STAT 33B Workbook 2

Ming Fong (3035619833)

Sep 10, 2020

This workbook is due **Sep 10, 2020** by 11:59pm PT.

The workbook is organized into sections that correspond to the lecture videos for the week. Watch a video, then do the corresponding exercises *before* moving on to the next video.

Workbooks are graded for completeness, so as long as you make a clear effort to solve each problem, you'll get full credit. That said, make sure you understand the concepts here, because they're likely to reappear in homeworks, quizzes, and later lectures.

As you work, write your answers in this notebook. Answer questions with complete sentences, and put code in code chunks. You can make as many new code chunks as you like.

In the notebook, you can run the line of code where the cursor is by pressing `Ctrl` + `Enter` on Windows or `Cmd` + `Enter` on Mac OS X. You can run an entire code chunk by clicking on the green arrow in the upper right corner of the code chunk.

Please do not delete the exercises already in this notebook, because it may interfere with our grading tools.

You need to submit your work in two places:

- Submit this Rmd file with your edits on bCourses.
- Knit and submit the generated PDF file on Gradescope.

File Systems

Watch the "File Systems" lecture video.

Exercise 1

For each of the following paths, say whether the path is absolute or relative, and explain how you can tell.

1. `"./documents"`
2. `"/Users/Jun/doggos_to_pet.md"`
3. `"TODO.md"`

YOUR ANSWER GOES HERE:

1. Relative. The path starts with a `.` which searches for `"/documents"` from the working directory
2. Absolute. The path starts at the root `/`.
3. Relative. The path searches for `"TODO.md"` from the current working directory.

The R Working Directory

Watch the "The R Working Directory" lecture video.

Exercise 2

1. What's the root directory called on your computer?
2. What's the absolute path to the home directory on your computer?
3. Use R to count the total number of files in your home directory. Your code should return the result as a number.

YOUR ANSWER GOES HERE:

1. `"C:/"`
2. `""C:\\Users\\mingf\\Documents""`

```
normalizePath("~")
```

```
## [1] "C:\\Users\\mingf\\Documents"
```

3. There are 14 files in my home directory. See below:

```
length(list.files("~"))
```

```
## [1] 14
```

Data Frames

Watch the "Data Frames" lecture video.

Exercise 3

1. Load the dogs data set `dogs.rds` into R (this one has more than 10 rows).
2. What's the mean weight of the dogs? You can use the `na.rm` parameter in the `mean()` function to make the function ignore missing values.
3. The `which.min()` function returns the index of the minimum element of a vector. Which breed of dog has the shortest height?
4. Which breed of dog has the longest lifespan?

YOUR ANSWER GOES HERE:

1.

```
dogs = readRDS("data/dogs.rds")
```

2.

```
mean(dogs$weight, na.rm = TRUE)
```

```
## [1] 44.97093
```

3. Chihuahua

```
dogs[which.min(dogs$height), 1]
```

```
## [1] "Chihuahua"
```

4. Chihuahua

```
dogs[which.max(dogs$longevity), 1]
```

```
## [1] "Chihuahua"
```

Factors

Watch the "Factors" lecture video.

No exercises for this video. Get up, stretch, and take a break! :)

File Formats

Watch the "File Formats" lecture video.

Exercise 5

1. Load the volcano data set into R.
2. What are the column names? Use R to get these rather than typing them out yourself.
3. How many volcano eruptions are recorded in the data set?
4. What are the classes/types of the columns? *Hint: an earlier lecture mentioned a function that summarizes of this information.*
5. Are there any columns that contain categorical data? Are these columns factors? If not, what are their classes?

YOUR ANSWER GOES HERE:

1.

```
volcanoes = read.csv("data/volcano.csv")
```

2.

```
colnames(volcanoes)
```

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

3. 835

```
nrow(volcanoes)
```

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

4.

```r
str(volcanoes)
```

```
## 'data.frame':    835 obs. of  37 variables:
##  $ X                                 : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Year                              : int  -4360 -4350 -4050 -4000 -3580 -3550 -2420 -2040 -1900 -18
##  $ Month                             : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Day                               : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ TSU                               : chr  "" "" "" "" ...
##  $ EQ                                : chr  "" "" "" "" ...
##  $ Name                              : chr  "Macauley Island" "Kikai" "Masaya" "Pago" ...
##  $ Location                          : chr  "Kermadec Is" "Ryukyu Is" "Nicaragua" "New Britain-SW Pa
##  $ Country                           : chr  "New Zealand" "Japan" "Nicaragua" "Papua New Guinea" ...
##  $ Latitude                          : num  -30.2 30.78 11.98 -5.58 14 ...
##  $ Longitude                         : num  -178.5 130.3 -86.2 150.5 121 ...
##  $ Elevation                         : int  238 717 635 742 400 1486 1281 1280 1032 1905 ...
##  $ Type                              : chr  "Caldera" "Caldera" "Caldera" "Caldera" ...
##  $ Status                            : chr  "Holocene" "Historical" "Historical" "Historical" ...
##  $ Time                              : chr  "U" "D1" "D1" "D2" ...
##  $ VEI                               : int  6 7 6 6 6 6 5 6 6 6 ...
##  $ Agent                             : chr  "" "P" "" "T" ...
##  $ DEATHS                            : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ DEATHS_DESCRIPTION                : int  NA 3 NA 1 NA NA NA NA NA NA ...
##  $ MISSING                           : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ MISSING_DESCRIPTION               : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ INJURIES                          : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ INJURIES_DESCRIPTION              : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ DAMAGE_MILLIONS_DOLLARS           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ DAMAGE_DESCRIPTION                : int  NA 3 NA 1 NA NA NA NA NA NA ...
##  $ HOUSES_DESTROYED                  : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ HOUSES_DESTROYED_DESCRIPTION      : int  NA 3 NA NA NA NA NA NA NA NA ...
##  $ TOTAL_DEATHS                      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ TOTAL_DEATHS_DESCRIPTION          : int  NA 3 NA 1 NA NA NA NA NA NA ...
##  $ TOTAL_MISSING                     : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ TOTAL_MISSING_DESCRIPTION         : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ TOTAL_INJURIES                    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ TOTAL_INJURIES_DESCRIPTION        : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ TOTAL_DAMAGE_MILLIONS_DOLLARS     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ TOTAL_DAMAGE_DESCRIPTION          : int  NA 3 NA 1 NA NA NA NA NA NA ...
##  $ TOTAL_HOUSES_DESTROYED            : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ TOTAL_HOUSES_DESTROYED_DESCRIPTION: int  NA 3 NA NA NA NA NA NA NA NA ...
```

5.