STA_445_Assignment_6

Sophia Kubisiak

03/28/24

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

Problem 1

Convert the following to date or date/time objects.

    a. September 13, 2010.

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

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

    b. Sept 13, 2010.

```r
b <- "Sept 13, 2010"
mdy(b)
```

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

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

    c. Sep 13, 2010.

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

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

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

```r
d <- "S 13, 2010"
mdy(d)
```

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

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

The abbreviations that work are september and sep. Sept and S do not work.

    e. 07-Dec-1941.

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

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

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

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

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

The date could be 'day month year' or 'month day year', it is not clear.

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

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

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

The date must be read as 'day month year' , because there are not 21 months.

    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:

```r
bday <- "June 21, 2002"
birthday <- mdy(bday)
```

    a. Calculate the date of your 64th birthday.

```r
birthday + years(64)
```

```
## [1] "2066-06-21"
```

    b. Calculate your current age (in years).

```r
as.duration(birthday %--% now())
```

```
## [1] "687120500.701457s (~21.77 years)"
```

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

```r
(nextbday <- birthday + years(22))
```

```
## [1] "2024-06-21"
```

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

```r
as.period(nextbday %--% now(), unit="days")
```

```
## [1] "-83d -5H -11M -39.2851819992065S"
```

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

```r
as.period(nextbday %--% now(), unit="months")
```

```
## [1] "-2m -22d -5H -11M -39.2765278816223S"
```

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?

```r
phonecall <- (mdy_h("May 8, 2015 3PM" , tz="US/Arizona"))
(Auckland <- with_tz(phonecall , tz = "NZ"))
```

```
## [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.

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")

births <- Births78 %>%
  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?

There are a lot more births between July and October. Maybe this has to do with trends of pregnancy, and possibly marriage dates could have a small impact as well. Also, maybe the amount of planned pregnancies all wanted to give birth in later summer months and early fall in 1978.

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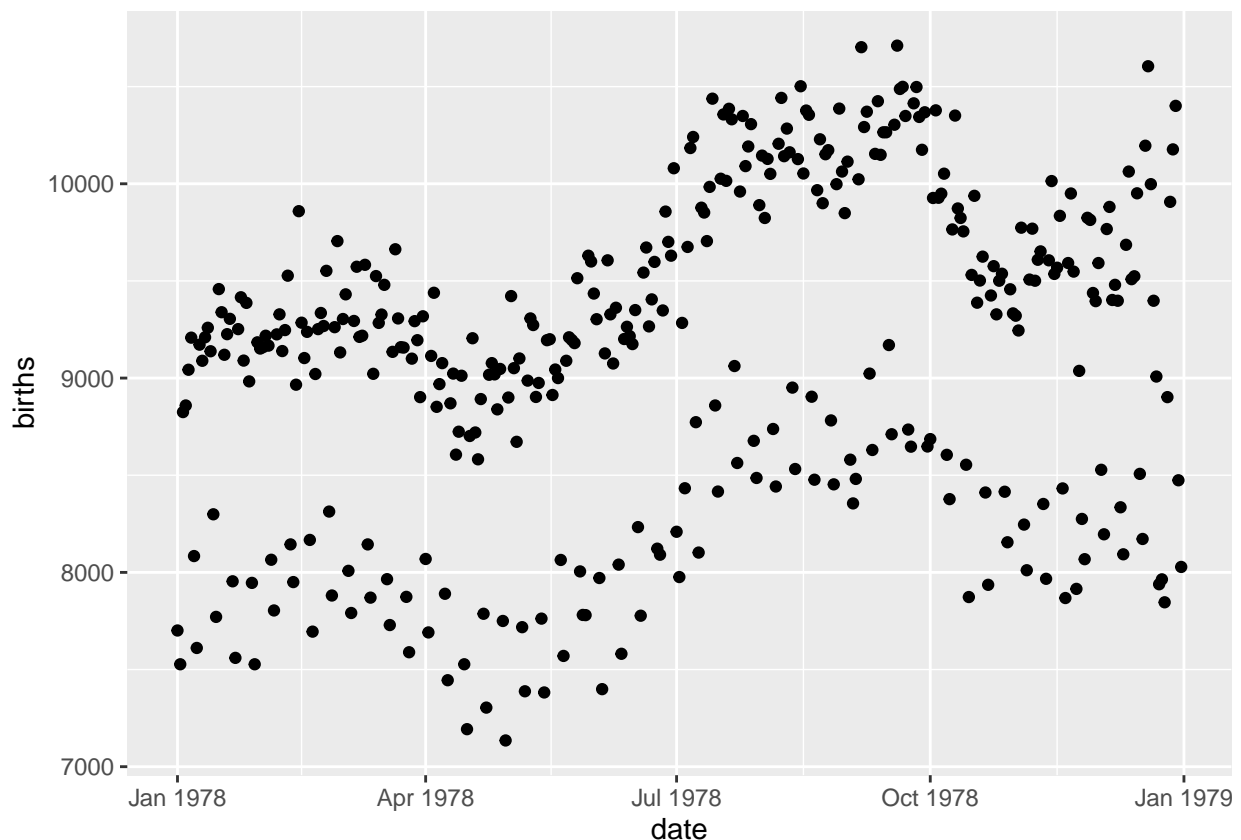

c. To test your assumption, we need to figure out the what day of the week each observation is. Use

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, label=TRUE))
```

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

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