

Will my flight arrive on-time at Seatac?

Dana Lindquist

October 31, 2018



Data

Flights arriving at Seatac

- Aug 2016 to July 2018 (2 years)
- 274,303 flights
- 19% of flights delayed

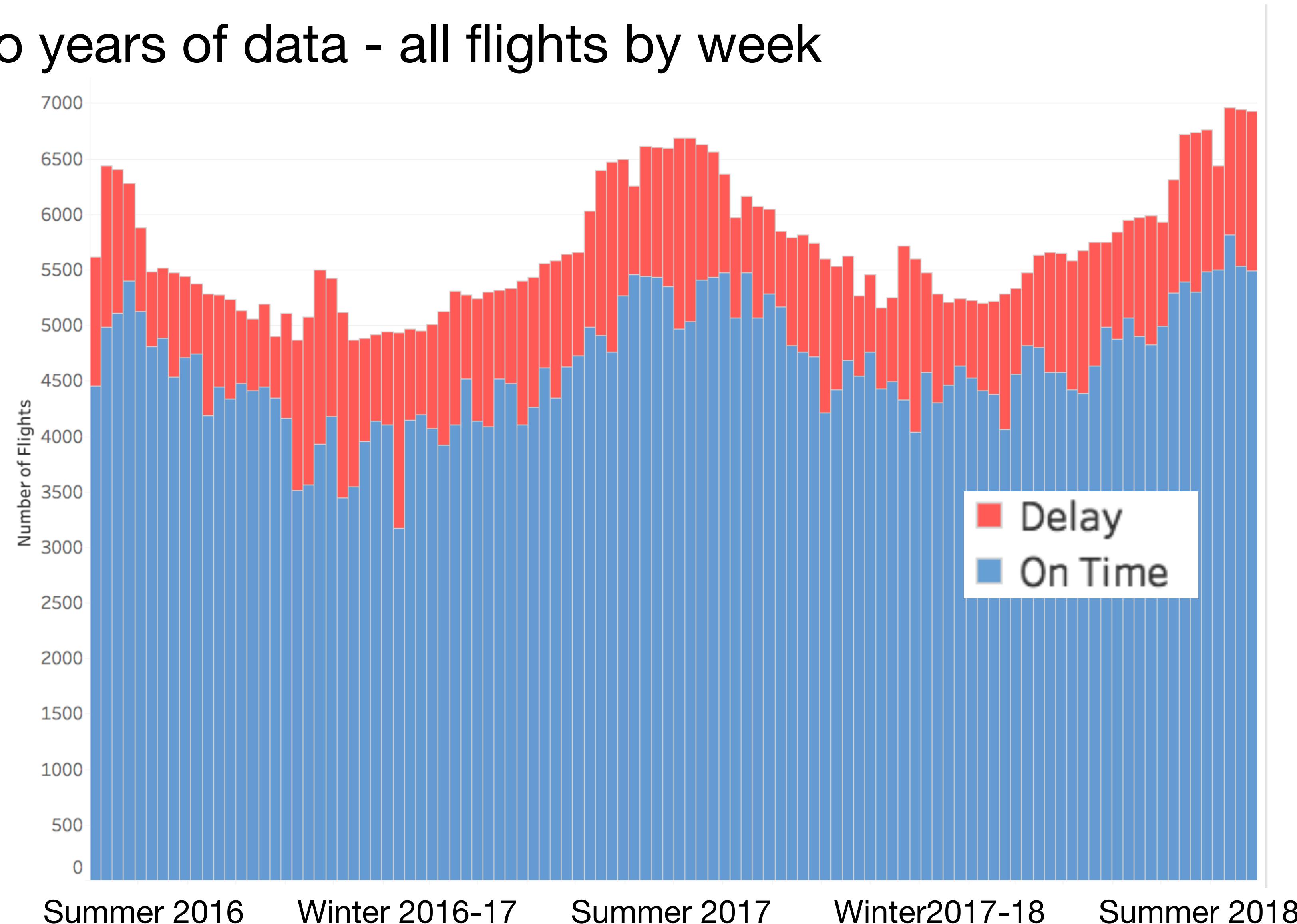
Weather from

- Seattle
- Chicago
- New York City

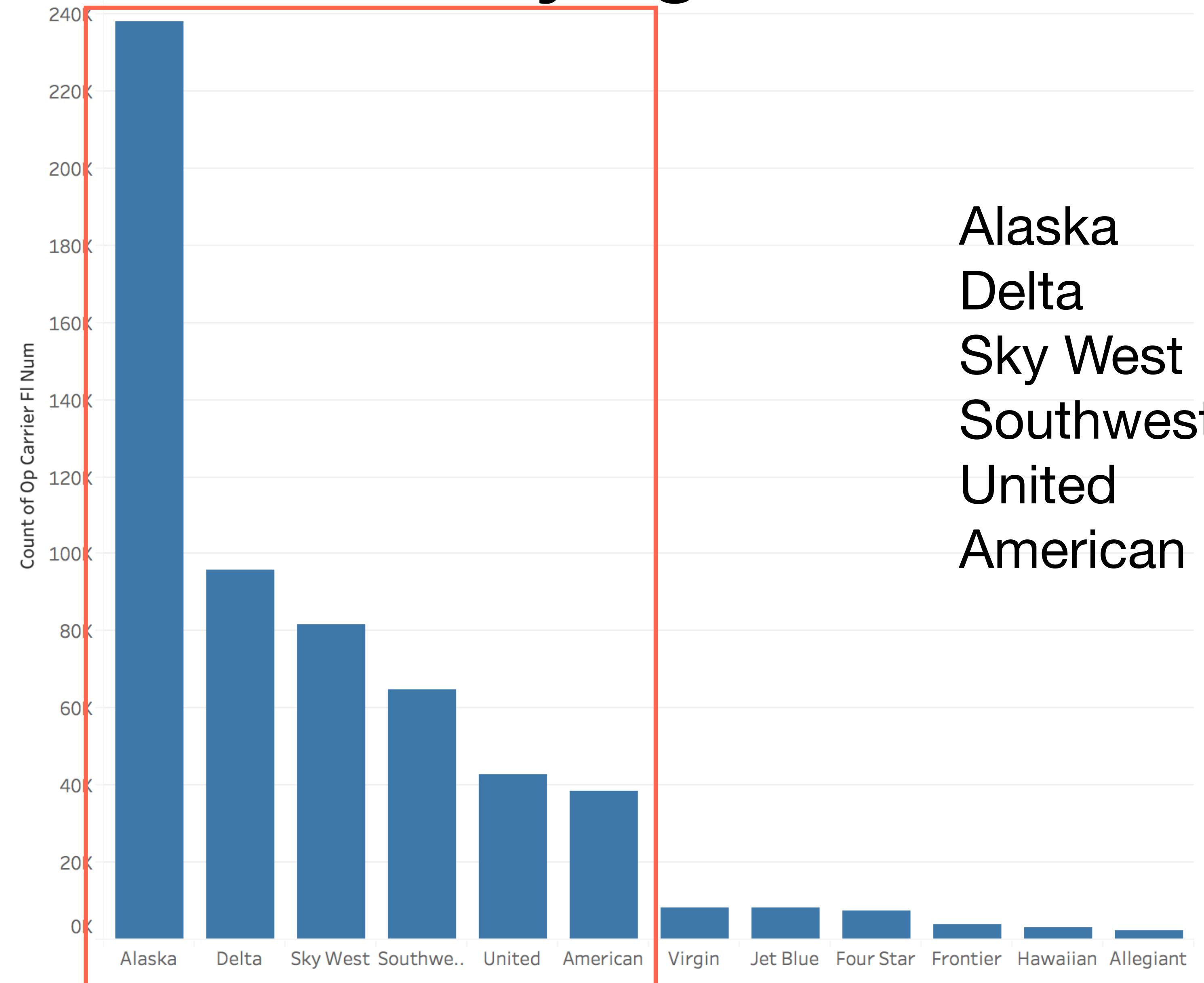


What do flight delays look like?

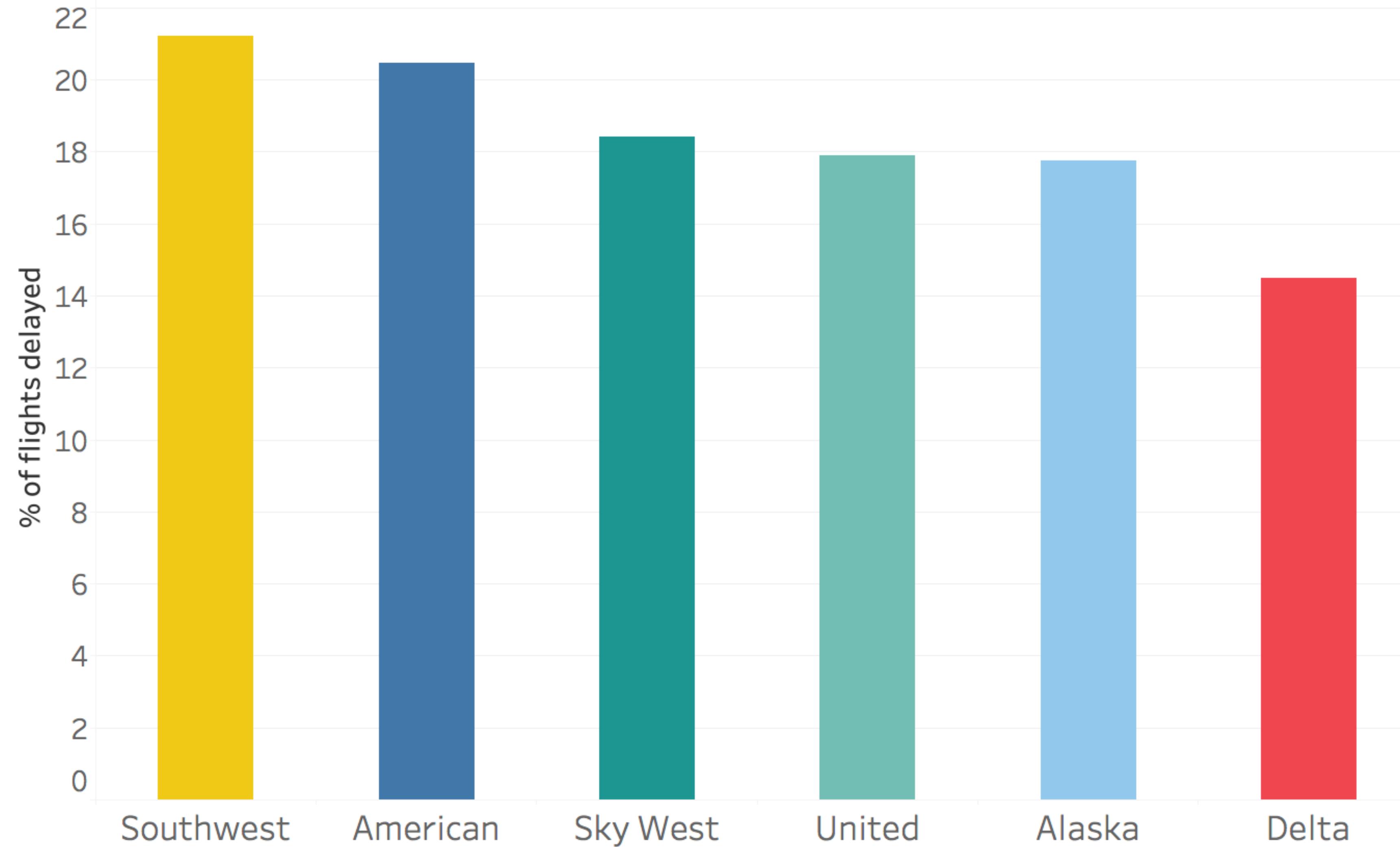
Two years of data - all flights by week



Which airlines are flying to Seatac?

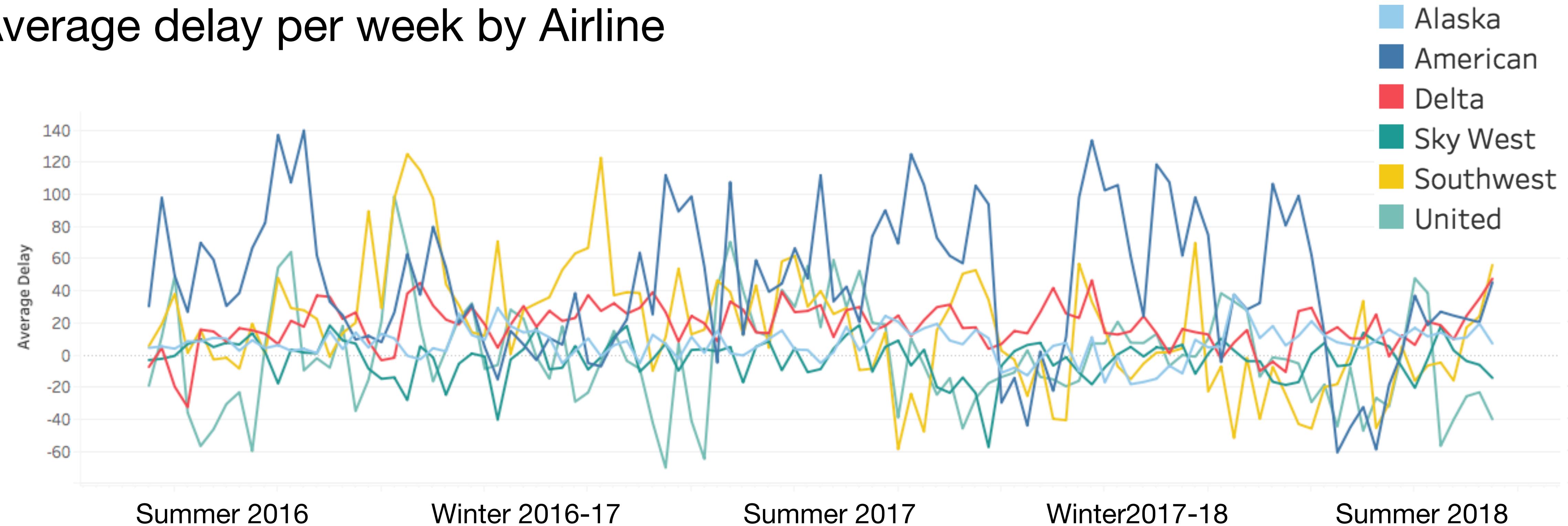


Is one airline worse than others?

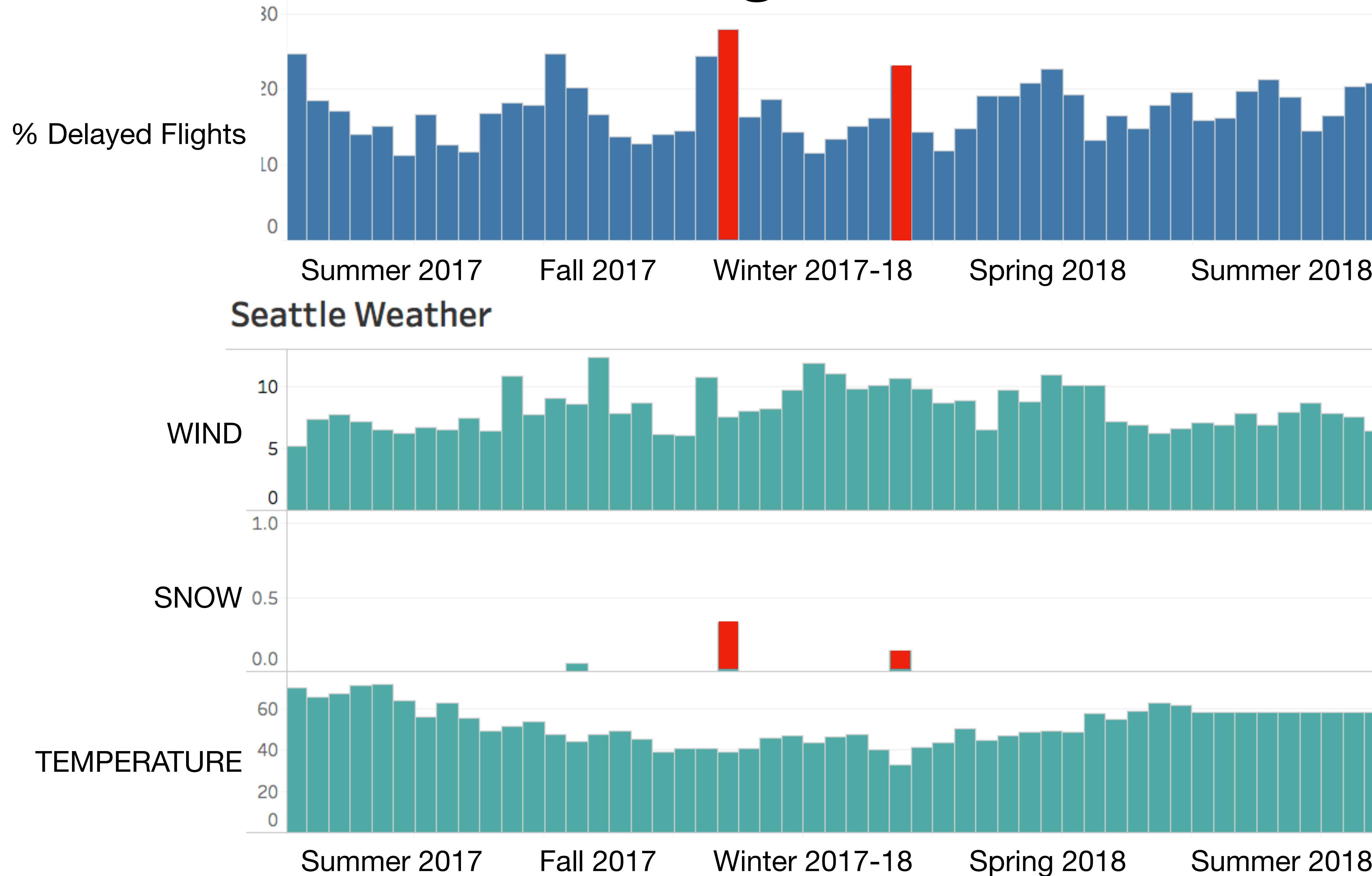


Is one airline worse than others?

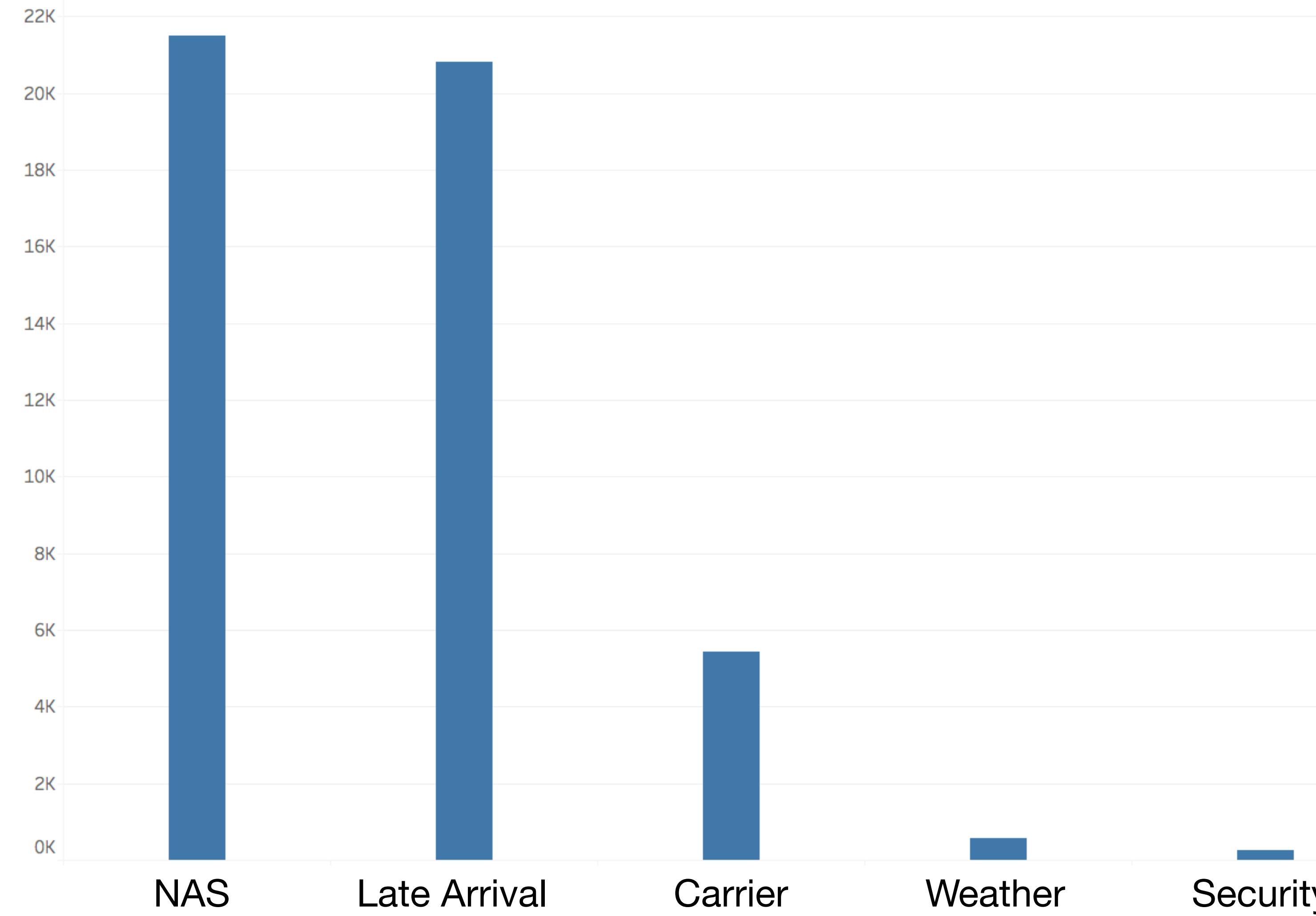
Average delay per week by Airline



Does weather effect flight on-time?



Why do flights arrive late?



You are traveling tomorrow.

You have

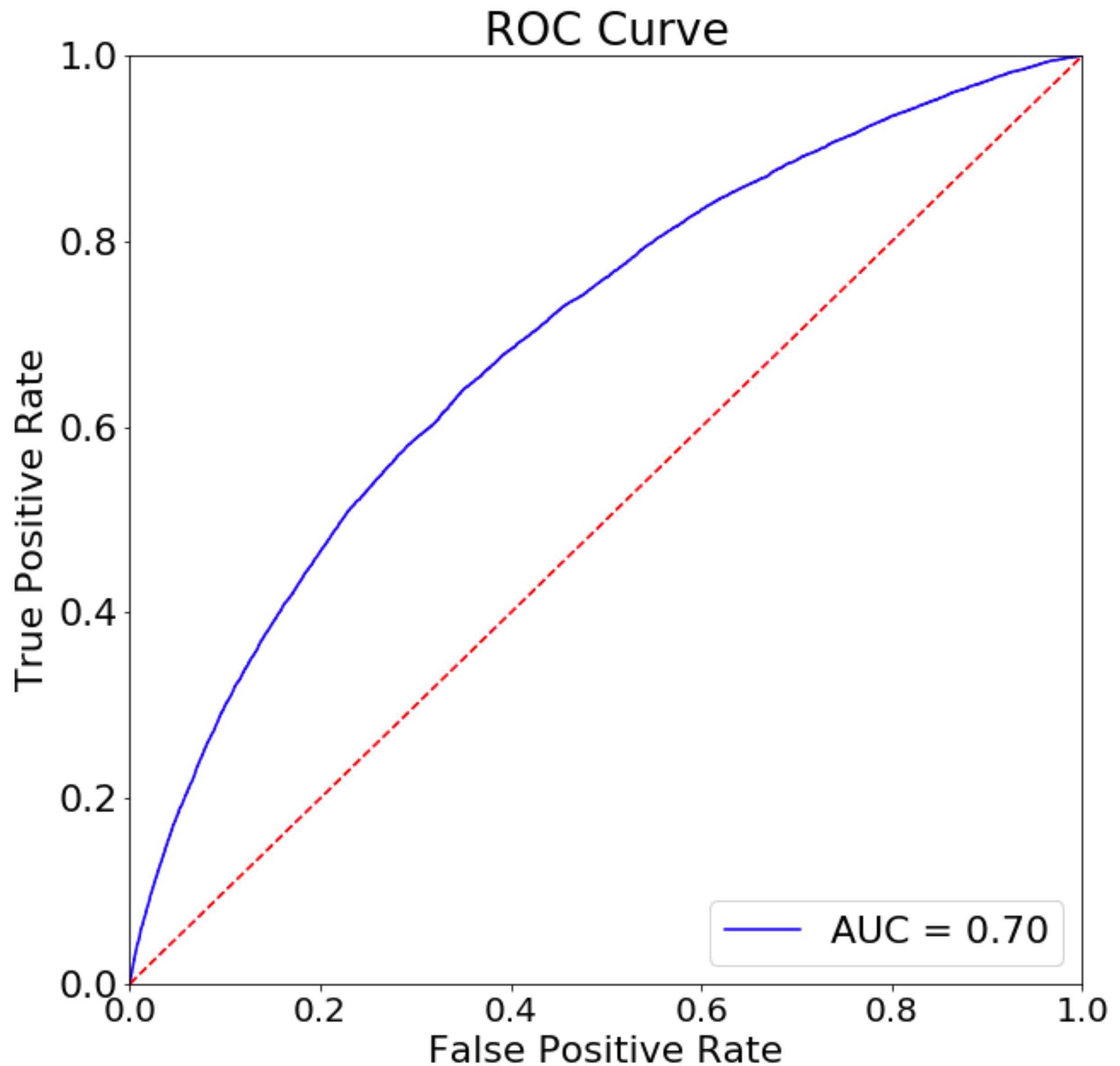
- Flight number
- Airline
- Departure time
- Weather forecast

What's the chance your flight will be on-time?



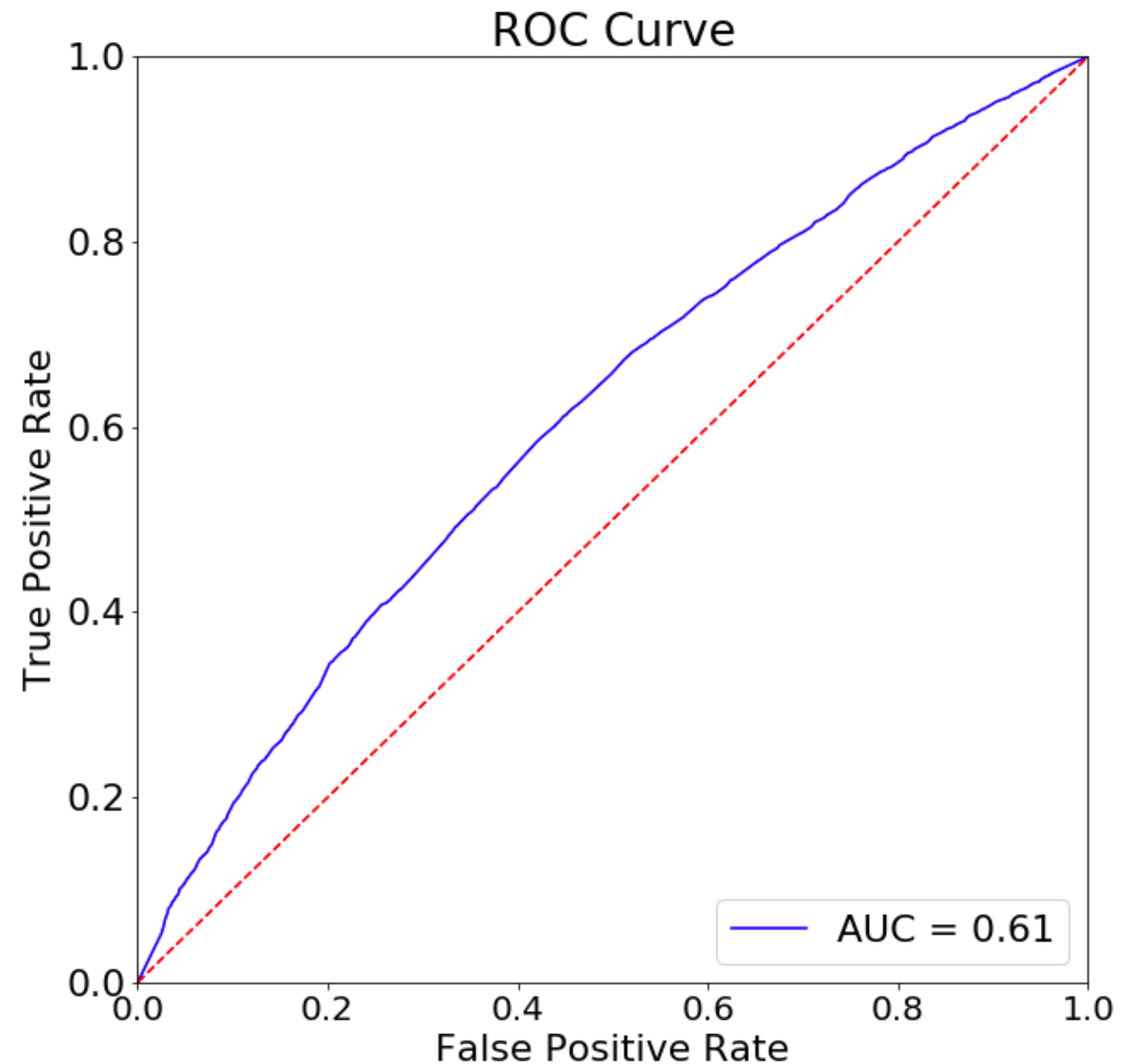
Random Forest

Accuracy	0.65
Precision	0.31
Recall	0.64
AUC	0.70



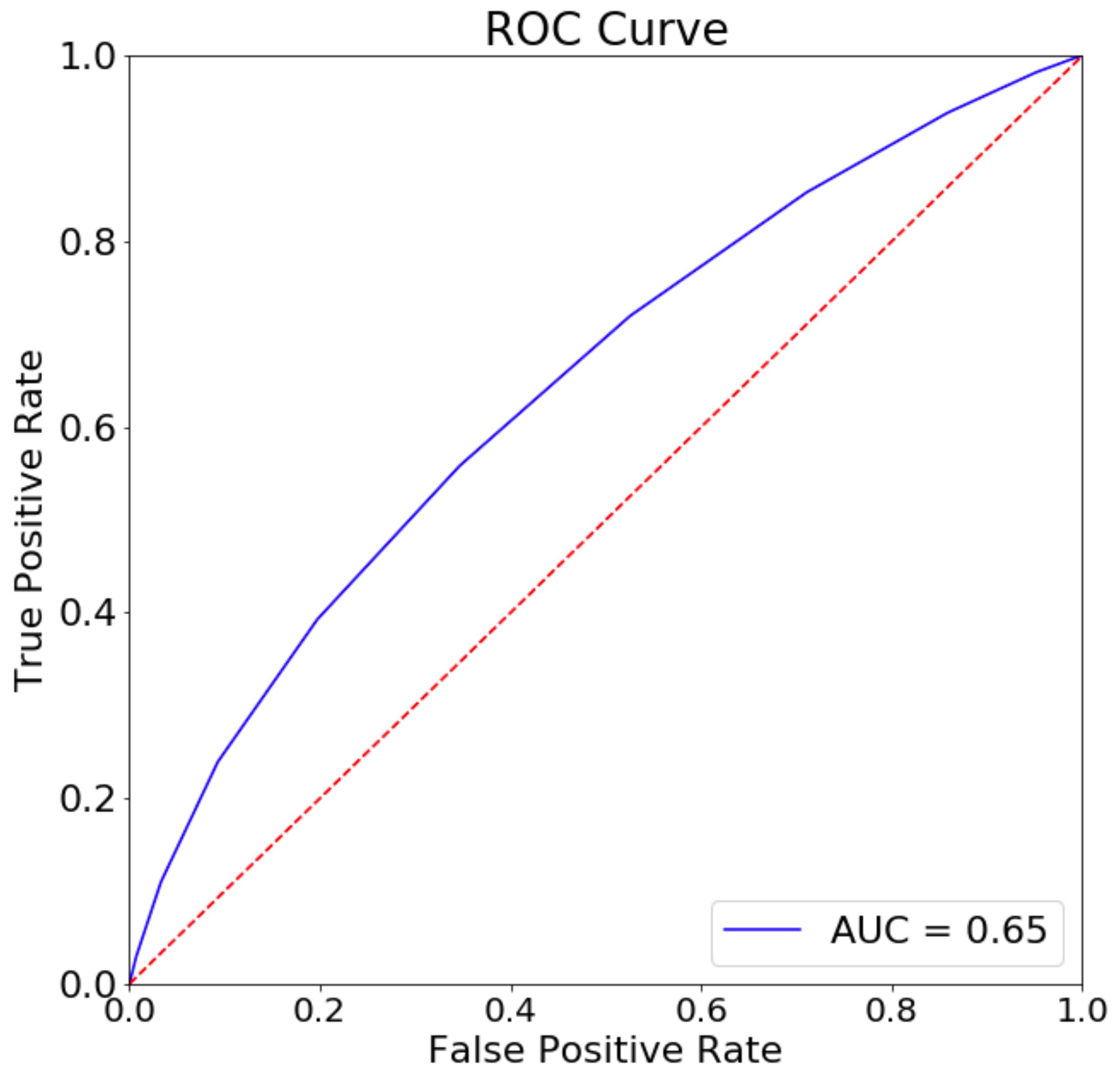
Decision Tree

Accuracy	0.58
Precision	0.25
Recall	0.59
AUC	0.61



KNN

Accuracy	0.63
Precision	0.28
Recall	0.56
AUC	0.65



Future Work

- Try a different city, maybe Chicago
- Add a class called “Very Delayed” for flights over 1 hour delayed
- Create a Flask app so you all can use my model the next time you fly

A photograph of a woman with long brown hair, smiling and looking towards the right. She is wearing a white sleeveless top, blue jeans, and a bright pink scarf with fringe. She is pulling a bright yellow suitcase with a black handle. A small orange tag is visible on the suitcase. She is also holding a small white object, possibly a ticket or a card.

Wishing you safe, timely travels

Thank you

Dana Lindquist

danalindquist@silverbeach.com

www.linkedin.com/in/danalindquist/

Appendix

Model

Features:

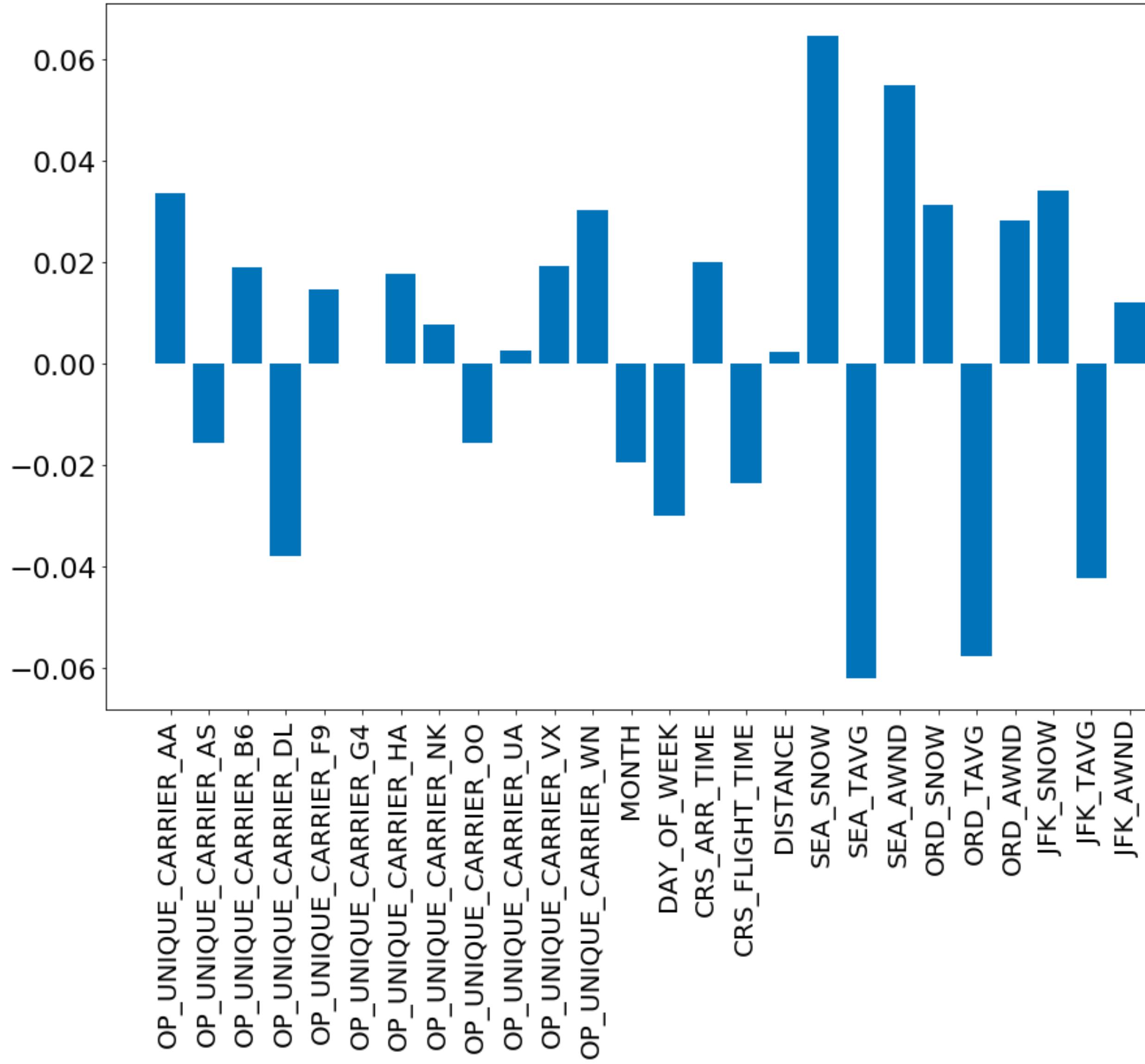
- Airline & Flight number
- Distance & planned flight duration
- Arrival
 - Time
 - Day of Week
 - Month
- Weather in Seattle, Chicago, NYC
 - Wind
 - Snow
 - Temperature

Trained on downsampled majority class (on-time) to match delays 50/50

Results from different models

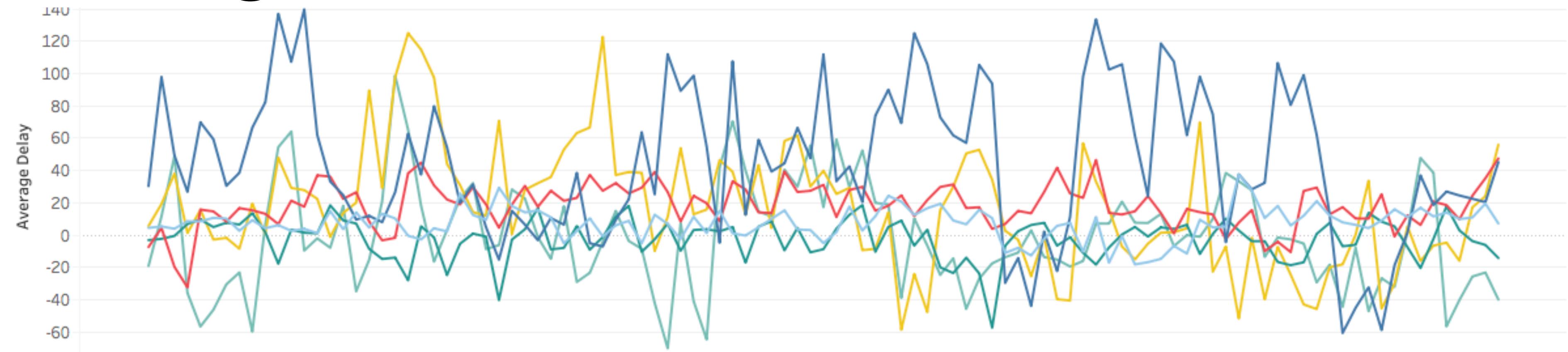
Model	Accuracy	Precision	Recall	F1	AUC	Notes
KNN	0.63	0.28	0.56	0.37	0.65	n_neighbors = 10
SVM	0.54	0.21	0.52	0.30	0.54	kernel='linear' C=1
Logistic Regression	0.56	0.23	0.56	0.33	0.58	C=1 penalty = L2
Decision Tree	0.58	0.25	0.59	0.35	0.61	max_depth = 14 min_samples_leaf = 6
Random Forest	0.65	0.31	0.64	0.41	0.70	n_estimators = 100 max_depth = 10 min_samples_leaf = 4

Correlation between Delay and Features

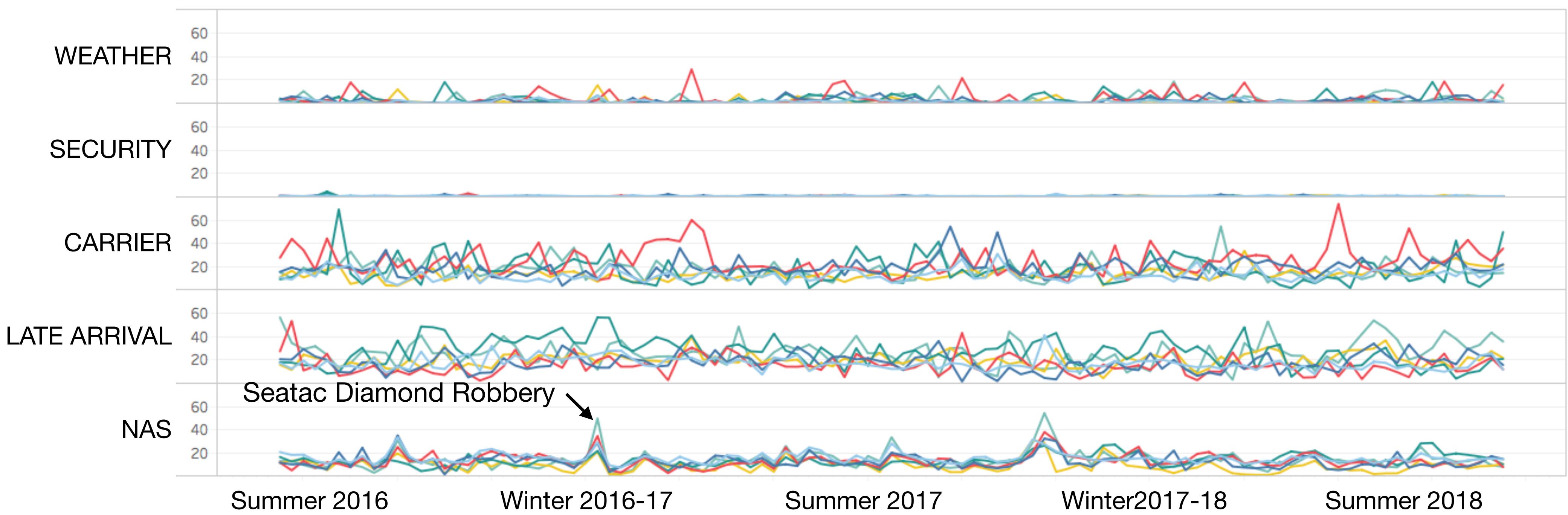


Features have been normalized
You can see correlation between delay and weather features

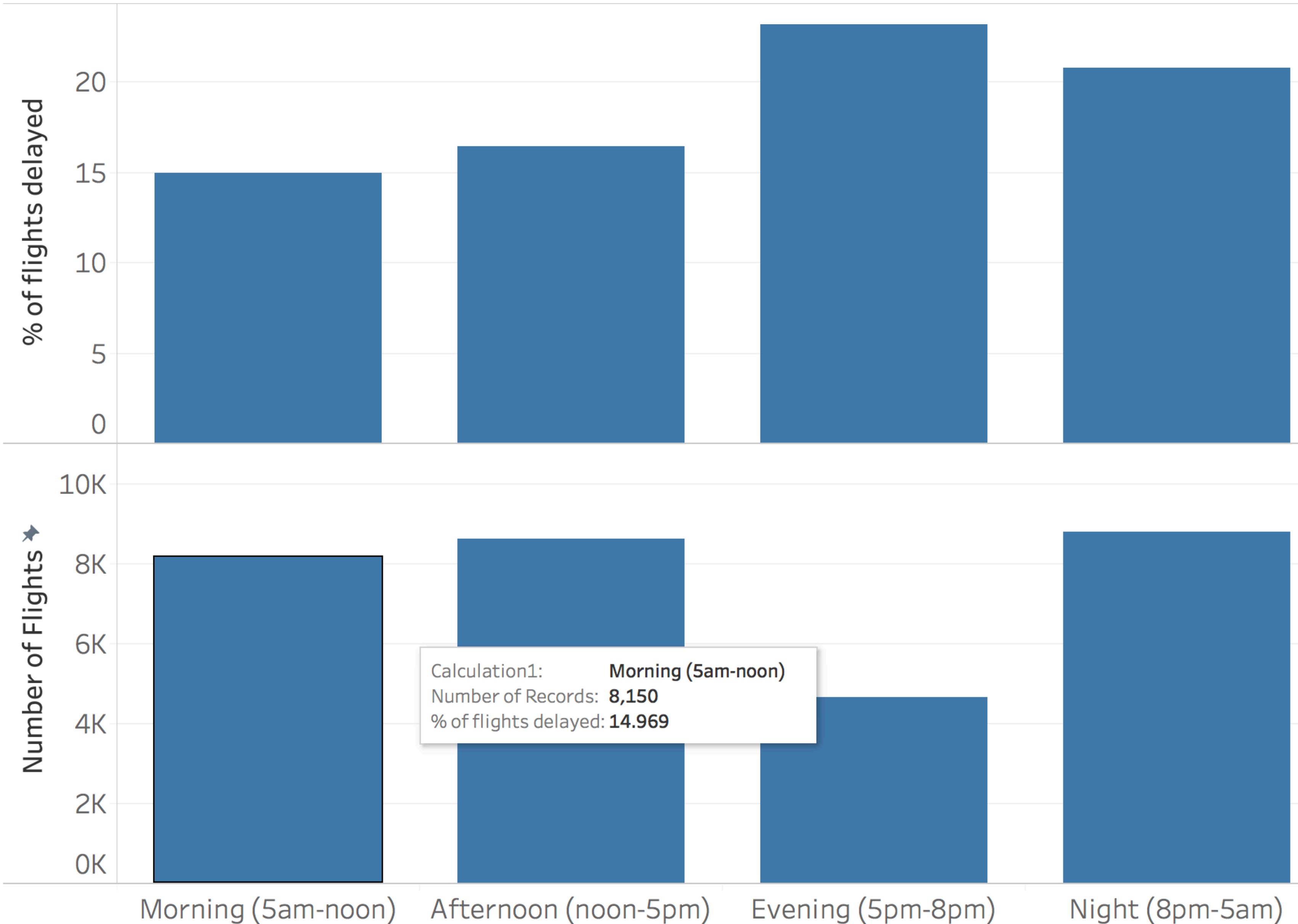
Why do flights arrive late?



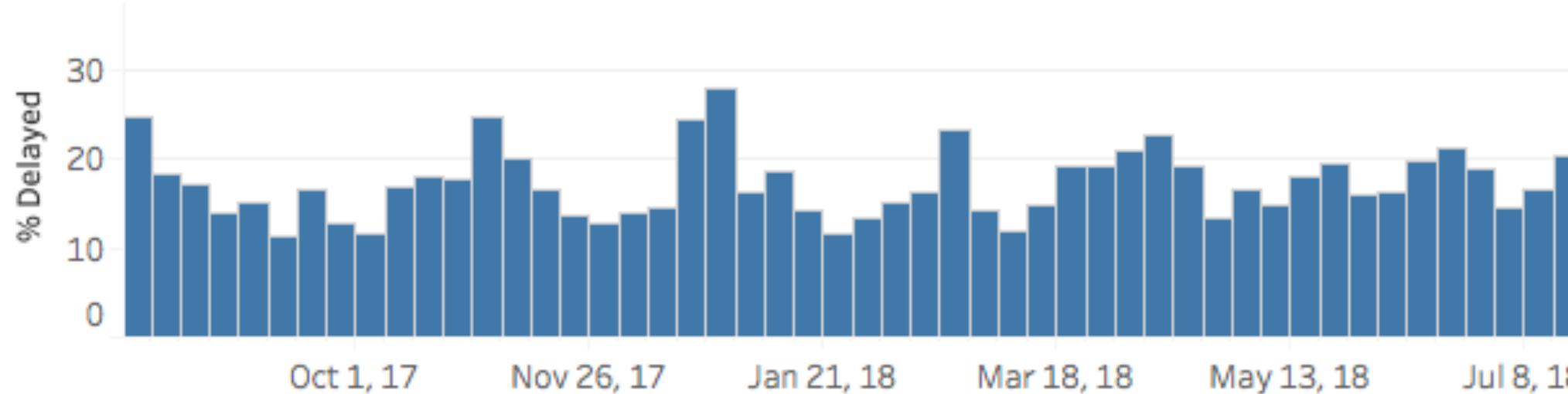
Reason for delay



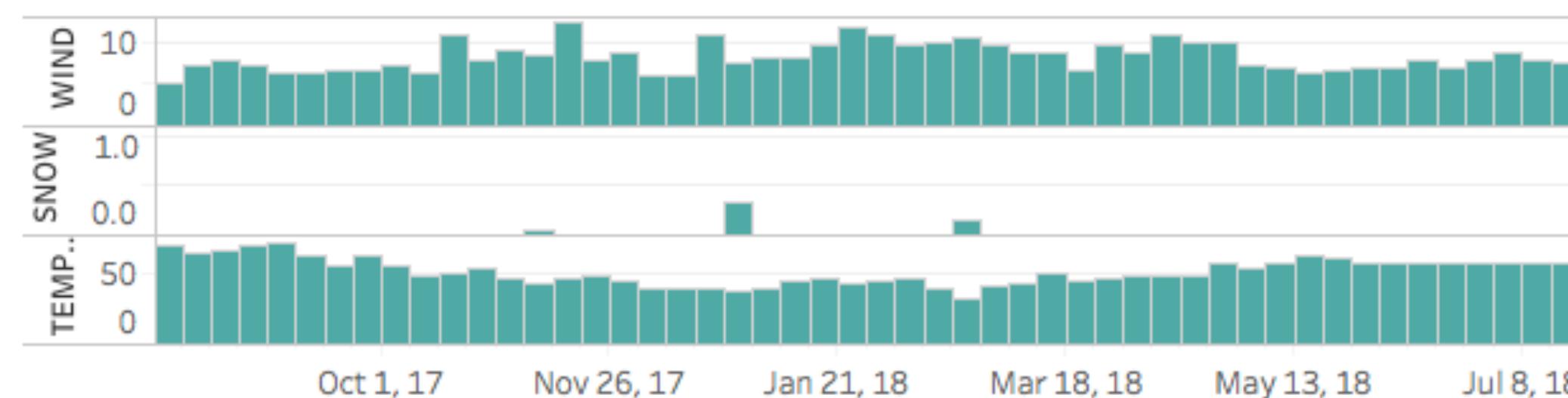
What time of day has the most delays?



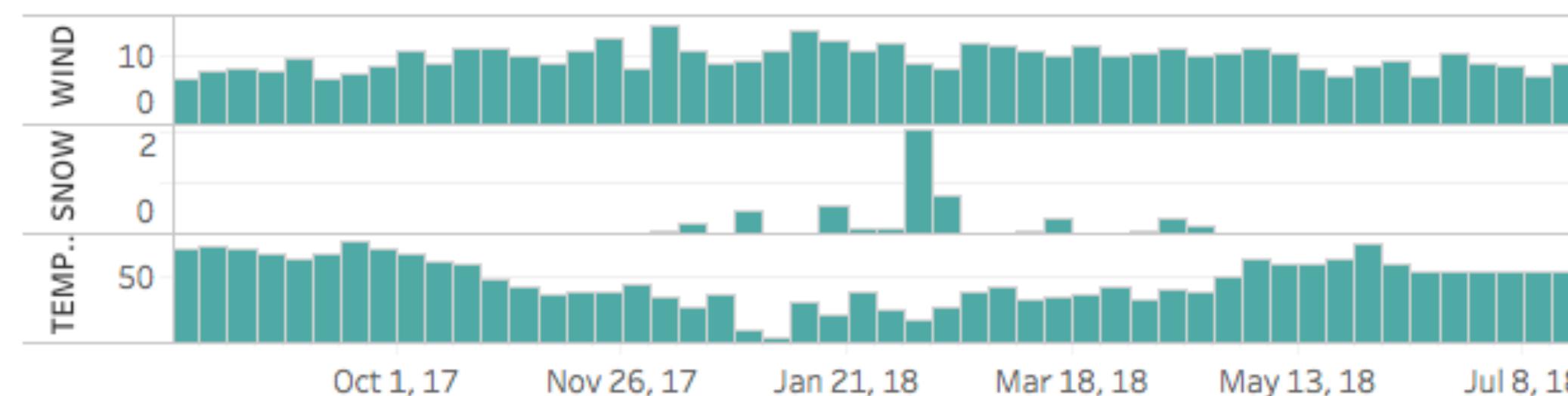
Percent of Flights Delayed



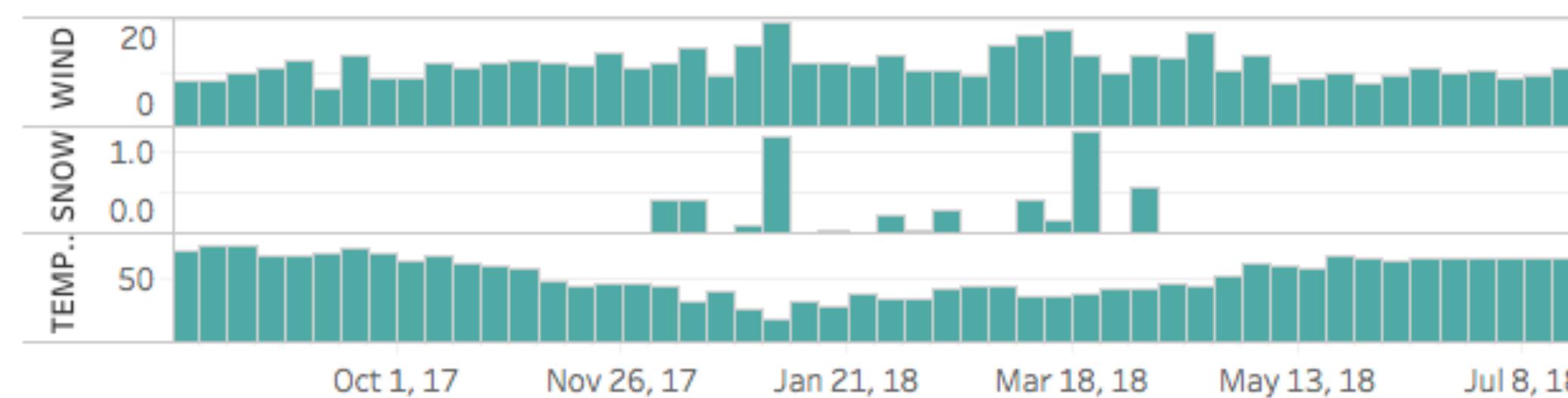
Seattle Weather



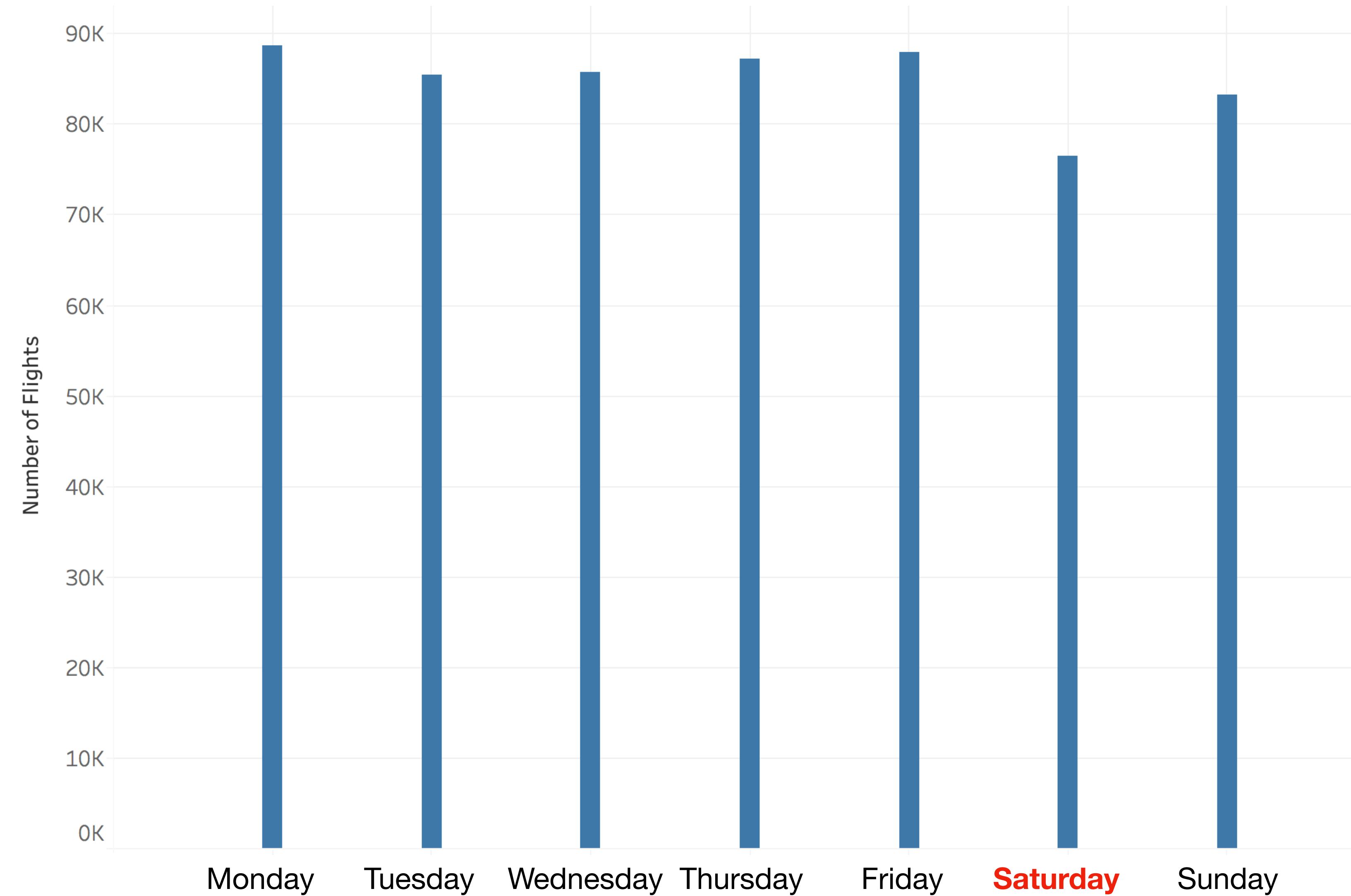
Chicago Weather



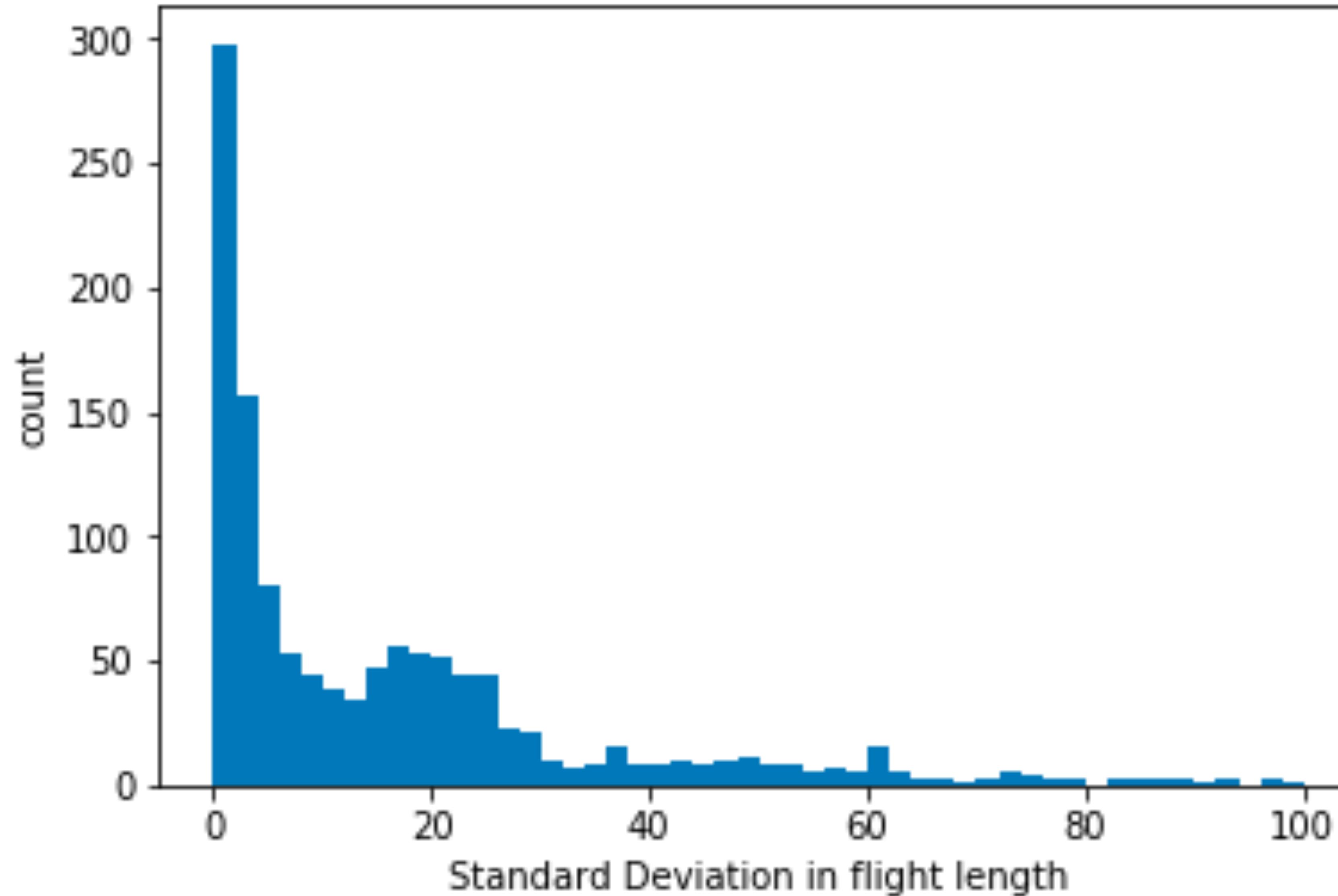
New York Weather



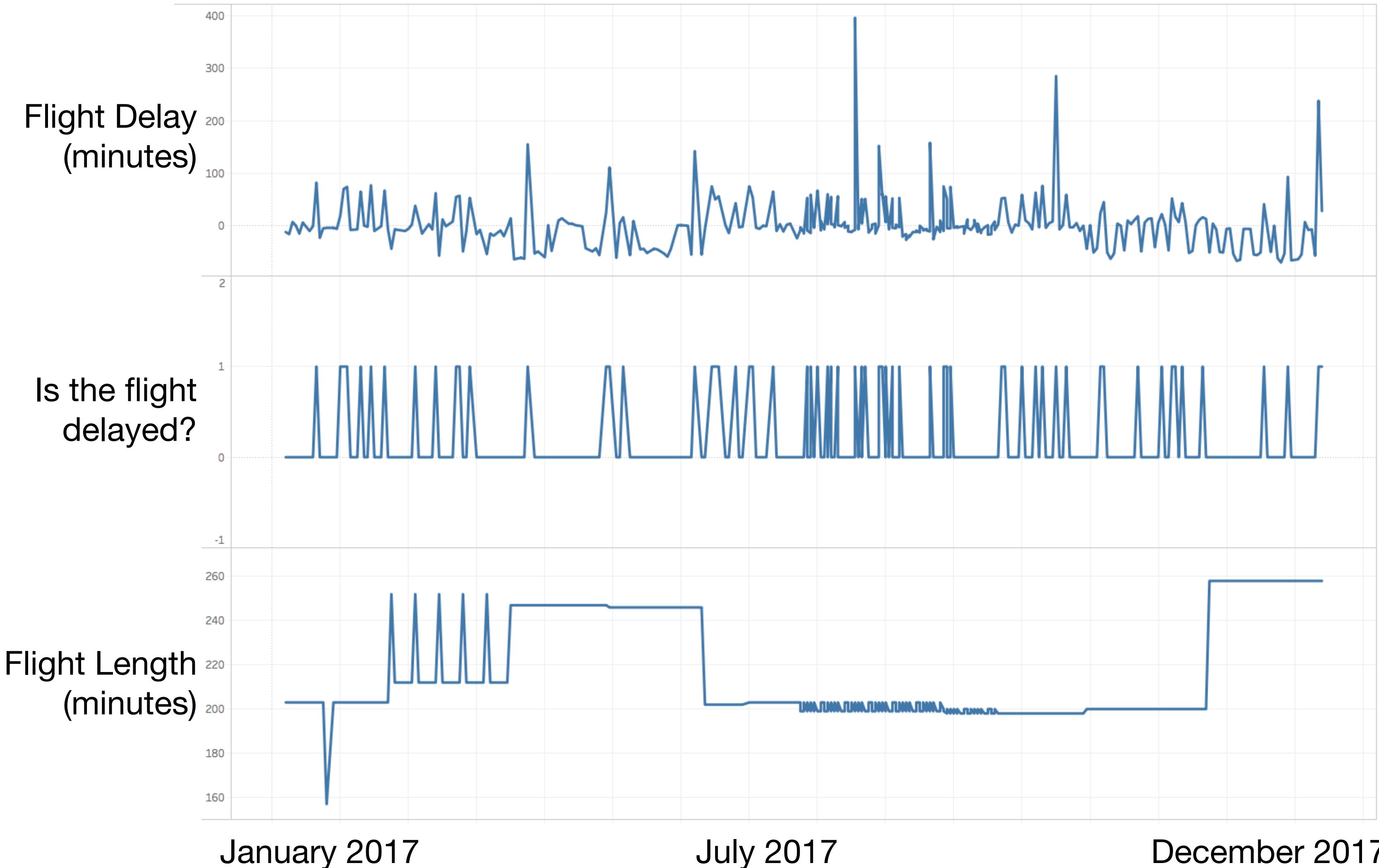
Does the number of flights vary by weekday?



Are airlines changing the flight time?



Are airlines changing the flight time?



Alaska Airlines
Flight 259
San Jose, CA to Seatac