

# Assignment 3: Volcanic Eruptions

MinJae Jo

2025-09-11

```
eruptions <- read_csv("eruptions_dataset.csv")
```

```
## Rows: 103 Columns: 16
## -- Column specification -----
## Delimiter: ","
## chr  (4): name, location, country, type
## dbl (12): year, month, day, latitude, longitude, elevation, VEI, deaths, mis...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

## Exercise 1

CSV (Comma-Separated Values) is a simple text format that stores data separated by commas. Each row represents one record (observation), and each column represents a variable (attribute). The file size is small and it is easy to read and write in various tools such as Excel, R, and Python, so it is highly compatible. However, if commas are included in the value, attention should be paid to quotation processing, encoding differences, and line modulation problems.

## Exercise 2

- i. 103 rows and 16 columns
- ii. Each rows shown volcanic eruption event that records of one eruption on a particular volcano on a particular year, month, or day.
- iii. Elevation is recorded in meters, so it is a metric system.

## Exercise 3

- i.

```
library(dplyr)

eruptions %>%
  select(name, elevation)
```

```
## # A tibble: 103 x 2
##   name          elevation
##   <chr>         <dbl>
## 1 Tungurahua      5023
## 2 Eyjafjallajokull 1651
## 3 Pacaya          2569
## 4 Zealandia Bank    0
## 5 Karangetang      1797
## 6 Sinabung         2460
## 7 Merapi           2910
## 8 Tungurahua      5023
## 9 Tengger Caldera  2329
## 10 Merapi          2910
## # i 93 more rows
```

ii. `eruptions %>% select(name:elevation)` ‘name:elevation’ selects all columns from ‘name’ to ‘elevation’ based on the order of the columns. Therefore, a total of six columns are output: ‘name, location, country, latitude, longitude, elevation’.

iii. `eruptions_stored <- eruptions %>% select(name, elevation)`

```
glimpse(eruptions_stored)
```

## Exercise 4

i.

```
eruptions %>%
  arrange(year, month)
```

```
## # A tibble: 103 x 16
##   year month   day name      location country latitude longitude elevation type
##   <dbl> <dbl> <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl> <chr>
## 1 2010     1    NA Tungur~ Ecuador Ecuador   -1.47    -78.4     5023 Stra~
## 2 2010     3    31 Eyjafj~ Iceland~ Iceland    63.6    -19.6     1651 Stra~
## 3 2010     5    27 Pacaya Guatemala~ Guatem~    14.4    -90.6     2569 Comp~
## 4 2010     5    29 Zealan~ Mariana~ United~    16.9     146.         0 Stra~
## 5 2010     8     6 Karang~ Sangihe~ Indone~     2.78    125.     1797 Stra~
## 6 2010     8    30 Sinabu~ Sumatra Indone~     3.17     98.4     2460 Stra~
## 7 2010    10    26 Merapi Java      Indone~    -7.54    110.     2910 Stra~
## 8 2010    11    NA Tungur~ Ecuador Ecuador   -1.47    -78.4     5023 Stra~
## 9 2010    12    28 Tengge~ Java      Indone~    -7.94    113.     2329 Stra~
## 10 2011     1     3 Merapi Java      Indone~    -7.54    110.     2910 Stra~
## # i 93 more rows
## # i 6 more variables: VEI <dbl>, deaths <dbl>, missing <dbl>, injuries <dbl>,
## #   damage <dbl>, houses_destroyed <dbl>
```

ii. `eruptions %>% arrange(month, year)`

iii. Depending on the order in which you use it, the sorting results will be different.

## Exercise 5

i.

```
eruptions %>%  
  arrange(desc(name))
```

```
## # A tibble: 103 x 16  
##   year month   day name      location country latitude longitude elevation type  
##   <dbl> <dbl> <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl> <chr>  
## 1 2010     5    29 Zealan~ Mariana~ United~    16.9      146.         0 Stra~  
## 2 2016     6     9 Yellow~ US-Wyom~ United~    44.4     -111.       2805 Cald~  
## 3 2019    12     9 Whakaa~ New Zea~ New Ze~   -37.5      177.        294 Stra~  
## 4 2019     6    26 Ulawun  New Bri~ Papua ~    -5.05     151.       2334 Stra~  
## 5 2013     9     1 Ubinas  Peru     Peru    -16.4     -70.9       5672 Stra~  
## 6 2010     1    NA Tungur~ Ecuador Ecuador   -1.47     -78.4       5023 Stra~  
## 7 2010    11    NA Tungur~ Ecuador Ecuador   -1.47     -78.4       5023 Stra~  
## 8 2011     4    NA Tungur~ Ecuador Ecuador   -1.47     -78.4       5023 Stra~  
## 9 2012    12    12 Tolbac~ Kamchat~ Russia    55.8      160.       3611 Shie~  
## 10 2010    12    28 Tengge~ Java     Indone~   -7.94     113.       2329 Stra~  
## # i 93 more rows  
## # i 6 more variables: VEI <dbl>, deaths <dbl>, missing <dbl>, injuries <dbl>,  
## #   damage <dbl>, houses_destroyed <dbl>
```

ii. `eruptions %>% filter(name == "Yellowstone") %>% select(year, month, day, name, location)`

## Exercise 6

- i. `eruptions_stored %>% mutate( elevation_yards = elevation * 1.093 )`
- ii. The new column is added to the last column with the name 'elevation\_yards'

## Exercise 7

- i. `eruptions %>% filter(damage > 0 & longitude < 0)`
- ii. `eruptions %>% filter(damage > 0 & longitude >= 0)`

## Exercise 8

- i. `eruptions %>% group_by(country) %>% summarize( average_deaths = mean(deaths, na.rm = TRUE) )`
- ii. `eruptions %>% group_by(country) %>% summarize( average_deaths = mean(deaths, na.rm = TRUE), total_deaths = sum(deaths, na.rm = TRUE) )`
- iii. `eruptions %>% group_by(country) %>% summarize( average_deaths = mean(deaths, na.rm = TRUE), total_deaths = sum(deaths, na.rm = TRUE) ) %>% arrange(average_deaths)`

## Exercise 9