MATH 216 Homework 1

Shannia Fu

Contents

Admistrative:

Please indicate

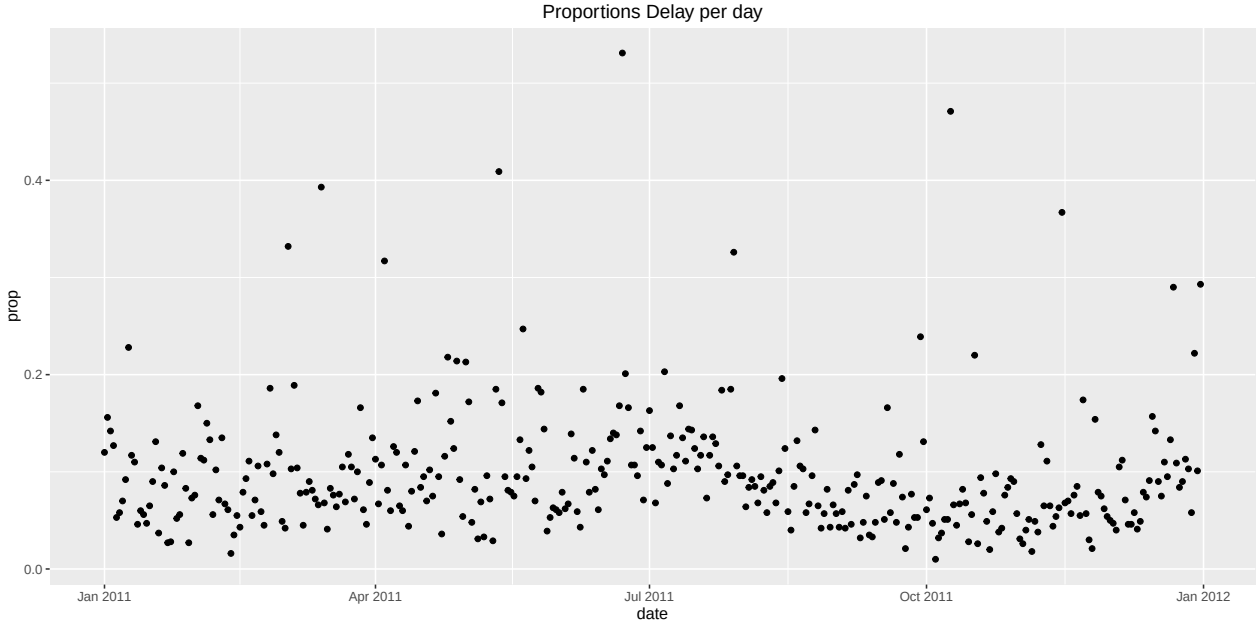
- Who you collaborated with: Brenda Li
- Roughly how much time you spent on this HW so far: 2.5 hours
- The URL of the RPubs published URL here.
- What gave you the most trouble: #3
- Any comments you have:

Question 1:

Plot a "time series" of the proportion of flights that were delayed by > 30 minutes on each day. i.e.

- the x-axis should be some notion of time
- the y-axis should be the proportion.

Using this plot, indicate describe the seasonality of when delays over 30 minutes tend to occur.



From the graph, it seems that delays typically occur more often in July-August (summer), and in December-January (winter). There's also a lower portion of delays just around October (fall).

Question 2:

Some people prefer flying on older planes. Even though they aren't as nice, they tend to have more room. Which airlines should these people favor?

```
## # A tibble: 10 × 2
##    carrier     n
##      <chr> <int>
## 1       WN 14957
## 2       CO  3942
## 3       US  3033
## 4       DL  1907
## 5       AA  1379
## 6       UA   785
## 7       OO   119
## 8       MQ    57
## 9       FL    16
## 10      YV     1
```

```
## # A tibble: 10 × 2
##    carrier           p
##      <chr>       <dbl>
## 1       MQ 1.000000000
## 2       AA 0.987822350
## 3       US 0.791906005
## 4       DL 0.729812476
## 5       UA 0.389965226
## 6       WN 0.334668397
## 7       CO 0.058138163
## 8       YV 0.012658228
## 9       FL 0.008171604
## 10      OO 0.007558915
```

I categorized "old planes" as planes that are from before 1995. Based on that assumption, the airline with the highest absolute number of flights on older planes (from Houston) is by far WN (Southwest), followed (not closely) by CO (Continental) and US (US Air). However, proportionally, WN comes in 6th with only 33.4% of its flights on older planes. Proportionally, then, the top 4 airlines to fly with are MQ (Envoy), AA (American Airlines), US (US Air), and DL (Delta), bearing in mind that MQ, for example, only has an absolute number of 57 flights on older planes.

Question 3:

- What states did Southwest Airlines' **flight paths** tend to fly to?
- What states did Southwest Airlines' **flights** tend to fly to?

For example, Southwest Airlines Flight 60 to Dallas consists of a single flight path, but since it flew 299 times in 2013, it would be counted as 299 flights.

```
## # A tibble: 19 × 2
##    state     n
##    <chr> <int>
## 1     TX   770
## 2     FL   297
## 3     LA   236
## 4     CA   219
```

```
## 5     OK   178
## 6     IL   163
## 7     NV   123
## 8     CO   121
## 9     TN   107
## 10    AZ   106
## 11    MO   101
## 12    NM    80
## 13    MD    79
## 14    MS    78
## 15    AL    52
## 16    SC    43
## 17    PA    35
## 18    AR    28
## 19    NJ    26
```

```
## # A tibble: 19 × 2
##    state     n
##    <chr> <int>
## 1     TX 17230
## 2     FL  3992
## 3     LA  3362
## 4     CA  2792
## 5     OK  2350
## 6     IL  2094
## 7     NV  1543
## 8     CO  1438
## 9     TN  1377
## 10    AZ  1369
## 11    MO  1334
## 12    MD  1227
## 13    NM  1019
## 14    MS  1016
## 15    AL   691
## 16    SC   588
## 17    PA   437
## 18    NJ   390
## 19    AR   365
```

Both flight paths and flights tended to fly to Texas, followed by Florida and Louisiana. In general, the frequency of flight paths and total flights follow the same order in frequency of destination.

Question 4:

I want to know proportionately what regions (NE, south, west, midwest) each carrier flies to/from Houston in the month of July. Consider the `month()` function from the `lubridate` package.

```
## Source: local data frame [34 x 5]
## Groups: carrier [15]
##
##    carrier  region total_in_region total_in_carrier      prop
##      <chr>   <chr>           <int>            <int>     <dbl>
## 1       AA   south             273              273 1.0000000
## 2       AS    west              31               31 1.0000000
## 3       B6      NE              62               62 1.0000000
```

```
## 4       CO midwest            713            6092 0.1170387
## 5       CO      NE           1060            6092 0.1739987
## 6       CO   south           2073            6092 0.3402823
## 7       CO    west           2246            6092 0.3686802
## 8       DL midwest             47             226 0.2079646
## 9       DL   south            179             226 0.7920354
## 10      EV midwest             95             164 0.5792683
## # ... with 24 more rows
```