**Dataset:**

600 students participated in a university-wide experiment to test if an educational intervention increased test performance. Half of the students were assigned to be in the control group and half of the students were assigned to be in the experimental group (the experimenters used matching techniques to keep the groups as random as possible). The experimental group was the only group that received the educational intervention. All participants were given exam 1 to assess their initial test-taking ability. Then, for the next 4 weeks, the professors began administering the educational intervention to only the experimental group. All participants then took a midterm. For the last four weeks of class, professors again administered the educational intervention to only the experimental group. All participants then took the final exam. Can it be concluded that the educational intervention worked (e.g., did the experimental group outperform the control group?)?

**Variables**:

* Gender (1 = male, 2 = female)
* Group (1 = control group, 2 = experimental group)
* 3 exam scores, ranging from 2-100 points (students received two points for writing their name and no extra credit was available). Decimals are possible!
  + Beginning
  + Midterm
  + Final exam

**Screen the dataset for the following problems:**

1. Accuracy:
   1. Include output and indicate how the data *are not* accurate.
   2. Include output to show how you fixed the accuracy errors, and describe what you did.
2. Missing data:
   1. Include output that shows you have missing data.
   2. Include output and a description that shows what you did with the missing data.
      1. Replace all participant data if they have less than 20% of missing data by row.
      2. You can leave out the other participants (i.e. you do not have to create filled in \_ missing).
3. Outliers:
   1. Include a summary of your mahal scores.
   2. What are the *df* for your Mahalanobis cutoff?
   3. What is the cut off score for your Mahalanobis measure?
   4. How many outliers did you have?
   5. Delete all outliers.
4. Assumptions:
   1. Additivity:
      1. Include the symnum bivariate correlation table of your continuous measures.
      2. Do you have issues with multicollinearity or singularity?
   2. Linearity:
      1. Include a picture that shows how you might assess multivariate linearity.
      2. Do you think you’ve met the assumption for linearity?
   3. Normality:
      1. Include a picture that shows how you might assess multivariate normality.
      2. Do you think you’ve met the assumption for normality?
      3. Include skew and kurtosis values for your exam scores (three sets).
   4. Homogeneity:
      1. Include a picture that shows how you might assess multivariate homogeneity.
      2. Do you think you’ve met the assumption for homogeneity?
   5. Homoscedasticity:
      1. Include a picture that shows how you might assess multivariate homoscedasticity.
      2. Do you think you’ve met the assumption for homoscedasticity?