



# Mineração de Dados

GRIMALDO OLIVEIRA

# Sobre Grimaldo

---



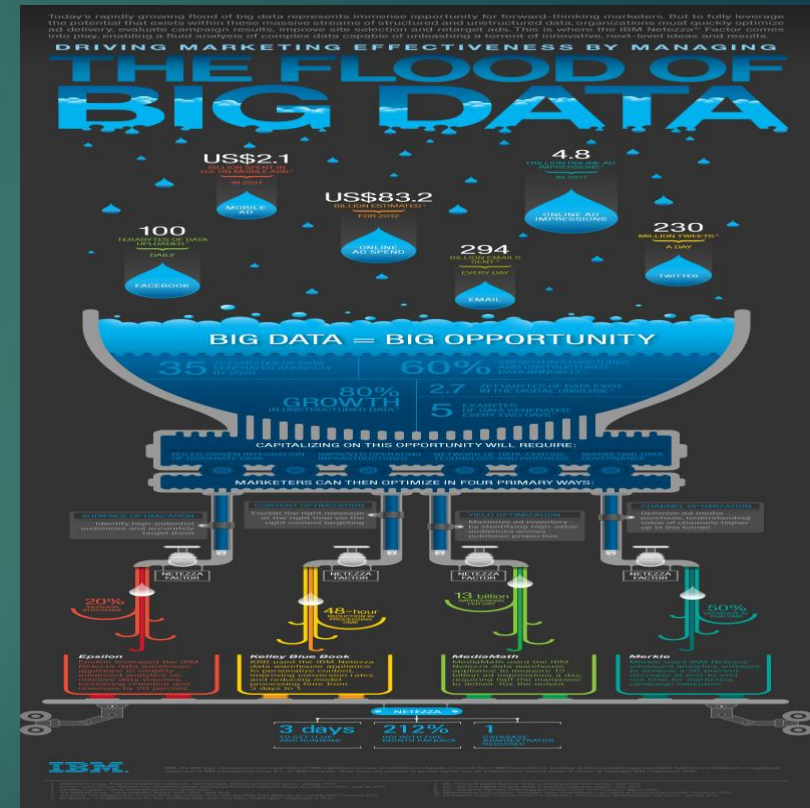
- Grimaldo Oliveira
  - [grimaldo\\_lopes@hotmail.com](mailto:grimaldo_lopes@hotmail.com)
- Formação
  - Mestre em Tecnologias Aplicadas a Educação Universidade do Estado da Bahia.
  - Especialização em Análise de Sistemas pela Faculdade Visconde de Cairu.
  - Estatístico pela Universidade Federal da Bahia.
- Atividades
  - Mais de 10 anos atuando como Consultor de Business Intelligence.
  - Projetos Governos Maranhão, Mato Grosso e Bahia.
  - Idealizador do Blog : BI com Vatapá – [bicomvatapa.blogspot.com](http://bicomvatapa.blogspot.com).
  - Livro: BI Como Deve Ser – [bicomodeveser.com.br](http://bicomodeveser.com.br)

# Agenda

- ▶ Conceito de Big Data
- ▶ O caminho através do cientista de dados
- ▶ Mineração Visual de Dados
  - ▶ Panorama visual
  - ▶ Análise exploratória visual
  - ▶ D3js - Biblioteca para mineração visual de dados

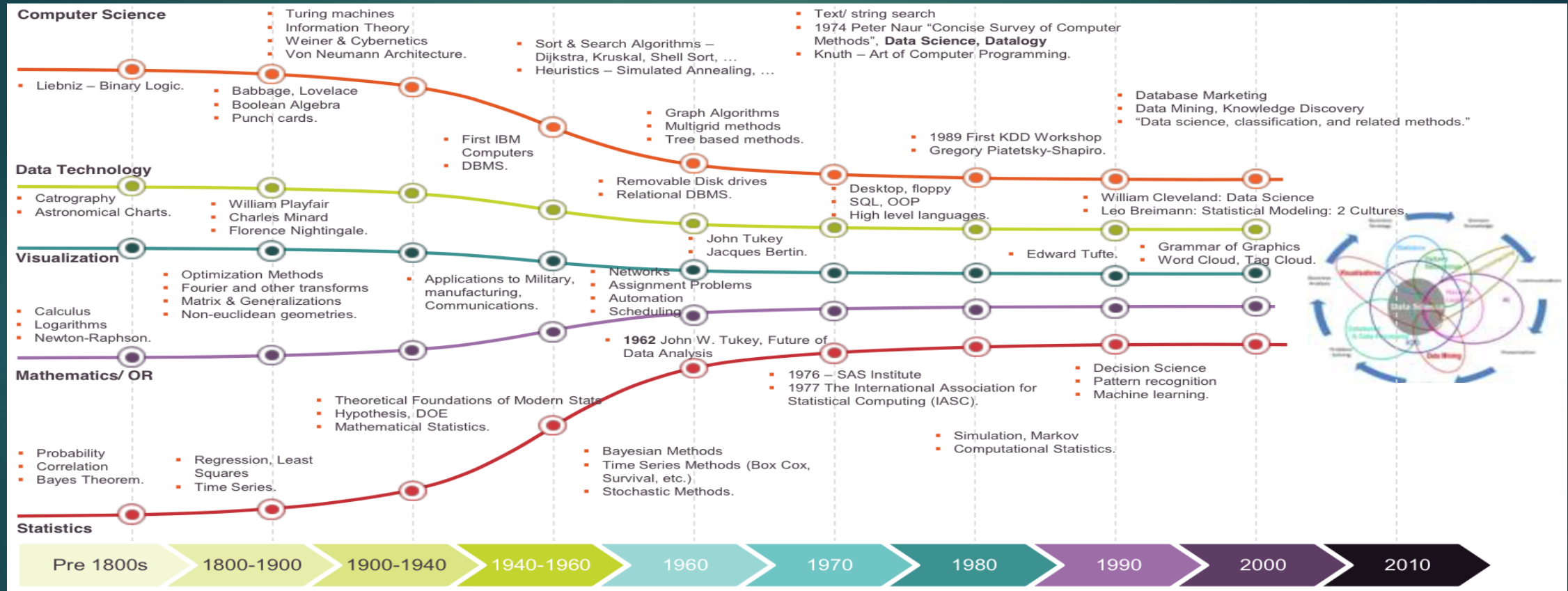
# Big Data – Enfrentando o futuro

- \* 83,2 bilhões dólares é o montante estimado gasto nos EUA em publicidade online;
- \* 230 milhões de tweets são enviados por dia ;
- \* 100 terabytes de dados são enviados diariamente para Facebook;
- \* **80 por cento de crescimento em dados não estruturados;**
- \* 2.7 zettabytes de dados existem no universo digital;



Fonte: IBM - 01 de maio 2012

# Infográfico – Cientista de Dados

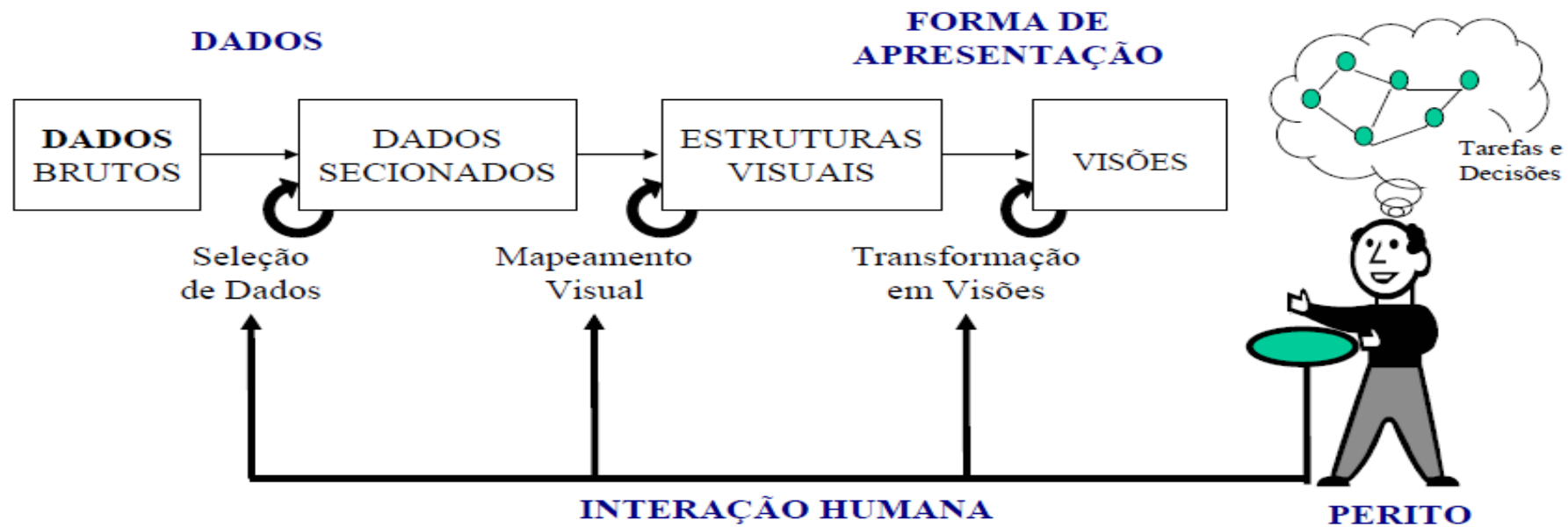


# Motivação

O uso de visualização e exploração interativa de dados ataca este problema diretamente;

- Visão é o sentido mais eficaz para se transferir informação para um ser humano;
- Bons paradigmas visuais jogam com as habilidades naturais dos seres humanos e com características do domínio de conhecimento, para a criar interfaces que podem ser facilmente aprendidas e aceitas por peritos;
- Esta abordagem retira os intermediários do processo de exploração dados, tornando-o muito mais interativo, rápido e barato.

# Modelo



*(Adaptado Card et al. 1999)*



# Exemplo: Porte de Hidroelétricas

Nome	Rio	Potência	Turbina	Estado
Itaipu	Paraná	11400	Kaplan	Paraná
Paulo Afonso	São Francisco	2700	Pelton	Bahia
Xingó	São Francisco	3200	Kaplan	Alagoas
Tucuruí	Tocantins	6300	Kaplan	Pará

**Dados Tabulares**



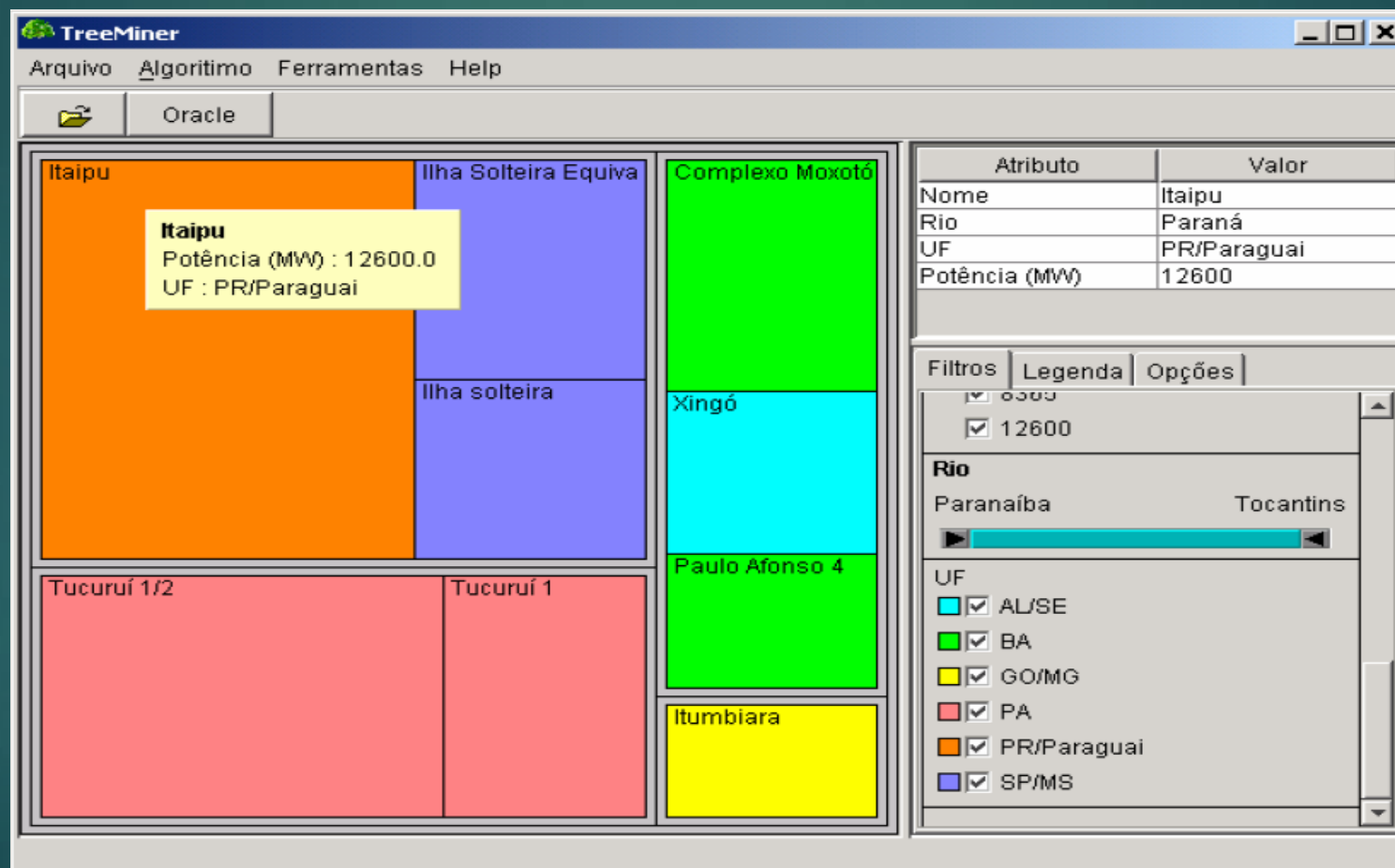
# Mapeamento dos Dados para Atributos Visuais

## Atributos Visuais

Etiqueta		Tamanho	Cor	Posição
↕		↕	↕	↕
Mapeamento				
Nome	Rio	Potência	Turbina	Estado
Itaipu	Paraná	11400	Kaplan	Paraná
Paulo Afonso	São Francisco	2700	Pelton	Bahia
Xingó	São Francisco	3200	Kaplan	Alagoas
Tucuruí	Tocantins	6300	Kaplan	Pará

## Dados

# Visualização dos Dados Mapeados



# Seleção dos Dados de Interesse

**application**

☐ All      **Selected**      ☐ None

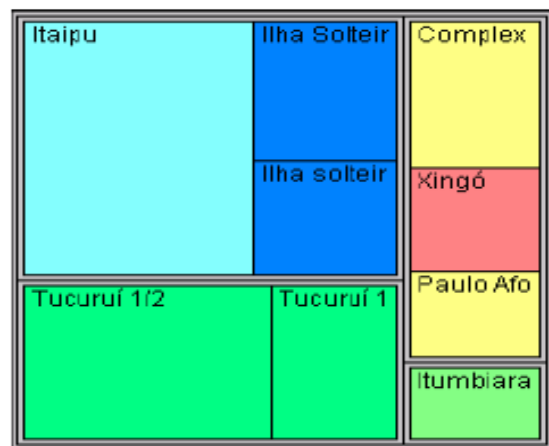
- ☒ Power Point
- ☒ program
- ☒ Web browser

**date**

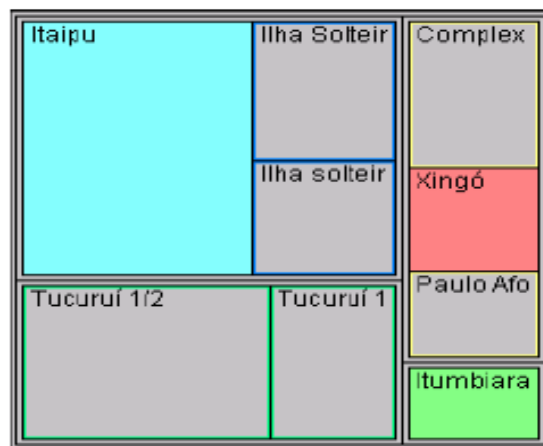
UF

- ☒ ☐ AL/SE
- ☒ ☐ BA
- ☒ ☐ GO/MG
- ☒ ☐ PA
- ☒ ☐ PR/Paraguai
- ☒ ☐ SP/MS

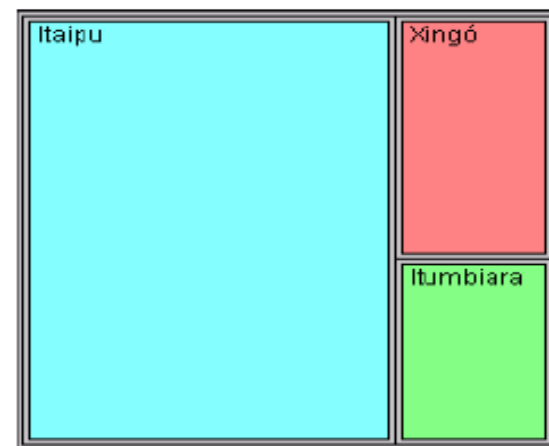
# Aplicando Filtros



(a)



(b)



(c)

- (a) Conjunto original.
- (b) Após a aplicação da seleção sem esconder os itens.
- (c) Após a aplicação da seleção escondendo os itens.

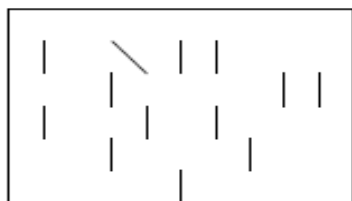
# Atributos Visuais - Outros

Existem vários tipos de atributos visuais que se pode utilizar para representar um atributo real:

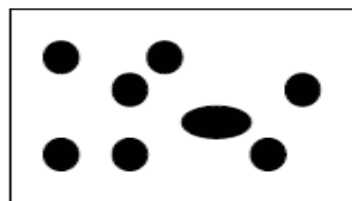
- **Forma** (ex.: largura, tamanho, curvatura, orientação)
- **Cor** (ex.: tonalidade, intensidade),
- **Movimento** (ex.: piscar, direção do movimento)
- **Posição Espacial** (ex.: côncavo/convexo, 2D, 3D)

Atributos podem ser exibidos simultaneamente representando múltiplas dimensões de um registro de dados

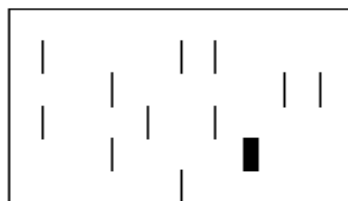
# Exemplo – Atributos Visuais



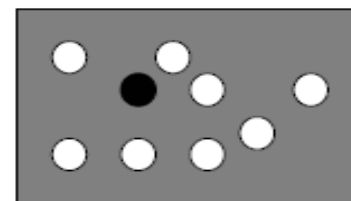
**ORIENTAÇÃO**



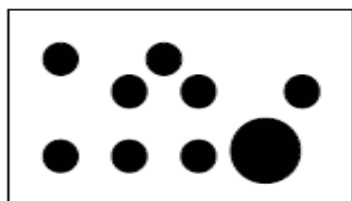
**FORMA**



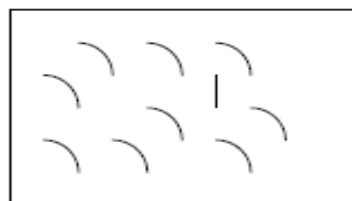
**FORMA**



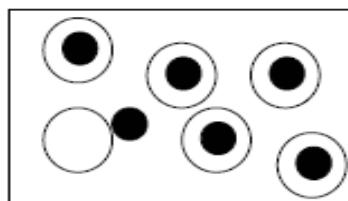
**VALOR**



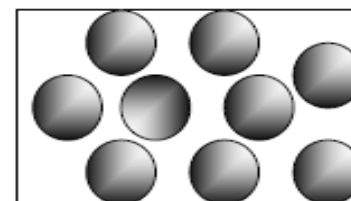
**TAMANHO**



**CURVATURA**



**DELIMITAÇÃO**



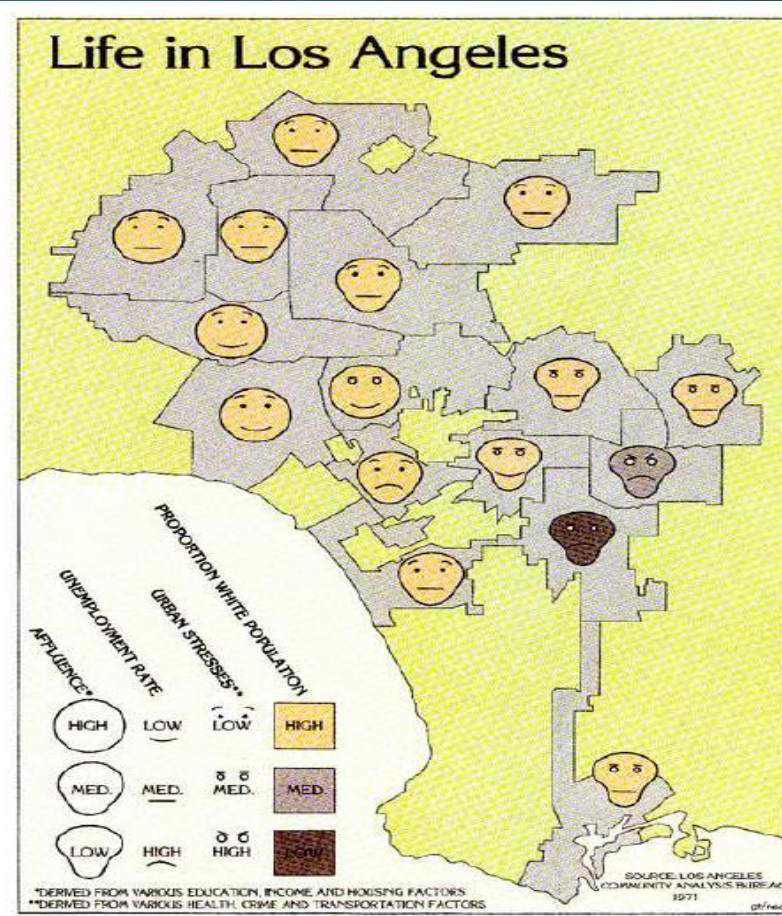
**CONCAVIDADE**

# Exemplo: População de Los Angeles

## Características

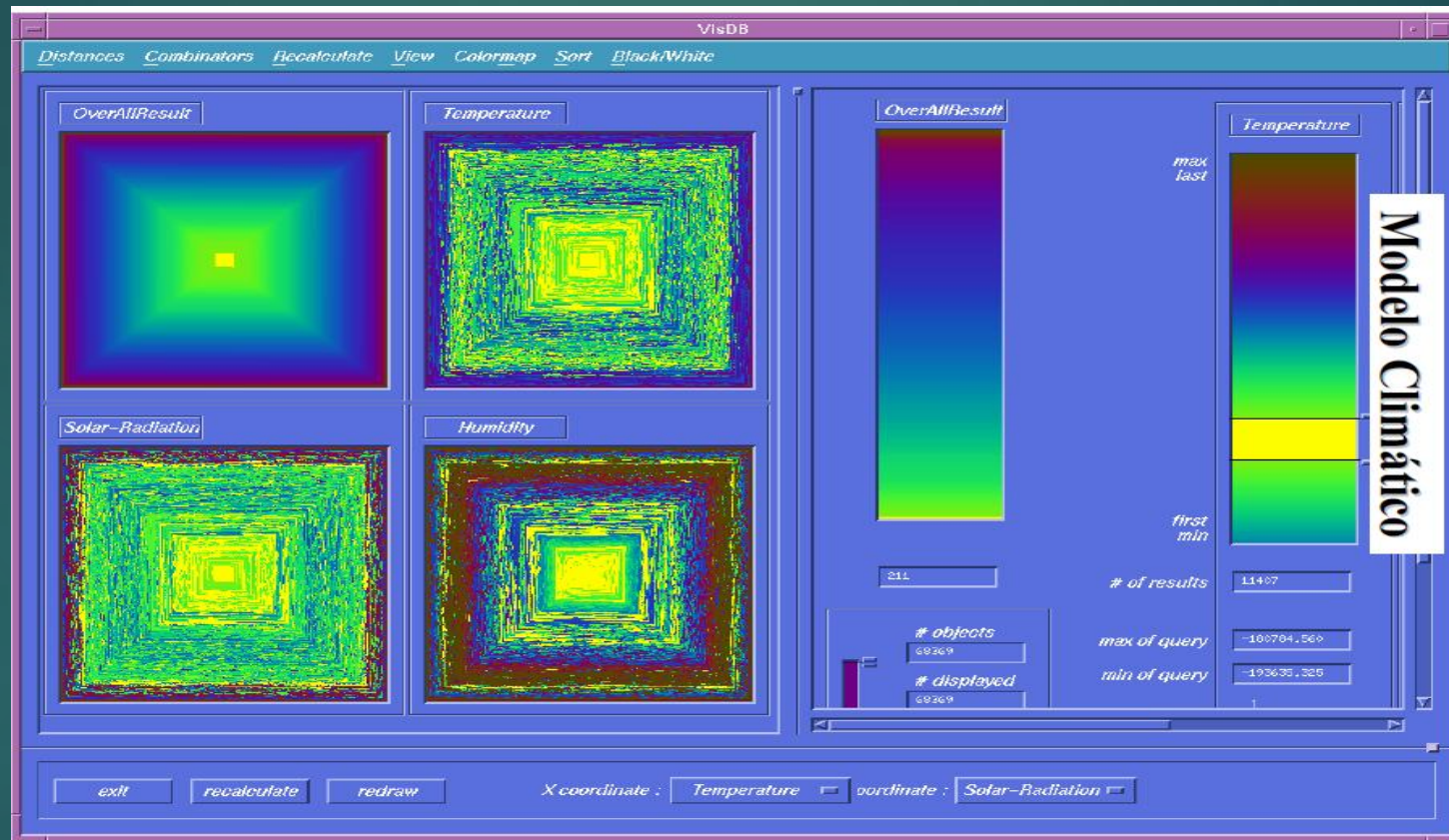
### Rostos de Chernoff (Che73, Tuf 83)

Propriedades das faces (formato do rosto, boca, olhos, nariz) são associados aos campos dos registros de dados

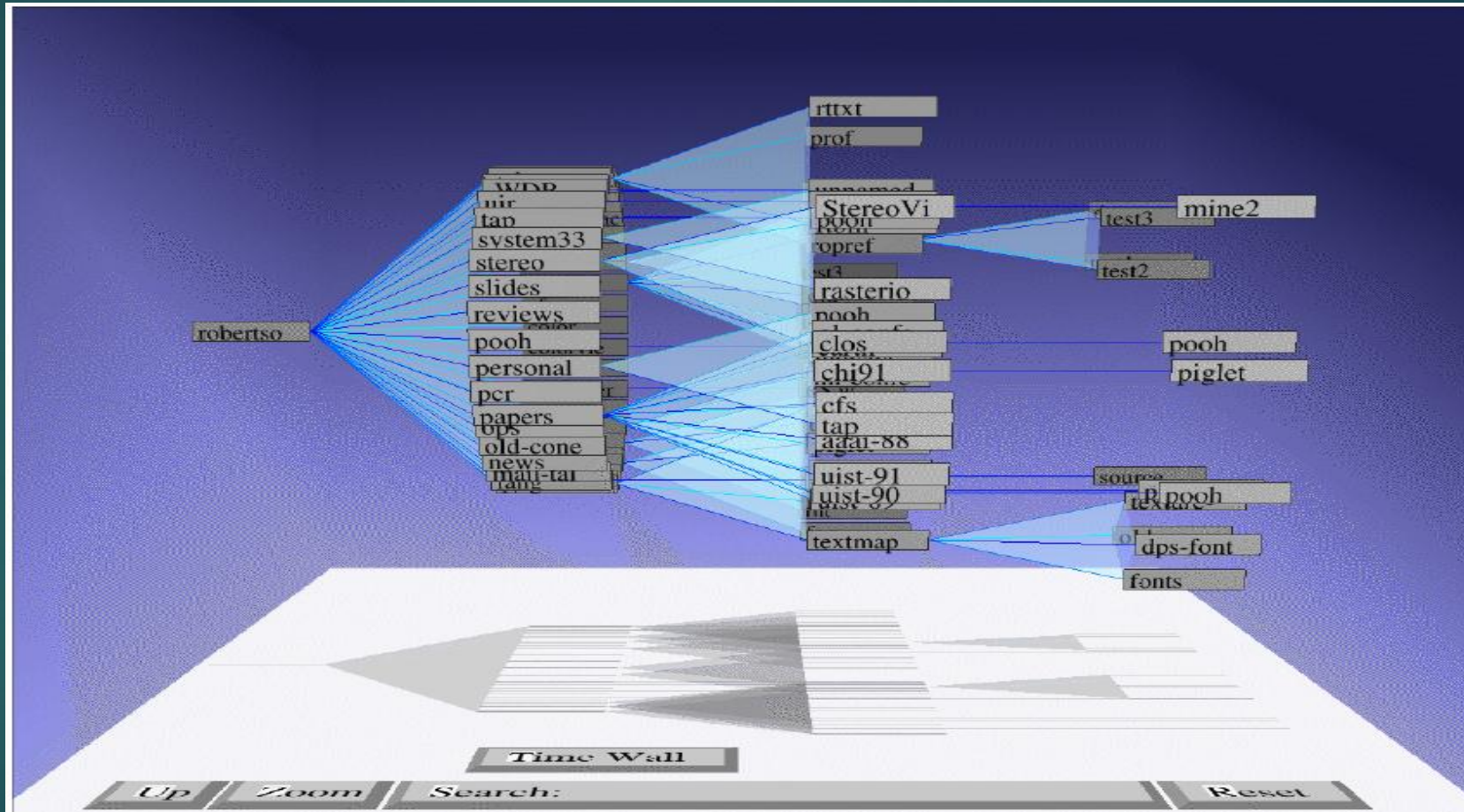




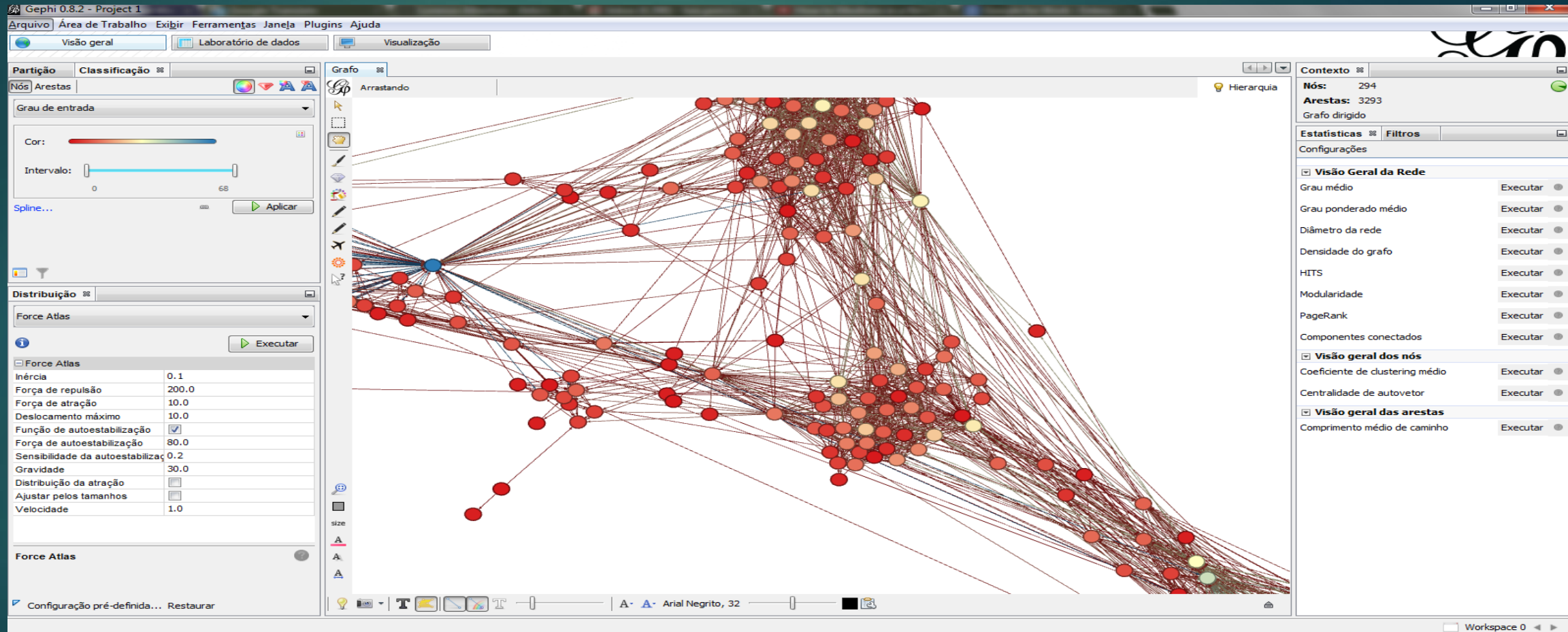
# Paradigma Visual Climático



# Visualização 3D - Hierárquica

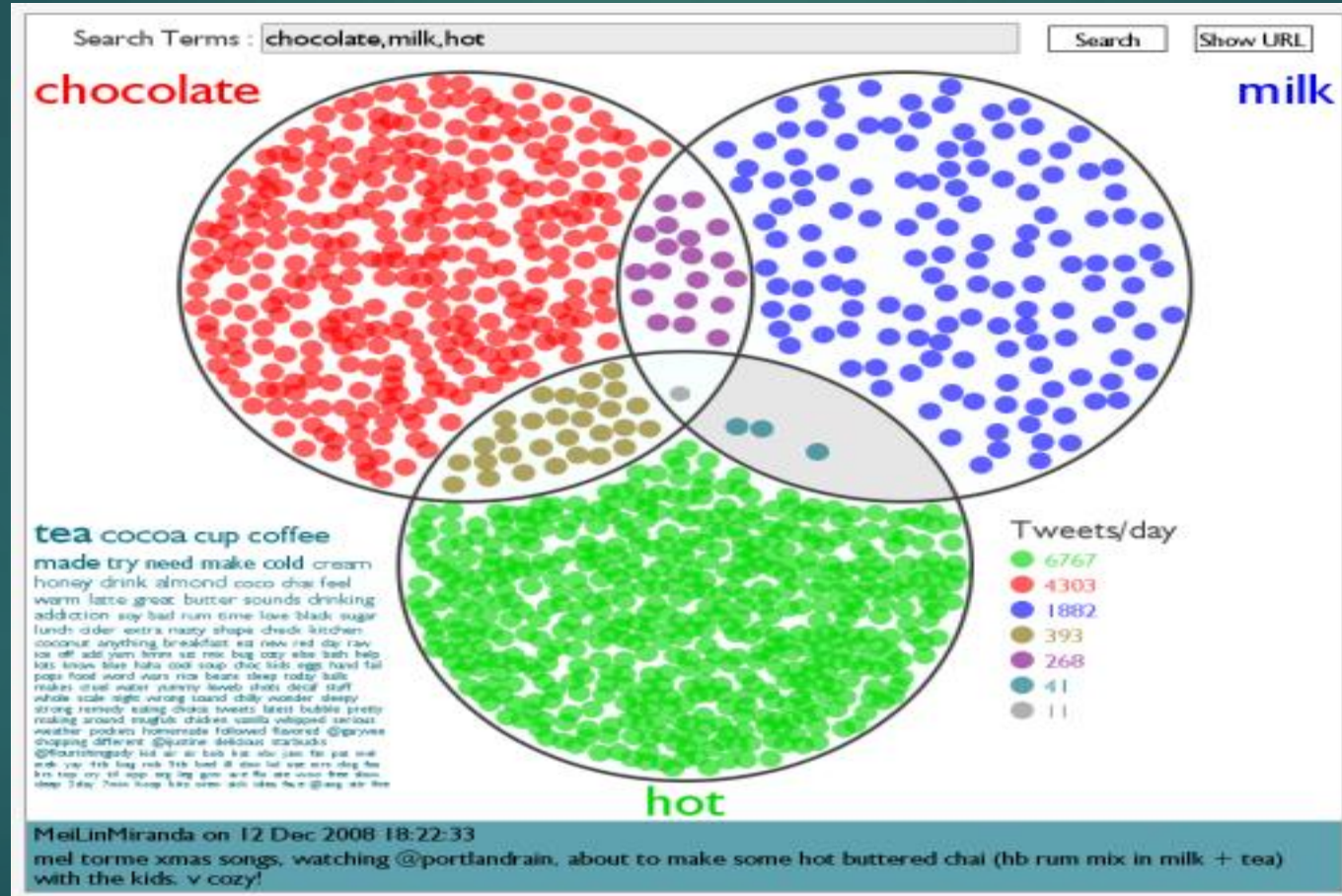


# Visualização Interação – Rede Relacionamento





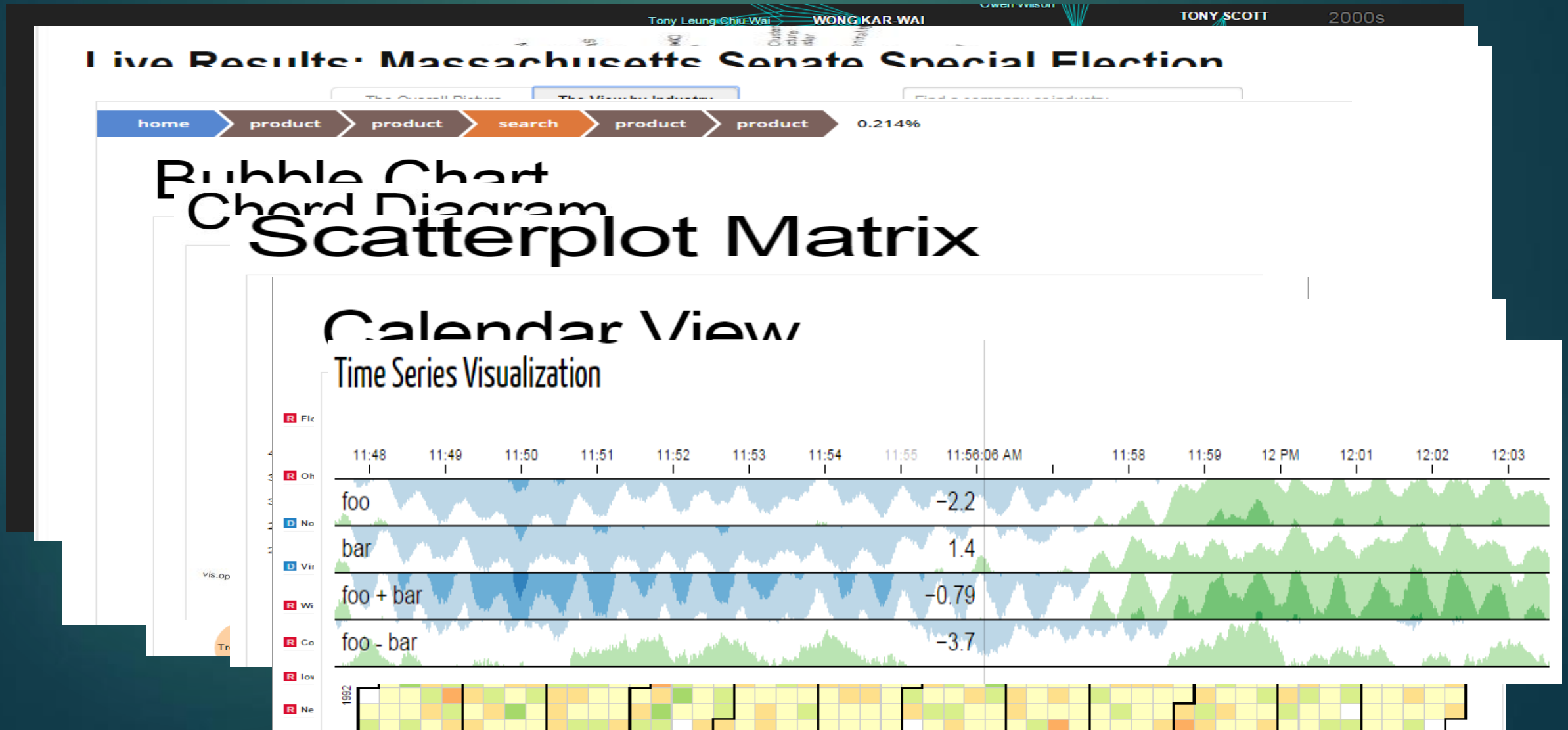
# Visualização Interação – Rede Relacionamento



*Neoformix*



# Mineração Visual – d3js





# Conclusões

- ▶ Data mining é um processo que permite compreender o comportamento dos dados.
- ▶ Data mining analisa os dados usando técnicas de aprendizagem para encontrar padrões e regularidades nestes conjuntos de dados.
- ▶ É um problema pluridisciplinar, envolve Inteligência Artificial, Estatística, Computação Gráfica, Banco de Dados.
- ▶ Pode ser bem aplicado em diversas áreas de negócios



# Projeto Completo...

**Finalidade:** Coleta de dados com os gestores para a construção do BI.

Fatos		Diária
Dimensões		
Hóspede		✓
Tipo Quarto		✓
Código Tipo Quarto		
Tipo Quarto	HISTÓRICO	
Classe Quarto		✓
Tempo (Data Registro Primeira Diária)		✓

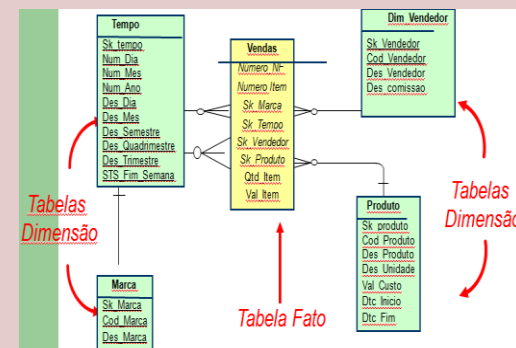
Tarefas de  
Mineração de  
Dados -Parte01

**Finalidade:** Levantamento dos relacionamentos e objetos que armazenam os dados da empresa.

DIMENSÕES	ORIGEM	
	TABELA/VISÃO	CAMPO
Hóspede		
Nome Hóspede	HOSPEDE	NOM_HOSPEDE
Cidade Hóspede	CIDADE_ORIGEM	NOM_CIDADE
País Hóspede	PAIS_ORIGEM	NOM_PAIS
Aeroporto Hóspede	AEROPORTO_SAIDA	DES_AEROPORTO
Local Aeroporto Saída	AEROPORTO_SAIDA	NOM_LOCALIDADE
Código Hóspede	HOSPEDE	COD_HOSPEDE

Tarefas de  
Mineração de  
Dados -Parte02

**Finalidade:** Modelo adequado para realizar as consultas nas bases que servirão ao BI



Mineração  
Visual

contato@bicomodeveser.com