Information Visualization

Jing Yang

Spring 2008

1

Introduction: What is Information Visualization

2

Data Explosion

- Society is more complex: computers, internet, web, ...
- How much data?
  - Between 1 and 2 exabytes of unique info produced per year
    - 1000000000000000000 bytes
    - 250 meg for every man, woman and child
    - Printed documents only .003% of total

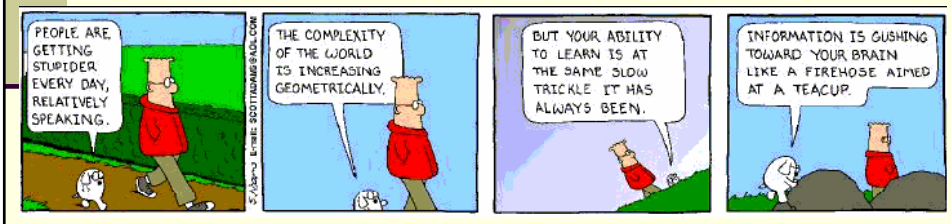
Peter Lyman and Hal Varian, 2000
Cal-Berkeley, Info Mgmt & Systems
www.sims.berkeley.edu/how-much-info

Slide courtesy of John Stasko

3

Data Overload

- Confound: How to make use of the data
  - How do we make sense of the data?
  - How do we harness this data in decision-making processes?
  - How do we avoid being overwhelmed?

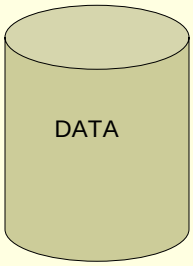


Slide courtesy of John Stasko

4

The Problem: Smell it, taste it?

Web,
Books,
Papers,
Game scores,
Scientific data,
Biotech,
Shopping,
Stock/finance,
News



Data Transfer

- Vision: 100 MB/s
- Ears: <100 b/s
- Smell
- Taste

Slide courtesy of John Stasko

5

Example

- A simple experiment

Count the number of 3s in the following text:

1243564289783012432567213524536912638137978021837459026

6

Visualization Makes Difference

- A simple experiment

Count the number of 3s in the following text:

1243564289783012432567213 52
453691263813797802183745902

7

Human Vision

- Highest bandwidth sense
- Fast, parallel
- Pattern recognition
- Pre-attentive
- Extends memory and cognitive capacity
- People think visually

Slide courtesy of John Stasko

The Challenge

- Transform the data into information (understanding, insight) thus making it useful to people

Slide courtesy of John Stasko

Example

Questions:

Which state has the highest income?

Is there a relationship between income and education?

Are there any outliers?

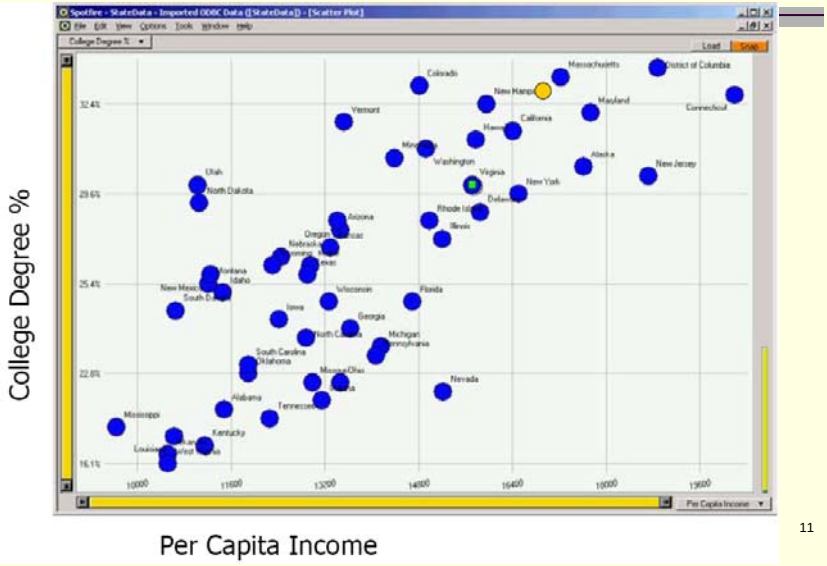
Table - StateData ()

Load | Snap

State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	33.9%	14821
Connecticut	33.8%	20189
Delaware	27.9%	15854
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
Idaho	25.2%	11457
Illinois	26.8%	15201
Indiana	20.9%	13149
Iowa	24.5%	12422
Kansas	26.5%	13300
Kentucky	17.7%	11153
Louisiana	19.4%	10635
Maine	25.7%	12957
Maryland	31.7%	17730
Massachusetts	34.5%	17224
Michigan	24.1%	14154
Minnesota	30.4%	14389
Mississippi	19.9%	9648
Missouri	22.3%	12989
Montana	25.4%	11213
Nebraska	26.0%	12452
Nevada	21.5%	15214
New Hampshire	32.4%	15959
New Jersey	30.1%	18714
New Mexico	25.5%	11246
New York	29.6%	16501
North Carolina	24.2%	12885
North Dakota	28.1%	11051
Ohio	22.3%	13461
Oklahoma	22.8%	11893
Oregon	27.5%	13418
Pennsylvania	23.2%	14068
Rhode Island	27.5%	14981
South Carolina	23.0%	11897
South Dakota	24.6%	10661
Tennessee	20.1%	12255
Texas	25.5%	12904
Utah	30.0%	11029
Vermont	31.5%	13527
Virginia	30.0%	15713
Washington	30.9%	14923
West Virginia	16.1%	10520
Wisconsin	24.9%	13276
Wyoming	25.7%	12311

Example courtesy of Chris North

Visualize the Data



11

Illustrate Our Approach

- Provide tools that present data in a way to help people understand and gain insight from it
- Cliches
  - "Seeing is believing"
  - "A picture is worth a thousand words"

Slide courtesy of John Stasko

12

Visualization

- Visualization - *the use of computer-supported, interactive visual representations of data to amplify cognition*

  From [Card, Mackinlay Shneiderman '98]

- Often thought of as process of making a graphic or an image
- Really is a cognitive process
  - Form a mental image of something
  - Internalize an understanding
- "The purpose of visualization is insight, not pictures"

  Insight: discovery, decision making, explanation

Slide courtesy of John Stasko

13

What is Information Visualization

Information Visualization is NOT Scientific Visualization

Scientific Visualization is primarily related to and represents something physical or geometric

- Examples:
  - Air flow over a wing
  - Stresses on a girder
  - Weather over Pennsylvania

14

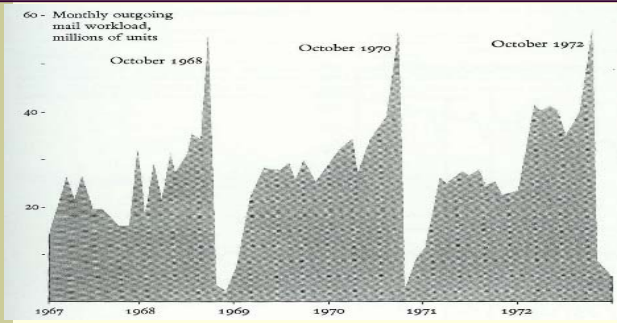
What is Information Visualization

- It is about abstract data

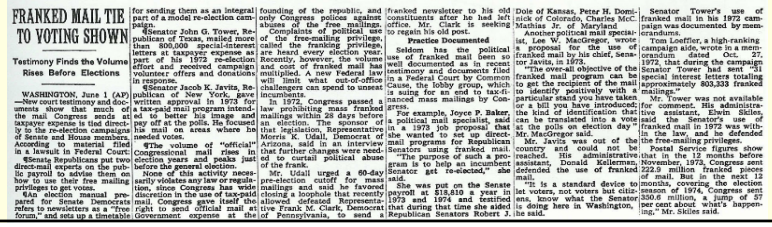
X

15

InfoVis Is About Numerical Data



This graphic is worth at least 700 words



FRANKED MAIL TIE TO VOTING SHOWN

Testimony Finds the Volume Rises Before Elections

WASHINGTON, June 1 (AP) —New court testimony and documents show that much of the mail Congress sends at taxpayer expense is tied directly to the re-election campaigns of Senate and House members. According to material filed in a lawsuit in Federal Court:

¶Senate Republicans put two direct-mail experts on the public payroll to advise them on how to use their free mailing privileges to get votes.

¶An election manual prepared for Senate Democrats refers to newsletters as a "free forum," and sets up a timetable for sending them as an integral part of a model re-election campaign.

¶Senator John G. Tower, Republican of Texas, mailed more than 800,000 special-interest letters at taxpayer expense as part of his 1972 re-election effort and received campaign volunteer offers and donations in response.

¶Senator Jacob K. Javits, Republican of New York, gave written approval in 1973 for a tax-paid mail program intended to better his image and pay off at the polls. He focused his mail on areas where he needed votes.

¶The volume of "official" Congressional mail rises in election years and peaks just before the general election.

None of this activity necessarily violates any law or regulation. Since Congress has wide discretion in the use of tax-paid mail Congress gave itself the right to send official mail at Government expense at the founding of the republic, and only Congress polices against abuses of the free mailings.

Complaints of political use of the free-mailing privilege, called the franking privilege, are heard every election year. Recently, however, the volume and cost of franked mail has multiplied. A new Federal law will limit what out-of-office challengers can spend to unseat incumbents.

In 1972, Congress passed a law prohibiting mass franked mailings within 28 days before an election. The sponsor of that legislation, Representative Morris K. Udall, Democrat of Arizona, said in an interview that further changes were needed to curtail political abuse of the frank.

Mr. Udall urged a 60-day pre-election cutoff for mass mailings and said he favored closing a loophole that recently allowed defeated Representative Frank M. Clark, Democrat of Pennsylvania, to send a franked newsletter to his old constituents after he had left office. Mr. Clark is seeking to regain his old post.

Practice Documented

Seldom has the political use of franked mail been so well documented as in recent testimony and documents filed in a Federal Court by Common Cause, the lobby group, which is suing for an end to tax-financed mass mailings by Congress.

For example, Joyce P. Baker, a political mail specialist, said in a 1973 job proposal that she wanted to set up direct-mail programs for Republican Senators using franked mail. "The purpose of such a program is to help an incumbent Senator get re-elected," she said.

She was put on the Senate payroll at $18,810 a year in 1973 and 1974 and testified that during that time she aided Republican Senators Robert J. Dole of Kansas, Peter H. Dominick of Colorado, Charles McC. Mathias Jr. of Maryland

Another political mail specialist, Lee W. MacGregor, wrote a proposal for the use of franked mail by his chief, Senator Javits, in 1970.

"The over-all objective of the franked mail program can be to get the recipient of the mail to identify positively with a particular stand you have taken or a bill you have introduced; the kind of identification that can be translated into a vote at the polls on election day," Mr. MacGregor said.

Mr. Javits was out of the country and could not be reached. His administrative assistant, Donald Kellerman, defended the use of franked mail.

"It is a standard device to let voters, not voters but citizens, know what the Senator is doing here in Washington, he said.

Senator Tower's use of franked mail in his 1972 campaign was documented by memorandums.

Tom Loeffler, a high-ranking campaign aide, wrote in a memorandum dated Oct. 27, 1972, that during the campaign Senator Tower had sent "31 special interest letters totaling approximately 803,333 franked mailings."

Mr. Tower was not available for comment. His administrative assistant, Elwin Skiles, said the Senator's use of franked mail in 1972 was within the law, and he defended the free-mailing privileges.

Postal Service figures show that in the 12 months before November, 1973, Congress sent 722.9 million franked pieces of mail. But in the next 12 months, covering the election season of 1974, Congress sent 350.6 million, a jump of 57 per cent about what's happening," Mr. Skiles said.

16

InfoVis Is About Ordinal Data

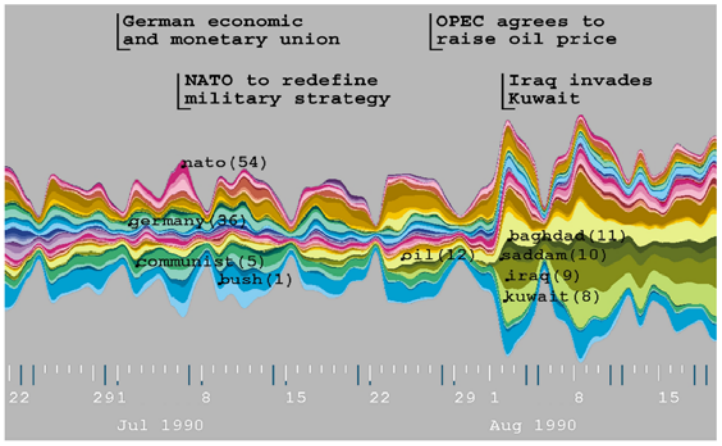
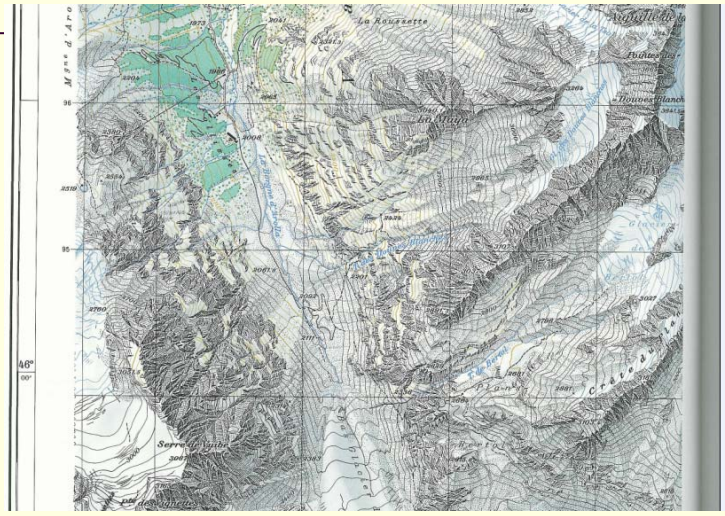


Figure 3: AP data from July - August 1990. A wide current in the river indicates heavy use of a topic, while changes in color distribution correlate to changes in themes.

ThemeRiver: Visualizing Theme Changes Over Time [Havre et al. Infovis 00]

17

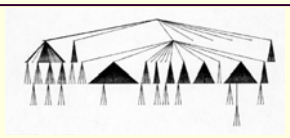
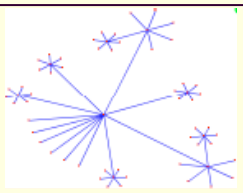
InfoVis Is About Nominal Data

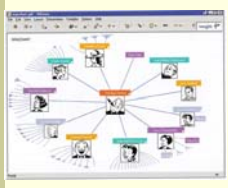


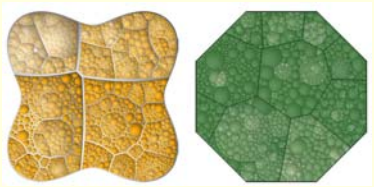
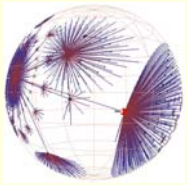
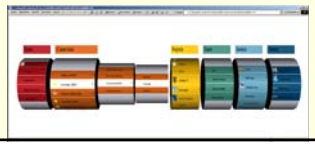
Swiss mountain map, L. Matterhorn, 1983

18

InfoVis Is About Structured Data

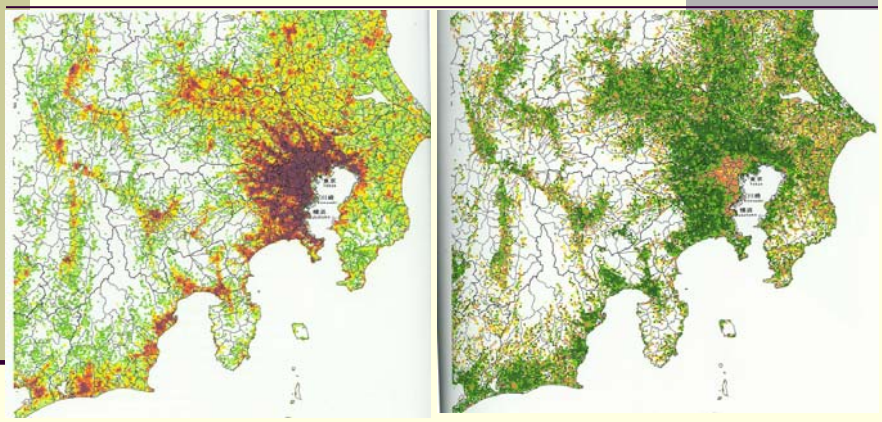






19

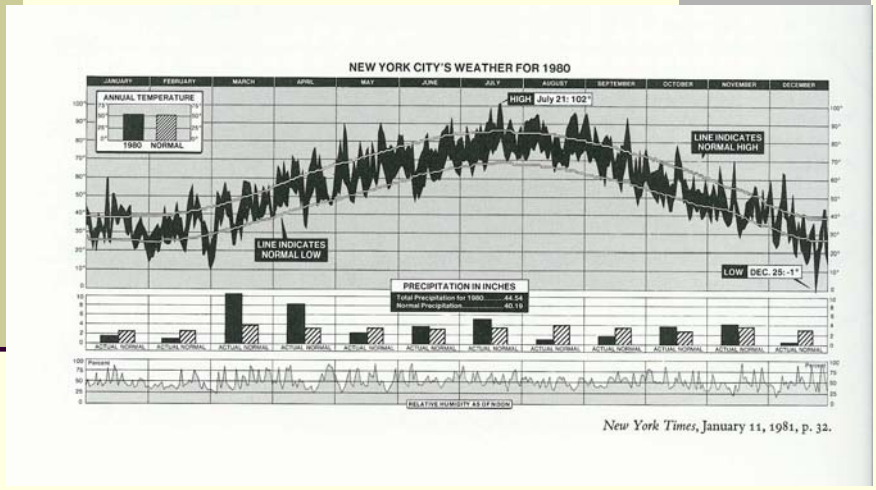
InfoVis Is About Space



Statistics Bureau, Tokyo, 1985

20

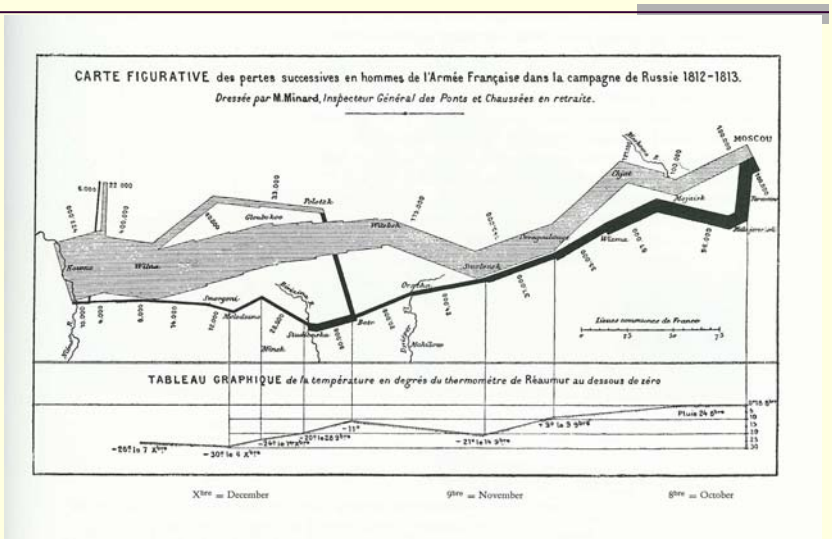
InfoVis Is About Time



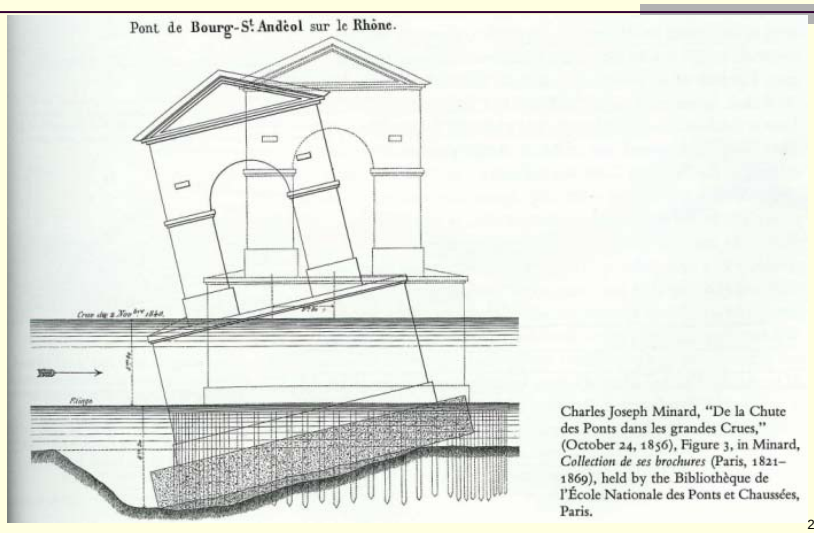
New York Times, January 1982

21

InfoVis Is About Space and Time



InfoVis Is About Change



Charles Joseph Minard, "De la Chute des Ponts dans les grandes Crues," (October 24, 1856), Figure 3, in Minard, *Collection de ses brochures* (Paris, 1821–1869), held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

23

InfoVis Is About Motion and Process

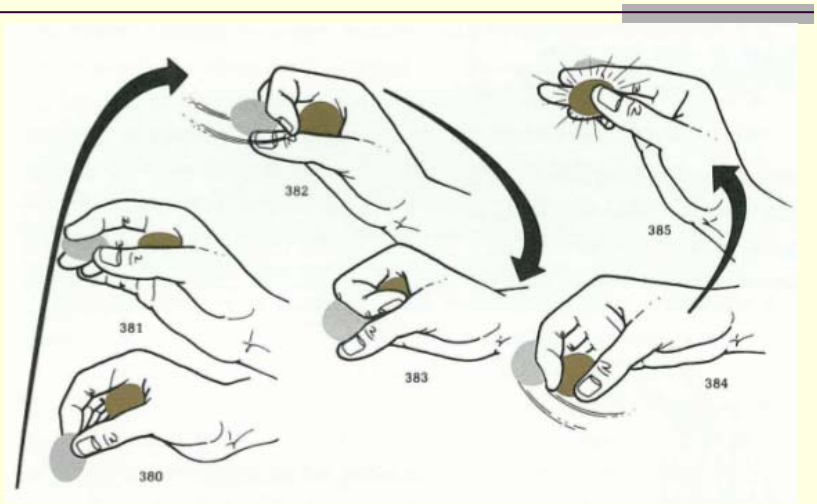
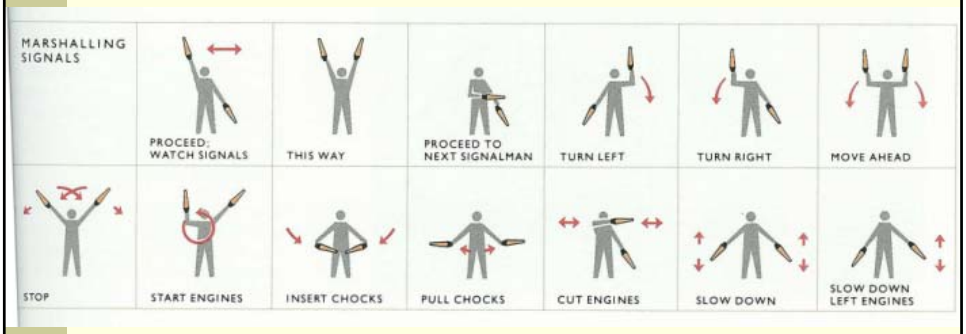


Illustration of magic turning a silver coin into a copper coin

24

InfoVis Is About All Kinds of Information



25

What is Information Visualization

- It is about analyzing, communicating, and decision making

Of all method for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful - E. Tufte

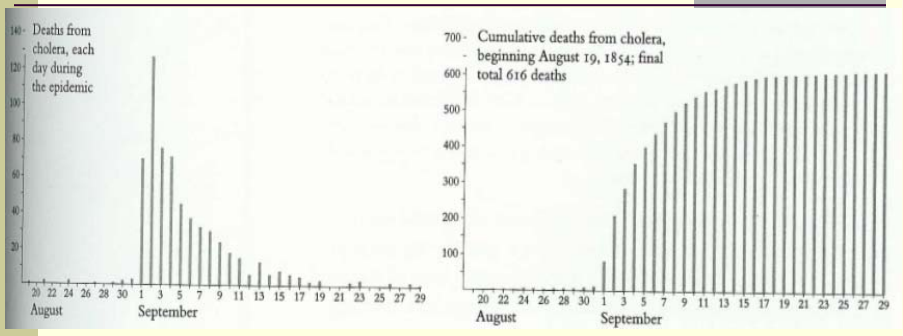
26

Story 1: Cholera

- Location: London
- Time: August and September 1854
- Event: Cholera broke out in the Broad Street area. Dr. John Snow suspected that the water supply was the cause.

27

Story 1: Cholera

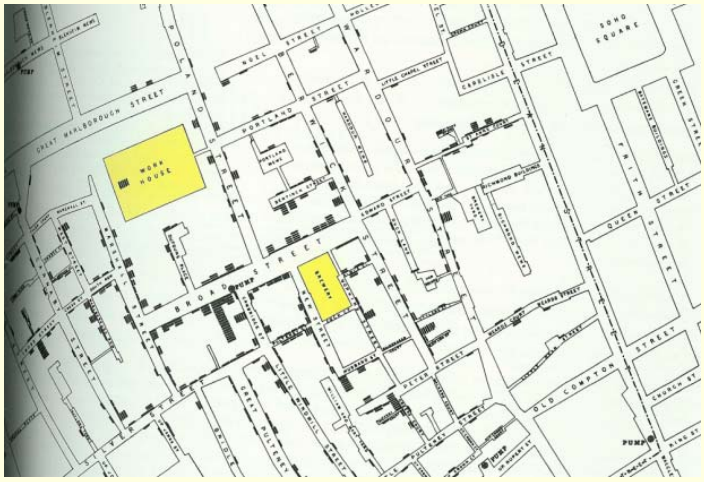


Will this figure help?

28

Story 1: Cholera

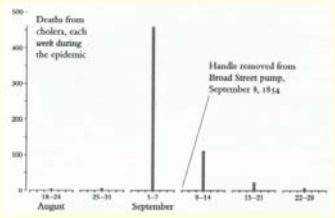
- Dr. Snow's approach

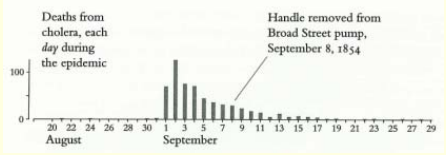


29

Story 1: Cholera

- Dr. Snow's conclusion
  - The water pump at the Broad Street was contaminated
- Consequence ?
  - The epidemic ended after the handle of that water pump was removed.





30

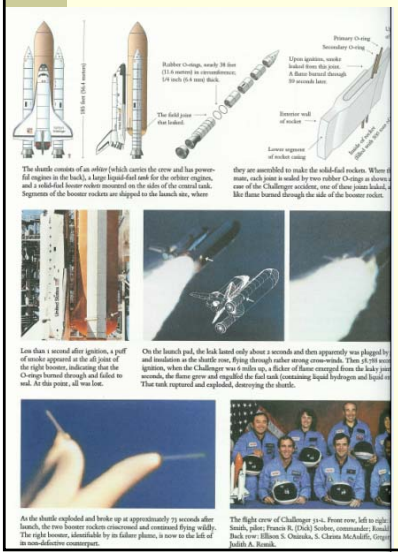
Story 1: Cholera

Lessons:

- 1. Place data in appropriate context for accessing cause and effect
- 2. Make quantitative comparisons
- 3. Consider alternative explanations and contrary cases

31

Story 2: Challenger



On January 28, 1986, the space shuttle Challenger exploded and seven astronauts died because two rubber O-rings leaked

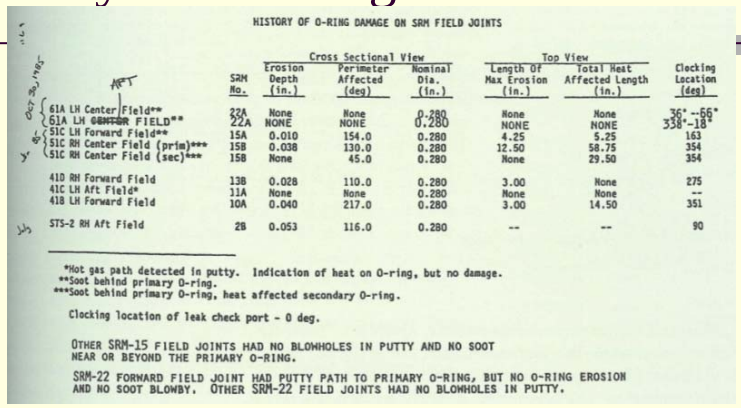
32

Story 2: Challenger

- Location: USA
- Time: January 27,1986
- Event: Engineers in Morton Thiokol suspected that rubber O-rings would fail due to cold weather the next day and required to delay the launch. They submitted 13 charts to NASA to illustrate their viewpoint.

33

Story 2: Challenger



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)
22A	None	None	0.280	None	None	36°-66°
22A	NONE	NONE	0.280	NONE	NONE	338°-18°
15A	0.010	154.0	0.280	4.25	5.25	163
15B	0.038	130.0	0.280	12.50	58.75	354
15B	None	45.0	0.280	None	29.50	354
13B	0.028	110.0	0.280	3.00	None	275
11A	None	None	0.280	None	None	--
10A	0.040	217.0	0.280	3.00	14.50	351
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.

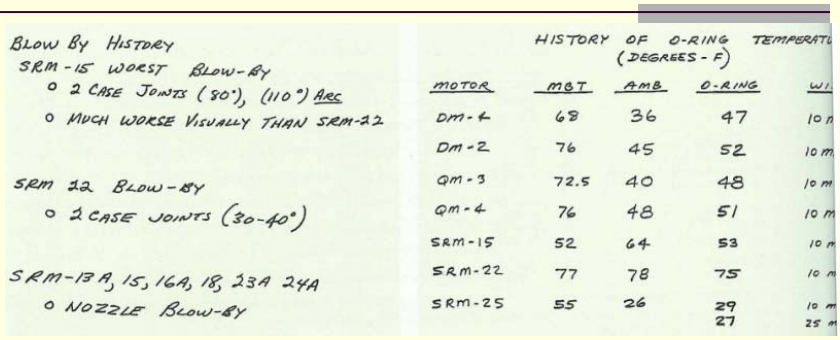
OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

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.

- The chart showed the effect, but does not shown the possible cause, temperature
- Six types of descriptions break the evidence up into stupefying fragments

34

Story 2: Challenger



Blow By History
SRM-15 WORST Blow-By
- 2 Case Joints (30°), (110°) Arc
- Much worse Visually than SRM-22

SRM 22 Blow-By
- 2 case joints (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A
- Nozzle Blow-By

	HISTORY OF O-RING (DEGREES-F)			TEMPERATU
MOTOR	MBT	AMB.	O-RING	WI
DM-1	68	36	47	10 n
DM-2	76	45	52	10 m
QM-3	72.5	40	48	10 m
QM-4	76	48	51	10 m
SRM-15	52	64	53	10 m
SRM-22	77	78	75	10 m
SRM-25	55	26	29 27	10 m 25 m

- Blow-by and temperature for two launches (there are other 22 cases left)

35

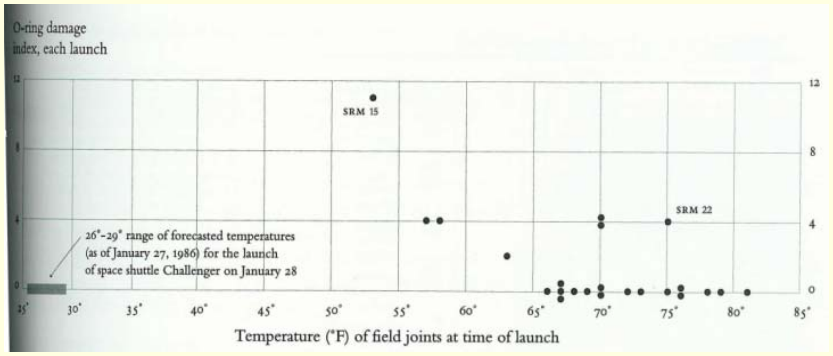
Story 2: Challenger

- Engineers' conclusion
  - O-rings would fail due to cold weather the next day.
- NASA officers' conclusion
  - They wouldn't fail.
- Consequence: Challenger was launched the next day. O-rings failed. ...

36

Story 2: Challenger

- What if this chart was used:



E. Tufte

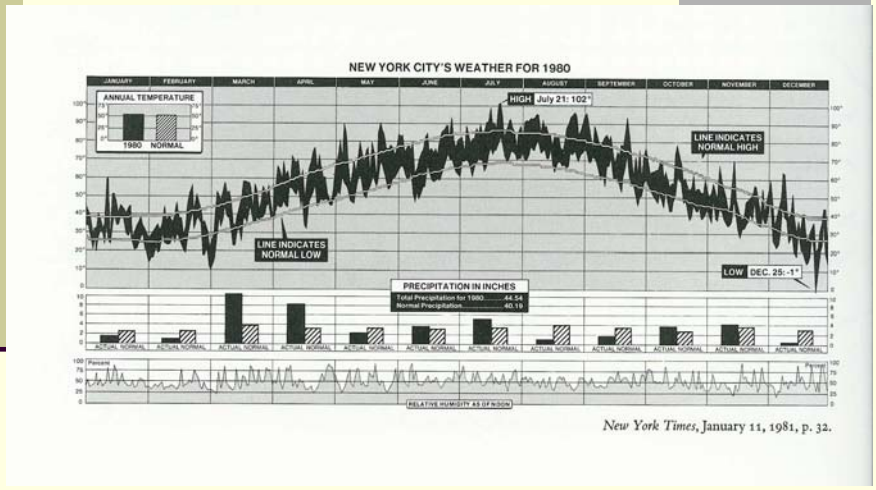
37

What is Information Visualization

- It is about scale and dimensionality
- Scale
  - Essential problem in reasoning is comparison
  - Comparisons must be enforced within scope of eye span
- Dimensionality
  - The world is multi-dimensional
  - The paper and computer is 2 dimensional
  - Escape from the flatland

38

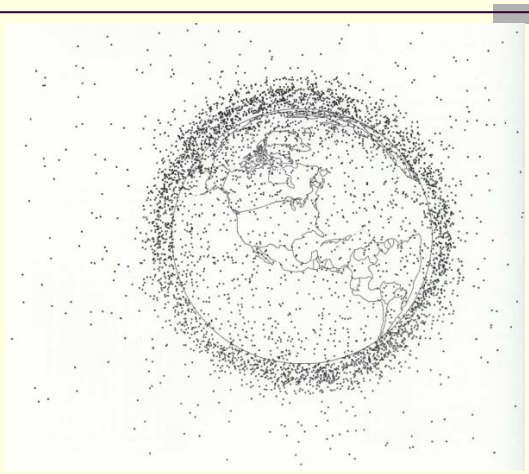
Weather of New York



The figure depicts 1,888 numbers

39

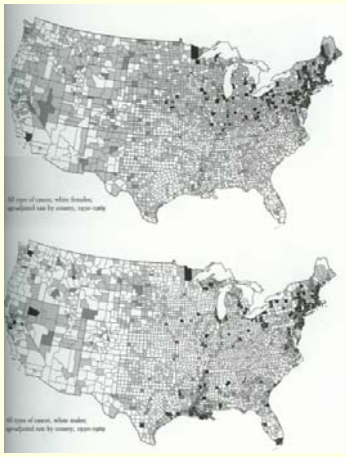
The Space Debris



7,000 pieces of space debris orbiting our world

40

The Cancer Maps

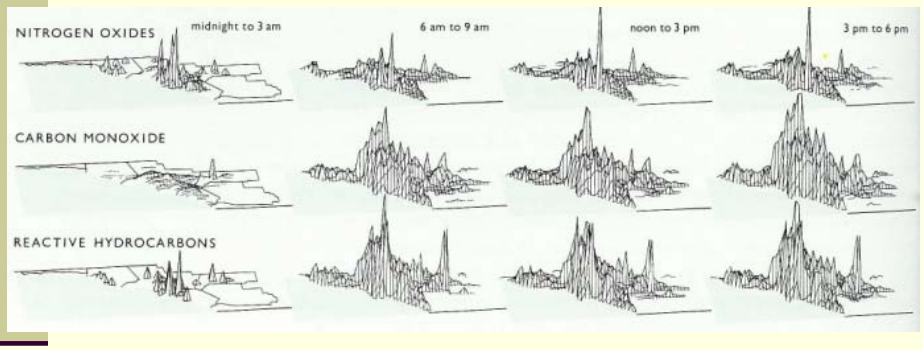


Each map portrays 21,000 numbers

41

R. Hoover et al. 1975

Levels of Air Pollutants



Six variables

42

What is Information Visualization

- It is about interactive exploration
  - Want to show multiple different perspectives on the data
  - A way to increase scalability and dimensionality

43

Interactive Exploration of Images

Video

44

What is Information Visualization

- It is about applications
  - Document, images, videos, multimedia
  - Financial/business data
  - Internet information
  - Software

45

Example 1 - Infocanvas

The Infocanvas project

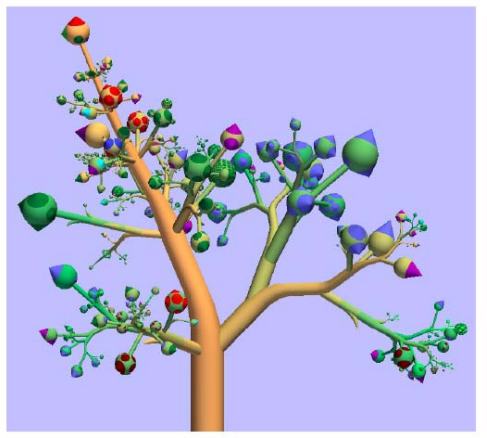
Team Members: John Stasko, Dave McColgin, Todd Miller, Chris Plaue, Zach Pousman

http://www.cc.gatech.edu/gvu/ii/infoart/

46

Example 2 – Botanical Tree

- The Unix home-directory of Dr. Kleiberg?



E. Kleiberg at.el. Infovis 2001

47

Example 3 – Stock Visualization

- Visualization 1:
  - Yahoo stock quotes for one stock
  - http://finance.yahoo.com/
- Visualization 2:
  - Smartmoney Map of Market
  - http://www.smartmoney.com/marketmap/

48

Example 4 – Home Finder

- Dynamic home finder
  - Human-Computer Interaction Lab / Univ. of Maryland
  - http://www.cs.umd.edu/hcil/pubs/products.shtml

49

Example 5 – Dynamic History

- Online American history textbook
  - http://www.digitalhistory.uh.edu/timeline/timelineO.cfm

50

The Need is There

In five years, 100 million people will be using an information-visualization tool on a near-daily basis. And products that have visualization as one of their top three features will earn $1 billion per year.

Ramana Rao, founder and chief technology officer, Inxight Software Inc., Sunnyvale, Calif.

http://www.computerworld.com/databasetopics/data/story/0,10801,80243,00.html

Slide courtesy of John Stasko

51

References:

- Edward Tufte:
  - The Visual Display of Quantitative Information
  - Envisioning Information
  - Visual Explanation
- Dr. John Stasko's Information Visualization class slides

52