Module 12: Markdown

Ian MacLeod

4/13/2020

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

## My Markdown

Using this template, I am now going to provide a rundown of the work I did for **Module #7: Distribution Analysis** in Dr. Friedman’s *Visual Analytics* class. The associated blog post can be found [here](https://ianmacusf.wordpress.com/2020/03/02/va-module-7-distribution-analysis/).

In that assignment, I had to conduct **distribution analysis**. I used the *mtcars* dataset to find the distribution of horsepower among the cars provided. Here’s the code:

avg <- mean(mtcars$hp)  
med <- median(mtcars$hp)  
  
library(grid)  
text1 <- "Dotted Line = Mean"  
annot1 <- grid.text(text1, x = 0.7, y = 0.9)  
text2 <- "Solid Line = Median"  
annot2 <- grid.text(text2, x = 0.7, y = 0.8)  
  
  
library(ggplot2)  
ggplot(mtcars, aes(x = hp)) +  
 geom\_histogram(binwidth = 25, color = "white", fill = "dark blue") +  
 labs(title = "Distribution of Horsepower", subtitle = "Bin Width = 25", x = "Horsepower", y = "Number of Cars") +  
 scale\_x\_continuous(breaks = seq(0,350,25)) +   
 scale\_y\_continuous(breaks = seq(0,9,1)) +   
 geom\_vline(xintercept = avg, linetype = "dotted", color = "red", size = 2) +   
 geom\_vline(xintercept = med, color = "red", size = 2) +  
 annotation\_custom(annot1) +   
 annotation\_custom(annot2)

## The Explanation

* Here’s what the code does:
  + The mean and median are calculated and assigned to *avg* and *med*
  + The annotations *annot1* and *annot2* are created and their coordinate positions are set
  + The “ggplot” function is called using the *mtcars* data-set and the horsepower (hp) data is assigned to the x-axis
  + “geom\_histogram” sets the bin width to 25, the color to white, and the fill to dark blue. This gives each bar a white outline
  + “labs” sets the title, the subtitle, and the labels for the x and y axis
  + “scale\_x\_continuous” sets the tick marks on the x-axis so that they are compatible with the bin width, making the graph easier to read
  + “scale\_y\_continuous” sets the tick marks on the y-axis so that there is a tick for each value in the range of data, making the graph easier to read
  + “geom\_vline” uses the previously calculated objects *avg* and *med* to create vertical lines representing the mean and the median of the data, respectively
  + “annotation\_custom” uses the previously specified character strings *annot1* and *annot2* to create the key in the upper right corner of the graph. The key denotes the difference between the mean and median lines, which should make the graph easier to read

## The Annotation and Plot

