

CLUES-WEASEL adjusted Rand index  
(mean: 0.2980)

**CLUES-WEASEL wins here  
[60W, 1T, 41L]**

**Shape-DBA wins here  
[41W, 1T, 60L]**

Shape-DBA adjusted Rand index  
(mean: 0.2314)

\* Dashed lines represent the median

Wilcoxon test for equality of medians:  $p\text{-value} \leq 1e-3$   
Paired t-test for equality of means:  $p\text{-value} \leq 1e-3$

