


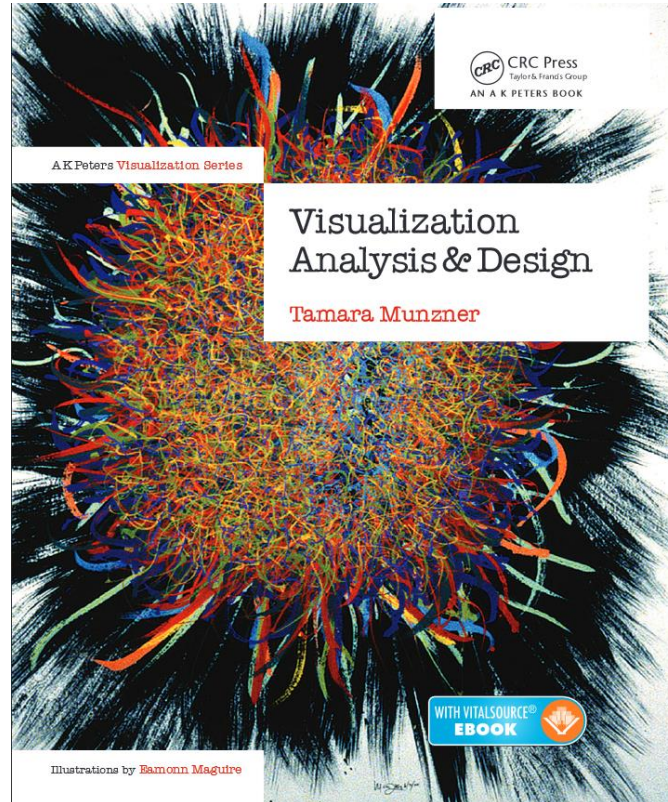
# Aula 08 – Visualização de Dados – Parte 01

Prof. Wellington Franco



# Conceitos Básicos de Visualização de Dados

# Bibliografia-base:



# O que é Visualização?

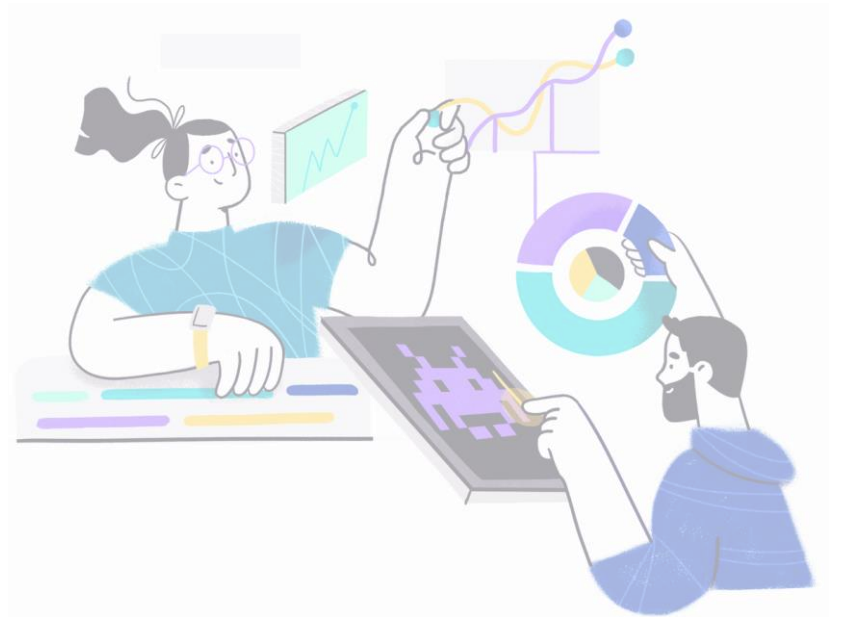
Visualização é uma área dedicada à **geração de imagens** que auxiliam seus usuários na **compreensão de dados e processos**.



# Quando usar Visualização?

A visualização é apropriada quando **há uma necessidade de aumentar as capacidades humanas e não substituir as pessoas com processos automatizados.**

**O humano está sempre envolvido!**



# E o que NÃO é Visualização?

É possível argumentar que qualquer coisa visual é de algum modo uma visualização... mas isso significa mesmo qualquer coisa?



“Se uma visualização é projetada para representar **dados visualmente**, fazendo isso de tal modo a **ganhar insights** sobre os dados, então ela deve ser chamada de **visualização pragmática**”.

Robert Kosara<sup>1</sup>, *Tableau Researcher*

# Visualização Pragmática

Três critérios mínimos que qualquer visualização deve respeitar para ser considerada pragmática:

1. Basear-se em dados (não-visuais);
2. Produzir uma imagem;
3. O resultado deve ser legível e reconhecível.



# Visualização Pragmática: analisando cada tópico

## 1. Basear-se em dados (não-visuais):

- Dados devem vir de alguma forma abstrata;
- Descartam-se fotografias ou processamento de imagem;
- Visualização transforma do invisível para visível.

## 2. Produzir uma imagem:

- Deveria ser óbvio, mas nem sempre é tão claro;
- O visual deve ser o modo primário da comunicação;
- Se a imagem é apenas uma parte do processo, então não é visualização.
  - Visualização é o resultado final.



# Visualização Pragmática: analisando cada tópico

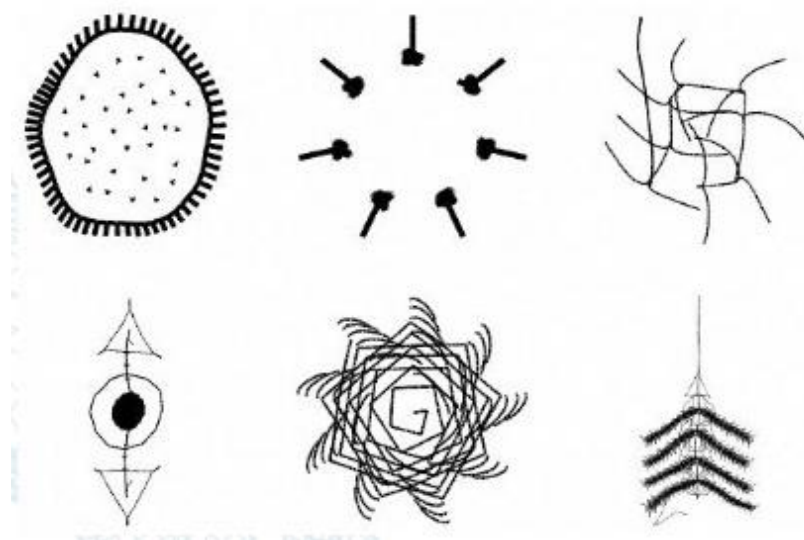
## 3. O resultado deve ser legível e reconhecível:

- É, sem dúvida, o critério mais importante;
- A visualização deve prover um aprendizado sobre os dados;
- Qualquer transformação de dados não-triviais em imagem vai deixar alguma informação de fora, mas devem permanecer aspectos relevantes dos dados que possam ser lidos.

# Exemplo do que NÃO é Visualização Pragmática:

- Dataset VISUAL-IDs:
  - Ajudam a diferenciar arquivos
  - Produzem imagens a partir de nome de arquivos para gerarem ícones visualmente semelhantes (ainda distintos) para arquivos.

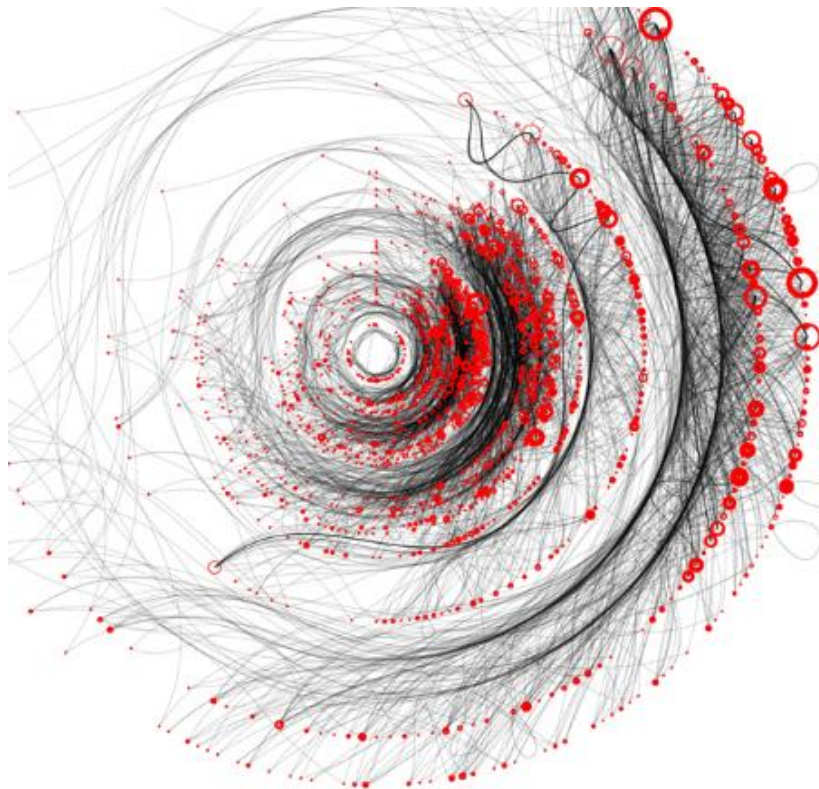
O Critério 3 da definição de visualização pragmática não é satisfeito, pois o resultado não é algo legível/ reconhecível.



# Exemplo do que NÃO é Visualização Pragmática:

- Poesia visual:
  - Poemas representados por valores numéricos, que por sua vez podem ser representados por círculos;
  - Círculos vermelhos representam um número.
    - Espessura do círculo depende da quantidade de palavras que são representadas pelo mesmo número

O Critério 3 da definição de visualização pragmática não é satisfeito, pois o resultado não é algo legível/reconhecível. Você consegue fazer uma interpretação dessa imagem?



# Visão Geral da Importância da Visualização

- A comunicação visual de dados já existe há bastante tempo;
- Hoje os dados são considerados objetos super valorizados;
- Com o bombardeio de informação, na hora de trabalharmos com um conjunto de dados, devemos atentar-se aos:
  - Ruídos;
  - Padrões;
  - Exceções;
  - A possível história contida em seu estado bruto.

# Importância da Visualização: O Experimento de Francis Anscombe

- Sejam quatro conjunto de dados;
- O que você consegue ver nesses conjuntos?

<b>x1</b>	<b>y1</b>	<b>x2</b>	<b>y2</b>	<b>x3</b>	<b>y3</b>	<b>x4</b>	<b>y4</b>
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

# Importância da Visualização: O Experimento de Francis Anscombe

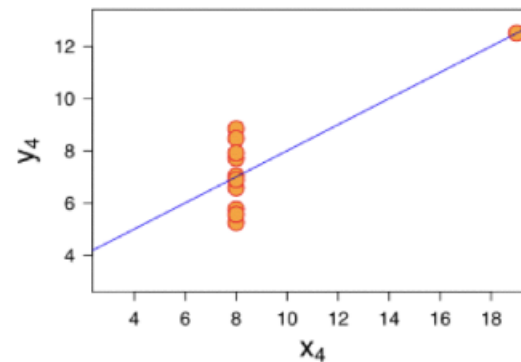
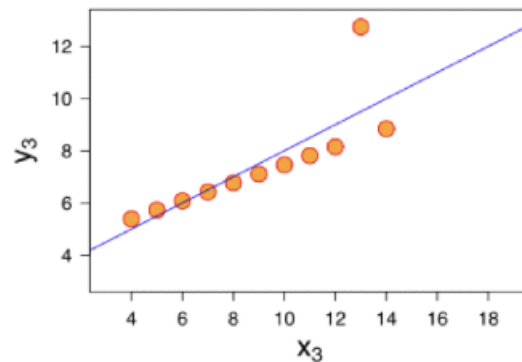
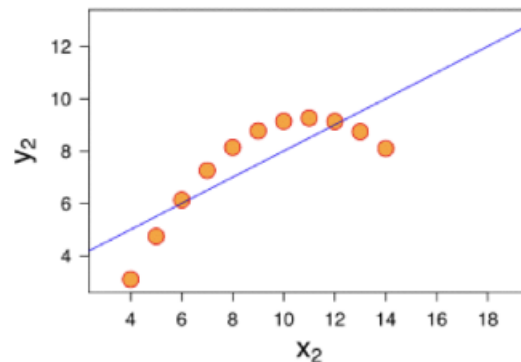
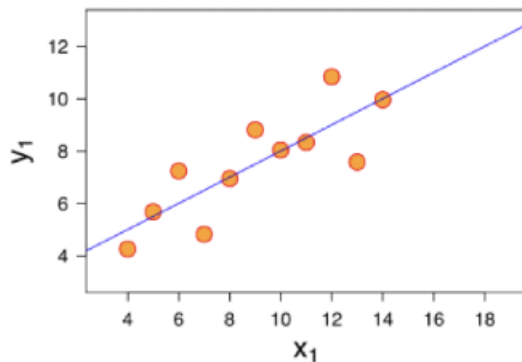
- Ao tirarmos a estatística geral dos dados (um *summary*: *média*, *variância*, *correlação*) de cada conjunto, temos o mesmo resultado para todos.

Identical statistics	
x mean	9
x variance	10
y mean	7,5
y variance	3,75
x/y correlation	0,816

Este experimento mostra que só confiar nesse sumário estatístico pode fazer com que haja perda de informação.

- A seguir, faremos uma plotagem dos dados em busca de algum

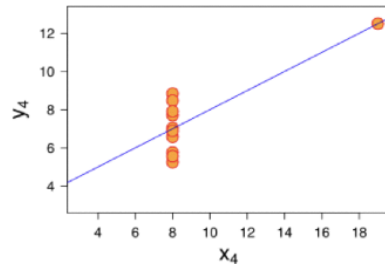
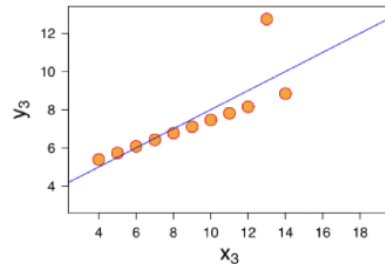
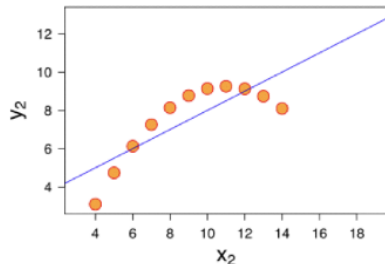
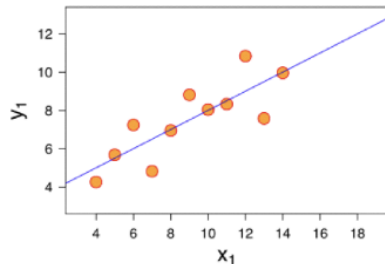
# Importância da Visualização: O Experimento de Francis Anscombe



# Importância da Visualização: O Experimento de Francis Anscombe

Agora conseguimos identificar os seguintes padrões:

- Uma certa tendência em relação a uma “linha de tendência” entre  $X_1$  e  $Y_1$ ;
- O padrão de curvatura entre  $X_2$  e  $Y_2$

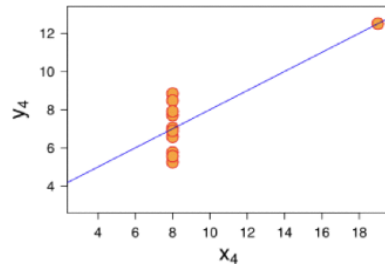
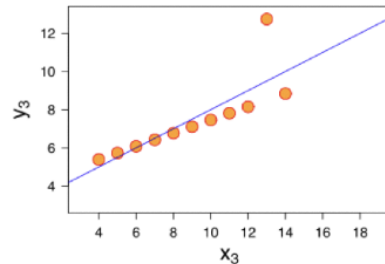
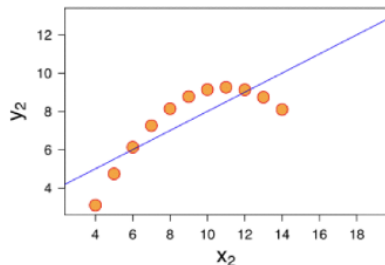
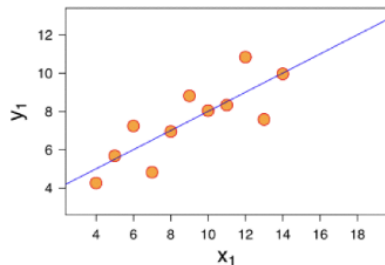




# Importância da Visualização: O Experimento de Francis Anscombe

Agora conseguimos identificar os seguintes padrões:

- Um forte padrão linear com apenas um *outlier* em  $X_3$  e  $Y_3$ ;
- Um forte padrão linear vertical também com apenas um *outlier* em  $X_4$  e  $Y_4$ .



# Importância da Visualização: O Experimento de Francis Anscombe

- Foi muito mais fácil descobrir e confirmar a presença (ou ausência) de padrões, relacionamentos e características físicas (*outliers*) através de recurso visual;
- Isso mostra que a visualização está fortemente relacionada ao processo de descoberta.

# VISFAIL: Quando a Visualização “Falha”

Um ponto que deve ser sempre levado em conta na hora de analisar uma visualização é tentar entender o que ela quer nos dizer (ou como estão fazendo com que determinado “*insight*” seja dito).



# VISFAIL: Quando a Visualização “Falha”

Exemplo: O que essa imagem quer nos dizer?



# VISFAIL: Quando a Visualização “Falha”

Exemplo: O que essa imagem quer nos dizer?

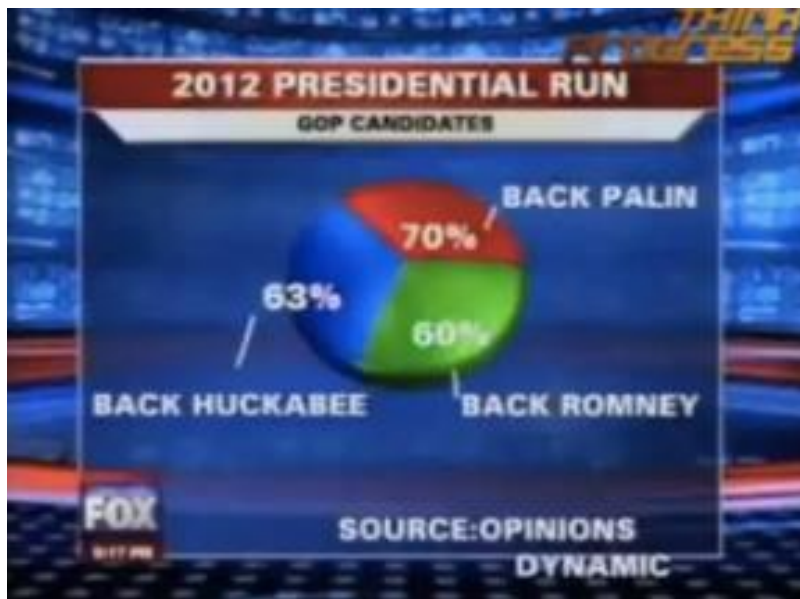


Pelo gráfico, podemos concluir que a inflação teve um aumento de 2010 a 2013; mas como isso seria possível se as porcentagens diminuíram de 5,92% para 5,91%?

A falta de proporcionalidade das barras nos induzem a tirar conclusões equivocadas.

# VISFAIL: Quando a Visualização “Falha”

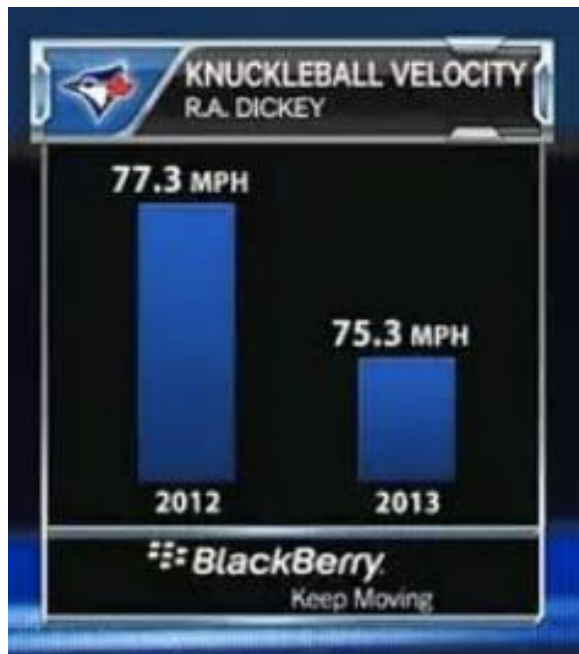
Exemplo: O que essa imagem quer nos dizer?



- Qual o sentido dessas porcentagens no gráfico de pizza?
- Por que o valor de 63% soa maior que o de 70%?

# VISFAIL: Quando a Visualização “Falha”

Exemplo: O que essa imagem quer nos dizer?



Neste caso da velocidade da bola em um taco de beisebol ter diminuído de 77.3 para 75.3, por que no gráfico de barras a redução foi pela metade?

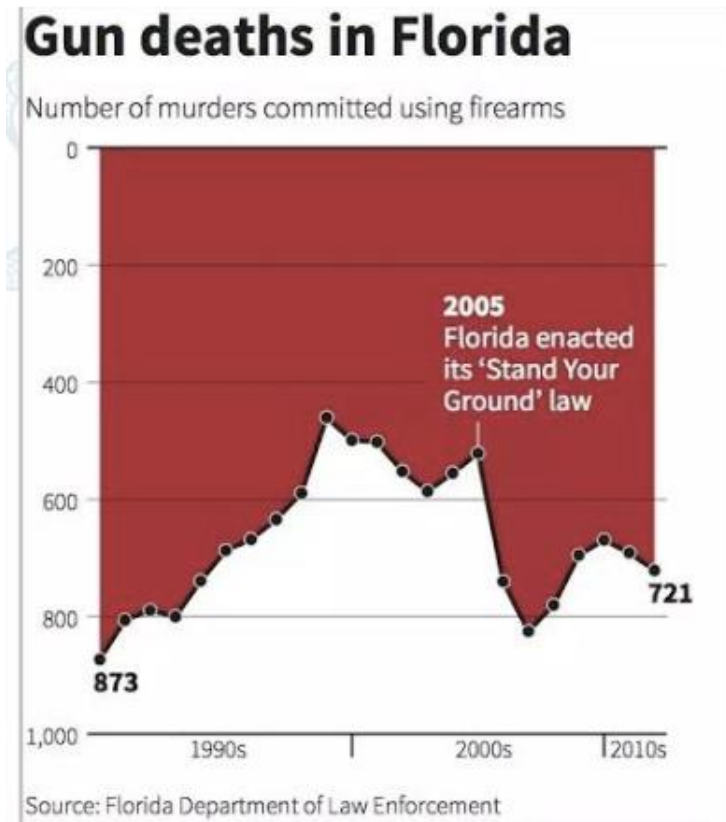
# VISFAIL: Quando a Visualização “Falha”

Ex: O que essa imagem quer nos dizer?

O gráfico mostra o número de mortes por arma de fogo na Flórida ao passar das décadas.

Numa rápida olhada, podemos concluir que na década mais recente houve um decréscimo.

Entretanto, o eixo Y está invertido.





# Definição de Visualização (*vis*)

Com tudo o que foi discutido até aqui, podemos trabalhar sob a seguinte definição do *estado-da-arte*:

**“Sistema de visualização baseado em computador que provém representações visuais de conjuntos de dados projetados para ajudar pessoas a realizarem tarefas mais efetivamente”.**

A seguir vamos destrinchar trechos da definição acima e discuti-las detalhadamente.

# Definição de Visualização (*vis*)

Visualização possibilita pessoas a analisarem dados quando elas não sabem, a priori, as perguntas que precisam ser feitas.

“Sistema de visualização baseado em computador que provém representações visuais de **conjuntos de dados** projetados para ajudar **pessoas** a realizarem tarefas mais efetivamente”.

Queremos a aumentar a capacidade das pessoas **e não substituí-las por processos automatizados.**

# Definição de Visualização (*vis*)


“Sistema de visualização baseado em computador que provém representações visuais de **conjuntos de dados** projetados para ajudar **pessoas** a realizarem tarefas mais efetivamente”.



- Auxilia na Análise Exploratória de Dados para checar hipóteses;
- Dá suporte às pessoas que querem explicar algo que elas compreendem bem.

# Definição de Visualização (*vis*)

“Sistema de visualização baseado em computador que provém **representações visuais** de conjuntos de dados projetados para ajudar pessoas a realizarem tarefas mais efetivamente”.



Representação que foca em substituir cognição por percepção.

- Pessoas transferem o uso de cognição e memória para o sistema perceptivo.
  - Ex: uso de diagramas para organizar as informações facilitando sua busca e reconhecimento.

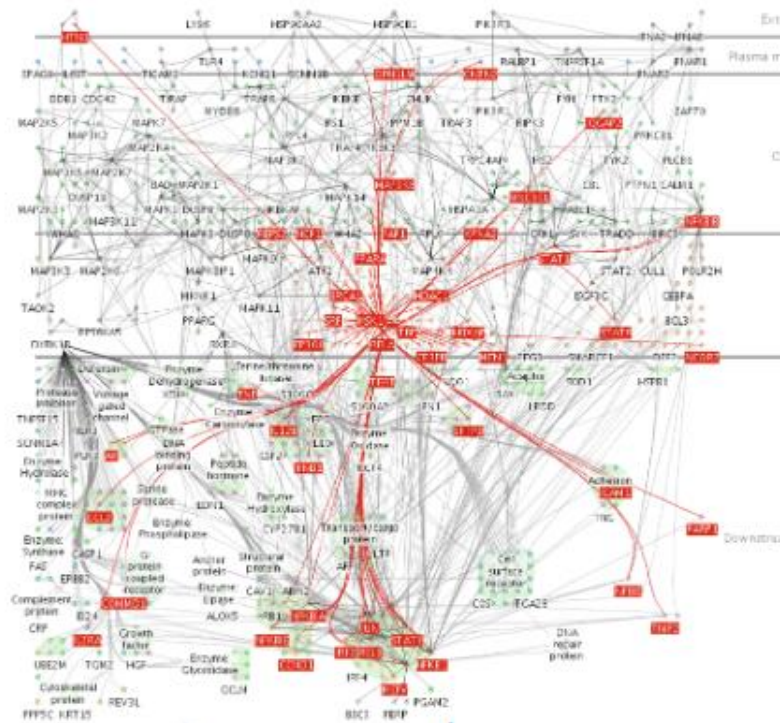
# Definição de Visualização (*vis*)

O uso de computadores é essencial dado o grande volume de dados




“Sistema de visualização **baseado em computador** que provém representações visuais de conjuntos de dados projetados para ajudar pessoas a realizarem tarefas mais efetivamente”.

Grafo com 760 nós e 1269 arestas



# Definição de Visualização (*vis*)

“Sistema de visualização baseado em computador que provém representações **visuais** de conjuntos de dados projetados para ajudar pessoas a realizarem tarefas mais efetivamente”.



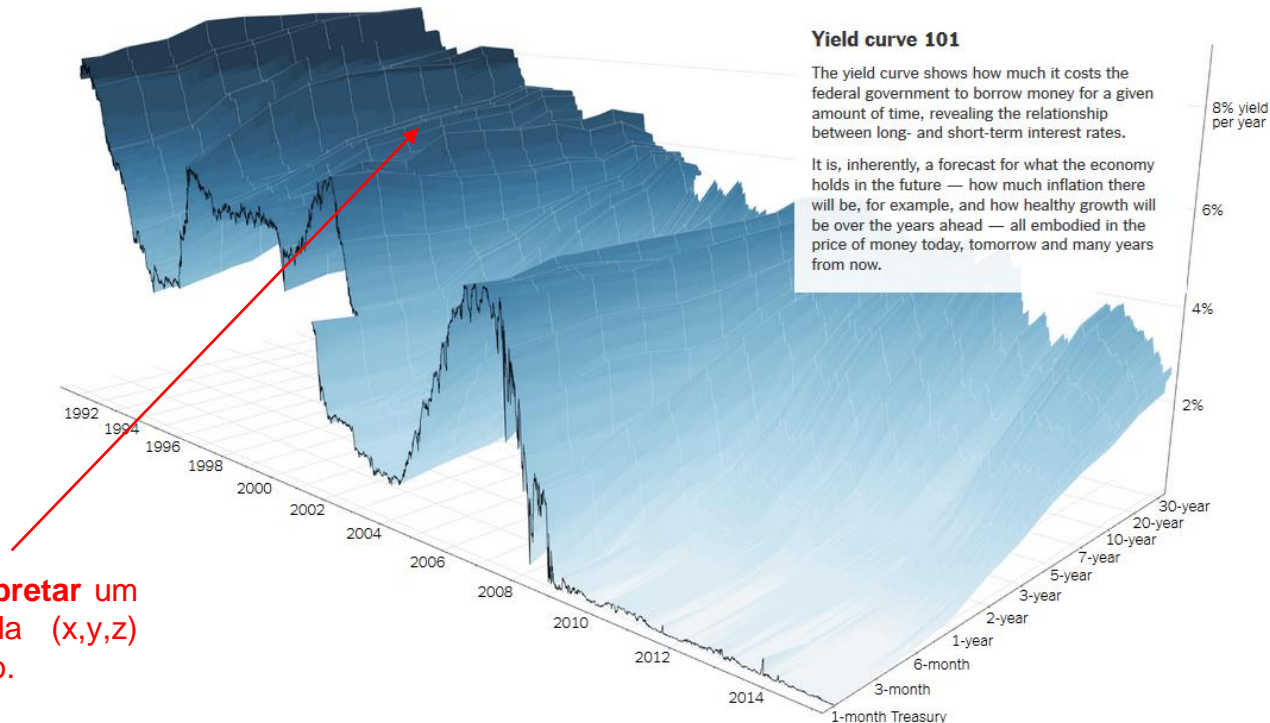
- O sistema visual humano é um canal de banda-larga para o cérebro
  - Uma quantidade significativa de processamento da informação visual ocorre em **paralelo** no nível pré-consciente (processamento pré-atentivo)
- O som não é apropriado para prover overviews de grandes quantidades de informação ao se comparar com a visão.
  - Canal de banda-estreita, **percepção sequencial**.

# Definição de Visualização (*vis*)

“Sistema de visualização baseado em computador que provém representações visuais de conjuntos de dados projetados para ajudar pessoas a realizarem **tarefas mais efetivamente**”.

- Realizar uma tarefa de visualização com efetividade tem a ver com **corretude, precisão e verdade**.
- Não se trata só de imagens bonitas (ver slide seguinte)

Este gráfico de linha em 3D embora bonito, torna-se bastante complicado de fazer interpretações...



Imagina ter que **interpretar** um ponto na coordenada (x,y,z) nesta região do gráfico.

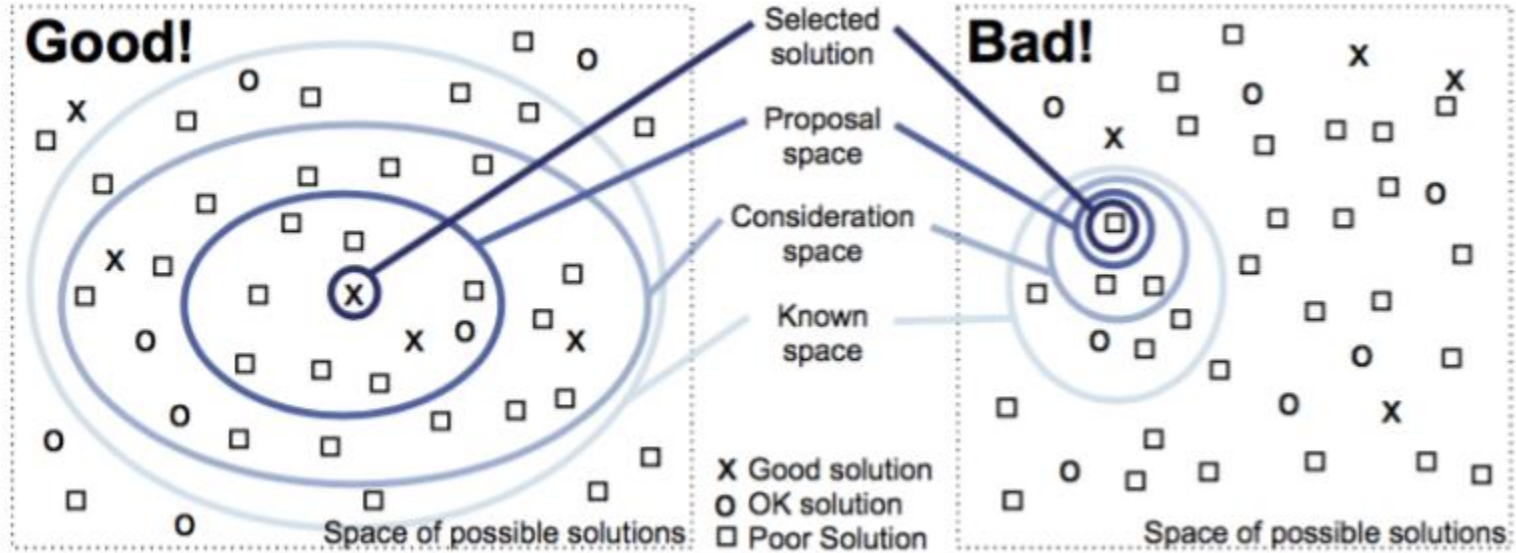


# Definição de Visualização (*vis*)

“Sistema de visualização baseado em computador que provém representações visuais de conjuntos de dados projetados para ajudar pessoas a realizarem **tarefas mais efetivamente**”.

- No entanto, imagem alguma pode comunicar toda a verdade;
- Corretude é complicado de verificar porque qualquer representação de dados é uma representação onde escolhas são feitas em relação a quais aspectos enfatizar.

# A Efetividade da Escolha



De acordo com a imagem acima, quanto melhor o espaço de conhecimento sobre o assunto, melhor a probabilidade de uma escolha acertada sobre o espaço a ser considerado para proposta e solução de uma análise.

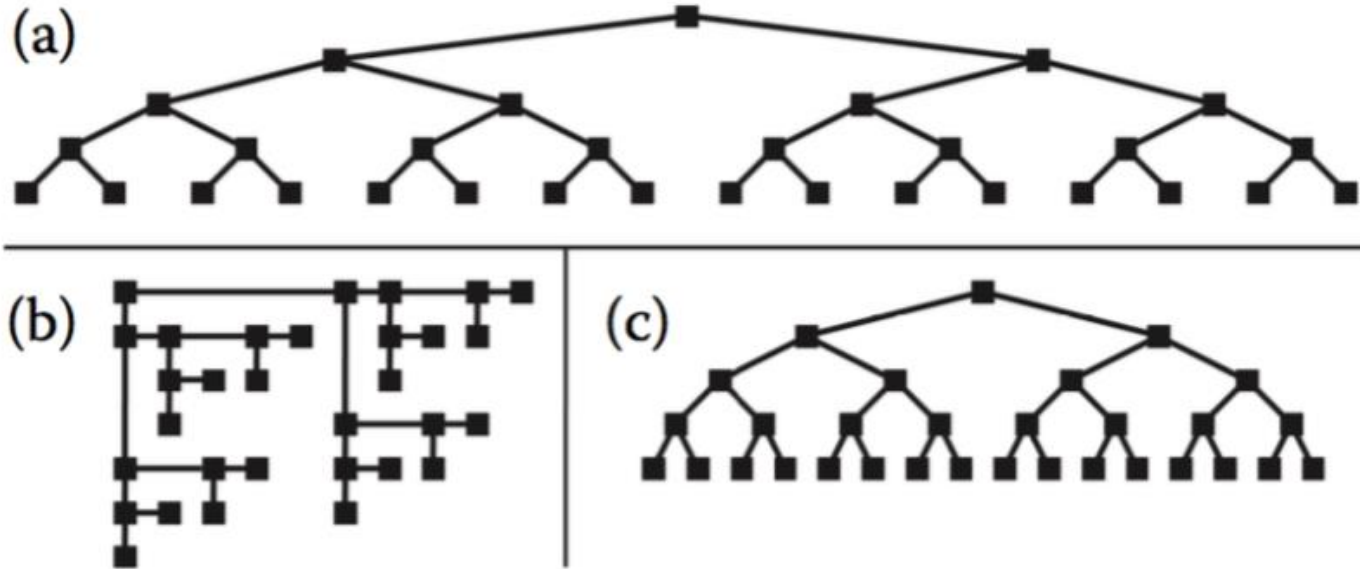
# Por Que a Validação é Difícil?

- Validar uma visualização é difícil porque existem muitas perguntas que poderíamos fazer ao considerar se uma visualização atingiu seus objetivos de análise;
- Como você argumenta que uma análise é melhor ou pior?

# Limitações de Recurso

- *Designers de Vis* devem levar em consideração três tipos de limitações de recursos:
  - Limitações computacionais;
    - Tempo de processamento;
    - Memória do sistema.
  - Limitação Humana;
    - Atenção e memória humana.
  - Limite de display.
    - Pixels são um recurso precioso e o recurso mais restrito.
    - **Densidade de informação:** razão do espaço usado para codificar a informação X espaço em branco (desperdício)

# Densidade de Informação



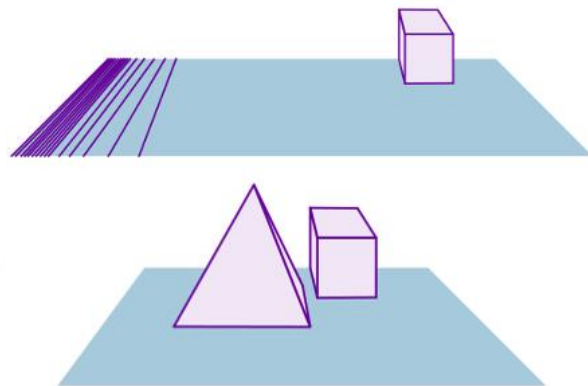
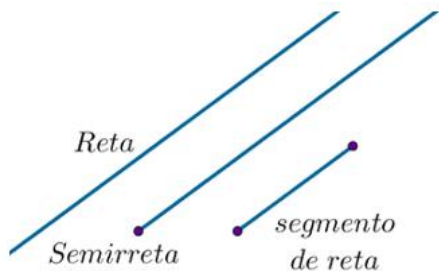
Como decidir quando usar bem a informação juntamente com espaço não utilizado (poluição X espaço desperdiçado)? (a), (b) e (c) representam a mesma árvore.

# Marcadores e Canais (*Marks and Channels*)

- **Marcações** e **Canais** são elementos básicos para analisarmos codificações visuais;
- A **combinação visual** geram dois aspectos:
  - **Marcações:** elementos gráficos;
  - **Canais visuais:** para controlar a aparência das marcações.

# Marcadores e Canais (*Marks and Channels*)

- Uma **marcação** é um elemento gráfico básico em uma imagem:
  - Elemento gráfico gerado a partir de primitivas geométricas.
    - Uma **primitiva geométrica** é o ponto, a reta, o plano ou o espaço.



# Marcadores e Canais (*Marks and Channels*)

- Um **canal** visual é uma maneira de controlar a aparência de marcações, independentemente da dimensionalidade da primitiva geométrica.

## → Position

→ Horizontal



→ Vertical



→ Both



## → Color



## → Shape



## → Tilt



## → Size

→ Length



→ Area



→ Volume





# Gráficos Quantitativos

- Exemplos de gráficos quantitativos mostrando um ou mais atributos;
  - Utilização de um ou mais canais visuais.
- Tente perceber nos gráficos a importância de combinar corretamente marcadores e canais para determinadas informações que se quer passar.



(a)



(b)



(c)



(d)

# Gráficos Quantitativos

- Exemplos de gráficos quantitativos mostrando um ou mais atributos;



(a)



(b)



(c)



(d)

Em (a), **bar chart (ou gráfico de barras)**, usamos dois atributos: um quantitativo (eixo y) e o outro categórico (eixo x), servindo apenas para espalhar as barras ao longo do eixo.

# Gráficos Quantitativos

- Exemplos de gráficos quantitativos mostrando um ou mais atributos;



(a)



(b)



(c)



(d)

Em (b), **scatterplot** (ou **gráfico de dispersão**), codificam dois atributos quantitativos usando marcas de ponto e posição espacial vertical e horizontal.

# Gráficos Quantitativos

- Exemplos de gráficos quantitativos mostrando um ou mais atributos;



(a)



(b)



(c)



(d)

Em **(c)**, a adição de mais um atributo através de recurso de canal de matiz (aspecto de cor).

# Gráficos Quantitativos

- Exemplos de gráficos quantitativos mostrando um ou mais atributos;



(a)



(b)



(c)

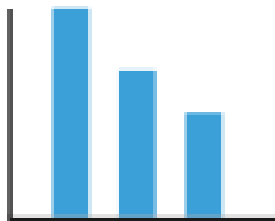


(d)

Em **(d)**, a adição de mais um atributo através de recurso de canal tamanho.

# Gráficos Quantitativos

- Resumo dos gráficos:



(a)

**Marcação:** Linha

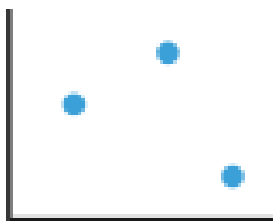
**Canal:**

Posição vertical

**Atributos:**

1 Quantitativo

1 Categórico



(b)

**Marcação:** Ponto

**Canal:**

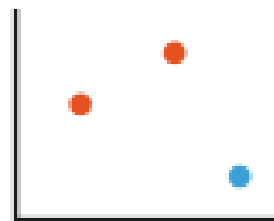
Posição vertical

Posição horizontal

**Atributos:**

1 Quantitativo

1 Categórico



(c)

**Marcação:** Ponto

**Canal:**

Posição vertical

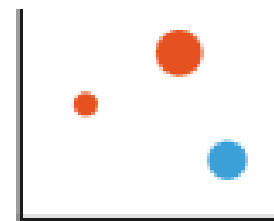
Posição horizontal

Matiz de cor

**Atributos:**

2 Quantitativos

1 Categórico



(d)

**Marcação:** Ponto

**Canal:**

Posição vertical

Posição horizontal

Matiz de cor

Tamanho (área)











**Atributos:**





3 Quantitativos

1 Categórico

# Tipos de Canais

- Podemos ter dois tipos de canais:
  - Magnitude;
  - Identidade.

Magnitude	
Position on common scale	
Position on unaligned scale	
Length (1D size)	
Tilt/angle	
Area (2D size)	
Depth (3D position)	
Color luminance	
Color saturation	
Curvature	
Volume (3D size)	

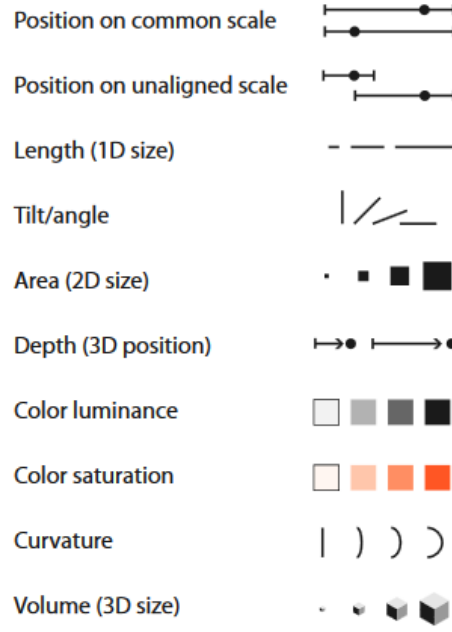
Identidade	
Spatial region	
Color hue	
Motion	
Shape	

↑  
*Glyphs*

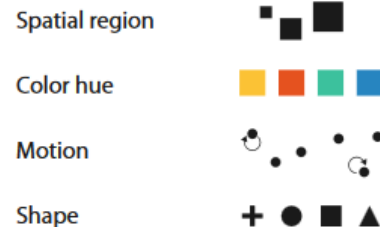
# Princípio da Expressividade e Efetividade

- Princípio da Efetividade:
  - Codifique os atributos mais importantes usando **canais de alto ranking**
  - Canais de alto ranking definidos no *estado-da-arte*.

## ➔ Magnitude Channels: Ordered Attributes



## ➔ Identity Channels: Categorical Attributes



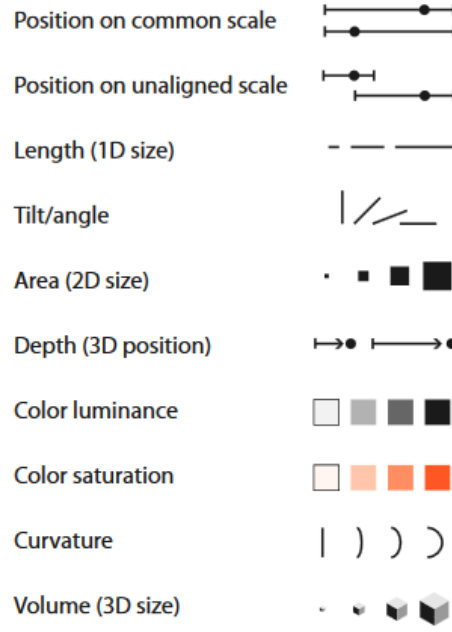


# Princípio da Expressividade e Efetividade

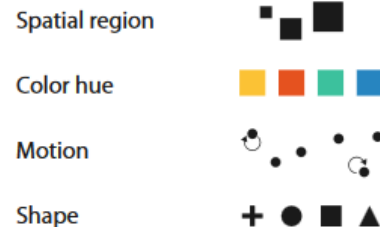
- Princípio da Expressividade:

- Combinar as características dos dados com as características do canal:
  - Canais de identidade: são a combinação perfeita para atributos categóricos.
  - Canais de magnitude: são apropriados para atributos ordenados (ordinais e quantitativos).

## ➔ Magnitude Channels: Ordered Attributes



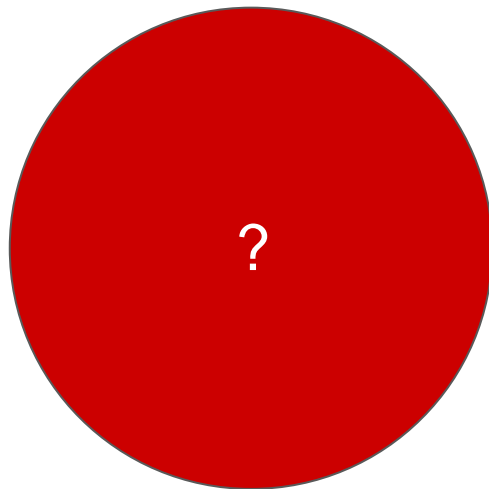
## ➔ Identity Channels: Categorical Attributes



# Princípio da Expressividade e Efetividade

- É preciso ter cuidado com o uso de alguns canais, pois alguns fornecem a informação de que há diferença na quantidade, mas que não há como quantificá-la.
  - Exemplos: Área e intensidade de cor.

Quantas vezes maior é o círculo da direita com relação ao da esquerda?



# Crítérios de Efetividade de Canais

Listaremos a seguir alguns dos principais critérios de efetividade dos canais necessário para entender as características dos canais visuais.

Esses critérios nos ajudam a responder perguntas do tipo:

- Por que o *designer de vis* escolheu este canal em particular?
- Por que alguns canais são melhores do que outros?

# Cr terios de Efetividade de Canais: Discriminabilidade

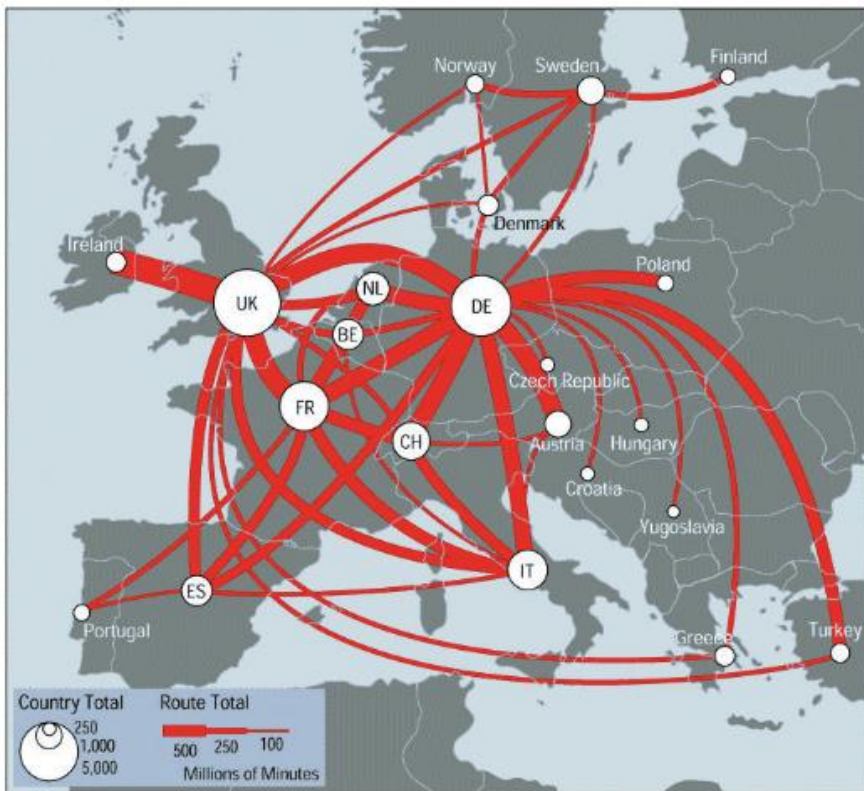
- A quest o da discriminabilidade  : se voc  codifica dados usando um canal visual espec fico, as diferen as entre os itens s o percept veis ao ser humano como pretendido?
- Alguns canais possuem um n mero limitado de n veis.
  - Exemplo: comprimento da linha (*linewidth*)
  - Comprimento da linha serve para at  3 ou quatro n veis de espessura diferentes. Mais que isso, fica invi vel.

# Crítérios de Efetividade de Canais: Discriminabilidade

## Exemplo:

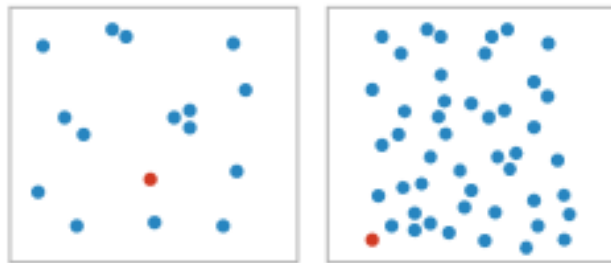
Aqui temos três níveis de espessura de linha (porque há três ranges diferentes).

Se houvesse um range maior a ser mostrado, provavelmente deveria-se escolher outro canal visual para exibir a informação.



# Cr terios de Efetividade de Canais: *Popout*

- *Popout*   quando nos referimos a um item no conjunto de dados que queremos destacar de outros;
- O grande valor da *popout*   que o tempo que levamos para localizar o objeto diferente n o depende do n mero de objetos distratores. Exemplo:



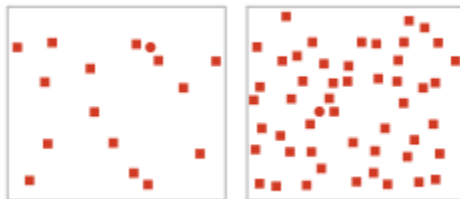
O tempo que leva para o  rculo vermelho sair do mar dos azuis   aproximadamente igual quando existem 15 azuis (esquerda) e 50 azuis (direita).

# Crítérios de Efetividade de Canais: *Popout*

- Velocidade não depende do número de *objetos distraidores*;



- Velocidade depende do canal e do quão diferente o objeto procurado é dos objetos distraidores.



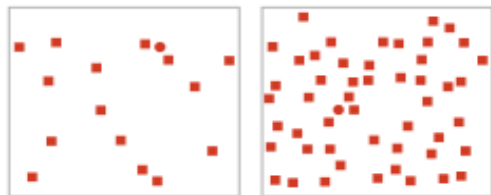
A diferença de cor é maior que a diferença de formato

# Crítérios de Efetividade de Canais: *Popout*



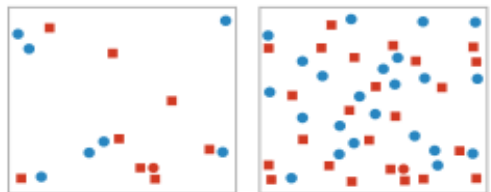
(a)

(b)



(c)

(d)



(e)

(f)

- (a) O círculo vermelho sai de um pequeno conjunto de círculos azuis;
- (b) O círculo vermelho sai de um grande conjunto de círculos azuis com a mesma rapidez;
- (c) O círculo vermelho também sai de um pequeno conjunto de formas quadradas, embora um pouco mais lento que nas cores;
- (d) o círculo vermelho também sai de um grande conjunto de quadrados vermelhos, só que mais lento ainda;
- (e) O círculo vermelho não leva muito tempo para ser encontrado em um pequeno conjunto de formas e cores misturadas.
- (f) Entretanto, se aumentada a quantidade de dados, é necessário uma busca um-a-um.



# Crítérios de Efetividade de Canais: *Popout*

A imagem ao lado mostra que muitos canais são eficientes para visualização de *popout*, exceto o último.

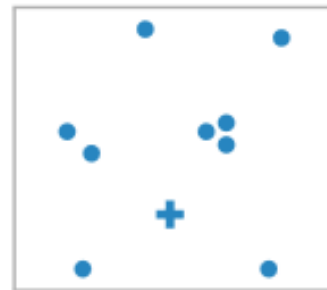
**Canais:** (a) Inclinação, (b) Tamanho, (c) Forma, (d) Proximidade, (e) Direção de Sombreamento e (f) Linhas Paralelas.



(a)



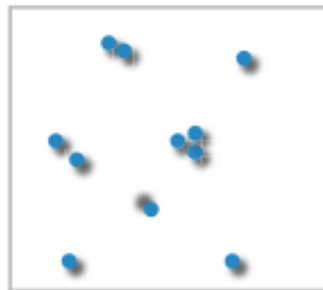
(b)



(c)



(d)



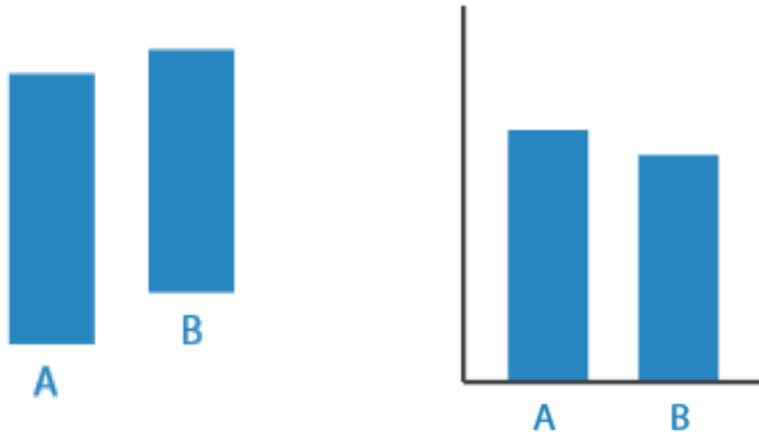
(e)



(f)

# Crítérios de Efetividade de Canais: *Julgamento*

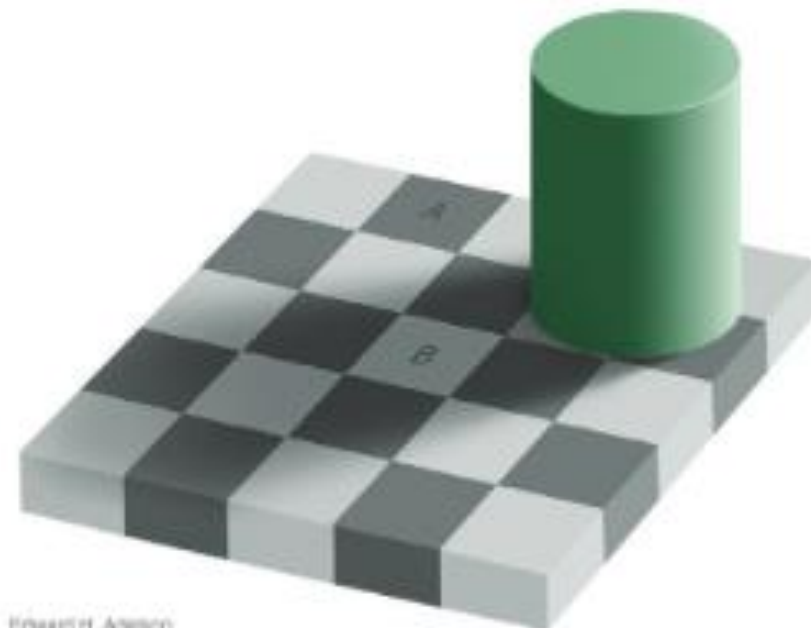
- O sistema perceptivo humano baseia-se na maioria das vezes em julgamentos relativos, não absolutos.
  - Para que haja uma maior acurácia em uma análise visual, o uso de escala visuais é um excelente suporte. Exemplo:



A figura da esquerda pode gerar diferentes julgamentos relativos, enquanto que com auxílio de uma escala podemos ter um julgamento absoluto.

# Cr terios de Efetividade de Canais: *Julgamento*

- A figura ao lado mostra que nossa percep  o de cores e lumin ncia   completamente contextual, baseada no contraste das cores ao redor.
- Quantas cores de cinza temos na imagem?



Edward H. Adelson

# Cr terios de Efetividade de Canais: *Julgamento*

- Ao tra armos duas retas paralelas (que servir o como escala) veremos que ambas as cores s o as mesmas.

