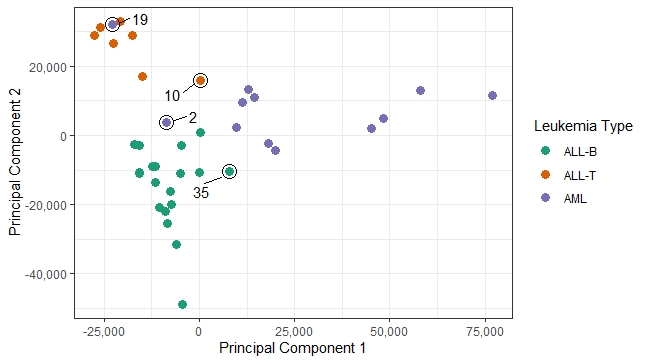
Email Subject: DNA for Leukemia Research - Incorrectly Labeled Samples

(Linear Algebra Homework Assignment)

Dear Olive,

Our investigations revealed that samples 2, 10, 19, and 35 require profiling to determine if they have been labeled correctly.

We conducted an analysis to condense the information from the 5000 gene expressions into two principal components. We plotted the 38 samples using one component as the x-axis and the second component as the y-axis. The three different colors represent the three Leukemia types present in the dataset. When plotted, the 38 samples are divided into approximately three clusters, one for each Leukemia type. We recommend profiling the samples circled, 2, 10, 19, and 35, as their placement in relation to the clusters indicates they may be mislabeled.



If you have any additional questions, please let us know.

Best Regards,

Data Team

(Orange Team 12 - Thomas Gow, Payal Patel, Keyur Shah, Wes Thompson, Sarah Wotus)