STA 445 Assignment #5

Chip Haskins

2024-10-28

Exercise 1

Convert the following to date or date/time objects.

a) September 13, 2010.

```
mdy("September 13, 2010")
```

```
## [1] "2010-09-13"
```

b) Sept 13, 2010.

```
mdy("Sept 13, 2010")
```

```
## Warning: All formats failed to parse. No formats found.
```

```
## [1] NA
```

c) Sep 13, 2010.

```
mdy("Sep 13, 2010")
```

```
## [1] "2010-09-13"
```

d) S 13, 2010. Comment on the month abbreviation needs.

```
mdy("S 13, 2010")
```

```
## Warning: All formats failed to parse. No formats found.
```

```
## [1] NA
```

Only "Sep" or "September" works. The parser doesn't recognize "S" or "Sept" as month names.

e) 07-Dec-1941.

```
dmy("07-Dec-1941")
```

```
## [1] "1941-12-07"
```

f) 1-5-1998. Comment on why you might be wrong.

```
dmy("1-5-1998")
```

```
## [1] "1998-05-01"
```

I arbitrarily chose "day-month-year" format, but this date could just as easily have been "month-day-year" format, and without other data points in the same format there is no way to know.

g) 21-5-1998. Comment on why you know you are correct.

```
dmy("21-5-1998")
```

```
## [1] "1998-05-21"
```

The first number must be the day, and the last number must be the year, so the only possible date format is "day-month-year."

h) 2020-May-5 10:30 am

```
ymd_hm("2020-May-5 10:30 am")
```

```
## [1] "2020-05-05 10:30:00 UTC"
```

i) 2020-May-5 10:30 am PDT (ex Seattle)

```
ymd_hm("2020-May-5 10:30 am", tz="US/Pacific")
```

```
## [1] "2020-05-05 10:30:00 PDT"
```

j) 2020-May-5 10:30 am AST (ex Puerto Rico)

```
ymd_hm("2020-May-5 10:30 am", tz="America/Puerto_Rico")
```

```
## [1] "2020-05-05 10:30:00 AST"
```

Exercise 3

Suppose you have arranged for a phone call to be at 3 pm on May 8, 2025 at Arizona time. However, the recipient will be in Auckland, NZ. What time will it be there?

```
with_tz(
  mdy_h("May 8, 2025 3 pm", tz="US/Arizona"),
  tzone="Pacific/Auckland"
)
```

```
## [1] "2025-05-09 10:00:00 NZST"
```

Exercise 5

It turns out there is some interesting periodicity regarding the number of births on particular days of the year.

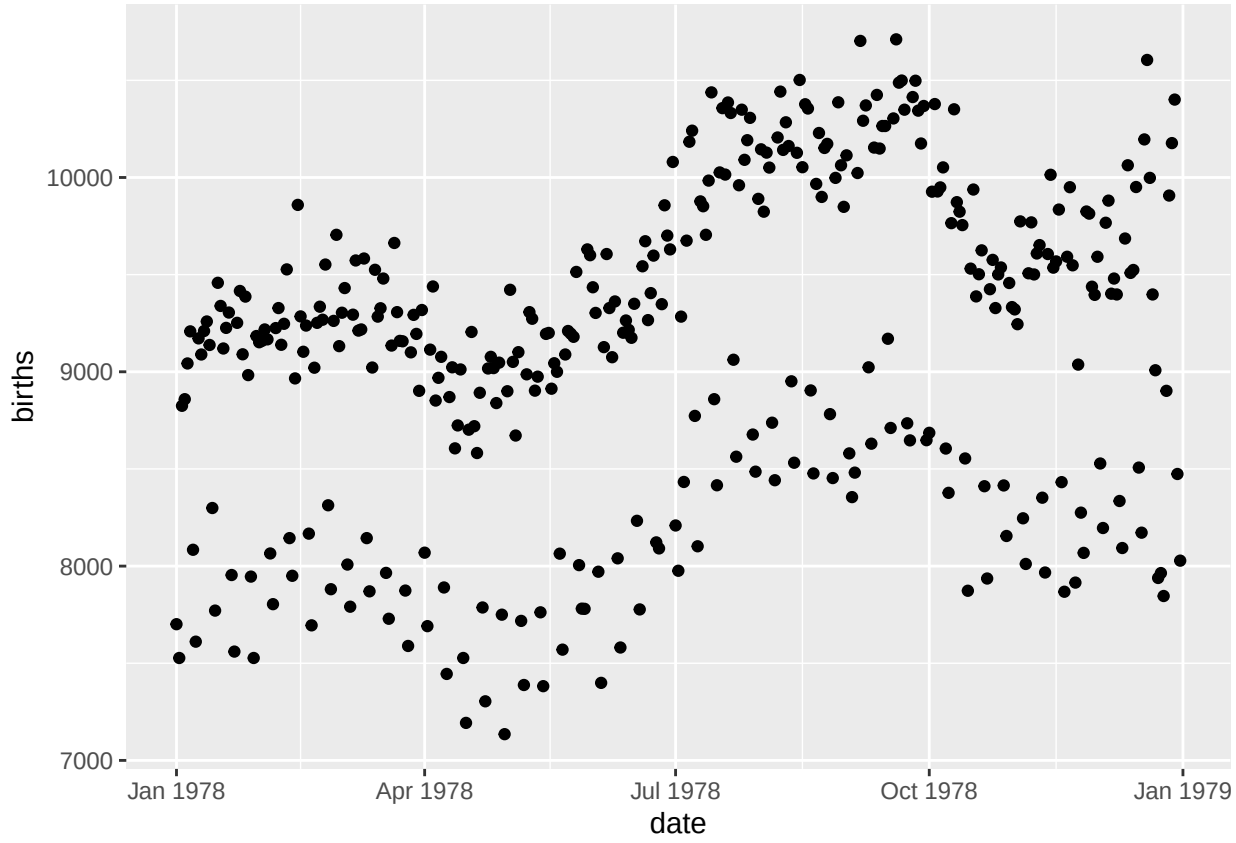
a) Using the `mosaicData` package, load the data set `Births78` which records the number of children born on each day in the United States in 1978. Because this problem is intended to show how to calculate the information using the `date`, remove all the columns *except* `date` and `births`.

```
data("Births78", package="mosaicData")
Births78 <- Births78 %>% select(date, births)
head(Births78)
```

```
##         date births
## 1 1978-01-01   7701
## 2 1978-01-02   7527
## 3 1978-01-03   8825
## 4 1978-01-04   8859
## 5 1978-01-05   9043
## 6 1978-01-06   9208
```

b) Graph the number of `births` vs the `date` with date on the x-axis. What stands out to you? Why do you think we have this trend?

```
ggplot(Births78, aes(x=date, y=births)) +
  geom_point()
```



There appear to be two groupings of data, one significantly lower but mirroring the shape of the other. This may be due to certain days of the week being favored by hospitals or mothers during labor.

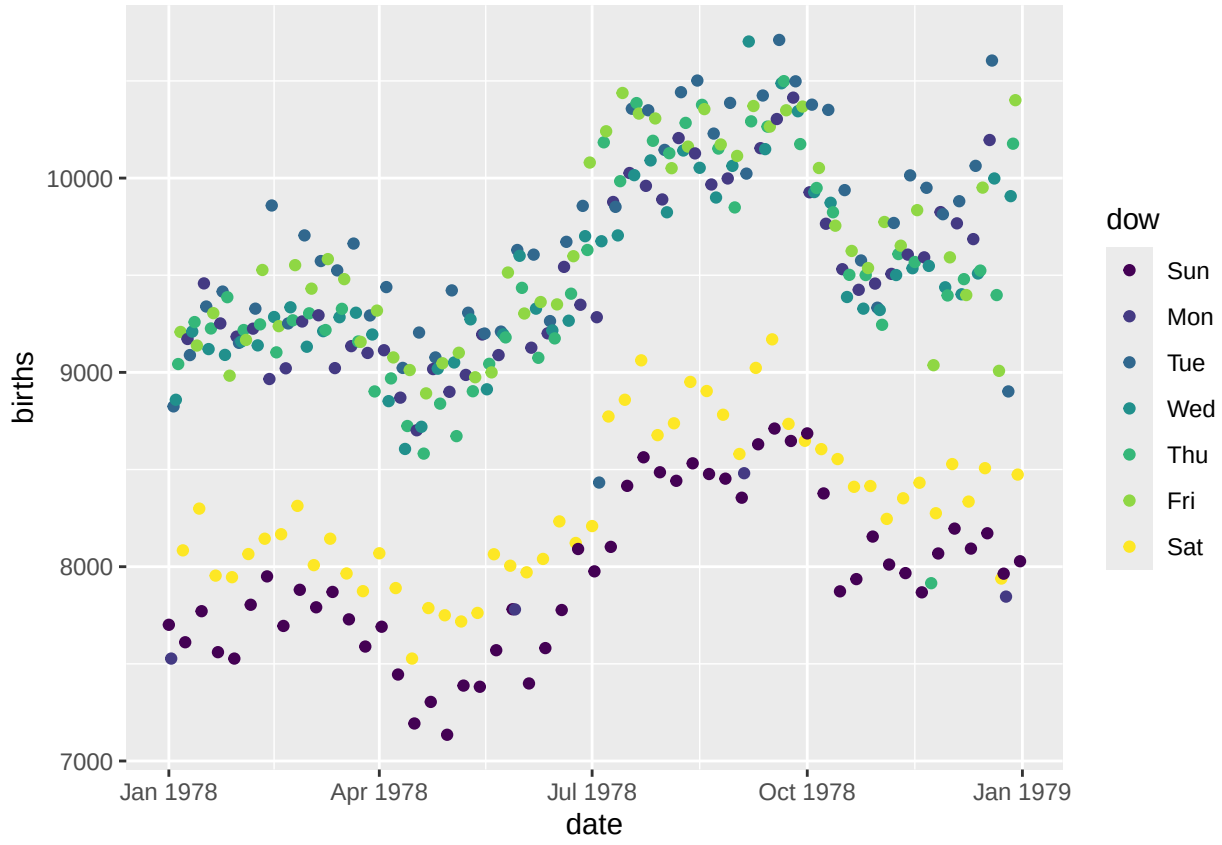
c) To test your assumption, we need to figure out the what day of the week each observation is. Use `dplyr::mutate` to add a new column named `dow` that is the day of the week (Monday, Tuesday, etc). This calculation will involve some function in the `lubridate` package and the `date` column.

```
Births78 <- Births78 %>% mutate(
  dow = wday(date, label = TRUE)
)
head(Births78)
```

```
##         date births dow
## 1 1978-01-01   7701 Sun
## 2 1978-01-02   7527 Mon
## 3 1978-01-03   8825 Tue
## 4 1978-01-04   8859 Wed
## 5 1978-01-05   9043 Thu
## 6 1978-01-06   9208 Fri
```

d) Plot the data with the point color being determined by the day of the week variable.

```
ggplot(Births78, aes(x=date, y=births)) +
  geom_point(aes(color=dow))
```



Optional Exercises

Exercise 2

Using your date of birth (ex Sep 7, 1998) and today's date calculate the following *Write your code in a manner that the code will work on any date after you were born.*:

```
birthdate <- ymd("2003-10-10")
```

a) Calculate the date of your 64th birthday.

```
birthdate + years(64)
```

```
## [1] "2067-10-10"
```

b) Calculate your current age (in years). *Hint: Check your age is calculated correctly if your birthday was yesterday and if it were tomorrow!*

```
age <- birthdate %--% today()
age <- as.period(age) %>% year()
age
```

```
## [1] 21
```

c) Using your result in part (b), calculate the date of your next birthday.

```
next.birthday <- birthdate + years(age + 1)
next.birthday
```

```
## [1] "2025-10-10"
```

d) The number of *days* until your next birthday.

```
today() %--% next.birthday %>% as.numeric("days")
```

```
## [1] 336
```

f) The number of *months* and *days* until your next birthday.

```
today() %--% next.birthday %>% as.period()
```

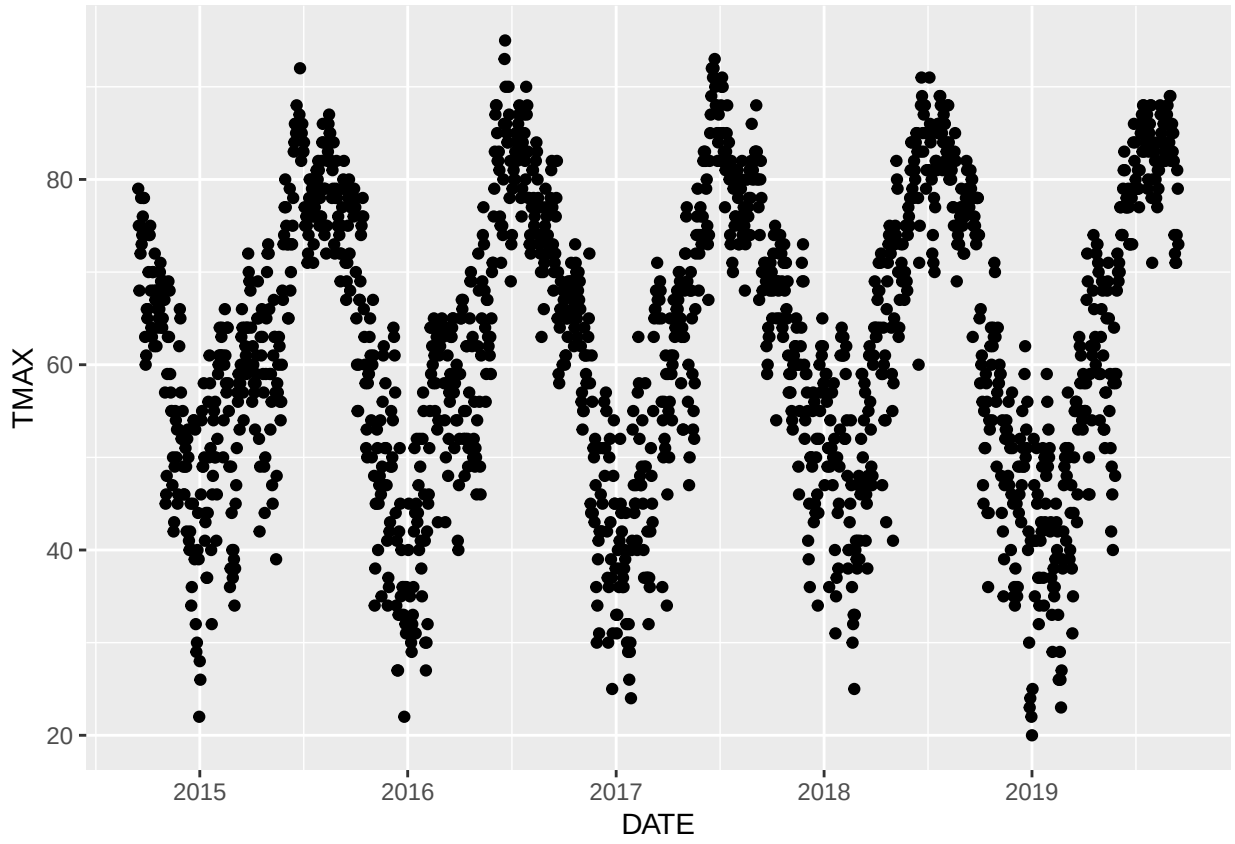
```
## [1] "11m 2d 0H 0M 0S"
```

Exercise 4

From this book's GitHub directory, navigate to the `data-raw` directory and then download the `Pulliam_Airport_Weather_Station.csv` data file. (*There are several weather station files. Make sure you get the correct one!*) There is a `DATE` column (is it of type `date` when you import the data?) as well as the Maximum and Minimum temperature. For the last 5 years of data included in the file, plot the time series of daily maximum temperature with date on the x-axis. Write your code so that it will work if I update the date set. *Hint: Find the maximum date in the data set and then subtract 5 years. Will there be a difference if you use* `dyears(5)` *vs* `years(5)`*? Which seems more appropriate here?*

```r
PulliamWeather <-
  read.csv(
    "https://raw.githubusercontent.com/BuscagliaR/STA_444_v2/main/data-raw/Pulliam_Airport_Weather_Stati
  select(DATE, TMAX) %>%
  mutate(DATE = ymd(DATE)) %>%
  filter(between(DATE, max(DATE) - years(5), max(DATE)))

ggplot(PulliamWeather, aes(x = DATE, y = TMAX)) +
  geom_point()
```



I used `years(5)` because the interval we care about is the five years before the *date* of the last observation; the exact number of seconds doesn't matter here. I used a period since subtracting 1 year as a period results in the exact same date (1 year in the past) and time, but subtracting 1 year as a duration may not.