**ASSIGNMENT 17euit134**

**1. Consider any dataset (other than Vehicle) and address the steps involved in handling Missing data and its impact on applying any technique on it (Both With and without NA and NULL values)**

**Correcting the errors**

**What to do with the missing values?**

First, thoroughly explore your dataset to identify and visualize the missing values.

**Three main choices:**

**Ignore:** Discard samples with missing values from your analysis.

**Impute:** "Fill in" the missing values with other values.

* Impute with the mean
* Impute with the median

**Accept:** Apply methods that are unaffected by the missing data.

The strategy depends on the type of missingness you have.

There are three types of missing data:

**MCAR:** Missing Completely At Random

**MAR:** Missing At Random

**MNAR:** Missing Not At Random

5 R packages popularly known for missing value imputation:

1. MICE
2. Amelia
3. missForest
4. Hmisc
5. mi

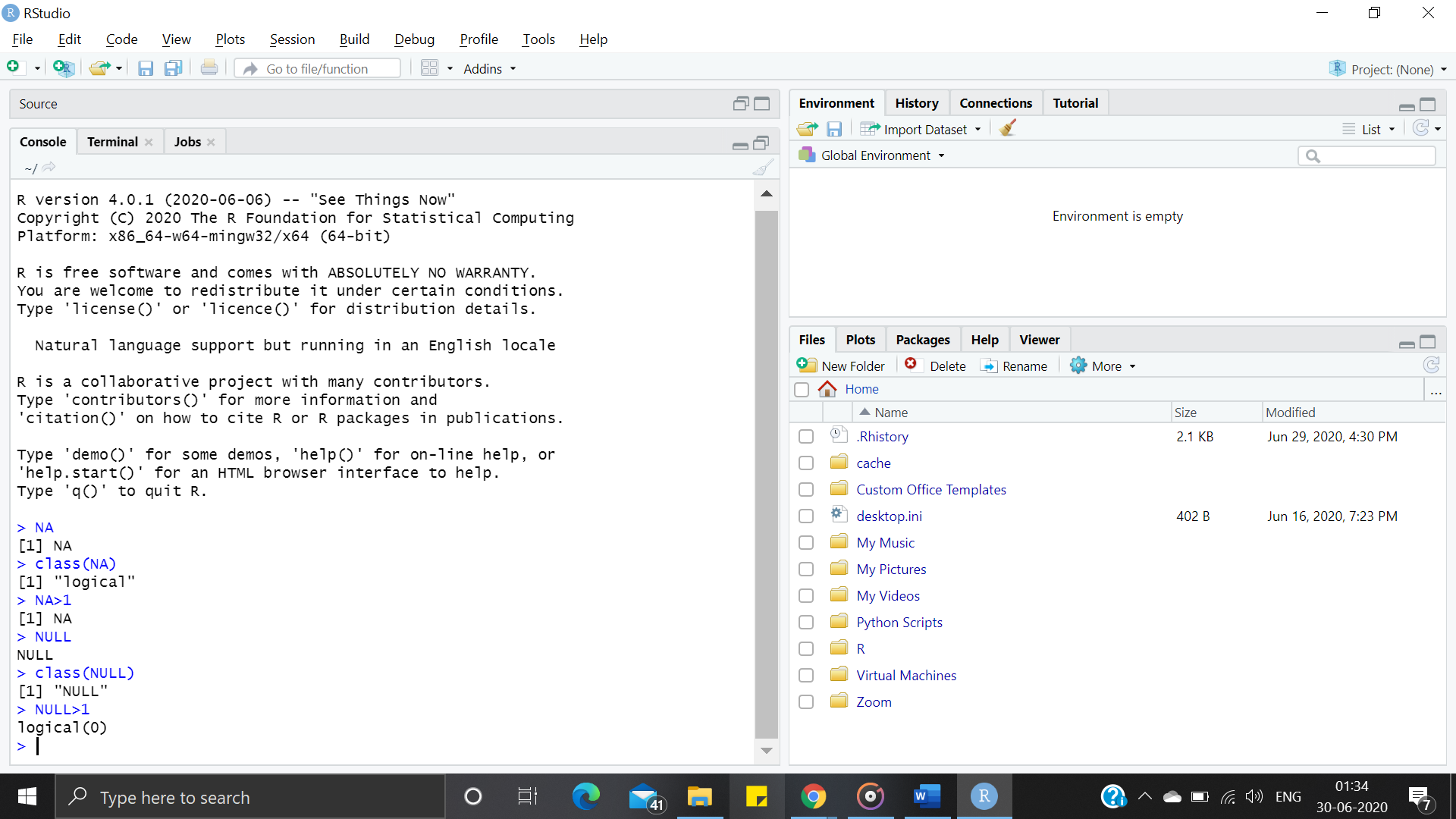
**NA AND NULL VALUES:**

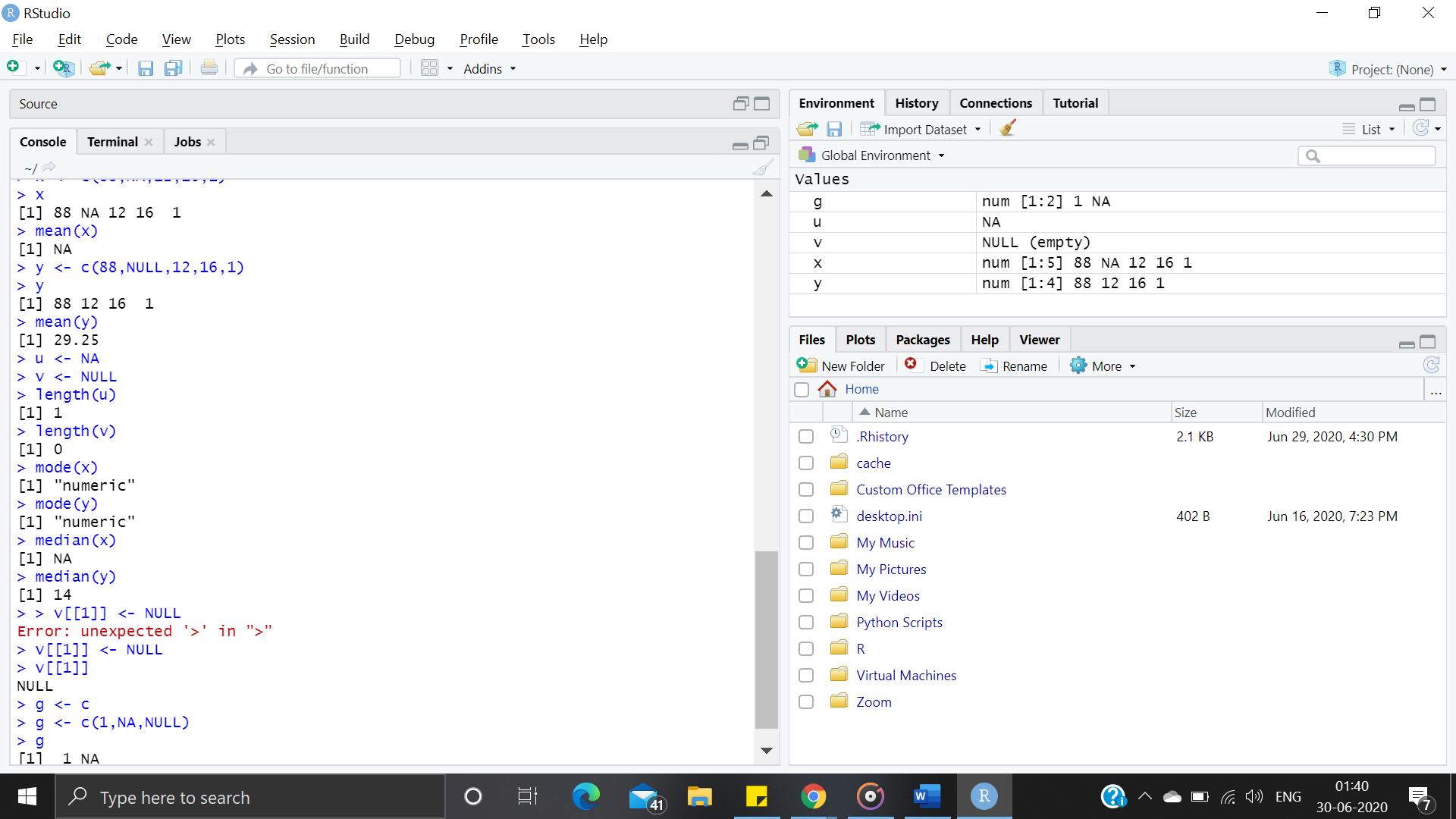
**NA:**

* Not Available (Missing data)
* It is a logical constant of length 1 which contains a missing value indicator.
* It can be coerced to any other vector type except raw.

**NULL:**

* Doesn’t exist cases, unknown
* It represents the null object in R: it is a reserved word.
* It is often returned by expressions and functions whose value is undefined.





**2. Create a sample R file and demonstrate the concepts like Variable creation, assignment operation and basic Data types available in R programming language**

