

Salvando o banco de dados

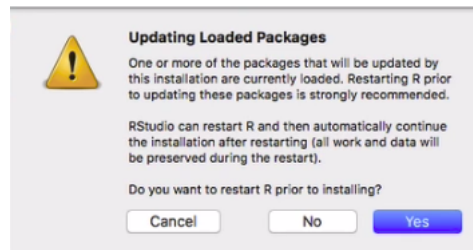
Transcrição

Precisamos salvar a informação sobre os cursos para utilizá-la adiante, já que anteriormente definimos que faríamos uma correlação da popularidade dos cursos e o tempo que eles levam para serem concluídos. Se fecharmos o programa agora, perderemos a tabela e os dados, que ainda não salvamos. Para isto, criaremos um **novo banco de dados** a partir de um objeto, que será uma contagem desses valores.

Utilizaremos a função `count()` para essa contagem, mas ela não está pré instalada no programa. Para utilizá-la, precisamos instalar uma biblioteca chamada `plyr`, que a contém:

```
install.packages("plyr")
```

Reparem que `plyr` foi inserida entre aspas e parênteses. Ao executá-lo, caso já tenham a biblioteca, receberão um aviso de que ela já está instalada na máquina, perguntando se desejamos fazer uma atualização dela.



Se ela não estiver instalada na máquina de vocês, o programa procederá a instalação, que demora alguns segundos e, uma vez feita, não será necessário instalar novamente. Agora, precisaremos carregar a biblioteca ao iniciarmos cada sessão, para informarmos ao programa que ela será utilizada. Para isso, utilizaremos a função `library()` e, entre parênteses, especificaremos `plyr`, sem aspas:

```
library(plyr)
```

Ao executarmos, a biblioteca será carregada, e pode-se usar a função de contagem. Então, criaremos um objeto para salvá-la, chamada de `auxiliar` — se preferirem, podem usar outro nome — e atribuiremos (`<-`) a ele o resultado da função `count()` como valor. O primeiro parâmetro dela será `aulas` e o segundo será a variável que queremos contar, `course_id`, que especificaremos em `vars`.

```
auxiliar <- count(aulas, vars = "course_id")
```

Assim, no comando, especificaremos que a contagem da variável `course_id`, do banco de dados `aulas`, deve ser atribuída ao objeto `auxiliar`. Ao executarmos, verificaremos que no canto superior direito do RStudio foi criado o objeto `auxiliar`. Se clicarmos nele, a visualização abrirá à esquerda:

The screenshot shows the RStudio interface. The top-left pane displays a data table with columns `course_id` and `freq`. The table contains 7 rows of data, with the first row being (1, 66) and the last row being (7, 170). The top-right pane shows the Environment pane with two objects: `aulas` (34829 obs. of 3 variables) and `auxiliar` (265 obs. of 2 variables). The bottom-left pane shows the Console with the following output:

```
z'  
Content type 'application/x-gzip' length 818975 bytes (799 KB)  
downloaded 799 KB  
  
tar: Failed to set default locale  
  
The downloaded binary packages are in  
/var/folders/kk/xpngpfkd6t5dwlgsdtdq00580000gn/T//Rtmpc3vgq9/downloaded_packages  
> library(plyr)  
> auxiliar <- count(aulas, vars = "course_id")  
> View(auxiliar)  
>
```

Nela teremos duas colunas: uma com `course_id` e outra com `freq`, de frequência. Ou seja, temos a quantidade de vezes que o curso foi visto. Na aba ao lado da visualização do objeto `auxiliar`, está o arquivo de R Script que estávamos editando. Retornaremos a ele, pois há um passo final. Se desligarmos a máquina agora, perderemos as informações; não temos um banco de dados salvo na máquina.

Precisaremos salvar o objeto `auxiliar` como um banco de dados, então utilizaremos a função `write`, que em inglês significa "escreva", seguida de `csv`, sigla de *comma-separated values* ("valores separados por vírgulas"). Dessa forma, criaremos um arquivo em que cada valor será separado por vírgula, e o objeto `auxiliar` será especificado entre parênteses.

Entre aspas, após `auxiliar` e vírgula (,) nomearemos o objeto como `popularidade.csv`, pois ele guardará a frequência dos cursos, e acrescentaremos a extensão `.csv` a ele:

```
write.csv(auxiliar, "popularidade.csv")
```

Executaremos o comando, e no diretório teremos um arquivo chamado `popularidade.csv`, com o qual trabalharemos a seguir.