Project Report

Title: U.S. Air Traffic Analysis

Notice: Dr. David Haynes
Author: Megan Marsolek

Date: 05/13/2022

Project Repository: https://github.com/MGIS-UMN/class-project-mmarsole

Abstract

This report is a simple exploration of flight statistics within the continental U.S. With the use of flight data (as points) in conjunction with weather and airspace locations I have computed the frequencies of flights within Class B airspace and around varying weather fronts and several other statistics. Assessments were completed with the use of ArcGIS Pro, PostGIS (SQL queries), QGIS, and Excel. Based on the results in this report I suggest future work include expanding the temporal and spatial extents and building automation processes to preprocess and load the data.

Introduction

Travel via flight has become a reliable and essential mode of transportation for today's society. With air travel's popularity it is important to keep track of flight trajectories especially around popular airport hubs (in the U.S. these often include Class B airspaces) since these areas tend to be busy with planes preparing for take-off or landing. Besides the need to monitor the real time location of flights, it is an educational opportunity to assess the general traffic trends of flights across the U.S. over time. Even simple statistics concerning frequencies of flights within certain areas is informative for pilots, airlines companies, and FAA regulators. Motivated to aid in informing these user's I am also excited to increase my experience in using computer languages and software (such as SQL, ArcGIS Pro, and QGIS) for statistical inquires.

This project is an exploratory statistical analysis of flight data. With access to Open Sky Network's current and archived Automatic Dependent Surveillance-Broadcast (ADS-B) data which provides an aircraft's position through GPS and other relevant flying metrics for air traffic control towers, I can derive simple statistics about these flights such as: the frequency of flights within U.S. Class B airspaces, average flight speed at varying altitudes, the length and frequency of flights within the continental U.S., and assess the behavior of flights near or around strong weather.

Problem Statement

The objective of this report is to provide a statistical summary analysis of flight paths within the continental U.S. on May 25th of 2020 from 12:00:10 am to 5:59:50 am GMT (May 24th of 2020 from 07:00:10 pm to 12:59:50 am CDT). Analysis primarily relies on ADS-B data but also utilizes Federal Aviation Administration (FAA) designated Class B airspace and archived Radar data as a measurement of the weather. Please see *Table 1* for more details concerning data utilized for this assessment.

Input Data

Table 1. Relevant Data

#	Title	Purpose in Analysis	Link to Source
1	ADS-B	ADS-B data records the location of a plane at a given time in points (lat., long. Alt.). Points are recorded as frequently as every 10 seconds (but are not always received). It also contains other attributes: make/model of aircraft, time, flight number, horizontal velocity, vertical velocity, etc.	Open Sky Network
2	IEM NexRad Mosaic	This is Raster data for weather (radar). It indicates where severe weather is located (which may influence flight paths). Data is recorded in 5 min time intervals. Originally measured in dBZ (decibel relative to Z) it is reprojected as pixel values ranging from 0 to 255 (8 bit-based data). See IEM Color Ramp Scale for conversion values between dBZ and 8 bit data scale. dBZ is proportional to the number of drops per unit volume and used to estimate the rain or snow intensity. NOAA has a classified scale to rank weather	Iowa State University

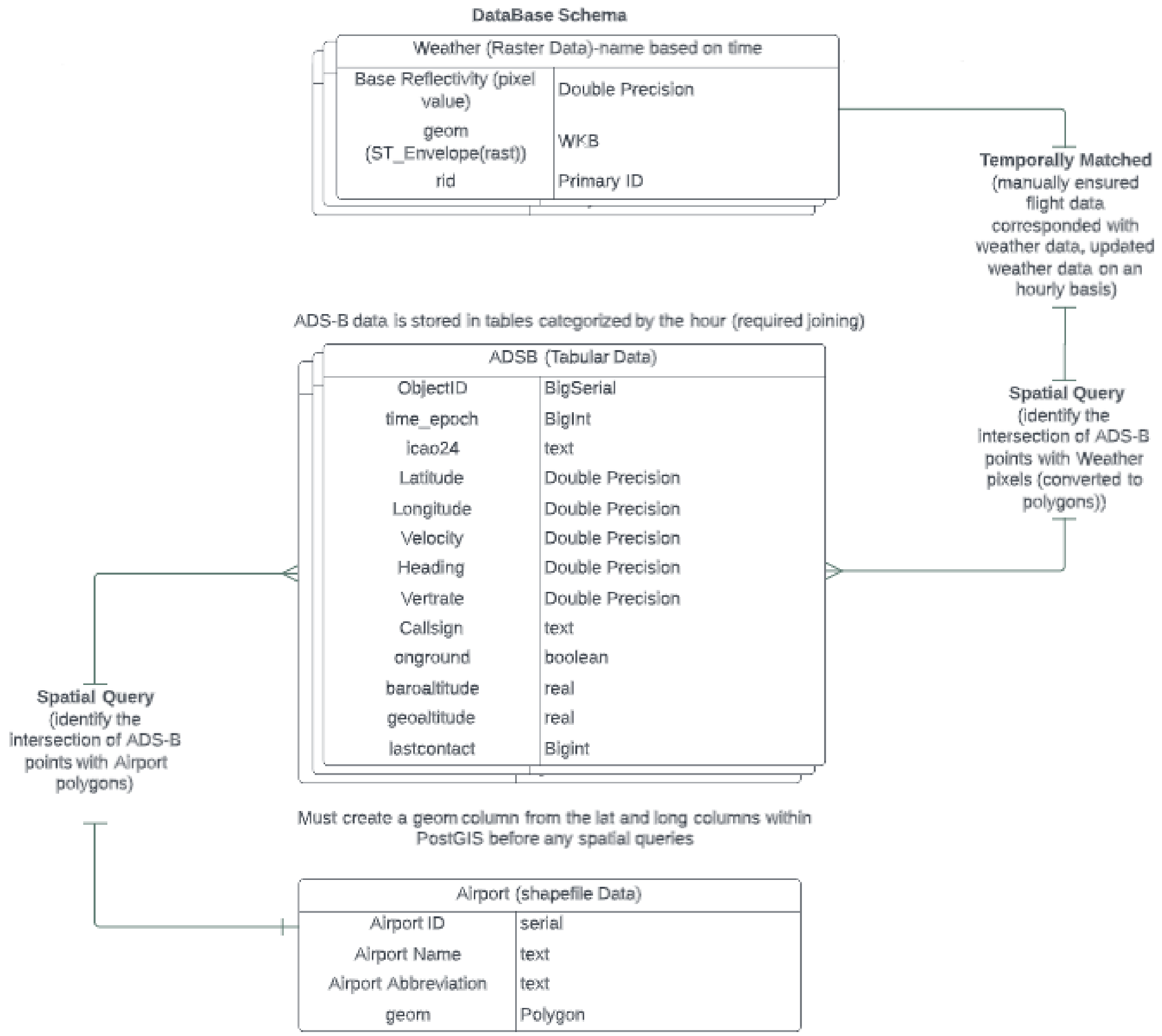
		severity based on dBZ values (see dBz (meteorology) for more information).	
3	Class B,C,D,E Airspace Shape Files	Defines where Airspace classifications are in 3D space (represented as 2D but contains attributes that describe its min and max altitudes). Class B airspace coincide with heavily used and trafficked Airports across the U.S.	Federal Aviation Administration (FAA)

Database Description

Data was loaded into PostGIS database as tabular, shapefile, and raster via tools in command prompt (such as psql, shp2pgsql, and raster2pgsql). For weather data (loaded as both raster and shapefile) and ADS-B data, there was a large number of files since each of these datasets were stored and sorted by time increments. For instance, weather data was recorded and stored every 5 mins, so I could update radar readings across the U.S. on a 5-minute interval (in practice I updated the weather data hourly), while ADS-B data was stored in tables based on one-hour increments. This meant, I loaded several individual ADS_B tables and joined them in SQL queries to assess flight data for longer time intervals which is essential since flights can fly for multiple hours (in this report I observed up to 6 hours of flight data).

As displayed in *Table 2*, I was able to make queries between weather and ADS-B data and ADS-B and Class B airspace. Using Spatial queries, I tracked the frequency of flights through various weather types (Light, Moderate, Heavy, or Extreme as designated by Air Traffic Control regulations), and their frequency among popular airport hubs (as indicated by Class B airspace). Additionally, I was able to perform some exploratory statistics based solely on ADS-B tables (such as frequencies of flights at varying altitude) and the weather data (Extracted the percentage of observed weather classes in a given weather raster).

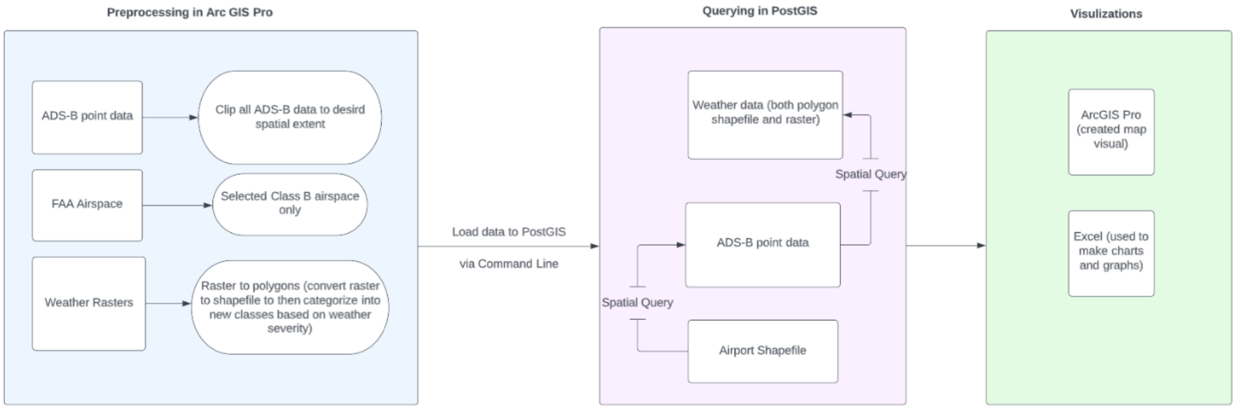
Table 2: Entity Relationship Diagram. Displays the relationships between the 3 main data tables loaded into PostGIS. It should be noted, there are 24 (1 hour) ADS-B tables per day and weather rasters for every 5 mins. The weather raster only covers the continental U.S., thus the Spatial extent studied in this report is limited to the extent of the weather raster.



Methods

Before any assessments could be made with the data, all data was first edited in ArcGIS Pro. For instance, the weather raster was converted to polygon data and reclassified into severity classes (Light, Moderate, Heavy, or Extreme based on pixel/dBZ values) to better interpret flight paths' interactions with weather data. By reclassifying the data, I was able to count the frequency of intersecting flights with more meaning (instead of counting values from 0 to 255 I counted interactions within 4 interpretable classes). I also clipped the ADS-B data to the same spatial extent as the weather data (the continental U.S.) to avoid long computational queries (even with the smaller extant there were still hundreds of thousands of points observed for each hour of ADS-B data). Additionally, FAA airspace shapefile contained Airspace designations for B, C, D, and E. Within ArcGIS Pro this data was trimmed to the relevant study of class B airspace. After editing, the data was loaded to PostGIS via command line. For raster and shapefiles using raster2pgsql and shp2pgsql requires minimal supplemental information to load data (needed to supply file path and SRID), but with tabular based data I assigned datatypes to each column within the ADS-B table. As tabular data, ADS-B data was loaded into PostGIS as a csv (without geometry) and needed additional SQL commands within PostGIS to add and create a new geometry column prior to any spatial queries.

Image 1: Data Flow diagram. Preprocessing within ArcGIS Pro (invovles cleaning the data and some minor conversion prior to loading in PostGIS).



Data analysis

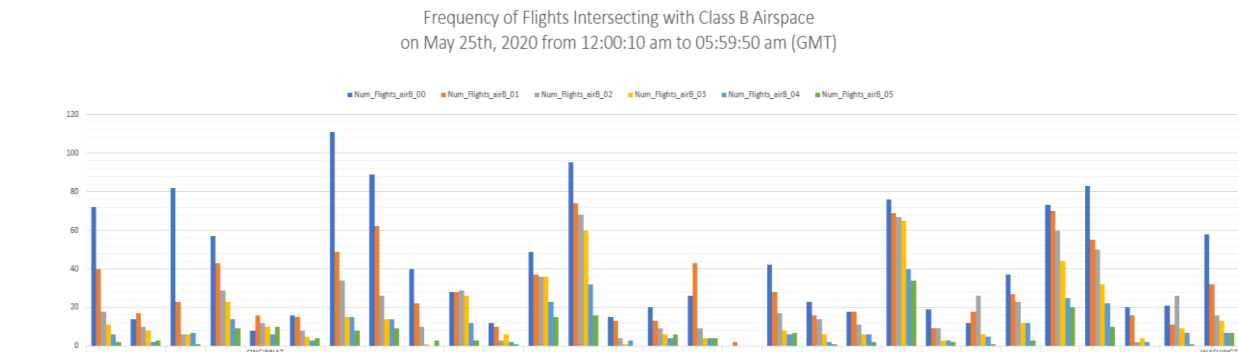
Spatial Queries were an essential method for comparing ADS-B's interactions around or within Class B and weather data (employing the use of ST_INTERSECT). In addition to spatial queries uniting ADS-B tables involved several unions, which unlike a join stacks rows together instead of appending columns. Many of the resulting statistics involved using functions like SUM, COUNT, DISTINCT, MIN, and MAX to return values used to build histograms and to find the length of flight times for each flight. For examples of SQL queries please see the Appendix at the end of this document.

Results

Utilizing SQL queries, I successfully counted the number of unique Flights intersecting with Class B airspace on May 25th for a 6-hour window. The map in *Image 2* displays the busiest airspaces (as indicated by the larger circle and darker purples) while the graph in *Image 3* displays the same frequencies but broken down to hourly increments. These two tables vary slightly in total number of unique flight because flights counted within each hour could be counted again in the next hour if the flight happens to be intersecting any class B airspace at the end or beginning or a new hour. Overall, Los Angeles seems the busiest during all six hours while Dallas is the busiest during the first observed hour but quickly tapers in subsequent hours. Conversely, New Orleans is the least busy, which could be direct result of extreme weather observed within its vicinity or it being an airport with less connecting flights.

Image 2: Map (created within ArcGIS Pro) displays the total Number of Unique Flights that traversed through each class B airspace on May 25th of 2020 from 12:00:10 am to 05:59:50 am (GMT). Note the size and coloring of each point for each respective airspace is based upon the number of flights observed during this timeframe. See Image 2 for a more detailed graphical interpretation of these results.

Image 3: This Graph contains the same information displayed in Image 2 but contains the hour-by-hour distribution of unique flights traversing through each respective Class B airspace.

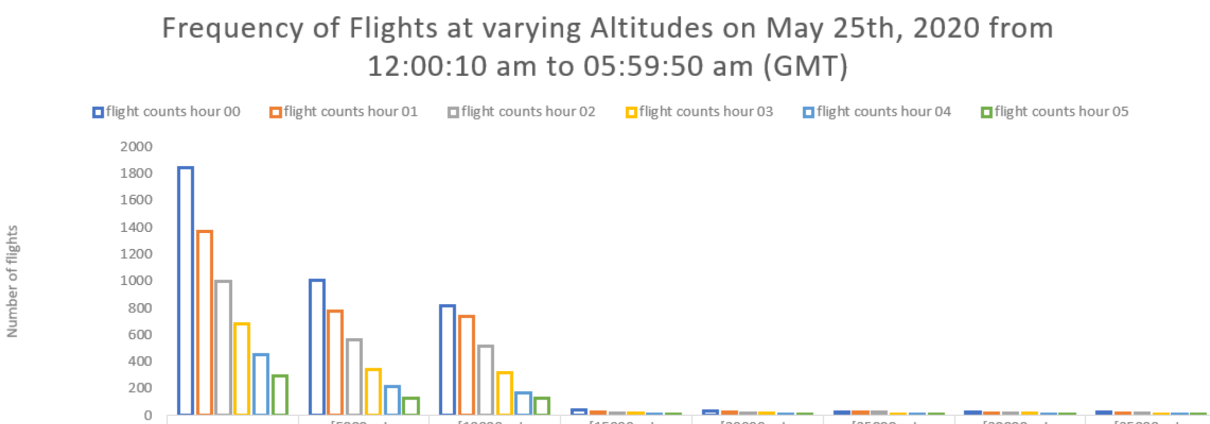


	ATLANTA CLASS B	BOSTON CLASS B	CHARLOTTE CLASS B	CHICAGO CLASS B	CINCINNATI/NORTHERN KY INTL CLASS B	CLEVELAND CLASS B	DALLAS CLASS B	DENVER CLASS B	DETROIT CLASS B	HOUSTON CLASS B	KANSAS CITY CLASS B	LAS VEGAS CLASS B	LOS ANGELES CLASS B	MEMPHIS CLASS B
Num_Flights_airB_00	72	14	82	57	8	16	111	89	40	28	12	49	95	15
Num_Flights_airB_01	40	17	23	43	16	15	49	62	22	28	10	37	74	13
Num_Flights_airB_02	18	10	6	29	12	8	34	26	10	29	3	36	68	4
Num_Flights_airB_03	11	8	6	23	10	5	15	14	1	26	6	36	60	1
Num_Flights_airB_04	6	2	7	14	6	3	15	14	0	12	2	23	32	3
Num_Flights_airB_05	2	3	1	9	10	4	8	9	3	3	1	15	16	0

	MIAMI CLASS B	MINNEAPOLIS CLASS B	NEW ORLEANS CLASS B	NEW YORK CLASS B	ORLANDO CLASS B	PHILADELPHIA CLASS B	PHOENIX CLASS B	PITTSBURGH CLASS B	SALT LAKE CITY CLASS B	SAN DIEGO CLASS B	SAN FRANCISCO CLASS B	SEATTLE CLASS B	ST LOUIS CLASS B	TAMPA CLASS B	WASHINGTON-TRI AREA CLASS B
Num_Flights_airB_00	20	26	0	42	23	18	76	19	12	37	73	83	20	21	58
Num_Flights_airB_01	13	43	2	28	16	18	69	9	18	27	70	55	16	11	32
Num_Flights_airB_02	9	9	0	17	14	11	67	9	26	23	60	50	2	26	16
Num_Flights_airB_03	6	4	0	8	6	6	65	3	6	12	44	32	4	9	13
Num_Flights_airB_04	4	4	0	6	2	6	40	3	5	12	25	22	2	7	7
Num_Flights_airB_05	6	4	0	7	1	2	34	2	1	3	20	10	0	1	7

In *Image 4* we have the number of flights flying through varying altitudes. It is clear that a majority of flights fly between 0 and 15,000 meters which corresponds with the typical commercial flight cruising between 31,000 and 38,000 feet (HACOBIAN). What is interesting is that there are flights observed flying between 35,000 and 40,000 meters, which is well above the cruising altitude for most planes and entering the Stratosphere. It is possible we are observing rockets or weather balloons, but I think this requires further investigation to make sure these aren't errors.

Image 4: Number of Flights that flew within each altitude broken down by the hour.



	[0m to 5000m)	[5000m to 10000m)	[10000m to 15000m)	[15000m to 20000m)	[20000m to 25000m)	[25000m to 30000m)	[30000m to 35000m)	[35000m to 40000m)
flight counts hour 00	1842	1003	814	42	33	25	28	22
flight counts hour 01	1373	780	736	26	28	29	19	19
flight counts hour 02	998	561	515	18	18	22	15	17
flight counts hour 03	682	342	320	18	18	10	15	4
flight counts hour 04	449	214	166	7	4	11	7	3
flight counts hour 05	291	125	130	6	7	8	4	3

Table 3 records the minimum, maximum, and average horizontal velocity recorded within each altitude. Again, I suspect the ADS-B is incomplete or contains errors, since I would not expect any flight to stop moving while in flight (as indicated by the minimum velocities observed between 5,000 and 20,000 meters). Based on the results from *Table 3* and *Image 4*, the ADS-B probably requires further processing to remove incomplete data.

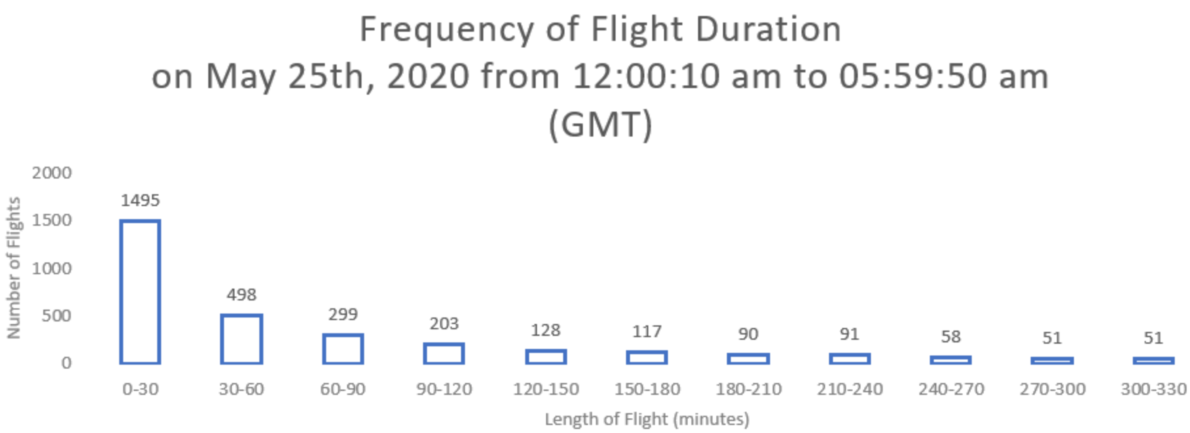
Table 3: Average Velocity of Flights within each respective altitude

Altitude (meters)	Altitude (feet)	Max Velocity (m/s)	Min Velocity (m/s)	Avg Velocity (m/s)
[0m to 5000m)	[0ft to 16,404ft)	424.1918	0	80.6390
[5000m to 10000m)	[16,404ft to 32,808ft)	285.5628	0	205.1380
[10000m to 15000m)	[32,808ft to 49,213ft)	295.9515	0	229.3006

[15000m to 20000m)	[49,213ft to 65,617ft)	281.4178	0	27.0541
[20000m to 25000m)	[65,617ft to 82,021ft)	262.0293	5.6822	90.0036
[25000m to 30000m)	[82,021ft to 98,425ft)	261.1658	0	73.6969
[30000m to 35000m)	[98,425ft to 114,829ft)	272.1370	1.4551	116.7134
[35000m to 40000m)	[114,829ft to 131,234ft)	256.2556	23.8649	117.6083

Next, *Image 5* displays the number of flights and their flight duration within the observed 6-hour window. It should be noted that these results are very skewed in favor of shorter flights since I have not been able to remove or identify flights beyond the temporal and spatial extant (there are flights already in flight when we started assessing ADS-B and other flights that are flying beyond the continental U.S.). To remove these flights, it might be best to remove flights that don't intersect with at least 2 class B airspaces (this way we can try to ensure we are recording the flight times of planes that are more likely to be within our spatial and temporal extant but this is not a full-proof way to isolate these flights).

Image 5: based on this Graph we'd be inclined to assume most flights within the continental US were 30 minutes or shorter, but it must be acknowledged we are counting flights beyond the spatial and temporal extant (thus their flight times appear to end and suggest short flights).



Lastly, *Table 4* counts the number of flights that pass through (maybe above) weather as classified by Air Traffic Control for pilots. The table doesn't sum the number of flights over the entire continenal U.S. because the process of identifying the intersection of flights with the entire Continental U.S. takes too long. As a proof of concept I cropped the weather spatial extant to Surrounding Houston Class B airspace. Based on *Table 4* we see flights do not traverse through 'Extreme' weather pheonmena which support the idea that flights tend to avoid 'Extreme' weather or that there is not as much weather classified as 'Extreme' and thus is less likely to be traversed by aircrafts.

Table 4: Frequencies of Weather pixels and Number of Flights that intersected with each pixel value. The pixel values range from 0 to 255 (8bit based values) and are stretched based on the intial dBz scaled used to measure rain or snow intensity.

Weather Classification	Number of flights	dBz values	Pixel values	% of Spatial Extent
Extreme	11	< 30	[1,125]	1.26
Heavy	30	[30, 40]	[126, 145]	16.30
Light	32	(40, 50]	[146, 165]	52.23
Moderate	39	>50	[166, 255]	30.20

Discussion and Conclusion

This report has successfully delivered simple summary statistics about flights on May 25th of 2020 but has the potential for more extensive exploration. For starters, it would be interesting to observe flight statistics beyond this one day. For instance, extracting the flight data once a month and comparing the flight statistics prior to and during covid over two years might be used to quantify how covid affected flight travel. Additionally, it would be interesting to observe if there is a correlation between countries flight frequencies and covid cases. This would require counting the number of domestic flight and international flights between countries (which is tricky since we'd have to identify planes that have landed and taken off within the observed timeframe).

As it stands, adding more temporal and spatial analysis will drastically increase the size of the data which will increase the challenges associated with handling Big Data (storage, computation time, and processing). During this project, I found managing the raster data in conjunction with the ADS-B data to be the most challenging. Loading multiple files to PostGIS is simple but time consuming when constantly utilizing GUI (for future projects I would encourage as much automation as possible when grabbing, loading, and preprocessing data).

Appendix

4 SQL Queries

a. A SQL query that involves only 1 table.

```
WITH pixeldata AS ( -- CTE: retruns identifiable pixel values
SELECT r.rid, (st_valuecount(r.rast)).* AS rast
FROM n0q_202005250030 r --Original raster file with pixel values ranging from 0 to 255
), histo AS ( --CTE returns a histogram for pixels per pixel value (currently counts are seperated by tiles)
SELECT value, sum(count) AS total_pixels --must SUM our count to merge counts from numerous tiles
FROM pixeldata
GROUP BY value --groups our counts by the assigned value
ORDER BY 1 ASC -- reorganizes order by value class
), count_pix AS ( --CTE untilizes subqueries to count the total number of pixels to two new categories
SELECT
    (SELECT  SUM(total_pixels) AS L_Weather_pix_cnt FROM histo WHERE value>=1 AND value <=125) AS L_pix_cnt,
    (SELECT  SUM(total_pixels) AS M_Weather_pix_cnt FROM histo WHERE value>=126 AND value <= 145) AS M_pix_cnt,
    (SELECT  SUM(total_pixels) AS H_Weather_pix_cnt FROM histo WHERE value>=146 AND value <= 165) AS H_pix_cnt,
    (SELECT SUM(total_pixels) AS  E_Weather_pix_cnt FROM histo WHERE value > 165) AS E_pix_cnt
FROM histo
Limit 1 --returns the number of pixels as severe and mild weather (only need 1 row, the rest are redundant)
)
Select
    L_pix_cnt/(L_pix_cnt + M_pix_cnt+H_pix_cnt+E_pix_cnt )* 100 AS percent_Light,
    M_pix_cnt/(L_pix_cnt + M_pix_cnt+H_pix_cnt+E_pix_cnt )* 100 AS percent_Moderate,
    H_pix_cnt/(L_pix_cnt + M_pix_cnt+H_pix_cnt+E_pix_cnt )* 100 AS percent_Heavy,
    E_pix_cnt/(L_pix_cnt + M_pix_cnt+H_pix_cnt+E_pix_cnt )* 100 AS percent_Extreme
FROM count_pix
--Returns the % of pixels classified either Light, Mod, Heavy, or Extreme weather (based on dBZ values used by Air Traffic Control)
```

b. A SQL query that involves 2 or more tables with a join.

```
With Merged_flights AS ( --merging flights from 5 differnt tables
SELECT time_epoch, icao24, geom
FROM states_2020_05_25_00
UNION
SELECT time_epoch, icao24, geom
FROM states_2020_05_25_01
UNION
SELECT time_epoch, icao24, geom
FROM states_2020_05_25_02
UNION
SELECT time_epoch, icao24, geom
FROM states_2020_05_25_03
UNION
SELECT time_epoch, icao24, geom
FROM states_2020_05_25_04
UNION
SELECT time_epoch, icao24, geom
FROM states_2020_05_25_05
), merged_B AS ( --CTE: merges the class b polygons based on their designated airspace
SELECT name, st_union(geom) AS geom
FROM class_b_repro
GROUP BY name
), pts_within AS ( --CTE: only returns the points that intersect with class b
SELECT merged_B.name, merged_B.geom, pts.geom as pts_geom, pts.icao24, pts.time_epoch
FROM merged_B, Merged_flights pts
```

```sql
WHERE ST_INTERSECTS(merged_B.geom, pts.geom) --alteration, returns results faster (spatial query is here)
), distinct_flights AS ( --CTE: selects only distinct flight paths
SELECT Distinct icao24, name
FROM pts_within
ORDER BY name
), last_CTE AS( --histogram that counts number of unique flights per airspace
SELECT name, count(name) AS Num_flights_AirB
FROM distinct_flights
GROUP BY name
ORDER BY name ASC
)--appends the geom for each Airspace to the histogram
SELECT cte.name, cte.Num_flights_AirB, b.geom
FROM merged_B b
JOIN last_CTE cte ON b.name = cte.name
--RETURNS A HISTOGRAM FOR FLIGHT FREQUENCIES PER CLASS B AIRSPACE WITH EACH AIRSPACE'S GEOM
```

c. A SQL query using a sub query or common table expression

```sql
WITH merged_flights AS ( --CTE returns all available flight data in one table (each table represents 1 hour of data, this merges 6 hours
SELECT time_epoch, icao24, baroaltitude, geom
FROM states_2020_05_25_00
UNION --using UNION because I am appending rows instead of columns (joins appends columns to already existing tables)
SELECT time_epoch, icao24, baroaltitude, geom
FROM states_2020_05_25_01
UNION
SELECT time_epoch, icao24, baroaltitude, geom
FROM states_2020_05_25_02
UNION
SELECT time_epoch, icao24, baroaltitude, geom
FROM states_2020_05_25_03
UNION
SELECT time_epoch, icao24, baroaltitude, geom
FROM states_2020_05_25_04
UNION
SELECT time_epoch, icao24, baroaltitude, geom
FROM states_2020_05_25_05
)
SELECT 0 + ((bucket-1) * (40000-0)/8) || '-' || (0 + (bucket) * (40000-0)/8) AS bins, cnt
FROM ( --this subquery returns a count of unique flights observed within each designated binned altitude
SELECT WIDTH_BUCKET(baroaltitude, 0, 40000, 8) AS bucket, COUNT(DISTINCT icao24) AS cnt
FROM merged_flights
GROUP BY bucket
ORDER BY bucket) x
```

d. A spatial SQL query

```sql
--counting the number of Unique flights that intersect with weather
WITH flight_paths AS ( --CTE grabs the flights from hour 00
SELECT time_epoch, icao24, geom
FROM states_2020_05_25_00
), flight_within AS ( --CTE idnetifies the intersection of flight pts with weather classes
Select w.weather_cl, f.icao24
FROM Weather_clip_2020_05_25_0030 w, flight_paths f
WHERE ST_INTERSECTS(w.geom, f.geom)
), distinct_flights AS ( --CTE grabs distinct flights from all pts in each weather class
SELECT Distinct icao24, weather_cl
FROM flight_within
ORDER BY weather_cl
) --Returns the number of unique flights found to traverse through each type of weather classification
SELECT weather_cl, COUNT(weather_cl) AS Num_flights_AirB
FROM distinct_flights
GROUP BY weather_cl
ORDER BY weather_cl ASC
```

Bibliography

HACOBIAN, CELINE. "Here's How High Planes Actually Fly, according to Experts." *Time*, 27 June 2018, time.com/5309905/how-high-do-planes-fly/#:~:text=Commercial%20aircraft%20typically%20fly%20between.