

# Universidade Cruzeiro do Sul

Code ▾

Rodrigo de Miranda Videira

28/11/2021

Hide

# Conclusão do problema

O problema de negócio foca na análise de dados de pacientes e valores de seguro liberados. Estes pacientes estão em uma faixa etária de 0 anos (bebês) até pessoas com mais idade, sendo a pessoa mais velha com 107 anos. A idade médias dos pacientes é de 57 anos, metade dos pacientes tem menos de 55 anos e a outra metade mais que 55. Temos 50% de pessoas com grande risco, e assim o valores totais liberados são maiores por este risco. Os valores totais liberados possuem outliers, podendo ser observado pelo gráfico de boxplot. Por estes dados estar com estes valores, não conseguiríamos aplicar uma análise preditiva para que pudessemos estimar valores para todos os tipos de pacientes.

Hide

# Análise Dados

- Importando bibliotecas

Hide

```
library(readr)
library(tidyverse)
```

- Carregando os dados

Hide

```
df <- read_csv("pacientes.csv", show_col_types = FALSE)
```

- Analisando quantidade de observações e variáveis

Hide

```
dim(df)
```

```
[1] 27846    4
```

Hide

```
# Observações = 27846
# Variáveis = 4
```

- Analisando as variáveis

Hide

```
names(df)
```

```
[1] "ID" "Idade do Segurado"
[3] "Código do Procedimento Principal" "Valor Total Liberado"
```

Hide

```
#ID -> Identificador
# Idade do Segurado -> (Int) idade dos segurados
# Código do Procedimento Principal -> (Str) Código
# Valor Total Liberado -> (Float) Valores liberado
```

- Acertando tipos de dados no arquivo

Hide

```
df$ID <- as.character(df$`Código do Procedimento Principal`)
df$`Código do Procedimento Principal` <- as.character(df$`Código do Procedimento Principal`)
```

- Tirando algumas medidas estatísticas

Hide

```
summary(df)
```

Warning: One or more parsing issues, see `problems()` for details

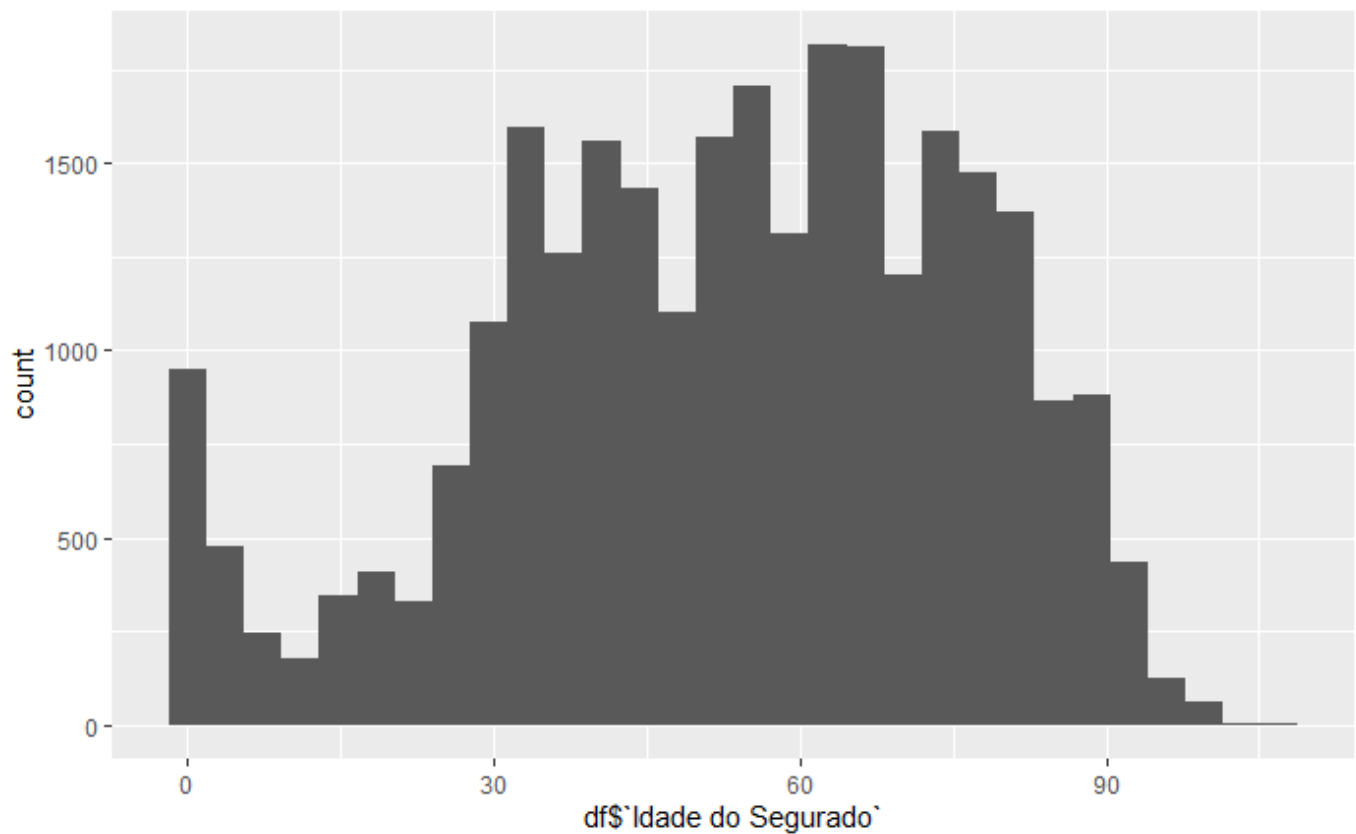
```
      ID      Idade do Segurado Código do Procedimento Principal
Length:27846   Min.   : 0.00   Length:27846
Class :character 1st Qu.: 37.00   Class :character
Mode  :character Median : 55.00   Mode  :character
              Mean  : 53.03
              3rd Qu.: 71.00
              Max.   :107.00

Valor Total Liberado
Min.   : 490.9
1st Qu.: 56620.5
Median : 71872.5
Mean   : 86512.1
3rd Qu.: 98470.8
Max.   :980639.4
NA's   :1
```

- Analisando a variável Idade do Segurado

Hide

```
ggplot(data=df, mapping = aes(df$`Idade do Segurado`)) +
  geom_histogram(bins = 30)
```

[Hide](#)

```
summary(df$`Idade do Segurado`)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	37.00	55.00	53.03	71.00	107.00

- Analisando variável “Valor Total Liberado”

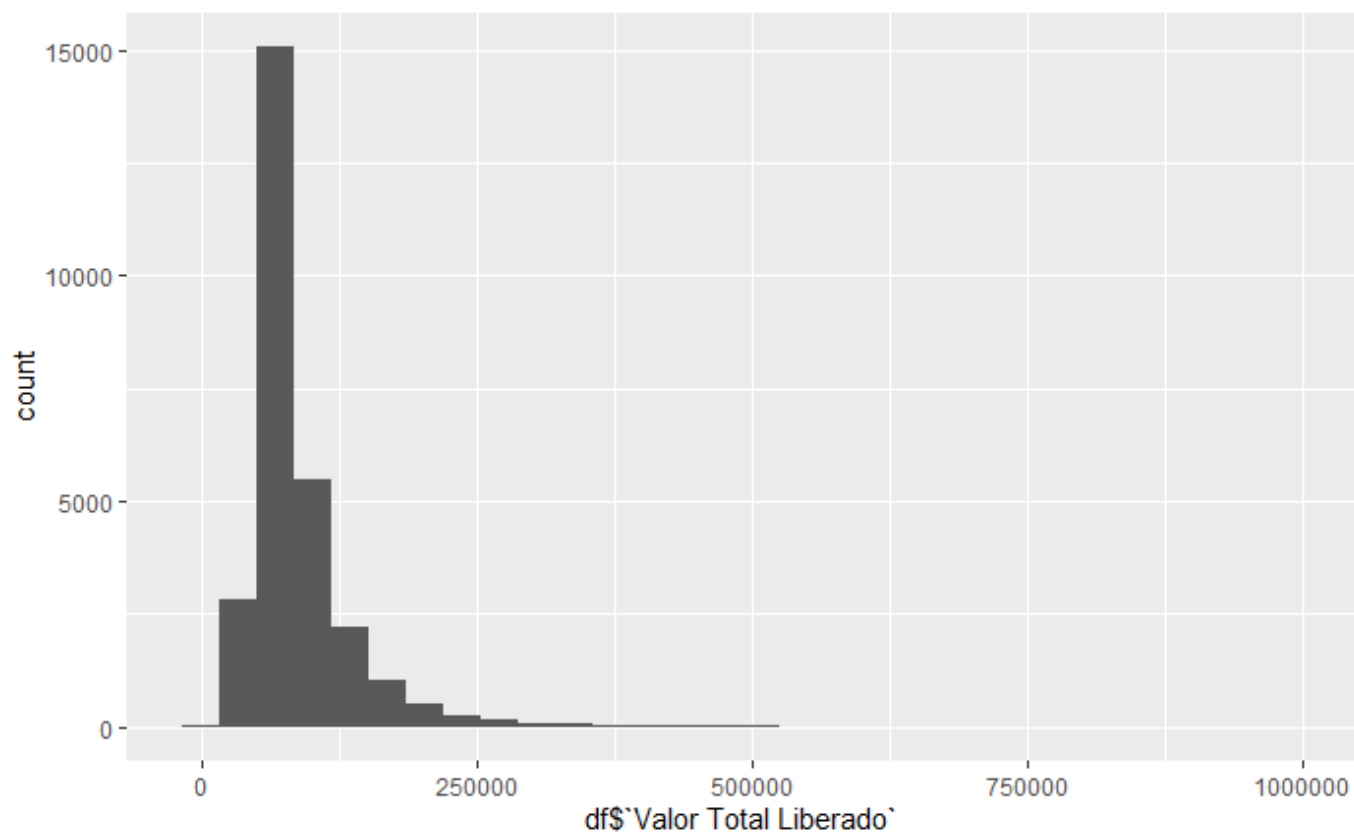
[Hide](#)

```
summary(df$`Valor Total Liberado`)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
490.9	56620.5	71872.5	86512.1	98470.8	980639.4	1

[Hide](#)

```
ggplot(data=df, mapping = aes(df$`Valor Total Liberado`)) +  
  geom_histogram(bins = 30, na.rm = TRUE)
```

[Hide](#)

```
ggplot(data=df, mapping = aes(df$`Valor Total Liberado`)) +  
  geom_boxplot(na.rm = TRUE)
```

