Where, oh where has the little plane gone?

Tidewater Big Data Enthusiasts
Chuck Cartledge
Developer

July 14, 2020

Contents

List of Tables

List of Algorithms

List of Figures

1 Introduction

We look at what information can be found starting with an aircraft tail number, an airline, and a starting departure airport. We construct a PostGreSQL database to hold the data as long term persistent storage, and use Bureau of Transportation Statistics (BTS) departure data to identify where a tail number is going to next.

After we have queried the BTS database to extract all of the airports that a tail number has departed from, we analyze the data in various ways to: 1) identify where the tail number has departed from, 2) what day of the week the tail number departed on, 3) what hour of the day the tail number took off, 4) where the domestic departure airport is located, 5) how long the airport has been in service, and 6) has the airport moved in its lifetime.

2 Purpose

The purpose of this exploration was to see what information could be obtained based on the tail number of a specific commercial aircraft that departed Norfolk International Airport (NORF) in January 2019. NORF was selected because it is local to the author. The tail number (N287WN) was selected because the author has traveled on Southwest Airlines from time to time. All explorations started from these few facts.

The initial quest was to find when the tail number departed which airports during 2019. 2019 was selected because it was the last full year's worth of data at the time of this report. When the departure times and locations were enumerated, focus was turned to finding out information about the various airports the tail number departed from, and from there to other airports.

In summary:

- We are interested in domestically where Southwest tail number N287WN departed from in 2019.
- We are interested in geographic information about the domestic airports used by tail number N287WN.
- By extension, we are interested in geographic information about domestic and foreign airports.

3 Program usage

The program has a number of command line arguments, and default values. It can be used in many different ways (see Table 1). Command line arguments are shown in (see Table 2). Most program control values (see Table 3) can be modified by command line arguments.

Table 1: Program usage.

Example usages
followAircraft.py
followAircraft.py [-?]
followAircraft.py [-option [optionArgument]]
followAircraft.py [-option [optionArgument] -option [optionArgument]]
followAircraft.py [-option [optionArgument] -option [optionArgument] -option [optionArgument]
and so on.
Airports, Airline, and TailNumber are lists, and can have multiple entries.
Lists can be reset by passing 'R' as an argument.
Options are processed in the order they are read from left to right.

Table 2: Program command line arguments.

Argument	**Meaning**
-?	Print program usage, and exit.
-a	Add this airport to the list.
-A	Add this airline to the list.
-C	Stop after processing this number of airports.
-D	Show current values.
-d	Set the download directory.
-R	Reset the entire database.
-S	Save the current values to this file.
-T	Add this tail number to the list.
-L	Load defaults from this file.
-r	Reset defaults to default values.
-Z	Airport location zip file.

Table 3: Default command line argument values. Most values can be modified by CLI arguments. Airport location data downloaded from `https://www.transtats.bts.gov/Fields.asp?Table_ID=288`

Variable	**Value**
DownloadDir	/home/chuck/Downloads
ResetTable	False
Airports	['ORF']
Airline	['WN']
TailNumber	['N287WN']
AirportLimit	sys.maxsize
AirportLocationZipFile	/home/chuck/CLC-Ent/Airplanes/Data/833374230_T_MASTER_CORD.zip

4 Finding airports

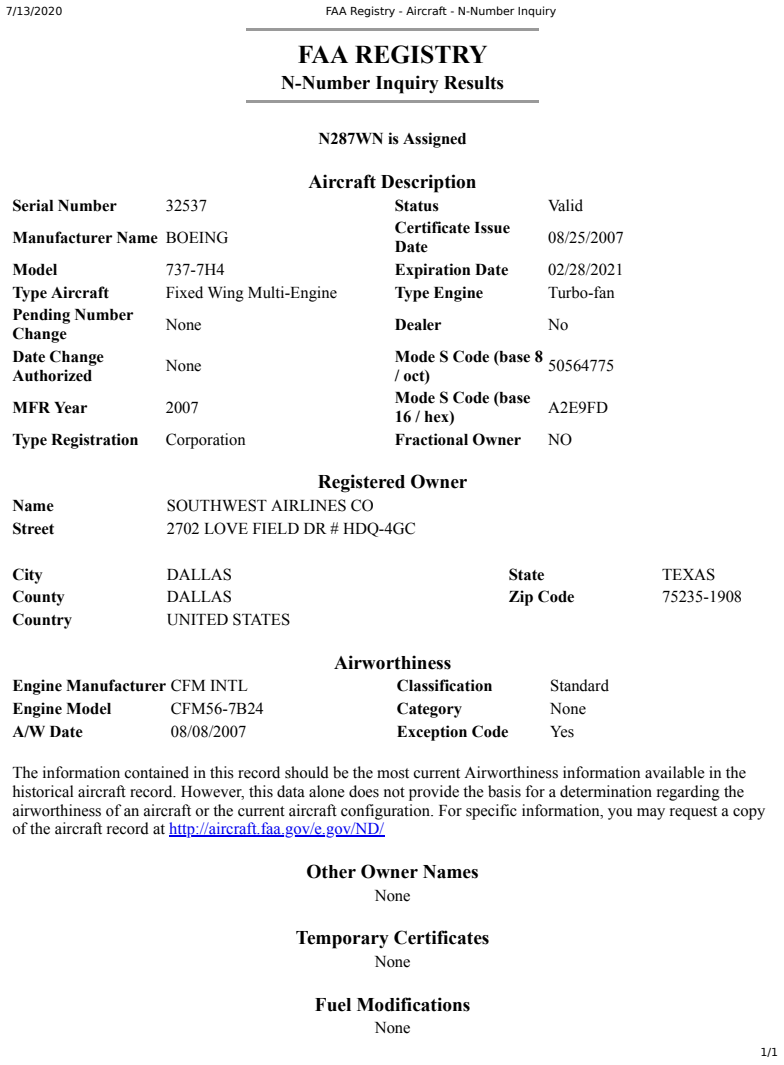
The Federal Aviation Administration (FAA) provides ownership information for aircraft based on their tail number (see Figure 1)[2].

The Bureau of Transportation Statistics (BTS) provides departure and arrival data for domestic airports based on airport, airline, and time of interest selections (see Figure 2)[1]. The requested data appears to return to the page from which it was requested (see Figure 3). Included in the data are the destination airports that each tail number is going to, as well as actual departure times.

Conceptually, the work that needs to be done to find all the airports that a tail number has left is relatively easy (see Algorithm 1). We view the database of departures as an unknown graph and use a breadth first search to explore the graph collecting airport departure data as we go.

A service is used to get departure data[1], and because the number of airports to be downloaded is unkown, a Python script was used to "drive" the service. The service does not appear to have a public API, so the driver script loads the request page, checks the appropriate boxes, and pushes the correct buttons to request data. Sometimes the script doesn't work. The failures do not appear to be predictable, so a human has to watch that the page is filled out correctly. The most common error is not selecting a year. Once the request page is submitted, the response page is presented. It appears that the driver program cannot control the response page, so the human has to press the CVS download button. In short, the human monitors the driver program's activities and takes action when necessary.

Figure 1: A web page showing tail number N287WN is owned by Southwest Airlines.



7/13/2020 FAA Registry - Aircraft - N-Number Inquiry

FAA REGISTRY

N-Number Inquiry Results

N287WN is Assigned

Aircraft Description

Serial Number	32537	**Status**	Valid
Manufacturer Name	BOEING	**Certificate Issue Date**	08/25/2007
Model	737-7H4	**Expiration Date**	02/28/2021
Type Aircraft	Fixed Wing Multi-Engine	**Type Engine**	Turbo-fan
Pending Number Change	None	**Dealer**	No
Date Change Authorized	None	**Mode S Code (base 8 / oct)**	50564775
MFR Year	2007	**Mode S Code (base 16 / hex)**	A2E9FD
Type Registration	Corporation	**Fractional Owner**	NO

Registered Owner

Name	SOUTHWEST AIRLINES CO		
Street	2702 LOVE FIELD DR # HDQ-4GC		
City	DALLAS	**State**	TEXAS
County	DALLAS	**Zip Code**	75235-1908
Country	UNITED STATES		

Airworthiness

Engine Manufacturer	CFM INTL	**Classification**	Standard
Engine Model	CFM56-7B24	**Category**	None
A/W Date	08/08/2007	**Exception Code**	Yes

The information contained in this record should be the most current Airworthiness information available in the historical aircraft record. However, this data alone does not provide the basis for a determination regarding the airworthiness of an aircraft or the current aircraft configuration. For specific information, you may request a copy of the aircraft record at http://aircraft.faa.gov/e.gov/ND/

Other Owner Names

None

Temporary Certificates

None

Fuel Modifications

None

1/1

Figure 2: Requesting data about Southwest Airlines departures from NORF.

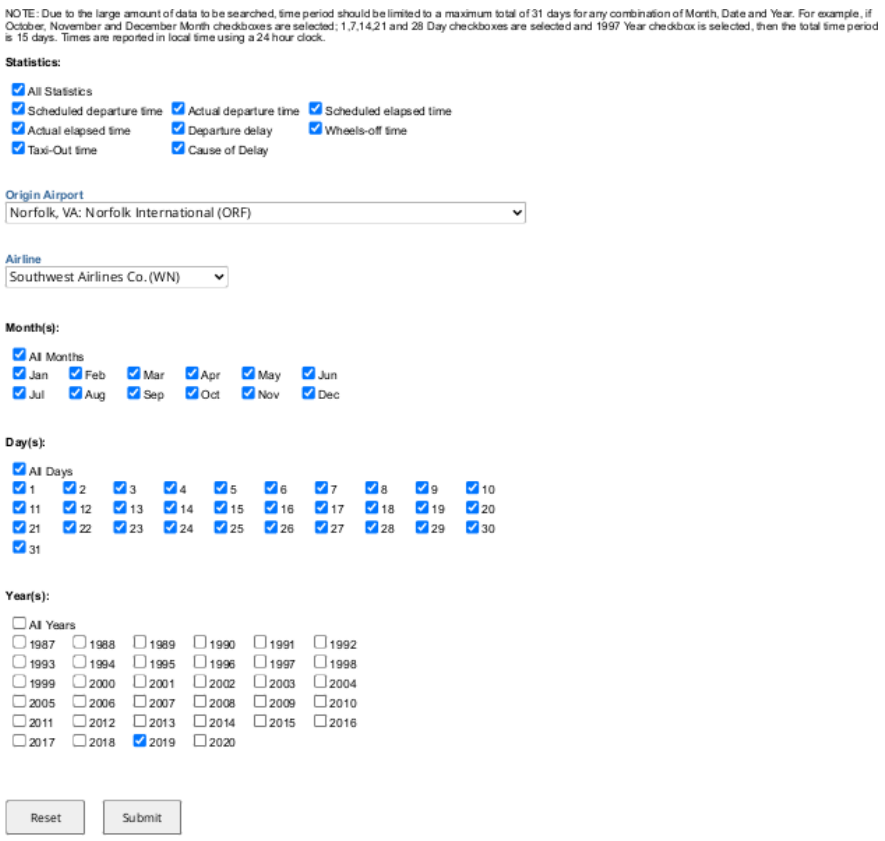


Figure 3: Requested data about Southwest Airlines departures from NORF. The data can be downloaded as an Excel, or CSV file.

NOTE: Due to the large amount of data to be searched, time period should be limited to a maximum total of 31 days for any combination of Month, Date and Year. For example, if October, November and December Month checkboxes are selected; 1,7,14,21 and 28 Day checkboxes are selected and 1997 Year checkbox is selected, then the total time period is 15 days. Times are reported in local time using a 24 hour clock.

Statistics:

All Statistics
Scheduled departure time | Actual departure time | Scheduled elapsed time
Actual elapsed time | Departure delay | Wheels-off time
Taxi-Out time | Cause of Delay

Origin Airport: Norfolk, VA: Norfolk International (ORF)

Airline: Southwest Airlines Co. (WN)

Month(s):

All Months
Jan | Feb | Mar | Apr | May | Jun
Jul | Aug | Sep | Oct | Nov | Dec

Day(s):

All Days
1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10
11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20
21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30
31

Year(s):

All Years
1987 | 1988 | 1989 | 1990 | 1991 | 1992
1993 | 1994 | 1995 | 1996 | 1997 | 1998
1999 | 2000 | 2001 | 2002 | 2003 | 2004
2005 | 2006 | 2007 | 2008 | 2009 | 2010
2011 | 2012 | 2013 | 2014 | 2015 | 2016
2017 | 2018 | 2019 | 2020

Reset | Submit

Origin Airport: Norfolk, VA: Norfolk International (ORF)
Airline: Southwest Airlines Co. (WN)
Month(s): January, February, March, April, May, June, July, August, September, October, November, and December
Day(s): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31
Year(s): 2019

NOTE: A complete listing of airline and airport abbreviations is available. Times are reported in local time using a 24 hour clock.
Airlines began reporting tarmac times for cancelled and diverted flights in October 2008. Tarmac times for cancelled or diverted flights operated prior to Oct. 1, 2008 are not available on this database beginning with flights operated in October 2008. For cause of delay data from June 2003, when cause of delay data was first reported, see BTS C Performance database. For an explanation of the Cause of Delay reporting, see Understanding the Reporting of Causes of Flight Delays and Cancellations.
All Cause of Delay (in minutes) are referring to the Arrival Delay.

Displaying Items 1 - 100 of 2566

Excel | CSV

Carrier Code	Date (MM/DD/YYYY)	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Taxi-Out time (Minutes)	Delay Carrier (Minutes)	Delay Weather (Minutes)
WN	01/01/2019	0165	N287WN	BWI	15:10	15:09	55	52	-1	15:19	10	0	0
WN	01/01/2019	0376	N741SA	BWI	12:35	12:29	55	51	-6	12:38	9	0	0
WN	01/01/2019	0913	N7704B	MCO	07:00	06:55	120	118	-5	07:02	7	0	0
WN	01/01/2019	0932	N228WN	BWI	06:30	06:26	55	44	-4	06:34	8	0	0

After the data is downloaded, it is cleaned and inserted into the database.

```
Data: Tail Number of interest;
Which airline owns the tail number;
Year of interest;
Starting airport;
Result: A database of all airports from which a tail number has departed
initialization;
ServiceQueue ← Push (starting airport);
while There are airports in ServiceQueue do
    Airport ← Pop (ServiceQueue);
    Get new departure data based on airport, airline, and year;
    Insert departure data into database;
    Destinations ← Push (tail number destination airports from new data);
    while Destinations do
        Airport ← Pop (Destinations);
        if Airport in database;
         then
            Ignore Airport
        else
            if Airport in ServiceQueue;
             then
                Ignore Airport
            else
                ServiceQueue ← Push (Airport);
            end
        end
    end
end
```

Algorithm 1: Searching for new airport departure data. An database of unknown size is queried until all airports are discovered.

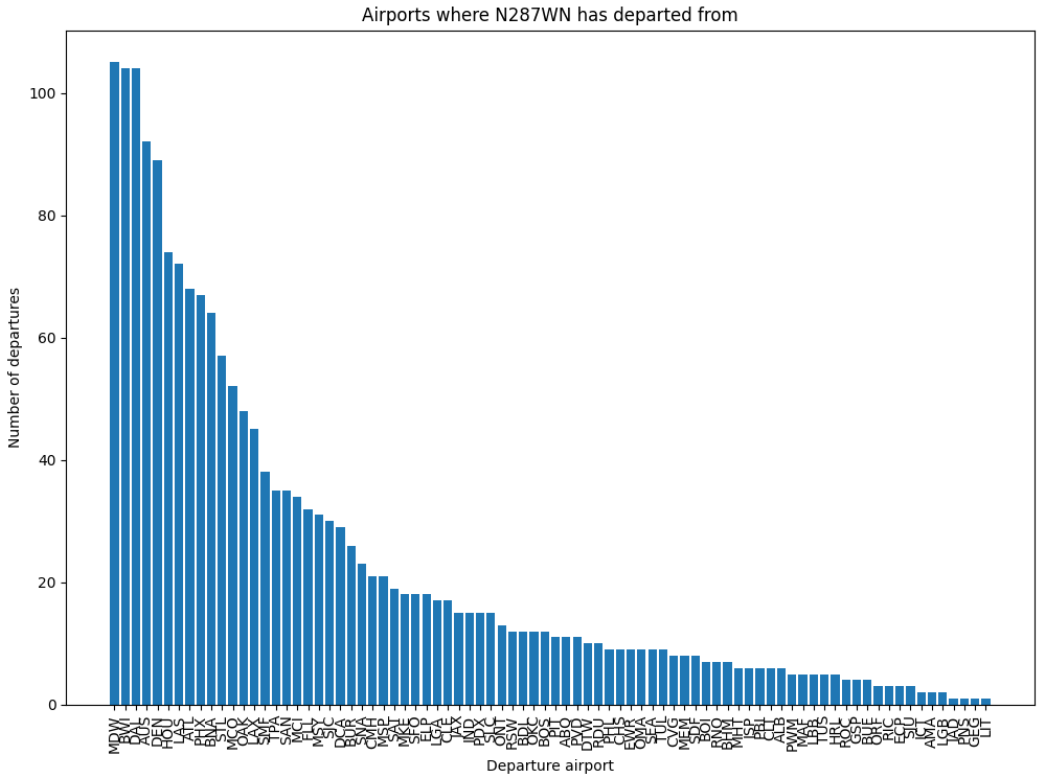


Figure 4: Histogram of where tail number N287WN has departed from. The five most departed airports are: MDW (Chicago Midway International), BWI (Baltimore/Washington International), DAL (Dallas Love Field), AUS (AustinBergstrom International), and DEN (Denver International).

5 Results

Using the program default values resulted in the following images. A histogram of where tail number N287WN departed from during 2019 (see Figure 4). There are 1,865 departures. A histogram of daily departures (see Figure 5) shows slight dips on Sundays and Tuesdays, perhaps evidence of fewer travlers on those days. The data was examined to see when tail number N287WN actually took off (see Figure 6). Sometimes it is easier to grasp where an airport is geographically rather than a list of acronyms (see Figure 7).

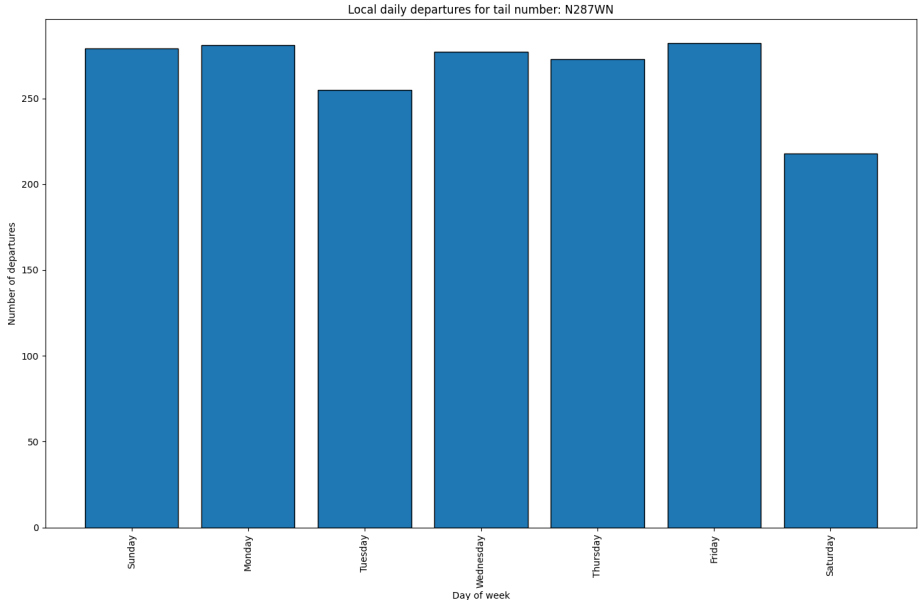


Figure 5: Histogram of tail number N287WN daily departures. There are slight dips on Sunday and Tuesday, suppporting the idea that there are fewer travelers on those days.

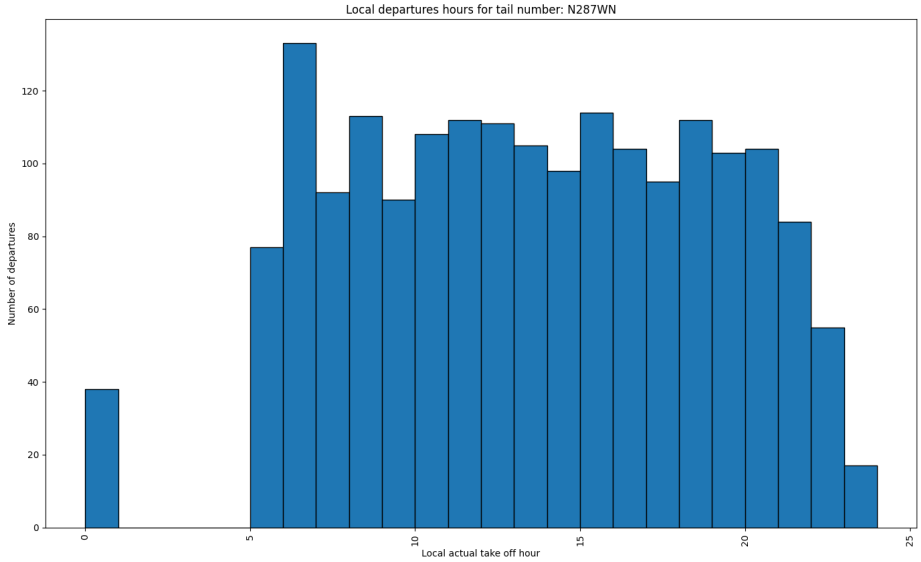


Figure 6: Histogram of tail number N287WN hourly takeoff times. It would seem that from 1AM to 5AM are good times to do plane maintenance.

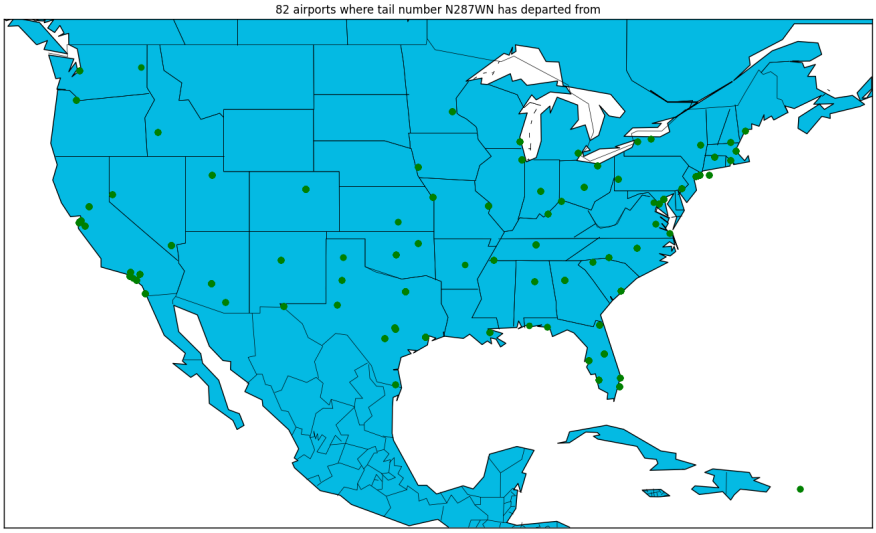


Figure 7: A map of where tail number N287WN has departed from.

6 Airports move

The airport location file contains 18,097 current and historical airport postions. For those airports with multiple position records, a positional uncertainty circle was computed. The first positional records was assumed to be the origin, and the great circle range and bearing to each of the other positions was computed and then converted to X and Y coordinates. The net effect was that each airport had a position record at (0,0) and then some number of other positions at other (X,Y) coordinates. The radius of the minimum enclosing circle for all the positions was reported. There are 6,387 airports that have current airport information records.

Data for some of the positional records is color coded (see Table 4).

Positional data was exmined both as to how many updates versus circular radius, and number of dayes between first and last updates versus circular size (see Figures 8, and 9). The radius error is relatively low for most of the data when looking at days between first and last updates, with a few outliers that might deserve closer investigation.

Creating a histogram of the circular radii, and limiting the histogram to 100 meters, it is apparent that approximately 1,750 airports have a radius of 0 (see Figure 10). This level of detail is lost when looking at all the data (see Figure 11). This change in perspective is also evident when limiting the data to Canadian (CA) airports (see Figures 12, and 13).

Table 4: Colors assigned to some positional records. Black is the default color for all countries.

Country	Color
US	red
GB	green
CA	orange
AU	orange
FR	blue

Figure 8: Number of positional updates, and circular radius. See text for explanation of the colors.

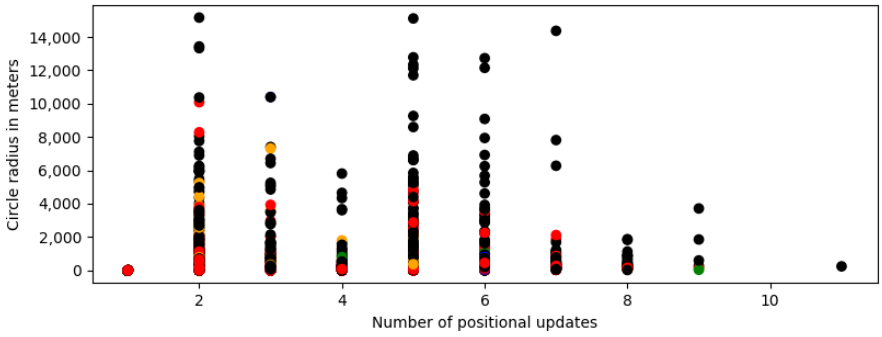


Figure 9: Days between positional updates, and circular radius. See text for explanation of the colors.

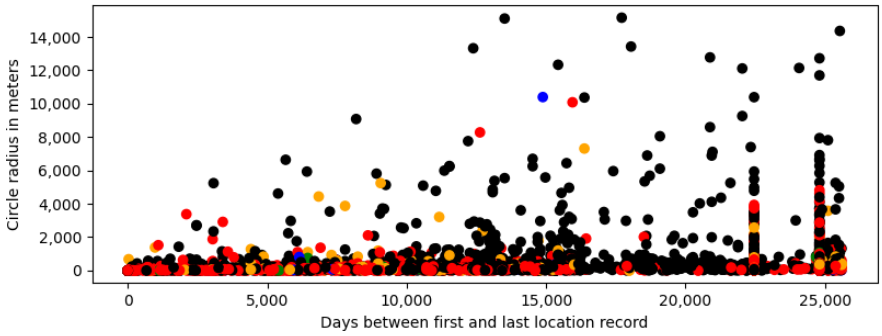


Figure 10: Radius histogram limited to 100 meters.

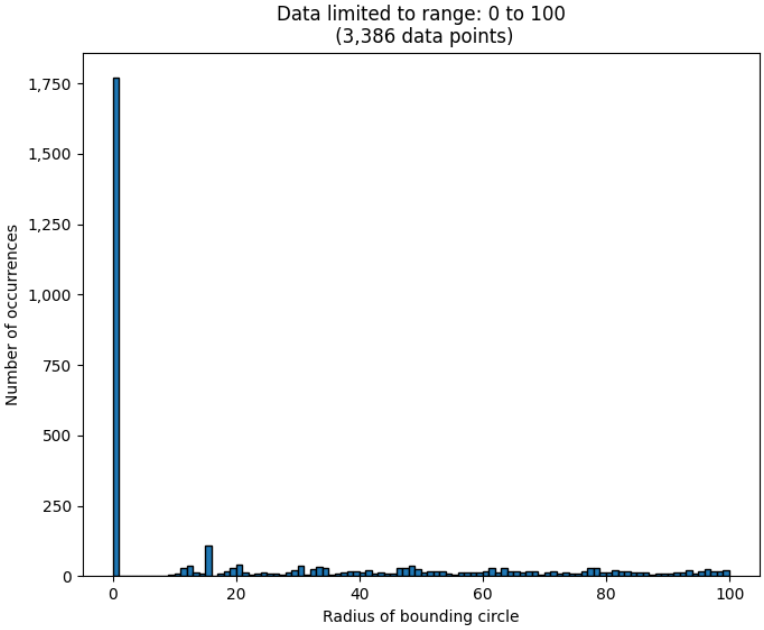


Figure 11: Radius histogram limited to 16,000 meters.

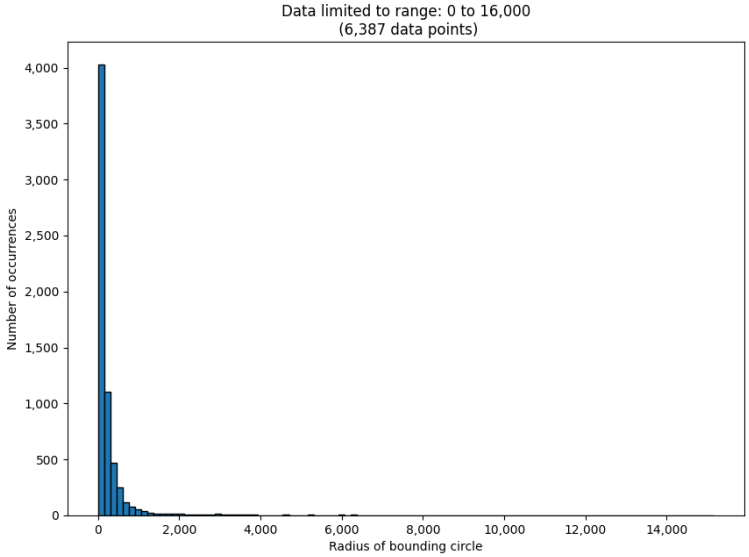


Figure 12: Radius histogram limited to 100 meters.

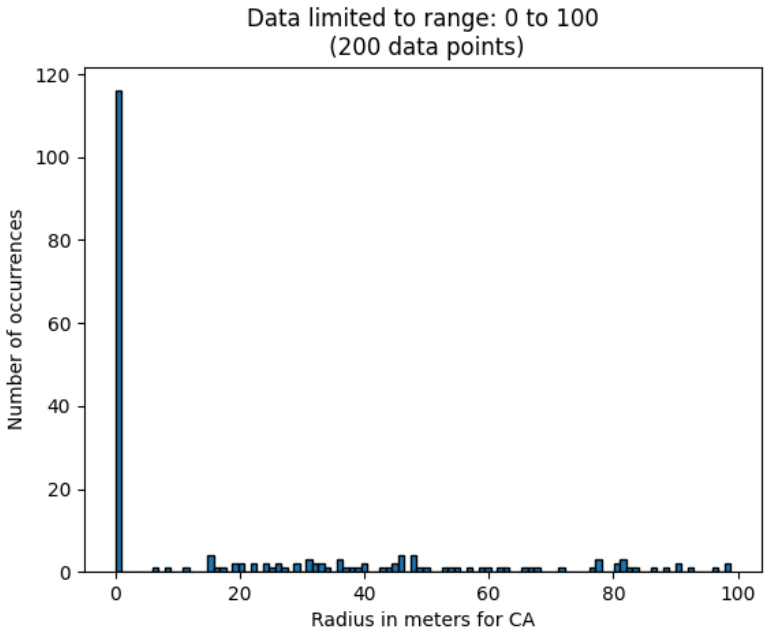
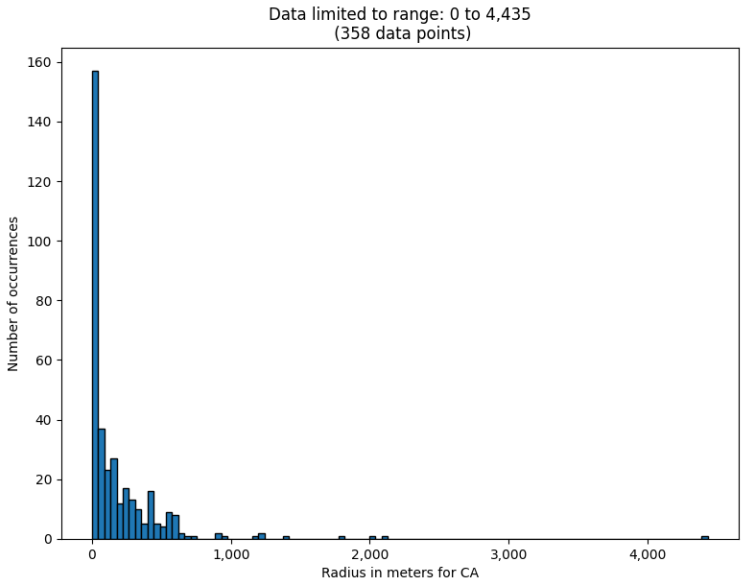


Figure 13: Radius histogram for CA limited to 4,435 meters.



7 Future work

There are a number of areas for future work:

- Compare the current 2019 year data with other years. 2020 might be an interesting year to compare against because of the marked decrease in air traffic due to COVID-19.
- Identify tail numbers that have international arrivals and departures, and see how their activities compare to their domestic counterparts.
- Explore why airports locations change (regulatory requirements, change in datums, continental drift[1]
- Write a GUI frontend to make command line argument usage transparent.
- Expand the command line arguments to include Postgres user name, password, and database name.
- Review the circular position versus days between first and last updates (see Figure 9) to understand/explain outliers.
- Expand the data collection to include arrival data. The current effort made the simplifing assumption that because a tail number departed an airport, that it arrived there just prior to departing. This assumption has not been verified.

8 Conclusion

In this exploration, we started with question of where did a particular aircraft (based on tail number) depart from during 2019. We answered that question by searching Bureau of Transportation Statistics data based for and finding the 82 domestic airports recording the departure of tail number N287WN. To have a better understanding of where these airports were located, geographic data was used to create a map of the airports. The geographic data had multiple date stamped positional reports for many of the airports, so a it was possible to see how large the change in position each airport experienced over the course of its existence.

We were able to programmatically query an on-line service to obtain airport departure data. We were able to parse that data to understand airport usage, daily take offs, and hourly take offs. Based on airport geographic data, we were able to construct a map of airport usage, evaluate changes in airport location, and identify data that deserves further investigation.

[1] Australia moves about 2.6cm per year. North America moves about 1cm per year.[3]

A Miscellaneous files

A collection of miscellaneous files mentioned in the report.

- followAircraft.py – The Python script used to analyze the data, and create the images in this report.
- 833374230_T_MASTER_CORD.zip – Airport location file. Data downloaded from `https://www.transtats.bts.gov/Fields.asp?Table_ID=288`

These files can be extracted from this report using something like the Adobe Acrobat Reader application. You may not be able to extract them using a Web browser.

B References

[1] Bureau of Transportation Statistics Staff, *Bts detailed statistics departures*, `https://www.transtats.bts.gov/ONTIME/Departures.aspx`, 2020.

[2] Federal Aviation Administration Staff, *Faa registry*, `https://registry.faa.gov/aircraftinquiry/NNum_Inquiry.aspx`, 2020.

[3] National Geographic Staff, *Continental drift*, `https://www.nationalgeographic.org/encyclopedia/continental-drift/`, 2020.