

Relatório da análise dos dados - Computacional I

Thierry Martins Ribeiro

15/07/2025

Introdução

As promoções com desconto têm sido observadas como uma forma eficaz de impulsionar as vendas e atrair consumidores na maioria dos setores. Jha et al. (2019) afirmaram que os descontos oferecidos pelas empresas geram um excelente aumento de curto prazo no volume de vendas, além de incentivar a experimentação do produto por novos clientes. Além disso, Guha et al. (2018) também observaram que uma política de descontos bem planejada não apenas aumenta as vendas no presente, mas também pode ajudar a elevar a participação de mercado ao oferecer o produto a um preço mais baixo e torná-lo mais competitivo.

Além disso, Bandyopadhyay et al. (2021) analisaram os efeitos dos descontos e concluíram que, em mercados extremamente competitivos, os descontos ajudam a distinguir a empresa na mente dos consumidores, funcionando como um “ímã” para atrair novos clientes e afastar consumidores que poderiam abandonar a marca. Jee (2021) confirmou que os programas de desconto influenciam positivamente o comportamento do consumidor, pois os clientes sentem que há mais valor nos produtos quando recebem um benefício monetário inicial, por exemplo, um preço reduzido.

Essas conclusões também se refletem no comportamento cotidiano, como promoções do tipo “leve 2 e pague 1”, descontos escalonados para compras em maior quantidade ou promoções-relâmpago em um determinado dia da semana. Grandes varejistas, como as Lojas Americanas, fazem amplo uso dessas estratégias, oferecendo preços promocionais semanais em categorias específicas de produtos e promoções do tipo “compre 3 e pague 2” para aumentar o volume de vendas. Da mesma forma, os supermercados também realizam promoções e feirões nos finais de semana para atrair consumidores e girar o estoque, sendo exemplos claros de como as políticas de desconto exercem uma influência favorável sobre o comportamento de compra.

Isso é confirmado pelo estudo de Rifqah Harahap e Anita Situmorang (2023) em uma empresa de equipamentos médicos sediada em Medan. Os autores observaram que promoções e descontos tiveram, cada um, um efeito positivo e significativo sobre as vendas, tanto de forma independente quanto combinada, explicando mais de 90% da variação nas vendas observadas durante o período do estudo.

Por fim, destaca-se a pesquisa realizada na Pelican Store, uma rede de lojas de roupas femininas nos Estados Unidos, que implementou uma campanha de cupons direcionada a clientes de outras marcas dentro do mesmo grupo corporativo. A análise dos dados da campanha revelou altas taxas de crescimento nas vendas, indicando que os programas de desconto podem ser especialmente bem-sucedidos quando acompanhados de uma comunicação precisa com o público-alvo relevante.

Bibliografia e link para os artigos citados

- https://www.researchgate.net/publication/371097465_Influence_of_Price_Promotion_and_Discounts_on_Sales_Study_at_A_Medical_Device_Company_in_Medan_City
- <https://journal.widyamangala.ac.id/index.php/jurnalaset/article/download/170/141/479>
- https://issuu.com/cengagebrasil/docs/cap_tulo_amostra_estatistica_aplicada_a_administra

Minha análise de dados feita configura em cima da disciplina Estatística I e Computacional I, onde foram utilizados os conceitos de estatística descritiva. Para ter algo mais bruto procurei saber superficialmente sobre a correlação entre as variáveis e como elas se relacionam, além de criar um modelo de regressão linear simples.

De início dividi em 2 arquivos principais, primeiro o `dados.r` que é responsável por carregar os dados e exibi-los, e o segundo `analise.r` que contém as análises estatísticas.

=====

Primeiro de tudo separei as variáveis que estão no dataset.

```
#=====
# Dicionário das Variáveis                                     /
#=====
# Customer           = Cliente                               (Quantitativa Discreta) /
# Type of customer   = Tipo de cliente                       (Qualitativa Nominal)  /
# Items              = Itens                                  (Quantitativa Discreta) /
# Net sales           = Vendas líquidas                      (Quantitativa Contínua) /
# Method of payment   = Método de pagamento                 (Qualitativa Nominal)  /
# Gender              = Gênero                                (Qualitativa Nominal)  /
# Marital status      = Estado civil                         (Qualitativa Nominal)  /
# Age                 = Idade                                 (Quantitativa Contínua) /
#=====
```

Carregamento dos dados no arquivo `dados.r`.

```
carregar_dados <- function() {
  dados_trabalho <- read.csv("C:/Users/Thierry/Documents/Computacional I/Trabalho/PelicanStores - Data.csv")
  return(dados_trabalho)
}
dados_trabalho <- carregar_dados()
View(dados_trabalho)
print(dados_trabalho)
```

Depois de carregar os dados, criei um arquivo `analise.r` que contém as análises estatísticas.

Carregamento dos dados no arquivo `analise.r` e as bibliotecas que utilizei.

```
source("C:/Users/Thierry/Documents/Computacional I/Trabalho/dados.r")
dados_trabalho <- carregar_dados()

library(ggplot2)
library(dplyr)
```

Fiz a inspeção dos dados para verificar se estavam corretos.

```
#=====
# Inspeccionar dados |
#=====
str(dados_trabalho) #
summary(dados_trabalho) #
head(dados_trabalho) #
tail(dados_trabalho) #
#=====
```

Em seguida fiz uma análise descritiva da variável Idade para entender melhor a distribuição dos dados. Comecei calculando a quantidade de observações, depois apliquei a regra de sturges para determinar o número de classes e dividi o intervalo de valores em classes. Com isso eu criei uma tabela de frequências para a variável Idade e gerei um histograma para visualizar a distribuição dos dados.

```
#=====
# Dados gerais |
#=====
qtd_observacoes <- nrow(dados_trabalho)
print(qtd_observacoes)

# Regra de Sturges para número de classes
k <- 1 + 3.322 * log10(qtd_observacoes)
arredonda <- ceiling(k)

#=====
# Analise da varivel idade |
#=====
idade_minima <- min(dados_trabalho$Age)
idade_maxima <- max(dados_trabalho$Age)

amplitude_total <- idade_maxima - idade_minima
amplitude_classe <- ceiling(amplitude_total / arredonda)

# Definir classes
classes_idade <- seq(idade_minima, idade_maxima + arredonda, by = amplitude_classe)
idade_classes <- cut(dados_trabalho$Age, breaks = classes_idade, right = FALSE)

# Tabela de frequência para idade
tabela_frequencia_idade <- data.frame(
  Classe = levels(idade_classes),
  Frequencia_Absoluta = as.numeric(table(idade_classes)),
  Frequencia_Relativa = as.numeric(prop.table(table(idade_classes))),
  Frequencia_Relativa_Porcentagem = as.numeric(prop.table(table(idade_classes))) * 100,
  Frequencia_Relativa_Acumulada = cumsum(as.numeric(prop.table(table(idade_classes))))
)
```

```
total_frequencia <- data.frame(
  Classe = "Total",
  Frequencia_Absoluta = sum(tabela_frequencia_idade$Frequencia_Absoluta),
  Frequencia_Relativa = 1,
  Frequencia_Relativa_Porcentagem = 100,
  Frequencia_Relativa_Acumulada = NA
)

tabela_idade <- rbind(tabela_frequencia_idade, total_frequencia)
```

Histograma gerado para a variável Idade.

```
source("analise.r")

#===== analise da variavel age =====#
hist(
  dados_trabalho$Age,
  main = "Idades dos clientes",
  xlab = "Idade",
  ylab = "Frequências",
  breaks = classes_idade,
  col = "violet"
)
```

Em seguida fiz uma análise descritiva das frequências das variáveis qualitativas Tipo de cliente, Método de pagamento, Gênero e Estado civil. Para não repetir de criação de tabelas, criei uma função que recebe o nome da variável e gera a tabela de frequências.

```
#=====
# Análise de Frequências e Estatísticas Descritivas /
#=====

criar_tabela_frequencia <- function(dados, variavel) {
  abs <- table(dados[[variavel]])
  rel <- prop.table(abs)
  rel_perc <- rel * 100
  data.frame(
    Categoria = names(abs),
    Frequencia_Absoluta = as.numeric(abs),
    Frequencia_Relativa = round(as.numeric(rel), 4),
    Frequencia_Relativa_Porcentagem = round(as.numeric(rel_perc), 2)
  )
}

adicionar_total <- function(tabela) {
  total <- data.frame(
    Categoria = "Total",
    Frequencia_Absoluta = sum(tabela$Frequencia_Absoluta),
```

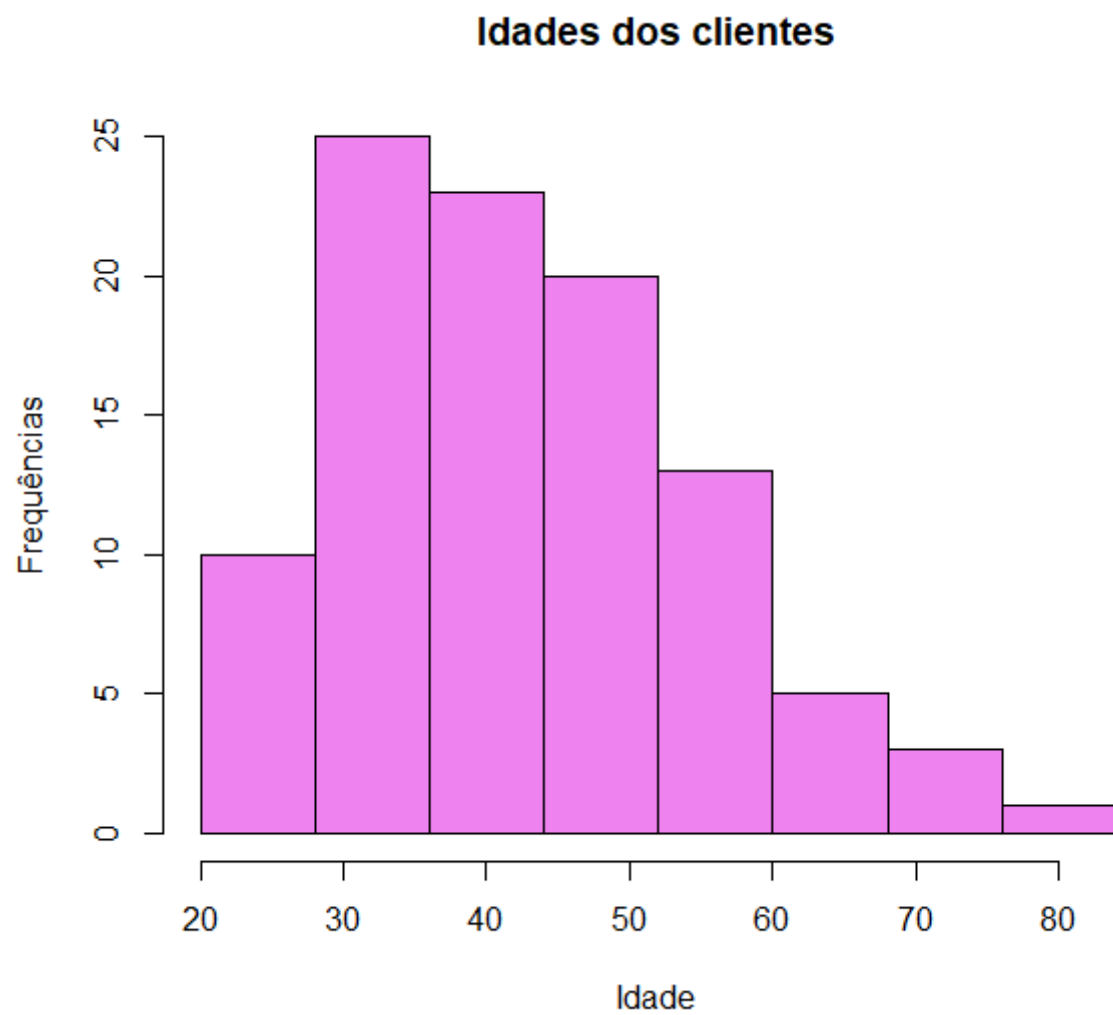


Figure 1: Alt text

```

    Frecuencia_Relativa = sum(tabela$Frecuencia_Relativa),
    Frecuencia_Relativa_Porcentagem = 100
  )
  rbind(tabela, total)
}

# Type of Customer
tabela_customer <- criar_tabela_frecuencia(dados_trabalho, "Type.of.Customer")
tabela_customer <- adicionar_total(tabela_customer)

# Method of Payment
tabela_payment <- criar_tabela_frecuencia(dados_trabalho, "Method.of.Payment")
tabela_payment <- adicionar_total(tabela_payment)

# Gender
tabela_gender <- criar_tabela_frecuencia(dados_trabalho, "Gender")
tabela_gender <- adicionar_total(tabela_gender)

# Marital Status
tabela_marital <- criar_tabela_frecuencia(dados_trabalho, "Marital.Status")
tabela_marital <- adicionar_total(tabela_marital)

#===== tabela frecuencia por cliente =====#

tabela_cliente_plot <- tabela_customer[tabela_customer$Categoria != "Total", ]
barplot(
  tabela_cliente_plot$Frecuencia_Absoluta,
  names.arg = tabela_cliente_plot$Categoria,
  main = "Frequência por Tipo de Cliente",
  col = "purple",
  las = 2,
  ylab = "Frequência",
  xlab = "Tipo de Cliente"
)

#===== tabela frecuencia por metodo de pagamento =====#

barplot(
  tabela_payment$Frecuencia_Absoluta[-nrow(tabela_payment)],
  names.arg = tabela_payment$Categoria[-nrow(tabela_payment)],
  main = "Método de Pagamento",
  col = "red",
  las = 2,
  ylab = "Frequência",
  xlab = "Método"
)

#===== tabela frecuencia por genero =====#

tabela_gender_plot <- tabela_gender[tabela_gender$Categoria != "Total", ]
barplot(
  tabela_gender_plot$Frecuencia_Absoluta,
  names.arg = tabela_gender_plot$Categoria,
  main = "Frequência por Gênero",
  col = "blue",

```

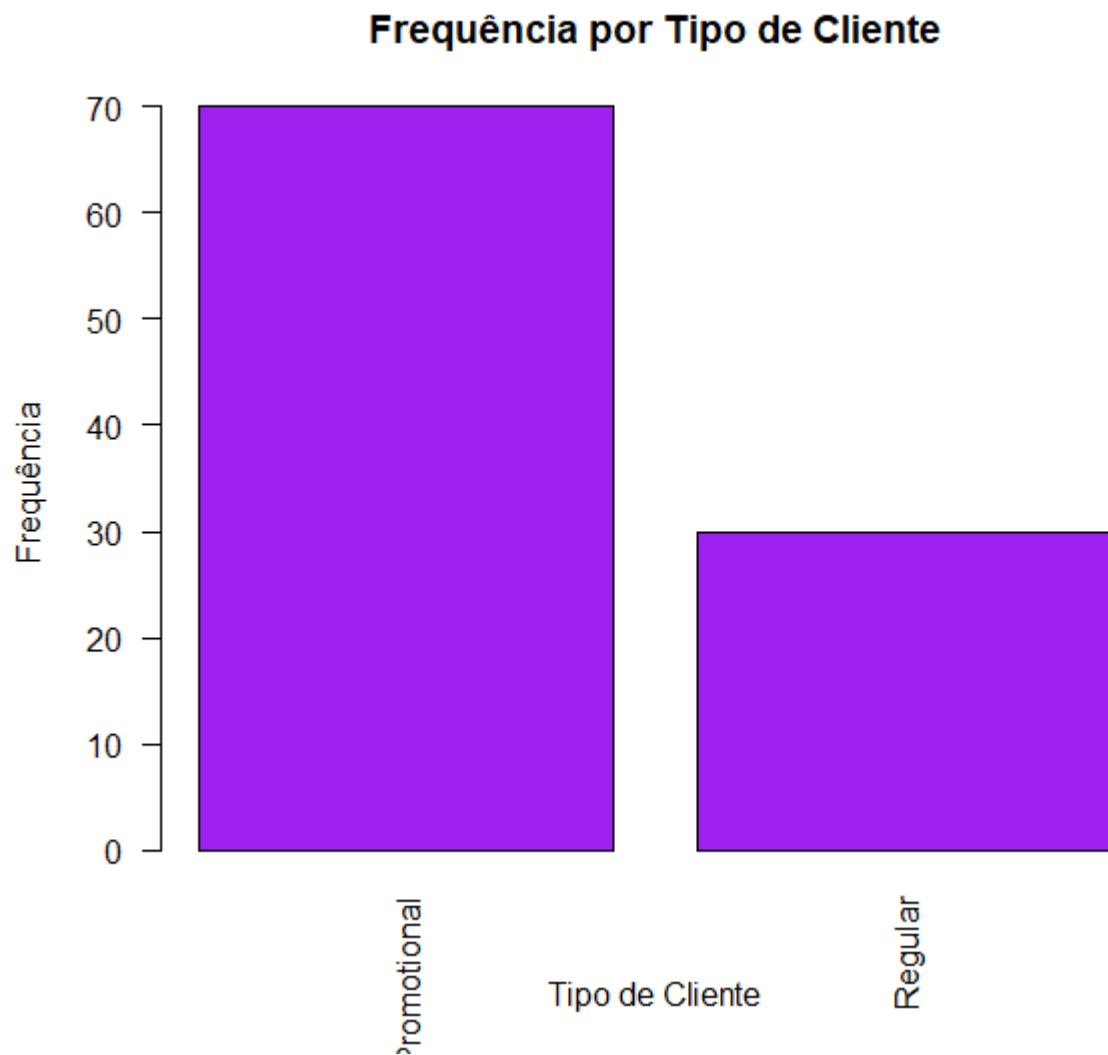
```

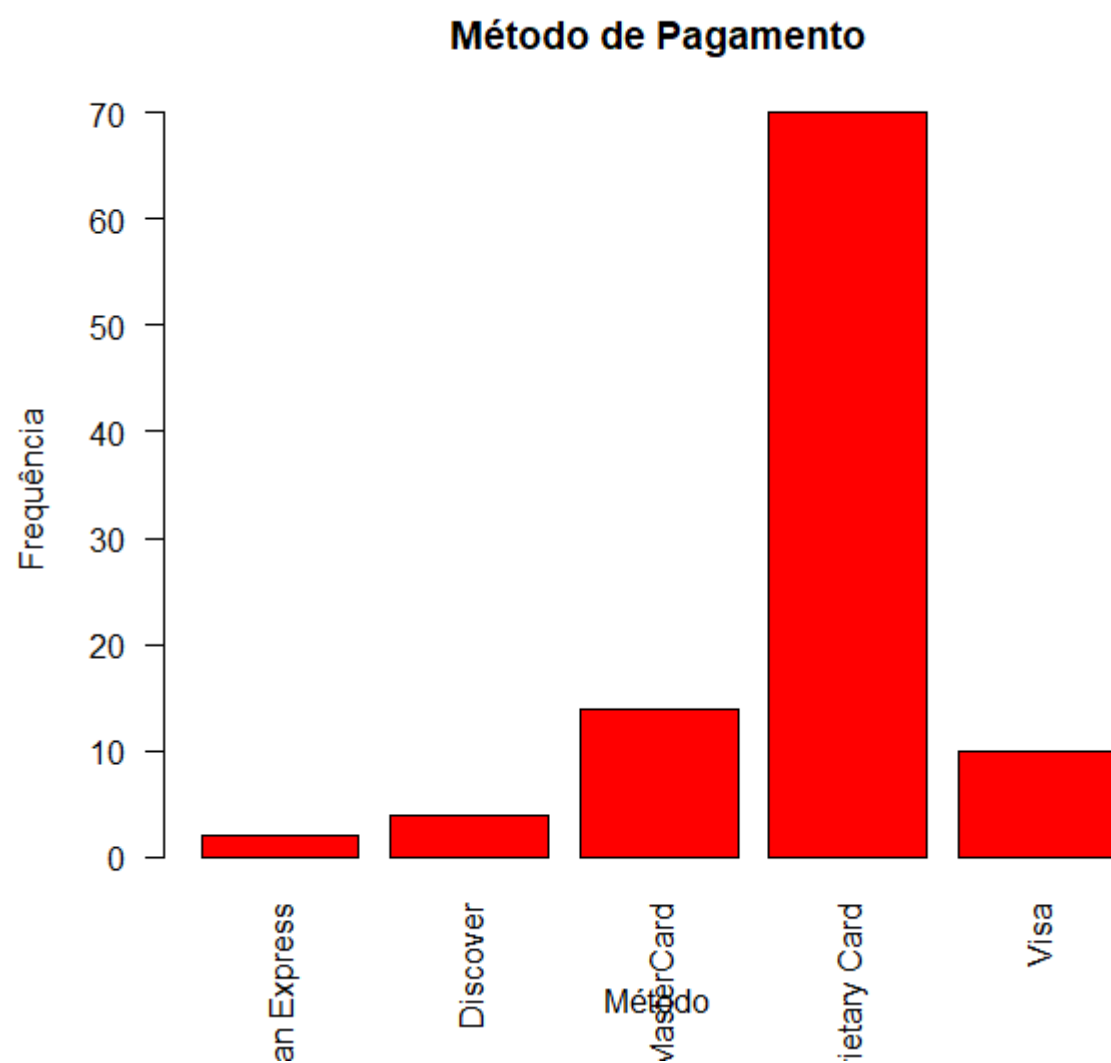
    las = 2,
    ylab = "Frequência",
    xlab = "Gênero"
)

##### tabela frequencia por estado civil #####

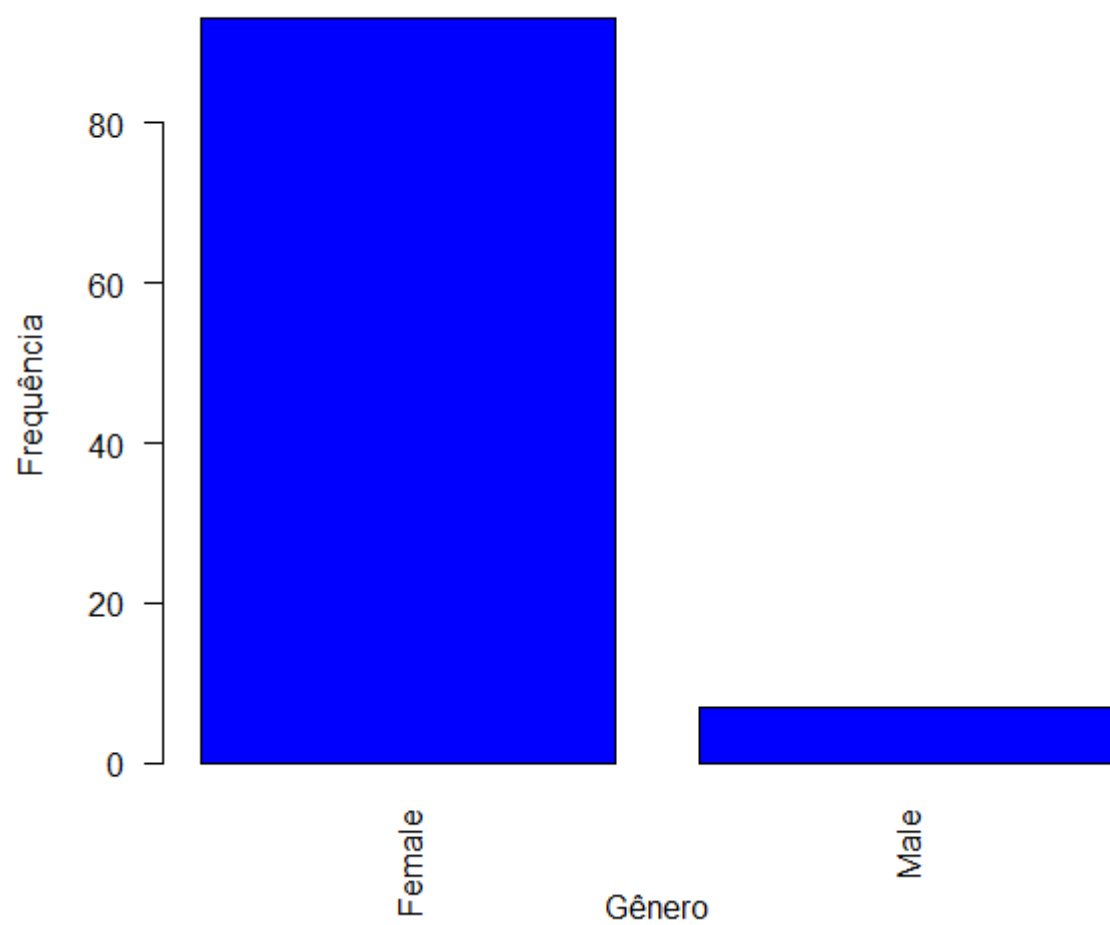
barplot(
  tabela_marital$Frequencia_Absoluta[-nrow(tabela_marital)],
  names.arg = tabela_marital$Categoria[-nrow(tabela_marital)],
  main = "Estado Civil",
  col = "green",
  las = 2,
  ylab = "Frequência",
  xlab = "Estado Civil"
)

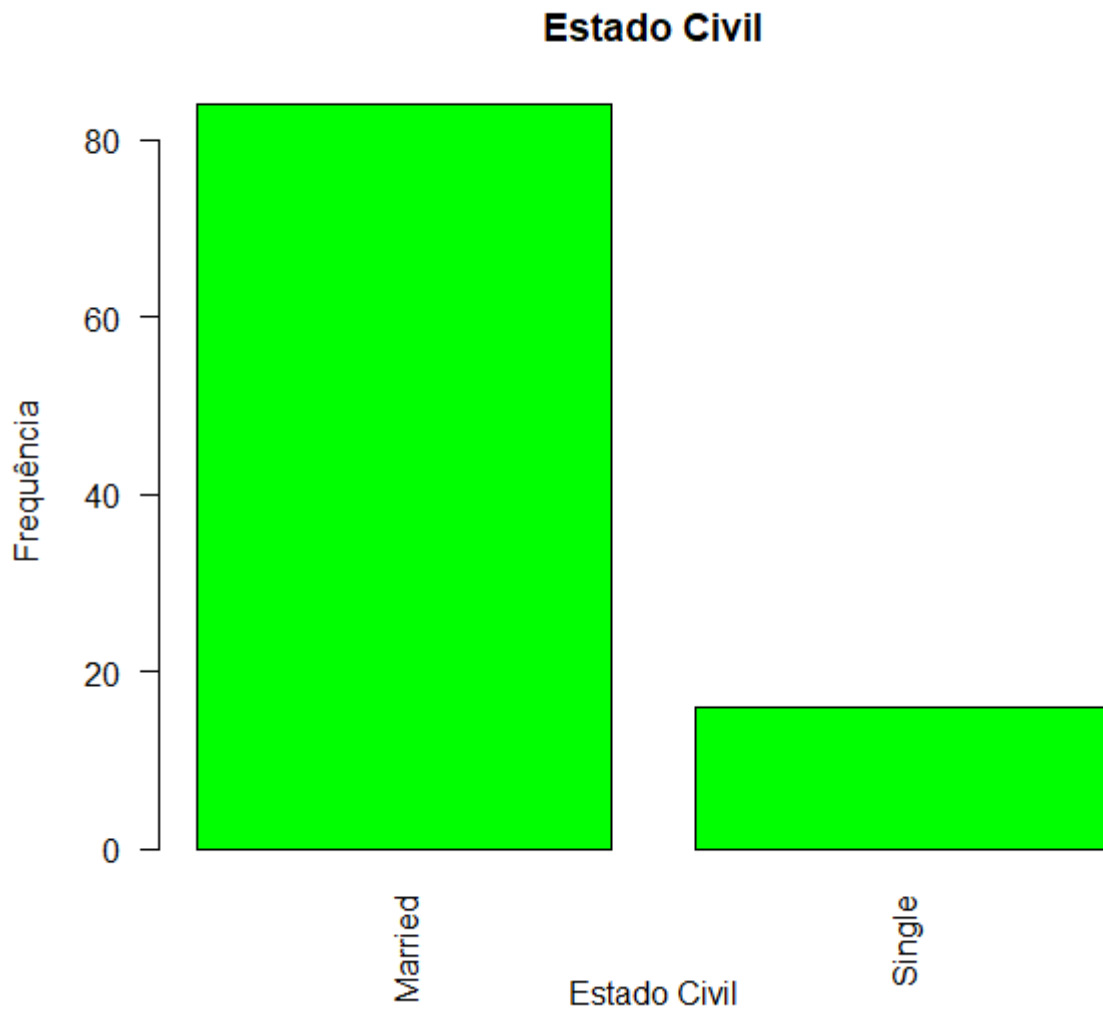
```





Frequência por Gênero





Fiz uma relação entre Type of Customer e Net Sales. A ideia é verificar se as vendas feitas para clientes promocionais apresentam valores maiores do que a venda para clientes regulares.

```
##### Estatísticas Descritivas #####
# Type of customer x Net sales /
# Média, mediana e desvio padrão /
#####
dados_trabalho$Type.of.Customer <- as.factor(dados_trabalho$Type.of.Customer)
dados_trabalho$Net.Sales <- as.character(dados_trabalho$Net.Sales)
dados_trabalho$Net.Sales <- gsub(",", ".", dados_trabalho$Net.Sales)
dados_trabalho$Net.Sales <- as.numeric(dados_trabalho$Net.Sales)

str(dados_trabalho)

# Estatísticas descritivas por grupo
```

```
media_cada_grupo <- aggregate(Net.Sales ~ Type.of.Customer, data = dados_trabalho, FUN = mean)
mediana_por_grupo <- aggregate(Net.Sales ~ Type.of.Customer, data = dados_trabalho, FUN = median)
desvio_padrao_por_grupo <- aggregate(Net.Sales ~ Type.of.Customer, data = dados_trabalho, FUN = sd)

# Teste t
teste <- t.test(Net.Sales ~ Type.of.Customer, data = dados_trabalho)
cat("Teste entre grupos:\n")
print(teste)
```

Criei um boxplot para visualizar a distribuição das vendas líquidas por tipo de cliente.

```
source("analise.r")

#===== boxplot vendas x tipo cliente =====#

boxplot(
  Net.Sales ~ Type.of.Customer,
  data = dados_trabalho,
  main = "Net Sales por tipo de cliente",
  ylab = "Vendas líquidas",
  xlab = "Tipo de clientes",
  col = c("#ff7fff", "#d64eff")
)
```

A saída para o teste t foi a seguinte:

```
Teste entre grupos:

      Welch Two Sample t-test

data:  Net.Sales by Type.of.Customer
t = 2.2883, df = 90.035, p-value = 0.02446
alternative hypothesis: true difference in means between group Promotional and group Regular is not equal to 0
95 percent confidence interval:
 2.939014 41.657653
sample estimates:
mean in group Promotional    mean in group Regular
          84.29000              61.99167
```

- Conclusão: Nesse caso de comparação tem uma diferença entre os grupos. Clientes Promotional gastaram mais em média que os Regular porque o p-valor baixo (< 0.05) e no intervalo de confiança que não inclui 0.

=====

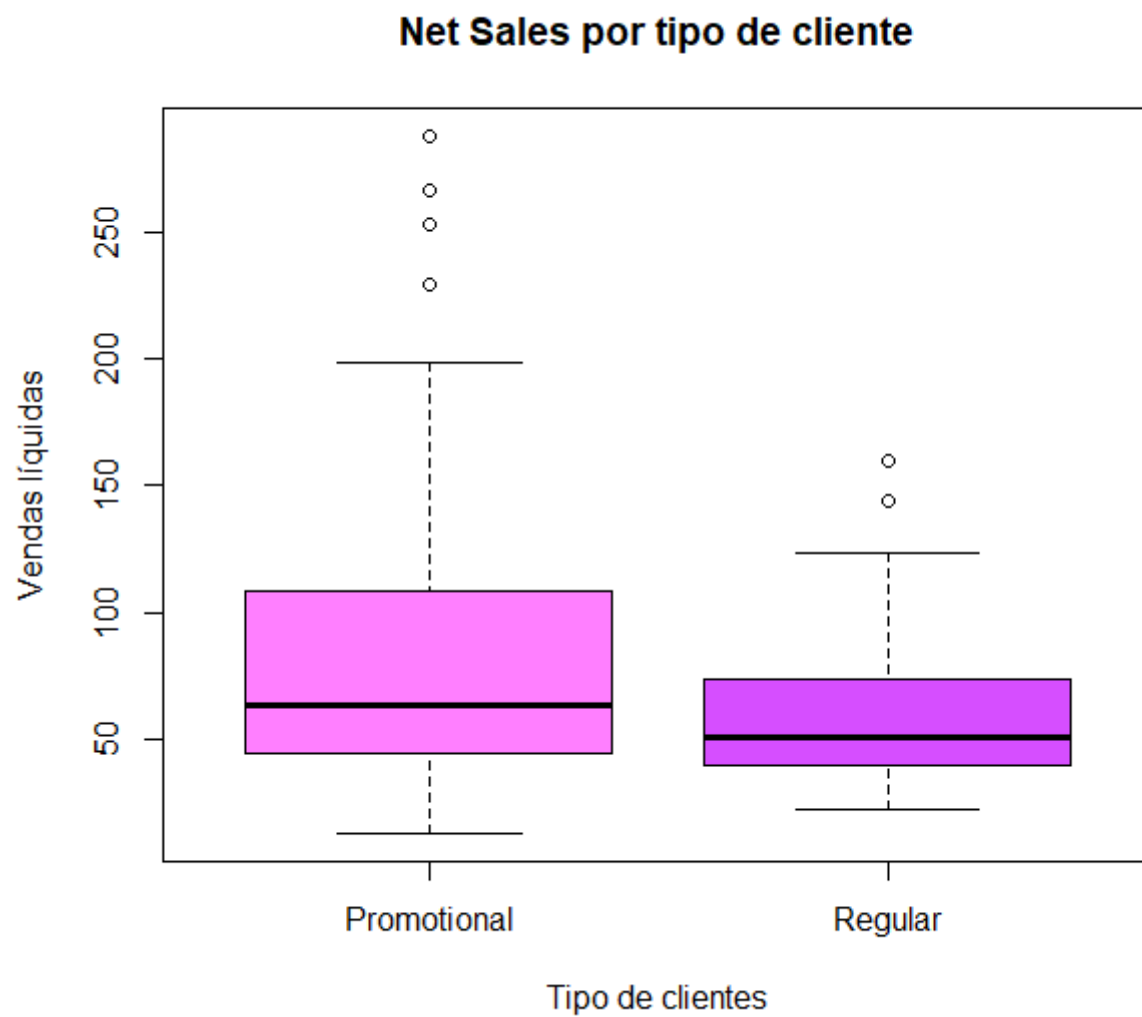


Figure 2: Alt text

Nesse parte eu fiz uma distribuição de itens na qual o objetivo é entender como os clientes se comportam em relação à quantidade de produtos adquiridos por compra

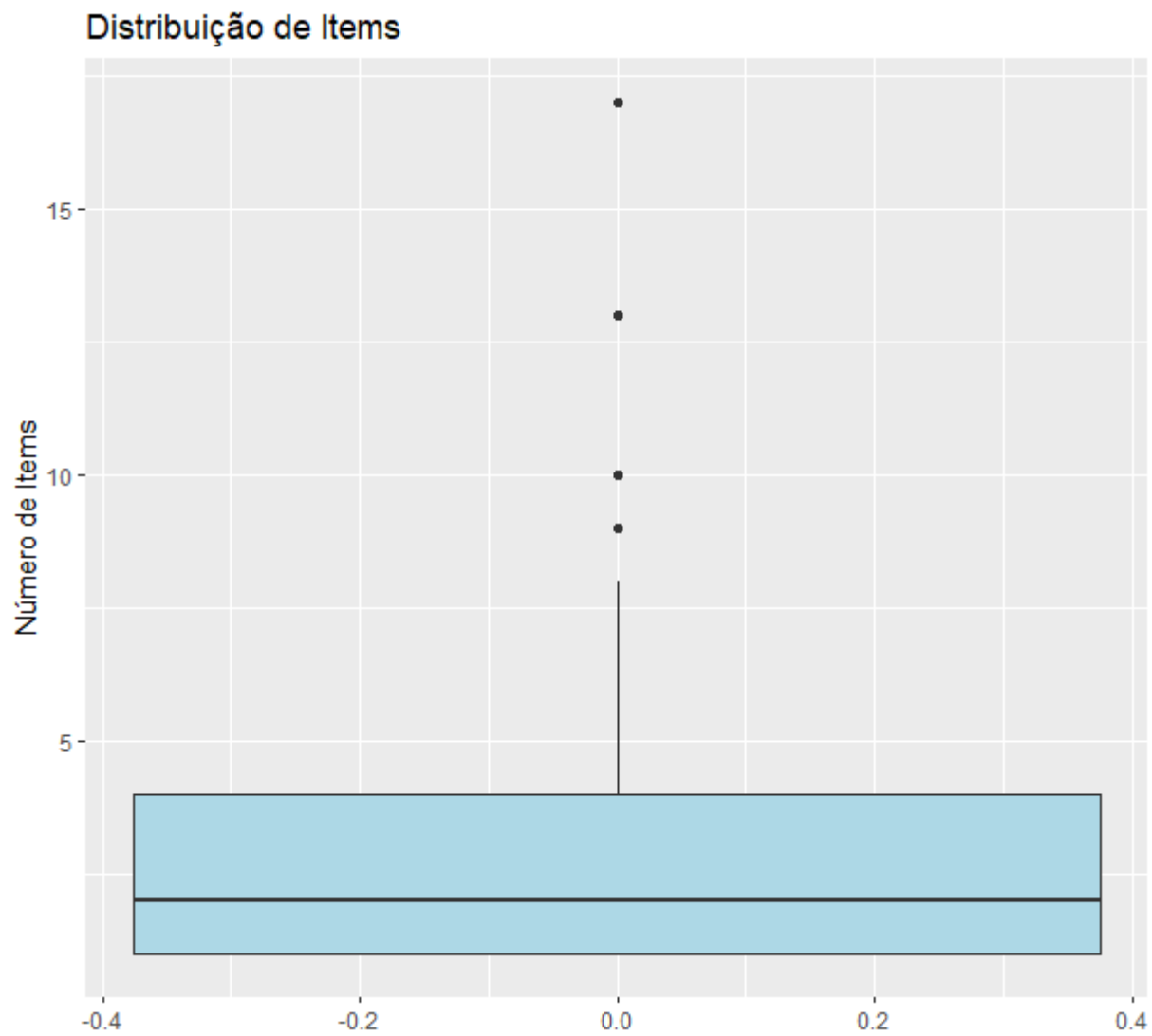
```
source("analise.r")

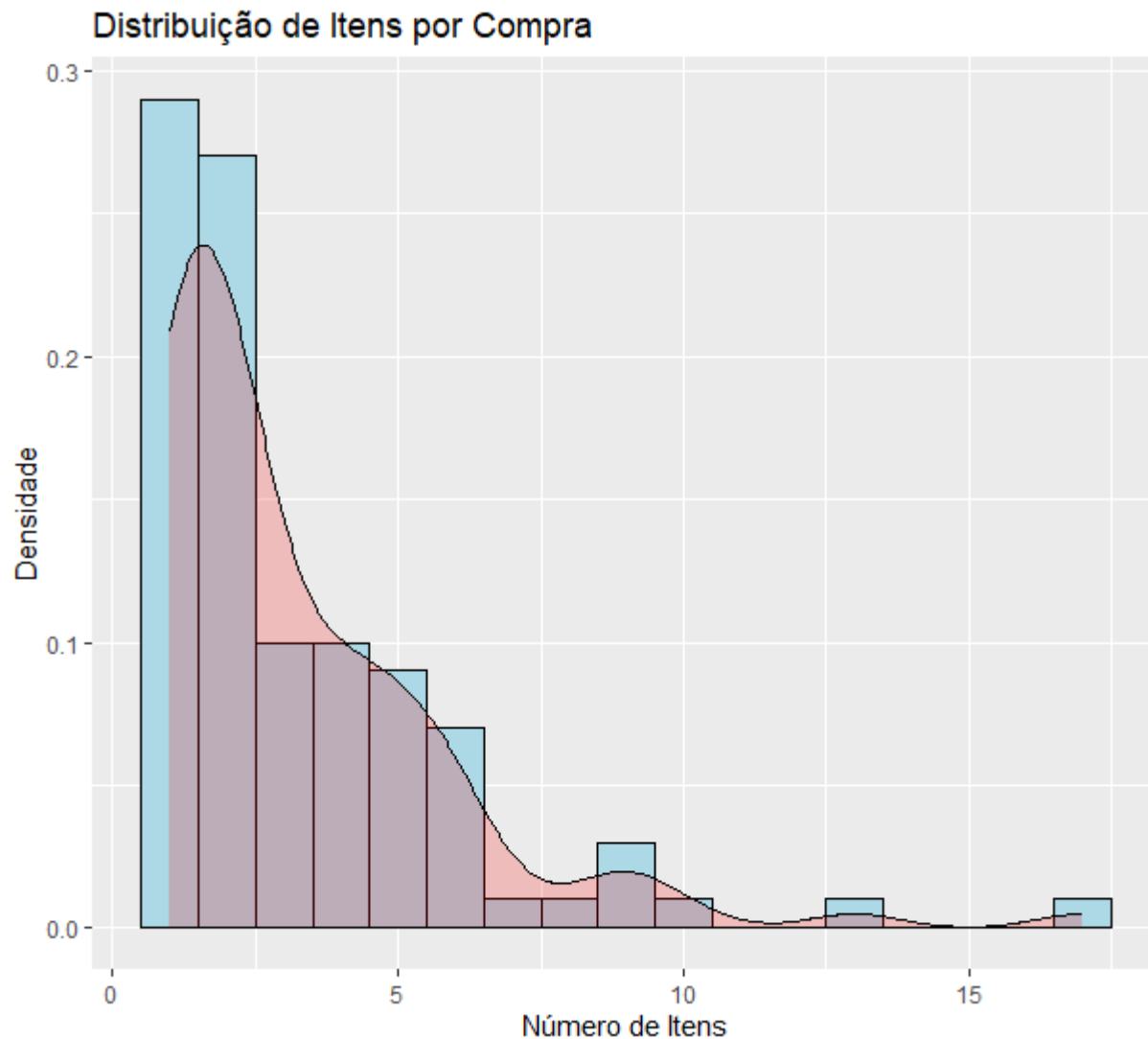
#===== distribuição de itens =====#
print(ggplot(dados_trabalho, aes(y = Items)) +
      geom_boxplot(fill = "lightblue") +
      labs(title = "Distribuição de Itens", y = "Número de Itens"))

source("analise.r")

#===== distribuição da freqência de itens =====#

print(ggplot(dados_trabalho, aes(x = Items)) +
      geom_histogram(aes(y = ..density..), binwidth = 1, fill = "lightblue", color = "black") +
      geom_density(alpha = 0.2, fill = "red") +
      labs(title = "Distribuição de Itens por Compra", x = "Número de Itens", y = "Densidade"))
```





Fiz também uma análise da quantidade de itens por características qualitativas dos clientes. O objetivo é verificar se há diferenças no comportamento de compra

```
#===== Itens x gênero =====#

print(
  ggplot(dados_trabalho, aes(x = Gender, y = Items, fill = Gender)) +
  geom_boxplot() +
  labs(title = "Distribuição de Itens por Gênero", x = "Gênero", y = "Número de Items") +
  theme_minimal()
)

#===== Itens x estado civil =====#

print(ggplot(dados_trabalho, aes(x = Marital.Status, y = Items, fill = Marital.Status)) +
  geom_boxplot() +
```



```

labs(title = "Distribuição de Items por Estado Civil", x = "Estado Civil", y = "Número de Items") +
theme_minimal()

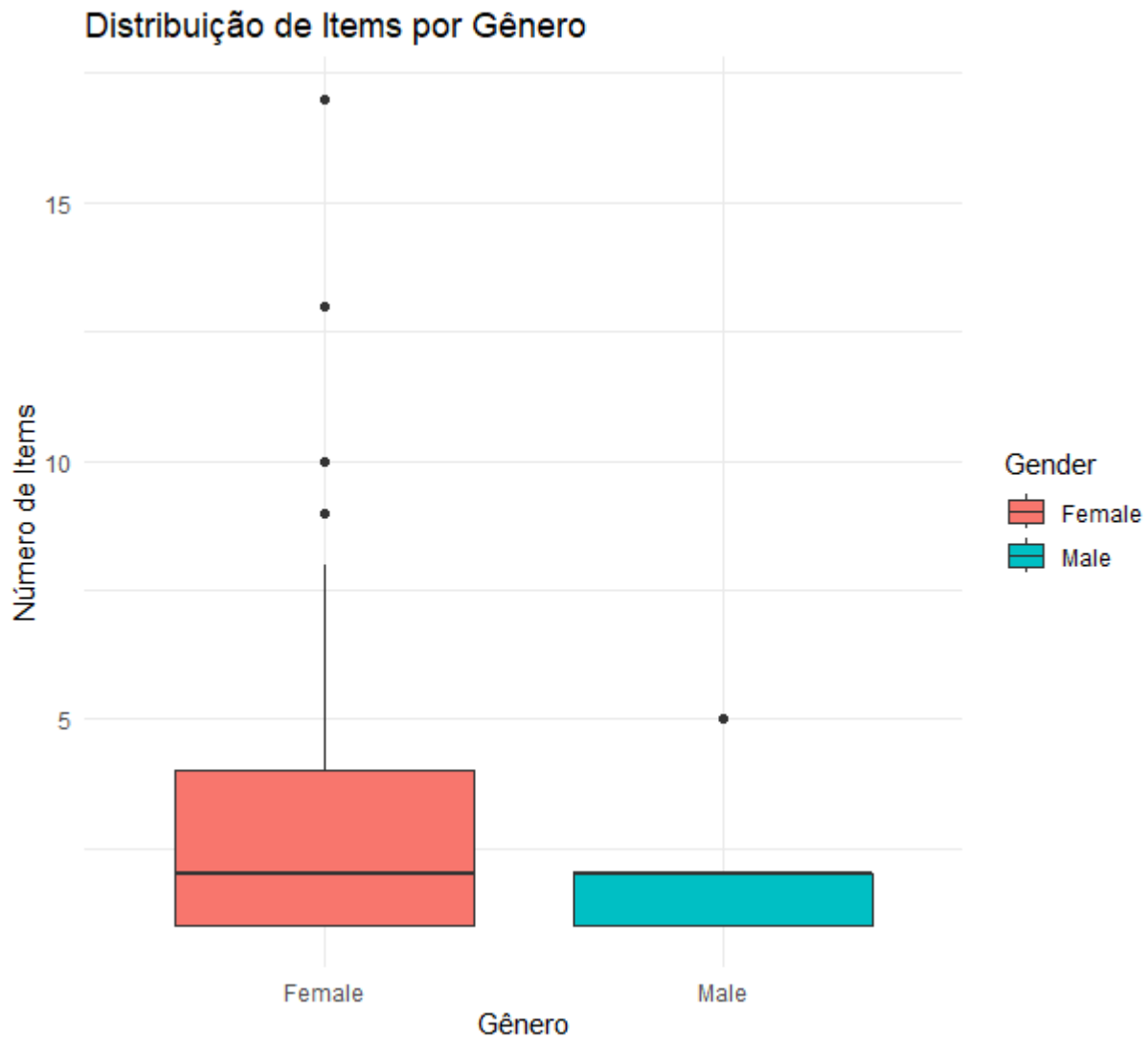
#===== Itens x tipo de cliente =====#

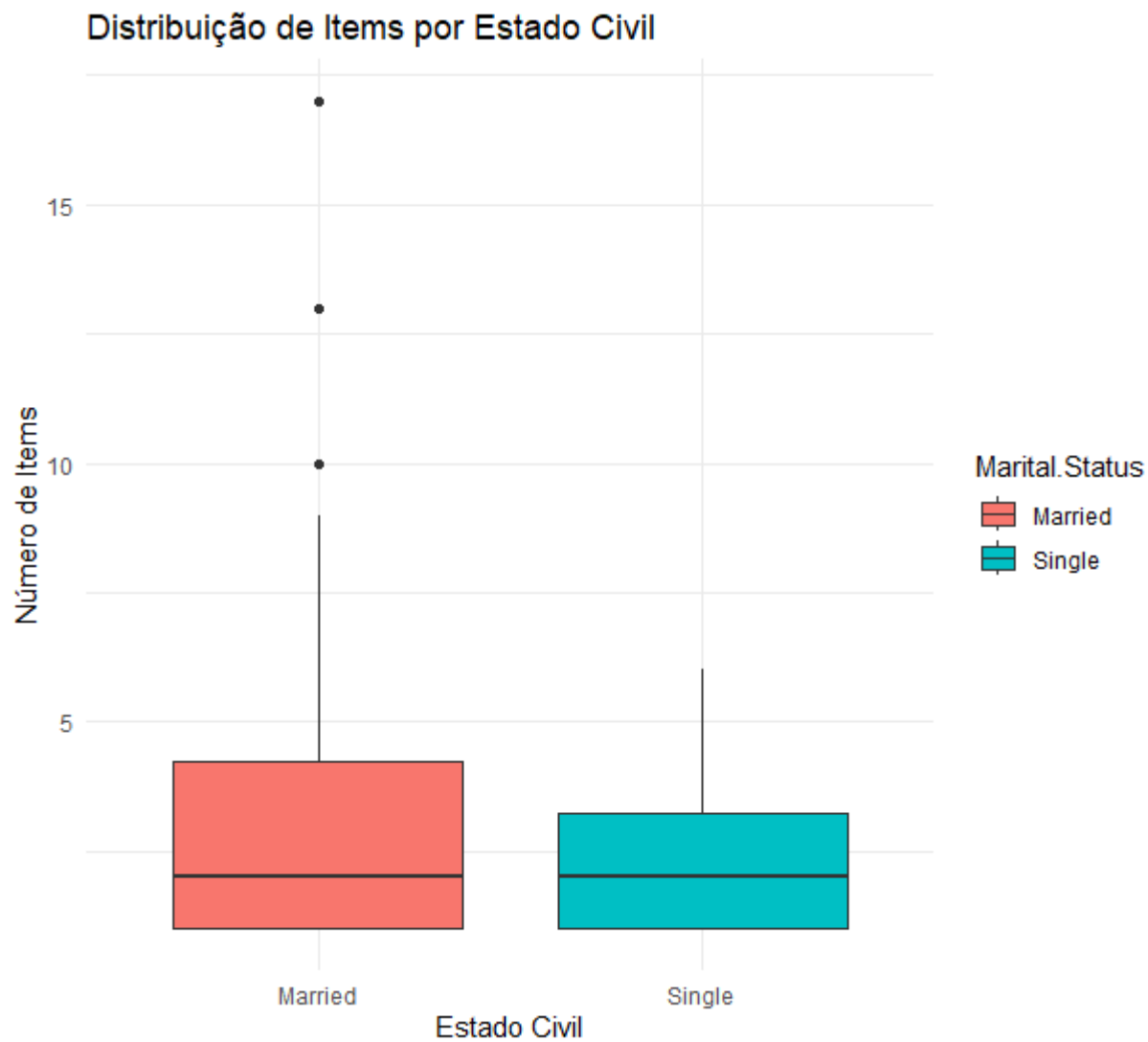
print(ggplot(dados_trabalho, aes(x = Type.of.Customer, y = Items, fill = Type.of.Customer)) +
  geom_boxplot() +
  labs(title = "Distribuição de Items por Tipo de cliente", x = "Tipo de cliente", y = "Número de Items") +
  theme_minimal())

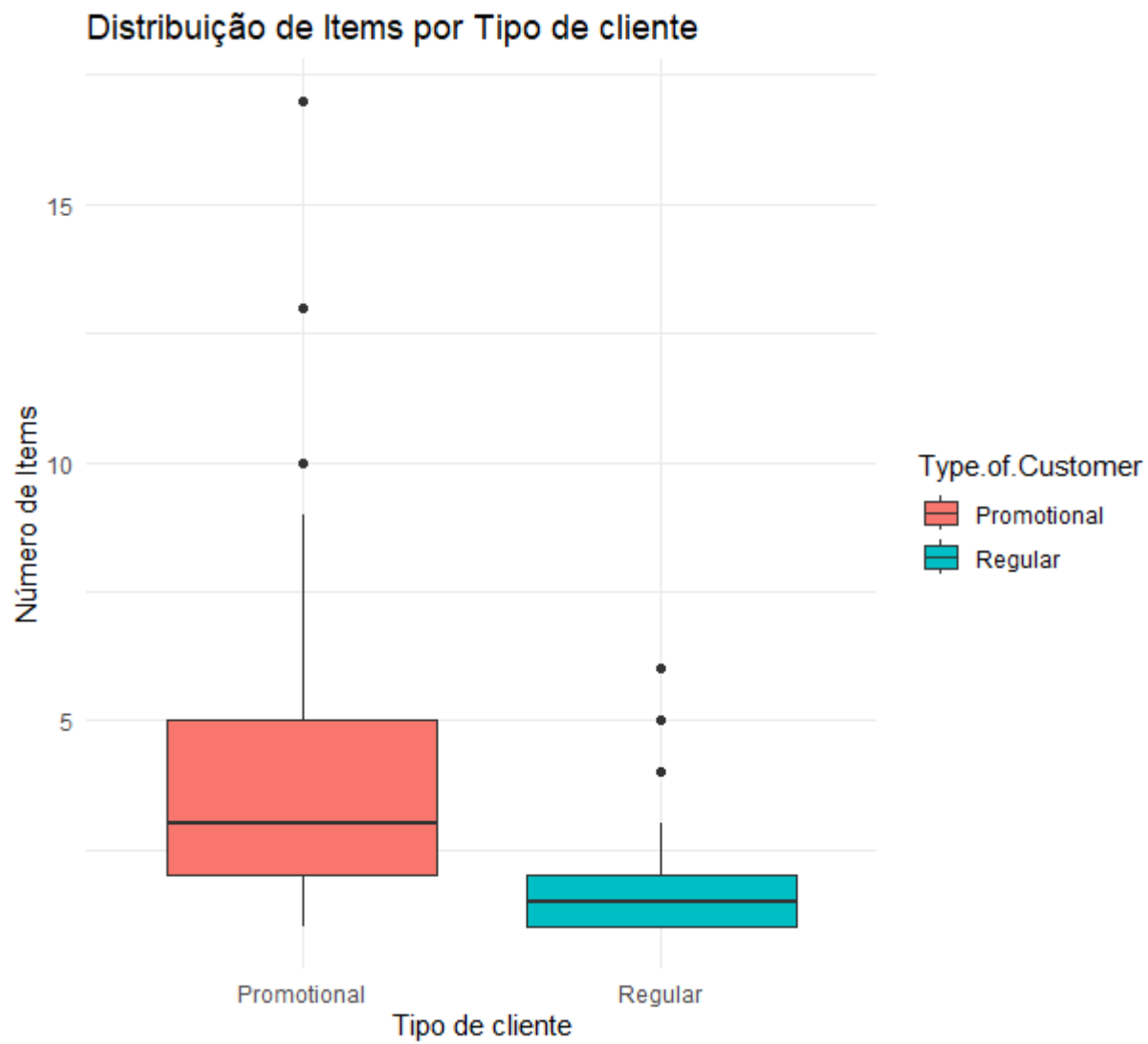
#===== Itens x método de pagamento =====#

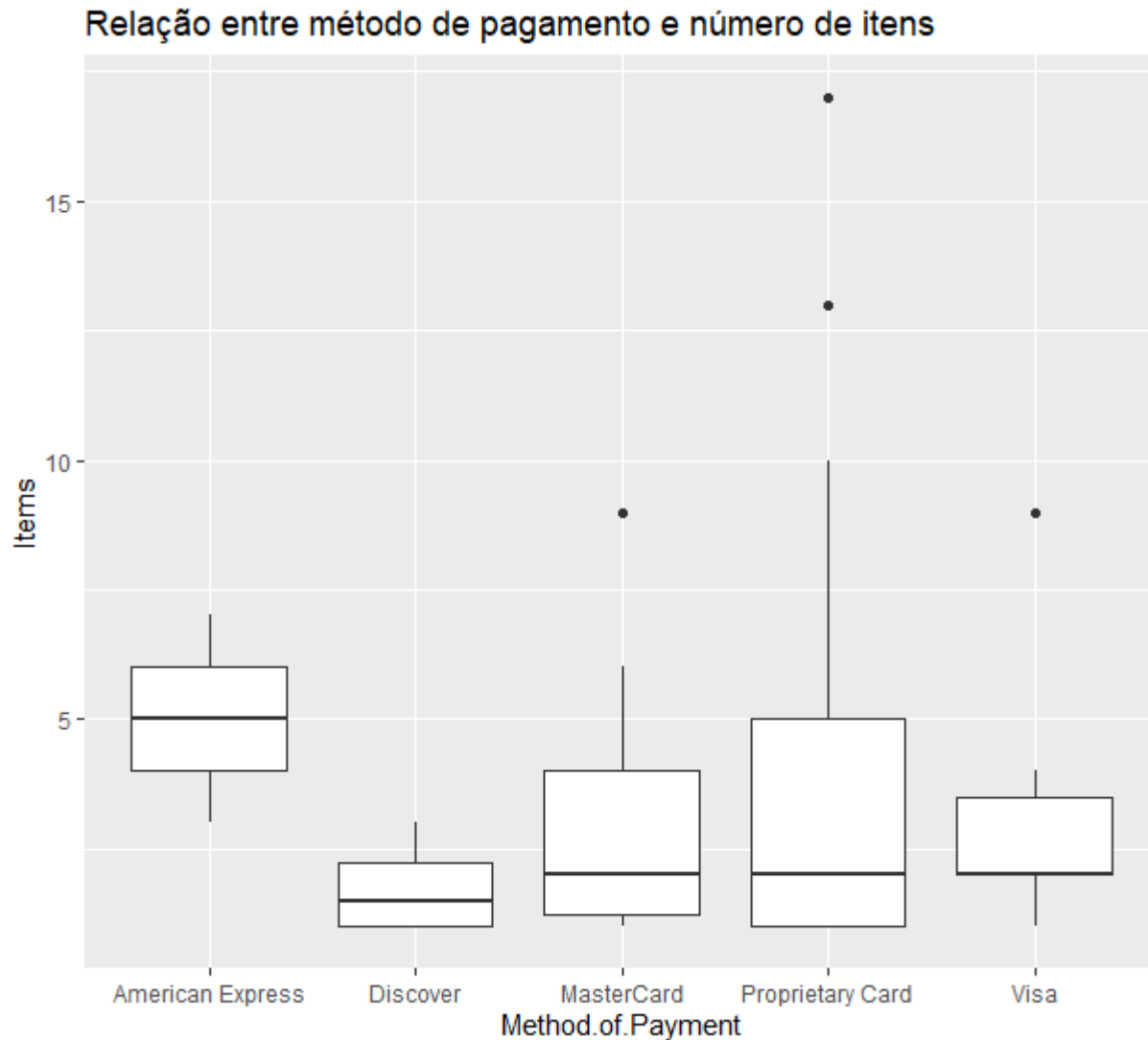
print(ggplot(dados_trabalho, aes(x = Method.of.Payment, y = Items)) +
  geom_boxplot() +
  labs(title = "Relação entre método de pagamento e número de itens"))

```









Análisei da relação entre número de itens e vendas líquidas. Nesta seção, eu analiso como o número de itens adquiridos em uma compra se relaciona com as vendas líquidas. Utilizei uma correlação, um modelo de regressão linear simples e visualização de um grafico.

Matriz de correlação - A matriz de correlação foi calculada entre as variáveis Net.Sales, Items e Age, para avaliar a força e a direção das relações lineares entre elas.

```
correlacao <- cor(dados_trabalho[, c("Net.Sales", "Items", "Age")])
diag(correlacao) <- NA
print(correlacao)
```

Modelo linear simples - Foi ajustado um modelo de regressão linear simples para prever as vendas líquidas a partir do número de itens. O resumo do modelo mostra o quanto cada item adicional impacta, em média, no valor das vendas.

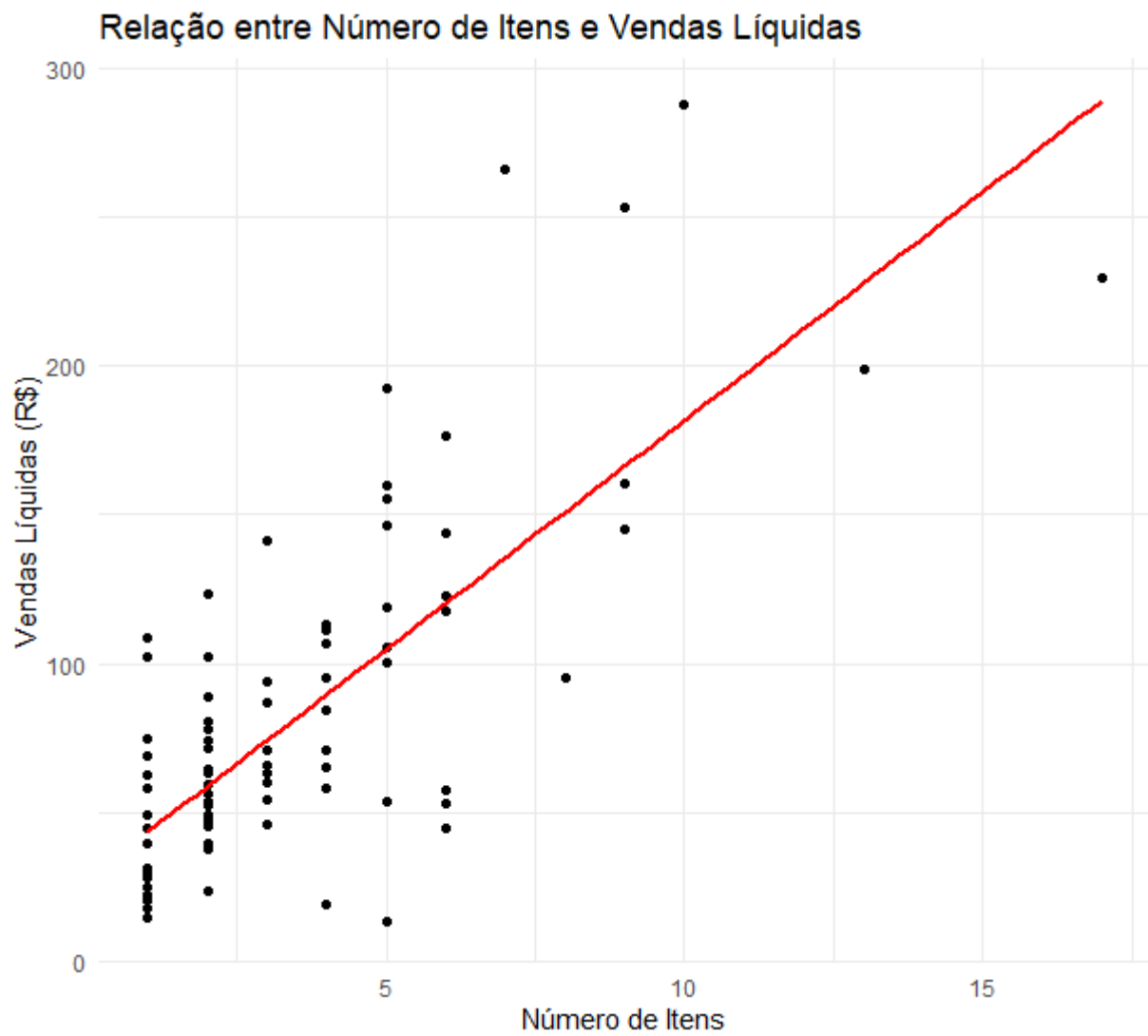
```
modelo_final <- lm(Net.Sales ~ Items, data = dados_trabalho)
summary(modelo_final)
```

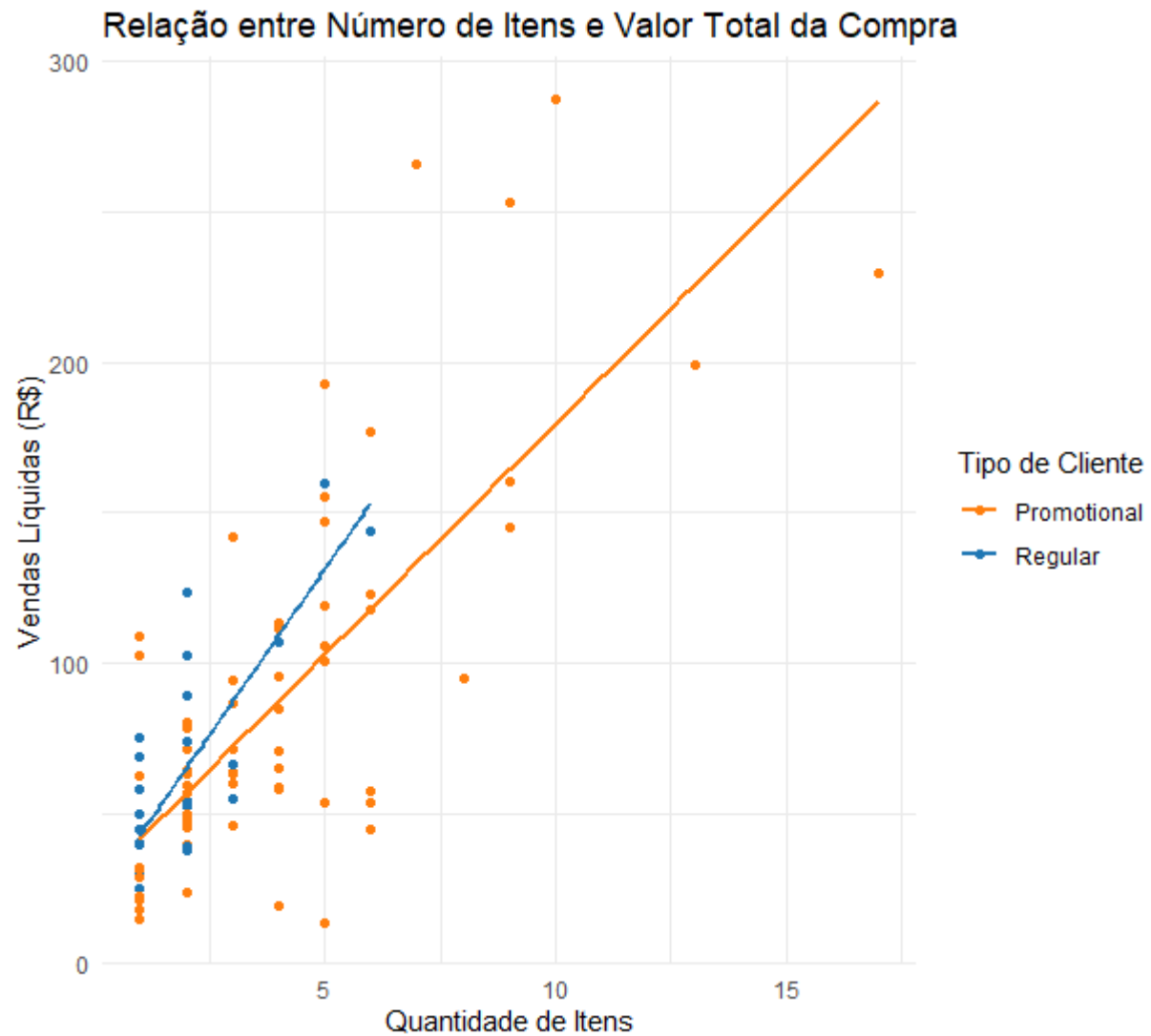
Gráfico de dispersão com reta de regressão - O gráfico abaixo mostra a relação entre o número de itens e as vendas líquidas, com a reta ajustada pelo modelo linear.

```
library(ggplot2)

ggplot(dados_trabalho, aes(x = Items, y = Net.Sales)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Relação entre Número de Itens e Vendas Líquidas",
       x = "Número de Itens",
       y = "Vendas Líquidas (R$)") +
  theme_minimal()

#===== Gráfico: Itens x Vendas Líquidas =====#
print(ggplot(dados_trabalho, aes(x = Items, y = Net.Sales, color = Type.of.Customer)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Relação entre Número de Itens e Valor Total da Compra",
    x = "Quantidade de Itens",
    y = "Vendas Líquidas (R$)",
    color = "Tipo de Cliente"
  ) +
  scale_color_manual(values = c("Regular" = "#1f77b4", "Promotional" = "#ff7f0e")) +
  theme_minimal())
```





Boxplot das vendas líquidas - Por fim, o boxplot abaixo apresenta a distribuição das vendas líquidas, destacando a mediana, a dispersão e possíveis outliers.

```
boxplot(dados_trabalho$Net.Sales,
        main = "Distribuição de Vendas Líquidas",
        col = "pink")
```

Aqui esta a saída esperada no terminal

```
Matriz de correlação:
      Net.Sales      Items      Age
Net.Sales      NA  0.75505939 -0.01063589
Items      0.75505939      NA -0.01661542
Age      -0.01063589 -0.01661542      NA
```

Distribuição de Vendas Líquidas

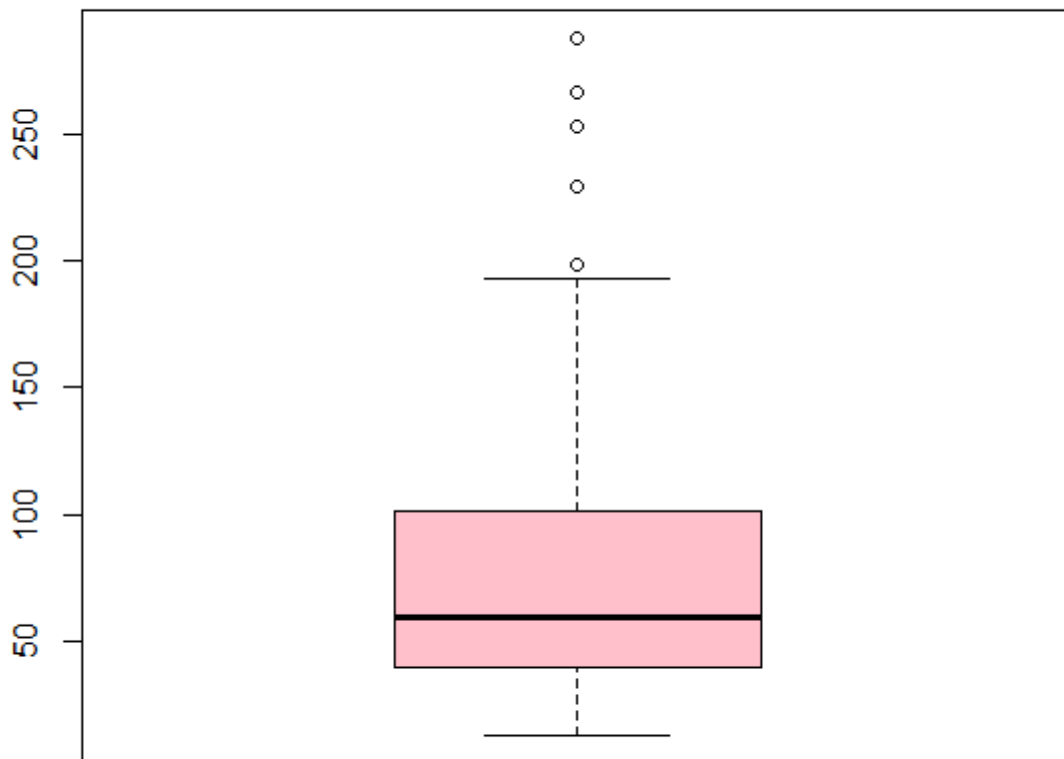


Figure 3: Alt text


```

Modelo final:

Call:
lm(formula = Net.Sales ~ Items, data = dados_trabalho)

Residuals:
    Min       1Q   Median       3Q      Max
-91.713 -19.450  -4.851  15.710 130.334

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   28.138     5.682   4.952 3.06e-06 ***
Items         15.361     1.347  11.400 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.68 on 98 degrees of freedom
Multiple R-squared:  0.5701,    Adjusted R-squared:  0.5657
F-statistic: 130 on 1 and 98 DF,  p-value: < 2.2e-16

```

Matriz de Correlação

Vemos na matriz de correlação entre as variáveis *Net.Sales*, *Items* e *Age* que a maior correlação é entre *Net.Sales* e *Items* (0,755), o que indica que as vendas líquidas aumentam à medida que mais itens são vendidos. As correlações com *Age* são todas próximas de zero, o que nos diz que a idade dos clientes não tem relação linear com as vendas nem com a quantidade de itens.

Regressão Linear de Vendas por Itens

Um modelo de regressão linear simples foi ajustado para prever as vendas (*Net.Sales*) a partir do número de itens vendidos (*Items*). O modelo ajustado foi:

$$\text{Net.Sales} = 28.14 + 15.36 \cdot \text{Items}$$

Ou seja, para cada item adicional vendido, a venda aumenta em média R\$ 15,36. O modelo explica cerca de 57% da variação nas vendas ($R^2 = 0.57$) e é estatisticamente significativo ($p < 0.001$), o que significa que a quantidade de itens vendidos é um bom preditor do valor das vendas.

=====

Como foi pedido no trabalho e lendo alguns artigos, decidi implementar uma variável aleatório de início Loyalty Score porém ela não se relacionou com uma precisão de venda boa, então decidi trocar por variável Ticket Médio, que é uma métrica importante para entender o comportamento de compra dos clientes. O Ticket Médio é calculado dividindo o total de vendas pelo número de transações realizadas. Essa métrica ajuda a identificar o valor médio gasto por cliente em cada compra. Essa variável foi muito útil para a análise, já que na minha perspectiva seria de aumentar o lucro da empresa.

Criação das novas variáveis

- *Age_Group*: você criou uma variável categórica com faixas etárias (*idade_classes*), para facilitar análises por idade.

- Ticket_Medio: calcula o valor médio pago por item em cada compra, dividindo Net.Sales pelo número de Items. Essa métrica ajuda a entender o “tamanho” médio do gasto por item.
- Cliente_premium: marca os clientes no quartil superior das vendas líquidas como Premium (acima do 75º percentil) e os demais como Regular. Essa segmentação permite analisar diferenças entre clientes de maior e menor valor.

```
dados_trabalho <- dados_trabalho %>%
  mutate(
    Age_Group = idade_classes,
    Ticket_Medio = Net.Sales / Items,
    Cliente_premium = ifelse(Net.Sales > quantile(Net.Sales, 0.75), "Premium", "Regular")
  )
```

Boxplot: Ticket Médio por Tipo de Cliente

Esse gráfico compara o Ticket Médio entre os dois tipos de cliente (Promotional e Regular), visualizando:

- Diferenças de mediana.
- Dispersão.
- Outliers (dados que se diferenciam drasticamente de todos os outros).

Serve para identificar se clientes regulares ou promocionais tendem a gastar mais por item.

```
ggplot(dados_trabalho, aes(x = Type.of.Customer, y = Ticket_Medio, fill = Type.of.Customer)) +
  geom_boxplot() +
  labs(
    title = "Nova análise: Ticket por Tipo de Cliente",
    y = "R$ por item"
  ) +
  theme_minimal()
```

Teste estatístico: diferença no Ticket Médio

```
t_test_result <- t.test(Ticket_Medio ~ Type.of.Customer, data = dados_trabalho)
```

- Testa se a média do Ticket Médio difere significativamente entre clientes Promotional e Regular.
- Se o p-valor < 0,05, você conclui que existe diferença estatisticamente significativa.

Regressão: Ticket Médio por Tipo de Cliente

```
modelo_estrategico_ticket_medio <- lm(Ticket_Medio ~ Type.of.Customer, data = dados_trabalho)
```

- Modelo simples para quantificar o impacto de ser Promotional ou Regular sobre o Ticket Médio.
- Complementa o t-test, mas com uma análise mais formal e possibilitando ver os coeficientes.

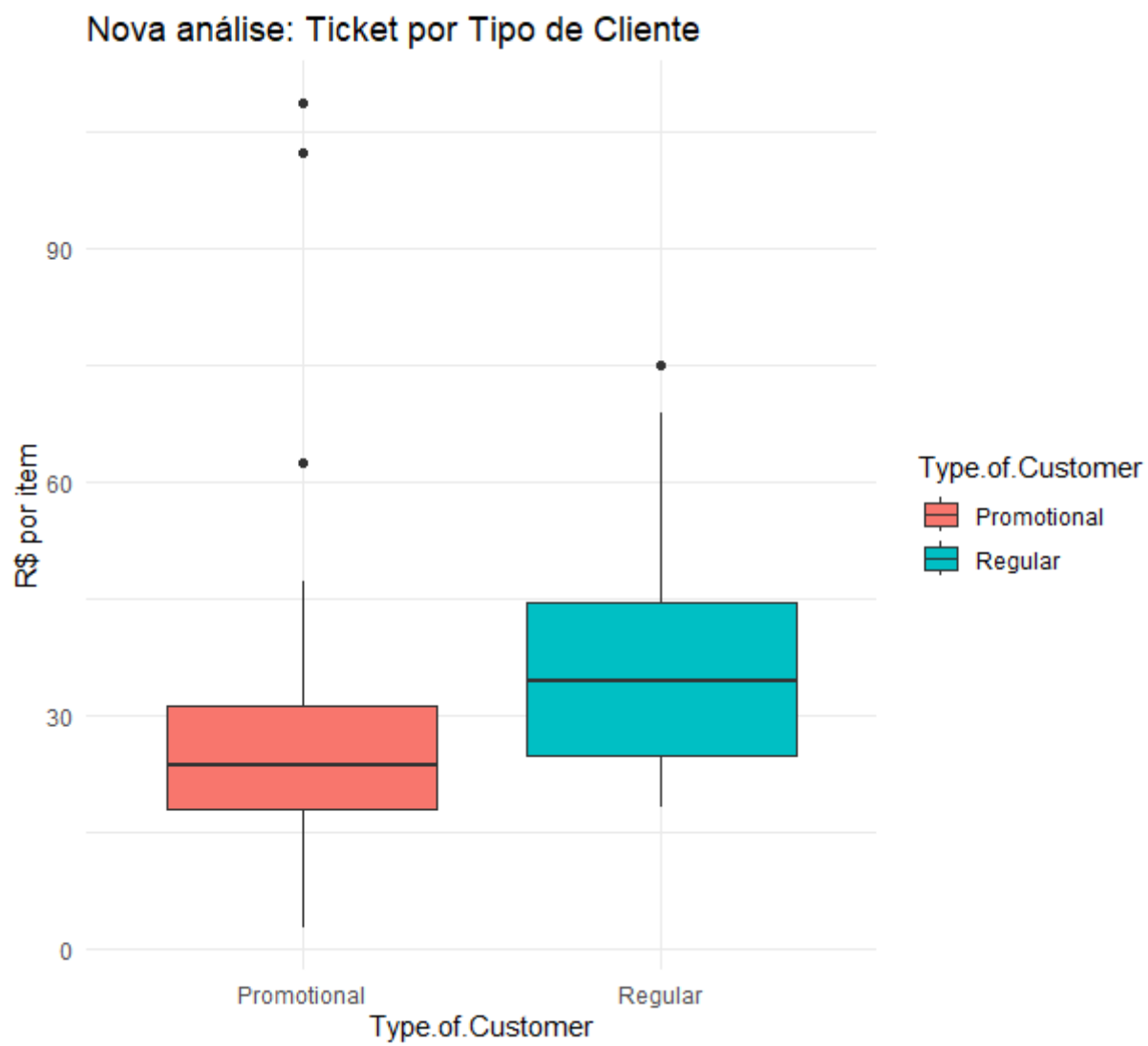


Figure 4: Alt text

Gráfico: Método de Pagamento por Status Premium

```
ggplot(dados_trabalho, aes(x = Method.of.Payment, fill = Cliente_premium)) +  
  geom_bar(position = "fill") +  
  labs(  
    title = "Nova análise: Métodos de Pagamento por Status Premium",  
    y = "Proporção"  
  ) +  
  theme_minimal() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

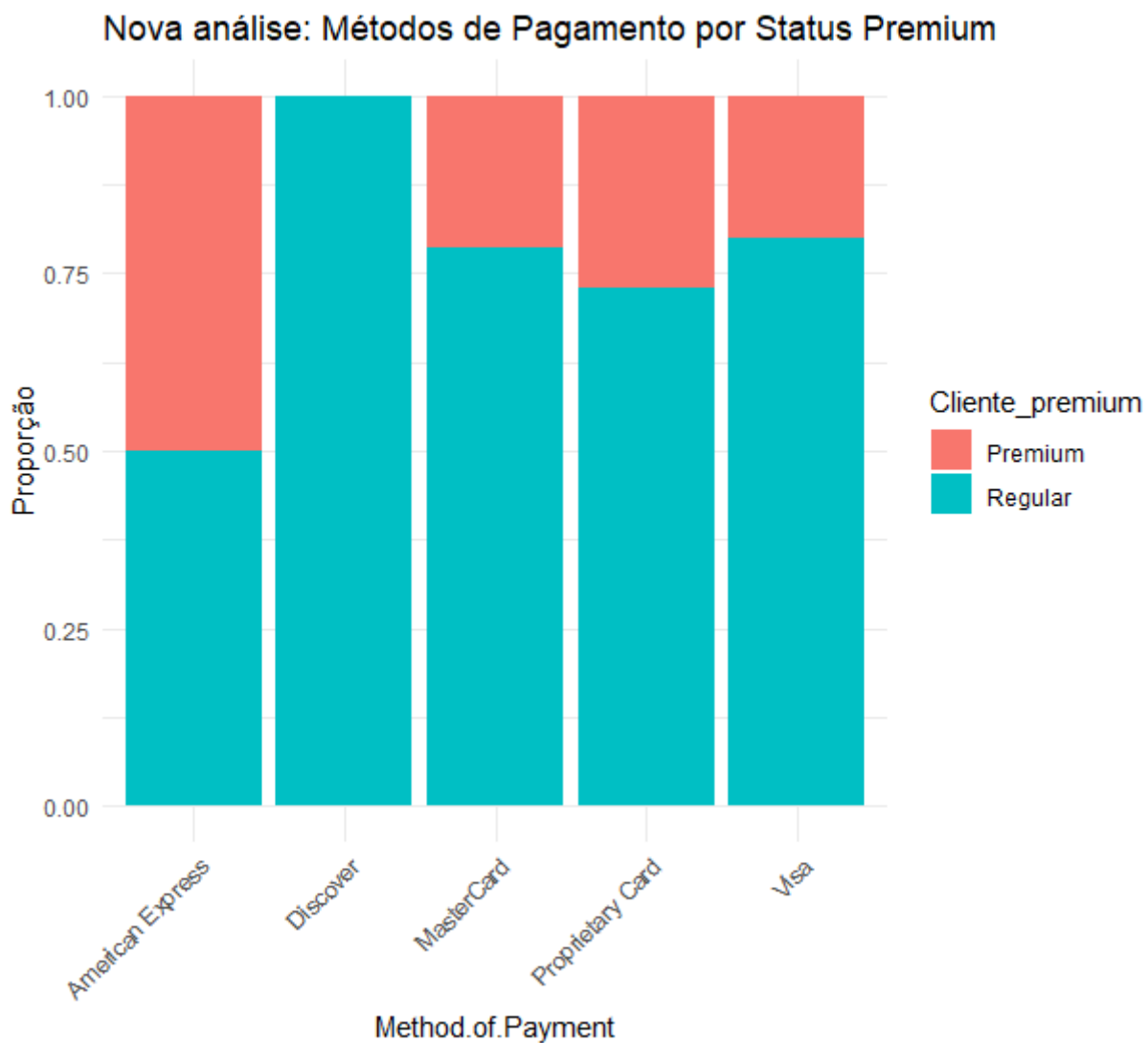


Figure 5: Alt text

- Mostra a proporção dos métodos de pagamento usados, separados entre clientes Premium e Regular.
- Ajuda a entender se clientes Premium têm preferência por algum método (ex.: cartão proprietário)

Modelo de regressão múltipla

```
modelo_estrategico <- lm(Net.Sales ~ Items + Ticket_Medio + Cliente_premium, data = dados_trabalho)
```

Modelo mais completo para prever as Net.Sales, considerando:

- Número de itens.
- Ticket médio.
- Status Premium.

Permite identificar quais dessas variáveis têm maior impacto nas vendas e se elas explicam bem a variação observada.

```
[1] "TESTE NOVO: Diferença no Ticket entre Clientes Regulares e Promocionais"
```

```
Welch Two Sample t-test
```

```
data: Ticket_Medio by Type.of.Customer
```

```
t = -2.9749, df = 61.267, p-value = 0.004189
```

```
alternative hypothesis: true difference in means between group Promotional and group Regular is not equal
```

```
95 percent confidence interval:
```

```
-17.114367 -3.356147
```

```
sample estimates:
```

```
mean in group Promotional    mean in group Regular
                26.55197                36.78722
```

```
[1] "Modelo Estratégico: Ticket Médio por Tipo de Cliente\n"
```

```
Call:
```

```
lm(formula = Ticket_Medio ~ Type.of.Customer, data = dados_trabalho)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-23.906 -10.294  -2.885   5.048  82.248
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      26.552      1.974   13.45 < 2e-16 ***
Type.of.CustomerRegular  10.235      3.604    2.84 0.00549 **
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 16.52 on 98 degrees of freedom
```

```
Multiple R-squared:  0.07603,    Adjusted R-squared:  0.0666
```

```
F-statistic: 8.064 on 1 and 98 DF,  p-value: 0.00549
```

```
[1] "MODELO NOVO: Impacto do Ticket e Status Premium nas Vendas"
```

```
Call:
```

```
lm(formula = Net.Sales ~ Items + Ticket_Medio + Cliente_premium,
```

```

data = dados_trabalho)

Residuals:
    Min       1Q   Median       3Q      Max
-51.746  -7.065  -3.679   9.479  94.165

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    44.3524    13.5910   3.263  0.00153 **
Items          13.0849     1.3325   9.820 3.63e-16 ***
Ticket_Medio     0.9444     0.1768   5.341 6.19e-07 ***
Cliente_premiumRegular -49.1486     7.8802  -6.237 1.20e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.18 on 96 degrees of freedom
Multiple R-squared:  0.8319,    Adjusted R-squared:  0.8267
F-statistic: 158.4 on 3 and 96 DF,  p-value: < 2.2e-16

```

Resultado das Análises Adicionais

1. Teste t para o Ticket Médio por Tipo de Cliente

Realizamos um teste t de Welch para verificar se o ticket médio (*venda média por item*) difere entre clientes **promocionais** e **regulares**.

Os resultados mostram:

- Média dos clientes **Promocionais**: 26,55
- Média dos clientes **Regulares**: 36,79
- Estatística t = -2,97
- p-valor = 0,004

Como o p-valor é menor que 0,05, concluímos que existe uma diferença estatisticamente significativa no ticket médio entre os dois grupos, indicando que clientes regulares tendem a gastar mais por item do que clientes promocionais.

2. Regressão Linear Simples: Ticket Médio por Tipo de Cliente

Ajustamos um modelo de regressão para quantificar a diferença no ticket médio entre os grupos. O modelo é dado por:

$$\text{Ticket Médio} = \beta_0 + \beta_1(\text{Cliente Regular})$$

- Intercepto (β_0): 26,55 → ticket médio dos **promocionais**
- Coeficiente para **Cliente Regular**: +10,24, com p-valor 0,005

Ou seja, ser um cliente regular está associado a um aumento médio de aproximadamente 10,24 unidades monetárias no ticket médio em relação aos promocionais. Esse resultado confirma o teste t.

O modelo explica cerca de 7,6% da variação no ticket médio ($R^2 \approx 0,076$).

3. Regressão Múltipla: Impacto do Ticket e Status Premium nas Vendas Líquidas

Por fim, ajustamos um modelo mais completo para explicar as **vendas líquidas**, incluindo como variáveis explicativas: número de itens comprados, ticket médio e status premium do cliente.

O modelo é dado por:

$$\text{Net Sales} = \beta_0 + \beta_1(\text{Items}) + \beta_2(\text{Ticket Médio}) + \beta_3(\text{Cliente Premium})$$

Os principais resultados:

- Para cada item adicional, as vendas líquidas aumentam, em média, 13,08.
 - Para cada unidade monetária a mais no ticket médio, as vendas líquidas aumentam, em média, 0,94.
 - Clientes regulares (em relação aos premium) têm em média 49,15 a menos em vendas líquidas.
 - O modelo tem ($R^2 \approx 0,83$), explicando mais de 80% da variação nas vendas líquidas, o que indica um bom ajuste.
-

Resumo

As análises revelam que:

- Clientes regulares tendem a ter um ticket médio maior do que os promocionais.
- O número de itens, o ticket médio e o status premium são fatores importantes para explicar as vendas líquidas.
- O modelo múltiplo mostra que essas variáveis, em conjunto, têm grande poder explicativo das vendas.