

Análise Exploratória de Dados Agrícolas

Fábio Marcos Pedroso Filho

2024-10-21

Carregar Dados e Resumo

```
# Carregar o conjunto de dados
df <- read.csv("/Users/pedrosof/Documents/FIAP/Trabalhos/Fase3_Cap14/Atividade_Cap_14_produtos_agricolas.csv")
summary(df)
```

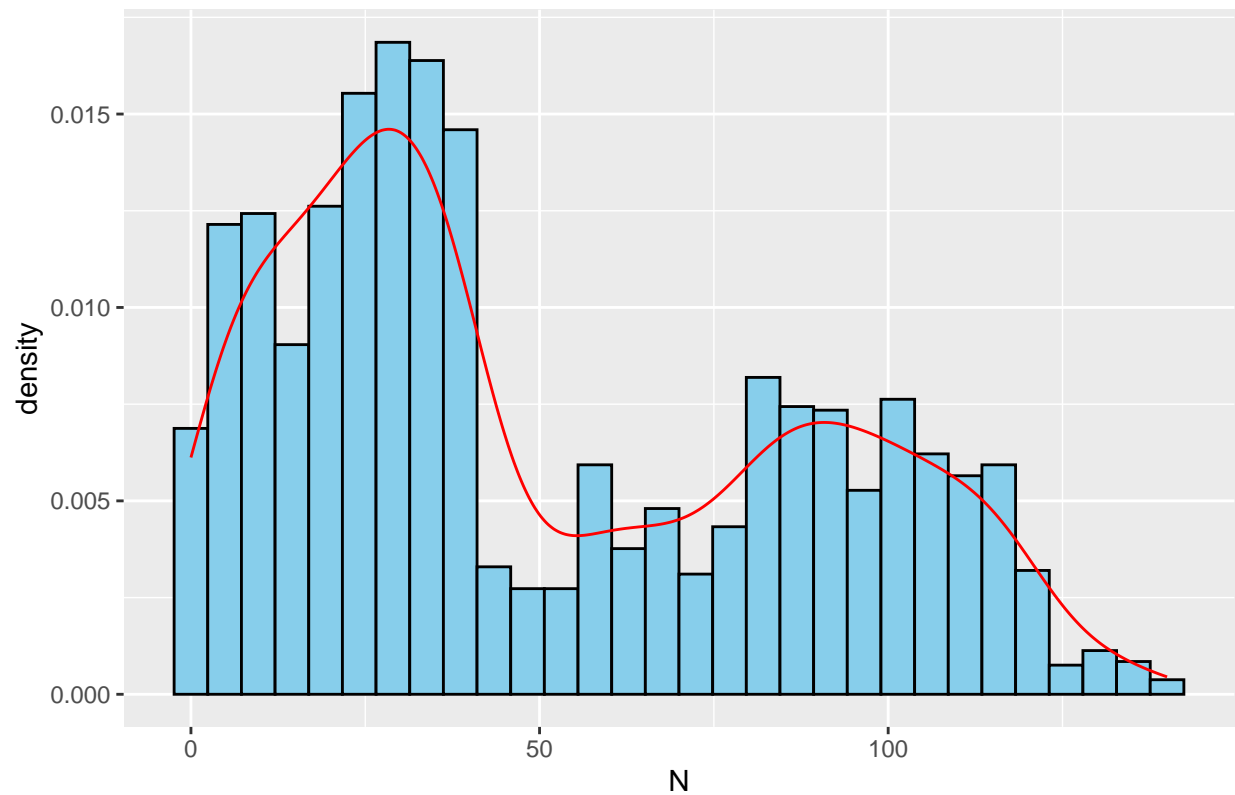
```
##           N                P                K          temperature
## Min.      : 0.00    Min.      : 5.00    Min.      : 5.00    Min.      : 8.826
## 1st Qu.: 21.00    1st Qu.: 28.00    1st Qu.: 20.00    1st Qu.:22.769
## Median : 37.00    Median : 51.00    Median : 32.00    Median :25.599
## Mean     : 50.55    Mean      : 53.36    Mean      : 48.15    Mean     :25.616
## 3rd Qu.: 84.25    3rd Qu.: 68.00    3rd Qu.: 49.00    3rd Qu.:28.562
## Max.     :140.00    Max.      :145.00    Max.      :205.00    Max.     :43.675
## humidity      ph      rainfall      label
## Min.      :14.26    Min.      :3.505    Min.      : 20.21    Length:2200
## 1st Qu.:60.26    1st Qu.:5.972    1st Qu.: 64.55    Class :character
## Median :80.47    Median :6.425    Median : 94.87    Mode  :character
## Mean      :71.48    Mean      :6.469    Mean      :103.46
## 3rd Qu.:89.95    3rd Qu.:6.924    3rd Qu.:124.27
## Max.      :99.98    Max.      :9.935    Max.      :298.56
```

Distribuição de Nitrogênio

```
ggplot(df, aes(x = N)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "skyblue", color = "black") +
  geom_density(color = "red") +
  ggtitle("Distribuição de Nitrogênio (N)")
```

```
## Warning: The dot-dot notation ('..density..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(density)' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

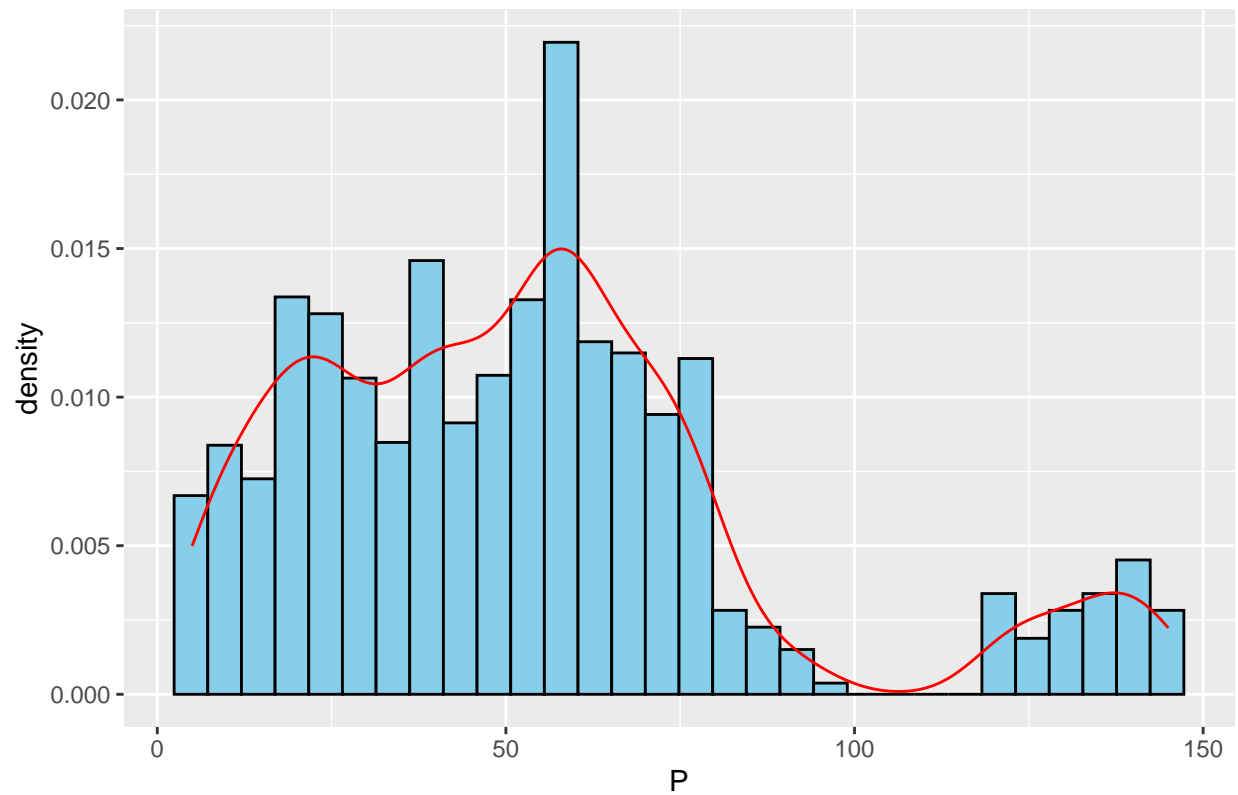
Distribuição de Nitrogênio (N)



Distribuição de Fósforo

```
ggplot(df, aes(x = P)) +  
  geom_histogram(aes(y = ..density..), bins = 30, fill = "skyblue", color = "black") +  
  geom_density(color = "red") +  
  ggtitle("Distribuição de Fósforo (P)")
```

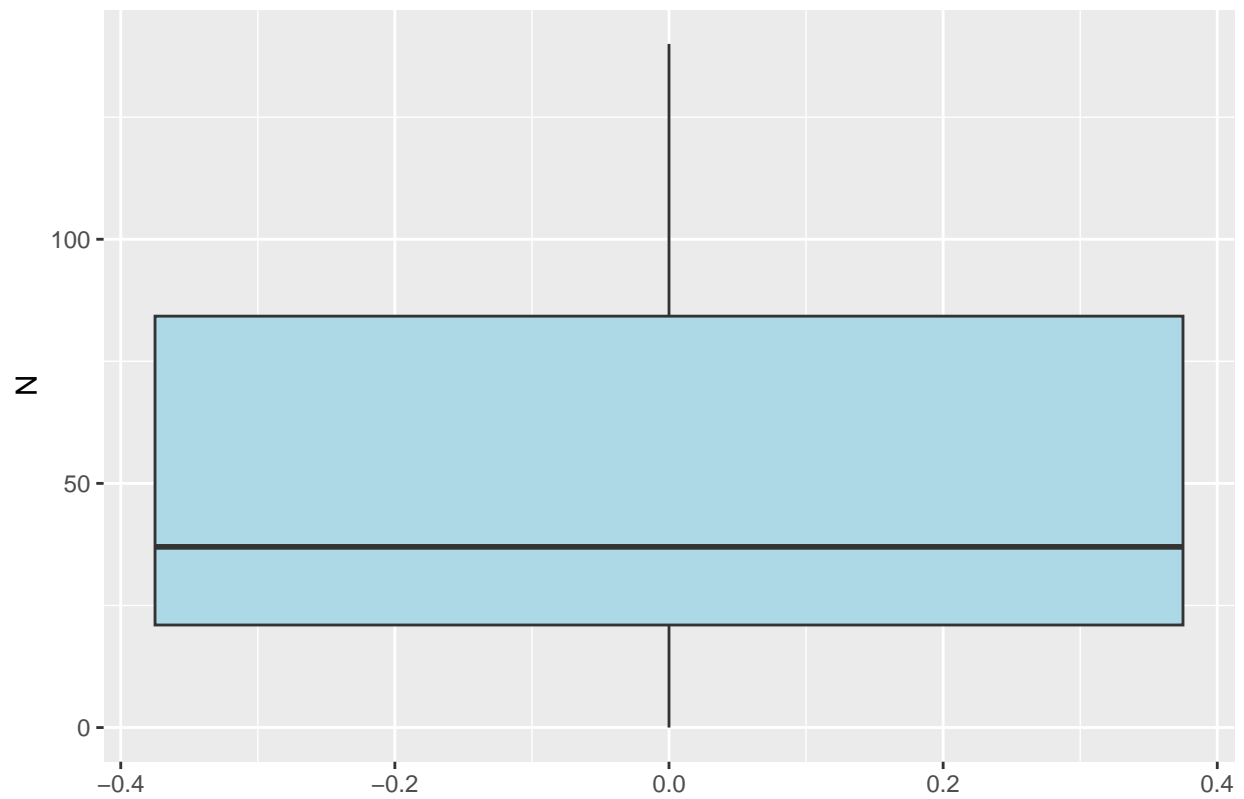
Distribuição de Fósforo (P)



Boxplot de Nitrogênio

```
ggplot(df, aes(y = N)) +  
  geom_boxplot(fill = "lightblue") +  
  ggtitle("Boxplot de Nitrogênio (N)")
```

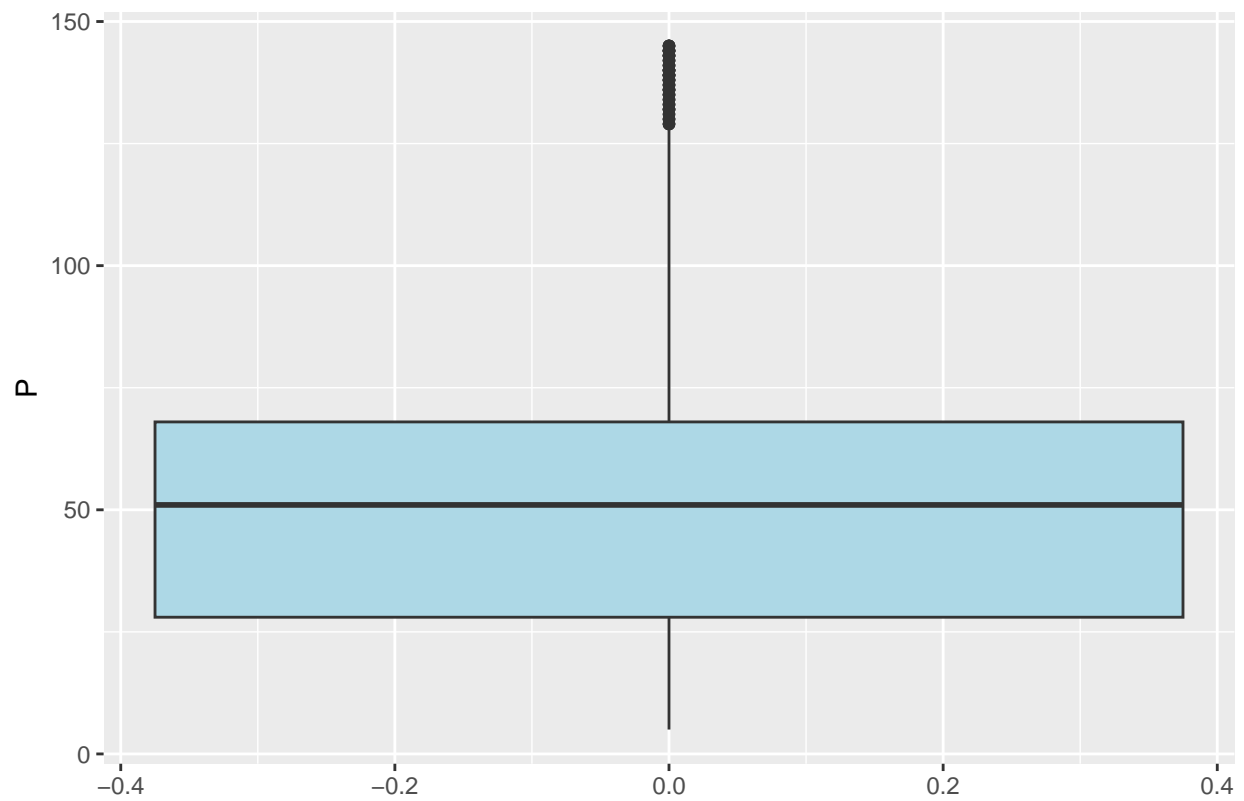
Boxplot de Nitrogênio (N)



Boxplot de Fósforo

```
ggplot(df, aes(y = P)) +  
  geom_boxplot(fill = "lightblue") +  
  ggtitle("Boxplot de Fósforo (P)")
```

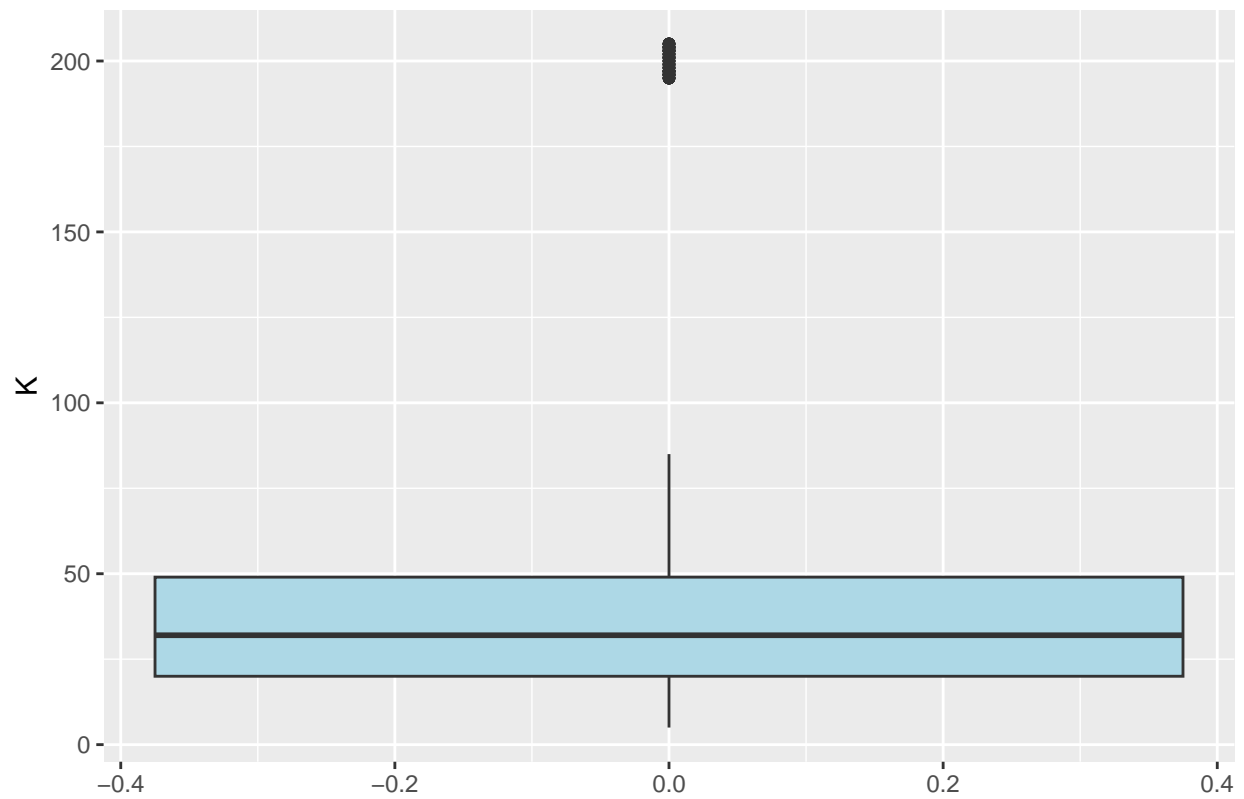
Boxplot de Fósforo (P)



Boxplot de Potássio

```
ggplot(df, aes(y = K)) +  
  geom_boxplot(fill = "lightblue") +  
  ggtitle("Boxplot de Potássio (K)")
```

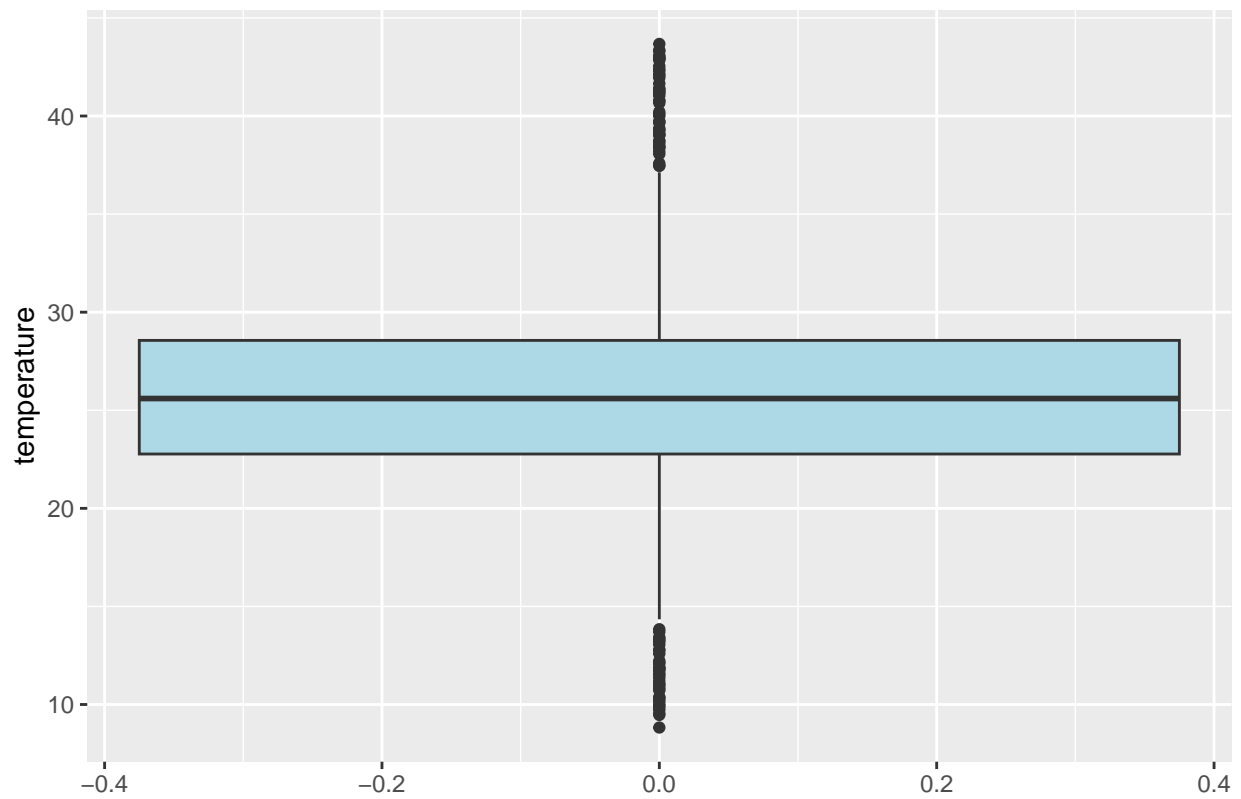
Boxplot de Potássio (K)



Boxplot de Temperatura

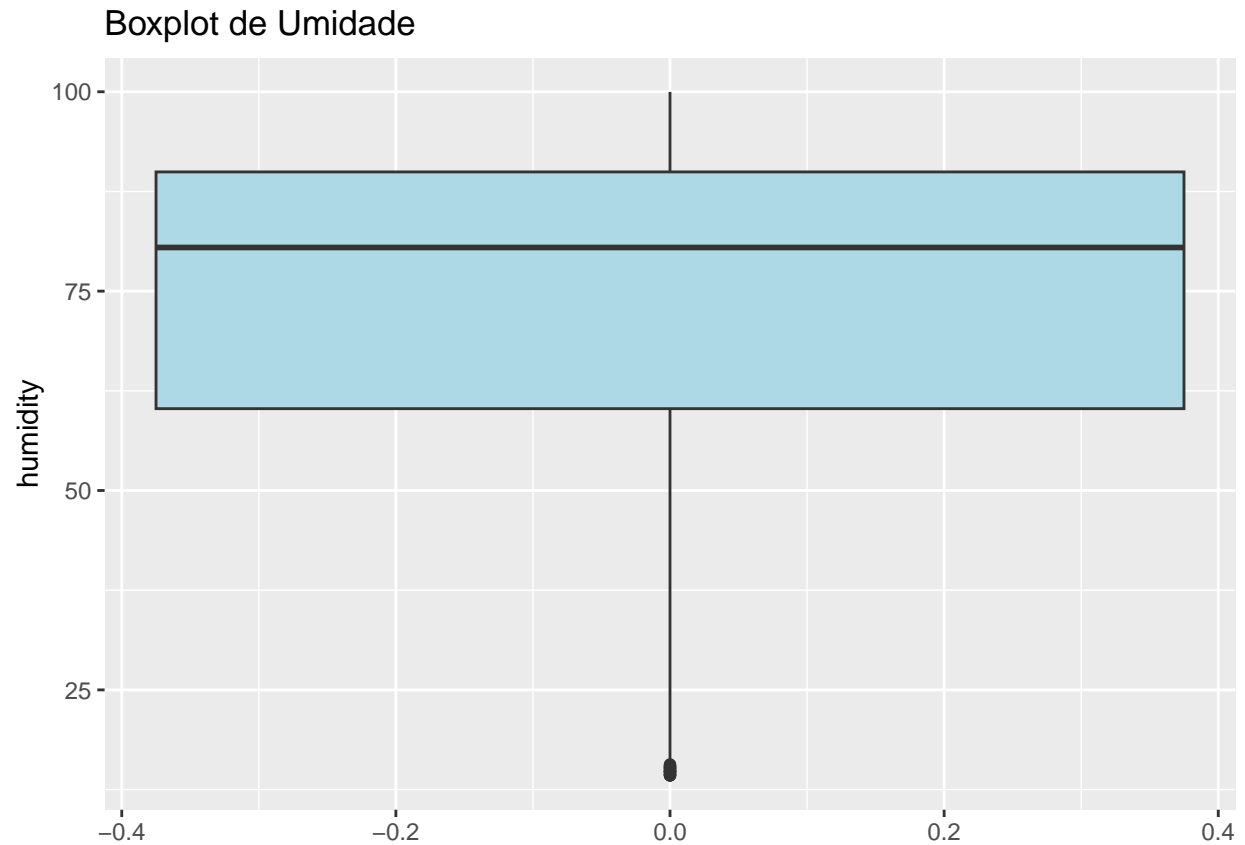
```
ggplot(df, aes(y = temperature)) +  
  geom_boxplot(fill = "lightblue") +  
  ggtitle("Boxplot de Temperatura")
```

Boxplot de Temperatura



Boxplot de Umidade

```
ggplot(df, aes(y = humidity)) +  
  geom_boxplot(fill = "lightblue") +  
  ggtitle("Boxplot de Umidade")
```

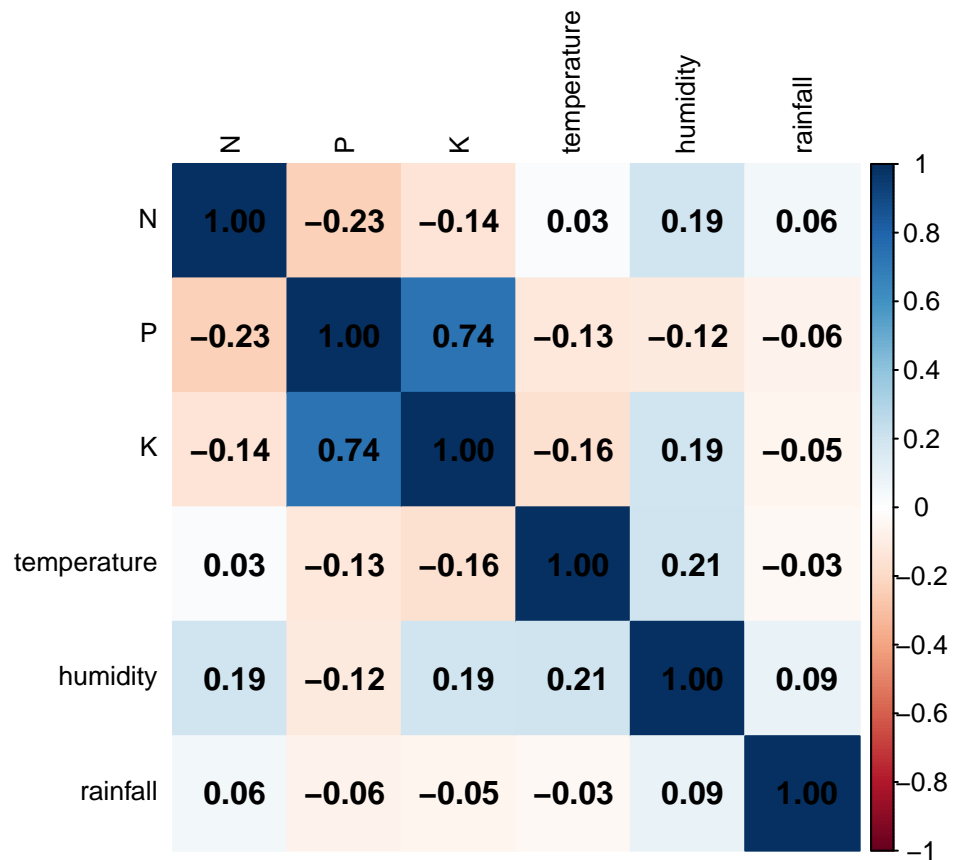


Matriz de Correlação

```
# Filtrar apenas as colunas numéricas
df_numeric <- df %>%
  select(N, P, K, temperature, humidity, rainfall)

# Criar a matriz de correlação
cor_matrix <- cor(df_numeric)

# Plotar a matriz de correlação
corrplot(cor_matrix, method = "color", addCoef.col = "black", tl.col = "black", tl.cex = 0.8)
```

Conclusão

Nesta análise, exploramos a distribuição das principais variáveis relacionadas às condições de solo e clima para diferentes culturas agrícolas. A análise da matriz de correlação nos ajudou a entender como essas variáveis se relacionam entre si. Esses insights podem servir como base para futuras análises preditivas.