

**UFRRJ**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM**  
**ESTATÍSTICA APLICADA**

**MONOGRAFIA**

**Análise Estatística sobre Pontos de Coleta de**  
**Materiais Recicláveis dentro do Estado do Rio de**  
**Janeiro**

**Daniel Farias Gomes**

**2023**



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA**

**ANÁLISE ESTATÍSTICA SOBRE PONTOS DE COLETA DE  
MATERIAIS RECICLÁVEIS DENTRO DO ESTADO DO RIO DE  
JANEIRO**

**DANIEL FARIAS GOMES**

*Sob a Orientação do professor*  
**Dr. Felipe Leite Coelho**

Monografia submetida como requisito parcial para conclusão do curso de Especialização em **Estatística Aplicada** “*latu sensu*”, do Departamento de Matemática da UFRRJ.

Seropédica, RJ  
Novembro de 2023

Universidade Federal Rural do Rio de Janeiro  
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada  
com os dados fornecidos pelo(a) autor(a)

G633a      Gomes, Daniel Farias, 1986-  
              Análise estatística sobre pontos de coleta de  
              materiais recicláveis dentro do Estado do Rio de  
              Janeiro / Daniel Farias Gomes. - Rio de Janeiro, 2023.  
              62 f.: il.

              Orientador: Felipe Leite Coelho da Silva.  
              Monografia(Especialização). -- Universidade Federal  
              Rural do Rio de Janeiro, Especialização em Estatística  
              Aplicada "latu sensu", 2023.

              1. reciclagem. 2. análise exploratória. 3. análise  
              de cluster. 4. estatística. I. Silva, Felipe Leite  
              Coelho da, 1981-, orient. II Universidade Federal  
              Rural do Rio de Janeiro. Especialização em Estatística  
              Aplicada "latu sensu" III. Título.

É permitida a cópia parcial ou total desta dissertação, desde que seja citada a fonte.

**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA APLICADA - PPGEAp**

**DANIEL FARIAS GOMES**

Monografia submetida como requisito parcial para obtenção do grau de **Especialista em Estatística Aplicada “*lato sensu*”**, do departamento de matemática da UFRRJ.

MONOGRAFIA APROVADA EM: 04/12/2023.

---

Felipe Leite Coelho da Silva Dr. UFRRJ  
(Orientador, UFRRJ)

---

Renan Vicente Pinto Dr. UFRRJ  
UFRRJ

---

Josiane S. Cordeiro Coelho Dra. UFRRJ  
UFRRJ

## DEDICATÓRIAS

*A Deus e Jesus Cristo*

*Dedico.*

## AGRADECIMENTOS

Agradeço a Deus pela fé, saúde, sabedoria, tempo, oportunidade, por cuidar da minha família, por ser casado com a Vivian, mulher da minha vida e nos dar um filho lindo e amado chamado Davi.

Agradeço a minha esposa Vivian, mulher da minha vida, por ser paciente comigo, cuidar do Davi enquanto estava em estudo, por me amar e ter fé que a vida pode ser melhor caminhando junto com Jesus e com alguém ao lado. Te amo.

Agradeço aos meus pais Edson e Jussirene que tanto amo, por me ensinar os ensinamentos de Cristo, por todo amor dedicado e sacrifício para dar o melhor que puderam para mim e meus irmãos.

Agradeço aos meus irmãos Anderson e Danielle, meus melhores amigos e que amam a reciclagem e o meio ambiente.

Agradeço a minha sogra amada Ana por ajudar a cuidar do Davi para que eu pudesse me dedicar ao curso e trabalho.

Agradeço ao meu tio Marcelo por todo apoio da carreira profissional.

Agradeço ao meu orientador Felipe Leite por todo o conhecimento, apoio, paciência e acreditar neste trabalho.

Agradeço aos meus professores do curso de especialização em Estatística Aplicada realizado em 2022-2023 e coordenado pela professora Adriana Andrade. Obrigado mestres por revelar o mundo que era desconhecido para mim sobre estatística e dar o ensino para que eu possa levar isto para a vida profissional e acadêmica transmitindo a outras pessoas pelo caminho.

Obrigado aos meus amigos e parceiros do curso de especialização em Estatística Aplicada Maura, Ana Carolina, Arnaldo, Luiz, Hugo, Magno, Melissa e Rafael. Vocês compartilharam o conhecimento e tempo para que eu pudesse compreender melhor a disciplina e seguir adiante.

Agradeço a toda Secretaria de Estado de Ambiente e Sustentabilidade do Rio de Janeiro, principalmente nas pessoas da Ana Asti e Irlaine Alvarenga por acreditarem na reciclagem como política pública de distribuição de renda e educação ambiental e pela disponibilidade concedida para que eu pudesse fazer a especialização e desenvolver a carreira.

Agradeço aos membros da Cooperativa Recria e Ecco Ponto pelos ensinamentos sobre reciclagem e por se dedicarem a missão de preservar o planeta.

Agradeço aos meus familiares e amigos que direta ou indiretamente contribuíram para que este trabalho se concretizasse. Muito obrigado.

## RESUMO

GOMES, Daniel Farias. **Análise estatística sobre pontos de coleta de materiais recicláveis dentro do Estado do Rio de Janeiro. 2023.** 62 p. Monografia (Especialização em Estatística Aplicada “*latu sensu*”). Instituto de Ciências Exatas, Departamento de Matemática, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2023.

A reciclagem de materiais tem um papel importante na despoluição e preservação do meio ambiente e atua como uma solução para redução de dióxido de carbono e renovação dos recursos naturais da Terra. Diante disto, buscou-se encontrar informações e dados sobre projetos e práticas dentro do setor da reciclagem que consigam revelar, através da análise estatística, padrões de comportamento dos dados que permitam compreender melhor como é a dinâmica do mercado e possíveis estratégias para aumentar o alcance da logística reversa de resíduos. Assim, o objetivo deste trabalho foi identificar quais as variáveis que mais contribuem para captação de materiais recicláveis dentro de uma operação de um ponto de coleta. A análise estatística foi sobre a base de dados de pontos de coleta de materiais recicláveis registrados dentro do Estado do Rio de Janeiro entre junho de 2022 à fevereiro de 2023. Foi aplicado como metodologia a análise exploratória de dados e análise multivariada, em particular análise de agrupamento hierárquico. A partir dos resultados da análise exploratória foi possível identificar que em uma operação de ponto de coleta o público predominante tem idade entre 41 a 60 anos, 74% são mulheres, que o plástico é o material mais comercializado nas operações, que o vidro e papel representam mais de 55% do peso comprado e são também os materiais mais baratos, com valor igual ou inferior a R\$ 0,20/kg. Além disso, o preço da mediana de todos os materiais fica entre R\$ 0,90/kg a R\$ 1,00/kg e 79% da receita do catador vem de materiais de massa leve como alumínio, PET I e PLÁSTICO I. Também foi possível perceber a homogeneidade de alguns pontos de coleta e a heterogeneidade entre outros.

**Palavras-chave:** reciclagem, análise exploratória, análise de *cluster*, estatística.

## ABSTRACT

GOMES, Daniel Farias. **Statistical analysis of collection points for recyclable materials within the State of Rio de Janeiro. 2023.** 62 p. Monograph (Specialization in Applied Statistics “latu sensu”). Institute of Exact Sciences, Department of Mathematics, Federal Rural University of Rio de Janeiro, Seropédica, RJ, 2023.

Recycling materials plays an important role in depolluting and preserving the environment and acts as a solution for reducing carbon dioxide and renewing the Earth's natural resources. In view of this, we sought to find information and data on projects and practices within the recycling sector that can reveal, through statistical analysis, data behavior patterns that allow us to better understand market dynamics and possible strategies to increase reach. of reverse waste logistics. Therefore, the objective of this work was to identify which variables contribute most to the capture of recyclable materials within a collection point operation. The statistical analysis was based on the database of collection points for recyclable materials registered within the State of Rio de Janeiro between June 2022 and February 2023. Exploratory data analysis and multivariate analysis were applied as methodology, in particular analysis of hierarchical grouping. From the results of the exploratory analysis, it was possible to identify that in a collection point operation the predominant public is aged between 41 and 60 years old, 74% are women, that plastic is the most commercialized material in the operations, that glass and paper represent more than 55% of the purchased weight and are also the cheapest materials, with a value equal to or less than R\$ 0.20/kg. Furthermore, the median price of all materials is between R\$ 0.90/kg and R\$ 1.00/kg and 79% of the collector's revenue comes from light mass materials such as aluminum, PET I and PLASTIC I. It was also possible to perceive the homogeneity of some collection points and the heterogeneity among others.

**Keywords:** recycling, exploratory analysis, cluster analysis, statistics.



## LISTA DE TABELAS

Tabela 1. Descrição das variáveis qualitativas. ....	17
Tabela 2. Descrição das variáveis quantitativas. ....	17
Tabela 3. Resumo da análise das variáveis categóricas por frequência.....	23
Tabela 4. Resumo da análise das variáveis quantitativas por medida de tendência central e posição. ....	24
Tabela 5. Valor que cada material contribui para o catador/morador e a gravimetria. ....	33
Tabela 6. Gravimetrias comparadas. ....	34

## LISTA DE FIGURAS

Figura 1. Um dos pontos de coleta operados pela Cooperativa Recria. ....	15
Figura 2. Um dos pontos de coleta operados pela Cooperativa Recria recebendo moradores. ....	15
Figura 3. Uma das moedas sociais utilizadas no projeto. ....	16
Figura 4. Modelo de extrato digital de movimentações de dados do celular de um participante do projeto. ....	16
Figura 5. Descrição dos métodos da Estatística Descritiva utilizada. ....	18
Figura 6. Desenho geométrico do método de ligação simples. ....	19
Figura 7. Desenho geométrico do método de ligação completa. ....	20
Figura 8. Desenho geométrico do método da média das distâncias. ....	20
Figura 9. Desenho geométrico do método do centroide. ....	21
Figura 10. Desenho geométrico do método do centroide. ....	22
Figura 11. Total de frequência por categoria. ....	24
Figura 12. <i>Boxplot</i> das variáveis quantitativas. ....	25
Figura 13. Total de material por idade. ....	25
Figura 14. Total de material por sexo. ....	25
Figura 15. Total de material por ecoponto. ....	26
Figura 16. Total de material por categoria. ....	26
Figura 17. Total de material por mês. ....	27
Figura 18. Gráfico de dispersão entre preço e peso. ....	27
Figura 19. Gráfico de dispersão entre preço e valor. ....	28
Figura 20. Gráfico de dispersão entre peso e valor. ....	28
Figura 21. Gráfico de dispersão entre categoria e peso. ....	29
Figura 22. Gráfico de dispersão entre categoria e preço. ....	30
Figura 23. Gráfico de dispersão entre categoria e valor. ....	30
Figura 24. <i>Boxplot</i> de material coletado por peso e idade. ....	31
Figura 25. <i>Boxplot</i> de material coletado por peso e sexo. ....	32
Figura 26. <i>Boxplot</i> de preço de material por mês. ....	32
Figura 27. <i>Boxplot</i> de preço de material por categoria. ....	33
Figura 28. Dendrograma pelo método de ligação simples. ....	35
Figura 29. Dendrograma pelo método de ligação completa. ....	35
Figura 30. Dendrograma das médias da distância. ....	36
Figura 31. Dendrograma do centroide. ....	36
Figura 32. Dendrograma do método de Ward. ....	37
Figura 33. Diagrama de influências de variáveis. ....	41

## Sumário

<b>1. INTRODUÇÃO.....</b>	<b>12</b>
<b>2. MATERIAIS E MÉTODOS .....</b>	<b>14</b>
<b>2.1 Descrição dos dados.....</b>	<b>14</b>
<b>2.2 Métodos.....</b>	<b>17</b>
2.2.1 Distância Euclidiana .....	18
2.2.2 Método de ligação simples .....	19
2.2.3 Método de ligação completa.....	19
2.2.4 Método da média das distâncias .....	20
2.2.5 Método do centroide .....	21
2.2.6 Método de Ward .....	21
<b>3. RESULTADOS OBTIDOS E DISCUSSÃO .....</b>	<b>23</b>
<b>3.1 Análise exploratória de dados.....</b>	<b>23</b>
<b>3.1 Aplicação dos métodos de agrupamentos hierárquico .....</b>	<b>34</b>
<b>4. CONCLUSÃO.....</b>	<b>38</b>
<b>5. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>44</b>
<b>6. ANEXO .....</b>	<b>47</b>

## 1. INTRODUÇÃO

Segundo publicação do Banco Mundial (KAZA, S. et al, 2018) em 2016 o mundo gerou 242 milhões de toneladas de resíduos plásticos. O consumo de plásticos está em crescimento e uma grande quantidade vai parar nos oceanos por falta de um sistema adequado de gerenciamento dos resíduos domiciliares ou, melhor, resíduos sólidos urbanos.

Entende-se por resíduos sólidos urbanos, de acordo com a Lei Federal 12.305/2010, os resíduos domiciliares, resíduos de limpeza urbana de origem de varrição, limpeza de logradouros e vias públicas e outros serviços de limpeza, e resíduos gerados por estabelecimentos comerciais e prestadores de serviços.

Conforme relatório publicado pelo banco, estimou-se em 2016 a emissão de 1,6 bilhão de toneladas de dióxido de carbono proveniente de resíduos sólidos. Se não houver alguma melhoria até 2050 a emissão pode alcançar 2,6 bilhões de toneladas.

De acordo com a *World Wildlife Fund* – WWF (2023) no dia 02 de agosto de 2023 foi considerado o Dia da Sobrecarga da Terra, ou seja, foi o dia em que a humanidade usou de recursos naturais uma quantidade maior que a capacidade da Terra de regenerar em um ano. A humanidade atualmente consome 1,7 planeta em recursos naturais por ano, em outras palavras, 70% a mais da capacidade de renovação.

Importante destacar que em março de 2022, na Assembleia das Nações Unidas para o Meio Ambiente, todos os 193 Estados-membro da ONU decidiram acabar com a poluição plástica e um acordo jurídico está em curso.

Além disso, a Agenda 2030 dos Objetivos de Desenvolvimento Sustentável, no ODS 12 – Consumo e Produção Sustentáveis, estabeleceu a meta 12.5 de reduzir substancialmente a geração de resíduos por meio da prevenção, redução, reciclagem e reuso.

Salienta-se que o art. 7º da Política Nacional de Resíduos Sólidos, dentre os seus objetivos, fomenta o uso de matéria-prima e insumos derivados de materiais recicláveis e reciclados, a capacitação técnica continuada na área de resíduos sólidos e a integração dos

catadores de materiais recicláveis que envolve a responsabilidade pelo ciclo de vida de produtos.

Segundo a Revista Exame (2023) cada brasileiro gerou 381 kg de resíduos por ano, ou seja, mais de um quilo por dia. Na publicação da Federação da Indústria do Estado do Rio de Janeiro - FIRJAN (2023), Mapeamento de recicláveis pós-consumo no Estado do Rio de Janeiro, o Estado gera 7,5 milhões de toneladas de resíduos sólidos urbanos anualmente e 2 milhões de toneladas (27%) que tinha potencial de reciclagem foi enviado para aterros. Isto representa mais de dois bilhões de reais de matéria-prima que se voltasse para a indústria poderia criar mais de 31,9 mil empregos diretos e indiretos.

Conforme dados da Pesquisa Científica BNDES FEP Nº 02/2010 (JUCÁ, J. F. T. et al, 2014) para cada 1 tonelada de material reciclado, mantêm-se 10 empregos em toda a cadeia produtiva da reciclagem. De acordo com o Instituto de Pesquisa Econômica Aplicada - IPEA (2017), o Brasil gera em torno de 160 mil toneladas diárias de resíduos e 30% a 40% são passíveis de reciclagem. Contudo apenas 4% a 5% desses materiais são enviados para reciclagem. Também afirma o Atlas Brasileiro da Reciclagem (ANCAT, 2022) que 90% do que é reciclado atualmente vem das mãos de catadores.

Bom é dizer que no 1º Congresso Internacional de Resíduos Sólidos em Búzios 2023 o coordenador geral de resíduos sólidos do Ministério do Meio Ambiente, Alberto da Rocha, afirmou que a meta do Governo é aumentar a taxa de reciclagem para 40% das embalagens.

Diante destes cenários torna-se importante uma investigação mais profunda de dados relacionados ao mercado da reciclagem para entender o seu funcionamento e perceber possibilidades disponíveis para diminuir o volume de resíduos jogado em aterros e no meio ambiente.

O objetivo deste trabalho foi realizar uma análise dos dados sobre reciclagem e identificar quais as variáveis que mais contribuem para captação de materiais recicláveis dentro de uma operação de um ponto de coleta.

## **2. MATERIAIS E MÉTODOS**

### **2.1 Descrição dos dados**

Por oportuno, para obtenção de dados, solicitou-se a disponibilidade de informações de algumas operações da cooperativa RECRIA RECICRIA – COOPERATIVA DE TRABALHO E VALORIZAÇÃO DE CATADORAS E CATADORES DE MATERIAIS RECICLÁVEIS ORGANIZADOS E AUTÔNOMOS LTDA, localizada no bairro de Realengo, Rio de Janeiro-RJ.

A Cooperativa Recria, criada desde 2018, administrou e operacionalizou um projeto de dez pontos de coleta de materiais recicláveis em diferentes bairros e comunidades do Estado do Rio de Janeiro durante alguns meses entre o período de junho de 2022 à agosto de 2023 (Figuras 1 e 2). Estes locais recebiam materiais recicláveis de moradores e catadores de rua e, em contrapartida, estas pessoas eram pagas em moeda digital de acordo com o peso de cada tipo de material e o valor recebido por elas era usado nos comércios locais cadastrados. Um projeto que monetizava os resíduos recicláveis para incentivar o engajamento das pessoas, promover educação ambiental, distribuir renda e ampliar o índice de reciclagem dentro do Estado (Figuras 3 e 4).



**Figura 1.** Um dos pontos de coleta operados pela Cooperativa Recria.

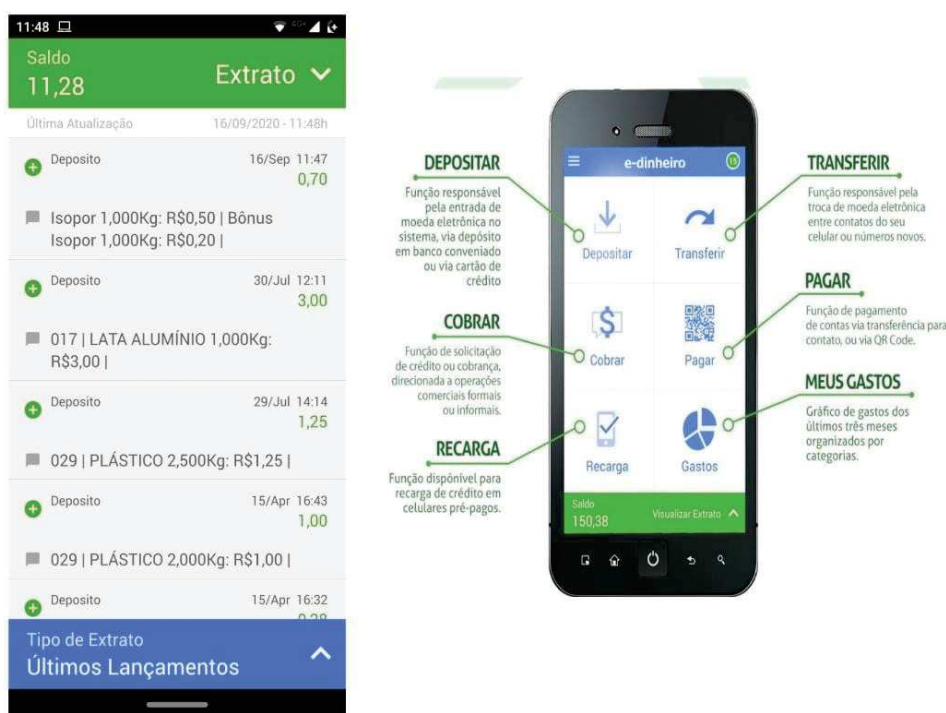


**Figura 2.** Um dos pontos de coleta operados pela Cooperativa Recria recebendo moradores.





**Figura 3.** Uma das moedas sociais utilizadas no projeto.



**Figura 4.** Modelo de extrato digital de movimentações de dados do celular de um participante do projeto.

A operação do projeto dos dez pontos de coleta gerou uma base de dados que foi disponibilizada pela Cooperativa Recria de forma anonimizada em respeito da Lei Geral de Proteção de Dados – LGPD. A base de dados utilizada possui 28848 linhas, 8 variáveis e registra as operações do mês de junho de 2022 à fevereiro de 2023. Foram ao todo coletados 140 toneladas de material e 1185 moradores e catadores envolvidos que receberam com a venda



dos materiais recicláveis a quantia média de 23 reais por mês. As variáveis utilizadas foram: idade, sexo, localização, categoria do material, mês, peso (kg), preço (R\$/kg) e valor da venda do material (R\$) (Tabela 1 e 2). Considera-se peso como material levado pelo morador ou catador que foi registrado em quilos na balança e o valor como o produto do peso e preço.

**Tabela 1.** Descrição das variáveis qualitativas.

<b>Tabela 1: Descrição das variáveis qualitativas</b>	
Variável	Definição
1 Idade	Variável qualitativa nominal simples categorizada por idade com intervalo de vinte anos.
2 Sexo	Variável qualitativa nominal simples categorizada por sexo masculino e feminino. Refere-se as pessoas que levavam o material até o ponto de coleta.
3 Localização	Variável qualitativa nominal simples categorizada por ponto de coleta.
4 Categoria	Variável qualitativa nominal simples categorizada por tipo de material.
5 Mês	Variável qualitativa nominal simples categorizada por mês.

Fonte: Elaborado pelo autor.

**Tabela 2.** Descrição das variáveis quantitativas.

<b>Tabela 2: Descrição das variáveis quantitativas</b>	
Variável	Definição
1 Peso (kg)	Variável quantitativa aleatória contínua informada em quilo. Refere-se ao peso obtido de resíduos por tipo ao final de cada transação.
2 Preço (R\$/kg)	Variável quantitativa aleatória discreta informada em decimal. Contém o preço do material comprado.
3 Valor (R\$)	Variável quantitativa aleatória contínua informada em valor em reais.

Fonte: Elaborado pelo autor.

## 2.2 Métodos

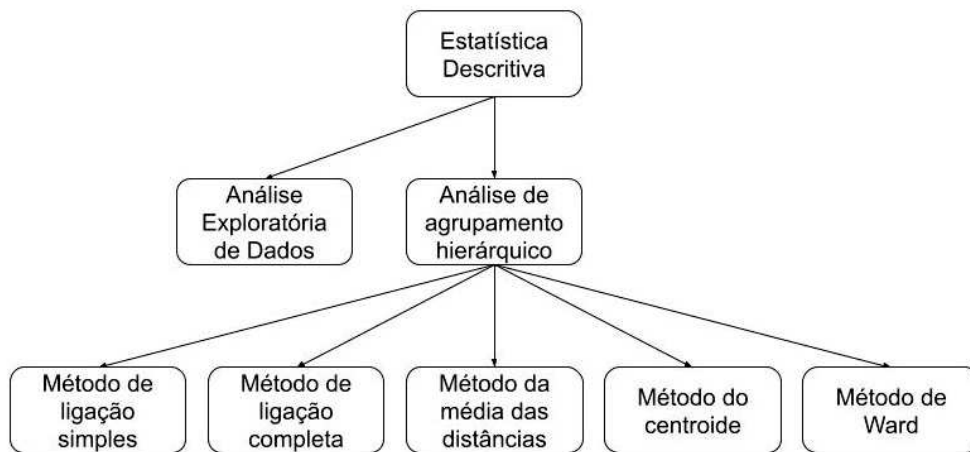
Primeiramente, foi realizada uma análise preliminar da base de dados para seleção de variáveis. Em seguida, foi realizada uma análise exploratória de dados para as variáveis selecionadas. Segundo Triola (1999), análise exploratória de dados tem o intuito de explorar os dados em um nível preliminar com o objetivo de entender a natureza dos dados e ver o que eles nos revelam. Na análise exploratória de dados são usadas as ferramentas das Estatística Descritiva que são as representações gráficas, medidas de tendência central, medidas de variação e medidas de posição.

Entre as representações gráficas se pode citar o histograma, tabela de frequências, gráfico de barra, *boxplot*, diagrama de dispersão entre outros. Já as medidas de tendência central se têm a média aritmética, a mediana e a moda. As medidas de variação trazem a amplitude, o desvio-padrão e a variância. E as medidas de posição tratam do escore padronizado, quartis, decis e percentis.

Por fim, foi utilizado a análise de agrupamento hierárquico sobre os dez pontos de coleta existentes com as variáveis peso, preço e valor do material. Na análise de agrupamento hierárquico, os objetos são aglomerados por suas associações e semelhanças, representados graficamente em dendrograma. Neste trabalho, foram utilizados os métodos de agrupamento hierárquico de ligação simples, ligação completa, média das distâncias, centroide e método de Ward.

A análise de agrupamento tem o intuito de agrupar as observações em agrupamentos (ou *cluster*) utilizando um índice de similaridade ou proximidade entre cada par de observações. Uma medida conveniente de proximidade é a distância entre duas observações. Neste trabalho foi utilizado a distância euclidiana.

Na Figura 5 abaixo está a descrição em fluxograma dos métodos utilizados para melhor compreensão.



**Figura 5.** Descrição dos métodos da Estatística Descritiva utilizada.

### 2.2.1 Distância Euclidiana

A distância euclidiana entre duas observações plurais dimensionais,  $\mathbf{x}^T = (x_1, x_2, \dots, x_p)$  e  $\mathbf{y}^T = (y_1, y_2, \dots, y_p)$ , é dada por

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}.$$

Assim,  $\mathbf{x}^T$  e  $\mathbf{y}^T$  são as representações de vetores colunas transpostos para linhas e a fórmula  $d(\mathbf{x}, \mathbf{y})$  representa a distância euclidiana entre estes dois vetores ou graficamente uma linha reta entre os dois pontos individuais que simboliza a distância.

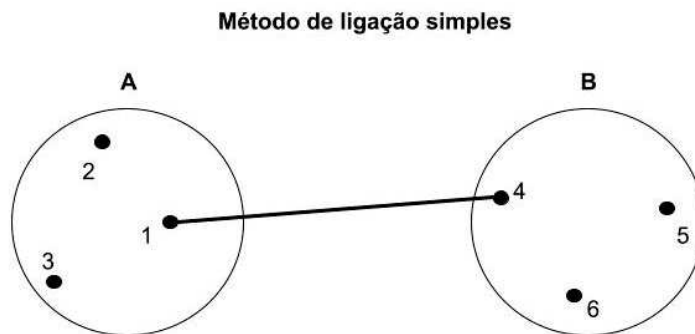
### 2.2.2 Método de ligação simples

No método de ligação simples, a distância entre dois *clusters* A e B é definida como a distância mínima entre um ponto em A e um ponto em B por

$$d(A, B) = \min \{d(y_i, y_j); y_i \in A \text{ e } y_j \in B\},$$

em que  $d(y_i, y_j)$  é a distância euclidiana. Esta abordagem também é chamada de método do vizinho mais próximo. Ela calcula a distância entre todos os pontos individuais  $(y_i, y_j)$ .

Na Figura 6 abaixo está a representação gráfica do método em que a linha preta representa a distância euclidiana que começa pelos pontos individuais mais próximos entre os grupos.



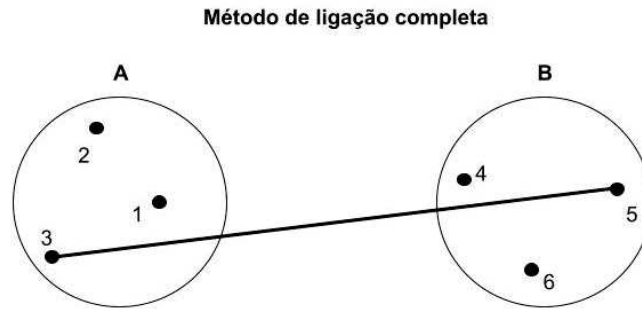
**Figura 6.** Desenho geométrico do método de ligação simples. Fonte: Adaptado de Oliveira (2022).

### 2.2.3 Método de ligação completa

No método de ligação completa, também chamada de método do vizinho mais distante, a distância entre dois *clusters* A e B é definida como a distância máxima entre um ponto em A e um ponto em B por

$$d(A, B) = \max \{d(y_i, y_j); y_i \in A \text{ e } y_j \in B\}.$$

Na Figura 7 está a representação do método de ligação completa. Como se pode ver ocorre um processo inverso do método de ligação simples, pois começasse o cálculo pelos pontos mais distantes.



**Figura 7.** Desenho geométrico do método de ligação completa. Fonte: Adaptado de Oliveira (2022).

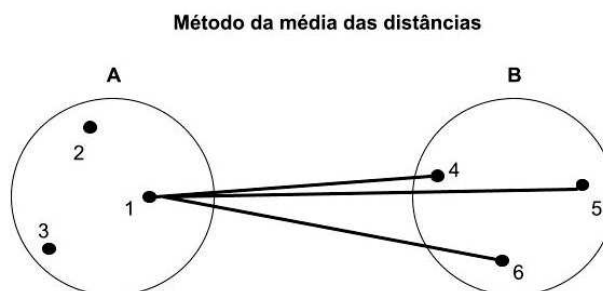
#### 2.2.4 Método da média das distâncias

O Método da média das distâncias é o método que usa os coeficientes de similaridade ou dissimilaridade médios dos objetos, sendo semelhantes os com as médias mais próximas.

$$d(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j).$$

Sendo  $d(A, B)$  a distância média entre os conjuntos A e B,  $n_A$  e  $n_B$  os números de pontos do conjunto A e conjunto B, esta equação resulta na média das distâncias entre todas as partições de  $y_i$  e  $y_j$ .

Na Figura 8 se verifica que um objeto do grupo A se interage com todos os objetos do grupo B para o cálculo da média das distâncias.



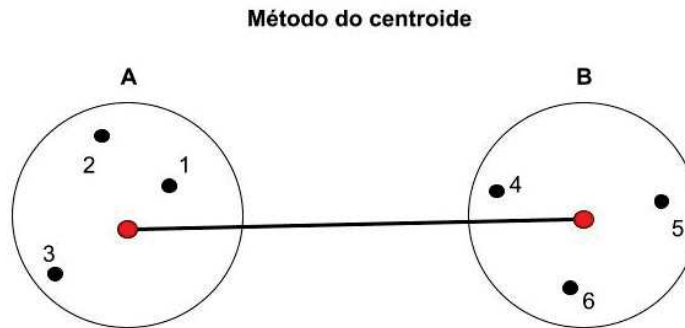
**Figura 8.** Desenho geométrico do método da média das distâncias. Fonte: Adaptado de Oliveira (2022).

### 2.2.5 Método do centroide

O método do centroide trabalha com os vetores das médias de dois grupos, ou seja, a distância euclidiana entre os centroides

$$d(A, B) = d(\bar{y}_A, \bar{y}_B).$$

Nota-se  $\bar{y}_A$  e  $\bar{y}_B$  representam as médias ou centroides dos pontos no conjunto A no conjunto B respectivamente. Na Figura 9 é possível visualizar a reta partindo do centroide ou ponto central do grupo.



**Figura 9.** Desenho geométrico do método do centroide. Fonte: Adaptado de Oliveira (2022).

### 2.2.6 Método de Ward

Segundo Ward (1963 apud EMBRAPA, 2020) o processo de classificação ocorre quando as parcelas são progressivamente reunidas dentro de grupos, minimizando uma função objetiva para cada etapa da fusão. É um método aglomerativo que usa a variância como critério para unir os grupos mais homogêneos, considerando que a variância é a média dos quadrados das diferenças de cada ponto dentro do grupo em relação ao centroide ou a média do grupo. Na equação abaixo, por exemplo, deve-se considerar AB como um grupo formado pela combinação de um grupo A e um grupo B. Sendo  $SSE_A$ ,  $SSE_B$  e  $SSE_{AB}$  a soma dos erros quadrados dos grupos A, B e AB com o objetivo de diminuir a dispersão ou variabilidade dos pontos.

$$SSE_A = \sum_{i=1}^{n_A} (y_i - \bar{y}_A)^T (y_i - \bar{y}_A),$$

$$SSE_B = \sum_{i=1}^{n_B} (y_i - \bar{y}_B)^T (y_i - \bar{y}_B),$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \bar{y}_{AB})^T (y_i - \bar{y}_{AB})$$

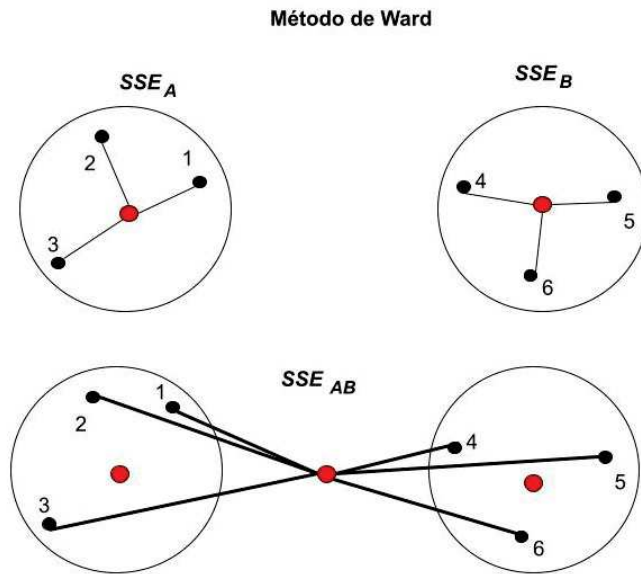
Em que  $y_{AB} = (n_A y_A + n_B y_B) / (n_A + n_B)$  descreve a média ponderada entre os centroides, sendo  $(n_{AB} = n_A + n_B)$  o número total de objetos depois da fusão dos grupos. O Método de Ward minimiza a perda associada a cada agrupamento ou o crescimento de  $SSE$  e pode ser descrito como

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B).$$

Sendo  $I_{AB}$  como a medida para avaliar a melhoria no resultado com a redução da variação da soma dos erros quadrados ao unir os grupos A e B em apenas um grupo. É preciso verificar se essa redução é significativa ou não.

O método de Ward tem um forte apelo estatístico, apresenta alta homogeneidade interna como o método de ligação completa e revela bons resultados em diversas distâncias, contudo é sensível à presença de *outliers*.

Na Figura 10 representa em desenho a operação com o método descrito acima.



**Figura 10.** Desenho geométrico do método do centroide. Fonte: Adaptado de Oliveira (2022).

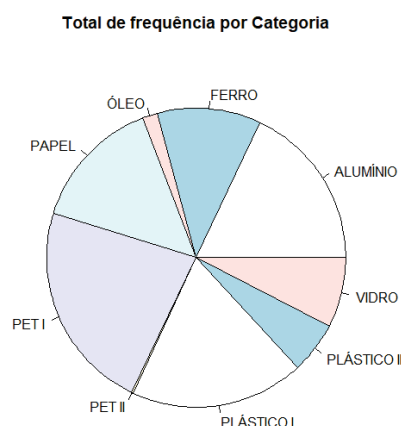
### 3. RESULTADOS OBTIDOS E DISCUSSÃO

#### 3.1 Análise exploratória de dados

Na análise exploratória de dados foi possível identificar uma predominância das variáveis categóricas “41-60” como o intervalo de idade de 41 à 60 anos que registra 15.994 operações de compra (Tabela 3). Importante dizer que se optou pelo intervalo de 20 anos para classificar as idades por ser um período usual para separar duas gerações e facilitar a visualização de grupos. Além disso, na variável Sexo as mulheres são a grande maioria, representando 74% das transações. O material PET I é o que foi mais presente nas transações (Figura 11) e o mês de julho 2022 foi o mais atuante em frequência de operações de compra de materiais recicláveis.

**Tabela 3.** Resumo da análise das variáveis categóricas por frequência.

<b>Tabela 3: Resumo da análise das variáveis categóricas por frequência</b>										
	Idade	Quantidade por idade	Sexo	Quantidade por sexo	Localização	Quantidade por localização	Categoria	Quantidade por categoria	Mês	Quantidade por mês
1	18-40	6717	F	21268	PONTO K	8972	PET I	6523	JUL	4564
2	41-60	15994	M	7580	PONTO A	7288	PLÁSTICO I	5424	AGO	4249
3	61 ou mais	6137			PONTO J	3271	ALUMÍNIO	5185	SET	4019
4					PONTO F	3011	PAPEL	4151	OUT	3844
5					PONTO C	2366	FERRO	3249	NOV	2783
6					(OUTROS)	3940	(OUTROS)	4316	(OUTROS)	9389
<b>Fonte: Elaborado pelo autor.</b>										



**Figura 11.** Total de frequência por categoria.

Importante destacar que deve haver uma diferenciação entre frequência e peso. Na Tabela 3 e na Figura 11, as informações são sobre as variáveis com as maiores frequências de operações ou transações, o que não quer dizer que também sejam as variáveis com os maiores pesos.

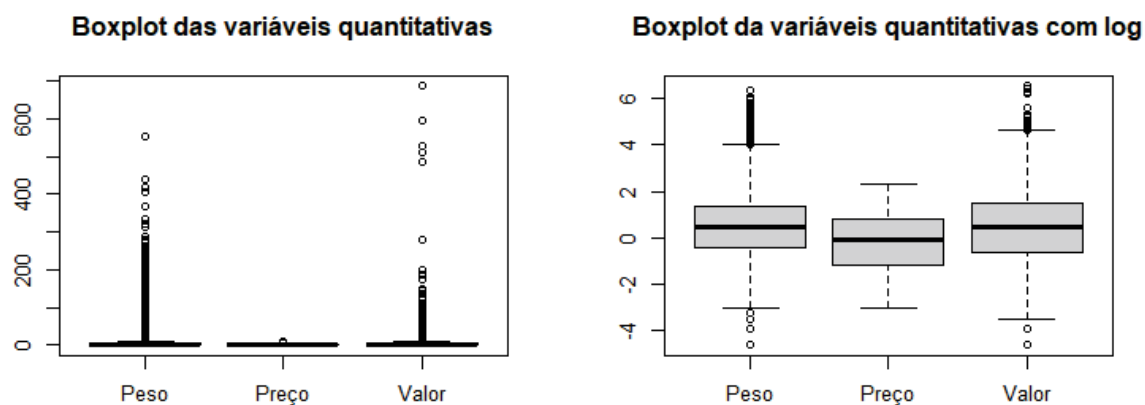
Também se aplicou a análise sobre as variáveis quantitativas e se pode perceber que há uma amplitude grande de valores nas variáveis peso, preço do material e o valor da compra, o que afasta a média da mediana. A Tabela 4 mostra algumas outras medidas estatísticas para as variáveis do estudo. A Figura 12 apresenta o *boxplot* de cada variável analisada. A distribuição da quantidade de material por faixa etária mostra uma predominância entre 41 à 60 anos (Figura 13). Os dados analisados mostraram que o sexo feminino apresentou a maior taxa de coleta de material reciclado (Figura 14).

**Tabela 4.** Resumo da análise das variáveis quantitativas por medida de tendência central e posição.

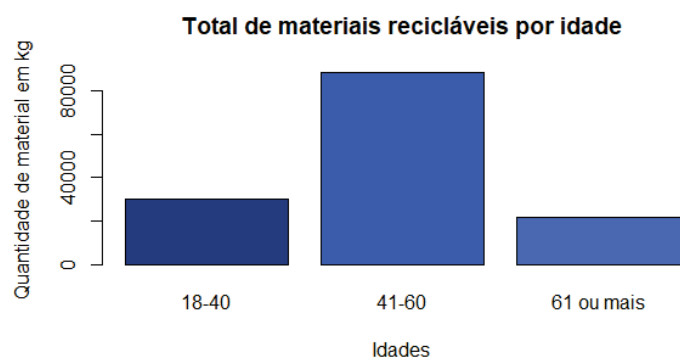
<b>Tabela 4: Resumo da análise das variáveis quantitativas por medida de tendência central e medida de posição</b>				
	Medidas	Peso (kg)	Preço (R\$/kg)	Valor (R\$)
1	Mínimo	0.010	0.00	0.00
2	1º quartil	0.650	0.30	0.52
3	Mediana	1.555	0.90	1.61
4	Média	4.859	1.81	4.08
5	3º quartil	3.850	2.30	4.37
6	Máximo	555.650	10.00	687.60

**Fonte:** Elaborado pelo autor.

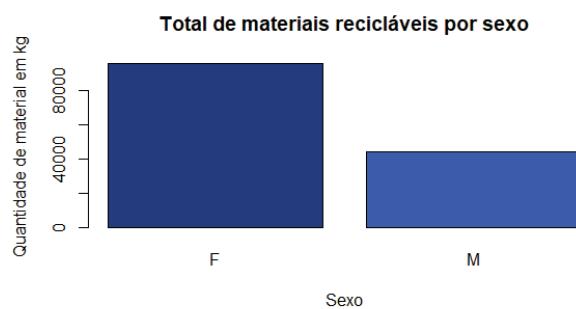




**Figura 12.** *Boxplot* das variáveis quantitativas.

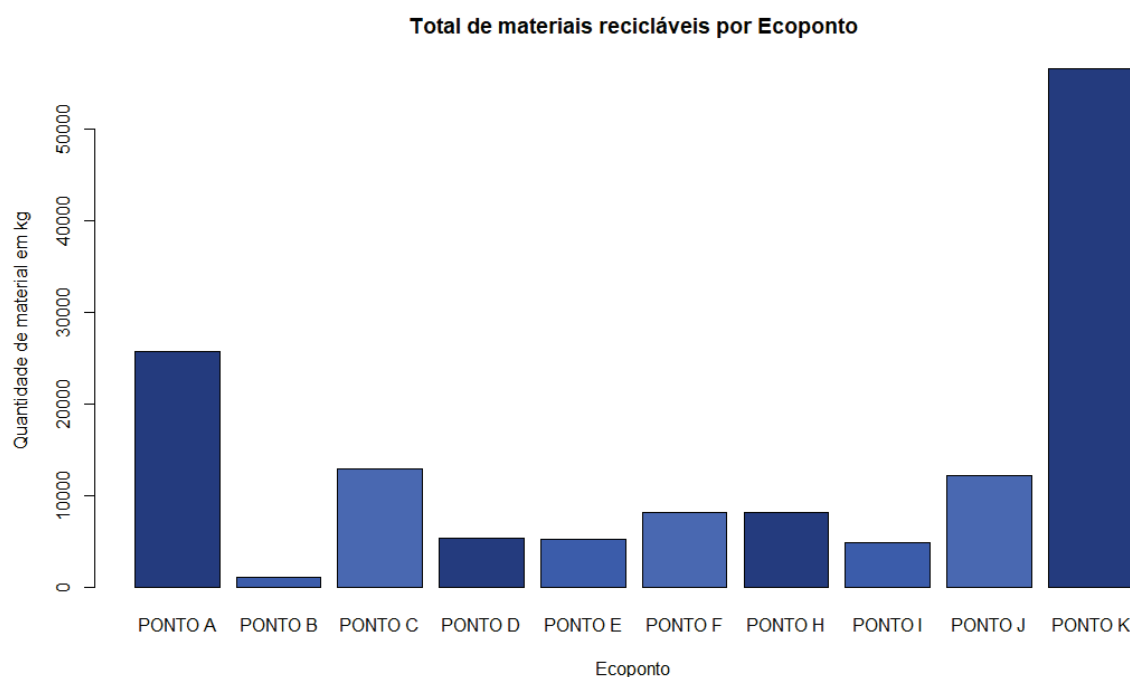


**Figura 13.** Total de material por idade.



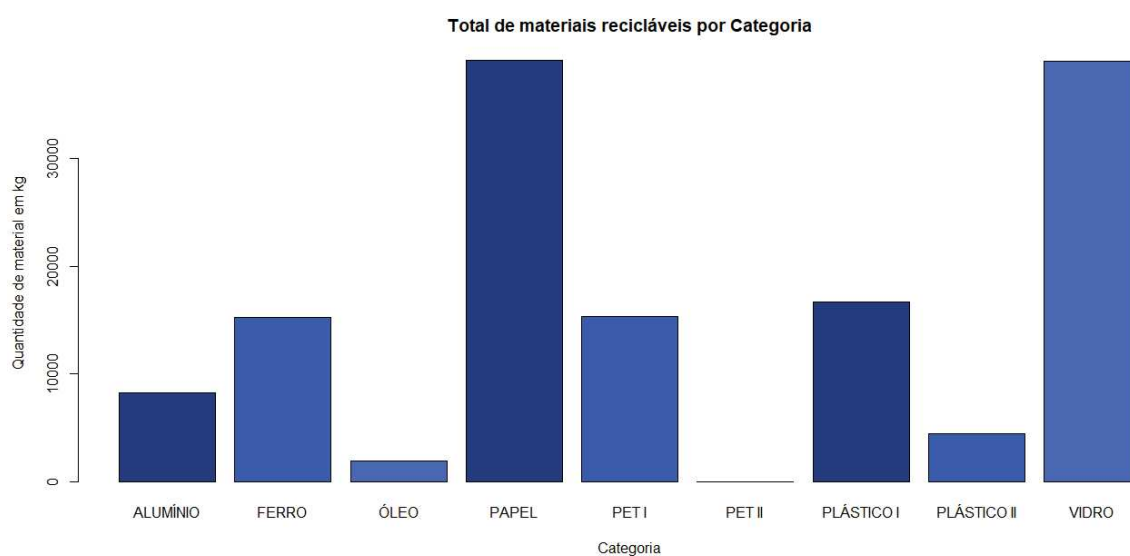
**Figura 14.** Total de material por sexo.

O ponto K foi o local com maior captação de materiais recicláveis da população, alcançando a marca acima de cinquenta toneladas durante nove meses (Figura 15).



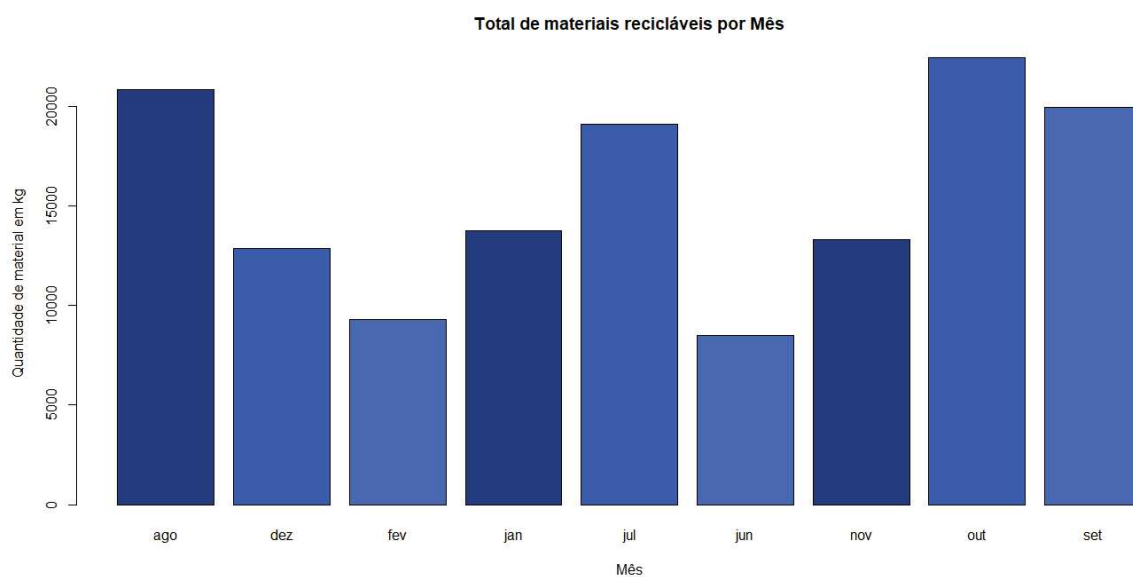
**Figura 15.** Total de material por eco ponto.

Já na análise dos pesos por categoria se notou que o papel e o vidro são os materiais predominantes, representando mais de 55% do total de peso (Figura 16). Por outro lado, o material PET II teve uma baixa coleta.



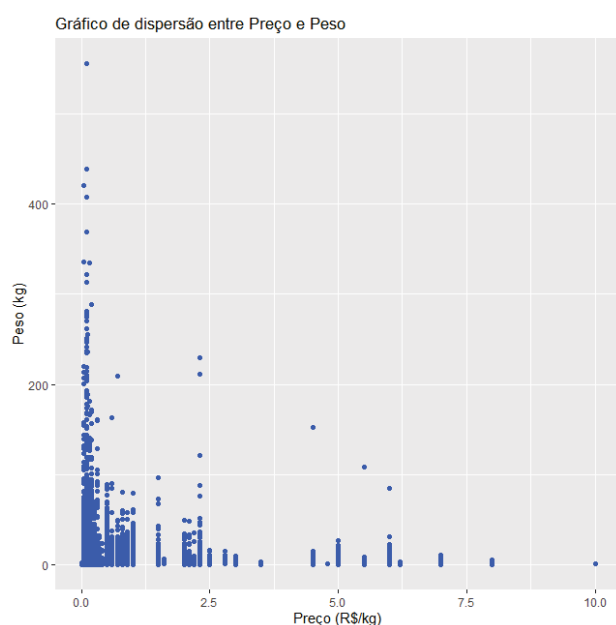
**Figura 16.** Total de material por categoria.

A Figura 17 mostra a distribuição de material coletado por mês e o mês de outubro de 2022 com maior volume de material no conjunto de dados analisado.



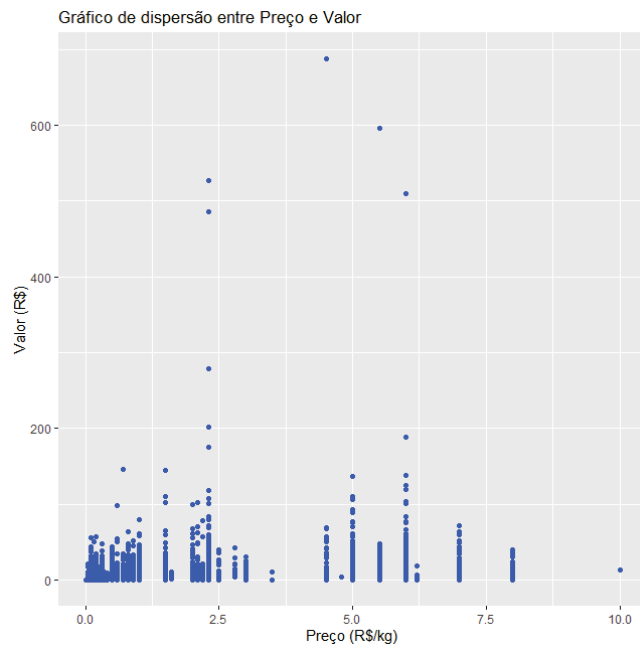
**Figura 17.** Total de material por mês.

A Figura 18 apresentou o gráfico de dispersão entre a variável peso e preço. Pelo gráfico, pode-se observar que existe uma associação linear baixa e negativa (-0,16) entre as variáveis, ou seja, quanto maior o preço pago no material menor o seu peso e quanto menor o preço maior o volume de operações com peso maior. Talvez a escassez maior de materiais compostos por matéria-prima mais cara nas embalagens de produtos de bens de consumo seja uma causa deste fato. Assim, pode-se concluir que quanto maior a matéria-prima disponível no mercado para embalagens, menor o preço comercial da embalagem reciclável para revenda.



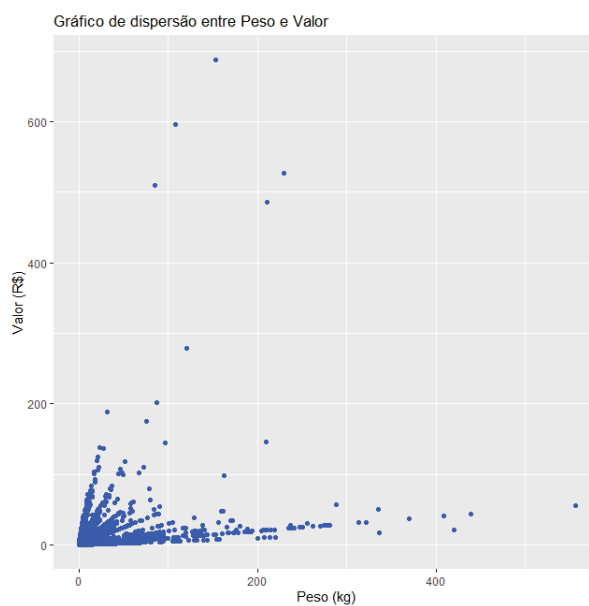
**Figura 18.** Gráfico de dispersão entre preço e peso.

A Figura 19 mostra o gráfico de dispersão entre as variáveis preço e o valor de compra do material. Pela figura, não há nenhuma correlação entre as variáveis.



**Figura 19.** Gráfico de dispersão entre preço e valor.

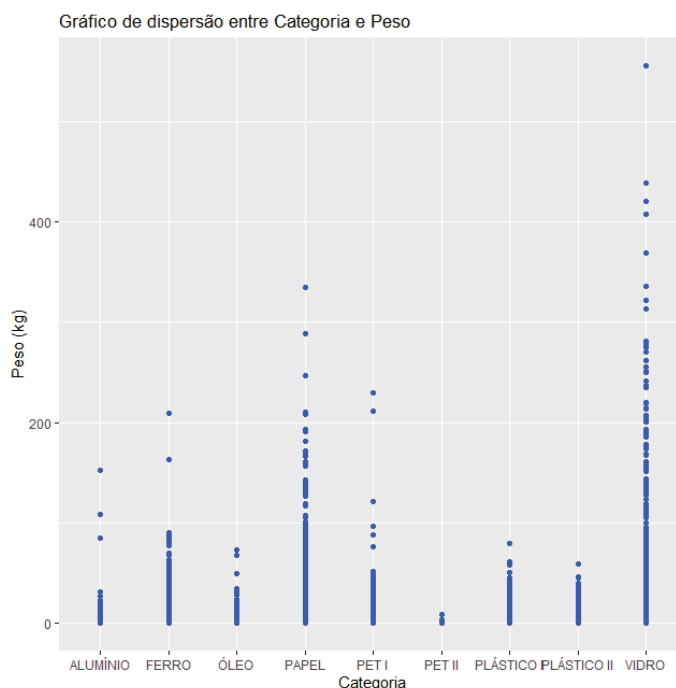
A Figura 20 apresenta o gráfico de dispersão entre as variáveis peso e valor do material. Pode-se observar que quanto menor o peso, menor o valor pago, mas em um peso crescente são poucas as operações com valor crescente, o que aproxima mais da neutralidade entre as variáveis.



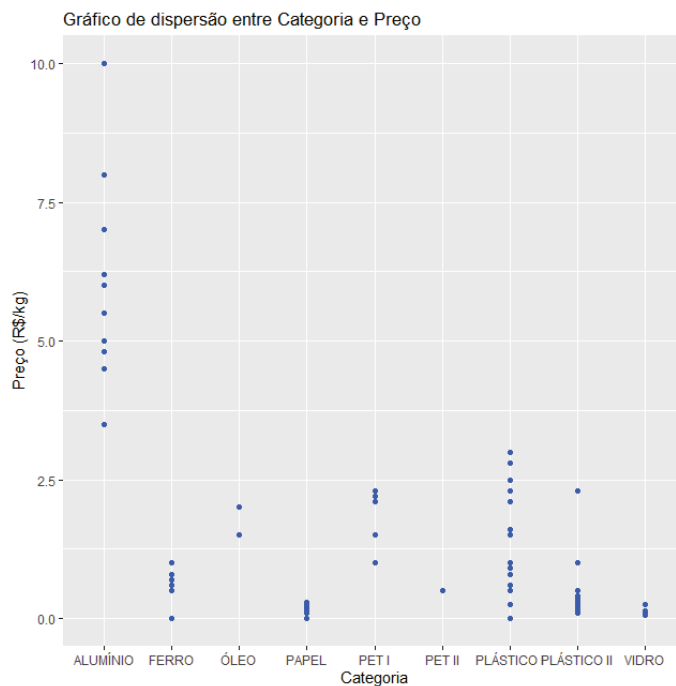
**Figura 20.** Gráfico de dispersão entre peso e valor.

A Figura 21, apesar de similar a Figura 16, ela consegue mostrar a distribuição de cada material reciclável e o peso do conjunto de dados analisado, se os pesos das operações estão concentrados em valores baixos ou altos. Pode-se observar nesta figura que os materiais papel e vidro estão com o maior peso e o material PET II com o menor peso, mas também é possível perceber que as operações, em sua maioria, se concentram em pesos abaixo de duzentos quilos.

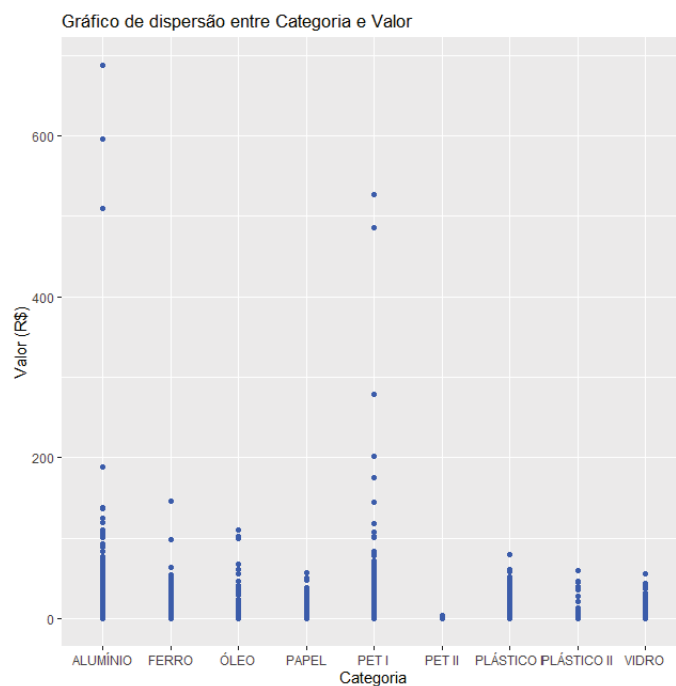
Ademais, como o vidro e o papel são os materiais mais baratos, isso fortalece a associação identificada na Figura 18, ou seja, os materiais mais baratos são os mais abundantes no mercado e os mais presentes nas embalagens dos produtos de bens de consumo. A Figura 22 apresenta a distribuição entre o material e o preço. Essa figura mostra que as operações são mais concentradas nos materiais mais baratos. A Figura 23 mostra a distribuição do material reciclado e o valor de compra. Os maiores valores de compra estão relacionados com o material de alumínio.



**Figura 21.** Gráfico de dispersão entre categoria e peso.



**Figura 22.** Gráfico de dispersão entre categoria e preço.

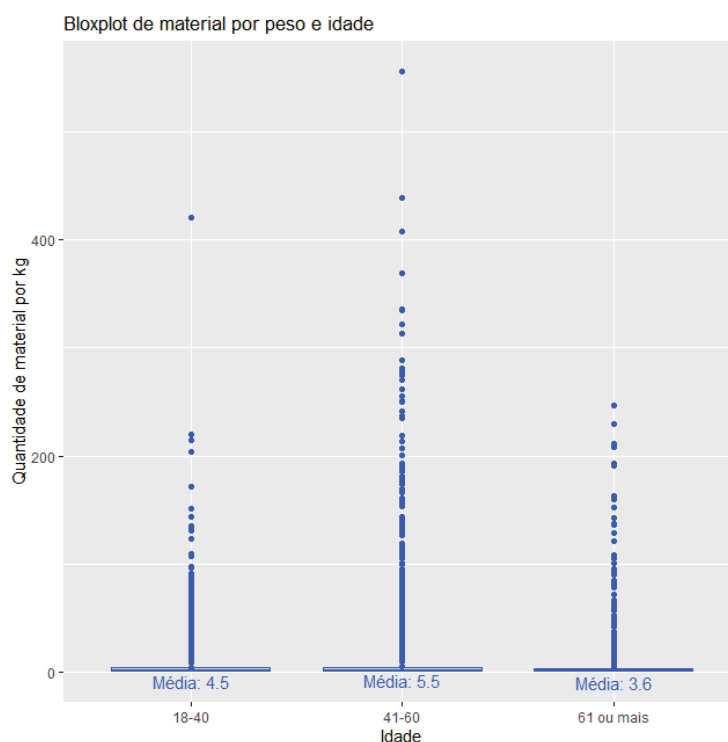


**Figura 23.** Gráfico de dispersão entre categoria e valor.

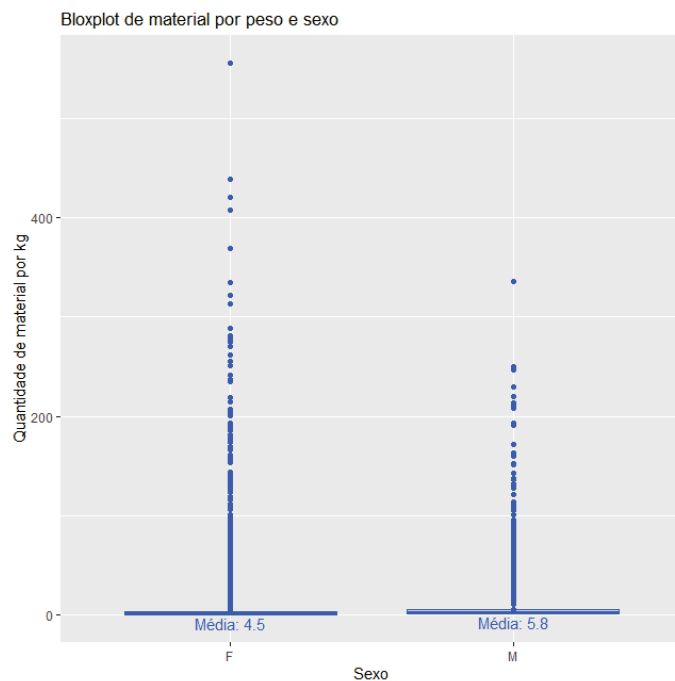
A Figura 24 mostra o gráfico de *boxplot* do peso do material coletado para as seguintes classes de idade. O maior público atendido nos pontos de coleta é de 41 a 60 anos e também é a faixa etária com maior peso médio por operação, 5,5 quilos, 53% maior que as pessoas da faixa de mais de 61 anos (3,6 quilos) e 22% maior que as pessoas da faixa de 18 à 40 anos (4,5

quilos), ou seja, as pessoas de 41 a 60 anos são as mais engajadas com a iniciativa de revender suas embalagens para a reciclagem. Nesta figura, pode-se observar um conjunto considerável de valores discrepantes para todas as classes de idade.

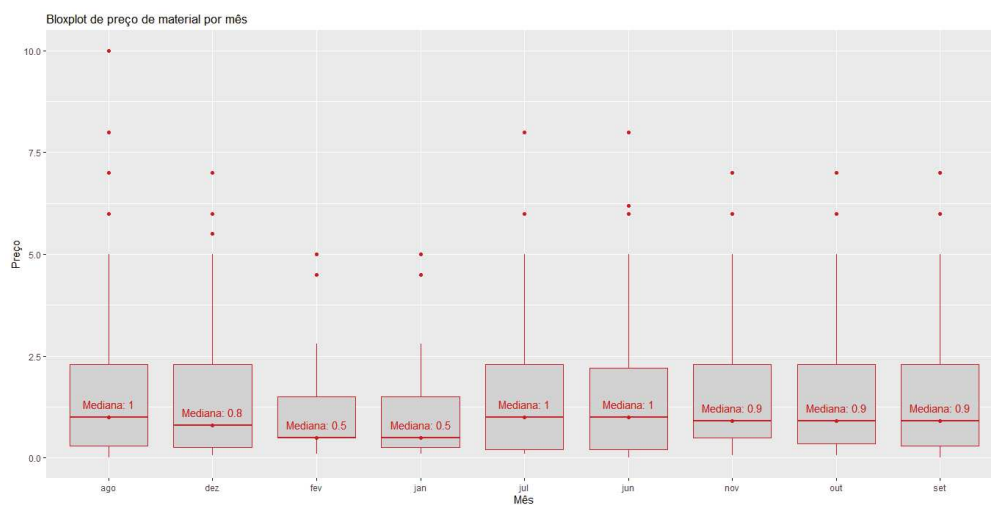
O gráfico de *boxplot* mostra um maior engajamento do público feminino com 74% das operações (Figura 25). Contudo o peso médio por transação foi 28% menor que o peso médio do público masculino. A Figura 26 mostra o gráfico de *boxplot* mensal com o preço mensal mediano de todos os materiais recicláveis comprados. Pode-se verificar que a mediana dos preços dos materiais ao longo de nove meses se manteve praticamente estável, variando principalmente entre 0,90 centavos por quilo a 1 real por quilo.



**Figura 24.** *Boxplot* de material coletado por peso e idade.



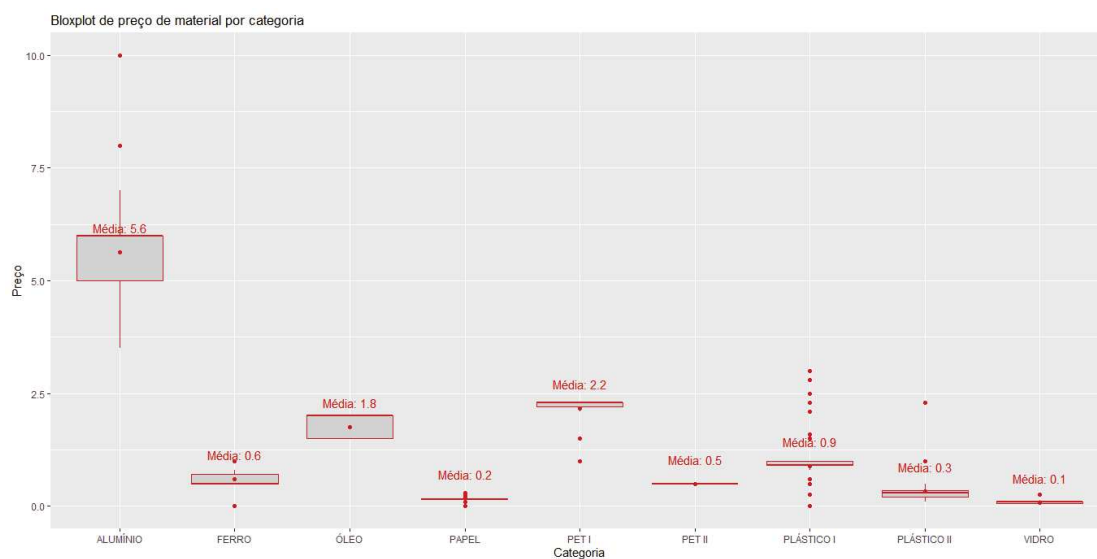
**Figura 25.** *Boxplot* de material coletado por peso e sexo.



**Figura 26.** *Boxplot* de preço de material por mês.

A Figura 27 apresenta os gráficos de *boxplot* para cada material reciclável deste estudo. Essa figura mostra que o preço dos materiais papel (R\$ 0,20/kg) e vidro (R\$ 0,10/kg) são os com menor preço médio e os materiais alumínio (R\$ 5,60/kg) e PET I (R\$ 2,20/kg) são os com maior preço médio.





**Figura 27.** Boxplot de preço de material por categoria.

Depois de todo o exposto, se pode concluir que os materiais com menor peso, mas com maior preço, são os que mais contribuem financeiramente para o morador ou catador. Isto pode ser visualizado na Tabela 5, a qual traz a participação financeira por tipo de material e a gravimetria dos resíduos. Importante dizer que a gravimetria é o percentual do peso de cada material sobre o peso total de todos eles. Observa-se na Tabela 5 que 79% da receita tem origem nos materiais com menos massa, porém sua matéria-prima tem mais valor comercial. E os materiais com mais massa (vidro e papel) só contribuem com 9% da receita.

**Tabela 5.** Valor que cada material contribui para o catador/morador e a gravimetria.

<b>Tabela 5: Valor que cada material contribui para o catador/morador e a gravimetria</b>			
	Material	Para cada 100 reais este material contribui com	Gravimetria da Cooperativa Recria
1	ALUMÍNIO	39	6%
2	PET I	28	11%
3	PLÁSTICO I	12	12%
4	FERRO	8	11%
5	PAPEL	6	28%
6	VIDRO	3	28%

**Fonte:** Elaborado pelo autor.

Ademais, com os dados disponíveis, foi possível construir a Tabela 6 em que se traz as gravimetrias comparadas da fração reciclável da Cooperativa Recria (2022-2023), da Companhia Municipal de Limpeza Urbana da Cidade do Rio de Janeiro – COMLURB (2021) e do Compromisso Empresarial para Reciclagem – CEMPRE (2019). Nota-se que nos plásticos há uma maior discrepância com a informação da COMLURB, uma diferença de 83% a mais, ou seja, na operação da Cooperativa existem mais plásticos que foi informado na pesquisa da COMLURB. No papel também a diferença é maior em relação aos dados da COMLURB, ou seja, existe 93% de diferença a menos no peso do papel. E no vidro existe diferença alta de gravimetria com a COMLURB e o CEMPRE, diferenças de 115% e 250% a mais, há mais vidro em peso no resíduo sólido urbano que a pesquisa que as instituições conseguiram identificar.

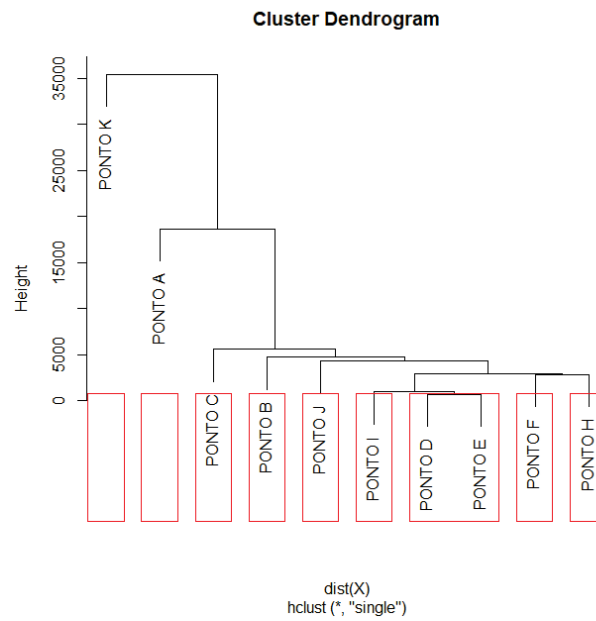
**Tabela 6.** Gravimetrias comparadas.

<b>Tabela 6: Gravimetrias comparadas</b>				
	Material	Gravimetria da Cooperativa Recria (2022- 2023)	Gravimetria da Comlurb (2021)	Gravimetria do CEMPRE (2019)
1	METAL	17%	16%	10%
2	PLÁSTICOS	33%	18%	42%
3	PAPEL	28%	54%	41%
4	VIDRO	28%	13%	8%

**Fonte:** Elaborado pelo autor.

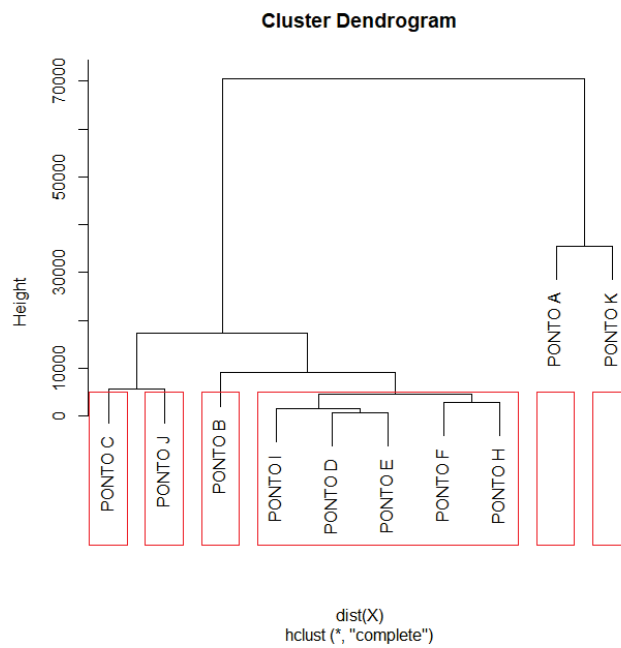
### 3.1 Aplicação dos métodos de agrupamentos hierárquico

O método de agrupamento foi realizado sobre os dez pontos de coleta e, como se pode ver no dendrograma (Figura 28), os pontos de coleta com resultados mais próximos são os pontos D e E, formando um único grupo. Já os outros locais formam grupos independentes, sendo o ponto K de maior peso. Assim se pode dizer que a técnica formou nove agrupamentos ao todo.



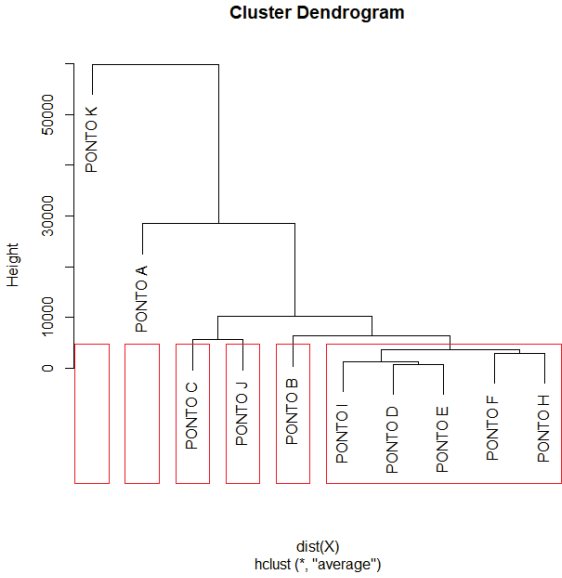
**Figura 28.** Dendrograma pelo método de ligação simples.

A Figura 29 mostra o dendrograma obtido pelo método de ligação completa. Esse método mostra a formação de seis grupos de dados, sendo presentes os pontos D, E, F, H e I em um único grupo por semelhança.



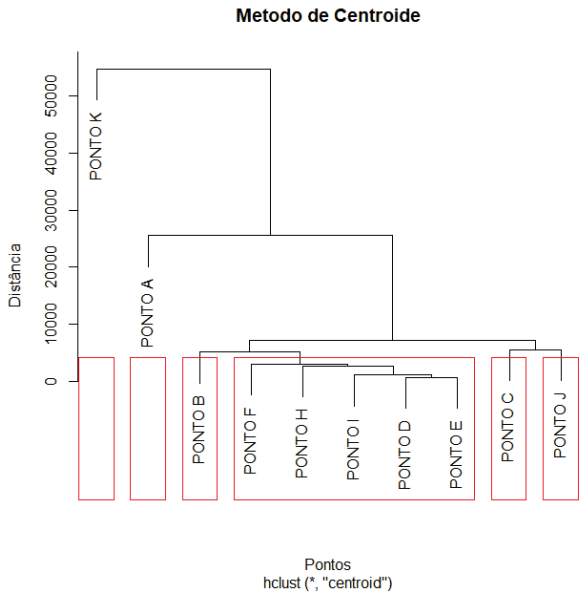
**Figura 29.** Dendrograma pelo método de ligação completa.

A Figura 30 mostra o dendrograma obtido via método da diferença de médias. Neste dendrograma são formados seis grupos e os pontos com maior associação entre si são mantidos os mesmos do caso anterior.



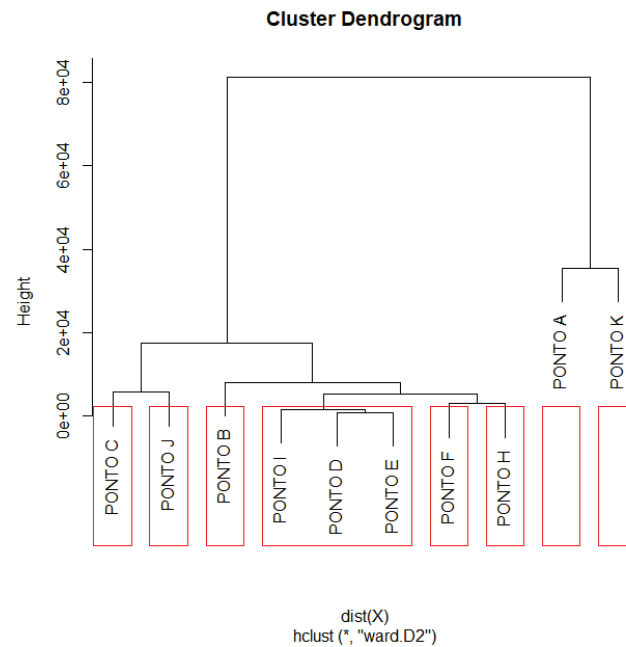
**Figura 30.** Dendrograma das médias da distância.

A Figura 31 apresenta o dendrograma obtido pelo método de agrupamento do centroide. Neste método também mantém a similaridade entre os pontos D, E, F, H e I. Ao todo também são seis grupos formados.



**Figura 31.** Dendrograma do centroide.

A Figura 32 apresenta o dendrograma obtido pelo método de Ward. Na aplicação sobre os dados formou-se oito grupos e o grupo que tinha cinco pontos semelhantes no método do centroide, passou a ter somente três pontos semelhantes (D, E e I).



**Figura 32.** Dendrograma do método de Ward.

#### 4. CONCLUSÃO

Neste trabalho foi desenvolvido um estudo com uma base de dados de materiais recicláveis com oito variáveis, sendo três quantitativas e cinco qualitativas. Estes dados foram coletados em dez locais diferentes que possuíam operações similares trabalhando próximo de comunidades e somente com moeda social e digital durante nove meses, envolvendo 1.185 participantes. Assim, com o intuito de analisar e avaliar os dados foram utilizadas uma análise exploratória de dados e análise de agrupamento (ou *cluster*) hierárquico.

Convém ressaltar que ao aplicar a análise exploratória se percebeu um padrão predominante de comportamento do público atendido e dos resultados do projeto que foi:

- Pessoas com idade de 41 à 60 anos (mais de 55% das pessoas).
- Mulheres (mais de 74% das pessoas).
- Pessoas entregavam mais frequentemente plástico em geral do que outros materiais (mais de 41% das operações).
- Recebiam R\$ 0,90/kg ganhando em média R\$ 23 com a venda mensal dos resíduos (27 kg).

Segundo a Cooperativa Recria o projeto ocorreu somente com moeda social digital e cerca de 80% dos participantes eram moradores locais e não catadores autônomos. Em outros pontos da Cooperativa que operam com dinheiro e não estão inclusos na base de dados estudada, a maioria dos participantes são catadores autônomos, muitos recebem valores mensais acima de 1400 reais com a venda do material. A Cooperativa afirma também que esses pontos de coleta trabalham com margem de contribuição de 35%, repassando 65% do valor de venda para pagar o material aos catadores autônomos.

Ressalta-se ainda que os pontos de coleta que usam somente moeda social e digital tiveram um volume médio mensal de 1.576 kg, enquanto que os pontos de coleta que trabalham com dinheiro, conforme entrevista com o presidente da Cooperativa, chegam a marca de 8,10, 20 toneladas ou mais por mês.

Algumas outras conclusões obtidas foram:

- Pessoas do ponto de coleta K levaram mais material para reciclagem (mais de 40% do peso total dos materiais)
- O mês de outubro foi o de maior volume de material.
- Os materiais de maior peso foram os materiais com menor preço.
- 79% da receita do catador/morador tem origem nos materiais com menos massa (alumínio, PET I e PLÁSTICO), porém sua matéria-prima tem mais valor comercial. Apenas 9% da receita vem de materiais com massa maior e menor preço (vidro e papel).
- Vidro e papel possuem a maior densidade de pesos de embalagens mesmo com menores números de operações e menores preços (mais de 55% do peso total de materiais e 22% do volume de operações). Enquanto que o alumínio, PET I e PLÁSTICO I, os três maiores preços de embalagens por quilo, representam 29% do peso total de materiais e 59% do volume de operações.

É provável que quanto maior a matéria-prima disponível no mercado para embalagens, menor o preço comercial da embalagem reciclável para revenda e o contrário também, quanto a menor a disponibilidade maior preço de revenda. Este fato é relatado como consequência do conceito econômico da lei da demanda ou da oferta e demanda.

Também foi possível concluir que:

- A maioria das operações dos pontos de coleta se concentraram na frequência dos materiais de maiores preços e valores finais. Contudo a maioria dos pesos coletados se concentraram nos materiais de menores preços e valores.

Importante destacar que pela dinâmica da economia de mercado, no setor da reciclagem, o preço da matéria-prima reciclada para venda usa como base referencial o preço da matéria-prima virgem. Em muitos casos a matéria-prima reciclada para fabricação de embalagens deve ser vendida para a indústria de bens e consumo com preço 30% menor que o preço da matéria-prima virgem, conforme artigo da Ambiente Brasil (2023).

Ao ser vendida com preço menor, a matéria-prima reciclada tem mais vantagem competitiva para concorrer com o material de origem primária, despertando a viabilidade financeira da indústria que pode vender o produto acabado mais barato para o consumidor ou somente reduzir o seu custo de embalagem, aumentando a sua margem de contribuição.

Salienta-se que existem algumas iniciativas de poucas indústrias para pagar um preço melhor pela matéria-prima reciclada para atender a legislação ambiental, justificar a recuperação de 22% das embalagens colocadas no mercado anualmente e/ou explorar o *marketing* de sustentabilidade. Este custo adicional é assumido porque representa muitas vezes

não mais que dez centavos no custo da embalagem, sendo facilmente repassado para consumidor final mais consciente, engajado no consumo de produtos sustentáveis e comprometidos com a logística reversa.

À maneira de exemplo, podemos citar uma garrafa PET de 40 gramas. Para fazer um quilo é preciso 25 garrafas, sendo estas garrafas vendidas cada uma a um preço de 10 centavos a mais para o consumidor, haveria um capital disponível de R\$ 2.500,00/tonelada para a indústria financiar a logística reversa e fortalecer a reciclagem além do preço comercial do material reciclado. Atualmente, segundo a Cooperativa Recria, ela recebe de crédito de reciclagem apenas R\$ 70,00/tonelada, uma diferença de R\$ 2.430,00.

Assim, pelo estudo no cenário atual se pode entender que a embalagem pós-consumo de material com maior preço na reciclagem (ex. alumínio, PET I e PLÁSTICO I) possui menor massa e é mais leve. Isto quer dizer que a embalagem usa menos recursos naturais como matéria-prima na sua fabricação, é mais fácil de transportar, aumenta o alcance da distribuição do conteúdo que armazena, diminui o custo logístico de entrega e remunera melhor o morador ou catador que faz coleta seletiva e vende por quilo.

E a embalagem pós-consumo de material com menor preço na reciclagem (ex. vidro e papel) possui maior massa e é mais pesada. Assim usa mais recursos naturais como matéria-prima na sua fabricação, é mais difícil de transportar, diminui o alcance da distribuição do conteúdo que armazena, aumenta o custo logístico de entrega e remunera menos o morador ou catador que faz coleta seletiva e vende por quilo.

A pergunta que deve ser respondida é: quais os critérios usados pelas indústrias para escolherem o tipo de matéria-prima com menor massa ou maior massa para suas embalagens?

Alguns desses critérios, excetuando a questão da qualidade, segurança, proteção, tamanho de armazenamento e *design* da embalagem, são os critérios baseados nos conceitos econômicos de *trade-off*, custo de oportunidade e lei da demanda. O *trade-off* é quando existe mais de uma opção de escolha disponível e você só pode optar por uma já que os seus recursos são limitados. O custo de oportunidade é o custo calculado da não escolha da opção descartada. Este custo deve ser o menor possível e a escolha deve ser a mais benéfica para ser considerada um *trade-off* correto. Já a lei da demanda afirma que quanto menor o preço de algo, maior será a sua procura pelos clientes e vice-versa.

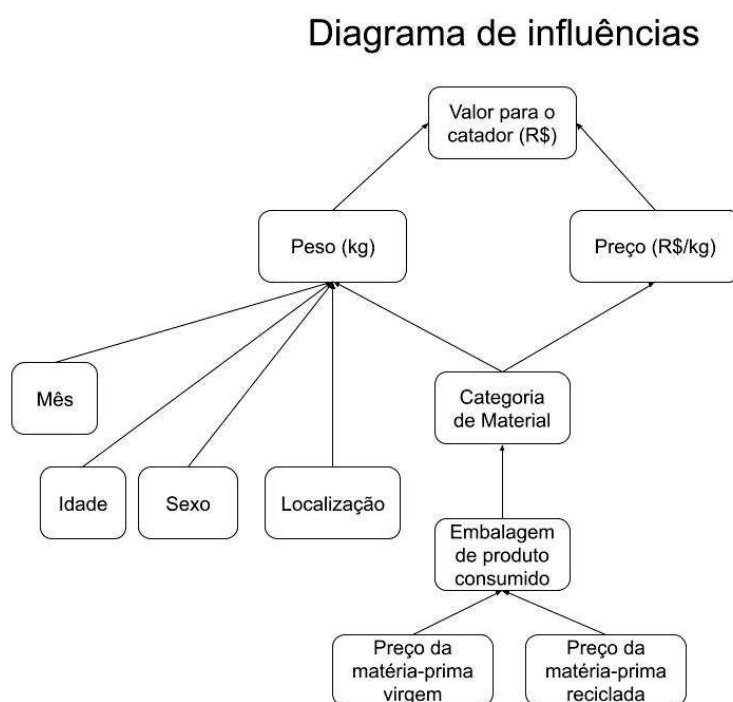
Neste sentido deve-se dizer que ao considerar a associação do preço da matéria-prima reciclada ser proporcional ao preço da matéria-prima virgem, o alumínio, o PET I e o



PLÁSTICO I possuem um valor da matéria-prima virgem maior que a matéria-prima virgem para fazer o vidro e o papel.

Posta assim a questão, é de se dizer que a indústria que compra a matéria-prima do alumínio, o PET I e PLÁSTICO I para fazer as suas embalagens, reconhece que para o seu produto é vantajoso pagar mais caro nesses tipos de materiais para ter compensação do custo logístico e de distribuição, pois são mais leves, e por ter mais valor a embalagem para o catador. E a indústria que compra a matéria-prima do vidro e papel para embalar seus produtos, vê isto como vantagem, pois reduz o seu custo de embalagem e este valor compensa o aumento do custo logístico e de distribuição que ela tem. Neste último caso, o valor pago ao catador não é tão relevante para a indústria.

Em virtude dessas considerações, para uma melhor percepção do encadeamento das variáveis mencionadas dentro da conclusão, na Figura 33 desenhou-se um fluxo chamado diagrama de influências em que a variável resposta Valor tem influência das variáveis Peso e Preço. Estas sofrem influência da Categoria de material. A Categoria de material é afetada pelo tipo de Embalagem de produto consumido e esta última tem influência do Preço de matéria-prima virgem e do Preço de matéria-prima reciclada. Também variável a Peso tem influência das variáveis qualitativas como Mês, Idade, Sexo e Localização.



**Figura 33.** Diagrama de influências de variáveis.

Outras lições aprendidas foram:

- Existe diferença alta de gravimetria entre os dados disponíveis publicamente com relação a fração reciclável dos resíduos sólidos urbanos, principalmente no vidro e papel.
- Pessoas com idade de 41 à 60 anos levaram em média 5,5 quilos de material por transação, um quilo a mais que pessoas com idade de 18 à 40 anos.
- O público masculino levou em média 5,8 quilos de material por transação, 1,3 quilo a mais que o público feminino.
- O preço médio dos materiais comprados se manteve praticamente estável variando a mediana entre R\$ 0,90/kg à R\$ 1,00/kg.
- Os materiais alumínio (R\$ 5,60/kg) e PET I (R\$ 2,20/kg) foram os materiais com os maiores preços médios. E o vidro (R\$ 0,10/kg) e o papel (R\$ 0,20/kg) foram os menores.

Registra-se, ainda, que ao aplicar o método de análise de *cluster* hierárquico sobre as variáveis peso, preço e valor de compra nos dez pontos de coleta, os resultados obtidos foram:

- No método de ligação simples encontrou-se a formação de nove grupos e homogeneidade interna dos pontos D e E.
- No método de ligação completa formaram-se seis grupos e teve homogeneidade interna entre os pontos D, E, F, H e I.
- No método da média das distâncias e no método do centroide os resultados foram similares ao método de ligação completa.
- No método de Ward criaram-se oito grupos e somente os pontos D, E e I tiveram homogeneidade interna.

Portanto, com os resultados da análise exploratória dos dados e da análise de *cluster* foi possível perceber o perfil de pessoas mais atuantes e não atuantes no projeto, quais as características mais relevantes dentro do ponto de coleta, os materiais predominantes, também quais os locais mais semelhantes, quais os mais dissimilares e quais os que tiveram um melhor ou pior desempenho.

As análises desenvolvidas neste trabalho podem contribuir com a área de *marketing* e educação ambiental para criar estratégias direcionadas para aumentar o engajamento do público por perfil. Também pode contribuir com a sociologia, antropologia e psicologia buscando mais informações para entender as motivações de cada grupo. A área de economia pode se beneficiar também entendendo melhor o funcionamento do mercado e a relação das pessoas com o dinheiro e resíduos.

Além disso, com o estudo, a área de engenharia de produção pode buscar características de eficiência e produtividade de cada ponto de coleta e a engenharia e gestão ambiental podem analisar impactos positivos e negativos dos materiais recicláveis. E outras áreas, como por exemplo administração pública e o direito, podem identificar oportunidades de criar novas formas de regular o mercado com políticas ambientais mais eficientes.

Para trabalhos futuros podem ser utilizadas outras técnicas, como os métodos de agrupamento não-hierárquico e as redes neurais para classificar os locais de coleta e, também, os indivíduos via o tipo de material reciclável coletado. Também é relevante fazer uma nova rodada de análise exploratória de dados e de agrupamento com dados de pontos de coleta que trabalham com dinheiro e comparar com este estudo. Outra sugestão seria fazer simulações de substituições do material das embalagens atuais por materiais com maior valor comercial na reciclagem e avaliar o impacto financeiro disto para o consumidor, catador, indústria e mercado, além do impacto ambiental e social. Além disso, um outro escopo de trabalho poderia ser o monitoramento e análise da série histórica dos preços de matéria-prima reciclada com os da matéria-prima virgem para verificar a variação da margem entre os dois preços.

“O grande objetivo de toda ciência...é cobrir o maior número possível de fatos empíricos por meio de deduções lógicas a partir do menor número possível de hipóteses ou axiomas.” A. Einstein

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

AMBIENTE BRASIL. (2023). Reciclagem de PET no Brasil. Ambiente Brasil, 2023. Disponível em: <[https://ambientes.ambientebrasil.com.br/residuos/reciclagem/reciclagem\\_de\\_pet\\_no\\_brasil.html](https://ambientes.ambientebrasil.com.br/residuos/reciclagem/reciclagem_de_pet_no_brasil.html)>. Acesso em 30 de novembro 2023.

ANCAT – Associação Nacional dos Catadores e Catadoras de Materiais Recicláveis. (2022). Atlas Brasileiro da Reciclagem. Reciclagem em números. ANCAT, 2022. Disponível em: <<https://atlasbrasileirodareciclagem.ancat.org.br/reciclagem-em-numeros>>. Acesso em 27 de Novembro de 2023.

BRAGA, A. A reciclagem de resíduos sólidos urbanos: um olhar a partir de três estudos de casos. São Paulo, 2018.

CEMPRE – Compromisso Empresarial para Reciclagem. (2019). CEMPRE Review 2019. CEMPRE, 2019. Disponível em: <<https://cempre.org.br/wp-content/uploads/2020/11/CEMPRE-Review2019.pdf>>. Acesso em 30 de novembro de 2023.

CIRS Búzios Economia Circular: Sustentabilidade Econômica. 2023. Disponível em: <[https://youtu.be/R4FsOvn1\\_jk?si=jEQRIoIxxKCoGOWB](https://youtu.be/R4FsOvn1_jk?si=jEQRIoIxxKCoGOWB)>. Acesso em 30 de novembro de 2023.

COMLURB - Companhia Municipal de Limpeza Urbana. (2021). Total do lixo recolhido através de coleta seletiva e total recuperado por tipo de material no Município do Rio de Janeiro entre 2002-2021. COMLURB, 2021. Disponível em: <<https://datariov2-perj.hub.arcgis.com/documents/4b74be782816403f9fda15df584d01f2/about>>. Acesso em 27 de novembro de 2023.

EXAME. (2023). Lixo de brasileiros em 2022 equivale a 85 milhões de carros populares. Revista Exame, 2023. Disponível em: <<https://exame.com/ciencia/lixo-de-brasileiros-em-2022-equivale-a-85-milhoes-de-carros-populares/>>. Acesso em 27 de novembro de 2023.

FERREIRA, R. et al. Análise de *cluster* não supervisionado em R: agrupamento hierárquico. Embrapa Territorial. Campinas, 2020.

FIRJAN – Federação das Indústrias do Estado do Rio de Janeiro. (2023). Mapeamento dos recicláveis pós-consumo no Estado do Rio de Janeiro. FIRJAN. Rio de Janeiro, 2023. Disponível em: <<https://www.firjan.com.br/publicacoes/manuais-e-cartilhas/mapeamento-dos->>

fluxos-de-reciclaveis-pos-consumo-no-estado-do-rio-de-janeiro.htm>. Acesso em 27 de novembro de 2023.

HAIR, J. F. et al. *Análise Multivariada de Dados*. Tradução: Adonai Schlup Sant'Anna e Anselmo Chaves Neto. 5ed. Porto Alegre: Bookman, 2005.

IPEA – Instituto de Pesquisa Econômica Aplicada. (25 de janeiro de 2017). Apenas 13% dos resíduos sólidos urbanos vão para reciclagem. IPEA, 2017. Disponível em: <<https://www.ipea.gov.br/portal/categorias/45-todas-as-noticias/noticias/2841- apenas-13-dos-residuos-solidos-urbanos-vao-para-reciclagem?highlight=WyJlY29ub21pYSIsImVjb25vbWlhJyIsInNvbGlkXHUwMGUxcmlhIiwZWNvbm9taWEgc29saWRhcmlhIl0=>> Acesso em 27 de novembro de 2023.

JOHNSON, R. A., WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 4ed. New Jersey: Prentice Hall, 1998.

JUCÁ, J. F. T. et al. *Análise das Diversas Tecnologias de Tratamento e Disposição Final de Resíduos Sólidos Urbanos no Brasil, Europa, Estados Unidos e Japão*. Jaboaão dos Guararapes, PE: Grupo de Resíduos Sólidos da UFPE e Banco Nacional do Desenvolvimento Econômico e Social - BNDES, 2014. 181p.

KAZA, S. et al. “What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050.” Overview booklet. World Bank, Washington, DC, 2018. License: Creative Commons Attribution CC BY 3.0 IGO.

MANLY, B. J. F.; ALBERTO, Jorge A. N. *Multivariate Statistical Methods: A Primer*. 4ed., Boca Raton: CRC Press, Includes bibliographical references and index. 2017.

MANLY, B. J. F. *Métodos Estatísticos Multivariados: Uma Introdução*. 3ed. Tradução: Sara Ianda Correa Carmona. Porto Alegre:Bookman, 2008.

OLIVEIRA, N. L. de. *Análise de agrupamento hierárquicos*. Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Fortaleza, CE, 2022.

REIS, T. Trade off: o que é? Entenda esse importante conceito econômico. Suno, 2018. Disponível em: < <https://www.suno.com.br/artigos/trade-off/>>. Acesso em 30 nov 2023.

RFB – República Federativa do Brasil. (2010). Política Nacional de Resíduos Sólidos - PNRS. Lei Federal nº 12.305/2010. RFB, 2010. Disponível em: < [https://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2010/lei/112305.htm](https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/lei/112305.htm)>. Acesso em 27 de novembro de 2023.

SILVA, A. R. da. *Métodos de análise multivariada em R*. Piracicaba: FEALQ, 2016.

SILVA, F. C. da. *Introdução à Análise Multivariada*. UFRRJ, 2023.

TRIOLA, M. F. *Introdução à Estatística*. 7 ed. Livros Técnicos e Científicos Editora, 1999.

UNEP - United Nations Environment Programme. (2023). Turning off the Tap. How the world can end plastic pollution and create a circular economy. Nairobi. UNEP, 2023.

Disponível em: <<https://www.unep.org/resources/turning-off-tap-end-plastic-pollution-create-circular-economy>>. Acesso em 27 de novembro de 2023.

VASCONCELOS, F. et al. Glossário ilustrado de Resíduos. Faculdade de Engenharia Química da Universidade Estadual de Campinas (UNICAMP), GAIA Social e Programa de Educação Tutorial (PET) do Ministério da Educação - MEC, 2020.

WWF - World Wildlife Fund. (2023). Dia da sobrecarga da Terra. WWF-Brasil, 2023. Disponível em: <<https://www.wwf.org.br/overshootday/>> Acesso em 27 de novembro de 2023.

## 6. ANEXO

### ANEXO A – Rotina do R

```
rm(list=ls(all=TRUE)) # Limpar a memoria

options(stringsAsFactors = FALSE)

setwd("C:/Users/Usuario/OneDrive/Documents/TCC de Estatística Aplicada")

library(readxl)

dados_tcc <- read_excel("dados_tcc.xlsx")

View(dados_tcc)

dados_tcc

dim(dados_tcc)

names(dados_tcc)

head(dados_tcc,5)

tail(dados_tcc,4)

str(dados_tcc)


# indicating the data type of
# each variable

sapply(dados_tcc, class)


# converting character type column
# to numeric and to factor

dados_tcc2<- transform(dados_tcc,

                        Peso_Kg = as.numeric(as.character(Peso_Kg)),
```

```

Preco_Tabela = as.numeric(as.character(Preco_Tabela)),

Valor = as.numeric(as.character(Valor)),

Idade_3 = as.factor(as.character(Idade_3)),

Sexo = as.factor(as.character(Sexo)),

Ecoponto = as.factor(as.character(Ecoponto)),

Categoria = as.factor(as.character(Categoria)),

Mes = as.factor(as.character(Mes)))

str(dados_tcc2)

#Estatísticas descritivas

round(apply(dados_tcc2[,1:3],2,var),4) # 2 = calcular variancia da coluna; 4 = n?mero de casas decimais, var = variancia

round(apply(dados_tcc2[,1:3],2,mean),4) # mean = media

summary(dados_tcc2)

#Boxplot

x11()

boxplot(dados_tcc2[,1:3])

summary(dados_tcc2)

colSums(dados_tcc2 ==0)

attach(dados_tcc2)

# Figura : Total de peso por Idade

barplot(tapply(dados_tcc2$Peso_Kg,dados_tcc2$Idade_3,sum),xlab="Idade", ylab="Quantidade de peso em kg",
        main="Total de peso por idade", col=c("royalblue4","royalblue3","royalblue2"))

# Figura : Total de peso por Sexo

barplot(tapply(dados_tcc2$Peso_Kg,dados_tcc2$Sexo,sum),xlab="Sexo", ylab="Quantidade de peso em kg",
        main="Total de peso por sexo", col=c("royalblue4","royalblue3","royalblue2"))

sum(dados_tcc2$Peso_Kg)

```



```
sum(dados_tcc2$Valor)
```

```
117582.7/140181.5
```

```
length(unique(dados_tcc2$Coleta))
```

```
140181.5/12261
```

```
140181.5/1185
```

```
118.2966/9
```

```
140181.5/9
```

```
tapply(dados_tcc2$Peso_Kg,dados_tcc2$Mes,sum)
```

```
tapply(dados_tcc2$Peso_Kg,dados_tcc2$Sexo,sum)
```

```
sum(dados_tcc2$Peso_Kg)
```

```
tapply(dados_tcc2$Peso_Kg,dados_tcc2$Sexo,sum)/28848
```

```
sum(dados_tcc2$Peso_Kg)
```

```
# count unique elements
```

```
length(unique(dados_tcc2$Sexo))
```

```
----
```

```
# Tabela 1: Descrição das variáveis qualitativas
```

```
library(tidyverse)
```

```
install.packages("gt")
```

```
library("gt")
```

```
install.packages("gapminder")
```

```
library(gapminder)
```

```
Variável<-c("Idade","Sexo","Localização","Categoria","Mês")
```

```
Definição<-c("Variável qualitativa nominal simples categorizada por idade com intervalo de vinte anos.",
```

"Variável qualitativa nominal simples categorizada por sexo masculino e feminino. Refere-se as pessoas que levavam o material até o ponto de coleta.",

"Variável qualitativa nominal simples categorizada por ponto de coleta.",

"Variável qualitativa nominal simples categorizada por tipo de material.",

"Variável qualitativa nominal simples categorizada por mês.")

```
Tabela<-data.frame(Variável,Definição)
```

Tabela

Tabela %>%

```
gt_preview() %>% tab_header(title = md("***Tabela 1: Descrição das variáveis qualitativas***"),
```

```
  subtitle = "") %>%
```

```
tab_options(heading.align = "left",
```

```
  column_labels.border.top.color = "black",
```

```
  column_labels.border.top.width = px(3),
```

```
  column_labels.background.color = "white") %>%
```

```
tab_source_note(md("***Fonte: Elaborado pelo autor.**"))
```

-----

# Tabela 2: Descrição das variáveis quantitativas

```
Variável<-c("Peso (kg)", "Preço (R$/kg)", "Valor (R$)")
```

Definição<-c("Variável quantitativa aleatória contínua informada em quilo. Refere-se ao peso obtido de resíduos por tipo ao final de cada transação.",

"Variável quantitativa aleatória discreta informada em decimal. Contém o preço do material comprado.",

"Variável quantitativa aleatória contínua informada em valor em reais.")

```
Tabela<-data.frame(Variável,Definição)
```

Tabela

Tabela %>%

```
gt_preview() %>% tab_header(title = md("***Tabela 2: Descrição das variáveis quantitativas***"),
                             subtitle = "") %>%
  tab_options(heading.align = "left",
              column_labels.border.top.color = "black",
              column_labels.border.top.width = px(3),
              column_labels.background.color = "white") %>%
  tab_source_note(md("***Fonte: Elaborado pelo autor.***"))
```

-----

## #1. ANÁLISE EXPLORATÓRIA DE DADOS

```
summary(dados_tcc2)
```

# Tabela 3: Resumo da análise das variáveis categóricas por frequência"

```
Idade<-c("18-40","41-60","61 ou mais","", "", "")
```

```
Quant.por.idade<-c(6717,15994,6137","", "", "")
```

```
Sexo<-c("F","M","", "", "", "")
```

```
Quant.por.sexo<-c(21268,7580","", "", "", "")
```

```
Localização<-c("PONTO K","PONTO A","PONTO J","PONTO F","PONTO C","(OUTROS)")
```

```
Quant.por.localização<-c(8972,7288,3271,3011,2366,3940)
```

```
Categoria<-c("PET I","PLÁSTICO I","ALUMÍNIO","PAPEL","FERRO","(OUTROS)")
```

```
Quant.por.categoria<-c(6523,5424,5185,4151,3249,4316)
```

```
Mês<-c("JUL","AGO","SET","OUT","NOV","(OUTROS)")
```

```
Quant.por.mês<-c(4564,4249,4019,3844,2783,9389)
```

```
tabela_categorica<-
data.frame(Idade,Quant.por.idade,Sexo,Quant.por.sexo,Localização,Quant.por.localização,Categoria,Quant.por.categoria,Mês,Quant.por.mês
)
```

```
tabela_categorica
```

```
names(tabela_categorica)[1:10] <- c("Idade", "Quantidade por idade", "Sexo","Quantidade por sexo","Localização","Quantidade por
localização","Categoria","Quantidade por categoria","Mês","Quantidade por mês")
```

```
tabela_categorica
```

```
library(gt)
```

```
tabela_categorica %>%
```

```
gt_preview() %>% tab_header(title = md("***Tabela 3: Resumo da análise das variáveis categóricas por frequência***"),
```

```
subtitle = "") %>%
```

```
tab_options(heading.align = "left",
```

```
column_labels.border.top.color = "black",
```

```
column_labels.border.top.width = px(3),
```

```
column_labels.background.color = "white") %>%
```

```
tab_source_note(md("***Fonte: Elaborado pelo autor.***"))
```

```
# Tabela 4: Resumo da análise das variáveis quantitativas por medida de tendência central e medida de posição"
```

```
summary(dados_tcc2)
```

```
Medidas<-c("Mínimo","1º quartil","Mediana","Média","3º quartil","Máximo")
```

```
Peso<-c(0.010,0.650,1.555,4.859,3.850,555.650)
```

```
Preço<-c(0.00,0.30,0.90,1.81,2.30,10.00)
```

```
Valor<-c(0.00,0.52,1.61,4.08,4.37,687.60)
```

```
tabela_quantitativa<-data.frame(Medidas,Peso,Preço,Valor)
```

```
tabela_quantitativa
```

```
names(tabela_quantitativa)[2:4] <- c("Peso (kg)", "Preço (R$/kg)", "Valor (R$)")
```

```
tabela_quantitativa
```

```
library(gt)
```

```
tabela_quantitativa %>%
```

```
gt_preview() %>% tab_header(title = md("**Tabela 4: Resumo da análise das variáveis quantitativas por medida de tendência central e  
medida de posição**"),
```

```
      subtitle = "") %>%
```

```
tab_options(heading.align = "left",
```

```
      column_labels.border.top.color = "black",
```

```
      column_labels.border.top.width = px(3),
```

```
      column_labels.background.color = "white") %>%
```

```
tab_source_note(md("**Fonte: Elaborado pelo autor.**"))
```

```
summary(dados_tcc2)
```

```
colSums(dados_tcc2 ==0)
```

```
colSums(dados_tcc2 !=0)
```

```
attach(dados_tcc2)
```

```
dados2 = as.data.frame(apply(dados_tcc2[,1:3], 2, as.numeric))
```

```
x11();boxplot(dados2, main="Boxplot das variáveis quantitativas")
```

```
x11();boxplot(log(dados2),main="Boxplot da variáveis quantitativas com log")
```

```
# Figura 2: Total de material por idade
```

```
barplot(tapply(dados_tcc2$Peso_Kg,dados_tcc2$Idade_3,sum),xlab="Idades", ylab="Quantidade de material em kg",
        main="Total de materiais recicláveis por idade", col=c("royalblue4","royalblue3","royalblue2"))
```

#Figura 3: Total de material por sexo

```
barplot(tapply(dados_tcc2$Peso_Kg,dados_tcc2$Sexo,sum),xlab="Sexo", ylab="Quantidade de material em kg",
        main="Total de materiais recicláveis por sexo", col=c("royalblue4","royalblue3","royalblue2"))
```

#Figura 4: Total de material por Ecoponto

```
x11()
barplot(tapply(dados_tcc2$Peso_Kg,dados_tcc2$Ecoponto,sum),xlab="Ecoponto", ylab="Quantidade de material em kg",
        main="Total de materiais recicláveis por Ecoponto", col=c("royalblue4","royalblue3","royalblue2"))
```

#Figura 5: Total de material por Categoria

```
x11()
barplot(tapply(dados_tcc2$Peso_Kg,dados_tcc2$Categoria,sum),xlab="Categoria", ylab="Quantidade de material em kg",
        main="Total de materiais recicláveis por Categoria", col=c("royalblue4","royalblue3","royalblue2"))
```

```
x11();pie(table(dados_tcc2$Categoria), main="Total de frequência por Categoria") #grafico pizza
```

#Figura 6: Total de material por Mês

```
x11()
barplot(tapply(dados_tcc2$Peso_Kg,dados_tcc2$Mes,sum),xlab="Mês", ylab="Quantidade de material em kg",
        main="Total de materiais recicláveis por Mês", col=c("royalblue4","royalblue3","royalblue2"))
```

#Média/Mediana/desvio-padrão de material coletados por kg

```
mean(dados_tcc2$Peso_Kg)
median(dados_tcc2$Peso_Kg)
sd(dados_tcc2$Peso_Kg)
```

```
#Média/Mediana/desvio-padrão de valor por operação
```

```
mean(dados_tcc2$Valor)
```

```
median(dados_tcc2$Valor)
```

```
sd(dados_tcc2$Valor)
```

```
library(ggplot2)
```

```
dados_tcc2
```

```
x11()
```

```
ggplot(data = dados_tcc2, aes(x = Preco_Tabela, y= Peso_Kg)) +
```

```
  geom_point(color="royalblue3", fill="royalblue3") +
```

```
  ggtitle("Gráfico de dispersão entre Preço e Peso") + xlab("Preço (R$/kg)") + ylab("Peso (kg)")
```

```
dados_tcc2
```

```
ggplot(data = dados_tcc2, aes(x = Preco_Tabela, y= Valor)) +
```

```
  geom_point(color="royalblue3", fill="royalblue3") +
```

```
  ggtitle("Gráfico de dispersão entre Preço e Valor") + xlab("Preço (R$/kg)") + ylab("Valor (R$)")
```

```
dados_tcc2
```

```
ggplot(data = dados_tcc2, aes(x = Peso_Kg, y= Valor)) +
```

```
  geom_point(color="royalblue3", fill="royalblue3") +
```

```
  ggtitle("Gráfico de dispersão entre Peso e Valor") + xlab("Peso (kg)") + ylab("Valor (R$)")
```

```
dados_tcc2
```

```
ggplot(data = dados_tcc2, aes(x = Categoria, y= Peso_Kg)) +
```

```
  geom_point(color="royalblue3", fill="royalblue3") +
```

```
  ggtitle("Gráfico de dispersão entre Categoria e Peso") + xlab("Categoria") + ylab("Peso (kg)")
```

```

ggplot(data = dados_tcc2, aes(x = Categoria, y= Preco_Tabela)) +

geom_point(color="royalblue3", fill="royalblue3") +

ggtitle("Gráfico de dispersão entre Categoria e Preço") + xlab("Categoria") + ylab("Preço (R$/kg)")


ggplot(data = dados_tcc2, aes(x = Categoria, y= Valor)) +

geom_point(color="royalblue3", fill="royalblue3") +

ggtitle("Gráfico de dispersão entre Categoria e Valor") + xlab("Categoria") + ylab("Valor (R$)")


ggplot(data = dados_tcc2, aes(x = Mes, y= Preco_Tabela)) +

geom_point(color="royalblue3", fill="royalblue3") +

ggtitle("Gráfico de dispersão entre Mês e Preço")


#Bloxplot material coletado por peso e idade

ggplot(data = dados_tcc2, aes(x = Idade_3, y= Peso_Kg)) +

geom_boxplot(color="royalblue3", fill="lightgray") +

ggtitle("Bloxplot de material por peso e idade") +

xlab("Idade") + ylab("Quantidade de material por kg") +

stat_summary(fun = mean, geom = "point", col="royalblue3") +

stat_summary(fun = mean, geom = "text", col = "royalblue3", vjust = 1.5, aes(label = paste("Média:",round(..y..,digits=1))))


#Bloxplot material coletado por peso e sexo

ggplot(data = dados_tcc2, aes(x = Sexo, y= Peso_Kg)) +

geom_boxplot(color="royalblue3", fill="lightgray") +

ggtitle("Bloxplot de material por peso e idade") +

xlab("Idade") + ylab("Quantidade de material por kg") +

stat_summary(fun = mean, geom = "point", col="royalblue3") +

stat_summary(fun = mean, geom = "text", col = "royalblue3", vjust = 1.5, aes(label = paste("Média:",round(..y..,digits=1))))


#Bloxplot de preço de material por mês

```



```
ggplot(data = dados_tcc2, aes(x = Mes, y= Preco_Tabela)) +

geom_boxplot(color="red3", fill="lightgray") +

ggtitle("Bloxplot de preço de material por mês") +

xlab("Mês") + ylab("Preço") +

stat_summary(fun = median, geom = "point", col="red3") +

stat_summary(fun = median, geom = "text", col = "red3", vjust = -1.0, aes(label = paste("Mediana:",round(..y...,digits=1))))
```

#Bloxplot de preço de material por categoria

```
ggplot(data = dados_tcc2, aes(x = Categoria, y= Preco_Tabela)) +

geom_boxplot(color="red3", fill="lightgray") +

ggtitle("Bloxplot de preço de material por categoria") +

xlab("Categoria") + ylab("Preço") +

stat_summary(fun = mean, geom = "point", col="red3") +

stat_summary(fun = mean, geom = "text", col = "red3", vjust = -2.0, aes(label = paste("Média:",round(..y...,digits=1))))
```

#Bloxplot de peso de material por categoria

```
ggplot(data = dados_tcc2, aes(x = Categoria, y= Peso_Kg)) +

geom_boxplot(color="red3", fill="lightgray") +

ggtitle("Bloxplot de peso de material por categoria") +

xlab("Categoria") + ylab("Preço") +

stat_summary(fun = mean, geom = "point", col="red3") +

stat_summary(fun = sum, geom = "text", col = "red3", vjust = -2.0, aes(label = paste("Sum",round(..y...,digits=1))))
```

```
sum(dados_tcc2$Peso_Kg)

sum(dados_tcc2$Peso_Kg[dados_tcc2$Categoria=="VIDRO"])

sum(dados_tcc2$Peso_Kg[dados_tcc2$Categoria=="PAPEL"])
```

78/140

#Participação financeira no rendimento do catador por material

```
summary(dados_tcc2$Valor)

sum(dados_tcc2$Valor)
```

```
tapply(dados_tcc2$Valor,dados_tcc2$Categoria,sum)
```

```
#Gravimetria por tipo de material
```

```
tapply(dados_tcc2$Peso_Kg,dados_tcc2$Categoria,sum)
```

```
sum(dados_tcc2$Peso_Kg)
```

```
#Tabela 5 Contribuição financeira para os moradores/catadores por material
```

```
Material<-c("ALUMÍNIO","PET I","PLÁSTICO I","FERRO","PAPEL","VIDRO")
```

```
Valor<-c(39,28,12,8,6,3)
```

```
Gravimetria_coop<-c("6%","11%","12%","11%","28%","28%")
```

```
tabela_fin<-data.frame(Material,Valor,Gravimetria_coop)
```

```
names(tabela_fin)[2:3] <- c("Para cada 100 reais este material contribui com","Gravimetria da Cooperativa Recria")
```

```
tabela_fin
```

```
library(gt)
```

```
tabela_fin %>%
```

```
gt_preview() %>% tab_header(title = md("***Tabela 5: Valor que cada material contribui para o catador/morador e a gravimetria***"),
```

```
subtile = "") %>%
```

```
tab_options(heading.align = "left",
```

```
column_labels.border.top.color = "black",
```

```
column_labels.border.top.width = px(3),
```

```
column_labels.background.color = "white") %>%
```

```
tab_source_note(md("***Fonte: Elaborado pelo autor.**"))
```

```
#Tabela 6 Gravimetrias comparadas
```

```
Material<-c("METAL","PLÁSTICOS","PAPEL","VIDRO")
```

```
Gravimetria_coop<-c("17%","33%","28%","28%")
```

```
Gravimetria_comlurb<-c("16%","18%","54%","13%")
```

```
Gravimetria_cempre<-c("10%","42%","41%","8%")
```

```
tabela_grav<-data.frame(Material,Gravimetria_coop, Gravimetria_comlurb,Gravimetria_cempre)
```

```
names(tabela_grav)[2:4] <- c("Gravimetria da Cooperativa Recria (2022-2023)","Gravimetria da Comlurb (2021)", "Gravimetria do CEMPRE (2019)")
```

```
tabela_grav
```

```
library(gt)
```

```
tabela_grav %>%
```

```
gt_preview() %>% tab_header(title = md("***Tabela 6: Gravimetrias comparadas***"),
```

```
subtile = "") %>%
```

```
tab_options(heading.align = "left",
```

```
column_labels.border.top.color = "black",
```

```
column_labels.border.top.width = px(3),
```

```
column_labels.background.color = "white") %>%
```

```
tab_source_note(md("***Fonte: Elaborado pelo autor.**"))
```

```
----
```

```
----
```

```
#Análise de Cluster
```

```
rm(list=ls()) #Limpar memoria
```

```
cat("\014") #limpar console
```

```
gc(reset = TRUE) #reset
```

```
setwd("C:/Users/Usuario/OneDrive/Documents/TCC de Estatística Aplicada")
```

```
library(readxl)
```

```

dados = read_xlsx('dados_tcc3.xlsx') # Variáveis

#convert tibble to data frame

str(dados)

dados1<- transform(dados,

                    Peso_Kg = as.numeric(as.character(Peso_Kg)),

                    Preco_Tabela = as.numeric(as.character(Preco_Tabela)),

                    Valor = as.numeric(as.character(Valor)))

str(dados1)

dados2 = as.data.frame(dados1)

str(dados2)

View(dados2)

pontos = dados2[, 1]

X = dados2[, -1]

row.names(X) = pontos

View(X)

str(X)


#a)

#### Metodo Ligacao Simples #####

x11()

hr=hclust(dist(X),method="single")

plot(hr)


grupos=round(cbind(rep((nrow(X)-1):1),hr$merge,hr$height, (1-hr$height/max(dist(X)))*100),3)

colnames(grupos) <- c("No de Grupos", "Grupos", "Unidos", "Distancia", "Similaridade")

grupos


groups <- cutree(hr, k=9)

rect.hclust(hr, k=9, border="red")


#### Metodo Ligacao Completa #####

x11()

hr=hclust(dist(X),method="complete")

```

```
plot(hr)
```

```
grupos=round(cbind(rep((nrow(X)-1):1),hr$merge,hr$height, (1-hr$height/max(dist(X)))*100),3)
```

```
colnames(grupos) <- c("No de Grupos", "Grupos", "Unidos", "Distancia", "Similaridade")
```

```
grupos
```

```
groups <- cutree(hr, k=6)
```

```
rect.hclust(hr, k=6, border="red")
```

```
#### Metodo da Media #####
```

```
x11()
```

```
hr=hclust(dist(X),method="average")
```

```
plot(hr)
```

```
grupos=round(cbind(rep((nrow(X)-1):1),hr$merge,hr$height, (1-hr$height/max(dist(X)))*100),3)
```

```
colnames(grupos) <- c("No de Grupos", "Grupos", "Unidos", "Distancia", "Similaridade")
```

```
grupos
```

```
groups <- cutree(hr, k=6)
```

```
rect.hclust(hr, k=6, border="red")
```

```
#### Metodo da Centroide #####
```

```
x11()
```

```
hr=hclust(dist(X),method="centroid")
```

```
plot(hr,main="Metodo de Centroide",xlab="Pontos",ylab="Distância")
```

```
grupos=round(cbind(rep((nrow(X)-1):1),hr$merge,hr$height, (1-hr$height/max(dist(X)))*100),3)
```

```
colnames(grupos) <- c("No de Grupos", "Grupos", "Unidos", "Distancia", "Similaridade")
```

```
grupos
```

```
groups <- cutree(hr, k=6)
```

```
rect.hclust(hr, k=6, border="red")
```

```
#### Metodo de Ward #####

x11()

hr=hclust(dist(X),method="ward.D2")

plot(hr)


grupos=round(cbind(rep((nrow(X)-1):1),hr$merge,hr$height, (1-hr$height/max(dist(X)))*100),3)

colnames(grupos) <- c("No de Grupos", "Grupos", "Unidos", "Distancia", "Similaridade")

grupos


groups <- cutree(hr, k=8)

rect.hclust(hr, k=8, border="red")


#fim
```