dottable homework

Gabriela C. Lebron & Alasya Zeweldi

2025-03-04

Go to GitHub and download the data nycdata.csv and then answer the following questions.

```
library(RCurl)
```

```
## Warning: package 'RCurl' was built under R version 4.4.1
```

```
##
## Attaching package: 'RCurl'
```

```
## The following object is masked from 'package:tidyr':
##
##     complete
```

```
x <- getURL("https://raw.githubusercontent.com/semiyarih/DATA613/refs/heads/main/dottable/nycdata.csv")
flights <- read.csv(text = x)
flights.data <- data.table(flights)
head(flights, 6)
```

```
##   year month day dep_delay arr_delay carrier origin dest air_time distance hour
## 1 2014     1   1        14        13      AA    JFK  LAX      359     2475    9
## 2 2014     1   1        -3        13      AA    JFK  LAX      363     2475   11
## 3 2014     1   1         2         9      AA    JFK  LAX      351     2475   19
## 4 2014     1   1        -8       -26      AA    LGA  PBI      157     1035    7
## 5 2014     1   1         2         1      AA    JFK  LAX      350     2475   13
## 6 2014     1   1         4         0      AA    EWR  LAX      339     2454   18
```

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

```
flights.data[, .(year, month, day, hour)]
```

```
##          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

```
flights.data[carrier == "DL" & origin == "JFK" & dest == "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
##   ---
## 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.

```
flights.data[carrier == "UA" & month == 3 & air_time < 330]
```

```
##        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.

```
flights %>%
  filter(carrier == "UA" & month == 3 & air_time < 330) %>%
  head(6)
```

```
##   year month day dep_delay arr_delay carrier origin dest air_time distance hour
## 1 2014     3   1        11        43      UA    EWR  STT      209     1634    9
## 2 2014     3   1        47        13      UA    EWR  PBI      133     1023   19
## 3 2014     3   1        39        10      UA    EWR  MIA      139     1085   17
## 4 2014     3   1        -2       -12      UA    EWR  IAH      197     1400    5
## 5 2014     3   1        34        36      UA    EWR  DEN      256     1605   16
## 6 2014     3   1        -2       -16      UA    EWR  TPA      139      997   13
```

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.

```
flights.data[,speed := distance/hour]
head(flights.data)
```

```
##     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     9 275.0000
## 2:     2475    11 225.0000
## 3:     2475    19 130.2632
## 4:     1035     7 147.8571
## 5:     2475    13 190.3846
## 6:     2454    18 136.3333
```

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

```
flights %>%
  mutate(speed = distance/hour) %>%
  head(6)
```

```
##   year month day dep_delay arr_delay carrier origin dest air_time distance hour
## 1 2014     1   1        14        13      AA    JFK  LAX      359     2475    9
## 2 2014     1   1        -3        13      AA    JFK  LAX      363     2475   11
## 3 2014     1   1         2         9      AA    JFK  LAX      351     2475   19
## 4 2014     1   1        -8       -26      AA    LGA  PBI      157     1035    7
## 5 2014     1   1         2         1      AA    JFK  LAX      350     2475   13
## 6 2014     1   1         4         0      AA    EWR  LAX      339     2454   18
##      speed
## 1 275.0000
## 2 225.0000
## 3 130.2632
## 4 147.8571
## 5 190.3846
## 6 136.3333
```

7. Show and use coding to change the carrier abbreviation of UA to UniitedAir,

a. data.table method

```
flights.data[carrier == "UA", carrier := "UniitedAir"]
unique(flights.data$carrier)
```

```
##  [1] "AA"         "AS"         "B6"         "DL"         "EV"
##  [6] "F9"         "FL"         "HA"         "MQ"         "VX"
## [11] "WN"         "UniitedAir" "US"         "OO"
```

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

```
flights %>%
  mutate(carrier = ifelse(carrier == "UA", "UniitedAir", carrier)) %>%
  filter(carrier == "UniitedAir") %>%
  head(6)
```

```
##   year month day dep_delay arr_delay    carrier origin dest air_time distance
## 1 2014     1   1         9        -2 UniitedAir    EWR  HNL      630     4963
## 2 2014     1   1        25        17 UniitedAir    EWR  TPA      149      997
## 3 2014     1   1        49        57 UniitedAir    EWR  TPA      157      997
## 4 2014     1   1         0         9 UniitedAir    EWR  TPA      171      997
## 5 2014     1   1         8        -1 UniitedAir    EWR  SAT      235     1569
## 6 2014     1   1        43        42 UniitedAir    EWR  MIA      155     1085
##   hour
## 1    9
## 2   12
## 3   18
## 4   20
## 5   17
## 6   15
```