**IST 687 PREP EXERCISE 05**

**Student Name: Bhavish Kumar**

**Prep Exercise No: 05**

**Date Due: 25th September 2019**

1. **Reading JSON**
2. For this week’s homework we’ll practice be using vehicle accident data from the State of Maryland. First, get the following lines of code to work, you might need the following packages installed and libraried: RCurl, jsonlite.

dataset <- getURL("https://opendata.maryland.gov/resource/pdvh-tf2u.json")

df <- jsonlite::fromJSON(dataset)  
  
*Note that there are many different versions of fromJSON, so the syntax jsonlite::fromJSON ensures we are using the fromJSON function from the jsonlite package.*

1. **Explore the dataframe and answer each of the following questions!**
2. Using the VIEW command, explore the data. You may also want to use the ?command. You may even have to do additional research to help you understand the elements of the dataset.

View(df)

?df

str(df)

summary(df)

1. What does getURL do?

getURL function is part of the RCurl package that is used for retrieving the source of a webpage that helps in scrapping data from a webpage. The JSON structure of a webpage can be obtained using the getURL function.

1. Describe why we use fromJSON.

fromJSON function is used to convert the data from JSON format to R object format, i.e. to the format of a dataframe.

1. How many accidents are in the dataset?

length(unique(df$case\_number)) ### returns the number of unique values in the column case\_number of the dataframe df

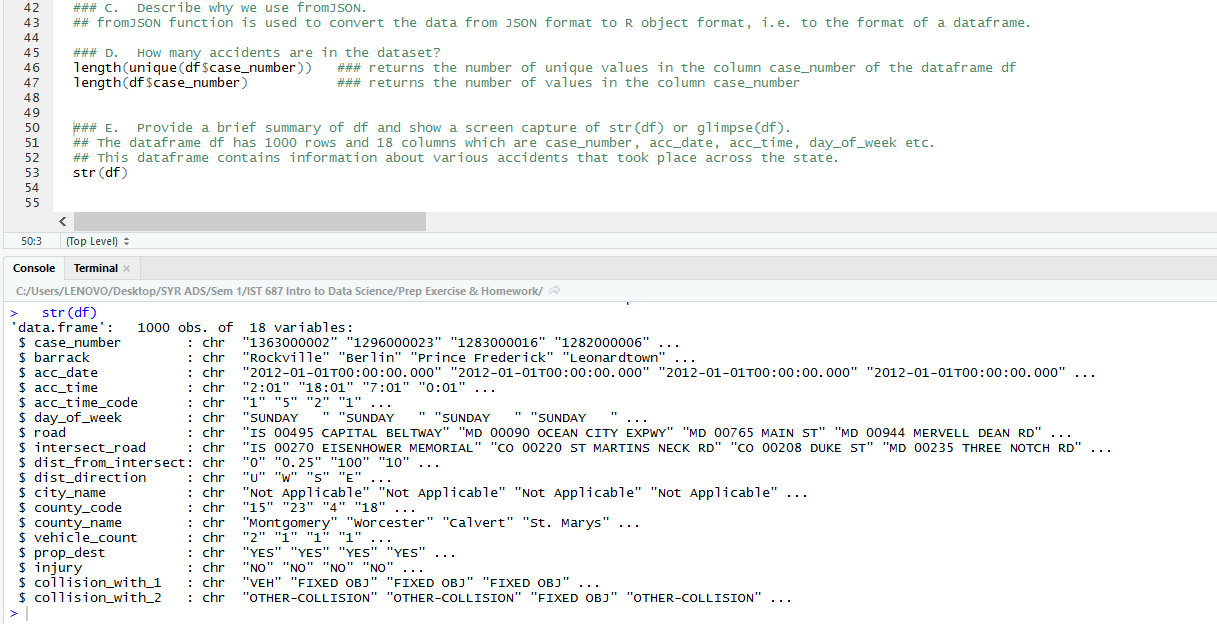
length(df$case\_number) ### returns the number of values in the column case\_number

There are 1000 rows in the table recording 1000 accident cases of which 999 are distinct cases.

1. Provide a brief summary of df and show a screen capture of str(df) or glimpse(df).

The dataframe df has 1000 rows and 18 columns which are case\_number, acc\_date, acc\_time, day\_of\_week etc. This dataframe contains information about various accidents that took place across the state.

str(df)



glimpse(df)

