Programming Exercise 4

Crime Analytics (CJUS 6106)

Instructions

For this assignment, you will be using the graphing functions reviewed in the Data Visualization Part I - Exploratory Graphics lecture and the NLSY97 data set I provided for an earlier exercise (the one you didn't need to create). To complete this assignment, you will need to use variables within this data frame (or newly created/recoded ones) to create a series of plots like those reviewed during lecture. You should submit both a .rmd file and knitted PDF by 11:59pm on the due date indicated on Canvas.

Question 1

Load the ggplot2 library and then make a copy of the nlsy97 data frame.

Then, make sure that all categorical variables in the copied data frame are stored as factors.

Question 2

Create a barplot that counts the number of youth who belong to each category of household structure. Be sure to provide appropriate axis and graph titles and to create bar labels for the total counts within each bar. Your final bar plot should look like the final graph for the transmission types in my lecture. **Note**: You do not need to create the graph iteratively like I do in lecture - only one code chunk is necessary to complete this part of the question.

Next, add the race variable to the plot like I add the number of cylinders in my example. Be sure to adjust the graph title as appropriate. **Note** - you can adjust the size of the bar labels using the size= option within the geom_text() function (I used a value of 2.5). Some bars will be very small and their labels won't have space to appear - change the color of the labels to "black" to partially account for this.

Provide a short, 2-3 sentence interpretation for each graph.

Question 3

Create a variable within the copied data frame called crime_freq that is the sum of all the crime count variables measured at the 1999 interview (five variables, in total).

Create a scatterplot with weeks employed along the x-axis and crime_freq along the y-axis. Be sure to apply the appropriate graph and axis titles and include a best fit line (with proper formatting) using the geom_smooth() function.

Next, create a new version of this graph where the color of the points is represented by the variable sex.

Provide a short, 2-3 sentence interpretation for each graph.

Question 4

Create a histogram that depicts the distribution of weeks employed. Be sure to provide the appropriate formatting (i.e., graph and axis titles).

Next, change the colors of the bars to coincide with the household structure variable. Be sure to adjust the appearance of the bars so they are all visible instead of stacked on top of one another in solid colors.

Provide a short, 2-3 sentence interpretation for each graph.

Question 5

Take the last graph you completed in Question 4 and add a density plot to it. Be sure to adjust the appearance of the plot as necessary (including the appearance of bars and titles).

Provide a short, 2-3 sentence interpretation for this graph.

Question 6

Create a boxplot where the x-axis is household structure and the y-axis is number of weeks employed. Be sure to provide the appropriate formatting options for this figure and to include the data points as dots as I do in lecture (use jitter=0.2 alpha=0.05).
Finally, add the race variable as the fill for the boxes. This is a little different from lecture, but the process is like the other graphs (this time, it just adds a box for each household structure and race combination).

Provide a short, 2-3 sentence interpretation for each graph.