

Regression by Eye

Replication of “Regression by Eye: Estimating Trends in Bivariate Visualizations” by
Michael Correll and Jeffrey Heer

Motivation & research question



Understand how different data visualizations bias the participants' estimations



How do the type of chart, the bandwidth of the residuals and trend type affect the estimated trend of the bivariate visualization by an individual?

Hypotheses



H1 - No bias in estimations from under- or overshooting

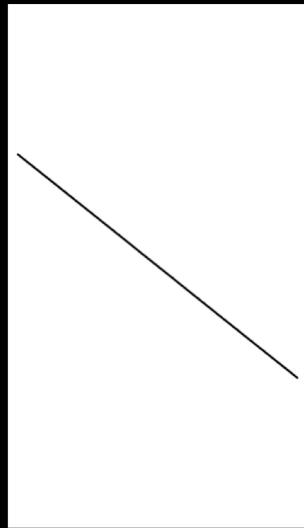


H2 - Larger residuals will result in less accuracy



H3 - The type of trend will have no statistically significant effect

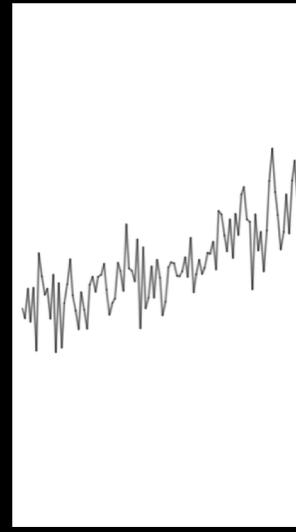
Experimental design



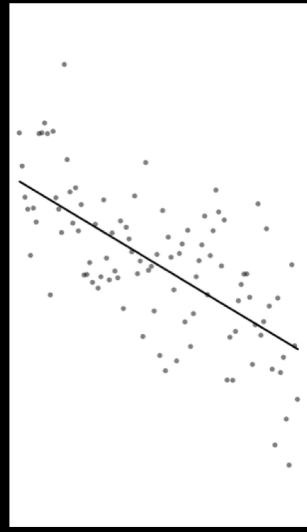
8 slopes



4 bandwidths of
residuals



3 chart types



3 trend types

Experimental design

96 combinations

(incl. 32 randomly assigned trend types)

4 validation tasks (100 overall)

Suitable for multiple of 8 participants

16 participants → 100% effect coverage

The screenshot shows a software interface for experimental design, specifically for a 'Latin square' experiment. The title bar reads 'Regression by Eye - Experiment 1'. The interface is organized into several sections:

- CT Chart Type:** A stack of three items: 'Scatter plot' (selected), 'Line chart', and 'Area chart'.
- Random of 1 replication(s) not serial:** A dropdown menu indicating the current settings.
- BR Bandwidth of Residuals:** A stack of four items: 0.05, 0.1, 0.15 (selected), and 0.2.
- Latin square of 1 replication(s) not serial:** A dropdown menu indicating the current settings.
- S Slope:** A stack of nine items: -0.8, -0.4, -0.2, -0.1, 0.1, 0.2, 0.4, and 0.8 (selected).
- Latin square of 1 replication(s) not serial:** A dropdown menu indicating the current settings.

Participants



Age:
20-25: 11
26-29: 4
30+: 1



Gender:
Female: 8
Male: 8



Nationality:
Swiss: 13
Other: 3



Faculty:
PF: 7
OEC: 9



Level of education:
BA: 10
MA: 5
PhD: 1



Graphs once a month:
Yes: 11
No: 5



Read paper:
Yes: 0
No: 16



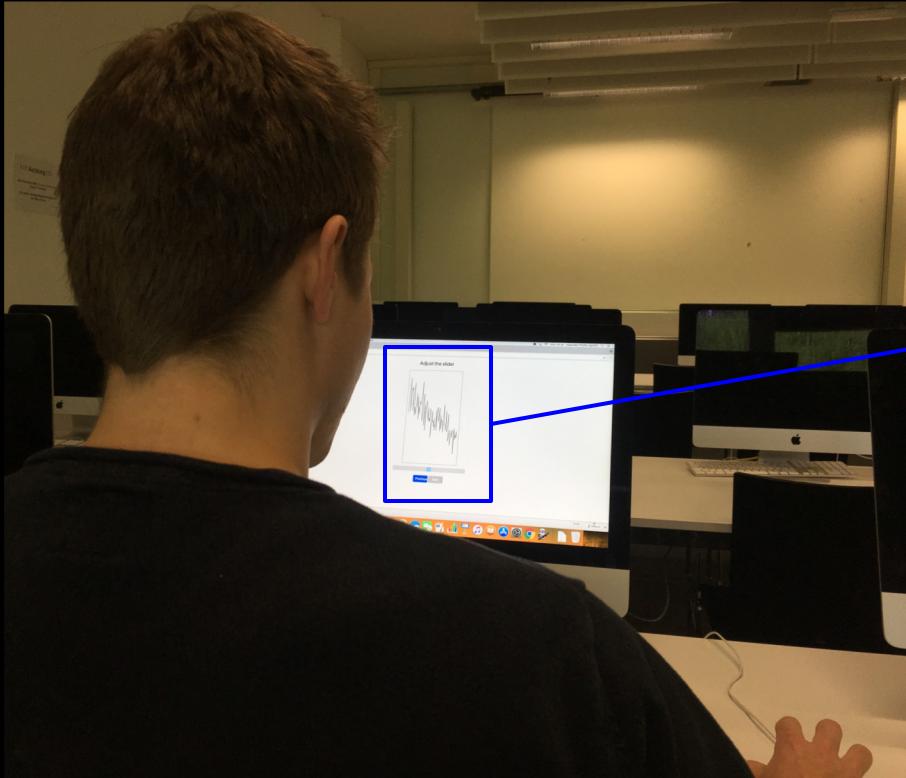
Visual impairments:
Yes: 0
No: 16



Computers once a week:
Yes: 16
No: 0

Adjust the slider

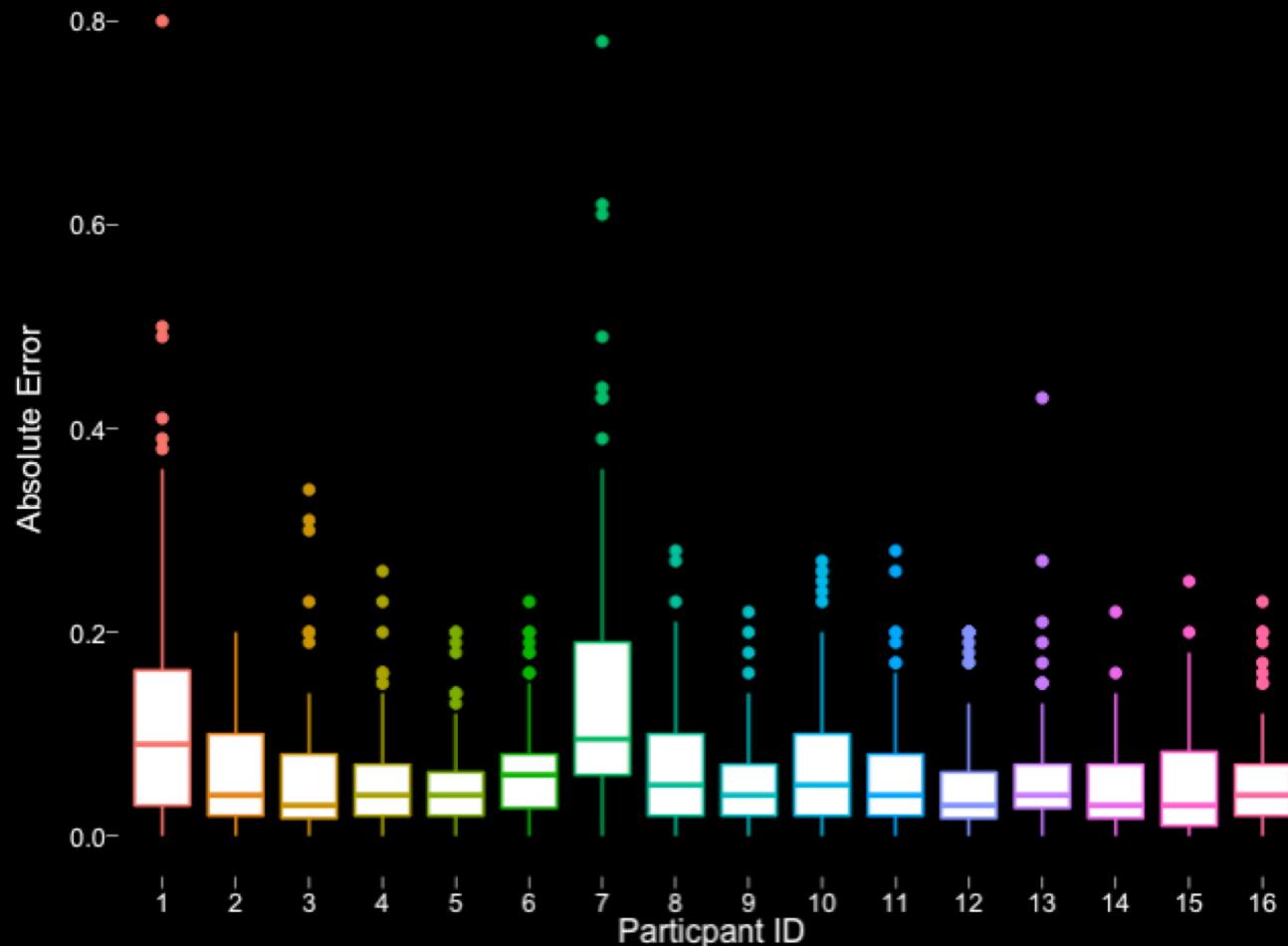
Software and tasks



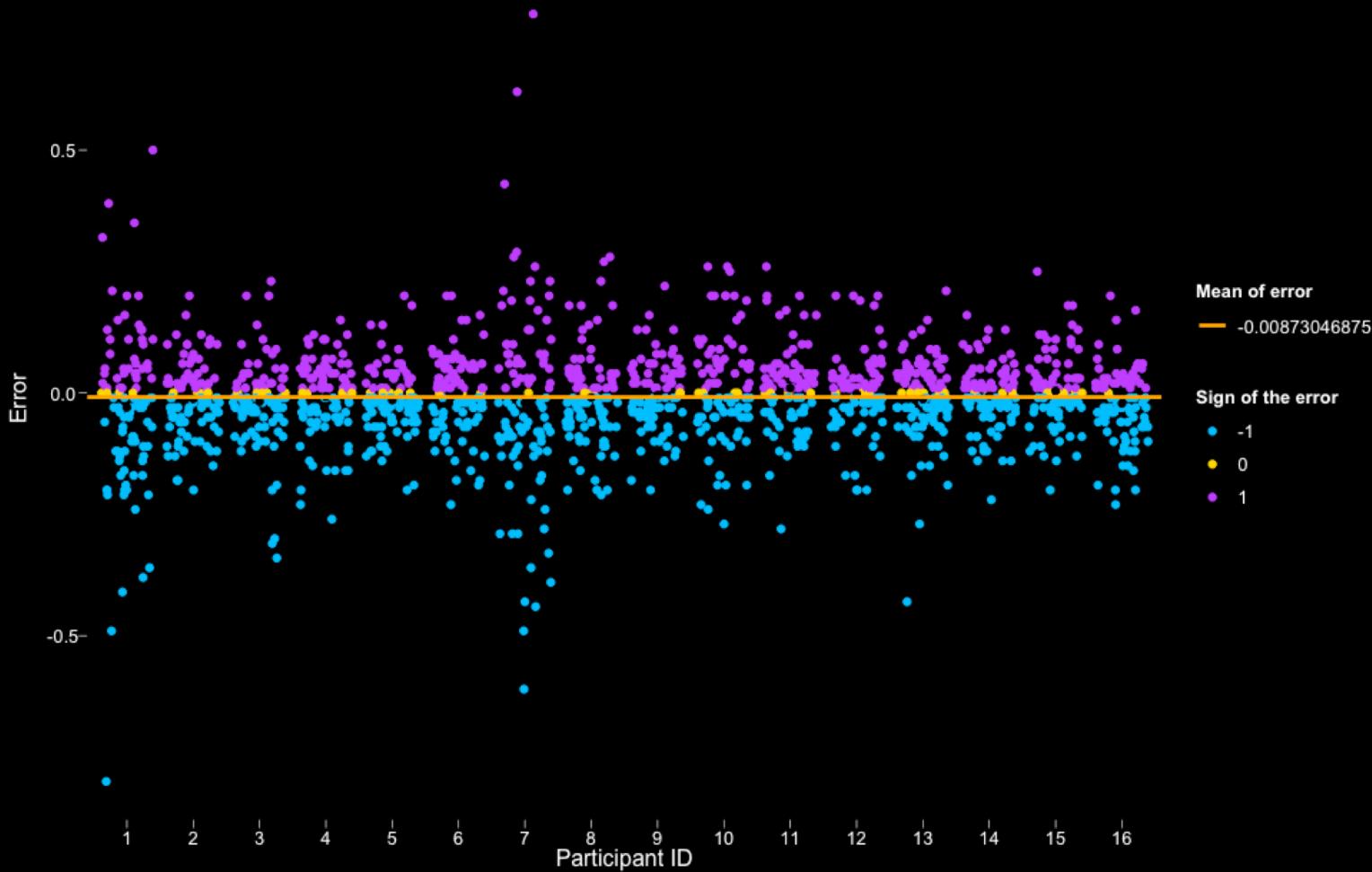
Setup



Errors per participant: within subject variation



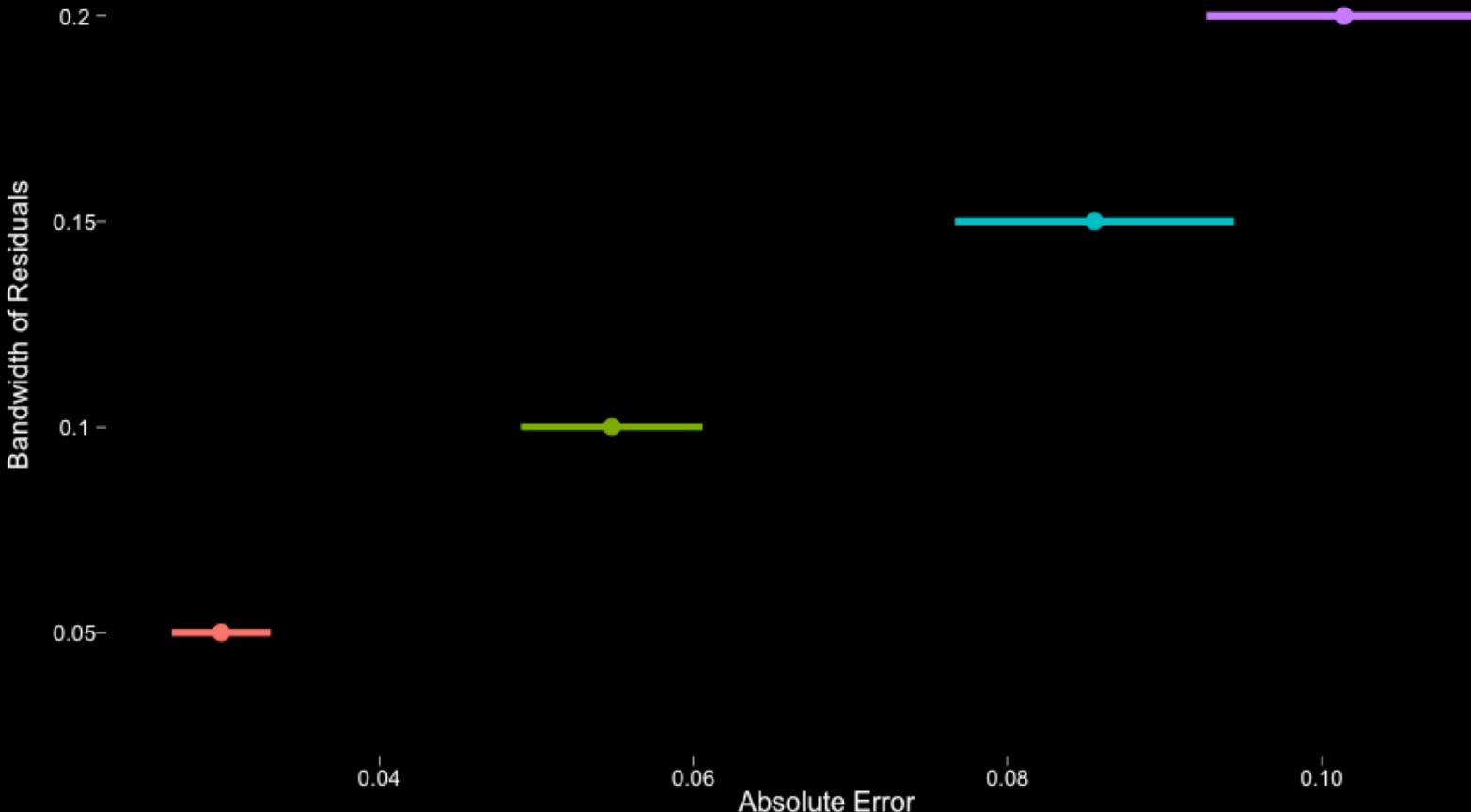
Hypothesis 1 - no over- or underestimation



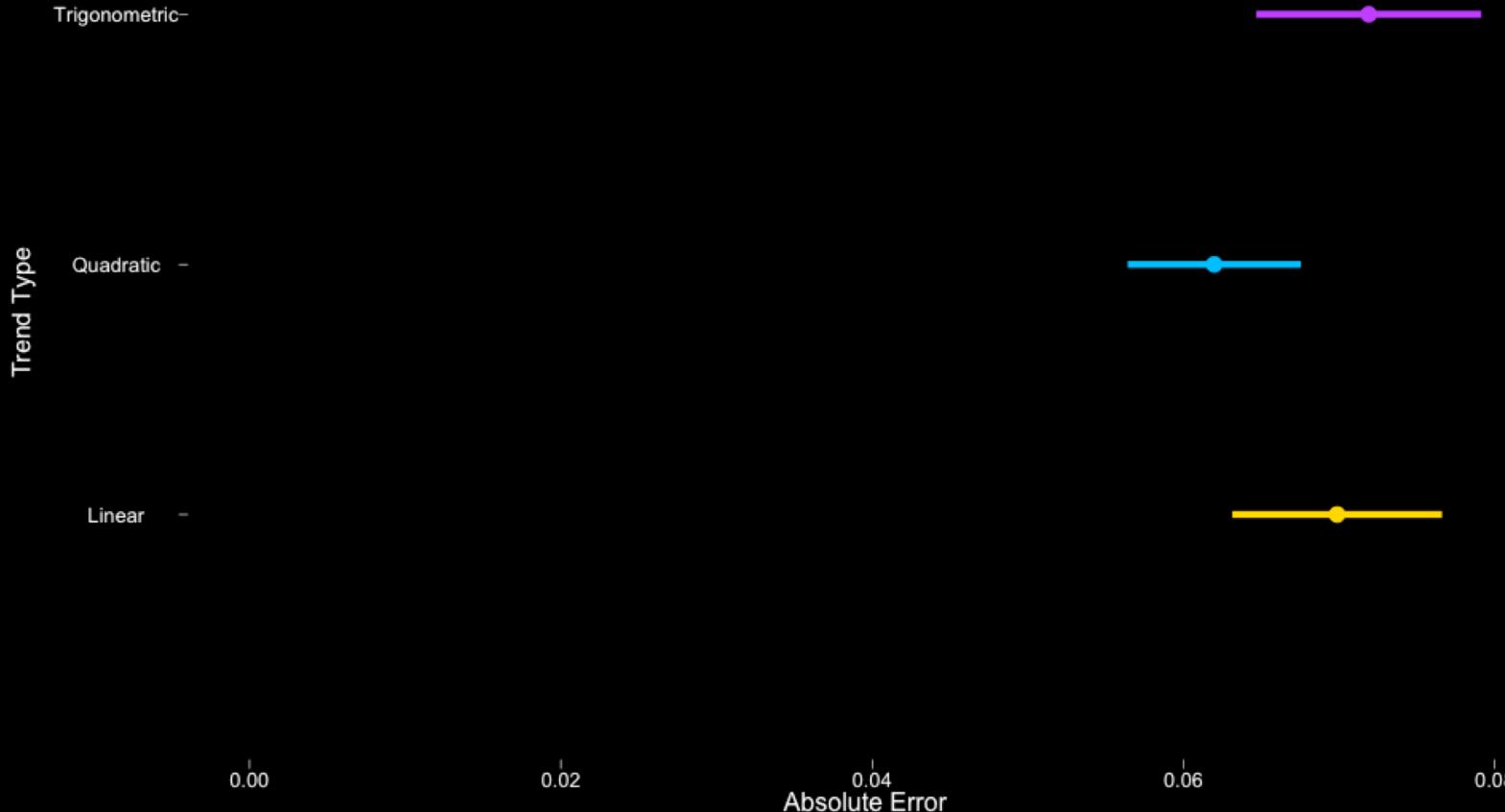
H1

Graph use	T Test (mean of x, p-value)
<i>Original Study</i>	$0.001, p = 0.7$
All participants	-0.009, $p < 0.001$ — undershooting
> once a month	-0.012, $p < 0.001$ — undershooting
< once a month	-0.002, $p = 0.7$

Hypothesis 2 - the larger the residuals, the larger the error



Hypothesis 3 - no statistically significant effect of the trend type



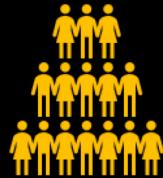
Results

	H1	H2	H3
Original	$p = 0.7$ mean = 0.0008	$p < 0.001$	$p > 0.05$ No pairwise interact.
Our Experiment	$p < 0.001$ mean = -0.0087	$p < 0.001$	$p > 0.1$ No pairwise interact.
Confirmed	No	Yes	Yes

Discussion

- What do you think about the setup?
- Any criticism toward the original paper?
- What could be some causes of variation to the original study?

Limitations



Number of observations (48 vs. 16)



Education bias (83% vs. 100% higher educated)



Gender proportion (70-30 vs. 50-50 M-F)

Acknowledgement

Michael Correll



Source: correl.io

Jeffrey Heer



Source: acm.org



Source: lecture slides



Source: develop.touchstone2.org



Thank you for attention!
Do not hesitate to ask any questions!

