

Forecasting

*** O que é? ***

Forecasting é a técnica de você prever, ou tentar prever, um determinado comportamento de algo ao qual você analisa. Você agir por puro impulso pode te trazer prejuízos de várias formas.

A habilidade de prever algo pode te ajudar a longo prazo (um investimento de alguma empresa, que pode te dar algum retorno financeiro somente após alguns anos), prever qual será a demanda de mercadorias no mês seguinte, etc. Conforme há um comportamento temporal, você **TALVEZ** possa prever o comportamento e te ajudar o planejamento.

Alguns modelos são muito mais fáceis do que outros. Já a questão do “talvez” se deve a alguns fatores que têm que ser levados em conta da hora de querer criar algum modelo de previsão. Estas condições são:

- O quão bem você entende dos fatores que contribuem para o comportamento;
- Quanta informação está disponível;
- Se o modelo preditivo pode influenciar o comportamento daquilo que você quer prever;

Os exemplos citados pelo livro são:

- **Projeção de demanda de eletricidade:** Temos boa ideia de quais fatores são os que mais influenciam em seu comportamento: temperatura e uma pequena influência de datas especiais (feriados, etc). Podemos usar um exemplo brasileiro que é a demanda elétrica das cidades litorâneas, que têm uma população pequena e, com datas festivas recebem uma grande população temporária. Juntando isso com os dados de demanda e de condições de temperatura, podemos ter um modelo preciso sobre o comportamento;
- **Projeção de preços de moedas:** A quantidade de informação da flutuação da moeda nós temos (é muito fácil procurar no Google o histórico da moeda X), em compensação, não temos plena ideia de quais são os fatores que influenciam nos preços, pois os fatores são diversos e o fato de termos um modelo de previsão pode mudar o comportamento do mercado. Exemplo: digamos que eu crio um modelo de preço do dólar e informo isso ao mercado (consideremos que eu tenho uma grande influência no mercado de dólar e afins). Quando os investidores souberem que o dólar vai aumentar, eles irão se comportar como se isso fosse realmente ocorrer. O contrário seria igual (falar que o preço do dólar iria diminuir). Por isso que há punições severas para quem divulga esse tipo de informação.

Boas previsões capturam o comportamento dos dados que existem no histórico, mas não os tenta replicar, já que são algo que já ocorreram.

1.2 – Forecasting, planning and goals

Predizer um comportamento de algo é uma tarefa comum em um ambiente de negócios, mas muita gente confunde **previsão, planejamento e objetivos**:

- **Predição (Forecasting)** – é sobre prever o futuro da forma mais precisa possível, usando todas as informações possíveis, incluindo dados históricos e informações futuras sobre eventos que podem alterar a predição;
- **Objetivos (Goals)** – É o que é desejável que ocorra. Os objetivos devem estar ligados com o planejamento e a predição, mas nem sempre isso ocorre. Muitas vezes os objetivos são concebidos sem qualquer plano ou predição do comportamento daquilo que quer que ocorra (objetivo);
- **Planejamento (planning)** – é a resposta da predição e dos objetivos. Planejar envolve tomar determinadas ações para que se chegue no objetivo desejado.

Predição deve ser o carro chefe das atividades de tomadas de decisão. Organizações modernas costumam classificar as predições em **curto, médio e longo prazo**. Alguns exemplos são:

- **Curto-prazo** – para gerenciamento de pessoas, produção e transporte. Também a predição de demanda é importante. Um exemplo seria o quanto que o boteco do Jailson deveria alocar o Chico-Butico da limpeza para o balcão ou o quanto de costela de boteco deveria deixar no forno;
- **Médio-prazo** – para planejamento de recursos futuros, tais como estoque, contratar mais pessoas, comprar equipamentos, etc. Voltando para o boteco do Jailson, ele poderia verificar o quanto de pinga seria necessário para a próxima temporada da 4ª divisão do futebol local;
- **Longo-prazo** – usado para planejamento de mercado, novas oportunidades, etc. Jailson já poderia entender que ele deve reformar o boteco em, pelo menos, 5 anos, já que está abrindo um condomínio gigantesco no mesmo quarterão e que ele teria muito mais lucro abrindo um mercado do que continuar com o copo-sujo.

1.3 – Determining what to forecast

Quando começamos um projeto de projeção, devemos verificar o que deverá ser predito. Por exemplo, se a predição for de alguns itens a serem manufaturados (digamos que sejam móveis de madeira), é necessário perguntar:

- A predição será para toda a linha de produção ou só para algum produto específico?
- Na questão das vendas, será para todos os lugares ou em algum lugar específico?
- O período será diário? Semanal? Anual?

É necessário também verificar qual será o horizonte da predição. O modelo vai precisar ter um mês de antecedência? Seis meses? Um ano? O problema é que para espaço temporal, deveremos usar um modelo diferente.

Temos que verificar também o quão frequente será a predição. Se for uma predição que necessita ser feita com muita frequência, é melhor que ela seja feita com de uma forma automatizada.

Após verificar se esse modelo terá algum uso por alguém, temos que verificar onde armazenaremos os dados e onde é que obteremos os dados para o modelo.

1.4 - Forecasting data and methods

O tipo de predição depende muito da quantidade de dados disponíveis.

Se não temos informações ou se os dados que possuímos não são tão relevantes, podemos usar modelos qualitativos. Modelos qualitativos não são puro “achismo”, mas são modelos que possuem uma estrutura bem definida.

Já modelos quantitativos podem ser criados se duas condições forem satisfeitas

1. Informação numérica sobre eventos passados estejam disponíveis;
2. Podermos assumir que certos aspectos do comportamento continuem em tempos futuros;

1.6 – The basic steps in forecasting task

Não podemos criar qualquer modelo “do nada”. Para que evitemos que isso ocorra, temos que seguir alguns passos:

Definição do problema: Sendo a parte mais difícil do processo e a mais importante, precisamos definir quem precisa do modelo, quem irá usar o modelo de projeção e como o modelo cabe no processo da organização ao qual será usado. Aquele que montará o modelo precisa estar ciente das necessidades das pessoas que irão usar o modelo e de quais problemas o modelo irá ajudar a resolver;

Obtendo informações: nesta parte temos que obter a parte estatística dos dados e a expertise das pessoas envolvidas no negócio. A expertise é tão importante quanto, pois ela pode apontar informações que podem estar fora do banco de dados;

Análise exploratória preliminar: Após obter os dados e algum conhecimento sobre o negócio, temos que realizar uma pequena exploração nos dados obtidos. Através de algum gráfico, temos que ver se há alguma sazonalidade, ciclo ou se não há qualquer comportamento diferente nos dados;

Escolhendo o modelo adequado: Após verificar os dados, temos que ver qual é a necessidade para podemos ajustar o modelo que mais se encaixe na resolução do problema;

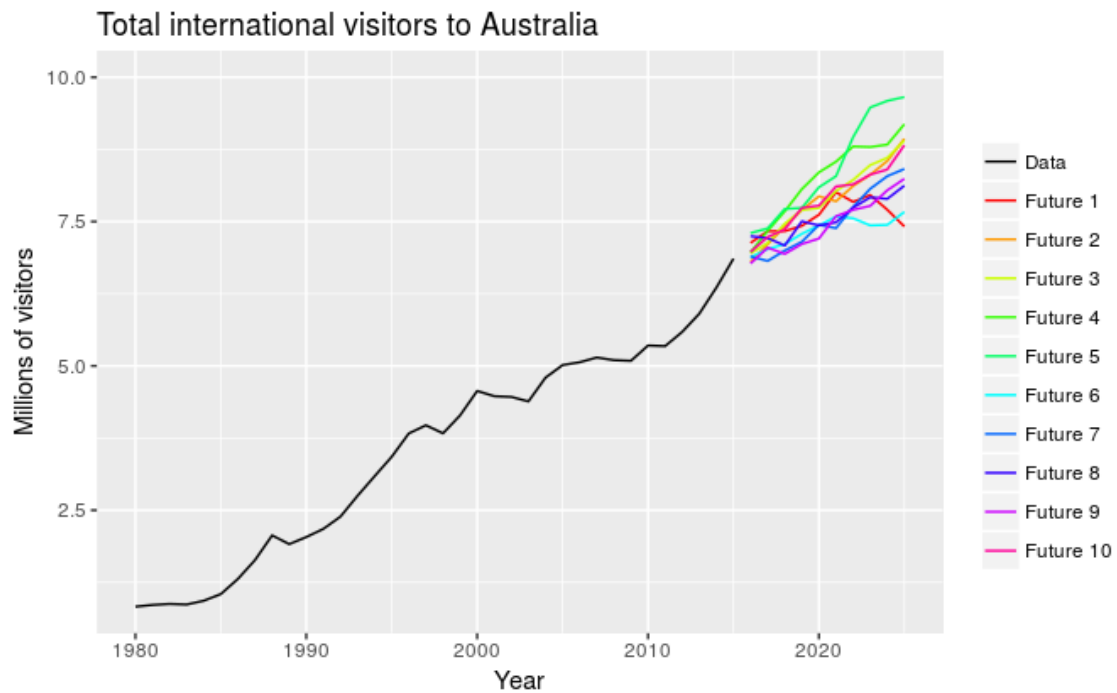
Usando e avaliando o modelo preditivo: após criar o modelo, temos que avaliar se o modelo criado possui alguma credibilidade. As vezes o modelo somente poderá ser testado após os eventos preditos ocorrerem.

1.7 – The statistical forecasting perspective

Quando estamos criando uma predição, temos que ter em mente que este comportamento, mesmo com os dados anteriores, ainda é desconhecido. Predições de curto prazo, como a venda do dia seguinte, por exemplo, são tecnicamente fáceis de se fazer, projetar as vendas do mesmo mês daqui um ano já não tem a mesma facilidade. Então até que conheçamos as vendas do dia seguinte ou do ano que vem, estas informações ainda são números aleatórios.

Então, partindo do pressuposto acima, quanto maior a distância da predição, maior é a dificuldade de prever um comportamento.

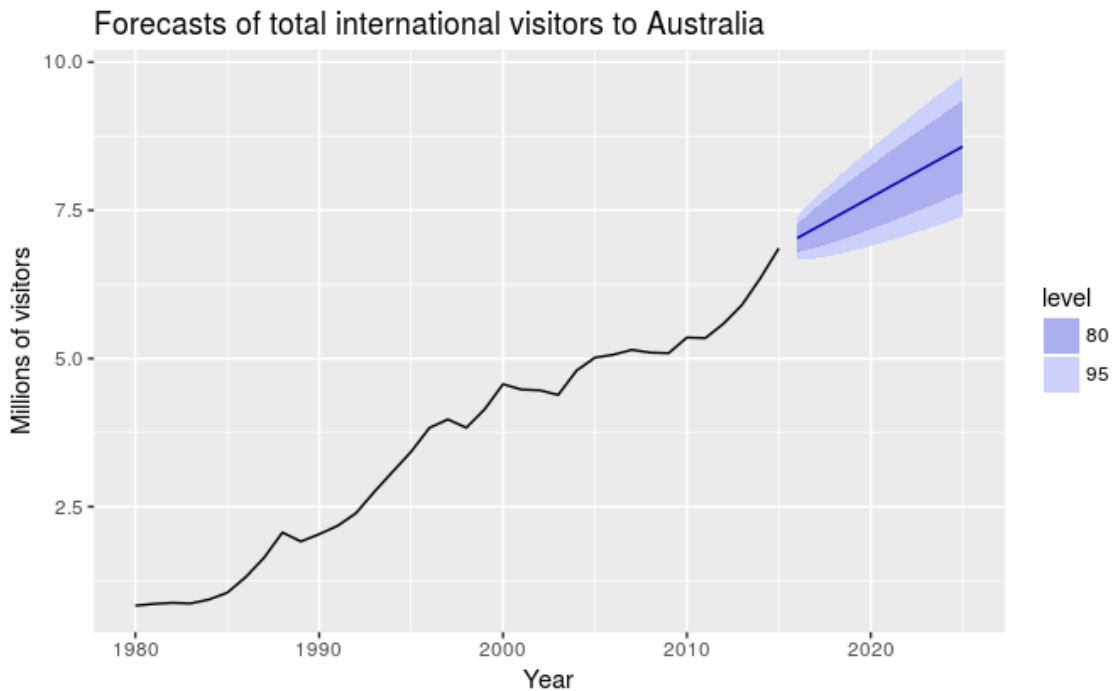
Quando criamos uma predição, não temos somente um único cenário, mas um conjunto de cenários.



A imagem acima demonstra que os dados futuros são diversos. Quanto mais futuros predizemos, mais cenários averão. Estes números não são gerados do nada, mas usando de alguns recursos matemáticos, tais como modelos **white noise**. Analisando de perto, isto nada mais é que uma distribuição probabilística.

Como essa distribuição possui uma faixa de valores, podemos assumir, assim como em uma distribuição normal, um intervalo de valores que são aqueles que estarão na nossa predição. Costumemente este valor está na casa dos 95%.

Ao invés de ficar plotando linha por linha, plotamos o intervalo de probabilidades de que tais intervalos ocorram.



Podemos ver três partes no gráfico:

- **Espaço de azul claro e escuro:** Neste espaço estão os 95% e os 80% do intervalo de confiança. Isso quer dizer que há a probabilidade de 95% de chance de o cenário futuro estar entre o espaço azul claro e 80% de chance de os valores futuros estarem no espaço azul escuro.
- **Linha azul:** nada mais é que a média dos valores, ao qual chamamos de **ponto de previsão**;

Usaremos a subscrito t para denotar tempo. Por exemplo, y_t denotará a observação no período t . Supondo que tenhamos observado o evento I e queiramos prever y_t . Quando escrevermos $I|y_t$, quer dizer que **“a variável aleatória y_t dado que conhecemos I ”** Os valores derivados disso terão uma distribuição probabilística, que é chamada de **“distribuição de previsão”**.

Quando falamos sobre previsão, na maioria das vezes queremos dizer a média da distribuição. Para representar isso, usamos o \hat{y} , que representa “a média dos valores possíveis de \hat{y} ”.

2 – Time Series graphics

Para que possamos criar os gráficos e os modelos, precisamos usar um tipo de dado do **R** que é próprio para isso: **ts**

2.1 – ts objects

Quando temos um determinado comportamento temporal, podemos instanciar o objeto **ts** para que armazene estes dados. O próprio objeto ts já possui algumas características, tais quais o intervalo de ocorrência entre os fatos. No código: **`ts(c(10, 20, 30, 40, 50), start = 2012)`**, teremos um objeto que já numerará os dados desde 2012 com intervalos de 1 em 1 ano.

```
> ts(c(10,20,30,40,50), start=2012)
Time Series:
Start = 2012
End = 2016
Frequency = 1
[1] 10 20 30 40 50
```

Figura 1 - Resultado do código do exemplo

Existem alguns parâmetros ao instanciar o objeto **ts**:

- **start** – Em qual período a série histórica começa
- **frequency** – É a frequência que o objeto irá interpretar e dividir os dados

Na questão da frequência, o **ts** interpretará de acordo com o valor passado:

1 = Anual

4 = Quarter (1/4 de um ano = 3 meses)

12 = Mensal

52 = Semanal

Como bem apontado [aqui](#), o ano não é dividido exatamente em 52 semanas, mas em 52,18 em média.

E na questão de divisão de tempos, não existe só a parte anual. Podemos dividir por dias (365,25 por ano), horas, minutos, segundos e assim por diante. O importante é ter uma divisão entre observações.

2.2 – Time plots

Quando nós tratamos sobre análise, temos que poder verificar isso de uma forma visual. Para isso utilizaremos de gráficos. Pelo fato do objeto **ts** ser sobre intervalos temporais, já surgirá um gráfico de linha ao utilizar o **autoplot**.

```
autoplot(melsyd[, "Economy.Class"]) +  
  ggtitle("Economy class passengers: Melbourne-Sydney") +  
  xlab("Year") + ylab("Thousands")
```

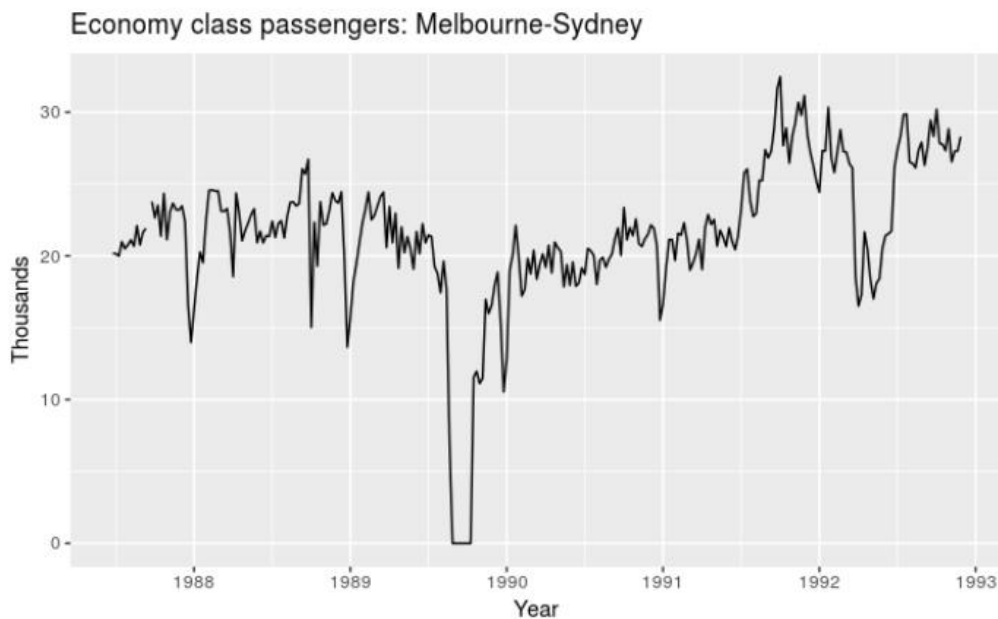


Figure 2.1: Weekly economy passenger load on Ansett Airlines.

Figura 2 - Resultado do código acima

Veja que o próprio **ggplot** já interpreta o objeto e já cria o gráfico com os intervalos temporais e suas respectivas observações.

2.3 – Time series patterns

Grandes coisas simplesmente olharmos para um gráfico e não conseguirmos interpretar o bendito. Existem três conceitos importantes para começarmos a brincar com análise de séries temporais:

1. **Tendência (trend)** – O trend é quando há um crescimento ou uma queda de longo prazo. A questão não é só a queda ou alta, mas o comportamento prolongado (sei que fui prolixo, mas enfim). Em um gráfico com algum comportamento temporal, veremos crescimentos e quedas com intervalos curtos, mas a tendência é quando vemos aquela linha subindo ou descendo por um determinado tempo;
2. **Sazonal (Seasonal)** – Um comportamento sazonal por algum ou alguns fatores da natureza ou algo que pode influenciar no comportamento do que está sendo observado. Pode ser alguma data especial (feriado, férias escolares), alguma adversidade da natureza (estações), etc;
3. **Cíclico (cyclic)** – O comportamento cíclico ocorre quando a data exibe uma alta e uma queda sem uma frequência fixa. Um bom exemplo de um ciclo são os ciclos econômicos (boom econômico, estabilidade, depressão, boom econômico...). A flutuação deste tipo de comportamento geralmente dura uns dois anos.

Como [apontado no livro](#), não podemos confundir um comportamento cíclico com um comportamento sazonal. Um comportamento **SAZONAL** é previsível, um **CÍCLICO** não tem a mesma previsibilidade.

2.4 – Seasonal Plots

Podemos identificar um comportamento sazonal quando há padrões de comportamento em determinado ponto da série histórica.

```
ggseasonplot(a10, year.labels=TRUE, year.labels.left=TRUE) +  
  ylab("$ million") + ggtitle("Seasonal plot: antidiabetic drug sales")
```

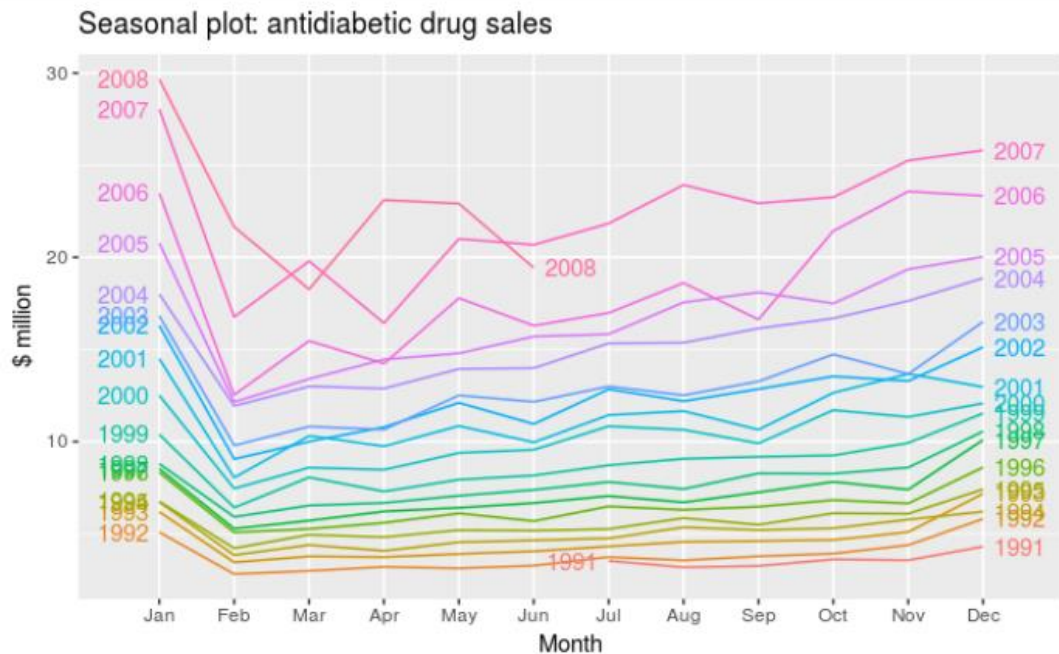


Figure 2.4: Seasonal plot of monthly antidiabetic drug sales in Australia.

Figura 3 - Gráfico de sazonalidade

Podemos ver que há uma queda de vendas em fevereiro, que interrompe a tendência ascendente das vendas. Como [explicado no livro](#), o comportamento se dá pelas vendas não contabilizadas de dezembro que só são contabilizadas em janeiro. Mesmo isso ocorrendo, podemos ver que o comportamento é quase fixo: temos uma tendência ascendente nas vendas e o **boom** em janeiro. Isso caracteriza um comportamento sazonal.

Uma variação deste gráfico é o gráfico polar


```
ggseasonplot(a10, polar=TRUE) +  
  ylab("$ million") + ggtitle("Polar seasonal plot: antidiabetic drug sales")
```

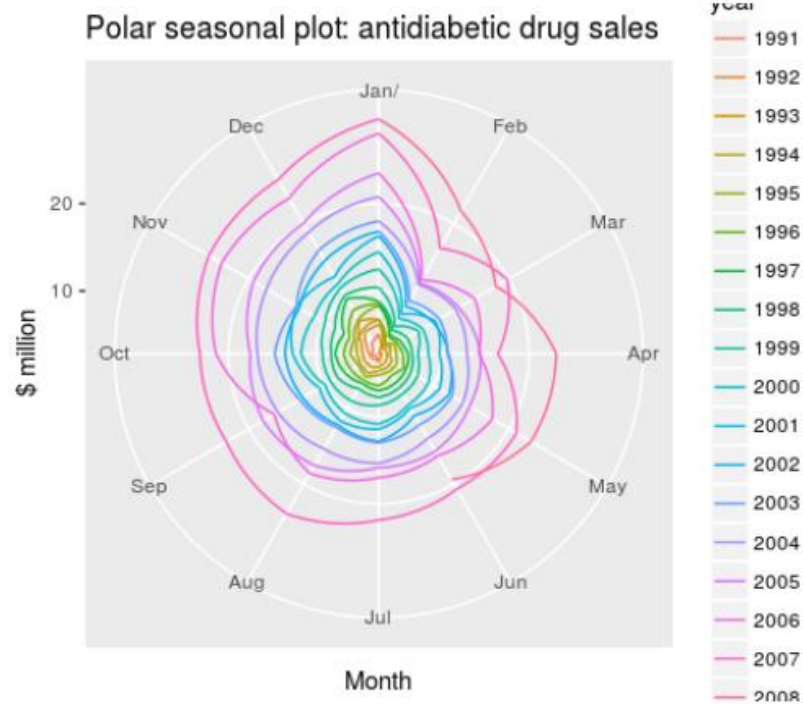


Figure 2.5: Polar seasonal plot of monthly antidiabetic drug sales in Australia.

Figura 4 - Gráfico polar. Note o formato parecido em todos os anos.

2.5 – Seasonal subseries plot

Também podemos usar um gráfico de subséries, onde veremos o comportamento de cada série temporal de uma forma diferente.

```
ggsubseriesplot(a10) + ylab("$ million") +  
  ggtitle("Seasonal subseries plot: antidiabetic drug sales")
```

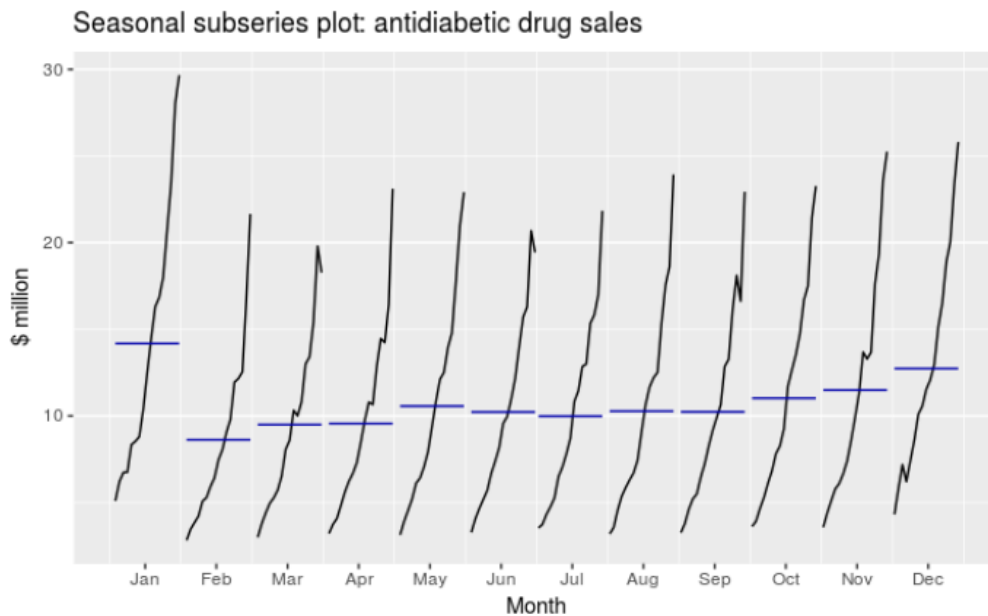


Figure 2.6: Seasonal subseries plot of monthly antidiabetic drug sales in Australia.

Figura 5 - Subséries das vendas mensas de medicamentos contra a diabetes. Note o pico do valor de janeiro

Links para estudo

Forecasting: Principles and Pactice - <http://otexts.org/fpp2/>

Séries históricas IBGE - <https://seriesestatisticas.ibge.gov.br/>

Conectando R ao banco de dados - <https://support.rstudio.com/hc/en-us/articles/214510788-Setting-up-R-to-connect-to-SQL-Server-https://github.com/r-dbi/odbc>