Statistics 3080
Homework 4

Note: Unless otherwise specified in a problem's instructions, **the final result must always be printed for all parts of all problems**.

Note: Do not use any loops.

Problem 1: The data file *fatal accidents.csv* contains a record of each fatal accident in Virginia and surrounding states in 2019. Each record details the number of vehicles involved, the number of people involved in each accident - both inside and outside the vehicle, and timing information for each fatal accident. The day of the week begins with 1 for Sunday.

a. Read in the file and save the data as a data frame. Print the first few rows of the resulting data frame.

b. Create a list called `state.list` that separates the data set by state. Each element of the list should contain a data frame subset to a single state. Each element should be named with the name of the state.
Note: Do NOT print the list.

c. Use an appropriate `apply()` function to show the first three rows of each element of the list created in part b.

d. Use an appropriate `apply()` function with `state.list` to create a list that contains tables that show the percentage of accidents in 2019 that occured on each day of the week. Each element of the list should contain the table for a single state.
Note: The percentage values should be rounded to one decimal. For example, 95.1% will be displayed as 95.1.

e. What observations can you draw from the tables created in part d?

f. Using an appropriate `apply()` function with `state.list`, create a list that contains tables that show the number of accidents in 2019 that occurred on each day of the week (rows) and the number of vehicles involved (columns). Each element of the list should contain the table for a single state.

g. What changes to the presentation of the tables in part f might improve ease of drawing observations from them?

Problem 2: Continue using the data file *fatal accidents.csv* described in problem 1. For the following, use functions in the `tidyverse`.

a. Add a column to the data frame read in for problem 1 part a called `People.count` that contains the total number of people involved in each accident. Print the first few rows of the resulting data frame.
Note: The function for printing the first few rows does not need to be in the `tidyverse`.

b. Display the average number of vehicles and the average number of people (columns) involved in accidents in 2019 in each state (rows) in a single two-dimensional data structure.
Note: The function for calculating the averages does not need to be in the `tidyverse`.

c. Display the minimum, average, and maximum number of vehicles (columns) involved in accidents in 2019 in each state (rows) in a single two-dimensional data structure.
Note: The functions for calculating the minimum, average, and maximum do not need to be in the `tidyverse`.

d. What observations can you draw from the tables created in parts b and c?

e. Display the number of accidents (columns) that occurred in Virginia in each month (rows) of 2019 in a single two-dimensional data structure.

f. Display the median and mean number of vehicles (columns) involved in accidents that occurred in Virginia on each day of the week in each of the summer months (June through August) (rows) of 2019 in a single two-dimensional data structure.
Note: The functions for calculating the median and mean do not need to be in the `tidyverse`.
Note: All rows should be displayed.

g. What observations can you draw from the tables created in parts e and f?