Homework 8

Benjamin Anderson II

```
library(tidyverse)
library(nycflights13)
```

Tasks that require an answer are bolded (inside ** in the .qmd file). For any task that includes a question (i.e. it ends with "?"), you should also answer the question in sentence form.

Logical Vectors

The first part of the homework uses the `flights` and `airports` data in the nycflights13 package.

1.

(2 pts)

For each of the following, filter `flights` to **find the number of flights that match the criteria**. An easy way to return just the number of flights is to pipe your resulting table into `nrow()`.

a) **Flights that flew from from JFK to PDX.**

```
flights <- nycflights13::flights
flights |>
  filter(origin == "JFK" & dest == "PDX") |>
  nrow()
```

```
[1] 783
```

There are 783 flights that flew from JFK to PDX

b) **Flights that arrive early, but have a arrival time that is greater than the scheduled arrival time.**

```
flights |>
  filter(arr_delay < 0 & arr_time > sched_arr_time) |>
  nrow()
```

```
[1] 777
```

There are 777 flights where the flight arrived early, yet had an arrival time greater than the scheduled arrival time.

c) **Flights that depart in a odd numbered month.**

```
flights |>
  filter(month %% 2 == 1)  |>
  nrow()
```

```
[1] 168901
```

There are 168901 flights that departed on an odd month

d) **Flights that departed on Friday the 13th.** *(In 2013, Friday the 13th occurred in September and December).*

```
flights |>
  filter(wday(paste(year, month, day, sep = "-")) == 5 & day == 13) |>
  nrow()
```

```
[1] 989
```

There were 989 flights that departed on Friday the 13th.

2. How many flights a month went Hawaii?

(2 pts)

First, you'll need to find all the airport codes for airports in Hawaii. **Filter the `airports` data to airports where the `tz` column has the value `-10`, and then, `pull()` the `faa` column. Store the results in `hawaii_codes`.**

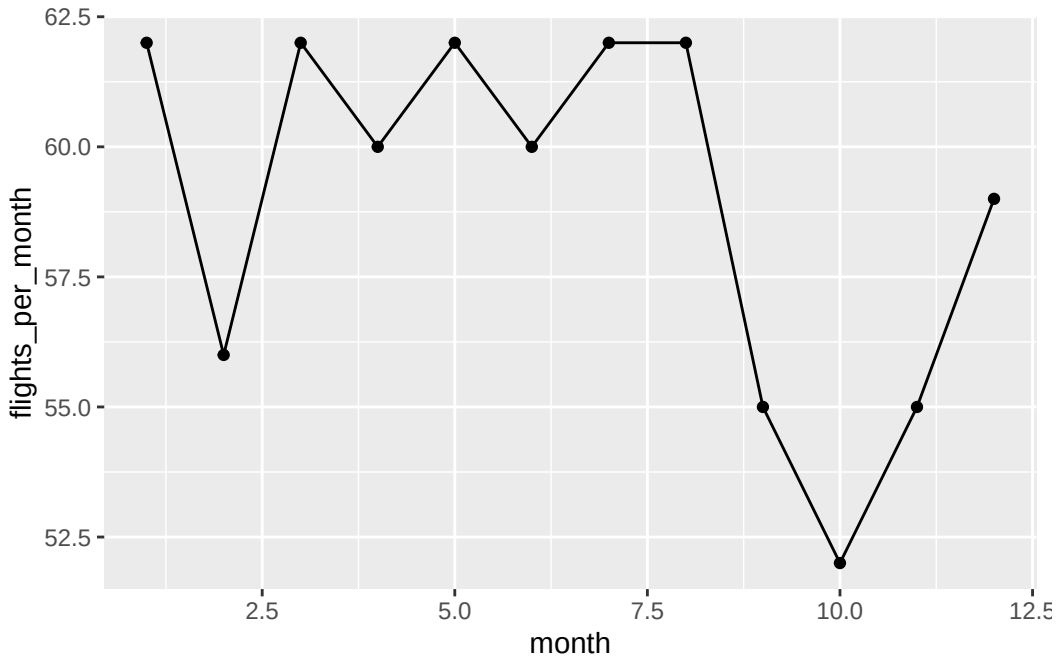
```r
airports <- nycflights13::airports
hawaii_codes <- airports |>
  filter(tz == -10) |>
  pull("faa")
```

Extract the rows in `flights` with a destination airport in the `hawaii_codes` vector. Store the result in `hawaii_flights`. Hint: the `%in%` operator is a logical operator (like `==` or `<`) that can be used to check if a value is equal to one of the values in a vector.

```r
hawaii_flights <- flights |>
  filter(dest %in% hawaii_codes)
```

Summarise `hawaii_flights` by counting the number of flights in each month, and then, produce a scatterplot of the number of flights by month.

```r
hawaii_flights |>
  group_by(month) |>
  summarize(flights_per_month = n()) |>
  ggplot(mapping = aes(x = month, y = flights_per_month)) +
  geom_point() +
  geom_line()
```



I know a scatterplot is asked of, but I feel the line graph is easier to read. The underlying scatterplot is still present though.

3. What proportion of flights went to Hawaii?

(2 pts)

A criticism of the previous plot might be that it just shows the absolute number of flights to Hawaii, not what proportion of all flights head to Hawaii.

Rather than filtering `flights` to flights heading to Hawaii, add a column called `hawaii` to `flights` that is `TRUE` if the flight has a destination in Hawaii.

```
flights <- flights |>
  mutate(hawaii = ifelse(dest %in% hawaii_codes, TRUE, FALSE))
```

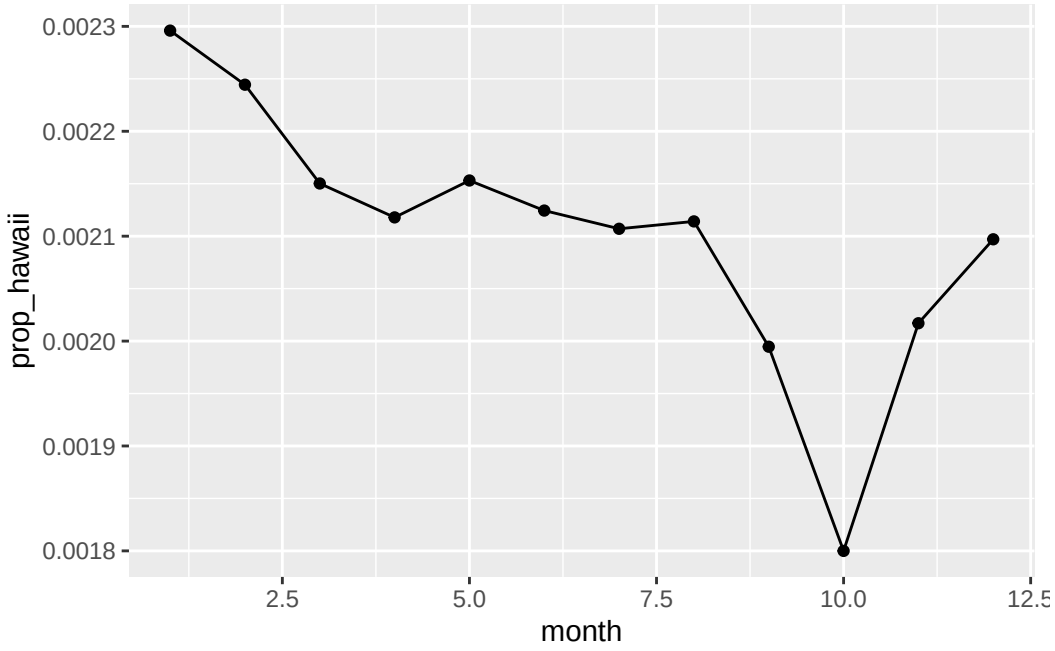
Summarise `flights` by computing the `mean()` of the `hawaii` column in each month. What does this value represent?

```
mean_hawaii <- flights |>
  group_by(month) |>
  summarize(prop_hawaii = mean(hawaii))
```

This value represents the proportion of flights per month that go to Hawaii in the dataset.

Create a scatterplot of the `mean()` of the `hawaii` column by month. Does it tell a different story to your previous plot?

```
mean_hawaii |>
  ggplot(mapping = aes(x = month, y = prop_hawaii)) +
  geom_point() +
  geom_line()
```



This plot does tell a different story, though not terribly different. There is still the least amount of travel to Hawaii during October, but rather than having the highest percentage of flights go to hawaii during Spring and Summer it happens during Winter, January specifically.

Lists

4.

(2 pts)

You saw that tibbles are also considered lists, so this means you can use the same subsetting tools (`[`, `[[`, `$`) on them as you can on lists.

Extract the 1st element of `starwars` with `[`.

```
starwars[1]
```

```
# A tibble: 87 x 1
   name
   <chr>
 1 Luke Skywalker
```

```
 2 C-3PO
 3 R2-D2
 4 Darth Vader
 5 Leia Organa
 6 Owen Lars
 7 Beru Whitesun Lars
 8 R5-D4
 9 Biggs Darklighter
10 Obi-Wan Kenobi
# i 77 more rows
```

```
# OR starwars["name"]
```

Extract the 1st element of starwars with [[.

```
starwars[[1]]
```

```
 [1] "Luke Skywalker"        "C-3PO"                 "R2-D2"
 [4] "Darth Vader"           "Leia Organa"           "Owen Lars"
 [7] "Beru Whitesun Lars"    "R5-D4"                 "Biggs Darklighter"
[10] "Obi-Wan Kenobi"        "Anakin Skywalker"      "Wilhuff Tarkin"
[13] "Chewbacca"             "Han Solo"              "Greedo"
[16] "Jabba Desilijic Tiure" "Wedge Antilles"        "Jek Tono Porkins"
[19] "Yoda"                  "Palpatine"             "Boba Fett"
[22] "IG-88"                 "Bossk"                 "Lando Calrissian"
[25] "Lobot"                 "Ackbar"                "Mon Mothma"
[28] "Arvel Crynyd"          "Wicket Systri Warrick" "Nien Nunb"
[31] "Qui-Gon Jinn"          "Nute Gunray"           "Finis Valorum"
[34] "Padmé Amidala"         "Jar Jar Binks"         "Roos Tarpals"
[37] "Rugor Nass"            "Ric Olié"              "Watto"
[40] "Sebulba"               "Quarsh Panaka"         "Shmi Skywalker"
[43] "Darth Maul"            "Bib Fortuna"           "Ayla Secura"
[46] "Ratts Tyerel"          "Dud Bolt"              "Gasgano"
[49] "Ben Quadinaros"        "Mace Windu"            "Ki-Adi-Mundi"
[52] "Kit Fisto"             "Eeth Koth"             "Adi Gallia"
[55] "Saesee Tiin"           "Yarael Poof"           "Plo Koon"
[58] "Mas Amedda"            "Gregar Typho"          "Cordé"
[61] "Cliegg Lars"           "Poggle the Lesser"     "Luminara Unduli"
[64] "Barriss Offee"         "Dormé"                 "Dooku"
[67] "Bail Prestor Organa"   "Jango Fett"            "Zam Wesell"
[70] "Dexter Jettster"       "Lama Su"               "Taun We"
[73] "Jocasta Nu"            "R4-P17"                "Wat Tambor"
```

```
[76] "San Hill"              "Shaak Ti"              "Grievous"
[79] "Tarfful"               "Raymus Antilles"       "Sly Moore"
[82] "Tion Medon"            "Finn"                  "Rey"
[85] "Poe Dameron"           "BB8"                   "Captain Phasma"
```

```
# OR starwars[["name"]]
```

Extract the 1st element of `starwars` with `$`. (You'll have to figure out what this element is called to use `$`).

```
starwars$name
```

```
 [1] "Luke Skywalker"        "C-3PO"                 "R2-D2"
 [4] "Darth Vader"           "Leia Organa"           "Owen Lars"
 [7] "Beru Whitesun Lars"    "R5-D4"                 "Biggs Darklighter"
[10] "Obi-Wan Kenobi"        "Anakin Skywalker"      "Wilhuff Tarkin"
[13] "Chewbacca"             "Han Solo"              "Greedo"
[16] "Jabba Desilijic Tiure" "Wedge Antilles"        "Jek Tono Porkins"
[19] "Yoda"                  "Palpatine"             "Boba Fett"
[22] "IG-88"                 "Bossk"                 "Lando Calrissian"
[25] "Lobot"                 "Ackbar"                "Mon Mothma"
[28] "Arvel Crynyd"          "Wicket Systri Warrick" "Nien Nunb"
[31] "Qui-Gon Jinn"          "Nute Gunray"           "Finis Valorum"
[34] "Padmé Amidala"         "Jar Jar Binks"         "Roos Tarpals"
[37] "Rugor Nass"            "Ric Olié"              "Watto"
[40] "Sebulba"               "Quarsh Panaka"         "Shmi Skywalker"
[43] "Darth Maul"            "Bib Fortuna"           "Ayla Secura"
[46] "Ratts Tyerel"          "Dud Bolt"              "Gasgano"
[49] "Ben Quadinaros"        "Mace Windu"            "Ki-Adi-Mundi"
[52] "Kit Fisto"             "Eeth Koth"             "Adi Gallia"
[55] "Saesee Tiin"           "Yaarael Poof"          "Plo Koon"
[58] "Mas Amedda"            "Gregar Typho"          "Cordé"
[61] "Cliegg Lars"           "Poggle the Lesser"     "Luminara Unduli"
[64] "Barriss Offee"         "Dormé"                 "Dooku"
[67] "Bail Prestor Organa"   "Jango Fett"            "Zam Wesell"
[70] "Dexter Jettster"       "Lama Su"               "Taun We"
[73] "Jocasta Nu"            "R4-P17"                "Wat Tambor"
[76] "San Hill"              "Shaak Ti"              "Grievous"
[79] "Tarfful"               "Raymus Antilles"       "Sly Moore"
[82] "Tion Medon"            "Finn"                  "Rey"
[85] "Poe Dameron"           "BB8"                   "Captain Phasma"
```

Compare the subsetting above to the tidyverse tools you know. Which correspond to `select()` and which to `pull()`?

`select()` corresponds to the `[` operator, while `pull()` corresponds to both the `[[` and `$` operators, as they both produce the same result.

5.

(2 pts)

Data frames are also considered lists. Read the section comparing Tibbles vs. data.frame in R for Data Science. Pay close attention to the Extracting variables, and Subsetting sections.

How do data frames behave differently to tibbles in terms of extracting or subsetting variables?

In terms of subsetting, tibbles do not allow for partial matching when using the `[`, `[[`, and `$` operators. They will also generate a warning if the column specified within or after the operator does not exist. Also, when working with base R data frames the `[` operator will sometimes return a data frame and sometimes a vector, while a tibble will only ever return another tibble when the `[` operator is used.