Figures

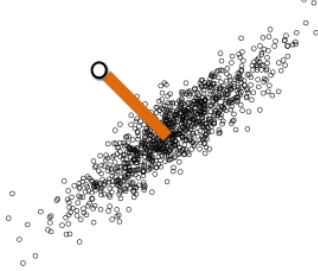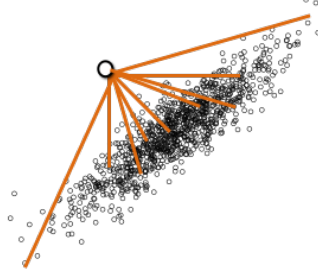**Mahalanobis distance**

$$D_M(x_i) = \sqrt{(x_i - \bar{x})^T S^{-1}(x_i - \bar{x})}$$

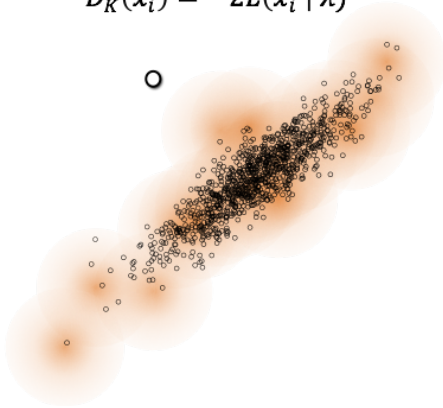**Harmonic mean distance**

$$D_H(x_i) = N \left[\sum_{j \neq i} \left[(x_i - x_j) S^{-1}(x_i - x_j) \right]^{-1/2} \right]^{-1}$$

Kernel Density Deviance

$$L(x_i \mid \lambda) = \log\left(\frac{1}{N-1} \sum_{j \neq i} G(x_i \mid x_j, \lambda^2 S) \right)$$
$$D_K(x_i) = -2L(x_i \mid \lambda)$$

Nearest neighbor distance

$$D_N(x_i) = \min_{j \neq i}\left(\sqrt{(x_i - x_j) S^{-1}(x_i - x_j)} \right)$$
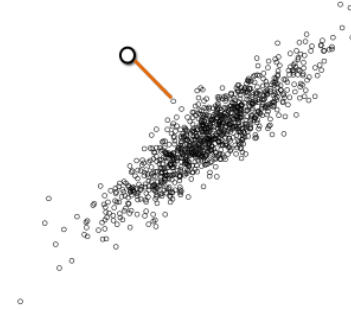
FIGURE 1. Conceptual examples of the four multivariate statistics compared in this study.

TO DO: Add ABCD.

TEAM: Would anyone like to take on making a figure like this one, but better for publication? I made these in power point, which is why they are awful. I think the kernel deviance could be challenging to visualize, because technically the bandwidth should be proportional to the variance in each dimension. The harmonic mean could also be challenging, because that is a lot of overlapping lines to draw. It might be good to draw the lines first with transparency=0.1 and lay the points on top.
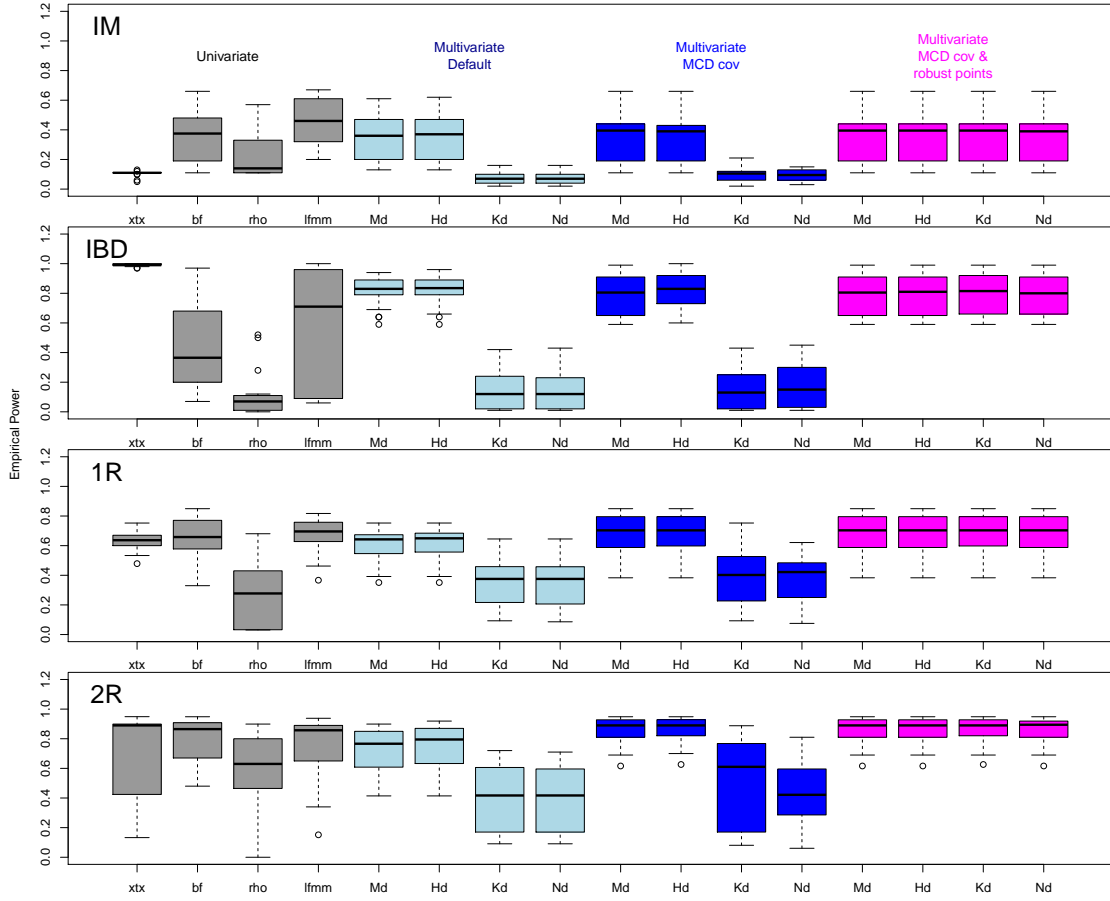
FIGURE 2. Comparison of univariate and multivariate statistics for their power to separate signals of selection from neutrality. Each row represents one of the simulated demographies (IM = Island Model, IBD = Isolation By Distance, 1R = range expansion from one refuge, 2R = range expansion from two refugia). The four univariate statistics (grey box plots) evaluated were: $X^T X$ (xtx), Bayes Factors (bf), Spearman's ρ, and Z-scores (lfmm).
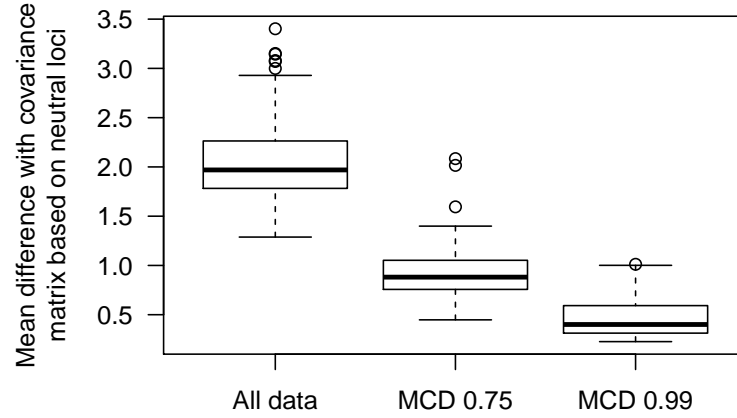
FIGURE 3. Comparison of the mean absolute difference between the variance-covariance matrix among the univariate statistics calculated using (i) all the data, (ii) the MCD estimate using 75% of the dataset, and (iii) the MCD estimate using 99% of the dataset. Boxplots are summarized over all demographies and sampling designs.
TO DO: Should I switch MCD99 and MCD75 here, since MCD99 will be closer to "all data" than MCD75?
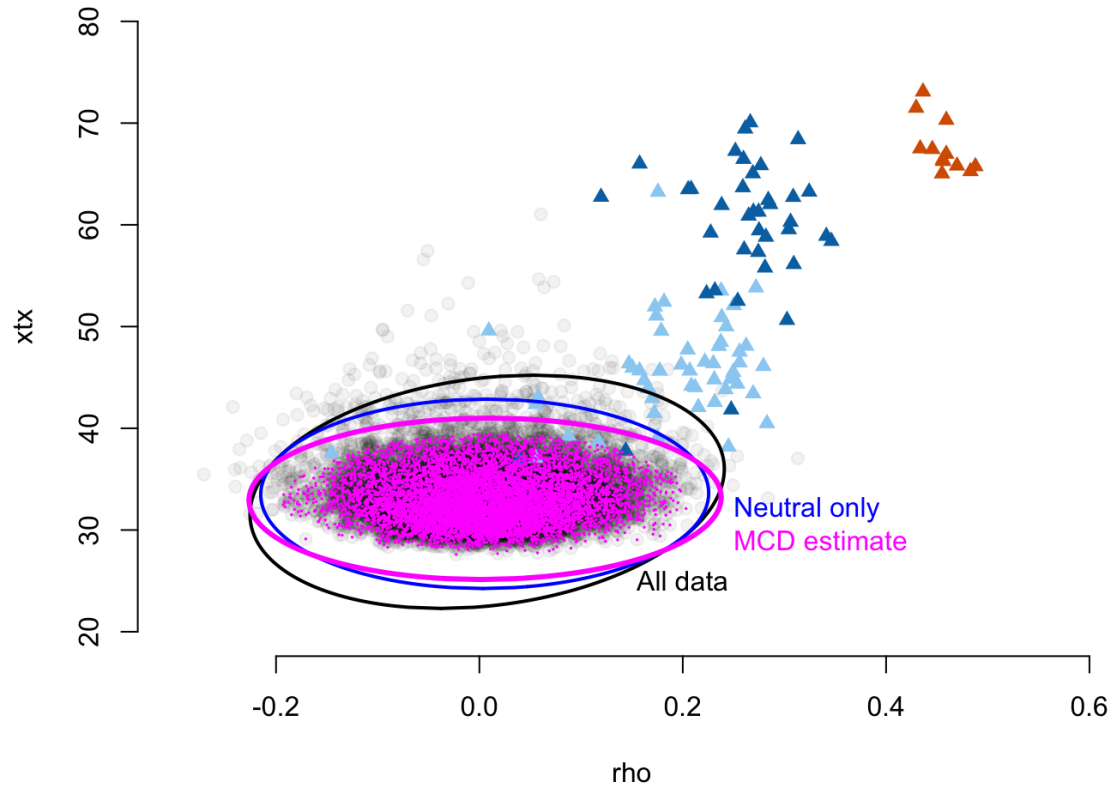
FIGURE 4. Comparison of 99% confidence intervals on the two dimensional ellipse given by the determinant on the covariance matrix between these two variables calculated from: (i) all the data, (ii) neutral loci only, and (iii) the MCD estimate.

TEAM: does this visualize that the robust points are neutral or is it unclear?

TO DO: Need to add legend to this plot (neutral = grey, trianges = different strengths of selection, pink dots = robust points). Replace "xtx" with $X^T X$ and rho with Spearman's ρ. Maybe change the colors so that neutral ellipse is black, and change the lines so that it works in BW.
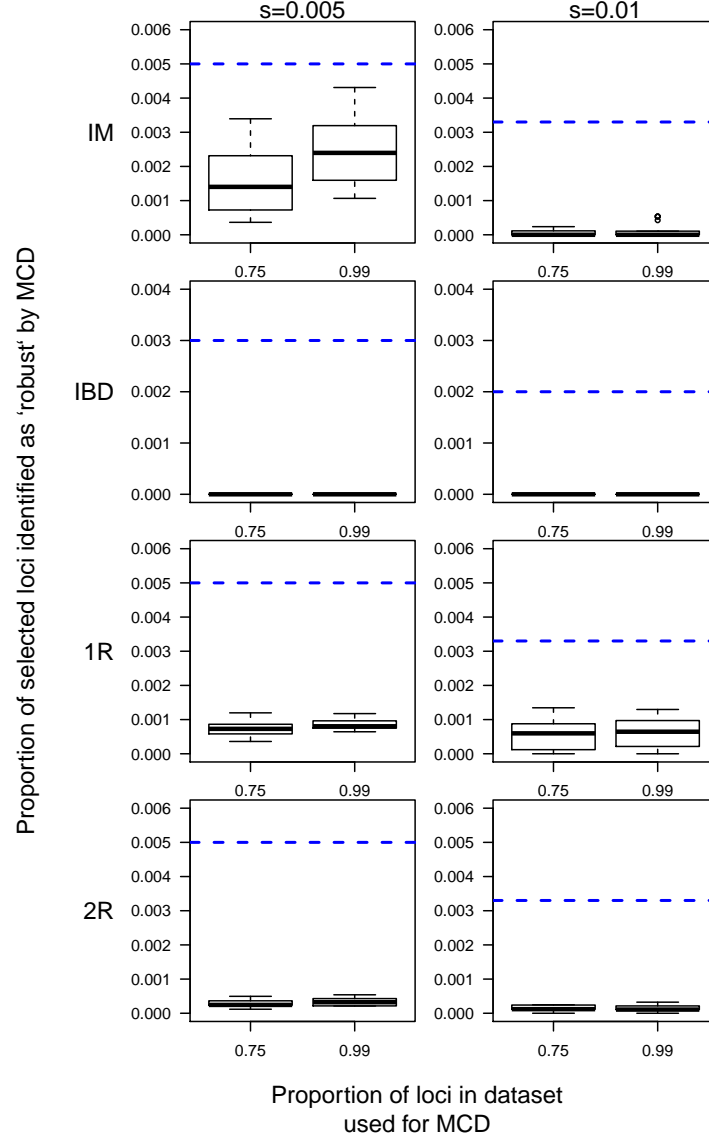
FIGURE 5. Evaluation of whether loci affected by selection were identified as 'robust' points (e.g., not outliers) by the MCD algorithm. Loci under relatively weak selection (s=0.005) are in the left column, well loci under moderate selection are in the right column (s=0.01). The horizontal dashed line represents the null expectation, based on the proportion of loci under that strength of selection in the entire dataset. The MCD never identified loci simulated under strong selection (s=0.1) as robust (expected = 0.001 in IBD and 0.0017 in IM, 1R, and 2R; observed = 0 across all simulations).
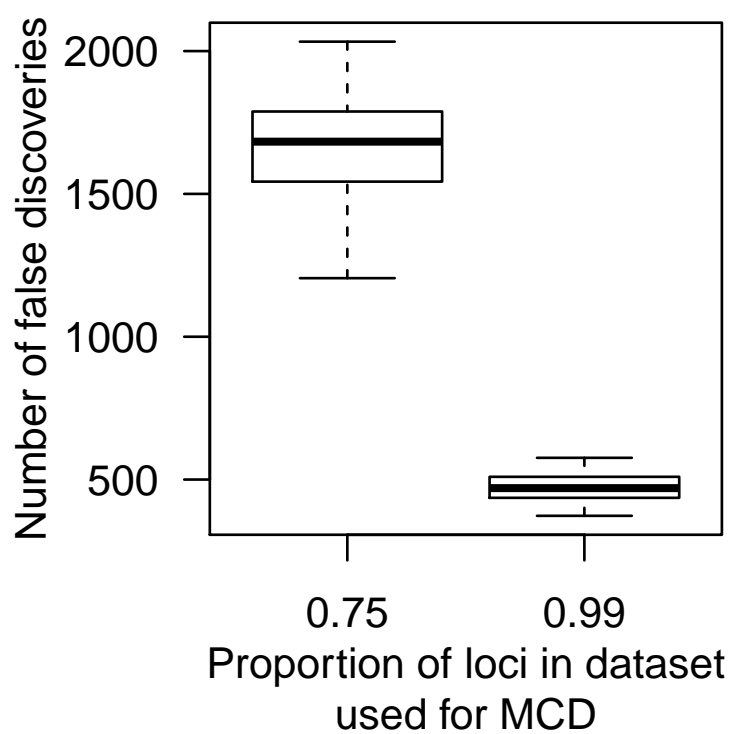
FIGURE 6. The number of false discoveries (neutral loci inferred to be outliers, out of 9900 total neutral loci in the dataset) resulting if robust points identified by the MCD are used as an empirical null distribution.