

# Data Frames

- Os *data frames* são objectos usados para guardar tabelas com dados sobre um problema.
  - Uma tabela de dados é formada por um conjunto linhas cada qual representando uma observação (um caso).
  - Cada observação é descrita por um conjunto de propriedades (atributos)
    - Um data frame é algo muito parecido com uma tabela do Access.
- Os data frames criam-se indicando ao R o nome de cada coluna da tabela de dados e respectivo conteúdo.

```

> notas.inform <- data.frame(
+   nros=c(2355,3456,2334,5456),
+   turma=c('tp1','tp1','tp2','tp3'),
+   notas = c(10.3,9.3,14.2,15))
> notas.inform
  nros turma notas
1 2355  tp1  10.3
2 3456  tp1   9.3
3 2334  tp2  14.2
4 5456  tp3  15.0
> notas.inform[2,2]
[1] tp1
Levels: tp1 tp2 tp3
> notas.inform[4,'turma']
[1] tp3
Levels: tp1 tp2 tp3
> notas.inform[2,]
  nros turma notas
2 3456  tp1   9.3
> notas.inform[, 'nros']
[1] 2355 3456 2334 5456

```

A diferença fundamental entre um data frame e uma matriz é que num data frame os dados não precisam de ser todos do mesmo tipo.

Notar que a função `data.frame()` transformou a coluna “turma” num factor.

Existe uma alternativa para aceder ao conteúdo de uma coluna:

```

> notas.inform$nros
[1] 2355 3456 2334 5456

```

# Exemplos de indexação em data frames

```
> notas.inform[notas.inform$notas > 10,]
```

```
  nros turma notas
1 2355   tp1  10.3
3 2334   tp2  14.2
4 5456   tp3  15.0
```

```
> notas.inform[notas.inform$notas > 14, 'nros']
```

```
[1] 2334 5456
```

```
> notas.inform[notas.inform$turma == 'tp1', c('nros', 'notas')]
```

```
  nros notas
1 2355  10.3
2 3456   9.3
```

Operadores de comparação:

>, <, >=, <=, ==, !=

# Simplificação do acesso às colunas

```
> attach(notas.inform)
```

```
> notas.inform[notas > 14,]
```

```
  nros turma notas
```

```
3 2334   tp2  14.2
```

```
4 5456   tp3  15.0
```

```
> turma
```

```
[1] tp1 tp1 tp2 tp3
```

```
Levels: tp1 tp2 tp3
```

```
> detach(notas.inform)
```

```
> turma
```

```
Error: Object "turma" not found
```

A função `attach()` permite aceder directamente às colunas sem ter que usar a sintaxe:

```
notas.inform$notas
```

A função `detach()` retira esse acesso directo

- É possível acrescentar novas colunas a um data frame existente:

```
> notas.inform$resultado <- c('aprovado','oral','aprovado','reprovado')
```

```
> notas.inform
```

```
  nros turma notas resultado
1 2355  tp1  10.3  aprovado
2 3456  tp1   9.3     oral
3 2334  tp2  14.2  aprovado
4 5456  tp3  15.0 reprovado
```

### Restrição:

A nova coluna tem que ter tantos valores quantas as linhas do data frame.

- O mesmo se passa com novas linhas:

```
> notas.inform[5,] <- c(5466,'tp2',5.6,'oral')
```

```
> notas.inform
```

```
  nros turma notas resultado
1 2355  tp1  10.3  aprovado
2 3456  tp1   9.3     oral
3 2334  tp2  14.2  aprovado
4 5456  tp3   15  reprovado
5 5466  tp2   5.6     oral
```

### Restrição:

Não se podem usar valores novos para as colunas que são factores.

- Podem-se remover colunas ou linhas de um data frame tirando partido das capacidades de indexação do R, e do facto de a atribuição ser destrutiva, como já foi referido:

```
> notas.inform <- notas.inform[, -4]
```

```
> notas.inform
```

```
  nros turma notas
1 2355   tp1  10.3
2 3456   tp1   9.3
3 2334   tp2  14.2
4 5456   tp3   15
5 5466   tp2   5.6
```

```
> notas.inform <- notas.inform[-5,]
```

```
> notas.inform
```

```
  nros turma notas
1 2355   tp1  10.3
2 3456   tp1   9.3
3 2334   tp2  14.2
4 5456   tp3   15
```

- As seguintes funções dizem-nos quantas linhas e colunas tem um data frame:

```
> nrow(notas.inform)
```

```
[1] 3
```

```
> ncol(notas.inform)
```

```
[1] 4
```

- Para alterar alguns valores particulares de um data frame pode ser mais fácil usar:

```
> notas.inform <- edit(notas.inform)
```

- Isto também pode ser usado para criar um novo data frame:

```
> novas.notas <- edit(data.frame())
```

	nros	turma	notas	var4	var5
1	2355	tp1	10.3		
2	3456	tp1	9.3		
3	2334	tp2	13.5		
4	5456	tp3	15		
5	43556	tp2	12.6		
6	5666	tp4	12.4		
7					
8					

- A função `summary()` pode ser usada para obter alguma informação estatística sobre os dados armazenados num data frame:

```
> summary(notas.inform)
```

nros		turma	notas	
Min.	:2334	tp1:2	Min.	: 9.30
1st Qu.	:2350	tp2:1	1st Qu.	:10.05
Median	:2906	tp3:1	Median	:12.25
Mean	:3400		Mean	:12.20
3rd Qu.	:3956		3rd Qu.	:14.40
Max.	:5456		Max.	:15.00



# Conjuntos de dados disponíveis no R

- O R vem com um conjunto grande de dados que podem ser usados para praticar algumas das noções estudadas.
- Para obter uma lista dos conjuntos de dados disponíveis:  
> `data()`
- Pode-se obter mais informação específica sobre um conjunto em particular fazendo:  
> `? longley`
- Para usar um dos conjuntos de dados:  
> `data(longley)`
- Isto cria um data frame chamado “longley” com os dados  
> `summary(longley)`