Representation & Analytic Tasks

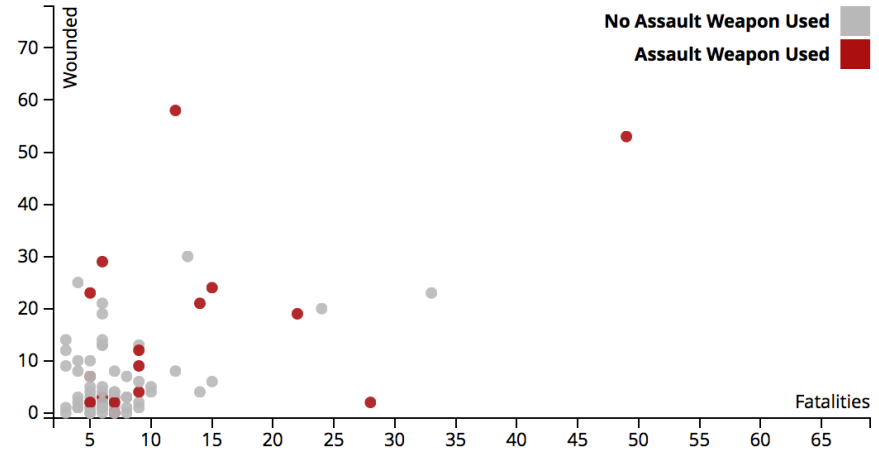
James EAGAN

TELECOM ParisTech

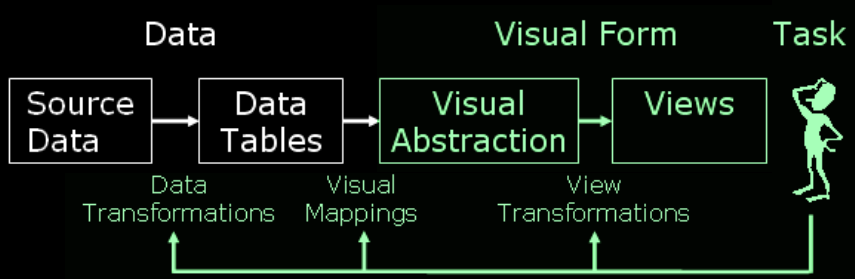
Slides adapted from John Stasko (Georgia Tech),
Petra Isenberg & Jean-Daniel Fekete (INRIA),
Chris North (Virginia Tech)

Critique du moment

Source: Mother Jones
Graphic: Carla Astudillo / International Business Times



InfoVis Pipeline



Low-level analytic tasks

- Retrieve value
- Filter
- Compute derived value
- Find extremum
- Sort
- Determine range
- Characterize distribution
- Find anomalies
- Cluster
- Correlate

Retrieve value

- For a specific data case, find its attributes

Filter

- Show only the data cases matching a set of conditions based on attribute values

Compute derived value

- Compute an aggregate value for a set of data cases

Find extremum

- Find data cases at either extreme of the data set for a given attribute

Sort

- Rank data cases according to some ordinal metric

Determine range

- Identify the span of values for a given set of data cases and attributes

Characterize distribution

- For a given set of cases and attributes of interest, characterize the distribution of that value over the set

Find anomalies

- Identify any anomalies in a set of data cases with respect to a given relationship or expectation (*e.g.* statistical outliers)

[Amar, Eagan, & Stasko, InfoVis '05]

Cluster

- Within a set of cases, identify any clusters of similar attribute values

[Amar, Eagan, & Stasko, InfoVis '05]

Correlate

- For a given set of cases and two attributes, determine useful relationships between the values of those attributes

[Amar, Eagan, & Stasko, InfoVis '05]

Low-level analytic tasks

- Retrieve value
- Filter
- Compute derived value
- Find extremum
- Sort
- Determine range
- Characterize distribution
- Find anomalies
- Cluster
- Correlate

[Amar, Eagan, & Stasko, InfoVis '05]

Kinds of interaction

Select: mark something as interesting

Explore: show something else

Reconfigure: show a different arrangement

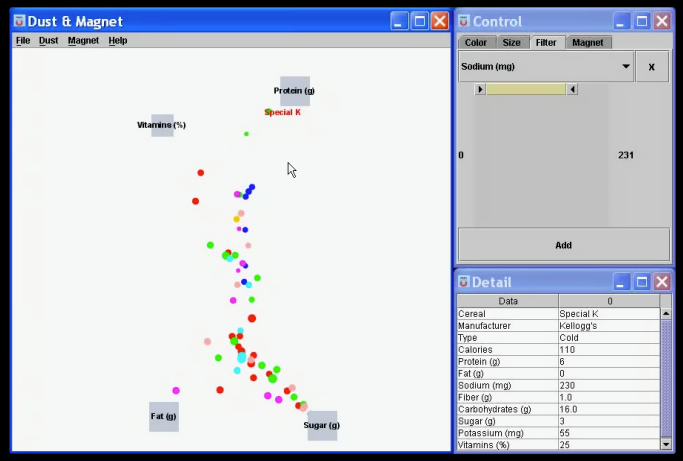
Encode: show a different representation

Abstract/Elaborate: show less or more detail

Filter: show something conditionally

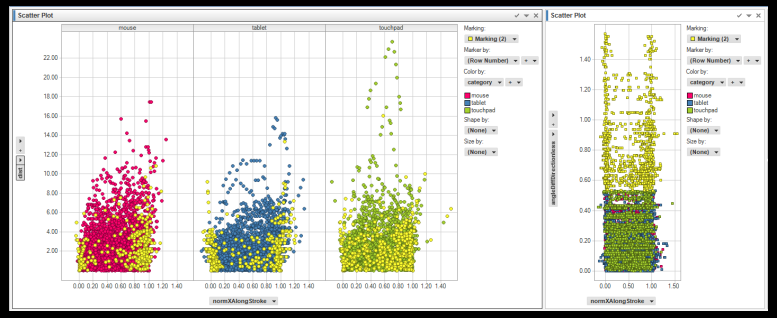
Connect: show related items

[Yi et al., TVCG '07]



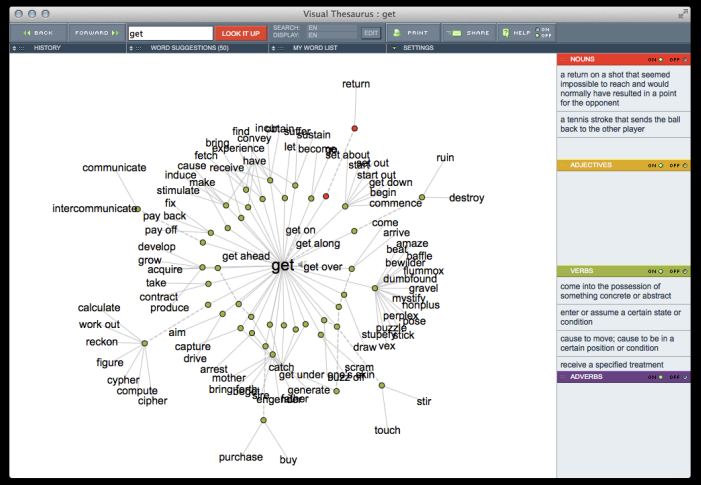
Select

[Yi et al., InfoVis '05]



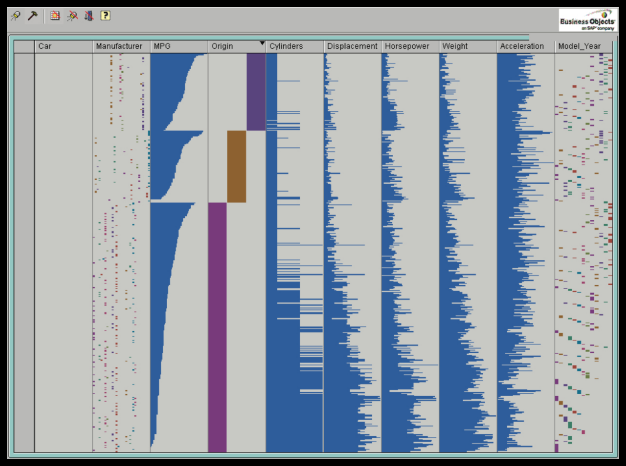
Select

[Ahlberg, SIGMOD Record '96]



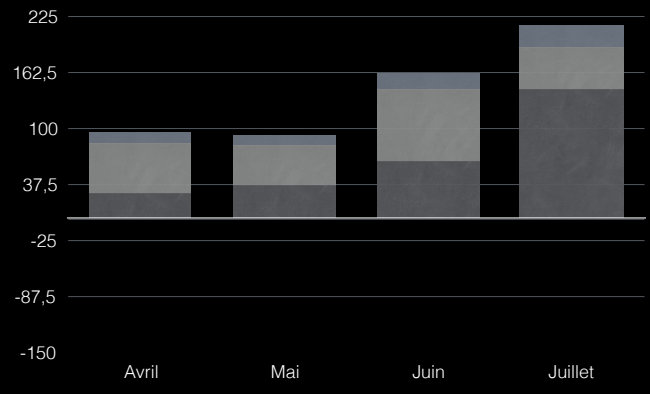
Explore

[www.visualthesaurus.com]



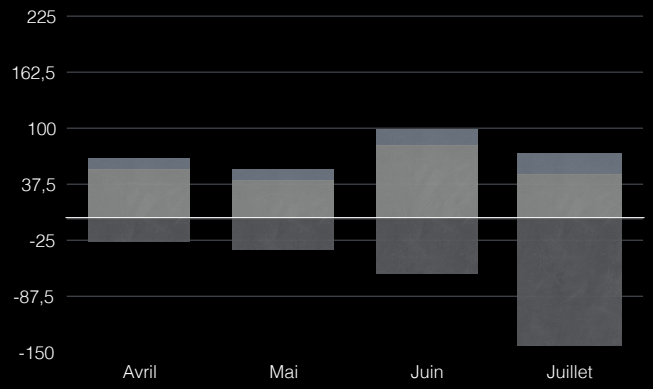
Reconfigure

[Rao & Card, CHI '94]



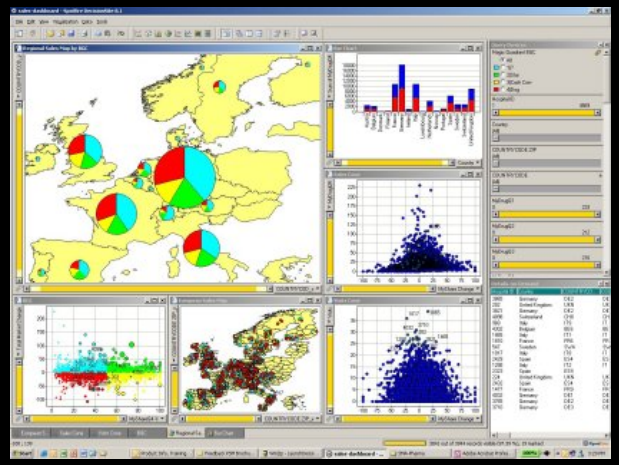
Reconfigure

[Dix & Ellis, AVI '98]



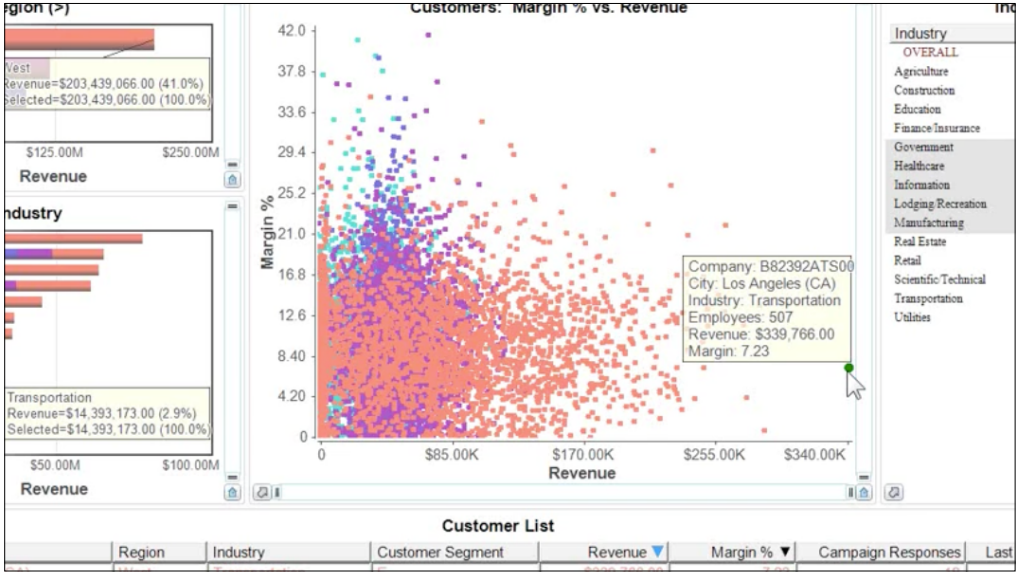
Reconfigure

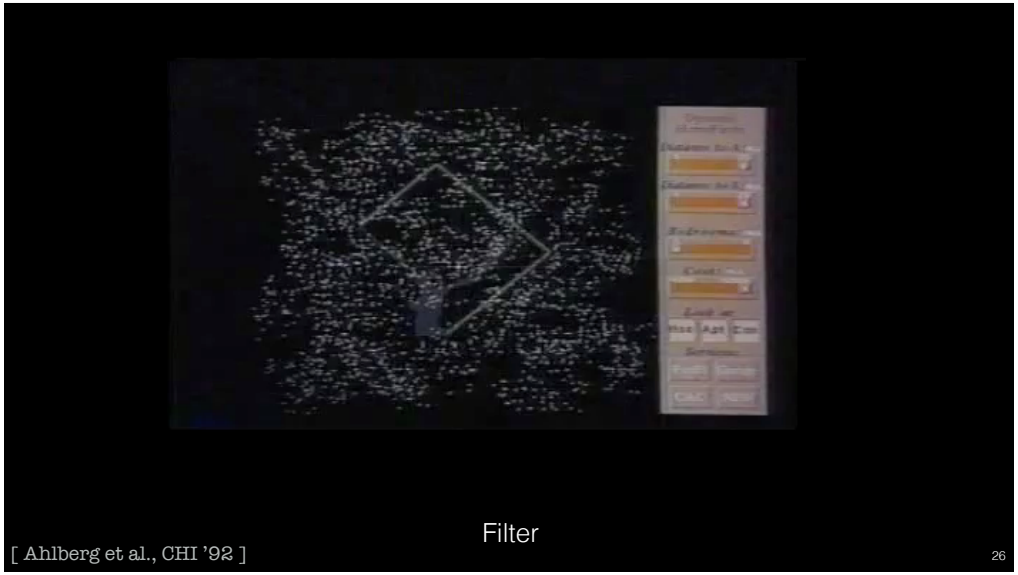
[Dix & Ellis, AVI '98]



Encode

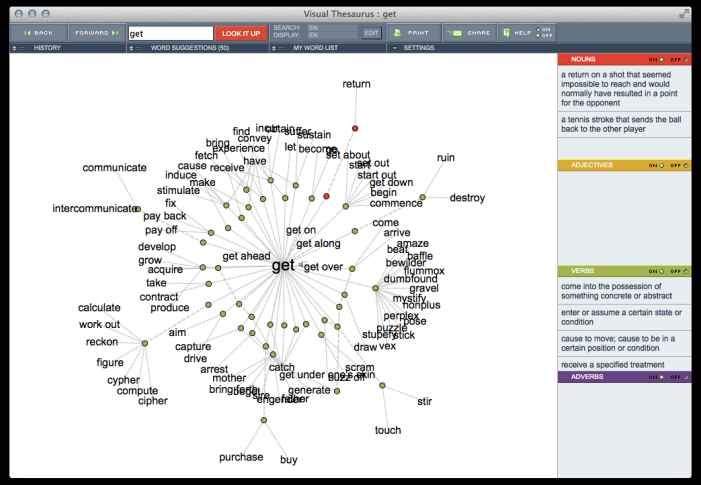
[Ahlberg, SIGMOD Record '96]





Filter

[Ahlberg et al., CHI '92]



Connect

[www.visualthesaurus.com]

Kinds of interaction

Select: mark something as interesting

Explore: show something else

Reconfigure: show a different arrangement

Encode: show a different representation

Abstract/Elaborate: show less or more detail

Filter: show something conditionally

Connect: show related items

[Yi et al., TVCG '07]

Response time

- *0.1 sec — animation, visual continuity*
- *1 sec — system response, conversation break*
- *10 sec — cognitive response*

Putting it all together

- *Problem: Finding a home*

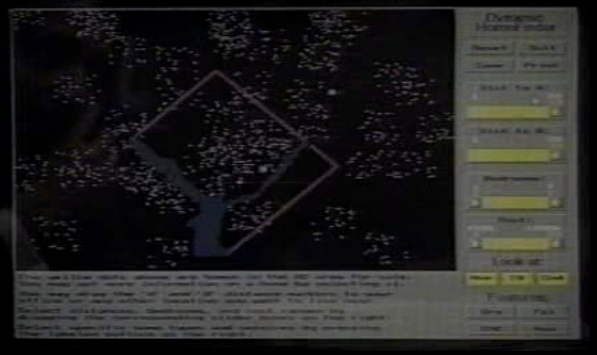
Dynamic queries

- Query language
- **Select** house-address
  From sf-realty-db
  Where price >= 200,000 **and**
  price <= 400,000 **and**
  bathrooms >= 3 **and**
  garage == 2 **and**
  bedrooms >= 4

Dynamic queries

- Query language
- **Select** house-address
  From sf-realty-db
  Where price >= 200,000 **and**
  price <= 400,000 **and**
  bathrooms >= 3 **and**
  garage == 2 **and**
  bedrooms >= 4

Dynamic queries



[Ahlberg & Shneiderman, '92]

What did we see?

- *Strengths?*
- *Weaknesses?*

Shneiderman's mantra

- Overview first, zoom & filter, details on demand
- Overview first, zoom & filter, details on demand
- Overview first, zoom & filter, details on demand

Big Data

- Techniques to reduce data
  - Sampling
  - Aggregation

Twitter logs

- Suppose you have logs of user events on Twitter

Component	Description	Example
client	client application	`web`, `iphone`, `android`
page	page or functional grouping	`home`, `profile`, `who_to_follow`
section	tab or stream on a page	`home`, `mentions`, `retweets`, `searches`
component	component, object, or objects	`search_box`, `tweet`
element	UI element within the component	`button`, `avatar`
action	actual user or application action	`impression`, `click`, `hover`

Field	Description
`event_name`	event name
`user_id`	user id
`ip`	user's IP address
`timestamp`	timestamp
`event_details`	event details

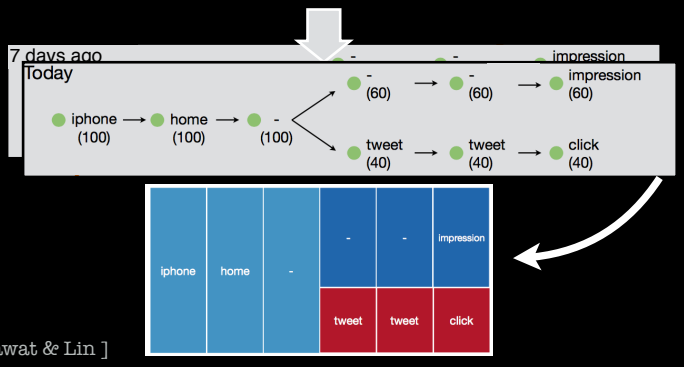
[Wongsuphasawat & Lin, VAST '14]

web:home:mentions:stream:avatar:profile_click

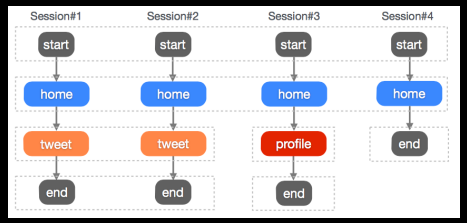
web:home:mentions:stream:avatar:profile_click

Task: See changes in patterns

client	page	section	component	element	action	count
iphone	home	-	-	-	impression	60
iphone	home	-	tweet	tweet	click	40



Overview

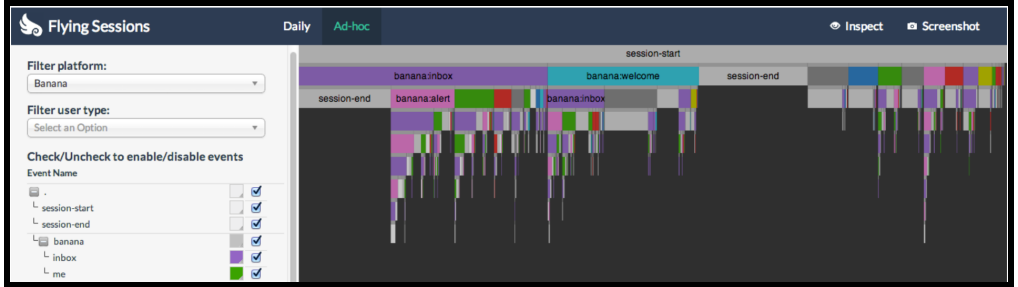


Aggregation

web:home:mentions:stream:avatar:profile_click

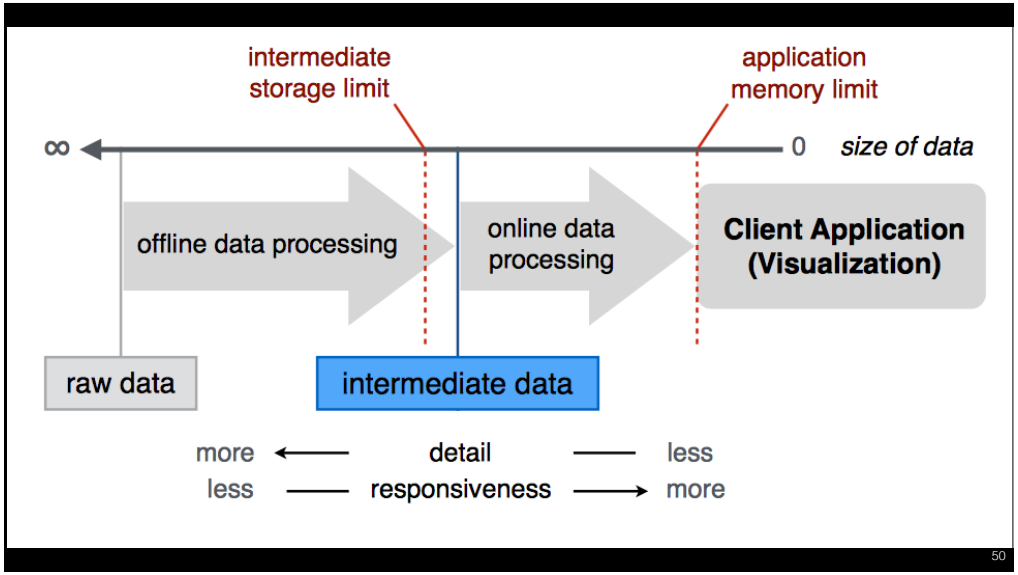
ABCDEF → ABCDEF

ABC	(2000)	ABC	(2000)	ABC	(2000)
ABCD	(80)	ABCD	(80)	ABCx	(102)
ABCE	(20)	ABCE	(20)		
ABCDF	(1)	ABCDx	(2)		
ABCDG	(1)				



[Wongsuphasawat & Lin, VAST '14]

Interactivity



USING VISUALIZATIONS
TO MONITOR CHANGES AND HARVEST INSIGHTS
FROM A GLOBAL-SCALE LOGGING INFRASTRUCTURE
AT TWITTER

Krist Wongsuphasawat & Jimmy Lin
@kristw @lintool

IEEE VAST 2014
Paris, France