Background Notes on Experimental Design & Evaluation

1. What are some different types of questions you might want to answer through evaluation?
   (In the brackets I suggest possible evaluation approaches for each of the questions.)

   Is my new visualization technique usable? Can users do the tasks I intended?
   (This question suggests doing a usability study, where you ask people to do a set of tasks and identify places where they have problems.)

   Is visualization technique A faster than technique B for a particular task? Are users more accurate with technique A or B?
   (Do a quantitative experiment to compare A to B for one or more tasks.)

   What features of a visualization help users gain insight into problem X?
   (Case study. Do a qualitative study where you provide experienced domain experts with one or more visualization tools and observe how they work with each one to examine their own data. Interview them extensively to determine how they work and what features were helpful.)

   Is visualization software suite A (e.g. Spotfire) better than visualization software suite B (e.g. Tableau) for a particular domain and task?
   (Although there are several quantitative experiments out there that do comparisons like this, the quantitative results are not terribly helpful because there are many differences between the software tools. However, a qualitative comparison of different tools to determine what features of each are beneficial could be very insightful.)

   What are the needs of users in domain X for problem Y?
   (Qualitative pre-design study. Conduct interviews, contextual interviews, and / or observation of users to determine current practice and challenges they encounter. Often these studies make use of a qualitative research approach called grounded theory.)

2. What is the difference between qualitative and quantitative evaluation?

   Quantitative: carefully controlled experimental conditions to enable you to statistically compare numeric results. Typically involves simplistic tasks.

   Qualitative: focuses on non-numeric results. More concerned with examining realistic data and tasks and understanding the reasons why different features are beneficial. May focus on understanding / characterizing a problem domain rather than determining which solution is best.

3. Explain some ways in which an experiment could be biased.
   • Experimenter may indicate (possibly without realizing it) which visualization technique they prefer. (E.g. naming one technique the “Baseline Method”.)
• Participants may know that one visualization technique was developed by the experimenter (or by the experimenter’s organization) and the other was not.
• Observers will tend to see what they expect to see or want to see, so if only one live observer is used, the results may be skewed.
• Experimenter may ask leading questions (e.g., “What was bad about visualization technique A?” as compared to “What did you like and dislike about technique A?”)

4. Experimental terms:

**Hypothesis** A clear and testable statement of what you expect the result of an experiment to be.

**Subject (or Participant)** A person who takes part in an experiment (and is not an experimenter).

**Within-subjects design** All participants participate in all experimental conditions. This requires fewer participants and enables each participant to compare the conditions so they can assess which one they prefer. However, there are ordering effects that must be controlled (see below).

**Between-subjects design** Each participant takes part in only one experimental condition. This requires more participants and does not enable participants to compare the conditions, but there are no order effects.

**Independent variable** An experimental variable that is explicitly varied by the experimenter. (E.g. in an experiment comparing greyscale to a rainbow scale, colour scale is an independent variable).

**Dependent variable** An experimental variable that is measured. (E.g., you might measure time to complete a task, number of errors, distance traveled by the mouse, etc.)

**Control** A control condition is a condition that receives no treatment (e.g. a placebo in a drug trial). In visualization, this might be something like showing no overview at all in an experiment comparing different types of overview windows.

**Confounding factor** Any factor that varies systematically with an independent variable and therefore could provide an alternate explanation for any results.

**Task** The thing you ask a participant to do during an experiment. (E.g. find an item of type X, or compare items X and Y to see which is larger, etc.)

**Likert scale** A numbered rating scale used by participants to answer a structured question. Likert scales usually range from 1-5, 1-7, or 1-9, where the numbers represent answers such as “strongly disagree”, “disagree”, “agree”, “strongly agree”, etc.
**Structured interview** An interview in which all the questions have been decided beforehand by the interviewer. A semi-structured interview is one in which some questions have been decided but the interviewer may also ask additional questions as they see fit.

5. **What are some things that must be controlled in an experiment? How might you do this?**

   - Who gets what condition in a between-subjects’ design. People should be assigned randomly to groups.
   - When and where each condition is given. Should be kept consistent across all participants.
   - Instructions. Should be consistent across all participants.
   - Order of items:
     - If the items do not need to be grouped together temporally, then you can randomize the order. E.g., in a target search experiment you can randomize the number of distractor items.
     - If the items do need to be grouped (aka “blocked”), you can counter-balance the order. E.g., Half the participants might use visualization A first, then visualization B. The other half do the opposite. This gets more difficult with more than a few conditions because there are many possible orders (but there are ways to do partial counterbalancing – e.g. using Latin Squares).

Generally, all factors in an experiment should either be controlled or systematically varied. To control a variable you either:

   - Keep it consistent across all groups.
   - Randomize it.
   - Counter-balance it.