# Assignment 5

Reagan Anderson

2024-11-01

## 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
```

The month needs to be 'Sep' (the first three letter of the month) to properly work mdy()

**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 because the 1 and 5 are unspecified on which is the month and which is the day, so it could be dmy() instead.

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

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

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

I know I am right because there are only 12 months, and 21 is larger than 12 and therefore a day and not a month.

**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 = 'America/Los_Angeles')
```

```
## [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?

```
az_time <- mdy_hm('May 8 2025 3:00 pm', tz = 'US/Arizona')
with_tz(az_time, tz='Pacific/Auckland')
```

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

### Exercise 5:

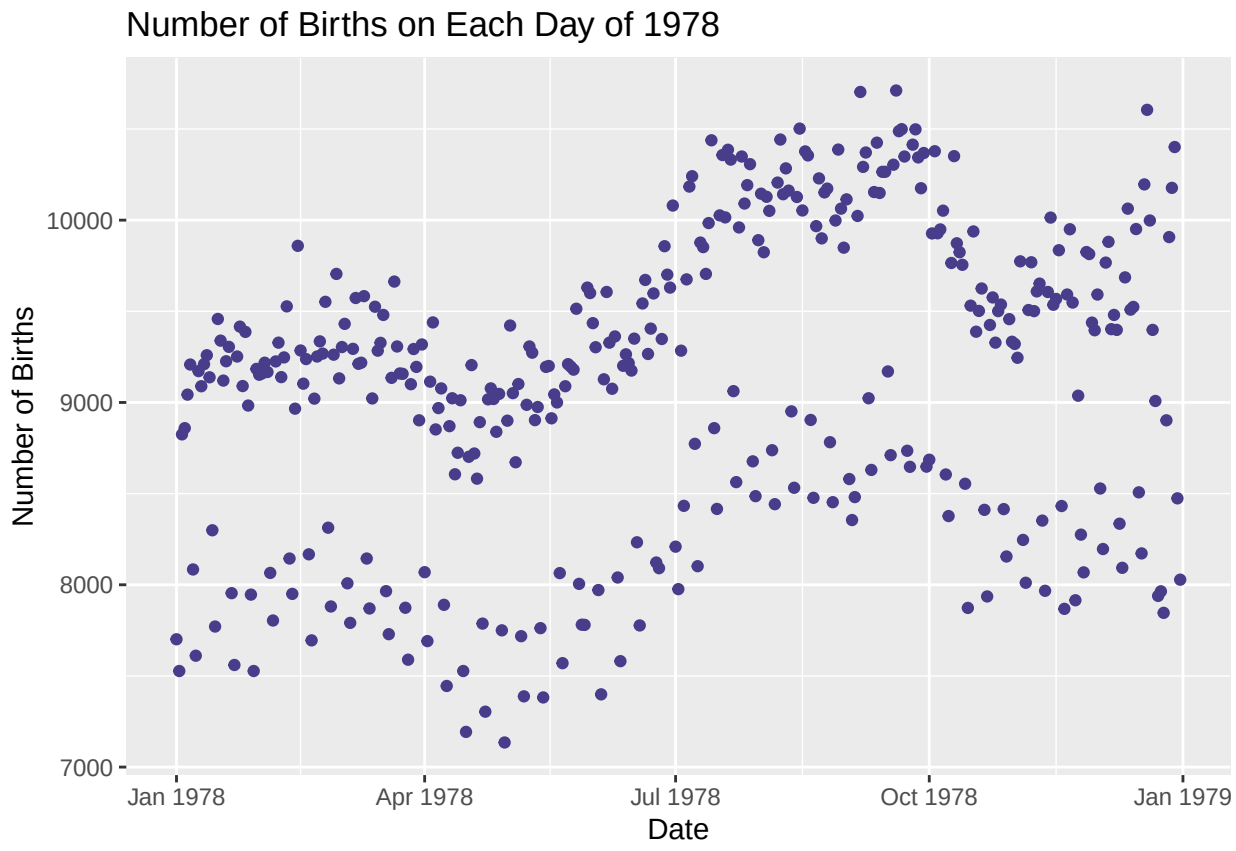
It turns out there is some interesting periodicity regarding the number of births on particular days of the year.

**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')
Births78_date <- Births78 %>% select( starts_with(c('date', 'births')) )
```

**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_date,
       aes(x=date, y=births) ) +
  geom_point(color='darkslateblue') +
  labs(title = "Number of Births on Each Day of 1978") +
  labs( x="Date", y="Number of Births")
```




What stands out immediately is the appearance of two separate flows of data. However, there is only one and it shows a fluctuation in the births on each day of the year 1978.

**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_date$dow <- wday(Births78_date$date, label = TRUE)
```

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

```
ggplot(Births78_date,
       aes(x=date, y=births) ) +
  geom_point( aes(color = dow) ) +
  labs(title = "Number of Births on Each Day of 1978") +
  labs( x="Date", y="Number of Births")
```