

Final Project: Flights During COVID-19

Data Science I (STAT 301-1)

Noah Holubow

6 December 2020

Contents

Overview & Purpose	1
Data Sets	1
Importing & Cleaning Data	2
Initial View: Comparisons	2
Month by Month Overview	4
Airlines	7
Maps	8
COVID-19 Cases	10
Segment Characteristics	11
Conclusion	14

Overview & Purpose

COVID-19 has impacted our our way of life in countless ways. But one particularly affected industry was the airline industry. When the virus struck the United States, passenger counts undeniably took a nose dive across the entire country and world. As someone who is particularly interested in transportation and the airline industry, this project was a fun way to explore the specific impacts of COVID-19 on the airline industry from a variety of standpoints. In this project I focused on month-by-month flight segment data in the United States, including both domestic and international flights, to get a snapshot on passenger counts, freight, mail, and the like by region, segment, airline, and many other facets. I plan to explore flight data between 2019 and 2020 in the United States. This includes both domestic and international flights. Obviously COVID-19 has decimated the airline industry in many ways, but I am curious about how this impacted flights in multiple dimensions in the early months of the pandemic.

Data Sets

Eight data sets were used as part of this project. The primary data sets contained flight segment data from the US Department of Transportation (DOT)'s Bureau of Transportation Statistics (BTS).

- **2019 & 2020 T-100 all carrier segment data.** This data came from DOT's BTS and reports air carrier statistics (DOT Form 41 Traffic) for all carriers. The data tables include domestic and international segment data for all carriers, including the aircraft type/service class; passenger, freight, and mail counts; number of departures performed; aircraft hours; and load factor. Because the data is only available by years, I included 2019 and 2020. The 2020 data went through May, so all 2019 data was restricted through May as well. Both sets were merged and analyzed together. Each observation is a particular month's segments (route) data for a given airline, aircraft, etc. Data can be found [here](#).
 - 2019 All Carrier Data
 - 2020 All Carrier Data

- **Data code tables from BTS.** Many of the columns in the carrier segment data include codes, such as carrier group, aircraft group, aircraft configuration, and aircraft type. To ease the data analysis, I obtained the code keys as individual data sets to merge in the main data. These data sets come from the same source as above.
 - Carrier Group
 - Aircraft Group
 - Aircraft Config
 - Aircraft Type
- **COVID-19 cases by state.** Because I am primarily investigating the pandemic’s effect on flights, I wanted to include case counts by state. I obtained daily case counts by state from the New York Times’ COVID-19 Github repository. I then was able to determine the daily and monthly totals. The data set can be found [here](#).
- **State population count.** It is important to recognize that larger states (e.g., New York, California, Texas) will naturally have larger case counts. As such, I obtained the state population counts in 2019 (as a metric extrapolated from the 2010 Census and other sources). This information was obtained from the US Census Bureau at [this link](#).

Importing & Cleaning Data

The following section imports and cleans the data sets to be used for analysis later on but is hidden for purposes of keeping the report tidy. A significant portion of the project worked on cleaning the data, manipulating it, and joining in other data sets to make it more comprehensive.

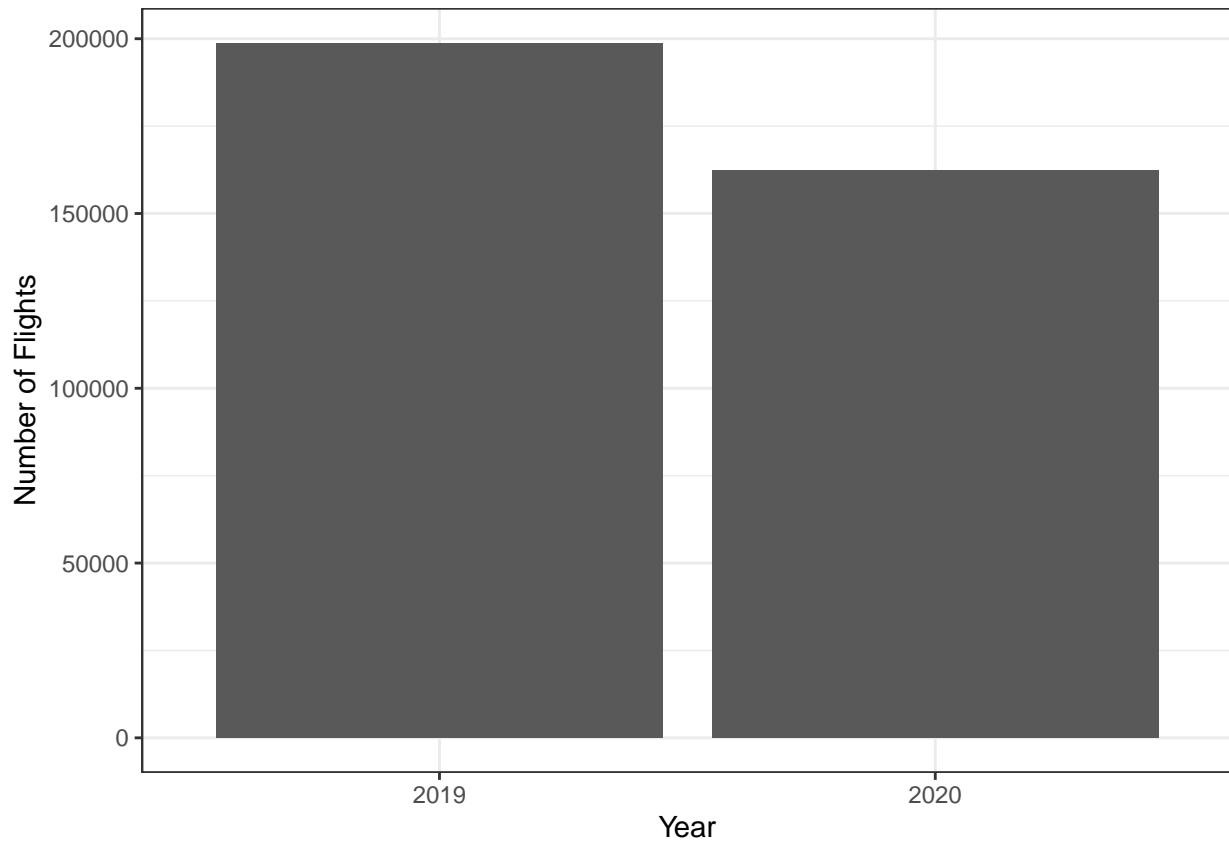
Prior to walking through charts and graphs, it’s worth taking a moment to look briefly at a condensed version of the cleaned data set I used for the majority of this project:

Month	Year	Carrier	Origin	Destination	Seats	Passengers	Freight	Mail
3	2019	United Air Lines Inc.	SFO	YVR	142	142	0	0
3	2019	United Air Lines Inc.	SLC	PHL	166	166	27	0
3	2019	United Air Lines Inc.	SMF	SFO	150	149	0	0
3	2019	United Air Lines Inc.	SNA	CLE	126	126	1	0
3	2019	United Air Lines Inc.	SPN	GUM	166	115	0	0
3	2019	United Air Lines Inc.	SRQ	BWI	179	62	9263	0

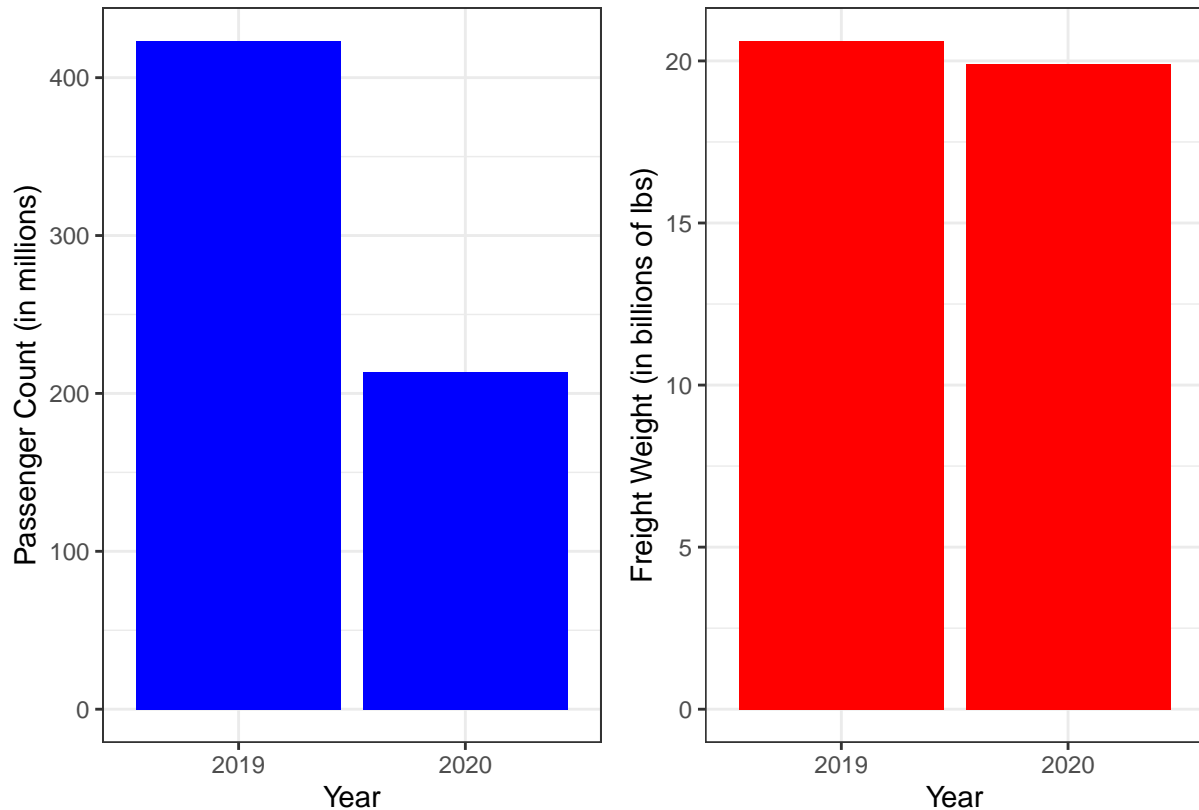
As you can see, this set includes flight segments by month, year, carrier, origin, destination, and other parameters that define what a “segment” is. The key metrics include variables seen in the preview, such as departures scheduled and payload. These allow us to view statistics on how a given segment performed during that month/year.

Initial View: Comparisons

First, I want to confirm my original hypothesis, that flight counts (i.e., total number of flights) are indeed down this year. To do this, we can use a simple bar graph just to compare between 2019 and 2020.



At first glance, this doesn't look terribly informative. Flights are definitely down to an extent, but by eyeballing it only appears to be something like 20% down. Still, it's a significant decrease. But going into more specific details is helpful. My original guess was that passengers were more significantly impacted than freight since people started staying home as states increasingly shut down, but there were huge demands for orders, including PPE, home deliveries, and critical supplies. Comparing passengers (left) and freight (right) confirms this.



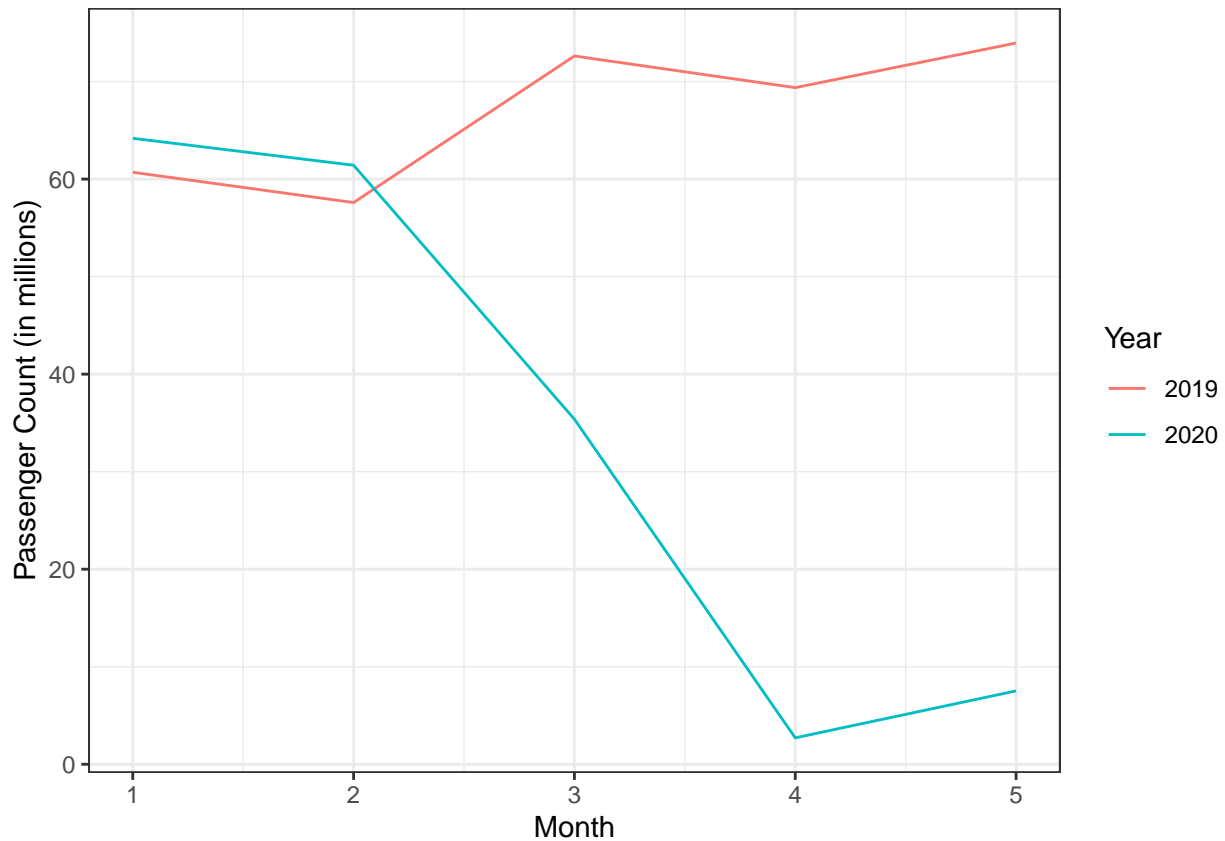
These graphs indicate that passenger counts were hit more heavily than freight. In fact, freight being moved to/within the US was not appreciably lower compared to 2019. But it's also important to note that during the first two months of 2020, there was no significant change in lifestyle in the US, which means that the real bulk of the change likely occurred in March, April, and May. This is covered in the next section.

Month by Month Overview

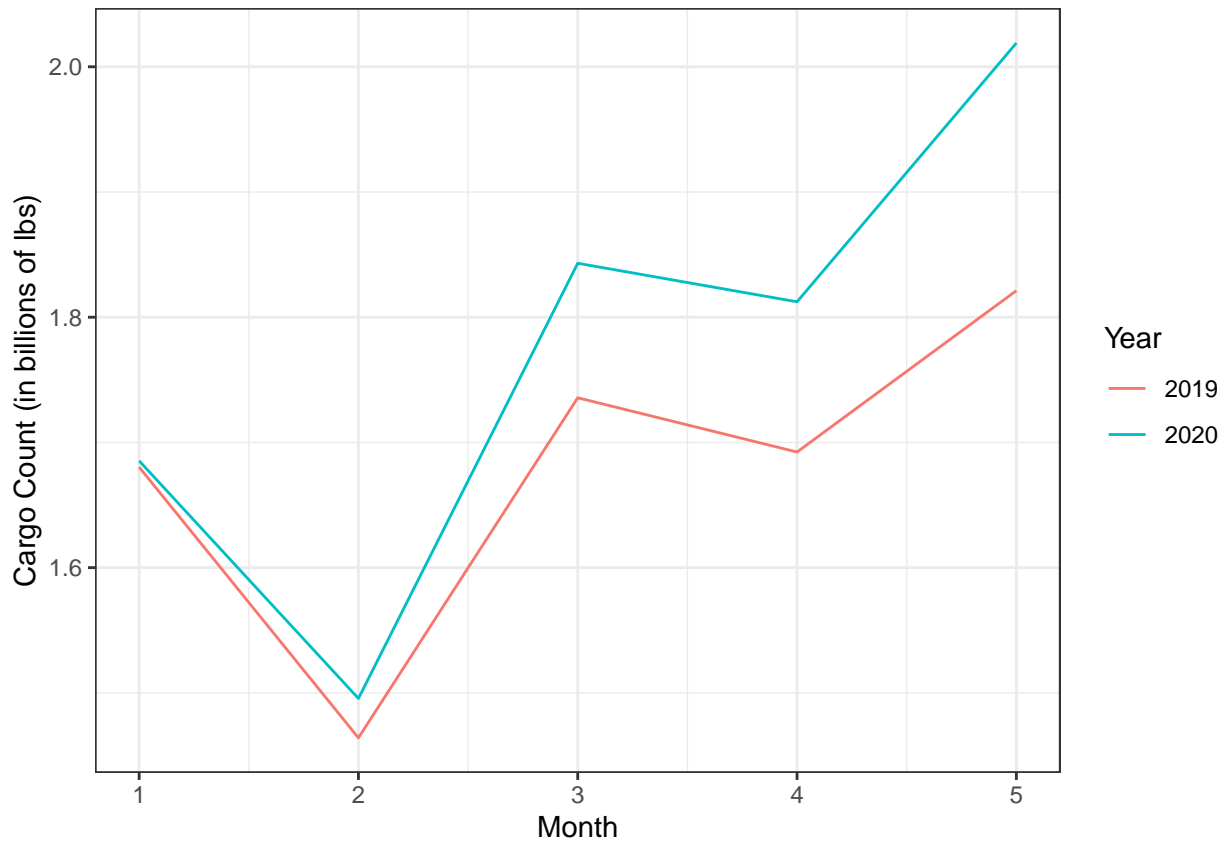
According to the CDC, the first lab-confirmed case of COVID-19 was confirmed on January 20, 2020. On January 31, the Department of Homeland Security and the CDC restricted flights arriving from China. The first coronavirus-related death occurred on February 29, 2020 in Kirkland, Washington. On March 16, President Trump issued guidelines urging people to avoid social gatherings of more than ten people and to refrain from discretionary travel. And on March 18, California Governor Gavin Newsom issued the US' first statewide stay-at-home order. This was followed by similar measures in many other states (e.g., South Carolina on March 19 by Governor Henry McMaster and Illinois & New York by Governors JB Pritzker and Andrew Cuomo, respectively, on March 20). For a detailed day-by-day timeline of the COVID-19 pandemic in the US, I would recommend checking out this [Wikipedia article](#).

Thus it would make sense that travel would dip in March or shortly thereafter, perhaps somewhat counteracted briefly in March due to many students and individuals abroad traveling home.

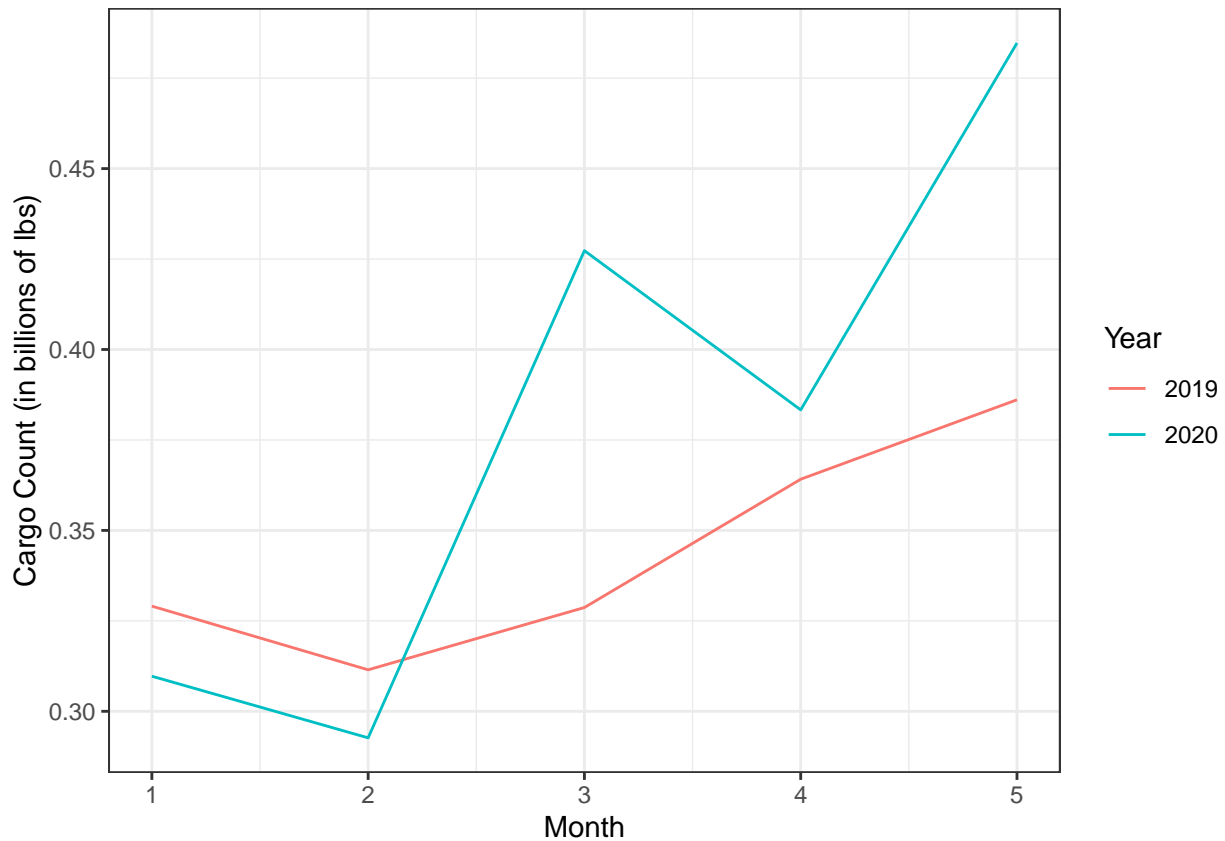
To confirm this hypothesis, we can plot passenger counts on mainline carriers (e.g., Alaska, Allegiant, American, Delta, Frontier, Hawaiian, JetBlue, Southwest, Spirit, United) traveling on passenger revenue-scheduled flights.



A first glance at the graph is quite shocking. But going from left to right, it's actually interesting to note that travel was up slightly in 2020 for the first two months of the year. There isn't a clear explanation for that, but at least it is clear that it was not deterred entirely by the fear of the pandemic initially since it was still greater than in 2019. But then there is a clear drop of about 50% from February 2020 levels. Referring back to the brief timeline, this would make sense since the situation got worse in March and states began to shut down. But the significant shutdowns and the time for them to take effect occurred in the latter half of March, which is why passenger counts still didn't fall as much until April, when they nearly hit rock bottom (i.e., nearly 0 passengers). There was a significant spread between April 2019 and April 2020 passenger counts (nearly 70 million). Travel then picked up slightly in May but not by much. We can see that a different trend is true for cargo.



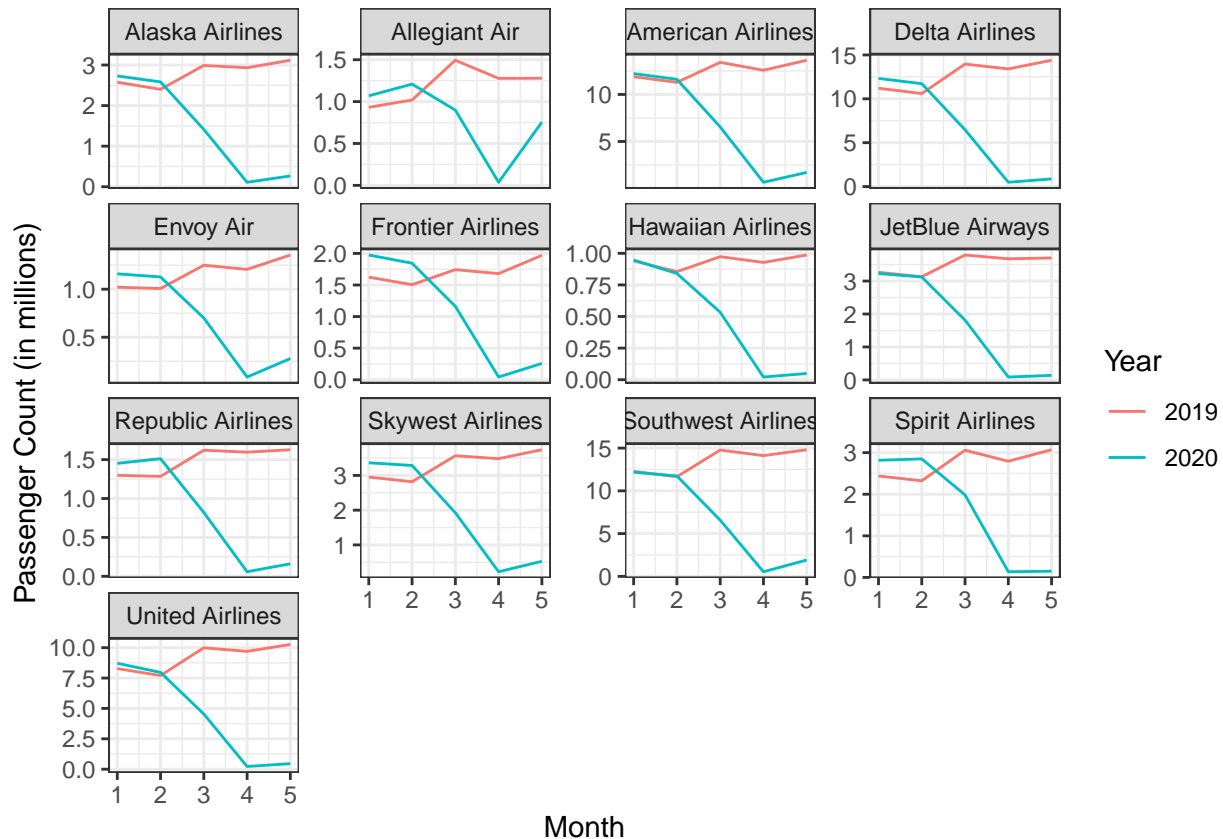
Whereas passenger counts were down, cargo count by weight was actually up, increasingly so in March, April, and May. I believe that this can be explained by the fact that with stay-at-home orders, resulting in reduced mobility and fear of shopping, more people and companies placed orders. The original graph of freight above showed freight from all flights. It is common for commercial, passenger revenue flights to carry some freight along with baggage and mail. But looking at the original chart of freight, that there was only a very slight increase in freight counts. In this graph I am looking at flights filed with the DOT as “scheduled cargo service” or “non-scheduled civilian all cargo service.” This way, we are looking at cargo carried on cargo-exclusive flights. It is true that many of these carried essential PPE from China. I was curious about how reliant we became on other countries, so I looked at all cargo flights that did not originate in the US.



This shows that, compared to the general cargo plot shown above, the increase is much higher in 2020 (particularly in March), likely indicating frantic efforts to import supplies and goods when the pandemic actually became a daily reality in the US (again, around that time).

Airlines

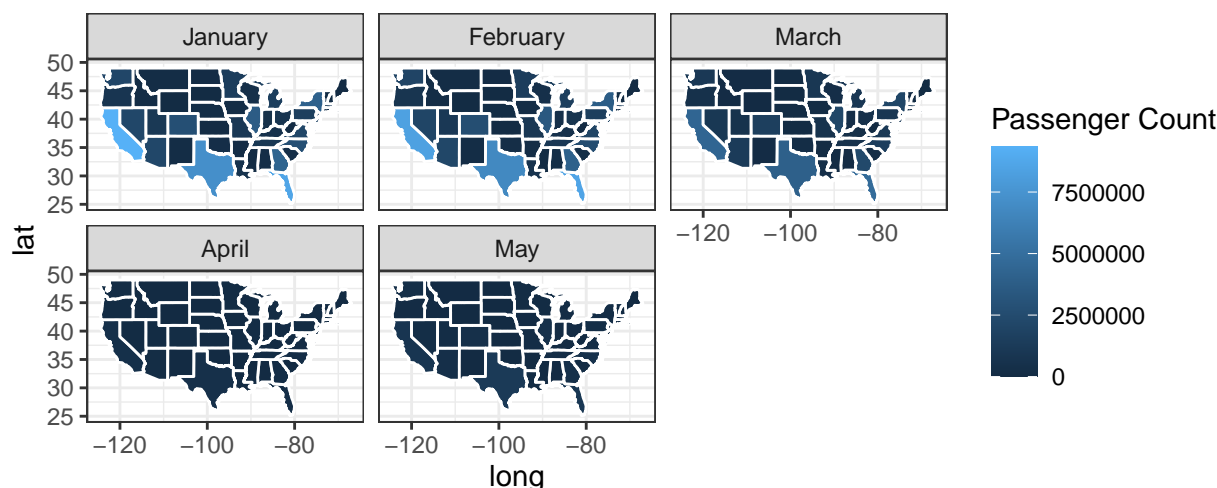
Passenger travel was unambiguously down. But I wondered if some airlines saw more success than others. Using a similar approach to look at passengers on mainline commercial carriers, I sorted by each airline:



It appears that no airline was immune to the COVID-19 pandemic. Most have a similar trend. They had apparently more success in 2020 during January and February but then took a hit beginning in March. Recovery was mostly minimal. It is interesting that Allegiant seemed to have a much more impressive recovery in May 2020 compared to all other airlines. A bit of background research on the company shows that Allegiant is a low-cost, very leisure-focused airline with most of its routes in small and mid-sized airports. This makes sense because a large portion of passenger count, particularly in off-summer months, comes from business travelers. Business travel is not picking up quickly, and it certainly was not doing so in the early months of the pandemic (i.e., May). But perhaps a reason Allegiant has managed to increase its passenger counts so quickly is because people are still flying for leisure to an extent and in greater numbers/proportion than business travelers. Allegiant is also not an international carrier, so there was not an international market it had to recover.

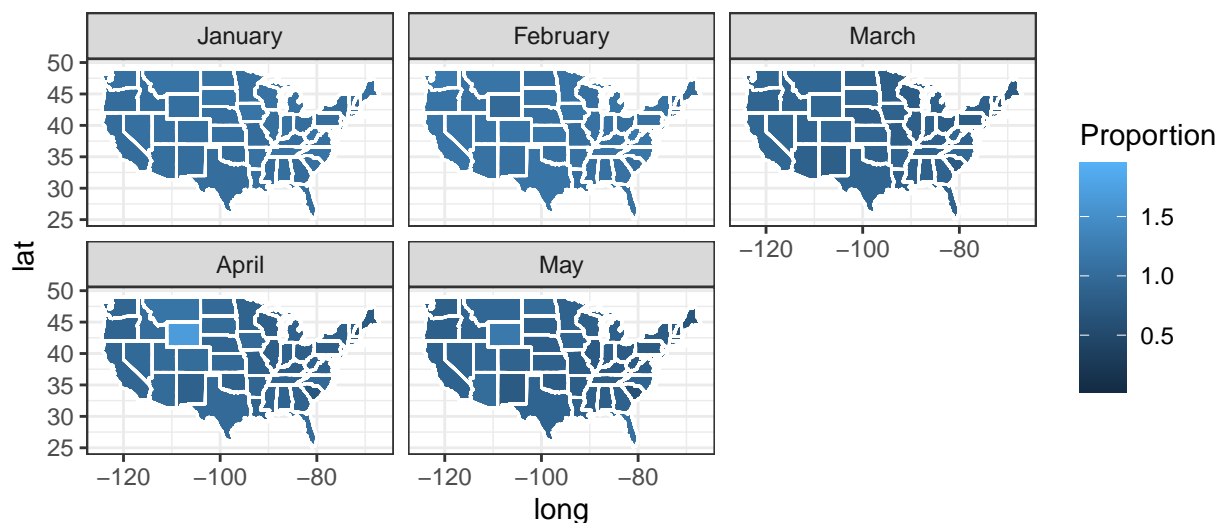
Maps

I thought it would then be interesting to look at flights by state. First, we will look at states' flight counts by month, where counts are the number of flights departing a given state.

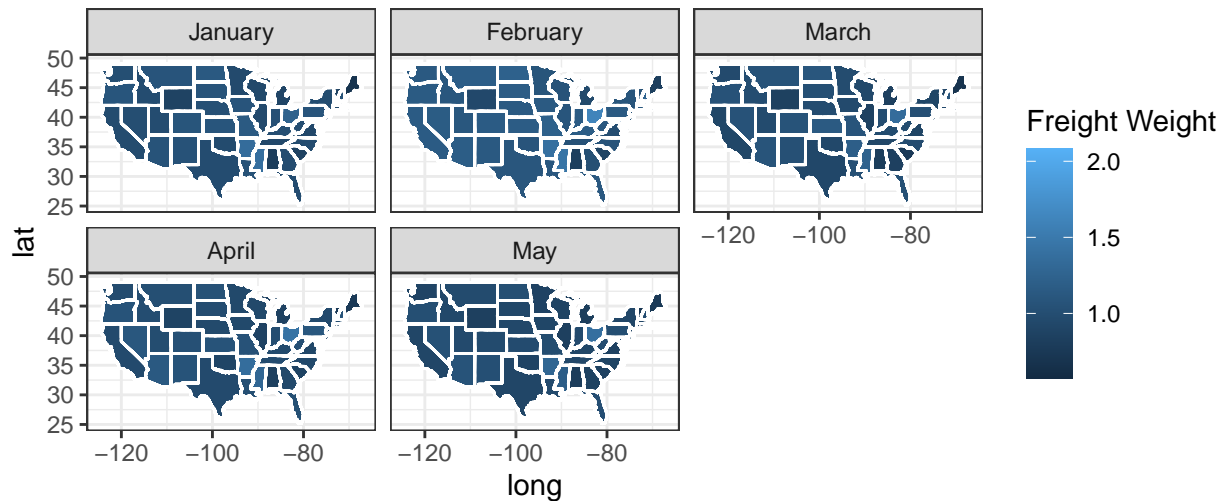


The results here are not terribly surprising. Because the states are shaded according to total passenger counts, it is no surprise that a state like North Dakota is always a very dark color since, in normal circumstances, it has significantly fewer flights and passengers compared to states like California, Texas, or Florida, which stand out in bright blue. Starting in March, we see those larger states look more like the smaller states (i.e., now California has similar shading as North Dakota). It looks like some states pick up a bit in May, such as Texas. But this graph isn't fully telling because it looks at mere passenger counts and is hard to differentiate differences in smaller states.

We now shade states by proportions rather than counts. The proportion represents the fraction of 2020's flight metrics (in this case passenger count) in a given state on a given month divided by the same metrics in 2019. For example, if in 2020 5 passengers flew to a state compared to 10 the previous year, the ratio would be $\frac{5}{10}$, or 0.5. We apply this each month.



This plot is much clearer because we can see that most states experienced similar trends. I apply the same type of graph with freight:



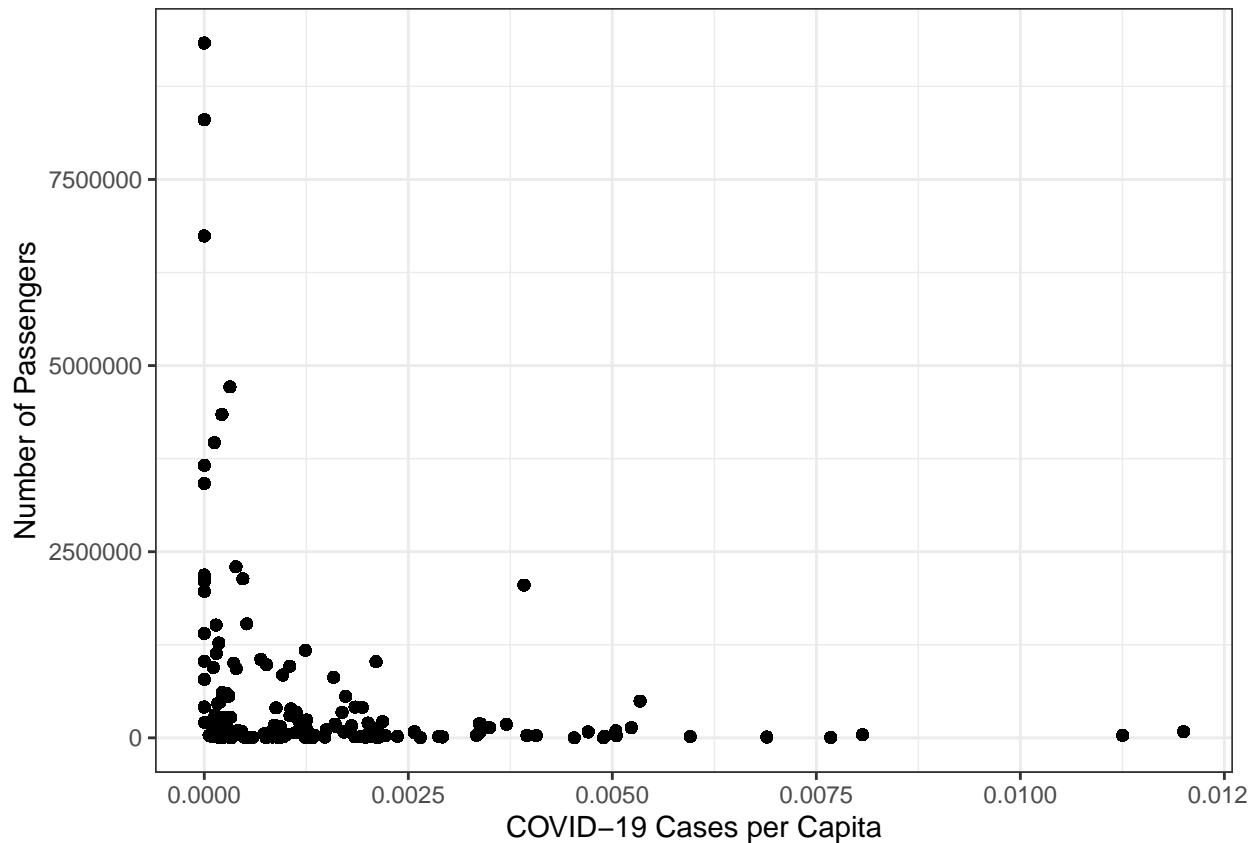
This chart brings up an interesting question. Why would larger states, like California and Texas, have lower ratios than some of the smaller states? I didn't really think this would be the case beforehand, but it makes sense that larger states that have lots of distribution centers may see a decline in those services/products. Smaller states that don't normally high freight imports likely had special shipments of PPE and other essential goods that they were not used to receiving. It's not that larger states didn't get these essential items, but comparatively they lost some of the imports since the large 2019 numbers were likely based mostly on economic goods for sale.

We can verify this by looking at the individual statistics for the 5 highest proportions in March, April, and May:

Month	Year	Destination	2019 Passengers	2020 Passengers	Proportion
4	2019	wyoming	31701	61877	1.951894
5	2019	wyoming	44364	53210	1.199396
4	2019	montana	153794	177529	1.154330
5	2019	florida	7967170	8556311	1.073946
4	2019	south dakota	71033	75490	1.062745

COVID-19 Cases

As mentioned earlier, I also obtained COVID-19 case information by state from the CDC. Putting this together, I investigated the relationship between number of cases and passenger travel, again for passengers on mainline flights. I calculated the per capita rate number of cases per state and plotted it against the number of passengers.



Though the trend isn't clearly linear, it definitely appears that for most states, more travel occurred in states that exhibited lower per-capita COVID-19 numbers. This is a good sign, although it is clear that there are some outliers. The one major outlier is in between 0.0025 and 0.0050 for covid per capita and just below 2,500,000 passengers. Further inspection reveals the state:

Month	Destination	2020 Passengers	Total Cases	COVID-19 Per Capita
3	new york	2050135	76211	0.0039176

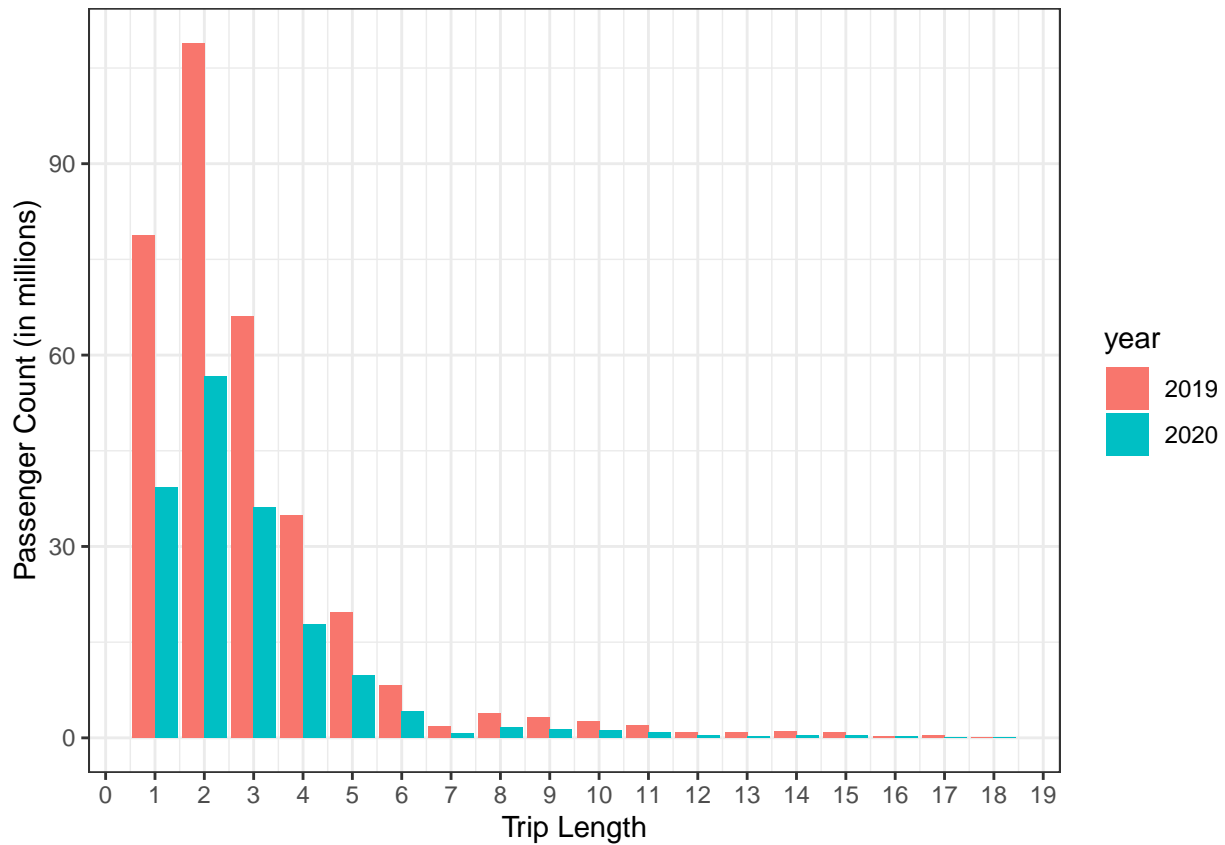
We see that New York had an abnormally high number of passengers in March considering its covid per capita value. But this makes sense considering a few factors. For one, New York City was by far the first worst-hit city in the US. So it had the virus spreading rapidly before people even knew not to travel or what exactly to do since the pandemic had really only become a reality then or soon before then. New York City is also incredibly dense and is the center for lots of business and finance. Therefore, before people realized the necessity to absolutely not travel, there was still a fair bit of travel.

Segment Characteristics

The BTS provides information as part of the T-100 data set on various characteristics/features of each segment. These include trip length, region, and aircraft type/number of jet engines. I look at trip length and region in detail.

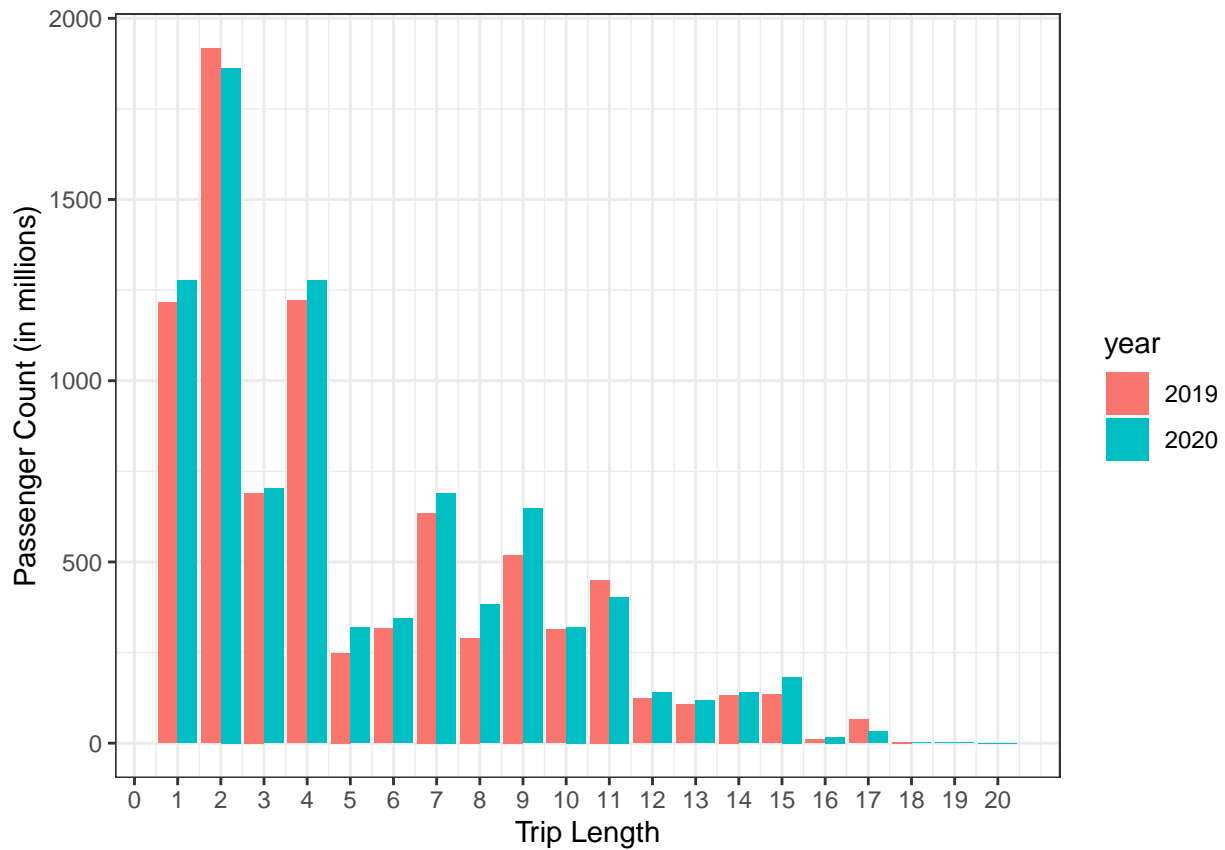
Commercial Mainline Carriers

Here I look at commercial mainline carriers by trip length:



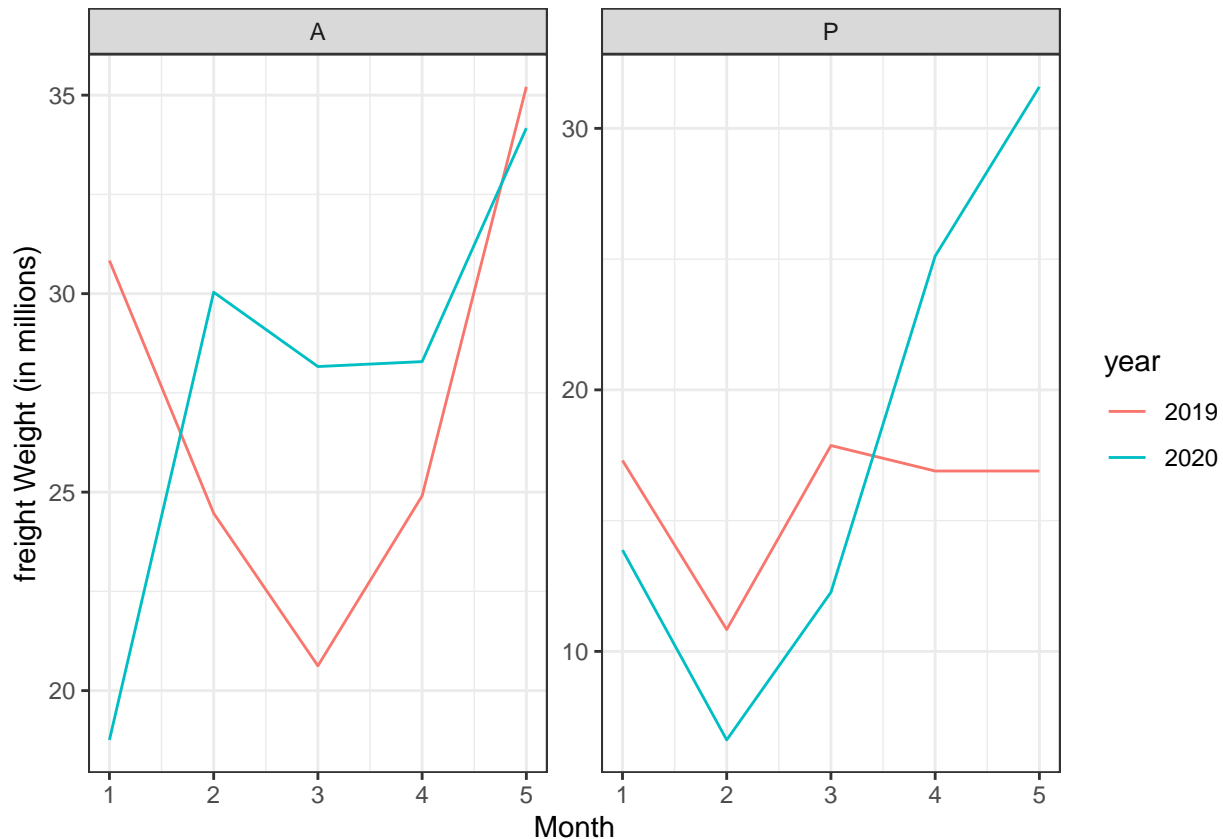
The results are not entirely surprising. The “trip length” column is a categorical, ordinal variable that identifies the distance of the trip. Exact ranges for each value are available from the BTS but are trivial for this explanation since they are listed in ascending order. Each distance group observes a decrease of about 50%.

However, it is interesting to look at the same type of graph for freight, again on scheduled cargo flights.



The most notable differences in this graph occur in distance groups 7-9. Originally I thought that this might be China to California, but the distance from Shanghai, which is one of the largest shipping cities, to Los Angeles is approximately 6479 miles. Region 9 ends at 4499 miles. However, doing a bit of research and prior knowledge indicated this uptick was likely was cargo going to Alaska. Shanghai to Anchorage is 4304 miles, which falls right within distance group 9. This was the largest increase among distance groups, and it makes sense why this would be the case. Overall, cargo amounts seemed to mostly be up for all distance groups, so this confirms what we found earlier, but it is interesting to note exactly where some of the largest increases were coming from.

We can also observe Asian vs. European imports, taking advantage of the “region” column of the segment data. I specifically looked at inbound Atlantic and Pacific, indicating European and Asian cargo imports, respectively.



What we see is that for Atlantic flights, there is a not a terribly noticeable trend. It is true that February-April saw a sustained increase over 2019. But this trend eroded in May. For Pacific cargo flights, March was fairly similar and a bit below, but there was a very steep increase in cargo than normal, likely indicating a much heavier than normal reliance on Asia for PPE. We know there was a shortage around this time, but it looks like the US was getting as many goods (both PPE and non-PPE) as possible.

Conclusion

In conclusion, I don't think the fact that less passengers flew and there were greater amounts of cargo imported is all that exciting or profound. But what is interesting is the set of insights that these statistics and graphs allowed us to understand about what the COVID-19 pandemic has done to the airline industry. In particular, some highlights included Allegiant Airlines coming out in the strongest position for recovery based on certain characteristics of its flights, the effects of being the first "major-hit" city as far as not realizing the dangers of traveling (though importantly not necessarily flying), and the reliance on Anchorage as a key hub for imports from Asia, as well as Asia being critical in the early months of the pandemic. It is my hope that we can soon return to the normal ways of flying for business and leisure and see many of these observed effects reversed, but it cannot be denied that airlines will likely struggle until then and there will be a lot of hesitancy by passengers (including myself) to once again take to the skies.