Air Traffic Delays

MIDS Bridge Project II
Laura Williams Geoff Stirling John Schoen
Source: US DOT

Overview

Air traffic delays are not just an inconvenience to passengers. Last year, airline flight delays and cancellations cost American passengers more than $2 billion in out-of-pocket costs for hotels, meals, alternative travel arrangements. We decided to take a closer look at the nature and potential causes of these delays with an analysis of U.S. Department of Transportation data on flight delays for the month of July, 2015, at the height of the summer travel season.

SOURCE:
(https://www.rita.dot.gov/bts/data_and_statistics/by_mode/airline_and_airports/airline_delay.html)

Methodology

We applied a very simple analysis to the dataset, which covered 500,000+ unique flights between two U.S. airports at some time during the month. The dataset initially included 110 columns, which record arrival and departure times, unique IDS for airports, airlines, departure and arrival cities, airport codes, taxi times, inflight times, flight distance, etc.

A quick analysis demonstrated that the difference between departure delays and arrival delays was very small, so we chose to focus on arrival delays as the basic measure of on time performance.

We were also able to eliminate duplicate columns (FIPS codes, multiple carrier IDS, etc.) to simplify or dataset and improve the speed and performance of our code blocks.

Columns used for this analysis:

Year
Month
DayofMonth
DayOfWeek
UniqueCarrier
AirlineID
TailNum
FlightNum
OriginAirportID
OriginCityMarketID
Origin
OriginCityName
OriginState
DestAirportID
DestCityMarketID
Dest
DestCityName
DestState
DepDelay
DepDelayMinutes
DepDel15
ArrDelay
ArrDelayMinutes
ArrDel15
ArrTimeBlk
CarrierDelay
WeatherDelay
NASDelay
SecurityDelay
LateAircraftDelay

Results

Worst delayed flights

One of the simplest and most basic questions is also one of the most interesting: which poor traveling souls were subject to the worst inconvenience in trying to get to their destinations? By reading our table into a pandas DataFrame we were able to run a simple sort on the ArrDelay column.

American Airlines came in with the five worst delays for the month, the longest lasting 1636 minutes – some 27.25 hours or more than a full day.

	DayofMonth	DayOfWeek	UniqueCarrier	FlightNum	Origin	OriginCityName	Dest	DestCityName	DepDelayMinutes	ArrDelayMinutes	CarrierDelay	WeatherDelay	NASDelay	SecurityDelay	LateAircraftDelay
47849	9	4	AA	16	SFO	San Francisco, CA	JFK	New York, NY	1176	1167	1167	0	0	0	0
38485	28	2	AA	2334	ORD	Chicago, IL	DFW	Dallas/Fort Worth, TX	1223	1203	804	0	0	0	399
15000	8	3	AA	162	HNL	Honolulu, HI	LAX	Los Angeles, CA	1170	1212	1170	0	42	0	0
3115	18	6	AA	210	LAS	Las Vegas, NV	JFK	New York, NY	1393	1364	1364	0	0	0	0
22263	13	1	AA	1319	IND	Indianapolis, IN	LAX	Los Angeles, CA	1625	1636	1625	0	11	0	0

Outlier data like these need to be verified due to common occurrences of "dirty data."

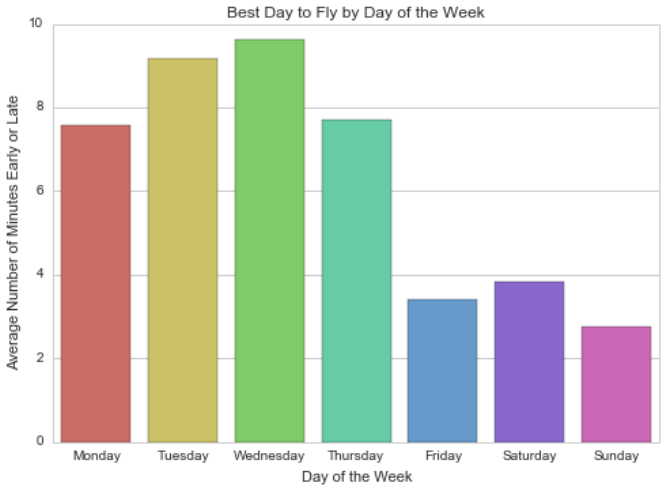
In this case, the entry could have recorded the wrong arrival and departure dates turning what may have been a two-hour delay into an epic travel tragedy. (Note: for this analysis we encountered a number of NaN records, which could also indicate problems with the reliability of this data set.)

Beating the clock

While American recorded the five worst delays, the airline also logged four of the best early flights for the month. This should not be surprising; as one of the largest airlines with the most flights, the distribution of Early/Late flights will be wider than other, smaller carriers. A better analysis would normalize for the number of flights. (We tried running the .mean method, but the processing time proved excessive.)

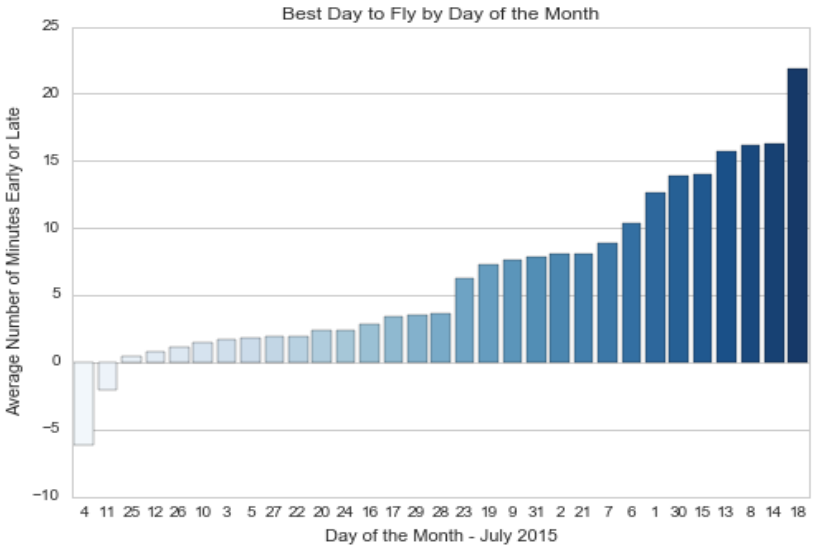
	DayofMonth	DayOfWeek	UniqueCarrier	FlightNum	Origin	OriginCityName	Dest	DepDelay	ArrDelay
76729	4	6	AS	3	DCA	Washington, DC	SEA	-13	-73
80490	10	5	AS	729	IAD	Washington, DC	SEA	-10	-68
47515	19	7	AA	2	LAX	Los Angeles, CA	JFK	-4	-68
245	11	6	AA	45	JFK	New York, NY	SEA	-3	-65
86507	28	2	AS	11	EWR	Newark, NJ	SEA	-10	-64

Best Day To Fly



The US air traffic system experiences periods of congestion that will delay flights no matter how well the airline or system is managed. The travelers determine which days are most prone to congestion, we looked at the average delay for all flights on a given weekday. (The sum of minutes would give an erroneous result because there were the month of July 2015 included five Wednesdays, Thursdays and Fridays and only four occurrences of the remaining weekdays.) Wednesday was the worst day for delays, with an average of nearly 10 minutes per flight, while Sunday was the best day, with less than three minutes delay per flight.

Arrive early: Fly on the Fourth of July

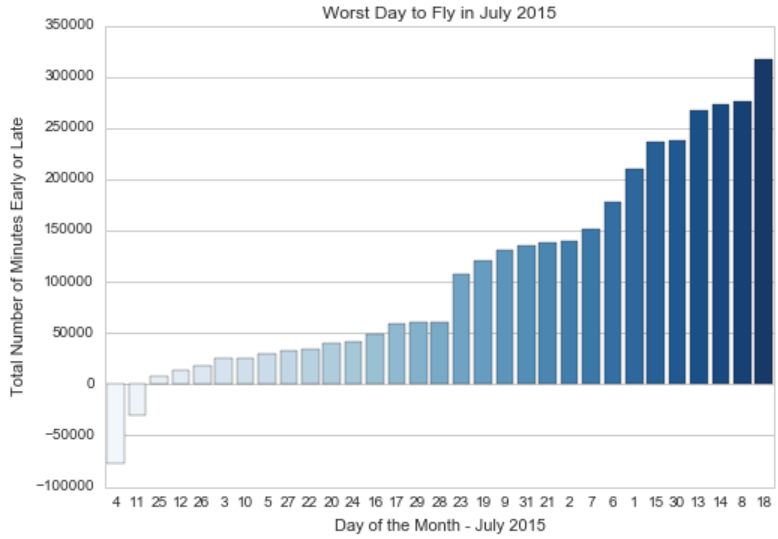


It turned out that the ArrDelay series actually records the variance from published schedules, which lets us determine when flights arrived *early.* By taking the aggregate of ArrDelay for each day of the month, we were able to see which July day produced the worst delays and which one included the best on time performance. By far, the best day to fly in July is the Fourth when most Americans are busy swilling beer at a backyard barbeque. On that date last year, U.S. airlines delivered their passengers to their destinations a total of more than 77,000 minutes early or an average of seven minutes per flight.

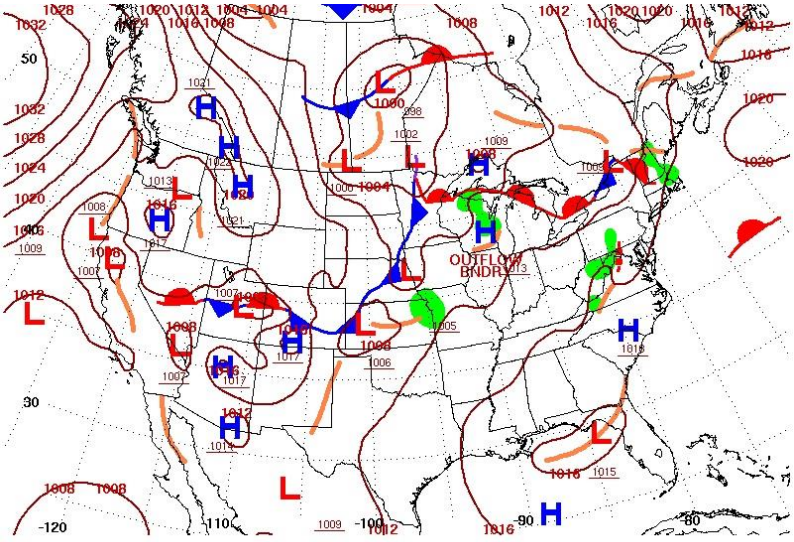
July 4, 2015 also produced the best single "undertime" record in our dataset. On that date, passengers flying Alaska Air from Washington, DC to Seattle lifted off the runway 13 minutes early and landed a full 73 minutes ahead of schedule.

Watch out for thunderstorms

The worst total delays occurred on the 18th of July, a Saturday, which is typically one of the least delayed days of the week. So what happened on the 18th?

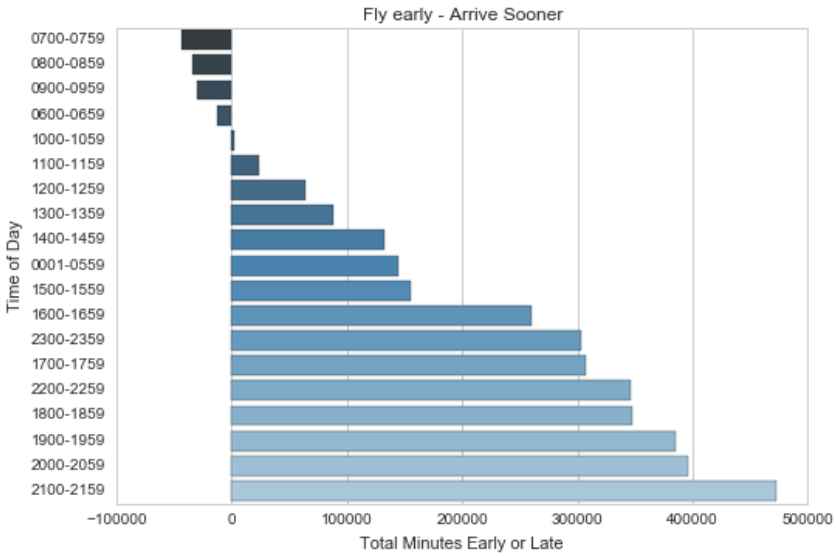


Turns out that was a bad day for thunderstorms, based on the NOAA map for that day.



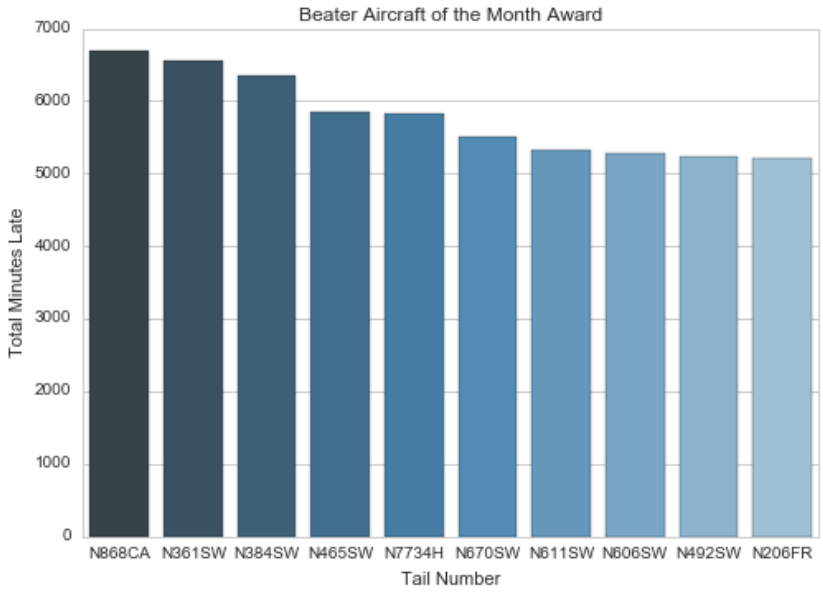
To get there early, get up early

No matter what day of the week you want to fly, you'll get their sooner, on average, if you book the earliest flight possible.



Beater airplanes

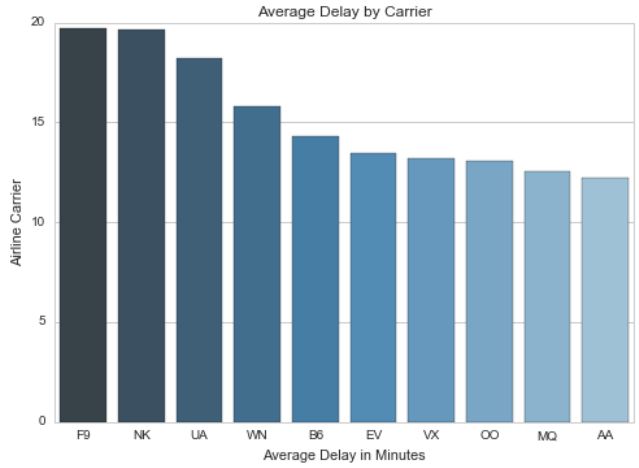
Equipment problems are also a major cause of flight delays. So we wanted to know which aircraft had the worst ontime records. Here are the tail numbers to watch out for. (Not that you can look up tail numbers on Expedia.)



Worst airline

All airlines have to deal with bad weather, maintenance issues and systemwide congestion. Some handle it better than others. Looking at the net minutes of early/late arrivals, Southwest wins the prize for the biggest cumulative delays more than twice as much as the runner-up, United. In July 2015, Southwest kept its passengers behind schedule by a net total of some 1.2 million minutes or 20,000 hours, 833 days, and 2.28 years. All in one month.

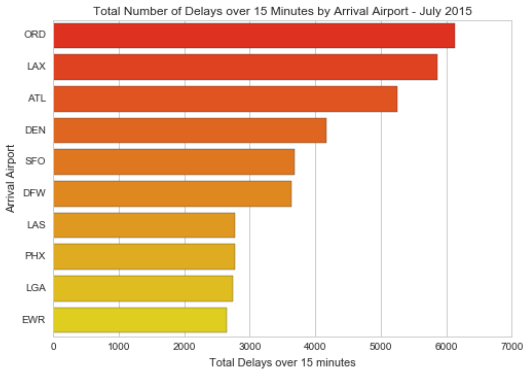
On the other hand, looking at average delays per flight, Frontier leads the pack.

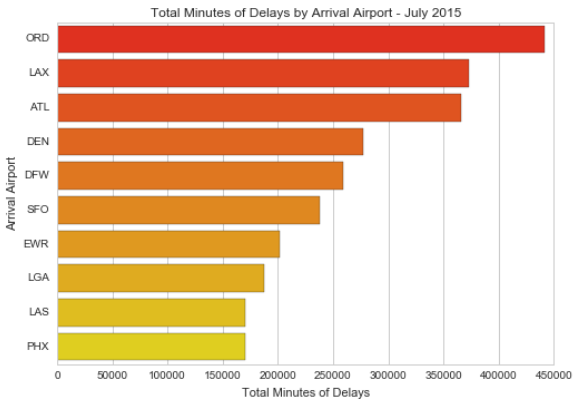


Worst Airport

Given a choice, most air travelers would prefer to avoid airports that routinely delay their arrival at their desired destination. Are some airports reliably network bottlenecks?

The simplest data filter suggests that Chicago's O'Hare was the worst offender in July 2015, ranking at the top for both total delays more than 15 mins and total minutes of delay.





But ORD was the second busiest airports that month (with 28,183 flights) after Atlanta (33,735). So a better measure of performance would normalize for delays per flight. On average, we found a much higher chance of late arrival at much smaller airports.

If you're traveling to Gunnison, Colorado, allow a little extra time. The town was named for John W. Gunnison, the first European arrival who, in 1853, searching for a transcontinental railroad route, was delayed there for three days before traveling west to Utah. Thanks to major advances in travel technology, you won't have to wait nearly as long. But expect to spend an extra 30 mins (on average) arriving at this destination.

Based on the share of flights delayed 15 mins or more, Guam had the worst record in July, 2015. That month, the local air traffic infrastructure of this American territory landed more than half of all flights.