Séries Temporais

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Introdução

- O que é uma série temporal?
- O que é a análise de uma série temporal?
- Quais os objetivos em analisar séries temporais?

- Uma *série temporal* é uma coleção de observações feitas sequencialmente ao longo do tempo.
- A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência.
- A ordem das observações é crucial para a análise.
- Tempo pode ser substituido por outra variável como espaço, profundidade, etc.

Características particulares

- Dados correlacionados são mais difíceis de analisar e requerem técnicas específicas.
- Precisamos levar em conta a *ordem temporal* das observações.
- Fatores complicadores como presença de tendências e variação sazonal ou cíclica podem ser difíceis de estimar ou remover.
- A seleção de modelos pode ser bastante complicada, e as ferramentas podem ser de difícil interpretação.
- É mais difícil de lidar com observações perdidas e dados discrepantes devido à natureza sequencial.
- Representação gráfica da série temporal é sempre o ponto de partida da análise.

Terminologia

- Uma série temporal é *contínua* quando as observações são feitas continuamente no tempo. A série temporal será denotada por $\{X(t) : t \in T\}$, $T = \{t : t_1 < t < t_2\}$.
- Uma série temporal é *discreta* se as observações são feitas em tempos específicos, geralmente equiespaçados. A série temporal será denotada por $\{X_t : t \in T\}$, $T = \{t_1, \dots, t_n\}$.
- A variável observada X pode assumir valores discretos ou contínuos.
- X pode ser discreta por definição mas ser tratada como continua se os valores observados não forem muito pequenos.

- Séries contínuas podem ser *discretizadas* se os valores são registrados a certos intervalos de tempo.
- Séries podem ser valores agregados ou acumulados em intervalos de tempo, e.g. exportações medidas mensalmente ou quantidade de chuva medida diariamente.
- Algumas séries são inerentemente discretas, e.g. dividendos pagos por uma empresa aos seus acionistas em anos sucessivos.
- A série temporal é multivariada se k variáveis são observadas a cada tempo $\{X_{1t}, \dots, X_{kt}, t \in T\}$. As séries correlacionadas devem ser analisadas conjuntamente e em cada tempo tem-se um vetor de observações.

Objetivos

- Descrever propriedades da série, e.g. o padrão de tendência, existência de variação sazonal ou cíclica, observações discrepantes (*outliers*), alterações estruturais (e.g. mudanças no padrão da tendência ou da sazonalidade), etc.
- Usar a variação em uma série para explicar a variação em outra série.
- Predizer valores futuros com base em valores passados. Assume-se que o futuro envolve incerteza e as previsões não são perfeitas. Porém devemos tentar reduzir os erros de previsão.

Abordagens

- Técnicas Descritivas. Técnicas gráficos, identificação de padrões, etc.
- Modelos Probabilísticos. Seleção, comparação e adequação de modelos, estimação, predição. Abordagens Clássica e/ou Bayesiana.
- Análise espectral.
- Métodos não paramétricos (alisamento ou suavização).
- Outras Abordagens. Modelos de espaço de estados, modelos não lineares, séries multivariadas, estudos longitudinais, processos de longa dependência, modelos para volatilidade, etc.

Sazonalidade

- Comportamento da série temporal tende a se repetir a cada s períodos de tempo.
- Por exemplo, vendas mensais de brinquedos terão um pico no mês de dezembro e talvez um pico secundário em outubro.
- Este padrão possivelmente se repetirá ao longo de vários anos.

- Sazonalidade deterministica. Variáveis *dummies* (binárias). O coeficiente de cada variável *dummy* representa o fator sazonal do respectivo mês, trimestre, etc.

- Funções trigonométricas.

- Sazonalidade estocástica:
    - Variável endógena com defasagem sazonal no modelo (modelos ARMA periódicos),
    - modelo ARMA sazonal.

Tipos de Sazonalidade

- Aditiva. A série apresenta flutuações sazonais mais ou menos constantes não importando o nível global da série.
- Multiplicativa. O tamanho das flutuações sazonais varia dependendo do nível global da série.

Tendência

Globalmente, uma série pode exibir tendência de crescimento (ou decrescimento) com vários possíveis padrões.

- Crescimento linear. Por exemplo, a cada ano o aumento esperado nas vendas de um certo brinquedo é de 1 milhão de reais.
- Crescimento exponencial. Por exemplo, a cada ano as vendas de um certo brinquedo aumentam de um fator 1.3.
- Crescimento amortecido. Por exemplo, as vendas de um certo brinquedo tem uma aumento esperado de 70% sobre o ano anterior. Se o aumento esperado for de 1 milhão de reais no primeiro ano, no segundo ano será de 700 mil reais, no terceiro ano será de 490 mil reais e assim por diante.

Um modelo é uma simplificação da realidade
(e alguns são úteis)

Quantidades observáveis
(podem ser medidas)

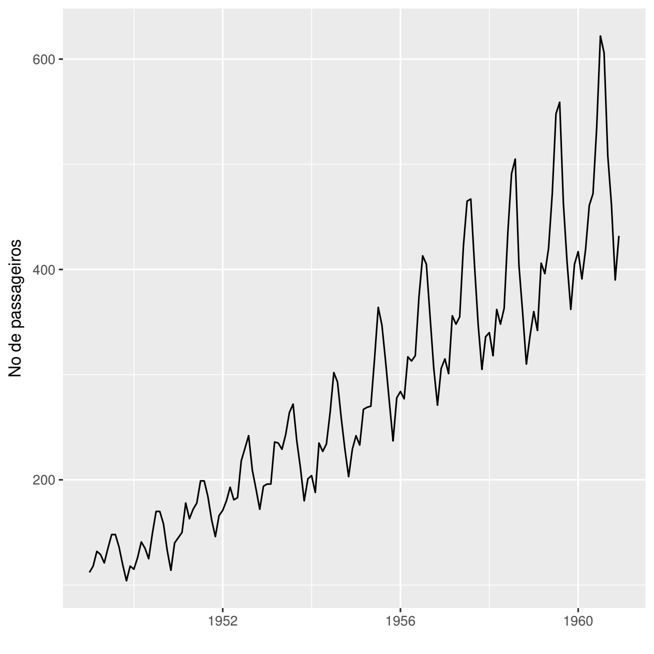
Quantidades não observáveis
(parâmetros e variáveis latentes)

Abordagens: Clássica e Bayesiana

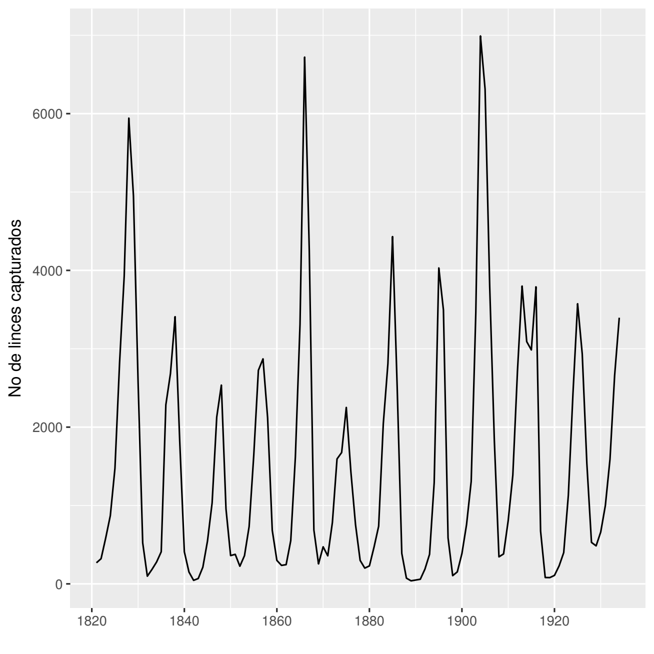
Dados: os valores observados das quantidades observáveis.

Exemplos de Séries Temporais

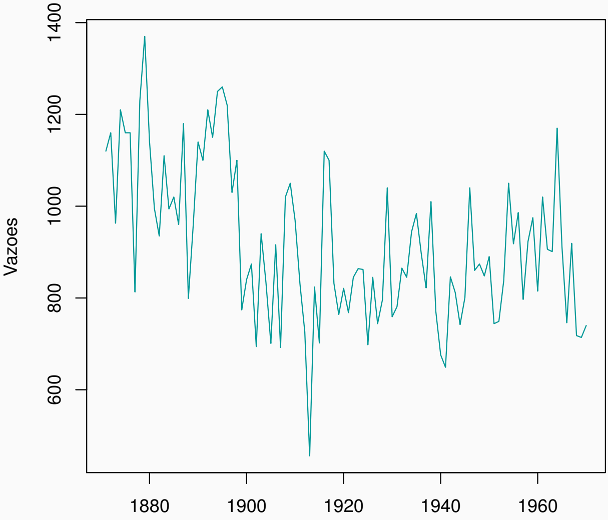
Totais mensais de passageiros em linhas aéreas internacionais nos EUA



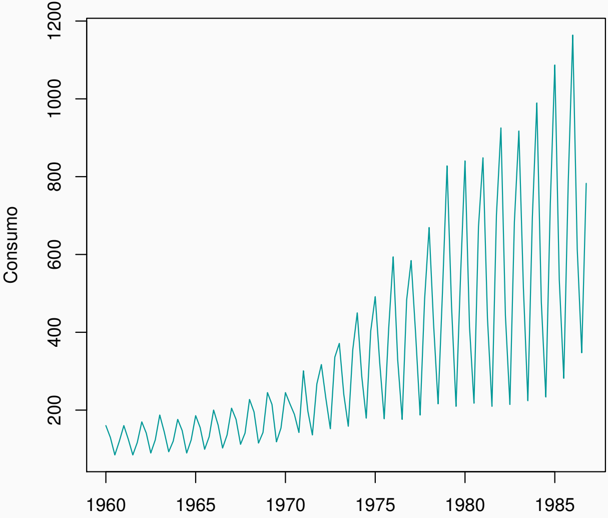
Número anual de linces capturados em armadilhas 1821-1934 no Canadá



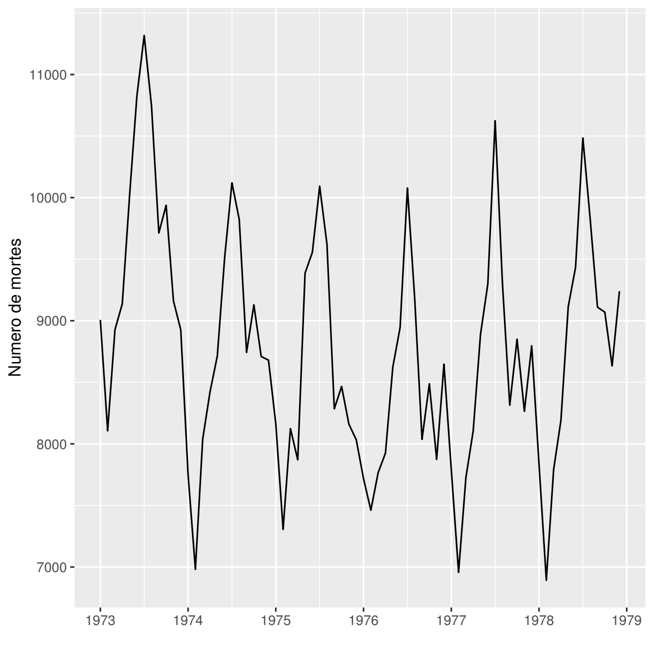
Medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970



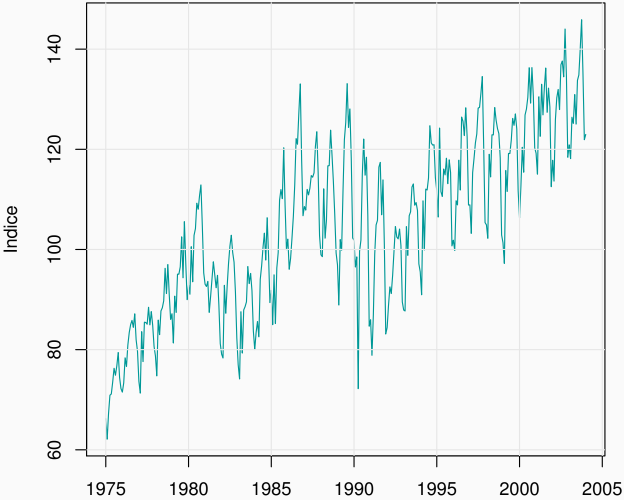
Consumo de gás no Reino Unido entre o primeiro trimestre de 1960 e o quarto trimestre de 1986



Totais mensais de mortes por acidente nos EUA jan/1973 a dez/1978

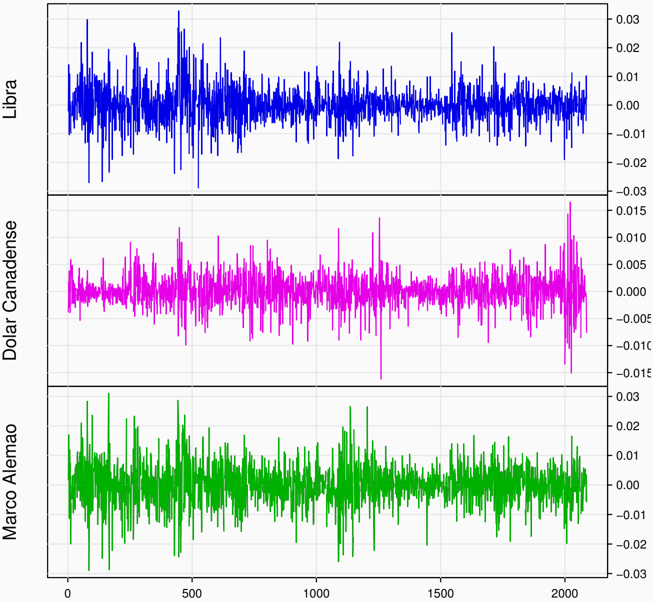


Indice da Produção Industrial do Brasil (indústria geral, média 1991 = 100). Janeiro de 1975 a Janeiro de 2004. Fonte: IBGE

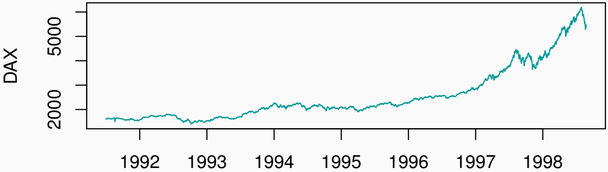


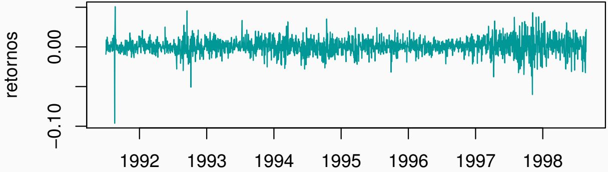
Taxas de câmbio diárias em relação ao Dolar Americano, de janeiro/1991 a dezembro/1998.

Retornos diários das taxas de câmbio em relação ao Dolar Americano, de janeiro/1991 a dezembro/1998.

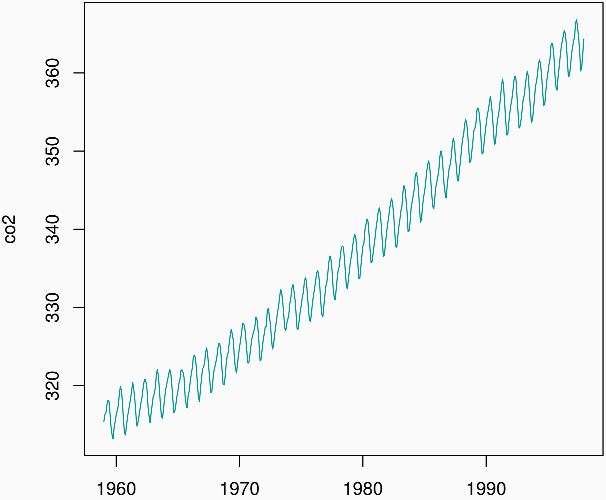


Preços diários no fechamento de um indice de mercado da Alemanha, 1991-1998 e respectivos retornos

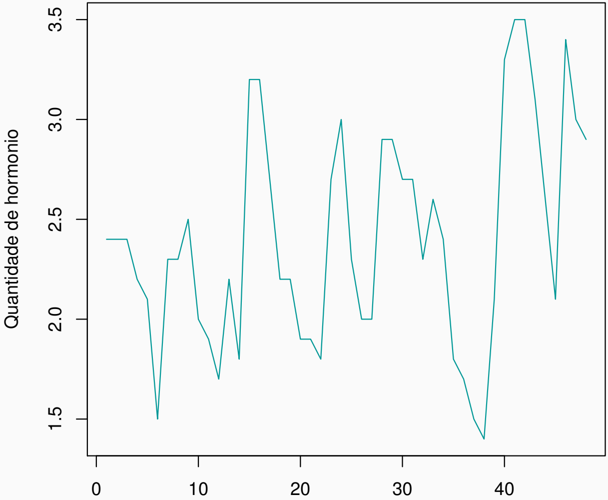




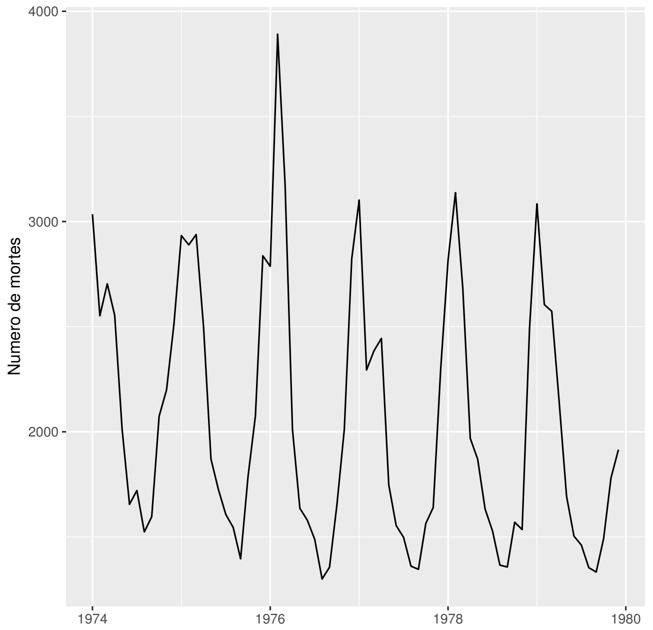
Concentrações atmosfericas de CO2 (em ppm) janeiro/1953 a dezembro/1997

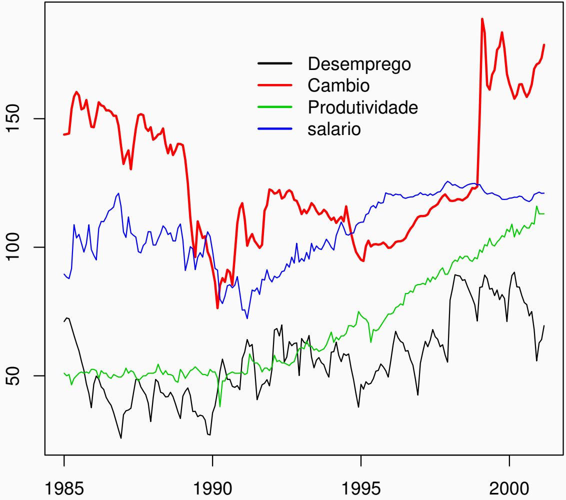


Quantidades de um tipo de hormonio em amostras de sangue coletadas a cada 10 minutos de uma pessoa do sexo feminino



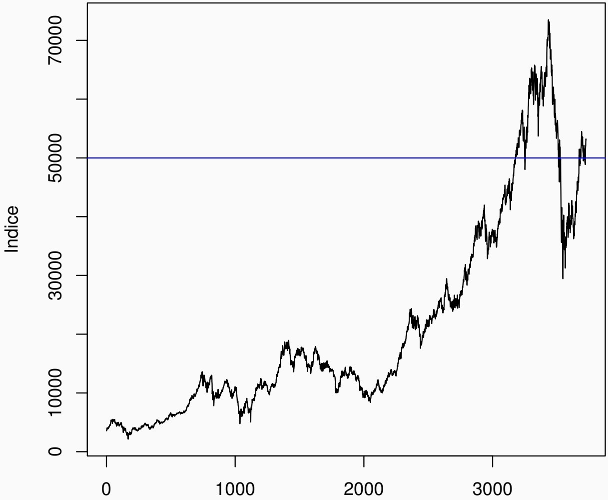
Números mensais de mortes por doenças do pulmão (bronquite, efisema e asma) no Reino Unido entre janeiro de 1974 e dezembro de 1979



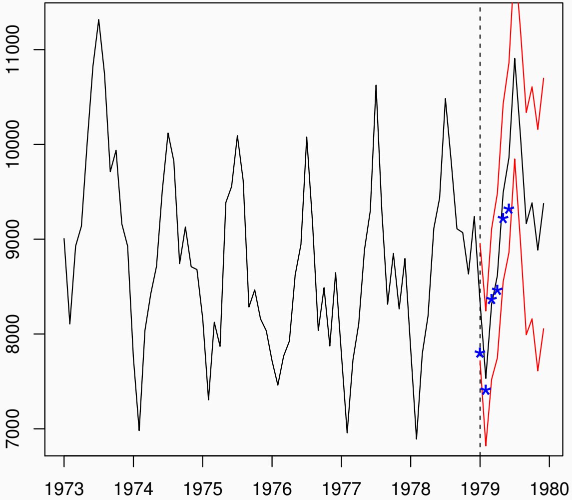


Desemprego
Cambio
Produtividade
salario
50
100
150
1985
1990
1995
2000

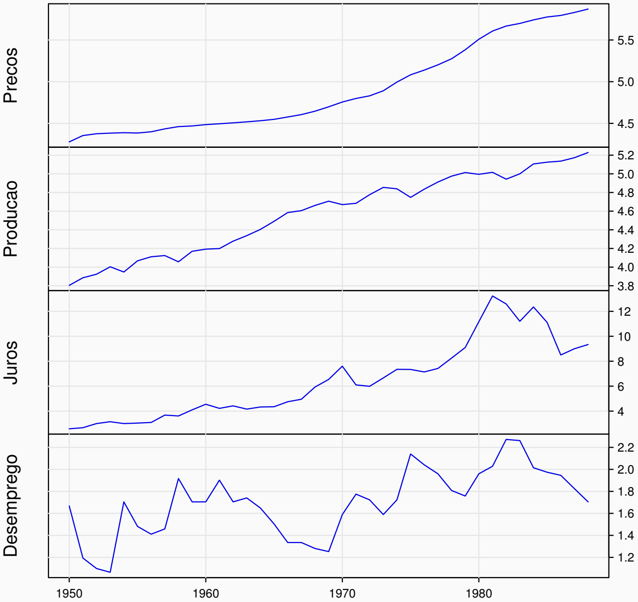
Indice diário da Bolsa de Valores de São Paulo (fechamento) de 04/07/1994 a 22/07/2009



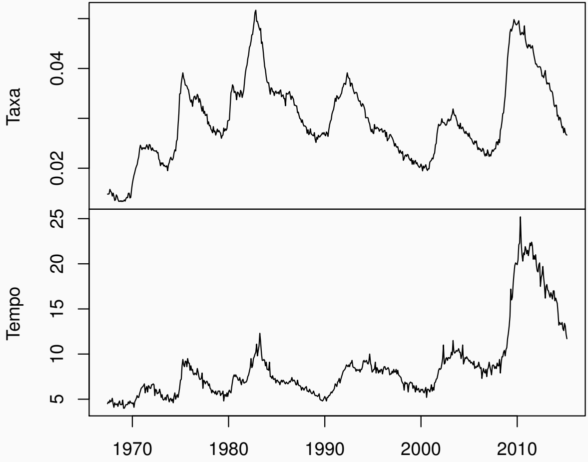
Totais mensais de mortes por acidente nos EUA jan/1973 a dez/1978 e previsões para 1979



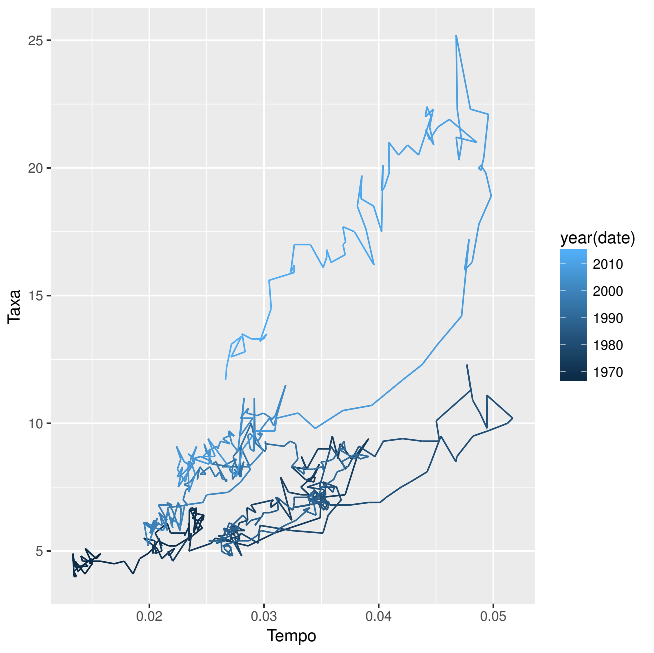
Indice de preços, produção industrial, taxa de juros e desemprego, EUA.
Fonte: *http://www.amstat.org/publications/jbes/*



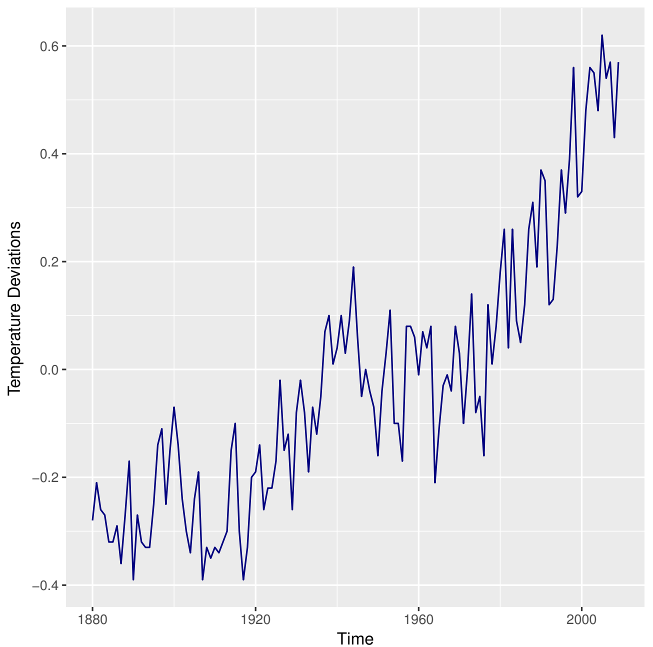
Taxa de desemprego e número mediano de semanas sem emprego (EUA 1967-2007)

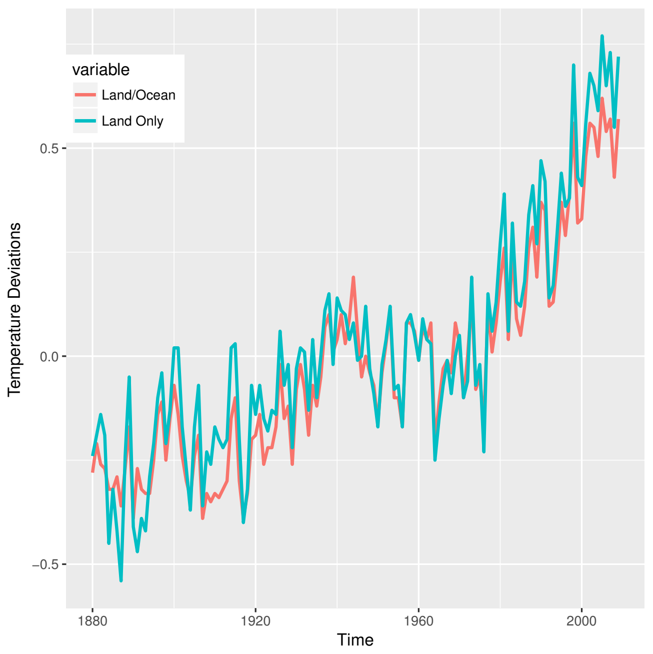


Taxa de desemprego versus tempo de desemprego

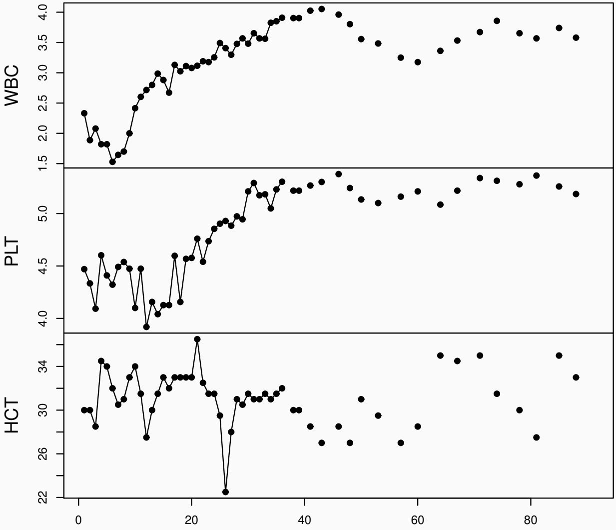


Desvios médios globais de temperatura, terra e oceânos (em graus Celcius)




variable
Land/Ocean
Land Only
Temperature Deviations
0.5
0.0
−0.5
1880
1920
1960
2000
Time

Contagens de células brancas, plaquetas e hematócritos coletados em 91 dias com dados perdidos



Diferenciação

- Um tipo de filtro para remover componente de tendência polinomial, consiste em *diferenciar* a série até que ela se torne *estacionária* (este conceito será formalizado adiante.
- Para dados não sazonais, a primeira diferença é em geral suficiente para induzir estacionariedade aproximada.
- A nova série $y_2, \dots, y_n$ é formada a partir da série original $x_1, \dots, x_n$ como

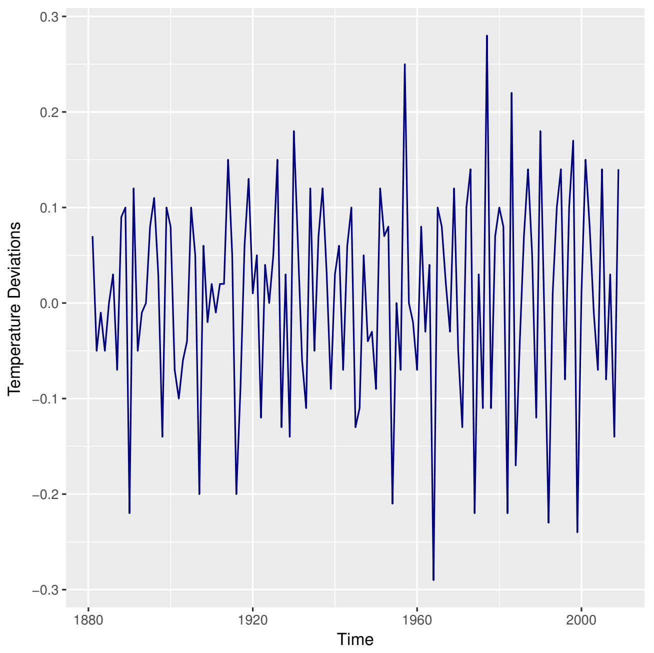
$$y_t = x_t - x_{t-1} = \nabla x_t.$$

- Diferenciação de primeira ordem é a mais utilizada sendo que ocasionalmente uma diferenciação de segunda ordem pode ser requerida,

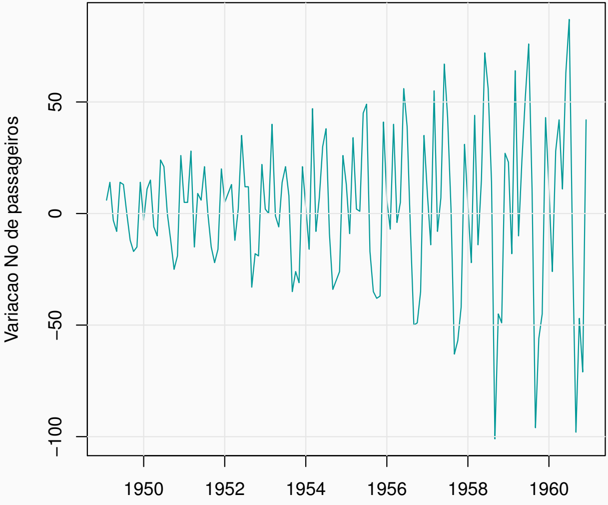
$$y_t = \nabla^2 x_t = \nabla(x_t - x_{t-1}) = x_t - 2x_{t-1} + x_{t-2}.$$

- Diferenciação pode ser útil como ferramenta exploratória. Observações discrepantes podem ter um efeito dramático na série diferenciada

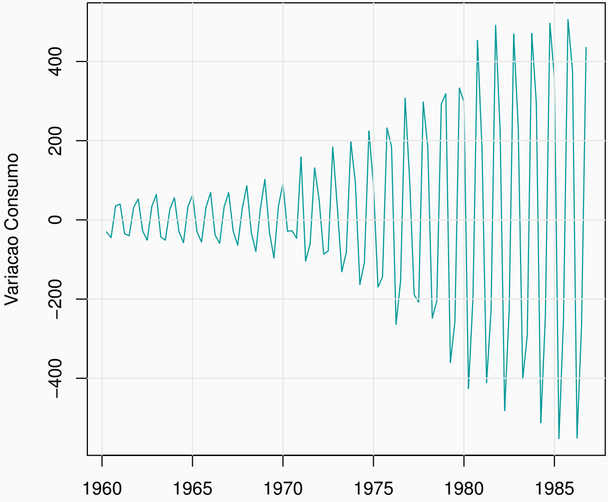
1a diferença dos desvios médios globais de temperatura



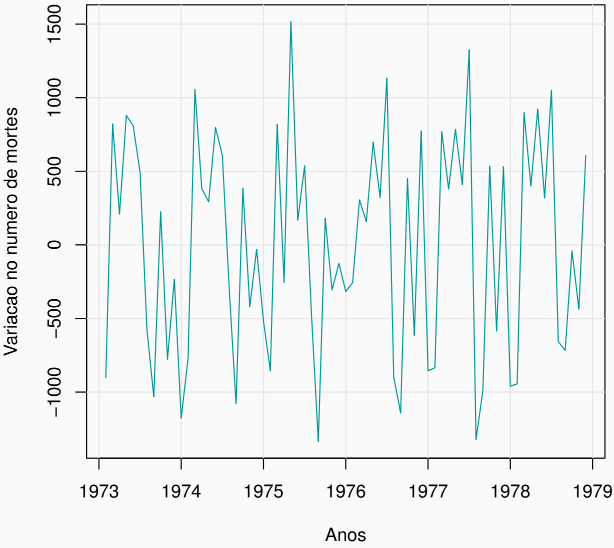
Primeira diferença dos totais mensais de passageiros em linhas aéreas internacionais nos EUA entre 1949 e 1960



Primeira diferença no consumo de gás no Reino Unido entre o primeiro trimestre de 1960 e o quarto trimestre de 1986



Primeira diferença do número mensal de mortes por acidente nos EUA.



Totais mensais de mortes por acidente nos EUA

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1973	9007	8106	8928	9137	10017	10826	11317	10744	9713	9938	9161	8927
1974	7750	6981	8038	8422	8714	9512	10120	9823	8743	9129	8710	8680
1975	8162	7306	8124	7870	9387	9556	10093	9620	8285	8466	8160	8034
1976	7717	7461	7767	7925	8623	8945	10078	9179	8037	8488	7874	8647
1977	7792	6957	7726	8106	8890	9299	10625	9302	8314	8850	8265	8796
1978	7836	6892	7791	8192	9115	9434	10484	9827	9110	9070	8633	9240

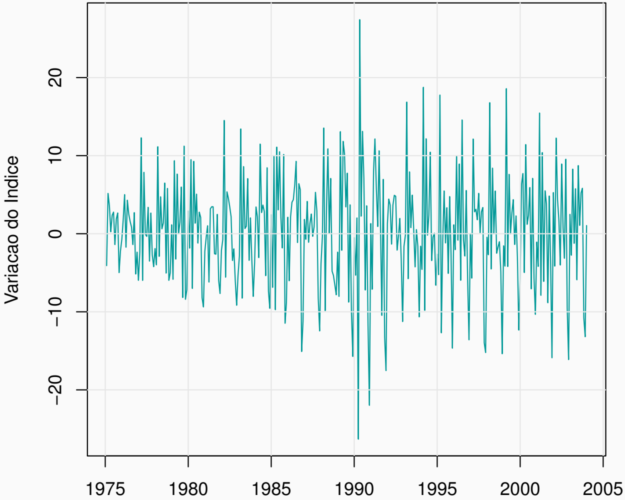
Primeira diferença dos totais mensais de mortes por acidente nos EUA

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1973		-901	822	209	880	809	491	-573	-1031	225	-777	-234
1974	-1177	-769	1057	384	292	798	608	-297	-1080	386	-419	-30
1975	-518	-856	818	-254	1517	169	537	-473	-1335	181	-306	-126
1976	-317	-256	306	158	698	322	1133	-899	-1142	451	-614	773
1977	-855	-835	769	380	784	409	1326	-1323	-988	536	-585	531
1978	-960	-944	899	401	923	319	1050	-657	-717	-40	-437	607

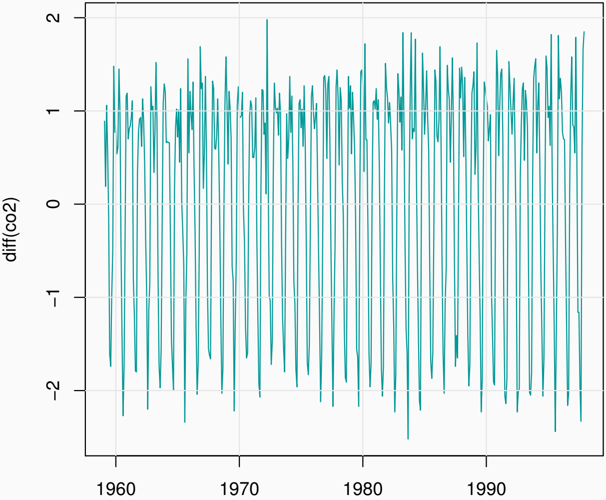
Primeira defasagem dos totais mensais de mortes por acidente nos EUA

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1972												9007
1973	8106	8928	9137	10017	10826	11317	10744	9713	9938	9161	8927	7750
1974	6981	8038	8422	8714	9512	10120	9823	8743	9129	8710	8680	8162
1975	7306	8124	7870	9387	9556	10093	9620	8285	8466	8160	8034	7717
1976	7461	7767	7925	8623	8945	10078	9179	8037	8488	7874	8647	7792
1977	6957	7726	8106	8890	9299	10625	9302	8314	8850	8265	8796	7836
1978	6892	7791	8192	9115	9434	10484	9827	9110	9070	8633	9240	

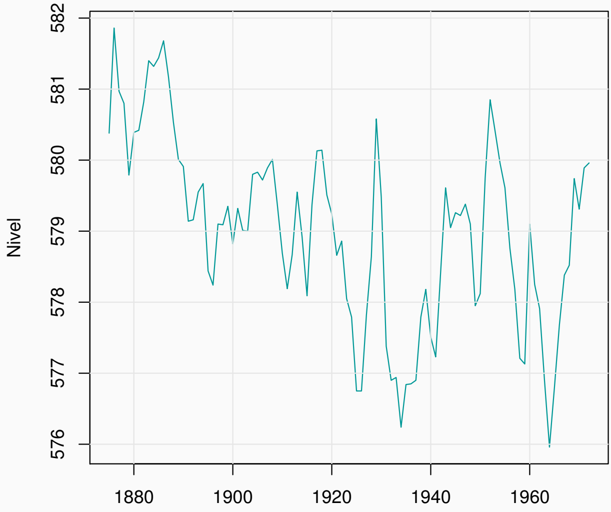
Primeira diferença do Indice da Produção Industrial do Brasil (indústria geral, média 1991 = 100). Jan/1975 a Jan/2004



Primeira diferença nas concentrações atmosfericas de CO2 (em ppm) de janeiro de 1953 a dezembro de 1997



Medições anuais (em pés) do nivel do Lago Huron, 1875-1972



Defasagens e primeira diferença do nivel do Lago Huron

	x(t)	x(t-1)	x(t-2)	x(t-3)	x(t-4)	x(t)-x(t-1)
1					578.25	
2				578.25	577.91	
3			578.25	577.91	576.89	
4		578.25	577.91	576.89	575.96	
5	578.25	577.91	576.89	575.96	576.80	
6	577.91	576.89	575.96	576.80	577.68	-0.34
7	576.89	575.96	576.80	577.68	578.38	-1.02
8	575.96	576.80	577.68	578.38	578.52	-0.93
9	576.80	577.68	578.38	578.52	579.74	0.84
10	577.68	578.38	578.52	579.74	579.31	0.88
11	578.38	578.52	579.74	579.31	579.89	0.70
12	578.52	579.74	579.31	579.89	579.96	0.14
13	579.74	579.31	579.89	579.96		1.22
14	579.31	579.89	579.96			-0.43
15	579.89	579.96				0.58
16	579.96					0.07

Retornos

Em Finanças, o risco é frequentemente medido em termos de variações de preços de ativos. Seja P_t o preço de um ativo no tempo t.

- A variação de preços entre $t-1$ e t (sem dividendos pagos) é, $P_t - P_{t-1}$.
- Variação relativa de preços,

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1.$$

- Retorno composto continuamente, ou log-retorno,

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(1 + R_t).$$

Alguns fatos estilizados sobre retornos

- Retornos tendem a apresentar caudas pesadas com pico mais alto em torno da média.

- Retornos tendem a apresentar variabilidades diferentes ao longo do tempo.

- Retornos tendem a apresentar variabilidades agrupadas. Retornos grandes tendem a ser seguidos por retornos grandes e retornos pequenos tendem a ser seguidos por retornos pequenos.

- A variabilidade tende a crescer mais seguindo uma queda de preço (retorno negativo) do que após um aumento de preço de mesma magnitude (efeito alavancagem).