The data set I choose is called Diagnosed Diabetes by Age. The data set groups ages by 18-44, 45-64, 65-74, and 75+ and shows the percentage of survey respondents that reportedly have diabetes. In addition, the upper and lower limits of the 95% confidence interval for each age group are also displayed. This data is collected annually by the Center for Disease Control and Prevention. This data was taken from the National Health Interview Survey (NHIS). The NHIS Health Interview Survey is extensive, and covers a variety of topics related to disease, illness, and health care usage. The data set I choose displays the time series of the data specific to diabetes cases. The CDC is responsible for creating this specific data set from the NHIS survey. This survey is conducted using a representative sample and by both in person and online survey means. The survey data is collected annually. However, the model used to determine how to make the representative sample is updated every ten years. It is important to note that when looking at the percentage of diabetes cases in the population, this includes both type one and type two diabetes. This diabetes percentage by population does not reflect gestational diabetes. Surveyors changed how they asked Female respondents to respond to this survey in order to exclude gestational diabetes.

I choose this dataset because I myself am type one diabetic and find learning diabetes trends interesting. One trend I have seen in my personal life is that most type one diabetes products are now becoming FDA approved for type two diabetics. This often means that as the number of diabetics increase, the funding and technology often improve. I am interested to learn more about the diabetes trend in specific reference to age.