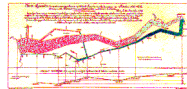


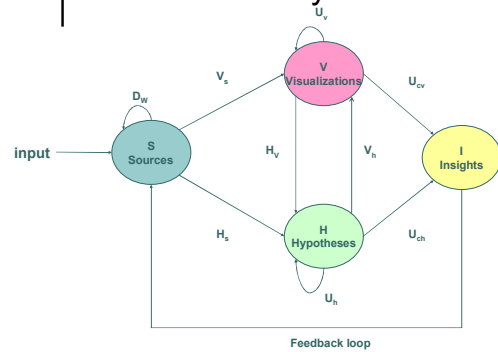
## Visual Analytics

Lecture 2: Software and Challenges

Marcel Worring



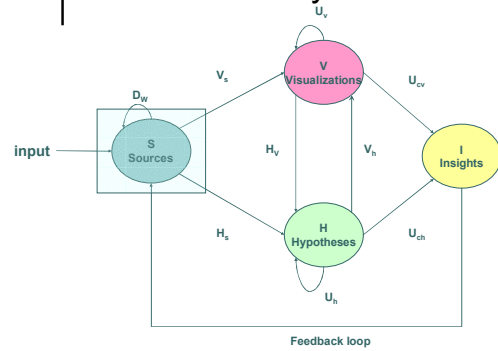
## The visual analytics model



## Visual Analytic Systems

- o Needed
  - Support for data management
  - Support for Information Visualization
  - Support for GUI management
  - Support for the analytic process
  - Support for data analysis

## The visual analytics model



## Data source representation

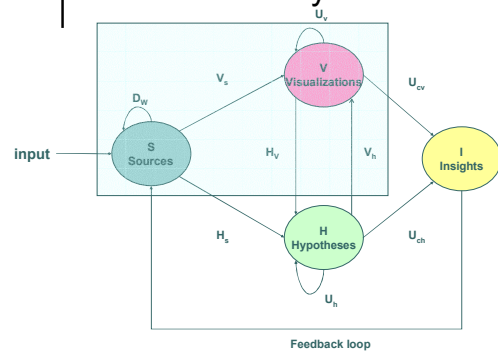
Item	Attribute <sub>i</sub>	Attribute <sub>j</sub>	Attribute <sub>k</sub>	
Instantiation <sub>x</sub>	Value <sub>ix</sub>	Value <sub>jx</sub>	Value <sub>kx</sub>	...
Instantiation <sub>y</sub>	Value <sub>iy</sub>	Value <sub>jy</sub>	Value <sub>ky</sub>	...
...	...	...	...	...

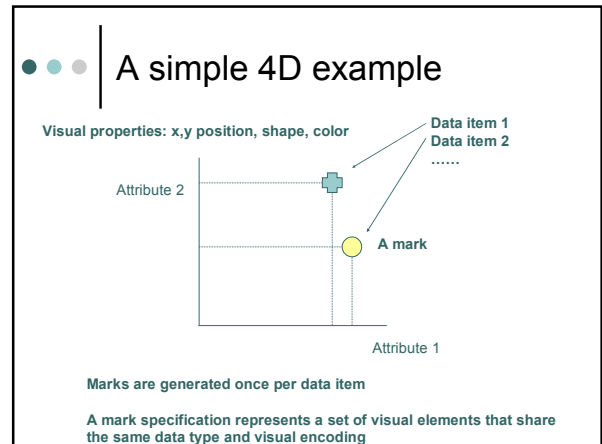
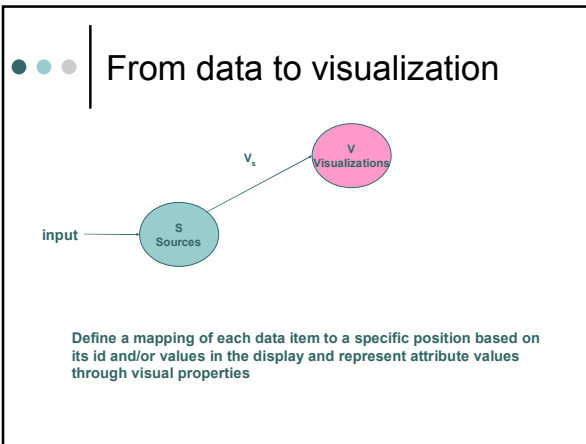
We assume one homogeneous set of items where for each item the same set of attributes has been defined.

The attributes do NOT have to be homogeneous and can be several different types including all 7 of Schneiderman

We can have multiple data source tables in an application

## The visual analytics model



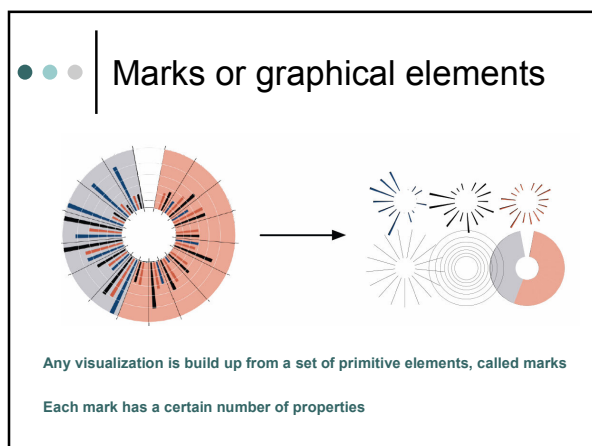


Bertin, picture from Krygier and Wood

## The properties

	Points	Lines	Areas	Best to show
Shape				qualitative differences
Size				quantitative differences
Color Hue				qualitative differences
Color Value				quantitative differences
Color Intensity				qualitative differences
Texture				qualitative & quantitative differences

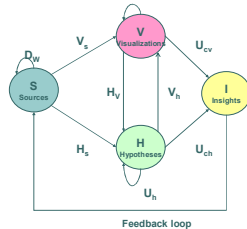
- ## Properties
- Can be static
    - A constant ignoring the value of the data item
  - Can depend on the data
    - Derived from the value or the index of the data item
  - Can change dynamically



- ## Software
- Support for data management
  - Support for information visualization
    - Support for drawing graphical elements independent of the data
    - Support for drawing graphical elements as function of the data
    - Support for interaction with the graphical elements
    - Support for animating the graphical elements
  - Support for data analysis

## Graphical elements and data

- o Needed for different components
  - Data
  - Models
  - Hyptheses



## Software

- o Systems
  - Complete solutions in which you can enter data
- o Toolkits
  - Programming based
  - Scripting based

## Systems

- o Systems
  - [ManyEyes](#)
  - [Tableau](#)
  - [Google Analytics](#)

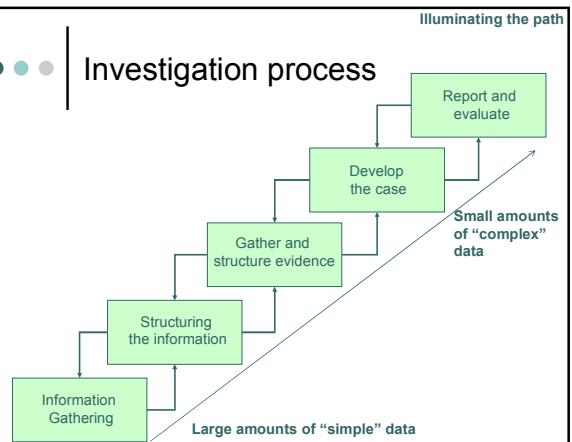
## Toolkits

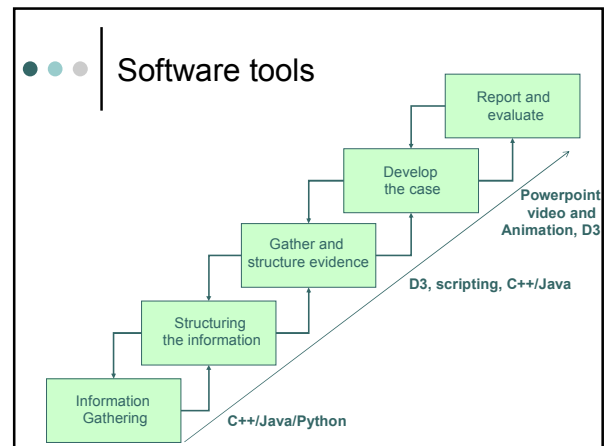
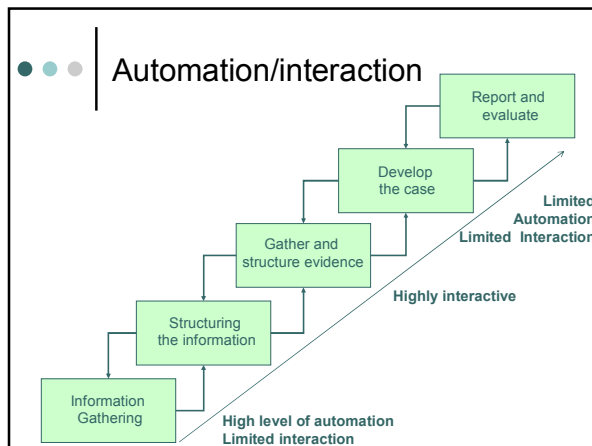
- o Java Based
  - [Prefuse](#)
  - [Processing](#)
  - [The InfoVis toolkit](#)
  - [Mondrian](#)

## Toolkits

- o Scripting Based
  - Actionscript/Flash
    - [Flare](#)
  - JavaScript
    - [Protovis](#)
    - [D3](#)

## Investigation process





## ProtoVis

- ProtoVis
  - An extensible toolkit for constructing visualizations by composing simple graphical primitives
  - A nice balance between high-level means for expression and fine-grained control

## Example properties in Protovis

bar

data	[1, 1.2, 1.7, 1.5, .7]
visible	true
left	$\lambda: i * 25$
bottom	0
height	$\lambda: d * 80$
width	20
fillStyle	blue
strokeStyle	black
lineWidth	1.5
...	...

## Properties in ProtoVis

```
vis.add(pv.Bar)
  .data([1, 1.2, 1.7, 1.5, .7])
  .visible(true)
  .left(function() this.index * 25)
  .bottom(0)
  .height(function(d) d * 80)
  .width(20)
  .fillStyle("blue")
  .strokeStyle("black")
  .lineWidth(1.5);
```

## D3

- Data Driven Documents
  - Further development of ProtoVis
    - Improved interaction
    - Improved efficiency

## D3: basics is html

Style controlled by .css file

```

HTML page
<html>
<h1> header 1 </h1>
<p> paragraph </p>
<p> paragraph </p>
.....
.....
</html>

```

Described by hierarchical document object model

## D3: html might contain SVG

Style can and should also be controlled by .css file

```

HTML page
<html>
<h1> header 1 </h1>
<svg width="100px" height="100px">
<g>
  <rect x="10" y="10"
    width="40" height="40"
    fill="blue" stroke="#0ebb00"
    stroke-width="4" />
</g>
</svg>
</html>

```

Scalable Vector Graphics

## D3: html might contain JavaScript

```

HTML page
<html>
<h1> header 1 </h1>
.....
.....
<javascript>
  DOM manipulation functions
  Computations
  Interaction control
</javascript>
.....
.....
</html>

```

Scalable Vector Graphics

## D3

- Builds on top of
  - HTML
  - SVG
  - Javascript
- Providing
  - Visualization specific graphics
  - DOM manipulation functions
    - Directly using HTML/SVG not through intermediate language
  - Several helper modules

## D3: basic elements

**Data joins:**  
 bind input data to elements, enabling functional operators that depend on the data

**Document**  
 Containers for visualization elements

**Selections:**  
 A filtered set of elements queried from the current document

**Operations:**  
 Act on selections, modifying content


**Transitions:** interpolate Attributes and styles smoothly over time

## D3: container elements

Document

## D3: basic elements

**Data joins:**  
bind input data to elements, enabling functional operators that depend on the data



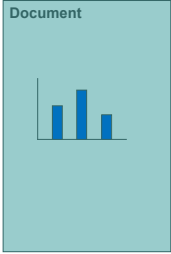
## D3: basic elements

**Data joins:**  
bind input data to elements, enabling functional operators that depend on the data

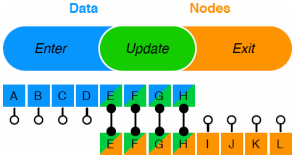
**Selections:**  
All rectangles

**Operators:**  
Make them blue

**Transitions:**  
interpolate from green to blue

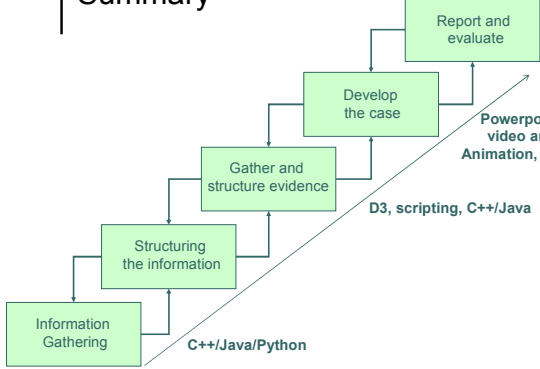


## Dynamics in D3



When new data arrives (the blue part) and are joined with existing nodes (graphical elements depicted in orange), three subselections result

## Summary



## The visual analytics challenge

- The challenge is organized yearly
  - To promote scientific progress
  - To benchmark current systems
  - To provide a set of challenging and realistic problems
- The data
  - Combination of real anonymized data with artificial elements added

## Challenge 2011

- Mini challenge 1
  - Characterization of an epidemic spread
    - Based on geo-coded microblog messages
- Mini challenge 2
  - Computer networking operation
  - Network data, firewall logs, IDS logs



## Challenge 2011

- Mini challenge 3
  - Investigation into terrorist activity
    - Based on a text corpus with news reports
- The grand challenge
  - All the three challenges are in some way related. Can you find out whether the epidemic, network intrusions, and terrorist activity are related?



## Visual Analytics

- Challenge MC1
  - Colorado
  - Bangor
  - Stuttgart
  - Purdue



## Visual Analytics

- Challenge MC2
  - Charlotte



## Visual Analytics

- Challenge MC3
  - Oculus
  - Virginia Tech



## Visual Analytics

- Grand Challenge
  - Konstanz