Visualising distributions

Reading group Statistics Communication
Sanne Willems
March 4, 2024
Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering

Jessica Hullman¹*, Paul Resnick², Eytan Adar²

1. Information School, University of Washington, Seattle, WA, United States of America,
2. School of Information, University of Michigan, Ann Arbor, MI, United States of America
My rating

- Experimental set-up
- Analysis, results, usefulness conclusion
Visualisations

Introduction
Three ways of visualising distributions

- Error Bars
- Violin Plot
- Hypothetical Outcome Plots (selected frames)
Error bars

• Static
• Represents range around central tendency (e.g. mean/median)

• Pro/con:
  • - Meaning of the bars often unclear
    • Standard error?
    • 95% confidence interval?
Violin plots

• Static
• Represents full distribution
• Pro/con:
  • + Less prone to misunderstanding the meaning of the bars
Hypothetical Outcome Plots

- Dynamic (and interactive)
- Each frame is a draw from a sample

Pro/con:
- - sampling error (finite sample)
- - reader needs to process information from frames
- + reader can think in frequencies instead of percentages
- + no need to understand encoding
Experiment set-up

Overview
Overview of the experiment

Scientists have measured the concentration of some chemical solutes (measured in parts-per-million) in many samples of sea water. We will show you plots based on their measurements and ask you questions about them.

9 tasks:
Participants

Study set-up
Participants

- Amazon Mechanical Turk
- 96 subjects per visualisation (randomly assigned)
- Standard criteria
  - U.S. workers
  - 95% rating or higher
- Payment:
  - Base reward
  - + bonus for correct answers

Exclusion criteria:
- Quality control measure
  - give value of red dot
  - > 30 from value, delete participant
- Incomplete answers
Participants

• Amazon Mechanical Turk
• 96 subjects per visualisation (randomly assigned)
• Standard criteria
  • U.S. workers
  • 95% rating or higher
• Payment:
  • Base reward
  • + bonus for correct answers

Exclusion criteria:
• Quality control measure
  • give value of red dot
  • > 30 from value, delete participant
• Incomplete answers
Instruments - visualisations

Study set-up
Sampled data to visualise

• Normal distribution
• Sample size = 5,000
• All visualisations show the distribution of the data instead of the sampling distribution (e.g. distribution of the mean)
Sampled data to visualise

• Normal distribution
• Sample size = 5,000
• All visualisations show the distribution of the data instead of the sampling distribution (e.g. distribution of the mean)

• Most optimal choices for error bar and violin plot (symmetry) 
  -> no disadvantage created for those plots
Error bars

The blue line shows the average amount of solute in all the seawater vials. The dashed lines show an error bar, a range above and below the average. 95% of the collected vials fall in the range defined by the dashed lines.
Violin plots

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.
Hypothetical Outcome Plots

Each plot shows the quantity of solute in one vial of seawater. Use the buttons at the top of the plot to pause, play, or step forward and back through the plots if you want to see them at your own pace.
Hypothetical Outcome Plots

Each plot shows the quantity of solute in one vial of seawater. Use the buttons at the top of the plot to pause, play, or step forward and back through the plots if you want to see them at your own pace.

• 5,000 frames, 400ms per frame
Hypothetical Outcome Plots

Each plot shows the quantity of solute in one vial of seawater. Use the buttons at the top of the plot to pause, play, or step forward and back through the plots if you want to see them at your own pace.

- 5,000 frames, 400ms per frame
- = ± 33 minutes
- so most watched just ± 101 frames
  -> less information compared to other visualisation types
Experiment

One set at the time
Overview

One variable

Two variables

Three variables
One variable

Hypothesis, set-up and results
Questions

1. What is the average measurement of solute in parts-per-million (ppm)?
2. How often are the measurements above the value of the red dot?
3. How often will the measurements lie between k2 and k3 ppm?

Outcome: Mean absolute error
Questions

1. What is the average measurement of solute in parts-per-million (ppm)?
2. How often are the measurements above the value of the red dot?
3. How often will the measurements lie between $k_2$ and $k_3$ ppm?

Outcome: Mean absolute error

Why not full distribution?
Distributions

• Distribution settings:

  | Table 1. One-Variable Distribution Types. |
  |-----------------|-----------------|-----------------|
  | Distr. | σ | \(D = |\mu - k_1|\) | \(Pr(X > \mu + D)\) |
  |-------|---|----------------|-----------------|
  | 1     | 3 | 5              | 5%               |
  | 2     | 3 | 20             | 0%               |
  | 3     | 17| 5              | 39%              |
  | 4     | 17| 20             | 12%              |

• \(\mu\): 131, 344, 523, and 672.
• 24 x 4 = 96 possible sequences of four trials
• randomly (\(Pr = 0.5\)) determine whether the red dot is above or below \(\mu\)
Distributions

- Distribution settings:

  - $\mu$: 131, 344, 523, and 672.
  - randomly assign to $24 \times 4 = 96$ possible sequences of four trials
  - randomly ($Pr = 0.5$) determine whether the red dot is above or below $\mu$
Hypothesis 1 + results

When estimating the mean of a single variable, subjects will have lower error rates using error bars and violin plots than HOPs where $\sigma$ is high, but not on distributions where $\sigma$ is low.
Hypothesis 1 + results

When estimating the mean of a single variable, subjects will have lower error rates using error bars and violin plots than HOPs where $\sigma$ is high, but not on distributions where $\sigma$ is low.

No statistical methods mentioned.
Probably ANOVA, but no info on how dealt with multiple measures.
Hypothesis 1 + results

When estimating the mean of a single variable, subjects will have lower error rates using error bars and violin plots than HOPs where $\sigma$ is high, but not on distributions where $\sigma$ is low.

Data agrees!
Hypothesis 2 + results

When estimating the probability of a random variable *above a threshold* or *between two thresholds*, subjects will have lower error rates using violin plots than error bars or HOPs.
Hypothesis 2 + results

When estimating the probability of a random variable *above a threshold* or *between two thresholds*, subjects will have lower error rates using violin plots than error bars or HOPs.

No significant differences, no clear pattern.
Hypothesis 2 + results

When estimating the probability of a random variable \textit{above a threshold} or \textit{between two thresholds}, subjects will have lower error rates using violin plots than error bars or HOPs.

No clear winner.
Two variables

Hypothesis, set-up and results
Question

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.

*Outcome: Mean absolute error*
Example

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.
Example

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.

Eh...?
Also depends on correlation?
Example

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.

Eh...?
Also depends on correlation?
Probably much easier for HOPs?
Hypothesis 3 + results

When estimating $Pr(B > A)$ in a bivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.
Hypothesis 3 + results

When estimating $\Pr(B > A)$ in a bivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.

- 3x no correlation, 1x high correlation
Hypothesis 3 + results

When estimating $\Pr(B > A)$ in a bivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.

- 3x no correlation, 1x high correlation
- Does not result in different plots for error bars / violin plots? How do participants know?
Hypothesis 3 + results

When estimating $\Pr(B > A)$ in a bivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.

Data agrees!
Hypothesis 3 + results

When estimating $\Pr(B > A)$ in a bivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.

Data agrees!
But this just says that the HOPs can convey more info?
Three variables

Hypothesis, set-up and results
Question

How often is the measurement of solute B larger than both the measurement of solute A and the measurement of solute C?

Answer in terms of the number of times out of 100.

*Outcome: Mean absolute error*
Hypothesis 4 + results

When estimating $\Pr(B > A \text{ and } B > C)$ in a trivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.
Hypothesis 4 + results

When estimating $\text{Pr}(B > A \text{ and } B > C)$ in a trivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.

Same issues..
Hypothesis 4 + results

When estimating $\text{Pr}(B > A \text{ and } B > C)$ in a trivariate distribution plot, subjects will have lower error rates using HOPs than violin plots or error bars.

No surprise, data agrees.
My conclusions
My conclusions

• Error bars and violin plots are good at conveying exact values in distributions,
• HOPs are pretty good at it too, unless variance is too large.
• HOPs are useful to compare (correlated) variables.
  • But error bars / violins can probably do that too if you plot composite variable B - A