# SWIRL ggplot2 Tutorial

* The ggplot2 package "is composed of a set of independent components that can be composed in many different ways. ... you can create new graphics that are precisely tailored for your problem." These components include aesthetics which are attributes such as colour, shape, and size, and geometric objects or geoms such as points, lines, and bars.
* ggplot2 combines the best of base and lattice. It allows for multipanel (conditioning) plots (as lattice does) but also post facto annotation (as base does), so you can add titles and labels. It uses the low-level grid package (which comes with R) to draw the graphics. As part of its grammar philosophy, ggplot2 plots are composed of aesthetics (attributes such as size, shape, color) and geoms (points, lines, and bars), the geometric objects you see on the plot.
* **qplot()** is the basic workhouse function of ggplot2, and **gplot()** is the more advanced workhouse function. The latter is more flexible.
* the **str()** function gives and idea of what a data set contains
* smoothing function to produce trend lines, one for each color? Just add a fifth argument, geom, and using the R function c(), set it equal to the concatenation of the two strings "point" and "smooth". The first refers to the data points and second to the trend lines we want plotted.
* The all-purpose qplot can also create box and whisker plots. Call qplot now with 4 arguments. First specify the variable by which you'll split the data, in this case drv, then specify the variable which you want to examine, in this case hwy. The third argument is data (set equal to mpg), and the fourth, the geom, set equal to the string "boxplot".
* histograms. These display frequency counts for a single variable. Let's start with an easy one. Call qplot with 3 arguments. First specify the variable for which you want the frequency count, in this case hwy, then specify the data (set equal to mpg), and finally, the aesthetic, fill, set equal to drv. Instead of a plain old histogram, this will again use colors to distinguish the 3 different drive factors.
  + It's cool that qplot can do this so easily, but some people may find this multi-color histogram hard to interpret. Instead of using colors to distinguish between the drive factors let's use facets or panels. (That's what lattice called them.) This just means we'll split the data into 3 subsets (according to drive) and make 3 smaller individual plots of each subset in one plot (and with one call to qplot).

## Using **ggplot()**:

* there's a DATA FRAME which contains the data you're trying to plot. Then the AESTHETIC MAPPINGS determine how data are mapped to color, size, etc. The GEOMS (geometric objects) are what you see in the plot (points, lines, shapes) and FACETS are the panels used in conditional plots. You've used these or seen them used in the first ggplot2 (qplot) lesson.
* There are 3 more. STATS are statistical transformations such as binning, quantiles, and smoothing which ggplot2 applies to the data. SCALES show what coding an aesthetic map uses (for example, male = red, female = blue). Finally, the plots are depicted on a COORDINATE SYSTEM. When you use qplot these were taken care of for you.
* As in the base plotting system (and in contrast to the | lattice system), when building plots with ggplot2, the | plots are built up in layers, maybe in several steps. You | can plot the data, then overlay a summary (for instance, a | regression line or smoother) and then add any metadata and | annotations you need.