



Clube do Código 28

Extraindo tabelas de arquivos PDF com o R

Vítor Wilher
Mestre em Economia
analisemacro.com.br

Pacotes e Scripts Externos

```
library(tabulizer)
```

```
library(stargazer)
```

1 Extraindo tabelas de arquivos pdf

Nessa edição do [Clube do Código](#), vamos mostrar como extrair tabelas de arquivos PDF com o uso do R. Para isso, vamos fazer uso do pacote `tabulizer` e da função `extract_tables`, aplicando a mesma ao Relatório de Inflação do Banco Central. O código abaixo faz isso.

```
url <- 'http://www.bcb.gov.br/htms/relinf/port/2017/03/ri201703c2p.pdf'
out <- extract_tables(url)
```

2 Limpando e apresentando os dados

Observe que o objeto `out` é uma lista com sete elementos, como se pode ver abaixo.

```
str(out)

## List of 7
## $ : chr [1:6, 1:6] "" "" "" "Cenário do Copom1/" ...
## $ : chr [1:16, 1:13] "Período" "" "" "" ...
## $ : chr [1:11, 1:4] "Período" "2017 I" "2017 II" "2017 III" ...
## $ : chr [1:13, 1:9] "Período" "" "" "2017 1" ...
## $ : chr [1:13, 1:5] "" "Período" "" "" ...
## $ : chr [1:12, 1:8] "Período" "" "" "2017 1 4,6" ...
## $ : chr [1:14, 1:8] "Período" "" "" "2017 1 4,6" ...
```

Vamos, então, pegar a lista 2, que é a tabela 2.3 do Relatório de Inflação. Limpamos os dados que não precisamos e geramos então uma tabela com a função `stargazer`, como no código abaixo.

```
table <- out[[2]][6:16,1:8]
colnames(table) <- c('Trimestre', '50% Lower', '30% Lower',
                    '10% Lower', 'Central',
                    '10% Upper', '30% Upper', '50% Upper')
stargazer(table,
           title='Projeção da inflação no cenário com taxas de juros e câmbio da pesquisa',
           notes='Fonte: Relatório de Inflação BCB - Março de 2017.',
           font.size = 'footnotesize')
```

Tabela 1: Projeção da inflação no cenário com taxas de juros e câmbio da pesquisa Focus

Trimestre	50% Lower	30% Lower	10% Lower	Central	10% Upper	30% Upper	50% Upper
2017 1	4,6	4,6	4,6	4,6	4,6	4,6	4,6
2017 2	3,5	3,7	3,8	3,9	4,0	4,1	4,3
2017 3	2,8	3,1	3,3	3,4	3,5	3,7	4,0
2017 4	3,2	3,6	3,9	4,0	4,1	4,4	4,8
2018 1	3,5	3,9	4,2	4,4	4,6	4,9	5,3
2018 2	3,2	3,7	4,0	4,2	4,4	4,7	5,2
2018 3	3,2	3,7	4,0	4,2	4,4	4,7	5,2
2018 4	3,5	4,0	4,3	4,5	4,7	5,0	5,5
2019 1	3,6	4,1	4,4	4,6	4,8	5,1	5,6

Fonte: Relatório de Inflação BCB – Março de 2017.