r4ds Ex 5.5.2

MW

2019/05/24

5.5.2

1

Currently `dep_time` and `sched_dep_time` are convenient to look at, but hard to compute with because they're not really continuous numbers. Convert them to a more convenient representation of number of minutes since midnight.

```
tomin <- function(nm){
    ifelse(is.na(nm), NA,
    ifelse(nchar(nm)<=2, nm,
         (nm %>% str_sub(start=1, end=-3) %>% as.numeric())*60 +
            nm %>% str_sub(start=-2, end=-1) %>% as.numeric()
    )
    )
}

flights %>% select(dep_time, sched_dep_time) %>%
    mutate(dep_min=tomin(dep_time), sched_dep_min=tomin(sched_dep_time))
```

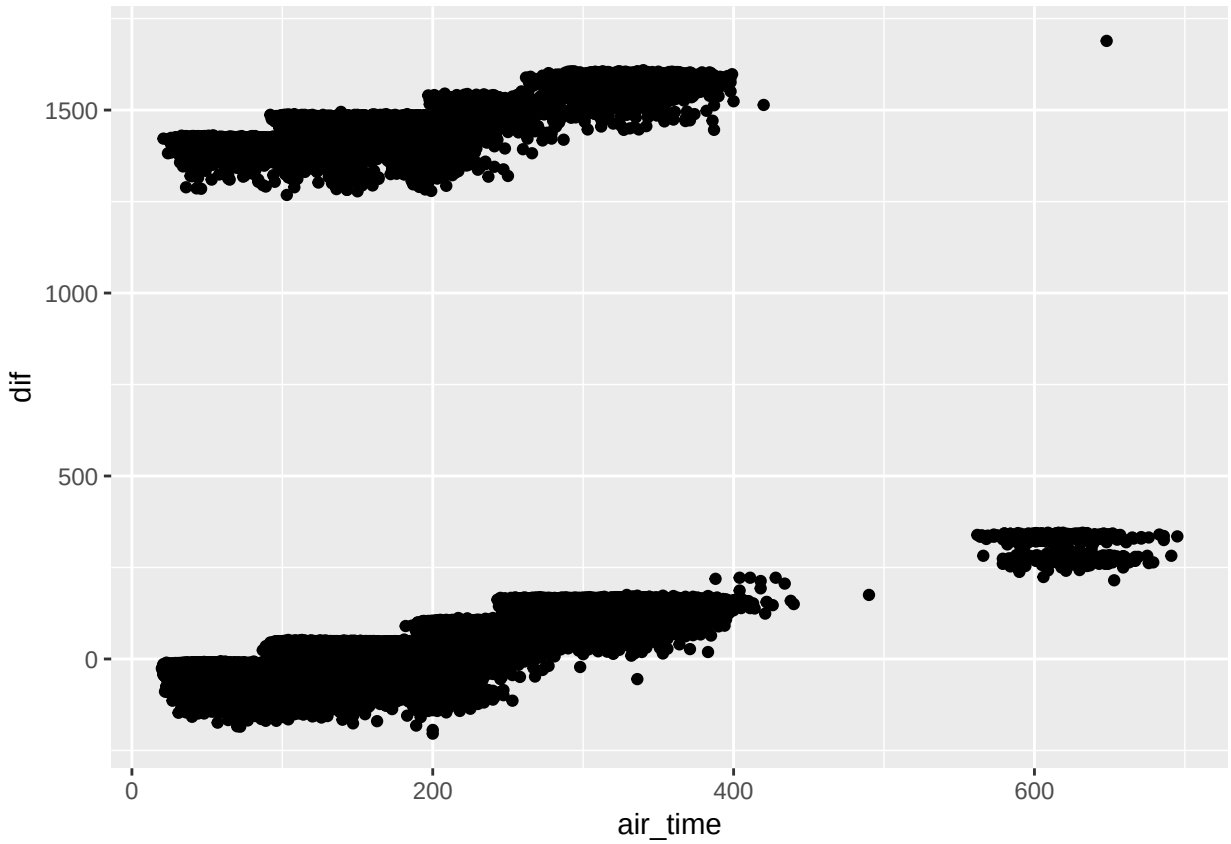
```
## # A tibble: 336,776 x 4
##    dep_time sched_dep_time dep_min sched_dep_min
##       <int>          <int>   <dbl>         <dbl>
##  1      517            515     317           315
##  2      533            529     333           329
##  3      542            540     342           340
##  4      544            545     344           345
##  5      554            600     354           360
##  6      554            558     354           358
##  7      555            600     355           360
##  8      557            600     357           360
##  9      557            600     357           360
## 10      558            600     358           360
## # ... with 336,766 more rows
```

2

Compare `air_time` with `arr_time - dep_time`. What do you expect to see? What do you see? What do you need to do to fix it?

```
smpl2 <- flights %>%
    mutate(arr_dep=(arr_time %>% tomin)-(dep_time %>% tomin), dif=air_time-arr_dep) %>%
    select(air_time, arr_dep, dif)
smpl2 %>% ggplot(aes(x=air_time, y=dif)) +
    geom_point()
```

```
## Warning: Removed 9430 rows containing missing values (geom_point).
```



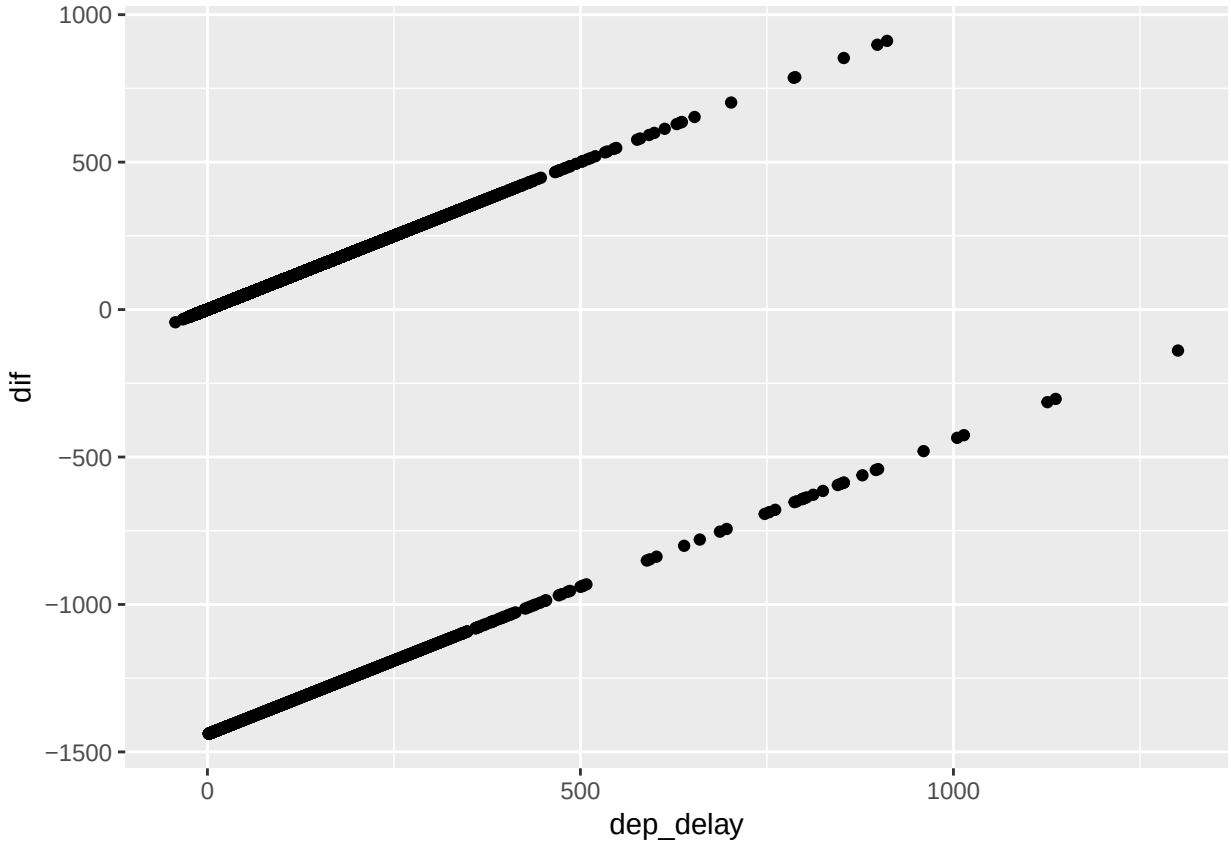
If `air_time` and `arr_dep` is equal, `y = x` line is expected. Negative y-axis values show to span Time Zone.

3

Compare `dep_time`, `sched_dep_time`, and `dep_delay`. How would you expect those three numbers to be related?

```
smpl3 <- flights %>%
    mutate(dif=tomin(dep_time)-tomin(sched_dep_time)) %>%
    select(dif, dep_delay)
smpl3 %>% ggplot(aes(x=dep_delay, y=dif)) +
    geom_point()
```

```
## Warning: Removed 8255 rows containing missing values (geom_point).
```



If `dep_min` and `sched_dep_min` is equal, `y = x` line is expected. Negative y-axis values show to span Time Zone.

4

Find the 10 most delayed flights using a ranking function. How do you want to handle ties? Carefully read the documentation for `min_rank()`.

```
flights %>%
    select(contains("delay")) %>%
    mutate(dep_min=min_rank(desc(dep_delay)), arr_min=min_rank(desc(arr_delay))) %>%
    filter(dep_min<=10 | arr_min<=10)
```

```
## # A tibble: 11 x 4
##    dep_delay arr_delay dep_min arr_min
##        <dbl>     <dbl>   <int>   <int>
##  1      1301      1272       1       1
##  2      1126      1109       3       3
##  3       896       878      10       9
##  4       911       915       7       7
##  5       960       931       6       6
##  6       878       875      11      10
##  7      1137      1127       2       2
##  8       899       850       8      14
##  9      1005       989       5       5
## 10       898       895       9       8
## 11      1014      1007       4       4
```

5

What does `1:3 + 1:10` return? Why?

The process between Vector1 and Vector2 can run till Vector2's loop finish. But the size doesn't match, then Vector1 repeats.

6

What trigonometric functions does R provide?

```
?trigonometric
```