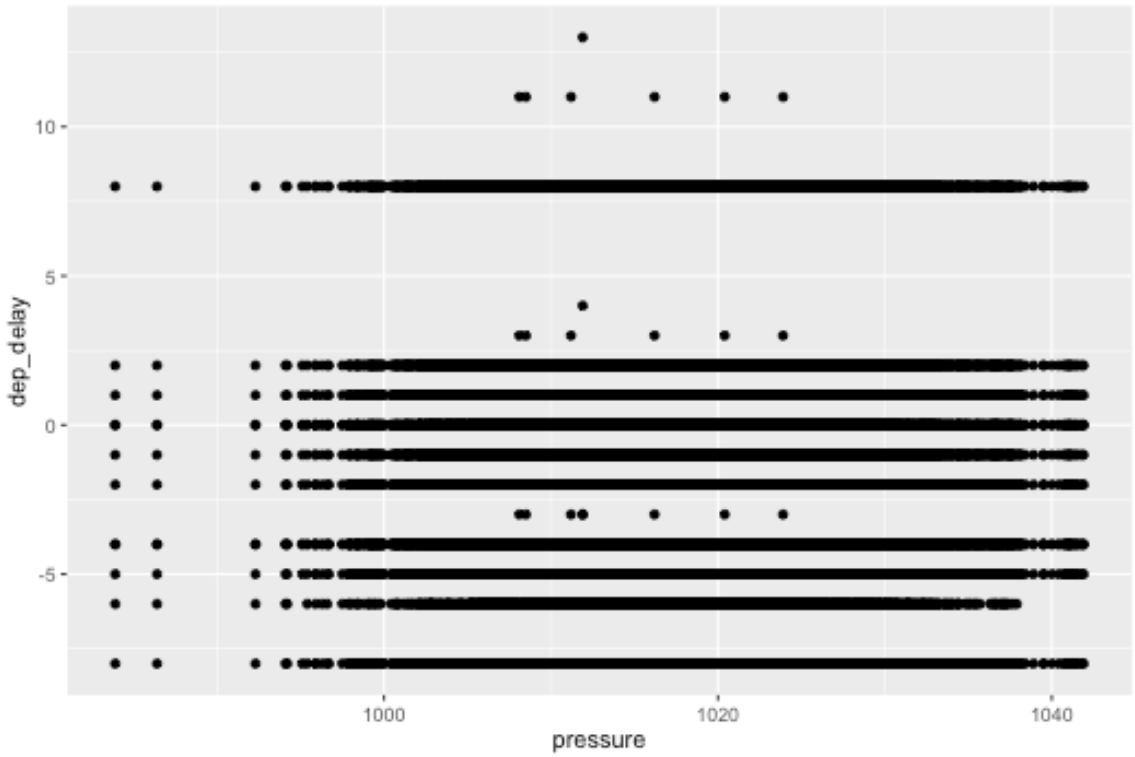
Final
Yanchi Li

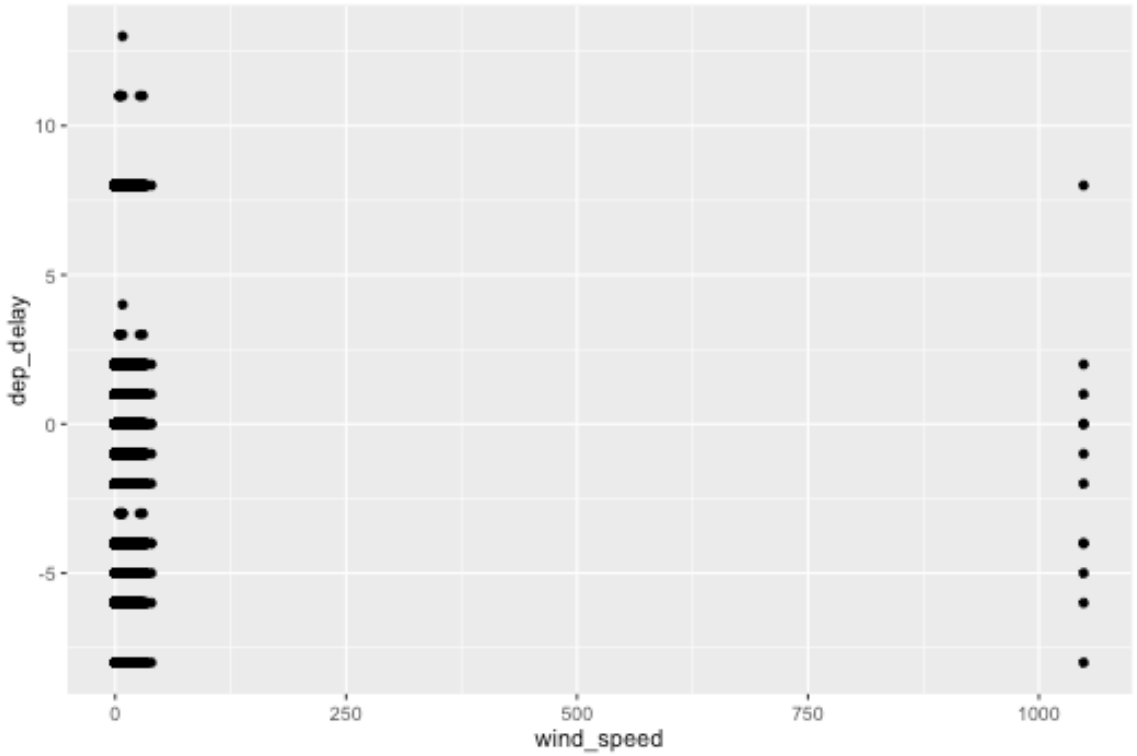
#First, we should install"nycflights13 and RSQlite packages, and install all the packages below to finish the problem, it takes a long time because of my poor internet!

```
library(nycflights13)
library(dplyr)
library(ggplot2)
flights_sqlite<-tbl(nycflights13_sqlite(),"flights")
flights_sqlite

#Source: sqlite 3.8.6
#[/var/folders/77/46r702mj3yq5j2rdzxxsn6w00000gn/T//Rtmpymd
#1QJ/nycflights13.sqlite]
#From: flights [336,776 x 16]

#   year month   day dep_time dep_delay arr_time
#  (int) (int) (int)    (int)     (dbl)    (int)
#1   2013     1     1      517         2      830
#2   2013     1     1      533         4      850
#3   2013     1     1      542         2      923
#4   2013     1     1      544        -1     1004
#5   2013     1     1      554        -6      812
#6   2013     1     1      554        -4      740
#7   2013     1     1      555        -5      913
#8   2013     1     1      557        -3      709
#9   2013     1     1      557        -3      838
#10  2013     1     1      558        -2      753
#..   ...   ...   ...      ...       ...      ...
#Variables not shown: arr_delay (dbl), carrier (chr),
#  tailnum (chr), flight (int), origin (chr), dest
#  (chr), air_time (dbl), distance (dbl), hour (dbl),
#  minute (dbl)

#a)weater
df <- flights_sqlite %>% left_join(weather, by = "origin",
copy = TRUE)
df <- as.data.frame(df)
```

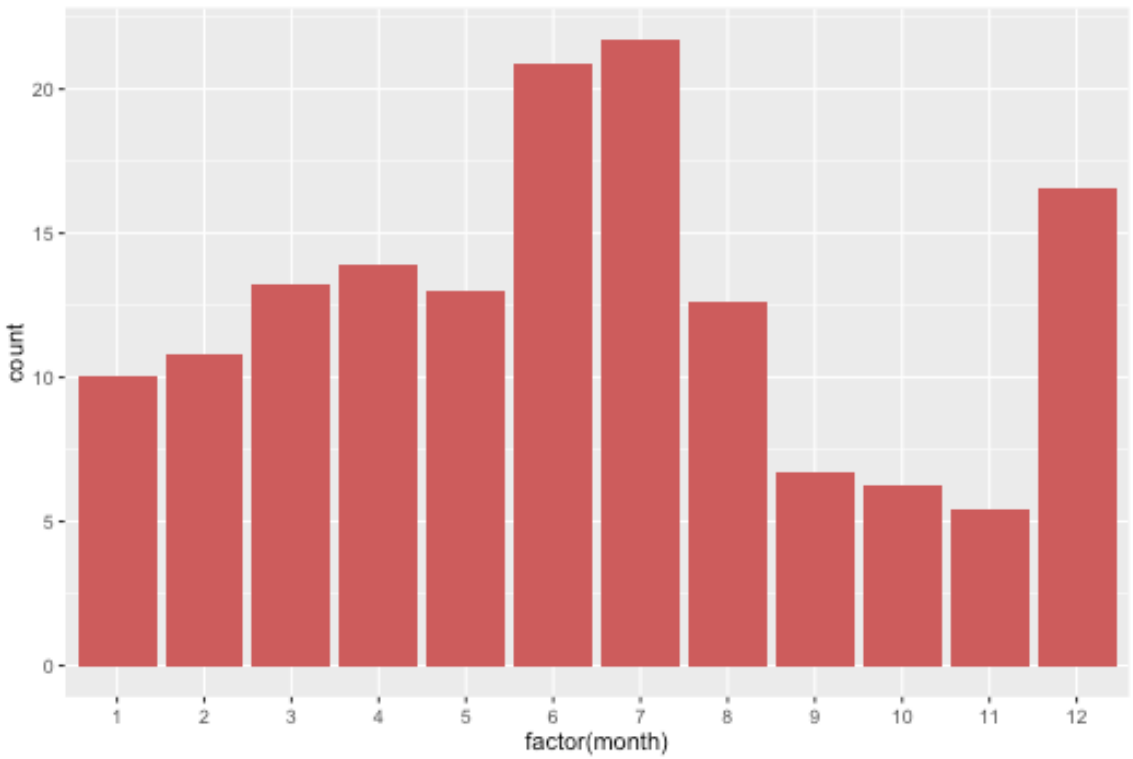
```
p <- ggplot(df, aes(x= pressure,y=dep_delay)) +
geom_point()
p
```



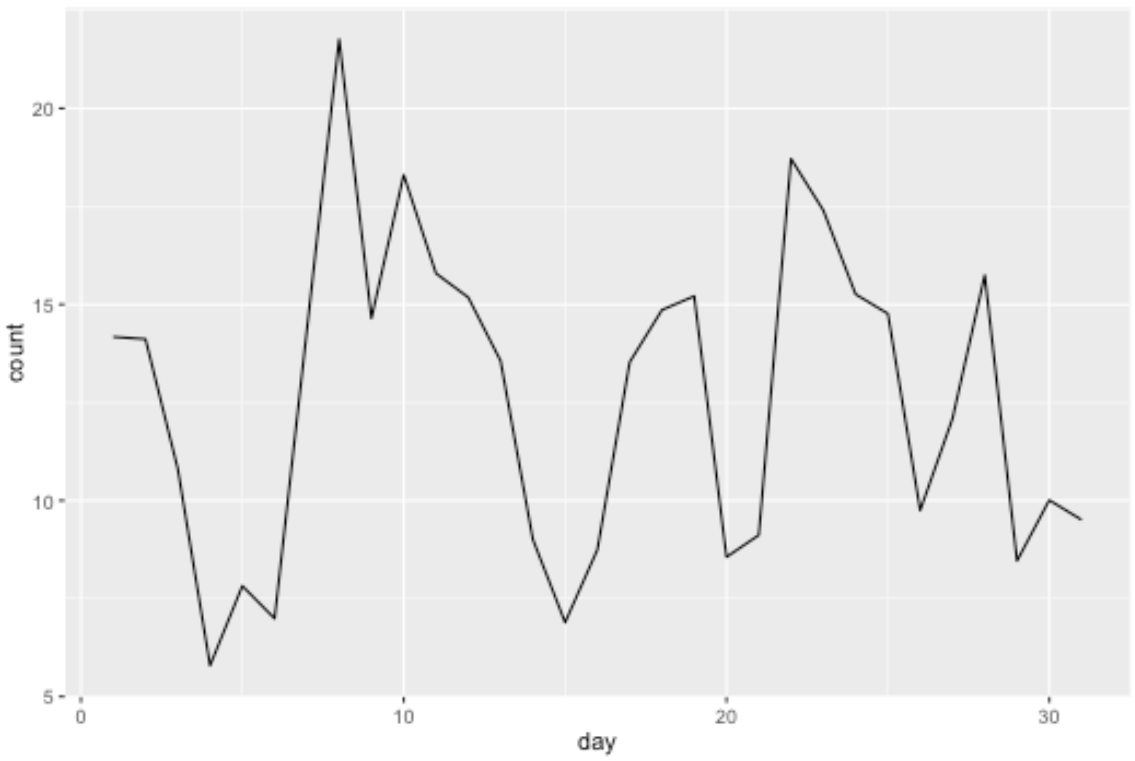

```
# So dep_delay seems not correlated with pressure.
# we next plot wind speed to see if there is any
correlations between wind speed and delay

df <- na.omit(df)
p <- ggplot(df, aes(x= wind_speed,y=dep_delay)) +
geom_point()
p
```



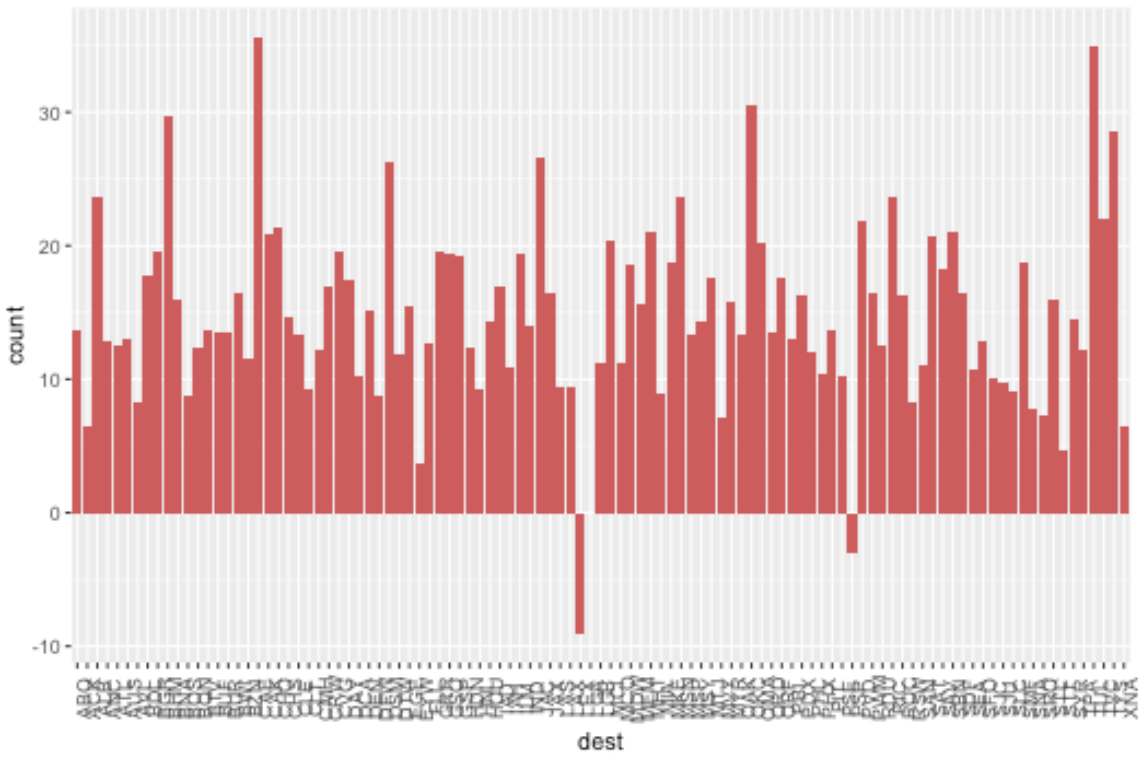

```
# So dep_delay seems also not correlated with wind_speed.

# next, we will test time
#b)time of day, day of week, time of year, and any other
aspect of time

year_bar <- group_by(flights_sqlite, year) %>%
summarise(count = mean(dep_delay))
df <- group_by(flights_sqlite, month) %>%
summarise(count = mean(dep_delay))
df <- as.data.frame(df)
p <- ggplot(df, aes(x= factor(month),y=count)) +
geom_bar(stat="identity", fill = "indianred")
p
```



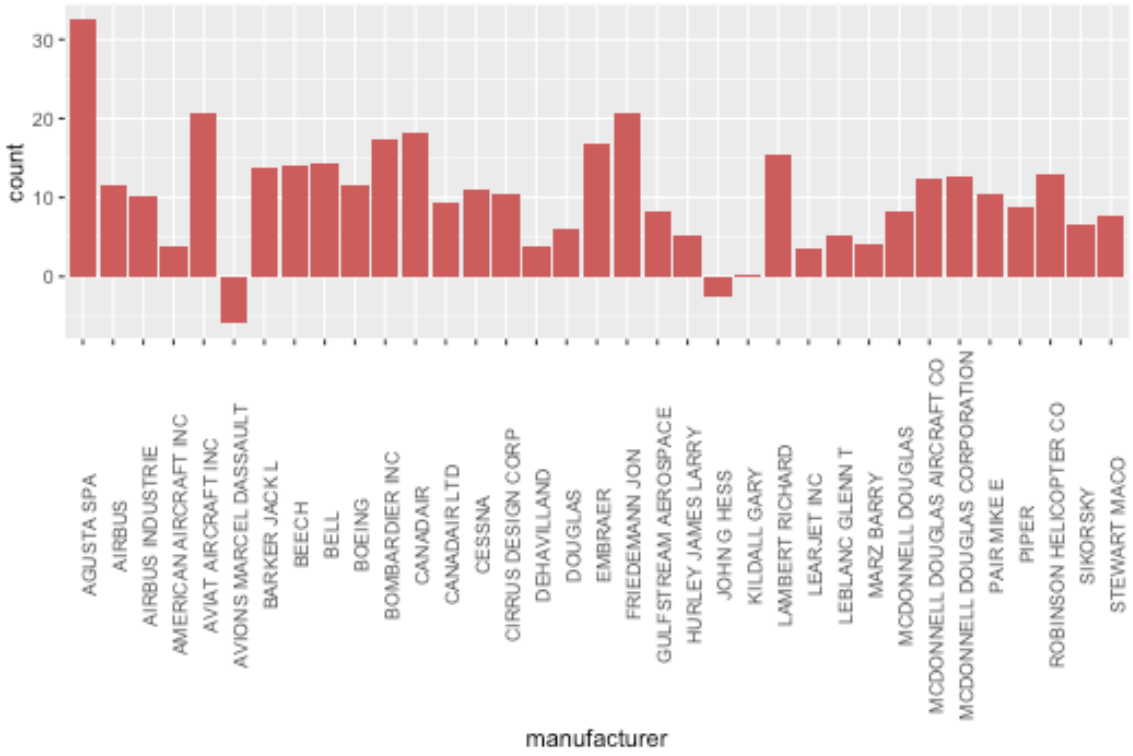

```
df <- group_by(flights_sqlite, day) %>% summarise(count =
mean(dep_delay))
df <- as.data.frame(df)
p <- ggplot(df, aes(x= day,y=count)) + geom_line()
p
```



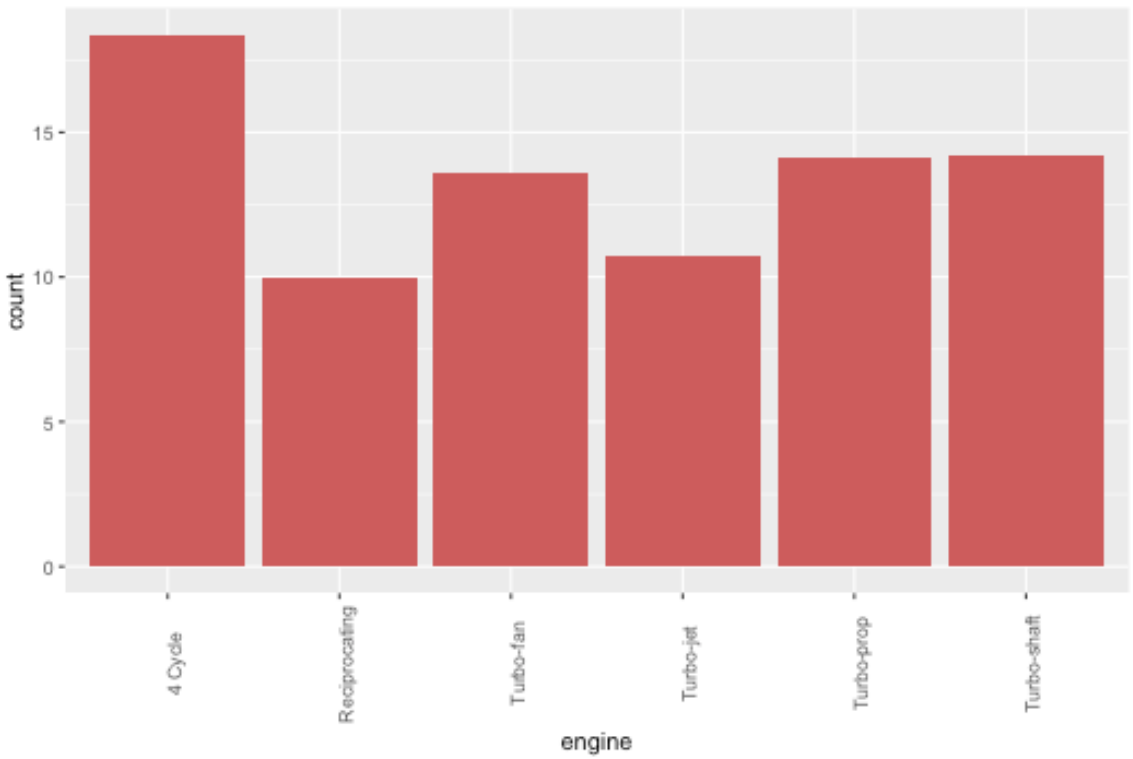

```
# from the picture above, we can see that month 6,7 and 12
#have the most delay, and 9,10 and 11 have the least. Some
#waves also in different days of week.

# c)airport destionation

df <- group_by(flights_sqlite, dest) %>%
summarise(count = mean(dep_delay))
df <- as.data.frame(df)
p <- ggplot(df, aes(x= dest,y=count)) +
geom_bar(stat="identity", fill = "indianred") +
theme(axis.text.x=element_text(angle = 90))
p
```



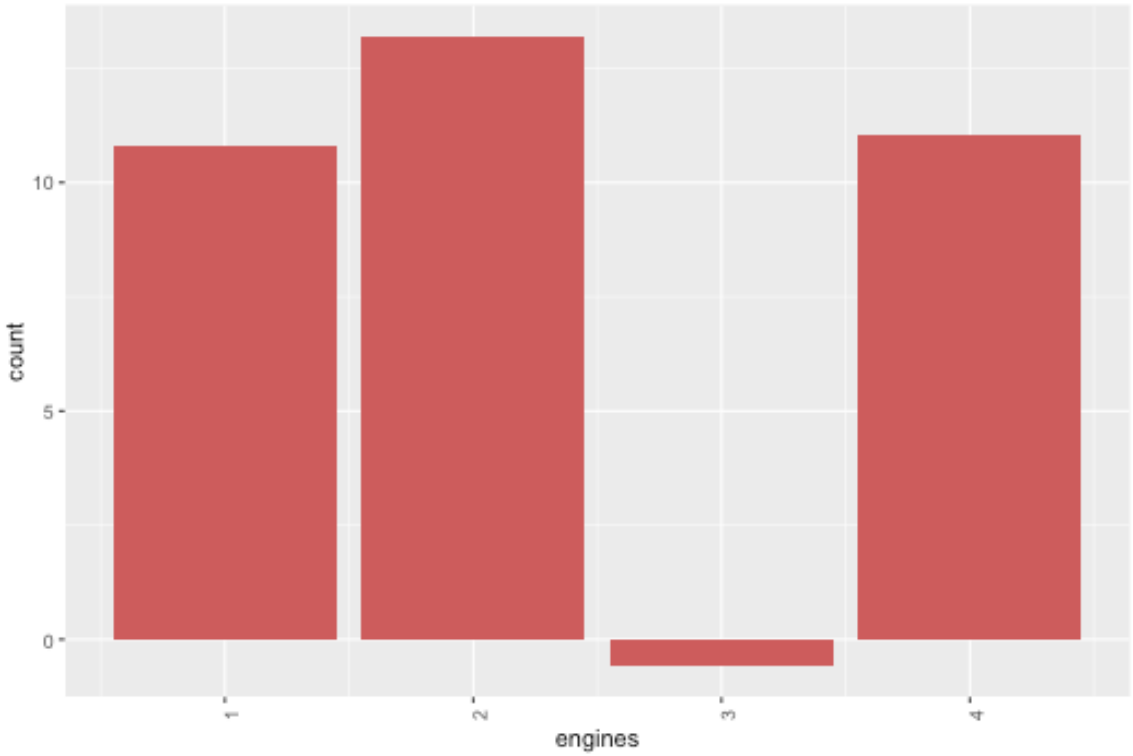

```
# It is obviously can be seen that, there are 2
#destinations LEX #and PSP which have negative mean value
#of dep_delay, it means that these destinations are almost
early every time, especially for LEX, the first negative
#line.

#d)characterisitcs of the plane

df <- flights_sqlite %>% left_join(planes, by =
"tailnum", copy = TRUE)
df2 <- group_by(df, manufacturer) %>% summarise(count =
mean(dep_delay))
df2 <- as.data.frame(df2)
p <- ggplot(df2, aes(x= manufacturer,y=count)) +
geom_bar(stat="identity", fill = "indianred") +
theme(axis.text.x=element_text(angle = 90))
p
```




```
df2 <- group_by(df, engine) %>% summarise(count =
mean(dep_delay))
df2 <- as.data.frame(df2)
p <- ggplot(df2, aes(x= engine,y=count)) +
geom_bar(stat="identity", fill = "indianred") +
theme(axis.text.x=element_text(angle = 90))
p
```




```
df2 <- group_by(df, engines) %>% summarise(count =
mean(dep_delay))
df2 <- as.data.frame(df2)
p <- ggplot(df2, aes(x= engines,y=count)) +
geom_bar(stat="identity", fill = "indianred") +
theme(axis.text.x=element_text(angle = 90))
p
```



from the first pictrure, It can be seen that planes made by AVIONS MARCEL DASSAULT and JOHN G HESS two manufacturers have negtive mean value of dep_delay, it means that these planes are almost early every time, especially for planes made by AVIONS MARCEL DASSAULT.

from the second, planes use engine 4 Cycle are most likely departure delay.

from the last, planes use 3 engines are most likely not deley because they have the minimum mean value of dep_delay.

End.