Assignment 5

Jack Scherer

2024-10-30

Exercise 1

Convert the following to date or date/time objects.

a) September 13, 2010.

```
(a <- mdy( "September 13, 2010" ))
```

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

b) Sept 13, 2010.

```
date.b <- "Sept 13, 2010"
date.b <- str_replace( date.b, 'Sep\\D*', 'Sep ' )
( b <- mdy( date.b ) )
```

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

c) Sep 13, 2010.

```
(c <- mdy( "Sep 13, 2010" ))
```

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

d) S 13, 2010. Comment on the month abbreviation needs. The month abbreviation needs to be three letters.

```
date.c <- "S 13, 2010"
date.c <- str_replace( date.c, 'S', 'Sep ' )
( b <- mdy( date.c ) )
```

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

e) 07-Dec-1941.

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

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

f) 1-5-1998. Comment on why you might be wrong. I do not know whether the order is day-month-year or month-day-year.

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

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

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

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

g) 21-5-1998. Comment on why you know you are correct. The highest numeric value for a month is 12, so 21 must indicate the day.

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

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

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.dt <- ymd_hm( '2025-05-08 15:00', tz='US/Arizona' )

nz.dt <- with_tz( az.dt, tzone='Pacific/Auckland' )
nz.dt
```

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

Exercise 5

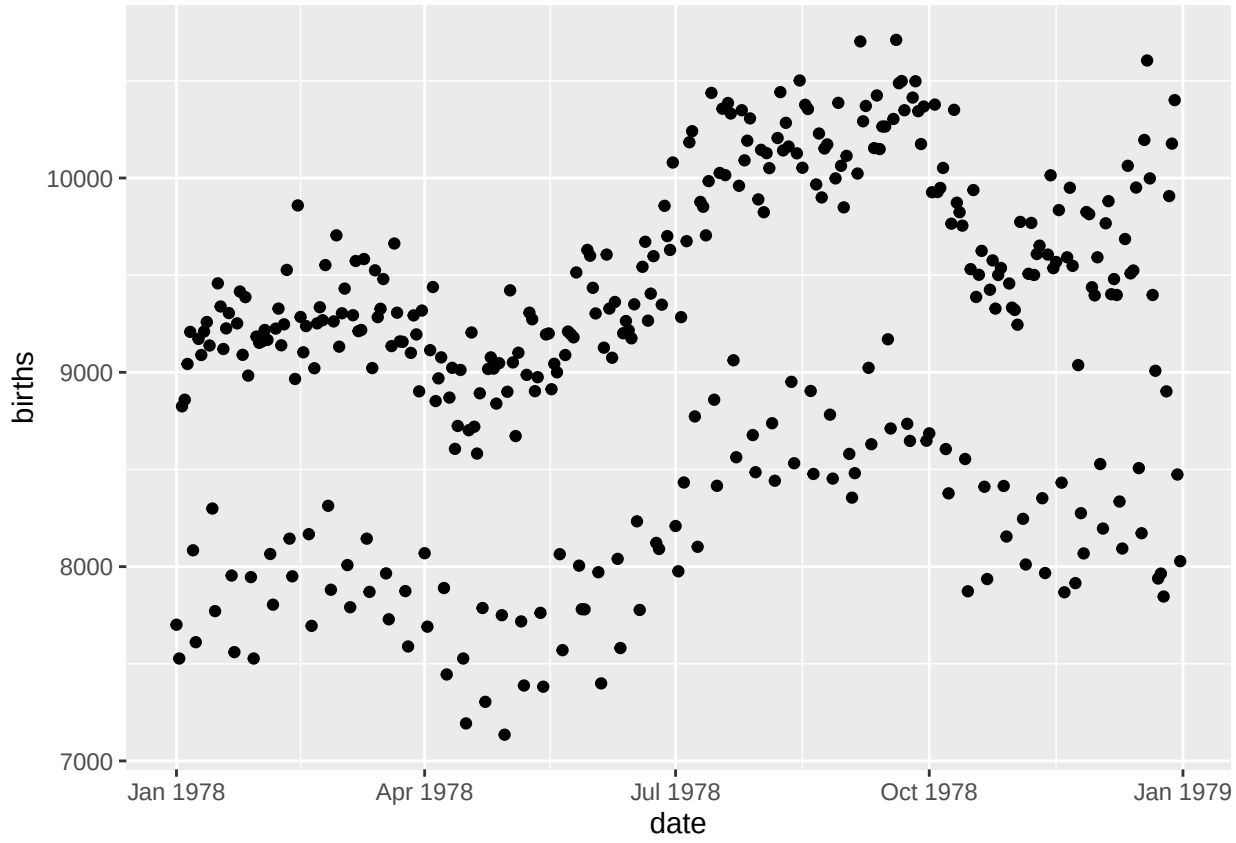
It turns out there is some interesting periodicity regarding the number of births on particular days of the year.

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`.

```
Births <- mosaicData::Births78
Births <- Births %>%
  select( 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( Births, aes(x=date, y=births) ) +
  geom_point()
```



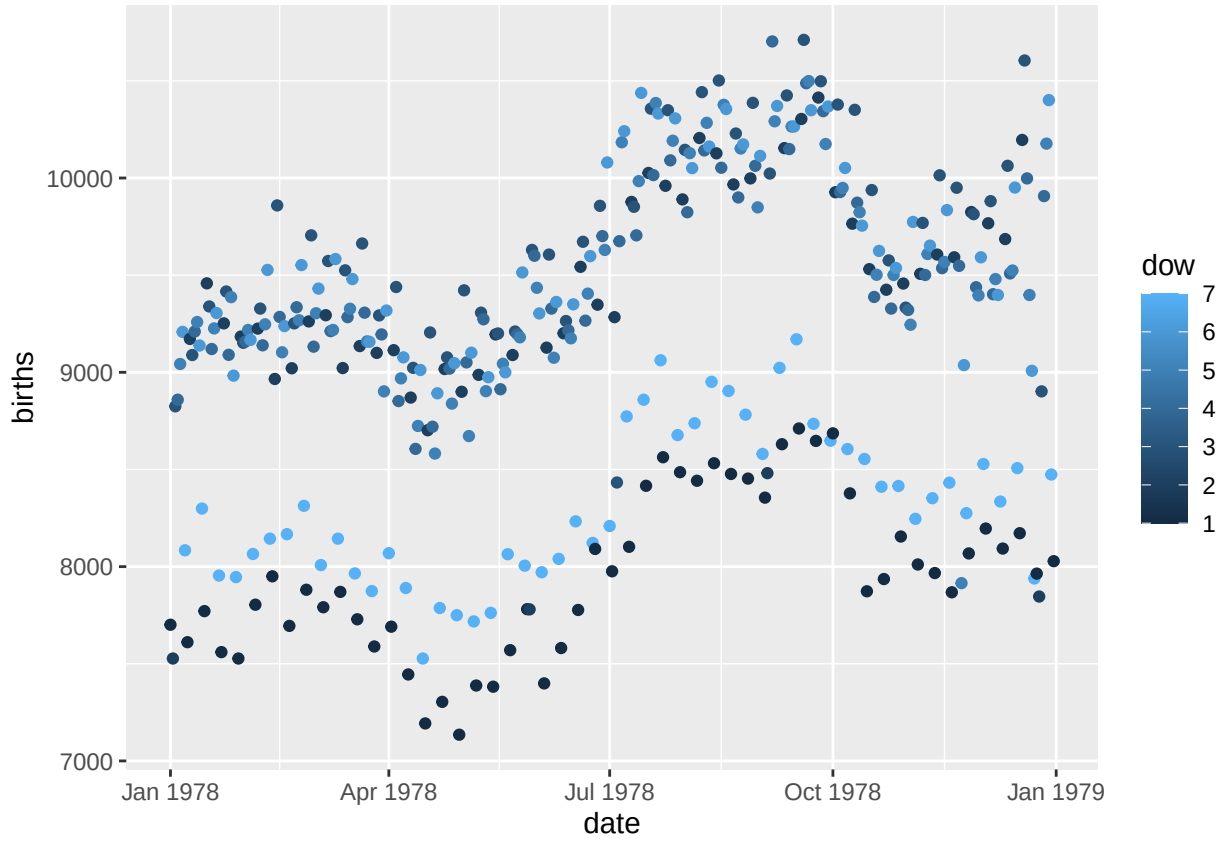
It appears that births are most common during the late summer and early fall. Also it appears that there is some day of the week where births are less common, but the plot is too zoomed out to figure out which one.

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.

```
Births <- Births %>%
  mutate( dow=wday(date) )
```

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

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



It appears that Sunday and Saturday have less births than the weekdays.