Name _____

Psychology 516
Applied Multivariate Analysis
Homework 3
Due September 25, 2018

The file, Set_3.csv, contains the data from a study in which 500 high school students completed a measure of scholastic aptitude: Grammar, Paragraph Comprehension, Vocabulary, Sentence Completion, Geometry, Algebra, Numerical Puzzles, Series Completion, Practical Problem Solving, Symbol Manipulation, Analytical Ability, and Formal Logic. Answer the following questions about these data:

1. What evidence do you have that these data should be subjected to a principal components analysis?
2. How many principal components should be extracted?
3. How much variance do these extracted components account for in the original data?
4. How much variance in the original Geometry variable is accounted for by these extracted components?
5. Now screen the data for unusual cases and determine if your conclusions change when any such cases are excluded from the analysis.

   If you believe there is more than one outlier in the data, follow a sequential approach to determining how many to exclude. This means that you will identify the worst offender, exclude that case, and then repeat your diagnostics to determine if other outliers are present. If so, again exclude the worst one, and repeat the diagnostics to determine if an additional outlier is present. Keep cycling through these steps until you are satisfied you have all outliers identified and excluded. Then conduct the principal components analysis. This iterative approach is necessary for multivariate diagnostics such as Mahalanobis distance because the presence of one outlier can influence the apparent presence of others via their joint influence on the covariance matrix. Removing them one at a time insures you don't miss any or mistakenly remove cases that are not really outliers.