CT248 Assignment 6

Exploring Flights Data

Consider the following data set (Flights.csv), and generate the following analysis. When processing the day, make sure to remove records with missing values.

1. Read in the file, and confirm the number of records (336,766)

```
>> flights

flights =

  336776×9 table
```

YearDay	Month	Day	origin	dest	hour	minute	dep_delay	distance
1	1	1	{'EWR'}	{'IAH'}	5	15	2	1400
1	1	1	{'LGA'}	{'IAH'}	5	29	4	1416
1	1	1	{'JFK'}	{'MIA'}	5	40	2	1089
1	1	1	{'JFK'}	{'BQN'}	5	45	-1	1576
1	1	1	{'LGA'}	{'ATL'}	6	0	-6	762
:	:	:	:	:	:	:	:	:
273	9	30	{'JFK'}	{'DCA'}	14	55	NaN	213
273	9	30	{'LGA'}	{'SYR'}	22	0	NaN	198
273	9	30	{'LGA'}	{'BNA'}	12	10	NaN	764
273	9	30	{'LGA'}	{'CLE'}	11	59	NaN	419
273	9	30	{'LGA'}	{'RDU'}	8	40	NaN	431

2. Convert “origin” and “dest” to strings (from cell type)

```
>> flights

flights =

  336776×9 table
```

YearDay	Month	Day	origin	dest	hour	minute	dep_delay	distance
1	1	1	"EWR"	"IAH"	5	15	2	1400
1	1	1	"LGA"	"IAH"	5	29	4	1416
1	1	1	"JFK"	"MIA"	5	40	2	1089
1	1	1	"JFK"	"BQN"	5	45	-1	1576
1	1	1	"LGA"	"ATL"	6	0	-6	762
:	:	:	:	:	:	:	:	:
273	9	30	"JFK"	"DCA"	14	55	NaN	213
273	9	30	"LGA"	"SYR"	22	0	NaN	198
273	9	30	"LGA"	"BNA"	12	10	NaN	764
273	9	30	"LGA"	"CLE"	11	59	NaN	419
273	9	30	"LGA"	"RDU"	8	40	NaN	431

3. Check the number of missing values for the departure time.

```
dep_delay: 336776×1 double

    Values:

        Min             -43
        Median          -2
        Max             1301
        NumMissing      8255
```

4. Filter all the missing values from the departure delay. Use the MATLAB function **isnan()** to help with this. Check the difference in the number of records.

```
>> flights_clean

flights_clean =

  328521×9 table

    YearDay    Month    Day    origin    dest     hour    minute    dep_delay    distance
    _______    _____    ___    ______    _____    ____    ______    _________    ________

       1         1       1     "EWR"     "IAH"     5       15           2          1400
       1         1       1     "LGA"     "IAH"     5       29           4          1416
       1         1       1     "JFK"     "MIA"     5       40           2          1089
       1         1       1     "JFK"     "BQN"     5       45          -1          1576
       1         1       1     "LGA"     "ATL"     6        0          -6           762

      :         :       :       :         :       :        :           :            :

     273        9      30     "JFK"     "SYR"    22       45          -5           209
     273        9      30     "JFK"     "BUF"    22       50         -10           301
     273        9      30     "JFK"     "ROC"    22       46          -5           264
     273        9      30     "JFK"     "BOS"    22       55          12           187
     273        9      30     "JFK"     "PSE"    23       59         -10          1617

    Display all 328521 rows.
```

5. Confirm the difference in records between the two tables

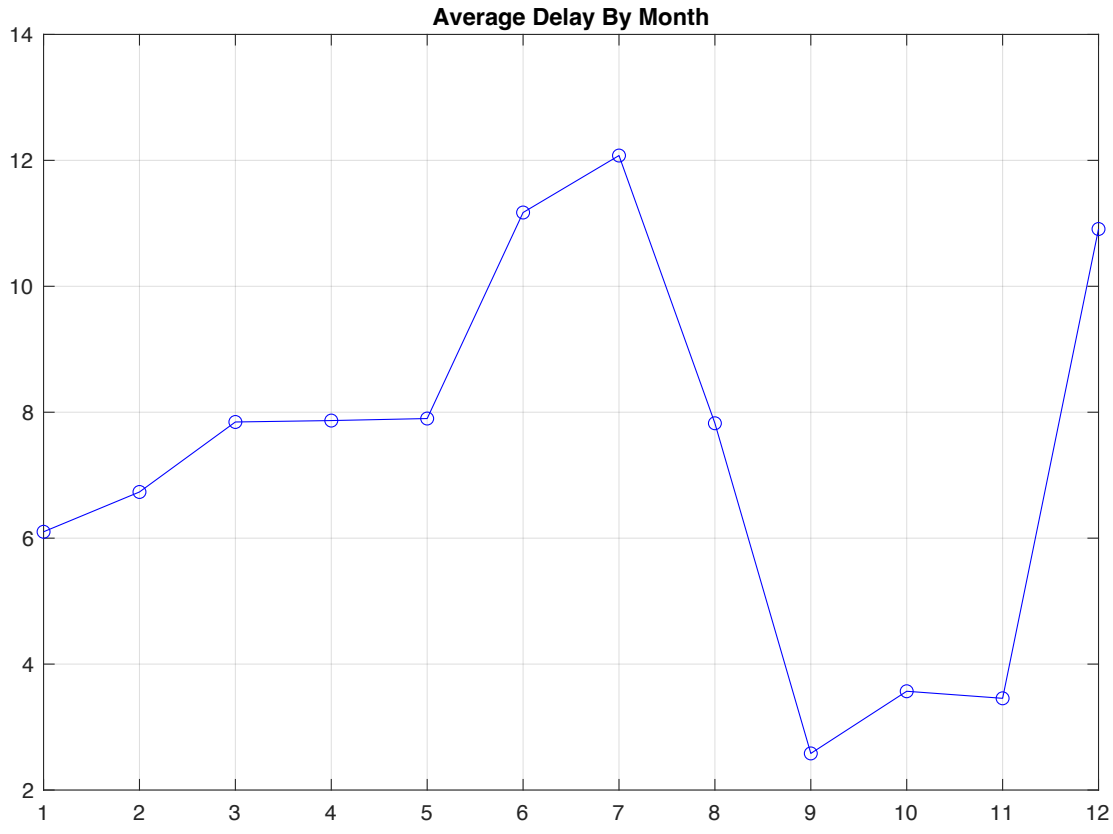
```
>> height(flights) - height(flights_clean)

ans =

        8255
```

6. Remove any departure delay greater than 2 hours (120 minutes). This leaves 318,798 observations.

```
flights_final =

  318798×9 table

    YearDay    Month    Day    origin    dest     hour    minute    dep_delay    distance
    _______    _____    ___    ______    _____    ____    ______    _________    ________

        1          1      1    "EWR"     "IAH"      5        15           2          1400
        1          1      1    "LGA"     "IAH"      5        29           4          1416
        1          1      1    "JFK"     "MIA"      5        40           2          1089
        1          1      1    "JFK"     "BQN"      5        45          -1          1576
        1          1      1    "LGA"     "ATL"      6         0          -6           762

       :          :      :      :         :        :         :           :            :

      273          9     30    "JFK"     "SYR"     22        45          -5           209
      273          9     30    "JFK"     "BUF"     22        50         -10           301
      273          9     30    "JFK"     "ROC"     22        46          -5           264
      273          9     30    "JFK"     "BOS"     22        55          12           187
      273          9     30    "JFK"     "PSE"     23        59         -10          1617

    Display all 318798 rows.
```

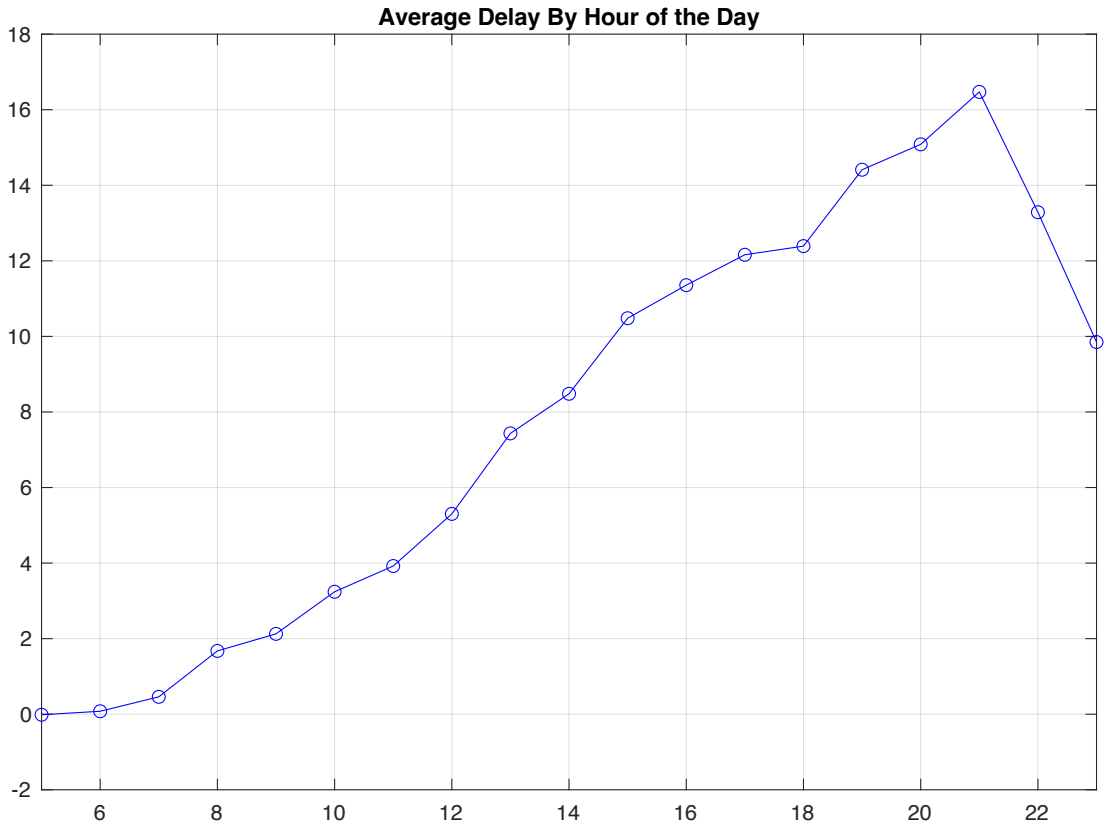
7. Generate the following table and graph, showing the average delay per month.

```
res1 =

  12×2 table

    Month    AvrDelayMonth
    _____    _____________

      1          6.1014
      2          6.7335
      3          7.8451
      4          7.8675
      5          7.8998
      6          11.172
      7          12.076
      8          7.8246
      9          2.5813
     10          3.5686
     11          3.4567
     12           10.91
```



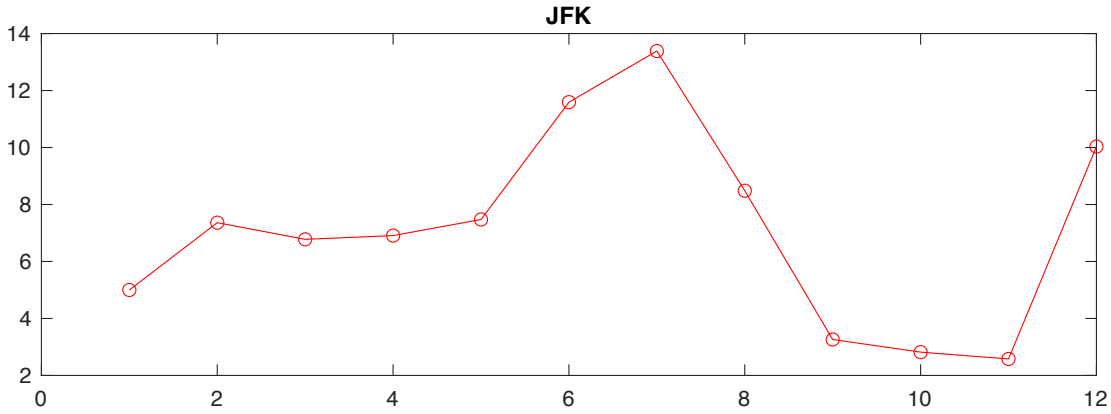
8. Generate the following table and graph, showing the average delay per hour.

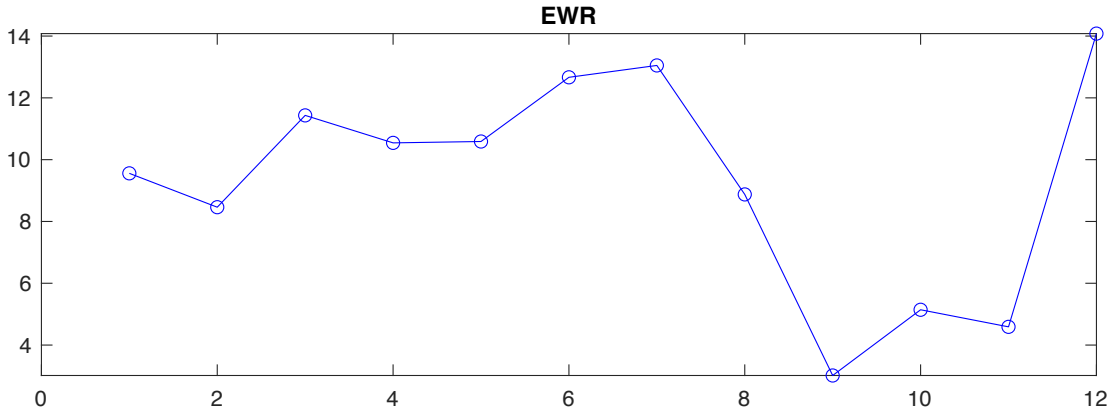
```
res2 =

  19×2 table

    Hour    AvrDelayHour
    ____    ____________

      5       -0.013953
      6        0.077367
      7         0.45981
      8          1.6756
      9          2.1263
     10          3.2408
     11          3.9218
     12          5.3017
     13           7.433
     14          8.4813
     15          10.481
     16          11.357
     17          12.161
     18           12.39
     19          14.412
     20          15.083
     21          16.472
     22          13.287
     23          9.8507
```

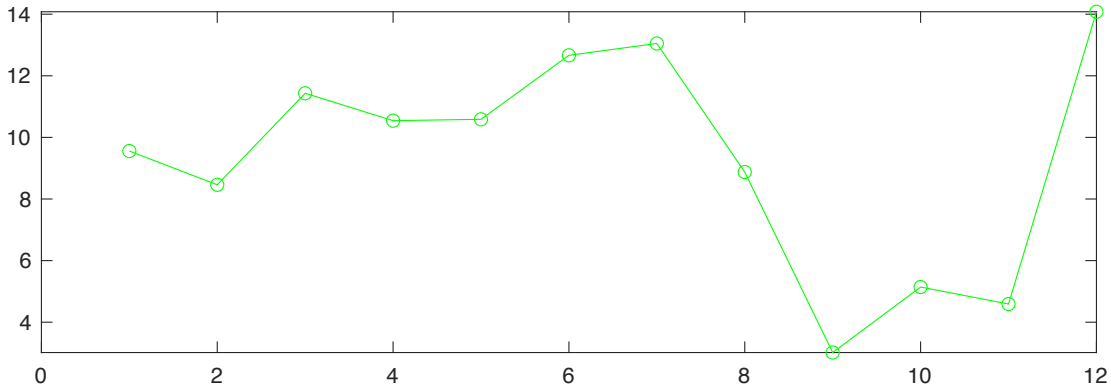


9. Generate the following table and graph, showing the average delay by month and by origin

```
res1 =

  36×3 table

    Month    Origin    AvrDelayMonthOrigin
    _____    ______    ___________________

      1      "EWR"           9.5572
      1      "JFK"           5.0003
      1      "LGA"           3.1567
      2      "EWR"           8.4617
      2      "JFK"           7.3617
      2      "LGA"           3.9428
      3      "EWR"           11.433
      3      "JFK"            6.777
      3      "LGA"           4.8108
      4      "EWR"           10.543
      4      "JFK"           6.9087
      4      "LGA"           5.6536
      5      "EWR"           10.588
      5      "JFK"           7.4736
      5      "LGA"           5.1239
      6      "EWR"           12.665
      6      "JFK"           11.592
      6      "LGA"           8.9138
      7      "EWR"           13.049
      7      "JFK"           13.387
      7      "LGA"           9.4009
      8      "EWR"           8.8749
      8      "JFK"           8.4828
      8      "LGA"           5.8566
      9      "EWR"           3.0154
      9      "JFK"           3.2613
      9      "LGA"           1.4469
     10      "EWR"           5.1424
     10      "JFK"           2.8182
     10      "LGA"           2.6456
     11      "EWR"           4.5897
     11      "JFK"           2.5788
     11      "LGA"           3.0805
     12      "EWR"           14.079
     12      "JFK"           10.033
     12      "LGA"           8.4138
```


JFK
14
12
10
8
6
4
2
0
2
4
6
8
10
12


EWR
14
12
10
8
6
4
0
2
4
6
8
10
12


14
12
10
8
6
4
0
2
4
6
8
10
12