Airport-activity

Christopher Eshleman

8/18/2019

Airport activity:

Building a quick profile of daily this and that.

(Code chunk shortcut: Command+Option+I)

Lemme create a palette. Clear my working space, load some useful packages, and set my working drive.

```
cat("\014") # clear the console
```

```
rm(list=ls())
dev.off()
```

```
## null device
##           1
```

```
library(reshape2)
library(stringr)
library(plyr)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(formattable)
library(doBy)
library(ggplot2)

setwd("~/Dropbox/Work and research/Airport-activity/aviation-activity")
```

Colleagues in the Aviation Department drummed up data on weekday activity for a few random weekdays and weekends across a couple of typical years (2015 and 2016) for activity at the airport I'm looking at.

The weekday data:

```
list.files()
```

```
##  [1] "Airport-activity.Rmd"
##  [2] "All runway data 20190818.csv"
##  [3] "aviation-activity.Rproj"
##  [4] "EWR data.R"
##  [5] "EWR_FD_150914_150916.csv"
##  [6] "EWR_FD_160914_160916.csv"
##  [7] "EWR_FD_2015_WKND.csv"
##  [8] "EWR_FD_2016_WKND.csv"
##  [9] "EWR_FD_20160917-18_20150919-20.xlsx"
## [10] "README.md"
## [11] "Save weekday 20190818.csv"
## [12] "Save weekend 20190818.csv"
## [13] "Time of day.csv"
```

```
a1 = read.csv("./EWR_FD_150914_150916.csv", skip = 7)
b1 = read.csv("./EWR_FD_160914_160916.csv", skip = 7)
ab1 = rbind(a1,b1)
  names(ab1) = tolower(names(ab1))
ab1$runway.occupancy.time=NULL
ab1$taxiway.used.to.enter.exit.runway=NULL
ab1$daytype = "weekday"
rm(a1,b1)
```

Locked and loaded.

The weekend data:

```
a = read.csv("./EWR_FD_2016_WKND.csv") #read.csv("./EWR_FD_150914_150916.csv", skip = 7)
head(a)
```

```
##      Carrier.Group Call.Sign Registration Model Operation
## 1  United Airlines   UAL1960       N34131  B752   Arrival
## 2  United Airlines    UAL213       N587UA  B752   Arrival
## 3 General Aviation   AAL1384       N845NN  B738   Arrival
## 4  United Airlines   RPA3540       N644RW  E170   Arrival
## 5  United Airlines   UAL1261       N14228  B738   Arrival
## 6  United Airlines   UAL1186       N411UA  A320   Arrival
##   Origination.Airport Destination.Airport Gate.Assigned..Aerobahn.
## 1                 LAX                 EWR                 Gate_110
## 2                 SFO                 EWR                 Gate_104
## 3                 CLT                 EWR                  Gate_37
## 4                 MEM                 EWR                  Gate_87
## 5                 ORD                 EWR                 Gate_134
## 6                 LAS                 EWR                 Gate_132
##       Gate Runway.Assigned..Aerobahn. Runway First.Fix
## 1 Gate_110                       22L    22L       DAG
## 2 Gate_104                       22L    22L     TIPRE
## 3  Gate_37                       22L    22L     AUDII
## 4  Gate_87                       22L    22L     DIYAB
## 5 Gate_134                       22L    22L     DUFEE
## 6 Gate_132                       22L    22L       DVC
##   International.or.Domestic.Indicator   Event.Time Flight.Origination.Date
## 1                            Domestic 9/17/16 0:02
## 2                            Domestic 9/17/16 0:05
## 3                            Domestic 9/17/16 0:08
## 4                            Domestic 9/17/16 0:11
```

```
## 5                            Domestic 9/17/16 0:12
## 6                            Domestic 9/17/16 0:25
##   Scheduled.Off.Block.Time..Aerobahn. Actual.Off.Block.Time..Aerobahn.
## 1                      9/16/16 19:06                    9/16/16 19:11
## 2                      9/16/16 19:00                    9/16/16 18:55
## 3                      9/16/16 22:15                    9/16/16 22:15
## 4                      9/16/16 20:46                    9/16/16 21:55
## 5                      9/16/16 22:10                    9/16/16 22:30
## 6                      9/16/16 19:59                    9/16/16 19:59
##   Movement.Area.Entrance.Time Actual.Take.Off.Time..Aerobahn.
## 1                                                9/16/16 19:23
## 2                                                9/16/16 19:11
## 3                                                9/16/16 22:48
## 4                                                9/16/16 22:04
## 5                                                9/16/16 22:41
## 6                                                9/16/16 20:15
##   Actual.Landing.Time..Aerobahn. Movement.Area.Exit.Time
## 1                  9/17/16 0:02           9/17/16 0:06
## 2                  9/17/16 0:05           9/17/16 0:12
## 3                  9/17/16 0:08           9/17/16 0:11
## 4                  9/17/16 0:11           9/17/16 0:15
## 5                  9/17/16 0:12           9/17/16 0:18
## 6                  9/17/16 0:25           9/17/16 0:32
##   Scheduled.In.Block.Time..Aerobahn. Actual.In.Block.Time..Aerobahn.
## 1                      9/17/16 0:21                   9/17/16 0:07
## 2                      9/17/16 0:13                   9/17/16 0:13
## 3                      9/17/16 0:04                   9/17/16 0:12
## 4                     9/16/16 23:38                   9/17/16 0:15
## 5                      9/17/16 0:19                   9/17/16 0:19
## 6                      9/17/16 0:49                   9/17/16 0:33
##   Total.Taxi.Time
## 1         0:05:28
## 2         0:07:32
## 3         0:03:48
## 4         0:04:28
## 5         0:06:41
## 6         0:07:48
```

```
b = read.csv("./EWR_FD_2015_WKND.csv") #read.csv("./EWR_FD_160914_160916.csv", skip = 7)
ab = rbind(a,b)
names(ab) = tolower(names(ab))
ab$daytype = "weekend"
rm(a,b)
```

A little messy, so rename variables and then consolidate.

```
names(ab1) = names(ab)
ab = rbind(ab1,ab)
rm(ab1)
```

Preparation.

The airport tracks all kind of activity. But I'm interested explicitly in arrivals and departures.

```
table(ab$operation)
```

```
##
##           Arrival         Departure   Missed Approach Rejected Take-off
##              5573              5551               154                 8
```

```
ab = subset(ab, ab$operation=="Arrival" | ab$operation == "Departure")
```

The underlying .csvs and the imports didn't recognize time for what it was. (Times and dates usually require a little extra attention.)

```
ab$day.time = strptime(ab$event.time, "%m/%d/%y %H:%M", tz = "EST5EDT")
```

```
## Warning in strptime(ab$event.time, "%m/%d/%y %H:%M", tz = "EST5EDT"):
## unknown timezone 'zone/tz/2019b.1.0/zoneinfo/America/New_York'
```

```
ab$date = format(as.POSIXct(ab$day.time,format='%m/%d/%Y %H:%M:%S'),format='%m/%d/%Y')
```

```
## Warning in as.POSIXct.POSIXlt(ab$day.time, format = "%m/%d/%Y %H:%M:%S"):
## unknown timezone 'zone/tz/2019b.1.0/zoneinfo/America/New_York'
```

```
## Warning in as.POSIXlt.POSIXct(x, tz): unknown timezone 'zone/tz/2019b.1.0/
## zoneinfo/America/New_York'
```

```
ab$date = as.Date(ab$date, format="%m/%d/%Y")

ab$weekday = weekdays(ab$date)
```

Specifically, I want to be able to treat time down to the minute, so I'll just create a time variable that lets me do it easily.

```
ab$time = as.POSIXct(as.numeric(as.POSIXct(ab$day.time)) %% 86400, origin = "2000-01-01")
```

```
## Warning in as.POSIXct.POSIXlt(ab$day.time): unknown timezone 'zone/tz/
## 2019b.1.0/zoneinfo/America/New_York'
```

```
ab$time2 = as.numeric(as.POSIXct(ab$day.time)) %% 86400  # 60*60*24
```

```
## Warning in as.POSIXct.POSIXlt(ab$day.time): unknown timezone 'zone/tz/
## 2019b.1.0/zoneinfo/America/New_York'
```

```
benchpoints = c("2001-01-01 05:30:00", "2001-01-01 06:30:00", "2001-01-01 9:30:00", "2001-01-01 16:30:0
benchpoints2 = as.numeric(as.POSIXct(benchpoints)) %% 86400

# Effective hours are everything but 11pm-6am. 86400 seconds in a day.
ab$timecat = ifelse(ab$time2>=82800 | ab$time2<=21600,"Night","SixAMtoElevenPM") #18000
ab = droplevels(ab) #Quick cleanup to trim unused data levels and
```

OK, the analysis is going to consider expected delays by time of day. There are some rule-of-thumb level windows of time that the airport uses as categories for high-level analysis. I'll borrow them and shape the data into tables. I'll treat 6 AM to 11 PM as peak and the remaining hours as "night".

```
table(ab$timecat)
```

```
##
##           Night SixAMtoElevenPM
##            2784            8340
```

```
table(ab$timecat,ab$operation)
```

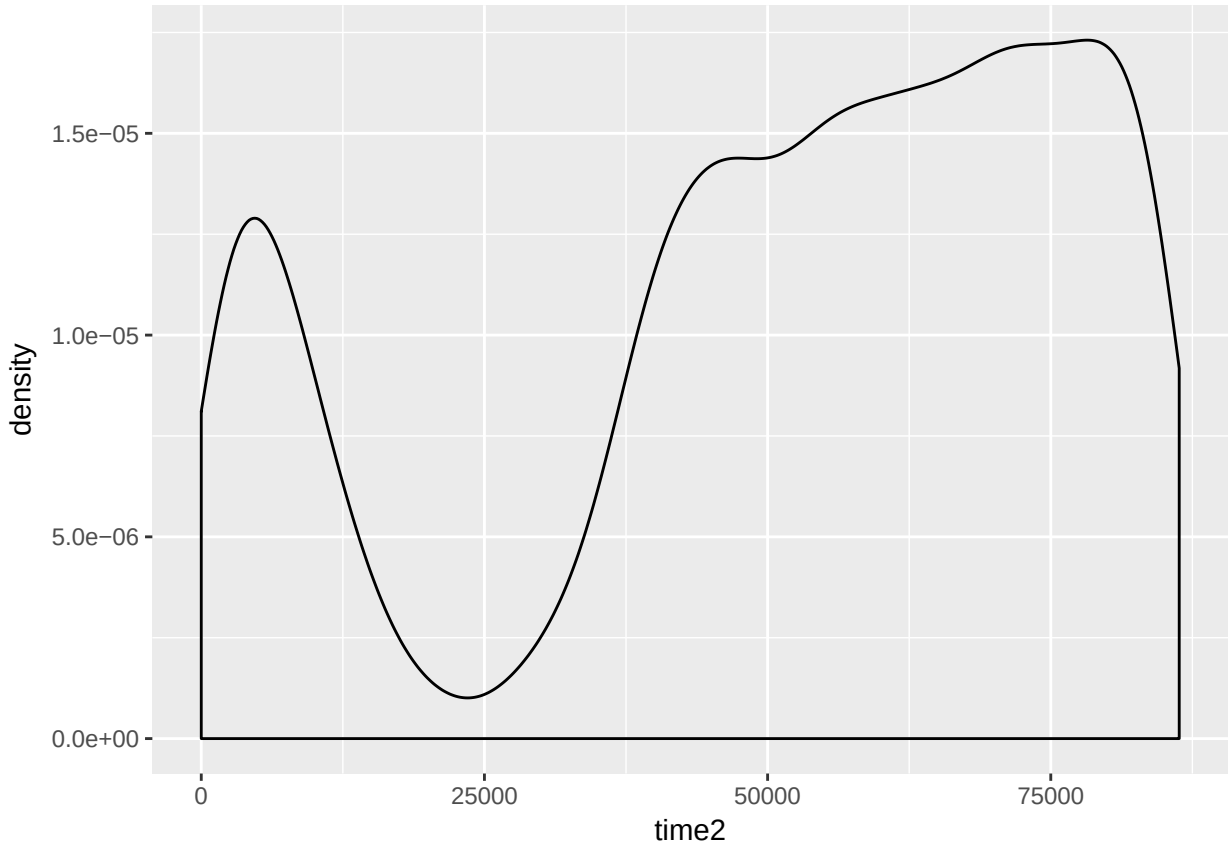
```
##
```

```
##                   Arrival Departure
##   Night              1437      1347
##   SixAMtoElevenPM    4136      4204
```

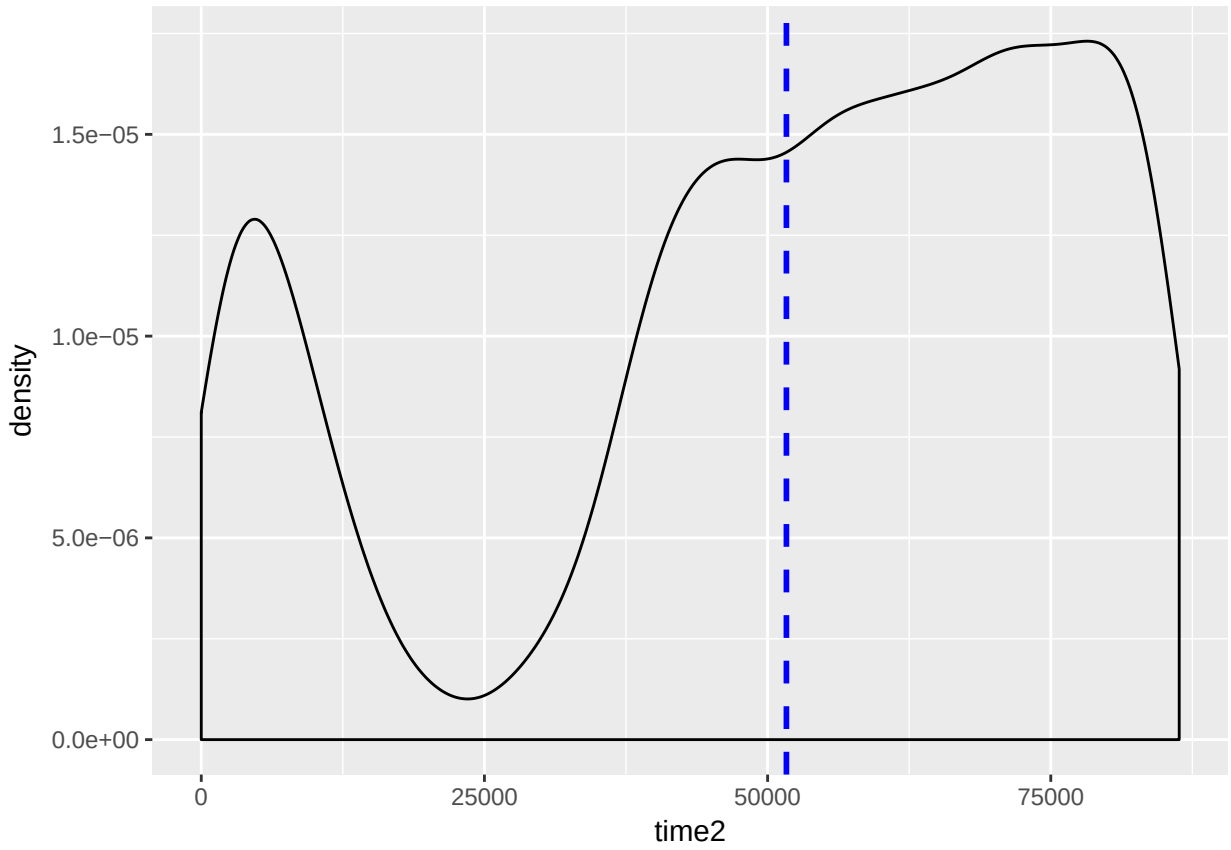
```
ab_wknd = subset(ab,ab$daytype=="weekend")
ab_wkdy = subset(ab,ab$daytype=="weekday")
ab_wknd = table(ab_wknd$timecat,ab_wknd$runway)
ab_wkdy = table(ab_wkdy$timecat,ab_wkdy$runway)
```

So what DOES activity at the airport look like, anyway?

```
# Basic density
p = ggplot(ab, aes(x=time2)) +
  geom_density()
p
```

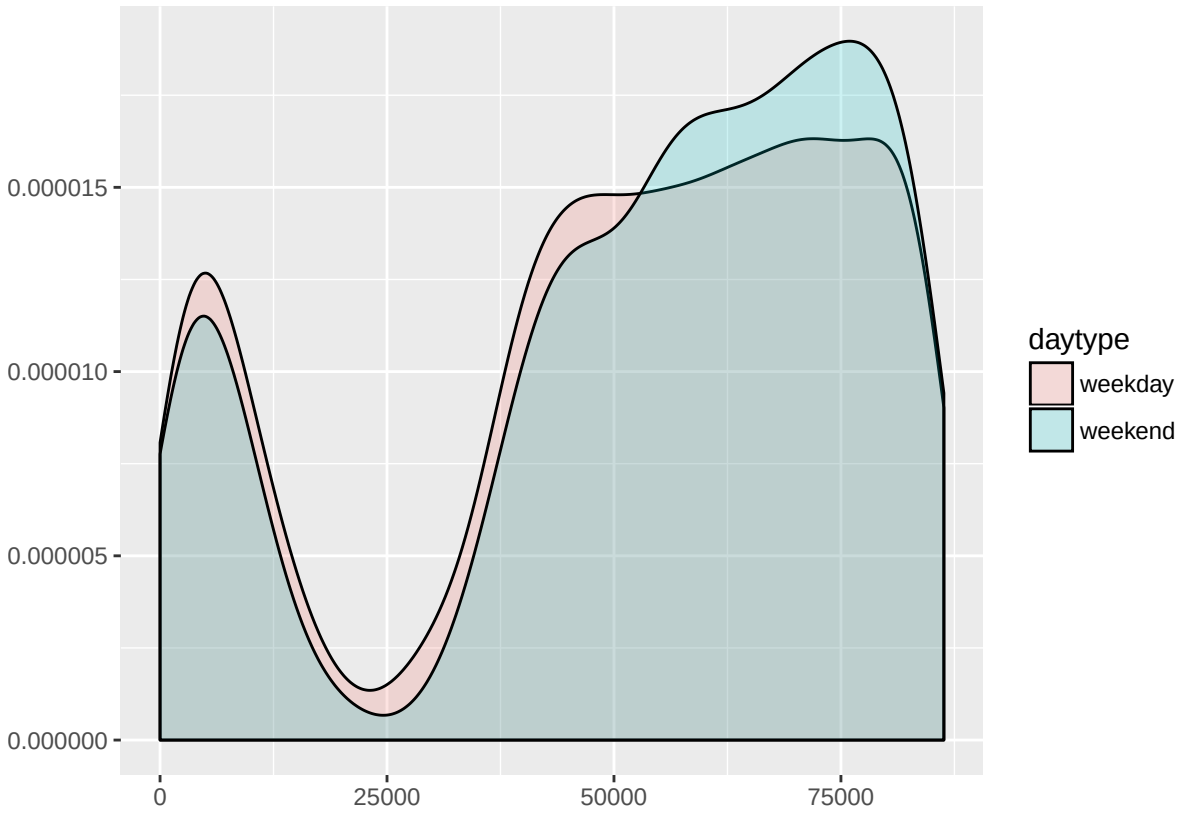



```
# Add mean line
p + geom_vline(aes(xintercept=mean(time2)),
               color="blue", linetype="dashed", size=1)
```



These will be the tables I use to apply expected delays, which we will estimate separately using the same data but during periods of similar delays in other years. First I'm curious as to how different the airport's weekday profile is from the weekend:

```
options(scipen=5)
pq = ggplot(ab, aes(x=time2, fill=daytype)) +
  geom_density(alpha=0.2) + labs(x = "", y = "")
pq
```




```
ggsave(pq, file="./Activity_EWR.png", height=4,width=6)
```

The distributions are fairly similar.

Saving the data:

```
write.csv(ab_wknd,"./Save weekend 20190818.csv")
write.csv(ab_wkdy,"./Save weekday 20190818.csv")
write.csv(ab,"./All runway data 20190818.csv")
```