

# Guia de Bolso: ggplot2

*Gustavo B Paterno*

*2018-10-09*



# Contents

<b>Sobre este livro</b>	<b>5</b>
<b>Sobre o autor</b>	<b>7</b>
<b>1 Introdução</b>	<b>9</b>
1.1 Pré-requisitos . . . . .	9
1.2 Como utilizar este livro? . . . . .	10
1.3 Ajude a melhorar o livro . . . . .	10
<b>2 Histogramas</b>	<b>11</b>
2.1 Histograma (simples) . . . . .	11
2.2 Número de classes . . . . .	12
2.3 Múltiplas categorias . . . . .	13
2.4 Histograma (bonito) . . . . .	13
<b>3 Density Plot</b>	<b>15</b>
3.1 Density plot (simples) . . . . .	15
3.2 Density plot (colorido) . . . . .	16
3.3 Múltiplas categorias . . . . .	17
3.4 Density plot (bonito) . . . . .	17
<b>4 Boxplot</b>	<b>19</b>
4.1 Boxplot (simples) . . . . .	19
4.2 Destaque para os outliers . . . . .	20
4.3 Colorindo por categoria . . . . .	22
4.4 Histograma (bonito) . . . . .	23





# Sobre este livro

Este livro é um guia de bolso para consultar como fazer diferentes tipos de gráficos utilizando o programa estatístico R e o pacote ggplot2.



# Sobre o autor

Gustavo Brant de Carvalho paterno é doutor em Ecologia pela UFRN, com experiência em Ecologia Vegetal, Ecologia Evolutiva e análises estatísticas no R.





# Chapter 1

## Introdução

### 1.1 Pré-requisitos

#### Programas necessários

Para utilizar este livro é fundamental instalar os programas R e R studio. Para instalar o R e o R studio, visite os sites citados abaixo, siga as instruções de acordo com o seu sistema operacional (compatível com Windows, MAC e Linux).

- O programa R



- O programa R Studio



**Pacotes do R necessários**

- O pacote de R ggplot2
- O pacote de R dplyr

Para instalar os pacotes necessários, abra o R studio e digite os comandos abaixo no console:

```
install.packages("ggplot2")  
install.packages("dplyr")
```

**Versões dos programas e pacotes utilizadas no livro**

Para que os exemplos descritos ao longo deste livro, verifique as versões dos programas e pacotes que estão instalados no seu computador.

1. R version 3.4.4 (2018-03-15)
  2. Versão do R Studio: 1.1.4 (Linux - Ubuntu)
  3. Versão do ggplot2: 3.0.0
  4. Versão do dplyr: 0.7.6
- 

## 1.2 Como utilizar este livro?

## 1.3 Ajude a melhorar o livro

## Chapter 2

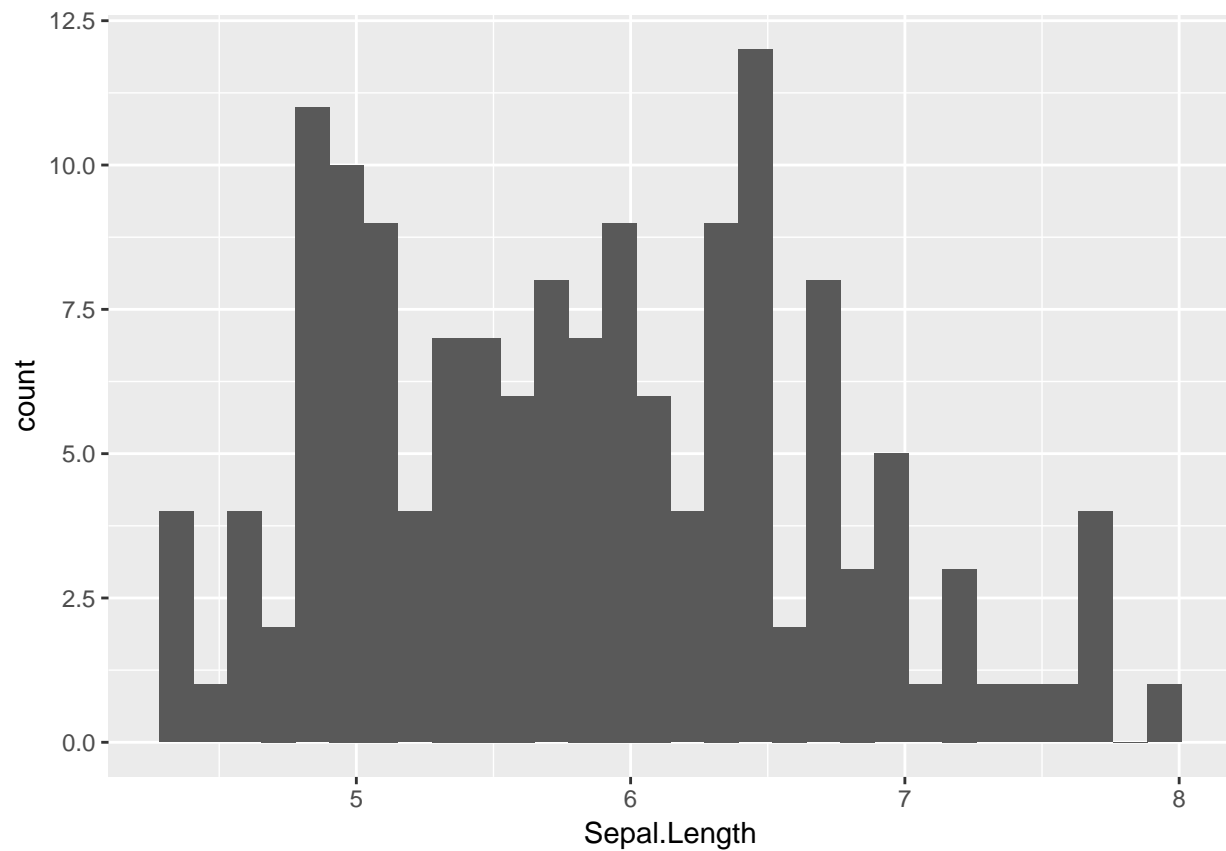
# Histogramas

Neste capítulo você irá aprender a criar histogramas no R utilizando o ggplot2.

### 2.1 Histograma (simples)

Utilize o banco de dados `iris`, para plotar a distribuição da variável `Sepal.Length` em um gráfico de histograma. A função `geom_histogram()` plota uma variável contínua no eixo x e a frequência de cada categoria no eixo y.

```
# Carregue um banco de dados  
data("iris")  
  
# Histograma de uma variável contínua  
ggplot(iris, aes(x = Sepal.Length)) +  
  geom_histogram()
```

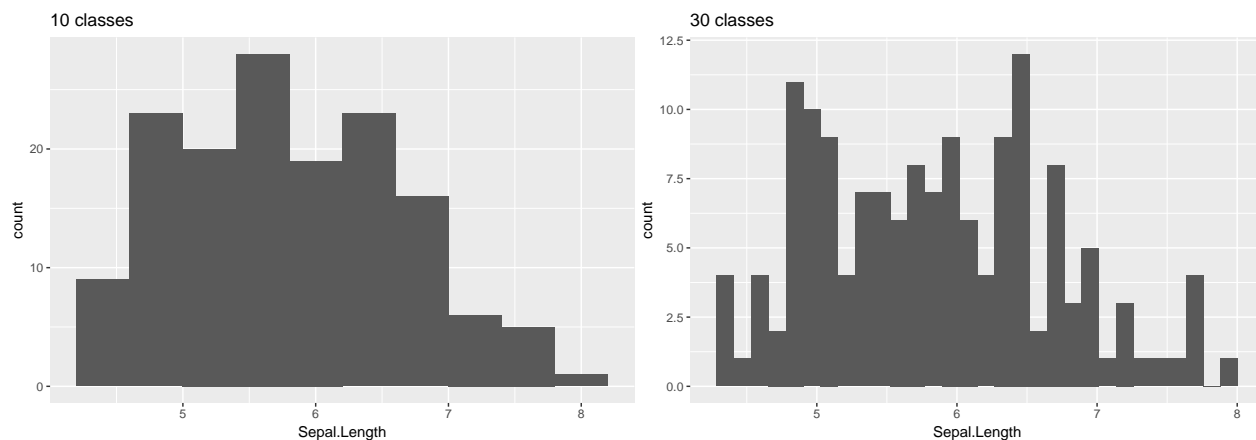


## 2.2 Número de classes

Utilize o **argumento** `bins` para definir em quantas classes a variável `x` deve ser dividida.

```
# Histograma com 10 classes:
ggplot(iris, aes(x = Sepal.Length)) +
  geom_histogram(bins = 10) +
  ggtitle("10 classes")

# Histograma com 30 classes:
ggplot(iris, aes(x = Sepal.Length)) +
  geom_histogram(bins = 30) +
  ggtitle("30 classes")
```

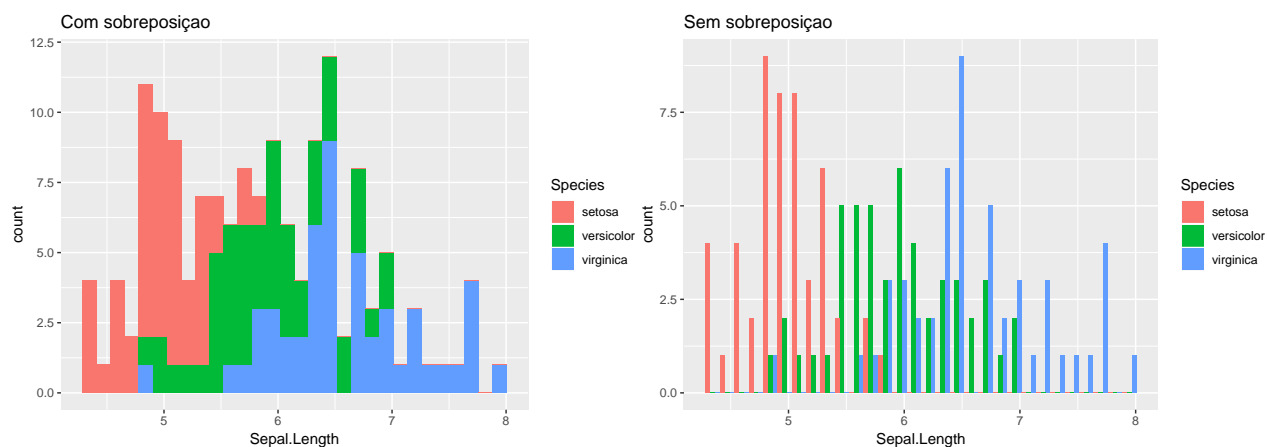


## 2.3 Múltiplas categorias

Se você quiser comparar a distribuição de uma variável contínua entre diferentes categorias, você pode utilizar o argumento `fill` para colorir o gráfico. No exemplo abaixo, utilizamos cores diferentes para ilustrar a distribuição da variável `x` entre espécies diferentes (`fill = Species`)

```
# Histograma com cores para diferentes categorias com sobreposição
ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
  geom_histogram() +
  ggtitle("Com sobreposição")

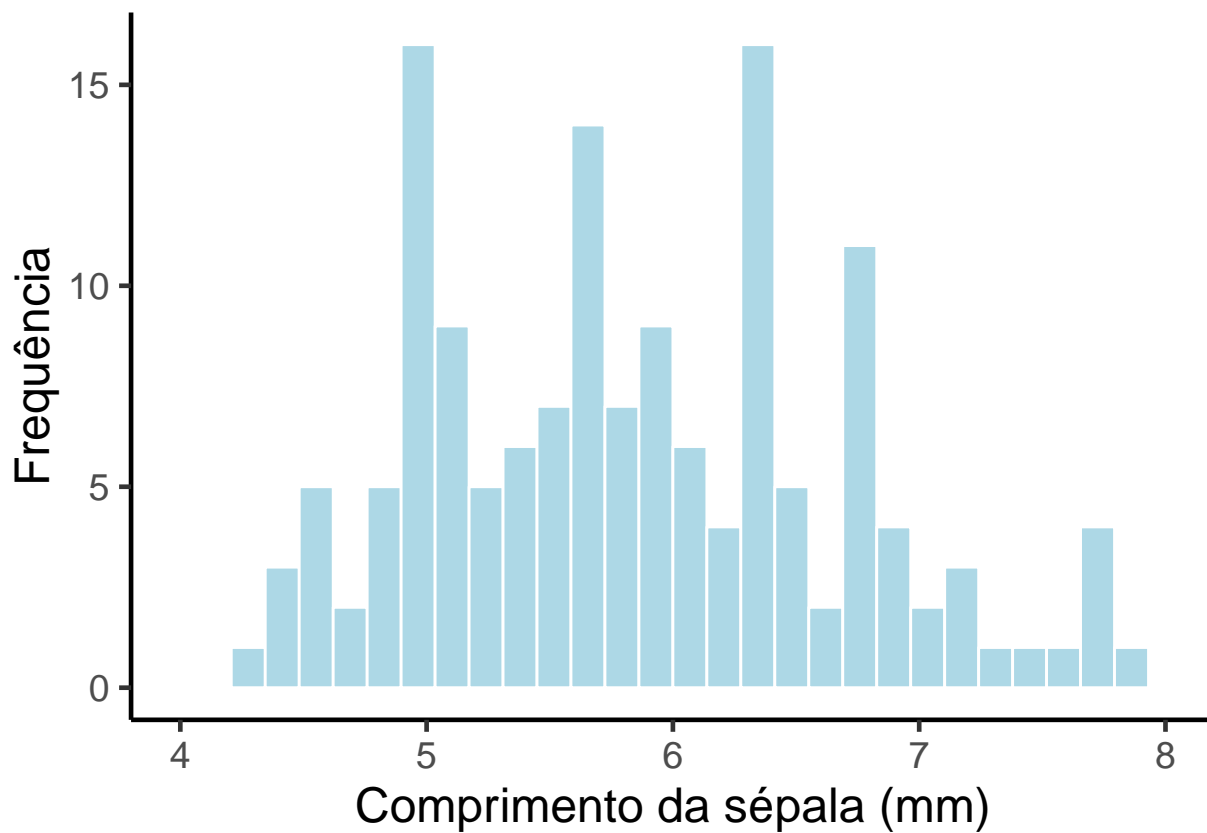
# Histograma com cores para diferentes categorias sem sobreposição
ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
  geom_histogram(position = "dodge") +
  ggtitle("Sem sobreposição")
```



## 2.4 Histograma (bonito)

Todo o final de capítulo temos a demonstração de um gráfico mais elegante. Veja o capítulo XXX para mais detalhes de como manipular cada detalhe do seu gráfico com o `ggplot2`, mas ao observar o script abaixo você pode ir aprendendo com cada exemplo.

```
ggplot(iris, aes(x = Sepal.Length)) +  
  geom_histogram(color = "white", fill = "lightblue") +  
  theme_classic(base_size = 18) +  
  scale_x_continuous(breaks = seq(from = 4,to = 8,by = 1), limits = c(4,8)) +  
  xlab("Comprimento da sépala (mm)") +  
  ylab("Frequência")
```



## Chapter 3

# Density Plot

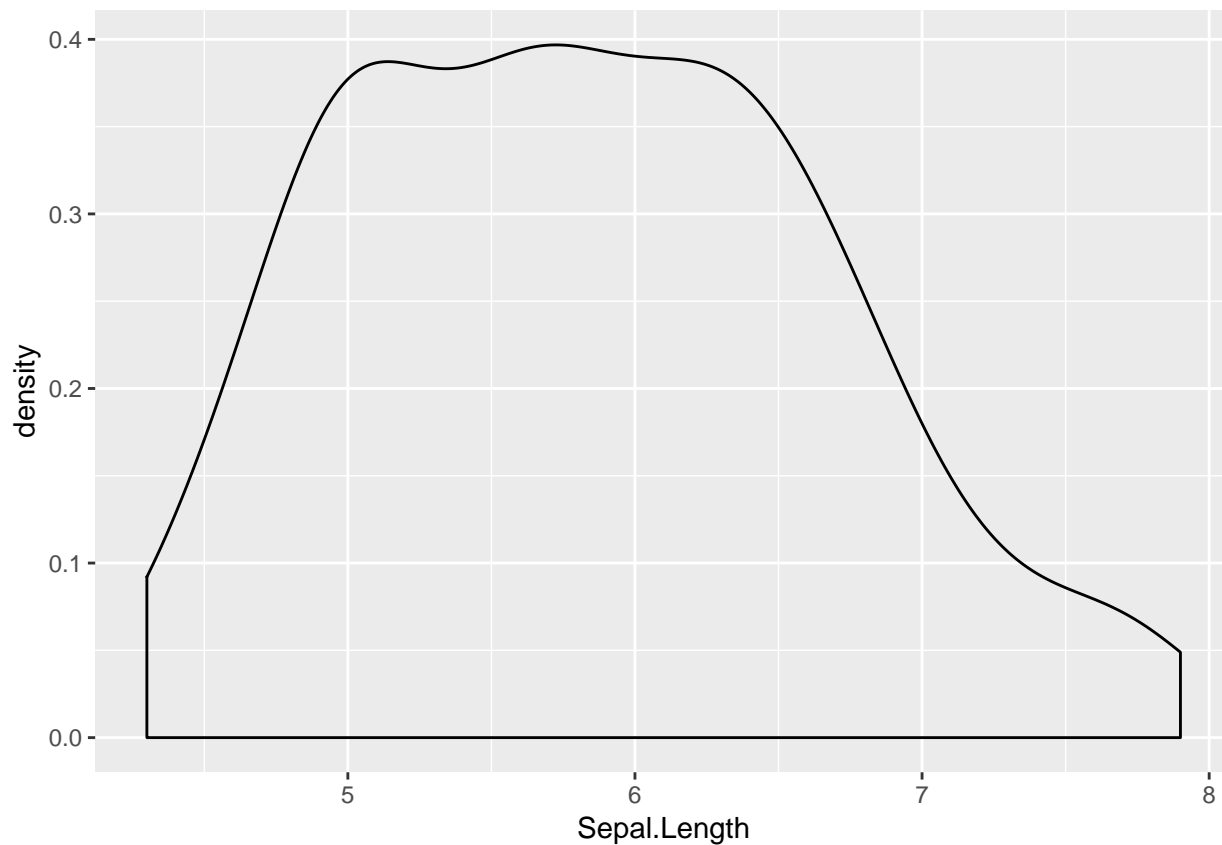
Neste capítulo você irá aprender a criar Density Plots no R utilizando o ggplot2.

O **Density plot** é utilizado para visualizar a distribuição de uma variável numérica em um intervalos contínuos. Esse gráfico é uma variação do gráfico de Histogramas(ver capítulo 2) que utiliza Kernel Smoother.

### 3.1 Density plot (simples)

Utilize o banco de dados `iris`, para plotar a distribuição da variável `Sepal.Length` em um Density Plot. Utilize a função `geom_density()` para plotar uma variável no eixo x.

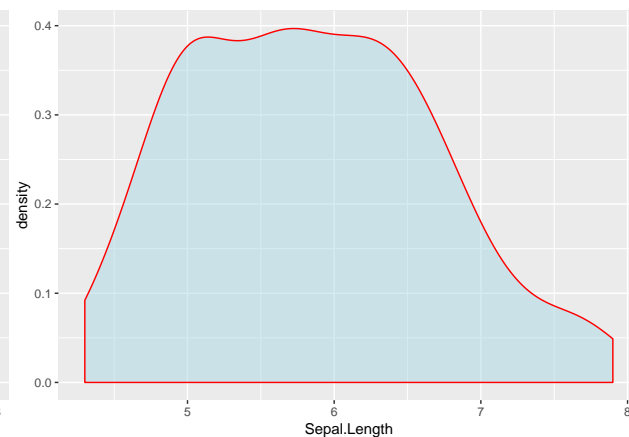
```
# Carregue um banco de dados  
data("iris")  
  
# Density plot de uma variável contínua  
ggplot(iris, aes(x = Sepal.Length)) +  
  geom_density()
```



### 3.2 Density plot (colorido)

Utilize o argumento `fill` para definir a cor de preenchimento do seu gráfico e o argumento `alpha` para definir a transparência do preenchimento. Utilize o argumento `color` para definir a cor da linha.

```
# Density plot de uma variável contínua
ggplot(iris, aes(x = Sepal.Length)) +
  geom_density(fill = "tomato")
# Density plot de uma variável contínua
ggplot(iris, aes(x = Sepal.Length)) +
  geom_density(fill = "lightblue", alpha = .5, color = "red")
```



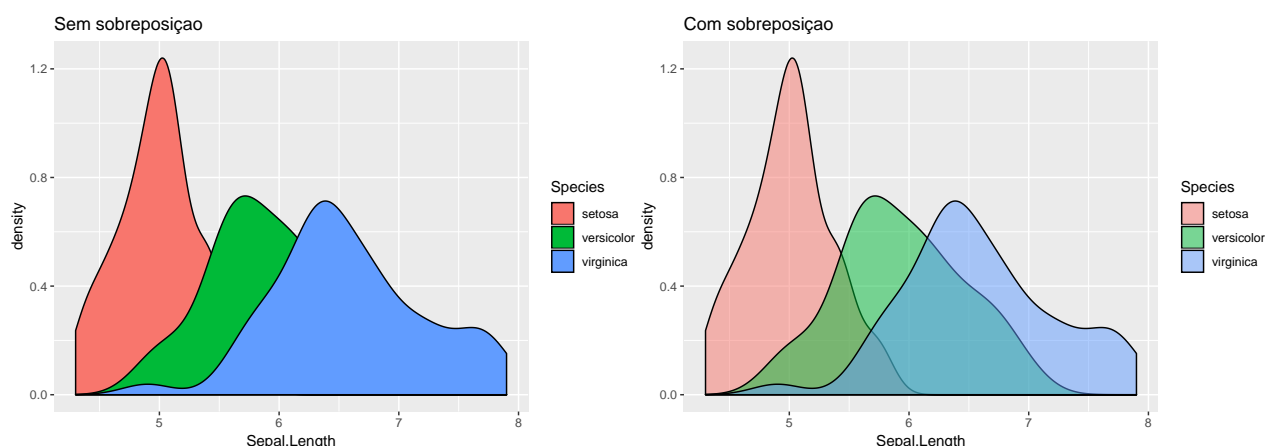


### 3.3 Múltiplas categorias

Se você quiser comparar a distribuição de uma variável contínua entre diferentes categorias, você pode utilizar o argumento `fill` para colorir o gráfico. No exemplo abaixo, utilizamos cores diferentes para ilustrar a distribuição da variável `x` entre espécies diferentes (`fill = Species`)

```
# Desnity plot com cores para diferentes categorias sem sobreposição
ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
  geom_density() +
  ggtitle("Sem sobreposição")

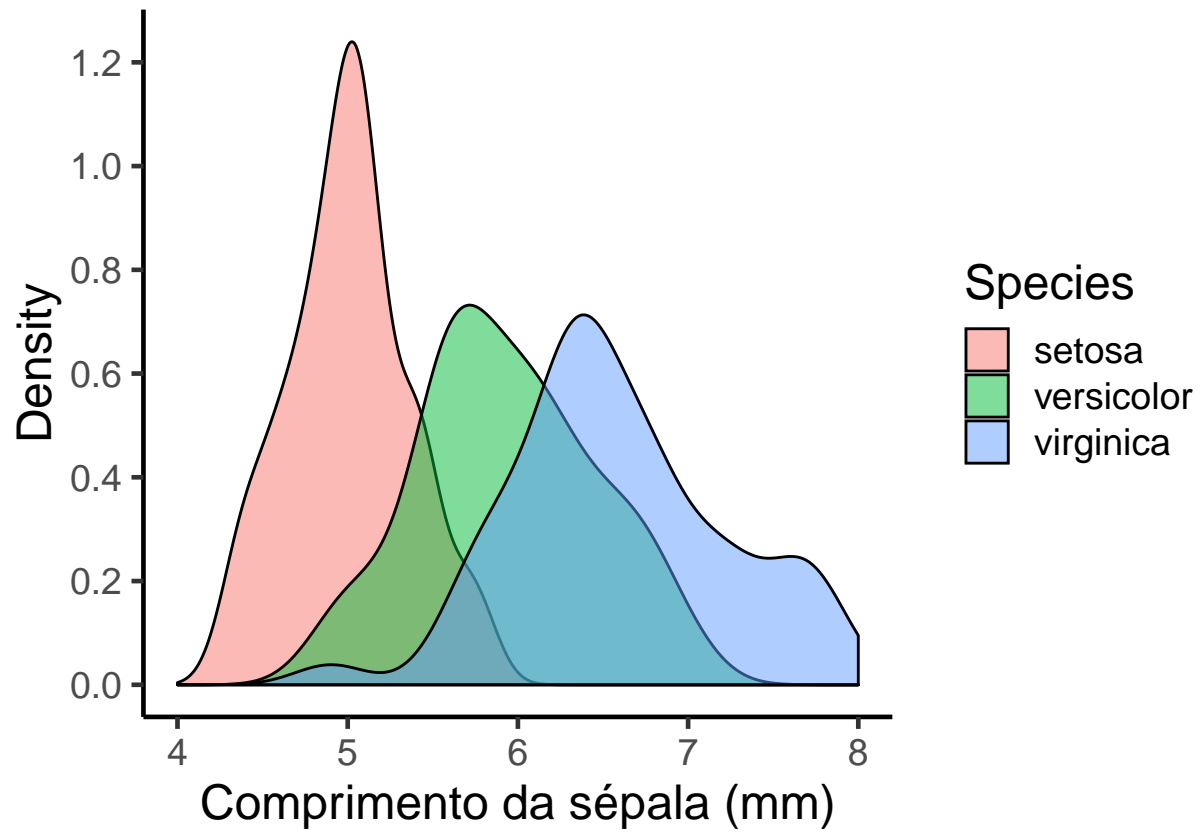
# Desnity plot com cores para diferentes categorias com sobreposição
ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
  geom_density(alpha = .5) +
  ggtitle("Com sobreposição")
```



### 3.4 Density plot (bonito)

Todo o final de capítulo temos a demonstração de um gráfico mais elegante. Veja o capítulo XXX para mais detalhes de como manipular cada detalhe do seu gráfico com o `ggplot2`, mas ao observar o script abaixo você pode ir aprendendo com cada exemplo.

```
ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
  geom_density(alpha = .5) +
  theme_classic(base_size = 18) +
  scale_x_continuous(breaks = seq(from = 4, to = 8, by = 1), limits = c(4, 8)) +
  scale_y_continuous(breaks = seq(from = 0, to = 1.4, by = .2)) +
  xlab("Comprimento da sépala (mm)") +
  ylab("Density")
```



## Chapter 4

# Boxplot

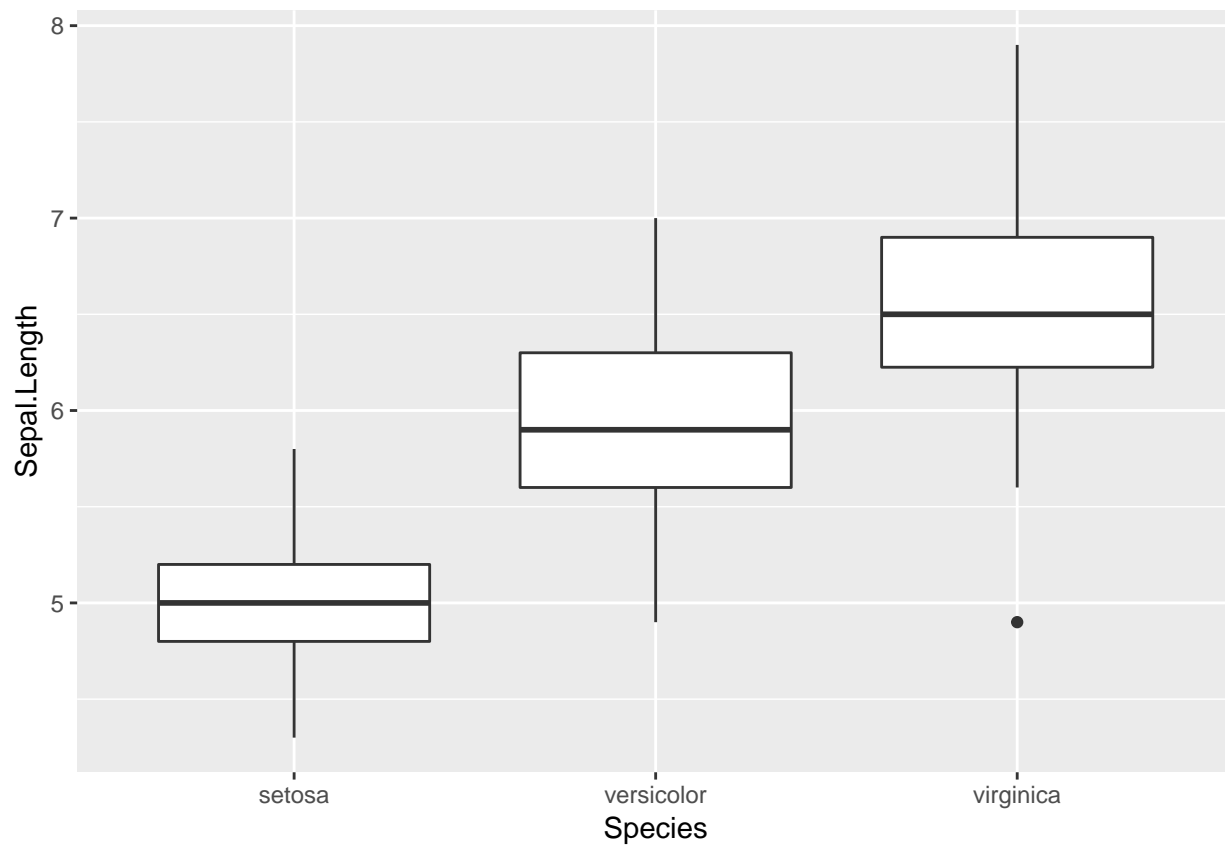
Neste capítulo você irá aprender a criar boxplots no R utilizando o ggplot2.

### 4.1 Boxplot (simples)

Utilize o banco de dados `iris`, para plotar um diagrama de caixa (boxplot). A função `geom_boxplot()` plota uma variável contínua (`Sepal.Length`) no eixo y em função de uma variável categórica no eixo x (`Species`). A definição de qual coluna do bando de dados é a x e qual é a y é feita dentro do comendo `aes()`.

```
# Carrgue um banco de dados
data("iris")

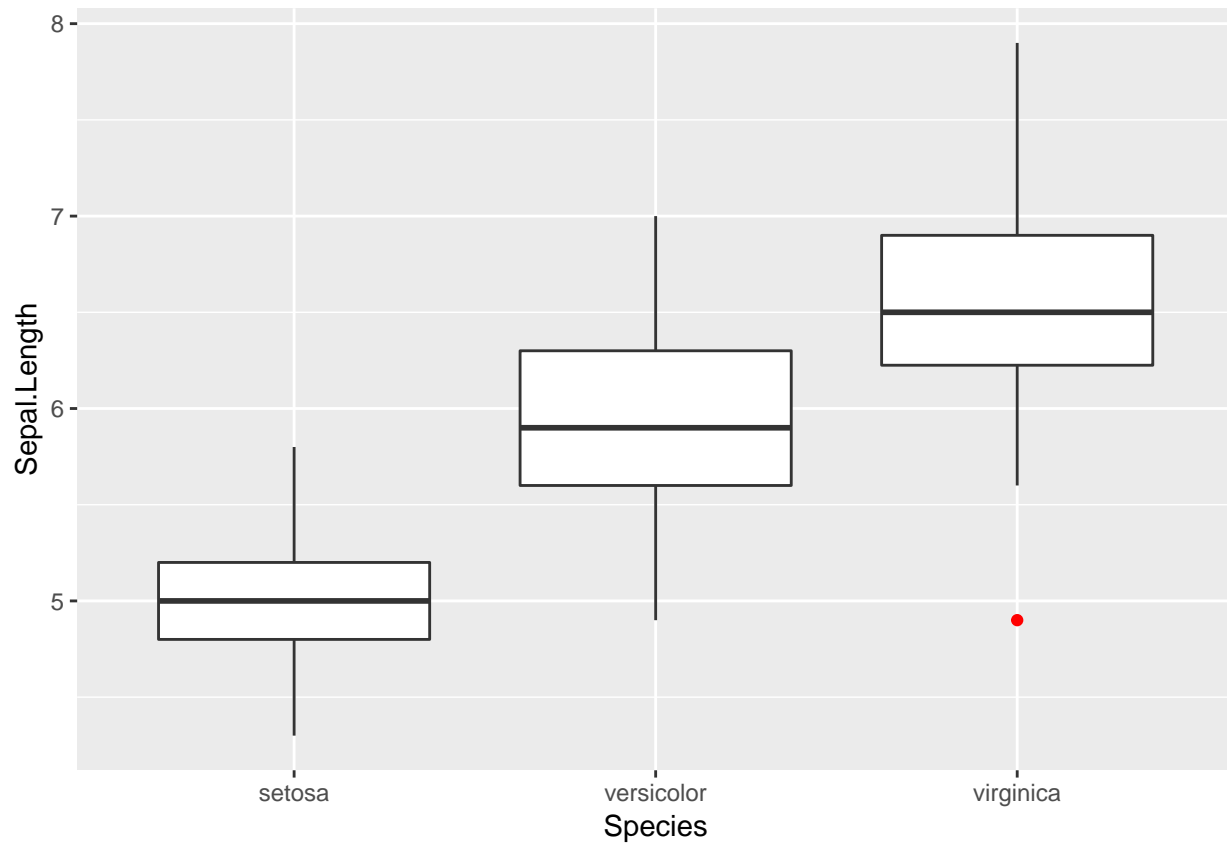
# Boxplot
ggplot(iris, aes(y = Sepal.Length, x = Species)) +
  geom_boxplot()
```



## 4.2 Destque para os outliers

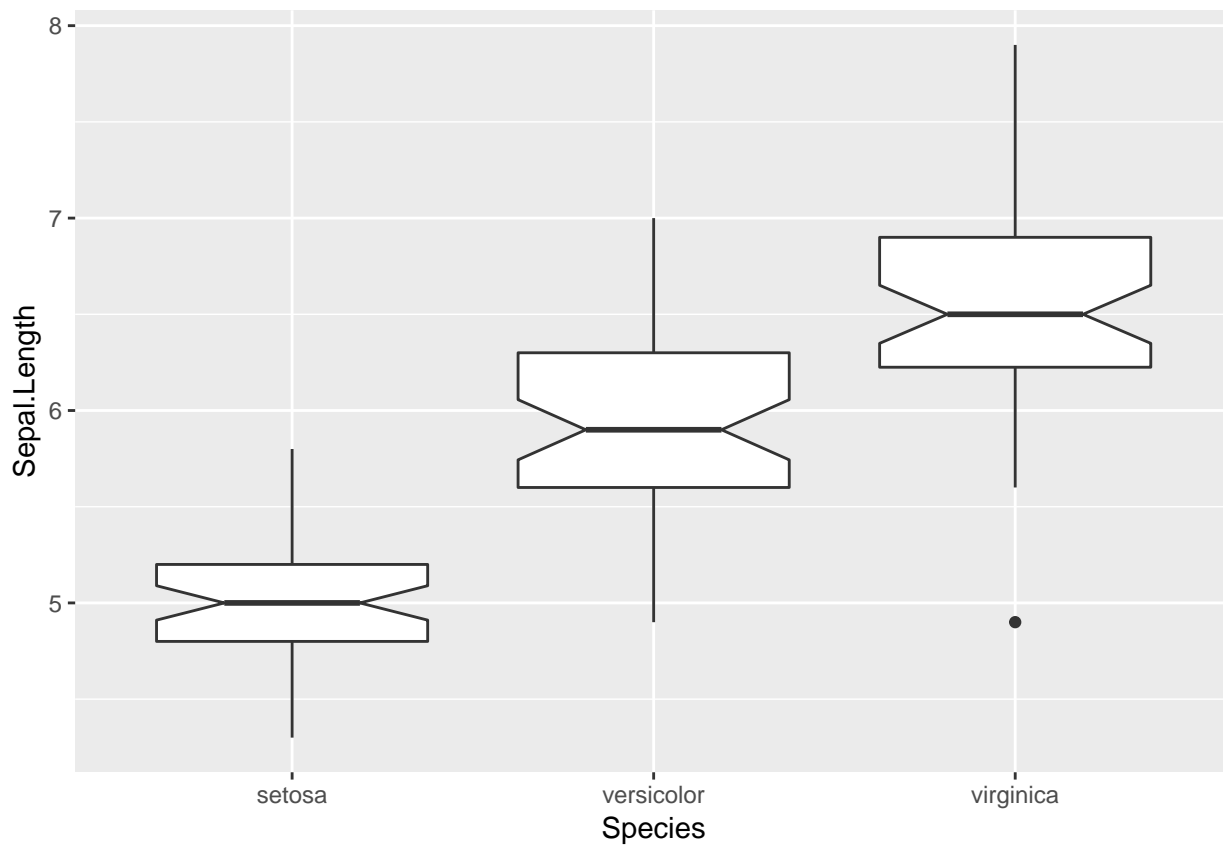
Utilize o **argumento** `outlier.color` para definir a cor dos outliers.

```
ggplot(iris, aes(y = Sepal.Length, x = Species)) +  
  geom_boxplot(outlier.color = "red")
```



Utilize o **argumento** `notch = T` para produzir diagramas de caixa entalhados (notched). Estes diagramas são úteis para inferir de forma aproximada se existe diferença significativa entre as medias dos grupos.

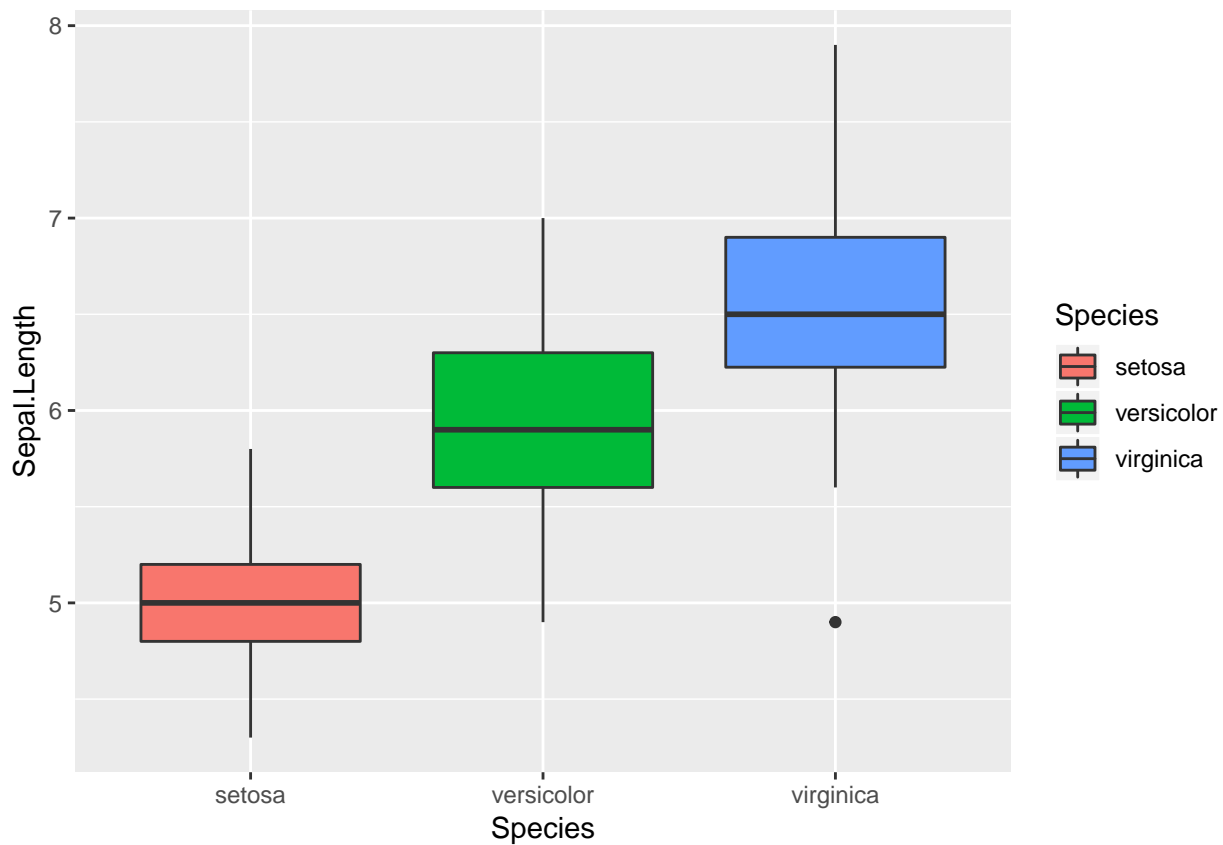
```
ggplot(iris, aes(y = Sepal.Length, x = Species)) +  
  geom_boxplot(notch = T)
```



### 4.3 Colorindo por categoria

No exemplo abaixo, utilizamos cores diferentes para ilustrar espécies diferentes através do argumento `fill = Species`.

```
# Histograma com cores para diferentes categorias com sobreposição
ggplot(iris, aes(y = Sepal.Length, x = Species, fill = Species)) +
  geom_boxplot()
```



## 4.4 Histograma (bonito)

Todo o final de capítulo temos a demonstração de um gráfico mais elegante. Veja o capítulo XXX para mais detalhes de como manipular cada detalhe do seu gráfico com o ggplot2, mas ao observar o script abaixo você pode ir aprendendo com cada exemplo.

```
ggplot(iris, aes(y = Sepal.Length, x = Species, fill = Species)) +  
  geom_boxplot(show.legend = F, alpha = .5) +  
  scale_y_continuous(limits = c(4,8), breaks = seq(4,8,1)) +  
  theme_classic(base_size = 18) +  
  xlab("Espécie") +  
  ylab("Comprimento da sépala (mm)")
```

