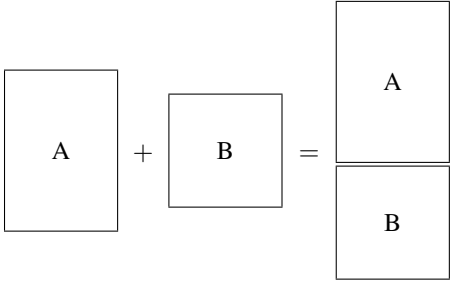


23 Combining datasets

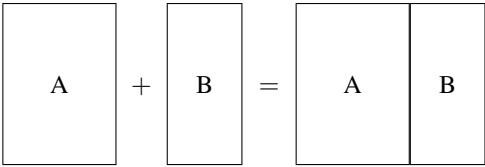
You have two datasets that you wish to combine. Below, we will draw a dataset as a box where, in the box, the variables go across and the observations go down.

See [D] **append** if you want to combine datasets vertically:



append adds observations to the existing variables. That is an oversimplification because **append** does not require that the datasets have the same variables. **append** is appropriate, for instance, when you have data on hospital patients and then receive data on more patients.

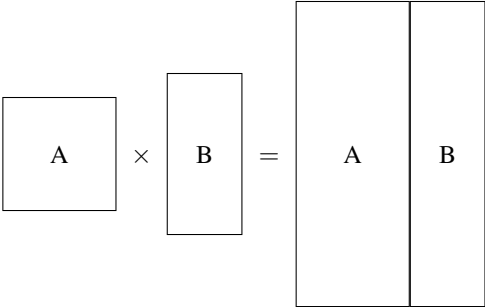
See [D] **merge** if you want to combine datasets horizontally:



merge adds variables to the existing observations. That is an oversimplification because **merge** does not require that the datasets have the same observations. **merge** is appropriate, for instance, when you have data on survey respondents and then receive data on part 2 of the questionnaire.

There is another way to combine datasets horizontally, or more precisely, hierarchically, by loading them into separate frames and linking them. See [D] **frlink** for a discussion of when you might want to use **merge** versus **frlink**.

See [D] **joinby** when you want to combine datasets horizontally but form all pairwise combinations within group:



`joinby` is similar to `merge` but forms all combinations of the observations where it makes sense. `joinby` would be appropriate, for instance, where A contained data on parents and B contained data on their children. `joinby familyid` would form a dataset of each parent joined with each of his or her children.

Also see [D] `cross` for a less frequently used command that forms every pairwise combination of two datasets.

See Mitchell (2020, chap. 7) for more information on combining datasets in Stata.

References

- Golbe, D. L. 2010. *Stata tip 83: Merging multilingual datasets*. *Stata Journal* 10: 152–156.
- Gould, W. W. 2011a. Merging data, part 1: Merges gone bad. *The Stata Blog: Not Elsewhere Classified*. <http://blog.stata.com/2011/04/18/merging-data-part-1-merges-gone-bad/>.
- . 2011b. Merging data, part 2: Multiple-key merges. *The Stata Blog: Not Elsewhere Classified*. <http://blog.stata.com/2011/05/27/merging-data-part-2-multiple-key-merges/>.
- Mitchell, M. N. 2020. *Data Management Using Stata: A Practical Handbook*. 2nd ed. College Station, TX: Stata Press.