Airport-activity

Christopher Eshleman

8/18/2019

# Airport activity:

## Taking a quick snapshot of daily comings and goings.

A number of departments collaborated this spring to evaluate plans for a runway reconstruction at Newark Airport. The question at focus was whether it’d be cheaper, over the life of the investment (around four decades), to build with asphalt, concrete or a designed hybrid.

One step of the evaluation involved an analysis of the expected impact to users. (By users we meant airlines and the traveling public, who we need to consider in slightly different ways.) The runway is one of two major runways and will need to be shut down for months, which will force activity for part or all (depending on how the project is designed) of a given day’s activity onto the second major runway. (There’s also a third, smaller runway but we’re omitting it from the analysis to keep things simple and generalizable.)

We ran a basic analysis to inform the analysis, knowing we could return and build something more prescriptive if the alternatives were close enough in expected costs and impacts to merit real surgery.

A basic but critical step is to summarize activity, by time of day, for a typical day. We used R and a few handfuls of daily activity data from the Aviation’s activity database.

First, create a palette. Clear the working space, load some useful packages, and set the working drive.

Colleagues in the Aviation Department drummed up data for a handful of typical weekdays and weekends across a couple of typical years (2015 and 2016) at the airport.

The weekday data:

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)

It looks like the variables were named by hand or the terminology changed a little from one year to the next, so I’ll rename variables before I consolidate the data.

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

## Preparation.

The airport categorizes activity within a few buckets but we’re explicitly interested 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")

R doesn’t usually recognize time and day variables for what they are. (Times and dates require a little extra attention in all statistical software.)

ab$day.time = strptime(ab$event.time, "%m/%d/%y %H:%M", tz = "EST5EDT")   
ab$date = format(as.POSIXct(ab$day.time,format='%m/%d/%Y %H:%M:%S'),format='%m/%d/%Y')  
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. I can convert things to seconds, minutes, et cetera. The airport’s peak hours are everything but 11pm-6am. (There are 86400 seconds in a day, 60 minutes x 60 minutes x 24 hours.)

ab$time = as.POSIXct(as.numeric(as.POSIXct(ab$day.time)) %% 86400, origin = "2000-01-01")   
ab$time2 = as.numeric(as.POSIXct(ab$day.time)) %% 86400   
benchpoints = c("2001-01-01 05:30:00", "2001-01-01 06:30:00", "2001-01-01 9:30:00", "2001-01-01 16:30:00", "2001-01-01 19:30:00", "2001-01-01 23:00:00")   
benchpoints2 = as.numeric(as.POSIXct(benchpoints)) %% 86400   
ab$timecat = ifelse(ab$time2>=82800 | ab$time2<=21600,"Night","SixAMtoElevenPM")  
ab = droplevels(ab) #Quick cleanup to trim unused data levels and

OK, the analysis is going to consider expected delays by time of day. We did that delay calculation separately (using the same data but for a year when simiilar construction occurred).

(Come to think of it, it would have been nice to replicate that work and bake it into this script.)

Here, I aggregate the data for export and apply the delays in a spreadsheet consolidating the larger project, which is done by multiple people across teams with various analytical tools (largely Excel).

table(ab$timecat)

##   
## Night SixAMtoElevenPM   
## 2784 8340

table(ab$timecat,ab$operation)

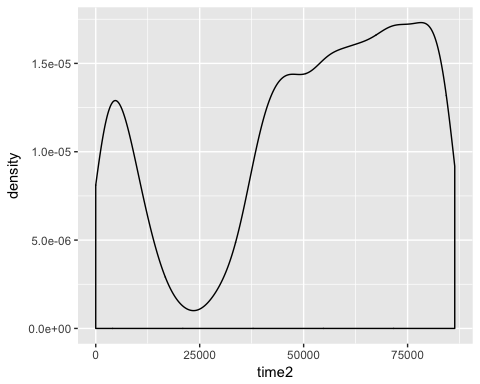
##   
## Arrival Departure  
## Night 1437 1347  
## SixAMtoElevenPM 4136 4204

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)

## Exploration and visualization.

So what DOES activity at the airport look like, anyway? Here are some basic density maps.

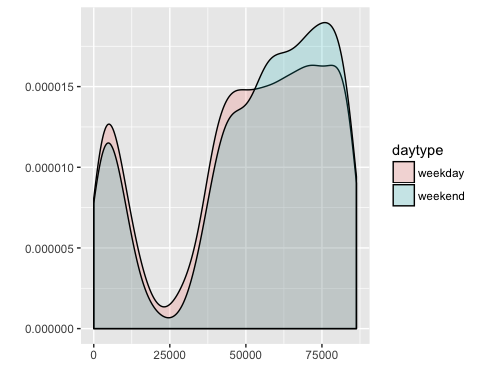
p = ggplot(ab, aes(x=time2)) +   
 geom\_density()  
p



pp = p + geom\_vline(aes(xintercept=mean(time2)),  
 color="blue", linetype="dashed", size=1) # Add mean line  
ggsave(p, file="./Raw\_EWR.png", height=4,width=6)   
ggsave(pp, file="./Raw\_w\_mean\_EWR.png", height=4,width=6)

The airport’s weekday and weekend profiles aren’t really that different, distributionally (below) or in raw volume:

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)

I saved the plot separately in case I want to use it in presentations.

## Save and clean up.

Ultimately the tables are what I’ll use to develop expected delays. 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")  
options(warn=0) # Turn warnings back on for next project.