Homework 4: Biggish Data dottable

Sloane Johns and Avery Rathman

2025-03-04

```
# libraries
library(tidyverse)
library(dplyr)
library(data.table)

# data
nycdata <- fread("~/hw_dottable/nycdata.csv")
```

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

```
head(nycdata[, .(year, month, day, hour)], 6)
```

```
##     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
```

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

```
head(nycdata[carrier == "DL" & origin == "JFK" & dest == "SEA"], 6)
```

```
##     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
```

Question 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.

```
head(nycdata[carrier == "UA" & month == 3 & air_time < 330], 6)
```

```
##     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
```

Question 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) %>% head(6)
```

```
##     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
```

Question 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[, c("speed") := .(distance / (air_time / 60))]
head(nycdata, 6)
```

```
##     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
```

Question 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 %>% mutate(speed = distance / (air_time / 60)) %>% head(6)
```

```
##     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
```

Question 7: Show and use coding to change the carrier abbreviation of UA to UnitedAir.

```
# data.table method
nycdata_newname <- nycdata[carrier == "UA", carrier := "UnitedAir"]
sort(unique(nycdata_newname$carrier))
```

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

```r
# tidyverse method
nycdata %>% mutate(carrier = case_match(carrier, "UA" ~ "UnitedAir", .default = carrier)) %>%
  filter(carrier == "UnitedAir") %>% head(6)
```

```
##     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         9        -2 UnitedAir    EWR    HNL      630
## 2:  2014     1     1        25        17 UnitedAir    EWR    TPA      149
## 3:  2014     1     1        49        57 UnitedAir    EWR    TPA      157
## 4:  2014     1     1         0         9 UnitedAir    EWR    TPA      171
## 5:  2014     1     1         8        -1 UnitedAir    EWR    SAT      235
## 6:  2014     1     1        43        42 UnitedAir    EWR    MIA      155
##    distance  hour    speed
##       <int> <int>    <num>
## 1:     4963     9 472.6667
## 2:      997    12 401.4765
## 3:      997    18 381.0191
## 4:      997    20 349.8246
## 5:     1569    17 400.5957
## 6:     1085    15 420.0000
```