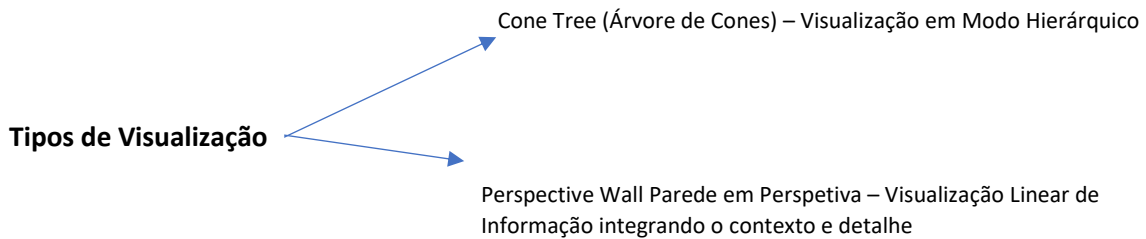


Outline – Representações Gráficas para diferentes tipos de Conjuntos de Dados

Onde surge a expressão:

“Visualização de Informação”

A expressão visualização de informação foi expressa em 1991 associada à visualização de dados onde não existiam modos adequados de serem representados nos sistemas de visualização científica.



A visualização de informação emerge como extensão da visualização científica, uma vez que existem dados que:

- Requerem apresentações não fornecidas pela Visualização Científica
- Não dispõem de primitivas georreferenciadas das grelhas, não tendo um posição espacial predefinida.

A localização espacial é um dos problemas da visualização de dados que maior parte dos casos para definir uma representação gráfica é necessária a especificação dos elementos gráficos e as suas posições.

A visualização de informação era associada à Visualização de dados abstratos.

Dados vs Informação (Temos os dados extrai-se info.)

A visualização de informação visa representar os dados e extrair informação.

Tipos de Dados de Visualização de Informação:

Podemos classificar os dados por:

- **Nível de Atributo:**
 - Quantitativo (Quantidade)
 - Categoria → Ex: Quais os tipos de transformação de Fourier são possíveis?
 - Classificador
- **Nível de Conjunto de Dados**

Tipo de Atributos

Sem ordem....

Categoria: Não tem uma ordem implícita dos dados, apenas se distinguem pelo princípio de igualdade, exemplo 1 maçã = 1 maçã, 1 maçã ≠ 1 Laranja.

Com ordem...

Quantitativos: Dados, por exemplo idade que corresponde à magnitude da amostras

Ordinal: Dados com ordem bem definida.

Exemplos: →

Name	ID	Number of boxes	Size
Persil	253890	10	L
Persil	253750	8	M
Skip	360920	15	L
Xau	420840	14	S

Number that represents a **code**
It is a **category** that happens to be a **number** rather than a **textual name**
Categorical data

Number that represents a **quantity**
Adding two quantities together makes sense
Ordered data

Textual name, but it has **order**

Tipos de Conjuntos de Dados

Os conjuntos de Dados podem ser classificados por:

- Estruturas Tabulares
 - Linear
 - 2D e 3D
 - Multidimensional e Multivariável
- Estruturas Complexas
 - Cone Trees (Estruturas Hierárquicas)
 - Graphs
- Texto

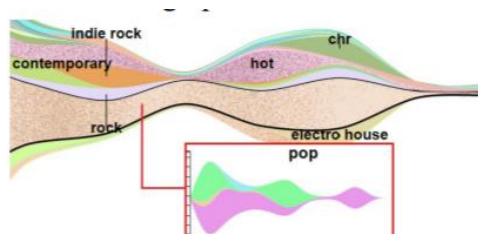
Estruturas Tabulares

Dados Lineares

Dados lineares são definidos como dados mono dimensionais espacialmente. Os dados mais comuns lineares correspondem a observações ao longo do tempo.

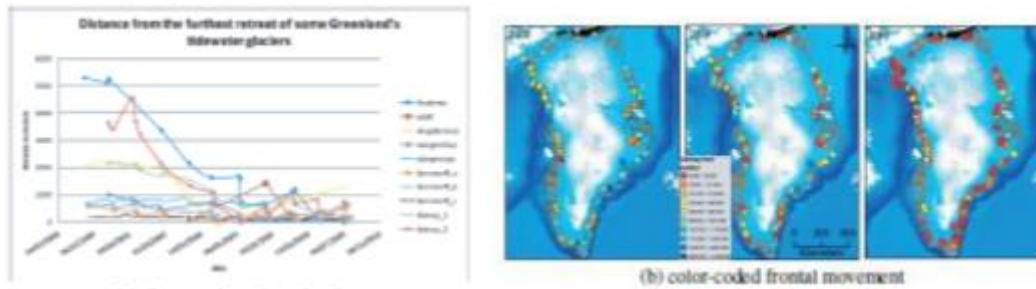
THEME RIVER

É definido como forma de visualização de evolução temporal do conteúdo temático de uma coleção de documentos. Os tópicos dos documentos são representados em bandas horizontais justa posicionadas. Cada cor de banda corresponde a um tema. A largura da banda varia pelo tempo em acordo com o número de documentos de um dado tema. Theme River é um caso especial de Stacked Graphs.



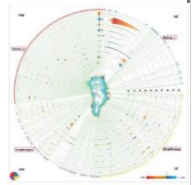
Visualização Espaço-Tempo (2D) e Temporal (1D)

Representam-se glifos pela cor de representação a diferentes escalas, classificando pela cor dada a magnitude do classificador.



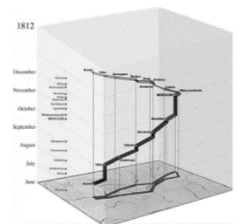
Space-Time Visualization 2D

São representadas linhas radiais que correspondem às diferentes costas de fronteira. São dispostos anéis alinhados representando a evolução do tempo pela dimensão do raio interno para o exterior.



Space-Time Cube (3D)

Representa trajetórias de movimento do objeto pela orientação tridimensional das componentes.



DADOS MULTIVARÁVEIS OU DADOS MULTIDIMENSIONAIS

Dados Multivariados – São dados que podem ser definidos num espaço de n dimensões ou que contrariamente não tem localização explícita. São dados relativos a várias variáveis. Não são necessárias informações de dependência entre as variáveis possa ainda assim haver correlação entre as mesmas.

Dados Multidimensionais – São dados que são definidos no espaço euclidiano n dimensional no qual a posição de um ponto é dada por n coordenadas, havendo por isso dependência entre as variáveis.

Objetivos

Dados Multivariados – visam visualizar diferentes variáveis simultaneamente.

Dados Multidimensionais – visam visualizar o valor de correlação de uma variável para outra.

Notas

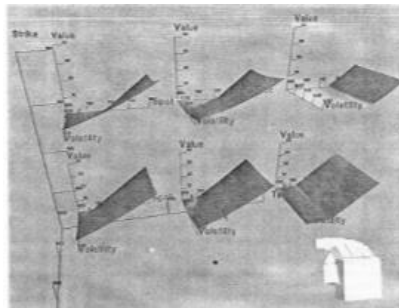
Tanto na visualização de dados multidimensionais como de multivariados o espaço de representação pode ser menor que o espaço onde os valores são definidos.

Pelo uso de animação é possível visualizar uma variável extra às que são inicialmente apresentadas, visando a componente 3dimensional tornar-se numa 4ª.

Para dimensões maiores é necessário uso de técnicas para reduzir o número de dimensões e poderem ser visualizadas simultaneamente.

Exemplos de Dados Multidimensionais:

Mundos dentro de Mundos, visam reduzir a complexidade de uma função dependente das variáveis.



O espaço n-dimensional é reduzido a 3 dimensões, usando cada variável também definida a 3 dimensões. Existe uma imersão do espaço 3D num novo espaço 3D orientado pela referência do espaço inercial com a relação do valor das variáveis.

Dados Multivariados

Tipos de Representação:

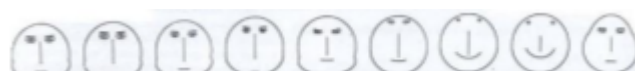
1. Ícones ou Glifos
2. Texturas
3. Scatter Plots – Nuvens de Pontos
4. Parallel Coordinates – Coordenadas Paralelas

1. Ícones (Glifos)

Glifo ou ícone é uma representação visual cujos atributos geométricos e gráficos podem refletir valores de variáveis em estudo.

Ícones são uma representação comum para os dados multivariados. Acabam por explorar a capacidade humana de identificar relação espacial e o comportamento de relação entre múltiplas variáveis simultaneamente.

Exemplo de visualização simultânea de um conjunto de variáveis delimitadas a um número de pontos, associando cada variável a correspondentes expressões faciais - **Chernof Faces**



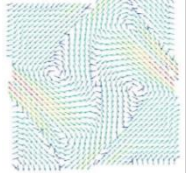
TimeWheel e InfoBug – Ícones usados para representar conjuntos de dados em projetos de gestão de software.

Ícones: Glifos, Chernof Faces, TimeWheel, Info Bug

Textures

Em alguns casos, o uso de glifos corresponde a analisar texturas. A sua visualização distribui-se pelas instâncias e respetivos valores de uma amostra.

A textura pode variar em função: **Escala, Regularidade e Intensidade**



Scatter Plotts – Nuvem de Pontos

A nuvem de Pontos é uma representação 3D ou 2D. Cada variável é associada a uma posição (x,y) ou (x,y,z) pelos valores e magnitude que ocuparem.

Scatter Plotts são combinados com glifos (e ainda animação) ----> Bubble Charts

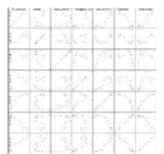
Bubble Chart é similar à nuvem de pontos, contudo a Scatter Plotts representa 2 variáveis (x,y), uma bubble chart pode representar uma terceira variável pelo raio do seu círculo.

A nuvem de pontos pode ser estendida a dados multivariados pelas Visualizações:

- Scatter Plot Matrix
- Hyper Box

Scatter Plot Matrix – Matriz de Nuvem de Pontos

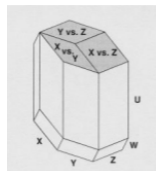
Havendo n variáveis, o ecrã é dividido em grelha n*n, e cada variável é associada a uma posição linha e coluna. Em cada área de grelha é desenhada a scatter plot respetiva que relaciona os valores de 2 variáveis associadas a uma linha ou coluna que se interseitam.



Desvantagens - - - - - > Apenas podem-se comparar variáveis por par e não as variáveis num modelo global, tornando toda a visualização e interação menos eficiente.

HyperBox

A hyperbox é uma representação 2D sólida de n dimensões onde estão as fronteiras com diferentes direções. Cada variável é associada a uma direção de fronteira.



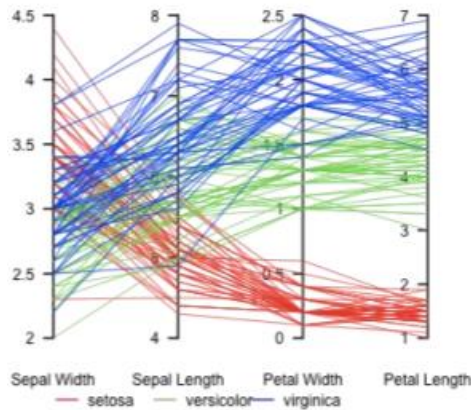
PARALLEL COORDINATES – Coordenadas Paralelas

No espaço 2d são consideradas linhas equidistantes paralelas ao eixo Y formando um sistema de coordenadas paralelo.

Os dados são tratados como um conjunto de pontos num espaço n dimensional onde cada componente é associada a um eixo. O número de eixos é portanto de n dimensões limitados à tela de ecrã.

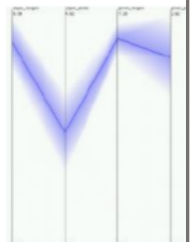
1 Ponto é representado por uma linha poligonal conectando as posições determinadas aos valores de cada eixo do respetivo componente.

À medida que o número de dados aumenta torna-se difícil de distinguir as linhas e respetivos valores de variável.



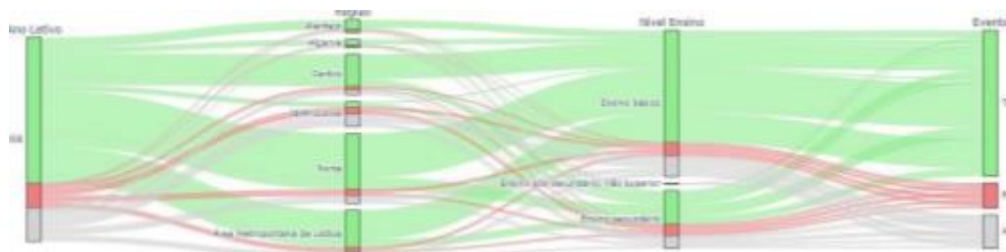
Hierarquical Parallel Coordinates (Coordenadas Paralelas Hierárquicas)

Para a visualização da grandes bases de dados é usada a técnica de clustering. Os pontos de um cluster são representados pela linha de valor médio, acompanhada de uma região inferior e superior alusivos a intervalos de valores máximos (acima) e mínimos (abaixo). Estas regiões são de opacidade e transparência progressiva.



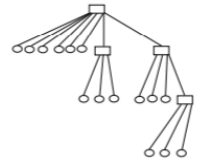
Parallel Sets

Conjuntos de Dados para visualização de dados categóricos que mostram a frequência de dados em vez de valores individuais. O método é baseado no layout do eixo de coordenadas paralelas, com caixas a representar categorias e paralelogramas entre os eixos, mostrando a relação entre cada categoria.



Hierarquical Structures

A visualização de estruturas hierárquicas muitas das vezes causa dificuldades na gestão do espaço de tela, que pioram à medida que os elementos e dados da estrutura aumentem



Uma das representações mais populares de estruturas hierárquicas é o gráfico bidimensional

Para hierarquias com um grande número de nós não é possível visualizar toda a estrutura imediatamente na área de desenho.

CONE TREES

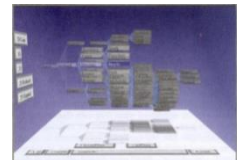
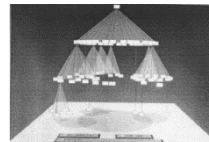
São um modelo 3dimensional. A hierarquia pode ser representada no plano vertical (**conetree**) ou horizontal (**camtree**)

A dimensão dos cones, tanto a sua altura como o seu diâmetro são calculados pela ordem de representação na estrutura.

Os cones são transparente por forma a poder ser perspectivada existência de cones por detrás ou escondidos.

A interatividade permite:

- Tornar invisíveis alguns dos ramos da árvore
- Restrutura a arvore pela manipulação dos cones
- Rodar elementos
- Procurar informação

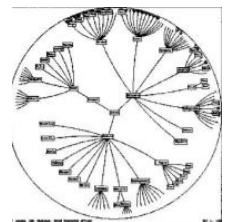


Hiperbolic Trees

Ao contrário do espaço euclideano, no epaço hiperbólico as linhas paralelas não são equidistantes. É possível construir duas linhas que não se cruzam e cujas distâncias aumentam progressivamente à medida que se afastem. O espaço hiperbólico pode ser projetado no espaço euclideano em 2 modos:

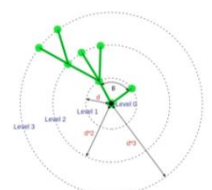
1 – Mantendo as Linhas retas e os angulos de inclinação serem definidos --- 3D

2 - Deformando as linhas e mantendo os angulos ----- 2D



Radial Trees

Este tipo de representação tem similaridades com a Hiperbolic Tree, contudo a diferença reside em que as Hiperbolic Trees são baseadas na geometria hiperbólica enquanto a Radial Tree a distância entre órbitas é relativamente linear.



Tree Map

Tree Map é um modelo bidimensional que visa utilizar de forma eficiente a área de desenho. A representação da estrutura hierárquica é constituída através do preenchimento de campos retangulares:

- A raiz da hierarquia corresponde ao maior retângulo que ocupa a área de desenho, reservada para a visualização da estrutura.
- A raiz deste retângulo é dividida verticalmente em retângulos correspondentes ao nível abaixo da raiz.
- A área de cada retângulo é depois dividida, mas agora horizontalmente em retângulos correspondentes aos descendentes, e assim sucessivamente.

A posição, tamanho e cor de cada retângulo irão corresponder às propriedades dos elementos que formam a hierarquia.

Treemap (1)



CHEOPS

Representação hierárquica 2D baseada na compressão overlapping dos nós.



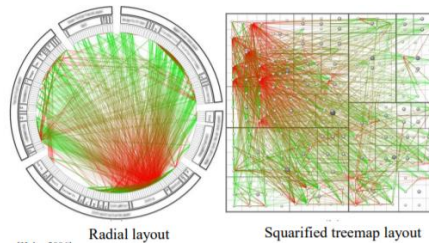
Estruturas Hierárquicas

- **Tridimensional** – Mostram a estrutura, porém têm problemas de oclusão
- **Hyperbolic** – Permitem representar um largo nº de nós porém geram distorção
- **Radial** – A raiz da hierarquia é o centro do círculo
- **By Filling Areas** – Mostram árvores hierárquicas sem problemas de oclusão porém de difícil identificação da estrutura
- **Incomplete** – Não mostram nós simultaneamente, porém nunca perdem a organização dos nós

Hierarchical Edge Bundles

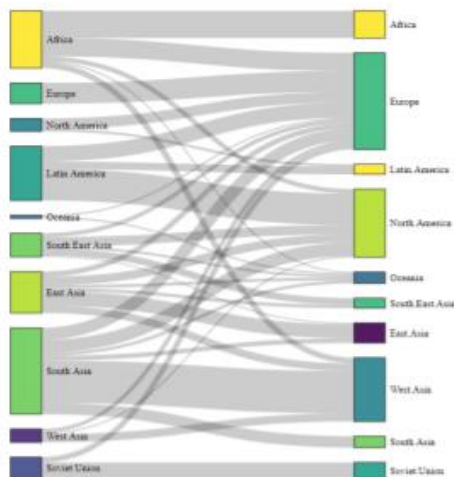
Pacotes de Fronteira Hierárquica – Visualização de Relações de Adjacência em Dados Hierárquicos

Há uma grande classe de conjunto de dados que contem componentes dos dados hierárquicos ou seja pela relação pai-filho entre dados, bem como de componentes não hierárquicos que representam relações entre os dados, denominadas relações de adjacência.



Sankey Diagrams

Diagramas Sankey são uma técnica de visualização que permite a representação de fluxos. Um número de nós é representado por retângulos e são ligados por setas e arcos com larguras proporcionais à importância de fluxo.



Grafos e Dados Georreferenciados

Exemplo de Mapa Suburbano de Londres,

Regras De BECK



- Linhas de Transporte sinuosas são endireitadas e restritas horizontalmente, verticalmente ou diagonais a 45º
- Escalas em área central cotada é maior do que nos subúrbios menos densos
- Apesar da distorção a topologia da rede e o senso de geometria mantém-se.

Dados Georreferenciados - Color Mapping /// Cloropleth Maps



Fornecem uma maneira de visualizar valores de uma dada área geográfica que pode mostrar variação ou padrões distribuídos geograficamente.

- A variável de dados será a progressão de cores para representar em cada região do mapa.
- Uma desvantagem do uso de cores é que não existe possibilidade de serem comparados os valores e instâncias nos mapas.
- Outro problema é que as regiões maiores apareceram mais enfatizadas que as menores, perturbando a percepção do visualizador do sombreado e progressão de cor entre regiões periféricas.

Um erro comum ao produzir mapas Cloropleth é codificar dados brutos (ex: População absoluta) em vez de usar valores normalizados.

TEXT

Existem dois tipos de abordagens textuais:

- 1- A visualização do texto estruturado em sequências de linhas (Listar Linhas)
- 2- A visualização de conteúdo semântico de um ou mais documentos

Colection of Documents

Para visualizar uma coleção de documentos, precisamos de metadados sobre os documentos, exemplo:

Sujeito;

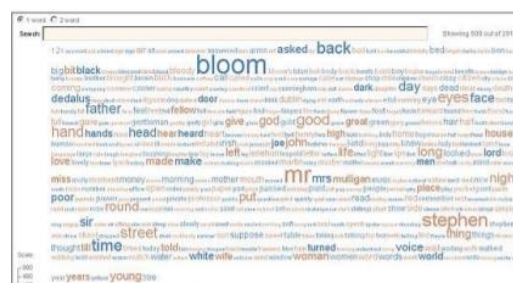
Palavras Chaves

Referencias

Por isso é necessário um pré-processamento inicial do texto antes da geração da sua visualização. Para estabelecer a ligação entre os documentos é necessário adotar medidas para determinar o seu grau de similaridade.

TAG CLOUD - Nuvem de Termos

Representa a frequência de palavras num texto. O tamanho de cada palavra é ampliado/ diminuído com forme a frequência da palavra no texto.

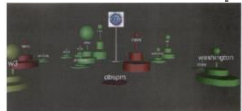


Visualizing the Internet

Para visualizar a internet considera-se:

- A – A visualização de um conjunto de documentos obtidos via internet
- B – A visão dos caminhos efetuados
- C – A visão das estruturas dos Sites

Visualization Landscape ---- Paisagem de Informação: Consiste em colocar no plano 3dimensional objetos do plano bidimensional



Dimensão da Representação Espacial (2d, 3d)

O resultado final é sempre uma imagem bidimensional, contudo o espaço onde definida e gerada pode ser 2D ou 3D.

O que distingue a representação 3D?

A coordenada espacial do utilizador é 3dimensional. Podem ser usados cones, cubos, figuras volumétricas.

Para ser realizada a transformação do espaço 3d para o espaço 2D gráfico é necessário

- **Selecionar o Tipo de Projeção**
- **Selecionar a posição do observador**

A interação dá ao utilizador a ilusão 3dimensional no espaço.

Representação 2 (½) D

Consiste em utilizar objetos 3dimensionais num plano 3d, quando é necessário esta representação usa 1 ou + técnicas para fornecer a perceção de profundidade, como por exemplo oclusão de elementos, projeção de sombras. Acaba por ser mais eficiente interactivamente que a representação 3d.

Data Space vs Representation Space – O mesmo espaço de dado pode ser convertido em espaços de representação diferentes.

Data Mountain ---- > É um sistema de visualização de páginas obtidas via internet onde o ecrã é distribuído pela dimensão das janelas.