

# RT01: Uso do pacote `ggplot2`

Estudo dirigido e Relatórios Técnicos em R Markdown.

Lucas Henrique Nogueira

2024-06-09

## Contents

<b>1</b>	<b>Introdução</b>	<b>3</b>
1.1	O pacote <code>ggplot2</code> . . . . .	3
1.1.1	Principais Características do <code>ggplot2</code> . . . . .	3
1.2	Conjunto de dados <code>gapminder</code> . . . . .	3
1.2.1	Estrutura do Conjunto de Dados . . . . .	3
1.3	Conjunto de dados <code>mtcars</code> . . . . .	4
1.3.1	Estrutura do Conjunto de Dados . . . . .	4
1.4	Conjunto de dados <code>diamonds</code> . . . . .	4
1.4.1	Estrutura do Conjunto de Dados . . . . .	4
<b>2</b>	<b>Realização dos exercícios propostos</b>	<b>5</b>
2.1	Questão 1 - Adicionando texto a gráfico . . . . .	5
2.1.1	Gráfico . . . . .	5
2.1.2	Script do gráfico . . . . .	5
2.2	Questão 2 - Diagrama de dispersão . . . . .	6
2.2.1	Gráfico . . . . .	6
2.2.2	Script do gráfico . . . . .	6
2.3	Questão 3 - Formatos dos símbolos dos dados . . . . .	7
2.3.1	Gráfico . . . . .	7
2.3.2	Script do gráfico . . . . .	7
2.4	Questão 4 - Atribuindo função <code>ggplot2</code> a objeto . . . . .	8

2.4.1	Gráfico . . . . .	8
2.4.2	Script do gráfico . . . . .	8
2.5	Questão 5 - Criando gráfico de área . . . . .	9
2.5.1	Gráfico . . . . .	9
2.5.2	Script do gráfico . . . . .	9
<b>3</b>	<b>Considerações finais</b>	<b>10</b>
<b>4</b>	<b>Links</b>	<b>11</b>

# 1 Introdução

## 1.1 O pacote `ggplot2`

O pacote `ggplot2` é uma das bibliotecas mais populares para visualização de dados no R. Desenvolvido por Hadley Wickham, ele implementa a abordagem de gramática de gráficos (*grammar of graphics*), permitindo a criação de visualizações complexas de maneira concisa e flexível.

### 1.1.1 Principais Características do `ggplot2`

- **Gramática de Gráficos:** Baseia-se na gramática de gráficos, que permite a construção incremental de gráficos por meio da adição de camadas de componentes, como dados, geometrias, estatísticas, escalas, coordenadas e facetas.
- **Abordagem Declarativa:** Em vez de especificar passo a passo como um gráfico deve ser desenhado, descreve-se a relação entre dados e as representações visuais.
- **Suporte a Diversos Tipos de Gráficos:** Suporta uma ampla variedade de gráficos, incluindo gráficos de dispersão, linhas, barras, boxplots, histogramas, gráficos de densidade, entre outros.
- **Flexibilidade e Extensibilidade:** Permite criar gráficos simples ou altamente customizados com facilidade. Integra-se bem com outros pacotes do ecossistema tidyverse.

## 1.2 Conjunto de dados `gapminder`

O conjunto de dados `gapminder` é amplamente utilizado para ilustrar a evolução de diversos indicadores socioeconômicos e de saúde ao longo do tempo. Este conjunto de dados contém informações sobre a expectativa de vida, PIB per capita, e população de vários países do mundo em diferentes anos, facilitando análises de tendências e comparações entre nações.

### 1.2.1 Estrutura do Conjunto de Dados

O conjunto de dados `gapminder` possui as seguintes colunas principais:

- **country:** O nome do país.
- **continent:** O continente ao qual o país pertence.
- **year:** O ano de registro dos dados.
- **lifeExp:** Expectativa de vida ao nascer (em anos).
- **pop:** População total do país.
- **gdpPercap:** PIB per capita (Produto Interno Bruto per capita).

## 1.3 Conjunto de dados `mtcars`

O conjunto de dados `mtcars` é um conjunto de dados integrado no R, derivado da revista Motor Trend US de 1974. Ele contém dados sobre o consumo de combustível e outras 10 características de design e desempenho de 32 carros. Este conjunto de dados é amplamente utilizado para ensinar técnicas de análise de dados e visualização em R.

### 1.3.1 Estrutura do Conjunto de Dados

O `mtcars` possui 32 observações (carros) e 11 variáveis. As variáveis são:

- `mpg` (Milhas por Galão): Consumo de combustível.
- `cyl` (Número de Cilindros): Número de cilindros no motor.
- `disp` (Deslocamento): Volume do motor (polegadas cúbicas).
- `hp` (Potência): Potência do motor (cavalos de potência).
- `drat` (Relação do Eixo Traseiro): Relação do eixo traseiro.
- `wt` (Peso): Peso do carro (em milhares de libras).
- `qsec` (Tempo de 1/4 de Milha): Tempo para percorrer 1/4 de milha (em segundos).
- `vs` (Forma do Motor): Tipo de motor (0 = em V, 1 = reto).
- `am` (Transmissão): Tipo de transmissão (0 = automática, 1 = manual).
- `gear` (Número de Marchas): Número de marchas.
- `carb` (Número de Carburadores): Número de carburadores.

## 1.4 Conjunto de dados `diamonds`

O conjunto de dados `diamonds` é um dos datasets incluídos no pacote `ggplot2` e é amplamente utilizado para demonstrações de visualização de dados. Este dataset contém informações sobre 53.940 diamantes, incluindo várias características que influenciam o preço dos diamantes.

### 1.4.1 Estrutura do Conjunto de Dados

O conjunto de dados `diamonds` possui as seguintes variáveis:

- `carat`: Peso do diamante (em quilates).
- `cut`: Qualidade do corte do diamante (Fair, Good, Very Good, Premium, Ideal).
- `color`: Cor do diamante, de J (pior) a D (melhor).
- `clarity`: Clareza do diamante, com categorias de I1 (pior) a IF (melhor).
- `depth`: Profundidade total do diamante como uma porcentagem da média do diâmetro.
- `table`: Largura da parte superior do diamante em relação à maior parte do diâmetro.
- `price`: Preço do diamante (em dólares).
- `x`: Comprimento do diamante (em milímetros).
- `y`: Largura do diamante (em milímetros).
- `z`: Profundidade do diamante (em milímetros).

## 2 Realização dos exercícios propostos

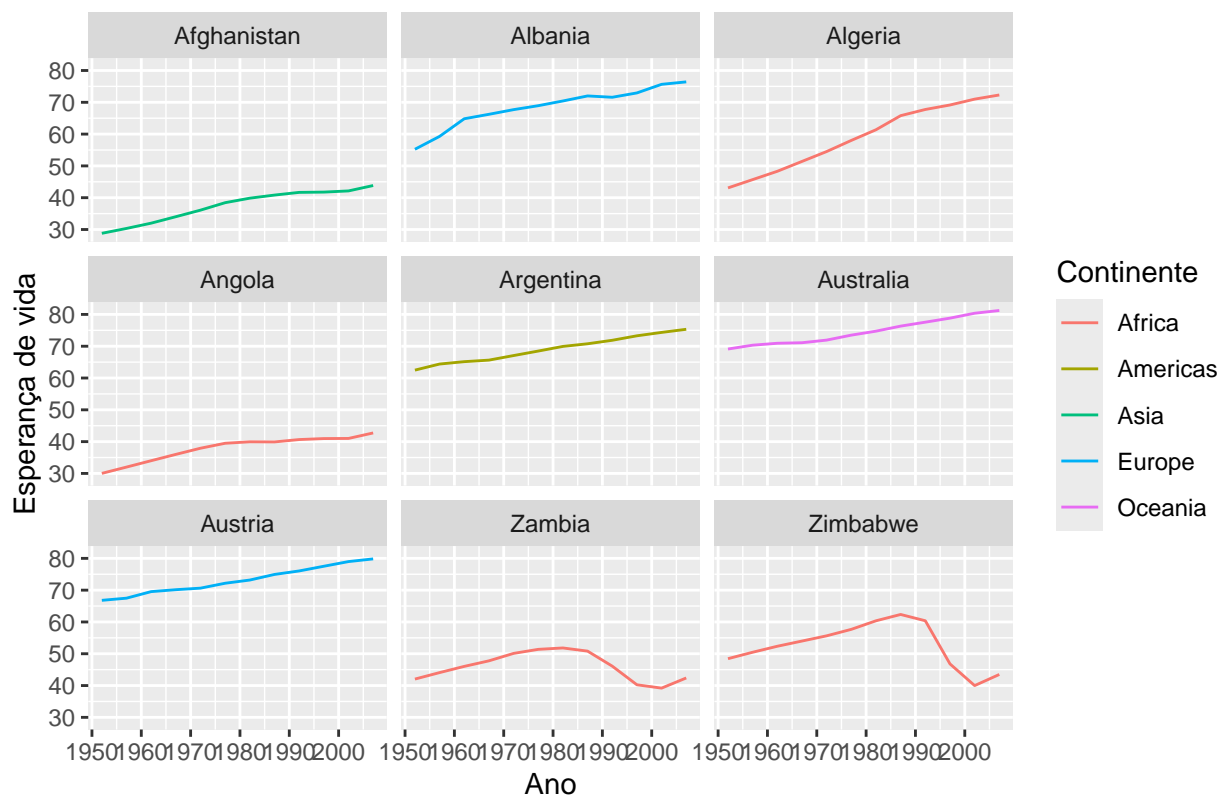
### 2.1 Questão 1 - Adicionando texto a gráfico

Proposta: Adicionar texto a um gráfico criado com o pacote `ggplot2` em R com a utilização do conjunto de dados `gapminder`.

```
library(gapminder); library(ggplot2)
```

#### 2.1.1 Gráfico

Figura 1



#### 2.1.2 Script do gráfico

```
comeca.com <- substr(gapminder$country, start = 1, stop = 1)
países.az <- gapminder[comeca.com %in% c("A", "Z"), ]

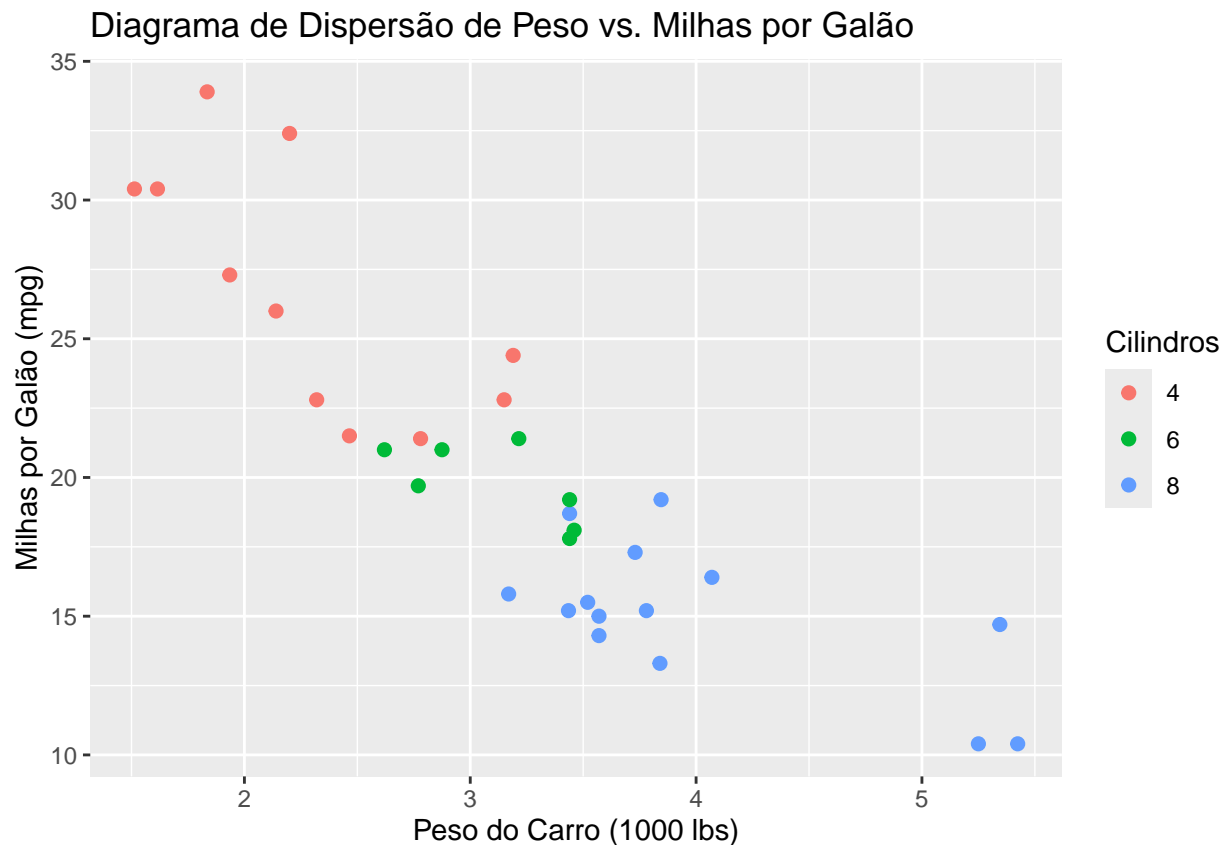
ggplot(data = países.az, aes(x = year, y = lifeExp, color = continent)) +
  geom_line() + facet_wrap(~ country) + ggtitle("Figura 1") + xlab("Ano") +
  ylab("Esperança de vida") + scale_colour_discrete(name = "Continente")
```

## 2.2 Questão 2 - Diagrama de dispersão

Proposta: Criar um diagrama de dispersão com o conjunto de dados `mtcars` e o pacote `ggplot2`.

```
data(mtcars)
```

### 2.2.1 Gráfico



### 2.2.2 Script do gráfico

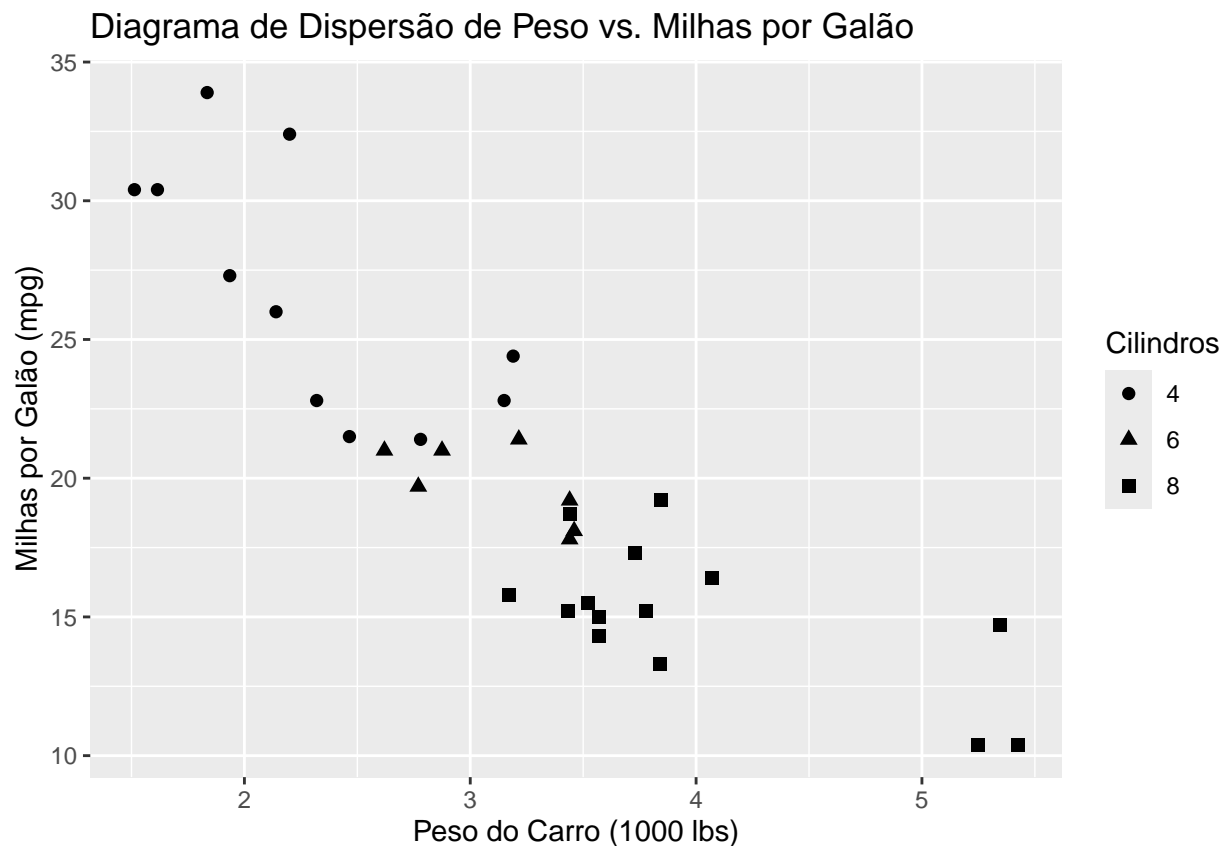
```
mtcars$cyl <- factor(mtcars$cyl)

ggplot(mtcars, aes(x = wt, y = mpg, color = cyl)) +
  geom_point(size = 2) +
  labs(title = "Diagrama de Dispersão de Peso vs. Milhas por Galão",
       x = "Peso do Carro (1000 lbs)",
       y = "Milhas por Galão (mpg)",
       color = "Cilindros")
```

## 2.3 Questão 3 - Formatos dos símbolos dos dados

Proposta: Criar um novo diagrama de dispersão com os dados do exercício anterior, usando formatos diferentes de símbolos para distinguir as diversas quantidades de cilindros do motor.

### 2.3.1 Gráfico



### 2.3.2 Script do gráfico

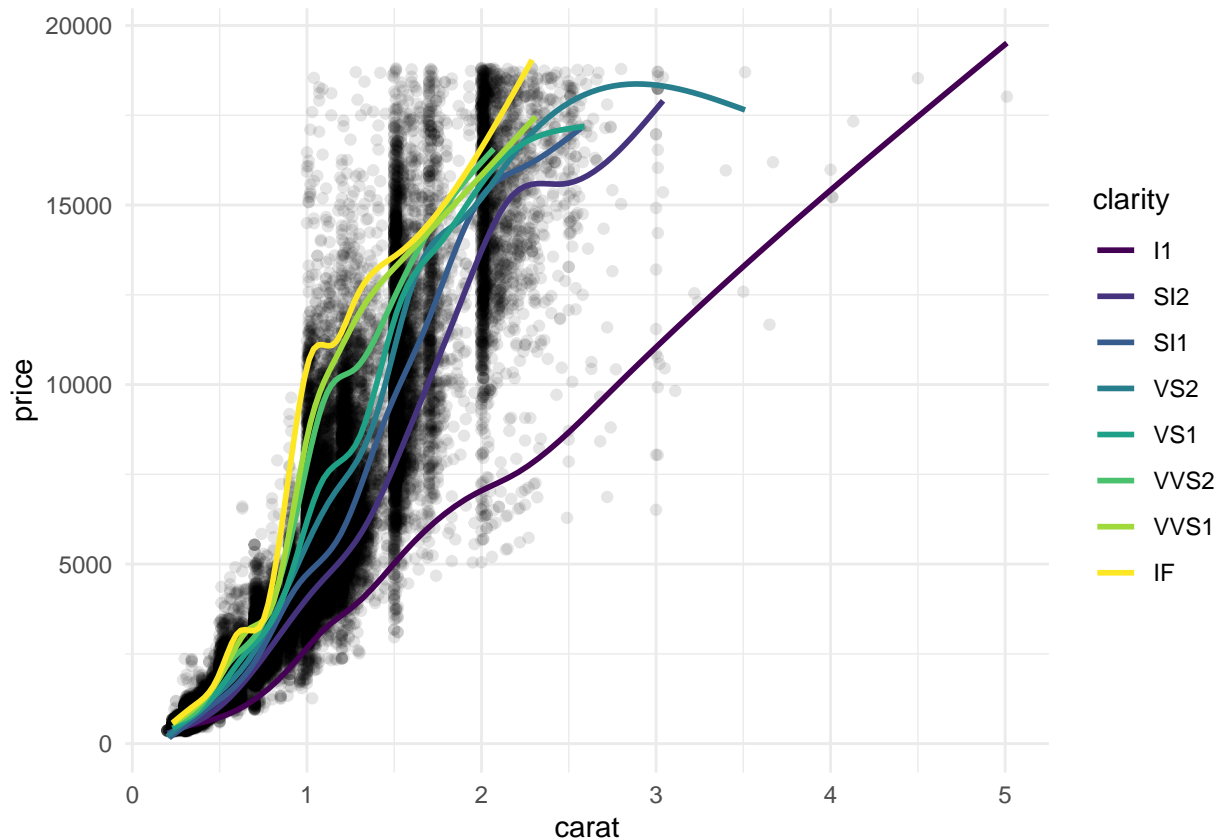
```
ggplot(mtcars, aes(x = wt, y = mpg, shape = cyl)) +  
  geom_point(size = 2) +  
  labs(title = "Diagrama de Dispersão de Peso vs. Milhas por Galão",  
        x = "Peso do Carro (1000 lbs)",  
        y = "Milhas por Galão (mpg)",  
        shape = "Cilindros")
```

## 2.4 Questão 4 - Atribuindo função ggplot2 a objeto

Proposta: Criar um diagrama de dispersão do conjunto de dados `diamonds`, atribuindo a função `ggplot` a um objeto.

```
data(diamonds)
```

### 2.4.1 Gráfico



### 2.4.2 Script do gráfico

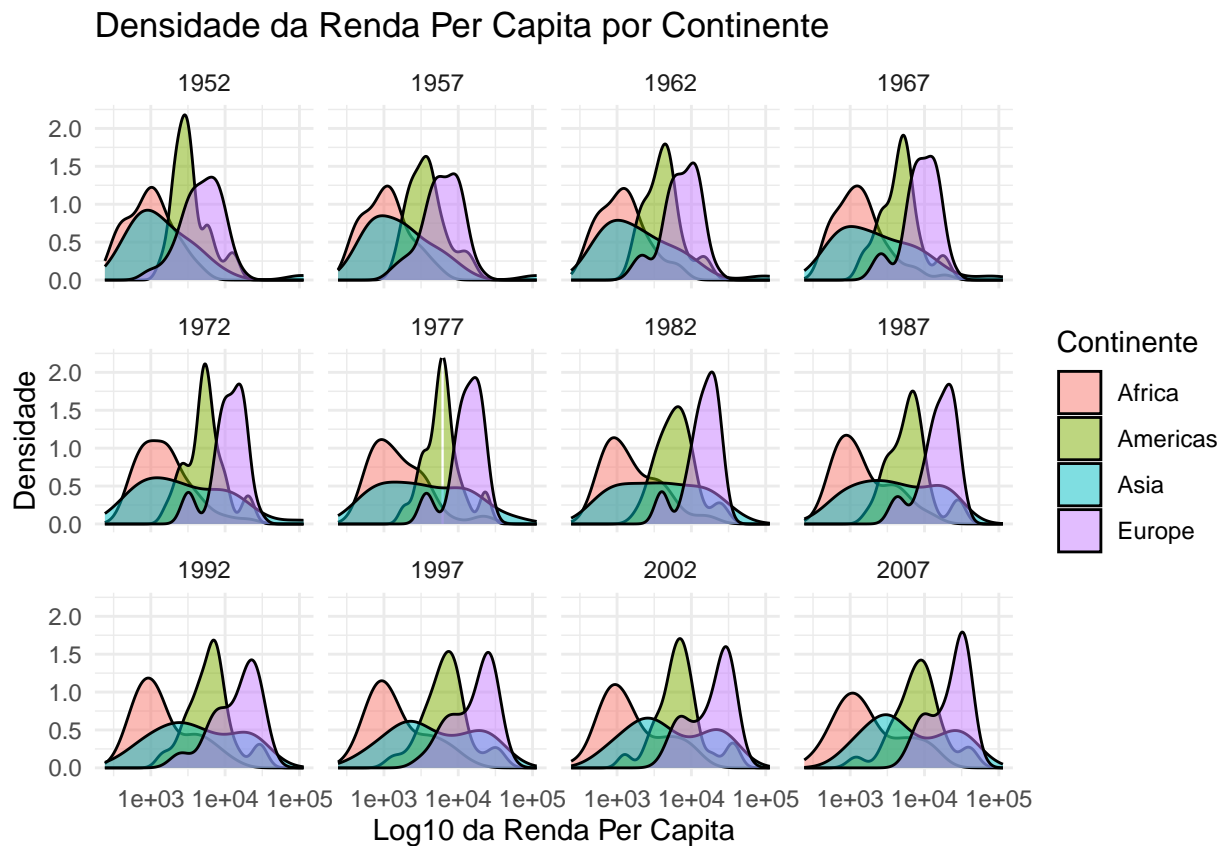
```
meu.grafico <- ggplot(diamonds, aes(x = carat, y = price)) +  
  geom_point(alpha = 0.1) +  
  geom_smooth(aes(color = clarity), se = FALSE) + theme_minimal()  
  
# OBS: Estava aparecendo a seguinte mensagem ao exibir o gráfico:  
# 'geom_smooth()' using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'  
# Por isso utilizei a função `suppressMessages(print())`  
suppressMessages(print(meu.grafico))
```



## 2.5 Questão 5 - Criando gráfico de área

Proposta: Criar gráficos de densidade da renda per capita, preenchidos com cores de acordo com o continente através do conjunto de dados `gapminder` e do pacote `ggplot2`.

### 2.5.1 Gráfico



### 2.5.2 Script do gráfico

```
# Removendo o continente da Oceania
gapminder_filtered <- gapminder[gapminder$continent != "Oceania", ]

ggplot(gapminder_filtered, aes(x = gdpPercap, fill = continent)) +
  geom_density(alpha = 0.5) + facet_wrap(~ year, nrow = 3, ncol = 4) +
  labs(title = "Densidade da Renda Per Capita por Continente",
       x = "Log10 da Renda Per Capita",
       y = "Densidade", fill = "Continente") + scale_x_log10() +
  scale_y_continuous(limits = c(0, 2.2)) +
  theme_minimal()
```

### 3 Considerações finais

O pacote `ggplot2` é uma ferramenta versátil e robusta para visualização de dados no ambiente R. Com sua ampla gama de funcionalidades, é possível criar uma variedade de gráficos com uma sintaxe simples e intuitiva, simplificando assim a análise e a tendências nos dados.

Ao longo das questões fornecidas, foi possível explorar várias funcionalidades do `ggplot2` e dos conjuntos de dados `gapminder`, `mtcars` e `diamonds`. Esses conjuntos de dados forneceram uma base sólida para a criação de gráficos informativos e aprofundados, permitindo explorar diferentes aspectos dos dados e extrair insights valiosos para análises.

## 4 Links

- Documentação do pacote `ggplot2`
- Documentação do conjunto de dados `gapminder`
- Documentação do conjunto de dados `mtcars`
- Documentação do conjunto de dados `diamonds`