

# Airline Delays in the First Course (part 2: analyzing the data from Boston)

Nicholas Horton, [nhorton@amherst.edu](mailto:nhorton@amherst.edu), July 31, 2014

```
require(mosaic)
```

## Accessing the data

```
## Warning: replacing previous import by 'dplyr::select' when loading  
## 'mosaic'
```

```
options(digits=3)  
trellis.par.set(theme=col.mosaic()) # get a better color scheme for lattice  
load("Meetup-talk/jsm2014/JSM-BOS.rda")  
names(bosFlights)
```

```
## [1] "DayofMonth" "Month" "Year" "Origin"  
## [5] "Dest" "UniqueCarrier" "TailNum" "CRSDepTime"  
## [9] "ArrDelay" "Cancelled"
```

```
oneday = filter(bosFlights, Year==2013 & DayofMonth==8) %>% # Thursday, August 8th, 2013  
  select(DayofMonth, Month, Year, Origin, Dest, CRSDepTime, ArrDelay, Cancelled)  
head(oneday)
```

```
## DayofMonth Month Year Origin Dest CRSDepTime ArrDelay Cancelled  
## 1 8 8 2013 BOS JFK 600 -14 0  
## 2 8 8 2013 BOS MEM 1635 34 0  
## 3 8 8 2013 BOS CVG 630 -6 0  
## 4 8 8 2013 BOS CVG 1040 -12 0  
## 5 8 8 2013 BOS ORF 1905 -17 0  
## 6 8 8 2013 BOS ORF 840 -24 0
```

```
bosFlights = mutate(bosFlights, RealDelay = ifelse(ArrDelay < 0, 0, ArrDelay))  
bosFlights = mutate(bosFlights, TimeOfDay = cut(CRSDepTime, breaks=c(0, 1200, 1800, 2400),  
  labels=c("morning", "afternoon", "evening")))  
bosFlights = mutate(bosFlights,  
  DelayOrCancel = ifelse(is.na(ArrDelay) | ArrDelay > 15, 1, 0))
```

```
favstats(~ DelayOrCancel, data=bosFlights)
```

```
## min Q1 median Q3 max mean sd n missing  
## 0 0 0 0 1 0.243 0.429 29442 0
```

```
favstats(DelayOrCancel ~ TimeOfDay, data=bosFlights)
```

```
##      .group min Q1 median Q3 max mean  sd    n missing
## 1  morning  0  0      0  0   1 0.163 0.369 14130      0
## 2 afternoon 0  0      0  1   1 0.285 0.451 10004      0
## 3  evening  0  0      0  1   1 0.376 0.485  5308      0
```

```
favstats(DelayOrCancel ~ UniqueCarrier, data=bosFlights)
```

```
##      .group min Q1 median Q3 max mean  sd    n missing
## 1      9E  0  0      0 0.0   1 0.115 0.320  182      0
## 2      AA  0  0      0 0.0   1 0.196 0.397 2835      0
## 3      AS  0  0      0 0.0   1 0.123 0.329  334      0
## 4      B6  0  0      0 1.0   1 0.312 0.463 9275      0
## 5      CO  0  0      0 1.0   1 0.386 0.487  515      0
## 6      DL  0  0      0 0.0   1 0.190 0.392 3145      0
## 7      EV  0  0      0 0.0   1 0.188 0.391  538      0
## 8      F9  0  0      0 0.0   1 0.140 0.349   86      0
## 9      FL  0  0      0 0.0   1 0.208 0.406 1487      0
## 10     MQ  0  0      0 1.0   1 0.365 0.482  356      0
## 11     OO  0  0      0 0.5   1 0.267 0.458   15      0
## 12     UA  0  0      0 0.0   1 0.220 0.414 3175      0
## 13     US  0  0      0 0.0   1 0.200 0.400 4903      0
## 14     VX  0  0      0 0.0   1 0.210 0.408  390      0
## 15     WN  0  0      0 0.0   1 0.239 0.427 2023      0
## 16     XE  0  0      0 0.0   1 0.150 0.360   60      0
## 17     YV  0  0      0 0.5   1 0.252 0.436  123      0
```

```
favstats(DelayOrCancel ~ Dest, data=bosFlights)
```

```
##      .group min Q1 median Q3 max mean  sd    n missing
## 1     ACK  0  0      0 0.00   1 0.210 0.410   62      0
## 2     ATL  0  0      0 0.00   1 0.219 0.414 1519      0
## 3     AUS  0  0      0 0.00   1 0.194 0.397   93      0
## 4     BNA  0  0      0 0.00   1 0.139 0.348   79      0
## 5     BUF  0  0      0 0.00   1 0.245 0.430  433      0
## 6     BWI  0  0      0 1.00   1 0.252 0.434 1616      0
## 7     CAK  0  0      0 0.00   1 0.242 0.430  161      0
## 8     CHS  0  0      0 0.00   1 0.222 0.422   36      0
## 9     CLE  0  0      0 0.00   1 0.168 0.374  197      0
## 10    CLT  0  0      0 1.00   1 0.265 0.441 1008      0
## 11    CMH  0  0      0 0.00   0 0.000 0.000    2      0
## 12    CVG  0  0      0 0.00   1 0.202 0.403  178      0
## 13    DCA  0  0      0 0.00   1 0.204 0.403 2123      0
## 14    DEN  0  0      0 0.00   1 0.224 0.417  729      0
## 15    DFW  0  0      0 0.00   1 0.226 0.418  926      0
## 16    DTW  0  0      0 0.00   1 0.161 0.368  596      0
## 17    EWR  0  0      0 1.00   1 0.303 0.460 1295      0
## 18    FLL  0  0      0 1.00   1 0.405 0.491  494      0
## 19    HOU  0  0      0 0.00   1 0.202 0.403  114      0
## 20    IAD  0  0      0 0.00   1 0.230 0.421  822      0
```

```

## 21   IAH   0 0      0 0.00   1 0.238 0.427 499      0
## 22   IND   0 0      0 0.00   1 0.213 0.414  47      0
## 23   JAX   0 0      0 1.00   1 0.344 0.477 180      0
## 24   JFK   0 0      0 0.00   1 0.239 0.427 1319     0
## 25   LAS   0 0      0 1.00   1 0.327 0.470 211      0
## 26   LAX   0 0      0 1.00   1 0.258 0.438 960      0
## 27   LGA   0 0      0 0.00   1 0.205 0.404 1804     0
## 28   LGB   0 0      0 1.00   1 0.269 0.446  93      0
## 29   MCI   0 0      0 1.00   1 0.281 0.453  57      0
## 30   MCO   0 0      0 1.00   1 0.295 0.456 878      0
## 31   MDW   0 0      0 0.00   1 0.237 0.426 455      0
## 32   MEM   0 0      0 0.00   1 0.103 0.307  58      0
## 33   MIA   0 0      0 0.00   1 0.238 0.426 651      0
## 34   MKE   0 0      0 0.00   1 0.139 0.347 380      0
## 35   MSP   0 0      0 0.00   1 0.147 0.354 464      0
## 36   MSY   0 0      0 1.00   1 0.341 0.477  88      0
## 37   MYR   0 0      0 0.00   0 0.000  NA    1      0
## 38   OAK   0 0      0 1.00   1 0.467 0.502  90      0
## 39   ORD   0 0      0 0.00   1 0.219 0.413 1945     0
## 40   ORF   0 0      0 0.00   1 0.158 0.368  57      0
## 41   PBI   0 0      0 1.00   1 0.442 0.498 217      0
## 42   PDX   0 0      0 0.25   1 0.250 0.434 212      0
## 43   PHF   0 0      0 0.00   1 0.242 0.432  62      0
## 44   PHL   0 0      0 0.00   1 0.231 0.422 1680     0
## 45   PHX   0 0      0 0.00   1 0.199 0.399 418      0
## 46   PIT   0 0      0 0.00   1 0.219 0.414 302      0
## 47   RDU   0 0      0 0.00   1 0.250 0.433 629      0
## 48   RIC   0 0      0 0.00   1 0.199 0.400 216      0
## 49   RSW   0 0      0 1.00   1 0.371 0.484 186      0
## 50   SAN   0 0      0 1.00   1 0.267 0.444 217      0
## 51   SEA   0 0      0 0.00   1 0.192 0.394 427      0
## 52   SFO   0 0      0 1.00   1 0.296 0.457 1114     0
## 53   SJC   0 0      0 1.00   1 0.323 0.470  93      0
## 54   SJU   0 0      0 1.00   1 0.399 0.491 303      0
## 55   SLC   0 0      0 0.00   1 0.211 0.410 123      0
## 56   SRQ   0 0      0 0.00   1 0.205 0.409  39      0
## 57   STL   0 0      0 0.00   1 0.220 0.415 205      0
## 58   TPA   0 0      0 1.00   1 0.401 0.491 279      0

```

```

destDelay = bosFlights %>%
  filter(!is.na(RealDelay)) %>%
  group_by(DayofMonth, Month, Year, Dest) %>%
  summarise(delay = mean(RealDelay), n = n()) %>%
  filter(n > 10)

```

```

filter(destDelay, Dest=="ATL") %>% head(.)

```

```

## Source: local data frame [6 x 6]
## Groups: DayofMonth, Month, Year
##
##   DayofMonth Month Year Dest delay  n
## 1           1     8 2011 ATL  34.4 16
## 2           1     8 2012 ATL  10.7 16

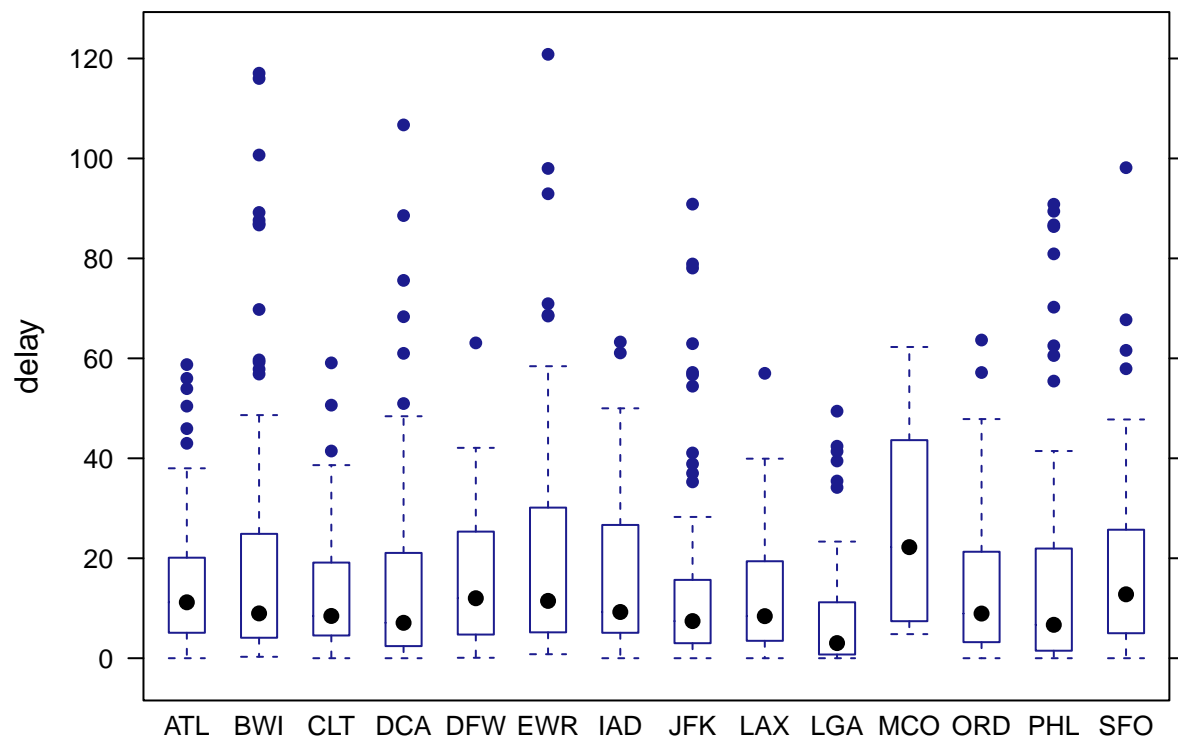
```

```
## 3      1      8 2013  ATL  16.8 16
## 4      2      8 2011  ATL  30.5 15
## 5      2      8 2012  ATL  58.8 16
## 6      2      8 2013  ATL  11.6 16
```

```
mean(delay ~ Dest, data=destDelay)
```

```
##  ATL  BWI  CLT  DCA  DFW  EWR  IAD  JFK  LAX  LGA  MCO  ORD
## 15.01 20.65 13.32 14.46 16.00 21.79 18.09 14.96 12.41  8.51 26.70 14.36
##  PHL  SFO
## 16.61 18.34
```

```
bwplot(delay ~ Dest, data=destDelay)
```



```
airlineDelay = bosFlights %>%
  group_by(DayofMonth, Month, Year, UniqueCarrier) %>%
  summarise(delay = mean(DelayOrCancel), n = n()) %>%
  filter(n > 10)
```

```
filter(airlineDelay, UniqueCarrier=="B6") %>% head(.)
```

```
## Source: local data frame [6 x 6]
## Groups: DayofMonth, Month, Year
##
##   DayofMonth Month Year UniqueCarrier  delay  n
## 1           1     8 2011             B6 0.3370 92
## 2           1     8 2012             B6 0.4227 97
## 3           1     8 2013             B6 0.5128 117
```

```
## 4      2      8 2011      B6 0.3478 92
## 5      2      8 2012      B6 0.0928 97
## 6      2      8 2013      B6 0.3932 117
```

```
mean(delay ~ UniqueCarrier, data=airlineDelay)
```

```
##   AA   B6   CO   DL   FL   MQ   UA   US   WN
## 0.198 0.314 0.397 0.190 0.210 0.361 0.223 0.203 0.230
```

```
bwplot(delay ~ UniqueCarrier, data=airlineDelay) # B6 is JetBlue
```

