

# Dottable Homework

Collin Brown and Jesse Brandt

2025-03-04

## Libraries

```
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  1.9.4      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 become errors
```

```
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
```

## Load data

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

```
glimpse(nycdata)
```

```
## Rows: 253,316
```

```
## Columns: 11
## $ year      <int> 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, ~
## $ month     <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ day       <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ dep_delay <int> 14, -3, 2, -8, 2, 4, -2, -3, -1, -2, -5, 7, 3, 142, -5, 18, ~
## $ arr_delay <int> 13, 13, 9, -26, 1, 0, -18, -14, -17, -14, -17, -5, 1, 133, --
## $ carrier   <chr> "AA", "AA", "AA", "AA", "AA", "AA", "AA", "AA", "AA", "AA", ~
## $ origin    <chr> "JFK", "JFK", "JFK", "LGA", "JFK", "EWR", "JFK", "JFK", "JFK~
## $ dest      <chr> "LAX", "LAX", "LAX", "PBI", "LAX", "LAX", "LAX", "LAX", "MIA~
## $ air_time  <int> 359, 363, 351, 157, 350, 339, 338, 356, 161, 349, 161, 365, ~
## $ distance  <int> 2475, 2475, 2475, 1035, 2475, 2454, 2475, 2475, 1089, 2422, ~
## $ hour      <int> 9, 11, 19, 7, 13, 18, 21, 15, 15, 18, 16, 17, 12, 19, 17, 17~
```

## Problem 1

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

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

```
##      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
##      ---
## 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
```

## Problem 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

```
nycdata[carrier == "DL" & origin == "JFK" & dest == "SEA"]
```

```
##      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
##      ---
## 1074: 2014    10    30         -3        -15      DL   JFK   SEA       339
## 1075: 2014    10    31         -6        -26      DL   JFK   SEA       317
## 1076: 2014    10    31         -1         -8      DL   JFK   SEA       338
```

```
## 1077: 2014 10 31 -1 -23 DL JFK SEA 326
## 1078: 2014 10 31 4 -27 DL JFK SEA 318
##      distance hour
##      <int> <int>
## 1: 2422 9
## 2: 2422 18
## 3: 2422 15
## 4: 2422 7
## 5: 2422 18
## ---
## 1074: 2422 18
## 1075: 2422 9
## 1076: 2422 6
## 1077: 2422 15
## 1078: 2422 18
```

### Problem 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.

```
nycdata[carrier == "UA" & month == 3 & air_time < 330]
```

```
##      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
## ---
## 3785: 2014 3 31 6 -8 UA EWR FLL 155
## 3786: 2014 3 31 7 -9 UA EWR PBI 135
## 3787: 2014 3 31 1 -21 UA EWR RSW 145
## 3788: 2014 3 31 0 -19 UA EWR IAH 196
## 3789: 2014 3 31 18 -7 UA EWR ORD 108
##      distance hour
##      <int> <int>
## 1: 1634 9
## 2: 1023 19
## 3: 1085 17
## 4: 1400 5
## 5: 1605 16
## ---
## 3785: 1065 16
## 3786: 1023 10
## 3787: 1068 14
## 3788: 1400 16
## 3789: 719 6
```

## Problem 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_4 <- nycdata |>
  filter(carrier == "UA", month == 3, air_time < 330)
nycdata_4
```

```
## Index: <dest__origin__carrier>
##      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
##    ---
## 3785: 2014     3    31         6         -8     UA   EWR   FLL      155
## 3786: 2014     3    31         7         -9     UA   EWR   PBI      135
## 3787: 2014     3    31         1        -21     UA   EWR   RSW      145
## 3788: 2014     3    31         0        -19     UA   EWR   IAH      196
## 3789: 2014     3    31        18         -7     UA   EWR   ORD      108
##      distance  hour
##      <int> <int>
##  1:    1634     9
##  2:    1023    19
##  3:    1085    17
##  4:    1400     5
##  5:    1605    16
##    ---
## 3785:    1065    16
## 3786:    1023    10
## 3787:    1068    14
## 3788:    1400    16
## 3789:     719     6
```

## Problem 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_5 <- nycdata
nycdata_5[, speed := (distance / air_time * 60)]
nycdata_5
```

```
## Index: <dest__origin__carrier>
##      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
```

```
##      ---
## 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
##      <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
##      ---
## 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
```

## Problem 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_6 <- nycdata |>
  mutate(speed = (distance / air_time) * 60)
nycdata_6
```

```
## Index: <dest__origin__carrier>
##      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
##      ---
## 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
##      <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
##      ---
## 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
```

## Problem 7a

Show and use coding to change the carrier abbreviation of UA to UnitedAir (sic), 7a. data.table method

```
nycdata_7a <- nycdata
nycdata_7a[carrier == "UA", carrier := "UnitedAir"]
unique(nycdata_7a$carrier) # list all carriers
```

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

```
nycdata_7a
```

```
##      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
##    ---
## 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
##      <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
##    ---
## 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
```

## Problem 7b

Show and use coding to change the carrier abbreviation of UA to UnitedAir (sic), 7b. tidyverse method (Use a sequence of dplyr commands so that you can see the change in your table)

```
nycdata_7b <- nycdata |>
  mutate(carrier = case_when(carrier == "UA" ~ "UnitedAir", TRUE ~ carrier))
unique(nycdata_7b$carrier) # list all carriers
```

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

nycdata\_7b

```
##      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
##    ---
## 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
##      <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
##    ---
## 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
```