

## Data Exploration by Using Graphics:

Once you bring your data in STATA the next step is getting comfortable with your data.

### Example:

We will begin by loading `hs0.dta`, a dataset saved in Stata's format. Stata data files end with the `dta` extension. Stata data files are loaded into memory using the `use` command.

```
. use http://www.ats.ucla.edu/stat/data/hs0, clear
```

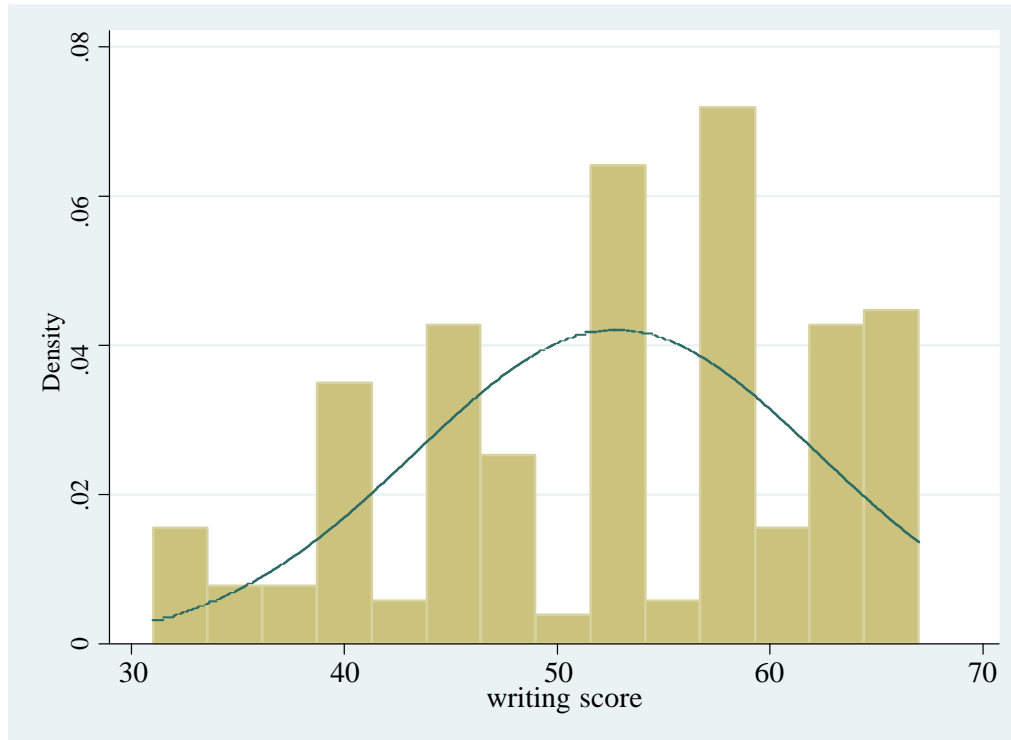
In this file we will use some graphics commands to look at our data. We will begin with `stem` which generates an ASCII stem-and-leaf plot. We will also use the `graph` command with the `hist` (histogram) and `box` (boxplot) options. We also show the `kdensity` command which produces a smoothed density plot.

#### stem-and-leaf plot:

```
stem-and-leaf plot for write (writing score)
3*  | 1111
3t  | 3333
3f  | 55
3s  | 66777
3.  | 899999
4*  | 00011111111111
4t  | 223
4f  | 44444444444445
4s  | 66666666677
4.  | 99999999999
5*  | 00
5t  | 222222222222223
5f  | 4444444444444444555
5s  | 77777777777
5.  | 9999999999999999999999999999999
6*  | 00001111
6t  | 222222222222222222223333
6f  | 5555555555555555
6s  | 7777777
```

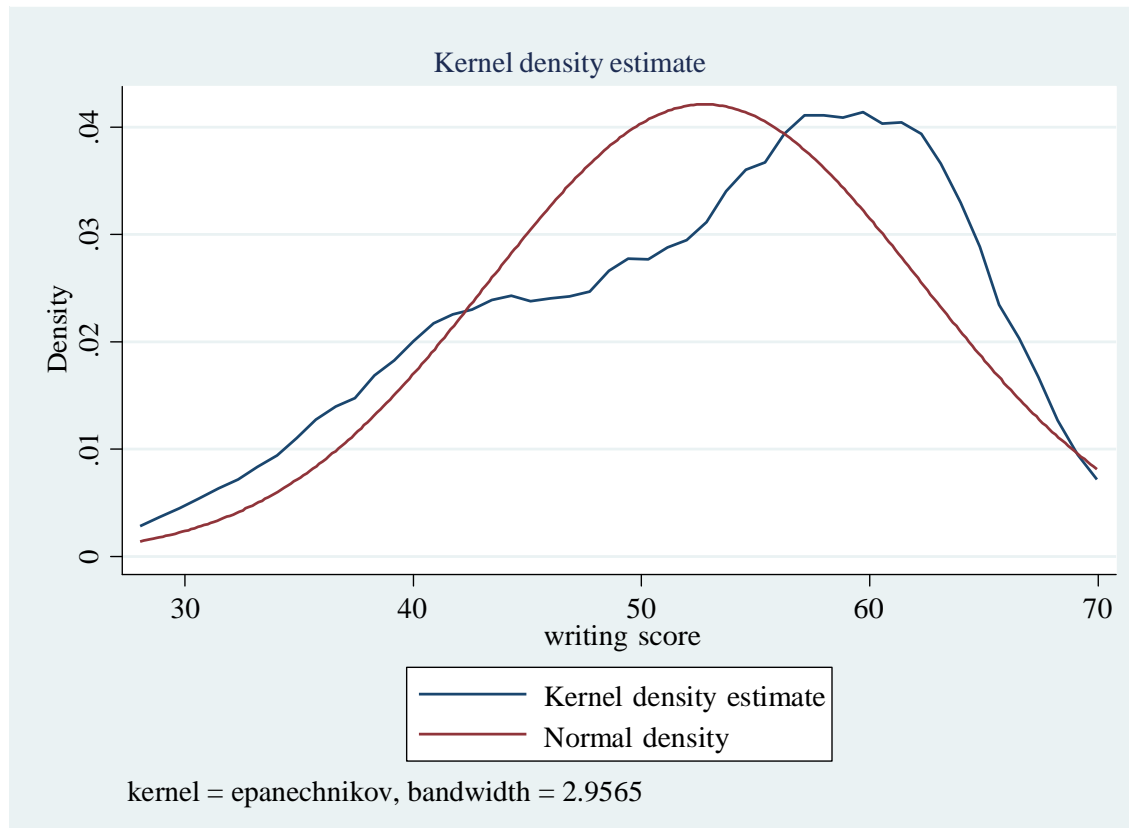
**Histogram:**

Type `histogram write, normal` in the command window to get the histogram for each individual variables in the data.



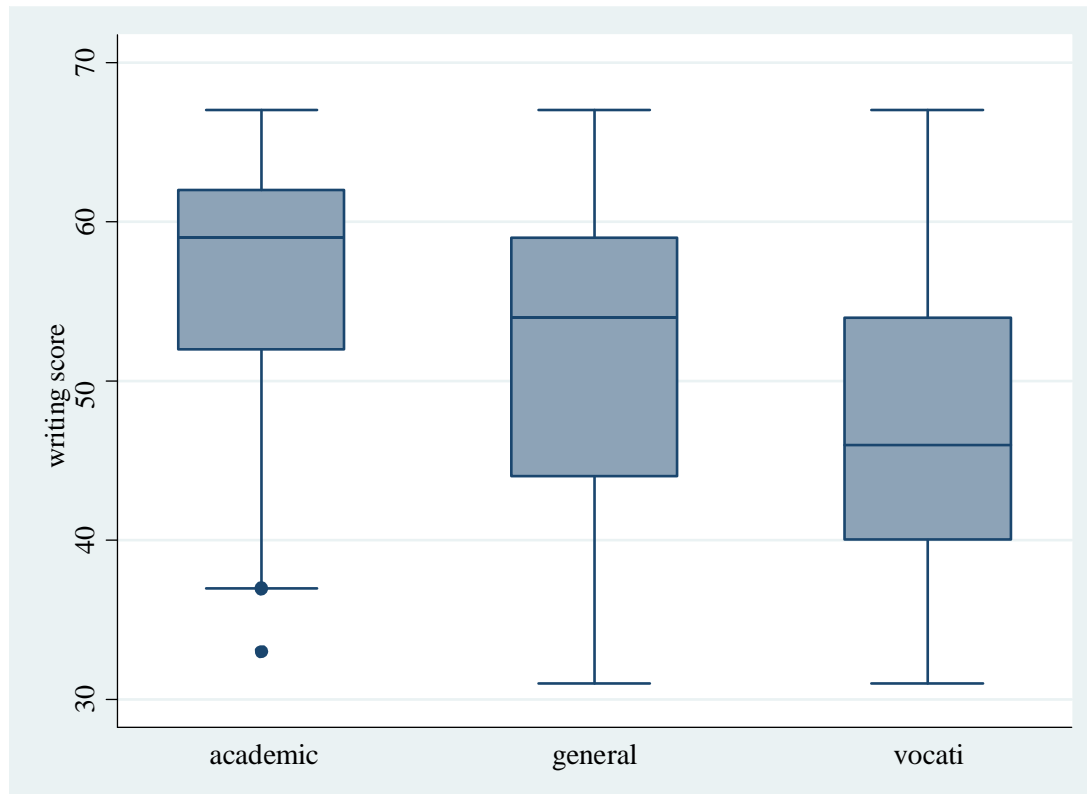
**smoothed density plot:**

Type `kdensity write, normal` in the command window to produce a smoothed density plot for each individual variable in the data.



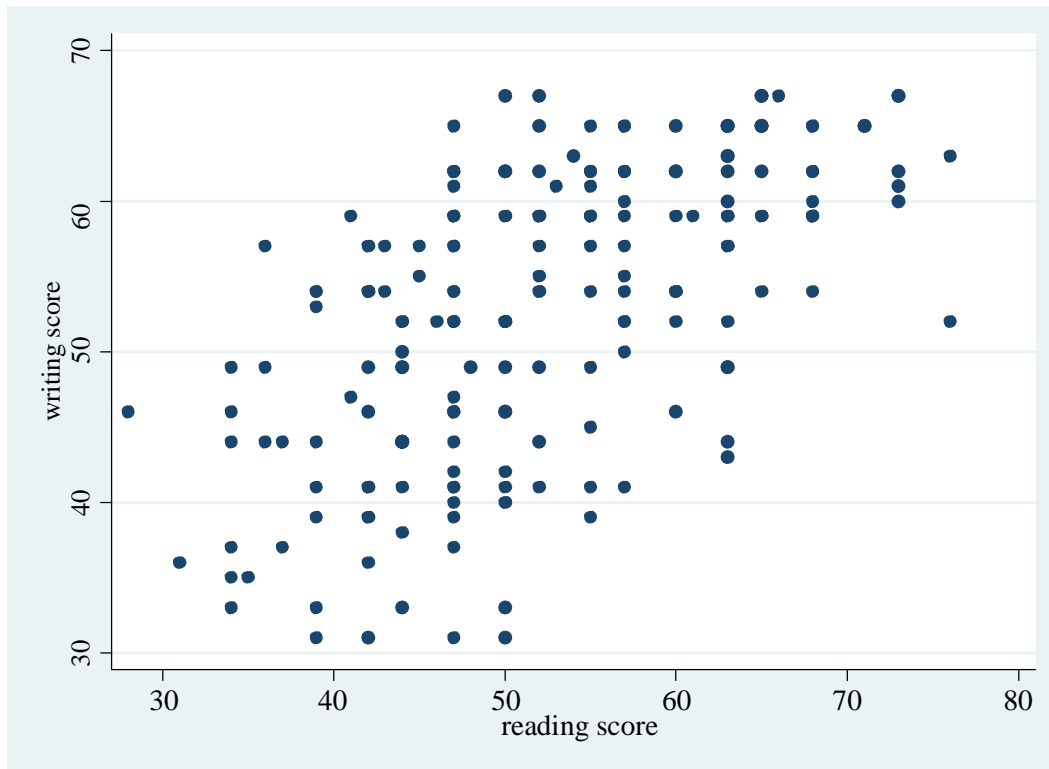
**Box plot:**

Type `graph box write, over(prgtype)` in the command window to produce the box plot for the writing score of students by programs.



**Scatter plot:**

Type `scatter write read` in the command window to produce the scatter plot between reading and writing scores of students. A scatter plot is good for exploring if there is an association between the two variables.



The code relevant to the above discussed methods is as follows:

```
use http://www.ats.ucla.edu/stat/data/hs0, clear
describe
codebook

capture log close
log using example1.txt, text replace

summarize read math science write
tabstat read write math, by(prgtype) stat(n mean sd)
tabstat write, by(prgtype) stat(n mean sd p25 p50 p75)

// # 3 Exploration by graphics

stem write
histogram write, normal
kdensity write, normal
graph box write, over(prgtype)
scatter write read

log close
```