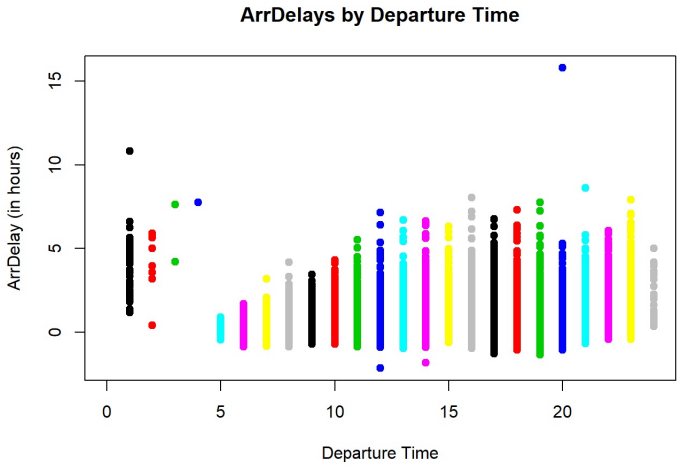
Flights at ABIA

```
library(ggplot2)
par(mfrow=c(2,2))
data=read.csv('C:/Users/rache/OneDrive/Documents/MSBA/Academic/next session/STA380-master/STA380-master/data/ABIA.csv')
data$Austin=ifelse(data$Origin=='AUS',1,0)
```

What is the best time of day to fly to minimize delays?

```
data$hour=round(data$DepTime/100)
plot(data$hour, data$ArrDelay/60, pch=19, col=data$hour, xlab='Departure Time', ylab='ArrDelay (in hours)', main='ArrDelays by Departure Time')
```



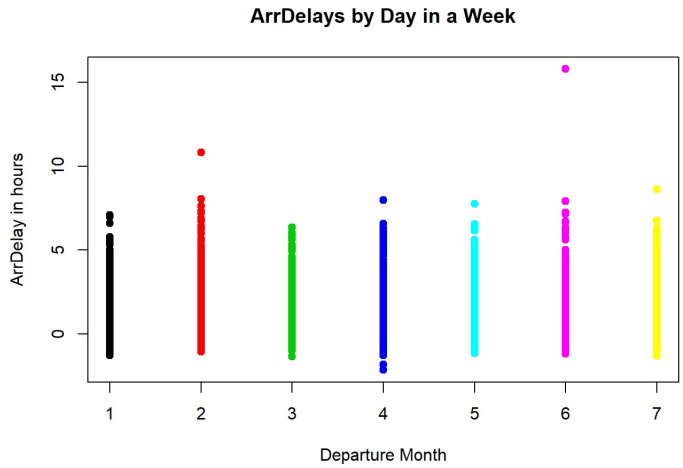

```
median_day=rep(NA,24)
std_day=rep(NA, 24)
for(i in 1:24){
  newsub=subset(data, hour==i)
  median_day[i]<-median(newsub$ArrDelay/60, na.rm=TRUE)
  std_day[i]<-sd(newsub$ArrDelay/60, na.rm=TRUE)

}
choose_day=data.frame(col1=median_day,col2=std_day)
choose_day[
  with(choose_day, order(choose_day$col2),decreasing=FALSE),
]
```

```
##           col1      col2
## 5  -0.01666667 0.1750175
## 6  -0.08333333 0.1908791
## 7  -0.06666667 0.2293273
## 8  -0.06666667 0.2724264
## 9  -0.03333333 0.3566187
## 11 -0.03333333 0.4266867
## 10 -0.03333333 0.4345897
## 12 -0.01666667 0.4677728
## 13 -0.05000000 0.4838855
## 15 -0.03333333 0.5049383
## 14 -0.03333333 0.5275240
## 17 -0.01666667 0.5860976
## 16  0.00000000 0.5875870
## 18  0.03333333 0.6161010
## 19  0.01666667 0.6354975
## 20  0.05000000 0.7367572
## 21  0.16666667 0.8395833
## 22  0.41666667 1.0999831
## 24  1.46666667 1.2421436
## 23  0.26666667 1.4330274
## 1   3.75000000 1.7487295
## 2   5.00000000 1.7973188
## 3   4.21666667 1.9726134
## 4   7.76666667        NA
```

The table above indicates the standard deviation and median of different Arrival Delays in hours in each departuring hour. Here, we could see that 5 am has the lowest standard deviation and low median. ## What is the best day of week to fly to minimize delays?

```
plot(data$DayOfWeek, data$ArrDelay/60, pch=19, col=data$DayOfWeek, xlab='Departure Month', ylab='ArrDelay in hours', main='ArrDelays by Day in a Week')
```



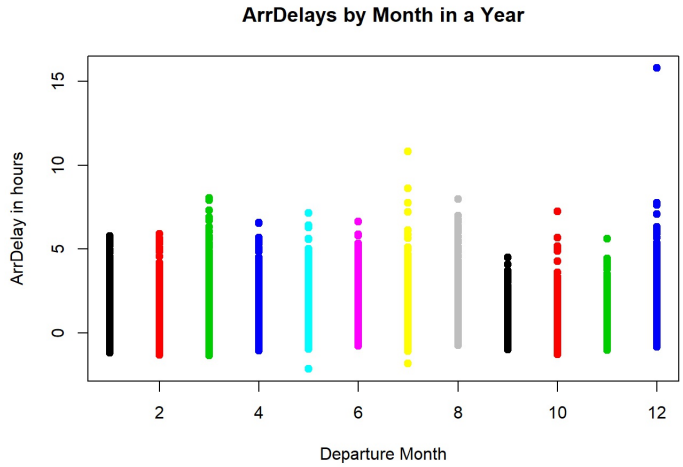

```
median_week=rep(NA,7)
std_week=rep(NA, 7)
for(i in 1:7){
  newsub=subset(data, DayOfWeek==i)
  median_week[i]<-median(newsub$ArrDelay/60, na.rm=TRUE)
  std_week[i]<-sd(newsub$ArrDelay/60, na.rm=TRUE)

}
choose_week=data.frame(col1=median_week,col2=std_week)
choose_week[
  with(choose_week, order(choose_week$col2),decreasing=FALSE),
]
```

```
##          col1      col2
## 3 -0.03333333 0.5035203
## 4 -0.01666667 0.5429525
## 6 -0.05000000 0.5654067
## 1 -0.01666667 0.5684814
## 5  0.00000000 0.5758807
## 2 -0.03333333 0.5830499
## 7 -0.03333333 0.5858247
```

The table above indicates the standard deviation and median of different Arrival Delays in hours in each day of the week. Here, we could see that Wednesday has the lowest standard deviation and low median. ## What is the best month of year to fly to minimize delays?

```
plot(data$Month, data$ArrDelay/60, pch=19, col=data$Month, xlab='Departure Month', ylab='ArrDelay in hours', ma
in='ArrDelays by Month in a Year')
```



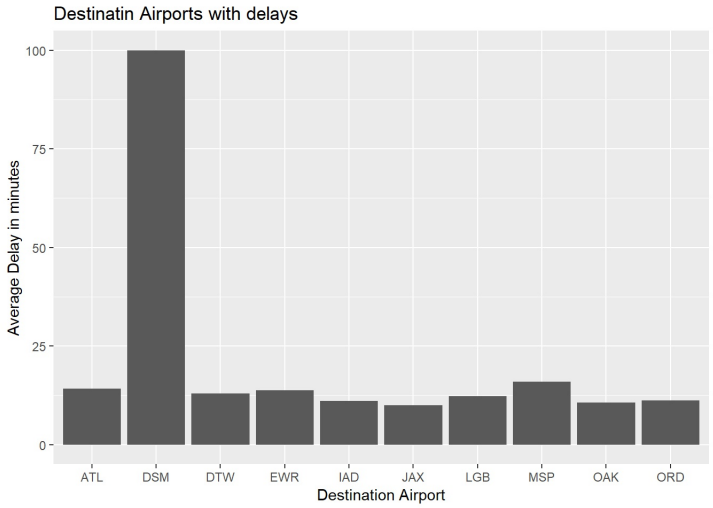

```
median_month=rep(NA,12)
std_month=rep(NA, 12)
for(i in 1:12){
  newsub=subset(data, Month==i)
  median_month[i]<-median(newsub$ArrDelay/60, na.rm=TRUE)
  std_month[i]<-sd(newsub$ArrDelay/60, na.rm=TRUE)

}
choose_month=data.frame(col1=median_month,col2=std_month)
choose_month[
  with(choose_month, order(choose_month$col2),decreasing=FALSE),
]
```

```
##           col1      col2
## 9  -0.06666667 0.3734475
## 10 -0.06666667 0.4066320
## 11 -0.06666667 0.4213849
## 1  -0.03333333 0.5306185
## 5   0.00000000 0.5317505
## 4  -0.01666667 0.5465534
## 2  -0.01666667 0.5549020
## 8  -0.01666667 0.6064537
## 7  -0.03333333 0.6136267
## 6   0.01666667 0.6183460
## 3   0.01666667 0.6417106
## 12  0.00000000 0.7068274
```

The table above indicates the standard deviation and median of different Arrival Delays in hours in each Month. Here, we could see that september has the lowest standard deviation and low median. ## What are the bad airports to fly to?

```
airport=aggregate(data$ArrDelay, by=list(Category=data$Dest), FUN=mean, na.rm='True')
airport=airport[order(airport$x,decreasing = TRUE),]
airport=subset(airport,x>10)
ggplot(aes(x = Category, y = x), data = airport ) +geom_bar(stat = "identity") + xlab("Destination Airport") +
  ylab("Average Delay in minutes") + ggtitle("Destinatin Airports with delays")
```



From the all the plots above, we can see that the best time to fly to minimize delay time is round 5 in the morning on Wednesday during September. \ In addition, we found that the Des Moines International Airport (DSM) airport has the most delay time as a destination airport from Austin Airport.\

How do patterns of flights to different destinations or parts of the country change over the course of the year?

Here, we could take a look at the pattern of the flight (to Austin or from Austin in January, June and December)

```
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

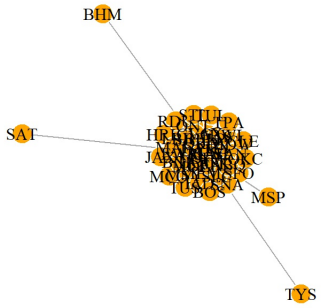
```
index=c(2,17,18)
network_flight=data[index]
```

```
##looking at January
network_1=subset(network_flight, Month==1 )
index_1=c(2,3)
network_1=network_1[index_1]
network_1=as.matrix(network_1)
#####creating a graph object####
graph_network_1=graph.edgelist(network_1, directed=FALSE)
sort(degree(graph_network_1),decreasing = TRUE)
```

```
##  AUS  DAL  DFW  IAH  PHX  ATL  ORD  HOU  DEN  JFK  LAX  ELP  LAS  MEM  MCI
## 8726 1194  906  609  469  421  416  404  392  242  240  234  210  186  179
##  BNA  SJC  CVG  EWR  BWI  MDW  LBB  SAN  MSY  MAF  ONT  ABQ  SLC  IAD  CLE
##  167  164  159  157  124  123  123  122  118  118  117  116  108   93   64
##  TUL  TUS  SFO  SNA  BOS  TPA  JAX  HRL  MCO  STL  OKC  RDU  MSP  SAT  TYS
##   64   62   62   62   62   62   62   62   62   62   61   54   11    1    1
##  BHM
##    1
```

```
#start ploting the network

V(graph_network_1)$color = "orange"
V(graph_network_1)$frame.color = 0
V(graph_network_1)$label.color = "black"
plot(graph_network_1, edge.curved=FALSE)
```

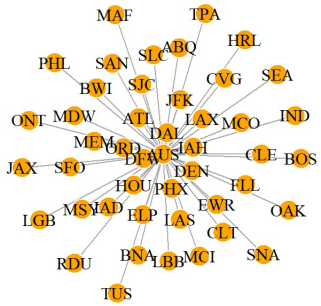



```
##looking at June
network_6=subset(network_flight, Month==6 )
index_1=c(2,3)
network_6=network_6[index_1]
network_6=as.matrix(network_6)
#####creating a graph object####
graph_network_6=graph.edgelist(network_6, directed=FALSE)
sort(degree(graph_network_6),decreasing = TRUE)
```

```
##  AUS  DAL  DFW  IAH  DEN  PHX  ORD  ATL  HOU  LAX  JFK  ELP  LAS  MEM  SJC
## 9090 1149  889  623  493  472  464  402  381  299  236  222  214  180  172
##  IAD  EWR  MCO  BWI  CLT  MDW  SFO  SAN  FLL  LBB  ABQ  MSY  MCI  SLC  BNA
##  170  161  128  120  120  120  120  120  120  112  112  112  111  110  110
##  CVG  CLE  OAK  MAF  ONT  BOS  SNA  TPA  PHL  LGB  TUS  IND  RDU  HRL  SEA
##  109  102   60   60   60   60   60   60   60   60   60   60   60   60   59
##  JAX
##   58
```

```
#start ploting the network

V(graph_network_6)$color = "orange"
V(graph_network_6)$frame.color = 0
V(graph_network_6)$label.color = "black"
plot(graph_network_6, edge.curved=FALSE)
```

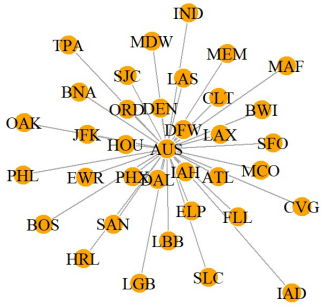



```
##looking at December
network_12=subset(network_flight, Month==12 )
index_1=c(2,3)
network_12=network_12[index_1]
network_12=as.matrix(network_12)
#####creating a graph object####
graph_network_12=graph.edgelist(network_12, directed=FALSE)
sort(degree(graph_network_12),decreasing = TRUE)
```

```
##  AUS  DFW  DAL  IAH  PHX  DEN  HOU  ORD  LAX  ATL  ELP  CLT  LAS  JFK  EWR
## 7248  974  714  622  484  424  377  308  292  276  215  182  179  178  156
##  SJC  MCO  SFO  BWI  SAN  FLL  LBB  BNA  MDW  MEM  BOS  MAF  SLC  OAK  PHL
##  154  130  124  124  124  124  120  116  106   94   63   62   62   62   62
##  CVG  IND  HRL  LGB  TPA  IAD
##   62   62   62   62   62   30
```

```
#start ploting the network

V(graph_network_12)$color = "orange"
V(graph_network_12)$frame.color = 0
V(graph_network_12)$label.color = "black"
plot(graph_network_12, edge.curved=FALSE)
```



From the graph, we can see that all other flights flew frequently from or too Austin except TYS, SAT, MSP, BHM in January, and the top 5 airports having high degree with Austin are DAL, DFW, IAH, PHX, and ATL. The top 5 airports having high degree with Austin in June are DAL, DFW, IAH, DEN, and PHX. The top 5 airports having high degree with Austin in December are DFW, DAL, IAH, PHX, and DEN ```