## Outliers

Code obtained from lecture material.

Mahalanobis Distance is a depth-based method for removing outliers. As values in our dataset are highly correlated, this method of outlier removal was chosen as this method accounts for highly correlated features within a dataset and standardises the data.

Critical values were obtained by calculating the Hotellings T-Squared from n rows with k columns. The Mahalanobis distance for each row was calculated from the overall multivariate mean. The covariance matrix standardises the data, ensuring that highly correlated values do not bias the outcome.  Only numeric columns can be selected for covariance calculation. Therefore, only columns containing numeric values have their outliers removed.

Outliers detected shown in graph:

A graph of a number of data

AI-generated content may be incorrect.

Removing outliers removed 560 rows of data. Boxplots confirm that outliers have been removed.