




Introdução ao uso do



em Ciências da Saúde

3. Leitura, limpeza e manejo de dados b. Limpeza e manejo de dados

Fúlvio Borges Nedel

Departamento de Saúde Pública – SPB
Centro de Ciências da Saúde – CCS
Universidade Federal de Santa Catarina – UFSC

Grups de Recerca d'Amèrica i Àfrica Llatines – GRAAL
<http://graal.uab.cat>

19 de dezembro de 2017



Introdução ao R
3.b Manejo de
dados
2/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

1 Introdução

2 Modificar a classe de variáveis

3 Definir *missings*



- 1 Verificação de inconsistências
- 2 Definição de *missings* e análise de sua ocorrência: algumas variáveis são particularmente afetadas?
- 3 Criação de novas variáveis
 - computação a partir de outras (idade, IMC)
 - recodificação (estado nutricional, faixa etária, ABEP)
- 4 Criar ou redefinir rótulos de variáveis e categorias



- 1 Verificação de inconsistências
- 2 Definição de *missings* e análise de sua ocorrência: algumas variáveis são particularmente afetadas?
- 3 Criação de novas variáveis
 - computação a partir de outras (idade, IMC)
 - recodificação (estado nutricional, faixa etária, ABEP)
- 4 Criar ou redefinir rótulos de variáveis e categorias

Abra o arquivo de sintaxe criado anteriormente e adicione uma nova linha com um comentário do tipo:

```
# Limpeza e manejo de dados  
# -----  
...
```



Listar os arquivos .RData no diretório

```
dir(patt='.rdata', ignore.case = TRUE)
[1] "cursoR.RData"    "usuarios.Rdata"
# Para tornar a busca "case-insensitive", use o argumento
# ignore.case = T
```

Carregar o arquivo

```
load("cursoR.RData")
```

Verificar sua presença no espaço de trabalho

```
ls()
[1] "cursoR"
```



Observar o banco



Introdução ao R
3.b Manejo de
dados
5/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
head(cursorR) # observar os primeiros registros
```

	peso	altura	sexo	dataentr	datanasc	abepcls	grupohas	grupodm
1	64	1.44	Feminino	2011-03-19	1932-07-08	C2	Não	<NA>
2	50	1.5	Feminino	2011-03-19	1951-11-10	C1	Não	Não
3	69	1.6	Feminino	2011-03-20	1947-11-14	C2	Não	<NA>
4	75	1.6	Feminino	2011-03-20	1930-03-09	C2	Não	Não
5	60	1.58	Feminino	2011-03-20	1960-08-13	B1	Não	<NA>
6	73	1.6	Feminino	2011-03-20	1942-04-15	D	Não	<NA>

```
tail(cursorR) # observar os últimos registros
```

	peso	altura	sexo	dataentr	datanasc	abepcls	grupohas	grupodm
295	68	1.72	Masculino	2011-03-13	1936-08-26	C1	Não	<NA>
296	75	1.66	Feminino	2011-03-12	1955-07-21	C1	Não	<NA>
297	70	1.65	Feminino	2011-03-12	1936-06-28	C1	Não	<NA>
298	57	1.48	Feminino	2011-03-14	1934-03-25	B2	Não	<NA>
299	70	1.6	Feminino	2011-03-14	1942-11-12	<NA>	Não	<NA>
300	90	1.62	Feminino	2011-03-14	1958-05-11	C2	Não	<NA>



A estrutura do objeto 'cursoR'



Introdução ao R
3.b Manejo de
dados
6/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
str(cursoR)
```

```
'data.frame': 300 obs. of 8 variables:
 $ peso      : Factor w/ 74 levels "40.5","44.5",...: 21 6 30 37 17 35 28 1
 ..- attr(*, "label")= Named chr "u47. Qual o seu peso?"
 .. ..- attr(*, "names")= chr "peso"
 $ altura    : Factor w/ 46 levels "1.3","1.36","1.4",...: 4 9 19 19 17 19
 ..- attr(*, "label")= Named chr "u48. Qual a sua altura?"
 .. ..- attr(*, "names")= chr "altura"
 $ sexo      : Factor w/ 2 levels "Feminino ","Masculino": 1 1 1 1 1 1 1 1
 ..- attr(*, "label")= Named chr "u8. Sexo:"
 .. ..- attr(*, "names")= chr "sexo"
 $ dataentr  : Date, format: "2011-03-19" "2011-03-19" ...
 $ datanasc  : Date, format: "1932-07-08" "1951-11-10" ...
 $ abepcls   : Factor w/ 7 levels "A2","B1","B2",...: 5 4 5 5 2 6 4 2 4 6
 ..- attr(*, "label")= Named chr "Classificação socioeconômica ABEP mod
 .. ..- attr(*, "names")= chr "abepcls"
 $ grupohas  : Factor w/ 2 levels "Não","Sim": 1 1 1 1 1 1 1 1 1 1
 ..- attr(*, "label")= Named chr "u53. Desde <6 MESES ATRÁS> o(a) Sr.(a)
 .. ..- attr(*, "names")= chr "grupohas"
 $ grupodm   : Factor w/ 2 levels "Não","Sim": NA 1 NA 1 NA NA 1 NA NA NA
 ..- attr(*, "label")= Named chr "u63. Desde <6 MESES ATRÁS> o(a) Sr.(a)
 .. ..- attr(*, "names")= chr "grupodm"
```



A estrutura do objeto 'cursoR'



PROBLEMA

```
str(cursoR)
```

```
'data.frame': 300 obs. of 8 variables:
```

```
$ peso : Factor w/ 74 levels "40.5","44.5",...: 21 6 30 37 17 35 28 1
```

```
..- attr(*, "label")= Named chr "u47. Qual o seu peso?"
```

```
.. ..- attr(*, "names")= chr "peso"
```

```
$ altura : Factor w/ 46 levels "1.3","1.36","1.4",...: 4 9 19 19 17 19
```

```
..- attr(*, "label")= Named chr "u48. Qual a sua altura?"
```

```
.. ..- attr(*, "names")= chr "altura"
```

```
$ sexo : Factor w/ 2 levels "Feminino ","Masculino": 1 1 1 1 1 1 1 1
```

```
..- attr(*, "label")= Named chr "u8. Sexo:"
```

```
.. ..- attr(*, "names")= chr "sexo"
```

```
$ dataentr: Date, format: "2011-03-19" "2011-03-19" ...
```

```
$ datanasc: Date, format: "1932-07-08" "1951-11-10" ...
```

```
$ abepcls : Factor w/ 7 levels "A2","B1","B2",...: 5 4 5 5 2 6 4 2 4 6
```

```
..- attr(*, "label")= Named chr "Classificação socioeconômica ABEP mod
```

```
.. ..- attr(*, "names")= chr "abepcls"
```

```
$ grupohas: Factor w/ 2 levels "Não","Sim": 1 1 1 1 1 1 1 1 1 1
```

```
..- attr(*, "label")= Named chr "u53. Desde <6 MESES ATRÁS> o(a) Sr.(a)
```

```
.. ..- attr(*, "names")= chr "grupohas"
```

```
$ grupodm : Factor w/ 2 levels "Não","Sim": NA 1 NA 1 NA NA 1 NA NA NA
```

```
..- attr(*, "label")= Named chr "u63. Desde <6 MESES ATRÁS> o(a) Sr.(a)
```

```
.. ..- attr(*, "names")= chr "grupodm"
```




Introdução ao R
3.b Manejo de
dados
7/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

1 Introdução

2 Modificar a classe de variáveis

3 Definir *missings*



Corrigir a classe de peso e altura

```
str(cursorR[1:2])
```

```
'data.frame': 300 obs. of 2 variables:
```

```
$ peso : Factor w/ 74 levels "40.5","44.5",...: 21 6 30 37 17 35 28 16
```

```
..- attr(*, "label")= Named chr "u47. Qual o seu peso?"
```

```
.. ..- attr(*, "names")= chr "peso"
```

```
$ altura: Factor w/ 46 levels "1.3","1.36","1.4",...: 4 9 19 19 17 19 19
```

```
..- attr(*, "label")= Named chr "u48. Qual a sua altura?"
```

```
.. ..- attr(*, "names")= chr "altura"
```

```
cursorR$peso = as.numeric(as.character(cursorR$peso))
```

```
cursorR$altura = as.numeric(as.character(cursorR$altura))
```

```
str(cursorR[1:2])
```

```
'data.frame': 300 obs. of 2 variables:
```

```
$ peso : num 64 50 69 75 60 73 68 59.5 65 67 ...
```

```
$ altura: num 1.44 1.5 1.6 1.6 1.58 1.6 1.6 1.6 1.6 1.72 1.68 ...
```

Note que

- é preciso primeiro converter o factor em character
- ao fazer a conversão, perde-se o rótulo da variável
- para mantê-lo, guarde-o antes e depois o destine novamente à variável



Corrigir a classe de peso e altura

```
load("cursoR.RData")
str(cursoR[1:2])

'data.frame': 300 obs. of 2 variables:
 $ peso : Factor w/ 74 levels "40.5","44.5",...: 21 6 30 37 17 35 28 16 23 26 ...
 ..- attr(*, "label")= Named chr "u47. Qual o seu peso?"
 .. ..- attr(*, "names")= chr "peso"
 $ altura: Factor w/ 46 levels "1.3","1.36","1.4",...: 4 9 19 19 17 19 19 19 32 28
 ..- attr(*, "label")= Named chr "u48. Qual a sua altura?"
 .. ..- attr(*, "names")= chr "altura"

rotulopeso = attributes(cursoR$peso)$label
rotuloaltura = attr(cursoR$altura, "label")

cursoR$peso = as.numeric(as.character(cursoR$peso))
cursoR$altura = as.numeric(as.character(cursoR$altura))

attr(cursoR$peso, "label") <- rotulopeso
attributes(cursoR$altura)$label <- rotuloaltura

str(cursoR[1:2])

'data.frame': 300 obs. of 2 variables:
 $ peso : atomic 64 50 69 75 60 73 68 59.5 65 67 ...
 ..- attr(*, "label")= Named chr "u47. Qual o seu peso?"
 .. ..- attr(*, "names")= chr "peso"
 $ altura: atomic 1.44 1.5 1.6 1.6 1.58 1.6 1.6 1.6 1.72 1.68 ...
 ..- attr(*, "label")= Named chr "u48. Qual a sua altura?"
 .. ..- attr(*, "names")= chr "altura"
```



Veja também a função **describe**, do pacote **Hmisc**

summary(cursoR)

peso	altura	sexo	dataentr
Min. : 40.50	Min. :1.30	Feminino :198	Min. :2011-03-12
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102	1st Qu.:2011-03-15
Median : 72.00	Median :1.63		Median :2011-03-18
Mean : 76.93	Mean :1.63		Mean :2011-03-17
3rd Qu.: 83.00	3rd Qu.:1.70		3rd Qu.:2011-03-19
Max. :999.00	Max. :1.92		Max. :2011-03-20

datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08

grupohas	grupodm
Não :271	Não : 88
Sim : 10	Sim : 7
NA's: 19	NA's:205

summary(cursoR\$abepcls)

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Um breve sumário do banco



Veja também a função **describe**, do pacote **Hmisc**

Introdução ao R
3.b Manejo de
dados
9/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
summary(cursorR)
```

peso	altura	sexo	dataentr
Min. : 40.50	Min. :1.30	Feminino :198	Min. :2011-03-12
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102	1st Qu.:2011-03-15
Median : 72.00	Median :1.63		Median :2011-03-18
Mean : 76.93	Mean :1.63		Mean :2011-03-17
3rd Qu.: 83.00	3rd Qu.:1.70		3rd Qu.:2011-03-19
Max. :999.00	Max. :1.92		Max. :2011-03-20

```

datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08

```

```

grupohas grupodm
Não :271 Não : 88
Sim : 10 Sim : 7
NA's: 19 NA's:205

```

```
summary(cursorR$abepcls)
```

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Um breve sumário do banco



Veja também a função **describe**, do pacote **Hmisc**

Introdução ao R
3.b Manejo de
dados
9/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
summary(cursorR)
```

peso	altura	sexo	dataentr
Min. : 40.50	Min. :1.30	Feminino :198	Min. :2011-03-12
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102	1st Qu.:2011-03-15
Median : 72.00	Median :1.63		Median :2011-03-18
Mean : 76.93	Mean :1.63		Mean :2011-03-17
3rd Qu.: 83.00	3rd Qu.:1.70		3rd Qu.:2011-03-19
Max. :999.00	Max. :1.92		Max. :2011-03-20

```

datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08

```

```

grupohas grupodm
Não :271 Não : 88
Sim : 10 Sim : 7
NA's: 19 NA's:205

```

```
summary(cursorR$sabepcls)
```

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Um breve sumário do banco



Veja também a função **describe**, do pacote **Hmisc**

Introdução ao R
3.b Manejo de
dados
9/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
summary(cursorR)
```

peso	altura	sexo	dataentr
Min. : 40.50	Min. :1.30	Feminino :198	Min. :2011-03-12
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102	1st Qu.:2011-03-15
Median : 72.00	Median :1.63		Median :2011-03-18
Mean : 76.93	Mean :1.63		Mean :2011-03-17
3rd Qu.: 83.00	3rd Qu.:1.70		3rd Qu.:2011-03-19
Max. :999.00	Max. :1.92		Max. :2011-03-20

```

datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08

```

```

grupohas grupodm
Não :271 Não : 88
Sim : 10 Sim : 7
NA's: 19 NA's:205

```

```
summary(cursorR$sabepcls)
```

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Um breve sumário do banco



Veja também a função **describe**, do pacote **Hmisc**

Introdução ao R
3.b Manejo de
dados
9/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
summary(cursorR)
```

peso	altura	sexo
Min. : 40.50	Min. :1.30	Feminino :198
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102
Median : 72.00	Median :1.63	
Mean : 76.93	Mean :1.63	
3rd Qu.: 83.00	3rd Qu.:1.70	
Max. :999.00	Max. :1.92	

dataentr
Min. :2011-03-12
1st Qu.:2011-03-15
Median :2011-03-18
Mean :2011-03-17
3rd Qu.:2011-03-19
Max. :2011-03-20

datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08

grupohas	grupodm
Não :271	Não : 88
Sim : 10	Sim : 7
NA's: 19	NA's:205

```
summary(cursorR$sabepcls)
```

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Um breve sumário do banco



Veja também a função **describe**, do pacote **Hmisc**

Introdução ao R
3.b Manejo de
dados
9/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
summary(cursorR)
```

peso	altura	sexo	dataentr
Min. : 40.50	Min. :1.30	Feminino :198	Min. :2011-03-12
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102	1st Qu.:2011-03-15
Median : 72.00	Median :1.63		Median :2011-03-18
Mean : 76.93	Mean :1.63		Mean :2011-03-17
3rd Qu.: 83.00	3rd Qu.:1.70		3rd Qu.:2011-03-19
Max. :999.00	Max. :1.92		Max. :2011-03-20

```
datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08
```

```
grupohas grupodm
Não :271 Não : 88
Sim : 10 Sim : 7
NA's: 19 NA's:205
```

```
summary(cursorR$abepcls)
```

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Um breve sumário do banco



Veja também a função **describe**, do pacote **Hmisc**

Introdução ao R
3.b Manejo de
dados
9/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
summary(cursorR)
```

peso	altura	sexo	dataentr
Min. : 40.50	Min. :1.30	Feminino :198	Min. :2011-03-12
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102	1st Qu.:2011-03-15
Median : 72.00	Median :1.63		Median :2011-03-18
Mean : 76.93	Mean :1.63		Mean :2011-03-17
3rd Qu.: 83.00	3rd Qu.:1.70		3rd Qu.:2011-03-19
Max. :999.00	Max. :1.92		Max. :2011-03-20

```
datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08
```

```
grupohas grupodm
Não :271 Não : 88
Sim : 10 Sim : 7
NA's: 19 NA's:205
```

```
summary(cursorR$abepcls)
```

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Um breve sumário do banco



Veja também a função **describe**, do pacote **Hmisc**

Introdução ao R
3.b Manejo de
dados
9/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
summary(cursorR)
```

peso	altura	sexo	dataentr
Min. : 40.50	Min. :1.30	Feminino :198	Min. :2011-03-12
1st Qu.: 64.00	1st Qu.:1.57	Masculino:102	1st Qu.:2011-03-15
Median : 72.00	Median :1.63		Median :2011-03-18
Mean : 76.93	Mean :1.63		Mean :2011-03-17
3rd Qu.: 83.00	3rd Qu.:1.70		3rd Qu.:2011-03-19
Max. :999.00	Max. :1.92		Max. :2011-03-20

```

datanasc
Min. :1916-05-30
1st Qu.:1936-10-18
Median :1946-04-09
Mean :1946-08-23
3rd Qu.:1955-03-25
Max. :1988-01-08

```

```

grupohas grupodm
Não :271 Não : 88
Sim : 10 Sim : 7
NA's: 19 NA's:205

```

```
summary(cursorR$sabepcls)
```

A2	B1	B2	C1	C2	D	E	NA's
1	11	61	102	76	30	1	18



Introdução ao R
3.b Manejo de
dados
10/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

1 Introdução

2 Modificar a classe de variáveis

3 Definir *missings*



A variável **peso** tem um valor máximo de **999 Kg**, o que não é um valor válido, foi um código para indicar a falta de informação.

Vamos definir esses valores como missings (**NA**):

```
cursoR$peso[cursoR$peso==999] <- NA  
summary(cursoR$peso)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
40.50	64.00	72.00	73.85	83.00	126.00	1

A função **attach**

Podemos evitar o trabalho de chamar o nome do data frame antes de cada variável. **Veja as funções *attach* e *detach*. Use com moderação!**

```
attach(cursoR)
```



Um caso mais complexo

- As variáveis `grupodm` e `grupohas` referem-se à participação em grupos de promoção da saúde para pessoas com, respectivamente, DM ou HAS.
- Se a pessoa não tem DM, a resposta para `grupodm` será um missing e assim para HAS e `grupohas`.
- A definição de missings nesses casos é um pouco mais complexa. Em seguida apresentam-se três formas de realizá-la:

As combinações possíveis

	grupohas	grupodm
1	Sim	Sim
2	Sim	<NA>
3	<NA>	Sim
4	Não	Não
5	Não	<NA>
6	<NA>	Não
7	<NA>	<NA>



Participa em grupos de hipertensos *ou* diabéticos?

```
grupo = rep(NA, 300)
grupo[grupodm == 'Sim' | grupohas == 'Sim'] = "Sim"
grupo[(grupodm == 'Não' & is.na(grupohas))] = "Não"
grupo[(grupohas == 'Não' & is.na(grupodm))] = "Não"
grupo[grupodm == 'Não' & grupohas == 'Não'] = "Não"
table(grupo, useNA = 'ifany')

grupo
  Não  Sim <NA>
285   13    2
```



Participa em grupos de hipertensos *ou* diabéticos?

```
grupo = rep(NA, 300)
grupo[grupodm == 'Sim' | grupohas == 'Sim'] = "Sim"
grupo[(grupodm == 'Não' & is.na(grupohas))] = "Não"
grupo[(grupohas == 'Não' & is.na(grupodm))] = "Não"
grupo[grupodm == 'Não' & grupohas == 'Não'] = "Não"
table(grupo, useNA = 'ifany')
```

```
grupo
Não Sim <NA>
285 13 2
```

```
str(grupo)
chr [1:300] "Não" "Não" "Não" "Não" "Não" "Não" "Não" "Não" "Não" ...
grupo <- factor(grupo, levels = c("Sim", "Não"),
               labels = c("Sim", "Não"))
```

```
str(grupo)
Factor w/ 2 levels "Sim","Não": 2 2 2 2 2 2 2 2 2 2 ...
table(grupo, useNA = 'ifany')
```

```
grupo
Sim Não <NA>
13 285 2
```




Definição de *missings*

Um caso mais complexo



Introdução ao R
3.b Manejo de
dados
14/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

```
grupo2 = rep("Não", 300)
grupo2[grupodm == 'Sim' | grupohas == 'Sim'] = "Sim"
grupo2[which(is.na(grupodm) & is.na(grupohas))] = NA
table(grupo2, useNA = 'ifany')

grupo2
  Não  Sim <NA>
285   13     2

#
grupo3 = grupohas
grupo3[grupodm == 'Sim' | grupohas == 'Sim'] = "Sim"
grupo3[is.na(grupohas)] = grupodm[is.na(grupohas)]
table(grupo3, useNA = 'ifany')

grupo3
  Não  Sim <NA>
285   13     2
```



Definição de *missings*

Um caso mais complexo



Introdução ao R
3.b Manejo de
dados
15/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

Podemos comprovar que o resultado é o desejado comparando as variáveis em todos os casos possíveis:

```
# Comprovar
```

```
x = cbind(cursor[c('grupohas', 'grupodm')], " " = '-->',  
           grupo, grupo2, grupo3)
```

```
x[c(77,11,23,2,1,21,25), ]
```

	grupohas	grupodm	grupo	grupo2	grupo3
77	Sim	Sim	-->	Sim	Sim
11	Sim	<NA>	-->	Sim	Sim
23	<NA>	Sim	-->	Sim	Sim
2	Não	Não	-->	Não	Não
1	Não	<NA>	-->	Não	Não
21	<NA>	Não	-->	Não	Não
25	<NA>	<NA>	-->	<NA>	<NA>



Definição de *missings*

Um caso mais complexo



Introdução ao R
3.b Manejo de
dados
15/16

Fúlvio Nedel
SPB/UFSC

Introdução

Modificar a classe
de variáveis

Definir *missings*

Podemos comprovar que o resultado é o desejado comparando as variáveis em todos os casos possíveis:

```
# Comprovar
x = cbind(cursoR[c('grupohas', 'grupodm')], " " = '-->',
          grupo, grupo2, grupo3)
x[c(77,11,23,2,1,21,25), ]
```

	grupohas	grupodm	grupo	grupo2	grupo3
77	Sim	Sim	-->	Sim	Sim
11	Sim	<NA>	-->	Sim	Sim
23	<NA>	Sim	-->	Sim	Sim
2	Não	Não	-->	Não	Não
1	Não	<NA>	-->	Não	Não
21	<NA>	Não	-->	Não	Não
25	<NA>	<NA>	-->	<NA>	<NA>

Para encontrar os casos, usei as funções *which* e *is.na*

```
which(x$grupohas == 'Sim' & x$grupodm == 'Sim')[1]
```

```
[1] 77
```

```
which(x$grupohas == 'Sim' & is.na(x$grupodm))[1]
```

```
[1] 11
```

etc.



Guardar as modificações para continuar mais tarde

```
detach(cursoR) # "detachar" o banco
```

```
ls() # listar os objetos da sessão de trabalho
```

```
[1] "cursoR"          "grupo"           "grupo2"          "grupo3"  
[5] "rotuloaltura"    "rotulopeso"      "x"
```

```
# guardar só o que interessa
```

```
save(cursoR, grupo, file="cursoR.RData")
```