1. When analyzing cranial data, biological anthropologists often use size-adjusted data. The method most-often used is to divide each measurement by the geometric mean for that individual. The geometric mean is computed by multiplying all the values together and then taking the \( n \)th root (where \( n \) is the number of measurements for that individual). For example the geometric mean of 5, 35, and 50 is \( (5 \times 35 \times 50)^{(1/3)} \) = 20.6 so the size-adjusted values would be 5/20.6, 35/20.6, 50/20.6 or .24, 1.70, 2.43. Because the numbers get very big quickly, it is usually better to compute the geometric mean using logarithms since multiplication becomes addition and taking roots becomes division. If \( a = c(5, 35, \text{and } 50) \), then the geometric mean, \( \text{gm} = \exp(\text{mean}(\log(a))) \) and the size-corrected values are \( \exp(\log(a) – \text{mean}(\log(a))) \). For this problem we are going to size-correct the measurements in the Howells3Pops data set and then conduct a Principal Components Analysis. Load the Howells3Pops data set and then type and submit the following four commands:

\[
\begin{align*}
\text{H3PopsStd} & \leftarrow \text{data.frame(Howells3Pops[,1:2], log(Howells3Pops[,3:18]))} \\
\text{GMean} & \leftarrow \text{apply(H3PopsStd[,3:18], 1, mean)} \\
\text{H3PopsStd} & \leftarrow \text{data.frame(H3PopsStd[,1:2], sweep(H3PopsStd[,3:18], 1, GMean))} \\
\text{H3PopsStd} & \leftarrow \text{data.frame(H3PopsStd[,1:2], exp(H3PopsStd[,3:18]))}
\end{align*}
\]

Now use Rcmdr to run the analysis and add five principal component scores to \( \text{H3PopsStd} \). Rerun the command analysis by highlighting the command that creates \( .PC \) and create a biplot with \( \text{biplot(.PC, xlabs=unclass(H3PopsStd$Population))} \). Which five variables are most important in principal components 1 and 2 (five in each component)?

2. Now use the Scatterplot matrix in Rcmdr to plot the five components against one another. Turn off the least squares and smooth lines and use Population as the Plot by groups variable. Which principal components seem to separate the four groups the best? Finally do a 3D scatterplot using the first three principal components. Turn off the linear least squares and plot by Population. There is no legend, but the blue points are the Australian population, the green are the Buriat, and the orange (more like gold on my display) are the Bushman. Rotate the graph and look at the separation between the groups. Which one is most distinct (has the least overlap with either of the other two)?

3. The data set ObsidianSources comes from Michelle Woodward’s master's thesis. She collected obsidian from three localities in Guatemala. The data consist of the common logarithms (base 10) of 20 different elements (so “lal” is the log10 of Aluminum). Use principal components analysis. Print the results and a biplot using \( \text{biplot(.PC, xlabs=ObsidianSources$group)} \). How well are the three sources separated? Which elements are most important in defining components 1 and 2?