# Prep Exercise (PE05) JSON and Lists (External datasource Connections)

### General Instructions

1. For this exercise you will answer all of the questions in this document and turn it in to Blackboard.
2. Before you get started make sure to read Chapters 11 of *An Introduction to Data Science* and execute the code throughout the chapter to gain familiarity.
3. Note that up until now we have been working with quite simple data structures. For example, a vector is just a simple list of values; a data frame is a rectangular data structure with well-defined columns. This week we will start using more complex datasets.

# IST 687, Standard Homework Heading

#

# Student name: Thadhani Hitesh Chandrakumar

# Homework number: PE05

# Date due: 25th Sep 2019 11:59PM

#

# Attribution statement: (choose the statements that are true)

# 1. I did this work by myself, with help from the book and the professor

# 2. I did this homework with help from the book and the professor and these Internet sources: <provide the urls>

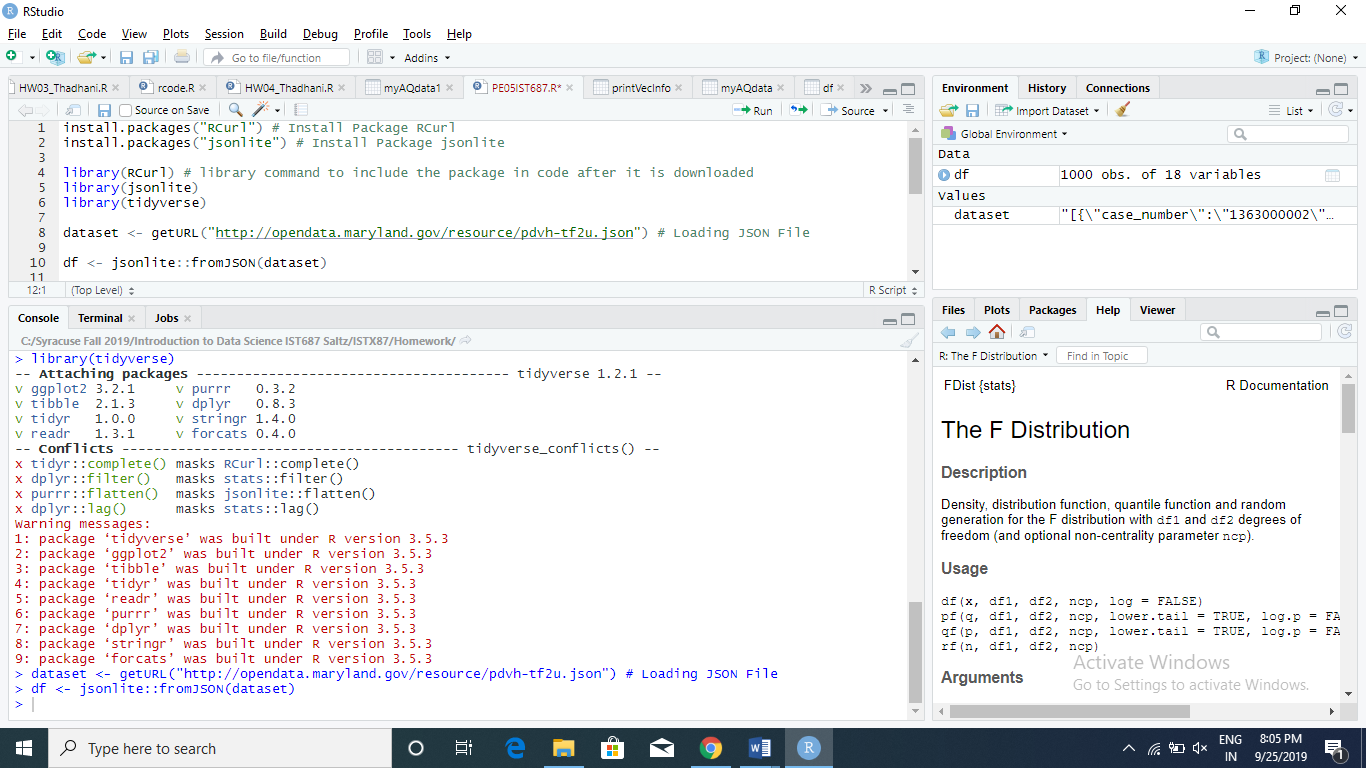
# 3. I did this homework with coaching from <Name of another student> but did not cut and paste any code

### Prep Exercise

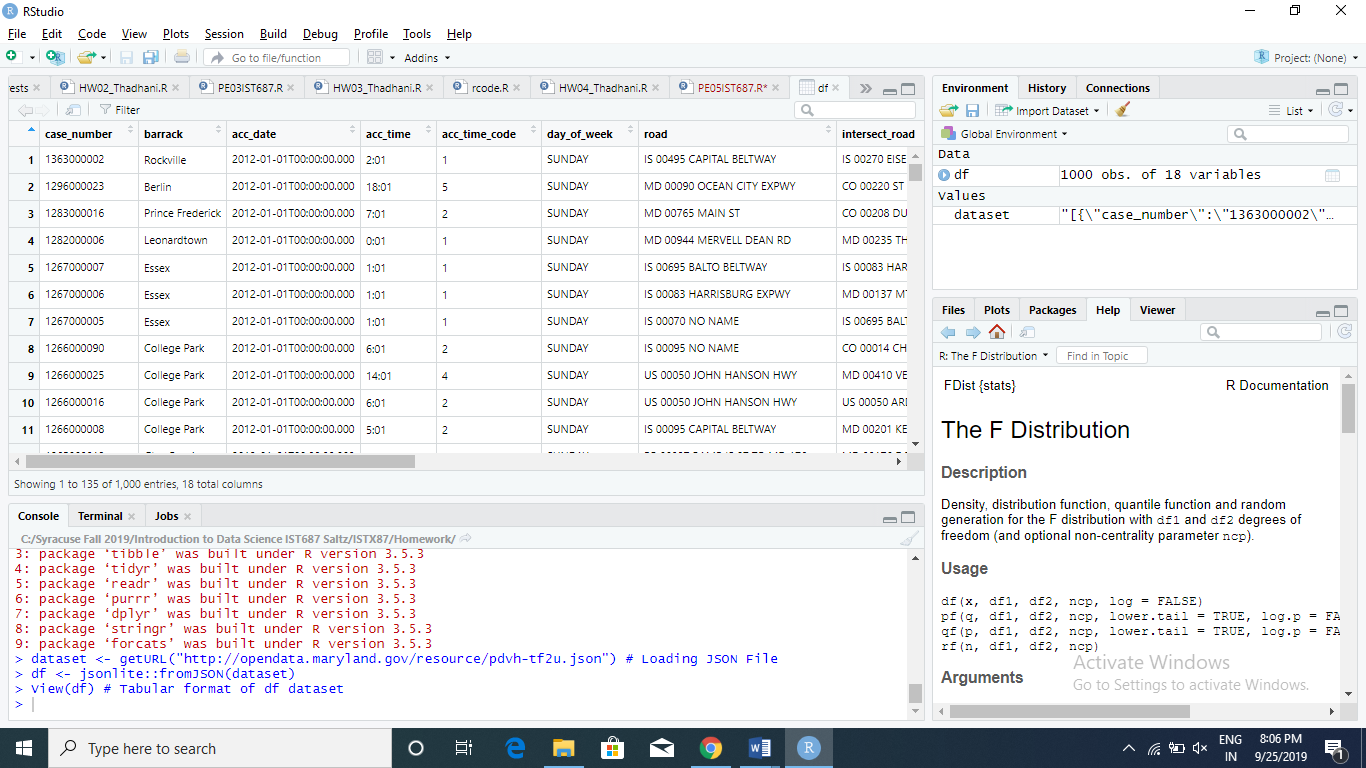
1. **Reading JSON**
2. For this week’s homework we’ll practice be using vehicle accident data from the State of Maryland. First, get the following lines of code to work, you might need the following packages installed and libraried: RCurl, jsonlite.

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.



1. What does getURL do?

getURL Command loads the JSON, CSV file from the website which is specified in URL into R.

1. Describe why we use fromJSON.

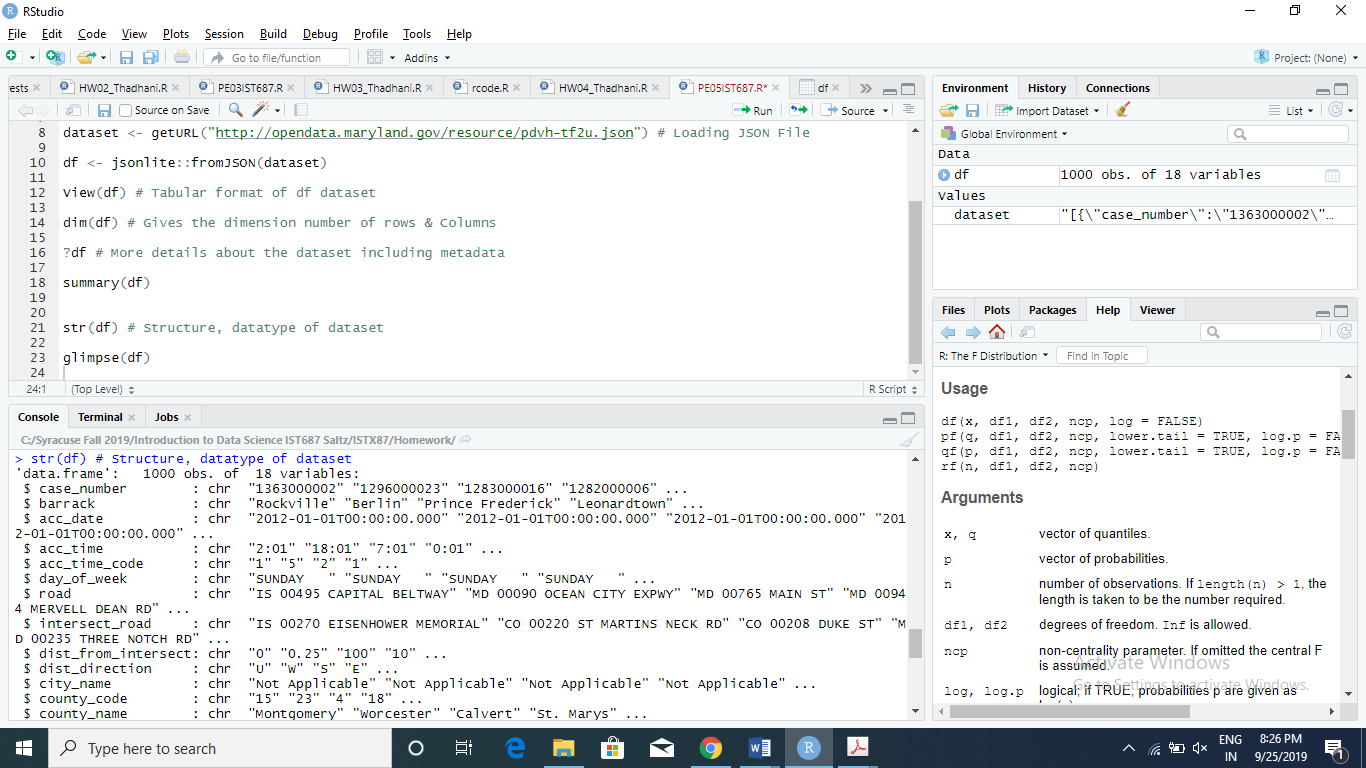
fromJSON is used to convert the data in JSON Format to R readable format.

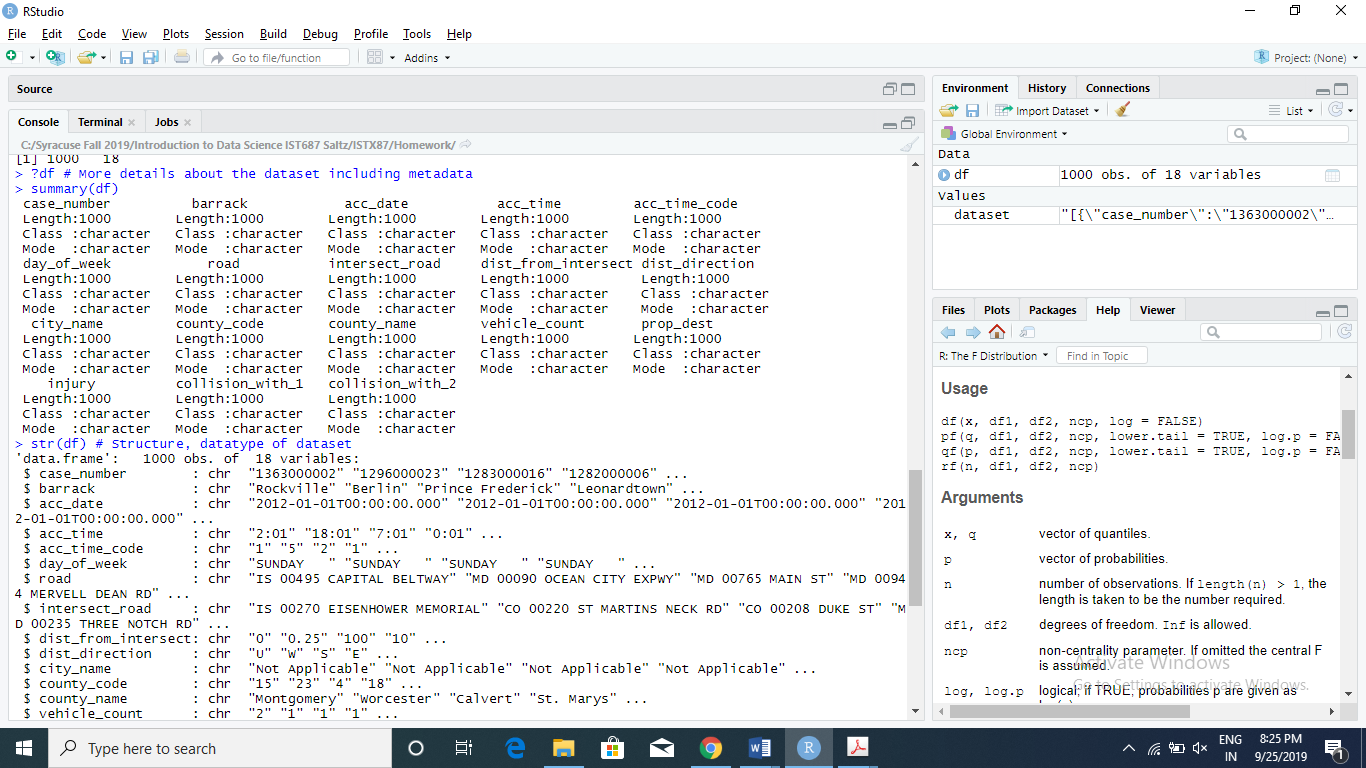
1. How many accidents are in the dataset?

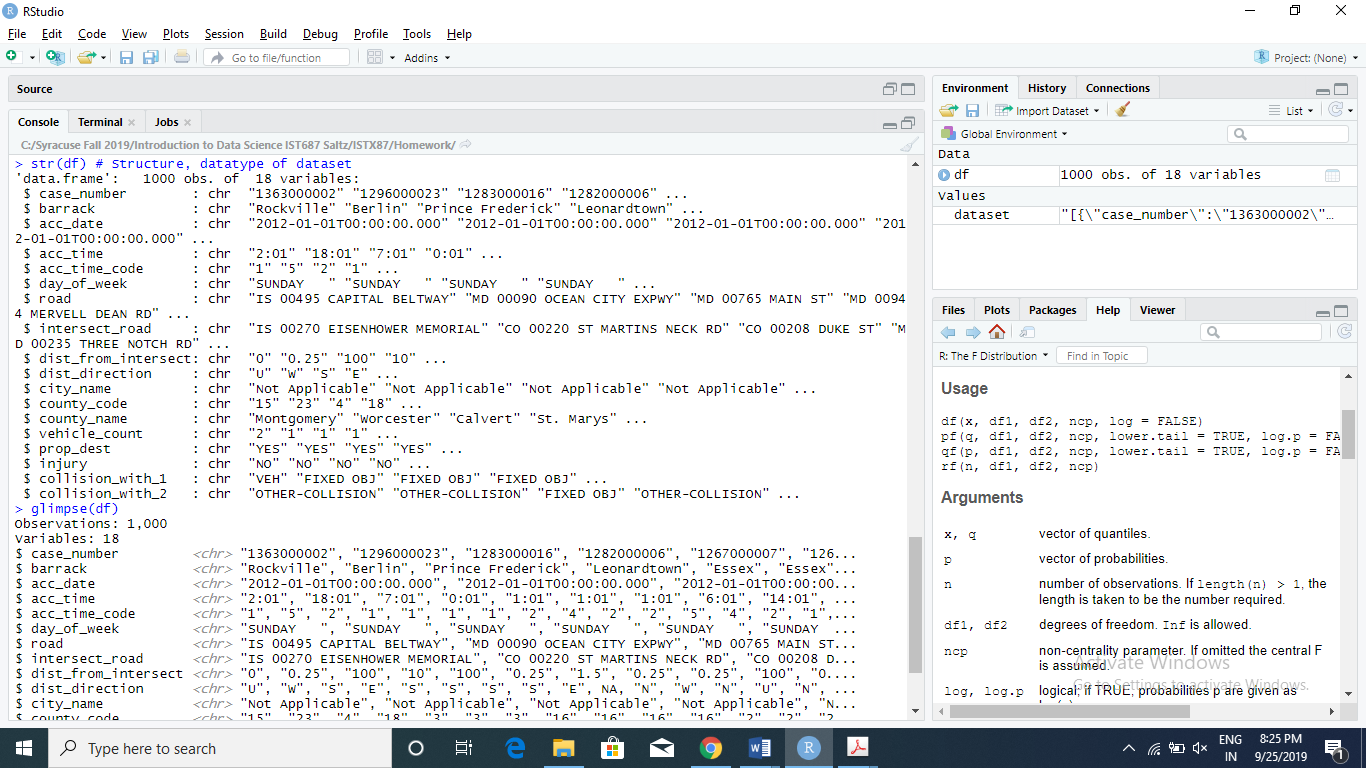
There are 1000 accidents which can be found out by dim(df) command giving 1000 rows and 18 variables.

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

df dataset contains the Accident data with different attributes such as accident date, accident time, case number etc.







1. **List any additional resources you used here.**
2. **Be sure to save your R file as this will become the starting code for your homework.**

***You must submit all Prep Exercises to blackboard prior to the deadline specified for each assignment.*** PE assignments are due on the evening prior to the lecture class. Late PE assignments will not be accepted for credit.