hw_dottable

Jack Kuo Chang

2025-03-02

```
setwd("~/hw_dottable")
library(tidyverse)
library(dplyr)
library(data.table)
nycdata <- fread("nycdata.csv")
#link: https://github.com/jkuoc183/hw_dottable
#ssh: git@github.com:jkuoc183/hw_dottable.git
```

```
#1. Use and show data.table code to select the variables year, month, day, and hour from the imported f
nycdata_col <- nycdata[ , .(year, month, day, hour)]
head(nycdata_col)
```

```
##     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
## 6:  2014     1     1    18
```

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

dl_jfk_sea <- nycdata[carrier == "DL" & origin =="JFK" & dest =="SEA"]
head(dl_jfk_sea)
```

```
##     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
## 6:  2014     1     3       579       556      DL    JFK    SEA      327
##    distance  hour
##       <int> <int>
## 1:     2422     9
## 2:     2422    18
## 3:     2422    15
## 4:     2422     7
## 5:     2422    18
## 6:     2422     0
```

```
#3. Use and show data.table code to produce a table that shows a carrier of UA, a month of March, and a
ua_march_airtime <- nycdata[carrier == "UA" & month == 3 & air_time < 330]
head(ua_march_airtime)
```

```
##     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
## 6:  2014     3     1        -2       -16      UA    EWR    TPA      139
##    distance  hour
##       <int> <int>
## 1:     1634     9
## 2:     1023    19
## 3:     1085    17
## 4:     1400     5
## 5:     1605    16
## 6:      997    13
```

```
#4. Use and show tidyverse code to produce a table that shows a carrier of UA, a month of March, and an
ua_march_airtime_tidy <-  nycdata |>
  filter(carrier == "UA", month == 3, air_time < 330)
head(ua_march_airtime_tidy)
```

```
##     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
## 6:  2014     3     1        -2       -16      UA    EWR    TPA      139
##    distance  hour
##       <int> <int>
## 1:     1634     9
## 2:     1023    19
## 3:     1085    17
## 4:     1400     5
## 5:     1605    16
## 6:      997    13
```

```
#5. Use the data.table method to add a variable called speed that is the average air speed of the plane
nycdata[ , speed := round(distance/(air_time/60),2)]
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     9 413.65
## 2:     2475    11 409.09
## 3:     2475    19 423.08
## 4:     1035     7 395.54
## 5:     2475    13 424.29
## 6:     2454    18 434.34
```

```
#6. Use the tidyverse method to add a variable called speed that is the average air speed of the plane
nycdata<- nycdata |>
  mutate(speed = round(distance/ (air_time/60),2))
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     9 413.65
## 2:     2475    11 409.09
## 3:     2475    19 423.08
## 4:     1035     7 395.54
## 5:     2475    13 424.29
## 6:     2454    18 434.34
```

```
#7. Show and use coding to change the carrier abbreviation of UA to UnitedAir,
#7a. data.table method
nycdata[carrier == "UA", carrier := "UnitedAir"]
unique(nycdata$carrier)
```

```
##  [1] "AA"        "AS"        "B6"        "DL"        "EV"        "F9"
##  [7] "FL"        "HA"        "MQ"        "VX"        "WN"        "UnitedAir"
## [13] "US"        "OO"
```

```
#7b. tidyverse method (Use a sequence of dplyr commands so that you can see the change in your table)

nycdata <- nycdata |>
 mutate(carrier = ifelse(carrier == "UA", "UnitedAir",carrier))
nycdata |>
  count(carrier)
```

```
##       carrier     n
##        <char> <int>
##  1:        AA 26302
##  2:        AS   574
##  3:        B6 44479
##  4:        DL 41683
##  5:        EV 39819
##  6:        F9   473
##  7:        FL  1251
```

```
##  8:        HA   260
##  9:        MQ 18559
## 10:        OO   200
## 11:        US 16750
## 12: UnitedAir 46267
## 13:        VX  4797
## 14:        WN 11902
```