

# Séries de Tempo

## Aula 2 - Sazonalidade e Tendência

Regis A. Ely

Departamento de Economia  
Universidade Federal de Pelotas

10 de julho de 2020

# Conteúdo

Estatísticas e gráficos de séries de tempo

Exemplo no R: Turismo na Austrália

Estatísticas descritivas

Gráficos de séries de tempo

Componentes de uma série de tempo

Sazonalidade

Inspeção

Características

Estimação

Tendência

Inspeção

Características

Estimação

Diferenciação

Testes de hipótese

# Gráficos de séries de tempo

- Séries de tempo podem ter diferentes frequências: *diária, semanal, mensal, trimestral, anual, etc*
- Para visualizar séries de tempo são utilizados gráficos de linha com as datas no eixo das abcissas (*eixo x*)
- Com a inspeção gráfica de uma série de tempo é possível identificar rapidamente componentes como:
  - *Outliers*
  - Tendências
  - Sazonalidades
  - Quebras estruturais
  - Ciclos

# Carregando os pacotes

Nos exemplos desta aula, utilizaremos alguns pacotes do R úteis para lidar com séries de tempo:

- `tidyverse`: uma coleção de pacotes para manipular e plotar gráficos, incluindo o pacote `ggplot2` (*The Grammar of Graphics*)
- `lubridate`: cria e manipula datas no R
- `tsibble`: manipulação de séries de tempo e possibilidade de agrupar várias séries em uma variável no R, através de identificadores
- `feasts`: estatísticas descritivas e modelos de sazonalidade

# Carregando os pacotes

Para carregar os pacotes usamos a função `library`. Lembre que você deve instalar estes pacotes antes de utilizá-los, por exemplo, `install.packages("tsibble")`<sup>1</sup>

```
library(tidyverse)
library(lubridate)
library(tsibble)
library(feasts)
```

---

<sup>1</sup>Para esta aula é importante atualizar o R para a última versão. Ver pacote `installr`.

# Turismo na Austrália

Vamos trabalhar com algumas séries de tempo que contém informações sobre o número de viagens domésticas na Austrália:

```
glimpse(tourism)
```

```
## Rows: 24,320
## Columns: 5
## Key: Region, State, Purpose [304]
## $ Quarter <qtr> 1998 Q1, 1998 Q2, 1998 Q3, 1998 Q4...
## $ Region <chr> "Adelaide", "Adelaide", "Adelaide"...
## $ State <chr> "South Australia", "South Australi...
## $ Purpose <chr> "Business", "Business", "Business"...
## $ Trips <dbl> 135.0777, 109.9873, 166.0347, 127....
```

## Como definir um tsibble

A variável `tourism` já está em formato `tsibble`, mas se tivéssemos uma base de dados que não está neste formato, deveríamos especificar os identificadores e a variável de tempo com o comando `as_tsibble`:

```
data <- tourism %>%  
  as_tsibble(  
    key = c("Region", "Purpose"),  
    index = Quarter  
  )
```

Assim, definimos os identificadores de Região e Propósito da viagem, e a variável de tempo como o Trimestre em que ocorreu a viagem, gravando na variável `data`

# Propósitos das viagens

Nesta base de dados, existe uma série de tempo para cada região e propósito. Ao todo são 4 propósitos:

```
tourism %>% distinct(Purpose)
```

```
## # A tibble: 4 x 1
##   Purpose
##   <chr>
## 1 Business
## 2 Holiday
## 3 Other
## 4 Visiting
```

## Soma das viagens por propósito

Vamos somar o número de viagens nas diferentes regiões, agrupando pelo propósito da viagem:

```
data <- tourism %>%  
  group_by(Purpose) %>%  
  summarise(Trips = sum(Trips), .groups = "drop")
```

Agora temos uma série de tempo para cada propósito da viagem

# Estatísticas descritivas

A função `features` pode ser usada para reportar estatísticas em relação a cada uma das séries por grupo:

```
data %>%  
  features(Trips, quantile)
```

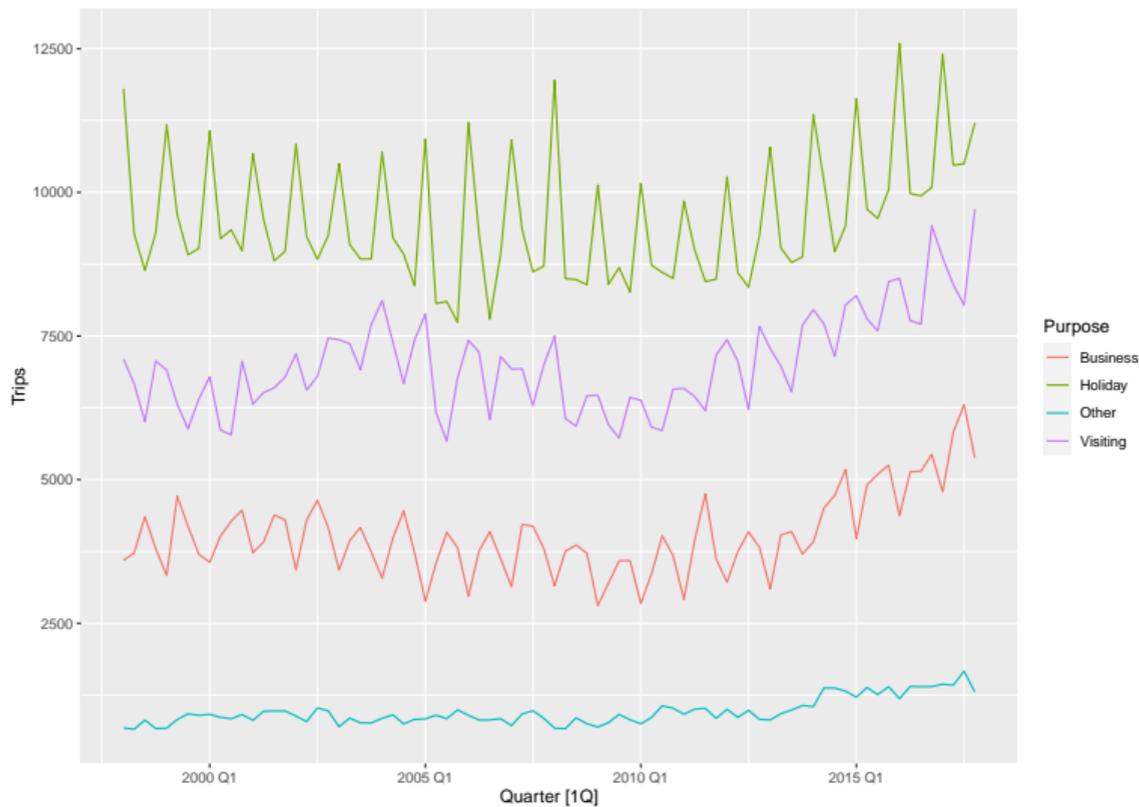
```
## # A tibble: 4 x 6  
##   Purpose    `0%`  `25%`  `50%`  `75%`  `100%`  
##   <chr>      <dbl> <dbl> <dbl>  <dbl> <dbl>  
## 1 Business 2807. 3616. 3926. 4361. 6305.  
## 2 Holiday 7737. 8730. 9235. 10216. 12593.  
## 3 Other    656.  818.  899. 1020. 1665.  
## 4 Visiting 5671. 6425. 6954. 7521. 9703.
```

# Gráficos de séries de tempo no R

A função `autoplot` utiliza o pacote `ggplot2` para plotar as séries de tempo, já considerando os grupos diferentes definidos com a função `as_tsibble`:

```
data %>% autoplot(Trips)
```

# Gráficos de séries de tempo no R

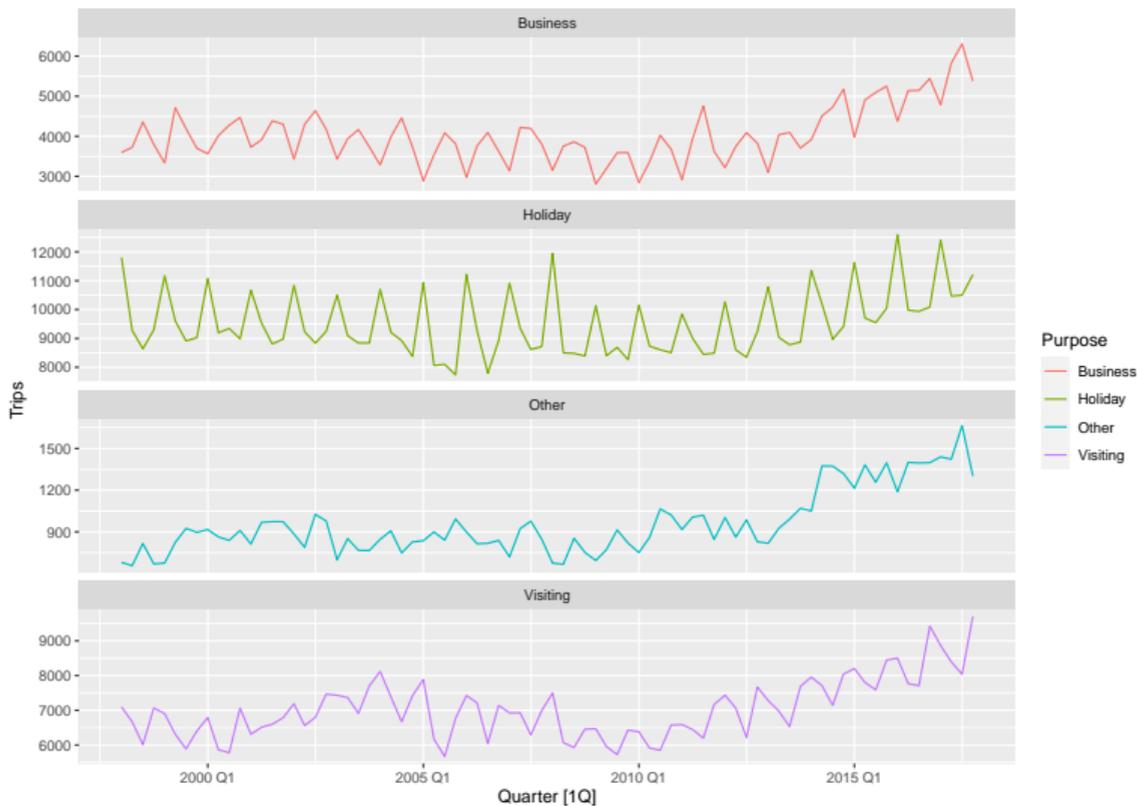


# Gráficos de séries de tempo no R

Podemos dividir o gráfico em 4 painéis. Isso é útil quando os valores das séries de tempo possuem diferenças significativas

```
data %>%  
  autoplot(Trips) +  
  facet_wrap(~Purpose, scales = "free_y", ncol = 1)
```

# Gráficos de séries de tempo no R



# Gráficos de séries de tempo no R

Com os gráficos anteriores, podemos observar que:

1. As séries possuem sazonalidade anual:
  - O número de viagens de lazer e visitas a amigos e família costuma ser maior no primeiro trimestre
  - O número de viagens a negócio é maior no segundo trimestre
2. Houve um crescimento no número de viagens para todos os tipos no período mais recente da amostra

# Componentes de uma série de tempo

Uma série de tempo pode ser descrita por três componentes<sup>2</sup>:

$$Y_t = T_t + S_t + \varepsilon_t$$

Onde:

- $T_t$  é um componente de tendência
- $S_t$  é um componente sazonal
- $\varepsilon_t$  é um componente aleatório, com média zero e variância constante

Os componentes também podem ser multiplicativos:  $Y_t = T_t S_t \varepsilon_t$

---

<sup>2</sup>Estes componentes são dependentes entre si, de modo que a maneira como a tendência é especificada pode afetar a estimação da sazonalidade. Por isso será necessário estimá-los conjuntamente.

# Decomposição de séries de tempo no R

- O método clássico de decomposição de séries de tempo é um procedimento que utiliza médias móveis para estimar cada um dos componentes
- No R podemos estimá-lo através do código:

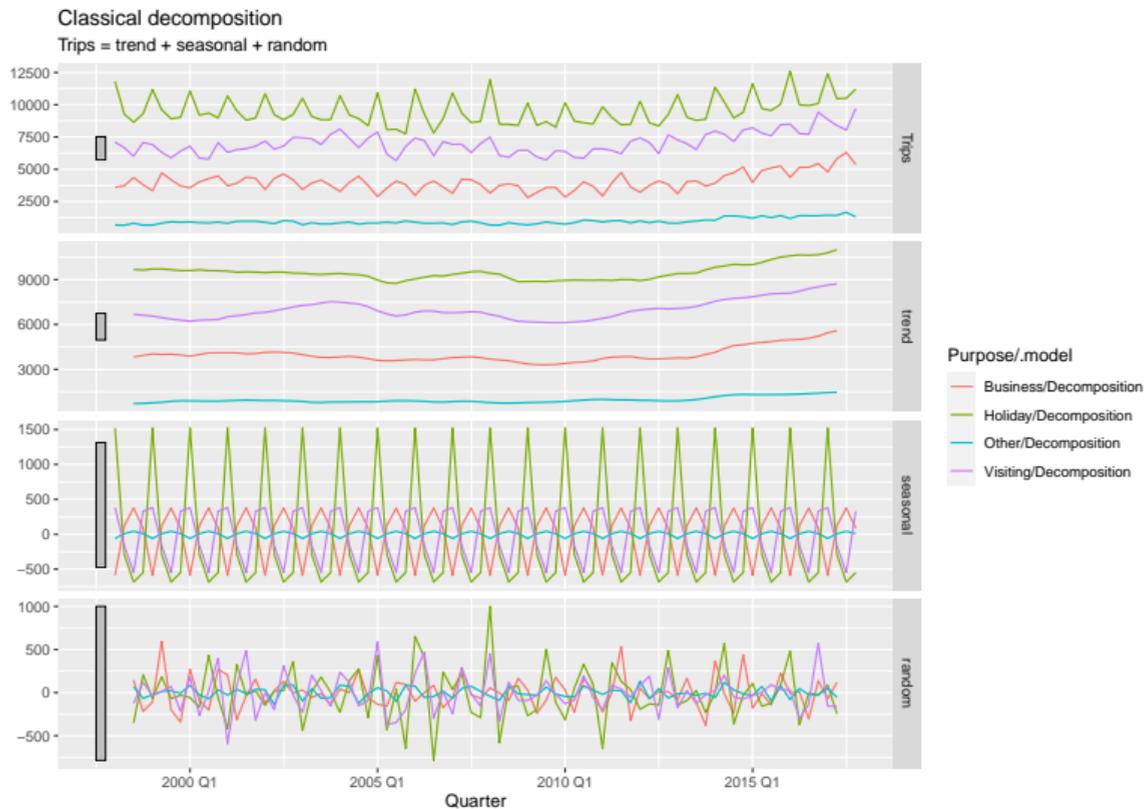
```
data %>%  
  model(  
    classical_decomposition(Trips, type = "additive")  
  ) %>%  
  components() %>%  
  autoplot()
```

# Decomposição de séries de tempo no R

No código do slide anterior:

- A função `model` é utilizada sempre que quisermos estimar algum modelo para todas as séries de um `tsibble`
- A função `classical_decomposition` utiliza a método clássico de decomposição na variável `Trips`
- A função `components` retorna os componentes estimados pelo modelo clássico
- A função `autoplot` plota os componentes das séries de tempo

# Decomposição de séries de tempo no R



# Sazonalidade

Para construirmos uma série de tempo com ajuste sazonal, devemos estimar um componente sazonal,  $\hat{S}_t$ , e então subtrair da série original:

$$Y_t^{SA} = Y_t - \hat{S}_t$$

Observações:

- Existem diversos métodos de estimação de  $\hat{S}_t$
- A estimação do componente de tendência pode afetar a estimação de  $\hat{S}_t$
- Não podemos isolar um dos componentes sem isolar o outro
- Assim, os métodos que veremos costumam estimar tanto a sazonalidade quanto a tendência

# Inspeção de sazonalidade no R

Para avaliar a sazonalidade no R, podemos plotar um gráfico com a evolução de cada série por ano<sup>3</sup>:

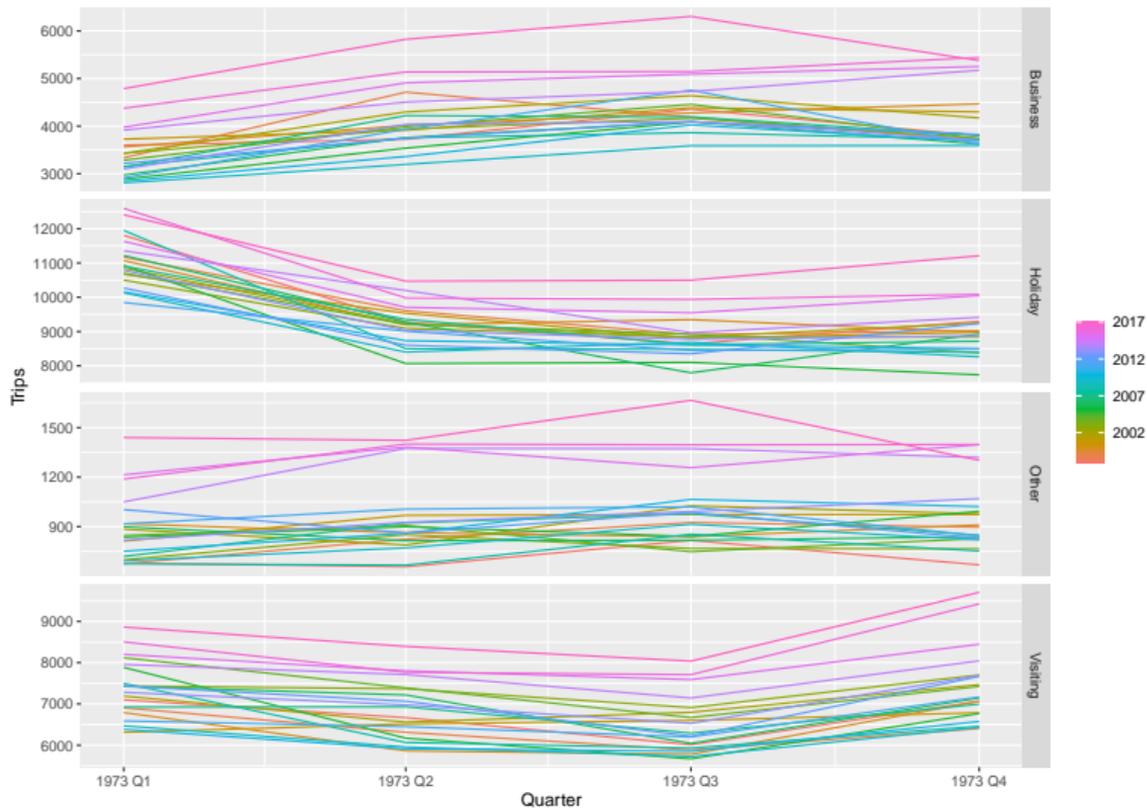
```
data %>%  
  gg_season(Trips)
```

Assim, podemos confirmar que as viagens de lazer e de visitas são maiores no primeiro e quarto trimestres para quase todos os anos, enquanto que as viagens a negócio são maiores segundo e terceiro trimestres

---

<sup>3</sup>O formato `tsibble` identifica automaticamente o período sazonal, entretanto você pode especificar manualmente pelo argumento `period` nas várias funções que utilizaremos, especialmente no caso de sazonalidade múltipla

# Inspeção de sazonalidade no R

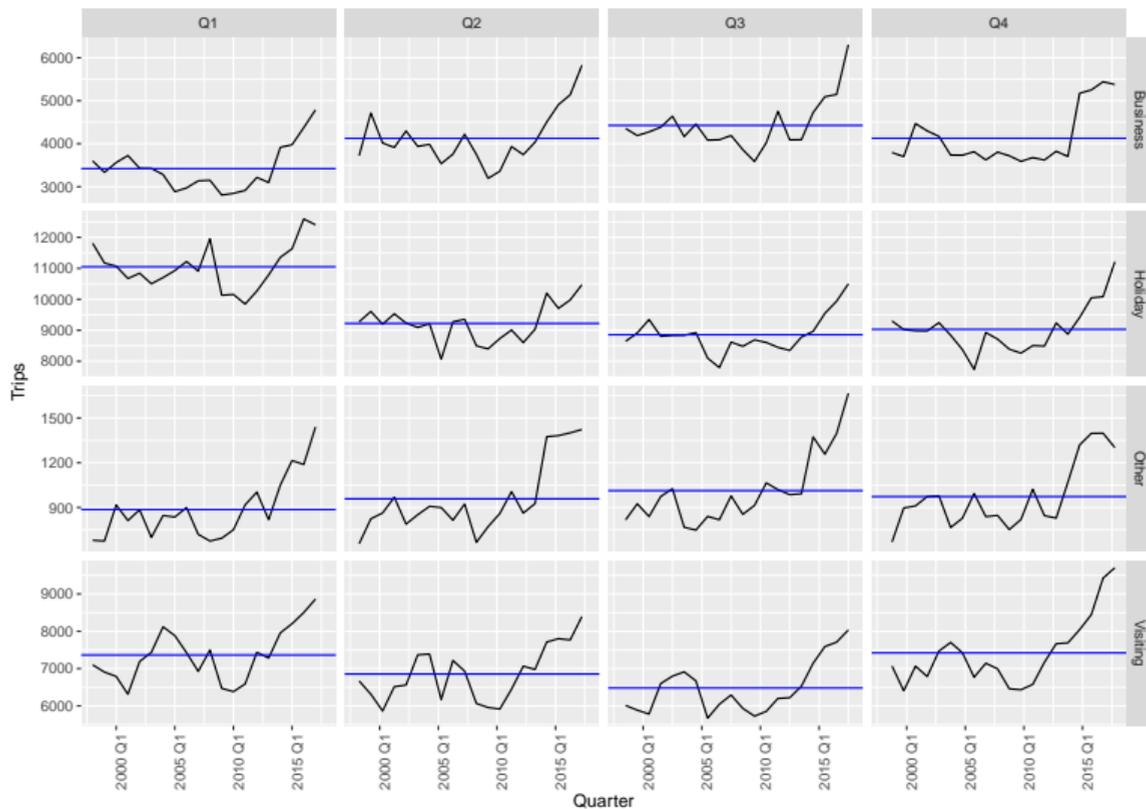


# Inspeção de sazonalidade no R

Outra alternativa é plotar a evolução anual de cada série fixando o trimestre. Podemos fazer isso com a função `gg_subseries`:

```
data %>%  
  gg_subseries(Trips)
```

# Inspeção de sazonalidade no R



# Períodos sazonais

- O período sazonal se refere ao tempo necessário para que um ciclo inteiro do padrão sazonal se repita. Os períodos sazonais mais comuns são:
  - *Sazonalidade Anual*: usual com dados de frequência mensal ou trimestral
  - *Sazonalidade Semanal*: usual com dados de frequência diária
  - *Sazonalidade Diária*: usual com dados de frequência intra-diária
- Uma série de tempo pode ter sazonalidade múltipla
  - O caso mais comum é de sazonalidades anual e semanal em uma série com frequência diária

# Tipos de sazonalidade

Podemos diferenciar a sazonalidade em dois tipos:

- **Sazonalidade determinística:** quando o componente sazonal é constante (*ex: médias sazonais constantes*)
- **Sazonalidade estocástica:** componente sazonal varia com o tempo

# Estimação de sazonalidade

Os métodos mais comuns para estimar sazonalidade são:

- **Métodos de regressão:**  $Y_t = \alpha + \sum_{i=1}^{s-1} \gamma_i D_{it} + \varepsilon_t$ 
  - $D_{it}$  são *dummies* sazonais e  $s$  é o número de observações para completar um ciclo sazonal inteiro (ex: 12 meses)
  - O valor previsto desta regressão será a sazonalidade estimada,  $\hat{S}_t$
  - Este método é útil para sazonalidade determinística
  - Se a série tiver tendência, ela deve ser especificada na regressão
- **Método de médias móveis:**  $\hat{S}_j = \bar{Y}^s - \bar{Y}$ 
  - $\bar{Y}^s$  é a média sazonal,  $\bar{Y}^s = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}$  para  $j = 1, \dots, 12$
  - $\bar{Y}$  é a média temporal
  - Este método é útil para sazonalidade estocástica
  - Se a série tiver tendência, ela deve ser removida antes do cálculo
  - Este método é semelhante a decomposição clássica

# Estimação de sazonalidade

- **STL**: método de decomposição baseado em suavização local por LOESS<sup>4</sup> (Cleveland et al., 1990)
- **X-11-ARIMA**: método original de ajuste sazonal desenvolvido por *US Census Bureau* e *Canada National Statistical Agency*
- **X-12-ARIMA**: evolução do método X-11-ARIMA desenvolvido pelo *US Census Bureau*
- **X-13ARIMA-SEATS**: procedimento desenvolvido pelo Banco da Espanha<sup>5</sup>
- **SARIMA**: método que incorpora a sazonalidade na estimação de modelos ARIMA<sup>6</sup>

---

<sup>4</sup>Locally Estimated Scatterplot Smoothing

<sup>5</sup>Os métodos X11 a X13 são utilizados para séries mensais ou trimestrais

<sup>6</sup>Maiores detalhes sobre estes métodos em Dagum e Bianconcini (2016)

# Estimação de sazonalidade

Vamos estimar alguns destes modelos no R e salvar em novas variáveis:

```
data_stl <- data %>%  
  model(STL = STL(Trips)) %>%  
  components()  
data_x11 <- data %>%  
  model(X11 = feasts::X11(Trips)) %>%  
  components()  
data_seats <- data %>%  
  model(SEATS = feasts::SEATS(Trips)) %>%  
  components()
```

# Estimação de sazonalidade

Podemos unir a base de dados original com cada uma das séries ajustadas sazonalmente:

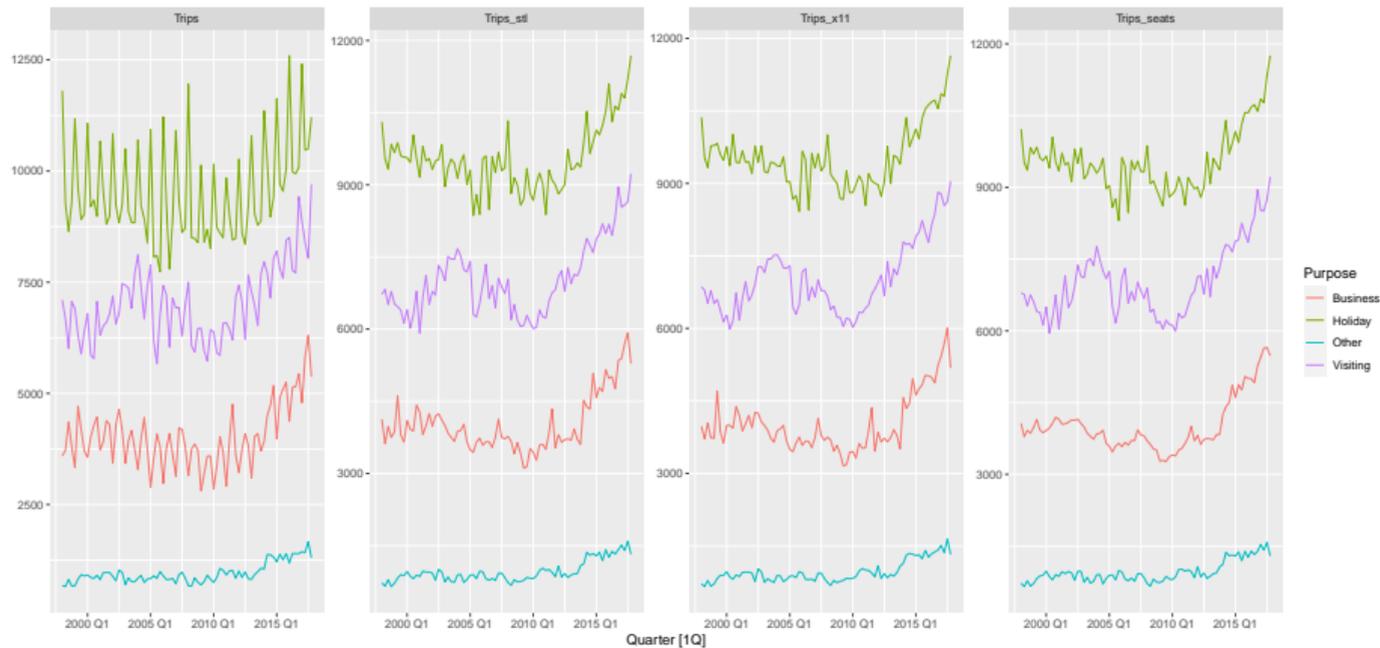
```
data_adj <- data %>%  
  bind_cols(  
    as_tibble(data_st1) %>%  
      select(Trips_st1 = season_adjust),  
    as_tibble(data_x11) %>%  
      select(Trips_x11 = season_adjust),  
    as_tibble(data_seats) %>%  
      select(Trips_seats = season_adjust)  
  )
```

# Estimação de sazonalidade

Podemos plotar as séries originais e os ajustes sazonais para cada grupo e método utilizando a função `autoplot`:

```
data_adj %>%  
  autoplot(  
    vars(Trips, Trips_st1, Trips_x11, Trips_seats)  
  )
```

# Estimação de sazonalidade



# Tendência

Para removermos a tendência de uma série de tempo devemos estimar o componente  $\hat{T}_t$ , e então subtrair da série original:

$$Y_t^{TA} = Y_t - \hat{T}_t$$

Observações:

- Podemos estimar  $\hat{T}_t$  de diversas maneiras, dependendo dos pressupostos sobre o comportamento da tendência
- A estimação do componente sazonal pode afetar a estimação do componente de tendência,  $\hat{T}_t$
- Os métodos que vimos para estimar sazonalidade também estimam o componente de tendência conjuntamente

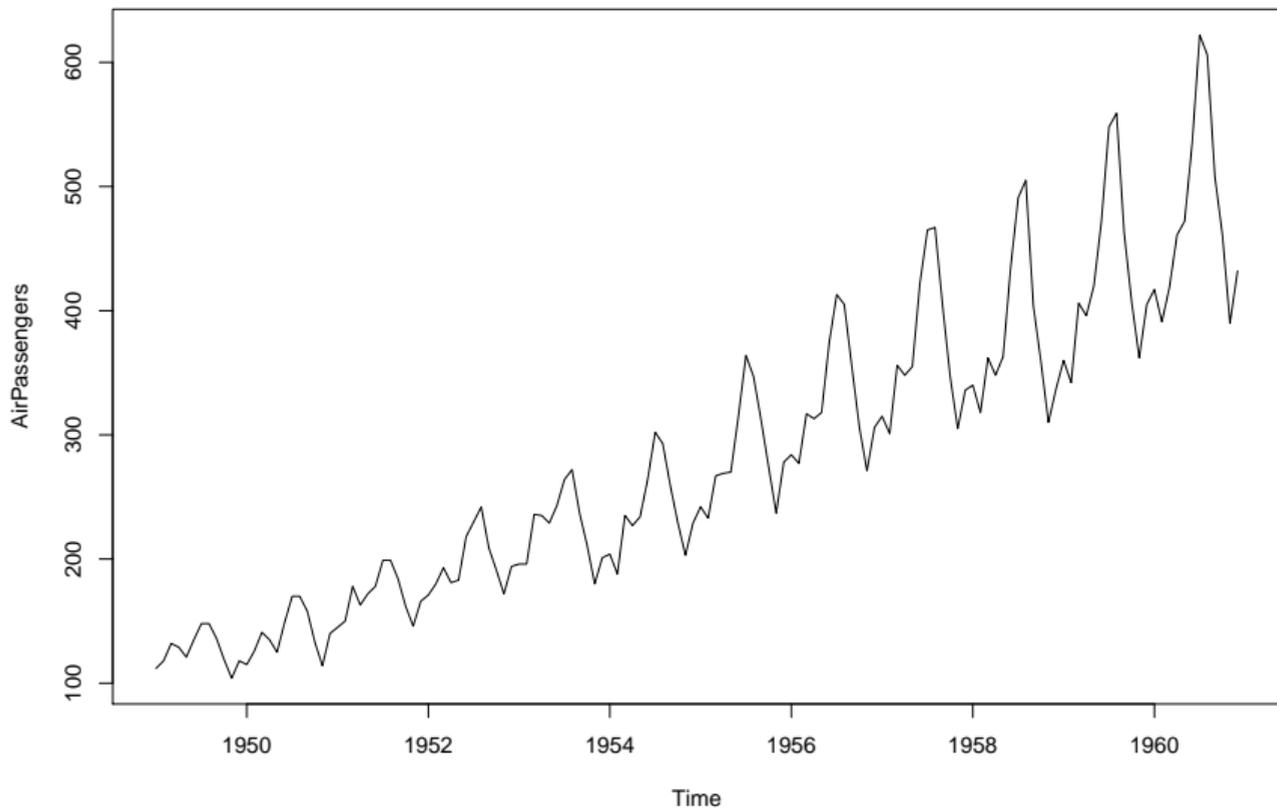
# Inspeção de tendência no R

O método mais rápido de verificar se uma série apresenta tendência é através da inspeção gráfica

- Vamos plotar uma série do número de passageiros mensais de empresas aéreas internacionais utilizada no livro de Box et al. (1976)

```
plot(AirPassengers)
```

# Inspeção de tendência no R



# Inspeção de tendência no R

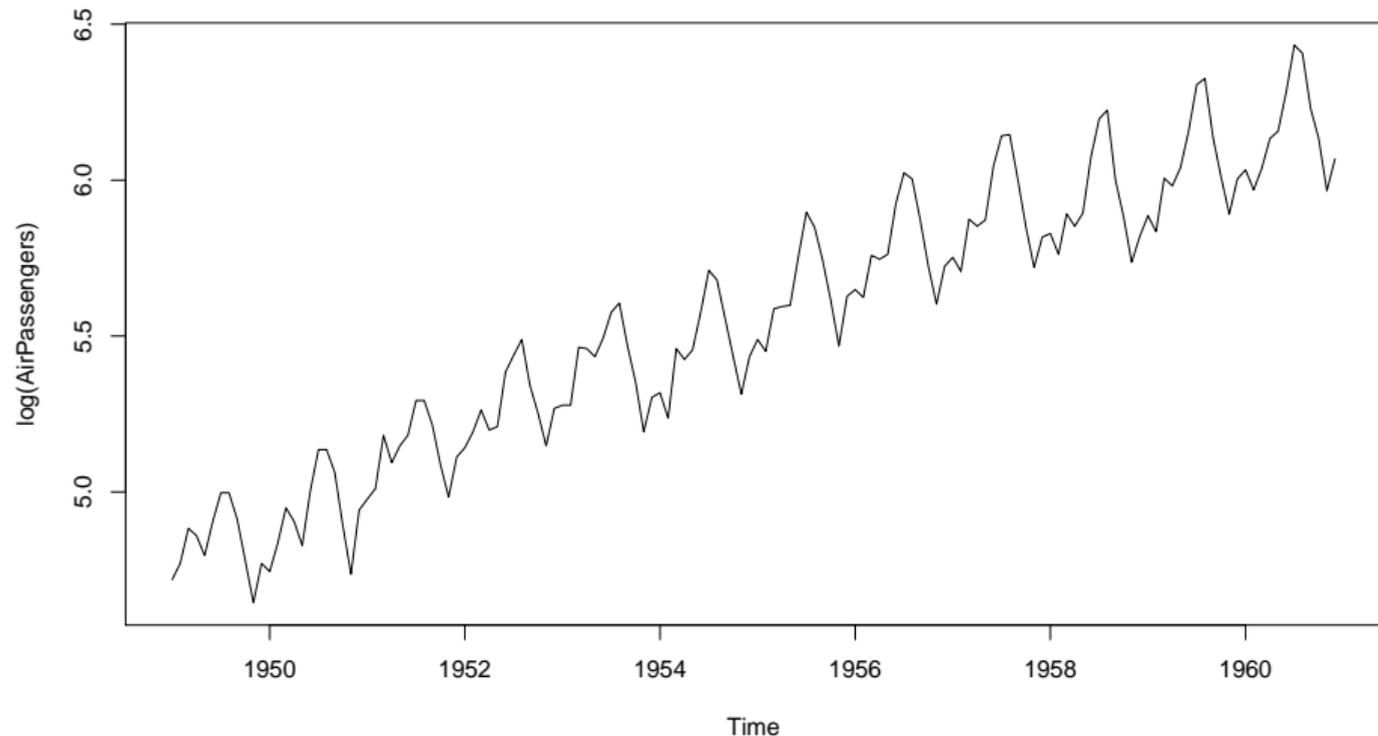
A partir do gráfico anterior, podemos observar que:

1. A série aumenta significativamente ao longo do tempo, deixando clara a presença de uma tendência,
2. A tendência parece aumentar de maneira mais rápida do que linearmente,
3. A variabilidade (variância) dos dados aumenta ao longo do tempo.

Uma maneira de lidar com os pontos 2 e 3 é utilizando o logaritmo natural, que suaviza a variância dos dados e transforma tendências exponenciais em lineares

# Inspeção de tendência no R

```
plot(log(AirPassengers))
```



# Inspeção de tendência no R

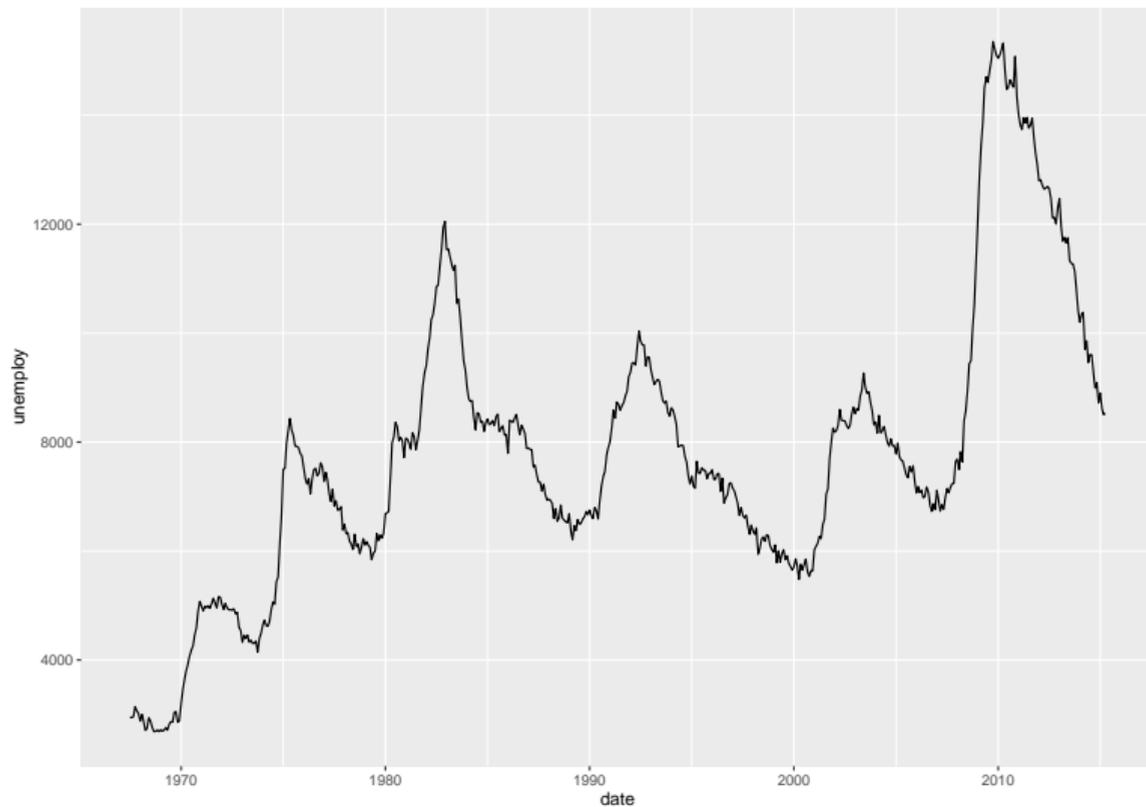
Outro exemplo de tendência mais errática pode ser observada na série do número de desempregados nos Estados Unidos

- Esta série pode ser acessada através da coluna `unemploy` da variável `economics`, carregada junto com o pacote `ggplot2`

```
economics %>%  
  ggplot(aes(x = date, y = unemploy)) +  
  geom_line()
```

No gráfico do próximo slide pode-se ver que a tendência desta série parece mudar ao longo do tempo, tendo um comportamento estocástico

# Inspeção de tendência no R



# Características

Há dois tipos principais de tendência que podemos observar em séries de tempo:

1. **Tendência determinística:** comportamento regular ao longo do tempo, normalmente caracterizado por tendências lineares ou polinomiais
2. **Tendência estocástica:** tendências mudam ao longo do tempo, devendo ser removidas através de médias móveis, regressões locais ou diferenças

# Estimação de tendência determinística

Supondo que a série não tenha nenhum componente sazonal,  $Y_t = T_t + \varepsilon_t$ , podemos estimar uma tendência determinística utilizando uma regressão linear de  $Y_t$  contra uma função polinomial no tempo<sup>7</sup>:

$$Y_t = \beta_0 + \beta_1 t + \cdots + \beta_k t^k + \varepsilon_t$$

Assim, o valor previsto desta regressão será a tendência estimada,  $\hat{T}_t$

---

<sup>7</sup>A variável  $t$  tem valores  $1, 2, \dots, T$ , de acordo com a unidade de tempo da observação

# Estimação de tendência determinística

Para estimar uma tendência linear ao logaritmo da série de tempo `AirPassengers` no R, devemos:

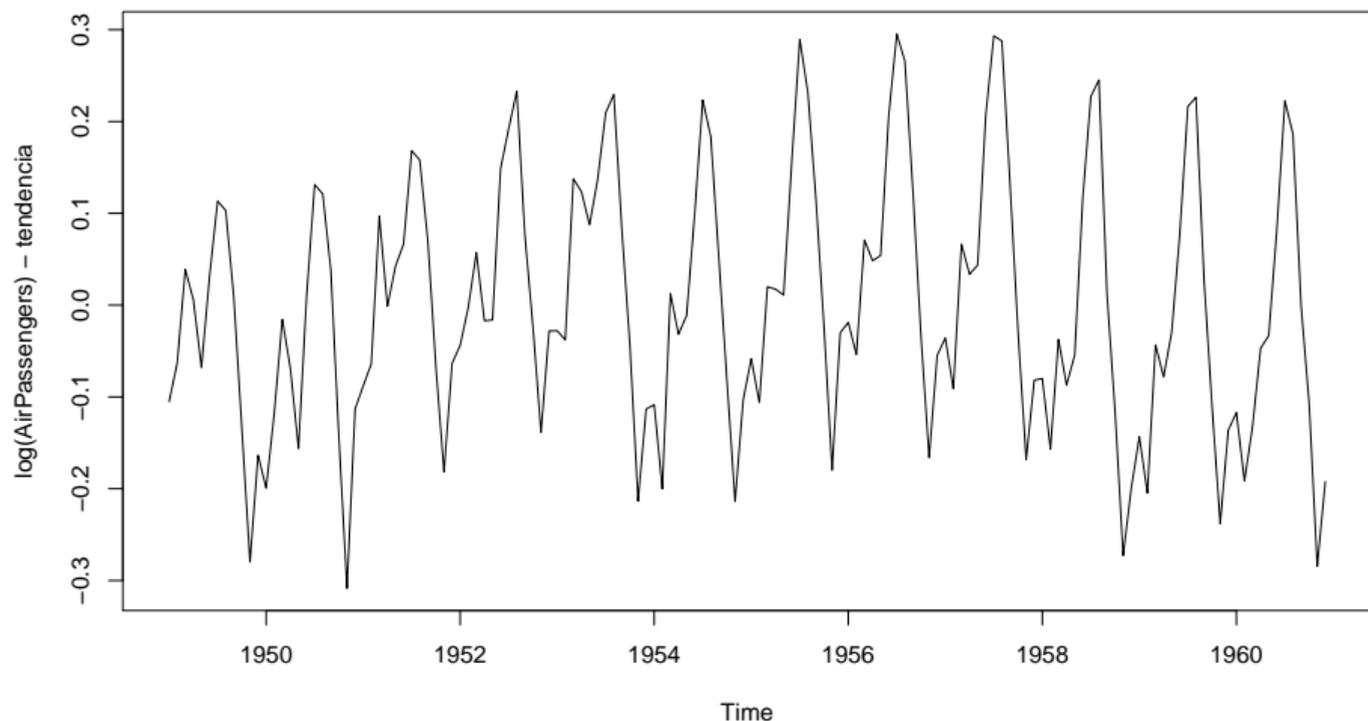
1. Criar uma variável temporal, que vai de 1 até T;
2. Estimar uma regressão com a função `lm`;
3. Obter os valores previstos dessa regressão

```
tempo <- c(1:length(AirPassengers))  
trend_reg <- lm(log(AirPassengers) ~ tempo)  
tendencia <- trend_reg$fitted
```

Agora podemos calcular a diferença entre a série original e a tendência para obter a série ajustada

# Estimação de tendência determinística

```
ts.plot(log(AirPassengers) - tendencia)
```



# Estimação de tendência determinística

Observando o gráfico anterior, podemos notar que:

1. A tendência linear parece não captar perfeitamente as mudanças ao longo do tempo
2. O componente sazonal da série pode estar influenciando na estimação da tendência

# Estimação de tendência estocástica

Entre os métodos para estimar tendências estocásticas podemos citar:

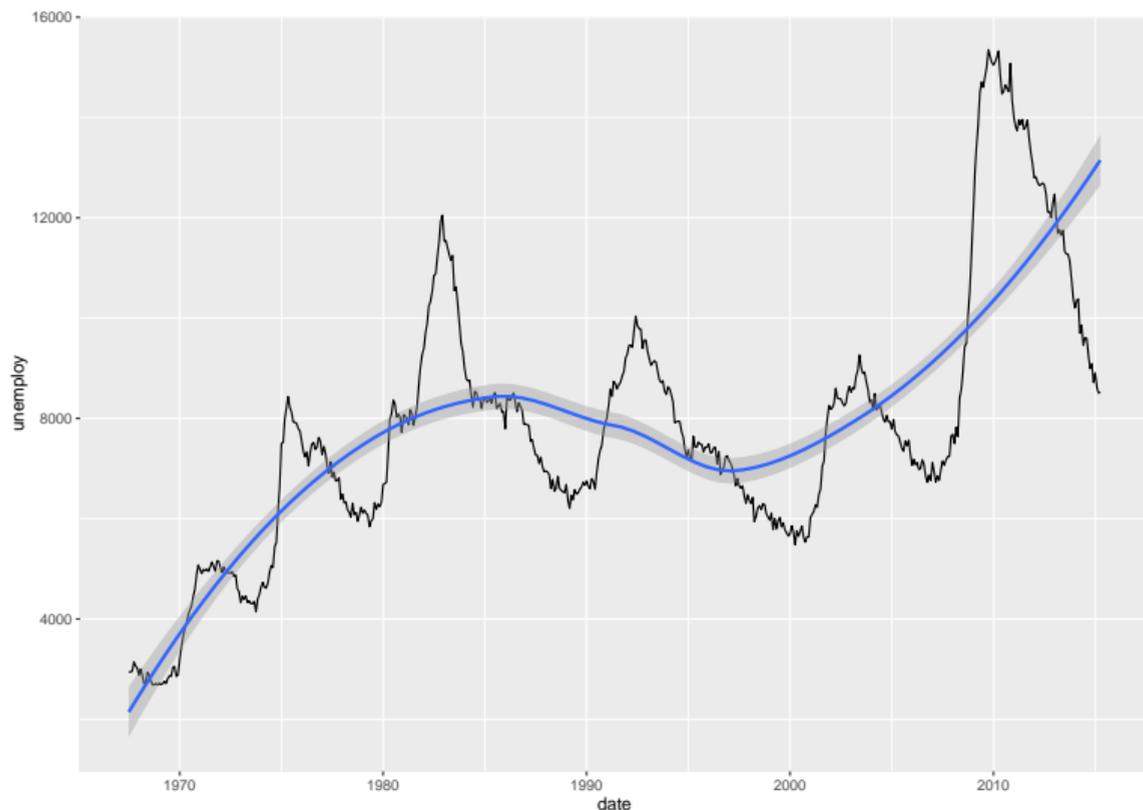
1. **Suavização**: métodos de aproximação com médias móveis ou regressões locais, como LOESS, LOWESS, splines, entre outros
2. **Diferenciação**: utilização da diferença de primeira ordem ou ordem maiores

# Estimação de tendência estocástica

Para estimar a tendência estocástica da série do número de desempregados nos Estados Unidos, podemos adicionar a função `geom_smooth()` ao gráfico dos dados

```
economics %>%  
  ggplot(aes(x = date, y = unemploy)) +  
  geom_line() +  
  geom_smooth(method = "loess")
```

# Estimação de tendência estocástica



# Estimação de tendência estocástica

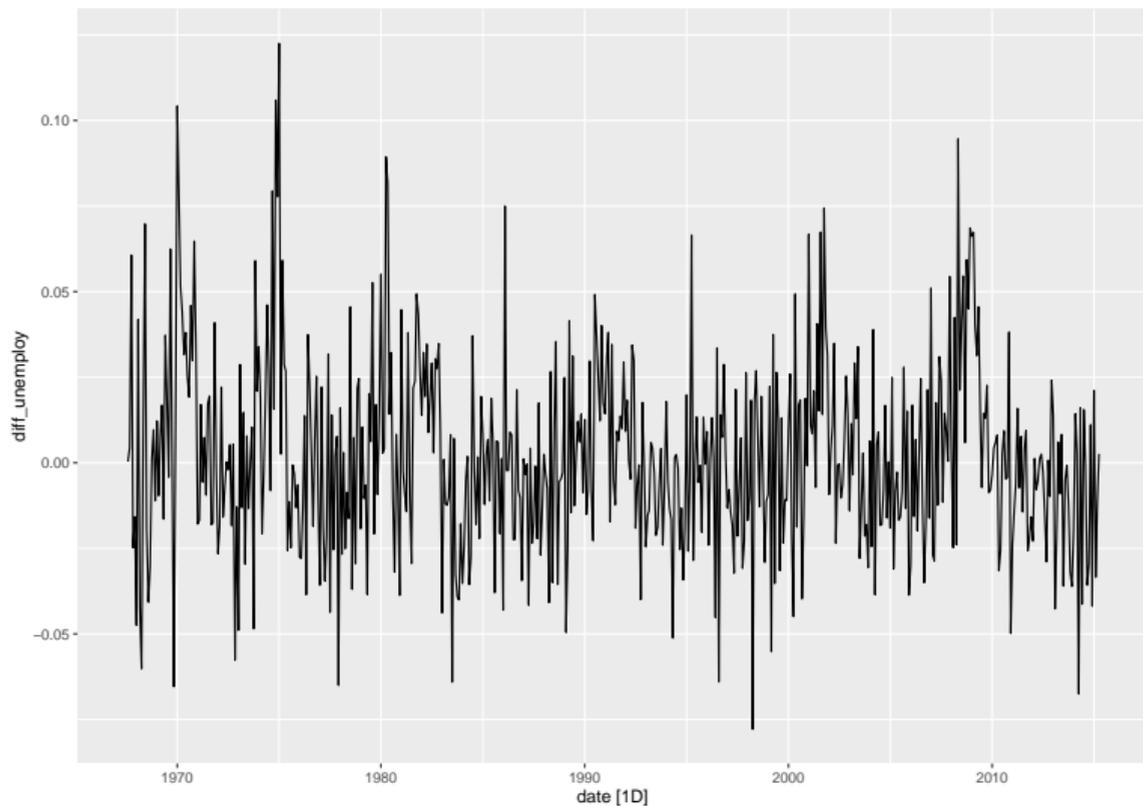
Para calcularmos a diferença da série de desemprego iremos primeiro passar o logaritmo nos dados para suavizar a variância, e depois utilizar a função `difference` do pacote `tsibble`<sup>8</sup>

```
economics %>%  
  as_tsibble(index = date) %>%  
  mutate(diff_unemploy = difference(log(unemploy))) %>%  
  autoplot(diff_unemploy)
```

---

<sup>8</sup>Também é possível calcular a diferença dos logs utilizando a função `lag`, através do comando `log(unemploy) - lag(log(unemploy))`

# Estimação de tendência estocástica



# Estimação de tendência estocástica

- Em geral, o método da diferenciação é a maneira mais usual de remover tendências em séries de tempo
- Em algumas situações, a interpretação da série diferenciada pode ser prejudicada (taxa de juros, razões contábeis, proporções, etc)
- A decisão entre tirar a diferença ou não da série deve depender do modelo econométrico e dos objetivos do pesquisador
- É importante confirmar que a tendência é determinística antes que recorrer a métodos polinomiais (usualmente por inspeção gráfica)

# Diferenciação

A primeira diferença remove tendências lineares. Considere

$T_t = \beta_0 + \beta_1 t$ , então:

- $\Delta T_t = T_t - T_{t-1} = (\beta_0 + \beta_1 t) - [\beta_0 + \beta_1(t - 1)] = \beta_1$

Em geral,  $d$  diferenças eliminam tendências polinomiais de ordem  $d$ :

$$\Delta^d T_t = \begin{cases} d! \beta_d & \text{se } m = d \\ 0 & \text{se } m < d \end{cases}$$

onde  $m$  é o grau do polinômio da tendência

- Se tomarmos menos do que  $d$  diferenças, a série continuará com tendência

# Diferenciação

Ao tomarmos diferenças removemos tendências mas introduzimos autocorrelação. Suponha um ruído branco  $Z_t = \varepsilon_t \sim N(0, \sigma^2)$ , então:

- $\Delta Z_t = \varepsilon_t - \varepsilon_{t-1}$
- $\Delta Z_{t-1} = \varepsilon_{t-1} - \varepsilon_{t-2}$

Assim,  $Cov(\Delta Z_t, \Delta Z_{t-1}) = -\sigma^2$ .<sup>9</sup>

- Séries econômicas apresentam um comportamento exponencial com frequência, por isso é comum aplicarmos o logaritmo antes de realizarmos a diferenciação. Isso ajuda a estabilizar a variância, além de termos uma boa interpretação dos valores (variação percentual continuamente composta)

---

<sup>9</sup>Esta autocorrelação será tratado nos modelos ARIMA

# Teste de sazonalidade

Caso tenhamos dúvidas sobre a presença de sazonalidade após a inspeção gráfica, podemos realizar um teste de sazonalidade com o pacote `seastests`:<sup>10</sup>

```
library(seastests)
data %>% # Valores e P-valores do teste
  group_by(Purpose) %>%
  group_map(~wo(.x$Trips, freq = 4))
data %>% # Vetor TRUE/FALSE
  group_by(Purpose) %>%
  group_map(~isSeasonal(.x$Trips, freq = 4)) %>%
  `names<-` (unique(data$Purpose))
```

---

<sup>10</sup>Este teste é um combinação de vários testes. Maiores detalhes em Ollech e Webel (forthcoming), e Webel and Ollech (forthcoming)

# Teste de sazonalidade

```
## $Business
## [1] TRUE
##
## $Holiday
## [1] TRUE
##
## $Other
## [1] FALSE
##
## $Visiting
## [1] TRUE
```

# Testes de tendência

- Os testes de tendência mais comuns são não paramétricos
  - Baseados em diferenças, sinais ou ranqueamento das observações
- Dois testes de tendência comuns são:
  - *Cox and Stuart Trend Test*: divide a série em três partes e compara se os valores da primeira parte são maiores ou menores que os valores da terceira parte
  - *Mann-Kendall Trend Test*: teste baseado na soma dos sinais das diferenças, que tendem a apresentar pouca variação caso a série tenha tendência
- A hipótese nula de ambos os testes é de que a série não apresenta tendência

# Testes de tendência: Cox-Stuart

O pacote trend contém diversos testes de tendência

```
library(trend)
cs.test(AirPassengers) # Cox and Stuart Trend test

##
## Cox and Stuart Trend test
##
## data: AirPassengers
## z = 6.9282, n = 144, p-value = 4.262e-12
## alternative hypothesis: monotonic trend
```

## Testes de tendência: Mann-Kendall

```
mk.test(AirPassengers) # Mann-Kendall Trend Test

##
## Mann-Kendall trend test
##
## data:  AirPassengers
## z = 14.382, n = 144, p-value < 2.2e-16
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##           S           varS           tau
## 8.327000e+03 3.351643e+05 8.098232e-01
```

# Referências

- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1976) Time Series Analysis, Forecasting and Control. Third Edition. Holden-Day. Series G.
- Cleveland, R. B.; Cleveland, W. S.; McRae J.E., e Terpenning I. (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics, 6, 3–73.
- Dagum, E. B. e Bianconcini, S. (2016). Seasonal adjustment methods and real time trend-cycle estimation. Springer.
- Ollech, Daniel and Webel, Karsten (forthcoming). A random forest-based approach to identifying the most informative seasonality tests. Bundesbank Discussion Paper
- Webel, Karsten and Ollech, Daniel (forthcoming). An overall seasonality test. Bundesbank Discussion Paper