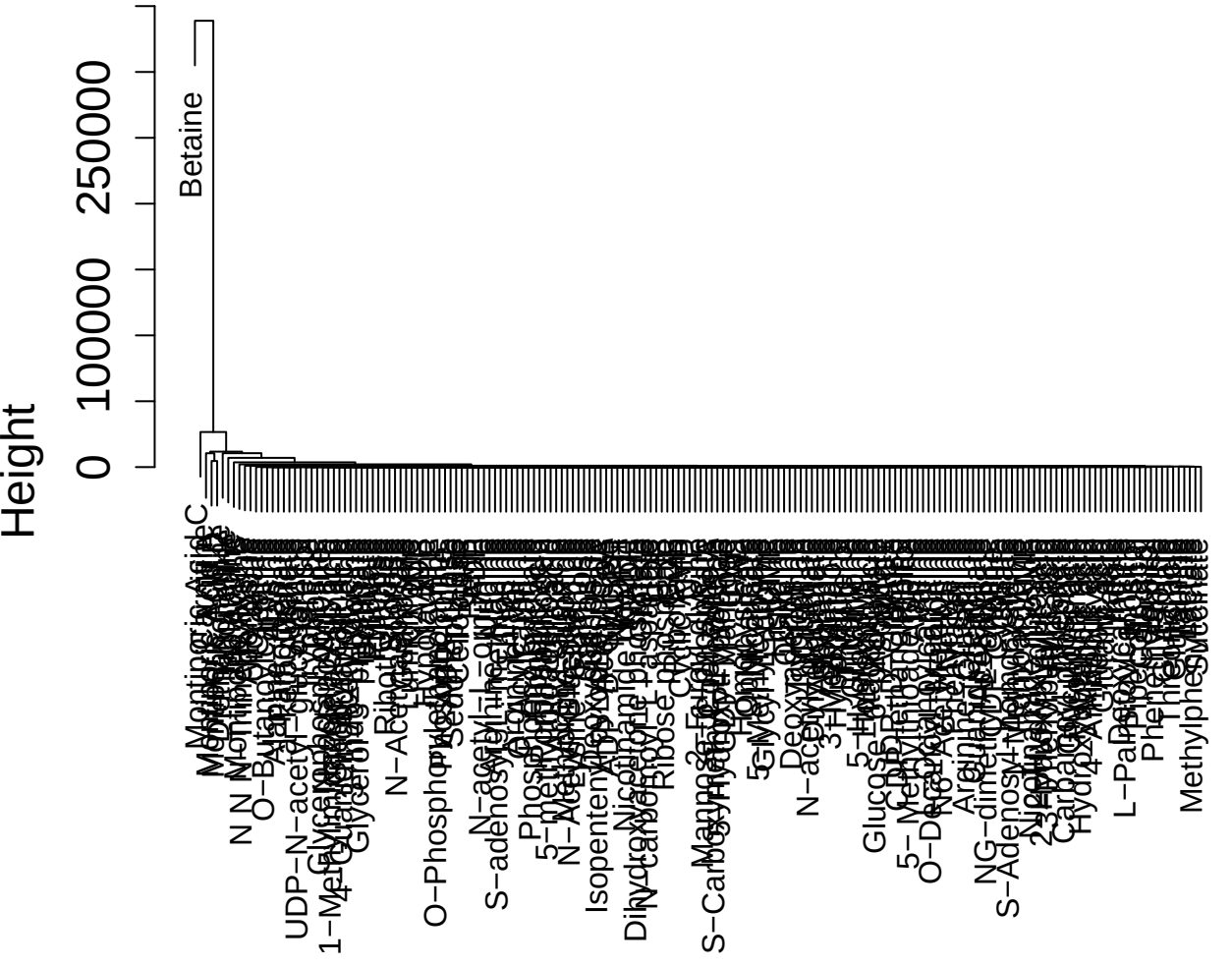Sample clustering to detect outliers