

Homework 1

Stats 32: Introduction to R for Undergraduates

Due: Thursday, April 11, 2024, 11:59 am PT

Instructions

You may (in fact, are encouraged to) use the Internet (including AI assistants, like ChatGPT) to search up any information to help you with this assignment, though you must cite any external (i.e. non-course related) resources that you use. Similarly, *after attempting this assignment by yourself*, you may collaborate with other students in the course, but you must each write your own code and acknowledge all students with whom you collaborated *for each problem* (you don't need to cite by subpart). However, you may not post on Internet forums (e.g. Stack Exchange) for help with this assignment; doing so is considered an Honor Code violation. You also may not copy verbatim any significant amount of code from the Internet (including AI assistants, like ChatGPT), even with citation. Feeding in the problems directly into AI assistants (or substantively paraphrased version) is also not permitted.

Please provide your code responses to each problem in the `.Rmd` file in the R code chunks directly below each subpart, inserting additional R code chunks if needed. Any text response can go right underneath the corresponding question.

On Gradescope, please submit a single `.pdf` file created by knitting the document with your responses. Problem 0 will provide guidance on how to do this.

Credit is given based on the approach and code, not necessarily the final answer.

Problem 0

Read Sections 2.2-2.6 of R Markdown: The Definitive Guide. It will introduce you to using R markdown. You may skip subsection 2.5.3 if you are not familiar with LaTeX. It is not needed for this class.

Problem 1

- (a) [1 point] Load the `nycflights13` package.
- (b) [2 points] Consider the `airports` tibble (data frame). It is automatically loaded into the environment when you complete part (a). How many airports are in the tibble?
- (c) [2 points] What is the mean altitude of all the airports in the tibble? Give the units.
- (d) [2 points] How many different time zones are represented among the airports in the tibble? Hint: Look at the help page for `unique()`.
- (e) [2 points] Select the `name`, `lat`, `lon`, and `alt` of the 3rd through 10th airports in the tibble. The result should be a 8 x 4 tibble.
- (f) [2 points] How many airports are in the “America/New_York” time zone?

Problem 2

- (a) [4 points] Write a function called `percent_decrease` that takes in two arguments — `num` and `percent` — and returns the value when `num` is decreased by `percent` percent. For instance, if `num` is 10 and `percent` is 20, your function should return 8.
- (b) [2 points] Use your function from (a) to calculate the cost of an item of clothing which originally sold for \$49.95, but is on sale for 35% off. Use `round()` to print out the price to the nearest **cent**.

Problem 3

- (a) [4 points] Run the following three lines of code:

```
scores <- c(89, 72, 96, 93, 100, 68, 82, 69, 77, 79, 78, 90)
x <- scores > 90
y <- scores[scores > 90]
```

Explain what kind of data structure `x` is, the dimension(s) of that data structure, and the data type(s) of the objects (variables) within that data structure. Repeat for `y`. Use code to justify your answers.

- (b) [2 points] Suppose that `scores` is a vector of exam scores for students in a recent class. In light of this, interpret (in words) the content of the objects `x` and `y` above.
- (c) [2 points] Describe (in words) what the following line of code does.

```
scores[-2]
```

```
## [1] 89 96 93 100 68 82 69 77 79 78 90
```