Data Frames





• Há várias maneiras para identificar os elementos de um data frame

```
meubanco[2:3]  # columas 2 e3 do data frame
meubanco[c("ID", "verificou")] # columns ID e do data frame
meubanco$Cor  # variable verificou do data frame
```

Laboratório de Estatística - 2020



Data Frames



- É mais geral que uma matriz
 - √ Colunas diferentes de mesmo comprimento podem ter diferentes tipos de dados (numérico, caracter, lógico, fator, etc.)

```
d <- c(1,2,3,4)
e <- c("vermelho", "branco", "vermelho", NA)
f <- c(TRUE,TRUE,TRUE,FALSE)
meubanco <- data.frame(d,e,f)
names(meubanco) <- c("ID","Cor","verificou") # nome das variáveis</pre>
```

Laboratório de Estatística - 2020

84



Exemplo – Consumo de Energia



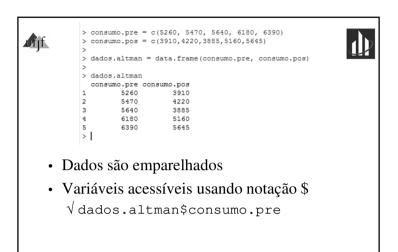
- Consumo de energia pré-menstrual:
 - $\sqrt{\text{consumo.pre}} = c(5260, 5470, 5640, 6180, 6390)$
- Consumo de energia pós-menstrual:
 - $\sqrt{\text{consumo.pos}} = c(3910, 4220, 3885, 5160, 5645)$
- Conjunto de dados (Data frame):
 - $\sqrt{\text{dados.altman=data.frame(consumo.pre,consumo.pos)}}$
- Tamanho do conjunto de dados:

√ dim(dados.altman)

86

Laboratório de Estatística - 2020

Fif



Laboratório de Estatística - 2020

Arie Consumo pré-menstrual todas exceto mulheres # 1, 3 e 5 $\sqrt{\text{consumo.pre}[-c(1,3,5)]}$ > consumo.pre[5] [1] 6390 > consumo.pre[c(3,5,7)] [1] 5640 6390 NA > consumo.pre[1:3] [1] 5260 5470 5640 > consumo.pre[-c(1,3,5)] [1] 5470 6180 > impar=c(1,3,5) > consumo.pre[-impar] [1] 5470 6180 > dados.altman\$consumo.pre[5] [1] 6390 89 Laboratório de Estatística - 2020



Indexação



- Consumo pré-menstrual mulher #5 √consumo.pre[5]
- Consumo pré-menstrual mulheres #1, 3, 5 $\sqrt{\text{consumo.pre}[c(1,3,5)]}$
- Consumo pré-menstrual 3 primeiras mulheres $\sqrt{\text{consumo.pre}[1:3]}$

Laboratório de Estatística - 2020



• Pode-se indexar pelo do Data frame



√ dados.altman\$consumo.pre[5]

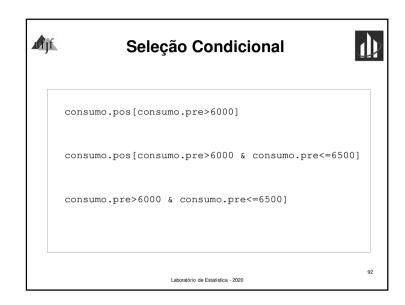
- Todas as medidas da mulher #5: $\sqrt{\text{dados.altman}[5,]}$
- Consumos de energia pós-menstrual 3 primeiras mulheres

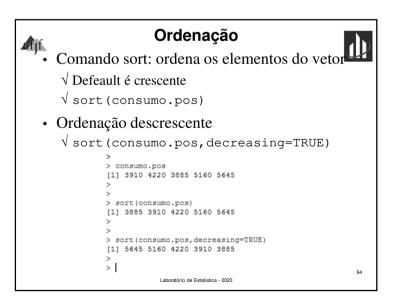
 $\sqrt{\text{dados.altman}[1:3,2]}$

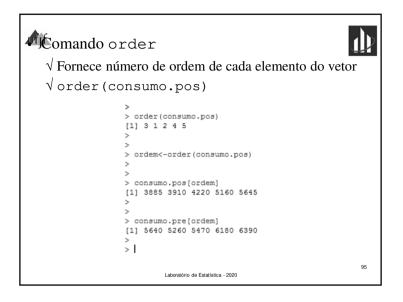
Laboratório de Estatística - 2020

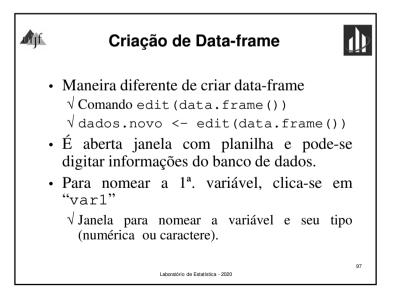
90

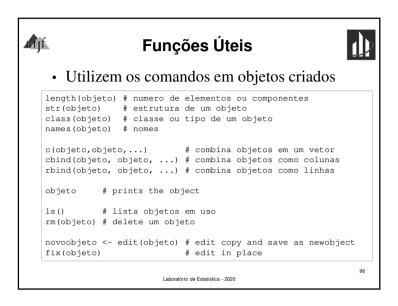
```
> consumo.pos[consumo.pre>6000]
[1] 5160 5645
> consumo.pos[consumo.pre>6000 & consumo.pos <=6500]
[1] 5160 5645
> consumo.pre>6000 & consumo.pos <=6500
[1] FALSE FALSE FALSE TRUE TRUE
> condicao=consumo.pre>6000 & consumo.pos <=6500
> consumo.pos[condicao]
[1] 5160 5645
> |
```

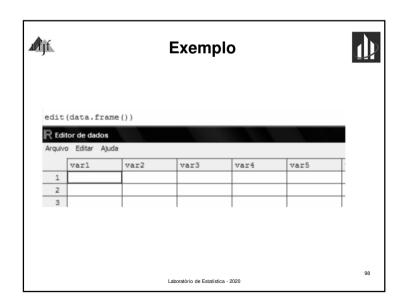














Carregando Pacotes



- · Pacotes:
 - √ Conjuntos de funções específicas do R
 - √ No repositório do R está armazenada uma quantidade muito grande de pacotes que tem funções para um certo conjunto de tarefas
 - $\sqrt{\text{Para usar um pacote:}}$
 - baixar o pacote (download) do repositório
 - carregar o pacote na sua área de trabalho.

Laboratório de Estatística - 2020

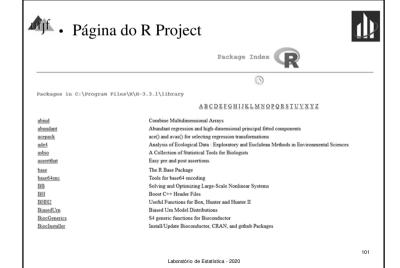




Pacotes



- · Quais pacotes estão disponíveis na sua instalação de R?
 - √library()
- Interface hipertexto de ajuda:
 - √help.start()
 - √ Escolher o link "Packages"
 - $\sqrt{\text{Clique no nome de um dos pacotes}}$
 - Lista todos os objetos que este pacote contém.





· Quais pacotes estão carregados na sua sessão?



100

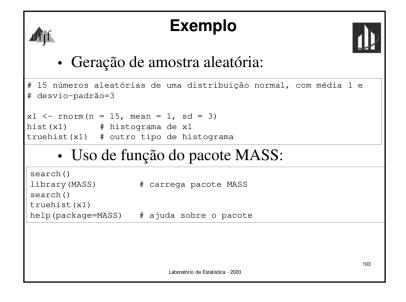
√search()

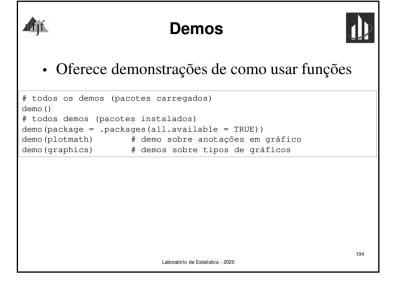
• Instalação de pacote direto do R

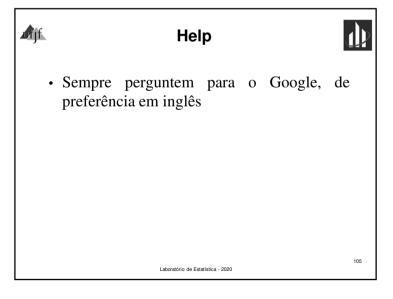
√install.packages("ISwR")

Laboratório de Estatística - 2020

102









Conjuntos de Dados do R

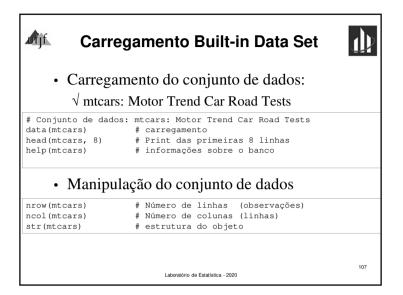


106

- R traz vários conjuntos de dados internos, que são geralmente usados em demos ou exemplos
- Comando para ver a lista dos conjuntos de dados carregados:

√data()

Laboratório de Estatística - 2020



```
• Colunas como vetores
mtcars$mpg
                            # acesso à variavel mpg
mtcars$disp
                            # acesso à variavel disp
pairs(mtcars, main = "Conjunto mtcars")
plot (mpg, disp)
# acesso às variáveis sem especificar banco
attach (mtcars)
plot(mpg, disp)
                            # plot das variáveis mpg e disp
objeto <- lm(disp ~ mpg)
                         # regressão linear
objeto
                            # resultados
summary (objeto)
                            # ajuste do modelo de regressão
abline(lm(disp ~ mpg))  # plot da reta de regressão
# estratificação do plot pela variável cyl
coplot(mpg ~ disp | as.factor(cyl), data = mtcars,
      panel = panel.smooth, rows = 1)
dettach (mtcars)
                    # interromper acesso (final do trabalho)
                                                                 108
                           Laboratório de Estatística - 2020
```

Prof. Lupércio F. Bessegato - UFJF