HW4\_DevinAdams

Ray Nelson, edited by Devin Adams

1/16/2019

## Concepts

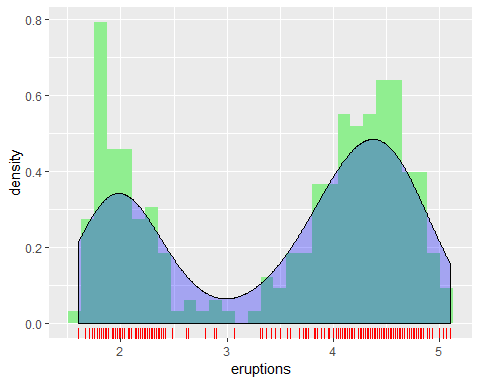
Explain the sequence whereby a word file is created from R Markdown document. What is the purpose of the YAML header?  
To create a word file from an R Markdown document, the code is Knit to produce a report. That sends the R markdown to the knitr which generates a markdown file that’s processed by pandoc which creates the final word file. This process allows the creation of a wide range of output formats.  
The YAML (Yet another markup language) is used to set parameters for the whole document and details of the output. For example, the document parameters and the bibliography.

## Interpretation of code

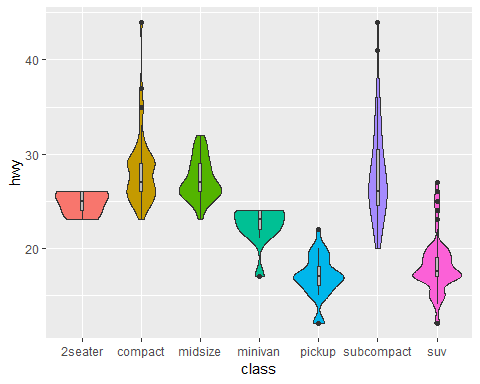
Interpret the following R code:

ggplot(data = faithful, mapping = aes(x = eruptions)) +  
 geom\_histogram(aes(y = stat(density)), fill = "lightgreen") +  
 geom\_density(fill = "blue", alpha = 0.3) +  
 geom\_rug(color = "red")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggplot(data = mpg, mapping = aes(x = class, y = hwy, fill = class)) +  
 geom\_violin(show.legend = FALSE) +  
 geom\_boxplot(show.legend = FALSE, width = 0.05, fill = "grey")

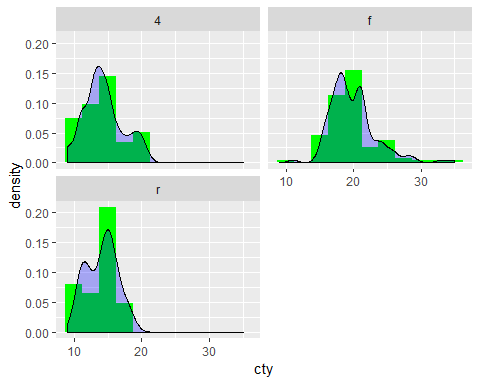


Line 24: Calls ggplot and defines the data to be plotted, faithful. It then mapps eruptions to the x-axis.  
Line 25: Creates a histogram on the first layer. The default stat is overridden using density instead and the bars are then colored lightgreen  
Line 26: Creates a density plot of erruptions on the second layer. The fill is set as blue and it’s given a 30% transparency.  
Line 27: A rug plot is created for erruptions and is colored red.  
Line 29: Another ggplot is created with data for mpg. The class and hwy variablese are mapped to the x and y axes respectively and filled according to their class.  
Line 30: A violin plot is created on the first layer from the mpg data defined above and the legend is turned off.  
Line 31: A boxplot is created on the second layer using the same data as the violin plot, its legend is turned off, the width is set to 0.05, and the fill color is grey.

## Compare the distribution of city mileage from the mpg dataframe for each of the three different types of drive trains using:

1. Facets with a density superimposed on top of a histogram
2. Violin plots

ggplot(data = mpg, mapping = aes(x = cty)) +  
 geom\_histogram(aes(y = stat(density)), fill = "green", binwidth = 2.5)+  
 geom\_density(fill = "blue", alpha = 0.3)+  
 facet\_wrap(~ drv, nrow=2)



ggplot(data = mpg, mapping = aes(x = drv, y = cty, fill = drv)) +  
 geom\_violin(show.legend = FALSE) +  
 geom\_boxplot(show.legend = FALSE, width = 0.05, fill = "grey")

