

AULA 03

Resumos e Gráficos de Dados

Ernesto F. L. Amaral

16 de agosto de 2011
Metodologia de Pesquisa (DCP 854B)

Fonte:

Triola, Mario F. 2008. “Introdução à estatística”. 10^a ed. Rio de Janeiro: LTC. Capítulo 2 (pp.32-59).

ESQUEMA DA AULA

- Distribuição de frequência.
- Histogramas.
- Gráficos estatísticos.

VISÃO GERAL

- Apresentação de métodos para organização, resumo e obtenção de gráficos de conjuntos de dados.
- Objetivo principal é a compreensão dos dados e não simplesmente obtenção de tabelas e gráficos.
- Ao descrever, explorar ou comparar dados é bom saber:
 - **Centro:** valor representativo ou médio, que indica o meio do conjunto de dados.
 - **Variação:** medida de quanto valores variam no banco.
 - **Distribuição:** forma da distribuição dos dados.
 - **Valores discrepantes (*outliers*):** valores amostrais que se localizam longe da grande maioria dos outros valores.
 - **Tempo:** características dos dados que mudam com o tempo.

SUGESTÃO

- Com o uso de programas computacionais e calculadoras, não é preciso memorizar fórmulas e realizar cálculos aritméticos manuais.
- É mais importante entender o sentido prático desses cálculos e resultados, a partir de um pensamento crítico.
- Em alguns casos serão colocados passos detalhados, mas não é preciso dominá-los completamente.
- Porém, é importante entender a lógica dos cálculos e, até mesmo, realizar cálculos manuais, aumentando a compreensão dos resultados obtidos.

DISTRIBUIÇÃO DE FREQUÊNCIA

DISTRIBUIÇÃO DE FREQUÊNCIA

- Uma distribuição de frequência (tabela de frequência) lista os valores dos dados (individualmente ou por grupos de intervalos), juntamente com suas frequências correspondentes.
- Porque construir tabelas de frequência?
 - São úteis quando trabalhamos com grandes conjuntos de dados, os quais podem ser resumidos.
 - Permitem compreensão sobre natureza dos dados.
 - Servem como base para pensar e construir gráficos.
- Os detalhes de construção das tabelas podem não ser tão importantes, mas sim o que elas nos dizem sobre os conjuntos de dados.

IDADE (EM ANOS) DAS MELHORES ATRIZES DOS PRÊMIOS DA ACADEMIA (OSCAR)

22	37	28	63	32	26	31	27	27	28
30	26	29	24	38	25	29	41	30	35
35	33	29	38	54	24	25	46	41	28
40	39	29	27	31	38	29	25	35	60
43	35	34	34	17	37	42	41	36	32
41	33	31	74	33	50	38	61	21	41
26	80	42	19	33	35	45	49	39	34
26	25	33	35	35	28				

DISTRIBUIÇÃO DE FREQUÊNCIA: IDADE DAS MELHORES ATRIZES

Idade das Atrizes	Frequência
21-30	28
31-40	30
41-50	12
51-60	2
61-70	2
71-80	2

TERMOS PADRÕES

- **Limites inferiores de classe** são os menores números que podem pertencer às diferentes classes (21, 31, ..., 71).
- **Limites superiores de classe** são os maiores números que podem pertencer às diferentes classes (30, 40, ..., 80).
- **Fronteiras de classe** são os números usados para separar as classes, mas sem os saltos criados pelos limites de classe (30,5, 40,5, ..., 70,5). São úteis para construir histogramas.
- **Pontos médios das classes** são os pontos médios dos intervalos que determinam cada classe (25,5, 35,5, ..., 75,5). Igual a soma do limite inferior e superior, dividido por 2.
- **Amplitude de classe** é a diferença entre dois limites inferiores de classe consecutivos ou duas fronteiras inferiores de classe consecutivas (no exemplo, igual a 10).

CONSTRUINDO TABELA DE FREQUÊNCIA MANUALMENTE

- Escolher o número de classes mais apropriado à análise.
- Amplitude $\approx [(\text{valor máximo}) - (\text{valor mínimo})] / n^{\circ}$ classes
- Ponto inicial para limite inferior da primeira classe (pode ser o valor mínimo ou um valor um pouco menor).
- Com limite inferior da primeira classe e amplitude de classe, defina outros limites inferiores.
- Defina os limites superiores de classe.
- Percorra o banco de dados para identificar à qual classe o valor pertence.
- Observações:
 - Classes não devem se sobrepor.
 - Apresentar todas classes, mesmo as com nenhum valor.
 - Tentar usar mesma amplitude para classes.

DISTRIBUIÇÃO DE FREQUÊNCIA RELATIVA

- As frequências relativas são encontradas pela divisão entre cada frequência de classe pelo total de todas frequências.
- Frequências relativas são geralmente expressas como percentuais.
- A soma das frequências relativas deve totalizar 1 (ou 100%), com possibilidade de discrepância devido a arredondamento.
- Frequências relativas facilitam entender distribuição dos dados e comparar diferentes conjuntos.

Idade das Atrizes	Frequência	Frequência relativa
21-30	28	37%
31-40	30	39%
41-50	12	16%
51-60	2	3%
61-70	2	3%
71-80	2	3%
Total	76	101%

DISTRIBUIÇÃO DE FREQUÊNCIA ACUMULADA

- A frequência acumulada para uma classe é a soma da frequência daquela classe mais as frequências de todas as classes anteriores.

Idade das Atrizes	Frequência	Frequência relativa	Frequência acumulada
21-30	28	37%	28
31-40	30	39%	58
41-50	12	16%	70
51-60	2	3%	72
61-70	2	3%	74
71-80	2	3%	76
Total	76	101%	---

INTERPRETANDO DISTRIBUIÇÕES DE FREQUÊNCIA

- Um objetivo importante na realização de tabelas de frequência é a identificação da natureza da distribuição.
- Procuramos dados com distribuição normal, o que é importante na estatística.
- Podemos usar tabela de frequência para ajudar a determinar se dados têm distribuição normal.
- Característica essencial da distribuição normal é que seu gráfico tem forma de “sino”, com frequências que começam baixas, crescem até um máximo e depois decrescem.

DETERMINANDO SE DISTRIBUIÇÃO É NORMAL

- Frequências começam baixas, crescem até uma frequência máxima e depois decrescem para uma frequência baixa.
- Distribuição deve ser aproximadamente simétrica, com frequências igualmente distribuídas em ambos os lados da frequência máxima.

HISTOGRAMAS

HISTOGRAMAS

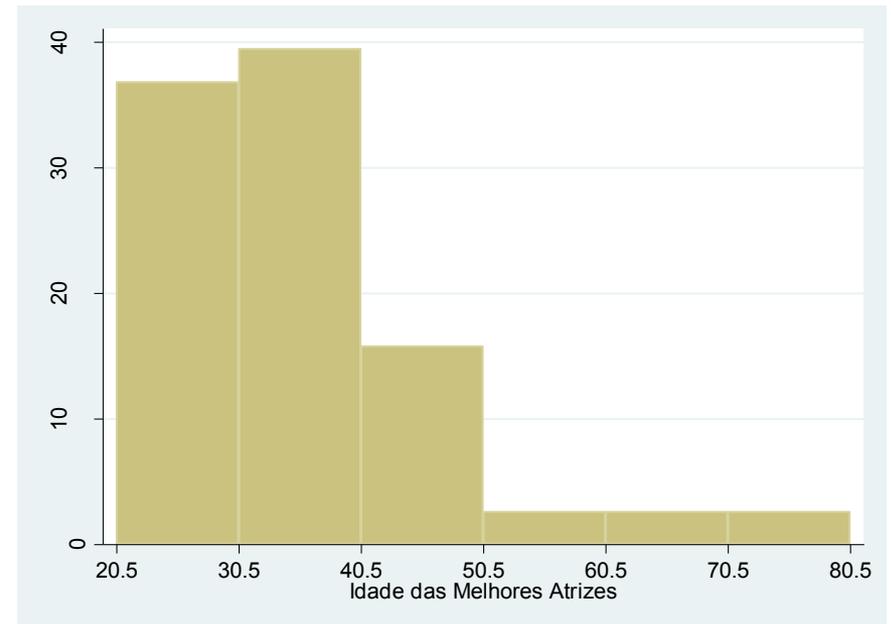
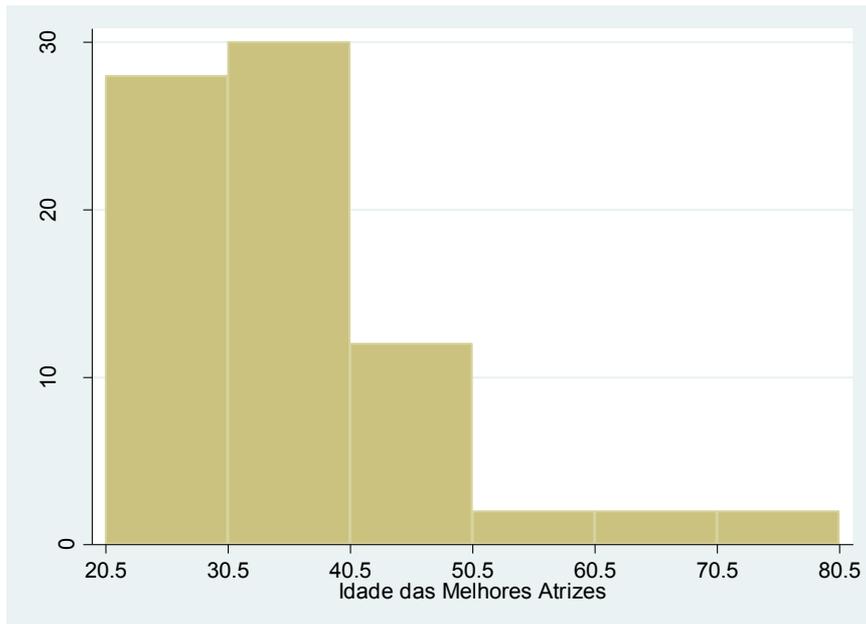
- Histograma é um gráfico de barras em que a escala horizontal representa classes de valores de dados e a escala vertical representa frequências.
- As alturas das barras correspondem aos valores das frequências, e as barras são desenhadas adjacentes umas às outras (sem separação).
- Não é importante dominar os procedimentos mecânicos para sua construção, já que programas de estatística fazem isso automaticamente.
- Devemos utilizar histogramas para compreender a natureza da distribuição dos dados.
- Histogramas podem apresentar frequências relativas (soma das alturas igual a 1 ou 100%) ou densidade (área igual a 1).

CONSTRUINDO HISTOGRAMAS

- O histograma é uma versão gráfica da tabela de frequência.
- Na escala horizontal, cada barra do histograma é marcada com seu limite inferior de classe à esquerda e seu limite superior de classe à direita.
 - Também podemos usar fronteiras de classe ou pontos médios das classes.
- Quanto às escalas dos eixos:
 - A frequência máxima deve sugerir o valor para o topo da escala vertical, sendo que zero deve estar na base.
 - Escala horizontal deve ser subdividida de modo a permitir que todas classes se encaixem bem.
 - Eixos devem ser identificados.
- Número de classes deve ficar entre 5 e 20.

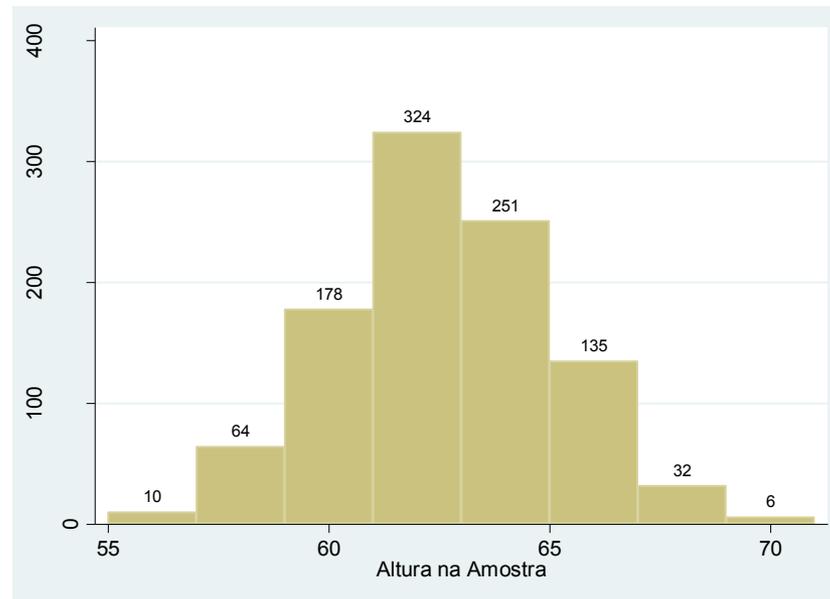
HISTOGRAMAS COM BASE NAS FREQUÊNCIAS

Idade das Atrizes	Frequência	Frequência relativa
21-30	28	37%
31-40	30	39%
41-50	12	16%
51-60	2	3%
61-70	2	3%
71-80	2	3%
Total	76	101%



HISTOGRAMA COM DISTRIBUIÇÃO NORMAL

- Frequências aumentam, atingem máximo e depois caem.
- Há simetria entre duas metades dos gráficos.
- Muitos métodos estatísticos exigem que dados amostrais sejam provenientes de uma população que tenha uma distribuição próxima da normal.
- Histograma de 1000 alturas de mulheres selecionadas aleatoriamente:



GRÁFICOS ESTADÍSTICOS

GRÁFICOS ESTATÍSTICOS

- O principal objetivo de gráficos é de entender melhor um conjunto de dados, revelando de forma eficaz características importantes destes dados.
- É importante ser também inovador na criação de gráficos que captem características-chave dos dados.
- Alguns destes gráficos são polígonos de frequência, ogivas, gráficos de pontos, ramo e folhas, gráficos de Pareto, gráficos de setores, diagramas de dispersão e gráficos de séries temporais.

EXEMPLOS DE GRÁFICOS

- **Polígono de frequência** usa segmentos de reta ligados a pontos localizados diretamente acima dos valores dos pontos médios de classe.
- Alturas dos pontos correspondem às frequências ou frequências relativas das classes.

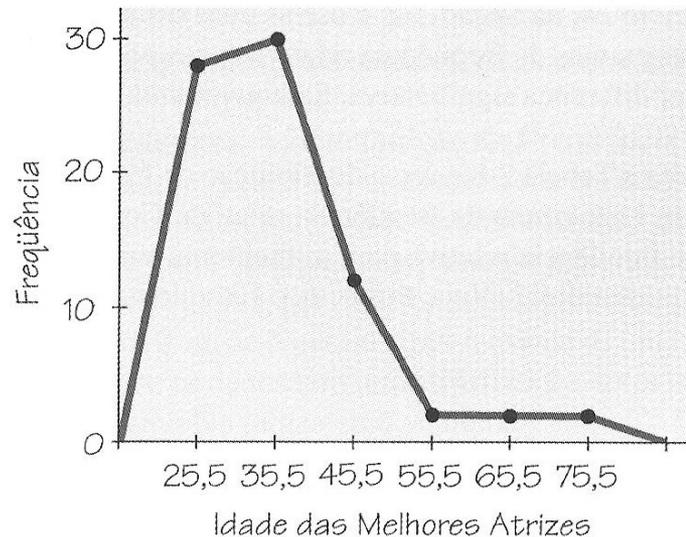


FIGURA 2-4 Polígono de Frequências

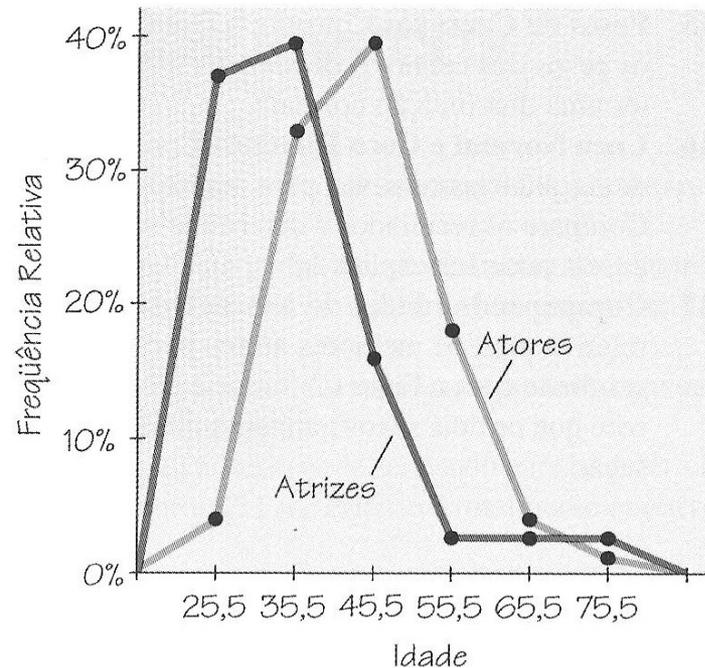
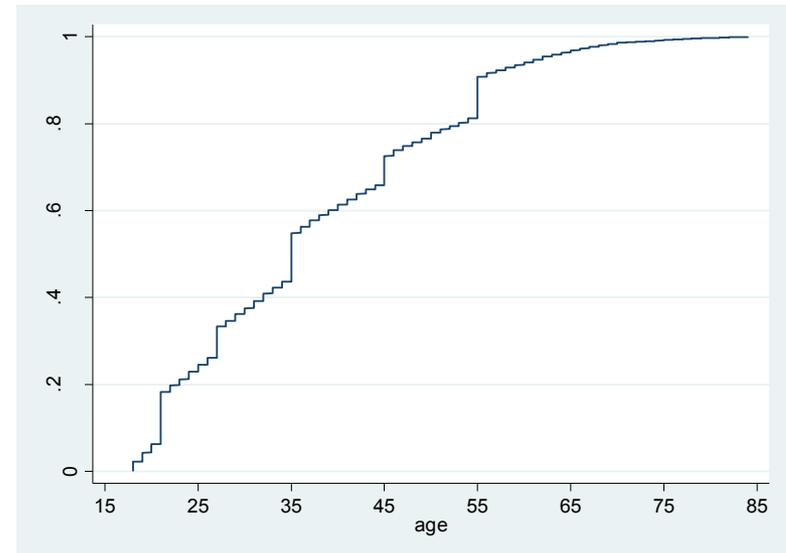


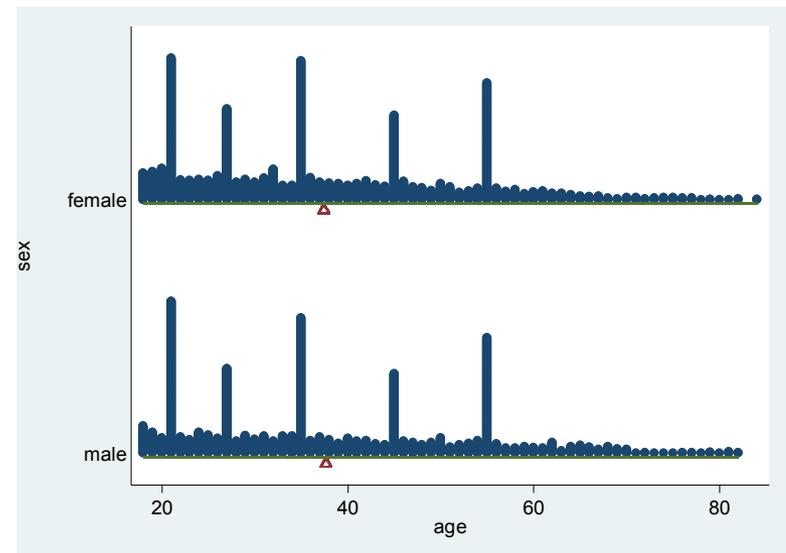
FIGURA 2-5 Polígonos de Frequências Relativas

MAIS GRÁFICOS...

– **Ogiva** é um gráfico de linha que representa frequências acumuladas. Útil na determinação de valores abaixo de certo valor.

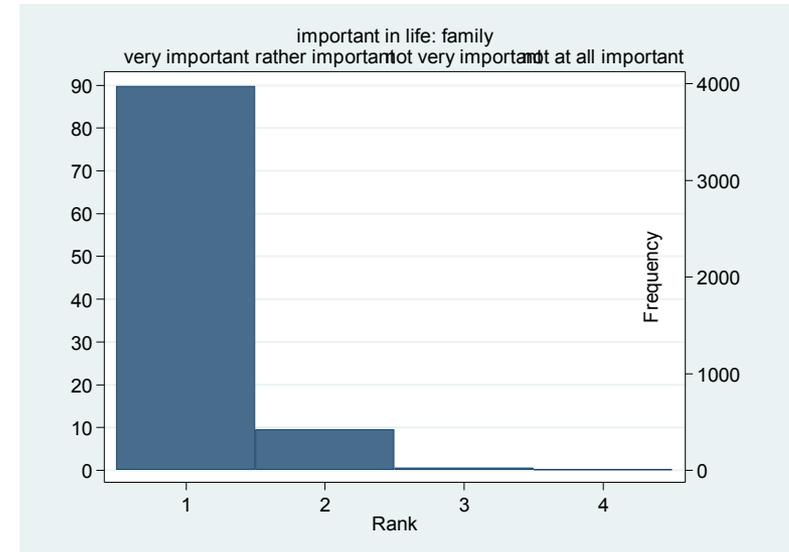


– **Gráfico de pontos** é um gráfico em que cada valor é ilustrado como um ponto ao longo de uma escala de valores. Pontos que representam valores iguais são empilhados.

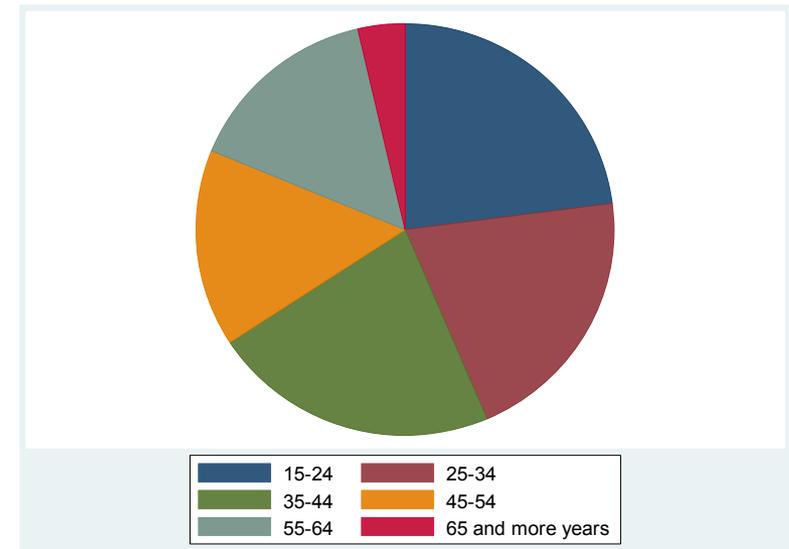


MAIS UM POUCO...

– **Gráfico de Pareto** é um gráfico de barras para dados qualitativos, com as barras dispostas em ordem pela frequência.

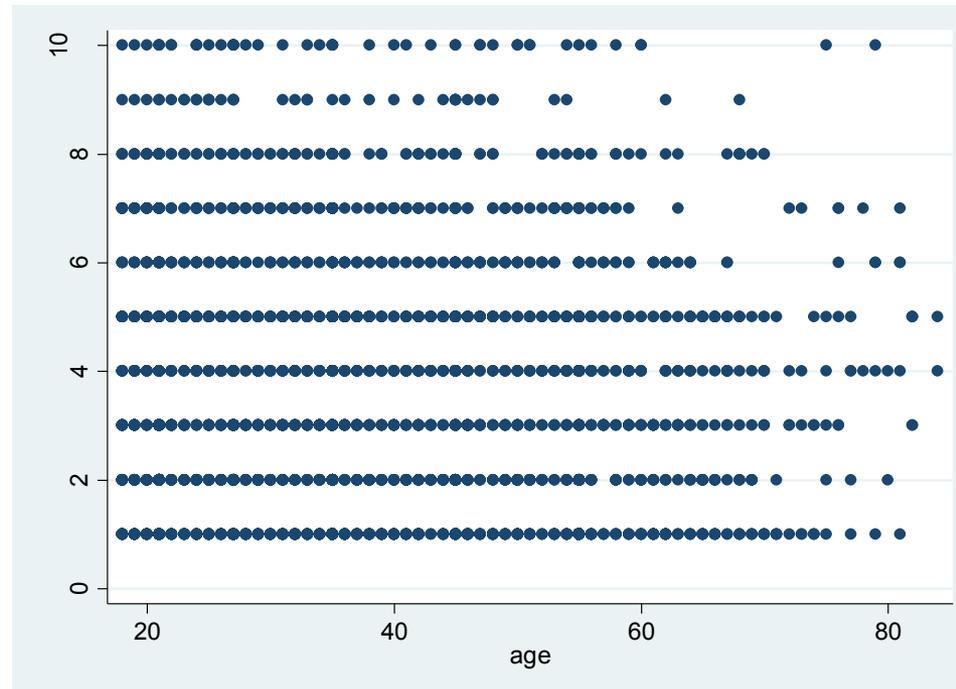


– **Gráfico de setores** são usados para retratar dados qualitativos como setores de um círculo (gráfico de pizza).



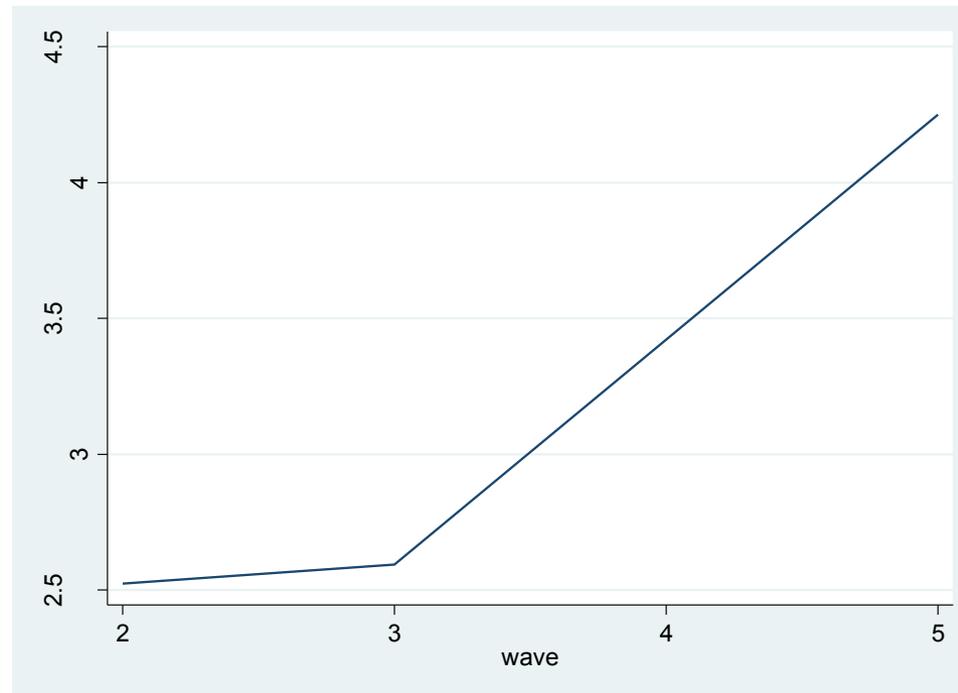
DIAGRAMAS DE DISPERSÃO

- Diagrama de dispersão é um gráfico de pares de dados (x, y), com um eixo x horizontal e um eixo y vertical.
- Dados são colocados em pares que combinam cada valor de um conjunto de dados com um valor correspondente de um segundo conjunto de dados.
- Exemplo ilustra renda (y) por idade (x):



GRÁFICOS DE SÉRIES TEMPORAIS

- Gráfico de série temporal ilustra mudanças nos valores de uma determinada variável ao longo do tempo.
- O exemplo abaixo mostra a média de renda (y) por diferentes anos (x).



DICAS IMPORTANTES

- Para pequenos conjuntos de dados, use uma tabela ao invés de gráfico.
- Gráfico deve fazer o observador concentrar-se na verdadeira natureza dos dados, e não em elementos que chamam a atenção, mas não elucidam a análise.
- Não distorça os dados. Construa gráficos para revelar a verdadeira natureza dos dados.
- É preciso ter criatividade para utilizar o gráfico como ferramenta de descrição, exploração e comparação de dados.

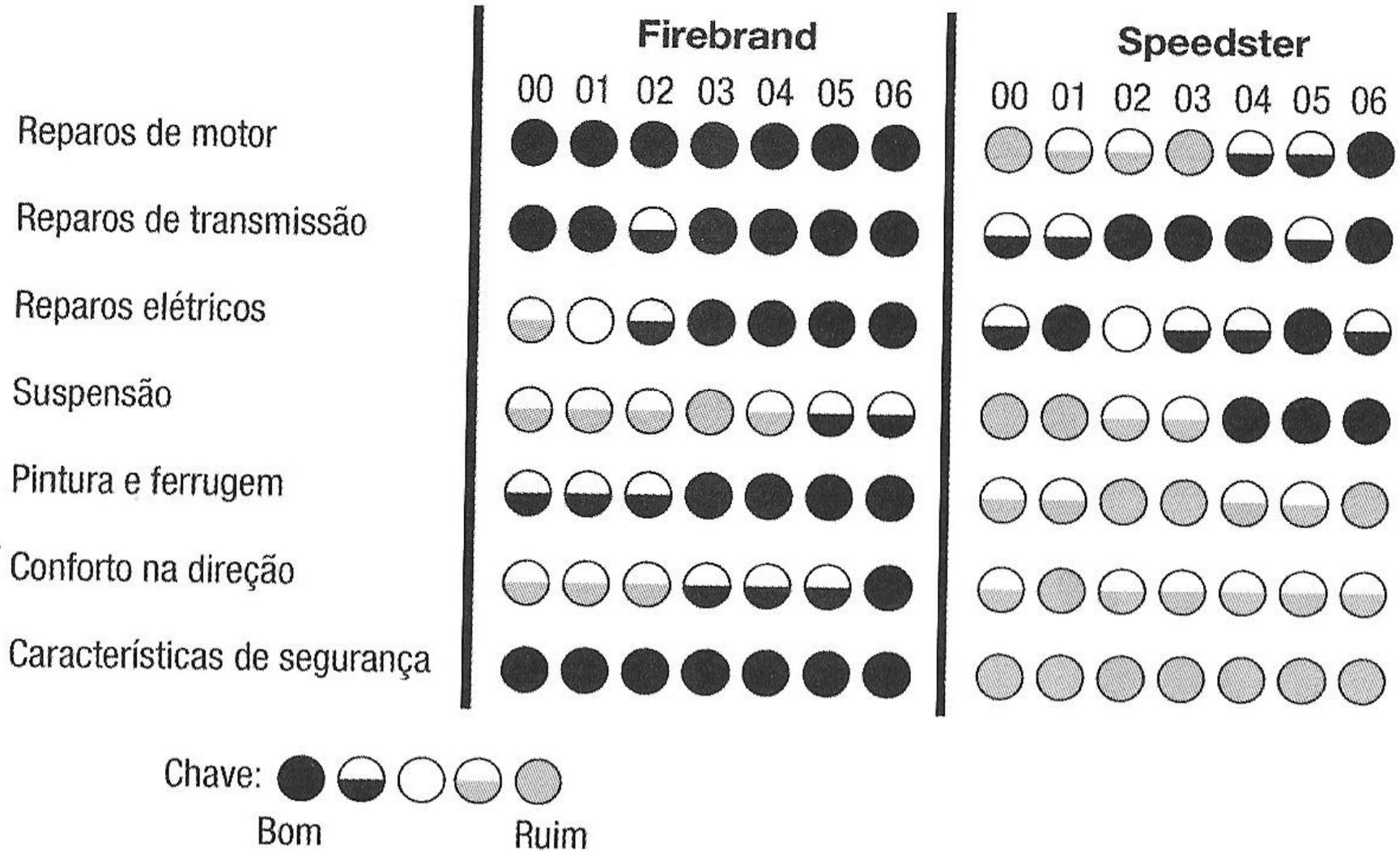
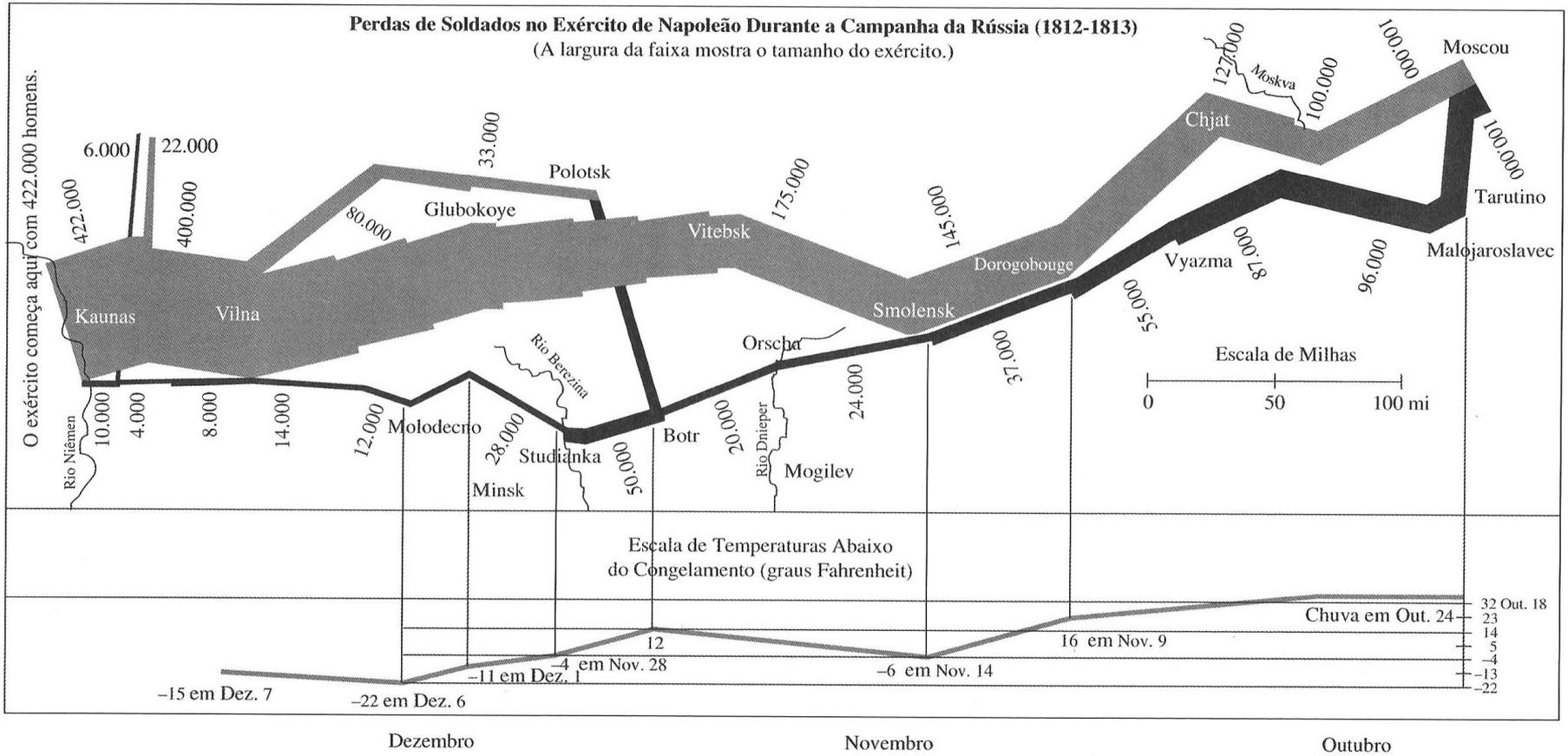


FIGURA 2-9 Dados sobre Confiabilidade de Carros por ano



Crédito: Redesenhado conforme Edward R. Tufte, *The Visual Display of Quantitative Information*, (Cheshire, CT: Graphics Press, 1983). Reproduzido com permissão.

FIGURA 2-10 Mortes nos Hospitais Militares Britânicos durante a Guerra da Criméia

Parte externa: Mortes devidas a doenças evitáveis.
Região do meio: Mortes devidas a outras causas que não doenças evitáveis ou ferimentos de guerra.
Região interna: Mortes devidas a ferimentos de guerra.

