

# InfoVis Systems & Toolkits



CS 7450 - Information Visualization  
February 15, 2011  
John Stasko

## Background



- In previous classes, we have examined different techniques for presenting multivariate data
  - We'll continue to show more later too
- Today we look at systems that implement these ideas and provide some of their own new visualization techniques

# Agenda



- Systems providing a view or views
  - Improvise, ILOG Discovery, Many Eyes, Polaris
- Toolkits that can be used to build systems
  - InfoVis Toolkit, prefuse, Piccolo, protovis
- Commercial systems (to come in a future class)
  - Spotfire, Inxight, InfoZoom, InfoScope, Tableau, ...

# Systems/Tools



- Primarily examining academic systems that provide preexisting views
  - Commercial systems next time

# Improvise



- System used to build visualization tool
- Modular library of visualization components
- Multiple views key
- Coordination model coupled with data processing language at its heart

Weaver  
InfoVis '04

Spring 2011

CS 7450

5

# Improvise



- Visualization authoring
  - Put together combinations of well-known views and coordinations
  - Designed for people to build these combinations, but that process is non-trivial (Chris admits)

Spring 2011

CS 7450

6

# Key Components



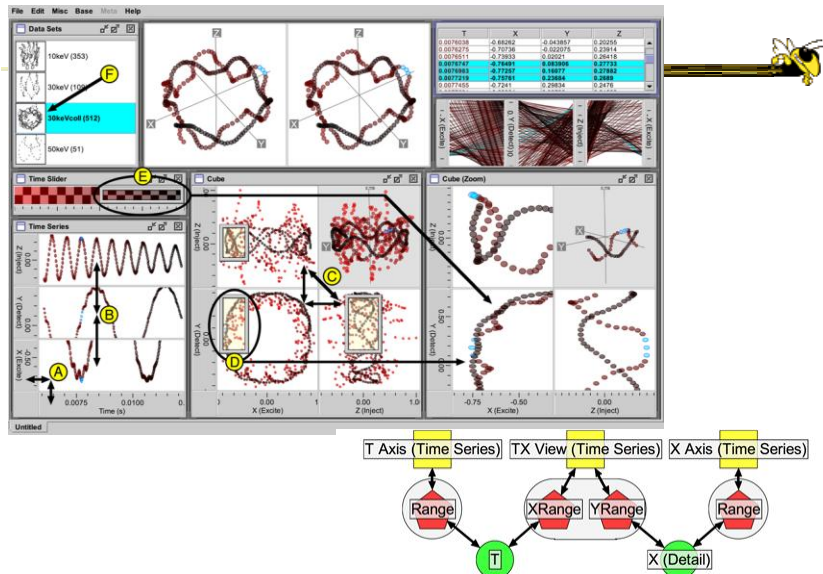
- Live Properties
  - For coordinating controls (eg, views, sliders)
  - Control defines live properties that bind to variables
- Coordinated Queries
  - Visual abstraction language
  - Visually encode data records into graphical attributes

Spring 2011

CS 7450

7

<http://www.cs.ou.edu/~weaver/improvise/>



Spring 2011

CS 7450

8

# Application

Weaver et al  
*Information Visualization '07*



- Understand spatial and temporal patterns of movements of people and groups of people, understand commerce and travel patterns
- Focus: Social networks of people and historical data
- Alternate: Understand actions and movements of people in terrorist groups

Spring 2011

CS 7450

9

# Particulars

- Who – Historical geographers, geneologists
- Problem – Provide platform where these people can do research and more easily understand the movement patterns
- Data – Old hotel records & registers



Figure 1 The National Hotel in Roaring Spring, PA.

Spring 2011

CS 7450

10

## Data Collection

---



Figure 2 The Rebersburg Hotel Register.

- Two hotel registers in PA
  - ~10k guest entries
  - Hundreds of place of origin (mostly NE)
- Transcribed into spreadsheets (400 hours)
- Many repeat visitors
  
- 6% of dataset not legible/usable

## Preliminary Data Analysis

---



- Simple tables generated
- GIS software for mapping geographic locations

# Visualization

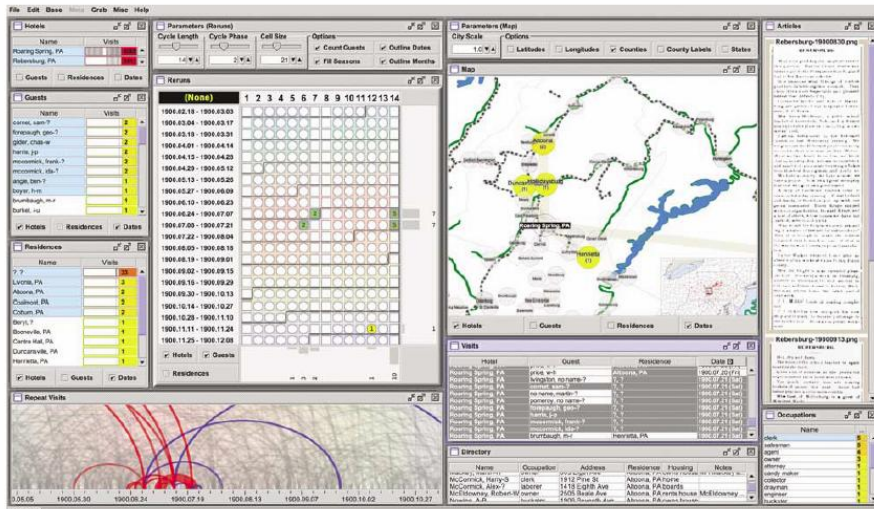


- How?
  - Consult with historical geographers
  - Iterative design
- Multiple, coordinated views

Spring 2011

CS 7450

13

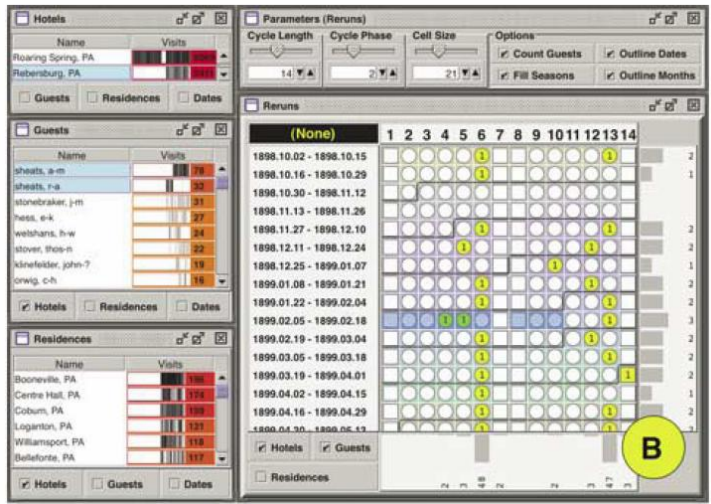


User interface

Spring 2011

CS 7450

14

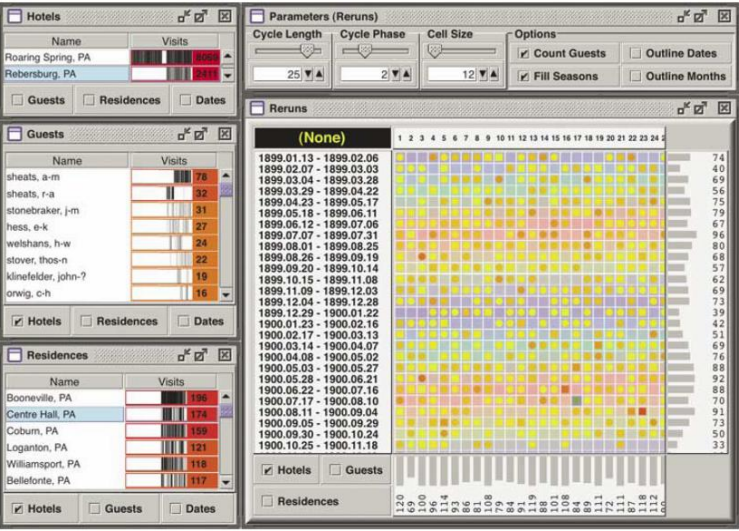


Looking for patterns of visitation

Spring 2011

CS 7450

15



Showing time of year of visits

Spring 2011

CS 7450

16



# Video



Spring 2011

CS 7450

17

# Data Analysis



- Through interaction with views, analysts can pose queries and look for answers
- “For these people on this date, did they ever meet prior to that?”
- Can also ask questions for unknown people (match a ?)

Spring 2011

CS 7450

18

# Data Attributes Explored



- Weekly circuits
- Weather effects
- Biweekly circuits
- Particular hotel characteristics
- Groups & gatherings
- Non-periodic regular visits

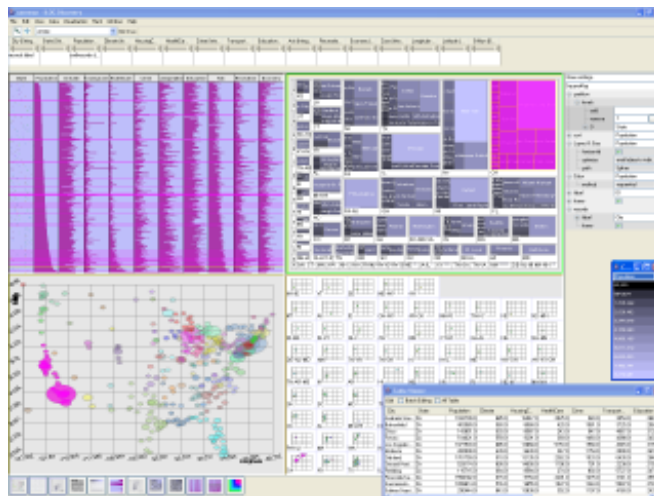
Spring 2011

CS 7450

19

<http://www2.ilog.com/preview/Discovery/>

# ILOG Discovery



Spring 2011

CS 7450

20

# Characteristics



- Large collection of views including scatterplots, parallel coordinates, treemaps, heatmaps, bar charts, ...

Spring 2011

CS 7450

21

<http://www.graphics.stanford.edu/projects//polaris/>

## Polaris

<http://www.tableausoftware.com>



**Database Schema:** The user drags fields from the database schema to shelves to define the visual specification.

**Layer Tabs:** Each layer has its own tab; different transformations and mappings can be specified for each layer.

**Axis Shelves:** The fields placed here determine the structure of the table and the types of graphs in each table pane.

**Context Menu:** The context menu provides access to the data transformation and interaction capabilities of Polaris such as sorting, filtering, and aggregation.

**Layer Shelf:** The fields placed here determine how records are partitioned into layers.

**Grouping and Sorting Shelves:** The fields placed here determine how records are grouped and sorted within the table panes.

**Mark Pull-downs:** Relations in each pane are mapped to marks of the selected type.

**Retinal Property Shelves:** The fields placed here determine how data is encoded in the retinal properties of the marks.

**Legends:** Legends enable the user to see and modify the mappings from data to retinal properties.

Video

Stolte et al  
TVCG '02

Spring 2011

CS 7450

22

# Basis



- Relational databases
- Pivot tables from spreadsheets
- N-dimensional data cubes
  
- Analytic approach is fundamental
- Provides visual representations of these concepts

Spring 2011

CS 7450

23

# Visualization



- Table of data (rows, columns)
- Each axis may have nested dimensions
- Table entry is a pane, and has visual marks to represent data
  
- Analyst drags and drops fields from database schema onto shelves of display
- Much interaction supported

Spring 2011

CS 7450

24

# Graphics



- Formal table algebra provided to describe data
- Visual mappings established from data types to appropriate (good) markings and encodings

# Interaction



- Four main types
  - Deriving additional fields
  - Sorting and filtering
  - Brushing and tooltips
  - Undo and redo

# Many Eyes



- InfoVis on the web
- Website developed from IBM's infovis group
- Motivating infovis challenges:
  - Difficulty of creating new visualizations
  - How do you discuss the visualizations?
- Project goals:
  - Enable end-user creation of visualizations
  - Foster large-scale collaborative usage

Viégas et al  
*TVCG '07*  
27

Spring 2011

CS 7450

# Features

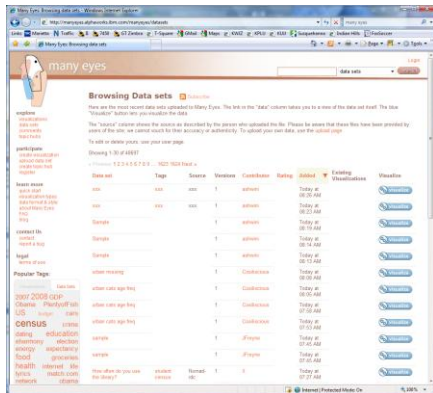


- Provides data upload capabilities and choice from library of visualizations
- Includes
  - Gallery of recently uploaded visualizations for browsing
  - Chosen highlighted visualizations
  - Attached discussion forums for each vis

Spring 2011

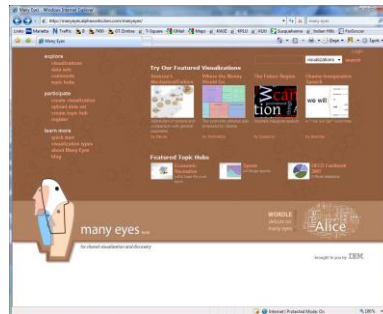
CS 7450

28



Recent gallery

## Featured visualizations



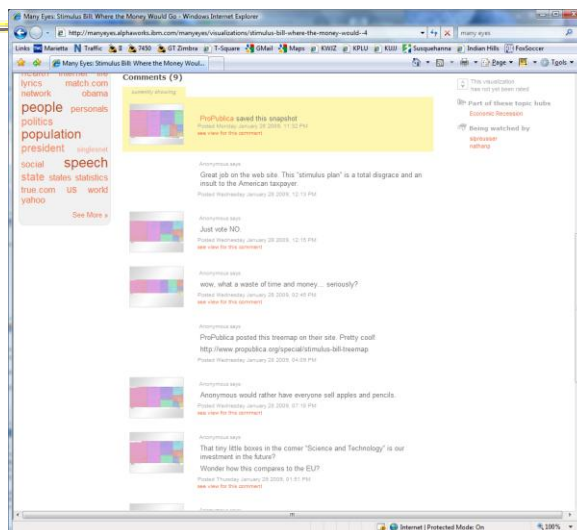
Spring 2011

CS 7450

29



## Discussion forums



Spring 2011

CS 7450

30

# Data



- Users upload their own data sets
  - All become public
- Format: table or unstructured text
  - Metadata allowed
- Immutable once uploaded

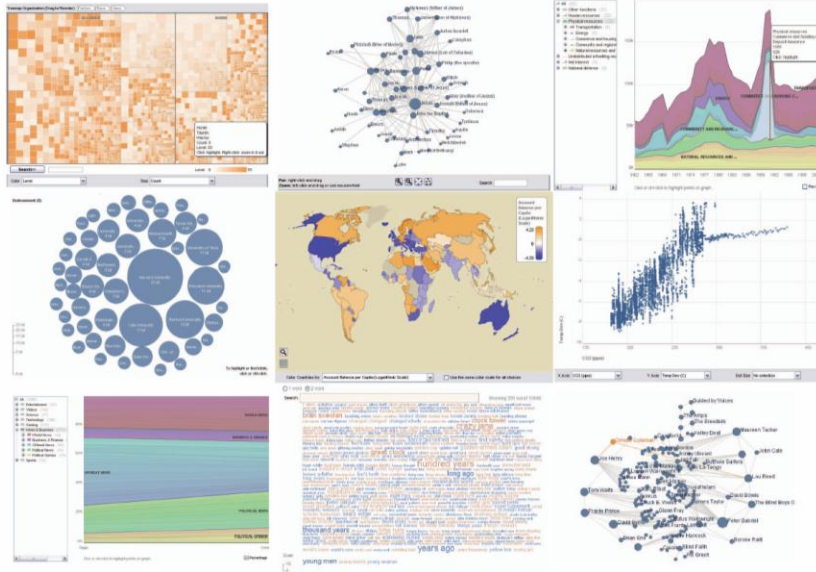
# Visualizations



- Preloaded visualization types
  - Has grown over time
- User chooses one to combine with their data
- Provides named, typed slots that the user maps particular pieces of data to
  - System makes some reasonable guesses too



Originals

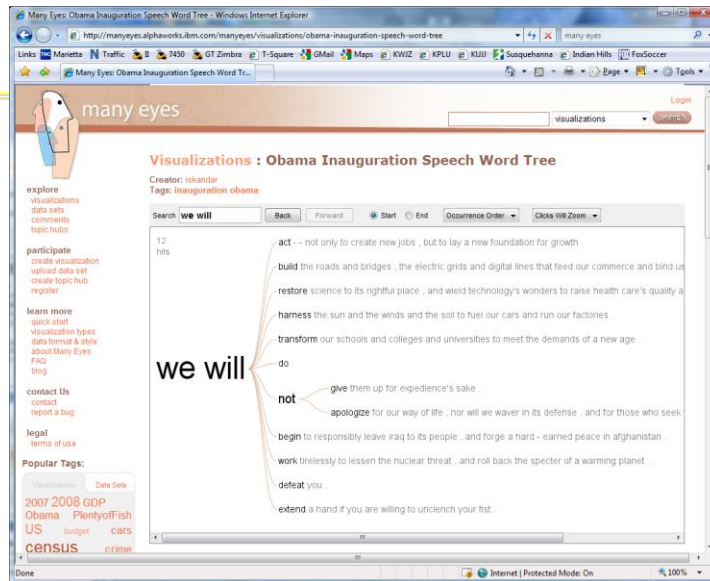


Spring 2011

CS 7450

33

WordTree



Spring 2011

CS 7450

34

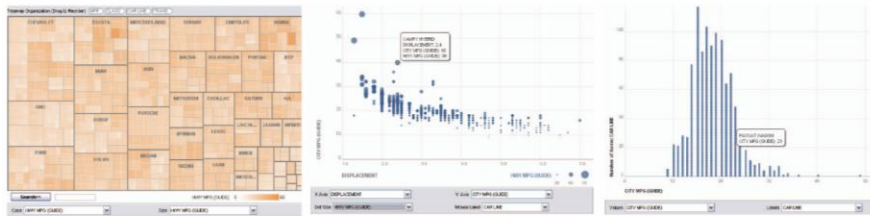


Fig. 3. Three user generated visualizations offering different perspectives on the same dataset on car fuel economy. The grey areas on the top and bottom are automatically generated by the application and allow the user to browse through different dimensions in the data.

Allows the user to control the mapping from data to image

## Social Aspects



- Users identified by login ID
- Can leave comments about different visualizations
- Can take snapshot of visualization state
  - Unique URL
- “Blog this” button

# Evaluation

---



- Quantitative, objective
- 1895 posts of March '07
- Wide variety of topics of visualizations and motivations for creating visualizations
- Does seem to be fostering discussion

Viégas et al  
*HICSS '08*

Spring 2011

CS 7450

37

# Evaluation

---



- Qualitative, subjective
- In-depth interviews with some ME users
- Visualizations used largely for communication and collaboration (not necessarily analysis)
  - Privacy and audience management a concern
- Highlights a number of interesting, non-expected uses of the technology

Danis et al  
*CHI '08*

Spring 2011

CS 7450

38

## Watch More



- Video of talk from Stanford HCI seminar
- <http://hci.stanford.edu/cs547/abstracts/07-08/080201-viegas-wattenberg.html>

Spring 2011

CS 7450

39

## Thoughts?



- What do you think of the design choices they made?

Spring 2011

CS 7450

40

# Sense.us



- Related, follow-up to Many Eyes
- Discussion and visualization of US Census data
- Go beyond Many Eyes in terms of annotation, collaboration, and discussion

Heer et al  
CHI '07

Spring 2011

CS 7450

41

# Components

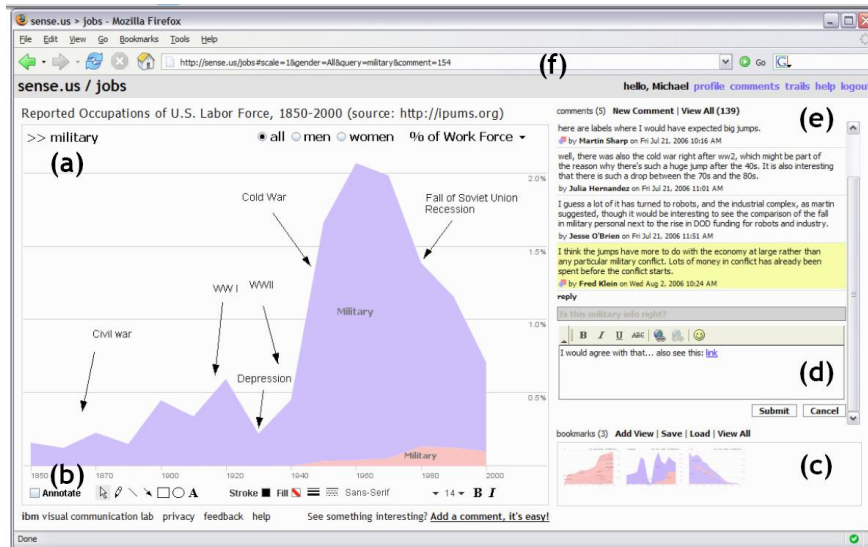


- Doubly-linked discussion
  - Can go from visualization to threaded discussion items or vice-versa
- Graphical annotation
  - Simple graphics editor and comments (like transparent layer)
- Bookmark trail
  - Small strip of views
- Comment listings and social navigation
  - Searchable, sortable indices and links

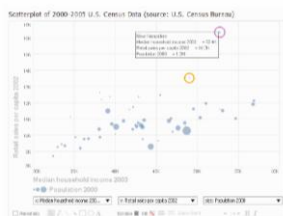
Spring 2011

CS 7450

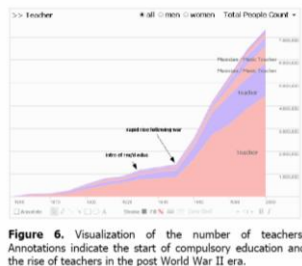
42



**Figure 1. The sense.us collaborative visualization system.** (a) An interactive visualization applet, with a graphical annotation for the currently selected comment. The visualization is a stacked time-series visualization of the U.S. labor force, broken down by gender. Here the percentage of the work force in military jobs is shown. (b) A set of graphical annotation tools. (c) A bookmark trail of saved views. (d) Text-entry field for adding comments. Bookmarks can be dragged onto the text field to add a link to that view in the comment. (e) Threaded comments attached to the current view. (f) URL for the current state of the application. The URL is updated automatically as the visualization state changes.



**Figure 5.** Scatterplot of U.S. states showing median household income (x-axis) vs. retail sales per capita (y-axis). New Hampshire and Delaware have the highest retail sales.



**Figure 6.** Visualization of the number of teachers. Annotations indicate the start of compulsory education and the rise of teachers in the post World War II era.



**Figure 7.** Annotated view of stock brokers. The attached comment reads "Great depression killed a lot of brokers".

## Sample annotations

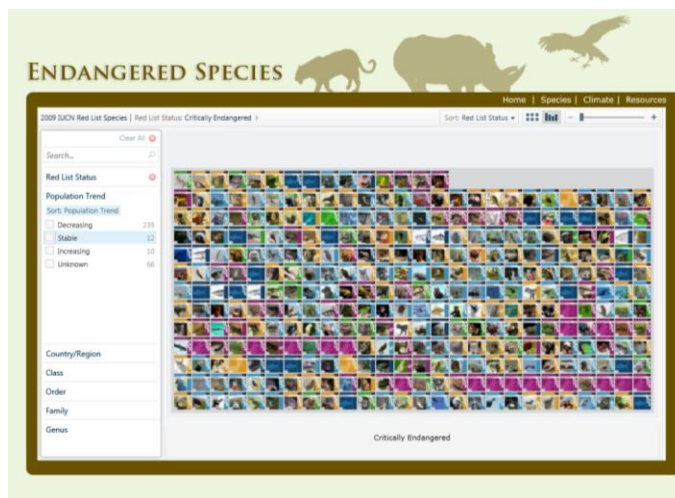
# Microsoft Pivot



- System unveiled to great publicity from LiveLabs
  - After shut-down, has mostly disappeared
- Faceted browsing/ filtering with image-based UI



# Demo



# Toolkits & Infrastructures



- Set of components or capabilities that allow others to put together visualization systems
- Growing trend

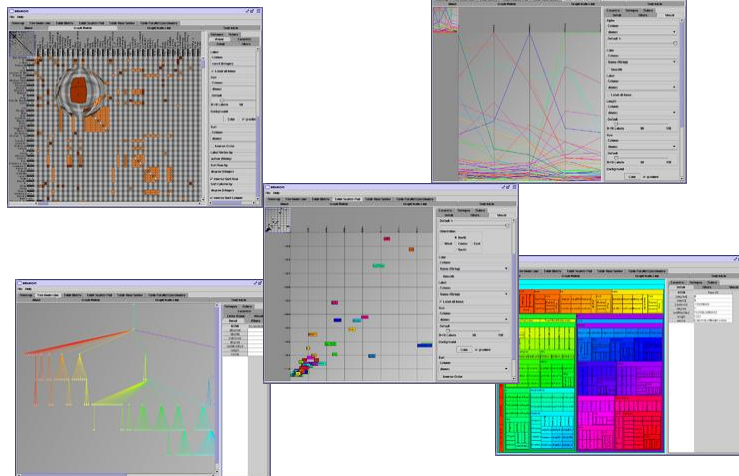
Spring 2011

CS 7450

47

## InfoVis Toolkit

<http://ivtk.sourceforge.net>



Demo

Fekete  
InfoVis '04

Spring 2011

CS 7450

48



# Characteristics



- Coherent software architecture and set of Java components
  - Data structure is a table
  - Views: time series, parallel coordinates, scatterplots, node-link diagrams, treemaps
  - Added capabilities in color management, labeling, dynamic queries, ...

# Critique



- Primarily limited to views it provides, but good & powerful support for them
- Documentation?

# Prefuse



- Toolkit in Java for building information visualizations



Heer et al  
CHI '05

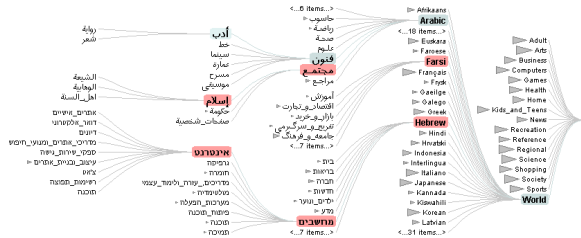
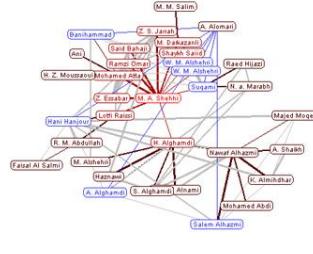
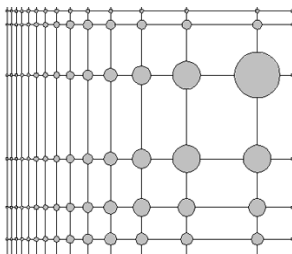
Spring 2011

CS 7450

51

# Prefuse

<http://prefuse.org>



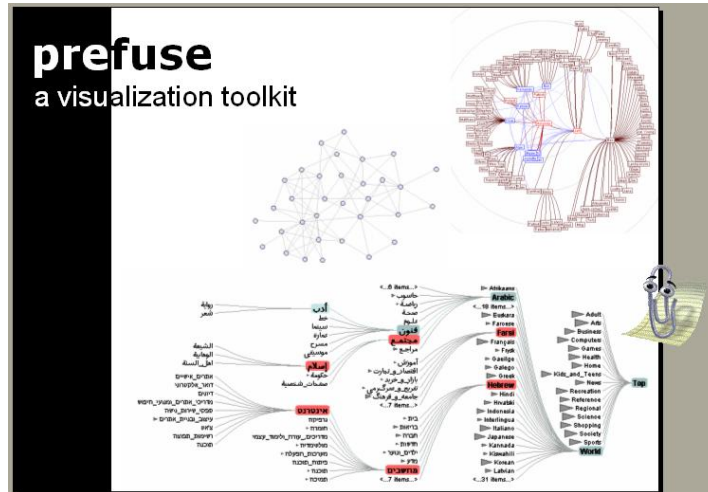
Demo  
Video

Spring 2011

CS 7450

52

# Tutorial Slides



Spring 2011

CS 7450

53

# Characteristics



- Fine-grained building blocks for constructing visualizations (as opposed to pre-made views)
- Data model is a graph (entities & relations)
- Includes library of layout algorithms, navigation and interaction techniques
- Written in Java2d
- Follows CMS vis process model

Spring 2011

CS 7450

54

# Applications



- DOI Trees (AVI '04)
- Vizster (InfoVis '05)
- Many more

Spring 2011

CS 7450

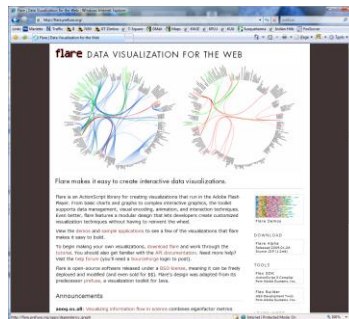
55

<http://flare.prefuse.org>

## Flare



- New version of prefuse implemented in ActionScript for running in Flash player



Demos

Spring 2011

CS 7450

56

# Protovis

<http://vis.stanford.edu/protovis>



- New toolkit following a declarative model
  - Simple grammar of graphical primitives called *marks*
  - Marks are associated with data, using mapping to physical *properties*
  - Properties can be dynamic or static
  - Register *event handlers* to provide interactivity

Bostock & Heer  
*TVCG '09*

Heer & Bostock  
*TVCG '10*

Spring 2011

CS 7450

57

## Built-in Mark Types

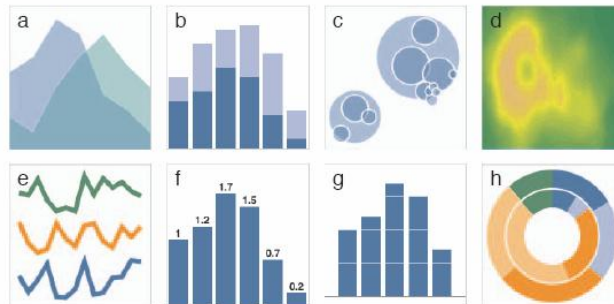


Fig. 5. Examples of built-in mark types. (a-h) Area; Bar; Dot; Image; Line; Label and Bar; Rule and Bar; Wedge.

Spring 2011

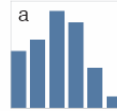
CS 7450

58

# Examples



```
new pv.Panel().canvas("fig3a")
  .add(pv.Bar)
  .data([1, 1.2, 1.7, 1.5, .7, .2])
  .bottom(0).width(20)
  .height(function(d) d * 80)
  .left(function() this.index * 25)
  .root.render();
```



```
new pv.Panel().canvas("fig3b")
  .data([[1, 1.2, 1.7, 1.5, .7],
        [1.5, 1, .8, 1.1, 1.3],
        [.2, .5, .8, .9, 1]])
  .add(pv.Area)
  .data(function(d) d)
  .fillStyle(pv.Colors.category19.parent)
  .bottom(function() let (c = this.cousin())
    c ? (c.bottom + c.height) : 0)
  .height(function(d) d * 40)
  .left(function() this.index * 35)
  .root.render();
```

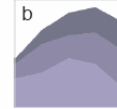
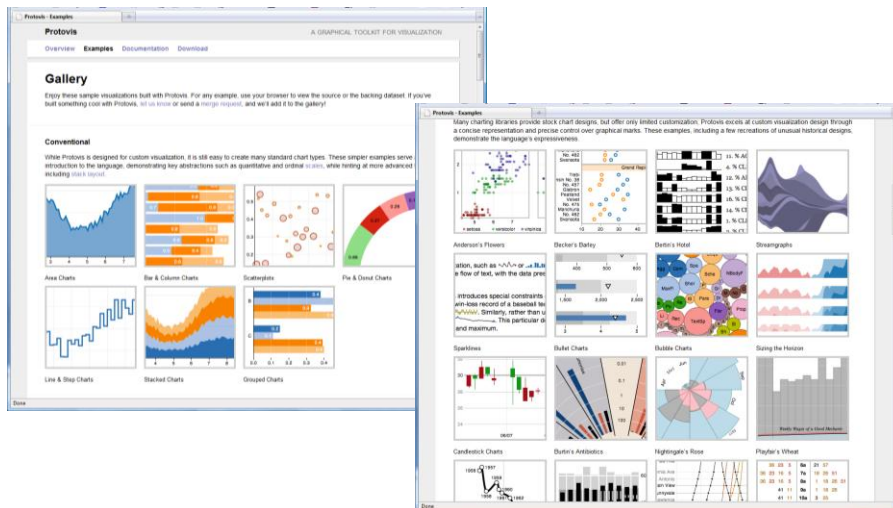


Fig. 2. Specifying two simple charts. (a) Bar. (b) Stacked area.

# Gallery



# Design Patterns



- Heer used his experience with prefuse etc to describe common design patterns for infovis

- Reference model
- Data column
- Cascaded table
- Relational graph
- Proxy tuple
- Expression
- Scheduler
- Operator
- Renderer
- Production rule
- Camera
- Dynamic query binding

Heer & Agrawala  
*TVCG '06*

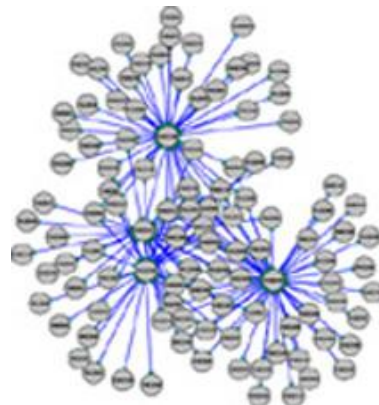
Spring 2011

CS 7450

61

# Piccolo

<http://code.google.com/p/piccolo2d>



Spring 2011

CS 7450

62

# Characteristics



- Graphics toolkit with built-in zooming and panning support
- Very nice for implementing infovis
- Will discuss more later in course...

Spring 2011

CS 7450

63

# InfiView

<http://www.infiview.com>



- Environment for developing web-based data visualizations
- Maps, charts, graphs, ...
- Ajax-powered
- Focus on speed and connecting to large databases



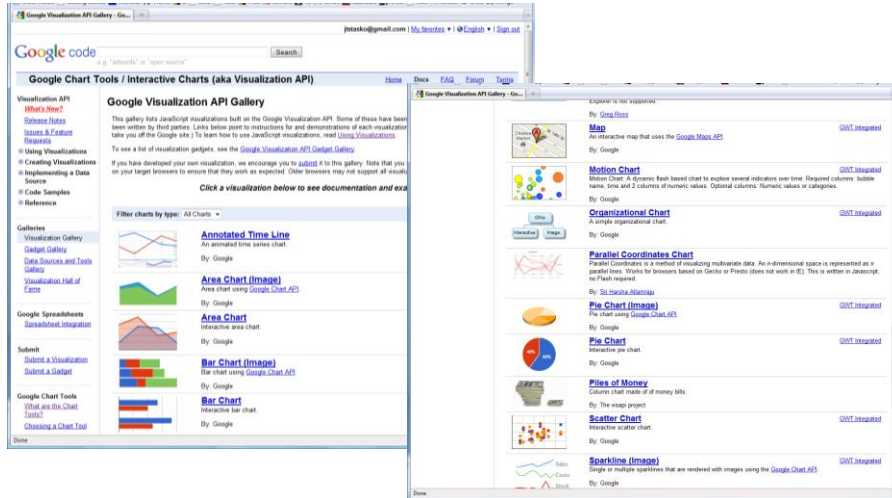
Spring 2011

CS 7450

64



# Google Chart Tools



Spring 2011

CS 7450

65

## Characteristics



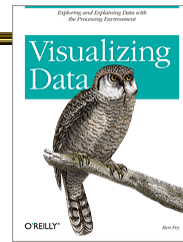
- Javascript-based
- Gallery of contributed code segments
- Visualizations are interactive
- Evolving API

Spring 2011

CS 7450

66

# Processing



- Java based
- Unlike prefuse & flare, not specifically designed for InfoVis
  - Data Reader? Layout algorithm?
  - But can definitely be used to build visualizations!
- Well documented, lots of tutorials with contributions from many people and even books

Ben Fry

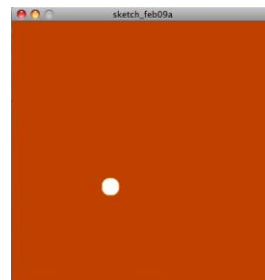
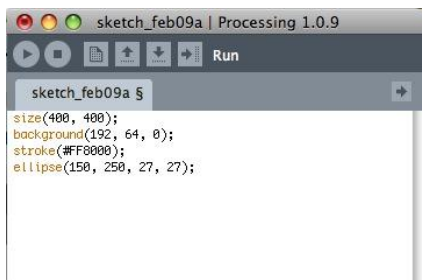
Spring 2011

CS 7450

67

## Processing: the idea

- Programming as scripting
  - PDE: processing development environment
  - A program is called a *sketch*
  - written as a list of statements



Spring 2011

CS 7450

68

# The Pain of using Java2D



- Graphics repaint
  - Animation
  - ActionListener, EventListener ...
  - Long-winded Java syntax
- 
- Processing makes many of these things much easier

Spring 2011

CS 7450

69

# Processing: Useful Functions



```
void setup() {  
    //your own code here  
}
```

```
void draw() {  
    //your own code here  
}
```

- These are built-in functions that are called automatically.
  - The `setup()` block runs once.
  - The `draw()` block runs repeatedly: good for animation

Spring 2011

CS 7450

70

# Reflection



- What would you seek in a good infovis system or toolkit?

Spring 2011

CS 7450

71

# Toolkit Tradeoffs



- InfoVis-focused
  - Many fundamental techniques built-in
  - Can be faster to get something going
  - Often more difficult to implement something “different”
  - Doc is often poor
- Generic graphics
  - More flexible
  - Can customize better
  - Big learning curve
  - Doc is often better
  - Can take a long time to (re)implement basic techniques

Spring 2011

CS 7450

72

## More Next Time...



- Demos of commercial infovis systems
  - Spotfire
  - Table Lens/Eureka
  - SeeIt
  - InfoZoom
  - InfoScope
  - Tableau

Spring 2011

CS 7450

73

## Upcoming



- Commercial systems
  - Reading  
Spence paper
- Tufte's Principles

Spring 2011

CS 7450

74