dottable Assignment

AUTHOR
Jin Sook Song

Classwork Biggish Data

Question 1.

```
library(data.table)
dt <- fread("nycdata.csv")
dt_q1 <- dt[, .(year, month, day, hour)]
head(dt_q1)
```

```
    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
6:  2014     1     1    18
```

Question 2.

```
library(data.table)
dt <- fread("nycdata.csv")
dt_q2 <- dt[carrier == "DL" & origin == "JFK" & dest == "SEA"]
head(dt_q2)
```

```
    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
6:  2014     1     3       579       556      DL    JFK    SEA      327
   distance  hour
      <int> <int>
1:     2422     9
2:     2422    18
3:     2422    15
4:     2422     7
5:     2422    18
6:     2422     0
```

Question 3.

```
library(data.table)
dt <- fread("nycdata.csv")
dt_q3 <- dt[carrier == "UA" & month == 3 & air_time < 330]
head(dt_q3)
```

```
    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
```

Question 4.

```r
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.4     
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::between()     masks data.table::between()
✖ dplyr::filter()      masks stats::filter()
✖ dplyr::first()       masks data.table::first()
✖ lubridate::hour()    masks data.table::hour()
✖ lubridate::isoweek() masks data.table::isoweek()
✖ dplyr::lag()         masks stats::lag()
✖ dplyr::last()        masks data.table::last()
✖ lubridate::mday()    masks data.table::mday()
✖ lubridate::minute()  masks data.table::minute()
✖ lubridate::month()   masks data.table::month()
✖ lubridate::quarter() masks data.table::quarter()
✖ lubridate::second()  masks data.table::second()
✖ purrr::transpose()   masks data.table::transpose()
✖ lubridate::wday()    masks data.table::wday()
✖ lubridate::week()    masks data.table::week()
✖ lubridate::yday()    masks data.table::yday()
✖ lubridate::year()    masks data.table::year()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
df <- read_csv("nycdata.csv")
```

```
Rows: 253316 Columns: 11
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): carrier, origin, dest
dbl (8): year, month, day, dep_delay, arr_delay, air_time, distance, hour

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
df_q4 <- df %>%
  filter(carrier == "UA", month == 3, air_time < 330)
```

```r
head(df_q4)
```

```
# A tibble: 6 × 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
# ℹ 1 more variable: hour <dbl>
```

Question 5.

```r
library(data.table)
dt <- fread("nycdata.csv")
dt[, speed := (distance / air_time) * 60]
head(dt)
```

```
    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
```

Question 6.

```r
library(tidyverse)
df <- read_csv("nycdata.csv")
```

```
Rows: 253316 Columns: 11
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): carrier, origin, dest
dbl (8): year, month, day, dep_delay, arr_delay, air_time, distance, hour

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
df <- df %>%
  mutate(speed = (distance / air_time) * 60)
head(df)
```

```
# A tibble: 6 × 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
# ℹ 2 more variables: hour <dbl>, speed <dbl>
```

Question 7a.

```
        library(data.table)
        dt <- fread("nycdata.csv")
        unique(dt$carrier) # before change
```

```
 [1] "AA" "AS" "B6" "DL" "EV" "F9" "FL" "HA" "MQ" "VX" "WN" "UA" "US" "OO"
```

```
        dt[carrier == "UA", carrier := "UnitedAir"]
        unique(dt$carrier) # after change
```

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

Question 7b.

```
        library(tidyverse)
        library(dplyr)
        df <- read_csv("nycdata.csv")
```

```
Rows: 253316 Columns: 11
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): carrier, origin, dest
dbl (8): year, month, day, dep_delay, arr_delay, air_time, distance, hour

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
        df %>% pull(carrier) %>% unique() # before change
```

```
 [1] "AA" "AS" "B6" "DL" "EV" "F9" "FL" "HA" "MQ" "VX" "WN" "UA" "US" "OO"
```

```
        df <- df %>%
          mutate(carrier = ifelse(carrier == "UA", "UnitedAir", carrier))
        df %>% pull(carrier) %>% unique() # after change
```

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