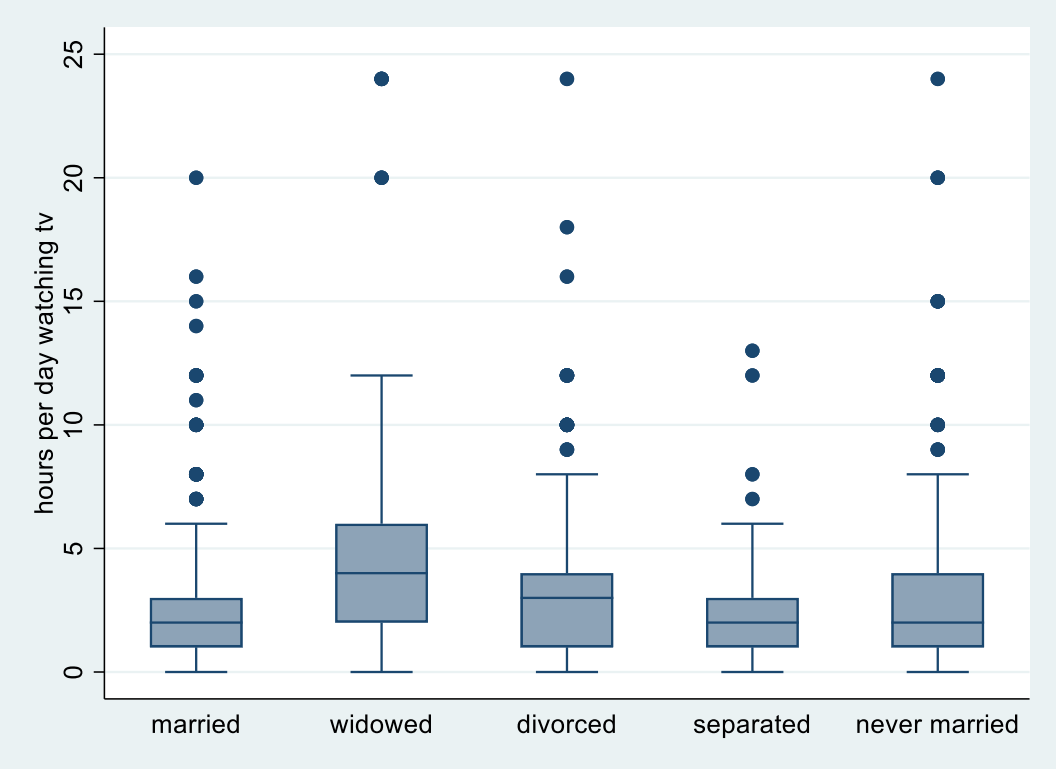
Revised July 2018

Boxplot Comparison

This handout explains to use Stata to construct a chart that compares boxplots of a quantitative variable across two or more groups. It assumes that you have set Stata up on your computer (see the "Getting Started with Stata" handout), and that you have read in the set of data that you want to analyze (see the "Reading in Stata Format (.dta) Data Files" handout).

In Stata, most tasks can be performed either by issuing commands within the "Stata command" window, **or** by using the menus. These notes illustrate both approaches, using the data file "GSS2016.DTA" (this data file is posted here: https://canvas.harvard.edu/courses/53958).

The process for constructing a boxplot comparison parallels that for constructing a bar chart. We illustrate with a graph on differences in television viewing (hours per day) among persons of different marital status. The completed graph looks like this:



Simple boxplot comparisons can be constructed easily via the Stata command line. The general command structure is

```
graph box <varname1>, over(<varname2>)
```

where you enter a quantitative variable in place of "varname1" and a categorical variable in place of "varname2".

Using the variables in the above graph, we would issue the command

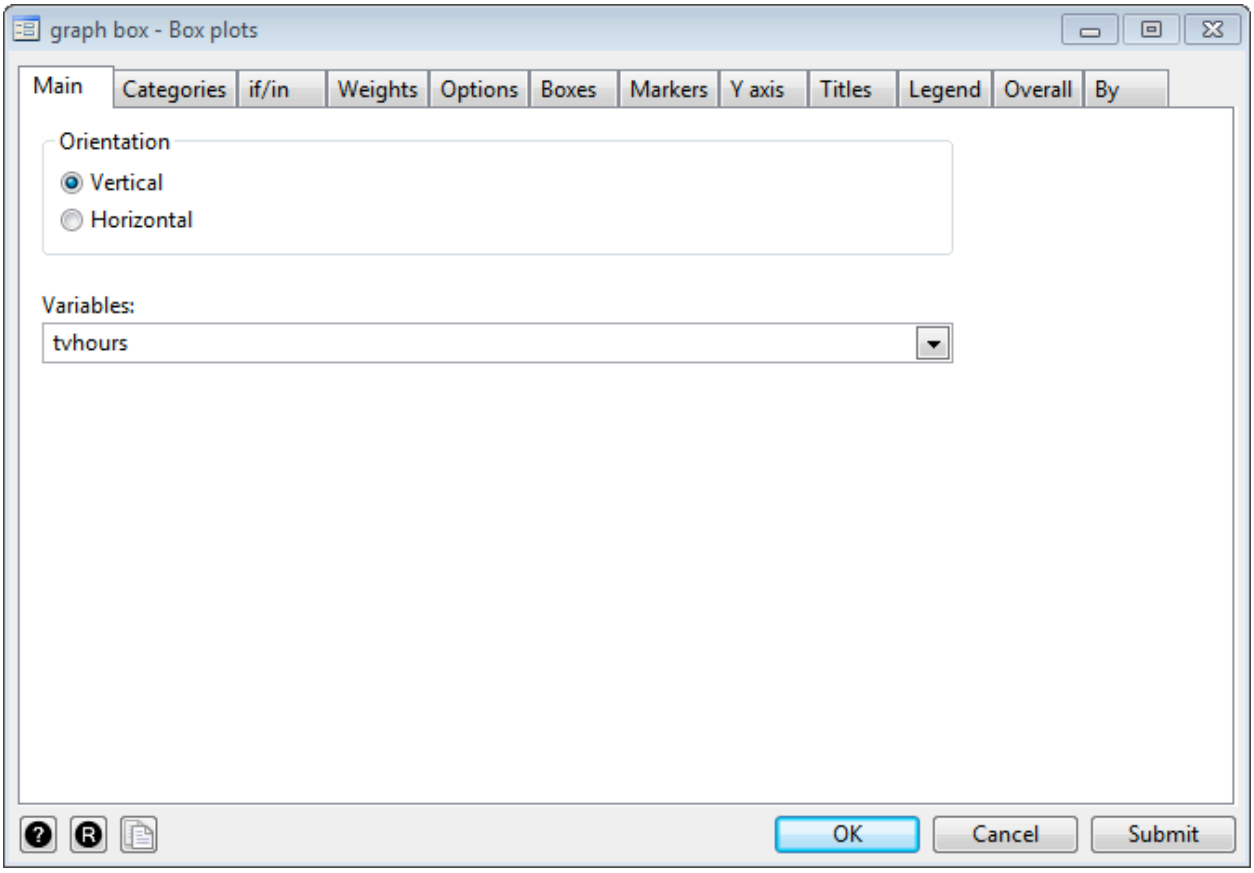
```
graph box tvhours, over(marital)
```

This command produces a bare-bones graph with no enhancements.

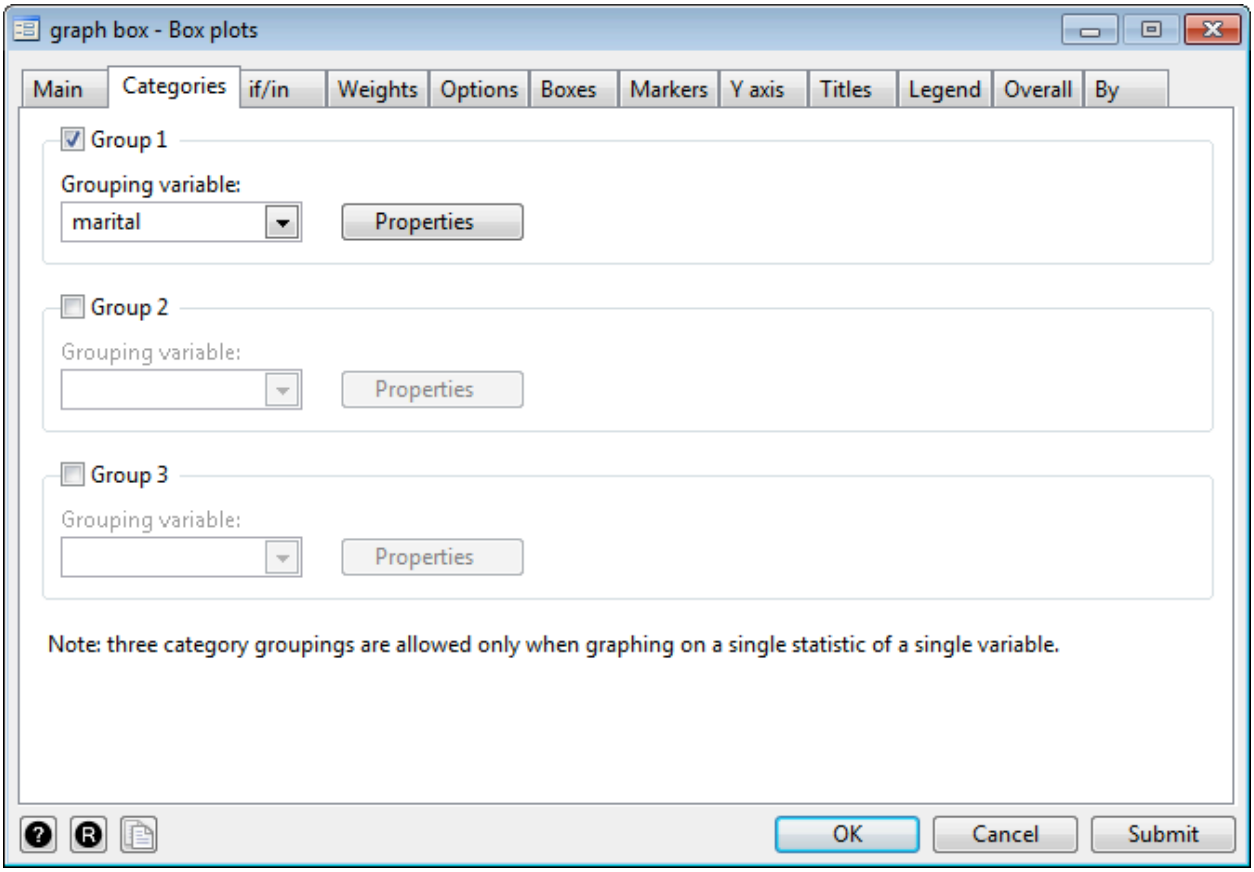
To enhance and improve the display, it is probably easier to use the Stata menus. For a box plot,

Click on “Graphics”
Click on “Box plot”

When you do that, the following window opens up:



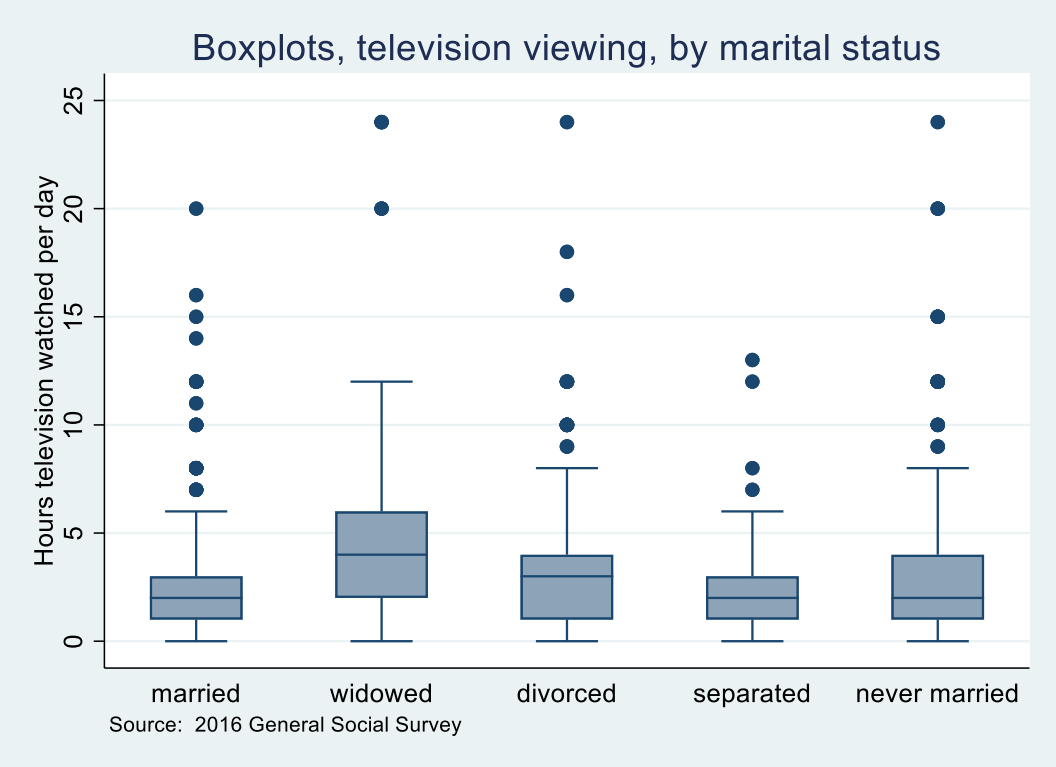
Enter the name of your quantitative variable in the “Variable(s):” box here, and then click on the “Categories” tab to open the following window:



Click the "Group 1" box and then fill in the name of your categorical (group) variable in the "Grouping Variable:" box. Clicking "OK" at this point will produce a basic boxplot comparison like the one shown earlier. Or, you can click "Submit" to display the current graph, and then decide whether you want to make further modifications or enhancements using other tabs.

Available enhancements include the possibility of adding a title, subtitle, caption and/or via the "Titles" tab. You can use the "Y axis" tab to change the labeling of the vertical axis. There are many other choices, too, but the basic display shown earlier will often be sufficient. For more information about these enhancements, see the "Bar Charts" handout.

To illustrate the use of these, the same boxplot comparison is shown on the following page, with a title (at top) and a caption (bottom) added, and with a different title for the vertical (Y) axis.



Saving Graphs

To save your boxplot(s), see the "Saving Graphs" handout.