

Apresentação gráfica

Tipos de variáveis, de gráficos e dicas práticas

André Gustavo dos Santos

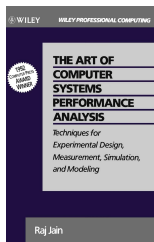
Departamento de Informática
Universidade Federal de Viçosa

INF222 - 2022/2

– Fonte do material

O conteúdo da seção 2 foi preparado e gentilmente cedido para uso nesta disciplina pela prof^a Elizabeth Wanner, do PPGMMC, CEFET-MG.

O conteúdo e as figuras das demais são do livro texto da disciplina (10.1–10.4):



Jain, Raj.
The art of computer systems performance analysis:
techniques for experimental design, measurement,
simulation, and modeling.
John Wiley & Sons, 1990

Tópicos da aula

1 Tipos de variáveis

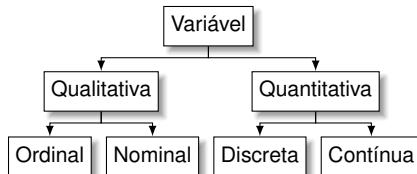
2 Tipos de gráficos

3 Sugestões

4 Erros comuns

5 Checklist

Tipos de variáveis



- Qualitativa: não expressas numericamente (qualidade, categoria)
 - Ordinal: podem ser ordenadas (melhor que, vem antes de, ...)
 - Nominal: não têm relação de ordem
- Quantitativa: expressas numericamente (quantidade, contagem, mensuração)
 - Discreta: valores enumeráveis
 - Contínua: valores em um intervalo de números reais

Tipos de variáveis – exemplos

- Qualitativa ordinal
 - Tipo de computador: supercomputador, minicomputador, microcomputador
- Qualitativa nominal
 - Tipos de arquivos: imagens, vídeos, documentos, aplicativos
 - Carga horária de trabalho: ensino, pesquisa, extensão
- Quantitativa discreta
 - Número de processadores de um computador
 - Tamanho dos blocos de disco
 - Quantidade de periféricos
- Quantitativa contínua
 - Tempo de resposta de uma tarefa em um sistema
 - Peso de um computador portátil

Representação Gráfica

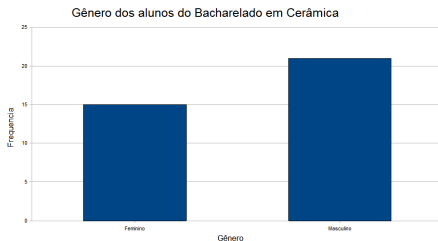
- Além de utilizar tabelas para resumir um conjunto de dados, gráficos fornecem um impacto visual alternativo.
- Tipos mais comuns:
 - Barras ou colunas: para dados discretos.
 - Setores: para se enfatizar proporções.
 - Linhas: para dados contínuos.

Gráfico de barras e colunas

- Tipo mais utilizado para representação gráfica.
- Cada categoria é representada por uma barra retangular distinta, sendo a frequência indicada pelo comprimento ou altura da barra.
- Em geral se chama gráfico de barras ou colunas, se são horizontais ou verticais.
- Pode ser utilizado por dados categóricos e numéricos discretos.
- Pode ser simples (um grupo) ou composto (mais de um grupo de dados), sendo múltiplo ou empilhado.

Gráfico de barras e colunas

Colunas simples



Barras simples com frequências

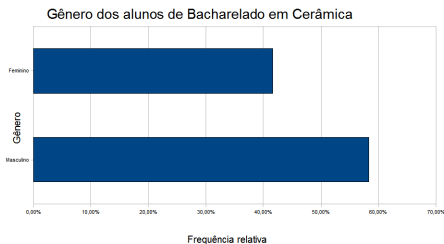
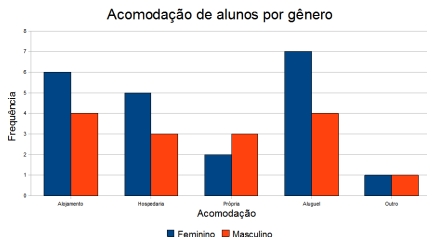


Gráfico de barras e colunas

Colunas múltiplas



Colunas múltiplas empilhadas

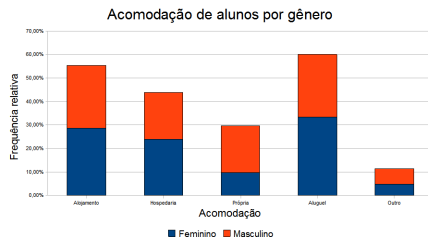


Gráfico de Setores

- Também conhecido por gráfico de pizza.
- Todo o conjunto de dados é representado por um círculo e cada categoria é representada por uma parte desse círculo, ou seja, por um setor.
- A frequência é representada pelo ângulo e 360 graus representa o total de dados.
- Utilizado para mostrar a que proporção do todo é tomada por cada categoria, "cortando a fatia" do gráfico principal.

Número de qualificações pós-médio

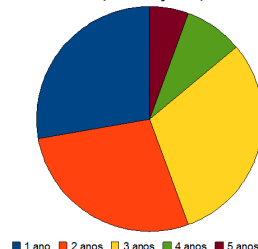
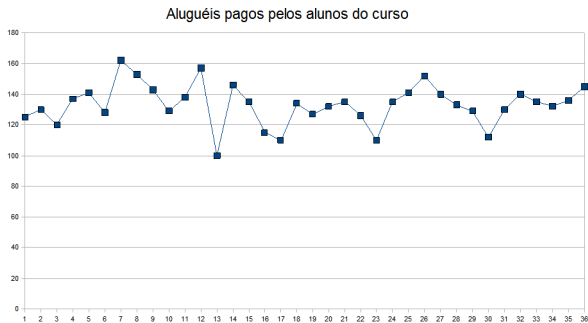


Gráfico de linhas

- Utilizado para representar dados numéricos contínuos e séries temporais: um conjunto de dados (vendas, lucros, custos) discriminados em função de intervalos de tempo variáveis.
- Um gráfico é de uma série temporal quando os dados foram coletados ao longo do tempo, representado pelo eixo horizontal x .
- O valor de cada variável é marcada com um ponto especial, e estes são ligados, ao longo do tempo, por uma linha.
- Se dois ou mais dados estiverem sendo exibidos num mesmo gráfico, utiliza-se cores diferentes.

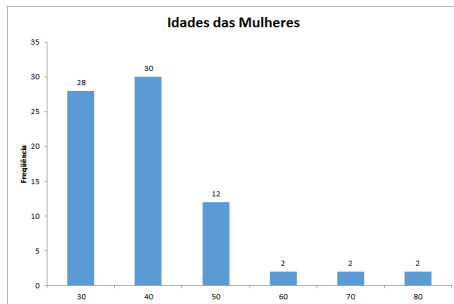
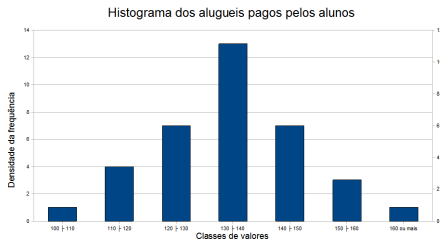
Gráfico de linhas



Histograma

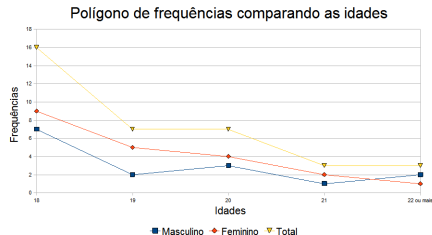
- É um tipo de gráfico de barras, que fornece uma representação comparativa dos dados.
- Os dados são agrupados em classes e mostra-se a frequência de cada uma.
- A base do gráfico é proporcional aos intervalos das classes e área de cada retângulo é proporcional à respectiva frequência.
- Procura informar como os dados estão distribuídos em relação à frequência.

Histograma



Polígono de Frequência

- Usa segmentos de retas ligando os pontos médios das classes.
- Representa a mesma informação de um histograma, mas utilizando gráficos de linhas.
- Ideal para se comparar a distribuição (histograma) de dois ou mais grupos de dados.



Ogiva

- Também chamado de gráfico de frequência acumulada.
- É um gráfico de linhas que fornece uma representação proporcional dos dados.
- É preciso primeiramente calcular a frequência acumulada de cada classe, que é o número total de observações que são menores que uma determinada observação.
- O gráfico é da forma frequência acumulada em função dos valores das classes.

Ogiva

Ogiva mostrando os aluguéis dos alunos

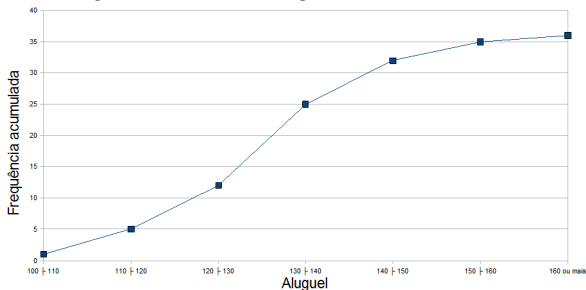
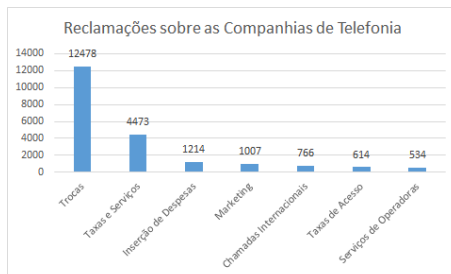


Gráfico de Pareto

- É um gráfico de barras para dados qualitativos, com as barras dispostas em ordem pela frequência.
- A barra mais alta fica sempre à esquerda e as barras menores se afastam para a direita.
- O gráfico chama a atenção para as categorias mais importantes.
- As escalas verticais podem representar as frequências absolutas ou frequências relativas.



Boxplot

- É um gráfico de um conjunto de dados que consiste em uma linha que se estende do valor mínimo ao valor máximo, em uma caixa com linhas traçadas no primeiro quartil, na mediana e no terceiro quartil.
- Revela o centro, a dispersão e a distribuição dos dados, além da presença de outliers.
- Necessita do cálculo do valor mínimo, valor máximo e os quartis.

Quartis

Assim como a mediana divide os dados em duas partes iguais, os três quartis, representados por Q_1 , Q_2 e Q_3 , dividem os valores ordenados em quatro partes iguais.

- Q_1 : No mínimo 25 % dos valores ordenados são menores do que ou iguais a Q_1 e 75 % dos valores são maiores do que ou iguais a Q_1 .
- Q_2 : O mesmo que a mediana: separa os 50 % inferiores dos valores ordenados dos 50 % superiores.
- Q_3 : No mínimo 75 % dos valores ordenados são menores do que ou iguais a Q_3 e 25 % dos valores são maiores do que ou iguais a Q_3 .

Percentis

- Os percentis são medidas que dividem a amostra (por ordem crescente dos dados) em 100 partes, cada uma com uma percentagem de dados aproximadamente igual.
- O k -ésimo percentil P_k é o valor $x(xk)$ que corresponde à frequência cumulativa de $N \cdot k/100$, em que N é o tamanho da amostra.
- Assim como há três quartis que separam um conjunto de dados em quatro partes, há 99 percentis, representados por P_1, P_2, \dots, P_{99} , que dividem os valores ordenados em 100 grupos com cerca de 1% dos valores em cada um.
- Portanto: o 1º percentil determina o 1% menor dos dados, o 25º percentil é o primeiro quartil e o 50º percentil é a mediana.

Idade das Melhores Atrizes e dos Melhores Atores

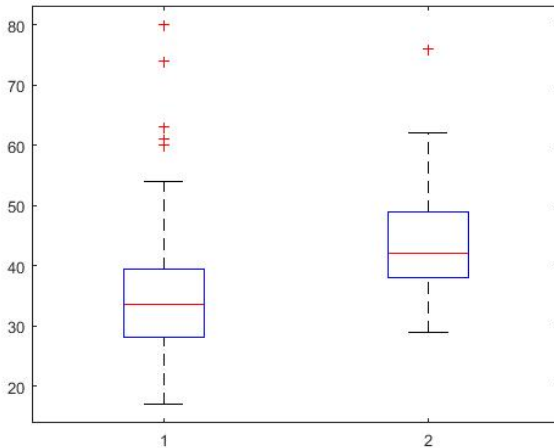
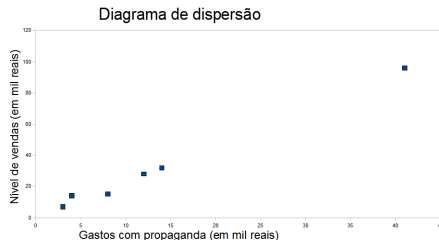


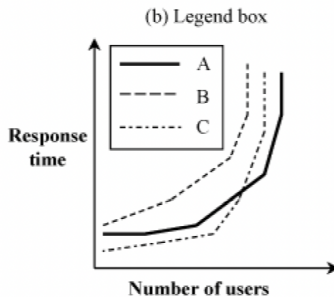
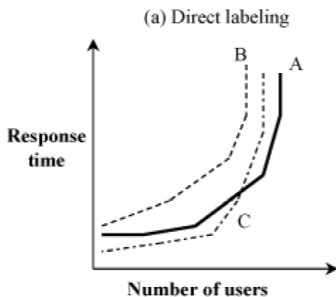
Diagrama de dispersão

- Os dados são colocados em pares ordenados que combinam cada valor de um conjunto de dados com um valor correspondente de um segundo conjunto de dados.
- Os pontos dos pares ordenados (x, y) são marcados no gráfico, sendo que a primeira variável é representada no eixo horizontal e a segunda variável é representada no eixo vertical.



Sugestões para bons gráficos

- Exigir esforço mínimo do leitor

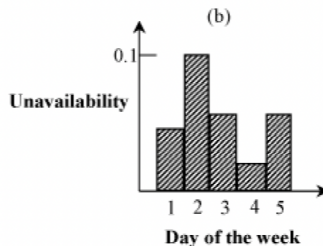
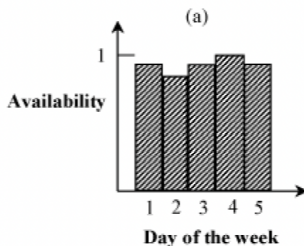


- Maximizar informação
 - Usar palavras em vez de símbolos
 - Rotular claramente os eixos

Sugestões para bons gráficos

■ Minimizar “tinta”

- Informação desnecessária torna o gráfico menos interessante (linhas de grade?)
- Gráficos diferentes podem mostrar mais informação dos mesmos dados



■ Usar padrões comumente aceitos

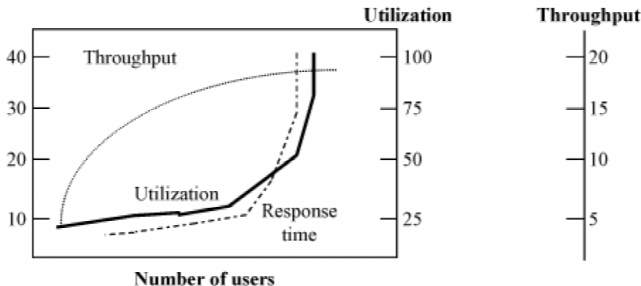
- Origem no (0, 0)
- Variáveis independentes (causa) no eixo x , variáveis dependentes (efeito) no eixo y
- Escala linear, crescente, igualmente espaçada, ...

■ Evitar ambiguidade

- Mostrar escalas, origem, identificar curvas, barras, ...

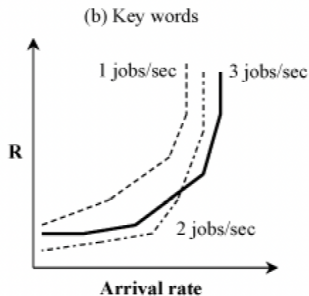
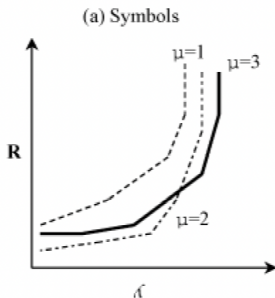
Erros comuns

- Apresentar muitas alternativas no mesmo gráfico
 - Regra dos 5: é fácil contar até 5, mais que isso começa a confundir
 - Não mais que 6 linhas, 8 setores, 10 barras
- Apresentar muitas variáveis y no mesmo gráfico



Erros comuns

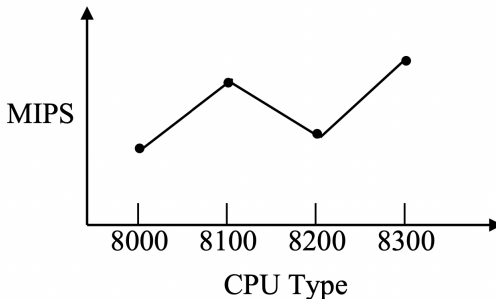
- Usar símbolos no lugar de texto



- Colocar informação extra
- Selecionar intervalos de escala inadequados

Erros comuns

- Usar gráfico de linha para variáveis qualitativas
 - Ligar pontos consecutivos em gráfico de linha indica que os valores intermediários podem ser aproximados por interpolação
 - Ferramentas que geram gráficos podem interpretar os valores numéricos como variáveis quantitativas e sugerir gráfico de linha, mas nesse caso são tipo de CPU

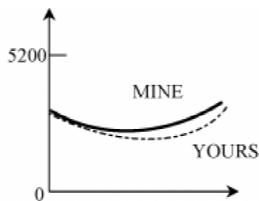
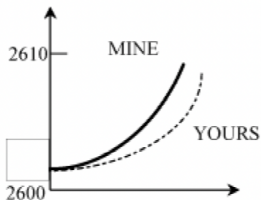


“Enganos” propositais

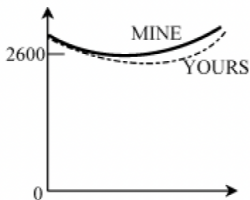
- Já que uma imagem vale mais que mil palavras, pode-se enganar com uma imagem o que necessitaria mil palavras
- Algumas técnicas têm sido usadas para deliberadamente esconder a verdade e iludir os leitores
- Não são erros por descuido ou inexperiência (como os anteriores), mas usados de propósito para exagerar os resultados

“Enganos” propositais

- Mudar origem para enfatizar diferenças

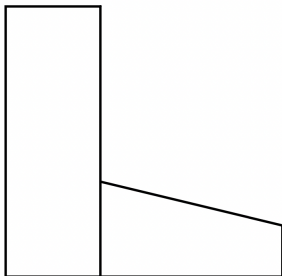


- Regra dos 3/4: altura/largura $\geq 3/4$

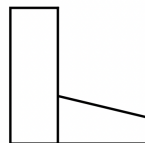


“Enganos” propositais

- Similar a alterar escala de imagens para enfatizar diferenças
- Dobrar a altura de um pictograma acaba dando impressão de 4 vezes maior



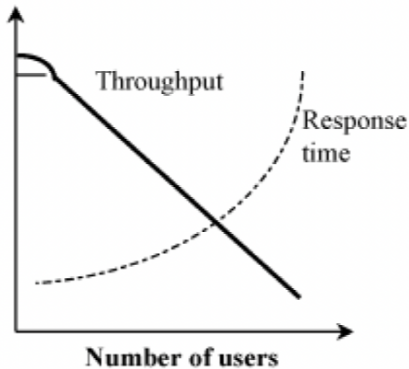
Mine
Performance = 2



Yours
Performance = 1

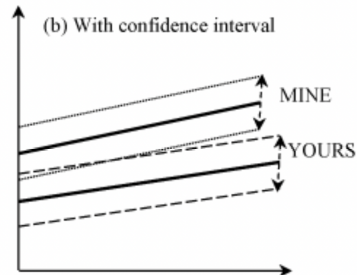
“Enganos” propositais

- Usar “golpe duplo” para dramatizar
 - Duas curvas no mesmo gráfico causam duplo impacto
 - Isto pode ser usado para amplificar resultados bons ou ruins
 - Mas se mostram dados relacionados (um prediz o outro) está apenas exagerando informação



“Enganos” propositais

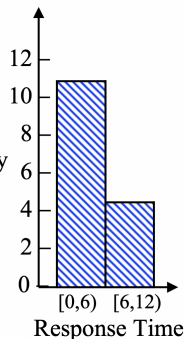
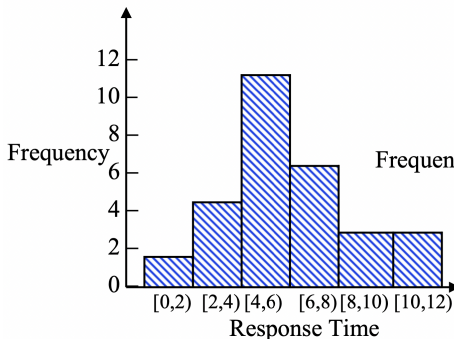
- Mostrar resultados de amostras aleatórias sem intervalo de confiança
 - Algumas medidas podem resultar valores diferentes em cada observação
 - Nesse caso devem ser feitas muitas repetições do experimento para achar a média
 - A média sozinha, entretanto, não é suficiente para comparação de duas quantidades
 - É necessário mostrar também o intervalo de confiança



“Enganos” propositais

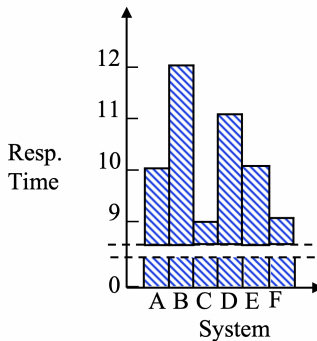
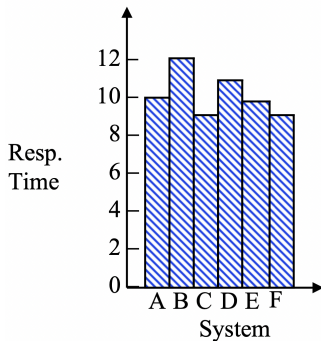
■ Usar intervalos inadequados em histogramas

- É geralmente difícil escolher um tamanho apropriado para os intervalos
- Histogramas para dados reais requerem mais de uma tentativa
- Para intervalos muito grande, os dados se concentram em algumas poucas células
- Para intervalos muito pequenas, o histograma perde suavidade
- Selecionando apropriadamente o tamanho da célula, pode-se fazer com que os dados se pareçam com a distribuição desejada...



“Enganos” propositais

- Quebrar escalas em gráficos de barras
 - Isso pode amplificar exageradamente diferenças que na verdade são insignificantes
 - Um efeito similar à mudança de origem



Checklist para bons gráficos

- 1 Os dois eixos de coordenadas são mostrados e rotulados?
- 2 Os rótulos dos eixos são autoexplicativos e concisos?
- 3 As escalas e divisões são mostradas em ambos os eixos?
- 4 O mínimo e o máximo dos intervalos mostrados nos eixos são apropriados para apresentar o máximo de informações?
- 5 O número de curvas é razoavelmente pequeno?
(um gráfico de linhas geralmente não deve ter mais de seis curvas)
- 6 Todos os gráficos usam a mesma escala?
(várias escalas no mesmo gráfico causam confusão)
- 7 Não há curva que possa ser removida sem reduzir a informação?
- 8 As curvas em um gráfico de linhas são rotuladas individualmente?
- 9 As células em um gráfico de barras são rotuladas individualmente?

Checklist para bons gráficos

- 10 Todos os símbolos são acompanhados de explicações textuais apropriadas?
- 11 Se curvas se cruzam, os padrões das linhas são diferentes para evitar confusão?
- 12 As unidades de medida estão indicadas?
- 13 A escala horizontal está aumentando da esquerda para a direita?
- 14 A escala vertical está aumentando de baixo para cima?
- 15 As linhas de grade estão auxiliando na leitura das curvas?
(se não, as linhas de grade não devem ser mostradas)
- 16 Todo este gráfico contribui para a informação disponível ao leitor?
- 17 As escalas são contíguas?
(quebras na escala devem ser evitadas ou claramente mostradas)
- 18 A ordem das barras em um gráfico de barras é sistemática?
(ordem alfabética, temporal, do melhor para o pior deve ser preferida à aleatória)

Checklist para bons gráficos

- 19 Se o eixo vertical representa uma quant. aleatória, os intervalos de confiança são mostrados?
- 20 Não há curvas, símbolos ou textos que possam ser removidos sem afetar a informação?
- 21 Existe um título para todo o gráfico?
- 22 O título do gráfico é autoexplicativo e conciso?
- 23 Para gráficos de barras com intervalo de classe desigual, a área e a largura são representativas da frequência e do intervalo, respectivamente?
- 24 As variáveis plotadas neste gráfico fornecem mais informações do que outras alternativas?
- 25 O gráfico traz claramente a mensagem pretendida?
- 26 A figura é referenciada e discutida no texto do relatório?