

**Data management** — Introduction to data management commands[Description](#)   [References](#)   [Also see](#)

## Description

This manual, called [D], documents Stata's data management features. See [Mitchell \(2020\)](#) for additional information and examples on data management in Stata.

Data management for statistical applications refers not only to classical data management—sorting, merging, appending, and the like—but also to data reorganization because the statistical routines you will use assume that the data are organized in a certain way. For example, statistical commands that analyze longitudinal data, such as `xtreg`, generally require that the data be in long rather than wide form, meaning that repeated values are recorded not as extra variables, but as extra observations.

Here are the basics everyone should know:

[D] <a href="#">use</a>	Load Stata dataset
[D] <a href="#">sysuse</a>	Use shipped dataset
[D] <a href="#">webuse</a>	Use dataset from Stata website
[D] <a href="#">save</a>	Save Stata dataset
[D] <a href="#">describe</a>	Describe data in memory or in a file
[D] <a href="#">codebook</a>	Describe data contents
[D] <a href="#">inspect</a>	Display simple summary of data's attributes
[D] <a href="#">count</a>	Count observations satisfying specified conditions
[D] <a href="#">Data types</a>	Quick reference for data types
[D] <a href="#">Missing values</a>	Quick reference for missing values
[D] <a href="#">Datetime</a>	Date and time values and variables
[D] <a href="#">list</a>	List values of variables
[D] <a href="#">edit</a>	Browse or edit data with Data Editor
[D] <a href="#">varmanage</a>	Manage variable labels, formats, and other properties
[D] <a href="#">rename</a>	Rename variable
[D] <a href="#">format</a>	Set variables' output format
[D] <a href="#">label</a>	Manipulate labels
[D] <a href="#">frames intro</a>	Introduction to frames

To work with multiple datasets in memory, see

[D] <b>frames intro</b>	Introduction to frames
[D] <b>frames</b>	Data frames
[D] <b>frame change</b>	Change identity of current (working) frame
[D] <b>frame copy</b>	Make a copy of a frame
[D] <b>frame create</b>	Create a new frame
[D] <b>frame drop</b>	Drop frames from memory
[D] <b>frame prefix</b>	The frame prefix command
[D] <b>frame put</b>	Copy selected variables or observations to a new frame
[D] <b>frame pwf</b>	Display name of current (working) frame
[D] <b>frame rename</b>	Rename existing frame
[D] <b>frames dir</b>	Display names of all frames in memory
[D] <b>frames reset</b>	Drop all frames from memory
[D] <b>frames save</b>	Save a set of frames on disk
[D] <b>frames use</b>	Load a set of frames from disk
[D] <b>frames describe</b>	Describe frames in memory or in a file
[D] <b>frget</b>	Copy variables from linked frame
[D] <b>frlink</b>	Link frames

You will need to create and drop variables, and here is how:

[D] <b>generate</b>	Create or change contents of variable
[D] <b>egen</b>	Extensions to generate
[D] <b>drop</b>	Drop variables or observations
[D] <b>clear</b>	Clear memory

For inputting or importing data, see

[D] <b>use</b>	Load Stata dataset
[D] <b>sysuse</b>	Use shipped dataset
[D] <b>webuse</b>	Use dataset from Stata website
[D] <b>input</b>	Enter data from keyboard
[D] <b>import</b>	Overview of importing data into Stata
[D] <b>import dbase</b>	Import and export dBase files
[D] <b>import delimited</b>	Import and export delimited text data
[D] <b>import excel</b>	Import and export Excel files
[D] <b>import fred</b>	Import data from Federal Reserve Economic Data
[D] <b>import haver</b>	Import data from Haver Analytics databases
[D] <b>import sas</b>	Import SAS files
[D] <b>import sasxport5</b>	Import and export data in SAS XPORT Version 5 format
[D] <b>import sasxport8</b>	Import and export data in SAS XPORT Version 8 format
[D] <b>import spss</b>	Import and export SPSS files
[D] <b>infile (fixed format)</b>	Import text data in fixed format with a dictionary
[D] <b>infile (free format)</b>	Import unformatted text data
[D] <b>infix (fixed format)</b>	Import text data in fixed format
[D] <b>jdbc</b>	Load, write, or view data from a database with a Java API
[D] <b>odbc</b>	Load, write, or view data from ODBC sources
[D] <b>hexdump</b>	Display hexadecimal report on file
[D] <b>icd9</b>	ICD-9-CM diagnosis codes
[D] <b>icd9p</b>	ICD-9-CM procedure codes
[D] <b>icd10</b>	ICD-10 diagnosis codes
[D] <b>icd10cm</b>	ICD-10-CM diagnosis codes
[D] <b>icd10pcs</b>	ICD-10-PCS procedure codes

and for exporting data, see

[D] <b>save</b>	Save Stata dataset
[D] <b>export</b>	Overview of exporting data from Stata
[D] <b>outfile</b>	Export dataset in text format
[D] <b>import dbase</b>	Import and export dBase files
[D] <b>import delimited</b>	Import and export delimited text data
[D] <b>import excel</b>	Import and export Excel files
[D] <b>import sasxport5</b>	Import and export data in SAS XPORT Version 5 format
[D] <b>import sasxport8</b>	Import and export data in SAS XPORT Version 8 format
[D] <b>import spss</b>	Import and export SPSS files
[D] <b>jdbc</b>	Load, write, or view data from a database with a Java API
[D] <b>odbc</b>	Load, write, or view data from ODBC sources

The ordering of variables and observations (sort order) can be important; see

[D] <b>order</b>	Reorder variables in dataset
[D] <b>sort</b>	Sort data
[D] <b>gsort</b>	Ascending and descending sort

To reorganize or combine data, see

[D] <b>append</b>	Append datasets
[D] <b>merge</b>	Merge datasets
[D] <b>frlink</b>	Link frames
[D] <b>frget</b>	Copy variables from linked frame
[D] <b>reshape</b>	Convert data from wide to long form and vice versa
[D] <b>collapse</b>	Make dataset of summary statistics
[D] <b>contract</b>	Make dataset of frequencies and percentages
[D] <b>fillin</b>	Rectangularize dataset
[D] <b>expand</b>	Duplicate observations
[D] <b>expandcl</b>	Duplicate clustered observations
[D] <b>stack</b>	Stack data
[D] <b>joinby</b>	Form all pairwise combinations within groups
[D] <b>xpose</b>	Interchange observations and variables
[D] <b>cross</b>	Form every pairwise combination of two datasets

In the above list, we particularly want to direct your attention to [D] **reshape**, a useful command that beginners often overlook.

For random sampling, see

[D] <b>sample</b>	Draw random sample
[D] <b>splitsample</b>	Split data into random samples
[D] <b>drawnorm</b>	Draw sample from multivariate normal distribution

For file manipulation, see

[D] <b>type</b>	Display contents of a file
[D] <b>erase</b>	Erase a disk file
[D] <b>copy</b>	Copy file from disk or URL
[D] <b>cd</b>	Change directory
[D] <b>dir</b>	Display filenames
[D] <b>mkdir</b>	Create directory
[D] <b>rmdir</b>	Remove directory
[D] <b>cf</b>	Compare two datasets
[D] <b>changeool</b>	Convert end-of-line characters of text file
[D] <b>filefilter</b>	Convert ASCII or binary patterns in a file
[D] <b>checksum</b>	Calculate checksum of file
[D] <b>zipfile</b>	Compress and uncompress files and directories in zip archive format

For handling Unicode strings, see

[D] <b>unicode</b>	Unicode utilities
[D] <b>unicode translate</b>	Translate files to Unicode
[D] <b>unicode encoding</b>	Unicode encoding utilities
[D] <b>unicode locale</b>	Unicode locale utilities
[D] <b>unicode collator</b>	Language-specific Unicode collators
[D] <b>unicode convertfile</b>	Low-level file conversion between encoding

The entries above are important. The rest are useful when you need them:

[D] <b>datasignature</b>	Determine whether data have changed
[D] <b>type</b>	Display contents of a file
[D] <b>notes</b>	Place notes in data
[D] <b>label language</b>	Labels for variables and values in multiple languages
[D] <b>labelbook</b>	Label utilities
[D] <b>encode</b>	Encode string into numeric and vice versa
[D] <b>recode</b>	Recode categorical variables
[D] <b>ipolate</b>	Linearly interpolate (extrapolate) values
[D] <b>destring</b>	Convert string variables to numeric variables and vice versa
[D] <b>mvencode</b>	Change missing values to numeric values and vice versa
[D] <b>pctile</b>	Create variable containing percentiles
[D] <b>range</b>	Generate numerical range
[D] <b>by</b>	Repeat Stata command on subsets of the data
[D] <b>statsby</b>	Collect statistics for a command across a by list
[D] <b>dyngen</b>	Dynamically generate new values of variables
[D] <b>compress</b>	Compress data in memory
[D] <b>recast</b>	Change storage type of variable
[D] <b>Datetime display formats</b>	Display formats for dates and times
[D] <b>Datetime conversion</b>	String to numeric date conversion functions
[D] <b>Datetime durations</b>	Obtaining and working with durations
[D] <b>Datetime relative dates</b>	Datetime relative dates
[D] <b>Datetime values from other software</b>	Date and time conversion from other software
[D] <b>bcal</b>	Business calendar file manipulation
[D] <b>Datetime business calendars</b>	Business calendars
[D] <b>Datetime business calendars creation</b>	Business calendars creation

[D] <b>assert</b>	Verify truth of claim
[D] <b>assertnested</b>	Verify variables nested
[D] <b>clonevar</b>	Clone existing variable
[D] <b>compare</b>	Compare two variables
[D] <b>corr2data</b>	Create dataset with specified correlation structure
[D] <b>ds</b>	Compactly list variables with specified properties
[D] <b>duplicates</b>	Report, tag, or drop duplicate observations
[D] <b>insobs</b>	Add or insert observations
[D] <b>isid</b>	Check for unique identifiers
[D] <b>lookfor</b>	Search for string in variable names and labels
[D] <b>memory</b>	Memory management
[D] <b>putmata</b>	Put Stata variables into Mata and vice versa
[D] <b>obs</b>	Increase the number of observations in a dataset
[D] <b>rename group</b>	Rename groups of variables
[D] <b>separate</b>	Create separate variables
[D] <b>shell</b>	Temporarily invoke operating system
[D] <b>snapshot</b>	Save and restore data snapshots
[D] <b>split</b>	Split string variables into parts
[D] <b>vl</b>	Manage variable lists
[D] <b>vl create</b>	Create and modify user-defined variable lists
[D] <b>vl drop</b>	Drop variable lists or variables from variable lists
[D] <b>vl list</b>	List contents of variable lists
[D] <b>vl rebuild</b>	Rebuild variable lists
[D] <b>vl set</b>	Set system-defined variable lists

There are some real jewels in the above, such as [D] **notes**, [D] **compress**, and [D] **assert**, which you will find particularly useful.

## References

- Hoffmann, J. P. 2017. *Principles of Data Management and Presentation*. Oakland, CA: University of California Press.
- Mitchell, M. N. 2020. *Data Management Using Stata: A Practical Handbook*. 2nd ed. College Station, TX: Stata Press.

## Also see

[D] **Intro** — Introduction to data management reference manual

[R] **Intro** — Introduction to base reference manual

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).