

atividade ponderada

2024-08-12

1. Carregamento e Preparação dos Dados (20%)

```
# Instalar pacotes necessários
install.packages(c("tidyverse", "GGally", "corrplot", "gridExtra"))

## Installing packages into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

# Carregar pacotes
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2     3.5.1      v tibble     3.2.1
## v lubridate  1.9.3      v tidyr      1.3.1
## v purrr      1.0.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

library(corrplot)

## corrplot 0.92 loaded

library(gridExtra)

##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##   combine

library(ggplot2)
```

Carregamento dos dados (7%): O aluno deve carregar corretamente o conjunto de dados fornecido.

Visualização das primeiras linhas (7%): Utilização da função `head()` ou equivalente para mostrar as primeiras linhas do conjunto de dados.

```
head(dados)

##           cod_prod           nome_abrev
```

```
## 1 603078352615      Acne Solutions Cleansing Bar
## 2 131384024501      Pep Start Daily UV
## 3 85873708529       Deep Comfort Body Wash
## 4 579926605113      Lipikar Wash Gentle Foaming
## 5 397769055946 Anthelios AOX Daily Antioxidant
## 6 981647487600      Basil White Tea Body
##                               nome_completo
## 1      Acne Solutions Cleansing Bar for Face & Body
## 2 Pep Start Daily UV Protector Broad Spectrum SPF 50
## 3      Deep Comfort Body Wash
## 4 Lipikar Wash AP+ Gentle Foaming Moisturizing Wash
## 5 Anthelios AOX Daily Antioxidant Face Serum SPF 50
## 6      Basil & White Tea Body Wash
##
##                               descricao
## 1      Sabonete líquido Clinique Acne Solutions Cleansing Bar for Face & Body para Corpo
## 2      Protetor solar Clinique Pep Start Daily UV Protector Broad Spectrum SPF 50 para Corpo
## 3      Sabonete líquido Clinique Deep Comfort Body Wash para Corpo
## 4 Sabonete líquido La Roche-Posay Lipikar Wash AP+ Gentle Foaming Moisturizing Wash para Corpo
## 5      Protetor solar La Roche-Posay Anthelios AOX Daily Antioxidant Face Serum SPF 50 para Corpo
## 6      Sabonete líquido Nivea Basil & White Tea Body Wash para Corpo
## categoria      sub_categoria      marca      conteudo_valor      conteudo_medida
## 1      Corpo Sabonete líquido      Clinique      156      ml
## 2      Corpo Protetor solar      Clinique      30      ml
## 3      Corpo Sabonete líquido      Clinique      204      ml
## 4      Corpo Sabonete líquido La Roche-Posay      405      ml
## 5      Corpo Protetor solar La Roche-Posay      30      ml
## 6      Corpo Sabonete líquido      Nivea      600      ml
```

Verificação da estrutura dos dados (6%): Uso da função `str()` ou equivalente para verificar a estrutura dos dados.

```
# Estrutura dos dados
str(dados)
```

```
## 'data.frame':    563 obs. of  9 variables:
## $ cod_prod      : num  6.03e+11 1.31e+11 8.59e+10 5.80e+11 3.98e+11 ...
## $ nome_abrev    : chr   "Acne Solutions Cleansing Bar" "Pep Start Daily UV" "Deep Comfort Body Wash"
## $ nome_completo : chr   "Acne Solutions Cleansing Bar for Face & Body" "Pep Start Daily UV Protector"
## $ descricao     : chr   "Sabonete líquido Clinique Acne Solutions Cleansing Bar for Face & Body"
## $ categoria     : chr   "Corpo" "Corpo" "Corpo" "Corpo" ...
## $ sub_categoria : chr   "Sabonete líquido" "Protetor solar" "Sabonete líquido" "Sabonete líquido"
## $ marca         : chr   "Clinique" "Clinique" "Clinique" "La Roche-Posay" ...
## $ conteudo_valor: num    156 30 204 405 30 600 600 600 252 201 ...
## $ conteudo_medida: chr    "ml" "ml" "ml" "ml" ...
```

2. Resumo Estatístico e Descrição dos Dados (20%)

Resumo estatístico (10%): Utilização da função `summary()` ou equivalente para obter um resumo estatístico das variáveis.

```
# Resumo estatístico das variáveis
summary(dados)
```

```
##      cod_prod      nome_abrev      nome_completo      descricao
## Min.   :6.587e+09 Length:563      Length:563      Length:563
## 1st Qu.:2.506e+11 Class :character Class :character Class :character
## Median :4.800e+11 Mode  :character Mode  :character Mode  :character
```

```
## Mean      :5.024e+11
## 3rd Qu.   :7.696e+11
## Max.      :9.970e+11
## categoria      sub_categoria      marca      conteudo_valor
## Length:563      Length:563      Length:563      Min.       : 0.02
## Class :character Class :character Class :character 1st Qu.    : 3.50
## Mode  :character Mode  :character Mode  :character Median      : 15.00
##                                     Mean       : 69.28
##                                     3rd Qu.    : 61.50
##                                     Max.       :1440.00
## conteudo_medida
## Length:563
## Class :character
## Mode  :character
##
##
##
```

Descrição das variáveis (10%): Descrição clara e concisa das variáveis presentes no conjunto de dados.

cod_prod:

Descrição: Código identificador único de cada produto. Tipo: Numérico (inteiro) Função: Serve para distinguir um produto do outro e pode ser utilizado como chave primária em um banco de dados. nome_abrev:

Descrição: Nome abreviado do produto. Tipo: Caractere Função: Nome curto e fácil de lembrar do produto, utilizado para referenciar o produto de forma concisa. nome_completo:

Descrição: Nome completo do produto. Tipo: Caractere Função: Descrição detalhada do produto, contendo informações sobre sua função e características. descricao:

Descrição: Descrição adicional do produto. Tipo: Caractere Função: Fornece informações complementares sobre o produto, como textura, fragrância, etc. categoria:

Descrição: Categoria a qual o produto pertence. Tipo: Caractere Função: Agrupa produtos com características semelhantes (ex: produtos para pele, para cabelo, etc.). sub_categoria:

Descrição: Subcategoria do produto. Tipo: Caractere Função: Divide a categoria em grupos mais específicos (ex: sabonete líquido, protetor solar, etc.). marca:

Descrição: Marca do produto. Tipo: Caractere Função: Indica o fabricante do produto. conteudo_valor:

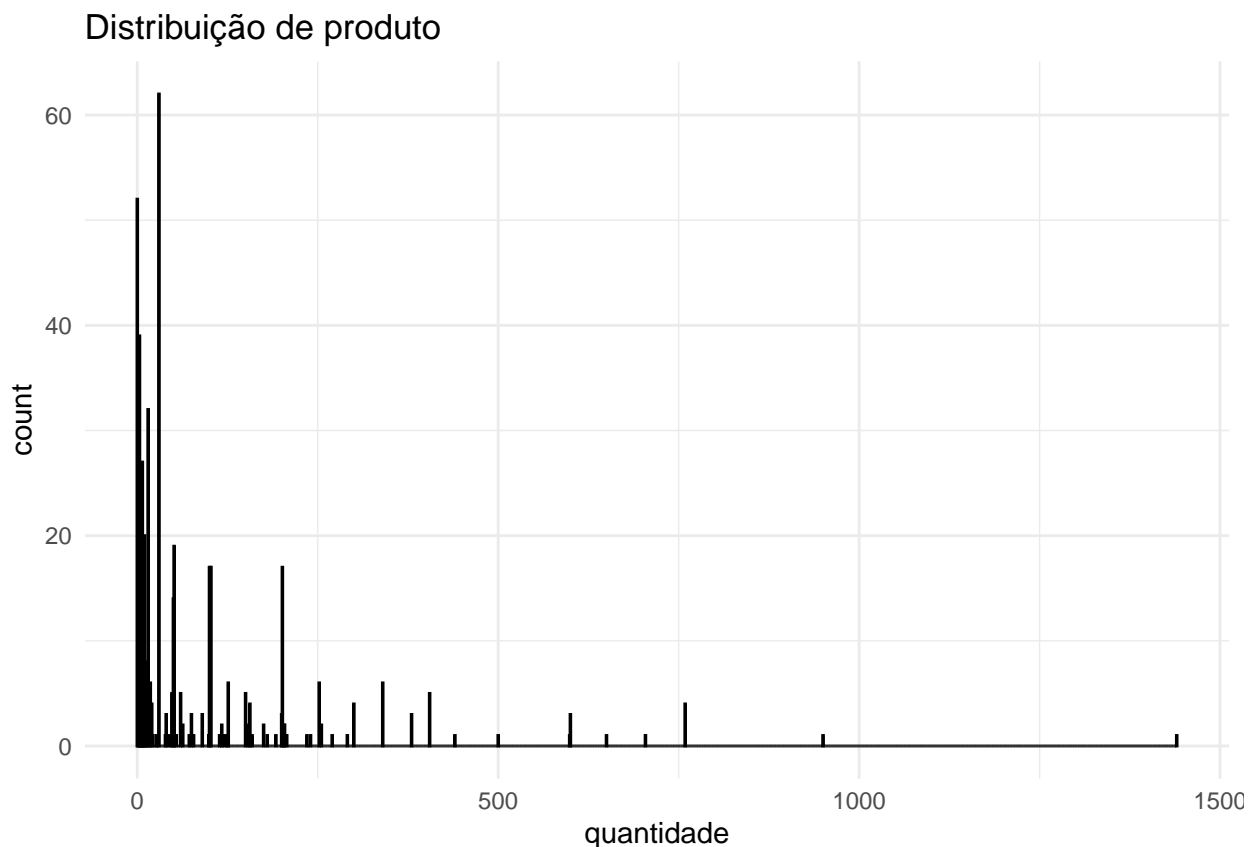
Descrição: Quantidade do produto. Tipo: Numérico Função: Indica a quantidade do produto em uma unidade específica (ex: mililitros). conteudo_medida:

Descrição: Unidade de medida da quantidade do produto. Tipo: Caractere Função: Especifica a unidade utilizada para medir a quantidade do produto (ex: ml, g, etc.).

3. Análise Univariada (25%)

Visualização das distribuições (13%): Criação de histogramas, gráficos de densidade ou boxplots para as variáveis principais.

```
# Histograma da variável 'quality'
ggplot(dados, aes(x = conteudo_valor)) +
  geom_histogram(binwidth = 1, fill = "blue", color = "black") +
  theme_minimal() +
  labs(title = "Distribuição de produto", x = "quantidade")
```



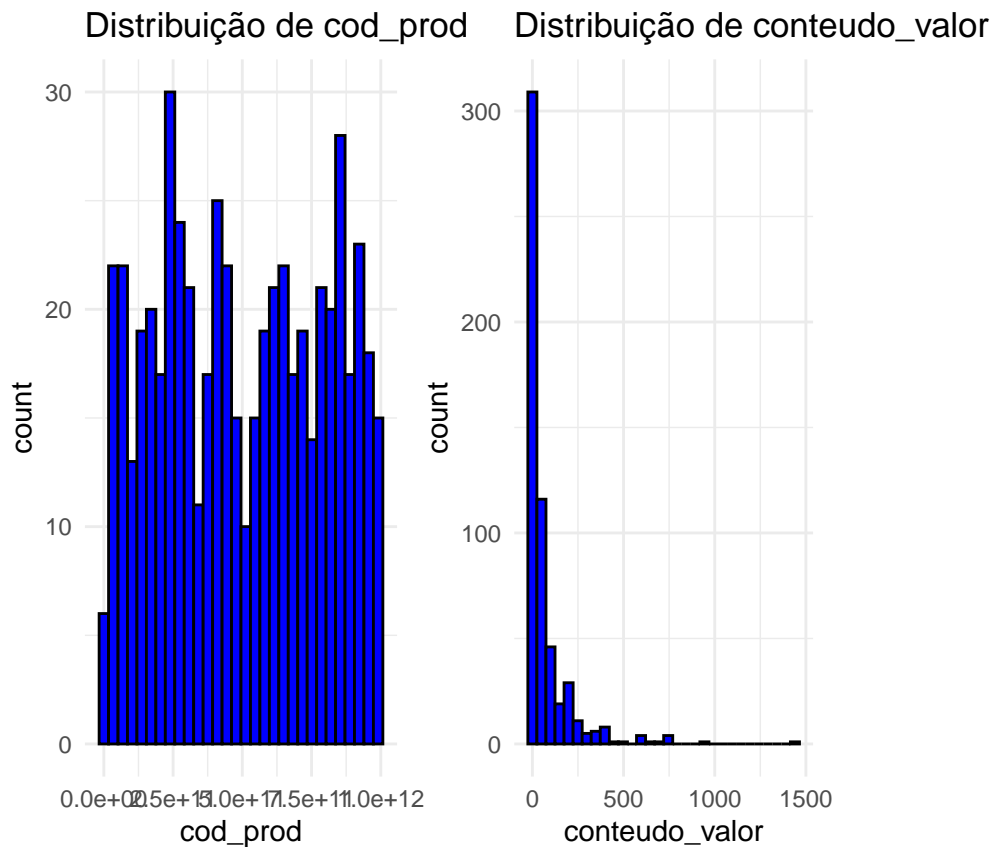
Identificação de outliers (12%): Identificação e discussão sobre outliers presentes nas variáveis.

```
# Seleciona apenas as variáveis numéricas
numeric_vars <- dados %>% select_if(is.numeric)

# Cria os histogramas para cada variável numérica
p <- lapply(names(numeric_vars), function(var) {
  ggplot(dados, aes_string(x = var)) +
    geom_histogram(fill = "blue", color = "black", bins = 30) +
    theme_minimal() +
    labs(title = paste("Distribuição de", var), x = var, )
})
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

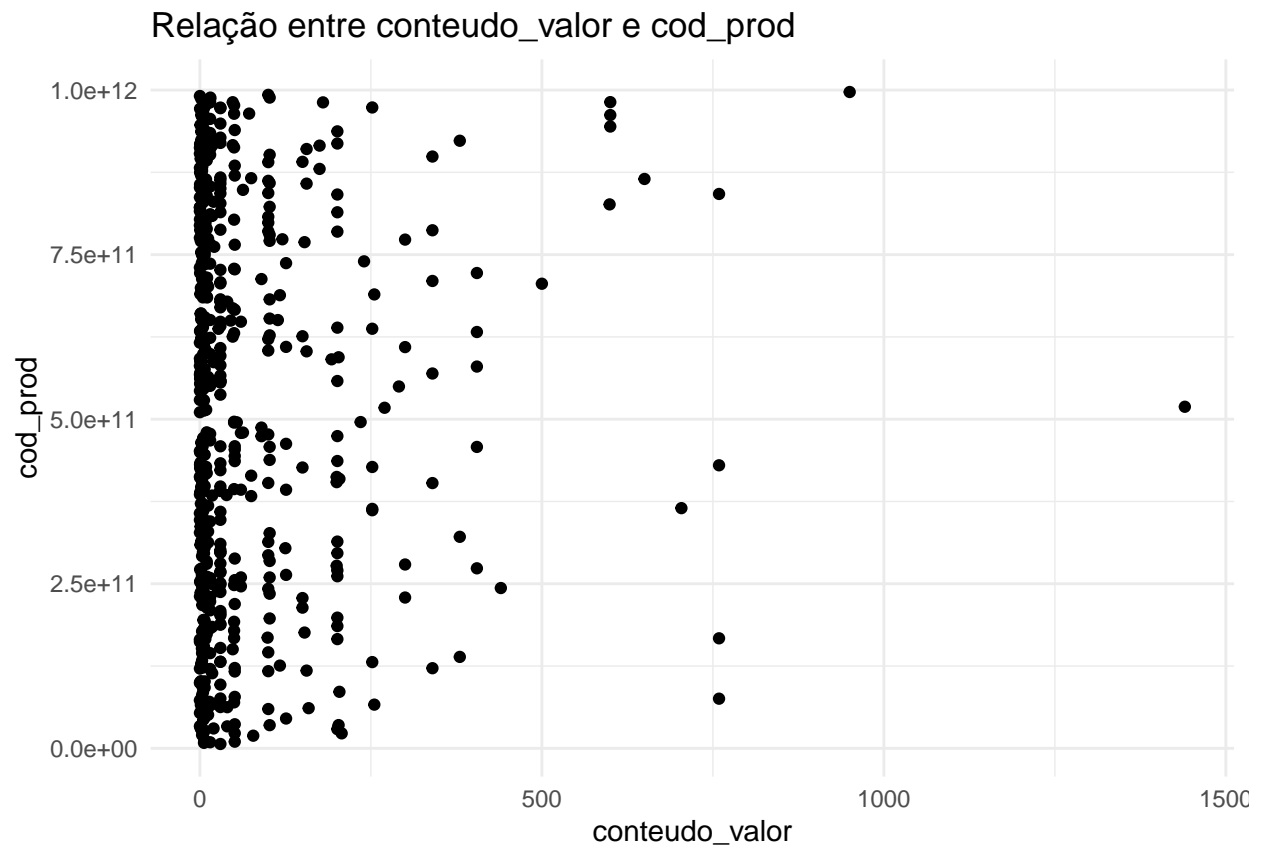
```
# Exibe os gráficos em um layout com 3 colunas
grid.arrange(grobs = p, ncol = 3)
```



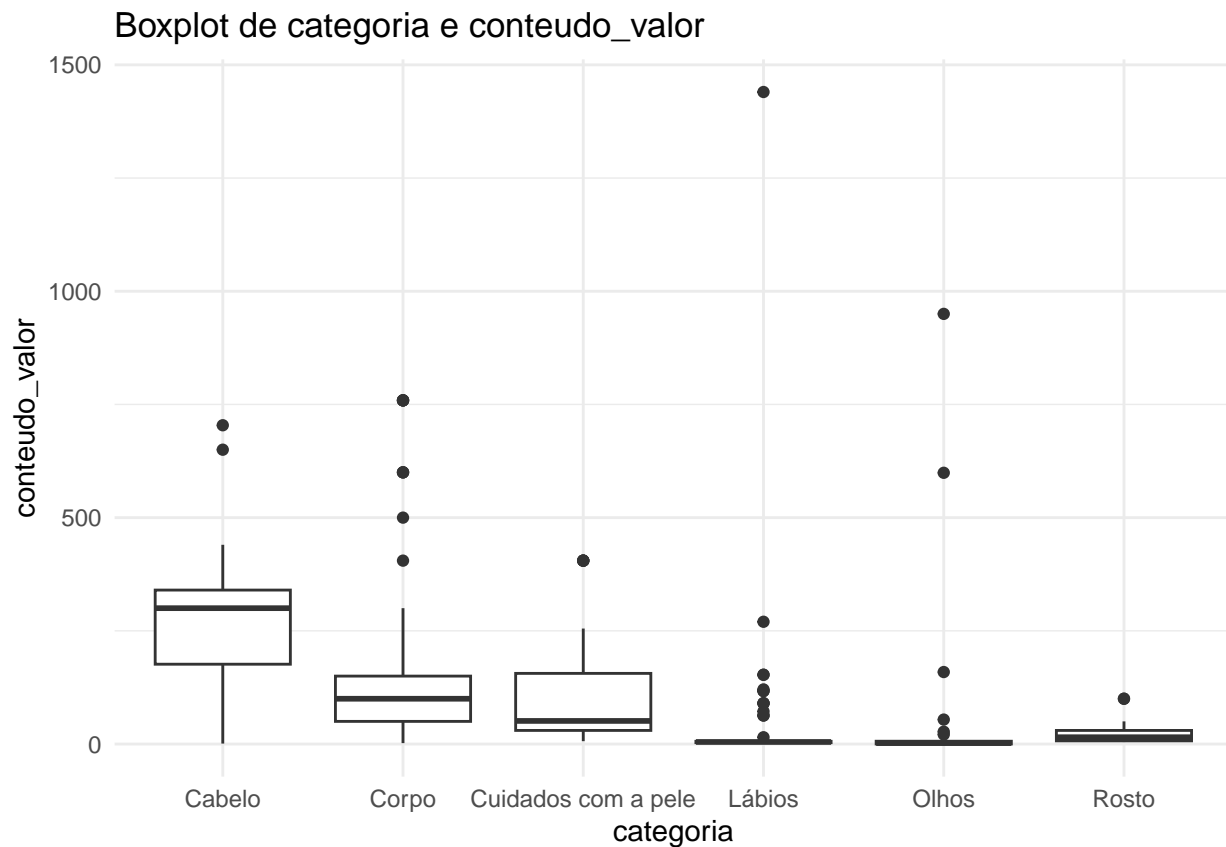
4. Análise Bivariada (25%)

Visualização de relações entre variáveis (13%): Criação de gráficos de dispersão (scatter plots) ou gráficos de barras para explorar relações entre variáveis.

```
# Scatter plot para duas variáveis numéricas
ggplot(dados, aes(x = conteudo_valor, y = cod_prod)) +
  geom_point() +
  theme_minimal() +
  labs(title = "Relação entre conteudo_valor e cod_prod", x = "conteudo_valor", y = "cod_prod")
```

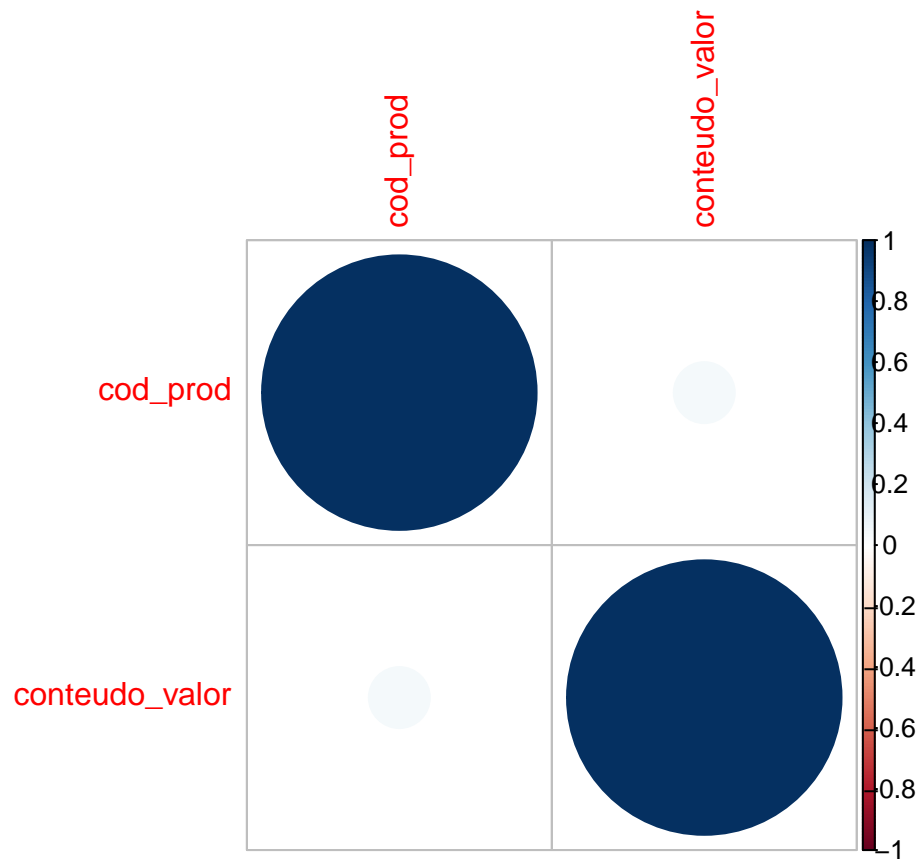


```
# Boxplot para uma variável categórica e uma numérica  
ggplot(dados, aes(x = categoria, y = conteudo_valor)) +  
  geom_boxplot() +  
  theme_minimal() +  
  labs(title = "Boxplot de categoria e conteudo_valor", x = "categoria", y = "conteudo_valor")
```

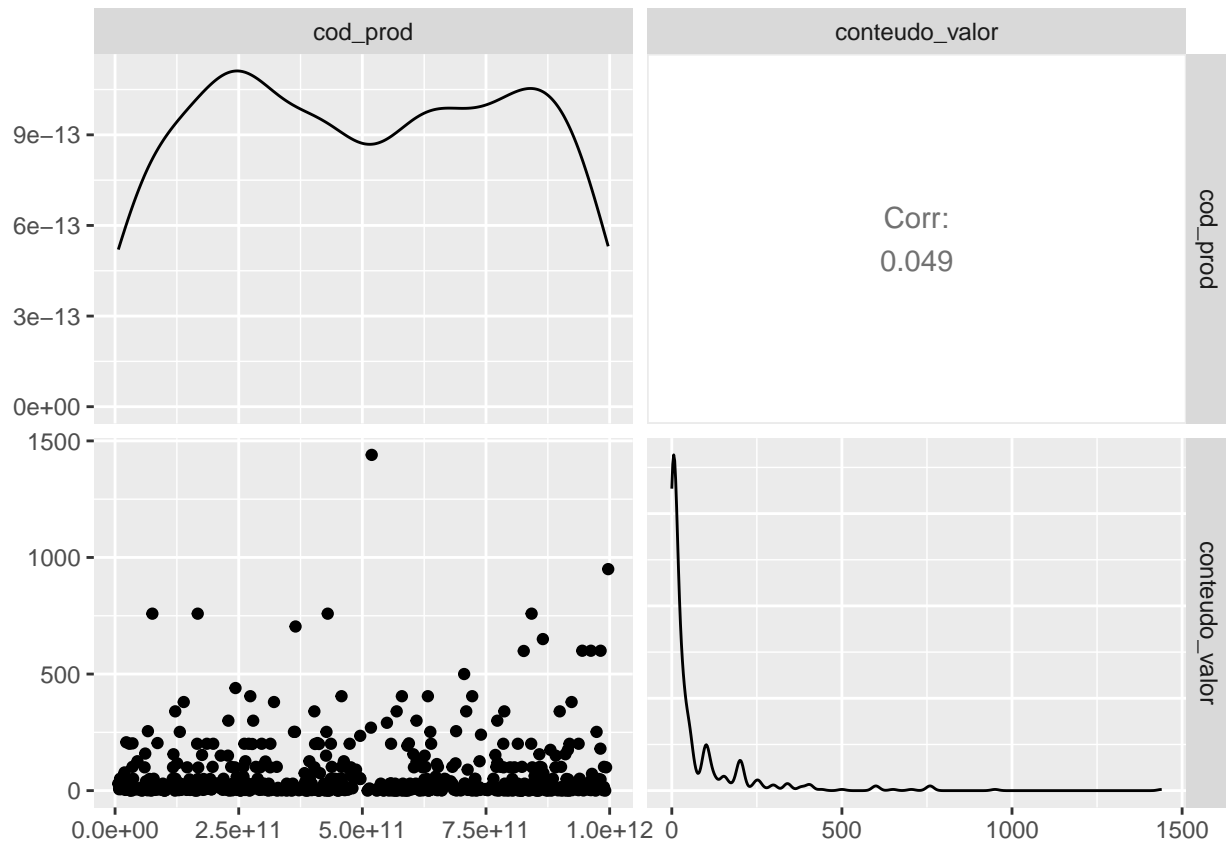


Análise de correlação (12%): Cálculo e visualização da matriz de correlação entre as variáveis, usando `cor()` e `corrplot()`.

```
# Matriz de correlação
cor_matrix <- cor(dados %>% select_if(is.numeric))
corrplot(cor_matrix, method = "circle")
```



```
# Scatterplot matrix  
ggpairs(dados %>% select_if(is.numeric))
```

5. Análise Multivariada (15%)

Análise de Componentes Principais (PCA) (10%): Implementação de PCA utilizando FactoMineR ou prcomp(), e visualização dos resultados.

```
library(ggfortify)

# Regressão linear simples
model <- lm(cod_prod ~ conteudo_valor, data = dados)
summary(model)

##
## Call:
## lm(formula = cod_prod ~ conteudo_valor, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.990e+11 -2.487e+11 -2.198e+10  2.654e+11  4.958e+11
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.951e+11  1.376e+10   35.98  <2e-16 ***
## conteudo_valor 1.044e+08  8.926e+07    1.17   0.242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.917e+11 on 561 degrees of freedom
## Multiple R-squared:  0.002435,    Adjusted R-squared:  0.0006566
```

```
## F-statistic: 1.369 on 1 and 561 DF, p-value: 0.2424
```

```
# Análise de componentes principais (PCA)
```

```
pca <- prcomp(dados %>% select_if(is.numeric), scale = TRUE)
```

```
summary(pca)
```

```
## Importance of components:
```

```
##          PC1      PC2
```

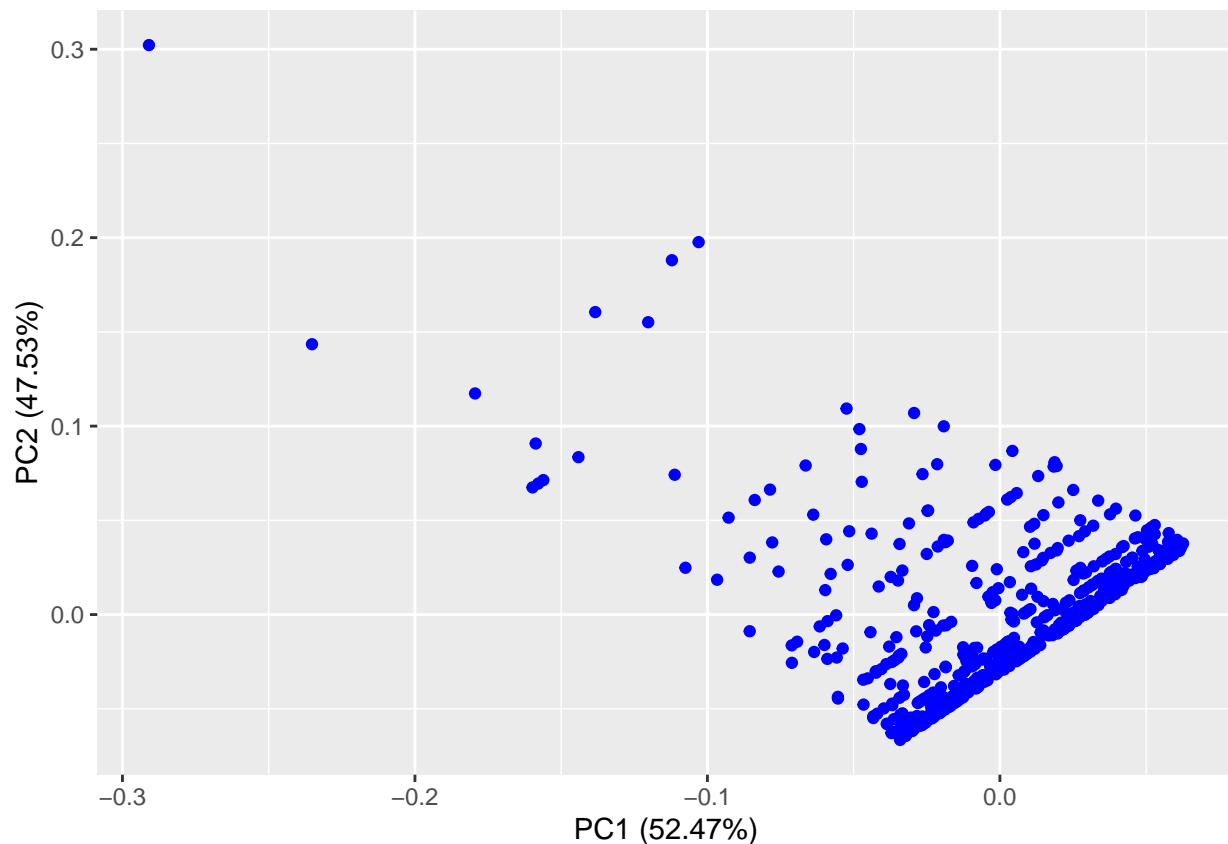
```
## Standard deviation    1.0244 0.9750
```

```
## Proportion of Variance 0.5247 0.4753
```

```
## Cumulative Proportion 0.5247 1.0000
```

```
# Visualização dos componentes principais
```

```
autoplot(pca, data = dados, colour = 'blue')
```



Interpretação dos componentes (5%): Interpretação e discussão dos componentes principais e sua variância explicada.

6. Conclusão e Discussão (15%)

Sumário das descobertas (8%): Resumo das principais descobertas da análise.

Discussão sobre limitações e possíveis melhorias (7%): Discussão sobre as limitações do estudo e sugestões para análises futuras.