Code ▾

Assignment 2 - Carrier Delay Visualization

Gordon Wall (gwall2)

workspace set-up

import relevant data

Hide

```
carrier.df <- fread("On_Time_Marketing_Carrier_On_Time_Performance_(Beginning_January_2018)_2018_1.csv")
```

```
|--------------------------------------------------|
|==================================================|
```

examine data

Hide

```
skim_without_charts(carrier.df)
```

```
-- Data Summary ------------------------
                           Values
Name                       carrier.df
Number of rows             621461
Number of columns          120
_______________________
Column type frequency:
  character                25
  logical                  25
  numeric                  70
________________________
Group variables            None

-- Variable type: character --------------------------------------------------------------------
-----------------------------
# A tibble: 25 x 8
   skim_variable                                       n_missing complete_rate   min   max  empt
y n_unique whitespace
 * <chr>                                                   <int>         <dbl> <int> <int>  <int
>    <int>      <int>
 1 "FlightDate"                                                0             1    10    10
0       31          0
 2 "Marketing_Airline_Network"                                 0             1     2     2
0       11          0
 3 "Operated_or_Branded_Code_Share_Partners"                   0             1     2    12
0       16          0
 4 "IATA_Code_Marketing_Airline"                               0             1     2     2
0       11          0
 5 "Originally_Scheduled_Code_Share_Airline"                   0             1     0     2 62130
1        9          0
 6 "IATA_Code_Originally_Scheduled_Code_Share_Airline"         0             1     0     2 62130
1        9          0
 7 "Operating_Airline "                                        0             1     2     2
0       28          0
 8 "IATA_Code_Operating_Airline"                               0             1     2     2
0       28          0
 9 "Tail_Number"                                               0             1     0     6   275
9     5637          0
10 "Origin"                                                    0             1     3     3
0      351          0
11 "OriginCityName"                                            0             1     8    34
0      345          0
12 "OriginState"                                               0             1     2     2
0       52          0
13 "OriginStateName"                                           0             1     4    46
0       52          0
14 "Dest"                                                      0             1     3     3
0      351          0
15 "DestCityName"                                              0             1     8    34
0      345          0
16 "DestState"                                                 0             1     2     2
0       52          0
17 "DestStateName"                                             0             1     4    46
```

```
0       52          0
18 "DepTimeBlk"                                                0             1     9     9
0       19          0
19 "ArrTimeBlk"                                                0             1     9     9
0       19          0
20 "CancellationCode"                                          0             1     0     1 60248
5        5          0
21 "Div1Airport"                                               0             1     0     3 61991
0      205          0
22 "Div1TailNum"                                               0             1     0     6 62042
3      926          0
23 "Div2Airport"                                               0             1     0     3 62142
8       28          0
24 "Div2TailNum"                                               0             1     0     6 62145
1       11          0
25 "Duplicate"                                                 0             1     1     1
0        1          0

-- Variable type: logical -----------------------------------------------------------------------
-----------------------------
# A tibble: 25 x 5
   skim_variable    n_missing complete_rate  mean count
 * <chr>                <int>         <dbl> <dbl> <chr>
 1 Div3Airport         621461             0   NaN ": "
 2 Div3AirportID       621461             0   NaN ": "
 3 Div3AirportSeqID    621461             0   NaN ": "
 4 Div3WheelsOn        621461             0   NaN ": "
 5 Div3TotalGTime      621461             0   NaN ": "
 6 Div3LongestGTime    621461             0   NaN ": "
 7 Div3WheelsOff       621461             0   NaN ": "
 8 Div3TailNum         621461             0   NaN ": "
 9 Div4Airport         621461             0   NaN ": "
10 Div4AirportID       621461             0   NaN ": "
11 Div4AirportSeqID    621461             0   NaN ": "
12 Div4WheelsOn        621461             0   NaN ": "
13 Div4TotalGTime      621461             0   NaN ": "
14 Div4LongestGTime    621461             0   NaN ": "
15 Div4WheelsOff       621461             0   NaN ": "
16 Div4TailNum         621461             0   NaN ": "
17 Div5Airport         621461             0   NaN ": "
18 Div5AirportID       621461             0   NaN ": "
19 Div5AirportSeqID    621461             0   NaN ": "
20 Div5WheelsOn        621461             0   NaN ": "
21 Div5TotalGTime      621461             0   NaN ": "
22 Div5LongestGTime    621461             0   NaN ": "
23 Div5WheelsOff       621461             0   NaN ": "
24 Div5TailNum         621461             0   NaN ": "
25 V120                621461             0   NaN ": "

-- Variable type: numeric -----------------------------------------------------------------------
-----------------------------
# A tibble: 70 x 10
   skim_variable                                      n_missing complete_rate          mean
sd      p0      p25
```

```
 * <chr>                                                  <int>         <dbl>         <dbl>
<dbl>   <dbl>    <dbl>
 1 Year                                                       0     1            2018
0         2018    2018
 2 Quarter                                                    0     1               1
0            1       1
 3 Month                                                      0     1               1
0            1       1
 4 DayofMonth                                                 0     1              15.9
8.98         1       8
 5 DayOfWeek                                                  0     1               3.74
1.99         1       2
 6 DOT_ID_Marketing_Airline                                   0     1           19833.          2
93.       19393   19790
 7 Flight_Number_Marketing_Airline                            0     1            2725.         19
13.           1    1044
 8 DOT_ID_Originally_Scheduled_Code_Share_Airline        621301     0.000257    20373.
99.6      20046   20366
 9 Flight_Num_Originally_Scheduled_Code_Share_Airline    621301     0.000257     5628.          9
65.        2836    5558.
10 DOT_ID_Operating_Airline                                   0     1           20024.          4
11.       19393   19790
11 Flight_Number_Operating_Airline                            0     1            2725.         19
13.           1    1044
12 OriginAirportID                                            0     1           12683.         15
17.       10135   11292
13 OriginAirportSeqID                                         0     1         1268323.       1517
07.     1013505 1129202
14 OriginCityMarketID                                         0     1           31769.         13
07.       30070   30721
15 OriginStateFips                                            0     1              27.1
16.5          1      12
16 OriginWac                                                  0     1              54.2
26.4          1      34
17 DestAirportID                                              0     1           12683.         15
17.       10135   11292
18 DestAirportSeqID                                           0     1         1268301.       1517
00.     1013505 1129202
19 DestCityMarketID                                           0     1           31769.         13
07.       30070   30721
20 DestStateFips                                              0     1              27.1
16.5          1      12
21 DestWac                                                    0     1              54.2
26.4          1      34
22 CRSDepTime                                                 0     1            1327.          4
84.           1     916
23 DepTime                                                18571     0.970        1333.          4
93.           1     924
24 DepDelay                                               19062     0.969           9.70
47.5      -1280      -6
25 DepDelayMinutes                                        19062     0.969          13.3
46.2          0       0
26 DepDel15                                               19062     0.969           0.180
0.384        0       0
```

```
27 DepartureDelayGroups                                   19062     0.969           0.0141
2.21        -2      -1
28 TaxiOut                                                19708     0.968          17.9
10.7          1      11
29 WheelsOff                                              19699     0.968        1360.         4
93.           1     941
30 WheelsOn                                               20233     0.967        1477.         5
15.           1    1059
31 TaxiIn                                                 20242     0.967           7.49
5.91         0       4
32 CRSArrTime                                                 0     1            1493.         5
07.           1    1109
33 ArrTime                                                19381     0.969        1482.         5
19.           1    1103
34 ArrDelay                                               20604     0.967           3.17
49.6      -1290     -17
35 ArrDelayMinutes                                        20604     0.967          13.1
45.7          0       0
36 ArrDel15                                               20604     0.967           0.179
0.384        0       0
37 ArrivalDelayGroups                                     20604     0.967          -0.302
2.35        -2      -2
38 Cancelled                                                  0     1               0.0305
0.172        0       0
39 Diverted                                                   0     1               0.00226
0.0475       0       0
40 CRSElapsedTime                                             0     1             139.
73.9        -90      87
41 ActualElapsedTime                                      20402     0.967         133.
71.6      -1228      82
42 AirTime                                                21263     0.966         108.
69.9      -1244      57
43 Flights                                                    0     1               1
0            1       1
44 Distance                                                   0     1             761.         5
83.          16     337
45 DistanceGroup                                              0     1               3.52
2.30         1       2
46 CarrierDelay                                          513669     0.173          21.8
66.7          0       0
47 WeatherDelay                                          513669     0.173           4.70
36.1          0       0
48 NASDelay                                              513669     0.173          14.0
31.8          0       0
49 SecurityDelay                                         513669     0.173           0.0955
3.27         0       0
50 LateAircraftDelay                                     513669     0.173          26.9
53.2          0       0
51 FirstDepTime                                          617762     0.00595      1256.         5
00.           1     818.
52 TotalAddGTime                                         617762     0.00595        35.3
29.7          1      17
53 LongestAddGTime                                       617762     0.00595        34.6
28.1          1      16
```

```
54 DivAirportLandings                                         0     1               0.00439
0.145        0       0
55 DivReachedDest                                        620054     0.00226         0.719
0.450        0       0
56 DivActualElapsedTime                                  620449     0.00163       393.         2
37.          90     251.
57 DivArrDelay                                           620449     0.00163       259.         2
47.           0     128
58 DivDistance                                           620054     0.00226        70.0        2
00.           0       0
59 Div1AirportID                                         619910     0.00250     12777.        15
48.        10135   11298
60 Div1AirportSeqID                                      619910     0.00250   1277727.      1548
08.     1013505 1129806
61 Div1WheelsOn                                          619914     0.00249      1397.         5
45.           1    1000
62 Div1TotalGTime                                        619914     0.00249        29.4
30.7          2       9
63 Div1LongestGTime                                      619914     0.00249        23.8
26.1          2       8
64 Div1WheelsOff                                         620426     0.00167      1400.         5
27.           1    1056
65 Div2AirportID                                         621428     0.0000531   12174.        14
99.        10397   10990
66 Div2AirportSeqID                                      621428     0.0000531 1217380.      1499
48.     1039707 1099005
67 Div2WheelsOn                                          621428     0.0000531    1241.         6
78.          34     954
68 Div2TotalGTime                                        621428     0.0000531      20.1
20.8          4       7
69 Div2LongestGTime                                      621428     0.0000531      17.2
16.6          4       6
70 Div2WheelsOff                                         621451     0.0000161    1252.         3
66.         803     975.
        p50      p75    p100
 *    <dbl>    <dbl>   <dbl>
 1     2018     2018    2018
 2        1        1       1
 3        1        1       1
 4       16       24      31
 5        4        5       7
 6    19805    19977   21171
 7     2237     4444    9366
 8    20378    20378   21167
 9     6068     6209    6344
10    19977    20378   21171
11     2237     4443    9375
12    12889    14057   16218
13 1288903  1405702  1621801
14    31453    32575   36133
15       26       42      78
16       44       81      93
17    12889    14057   16218
18 1288903  1405702  1621801
```

```
19   31453    32575    36133
20      26       42       78
21      44       81       93
22    1320     1730     2359
23    1329     1737     2400
24      -3        6     2007
25       0        6     2007
26       0        0        1
27      -1        0       12
28      15       21     1394
29    1343     1753     2400
30    1511     1909     2400
31       6        9      258
32    1520     1915     2359
33    1515     1914     2400
34      -8        6     2023
35       0        6     2023
36       0        0        1
37      -1        0       12
38       0        0        1
39       0        0        1
40     120      170     1645
41     115      164      728
42      89      138      683
43       1        1        1
44     599     1005     4983
45       3        5       11
46       0       19     2007
47       0        0     1682
48       2       19     1346
49       0        0      593
50       2       32     1648
51    1211     1652     2400
52      28       44      353
53      27       43      232
54       0        0        9
55       1        1        1
56     319      423     1514
57     176.     268.    2524
58       0       55     2556
59   12891    14107    15919
60 1289102  1410702  1591904
61    1350     1808.    2400
62      20       37      308
63      14       28      194
64    1345     1818.    2358
65   11577    13930    14869
66 1157706  1393006  1486903
67    1429     1704     2243
68      10       27       96
69      10       20       75
70    1225     1442.    1950
```

A lot of columns have nearly all entries missing

these columns provide nothing for analysis and can
be removed
I will filter for variables that have a complete_rate
of less than 1% and create a column names index from this
which I will then use to remove them from the data frame

separate drop-worthy columns

Hide

```
drops <- carrier.df %>% skim() %>% dplyr::filter(complete_rate < 0.01)

drop.index <- drops[,"skim_variable"]

t <- as.vector(drop.index$skim_variable)
```

drop by drop.index

Hide

```
carrier.clean <- select(carrier.df, -t)
```

the dataframe is much cleaner now, free of noisey variables
for further example, things like year and quarter can be dropped for this
analysis since every observation is from january, 2018 (1st quarter)
and all entries will be the same in these columns
other variables will be dropped with this same logic

separate drop-worthy columns (second time thru)

Hide

```
### dropping columns with more than 600,000 empty data entries

drops2 <- carrier.clean %>% skim() %>% dplyr::filter(character.empty > 600000)

drop.index2 <- drops2[,"skim_variable"]

q <- as.vector(drop.index2$skim_variable)
```

drop by second drop.index

Hide

```
carrier.clean <- select(carrier.clean, -q)
```

Hide

```
### dropping year, month, quarter, day of month columns

carrier.clean <- carrier.clean[,-c("Year", "Quarter", "Month", "DayofMonth")]
```

there are variables described in the readme.html

file stating that some columns have codes which could’ve

been used for multiple different carriers

these destroy the integriy of unique IDs and will be dropped now

Hide

```
### dropping columns that start with IATA
### the non-unique ones

carrier.clean <- carrier.clean %>% select(-starts_with("IATA"))
```

Hide

```
### dropping other irrelevant/redundant columns

carrier.clean <- carrier.clean[,!c("OriginCityName", "OriginStateName")]

carrier.clean <- carrier.clean[,!c("Duplicate")]

carrier.clean <- carrier.clean[,!c("DestStateName", "DestCityName")]

carrier.clean <- carrier.clean[,!c("Marketing_Airline_Network", "Operated_or_Branded_Code_Share_
Partners", "Tail_Number")]
```

convert to proper variable types

Diverted <dbl>
0
1

2 rows

Final Check of Clean Dataset

Hide

```
skim_without_charts(carrier.clean)
```

```
-- Data Summary ------------------------
                           Values
Name                       carrier.clean
Number of rows             621461
Number of columns          53
_______________________
Column type frequency:
  Date                     1
  factor                   11
  numeric                  41
________________________
Group variables            None

-- Variable type: Date -------------------------------------------------------------------------
-----------------------------
# A tibble: 1 x 7
  skim_variable n_missing complete_rate min        max        median     n_unique
* <chr>             <int>         <dbl> <date>     <date>     <date>        <int>
1 FlightDate            0             1 2018-01-01 2018-01-31 2018-01-16       31

-- Variable type: factor -----------------------------------------------------------------------
-----------------------------
# A tibble: 11 x 6
   skim_variable     n_missing complete_rate ordered n_unique top_counts
 * <chr>                 <int>         <dbl> <lgl>      <int> <chr>
 1 DayOfWeek                 0             1 FALSE          7 1: 103449, 3: 102953, 2: 101470,
5: 84898
 2 operating.airline         0             1 FALSE         28 WN: 109676, AA: 73598, DL: 71254,
OO: 62181
 3 Origin                    0             1 FALSE        351 ATL: 30729, ORD: 29921, DFW: 2243
4, DEN: 20485
 4 OriginState               0             1 FALSE         52 CA: 69059, TX: 61074, FL: 48715, I
L: 38592
 5 Dest                      0             1 FALSE        351 ATL: 30731, ORD: 29905, DFW: 2244
2, DEN: 20477
 6 DestState                 0             1 FALSE         52 CA: 69082, TX: 61080, FL: 48659, I
L: 38581
 7 DepTimeBlk                0             1 FALSE         19 060: 44327, 170: 42357, 070: 4179
2, 080: 41608
 8 ArrTimeBlk                0             1 FALSE         19 160: 41012, 140: 39644, 210: 3935
9, 180: 38767
 9 Cancelled                 0             1 FALSE          2 0: 602485, 1: 18976
10 Diverted                  0             1 FALSE          2 0: 620054, 1: 1407
11 DistanceGroup             0             1 FALSE         11 2: 160553, 3: 120757, 1: 95568, 4:
88897

-- Variable type: numeric ----------------------------------------------------------------------
-----------------------------
# A tibble: 41 x 10
   skim_variable                     n_missing complete_rate          mean          sd      p0
p25     p50     p75    p100
 * <chr>                                 <int>         <dbl>         <dbl>      <dbl>    <dbl>   <d
bl>   <dbl>   <dbl>   <dbl>
```

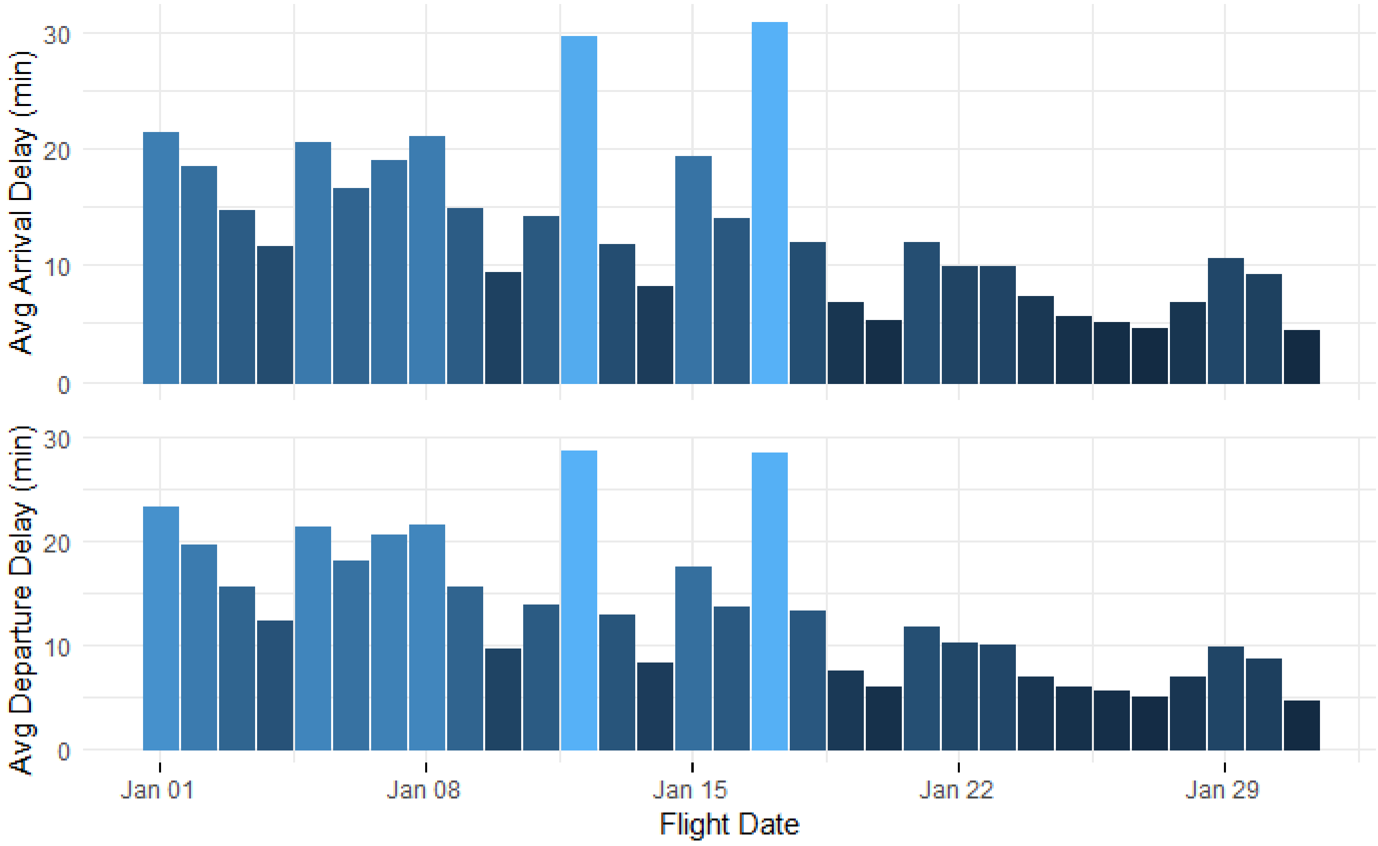
```
 1 DOT_ID_Marketing_Airline                0         1       19833.        293.      19393   19
790   19805   19977   21171
 2 Flight_Number_Marketing_Airline         0         1        2725.       1913.          1    1
044    2237    4444    9366
 3 DOT_ID_Operating_Airline                0         1       20024.        411.      19393   19
790   19977   20378   21171
 4 Flight_Number_Operating_Airline         0         1        2725.       1913.          1    1
044    2237    4443    9375
 5 OriginAirportID                         0         1       12683.       1517.      10135   11
292   12889   14057   16218
 6 OriginAirportSeqID                      0         1     1268323.     151707.    1013505 1129
202 1288903 1405702 1621801
 7 OriginCityMarketID                      0         1       31769.       1307.      30070   30
721   31453   32575   36133
 8 OriginStateFips                         0         1          27.1        16.5         1
12       26      42      78
 9 OriginWac                               0         1          54.2        26.4         1
34       44      81      93
10 DestAirportID                           0         1       12683.       1517.      10135   11
292   12889   14057   16218
11 DestAirportSeqID                        0         1     1268301.     151700.    1013505 1129
202 1288903 1405702 1621801
12 DestCityMarketID                        0         1       31769.       1307.      30070   30
721   31453   32575   36133
13 DestStateFips                           0         1          27.1        16.5         1
12       26      42      78
14 DestWac                                 0         1          54.2        26.4         1
34       44      81      93
15 CRSDepTime                              0         1        1327.        484.          1
916    1320    1730    2359
16 DepTime                             18571         0.970    1333.        493.          1
924    1329    1737    2400
17 DepDelay                            19062         0.969       9.70       47.5     -1280
-6      -3       6    2007
18 DepDelayMinutes                     19062         0.969      13.3        46.2         0
0       0       6    2007
19 DepDel15                            19062         0.969       0.180       0.384       0
0       0       0       1
20 DepartureDelayGroups                19062         0.969       0.0141      2.21       -2
-1      -1       0      12
21 TaxiOut                             19708         0.968      17.9        10.7         1
11      15      21    1394
22 WheelsOff                           19699         0.968    1360.        493.          1
941    1343    1753    2400
23 WheelsOn                            20233         0.967    1477.        515.          1    1
059    1511    1909    2400
24 TaxiIn                              20242         0.967       7.49        5.91        0
4       6       9     258
25 CRSArrTime                              0         1        1493.        507.          1    1
109    1520    1915    2359
26 ArrTime                             19381         0.969    1482.        519.          1    1
103    1515    1914    2400
27 ArrDelay                            20604         0.967       3.17       49.6     -1290
-17      -8       6    2023
```

```
28 ArrDelayMinutes                   20604         0.967      13.1          45.7         0
0       0       6    2023
29 ArrDel15                          20604         0.967       0.179         0.384       0
0       0       0       1
30 ArrivalDelayGroups                20604         0.967      -0.302         2.35       -2
-2      -1       0      12
31 CRSElapsedTime                        0         1         139.          73.9       -90
87     120     170    1645
32 ActualElapsedTime                 20402         0.967     133.          71.6     -1228
82     115     164     728
33 AirTime                           21263         0.966     108.          69.9     -1244
57      89     138     683
34 Flights                               0         1           1           0           1
1       1       1       1
35 Distance                              0         1         761.         583.         16
337     599    1005    4983
36 CarrierDelay                     513669         0.173      21.8          66.7         0
0       0      19    2007
37 WeatherDelay                     513669         0.173       4.70         36.1         0
0       0       0    1682
38 NASDelay                         513669         0.173      14.0          31.8         0
0       2      19    1346
39 SecurityDelay                    513669         0.173       0.0955        3.27        0
0       0       0     593
40 LateAircraftDelay                513669         0.173      26.9          53.2         0
0       2      32    1648
41 DivAirportLandings                    0         1           0.00439       0.145       0
0       0       0       9
```

Data Visualization

Question 1 & 2

What is the pattern of arrival traffic and departure traffic delays with respect to days and weeks?

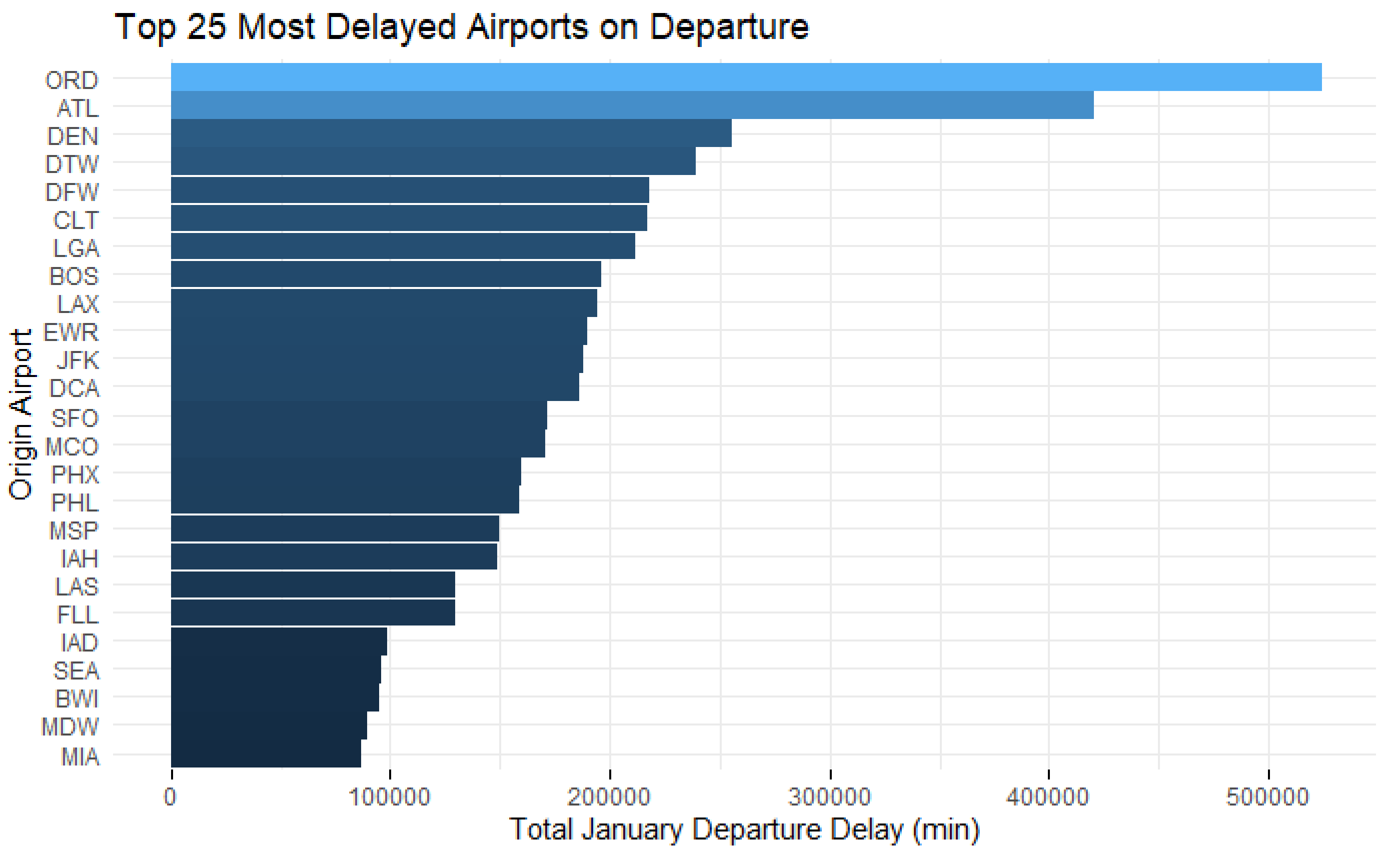


Can you interpret the traffic delays?

The graph is of Avg Arrival/Departure Delays across the date range of January 1st thru 31st. The bars outline the avg delay on each day and the x axis ticks section off every 7 days, with each tick being the start of a new seven-day cycle (week). Interpreting the traffic delays in this format reveal that the month of January, on average, decreases steadily in traffic over the duration of the month, with the middle of the month seeing a few peak high traffic delay spikes in the second and third weeks. Further interpretation could reveal that these spikes have something to do with winter storms in the heart of cold January…

Question 3

Which Airport ('Origin Airport') has highest departure delay?



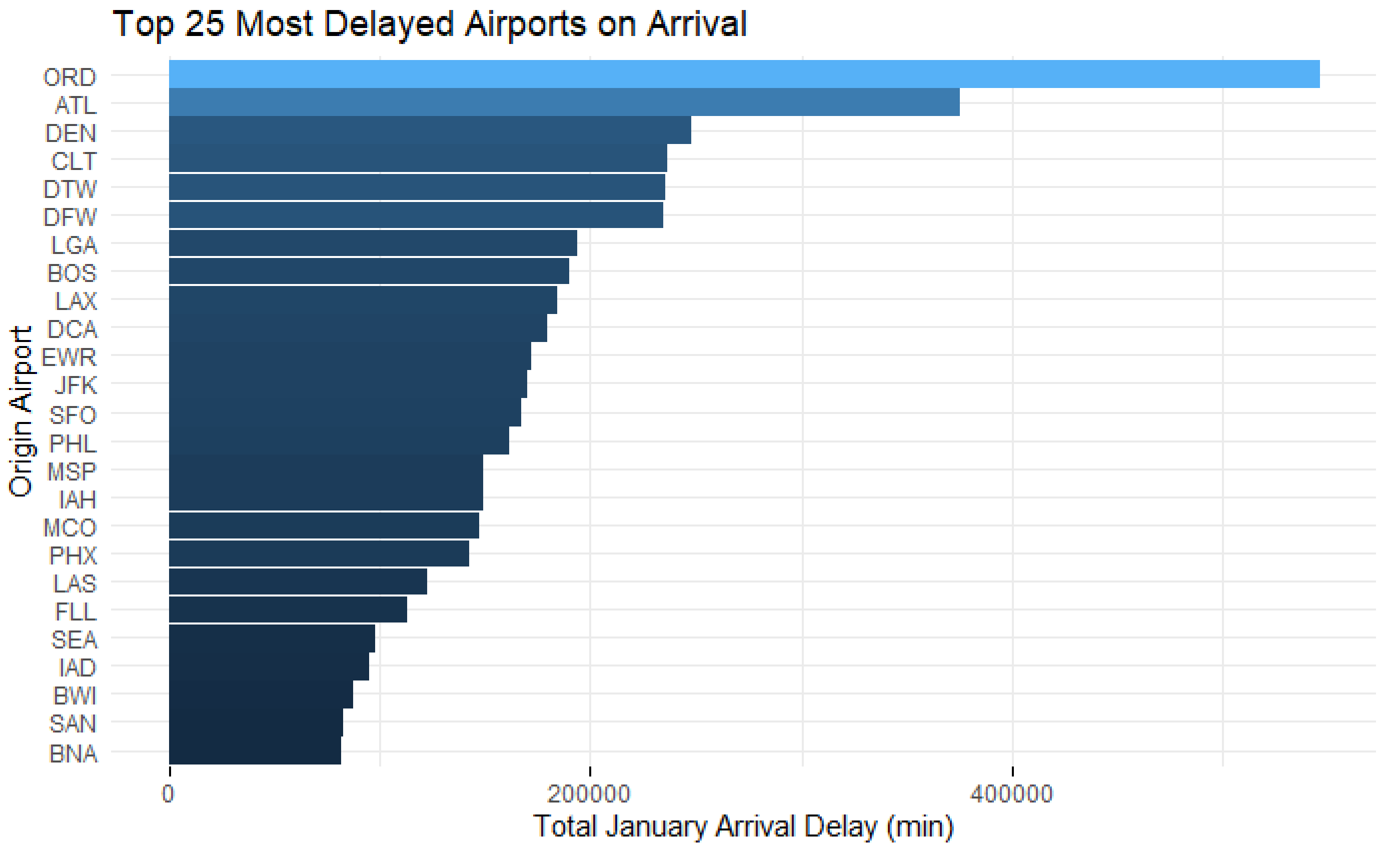
Chicago has the highest departure delay.

Question 4

Which Airport has highest arrival delay?

Hide

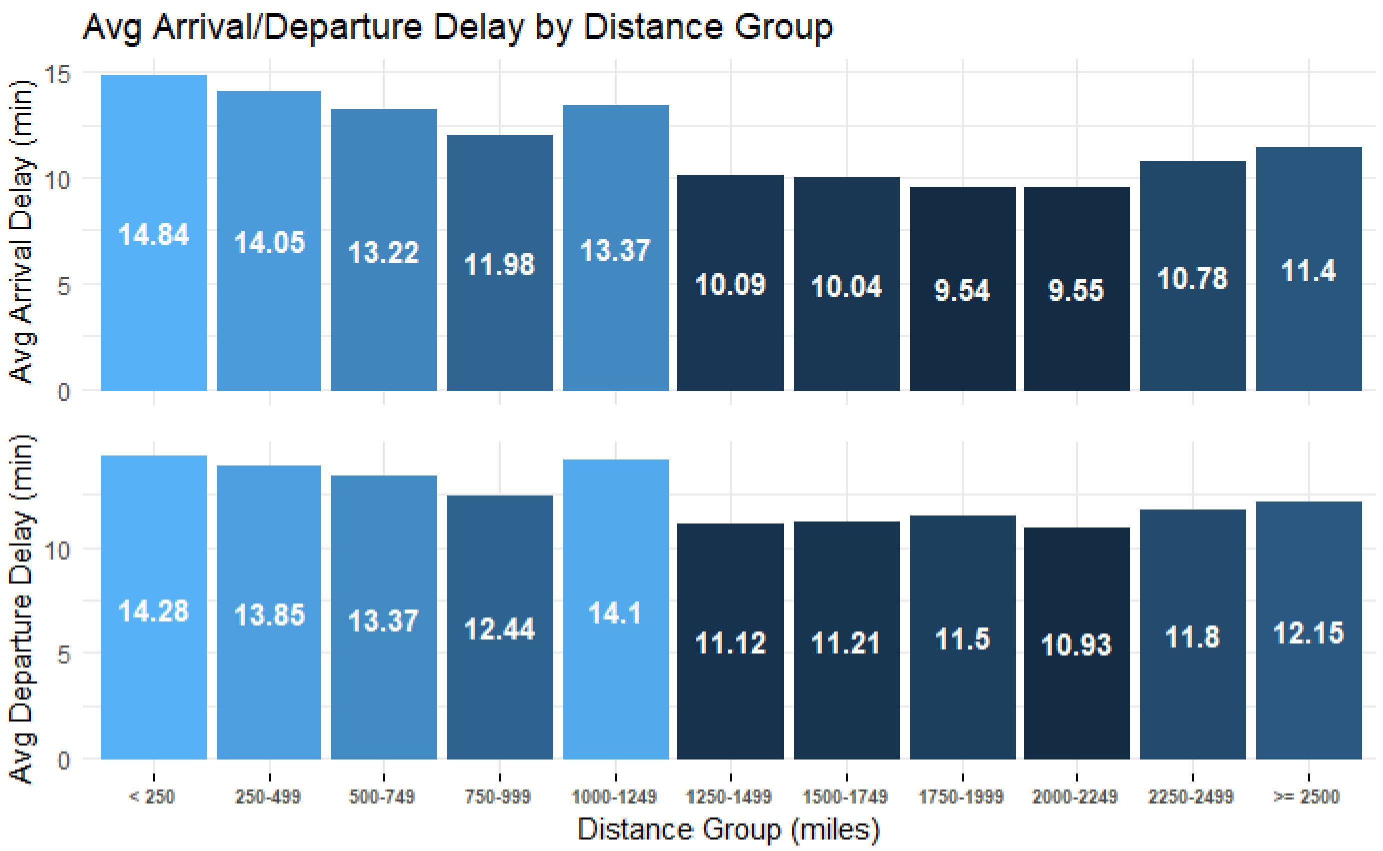
```
subset4 <- carrier.clean %>%
  select(Origin, ArrDelayMinutes) %>%
  group_by(Origin) %>%
  summarise(sum.airportdelay = sum(ArrDelayMinutes, na.rm = TRUE)) %>%
  arrange(desc(sum.airportdelay)) %>%
  slice(1:25)

plot4 <- ggplot(subset4, aes(reorder(Origin, sum.airportdelay), sum.airportdelay)) +
  geom_bar(stat = "identity", aes(col = sum.airportdelay, fill = sum.airportdelay)) +
  coord_flip() +
  ylab("Total January Arrival Delay (min)") +
  xlab("Origin Airport") +
  ggtitle("Top 25 Most Delayed Airports on Arrival") +
  theme_minimal() +
  theme(axis.ticks.x = element_line(color = "black"), legend.position = "none")


plot4
```



Chicago also has the highest arrival delay.

Question 5

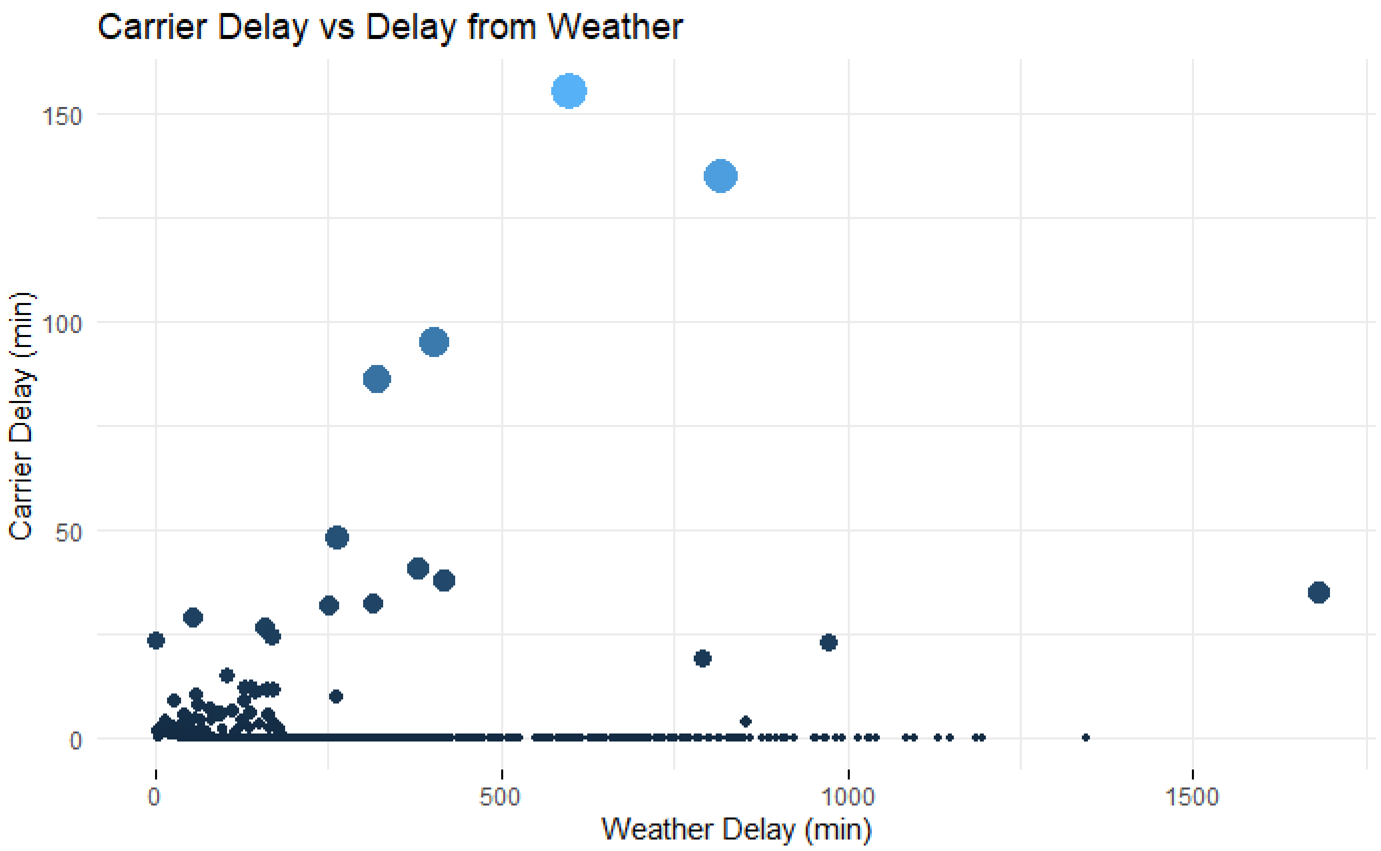
How do you relate the delay pattern to the distance travelled?



It appears that distance travelled has a relatively uncorrelated effect on delay, as shown above. I originally speculated that delay would increase with distance, but this was a naive assumption at best. Turns out the highest average delays come from the shortest three-five mileage categories. Maybe this has something to do with the fact that longer flights have the ability to make up for lost delay time by reaching and maintaing cruising altitude at a faster speed…

Question 6

Is there any correlation between weather delay and carrier delay?



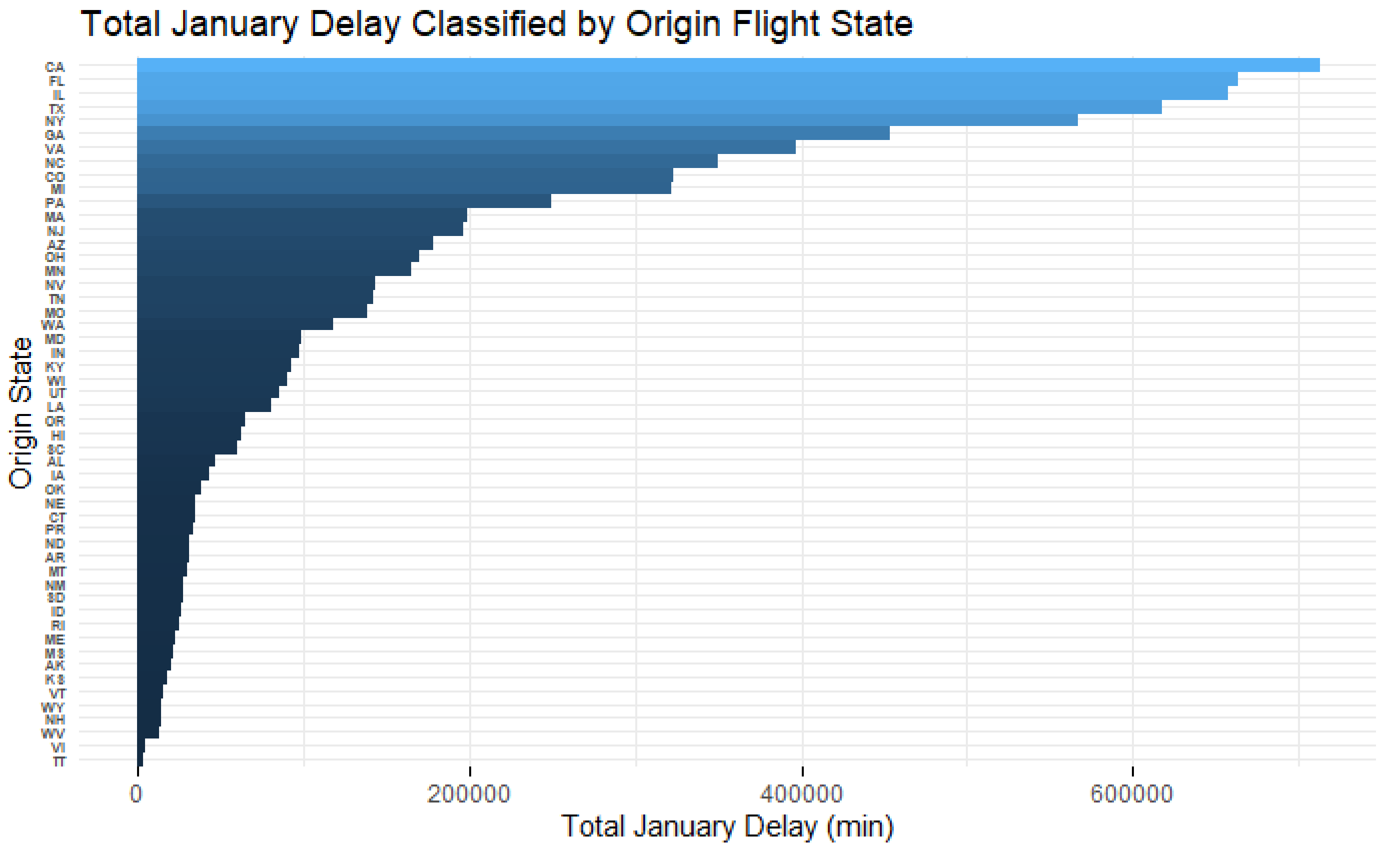
From the above visual, it seems that although some weather delays have impacted carrier delay, the vast majority of data shows that carrier delay has been virtually zero even in the presence of increasing weather delay. You can observe some spikes in carrier delay, which seem random and uncorrelated with weather delay. This dataset would have to be cross-referenced with storm data in the same date-range to determine if recorded heavy storms correlated with the spikes in carrier delay…

Question 7

What is the delay pattern you can find in respective states?

Hide

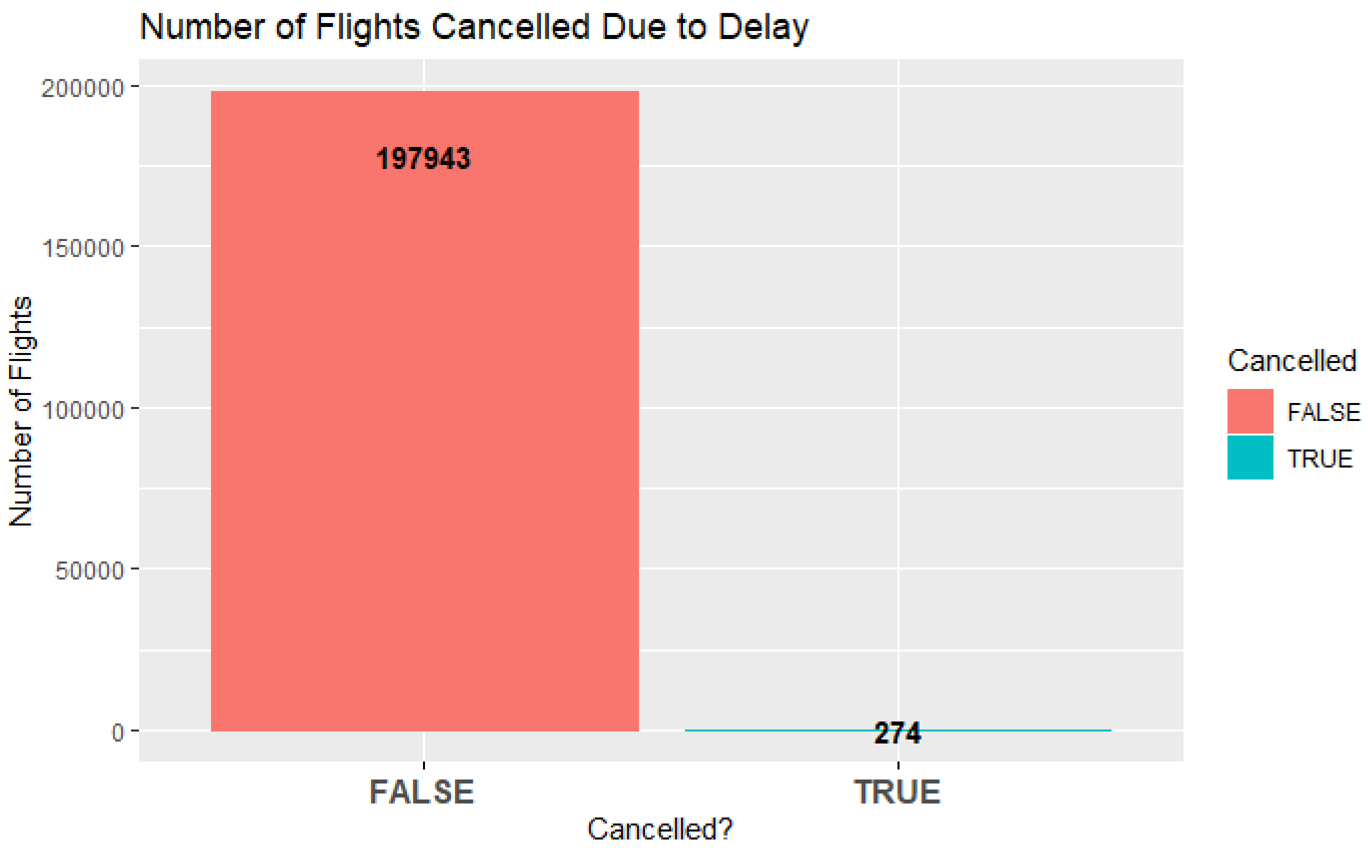
```
subset7 <- carrier.clean %>%
  select(OriginState, DepDelayMinutes) %>%
  group_by(OriginState) %>%
  summarise(sum.statedelay = sum(DepDelayMinutes, na.rm = TRUE)) %>%
  arrange(desc(sum.statedelay))

plot7 <- ggplot(subset7, aes(reorder(OriginState, sum.statedelay), sum.statedelay)) +
  geom_bar(stat = "identity", aes(col = sum.statedelay, fill = sum.statedelay)) +
  coord_flip() +
  ylab("Total January Delay (min)") +
  xlab("Origin State") +
  ggtitle("Total January Delay Classified by Origin Flight State") +
  theme_minimal() +
  theme(axis.ticks.x = element_line(color = "black"), legend.position = "none", axis.text.y = el
ement_text(size = 5, face = "bold"))


plot7
```



It seems from the chart above that delay is typically higher in more trafficked and popular transportation states. The largest delays come from states like California (highest), Illinois (third) and New York (fifth), whereas the lowest delays are sported by more remote states like Arkansas.

Question 8

How many delayed flights were cancelled? (approximation)



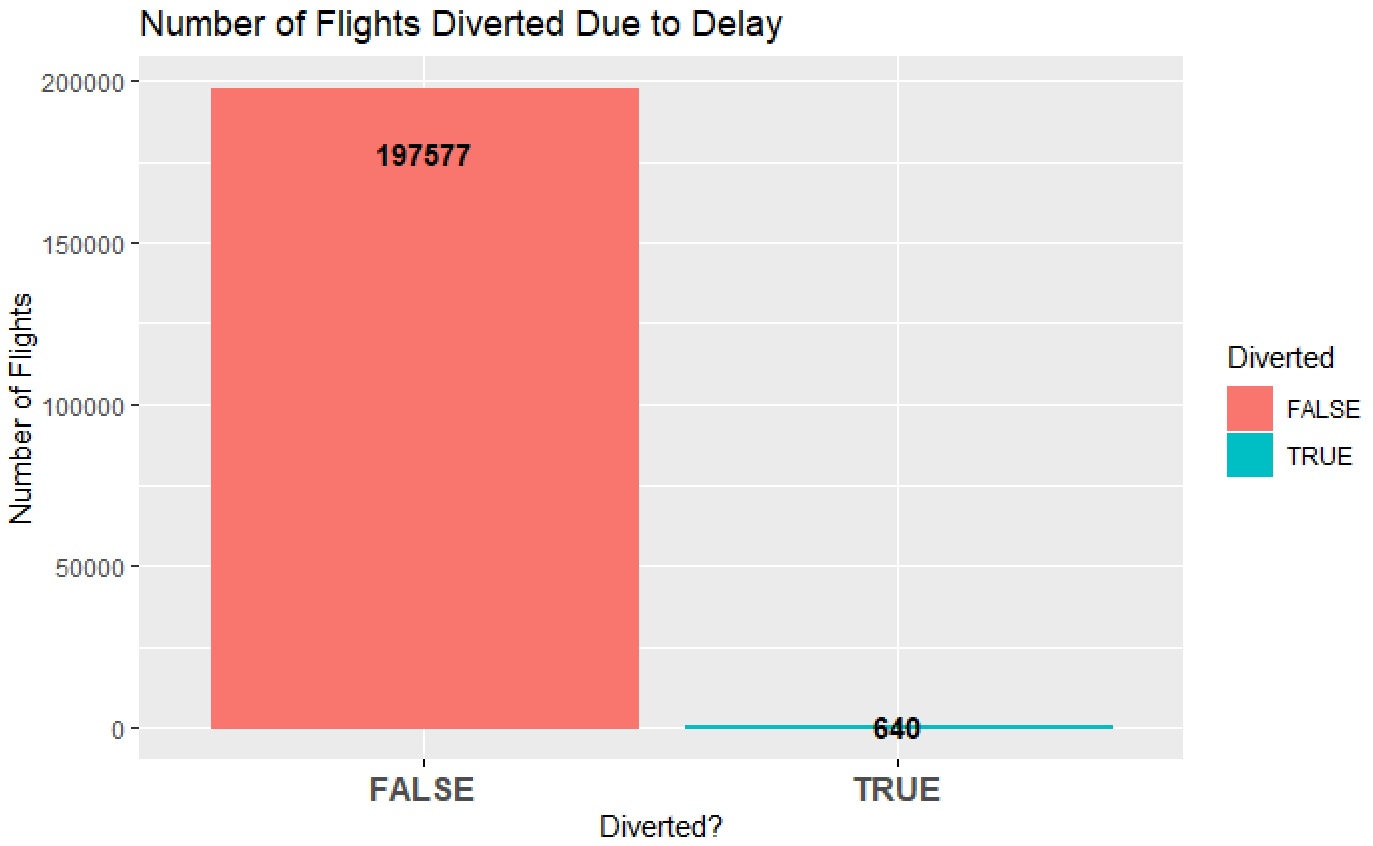
As shown by the chart above of all delayed flights, only 274 flights were actually cancelled because of the delay. Likely due to the fact that the economy of carrier shipping would rather take a small loss from delay than cancel all together. The people of the USA need their products after all!!

Question 9

How many delayed flights were diverted? (approximation)

Hide

```
subset9 <- carrier.clean %>%
  mutate(Delayed = if_else(DepDelayMinutes>0, TRUE, FALSE), Diverted = if_else(Diverted == 1, TRUE, FALSE)) %>%
  select(Delayed, Diverted) %>%
  na.omit()

subset9 <- subset9 %>%
  group_by(Delayed, Diverted) %>%
  summarise(total = n()) %>%
  mutate(prop = total/sum(total)) %>%
  filter(Delayed == TRUE)

plot9 <- ggplot(subset9, aes(Diverted, total)) +
  geom_bar(stat = "identity", position = "dodge", aes(col = Diverted, fill = Diverted)) +
  geom_text(label = subset9$total, col = "black", fontface = "bold", position = position_stack(vjust = 0.9)) +
  ylab("Number of Flights") +
  xlab("Diverted?") +
  ggtitle("Number of Flights Diverted Due to Delay") +
  theme(axis.ticks.x = element_line(color = "black"), legend.position = "right", axis.text.x = element_text(size = 12, face = "bold"))


plot9
```



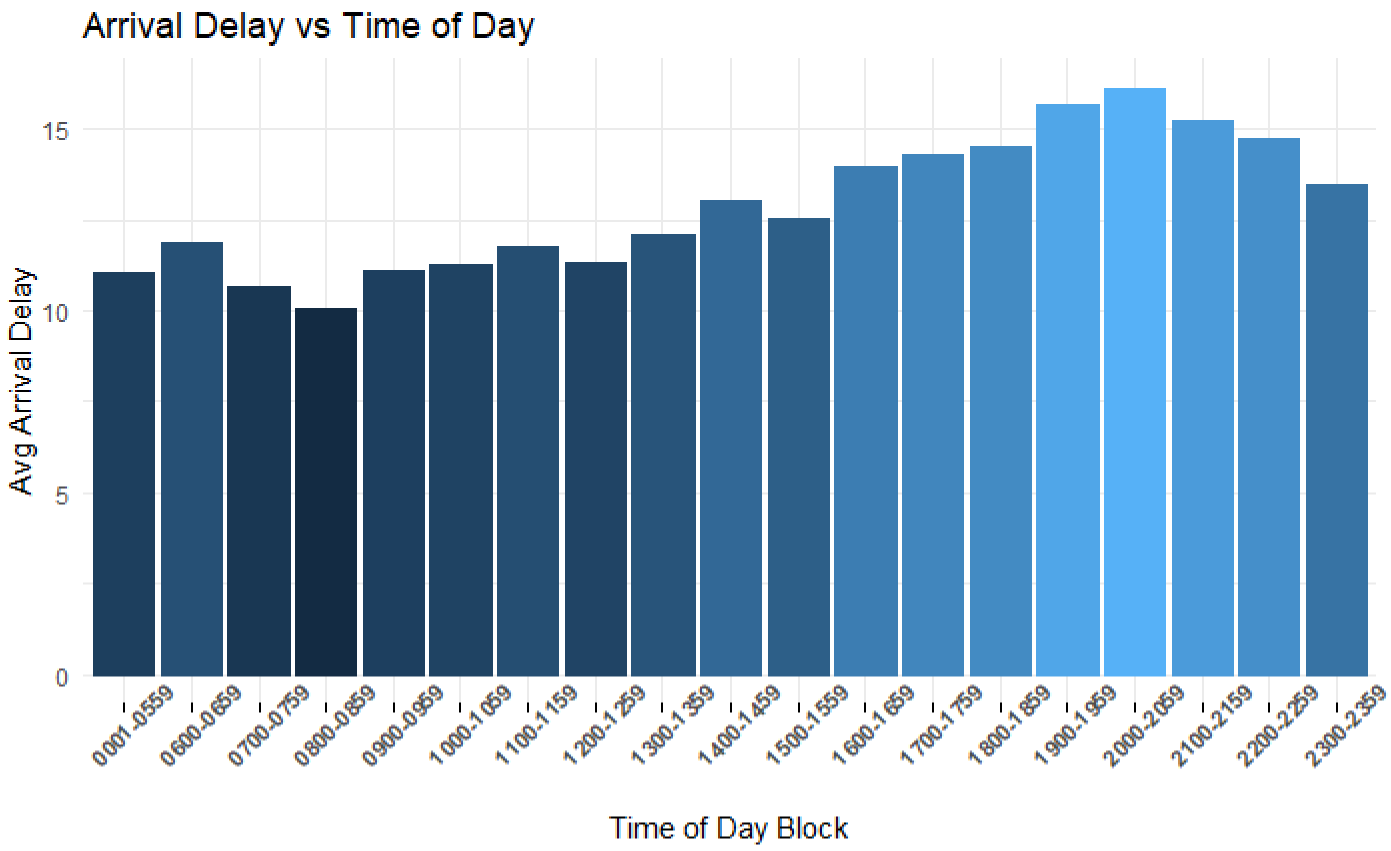
Similar findings to question 8. Only 640 flights were diverted in the face of delay

Question 10

What time of the day do you find arrival delays?

Hide

```
subset10 <- carrier.clean %>%
  select(ArrTimeBlk, ArrDelayMinutes) %>%
  group_by(ArrTimeBlk) %>%
  summarise(mean.arrdelay = round(mean(ArrDelayMinutes, na.rm = TRUE), 4)) %>%
  arrange(ArrTimeBlk)

plot10 <- ggplot(subset10, aes(ArrTimeBlk, mean.arrdelay)) +
  geom_bar(stat="identity", aes(col = mean.arrdelay, fill = mean.arrdelay)) +
  ylab("Avg Arrival Delay") +
  xlab("Time of Day Block") +
  ggtitle("Arrival Delay vs Time of Day") +
  theme_minimal() +
  theme(axis.ticks.x = element_line(color = "black"), legend.position = "none", axis.text.x = el
ement_text(size = 8, face = "bold", angle = 45))

plot10
```



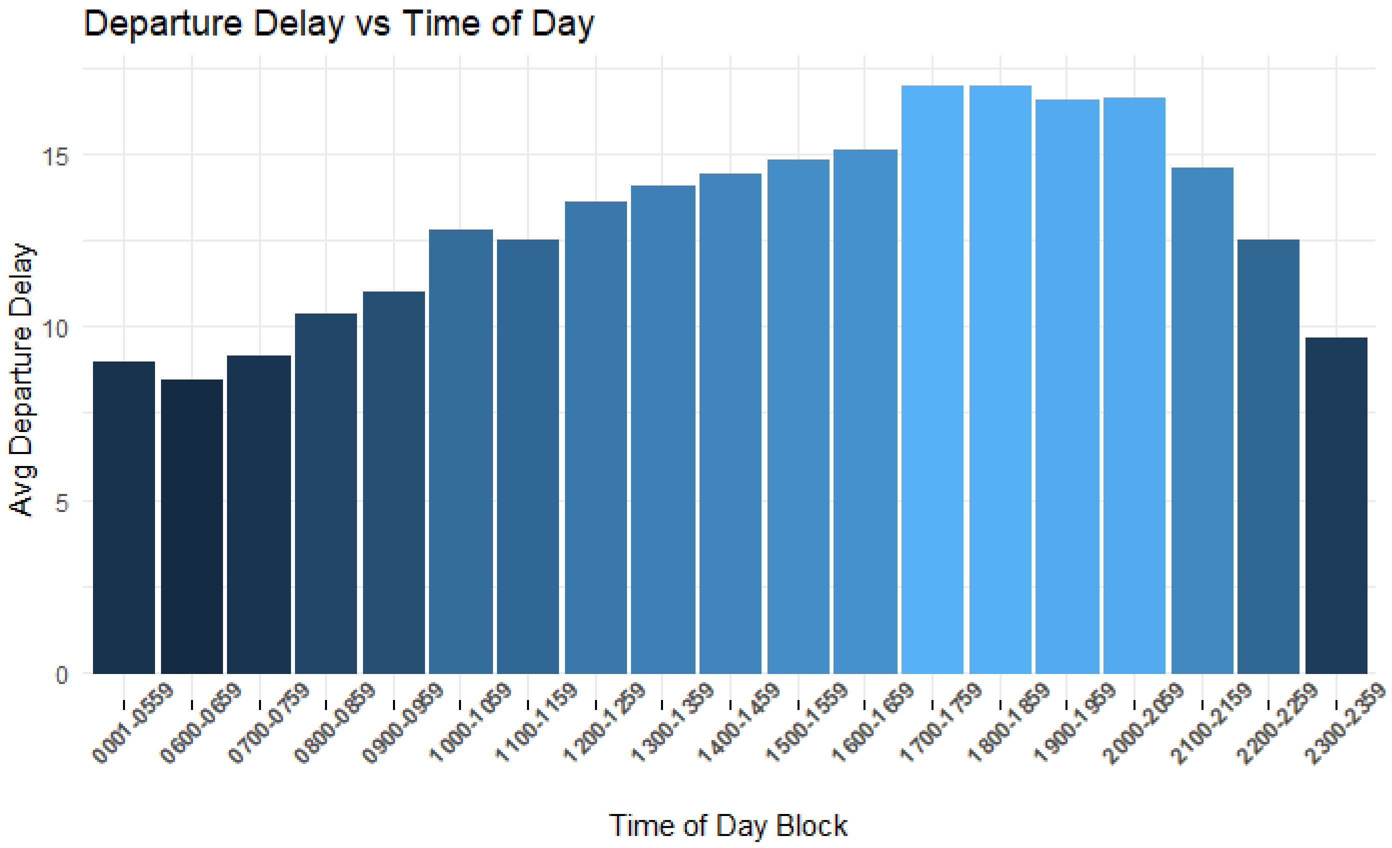
It appears that arrival delay peaks around the end of the 24-hour cycle; within the 8-10 PM range

Question 11

What time of the day do you find departure delays?

Hide

```
subset11 <- carrier.clean %>%
  select(DepTimeBlk, DepDelayMinutes) %>%
  group_by(DepTimeBlk) %>%
  summarise(mean.depdelay = round(mean(DepDelayMinutes, na.rm = TRUE), 4)) %>%
  arrange(DepTimeBlk)

plot11 <- ggplot(subset11, aes(DepTimeBlk, mean.depdelay)) +
  geom_bar(stat="identity", aes(col = mean.depdelay, fill = mean.depdelay)) +
  ylab("Avg Departure Delay") +
  xlab("Time of Day Block") +
  ggtitle("Departure Delay vs Time of Day") +
  theme_minimal() +
  theme(axis.ticks.x = element_line(color = "black"), legend.position = "none", axis.text.x = el
ement_text(size = 8, face = "bold", angle = 45))

plot11
```



Similar findings to question 10.

It appears that arrival delay peaks and flattens off

within the 7-10 PM range.