

Tarea - Calculando nuevas variables con dplyr

Jesus Mudarra Luján

2022-10-31

Pregunta 1

El dataset de vuelos tiene dos variables, `dep_time` y `sched_dep_time` muy útiles pero difíciles de usar por cómo vienen dadas al no ser variables continuas. Fíjate que cuando pone 559, se refiere a que el vuelo salió a las 5:59...

Convierte este dato en otro más útil que represente el número de minutos que han transcurrido desde media noche.

```
mutate(flights,
  dep_time = paste0((dep_time%%100), ":", dep_time%%100),
  sched_dep_time = paste0((sched_dep_time%%100), ":", sched_dep_time%%100))
```

```
## # A tibble: 336,776 x 19
##   year month   day dep_time sched_de~1 dep_d~2 arr_t~3 sched~4 arr_d~5 carrier
##   <int> <int> <int> <chr>    <chr>         <dbl> <int>    <int>    <dbl> <chr>
## 1  2013     1     1 5:17      5:15           2     830     819     11 UA
## 2  2013     1     1 5:33      5:29           4     850     830     20 UA
## 3  2013     1     1 5:42      5:40           2     923     850     33 AA
## 4  2013     1     1 5:44      5:45          -1    1004    1022    -18 B6
## 5  2013     1     1 5:54      6:0          -6     812     837    -25 DL
## 6  2013     1     1 5:54      5:58          -4     740     728     12 UA
## 7  2013     1     1 5:55      6:0          -5     913     854     19 B6
## 8  2013     1     1 5:57      6:0          -3     709     723    -14 EV
## 9  2013     1     1 5:57      6:0          -3     838     846     -8 B6
## 10 2013     1     1 5:58      6:0          -2     753     745      8 AA
## # ... with 336,766 more rows, 9 more variables: flight <int>, tailnum <chr>,
## #   origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #   minute <dbl>, time_hour <dtm>, and abbreviated variable names
## #   1: sched_dep_time, 2: dep_delay, 3: arr_time, 4: sched_arr_time,
## #   5: arr_delay
```

Pregunta 2

Compara las variables `air_time` contra `arr_time - dep_time`.

- ¿Qué esperas ver?
- ¿Qué ves realmente?
- ¿Se te ocurre algo para mejorarlo y corregirlo?

Pregunta 3

Compara los valores de `dep_time`, `sched_dep_time` y `dep_delay`. Cómo deberían relacionarse estos tres números? Compruébalo y haz las correcciones numéricas que necesitas.

Pregunta 4

Usa una de las funciones de ranking para quedarte con los 10 vuelos más retrasados de todos.

Pregunta 5

Aunque la ejecución te dé una advertencia, qué resultado te da la operación

`1:6 + 1:20`

```
1:6 + 1:20
```

```
## Warning in 1:6 + 1:20: longitud de objeto mayor no es múltiplo de la longitud de  
## uno menor
```

```
## [1] 2 4 6 8 10 12 8 10 12 14 16 18 14 16 18 20 22 24 20 22
```

Cuando acaba la secuencia 1:6 vuelve a empezar, es decir, que a partir del séptimo elemento se sumaría 1 + 7, 2 + 8, 3 + 9, ..., 2 + 20.

Pregunta 6

Además de todas las funciones que hemos dicho, las trigonométricas también son funciones vectoriales que podemos usar para hacer transformaciones con `mutate`. Investiga cuales trae R y cual es la sintaxis de cada una de ellas.