

FINAL PROJECT: AIRLINE DELAYS

AUTHOR
Tess Downing and Denali Stevens

INTRODUCTION: AIRLINE DELAYS

Everyday, thousands of Americans use Airports as a means of travel. Many of these people are flying because it is the fastest mode of transportation. However, plans can be ruined when delays happen. Using the `airline_delay` data set from OpenIntro (https://www.openintro.org/data/index.php?data=airline_delay), we set out to find which Airports in the US are the worst to fly out of based off of their delays. The data set contains information from December 2019 to December 2020. Once the airports are determined we plan on investigating the length of delays and the breakdown (weather, security, and carriers) of what is causing the delays. To calculate accurate data, we will use the proportions of outgoing flights to account for some airports being larger than others. However, the data set does not have a variable for outgoing flights, it only has one for incoming flights (`arr_flights`). To aid this, we will assume that the number of incoming flights is the same as the amount of outgoing flight. We are also only going to look at airports that have more than 1000 flights in 2019 and more than 800 flights in 2020 because these are the airports that people are more likely to fly out of and they have less skewed data because there are more entries.

To do so we will examine the following variables:

- `arr_flights`, Number of flights arriving at airport
- `year`, Year data collected
- `carrier_name`, Carrier Name
- `airport_name`, the name of the airport
- `airport`, Airport code
- `carrier_ct`, Number of flights delayed due to air carrier. (e.g. no crew)
- `weather_ct`, Number of flights due to weather
- `security_ct`, Number of flights canceled due to a security breach
- `carrier_delay`, Total time (minutes) of delay due to air carrier
- `weather_delay`, Total time (minutes) of delay due to inclement weather
- `security_delay`, Total time (minutes) of delay as a result of a security issue

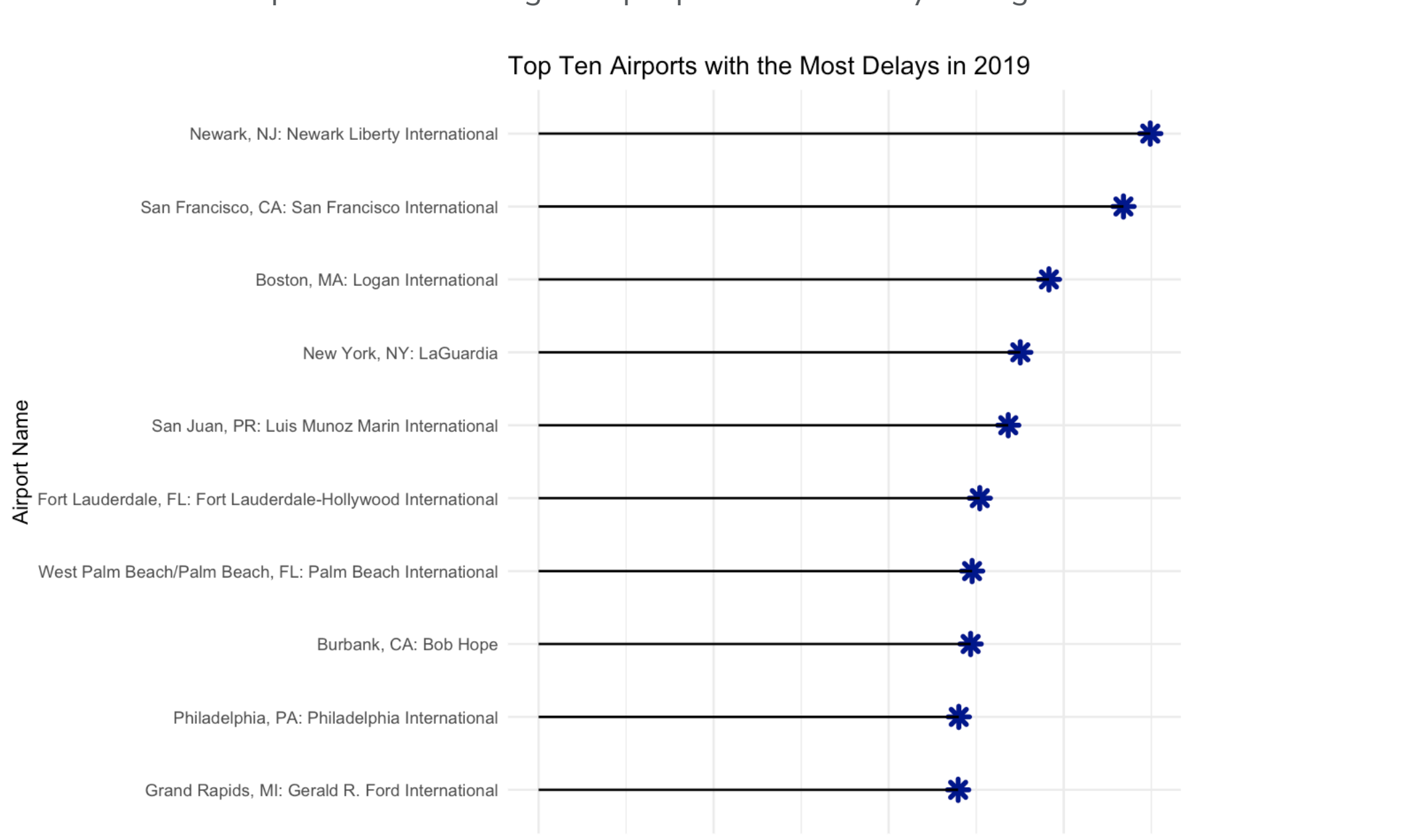
CONTENT:

- Find the total amount of delays for each airport per year and the proportion of outgoing flights that are delayed. Exclude airports that had less than 1000 outgoing flights in 2019, and less than 800 outgoing flights in 2020.

```
# A tibble: 93 × 4
# Groups:   airport airport_name [93]
  year airport airport_name
<dbl> <chr>   <chr>
1 2019 EWR    Newark, NJ: Newark Liberty International
2 2019 SFO    San Francisco, CA: San Francisco International
3 2019 BOS    Boston, MA: Logan International
4 2019 LGA    New York, NY: LaGuardia
5 2019 SJU    San Juan, PR: Luis Munoz Marin International
6 2019 FLL    Fort Lauderdale, FL: Fort Lauderdale-Hollywood Intern...
7 2019 PBI    West Palm Beach/Palm Beach, FL: Palm Beach Internation...
8 2019 BUR    Burbank, CA: Bob Hope
9 2019 PHL    Philadelphia, PA: Philadelphia International
10 2019 GRR    Grand Rapids, MI: Gerald R. Ford International
# ... with 83 more rows, and abbreviated variable name 'prop_delayed'

# A tibble: 79 × 4
# Groups:   airport airport_name [79]
  year airport airport_name
<dbl> <chr>   <chr>
1 2020 DFW    Dallas/Fort Worth, TX: Dallas/Fort Worth International
2 2020 BDL    Hartford, CT: Bradley International
3 2020 PBI    West Palm Beach/Palm Beach, FL: Palm Beach Internation...
4 2020 JFK    New York, NY: John F. Kennedy International
5 2020 SAT    San Antonio, TX: San Antonio International
6 2020 ANC    Anchorage, AK: Ted Stevens Anchorage International
7 2020 FLL    Fort Lauderdale, FL: Fort Lauderdale-Hollywood Interna...
8 2020 DEN    Denver, CO: Denver International
9 2020 AUS    Austin, TX: Austin – Bergstrom International
10 2020 BOS    Boston, MA: Logan International
# ... with 69 more rows, and abbreviated variable name 'prop_delayed'
```

- Which 10 airports had the highest proportion of delayed flights in 2019?



- Which 10 airports had the highest proportion of delayed flights in 2020?



- Find the airports that were in the top 10 proportion of delays for 2019 and 2020.

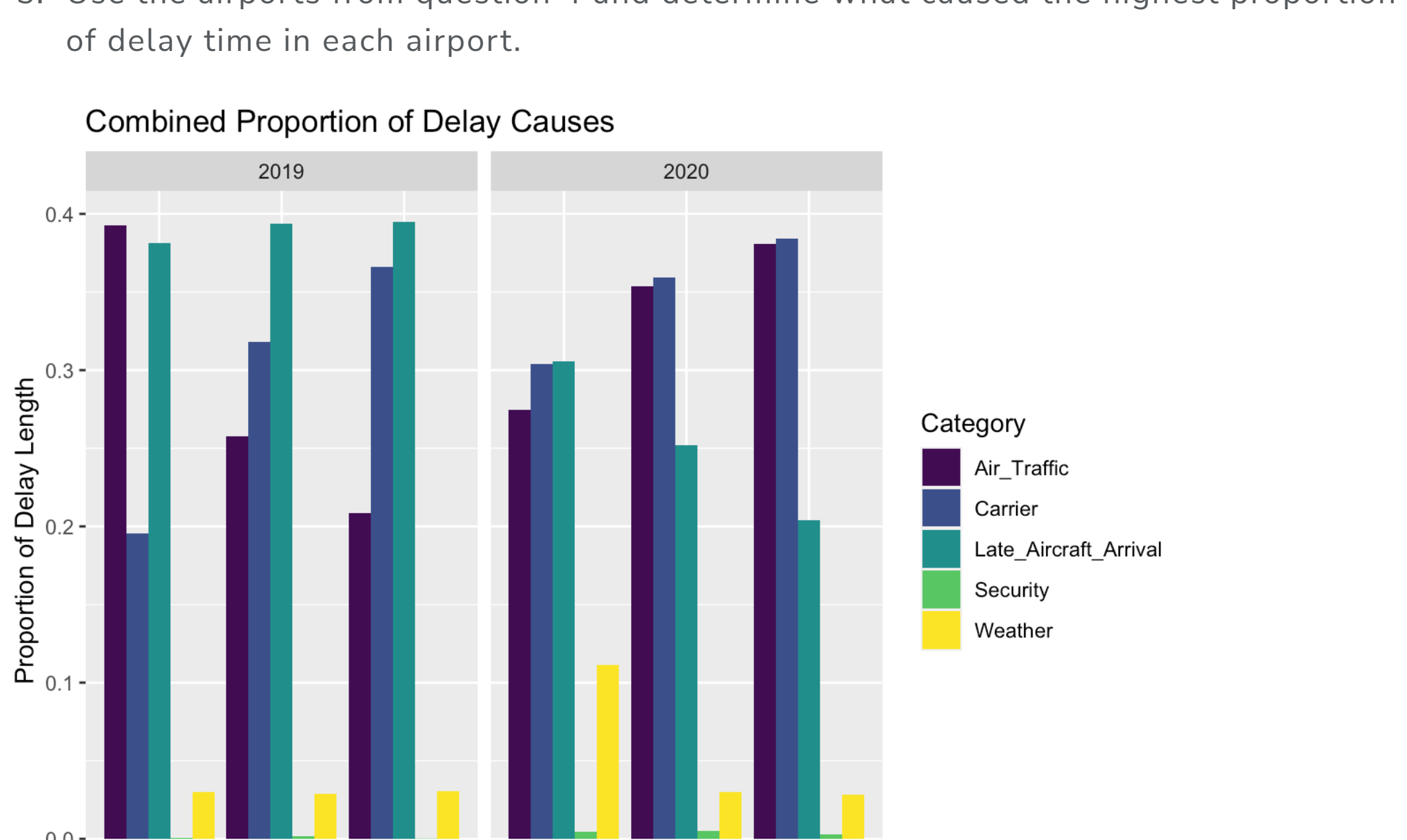
AIRPORT_NAME
Boston, MA: Logan International
Fort Lauderdale, FL: Fort Lauderdale-Hollywood International
West Palm Beach/Palm Beach, FL: Palm Beach International

Boston, MA: Logan International, Fort Lauderdale, FL: Fort Lauderdale-Hollywood International, West Palm Beach/Palm Beach, FL: Palm Beach International all appear in both 2019 and 2020 with some of the highest proportions of delayed flights.

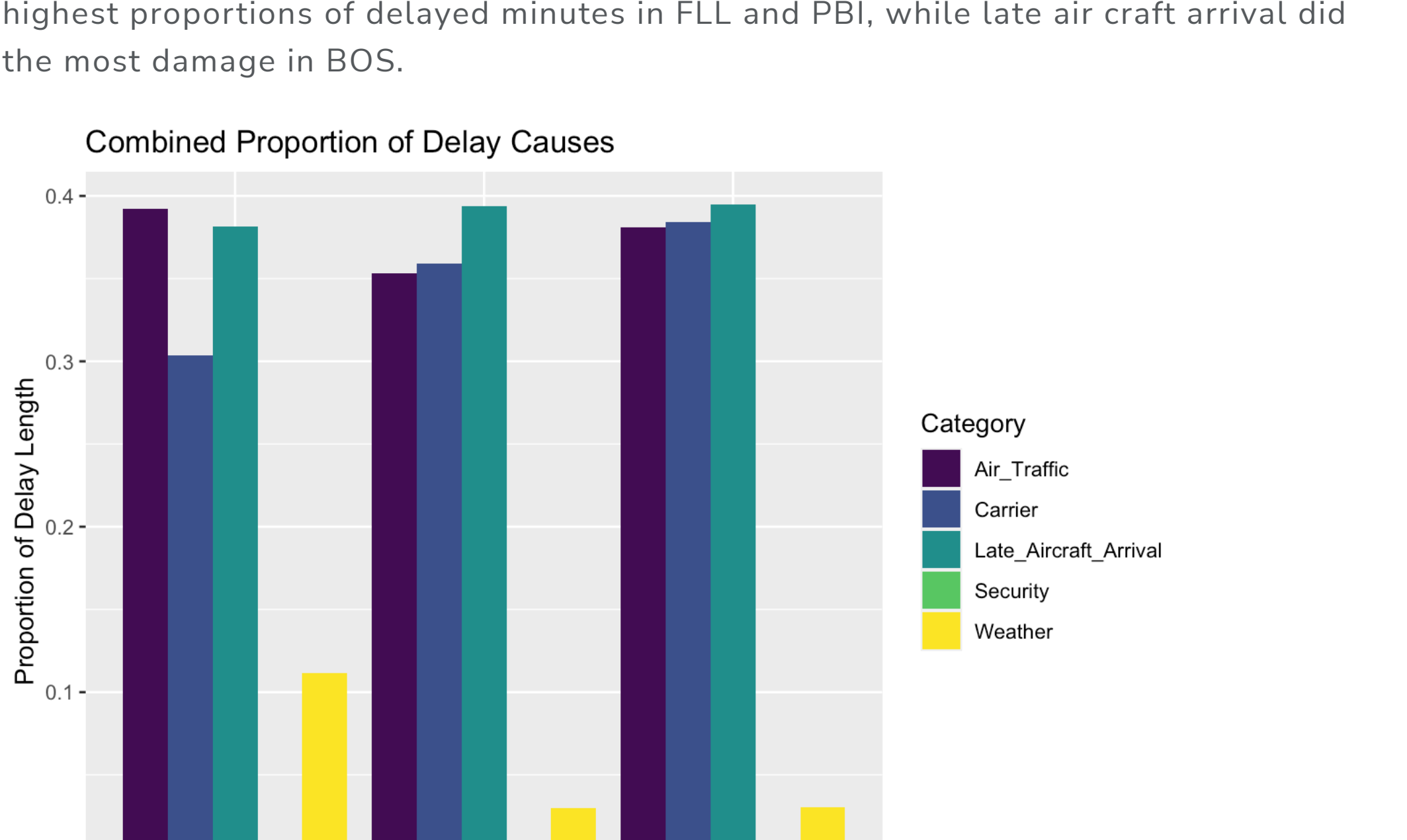
- Interpret what it might mean for an airport to be on this list.

For an airport to be on this list means their proportion of delayed flights were top 10 in the country for 2019 and 2020. This means given the limited data we have, we can say these airports have some of the highest proportion of delays over a span of multiple years, making them poor airports to travel through.

- Use the airports from question 4 and determine what caused the highest proportion of delay time in each airport.



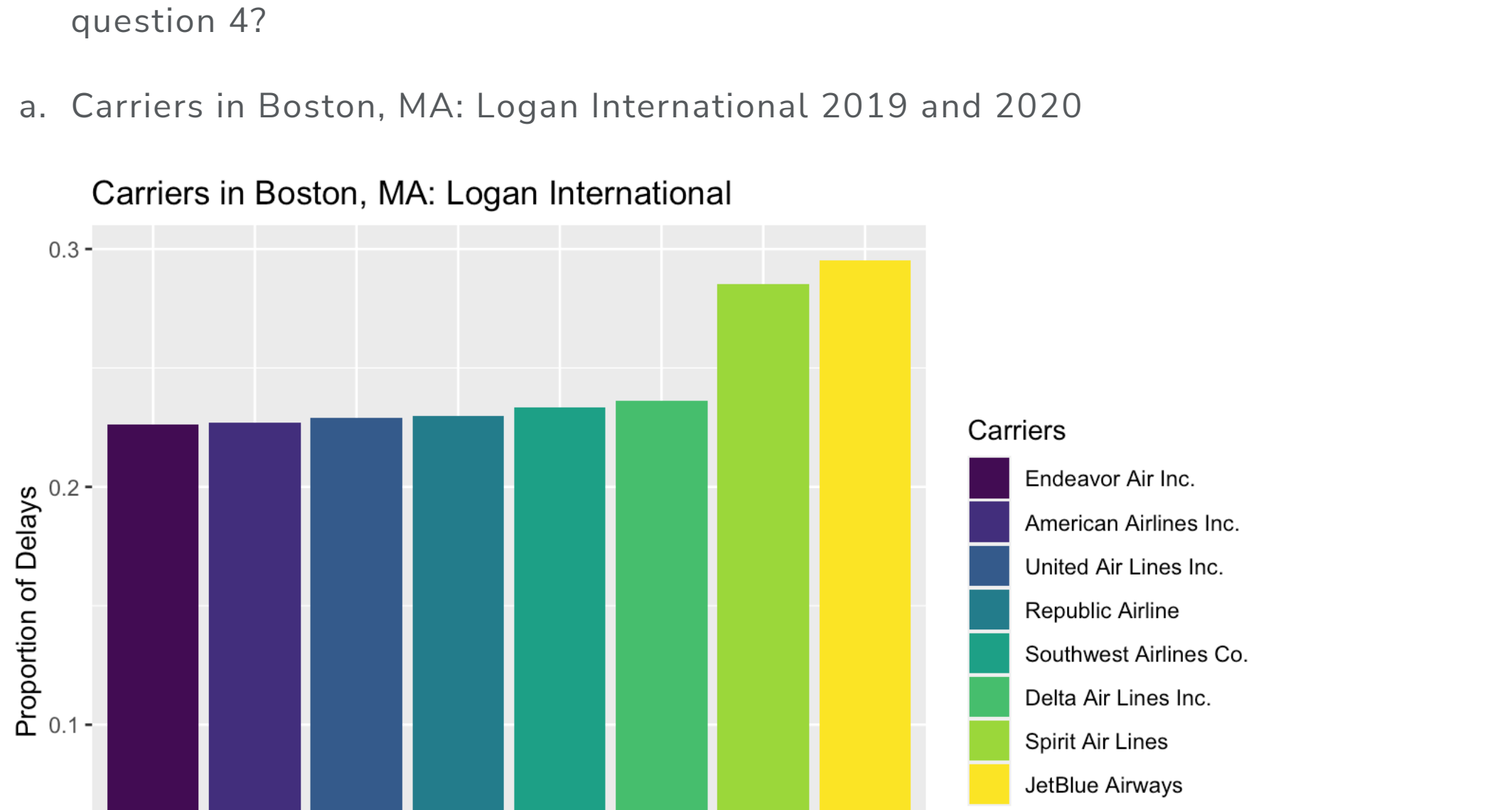
In 2019 late aircraft arrival caused the highest proportions of delayed minutes in FLL and PBI, while air traffic was the leading category in BOS. In 2020 carrier caused the highest proportions of delayed minutes in FLL and PBI, while late air craft arrival did the most damage in BOS.



With the years combined, late air craft arrival accounted for the highest proportions of delayed minutes in FLL and PBI, and air traffic caused the highest proportions of delayed minutes in BOS.

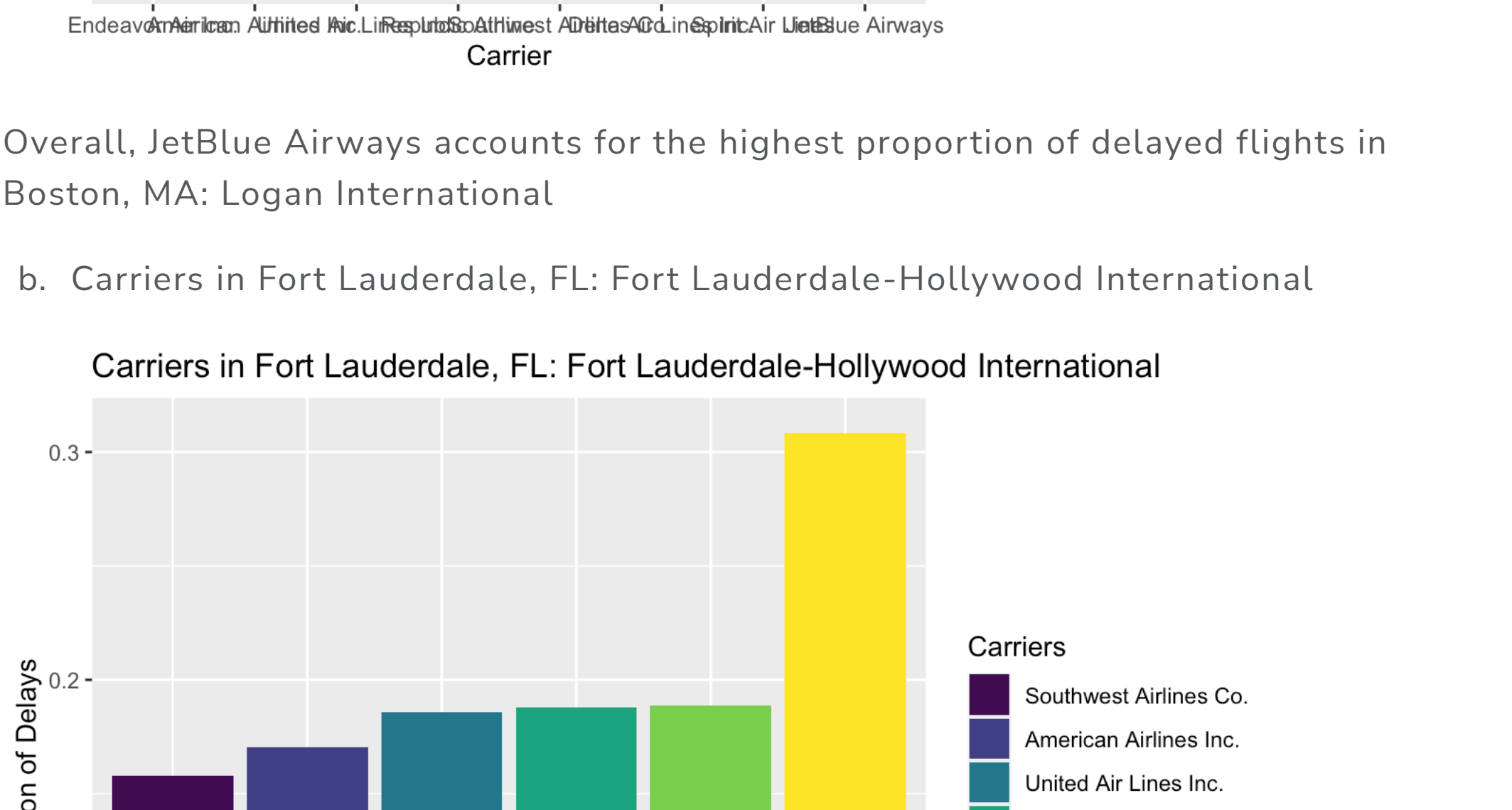
- Which carrier accounts for the highest proportion of delays at each airport from question 4?

- Carriers in Boston, MA: Logan International 2019 and 2020



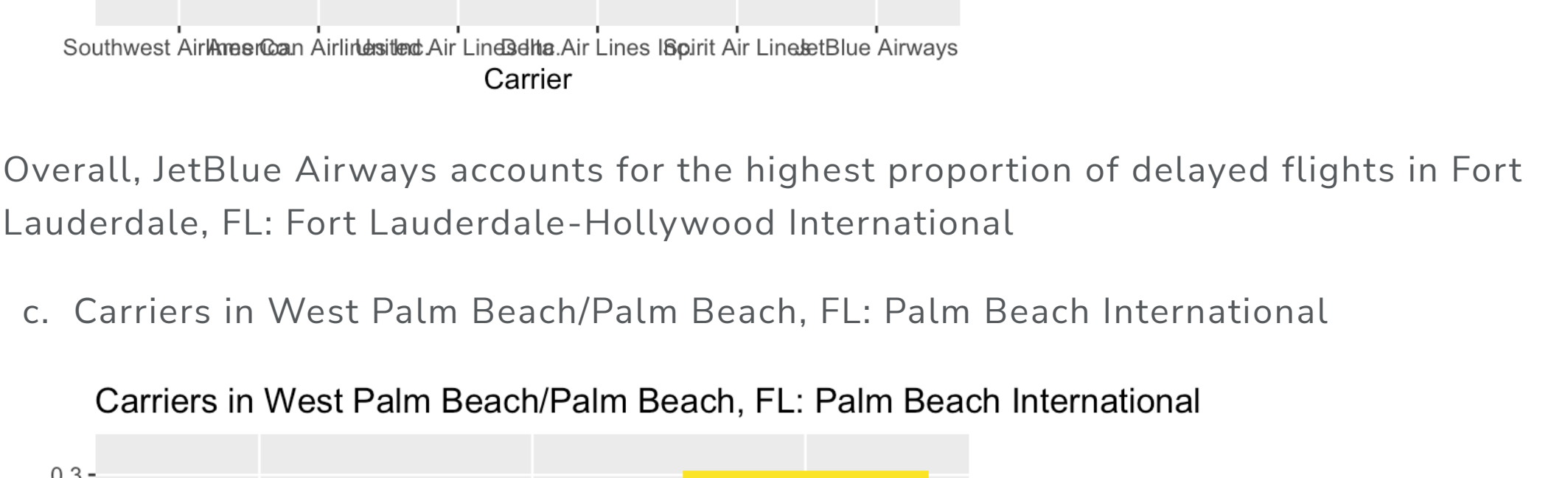
Overall, JetBlue Airways accounts for the highest proportion of delayed flights in Boston, MA: Logan International

- Carriers in Fort Lauderdale, FL: Fort Lauderdale-Hollywood International



Overall, JetBlue Airways accounts for the highest proportion of delayed flights in Fort Lauderdale, FL: Fort Lauderdale-Hollywood International

- Carriers in West Palm Beach/Palm Beach, FL: Palm Beach International



Overall, JetBlue Airways accounts for the highest proportion of delayed flights in West Palm Beach/Palm Beach, FL: Palm Beach International

- What is the difference between the national average proportion of delayed flights in 2019 and 2020

YEAR	NAT_AVERAGE
2019	0.2036
2020	0.1159

[1] 0.08769787

CONCLUSION

As in all studies there were some limitations in ours. One limitation was COVID 19. There were significantly less flights over all in 2020 due to the pandemic. This is why we only looked at airports that had more than 800 flights for that year. Also if we did not take year into account, none of the data from 2020 would have made it into the highest proportion of delays because there were so many less flights, so we decided to keep them separate when asking questions. Another limitation was that the data was only from two years. If there was more data from more years, we could have gotten more accurate results about which airports have the worst delays.

WHAT AIRPORT IS THE WORST TO FLY OUT OF IN THE UNITED STATES OF AMERICA?

Based off of the different questions we explored, we have determined that Boston, MA: Logan International is the worst American airport to fly out of. This airport had 29.2% of it's departing flights delayed in 2019 and 14.9% delayed in 2020. The amount of carrier delays in proportion to the amount of outgoing flights for 2019 and 2020 were 0.0725 and 0.0458, respectively. The duration of delays were 28.9 minutes in 2019 and 13.1 minutes in 2020. The proportion of minutes delayed due to security in 2019 was 0.000140 and 0.00800 in 2020. The proportion of minutes delayed due to weather in 2019 was 0.0277 and 0.0109 in 2020.