

Exploratory Data Analysis of US Flights

Anant Agarwal

August 17, 2018

The dataset provides US flight data for the years 1990 through 2017. There were 85199 routes during this time period, and there were 5599602 operating during this period. Here is a quick graphical summary of how many flights operated each year:

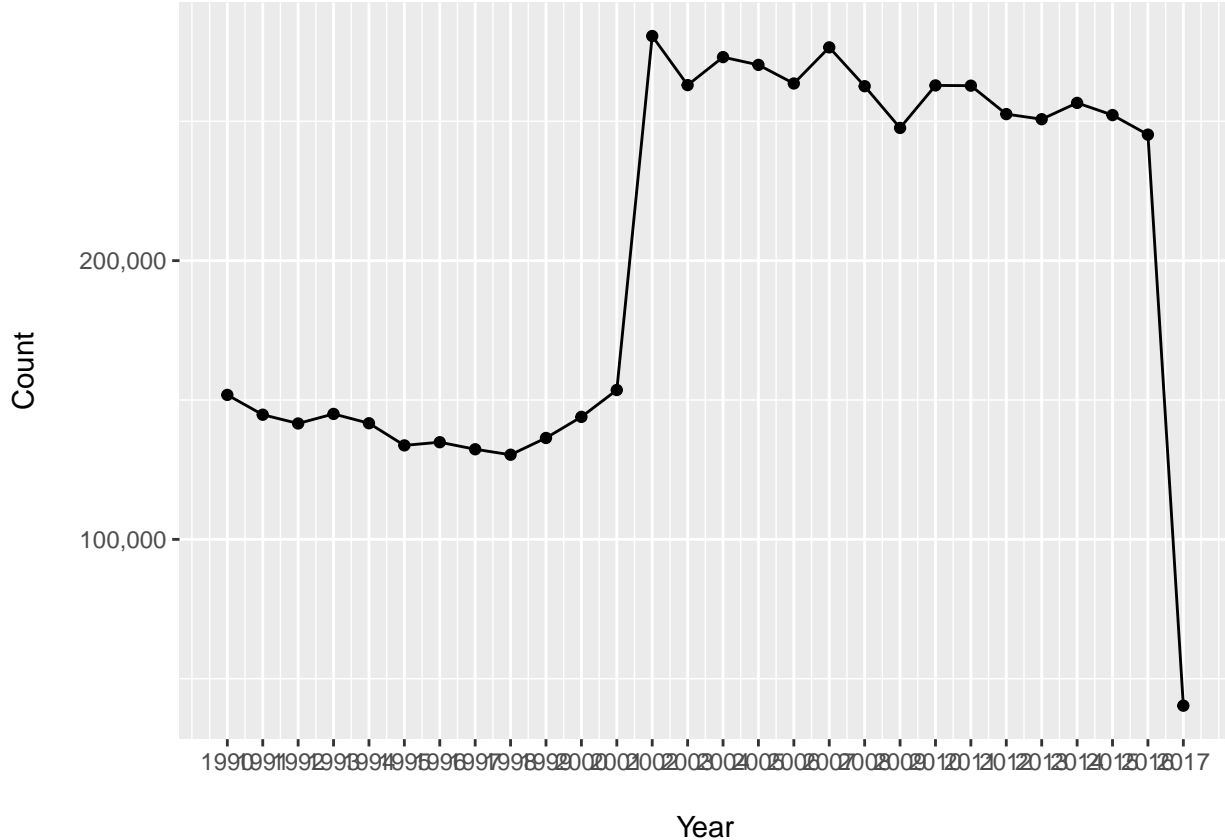


Figure 1: The total number of US flights operated over the time period 1990-2017 worldwide

We observe that there is a sudden jump in the total number of flights in year 2002, which we will investigate later.

The airlines with the highest number of flights are as follows:

Carriers	Frequency (1990-2017)
Southwest Airlines Co.	584,490
Delta Air Lines Inc.	382,952
United Air Lines Inc.	360,937
American Airlines Inc.	304,131
Northwest Airlines Inc.	237,982
US Airways Inc.	187,208
Hageland Aviation Service	181,154
Continental Air Lines Inc.	179,991
ExpressJet Airlines Inc.	179,178

Carriers	Frequency (1990-2017)
Federal Express Corporation	155,514

Now, let's explore which of the se airlines had the most number of flights in each year, starting from 1990 to 2017.

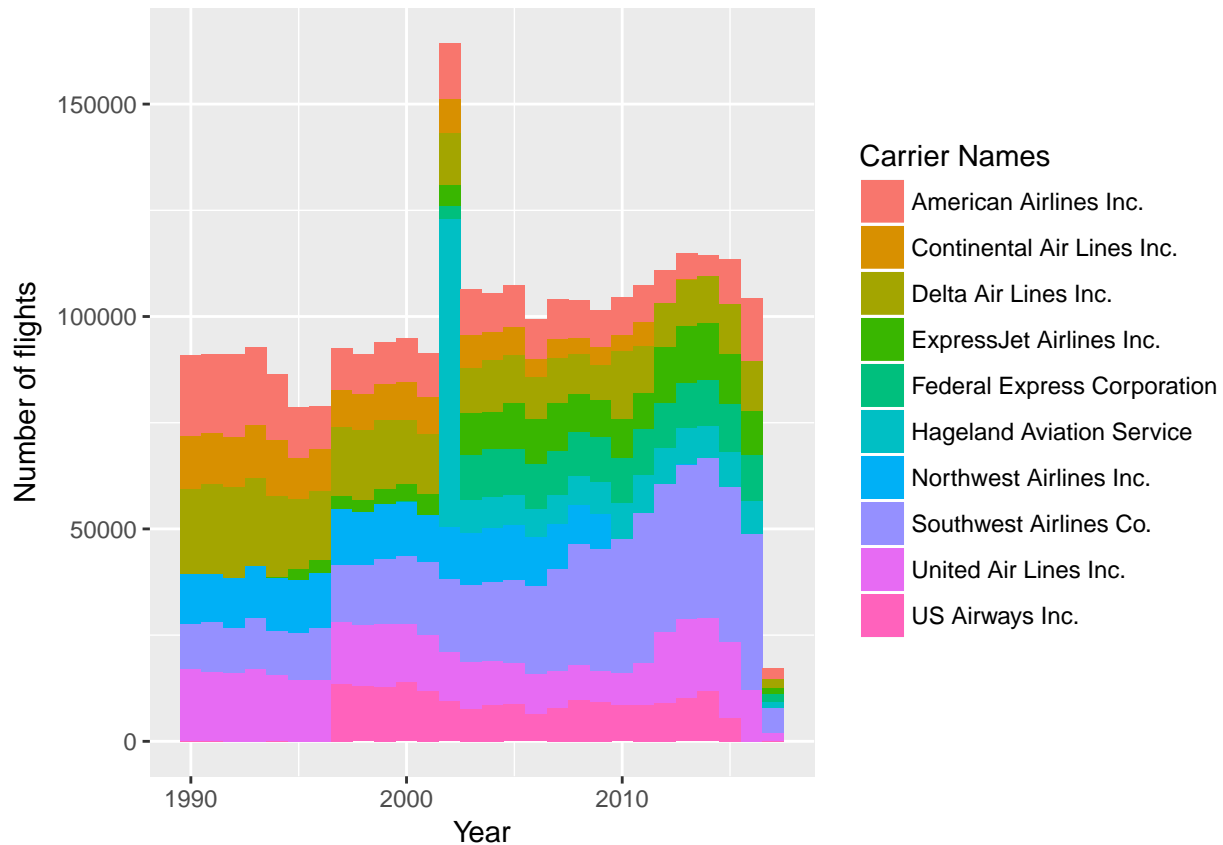


Figure 2: The number of flights operated yearly by the biggest overall carriers from 1990 to 2017

Here, we can clearly see that *Hageland Aviation Service* carrier had a sudden jump in 2002, maybe because it started operating in that year. After that, it went down and the total counts as seen above still don't explain why the jump occurred in Figure 1 which continued 2002 onwards.

We also notice that *Delta Airlines* and *US Air* were the largest carriers in the early 1990s. *Hageland Aviation Service* was the largest operating carrier in the year 2002. *Southwest Airlines Co.* became the largest airline from 2003 onwards.

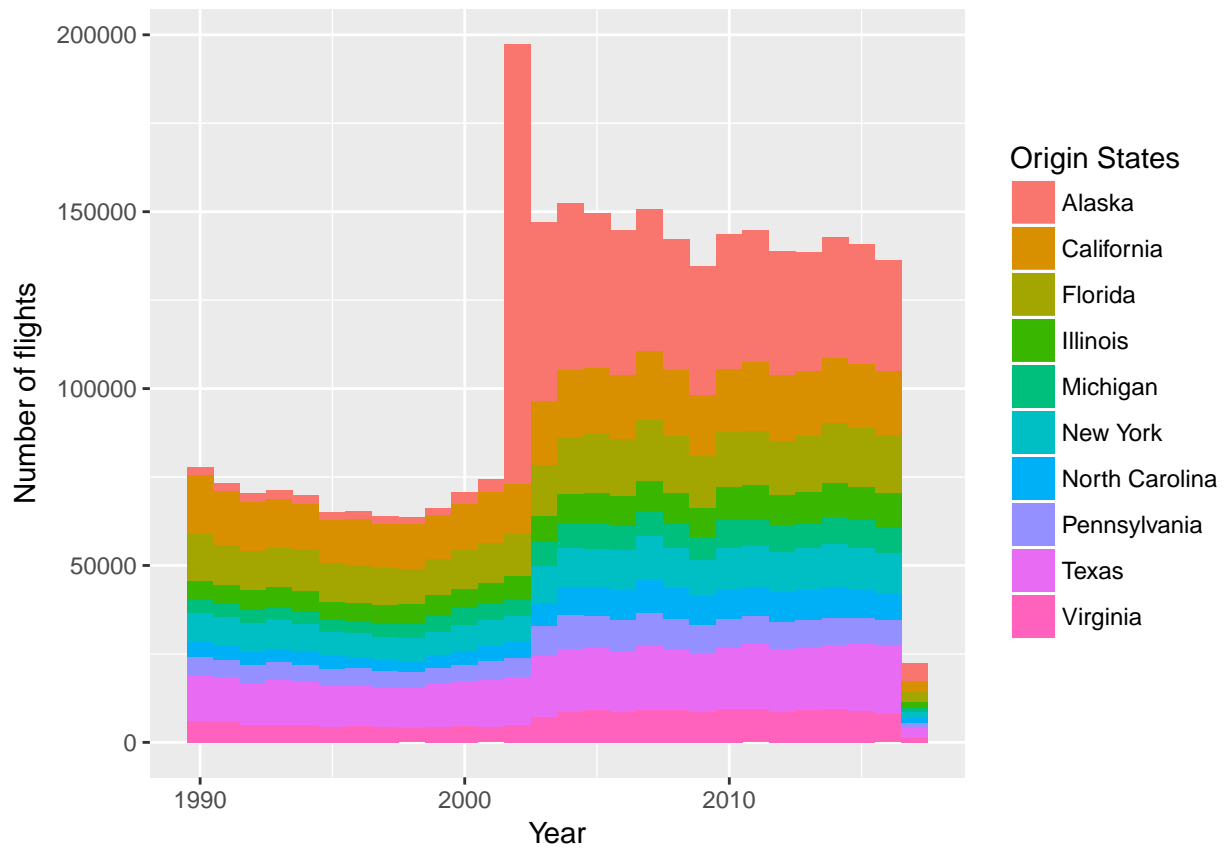
Now, let's see which states had the most originating and arriving flights.

Origin States	Frequency (1990-2017)
Alaska	695,195
California	437,959
Texas	412,686
Florida	370,848
New York	256,794
Illinois	195,898
Virginia	188,639
Pennsylvania	179,642
North Carolina	164,889

Origin States	Frequency (1990-2017)
Michigan	155,284

Destination States	Frequency (1990-2017)
Alaska	695,604
California	446,391
Texas	412,406
Florida	362,781
New York	253,848
Illinois	195,914
Virginia	186,974
Pennsylvania	179,738
North Carolina	161,590
Michigan	153,555

Now, let's explore which of these states had the most number of originating flights in each year, starting from 1990 to 20



Now, let's explore which of these states had the most number of arriving flights in each year, starting from 1990 to 2017.

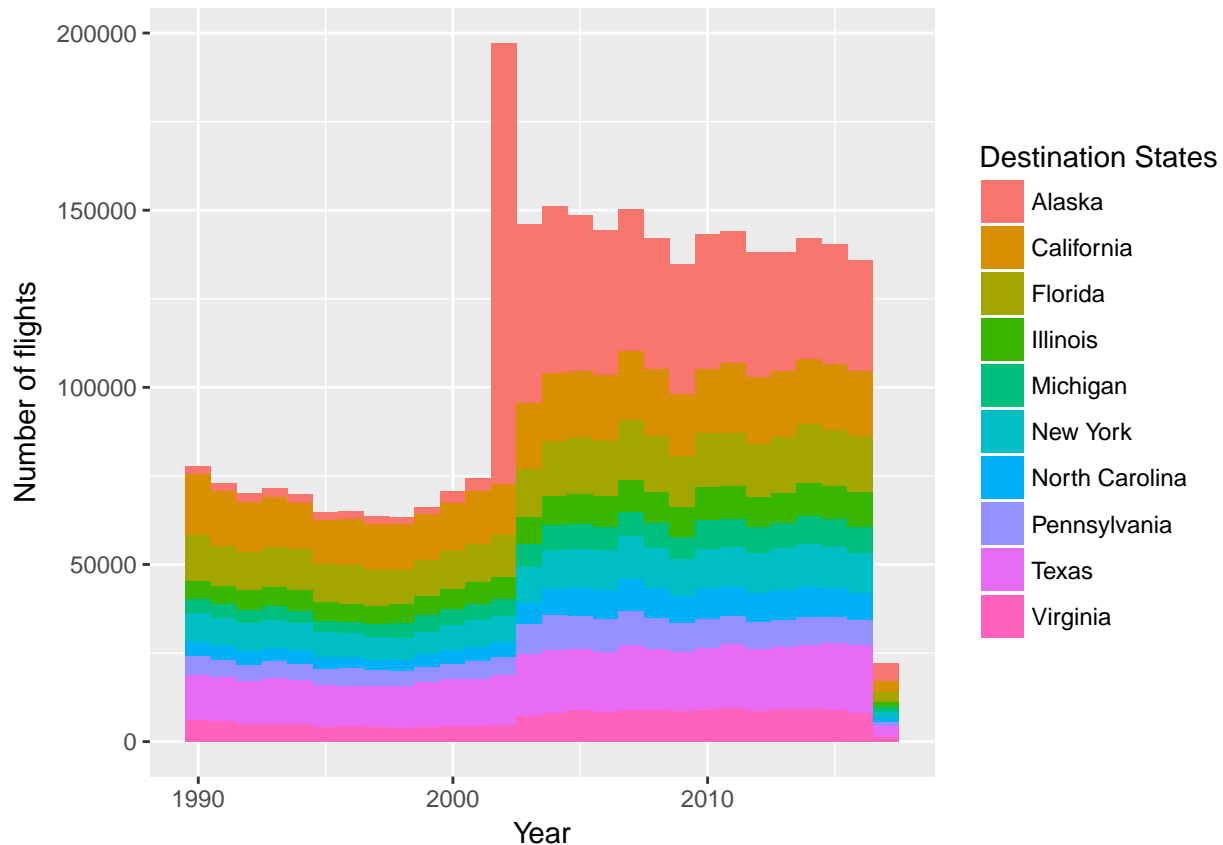


Figure 4: The number of incoming flights operated yearly by destination states

We notice surprisingly that the distribution of origin states and destination states for airlines is almost identical. Before 2002, *California* was the state with the highest number of incoming and outgoing flights.

We notice that carriers started operating in *Alaska* from 2002 onwards, after which Alaska became the state with the highest incoming and outgoing flights.

This is strange! There is a spike in number of flights in 2002 as well, when *Hageland Aviation Service* started operating. Let's explore this further by finding out where did *Hageland Aviation Service* start operating.

We can see that the total number of *Hageland Aviation Service* flights operating in 2002, the number of flights operating within Alaska that year of *Hageland Aviation Service* and the total number of flights operating within Alaska that year respectively were as follows:

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1 72623

## # A tibble: 1 x 1
##       n
##   <int>
## 1 72602

## # A tibble: 1 x 1
##       n
##   <int>
## 1 123372
```

Since Alaska has the most operating flights, let us provide a summary of which flights operate the most

within Alaska and try to explain the constant large number of operating flights from 2002 onwards, as seen in Figure 1.

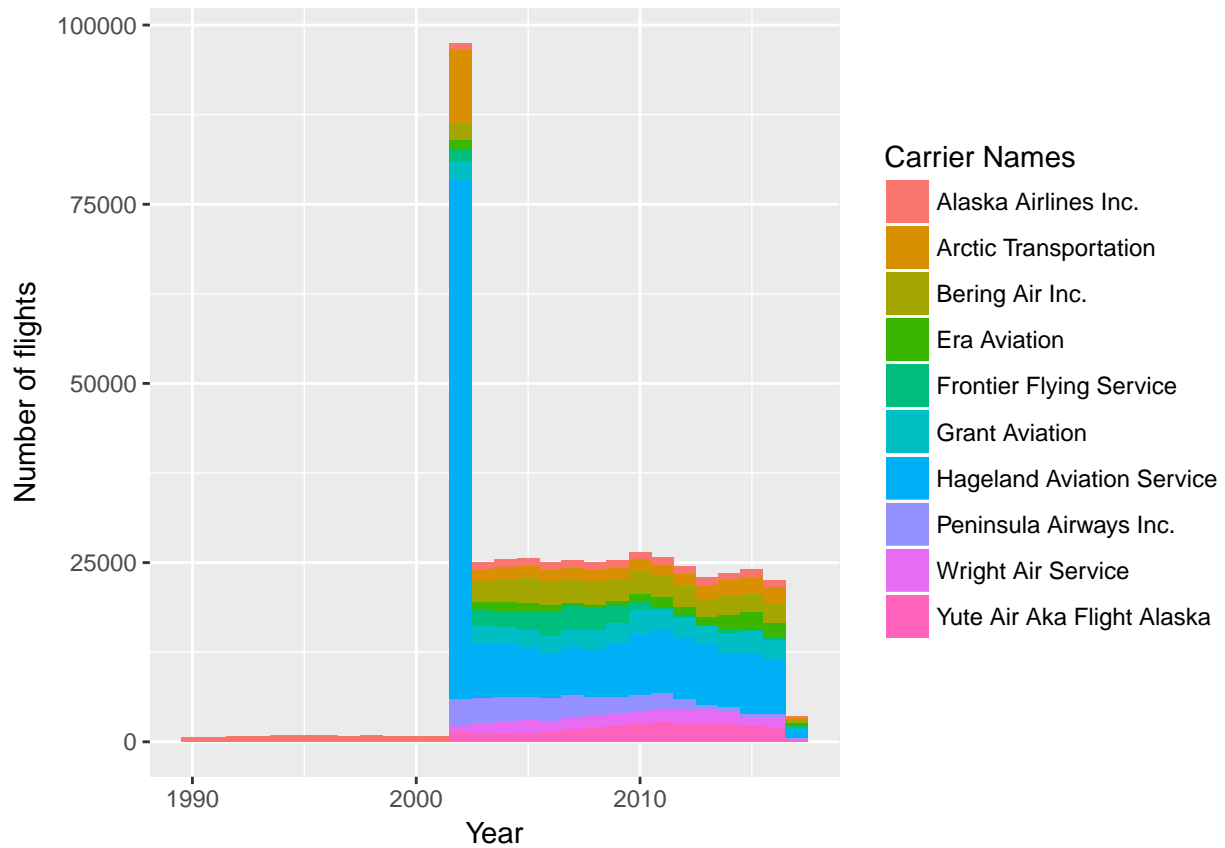


Figure 5: The number of flights operating within Alaska of different carriers

Here too, we see an anomaly in 2002, when the number of flights being operated by *Hageland Aviation Service* were anomalously high!

Let us now explore other data in the airlines dataset. Let us explore during which quarters did passengers travel the most during 1990-2017.

Quarter	Number
3	4,125,732,866
2	4,073,941,246
4	3,869,384,297
1	3,737,795,958

We can see that passengers traveled the most during Quarter 3. Let us see during which month the most amount of travel occurred.

Month	Number
7	1,468,977,397
8	1,445,326,860
6	1,411,623,287
3	1,370,378,671
5	1,351,506,757
10	1,328,177,409

Month	Number
4	1,310,811,202
12	1,281,606,200
11	1,259,600,688
9	1,211,428,609
1	1,197,957,428
2	1,169,459,859
We can see that the most amount of travel occurs in July and August.	

Let us see during which month did the most travel occur over 1990 to 2017.

Month and year	Number
7/2016	66,261,322
7/2015	65,296,846
6/2016	64,883,680
8/2016	63,651,050
7/2007	63,642,342

Thus, we observe that passengers traveled the most since 1990 in July 2016, followed by July 2015, followed by June 2016.

The year in which the most travel occurred is as follows:

Year	Number
2,016	720,910,688
2,015	698,121,080
2,007	681,492,975
2,014	665,149,401
2,006	660,642,163
We can see that the highest amount of travel occurred in 2016, followed by 2015 and 2007.	

Thus, passengers traveled the most in Quarter 3, in months July and August, specifically the highest ever in July 2016 and in the year 2016.