

# Airline Delays in the First Course (part 4: analyzing the data from Flagstaff)

Nicholas Horton, [nhorton@amherst.edu](mailto:nhorton@amherst.edu), July 15, 2014

```
require(mosaic)
options(digits=3)
trellis.par.set(theme=col.mosaic()) # get a better color scheme for lattice
load("Meetup-talk/ICOTS-FLG.rda")
names(ds)
```

## Accessing the data

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

```
filter(ds, Year==2013 & DayofMonth==20) %>% # Saturday, July 20th, 2013
  select(DayofMonth, Month, Year, Origin, Dest, CRSDepTime, ArrDelay, Cancelled) %>%
  arrange(CRSDepTime)
```

```
## DayofMonth Month Year Origin Dest CRSDepTime ArrDelay Cancelled
## 1 20 7 2013 FLG PHX 650 -4 0
## 2 20 7 2013 FLG PHX 1030 14 0
## 3 20 7 2013 FLG PHX 1200 -8 0
## 4 20 7 2013 FLG PHX 1500 -8 0
## 5 20 7 2013 FLG PHX 2105 152 0
```

```
ds = mutate(ds, TimeOfDay = cut(CRSDepTime, breaks=c(0, 1200, 1800, 2400),
  labels=c("morning", "afternoon", "evening")))
ds = mutate(ds, delayorcancel = ifelse(is.na(ArrDelay) | ArrDelay > 15, "yes", "no"))
```

```
favstats(~ CRSDepTime, data=ds)
```

```
## min Q1 median Q3 max mean sd n missing
## 539 1030 1200 1655 2115 1340 443 553 0
```

```
tally(~ TimeOfDay, data=ds)
```

```
##
## morning afternoon evening
## 279 177 97
```

```
tally(~ Cancelled, format="percent", data=ds)
```

```
##  
##      0      1  
## 96.93  3.07
```

```
tally(~ delayorcancel, format="percent", data=ds)
```

```
##  
##   no  yes  
## 86.3 13.7
```

```
tally(~ Cancelled | TimeOfDay, format="percent", data=ds)
```

```
##           TimeOfDay  
## Cancelled morning afternoon evening  
##           0  98.21    94.35    97.94  
##           1   1.79     5.65     2.06
```

```
tally(~ delayorcancel | TimeOfDay, format="percent", data=ds)
```

```
##           TimeOfDay  
## delayorcancel morning afternoon evening  
##           no    90.0    79.7    87.6  
##           yes   10.0    20.3    12.4
```

```
favstats(~ ArrDelay, data=ds)
```

```
## min Q1 median Q3 max mean  sd  n missing  
## -32 -13    -7  1 300 1.18 34.2 535     18
```

```
ds = mutate(ds, ActDelay = ifelse(ArrDelay < 0, 0, ArrDelay))  
favstats(~ ActDelay, data=ds)
```

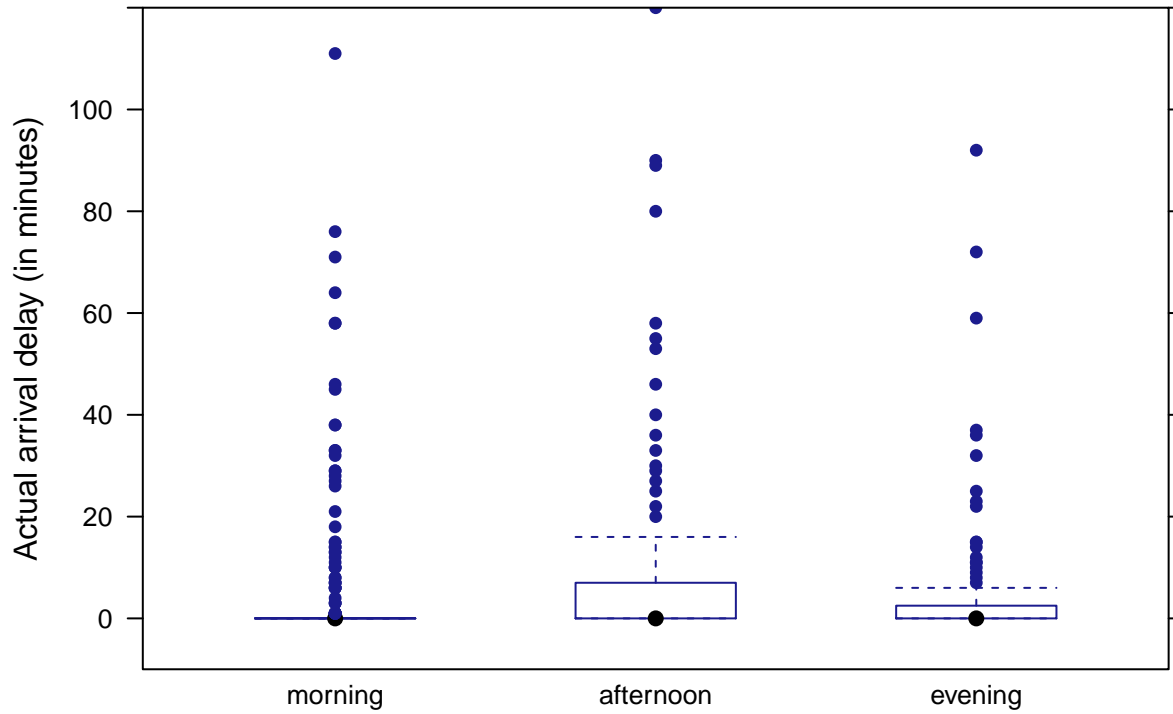
```
## min Q1 median Q3 max mean  sd  n missing  
##   0  0     0  1 300 8.99 31.2 535     18
```

```
favstats(ActDelay ~ TimeOfDay, data=ds)
```

```
##   .group min Q1 median Q3 max mean  sd  n missing  
## 1 morning  0  0     0  0.0 250  5.95 22.9 274     5  
## 2 afternoon 0  0     0  6.5 300 15.01 44.5 166    11  
## 3 evening  0  0     0  2.5 152  7.24 21.1  95     2
```

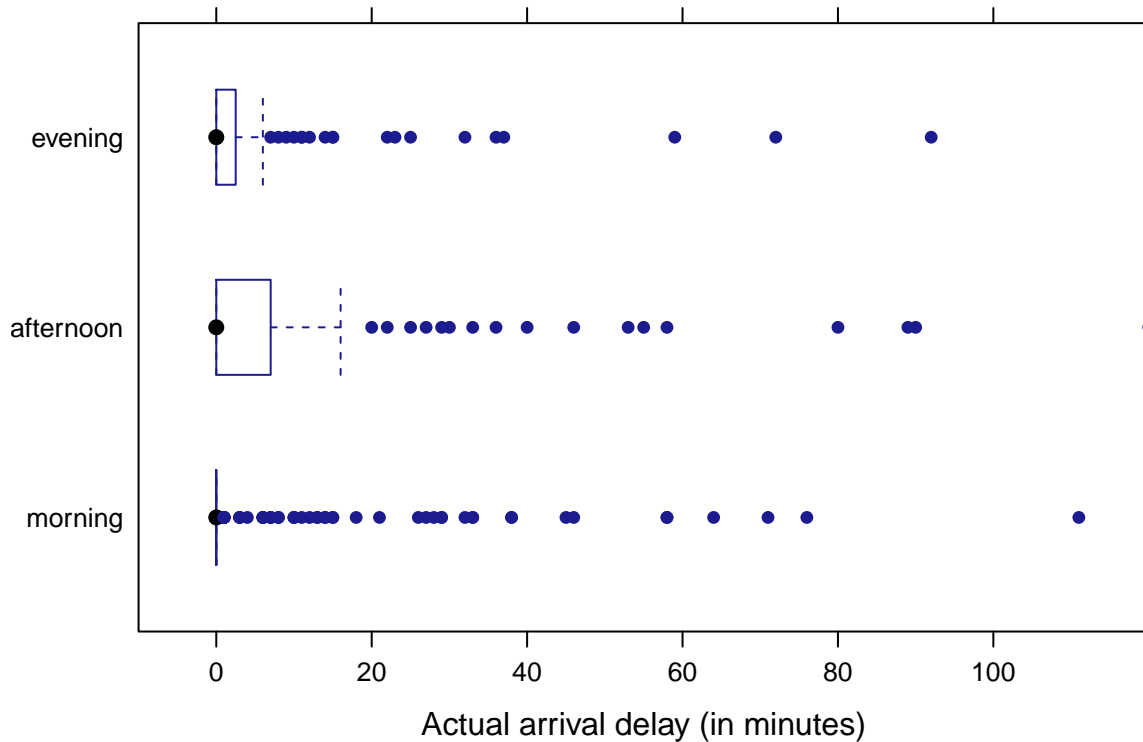
```
bwplot(ActDelay ~ TimeOfDay, ylim=c(-10, 120), main="July flights from Flagstaff, 2011-2013", ylab="Act")
```

### July flights from Flagstaff, 2011–2013

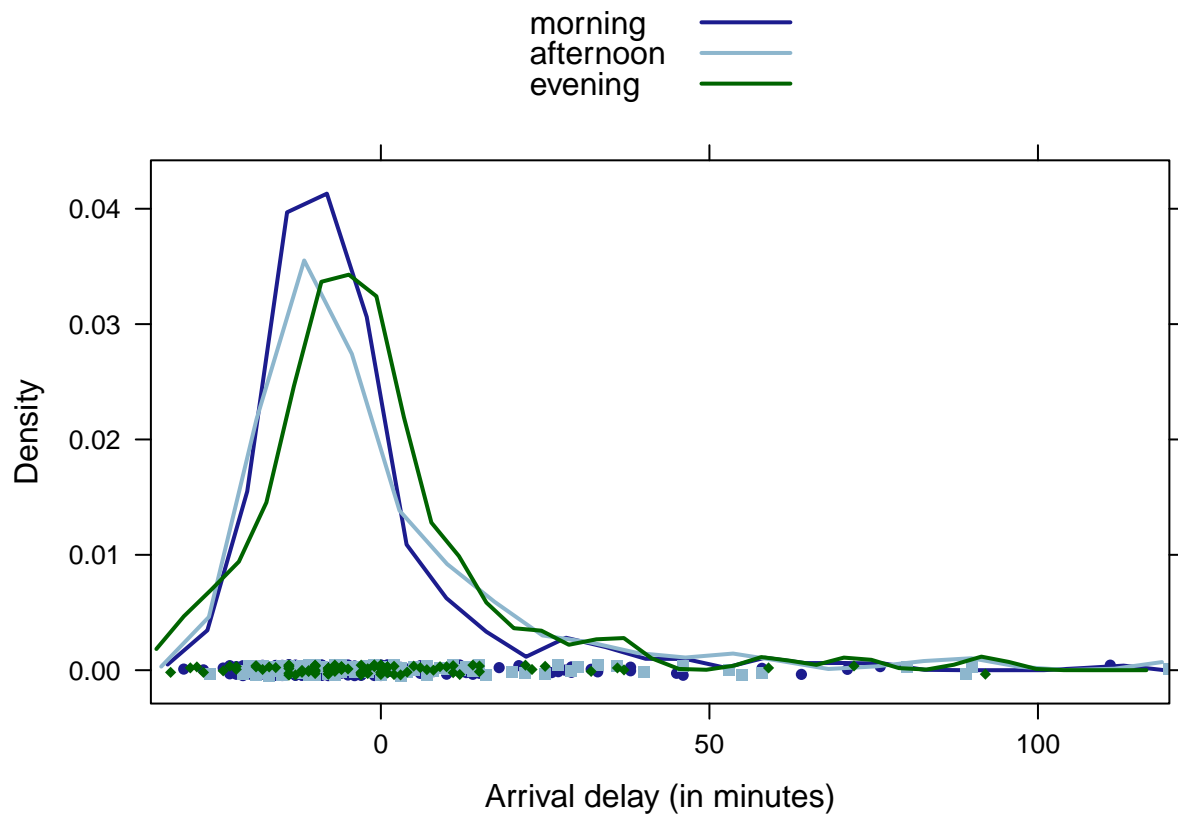


```
bwplot(TimeOfDay ~ ActDelay, xlim=c(-10, 120), main="July flights from Flagstaff, 2011-2013", xlab="Act
```

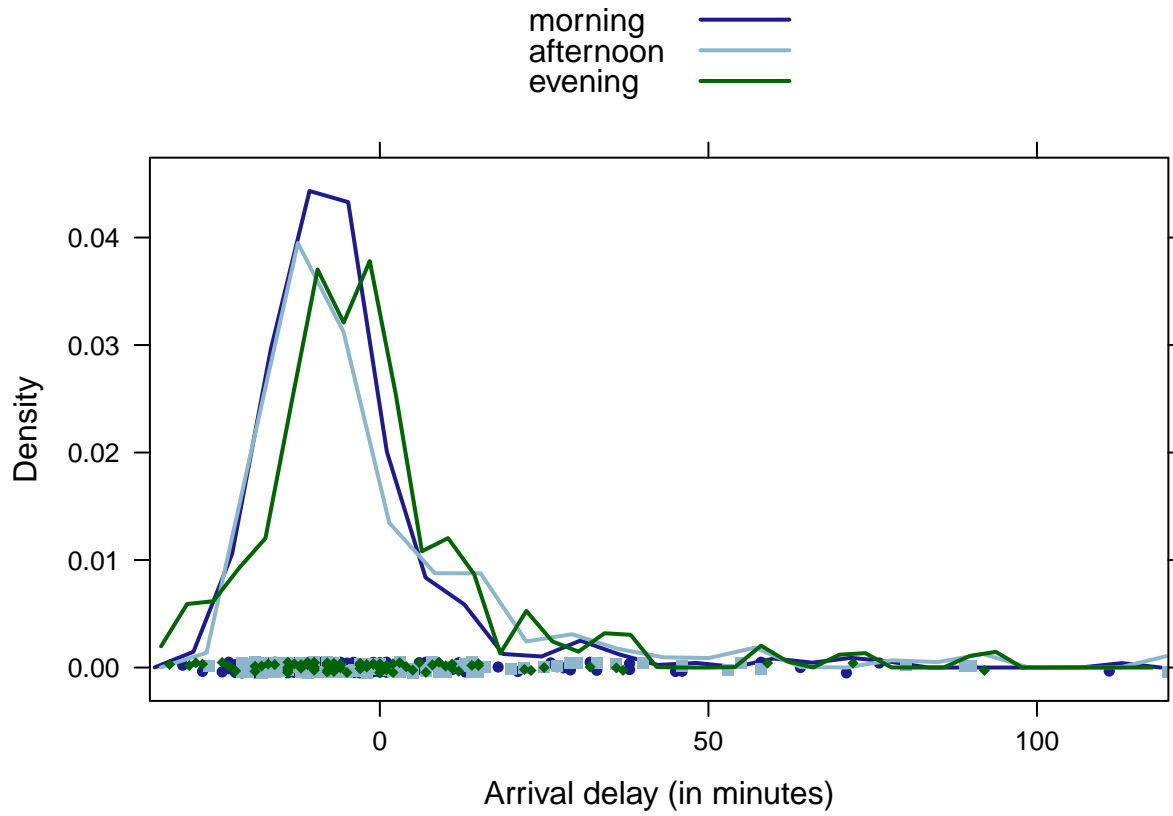
### July flights from Flagstaff, 2011–2013



```
densityplot(~ ArrDelay, groups=TimeOfDay, auto.key=TRUE,  
            xlab="Arrival delay (in minutes)", xlim=c(-35, 120), data=ds)
```



```
densityplot(~ ArrDelay, groups=TimeOfDay, adjust=1/2, auto.key=TRUE,  
            xlab="Arrival delay (in minutes)", xlim=c(-35, 120), data=ds)
```



```
ds2 = filter(ds, Year==2013)
sort(tally(~ TailNum, data=ds2), decreasing = TRUE)
```

```
##
## N874AS N496CA N468CA N885AS N944SW N875AS N495CA N907SW N886AS N906SW
##      23      17      15      14      14      13      12      12      11      11
## N889AS N821AS N506CA N955SW
##      9      8      7      4
```