

## Appendix: Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences About Reliability of Variable Ordering

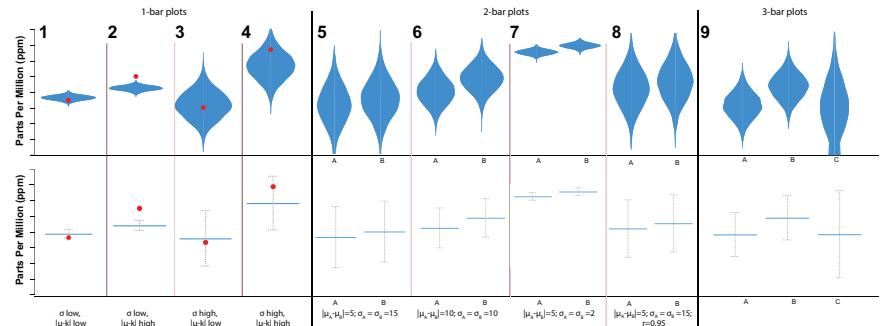
Jessica Hullman<sup>1,\*</sup>, Paul Resnick<sup>2</sup>, Eytan Adar<sup>2</sup>,

**1 Information School, University of Washington, Seattle, WA, USA**

**2 School of Information, University of Michigan, Ann Arbor, MI, USA**

\* [jhullman@uw.edu](mailto:jhullman@uw.edu)

### 1 Visualization Stimuli



**Figure 1.** The eight tasks that subjects completed, shown as violin plots and error bars. Click the image to play the HOPs animation.

### 2 Study Apparatus

Each Mechanical Turk worker who participated in our study completed 4 one-variable plot screens, 4 two-variable plot screens, and 1 three-variable plot screen. An introductory screen preceded the task, and a final screen asked follow-up questions to check understanding.

A testable interface to the study is available [here](#). All interface code, including the software for generating all visualization stimuli, is also included as supplemental materials.

Additionally, an example of each screen is shown below.

**"Interpret Data Charts" (with possible \$0.90 bonus)**

You will see charts showing quantities and giving you information on the uncertainty of the quantities. Here are examples of what the questions are like:

- What is the value of the quantity at a marked point on the chart?
- What is the average value of the quantity?
- How often is the quantity above a marked point on the chart?
- How often is the quantity between two values marked on the chart?

**Scoring**

You will answer one or more questions for each of 9 total charts. We will randomly choose one of your answers for each chart, and compare your answer to the true answer for that question. If your answer is close to the true answer, we will add an additional \$0.15 to your bonus (up to \$1.35 total bonus).

**Participation and Consent**

This HIT is part of a research project to understand how individuals make decisions about visualized information. We expect that this task will take no more than 40 minutes to complete. Your participation is completely voluntary. If at any time you wish to stop participating in the study, you may return the HIT. No personally identifying information will be recorded if you participate or if you discontinue participation.

If you wish to speak with the researcher contact the requester Jessica Hullman. For other questions about your rights as a research participant, or to obtain information, ask questions or discuss any concerns about this study with someone other than the researcher(s), please contact the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board, 540 E Liberty St., Ste 202, Ann Arbor, MI 48104-2210, (734) 936-0933, toll free (866) 936-0933, irbhsbs@umich.edu.

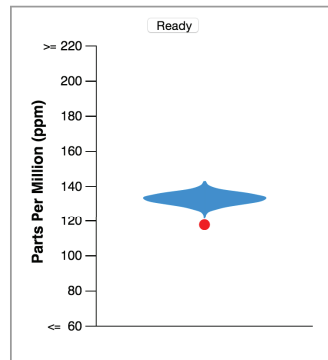
By clicking continue, you consent to be part of this study.

Continue

**Figure 2.** Task introduction screen.

**"Interpret Data Charts" - Number 1 out of 9**

The chart summarizes data that scientists have gathered by measuring the concentration of a chemical solute (measured in parts per million) in many vials of sea water.



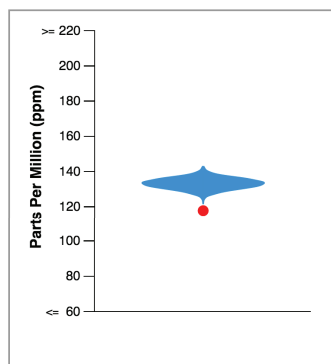
The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

When you are ready to start, press Ready at the top of the plot

**Figure 3.** First one-variable screen. Subject must press the 'Ready' button to see the question. In the HOPs treatment, pressing 'Ready' also starts the animation.

"Interpret Data Charts" - Number 1 out of 9

The chart summarizes data that scientists have gathered by measuring the concentration of a chemical solute (measured in parts per million) in many vials of sea water.



The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Q1: What is the value of solute in parts per million (ppm) at the red dot?

Q2: What is the average measurement of solute in parts per million (ppm)?

Q3: How often are the measurements above the value of the red dot? Answer in terms of the number of times out of 100.

 times out of 100

Q4: How often will the measurements lie between 130 and 160? Answer in terms of the number of times out of 100.

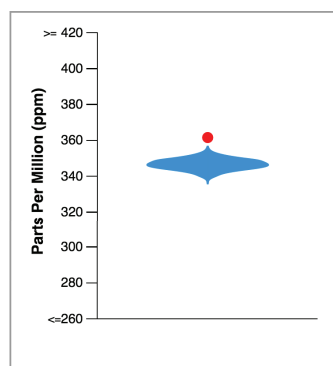
 times out of 100

Continue

Figure 4. First one-variable screen, after subject has hit 'Ready' button.

"Interpret Data Charts" - Number 2 out of 9

The chart summarizes data that scientists have gathered by measuring the concentration of a chemical solute (measured in parts per million) in many vials of sea water.



The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Q1: What is the value of solute in parts per million (ppm) at the red dot?

Q2: What is the average measurement of solute in parts per million (ppm)?

Q3: How often are the measurements above the value of the red dot? Answer in terms of the number of times out of 100.

 times out of 100

Q4: How often will the measurements lie between 340 and 370? Answer in terms of the number of times out of 100.

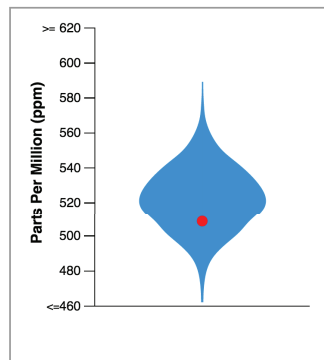
 times out of 100

Continue

Figure 5. Second one-variable screen.

"Interpret Data Charts" - Number 3 out of 9

The chart summarizes data that scientists have gathered by measuring the concentration of a chemical solute (measured in parts per million) in many vials of sea water.



The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Q1: What is the value of solute in parts per million (ppm) at the red dot?

Q2: What is the average measurement of solute in parts per million (ppm)?

Q3: How often are the measurements above the value of the red dot? Answer in terms of the number of times out of 100.

 times out of 100

Q4: How often will the measurements lie between 520 and 550? Answer in terms of the number of times out of 100.

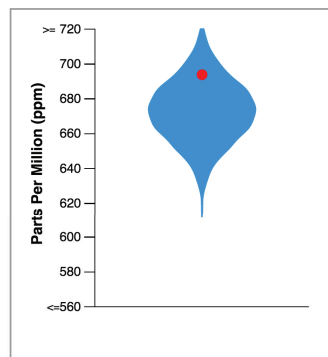
 times out of 100

Continue

Figure 6. Third one-variable screen.

"Interpret Data Charts" - Number 4 out of 9

The chart summarizes data that scientists have gathered by measuring the concentration of a chemical solute (measured in parts per million) in many vials of sea water.



The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Q1: What is the value of solute in parts per million (ppm) at the red dot?

Q2: What is the average measurement of solute in parts per million (ppm)?

Q3: How often are the measurements above the value of the red dot? Answer in terms of the number of times out of 100.

 times out of 100

Q4: How often will the measurements lie between 670 and 700? Answer in terms of the number of times out of 100.

 times out of 100

Continue

Figure 7. Fourth one-variable screen.

"Interpret Data Charts"

You will now see charts showing two quantities and giving you information on the uncertainty of the quantities. You will answer one question:

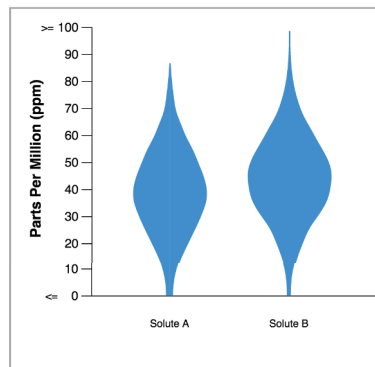
- How often is one of the quantities larger than the other quantity?

Continue

Figure 8. Introduction to two-variable plots.

"Interpret Data Charts" - Number 5 out of 9

The chart summarizes two data sets that scientists have gathered by measuring the concentration of two chemical solutes (A and B, each measured in parts per million) in many vials of sea water.



Q: How often is the measurement of solute B larger than the measurement of solute A? Answer in terms of the number of times out of 100.

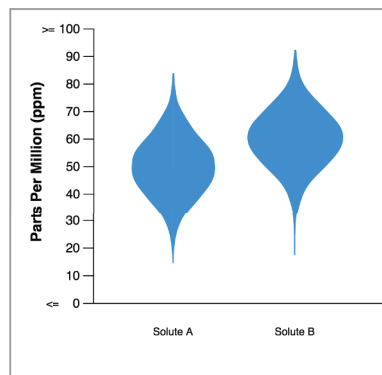
The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Continue

Figure 9. First two-variable screen.

"Interpret Data Charts" - Number 6 out of 9

The chart summarizes two data sets that scientists have gathered by measuring the concentration of two chemical solutes (A and B, each measured in parts per million) in many vials of sea water.



Q: How often is the measurement of solute B larger than the measurement of solute A? Answer in terms of the number of times out of 100.

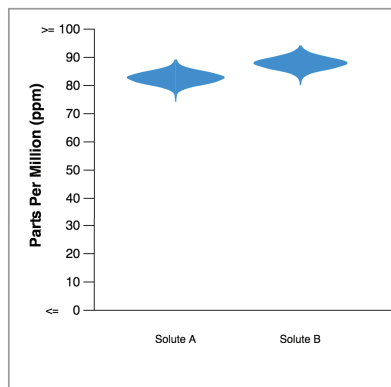
The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Continue

Figure 10. Second two-variable screen.

"Interpret Data Charts" - Number 7 out of 9

The chart summarizes two data sets that scientists have gathered by measuring the concentration of two chemical solutes (A and B, each measured in parts per million) in many vials of sea water.



Q: How often is the measurement of solute B larger than the measurement of solute A? Answer in terms of the number of times out of 100.

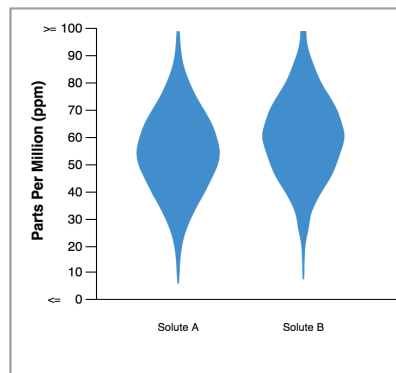
The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Continue

Figure 11. Third two-variable screen.

"Interpret Data Charts" - Number 8 out of 9

The chart summarizes two data sets that scientists have gathered by measuring the concentration of two chemical solutes (A and B, each measured in parts per million) in many vials of sea water.



Q: How often is the measurement of solute B larger than the measurement of solute A? Answer in terms of the number of times out of 100.

The width of the colored area at each level shows how many vials of sea water were found to have that particular amount of the chemical solute.

Continue

Figure 12. Fourth two-variable screen.

### "Interpret Data Charts"

You will now see a final chart showing three quantities and giving you information on the uncertainty of the quantities. You will answer one question:

- How often is one of the quantities larger than the other two quantities?

**Figure 13.** Introduction to three-variable screen.

### "Interpret Bar Charts" - Final Feedback

You are almost done with this HIT. Please fill out the following questions, then submit your work to finish this HIT. Your information will not be connected to your workerid in any way in analysis.

Think about your experience using the visualization to answer the questions. How difficult was it to make your estimates using the visualization?

- Very easy
- Somewhat easy
- A little easy
- Neutral
- A little difficult
- Somewhat difficult
- Very difficult

Was it clear to you what you were seeing in the charts? Please describe any trouble you had understanding the charts or task instructions.

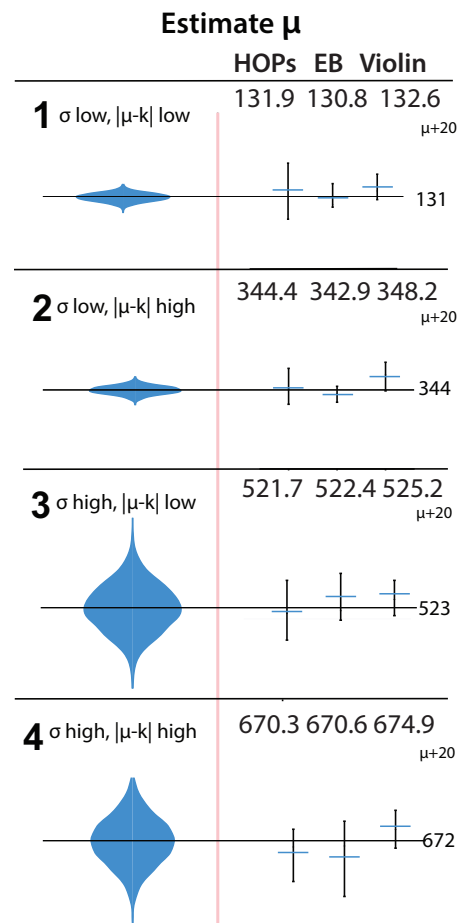
How often do you use any types of graphs, charts, or data maps?

- Daily
- 1 - 3 times per week
- 1 - 5 times per month
- Less than once a month

**Figure 14.** Final screen. Subjects are asked to rate the difficulty of the task and provide feedback on specific difficulties they encountered. Information is gathered about the subject's familiarity with data visualizations.

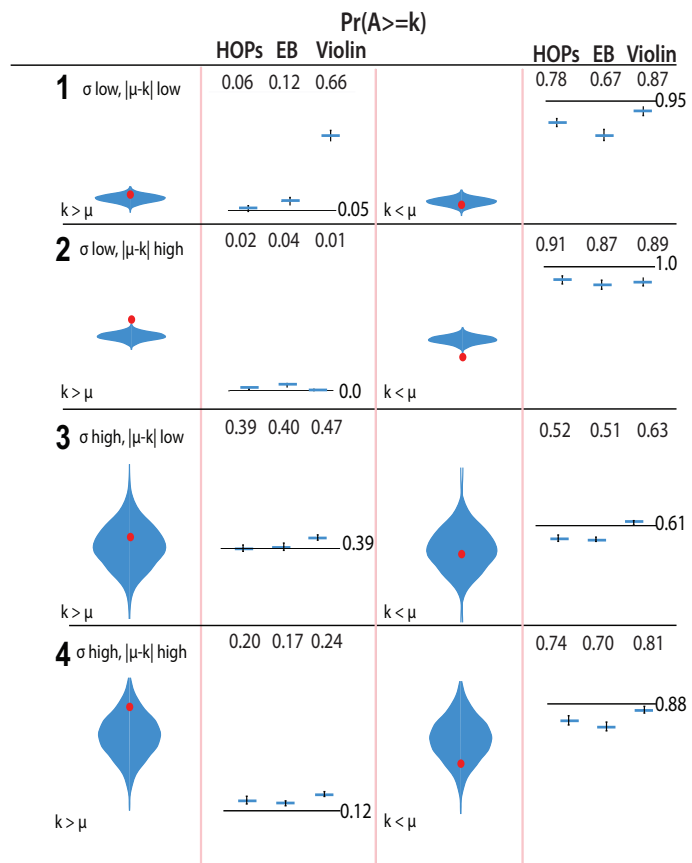
### 3 Detailed Results

#### 3.1 One-Variable Plots

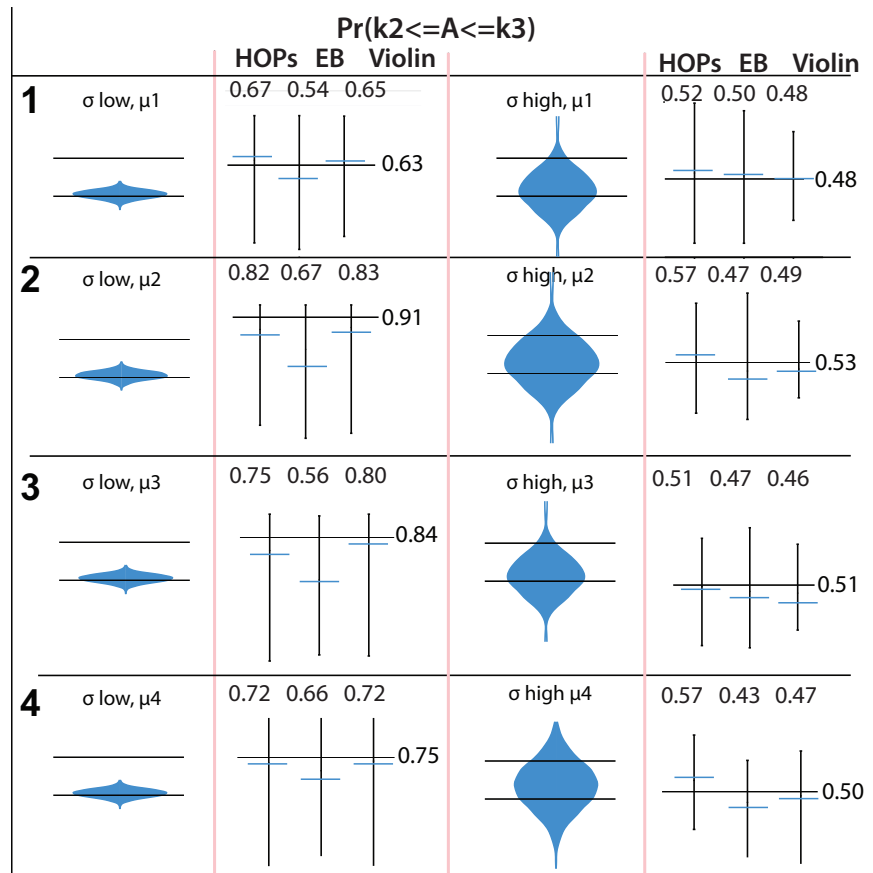


**Figure 15.** Stimuli (left) and raw responses (right) for estimation of  $\mu$ . The correct response is marked with a horizontal line. Subjects who used the violin plot consistently overestimated  $\mu$ . Click the image to play the HOPs animation.



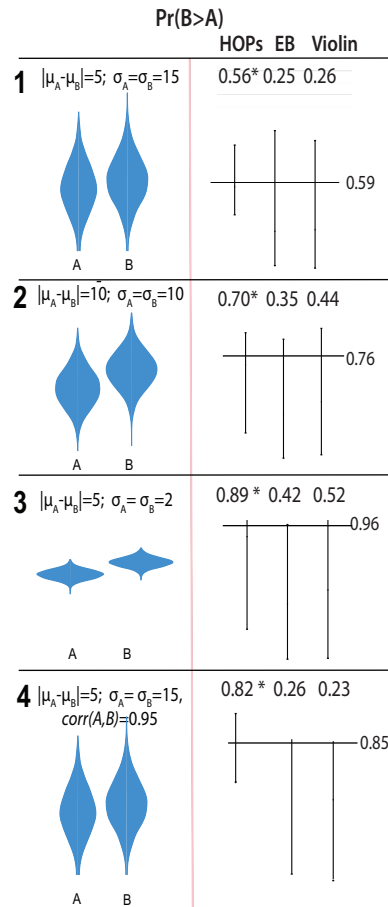


**Figure 16.** Stimuli (left) and raw responses (right) for estimate of  $Pr(A \geq k)$ . The correct response is marked with a horizontal line. Across all three conditions, subjects show a tendency to under- or overestimate  $Pr(A \geq k)$  corresponding to the relative position of the red dot. Click the image to play the HOPs animation.



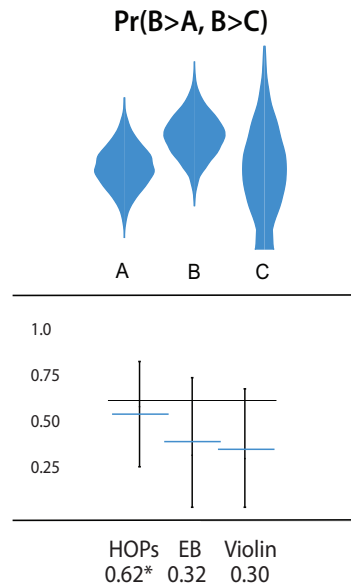
**Figure 17.** Stimuli (left) and raw responses (right) for estimate of  $Pr(k_2 \leq A \leq k_3)$ . The correct response is marked with a horizontal line. Subjects who used error bars and violin plots show a consistent tendency to underestimate  $Pr(k_2 \leq A \leq k_3)$ . Click the image to play the HOPs animation.

### 3.2 Two-Variable Tasks



**Figure 18.** Stimuli (left) and raw responses (right) for estimate of  $Pr(B > A)$ . The correct response is marked with a horizontal line. With the exception of HOPs subjects for distribution 1, subjects across all three conditions tended to underestimate  $Pr(B > A)$ . Click the image to play the HOPs animation.

### 3.3 Three-Variable Task



**Figure 19.** Stimuli (top) and raw responses (bottom) for estimate of  $\Pr(B > A, B > C)$ . The correct response is marked with a horizontal line. Subjects across all three conditions show a consistent tendency to underestimate  $\Pr(B > A, B > C)$ . Click the image to play the HOPs animation.