Sim\_resuls\_df.txt

Columns:

**analysis\_id:** unique string identifier for each analysis

**parameter:** ‘p’ or ‘r’, denoting whether p-value or correlation coefficient was used as decision boundary

**fold\_value:** value of fold value parameter chosen

**stat:** ‘pearson’, ‘rpearson’ etc. where the r-prefix indicates TN/FN separation as opposed to TP/FP separation

**cooksd:** whether Cook’s D was used instead of CUTIE

**seed:** seed used for that individual scatterplot

**class:** label of correlation (NP, FP, FN, CD)

**sample\_size:** number of points in scatterplot (n = 25, 50 or 100)

**corr\_strength:** underlying distribution parameter for correlation strength

**indicator:** 1 or 0 whether correlation is ‘true’ (TP/FN) or ‘false’ (TN/FP) respectively

Real\_results\_df.txt

**analysis\_id:** unique string identifier for each analysis

**parameter:** ‘p’ or ‘r’, denoting whether p-value or correlation coefficient was used as decision boundary

**dataset:** real-world data type used in analysis

**statistic:** ‘pearson’, ‘rpearson’ etc. where the r-prefix indicates TN/FN separation as opposed to TP/FP separation

**mc\_used:** whether multiple corrections adjustment was used (nomc = none, FDR = false discovery rate)

**fold\_value:** value of fold value parameter chosen

**cooksd:** whether Cook’s D was used instead of CUTIE

**initial\_corr:** number of correlations p <  (in TP/FP separation) or p >  (in TN/FN separation)

**true\_corr(TP\_FN):** number of correlations that were classified as TP (in TP/FP separation) or FN (in TN/FN separation)

**false\_corr(FP\_TN):** number of correlations classified as FP (in TP/FP separation) or TN (TN/FN separation)

**rs\_true\_corr\_TP\_FN:** number of correlations with reversed sign in TP or FN set **rs\_false\_corr\_FP\_TN:** number of correlations with reversed sign in FP or TN set runtime: runtime of analysis in seconds