**Workflow for R and presentation:**S

Dataset Introduction

Background of the information

Outlier samples with visual proof, and remove the outliers

Filter out genes with low expression values so we only get the best numbers

Conduct feature selection with stat test or other machine learning method(depends on factors that are included in the data set)(ex. 2 conditions = two-sample test)

Adjust for multiplicity and provide # of genes retained with associated score(p-val) and the threshold used

Plot these scores into a histogram

Next, subset data by the genes determined, use clustering/dim red. Methods to see in 2D space(in dendrogram like the lab?)

Use linear projections of the original data(cluster centroids/latent variables), then use a classification method to classify the samples to their respective classes(color for PREDICTED CLASS MEMBERSHIP, dif symbol points for ACTUAL class membership)

Use top 5 discriminant genes in the positive and negative direction(10 total), and go to NCBI DAVID and look up their gene info(include name and functional information like assoc. pathways)

Conclusions

References

Text file with all the code for the project