Stat 450 Dates and Times

Prof. Eric A. Suess

December 3, 2018

# Chapter 16 Dates and Times

library(tidyverse)

## ── Attaching packages ──────────────────────────────────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.1.0 ✔ purrr 0.2.5
## ✔ tibble 1.4.2 ✔ dplyr 0.7.7
## ✔ tidyr 0.8.2 ✔ stringr 1.3.1
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## ── Conflicts ─────────────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag() masks stats::lag()

library(lubridate)

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
## date

library(nycflights13)

today()

## [1] "2018-12-03"

now()

## [1] "2018-12-03 10:31:25 PST"

Dates from strings

ymd("2017-01-31")

## [1] "2017-01-31"

mdy("January 31st, 2017")

## [1] "2017-01-31"

dmy("31-Jan-2017")

## [1] "2017-01-31"

Dates with times

ymd\_hms("2017-01-31 20:11:59")

## [1] "2017-01-31 20:11:59 UTC"

mdy\_hm("01/31/2017 08:01")

## [1] "2017-01-31 08:01:00 UTC"

flights %>%
 select(year, month, day, hour, minute)

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

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

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),
 sched\_dep\_time = make\_datetime\_100(year, month, day, sched\_dep\_time),
 sched\_arr\_time = make\_datetime\_100(year, month, day, sched\_arr\_time)
 ) %>%
 select(origin, dest, ends\_with("delay"), ends\_with("time"))

flights\_dt

## # A tibble: 328,063 x 9
## origin dest dep\_delay arr\_delay dep\_time sched\_dep\_time
## <chr> <chr> <dbl> <dbl> <dttm> <dttm>
## 1 EWR IAH 2 11 2013-01-01 05:17:00 2013-01-01 05:15:00
## 2 LGA IAH 4 20 2013-01-01 05:33:00 2013-01-01 05:29:00
## 3 JFK MIA 2 33 2013-01-01 05:42:00 2013-01-01 05:40:00
## 4 JFK BQN -1 -18 2013-01-01 05:44:00 2013-01-01 05:45:00
## 5 LGA ATL -6 -25 2013-01-01 05:54:00 2013-01-01 06:00:00
## 6 EWR ORD -4 12 2013-01-01 05:54:00 2013-01-01 05:58:00
## 7 EWR FLL -5 19 2013-01-01 05:55:00 2013-01-01 06:00:00
## 8 LGA IAD -3 -14 2013-01-01 05:57:00 2013-01-01 06:00:00
## 9 JFK MCO -3 -8 2013-01-01 05:57:00 2013-01-01 06:00:00
## 10 LGA ORD -2 8 2013-01-01 05:58:00 2013-01-01 06:00:00
## # ... with 328,053 more rows, and 3 more variables: arr\_time <dttm>,
## # sched\_arr\_time <dttm>, air\_time <dbl>

flights\_dt %>%
 ggplot(aes(dep\_time)) +
 geom\_freqpoly(binwidth = 86400) # 86400 seconds = 1 day



flights\_dt %>%
 filter(dep\_time < ymd(20130102)) %>%
 ggplot(aes(dep\_time)) +
 geom\_freqpoly(binwidth = 600) # 600 s = 10 minutes



Getting the components

datetime <- ymd\_hms("2016-07-08 12:34:56")

year(datetime)

## [1] 2016

month(datetime)

## [1] 7

mday(datetime)

## [1] 8

yday(datetime)

## [1] 190

wday(datetime)

## [1] 6

flights\_dt %>%
 mutate(wday = wday(dep\_time, label = TRUE)) %>%
 ggplot(aes(x = wday)) +
 geom\_bar()



flights\_dt %>%
 mutate(minute = minute(dep\_time)) %>%
 group\_by(minute) %>%
 summarise(
 avg\_delay = mean(arr\_delay, na.rm = TRUE),
 n = n()) %>%
 ggplot(aes(minute, avg\_delay)) +
 geom\_line()



sched\_dep <- flights\_dt %>%
 mutate(minute = minute(sched\_dep\_time)) %>%
 group\_by(minute) %>%
 summarise(
 avg\_delay = mean(arr\_delay, na.rm = TRUE),
 n = n())

ggplot(sched\_dep, aes(minute, avg\_delay)) +
 geom\_line()



Time zones

Sys.timezone()

## [1] "America/Los\_Angeles"

All of the time zones

length(OlsonNames())

## [1] 606

head(OlsonNames())

## [1] "Africa/Abidjan" "Africa/Accra" "Africa/Addis\_Ababa"
## [4] "Africa/Algiers" "Africa/Asmara" "Africa/Asmera"

(x1 <- ymd\_hms("2015-06-01 12:00:00", tz = "America/New\_York"))

## [1] "2015-06-01 12:00:00 EDT"

(x2 <- ymd\_hms("2015-06-01 18:00:00", tz = "Europe/Copenhagen"))

## [1] "2015-06-01 18:00:00 CEST"

(x3 <- ymd\_hms("2015-06-02 04:00:00", tz = "Pacific/Auckland"))

## [1] "2015-06-02 04:00:00 NZST"