Intro to ML Exam

Sahil Natu

8/1/2021

Book Problems

Question 2

Visual story telling part 2: flights at ABIA**

The flight data contains data for all flights that departed from and arrived at the Austin airport. We will analyze this data to search for patterns and answer some questions.

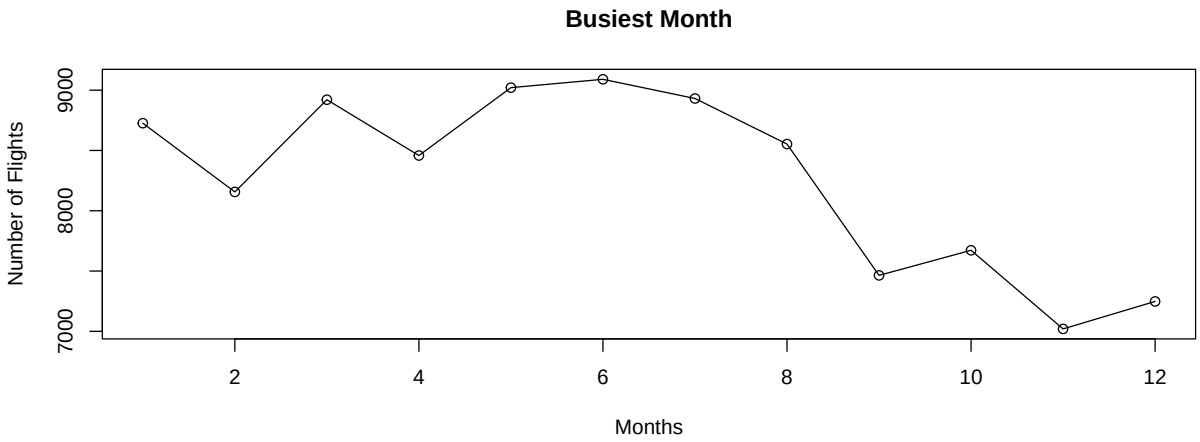
```
## [1] 99260    29
```

```
##  [1] "Year"              "Month"             "DayofMonth"
##  [4] "DayOfWeek"         "DepTime"           "CRSDepTime"
##  [7] "ArrTime"           "CRSArrTime"        "UniqueCarrier"
## [10] "FlightNum"         "TailNum"           "ActualElapsedTime"
## [13] "CRSElapsedTime"    "AirTime"           "ArrDelay"
## [16] "DepDelay"          "Origin"            "Dest"
## [19] "Distance"          "TaxiIn"            "TaxiOut"
## [22] "Cancelled"         "CancellationCode"  "Diverted"
## [25] "CarrierDelay"      "WeatherDelay"      "NASDelay"
## [28] "SecurityDelay"     "LateAircraftDelay"
```

The data consists of 99260 rows and 29 variables. The variable names have been listed above.

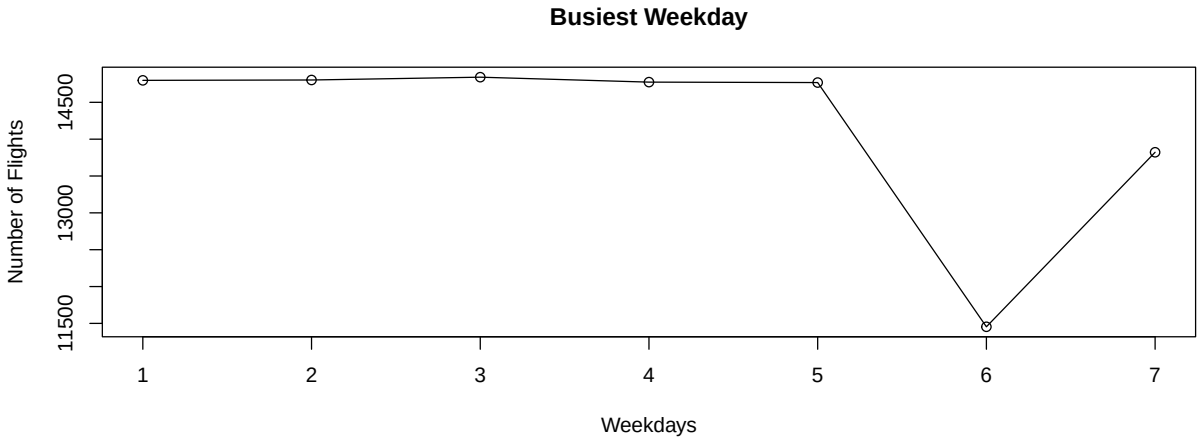
Busiest Month

We will begin by understanding the temporal distribution of flights across the year to find the busiest month. Spring and Summer are the busiest times of the year at the Austin airport with a maximum of about 9,000 flights handled in the month of June.



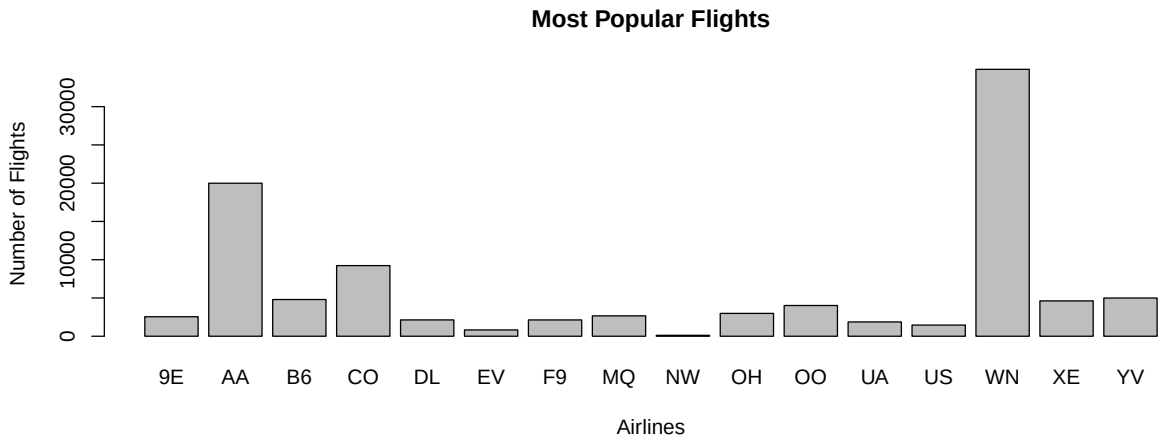
Busiest Weekday

Next up, we will have a look at the temporal distribution of flights across all weeks to find the busiest weekday.
All 5 working days, Monday to Friday are equally busy at the airport with a fall in flights on Saturday and a slight bounce back on Sunday.



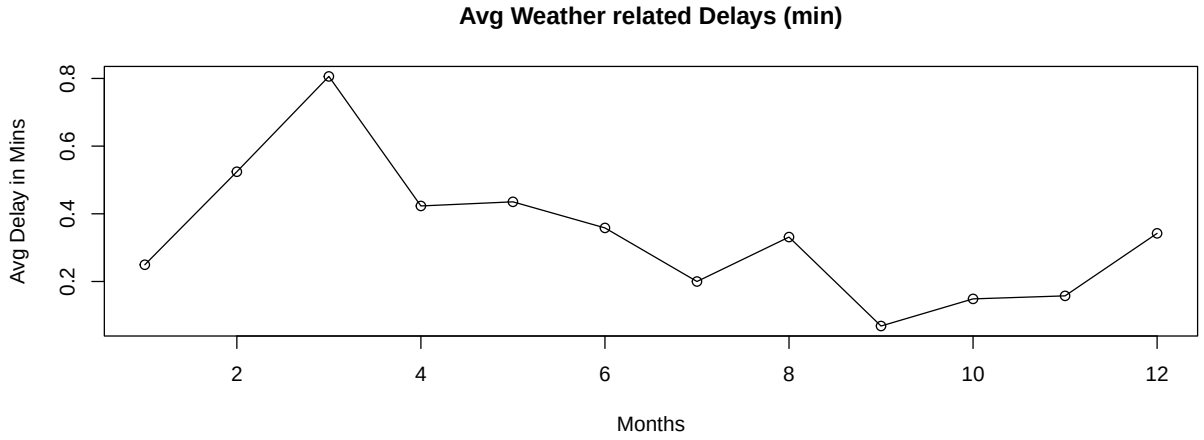
Most Popular Airline

Next up, we look at distribution of flights across various carriers to find the most popular airlines.
Southwest Airlines (WN) is the most popular airlines with over 30,000 flights in 2008. This is followed by American Airlines (AA) and Continental Airlines (CO).

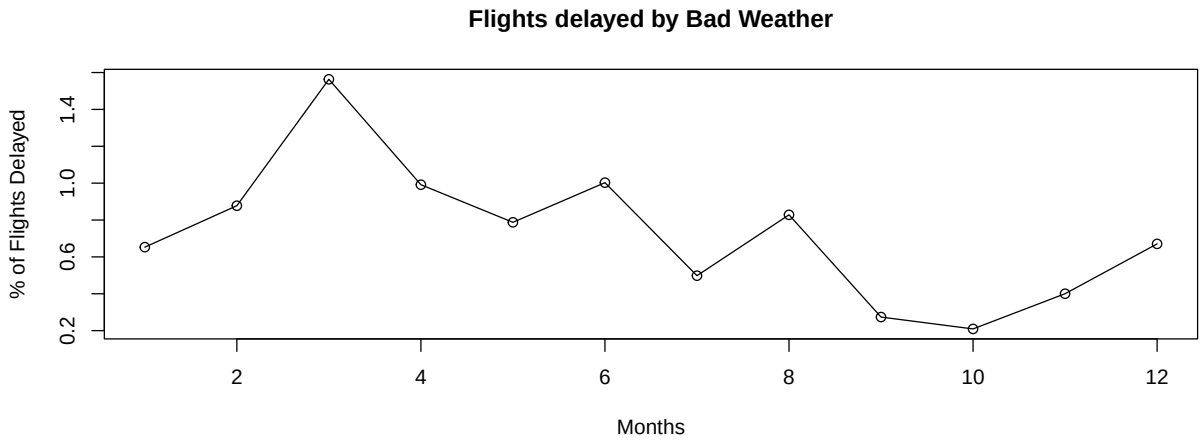


Departure Delays due to Bad Weather

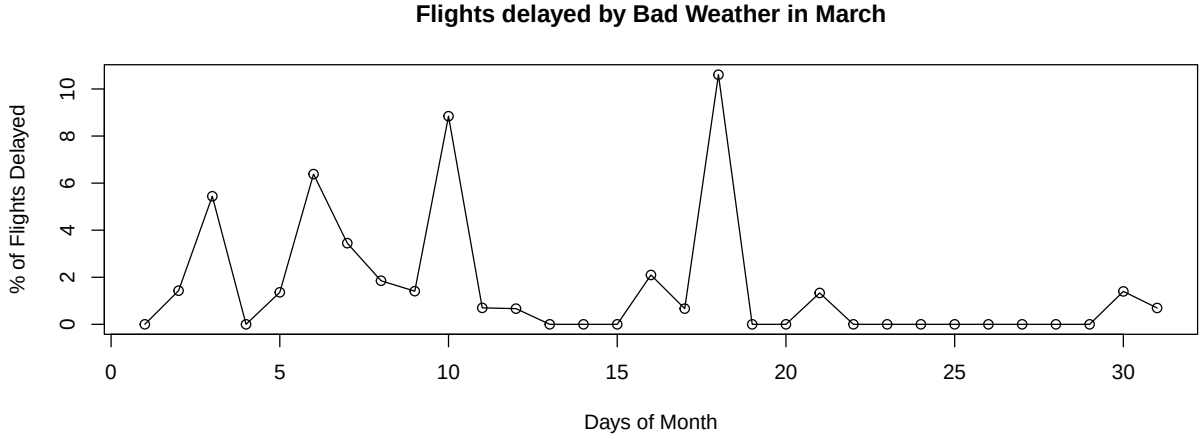
We will analyze the data for delayed departures due to bad weather. We start with checking for delay in departure due to bad weather across multiple months.



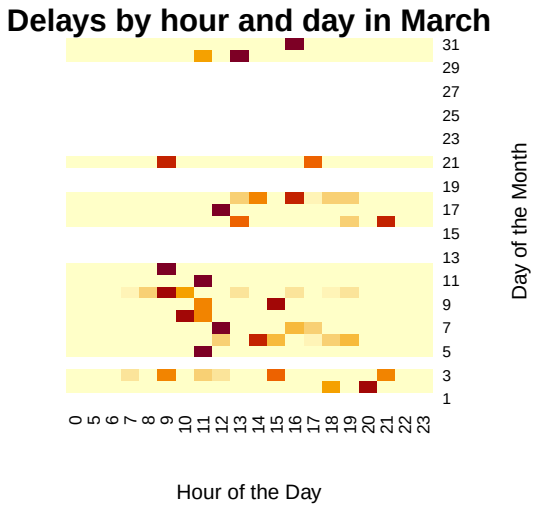
While this provides us with average delay in minutes due to bad weather across all months, we want to further see the frequency of weather related delay occurrences in each month.



We find similar results to the previous graph with a maximum delays occurring in March. There is 1 weather related delay among every 65 flights in March. With March being the month with most weather related delays, we check for frequency of delays in March.

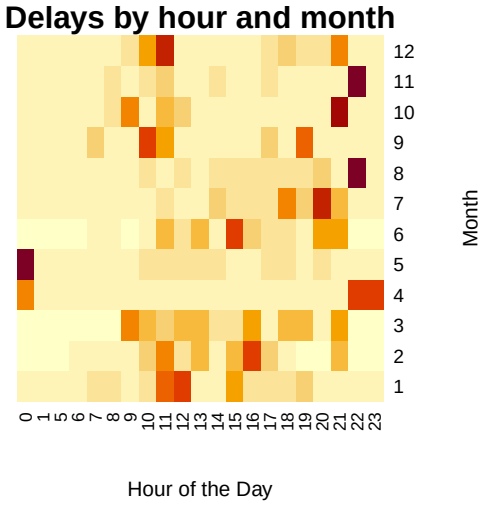


Since flights were delayed for almost all days across the month (barring a week at the end), the delays cannot be attributed to a standalone weather event. Thus, we can conclude that March has the highest number of weather related delays at the Austin airport. We can further break down weather related delays in March by time of the day to figure out flights at which hours are most likely to be affected by bad weather.



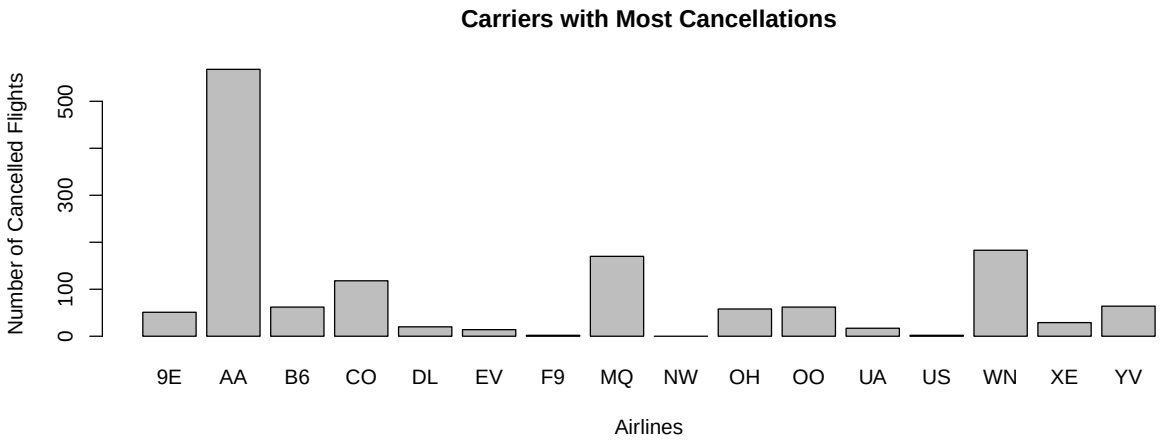
We can infer from the heat map that majority of the weather related delays in March have happened before noon and none have happened during the night time. We can plot a similar heat map for all months.

When looking at data for all the months, we still see a trend of no weather related delays during the night hours with delays concentrated in the morning hours right before noon and in the late evening hours.

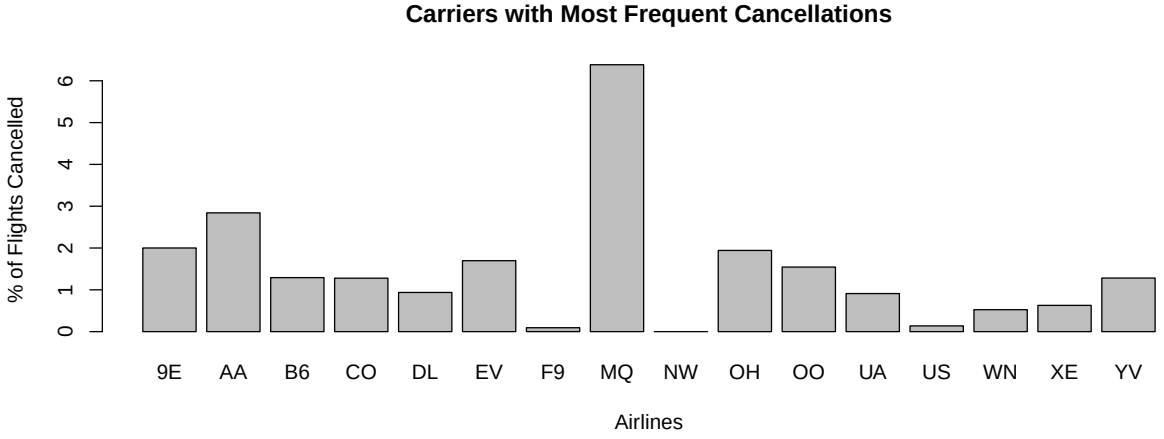


Flight Cancellations

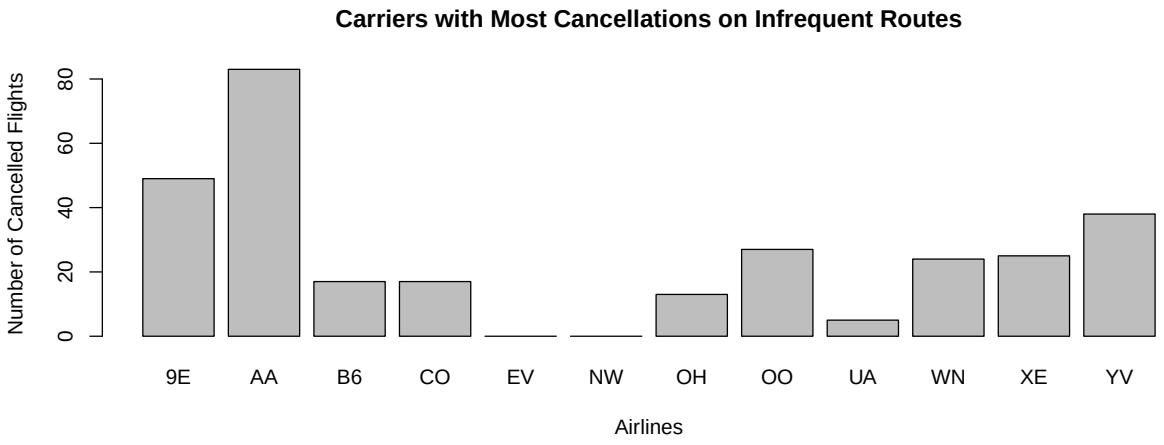
Here, we begin by looking at flight cancellations by various carriers.



American Airlines (AA) has the dubious distinction of being the carrier with most cancelled flights. However, this is an incomplete picture since all carriers may not be flying as many flights in and out of Austin as American. Hence, we look at cancellation frequency among the carriers.

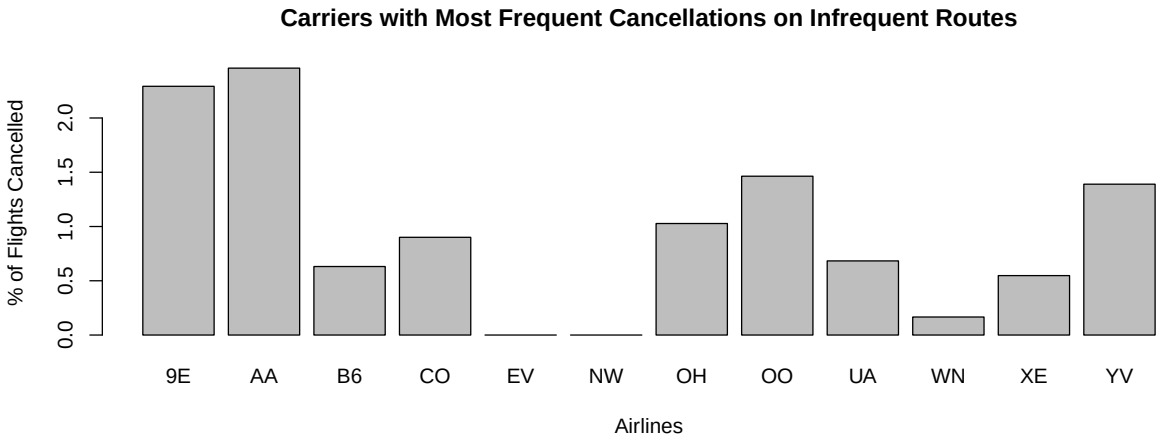


It is now apparent that Envoy Air (MQ) has the highest frequency of cancellations with 1 in every 16 of its flights getting cancelled. Another important metric to look at when studying flight cancellations is frequency of cancellation on routes that do not have more than 1 daily connection. A cancelled flight on such a route means waiting at least a day before the next flight. We will call such routes as *infrequent routes*.



Again, American Airlines (AA) pops up as the carrier with most cancelled flights. We further look at carrier wise flight cancellation frequency for infrequent routes.

American Airlines (AA) and Endeavor Air (9E) are carriers with highest cancellation frequencies for infrequent routes. Surprisingly, Envoy Air (MQ), which had the worst cancellation rate overall, does not feature here as it does not serve any infrequent route. ExpressJet Airlines (EV) and Northwest Airlines (NW) have had no cancellations serving infrequent routes, thus these carriers can be relied upon when traveling to or from uncommon destinations.

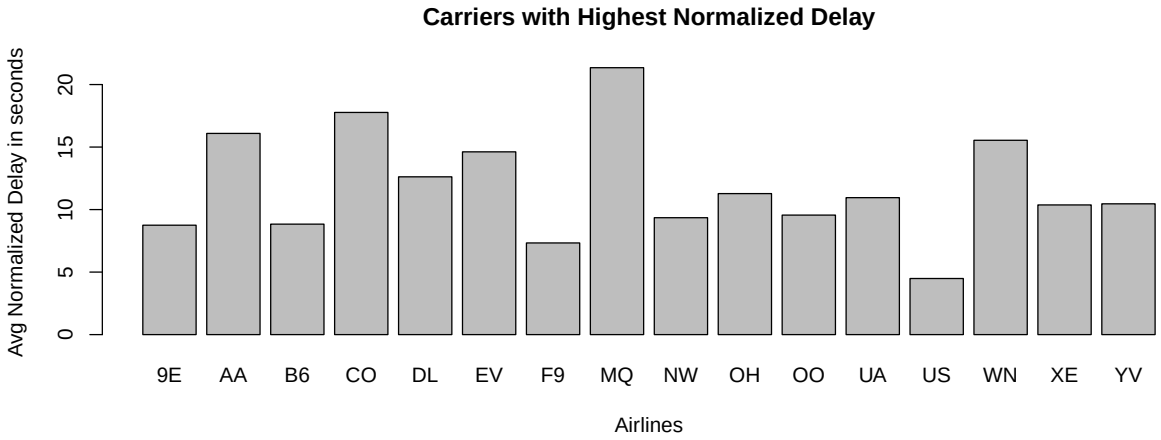


Normalized Delay

Let us now look at Normalized Delay (with respect to total flight time) across multiple variables. Not all delays are equal. A 30-minute delay on a flight to Anchorage is not the same as a 30-minute delay on a short flight like Austin to Dallas or Austin to Houston. We will normalize the delay by dividing it by total flight duration. This gives us delay per unit route duration.

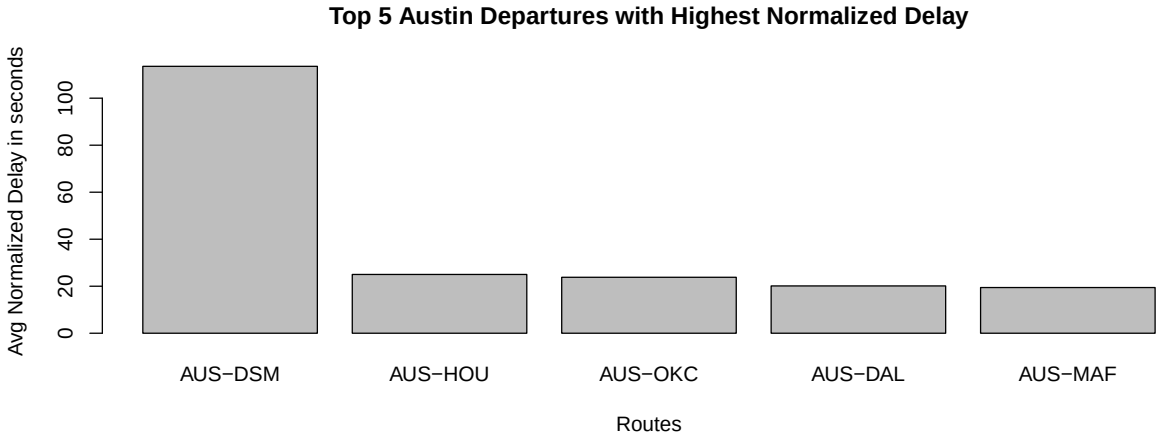
Note: We calculate total normalized delay as sum of normalized departure and arrival delays. While these 2 delays cannot be added directly as a delay in departure will inadvertently lead to delayed arrival, we

still proceed with this formula to penalize flights that were not able to compensate their delayed departure in-flight as compared to those flights which had a delayed departure but were able to compensate for the delay by arriving before the estimated time of arrival based on the route flight time.



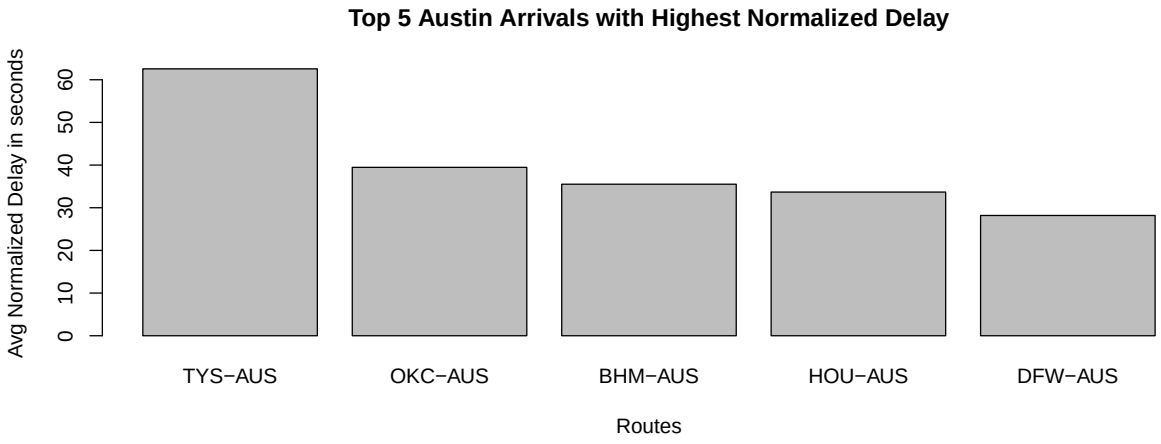
As evident from the bar plot, Envoy Air (MQ) has the highest normalized delays to the tune of 20 seconds for every 1 minute of flight time. US Airways (US) has the least delays. We can further look at routes with the most amount of normalized delays, to and from Austin.

1) Routes with highest delays flying out of Austin
The plot throws up Austin to Des Moines as the most delayed route with a delay of 2 minutes for every 1 minute of flight time. This looks fishy, and hence upon further inspection, it turns out that the data has just 1 flight that has flown this route, and the extraordinary delay belongs to this flight. Since we cannot generalize with such a small data sample, Austin to Houston is the next most delayed route with high number of routine flights and an average delay of 20 seconds for every 1 minute of flight time.

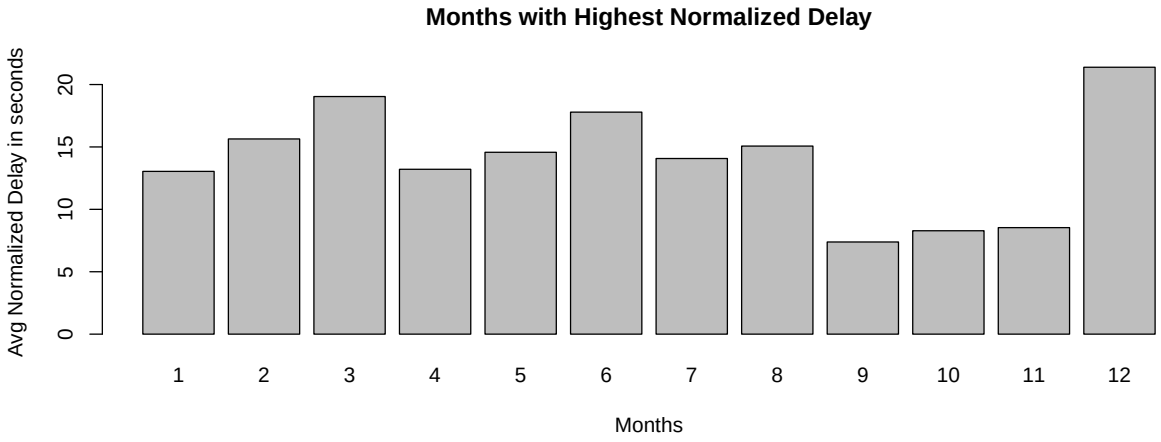


2) Routes with highest delays flying into Austin
As we can see, flights from McGhee Tyson Airport (Knoxville,TN) have the highest delays averaging over 1 minute of delay for every 1 minute of flight time. However, there were only 3 flights from this airport to

Austin in the year 2008, and hence the sample is too small to generalize that route as the most delayed at arrival in Austin. Hence, we look at the next route from Oklahoma City to Austin. Flights from Oklahoma City to Austin are delayed by an average of 40 seconds for every 1 minute of flight time. Thus, this is the most delayed route coming in to Austin.
Oklahoma City and Houston have featured in both the plots and thus we can say with confidence that flights to and from Oklahoma City and Houston are the most delayed relative to their flight times.



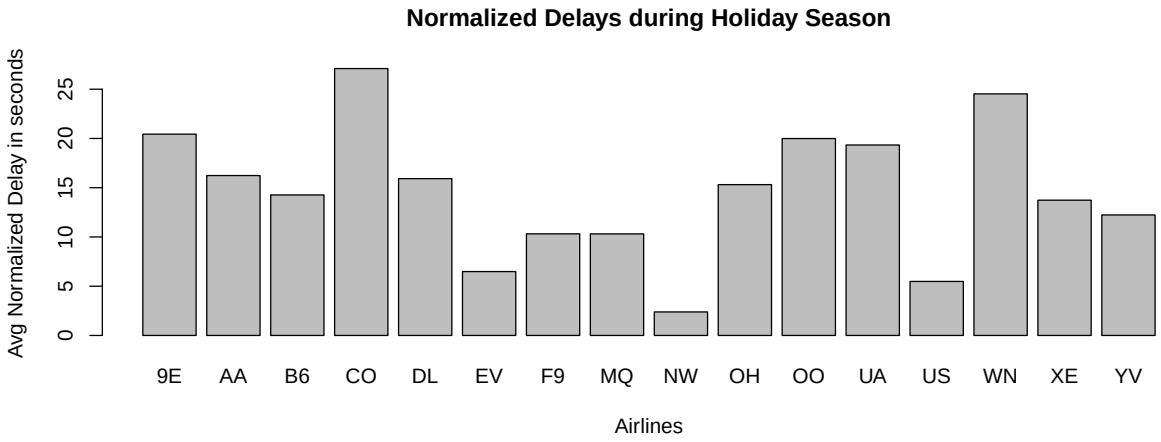
We can further look at normalized delays across various months.
The average normalized delays are the highest in December and March, i.e. during the holiday season and in Spring.



Answering some questions for travelers -

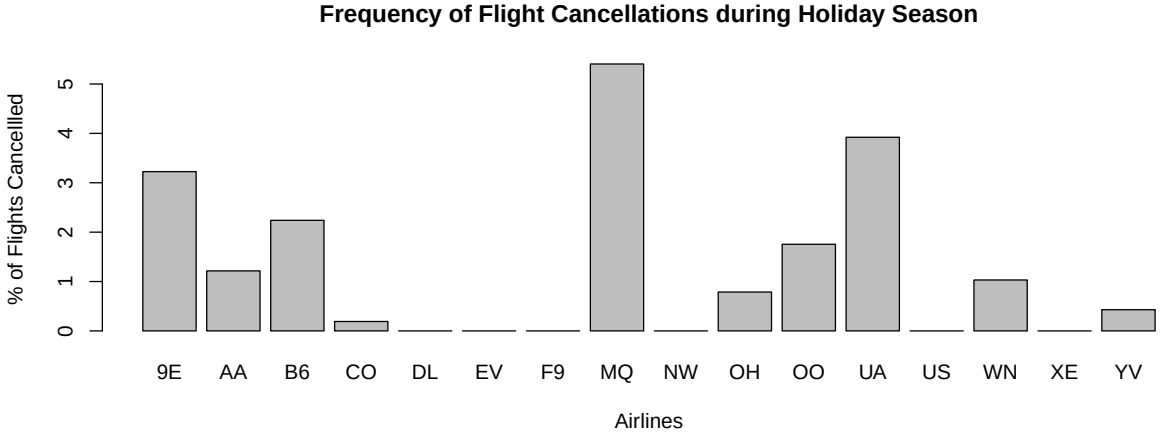
a) Which carriers could be avoided during the Holiday Season?

Any flight that gets cancelled or delayed during the Holiday Season puts a damper on one's spirits. Hence a traveler might want to avoid such flights. We specifically look for cancellations and/or delays between December 20 and January 10 to answer this question.



Continental Airlines (CO) has the highest delays, an average of over 25 seconds per 1 minute of flight time. Northwest Airlines (NW) has the least average delay. The observations are different from those we obtained earlier when we had considered the data for all months. Envoy Air (MQ) has been the worst performing carrier overall, but that has been replaced by Continental Airlines (CO) during the holiday season, followed by Southwest Airlines (WN).

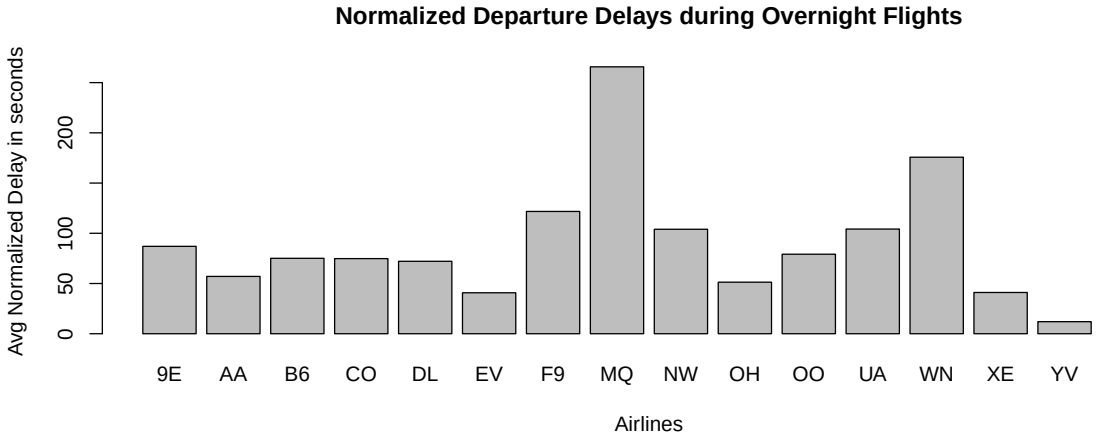
Upon looking at flight cancellations, it is evident that Envoy Air (MQ) has the highest cancellations at 5% of their total scheduled flights, followed by United Airlines (UA). Thus some of the carriers to avoid during the holiday season are Continental Airlines, Southwest Airlines, Envoy Air and United Airlines. One can expect a smoother start or end to their holiday travels by hopping onto a flight operated by Northwest Airlines, US Airways or ExpressJet Airlines.



b) Which overnight flights could be avoided?

Overnight flights or red-eye flights are the ones that have a substantial part of their journey during the night hours. Travelers generally prefer these flights to avoid spending precious daylight hours in traveling. Any delay in these flights not only throws off the plans, but also causes sleep deprivation as a traveler might

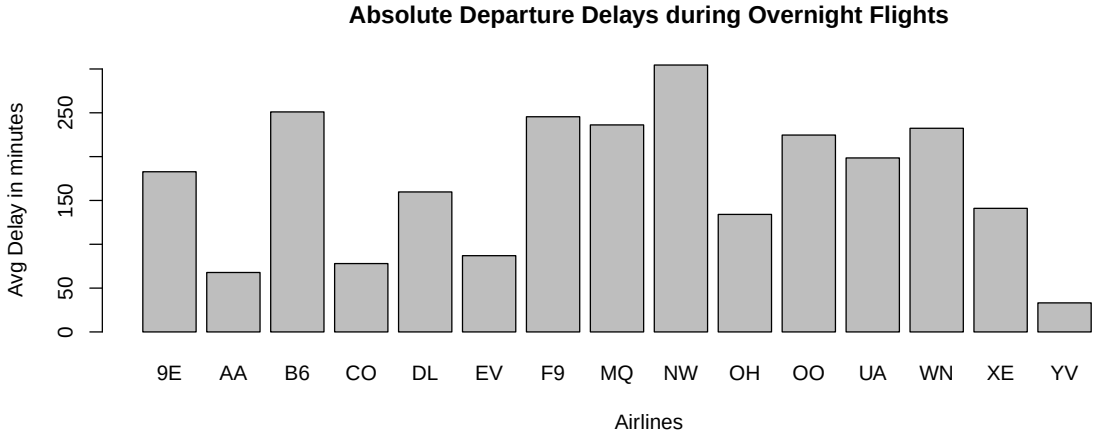
be expecting to sleep during these hours in the flight. We look at departure delays in flights that depart between 11:00 PM and 3:59 AM to answer this question.



Envoy Air (MQ) and Southwest Airlines (WN) have the highest normalized delays at departure for overnight flights. Departure delays in overnight flights are a cause of concern irrespective of the duration of the flight that is to be boarded. Thus, we also need to look at absolute departure delays in overnight flights and not just the normalized delays.

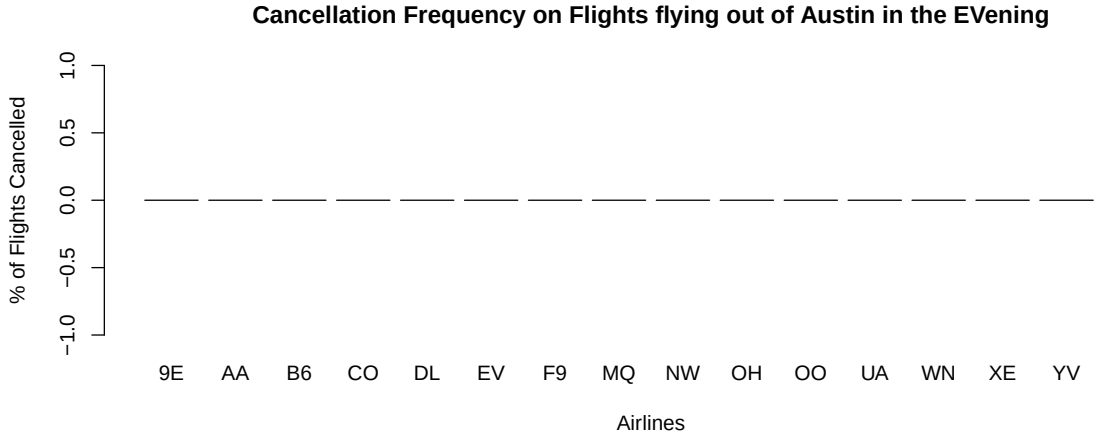
With absolute delays considered, Northwest Airlines (NW) is the worst performing airline with highest average delay per flight at 5 hours, followed by JetBlue Airways (B6) and Frontier Airlines (F9). The best performing airlines for this metric are Mesa Airlines (YV), American Airlines (AA), and Continental Airlines (CO). Envoy Air (MQ) and Southwest Airlines (WN) still show up average delays on the higher side (~4 hours).

Thus, the carriers to avoid on overnight flights are Northwest Airlines, JetBlue Airways, Frontier Airlines, Southwest Airlines and Envoy Air.



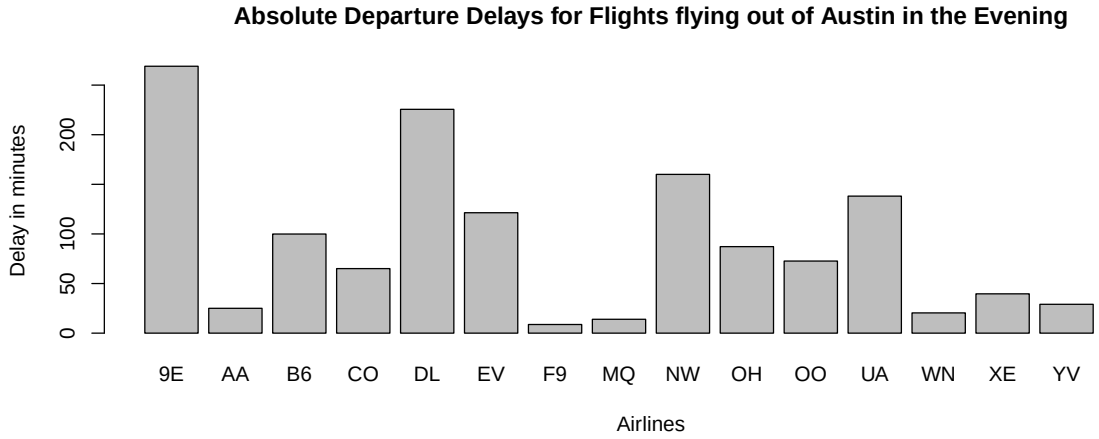
c) Which flights could be avoided if you are flying out of Austin at the end of the day?

Many travelers take a flight out of a city at the end of the day to save on accommodation costs for the night. A person can spend the entire day in the city and fly out in late evening hours. However, if such a flight is cancelled, the traveler seldom has options apart from staying the night in the city, which entails additional costs for accommodation over and above their planned trip budget. To answer this question, we will be looking at flight cancellations data for departures from Austin between 7:00 PM and 10:59 PM.



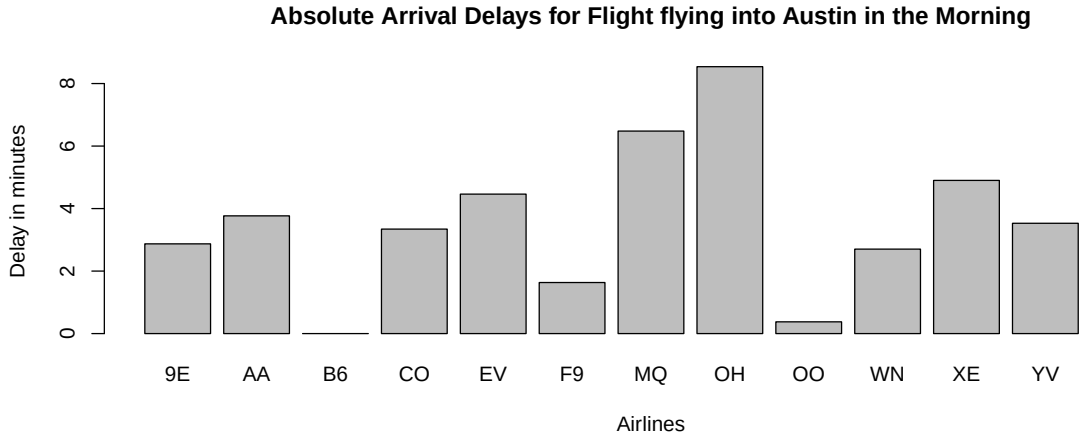
It turns out that no flights flying out of Austin during these hours were cancelled in 2008. To further broaden our problem, we will look at delays on flights flying out of Austin during these hours. Again, we will look at absolute departure delay and not the normalized departure delay as any delay, irrespective of the flight-time upon boarding, is a cause of concern when flying out of Austin at the end of the day.

Endeavor Air (9E) has the highest average delay of about 4 and a half hours per flight when flying out of Austin in late evening. This is followed by Delta Air Lines (DL). These carriers could be avoided when flying out of Austin at the end of the day.



d) Which flights to avoid if flying into Austin for work in the morning?

Many professionals that work in Austin fly into the city during the morning hours. Any cancellations or delays on these flights would be detrimental to the objectives of these flyers. Thus, we look at the data for cancellations and delays on flights arriving into Austin between 8:00 AM and 10:59 AM.



PSA Airlines (OH) has the highest average delay of 8 minutes on flights arriving into Austin in the Morning. This is followed by Envoy Air (MQ).

No flights flying into Austin in the Morning hours were cancelled in 2008. Thus, based on delays alone, professionals can avoid flying PSA Airlines and Envoy Air when flying into Austin in the Morning hours. However, the delays here are not as significant as those seen during other hours of the day.