The Value of Visualization for Understanding Data and Making Decisions

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

JISIC 2014
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
Data Visualization

Making pretty pictures?
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process
Gain an understanding
Visualization

“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visually help us think
Provide a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Applications of Visualization

Presentation

Analysis
1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Infographics
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE

- First Atomic Bomb on Hiroshima: 1945
- First American H Bomb: 1949
- First Soviet A Bomb: 1952
- First Soviet H Bomb: 1953
- USMC submarine launched: 1957
- SALT I talks on reducing nuclear weapons: 1969
- SALT II talks on reducing nuclear weapons: 1979
- Reykjavik summit on reducing nuclear weapons: 1986
- USA and USSR Intercontinental Ballistic Missiles program: 1989
- USA withdraw from SALT II: 1991
- USSR collapses: 1991

NUCLEAR STOCKPILE

- Year 45: 10,000
- Year 50: 20,000
- Year 60: 30,000
- Year 70: 40,000
- Year 80: 50,000
- Year 91: 60,000

ROCKET MODELS

- SS-9 (USSR): 1966
- Titan II (USA): 1962
- Warhead: 10 Mt for USSR, 9 Mt for USA
- Operational range: 16,000 km for USSR, 15,000 km for USA

AIR CONTROL

- Intercontinental Ballistic Missiles: USA 8,000, USSR 7,000
- Planes: USA 4,000, USSR 5,000

A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states. The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.

- Troops: NATO 2.5 million, Warsaw Pact 4 million
- Tanks: NATO 13,000, Warsaw Pact 42,500
- Artillery: NATO 10,750, Warsaw Pact 31,500

Produced by WatchTheAmericans.com a Fan Website

http://visual.ly/nuclear-arms-race
Gay rights in the US, state by state
Gay rights laws in America have evolved to allow — but in some cases ban — rights for gay, lesbian and transgender people on a range of issues, including marriage, hospital visitation, adoption, housing, employment and school bullying. The handling of gay rights issues vary by state and follow trends by region.

- Obama supports same-sex marriage: share with us what it means to you
- Gay rights: five activists reflect on the history of the movement in the US
- President Obama endorses gay marriage
Simply presenting data \textit{visually} can have a profound impact
Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri
Gun ownership in New York counties

The gun owner next door: What you don't know about the weapons in your neighborhood

http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1
http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood?-gcheck=1&nclick_check=1
Frequent presentation goals

Clarify
Focus
Highlight
Simplify

May just show a few variables and/or a subset of the data cases
2. Analysis
Explore the data
Assess a situation
Determine how to proceed
Decide what to do
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland

*The Elements of Graphing Data*
Frequent analysis goals

Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility
Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
So what is visualization most useful for?
1. **Visualization** is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context
Investigative analysis

Context

Trust

Learning

Awareness

Visual Analytics Support for Intelligence Analysis

Görg, Kang, Kiu, Stasko
IEEE Computer ‘13
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

**Q & A dialog through interaction**
Engage in a dialog with your data

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect
3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly
Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Visualization of different data types
Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

32
Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?
Example Tasks and Goals

• Which documents contain text on topic XYZ?
• Which documents are of interest to me?
• Are there other documents that are similar to this one (so they are worthwhile)?
• How are different words used in a document or a document collection?
• What are the main themes and ideas in a document or a collection?
• Which documents have an angry tone?
• How are certain words or themes distributed through a document?
• Identify “hidden” messages or stories in this document collection.
• Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
• Find connections between documents.
The Challenge

Text is nominal data
Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data
- Bar charts, line charts, scatterplots, etc.
Jigsaw
Computational analysis of documents’ text
  Entity identification, document similarity, clustering, summarization, sentiment
Multiple visualizations of documents, analysis results, entities, and their connections
  Views are highly coordinated

“Putting the pieces together”
Help “investigators” explore, analyze and understand large document collections

Academic Papers

Product Reviews

Health Forums

Police Reports

2010 Hyundai Genesis Sedan - Consumer Reviews

Vehicle
2010 Hyundai Genesis 3.8-litre Sedan (16/23/24) 4.6 (4.6L 6cyl 6A)

Review
After driving 4 US series Leous's decided on a change due to price. Toyota/Lexus were having all the recall problems at the automotive reviews, I took a Genesis for a test drive. What I say that the car handles beautifully in all driving conditions, I found proof, I would highly recommend this car to anyone.
Visual analytics of large text document collections

Görg, Liu, Kihm, Choo, Park, & Stasko

TVCG ‘13
Example

Made each page into a separate “document”
585 in total
Entities: Person, Location, Organization, Date, Money

Demo
Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw
Applications

• Intelligence & law enforcement
  – Police cases
  – Won 2007 VAST Contest
  – Stasko et al, *Information Visualization* ’08
• Academic papers, PubMed
  – All InfoVis & VAST papers
  – CHI papers
  – Görg et al, KES ‘10
• Investigative reporting
• Fraud
  – Finance, accounting, banking
• Grants
  – NSF CISE awards from 2000
• Topics on the web (medical condition)
  – Autism
• Consumer reviews
  – Amazon product reviews, edmunds.com, wine reviews
  – Görg et al, HCIR ’10
• Business Intelligence
  – Patents, press releases, corporate agreements, ...
• Emails
  – White House logs
• Software
  – Source code repositories
  – Ruan et al, SoftVis ‘10
Further benefits of visualization

Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
EuroVis ‘14 Capstone Talk

The Value of Visualization... and Why Interaction Matters

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

Abstract: Visualization researchers need to do a better job communicating the value of our field externally. Visualization, by its very nature, provides inherent challenges to doing this. In this talk I will explain these challenges and articulate my views on the value of visualization, including its unique capabilities for data presentation and analysis. I will describe the advantages of interaction, and discuss in depth why interaction is so important to our field and how it has been under-utilized to date. Finally, I will present a number of new interaction ideas and techniques that can be integrated into our future systems.

http://vimeo.com/98986594
Take Aways
Presentation & analysis

Interaction provides the power

Exploring & developing questions
Acknowledgments

• Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

• Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

• Supported by DARPA’s XDATA Program
My Research

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