15.3.5   Create Advanced Boxplots in ggplot2

> **Jeremy** and Colleen are excited about how their team is progressing. They take a break for lunch and start to brainstorm about how to really impress their CEO. Colleen suggests moving beyond basic charts to some more advanced visualizations, and Jeremy suggests a joint presentation to give Colleen a chance to shine. While Colleen practices her public speaking, Jeremy digs back into R.

Beyond the basic bar, line, and scatter plots, there are a number of more advanced ggplot2 visualizations that can be used to describe specific features of a dataset.

For instance, when performing statistical analysis, we may want to visualize summary statistics using boxplots, or unpack the relationship across multiple variables using a heatmap. Fortunately, ggplot2 has functions such as `geom_boxplot()` and `geom_tile()` that generate more advanced visualizations with ease.

The boxplot is also known as a box-and-whisker-plot, named for the lines extending from the boxes. It's used to visualize a variety of summary statistics for a continuous numerical vector. Boxplots are very common in data science due to the density of information contained within a single visualization, as well as the boxplot's ability to compare measurements across grouping factors.

To generate a boxplot in ggplot2, we must supply a vector of numeric values. For example, if we want to generate a boxplot to visualize the highway fuel efficiency of our mpg dataset, our R code would look as follows:

```
> plt <- ggplot(mpg,aes(y=hwy)) #import dataset into ggplot2
> plt + geom_boxplot() #add boxplot
```
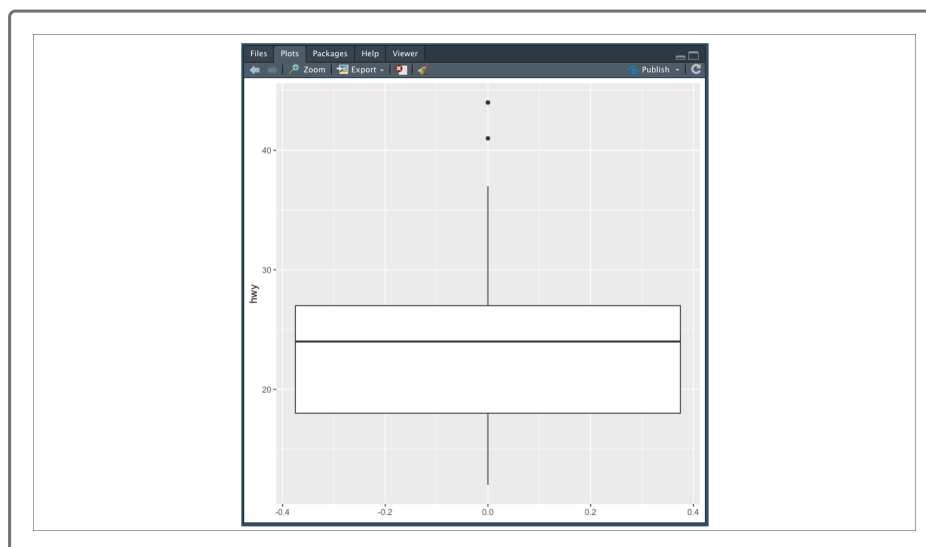
Unlike the previous `ggplot` objects, `geom_boxplot()` expects a numeric vector assigned to the y-value. This is due to the ggplot accounting for multiple boxplots in a single figure. If we supply our categorical grouping factor to x, we can create a boxplot that compares measurements from a variety of groups.

Expanding on our previous example, if we want to create a set of boxplots that compares highway fuel efficiency for each car manufacturer, our new R code would look as follows:

```
> plt <- ggplot(mpg,aes(x=manufacturer,y=hwy)) #import dataset into ggplot2
> plt + geom_boxplot() + theme(axis.text.x=element_text(angle=45,hjust=1)) #
```
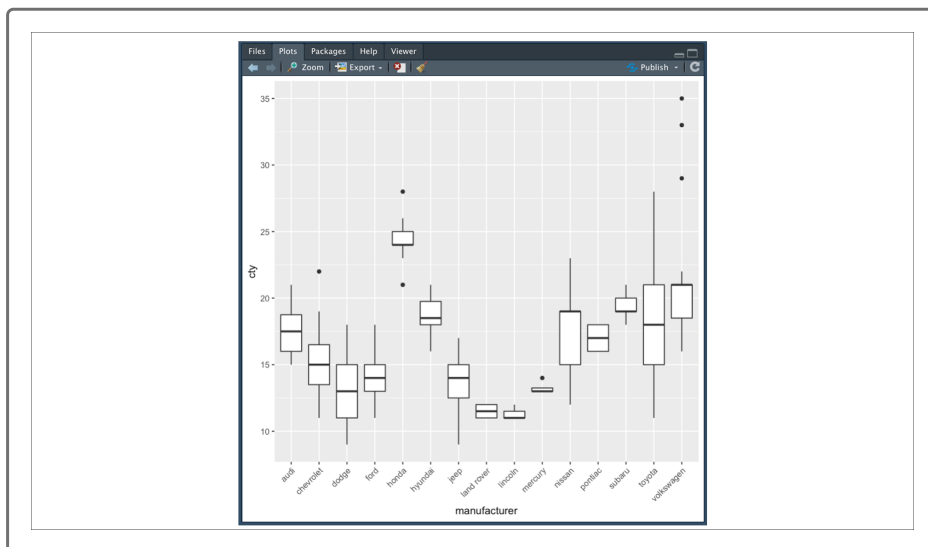

These grouped boxplots are fantastic to use in technical reports and presentations due to how easy they are to read and interpret as well as how much information can be conveyed.

SKILL DRILL

Customize the boxplot to be more aesthetic by adding some color and using dotted instead of solid lines.

Hint: Reference the **geom_boxplot documentation (https://ggplot2.tidyverse.org/reference/geom_boxplot.html#aesthetics)** for assistance.

NOTE

You can apply numerous custom aesthetics to boxplots to improve visualizations so they are easier to read. Be sure to check out the **ggplot2 boxplot documentation (https://ggplot2.tidyverse.org/reference/geom_boxplot.html)** .