

Séries Temporais

James D Santos

2025-04-09

Table of contents

Prefácio

Estas são as notas de aula estão sendo produzidas para a utilização na disciplina Séries Temporais, no Bacharelado de Estatística.

Considere-as um rascunho, e portanto, sujeita a erros.

Dúvidas e sugestões podem ser enviadas para o e-mail james@ufam.edu.br

1 Summary

In summary, this book has no content whatsoever.

$1 + 1$

[1] 2

2 Introdução

2.1 Notações

Serão utilizadas letras minúsculas para designar tanto variáveis aleatórias quanto seus respectivos valores observados, entendendo a diferença clara no contexto. Exemplo: em

$$x_t \sim \text{Normal}(0, 1),$$

x_t representa uma variável aleatória, enquanto que em $x_t = 0$ é um valor observado.

Vetores serão denotados por negritos e sempre serão vetores-coluna. Exemplo

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}.$$

O vetor \boldsymbol{x}' é o transposto de \boldsymbol{x} .

Para $\mathcal{T} = \{1, 2, \dots\}$,

- Se $A \subset \mathcal{T}$. Então $\boldsymbol{x}_A = \{x_t, t \in A\}$.
- $\boldsymbol{x}_{a:b} = x_a, x_{a+1}, \dots, x_{b-1}, x_b$.
- Um vetor de dimensão q observado no tempo t é escrito como

$$\boldsymbol{x}_t = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix}_t.$$

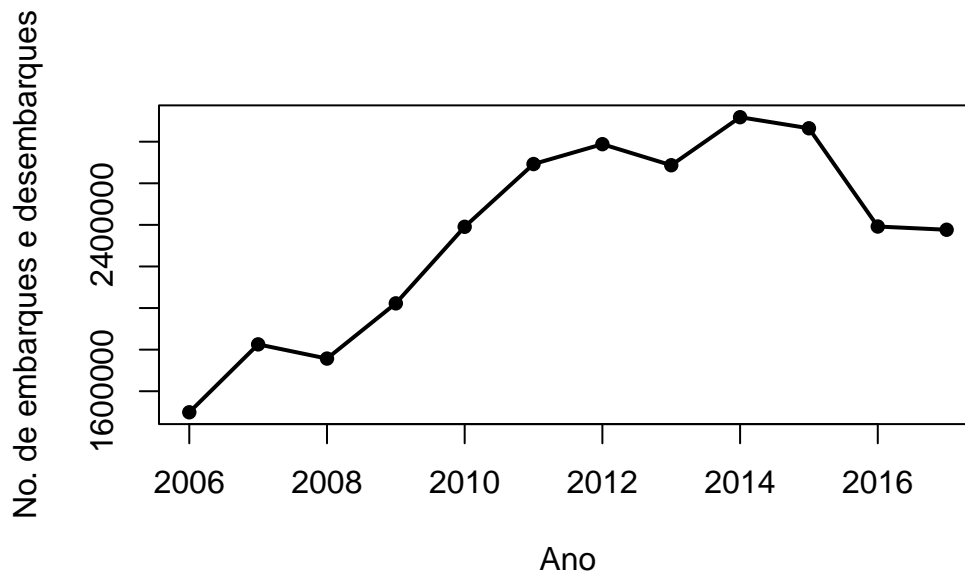
2.2 O que é uma análise de séries temporais?

Considera-se que uma série temporal é uma coleção de observações realizadas ao longo do tempo. Será utilizada a notação x_t para designar o valor registrado no tempo t e $\mathcal{D}_t = \{x_1, \dots, x_t\}$ representará a série observada até o tempo t .

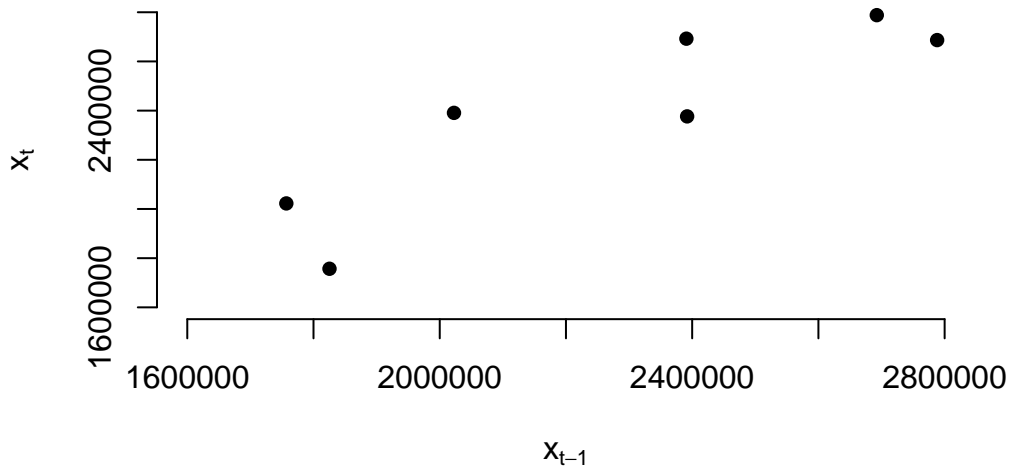
Existem três objetivos principais no estudo de séries temporais

- *Previsão*: Dado \mathcal{D}_t a previsão trata do problema de realizar inferências sobre x_{t+h} , com $h > 0$.
- *Suavização (ou alisamento)*: Dado \mathcal{D}_t a suavização trata do problema de realizar inferências baseadas x_{t-h} , com $h > 0$
- *Monitoramento*: detectar em tempo real as mudanças ou discrepâncias no comportamento do processo.

Note que tais objetivos só fazem sentido se há alguma estrutura de dependência entre as variáveis que compõe a série temporal. Para ilustrar, considere a figura abaixo representa o gráfico a série temporal com o número anual de embarques e desembarques de passageiros em vôos domésticos no aeroporto Eduardo Gomes.



Ainda considerando a série acima, seja x_t o número de embarques e desembarques registrado no ano t . A figura abaixo mostra o diagrama de dispersão entre x_t e x_{t-1} , de onde é possível observar a correlação positiva, estimada em 0,86.



De posse desses resultados, pode-se imaginar um primeiro modelo, no qual a relação entre o presente e o passado imediato é ditado por uma regressão linear simples, gerando a equação

$$\hat{x}_t = 7,589 \times 10^5 + 0,7109x_{t-1}.$$

Sabendo que $x_{2017} = 2.376.505$, uma previsão para 2018 seria $\hat{x}_{2018} = 2.448.357$. O valor observado em 2018 foi 2.572.159, gerando um erro de previsão igual a $x_{2018} - \hat{x}_{2018} = 195.654$ embarques e desembarques domésticos.

2.3 Exemplos de séries temporais

2.3.1 Eletrocardiograma

```
ts.plot(ECG)
```

2.3.2 Produto Interno Bruto Brasileiro

```
ts.plot(PIB)
```

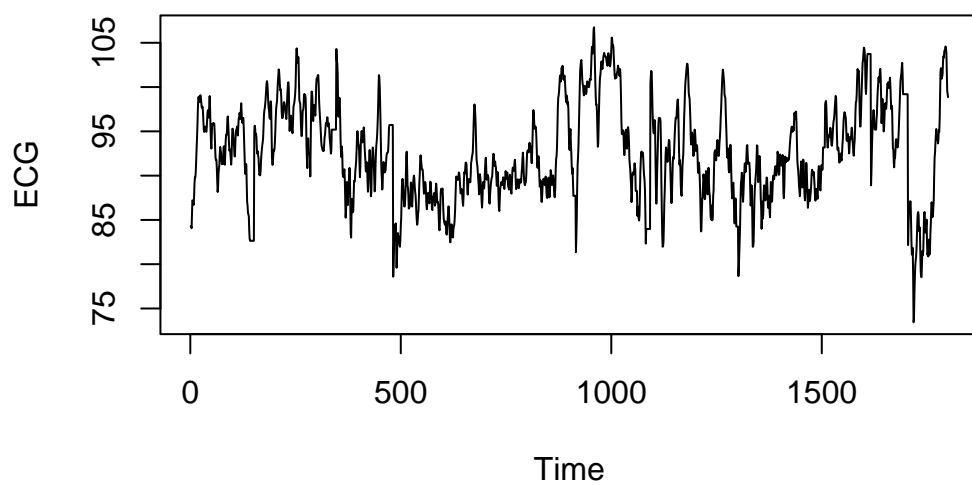


Figure 2.1: 1800 medidas da taxa cardíaca instantânea, em batidas por minuto, de um indivíduo.

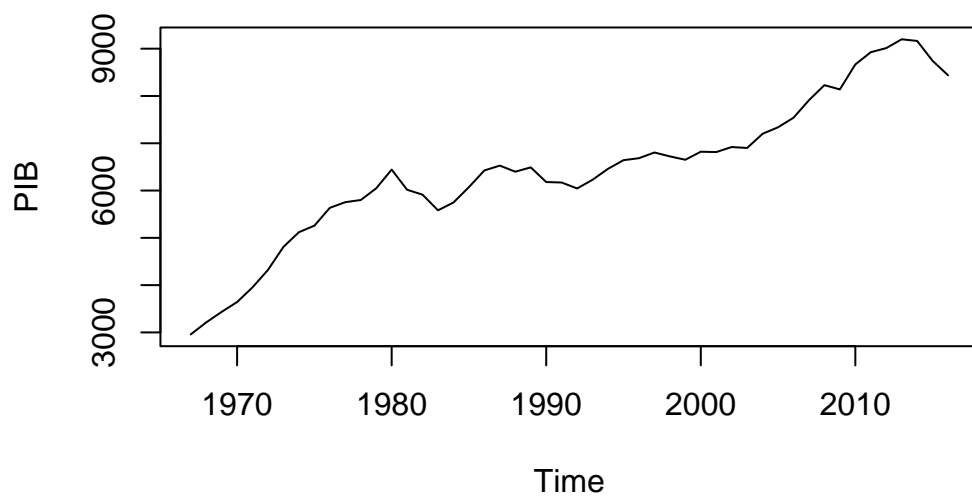


Figure 2.2: PIB entre 1967 e 2014 corrigidos pelo valor do dólar em 4/2015.

2.3.3 Mortes por doenças pulmonares no Reino Unido

```
ts.plot(ldeaths)
```

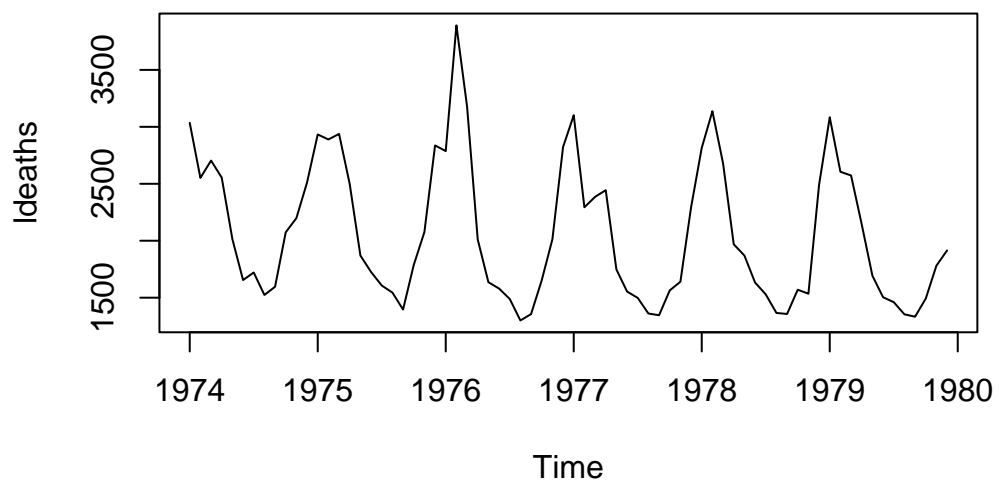


Figure 2.3: PIB entre 1967 e 2014 corrigidos pelo valor do dólar em 4/2015.

3 Criando séries no R

Esta seção tem por objetivo mostrar algumas funções em R para a criação e análise exploratória de séries temporais.

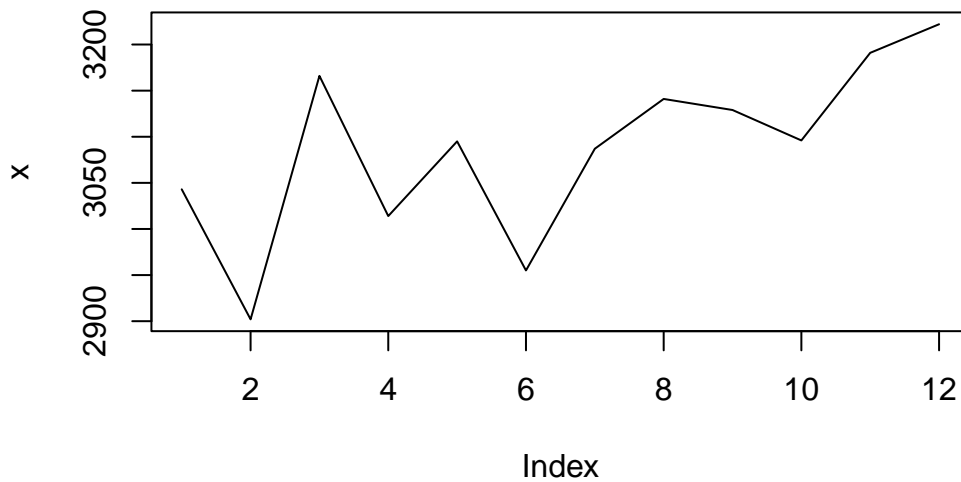
3.1 A classe `ts`

Para todos os efeitos, uma série temporal é um vetor numérico. O vetor abaixo armazena o número de nascidos vivos por mês na cidade de Manaus em 2021, sendo `x[1]` o mês de janeiro e assim sucessivamente.

```
x <- c( 3043, 2902, 3166, 3014, 3095, 2955, 3087, 3141,  
       3129, 3096, 3191, 3222)
```

Por sua vez, o gráfico da série temporal pode ser construído utilizando a função `plot`, com o argumento `type='l'`.

```
plot(x, type = 'l')
```



Contudo, é útil construir a série temporal como um objeto da classe **ts**. Tal função possui dois argumentos importantes:

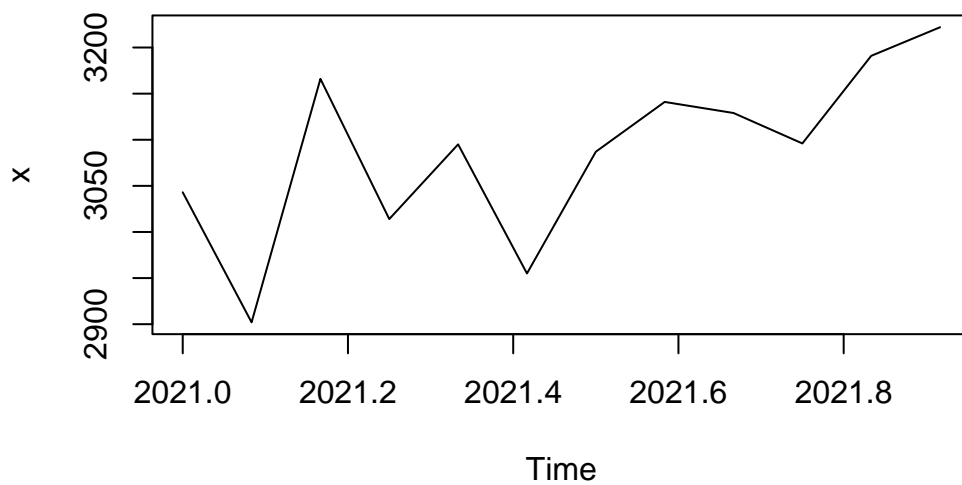
- **frequency**: representa o número de observações por unidade de tempo. Por exemplo, se tempo está sendo contado em anos, mas os dados são mensais, então **frequency=12**; se os dados forem trimestrais, **frequency=4** e assim por diante.
- **start**: representa o tempo da primeira observação. Pode ser representado por um único número ou por um vetor de comprimento dois. Esse último caso só é utilizado quando **frequency** é diferente de 1 e representa a ordem, em relação à frequência, da primeira observação. Por exemplo, com **frequency=12**, o vetor **start=c(1996,2)** implica que a primeira observação data de fevereiro de 1996.

No código abaixo, o vetor criado anteriormente é colocado com um objeto **ts**

```
x <- ts( x, start = c(2021,1), frequency = 12)
x
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2021	3043	2902	3166	3014	3095	2955	3087	3141	3129	3096	3191	3222

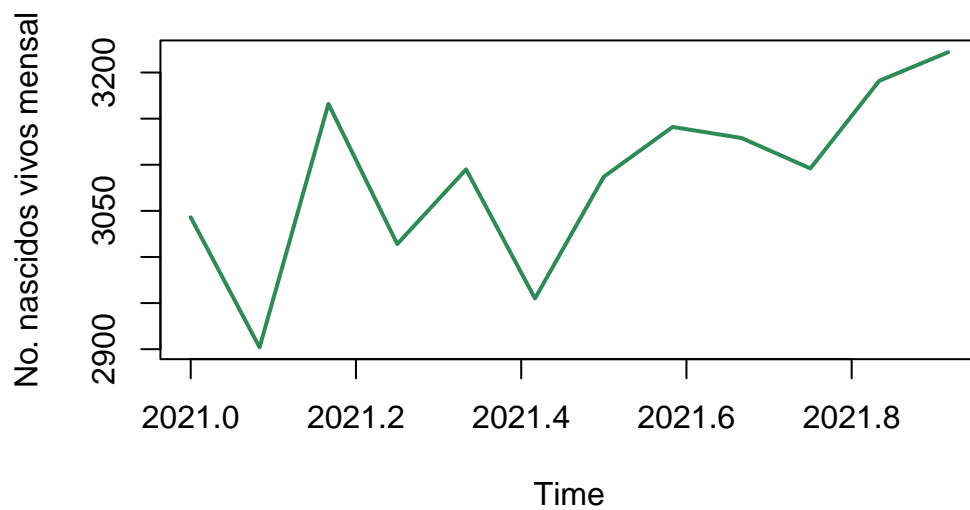
```
ts.plot(x)
```



No gráfico acima, a parte decimal no eixo x representa a fração do tempo entre de um ano (começando em 0 e acumulando $1/12$ para cada mês subsequente).

O gráfico pode ser customizado do mesmo modo que um `plot`. Abaixo segue um exemplo.

```
plot(x, ylab = 'No. nascidos vivos mensal', lwd = 2, col = 'seagreen')
```



A função `start` retorna o início da série, `end` seu fim e `frequency` o número de observações por unidade de tempo. Observe o exemplo abaixo.

```
start(x)
```

```
[1] 2021    1
```

```
end(x)
```

```
[1] 2021   12
```

```
frequency(x)
```

```
[1] 12
```

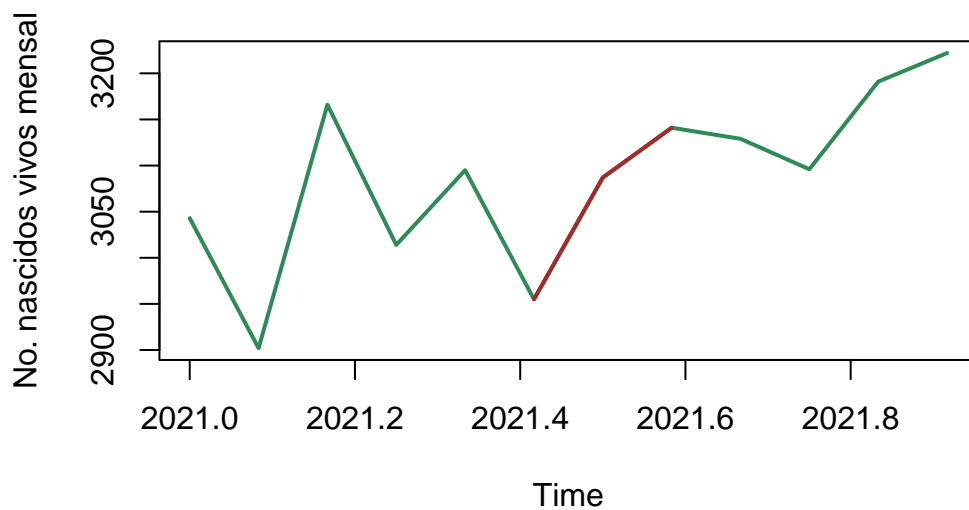
A partir das informações acima, sabe-se a série `x` é mensal (`frequency=12`), que sua primeira observação data de janeiro de 2021 e a última de dezembro de 2021.

3.2 A função window

A função `window` seleciona um subconjunto da série temporal. Abaixo foram selecionados apenas os nascimentos entre Junho e Agosto e estes valores foram registrados no gráfico.

```
z <- window(x, start=c(2021,6), end = c(2021,8))

plot(x, ylab = 'No. nascidos vivos mensal', lwd = 2, col = 'seagreen')
lines(z, col = 'brown', lwd= 2)
```



3.3 O pacote data.table

Assim como números e textos possuem classes específicas, as datas no ambiente R também possuem sua própria classe, denominada `Date`.

```
# 3 de agosto de 1998 (formato americano)
x <- '1998/8/3'
as.Date(x)
```

```
[1] "1998-08-03"
```

Para que o R entenda uma data fora do padrão americano, é necessário passar o formato para o argumento `format`. Seguem alguns exemplos:

```
# 3 de agosto de 1998 (formato nacional)
x <- '3/8/1998'
as.Date(x, format = '%d/%m/%Y')
```

```
[1] "1998-08-03"
```

```
x <- '3-8-1998'
as.Date(x, format = '%d-%m-%Y')
```

```
[1] "1998-08-03"
```

```
# agosto de 1998
x <- '8/1998'
as.Date(x, format = '%m/%Y')
```

```
[1] NA
```

Ao se trabalhar com fontes originais, é comum ter como unidade amostral um evento com sua data registrada. Em geral, nosso objetivo é determinar a quantidade de eventos dentro de dias, semanas, meses ou anos. O pacote `data.table` permite lidar com esse problema de modo rápido, criando um objeto deste tipo utilizando a função `fread`.

Para ilustrar, será utilizada a base de dados de acidentes com aeronaves, mantida pela Força Aérea Brasileira, que registra diariamente o número de acidentes com aeronaves.

```
library(data.table)
url <- 'https://drive.google.com/uc?authuser=0&id=1iYrnwXgmLK07x8b330aD73sc0VruZEuz&export=csv'

aereo <- fread(url, encoding = 'Latin-1')
aereo$ocorrencia_dia <- as.Date(aereo$ocorrencia_dia, '%d/%m/%Y')
```

Um objeto do tipo `data.table` permite uma série de consultas. Em geral, pode-se fazer `aereo[a,b,c]`, onde `a` é uma consulta/função nas linhas, `b` nas colunas e `c` é um agrupador. Uma excelente introdução pode ser vista em [Introduction to data.table](#).

Abaixo, foi selecionada a coluna de interesse `ocorrencia_dia`.

```
fab_dia <- aereo[, 'ocorrencia_dia', ]
head(fab_dia)
```

```
      ocorrencia_dia
1:      2023-04-05
2:      2023-06-24
3:      2023-06-27
4:      2023-06-30
5:      2023-06-25
6:      2023-06-23
```

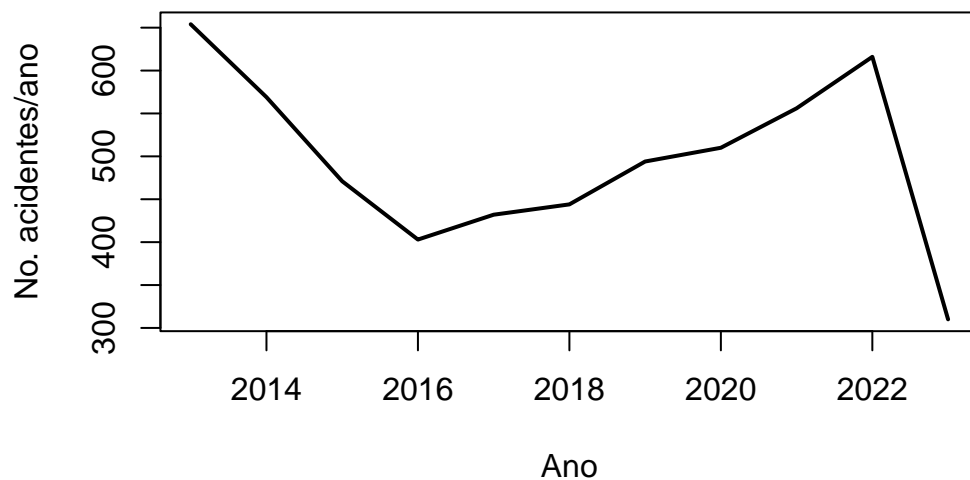
Ao utilizar o operador `.N` em `[, .N, c]`, é retornado o número de linhas que possuem o agrupamento em `c`. Abaixo, as datas do banco são agrupadas por ano.

```
fab_ano <- fab_dia[, .N, by=(year(ocorrencia_dia))]
fab_ano <-fab_ano[ order(year) ]
head(fab_ano)
```

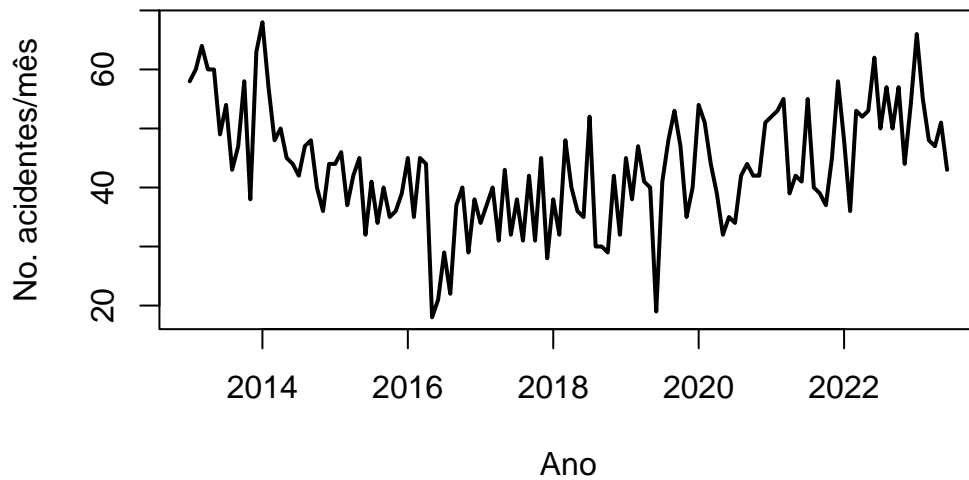
```
      year    N
1: 2013  654
2: 2014  569
3: 2015  471
4: 2016  403
5: 2017  432
6: 2018  444
```

Os comandos a seguir criam dois objetos do tipo `ts`, sendo um para o número anual de acidentes e outro para o mensal

```
fab_ano <- ts( fab_ano, start = 2013)
plot(fab_ano[,2], lwd = 2, ylab = 'No. acidentes/ano', xlab = 'Ano')
```

```
fab_mes <- fab_dia[, .N, by=.(year(ocorrencia_dia), month(ocorrencia_dia))]  
  
fab_mes <-fab_mes[ order(year, month ) ]  
fab_mes <- ts( fab_mes[,3], start = c(2013, 1), frequency = 12)  
plot(fab_mes, lwd = 2, ylab = 'No. acidentes/mês', xlab = 'Ano')
```



3.4 Exercício

Exercício 1

A série abaixo contém a data dos óbitos maternos no Brasil a partir de 2010.

```
url <- 'https://drive.google.com/uc?authuser=0&id=1tYFFT9L2iopKmBDUI3P8qNIRa0nMYj7d&export'
```

Crie uma série temporal com o número de óbitos mensal e faça um gráfico. Crie uma janela para colocar no gráfico o período da pandemia de COVID-19.