# 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.

### 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.
3. What does getURL do?

**getURL is used to retrieve the result of a webpage.**

1. Describe why we use fromJSON.

**fromJSON is used to convert JSON data into R object. In this case the JSON data is converted into a dataframe.**

1. How many accidents are in the dataset?

**Totally there are 1000 accident entries in the dataset. 999 entries have unique case number and 1 case number is duplicated.**

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

**The dataset contains motor vehicle accident data from Maryland state between 01/01/2012 and 01/21/2012.**

**The df dataframe has 18 attributes and 1000 instances. All the attributes are of character data type.**

**case\_number is the unique number given to each accident that is reported**

**barrack is one of the 23 Maryland state police barracks**

**acc\_date is the date when the particular accident occurred**

**acc\_time is the time when the accident took place**

**acc\_time\_code determines the code based on the accident**

**day\_of\_week is the day of the week when the accident occurred**

**road is the location of the accident**

**intersect\_road is any junction closer to the accident spot**

**dist\_from\_intersect is the distance of the junction from the accident spot in meters**

**dist\_direction is the direction in which the distance is measured from the accident location**

**city\_name is the city where the accident took place**

**county\_code is the code of the county**

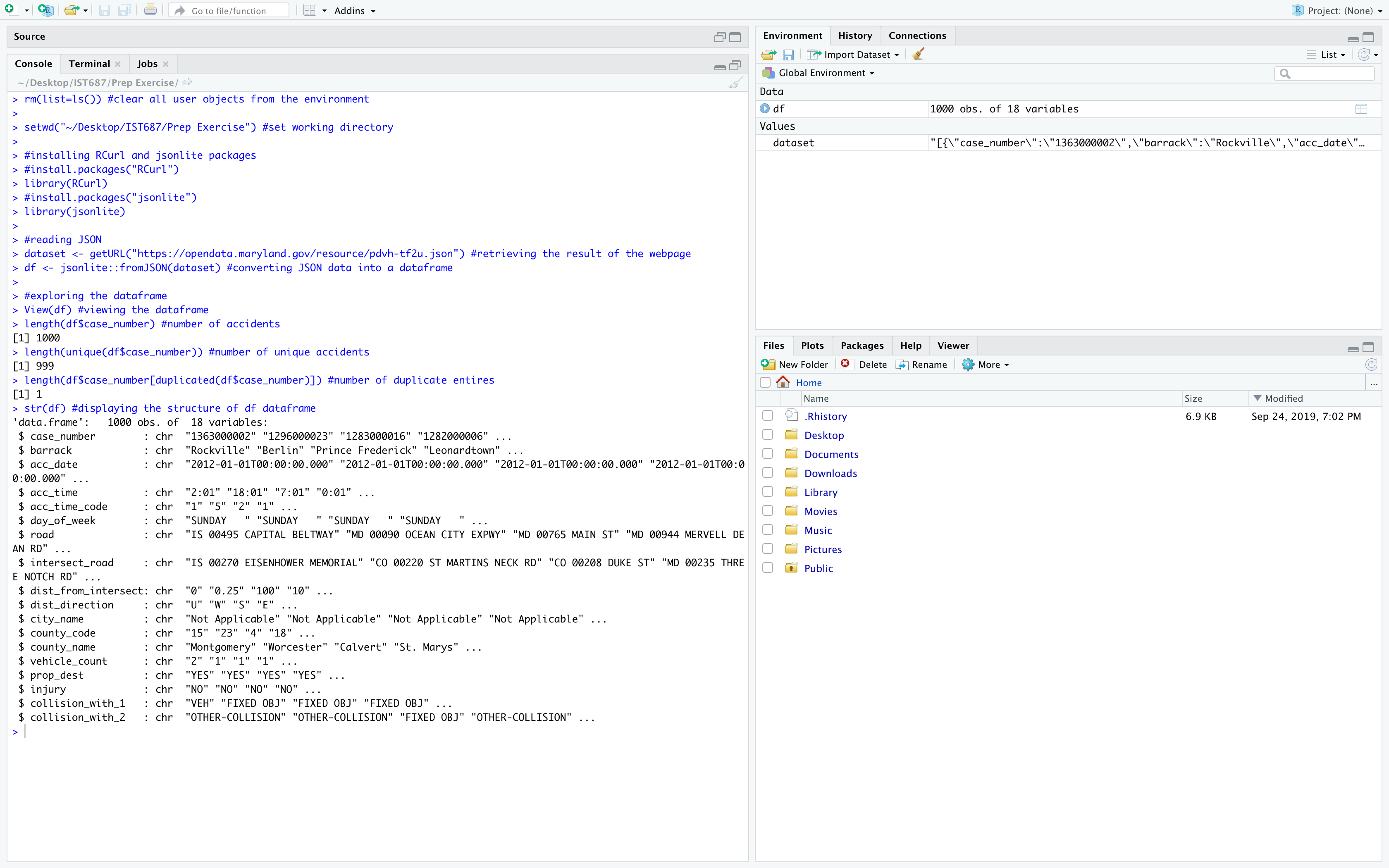
**county\_name is the name of the county**

**vehicle\_count determines the number of vehicles involved in the accident**

**prop\_dest specifies if any property was destroyed because of the accident**

**injury determines if anyone was injured in the accident**

**collision\_with\_1 and collision\_with\_2 specifies objects that were involved in the accident**

****

1. **List any additional resources you used here.**

<https://opendata.maryland.gov/Public-Safety/2012-Vehicle-Collisions-Investigated-by-State-Poli/pdvh-tf2u>

1. **Be sure to save your R file as this will become the starting code for your homework.**
2. ***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.