

2008 Arrival Delay Analysis

American Airlines vs. Everyone Else

Brandon Croarkin, Michelle Mak, TS Yeap
IST 687 - Applied Data Science
June 27, 2018

[introduction/problem statement]

In 2008, American Airlines experienced a total of **172,197 arrival delays**, which was the second highest number of delays among its **competing class of 20 airlines**.

We used data analysis techniques to:

- Find where the biggest pain points were for American Airlines
- Compare findings to the industry average.

We ultimately wanted to **minimize future arrival delays** by using the data to predict the probability of experiencing future delays.



[essential research questions]

Type 1 Questions: Initial Discoveries

- Which airline experiences the most delays?
- What is the average length of a delay?
- Which airline experiences the longest delays?
- When do delays occur most frequently?
- Which airport has the most delays?

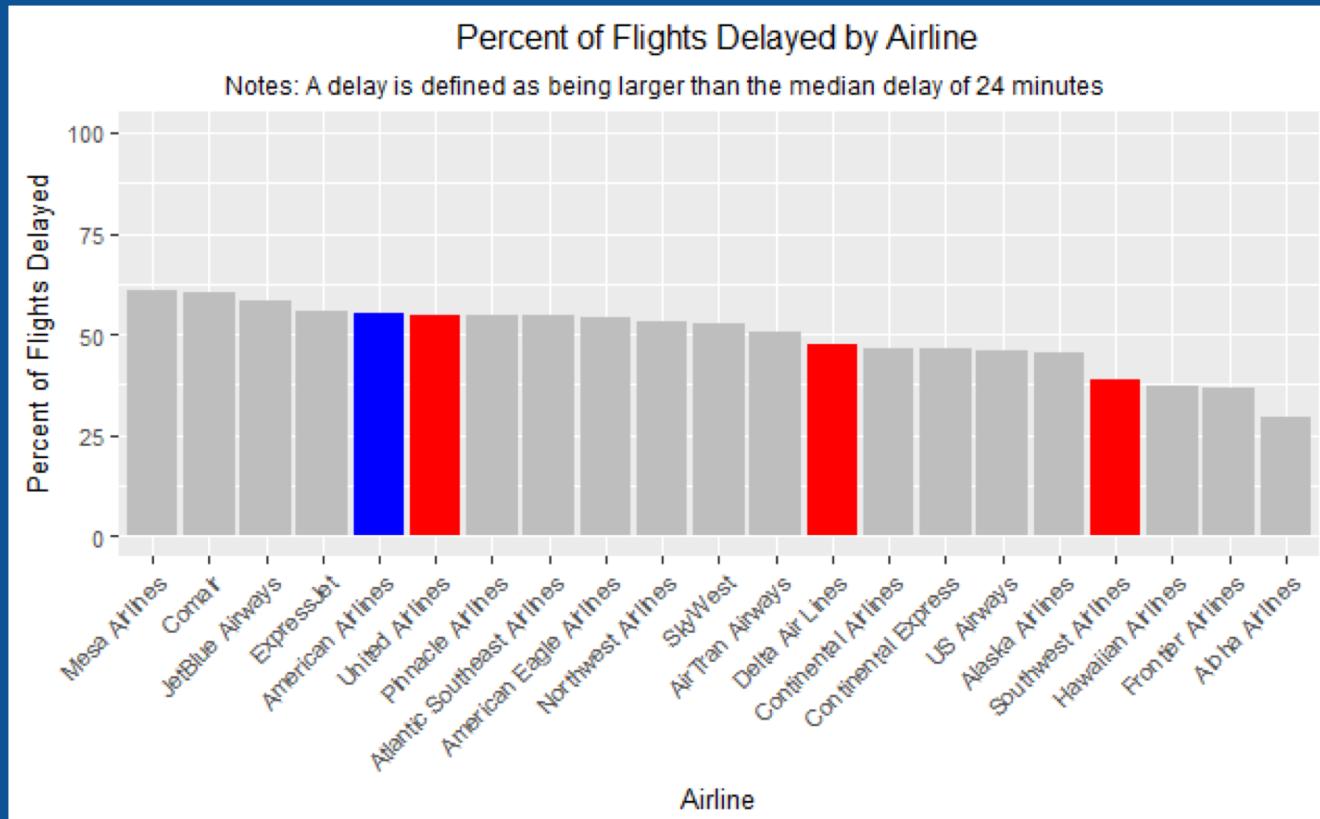
Type 2 Questions: Exploratory Analysis

- What causes airline delays?
- Can we predict the length of a delay?



[initial findings]

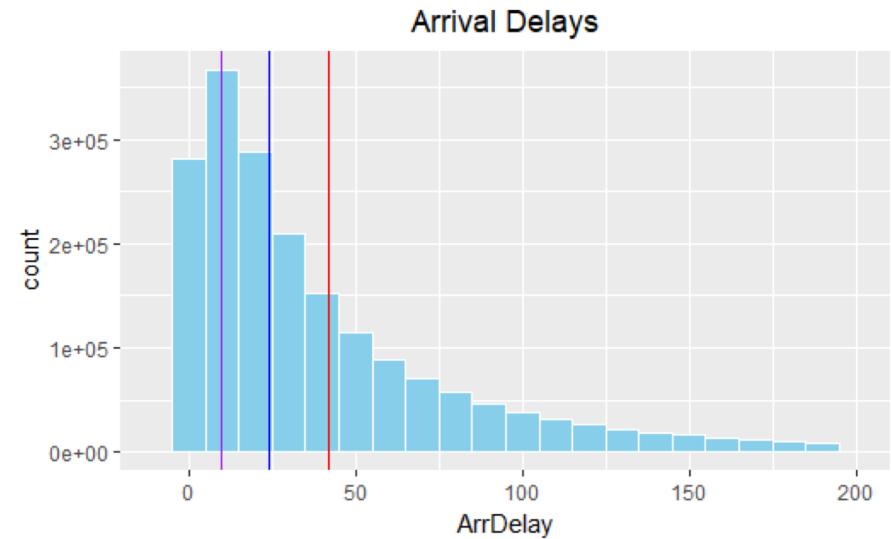
[which airline experiences the most delays?]



[average length of a delay]

AMERICAN AIRLINES

Mean: 42.2 minutes
Median: 29 minutes
Mode: 10 minutes

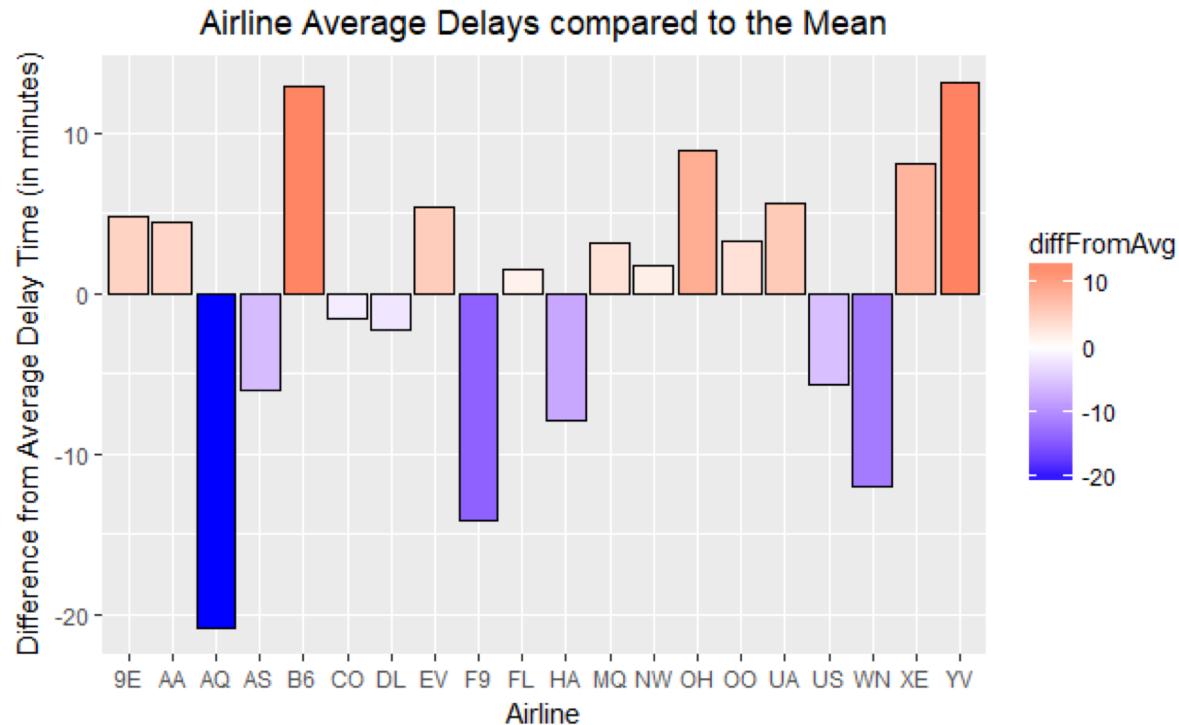


INDUSTRY AVERAGE

Mean: 42.8 minutes
Median: 24 minutes
Mode: 10 minutes

[which airline experience the longest delays?]

- American Airlines (AA) has above average delays
- Airlines that experience the LONGEST delays
 - Mesa Airlines (YV)
 - JetBlue (B6)



[when do delays occur
most frequently?]

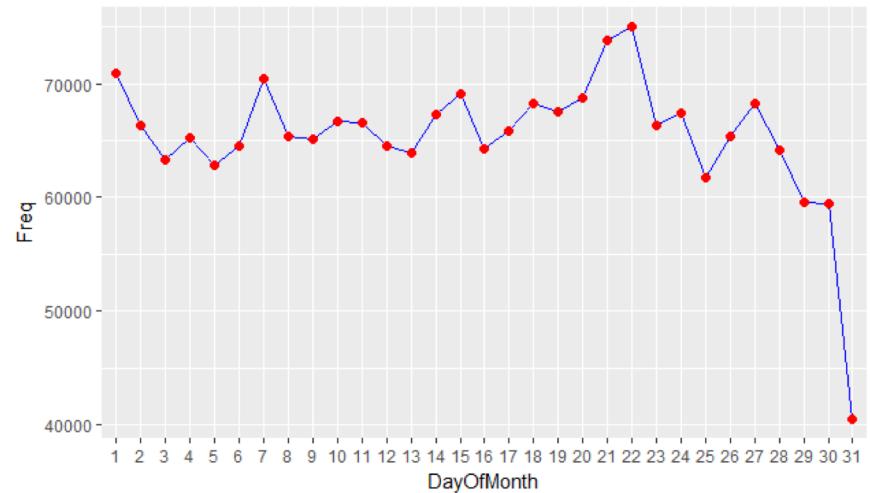
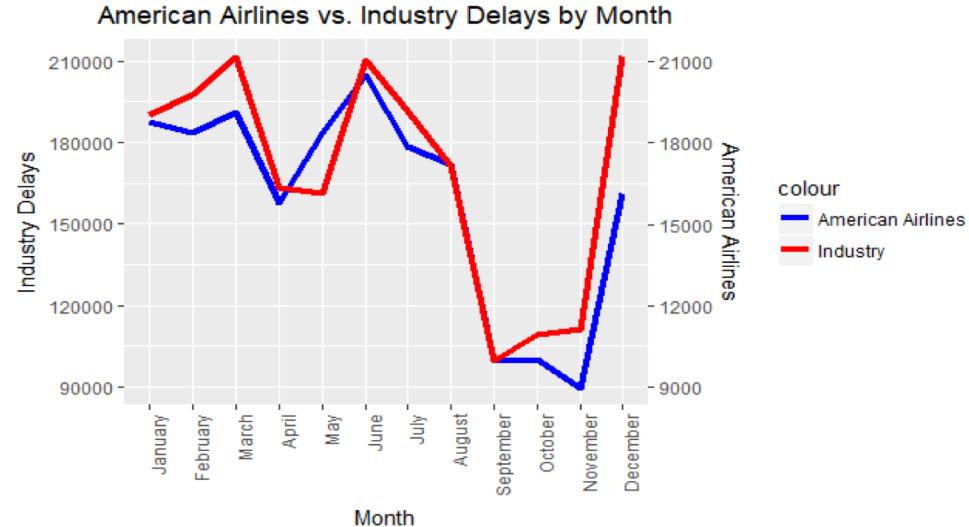
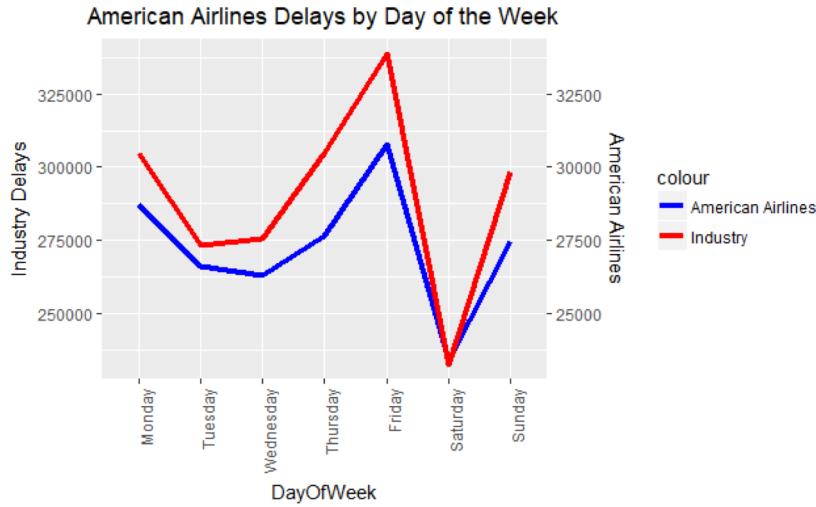
By day of week:

- American Airlines: Friday
- Industry Average: Friday

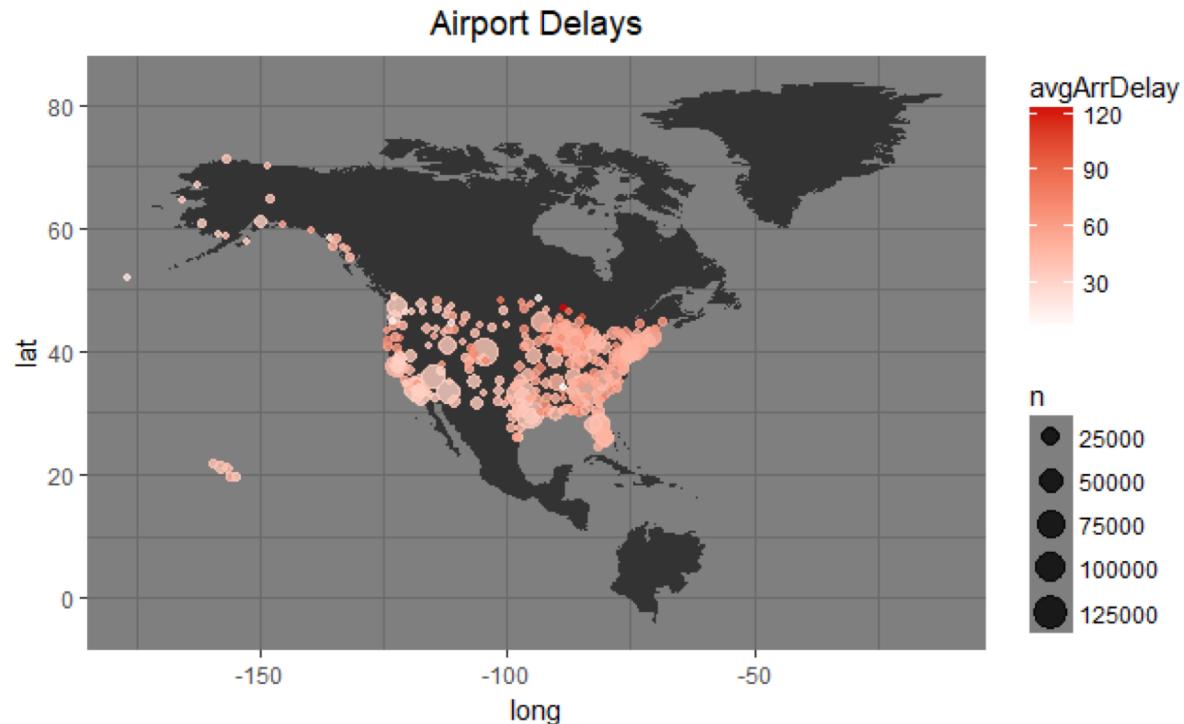
By month of year:

- American Airlines: June
- Industry Average: December

[when do delays occur most frequently?]



[which airport experiences the most delays?]



Arrival delays map by origin airport

Industry Average

Top 5 Worst & Best Airports

Worst Airports:

1. CMX (Hancock, MI)
2. PLN (Pellston, MI)
3. SPI (Springfield, IL)
4. MQT (Marquette, MI)
5. ALO (Waterloo, IO)

Best Airports:

1. TUP (Tupelo, MS)
2. INL (International Falls, MN)
3. SLE (Salem, OR)
4. WYS (West Yellowstone, MT)
5. ADK (Adak Island, AK)

American Airlines

Top 5 Worst & Best Airports

Worst Airports:

1. GUC (Gunnison, CO)
2. HDN (Hayden, CO)
3. EGE (Vail, CO)
4. KOA (Kailua, HI)
5. JAC (Jackson, WY)

Best Airports:

1. OAK (Oakland, CA)
2. FAT (Fresno, CA)
3. BWI (Baltimore, MD)
4. CMH (Columbus, OH)
5. BUR (Burbank, CA)



[exploratory analysis]

[what causes delays?]

The data set came with five predetermined causes:

- Carrier delay
- Weather delay
- National Air Security delay
- Security issue delay
- Late arrivals delay

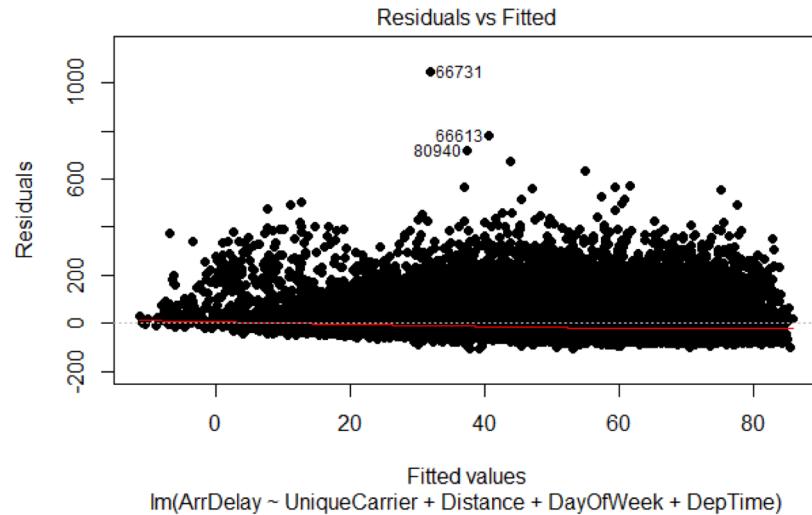
Cause	Avg. Delay (min)
Carrier (Industry)	19.18
Carrier (AA)	21.34
Weather (Industry)	3.70
Weather (AA)	3.17
NAS (Industry)	15.02
NAS (AA)	16
Security (Industry)	0.09
Security (AA)	0.05
Late Arr (Industry)	25.3
Late Arr (AA)	25.21

[can we predict delays?]

Based on our initial findings, we used these five variables in our multiple linear regression model to try to predict delays:

- Unique Carrier
- Departure Hour
- Month
- Distance
- Weekend

We were unable to achieve a great R-squared in this model. An analysis of the residuals gives us some insight on why this is. We have such a large spread in residuals. It appears that there is a lot of inherent error in the model that makes predicting airline delays hard.



[takeaways]

IN CONCLUSION

We ultimately cannot predict delays or length of delays -- too many outlying factors

BUT WE DID LEARN

- Southwest Airlines is the bee's knees
 - Most flights and (one of the) least delays
- Delays happen a lot on Friday
 - So plan your long weekends accordingly!
- Factor in **extra** time when traveling in December

Q & A

Easy questions only please

THE END

Thanks for flying with American Airlines!