DATA 413: Classwork/Lab 6

Frankie Tyndall

2022-07-19

## Dataset by data.table

flight3

## year month day dep\_delay arr\_delay carrier origin dest air\_time  
## 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  
## ---   
## 253312: 2014 10 31 1 -30 UA LGA IAH 201  
## 253313: 2014 10 31 -5 -14 UA EWR IAH 189  
## 253314: 2014 10 31 -8 16 MQ LGA RDU 83  
## 253315: 2014 10 31 -4 15 MQ LGA DTW 75  
## 253316: 2014 10 31 -5 1 MQ LGA SDF 110  
## distance hour  
## 1: 2475 9  
## 2: 2475 11  
## 3: 2475 19  
## 4: 1035 7  
## 5: 2475 13  
## ---   
## 253312: 1416 14  
## 253313: 1400 8  
## 253314: 431 11  
## 253315: 502 11  
## 253316: 659 8

## Dataset by Tidyverse

You can also embed plots, for example:

flight4

## # A tibble: 253,316 × 11  
## year month day dep\_delay arr\_delay carrier origin dest air\_time distance  
## <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>  
## 1 2014 1 1 14 13 AA JFK LAX 359 2475  
## 2 2014 1 1 -3 13 AA JFK LAX 363 2475  
## 3 2014 1 1 2 9 AA JFK LAX 351 2475  
## 4 2014 1 1 -8 -26 AA LGA PBI 157 1035  
## 5 2014 1 1 2 1 AA JFK LAX 350 2475  
## 6 2014 1 1 4 0 AA EWR LAX 339 2454  
## 7 2014 1 1 -2 -18 AA JFK LAX 338 2475  
## 8 2014 1 1 -3 -14 AA JFK LAX 356 2475  
## 9 2014 1 1 -1 -17 AA JFK MIA 161 1089  
## 10 2014 1 1 -2 -14 AA JFK SEA 349 2422  
## # … with 253,306 more rows, and 1 more variable: hour <dbl>

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

variables\_ymd <- flight3[, .(year, month, day, hour)]  
variables\_ymd

## 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

## 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

carrier\_fl <- flight3[origin == "JFK" & dest == "SEA" & carrier == "DL"]  
carrier\_fl

## year month day dep\_delay arr\_delay carrier origin dest air\_time distance  
## 1: 2014 1 1 86 79 DL JFK SEA 347 2422  
## 2: 2014 1 1 -2 -4 DL JFK SEA 347 2422  
## 3: 2014 1 2 0 11 DL JFK SEA 339 2422  
## 4: 2014 1 2 -3 9 DL JFK SEA 337 2422  
## 5: 2014 1 2 21 19 DL JFK SEA 337 2422  
## ---   
## 1074: 2014 10 30 -3 -15 DL JFK SEA 339 2422  
## 1075: 2014 10 31 -6 -26 DL JFK SEA 317 2422  
## 1076: 2014 10 31 -1 -8 DL JFK SEA 338 2422  
## 1077: 2014 10 31 -1 -23 DL JFK SEA 326 2422  
## 1078: 2014 10 31 4 -27 DL JFK SEA 318 2422  
## hour  
## 1: 9  
## 2: 18  
## 3: 15  
## 4: 7  
## 5: 18  
## ---   
## 1074: 18  
## 1075: 9  
## 1076: 6  
## 1077: 15  
## 1078: 18

## 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.

airtime\_330 <- flight3[carrier == "UA" & month == 3 & air\_time < 330]  
airtime\_330

## year month day dep\_delay arr\_delay carrier origin dest air\_time distance  
## 1: 2014 3 1 11 43 UA EWR STT 209 1634  
## 2: 2014 3 1 47 13 UA EWR PBI 133 1023  
## 3: 2014 3 1 39 10 UA EWR MIA 139 1085  
## 4: 2014 3 1 -2 -12 UA EWR IAH 197 1400  
## 5: 2014 3 1 34 36 UA EWR DEN 256 1605  
## ---   
## 3785: 2014 3 31 6 -8 UA EWR FLL 155 1065  
## 3786: 2014 3 31 7 -9 UA EWR PBI 135 1023  
## 3787: 2014 3 31 1 -21 UA EWR RSW 145 1068  
## 3788: 2014 3 31 0 -19 UA EWR IAH 196 1400  
## 3789: 2014 3 31 18 -7 UA EWR ORD 108 719  
## hour  
## 1: 9  
## 2: 19  
## 3: 17  
## 4: 5  
## 5: 16  
## ---   
## 3785: 16  
## 3786: 10  
## 3787: 14  
## 3788: 16  
## 3789: 6

## 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.

airtime\_fl <- flight4 %>%  
 filter(carrier == "UA", month == 3, air\_time < 330)  
airtime\_fl

## # A tibble: 3,789 × 11  
## year month day dep\_delay arr\_delay carrier origin dest air\_time distance  
## <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>  
## 1 2014 3 1 11 43 UA EWR STT 209 1634  
## 2 2014 3 1 47 13 UA EWR PBI 133 1023  
## 3 2014 3 1 39 10 UA EWR MIA 139 1085  
## 4 2014 3 1 -2 -12 UA EWR IAH 197 1400  
## 5 2014 3 1 34 36 UA EWR DEN 256 1605  
## 6 2014 3 1 -2 -16 UA EWR TPA 139 997  
## 7 2014 3 1 1 -2 UA EWR ORD 123 719  
## 8 2014 3 1 251 205 UA EWR TPA 127 997  
## 9 2014 3 1 5 17 UA EWR DEN 243 1605  
## 10 2014 3 1 19 -5 UA EWR MIA 140 1085  
## # … with 3,779 more rows, and 1 more variable: hour <dbl>

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

flight3

## year month day dep\_delay arr\_delay carrier origin dest air\_time  
## 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  
## ---   
## 253312: 2014 10 31 1 -30 UA LGA IAH 201  
## 253313: 2014 10 31 -5 -14 UA EWR IAH 189  
## 253314: 2014 10 31 -8 16 MQ LGA RDU 83  
## 253315: 2014 10 31 -4 15 MQ LGA DTW 75  
## 253316: 2014 10 31 -5 1 MQ LGA SDF 110  
## distance hour  
## 1: 2475 9  
## 2: 2475 11  
## 3: 2475 19  
## 4: 1035 7  
## 5: 2475 13  
## ---   
## 253312: 1416 14  
## 253313: 1400 8  
## 253314: 431 11  
## 253315: 502 11  
## 253316: 659 8

sp <- flight3[, c("speed") := .(distance/(air\_time/60))]  
sp

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

perhour\_tideverse <- flight4 %>%  
 mutate(speed = (distance/(air\_time/60)))  
perhour\_tideverse

## # A tibble: 253,316 × 12  
## year month day dep\_delay arr\_delay carrier origin dest air\_time distance  
## <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>  
## 1 2014 1 1 14 13 AA JFK LAX 359 2475  
## 2 2014 1 1 -3 13 AA JFK LAX 363 2475  
## 3 2014 1 1 2 9 AA JFK LAX 351 2475  
## 4 2014 1 1 -8 -26 AA LGA PBI 157 1035  
## 5 2014 1 1 2 1 AA JFK LAX 350 2475  
## 6 2014 1 1 4 0 AA EWR LAX 339 2454  
## 7 2014 1 1 -2 -18 AA JFK LAX 338 2475  
## 8 2014 1 1 -3 -14 AA JFK LAX 356 2475  
## 9 2014 1 1 -1 -17 AA JFK MIA 161 1089  
## 10 2014 1 1 -2 -14 AA JFK SEA 349 2422  
## # … with 253,306 more rows, and 2 more variables: hour <dbl>, speed <dbl>

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

# data.table method  
flight3

## year month day dep\_delay arr\_delay carrier origin dest air\_time  
## 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  
## ---   
## 253312: 2014 10 31 1 -30 UA LGA IAH 201  
## 253313: 2014 10 31 -5 -14 UA EWR IAH 189  
## 253314: 2014 10 31 -8 16 MQ LGA RDU 83  
## 253315: 2014 10 31 -4 15 MQ LGA DTW 75  
## 253316: 2014 10 31 -5 1 MQ LGA SDF 110  
## distance hour speed  
## 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  
## ---   
## 253312: 1416 14 422.6866  
## 253313: 1400 8 444.4444  
## 253314: 431 11 311.5663  
## 253315: 502 11 401.6000  
## 253316: 659 8 359.4545

ua <- flight3[carrier == "UA", carrier := "UniitedAir"]  
ua[carrier == "UniitedAir"]

## year month day dep\_delay arr\_delay carrier origin dest air\_time  
## 1: 2014 1 1 9 -2 UniitedAir EWR HNL 630  
## 2: 2014 1 1 25 17 UniitedAir EWR TPA 149  
## 3: 2014 1 1 49 57 UniitedAir EWR TPA 157  
## 4: 2014 1 1 0 9 UniitedAir EWR TPA 171  
## 5: 2014 1 1 8 -1 UniitedAir EWR SAT 235  
## ---   
## 46263: 2014 10 31 427 393 UniitedAir EWR ORD 100  
## 46264: 2014 10 31 10 -27 UniitedAir EWR LAX 326  
## 46265: 2014 10 31 18 -14 UniitedAir EWR LAS 291  
## 46266: 2014 10 31 1 -30 UniitedAir LGA IAH 201  
## 46267: 2014 10 31 -5 -14 UniitedAir EWR IAH 189  
## distance hour speed  
## 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  
## ---   
## 46263: 719 21 431.4000  
## 46264: 2454 10 451.6564  
## 46265: 2227 16 459.1753  
## 46266: 1416 14 422.6866  
## 46267: 1400 8 444.4444

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

# tidyverse method   
  
changeUA <- flight4 %>%  
 mutate(carrier = recode(carrier, "UA" = "UniitedAir")) %>%  
 filter(carrier == "UniitedAir")  
changeUA

## # A tibble: 46,267 × 11  
## year month day dep\_delay arr\_delay carrier origin dest air\_time distance  
## <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>  
## 1 2014 1 1 9 -2 Uniited… EWR HNL 630 4963  
## 2 2014 1 1 25 17 Uniited… EWR TPA 149 997  
## 3 2014 1 1 49 57 Uniited… EWR TPA 157 997  
## 4 2014 1 1 0 9 Uniited… EWR TPA 171 997  
## 5 2014 1 1 8 -1 Uniited… EWR SAT 235 1569  
## 6 2014 1 1 43 42 Uniited… EWR MIA 155 1085  
## 7 2014 1 1 10 4 Uniited… EWR PBI 155 1023  
## 8 2014 1 1 0 11 Uniited… EWR TPA 162 997  
## 9 2014 1 1 5 -3 Uniited… EWR RSW 165 1068  
## 10 2014 1 1 0 8 Uniited… LGA ORD 131 733  
## # … with 46,257 more rows, and 1 more variable: hour <dbl>