**שאלות**

1. מה קורה אם מנסים לעשות parse ל-string שמכיל תאריכים לא תקפים?

ymd(c("2010-10-10", "bananas"))

1. מה התפקיד של הארגומנט tzone בפונקציה today? מדוע הוא חשוב?
2. השתמש בפונקציות המתאימות ב-lubridate כדי לעשות parse לתאריכים הבאים:

d1 <- "January 1, 2010"

d2 <- "2015-Mar-07"

d3 <- "06-Jun-2017"

d4 <- **c**("August 19 (2015)", "July 1 (2015)")

d5 <- "12/30/14" *# Dec 30, 2014*

1. איך ההתפלגות של שעות הטיסה משתנות לאורך השנה?
2. באיזה יום בשבוע כדאי לטוס אם רוצים למזער את ההסתברות לאיחור?

**פתרונות**

1.

ret <- **ymd**(**c**("2010-10-10", "bananas"))

**print**(**class**(ret))

ret

It produces an NA and a warning message.

2. It determines the time-zone of the date. Since different time-zones can have different dates, the value of today() can vary depending on the time-zone specified.

3.

**mdy**(d1)

**ymd**(d2)

**dmy**(d3)

**mdy**(d4)

**mdy**(d5)

4. Let’s try plotting this by month:

flights\_dt %>%

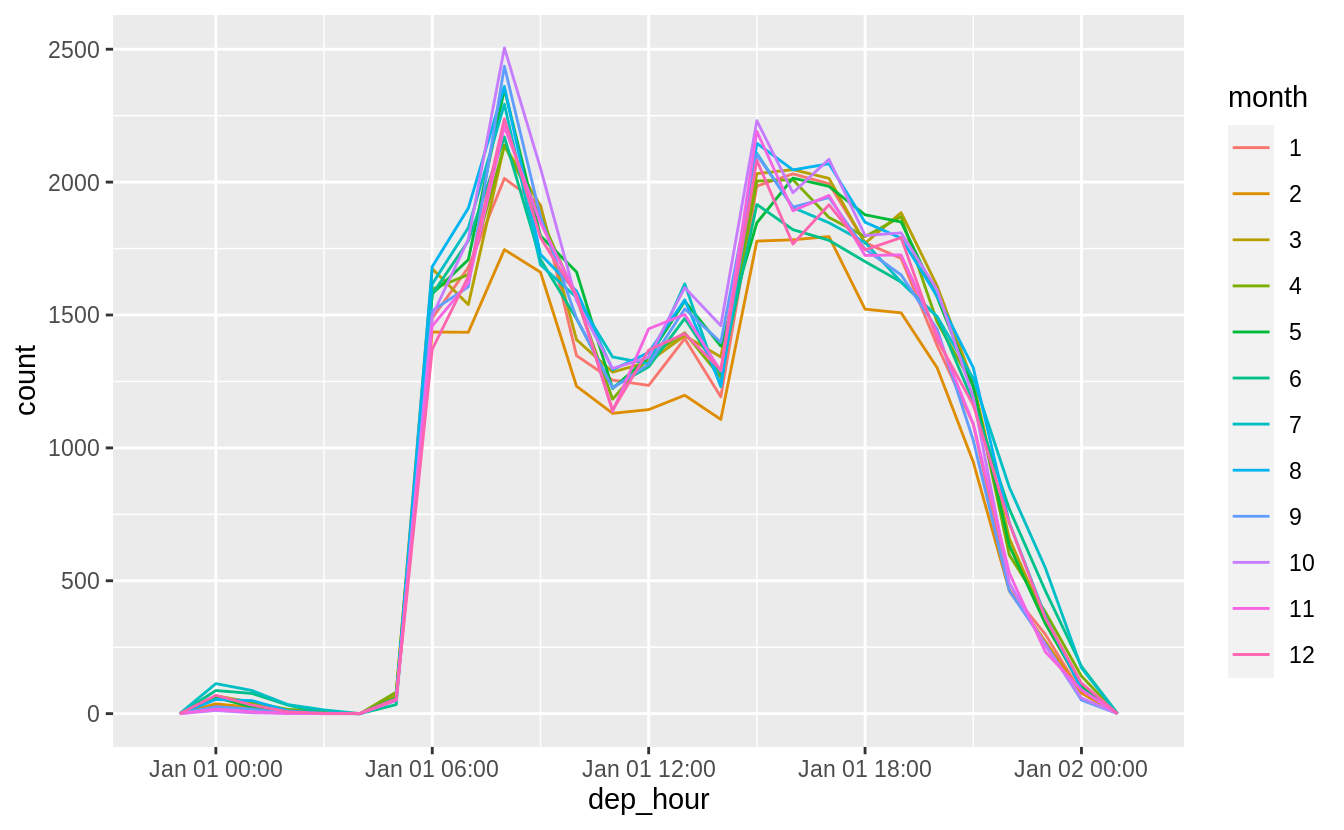
**filter**(!**is.na**(dep\_time)) %>%

**mutate**(dep\_hour = **update**(dep\_time, yday = 1)) %>%

**mutate**(month = **factor**(**month**(dep\_time))) %>%

**ggplot**(**aes**(dep\_hour, color = month)) +

**geom\_freqpoly**(binwidth = 60 \* 60)



This will look better if everything is normalized within groups. The reason that February is lower is that there are fewer days and thus fewer flights.

flights\_dt %>%

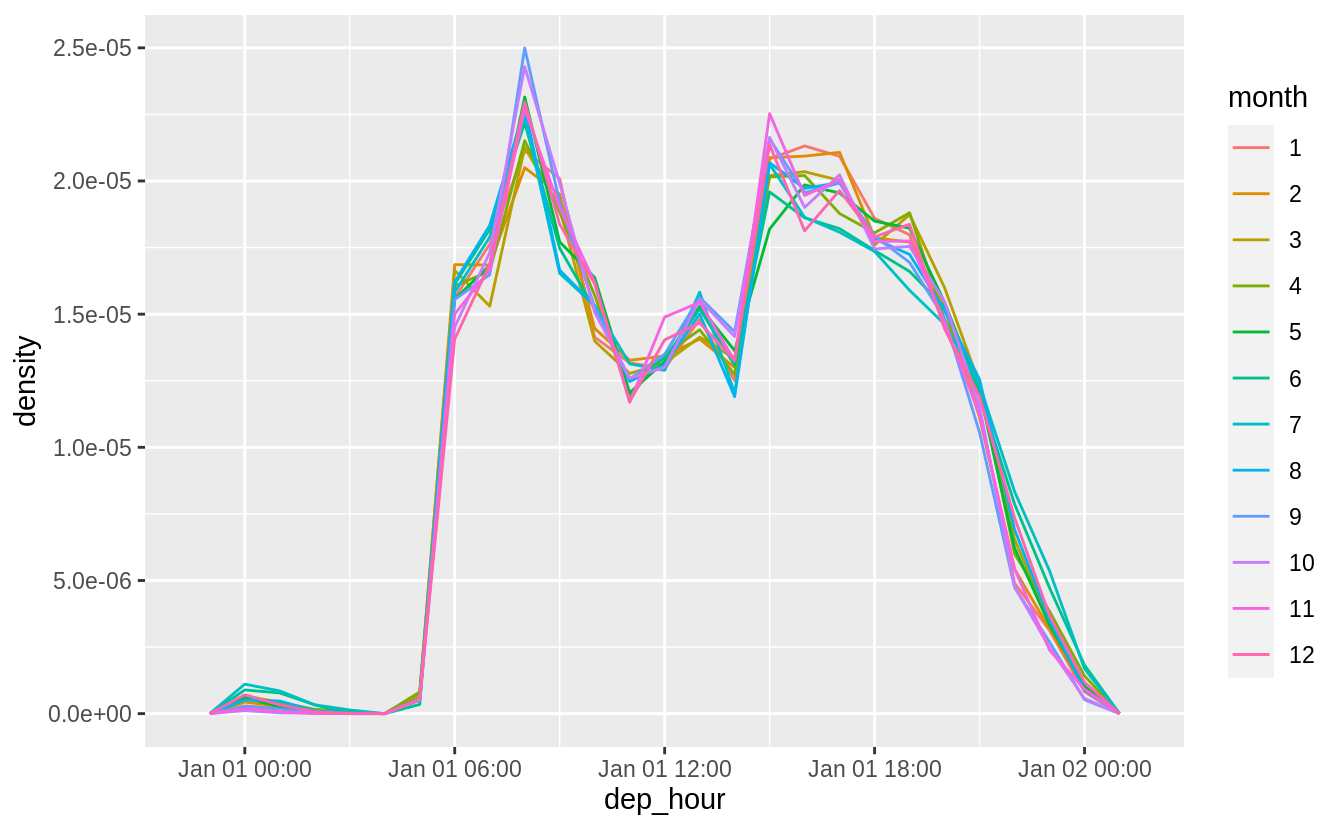
**filter**(!**is.na**(dep\_time)) %>%

**mutate**(dep\_hour = **update**(dep\_time, yday = 1)) %>%

**mutate**(month = **factor**(**month**(dep\_time))) %>%

**ggplot**(**aes**(dep\_hour, color = month)) +

**geom\_freqpoly**(**aes**(y = ..density..), binwidth = 60 \* 60)



At least to me there doesn’t appear to much difference in within-day distribution over the year, but I maybe thinking about it incorrectly.

5. Saturday has the lowest average departure delay time and the lowest average arrival delay time.

flights\_dt %>%

**mutate**(dow = **wday**(sched\_dep\_time)) %>%

**group\_by**(dow) %>%

**summarise**(

dep\_delay = **mean**(dep\_delay),

arr\_delay = **mean**(arr\_delay, na.rm = TRUE)

) %>%

**print**(n = Inf)

flights\_dt %>%

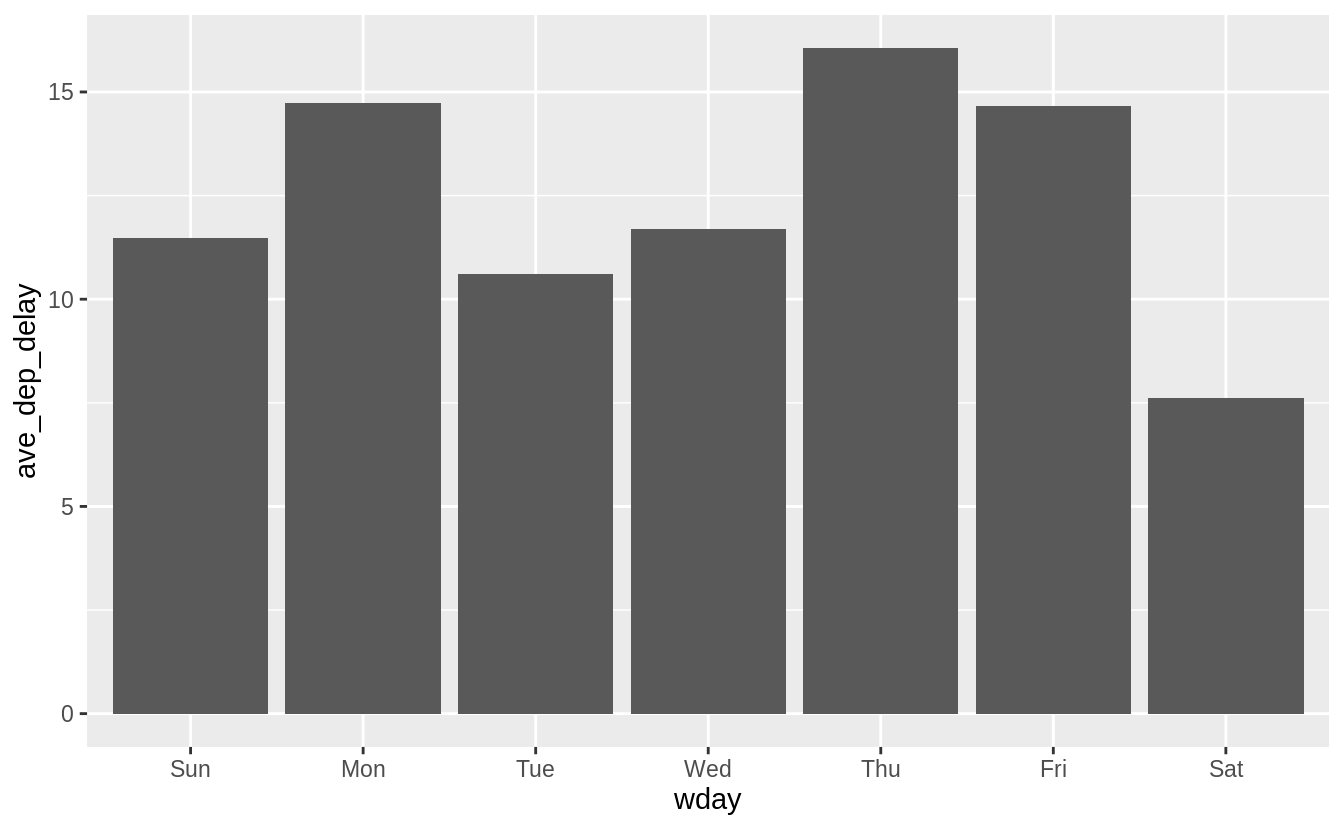
**mutate**(wday = **wday**(dep\_time, label = TRUE)) %>%

**group\_by**(wday) %>%

**summarize**(ave\_dep\_delay = **mean**(dep\_delay, na.rm = TRUE)) %>%

**ggplot**(**aes**(x = wday, y = ave\_dep\_delay)) +

**geom\_bar**(stat = "identity")



flights\_dt %>%

**mutate**(wday = **wday**(dep\_time, label = TRUE)) %>%

**group\_by**(wday) %>%

**summarize**(ave\_arr\_delay = **mean**(arr\_delay, na.rm = TRUE)) %>%

**ggplot**(**aes**(x = wday, y = ave\_arr\_delay)) +

**geom\_bar**(stat = "identity")

