Nikhil Gopal

Explain how I might record the data, if variable is observation or ancillary, if its relevant or only to indicate structure, if its more natural to think of variable as indicating shared sources of variability or different conditional distributions.

1. The patient being cured or not is an observation, as is the treatment experienced. The goal of the study would be to see if type of treatment (placebo vs drug) affects if the patient is cured, so both of these values are crucial to the experiment. I would have rows represent individual study participants, and columns for type of treatment received, patient ID and if they were cured.
2. For this study, I would use two columns for cure status instead of one, since patients will be tested twice. The columns would correspond to the first and second tests accordingly. Additionally, the type of treatment wouldn’t be placebo vs drug but would instead be placebo first vs drug first. This way, researchers could compare individuals to themselves using the patient ID, and could also compare groups to each other (placebo first vs treatment first). As with the first question, all of the variables recorded are observations and not ancillaries.
3. For this study, I would use rows for individuals, and I would create separate columns for subject ID, systolic blood pressure, diastolic blood pressure, and date and city. I would say that all variables are observations except the date, which is an ancillary. Since we are trying to compare results between SF/NYC, city is an extremely important variable. Date is less important, but it could be useful in case you are testing the same subjects on multiple days. In a random sample with a big enough sample size, ideally variability stemming from a person being extra stressed out on a particular day would be minimized by the randomization of the entire population.
4. For this data, I would make the date/time the rows and I would make the recorded measurements of each researcher the columns, I would also add a column for weather. The measurements are definitely observations and weather would be an ancillary, since weather could slightly alter measurements of the asteroid’s position, but likely does not influence that much.
5. Since the scientists are interested in the mass of the individual planets, I would say that each row should be a different planet. For columns, I would record position, time of recording, how catapults were setup, height of objects used and velocity. For this data, position and times are the most important variables, as they will be used to calculate the mass of the planet. Everything else are ancillaries, as they won’t directly factor into the mass calculations.s