

Estatística I

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Universidade Federal de Viçosa
Campus UFV - Viçosa



Sumário

- 1 Introdução
- 2 Gráficos para Variáveis Qualitativas
 - Gráfico de Barras
 - Gráfico de Pareto
 - Gráfico de Barras Empilhadas
 - Gráfico de Setores

Visualizações Gráficas

Gráficos e tabelas são uma constante em periódicos como jornais diários, revistas, periódicos técnicos e relatórios, acadêmicos ou não. Apesar disso, não existe uma teoria complexa sobre gráficos nos livros de Matemática e/ou Estatística. Na verdade, não existe muita teoria. No entanto, essa é uma parte essencial na formação de qualquer profissional. Na verdade, é essencial para a formação de qualquer cidadão.

Visualizações Gráficas

As técnicas, os conceitos e o conteúdo sobre visualizações gráficas fazem parte da Estatística Descritiva e, conforme **Unwin (2015)**, é parte essencial da Análise de Dados.

Visualizações Gráficas

As técnicas, os conceitos e o conteúdo sobre visualizações gráficas fazem parte da Estatística Descritiva e, conforme **Unwin (2015)**, é parte essencial da Análise de Dados.

Estatística Descritiva

A Estatística Descritiva emprega métodos numéricos e gráficos para investigar padrões em um conjunto de dados, resumir informações e apresentar resultados de maneira apropriada.

Visualizações Gráficas

As técnicas, os conceitos e o conteúdo sobre visualizações gráficas fazem parte da Estatística Descritiva e, conforme **Unwin (2015)**, é parte essencial da Análise de Dados.

Estatística Descritiva

A Estatística Descritiva emprega métodos numéricos e gráficos para investigar padrões em um conjunto de dados, resumir informações e apresentar resultados de maneira apropriada.

Um Gráfico Estatístico é uma representação visual dos dados e, tem a vantagem de, rápida e concisamente, informar sobre sua variabilidade.

Cuidado!

Existem vários tipos e formatos de gráficos e, tanto a escolha quanto a forma como são visualizados podem ter uma influência importante nas conclusões tiradas em relação a análise dos dados. Não há limites para o número de possibilidades de interpretações. Isso significa que você precisa adquirir experiência na criação e visualização de gráficos para aprender a apreciar o que eles podem e não podem mostrar.

Cuidado!

Existem vários tipos e formatos de gráficos e, tanto a escolha quanto a forma como são visualizados podem ter uma influência importante nas conclusões tiradas em relação a análise dos dados. Não há limites para o número de possibilidades de interpretações. Isso significa que você precisa adquirir experiência na criação e visualização de gráficos para aprender a apreciar o que eles podem e não podem mostrar.

Não é o que você olha que importa, é o que você vê.

Henry David Thoreau

Visualizações Gráficas

Assim, a depender do tipo de variável considerada, temos diferentes tipos de gráficos. Veremos alguns a partir de agora!

Tomemos como ilustração a variável Y: grau de instrução da Tabela [CompanhiaMB](#). Para organizar os dados provenientes de uma variável qualitativa, é usual fazer uma tabela de frequências, como a Tabela abaixo, antes de construir os gráficos.

Distribuições de Frequências

Tabela: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução.

Grau de instrução	Frequência n_i	Proporção f_i	Porcentagem $100f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

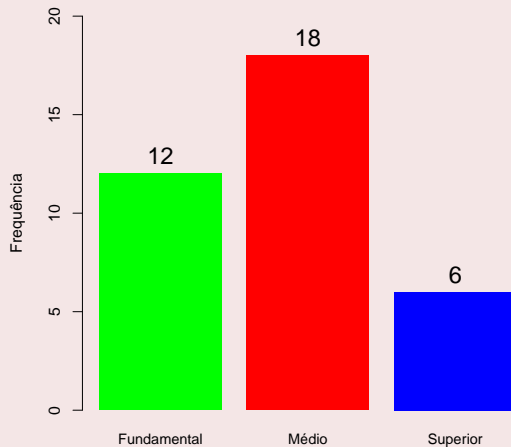
Fonte: Morettin and Bussab (2017)

Gráfico de Barras

O gráfico em barras consiste em construir retângulos ou barras, em que uma das dimensões é proporcional à magnitude a ser representada (n_i ou f_i), sendo a outra arbitrária, porém igual para todas as barras. Essas barras são dispostas paralelamente umas às outras, horizontal ou verticalmente. Na próxima Figura temos o gráfico em barras (verticais) para a variável “Grau de Instrução”.

Gráfico de Barras

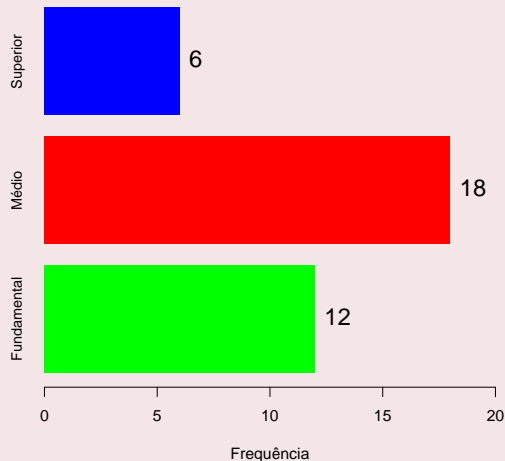
FIG 1: Gráfico em barras para a variável Y: grau de instrução.



Fonte: Morettin (2017)

Gráfico de Barras

FIG 1: Gráfico em barras para a variável Y: grau de instrução.



Fonte: Morettin (2017)

Gráficos de Barras

O conjunto de dados Fleiss93 do pacote meta contém detalhes de sete estudos sobre o uso de aspirina após infarto do miocárdio. A Figura abaixo representa um gráfico de barras dos tamanhos dos estudos, com os estudos ordenados pelo número total de pacientes.

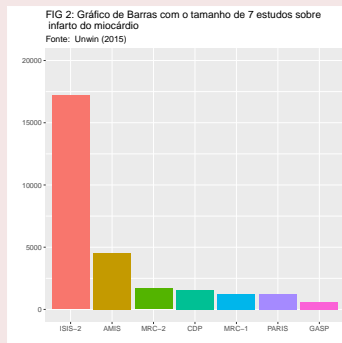
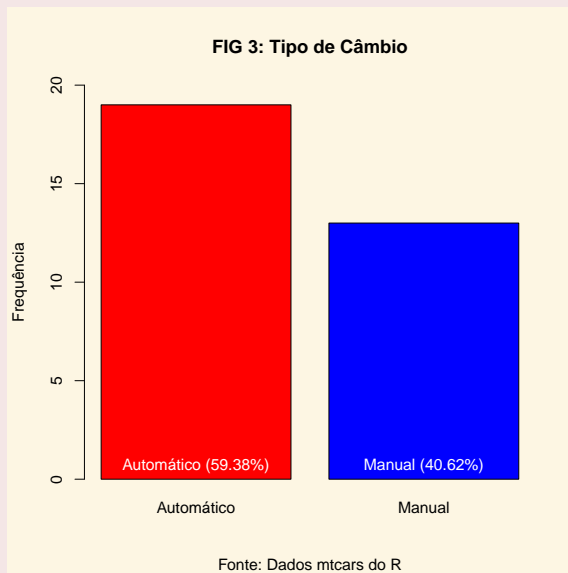


Gráfico de Barras



Gráficos de Barras

Os gráficos de barras são talvez o tipo de visualização de dados mais comumente usado. No entanto, para a representação de variáveis qualitativas, há também o gráfico de setores, popularmente conhecido como gráfico de pizza.

Gráfico de Pareto

Um gráfico de Pareto é um gráfico de barras em que as barras são ordenadas da maior frequência de ocorrência para a menor frequência de ocorrência. No gráfico de Pareto também acrescentamos uma linha acima das barras com a frequência acumulada da variável.

Gráfico de Pareto

FIG 2: Gráfico de Pareto para a variável grau de instrução

Fonte: Elaborado pelo Autor

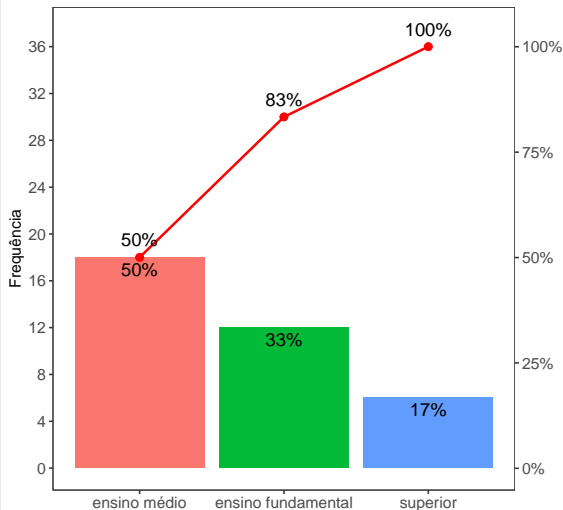


Gráfico de Barras Empilhadas

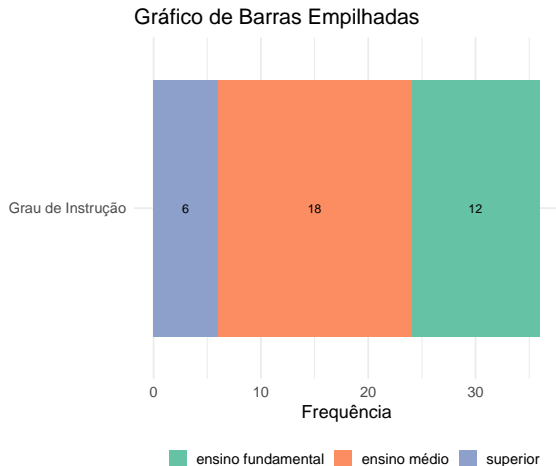


Gráfico de Barras Empilhadas

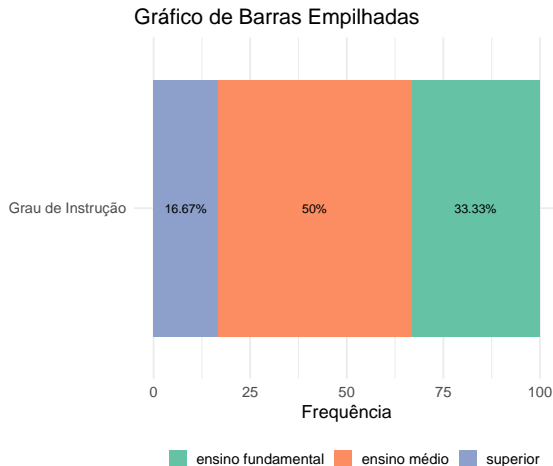
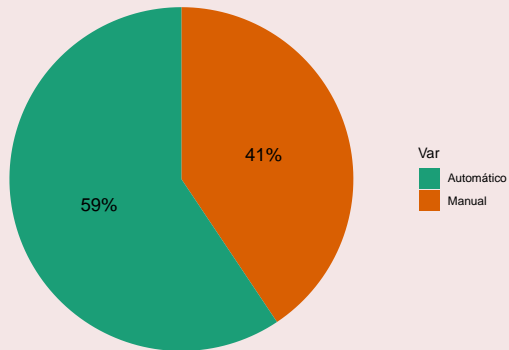


Gráfico de Setores

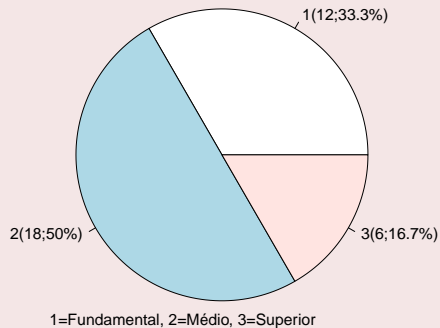
O gráfico em setores é comumente utilizado para representar parte de um todo, geralmente em percentagens. Ele é bastante apropriado para mostrar frequências de ocorrências de variáveis qualitativas.

Gráfico de Setores

Tipo de Câmbio (Pacote mtcars)



**Gráfico em setores para a variável
Y: grau de instrução.**



Fonte: Morettin (2017)

Um procedimento alternativo para resumir um conjunto de valores, com o objetivo de se obter uma idéia da forma de sua distribuição, é o ramo-e-folhas. Uma vantagem deste diagrama é que não perdemos (ou perdemos pouca) informação sobre os dados em si.

Diagrama de ramos e folhas para variáveis contínuas

Quando o número de observações é relativamente grande, este diagrama pode ser útil.

Tabela: Diagrama de Ramos e Folhas da idade

Ramo	Folhas																
2	0	3	5	6	6	7	8	9									
3	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	9
4	0	0	1	1	2	3	3	4	6	8							

Tabela: Diagrama de Ramos e Folhas dos Salários (\times sal. Min)

Ramo	Folhas			
4	00	56		
5	25	73		
6	26	66	86	
7	39	44	59	
8	12	46	74	95
9	13	35	77	80
10	53	76		
11	06	59		
12	00	79		
13	23	60	85	
14	69	71		
15	99			
16	22		61	
17	26			
18	75			
19	40			
20				
21				
22				
23		30		

Algumas informações que se obtêm deste ramo-e-folhas são:

- 1 Há um destaque grande para o valor 23,30.

Algumas informações que se obtêm deste ramo-e-folhas são:

- 1 Há um destaque grande para o valor 23,30.
- 2 Os demais valores estão razoavelmente concentrados entre 4,00 e 19,40.

Algumas informações que se obtêm deste ramo-e-folhas são:

- 1 Há um destaque grande para o valor 23,30.
- 2 Os demais valores estão razoavelmente concentrados entre 4,00 e 19,40.
- 3 Um valor mais ou menos típico para este conjunto de dados poderia ser, por exemplo, 10,00.

Algumas informações que se obtêm deste ramo-e-folhas são:

- 1 Há um destaque grande para o valor 23,30.
- 2 Os demais valores estão razoavelmente concentrados entre 4,00 e 19,40.
- 3 Um valor mais ou menos típico para este conjunto de dados poderia ser, por exemplo, 10,00.
- 4 Há uma leve assimetria em direção aos valores grandes; a suposição de que estes dados possam ser considerados como amostra de uma população com distribuição simétrica, em forma de sino (a chamada distribuição normal), pode ser questionada.

Referências

- P. A. Morettin and W. O. Bussab. *Estatística básica*. Saraiva Educação, 2017.
- A. Unwin. *Graphical data analysis with R*, volume 27. CRC Press, 2015.