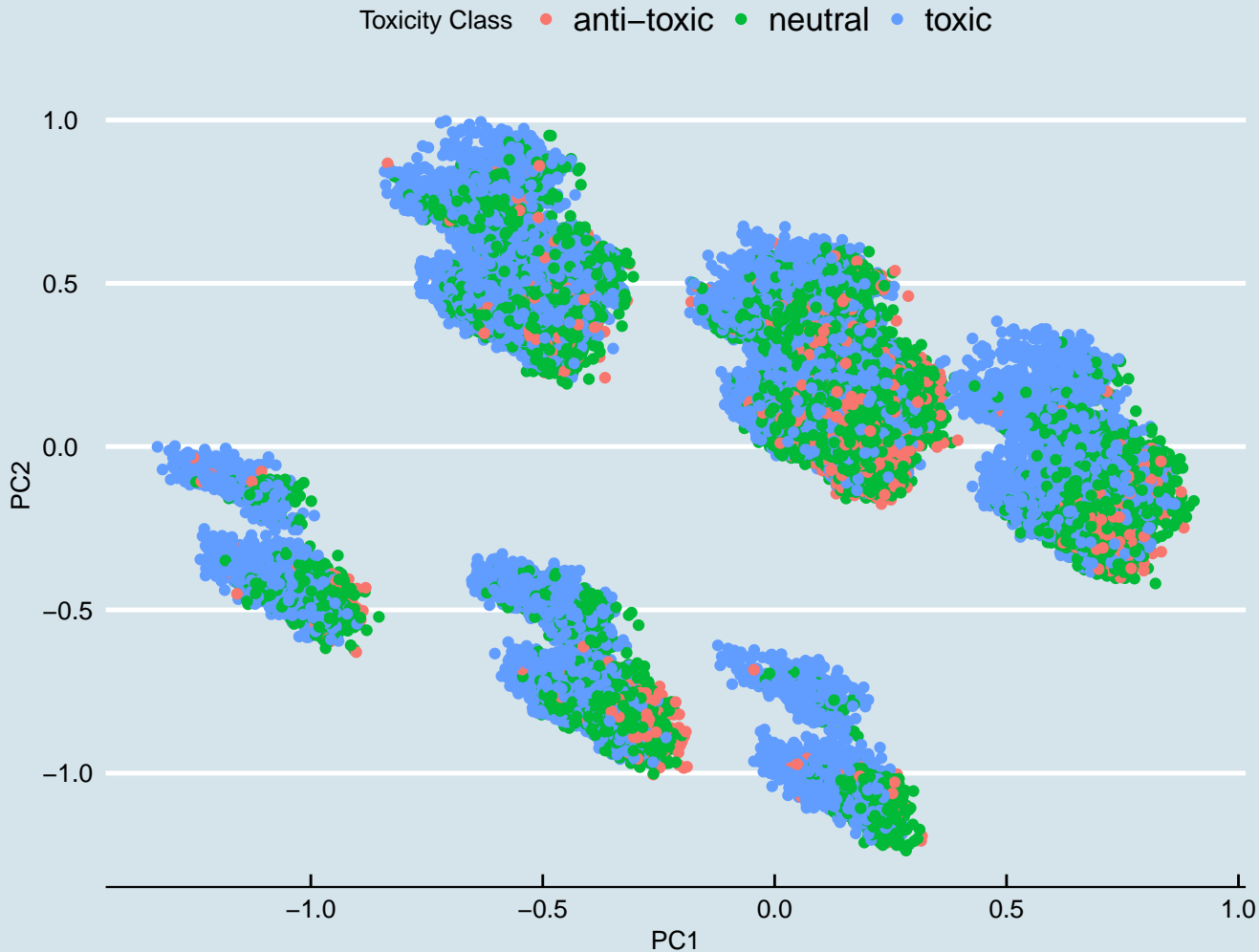**Principal Component Analysis Plot**

Looking at the plot of the first two principal components, we see that the data clusters into about 12 distinct groups. Data points are colored by their toxicity class to visualize the relative frequencies of each tox class in each cluster. Most clusters contain roughly equal numbers of Toxic and Neutral peptides. However, there are some clusters (top right, bottom right) that are predominantly Toxic. This supports our hypothesis that there are patterns in the sequence data that confer toxicity.