Data Screening Class Assignment

Download the data screening file to analyze. You want to screen the whole dataset at once to look for issues, since all the variables will be used in several different hypotheses.

**Study:** This dataset includes a male body dissatisfaction scale with the following questions:

1. I think my body should be leaner
2. I am concerned that my stomach is too flabby
3. I feel dissatisfied with my overall body build
4. I think I have too much fat on my body
5. I think my abs are not thin enough
6. I feel satisfied with the size and shape of my body
7. Has eating sweets, cakes, or other high calorie food made you feel fat or weak?
8. Have you felt excessively large and rounded (i.e., fat)?
9. Have you felt ashamed of your body size or shape?
10. Has seeing your reflection (e.g., in a mirror or window) made you feel badly about your size or shape?
11. Have you been so worried about your body size or shape that you have been feeling that you ought to diet?

Check the data for the following:

**Accuracy:**

1. Check the data for out of range scores. The scale ranges from 1 (never) – 6 (always).
   1. Include a summary showing you do/do not have out of range scores.
   2. If necessary, fix the out of range scores.
      1. Indicate what the problems were in the dataset.
      2. Make all out of range values NA.
      3. Include a summary showing that you fixed the accuracy issues.

**Missing data:**

1. What type of missing data do you appear to have?
2. Include a summary of the missing data by participant.
   1. For example, you can do table(missing) to see the percentages by participant.
3. Include a table of the missing data by column after you exclude participants with too much missing data.
4. If necessary, use *mice* to impute the data.
   1. Include a summary of the data that shows that you might have NA values for categorical columns but not continuous columns.

**Outliers:**

1. Calculate Mahalanobis distance scores for your data.
   1. What is your *df* for the cut off score?
   2. What is the cut off score?
   3. How many outliers did you have? You can include the summary of the mahal < cutoff.
   4. Delete the outliers.

**Additivity:**

1. Include a symnum table of the continuous variables.
2. Are any of the variables too highly correlated?

**Normality:**

1. Include the multivariate normality histogram.
2. Interpret the graph. Does it indicate multivariate normality?

**Linearity:**

1. Include the multivariate QQ plot.
2. Interpret the graph. Does it indicate multivariate linearity?

**Homogeneity/Homoscedasticity:**

1. Include the multivariate residuals plot.
2. Interpret the graph. Does it indicate homogeneity?
3. Interpret the graph. Does it indicate homoscedasticity?