Explorando os dados traumas

Felipe

25/02/2022

Carregando bibliotecas para manipulação de dados

## Warning: package 'pastecs' was built under R version 4.1.3

Carregando dados do arquivo “traumas.csv”:

dados\_traumas <- read.csv("traumas.csv", na.strings = "NA", sep = ";")   
names(dados\_traumas) #execute este comando para ver os nomes das colunas/variáveis presentes no banco de dados

## [1] "autor\_ano" "pais"   
## [3] "regiao" "sitio"   
## [5] "long\_leste\_m" "lat\_sul\_m"   
## [7] "elevacao\_m" "cultura"   
## [9] "cronologia" "data\_absBP\_.inicial"   
## [11] "data\_absBP\_final" "sexo"   
## [13] "n\_id" "n\_id\_traumas"   
## [15] "sem\_traumas" "fr\_traumas"   
## [17] "n\_id\_antimortem" "fr\_antimortem"   
## [19] "n\_id\_perimortem" "fr\_perimortem"   
## [21] "n\_tr\_anterior" "nasal"   
## [23] "frontal" "n\_tr\_posterior"   
## [25] "n\_tr\_pariental" "parietal\_e"   
## [27] "pariental\_d" "n\_tr\_lateral"   
## [29] "n\_antimortem\_lateral" "n\_tr\_perimortem\_lateral"  
## [31] "lateral\_e" "lateral\_d"   
## [33] "ev\_fort" "estr\_sociopolitica"   
## [35] "clima" "carc\_geografica"

#View(dados\_traumas) #execute para visualizar todo o conjunto de dados em uma tabela separada  
  
head(dados\_traumas)

## autor\_ano pais regiao sitio  
## 1 Benyon and Siegel 1981 Peru South Peru\_Paracas Santo\_Domingo\_Pampa  
## 2 Standen et al 2020 Chile Arica Arica\_Acha3C4\_Acha3C1  
## 3 Standen et al 2020 Chile Arica Arica\_Morro\_Praia Miller  
## 4 Standen et al 2020 Chile Arica Arica  
## 5 Standen et al 2020 Chile Arica Arica  
## 6 Standen et al 2010 Chile Vale de Azapa AZ\_146  
## long\_leste\_m lat\_sul\_m elevacao\_m cultura cronologia data\_absBP\_.inicial  
## 1 365028.4 8467514 17 NID ARC 8340  
## 2 360360.0 7955939 71 Chinchorro ARC 10000  
## 3 360360.0 7955939 71 Chinchorro ARC 8000  
## 4 360360.0 7955939 71 Chinchorro ARC 6000  
## 5 360360.0 7955939 71 Chinchorro ARC 6000  
## 6 375393.0 7951291 278 NID FF 2740  
## data\_absBP\_final sexo n\_id n\_id\_traumas sem\_traumas fr\_traumas  
## 1 6220 M 2 0 2   
## 2 8000 M 4 2 2 0,50  
## 3 6000 M 12 3 9 0,25  
## 4 4000 M 63 20 43 0,32  
## 5 4000 F 57 11 46 0,19  
## 6 1930 M 3 3 0 1,00  
## n\_id\_antimortem fr\_antimortem n\_id\_perimortem fr\_perimortem n\_tr\_anterior  
## 1 0 0 0  
## 2 2 0,50 0 1  
## 3 2 0,17 1 0,08 3  
## 4 18 0,29 2 0,03 NA  
## 5 11 0,19 0 NA  
## 6 0 3 1,00 3  
## nasal frontal n\_tr\_posterior n\_tr\_pariental parietal\_e pariental\_d  
## 1 0 0 0 0 0 0  
## 2 NA NA 0 NA NA NA  
## 3 NA NA 0 NA NA NA  
## 4 NA NA NA NA NA NA  
## 5 NA NA NA NA NA NA  
## 6 NA NA 0 NA NA NA  
## n\_tr\_lateral n\_antimortem\_lateral n\_tr\_perimortem\_lateral lateral\_e lateral\_d  
## 1 0 NA NA NA NA  
## 2 1 NA NA NA NA  
## 3 0 NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 3 NA NA NA NA  
## ev\_fort estr\_sociopolitica clima carc\_geografica  
## 1 0 deserto costa  
## 2 0 Tribo deserto costa  
## 3 0 Tribo deserto costa  
## 4 0 Tribo deserto costa  
## 5 0 Tribo deserto costa  
## 6 0 Tribo deserto vale

Filtrando dados do Peru que contenha somente os 3 seguintes períodos: Períoodo Intermédio Primeiro (PIP), Horizonte Médio (HM), e Período Intermédio Tardio (PIT):

# usando o operador pipe " %>% " da bilioteca tydiverse para manipular os dados:  
  
data.tr.peru <- dados\_traumas %>% filter(pais%in%c("Peru"))%>% slice(-1) #a função slice é usada para remover a primeira linha que continha um dado do Arcaico  
   
#View(data.tr.peru)  
#names(data.tr.peru)  
head(data.tr.peru)

## autor\_ano pais regiao sitio long\_leste\_m  
## 1 Tung 2007 e Tung 2012 Peru Ayacucho Conchopata 579248.9  
## 2 Tung 2007 e Tung 2012 Peru Ayacucho Conchopata 579248.9  
## 3 Tung 2007 e Tung 2012 Peru Ayacucho Conchopata 579248.9  
## 4 Tung 2007, Tung 2012, Tung 2008 Peru Ayacucho Conchopata 579248.9  
## 5 Tung 2007 e Tung 2012 Peru Ayacucho Conchopata 579248.9  
## 6 Tung 2007 e Tung 2012 Peru Ayacucho Conchopata 579248.9  
## lat\_sul\_m elevacao\_m cultura cronologia data\_absBP\_.inicial  
## 1 8545412 2839 Pré\_Wari\_Huarpa PIP 1700  
## 2 8545412 2839 Pré\_Wari\_Huarpa PIP 1700  
## 3 8545412 2839 Pré\_Wari\_Huarpa PIP 1700  
## 4 8545412 2839 Wari HM 1300  
## 5 8545412 2839 Wari HM 1300  
## 6 8545412 2839 Wari HM 1300  
## data\_absBP\_final sexo n\_id n\_id\_traumas sem\_traumas fr\_traumas  
## 1 1350 M 1 1 0 1,00  
## 2 1350 F 4 2 2 0,50  
## 3 1350 NID 6 0 6 0,00  
## 4 1150 M 14 4 10 0,29  
## 5 1150 F 25 6 19 0,24  
## 6 1150 NID 5 0 5 0,00  
## n\_id\_antimortem fr\_antimortem n\_id\_perimortem fr\_perimortem n\_tr\_anterior  
## 1 1 1,00 0 0  
## 2 2 0,50 0 0  
## 3 0 0,00 0 0  
## 4 4 0,29 0 2  
## 5 6 0,24 0 0  
## 6 0 0 0  
## nasal frontal n\_tr\_posterior n\_tr\_pariental parietal\_e pariental\_d  
## 1 0 0 0 0 0 0  
## 2 0 0 0 0 0 0  
## 3 0 0 0 0 0 0  
## 4 NA NA 4 NA NA NA  
## 5 NA NA 7 NA NA NA  
## 6 0 0 0 0 0 0  
## n\_tr\_lateral n\_antimortem\_lateral n\_tr\_perimortem\_lateral lateral\_e lateral\_d  
## 1 1 NA NA NA NA  
## 2 0 NA NA NA NA  
## 3 0 NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 0 NA NA NA NA  
## ev\_fort estr\_sociopolitica clima carc\_geografica  
## 1 0 frio\_de\_montanha vale  
## 2 0 frio\_de\_montanha vale  
## 3 0 frio\_de\_montanha vale  
## 4 0 Estado frio\_de\_montanha vale  
## 5 0 Estado frio\_de\_montanha vale  
## 6 0 Estado frio\_de\_montanha vale

## *Descrição estatística dos dados:*

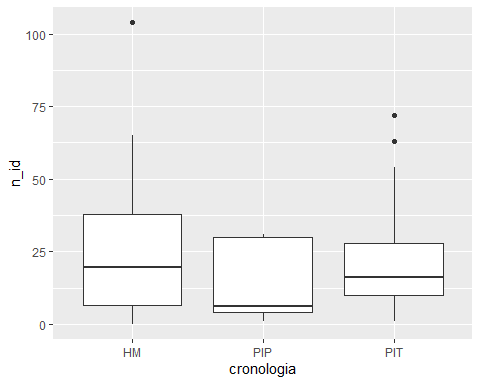
Realizando descrição estatística dos dados númericos de interesse do estudo pela biblioteca “pastecs”

# A função summary também pode ser utilizada  
# n\_id = número de individuos,   
#n\_id\_traumas = número de indivíduos com traumas,   
#sem\_traumas = número de indivíduos sem traumas,   
#n\_id\_antimortem = número de indivíduos com traumas antimortem,   
#n\_id\_perimortem = número de indivíduos com traumas perimortem  
  
data\_descr <- data.tr.peru %>% select(n\_id,n\_id\_traumas,sem\_traumas,n\_id\_antimortem,n\_id\_perimortem)  
  
descr <- stat.desc(data\_descr)  
round(descr,2)

## n\_id n\_id\_traumas sem\_traumas n\_id\_antimortem n\_id\_perimortem  
## nbr.val 36.00 36.00 37.00 26.00 27.00  
## nbr.null 1.00 5.00 5.00 6.00 13.00  
## nbr.na 1.00 1.00 0.00 11.00 10.00  
## min 0.00 0.00 0.00 0.00 0.00  
## max 104.00 52.00 72.00 48.00 9.00  
## range 104.00 52.00 72.00 48.00 9.00  
## sum 853.00 387.00 466.00 212.00 53.00  
## median 15.50 6.00 7.00 4.00 1.00  
## mean 23.69 10.75 12.59 8.15 1.96  
## SE.mean 3.95 2.21 2.30 2.26 0.51  
## CI.mean.0.95 8.03 4.49 4.66 4.65 1.04  
## var 562.62 176.42 195.41 132.46 6.96  
## std.dev 23.72 13.28 13.98 11.51 2.64  
## coef.var 1.00 1.24 1.11 1.41 1.34

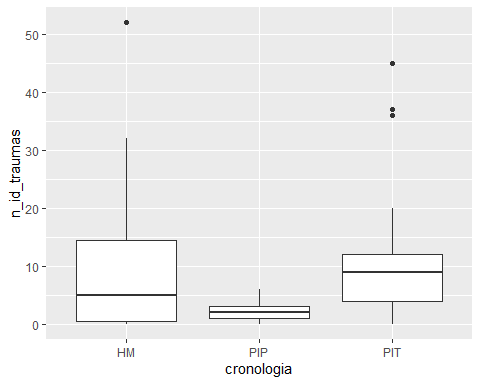
Visualizando distribuição plotando boxplots Por período:

#plotando número de indivíduos por período:  
library(ggplot2)  
data.tr.peru %>% filter(!is.na(n\_id))%>% #retirando valores "NA"   
 group\_by(cronologia) %>% ggplot+geom\_boxplot(aes(x=cronologia,y=n\_id))

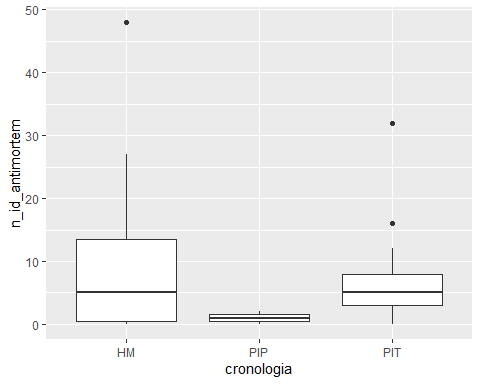


#falta alterar títulos e aparência

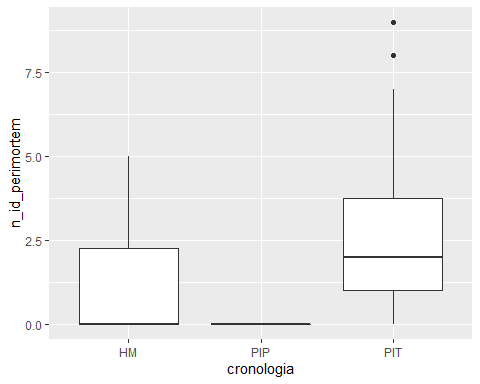
#plotanto número de indivíduos com traumas:  
data.tr.peru %>% filter(!is.na(n\_id\_traumas))%>%  
 group\_by(cronologia) %>% ggplot+geom\_boxplot(aes(x=cronologia,y=n\_id\_traumas))



#plotando número de indivíduos com traumas anitmortem  
  
data.tr.peru %>% filter(!is.na(n\_id\_antimortem))%>%  
 group\_by(cronologia) %>% ggplot+geom\_boxplot(aes(x=cronologia,y=n\_id\_antimortem))

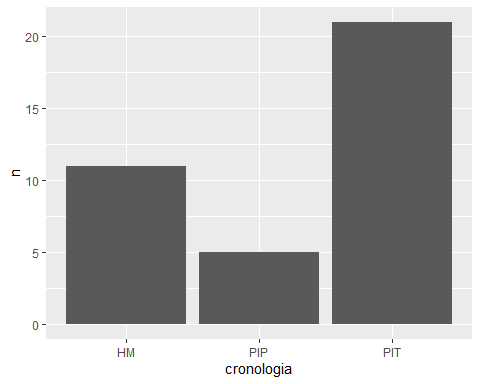


#plotando número de indivíduos com traumas perimortem  
  
data.tr.peru %>% filter(!is.na(n\_id\_perimortem))%>%  
 group\_by(cronologia) %>% ggplot+geom\_boxplot(aes(x=cronologia,y=n\_id\_perimortem))



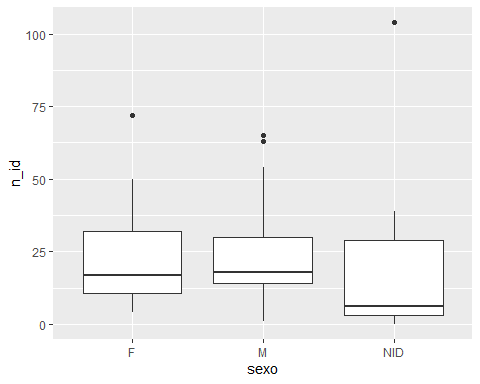
Plotando gráfico de barras com a contagem de amostras por período

data.tr.peru %>% count(cronologia) %>% top\_n(10,n) %>% ggplot()+geom\_col(aes(x=cronologia,y=n))

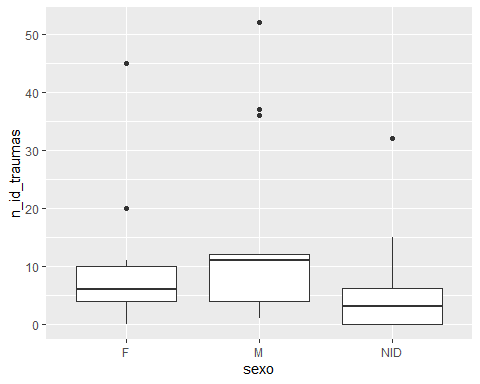


Por sexo:

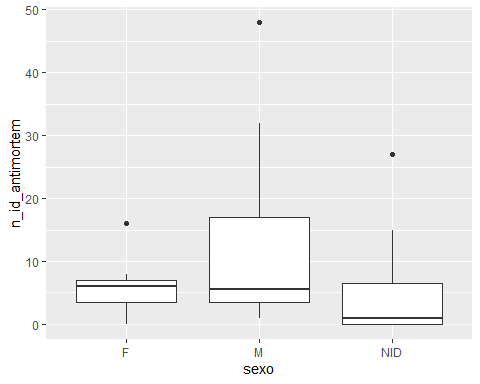
# F = feminino  
# M = masculino  
# NID = não identificado  
#número de indivíduos  
  
data.tr.peru %>% filter(!is.na(n\_id))%>%  
 group\_by(sexo) %>% ggplot+geom\_boxplot(aes(x=sexo,y=n\_id))



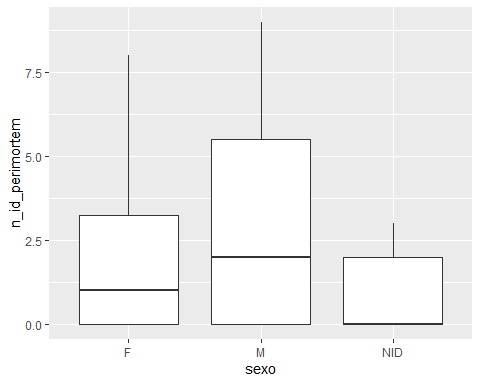
#número de indivíduos com traumas  
data.tr.peru %>% filter(!is.na(n\_id\_traumas))%>%  
 group\_by(sexo) %>% ggplot+geom\_boxplot(aes(x=sexo,y=n\_id\_traumas))



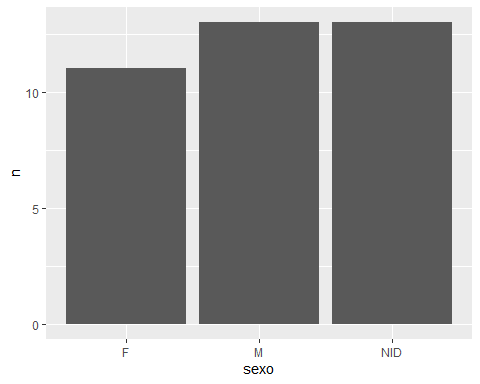
#número de indivíduos com traumas antimortem  
data.tr.peru %>% filter(!is.na(n\_id\_antimortem))%>%  
 group\_by(sexo) %>% ggplot+geom\_boxplot(aes(x=sexo,y=n\_id\_antimortem))



#número de indivíduos com traumas perimortem  
data.tr.peru %>% filter(!is.na(n\_id\_perimortem))%>%  
 group\_by(sexo) %>% ggplot+geom\_boxplot(aes(x=sexo,y=n\_id\_perimortem))

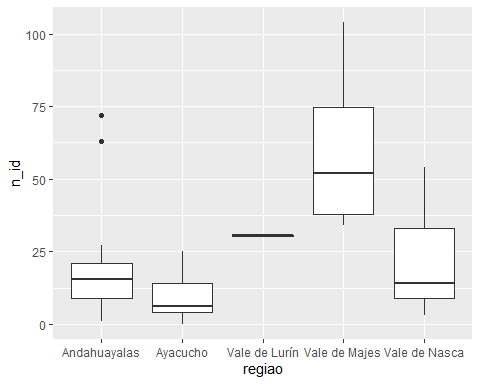


data.tr.peru %>% count(sexo) %>% top\_n(20,n) %>% ggplot()+geom\_col(aes(x=sexo,y=n))

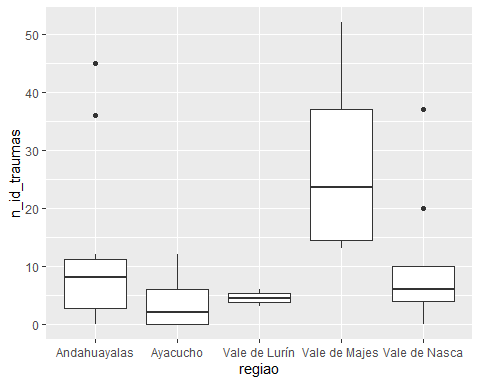


Por região:

# númeoro de indivíduos  
data.tr.peru %>% filter(!is.na(n\_id))%>% ggplot+geom\_boxplot(aes(x=regiao,y=n\_id))

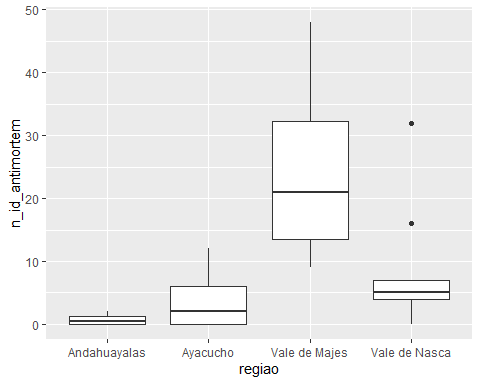


#número de indivíduos com traumas  
data.tr.peru %>% filter(!is.na(n\_id\_traumas))%>% ggplot+geom\_boxplot(aes(x=regiao,y=n\_id\_traumas))

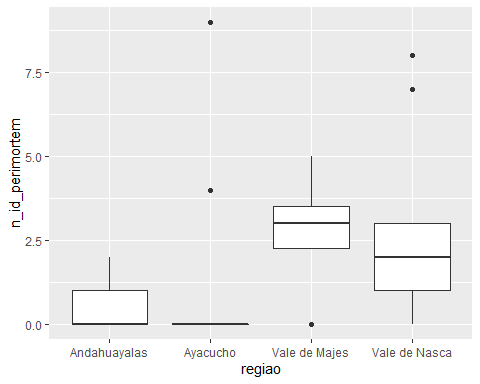


# os resultados mostram que o vale de majes concentra mais casos de traumas, no entanto, só pussui 4 amostra sendo desproporcional em relação ao restante

#número de indivíduos com traumas antimortem  
data.tr.peru %>% filter(!is.na(n\_id\_antimortem))%>% ggplot+geom\_boxplot(aes(x=regiao,y=n\_id\_antimortem))



#número de indivíduos com traumas perimortem  
data.tr.peru %>% filter(!is.na(n\_id\_perimortem))%>% ggplot+geom\_boxplot(aes(x=regiao,y=n\_id\_perimortem))



Testando a associação entre as variáveis pelo teste Chiquadrado. Os dados foram previamente preparados em forma de tabelas de contigência para permitir a aplicação do teste:

obs: as tabelas referentes a somente a região e o período ainda não foram analisadas

#carregando dados em tabela de contigencia para aplicar o chiquadrado  
  
#por região  
regiao <- read.csv("regiao\_contigent.csv", sep = ";",row.names = 1) %>% slice(-3) #usando a função slice para retirar as linhas 3 correspondentes a região do vale de Lurín que contém dados faltantes   
View(regiao)  
#por região e sexo:

#por período  
periodo <- read.csv("periodo\_contigent.csv", sep = ";")  
View(periodo)  
  
#por periodo e sexo:

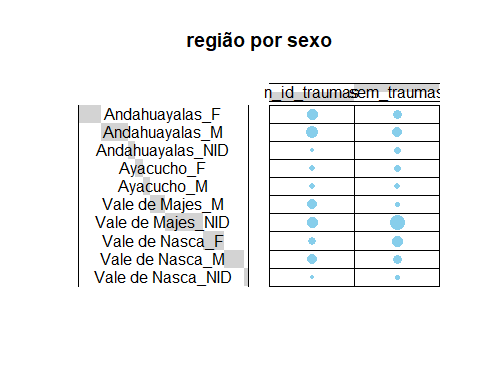
#analisando dados de região por sexo:  
  
regiao\_sexo <- read.csv("regiao\_sexo\_contigent.csv", sep = ";", row.names = 1) %>% slice(-6,-7,-8) # aplicando a função slice para retirar as linhas 7 e 8 correspondentes a região do vale de Lurin que nC#o possui dados pareados por sexo  
#View(regiao\_sexo)  
  
  
library("gplots") #usando o pacote de visualizaC'C#o dos dados matricial

## Warning: package 'gplots' was built under R version 4.1.3

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

#selecionando apenas as colunas de número de amostras, traumas e sem traumas (há ausência de dados nas variáveis antimortem e perimortem) e convertendo dados em matriz:  
  
regiao\_sexo\_at <- regiao\_sexo %>% select(n\_id\_traumas,sem\_traumas)  
  
rg.sexo.matriz <- as.table(as.matrix(regiao\_sexo\_at)) #transformando os dados em tipo matriz  
  
#plotando gráfico balloonplot   
  
balloonplot(t(rg.sexo.matriz), main ="região por sexo", xlab ="", ylab="",label = FALSE, show.margins = FALSE)



#aplicando teste chiquadrado para independencia das amostras  
chisq <- chisq.test(rg.sexo.matriz)  
chisq

##   
## Pearson's Chi-squared test  
##   
## data: rg.sexo.matriz  
## X-squared = 68.936, df = 9, p-value = 2.461e-11

*O resultado aponta que existe dependência entre as amostras, resta saber em quais pares de amostras especificamente*

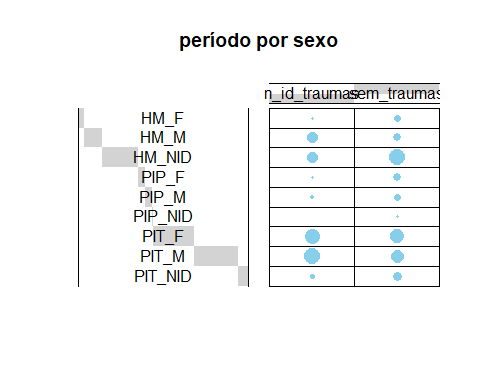
periodo\_sexo <-read.csv("periodo\_sexo\_contigent.csv", sep = ";", row.names = 1)#row.names = 1) %>% select(n\_id,n\_id\_traumas,sem\_traumas) %>% slice(-6)  
names(periodo\_sexo)

## [1] "n\_id" "n\_id\_traumas" "sem\_traumas" "n\_id\_antimortem"  
## [5] "n\_id\_perimortem"

#View(periodo\_sexo)  
head(periodo\_sexo)

## n\_id n\_id\_traumas sem\_traumas n\_id\_antimortem n\_id\_perimortem  
## HM\_F 30 6 24 6 0  
## HM\_M 90 58 32 54 5  
## HM\_NID 182 60 122 51 6  
## PIP\_F 35 5 30 2 0  
## PIP\_M 31 7 24 1 0  
## PIP\_NID 6 0 6 0 0

periodo\_sexo\_at <- periodo\_sexo %>% select(n\_id\_traumas,sem\_traumas) #selecionando apenas as colunas que com valores númericos pareados  
  
pr.sexo.matriz <- as.table(as.matrix(periodo\_sexo\_at))  
  
#plotando gráfico balloonplot   
  
balloonplot(t(pr.sexo.matriz), main ="período por sexo", xlab ="", ylab="",  
 label = FALSE, show.margins = FALSE)



#aplicando teste chiquadrado para independencia das amostras  
chisq <- chisq.test(pr.sexo.matriz)

## Warning in chisq.test(pr.sexo.matriz): Chi-squared approximation may be  
## incorrect

chisq

##   
## Pearson's Chi-squared test  
##   
## data: pr.sexo.matriz  
## X-squared = 84.425, df = 8, p-value = 6.263e-15