

Acertando as séries do Yahoo! Finance no R

As séries do Yahoo! Finance são invertidas, vão das datas mais recentes para as mais antigas, de forma que é necessário realizar esta inversão antes de começar a manipula-las para não causar erros no cálculo dos retornos. Usualmente esta inversão é realizada no Excel, o que é um desperdício tendo a disposição uma ferramenta como R. Vejamos como realizar esta inversão em algumas linhas de código R.

Carregando a série no R

Baixamos a série de preços mensais da Vale do Rio Doce (VALE5 PNA N1) no arquivo VALE5.monthly.raw.csv. O arquivo do Yahoo! Finance vem no formato CSV separado por vírgulas (,), não no formato pt-BR separado por ponto-e-vírgula (;).

```
# carregar o arquivo com as series
vale5 = read.csv("VALE5.monthly.raw.csv", header = TRUE, stringsAsFactors = FALSE)
str(vale5)
```

```
## 'data.frame':    118 obs. of  7 variables:
## $ Date      : chr  "2012-10-01" "2012-09-03" "2012-08-01" "2012-07-02" ...
## $ Open      : num  35.4 33.4 36.6 39 36 ...
## $ High      : num  37.5 38.6 38 40.7 40.3 ...
## $ Low       : num  34.1 31.7 31.9 33.9 35.6 ...
## $ Close     : num  36 35.3 33.2 36.4 39.2 ...
## $ Volume    : int  17504000 22002200 18063800 17227000 15314600 17371400 17403400
17242700 15497000 16927000 ...
## $ Adj.Close: num  36 35.3 33.2 36.4 39.2 ...
```

```
head(vale5)
```

```
##           Date  Open  High  Low Close  Volume Adj.Close
## 1 2012-10-01 35.38 37.46 34.07 35.95 17504000      35.95
## 2 2012-09-03 33.40 38.63 31.73 35.27 22002200      35.27
## 3 2012-08-01 36.62 38.05 31.95 33.18 18063800      33.18
## 4 2012-07-02 39.05 40.71 33.92 36.40 17227000      36.40
## 5 2012-06-01 35.98 40.32 35.65 39.16 15314600      39.16
## 6 2012-05-01 41.45 42.24 34.51 36.72 17371400      36.72
```

Como podemos observar a série do Yahoo! começa com as datas mais recentes. Ainda é possível notar que a coluna Date vem com tipo chr, ou seja, é uma *string* com formato de data. Vamos converter esta coluna para o tipo Date.

```
# converter para o tipo *Date*
vale5$Date = as.Date(vale5$Date)
str(vale5)
```

```
## 'data.frame': 118 obs. of 7 variables:
## $ Date : Date, format: "2012-10-01" "2012-09-03" ...
## $ Open : num 35.4 33.4 36.6 39 36 ...
## $ High : num 37.5 38.6 38 40.7 40.3 ...
## $ Low : num 34.1 31.7 31.9 33.9 35.6 ...
## $ Close : num 36 35.3 33.2 36.4 39.2 ...
## $ Volume : int 17504000 22002200 18063800 17227000 15314600 17371400 17403400
17242700 15497000 16927000 ...
## $ Adj.Close: num 36 35.3 33.2 36.4 39.2 ...
```

```
head(vale5)
```

```
##           Date  Open  High  Low Close  Volume Adj.Close
## 1 2012-10-01 35.38 37.46 34.07 35.95 17504000      35.95
## 2 2012-09-03 33.40 38.63 31.73 35.27 22002200      35.27
## 3 2012-08-01 36.62 38.05 31.95 33.18 18063800      33.18
## 4 2012-07-02 39.05 40.71 33.92 36.40 17227000      36.40
## 5 2012-06-01 35.98 40.32 35.65 39.16 15314600      39.16
## 6 2012-05-01 41.45 42.24 34.51 36.72 17371400      36.72
```

Para enfim ordenar pelas datas de forma ascendente.

```
# colocar a serie na ordem correta -- data crescente
vale5 = vale5[order(vale5$Date), ]
head(vale5)
```

```
##           Date  Open  High   Low  Close  Volume  Adj.Close
## 118 2003-01-01 24.25 24.40 21.85 22.98  626900    5.54
## 117 2003-02-03 23.12 25.20 22.89 24.50  768500    5.91
## 116 2003-03-05 24.61 24.70 21.62 21.90  998700    5.28
## 115 2003-04-01 21.90 22.25 19.20 19.36 1460600    4.67
## 114 2003-05-02 19.38 21.20 19.21 20.77 1884300    5.01
## 113 2003-06-02 20.75 20.97 19.58 19.75 1095700    4.77
```

Temos portanto um `data.frame` com as séries temporais ordenadas da forma adequada para o cálculo dos retornos.