Lab\_March24

Elizabeth Marge

3/24/2021

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.5 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(dplyr)  
library(data.table)

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

flights1 <- fread("nyc14.csv")  
flights1

## 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

flights2 <- read\_csv("nyc14.csv") # produces a tibble

## Parsed with column specification:  
## cols(  
## year = col\_double(),  
## month = col\_double(),  
## day = col\_double(),  
## dep\_delay = col\_double(),  
## arr\_delay = col\_double(),  
## carrier = col\_character(),  
## origin = col\_character(),  
## dest = col\_character(),  
## air\_time = col\_double(),  
## distance = col\_double(),  
## hour = col\_double()  
## )

flights2

## # A tibble: 253,316 x 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>

class(flights2)

## [1] "spec\_tbl\_df" "tbl\_df" "tbl" "data.frame"

## Question 1

flights1[, .(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

## Question 2

flights1[origin == "JFK" & dest == "SEA" & carrier=="DL"]

## 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

## Question 3

flights1 %>% filter(air\_time< 330, month=="3", carrier=="UA")

## 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

## Question 4

flights1 %>%  
 select(carrier, month, air\_time)%>%  
 filter(carrier == "UA", month == 3, air\_time < 330)

## 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

## Question 5

## 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[,c("speed"):= .(distance / hour)]  
flights1

## 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 9 275.00000  
## 2: 2475 11 225.00000  
## 3: 2475 19 130.26316  
## 4: 1035 7 147.85714  
## 5: 2475 13 190.38462  
## ---   
## 253312: 1416 14 101.14286  
## 253313: 1400 8 175.00000  
## 253314: 431 11 39.18182  
## 253315: 502 11 45.63636  
## 253316: 659 8 82.37500

## Question 6

## 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(mph = (distance/hour))

## # A tibble: 253,316 x 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>, mph <dbl>

## Question 7

## 7) Show and use coding to change the carrier abbreviation of UA to UniitedAir  
  
flights1

## 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 9 275.00000  
## 2: 2475 11 225.00000  
## 3: 2475 19 130.26316  
## 4: 1035 7 147.85714  
## 5: 2475 13 190.38462  
## ---   
## 253312: 1416 14 101.14286  
## 253313: 1400 8 175.00000  
## 253314: 431 11 39.18182  
## 253315: 502 11 45.63636  
## 253316: 659 8 82.37500

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

## 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 UnitedAir LGA IAH 201  
## 253313: 2014 10 31 -5 -14 UnitedAir 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 9 275.00000  
## 2: 2475 11 225.00000  
## 3: 2475 19 130.26316  
## 4: 1035 7 147.85714  
## 5: 2475 13 190.38462  
## ---   
## 253312: 1416 14 101.14286  
## 253313: 1400 8 175.00000  
## 253314: 431 11 39.18182  
## 253315: 502 11 45.63636  
## 253316: 659 8 82.37500

## (7) Show and use coding to change the carrier abbreviation of UA to UniitedAir  
  
flights2 %>%  
 mutate(carrier = recode(carrier, "UA" = "UnitedAir")) ->  
 flights2  
flights2

## # A tibble: 253,316 x 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>

View(flights2)