



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

Novo paradigma e importância da representação gráfica de dados

Atualmente, devido aos sucessivos e rápidos avanços computacionais, são gerados e armazenados computacionalmente grandes volumes de dados, cuja análise é cada vez mais difícil e complexa.

Os dados são muitas vezes gravados automaticamente através de sensores e sistemas de monitorização, havendo um armazenamento exaustivo dessas informações. A internet tem também revolucionado esta área, possibilitando a recolha de dados em tempo real (e.g., temperatura, chegada de transportes públicos, padrões de tráfego num website), pelo que a visualização de informação passa, assim, a lidar com dados do mundo real, que estão em constante atualização. Até mesmo atividades tão simples como o pagamento através de cartão de crédito ou a utilização do telefone são registadas pelos computadores. Este fenómeno é conhecido como “internet das coisas”. É de realçar, ainda, que o registo de um número considerável de parâmetros dá origem a dados com elevada dimensionalidade e que a visualização de dados multivariados é um campo de investigação, cada vez mais ativo, com aplicações nas mais variadas áreas do conhecimento.

Na era do Big Data (termo utilizado para descrever grandes volumes de dados, complexos, cujo processamento e análise dificilmente podem ser efetuados utilizando programas e processos usuais), a visualização gráfica de dados é de crucial importância. Face à produção diária de conteúdos digitais e à disponibilização massiva de informação, o tratamento gráfico destes dados torna-se imprescindível, de forma a possibilitar a extração de novos conhecimentos úteis, o que impulsiona a pesquisa contínua na área da visualização de informação.

A utilização de técnicas gráficas eficazes na transmissão de informação (quantitativa ou qualitativa) enriquece a análise de dados, destacando tendências e exceções, eliminando elementos irrelevantes dos dados e fazendo sobressair informações relevantes, o que tem uma especial utilidade no caso da representação de dados em larga escala que se estendem por um longo período de tempo. Na verdade, a deteção de padrões através de tabelas, sem recurso a técnicas de visualização gráfica, é muitas vezes impraticável (caso de tabelas de dados com um elevado número de linhas e/ou de colunas).

As imagens têm sido usadas como um mecanismo de comunicação, mesmo antes da formalização da linguagem escrita, e podem ser geralmente interpretadas independentemente da língua. As técnicas de visualização desempenham um papel fulcral a nível do tratamento e organização dos dados, proporcionando uma rápida perceção visual da informação estatística a que se reportam, auxiliando os processos cognitivos, e têm a sua origem nos diagramas geométricos, nas tabelas de posição das estrelas e nos mapas. É de referir, ainda, que no século XVI, com a expansão marítima da Europa, foram desenvolvidos novos instrumentos e téc-



nicas, os quais estão na gênese de novas e mais precisas formas de apresentação visual do conhecimento.

A apresentação abstrata da informação através de gráficos foi antecedida por diversas descobertas, tais como o sistema de coordenadas introduzido pelos egípcios em 3200 a.C. e reintroduzido por Descartes no século XVII e as representações simbólicas em áreas tais como a Música e a Astronomia, que contribuíram para a implantação de um modelo abstrato de pensamento, materializado em representações visuais (e.g., notação musical, movimentação planetária). Contudo, o maior avanço conceptual no que diz respeito à representação gráfica deu-se com a publicação do Atlas de William Playfair (1786), na qual as dimensões espaciais foram, pela primeira vez, usadas para representar dados quantitativos não espaciais, iniciando-se uma nova era a nível da difusão de informação estatística, em que se torna fundamental a comunicação de informação de forma eficiente e eficaz. É de referir, ainda, que Playfair inventou a maioria dos gráficos mais conhecidos atualmente (gráfico de barras, gráfico de linhas e gráfico circular).

Entre 1835 e 1876 foram realizados oito congressos internacionais de Estatística, em que esteve patente a discussão e a promoção da utilização de métodos gráficos. É, ainda, de sublinhar que o grande desenvolvimento dos métodos gráficos em meados do séc. XIX resultou do reconhecimento, por parte da maioria dos países, da importância da informação numérica a nível do planeamento do bem-estar geral da população. Este período marcou também o nascimento de novas disciplinas, tais como a Estatística, a Geologia, a Biologia e a Economia, tendo esta época sido considerada como a idade do entusiasmo, mais conhecida por “Época Áurea”.

Os avanços computacionais (desenvolvimento das capacidades de processamento, armazenamento e interação com o utilizador) ocorridos a partir da década de setenta do século XX têm possibilitado a aplicação e o desenvolvimento de novas técnicas gráficas e de novos métodos de análise de dados, utilizando software interativo e de fácil utilização, revolucionando a área da visualização de informação. Na verdade, com o desenvolvimento da informática, a representação gráfica está ao alcance de todos, mas é fundamental que seja evitada a utilização de gráficos enganosos ou pouco apelativos, já que as boas representações gráficas encorajam questões, mas as más escondem mais do que mostram. Um gráfico bem construído, mesmo que seja extremamente simples, deverá conseguir transmitir uma ideia impactante e fazer justiça ao ditado popular de que “uma imagem vale mais que 1000 palavras”, estabelecendo um equilíbrio entre a estética e a funcionalidade.

Atualmente, são utilizados diversos tipos de gráficos, sendo os mais frequentes os seguintes: Diagramas: Gráficos geométricos dispostos em, no máximo, duas dimensões (e.g., gráfico de linhas; gráfico de colunas ou de barras; gráfico de colunas ou de barras múltiplas; gráfico circular ou gráfico de sectores ou gráfico de pizza (um círculo é dividido em sectores, de acordo com o número de parcelas, sendo os 360° repartidos proporcionalmente (regra de 3 simples), onde a soma de todas as parcelas corresponde a 360°); histograma; polígono de frequências acumuladas (Ogiva de Galton); diagrama de dispersão; diagrama de extremos e quartis ou “Caixa-de-bogides”; gráfico de caule-e-folhas.

Gráfico polar: Representação de uma série por meio de um polígono (uma circunferência é dividida em tantos arcos iguais quantos forem

os dados a representar. Posteriormente, em cada raio, que liga o centro da circunferência a um dos vértices do polígono, é representado um valor da série, marcando-se um ponto cuja distância ao centro é proporcional a esse valor, unindo-se seguidamente os pontos assim determinados).

Cartogramas: Ilustrações relativas a cartas geográficas, utilizadas frequentemente em algumas áreas científicas, tais como a Geografia, a História e a Demografia;

Pictogramas: Processo gráfico em que são utilizadas figuras/imagens sugestivas.

No caso de dados multivariados merecem uma especial referência os mapas percutuais, resultantes, por exemplo, da Análise em Componentes Principais e da Análise de Correspondências, simples e múltipla; os dendrogramas obtidos através de algoritmos de Análise Classificatória Hierárquica; e as faces de Chernoff (cada ponto é associado a um esboço de face humana).

Em 1995, Tufte publica um livro intitulado “The Visual Display of Quantitative Information” que contém diversos princípios que devem nortear os designers de visualizações de informação. Alguns princípios da excelência dos gráficos, segundo Tufte, são os seguintes: bom design; clareza, precisão e eficácia; fornecer o maior número de ideias no menor tempo possível, com a menor quantidade de tinta e no menor espaço possível; utilização de diversas variáveis; e registar a verdade sobre os dados.

Os principais princípios da integridade dos gráficos são os seguintes: os números devem ser diretamente proporcionais às quantidades numéricas a representar; devem conter legendas; devem mostrar a variação na informação, não no design; a apresentação de dados monetários em gráficos em função do tempo (time-series) deve ser feita através de unidades monetárias padrão; O número de dimensões da informação representada não deve exceder o número de dimensões dos dados; e não devem fazer alusão a dados fora do contexto.

Dado o exposto, neste artigo, facilmente se compreende que o desenvolvimento de competências a nível da representação gráfica é essencial, a par de outras competências a nível da Estatística, não só para os funcionários de algumas empresas, públicas ou privadas, que necessitam de tomar decisões apoiadas nos dados, mas também para que os cidadãos em geral possam compreender melhor as informações difundidas pelos meios de comunicação social. Em jeito de síntese, uma das formas mais eficientes de explorar e tentar compreender um grande conjunto de dados é através da visualização de informação. A aposta na apresentação dos resultados de forma visual pode tornar a tomada de decisão mais rápida e eficaz. No entanto, para que isso aconteça, os gráficos devem ser apelativos e acima de tudo respeitarem os princípios da integridade dos gráficos, sobretudo o de não distorcer a realidade.