Dallas/Fort Worth Airport (DFW) flight data

Junzhao Hu

Part 1: Introduction

Below is the link for the data, and I downloaded data for November and December of 2013.

Links:

- [http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time]

I am interested in several things:

1 Is there a day of the week/ time of day effect on departure or arrival delays for the flights.

2 Which airports cancel or delay the most from Dallas/Fort Worth Airport (DFW).

3 Which Carrier delays or cancels most flights.

4 What kind of factors affect the delay time of the flights (like distance, depart time, arrive time)

5 What's the main reason for the delay and cancelation.

Part 2: Data analysis

It contains about a million rows of fights and more than a hundred variables.

I just choose the rows that have DFW as the Origin/Destination and the variables that are shown in the table 1.

Table 1

Table : Description of Table Columns

Variable	Description
DayofMonth	Day of Month
DayOfWeek	Day of Week
FlightDate	Flight Date (yyyymmdd)
Carrier	Code assigned by IATA
TailNum	Tail Number
Origin	Origin Airport
Dest	Destination Airport
DepTime	Actual Departure Time (local time: hhmm)
DepDelay	Difference in minutes between scheduled and actual departure time.
ArrDelay	Difference in minutes between scheduled and actual arrival time.
Cancelled	Cancelled Flight Indicator (1=Yes)
Distance	Distance between airports (miles)

With the data in the flights database, we can now use dplyr package to begin analysis.

First, we use Dallas/Fort Worth Airport (DFW) as origin.

Figure 1 is the relation between the average distance of the flights and the average delay of the flights, and it's shown that most flights are middle distance or short distance flights, and the average delayed time for time for most flights are less than 20 minutes.

Also the average delay time of middle (600-900) and short (<600) average distance flights is more than that of long average distance (>1000).

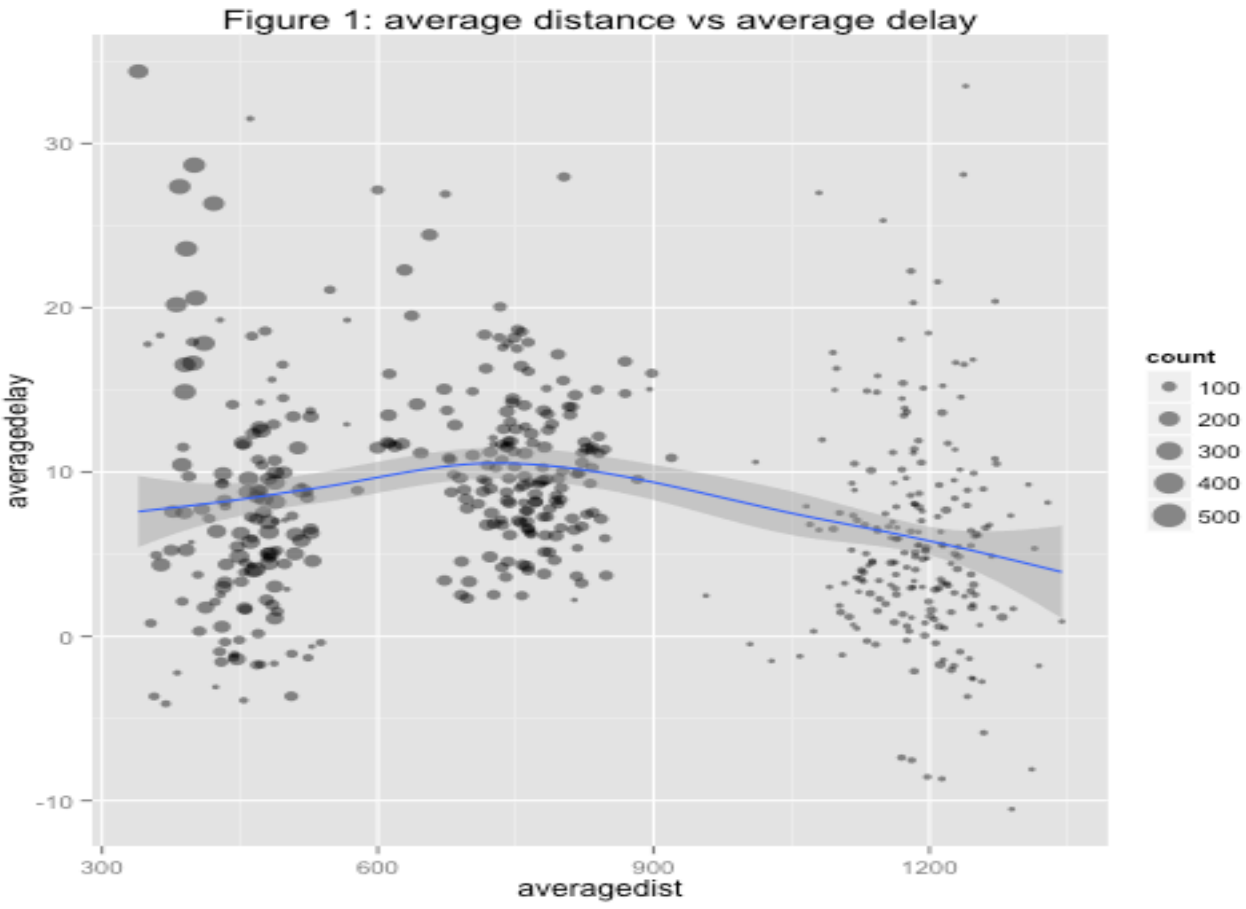


Table 2 below is the top 10 destinations from DFW, and Table 3 below is the top 10 origins to DFW.

From the tables, we see that the top 10 origin/destinations are near to Dallas, and the cities in which the 10 airports have very close economic relation with Dallas.

Table 2

	Destination	# of Planes	# of flights
1	LAX	465	1348
2	ATL	350	1201
3	ORD	344	1140
4	DEN	507	1087
5	SFO	435	1017
6	IAH	447	900
7	PHX	360	868
8	SAT	164	862
9	CLT	234	807
10	AUS	240	802

Table 3

	Origin	# of Planes	# of flights
1	LAX	464	1348
2	ATL	350	1201
3	DEN	320	1126
4	ORD	529	1103
5	SFO	452	1013
6	PHX	353	868
7	SAT	164	862
8	IAH	387	843
9	CLT	234	807
10	AUS	238	803

Figure 2 shows that most cancelation rates from are from 4% to 10%.

In Figure 3, for the destination flights number more than 500 within two months, flights go to HOU airport have the highest cancelation rate (11%) from DFW airport.

Figure 2: Cancel rate of origin airport from DFW

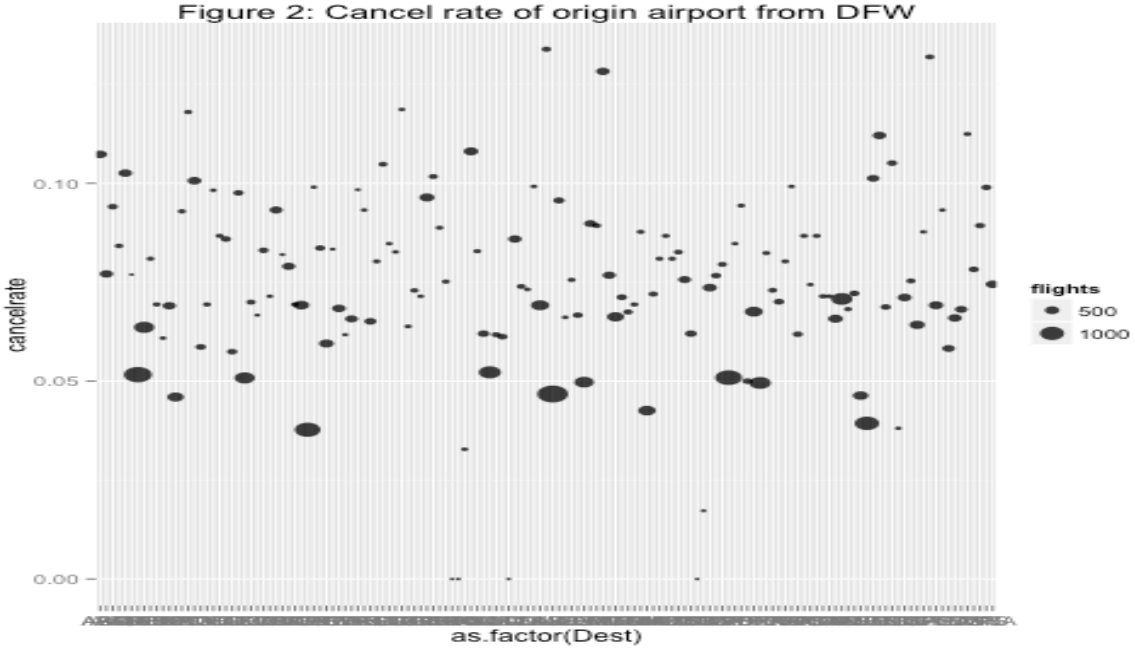
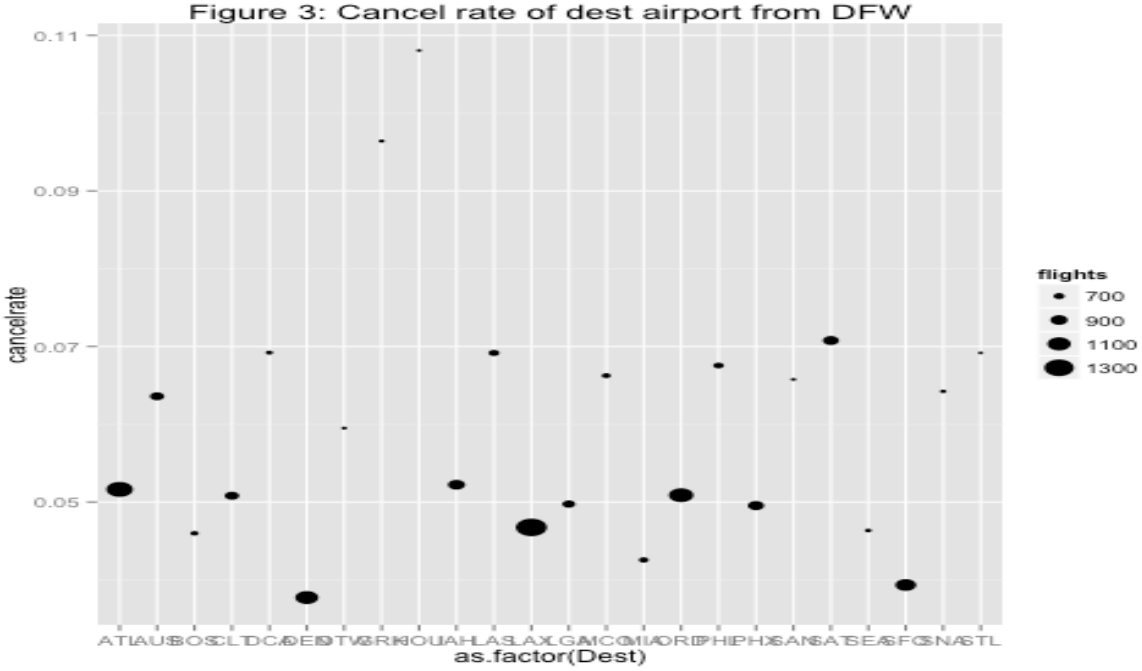
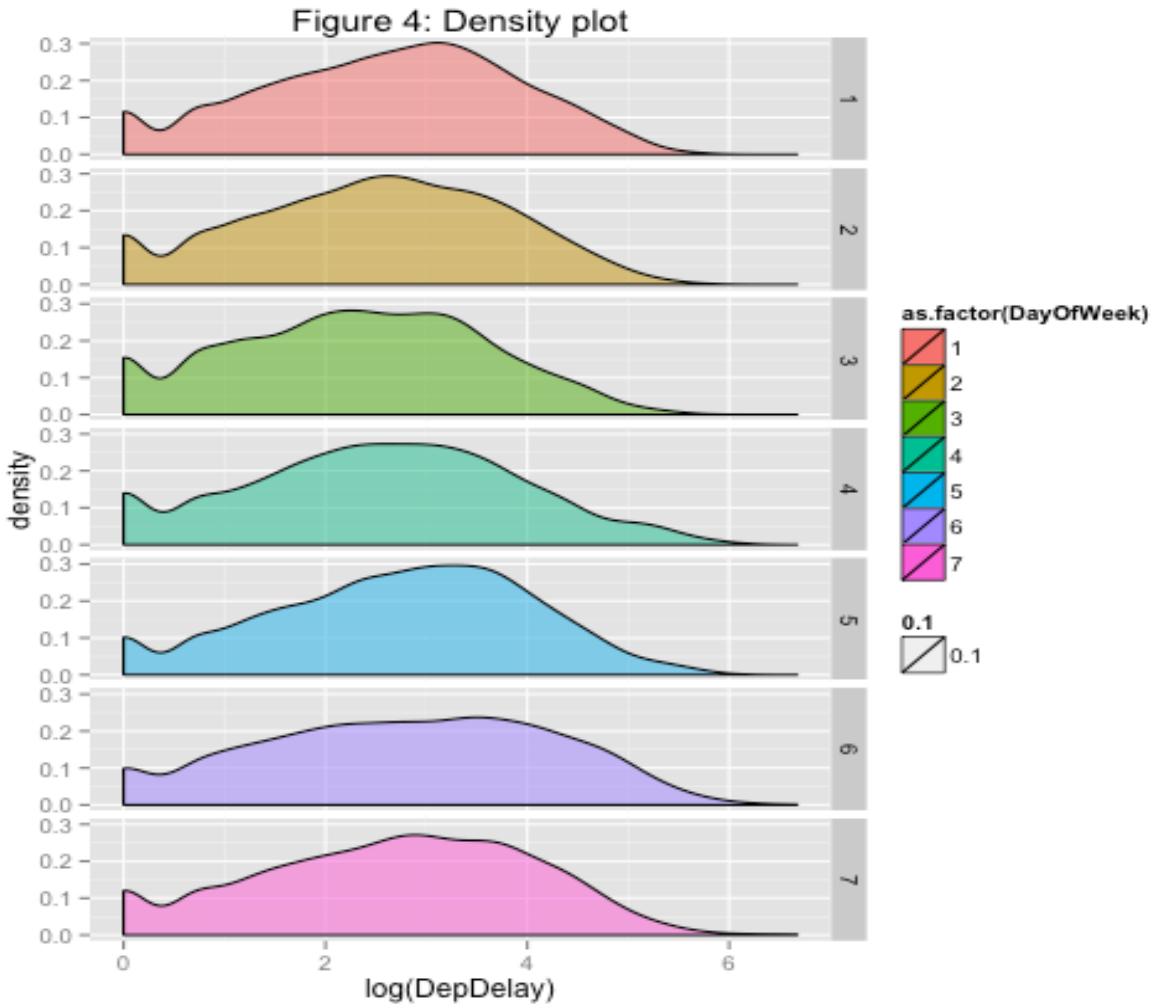


Figure 3: Cancel rate of dest airport from DFW

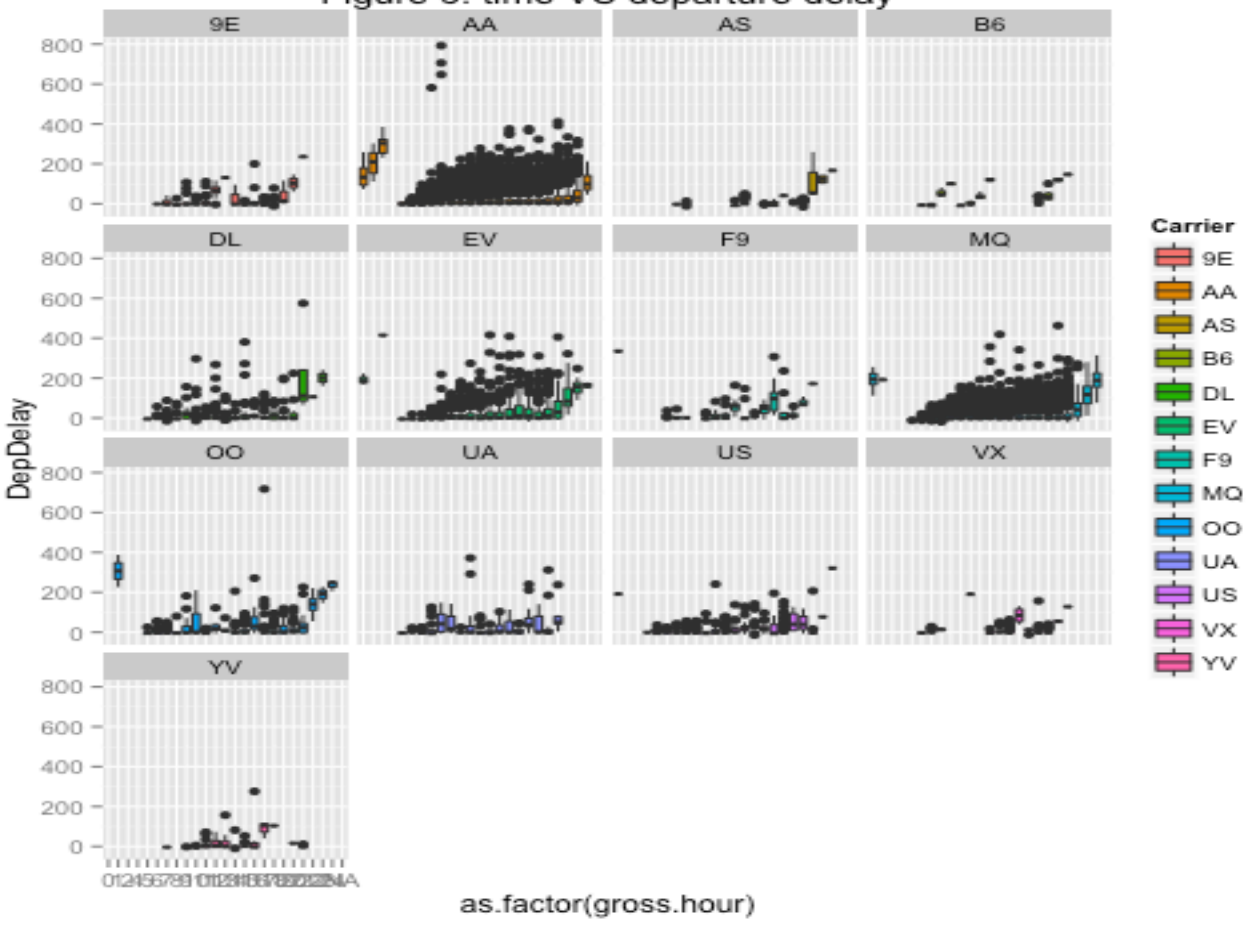


In Figure 4, the density plot shows that the delay time for most delayed flights are less than 150 minutes (e^5).



In Figure 5, notice from the boxplots that the time of delay grows for most airlines over the course of the day. This may be due to a domino effect as one late plane causes backups at terminals and on the runway leading to more delays for the airline. This causes the early morning flights to generally have a delay of more than one hour. The carriers most affected by this seem to be MQ (whose delays grow quickly at the end of the day), AA (which has the worst late flight delays and seems to be most effected at night) and OO which has no on time flights after 10:00 PM).

Figure 5: time VS departure delay



In Figure 6 and 7, flights that occur later in the week and on the weekend are more often cancelled, with airlines EV, MQ and 9E cancelling more than 15% of flights. 9E is more likely to cancel long flights while EV is more likely to cancel shorter flights. The rest of the airlines seem to have no relationship between flight length and cancellation.

Figure 6: carrier VS cancel rate

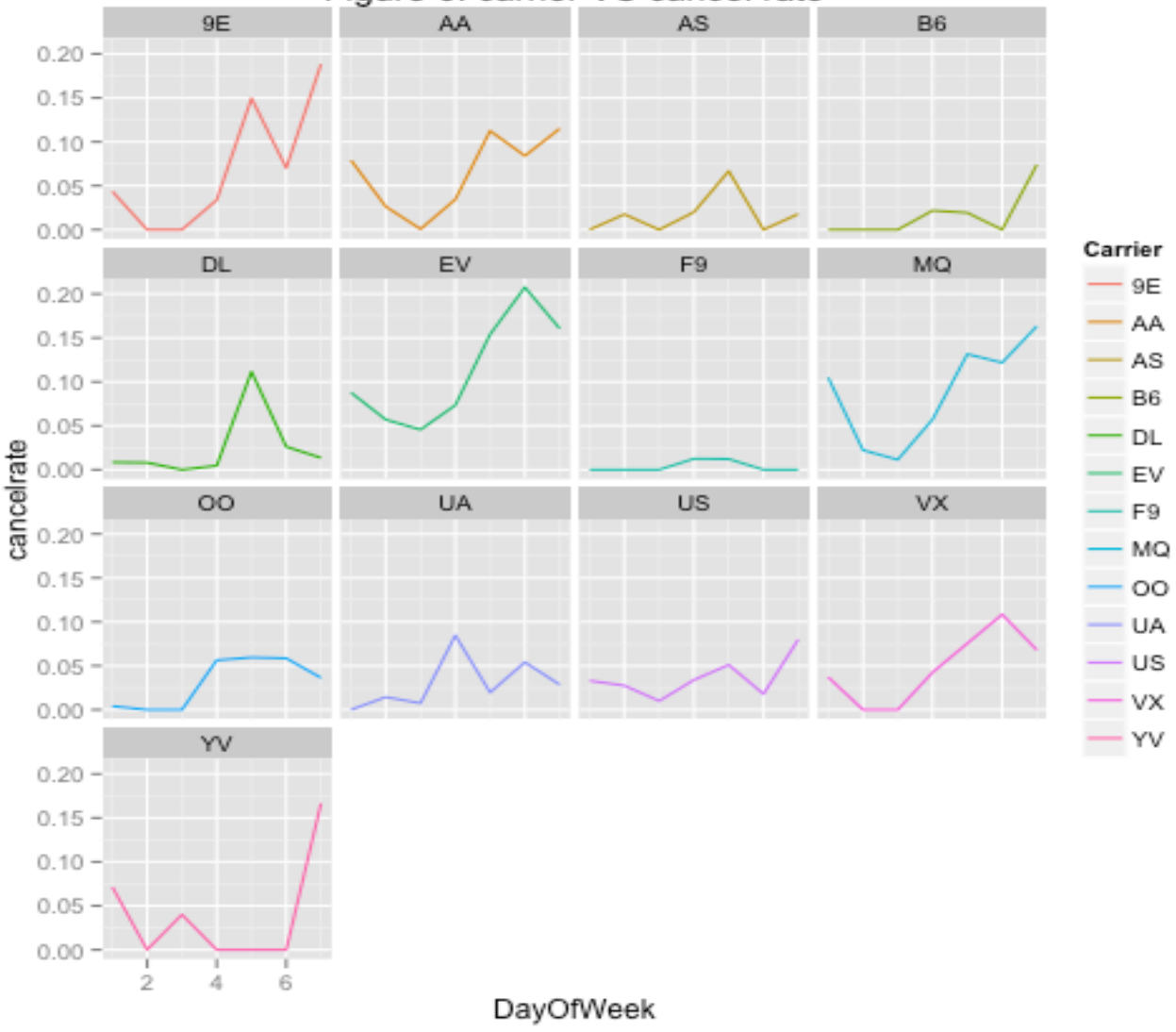
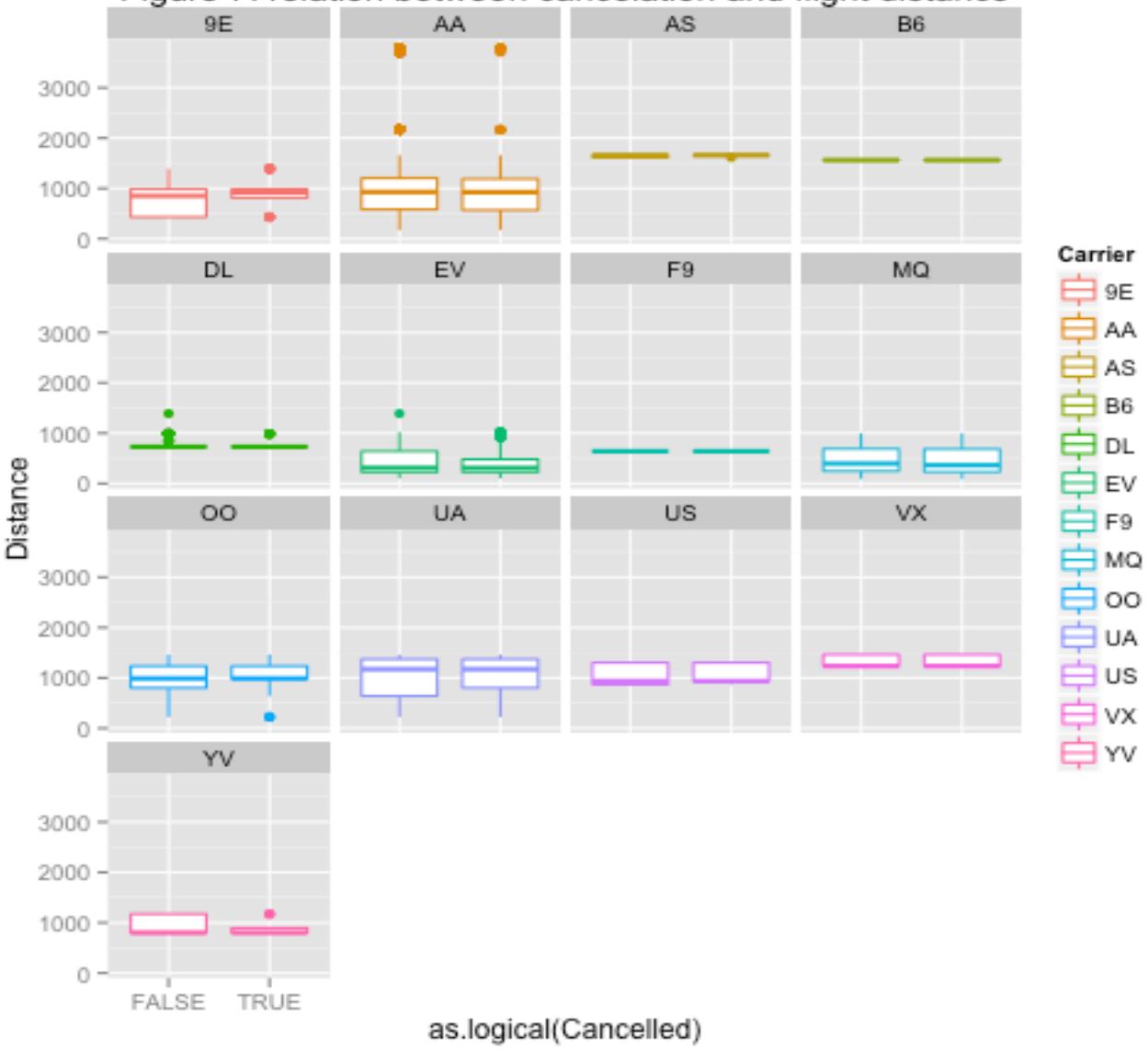
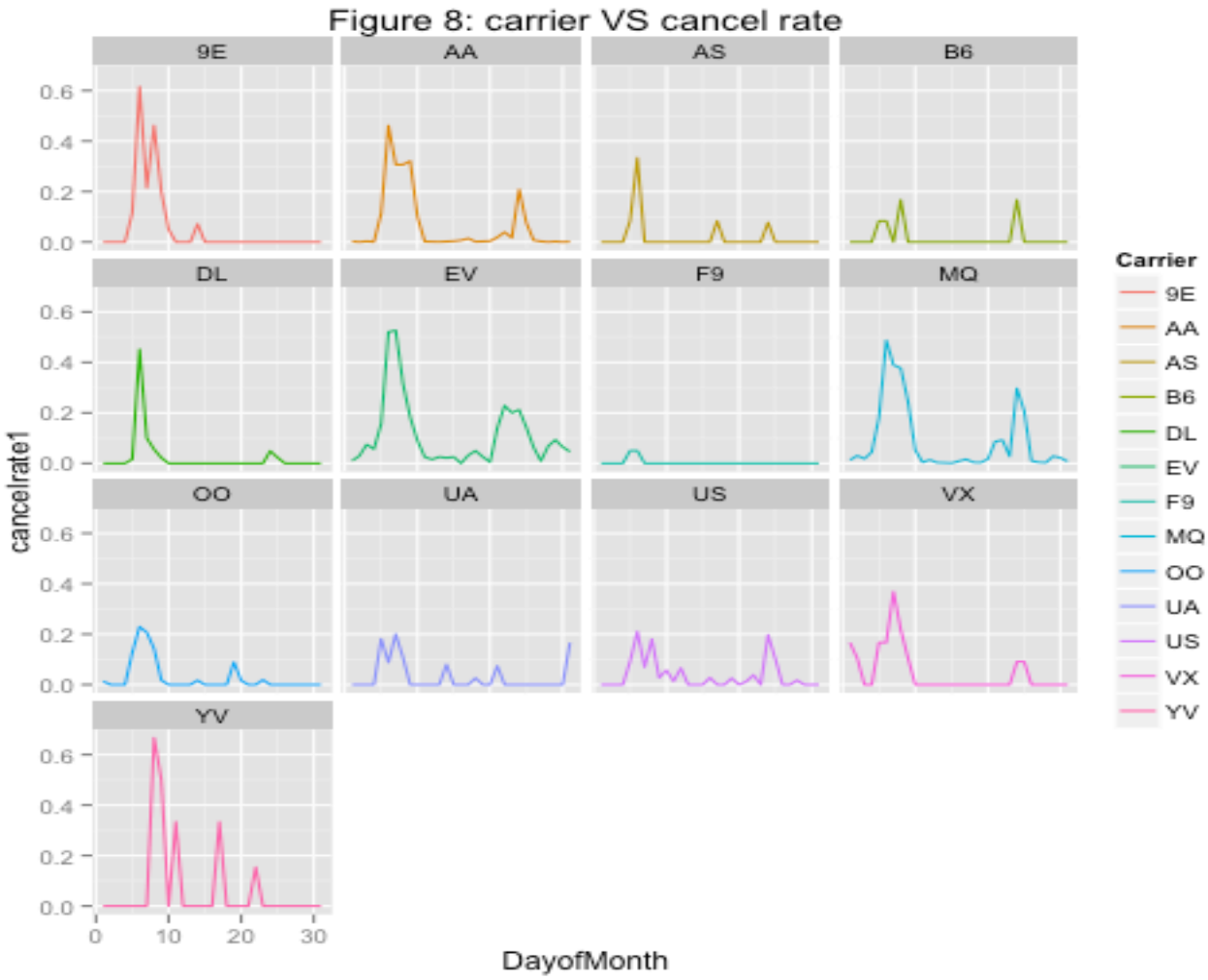


Figure 7: relation between cancelation and flight distance



What's the reason for so many delays and cancelations? Is there a day that has very bad weather?



From the graph, we see that on November 7th or December 7th,there is one day that greatly influenced the cancelation.

After checking the news, It's said that there is a big ice storm on December 6th, and continued on December 7th, and this leads to delay of thousands of flights and nearly a thousand flights canceled.

Part 3: Conclusion:

- Most flights are short distance and middle distance flights which are less 1000 miles, and most flights goes to (or from) nearby big cities which have close economic relation with Dallas.

- Most cancelation rates are from 4% to 10%.

- But the cancelation rate is very high during the weekend (EV, MQ and 9E have cancelation rate more than 15%), that is because there is a ice storm in December 7th, which just happened in Saturday, and lead to cancelation of nearly a thousand flights.

- The average delay time for long distance flights (>1000) is shorter than that of short and middle distance flights.

- Although there are many delays, but most of the delay time for the flights are less than two and half hours.