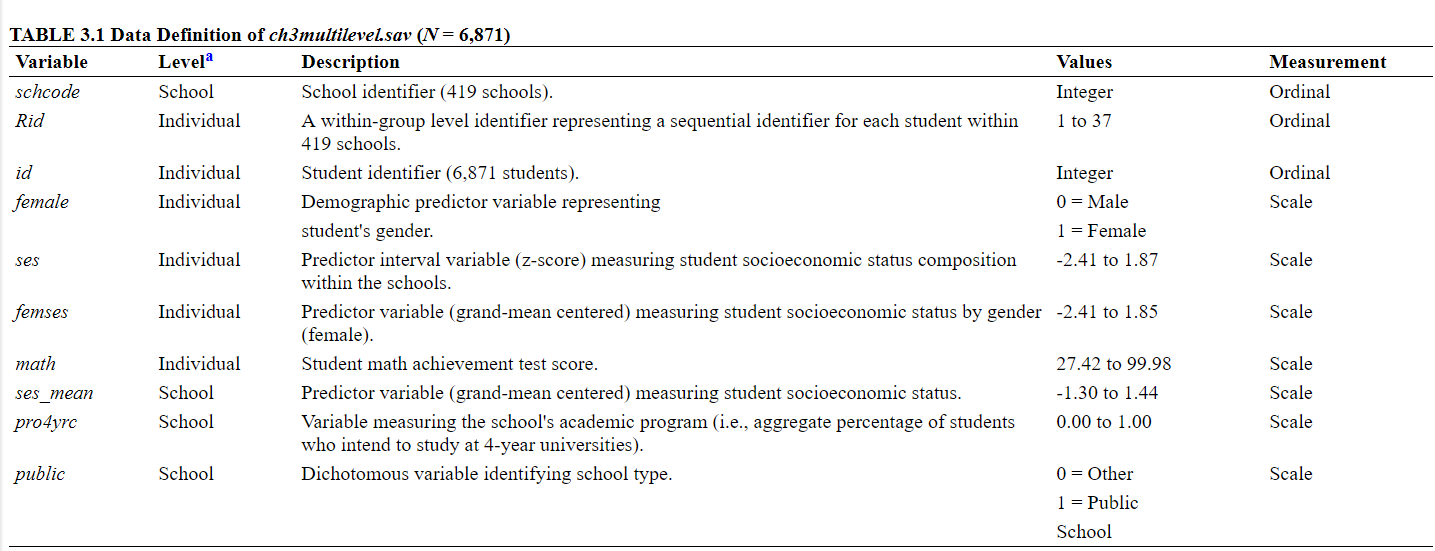
In this chapter, we will run our first multilevel model to account for the clustered nature of our data and begin visualizing and understanding the variance in our data.

The data for Module 3 were taken from chapter 3 of Heck, R. H., Thomas, S. L., & Tabata, L. N. (2011). *Multilevel and Longitudinal Modeling with IBM SPSS*: Taylor & Francis. These are the variables in the dataset:



1. Load the data stored in the heck2011.csv file and the packages we’ll be using for this data demo: dplyr, ggplot2, lme4, lmerTest and performance.
2. In Module 2, we created a scatterplot with math achievement (“math”) on the y-axis and socioeconomic status (“ses”) on the x-axis. Re-create that graph, ignoring the clustered data structure, and adding a regression line.
3. Create the same scatterplot, this time by school (“schcode”). Describe the intercepts and slopes for different schools.
   1. Do you see any issue with treating the data like observations are independent?
4. Write out the equation for the random intercept/null model
   1. L1, L2, and combined
5. Key output to note/interpret is
   1. Number of parameters
   2. Estimates of fixed effects
   3. Estimates of covariance parameters
      1. Residual
      2. Intercept
6. Run the random intercept model/ null model
7. Calculate the ICC
   1. What is the total variance?
   2. What is the % of variance between schools?
8. Describe how much the schools vary in math achievement by using the estimated between school variance and the intercept (i.e., calculate the 95% plausible values range)
9. Estimate the empirical Bayes estimates of the school specific intercept random effects
   1. Create a histogram of the EB estimates
   2. What is the mean of these EB estimates of the Uojs?