Slicing_with_dplyr.R

dieudon neoued raogo

Sun Mar 10 20:39:05 2019

```
library(nycflights13)
library(tidyverse)
## -- Attaching packages ------ tidyverse 1.2.1 --
## v ggplot2 3.1.0
                      v purrr
                               0.2.5
## v tibble 2.0.1
                      v dplyr
                               0.7.8
## v tidyr
            0.8.2
                      v stringr 1.3.1
            1.3.1
                     v forcats 0.3.0
## v readr
## Warning: package 'tibble' was built under R version 3.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()
                   masks stats::lag()
flights
## # A tibble: 336,776 x 19
##
      year month
                   day dep_time sched_dep_time dep_delay arr_time
##
                                                  <dbl>
     <int> <int> <int>
                         <int>
                                        <int>
   1 2013
                                                     2
              1
                    1
                           517
                                          515
                                                            830
   2 2013
##
                           533
                                          529
                                                     4
                                                            850
               1
                     1
## 3 2013
               1
                    1
                           542
                                          540
                                                     2
                                                            923
## 4 2013
               1
                    1
                           544
                                          545
                                                    -1
                                                           1004
## 5 2013
                    1
                           554
                                          600
                                                    -6
                                                            812
               1
## 6 2013
                           554
                                          558
                                                    -4
                                                            740
               1
                    1
##
   7 2013
               1
                    1
                           555
                                          600
                                                    -5
                                                            913
  8 2013
                                                    -3
                                                            709
##
               1
                     1
                           557
                                          600
## 9 2013
                           557
                                          600
                                                    -3
                                                            838
               1
                     1
## 10 2013
               1
                     1
                           558
                                          600
                                                    -2
                                                            753
## # ... with 336,766 more rows, and 12 more variables: sched_arr_time <int>,
      arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
## #
      origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
      minute <dbl>, time hour <dttm>
####Filter is used for rows
day1<-filter(flights, month == 1, day == 1)</pre>
day1
## # A tibble: 842 x 19
                   day dep_time sched_dep_time dep_delay arr_time
##
      year month
##
     <int> <int> <int>
                                                  <dbl>
                         <int>
                                        <int>
                                                          <int>
##
   1 2013
               1
                     1
                           517
                                          515
                                                     2
                                                            830
##
   2 2013
               1
                     1
                           533
                                          529
                                                     4
                                                            850
   3 2013
                    1
                           542
                                          540
                                                     2
                                                            923
##
               1
## 4 2013
                    1
                           544
                                          545
                                                    -1
                                                           1004
               1
  5 2013
               1
                    1
                           554
                                          600
                                                    -6
                                                            812
##
  6 2013
               1
                     1
                           554
                                          558
                                                    -4
                                                            740
  7 2013
               1
                     1
                           555
                                          600
                                                    -5
                                                            913
```

```
## 8 2013
                              557
                                             600
                                                         -3
                                                                 709
                1
                      1
## 9 2013
                      1
                              557
                                             600
                                                         -3
                                                                 838
                1
## 10 2013
                              558
                      1
                                             600
                                                         -2
                                                                 753
## # ... with 832 more rows, and 12 more variables: sched_arr_time <int>,
       arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
       origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
       minute <dbl>, time hour <dttm>
(day25<-filter(flights, month == 1, day == 25))</pre>
## # A tibble: 922 x 19
##
       year month
                    day dep_time sched_dep_time dep_delay arr_time
##
      <int> <int> <int>
                            <int>
                                           <int>
                                                      <dbl>
   1 2013
                                                                 208
##
                     25
                               15
                                            1815
                                                        360
                1
   2 2013
##
                1
                     25
                               17
                                            2249
                                                        88
                                                                 119
## 3 2013
                     25
                               26
                                                        336
                                                                 225
                                            1850
                1
##
   4 2013
                1
                     25
                             123
                                            2000
                                                        323
                                                                 229
## 5 2013
                     25
                              123
                                            2029
                                                        294
                                                                 215
                1
##
   6 2013
                     25
                              456
                                             500
                                                                 632
                1
                                                         -4
##
  7 2013
                     25
                              519
                                             525
                                                         -6
                                                                 804
                1
   8 2013
                                                                 820
##
                1
                     25
                              527
                                             530
                                                         -3
## 9 2013
                                                                 826
                1
                     25
                              535
                                             540
                                                         -5
## 10 2013
                1
                     25
                              539
                                             540
                                                         -1
                                                                1006
## # ... with 912 more rows, and 12 more variables: sched_arr_time <int>,
       arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
       origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #
       minute <dbl>, time hour <dttm>
filter(flights, month == 11 | month == 12)
## # A tibble: 55,403 x 19
                    day dep_time sched_dep_time dep_delay arr_time
##
       year month
##
      <int> <int> <int>
                            <int>
                                           <int>
                                                      <dbl>
                                                               <int>
  1 2013
##
                               5
                                            2359
                                                                 352
               11
                      1
                                                          6
## 2 2013
               11
                      1
                               35
                                            2250
                                                        105
                                                                 123
## 3 2013
                              455
                                             500
                                                         -5
                                                                 641
               11
                      1
## 4 2013
               11
                      1
                              539
                                             545
                                                         -6
                                                                 856
## 5 2013
                             542
                                                         -3
                                                                 831
               11
                      1
                                             545
  6 2013
##
               11
                      1
                             549
                                             600
                                                        -11
                                                                 912
   7 2013
##
               11
                      1
                             550
                                             600
                                                        -10
                                                                 705
##
  8 2013
               11
                      1
                              554
                                             600
                                                         -6
                                                                 659
## 9 2013
               11
                              554
                                             600
                                                         -6
                                                                 826
## 10 2013
                                                         -6
                              554
                                             600
                                                                 749
               11
                      1
## # ... with 55,393 more rows, and 12 more variables: sched arr time <int>,
       arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
       origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
       minute <dbl>, time_hour <dttm>
nov_dec <- filter(flights, month %in% c(11, 12))</pre>
nov_dec
## # A tibble: 55,403 x 19
                    day dep_time sched_dep_time dep_delay arr_time
       year month
##
      <int> <int> <int>
                           <int>
                                           <int>
                                                      <dbl>
                                                               <int>
  1 2013
               11
                      1
                                5
                                            2359
                                                          6
                                                                 352
## 2 2013
                               35
                                            2250
                                                                 123
               11
                      1
                                                        105
```

```
500
##
    3 2013
                11
                       1
                              455
                                                          -5
                                                                   641
##
    4 2013
                       1
                              539
                                              545
                                                          -6
                                                                   856
                11
##
    5 2013
                11
                       1
                              542
                                              545
                                                          -3
                                                                   831
   6 2013
                              549
##
                                              600
                                                                   912
                11
                       1
                                                         -11
##
    7
       2013
                11
                       1
                              550
                                              600
                                                         -10
                                                                   705
##
   8 2013
                                              600
                                                                   659
                11
                       1
                              554
                                                          -6
    9 2013
                               554
                                              600
                                                          -6
                                                                   826
##
                11
                       1
## 10 2013
                              554
                                                          -6
                                                                   749
                11
                       1
                                              600
## # ... with 55,393 more rows, and 12 more variables: sched_arr_time <int>,
       arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
       origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #
       minute <dbl>, time_hour <dttm>
filter(flights, !(arr_delay > 120 | dep_delay > 120))
## # A tibble: 316,050 x 19
##
       vear month
                     day dep_time sched_dep_time dep_delay arr_time
##
      <int> <int> <int>
                            <int>
                                            <int>
                                                       <dbl>
                                                                 <int>
    1 2013
                               517
                                              515
                                                           2
                                                                   830
##
                 1
                       1
##
    2 2013
                                              529
                                                           4
                                                                   850
                 1
                       1
                               533
   3 2013
                                                           2
##
                       1
                              542
                                              540
                                                                   923
                 1
    4 2013
##
                 1
                       1
                              544
                                              545
                                                          -1
                                                                  1004
##
   5 2013
                 1
                       1
                              554
                                              600
                                                          -6
                                                                   812
##
   6 2013
                       1
                              554
                                              558
                                                          -4
                                                                   740
                 1
   7 2013
##
                 1
                       1
                              555
                                              600
                                                          -5
                                                                   913
   8 2013
                              557
                                              600
                                                                   709
##
                 1
                       1
                                                          -3
##
   9 2013
                       1
                              557
                                              600
                                                          -3
                                                                   838
                 1
## 10 2013
                       1
                              558
                                              600
                                                          -2
                                                                   753
## # ... with 316,040 more rows, and 12 more variables: sched_arr_time <int>,
       arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
## #
       origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
       minute <dbl>, time_hour <dttm>
filter(flights, arr_delay <= 120, dep_delay <= 120)</pre>
## # A tibble: 316,050 x 19
##
       year month
                     day dep_time sched_dep_time dep_delay arr_time
##
                                                       <dbl>
      <int> <int> <int>
                            <int>
                                            <int>
                                                                 <int>
##
   1 2013
                 1
                       1
                              517
                                              515
                                                           2
                                                                   830
    2 2013
                              533
                                              529
                                                           4
                                                                   850
##
                 1
                       1
##
   3 2013
                       1
                              542
                                              540
                                                           2
                                                                   923
                 1
##
   4 2013
                 1
                       1
                              544
                                              545
                                                          -1
                                                                  1004
   5 2013
##
                              554
                                              600
                                                          -6
                                                                   812
                 1
                       1
    6 2013
##
                 1
                       1
                              554
                                              558
                                                          -4
                                                                   740
##
   7 2013
                 1
                       1
                              555
                                              600
                                                          -5
                                                                   913
##
   8 2013
                 1
                       1
                              557
                                              600
                                                          -3
                                                                   709
##
   9 2013
                              557
                                              600
                                                          -3
                                                                   838
                 1
                       1
## 10 2013
                       1
                              558
                                              600
                                                          -2
                                                                   753
                 1
## # ... with 316,040 more rows, and 12 more variables: sched arr time <int>,
       arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
       origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #
       minute <dbl>, time_hour <dttm>
###############################
                             Dealing with NA
                                                ##################
df \leftarrow tibble(x = c(1, NA, 3))
filter(df, x > 1)
```

```
## # A tibble: 1 x 1
##
        x
##
     <dbl>
## 1
         3
filter(df, is.na(x) | x > 1)
## # A tibble: 2 x 1
##
        х
##
     <dbl>
## 1
       NA
## 2
arrange(flights, year, month, day)
## # A tibble: 336,776 x 19
                    day dep_time sched_dep_time dep_delay arr_time
##
       year month
##
      <int> <int> <int>
                                                    <dbl>
                           <int>
                                          <int>
   1 2013
                             517
                                                        2
                                                               830
##
                1
                      1
                                            515
##
   2 2013
                      1
                             533
                                            529
                                                        4
                                                               850
                1
   3 2013
                             542
                                            540
##
                1
                      1
                                                        2
                                                               923
##
   4 2013
                             544
                                            545
                                                              1004
                1
                      1
                                                       -1
   5 2013
##
                      1
                             554
                                            600
                                                       -6
                                                               812
                1
   6 2013
##
                1
                      1
                             554
                                            558
                                                       -4
                                                               740
##
   7 2013
                1
                      1
                             555
                                            600
                                                       -5
                                                               913
##
   8 2013
                1
                      1
                             557
                                            600
                                                       -3
                                                               709
##
  9 2013
                      1
                             557
                                            600
                                                       -3
                                                               838
                1
                                                       -2
## 10 2013
                1
                      1
                             558
                                            600
                                                               753
## # ... with 336,766 more rows, and 12 more variables: sched_arr_time <int>,
      arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
      origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #
## #
      minute <dbl>, time_hour <dttm>
####In descending order
arrange(flights, desc(arr_delay))
## # A tibble: 336,776 x 19
##
      year month
                    day dep_time sched_dep_time dep_delay arr_time
##
      <int> <int> <int>
                           <int>
                                          <int>
                                                    <dbl>
                                                             <int>
##
   1 2013
                             641
                                            900
                                                     1301
                                                              1242
                1
                      9
   2 2013
                     15
##
                6
                            1432
                                           1935
                                                     1137
                                                              1607
   3 2013
##
                     10
                            1121
                                           1635
                                                     1126
                                                              1239
                1
##
   4 2013
                9
                     20
                            1139
                                           1845
                                                     1014
                                                              1457
##
   5 2013
                7
                     22
                             845
                                           1600
                                                     1005
                                                              1044
##
   6 2013
                4
                     10
                            1100
                                           1900
                                                      960
                                                              1342
   7 2013
##
                     17
                            2321
                                                      911
                                                               135
                3
                                            810
                7
##
   8 2013
                     22
                            2257
                                            759
                                                      898
                                                               121
##
   9 2013
                                                      896
                                                              1058
               12
                      5
                             756
                                           1700
## 10 2013
                5
                      3
                            1133
                                           2055
                                                      878
                                                              1250
## # ... with 336,766 more rows, and 12 more variables: sched_arr_time <int>,
## #
      arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
      origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #
      minute <dbl>, time_hour <dttm>
```

```
####NA will be at the end
df \leftarrow tibble(x = c(5, 2, NA))
arrange(df, x)
## # A tibble: 3 x 1
## x
## <dbl>
## 1 2
## 2
      5
## 3 NA
arrange(df, desc(x))
## # A tibble: 3 x 1
## x
## <dbl>
## 1 5
## 2
      2
## 3 NA
```