

Introdução à Visualização: um Panorama

Maria Cristina F. de Oliveira
Rosane Minghim
Depto. de Ciências de Computação
ICMC, USP

2010/2011/2012

Panorama

- Motivação
- Visualização Científica
- Visualização de Informação
- Mineração Visual e Analítica Visual
- Desafios
- Bibliografia

3

Desafio

- É preciso tornar os dados acessíveis e fornecer às pessoas recursos adequados para interpretá-los
 - <http://www.gapminder.org/>
 - Ferramentas de visualização de dados
 - Organização com fins não lucrativos cujo objetivo é tornar acessíveis e transparentes dados sobre o desenvolvimento humano mundial
 - Fontes: UM, UNESCO, ONGs,
 - *Search statistics through Google and watch it move with Gapminder*

2

Motivação

Metáforas visuais fazem parte do processo cognitivo humano...

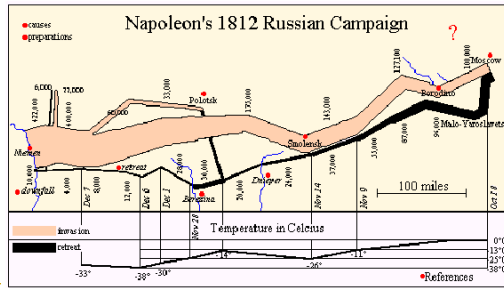
Cognição: a aquisição ou o uso de conhecimento

“Visualizar: construir na mente uma representação visual”
Shorter Oxford English Dictionary

Definição “tradicional”

4

<http://www.edwardtufte.com/tufte/minard>



5

Visualização

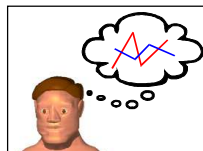
"Uma representação gráfica de dados, ou conceitos"
C. Ware, Information Visualization, Perception for Design

"O uso de representações visuais de dados interativas e apoiadas por computador, para ampliar a cognição"
Card et al., Readings in Information Visualization

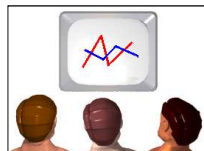
Definições "contemporâneas"

7

Visualização como Artefato



Construção Interna



Artefato Externo

Suporte à tomada de
Decisão

6

Visualização

- Área de pesquisa que estuda o uso de representações gráficas (metáforas visuais) para apoiar tarefas de exploração, análise e interpretação de dados
- Modelos gráficos acoplados a estratégias de interação: exploração dinâmica de dados
- A partir de 'dados', gera representações gráficas (imagens) interpretáveis pelas pessoas

8

Visualização

Hamming 1973: "*the purpose of computation is insight, not numbers*"

Card et al. 1999: "*the purpose of visualization is insight, not pictures*"

Principais objetivos desse "*insight*": descoberta, verificação de hipóteses, tomada de decisões, explicação

A Visualização é útil na medida em que amplia a nossa capacidade de executar essas e outras tarefas cognitivas

9

Visualização

■ Contraposição com

□ Visão Computacional

- Ponto de partida são imagens adquiridas, que são processadas e analisadas para extrair informação útil
- Análise das imagens pelo sistema computacional (processo automático)
- Objetivo seria dotar um sistema de discernimento visual comparável ao do ser humano

11

Visualização

■ Motivação

- Muito fácil coletar e armazenar dados
- Muito difícil processar, analisar e interpretar todos os dados coletados, identificar o que é relevante
- Volume dos dados muito grande
- Dimensionalidade dos dados muito alta
- Natureza dos dados muito diversa
 - Registros, textos, imagens, vídeos, voz, ...

■ Desafios para pesquisadores...

10

Visualização

■ Contraposição com

□ Computação Gráfica

- Ponto de partida são modelos geométricos, a partir dos quais gera imagens 'realistas' – foco é 'ilusão', não interpretação
- Entretanto, modelos geométricos precisos as vezes são gerados a partir de dados 'reais'... animações realistas podem ser geradas a partir de simulações de processos físicos...
- CG ou Visualização??!

12

Reconstrução

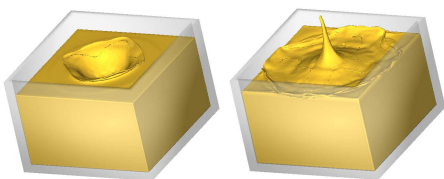


Créditos: Helton Biscaro, LCAD, 2005

Visualização

- Visualização no Processo de Análise de Dados
- Conceitos básicos
 - Comp. Gráfica, Dados, Malhas, Técnicas, Pipeline de Visualização Exemplos
- Visualização Planar e Volumétrica
 - Escalares, Vetoriais, Tensores
 - Imagens e Volumes
 - Visualização Médica
- Representação de Dados
- Visualização de Informação
 - Pontos, Atributos, Grafos, Projeções, Aplicações
- Mineração Visual de Dados

Escoamentos



Computing & Visualization in Science, 2000
Int. J. Numerical Methods in Fluids, 2002

Visualização

- Periódicos
 - The Visual Computer, Computer Graphics Forum, IEEE Computer Graphics and Applications, IEEE Trans. Visualization and Computer Graphics, Computers & Graphics, Information Visualization
- Conferences
 - Eurovis/EuroVa
 - IEEE Visweek (Vis/Infovis/VAST, etc.)
 - IV (Londres, Julho)
- Livros (provisório)
 - Principais
 - Telea, A. Data Visualization – Principles and Practice, A.K.Peters, 2008.
 - MINGHIM, R.; OLIVEIRA, M. C. F.; JAI 03 - "Uma Introdução à Visualização Computacional", JAI'97 - Jornadas de Atualização em Informática, XVII Congresso da SBC, Brasília, 2 a 8 de Agosto de 1997, pp.85-131.
 - SCHROEDER, W.J.; MARTIN, K.; LORENSEN, W. - The Visualization Toolkit - An Object-Oriented Approach to 3D Graphics, Prentice-Hall, 4ª edição 2006.

Visualização

■ Livros:

□ Leitura Complementar

Mingham, R.; Levkowitz, H – Visual Mining of Text Collections, Eurographics Tutorial Notes 11, 2007.

Maria Cristina Ferreira de Oliveira, Haim Levkowitz: From Visual Data Exploration to Visual Data Mining: A Survey. IEEE Trans. Vis. Comput. Graph. 9(3): 378-394 (2003).

Spence, R. Information Visualization, Addison Wesley, 2001.

Usama Fayyad, Georges G. Grinstein, Andreas Wierse (eds.) Information visualization in data mining and knowledge discovery, Morgan Kaufmann, c2002.

Chen, Chaomei – Information Visualization, Springer 2004.

Vladimir Geroimenko and Chaomei Chen - Visualizing the Semantic Web : XML-based Internet and information visualization., Springer, 2006.

... Demais livros: no site e na biblioteca

Site: wiki.icmc.usp.br

Visualização

■ Científica

Dados Médicos, Meteorológicos, Fluidos

■ de Informação

Bases de Dados, Web, Carga em Redes, Relações Humanas

■ de Software

Projeto, Depuração, Teste

Visualização Avaliação 2012

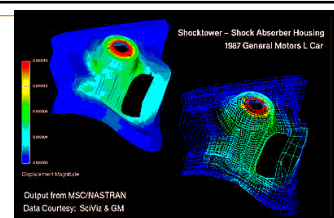
Obs: (ver syllabus/plano de curso no site wiki.icmc.usp.br para instruções mais detalhadas)

1. Resumos bibliográficos (10%): Dois por semana, revisados e submetidos no horário da aula. Bibtex deve acompanhar.
2. Duas apresentações curtas em classe (15%,15%): Vinte minutos cada, sobre temas associados aos assuntos do curso. Meio e fim do curso.
3. Dois 'surveys' curtos (15% 15%): Podem acompanhar as apresentações no que se refere ao tema, mas deve incluir conhecimento adicional ao apresentado.
4. Projeto (30%): Implementação significativa. Sugerir temas rapidamente e conferir com o ministrante antes de iniciar.
5. Adicionais por participação em discussões e questões em aula.

- **First Assignment (entrega 21/08 prox.):** Procurar nos *Proceedings* do EuroVis 2012/2011 e IEEE VisWeek (Vis/InfoVis/VAST), paper(s) de autores que deixem o código e/ou dados disponíveis. Recuperar esse código e/ou dados para deixar disponível numa página específica. Fazer o resumo conforme item 1. acima.

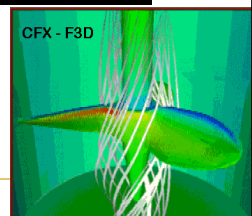
ViSC

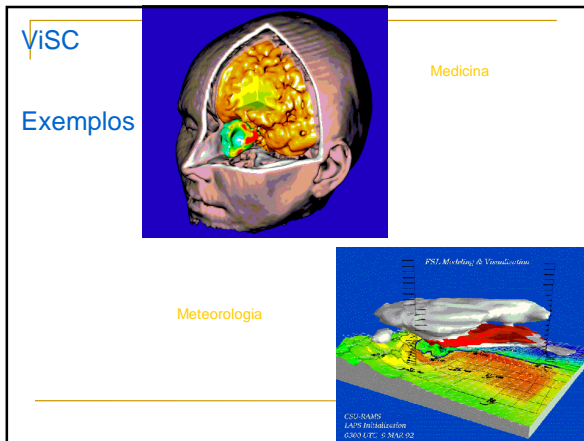
Exemplos



Engenharia

Simulação - CFD





Visualização Científica

- Dados 'científicos'
 - Têm uma representação geométrica conhecida, real ou imaginária. P.ex., órgãos do corpo humano, moléculas, ...
 - Resultantes de processos físicos, medidos, coletados ou simulados: atributos espaciais (e temporais) são determinantes para a visualização
- Interação com CG, HCI, matemática...
 - 'Reconstrução' de modelos geométricos, simulação numérica, *rendering* adequado dos modelos

Visualização de Informação

- Dados 'abstratos'
 - Não têm uma representação geométrica (espacial) inerente; qualquer representação adotada é arbitrária (atribuída)
 - Ex.: transações de clientes em uma base de dados, acessos a um *site* na Web, movimentação financeira na bolsa de valores, hierarquia de diretórios, coleções de textos, ...
 - Em geral, coletados, medidos ou 'criados'
- Interação com HCI, 'design' gráfico, semiótica, estatística, mineração...

Visualização de Informação

- Visualização Multidimensional
- Visualização Exploratória
- Metáforas visuais para ajudar pessoas a explorar/analisar dados
- Modelos gráficos acoplados a estratégias de interação ⇒ processos dinâmicos de exploração

25

Visualização de Informação

Exemplos de Produtos Comerciais

- Inxight TableLens e Star Tree
 - <http://www.inxight.com/products/sdks/tl/>
 - <http://www.inxight.com/products/vizserver/>
 - Demos
- SpotFire DecisionSite
 - <http://www.spotfire.com>

27

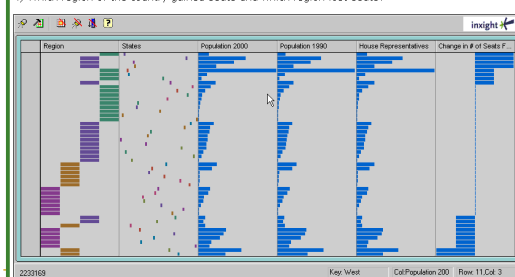
Visualização de Informação

- Dados estruturados
 - Tabelas de registros com múltiplos atributos (numéricos ou categóricos)
- Dados não estruturados
 - Textos, imagens, ...

26

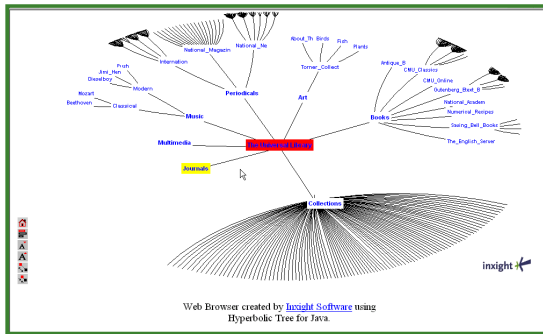
This data includes the resident population of the 50 States from Census 2000 and from Census 1990. It also includes the changes in the number of House Representatives in each state.

- 1) Which states gained the most seats due to the 2000 Census?
- 2) Which states lost the most seats?
- 3) Which 5 states are the most populated? How about 10 years ago?
- 4) Which region of the country gained seats and which region lost seats?



28

Exemplo: Hyperbolic Tree



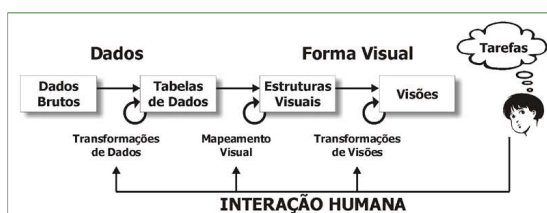
29

Comparação ViSc x InfoVis

- Em ambos os casos, ocorre um processo de *'especialização'* da informação: dados são mapeados em um espaço 2D ou 3D
- Especialização determina a representação geométrica visível com a qual o usuário interage
- *ViSc*: geometria do modelo (explícita ou não) **determinada** pelo domínio
- *InfoVis*: geometria do modelo **atribuída** pelo *'designer'* da representação visual

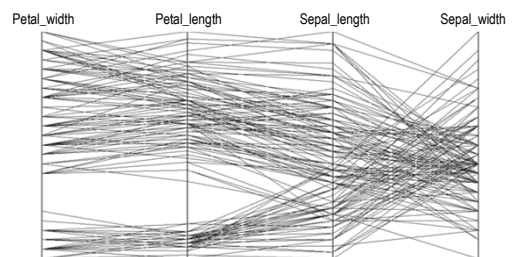
31

Processo de Visualização: Modelo de Referência (Card et al. 1999)



30

Exemplo: Coordenadas Paralelas

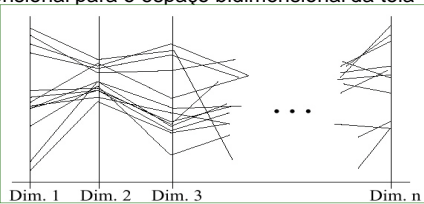


Iris flower data: 3 classes de flores

32

Exemplo: Coordenadas Paralelas

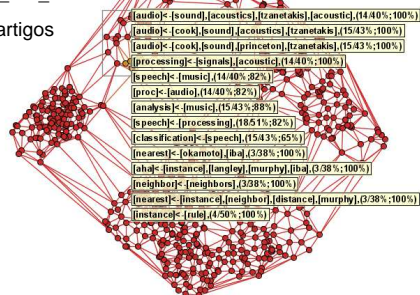
- Inselberg (1985) – geometria computacional
- Projeta instâncias definidas em um espaço n-dimensional para o espaço bidimensional da tela
- n-eixos
- dos eixos
- Eixos de valores
- Instâncias poligonais



33

CBR_ILP_IR

574 artigos



<http://www.lcad.icmc.usp.br/~paulovich/>
Paulovich & Minghim, IV 2006, p. 245-251

35

Exemplo: *Hyperbolic Tree*

- Demo
- <http://www.inxight.com/products/sdks/st/>

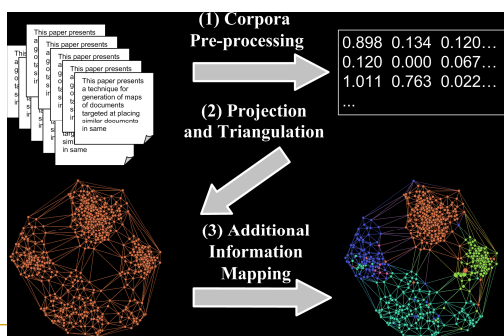
34

Mapas de documentos

- projeção 2D de instâncias de dados definidas em um espaço multidimensional (nD)
- Instâncias são corpus de documentos
 - Cada texto representado por um vetor nD
 - n determinado pelo número de termos na coleção
 - Em geral, n é grande, i.e., dimensionalidade alta
- Pontos próximos são ‘similares’
 - Definição de ‘similaridade’ entre documentos
 - P.ex., medidas de distância entre os vetores nD

36

Mapas de documentos



Paulovitch & Minghim, IV 2006, p. 245-251

37

Mapas de documentos

- Projeção: redução de dimensionalidade
 - Pontos próximos no espaço nD devem ser mantidos próximos no espaço 2D
 - Várias maneiras de projetar...
- Mapas: exploração interativa de coleções de documentos texto
 - IDMAP, ProjClus, LSP (*Least Squares Projection*)
 - PEX: Projection Explorer, e PEX-Web
 - R. Minghim, F. Paulovitch e R. Pinho

39

The screenshot shows the PEX-Web interface. On the left, there is a sidebar with a list of documents, including 'Creative Visualization - Attracting Success with...', 'The Power of Visualization', and 'The Healing Power of the Mind and Visualization'. The main area displays a network graph with many red nodes and edges. A search bar at the top contains the word 'creative'. The interface includes various controls for visualization, such as 'Color', 'Size', 'Edges', and 'Number documents'.

<http://infoserver.lcad.icmc.usp.br/infovis2/PEXWeb>

38

Visualização

- Ferramentas de Análise de Dados
 - Estatística, Mineração, Visualização
- Por que Visualização
 - Habilidade de expressar muita informação
 - Percepção de propriedades não antecipadas
 - Facilita a percepção simultânea de características dos dados em grande e pequena escala
 - Apoio a processos de formação de hipóteses
 - Apoio a tarefas de pré-processamento dos dados
 - Detecção de problemas, limpeza, seleção, ...

40

Visualização - Desafios

- Processos cognitivos difíceis de serem modelados
 - Ainda falta muito para que técnicas sejam usadas de forma efetiva por uma gama ampla de usuários
- Criar "boas" representações visuais não é trivial
- Disponibilizar ferramentas efetivas não é simples (funcionalidades e interação)
- Escalabilidade

41

Visualização - Desafios

- Mineração
- Interação

43

Visualização - Desafios

- Excesso de dados, de natureza complexa
 - Enormes volumes, dimensionalidade alta (muitos atributos), diferentes tipos de dados, diferentes organizações, diferentes mídias, *streaming data*...
- Como tratar tanto volume e variedade?
- Como saber o que é realmente relevante?

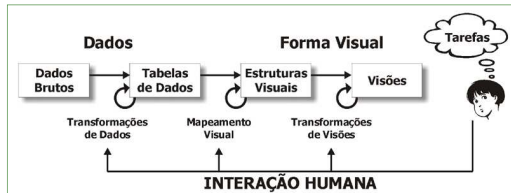
42

Papel da interação

- Aumentar a escalabilidade visual, dada a limitação física dos dispositivos
 - Limitar a quantidade de informação
 - Manter o contexto global
 - Visão geral e detalhe
 - Filtragem
- Viabilizar a visualização simultânea de diferentes representações
 - Coordenação entre múltiplas visualizações
 - *Link-and-brush*

44

Papel da interação



45

Papel da Mineração

Visualização direta dos dados
vs
Visualização de conteúdo/informação
'relevante' embutido nos dados

47

Papel da interação

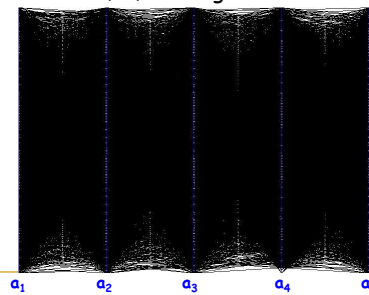
- *Visual Information Seeking mantra*
- Overview first, zoom and filter, then details-on-demand [Shneiderman, 1996]

46

Problema

Dados sintéticos, 7,500 registros 5-d

Como analisar???



48

Problema

- Instância do problema (mais genérico) de mapear individualmente itens de informação
 - Sobreposição e sobrecarga visual
 - Várias abordagens na literatura buscam identificar e realçar informação relevante em visualizações 'sobrecarregadas'

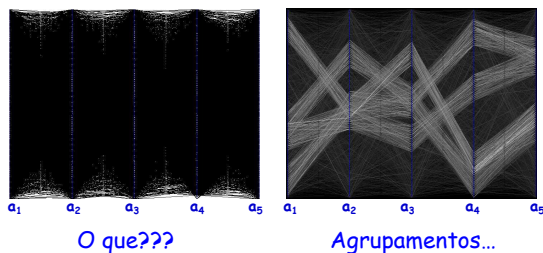
49

Mineração Visual

- Contraposição com Mineração
 - Etapa do problema (mais geral) de extrair conhecimento de dados (Aprendizado de Máquina)
 - Ponto de partida são arquivos de dados, processados *automaticamente* com o objetivo de extrair modelos que 'explicam' os dados
 - Modelos têm por objetivo descrever ou prever o comportamento dos dados
 - Exemplos: classificar, agrupar, identificar associações, ...

51

IPC Plots



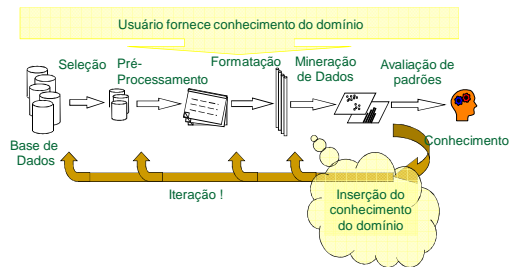
50

Mineração Visual

- Convergência de Mineração e Visualização
- Metáforas visuais para apoiar usuários de tarefas de mineração
- Motivação
 - 'Minerar' dados automaticamente é difícil, e nem sempre produz resultados compreensíveis/úteis
 - Visualizar dados brutos (sem minerar, ou extrair modelos) nem sempre é possível/interessante/útil

52

Processo de Descoberta "Centrado no Usuário"



Ankerst 2000

53

Categorias de VDM (Ankerst 2000)

- Visualização dos dados
 - Usuário tem total controle sobre a busca por padrões
 - Focalizar/delimitar espaço de busca
- Visualização dos resultados de uma mineração
 - Apoiar a interpretação dos modelos extraídos
- Visualização acoplada ao processo de mineração
 - Direcionar a busca
 - Fornecer conhecimento sobre o domínio, por exemplo, para adaptar um núcleo genérico (para diferentes aplicações) com a intervenção do usuário
- Mineração acoplada ao processo de visualização?

55

Mineração Visual de Dados – Definição (Ankerst 2000)

- “VDM é um passo no processo de extração de conhecimento (KDD) que utiliza a visualização como um canal de comunicação entre computador e usuário para apoiar a identificação de padrões novos e interpretáveis”
 - Posicionamento nas duas últimas fases do processo: mineração de dados e avaliação
 - Identifica três categorias de VDM

54

Categorias de VDM (Wong 1999)

- Fracamente acoplada
 - Visualização "intercalada" com estratégias analíticas de mineração
 - Apoiar pré-processamento, interpretação de resultados,...
 - Abordagem limitada: reforça limitações de uma e de outra...
- Fortemente acoplada
 - Visualização "integrada" em estratégias analíticas de mineração
 - Dar ao usuário maior controle e entendimento sobre o processo analítico, apoiando a tomada de decisões
 - Criação de representações visuais do espaço de busca

56

Desafios, segundo, segundo D. Keim (IV 2006)

- Mantra modificado
- “*Analysis first, show the important, zoom, filter & analysis, details on demand.*”
- Análise em tempo real, de informação em larga escala e dinâmica, vinda de fontes diversas (precisa ser integrada), em diferentes formatos e resoluções (precisa ser processada na resolução adequada)

57

“Visual Analytics is the science of analytical reasoning facilitated by interactive user interfaces”

“Visual analytics requires interdisciplinary science beyond traditional scientific and information visualization to include statistics, mathematics, knowledge representation, management and discovery technologies, cognitive and perceptual sciences, decision sciences”

Illuminating the Path: The Research and Development Agenda for Visual Analytics, IEEE Press, 2005

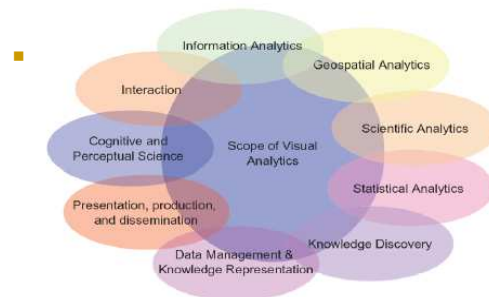
59

Desafios: *Visual Analytics*

- Science of Analytical Reasoning
- Ampliar a capacidade humana de análise de informação é problema estratégico
 - 2004: National Visualization and Analytics Center, EUA (<http://nvac.pnl.gov>)
 - *Department of homeland security*
 - Foco: contra-ataque ao terrorismo

58

Desafios: *Visual Analytics*



<http://nvac.pnl.gov/>

60

Bibliografia

- Card, S.K.; Mackinlay, J.D.; Shneiderman, B. (eds.) Readings in Information Visualization, Using Vision to Think. Morgan Kaufmann, San Francisco, CA, 1999.
- Keim, D.A. Visual Exploration of Large Databases. Communications ACM, Vol. 44(8), 2001.
- Keim, D.A. Information Visualization and Visual Data Mining. IEEE Trans. Visualization & Computer Graphics, Vol 8(1), pp. 1-8, 2002.
- Wong, P.C.; Bergeron, R.D. 30 Years of Multidimensional Multivariate Visualization. In G.M. Nielson et al. (eds.): Scientific Visualization Overviews, Methodologies and Techniques, IEEE Computer Society Press, Los Alamitos, CA, pp. 3-33, 1997.

61

Bibliografia

- Friedhoff, R.M. Visualization, The Second Computer Revolution, 1991.
- Keller & Keller. Visual Cues, Practical Data Visualization, IEEE CS Press, 1992.
- Spence, R. Information Visualization, 1. ed., ACM Press, 2000.
- Ware, C. Information Visualization, Perception for Design. Morgan Kaufman Pub., Academic Press, 2000. 438p.
- Illuminating the Path: The Research and Development Agenda for Visual Analytics, IEEE Press, 2005

63

Bibliografia

- Oliveira, M.C.F.; Minghim, R. Uma Introdução à Visualização Computacional, JAI Jornadas de Atualização em Informática, XXVII Congresso da SBC, 1997, 42p.
- Freitas, C.M.D.S.; Chubachi, M.O.; Luzzardi, P.R.G.; Cava, R.A. Introdução à Visualização de Informações. Revista de Informática Teórica e Aplicada, Vol. 8(2), pp. 143-158, 2001.
- Nascimento, H.A.D.; Ferreira, C.B.R. Visualização de Informações, Uma Abordagem Prática. JAI Jornadas de Atualização em Informática, XXV Congresso da SBC, 2006, 51p.

62

Bibliografia

- Hibbard's Top Ten Visualization Problems, em <http://www.siggraph.org/publications/newsletter/v33n2/columns/hibbard.html>
- Johnson's Top Scientific Visualization Research Problems, IEEE Computer Graphics & Applications, July-Aug. 2004 24(4), pp. 13- 17.
- Conferências: *IEEE Symposium on Information Visualization* e *Information Visualization* (acessíveis na IEEE Digital Library) e *IS&T/SPIE Visual Data Analysis*

64