**Which weather conditions factors for the delay of flights from New York city?**

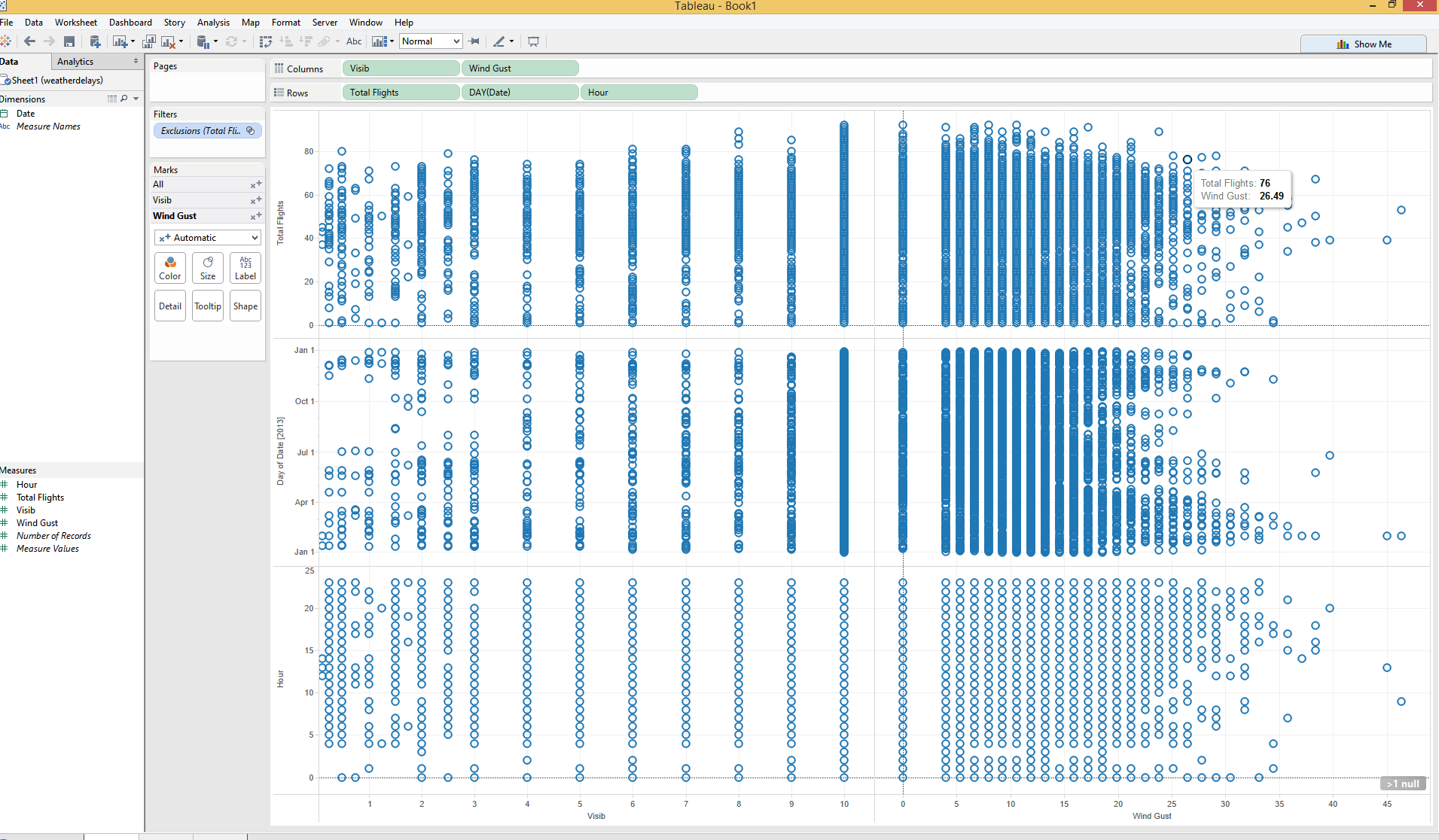
Have tested the data for Wind Gust and Visibility, it does correlate well with visibility as shown in the chart below. Other weather conditions can also be plotted with the same query.

Select w.year,w.month,w.day,w.hour,count(f.tailnum) as Total\_Flights, b.wind\_gust,b.visib from flights w join flights f on w.year=f.year and w.month=f.month and w.day=f.day

and w.hour=f.hour and w.tailnum=f.tailnum and w.dep\_delay =f.dep\_delay join weather b on w.year=b.year and w.month=b.month and w.day=b.day and w.hour=b.hour where w.origin in('LGA','EWR','JFK') and w.year=f.year and w.month=f.month and w.day=f.day

and w.hour=f.hour and w.tailnum=f.tailnum and w.dep\_delay =f.dep\_delay and w.year=b.year and w.month=b.month and w.day=b.day

and w.hour=b.hour group by w.year,w.month,w.day,w.hour,b.wind\_gust,b.visib order by count(f.dep\_delay) DESC



**Are old planes are also responsible for the delay?**

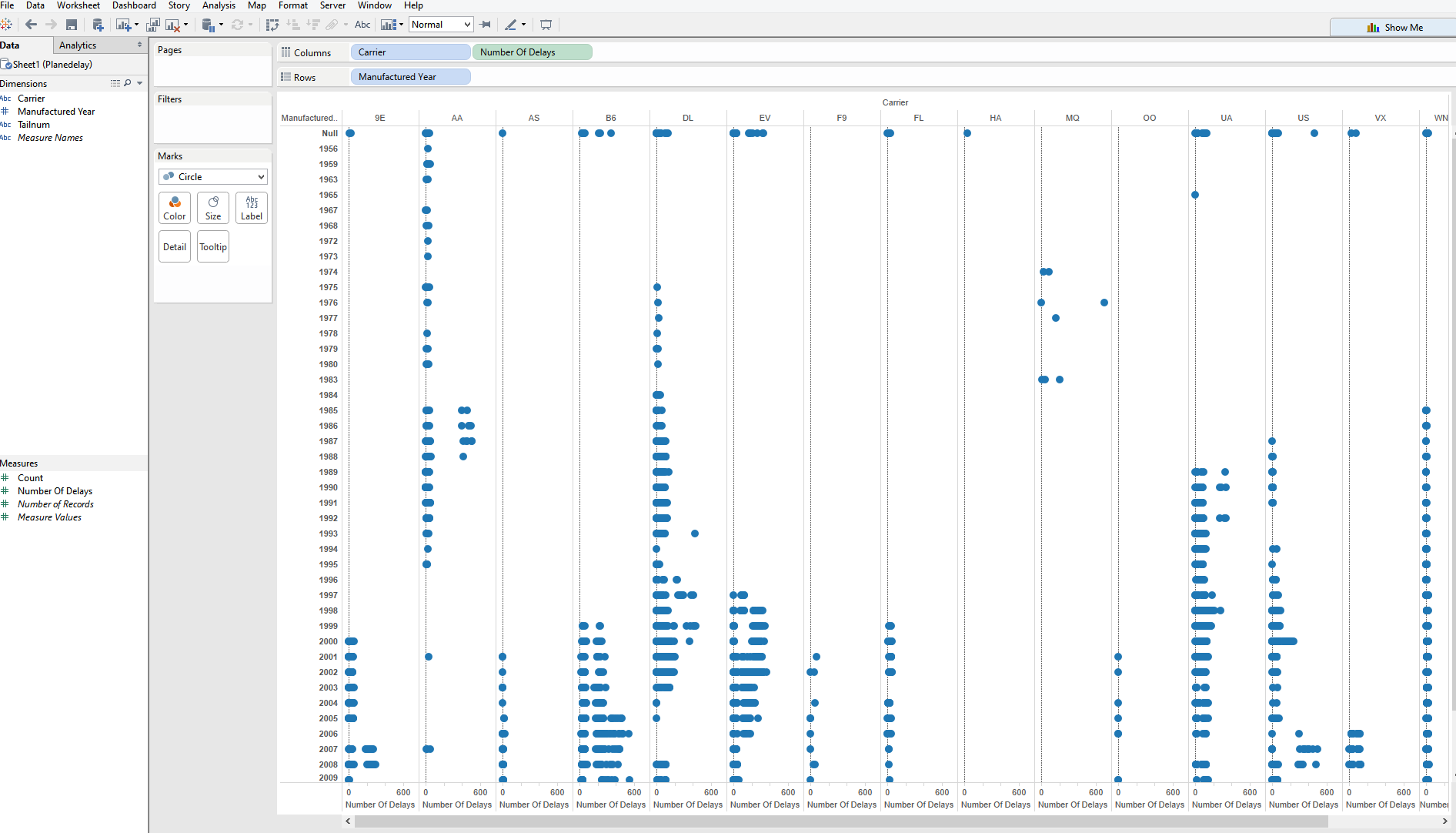
**After plotting the data it seems like Old planes are not the cause of the delay as most of the airlines which are manufactured in 1956 to 1986 have less number of delays than the later part. Mostly aggregated in 1998 to2002.**

Select a.carrier, a.tailnum, p.year as Manufactured\_Year, count(b.tailnum),count(b.dep\_delay) as Number\_Of\_delays from flights a join flights b

on a.tailnum=b.tailnum and a.dep\_delay=b.dep\_delay and a.arr\_delay=b.arr\_delay and a.carrier=b.carrier left join planes p on a.tailnum=p.tailnum where a.tailnum=b.tailnum and a.tailnum=p.tailnum

and a.origin=b.origin and a.origin in ('LGA','EWR','JFK')

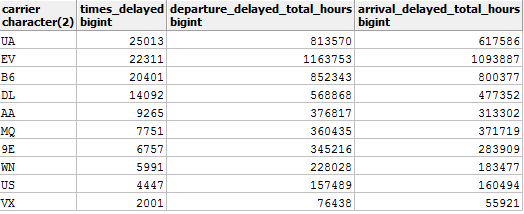
group by a.tailnum,p.year,a.carrier, a.origin ORDER BY Manufactured\_Year ASC



Other Questions which could arise from this data set are:

**Top 10 airlines which are mostly delayed ?**

select carrier,count(carrier) as Times\_Delayed, sum(dep\_delay) as Departure\_Delayed\_Total\_Hours,sum(arr\_delay) as Arrival\_Delayed\_Total\_Hours from flights where origin in('LGA','EWR','JFK') and dep\_delay>1 group by carrier order by Count(carrier) desc limit 10



**Which is the most busiest route for flights and which airline has maximum number of flights in that route?**

Select a.carrier,a.tailnum, count(b.tailnum), a.origin, a.dest from flights a join flights b

on a.tailnum=b.tailnum and a.origin=b.origin and a.dest=b.dest and a.carrier=b.carrier where a.tailnum=b.tailnum group by a.tailnum,a.origin,a.dest,a.carrier ORDER BY COUNT(B.TAILNUM) DESC