Error Bars Considered Harmful

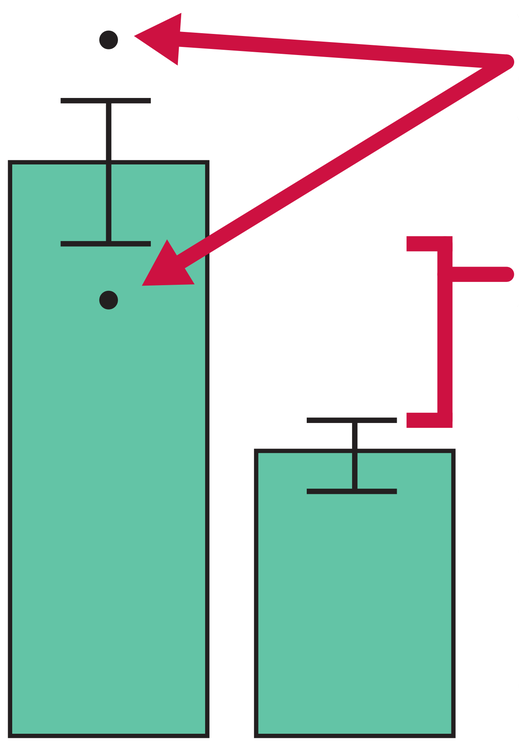
Michael Correll and Michael Gleicher
University of Wisconsin-Madison

Problem Statement

Statistical certainty is often displayed using bar charts with attached error bars. These types of displays are not ideal for displaying statistical inferences, which can be complex and counter-intuitive.

How can we present sample means and error in a more comprehensible way?

Can the choice of encoding improve the inferences of non-statistical general audiences ?



Values within the bar are perceived as liklier samples than values outside the bar

Even experts misinterpret how p-value is connected to the separation between error bars.

Error bars are ambiguous - can mean confidence interval, standard error, 1.5xIQR, ...

Error bars have no semiotic connection with uncertainty.

Conditions

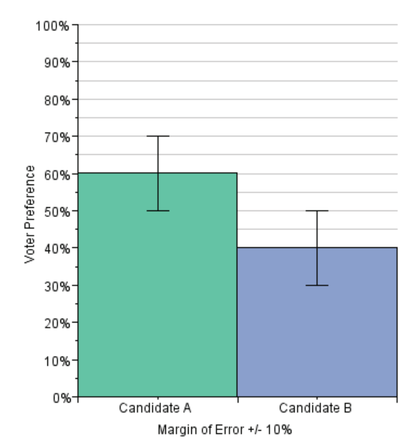
Bar charts with error bars are the standard encoding for mean+error data.

Violin charts directly encode a statistical distribution using width.

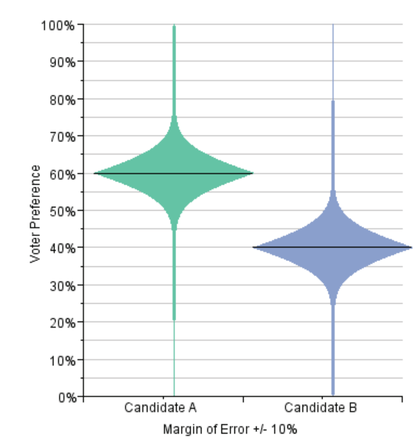
Gradient plots encode the likely location of a population mean with alpha value.

Pangloss plots do not encode the sample mean directly, but are instead a "quilt" of simulated "elections" based on a t distribution of sample means.

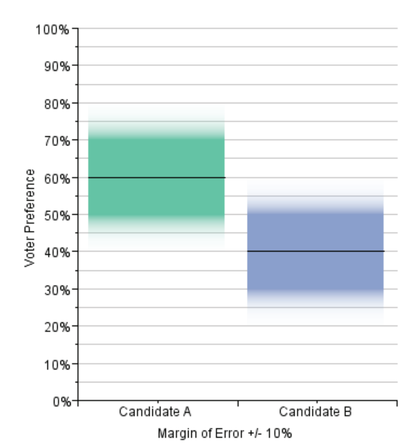
Bar Chart



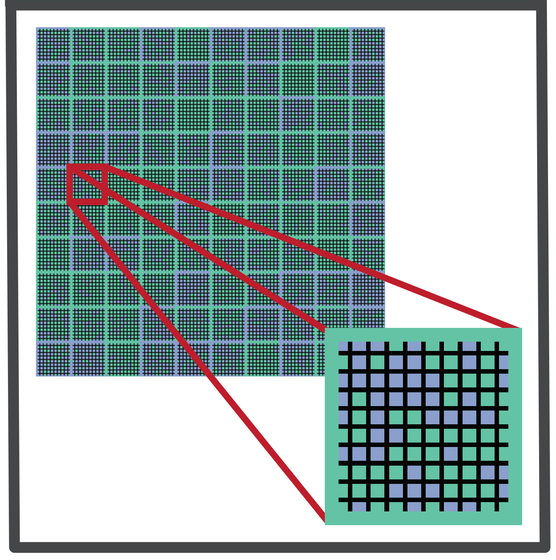
Violin Plot



Gradient Plot



Pangloss Plot



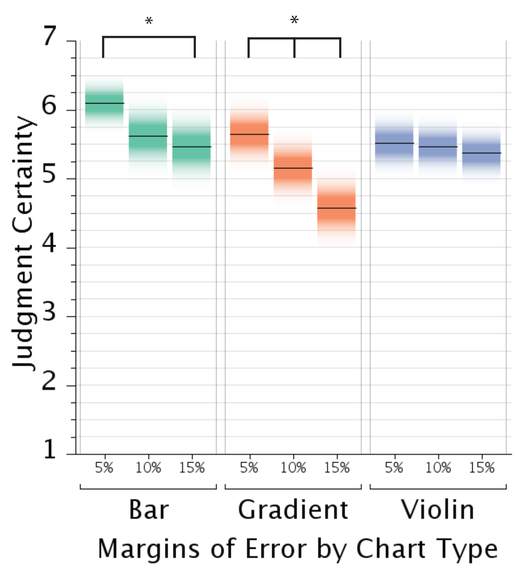
Methodology

We performed a study on Amazon's Mechanical Turk to determine if alternate encodings of mean and error would result in human predictions better aligned with statistical predictions.

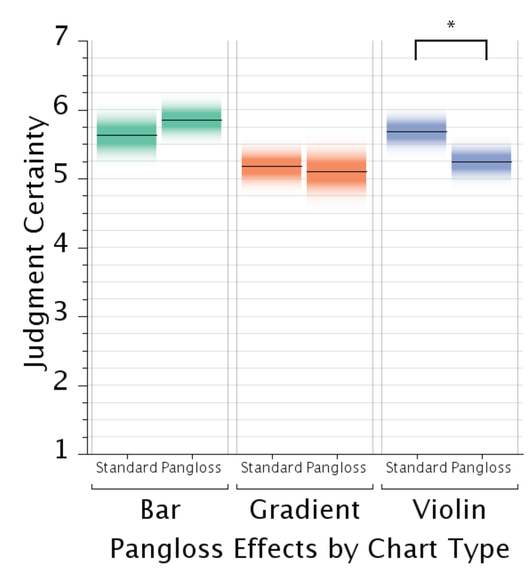
36 participants were each given 21 polls from fictional elections and asked to predict who would win the final election, and how confident they were in their prediction. Polling data was presented as one of three types of charts, with half of participants seeing a "Pangloss plot" in additon to a main chart type.

Results

Only for gradient plots does the margin of error consistently and significantly impact uncertainty in predictions



Adding simulated outcomes does not consistently impact uncertainty in predictions.



Results presented as a gradient plot. Fully saturate areas are within a 95% t-confidence interval.

This work was supported in part by NSF award IIS-1162037 and NIH award R01 AU974787.

Please consult poster summary for related works