Dates

Getting and Cleaning Data

http://lubridate.tidyverse.org/

date objects

date objects

```
> ymd("1988-09-29")
[1] "1988-09-29"
>
> mdy("September 29th, 1988")
[1] "1988-09-29"
>
> dmy("29-Sep-1988")
[1] "1988-09-29"
```

date-time object

```
> ymd_hms("1988-09-29 20:11:59")
[1] "1988-09-29 20:11:59 UTC"
```

```r
# Install the nycflights13 dataset
install.packages('nycflights13')
library(nycflights13)
```


Flights data

Description

On-time data for all flights that departed NYC (i.e. JFK, LGA or EWR) in 2013.

Usage

```
flights
```

Format

Data frame with columns

year,month,day

: Date of departure

dep_time,arr_time

: Actual departure and arrival times (format HHMM or HMM), local tz.

sched_dep_time,sched_arr_time

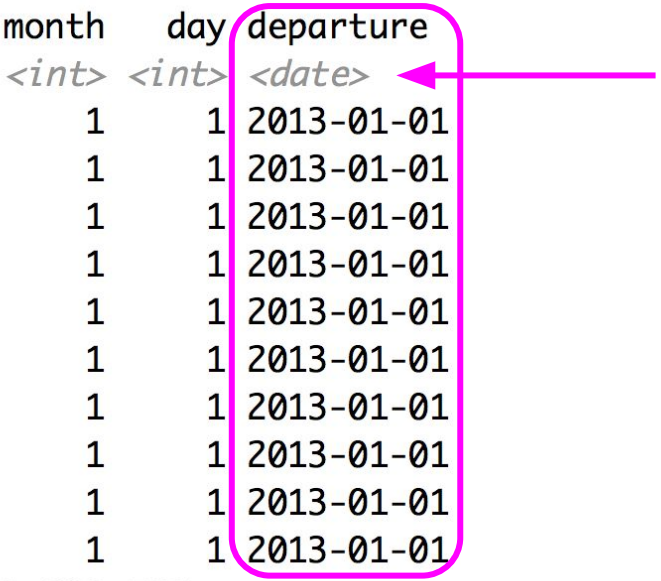
: Scheduled departure and arrival times (format HHMM or HMM), local tz.

dep_delay,arr_delay

: Departure and arrival delays, in minutes. Negative times represent early departures/arrivals.

```r
# make_date() creates a date object
# from information in separate columns
flights %>%
  select(year, month, day) %>%      # So only display the date columns during this test
  mutate(departure = make_date(year, month, day))    # Make a new column combining the date info
```

```
# A tibble: 336,776 x 4
    year month   day departure
   <int> <int> <int> <date>
 1  2013     1     1 2013-01-01
 2  2013     1     1 2013-01-01
 3  2013     1     1 2013-01-01
 4  2013     1     1 2013-01-01
 5  2013     1     1 2013-01-01
 6  2013     1     1 2013-01-01
 7  2013     1     1 2013-01-01
 8  2013     1     1 2013-01-01
 9  2013     1     1 2013-01-01
10  2013     1     1 2013-01-01
# ... with 336,766 more rows
```



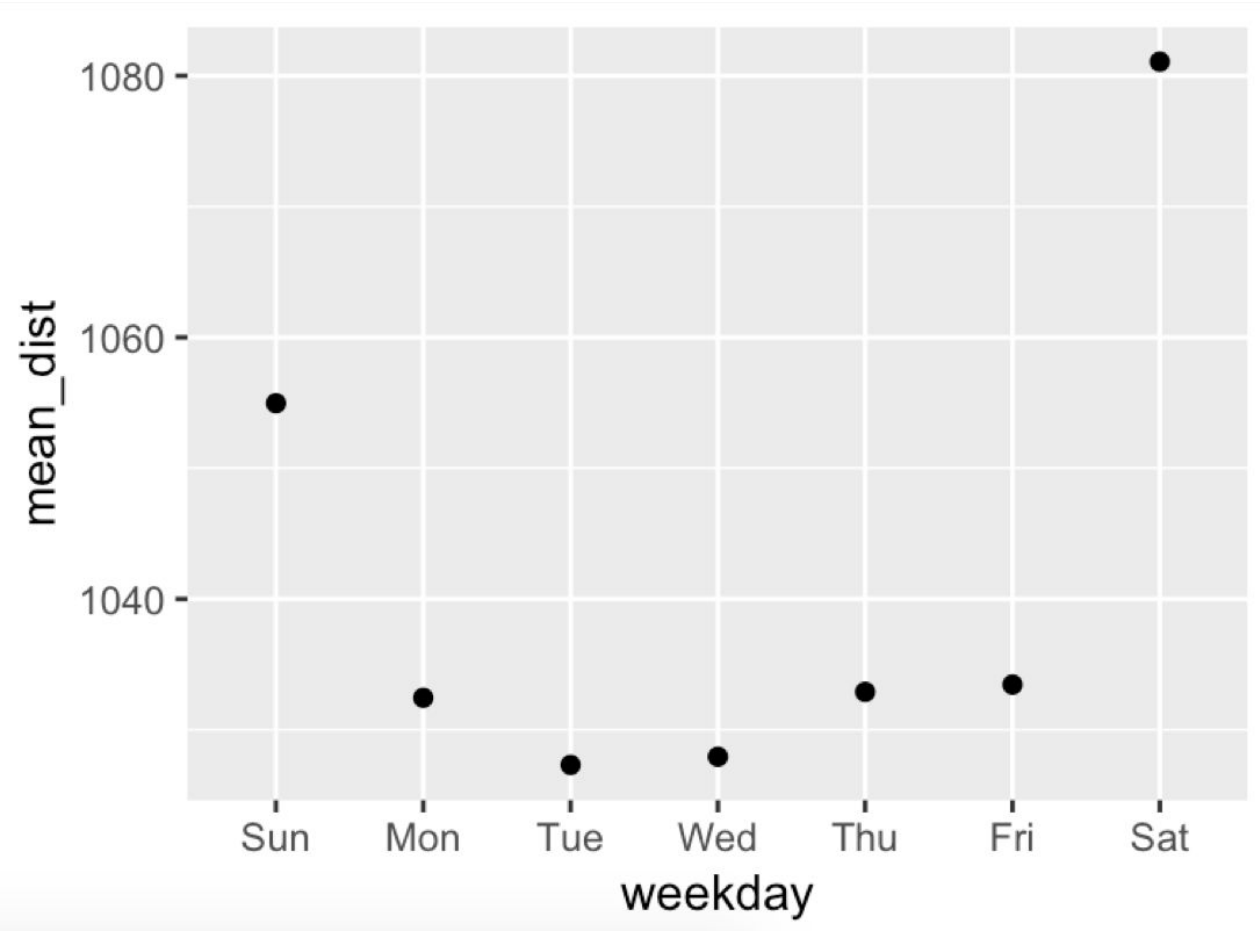
```r
# make_datetime() creates a date-time object
# from information in separate columns
flights %>%
  select(year, month, day, hour, minute) %>%
  mutate(departure = make_datetime(year, month, day, hour, minute))
```

```
# A tibble: 336,776 x 6
    year month   day  hour minute departure
   <int> <int> <int> <dbl>  <dbl> <dttm>
 1  2013     1     1     5     15 2013-01-01 05:15:00
 2  2013     1     1     5     29 2013-01-01 05:29:00
 3  2013     1     1     5     40 2013-01-01 05:40:00
 4  2013     1     1     5     45 2013-01-01 05:45:00
 5  2013     1     1     6      0 2013-01-01 06:00:00
 6  2013     1     1     5     58 2013-01-01 05:58:00
 7  2013     1     1     6      0 2013-01-01 06:00:00
 8  2013     1     1     6      0 2013-01-01 06:00:00
 9  2013     1     1     6      0 2013-01-01 06:00:00
10  2013     1     1     6      0 2013-01-01 06:00:00
# ... with 336,766 more rows
```

date-time object

```
> mydate <- ymd("1988-09-29")
>
> # Extract year information
> year(mydate)
[1] 1988
>
> # Extract day of the month
> mday(mydate)
[1] 29
>
> # Extract weekday information
> wday(mydate)
[1] 5
>
> # Label with actual day of the week
> wday(mydate, label = TRUE)
[1] Thu
Levels: Sun < Mon < Tue < Wed < Thu < Fri < Sat
```

```r
flights %>%
  mutate(departure = make_datetime(year, month, day, hour, minute)) %>%
  mutate(weekday = wday(departure, label = T)) %>%
  select(weekday, distance) %>%
  group_by(weekday) %>%
  summarize(mean_dist = mean(distance)) %>%
  ggplot(aes(x = weekday, y = mean_dist)) +
  geom_point()
```


1080
1060
1040
mean_dist
Sun
Mon
Tue
Wed
Thu
Fri
Sat
weekday

```r
# How old is someone born on Sept 29, 1988?

# Save their birthdate
birthdate <- ymd("1988-09-29")

# Subtract their birthday from today's date
age <- today() - birthdate
age # in days
```

```
Time difference of 11164 days
```

```r
# A duration object can get this information in years
as.duration(age) # in years
```

```
[1] "964569600s (~30.57 years)"
```

Summarizing: Dates

Getting and Cleaning Data