

# How to Easily Generate and Customize Box Plots in Stata

Authored by  
**stats writer**

December 28, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Generate and Customize Box Plots in Stata*.  
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=109571>

The creation and customization of box plots in the statistical software package Stata is a fundamental skill for effective data visualization and exploratory data analysis. Box plots offer a powerful, yet concise, visual summary of the distribution of a continuous variable, highlighting key statistical measures and potential outliers. To generate these plots, Stata utilizes the straightforward `graph box` command, which requires specifying the variables and any desired options. Subsequent modifications to the plot's aesthetics, such as size, color, and labeling, can be achieved by appending various options to the initial command.

A box plot, often referred to as a box-and-whisker plot, is a graphical method designed to display the distribution of data based on the five number summary. This summary provides a standardized way to compare distributions across different datasets or groups, offering insights into central tendency, spread, and skewness.

The five number summary includes the following statistical measures:

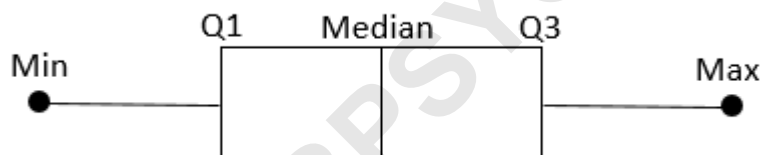
The minimum value (the smallest observation, excluding defined outliers).

The first quartile ( $Q_1$ , representing the 25th percentile).

The median ( $Q_2$ , representing the 50th percentile).

The third quartile ( $Q_3$ , representing the 75th percentile).

The maximum value (the largest observation, excluding defined outliers).



This comprehensive tutorial outlines the practical steps necessary to create various types of box plots and apply detailed modifications to enhance their presentation quality in Stata.

### Example Setup: Utilizing the Auto Dataset

To demonstrate the versatility of box plots in Stata, we will use the commonly referenced built-in dataset, `auto`. This dataset contains information on 74 cars from 1978 and is ideal for illustrating single-variable and comparative visualizations. Before generating any plots, the data must be loaded into the memory.

Load the data by typing the following command into the Stata Command window and executing it:

use <http://www.stata-press.com/data/r13/auto>

Once the dataset is loaded, we can proceed with generating our initial visualizations based on variables such as `mpg` (miles per gallon).

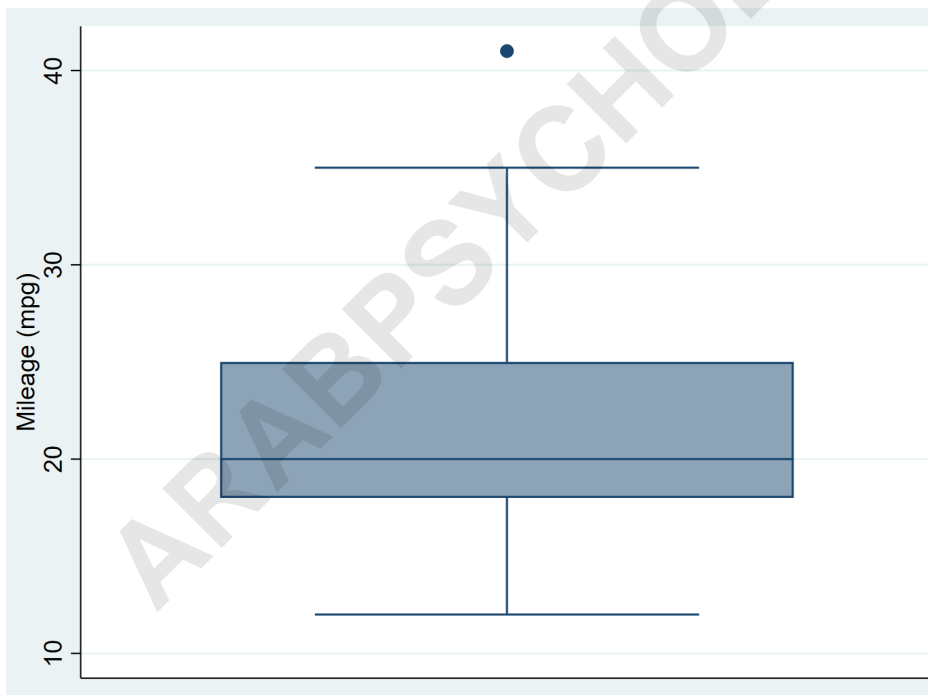
## Vertical Box Plots: The Standard Visualization

The standard box plot generated in Stata is oriented vertically, displaying the scale on the Y-axis. This orientation is generally preferred for simple distribution checks. To create a vertical box plot for the variable `mpg`, we use the core command `graph box` followed by the variable name:

We can create a vertical box plot for the variable `mpg` by using the **graph box** command:

### **graph box mpg**

This command generates a plot that immediately visualizes the central tendency (median line) and the dispersion (interquartile range, or the box itself) of the mileage data, clearly identifying any potential outliers that fall outside the whisker range.



## Horizontal Box Plots: An Alternative Orientation

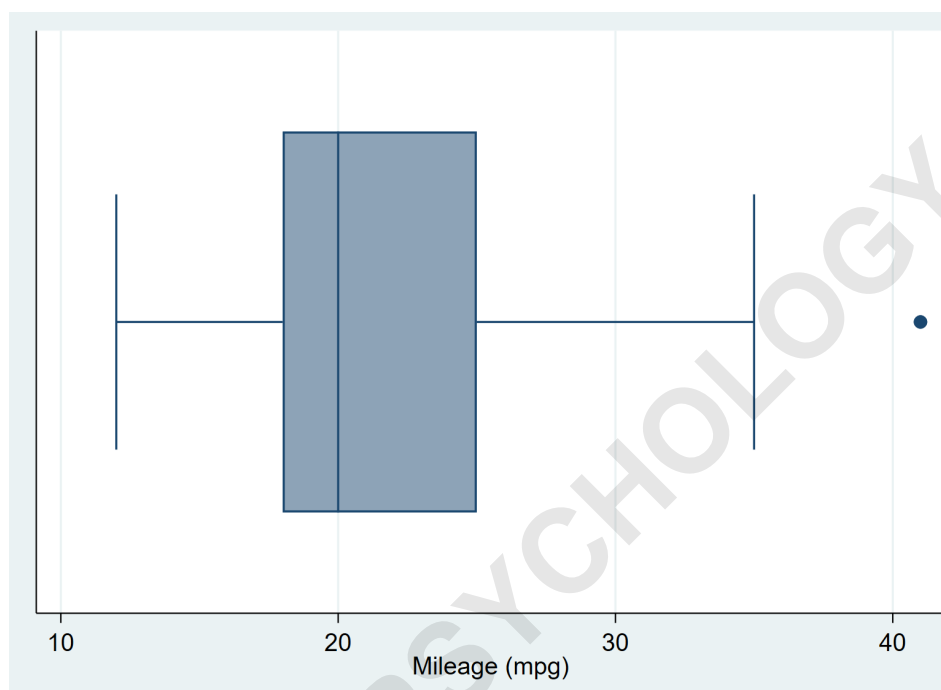
In certain contexts, particularly when comparing many groups or when space constraints require a different layout, a horizontal orientation may be more suitable. Stata provides a specific command, `graph hbox`, to easily switch the plot's orientation. This results in the variable scale being

displayed along the X-axis.

To generate a horizontal box plot for the `mpg` variable, execute the following command:

```
graph hbox mpg
```

The statistical information conveyed by the horizontal plot remains identical to the vertical version; only the visual presentation is altered. Choosing between vertical and horizontal plots often comes down to personal preference or specific publication requirements.



## Comparative Box Plots by Categorical Variable

One of the most valuable applications of the `box plot` is comparing the distributions of a continuous variable across different groups defined by a categorical variable. In the `auto` dataset, we can compare the distribution of `mpg` between domestic and foreign cars using the `over()` option.

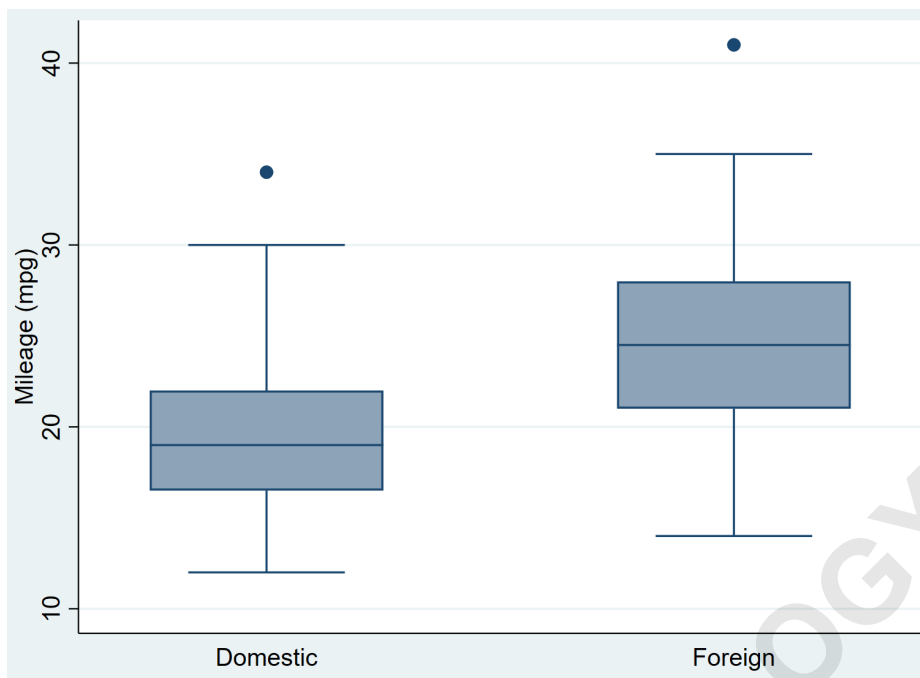
The `over()` option instructs Stata to generate separate box plots for each level within the specified grouping variable. This technique is highly effective for quickly identifying differences in median, variance, and symmetry between groups.

To compare the `mpg` distribution based on the `foreign` variable, use the following code:

```
graph box mpg, over(foreign)
```

This command produces two side-by-side box plots, allowing for a direct visual comparison of fuel

efficiency distributions between the two car origins.



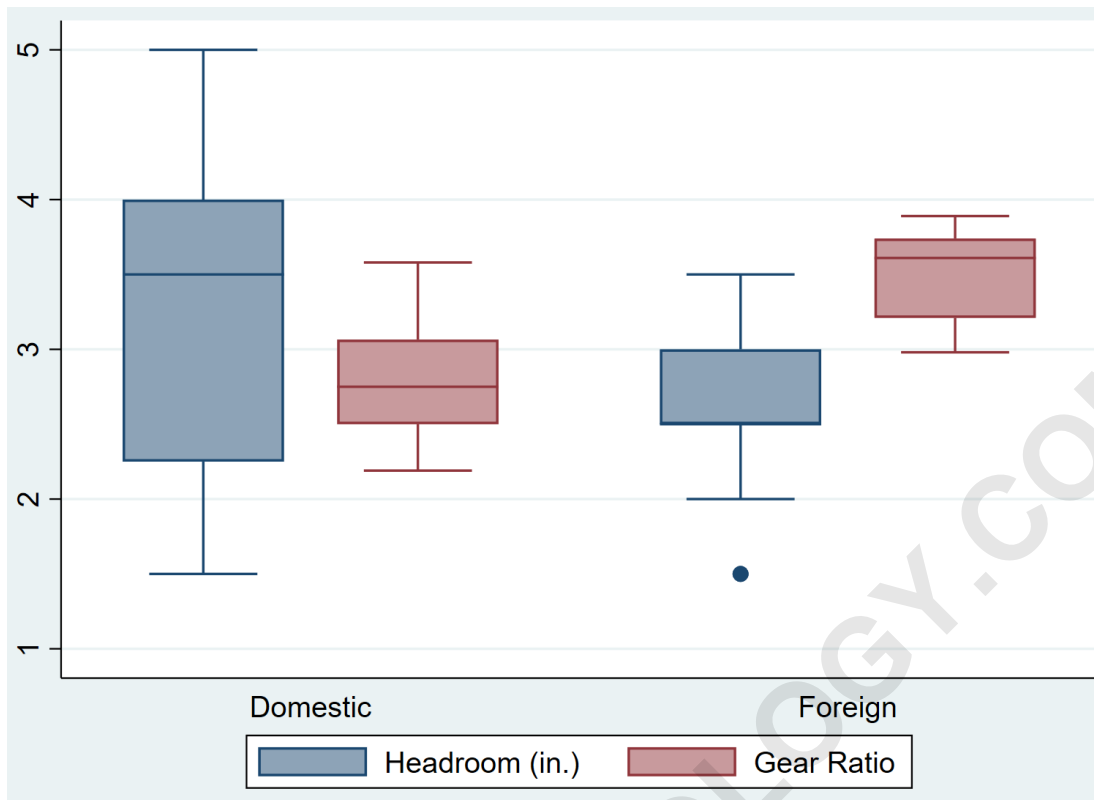
## Visualizing Multiple Variables Categorically

Beyond comparing a single variable across categories, Stata enables users to compare the distributions of multiple continuous variables simultaneously, while still grouping them by a single categorical variable. This provides a comprehensive overview of how several key metrics vary across defined populations.

To illustrate, we can create box plots for both `headroom` and `gear_ratio`, grouped by the `foreign` variable. Simply list the continuous variables before the comma, followed by the `over()` option:

```
graph box headroom gear_ratio, over(foreign)
```

The resulting graph displays four individual box plots (two variables times two categories), organized logically for easy interpretation. This is particularly useful in multivariate analysis where researchers need to assess how different structural features compare across subgroups.



## Enhancing Plot Aesthetics and Readability

While the default `Stata` output is statistically sound, customizing the appearance of box plots is essential for generating professional-grade reports and presentations. `Stata` offers a robust set of options to modify titles, subtitles, notes, and element colors, significantly improving the plot's communicative power.

The most crucial modification involves adding descriptive text elements. Titles, subtitles, and notes provide context, methodology details, and source information, ensuring the visualization stands alone as an informative piece of content. These options are appended directly to the main `graph box` command.

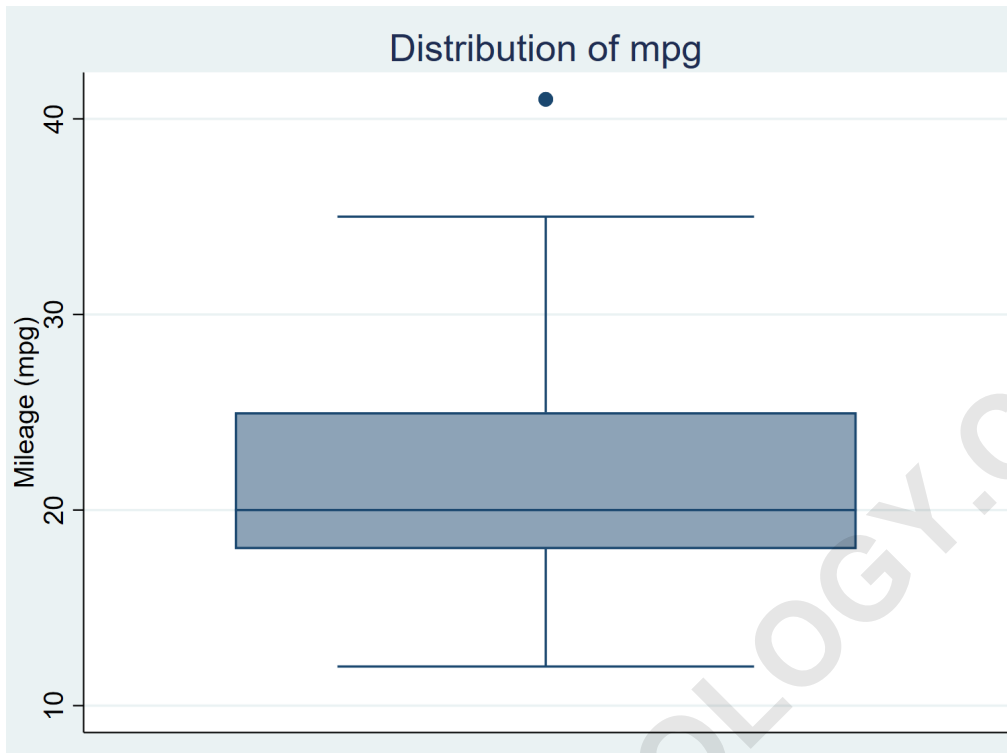
We begin by focusing on descriptive text elements, followed by visual modifications such as color adjustments.

## Adding Descriptive Titles and Notes

Graph titles are indispensable for clearly stating the visualization's purpose. We use the `title()` option to place a primary title above the plot area.

To add a title describing the distribution of `mpg`:

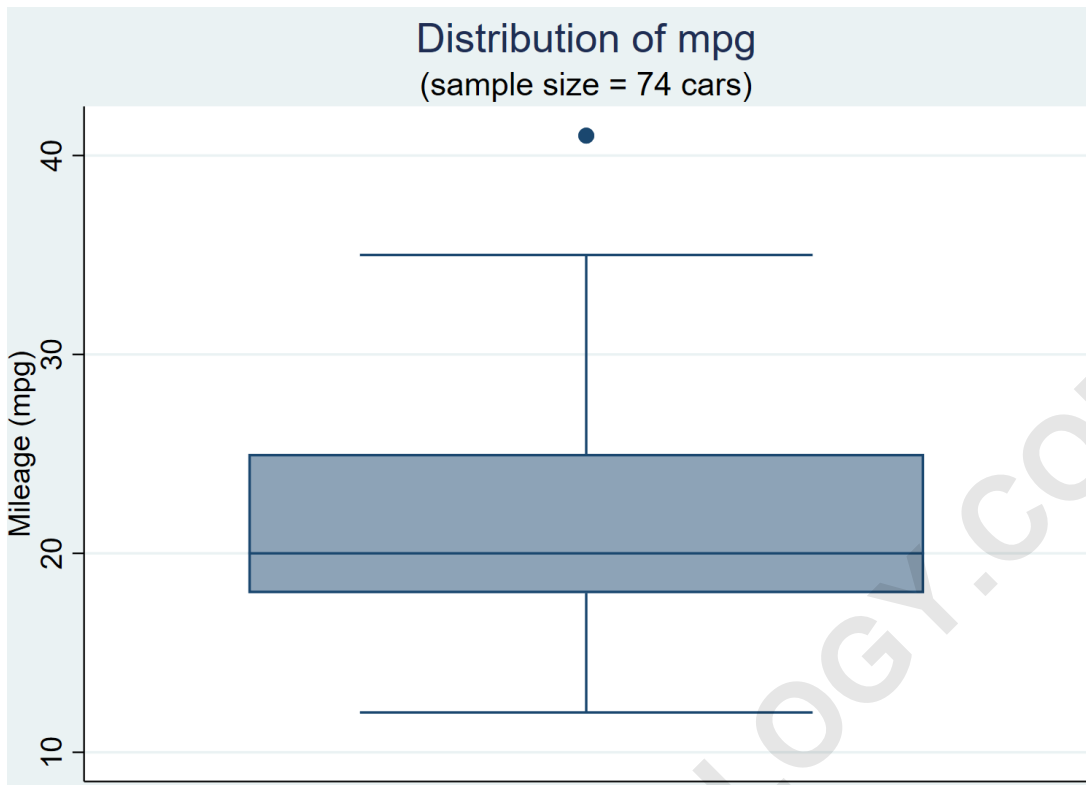
```
graph box mpg, title("Distribution of mpg")
```



For additional details or secondary information that supports the main title--such as the sample size or calculation method--the `subtitle()` option can be employed. The subtitle appears immediately below the main title, offering hierarchical organization of textual information.

To include the sample size as a subtitle:

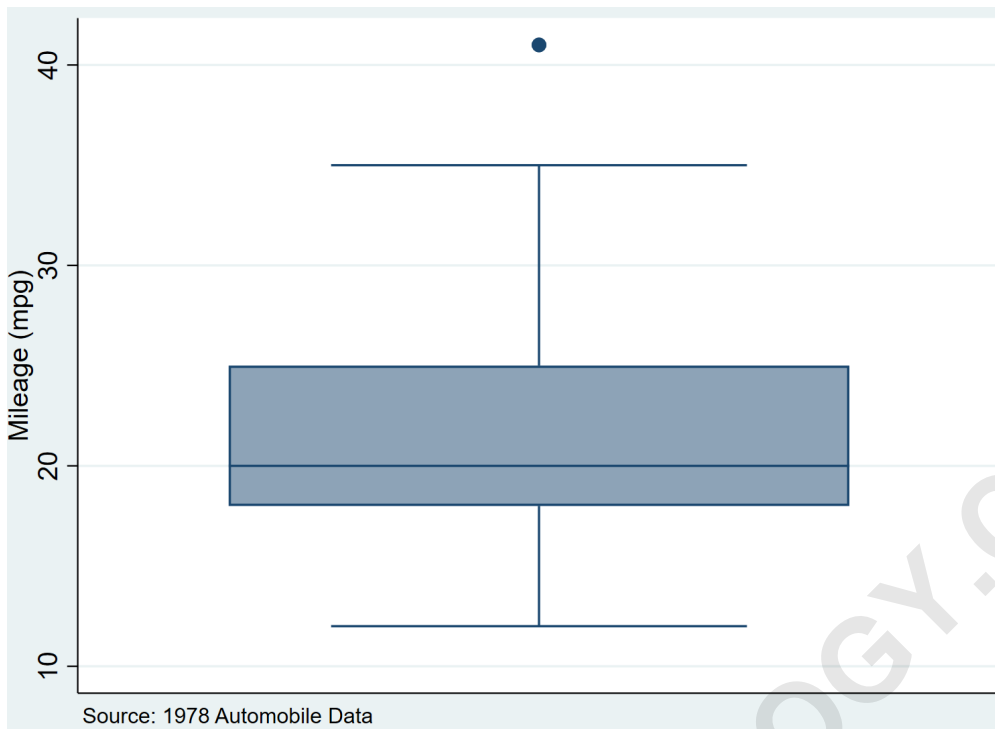
```
graph box mpg, title("Distribution of mpg") subtitle("(sample size = 74 cars)")
```



Finally, for essential source documentation or methodological comments, the `note()` command places text at the bottom of the graph. This is standard practice in academic and professional reporting to ensure transparency and traceability of the data source.

To add a source note to the plot:

```
graph box mpg, note("Source: 1978 Automobile Data")
```



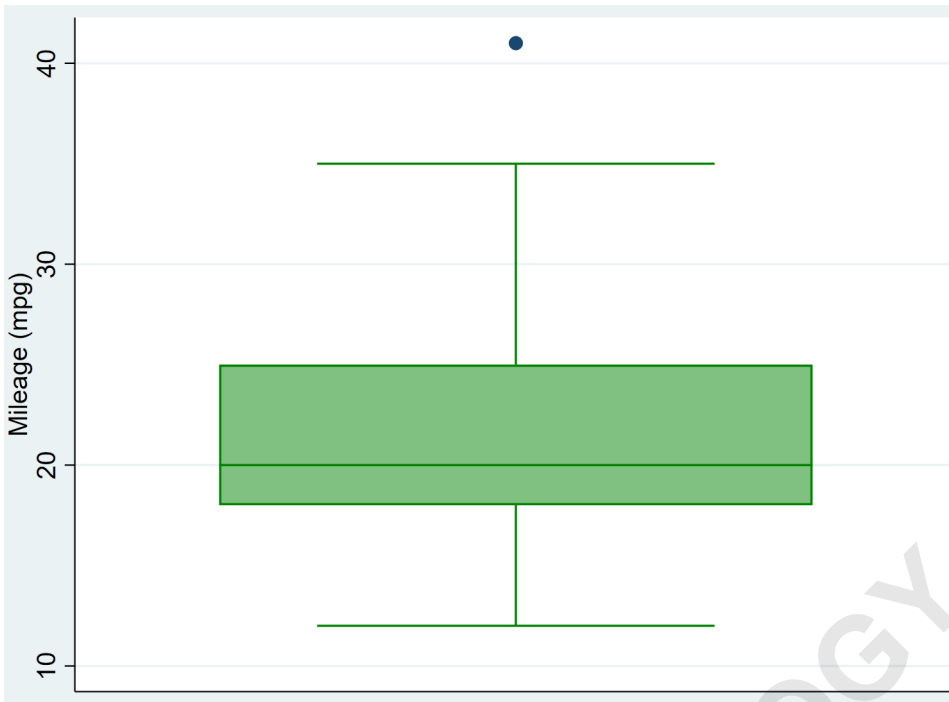
## Customizing Color and Style

Visual appeal plays a significant role in data communication. Stata allows granular control over the colors of plot elements. For box plots, the color of the box itself can be modified using the `box()` option, which takes arguments for the box number (for plots with multiple boxes) and the desired color.

The syntax requires specifying the box element being modified (e.g., `1` for the first box in a simple plot), followed by the style modifier, such as `color()`.

To change the color of the `mpg` box plot to green:

```
graph box mpg, box(1, color(green))
```



A comprehensive list of supported color names and schemes in [Stata](#), including standard colors and RGB definitions, can be found in the official [Stata Graphics Manual](#) documentation.

Mastering these commands allows users to efficiently create informative, comparative, and aesthetically tailored box plots suitable for any rigorous statistical presentation.