

Airline Delays in the First Course (part 1: accessing the data)

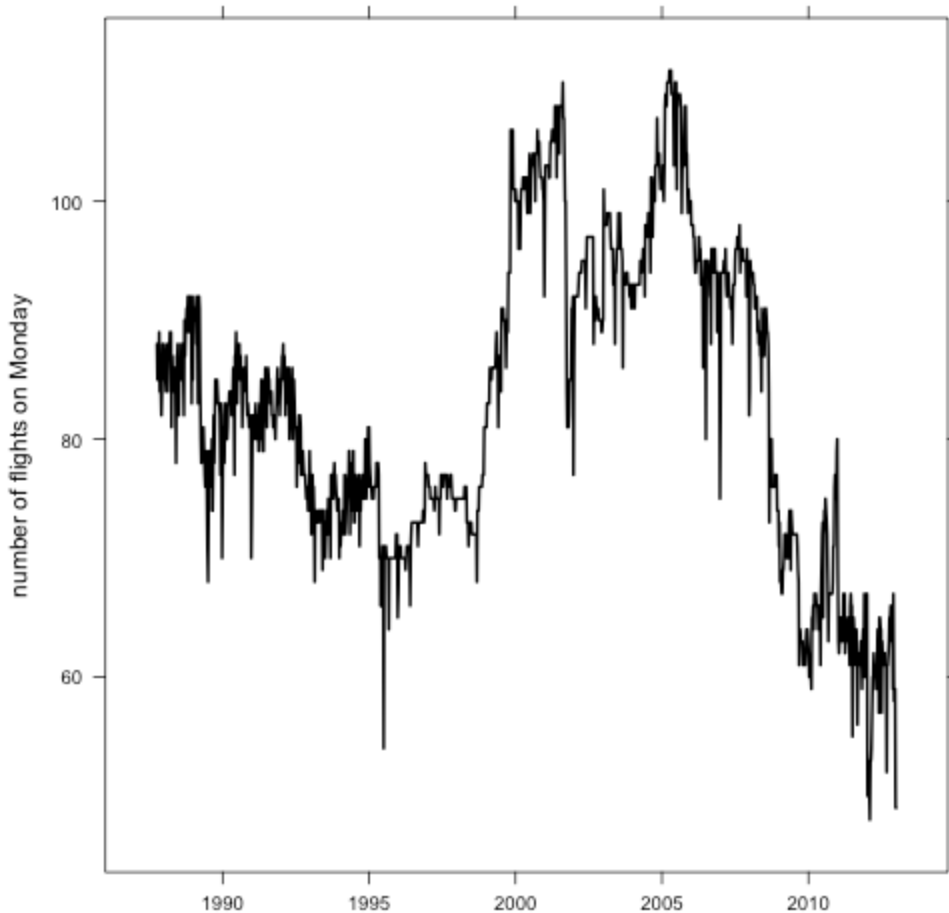
Nicholas Horton, nhorton@amherst.edu,
January 17, 2014

Accessing the data

```
## Loading required package: RMySQL
## Loading required package: DBI
```

```
require(RMySQL)
# also need to provide the password SQLite can be used to
access these data
# without a MySQL server
con = dbConnect(MySQL(), user = "mth292", host =
"rucker.smith.edu", dbname = "airlines")
```

```
ds = dbGetQuery(con, "SELECT DayofMonth, Month, Year,
Origin,\n  sum(1) as numFlights FROM ontime WHERE
Origin='BDL'\n  GROUP BY DayofMonth,Month,Year")
# returns a data frame with 7,763 rows and 5 columns
ds = transform(ds, date = as.Date(paste(Year, "-", Month, "-",
DayofMonth, sep = "")))
ds = transform(ds, weekday = weekdays(date))
ds = ds[order(ds$date), ]
mondays = subset(ds, weekday == "Monday")
library(lattice)
xyplot(numFlights ~ date, xlab = "", ylab = "number of flights
on Monday", type = "l",
  col = "black", lwd = 2, data = mondays)
```



```
ds2 = dbGetQuery(con, "SELECT UniqueCarrier, ArrDelay, Month,
Year, Origin, Dest FROM ontime WHERE Origin='GRB' AND
Dest='ORD' AND Year=2005")
dim(ds2)
```

```
## [1] 2166    6
```

```
head(subset(ds2, UniqueCarrier == "MQ"))
```

```
##   UniqueCarrier ArrDelay Month Year Origin Dest
## 1             MQ         41     1 2005  GRB  ORD
## 2             MQ          3     1 2005  GRB  ORD
## 3             MQ        144     1 2005  GRB  ORD
## 4             MQ          9     1 2005  GRB  ORD
## 5             MQ        168     1 2005  GRB  ORD
## 6             MQ          5     1 2005  GRB  ORD
```

```
head(subset(ds2, UniqueCarrier == "00"))
```

```
##      UniqueCarrier ArrDelay Month Year Origin Dest
## 1365             00         18   10 2005    GRB  ORD
## 1366             00         28   10 2005    GRB  ORD
## 1367             00         44   10 2005    GRB  ORD
## 1368             00          7   10 2005    GRB  ORD
## 1369             00          7   10 2005    GRB  ORD
## 1370             00        -19   10 2005    GRB  ORD
```

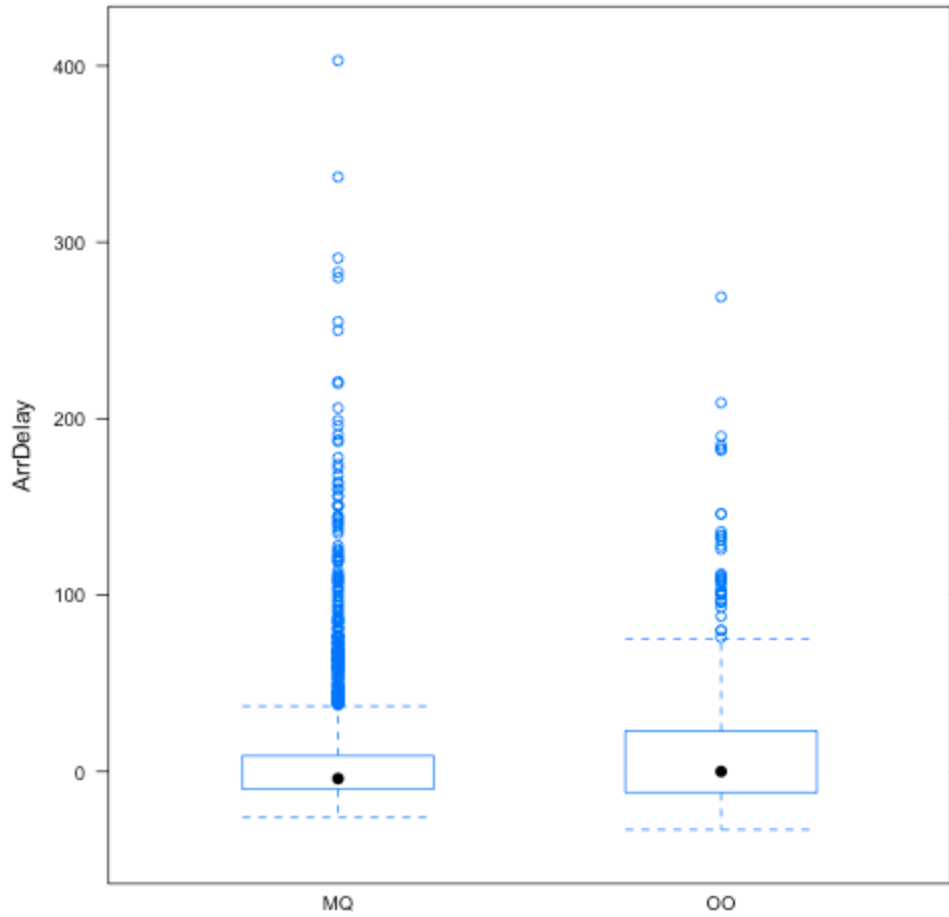
```
tally(~UniqueCarrier, data = ds2)
```

```
## Error: could not find function "tally"
```

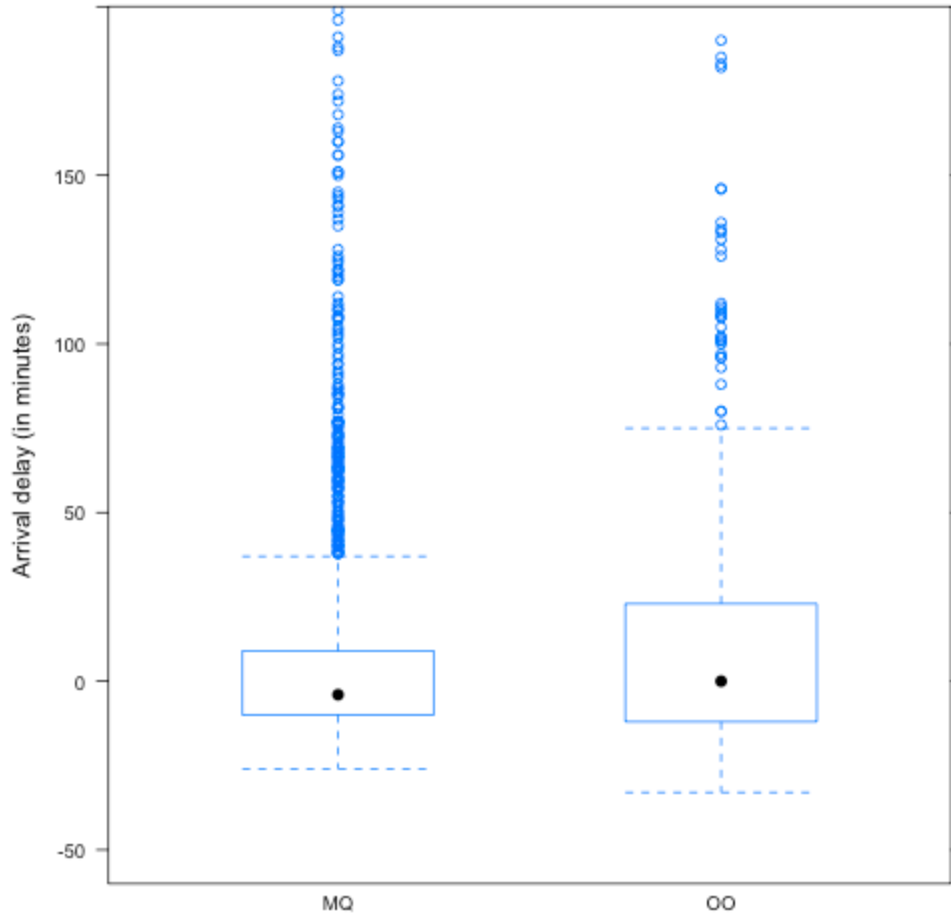
```
favstats(ArrDelay ~ UniqueCarrier, data = ds2)
```

```
## Error: could not find function "favstats"
```

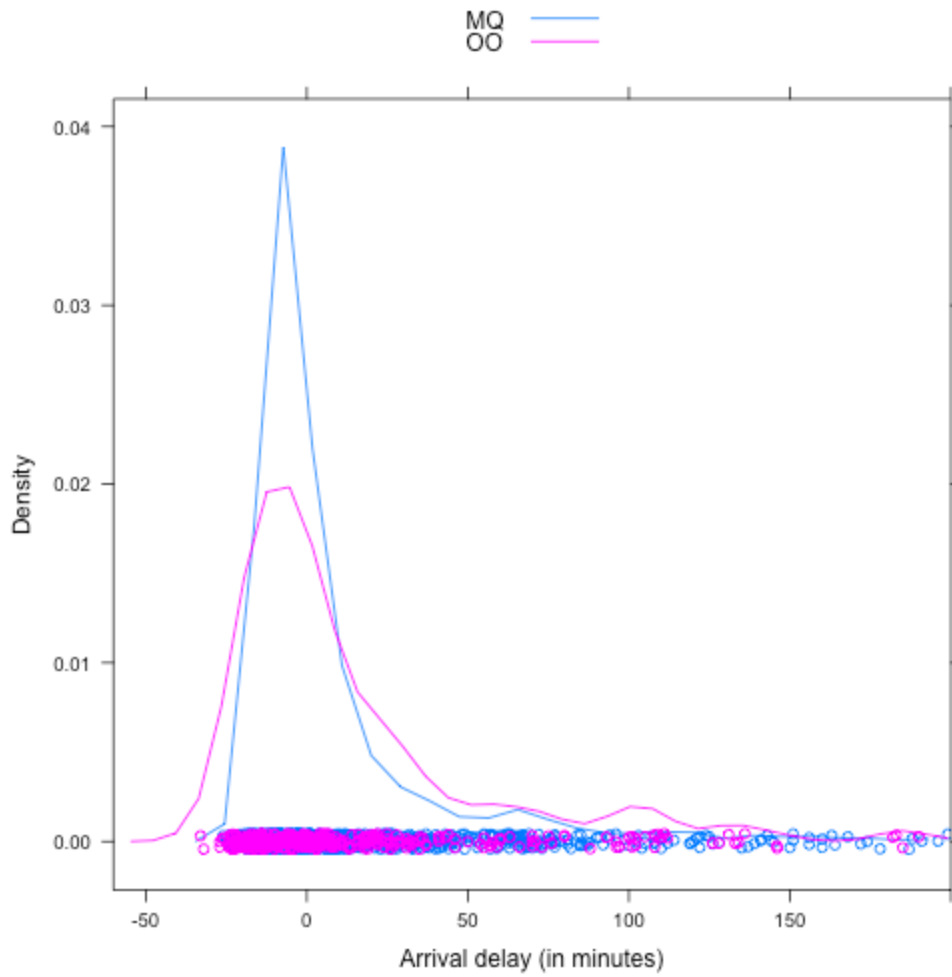
```
bwplot(ArrDelay ~ UniqueCarrier, data = ds2)
```



```
bwplot(ArrDelay ~ UniqueCarrier, ylim = c(-60, 200), ylab =  
"Arrival delay (in minutes)",  
data = ds2)
```



```
densityplot(~ArrDelay, groups = UniqueCarrier, auto.key =  
TRUE, xlab = "Arrival delay (in minutes)",  
xlim = c(-60, 200), data = ds2)
```



```
# save(ds2, file='GRB.Rd')
```