Stat 604
Assignment 06 R

Scope:

This assignment reinforces the R material covered through R Lesson 6 (Lecture 7). You will practice using higher level graphics and adding objects to graphs

Specific Instructions for this Assignment:

We will continue to work with the Oklahoma school data that we have been using in the previous R assignments.

In addition to the NetID_HW06_console.pdf and NetID_HW06_script.pdf like you submitted in the previous assignment, this assignment will require a NetID_HW06_graph.pdf file containing the graphics output of this assignment. There are three "questions" on WebAssign; one for each of your homework files from this assignment. Make sure each file is properly uploaded to the appropriate question.

Perform in R each of the exercises listed below. Include the required header information and a comment line in your script above the section for each step so that each is clearly identified. The steps in this assignment may require multiple lines of R code to obtain the desired results.

A PDF file containing the graphics produced by the assignment code has been posted on eCampus for your reference. Except for the time stamp, your output should match the sample output provided.

1. At the top of your script, include housekeeping steps to first list the contents of the workspace and then clear it out.
2. Load the HW04 workspace that you downloaded from eCampus for the previous assignments. You may use the file menu on the R console to browse for and load the workspace. When this is completed, copy the line of code from the console window and paste it into your script so you can load the workspace automatically the next time you want to run the script. After loading the workspace, include a step that lists in the console the contents of the workspace again to confirm the desired data file has been loaded.
3. Include a statement that will cause R to send all of the graphics in this assignment directly to the PDF file that you will be posting on WebAssign. NOTE: You may want to wait until you have all of your graphics working properly before including this step so you can see the graphics in the console while you are working on them.
4. Create histograms of the PTRatio in the Oklahoma schools. Make sure the layout, labels, and title all match the sample output.
   a. Create the first histogram using the default number of breaks.
   b. Create a vector that can be used to extend the length of the X axis and force breaks every 5 pupils/teacher. You may find it helpful to use the statistics on the PTRatio column to determine how far your sequence should go.
   c. Create a new histogram with break points at 5 pupils/teacher.
5. Add a maroon line to the graph that shows the normal distribution density of the PTRatio numbers. Feel free to use separate lines of code as necessary to create the vectors required for this step.

6. Draw a yellow vertical line on the previous histogram at the average PTRatio. Use a function to determine the position of the line instead of hard coding in the mean value. Use a value from the color palette to specify the color instead of using the color name.

7. Create a new plot showing the number of teachers compared to the Pupil Teacher Ratio (PTRatio). Use the plot character shown in the sample output on eCampus. The color of the plot character is purple. Force the x axis to extend to 140.

8. Add a colored fit line to the plot. Use a hex value to mix a color for the line that is FF parts Red, 99 parts Green and no Blue. In the console, show the summary statistics for the model that created the fit line.

9. Imbed the date and time of creation near the upper right hand corner of the graph area. The coordinates of the time stamp location need only be close to the ones used for the sample output. This time stamp must be based on system time so that it updates whenever your code is run.

10. Create a boxplot of the number of students in each grade (Columns Grade7 through Grade12). Make sure the labels, etc. match the posted output. The inside of the boxes is light green.

11. Add diamonds that can be colored and filled that represent the average number of students in each grade from Tulsa County. You may create the vector of averages first to simplify the function that is used to plot the diamonds. The background of the diamonds is dark green and the border is red. Make the size of the diamonds larger than the default similar to what is shown in the sample output. (TIP: Do not over-complicate the plotting of the diamonds. You can plot a vector without supplying the x component of the points if the vector labels match the data used to create the original graph as it does in this case.) You may find it helpful to consult the R help on points in order to find out which plot character to use and how to specify the background color.