Midterm Term Project
Due April 1 by 11:59 PM

Steps:

1) Unzip the files in the the ct_crash.zip dataset. They are 12 datasets of accidents in Connecticut over the past several years. Each file is a record of all the accidents involving a different type/cause of accident. So the DUI file is a record of all the accidents where a driver was intoxicated. The Distracted Driver file is a record of all the accidents that involved a distracted driver. **An unique accident (identified by CrashID) can be in more than one dataset.**
2) Open the files in R and combine into one dataset. This dataset should not have any duplicated accidents!
3) For each of the 12 causes/types of accidents, created a new variable, that identifies whether or not the accident was caused or involved by the cause/type of accident in question. For example, created *Distracted* and if an accident was in the distracted dataset it gets a "yes" and if it was not, it gets a "no". You should end up with a dataset with 47 columns, the 35 original columns in each file, plus the 12 new variables that you created.
4) Take the first letter of your surname. Find the city that starts with that letter with the most accidents. If your surname starts with an "I" you can use the letter "H", if it starts with a "J" you can use the letter "K" and if it starts with "X", "Y" or "Z" you can use "W". So for me I would get "Stamford" since Stamford's 22k accidents exceeds Stratford's 8k. *Hint: You can use the function starts With to give you a TRUE/FALSE whether a character string starts with a character.*
5) Once you have your dataset, conduct a comprehensive EDA (exploratory data analysis) of your dataset. What types of car accidents occur in your town? What factors seems to be involved (type of route, weather, school bus related etc.). How severe are the types of accidents in your town. Are there any notable temporal factors (accidents are increasing/decreasing over time?

Accidents occur more often at what time of the day etc.?). Put yourself in the shoes of a local government analyst or consultant hired to conduct a review of car accidents in the town.

6) Display your analysis using ggplot2, plotly and DT in flexdashboard. Be sure to label your graphs professionally.

Various Additional Hints that Maybe Helpful

1) You can use the "$" to treat a data.frame column as if it were a vector.

2) There are multiple ways to convert a date-time based variable that is misidentified as a character.

3) The data comes from the Connecticut Department of Transportation and a code value document can be found at https://gis.cti.uconn.edu/software/Crash_Coded_Values.xlsx.