Stata merging – applied examples:

1:1 MERGE

- **Context:**
  - your original data set ("left hand side" or "master") is unique by the variable you want to merge on, so there is **one row per ID**.
  - your second data set ("right hand side" or "using") is also unique by that same variable, so there is **one row per ID**.
- **Example:** Merging the alcohol use questionnaire data ("using"/second data) from NHANES to the demographic file ("master"/original data). Each is unique by person ID.

1:M MERGE

- **Context:**
  - your original data set ("left hand side" or "master") is unique by the variable you want to merge on, so there is **one row per ID**.
  - your second data set ("right hand side" or "using") is not unique by that variable, so there are **many rows per ID**.
- **Example:** Merging all current medications ("using"/second data) to a list of visits in EHR data ("master"/original data). The visit data is unique by visit ID, but there can be many medications per visit ID in the medications file.

M:1 MERGE

- **Context:**
  - your original data set ("left hand side" or "master") is not unique by the variable you want to merge on, so there are **many rows per ID**.
  - your second data set ("right hand side" or "using") is unique by that same variable, so there is **one row per ID**.
- **Example:** Merging county-level poverty rates from the Census ("using"/second data) to hospital discharge data ("master"/original data). The discharge data can have *many* records in the same county, but there is only one record per county in the Census data.

M:M MERGE ***(rare and almost never recommended – see joinby command if you think you need a M:M merge since the merge m:m can give unexpected results. You do not need to know or practice m:m merges for this course/exam)***

- **Context:**
  - your original data set ("left hand side" or "master") is not unique by the variable you want to merge on, so there are **many rows per ID**.
  - your second data set ("right hand side" or "using") is also not unique by that variable, so there are **many rows per ID**.
- **Example:** Merging a list of hospitals that all NY physicians work at ("using"/second) to a list of specialties each physician is certified in (master/original). Physicians can work at multiple hospitals and can have multiple specialties, so neither data set is unique by physician ID. If you tried to do an m:m merge, Stata would likely only keep the first matching record for each. So if we had Dr. Smith certified in Internal Medicine and Cardiology and she worked at NY Presbyterian and Bellevue Hospital, we'd probably get a data set that said:
  - Dr. Smith – Internal Med – NY Pres
  - Dr. Smith – Cardiology – Bellevue

  If you wanted all possible combinations, it is recommended you use "joinby" instead of merge. But even then, m:m situations are rare, and you should ask yourself if the situation really calls for an m:m. Here is what all possible combinations in a joinby would look like:
  - Dr. Smith – Internal Med – NY Pres
  - Dr. Smith – Cardiology – Bellevue
  - Dr. Smith – Cardiology – NY Pres
  - Dr. Smith – Internal Med – Bellevue