Math 300 Lesson 1 Notes

Data with R

YOUR NAME HERE

June, 2022

Contents

Objectives

1. Install, load, and use R packages.
2. Explore data sets with R functions to include glimpse(), View(), kable() and $.
3. Identified and justify if a variable is for identification or measurement and if measurement if it is categorical or quantitative.
4. Understand and practice the tips on learning to code to include using help functions and reading error messages

Reading

Chapter 1

Lesson

Work through the learning checks LC1.1 - LC1.7. Complete the code when necessary.

LC 1.1 (Objective 1)

(LC 1.1) Repeat the above installing steps, but for the `dplyr`, `nycflights13`, and `knitr` packages. This will install the earlier mentioned `dplyr` package, the `nycflights13` package containing data on all domestic flights leaving a NYC airport in 2013, and the `knitr` package for writing reports in R.

- These may be installed. Know how to load a package using the Packages tab in RStudio. *

LC 1.2 (Objective 1)

(LC 2.1) "Load" the `dplyr`, `nycflights13`, and `knitr` packages as well by repeating the above steps.

Solution: Complete the code by removing the comment symbol # and typing in the appropriate library. Does the order matter?

```
#library(______)
#library(______)
#library(_______)
```

LC 1.3 (Objective 2)

(LC 1.3) What does any *ONE* row in this `flights` dataset refer to?

- A. Data on an airline
- B. Data on a flight
- C. Data on an airport
- D. Data on multiple flights

Solution:

LC 1.4 (Objective 3, 4)

(LC 1.4) What are some examples in this dataset of **categorical** variables? What makes them different than **quantitative** variables?

Solution: Hint: Type `?flights` in the console to see what all the variables mean!

- Categorical:
- Quantitative:

LC 1.5 (Objective 3)

(LC 1.5) What properties of the observational unit do each of `lat`, `lon`, `alt`, `tz`, `dst`, and `tzone` describe for the `airports` data frame? Note that you may want to use `?airports` to get more information.

Solution:

LC 1.6 (Objective 3)

(LC 1.6) Provide the names of variables in a data frame with at least three variables in which one of them is an identification variable and the other two are not. In other words, create your own tidy dataset that matches these conditions.

Solution: Complete the code by removing the comment symbol # and entering what you think an appropriate value would be for the missing value of the variable.

```
# LC6 <- tibble(id=c(1,2,____),gpa=c(__________,2.7,3.6),pea=c(2.7,_______,3.3))
```

```
# glimpse(LC6)
```

- In the example,`id` is an XXXX variable as it identifies the observation in question.
- Anything else pertains to XXXXX.

We can also look at the `weather` data object.

```
# glimpse(weather)
```

- The combination of XXXX, XXXX, XXXXX, XXXX, XXXXX are identification variables as they identify the observation in question.
- Anything else pertains to measurements.

LC 1.7 (Objective 2, 4)

(LC 1.7) Look at the help file for the `airports` data frame. Revise your earlier guesses about what the variables `lat`, `lon`, `alt`, `tz`, `dst`, and `tzone` each describe.

Solution:

Documenting software

- File creation date: 2022-06-04
- R version 4.1.3 (2022-03-10)
- `dplyr` package version: 1.0.9
- `nycflights13` package version: 1.0.2