STAT 33B Lec Workbook Wk 11

Won Shil Park (3033452021)

Apr 3, 2021

This workbook is due **Apr 7, 2021** by 11:59pm PT.

- Knit and submit the generated PDF file on Gradescope.

File Formats

Exercise 1

The `volerup.txt` file that contains the volcano data can be found in bcourses in the "data" directory. Download the file to your local computer and inspect it with a plain text editor to confirm the file format.

Read the data set into R as a data frame. Use both `read.delim()` and `read.csv()` and `read.table()` to read the data into R. Save each data frame into `vol.delim`, `vol.csv`, and `vol.table`, respecitvely.

The data frames should have 36 columns and 835 rows/eruptions. You may need to provide arguments to the function calls, in addition to the location of the file.

HINT: Consider the values for the 'header', 'sep', and 'quote' parameters in your function calls.

```
vol.delim = read.delim("~/Downloads/volerup.txt")
vol.csv = read.csv("~/Downloads/volerup.txt", sep = "\t")
vol.table = read.table("~/Downloads/volerup.txt", header = TRUE, sep = "\t", quote = "\"")
```

Confirm that the three data frames have the same dimension, and that they columns are of the same class. Are these three data frames identical?

```
# check dim
dim(vol.delim)
```

```
## [1] 835  36
```

```
dim(vol.csv)
```

```
## [1] 835  36
```

```
dim(vol.table)
```

```
## [1] 835  36
```

```
# check names
names(vol.delim)
```

```
##  [1] "Year"                                "Month"
##  [3] "Day"                                 "TSU"
##  [5] "EQ"                                  "Name"
##  [7] "Location"                            "Country"
##  [9] "Latitude"                            "Longitude"
## [11] "Elevation"                           "Type"
## [13] "Status"                              "Time"
## [15] "VEI"                                 "Agent"
## [17] "DEATHS"                              "DEATHS_DESCRIPTION"
## [19] "MISSING"                             "MISSING_DESCRIPTION"
## [21] "INJURIES"                            "INJURIES_DESCRIPTION"
## [23] "DAMAGE_MILLIONS_DOLLARS"             "DAMAGE_DESCRIPTION"
## [25] "HOUSES_DESTROYED"                    "HOUSES_DESTROYED_DESCRIPTION"
## [27] "TOTAL_DEATHS"                        "TOTAL_DEATHS_DESCRIPTION"
## [29] "TOTAL_MISSING"                       "TOTAL_MISSING_DESCRIPTION"
## [31] "TOTAL_INJURIES"                      "TOTAL_INJURIES_DESCRIPTION"
## [33] "TOTAL_DAMAGE_MILLIONS_DOLLARS"       "TOTAL_DAMAGE_DESCRIPTION"
## [35] "TOTAL_HOUSES_DESTROYED"              "TOTAL_HOUSES_DESTROYED_DESCRIPTION"
```

```
names(vol.csv)
```

```
##  [1] "Year"                                "Month"
##  [3] "Day"                                 "TSU"
##  [5] "EQ"                                  "Name"
##  [7] "Location"                            "Country"
##  [9] "Latitude"                            "Longitude"
## [11] "Elevation"                           "Type"
## [13] "Status"                              "Time"
## [15] "VEI"                                 "Agent"
## [17] "DEATHS"                              "DEATHS_DESCRIPTION"
## [19] "MISSING"                             "MISSING_DESCRIPTION"
## [21] "INJURIES"                            "INJURIES_DESCRIPTION"
## [23] "DAMAGE_MILLIONS_DOLLARS"             "DAMAGE_DESCRIPTION"
## [25] "HOUSES_DESTROYED"                    "HOUSES_DESTROYED_DESCRIPTION"
## [27] "TOTAL_DEATHS"                        "TOTAL_DEATHS_DESCRIPTION"
## [29] "TOTAL_MISSING"                       "TOTAL_MISSING_DESCRIPTION"
## [31] "TOTAL_INJURIES"                      "TOTAL_INJURIES_DESCRIPTION"
## [33] "TOTAL_DAMAGE_MILLIONS_DOLLARS"       "TOTAL_DAMAGE_DESCRIPTION"
## [35] "TOTAL_HOUSES_DESTROYED"              "TOTAL_HOUSES_DESTROYED_DESCRIPTION"
```

```
names(vol.table)
```

```
##  [1] "Year"                                "Month"
##  [3] "Day"                                 "TSU"
##  [5] "EQ"                                  "Name"
##  [7] "Location"                            "Country"
##  [9] "Latitude"                            "Longitude"
## [11] "Elevation"                           "Type"
## [13] "Status"                              "Time"
```

```
## [15] "VEI"                                "Agent"
## [17] "DEATHS"                             "DEATHS_DESCRIPTION"
## [19] "MISSING"                            "MISSING_DESCRIPTION"
## [21] "INJURIES"                           "INJURIES_DESCRIPTION"
## [23] "DAMAGE_MILLIONS_DOLLARS"            "DAMAGE_DESCRIPTION"
## [25] "HOUSES_DESTROYED"                   "HOUSES_DESTROYED_DESCRIPTION"
## [27] "TOTAL_DEATHS"                       "TOTAL_DEATHS_DESCRIPTION"
## [29] "TOTAL_MISSING"                      "TOTAL_MISSING_DESCRIPTION"
## [31] "TOTAL_INJURIES"                     "TOTAL_INJURIES_DESCRIPTION"
## [33] "TOTAL_DAMAGE_MILLIONS_DOLLARS"      "TOTAL_DAMAGE_DESCRIPTION"
## [35] "TOTAL_HOUSES_DESTROYED"             "TOTAL_HOUSES_DESTROYED_DESCRIPTION"
```

```
# check if identical
identical(vol.delim, vol.csv)
```

```
## [1] TRUE
```

```
identical(vol.delim, vol.table)
```

```
## [1] TRUE
```

Exercise 2

Write the `vol.delim` data freme out as a CSV file called `vol2.csv`. use the function `write.csv()` to do this.

```
write.csv(vol.delim, file = "~/Downloads/vol2.csv", row.names = FALSE)
```

Read the csv file that you created back into R and save it as `vol2.csv`. Use `identical()` to confirm that `vol2.csv` hasn't changed from `vol.delim`. If it has, investigate what changed, and fix your call by specifying alternative values to some of the parameters.

```
vol.csv2 = read.csv("~/Downloads/vol2.csv")

identical(vol.csv2, vol.delim)
```

```
## [1] TRUE
```

```
dim(vol.csv2)
```

```
## [1] 835  36
```

```
names(vol.csv2)
```

```
##  [1] "Year"                                "Month"
##  [3] "Day"                                 "TSU"
##  [5] "EQ"                                  "Name"
##  [7] "Location"                            "Country"
##  [9] "Latitude"                            "Longitude"
```

```
## [11] "Elevation"                          "Type"
## [13] "Status"                             "Time"
## [15] "VEI"                                "Agent"
## [17] "DEATHS"                             "DEATHS_DESCRIPTION"
## [19] "MISSING"                            "MISSING_DESCRIPTION"
## [21] "INJURIES"                           "INJURIES_DESCRIPTION"
## [23] "DAMAGE_MILLIONS_DOLLARS"            "DAMAGE_DESCRIPTION"
## [25] "HOUSES_DESTROYED"                   "HOUSES_DESTROYED_DESCRIPTION"
## [27] "TOTAL_DEATHS"                       "TOTAL_DEATHS_DESCRIPTION"
## [29] "TOTAL_MISSING"                      "TOTAL_MISSING_DESCRIPTION"
## [31] "TOTAL_INJURIES"                     "TOTAL_INJURIES_DESCRIPTION"
## [33] "TOTAL_DAMAGE_MILLIONS_DOLLARS"      "TOTAL_DAMAGE_DESCRIPTION"
## [35] "TOTAL_HOUSES_DESTROYED"             "TOTAL_HOUSES_DESTROYED_DESCRIPTION"
```