Homework #4

Derrik 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?

A word file is created if the output section in the YAML has “word\_document” there. When the document is then knit, the .Rmd file is sent to the knitr where the code chunks are executed and a new markdown document is created. The new markdown file is then processed by pandoc, which creates the finished file in the desired output format.

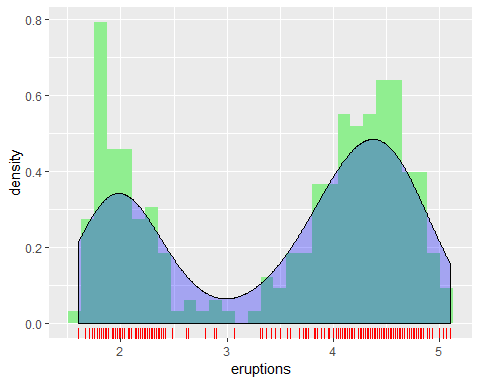
The purpose of the YAML header is to control the entire document settings with just a few parameters and make doing so easy for people to read and write.

## Interpretation of code

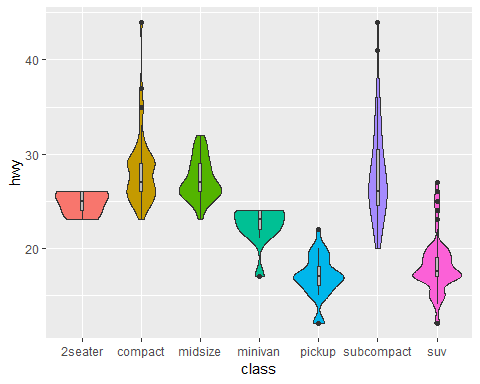
Interpret the following R code:

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")



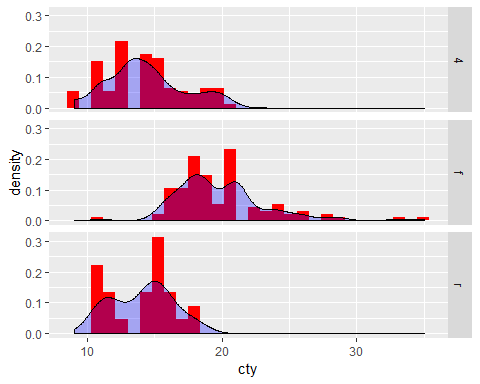
1. Uses ggplot on the faithful data and sets the x-axis to show eruption time.
2. Creates a histogram where the y-axis is the density of the frequency for each point of data within that bin and colors it green 1. Creates a density plot filled blue with a transparency of .3 to allow you to see the histogram behind it
3. Creates a rug plot of the eruption times and colors it red. All plots are plotted over each other.
4. Uses ggplot on the mpg data and defines the x-axis as the class of cars and the y-axis as the hwy mpg. Fill then tells it to color according to class.
5. Creates a violin plot that doesn’t show the legend since it would be redundant.
6. Creates a boxplot without the legend again and plots it over the violin plot. It’s width is defined as .05 and the color for the boxes is defined as 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 = "red") +  
 geom\_density(fill = "blue", alpha = 0.3) +  
 facet\_grid(rows = vars(drv))

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



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

