The Value of Visualization... and Why Interaction Matters

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
Data Visualization
Making pretty pictures?
Data Visualization

Making pretty pictures
Data Visualization

A cognitive process
  Gain an understanding
Data Visualization

A cognitive process
Gain an understanding
Visuals help us think
Provide a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Visualization’s Value

Need to make the case better externally
The Value of Information Visualization

Jean-Daniel Fekete¹, Jarke J. van Wijk², John T. Stasko³, and Chris North⁴

¹ Université Paris-Sud, INRIA, Bât 490, F-91405 Orsay Cedex, France,
Jean-Daniel.Fekete@inria.fr,
http://www.aviz.fr/~fekete/
² Department of Mathematics and Computing Science,
Eindhoven University of Technology, P.O. Box 513,
5600 MB EINDHOVEN, The Netherlands,
vanwijk@win.tue.nl,
http://www.win.tue.nl/~vanwijk/
³ School of Interactive Computing, College of Computing & GVU Center,
Georgia Institute of Technology, 85 5th St., NW,
Atlanta, GA 30332-0760, USA,
stasko@cc.gatech.edu,
http://www.cc.gatech.edu/~john.stasko
⁴ Dept of Computer Science, 2202 Kraft Drive,
Virginia Tech, Blacksburg, VA 24061-0106, USA,
north@vt.edu,
http://people.cs.vt.edu/~north/

Abstract. Researchers and users of Information Visualization are convinced that it has value. This value can easily be communicated to others in a face-to-face setting, such that this value is experienced in practice. To convince broader audiences, and also, to understand the intrinsic qualities of visualization in our difficult domain. In this paper we use

Information Visualization: Human-Centered Issues and Perspectives
Springer, 2008
Need for Visualization

Articulating value requires identifying purpose

Three stories...
People - Faculty

Home  People  Faculty

People - Faculty

Faculty  |  Staff

Gregory Abowd
Distinguished Professor; Regents Professor

Hua Ai
Research Scientist II

Annie Antón
Professor and Chair, School of Interactive Computing

Alberto Apostolico
Professor

Ronald Arkin
Associate Dean for Research and Space Planning; Regents' Professor

Rosa Arriaga
Senior Research Scientist
Instagram Visualization

Number of photos per day (photos get from API, not the actual photo counts on Instagram)

Average likes per photos per filter

Average numbers of comments per photos per filter

Popular photos

Chloe Xie
“A man’s got to know his limitations.”

*Magnum Force*, 1973
Thought
If you can articulate very precisely what you’re seeking, visualization likely isn’t your best approach

OK, so what is visualization good for?
Applications of Visualization

Presentation  Explanatory
Analysis      Exploratory
1. Presentation
Communicate data and ideas
Explain and inform
Provide evidence and support
Influence and persuade
Infographics

All the rage...
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 Bombs on Hiroshima: 1945
- First Soviet A Bomb: 1949
- First US H Bomb: 1952
- First US missile submarine launched: 1953
- First US satellite 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: 91
- USA withdraws from SALT II: 86
- USSR collapses: 91

NUCLEAR STOCKPILE
- USA
- USSR

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

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.

Produced by WatchTheAmericans.com a Fan Website
<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>ABV</th>
<th>IBU</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Berliner Weisse</td>
<td>5.0</td>
<td>15</td>
</tr>
<tr>
<td>02</td>
<td>Belgian Wit</td>
<td>4.8</td>
<td>10</td>
</tr>
<tr>
<td>03</td>
<td>American Wheat</td>
<td>4.8</td>
<td>10</td>
</tr>
<tr>
<td>04</td>
<td>Weizen</td>
<td>5.0</td>
<td>15</td>
</tr>
<tr>
<td>05</td>
<td>Dunkelbier</td>
<td>4.5</td>
<td>9</td>
</tr>
<tr>
<td>06</td>
<td>Flanders Dubbel</td>
<td>5.6</td>
<td>40</td>
</tr>
<tr>
<td>07</td>
<td>Lambic / Guuze</td>
<td>5.6</td>
<td>40</td>
</tr>
<tr>
<td>08</td>
<td>Belgian Pale</td>
<td>5.0</td>
<td>15</td>
</tr>
<tr>
<td>09</td>
<td>American Pale Ale</td>
<td>5.0</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>English Pale Ale</td>
<td>4.5</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>Belgian Strong Pale</td>
<td>8.0</td>
<td>60</td>
</tr>
<tr>
<td>12</td>
<td>Belgian Strong Ale</td>
<td>8.0</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>Att. Red &amp; Amber Ale</td>
<td>7.0</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>Barleywine</td>
<td>6.5</td>
<td>40</td>
</tr>
<tr>
<td>15</td>
<td>Scotch Ale</td>
<td>4.0</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>Out Brun</td>
<td>8.0</td>
<td>60</td>
</tr>
<tr>
<td>17</td>
<td>Dubbel</td>
<td>5.5</td>
<td>40</td>
</tr>
<tr>
<td>18</td>
<td>Strong Ale</td>
<td>8.0</td>
<td>60</td>
</tr>
<tr>
<td>19</td>
<td>Brown Ale</td>
<td>6.0</td>
<td>40</td>
</tr>
<tr>
<td>20</td>
<td>Imperial Stout</td>
<td>5.0</td>
<td>9</td>
</tr>
<tr>
<td>21</td>
<td>Pale Ale</td>
<td>4.5</td>
<td>9</td>
</tr>
</tbody>
</table>

**Ale**

http://thebeermongers.com/beers/
Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

According to allegations in a book about Bonds, he began taking steroids before the 1990 season, hit 44 HR in the regular season, a record. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run pace for each player after age 34.

- Hank Aaron: Actual home runs slightly outpace projected home runs for five seasons.
- Babe Ruth: Averaged 42.4 home runs in a season from age 30 to 34. Averaged 42.5 for 1941 to 1945 seasons.
- Barry Bonds: From age 35 to 39, he averaged 43 more home runs, a season than projected.

Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (97th) and Pujols (267th):

- Hank Aaron (755): Hit 250 or more (M.L. most) in first five seasons.
- Babe Ruth (714): Averaged 62 from 1920 to 1924.
- Barry Bonds (708): No one hit more from 1950-59.
- Willie Mays (659): Three 60 homer seasons is a record.
- Sammy Sosa (589): Triple Crown in '98 (41, 122, 316).
- Frank Robinson (584): First to hit 70 in a season.
- Mark McGwire (583): Only McGwire had more in the 90's.
- Alex Rodriguez (480): Second most home run leader.
Where We Live...

Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie.
All of the goals which took the teams to Euro 2012

Euro 2012 kicks off tomorrow night and promises to keep soccer fans glued to their television sets for the rest of the month. The Post takes a look at how the 16 finalists put the ball in the back of the net in qualifying, giving an idea of who is most likely to be a scoring threat in the finals, and from where.

Group A
- Poland
- Greece
- Russia
- Czech Republic

Goals scored:
- Poland: 14
- Greece: 2
- Russia: 6
- Czech Republic: 4

Group B
- Netherlands
- Denmark
- Germany
- Portugal

Goals scored:
- Netherlands: 9
- Denmark: 3
- Germany: 7
- Portugal: 2

TICKET READER PREDICTIONS

More than 3,000 Times readers have gone online to give their predictions about how the World Cup finals will develop. Starting with the groups, readers have worked their way through the competition and their thoughts are collated below. Participants were asked which teams would qualify from the group stages and who would win the tournament overall. There was little disagreement, with only group E throwing up diverse viewpoints.

THE GROUP STAGES

Everyone thought Brazil would qualify from group A but were split over who join the hosts in the round of 16 by finishing second. Australia were the least fancied team to advance, a mere 1 per cent of people think they would get out of group B.

GROUP C

GROUP D

GROUP E

GROUP F

GROUP G

GROUP H

WHO WILL WIN THE WORLD CUP?

Brazil took most votes to win the World Cup on home ground. An optimistic 5 per cent plumped for England.
GROUP D

71% Italy

1% Costa Rica

53% Uruguay

76% England
GROUP G

90% Portugal

5% Ghana

6% USA

99% Germany
Frequent presentation goals

Clarify
Focus
Highlight
Simplify
Persuade

May just show a few variables and/or a subset of the data cases
Simply presenting data *visually* can have a profound impact
**My Class**

http://www.cc.gatech.edu/~stasko/7450

---

**CS 7450 - Information Visualization**

Instructor: John Stasko  
Fall 2013  
Mon, Wed 3:00 - 4:30 pm  
Whitaker Bldg. room 1103

Information visualization is a research area that focuses on the use of visualization techniques to help people understand and analyze data. While fields such as scientific visualization involve the presentation of data that has some physical or geometric correspondence, information visualization focuses on abstract data without such correspondences such as symbolic, tabular, networked, hierarchical, or textual information sources.

The objectives of the course are:

- Learn the principles involved in information visualization
- Learn about the variety of existing techniques and systems in information visualization
- Develop skills in critiquing different visualization techniques as applied to particular tasks
- Learn how to evaluate visualization systems
- Gain a background that will aid the design of new, innovative visualizations

The course will follow a lecture/seminar style with much discussion of assigned readings, as well as viewing of videos and hands-on experience with research and commercial information visualization tools.

We will be reading recent research papers about the different course topics. In addition, we will be using one book for the course: *Now You See It* by Stephen Few, Analytics Press 2009. Also highly recommended is *Envisioning Information* by Edward Tufte, Graphics Press 1990.
Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri
Gun ownership in New York counties

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
Hans Rosling
Gapminder

An Inconvenient Truth

Gore made extensive use of data graphics
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
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
At Its Heart: Uncertainty

Nothing is perfect
Humans trade off choices, compromise

Cost

Bedroom

Journey time

Attribute Explorer

Spence & Tweedie

Interacting with Computers ‘98
Thought

Even analysis is about effective communication
Visualization’s Value?
Visualization’s Value?

\[ V_{\text{value}} = T + I + E + C \]
Visualization’s Value?

\[ V \text{alue} = T + I + E + C \]

Ability to minimize the total \textbf{time} needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)
What kinds of questions?

“Low-level” tasks

Retrieve value
Filter
Compute derived value
Find extremum
Sort
Determine range
Characterize distribution
Find anomalies
Cluster
Correlate
Visualization’s Value?

\[ V_{\text{value}} = T + I + E + C \]

Ability to spur and discover \textbf{insights} or insightful questions about the data

(Would be very difficult with only the data)
What is Insight?

An individual observation about the data by the participant, a unit of discovery

Complex
Deep
Qualitative
Relevant
Unexpected

Saraiya, North, & Duca
TVCG ‘05

North
IEEE CG&A ‘06
What is Insight?

Sudden grasp of new relationships that are necessary to solve a problem and that were not learned in the past

Bernstein, Penner, Clarke-Stewart & Roy
Psychology, 6th edition
What is Insight?

Is not spontaneous “aha!” moments (eg, in cognitive science)
Is knowledge-building and model-confirmation
   Like a substance that people acquire with the aid of systems

Chang, Ziemkiewicz, Green, & Ribarsky
IEEE CG&A ‘09
Visualization’s Value?

\[ V_{\text{value}} = T + I + E + C \]

Ability to convey an overall essence or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)
### Overview and detail

#### Focus + context

<table>
<thead>
<tr>
<th>Name</th>
<th>Avg</th>
<th>Career Avg</th>
<th>Team</th>
<th>Salary 97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larry Herndon</td>
<td>0.24734903</td>
<td>0.27262676</td>
<td>Det.</td>
<td>225</td>
</tr>
<tr>
<td>Jesse Barfield</td>
<td>0.2986243</td>
<td>0.27263819</td>
<td>Tor.</td>
<td>1337.5</td>
</tr>
<tr>
<td>Jeffrey Leonar</td>
<td>0.2705238</td>
<td>0.27260458</td>
<td>S.F.</td>
<td>300</td>
</tr>
<tr>
<td>Donnie Hill</td>
<td>0.28318584</td>
<td>0.2725564</td>
<td>OAK.</td>
<td>275</td>
</tr>
<tr>
<td>Billy Sample</td>
<td>0.285</td>
<td>0.2718501</td>
<td>Atl.</td>
<td>NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Avg</th>
<th>Career Avg</th>
<th>Team</th>
<th>Salary 97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Howard Johnson</td>
<td>0.24545455</td>
<td>0.25232068</td>
<td>N.Y.</td>
<td>297.5</td>
</tr>
<tr>
<td>Andrea Thomas</td>
<td>0.250774</td>
<td>0.2561994</td>
<td>Atl.</td>
<td>75</td>
</tr>
<tr>
<td>Billy Hatcher</td>
<td>0.2575556</td>
<td>0.2551507</td>
<td>Hou.</td>
<td>110</td>
</tr>
<tr>
<td>Omar Moreno</td>
<td>0.2339833</td>
<td>0.2518029</td>
<td>Atl.</td>
<td>NA</td>
</tr>
<tr>
<td>Darnell Coles</td>
<td>0.2725523</td>
<td>0.25153375</td>
<td>Det.</td>
<td>105</td>
</tr>
</tbody>
</table>

Rao & Card

CHI ‘94
Visualization’s Value?

\[ V_{\text{alue}} = T + I + E + C \]

Ability to generate **confidence** and trust about the data, its domain and context

(Beneficial data analysis process side effects)
Some examples
E. Tufte
_The Visual Display of Quantitative Information_
1983

## Data Values

<table>
<thead>
<tr>
<th>Count</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>365</td>
<td>High temp for each day</td>
</tr>
<tr>
<td>365</td>
<td>Low temp for each day</td>
</tr>
<tr>
<td>365</td>
<td>Avg high temp for each day</td>
</tr>
<tr>
<td>365</td>
<td>Avg low temp for each day</td>
</tr>
<tr>
<td>365</td>
<td>Precipitation for each day</td>
</tr>
<tr>
<td>365</td>
<td>Humidity for each day</td>
</tr>
<tr>
<td>12</td>
<td>Precipitation for each month</td>
</tr>
<tr>
<td>12</td>
<td>Avg precipitation for each month</td>
</tr>
<tr>
<td>1</td>
<td>Precipitation for the year</td>
</tr>
<tr>
<td>1</td>
<td>Avg precipitation per year</td>
</tr>
<tr>
<td>1</td>
<td>Highest temp (&amp; day) for the year</td>
</tr>
<tr>
<td>1</td>
<td>Lowest temp (&amp;day) for the year</td>
</tr>
<tr>
<td>1</td>
<td>Avg daily temp for the year</td>
</tr>
<tr>
<td>1</td>
<td>Avg daily temp per year</td>
</tr>
</tbody>
</table>
Visualization’s Value

\[ \text{Value} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{alue}} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{alue}} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{value}} = T + I + E + C \]
A month-by-month look at Atlanta’s weather in 2013

Atlanta’s 2013 weather can be summed up in one word—soggy. The city ended the year 16.31 inches above normal in rainfall, with much of that rain falling during the waterlogged summer months, which had about twice as many rainy days as dry ones. Rain fell into the city’s official gauge at Hartsfield-Jackson International Airport on 17 days in June, 19 days in July and 24 days in August. The rainfall total for those three months was 23.29 inches, making the period the city’s fourth wettest summer on record. Daily rainfall records were set on April 28 (1.73 inches), May 4 (3.40 inches), June 5 (4.14 inches), Aug. 8 (2.60 inches) and Oct. 6 (2.12 inches). Summer 2013 will also be remembered for a lack of extreme heat, particularly compared to the sweltering summer of 2012. Temperatures climbed into the 90s only 20 times, topping out at 92 on June 28 and Aug. 29. That’s 32 less days of 90+ heat than in 2012, when the all-time record high of 106 was reached.

Daily temperatures

Atlanta saw record highs twice in 2013—on Jan. 12 and Dec. 22, when temperatures reached 76 and 71, respectively.

Monthly rainfall

Sources: National Weather Service, National Drought Mitigation Center

Atlanta Journal Constitution

Jan. 3, 2014
Map of the Market

http://www.marketwatch.com/tools/stockresearch/marketmap
Map of the Market (old school)
Visualization’s Value

\[ \text{Value} = \text{T} + \text{I} + \text{E} + \text{C} \]
Visualization’s Value

\[ V_{\text{value}} = T + I + E + C \]
Visualization’s Value

\[ V_{value} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{value}} = T + I + E + C \]
CiteVis

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

Demo

Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster ‘13
Visualization’s Value

\[ \text{Value} = \text{T} + \text{I} + \text{E} + \text{C} \]
Visualization’s Value

\[ V_{\text{value}} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{alue}} = T + I + E + C \]
Visualization’s Value

\[ V_{\text{alue}} = T + I + E + C \]
Value
Not evaluation in traditional sense

HCI: Benchmark tasks
Visualization: Vagueness or absence of specific tasks
More about exploration and understanding

Pffft on evaluation.
Just build cool stuff.
STAR Report: Set Visualization

Alsallakh, Micallef, Aigner, Hauser, Miksch & Rodgers
EuroVis ‘14
Data Visualization 101

Problem:
You have a lot of data (& attributes) to understand

Do you?

Pack all the data into one complex representation

Spread the data into multiple coordinated views

Use interaction to reveal different subsets of the data
Constituents

Two key aspects of data visualization
  Representation
  Interaction

“The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means.”

S. Few, *Now you see it*
Interaction is Vital

Engage in a dialog with your data

Fundamental nature: Equal sibling with representation or subordinate facilitator?
Interaction

Why interact?

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

Yi, Kang, Stasko & Jacko
TVCG (InfoVis) ‘07
Interaction
How manifested today?

- Tooltips & selection
- Get details
- Navigation
- Brushing & linking
Interaction
Can we do more?

Employ interaction in a more fundamental manner to strengthen the power of visualization
CiteVis

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

Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster ‘13
Dust and Magnet

Yi, Melton, Stasko & Jacko
Information Visualization ‘05
OnSet

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

Demo

Sadana, Major, Dove, & Stasko
Interaction

What are the tools of interaction?

Traditional – Desktop: keyboard, mouse

New – Tablet: fingers and multitouch
TouchWave

Baur, Lee & Carpendale
ITS ‘12

Selection Gestures

Willett, Lan & Isenberg
EuroVis ‘14
Moving to Tablets

Scatterplot

Video

Sadana & Stasko
AVI ‘14
Key Open Problems in Visualization
Key Open Problems in Visualization

1. Assess and communicate \textit{value}

\[ V_{\text{value}} = T + I + E + C \]
Key Open Problems in Visualization

1. Assess and communicate value

2. Make the construction of visualizations easier

SketchStory

Lee, Kazi & Smith
TVCG (InfoVis) ‘13

Lyra

Satyanarayan & Heer
EuroVis ‘14
Key Open Problems in Visualization

1. Assess and communicate value
2. Make the construction of visualizations easier
3. Address real world “big” problems
Take Aways
Visualization is not for a few precise, concrete tasks
Exploratory data analysis

Presentation & analysis, related but different

Value = T + I + E + C

Interaction provides power, use it
And finally...
Thank you
Acknowledgments

• Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

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

• Supported by DARPA’s XDATA program