

data.table Homework

Carson Cherniss and Ainsley Gallagher

2025-03-04

```
library(tidyverse)
library(data.table)
```

```
flights1 <- fread("nycddata.csv")
flights2 <- read_csv("nycddata.csv")
```

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

```
flights1[, c("year", "month", "day", "hour")]
```

	year	month	day	hour
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

```
flights1[carrier == "DL" & origin == "JFK" & dest == "SEA", c("carrier","origin","dest")]
```

	carrier	origin	dest
1:	DL	JFK	SEA
2:	DL	JFK	SEA
3:	DL	JFK	SEA
4:	DL	JFK	SEA
5:	DL	JFK	SEA

1074:	DL	JFK	SEA
1075:	DL	JFK	SEA
1076:	DL	JFK	SEA
1077:	DL	JFK	SEA
1078:	DL	JFK	SEA

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.

```
flights1[carrier == "UA" & month == 3 & air_time < 330, c("carrier", "month", "air_time")]
```

	carrier	month	air_time
1:	UA	3	209
2:	UA	3	133
3:	UA	3	139
4:	UA	3	197
5:	UA	3	256

3785:	UA	3	155
3786:	UA	3	135
3787:	UA	3	145
3788:	UA	3	196
3789:	UA	3	108

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.

```
flights2 |>
  select(carrier, month, air_time) |>
  filter(carrier == "UA", month == 3, air_time < 330)
```

```
# A tibble: 3,789 x 3
  carrier month air_time
  <chr>    <dbl>    <dbl>
1 UA         3      209
2 UA         3      133
3 UA         3      139
4 UA         3      197
5 UA         3      256
6 UA         3      139
7 UA         3      123
8 UA         3      127
9 UA         3      243
10 UA        3      140
# i 3,779 more rows
```

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.

```
flights1[, speed := (distance / air_time) * 60]
flights1[, c("distance", "air_time", "speed")]
```

```
      distance air_time    speed
1:      2475      359 413.6490
2:      2475      363 409.0909
3:      2475      351 423.0769
4:      1035      157 395.5414
5:      2475      350 424.2857
---
253312:    1416      201 422.6866
253313:    1400      189 444.4444
253314:     431       83 311.5663
253315:     502       75 401.6000
253316:     659      110 359.4545
```

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

```
flights2 |>
  mutate(speed = (distance / air_time) * 60) |>
  select(distance, air_time, speed)
```

```
# A tibble: 253,316 x 3
  distance air_time speed
  <dbl>     <dbl> <dbl>
1    2475       359  414.
2    2475       363  409.
3    2475       351  423.
4    1035       157  396.
5    2475       350  424.
6    2454       339  434.
7    2475       338  439.
8    2475       356  417.
9    1089       161  406.
10   2422       349  416.
# i 253,306 more rows
```

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

7a: data.table method

```
flights1[carrier == "UA", carrier := "UnitedAir"]

flights1[carrier == "UnitedAir", .(carrier, origin, dest)]
```

```
      carrier origin dest
1: UnitedAir    EWR  HNL
2: UnitedAir    EWR  TPA
3: UnitedAir    EWR  TPA
4: UnitedAir    EWR  TPA
5: UnitedAir    EWR  SAT
---
46263: UnitedAir    EWR  ORD
46264: UnitedAir    EWR  LAX
```

```
46265: UnitedAir    EWR  LAS
46266: UnitedAir    LGA  IAH
46267: UnitedAir    EWR  IAH
```

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

```
flights2 <- flights2 |>
  mutate(carrier = ifelse(carrier == "UA", "UnitedAir", carrier))

flights2 |>
  filter(carrier == "UnitedAir") |>
  select(carrier, origin, dest) |>
  head()
```

```
# A tibble: 6 x 3
  carrier origin dest
  <chr>    <chr> <chr>
1 UnitedAir EWR    HNL
2 UnitedAir EWR    TPA
3 UnitedAir EWR    TPA
4 UnitedAir EWR    TPA
5 UnitedAir EWR    SAT
6 UnitedAir EWR    MIA
```