

# Dados Econômicos no R

Aula ao Vivo 01 - 29 de setembro de 2020

Vítor Wilher

Cientista de Dados | Mestre em Economia



# Plano de Voo

Sobre o professor

Sobre a Análise Macro

O ambiente de trabalho clássico de um economista

Por que usar o R?

Exemplo 01: coletar dados de inflação do SIDRA/IBGE

Exemplo 02: tratamento de dados da Produção Industrial

Exemplo 03: dessazonalizar uma série econômica com o R

Exemplo 05: gerando um documento RMarkdown

## Sobre o professor

**Vítor Wilher** é Bacharel e Mestre em Economia, pela Universidade Federal Fluminense e Especialista em Data Science. Sua dissertação de mestrado foi na área de política monetária, intitulada “*Clareza da Comunicação do Banco Central e Expectativas de Inflação: evidências para o Brasil*”, defendida perante banca composta pelos professores Gustavo H. B. Franco (PUC-RJ), Gabriel Montes Caldas (UFF), Carlos Enrique Guanziroli (UFF) e Luciano Vereda Oliveira (UFF). Já trabalhou em grandes empresas, nas áreas de telecomunicações, energia elétrica, consultoria financeira e consultoria macroeconômica. Atualmente, é Sócio-fundador da **Análise Macro** e Palestrante.

Mais sobre o professor, visite: <https://vitorwilher.github.io/>

# Sobre a Análise Macro

A Análise Macro é um sonho acalentado durante muito tempo e de pé desde setembro de 2015.

Acreditamos que a teoria não pode estar dissociada da prática.

Queremos fazer parte de uma revolução: mostrar para as pessoas que a verdade está nos dados.

E para ter acesso a ela é preciso ter domínio sobre as etapas de coleta, tratamento, análise e apresentação de dados.

# Sobre a Análise Macro

Nossa missão é, portanto, lhe ajudar nesse processo. Para isso, oferecemos:

- cursos pré-formatados que são ofertados de maneira on-line;
- cursos na modalidade in company;
- consultoria customizada de análise de dados;
- o Clube do Código, local onde são divulgados todos os códigos de nossos exercícios de análise de dados.

Visite nosso site e saiba mais: <http://analisemacro.com.br>

# O ambiente de trabalho clássico de um economista



Figure 1: Vários programas para lidar com dados

# Por que usar o R?

- A existência de uma IDE bastante poderosa chamada **RStudio** que permite integrar as etapas de coleta, tratamento, análise e apresentação de dados;
- A existência de uma comunidade grande e bastante entusiasmada, que compartilha conhecimento todo o tempo. Se você sabe **R**, a fronteira do conhecimento está disponível para você;
- o **R** é gratuito, *open source*, de modo que você não precisa comprar licenças de software para instalá-lo;
- Tem inúmeras bibliotecas (pacotes) em estatística, econometria, *machine learning*, visualização, importação e tratamento de dados;
- Ferramentas poderosas para comunicação dos resultados da sua pesquisa, seja em forma de um website ou em pdf.

Mas chega de preâmbulo, vamos à prática?



## Exemplo 01: coletar dados de inflação do SIDRA/IBGE

Suponha que você esteja interessado em verificar o comportamento da inflação ao longo do tempo?

Qual estratégia você usaria para isso?

Talvez, ir ao site do SIDRA/IBGE, baixar um arquivo .csv ou .xls, abrir a planilha, produzir um gráfico. . .

# Exemplo 01: coletar dados de inflação do SIDRA/IBGE

No **R**, podemos usar um pacote para isso. Nesse caso específico, podemos usar o pacote `sidrar`, como abaixo:

```
library(sidrar)
library(tidyverse)
ipca =
  '/t/1737/n1/all/v/2266/p/all/d/v2266%2013' %>%
  get_sidra(api=.) %>%
  mutate(date = parse_date(`Mês (Código)`, format="%Y%m")) %>%
  rename(indice = Valor) %>%
  mutate(inflacao_mensal = (indice/lag(indice,1)-1)*100,
         inflacao_anual = (indice/lag(indice,12)-1)*100) %>%
  select(date, indice, inflacao_mensal, inflacao_anual) %>%
  as_tibble()

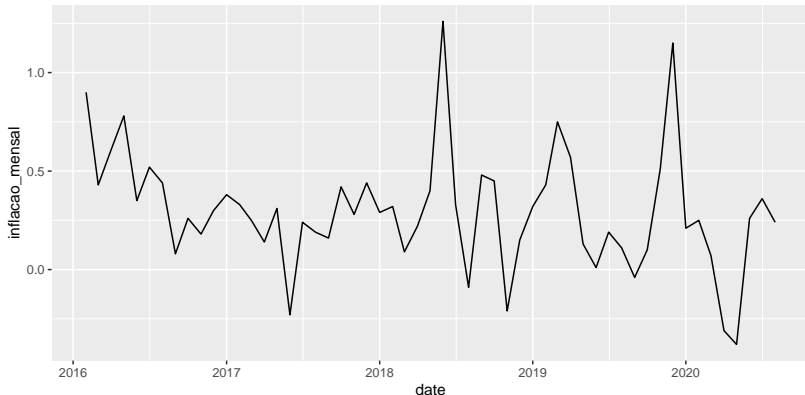
tail(ipca)
```

```
## # A tibble: 6 x 4
##   date      indice inflacao_mensal inflacao_anual
##   <date>    <dbl>         <dbl>         <dbl>
## 1 2020-03-01 5348.           0.0700          3.30
## 2 2020-04-01 5332.          -0.310          2.40
## 3 2020-05-01 5312.          -0.380          1.88
## 4 2020-06-01 5325.           0.260          2.13
## 5 2020-07-01 5345.           0.360          2.31
## 6 2020-08-01 5357.           0.240          2.44
```

# Exemplo 01: coletar dados de inflação do SIDRA/IBGE

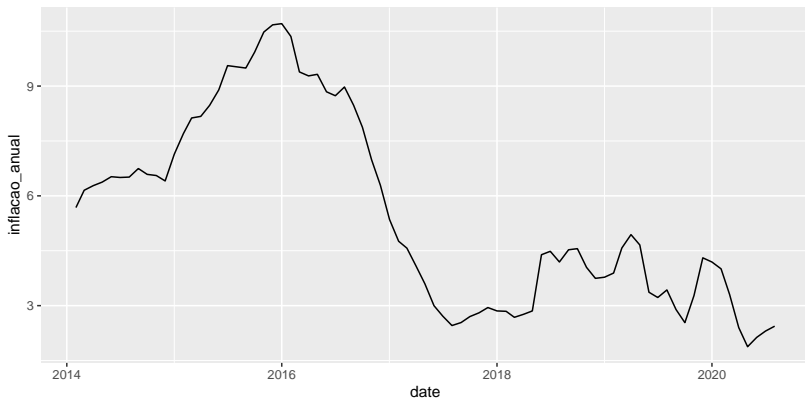
Com os dados devidamente importados, nós podemos gerar gráficos. . .

```
filter(ipca, date > '2016-01-01') %>%  
  ggplot(aes(x=date, y=inflacao_mensal))+  
  geom_line()
```



# Exemplo 01: coletar dados de inflação do SIDRA/IBGE

```
filter(ipca, date > '2014-01-01') %>%  
  ggplot(aes(x=date, y=inflacao_anual))+  
  geom_line()
```



# Exemplo 01: coletar dados de inflação do SIDRA/IBGE

Podemos também pegar *subgrupos* do IPCA, também diretamente para o **R**. Por exemplo, podemos estar interessados no comportamento da inflação de alimentos:

```
library(tstools)

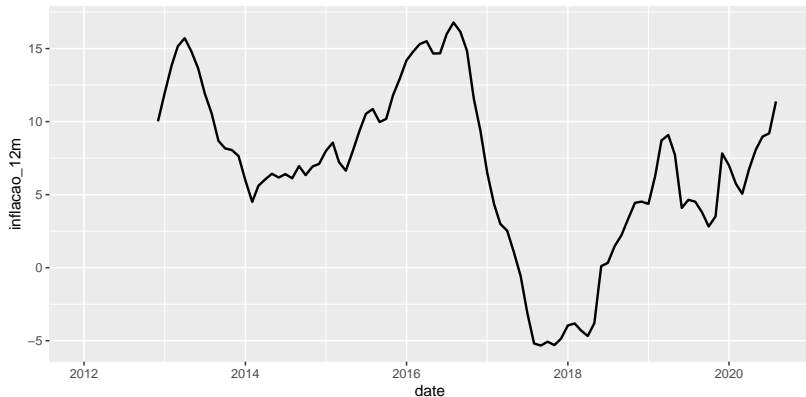
alim_dom_01 =
  get_sidra(api='/t/1419/n1/all/v/63/p/all/c315/7171/d/v63%202')

alim_dom_02 =
  get_sidra(api='/t/7060/n1/all/v/63/p/all/c315/7171/d/v63%202')

alim_dom = full_join(alim_dom_01, alim_dom_02) %>%
  mutate(date = parse_date(`Mês (Código)`, format="%Y%m")) %>%
  mutate(inflacao_12m = acum_p(Valor,12)) %>%
  select(date, Valor, inflacao_12m)
```

## Exemplo 01: coletar dados de inflação do SIDRA/IBGE

```
ggplot(alim_dom, aes(x=date, y=inflacao_12m))+  
  geom_line(size=.8)
```



## Exemplo 02: tratamento de dados da Produção Industrial

Os números-índices nos informam, em geral, a evolução de uma cesta de produtos ou serviços ao longo do tempo. O IPCA que pegamos anteriormente é um exemplo de número-índice. A partir dele, nós calculamos a *inflação*, a taxa de crescimento do nível geral de preços entre dois períodos distintos.

De forma geral, a partir desses índices, podemos criar *métricas de crescimento*, de modo a avaliar melhor o comportamento dele ao longo do tempo.

## Exemplo 02: tratamento de dados da Produção Industrial

Por exemplo, a variação na margem é aquela que é obtida pela razão entre o período  $t$  contra o período imediatamente anterior.

$$Var_{margem} = \left[ \left( \frac{I_t}{I_{t-1}} \right) - 1 \right] \times 100$$

```
pim_sa =  
  '/t/3653/n1/all/v/3134/p/all/c544/129314/d/v3134%201' %>%  
  get_sidra(api=.) %>%  
  mutate(date = parse_date(`Mês (Código)`, format='%Y%m')) %>%  
  select(date, Valor) %>%  
  mutate(var_margem = (Valor/lag(Valor,1)-1)*100) %>%  
  as_tibble()  
  
tail(pim_sa)
```

```
## # A tibble: 6 x 3  
##   date      Valor var_margem  
##   <date>    <dbl>    <dbl>  
## 1 2020-02-01  88.1      0.916  
## 2 2020-03-01  79.9     -9.31  
## 3 2020-04-01  64.3    -19.5  
## 4 2020-05-01  69.9      8.71  
## 5 2020-06-01  76.7     9.73  
## 6 2020-07-01  82.8     7.95
```



## Exemplo 02: tratamento de dados da Produção Industrial

A variação interanual é aquela que obtida pela razão entre o período  $t$  e o mesmo período no ano anterior.

$$Var_{interanual}^{Mensal} = \left[ \left( \frac{I_t}{I_{t-12}} \right) - 1 \right] \times 100$$

A variação interanual também pode ser feita em relação ao mesmo trimestre do ano anterior.

## Exemplo 02: tratamento de dados da Produção Industrial

```
pim = get_sidra(api='/t/3653/n1/all/v/3135/p/all/c544/129314/d/v3135%201')
pim =
  pim %>%
    mutate(date = parse_date(`Mês (Código)`, format='%Y%m')) %>%
    select(date, Valor) %>%
    mutate(var_interanual = (Valor/lag(Valor,12)-1)*100)

tail(pim)
```

```
##           date Valor var_interanual
## 218 2020-02-01  78.9      -0.2528445
## 219 2020-03-01  78.0      -3.7037037
## 220 2020-04-01  60.4     -27.5779376
## 221 2020-05-01  71.2    -21.7582418
## 222 2020-06-01  78.3     -8.7412587
## 223 2020-07-01  90.6     -3.1016043
```

## Exemplo 02: tratamento de dados da Produção Industrial

### Variação acumulada em determinado período

O cálculo da variação acumulada em determinado período é feita em três passos:

- Calcula-se a variação percentual em relação ao período imediatamente anterior;
- Transforme esses valores em um fator, i.e., some 1 ao valor percentual em notação decimal;
- Multiplica-se todos os fatores que se deseja acumular.

# Exemplo 02: tratamento de dados da Produção Industrial

## Variação acumulada em determinado período

No R, vamos utilizar a função `acum_i()` do pacote `tstools`, disponível no github.<sup>1</sup>

```
library(tstools)
pim =
  pim %>%
    mutate(var_anual = acum_i(Valor,12))
tail(pim)
```

```
##           date Valor var_interannual  var_anual
## 218 2020-02-01  78.9      -0.2528445 -1.1838242
## 219 2020-03-01  78.0      -3.7037037 -0.9802988
## 220 2020-04-01  60.4     -27.5779376 -2.8827797
## 221 2020-05-01  71.2     -21.7582418 -5.3690002
## 222 2020-06-01  78.3      -8.7412587 -5.6058728
## 223 2020-07-01  90.6     -3.1016043 -5.6755207
```

---

<sup>1</sup><https://github.com/leripio/tstools>

## Exemplo 02: tratamento de dados da Produção Industrial

```
data_pim = inner_join(pim_sa, pim, by='date') %>%  
  rename(pim_sa = Valor.x,  
         pim = Valor.y)  
  
tail(data_pim)
```

```
## # A tibble: 6 x 6  
##   date      pim_sa var_margem   pim var_interannual var_anual  
##   <date>    <dbl>    <dbl> <dbl>    <dbl>    <dbl>  
## 1 2020-02-01  88.1      0.916  78.9    -0.253    -1.18  
## 2 2020-03-01  79.9     -9.31   78      -3.70    -0.980  
## 3 2020-04-01  64.3    -19.5   60.4    -27.6     -2.88  
## 4 2020-05-01  69.9      8.71  71.2    -21.8     -5.37  
## 5 2020-06-01  76.7      9.73  78.3    -8.74     -5.61  
## 6 2020-07-01  82.8      7.95  90.6     -3.10     -5.68
```

## Exemplo 02: tratamento de dados da Produção Industrial

```
data_pim %>%  
  gather(variavel, valor, -date) %>%  
  filter(date > '2014-01-01') %>%  
  ggplot(aes(x=date, y=valor, colour=variavel))+  
  geom_line(size=.8)+  
  facet_wrap(~variavel, scales='free')+  
  theme(legend.position = 'none')
```



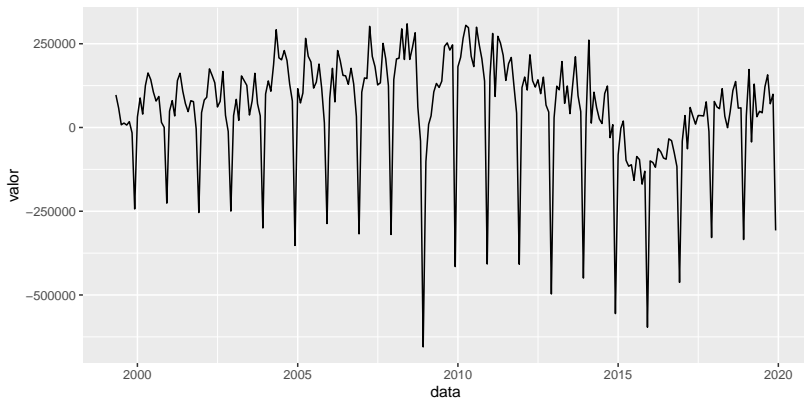
## Exemplo 03: dessazonalizar uma série econômica com o R

Um problema muito comum em séries econômicas é a existência de **sazonalidade**. Isto é, em um ano calendário, uma série pode apresentar um comportamento que se repete. Para ilustrar, vamos pegar os dados do CAGED antigo a partir do IPEADATA:

```
library(ecoseries)
## Baixar dados do CAGED
saldo_caged = series_ipeadata("272844966", periodicity = 'M')$serie_272844966
```

## Exemplo 03: dessazonalizar uma série econômica com o R

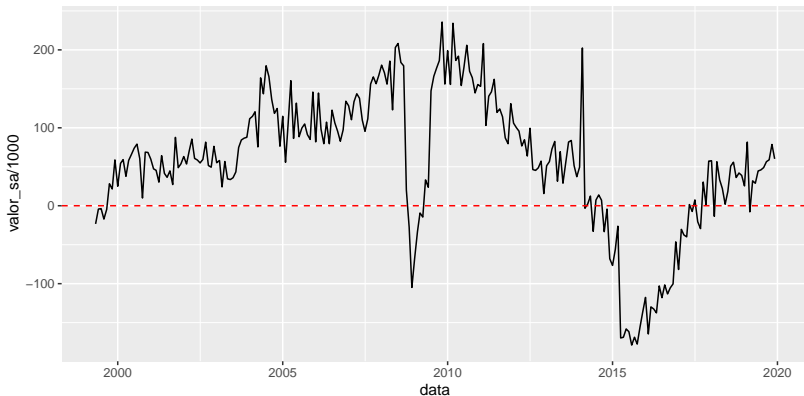
```
ggplot(saldo_caged, aes(x=data, y=valor))+  
  geom_line()
```





## Exemplo 03: dessazonalizar uma série econômica com o R

```
library(seasonal)
saldo_caged =
  saldo_caged %>%
  mutate(valor_sa = final(seas(ts(valor, start=c(1999,05),
                                   freq=12))))
ggplot(saldo_caged, aes(x=data, y=valor_sa/1000))+
  geom_line()+
  geom_hline(yintercept=0, colour='red', linetype='dashed')
```

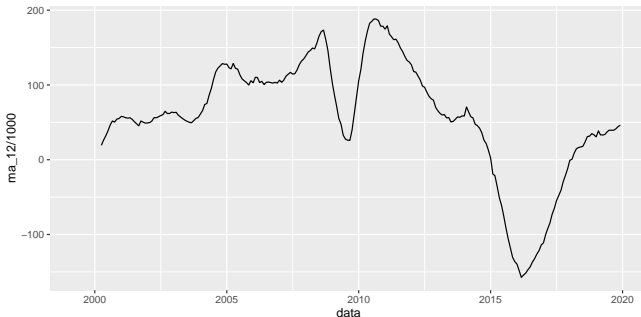


## Exemplo 04: criando uma média móvel sobre os dados

Podemos continuar suavizando os dados, construindo uma média móvel de 12 meses, como no exemplo abaixo.

```
library(zoo)
saldo_caged =
  saldo_caged %>%
    mutate(ma_12 = rollapply(valor_sa, 12, mean, align='right',
                              fill=NA))

ggplot(saldo_caged, aes(x=data, y=ma_12/1000))+
  geom_line()
```



Vamos agora terminar nossa aula, integrando  
em um único documento a coleta, o  
tratamento e a apresentação de dados?

## Exemplo 05: gerando um documento RMarkdown

No meu caso aqui, estou gerando uma apresentação para vocês. . .

Para isso, vamos dar uma olhada no arquivo .Rmd

