

Revisão da aula 3

Eduardo

Em nossa terceira aula vimos como importar conjuntos de dados para o R. Utilizamos a função *read.table* conforme o exemplo a seguir para importar dados de meteorologia obtidos da página do inmet.

```
dados <- read.table("dadosinmet",sep="," , head=TRUE)
head(dados)
```

```
##      codigo_estacao      data hora temp_inst temp_max temp_min umid_inst
## 1          A001 25/09/2016    0    21.7    22.0    21.0      66
## 2          A001 25/09/2016    1    21.4    21.7    20.9      71
## 3          A001 25/09/2016    2    21.0    21.5    21.0      73
## 4          A001 25/09/2016    3    20.7    21.2    20.5      79
## 5          A001 25/09/2016    4    20.9    21.0    20.3      74
## 6          A001 25/09/2016    5    21.3    21.3    20.2      71
##      umid_max umid_min pto_orvalho_inst pto_orvalho_max pto_orvalho_min
## 1          75      66          15.1          16.4          15.1
## 2          72      66          15.9          15.9          15.1
## 3          74      70          16.1          16.2          15.6
## 4          79      71          16.9          16.9          15.6
## 5          80      71          16.0          16.7          15.4
## 6          80      70          15.9          16.6          15.3
##      pressao pressao_max pressao_min vento_direcao vento_vel vento_rajada
## 1      888.4      888.4      888.0          1.8      343          4.6
## 2      888.7      888.8      888.4          1.6      251          6.3
## 3      888.2      888.7      888.2          1.3      305          2.9
## 4      887.7      888.3      887.7          1.3      346          2.7
## 5      887.1      887.7      887.1          1.2      284          2.6
## 6      887.1      887.1      886.7          1.9      260          5.2
##      radiacao precipitacao
## 1      -3.30          0.2
## 2      -3.28          0.0
## 3      -3.11          0.0
## 4      -3.28          0.0
## 5      -2.87          0.0
## 6      -2.58          0.0
```

```
summary(dados)
```

```
##      codigo_estacao      data      hora      temp_inst
## A001:135      25/09/2016:24      Min. : 0      Min. :16.50
##              26/09/2016:24      1st Qu.: 5      1st Qu.:19.30
##              27/09/2016:24      Median :11      Median :21.30
##              28/09/2016:24      Mean :11      Mean :22.26
##              29/09/2016:24      3rd Qu.:17      3rd Qu.:25.35
##              30/09/2016:15      Max. :23      Max. :29.50
##      temp_max      temp_min      umid_inst      umid_max
## Min. :17.20      Min. :16.00      Min. :24.00      Min. :29.00
## 1st Qu.:20.05      1st Qu.:18.70      1st Qu.:47.50      1st Qu.:53.50
## Median :21.80      Median :20.50      Median :66.00      Median :71.00
```

```
## Mean :23.07 Mean :21.42 Mean :63.21 Mean :67.76
## 3rd Qu.:26.40 3rd Qu.:24.50 3rd Qu.:79.00 3rd Qu.:82.50
## Max. :30.10 Max. :28.70 Max. :91.00 Max. :91.00
## umid_min pto_orvalho_inst pto_orvalho_max pto_orvalho_min
## Min. :23.00 Min. : 6.5 Min. : 8.40 Min. : 5.90
## 1st Qu.:42.50 1st Qu.:13.6 1st Qu.:14.30 1st Qu.:12.45
## Median :62.00 Median :14.6 Median :15.10 Median :13.90
## Mean :59.16 Mean :14.2 Mean :14.94 Mean :13.51
## 3rd Qu.:72.50 3rd Qu.:15.4 3rd Qu.:16.00 3rd Qu.:14.90
## Max. :90.00 Max. :17.2 Max. :17.90 Max. :16.80
## pressao pressao_max pressao_min vento_direcao
## Min. :883.8 Min. :883.9 Min. :883.7 Min. :0.500
## 1st Qu.:885.7 1st Qu.:886.0 1st Qu.:885.4 1st Qu.:1.500
## Median :886.6 Median :887.0 Median :886.5 Median :1.900
## Mean :886.6 Mean :886.9 Mean :886.4 Mean :2.058
## 3rd Qu.:887.7 3rd Qu.:887.9 3rd Qu.:887.4 3rd Qu.:2.650
## Max. :889.8 Max. :889.9 Max. :889.6 Max. :6.200
## vento_vel vento_rajada radiacao precipitacao
## Min. : 1.0 Min. : 1.400 Min. : -3.540 Min. : 0.0000
## 1st Qu.: 59.0 1st Qu.: 3.250 1st Qu.: -3.525 1st Qu.: 0.0000
## Median :202.0 Median : 4.500 Median : 2.135 Median : 0.0000
## Mean :189.5 Mean : 4.741 Mean : 745.613 Mean : 0.1704
## 3rd Qu.:315.0 3rd Qu.: 5.950 3rd Qu.:1402.500 3rd Qu.: 0.0000
## Max. :359.0 Max. :10.800 Max. :3397.000 Max. :13.8000
```

A função `read.table` foi utilizada com 3 argumentos, o primeiro deles foi o nome do arquivo a ser aberto, o R foi capaz de localizar esse arquivo localmente no computador pois o arquivo está posicionado no mesmo diretório deste arquivo *rmarkdown* de revisão. O segundo argumento (`sep=","`) é referente a tabulação existente no banco de dados que tem vírgulas separando cada uma das observações e o terceiro argumento (`head=TRUE`) é para indicar que o banco de dados tem um cabeçalho em sua primeira linha, que representa os nomes das variáveis.

Existem algumas alternativas para que o R possa localizar o arquivo que se deseja importar. Uma delas seria passar o endereço completo do arquivo em sua máquina conforme o seguinte exemplo que ilustra os diretórios em que estão organizados meu computador até a pasta que está o arquivo de interesse

```
dados <- read.table("/home/eduardo/Dropbox/AAAUNB/Extensao/cursoRsegundo2016/disponibilizar/aula4/dados.csv",
  sep = ",", head = TRUE)
```

Uma outra possibilidade é utilizar a função `file.choose()` que permite que o usuário escolha interativamente o arquivo que deseja importar, navegando e clicando pelas janelas do sistema operacional para encontrar o arquivo de interesse.

```
read.table(file.choose(), sep=",", head=TRUE)
```

Após a importação do arquivo foi utilizado o comando (`head(dados)`) que permite visualizar as primeiras linhas do banco de dados possibilitando ver as diferentes variáveis presentes no banco de dados e seus tipos, conferindo se a importação dos dados foi realizada de forma correta. Também foi ilustrado a utilização do comando (`summary(dados)`) que faz um resumo descritivo das diferentes variáveis presentes no banco de dados.

Além da importação de dados disponíveis localmente no computador que estamos utilizando vimos exemplo de importação de banco de dados diretamente de um endereço de internet.

```
dadosCamaroes <- read.table("https://raw.githubusercontent.com/cursosRunb/cursos_r_2016/master/aula3/dadosCamaroes.csv",
  head = TRUE, dec = ",", sep = " ")
head(dadosCamaroes)
```

```
##           Local Sexo    Ct    Cc    Ca  Peso
## 1 BaiaFormosa    M  94.6  24.8  64.8  9.67
## 2 BaiaFormosa    F 133.0  34.7  87.7 29.32
## 3 BaiaFormosa    F 113.5  32.2  75.6 18.40
## 4 BaiaFormosa    F 120.8  31.2  83.0 21.70
## 5 BaiaFormosa    F 116.4  33.3  80.2 21.60
## 6 BaiaFormosa    M  91.0  23.2  63.7  8.96
```

```
summary(dadosCamaroes)
```

```
##           Local    Sexo           Ct           Cc           Ca
## BaiaFormosa:37  F:85   Min.      : 55.80   Min.      :13.90   Min.      :39.60
## DiogoLopes :32  M:35   1st Qu.: 90.35   1st Qu.:23.18   1st Qu.:64.70
## Touros       :51           Median :102.60   Median :27.05   Median :74.50
##                                     Mean   :102.70   Mean   :27.19   Mean   :73.16
##                                     3rd Qu.:118.00   3rd Qu.:32.00   3rd Qu.:83.55
##                                     Max.    :138.20   Max.    :37.90   Max.    :99.20
##           Peso
## Min.      : 1.980
## 1st Qu.:  9.193
## Median :14.230
## Mean     :15.752
## 3rd Qu.:22.418
## Max.     :35.780
```

E nesse caso a função *read.table* foi passada com o endereço de internet do arquivo de interesse, a indicação que os dados tinham em sua primeira linha o cabeçalho com os nomes das variáveis (*head=TRUE*), a indicação que os números decimais estavam separados por vírgulas no banco de dados (*dec=","*), e finalmente com (*_sep=" "*) a indicação de tabulação que as observações das diferentes variáveis estão separadas por espaços.

É importante visualizar os dados em um editor de textos ou outra ferramenta antes de fazer a importação de dados para o R, para ver a forma em que se encontram os dados e poder determinar os parâmetros que devem ser passados para função *read.table* para que a mesma faça importação levando em conta as particularidades do banco de dados a ser utilizado.