

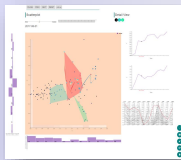
再配布禁止

# 情報の可視化

～情報科学概論 1～  
高間 康史（2回目講義）

1

## 高間研究で開発した可視化システム



視覚的データ  
分析



LOD 可視化



協調的可視化



モニタリング  
支援



2

# 可視化 (Visualization) とは

- 見えないものを見えるように
- 見えにくいものを見えやすく
- 人間の知覚能力の活用
- 計算機能力の向上による対象範囲の拡大
  - ◆ 処理対象データの大規模化
  - ◆ グラフィクス能力の向上
  - ◆ インタラクティブ性の向上

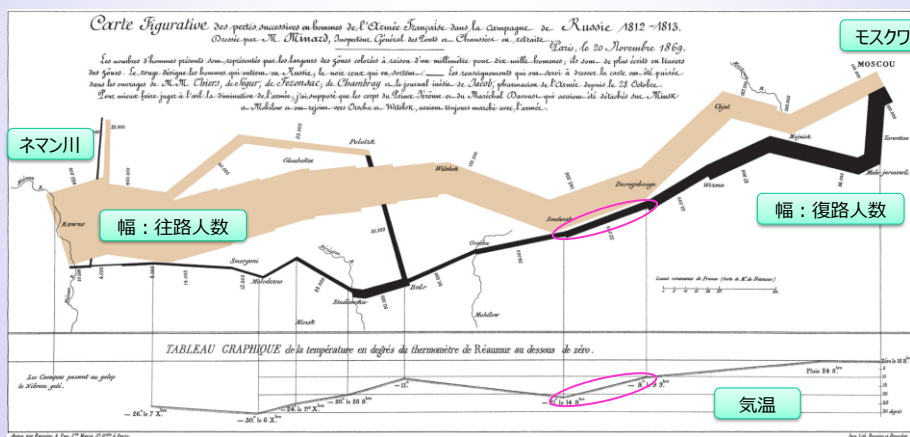
## 取り調べの可視化



3

## 可視化の例1：インフォグラフィック

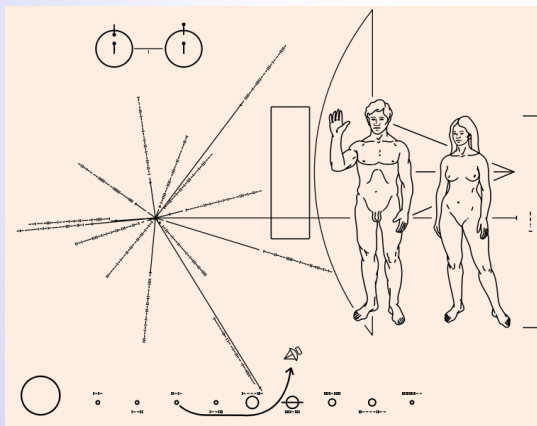
- 1869 by Charles Joseph Minard (フランス)
- ナポレオンのモスクワ遠征 (1812-13) の悲惨さを可視化
  - ◆ 軍団の進行方向, 通過場所, 兵数減少 (42万→1万), 気温



4

## 可視化の例2：インフォグラフィック

- 1972 パイオニア探査機に搭載された金属板
- 地球外生命体に理解してもらう事を意図



- 中性水素の超微細遷移  
(長さ, 時間の単位)
- 人間の身長  
(探査機との相対的サイズ)
- 太陽の相対的位置
- 探査機の太陽系における経路

5

## 可視化の効果：例題1

学生	数学	国語	英語
出雲太郎	85	80	90
首都大次郎	90	55	70
山田花子	55	70	60
東京一郎	50	30	40

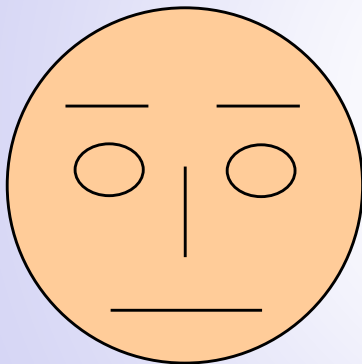
- 成績が良い／悪いのは誰？
- 全体的な傾向は？
- 科目数・学生数が増えると大変
- このデータを可視化すると. . .

6

# チャーノフの顔グラフ

■ 顔の各パーツに属性を割当て

- ◆ 眉毛：数学の得点
- ◆ 目：国語の得点
- ◆ 口：英語の得点

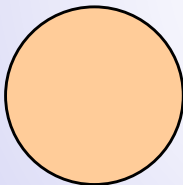


	○ (80-)	△ (60-79)	× (-59)
数学	↘ ↙	— —	↗ ↘
国語	○	○	—
英語	∨	—	∧

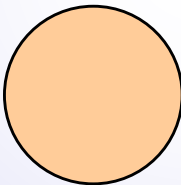
7

# 描いてみましょう

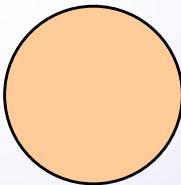
学生	数学	国語	英語
出雲太郎	○	○	○
首都大次郎	○	×	△
山田花子	×	△	△
東京一郎	×	×	×



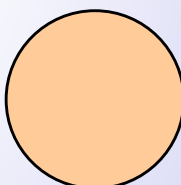
出雲太郎



都立大次郎



山田花子



東京一郎

8

## 可視化の効果：例題2

### ■ Iris plantsデータ

- ◆ アヤメの品種分類
- ◆ 特徴：4種類
  - がくの長さ(sepal length)
  - がくの幅(sepal width)
  - 花弁の長さ(petal length)
  - 花弁の幅(petal width)
- ◆ 種類：3種類
  - Iris setosa, iris versicolor, iris virginica

がく長	がく幅	花弁長	花弁幅	種類
5.1	3.5	1.4	0.2	Setosa
4.9	3.0	1.4	0.2	Setosa
7.0	3.2	4.7	1.4	Versicolor
6.4	3.2	4.5	1.5	Versicolor
5.9	3.0	5.1	1.8	Virginica
6.2	3.4	5.4	2.3	Virginica
...	...	...	...	...

### ■ データ数：150

- ◆ 各品種の特徴は？ ← どの特徴で分類可能？
- ◆ 特徴間の関係は？

9

## 散布図行列 (Scatter Matrix) による可視化

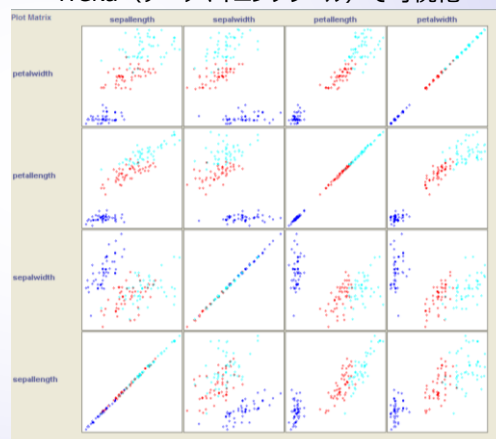
### ■ 散布図行列

- ◆ 散布図(Scatter plot)  
2変数について  
データ分布を可視化
- ◆ 多変数の全ての組合せに  
ついて散布図を描画

※色はクラスを表現

● Setosa  
● Versicolor  
● Virginica

Weka (データマイニングツール) で可視化



<http://www.cs.waikato.ac.nz/ml/weka/>

10

## 可視化の技術的展開

- データ解析 + コンピュータグラフィックス (CG)
- 科学・工学分野での利用
  - ◆ 気象学, 流体力学, 化学反応, 物理現象
  - ◆ 振動解析, 自動車設計, CAD
  - ◆ 医療分野: 器官の視覚化, CT画像処理
- より広い分野への展開
  - ◆ ビジネス分野: データマイニング, 統計解析, アンケート解析
  - ◆ 情報検索

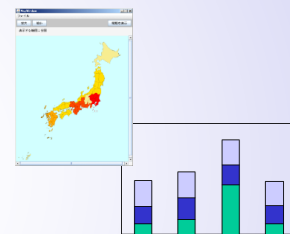


計算機の機能向上により適用範囲拡大

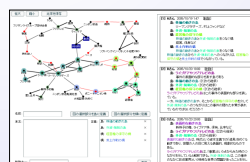
11

## 可視化対象の拡大

- Scientific Visualization
  - ◆ 流体力学
  - ◆ 地理データ (Geographic data)
- Information Visualization (情報可視化, InfoVis)
  - ◆ ビジネス応用 (データマイニング)
  - ◆ 情報検索システム
- 両者の違いは？



対象の  
拡大



12

# Scientific vs. Information Visualization

## ■ Scientific Visualization

- ◆ 本質的な構造（2,3次元）を持つデータが対象
- ◆ データが本来持つ構造をわかりやすく見せることが目的

How to visualize?

可視化技術  
自体は共通性あり

## ■ Information Visualization

- ◆ 抽象的なデータ
- ◆ 多次元データ
- ◆ 描画だけでなく、構造を与えることも必要

構造に関する  
自由度が高い

What to visualize?

How to visualize?

13

# 代表的な可視化技術

- 統計グラフ
- 平行座標（Parallel coordinates） [Inselberg01]
  - ◆ 多属性オブジェクトの可視化
- TreeMap
  - ◆ オブジェクト間関係（階層構造）の可視化 [Zhao05]
  - ◆ 空間的効率が高い可視化手法
- バネモデル
  - ◆ オブジェクト間関係（ネットワーク構造）の可視化
  - ◆ 多数の可視化ツールが存在

14

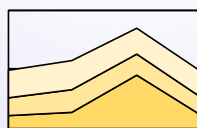
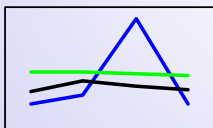
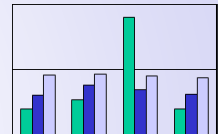
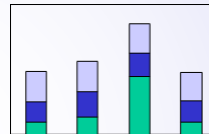
# 統計グラフ

- 可視化の基本
- 5種類に分類可能[上田05]
  - ◆ 棒グラフ系 (bar chart)
  - ◆ 線グラフ系 (line chart)
  - ◆ 帯グラフ系 (percentage bar chart)
  - ◆ 分布図系 (scatter plot)
  - ◆ レーダーチャート (distribution chart)

15

## 棒グラフ・線グラフ

- 棒グラフ系 (bar chart)
  - ◆ 大きさの比較
  - ◆ 組合せ, 積上げ (stacked), 統計地図 (statistical map)
- 線グラフ系 (line chart)
  - ◆ 間隔的, 比率的データの変化・傾向 (X:順番・間隔に意味, Y:差に意味)
  - ◆ 組合せ, (積上げ) 面グラフ (area chart)



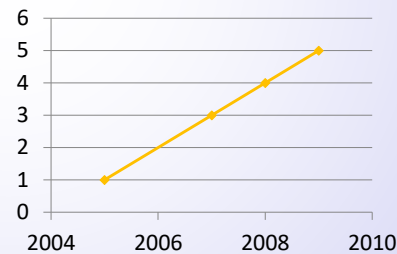
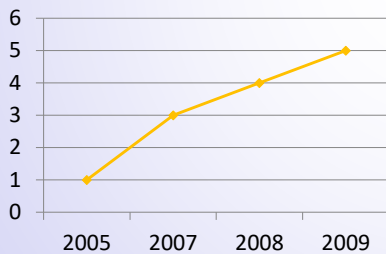
16



## 棒グラフ・線グラフの使い分け

### ■ X軸にマッピングする属性の違い

- ◆ 棒グラフ：順番に意味なし。名義的，順序的データも可
- ◆ 線グラフ：順番，間隔に意味  
→ 間隔的，比率的データである必要性



17

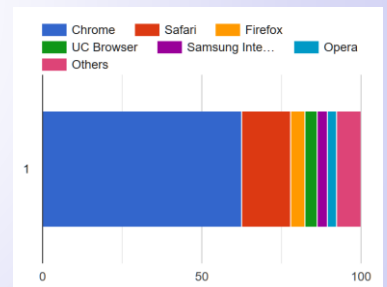
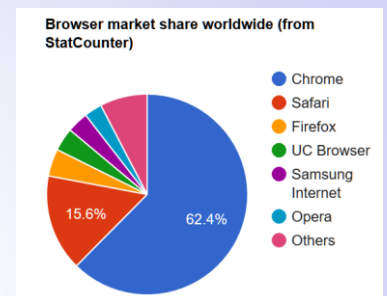
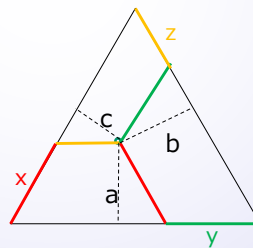
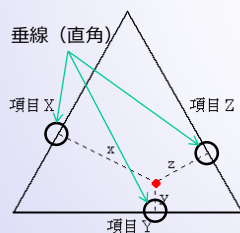
## 帯グラフ系

### ■ 属性毎の構成比率を可視化

- ◆ 円グラフ
- ◆ 帯グラフ（積上げ棒グラフの一種）

### ■ 三角グラフ

- ◆ データ (X,Y,Z):  $X+Y+Z=100\%$
- ◆ 二つのアプローチ



18

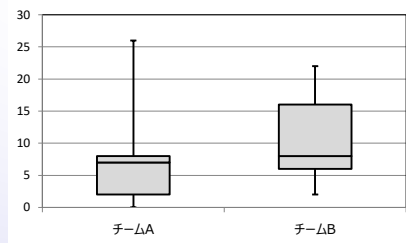
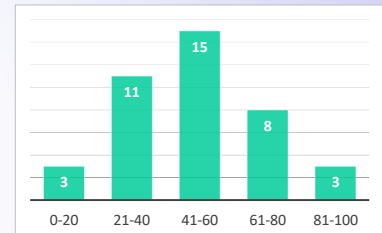
## 分布図系

### ■ データ（オブジェクト）の分布を可視化

- ◆ 散布図
- ◆ ヒストグラム

### ■ ボックスプロット

- ◆ n数要約, 箱ひげ図
- ◆ 分布の特徴を表す数値をグラフィカルに表示
- ◆ ヒストグラムの簡略的表示



19

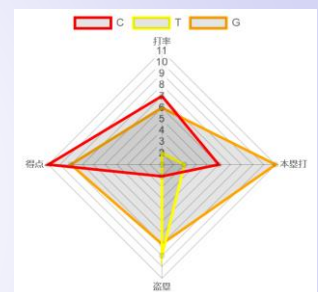
## レーダーチャート

### ■ 複数項目（属性）の比較

- ◆ 同一系列（オブジェクト）のデータ値を線で結ぶ
- ◆ 重ねて描画することにより比較

### ■ 他の可視化手法との関連

- ◆ Geometric projection techniques :  
軸を直交・平行に配置←レーダーチャートでは放射線状
- ◆ Iconographical techniques :  
属性を幾何オブジェクトのパラメータに写像  
…（属性, パラメータ）の組合せを変えると異なる結果  
←レーダーチャートではマッピングの違いによる変化は少ない



20

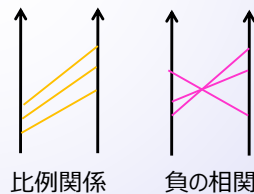
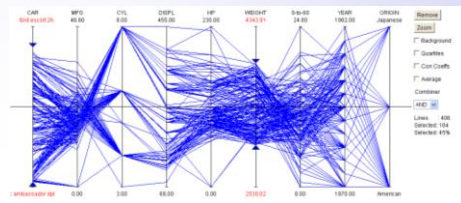
## 多次元データの可視化

- 各属性を次元（軸）に割当て
- 直交座標系：各軸が直交
  - ◆ 2属性 = 平面
  - ◆ 3属性 = 空間
  - ◆ 4属性以上…（同時に）描画は困難
- 多属性（ $\geq 4$ ）の場合の対策
  - ◆ 他の可視化属性の利用：色，サイズ，etc. → 属性毎の印象の違い
  - ◆ 次元の削減・選択 cf. 散布図行列
  - ◆ 次元の縮退：低次元空間への写像
  - ◆ 平行座標

21

## 平行座標 (Parallel Coordinates) [Inselberg01]

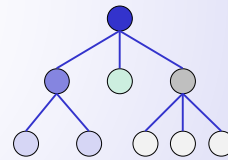
- 軸を平行に並べる
  - ◆ 多次元データに強い
  - ◆ 全属性を平等に扱うことが可能
- 属性間の関係を読み取り可能
  - ◆ 比例関係
  - ◆ 負の相関
- 注意点
  - ◆ 読み取りに慣れが必要
  - ◆ 軸（属性）の順序が重要



22

# 木構造の可視化

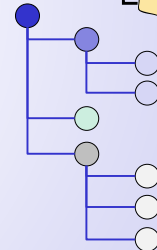
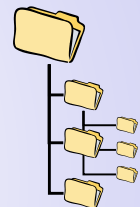
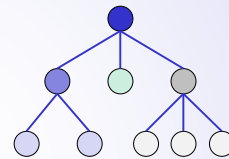
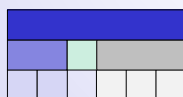
- オブジェクト間関係の可視化
  - ◆ 親子（上下）関係，兄弟関係
- よく見られる構造
  - ◆ 概念体系，シソーラス
  - ◆ フォルダ構造
  - ◆ （階層的）クラスタリング結果
  - ◆ ソフトウェア（モジュール，クラス間関係）
- 基本的構造
  - ◆ ノード（頂点，接点）とエッジ（リンク，枝）
  - ◆ ルート：一番上のノード
  - ◆ リーフ：末端のノード



23

## Node-Link diagram

- 木構造の最も一般的な可視化方法
  - ◆ 理解しやすい
  - ◆ 下層ほど見にくい
- 類似形式：
  - ◆ Indented outline :  
フォルダ構造の可視化に多い
  - ◆ Use of alignment & adjacency



(<https://observablehq.com/@d3/sunburst>)

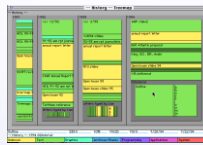
24

# Treemap

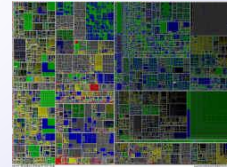
[Shneiderman91,Rekimoto93,Ito04]

- 「外側に展開するタイプの可視化」の問題点
  - ◆ 下の階層ほど煩雑 → 空間的効率性が低い
- Treemap: 入れ子構造による階層構造の可視化
  - スケーラビリティに優れる（数千規模も可能）
  - × 構造把握が困難
  - オーバビュー提示に適する

バリエーションあり  
 ・領域分割の変更  
 （Voronoi Treemaps）  
 [Balzer05]  
 ・配置の変更  
 （Radial Treemaps）

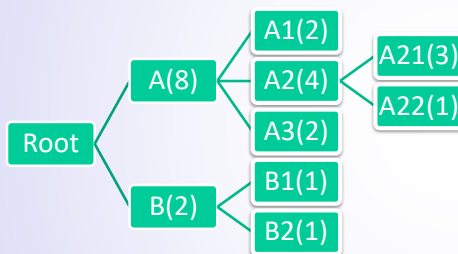


<http://www.cs.umd.edu/hcil/treemap-history/>



25

## Treemapの基本 ～Slice & Dice～



描画領域



26

## ネットワーク構造の可視化

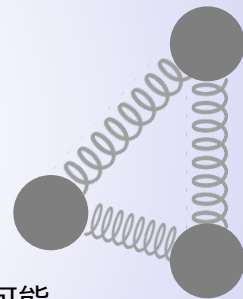
- ネットワーク構造の例
  - ◆ Web: ハイパーリンク構造
  - ◆ ソフトウェア（モジュール, クラス間関係）
  - ◆ ソーシャルネットワーク（人間関係）
  - ◆ 遺伝子ネットワーク（遺伝子間の相互作用）
  - ◆ 神経回路網
  - ◆ 通信網
- 可視化対象として近年有名
  - ◆ 複雑ネットワーク：現実世界の多様なネットワークの可視化・分析
  - ◆ ネットワーク可視化ツールの存在

27

## バネモデル

(Spring model / spring layout / spring embedded)

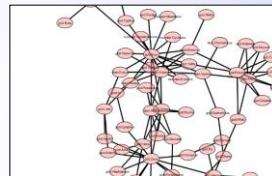
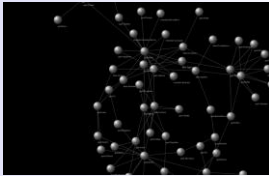
- 代表的な多数オブジェクト配置手法
  - ◆ 絶対位置ではなく, 相対位置に意味
  - ◆ 多数オブジェクト間の関係 ( $R$ ) を2,3次元に写像
- ノード間をバネで接続
  - ◆ バネ長より近い（遠い）  
→ 遠ざかる（近づく）方向に力
  - ◆ 弾性力エネルギー極小の配置を計算
  - ◆ 反復計算 → アニメーション, リアルタイム配置に利用可能
  - ◆ 他の力との組合せ → occlusion回避（分子間力）, 整列



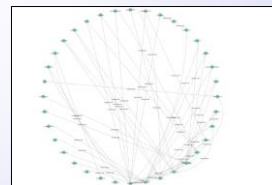
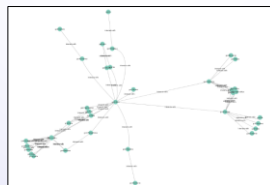
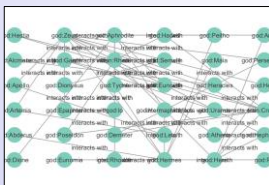
28

# Cytoscapeの可視化例

## スタイルの例



## レイアウトの例



グリッド（格子）

ばねモデル

サークル（円周）

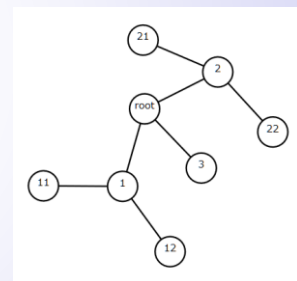
29

## 様々な力の利用



### ■ D3.js- Force layoutの場合

- ◆ ノード間距離（ばね長）
- ◆ ノード間引力（反発力）
- ◆ 重力：特定場所（中心など）に向かう力
- ◆ 摩擦力：移動方向と逆向きに作用。



<https://codepen.io/ytakama/pen/MxLvGe>

49



## 可視化の対象タスク

### ■ プレゼンテーション

- ◆ インフォグラフィクス
- ◆ データの提示
- ◆ 支援目的：Storytelling（データ閲覧・解釈の誘導）

### ■ 探索的データ分析

- ◆ EDA (Exploratory Data Analysis)
- ◆ データから知見・仮説を得る
- ◆ 支援目的：多様な観点からの分析・探索空間の拡大

### ■ モニタリング

- ◆ 日々発生するデータの効率的な確認
- ◆ 対象データ：オンラインニュース，2チャンネル，Twitter，etc.



31

## 探索的データ分析と可視化

### ■ 可視化＝探索的プロセス[Chen09,10]

- ◆ 満足するviewを見つける作業
- ◆ 膨大な探索空間をインタラクティブに

### ■ 視覚的分析（Visual analytics）

- ◆ 人間をデータ探索プロセスに組込む→知覚能力の活用
- ◆ Viewを介して人間とデータがインタラクション

### ■ 分析の目的（人間の役割）

- ◆ Insight: 洞察、知見の獲得
- ◆ Sense making : 「気づき」に対する解釈・意味づけ

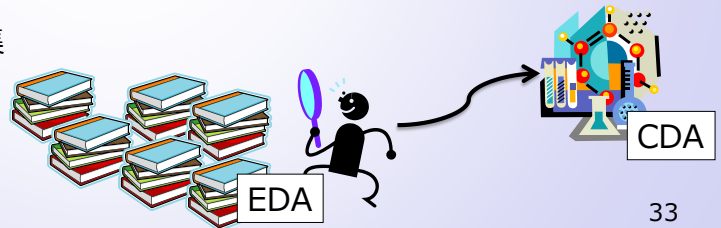
究極の役割：  
意思決定

32



## 探索的データ分析と検証的データ分析

- 探索的データ分析  
(Exploratory Data Analysis, EDA) [Tukey77]
  - ◆ 特定の前提を置くことを避けて分析を進める
  - ◆ データが先に存在 (e.g. データマイニング)
- ◆ 検証的データ分析  
(Confirmatory Data Analysis, CDA)
  - ◆ 数理統計. 仮説の検証
  - ◆ 分析のためにデータを収集



33

## 探索的データ分析支援のアプローチ

- インタラクションの導入
  - ◆ 効率から試行錯誤へ
  - ◆ 操作の結果を即座に反映
  - ◆ 「後戻り」を可能に→失敗の許容
- 比較作業を支援する
- 協調作業を支援する
  - ◆ 多様な観点・知識の導入→探索空間の拡大
  - ◆ 三人寄れば文殊の知恵

34

## Visual Analyticsのプロセス

- 分析者が行うべき作業
  - ◆ ビューから興味ある部分, 特徴的部分に気づく
  - ◆ 特異値, 変動点など
  - ◆ 気づいた点の背後にある理由を考察
  - ◆ 解釈, センスメイキング (sense making)
  - ◆ 気づきを積み上げ, 知見を得る
  - ◆ 洞察 (insight) , 包括的理解, アクションへ

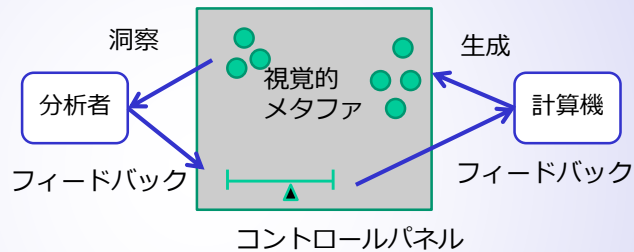
35

## Visual Analyticsの課題

- ドメイン専門家 ≠ AIの専門家
  - ◆ 学習モデル・(可視化手法) に対する知識なし
  - ◆ 結果の解釈困難
  - ◆ パラメータ調整困難
- 対策
  - ◆ 専門家が理解しやすいモデルの利用
    - Prophet : 専門家が理解しやすい時系列データのモデル [Taylor17]
  - ◆ Semantic Interaction [Endert12]

36

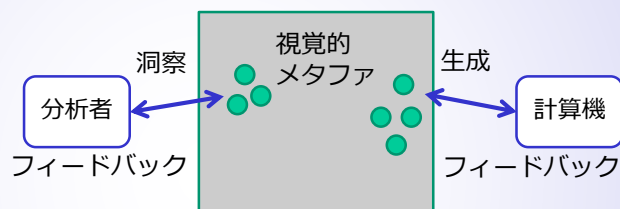
# Visual Analyticsにおけるインタラクションの課題



- 分析者：可視化から洞察
- 計算モデルに対する知識が必要
  - ◆ コントロールパネル：パラメータを直接調整

37

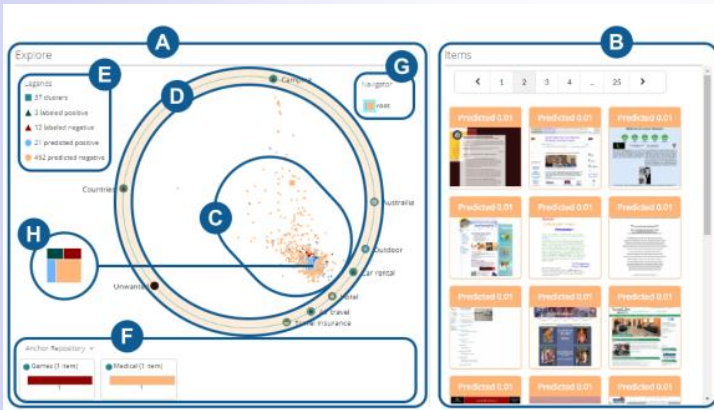
## Semantic Interaction



- ドメインの専門家を支援するためのインタラクション
- 操作の意味を解釈 → 計算モデルへのフィードバック
  - ◆ モデル更新結果は視覚的に逐次反映
  - ◆ モデル・パラメータの複雑さから分析者を保護
- 視覚的メタファの活用
  - ◆ 近い＝類似, 場所に意味, 直感的操作：移動, 重ねる, etc.

38

# AnchorVis [Chen18]



[Chen18] N. Chen et al., AnchorVis: Facilitating Classifier Error Discovery through Interactive Semantic Data Exploration, IUI2018, pp. 269-280, 2018.

- Interactive Machine Learning
  - ◆ Human in the loop
- 誤分類を発見, 正例・負例の指示
- 磁石のメタファ
  - ◆ D: アンカー : 類似データを惹きつける役割
  - ◆ 正例, 負例のグループを作成

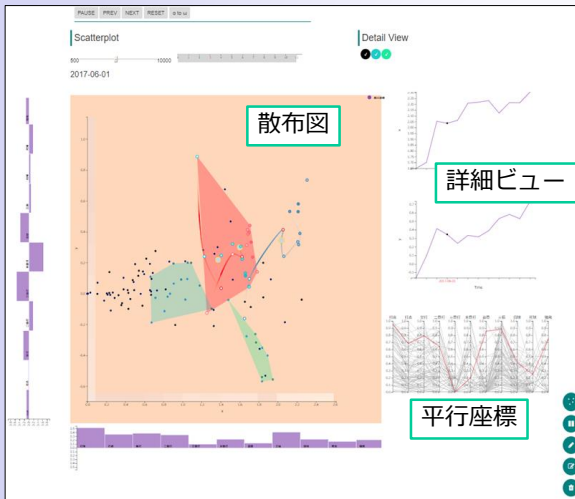
39

## 指標形成

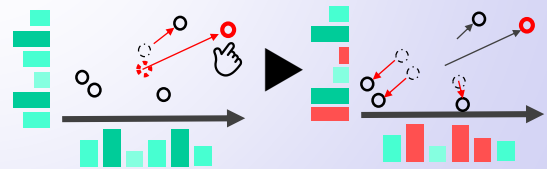
- 指標 : 評価の目安 (ものさし)
  - ◆ データからの獲得 : 統計学的根拠
  - ◆ 例 : セイバーメトリクス
- 視覚的分析のアウトプット
  - ◆ 知見・知識の一形態
  - ◆ データ分析 + 専門家の知識
- 分類モデルとの関係
  - ◆ 共通点 : 未知データの評価, 将来予測
  - ◆ 相違点 : 解釈可能性

40

## 開発したシステム



- 多次元時系列データの分析
- 軌跡表現：時間変化→空間的表現
  - ◆ 時間・空間軸衝突の回避 [Takami18]
- 散布図のXY軸：属性の線形和
  - ◆ PCAで計算
  - ◆ ユーザフィードバックを反映



41

## コラボレーション ～ソーシャルなデータ分析

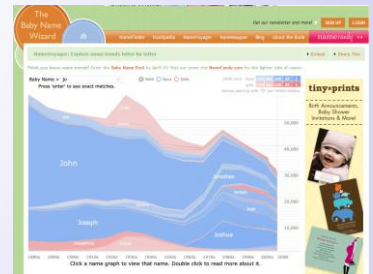
- 協調作業としてのデータ分析
  - ◆ 個人で大規模データ空間を探索することは困難
  - ◆ データ空間の深い探索・理解
- 協調的情報可視化[Heer07]
  - ◆ 可視化による センスメイキングの支援
  - ◆ 議論・探索のコンテキストを可視化
- 課題
  - ◆ コンテキスト共有の困難さ→どのビューを見ているか？
  - ◆ 同じ議論・質問が繰り返される
  - ◆ 議論の把握とコントロール（発散→収束）



42

# NameVoyager[Wattenberg06]

- 赤ちゃんの名前のトレンド（@米国）
  - ◆ サイト公開から2週間で50万visit達成, その後もコンスタントにアクセス
  - ◆ Stacked graph : X=年, Y=頻度. 男女で色分け
  - ◆ Dynamic query : アルファベットタイプでフィルタリング
- コラボレーションの発生
  - ◆ Blog, discussion forumでの引用
    - 画像ではなくキータイプ
  - ◆ 発見を他者に伝える, 質問 → 発見の共有
  - ◆ 他者からのコメント, 回答 → 意見交換



<https://www.babynamewizard.com/voyager>

43

## コラボレーションはどこで発生？

- NameVoyagerは可視化（グラフ）のみ
  - ◆ コメント機能, 掲示板機能なし
- Blog, discussion forumでの引用
  - ◆ 画像ではなくキータイプを引用：「aとタイプしてみて」
  - ◆ Dynamic query: 入力文字列でグラフが決定
- 多様な議論の発生
  - ◆ 気づく人, 解釈を与える人
  - ◆ データ空間の十分な探索, 深い理解につながる現象を観測
- より積極的に支援するには？

44

## ソーシャルなデータ分析

- 協調作業としてのデータ分析
  - ◆ 個人で大規模データ空間を探索することは困難
  - ◆ データ空間の深い探索・理解
- 協調的情報可視化[Heer07]
  - ◆ 可視化による sense-making の支援
  - ◆ 議論・探索のコンテキストを可視化
- 課題
  - ◆ コンテキスト共有の困難さ→どのビューを見ているか？
  - ◆ 同じ議論・質問が繰り返される
  - ◆ 議論の把握とコントロール（発散→収束）

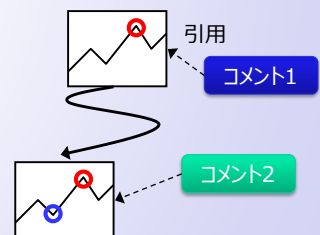
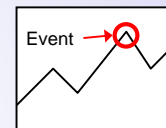


45

## ソーシャルなデータ分析の支援機能

- グラフィカル・アノテーション
  - ◆ ビューに補足コメント・マークを追加
  - ◆ 補足コメント：変動の原因となったイベントなど
  - ◆ 着目点：どこを見てコメントしているか？
  - ◆ （コメント中で引用）
- Visualization bookmark [Heer07][Viegas07]
  - ◆ ビューの状態にURLを割当て
  - ◆ （状態 = ビュー + アノテーション）
  - ◆ コメントをそれが書かれた時の状態にリンク
- コメントリスティング
  - ◆ ビューを引用するコメントの一覧

補足コメント

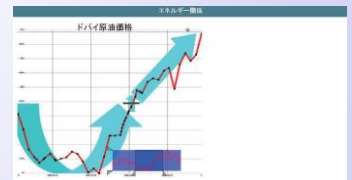
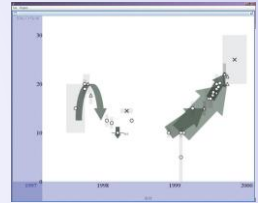


46



## アノテーションの例

- MuST（動向情報の要約と可視化に関するWS）
  - ◆ 新聞記事からの動向情報抽出
  - ◆ 基本：NLPによる統計データ（数値情報）の抽出・提示
  - ◆ +  $\alpha$ ：解釈に有効な情報の抽出・提示
- 動向情報と要因の可視化[山本06]
  - ◆ 変動要因に関する情報  
e.g. 内閣支持率…政治的イベント（政治家の発言など）
- STEND[松下06]
  - ◆ グラフ概形に関する情報（定性表現）  
e.g. 「下半期は安定傾向だった」、「7月をピークに下落」

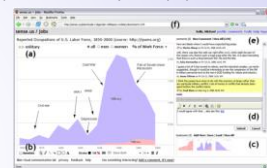


47

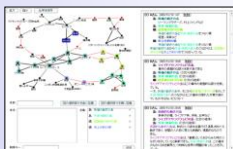
## ソーシャルなデータ分析を支援するシステム

- データ・Viewの共有
- コミュニケーション支援
  - ◆ アノテーション機能[Heer07]
  - ◆ コメント内での引用  
[Viegas07, Takama07, 高間09]

<http://vis.berkeley.edu/papers/sense.us/>



Sense.us [Heer07]



KGBBS  
[Takama07]



オンライン防犯マップシステム  
[高間09]



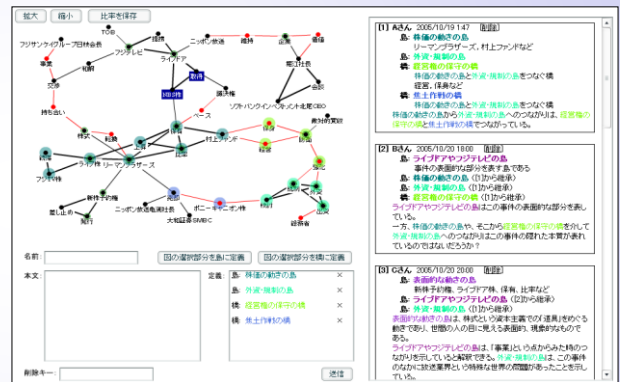
Many Eyes [Viegas07]

48



# KGBBS: BBSと可視化の融合による オンライン議論支援

- オンライン議論
- ○場所・時間の制約低減
- ×コンテキストの共有
- 可視化表現による  
コンテキスト共有



KeyGraph … チャンス発見で代表的な可視化手法

- 「島」と「橋」による解釈

KGBBS … スレッド毎に可視化表現を共有

- 「島」と「橋」の引用機能, 「島」「橋」に基づくシナリオ検索

49

## 地域安全マップ作成支援

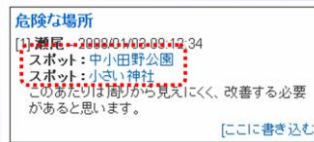
- マップ作成を通じ, 危険な場所を判断する能力を養う
- マップ作成の支援: 写真・アイコンの貼り付け
- マップ活用の支援: 防犯に関する議論支援



50

## 地域安全マップに基づく議論支援

- 写真・アイコン = アノテーション
- スポットを引用したコメント投稿 = ビューへのコメント
- スポットに対するコメント検索 = コメントリスティング



51

## おわりに：情報環境の進歩がもたらすもの

- (従来) 向上した計算機能力の活用
  - ◆ より高速に, 大規模に, 小型に, . . .
  - ◆ 人間が情報活用のボトルネックに
- 解決策：人間と計算機の「役割分担」の見直し
  - ◆ 人工知能：知的処理の代行
  - ◆ **可視化**：ボトルネックの解消



52