Problem 2: Capturing the essence of gemstones

The dataset `gemstones.txt` contains information on 300 different gemstones, each characterized by eight quantitative features, including area, perimeter, major axis length, minor axis length, eccentricity, convex area, equivalent diameter (diameter of the circle of the same area), and roundness (calculated with the formula $4\pi\text{area}/\text{perimeter}^2$), all obtained using high-resolution 2D imaging technology. The dataset also includes details about the type of gemstone.

a) Conduct a Principal Component Analysis (PCA) of the dataset, focusing on the quantitative variables only. Determine whether it is more appropriate to use the original variables or the standardized ones and proceed accordingly.

b) Report a plot of the loadings of the first two principal components and provide an interpretation. Report the scatter plot of the data along the first two principal components. Considering the categorical variable `variety`, interpret the results.

c) For gemstones of the `ruby` variety, construct a 95% confidence region for the mean of the vector whose components are the first two principal components. Describe this region by providing its center coordinates, axis directions, and semi-axes lengths. Provide a plot of the region. Introduce and assess the hypothesis of normality, which will here be tested at the 1% significance level..