

Comparing the effect of single outliers and outlier clusters on trend estimation in scatterplots

Özgür Saydır

Koç University, İstanbul, İstanbul, Turkey

Aysecan Boduroglu

Koc University, Istanbul, Turkey

Abstract

Scatterplots are commonly used data visualizations to depict relationships between variables. There are inconsistent findings in the literature regarding how outliers in scatterplots affect trendline estimates. Correll & Heer (2017) found no difference for trendline estimations between the no-outlier and the outlier conditions consisting of a separate group of items creating an outlier cluster. However, Ciccione et al. (2022) showed that single outlier points might be included in trendline estimations. To investigate whether an outlier cluster was perceived as a salient and separate unit and thus excluded from the remaining data points, we directly compared the effects of single and multiple outliers on trendline estimations, controlling for correlation strength, outlier position and trend direction. Participants drew trendlines. We found that participants included single outliers more than they've included outlier clusters into the trendlines; this pattern was similar across all other control variables; suggesting grouping might play a role in this process.