Name: Your name here
Due: 2024/09/16

Homework 1

Be sure to submit **both** the .pdf and .qmd file to Canvas by Monday, September 16th at 11:59 pm. Additionally, when possible, your answers should be denoted with a callout box. For example, if the question is, "What is the meaning of life?"

> 42

1. [1 pt] Run the following code so that I know what version of R you have installed.

```
sessionInfo()
```

```
R version 4.3.0 (2023-04-21 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default


locale:
[1] LC_COLLATE=English_United States.utf8
[2] LC_CTYPE=English_United States.utf8
[3] LC_MONETARY=English_United States.utf8
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.utf8

time zone: America/New_York
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
 [1] compiler_4.3.0  fastmap_1.1.1   cli_3.6.1       tools_4.3.0
 [5] htmltools_0.5.5 rstudioapi_0.14 yaml_2.3.7      rmarkdown_2.25
 [9] knitr_1.42      jsonlite_1.8.4  xfun_0.39       digest_0.6.31
[13] rlang_1.1.1     evaluate_0.21
```

2. [10 pt] For this assignment, we will use a data set describing weather station measurements at the Knapp State Airport, west of Barre, VT. This data set is publicly available and provided by the

state of Vermont. The purpose of this question is to import and manicure the data, so that we have a clean time series for analysis later in the assignment.

a) [1 pt] Import the Barre weather station data and print the first six rows.

b) [1 pt] Create a new data frame, called `barre_clean`, that includes a `year`, `month`, and `day` column.

c) [2 pt] It is always advised to inspect the data before manipulating it. Create a table of the number of observed days by year. Filter to years with fewer than 365 days and greater than 366 days and print the resulting data frame. Do you notice anything strange?

d) [2 pt] Create a new data frame, called `barre_monthly`, that summarizes the total sum of snowfall in each month (you may ignore any oddities you found in the previous question, for now). The resulting data frame should have a `sum_snowfall` column, a `year` column, and a `month` column. Arrange the data frame by year and month. Print the first and last six rows of `barre_monthly`.

e) [1 pt] Create a histogram of the `sum_snowfall` in February. Does anything strike you as odd?

f) [1 pt] What years contain 0 sum total snowfall in February? Does anything strike you as odd?

g) [2 pt] Create a `ts` object containing the sum total monthly snowfall between 1950 and 1996 (hopefully you are convinced that we should not include `1997:2015`), called `barre_ts`. What is the frequency of this time series?

3. [5 pt] You should now have a clean `ts` object, called `barre_ts`. Using this object, describe the sum total monthly snowfall (in) at the Knapp State Airport weather station. Be sure to reference any apparent trends or seasonality in your response, including figures that support your statements.