

Dataframes

Ivanildo Batista

13 de marco de 2021

Trabalhando com Dataframes

```
#criando dataframes
df=data.frame() #um dataframe vazio
class(df)#verificando a classe
```

```
## [1] "data.frame"
```

```
df #nao possui nem linhas nem colunas
```

```
## data frame with 0 columns and 0 rows
```

```
#criando vetores vazios
nomes = character() #vetor de caracteres
idades = numeric() #vetor de numeros
data = as.Date(character()) #vetor de datas
codigos = integer() #vetor de inteiros
```

```
#transformando esses vetores em um dataframe
df = data.frame(c(nomes,idades,data,codigos))
df
```

```
## [1] c.nomes..idades..data..codigos.
## <0 rows> (or 0-length row.names)
```

```
#criando vetores

pais = c('EUA','Dinamarca','Brasil','Espanha','Holanda')
nome = c('Alberto','Claudio','Joao','Jose','Gilberto')
altura = c(1.93,1.59,1.63,1.59,1.58)
codigo = c(5069,3258,6358,1258,4555)
```

```
#criando o dataframe
pesq = data.frame(pais,nome,altura,codigo)
pesq
```

```
##      pais      nome altura codigo
## 1      EUA   Alberto   1.93   5069
## 2 Dinamarca Claudio   1.59   3258
## 3   Brasil    Joao    1.63   6358
## 4   Espanha    Jose    1.59   1258
## 5   Holanda Gilberto   1.58   4555
```

```
#inserindo uma nova coluna
olhos = c('verde','castanho','azul','azul','castanho')

pesq = cbind(pesq,olhos)
pesq
```

```
##      pais      nome altura codigo   olhos
## 1      EUA   Alberto   1.93   5069   verde
## 2 Dinamarca Claudio   1.59   3258 castanho
## 3   Brasil    Joao    1.63   6358    azul
## 4   Espanha    Jose    1.59   1258    azul
## 5   Holanda Gilberto   1.58   4555 castanho
```

```
#informacoes sobre o dataframe

str(pesq) #informacoes gerais do dataframe
```

```
## 'data.frame':   5 obs. of  5 variables:
## $ pais : chr  "EUA" "Dinamarca" "Brasil" "Espanha" ...
## $ nome : chr  "Alberto" "Claudio" "Joao" "Jose" ...
## $ altura: num  1.93 1.59 1.63 1.59 1.58
## $ codigo: num  5069 3258 6358 1258 4555
## $ olhos : chr  "verde" "castanho" "azul" "azul" ...
```

```
dim(pesq) #dimensao do dataframe
```

```
## [1] 5 5
```

```
length(pesq) #comprimento do dataframe
```

```
## [1] 5
```

```
# Obtendo um vetor do dataframe
pesq$pais
```

```
## [1] "EUA"      "Dinamarca" "Brasil"    "Espanha"  "Holanda"
```

```
pesq$altura
```

```
## [1] 1.93 1.59 1.63 1.59 1.58
```

```
#Extraindo um valor unico do dataframe
```

```
pesq[1,1]
```

```
## [1] "EUA"
```

```
pesq[3,2]
```

```
## [1] "Joao"
```

```
#numero de linhas e de colunas do dataframe
```

```
nrow(pesq)
```

```
## [1] 5
```

```
ncol(pesq)
```

```
## [1] 5
```

```
#primeiros elementos do dataframe
```

```
head(pesq)
```

```
##      pais      nome altura codigo  olhos
## 1     EUA  Alberto   1.93   5069   verde
## 2 Dinamarca Claudio   1.59   3258 castanho
## 3   Brasil    Joao   1.63   6358    azul
## 4  Espanha    Jose   1.59   1258    azul
## 5  Holanda Gilberto  1.58   4555 castanho
```

```
head(mtcars)
```

```
##      mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710      22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant        18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
#ultimos elementos do dataframe
tail(pesq)
```

```
##      pais      nome altura codigo  olhos
## 1     EUA   Alberto   1.93   5069   verde
## 2 Dinamarca Claudio   1.59   3258 castanho
## 3   Brasil    Joao   1.63   6358   azul
## 4  Espanha    Jose   1.59   1258   azul
## 5  Holanda Gilberto  1.58   4555 castanho
```

```
tail(mtcars)
```

```
##      mpg cyl  disp  hp drat   wt  qsec vs am gear carb
## Porsche 914-2 26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa  30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino  19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E    21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

```
#filtrando o dataframe
pesq[altura<1.60]
```

```
##      nome codigo  olhos
## 1 Alberto   5069   verde
## 2 Claudio   3258 castanho
## 3 Joao     6358   azul
## 4 Jose     1258   azul
## 5 Gilberto  4555 castanho
```

```
pesq[altura<1.60, c('codigo','olhos')] #selecionada as colunas codigo e olhos, para os valores d
e altura menores que 1.60
```

```
##      codigo  olhos
## 2    3258 castanho
## 4    1258   azul
## 5    4555 castanho
```

```
#nomeando dataframes
```

```
names(pesq) = c('Pais','Nome','Altura','Codigos','Olhos')
pesq
```

```
##      Pais      Nome Altura Codigos  Olhos
## 1      EUA   Alberto   1.93   5069   verde
## 2 Dinamarca Claudio   1.59   3258 castanho
## 3   Brasil    Joao    1.63   6358   azul
## 4  Espanha    Jose    1.59   1258   azul
## 5  Holanda Gilberto   1.58   4555 castanho
```

```
colnames(pesq) = c('Var 1','Var 2','Var 3','Var 4','Var 5') #renomeando apenas colunas
pesq
```

```
##      Var 1      Var 2 Var 3 Var 4      Var 5
## 1      EUA   Alberto   1.93  5069   verde
## 2 Dinamarca Claudio   1.59  3258 castanho
## 3   Brasil    Joao    1.63  6358   azul
## 4  Espanha    Jose    1.59  1258   azul
## 5  Holanda Gilberto   1.58  4555 castanho
```

```
rownames(pesq) = c("Obs 1","Obs 2","Obs 3","Obs 4","Obs 5") #renomeando apenas as linhas
pesq
```

```
##      Var 1      Var 2 Var 3 Var 4      Var 5
## Obs 1      EUA   Alberto   1.93  5069   verde
## Obs 2 Dinamarca Claudio   1.59  3258 castanho
## Obs 3   Brasil    Joao    1.63  6358   azul
## Obs 4  Espanha    Jose    1.59  1258   azul
## Obs 5  Holanda Gilberto   1.58  4555 castanho
```

#formas de ler bases de dados:

```
##read.csv() - arquivos csv
##read.xls() - arquivos excel
##read.spss() - arquivos spss
##read.mtp() - arquivos minitab
##read.table() - arquivos txt
##read.delim() - leitura de arquivos delimitados
```

#dataframe de um arquivo csv

```
df2 = data.frame(read.csv(file = 'C:/Users/junio/dframe.csv',header = TRUE, sep = ','))
```

df2 #imprimindo do dataframe

```
##      ID      Nome Idade   Admdate Diabetes Status
## 1  1 Paciente 1    43 15/10/2015 Tipo 1   Ruim
## 2  2 Paciente 2    23 16/10/2015 Tipo 2    Bom
## 3  3 Paciente 3    56 23/10/2015 Tipo 2    Bom
## 4  4 Paciente 4    34 23/10/2015 Tipo 1   Ruim
## 5  5 Paciente 5    38 31/10/2015 Tipo 1  Medio
## 6  6 Paciente 6    37 28/10/2015 Tipo 1    Bom
## 7  7 Paciente 7    41 27/10/2015 Tipo 1   Ruim
```

```
head(df2) #primeiras linhas do dataframe
```

```
##      ID      Nome Idade   Admdate Diabetes Status
## 1  1 Paciente 1    43 15/10/2015 Tipo 1   Ruim
## 2  2 Paciente 2    23 16/10/2015 Tipo 2    Bom
## 3  3 Paciente 3    56 23/10/2015 Tipo 2    Bom
## 4  4 Paciente 4    34 23/10/2015 Tipo 1   Ruim
## 5  5 Paciente 5    38 31/10/2015 Tipo 1  Medio
## 6  6 Paciente 6    37 28/10/2015 Tipo 1    Bom
```

```
summary(df2) #sumario do dataframe
```

```
##           ID           Nome           Idade           Admdate
## Min.      :1.0    Length:7           Min.      :23.00    Length:7
## 1st Qu.:2.5    Class :character    1st Qu.:35.50    Class :character
## Median :4.0    Mode  :character    Median :38.00    Mode  :character
## Mean    :4.0                                Mean    :38.86
## 3rd Qu.:5.5                                3rd Qu.:42.00
## Max.     :7.0                                Max.     :56.00
##      Diabetes           Status
## Length:7           Length:7
## Class :character    Class :character
## Mode  :character    Mode  :character
##
##
##
```

```
df2$Diabete #coluna diabetes
```

```
## [1] "Tipo 1" "Tipo 2" "Tipo 2" "Tipo 1" "Tipo 1" "Tipo 1" "Tipo 1"
```

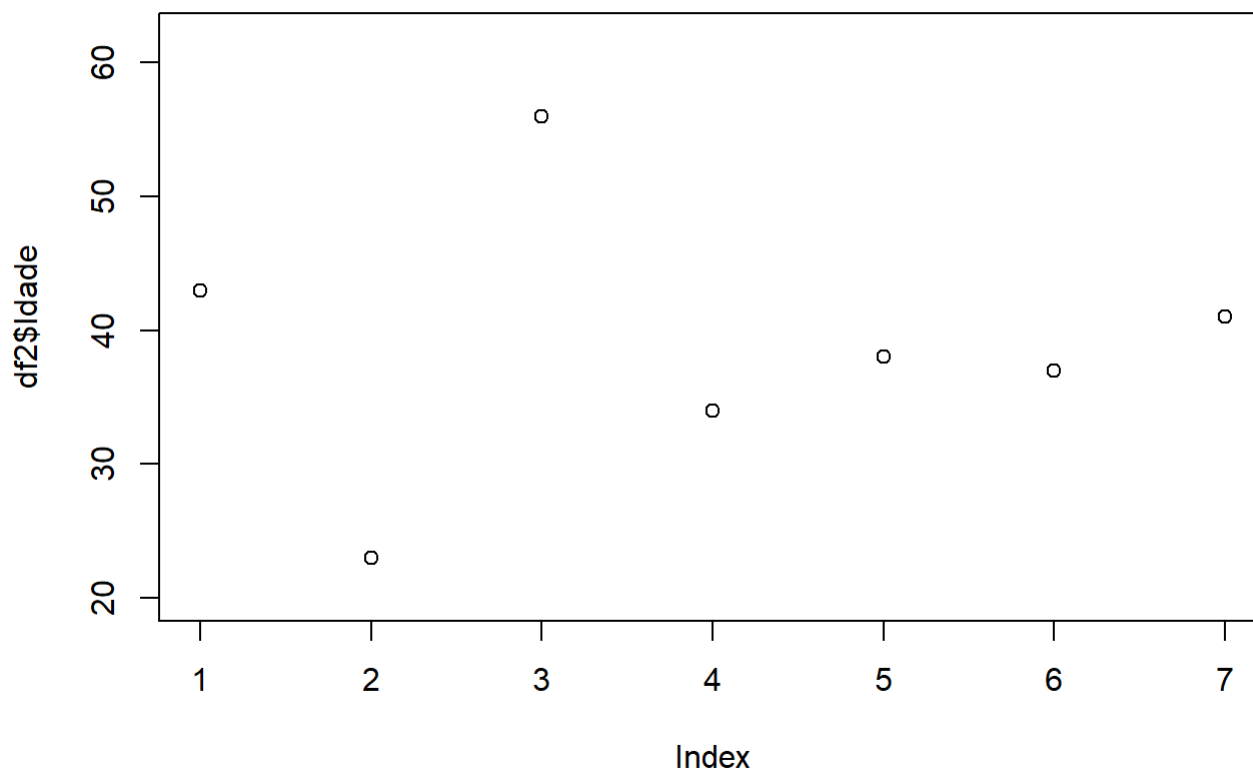
```
df2$status #retorna um valor nulo, pois o certo e Status
```

```
## NULL
```

```
df2$Status #forma correta
```

```
## [1] "Ruim" "Bom" "Bom" "Ruim" "Medio" "Bom " "Ruim"
```

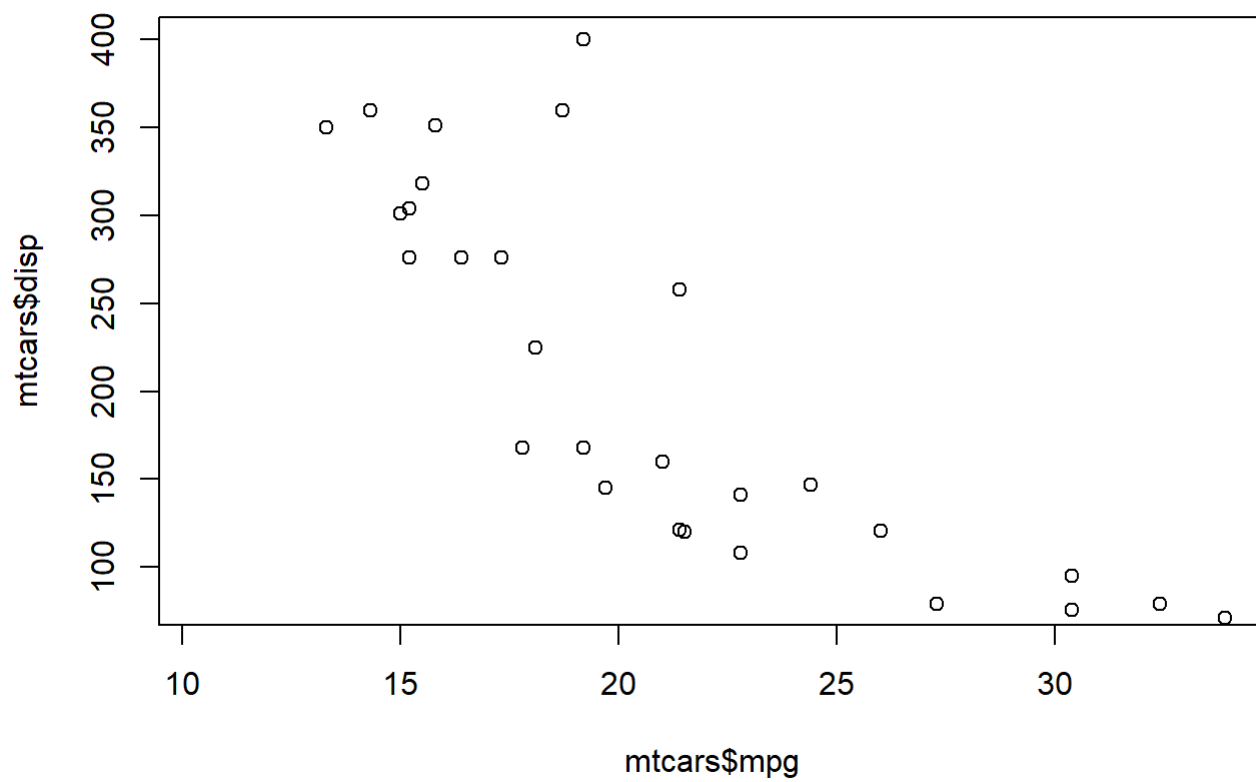
```
plot(df2$Idade, ylim = c(20,62)) #gerando um grafico da coluna selecionada
```



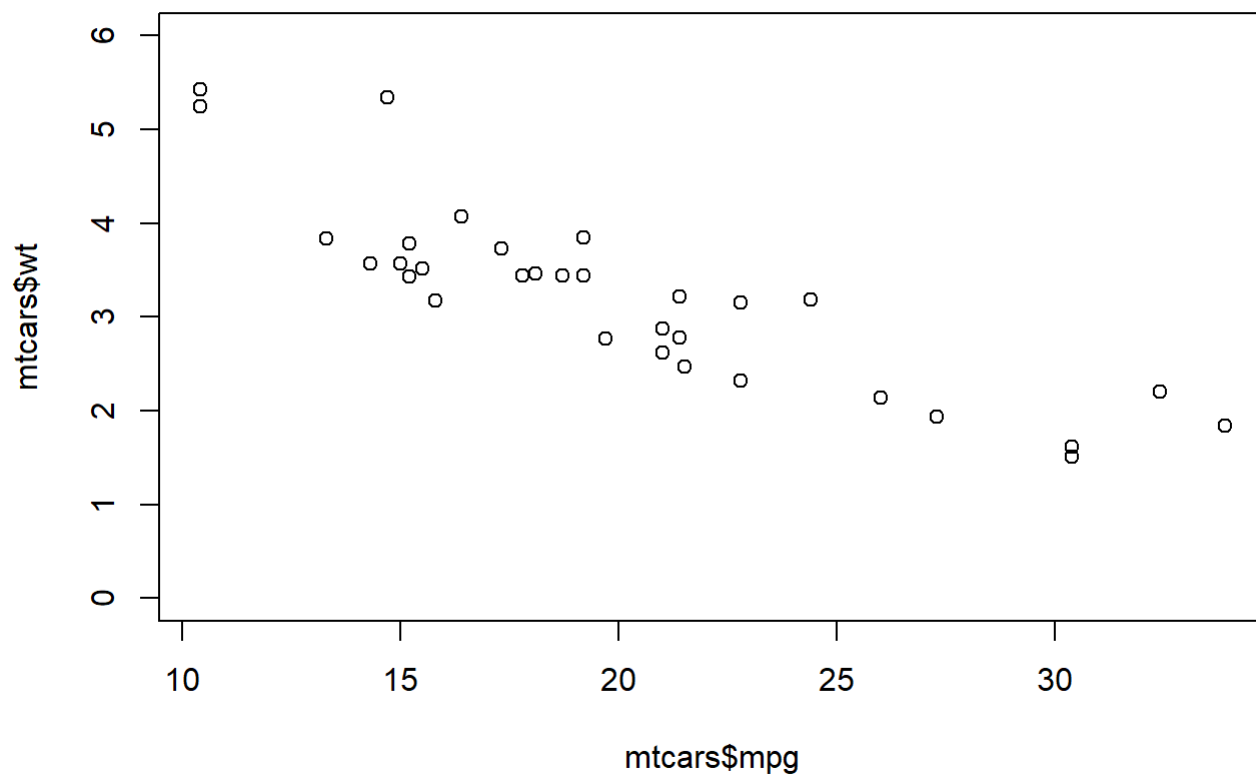
```
#usando a base mtcars
summary(mtcars$mpg)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  10.40   15.43   19.20   20.09   22.80   33.90
```

```
plot(mtcars$mpg, mtcars$disp, ylim = c(80,400)) #grafico de dispersao entre duas variaveis
```



```
plot(mtcars$mpg, mtcars$wt, ylim = c(0,6)) #outro grafico de dispersao entre duas variaveis
```

```
#combinando dataframes  
df3 = merge(pesq,df2)  
df3
```

##	Var 1	Var 2	Var 3	Var 4	Var 5	ID	Nome	Idade	Admdate
## 1	EUA	Alberto	1.93	5069	verde	1	Paciente 1	43	15/10/2015
## 2	Dinamarca	Claudio	1.59	3258	castanho	1	Paciente 1	43	15/10/2015
## 3	Brasil	Joao	1.63	6358	azul	1	Paciente 1	43	15/10/2015
## 4	Espanha	Jose	1.59	1258	azul	1	Paciente 1	43	15/10/2015
## 5	Holanda	Gilberto	1.58	4555	castanho	1	Paciente 1	43	15/10/2015
## 6	EUA	Alberto	1.93	5069	verde	2	Paciente 2	23	16/10/2015
## 7	Dinamarca	Claudio	1.59	3258	castanho	2	Paciente 2	23	16/10/2015
## 8	Brasil	Joao	1.63	6358	azul	2	Paciente 2	23	16/10/2015
## 9	Espanha	Jose	1.59	1258	azul	2	Paciente 2	23	16/10/2015
## 10	Holanda	Gilberto	1.58	4555	castanho	2	Paciente 2	23	16/10/2015
## 11	EUA	Alberto	1.93	5069	verde	3	Paciente 3	56	23/10/2015
## 12	Dinamarca	Claudio	1.59	3258	castanho	3	Paciente 3	56	23/10/2015
## 13	Brasil	Joao	1.63	6358	azul	3	Paciente 3	56	23/10/2015
## 14	Espanha	Jose	1.59	1258	azul	3	Paciente 3	56	23/10/2015
## 15	Holanda	Gilberto	1.58	4555	castanho	3	Paciente 3	56	23/10/2015
## 16	EUA	Alberto	1.93	5069	verde	4	Paciente 4	34	23/10/2015
## 17	Dinamarca	Claudio	1.59	3258	castanho	4	Paciente 4	34	23/10/2015
## 18	Brasil	Joao	1.63	6358	azul	4	Paciente 4	34	23/10/2015
## 19	Espanha	Jose	1.59	1258	azul	4	Paciente 4	34	23/10/2015
## 20	Holanda	Gilberto	1.58	4555	castanho	4	Paciente 4	34	23/10/2015
## 21	EUA	Alberto	1.93	5069	verde	5	Paciente 5	38	31/10/2015
## 22	Dinamarca	Claudio	1.59	3258	castanho	5	Paciente 5	38	31/10/2015
## 23	Brasil	Joao	1.63	6358	azul	5	Paciente 5	38	31/10/2015
## 24	Espanha	Jose	1.59	1258	azul	5	Paciente 5	38	31/10/2015
## 25	Holanda	Gilberto	1.58	4555	castanho	5	Paciente 5	38	31/10/2015
## 26	EUA	Alberto	1.93	5069	verde	6	Paciente 6	37	28/10/2015
## 27	Dinamarca	Claudio	1.59	3258	castanho	6	Paciente 6	37	28/10/2015
## 28	Brasil	Joao	1.63	6358	azul	6	Paciente 6	37	28/10/2015
## 29	Espanha	Jose	1.59	1258	azul	6	Paciente 6	37	28/10/2015
## 30	Holanda	Gilberto	1.58	4555	castanho	6	Paciente 6	37	28/10/2015
## 31	EUA	Alberto	1.93	5069	verde	7	Paciente 7	41	27/10/2015
## 32	Dinamarca	Claudio	1.59	3258	castanho	7	Paciente 7	41	27/10/2015
## 33	Brasil	Joao	1.63	6358	azul	7	Paciente 7	41	27/10/2015
## 34	Espanha	Jose	1.59	1258	azul	7	Paciente 7	41	27/10/2015
## 35	Holanda	Gilberto	1.58	4555	castanho	7	Paciente 7	41	27/10/2015
##	Diabete Status								
## 1	Tipo 1	Ruim							
## 2	Tipo 1	Ruim							
## 3	Tipo 1	Ruim							
## 4	Tipo 1	Ruim							
## 5	Tipo 1	Ruim							
## 6	Tipo 2	Bom							
## 7	Tipo 2	Bom							
## 8	Tipo 2	Bom							
## 9	Tipo 2	Bom							
## 10	Tipo 2	Bom							
## 11	Tipo 2	Bom							
## 12	Tipo 2	Bom							
## 13	Tipo 2	Bom							
## 14	Tipo 2	Bom							
## 15	Tipo 2	Bom							
## 16	Tipo 1	Ruim							

```
## 17 Tipo 1 Ruim
## 18 Tipo 1 Ruim
## 19 Tipo 1 Ruim
## 20 Tipo 1 Ruim
## 21 Tipo 1 Medio
## 22 Tipo 1 Medio
## 23 Tipo 1 Medio
## 24 Tipo 1 Medio
## 25 Tipo 1 Medio
## 26 Tipo 1 Bom
## 27 Tipo 1 Bom
## 28 Tipo 1 Bom
## 29 Tipo 1 Bom
## 30 Tipo 1 Bom
## 31 Tipo 1 Ruim
## 32 Tipo 1 Ruim
## 33 Tipo 1 Ruim
## 34 Tipo 1 Ruim
## 35 Tipo 1 Ruim
```