S&DS 220: Homework 1

Braeden Cullen

2024-01-19

Instructions

- This assignment is **due Friday January 19th, 11:59 pm**[1] in Gradescope. Download the R Markdown version of this document from Canvas (called "sds220_s24_hw01.rmd").
- Be sure to put you name in the "author" section of the YAML header above.
- This homework assignment consists of two parts.
    1. In Part I, you will work through R tutorials in the `swirl` package at your own pace. This will not be turned in, and you do not need to complete it before doing Part II.
    2. In Part II, you will be working the `nhtemp` data set, which is built into R. There are 5 questions, each worth 20 points. For each question, your solution will consist of writing R code (in a code chunk) or writing an explanation (outside of a code chunk after the prompt).
- Submit your solutions in Gradescope as a pdf file generated from this document by clicking the *Knit* button at the top of the window.

[1] There is a 48 hour grace period, so you have until the end of Sunday.

Part I: `swirl` package tutorials

First you need to install the `swirl` package. In the *console* window, type the command

```
install.packages("swirl")
```

This will install the `swirl` package on your computer. R packages are written by R users and shared with the R community through a central repository. They contain extra tools and expand the functionality base R. Some packages focus on specialized applications, including finance, sports, astronomy, and gene sequencing (and much much more).

Here is a list of current R packages. As you can see, the list is quite extensive. In addition to this, some users create packages and share them from their Github account, rather than using the central CRAN repository.

When you download a package, it goes into a folder on your computer called a *library*. However its contents are not available immediately. Each time you open a new R session, you can make a package available with the `library` function. Think of the folder that contains R packages as a library (that's what R calls it after all!), and the individual packages as books. The `library` function is like going to the library and checking out a book (package).

Load the `swirl` package by typing the following command on the command line in the console:

```
library(swirl)
```

You will be greeted with a welcome message and prompted to type in the command

```
swirl()
```

to start the tutorial. Do so, and follow the prompts. When you get to the list of courses, choose the first option:

1: R Programming: The basics of programming in R

This course consists of 15 lessons, each of which takes about 10-15 minutes to complete. Work through these lessons at your own pace. You do not need to complete them all before moving on to Part II of this assignment. I recommend doing several lessons each day (or every couple of days) over the next two weeks. You are welcome to do more and dive into the other swirl courses! The more time you put into learning R, the easier it will become and the more you'll be able to do.

Part II: Basic R

Question 1

Write R code to create a numeric vector that repeats the sequence `1, 1, 2, 3, 5` eight hundred times. Store the result as `fib`.

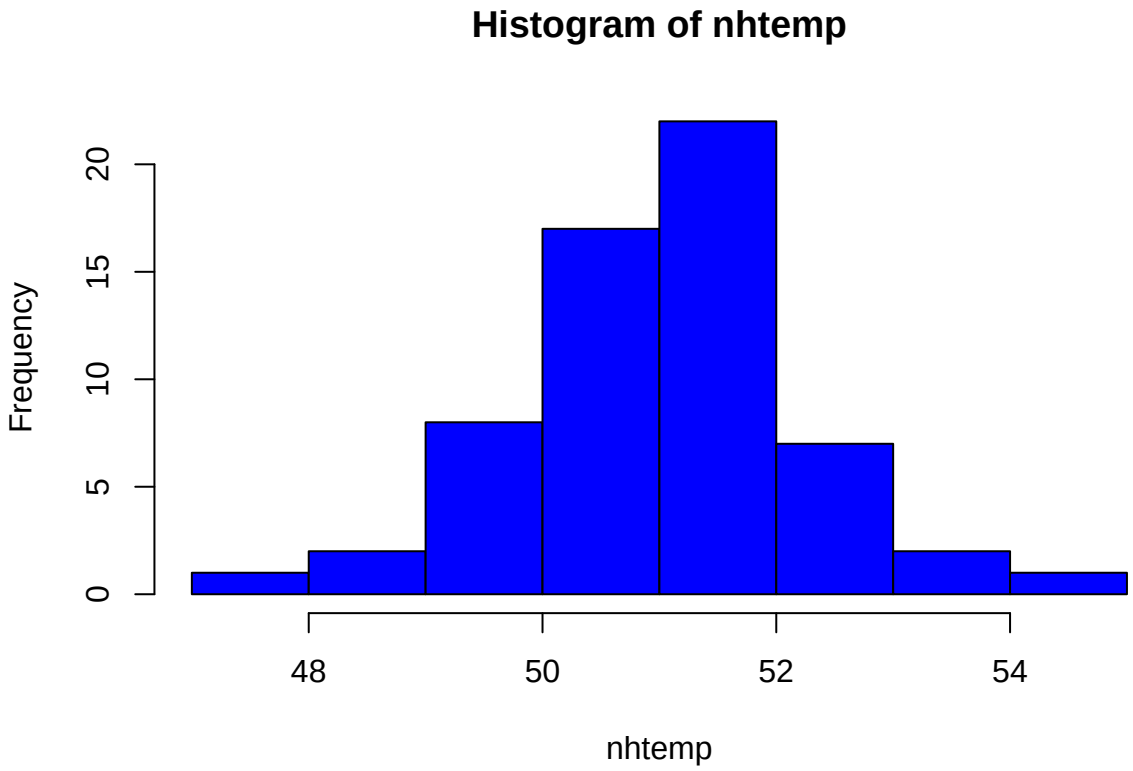
```r
# your answer here
fib <- rep(c(1, 1, 2, 3, 5), times = 800)
```

Question 2

`nhtemp` is a dataset that is built into R. Create a histogram of `nhtemp` using the `hist` function. Set the color of the histogram to a color that you like (but something other than the default).

```
# your answer here

hist(nhtemp, col="blue")
```



Question 3

In the console window call the `help` function on `nhtemp`. What is range of years for the data set?

Your answer here: 1912 to 1971

Call the function `str` on `nhtemp` to see what type of object it is.

```
# your answer here
str(nhtemp)
```

```
##  Time-Series [1:60] from 1912 to 1971: 49.9 52.3 49.4 51.1 49.4 47.9 49.8 50.9 49.3 51.9 ...
```

Question 4

What is the mean temperature in the `nhtemp` data?

```
# your answer here
mean(nhtemp)
```

```
## [1] 51.16
```

```
51.16
```

```
## [1] 51.16
```

What is the standard deviation of the temperatures in the `nhtemp` data?

```
# your answer here
sd(nhtemp)
```

```
## [1] 1.265608
```

```
1.265608
```

```
## [1] 1.265608
```

What is the maximum temperature in the `nhtemp` data?

```
# your answer here
max(nhtemp)
```

```
## [1] 54.6
```

```
54.6
```

```
## [1] 54.6
```

Which year had the maximum temperature in the `nhtemp` data? Do not manually inspect the data for the largest value. Use the function `which.max` and figure which year this corresponds to.

```
# your answer here
# which.max(nhtemp)
1912 + which.max(nhtemp) - 1 # -1 to account for 0 indexing
```

```
## [1] 1953
```

Question 5

- Call the `plot` function on `nhtemp`.
- Change the label on the x-axis to "Year" with the `xlab` argument.
- Change the label on the y-axis to "Degrees Fahrenheit" with the `ylab` argument.
- Give the plot the title "Temperatures in New Haven, CT from 1912 to 1971" with the `main` argument.

```
# your answer here
plot(nhtemp, xlab = "Year", ylab = "Degrees Fahrenheit", main = "Temperatures in New Haven, CT from 191
```