.Natureza, -Natureza, -Acções.Peão, -Lesões.a.30.dias) %>%
  cluster::pam(diss = F, k = 4, metric = "manhattan")
peoes_d.pam %>% tidy()

## # A tibble: 4 × 12
##   size max.diss avg.diss diameter separation avg.width cluster
##   <dbl>   <dbl>   <dbl>   <dbl>     <dbl>   <dbl> <fct>
## 1  5033     2 0.00874     2         1      0.987 1
## 2   184     1 0.0217     1         1      0.968 2
## 3   402     1 0.0299     1         1      0.957 3
## 4   141     1 0.128      2         1      0.768 4
## # i 5 more variables: Tipo.Natureza_Colisão <int>,
## #   Tipo.Natureza_Despiste <int>, `Lesões.a.30.dias_Ferido grave`
## #   <int>,
## #   Lesões.a.30.dias_Ileso <int>, Lesões.a.30.dias_Morto <int>
peoes_d.pam %>% plot()
```

`plot(cluster::pam(x = ., k = 4, diss = F, metric = "manl`



These two components explain 43.82 % of the point variab

**Silhouette plot of cluster::pam(x = ., k = 4, di**

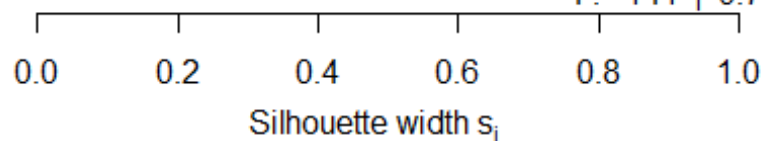
n = 5760

4 clusters  $C_j$

$j : n_j \mid \text{ave}_{i \in C_j} s_i$

1 : 5033 | 0.99

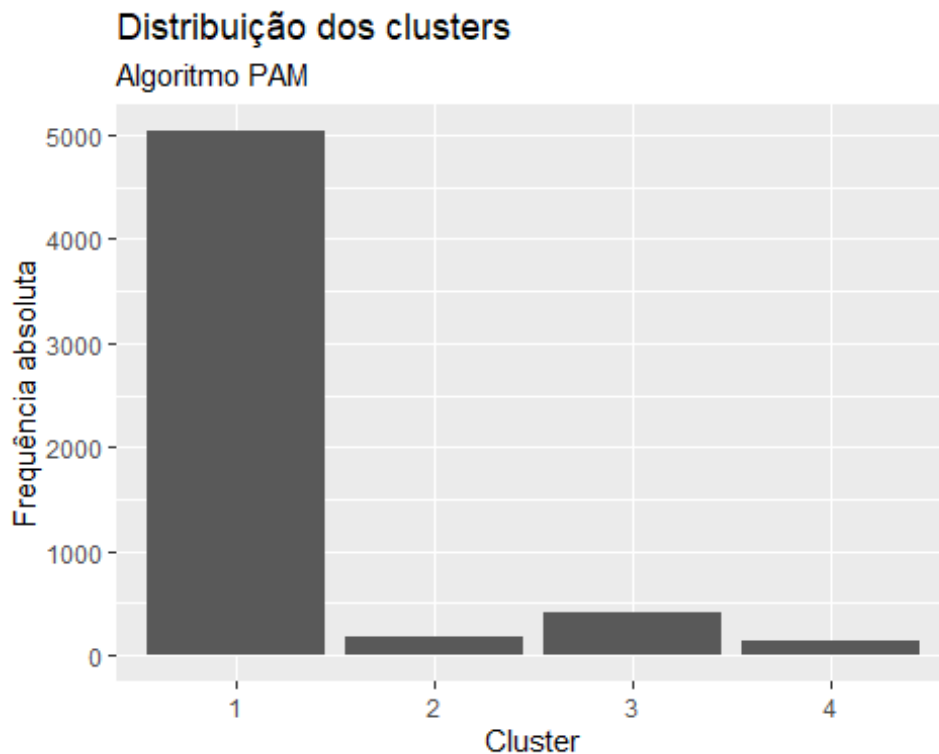
2 : 184 | 0.97  
3 : 402 | 0.96  
4 : 141 | 0.77



Average silhouette width : 0.98

No Quiz faça upload de um gráfico de barras ilustrando a sua distribuição (usando a frequência absoluta)

```
peoes_d.aug <- peoes_d.pam %>% augment(peoes_d)
peoes_d.aug %>%
  ggplot(aes(x = .cluster)) +
  geom_bar() +
  labs(x = "Cluster", y = "Frequência absoluta", title = "Distribuição
dos clusters", subtitle = "Algoritmo PAM")
```



**III.3) [5 valores] Caracterize os clusters obtidos usando as variáveis base relevantes. Para o efeito, calcule medidas de associação. No Quiz faça upload de gráficos de barras ilustrando as associações de interesse.**

*# simulate.p.value=TRUE para não aparecer o warning the aproximação de chi-squared*

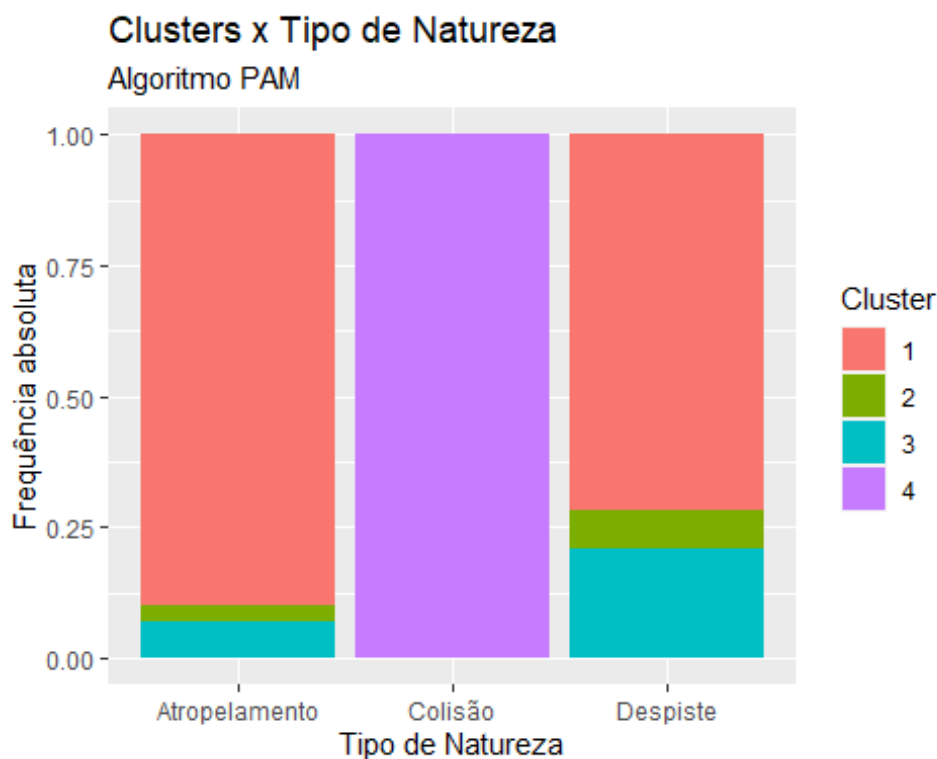
```
vc1 <- lsr::cramersV(peoes_d.aug$Tipo.Natureza, peoes_d.aug$.cluster,
simulate.p.value=TRUE)
vc2 <- lsr::cramersV(peoes_d.aug$Natureza, peoes_d.aug$.cluster,
simulate.p.value=TRUE)
vc3 <- lsr::cramersV(peoes_d.aug$Acções.Peão, peoes_d.aug$.cluster,
simulate.p.value=TRUE) ## não é significativo
vc4 <- lsr::cramersV(peoes_d.aug$Lesões.a.30.dias, peoes_d.aug$.cluster,
simulate.p.value=TRUE)
(cramersV_vct <- c("Tipo Natureza"=vc1,"Natureza"=vc2,"Acções
Peão"=vc3,"Lesões a 30 dias"=vc4) %>% round(4))
```

##	Tipo Natureza	Natureza	Acções Peão	Lesões a 30 dias
##	0.7084	0.5807	0.1429	0.8055

A única variável que não tem uma associação forte ( $\varphi > 0.4$ ) é a variável de Acções Peão. Por essa razão, mantemos as restantes na caracterização.

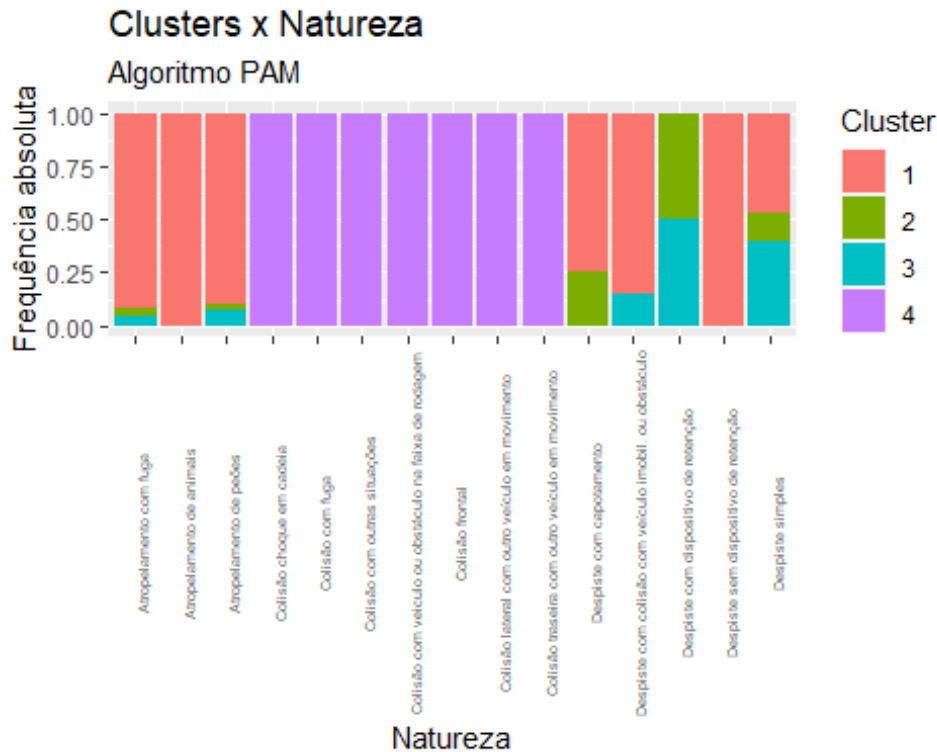
*# bar plot para caracterizar cluster 1*

```
peoes_d.aug %>%
  select(Tipo.Natureza, Natureza, Acções.Peão, Lesões.a.30.dias,
    .cluster) -> peoes_d.aug.noD
peoes_d.aug.noD %>%
  ggplot(aes(x = Tipo.Natureza, fill = .cluster)) +
  geom_bar(position = "fill") +
  labs(x = "Tipo de Natureza", y = "Frequência absoluta", title =
    "Clusters x Tipo de Natureza", subtitle = "Algoritmo PAM") +
  scale_fill_discrete(name = "Cluster", labels = c("1", "2", "3", "4"))
```



```
peoes_d.aug.noD %>%
  ggplot(aes(x = Natureza, fill = .cluster)) +
  geom_bar(position = "fill") +
  labs(x = "Natureza", y = "Frequência absoluta", title = "Clusters x
    Natureza", subtitle = "Algoritmo PAM") +
  scale_fill_discrete(name = "Cluster", labels = c("1", "2", "3", "4")) +
  theme(axis.text.x = element_text(angle = 90, size = 5))
```





```
# peoes_d.aug.noD %>%
#   ggplot(aes(x = Acções.Peão, fill = .cluster)) +
#   geom_bar(position = "fill") +
#   labs(x = "Acções do Peão", y = "Frequência absoluta", title =
# "Clusters x Acções do Peão", subtitle = "Algoritmo PAM") +
#   scale_fill_discrete(name = "Cluster", labels = c("1", "2", "3", "4"))
#   theme(axis.text.x = element_text(angle = 90, size = 5))
peoes_d.aug.noD %>%
  ggplot(aes(x = Lesões.a.30.dias, fill = .cluster)) +
  geom_bar(position = "fill") +
  labs(x = "Lesões a 30 dias", y = "Frequência absoluta", title =
"Clusters x Lesões a 30 dias", subtitle = "Algoritmo PAM") +
  scale_fill_discrete(name = "Cluster", labels = c("1", "2", "3", "4"))
```



- 1 - Peões Feridos leves e ilesos
- 2 - Peões Mortos
- 3 - Peões Feridos graves
- 4 - Colisões com carros

**III.4) [5 valores] Caracterize os clusters obtidos usando as restantes variáveis disponíveis em “peoes\_d”. No Quiz faça upload de um resumo dos perfis dos clusters em que também inclui todas as tabelas cruzadas que ilustram as associações entre os clusters obtidos e as variáveis de interesse (sejam ou não variáveis base de agrupamento).**

```
table(peoes_d.aug.noD$.cluster, peoes_d.aug.noD$Acções.Peão) %>% t() %>%
prop.table(.,2) %>% round(4)*100
```

```
##
##
##      1
## A sair ou entrar num veículo      2.01
## Atravessando em passagem sinalizada      39.92
## Atravessando em passagem sinalizada com desrespeito da sinalização semafórica      1.89
## Atravessando fora da passagem de peões a mais de 50m de uma passagem ou quando não exista passagem      9.36
## Atravessando fora da passagem de peões, a menos de 50m de uma passagem      12.14
## Em ilhéu ou refúgio na via      0.95
## Em plena faixa de rodagem      11.86
## Em trabalhos na via      0.81
## Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo      8.70
## Transitando pela berma ou passeio      7.43
## Transitando pela direita da faixa de rodagem      2.90
```

```
## Transitando pela esquerda da faixa de rodagem 2.03
##
## 2
## A sair ou entrar num veículo 0.54
## Atravessando em passagem sinalizada 18.48
## Atravessando em passagem sinalizada com desrespeito da sinalização semafórica 0.54
## Atravessando fora da passagem de peões a mais de 50m de uma passagem ou quando não exista passagem 9.24
## Atravessando fora da passagem de peões, a menos de 50m de uma passagem 13.04
## Em ilhéu ou refúgio na via 0.00
## Em plena faixa de rodagem 32.07
## Em trabalhos na via 0.54
## Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo 3.80
## Transitando pela berma ou passeio 13.59
## Transitando pela direita da faixa de rodagem 3.80
## Transitando pela esquerda da faixa de rodagem 4.35
##
## 3
## A sair ou entrar num veículo 2.99
## Atravessando em passagem sinalizada 28.36
## Atravessando em passagem sinalizada com desrespeito da sinalização semafórica 3.73
## Atravessando fora da passagem de peões a mais de 50m de uma passagem ou quando não exista passagem 11.69
## Atravessando fora da passagem de peões, a menos de 50m de uma passagem 12.94
## Em ilhéu ou refúgio na via 0.50
## Em plena faixa de rodagem 14.93
## Em trabalhos na via 1.99
## Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo 8.71
## Transitando pela berma ou passeio 7.71
## Transitando pela direita da faixa de rodagem 4.73
## Transitando pela esquerda da faixa de rodagem 1.74
##
## 4
## A sair ou entrar num veículo 4.96
## Atravessando em passagem sinalizada 25.53
## Atravessando em passagem sinalizada com desrespeito da sinalização semafórica 0.00
## Atravessando fora da passagem de peões a mais de 50m de uma passagem ou quando não exista passagem 3.55
## Atravessando fora da passagem de peões, a menos de 50m de uma passagem 0.71
## Em ilhéu ou refúgio na via 7.80
## Em plena faixa de rodagem 14.18
## Em trabalhos na via 4.96
## Surgindo inesperadamente na faixa de rodagem de trás de um obstáculo 1.42
## Transitando pela berma ou passeio 30.50
## Transitando pela direita da faixa de rodagem 4.26
## Transitando pela esquerda da faixa de rodagem 2.13
```

```
table(peoes_d.aug.noD$.cluster, peoes_d.aug.noD$Natureza) %>% t() %>%
prop.table(.,2) %>% round(4)*100
```

```
##
## 1 2 3 4
## Atropelamento com fuga 5.38 5.98 3.23 0.00
## Atropelamento de animais 0.02 0.00 0.00 0.00
## Atropelamento de peões 93.78 91.85 93.78 0.00
## Colisão choque em cadeia 0.00 0.00 0.00 1.42
## Colisão com fuga 0.00 0.00 0.00 1.42
## Colisão com outras situações 0.00 0.00 0.00 42.55
## Colisão com veículo ou obstáculo na faixa de rodagem 0.00 0.00 0.00 21.99
## Colisão frontal 0.00 0.00 0.00 2.13
## Colisão lateral com outro veículo em movimento 0.00 0.00 0.00 18.44
## Colisão traseira com outro veículo em movimento 0.00 0.00 0.00 12.06
## Despiste com capotamento 0.06 0.54 0.00 0.00
## Despiste com colisão com veículo imobil. ou obstáculo 0.58 0.00 1.24 0.00
## Despiste com dispositivo de retenção 0.00 0.54 0.25 0.00
## Despiste sem dispositivo de retenção 0.04 0.00 0.00 0.00
## Despiste simples 0.14 1.09 1.49 0.00
```

```
table(peoes_d.aug.noD$.cluster, peoes_d.aug.noD$Tipo.Natureza) %>% t()
%>% prop.table(.,2) %>% round(4)*100
```

```
##
## 1 2 3 4
```

```
## Atropelamento 99.19 97.83 97.01 0.00
## Colisão 0.00 0.00 0.00 100.00
## Despiste 0.81 2.17 2.99 0.00

table(peoes_d.aug.noD$.cluster, peoes_d.aug.noD$Lesões.a.30.dias) %>% t()
%>% prop.table(.,2) %>% round(4)*100

##
## 1 2 3 4
## Ferido grave 0.00 0.00 100.00 8.51
## Ferido leve 99.94 0.00 0.00 87.23
## Ileso 0.06 0.00 0.00 0.71
## Morto 0.00 100.00 0.00 3.55
```

No cluster 1 predominam atropelamentos, cerca de 99.2%, dos quais se destacam atropelamentos de peões e atropelamentos com fuga. Este cluster, tal como foi caracterizado, refere-se sobretudo a feridos leves e ocorrem principalmente em passagens sinalizadas, a menos de 50m de passagens de peões e em faixas de rodagem.

Tal como no cluster 1, nos clusters 2 e 3 predominam atropelamentos de peões, contudo nestes a originar na totalidade (100%) mortos e feridos graves, respetivamente. Para além disso, em termos de Ações de peão, predominam os mesmos que os mencionados para o cluster 1. Nestes três clusters também são presentes situações de despiste, mas em pequenas proporções.

O cluster 4 diverge dos restantes porque remete para colisões e tem ferimentos de todos os tipos, predominado por feridos leves. Quanto a Ações de peões, para este cluster, predominam ao atravessar passagens sinalizadas, em faixas de rodagem e transição pela berma ou passeio.