Tackling the Challenges of Big Data
Big Data Systems

David Karger

Professor

Massachusetts Institute of Technology

Tackling the Challenges of Big Data
Big Data Systems

David Karger

Professor

Massachusetts Institute of Technology

© 2014 Massachusetts Institute of Technology

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
The Big Picture

David Karger

Professor

Massachusetts Institute of Technology

© 2014 Massachusetts Institute of Technology

My Research Group

- **Mission**
  - Understand what makes it hard to manage data
  - Create tools to fix the problems we understand
  - Watch people use them to understand more
- **Technical Areas:**
  - Databases
  - Information Retrieval
  - Machine Learning
  - Semantic Web
  - Human Computer Interaction
- **Theme**
  - Empower end users to manage on their own

Why User Interfaces?

The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades, … because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Hal Varian, Google's Chief Economist
***The McKinsey Quarterly*, Jan 2009**

Why User Interfaces?

- **Computers can store and process information at super-human scale**
- **But people are still superior to computers in many information processing activities**
  - See patterns
  - Notice oddities
  - Impose order
  - Choose a suitable model
- **Combination is more powerful than either**
- **Even if computer does most of the work, need a way for human to tell it what work to do**

The Value of Interfaces

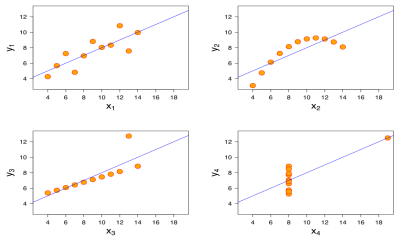
- **Analyze data to support reasoning**
  - Expand memory
  - Find patterns
  - Develop and assess hypotheses
  - Discover errors in data
- **Communicate information to others**
  - "Seeing is Believing", "Picture worth 1000 words"
  - Share and persuade
  - Collaborate and revise

Leveraging Human Insight

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

- **4 x-y data sets (Anscombe's quartet)**
- **Identical means and standard deviations**
- **What's the difference?**

Leveraging Human Insight

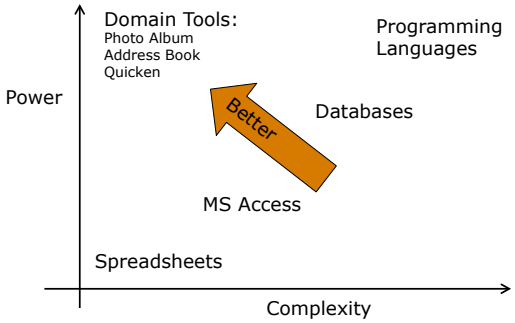


- **4 x-y data sets (Anscombe's quartet)**
- **Identical means and standard deviations**
- **What's the difference?**

User Interfaces for Data

- **User Interface Design is a huge subfield of HCI**
- **Focus here on UI for data**
  - large number of items
  - all following same schema
  - e.g., table of rows (items) with columns (schema)
  - or many items of a few types, where each type has a common set of properties
- **Data UIs exploit the regularity of the data**
  - spreadsheet shows each property in a column
  - scatter plot maps one property on each axis
  - web site uses 1 template for all items of same type
  - and facets to filter on a particular property

Spectrum of Interface Capabilities



Exclude

- **Image interfaces**

Photos removed due to copyright restrictions.

We are sorry for any inconvenience this may cause.

- **Text interfaces**
- **Audio/Signal interfaces**

Overview

- **Information visualization**
  - Leveraging human perception to understand data
- **Lying with visualizations**
  - What not to do, and why
- **Exploratory data analysis**
  - Interfaces to understand data through exploration
- **Faceted browsing**
  - The standard for data on the web
- **Research: beyond the spreadsheet**
- **Research: building your own interfaces**

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces
The Big Picture

THANK YOU

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger

Professor

Massachusetts Institute of Technology

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Information Visualization

David Karger

Professor

Massachusetts Institute of Technology

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data © 2014 Massachusetts Institute of Technology

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data © 2014 Massachusetts Institute of Technology

Find patterns: NYC weather

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

- **NY Times, 1981**
- **2200 numbers in a comprehensible aggregate**

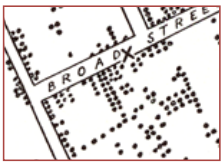
Catalhoyuk Map

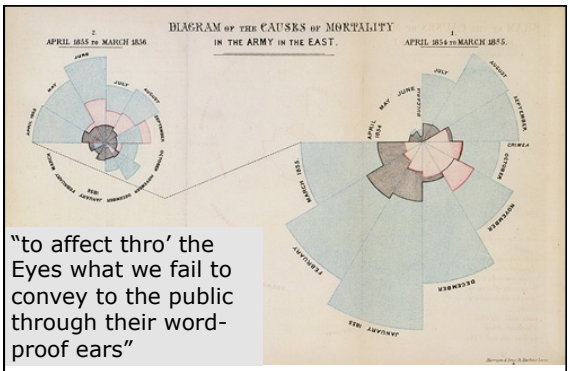
Circa 6200 BC

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

London Cholera outbreak

- **John Snow, 1854**
- **Map shows deaths clustered around pump**





"to affect thro' the Eyes what we fail to convey to the public through their word-proof ears"

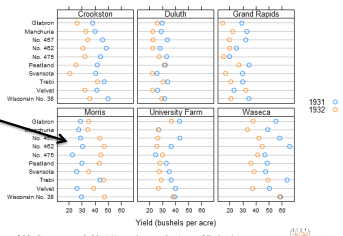
1856 Crimean War Death "Coxcomb", Florence Nightingale

Small Multiples

Many copies of a visualization

- All at same scales
- Indexed by another variable
- Enables side-by-side comparison

Colors flipped



What is visualization?

"Transformation of the symbolic into the geometric"

[McCormick et al. 1987]

"... finding the artificial memory that best supports our natural means of perception." [Bertin 1967]

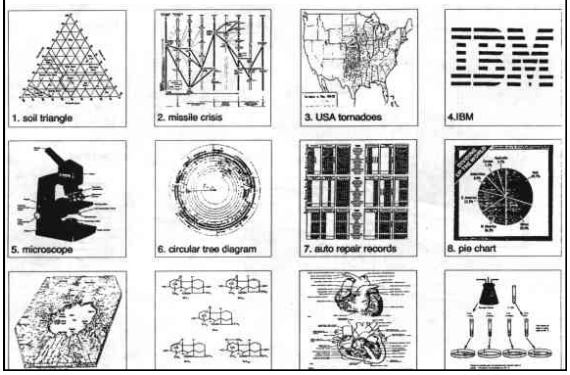
"The use of computer-generated, interactive, visual representations of data to amplify cognition."

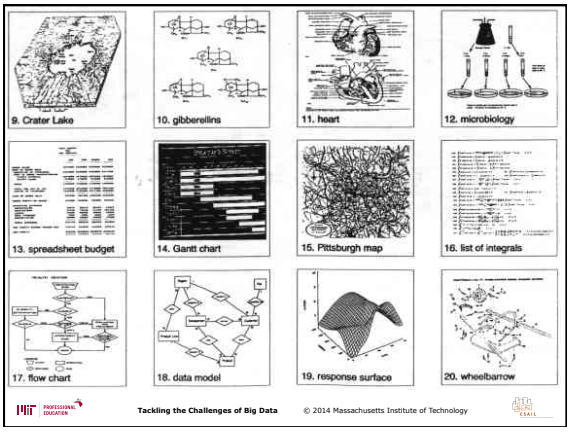
[Card, Mackinlay, & Shneiderman 1999]

Why Visualizations?

- **"A picture is worth a thousand words"**
- **We receive information through our senses**
- **Far higher bandwidth than thought**
  - React to an image much faster than thinking
  - Perceive and remember details without thinking
- **Transform abstract data to visual variables**
- **Leverage pre-attentive processing**
  - Information we process without thinking

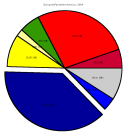
Visual Representations [Lohse '94]

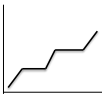
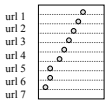




Common Graph Types





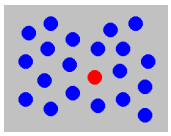
Visual Variables [Bertin '74]

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

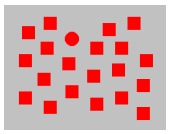
Pre-attentive [Treisman, '85]

- **pre-attentive processing of visual information is performed automatically on the entire visual field detecting basic features of objects in the display. Such basic features include colors, closure, line ends, contrast, tilt, curvature and size. These simple features are extracted from the visual display in the pre-attentive system and later joined in the focused attention system into coherent objects. pre-attentive processing is done quickly, effortlessly and in parallel without any attention being focused on the display.**

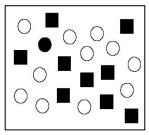
Pre-attentive Visual Variables



Color
[Healy '96]



Shape
[Chipman '96]



Combination prevents
[Healy '96]

Pre-attentive Visual Variables

[Healey '97]

length	**Triesman & Gormican [1988]**
width	**Julesz [1985]**
size	**Triesman & Gelade [1980]**
curvature	**Triesman & Gormican [1988]**
number	**Julesz [1985]; Trick & Pylyshyn [1994]**
terminators	**Julesz & Bergen [1983]**
intersection	**Julesz & Bergen [1983]**
closure	**Enns [1986]; Triesman & Souther [1985]**
color (hue)	**Nagy & Sanchez [1990, 1992]; D'Zmura [1991] Kawai et al. [1995]; Bauer et al. [1996]**
intensity	**Beck et al. [1983]; Triesman & Gormican [1988]**
flicker	**Julesz [1971]**
direction of motion	**Nakayama & Silverman [86]; Driver & McLeod [92]**
binocular lustre	**Wolfe & Franzel [1988]**
stereoscopic depth	**Nakayama & Silverman [1986]**
3-D depth cues	**Enns [1990]**
lighting direction	**Enns [1990]**

Text NOT pre-attentive

SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

Text NOT pre-attentive

SUBJECT	PUNCHED	QUICKLY	OXIDIZED	TCEJBUS	DEHCNUP	YLKCIUQ	DEZIDIXO
CERTAIN	QUICKLY	PUNCHED	METHODS	NIATREC	YLKCIUQ	DEHCNUP	SDOHTEM
SCIENCE	ENGLISH	RECORDS	COLUMNS	ECNEICS	HSILGNE	SDROCER	SNMULOC
GOVERNS	PRECISE	EXAMPLE	MERCURY	SNREVOG	ESICERP	ELPMAXE	YRUCREM
CERTAIN	QUICKLY	PUNCHED	METHODS	NIATREC	YLKCIUQ	DEHCNUP	SDOHTEM
GOVERNS	PRECISE	EXAMPLE	MERCURY	SNREVOG	ESICERP	ELPMAXE	YRUCREM
SCIENCE	ENGLISH	RECORDS	COLUMNS	ECNEICS	HSILGNE	SDROCER	SNMULOC
SUBJECT	PUNCHED	QUICKLY	OXIDIZED	TCEJBUS	DEHCNUP	YLKCIUQ	DEZIDIXO
CERTAIN	QUICKLY	PUNCHED	METHODS	NIATREC	YLKCIUQ	DEHCNUP	SDOHTEM
SCIENCE	ENGLISH	RECORDS	COLUMNS	ECNEICS	HSILGNE	SDROCER	SNMULOC

Elementary Abstract Data Types

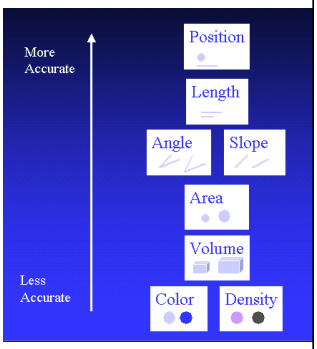
- **Nominal (qualitative)**
  - (no inherent order)
  - city names, types of diseases, ...
- **Ordinal (qualitative)**
  - (ordered, but not at measurable intervals)
  - cold, warm, hot; historical eras …
- **Quantitative**
  - Numeric
  - Some absolute (fixed 0): mass, length
  - Some relative (arbitrary 0): date, position
- **Relational**
  - Connections between items
  - Social network, subway map

Visual Variables [Bertin '74]

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Visual Variable Accuracy

- **[Mackinlay '88 from Cleveland & McGill]**
- **Based on pairwise comparisons**



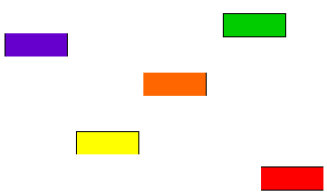
Ranking Utility of Visual Variables for Different Data Types

(Mackinlay '88, not empirically verified)

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density	Color Hue
Angle	Color Saturation	Texture
Slope	Color Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Color Saturation
Color Saturation	Length	Shape
Color Hue	Angle	Length

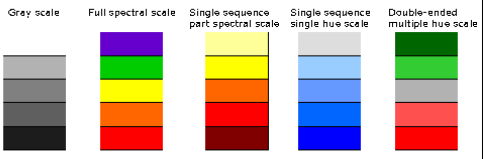
Color Schemes

Order these (low->hi)



1

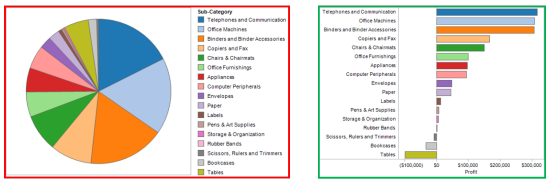
Color Schemes



1

John Tukey (more later)

There is no data that can be displayed in a pie chart, that cannot be displayed BETTER in some other type of chart.

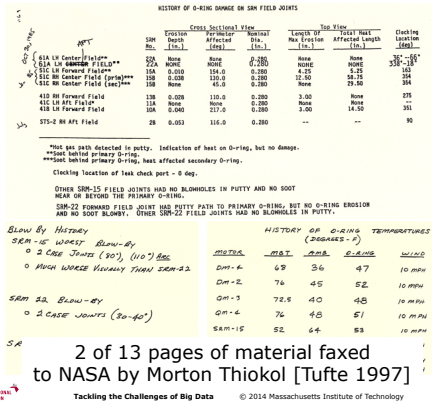


CASE STUDY: CHALLENGER SHUTTLE

Challenger Disaster

- Jan 28, 1987
- Challenger shuttle scheduled for launch
- History of problems with O-rings
- Unusually cold weather
- Engineers argued against launch
- NASA went ahead anyway

- Why?



HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A NONE	None	None	0.280	None	None	352 -165°
61A LH Center Field**	22A NONE	None	None	0.280	None	None	358 -18°
51C LH Forward Field**	15A 0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	15B 0.028	110.0	0.280	12.50	58.75	164
51C RH Center Field (sec)***	15B None	45.0	0.280	None	29.50	164
41D RH Forward Field	13B 0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A None	None	0.280	None	None	--
41B LH Forward Field	10A 0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B 0.053	116.0	0.280	--	--	90

*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

Blow By History
SRM-15 Worst Blow-By
- 2 Case Joints (30°), (110°) Asc
- Much Worse Visually than SRM-22

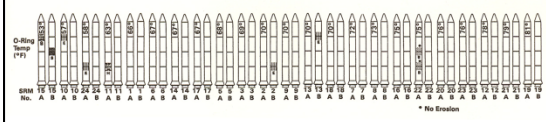
SRM 22 Blow-By
- 2 Case Joints (30-40°)

HISTORY OF O-RING TEMPERATURES (DEGREES-F)

MOTOR	MAT	AMB.	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SAM-15	52	64	53	10 MPH

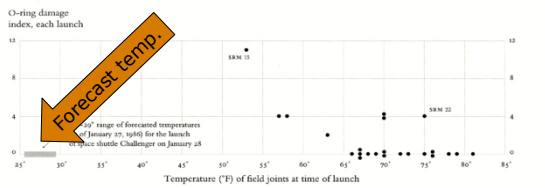
2 of 13 pages of material faxed to NASA by Morton Thiokol [Tufte 1997]

Make a decision: Challenger



- Visualization by Morton-Thiokol engineers
- Shows problems and non-problems
- Temperature not pre-attentive
  - Need to look close and read

Make a decision: Challenger



- **Visualization proposed by [Tufte 97]**
- **Highlight role of temperature**
- **Shows correlation**
- **Extended scale shows extreme current temp**

Information Visualization

- **Power of pre-attentive processing**
  - Absorb mass of information with little effort
- **Different visual variables have different accuracies for different information types**
  - Nominal
  - Ordinal
  - Quantitative
  - Relational (less studied)
- **The right visualization can save lives**
  - Cholera, Crimean War, Challenger
- **The wrong visualization can deceive….**

Tackling the Challenges of Big Data

Big Data Systems

User Interfaces for Data

Information Visualization

THANK YOU

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger

Professor

Massachusetts Institute of Technology

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Lying with Visualizations

David Karger

Professor

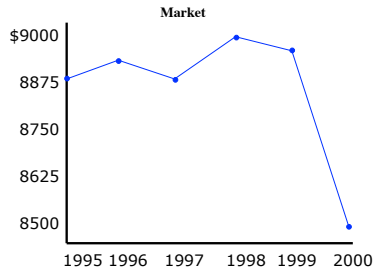
Massachusetts Institute of Technology

Lying Eyes

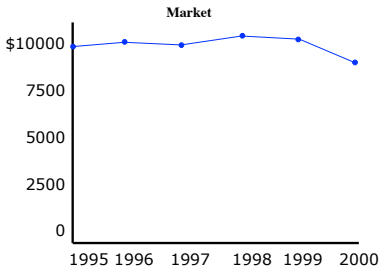
- **Lies, damn lies, statistics, and visualizations.**
- **One's convincing visualization is another's propaganda**
- **Visualization can fail to reflect correct values**
- **Worse, eyes can deceive even when the numbers are "right"**

Photos removed due to copyright restrictions. We are sorry for any inconvenience this may cause.

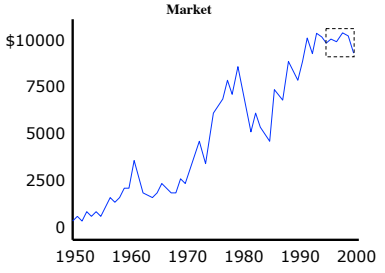
Stock Market Crash?!



Showing entire scale



Shown in context

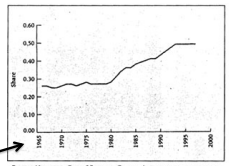


College Costs and Ranks

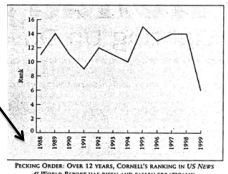
Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Flaws

- **Low rank = good!**
- **Different time scales**
  - Tuition from 1965
  - Rank from 1988
- **Not really tuition**
  - Relative to income
- **Artistic mood**



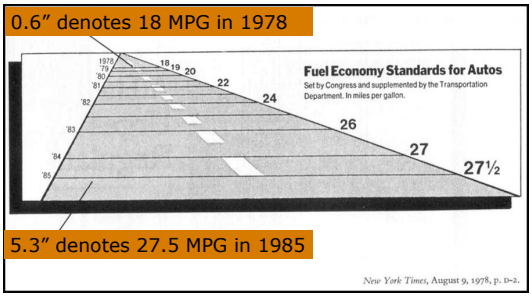
BY THE NUMBERS: OVER 35 YEARS, CORNELL'S TUITION HAS TAKEN AN INCREASINGLY LARGER SHARE OF ITS MEDIAN STUDENT FAMILY INCOME.



PECKING ORDER: OVER 12 YEARS, CORNELL'S RANKING IN *US NEWS & WORLD REPORT* HAS RISEN AND FALLEN ERRATICALLY.

False Perspective

0.6" denotes 18 MPG in 1978



5.3" denotes 27.5 MPG in 1985

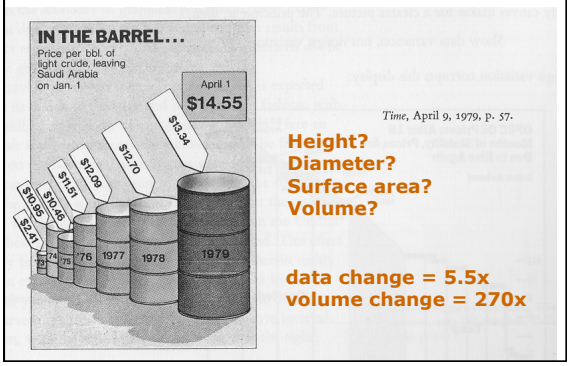
New York Times, August 9, 1978, p. D-2.

MPG change: +52% Length change: +780%

Size Encoding

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

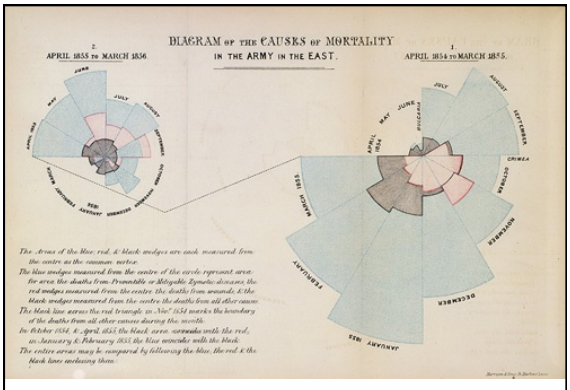
Volume encoding?



Time, April 9, 1979, p. 57.

Height?
Diameter?
Surface area?
Volume?

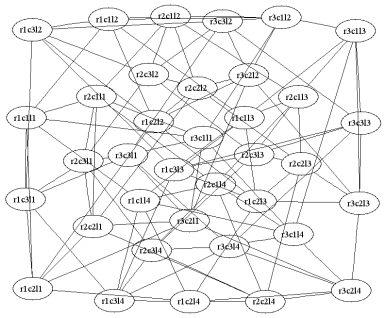
data change = 5.5x
volume change = 270x



1856 Crimean War Death "Coxcomb", Florence Nightingale

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Network Visualization



Core and outliers?

Symmetric 3x3x4 torus

Core artifact of layout algorithms

How not to lie

- **Show entire scale**
- **Show data in context**
- **Consistent, linear scale**
  - Log scale for log data
- **Use visual variables appropriate to data**
  - Don't falsely imply ordinal/quantitative relationships
- **Avoid size encoding**
  - Use height OR width
  - Don't use both for same data attribute
  - Avoid area, volume encoding

Tufte: Graphical Integrity

- **Lie factor = $\frac{\text{size of effect in graph}}{\text{size of effect in data}}$**
- **The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented**
- **Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.**

Tufte: Graphical Integrity

- **Show data variation, not design variation**
  - i.e., don't add decoration just for looks
- **In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units**
- **The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data**
- **Graphics must not quote data out of context**

STOP AND FRISK

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data

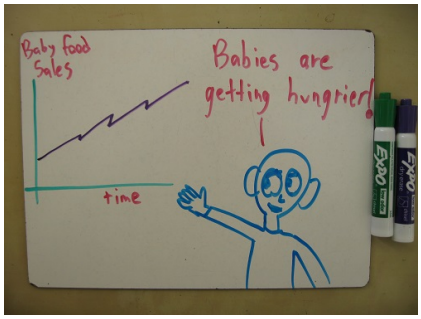
Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.



http://mathwithbaddrawings.com/

Summary

- **Visualizations are powerful communicators**
- **Use that power wisely**
  - Ask what your visualizations imply that isn't true
  - Show entire scale
  - Show data in context
  - Use appropriate visual variables (nominal vs. quant.)
- **Leave art to the artists**
- **Be skeptical of visualizations you see**

Interactivity

- **A visualization can raise questions**
  - Is this true or an artifact of the visualization?
  - Why….
  - What if….
- **If it's static, they're hard to answer**
- **Interactivity can help**
  - Look at data from multiple perspectives
  - Change data, see what happens to visualization
  - Test hypotheses
- **Better for analysis: more informative**
- **Better for communication: more convincing**

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Lying with Visualizations

THANK YOU

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger

Professor

Massachusetts Institute of Technology

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Exploring Data

David Karger

Professor

Massachusetts Institute of Technology

Interactivity

- **A visualization can raise questions**
  - Is this real, or an artifact of the visualization?
  - Why….
  - What if….
- **If it's static, they're hard to answer**
- **Interactivity can help**
  - Look at data from multiple perspectives
  - Change data, see what happens to visualization
  - Test hypotheses
- **Better for analysis: more informative**
- **Better for communication: more convincing**

Why?

- **Interaction lets you navigate a huge space of possible visualizations until you find the one that tells you what you want to know**

Plan

- **Mindset: Exploratory Data Analysis**
  - Roots of data interaction interfaces
- **Interface tactic: Direct Manipulation**
  - The right input mechanism for most users
- **Interface strategy:**
  - Overview first
  - Zoom and filter
  - Details on demand

Exploratory Data Analysis

John Tukey

- **Pioneered use of computer visualization in exploratory data analysis**
- **PRIM-9 System**

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Exploratory vs. Confirmatory

- **Confirmatory data analysis tests a hypothesis**
  - E.g., whether data fits a certain distribution
- **But how do you develop the hypothesis?**
  - Need to explore
- **Visualizations wrong for confirmatory analysis**
  - Too easy to deceive
- **But great for suggesting hypotheses to test**
  - Make sure to confirm in other ways

John Tukey

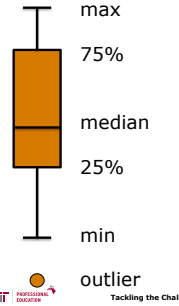
"If we need a short suggestion of what exploratory data analysis is, I would suggest that

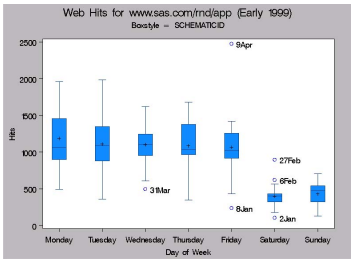
- it is an attitude AND
- a flexibility AND
- some graph paper (or transparencies, or both).

No catalogue of techniques can convey a willingness to look for what can be seen, whether or not anticipated. Yet this is at the heart of exploratory data analysis. The graph paper - and transparencies - are there, not as a technique, but rather as recognition that the picture-examining eye is the best finder we have of the wholly unanticipated."

John Tukey Box Plot

Small-multiples visualization of distributions





John Tukey

- **Computer tool for exploratory data analysis**
- **PRIM-9**
- **1972**

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

- **Can Project, Rotate, Isolate, Mask (filter)**
- **Via direct manipulation (next module)**

John Tukey

Invented FFT

(and "bit", Tukey Range Test, Tukey Lambda Distribution, Tukey's Lemma)

Summary

- **Exploratory data analysis**
  - looking at data without preconceptions
  - distinct from confirmatory analysis
- **Benefits from trying different visualizations**
  - Computer provides great support
- **Remember visualizations can deceive**
  - Need non-visual confirmatory tools
- **Next: what kind of interfaces help explore?**

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Exploring Data

THANK YOU

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger
Professor
Massachusetts Institute of Technology

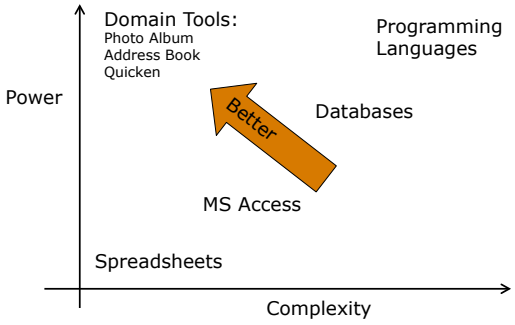
Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Direct Manipulation

David Karger
Professor
Massachusetts Institute of Technology

Goal

- **Want interfaces that permit user to explore**
- **Ideally, think about the data without thinking about the interface**
- **Information visualization user pre-attentive visual features to understand without thinking**
- **What is the analogue for interacting with the data?**

Spectrum of Interface Capabilities



Direct Manipulation Motivation

Advanced Users	Amateurs/Novices
• Think abstractly	• Think concretely
• Have a language of relevant actions	• Know what they want done, but not what to call it
• Can describe a sequence of actions (program syntax)	• Can show, not tell
• Can debug, try again	• Don't know what they did wrong
• Read the manual	• Learn by doing

Direct Manipulation Examples

Advanced Users	Amateurs/Novices
• DOS command line	• Desktop files & folders
• Hotkey shortcuts	• Menus
• HTML/LaTeX/WikiText source code	• WYSIWYG editors
• Write programs	• Record macros
• Write SQL database queries	• Excel & (some) Access
• Zork, Adventure	• Pong, Space Invaders

Direct Manipulation Paradigm

- **Continuous representation of the objects and actions of interest**
- **Physical actions or button presses instead of complex syntax**
- **Rapid**
  incremental
  reversible
  operations whose effect on objects is immediately visible

Shneiderman, Ben (1983). Direct Manipulation: A Step Beyond Programming Languages

Direct Manipulation Motivation

Advanced Users	Amateurs/Novices
• Think abstractly	• Think concretely
• Have a language of relevant actions	• Know what they want done, but not what to call it
• Can describe a sequence of actions (program syntax)	• Can show, not tell
• Can debug, try again	• Don't know what they did wrong
• Read the manual	• Learn by doing

Direct Manipulation Utility

DM interfaces	Amateurs/Novices
• Show data to explore	• Think concretely
• Shows available actions	• Know what they want done, but not what to call it
• Immediate feedback	
• Physical interactions directly manipulate data	• Can show, not tell
• Reversible	• Don't know what they did wrong
• Discoverable actions	• Learn by doing

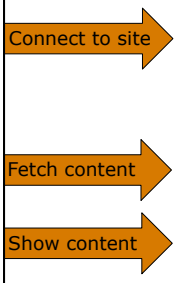
Direct Manipulation Tradeoff

- **Programming language/command line is usually more powerful**
- **Direct manipulation is easier to learn and use**
- **Even experienced users prefer direct manipulation when they can get it**
- **But resort to programming when the DM tools aren't powerful enough**

Example: The Web

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Example: The "Web" in 1987



Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Example: "The Web" in 1987

- **How did the web change things?**
- **No typing connect/fetch commands**
- **Instead, click on what you want**
  - Show, don't tell
- **The web didn't make anything new possible**
- **It just made something simple much simpler via direct manipulation**

Direct Manipulation for Data

Filmfinder: Interactive Scatterplots

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Filmfinder: Interactive Scatterplots

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Interface Choices

- **Scatter plot**
  - two quantitative feature axes: year and popularity
  - hues for (nominal, unordered) movie types
- **Direct manipulation**
  - Buttons for (nominal) categories
  - Sliders for (alphabetically ordered) names
  - Range slider for (quantitative) length
  - Zoom in to smaller date range
  - All actions immediately reversible
- **Initial view gives complete overview**
- **Selecting specific subsets shows subset of scatter points, preserving context**

Pre-attentive Interaction

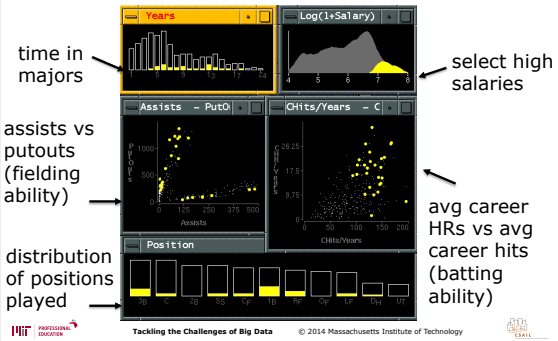
- **Goal of information visualization was output via pre-attentive features**
  - User acquires information by sight, without thought
- **Goal of direct manipulation is input via "pre-attentive" features**
  - User must make a decision to do something
  - But doesn't want to think about how to do it
  - (video games truly pre-attentive
  - perhaps data interfaces someday as well)
- **Direct manipulation tools for data is as important as the desktop and WYSIWYG editors were**

Data Linking and Brushing

- **Direct manipulation technique for filtering**
- **Link:**
  - Associate two or more visualizations of same data
- **Brush:**
  - Select a subset of data points
  - See role played by subset of points in linked views
- **Redoes visualization (output) as filter (input)**
- **Example systems**
  - Datadesk
  - Graham Wills' EDV system
  - Ahlberg & Shneiderman's IVEE (Spotfire)

Baseball: Scatterplots, Histograms

(from Eick & Wills 95)



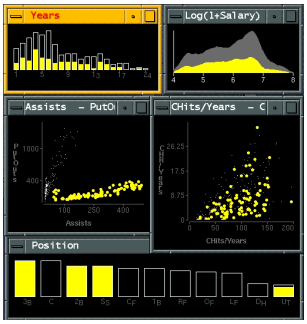
What was learned from interaction

- Seems impossible to earn a high salary in the first three years
- High salaried players have a bimodal distribution (peaking around 7 & 13 yrs)
- Hits/Year a better indicator of salary than HR/ Year
- High paid outlier with low HR and medium hits/year. Reason: person is player-coach
- There seem to be two differentiated groups in the put-outs/assists category (but not correlated with salary) Why?

Linking assists to position played
[Eick & Wills 95]

Select one "cluster" of assists scatterplot

Discover position correlation



Tackling the Challenges of Big Data
Big Data Analytics
User Interfaces for Data
Direct Manipulation

THANK YOU

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger

Professor

Massachusetts Institute of Technology

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Overview, Zoom, Filter, Details

David Karger

Professor

Massachusetts Institute of Technology

Interaction Strategy

- **[Schneiderman '96]**
- **Overview first**
  - Give user a starting point
- **Filter**
  - Ways to restrict to interesting sets of points
- **Pan & Zoom**
  - Let them hone in on an interesting area
  - Get more detail about that area
- **Details on demand**
  - Extreme zoom
  - See everything about one or a few items

Interaction Strategy

- **Overview**
  - Don't start user with a blank screen
  - Give them an idea of what the data is and what they can do
  - The first impression---what should it be?
- **Filter**
  - Saw buttons to filter on type
  - Sliders to choose ranges of values
  - Linking and brushing

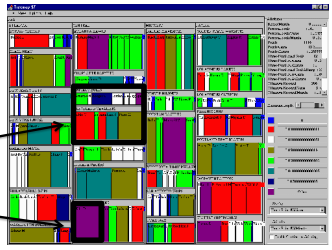
Pan & Zoom

- **Direct manipulation filtering**
- **Spatial metaphor**
  - Movement is pre-attentive?
- **Apply to features that have been mapped to spatial coordinates (positions)**
- **Shift in spatial coordinates defines shift in parameters**
- **Animation (smooth motion) helps preserve context as data points move**
- **Obvious application: spatial data (maps)**
- **But also applies to more abstract features**

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Treemap

- Johnson & Schneiderman '91
- Interface to hierarchical data
- Subgroups are rectangular regions
- Sub-subgroups inside
- Click area to zoom
- Hover for details



Treemap of a directory hierarchy

Treemap

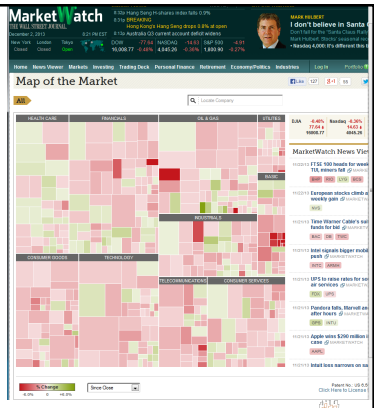
Display a hierarchical data set

Click on a sub-area to zoom in

Hover over one item to get details

Breadcrumb trail

2-color scale



Treemap

Display a hierarchical data set

Click on a sub-area to zoom in

Hover over one item to get details

Breadcrumb trail

2-color scale

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Treemap

- Display a hierarchical data set
- **Click on a sub-area to zoom in**
- Hover over one item to get details
- Breadcrumb trail
- 2-color scale

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Treemap

- Display a hierarchical data set
- Click on a sub-area to zoom in
- **Hover over one item to get details**
- Breadcrumb trail
- 2-color scale

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Treemap

- Display a hierarchical data set
- Click on a sub-area to zoom in
- Hover over one item to get details
- **Breadcrumb trail**
- 2-color scale

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Treemap

Display a hierarchical data set

Click on a sub-area to zoom in

Hover over one item to get details

Breadcrumb trail

2-color scale

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Overview, Zoom, Filter, Details

THANK YOU

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger

Professor

Massachusetts Institute of Technology

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Data Interfaces on the Web

David Karger

Professor

Massachusetts Institute of Technology

Overview

- **Many web sites are fronts for large data sets**
- **Common paradigm has emerged for presenting them**
  - Useful, because users know what to do
  - Matches Schneiderman's workflow
- **Views**
  - Overview of all items
- **Faceted browsing**
  - Direct-manipulation filtering
- **Templates**
  - Uniform presentation of individual item details

Views

- **Aggregate overview of all items**
- **Sortable list, map, table, grid…**

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Templates

- **Detail view on individual items**
- **In list or grid, may show for each item**
- **Popup on demand for maps, grids**

Templates

- **Uniform presentation of all items**
- **Easier for us to parse**
  - Quickly get used to which attribute goes where
  - Leverages power of (pre-attentive) position
- **Pleasant user experience**
  - Could show same info in a table, but that's ugly
- **Usually generated by templating engine**
  - Easy maintenance: just change the template

Faceted Browsing

- **Direct manipulation filtering**
- **Check-off values to filter**

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Faceted Browsing

- **Dominant filtering paradigm**
- **Give users a list of values for each attribute**
- **They pick some**
- **Collection restricts to items matching those values**

vs. Hierarchical Browsing

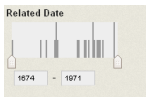
- **Original Yahoo site**
- **Navigate a hierarchy**
- **Links to children**
- **Forces user to descend hierarchy in order you defined**
- **Filters that don't match hierarchy unavailable**
- **Supports "zoom", but not filter**

Facet Benefits

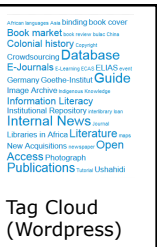
- **Direct manipulation**
  - Shows user what filtering options are available
  - User selects by click, unselects by another
  - Immediate undo if dislike outcome
- **Independent attributes**
  - Unlike hierarchy, user can mix filters arbitrarily
- **Apply in any order**
  - Result is intersection of all facet constraints
  - Final result independent of order of application

Facet Variety

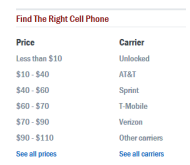
Icons (Amazon)



Slider+Histogram (Lib. of Congress



Tag Cloud (Wordpress)



Numeric Range, Enumerated List (CNET)

Matrix (Orbitz)

Common Practices

- **Hide values yielding no results**
  - Avoids frustrating dead ends
  - Update as other facets are applied
- **Even better, show count of matching results**
  - Lets user see what will happen if they select
  - Using numbers (in a list of values)
  - Or size (in a tag cloud)
- **Breadcrumbs**
  - Show user a trail of user selections
  - They can back up
  - Or easily remove specific constraints

Views+Facets+Templates

- **Users know what to expect**
- **So they know how to interact with new sites**
- **Are all data web sites done this way?**

General Enough?

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Summary

- **Most web data sites share common paradigm**
- **Standard workflow: overview, filter, details**
- **Collection overview (list, map, grid)**
- **Facets for filtering**
- **Templates show item details uniformly**

Impoverished Data Visualization?

- **Is sameness of all these sites a good thing?**
- **Most information presenters are not ambitious**
- **Carefully designed, domain- and task-specific information interactions will always be superior**
- **But powerful lowest common denominator**
- **People's experience of it makes it more powerful**
  - Leverage expectations
  - No need to learn new site

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Data Interfaces on the Web

THANK YOU

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger

Professor

Massachusetts Institute of Technology

PROFESSIONAL EDUCATION

CSAIL

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces
Can End Users Create Data Interfaces?

David Karger

Professor

Massachusetts Institute of Technology

Building the Data Web

- **Data-backed web sites created by professionals**
  - Back-end database engineers
  - Front-end designers
- **Regular users can't do the same**
  - They author static text pages
  - Or put data on sites others built
- **What if you want to present your data your way?**
  - E.g., create a combined presentation of photos and recipes organized by food in the pictures?

Idea

- **They're all the same**
  - So is programming really needed?
- **Extend HTML to describe standard elements**
  - Templates, Views, Facets
  - Each refers to specific data fields (columns)
- **Now creating a data visualization is just HTML authoring**
  - Edit the source code
  - Or copy and modify someone else's
  - Or use a WYSIWYG editor
- **Anyone can create their own visualizations**

Details

- **Data in a spreadsheet**
  - [first-name, last-name, age, home-latlng]
- **Templates**
  -
    ,

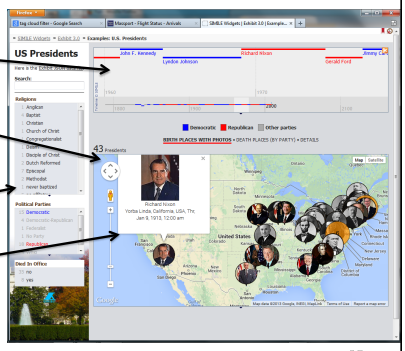
- **Views**
  -
  -
- **Facets**
  -

Prototype: Exhibit

- **Specific data/viz HTML vocabulary extension**
  - And a Javascript library to interpret it
- **Application independent**
  - Fits any tool that takes HTML, e.g. HTML editor
- **Pure client side**
  - No need to design/admin server
- **Freely interleave data with other HTML**
  - Complete control of design
  - Integration with whatever other elements you like
- **Author one HTML page, get interactive data visualization**

Exhibit: US Presidents

- **Timeline view**
- **Map view**
- **Facets**
- **Template**



Prototype: Exhibit

- **Javascript library**
  - Open source
  - Hosted at http://simile-widgets.org/exhibit
- **Currently scales to ~100,000 items**
- **Deployed 2007**
  - ~1900 exhibits on 800 domains
  - Millions of views

Music

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Cars

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Art

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Hobby Stores

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Science

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

PhD Thesis

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Art History

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Rental Apartments

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Data.gov

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Prarie Home Companion Tickets

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

NGOs

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Newspapers

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Libraries

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Sports

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Strange Hobbyists

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Restaurants

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Summary

- **Sameness of data web sites is an opportunity for simplification**
- **Instead of programming a site, standardize around basic elements**
  - Views, templates, facets
- **Any user can create their own visualization**
  - WYSIWYG html authoring
- **Permits more effective communication**

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces
Can End Users Create Data Interfaces?

THANK YOU

Haystack

- **Semantic Web app before the Semantic Web**
  - Research paid for by the machines I didn't buy
- **Entity-relation data model**
- **"Lenses" to display individual items**
  - Specification of which properties, and their layout
- **"Views" of collections**
  - E.g. lists, thumbnails, tabular
- **"Facets" to filter items**
- **When RDF invented, became Haystack model**
  - And haystack became an early "semantic desktop"

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Writing a Brain Research Paper

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Adding "Things to Do" Region

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Revised Environment

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces
Users Creating Interfaces

THANK YOU

Tackling the Challenges of Big Data
User Interfaces for Data

David Karger
Professor
Massachusetts Institute of Technology

Tackling the Challenges of Big Data
Big Data Systems
User Interfaces for Data
Beyond the Spreadsheet

David Karger
Professor
Massachusetts Institute of Technology

Databases

- **The dominant paradigm for data management**
- **Operate on tables**
  - Columns are attributes of interest
  - One row per "item"
    - Person: Name, Age, Employer
    - Company: Name, Sector, Earnings
- **Basic operations**
  - Select some rows by filtering value in columns
    - All rows where age=43
  - Join multiple tables
    - Find employees of companies earning more than $1M
- **SQL is the dominant programming language**

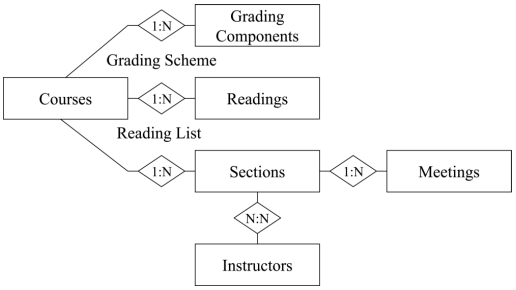
Spreadsheets

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Spreadsheets

- **Dominant database tool for end users**
  - Direct manipulation of rows
  - Study found most spreadsheets have no formulas!
- **But limited**
  - Only one table in view at a time
  - No joins
  - Hard to represent connections between tables
  - No support for many-many relationships
- **Goal: add power but preserve look and feel**
  - Nested views
  - Linkage/navigation between multiple tables

One-to-Many and Many-to-Many Relationships



Spreadsheet Solutions

- **Multiple worksheets**
  - One per table
  - Effective to hold data
  - But forces painful navigation
  - Little support for moving between tables
- **Comma-separated lists**
  - Offer only one level of multiplicity
  - If each of multiple items has multiple other items, trouble

Alternative: Related Worksheets

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Access a database with many-to-many relationships through a general-purpose, spreadsheet-like UI

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

"Related Worksheets" application at startup

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Creating a new worksheet

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

After entering some simple, tabular data

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Concept 1: **Data Types** for Worksheet Columns

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Concept 2: **Multiple-Value Types**

Concept 3: **Reference Types**
(Each cell in column refers to row in a different worksheet)

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Concept 3: **Reference Types**
Reference values displayed recursively

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

4th New Concept: Relationships are **bidirectional**

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

1

Concept 4: Relationships are **bidirectional**

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

2

Concept 4: Relationships are **bidirectional**

1

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

Teleport
(Press Ctrl+Space)

2

Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

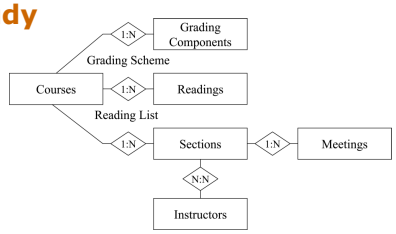
Related Worksheets track one-to-many/many-to-many relationships from within a spreadsheet-like user interface

User Study

- **Hypothesis: Excel-proficient users will be faster at lookup (read-only) tasks on a database stored in normalized form in our system vs. MS Excel**

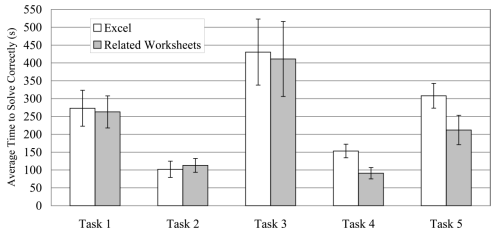
Photos removed due to copyright restrictions.
We are sorry for any inconvenience this may cause.

User Study



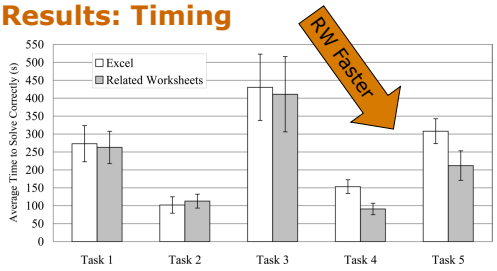
#	*Description*
1	Find a course that is taught by Harry Morrill.
2	In the course "MUS 105: Music Theory Through Performance and Composition," what percentage of the final grade is derived from the Midterm Exam?
3	Find a course that satisfies the "LA" (Literature and the Arts) distribution area, with a lecture (meetings belonging to a section denoted "L01") that starts after noon (after 12.00).
4	What is the e-mail address of the instructor who teaches "KOR 107: Intermediate Korean II"?
5	Who teaches the precept section of "HIS 383: The United States Since 1920" that meets on Wednesdays at 10am?

Results: Timing



#	*Description*
1	Find a course that is taught by Harry Morrill.
2	In the course "MUS 105: Music Theory Through Performance and Composition," what percentage of the final grade is derived from the Midterm Exam?
3	Find a course that satisfies the "LA" (Literature and the Arts) distribution area, with a lecture (meetings belonging to a section denoted "L01") that starts after noon (after 12.00).
4	What is the e-mail address of the instructor who teaches "KOR 107: Intermediate Korean II"?
5	Who teaches the precept section of "HIS 383: The United States Since 1920" that meets on Wednesdays at 10am?

Results: Timing



#	Description
1	Find a course that is taught by Harry Morrill.
2	In the course "MUS 105: Music Theory Through Performance and Composition," what percentage of the final grade is derived from the Midterm Exam?
3	Find a course that satisfies the "LA" (Literature and the Arts) distribution area, with a lecture (meetings belonging to a section denoted "L01") that starts after noon (after 12.00).
4	What is the e-mail address of the instructor who teaches "KOR 107: Intermediate Korean II"?
5	Who teaches the precept section of "HIS 383: The United States Since 1920" that meets on Wednesdays at 10am?

Correctness and Features Used

		Task #				
	N	*1*	*2*	*3*	*4*	*5*
Excel	18	18	18	14	17	12
RS	18	18	18	14	17	15

Table 4. Number of subjects who solved each task correctly.

Feature	*Excel*	*Related Worksheets*
Search	76%	53%
Mark Rows	20%	12%
Filter	73%	
Sort	7%	
Teleport		44%
Show/Hide Columns		18%

Table 5. Number of users observed to make use of specific features or techniques during the user study.

Correctness and Features Used

		Task #				
	N	*1*	*2*	*3*	*4*	*5*
Excel	18	18	18	14	17	12
RS	18	18	18	14	17	15

Fewer Errors

Table 4. Number of subjects who solved each task correctly.

Feature	*Excel*	*Related Worksheets*
Search	76%	53%
Mark Rows	20%	12%
Filter	73%	
Sort	7%	
Teleport		44%
Show/Hide Columns		18%

Table 5. Number of users observed to make use of specific features or techniques during the user study.

Conclusion

- **Spreadsheets are the dominant database tool**
- **But sacrifice key capabilities**
  - Multiple tables
  - Multiple values
  - Joins
- **Enhance spreadsheet paradigm with**
  - Multi-valued cells
  - Nested views of connected information
  - Teleportation to related worksheet cells
- **User Study shows improvements over Excel**
  - Faster
  - Fewer errors

Tackling the Challenges of Big Data

Big Data Systems

User Interfaces for Data

Beyond the Spreadsheet

THANK YOU

PROFESSIONAL EDUCATION

CSAIL

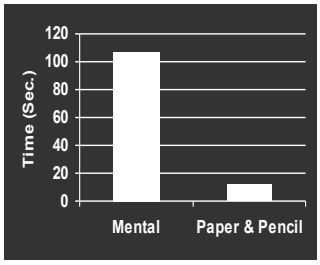
End

Expand memory: Multiplication

```
  34
x 72
----
```


Expand memory: Multiplication

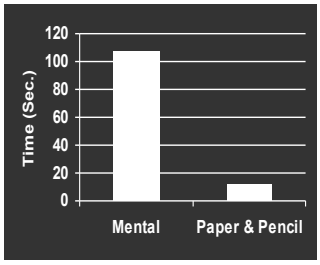
```
  34
x 72
---
  68
2380
---
2448
```



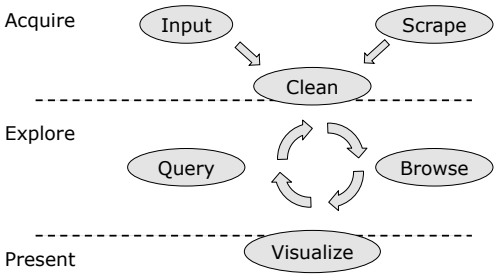
Expand memory: Multiplication

```
  34
x 72
---
  68
2380
---
2448
```

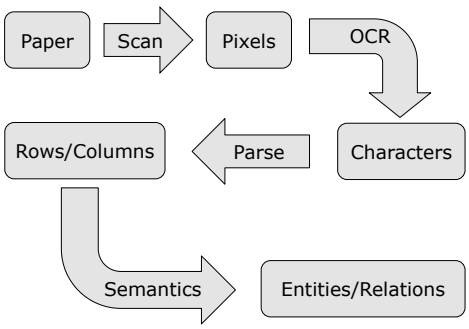
Picture of 2 Numbers



Stages of Information Interaction

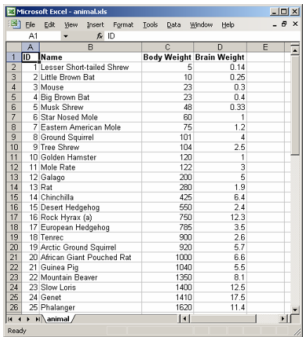


Adding Structure to Raw Data



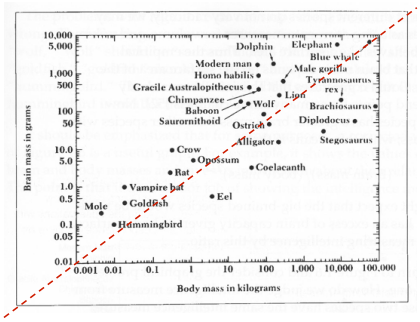
BRAINS

The most powerful brain?

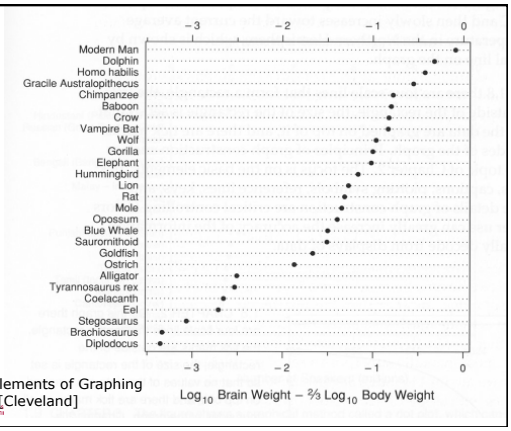


ID	Name	Body Weight	Brain Weight
1	Lesser Short-tailed Shrew	5	0.14
2	Little Brown Bat	10	0.25
3	Mouse	23	0.3
4	Big Brown Bat	23	0.4
5	Musk Shrew	48	0.33
6	Star Nosed Mole	60	1
7	Eastern American Mole	75	1.2
8	Ground Squirrel	101	4
9	Tree Shrew	104	2.5
10	Golden Hamster	120	1
11	Mole Rate	122	3
12	Galago	200	5
13	Rat	280	1.9
14	Chinchilla	425	6.4
15	Desert Hedgehog	550	2.4
16	Rock Hyrax (a)	750	12.3
17	European Hedgehog	785	3.5
18	Tenrec	900	2.6
19	Arctic Ground Squirrel	920	5.7
20	African Giant Pouched Rat	1000	6.6
21	Guinea Pig	1040	5.5
22	Mountain Beaver	1350	8.1
23	Slow Loris	1400	12.5
24	Genet	1410	17.5
25	Phalanger	1620	11.4

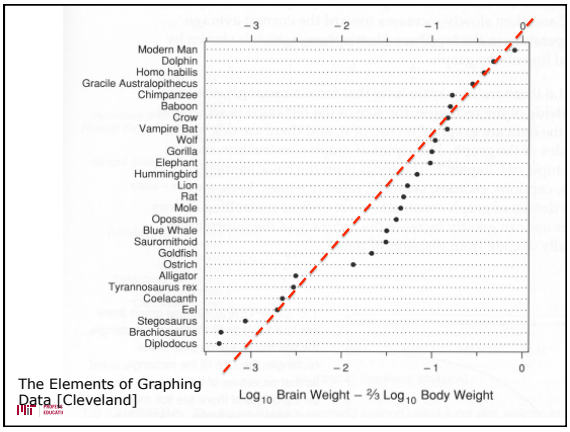
The most powerful brain?



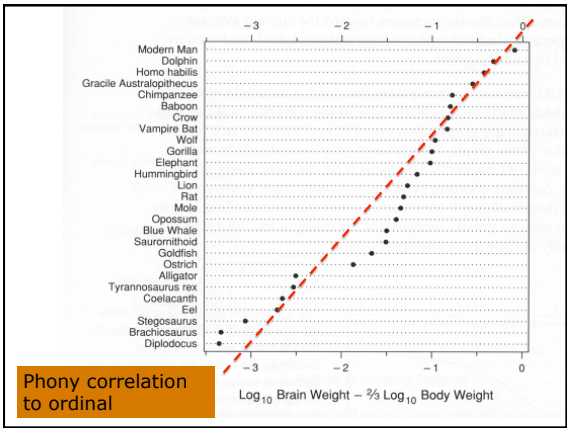
The Dragons of Eden [Sagan]



The Elements of Graphing Data [Cleveland]



The Elements of Graphing Data [Cleveland]



Phony correlation to ordinal

The Real Story

- **Really just a number line**
- **Marks spaced vertically for readability**