DonghanLiuHW1

Liu, Donghan, Donghan2

Solution

```
library(RSQLite)
```

Exercise 1

```
# install.packages('data.table')

library(data.table)
delay.con <- dbConnect(RSQLite::SQLite(), dbname = "AirlineDelay1980s.sqlite3")
delays87_89 <- dbGetQuery(delay.con,
                          "SELECT COUNT(*) , DayOfWeek FROM AirlineDelay1980s WHERE Year=1987 OR Year = 198


dayofweek = c('Mon','Tue', 'Wed', 'Thu','Fri', 'Sat', 'Sun')
max = data.frame(DayOfWeek = dayofweek[which(delays87_89['COUNT(*)'] == max(delays87_89['COUNT(*)']))],
row.names(max) = 'Heaviest Traffic Day of Week'
max = data.table(max,keep.rownames = TRUE)
max
```

```
##                              rn DayOfWeek NumOfFlights
## 1: Heaviest Traffic Day of Week       Thu      1685334
```

```
min = data.frame(DayOfWeek = dayofweek[which(delays87_89['COUNT(*)'] == min(delays87_89['COUNT(*)']))],
row.names(min) = 'Lightest Traffic Day of Week'
min = data.table(min,keep.rownames = TRUE)
min
```

```
##                              rn DayOfWeek NumOfFlights
## 1: Lightest Traffic Day of Week       Sat      1539245
```

Exercise 2

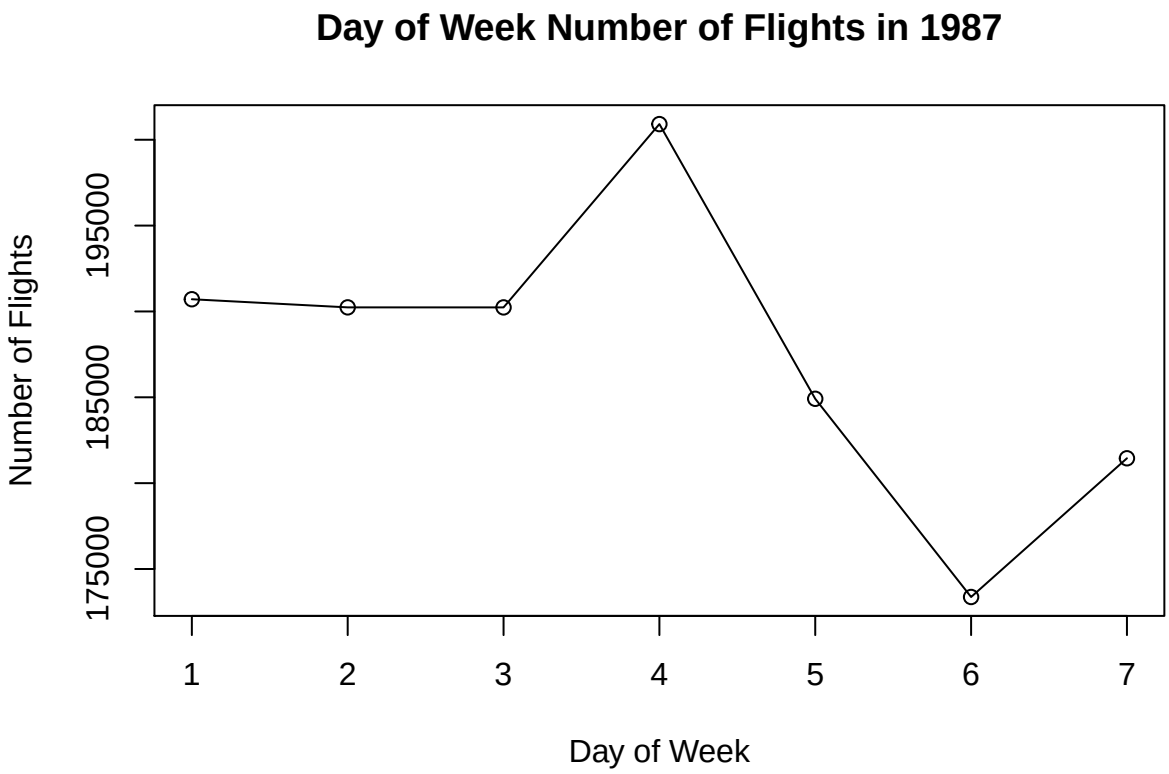
```
##1987 DATA
delays87 <- dbGetQuery(delay.con,
                          "SELECT COUNT(*),DayofWeek FROM AirlineDelay1980s WHERE Year=1987 GROUP BY DayofW

n87 = c(rep(0,7))
for (i in 1:7){
  n87[i] = unlist(delays87[i,1])
}

delays87 = data.frame(delays87,dayofweek)
colnames(delays87) = c('FlightNum87','NumofDayofWeek','DayofWeek')
delays87
```

```
##   FlightNum87 NumofDayofWeek DayofWeek
## 1      190711              1       Mon
## 2      190238              2       Tue
## 3      190235              3       Wed
## 4      200911              4       Thu
## 5      184913              5       Fri
## 6      173370              6       Sat
## 7      181448              7       Sun
```

```
plot(n87, main = 'Day of Week Number of Flights in 1987',xlab = 'Day of Week',ylab = 'Number of Flights
lines(n87)
```




```
# In 1987, based on the plot, Thursday has the highest flight number and Saturday has the lowest flight
# and from Monday to Wednesday, the amount of flights are approximately similar.
```

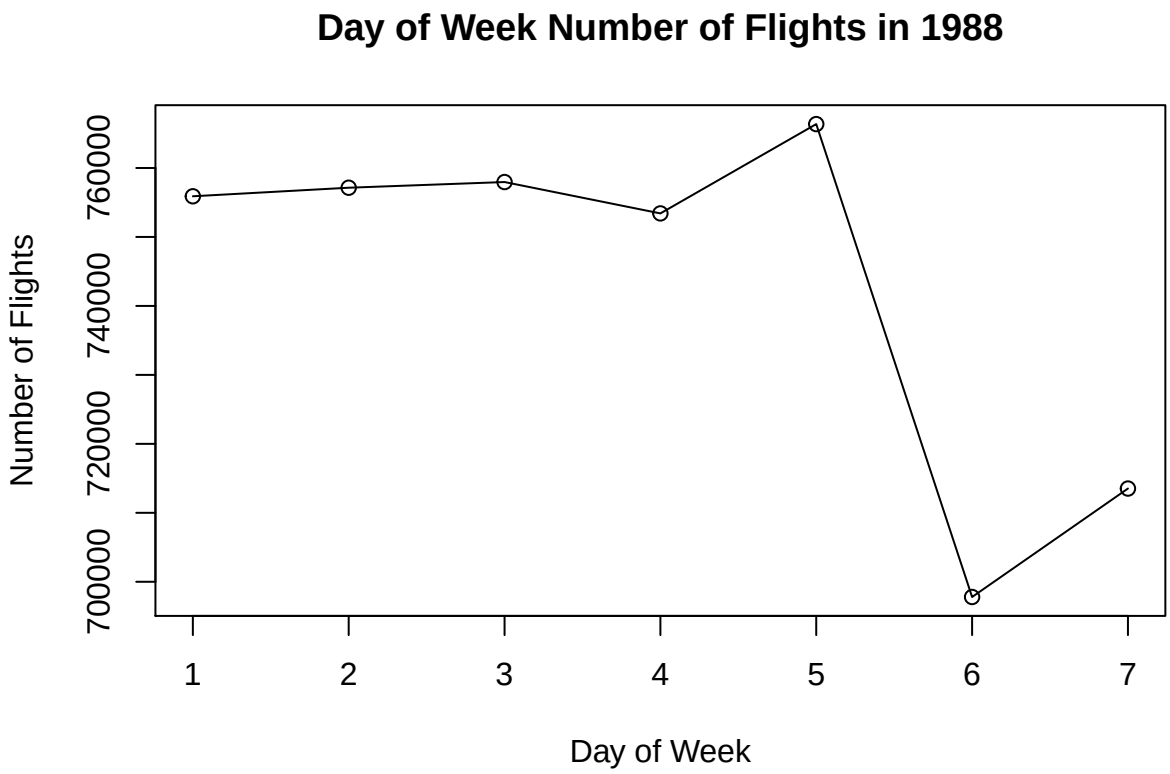
```
##1988 DATA
delays88 <- dbGetQuery(delay.con,
                       "SELECT COUNT(*),DayofWeek FROM AirlineDelay1980s WHERE Year=1988 GROUP BY DayofW

n88 = c(rep(0,7))
for (i in 1:7){
  n88[i] = unlist(delays88[i,1])
}

delays88 = data.frame(delays88,dayofweek)
colnames(delays88) = c('FlightNum88','NumofDayofWeek','DayofWeek')
delays88
```

```
##   FlightNum88 NumofDayofWeek DayofWeek
## 1      755898              1       Mon
## 2      757140              2       Tue
## 3      757963              3       Wed
## 4      753415              4       Thu
## 5      766364              5       Fri
## 6      697795              6       Sat
## 7      713521              7       Sun
```

```
plot(n88, main = 'Day of Week Number of Flights in 1988',xlab = 'Day of Week',ylab = 'Number of Flights
lines(n88)
```




```
# In 1988, the plot could give us an intuitional view about the change of number of flights regarding t
# day of week. From Monday to Thursday, the amount of flights number is approximately keep in the level
# 755000, however, there is a huge decreasing between Friday and Saturday, indicating the people are li
# flight in Friday rather than Saturday. Among the day of week, Friday is the day that has most flights
# Saturday is the least day.
```

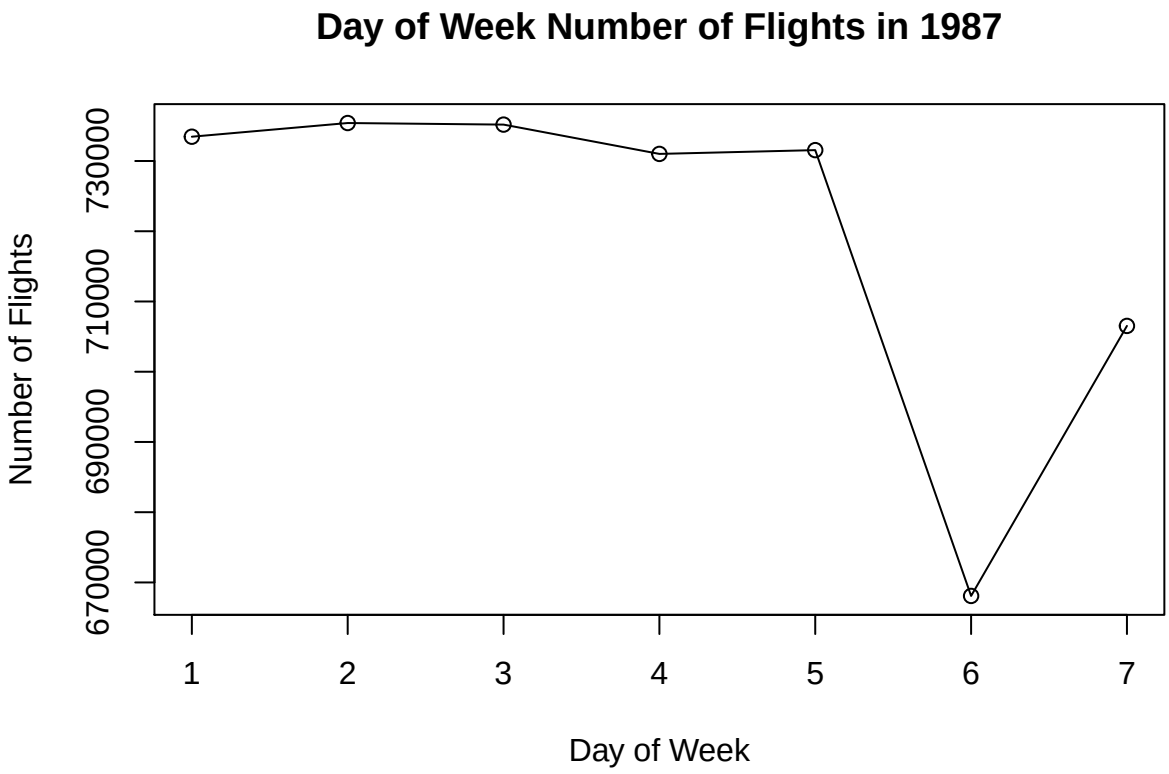
```
##1989 DATA
delays89 <- dbGetQuery(delay.con,
                       "SELECT COUNT(*),DayofWeek FROM AirlineDelay1980s WHERE Year=1989 GROUP BY DayofW

n89 = c(rep(0,7))
for (i in 1:7){
  n89[i] = unlist(delays89[i,1])
}
```

```
delays89 = data.frame(delays89,dayofweek)
colnames(delays89) = c('FlightNum87','NumofDayofWeek','DayofWeek')
delays89
```

```
##   FlightNum87 NumofDayofWeek DayofWeek
## 1      733459              1       Mon
## 2      735404              2       Tue
## 3      735180              3       Wed
## 4      731008              4       Thu
## 5      731548              5       Fri
## 6      668080              6       Sat
## 7      706521              7       Sun
```

```
plot(n89, main = 'Day of Week Number of Flights in 1987',xlab = 'Day of Week',ylab = 'Number of Flights
lines(n89)
```




```
# The day of week number of flights in 1989 has similar situation with 1988's, which is, people are lik
# flight in Friday and Saturday flight seems not a good option for traveler, based on the big gap from
# Saturday. In addition, in the weekday, the number of flights stay the similar amount.
```

Exercise 3

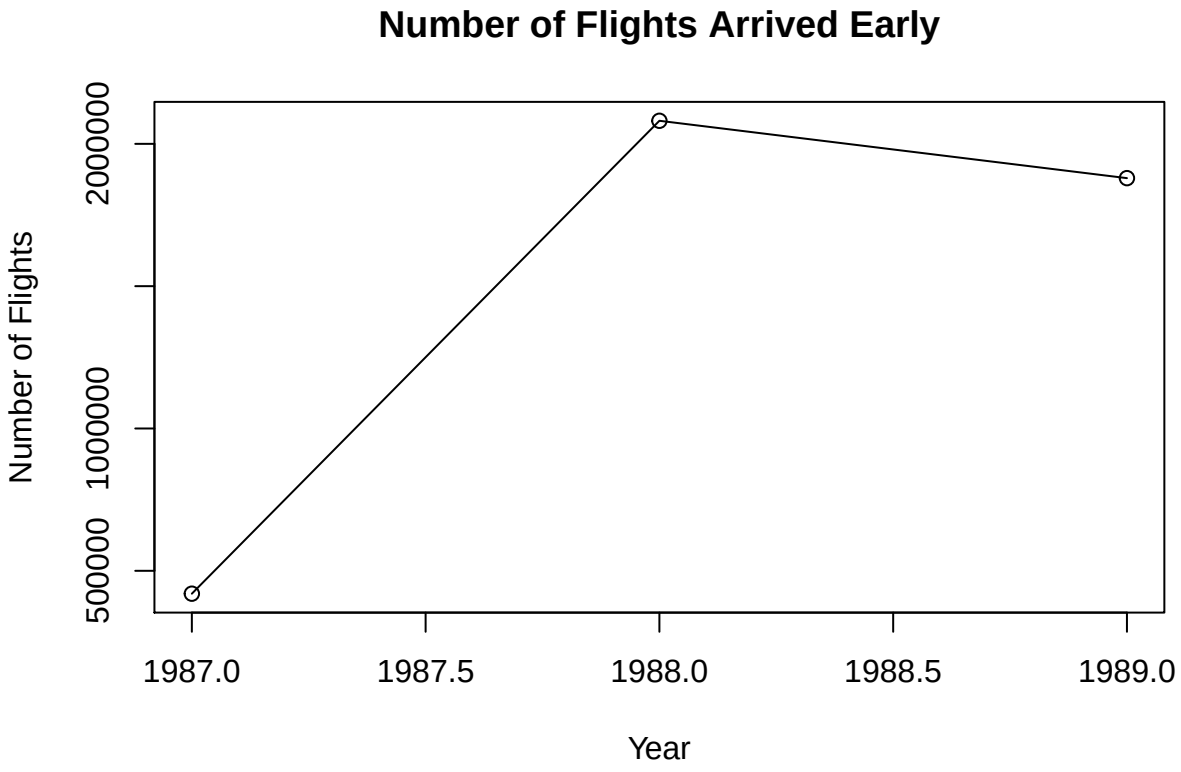
```
ArrEarly <- dbGetQuery(delay.con, "SELECT Count(*),Year FROM AirlineDelay1980s WHERE ArrDelay<0 GROUP B

colnames(ArrEarly) = c('Number of Flights','Year')
```

```
ArrEarly
```

```
##   Number of Flights Year
## 1            419697 1987
## 2           2080990 1988
## 3           1879528 1989
```

```
early = unlist(ArrEarly['Number of Flights'])
plot(x = c(1987,1988,1989),y = early, main = 'Number of Flights Arrived Early',xlab = 'Year',ylab = 'Nu
lines(x = c(1987,1988,1989),y = early)
```




```
# From the comparison of number of flights arrived early in these three years, we could see that there
# increasing from 1987 to 1988 and a slight decline from 1988 - 1989. Among the three years, 1988 is th
# that has the highest number of flights arrived early, whereas, 1987 is the least year.
```

Exercise 4

```
library(biganalytics)
```

```
## Loading required package: bigmemory
## Loading required package: foreach
## Loading required package: biglm
## Loading required package: DBI
```

```
data0708 = attach.big.matrix("air0708.desc")
nrow(data0708)
```

```
## [1] 14462943
```

```
data.frame(NumOfArrEarly = nrow(data0708[data0708[,"Year"] == 2007 & data0708[,"ArrDelay"]<0,]),Year = 2
```

```
##   NumOfArrEarly Year
## 1       3618468 2007
```

```
data.frame(NumOfArrEarly = nrow(data0708[data0708[,"Year"] == 2008 & data0708[,"ArrDelay"]<0,]),Year = 2
```

```
##   NumOfArrEarly Year
## 1       3690606 2008
```

```
# For year of 2007, number of flights that arrived early is 3618468.
# For 2008, there is 3690606 flights arrived early
```