Flight Paths and Number of Flights

4.4.1 Mapping Flight Paths

In the section below we explore, compare, and contrast the total number of flights and their paths by year, state, and airport.

We constructed from the *airlines* data several Hive tables that were useful in visualizing the frequency and direction of individual flight paths. From the Hive table of the original cleaned *airlines* data, the unique combinations of the variables *origin* and *destination* were extracted and joined with *lat* and *long* variables from the *airports* data set in order to get the coordinates for each airport for each possible flight path. Two joins were necessary to get the coordinates for both airports. This was done individually for each year as there could be differences in the specific routes completed in a given year. The count of the number of outgoing flights per airport was also calculated in a separate table that is also utilized for the following flight paths maps. Outgoing flights were used as a measure of an airports' busyness

	origin	dest	no_diverted
1	ORD	LGA	148
2	LGA	DFW	114
3	STL	LGA	99
4	DFW	LGA	91
5	DAL	HOU	85
6	ATL	LGA	82
7	MIA	LGA	74
8	DTW	DFW	72
9	PBI	LGA	71
10	SFO	SAN	71

and level of air traffic, as it was assumed this number would be in proportion to the number of incoming flights by airport as well. The first new Hive table is the format below:

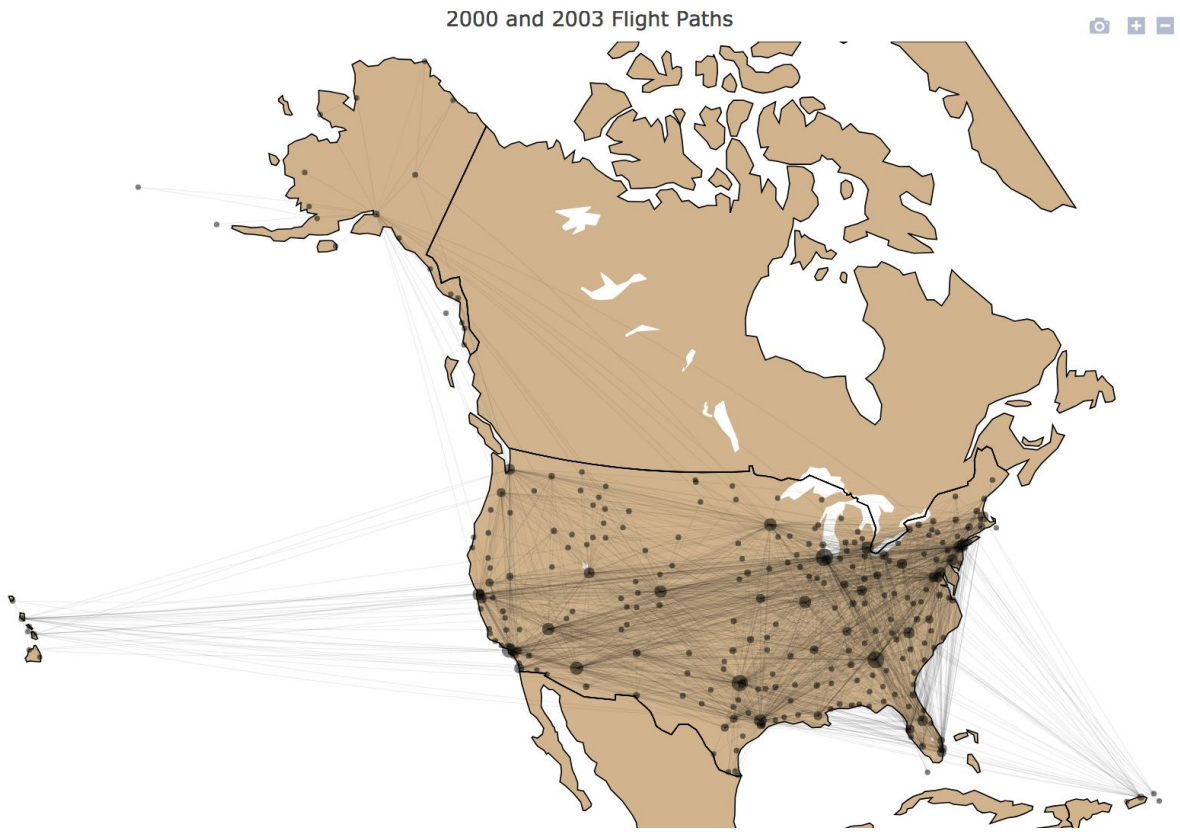
	flightpaths2.origin	flightpaths2.lat1	flightpaths2.long1	flightpaths2.dest	flightpaths2.lat	flightpaths2.long
1	ATL	33.64044444	-84.42694444	ABE	40.65236278	-75.44040167
2	CLE	41.41089417	-81.84939667	ABE	40.65236278	-75.44040167
3	CLT	35.21401111	-80.94312583	ABE	40.65236278	-75.44040167
4	DTW	42.21205889	-83.34883583	ABE	40.65236278	-75.44040167
5	IAD	38.94453194	-77.45580972	ABE	40.65236278	-75.44040167
6	MCO	28.42888889	-81.31602778	ABE	40.65236278	-75.44040167
7	MDT	40.19349528	-76.76340361	ABE	40.65236278	-75.44040167
8	ORD	41.979595	-87.90446417	ABE	40.65236278	-75.44040167
9	PIT	40.49146583	-80.23287083	ABE	40.65236278	-75.44040167
10	AEX	31.32737167	-92.54855611	ABI	32.41132	-99.68189722

And the second appears as:

	airports_complete.origin	airports_complete.lat	airports_complete.long	airports_complete.ct
1	ORD	41.979595	-87.90446417	634004
2	ATL	33.64044444	-84.42694444	616914
3	DFW	32.89595056	-97.0372	564045
4	LAX	33.94253611	-118.4080744	424716
5	PHX	33.43416667	-112.0080556	351597
6	IAH	29.98047222	-95.33972222	312803
7	DTW	42.21205889	-83.34883583	285717
8	MSP	44.88054694	-93.2169225	285474
9	LAS	36.08036111	-115.1523333	276947
10	DEN	39.85840806	-104.6670019	272590

This was sorted by the number of flights to give an idea of the busiest airports. The major hubs appear to be Chicago's O'Hare, Atlanta's Hartsfield–Jackson, Dallas/Fort Worth International Airport, and Los Angeles International Airport.

Using these two tables, the data was imported into R and visualized with ggplot2. The following maps highlight the sheer volume in the number of flights across and between the years we are looking at. The sizes of the circles at the airport locations (not every single airport location was plotted to reduce map clutter) represent the number of outgoing flights from that airport, an indicator of the number of patrons using an airport and its general busyness. The map for both 2000 and 2003 is below:

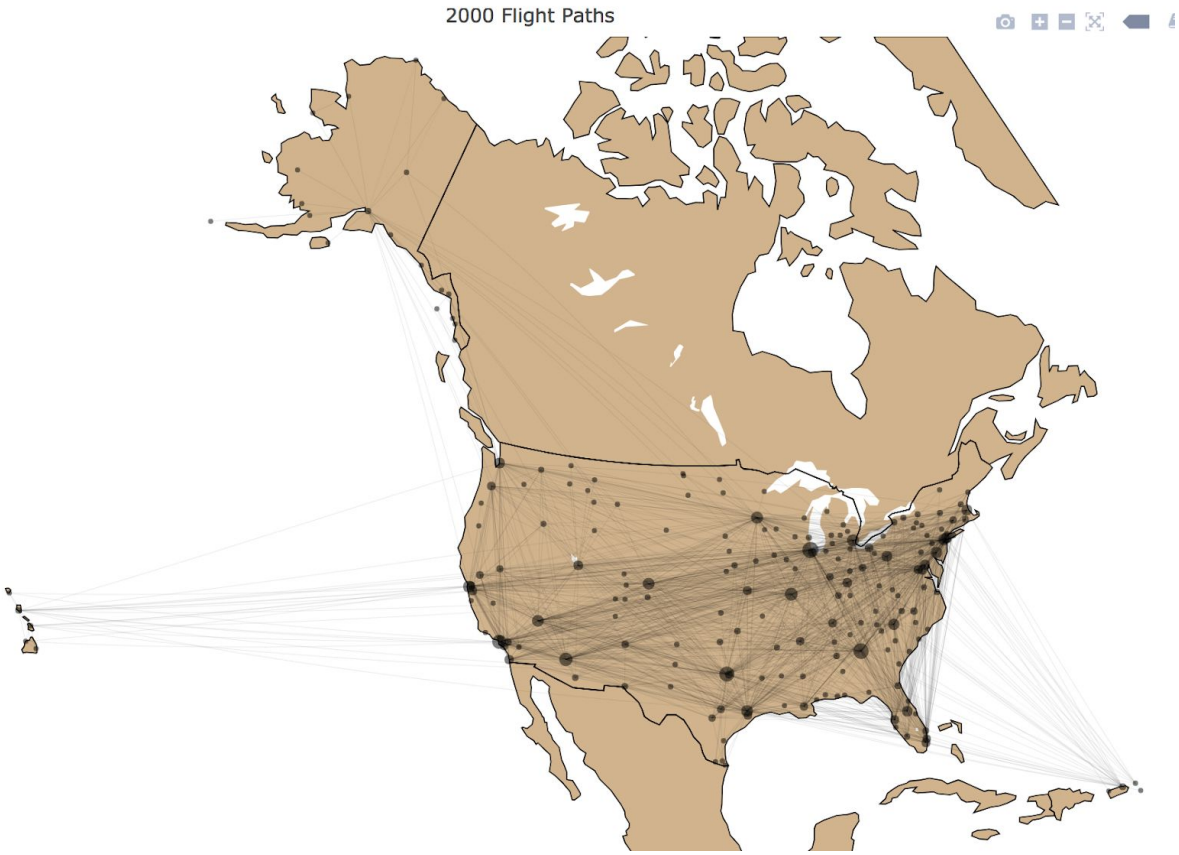


The alpha value of the flight path lines was altered to give a clearer view of the lanes of air travel receiving the most traffic. Looking at the map, it is evident that there are in general two major flyways: one east-west bound from the Northeast, which is densely populated by airports, to the Western of United States (Colorado, Arizona, Nevada, California, and by a lesser extent Hawaii), and another running north-south stretching from Chicago, Minnesota, and the cities of the east down to Florida, Atlanta, and the airports of Texas. This includes the area of the country that contains the highest density of flight paths: a triangle between Chicago, New York, and Atlanta. These were identified as hubs in the earlier Hive table, and others evident include DFW and IAH (Houston) in Texas, and several smaller circles in the west like LAX (LA), LAS (Las Vegas), and DEN (Denver), and multiple airports in the San Francisco bay area.

Also note the large number of flight paths running from the southwest to the northeast and vice versa, which appears to be as many if not more than those following a path from the northeast to the southeast. This could be explained by the large population centers in both of these corners of the country.

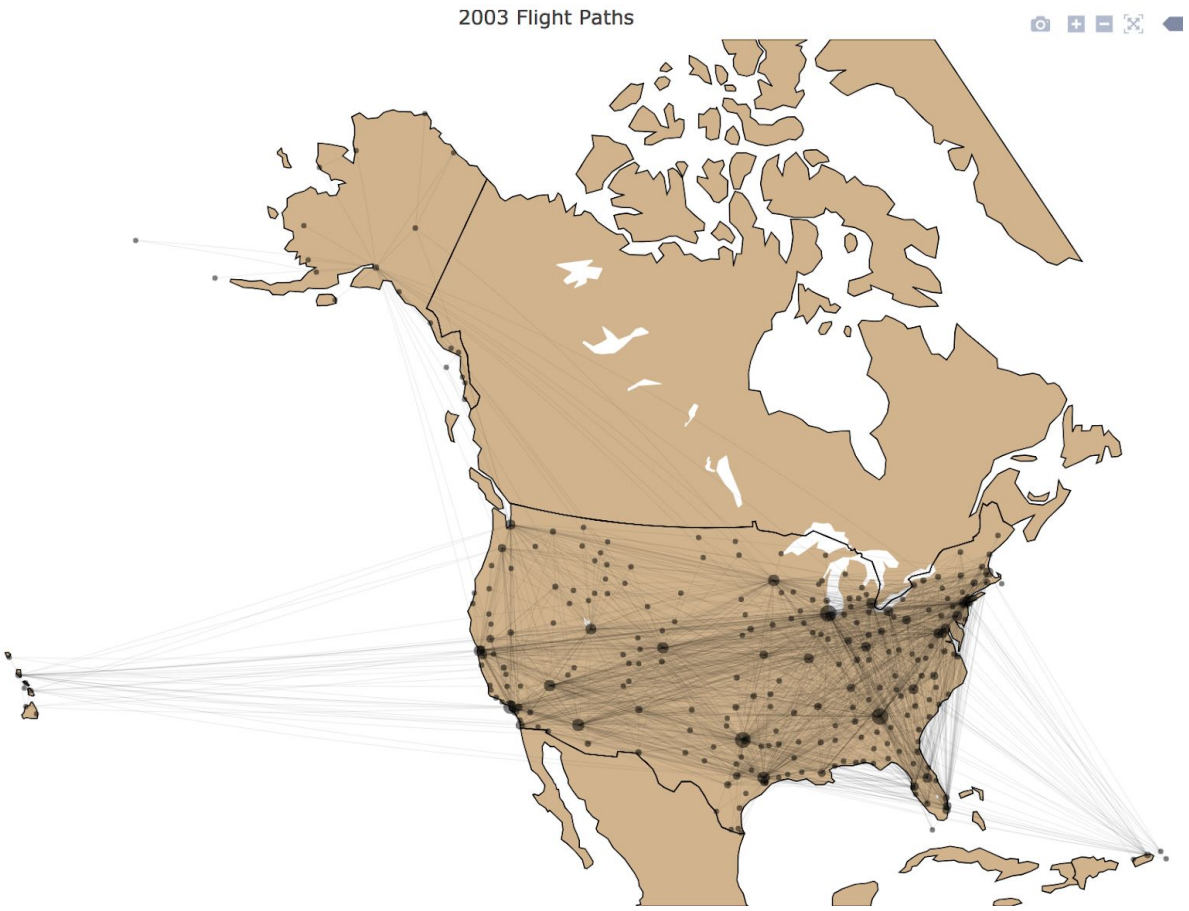
It is also interesting to note that the dataset included flights to US territories, including the US Virgin Islands and Puerto Rico. The longest flight paths are those between AK and HI and the contiguous US, with most of those to HI appearing to depart from the west coast in CA.

Specific flight path and airport data was extracted from these tables to plot flight paths of the individual years. The plot for flight paths in 2000 is below:



Compared to the overall flight paths map, the map for 2000 is obviously less dense with less flights and less airport destinations than the full map. This difference was queried and we found that there are 206 airports in 2000 and 282 in airports 2003 that appear either as a destination or origin. Long distance paths appear to be present in both. A noticeable missing airport in this map is EYW, or Key West International, that was present in the overall map. After investigating this, one possible explanation was that the airlines included in this dataset did not serve that airport during this year. This is somewhat doubtful, as we did not find available information on EYW that suggests flights were running abnormally during that year. This remains suspicious to us after some querying showed that there were 785 outgoing flights from EYW in 2003 but 0 in 2000 and research confirmed the airport was operational in that year.

The flight paths for 2003 is seen below:



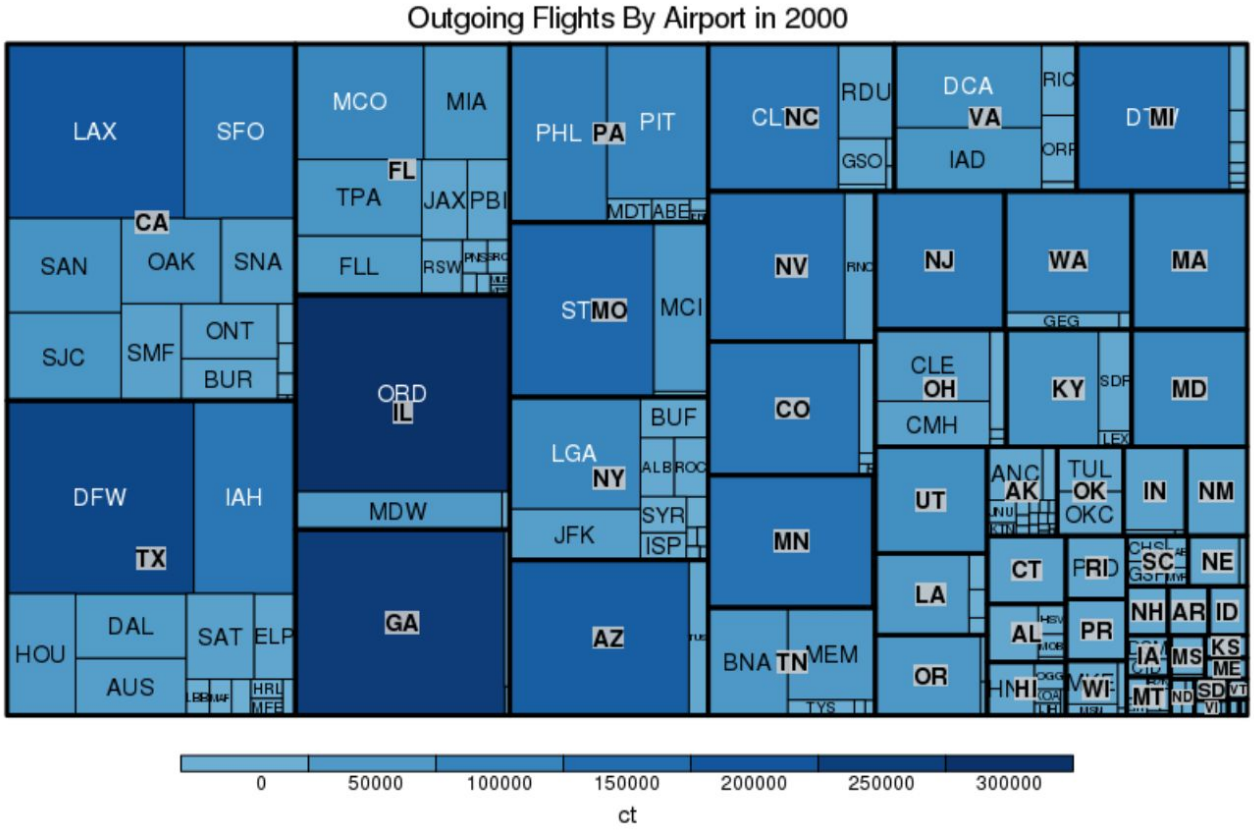
Comparing this map to the last one for 2000, this projection shows modest growth in the outgoing flights for most of the major hubs, like ORD, ATL, and DFW. Smaller airports by flight traffic also were identified as increasing, including Salt Lake City International Airport, Orlando's major airport, and the string of airports in the cities of the Northeast (Washington, New York, Boston). In terms of flight paths, this year has all of the same discernable destinations as the full map. It does a good job of highlighting the straight, uninterrupted flights going between those major airports to the west and east of the Great Plains over the commonly termed "flyover" states.

It is important to remember external variables or data that may not be included in this data set during our analysis. The impact of international flight numbers on flight patterns and number of flights is not represented here but can be assumed to account for a not insignificant portion of flights at the United States' many international airports (ORD, SFO (San Francisco), etc.). Also, airports nearest to the neighboring countries would intuitively harbor more international commuters. Thinking about these effects can give a more holistic view of the data that presented here.

4.4.2 Number of Flights By Airport and State

Next, we continued to investigate the number of flights, now looking to organize the volume of flights by airport and state. To accomplish this, Hive tables for each year were constructed to include the number of outgoing flights for each airport, obtained from the last analysis, with an included *state* variable. To visualize these tables, tree maps were constructed that used number of outgoing flights as the block size and color. This is an easier way to compare the sizes of the circles in the last sections' maps as well as gain any possible insights from grouping by *state*.

The tree map for the 2000 data is below:



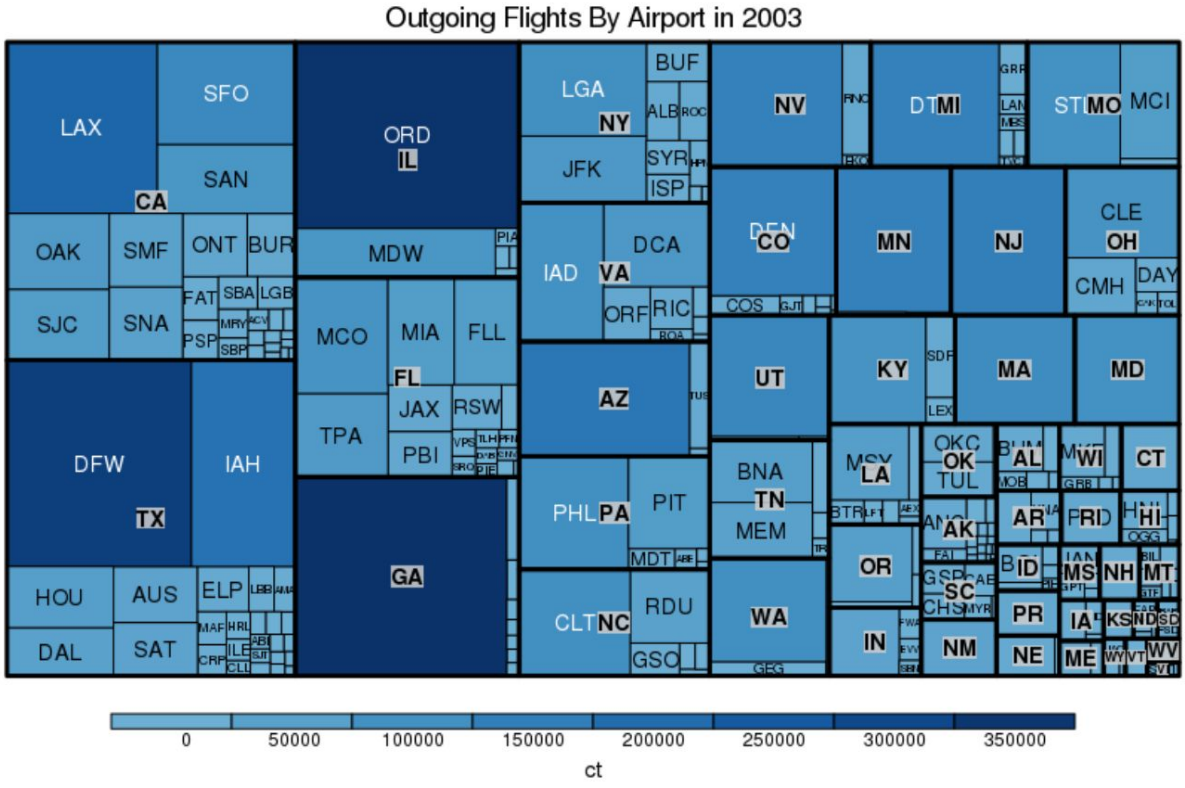
The hubs identified in the flight paths maps are noticeable here: ORD, ATL, DFW, and LAX dominate the count coloring which is not unexpected. CA and TX have the most number of outgoing flights, and combined they host more than all of the outgoing flights in FL, IL, and GA combined.

Looking at the blocks by state, we can identify several interesting characteristics. The top two airports for number of outgoing flights are not located in the states that lead in the number of outgoing flights. These are ORD in Chicago and ATL in Atlanta. This could be explained by the geography these airports share:

they have central locations and are presumably airports commonly used during layovers in multi-leg flights.

Many states, no matter their total amount of outgoing flights, are singly dominated by a large airport usually in a metropolitan area. Examples include: CA (LAX), TX (DFW), GA (ATL), IL (ORD), AZ (PHX), and states with a medium amount of air traffic like NV (LAS), CO (DEN), and MN (MSP). It should be noted that many of these dominating airports are also international and not regional airports. Even the states with the least amount of outgoing flights share this characteristic, for example, CT, OR, NH, AR, etc. It seems as if airports within states rarely share the share of flight traffic evenly. The only state that comes close would be FL, which has Orlando's airport with the most outgoing flights, followed by Miami's and Tampa's airports respectively. This could be explained by the state having several large population centers spread out across the state. Pennsylvania is also an exception, almost split evenly between Philadelphia and Pittsburgh, as well as Virginia, which has two major airports serving the Washington, D.C. Metropolitan area.

The tree map for the same data in 2003 is below:



Comparing this tree map to the last map, several differences are apparent. For overall state trends, the blocks for TX, CA, IL, FL, and GA are larger in general. This can be attributed to the increase in the number of flights and the rising populations of the areas in which these airports serve. An overall shrinking is observed for the state blocks for medium outgoing flight traffic, like MN, NJ, CO, UT, NV, MI, AZ, etc. I do not consider this as much as an effect of a lesser amount of outgoing flights in these states, but rather their proportion to the whole number of outgoing flights for all states has decreased, with the major hubs and states gaining more and more of a proportion.

One state in particular seemed to have shrunk in number of outgoing flights more than its neighbors had between years. Missouri's STL airport was probably in the top ten highest number of outgoing flights in 2000 but in 2003 it appeared to have shrunk drastically, bringing the state down with it. After some research the following information was discovered:

> "The September 11, 2001 terrorist attacks were a huge demand shock to air service nationwide, with total airline industry domestic revenue passenger miles dropping 20% in October 2001 and 17% in November 2001. Overnight, American no longer had the same need for a hub that bypassed its hubs at Chicago and Dallas, which suddenly became less congested. As a result of this and the ongoing economic recession, service at Lambert was subsequently reduced over the course of the next few years; to 207 flights by November 2003. Total passenger traffic dropped to 20.4 million that same year."

Source: https://en.wikipedia.org/wiki/St._Louis_Lambert_International_Airport

The decreased utility in the location of the STL airport explains the higher than normal drop in the size of the MO block as well as the proportion that the STL airport has within it.

In the NY block, the JFK airport appears to have gained a proportion of the states' traffic while LGA's traffic decreased. This could be due to LGA's problem with overcrowding delays, which was so bad in 2000 that Congress passed legislation to revoke the federal traffic limits on LaGuardia (https://en.wikipedia.org/wiki/LaGuardia_Airport). This could have resulted in more commuters using the nearby JFK airport.

Other trends included: CLE taking over a larger proportion of Ohio outgoing flights, and the same happened for PHL in Pennsylvania. The opposite occurred for Detroit's airport in Michigan, and MCO in Florida. These patterns could have many explanations, with one being that the wake of the 9/11 terrorist

attacks changed the dynamic of flight patterns in the country and thus changed the proportion of outgoing flights for many airports by state. Again, the true causes are dubious.