

Compte rendu projet Analyse donnée avec R

1 Les statistiques

1.1 Les vols (données flights)

Le dataframe flights se compose de 336 776 lignes qui correspondent à autant de vols.

Il y a 19 paramètres décrivant chaque vols :

Pour l'année 2013 au départ de 3 aéroports de New York

year	month	day	dep_time	sched_dep_time
Min. :2013	Min. : 1.000	Min. : 1.00	Min. : 1	Min. : 106
1st Qu.:2013	1st Qu.: 4.000	1st Qu.: 8.00	1st Qu.: 907	1st Qu.: 906
Median :2013	Median : 7.000	Median :16.00	Median :1401	Median :1359
Mean :2013	Mean : 6.549	Mean :15.71	Mean :1349	Mean :1344
3rd Qu.:2013	3rd Qu.:10.000	3rd Qu.:23.00	3rd Qu.:1744	3rd Qu.:1729
Max. :2013	Max. :12.000	Max. :31.00	Max. :2400	Max. :2359
			NA's :8255	

dep_delay	arr_time	sched_arr_time	arr_delay
Min. : -43.00	Min. : 1	Min. : 1	Min. : -86.000
1st Qu.: -5.00	1st Qu.:1104	1st Qu.:1124	1st Qu.: -17.000
Median : -2.00	Median :1535	Median :1556	Median : -5.000
Mean : 12.64	Mean :1502	Mean :1536	Mean : 6.895
3rd Qu.: 11.00	3rd Qu.:1940	3rd Qu.:1945	3rd Qu.: 14.000
Max. :1301.00	Max. :2400	Max. :2359	Max. :1272.000
NA's :8255	NA's :8713		NA's :9430

carrier	flight	tailnum	origin
Length:336776	Min. : 1	Length:336776	Length:336776
Class :character	1st Qu.: 553	Class :character	Class :character
Mode :character	Median :1496	Mode :character	Mode :character
	Mean :1972		
	3rd Qu.:3465		
	Max. :8500		

dest	air_time	distance	hour
Length:336776	Min. : 20.0	Min. : 17	Min. : 1.00
Class :character	1st Qu.: 82.0	1st Qu.: 502	1st Qu.: 9.00
Mode :character	Median :129.0	Median : 872	Median :13.00
	Mean :150.7	Mean :1040	Mean :13.18
	3rd Qu.:192.0	3rd Qu.:1389	3rd Qu.:17.00
	Max. :695.0	Max. :4983	Max. :23.00
	NA's :9430		

minute	time_hour
Min. : 0.00	Min. :2013-01-01 05:00:00
1st Qu.: 8.00	1st Qu.:2013-04-04 13:00:00
Median :29.00	Median :2013-07-03 10:00:00
Mean :26.23	Mean :2013-07-03 05:22:54
3rd Qu.:44.00	3rd Qu.:2013-10-01 07:00:00
Max. :59.00	Max. :2013-12-31 23:00:00

1.2 Statistiques du dataframe Weather

origin	year	month	day
Length:26115	Min. :2013	Min. : 1.000	Min. : 1.00
Class :character	1st Qu.:2013	1st Qu.: 4.000	1st Qu.: 8.00
Mode :character	Median :2013	Median : 7.000	Median :16.00
	Mean :2013	Mean : 6.504	Mean :15.68
	3rd Qu.:2013	3rd Qu.: 9.000	3rd Qu.:23.00
	Max. :2013	Max. :12.000	Max. :31.00

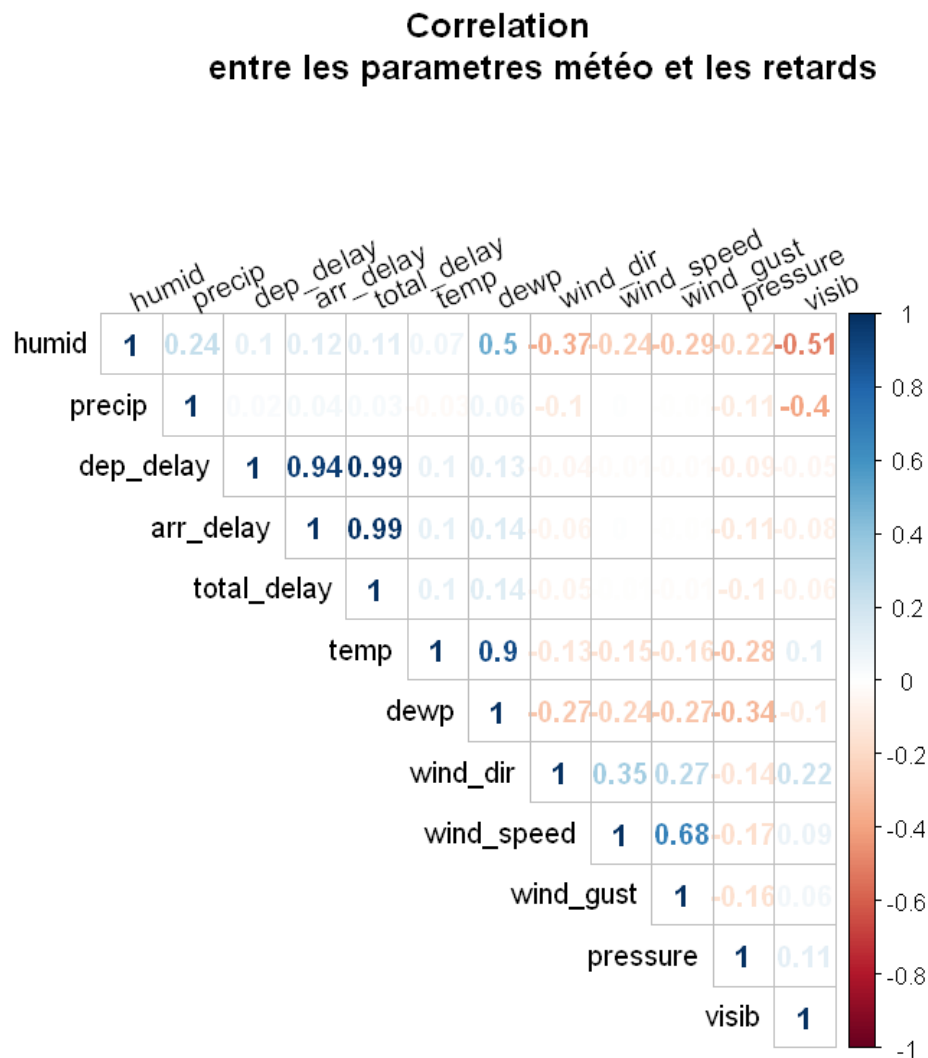
hour	temp	dewp	humid
Min. : 0.00	Min. : 10.94	Min. : -9.94	Min. : 12.74
1st Qu.: 6.00	1st Qu.: 39.92	1st Qu.:26.06	1st Qu.: 47.05
Median :11.00	Median : 55.40	Median :42.08	Median : 61.79
Mean :11.49	Mean : 55.26	Mean :41.44	Mean : 62.53
3rd Qu.:17.00	3rd Qu.: 69.98	3rd Qu.:57.92	3rd Qu.: 78.79
Max. :23.00	Max. :100.04	Max. :78.08	Max. :100.00
	NA's :1	NA's :1	NA's :1

wind_dir	wind_speed	wind_gust	precip
Min. : 0.0	Min. : 0.000	Min. : 0.000	Min. :0.000000
1st Qu.:120.0	1st Qu.: 6.905	1st Qu.: 0.000	1st Qu.:0.000000
Median :220.0	Median : 10.357	Median : 0.000	Median :0.000000
Mean :199.8	Mean : 10.518	Mean : 5.209	Mean :0.004469
3rd Qu.:290.0	3rd Qu.: 13.809	3rd Qu.: 0.000	3rd Qu.:0.000000
Max. :360.0	Max. :1048.361	Max. :66.745	Max. :1.210000
NA's :460	NA's :4		

pressure	visib	time_hour
Min. : 983.8	Min. : 0.000	Min. :2013-01-01 01:00:00
1st Qu.:1012.9	1st Qu.:10.000	1st Qu.:2013-04-01 21:30:00
Median :1017.6	Median :10.000	Median :2013-07-01 14:00:00
Mean :1017.9	Mean : 9.255	Mean :2013-07-01 18:26:37
3rd Qu.:1023.0	3rd Qu.:10.000	3rd Qu.:2013-09-30 13:00:00
Max. :1042.1	Max. :10.000	Max. :2013-12-30 18:00:00
NA's :2729		

2 Effets de la météo sur les retards

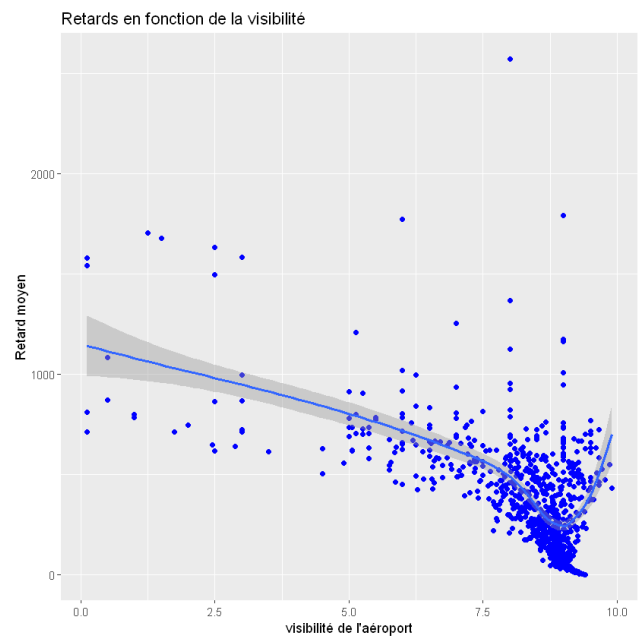
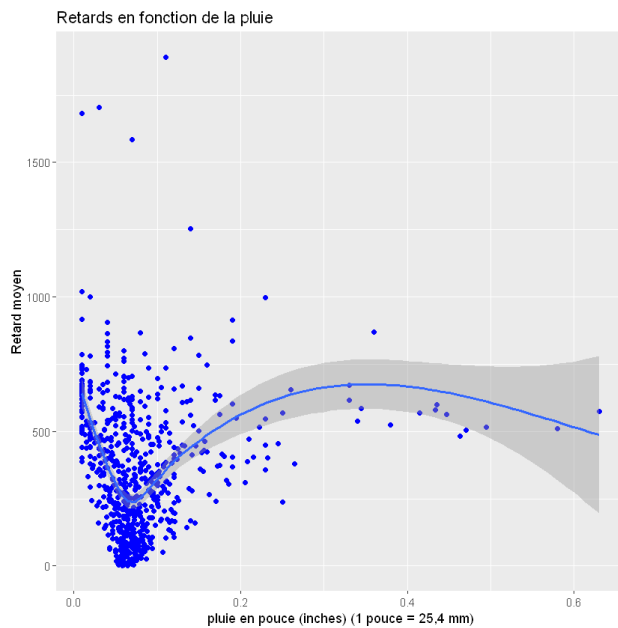
Suite à l'analyse des données des DataFrames nous avons cherché la corrélation entre les événements météorologiques et les retards.



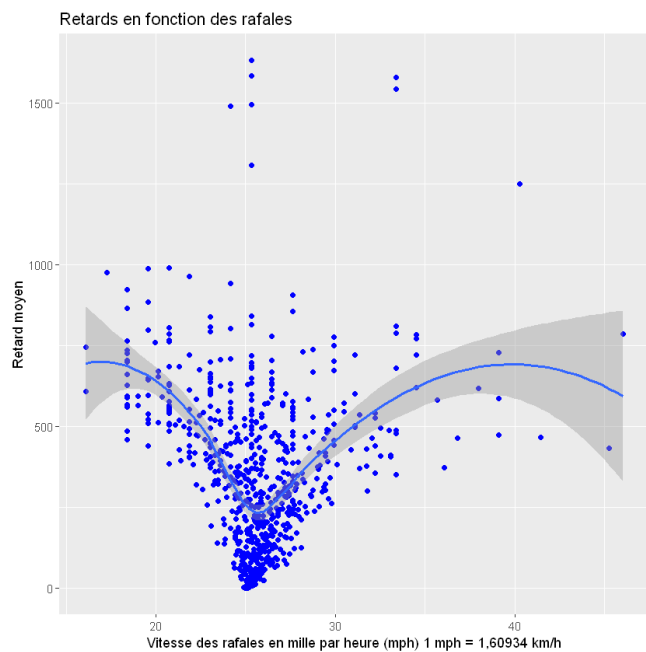
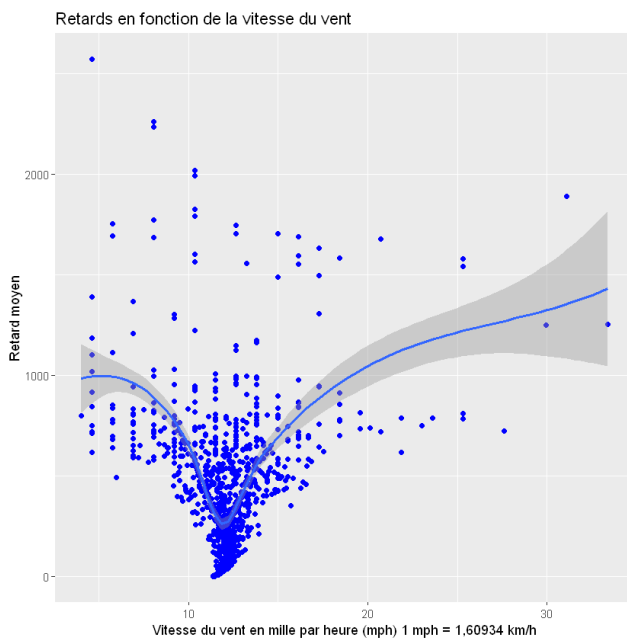
Ce graph ne met pas en évidence de corrélation entre les évènements météorologiques et les retards.

Nous analysons individuellement certain événement.

2.1 Retard en fonction de la pluie et de la visibilité



2.2 Retard en fonction du vent et des rafales



2.3 Conclusion

A partir des graphiques on constate que :

- Il n'y a pas de corrélation entre la pluie et les retards
- Il y a une légère corrélation entre la visibilité et des retards :
 - Plus il y a de visibilité moins il y a de retards
- Au-delà de 12 mph (19 km/h) de vent il y a une légère corrélation avec les retards :
 - Au-delà de 19 km/h, plus il y a de vents, plus il y a de retard
- Pour les vitesses de rafales au-delà de 26 mph (42 km/h) il y a une corrélation avec les retards