A study of flight delays from NY

by Mihai Munteanu

Contents

The data

~ 330 000 flights leaving from 3 NY airports in 2013

```
Total flights by origin
origin
EWR     145179
LGA     100663
JFK      72154
```

together with hourly weather data for the 3 airports

The problems with the data

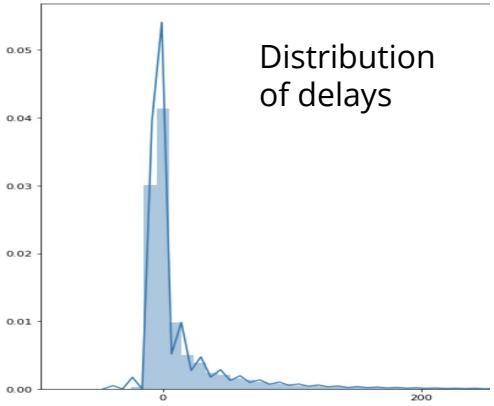
Weird outliers in the data:

	departure	scheduled_departure	flight_id	departure_delay
151	848	1835	3944	-587
7029	641	900	51	-139
8190	1121	1635	3695	-314
56746	603	1645	2042	-642

- Did Flight 3695 depart 5h early or 19h late?
- Are {'SJU', 'ERW', 'PSE', 'BQN', 'STT'} actually airports?
- The normal stuff: missing departure/arrival data, weather conditions data

Questions

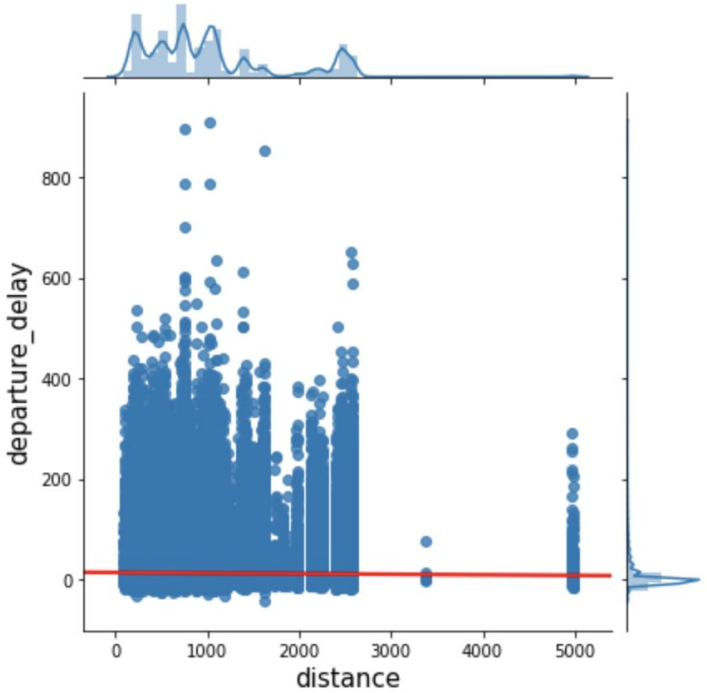
Main Goal: Figure out what influences flight delays and how



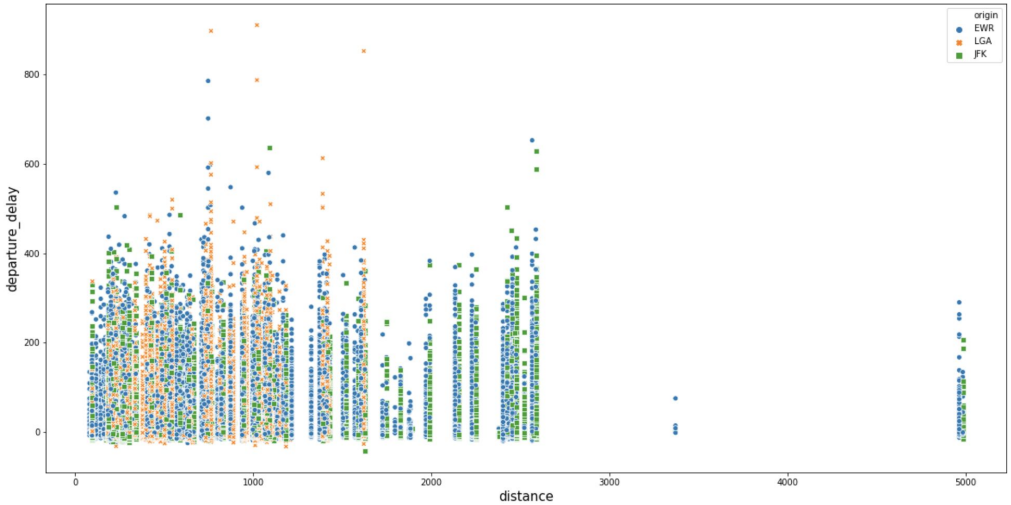
1. Do longer flights have longer delays than shorter flights?
2. Do evening flights have longer delays than morning flights?
3. Does big airport imply more crowding and hence longer delays?
4. Are some airlines more efficient than others in dealing with delays?
5. Is weather the main culprit for delays?
6. Any other questions/curiosities you have?

Let the data speak

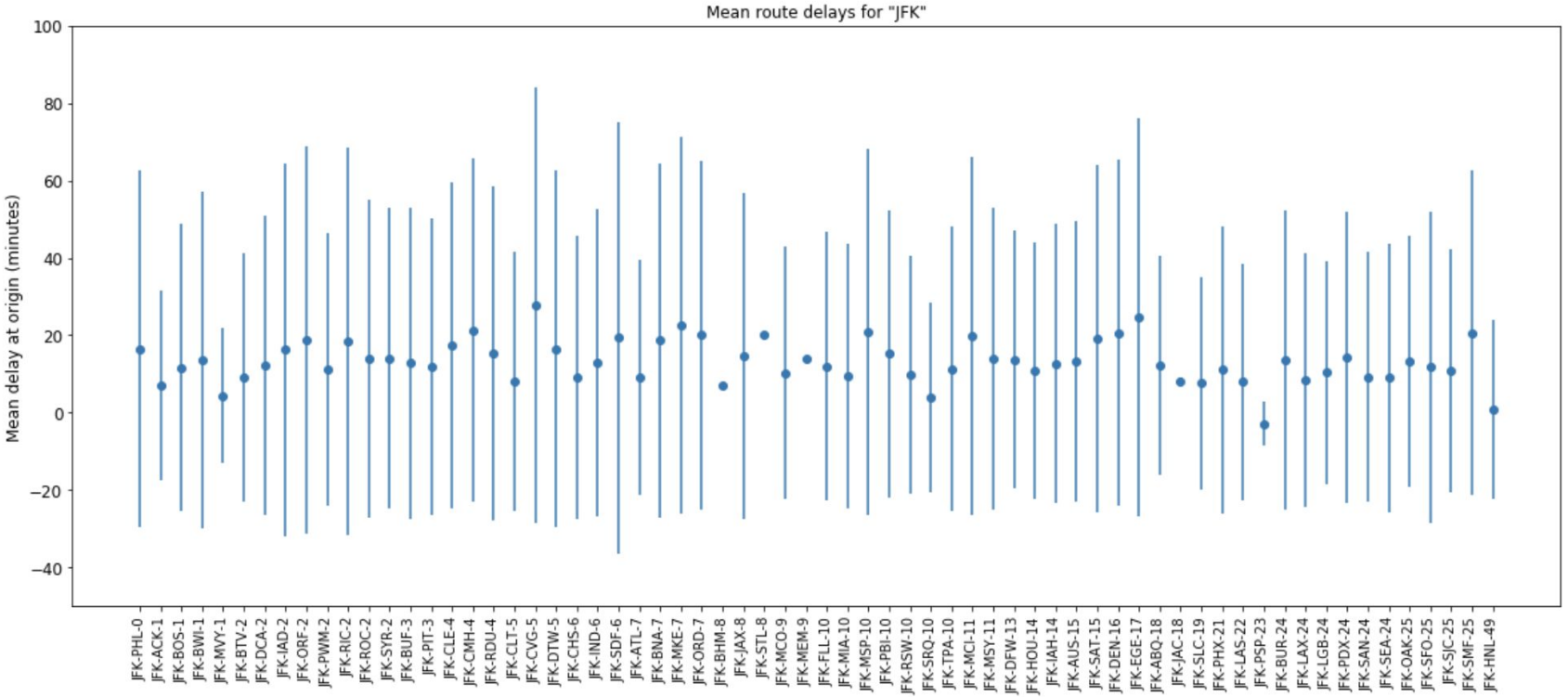
Distance of a flight does not have a significant impact on **delays**.




```
Percentages of delayed flights by origin
origin
EWR    0.434698
JFK    0.381919
LGA    0.331353
dtype: float64
```

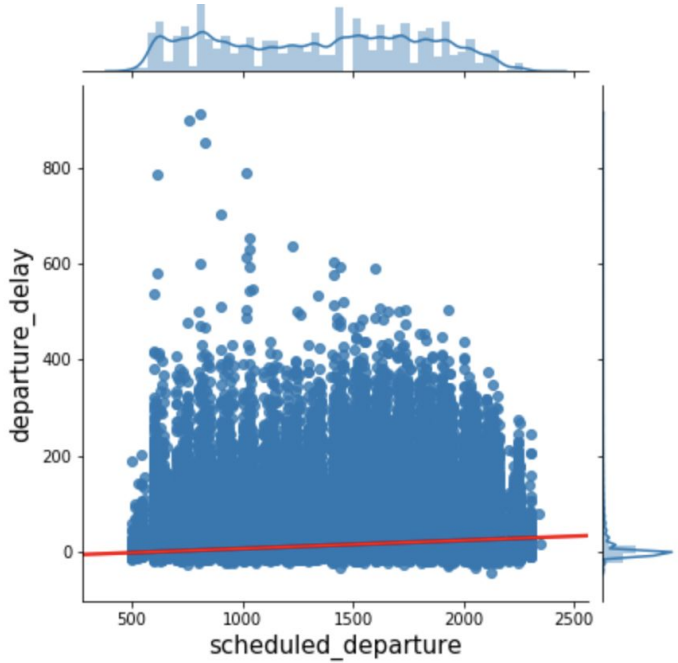


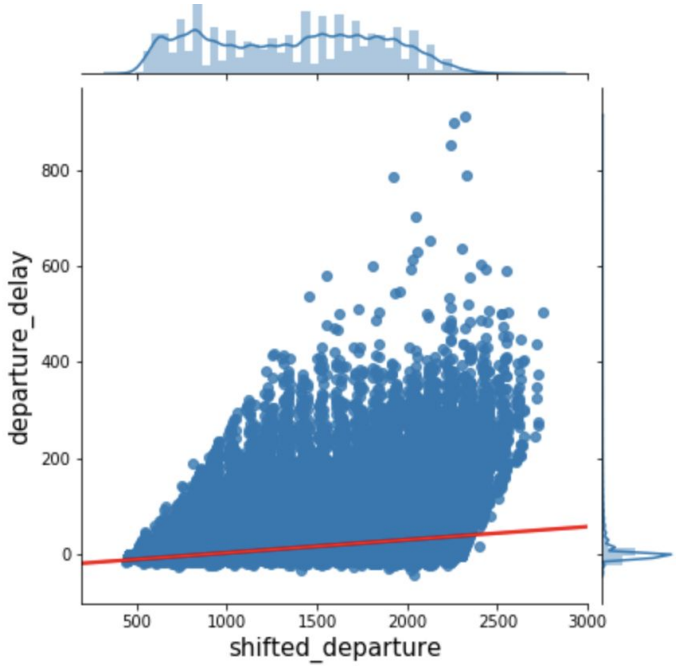
Let the data speak



Let the data speak

Generally, later **scheduled departure** times cause bigger **delays**.





Let the data speak

JFK is the best airport in town in regards to delays.

```
Stats for departure delay by origin
        min    max    count       mean        std
origin
EWR     1.0  786.0  63109.0  38.683944  51.745666
LGA     1.0  911.0  33355.0  41.413731  56.799874
JFK     1.0  636.0  27557.0  38.121457  50.913966
```

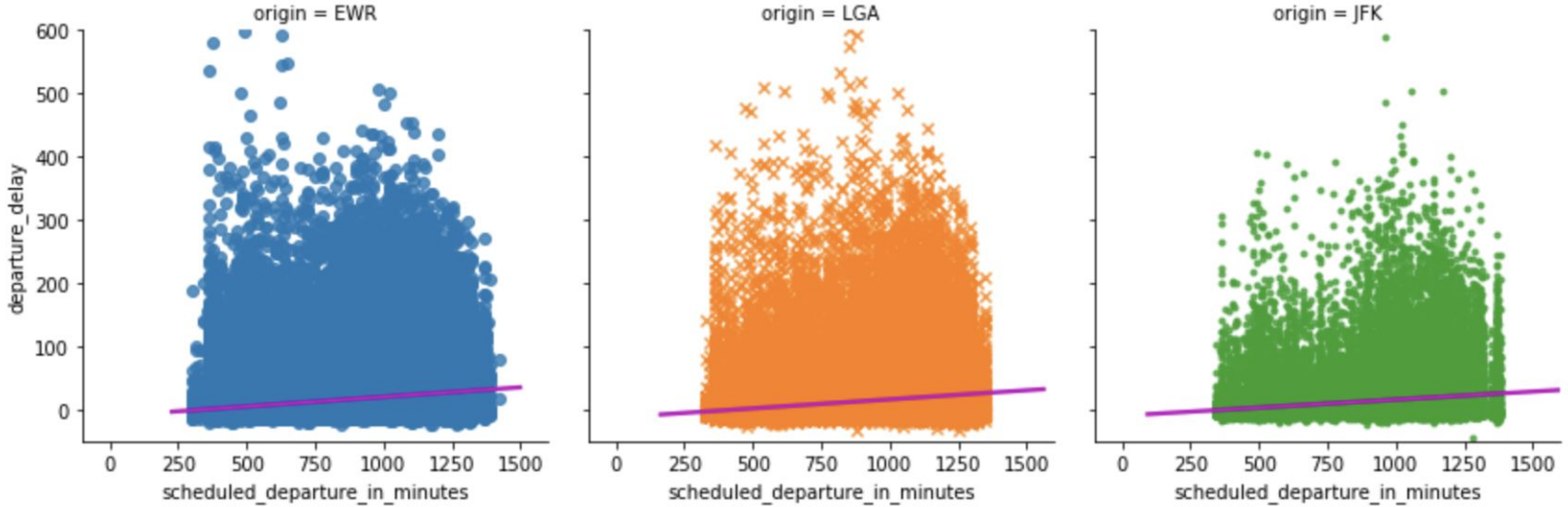
```
Percentages of delayed flights by origin
origin
EWR    0.434698
JFK    0.381919
LGA    0.331353
dtype: float64
```

Taking into account the mean delay time and the maximum delay time:

```
LGA is recommended for 38 destinations.
JFK is recommended for 56 destinations.
EWR is recommended for 6 destinations.
```

Let the data speak

The bigger the **airport** the bigger the **delay** caused by **scheduled departure**.

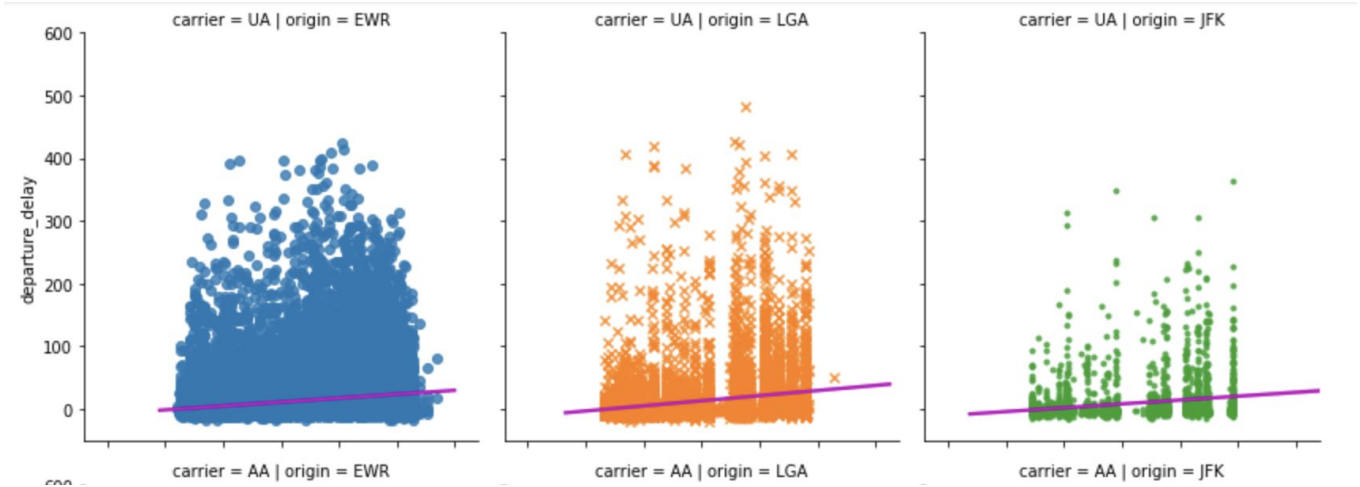


Slope	0.30	0.28	0.245
Total flights	145179	100663	72154

Let the data speak

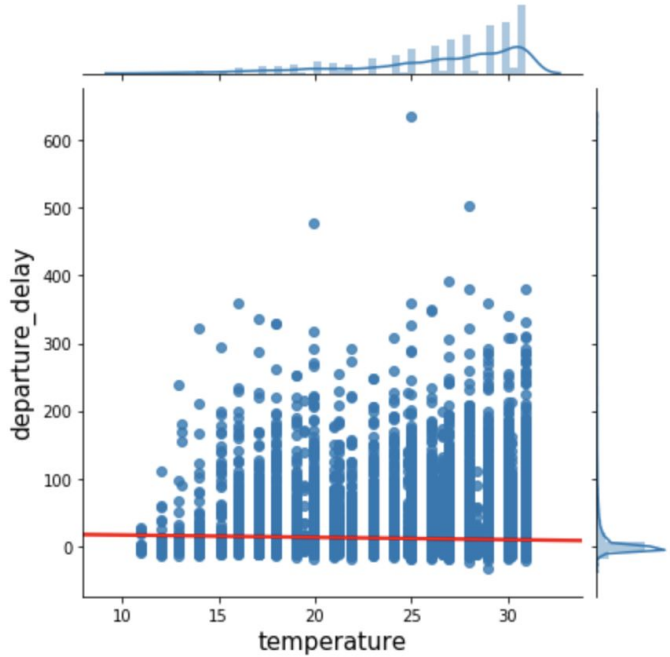
Airline influences **delay**, but the interplay with the **airport** is complicated.

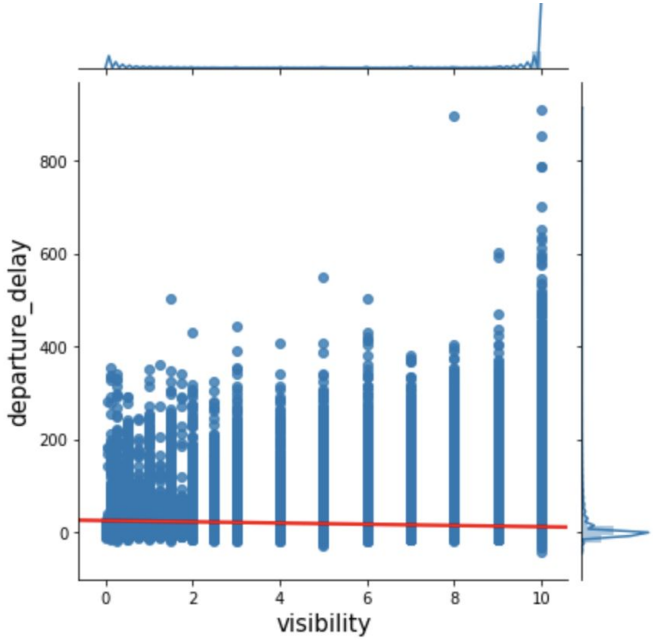
```
Percentages of delayed flights by carrier
carrier
WN    0.542602
FL    0.517591
F9    0.500739
UA    0.467530
EV    0.449564
VX    0.432953
YV    0.430427
9E    0.403788
B6    0.393612
DL    0.319517
MQ    0.318253
AS    0.316312
AA    0.315821
OO    0.310345
US    0.240708
HA    0.195846
dtype: float64
```



Let the data speak

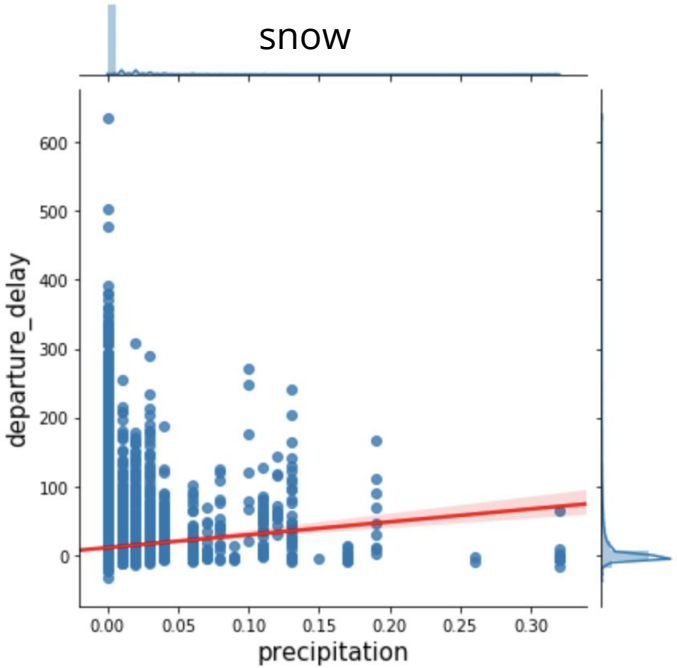
Cold **temperature** and **visibility** issues cause longer **delays** (but not by much).

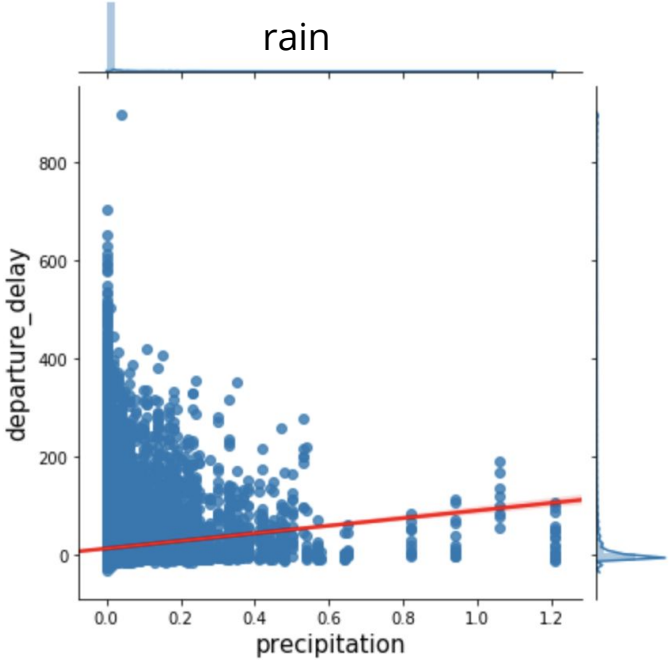




Let the data speak

As expected, the more **snow** the longer **delays** (the same for heavy **rain**).





Conclusions

If you are flying from NY and don't want to experience the dreaded delays:

- **Wake up early and fly from JFK!**
- Don't wish for snow! Why would you want to leave NY if it's snowing anyways?
- Avoid Newark.
- Reconsider whether it is time to let go your favourite airline.
- The distance of a flights and the time of a year don't really matter!
- Consider for how long you need to be in the NY traffic to get to your favourite airport (personal experience, the traffic study is for another time).