

STA_445_Assignment_6

Paige Hawkinson

2024-03-28

```
library(tidyverse)
library(lubridate)
library(ggplot2)
library(mosaicData)
```

Problem 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
```

We cannot run the code as R does not recognize Sept to be September. We would need to rewrite this to either Sep 13, 2020 or September 13, 2010.

- 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
```

This is similar to 1b where R does not recognize S as September. We would need to rewrite this in order for R to create the output we want.

e. 07-Dec-1941.

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

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

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

```
mdy('1-5-1998')
```

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

I might be wrong as I understood the date to be January 5th, 1998 but the date could also be day, month, year... May 1st, 1998.

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

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

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

I know I am correct because there are not 21 months, meaning the order 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"
```

Problem 2

Using just your date of birth (ex Sep 7, 1998) and today's date calculate the following:

- a. Calculate the date of your 64th birthday.

```
dob <- mdy('July 29, 2001')  
dob + years(64)
```

```
## [1] "2065-07-29"
```

- b. Calculate your current age (in years).

```
today's.date <- Sys.Date()  
current.age <- interval(dob, today's.date)  
(current.age.2 <- year(as.period(current.age)))
```

```
## [1] 22
```

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

```
(next.bday <- years(current.age.2 + 1) + dob)
```

```
## [1] "2024-07-29"
```

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

```
as.period(today's.date %--% next.bday, unit = "days")
```

```
## [1] "123d 0H 0M 0S"
```

- e. The number of *months* and *days* until your next birthday.

```
as.period(today's.date %--% next.bday, unit = "months")
```

```
## [1] "4m 1d 0H 0M 0S"
```

Problem 3

Suppose you have arranged for a phone call to be at 3 pm on May 8, 2015 at Arizona time. However, the recipient will be in Auckland, NZ. What time will it be there?

```
AZ.time <- ymd_hm('2015-May-8 03:00 pm', tz='US/Arizona')
with_tz(AZ.time, "Pacific/Auckland")
```

```
## [1] "2015-05-09 10:00:00 NZST"
```

Problem 4

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

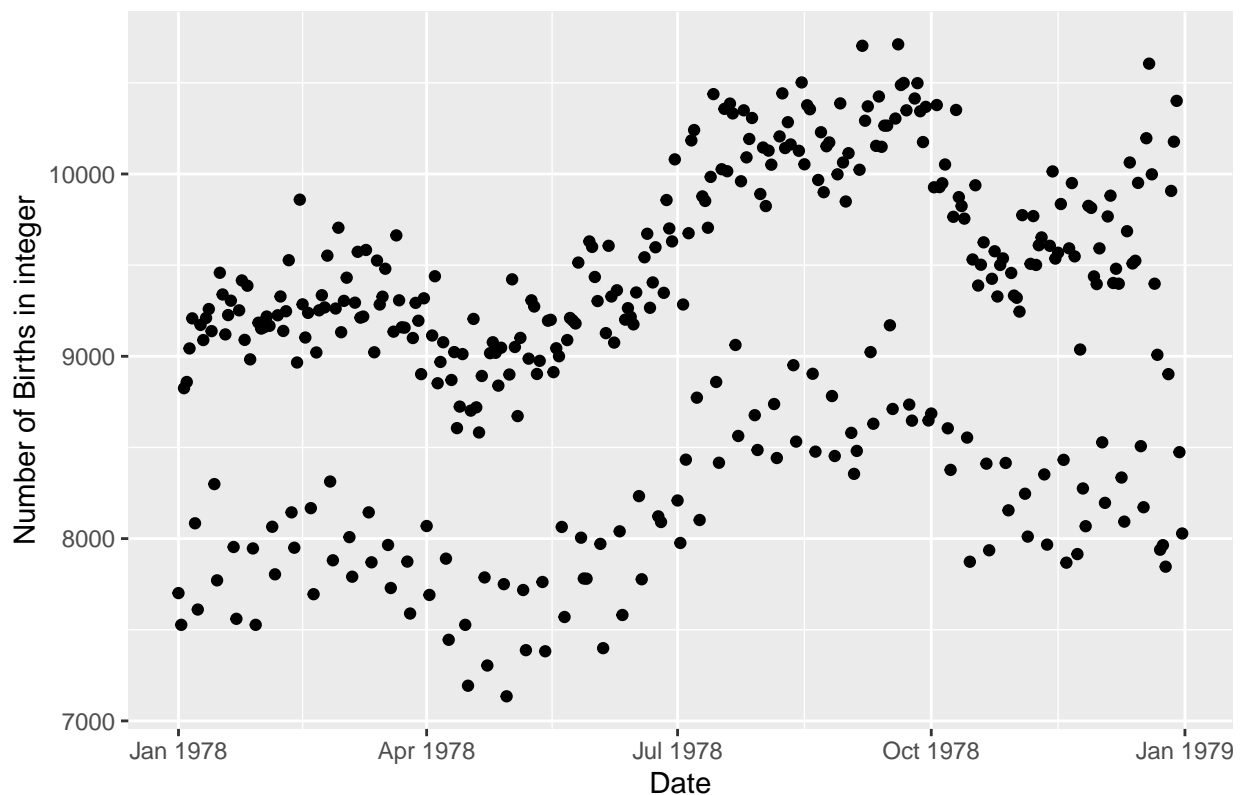
- Using the `mosaicData` package, load the data set `Births78` which records the number of children born on each day in the United States in 1978. Because this problem is intended to show how to calculate the information using the `date`, remove all the columns *except* `date` and `births`.

```
new.Births78 <- Births78 %>%
  select("date", "births")
```

- 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(data = new.Births78, aes(x = date, y = births)) +
  geom_point() +
  labs(title = 'Number of births vs date in 1978') +
  labs(x = "Date", y = "Number of Births in integer")
```

Number of births vs date in 1978



In the graph, there is a clear distinction in the data points, causing there to be two groups of samples. This trend could be due to the day of the week the baby was born on.

- c. To test your assumption, we need to figure out 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.

```
new.new.Births78 <- new.Births78 %>%  
  mutate(dow = wday(date, label = TRUE, abbr = FALSE))
```

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

```
ggplot(data = new.new.Births78, aes(x = date, y = births, color=dow)) +  
  geom_point() +  
  labs( title = 'Number of births vs date in 1978') +  
  labs( x = "Date", y = "Number of Births in integer")
```

