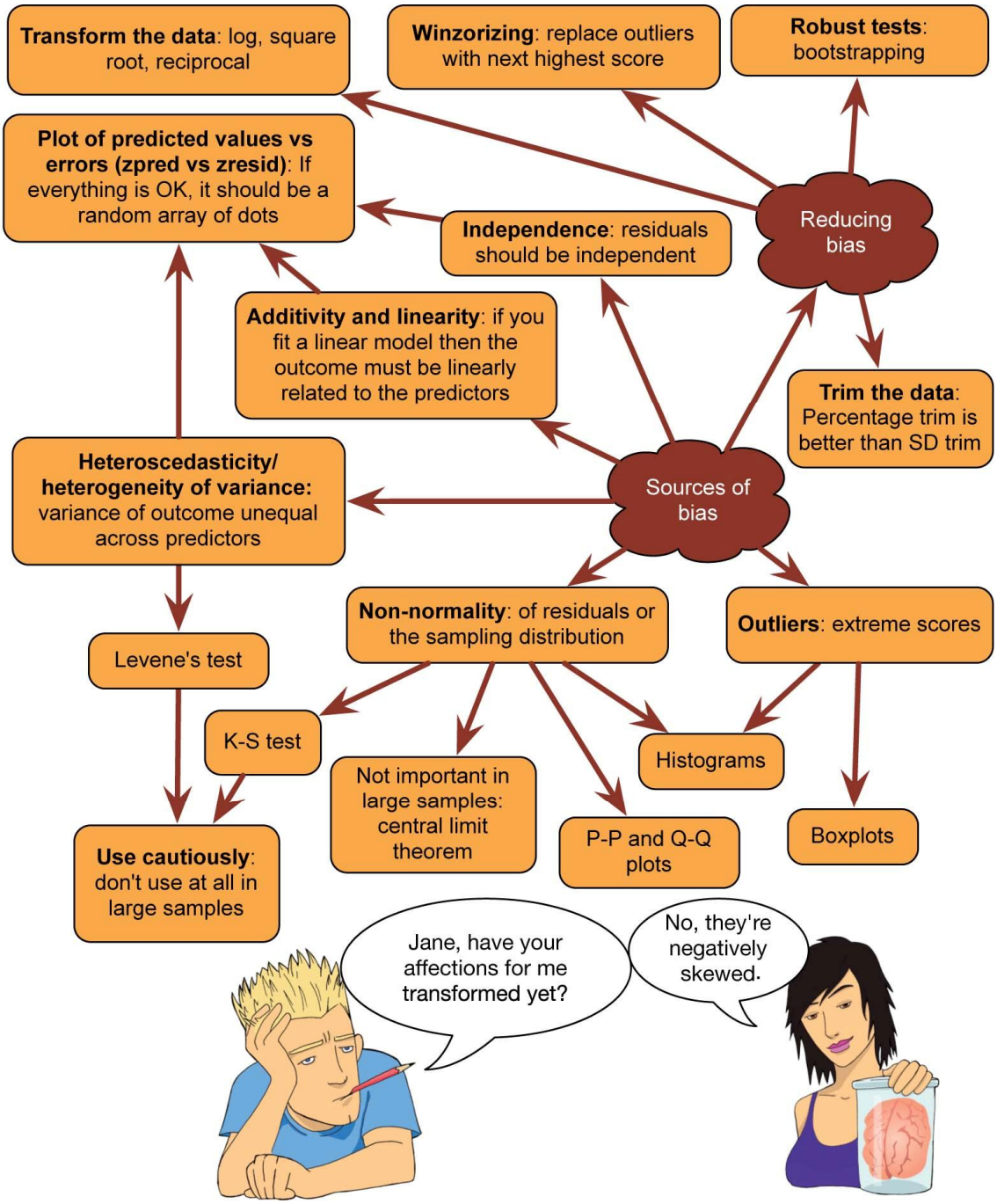
Transform the data: log, square root, reciprocal
Winzorizing: replace outliers with next highest score
Robust tests: bootstrapping
Plot of predicted values vs errors (zpred vs zresid): If everything is OK, it should be a random array of dots
Reducing bias
Independence: residuals should be independent
Additivity and linearity: if you fit a linear model then the outcome must be linearly related to the predictors
Trim the data: Percentage trim is better than SD trim
Heteroscedasticity/ heterogeneity of variance: variance of outcome unequal across predictors
Sources of bias
Non-normality: of residuals or the sampling distribution
Outliers: extreme scores
Levene's test
K-S test
Histograms
Not important in large samples: central limit theorem
P-P and Q-Q plots
Boxplots
Use cautiously: don't use at all in large samples
Jane, have your affections for me transformed yet?
No, they're negatively skewed.