

Capítulo 1

Introdução

De acordo com a IBM, criamos 2,5 quintilhões de *bytes* por dia e em 2016 90% dos dados no mundo foram gerados somente nos dois anos anteriores¹. Numa era em que os dados são produzidos a uma velocidade assustadora, estamos diante de uma nova situação desafiadora que consiste em extrair informações efetivas a partir de um grande volume de dados, muitas vezes incongruentes por se originarem de fontes distintas. Além de prover mecanismos apropriados para armazená-los, é necessário desenvolver mecanismos e ferramentas apropriados para acessá-los/minerá-los e exibí-los num formato inteligível de forma que possam efetivamente apoiar nas tomadas de decisão humanas.

A nossa visão é considerada a mais importante para reconhecer de forma efetiva o mundo físico em nossa volta, pois, dentre os cinco sentidos fundamentais (tato, paladar, olfato, visão e audição) ela é a mais apurada e desenvolvida na média da população. Embora os recursos visuais já sejam usados pelos homens de caverna para registrar os seus feitos através das pinturas rupestres e que na primeira metade do século XIX foram criados diferentes gráficos estatísticos (gráfico de barras, gráfico de pizza, gráfico de linhas, histograma, diagrama de extremos e quartis), o estatístico americano John Wilder Tukey foi o primeiro a mostrar formalmente em 1977 através do seu livro *Exploratory Data Analysis* como interagir visualmente com os dados brutos sintetizados em gráficos, diagramas e cartas para obter informações desejadas.

Na década de 1980, Ben Schneiderman introduziu a manipulação direta, no lugar da tradicional linguagem de programação, como uma forma inovadora de se interagir com um computador. A ideia consistia em tornar

¹<https://blog.storagecraft.com/inside-data-growth-explosion/>

mais simples as interações que ocorrem entre os objetos armazenados no computador e os seus usuários através de interfaces gráficas familiares e responsivas. Em paralelo, com o desenvolvimento das técnicas de renderização, emergiram na década dos 90s duas áreas de pesquisa que tem como objetivo exibir de forma inteligível os dados científicos tri-dimensionais (visualização científica) e os dados de natureza mais genérica ou abstrata (visualização de informação). A compreensão dos dados propiciada pelas técnicas de visualização incitaram a demanda por mecanismos que facilitem a exploração dos dados visualizados. Nasceram os primeiros sistemas de visualização exploratória de dados.

1.1 Analítica Visual

O termo **Analítica Visual**, ou em inglês *visual analytics*, foi usado em 2004 pelos editores Pak Chung Wong e Jim Thomas de uma edição especial da revista *IEEE Computer Graphics and Applications* para caracterizar uma coletânea de artigos selecionados que “*combine the art of human intuition and the science of mathematical deduction to directly perceive patterns and derive knowledge and insight from them*”². A característica comum destes trabalhos é a integração das técnicas de processamento e inferência de informações a partir de um volume grande de dados brutos e das técnicas de renderização e de interação para entender e analisar os dados brutos volumosos com a finalidade de proporcionar um melhor suporte às tomadas de decisão em problemas complexos³.

A tecnologia de Analítica Visual é, portanto, multidisciplinar. Ela integra ciência e tecnologia de muitas áreas, algumas bem consolidadas. Técnicas de gerenciamento e processamento de dados são aplicadas para manipular um conjunto volumoso de dados brutos. Técnicas de mineração de dados são usadas para extrair informações de interesse a partir destes dados brutos. Técnicas desenvolvidas na área de Visualização são úteis para exibir os dados numa forma inteligível, apropriada para a percepção e a cognição dos seus potenciais usuários. E, finalmente, técnicas de interação ajudam no desenvolvimento de uma interface através da qual um usuário possa explorar os dados a fim de buscar subsídios para o seu raciocínio lógico.

²Visual Analytics, IEEE Computer Graphics and Applications, 24:5, p. 20–21, Setembro 2004

³D. Keim e colegas. *Mastering the information age: solving problems with visual analytics*, 2010.

1.2 Aplicações de Analítica Visual

A maioria dos processos de exploração/mineração de um grande volume de dados tem como finalidade a solução de um problema. Numa era digital em que as máquinas conseguem substituir várias tarefas humanas, é natural perguntarmos em que situações um sistema de analítica visual venha a ser de fato relevante. Para responder esta pergunta, precisamos primeiro distinguir duas classes de problemas: problemas bem condicionados e problemas mal-condicionados. Problemas bem condicionados são aqueles que tem seus modelos matemáticos bem definidos com um comportamento bem estável, enquanto os problemas mal-condicionados são aqueles cujas respostas aos modelos existentes são totalmente imprevisíveis. Usualmente, isso decorre da escolha inadequada do modelo para representar o problema. Algumas vezes, conhecendo melhor os dados do problema através de uma exploração/mineração destes dados, pode-se chegar a um melhor modelo para o problema. Outras vezes, pode-se concluir que o problema em consideração é ainda um problema em aberto. Em ambos os casos, um sistema de analítica visual que proporcione diferentes percepções (*insights*) dos dados nos ajudam a organizar as ideias, perceber as nuances, fazer analogia, formular e reformular hipóteses, e até reconsiderar a forma como se olha para o problema.

Hoje em dia, é consenso que muitas áreas podem se beneficiar das tecnologias de analítica visual em diferentes níveis de abstração. Dentre elas destacam-se Astrofísica, Meteorologia, Gerenciamento de Emergências, Medicina, Genética, Bioquímica e Finanças.

1.3 Reengenharia

Veremos ao longo deste curso que um sistema de analítica visual é, essencialmente, um eficiente gerenciador, processador e analisador de dados volumosos acoplado a uma efetiva interface responsiva às possíveis intenções do seu potencial usuário, para que este possa **construir** do “zero” o seu modelo mental acerca do problema explorando os dados disponíveis, **formular e re-formular** suas hipóteses, **avaliá-las** através de provas de conceito, e, se possível, **solucioná-lo**. Note que as atividades construir, formular/reformular, avaliar e solucionar são exclusivas humanas. O papel da máquina é proporcionar um ambiente apropriado para estas tarefas. Além de dispor de um bom gerenciador de dados, ela deve suportar diferentes representações/modelos dos dados de diferentes níveis de abstração, um visua-

lizador adequado para estes diferentes níveis de abstração e, ainda, recursos de interação para o usuário transitar livremente entre os dados e modelos.

(Fig. 2.3 do livro de D. Keim e colegas⁴)

Se analisarmos a evolução nas áreas de pesquisa como Banco de Dados, Análise de Dados relacionada com o Aprendizado de Máquina, Computação Gráfica, Interface Homem-Máquina, veremos que as ferramentas mencionadas no parágrafo anterior já são pesquisadas, e algumas bem consolidadas. Devemos, então, considerar que um sistema de analítica visual pertence a uma área de desenvolvimento e não de pesquisa? Acredito que não. Um sistema de analítica visual tem algumas características peculiares que veremos ao longo deste curso. E o grande desafio que temos diante de nós é como recondicionar os recursos existentes para um sistema de analítica visual. Ou seja, no meu entender, é uma pesquisa relacionada com Reengenharia.

1.4 Exercícios

1. Numa era de crescimento exponencial do volume de dados, quais são as tecnologias desejadas para tirarmos proveito destes dados da melhor forma possível?
2. Por quê as técnicas automáticas de processamento, análise, inferência e tomadas de decisão (soluções *turnkey* e “inteligentes”) não são apropriadas para diversas aplicações?
3. O que é uma tecnologia de Analítica Visual (*visual analytics*)? Como se pretende aplicar esta tecnologia na solução de um problema?
4. Considera-se que a tecnologia de Analítica Visual seja uma tecnologia multidisciplinar, envolvendo quatro elementos de área de pesquisa distinta: visualização, processamento de dados, modelagem de dados e aplicação de conhecimentos. Qual é o papel de cada um destes elementos na solução do problema e como eles são divididos entre máquinas e homens?
5. O que você entende pelo trecho abaixo extraído do livro de D. Keim e colegas:

“Infrastructure is concerned with linking together all the processes, functions and services required by visual analytic applications so they

⁴D. Keim e colegas. *Mastering the information age: solving problems with visual analytics*, 2010.

work in harmony, in order to allow the user to undertake their data exploration tasks in an efficient and effective manner. This is difficult as the software infrastructures created by the different technologies are generally incompatible at a low level and this is further complicated as one of the fundamental requirements of visual analytics applications is high interactivity. For this reason, most visual analytics applications are currently custom-built stand-alone applications, using for example, in-memory data storage rather than database management systems. The design of system and software architectures is paramount in enabling applications to successfully utilise the most appropriate technologies. In addition, the reuse of many common components will result in applications being more adaptable and built much quicker.”

6. Com base nos seus conhecimentos, quais seriam os principais ingredientes da infra-estrutura de um sistema de analítica visual?
7. Pelo trecho abaixo extraído do livro de D. Keim e colegas, quais são as principais características de um problema mal-definido? Por quê uma solução baseada na analítica visual seja mais apropriada para tal problema?

“The latter (ill-defined problem) is where virtually no information about the problem and possible solutions are available, so the early stages of problem solving (recognition, definition, representation of a problem) are a challenging task. If the problem is well-defined, the emphasis of the problem solver’s activity is on the later stages (development of a solution strategy, progress monitoring, evaluation of the solution). In addition, the solution path can often be described by an algorithm, which is not possible with ill-defined problems because they usually necessitate radical changes in problem representation.”

8. Leia o artigo *Browsers to Support Awareness and Social Interaction* de Alison Lee, Andreas Girgensohn e Jun Zhang,⁵ e identifique: (1) o problema em foco; (2) aspecto impactante de um sistema de analítica visual na solução; (3) os componentes do sistema de analítica visual proposto.

⁵DOI: 10.1109/MCG.2004.24

Capítulo 2

Percepção e Cognição

Os conhecimentos de um usuário e as suas tomadas decisão constituem uma das características fundamentais de um sistema de Analítica Visual. Como se toma como o ponto de partida os estímulos visuais através dos gráficos estatísticos e/ou representações gráficas tri-dimensionais, é fundamental que entendamos o nosso processo de percepção visual e a nossa potencialidade e limitação cognitiva para podermos nos beneficiar no projeto de um sistema de Analítica Visual.

2.1 Percepção Visual

A visão humana é um sistema complexo que começa com os dois olhos. As radiações luminosas atravessam o cristalino (lente biconvexa flexível) e convergem na membrana interna da parte posterior do olho, denominada retina. A retina contém cerca de centenas de milhões de fotorreceptores (bastonetes e cones). Quando estes fotorreceptores são estimulados, ocorrem transições fotoquímicas gerando impulsos elétricos. Estes impulsos são transmitidos pelo nervo óptico ao cérebro e produzem sensações luminosas aproximáveis por uma convolução (Gaussiana) do sinal capturado por uma função de espalhamento pontual (PSF). Portanto, há uma diferença mínima entre dois elementos (cor, forma, tamanho) para que eles sejam percebidos como distintos, conhecida por **acuidade** visual. Outra característica é a ação inibidora fora do centro de campo receptivo de um fotorreceptor, de forma que nas regiões onde ocorrem variações abruptas de luz, elas aparentam ser mais claras ou mais escuras, criando ilusões como **Bandas de Mach**. Outra característica visual interessante é a conhecida **aberração cromática**. Ela é causada pela diferença dos pontos focais de radiações de comprimentos de