Notes on Assignment 2

- Draft presentations
- Critical participation
  - Group pairs will be up online before the presentation day.
  - One group is formally required to ask questions after the other finishes the presentation.
- Evaluation methods may be used
  - Heuristic evaluation: Informally evaluate against some guidelines for graphic design
  - Cognitive walkthrough
  - Interviews
  - Questionnaires

This lecture

- Recap: HCI evaluation
- Gestalt principles of perception
- Graph evaluation examples

Recap: HCI evaluation

- Visualization is more than making pretty pictures
- We visualize to:
  - Understand data
  - Communicate
- There are many ways of visualizing the same data
- To make better visualizations, we need
  - Knowledge of how people make sense of data
  - Evaluate the visualizations

Interview

- Qualitative technique
  - Gathering information about users by talking directly to them
  - A method for discovering facts and opinions of the users.
- Format:
  - It is usually done by one interviewer speaking to one user at a time.
  - Structured interviews: a pre-defined set of questions and users
  - Open-ended interviews: allows for an exploratory approach to uncover unexpected information.
- Problems:
  - The unstructured nature of the resulting data can be easily misinterpreted.
Questionnaire

- Qualitative technique
  - But results can be quantified
- Preparation
  - Keep questions simple, be clear and concise
  - Group questions appropriately & give explanation
- Pilot questionnaire before distributing it
  - It is still unreasonable to think that any one person can anticipate all the potential problems
- Problems
  - It is only as good as the questions it contains

Question types

- General:
  - On average, how much time per week do you spend on this system?
    1) less than 1 hour  3) 4 to less than 10 hours
    2) 1 to less than 4 hours  4) over 10 hours
- Open-ended:
  - What are the features you think helpful, if any?
  - What are the features you think can be improved, if any?

Question types

- Closed:
  - Which of the following have you used? (tick all that apply)
    1) word processor  2) database  3) spreadsheet
  - How easy was it to understand the drawing?
    1) Very easy  2) Easy  3) Average  4) Difficult  5) Very Difficult
- Scale:
  - Please indicate how much effort you devoted for this task based on a scale from 0-6?
    0 (extremely easy)  1  2  3  4  5  6 (extremely difficult)

Questionnaire

- Established questionnaires will give more reliable and repeatable results than ad hoc questionnaires.
- Three questionnaire for assessing the perceived usability of an interactive system:
  - Questionnaire for User Interface Satisfaction (QUIS) (1988)
  - Computer System Usability Questionnaire (CSUQ) (1995)
  - System Usability Scale (SUS) (1996)

Analytic inspection

- Benefits:
  - Generate results quickly with low cost.
  - Can be used early in the design phases
- Heuristic evaluation
  - Experts review the systems against a list of principles
- Cognitive walkthrough
  - Starts with a task analysis that specifies the sequence of steps or actions required by a user to accomplish a task.
  - Then work through the steps

Empirical evaluation

- Usability test
  - Think aloud, eye tracking
  - Formative: helps guide design
  - Single UI, early in design process
  - Few subjects
  - Identify usability problems
  - Qualitative feedback from users
- Controlled experiment
  - Summative: measure final result
  - Compare multiple UIs
  - Many subjects, strict protocol
  - Quantitative results, statistical significance
Gestalt principles

- Theories of visual perception developed by German psychologists in the 1920s.
- These theories attempt to describe how people tend to organize visual elements into groups.
- Similarity
  - Similar elements tend to be grouped together

Proximity
- Elements tend to be grouped together according to their nearness

Continuation
- Smooth continuity is preferred to abrupt changes of direction

Connectedness [Ware 2004]
- Connecting different graphical objects by lines can be a more powerful grouping principle than proximity, colour, size, or shape.

Figure-Ground relationship
- The eye differentiates an object from its surrounding area.

Closure
- We tend to see complete figures even when part of the information is missing.
Gestalt principles

- **Symmetry**
  - the whole of a figure is perceived rather than the individual parts which make up the figure.

- **Area**
  - The smaller of two overlapping figures is perceived as figure while the larger is regarded as ground.
Another example

- According to research at Cambridge University, it doesn't matter in what order the letters in a word are, the only important thing is that the first and last letter be at the right place. The rest can be a total mess and you can still read it without a problem. This is because the human mind does not read every letter by itself, but the word as a whole.

Graph Visualization

- Efficiency: The running time of algorithms should be reasonably fast.
- Elegance: algorithms should be easy to understand and easy to code; final drawings should be beautiful.
- Effectiveness: graph viewers should understand the underlying data quickly and correctly.

Effectiveness

- Visualization designers are often satisfied with the "coolness" of the technologies they introduced.
  - technology that looks "cool" to the designer might be too complex or superfluous for real users
- It is assumed that graphs should be effective when drawn conforming to some predefined criteria.
  - Maximise symmetry
  - Minimise edge crosses
  - Maximise angular resolution
- However, common senses and intuitions are not reliable.
  - Users, data sets, tasks are different
Which one is better?

• We do not know until we actually evaluate them.

History of graph visualization technology

Criteria assumed by designer

Theories from general psychology

Theories of how people read graphs

Derived criteria

Human tests

Maksa algorithms

(a) 1980's approach

(b) 1990's approach

(c) Ideal approach

Controlled experiment on performance

• Purchase et al. (1995) Validating graph drawing aesthetics. GD95, 435-446.

Design and analysis

• One dense graph
• IV1: bends, crosses, symmetry
• IV2: high/medium/low presence
• DV: accuracy in fixed time (45 seconds)
• Task: find the shortest path between two nodes
• Paper based, within-subjects, random order, a "filler" task
• Friedman ANOVA.

Results and discussion

• Bends and crossings are important. Symmetry needs further examination.
• A pioneering work that provides empirical evidence for intuition-based aesthetic criteria.

Limitations

• Graphs are difficult to control (confounding factors)
  • Elements are interconnected to each other.
  • Change in on criteria can lead to change in another.
• For example, making more crossings can also make more sharp turns on the paths.
  • Gestalt principle of continuity
An experiment without manipulating IVs


Design and analysis

- Create a set of random graphs, drawn with a spring algorithm
- Highlight two nodes, record the measurements of the predictor variables
  - Continuity, number of crossings, crossing angles, number of branches, shortest path length....
- Record the response variable
  - Response time
- Regress response variable on predictor variables to detect their relationships.
- Task: shortest path between two highlighted nodes
- Analysis: Correlation and multiple regression

Results and discussion

\[
rt = 0.414 sp + 0.406 con + 0.317 cr + 0.172 br
\]

- Path continuity (con) and number of crossings (cr) on the shortest path are important factors.

Beyond time and error (1)

- Response time and accuracy are common performance measures
- Given two different visualizations, one can achieve the same level of performance by devoting different cognitive efforts.
- Traditional performance measures alone may not be sensitive enough to detect this cognitive difference.
- An extra measure is needed to reflect this cognitive difference.

Beyond time and error (1)

- In graph evaluation, it is claimed that a visualization is better if people take less time and make fewer errors when performing a task.
- However,
  - It is not very often that a visualization can achieve a highest accuracy and a shortest time at the same time.
  - A unified measure is needed to judge the overall quality.

Mental effort and visualization efficiency

Mental effort and visualization efficiency

- **Mental effort** ($ME$) can be measured by obtaining effort devoted, based on scale from 1 to 9.
- **Visualization efficiency** ($E$): difference between response accuracy ($RA$) and cognitive cost ($ME + RT$)

$$E = \frac{RA - ME - RT}{\sqrt{3}}$$

where $RA$ - response accuracy
$ME$ - Mental effort
$RT$ - Response time

Design and analysis

- IV1: Network size (network 1, network 2, network 3)
- IV2: Visualization type (filtered, combined)
- IV3: Task complexity (simple, medium, complex)
- DV: Time, accuracy, and:
  - Mental effort devoted from 1 (very low) to 9 (very high)
- Within subjects
- ANOVA (w-s)

Results

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean</th>
<th>Mean</th>
<th>$F(1,29)$</th>
<th>$p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>11.39</td>
<td>0.21</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Filtered</td>
<td>10.77</td>
<td>0.97</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.95</td>
<td>0.95</td>
<td>0.16</td>
<td></td>
</tr>
</tbody>
</table>

* $p < 0.05$

- If only performance was measured, we would have concluded that the two visual versions were equivalent.

Discussion

- Individual measures have their own disadvantages and advantages.
- Visualization efficiency measure is more sensitive, therefore a better measurement.
- Visualization efficiency combines all individual measures, therefore it can be used for judging overall quality.

Beyond time and error (2)

- Where the time is spent and how the performance is affected?
- What is the mechanism of crossings affecting performance?
- Time and error performance logging
  - treat the human as a "black box", which tell us what, but not how and why
- Eye tracking may give insight as to how
- Post-interview and questionnaire tell us why

Extension to the formula

- Instead of difference, visualization efficiency can be measured as ratio.

$$E = \frac{RA - ME - RT}{\sqrt{3}}$$

$$E = \frac{RA}{ME + RT}$$
- Two exploratory eye tracking experiments
  - Ex1: small and sparse graphs
  - Ex2: larger and denser graphs
- Three confirmatory controlled experiments
  - Ex3a: existence of geodesic-path tendency
  - Ex3b: effects of geodesic-path tendency
  - Ex4: effects of crossing angles

Research Methodology

Top-down approach:
- two eye tracking experiments

- Develop theories based on performance measures
- Refine the theories through eye tracking
- Confirm

Bottom-up approach:
- three controlled experiments

Eye Tracker

Experiment 1

- Task: find the shortest path between two highlighted nodes.
- Time, error and eye movements were recorded.
- Questionnaires and interviews.

Examples of Stimuli

Results: Time and Error

- Overall, subjects spent significantly more time with crossing drawings than with non-crossings
- However, on some specific instances, this was not the case
Results: Eye Tracking Video Data

• Crossings had little impact on eye movements.
• Geodesic-path tendency: subjects seemed to follow the geodesic path between the current node and target node.

Possible Reasons for the Lack of Crossing Effects

• Crossing angles may inhibit readability [Ware et al. 2003].

Experiment 2

Crossing angle: graphs were drawn with three conditions:
– No crossings on the path
– Small-angle crossings
– Large-angle crossings

Stimuli

Results

• Effects of crossing angles were significant on time

Results: Eye Tracking Video Data

No crossings: eye movements were smooth and fast.

Large crossing angle: eye movements were still smooth, but slower.

Small crossing angle: eye movements were very slow and no longer smooth (back-forth moves at crossing points).
• Observations of eye-tracking videos need **confirmation**:
  - Existence of geodesic-path tendency (Ex3a).
  - Effects of geodesic-path tendency (Ex3b).
  - Effects of crossing angles (Ex4).

### Experiment 3a: Existence of geodesic-path tendency

- Two separate paths between the two highlighted nodes

### Results

- Subjects followed geodesic-closest path 75% of the time
- People have a “**geodesic-path tendency**”

### Ex3b: Effects of Geodesic-path Tendency

Each graph was drawn in two ways:
- **Difficult**: dead-end branches going toward target
- **Easy**: dead-end branches away from the target
Results

- Difficulty significantly affected response time and errors.

Discussion and an Example of Video

- Geodesic-path tendency affected performance of shortest path tasks significantly.

Ex4: Effects of Crossing Angles

- Linear component and a quadratic component in the relationship between time and angle.
- 70° degree angle was equivalent to that with no-crossings.

Conclusion

- Eye movements tell us how:
  - How crossings affect eye movements and performance
  - Impact of crossings differs with crossing angle and size of graphs
  - People have geodesic-path tendency in searching shortest paths
- This gives guidance to designers of graph visualization technology
- To obtain insights on why, post-task interviews were used.