

Universidade Cruzeiro do Sul

Professor: Rafael Guem Murakami

Aluno: Mateus Amorim Marques

RGM: 27227928

#carregando pacotes

```
library(dplyr)
```

```
library(readr)
```

#importação de banco de dados

```
df <- readr::read_csv("C:/Users/mateus/Desktop/Projetos_R/pacientes.csv")
```

#analizando tipo de dados e colunas

```
class(df)
```

```
str(df)
```

#contando quantidade de linhas

```
max(df$ID)
```

```
[1] 27846
```

#calculando média de Valor total liberado

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

```
[1] 86512.12
```

#tratando valores nulos(missing)

```
is.na(df)
```

#melhor desempenho que o anterior

```
any(is.na(df))
```

#excluir linhas com dados faltantes

```
df_1 <- na.omit(df)
```

#verificando quantas observações

```
nrow(df)
```

```
nrow(df_1)
```

#percentual de dados faltantes em cada coluna

#a função round arredonda

```
NAS <- round(colSums(is.na(df))*100/nrow(df),2)
```

```
NAS
```

```
NAS[NAS>0]
```

#observando quantidade de dados faltantes

```
colSums(is.na(df))
```

```
nrow(df)
```

#substituindo dados faltantes por média

#na.rm=TRUE ignora o valor faltante

```
df$`Valor Total Liberado`[is.na(df$`Valor Total Liberado`)] <- mean(df$`Valor Total Liberado`,  
na.rm = TRUE)  
mean(df$`Valor Total Liberado`)
```

#atualizando df_1 com dados tratados

```
df_1 <- na.omit(df)  
nrow(df)  
nrow(df_1)
```

#percentual de dados faltantes em cada coluna

```
NAS <- round(colSums(is.na(df_1))*100/nrow(df_1),2)  
NAS  
NAS[NAS>0]
```

```
any(is.na(df_1))
```

#calculando desvio padrão

```
sd(df_1$`Valor Total Liberado`)
```

#criando gráfico de dispersão com Valor Total Liberado

#mostrando valor médio liberado

```
library(ggplot2)
```

```
dados_0<- data.frame(valor = rnorm(27750,86508.76,49024.49))
```

```
ggplot(dados_0)+
```

```
  aes(x=valor)+
```

```
  geom_histogram(fill="lightblue",
```

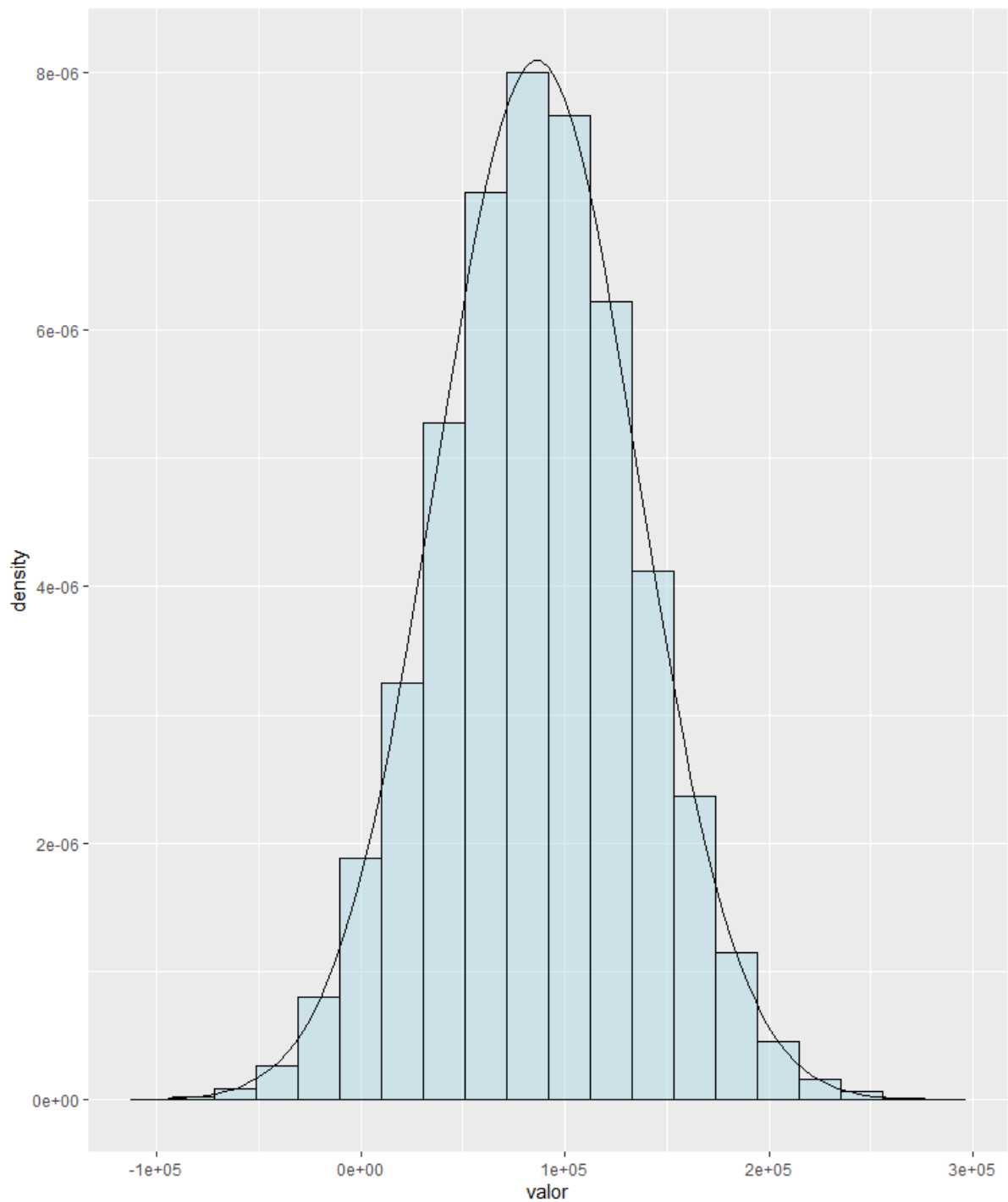
```
    col = "black",
```

```
    alpha = 0.5,
```

```
    bins = 20,
```

```
    aes(y=..density..))+
```

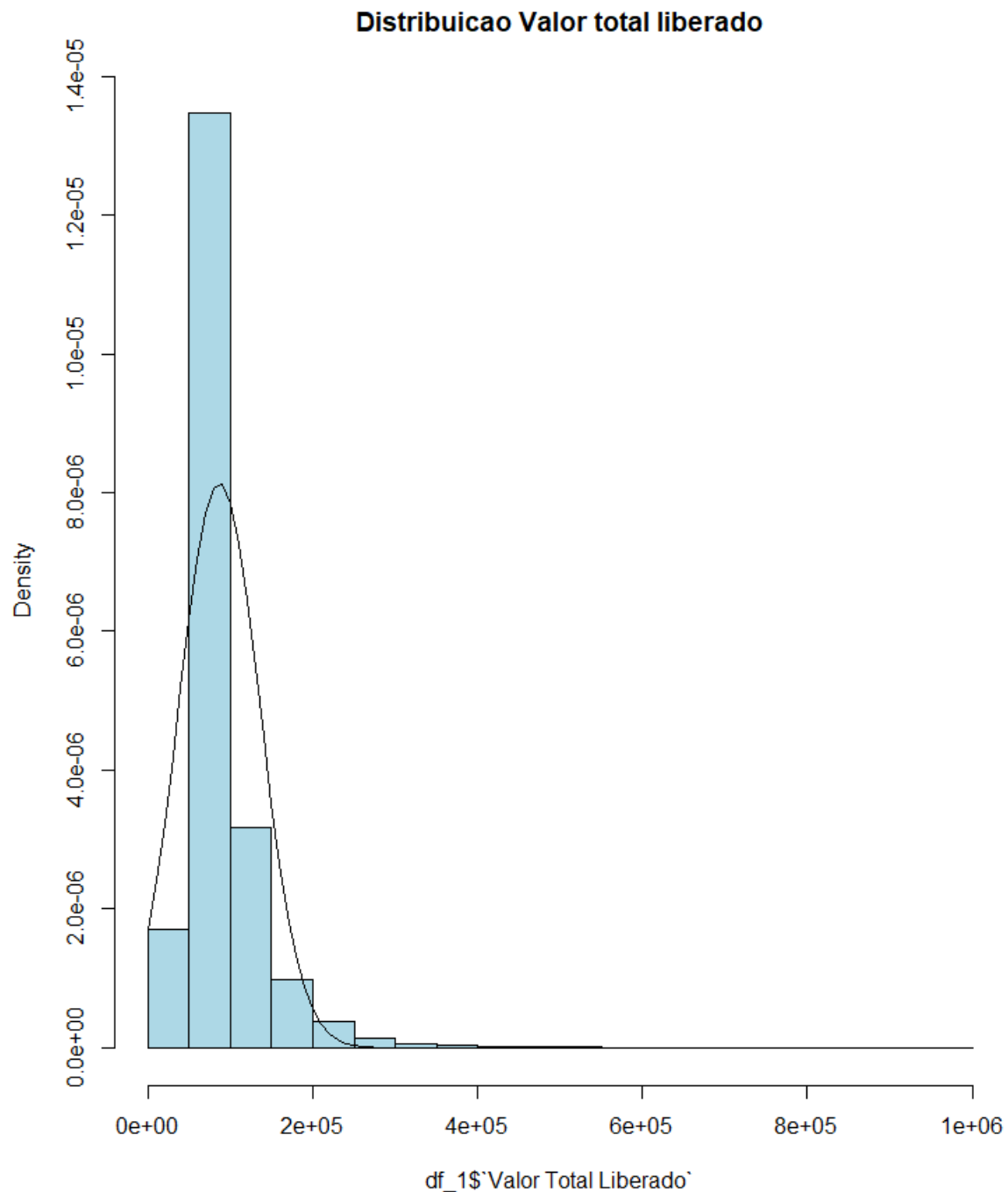
```
  stat_function(fun = dnorm, args = list(mean = mean(dados_0$valor),  
sd=sd(dados_0$valor)))
```



#plotando gráfico de valor total liberado com dados tratados

```
hist(df_1$`Valor Total Liberado`,  
     col = "lightblue",  
     freq = F,  
     main = "Distribuicao Valor total liberado",  
     xlabel = "Valor Total liberado",  
     breaks = 20)
```

```
curve(dnorm(x, mean = mean(df_1$`Valor Total Liberado`), sd = sd(df_1$`Valor Total Liberado`)), add = T)
```



#criacao de dataframe com valor por idade

```
valor_idade <- data.frame(df_1$`Idade do Segurado`,df_1$`Valor Total Liberado`)  
valor_idade
```

#criacao de novo gráfico valor_idade

```
hist(valor_idade,
```

```
col = "lightblue",  
freq = F,  
main = "Distribuicao Valor total liberado",  
xlabel = "Valor Total liberado",  
breaks = 20)  
curve(dnorm(x, mean = mean(valor_idade), sd = sd(valor_idade)), add = T)
```

