The dataset diagram, highlighted below in the red box, can be found near the top of an analysis page, below the study title.
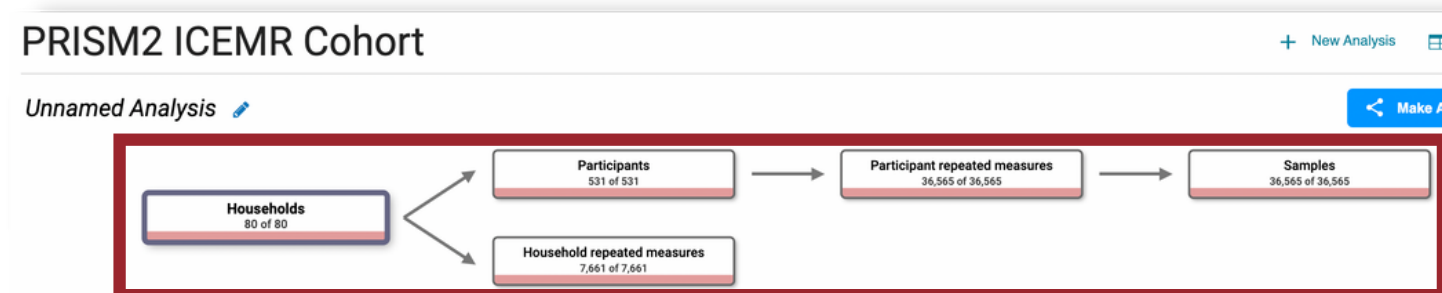
The dataset diagram shows us the different entities or categories that data was collected for - for instance, households, participants, and samples.

It also indicates TWO important things-
(a) **The sample size**; i.e., number of observations in each category, and
(b) **The shape of the dataset**, i.e., whether data was collected once or at more than one timepoint.

In the example above, the dataset diagram indicates that 80 households were enrolled in the study, and 531 participants, i.e., more than one participant per household.

The presence of **Households** and **Household repeated measures** indicates that some variables that collected data about the participant's household were asked only once, but other variables that collected household data were asked at more than one timepoint, which we group under repeated measures.

There were 531 participants, and 36,565 participant repeated measures, indicating a longitudinal study with repeated follow-up. Finally, there are 36,565 samples, indicating that one sample was collected at each repeated observation. Thus the dataset diagram provides an overview of the data.

Please contact **help@clinepidb.org** with feedback or questions!