dottable_hw

2025-03-04

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(data.table)
```

```
##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
##
## The following object is masked from 'package:purrr':
##
##     transpose
```

1) Use and show data.table code to select the variables year, month, day, and hour from the imported flights data

```
nycdata <- fread("nycdata.csv")

#use your own local folder. I was having issues with relative paths

filter1 <- nycdata[, .(year, month, day, hour)]

filter1
```

```
##          year month   day  hour
##         <int> <int> <int> <int>
##      1:  2014     1     1     9
##      2:  2014     1     1    11
##      3:  2014     1     1    19
##      4:  2014     1     1     7
##      5:  2014     1     1    13
##     ---
## 253312:  2014    10    31    14
## 253313:  2014    10    31     8
## 253314:  2014    10    31    11
## 253315:  2014    10    31    11
## 253316:  2014    10    31     8
```

2) Use and show data. table code to produce a table that shows a carrier of DL, an origin of JFK and a destination of SEA

```
filter2<- nycdata[carrier == "DL" & origin == "JFK" & dest == "SEA"]

filter2
```

```
##        year month   day dep_delay arr_delay carrier origin   dest air_time
##       <int> <int> <int>     <int>     <int>  <char> <char> <char>    <int>
##    1:  2014     1     1        86        79      DL    JFK    SEA      347
##    2:  2014     1     1        -2        -4      DL    JFK    SEA      347
##    3:  2014     1     2         0        11      DL    JFK    SEA      339
##    4:  2014     1     2        -3         9      DL    JFK    SEA      337
##    5:  2014     1     2        21        19      DL    JFK    SEA      337
##   ---
## 1074:  2014    10    30        -3       -15      DL    JFK    SEA      339
## 1075:  2014    10    31        -6       -26      DL    JFK    SEA      317
## 1076:  2014    10    31        -1        -8      DL    JFK    SEA      338
## 1077:  2014    10    31        -1       -23      DL    JFK    SEA      326
## 1078:  2014    10    31         4       -27      DL    JFK    SEA      318
##       distance  hour
##          <int> <int>
##    1:     2422     9
##    2:     2422    18
##    3:     2422    15
##    4:     2422     7
##    5:     2422    18
##   ---
## 1074:     2422    18
## 1075:     2422     9
## 1076:     2422     6
## 1077:     2422    15
## 1078:     2422    18
```

3) Use and show data.table code to produce a table that shows a carrier of UA, a month of March, and an airtime that is below 330.

```
filter3 <- nycdata[carrier == "UA" & month == 3 & air_time < 330]

filter3
```

```
##        year month   day dep_delay arr_delay carrier origin   dest air_time
##       <int> <int> <int>     <int>     <int>  <char> <char> <char>    <int>
##    1:  2014     3     1        11        43      UA    EWR    STT      209
##    2:  2014     3     1        47        13      UA    EWR    PBI      133
##    3:  2014     3     1        39        10      UA    EWR    MIA      139
##    4:  2014     3     1        -2       -12      UA    EWR    IAH      197
##    5:  2014     3     1        34        36      UA    EWR    DEN      256
##   ---
## 3785:  2014     3    31         6        -8      UA    EWR    FLL      155
## 3786:  2014     3    31         7        -9      UA    EWR    PBI      135
## 3787:  2014     3    31         1       -21      UA    EWR    RSW      145
## 3788:  2014     3    31         0       -19      UA    EWR    IAH      196
## 3789:  2014     3    31        18        -7      UA    EWR    ORD      108
##       distance  hour
##          <int> <int>
##    1:     1634     9
##    2:     1023    19
##    3:     1085    17
##    4:     1400     5
##    5:     1605    16
##   ---
## 3785:     1065    16
## 3786:     1023    10
## 3787:     1068    14
## 3788:     1400    16
## 3789:      719     6
```

4) Use and show tidyverse code to produce a table that shows a carrier of UA, a month of March, and an airtime that is below 330.

```
nycdata %>%
  filter(carrier == "UA", month == 3, air_time < 330) ->filter3

filter3
```

```
## Index: <dest__origin__carrier>
##        year month   day dep_delay arr_delay carrier origin   dest air_time
##       <int> <int> <int>     <int>     <int>  <char> <char> <char>    <int>
##    1:  2014     3     1        11        43      UA    EWR    STT      209
##    2:  2014     3     1        47        13      UA    EWR    PBI      133
##    3:  2014     3     1        39        10      UA    EWR    MIA      139
##    4:  2014     3     1        -2       -12      UA    EWR    IAH      197
##    5:  2014     3     1        34        36      UA    EWR    DEN      256
##   ---
## 3785:  2014     3    31         6        -8      UA    EWR    FLL      155
## 3786:  2014     3    31         7        -9      UA    EWR    PBI      135
## 3787:  2014     3    31         1       -21      UA    EWR    RSW      145
## 3788:  2014     3    31         0       -19      UA    EWR    IAH      196
```

```
## 3789:  2014     3    31        18        -7      UA    EWR    ORD      108
##       distance  hour
##          <int> <int>
##    1:     1634     9
##    2:     1023    19
##    3:     1085    17
##    4:     1400     5
##    5:     1605    16
##   ---
## 3785:     1065    16
## 3786:     1023    10
## 3787:     1068    14
## 3788:     1400    16
## 3789:      719     6
```

5) Use the data.table method to add a variable called speed that is the average air speed of the plane in miles per hour.

```
nycdata[, speed := (distance * 60) / air_time]

# Check if the new column is added
head(nycdata[, .(carrier, origin, dest, air_time, distance, speed)])
```

```
##    carrier origin   dest air_time distance    speed
##     <char> <char> <char>    <int>    <int>    <num>
## 1:      AA    JFK    LAX      359     2475 413.6490
## 2:      AA    JFK    LAX      363     2475 409.0909
## 3:      AA    JFK    LAX      351     2475 423.0769
## 4:      AA    LGA    PBI      157     1035 395.5414
## 5:      AA    JFK    LAX      350     2475 424.2857
## 6:      AA    EWR    LAX      339     2454 434.3363
```

6) Use the tidyverse method to add a variable called speed that is the average air speed of the plane in miles per hour.

```
nycdata <- nycdata %>%
  mutate(speed = distance / (air_time / 60)) # Speed = distance / time in hours

# Show the first few rows
head(nycdata)
```

```
##     year month   day dep_delay arr_delay carrier origin   dest air_time
##    <int> <int> <int>     <int>     <int>  <char> <char> <char>    <int>
## 1:  2014     1     1        14        13      AA    JFK    LAX      359
## 2:  2014     1     1        -3        13      AA    JFK    LAX      363
## 3:  2014     1     1         2         9      AA    JFK    LAX      351
## 4:  2014     1     1        -8       -26      AA    LGA    PBI      157
## 5:  2014     1     1         2         1      AA    JFK    LAX      350
## 6:  2014     1     1         4         0      AA    EWR    LAX      339
##    distance  hour    speed
##       <int> <int>    <num>
## 1:     2475     9 413.6490
```

```
## 2:     2475    11 409.0909
## 3:     2475    19 423.0769
## 4:     1035     7 395.5414
## 5:     2475    13 424.2857
## 6:     2454    18 434.3363
```

7) Show and use coding to change the carrier abbreviation of UA to UniitedAir, 7a. data.table method 7b. tidyverse method (Use a sequence of dplyr commands so that you can see the change in your table)

```
nycdata[carrier == "UA", carrier := "UnitedAir"]

unique(nycdata[, .(carrier)])
```

```
##       carrier
##        <char>
##  1:        AA
##  2:        AS
##  3:        B6
##  4:        DL
##  5:        EV
##  6:        F9
##  7:        FL
##  8:        HA
##  9:        MQ
## 10:        VX
## 11:        WN
## 12: UnitedAir
## 13:        US
## 14:        OO
```

```
# Problem 7b: Change "UA" to "UnitedAir" using tidyverse
nycdata <- nycdata %>%
  mutate(carrier = ifelse(carrier == "UA", "UnitedAir", carrier))

nycdata %>% select(carrier) %>% distinct()
```

```
##       carrier
##        <char>
##  1:        AA
##  2:        AS
##  3:        B6
##  4:        DL
##  5:        EV
##  6:        F9
##  7:        FL
##  8:        HA
##  9:        MQ
## 10:        VX
## 11:        WN
## 12: UnitedAir
## 13:        US
## 14:        OO
```