12 Deleting variables and observations

clear, drop, and keep

In this chapter, we will present the tools for paring observations and variables from a dataset. We saw how to do this using the Data Editor in [GSW] **6 Using the Data Editor**; this chapter presents the methods for doing so from the Command window.

There are three main commands for removing data and other Stata objects, such as value labels, from memory: `clear`, `drop`, and `keep`. Remember that they affect only what is in memory. None of these commands alter anything that has been saved to disk.

clear and drop _all

Suppose that you are working on an analysis or a simulation and that you need to clear out Stata's memory so that you can impute different values or simulate a new dataset. You are not interested in saving any of the changes you have made to the dataset in memory—you would just like to have an empty dataset. What you do depends on how much you want to clear out: at any time, you can have not only data but also metadata such as value labels, stored results from previous commands, and stored matrices. The `clear` command will let you carefully clear out data or other objects; we are interested only in simple usage here. For more information, see `help clear` and [D] **clear**.

If you type the command `clear` into the Command window, it will remove all variables and value labels. In basic usage, this is typically enough. It has the nice property that it does not remove any stored results, so you can load a new dataset and predict values by using stored estimation results from a model fit on a previous dataset. See `help postest` and [U] **20 Estimation and postestimation commands** for more information.

If you want to be sure that everything is cleared out, use the command `clear all`. This command will clear Stata's memory of data and all auxiliary objects so that you can start with a clean slate. The first time you use `clear all` while you have a graph or dialog open, you may be surprised when that graph or dialog closes; this is necessary so that Stata can free all memory being used.

If you want to get rid of just the data and nothing else, you can use the command `drop _all`.

drop

The `drop` command is used to remove variables or observations from the dataset in memory.
- If you want to drop variables, use `drop` *varlist*.
- If you want to drop observations, use `drop` with an `if` or an `in` qualifier or both.

We will use the `afewcarslab` dataset to illustrate `drop`:

```
. use afewcarslab
(A few 1978 cars)
. list
```

	make	price	mpg	weight	gear_r~o	foreign
1.	VW Rabbit	4697	25	1930	3.78	Foreign
2.	Olds 98	8814	21	4060	2.41	Domestic
3.	Chev. Monza	3667	.	2750	2.73	Domestic
4.		4099	22	2930	3.58	Domestic
5.	Datsun 510	5079	24	2280	3.54	Foreign
6.	Buick Regal	5189	20	3280	2.93	Domestic
7.	Datsun 810	8129	.	2750	3.55	Foreign

```
. drop in 1/3
(3 observations deleted)
. list
```

	make	price	mpg	weight	gear_r~o	foreign
1.		4099	22	2930	3.58	Domestic
2.	Datsun 510	5079	24	2280	3.54	Foreign
3.	Buick Regal	5189	20	3280	2.93	Domestic
4.	Datsun 810	8129	.	2750	3.55	Foreign

```
. drop if mpg > 21
(3 observations deleted)
. list
```

	make	price	mpg	weight	gear_r~o	foreign
1.	Buick Regal	5189	20	3280	2.93	Domestic

```
. drop gear_ratio
. list
```

	make	price	mpg	weight	foreign
1.	Buick Regal	5189	20	3280	Domestic

```
. drop m*
. list
```

	price	weight	foreign
1.	5189	3280	Domestic

These changes are only to the data in memory. If you want to make the changes permanent, you need to save the dataset.

keep

keep tells Stata to drop all variables except those specified explicitly or through the use of an `if` or `in` expression. Just like `drop`, `keep` can be used with *varlist* or with qualifiers but not with both at once. We use a `clear` command at the start of this example so that we can reload the `afewcarslab` dataset:

```
. clear
. use afewcarslab
(A few 1978 cars)
. list
```

	make	price	mpg	weight	gear_r~o	foreign
1.	VW Rabbit	4697	25	1930	3.78	Foreign
2.	Olds 98	8814	21	4060	2.41	Domestic
3.	Chev. Monza	3667	.	2750	2.73	Domestic
4.		4099	22	2930	3.58	Domestic
5.	Datsun 510	5079	24	2280	3.54	Foreign
6.	Buick Regal	5189	20	3280	2.93	Domestic
7.	Datsun 810	8129	.	2750	3.55	Foreign

```
. keep in 4/7
(3 observations deleted)
. list
```

	make	price	mpg	weight	gear_r~o	foreign
1.		4099	22	2930	3.58	Domestic
2.	Datsun 510	5079	24	2280	3.54	Foreign
3.	Buick Regal	5189	20	3280	2.93	Domestic
4.	Datsun 810	8129	.	2750	3.55	Foreign

```
. keep if mpg <= 21
(3 observations deleted)
. list
```

	make	price	mpg	weight	gear_r~o	foreign
1.	Buick Regal	5189	20	3280	2.93	Domestic

```
. keep m*
. list
```

	make	mpg
1.	Buick Regal	20

For suggested citations, see the FAQ on citing Stata documentation.