INFORMATION VISUALIZATION

Alvitta Ottley

Washington University in St. Louis
Announcements
Office Hour
Canceled Today
Assignment 1: Design Critique
Due: 01-24-2017, 11:59pm (midnight)

In this assignment, you will be looking for visualizations “in the wild” (in books, newspapers, magazines, on the internet, etc.). Specifically, you will be looking for two visualizations – one that you like, and one that you dislike. For the visualization that you like, you need to express why you like the visualization (what is it that makes the visualization good). For the visualization that you dislike, you need to provide a critique, as well as design a better alternate visualization.

Basic Requirements for this Assignment:

1. Find two visualizations in the wild
   a. One that you like
   b. One that you dislike

2. For the visualization that you like, provide a description of what makes the visualization good.

3. For the visualization that you dislike, explain why you dislike it.

4. In addition, for the visualization that you dislike, design an alternate visualization that is better than the original.

5. Explain your design and what problem(s) your design addresses.
Recap...
Why we need Visualization

- Cognition is limited
- Memory is limited
How does Visualization work?

- Uses perception to point out interesting things.
Reasons for creating visualizations

• answer questions
• generate hypotheses
• make decisions
• see data in context
• expand memory
• support computational analysis
• find patterns
• tell a story
• inspire
Today...
Today...

- Tufte’s Principles of Graphical Design
  - Graphical Integrity
  - Graphical Excellence

- Research that contradicts Tufte.
EDWARD TUFTE

- Evangelist for good visual design
- Most designs are static, but many principles apply to interactive (computer-based) visualization designs
- Take these design guidelines with a grain of salt
EDWARD TUFTE
TUFTE’S LESSONS

• Graphical Integrity
• Graphical Excellence
Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity.
MISSING SCALES

Tufte 2001
MISSING SCALES

What is the baseline?
MISSING SCALES

What is the baseline?

- $4,200,000

Tufte 2001
Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity.

“Above all else show the data”
THE LIE FACTOR

• Tufte coined the term “the lie factor”, which is defined as:

\[
\text{Lie\_factor} = \frac{\text{size of graphic}}{\text{size of data}}
\]

• “High” lie factor (LF) leads to:
  • Exaggeration of differences or similarities
  • Deception
  • Misinterpretation
THE LIE FACTOR

- The Lie Factor (LF) can be: \( \frac{\text{size of graphic}}{\text{size of data}} \)
  - LF > 1
  - LF < 1
- If LF is > 1, then size of graphic is greater than the size of data
  - This leads to exaggeration of the data (overstating the data)
- If LF < 1, then the size of the data is greater than the graphic
  - This leads to hiding the of data (understating the data)
The US Department of Transportation had set a series of fuel economy standards to be met by automobile manufacturers, beginning with 18 miles per gallon in 1978 and moving in steps up to 27.5 by 1985.
WHAT IS WRONG WITH THIS?

The line representing 18 miles per gallon in 1978, is 0.6 inches long.

The line representing 27.5 miles per gallon in 1985, is 5.3 inches long.
WHAT IS WRONG WITH THIS?

• The increase in real data between 1978 to 1985 (from 18 MPG to 27.5 MPG) is:

\[
\frac{27.5 - 18.0}{18.0} \times 100 = 53\%
\]

• The difference in length between 1978 to 1985 (from 0.6 inches to 5.3 inches) is:

\[
\frac{5.3 - 0.6}{0.6} \times 100 = 783\%
\]

• Lie Factor is:

\[
\frac{783}{53} = 14.8
\]
LIE FACTOR EXAMPLE

This design contains a lie factor of 9.4
This design contains a lie factor of **9.5**
OTHER WAYS TO LIE: ENCODING

Comparative Annual Cost per Capita for care of Insane in Pittsburgh City Homes and Pennsylvania State Hospitals.

- South Mountain: $147
- Pittsburgh: $172
- Harrisburg: $198
- Norristown: $213
- Warren: $214
OTHER WAYS TO LIE: DESIGN VARIATION

**OPEC Oil Prices: After 18 Months of Stability, Prices Are Due to Rise Again**

Dollars per barrel

- Jan. 1, 5% increase
- Apr. 1, 3.89% increase
- July 1, 2.29% increase
- Oct. 1, 2.69% increase

1979

Quarterly

Yearly

$13.34

$13.84

$14.16

$14.54

OTHER WAYS TO LIE: DESIGN VARIATION

Beware of the “3D” effect. It distorts the telling of the data.

• There are five vertical scales here:
  • 1073-1978: 1 inch = $8.00
  • Jan-Mar: 1 inch = $4.73
  • Apr – Jun: 1 inch = $4.37
  • Jul – Sep: 1 inch = $4.16
  • Oct – Dec: 1 inch = $3.92

• And two horizontal scales:
  • 1973-1978: 1 inch = 3.8 years
  • 1979: 1 inch = 0.57 years
OTHER WAYS TO LIE: THE 3D EFFECT
OTHER WAYS TO LIE: DOUBLE ENCODING
OTHER WAYS TO LIE: DOUBLE ENCODING

- Here, both width and height encode the same information. The effect is multiplicative.

\[ 0.44 \text{ (width)} \times 0.44 \text{ (height)} = 0.19 \]
OTHER WAYS TO LIE: UNINTENDED ENCODING
OTHER WAYS TO LIE: UNINTENDED ENCODING
OTHER WAYS TO LIE: ALIGNMENT

THE SHRINKING FAMILY DOCTOR
in California.

Percentage of Doctors Devoted Solely to Family Practice

<table>
<thead>
<tr>
<th>Year</th>
<th>1964</th>
<th>1975</th>
<th>1990</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27%</td>
<td>16.0%</td>
<td>12.0%</td>
</tr>
</tbody>
</table>

In 1964: 2,247
In 1975: 8,694
In 1990: 4,232

- 3,162
- 6,212
OTHER WAYS TO LIE: LIMITING CONTEXT

Connecticut Traffic Deaths, Before (1955) and After (1956)
Stricter Enforcement by the Police Against Cars Exceeding Speed limit
OTHER WAYS TO LIE: LIMITING CONTEXT

Connecticut Traffic Deaths, 1951–1959

Connecticut Traffic Deaths, Before (1955) and After (1956) Stricter Enforcement by the Police Against Cars Exceeding Speed Limit
OTHER WAYS TO LIE: LIMITING CONTEXT

Context: “Compared to what?”

100 combat deaths per month

Iraq

M. Ericson, New York Times
via Hampeter Pfister, Harvard
OTHER WAYS TO LIE: LIMITING CONTEXT
OTHER WAYS TO LIE: LIMITING CONTEXT

- World War II
- Vietnam
- Iraq

M. Ericson, New York Times via Hanipeter Pfister, Harvard
HOW TO NOT LIE

“Maximize the Data-Ink Ratio”
DATA-INK RATIO

\[
\text{ink used for data} \over \text{total ink in the graphic}
\]
DATA-INK RATIO

• The goal is to aim for high data-ink ratio

• Ink used for the data should be relatively large compared to the ink in the entire graphic
HIGH DATA-INK RATIO EXAMPLE

The length of an organism at the time of reproduction in relation to the generation time, plotted on a logarithmic scale.
LOW DATA-INK RATIO EXAMPLE
PREVIOUS EXAMPLE IMPROVED
ERASING NON-DATA INK

How many times is height encoded?
ERASING NON-DATA INK

Multiple encodings:

1. Height of the left line
2. Height of the right line
3. Height of shading
4. Position of top horizontal line
5. Position (placement) of the number
6. Value of the number
Results of a study indicating that one type of element always has a higher value under different experimental conditions.
ERASING NON-DATA INK EXAMPLE

After removing all non-data ink
ERASING NON-DATA INK EXAMPLE

The ink that has been removed
THOUGHTS ABOUT THIS?
THOUGHTS ABOUT THIS?
SUMMARY OF DESIGN PRINCIPLES

1. Above all else show the data
2. Maximize the data-ink ratio
3. Erase non-data-ink
4. Erase redundant data-ink
5. Revise and edit
GRAPHICAL EXCELLENCE

1. Graphical excellence is the well-designed presentation of interesting data – a matter of substance, of statistics, and of design.
2. Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.
3. Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink the smallest place.
4. Graphical excellence is nearly always multivariate.
5. And graphical excellence requires telling the truth about the data.
QUESTIONS?
EVIDENCE AGAINST TUFTE
A User Study of Visualization Effectiveness Using EEG and Cognitive Load

E. W. Anderson¹, K. C. Potter¹, L. E. Matzen², J. F. Shepherd², G. A. Preston³, and C. T. Silva¹

¹SCI Institute, University of Utah, USA
²Sandia National Laboratories, USA
³Utah State Hospital, USA

Abstract
Effectively evaluating visualization techniques is a difficult task often assessed through feedback from user studies and expert evaluations. This work presents an alternative approach to visualization evaluation in which brain activity is passively recorded using electroencephalography (EEG). These measurements are used to compare different visualization techniques in terms of the burden they place on a viewer’s cognitive resources. In this paper, EEG signals and response times are recorded while users interpret different representations of data distributions. This information is processed to provide insight into the cognitive load imposed on the viewer. This paper describes the design of the user study performed, the extraction of cognitive load measures from EEG data, and how those measures are used to quantitatively evaluate the effectiveness of visualizations.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: General—Human Factors, Evaluation, Electroencephalography
EXPERIMENT DESIGN

- Asked participants to choose the box plot with the largest range from a set
- Varied representations
- Measured cognitive load from EEG brain waves
RESULTS

The simplest box plot is the hardest to interpret.
Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT
Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ — visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
REDESIGNED CHARTS
RESULTS

1. No significant difference between interpretation accuracy

2. No significant difference in recall accuracy after a five-minute gap

3. Significantly better recall for Holmes charts of both chart topic and the details (categories and trend) after long-term gap (2-3 weeks).

4. Participants found the Holmes charts more attractive, more enjoyable, and were easiest and fastest to remember.
ASSIGNMENT 2
IS NOW AVAILABLE
NEXT TIME...
What Makes a Visualization Memorable?

Michelle A. Borkin, Student Member, IEEE, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Student Member, IEEE, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister, Senior Member, IEEE

Fig. 1. **Left:** The top twelve overall most memorable visualizations from our experiment (most to least memorable from top left to bottom right). **Middle:** The top twelve most memorable visualizations from our experiment when visualizations containing human recognizable cartoons or images are removed (most to least memorable from top left to bottom right). **Right:** The twelve least memorable visualizations from our experiment (most to least memorable from top left to bottom right).

**Abstract**—An ongoing debate in the Visualization community concerns the role that visualization types play in data understanding. In human cognition, understanding and memorability are intertwined. As a first step towards being able to ask questions about impact...