

**Vítor Wilher**  
Macroeconometrics & Data Analysis

# Uma introdução ao mundo dos dados e do R

Vítor Wilher  
([www.vitorwilher.com](http://www.vitorwilher.com))

24 de Setembro de 2015

# Plano de Voo

- 1 O mundo é feito de dados
- 2 As quatro operações
  - Capturar
  - Tratar
  - Analisar
  - Apresentar
- 3 Por que usar o R?
- 4 Contato

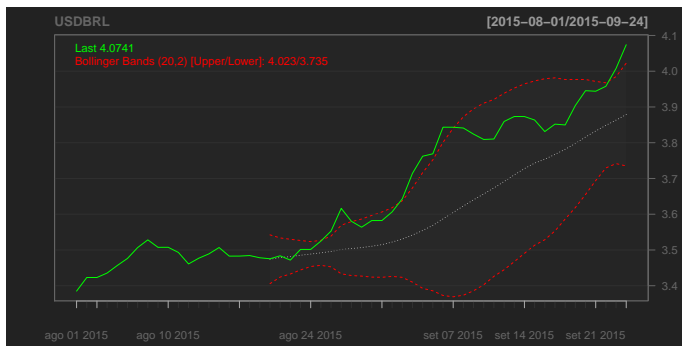
# O mundo é feito de dados

```
library(quantmod)
get <- getFX('USD/BRL', from='2015-08-01')
chart <- chartSeries(USDBRL)
```



# O mundo é feito de dados

```
addBBands ()
```



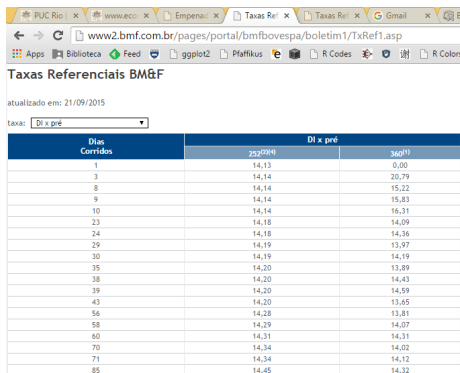
# O mundo é feito de dados

O índice BOVESPA direto do **Yahoo Finance**...



# O mundo é feito de dados

Que tal pegar dados direto da Bovespa?



Atualizado em: 21/09/2015

Taxa:

Dias Corridos	DI x pré	
	252(264)	360(1)
1	14,13	0,00
3	14,14	20,79
8	14,14	15,22
9	14,14	15,83
10	14,14	16,31
23	14,18	14,09
24	14,18	14,36
29	14,19	13,97
30	14,19	14,19
35	14,20	13,89
38	14,20	14,43
39	14,20	14,59
43	14,20	13,65
56	14,28	13,81
58	14,29	14,07
60	14,31	14,31
70	14,34	14,02
71	14,34	14,12
85	14,45	14,32

Figura: Dados da BMF Bovespa

O mundo é feito de dados  
As quatro operações  
Por que usar o R?  
Contato

## O mundo é feito de dados

O Wilson Freitas criou a função `get_curve`, que pega os dados da BMF Bovespa. Depois, é só criar um gráfico com o pacote `ggplot2`...

The screenshot displays the RStudio interface. The editor window shows R code for plotting data from the Bovespa. The console shows the execution of the code, including the use of `ggplot2` and `geom_line` to create a line plot. The plot, titled "Curvas de D11", shows the evolution of D11 curves from 2016 to 2022 for five different dates: 2015-01-02, 2015-03-02, 2015-05-04, 2015-07-01, and 2015-08-01. The y-axis is labeled "Taxa" and ranges from 12 to 14. The x-axis is labeled "Datas" and ranges from 2016 to 2022. The plot shows that the curves generally decrease over time, with the 2015-08-01 curve being the highest and the 2015-01-02 curve being the lowest.

```
111
112
113
114
115 \end{frame}
116
117 \begin{frame}
118 \frametitle{A matéria-prima do economista}
119
120 Que tal pegar dados direto da Bovespa?
121
122 \begin{figure}
123 \centering
124 \includegraphics[width=0.8\linewidth]{Figura02.png}
125 \caption{numbered}[dados da BM&F bovespa]
124:27 [Top Level] :
```

```
Environment: History
Global Environment+
Data
  tr_df      600 obs. of 3 variables
Values
  cal      list of 13
  curves   list of 5
  dates    date[1:5], format: "2015-01-02" "2015-03-02" "201...
  terms    num [1:120] 21 42 63 84 105 126 147 168 189 210 ...
Functions
```

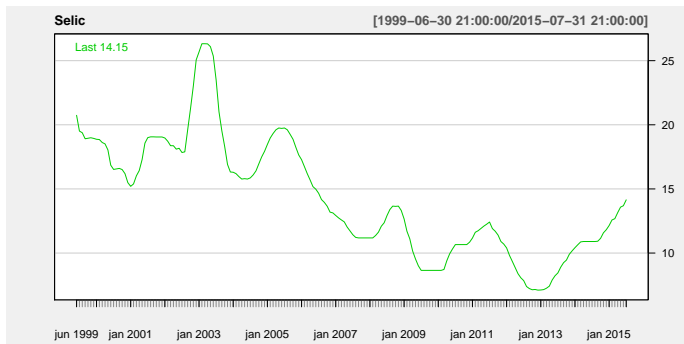
```
Console: Compile PDF x
~/palestra UFF sobre R/palestrauff.Rnw | View Log
Output Issues
Tabela: prdulo (with options)
List of 0
$ echo      : logi FALSE
$ results   : chr "asis"
$ fig.width : num 10
$ fig.height: num 5
$ fig.align : chr "center"
$ out.width : chr "1\linewidth"
$ warning   : logi FALSE
$ message   : logi FALSE
$ size      : chr "footnotesize"

|.....| 100%
ordinary text without R code

output file: palestrauff.tex
[1] "palestrauff.tex"
Houve 22 avisos (use warnings() para ver-los)
Running pdflatex.exe on palestrauff.tex...Failed
Issues: 1 warning
```

## O mundo é feito de dados

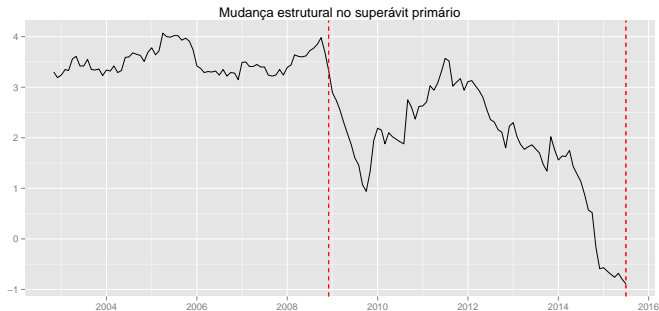
```
library(Quandl)  
Selic <- Quandl('BCB/4189', type='zoo', start_date='1999-07-01', end_date='2015-07-31')  
chartSeries(Selic, theme = 'white')
```





## O mundo é feito de dados

Importando dados diretamente do BCB com a função `getSeries`, elaborada pelo Alexandre Rademaker. Ademais, aplica-se a função `cpt.meanvar` do pacote `changepoint` para ver se há mudança na média e variância do superávit primário.



# As quatro operações

- 1 **Capturar** É preciso saber onde os dados estão...
- 2 **Tratar** É preciso tornar os dados analisáveis...
- 3 **Analisar** É preciso usar o conhecimento teórico para analisar os dados...
- 4 **Apresentar** É preciso saber divulgar os seus resultados...

## Onde estão os dados?

Há inúmeras fontes de dados por aí...

IBGE, BCB, IPEADATA, FRED St. Louis, Banco Mundial, Yahoo Finance, Google Finance, ANFAVEA, etc.

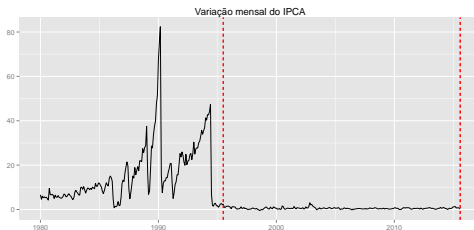
O **R** pode lhe ajudar nessa etapa. Que tal pegar um arquivo **zipado** no site do Banco Central sobre os dados do endividamento brasileiro, abrir o arquivo e trabalhar com ele direto no **R**?

```
temp <- tempfile()
download.file("http://www.bcb.gov.br/ftp/notaecon/Divggnp.zip",temp)
data <- unzip(temp, files='Divggnp.xls')
```

## Um exemplo

Imagine que você está interessado em prever a inflação de setembro de 2015. Você usa, então, a função `getSeries` e pega toda a série do Banco Central.

```
ipca <- getSeries(433, data.ini = '02/01/1980', data.fim = '01/08/2015')  
ipca <- ts(ipcavalor, start=c(1980,01), freq=12)  
autoplot(cpt.meanvar(ipca), main='Variação mensal do IPCA')
```



## É preciso tratar os dados

Observe que a série vem com todo o período da hiperinflação. Se estamos interessados em prever a inflação de setembro, precisamos tentar captar o **processo gerador** da série de inflação. O período da hiperinflação está, portanto, contaminando nosso dado...

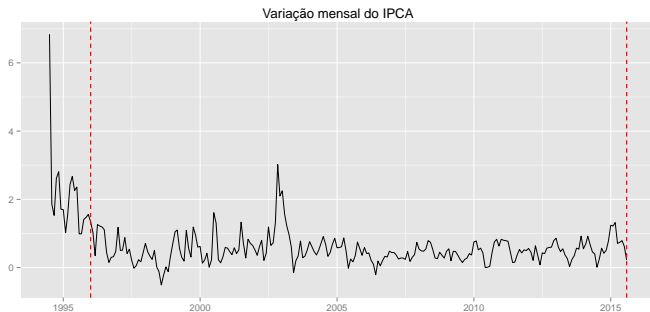
Você poderia, assim, ter pedido uma subamostra do dado. Isto porque, com o plano Real (ufa!), houve uma "quebra" na série, logo devemos tornar a nossa vida mais simples com o comando abaixo.

```
ipca.a1 <- window(ipca, start=c(1994,7), freq=12)
```

## É preciso tratar o dado...

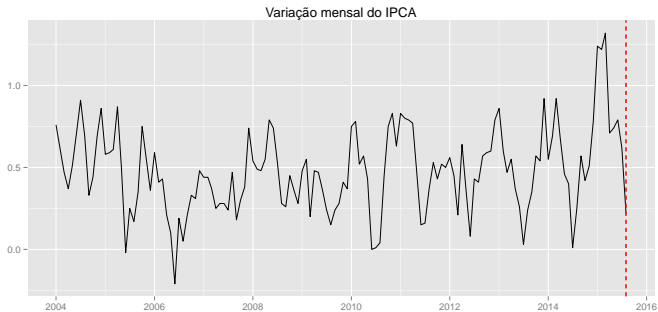
E agora...

```
autoplot(cpt.meanvar(ipca.a1), main='Variação mensal do IPCA')
```



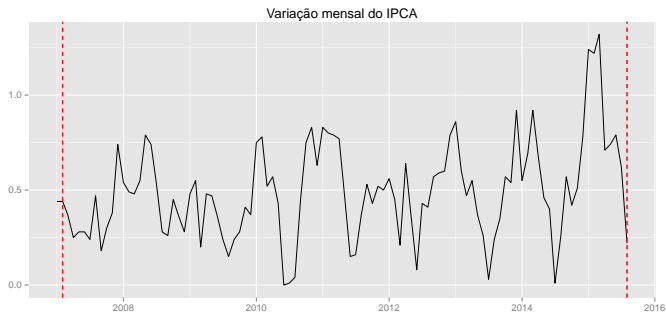
## É preciso tratar o dado...

Vamos continuar melhorando nosso dado, retirando da amostra o período da âncora cambial e o efeito Lula, começando a série a partir de 2004...



## É preciso tratar o dado...

Vamos iniciar a amostra a partir de 2007, retirando assim tanto o período da âncora cambial (1994-1998), quanto o efeito Lula...



Agora, parece que conseguimos...



Estamos prontos para começar o trabalho...

Uma vez capturado e tratado, o dado pode começar a ser analisado...

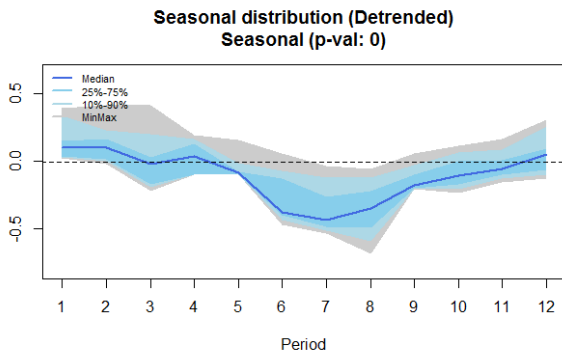
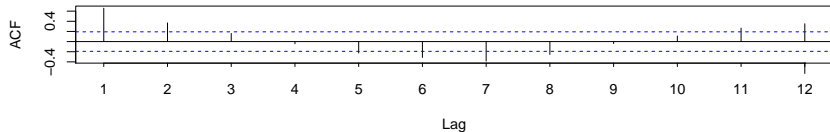


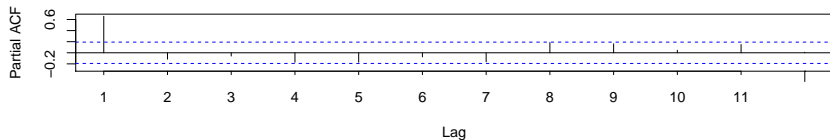
Figura: Sazonalidade do IPCA

# Correlogramas do IPCA

ACF



PACF



## Teste de Raiz Unitária

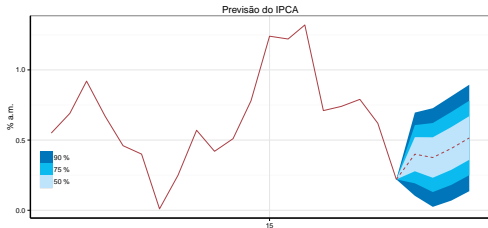
```
library(xtable)
library(urca)
lags.adf <- 3 Lag escolhido com o Ljung-Box
data.ct <- ur.df(ipca.a3, type='trend', lags=lags.adf)
trend <- cbind(t(data.ct@teststat), data.ct@cval)
xtable(trend)
```

	statistic	1pct	5pct	10pct
tau3	-5.20	-3.99	-3.43	-3.13
phi2	9.09	6.22	4.75	4.07
phi3	13.64	8.43	6.49	5.47

## Estimar o modelo<sup>1</sup>

Estimamos um  $SARIMA(1,0,0)(0,0,1)$  e fazemos a projeção 4 períodos à frente.

```
arima <- Arima(ipca.a3, order=c(1,0,0), seasonal=c(0,0,1))  
forecast <- forecast(arima, h=4, level=c(50,75, 90))
```

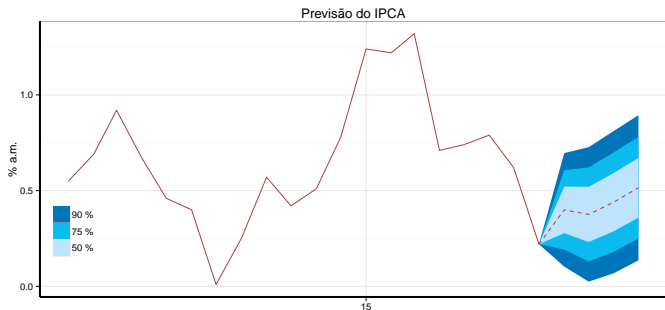


<sup>1</sup>O gráfico que ilustra o slide foi feito por Ricardo Lima.

## Apresentar os dados<sup>2</sup>

Abaixo colocamos uma tabela com os valores médios das projeções, bem como o gráfico. Uma vez estimado e feita a projeção, deve-se comparar com outros modelos, ver a qualidade do ajustamento, adequar a projeção à conjuntura, etc.

Mês	Média
Set/15	0,39
Out/15	0,37
Nov/15	0,44
Dez/15	0,51



<sup>2</sup>O gráfico que ilustra o slide foi feito por Ricardo Lima.

## Por que usar o R?

### Razões para entrar para o mundo do R

- 1 O R é gratuito;
- 2 O R é um substituto mais do que perfeito do EXCEL/VBA;
- 3 Há uma comunidade incrível trabalhando na plataforma;
- 4 É possível unificar as quatro operações no mesmo ambiente, automatizando rotinas;
- 5 Não é uma linguagem voltada apenas para estatística/econometria;
- 6 Liberdade para modificar objetos e apresentar resultados;
- 7 Torna o aprendizado da teoria econômica muito mais intuitivo;

Obrigado!

Liberte-se: venha para o mundo do **R**!

Conheça nossos cursos de **R**:

[www.vitorwilher.com/cursos-de-r](http://www.vitorwilher.com/cursos-de-r)