

4. ANÁLISE EXPLORATÓRIA

ANÁLISE EXPLORATÓRIA

Por quê?



Quando?



Como?



ANÁLISE EXPLORATÓRIA

Como?



- ▣ Medidas descritivas
 - ▾ Medidas de posição
 - ▾ Medidas de dispersão
- ▣ Gráficos

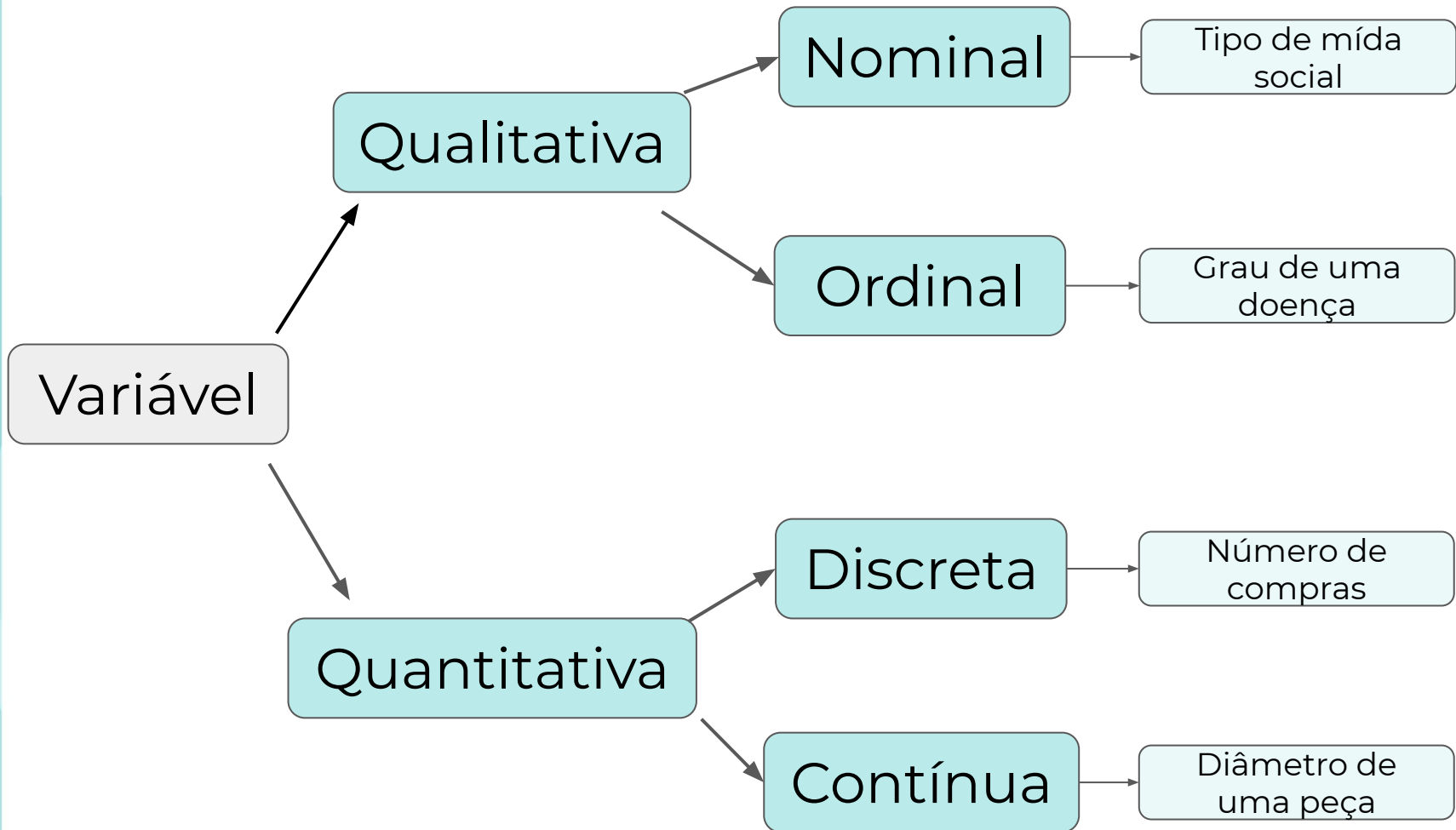
TIPOS DE VARIÁVEIS

4.1

CLASSIFICAÇÃO DE VARIÁVEIS

De forma mais geral: quantitativa ou qualitativa.

Ou mais especificamente: quantitativa discreta ou contínua, qualitativa nominal ou ordinal.



Variável qualitativa nominal: atributo que **não** pode ser ordenado naturalmente.

Variável qualitativa ordinal: atributo para o qual existe uma ordem.

Variável quantitativa discreta: valores assumidos são discretos (conjunto finito de valores ou enumerável).

Variável quantitativa contínua: valores assumidos são contínuos (intervalo de valores ou reunião de intervalos).

NOTAS

- As variáveis qualitativas são não numéricas mas podem ser representadas por números.
 - Grau de escolaridade: 1 → EF; 2 → EM; 3 → ES;
4 → pós-graduação.
- As variáveis quantitativas são numéricas.

POR QUE CLASSIFICAR AS VARIÁVEIS?

- Gráficos adequados.
- Modelos adequados.
 - Modelos para variáveis qualitativas ou quantitativas, discretas ou contínuas.

TABELA

O que é? É uma forma de apresentar informações que se assemelha a uma matriz, com a particularidade de que uma tabela deve ser autoexplicativa.

Para que serve? Facilitar o entendimento do que é apresentado.

Como? Colocar título que descreva a tabela. Nomear as colunas e linhas para que as informações sejam claras. Cores e linhas podem ser usadas para cumprir o objetivo.

Exemplo

Tabela: Avaliação da intensidade da dor em um período pós-operatório considerando dois métodos de analgesia.

Sexo	Método	Intensidade da dor					Totais
		Intolerável	Intensa	Moderada	Fraca	Ausente	
F	A	9	19	21	81	537	667
	B	17	33	37	134	445	666
M	A	6	9	13	53	586	667
	B	12	16	25	86	528	667

Fonte: Giolo, S. R. (2017). *Introdução à Análise de Dados Categóricos com Aplicações*. São Paulo: Editora Blucher, p. 198.

Outra escala de cor

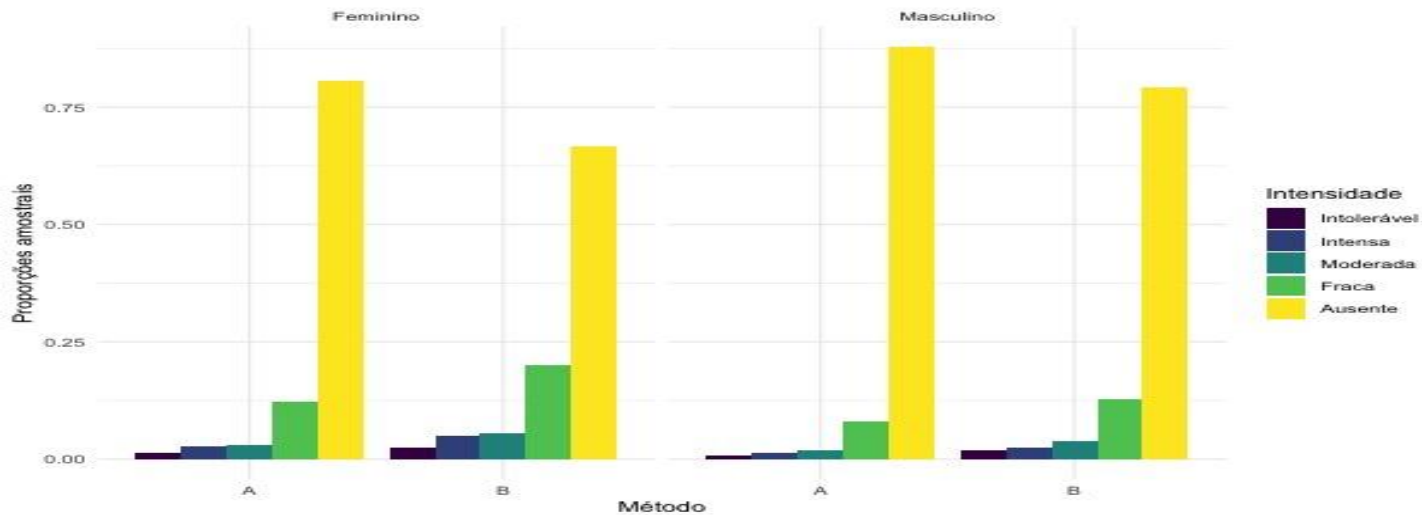


Figura: Avaliação da intensidade da dor em um período pós-operatório considerando dois métodos de analgesia.

TABELA *versus* GRÁFICO

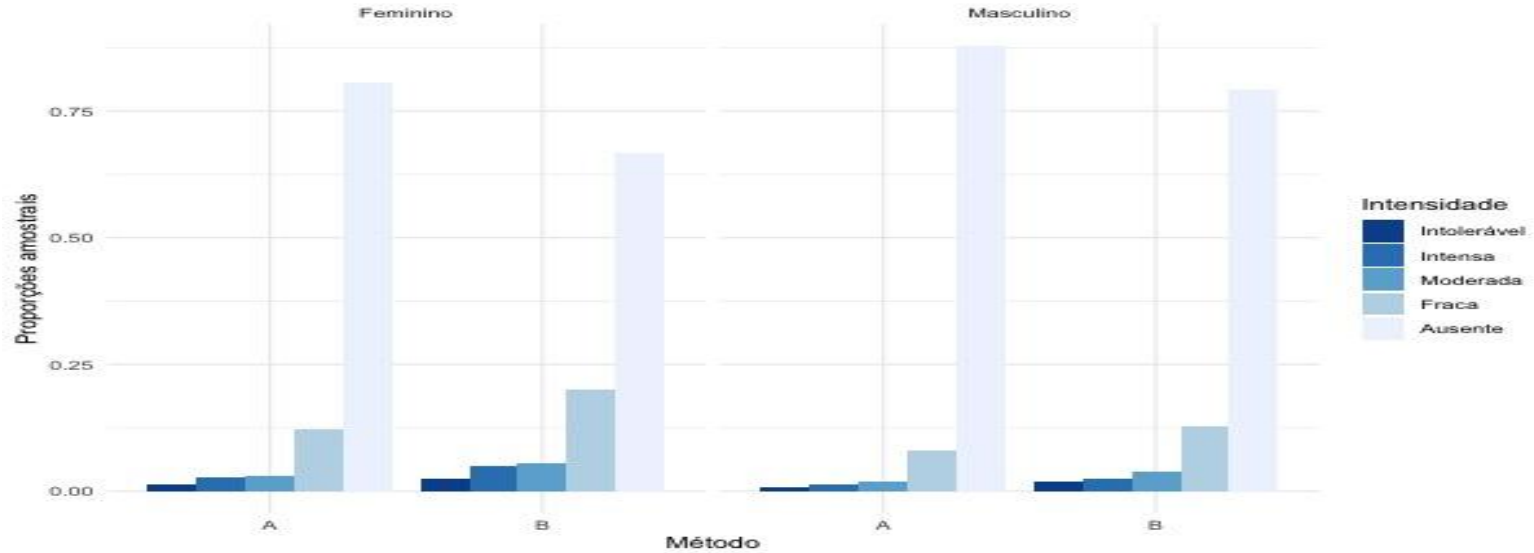


Figura: Avaliação da intensidade da dor em um período pós-operatório considerando dois métodos de analgesia.

CONCLUSÃO: Tabela e gráfico podem ser complementares ou um gráfico pode substituir uma tabela.

TABELA DE FREQUÊNCIAS

O que é? Tabela com pelo menos um dos tipos de frequências.

Para que serve? Apresentar as frequências das variáveis em estudo.

Como? Calculando as frequências e as dispondo organizadamente em uma tabela.

TIPOS DE FREQUÊNCIA

Frequência absoluta: quantidade de vezes que um certo valor ou atributo aparece no conjunto de dados.

Frequência relativa: proporção de certo valor, faixa de valores ou atributo no conjunto de dados.

Frequência acumulada: quando o valor, faixa de valores ou atributo é uma variável quantitativa ou qualitativa ordinal, a frequência acumulada existe e representa o acúmulo das quantidades de vezes que o valor, faixa de valores ou atributo juntamente com as quantidades dos valores ou atributos antecedentes (ou sucessores).

Frequência acumulada relativa: é a frequência acumulada em relação ao todo.

Tipos de frequência: Notação

É comum ser usada a seguinte notação

- ▮ Frequência absoluta: f_j
- ▮ Frequência relativa: f_{rj} .
- ▮ Frequência acumulada: F_j .
- ▮ Frequência acumulada relativa: F_{rj} .

Nota: j indica o valor, faixa de valores ou atributo.

Exemplo

Dados: Resposta sobre hábitos alimentares (vegano, vegetariano, onívoro, outros)

Onívoro	Onívoro	Vegetariano	Vegetariano	Onívoro	Vegetariano
Onívoro	Vegetariano	Onívoro	Onívoro	Onívoro	Outros
Onívoro	Onívoro	Onívoro	Onívoro	Onívoro	Onívoro
Onívoro	Onívoro	Onívoro	Onívoro	Onívoro	Onívoro
Onívoro	Onívoro	Onívoro	Onívoro	Onívoro	Onívoro
Onívoro	Onívoro	Onívoro	Vegetariano	Onívoro	Onívoro
Outros	Onívoro	Onívoro	Onívoro	Onívoro	Onívoro
Vegano	Vegetariano	Onívoro	Onívoro	Onívoro	Onívoro
Onívoro	Onívoro	Onívoro	Onívoro	Vegetariano	Onívoro
Onívoro	Onívoro	Onívoro	Onívoro	Onívoro	Onívoro
Outros	Vegano	Outros	Vegetariano	Onívoro	Onívoro
Onívoro	Onívoro	Vegetariano	Onívoro	Onívoro	Onívoro
Vegetariano	Onívoro	Onívoro	Onívoro	Onívoro	Onívoro
Vegano	Onívoro				

Exemplo: Tabela de frequências

Qualitativa nominal

Tabela: Frequências absoluta e relativa dos hábitos alimentares de uma amostra de frequentadores da cantina.

j	Hábito alimentar	f_j	f_{r_j}	f_{r_j} (%)
1	Onívoro	63	0,7875	78,75
2	Vegano	3	0,0375	3,75
3	Vegetariano	10	0,1250	12,50
4	Outros	4	0,0500	5,00
	Total	80	1,0000	100,00

Pergunta: Faz sentido obtermos as frequências acumuladas?

Resposta: Não, pois a variável *hábito alimentar* é qualitativa nominal.

Exemplo: Tabela de frequências

Quantitativa discreta

Tabela: Frequências do número de refeições de uma amostra de frequentadores da cantina.

j	Refeições	f_j	f_{rj}	f_{rj} (%)	F_j	F_{rj}	F_{rj} (%)
1	1	1	0,0125	1,25	1	0,0125	1,25
2	2	4	0,0500	5,00	5	0,0625	6,25
3	3	58	0,7250	72,50	63	0,7875	78,75
4	4	12	0,1500	15,00	75	0,9375	93,75
5	5 ou mais	5	0,0625	6,25	80	1,0000	100,00
	Total	80	1,0000	100,00			

Exemplo: Tabela de frequências

Quantitativa contínua

Massa corporal da amostra de frequentadores da cantina (kg)

81,2	89,5	67,0	77,2	66,8	65,1	69,9	92,6	49,7	74,7
51,9	77,9	75,4	60,8	84,5	87,5	69,4	69,2	73,9	60,0
94,8	57,7	72,8	86,4	77,6	98,1	81,1	80,9	75,1	88,0
68,5	70,1	106,6	89,5	85,9	67,2	78,9	59,2	75,4	90,2
86,3	68,8	87,8	67,0	84,6	88,8	76,9	80,7	60,4	100,5
82,1	80,1	60,5	73,2	62,5	77,7	70,6	79,8	70,2	76,3
81,9	80,4	75,9	87,6	70,0	75,8	89,2	77,1	75,1	83,2
93,0	95,2	73,5	65,0	74,0	73,2	75,1	78,5	76,9	69,6

TABELA DE FREQUÊNCIA POR CLASSES

É necessária a construção de classes para obter a tabela de frequências de uma variável contínua ou de uma variável discreta que tenha muitos diferentes valores observados.

Tabela de frequência por classes

Classes: faixa de valores.

Nomenclatura:

- ✓ Número de classes: k
- ✓ Limite inferior da classe j : l_j
- ✓ Limite superior da classe j : L_j
- ✓ Amplitude dos dados: $A = \text{máximo} - \text{mínimo} = x_{(n)} - x_{(1)}$
- ✓ Ponto médio da classe j :

$$x_j^* = \frac{l_j + L_j}{2}$$

Número de classes

Conveniência: Escolher um número, geralmente, entre 5 e 20 que seja conveniente.

Regra do quadrado: $k \approx \sqrt{n}$

Notas

- ✔ Ao invés de escolher o número de classes conveniente, pode-se escolher as classes por conveniência.
- ✔ Para análise da distribuição da variável, a escolha por conveniência pode não ser adequada.

Construção da tabela de frequências por classes

- ✔ Ordenação dos dados: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$.
- ✔ Amplitude dos dados: $A = x_{(n)} - x_{(1)}$
- ✔ Calcular o número de classes (arredondar conforme a conveniência)
- ✔ Calcular a amplitude das classes: A_j (arredondar conforme a conveniência)
- ✔ Calcular os limites inferiores e superiores das classes: l_j e L_j , $j = 1, \dots, k$.
- ✔ Verificar se é necessário aumentar L_k ou acrescentar uma classe. Também podemos arredondar l_1 .
- ✔ Contagem das observações em cada classe.

Exemplo (continuação)

Dados ordenados: Massa corporal da amostra de frequentadores da cantina (kg)

49,7	51,9	57,7	59,2	60,0	60,4	60,5	60,8
62,5	65,0	65,1	66,8	67,0	67,0	67,2	68,5
68,8	69,2	69,4	69,6	69,9	70,0	70,1	70,2
70,6	72,8	73,2	73,2	73,5	73,9	74,0	74,7
75,1	75,1	75,1	75,4	75,4	75,8	75,9	76,3
76,9	76,9	77,1	77,2	77,6	77,7	77,9	78,5
78,9	79,8	80,1	80,4	80,7	80,9	81,1	81,2
81,9	82,1	83,2	84,5	84,6	85,9	86,3	86,4
87,5	87,6	87,8	88,0	88,8	89,2	89,5	89,5
90,2	92,6	93,0	94,8	95,2	98,1	100,5	106,6

Tamanho da amostra:
 $n = 80$

Regra do quadrado:

$$k = \sqrt{80} \approx 8,94$$

Escolhemos:

$$k = 8$$

Amplitude:

$$x_{(80)} - x_{(1)} = 56,9$$

Amplitude de cada classe:


$$A_j = A/k \approx 7,11$$

Escolhemos:

$$A_j = 7$$

Exemplo (continuação)

j	Classe	f_j	F_j	x_j^*
1				
2				
3				
4				
5				
6				
7				
8				
	Total	80		



Observações

- ▷ $\mathbf{a} \vdash \mathbf{b}$ equivale a $[a,b)$
- ▷ $\mathbf{a} \dashv \mathbf{b}$ equivale a $(a,b]$
- ▷ $\mathbf{a} \dashv \vdash \mathbf{b}$ equivale a $[a,b]$
- ▷ O ponto médio de cada classe, x_j^* será útil para o cálculo de algumas medidas.

**ANÁLISE
DESCRITIVA DE
VARIÁVEIS
QUALITATIVAS**

4.2

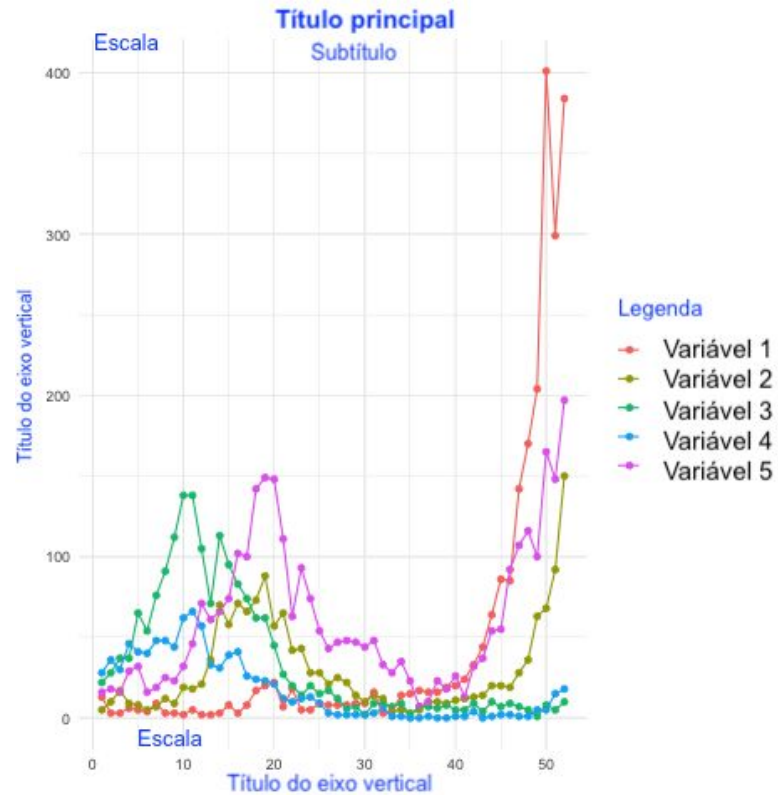
GRÁFICO

O que é? Representação visual de dados.

Para que serve? Facilitar o entendimento do que é apresentado (e também para analisar os resultados).

Como? Existem vários tipos de gráficos, e cada um é adequado a cada tipo de variável e situação.

ELEMENTOS DE UM GRÁFICO



TIPOS DE GRÁFICOS VARIÁVEIS QUALITATIVAS

- ✔ Gráfico de barras (*bar plot*)
- ✔ Gráfico de Pareto
- ✔ Gráfico de setores (pizza - *pie chart*)

NOTA

Muitas vezes mais de um gráfico é teoricamente adequado, mas apenas um deles é mais adequado na prática.

GRÁFICO DE BARRAS (OU COLUNAS)

É formado por retângulos na vertical (ou horizontal), cujas alturas (ou bases) são proporcionais às frequências dos atributos

NOTAS

- ✔ As frequências usadas podem ser a absoluta ou a relativa (em porcentagem ou em decimal).
- ✔ A largura (ou altura) da barra deve ser a conveniente.
- ✔ Podemos usar barras agrupadas, barras lado a lado, barras com um eixo em comum (espelho), barras empilhadas.
- ✔ Não é possível representar por um gráfico de barras uma variável quantitativa contínua. Se estiver em classes, pode ser representada por um outro tipo de gráfico.

Gráfico de barras para variável qualitativa

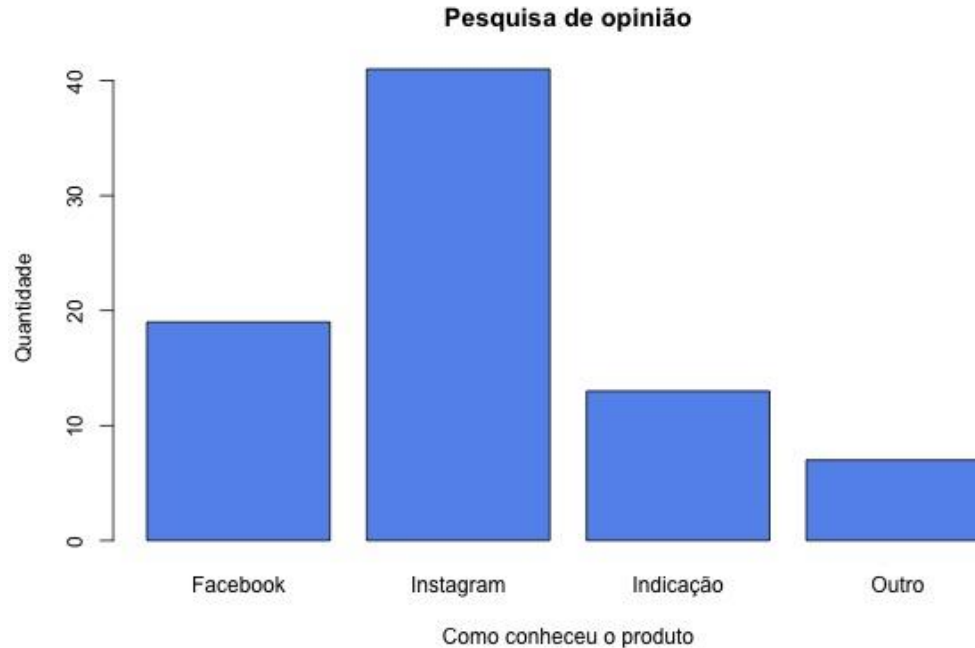


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto.

Gráfico de barras com a frequência absoluta e barras decrescentes

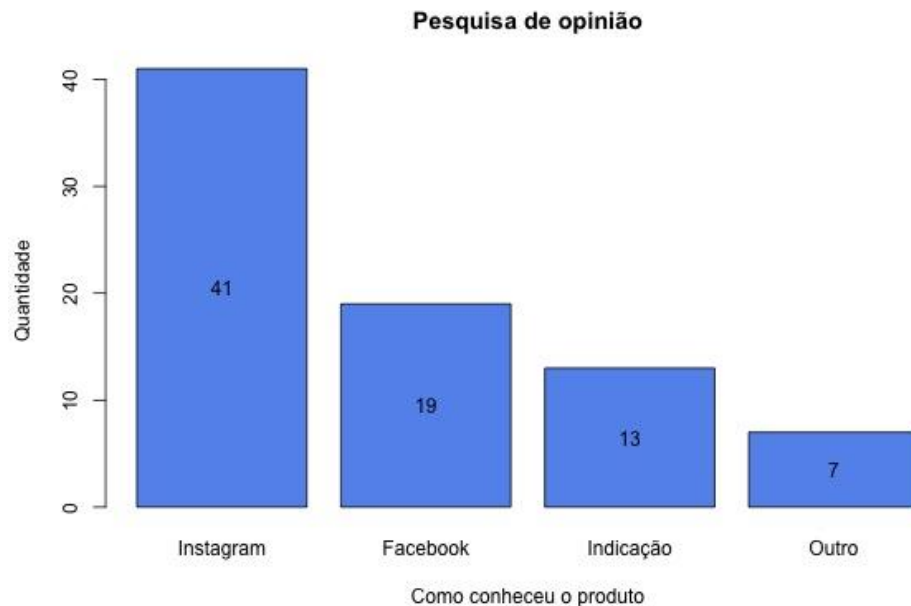


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto.

Gráfico de barras horizontais com a frequência relativa (%)

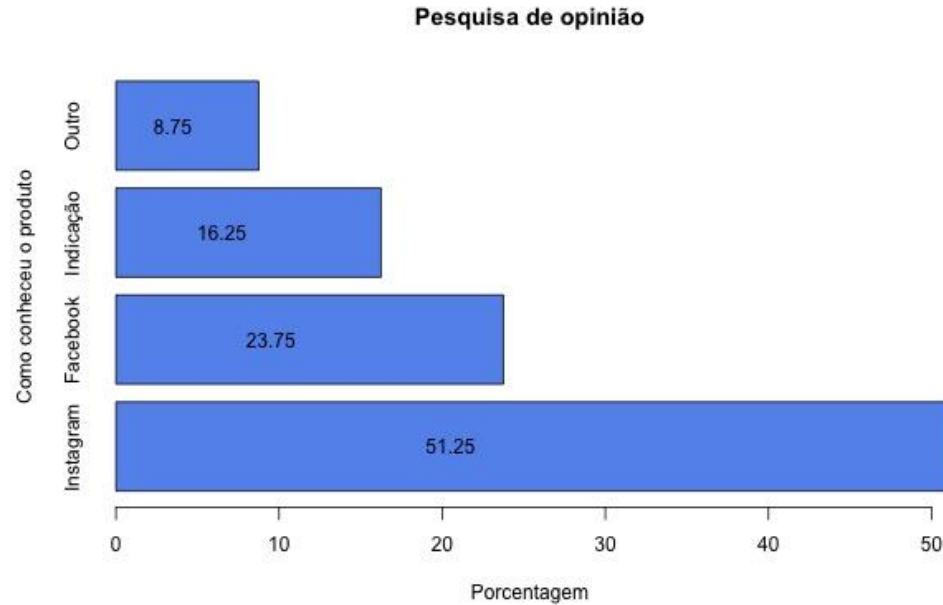


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto.

Gráfico de barras empilhadas

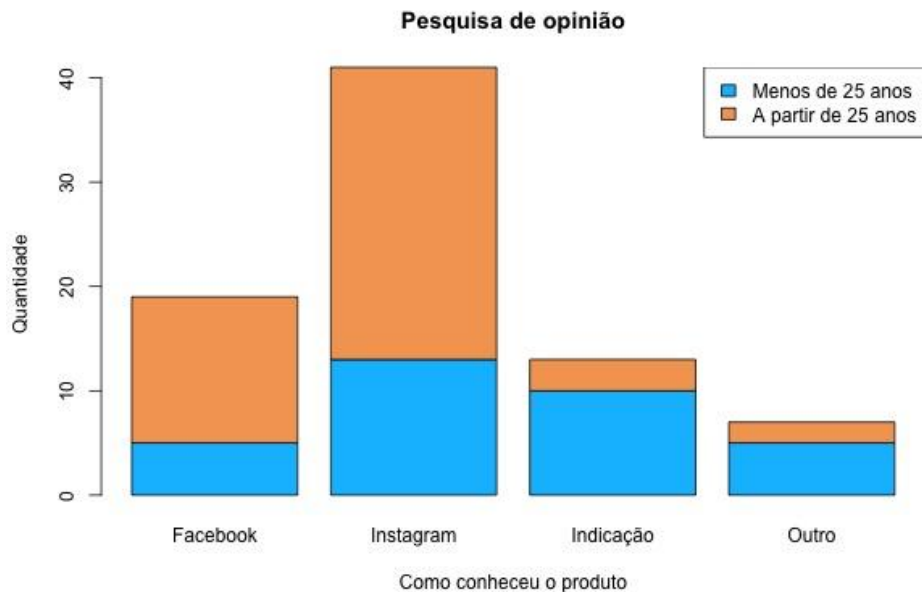


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto para diferentes faixas etárias.

Gráfico de barras empilhadas na horizontal (proporção)

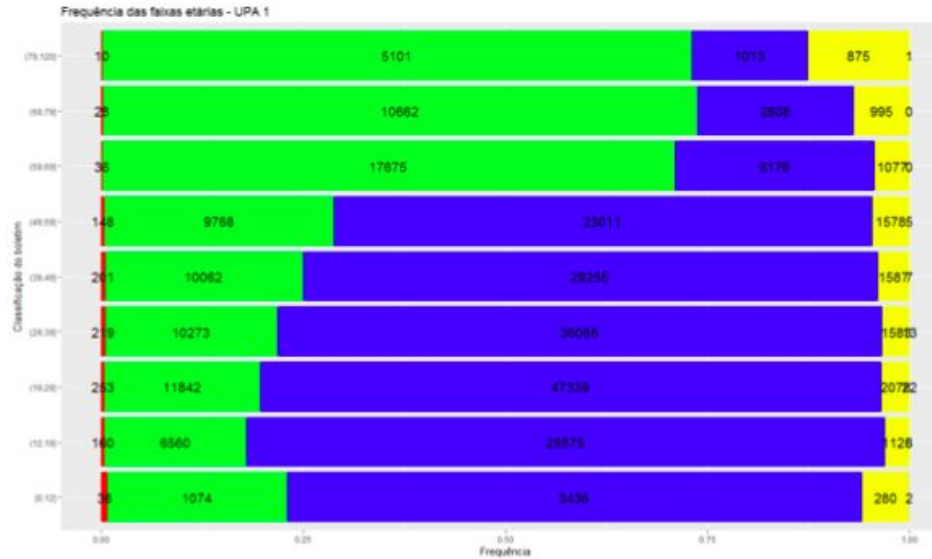


Figura: Números de atendimentos por faixa etária - UPA1 e UPA2.

Fonte: NEA por Matheus Toshio Hisatugu.

Gráfico de barras espelho

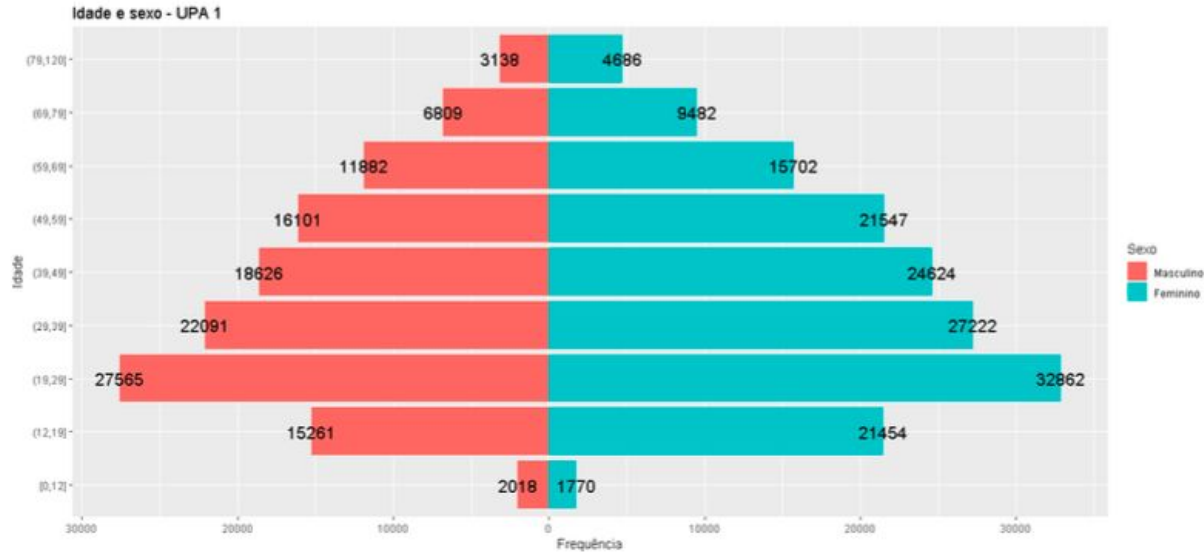


Figura: Idade e gênero - 2016 - UPA1, UPA2 e UPA3.

Fonte: NEA por Matheus Toshio Hisatugu.

Gráfico de barras espelho

Figura: Idade e gênero - 2016 - UPA1.

Fonte: NEA por Matheus Toshio Hisatugu.

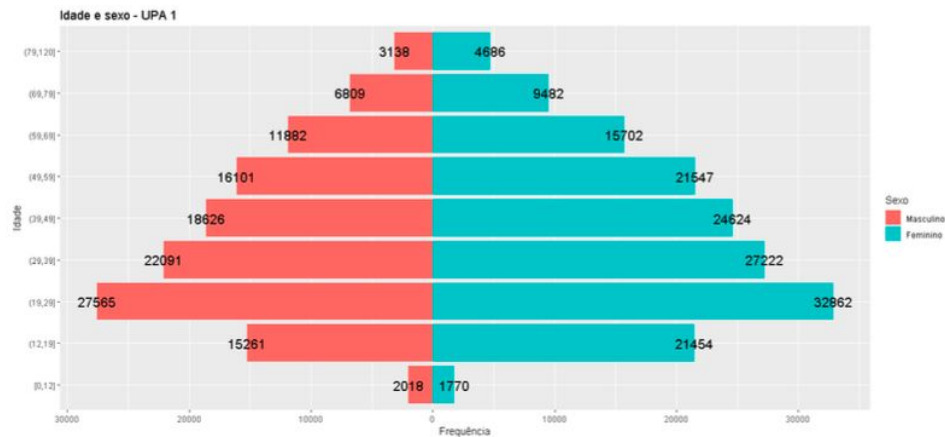


Figura: Idade e gênero - 2016 - UPA2.

Fonte: NEA por Matheus Toshio Hisatugu.

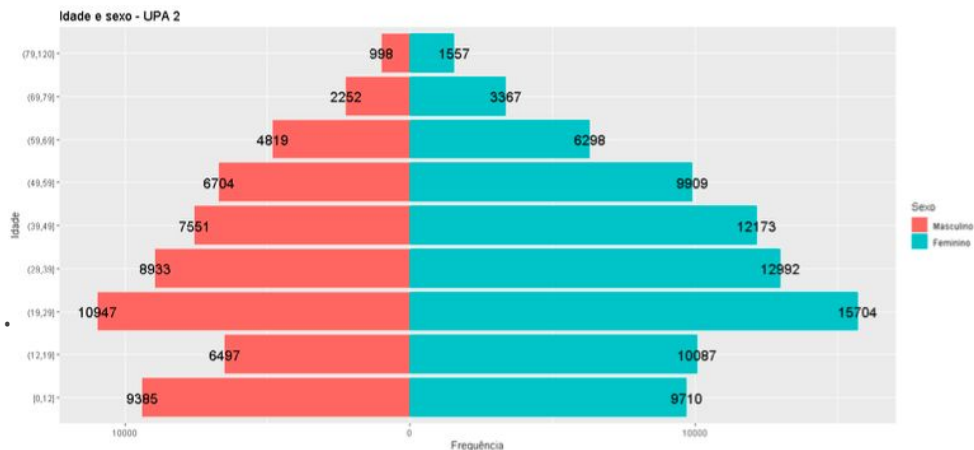


Gráfico de barras horizontais ordenadas por frequência

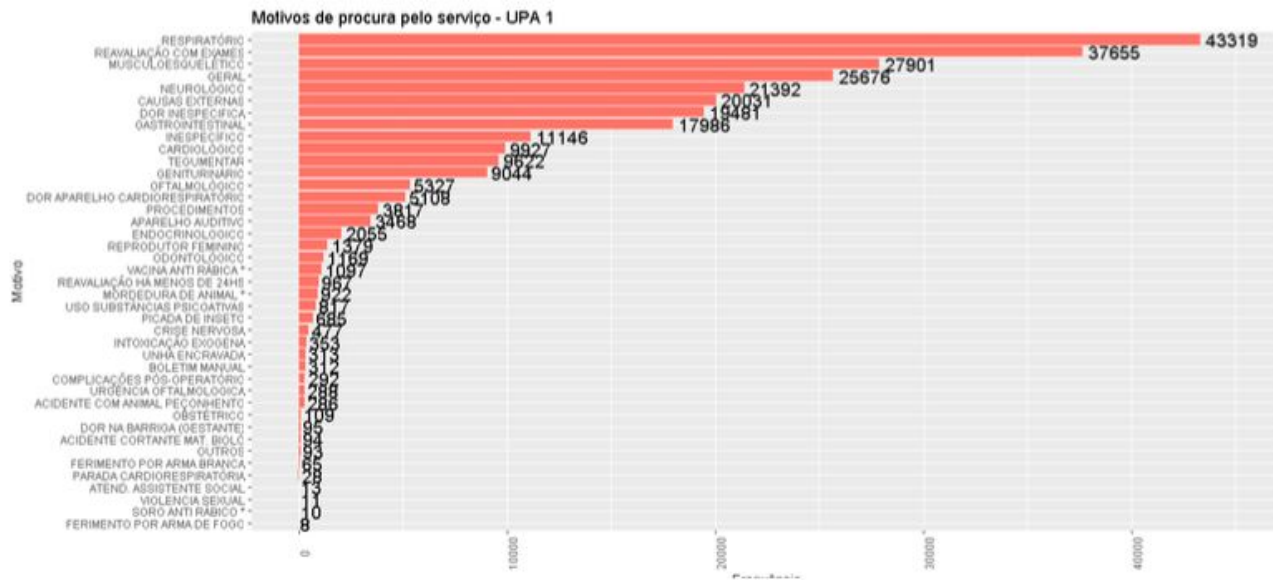


Figura: Motivos - 2016 - UPA1.3

Fonte: NEA por Matheus Toshio Hisatugu.

GRÁFICO DE PARETO

Gráfico de barras ordenadas de forma decrescente acrescido das frequências relativas acumuladas.

NOTAS

- ✔ Um eixo representa a frequência absoluta dos atributos e o outro eixo representa a frequência relativa, normalmente em porcentagem.
- ✔ O Gráfico de Pareto ficou famoso pelo Princípio de Pareto: 80% da riqueza de uma população pertencia a 20% da população.
- ✔ O Princípio de Pareto também é conhecido como Regra do 80/20, Lei dos Poucos Vitais ou Princípio de escassez do fator.
- ✔ A regra 80/20 mostra-se válida em diversas áreas: consumo, confiabilidade, finanças, etc.

Gráfico de Pareto

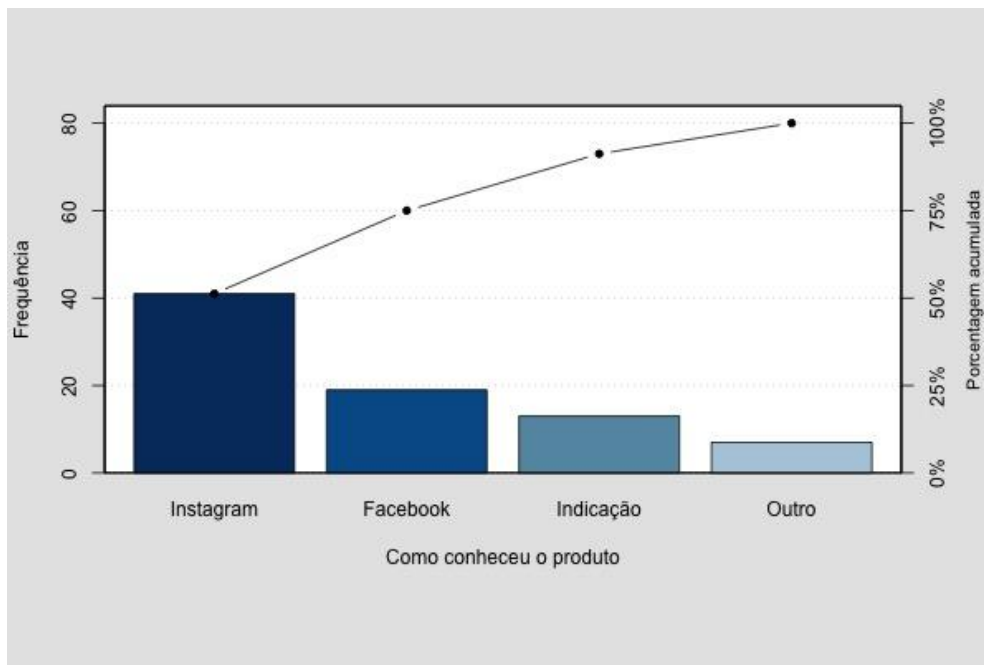


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto.

Exemplo

Tabela: Motivos das reclamações dos clientes de um restaurante.

Despesa	Quantidade R\$
Demora no atendimento inicial	3
Demora para trazer o pedido	92
Entrega de pedido incompleto ou errado	105
Conta errada	17
Falta de qualidade dos pratos	48
Temperatura da bebida	27
Atendimento ruim	20
Total	382

Exemplo

Tabela: Motivos das reclamações dos clientes de um restaurante.

Despesa	Quantidade R\$
Demora no atendimento inicial	3
Demora para trazer o pedido	92
Entrega de pedido incompleto ou errado	105
Conta errada	17
Falta de qualidade dos pratos	48
Temperatura da bebida	27
Atendimento ruim	20
Total	382

Exemplo: Gráfico de Pareto

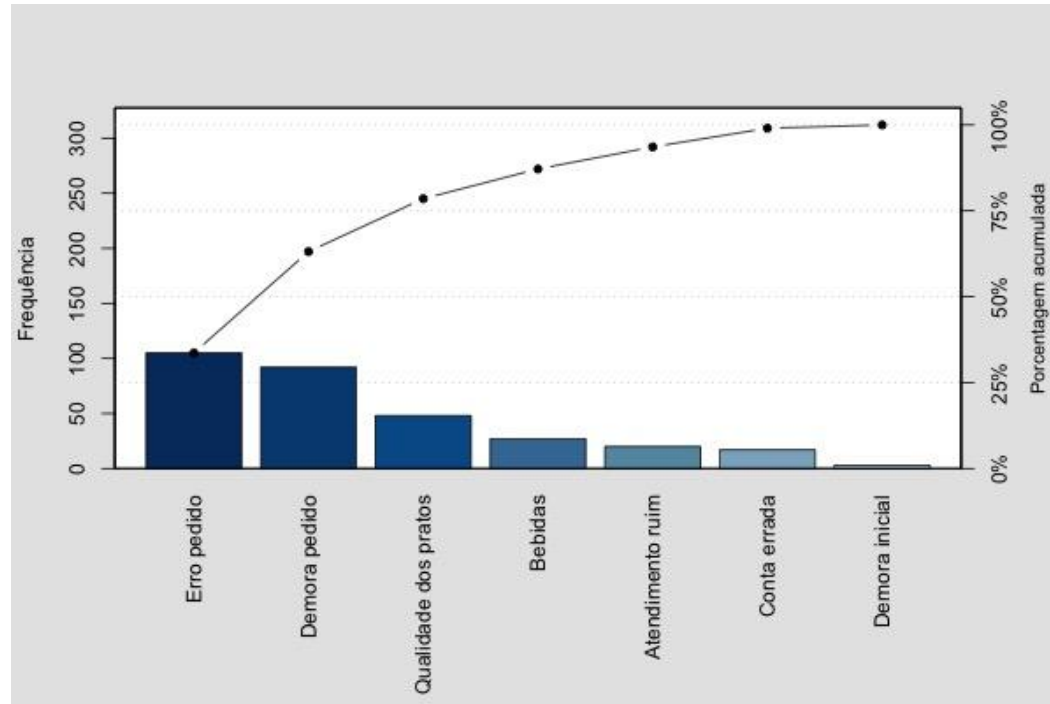


Figura: Motivos das reclamações dos clientes de um restaurante.

Exemplo

Tabela: Gastos mensais de um casal sem filhos.

Despesa	Valor R\$
Alimentação	1400,00
Cuidados com saúde (plano de saúde, academia, medicamentos, etc)	1260,00
Carro (combustível, seguro, pedágio, manutenção, etc)	1390,00
Diversão (viagens, passeios, restaurantes, etc)	1450,00
Casa (água, energia elétrica, aluguel, etc)	1400,00
Diversos (vestuário, <i>pet</i> , etc)	1500,00
Total	8400,00

Exemplo: Gráfico de Pareto

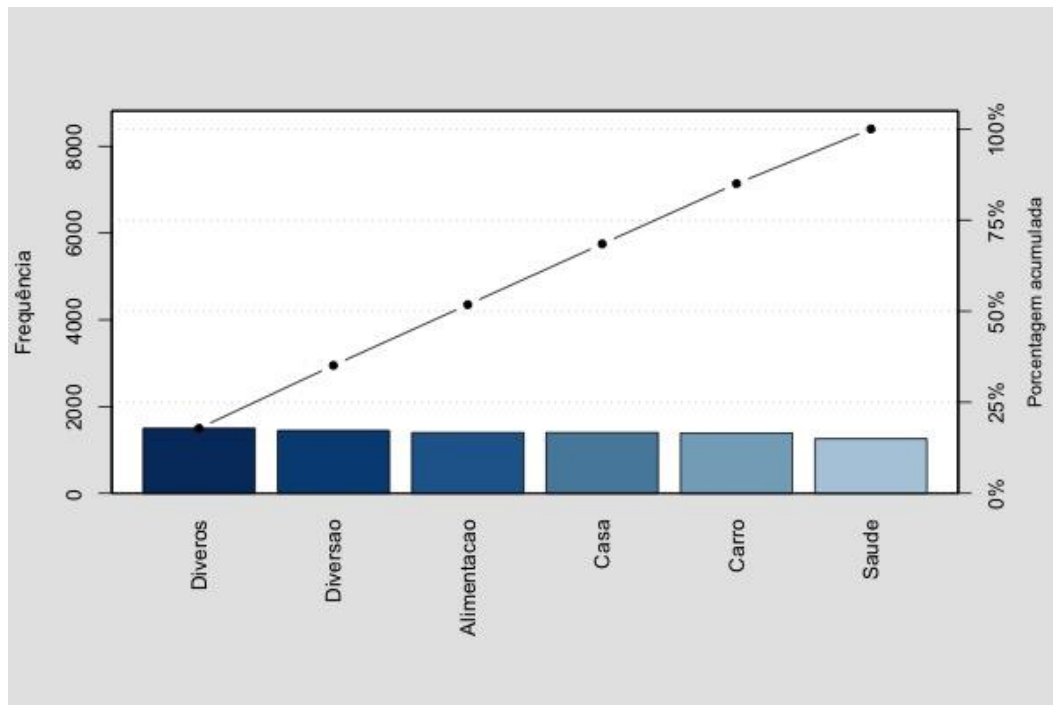


Figura: Gastos mensais de um casal.

NOTAS

- ✔ Quanto mais a curva da frequência acumulada está próxima do extremo superior esquerdo do gráfico, mais representativos são os atributos mais frequentes.
- ✔ Quanto mais a curva da frequência acumulada está próxima da reta identidade, mais igualmente frequentes são todos os atributos.

GRÁFICO DE SETORES

Gráfico circular sendo os atributos representados por arcos cujos ângulos (centrais) são proporcionais às suas frequências.

Gráfico de setores

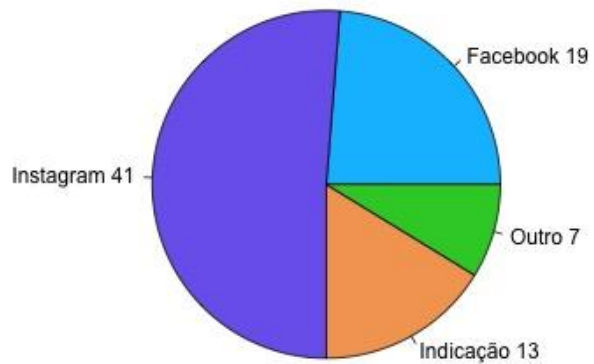


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto.

NOTAS

- ✔ Em geral, é difícil transformarmos a informação de setor em quantidade.
- ✔ Há uma grande dificuldade quando queremos comparar dois grupos, ou seja, dois gráficos de setores.
- ✔ Os atributos de menores frequências ficam pouco visíveis.

Gráfico de setores *versus* gráfico de barras

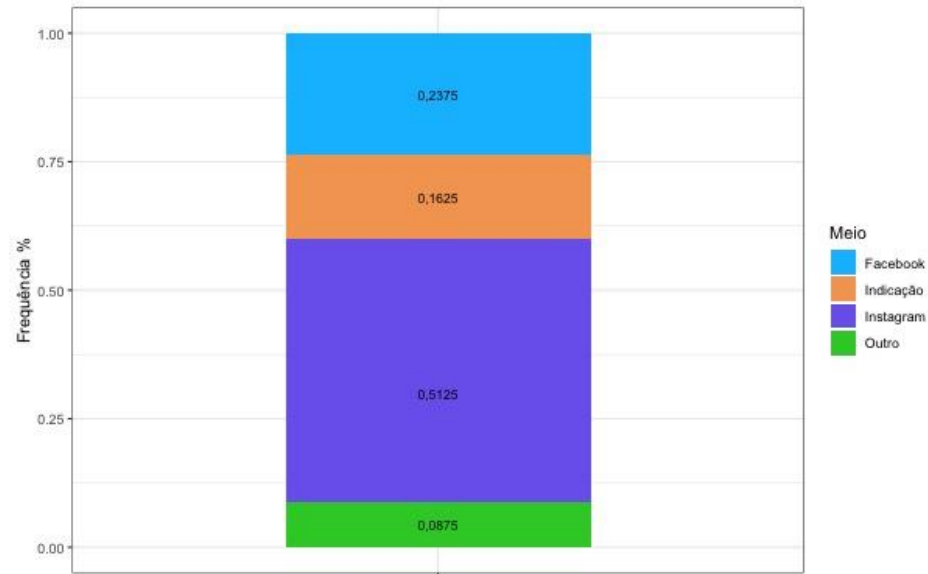


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto.

Gráfico de setores *versus* gráfico de barras

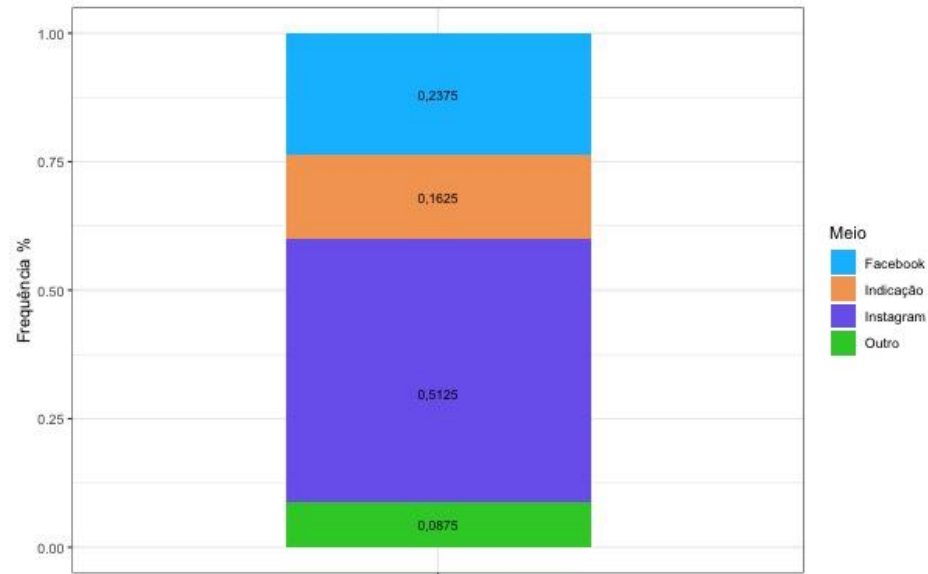


Figura: Meio pelo o qual o consumidor teve conhecimento ao produto.

Gráfico de setores *versus* gráfico de barras

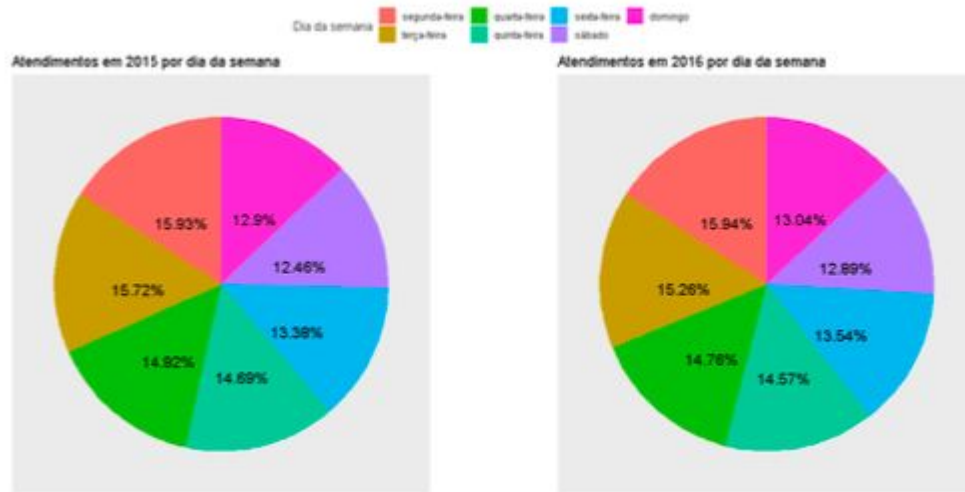


Figura: Proporções de atendimentos em 2015 e 2016 por dia da semana.

Fonte: Dados do Núcleo de Estatística Aplicada (NEA).

Gráfico de setores *versus* gráfico de barras

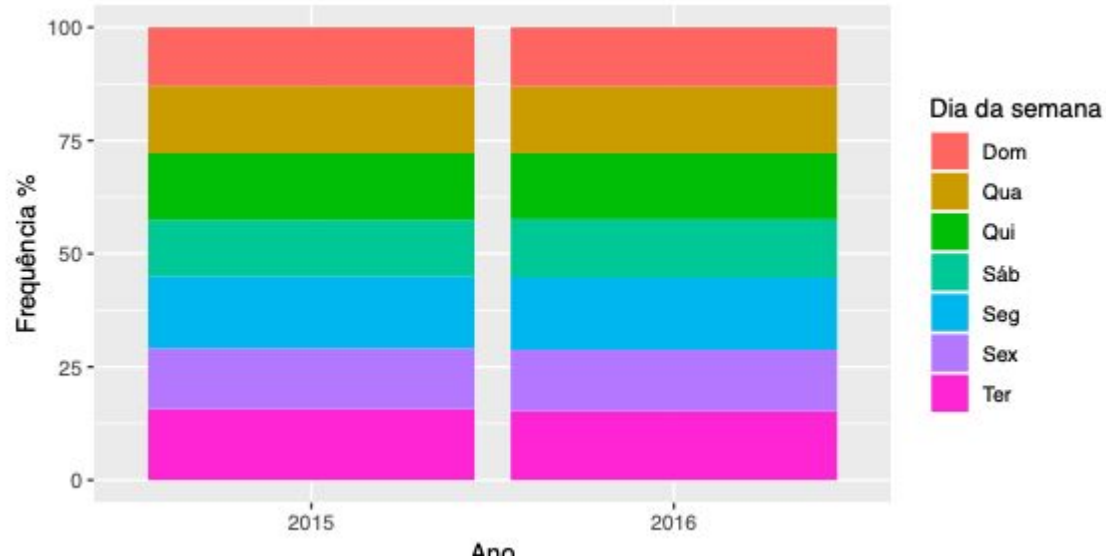


Figura: Proporções de atendimentos em 2015 e 2016 por dia da semana.

Fonte: Dados do Núcleo de Estatística Aplicada (NEA).

Cuidados com gráficos

Número de atendimentos em 2015

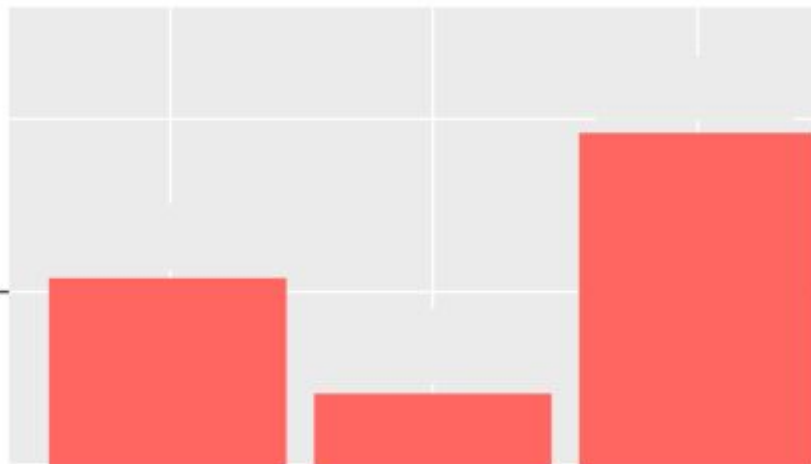


Figura: Número de atendimentos - 2015 - Janeiro a março.

Fonte: NEA por Matheus Toshio Hisatugu (adaptado).

Cuidados com gráficos

Número de atendimentos em 2015

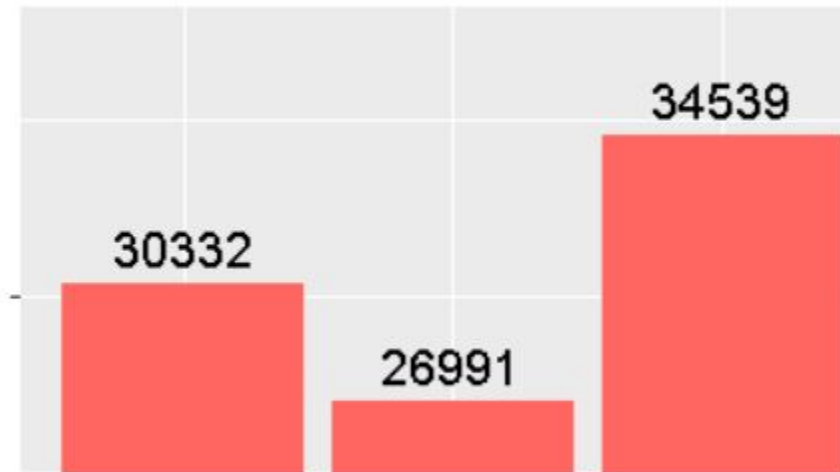


Figura: Número de atendimentos - 2015 - Janeiro a março.

Fonte: NEA por Matheus Toshio Hisatugu (adaptado).

Cuidados com gráficos

Figura: Número de atendimentos - 2015 - Janeiro a março.

Fonte: NEA por Matheus Toshio Hisatugu.

