trabalho01.R

Betman

2024-04-30

library(readr)  
library(doBy)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:doBy':  
##   
## order\_by

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

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

library(ggplot2)  
  
################################################################################  
##00 Tratamento de Dados #######################################################  
################################################################################  
  
## Leitura da Base de Dados  
stroke\_data <- read\_csv("healthcare-dataset-stroke-data.csv")

## Rows: 5110 Columns: 12

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (6): gender, ever\_married, work\_type, Residence\_type, bmi, smoking\_status  
## dbl (6): id, age, hypertension, heart\_disease, avg\_glucose\_level, stroke  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Remoção de Colunas não usadas  
stroke\_data = select(stroke\_data, -work\_type)  
stroke\_data = select(stroke\_data, -ever\_married)  
stroke\_data = select(stroke\_data, -Residence\_type)  
  
## Exclusão de objeto na Base de Dados  
stroke\_data = stroke\_data[stroke\_data$gender!="Other",]  
  
## Mudança de status das variáveis  
stroke\_data$hypertension = case\_match(stroke\_data$hypertension, 0~"Não", 1~"Sim")  
stroke\_data$heart\_disease = case\_match(stroke\_data$heart\_disease, 0~"Não", 1~"Sim")  
stroke\_data$stroke = case\_match(stroke\_data$stroke, 0~"Não", 1~"Sim")  
stroke\_data$gender = case\_match(stroke\_data$gender, "Female"~"Feminino", "Male"~"Masculino")  
stroke\_data$smoking\_status = case\_match(stroke\_data$smoking\_status, "formerly smoked"~"Fumou",  
 "never smoked"~"Nunca Fumou",  
 "smokes"~"Fuma", "Unknown"~"S/ info")  
  
## Mudança de tipagem da variavel IMC : chr --> float  
stroke\_data = stroke\_data %>%  
 mutate(imc = case\_when(bmi == "N/A" ~ -1,   
 .default = as.numeric(bmi)))

## Warning: There was 1 warning in `mutate()`.  
## ℹ In argument: `imc = case\_when(bmi == "N/A" ~ -1, .default =  
## as.numeric(bmi))`.  
## Caused by warning in `vec\_case\_when()`:  
## ! NAs introduzidos por coerção

stroke\_data = select(stroke\_data, -bmi)  
  
################################################################################  
##01 Análise de Dados ##########################################################  
################################################################################  
  
################################################################################  
# Funções Úteis Criadas:  
  
## Criação de uma Tabela de apenas uma variável  
one\_var\_table <- function(data\_var, name\_var){  
 ftabela = table(data\_var, useNA = "ifany")   
 ptabela = round(prop.table(ftabela)\*100,1)  
 tabela = data.frame(ftabela,ptabela)  
 tabela = tabela[,-3]  
 colnames(tabela) <- c(name\_var,"Frequencia","Porcentagem")  
   
 print(tabela)  
   
 return(tabela)  
}  
  
## Criação de uma Tabela de duas variáveis  
two\_var\_table <- function(var1, var2, name1, name2){  
 ftabela = table(var1, var2, useNA = "ifany")   
 ptabela = round(prop.table(ftabela,1)\*100,1)  
 tabela = data.frame(ftabela,ptabela)  
 tabela = tabela[,-c(4,5)]  
 colnames(tabela) <- c(name1,name2,"Frequencia","Porcentagem")  
   
 print(tabela)  
   
 return(tabela)  
}  
  
## Criação de uma Tabela relacionando uma variável à variável AVC  
stroke\_var\_table <- function(data\_var, name\_var){  
 tabela = two\_var\_table(data\_var, stroke\_data$stroke, name\_var, "AVC")  
   
 print(tabela)  
   
 return(tabela)  
}  
  
#################################################################################  
# MEDIDAS DE RESUMO PARA QUANTITATIVAS:  
  
## Idade  
summary(stroke\_data$age)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.08 25.00 45.00 43.23 61.00 82.00

## Nível médio de glicose  
summary(stroke\_data$avg\_glucose\_level)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 55.12 77.24 91.88 106.14 114.09 271.74

## IMC  
summary(stroke\_data$imc[stroke\_data$imc != -1])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10.30 23.50 28.10 28.89 33.10 97.60

################################################################################  
# TABELAS PARA VARIAVEIS QUALITATIVAS:  
  
## Gênero:  
tabela\_gender = one\_var\_table(stroke\_data$gender, "Gênero")

## Gênero Frequencia Porcentagem  
## 1 Feminino 2994 58.6  
## 2 Masculino 2115 41.4

tabela\_gender\_stroke = stroke\_var\_table(stroke\_data$gender, "Gênero")

## Gênero AVC Frequencia Porcentagem  
## 1 Feminino Não 2853 95.3  
## 2 Masculino Não 2007 94.9  
## 3 Feminino Sim 141 4.7  
## 4 Masculino Sim 108 5.1  
## Gênero AVC Frequencia Porcentagem  
## 1 Feminino Não 2853 95.3  
## 2 Masculino Não 2007 94.9  
## 3 Feminino Sim 141 4.7  
## 4 Masculino Sim 108 5.1

## Hipertensão  
tabela\_hypertension = one\_var\_table(stroke\_data$hypertension, "Hipertensão")

## Hipertensão Frequencia Porcentagem  
## 1 Não 4611 90.3  
## 2 Sim 498 9.7

tabela\_hypertension\_stroke = stroke\_var\_table(stroke\_data$hypertension, "Hipertensão")

## Hipertensão AVC Frequencia Porcentagem  
## 1 Não Não 4428 96.0  
## 2 Sim Não 432 86.7  
## 3 Não Sim 183 4.0  
## 4 Sim Sim 66 13.3  
## Hipertensão AVC Frequencia Porcentagem  
## 1 Não Não 4428 96.0  
## 2 Sim Não 432 86.7  
## 3 Não Sim 183 4.0  
## 4 Sim Sim 66 13.3

## Problema no Coração  
tabela\_heart = one\_var\_table(stroke\_data$heart\_disease, "Cardiaco")

## Cardiaco Frequencia Porcentagem  
## 1 Não 4833 94.6  
## 2 Sim 276 5.4

tabela\_heart\_stroke = stroke\_var\_table(stroke\_data$heart\_disease, "Cardiaco")

## Cardiaco AVC Frequencia Porcentagem  
## 1 Não Não 4631 95.8  
## 2 Sim Não 229 83.0  
## 3 Não Sim 202 4.2  
## 4 Sim Sim 47 17.0  
## Cardiaco AVC Frequencia Porcentagem  
## 1 Não Não 4631 95.8  
## 2 Sim Não 229 83.0  
## 3 Não Sim 202 4.2  
## 4 Sim Sim 47 17.0

## Histórico de fumante  
tabela\_smoke = one\_var\_table(stroke\_data$smoking\_status, "Cigarro")

## Cigarro Frequencia Porcentagem  
## 1 Fuma 789 15.4  
## 2 Fumou 884 17.3  
## 3 Nunca Fumou 1892 37.0  
## 4 S/ info 1544 30.2

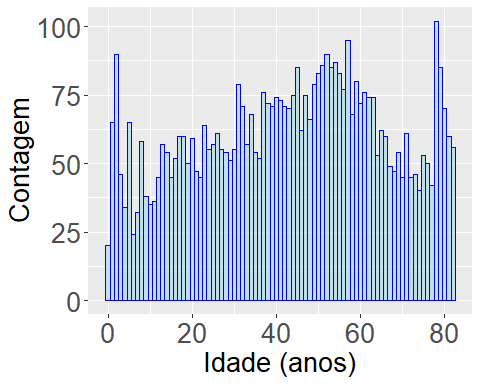
tabela\_smoke\_stroke = stroke\_var\_table(stroke\_data$smoking\_status, "Cigarro")

## Cigarro AVC Frequencia Porcentagem  
## 1 Fuma Não 747 94.7  
## 2 Fumou Não 814 92.1  
## 3 Nunca Fumou Não 1802 95.2  
## 4 S/ info Não 1497 97.0  
## 5 Fuma Sim 42 5.3  
## 6 Fumou Sim 70 7.9  
## 7 Nunca Fumou Sim 90 4.8  
## 8 S/ info Sim 47 3.0  
## Cigarro AVC Frequencia Porcentagem  
## 1 Fuma Não 747 94.7  
## 2 Fumou Não 814 92.1  
## 3 Nunca Fumou Não 1802 95.2  
## 4 S/ info Não 1497 97.0  
## 5 Fuma Sim 42 5.3  
## 6 Fumou Sim 70 7.9  
## 7 Nunca Fumou Sim 90 4.8  
## 8 S/ info Sim 47 3.0

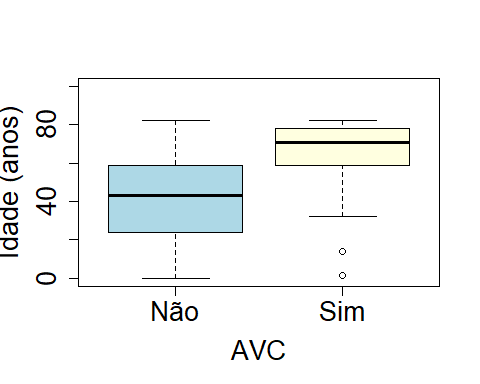
## AVC  
tabela\_stroke = one\_var\_table(stroke\_data$stroke, "AVC")

## AVC Frequencia Porcentagem  
## 1 Não 4860 95.1  
## 2 Sim 249 4.9

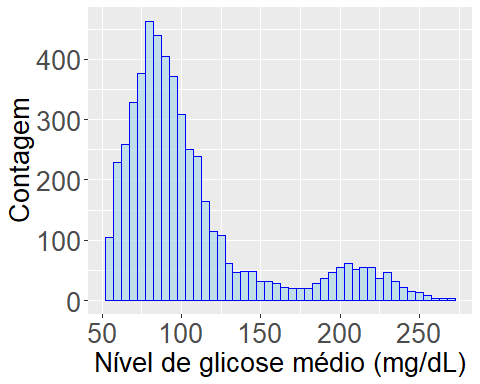
#################################################################################  
# ALTERAÇÃO DO TAMANHO DA FONTE DOS GRÁFICOS DE CAIXA  
par(cex.axis = 1.7, cex.lab = 1.7)  
  
#################################################################################  
# GRAFICOS DAS VARIAVEIS DE ESTUDO:  
  
## Idade  
ggplot(stroke\_data, aes(x=age)) +  
 geom\_histogram(binwidth=1, alpha=0.7, color="blue", fill="lightblue") +   
 labs(x="Idade (anos)", y="Contagem") +  
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))



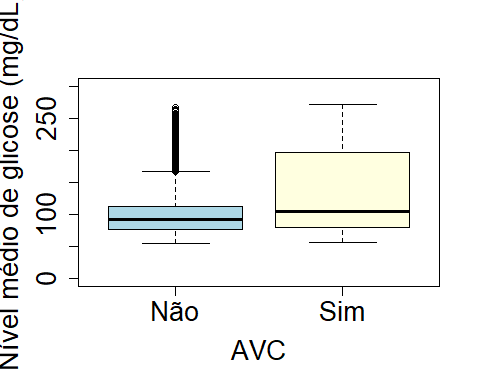
boxplot(age~stroke,  
 data = stroke\_data,  
 xlab = "AVC",  
 ylab = "Idade (anos)",  
 col = c("lightblue","lightyellow"),  
 ylim = c(0,100))



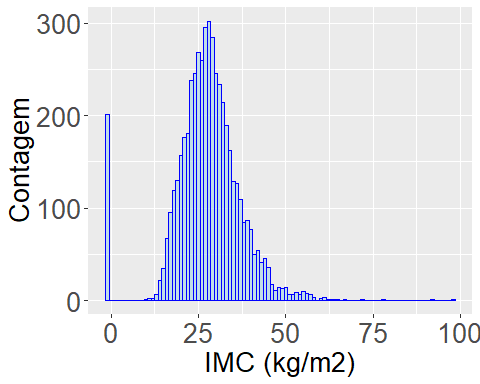
## Nível médio de glicose  
ggplot(stroke\_data, aes(x=avg\_glucose\_level)) +  
 geom\_histogram(binwidth=5, alpha=0.7, color="blue", fill="lightblue") +   
 labs(x="Nível de glicose médio (mg/dL)", y="Contagem") +   
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))



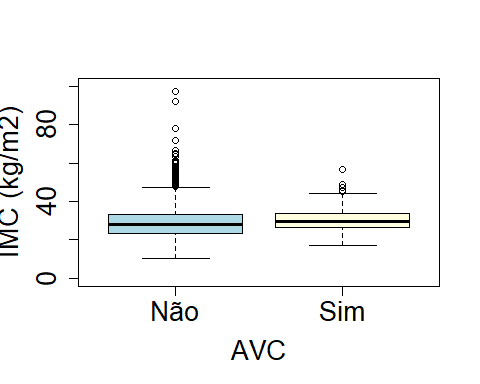
boxplot(avg\_glucose\_level~stroke,  
 data = stroke\_data,  
 xlab = "AVC",  
 ylab = "Nível médio de glicose (mg/dL)",  
 col = c("lightblue","lightyellow"),  
 ylim = c(0,300))



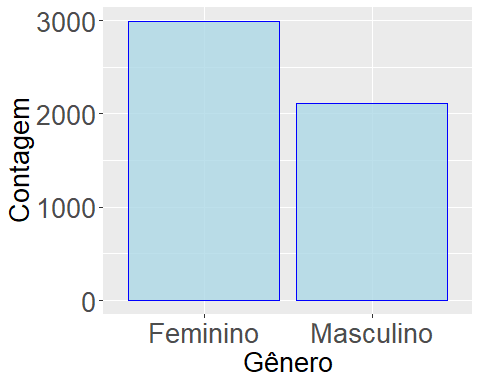
## IMC  
ggplot(stroke\_data, aes(x=imc)) +  
 geom\_histogram(binwidth=1, alpha=0.7, color="blue", fill="lightblue") +   
 labs(x="IMC (kg/m2)", y="Contagem") +   
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))



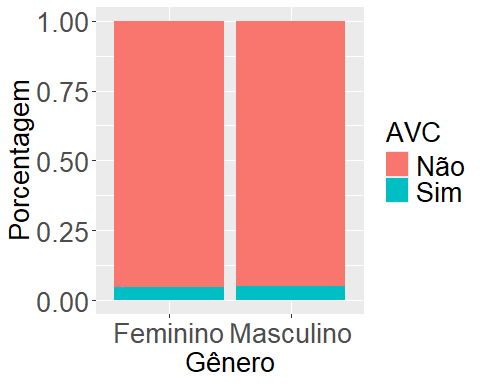
stroke\_data\_filtered = stroke\_data[stroke\_data$imc != -1,]  
boxplot(imc~stroke,  
 data = stroke\_data\_filtered,  
 xlab = "AVC",  
 ylab = "IMC (kg/m2)",  
 col = c("lightblue","lightyellow"),  
 ylim = c(0,100))



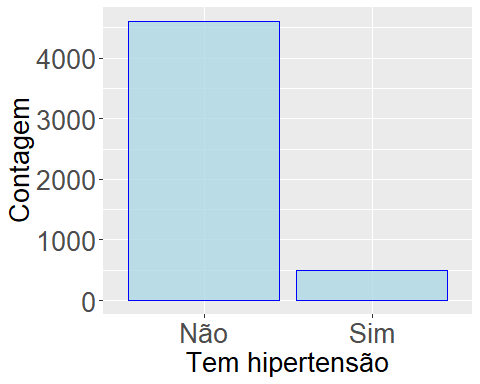
## Gênero  
ggplot(stroke\_data, aes(x=gender)) +  
 geom\_bar(fill="lightblue", color="blue", alpha=0.8) +  
 labs(x="Gênero", y="Contagem") +   
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))



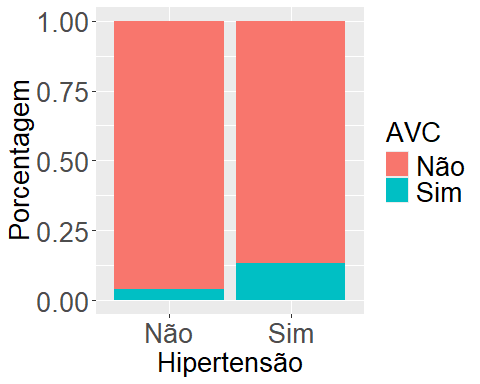
ggplot(stroke\_data, aes(x=gender, fill=stroke)) +   
 geom\_bar(position="fill") +  
 xlab("Gênero") +  
 ylab("Porcentagem") +   
 labs(fill="AVC") +  
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 20))



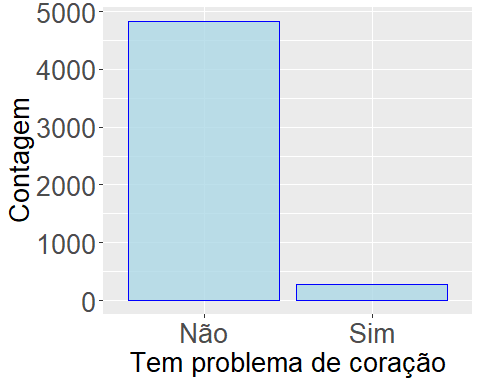
## Hipertensão  
ggplot(stroke\_data, aes(x=hypertension)) +  
 geom\_bar(fill="lightblue", color="blue", alpha=0.8) +  
 labs(x="Tem hipertensão", y="Contagem") +   
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))



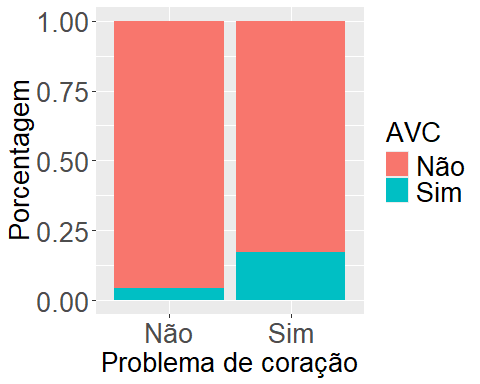
ggplot(stroke\_data, aes(x=hypertension, fill=stroke)) +   
 geom\_bar(position="fill") +  
 xlab("Hipertensão") +  
 ylab("Porcentagem") +   
 labs(fill="AVC") +  
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 20))



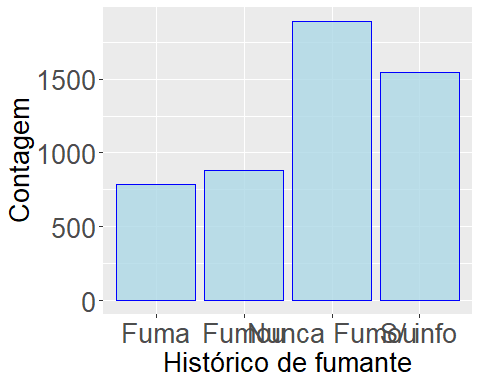
## Problema de coração  
ggplot(stroke\_data, aes(x=heart\_disease)) +  
 geom\_bar(fill="lightblue", color="blue", alpha=0.8) +  
 labs(x="Tem problema de coração", y="Contagem") +   
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))



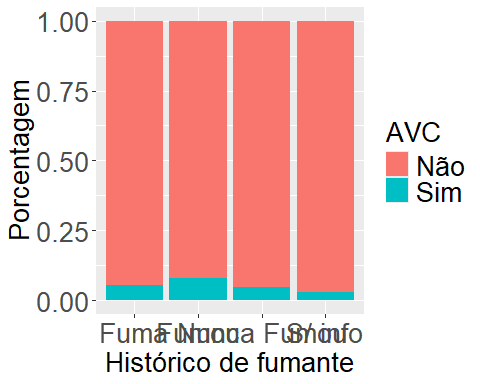
ggplot(stroke\_data, aes(x=heart\_disease, fill=stroke)) +   
 geom\_bar(position="fill") +  
 xlab("Problema de coração") +  
 ylab("Porcentagem") +   
 labs(fill="AVC") +  
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 20))



## Histórico de fumante  
ggplot(stroke\_data, aes(x=smoking\_status)) +  
 geom\_bar(fill="lightblue", color="blue", alpha=0.8) +  
 labs(x="Histórico de fumante", y="Contagem") +   
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))



ggplot(stroke\_data, aes(x=smoking\_status, fill=stroke)) +   
 geom\_bar(position="fill") +  
 xlab("Histórico de fumante") +  
 ylab("Porcentagem") +   
 labs(fill="AVC") +  
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 20))



## AVC  
ggplot(stroke\_data, aes(x=stroke)) +  
 geom\_bar(fill="lightblue", color="blue", alpha=0.8) +  
 labs(x="Teve AVC", y="Contagem") +   
 theme(axis.text.x = element\_text(size = 20),  
 axis.text.y = element\_text(size = 20),  
 axis.title.x = element\_text(size = 20),  
 axis.title.y = element\_text(size = 20))

