Psychological processes of graphical perception: Past findings and our research proposal

Mariya Shiyko
Masanao Yajima

Columbia University, G4160 March 10th, 2006
Outline

A model of graphical information-processing
1. Input: quantitative/qualitative information
2. Retinal image
3. Visual cortex
4. Short-term and long-term memory

Implications

Research proposal
1. Variables in the study
2. Hypotheses
3. Sample
4. Materials
5. Procedure

References
A model of graphical information-processing
(Kosslyn, Wilkinson)

Input (quantity & quality) → Retina → Visual Cortex → Short-term Long-term memory
Format

- **Line graphs**
  - emphasize $x - y$ trends
  (Carswell et al., 1987; Shan et al., 1999; Zacks & Tversky, 1999)

- **Bar graphs**
  - stimulate discrete comparisons
    (Shah & Shellhammer, 1999) and
  - reduce bias towards reporting only one variable

- **Pie charts**
  - preferable for comparing relative proportions
    (Kosslyn, 1994; Wilkinson, 1999)
Dimensionality

- 3D graphs enhance comparison of 3 variables (Wickens et al., 1994)
- 3D displays interfere with reading individual data points (Shah et al.)
- 3D objects require prior knowledge of reading similar graphs (Shah et al.)
- For two-variable relationship depicted in 3D, accuracy and speed of data reading decreases/does not change (Fischer, 2000; Spence, 1990)
Colors

- Colors may be misleading when used for continuous data \cite{Cleveland85}
- The use of colors may falsely suggest presence of categorical data \cite{Cleveland85}
- Colors help to group data in scatter plots \cite{Lewandowsky89}
- Colors reduce the load of short-term memory when used as referents and are preferred to legends \cite{Kosslyn94}
Rank ordering
(most accurate to least accurate)
of visual dimensions used to represent quantitative information
(adopted from Cleveland & McGill, 1984)

1. Position along a common scale
2. Position along nonaligned scales
3. Length, direction, angle
4. Area
5. Volume, curvature
6. Shading, color saturation
The retinal image

- Depends on lightning
- Depends on viewing position
- Depends on color contrasts
- Black & white colors are best suited for visual perception
The visual cortex

extracts spatial frequency, orientation, and other features needed to construct complex visual scenes

- Follows the principles of absolute & relative discriminability (Kosslyn, 1989)
- Most discriminable symbols are verticals, horizontals, & diagonals (Cleveland, 1993; Wilkinson, 1993): ‘0’, ‘-’, ‘!’, ‘X’
Principles of perceptual organization

- **Gestalt laws:**
  - a) good continuity: ‘---------’
  - b) proximity: ‘xx xxx xx’
  - c) similarity: ‘!!!OOO’
  - d) good form: ‘{ }’ vs. ‘{ ¥’

- **Integral vs. separable dimensions** (Kosslyn, 1989)

- **Processing priorities** (Kosslyn, 1989)

- **Processing limitations** (Kosslyn, 1989)
  - a) finite capacity (4-7 perceptual groups)
  - b) purposeful unit binding

- **Informative mapping** (e.g. labels match the axes)
Memory

- Short-term memory
- Long-term memory

- Discriminability
- Distortion
- Organization
- Priorities ...

Capacity limits

Reorganization

Knowledge

Encoding
Knowledge is stored in forms of **schemas** - mental structures that represent some aspect of the world. People use schemas to organize current knowledge and provide a framework for future understanding. (Wikipedia)
Knowledge/schemas about graphs that interfere with graph readings:

1. Viewers tend to favor 45 degree lines rather than steeper or flatter lines when required to remember graphical representations of data (Schiano & Tversky, 1992; Tversky & Schiano, 1989)

2. Viewers tend to expect independent variable plotted on X axis & dependent variable - on Y axis

3. Viewers confuse slope & height/value magnitude (Gattis & Holyoak, 1995 Bell and Janvier, 1981)

4. Novice viewers rely on their prior knowledge about the content rather than the information depicted in the graph (Shah 1995)
Content knowledge

- Influences on interpretation of graphs
- Influences on memory
- Is especially important for novice graph viewers

<table>
<thead>
<tr>
<th>Easy</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Distance</td>
<td>* Acceleration</td>
</tr>
<tr>
<td>* Position</td>
<td>* Velocity</td>
</tr>
<tr>
<td>* Time</td>
<td></td>
</tr>
</tbody>
</table>
Viewers tend to overestimate the correlation between variables when they believe that these variables are related (Freedman and Smith, 1996; Jennings et al., 1982).

Viewers more likely to make inferences about general trends (main effects) when viewing graphs with familiar relationships (Shah & Shellhammer, 1999).

Younger students tend to interpret abstract representation of data as an iconic representation of real event, e.g. height instead of growth, location instead of speed (Janvier 1981; Leinhardt et al., 1990).
Implications
Implications

- Choose the format depending on the communication goal
- Use multiple formats to communicate the same data
- Use 3D with caution
- Choose colors thoughtfully
- Reduce working memory demands
- Follow the principles of perception
- Be aware of your audience
Research proposal
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Prior knowledge about graphs</td>
<td></td>
</tr>
<tr>
<td>• Prior statistical knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• description of relationships b/w 3 variables</td>
</tr>
</tbody>
</table>
| | • Number of responses
| | • Correct
| | • Incorrect |
| | • Identification of misleading graphs |
Hypotheses

H1: people with prior knowledge of graphs will describe relationship b/w 3 variables
H2: people with prior knowledge of graphs will notice discrepancies in graphical representations
H3: people with prior knowledge of graphs will report more information than people with no such knowledge
H4: people with prior knowledge of graphs will have more correct responses compared to people with no such knowledge
Sample

- Students from the statistical graphing class
- Students from other classes (?)

We will differentiate students by:
- prior knowledge of graphs
- prior statistical knowledge
- gender
- age
- education
Materials

- 4 graphs adopted from other studies (and modified)
- All graphs represent relationship between 3 variables
- Some graphs contain misleading information about the relationship
Procedure

- Students will be given testing instructions
- Students will fill out personal information
- Testing will take 16 minutes (4 minutes per graph)
- Testing materials will be collected and processed
Your feedback

- About hypotheses
- About procedure
- About materials
- ???
Selected references