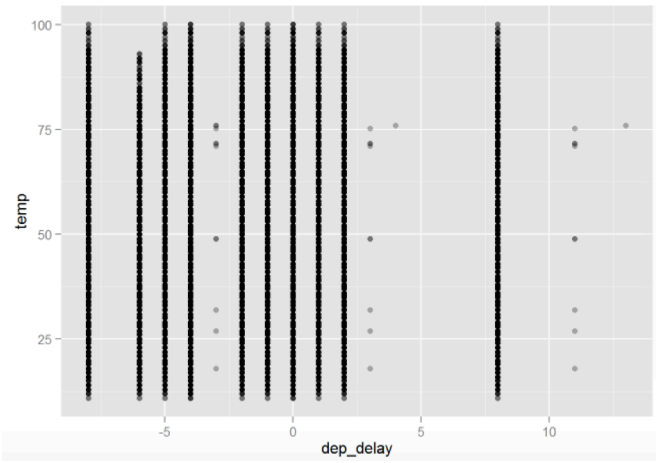
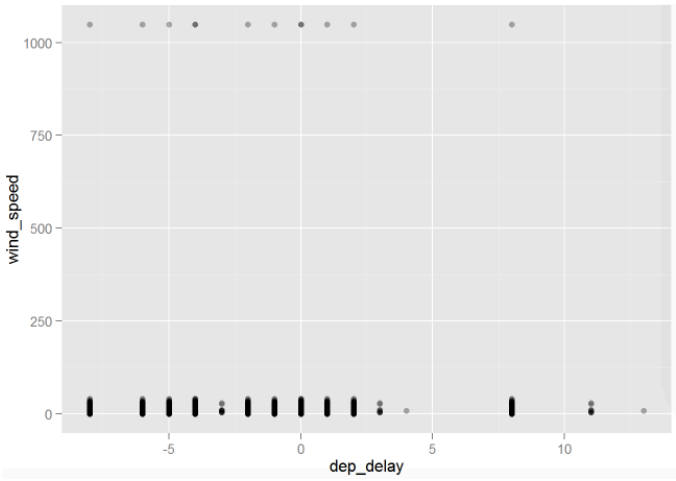
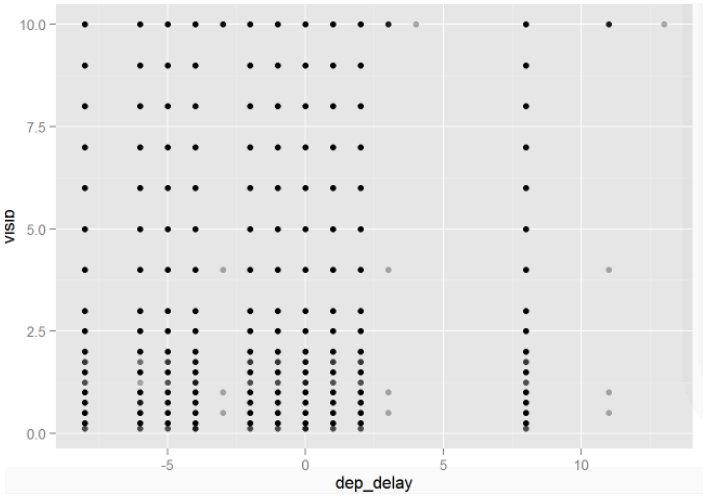
Junge Wu
Final exam

a) weather





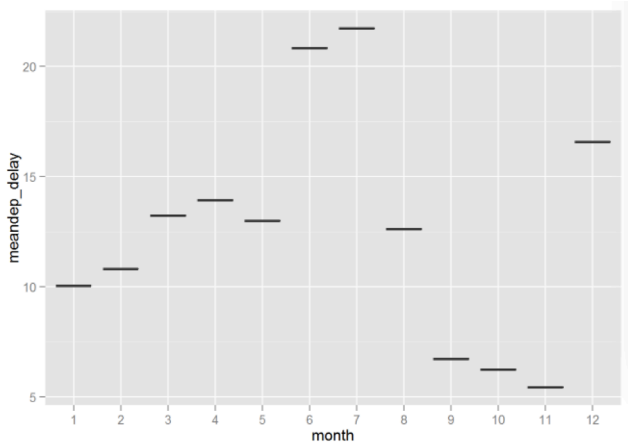


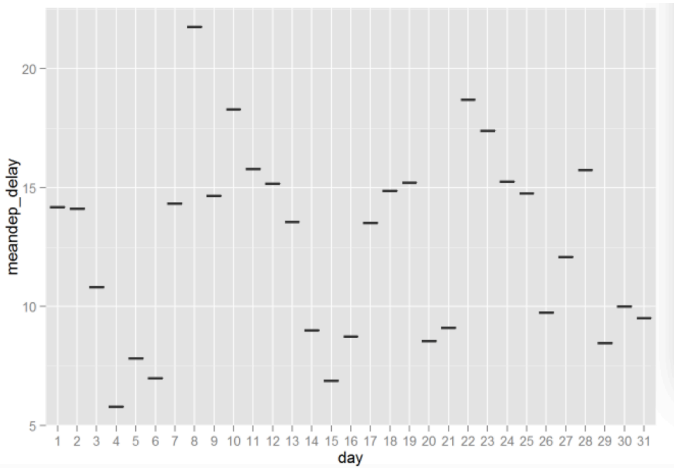

```
with(use_data,
cor(dep_delay,visib))# [1] 0.004420173
cor(na.omit(use_data[,c("dep_delay","temp")]))[1,2]
#[1] 0.01825393
cor(na.omit(use_data[,c("dep_delay","wind_speed")]
))[1,2]# [1] -0.002879336
```

Interpretation

From the scatter plot between departure delay and temp, wind speed and visible variables, it shows there is no obviously relationship between them, and the correlation coefficients also show that the correlation is very weak.

b)

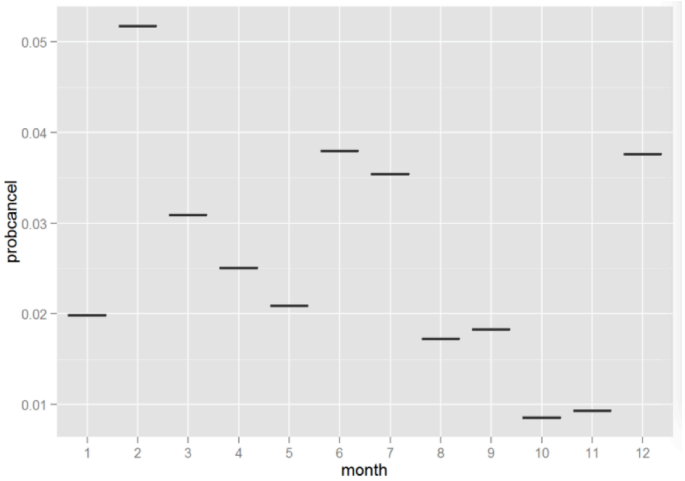


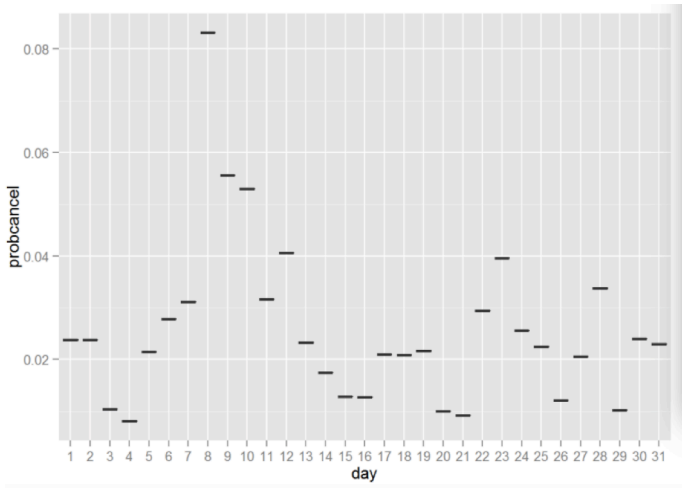


Interpretation:

We can see from the boxplots that some months have much higher average departure delay times such as month 6 and 7, and months like 10, 11 have a much lower average departure

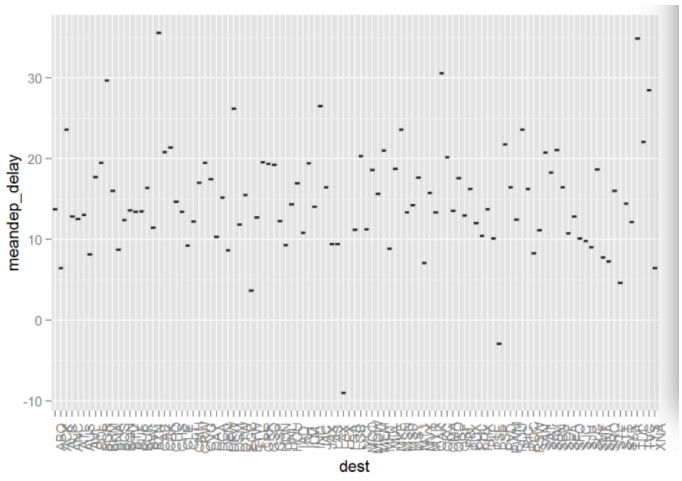
delay times. And from the day in month boxplot, it can be consider there is a correlation between day and departure delay, there appears periodicity, now we continue to see how cancellations correlated with month and day.

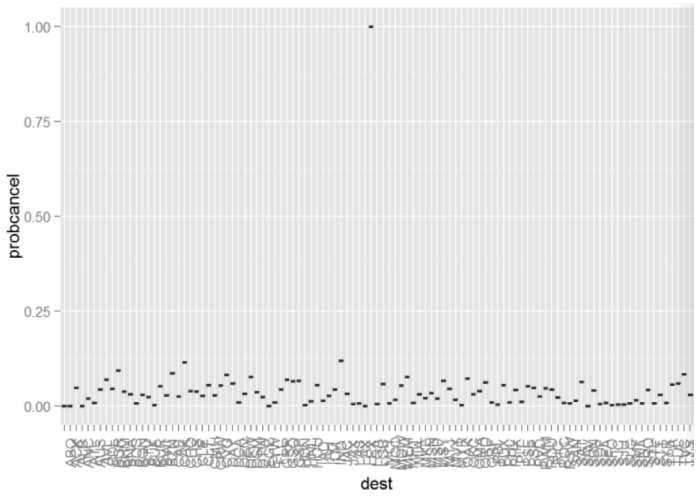




Now it can be seen that month 2 has the highest cancellations probability compared with other months and month 10 has the lowest cancellations probability, recall the average departure delay pattern, we can find that month 10 has lowest average departure delay time and lowest probability of cancellations at the same time.

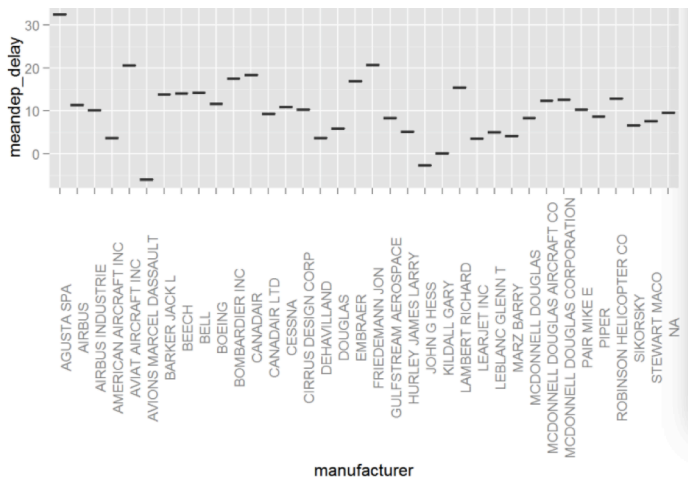
c) airport destination

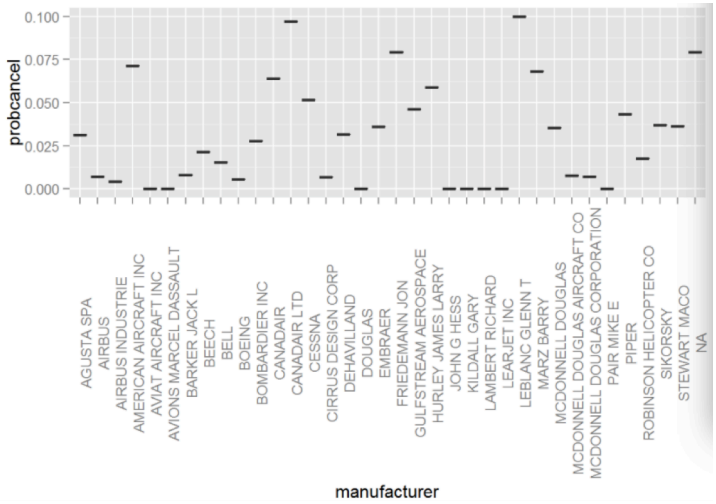




As we could see from the graph above, destinations LAX and SPE have early departure delays compared with others. Also, PSE have obviously early departure delays. After exploring the cancellations and we could find that the probabilities of cancellations are quite low. However, there is one destination called LEX achieves a 1 probability of cancellations, and we may think the data should be missed rather than cancellations

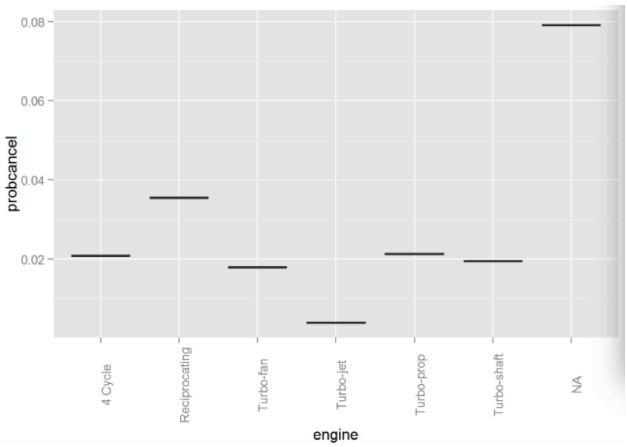
d) characteristics of the plane

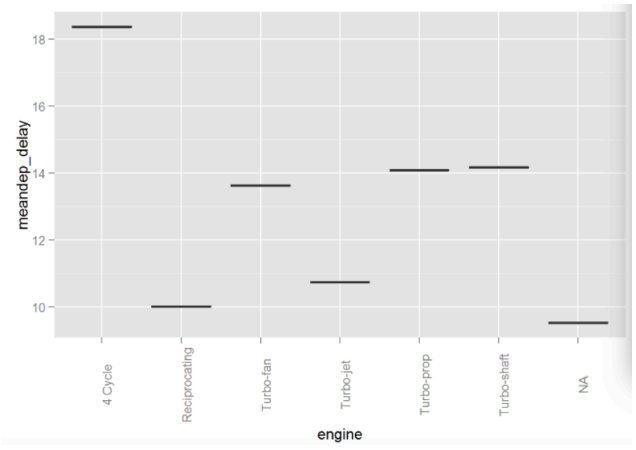




As we could see from the graphs above, it is obviously that different planes from different manufacturer would have different average departure delay time, such as AGUSTA SPA has the most high average departure delay time while AVIONS MARCEL DASSAULT has almost zero average departure delay time, then we continue to explore the cancellations and we could see manufacturer such as LEBLANC GLENN T has the highest probability of cancellations with the value about 0.1, and the AVIONS MARCEL DASSAULT also has the lowest probability of cancellations.
Now we use engine type to explore the relationship:

Compare the plots, it is easily to say that engine type 4 Cycle has the largest average departure delay time but its probability of cancellations is not high.





Finally, we can expect takeoff delays at New York City airports in 2013 that when month is 6 and 7 which also means in summer and Day is around weekend and planes made by AGUSTA SPA with engine type 4 Cycle and cancellations would be except when month is 2 and planes made by LEBLANC GLENNT.