

Os gráficos no apoio à tomada de decisões e à disseminação de informações estatísticas



Por: Áurea Sousa
Professora Auxiliar do Departamento de Matemática e Estatística
da Faculdade de Ciências e Tecnologia
da Universidade dos Açores
aurea.st.sousa@uac.pt

As tecnologias da comunicação e da informação possibilitam a recolha e o armazenamento (automático ou não) de dados, por vezes em tempo real, assim como a produção, cada vez mais frequente, de conteúdos digitais. Atualmente, muitas instituições, públicas e privadas, possuem bases de dados, algumas das quais armazenam grandes volumes de dados. Porém, para que possam ser extraídos conhecimentos úteis a partir destes dados, é necessário que estes sejam analisados e resumidos mediante a aplicação de métodos estatísticos apropriados. Neste contexto, têm surgido novos métodos de análise de dados, incluindo algumas representações gráficas, por vezes interactivas.

As imagens têm sido usadas, desde a antiguidade, como uma forma de comunicação, mesmo antes da consolidação da linguagem escrita, e geralmente tomam a comunicação de informações muito mais apelativa e explicativa, além de poderem ser compreendidas, qualquer que seja o idioma (linguístico primeiras formas de visualização de informações derivaram, fundamentalmente, da necessidade de utilização de mapas no decurso de viagens, assim como da necessidade de representação de dados relativos, por exemplo, ao comércio e à religião, e também para efeitos de comunicação).

A deteção de padrões através de tabelas, sem recurso a técnicas de visualização gráfica, é frequentemente impraticável, sobretudo no caso de bases de dados de elevado dimensão e no caso de dados em larga escala que abrangem um amplo período temporal. Atualmente, estas técnicas têm um papel de relevo a nível do tratamento e organização dos dados, facultando uma rápida perceção visual da informação estatística que comportam, fazendo sobressair as informações mais relevantes e realçando tendências e exceções. Assim, algumas informações são mais facilmente transmitidas e compreendidas visualmente do que através de palavras. Esta é a razão pela qual se diz, em português, que “uma imagem vale mais do que mil palavras” e, em inglês, que “a picture is worth a thousand words”, frase da autoria do filósofo chinês Confúcio (o seu verdadeiro nome era Chiu Kung), que enaltece o poder da comunicação através das imagens.

No séc. XVII foram registados grandes avanços a nível da visualização de informações estatísticas, impelidos por fatores, tais como o desenvolvimento da geometria analítica e do sistema de coordenadas e o aparecimento da teoria da probabilidade e da estatística demográfica, entre outros fatores.

William Playfair (1759–1823) produziu os primeiros gráficos de linhas, áreas e de barras, em 1786, e de setores, em 1801, os quais são muito utilizados atualmente. Para Playfair, a utilização de gráficos era preferível à utilização de tabelas, por mostrarem a forma dos dados numa perspetiva comparativa.

A importância da Estatística, em geral, e da representação gráfica de dados, em particular, tem sido reconhecida desde longa data. É de referir, por exemplo, que já em 1870, quase todos os países eu-

ropeus e os Estados Unidos recolhiam, analisavam e disseminavam sistematicamente estatísticas oficiais do governo relativas à população, ao comércio e a problemas políticos e sociais. Em 1872, é de realçar a utilização dos gráficos estatísticos, pelo governo dos Estados Unidos, nos relatórios dos Censos (Cartogramas de dados do nono Censo). É ainda de referir que o grande progresso a nível dos métodos gráficos, em meados do séc. XIX, derivou do reconhecimento da importância da informação quantitativa a nível do planeamento do bem-estar geral da população. Nesse período, surgiram novas disciplinas (e.g., Estatística, Geologia, Biologia, Economia), tendo esta época sido referida como a idade do entusiasmo, mais conhecida por “Época Áurea”.

Desde Playfair, a utilização e a divulgação dos gráficos estatísticos têm sido cada vez mais frequentes, embora muitos deles datem desse tempo (século XVIII/XIX).

Os sucessivos avanços computacionais ocorridos a partir da década de setenta do século XX têm revolucionando as técnicas de representação gráfica e incrementado, substancialmente, a sua utilização, em todas as áreas científicas, facilitada pelo acesso a software interativo e “userfriendly”. Os tipos de gráficos mais frequentemente utilizados na atualidade são os diagramas, que consideram, no máximo, duas dimensões (e.g., gráficos de linhas, gráficos de colunas ou de barras, gráficos de colunas ou de barras múltiplos, gráficos circulares ou de setores, histogramas, polígonos de frequências acumuladas (Ogiva de Galton), diagramas de dispersão, diagramas de extremos e quartis e os gráficos de caule-e-folhas); os gráficos polares; os cartogramas; e os pictogramas.

No caso da visualização de dados multivariados, a diversidade e a multiplicidade das variáveis, constituem um desafio adicional, pelo que nesse âmbito merecem uma especial ênfase os mapas



perceptuais, resultantes, por exemplo, da Análise em Componentes Principais e da Análise de Correspondências, simples e múltipla; os dendrogramas fornecidos por métodos de Análise Classificatória Hierárquica no âmbito da Análise de Clusters; e as faces de Chemoff (1973). Neste último tipo de gráfico, cada ponto de um espaço k -dimensional é associado a um esboço de face humana, sendo cada dimensão representada por uma característica da face.

Com a utilização cada vez mais frequente do Excel e de packages estatísticos, a utilização de gráficos está ao alcance de todos. Todavia, a sua utilização requer alguns cuidados para que as informações neles exibidas não sejam distorcidas. Um gráfico, por mais simples que seja, deverá ser apelativo, estar em conformidade com os princípios da integridade dos gráficos, sobretudo o de que não devem distorcer a realidade (isto, devem manter a verdade sobre os factos), e conciliar a funcionalidade e a estética.

A importância da aquisição e do desenvolvimento de competências a nível da Estatística, em geral, e da representação gráfica, em particular, é uma mais-valia, para todos os que necessita de produzir e/ou compreender resultados obtidos com base na análise estatística de dados. A aposta na apresentação dos resultados de forma visual pode tornar a tomada de decisões mais rápida, apelativa,

eficiente e eficaz.

Embora não seja possível descrever todos os tipos de gráficos, neste artigo, vou descrever de modo sucinto a forma como se processa a construção do diagrama de extremos e quartis, o qual foi desenvolvido e popularizado pelo matemático John Tukey (1915–2000).

O diagrama de extremos e quartis é um tipo de gráfico, em que o conjunto dos valores compreendidos entre o 1º quartil (Q1) e o 3º quartil (Q3) é representado por um retângulo, dentro do qual a posição da mediana é assinalada por um segmento de reta (barra). Este tipo de gráfico comporta, ainda, duas linhas que unem os meios dos lados do retângulo, determinados pelo 1º e 3º quartis, aos extremos da amostra (mínimo e máximo), e pode ser construído na vertical ou na horizontal. Esta representação gráfica permite obter informações relativas à forma como os dados se distribuem, mais especificamente em relação à sua maior ou menor dispersão; simetria; existência de outliers (valores aberrantes, isto é, muito grandes ou muito pequenos face aos restantes valores da série de dados), sendo muito útil a nível da comparação de vários conjuntos de dados (tanto menor for a altura da caixa mais concentrados, isto é, menos dispersos) estão os dados. Quando há outliers geralmente, em vez de se considerar o mínimo e o máximo da série estatística, utilizam-se o menor e o maior valor da amostra que não são outliers, sendo os outliers geralmente assinalados pelos símbolos * (no caso de serem severos) ou “o” (no caso de serem moderados).

Na realidade, os gráficos são cada vez mais utilizados pelos estudantes, gestores, investigadores, meios de comunicação social e pelas instituições governamentais, para a apresentação de informações estatísticas e apoio à tomada de decisões informadas. Nesse contexto, nunca é demais recordar que os resultados referentes à análise estatística de dados, incluindo os apresentados de forma gráfica, poderão ser decisivos a nível da definição de estratégias a serem implementadas no âmbito das políticas públicas, e até mesmo a nível de negociações. Porém, é essencial que os cidadãos apostem no desenvolvimento da literacia estatística e do raciocínio crítico, para que não se deixem enganar, caso se confrontem com gráficos mal construídos, de forma intencional ou não. Porém, quando bem construídos, os gráficos podem ser muito úteis a nível da tomada de decisões e da disseminação de informações estatísticas.

