Noah Kurrack
6 May 2022
CSCI-UA 476
Professor Ann Malavet

Aircraft Traffic Routes Analysis

Abstract

This application digested publicly available tracking data for commercial aircraft across the globe, and analyzed the most heavily trafficked geographic segments of the sky. Based on the frequency of travel by planes through certain segments (within a certain degree of precision), I established the most commonly used routes. Within the United States, the most commonly traversed segment was just north of the DFW airport in Texas, with 1048 unique aircraft traveling that segment in one day. I also found that larger aircraft are more likely to traverse segments over the Atlantic Ocean.

Introduction

In today's society, air travel is a major force that enables people and goods to move quickly around the globe. Every major metropolitan area has an airport to participate in this global network of travel. In order for aircraft to safely operate through the skies, governments have drawn paths—almost like highways in the sky—for pilots and air traffic controllers to utilize. Air traffic controllers use radar to track planes and determine the safest paths for them to take on their way to their destination.

While the many flight paths can be retrieved from government agencies like the Federal Aviation Administration, this information does not give any indication as to which paths aircraft often use in reality. With the help of open-sourced radar data, I analyzed actual aircraft location data to determine which segments of the sky are most frequently traveled by commercial aircraft in the United States.

Motivation

With the ubiquity of air travel today, there are tens of thousands of aircraft in the sky at any time. Thus, airlines, air traffic controllers, and pilots may be interested in the commonly traveled flight segments so that alternative routes can be established that may avoid congestion and unnecessary delays for travelers. Additionally, the general population may be interested to know where aircraft frequently travel because a heavily trafficked tract of sky likely experiences higher levels of air and noise pollution for residents who live there. This is particularly important in large metro areas because nearby airports facilitate the most amount of flights while affecting a large number of people due to the density of large cities.

Related Work

I found several paper that identify common air routes used by aircraft by applying a clustering algorithm called DBSCAN on the flight location data (Conde Rocha Murca et al. 2016). This algorithm works by clustering nearby points (aircraft location coordinates) into groups based on the density of the points. Dense cluster likely constitute frequently traveled segments (Lin et al., 2021). This approach has been applied to the US and Chinese aircraft patterns (Ren, P., & Li, L. 2018). Applying this machine learning algorithm to the flight data and comparing the results may provide valuable insight into the validity and accuracy of my results.

Dataset

ADS-B Exchange is a company that collects and publishes a wide array of aircraft tracking data. Their network of independently operated radar antenna constantly track the location (as well as many other metrics) of any aircraft operating around the world. While they charge a fee for full access to this unfiltered tracking data, they have released a free sample for experimentation such as this study. The sample consists of a full 24 hours of tracking data for all aircraft. The sample was recorded on February 1st of 2022.

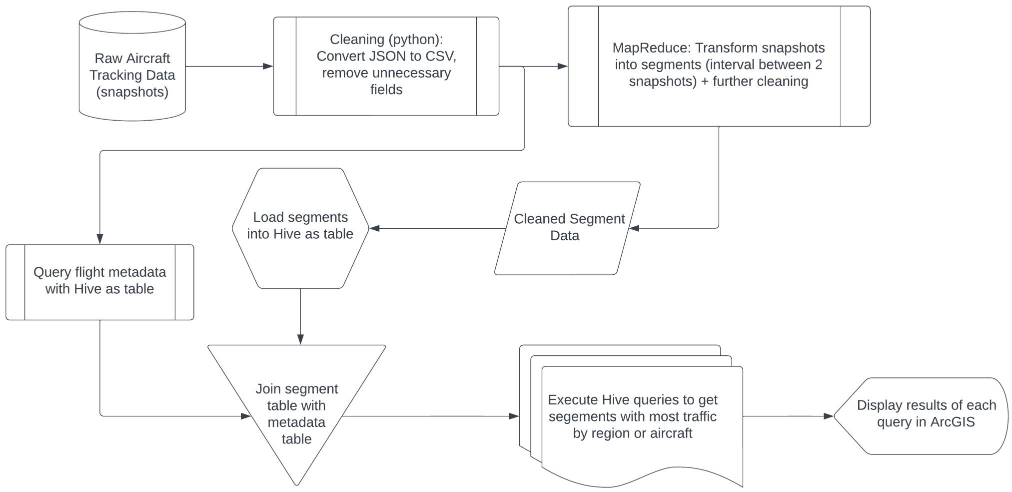
The dataset is published as JSON files in 5 second intervals for the entire 24 hour window, which was about 60 GB of data. Each JSON file contains a record for each aircraft being tracked at that time; each record contains the current value for over 30 different measures or attributes. I will refer to these records with current values for a particular timestamp as "snapshots." The relevant fields are enumerated in the following table:

hex	the 24-bit ICAO identifier of the aircraft, as 6 hex digits
flight	callsign, the flight name or aircraft registration as 8 chars
r	aircraft registration pulled from database
t	aircraft type pulled from database
lat	the aircraft position in decimal degrees
lon	the aircraft position in decimal degrees

The full schema can be found at https://www.adsbexchange.com/version-2-api-wip/.

Process

The process for cleaning and transforming the data is summarized in the following flowchart. This flowchart assumes that the JSON files have already been downloaded from the ADS-B Exchange server.



Step 1: I wrote a python script to load each JSON file into memory and write the relevant fields (described above) to a CSV file. Notably, I only wrote records to the CSV files that had non-null values for each of those relevant fields; otherwise the record was discarded.

Step 2: I wrote a MapReduce job (mapper only) to round the latitude and longitude coordinates of each records to 2 decimal places. I chose to do this because routing to 2 decimals places means the coordinates are accurate within 500 meters of error. This is sufficient for looking at 5 minute segments (discussed further in next step) so that the flight path is represented while allowing for similar paths (within 500 meters) to be aggregated together.

Step 3: I wrote a MapReduce job (mapper + reducer) to transform the "snapshots," an aircraft's current location at a timestamp, to "segments," an aircraft's start and end location over a specific interval. (I chose 5 minute intervals somewhat arbitrarily after trying different values and examining the segments that resulted from the transformation process.)

Step 4: I wrote a MapReduce job (mapper only) to remove flight segments where the start and end coordinates were the same, which meant that the aircraft was stationary. (This occurs often when plane are on the ground at an airport shortly before takeoff/after landing, but the radar is still tracking the plane's location.)

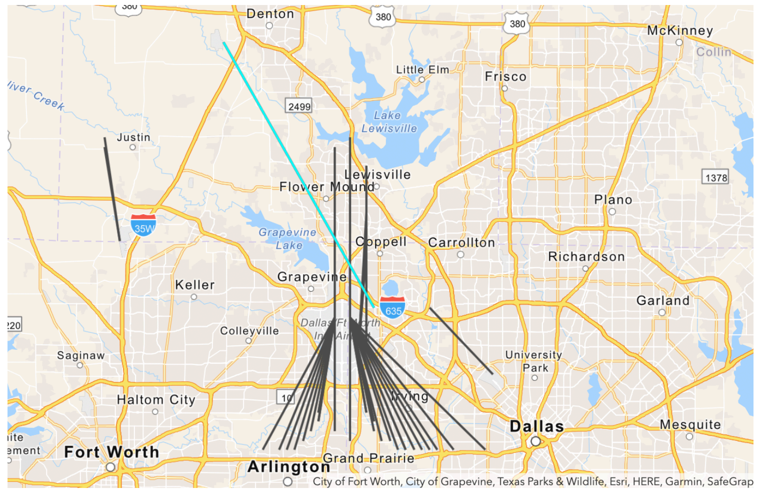
Step 5: I loaded the output of the last MapReduce job into Hive as an external table (simply pointing Hive to the HDFS directory of Step 4's output). I also created a "metadata" table in Hive from the output of Step 1 that contained static attributes of a flight (such as aircraft type). This was a "SELECT DISTINCT" query that yielded 1 row of metadata for every flight (identified by hex id).

Step 6: I executed several Hive queries on the joined data from both table described in Step 5. These queries returned the top 100 flight segments for certain conditions (such as in the NYC area) which were determined by counting the frequency of that segment in the dataset of all segments.

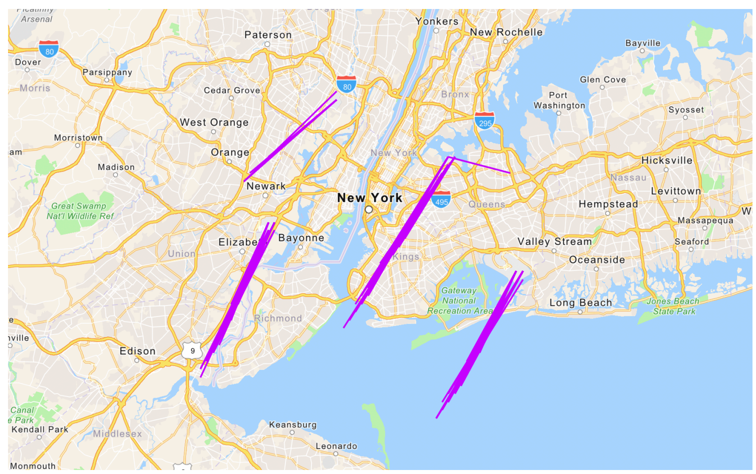
Step 7: I wrote a python script to transform the flight segments returned by the queries into geoJSON, which can be imported directly into ArcGIS online, a web-based mapping software. The maps with resulting segments are included in the next sections.

Visualizations

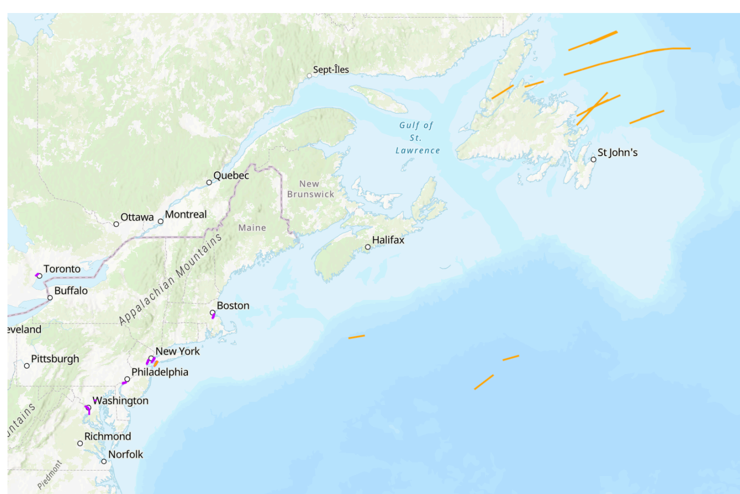
The most commonly traveled segment in February 1st, 2022 occurred just north of the Dallas-Ft. Worth International Airport as shown in the map on the right (shown in green). 1048 unique flights traversed this path during the 24 hour period. Other commonly traversed paths are also included in black.



The top 100 commonly traversed segments in the NYC area are shown in purple on the map to the right. These segments largely align with the approach paths to the 3 largest airports in the region: JFK, LaGuardia, and Newark.



Lastly, I looked at the top 100 segments in the northeast region of the US for large (orange) versus smaller (purple) commercial aircraft in the map on the right. For large, I filtered Boeing 747 and Airbus A380 aircraft, while for small, I filtered Boeing 737/757 and Airbus A320. As you can see, the larger aircraft heavily traveled along paths over northeast Canada which indicates trans-Atlantic flights.



Conclusion

Many of the segments found in our analyses were in the vicinity of large airports. This makes sense as many aircraft coalesce to travel the same paths as they approach or depart from the same exact location of the airports' runways. This data is useful for residents near major airports to know where they are likely to experience increase noise and air pollution. However, this is only one aspect of potential analysis. Thus, it would be beneficial to find a way to reliably filter out segments that are within a certain radius of an airport in order to capture frequently traveled segments in the middle of a flight. (One possible way is to only look at data points when the aircraft is above a certain altitude, which means it is at cruising altitude.)

References

Conde Rocha Murca, M., DeLaura, R., Hansman, R. J., Jordan, R., Reynolds, T., & Balakrishnan, H. (2016). Trajectory clustering and classification for characterization of Air Traffic Flows. *16th AIAA Aviation Technology, Integration, and Operations Conference*. https://doi.org/10.2514/6.2016-3760

Lin, Y., Li, L., Ren, P., Wang, Y., & Szeto, W. Y. (2021). From aircraft tracking data to network delay model: A data-driven approach considering en-route congestion. *Transportation Research Part C: Emerging Technologies*, *131*, 103329. https://doi.org/10.1016/j.trc.2021.103329

Ren, P., & Li, L. (2018). Characterizing air traffic networks via large-scale aircraft tracking data: A comparison between China and the US networks. *Journal of Air Transport Management*, *67*, 181–196. https://doi.org/10.1016/j.jairtraman.2017.12.005