dottable

Andrew Hwang

https://github.com/Hwangster1/dottable

```
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 1.9.3     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 becor
```

```
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
```

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

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

```
selected_flights <- nyc_data[, .(year, month, day, hour)]
str(selected_flights)
```

```
Classes 'data.table' and 'data.frame':  253316 obs. of  4 variables:
 $ year : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
 $ month: int  1 1 1 1 1 1 1 1 1 1 ...
 $ day  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ hour : int  9 11 19 7 13 18 21 15 15 18 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

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

```
filtered_flights <- nyc_data[carrier == "DL" & origin == "JFK" & dest == "SEA"]

str(filtered_flights)
```

```
Classes 'data.table' and 'data.frame':  1078 obs. of  11 variables:
 $ year     : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
 $ month    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ day      : int  1 1 2 2 2 3 3 4 4 5 ...
 $ dep_delay: int  86 -2 0 -3 21 579 370 22 72 142 ...
 $ arr_delay: int  79 -4 11 9 19 556 387 9 35 114 ...
 $ carrier  : chr  "DL" "DL" "DL" "DL" ...
 $ origin   : chr  "JFK" "JFK" "JFK" "JFK" ...
 $ dest     : chr  "SEA" "SEA" "SEA" "SEA" ...
 $ air_time : int  347 347 339 337 337 327 341 332 333 322 ...
 $ distance : int  2422 2422 2422 2422 2422 2422 2422 2422 2422 2422 ...
 $ hour     : int  9 18 15 7 18 0 13 15 19 17 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

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.

```
filtered_flights_2 <- nyc_data[carrier == "UA" & month == 3 & air_time < 330]
str(filtered_flights_2)
```

```
Classes 'data.table' and 'data.frame':  3789 obs. of  11 variables:
 $ year     : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
 $ month    : int  3 3 3 3 3 3 3 3 3 3 ...
 $ day      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ dep_delay: int  11 47 39 -2 34 -2 1 251 5 19 ...
 $ arr_delay: int  43 13 10 -12 36 -16 -2 205 17 -5 ...
 $ carrier  : chr  "UA" "UA" "UA" "UA" ...
 $ origin   : chr  "EWR" "EWR" "EWR" "EWR" ...
 $ dest     : chr  "STT" "PBI" "MIA" "IAH" ...
 $ air_time : int  209 133 139 197 256 139 123 127 243 140 ...
 $ distance : int  1634 1023 1085 1400 1605 997 719 997 1605 1085 ...
 $ hour     : int  9 19 17 5 16 13 13 23 7 21 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

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.

```
filtered_flights_3 <- nyc_data |>
  filter(carrier == "UA", month == 3, air_time < 330)
head(filtered_flights_3)
```

```
    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
```

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

```
flights1 <- nyc_data[, speed := (distance / air_time) * 60]
str(flights1)
```

```
Classes 'data.table' and 'data.frame':  253316 obs. of  12 variables:
 $ year     : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
 $ month    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ day      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ dep_delay: int  14 -3 2 -8 2 4 -2 -3 -1 -2 ...
 $ arr_delay: int  13 13 9 -26 1 0 -18 -14 -17 -14 ...
 $ carrier  : chr  "AA" "AA" "AA" "AA" ...
 $ origin   : chr  "JFK" "JFK" "JFK" "LGA" ...
 $ dest     : chr  "LAX" "LAX" "LAX" "PBI" ...
 $ air_time : int  359 363 351 157 350 339 338 356 161 349 ...
 $ distance : int  2475 2475 2475 1035 2475 2454 2475 2475 1089 2422 ...
 $ hour     : int  9 11 19 7 13 18 21 15 15 18 ...
 $ speed    : num  414 409 423 396 424 ...
 - attr(*, ".internal.selfref")=<externalptr>
 - attr(*, "index")= int(0)
  ..- attr(*, "__dest__origin__carrier")= int [1:253316] 185 858 3890 4544 5416 6246 7097 840
```

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

```
flights2 <- nyc_data |>
  mutate(speed = (distance / air_time) * 60)
head(flights2)
```

```
    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
```

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

7a. data.table method

```
united_1<- nyc_data[carrier == "UA", carrier := "UnitedAir"]
str(united_1)
```

```
Classes 'data.table' and 'data.frame':  253316 obs. of  12 variables:
 $ year     : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
 $ month    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ day      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ dep_delay: int  14 -3 2 -8 2 4 -2 -3 -1 -2 ...
 $ arr_delay: int  13 13 9 -26 1 0 -18 -14 -17 -14 ...
 $ carrier  : chr  "AA" "AA" "AA" "AA" ...
 $ origin   : chr  "JFK" "JFK" "JFK" "LGA" ...
 $ dest     : chr  "LAX" "LAX" "LAX" "PBI" ...
 $ air_time : int  359 363 351 157 350 339 338 356 161 349 ...
 $ distance : int  2475 2475 2475 1035 2475 2454 2475 2475 1089 2422 ...
 $ hour     : int  9 11 19 7 13 18 21 15 15 18 ...
 $ speed    : num  414 409 423 396 424 ...
 - attr(*, ".internal.selfref")=<externalptr>
 - attr(*, "index")= int(0)
```

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

```
united_2 <- nyc_data |>
  mutate(carrier = ifelse(carrier == "UA", "UnitedAir", carrier))
```