Introdução à Análise de Dados com Linguagem R

Aula 2

Analista Ambiental Robson Cruz

Contents


```
{r setup, include = FALSE} knitr::opts_chunk$set(fig.align = 'center', echo = TRUE)
```

1 Operadores Aritméticos e de Atribuição em R

Operador	Função
+	Soma
-	Subtração
/	Divisão
*	Multiplicação
%%	Resto da divisão
%/%	Parte inteira divisão
^	Potenciação
**	Potenciação
<-	Atribuição
=	Atribuição

1.1 Exercício Prático - Operadores Aritméticos e de Atribuição

Instrução 1/3 * Obter o resto da divisão entre os números inteiros 10 e 3.

Instrução 2/3 * Obter a parte interira da divisão de 10 por 3.

Instrução 3/3 * Obter o quadrado de um número inteiro qualquer.

2 Operadores de Comparação em R

Operador	Significado
==	igual a
!=	diferente de
>	maior que
<	menor que
>=	maior ou igual a
<=	menor ou igual a

TRUE

```
# Leitura de vetores de um inventário florestal
load('./data/dados_modulo_1.rda')

# Mostrar os objetos atualmente disponíveis no ambiente R
ls()
```

'altura'

'categoria'

'dap'

'nomes_cientificos'

3 Função `mapply()`

Versão multivariada das funções lapply e sapply, utilizada para iterar entre elementos de vetores ou listas.

```
# Definição dos Vetores a e b
a <- c(7, 12, 5, 2, 1)
b <- c(4, 2, 3, 5, 1)

# Nomes para os vetores
dias_semana <- c('Segunda', 'Terça', 'Quarta', 'Quinta', 'Sexta')

# Atribuir nome aos vetores
names(a) <- dias_semana
names(b) <- dias_semana

# Uso da função mapply() para retornar a soma
```

```
# entre os elementos dos vetores a e b
print(mapply(max, a, b))
```

Segunda

7

Terça

12

Quarta

5

Quinta

5

Sexta

1

4 Função *tapply()*

Aplica uma função sobre um vetor com agrupamento em outro vetor categórico. Recebe como parâmetros: um vetor numérico, um vetor categórico e uma função. O código a seguir aplica a função média sobre o vetor `volume` agrupado ao vetor `ut` (unidades de trabalho)

```
# Calcular o volume médio por unidade de trabalho
print(tapply(dap, categoria, mean))
```

Explorar

76.8872521591047

Remanescente

67.9580670429674

Substituta

71.9543683510638

5 Testes Lógicos com Vetores

- *any()* Testa se algum elemento do vetor atende a uma condição específica

Exemplo: Dado o vetor de nome `dap`, o qual armaneza dados de mensuração de diâmetro de milhares de árvores na Floresta Nacional de Altamira, teste se algum elemento é menor ou igual a 40.

```
any(dap >= 40)
```

TRUE

- *all()* Testa se todos os elementos de um vetor atendem a uma condição.

Exemplo: Dado o vetor de nome `dap`, testar se algum elemento é menor do que 0:

```
all(dap < 40)
```

FALSE

- `is.na( )` Testa se o vetor contém valores ausentes (*Not Availables*)

5.1 Exercício Prático - Índice de Vetor

Considerando o vetor de nome `dap`, o qual armaneza dados de mensuração de diâmetro de milhares de árvores na Floresta Nacional de Altamira, mostre:

Instrução 1/5 * Quantas árvores foram inventariadas.

Instrução 2/5 * **Apenas** o penúltimo elemento desse vetor.

Instrução 3/5 * O diâmetro mínimo de medição

Instrução 4/5 * O diâmetro máximo mensurado

Instrução 5/5 * O diâmetro médio mensurado

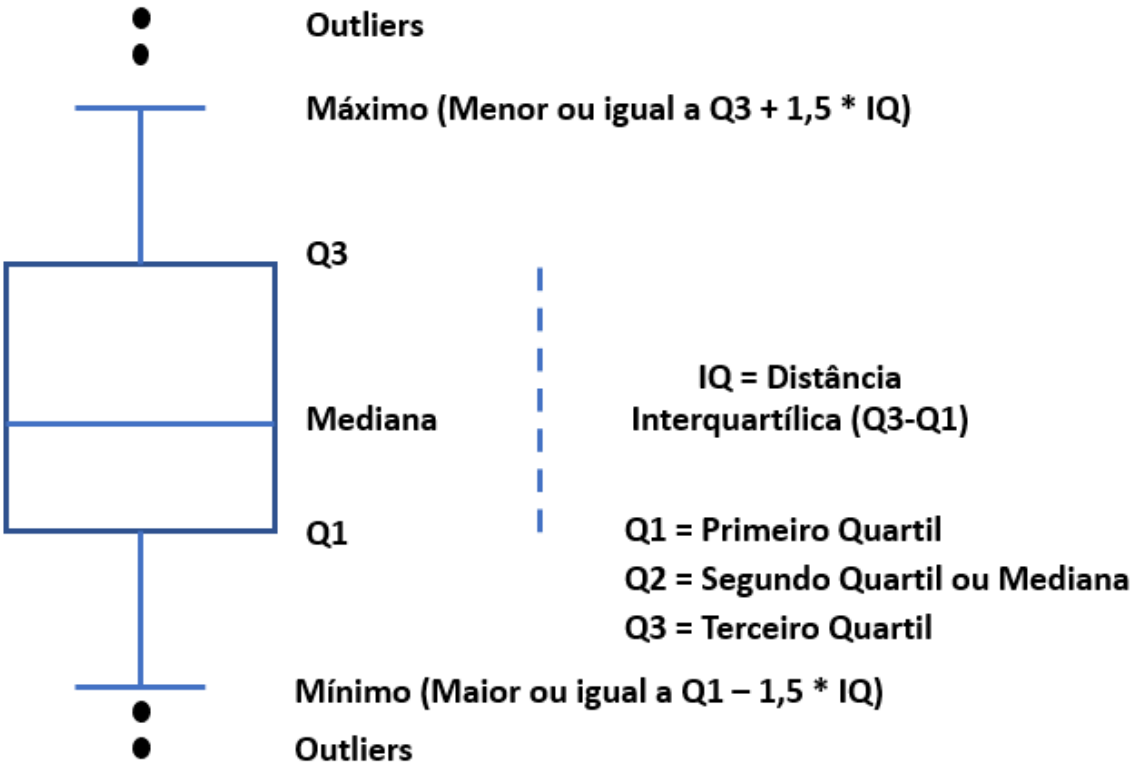
5.2 Visualização Gráfica de Vetores

Hora de tentar algo um pouco diferente. Até agora, você programou script e observou seus dados imprimindo-os. Para uma visualização mais informativa de dados, experimente uma saída gráfica.

Para este exercício, você irá trabalhar com dados de inventário florestal realizado em uma unidade de produção anual da Floresta Nacional de Altamira. Para tal utilizaremos apenas duas variáveis, a saber: diâmetro a altura do peito (DAP) e altura comercial.

5.2.1 Boxplot

O *boxplot* ou diagrama de caixa é uma ferramenta gráfica da estatística que nos permite visualizar a distribuição e valores discrepantes (outliers) de dados.



```
boxplot(altura)
```

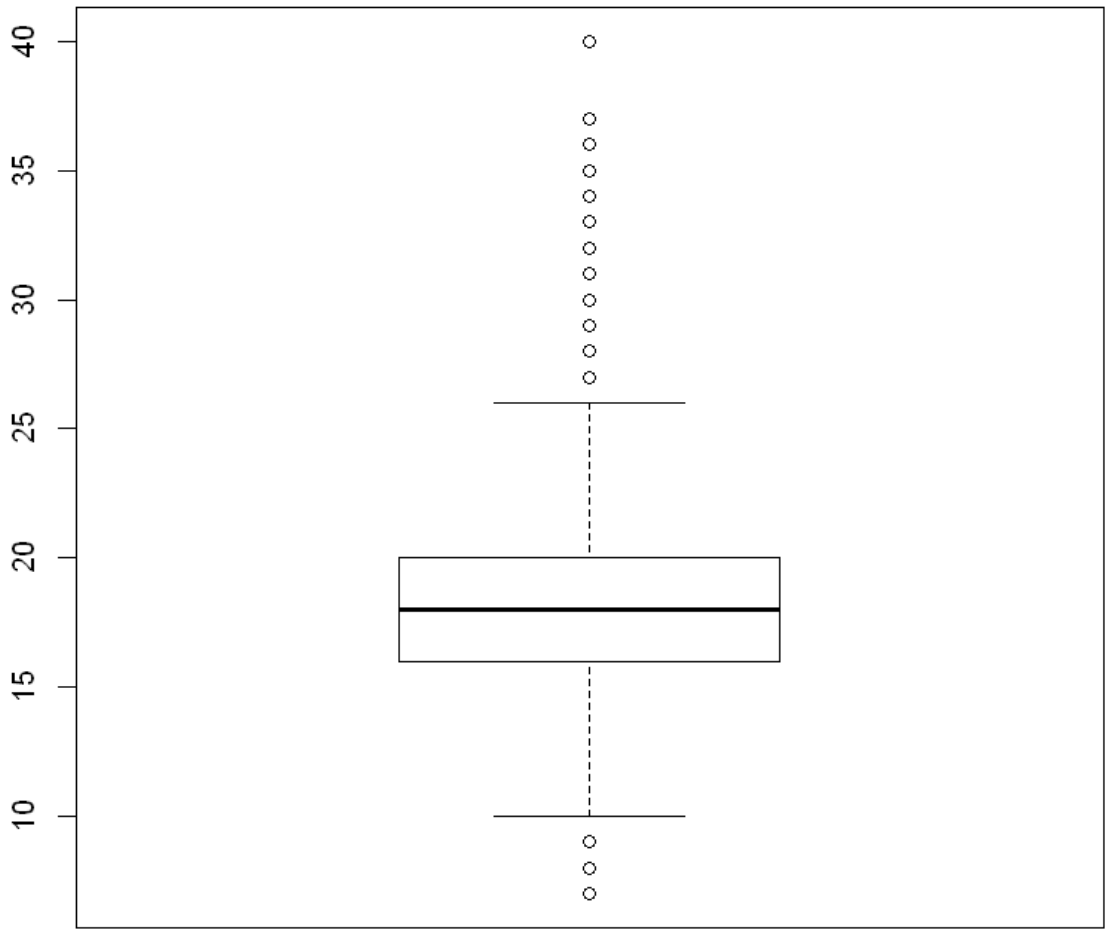


Figure 1: png

5.2.2 Histograma

Utilziamos histogramas para visualizar a distribuição de uma variável contínua. Em R o pacote "base" nos fornece a função `hist( )`.

Exercício Prático

- Mostrar o histograma para os dados da variável DAP disponível no vetor `dap`

```
hist(dap)
```

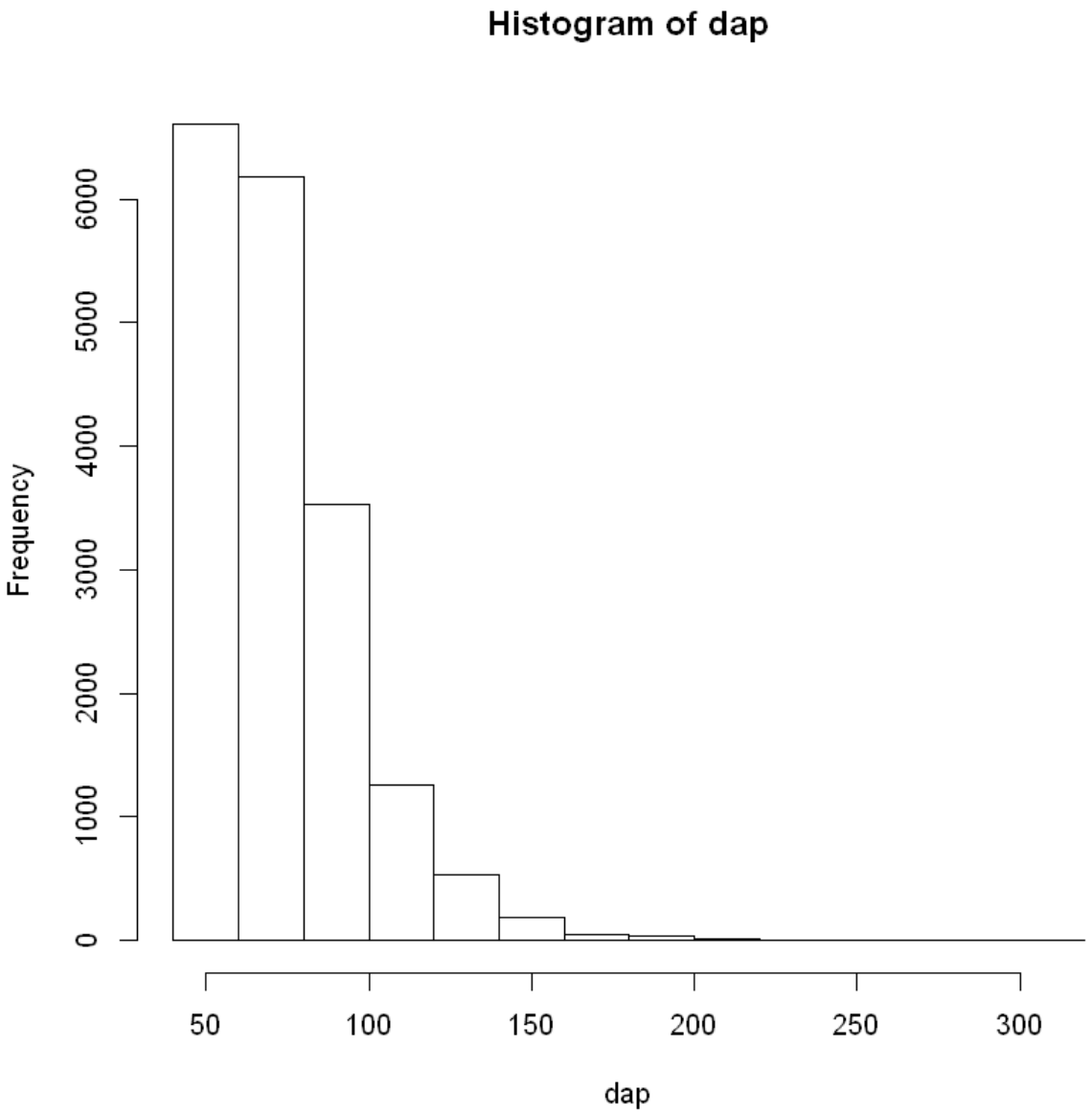


Figure 2: png

5.2.3 Gráfico Dispersão

O gráfico de dispersão é utilizado para visualizar a relação entre duas variáveis contínuas. Para gerar um gráfico de dispersão em R devemos utilizar a função `plot` do pacote "base".

Exercício Prático

Visualizar a relação entre a varíavel altura e diâmetro.

```
{r fig.align = 'center'} plot(altura, dap)
```

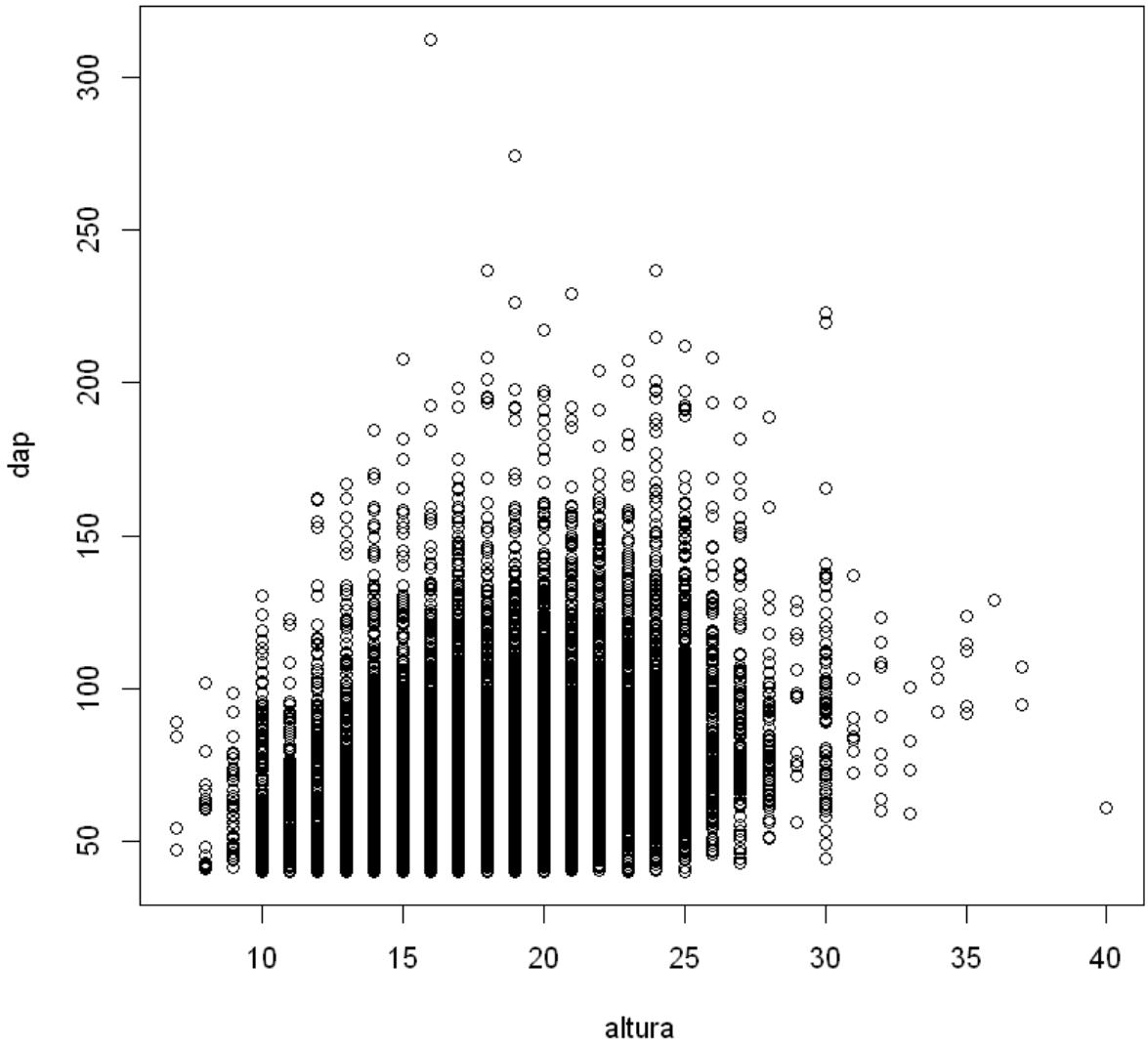


Figure 3: png

5.3 Modificar a Aparência dos Gráficos

A configuração dos parâmetros de estilo, tamanho e agrupamento dos gráficos pode ser obtida digitando o comando ? par. Para este curso introdutório utilizaremos apenas o básico da configuração da aparecência de gráficos no pacote "base" do R.

5.3.1 Personalizando Histogramas

- `main` Utilizado para atribuir ou modificar um título do gráfico

```
{r fig.align = 'center'}  # Modificar o título do gráfico hist(dap, main = 'Distribuição
da variável DAP (cm)')
```

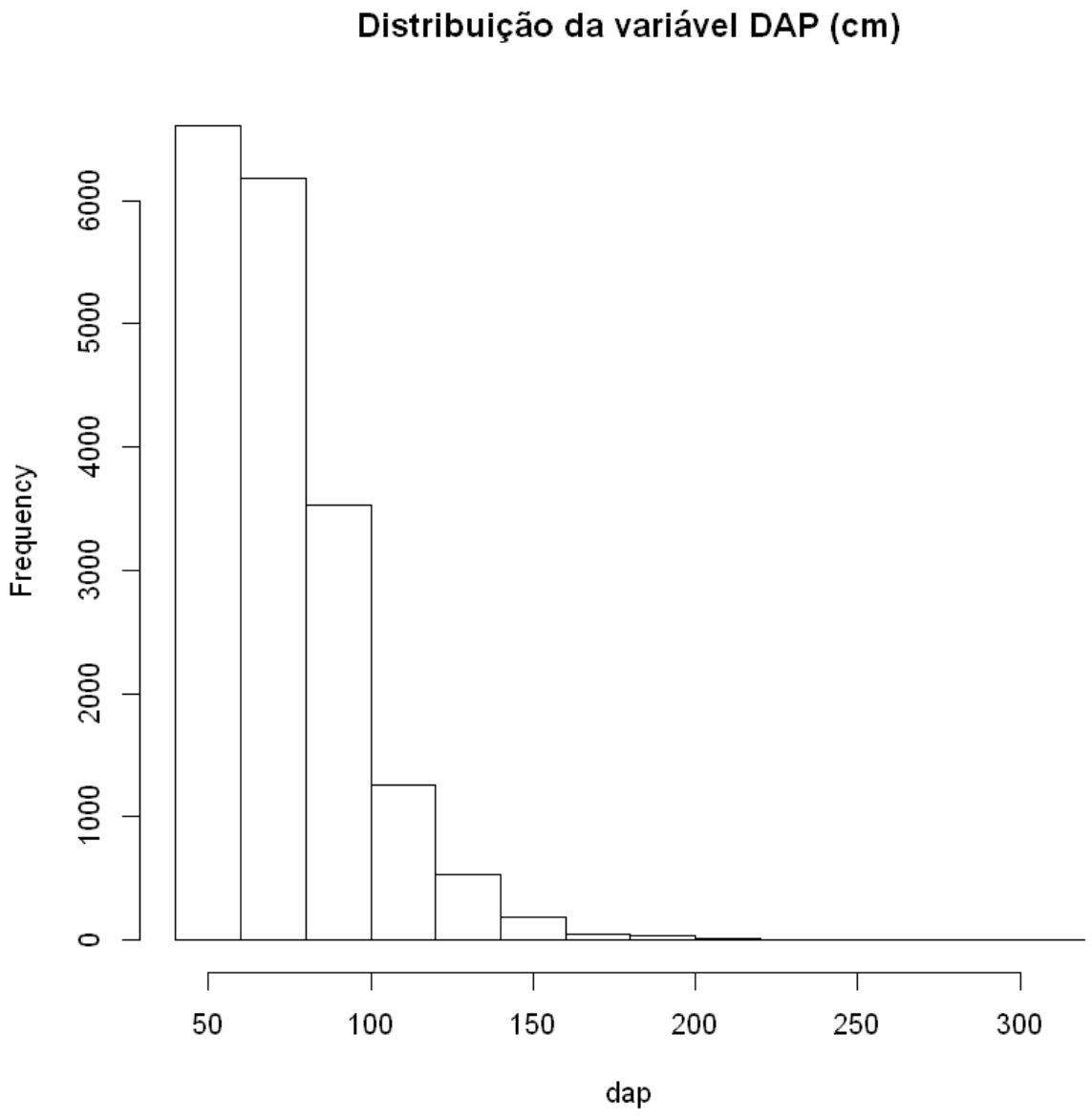


Figure 4: png

- `xlab` e `ylab` - Modificar os nomes dos eixos x e y.

```
{r fig.align = 'center'} # Modificar os rótulos dos eixos x e y hist(dap,       main =
'Distribuição da variável DAP', # título do gráfico     xlab = 'DAP (cm)',   # Rótulo do
eixo x     ylab = 'Frequência') # Rótulo do eixo y
```

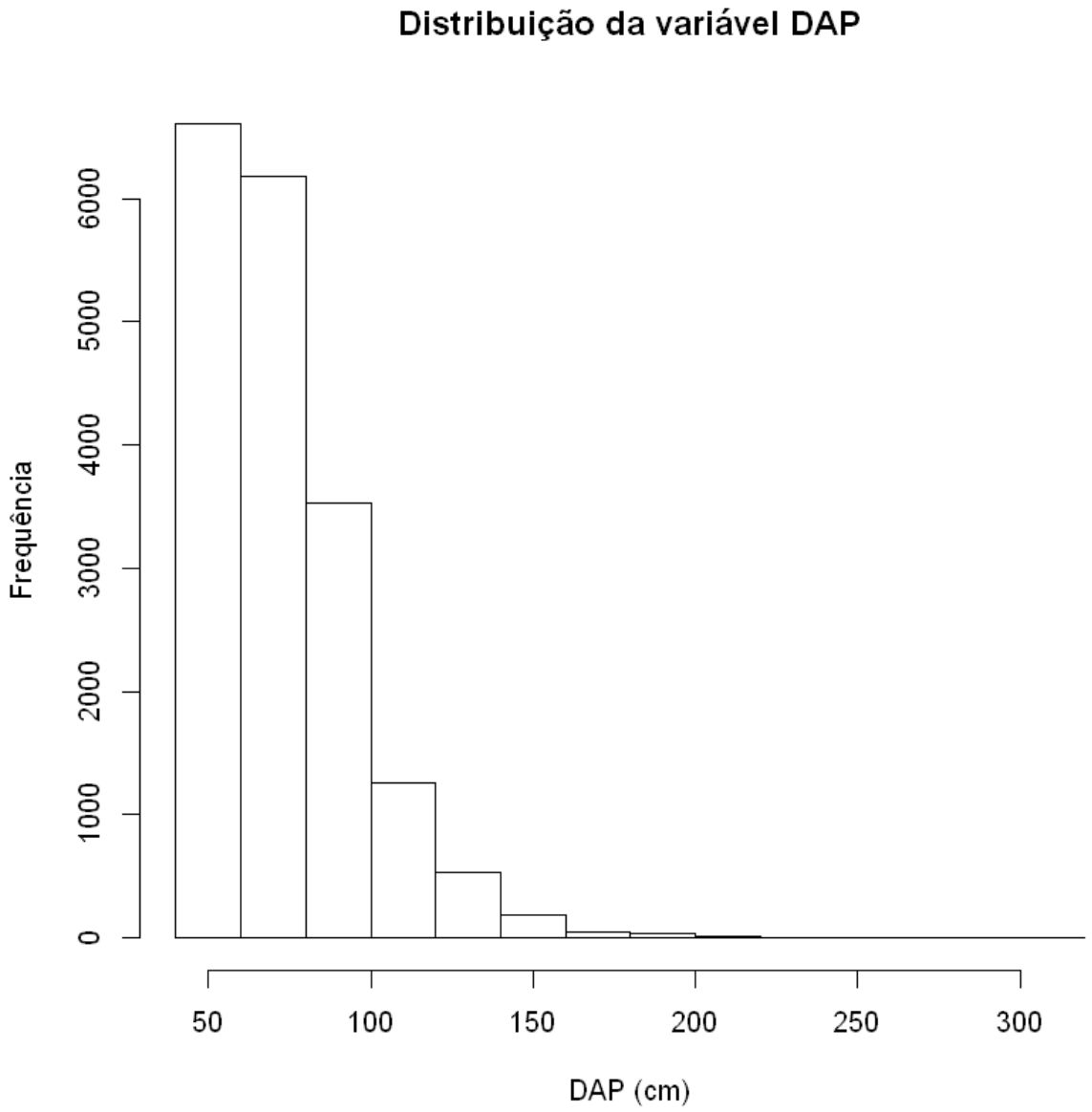


Figure 5: png

- `labels` - Mostra os valores de cada barra do histograma.

```
{r fig.align = 'center'} hist(dap,       main = 'Distribuição da variável DAP',     xlab
= 'DAP (cm)',     ylab = 'Frequência',     labels = TRUE)
```

- `col` - Muda a cor das barras do histograma.

```
{r fig.align = 'center'} hist(dap,       main = 'Distribuição da variável DAP',     xlab
```

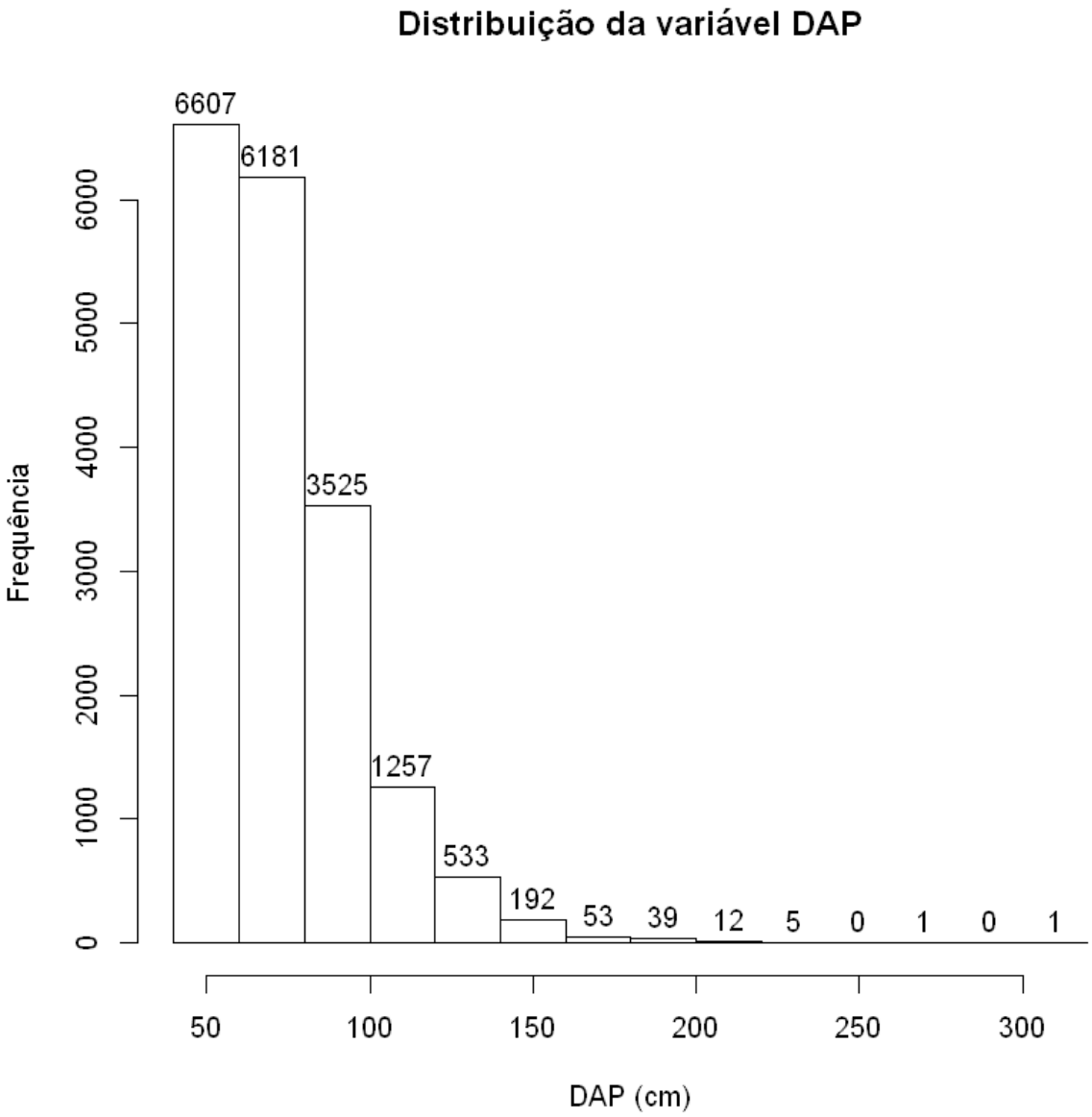

Distribuição da variável DAP
Frequência
0
1000
2000
3000
4000
5000
6000
6607
6181
3525
1257
533
192
53
39
12
5
0
1
0
1
50
100
150
200
250
300
DAP (cm)

Figure 6: png

```
= 'DAP (cm)',     ylab = 'Frequência',     labels = TRUE,     col = 'darkgreen')
```

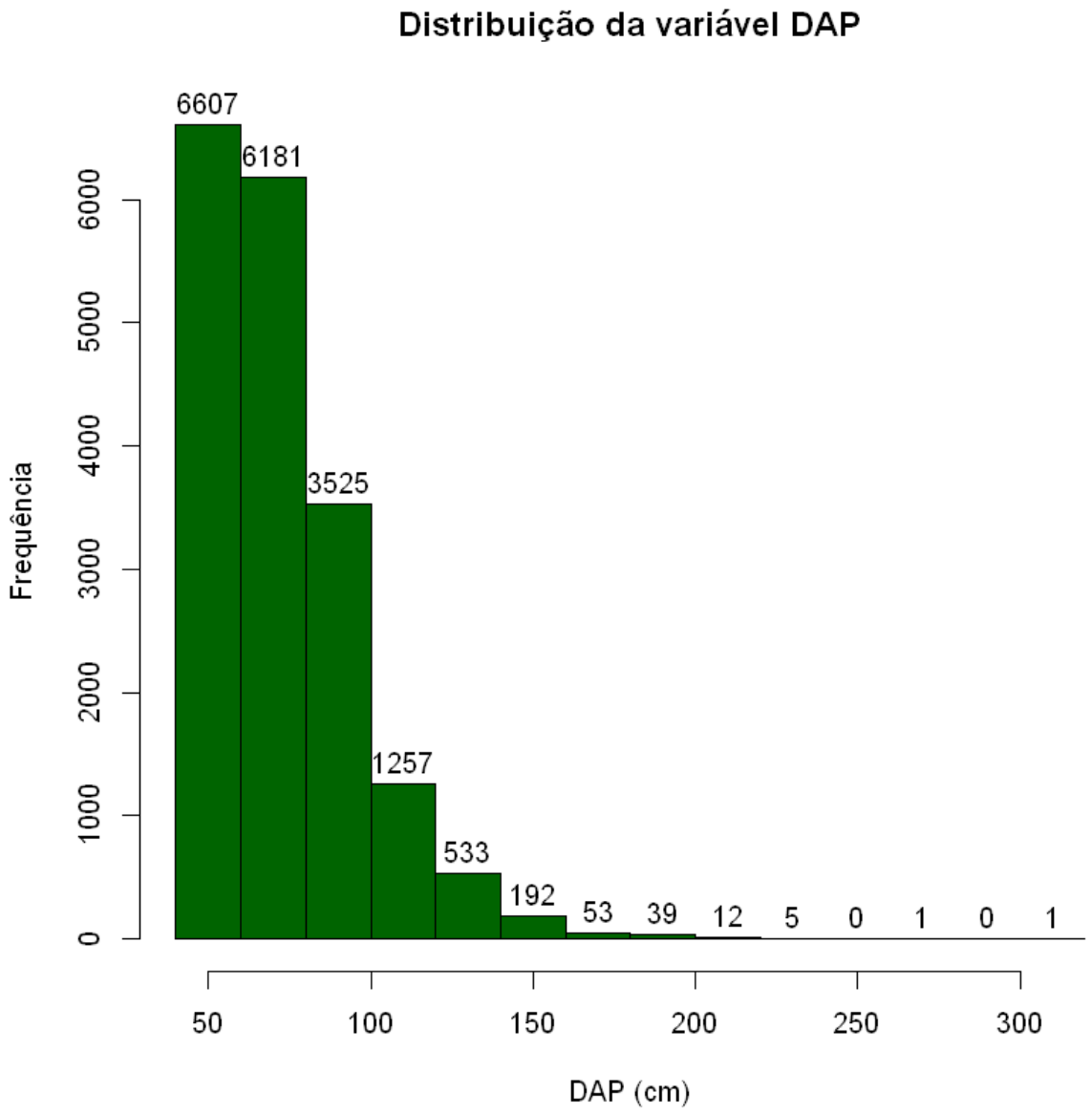


Figure 7: png

- `density` e `angle` - Mostram as barras do histograma hachuradas

```
{r fig.align = 'center'} hist(dap,       main = 'Distribuição da variável DAP',     xlab
= 'DAP (cm)',     ylab = 'Frequência',     labels = TRUE,     col = 'steelblue',     density
= 15,     angle = 60)
```

- `Abline( )` - Função para adicionar uma linha reta ao histograma. Para adicionar uma linha vertical deve-se utilizar o argumento `v` e para linha horizontal `h`. O tipo de linha é modificado através do argumento `lty` (*line type*) e a espessura da linha através do argumento `lwd` (*line width*).

O exemplo a seguir mostra como adicionar uma linha ao histograma para representar a média dos diâmetros.

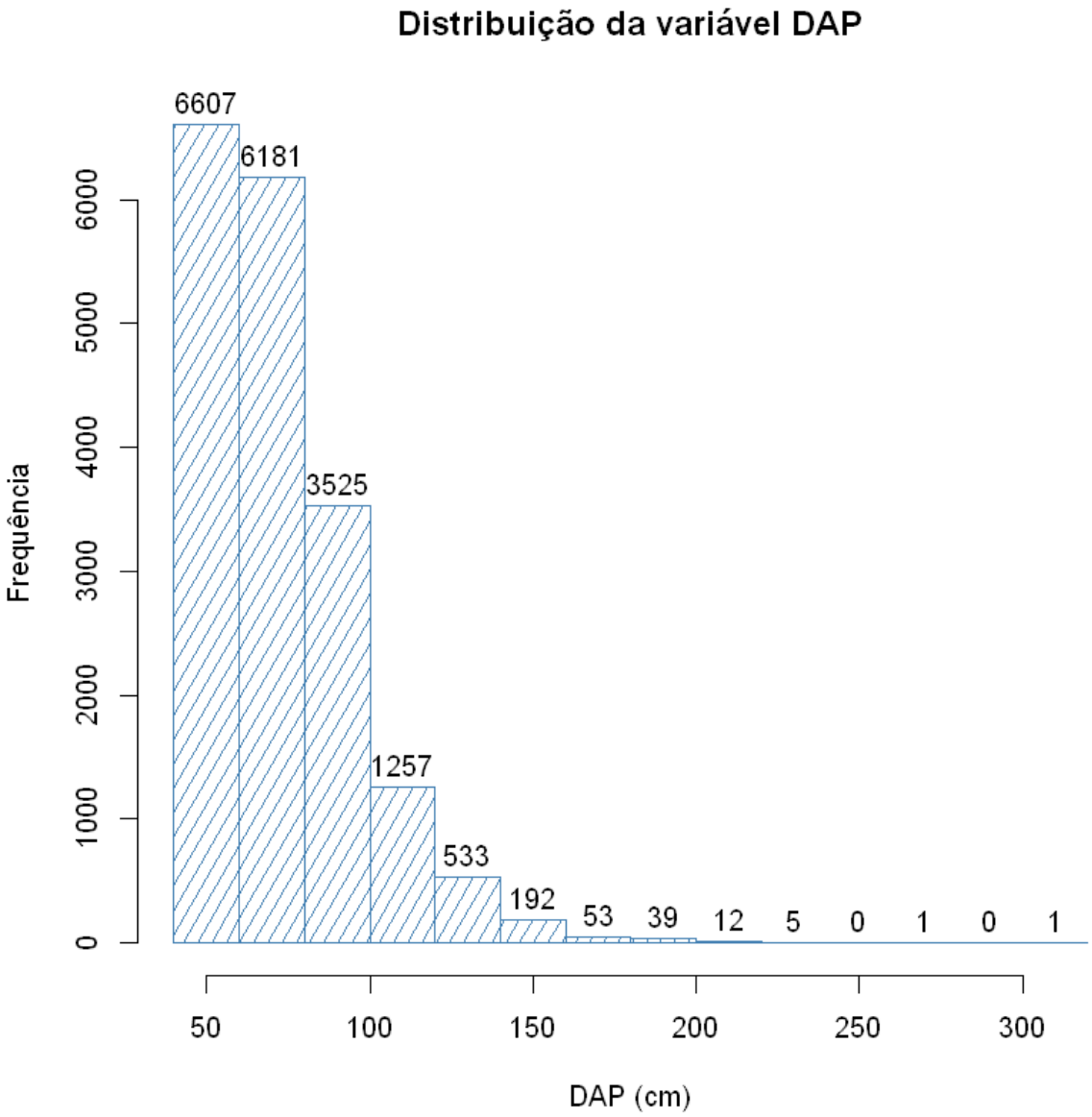

Distribuição da variável DAP
Frequência
0
1000
2000
3000
4000
5000
6000
6607
6181
3525
1257
533
192
53
39
12
5
0
1
0
1
50
100
150
200
250
300
DAP (cm)

Figure 8: png

“‘{r fig.align = ‘center’} hist(dap, main = ‘Distribuição da variável DAP’, xlab = ‘DAP (cm)’, ylab = ‘Frequência’, labels = TRUE, col = ‘steelblue’, density = 15, angle = 60)

abline(v = mean(dap), col = ‘red’, lty = 2, lwd = 2)

```
![png](output_58_0.png)
```

```
### Personalizando Gráficos de Dispersão

* `pch` - Altera o tipo de caractere dos pontos

![](./data/pch-symbols.png)

* Forma, Tamanho e Cor dos Pontos

Argumento | Saída
:--------:|:------:
`col` | Cor da borda do ponto.
`bg` | Cor do Fundo do ponto.
`cex` | Tamanho do ponto.
`lwd` | Espessura da Borda do Ponto.


```{r fig.align = 'center'}
# Alterar o tipo de caractere dos pontos
plot(altura, dap, pch = 25)
```

- `las` (*label style*) - Rotação dos rótulos dos eixos x e y.

`las`	Rótulo
0	Paralelo aos eixos
1	Sempre na Horizontal
2	Sempre na Perpendicular
3	Sempre na Vertical

{r fig.align = 'center'} plot(altura, dap, pch = 21, las = 0)

{r fig.align = 'center'} plot(altura, dap, pch = 21, las = 2)

5.3.2 Agrupar Gráficos em uma Única Figura

“‘{r fig.align = ‘center’} # Parâmetros gráficos par(mfcol = c(2, 2))

hist(dap, pin = c(12, 8)) hist(altura) plot(altura, dap, pch = 20) boxplot(dap) “‘
```

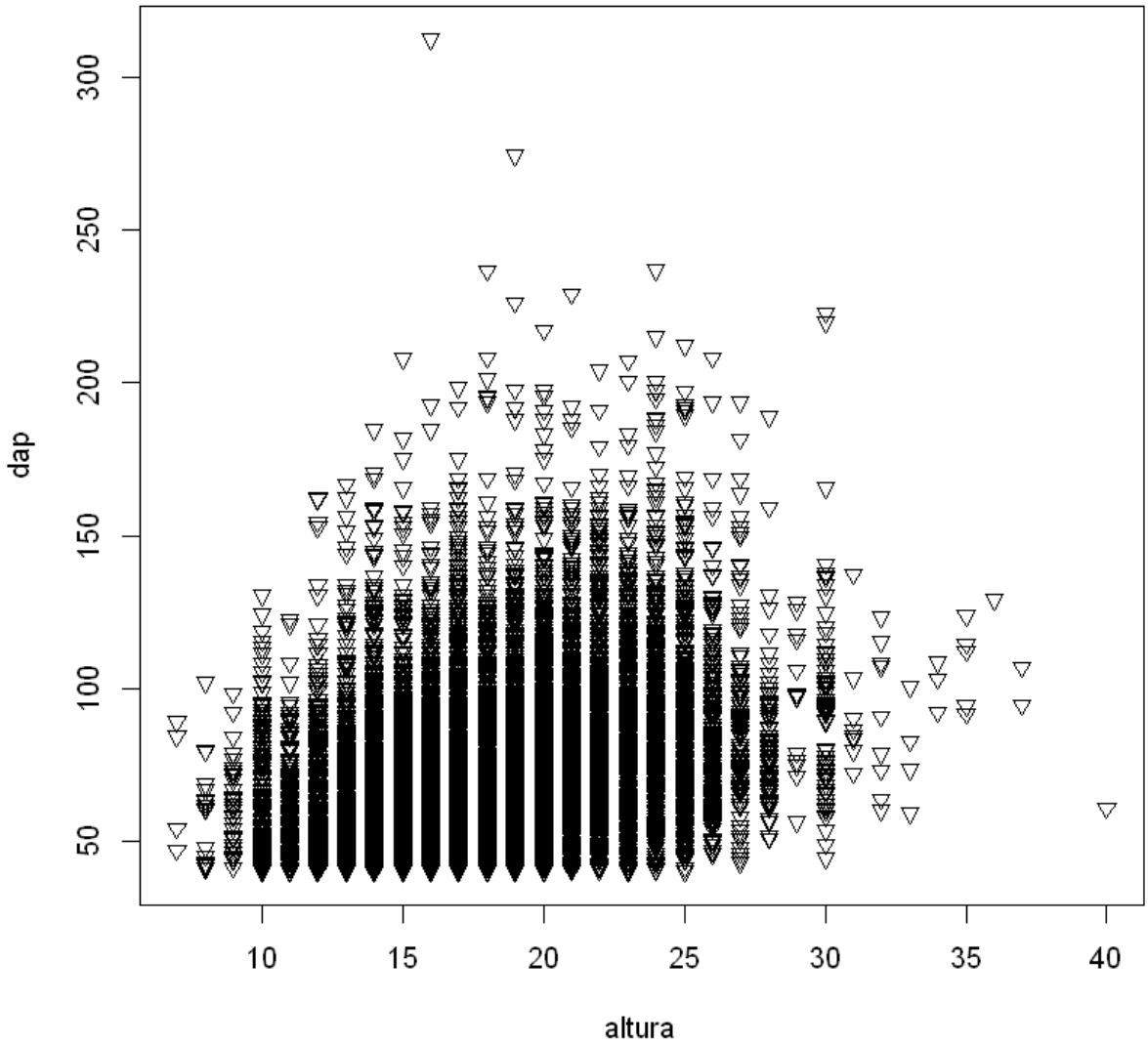


Figure 9: png

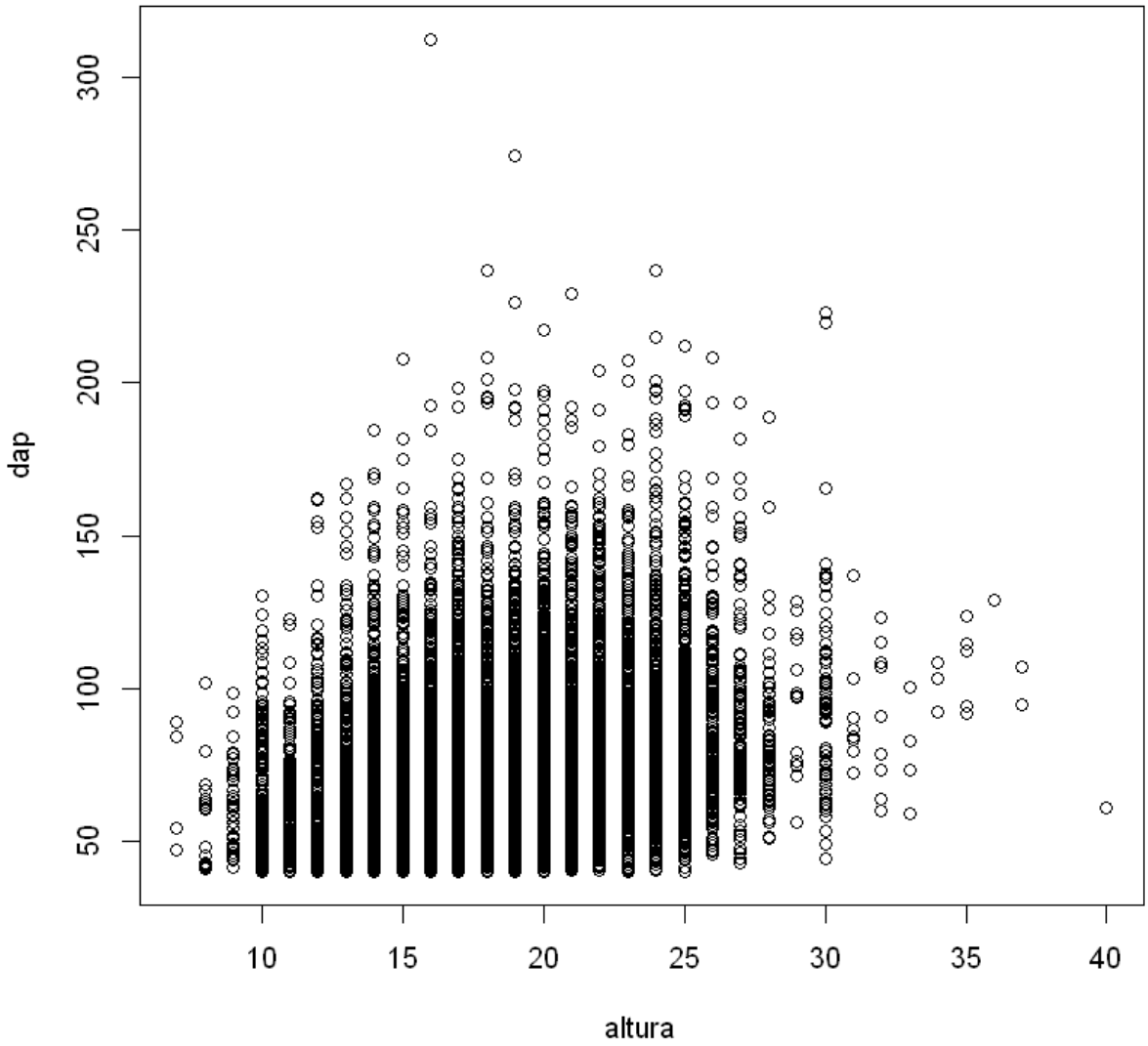


Figure 10: png

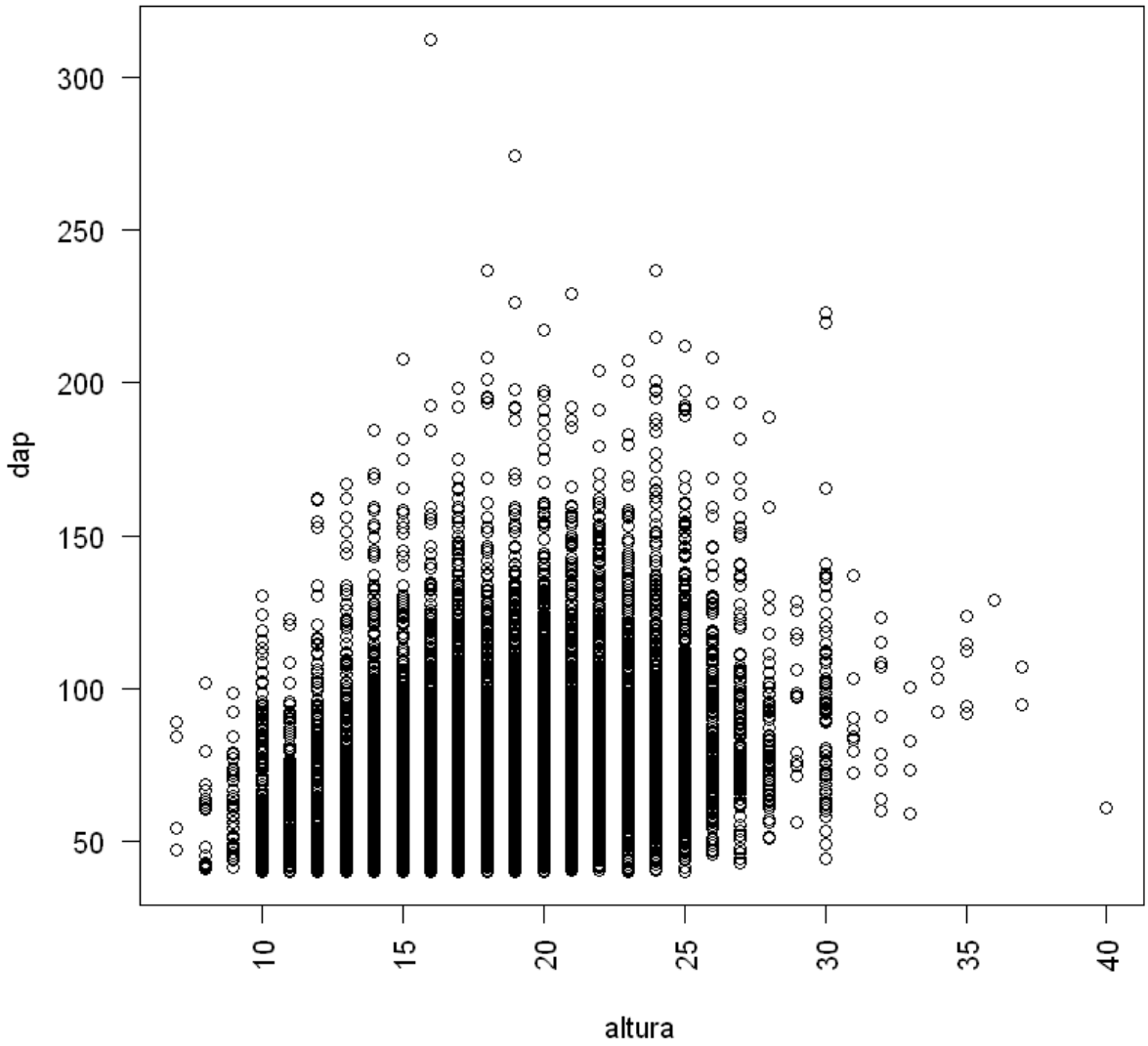


Figure 11: png

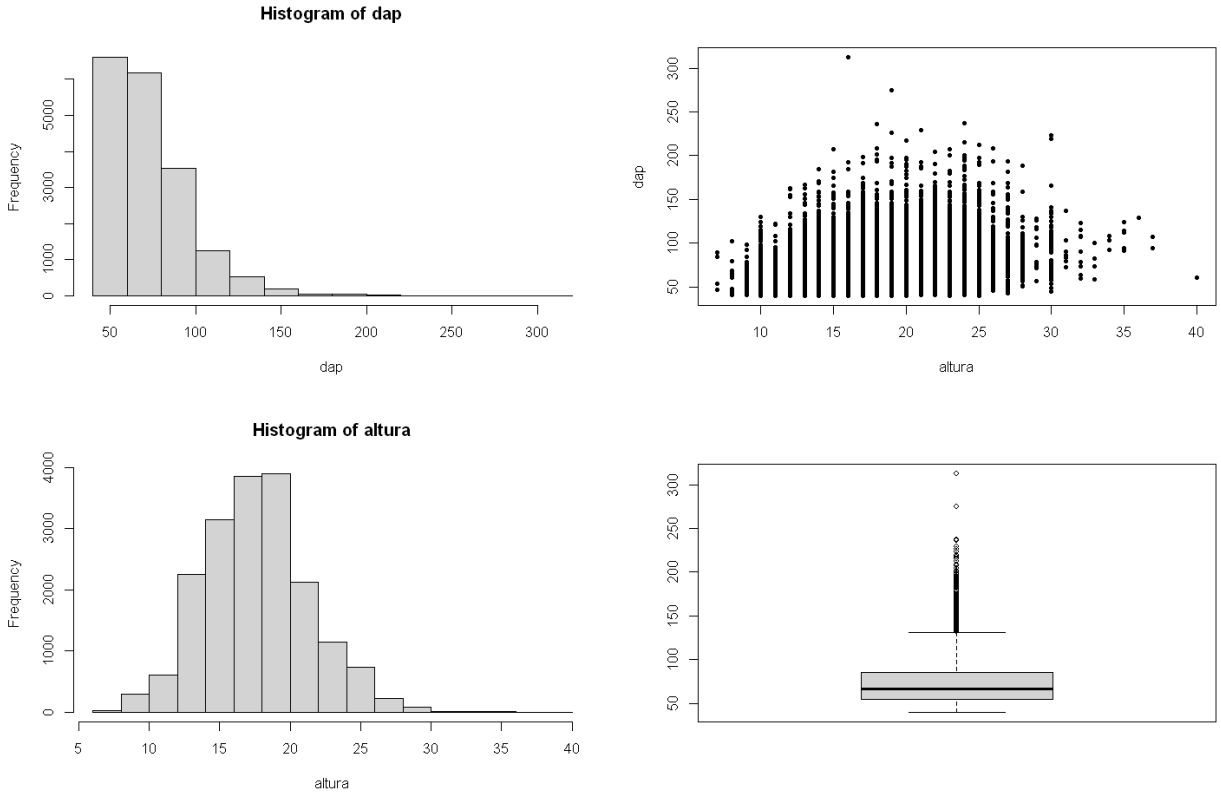


Figure 12: png