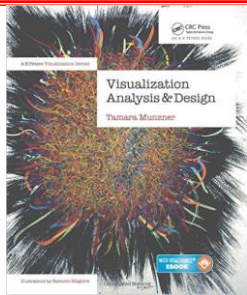


# Visualization Analysis & Design



José Remo Ferreira Brega  
remo.brega@unesp.br

12/03/2023

1

Visualization Analysis & Design	
Capítulo 1	O que é Visualização da Informação e por que fazê-la?
Capítulo 2	O que: Abstração de Dados
Capítulo 3	Por que: Abstração de Tarefas
Capítulo 4	Análise: Quatro Níveis de Validação
Capítulo 5	Marcas e Canais
Capítulo 6	Regras de Ouro
Capítulo 7	Organizar Tabelas
Capítulo 8	Organizar Dados Espaciais
Capítulo 9	Organizar Redes e Árvores
Capítulo 10	Cor do Mapa e Outros Canais
Capítulo 11	Manipulando a Visão
Capítulo 12	Facetando em Múltiplas Visões
Capítulo 13	Reduzir Itens e Atributos
Capítulo 14	Incorporar: Foco + Contexto
Capítulo 15	Estudos de caso de análise

2

2

# Visualization Analysis & Design

## O Que é Visualização da Informação e Por Que Fazê-la?

3

## Sumário

- Perspectiva Geral
- Por Que Um Ser Humano Está no Loop?
- Por Que Ter Um Computador no Loop?
- Por Que Usar Uma Representação Externa?
- Por Que Depender da Visão?
- Por Que Mostrar os Dados em Detalhes?
- Por Que Usar a Interatividade?
- Por Que o Espaço de Design do Idioma Vis é Enorme?
- Por Que Concentrar-se em Tarefas?
- Por Que Concentrar-se na Eficácia? (\*cont.\*)

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

4

4

## Sumário

- Por Que a Maioria dos Projetos é Ineficaz?
- Por Que a Validação é Difícil?
- Por Que Existem Limitações de Recursos?
- Por Que Analisar?

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

5

5

## Perspectiva Geral

- Sistemas de visualização baseados em computador fornecem representações visuais de conjuntos de dados projetados para ajudar as pessoas a realizar tarefas de forma mais eficaz.
- A visualização é adequada quando existe a necessidade de aumentar as capacidades humanas em vez de substituir as pessoas por métodos computacionais de tomada de decisão.
- O espaço de design de possíveis expressões vis é enorme, e inclui as considerações de como criar e como interagir com representações visuais.

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

6

6

## Perspectiva Geral

- Aspectos desta definição:
  - Por que um ser humano está no processo de tomada de decisão?
  - Por que um computador está no loop?
  - Por que usar uma representação externa?
  - Por que depender da visão?
  - Por que mostrar os dados em detalhe?
  - Por que usar a interatividade?
  - Por que o espaço de design visual é enorme?
  - Por que concentrar-se nas tarefas?
  - Por que a maioria dos projetos é ineficaz?
  - Por que concentrar-se com a eficácia?
  - Por que a validação é difícil?
  - Por que existem limitações de recursos?
  - Por que analisar vis?

7

## Por Que Um Ser Humano Está no Loop?

- A Vis permite que as pessoas analisem dados quando não sabem exatamente quais perguntas precisam fazer com antecedência.
- Alguns trabalhos que já foram feitos por seres humanos agora podem ser completamente automatizados com uma solução baseada em computador.
- Se uma solução totalmente automática tiver sido considerada aceitável, então não há necessidade de julgamento humano e, portanto, não há necessidade de você projetar uma ferramenta vis.

8

## Por Que Um Ser Humano Está no Loop?

- Muitos problemas de análise são mal especificados: as pessoas não sabem como abordar o problema.
  - Há muitas perguntas possíveis a fazer - em qualquer lugar de dezenas a milhares ou mais - e as pessoas não sabem qual dessas muitas perguntas são as corretas com antecedência.
- O melhor caminho a seguir é um processo de análise com um humano no loop, onde você pode explorar as poderosas propriedades de detecção de padrões do sistema visual humano em seu projeto.

9

## Por Que Um Ser Humano Está no Loop?

- Você pode projetar ferramentas vis para muitos tipos de usos.
  - Como um trampolim para obter uma compreensão mais clara dos requisitos de análise antes de desenvolver modelos matemáticos ou computacionais formais.
  - Para os designers de uma solução puramente computacional, para ajudá-los a refinar, depurar ou estender os algoritmos do sistema ou entender como os algoritmos são afetados por mudanças de parâmetros.
  - Para usuários finais em conjunto com outras decisões computacionais para iluminar se o sistema automático está fazendo a coisa certa de acordo com o julgamento humano.
  - Para análise exploratória em descobertas científicas.

10

## Por Que Um Ser Humano Está no Loop?

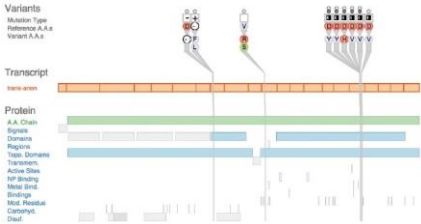


Figure 1.1. The Variant View vis tool supports biologists in assessing the impact of genetic variants by speeding up the exploratory analysis process. From [Ferstay et al. 13, Figure 1].

11

## Por Que Ter Um Computador no Loop?

- Pode-se criar ferramentas que permitem às pessoas explorar ou apresentar grandes conjuntos de dados que seriam completamente impossíveis de desenhar à mão, abrindo assim a possibilidade de ver como os conjuntos de dados mudam ao longo do tempo.
- O escopo do que as pessoas estão dispostas e capazes de fazer manualmente é fortemente limitado pela sua capacidade de atenção; Elas não são capazes de ir além de pequenos conjuntos de dados estáticos.

12

## Por Que Ter Um Computador no Loop?

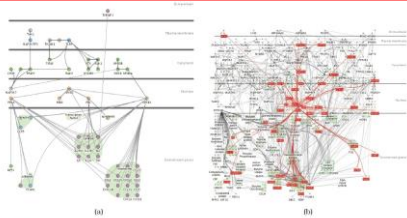


Figure 1.2. The Cerebral vis tool captures the style of hand-drawn diagrams in biology textbooks with vertical layers that correspond to places within a cell where interactions between genes occur. (a) A small network of 57 nodes and 74 edges might be possible to lay out by hand with enough patience. (b) Automatic layout handles this large network of 760 nodes and 1269 edges and provides a substrate for interactive exploration: the user has moved the mouse over the *MSX1* gene, so all of its immediate neighbors in the network are highlighted in red. From [Barsky et al. 07, Figures 1 and 5].

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

13

13

## Por Que Dependem da Visão?

- Baseia-se na exploração do sistema visual humano como um meio de comunicação.
- O sistema visual fornece um canal de largura de banda muito alta para nossos cérebros. Uma quantidade significativa de processamento de informação visual ocorre em paralelo no nível pré-consciente.
- Som é mal adaptado para fornecer visões gerais de grandes espaços de informação comparados com a visão.
- Os outros sentidos podem ser imediatamente descartados como canais de comunicação por causa das limitações tecnológicas.

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

15

15

## Por Que Mostrar os Dados em Detalhes?

Anscombe's Quartet: Raw Data									
	1	Y	2	Y	3	Y	4	Y	
100	8.04	100	9.14	103.1	7.46	8.0	6.58		
80	6.95	80	8.14	80	6.77	80	5.76		
130	7.58	130	8.74	130	12.74	80	7.71		
90	8.81	90	8.77	90	7.11	80	8.84		
110	8.33	110	9.26	110	7.81	80	8.47		
140	9.96	140	8.10	140	8.84	80	7.04		
60	7.24	60	6.13	60	6.08	80	5.25		
40	4.26	40	3.10	40	5.39	190	12.50		
120	10.84	120	9.13	120	8.15	80	5.56		
70	4.82	70	7.26	70	6.42	80	7.91		
50	5.68	50	4.74	50	5.73	80	6.89		
90	7.75	90	7.25	90	7.25	90	7.57		
Mean	90	7.75	90	7.25	90	7.25	90	7.57	
Variance	100	3.75	100	3.75	100	3.75	100	3.75	
Correlation	0.816		0.816		0.816		0.816		

Figure 1.3. Anscombe's Quartet is four datasets with identical simple statistical properties: mean, variance, correlation, and linear regression line. However, visual inspection immediately shows how their structures are quite different. After [Anscombe 73, Figures 1-4].

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

17

17

## Por Que Usar Uma Representação Externa?

- Representações externas aumentam a capacidade humana, permitindo-nos superar as limitações de nossa própria cognição e memória internas.
- Diagramas podem ser projetados para apoiar inferências perceptivas, que são muito fáceis para os seres humanos fazer.
- As vantagens dos diagramas como memória externa é que as informações podem ser organizadas por localização espacial, oferecendo a possibilidade de acelerar tanto a busca quanto o reconhecimento.

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

14

14

## Por Que Mostrar os Dados em Detalhes?

- As ferramentas Vis ajudam as pessoas em situações em que ver a estrutura do conjunto de dados em detalhes é melhor do que ver apenas um breve resumo do mesmo.
- Uma destas situações ocorre quando se exploram os dados para encontrar padrões, tanto para confirmar os esperados como para encontrar os inesperados.
- Outra ocorre quando se avalia a validade de um modelo estatístico, para julgar se o modelo de fato ajusta os dados.

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

16

16

## Por Que Usar a Interatividade?

- A interatividade é crucial para a construção de ferramentas vis que lidam com a complexidade.
- Quando conjuntos de dados são grandes o suficiente, as limitações de ambas as pessoas e exibe impedem apenas mostrar tudo de uma só vez.
- A interação é crucial:
  - Pode investigar em vários níveis de detalhe, desde uma visão de alto nível até vários níveis de resumo até uma visão detalhada de uma pequena parte dela.
  - Ele também pode apresentar diferentes formas de representar e resumir os dados de uma forma que suporte a compreensão das conexões entre essas alternativas.

Visualization Analysis & Design

O Que é Visualização da Informação e Por Que Fazê-la?

18

18

## Por Que o Espaço de Design do Idioma Vis é Enorme?

- Um idioma vis é uma abordagem distinta para criar e manipular representações visuais.
- Há muitas maneiras de criar uma codificação visual de dados como uma única imagem.
- O espaço de design de possibilidades fica ainda maior quando você considera como manipular uma ou mais dessas imagens com interação.
- As expressões estáticas simples incluem muitos tipos de gráficos (de dispersão, de barras e de linhas).
- Um idioma mais complicado pode unir múltiplos gráficos simples através da interação.

19

## Por Que o Espaço de Design do Idioma Vis é Enorme?

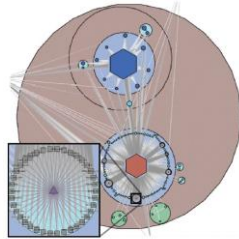


Figure 1.4. The Grouse vis tool features a complex idiom that combines visual encoding and interaction, supporting incremental layout of a network through interactive navigation. From [Archambault et al. 07a, Figure 5].

20

## Por Que Concentrar-se em Tarefas?

- Uma ferramenta que serve bem para uma tarefa pode ser mal adaptada para outra, exatamente para o mesmo conjunto de dados.
- A tarefa dos usuários é uma restrição igualmente importante para um designer vis como o tipo de dados que os usuários têm.
- Uma ferramenta vis pode suportar apresentação, ou descoberta, ou satisfação de informação; Ele também pode suportar a produção de mais informações para uso posterior.

21

## Por Que Concentrar-se na Eficácia?

- O foco na eficácia é um corolário da definição de vis para ter o objetivo de apoiar as tarefas do usuário.
- Este objetivo leva a preocupações sobre a correção, precisão e verdade desempenhando um papel muito central em vis.
- Nenhuma imagem pode comunicar a verdade, toda a verdade, e nada além da verdade.
- As preocupações de correção de um vis designer são complicadas pelo fato de que qualquer representação de dados é uma abstração onde as escolhas são feitas sobre quais os aspectos a enfatizar.

22

## Por Que a Maioria dos Projetos é Ineficaz?

- A razão mais fundamental que vis design é uma iniciativa difícil é que a grande maioria das possibilidades no espaço de design será ineficaz para qualquer contexto de uso específico.
- Em alguns casos, um possível design é uma combinação pobre com as propriedades dos sistemas perceptivos e cognitivos humanos.
- Em outros casos, o projeto seria compreensível por um ser humano em algum outro ambiente, mas é uma combinação ruim com a tarefa pretendida.

23

## Por Que a Maioria dos Projetos é Ineficaz?

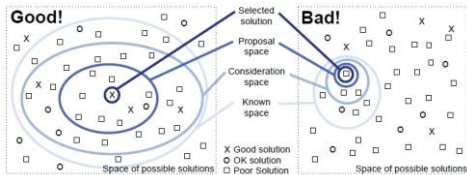


Figure 1.5. A search space metaphor for vis design.

24

## Por Que a Validação é Difícil?

- ❑ Como você sabe se funciona?
- ❑ Como você argumenta que um design é melhor ou pior do que outro para os usuários pretendidos?
- ❑ Por um lado, o que significa melhor?
- ❑ Os usuários fazem algo mais rápido?
- ❑ Eles se divertem mais fazendo isso?
- ❑ Podem trabalhar mais eficazmente?
- ❑ O que significa efetivamente?
- ❑ Como você mede insight ou compromisso?
- ❑ Qual é o design melhor do que?

25

## Por Que a Validação é Difícil?

- ❑ É melhor do que outro sistema vis?
- ❑ É melhor do que fazer as mesmas coisas manualmente, sem suporte visual?
- ❑ É melhor do que fazer as mesmas coisas completamente automaticamente?
- ❑ E que tipo de coisa faz melhor? Ou seja, como você decide que tipo de tarefa os usuários devem fazer ao testar o sistema? E quem é esse usuário? Um especialista que fez esta tarefa durante décadas, ou um iniciante que precisa da tarefa a ser explicado antes de começarem?

26

## Por Que a Validação é Difícil?

- ❑ Eles estão familiarizados com a forma como o sistema funciona com o uso por muito tempo, ou eles estão vendo-o pela primeira vez?
- ❑ Como você decide que tipo de dados de referência você deve usar ao testar o sistema?
- ❑ Você pode caracterizar quais classes de dados o sistema é adequado para?
- ❑ Como medir a qualidade de uma imagem vis?
- ❑ Quão bem alguma das métricas quantitativas de qualidade automaticamente calculadas corresponde a julgamentos humanos?

27

## Por Que Existem Limitações de Recursos?

- ❑ Ao projetar ou analisar um sistema vis, deve-se considerar pelo menos 3 tipos diferentes de limitações:
  - ❑ Capacidade computacional;
  - ❑ Capacidade perceptual e cognitiva humana; e
  - ❑ Capacidade de exibição.
- ❑ A escalabilidade é uma preocupação central.
- ❑ O aumento no tamanho do conjunto de dados ocorre:
  - ❑ Melhorias na capacidade do computador;
  - ❑ Melhorias na aquisição de dados e tecnologia de sensores; e
  - ❑ Crescente alcance da infra-estrutura computacional em todos os aspectos da vida.

28

## Por Que Existem Limitações de Recursos?

- ❑ O tempo e a memória do computador são recursos limitados, e muitas vezes há restrições suaves e duras na disponibilidade desses recursos.
- ❑ Em alguns cenários, os usuários não estão dispostos ou não podem esperar para o sistema pré-processar os dados antes que eles possam interagir com ele.
- ❑ O tamanho do conjunto de dados pode ultrapassar essa capacidade de memória do computador.
- ❑ Assim, a complexidade computacional dos algoritmos para pré-processamento de dados, transformação, layout e renderização é uma grande preocupação.

29

## Por Que Existem Limitações de Recursos?

- ❑ No lado humano, a memória e a atenção são recursos finitos.
- ❑ Existem poderes e limitações dos mecanismos preatentivos visuais de baixo nível que realizam o processamento massivamente paralelo do nosso campo visual atual.
- ❑ A memória humana para coisas que não são diretamente visíveis é notoriamente limitada.
- ❑ Nós armazenamos surpreendentemente pouca informação internamente na memória de trabalho visual, deixando-nos vulneráveis à cegueira da mudança.

30

## Por Que Existem Limitações de Recursos?

- ❑ A capacidade de exibição é um terceiro tipo de limitação a considerar.
- ❑ Vis designers muitas vezes ficam sem pixels; Ou seja, a resolução da tela não é suficiente para mostrar todas as informações desejadas simultaneamente.
- ❑ Há um trade-off entre os benefícios de mostrar tanto quanto possível de uma só vez, para minimizar a necessidade de navegação e exploração, e os custos de mostrar muito de uma vez, onde o usuário é oprimido pela desordem visual.

## Por Que Existem Limitações de Recursos?

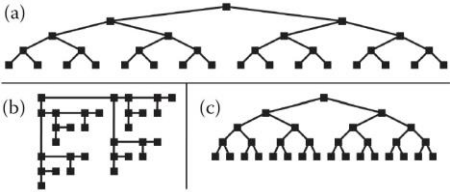


Figure 1.6. Low and high information density visual encodings of the same small tree dataset; nodes are the same size in each. (a) Low information density. (b) Higher information density, but depth in tree cannot be read from spatial position. (c) High information density, while maintaining property that depth is encoded with position. From [McGuffin and Robert 10, Figure 3].

## Por Que Analisar?

- ❑ A análise de sistemas existentes é um bom trampolim para projetar novos.
- ❑ Quando você é confrontado com um problema de vis como um designer, pode ser difícil decidir o que fazer.
- ❑ Muitas linguagens e ferramentas visuais baseadas em computador foram criadas nas últimas décadas e, considerando-as uma a uma, você se defronta com uma grande coleção de diferentes possibilidades.
- ❑ Mesmo com um candidato provável, você pode precisar escavar ainda mais na literatura para entender se há alguma evidência de que a ferramenta foi um sucesso.

## Por Que Analisar?



Figure 1.7. Three-part analysis framework for a vis instance: *why* is the task being performed, *what* data is shown in the views, and *how* is the vis idiom constructed in terms of design choices.

## Por Que Analisar?

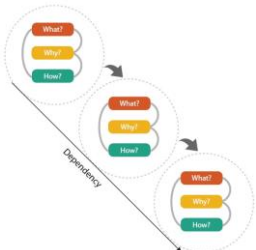


Figure 1.8. Analyzing vis usage as chained sequences of instances, where the output of one instance is the input to another.

## Visualization Analysis & Design

### O Que: Abstração de Dados

## Sumário

- Perspectiva Geral
- Por Que a Semântica de Dados e os Tipos São Importantes?
- Tipos de Dados
- Tipos de Conjuntos de Dados
- Tipos de Atributos
- Semântica

37

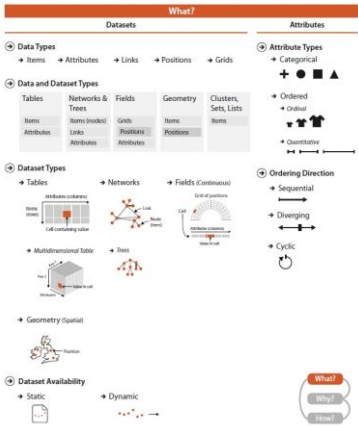


Figure 2.1. What can be visualized: data, datasets, and attributes.

38

38

## Perspectiva Geral

- Os quatro tipos básicos de conjuntos de dados são:
  - Tabelas, Redes, Campos e Geometria;
- Outras possíveis coleções de itens incluem:
  - Clusters, Conjuntos e Listas.
- Esses conjuntos de dados são compostos de diferentes combinações dos cinco tipos de dados:
  - Itens, Atributos, Links, Posições e Grades.
- O conjunto de dados completo:
  - Pode estar disponível imediatamente na forma de um arquivo estático; e
  - Podem ser dados dinâmicos.

39

## Perspectiva Geral

- O tipo de um atributo pode ser
  - Categórico ou
  - Ordenado, com uma divisão adicional:
    - Ordinal e
    - Quantitativa.
- A direção de ordenação dos atributos pode ser:
  - Sequencial,
  - Divergente ou
  - Cíclica.

40

## Por Que a Semântica de Dados e os Tipos São Importantes?

- Muitos aspectos do projeto vis são orientados pelo tipo de dados que você tem à sua disposição.
  - 14, 2.6, 30, 30, 15, 100001
  - Basil, 7, S, Pera
- Para ir além das suposições, é necessário conhecer duas partes de informações transversais dos termos: sua semântica e seus tipos.
- A semântica dos dados é o seu significado no mundo real.
  - Por exemplo, uma palavra representa um nome próprio humano, ou é a versão encurtada de um nome de companhia onde o nome cheio pode ser olhado acima em uma lista externa, ou é uma cidade, ou é um fruto?
- O tipo de dados é a sua interpretação estrutural ou matemática.
  - No nível de dados, que tipo de coisa é: um item, um link, um atributo? No nível do conjunto de dados, como esses tipos de dados são combinados em uma estrutura maior?

41

## Por Que a Semântica de Dados e os Tipos São Importantes?

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Table 2.1. A full table with column titles that provide the intended semantics of the attributes.

42



## Tipos de Dados

➔ Data Types

➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

Figure 2.2. The five basic data types: items, attributes, links, positions, and grids.

- Um **Item** é uma entidade individual que é discreta, como uma linha em uma tabela simples ou um nó em uma rede.
  - Por ex., os itens podem ser pessoas, estoques, cafeterias, genes ou cidades.
- Um **Atributo** é uma propriedade específica que pode ser medida, observada ou registrada.
  - Por ex., os atributos podem ser salário, preço, número de vendas, níveis de expressão proteica ou temperatura.

Visualization Analysis & Design

O Que: Abstração de Dados

43

43

## Tipos de Dados

- Um **Link** é uma relação entre itens, normalmente dentro de uma rede.
- Uma **Posição** é um dado espacial, proporcionando uma localização em espaço bidimensional (2D) ou tridimensional (3D).
  - Por ex., uma posição pode ser um par latitude-longitude descrevendo um local na superfície da Terra ou três números especificando um local dentro da região do espaço medido por um scanner médico.
- Uma **Grade** especifica a estratégia para amostrar dados contínuos em termos de relações geométricas e topológicas entre suas células.

Visualization Analysis & Design

O Que: Abstração de Dados

44

44

## Tipos de Conjuntos de Dados

- Um conjunto de dados é qualquer coleção de informações que é o alvo da análise.
- Os quatro tipos básicos de conjuntos de dados são:
  - Tabelas,
  - Redes,
  - Campos e
  - Geometria.
- Outras formas de agrupar itens juntos incluem:
  - Clusters,
  - Conjuntos e
  - Listas.
- Em situações do mundo real, combinações complexas desses tipos básicos são comuns.

Visualization Analysis & Design

O Que: Abstração de Dados

45

45

## Tipos de Conjuntos de Dados

➔ Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Figure 2.3. The four basic dataset types are tables, networks, fields, and geometry; other possible collections of items are clusters, sets, and lists. These datasets are made up of five core data types: items, attributes, links, positions, and grids.

Visualization Analysis & Design

O Que: Abstração de Dados

46

46

## Tipos de Conjuntos de Dados

➔ Dataset Types

➔ Tables ➔ Networks ➔ Fields (Continuous) ➔ Geometry (Spatial)

Attributes (columns)

Items (rows)

Cell containing value

Link

Node (item)

Grid of positions

Cell

Attributes (columns)

Values in cell

Position

➔ Multidimensional Table

➔ Trees

Key 1

Key 2

Value in cell

Attributes

Figure 2.4. The detailed structure of the four basic dataset types.

Visualization Analysis & Design

O Que: Abstração de Dados

47

47

## Tabelas

- Muitos conjuntos de dados vêm na forma de tabelas que são feitas de linhas e colunas, uma forma familiar para qualquer pessoa que tenha usado uma planilha.
- Para uma tabela plana simples, os termos usados são:
  - que cada linha representa um item de dados e
  - cada coluna é um atributo do conjunto de dados.
- Cada célula na tabela é totalmente especificada pela combinação de uma linha e uma coluna - um item e um atributo - e contém um valor para esse par.

Visualization Analysis & Design

O Que: Abstração de Dados

48

48



## Tabelas

A	B	C	D	E	F	G
Order ID	Order Date	Order Priority	Product	Container	Product Base Weight	Ship Date
5	10/14/08	3-Low	Large Box		0.8	10/21/08
6	2/12/08	4-Hot Specified	Small Pack		0.55	2/12/08
32	7/16/07	2-High	Small Pack		0.79	7/17/07
32	7/16/07	2-High	Large Box			7/17/07
32	7/16/07	2-High	Medium Box			7/18/07
35	10/23/07	4-Hot Specified	Wrap Bag		0.52	10/24/07
35	10/23/07	4-Hot Specified	Small Box		0.58	10/25/07
36	11/3/07	1-Urgent	Small Box		0.55	11/5/07
45	3/24/07	1-Urgent	Small Pack		0.43	3/25/07
68	4/24/08	3-Low	Wrap Bag		0.50	4/25/08
69	4/24/08	3-Low	Small Pack		0.44	4/25/08
69	4/24/08	4-Hot Specified	Wrap Bag		0.6	4/26/08
70	12/18/08	3-Low	Small Box		0.59	12/18/08
70	12/18/08	3-Low	Wrap Bag		0.82	12/18/08
96	4/17/05	2-High	Small Box		0.52	4/18/05
97	1/28/08	3-Medium	Small Box		0.58	1/30/08
109	11/18/08	3-Low	Small Box		0.57	11/18/08
130	5/8/08	2-High	Small Box		0.57	5/8/08
130	5/8/08	2-High	Medium Box		0.38	5/9/08
132	6/11/08	3-Medium	Small Box		0.6	6/11/08
132	6/11/08	3-Medium	Large Box		0.69	6/14/08
134	5/11/08	4-Hot Specified	Small Box		0.82	5/12/08
135	10/21/07	4-Hot Specified	Small Pack		0.84	10/23/07
166	8/12/07	2-High	Small Box		0.55	8/14/07
193	8/6/08	1-Urgent	Medium Box		0.57	8/10/08
194	4/23/08	3-Medium	Wrap Bag		0.42	4/27/08

Figure 2.5. In a simple table of orders, a row represents an item, a column represents an attribute, and their intersection is the cell containing the value for that pairwise combination.

## Redes e Árvores

- O tipo de conjunto de dados de redes é bem adequado para especificar que existe algum tipo de relação entre dois ou mais itens.
- Um item em uma rede é muitas vezes chamado de nó. Um link é uma relação entre dois itens.
- Exemplos:
  - Rede Social: os nós são pessoas e as ligações são amizades.
  - Rede de Computadores: nós são computadores e links capacidade de enviar mensagens.
- Os nós de rede podem ter atributos associados, assim como itens em uma tabela. Além disso, os próprios links também poderiam ser considerados como tendo atributos associados a eles; Estes podem ser parciais ou totalmente disjuntos dos atributos do nó.

## Árvores

- Redes com estrutura hierárquica são mais especificamente chamadas árvores.
- Em contraste com uma rede geral, as árvores não têm ciclos: cada nó filho tem apenas um nó pai apontando para ele.
- Um exemplo de uma árvore é o organograma de uma empresa, mostrando quem relata para quem;
- Outro exemplo é uma árvore mostrando as relações evolutivas entre as espécies na árvore biológica da vida, onde os nós filhos de humanos e macacos compartilham o mesmo nó pai de primatas.

## Campos

- O tipo de conjunto de dados de campo também contém valores de atributo associados a células.
- Cada célula em um campo contém medições ou cálculos de um domínio contínuo: existem valores conceitualmente infinitos que você pode medir, então você sempre poderia tomar uma nova medição entre quaisquer dois existentes.
- Os fenômenos contínuos que podem ser medidos no mundo físico ou simulados em software incluem temperatura, pressão, velocidade, força e densidade; Funções matemáticas também podem ser contínuas.
- Os dados contínuos exigem um tratamento cuidadoso para levar em conta as questões matemáticas da amostragem, a frequência das medições e a interpolação, de uma forma a não induzir em erro.

## Campos Espaciais

- Os dados contínuos são frequentemente encontrados na forma de um campo espacial, onde a estrutura celular do campo é baseada em amostragem em posições espaciais.
- A maioria dos conjuntos de dados que contêm dados inerentemente espaciais ocorrem no contexto de tarefas que requerem aspectos de compreensão de sua estrutura espacial, especialmente forma.
- O subcampo de visualização científica, ou scivis (abreviado), refere-se a situações em que a posição espacial é dada com o conjunto de dados.
- O subcampo de visualização de informação, ou infovis (abreviado), diz respeito a situações em que o uso do espaço em uma codificação visual é escolhido pelo designer.

## Tipos de Grade

- Quando um campo contém dados criados por amostragem em intervalos, as células formam uma grade uniforme. Não há necessidade de armazenar explicitamente a geometria de grade em termos de sua localização no espaço ou a topologia de grade em termos de como cada célula se conecta com suas células vizinhas.
- Uma grade retilínea suporta amostragem não uniforme, permitindo o armazenamento eficiente de informações que têm alta complexidade em algumas áreas e baixa complexidade em outras.
- Uma grade estruturada permite formas curvilíneas, onde a localização geométrica de cada célula precisa ser especificada.
- As informações topológicas sobre como as células se conectam entre si devem ser armazenadas explicitamente além de suas posições espaciais.

## Geometria

- ❑ O tipo de conjunto de dados de geometria especifica informações sobre a forma de itens com posições espaciais explícitas. Os itens podem ser pontos, linhas ou curvas unidimensionais, superfícies ou regiões 2D ou volumes 3D.
- ❑ Os conjuntos de dados de geometria são intrinsecamente espaciais e, como os campos espaciais, ocorrem tipicamente no contexto de tarefas que requerem compreensão da forma.
- ❑ Conjuntos de dados de geometria não têm necessariamente atributos, em contraste com os outros três tipos de conjunto de dados básicos. Muitas das questões de design em vis pertencem a perguntas sobre como codificar atributos.

55

## Outras Combinações

- ❑ Existem várias maneiras de agrupar vários itens:
  - ❑ Um conjunto é simplesmente um grupo não ordenado de itens.
  - ❑ Um grupo de itens com um pedido especificado pode ser chamado de lista.
  - ❑ Um cluster é um agrupamento baseado na similaridade de atributo, em que os itens dentro de um cluster são mais semelhantes entre si do que com os de outro cluster.
- ❑ Estruturas mais complexas construídas em cima do tipo de rede básica.
  - ❑ Um caminho através de uma rede é um conjunto ordenado de segmentos formados por links que conectam nós.
  - ❑ Uma rede composta é uma rede com uma árvore associada: todos os nós na rede são as folhas da árvore e os nós internos da árvore fornecem uma estrutura hierárquica para os nós que é diferente das ligações de rede entre eles.

56

## Disponibilidade de Dataset

- ❑ A abordagem padrão para vis assume que todo o conjunto de dados está disponível de uma só vez, como um arquivo estático. No entanto, alguns conjuntos de dados são, em vez disso, fluxos dinâmicos, onde a informação do conjunto de dados escorre no decorrer da sessão vis.

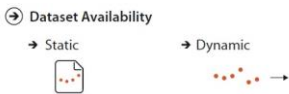


Figure 2.6. Dataset availability can be either static or dynamic, for any dataset type.

57

## Tipos de Atributos

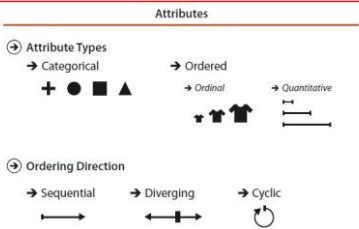


Figure 2.7. Attribute types are categorical, ordinal, or quantitative. The direction of attribute ordering can be sequential, diverging, or cyclic.

58

## Categórico

- ❑ O tipo de dados categóricos, como frutas ou nomes favoritos, não tem uma ordem implícita, mas muitas vezes tem estrutura hierárquica.
- ❑ Categorias só podem distinguir se duas coisas são as mesmas (maçãs) ou diferentes (maçãs versus laranjas).
- ❑ Naturalmente, qualquer ordenação arbitrária externa pode ser imposta sobre dados categóricos. Os frutos podiam ser ordenados alfabeticamente de acordo com o seu nome, ou pelo seu preço - mas apenas se essas informações auxiliares estivessem disponíveis.

59

## Ordenados

- ❑ **Ordinal e Quantitativo**
  - ❑ Todos os dados ordenados têm uma ordem implícita, ao contrário dos dados categóricos desordenados.
  - ❑ Esse tipo pode ser subdividido.
    - ❑ Com dados ordinais, como o tamanho da camisa, não podemos fazer aritmética completa, mas há uma ordem bem definida. Por exemplo, grande menos média não é um conceito significativo, mas sabemos que média cai entre pequenas e grandes.
    - ❑ Um subconjunto de dados ordenados são dados quantitativos, ou seja, uma medida de magnitude que suporta comparação aritmética. Por ex., a quantidade de 68 polegadas menos 42 polegadas é um conceito significativo, e a resposta de 26 polegadas pode ser calculada.

60

## Ordenados

- Sequencial versus Divergente
  - Os dados ordenados podem ser sequenciais, onde há uma faixa homogênea de um valor mínimo a um valor máximo, ou divergente, que pode ser desconstruída em duas sequências apontando em direções opostas que se encontram em um ponto zero comum.
  - Por ex. um conjunto de dados de altura de montanha é sequencial, quando medido de um ponto mínimo do nível do mar até um ponto máximo do Monte Everest.

61

## Ordenados

- Cíclico
  - Dados ordenados podem ser cíclicos, onde os valores envolvem volta para um ponto de partida em vez de continuar a aumentar indefinidamente.
  - Muitos tipos de medidas de tempo são cíclicos, incluindo a hora do dia, o dia da semana e o mês do ano.

62

## Atributos Hierárquicos

- Pode haver estrutura hierárquica dentro de um atributo ou entre vários atributos.
- Os preços diários de ações das empresas coletadas ao longo de uma década é um exemplo de um conjunto de dados de séries temporais, onde um dos atributos é o tempo.
- Pode haver padrões interessantes em escalas temporais múltiplas, como variações semanais muito fortes para o dia de semana versus fim de semana, ou padrões sutis mais anuais que mostram variações sazonais no verão versus inverno.

63

## Semântica

- Conhecer o tipo de um atributo não nos diz sobre sua semântica, porque essas duas questões são transversais: uma não dita a outra.
- A classificação aqui é fortemente focada na semântica de chaves versus valores, e as questões relacionadas de dados espaciais e contínuos versus dados não-espaciais e discretos, para coincidir com a estrutura de análise de escolha de design de idioma. Uma consideração adicional é se um atributo é temporal.

64

## Semântica Chave Versus Valor

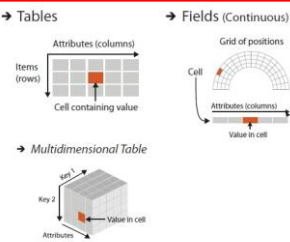


Figure 2.8. Key and value semantics for tables and fields.

65

## Semântica Chave Versus Valor

- Tabelas planas
  - Uma tabela plana simples tem apenas uma chave, onde cada item corresponde a uma linha na tabela e qualquer número de atributos de valor.
  - Neste caso, a chave pode estar completamente implícita, onde é simplesmente o índice da linha. Ele pode ser explícito, onde ele está contido na tabela como um atributo. Neste caso, não deve haver quaisquer valores duplicados dentro desse atributo.
  - Em tabelas, as chaves podem ser categóricas ou atributos ordinais, mas os atributos quantitativos são inadequados como chaves por permitirem a repetição.

66

## Semântica Chave Versus Valor

A	B	C	D	E	F	G
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date	
1	1/21/08	1-Low	Large Box	0.5	1/21/08	
2	2/21/08	4-High Specified	Small Pack	0.55	2/22/08	
3	2/14/07	2-High	Small Pack	0.79	2/17/07	
32	2/14/07	2-High	Junior Box	0.72	2/17/07	
33	2/14/07	2-High	Medium Box	0.6	2/18/07	
34	2/14/07	2-High	Medium Box	0.63	2/18/07	
35	10/23/07	4-High Specified	Small Box	0.52	10/24/07	
36	10/23/07	4-High Specified	Small Box	0.58	10/23/07	
43	11/2/07	1-Urgent	Small Pack	0.55	11/2/07	
45	3/18/07	1-Urgent	Small Pack	0.49	3/18/07	
46	1/26/08	5-Low	Wrap Box	0.56	1/26/08	
49	6/4/05	4-High Specified	Small Pack	0.44	6/4/05	
49	6/4/05	4-High Specified	Small Pack	0.5	6/4/05	
70	12/18/04	5-Low		0.58	12/23/04	
70	12/18/04	5-Low		0.82	12/23/04	
86	4/17/05	2-High		0.55	4/18/05	
87	1/28/04	3-Medium		0.38	1/28/04	
126	11/18/04	5-Low	Small Box	0.37	11/18/04	
126	5/8/08	2-High	Small Box	0.37	5/8/08	
130	5/8/08	2-High	Medium Box	0.38	5/11/08	
132	4/11/04	3-Medium	Small Box	0.6	4/12/04	
132	4/11/04	3-Medium	Medium Box	0.6	4/12/04	
132	6/11/04	4-High Specified	Junior Box	0.63	6/14/04	
131	10/21/07	4-High Specified	Large Box	0.64	10/23/07	
146	8/12/07	2-High	Small Box	0.53	8/14/07	
193	8/8/04	1-Urgent	Medium Box	0.57	8/10/04	
194	4/7/08	3-Medium	Wrap Box	0.42	4/7/08	

Figure 2.9. The order table with the attribute columns colored by their type; none of them is a key.

## Semântica Chave Versus Valor

- ❑ Tabelas multidimensionais
  - ❑ O caso mais complexo é uma tabela multidimensional, onde são necessárias várias teclas para procurar um item.
  - ❑ A combinação de todas as chaves deve ser exclusiva para cada item, mesmo que um atributo de chave individual possa conter duplicatas.
  - ❑ Por ex. uma tabela multidimensional comum a partir do domínio da biologia tem um gene como uma chave e tempo como outra chave, de modo que o valor em cada célula é o nível de atividade de um gene em um determinado momento.

## Semântica Chave Versus Valor

- ❑ Campos
  - ❑ Embora os campos diferem das tabelas de maneira fundamental porque representam dados contínuos e não discretos, as chaves e os valores ainda são preocupações centrais.
  - ❑ O vocabulário diferente para a mesma idéia básica é mais comum com dados de campo espacial, onde o termo variável independente é usado em vez de chave e variável dependente em vez de valor.
  - ❑ Os campos são estruturados por amostragem de forma sistemática, onde cada célula da grade é abrangida por um intervalo único de um domínio contínuo.

## Semântica Chave Versus Valor

- ❑ Campos escalares
  - ❑ Um campo escalar é univariável, com um único atributo de valor em cada ponto no espaço.
  - ❑ Um exemplo de um campo escalar 3D é a varredura médica variando no tempo; Outro é a temperatura em uma sala em cada ponto no espaço 3D. A intuição geométrica é que cada ponto em um campo escalar tem um único valor.
  - ❑ Um ponto no espaço pode ter vários números diferentes associados a ele; Se não houver nenhuma conexão subjacente entre eles, então eles são simplesmente múltiplos campos escalares separados.

## Semântica Chave Versus Valor

- ❑ Campos vetoriais
  - ❑ Um campo vetorial é multivariável, com uma lista de vários valores de atributos em cada ponto.
  - ❑ A intuição geométrica é que cada ponto em um campo de vetor tem uma direção e magnitude, como uma seta que pode apontar em qualquer direção e que pode ser qualquer comprimento.
  - ❑ O comprimento pode significar a velocidade de um movimento ou a força de uma força.
  - ❑ Um exemplo de um campo de vetor 3D é a velocidade do ar na sala em um ponto de tempo específico, onde há uma direção e velocidade para cada item.

## Semântica Chave Versus Valor

- ❑ Campos tensoriais
  - ❑ Um campo tensor tem uma matriz de atributos em cada ponto, representando uma estrutura matemática multivariada mais complexa do que a lista de números em um vetor.
  - ❑ Um exemplo físico é a tensão, que no caso de um campo 3D pode ser definido por nove números que representam forças atuando em três direções ortogonais.

## Semântica Chave Versus Valor

- ❑ **Semântica de campo**
  - ❑ Esta categorização de campos espaciais requer conhecimento da semântica de atributos e não pode ser determinada a partir de informações de tipo sozinho.
  - ❑ Se você receber um campo com vários valores medidos em cada ponto e nenhuma outra informação, não há maneira certa de conhecer sua estrutura.
  - ❑ Por ex., nove valores poderiam representar muitas coisas: nove campos escalares separados, ou uma mistura de vários campos vetoriais e campos escalares, ou um único campo tensor.

73

## Semântica Temporal

- ❑ Um atributo temporal é simplesmente qualquer tipo de informação que se relaciona com o tempo.
- ❑ Dados sobre o tempo são complicados de manipular por causa da estrutura hierárquica rica que usamos para raciocinar sobre o tempo e o potencial para a estrutura periódica.
- ❑ A hierarquia do tempo é profundamente multiescala: a escala de interesse pode variar de nanossegundos a horas a décadas a milênios.
- ❑ As palavras comuns tempo e data são uma maneira de especificar parcialmente a escala de interesse temporal.

74

## Semântica Temporal

- ❑ As tarefas de análise temporal envolvem frequentemente encontrar ou verificar a periodicidade, seja em uma escala predeterminada ou em alguma escala não conhecida antecipadamente.
- ❑ Além disso, as escalas temporais de interesse nem sempre se encaixam em uma hierarquia rígida.
- ❑ Os atributos temporais podem ter valor ou semântica chave. Exemplos de atributos temporais com semântica de valor dependente são a duração do tempo decorrido ou a data em que uma transação ocorreu.

75

## Semântica Temporal

- ❑ **Dados variáveis no tempo**
  - ❑ Um conjunto de dados tem semântica variável no tempo quando o tempo é um dos atributos-chave, ao contrário de quando o atributo temporal é um valor e não uma chave. Como com outras decisões sobre semântica, a questão de saber se o tempo tem chave ou valor semântica requer conhecimento externo sobre a natureza do conjunto de dados.
  - ❑ Um exemplo de um conjunto de dados com semântica variável no tempo é aquele criado com uma rede de sensores que rastreia a localização de cada animal dentro de um rebanho tomando medidas por segundo.

(\*cont.\*)

76

## Semântica Temporal

- ❑ A palavra dinâmica é frequentemente usada de forma ambígua para significar uma de duas coisas muito diferentes.
- ❑ Alguns usam isso para significar que um conjunto de dados tem semântica de tempo, em contraste com um conjunto de dados onde o tempo não é um atributo-chave, como discutido aqui.
- ❑ Outros usá-lo para significar um conjunto de dados tem tipo de fluxo, em contraste com um arquivo imutável que pode ser carregado de uma só vez.

77

## Visualization Analysis & Design

### Porque: Abstração de Tarefas

78

## Sumário

- Perspectiva Geral
- Por Que Analisar Tarefas de Forma Abstrata?
- Quem: Designer ou Usuário
- Ações
- Alvos
- Como: Uma Preview
- Analisando e Derivando: Exemplos

79

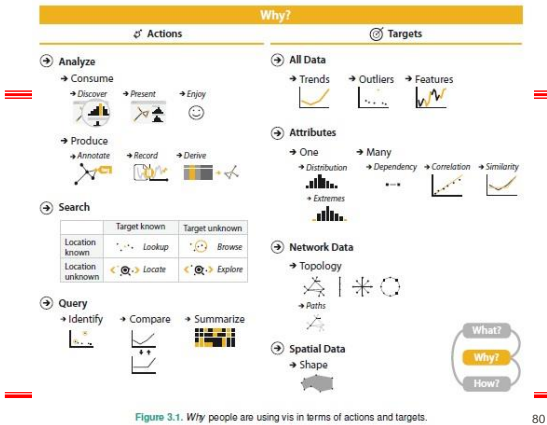


Figure 3.1. Why people are using vis in terms of actions and targets.

80

80

## Perspectiva Geral

- No nível mais alto, consumir ou produzir informações:
  - Os casos de consumo são apresentar, descobrir e desfrutar; Descoberta pode envolver gerar ou verificar uma hipótese.
- No nível médio, a pesquisa pode ser classificada de acordo com se a identidade e a localização dos alvos são conhecidos ou não:
  - Ambos são conhecidos com lookup, o alvo é conhecido mas sua localização não, o local é conhecido mas o alvo não é para navegar, e nem o destino nem a localização são conhecidos.
- No nível baixo, as consultas podem ter três escopos:
  - Identificar um destino, comparar alguns destinos e resumir todos os destinos.

81

## Por Que Analisar Tarefas de Forma Abstrata?

- Transformar descrições de tarefas de linguagem específica de domínio em forma abstrata permite raciocinar sobre semelhanças e diferenças entre elas.
- Há um monte de semelhanças no que as pessoas querem fazer uma vez que você tira as diferenças superficiais de linguagem.
- A estrutura de análise tem um pequeno conjunto de palavras cuidadosamente escolhidas para descrever por que as pessoas estão usando vis, projetada para ajudá-lo a distinguir decisivamente e concisamente entre diferentes objetivos.

82

## Por Que Analisar Tarefas de Forma Abstrata?

- Este conjunto tem verbos descrevendo ações e substantivos descrevendo alvos.
- É possível que você decida usar termos adicionais para descrever completa e precisamente os objetivos do usuário.
- A mesma ferramenta vis pode ser útil para muitos objetivos diferentes. Muitas vezes é útil considerar apenas um dos objetivos do usuário de cada vez, chegando a uma sequência encadeada de tarefas.
- Outra razão é como transformar os dados do usuário em diferentes formas, derivando novos dados.

83

## Quem: Designer ou Usuário

- As ferramentas Vis caem em algum lugar ao longo de um contínuo de específico para geral.
- No lado específico, as ferramentas são estreitas: o designer criou muitas escolhas no design da própria ferramenta de uma forma que o usuário não pode substituir.
- Do lado geral, as ferramentas são flexíveis e os usuários têm muitas opções para fazer. A amplitude das escolhas é tanto uma força quanto uma limitação: os usuários têm muito poder, mas também podem fazer escolhas ineficazes.

84

## Ações



Figure 3.2. Three levels of actions: analyze, search, and query.

85

## Analisar

- No nível mais alto, o framework distingue dois objetivos possíveis de pessoas que querem analisar dados usando uma ferramenta vis: os usuários podem querer apenas consumir informações existentes ou também produzir ativamente novas informações.
- O caso de uso mais comum para vis é para o usuário consumir informações que já foram geradas como dados armazenados em um formato passível de computação.
  - Descobrir
  - Apresentar
  - Apreciar

Visualization Analysis & Design

Porque: Abstração de Tarefas

86

## Analisar

- Descobrir**
  - A meta de descoberta refere-se ao uso de vis para encontrar novos conhecimentos que não eram conhecidos anteriormente.
  - A descoberta pode surgir da observação acidental de fenômenos inesperados, mas a investigação pode ser motivada por teorias, modelos, hipóteses ou palpites existentes.
  - Este uso inclui o objetivo de encontrar coisas completamente novas.
  - Ele também inclui o objetivo de descobrir se uma conjectura é verdadeira ou falsa.

Visualization Analysis & Design

Porque: Abstração de Tarefas

87

## Analisar

- Apresentar**
  - Refere-se ao uso de vis para a comunicação sucinta de informações, para contar uma história com dados ou orientar um público através de uma série de operações cognitivas.
  - A apresentação usando vis pode ocorrer dentro do contexto de tomada de decisão, planejamento, previsão e processos de instrução.
  - O ponto crucial sobre o objetivo atual é que vis está sendo usado por alguém para comunicar algo específico e já compreendido para um público.

Visualization Analysis & Design

Porque: Abstração de Tarefas

88

## Analisar

- Apreciar**
  - O objetivo de apreciar refere-se a encontros casuais com vis.
  - Nesses contextos, o usuário não é impulsionado por uma necessidade premente de verificar ou gerar uma hipótese, mas pela curiosidade que pode ser tanto estimulada e satisfeita pelo vis.
  - Encontros casuais com vis para o prazer pode ser fugaz, como quando se olha para um infográfico enquanto lê um post do blog.

Visualization Analysis & Design

Porque: Abstração de Tarefas

89

## Analisar

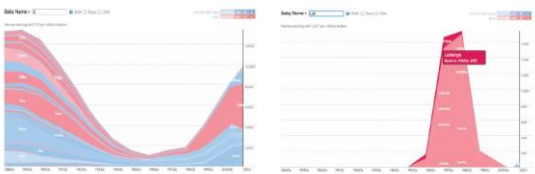


Figure 3.3. Name Voyager, a vis tool originally intended for parents focused deciding on what to name their expected baby, ended up being used by many nonparents to analyze historical trends for their own enjoyment. Left: Names starting with "O" had a notable dip in popularity in the middle of the century. Right: Names starting with "LAT" show a trend of the 1970s. After [Wattenberg 05, Figures 2 and 3], using <http://www.babynamewizard.com>.

Visualization Analysis & Design

Porque: Abstração de Tarefas

90



## Produzir

- Intenção do usuário é gerar novo material. Muitas vezes, o objetivo é produzir saída que é imediatamente usada, como entrada para uma próxima instância.
- Às vezes, o usuário pretende usar este novo material para alguma outra tarefa vis-related mais tarde, como descoberta ou apresentação.
- Às vezes, o uso pretendido do novo material é para alguma outra finalidade que não requer vis, como a análise a jusante usando ferramentas não-visuais.
- Existem três tipos de objetivos de produção: anotar, gravar e derivar.

91

## Produzir

- Anotar**
  - O objetivo de anotação refere-se à adição de anotações gráficas ou textuais associadas a um ou mais elementos de visualização preexistentes, tipicamente como uma ação manual pelo usuário.
  - Quando uma anotação está associada a itens de dados, a anotação pode ser considerada como um novo atributo para eles.
    - Por ex., o usuário pode anotar todos os pontos dentro de um cluster com um rótulo de texto.

92

## Produzir

- Registrar**
  - A meta de registrar salva ou captura os elementos de visualização como artefatos persistentes.
  - Esses artefatos incluem capturas de tela, listas de elementos ou locais marcados por livros, configurações de parâmetros, logs de interação ou anotações.
  - A opção de gravação salva um artefato persistente, em contraste com a anotação, que anexa informações temporariamente a elementos existentes; Uma anotação feita por um usuário pode ser posteriormente gravada.
  - Um exemplo interessante de um objetivo de registro é montar um histórico gráfico.

93

## Produzir



Figure 3.4. Graphical history recorded during an analysis session with Tableau. From [Heer et al. 08, Figure 1].

94

## Produzir

- Derivar**
  - A meta de derivar é produzir novos elementos de dados baseados em elementos de dados existentes.
  - Novos atributos podem ser derivados de informações contidas nos existentes, ou os dados podem ser transformados de um tipo para outro. A derivação de novos dados é uma parte crítica do processo de design vis.
  - O caso comum é que a obtenção de novos dados é uma escolha feita por designers vis, mas essa escolha também pode ser conduzida por um usuário de uma ferramenta vis.
  - Existe uma forte relação entre a forma dos dados - o atributo e os tipos de conjunto de dados - e que tipos de expressões vis são eficazes para exibi-lo. (\*cont.\*)

95

## Produzir

- A boa notícia é que suas mãos não são amarradas como um designer, porque você pode transformar os dados em uma forma mais útil para a tarefa em mãos.
- Não desenhe apenas o que é dado para você; Decida a coisa certa a mostrar que é, criá-la com uma série de transformações a partir do conjunto de dados original, e desenhe isso!

96

## Produzir



Figure 3.5. Derived attributes can be directly visually encoded. (a) Two original data attributes are plotted, imports and exports. (b) The quantitative derived attribute of trade balance, the difference between the two originals, can be plotted directly.

97

## Pesquisa

- ❑ Todos os casos de análise de alto nível exigem que o usuário procure por elementos de interesse dentro do vis como um objetivo de nível intermediário.
- ❑ A classificação da pesquisa em quatro alternativas é dividida de acordo com se a identidade e localização do alvo de pesquisa já é conhecida ou não.

98

## Pesquisa

- ❑ **Examinar**
  - ❑ Se os usuários já sabem o que eles estão procurando e onde ele está, então o tipo de pesquisa é simplesmente examinar.
  - ❑ Por ex., um usuário de uma árvore vis mostrando as relações ancestrais entre espécies de mamíferos pode querer olhar para os seres humanos, e pode chegar ao ponto certo rapidamente, lembrando como os seres humanos são classificados: eles estão no grupo de crias vivas ao invés de colocar ovos como um ornitorrinco ou tendo uma bolsa como cangurus, e dentro desse grupo seres humanos caem na categoria de primatas.

99

## Pesquisa

- ❑ **Localizar**
  - ❑ Para localizar um destino conhecido em um local desconhecido, o tipo de pesquisa é locate: ou seja, descobrir onde o objeto específico é.
  - ❑ Em um exemplo semelhante ao anterior, o mesmo usuário pode não saber onde encontrar coelhos, e teria que olhar ao redor em um número de lugares antes de localizá-los como lagomorfos (não roedores)!

100

## Pesquisa

- ❑ **Navegar**
  - ❑ Em contraste, a identidade exata de um alvo de busca pode não ser conhecida antecipadamente;
  - ❑ Em vez disso, ele pode ser especificado com base em características. Nesse caso, os usuários estão pesquisando um ou mais itens que se encaixam em algum tipo de especificação, como fazer correspondência com um intervalo específico de valores de atributos.
  - ❑ Quando os usuários não sabem exatamente o que estão procurando, mas eles têm uma localização em mente onde procurar, o tipo de pesquisa é navegar.

101

## Pesquisa

- ❑ **Explorar**
  - ❑ Quando os usuários não estão mesmo seguros da localização, o tipo de pesquisa é explorar. Ela implica a busca de características sem levar em conta a sua localização, muitas vezes a partir de uma visão geral de tudo.
  - ❑ Exemplos incluem a busca de valores discrepantes em um gráfico de dispersão, para picos anômalos ou padrões periódicos em um gráfico de linha de dados de séries temporais, ou por padrões espacialmente dependentes imprevistos em um mapa de calor.

102

## Consulta

- ❑ Uma vez que um alvo ou conjunto de destinos para uma pesquisa foi encontrado, um objetivo de usuário de nível baixo é consultar esses destinos em um dos três escopos: identificar, comparar ou resumir.
- ❑ A progressão destes três corresponde a um aumento na quantidade de alvos de busca em consideração: um, alguns ou todos. Ou seja, identificar refere-se a um único alvo, comparar se refere a várias metas, e resumir refere-se ao conjunto completo de possíveis alvos.

103

## Consulta

- ❑ **Identificar**
  - ❑ O escopo de identificar é um único alvo. Se uma pesquisa retornar alvos conhecidos, por pesquisa ou localizar, então identificar retorna suas características.
  - ❑ Por ex., um usuário de um mapa estático que representa os resultados das eleições dos EUA por codificação de cores cada estado vermelho ou azul, com o nível de saturação de qualquer matiz mostrando a proporção, pode identificar o partido vencedor e margem de vitória para o estado da Califórnia.

104

## Consulta

- ❑ **Comparar**
  - ❑ O escopo de comparação é múltiplas metas. Tarefas de comparação são geralmente mais difíceis do que identificar tarefas e exigem idiomas mais sofisticados para suportar o usuário.
  - ❑ Por ex., a capacidade de inspecionar um único alvo em detalhe é muitas vezes necessária, mas não suficiente, para comparação.

105

## Consulta

- ❑ **Resumir**
  - ❑ O escopo da tarefa resumir é todos os alvos possíveis.
  - ❑ Um sinônimo para resumir é uma visão geral, um termo é frequentemente usado na literatura vis como um verbo, onde significa fornecer uma visão abrangente de tudo, e como um substantivo, onde significa uma exibição sumária de tudo.
  - ❑ O objetivo de fornecer uma visão geral é extremamente comum na visualização.

106

## Alvos

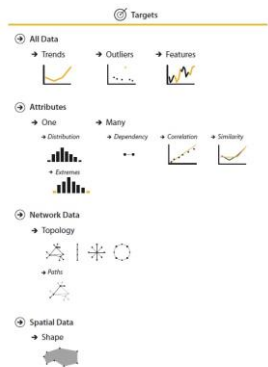


Figure 3.6: The goals of the user might be to find or understand specific aspects of the data: trends and outliers for all kinds of data; individual values, the minimum or maximum extremes of the range, or the entire distribution of a single attribute; or the dependencies, correlations, or similarities between multiple attributes; topology or paths for network data, and shape for spatial data.

107

107

## Alvos

- ❑ Três alvos de alto nível são muito amplamente relevantes, para todos os tipos de dados: tendências, outliers e características.
- ❑ Atributos são propriedades específicas que são visualmente codificadas. O alvo de nível mais baixo para um atributo é encontrar um valor individual. Outro alvo frequente de interesse é encontrar os extremos.
- ❑ Alguns alvos abrangem o escopo de vários atributos: dependências, correlações e semelhanças entre atributos.

108

### Como: Uma Preview

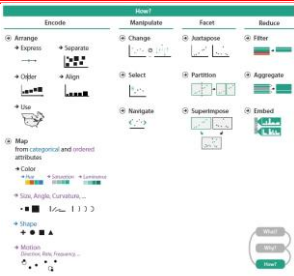


Figure 3.7. How to design via colors, encode, manipulate, facet, and reduce.

Visualization Analysis & Design

Porque: Abstração de Tarefas

109

109

### Analizando e Derivando: Exemplos

- Os três exemplos de análise e derivação são apresentados e dão uma amostra de como este quadro de "o que-por-como" pode ser usado imediatamente.
- O primeiro exemplo abrange a análise comparativa entre duas ferramentas vis.
- O segundo exemplo discute a derivação de um único atributo, uma medida de importância para as árvores decidir quais ramos mostrar para resumir sua estrutura topológica.
- O terceiro exemplo cobre a derivação de muitos novos atributos e o aumento de um conjunto de dados de dinâmica de fluidos espaciais criando espaços derivados em que características de interesse são fáceis de encontrar.

Visualization Analysis & Design

Porque: Abstração de Tarefas

110

110

### Exemplo: Comparando dois idiomas

- A estrutura de análise de "o quê por que" é útil para a análise comparativa, por exemplo, para examinar duas ferramentas vis diferentes que têm respostas diferentes para a questão de como o idioma é projetado quando usado exatamente para o mesmo contexto de por que e o que no nível abstração.
- O objetivo é apresentar um caminho traçado entre dois nós de interesse para um colega. Mais detalhadamente, ambas as ferramentas podem ser usadas para localizar caminhos entre nós e identificá-los.

Visualization Analysis & Design

Porque: Abstração de Tarefas

111

111

### Exemplo: Comparando dois idiomas

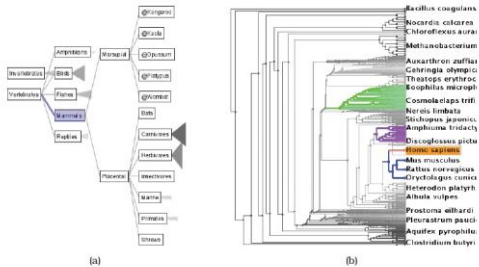


Figure 3.8. Comparing two idioms. (a) SpaceTree [Plaisant et al. 02]. (b) TreeJuxtaposer. From <http://www.cs.umd.edu/hcil/spacetrue> and [Munzner et al. 03, Figure 1].

Visualization Analysis & Design

Porque: Abstração de Tarefas

112

112

### Exemplo: Comparando dois idiomas

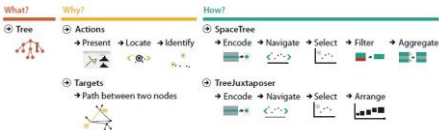


Figure 3.9. Analyzing what-why-how comparatively for the SpaceTree and TreeJuxtaposer idioms.

Visualization Analysis & Design

Porque: Abstração de Tarefas

113

113

### Exemplo: Derivando um atributo

- Numa vis que mostra uma rede ou uma árvore complexa, é útil ser capaz de filtrar a maior parte da complexidade desenhando uma imagem mais simples informando sua estrutura topológica.
- Uma maneira de apoiar esse tipo de resumo é calcular um novo atributo derivado que mede a importância de cada nó no gráfico e filtro com base nesse atributo.
- O número de Strahler é uma medida de importância de nó originalmente desenvolvida em hidrogeologia. Caracteriza a estrutura de ramificação de rios que foi adaptada e estendida, para uso em árvores e redes.

Visualization Analysis & Design

Porque: Abstração de Tarefas

114

114

### Exemplo: Derivando um atributo

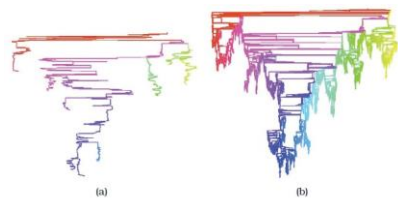


Figure 3.10. The derived quantitative attribute of Strahler numbers is used to filter the tree in order to create a recognizable summary. (a) The important skeleton of a large tree is visible when only 5000 of the highest-ranked nodes are drawn. (b) The full tree has over a half million nodes. From [Auber 02, Figures 10 and 13].

115

### Exemplo: Derivando um atributo

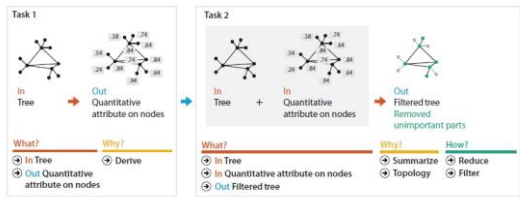


Figure 3.11. Analyzing a chained sequence of two instances where an attribute is derived in order to summarize a tree by filtering away the unimportant parts.

116

### Exemplo: Derivando muitos novos atributos

- As transformações de dados também podem lançar luz em dados espaciais. Em um exemplo da dinâmica de fluidos computacional, os espaços derivados ligados são usados para a detecção de características.
- O sistema vis ilustrado na Figura 3.12 permite que o usuário crie rapidamente gráficos de quaisquer duas variáveis originais ou derivadas a partir da paleta de variáveis mostradas no painel de campos derivado superior esquerdo.
- As vistas são ligadas juntamente com realce de cores.

117

### Exemplo: Derivando muitos novos atributos

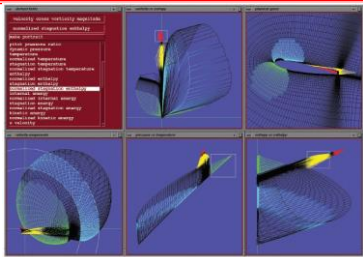


Figure 3.12. Computational fluid dynamics vis showing the list of many derived attributes (top left), one view of the original spatial field (top right), and four other views showing pairs of selected derived attributes. The multiple juxtaposed views are coordinated with shared colored highlights. From [Peters 96, Figure 15].

118

### Exemplo: Derivando muitos novos atributos

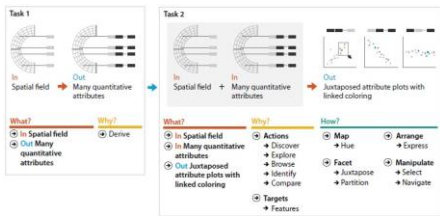


Figure 3.13. Analyzing a chained sequence, where many attributes are derived and visually encoded.

119

### Bibliografia

- Munzner, T. "Visualization Analysis & Design" 2014, Editora: AK Peters (1th Edition); ISBN-13: 978-1466508910

120