Assignment 2 - LMS Modules 3 & 4

YOUR NAME HERE

due 9/22/21

Use the attrition dataset to answer the following questions. As always, your assignment must be turned in as an .Rmd file and ONE knit file - PDF (preferred) or html (if unable to knit to PDF). *Please put your descriptive sentence(s) for each question OUTSIDE of your code chunks.*

First code chunk - set the file structure and open your libraries.

1. Create a plot that shows the distribution of monthly income (hint: this is a univariate graphic). Write a sentence or two describing what the plot tells you about the data.

2. Create a plot that shows the average level of monthly income by field of education. Write a sentence or two describing what the plot tells you about the data.

3. Create another plot that shows average level of monthly income by field of education and gender. Write a sentence or two describing what the plot tells you about the data.

4. Create a plot that shows average levels of monthly income by field of education, gender and job level (note - job level is on a scale of 1-5, highest-ranked employees are 5). Write a sentence or two describing what the plot tells you about the data.

For the following problems you'll need to open up, clean, and save datasets using the tools we've gone over in class. For each dataset, make sure that when you're done you have a nice, neatly labeled dataset that would be easy for you or another analyst to open and analyze. Some of the datasets may need no/minimal cleaning. Some may need some cleaning. That is at your discretion as the lead analyst. Save each dataset as an `RData` file. You will upload each saved R dataset to the LMS with your knit pdf and .Rmd file.

PLEASE NOTE: You can't just copy these links into your code. You need to figure out the format for the data. Take a look at the website to figure out what kind of data is there; then you can download the dataset and use the appropriate code in R to import it. Additionally, a CLEAN dataset means one that could be easily opened by another analyst and used immediately. You should tidy up as needed.

5. 2016 Health Education District Data:

https://www.cdc.gov/healthyyouth/data/shpps/data.htm

6. County level replication file for "Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States":

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BT3LXD

7. Airline dataset:

http://www.principlesofeconometrics.com/sas.htm

8. King county births:

http://courses.washington.edu/b517/Datasets/datasets.html