STA 380 HW2

Jack Cheng

8/15/2017

```r
library(ggplot2)
ABIA = read.csv("~/Documents/STA380/ABIA.csv")
```

First, we did data cleaning and dropped all the missing values in ArrDelay and DepDelay since we only care about the data that are with a status of delaying. Then we converted all months and day of week numbers to factor with names. In order to figure out when is the best time in a year to fly, we just assume that we only care about the flights that origin from Austin.

Here we add the departure delay and the arrival delay together as total dealy.

```r
# Data cleaning drop missing values and convert factors with month and day of week names.
ABIA_cleaned = ABIA[!is.na(ABIA$ArrDelay),][!is.na(ABIA$DepDelay),][ABIA$Origin=='AUS',]
ABIA_cleaned$Month<-factor(ABIA_cleaned$Month,levels=c(1,2,3,4,5,6,7,8,9,10,11,12),labels=c("Jan","Feb"
ABIA_cleaned$DayOfWeek<-factor(ABIA_cleaned$DayOfWeek,levels=c(1,2,3,4,5,6,7),labels=c("Mon","Tue","Wed

# For these delaied flights, we sum the departural and  arrival delay together to get the total delay.

ABIA_cleaned$TotDelay=ABIA_cleaned$ArrDelay+ABIA_cleaned$DepDelay
month_delay=ABIA_cleaned[,c('Month','TotDelay','DayOfWeek')]
month_avg_delay = aggregate(.~Month, data=month_delay, mean, na.rm=TRUE, na.action=NULL)
ggplot(data = month_avg_delay,mapping = aes(x= Month, y= TotDelay )) + geom_col()
```
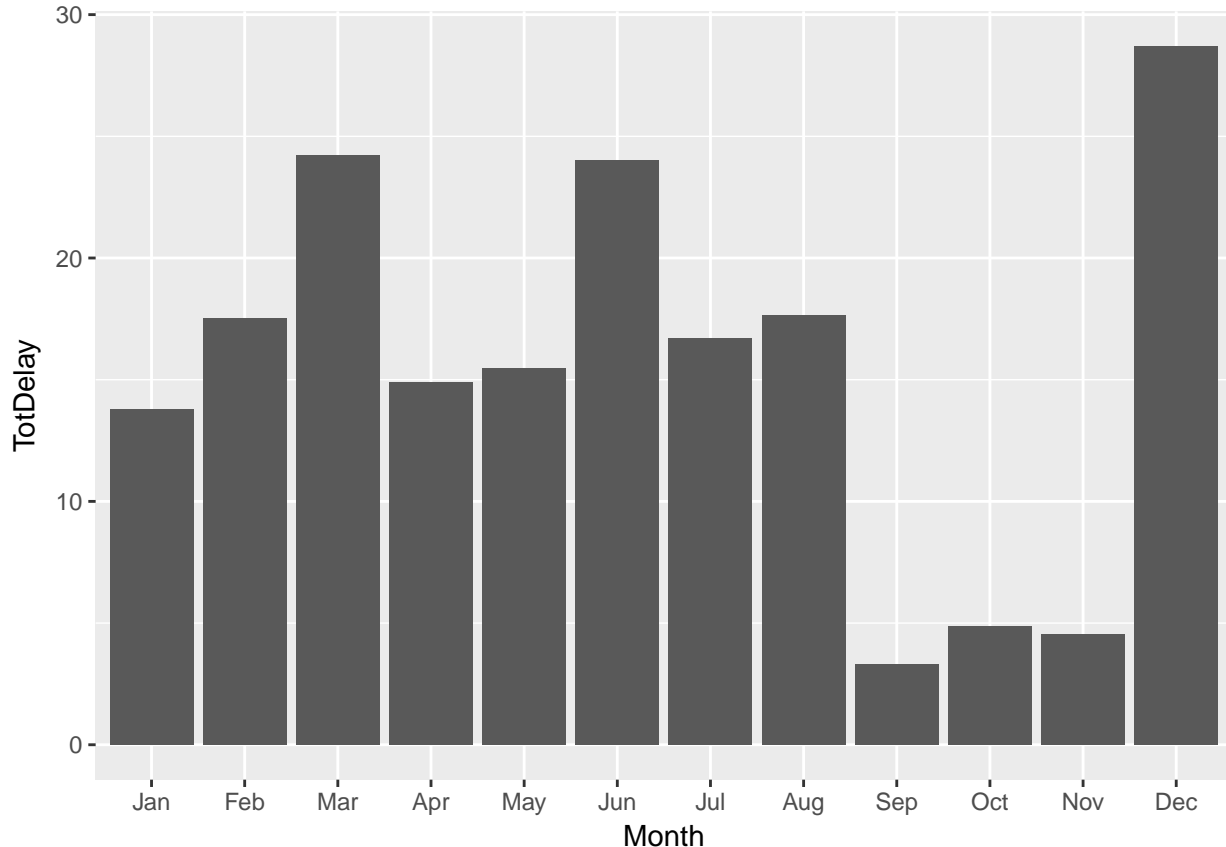
We can see from the graph that for departure and arrival delay, March, June, and December are three months with really high average total delay. We guess the reason is that a lot of students in Austin just seize this chance to go home or travel during spring break, summer holiday and the Christmas.

For the reasons above, we strongly suggest that if you have other options, just arrange your travel in some months like September, October, and November to minimize the delay.

However, in the real case, we are always not able to choose our flying time without limitation. Our team just imagine this is an advice for the students in UT like us so we may face the situation that we need to travel during the holidays. What we can do here is to better arrange the day of the week we leave. So we did a further analysis on average delay in these three months based on different day of a week.

```r
new = aggregate(month_delay$TotDelay, by=list(Month=month_delay$Month, DayOfWeek=month_delay$DayOfWeek)

list= c("Mar","Jun","Dec")

par(mfrow=c(1,3))
for (i in 1:length(list)){
  mon=new[which(new[,1]==list[i]),]
  monthstr=list[i]
  xlab=paste('Day of the Week in',monthstr)
  print(ggplot(data = mon,mapping = aes(x= DayOfWeek, y= x))+geom_col()+labs(x=xlab, y='Average Delay T

}
```
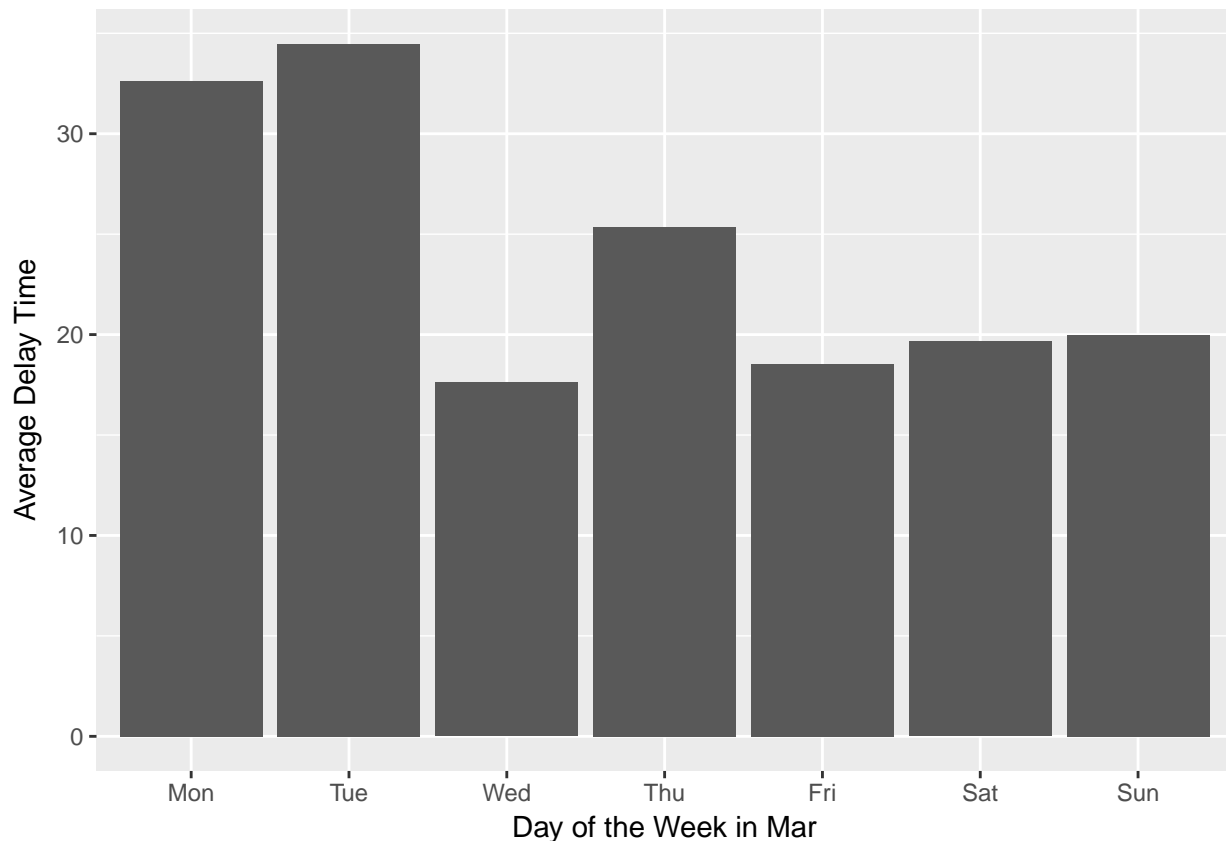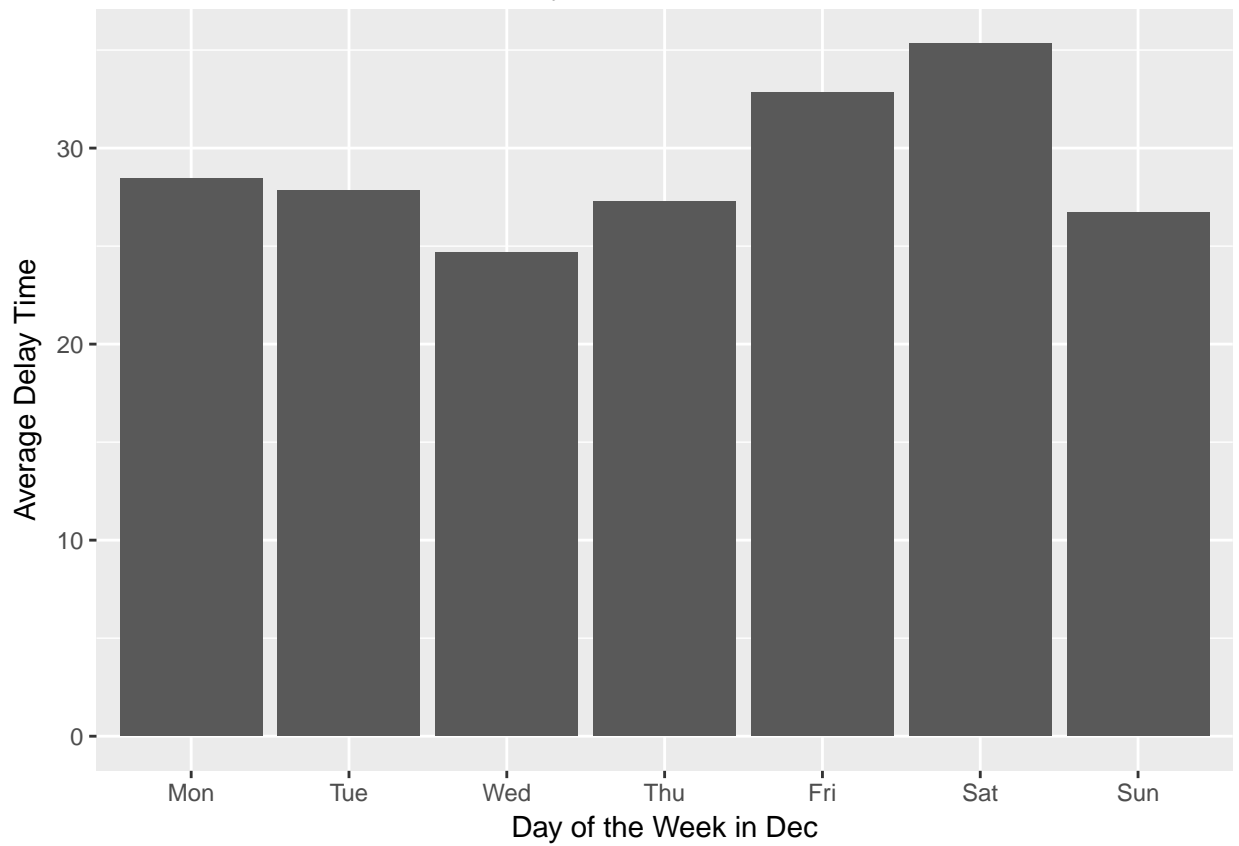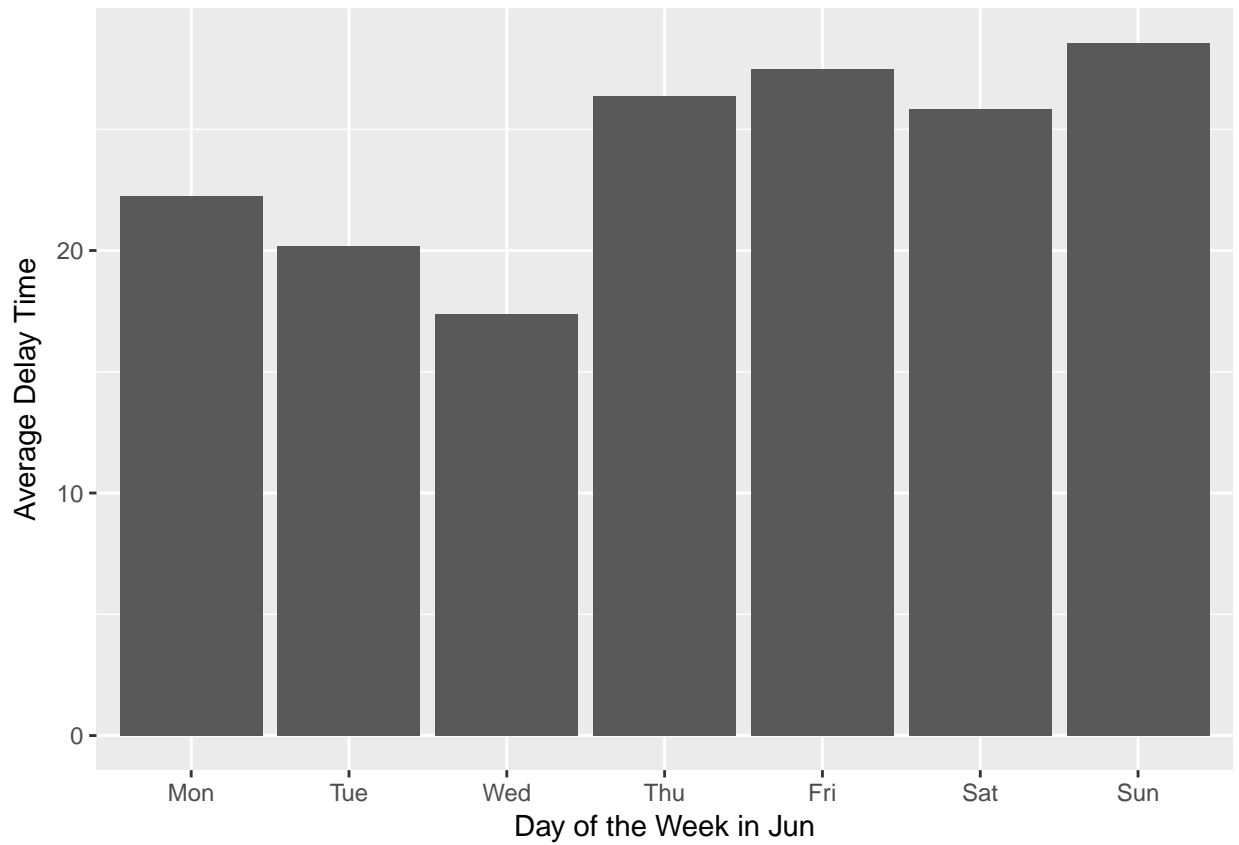
A good choice is to arrange the trip on Wednesday since it has the lowest average delay in the whole week. My team also hold an assumption that for different months, the flights for different distance will be influenced differently. Due to the weather and traffic factors, at certain months it would be easy for long distance (more than 750 miles) to re-accommodate their time by changing their speed. This kind of re-accommodating may require certain height and weather condition that related to month.

```
LongFlight=ABIA_cleaned[which(ABIA_cleaned$Distance>=750),c('Month','TotDelay','DayOfWeek')]
ShortFlight=ABIA_cleaned[which(ABIA_cleaned$Distance<750),c('Month','TotDelay','DayOfWeek')]



LongFlight_month = aggregate(.~Month, data=LongFlight, mean, na.rm=TRUE, na.action=NULL)
ShortFlight_month = aggregate(.~Month, data=ShortFlight, mean, na.rm=TRUE, na.action=NULL)

LongFlight_month$LongFlightSave=LongFlight_month$TotDelay-ShortFlight_month$TotDelay

LongFlight_month[which(LongFlight_month$LongFlightSave>=0),c(1,4)]
```

```
##    Month LongFlightSave
## 1    Jan     8.40689489
## 2    Feb     4.13520322
## 4    Apr     2.66006978
## 6    Jun     6.32394161
## 7    Jul     8.29463168
## 8    Aug     2.39986663
## 9    Sep     0.07839575
## 11   Nov     1.66358393
## 12   Dec     5.85167697
```

So for the months listed above, we recommend taking a long distance trip in order to minimize the delay. As for the rest of the year, a short distance trip would be better.