

DS 613, HW #1

Pete Magee

2023-01-25

```
library(tidyverse)
library(lubridate)
library(nycflights13)
```

```
# 1: How many days from July 6th 2020 to July 14th 2021?
```

```
dt <- mdy("July 6th 2020")
dt2 <- mdy("July 14th 2021")
```

```
((dt %--% dt2) / days(1))
```

```
## [1] 373
```

```
# 2: Confirm that 1988 was a leap year.
```

```
dt_leap <- ymd("1988-01-01")
dt_leap2 <- ymd("1989-01-01")
```

```
((dt_leap %--% dt_leap2) %/% days(1))
```

```
## [1] 366
```

```
# 3: Confirm that 1989 was NOT a leap year.
```

```
dt_nleap <- ymd("1989-01-01")
dt_nleap2 <- ymd("1990-01-01")
```

```
((dt_nleap %--% dt_nleap2) %/% days(1))
```

```
## [1] 365
```

#4: Why is there `months()` but no `dmonths()` (answer in 3-4 sentences)? Months are not a set interval. They can be 31, 30, 28 or 29 days, in the case of a 'leap February'. Therefore there is no set number of seconds in a month.

```
# 5: John was born April 11th, 1962. How old is John in years?
```

```
birthdate <- mdy("April 11th, 1962")
```

```
((birthdate %--% now()) / years(1))
```

```
## [1] 60.79405
```

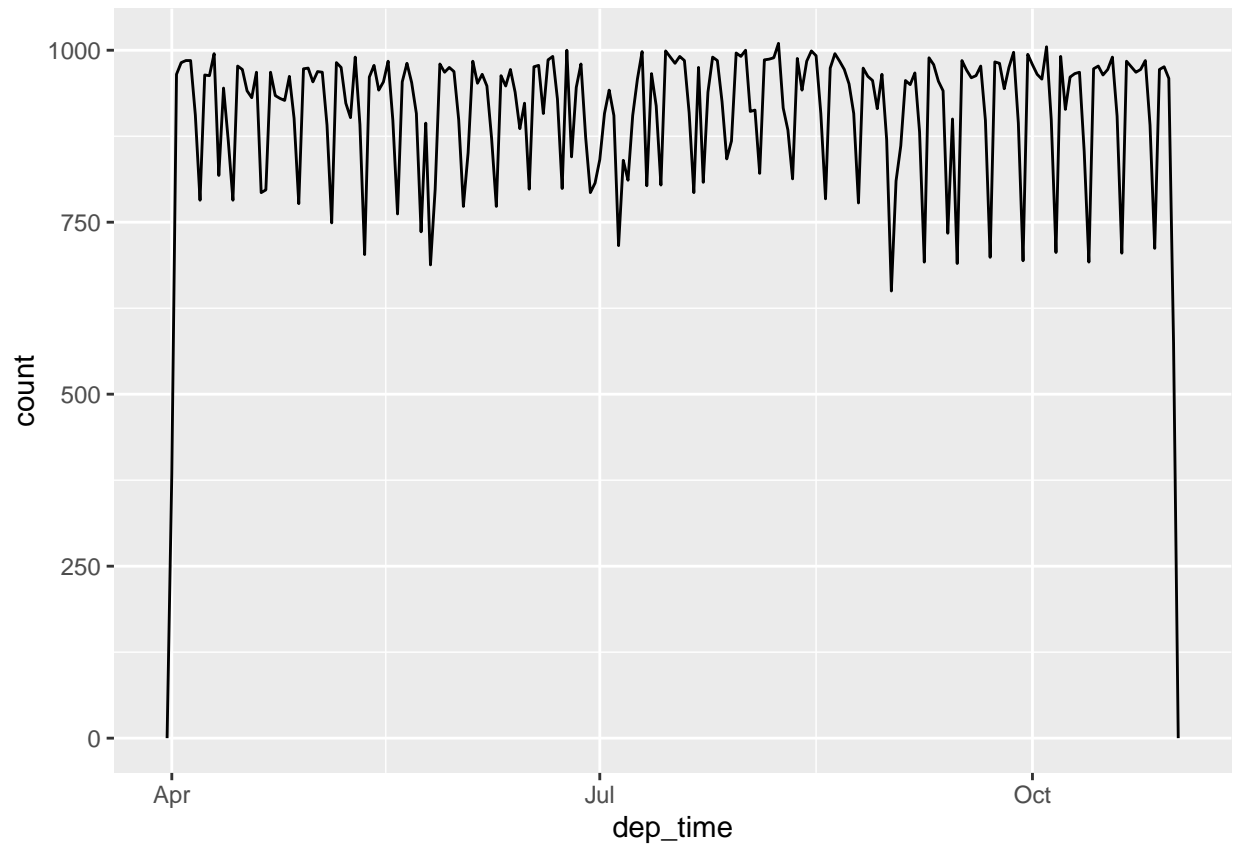
```
# 6: Obtain partial df
flights_dt <- flights %>%
  filter(!is.na(dep_time), !is.na(arr_time)) %>%
  mutate(
    dep_time = make_datetime(year, month, day, dep_time),
    arr_time = make_datetime(year, month, day, arr_time)
  ) %>%
  select(origin, dest, carrier, arr_time, dep_time)
flights_dt
```

```
## # A tibble: 328,063 x 5
##   origin dest carrier arr_time      dep_time
##   <chr>  <chr> <chr>   <dtm>      <dtm>
## 1 EWR    IAH    UA     2013-01-31 14:00:00 2013-01-22 13:00:00
## 2 LGA    IAH    UA     2013-01-31 10:00:00 2013-01-23 05:00:00
## 3 JFK    MIA    AA     2013-01-31 11:00:00 2013-01-23 14:00:00
## 4 JFK    BQN    B6     2013-01-31 20:00:00 2013-01-23 16:00:00
## 5 LGA    ATL    DL     2013-01-31 20:00:00 2013-01-24 02:00:00
## 6 EWR    ORD    UA     2013-01-31 20:00:00 2013-01-24 02:00:00
## 7 EWR    FLL    B6     2013-01-31 01:00:00 2013-01-24 03:00:00
## 8 LGA    IAD    EV     2013-01-30 13:00:00 2013-01-24 05:00:00
## 9 JFK    MCO    B6     2013-01-31 22:00:00 2013-01-24 05:00:00
## 10 LGA   ORD    AA     2013-02-01 09:00:00 2013-01-24 06:00:00
## # ... with 328,053 more rows
```

```
# 7: Produce frequency plot for April, July, and October for the year 2013.**
make_datetime_100 <- function(year, month, day, time) {
  make_datetime(year, month, day, time %/% 100, time %% 100)
}

flights_dt <- flights %>%
  filter(!is.na(dep_time), !is.na(arr_time)) %>%
  mutate(
    dep_time = make_datetime_100(year, month, day, dep_time),
    arr_time = make_datetime_100(year, month, day, arr_time)
  ) %>%
  select(month, origin, dest, carrier, arr_time, dep_time)

flights_dt %>%
  filter(dep_time >= ymd(20130401) & dep_time <= ymd(20131031)) %>%
  ggplot(aes(dep_time)) +
  geom_freqpoly(binwidth = 86400) # 86400 seconds = 1 day
```



```
# 8: arrival times for American Airlines at the Dallas Fort Worth Airport from the LaGuardia airport in
flights2 <- flights_dt %>%
  select(origin, dest, carrier, arr_time) %>%
  filter(
    carrier == "AA" &
    origin == "LGA" &
    dest == "DFW"
  )
flights2[115:125, ]
```

```
## # A tibble: 11 x 4
##   origin dest  carrier arr_time
##   <chr>  <chr>  <chr>   <dtm>
## 1 LGA    DFW    AA     2013-01-09 16:16:00
## 2 LGA    DFW    AA     2013-01-09 19:17:00
## 3 LGA    DFW    AA     2013-01-09 19:36:00
## 4 LGA    DFW    AA     2013-01-09 20:53:00
## 5 LGA    DFW    AA     2013-01-09 22:24:00
## 6 LGA    DFW    AA     2013-01-10 08:37:00
## 7 LGA    DFW    AA     2013-01-10 10:20:00
## 8 LGA    DFW    AA     2013-01-10 11:22:00
## 9 LGA    DFW    AA     2013-01-10 12:16:00
## 10 LGA   DFW    AA     2013-01-10 13:19:00
## 11 LGA   DFW    AA     2013-01-10 13:23:00
```

```
# 9: Prove a given time interval.
dt_1 <- ymd_hms("2013-01-09 16:16:00")
dt_2 <- ymd_hms("2013-01-09 19:17:00")

((dt_1 %--% dt_2) %/% minutes(1))
```

```
## [1] 181
```