HW 5, STAT 450

Due: Friday, November 15

Directions: This assignment should be completed using Quarto and submitted to Canvas as a self-contained HTML or PDF file.

Reading: Chapters 17 from R for Data Science (2e)

```
library(tidyverse)
library(lubridate)
```

Exercise 1

Use an appropriate `lubridate` function to parse each of the following dates. That is, convert each string to a date object in R.

```
t1 <- "10/31/2024"
t2 <- "31-10-2024"
t3 <- "October 31, 2024"
```

Exercise 2

Run the following code to load the 2023 San Francisco crime data discussed in Lecture 18:

```
sfcrimes2023 <- readRDS(url("https://ericwfox.github.io/data/sfcrimes2023.rds"))
glimpse(sfcrimes2023)
```

```
Rows: 132,914
Columns: 6
$ id              <dbl> 1231769, 1231872, 1233075, 1234053, 1234109, 1234242, ~
$ datetime        <chr> "2023/01/01 12:00:00 AM", "2023/01/01 12:00:00 AM", "2~
$ category        <chr> "Burglary", "Larceny Theft", "Other", "Larceny Theft",~
$ police_district <chr> "Northern", "Bayview", "Taraval", "Southern", "Souther~
$ latitude        <dbl> 37.77206, 37.74220, 37.71369, 37.82452, 37.77403, 37.7~
$ longitude       <dbl> -122.4271, -122.4004, -122.4984, -122.3756, -122.3956,~
```

a

Create a subsetted data frame that only contains crime incidents categorized as `Larceny Theft`. Then use an appropriate `lubridate` function to parse the date-time column (i.e., convert it from a character to date-time type). *Hint*: Use the pipe operator with `filter()` followed by `mutate()`.

b

Create a new data frame with columns for the date, month, day of week, and hour of day for the theft incidents.

c

Create a bar plot displaying the counts of theft incidents that occurred each hour of the day (0-23).

d

Create a bar plot displaying the counts of theft incidents that occurred each day of the week (Sun-Sat).

e

Create a time series plot displaying the counts of theft incidents that occurred on each day in 2023. Use `geom_smooth()` to add a trend line, and set `span = 0.2` to adjust the smoothness of the line.

f

Describe any interesting trends or features that you discovered in the plots created in parts **c**, **d**, and **e**.