

Error Bars Considered Harmful

Exploring Alternate Encodings for Mean and Error

Michael Correll

Michael Gleicher

University of Wisconsin-Madison

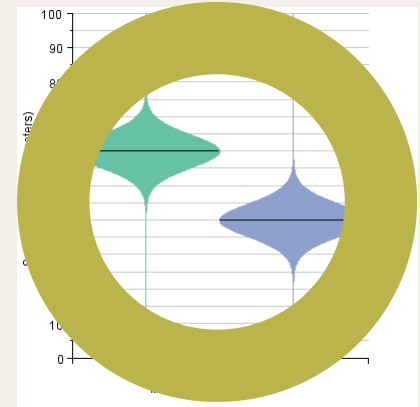
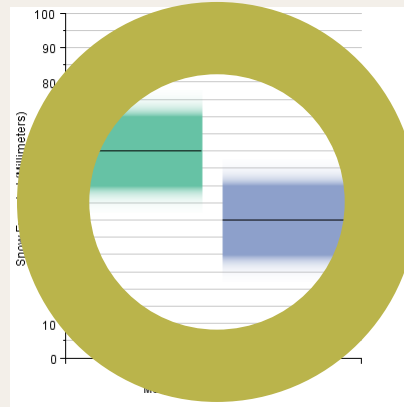


Don't Use Error Bars

They don't work as advertised



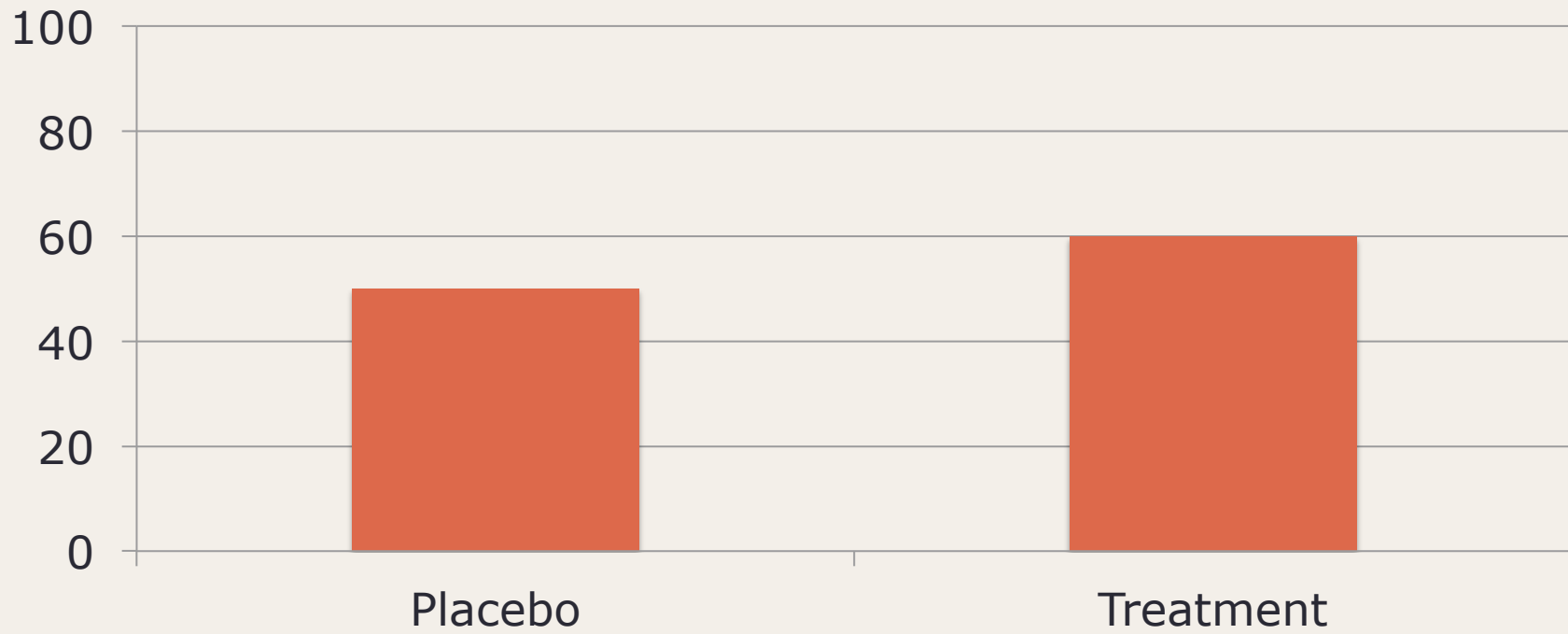
Try something else instead!

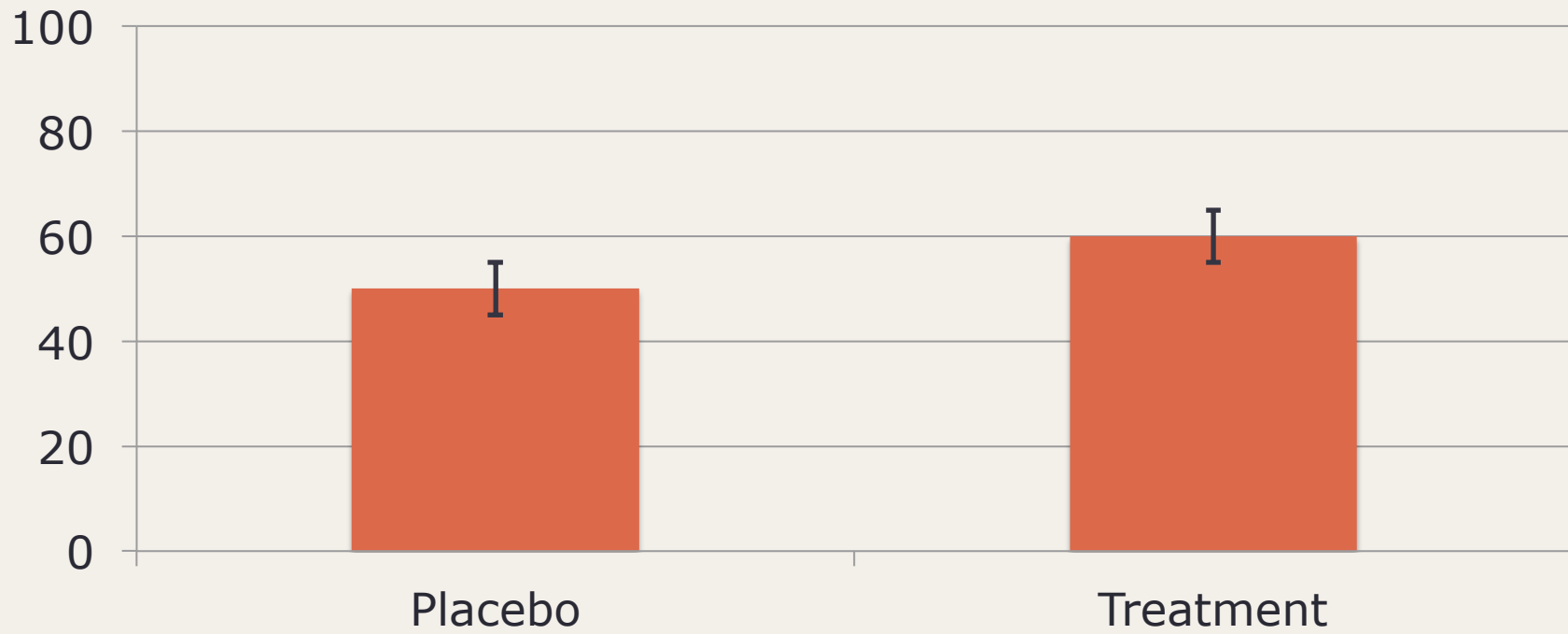


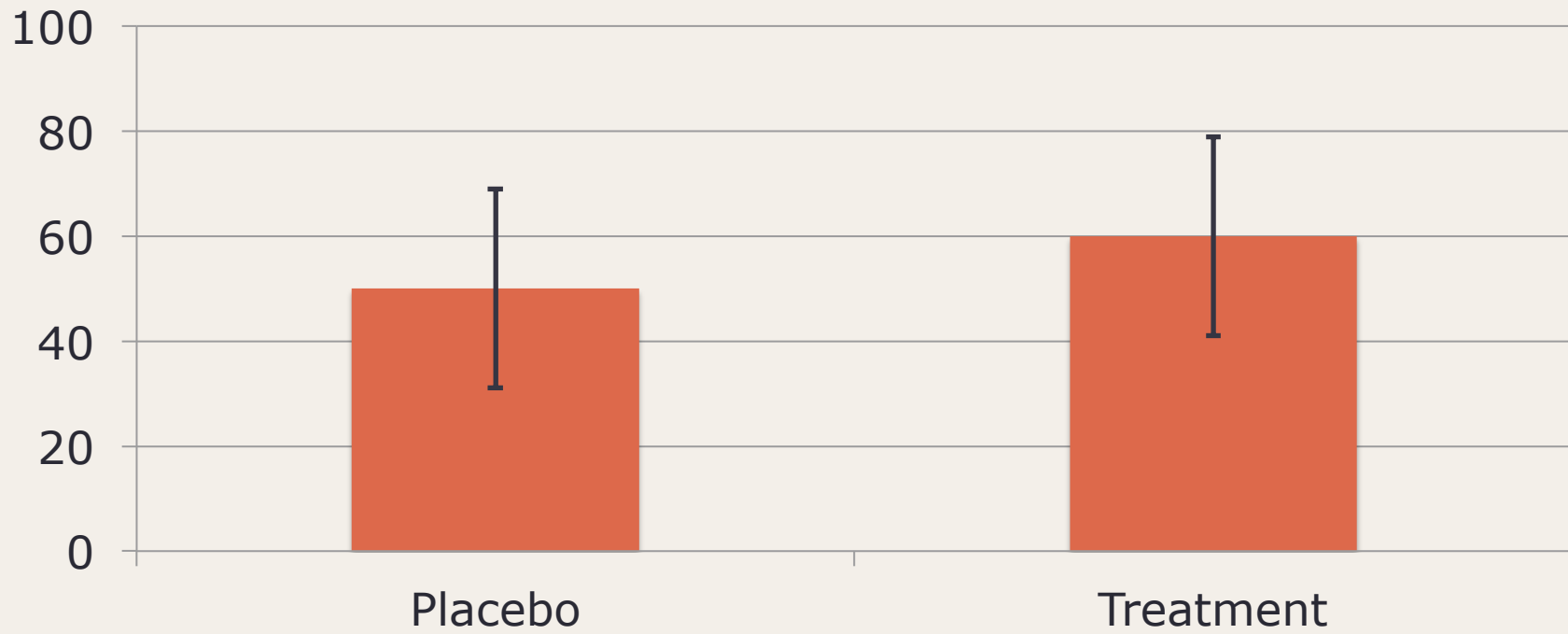
Don't Use Error Bars

They don't work as advertised

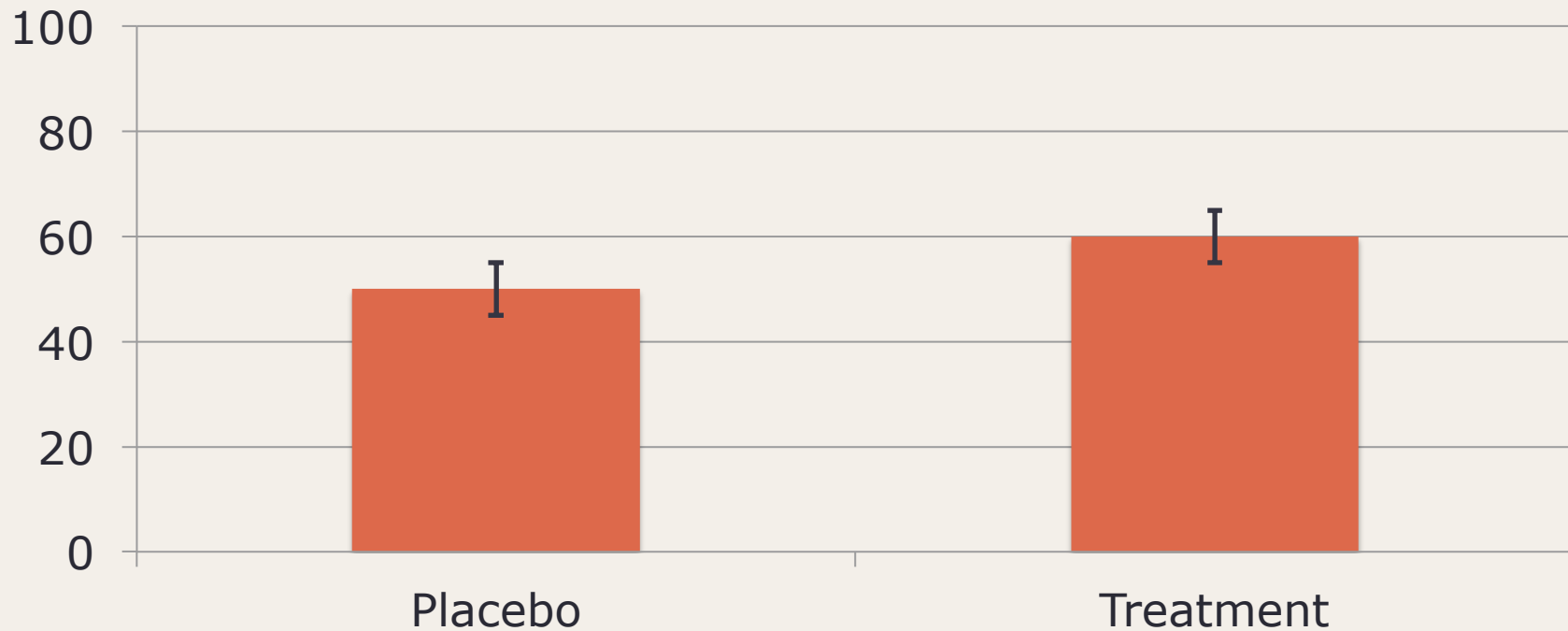


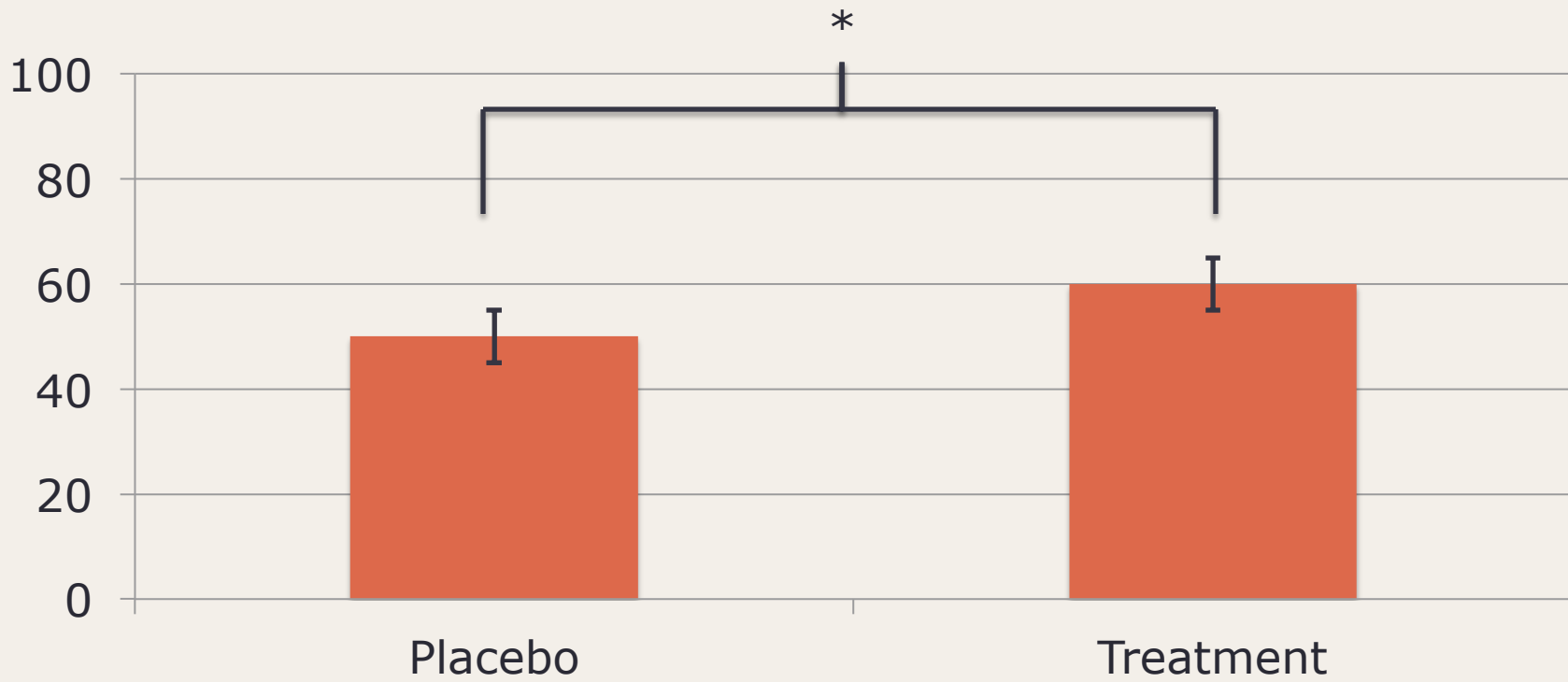






$p < 0.05$





Error Bars:

Are ambiguous

Are asymmetric

Are “all or nothing”

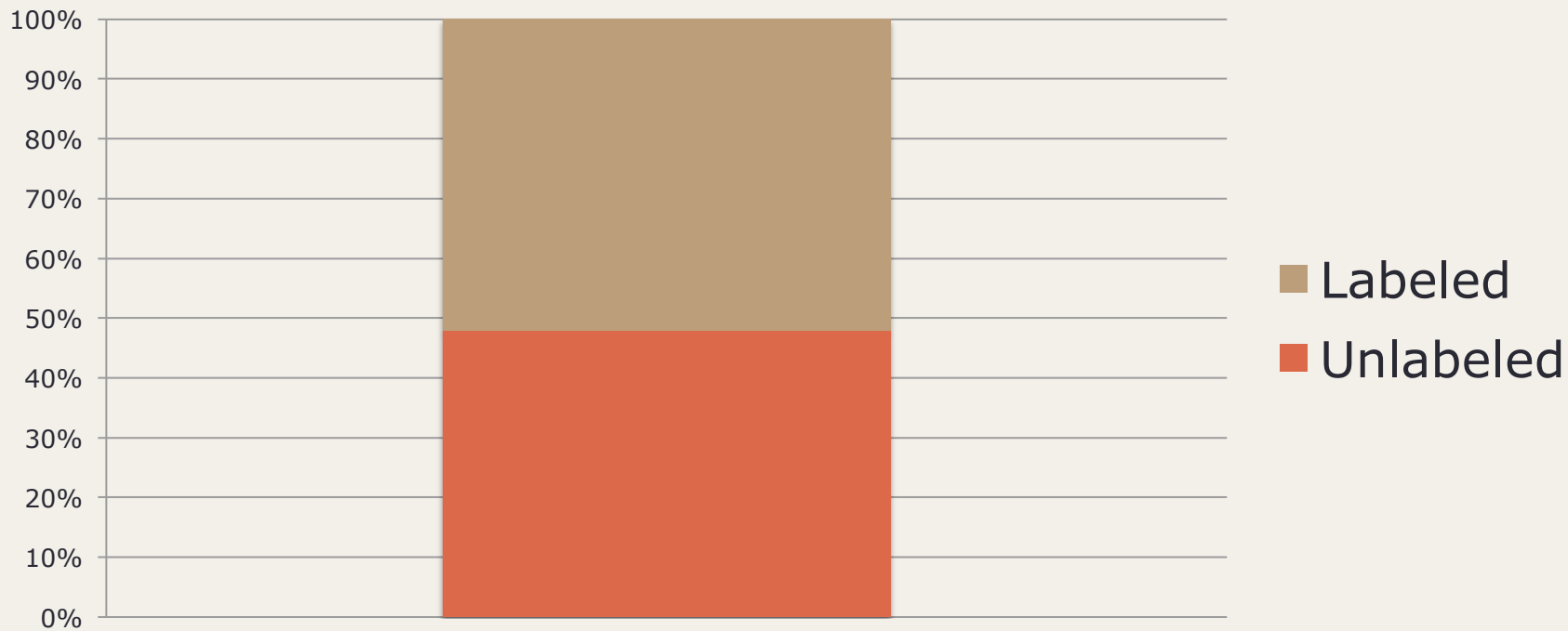
Error Bars:

Are ambiguous

Are asymmetric

Are “all or nothing”

InfoVis 2010-2013



InfoVis 2010-2013

Standard error

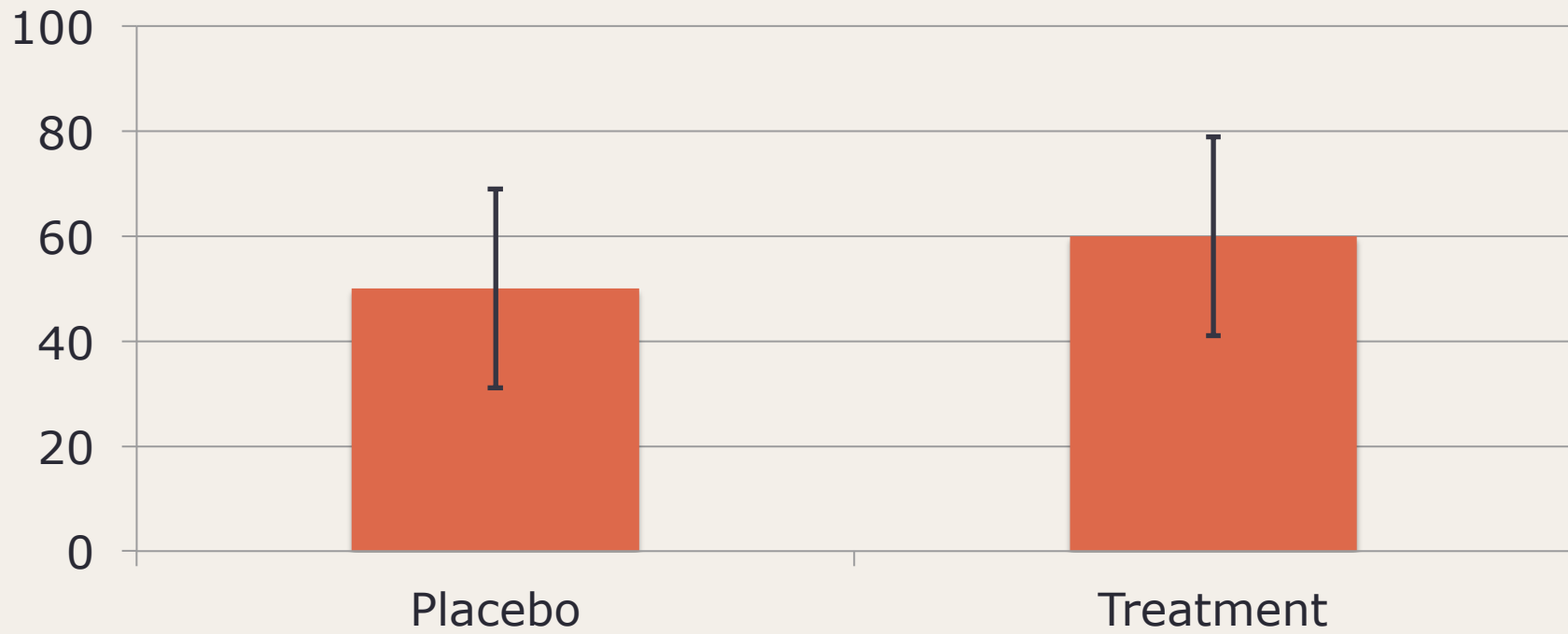
95% t confidence interval

Range

1.5 x interquartile range

Standard deviation

80% t confidence interval

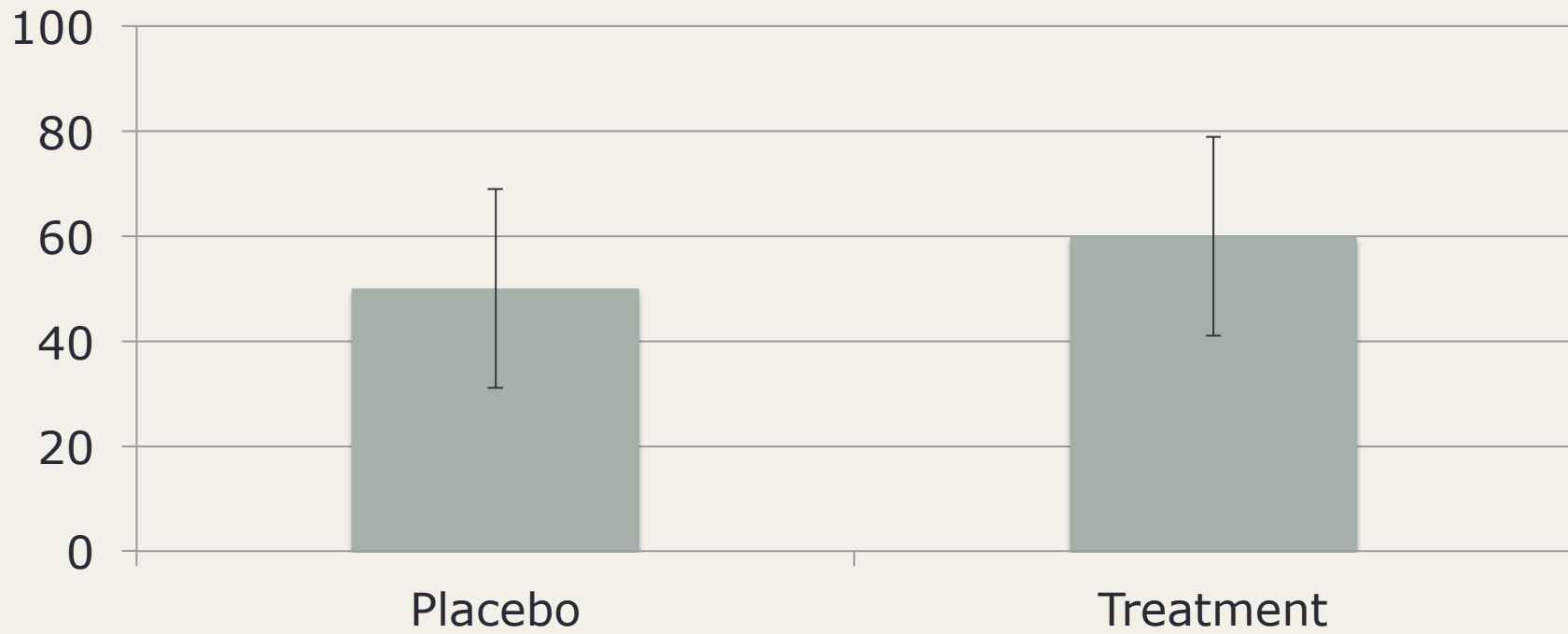


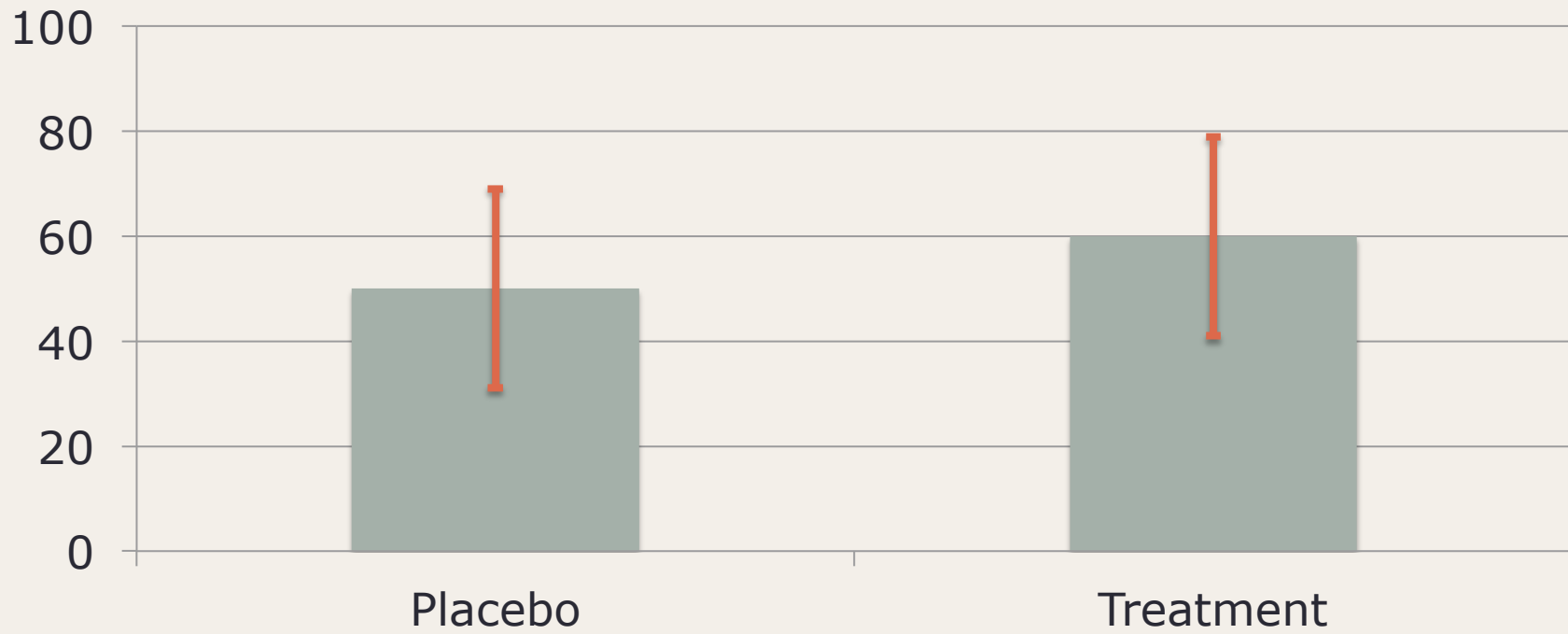
Error Bars:

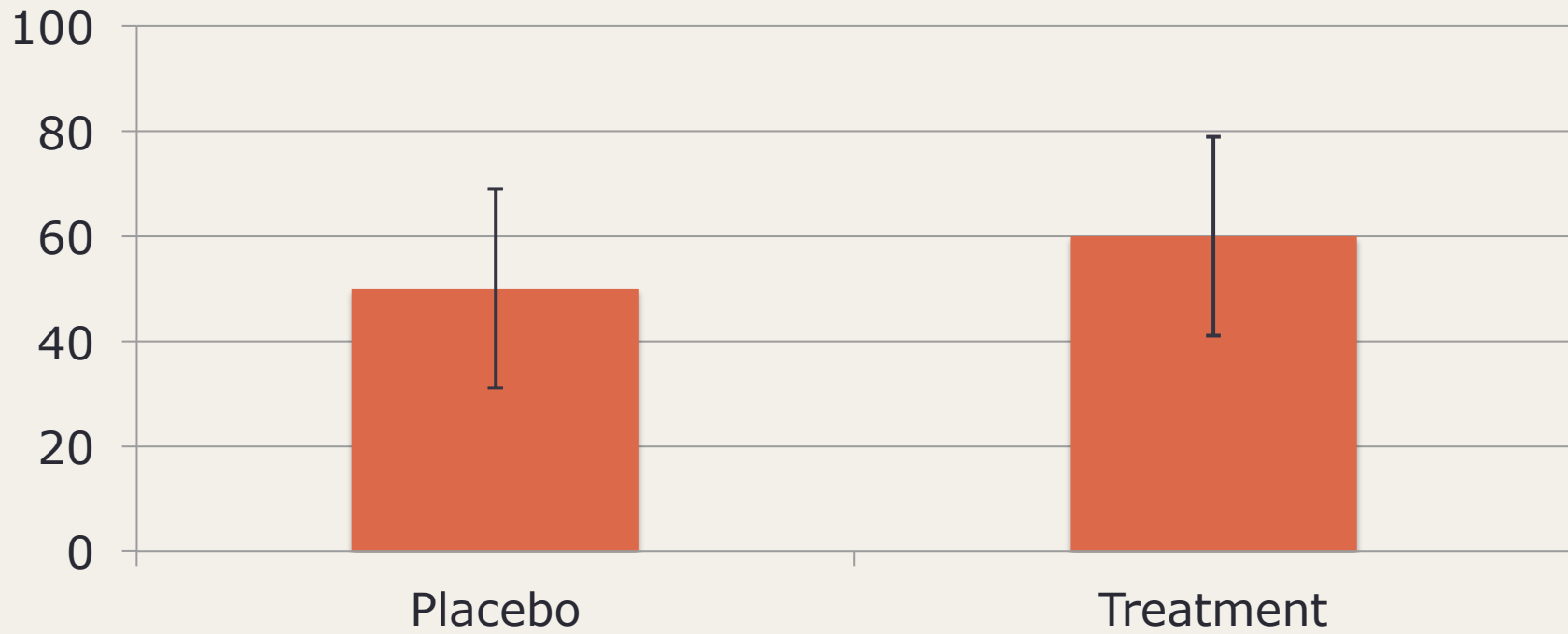
Are ambiguous

Are asymmetric

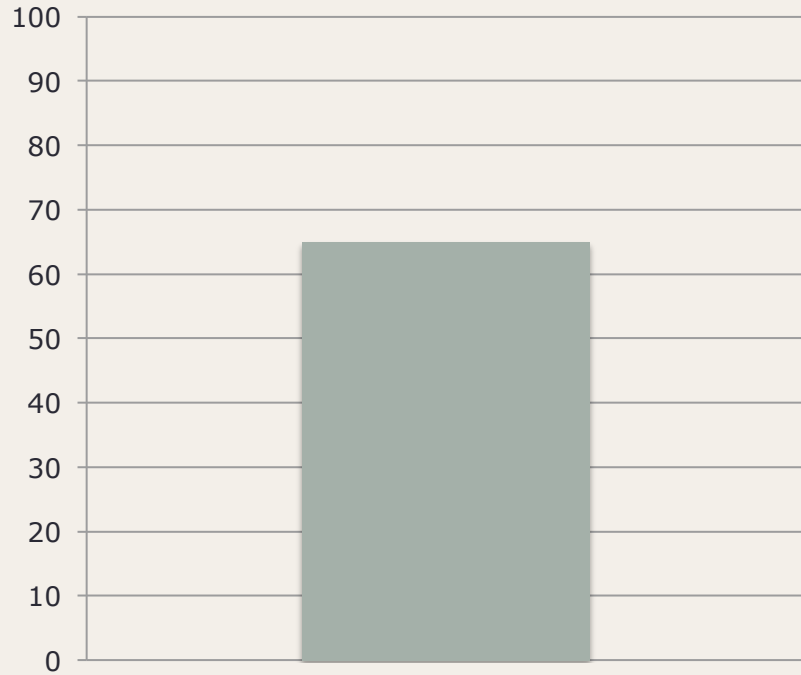
Are “all or nothing”





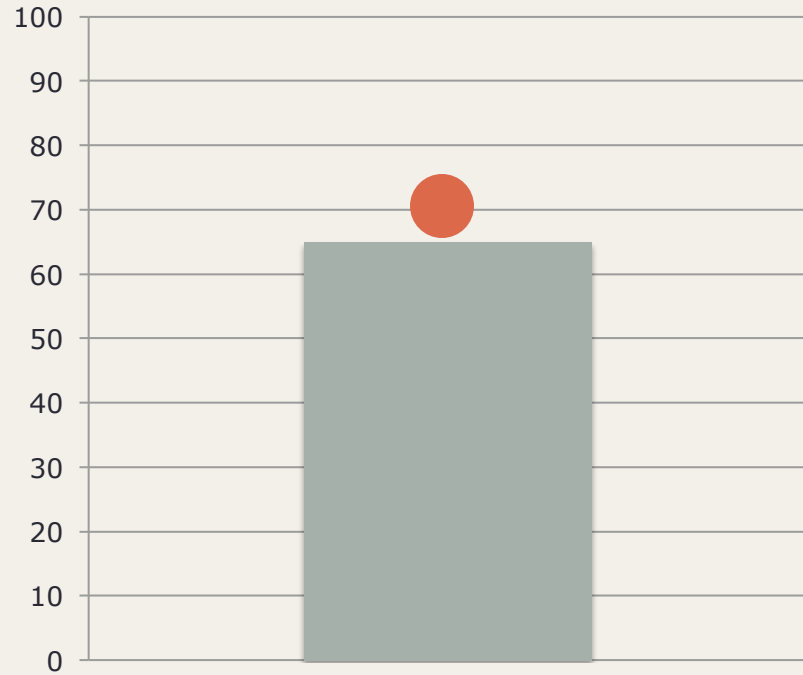
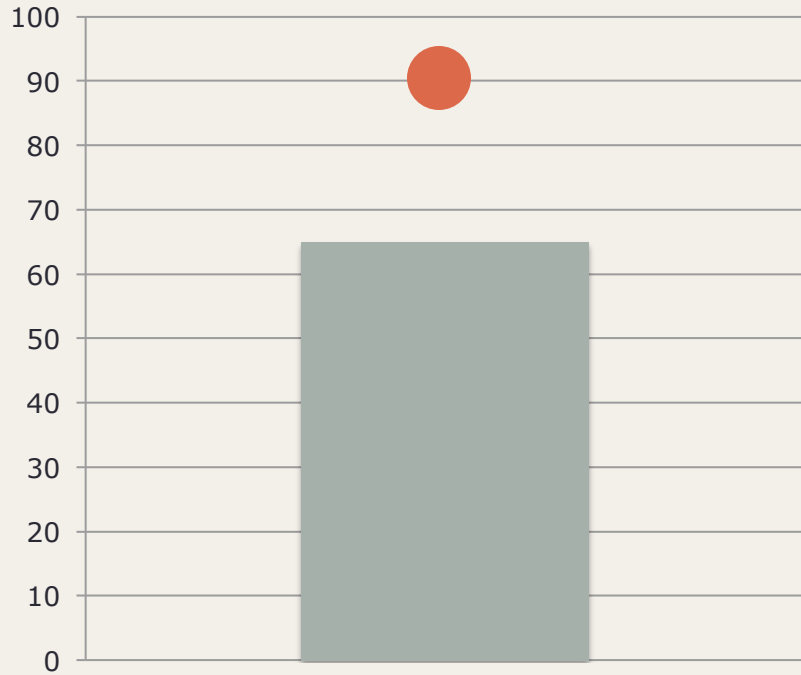


Within-the-bar bias

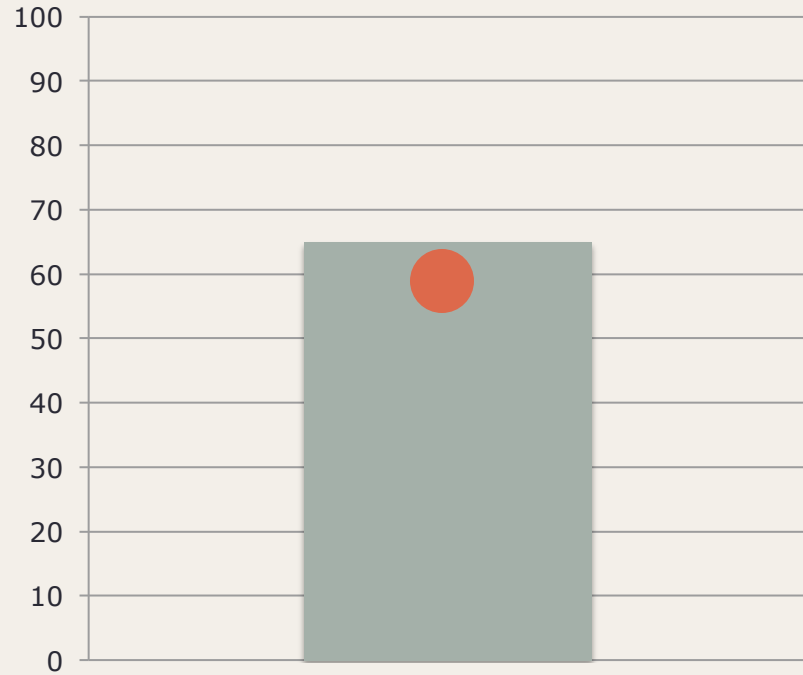
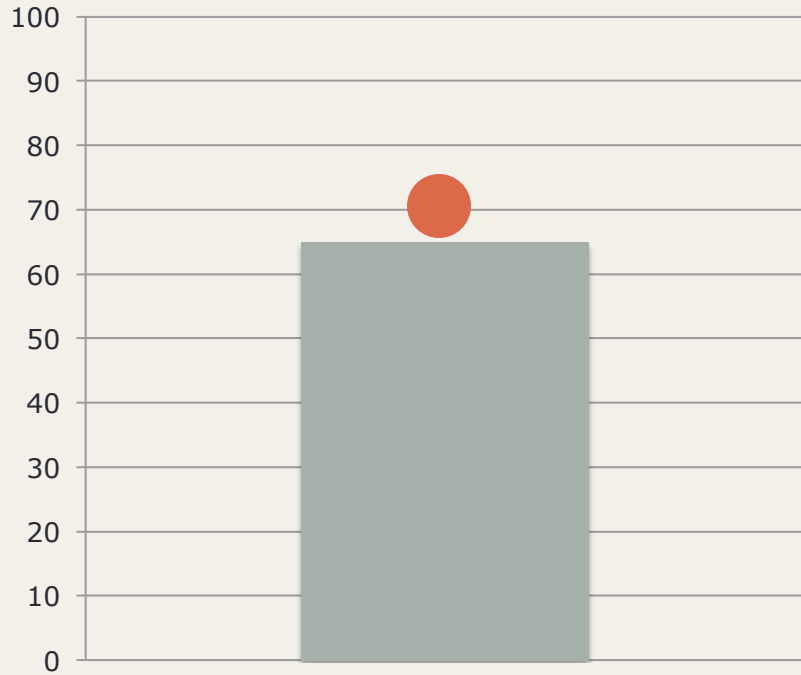


Newman, George E, and Brian J Scholl. "Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias." Psychonomic bulletin & review 19.4 (2012): 601-7.

Within-the-bar bias



Within-the-bar bias

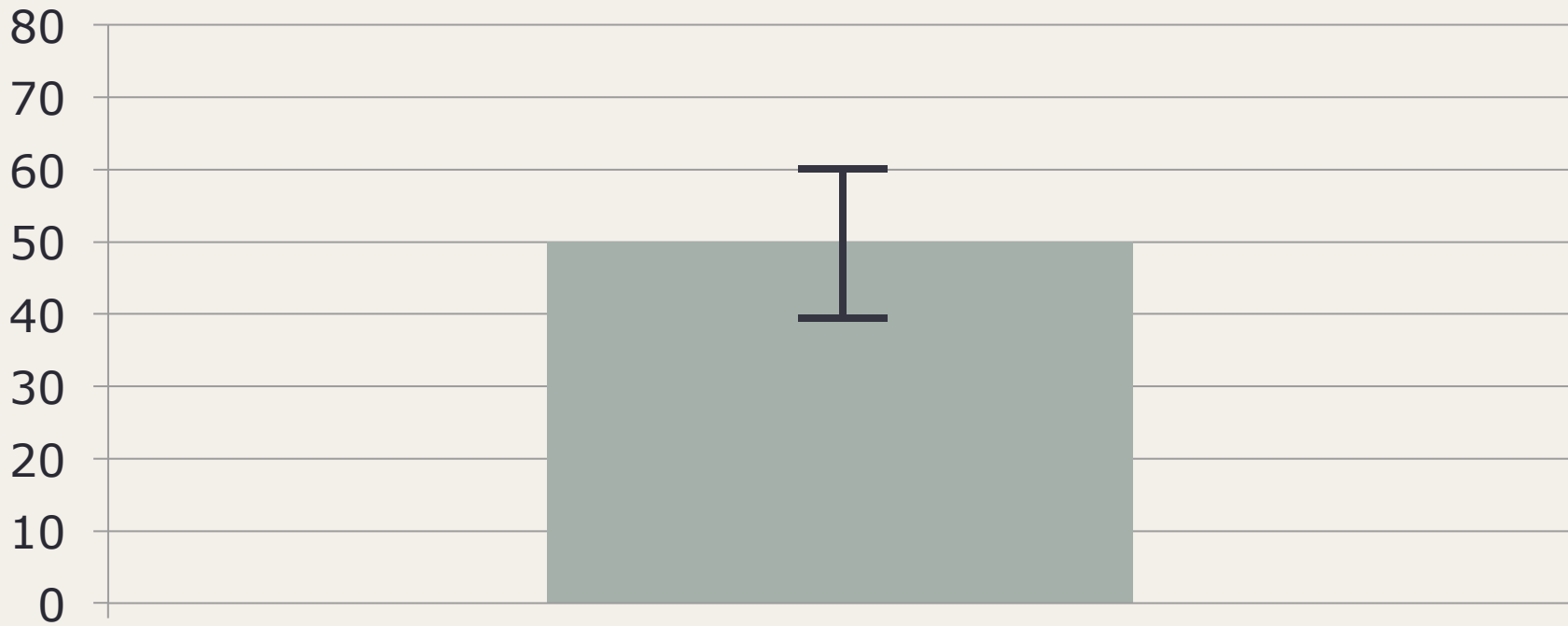


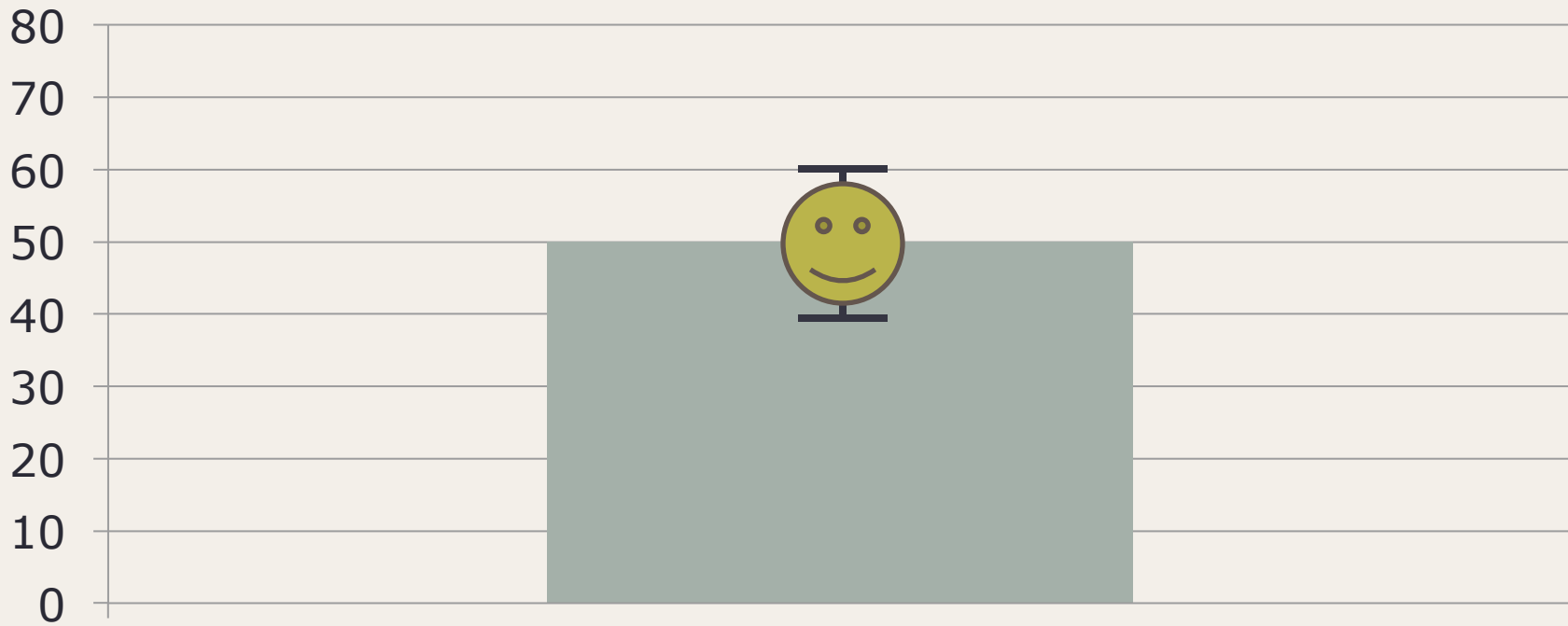
Error Bars:

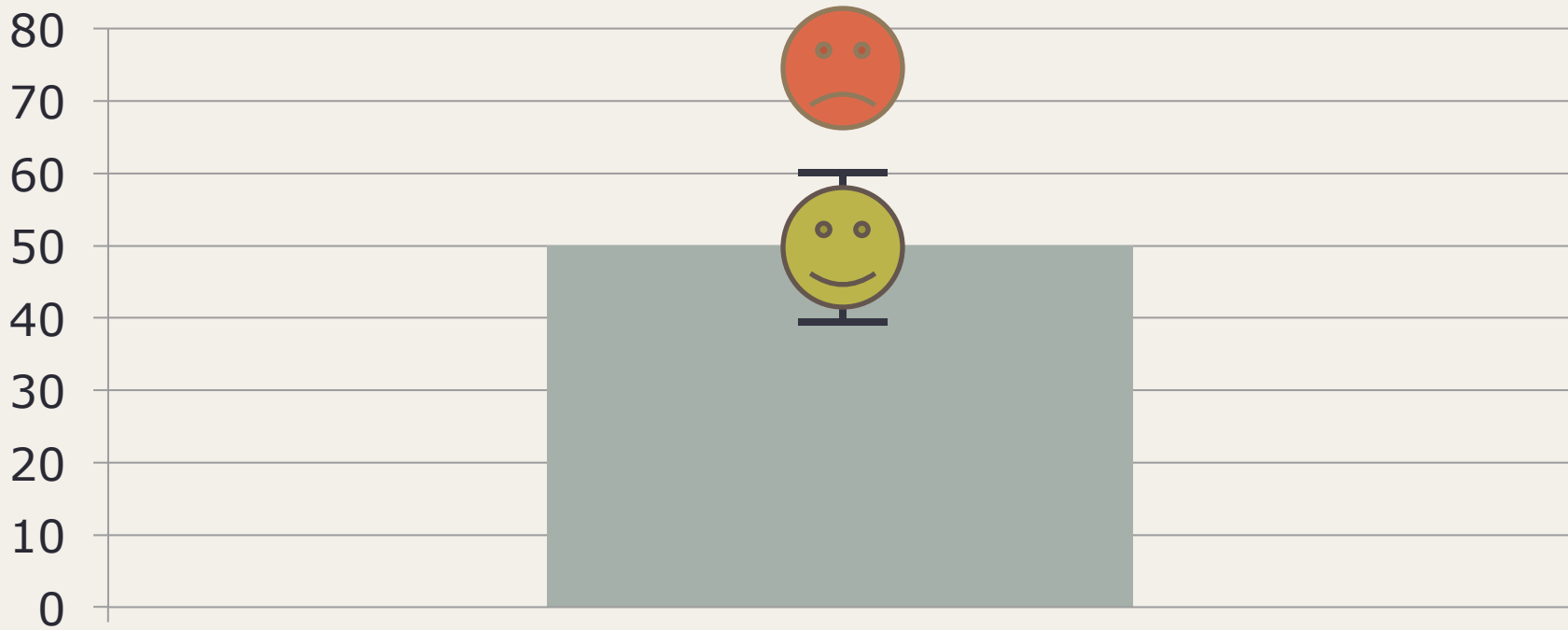
Are ambiguous

Are asymmetric

Are “all or nothing”





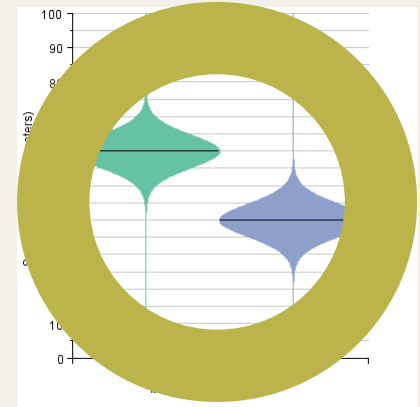
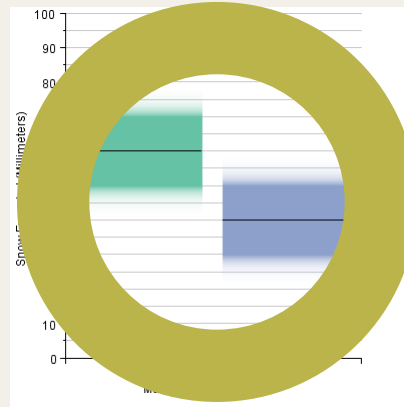


Don't Use Error Bars

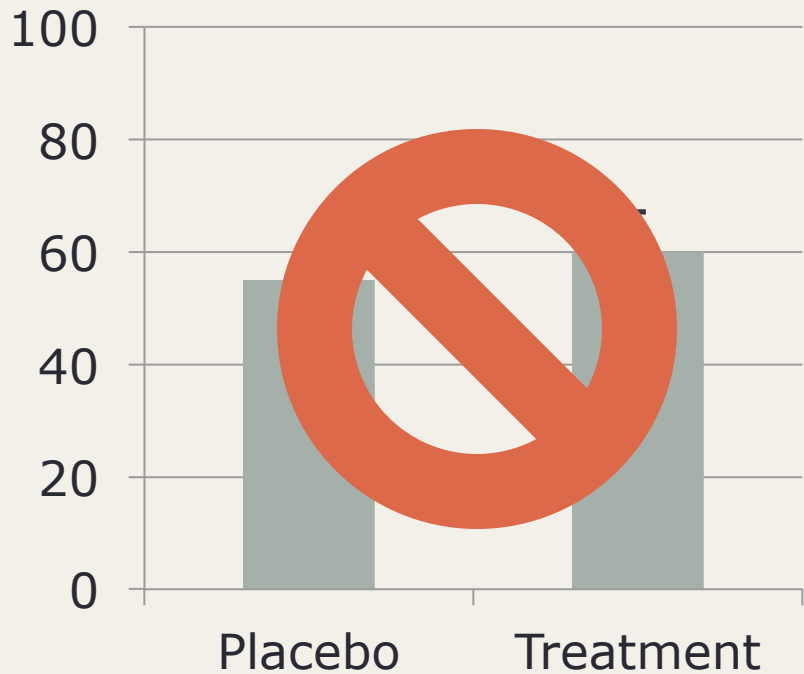
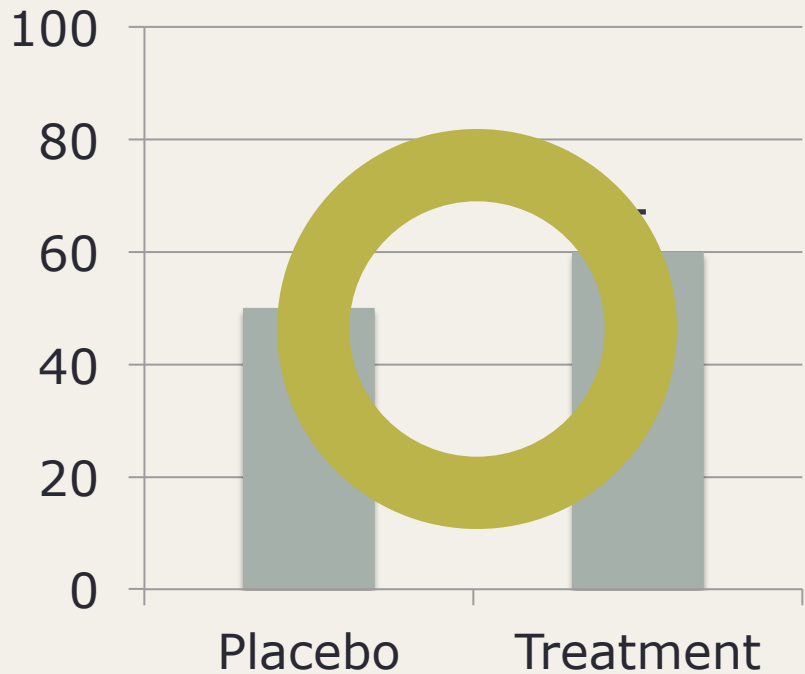
They don't work as advertised



Try something else instead!



A solution?

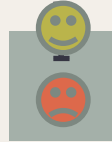


Design Requirements

Consistent

Symmetric

Continuous

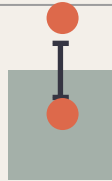


Design Requirements

Consistent

Symmetric

Continuous



Design Requirements

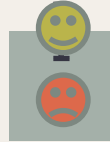
Consistent



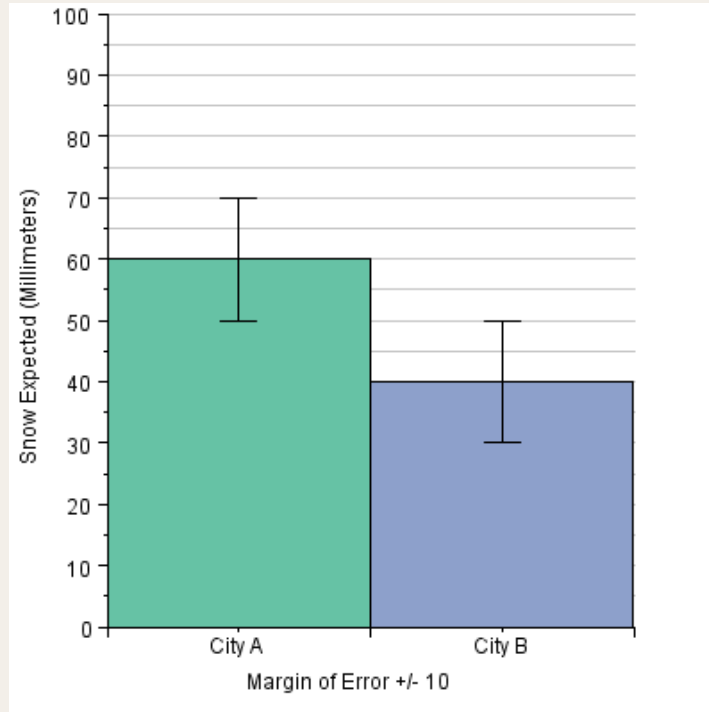
Symmetric



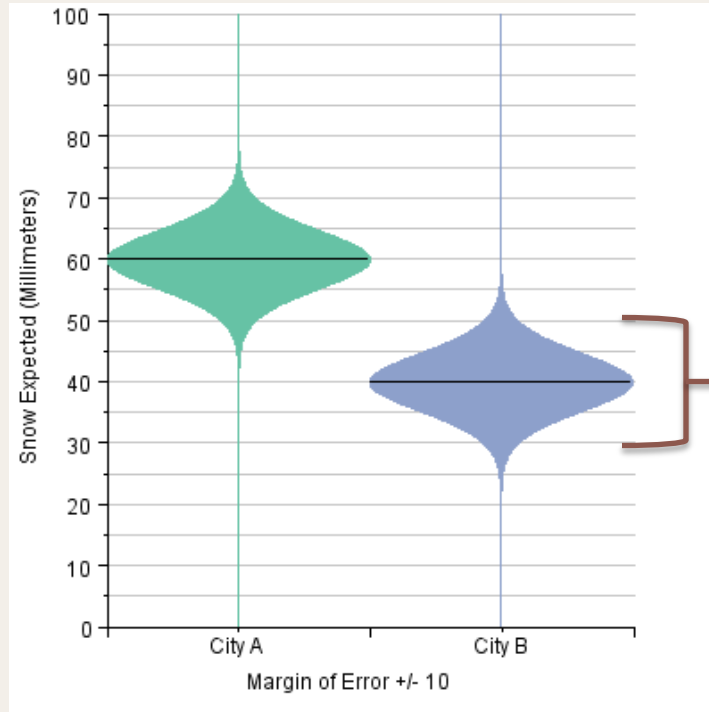
Continuous



Bar Chart



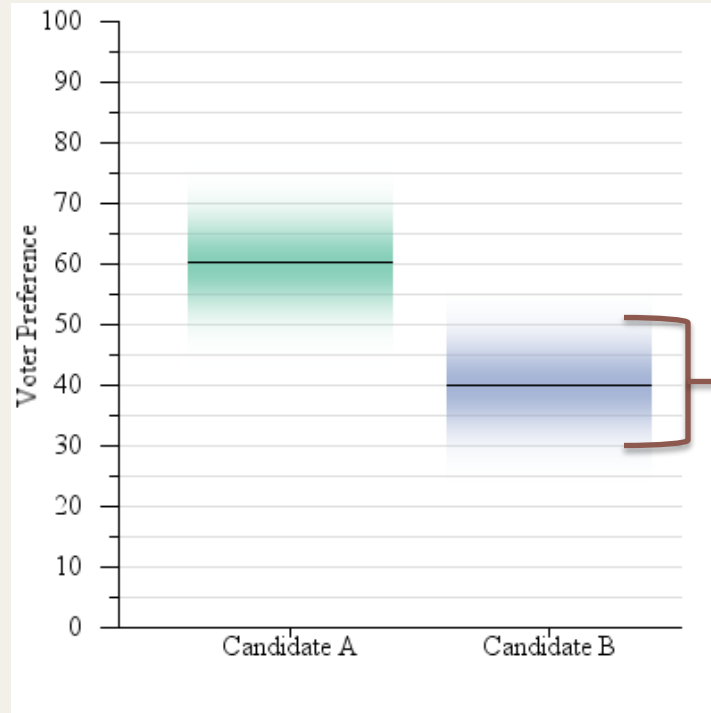
Violin Plot



95% t-confidence interval

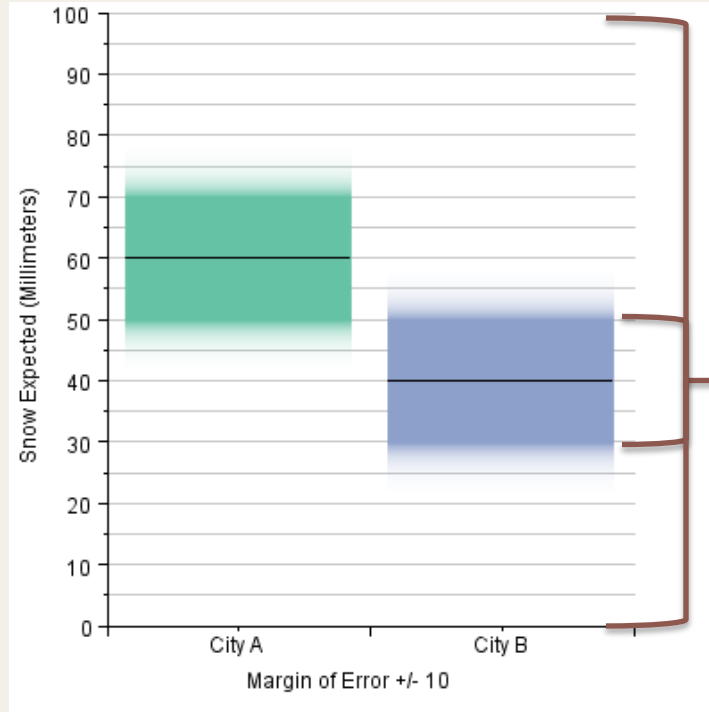
J. Hintze and R. Nelson. Violin plots: a box plot-density trace synergism. *The American Statistician*, 1998.

Gradient Plot?



95% t-confidence interval

Gradient Plot



"95% confidence interval"

Methods

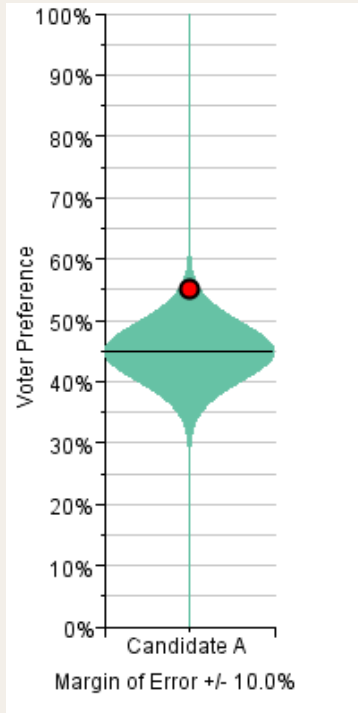
3 experiments on Amazon Mechanical Turk, 240 participants

3 problem frames (election polling, weather forecasting, financial modeling)

No prerequisite of statistical knowledge

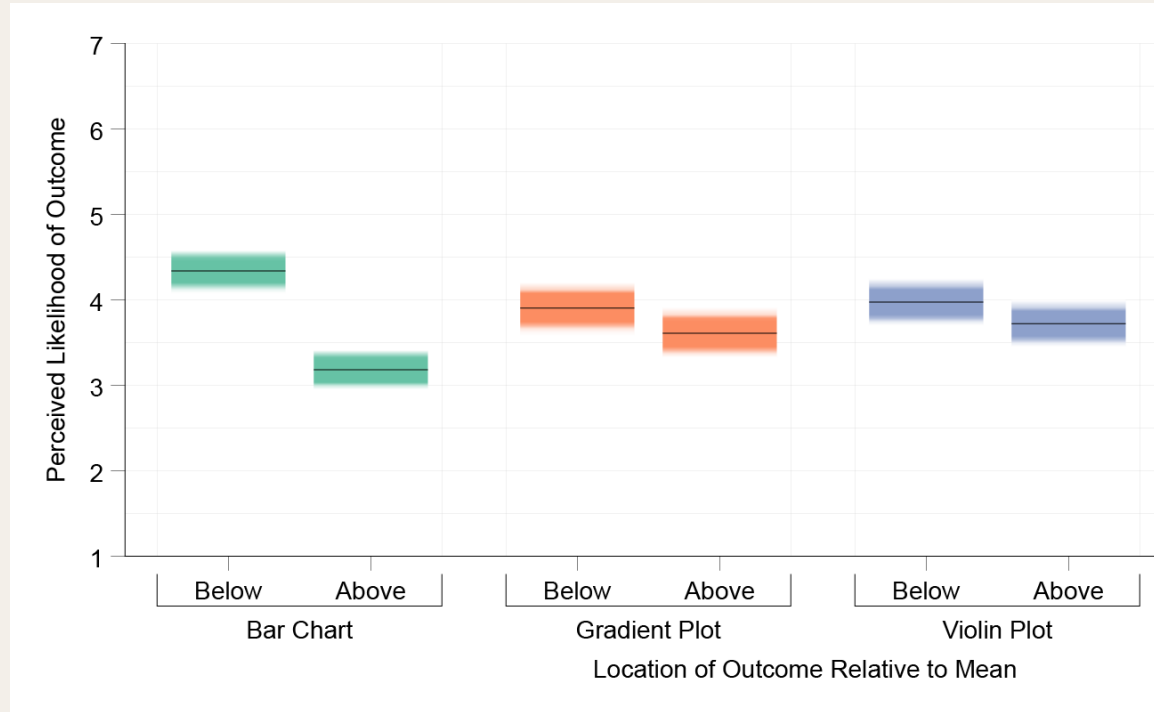
Participants gave a predictions as either binary forced choice, or on a Likert scale

One Sample Judgments



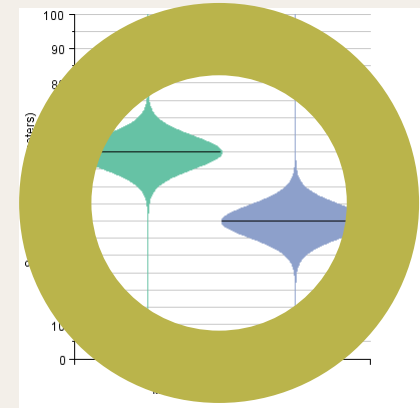
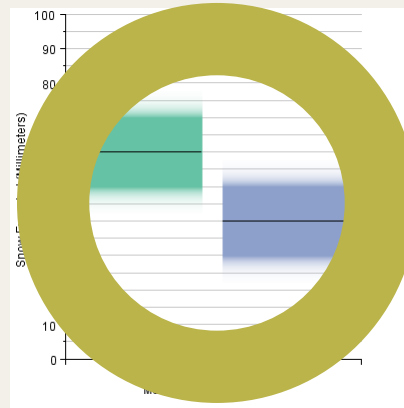
How likely (or how surprising) do you think the red potential outcome is, given the poll?

Results

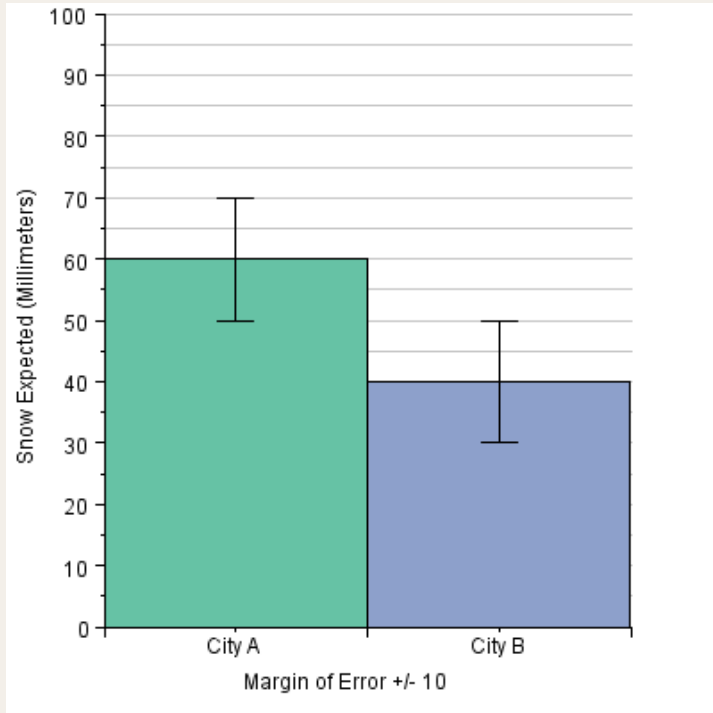


“Within the bar” bias

Error bars suffer from this bias... but other encodings don't



Two Sample Judgments

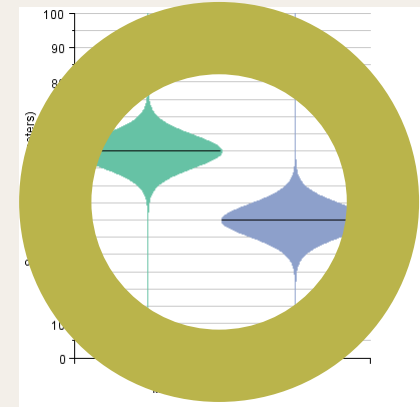
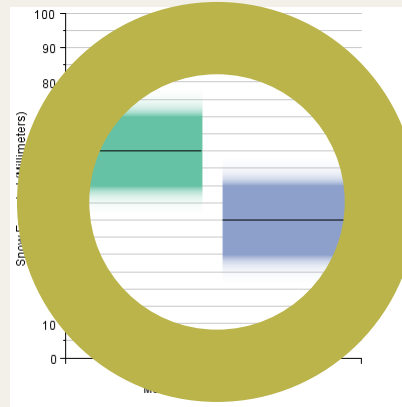


If forced to guess, which city do you predict will get more snow?

Overconfidence

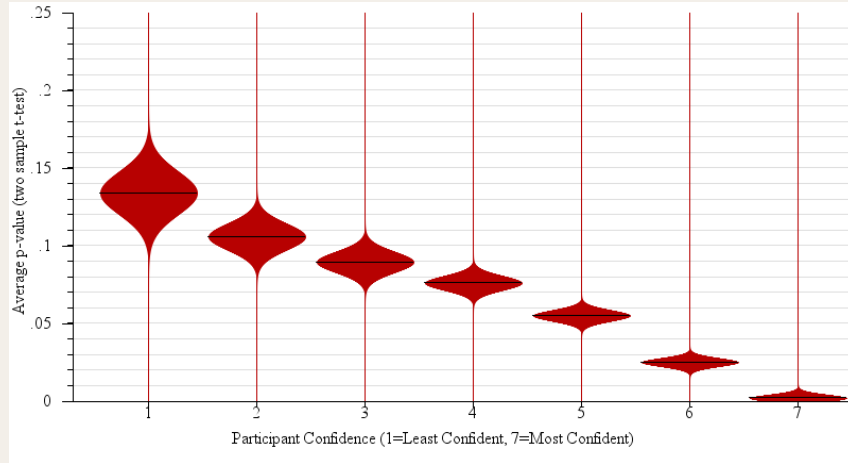
Error bars make people unjustifiably confident...

but other encodings don't

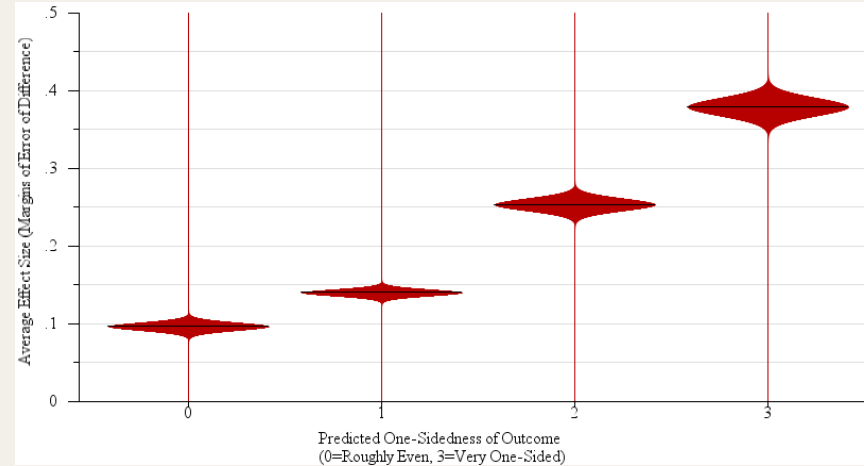


Costs are low

p-value



Effect size

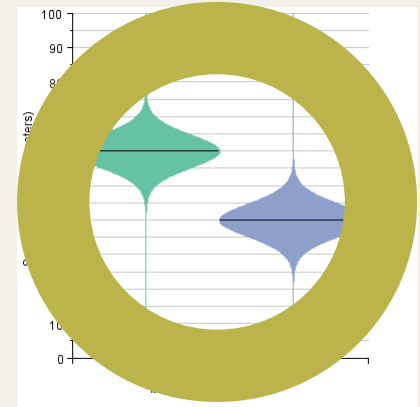
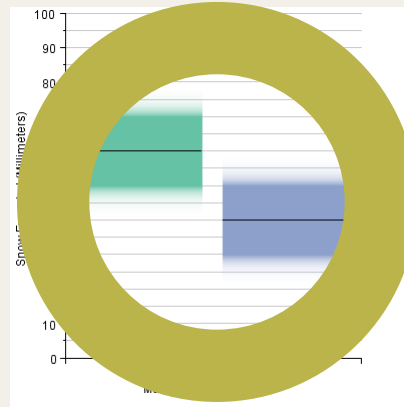


Don't Use Error Bars

They don't work as advertised



Try something else instead!



What's next?

More encodings

More testing

Real stakes

Make your own!

<http://graphics.cs.wisc.edu/Vis/ErrorBars/>

Acknowledgments

This work was supported in part by NSF award IIS-1162037, NIH award R01 AI077376, and ERC Advanced Grant "Expressive."

Thanks to Wei-Chen Chen for web generation code.

Visit:

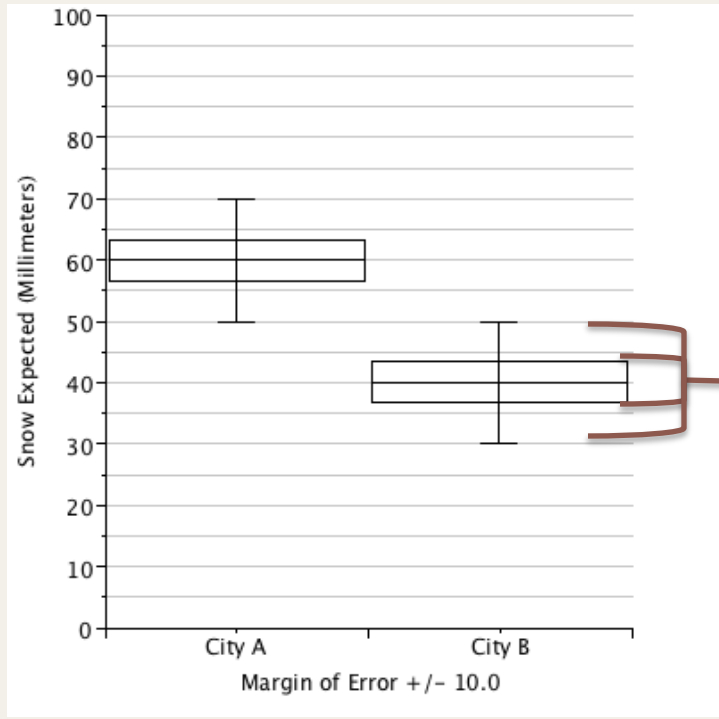
<http://graphics.cs.wisc.edu/Vis/ErrorBars/>

to make your own plots!

(and for data tables, stimuli, and sample experiments).

Contact: mcorrell@cs.wisc.edu

Box Plot



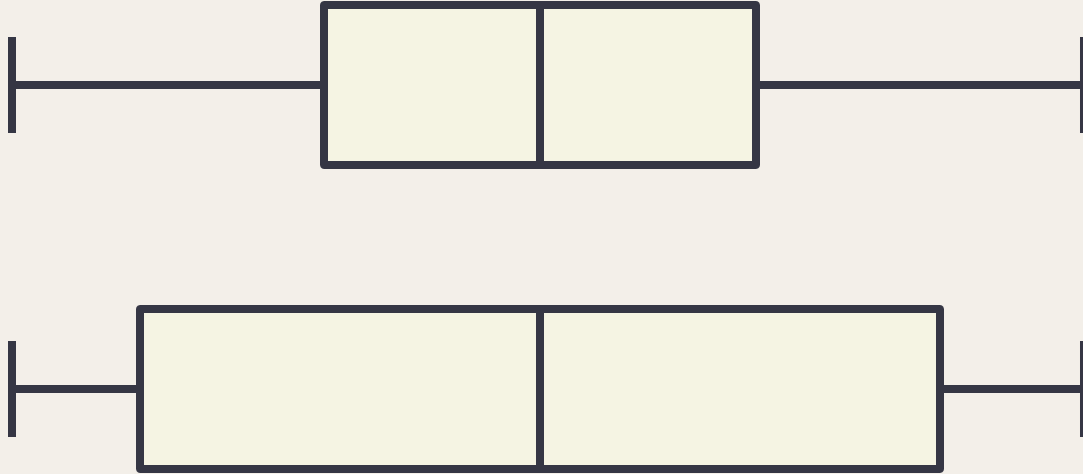
95% confidence interval

Müller-Lyer Illusion



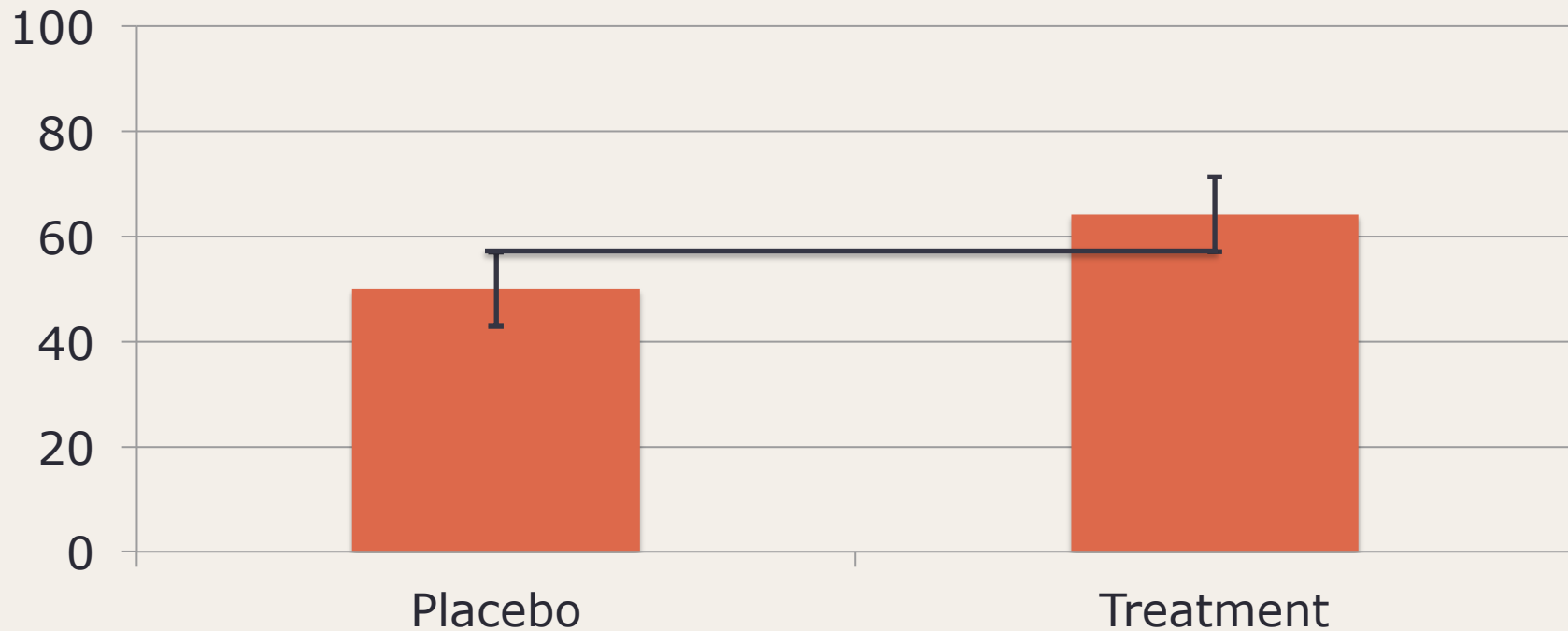
W. Stock and J. Behrens. Box, Line, and Midgap plots: Effects of Display Characteristics on the Accuracy and Bias of Estimates of Whiskey Length. *Journal of Educational and Behavioral Statistics*, 1991

Müller-Lyer Illusion

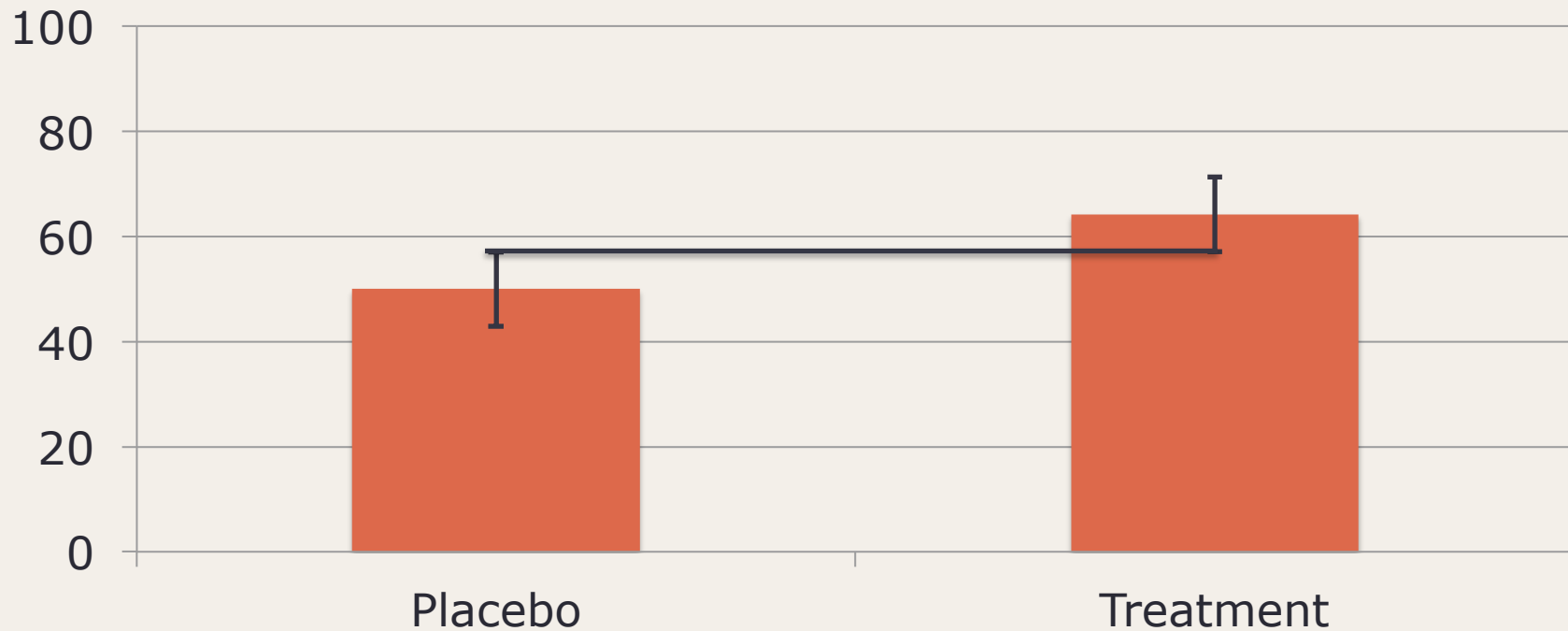


W. Stock and J. Behrens. Box, Line, and Midgap plots: Effects of Display Characteristics on the Accuracy and Bias of Estimates of Whiskey Length. *Journal of Educational and Behavioral Statistics*, 1991

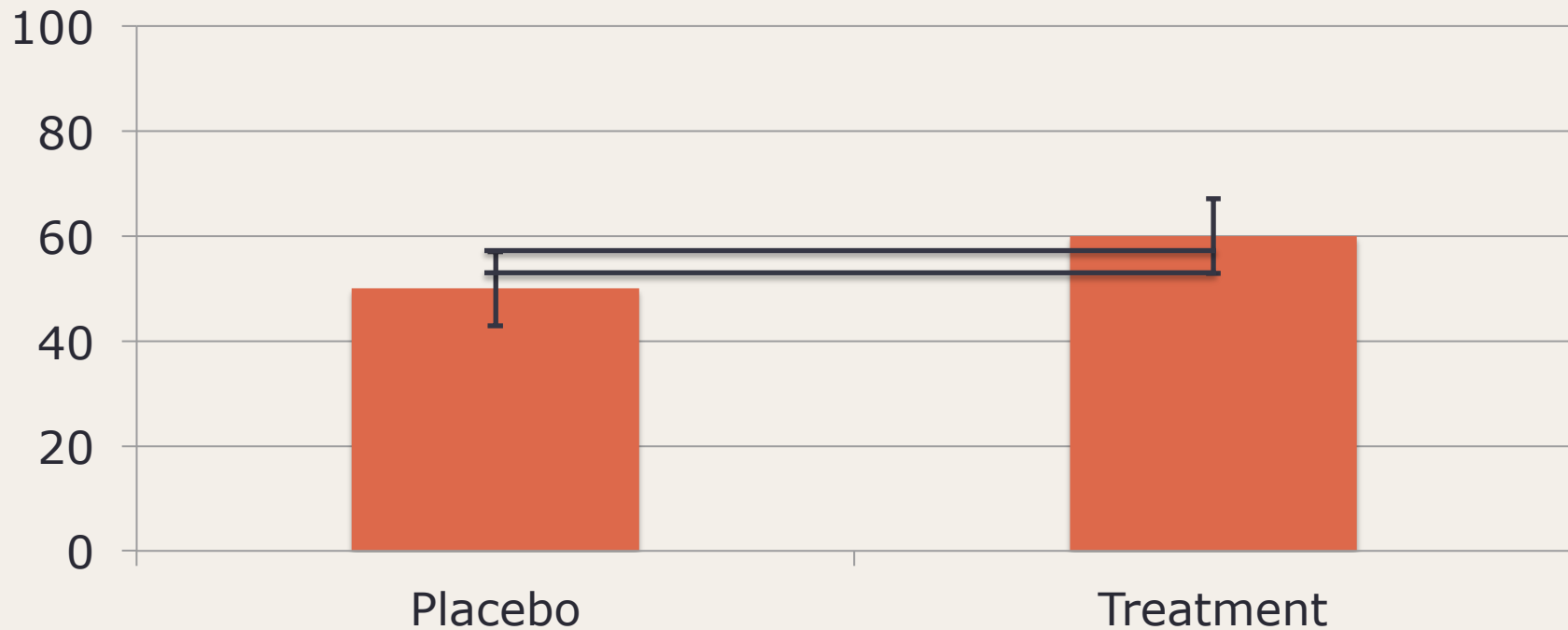
$p = .05?$



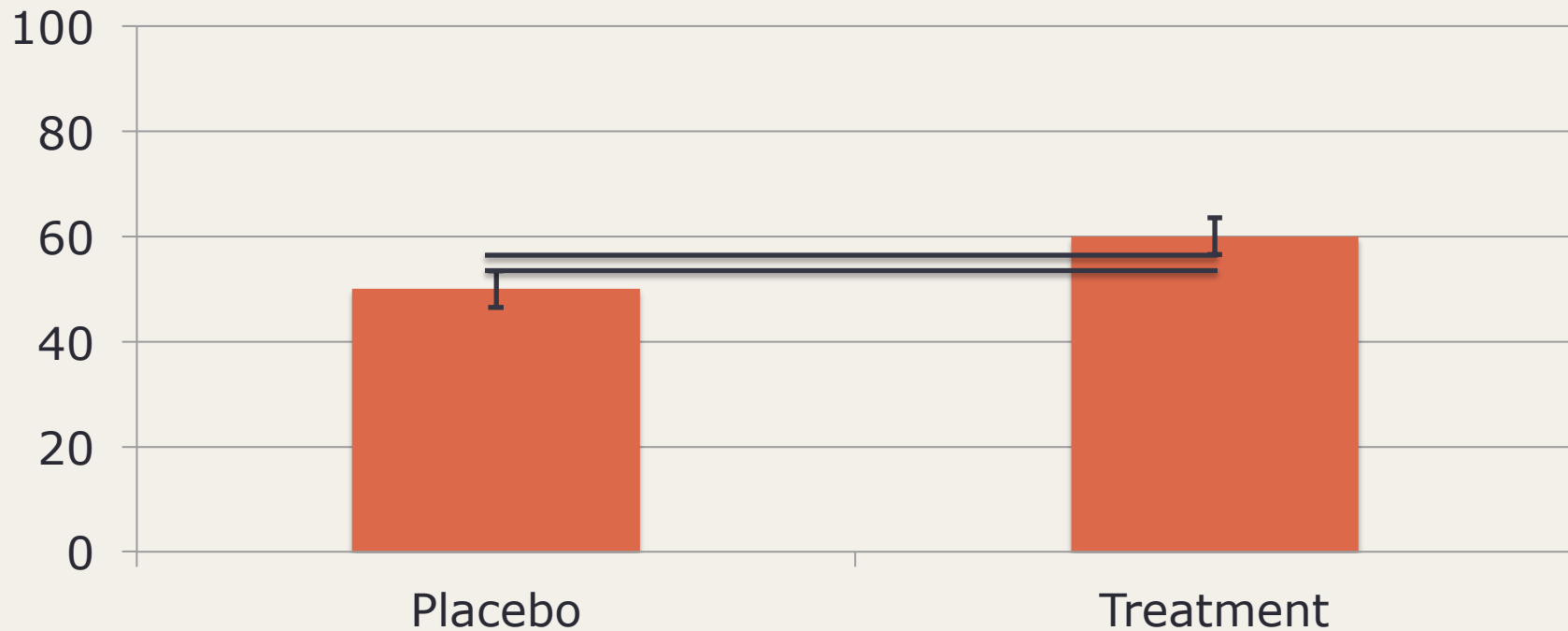
$p < 0.01$



$p = .05$



$p = .05$



“p-pdf”

