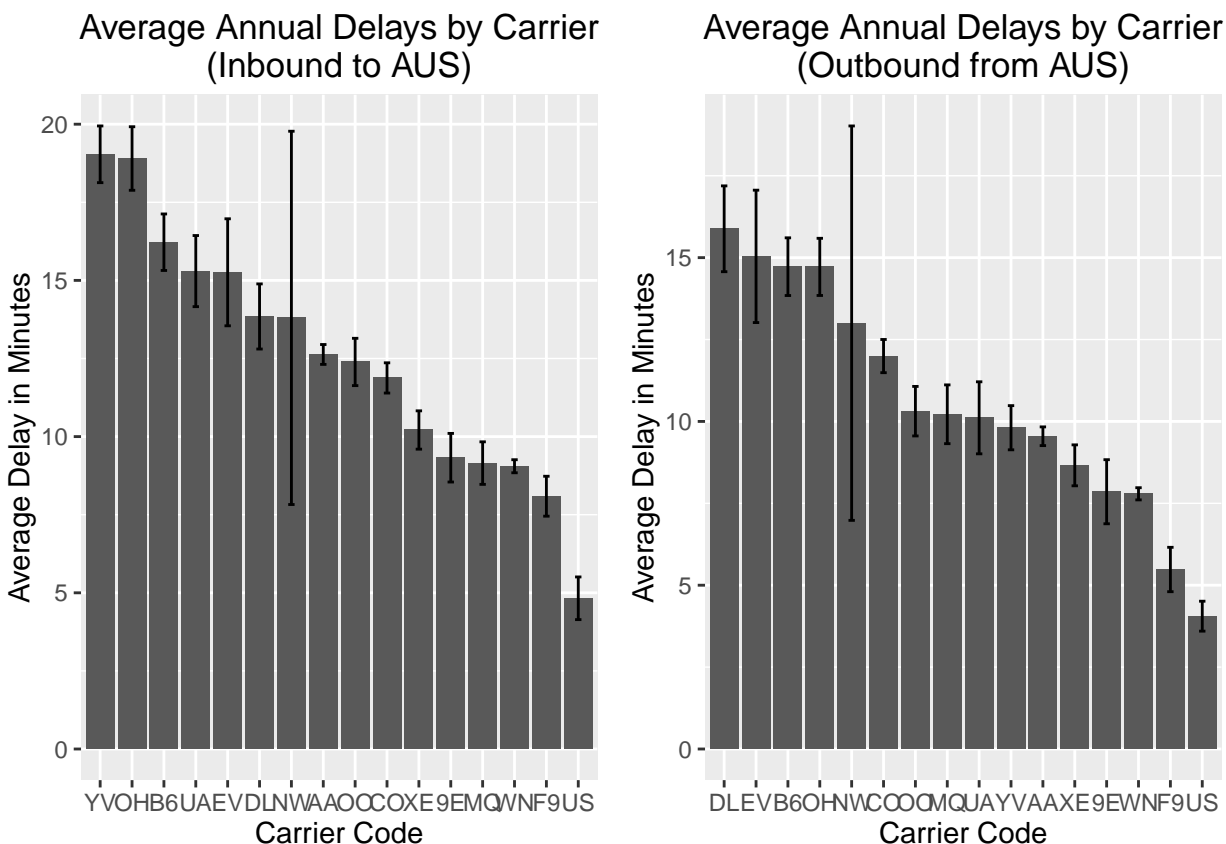


STA 380 Homework 2: Nicole Erich

Packages required: `tm`, `ggplot2`, `arules`, `reshape2`, `grid`, `e1071`, `nnet`

Question 1: Flights at ABIA

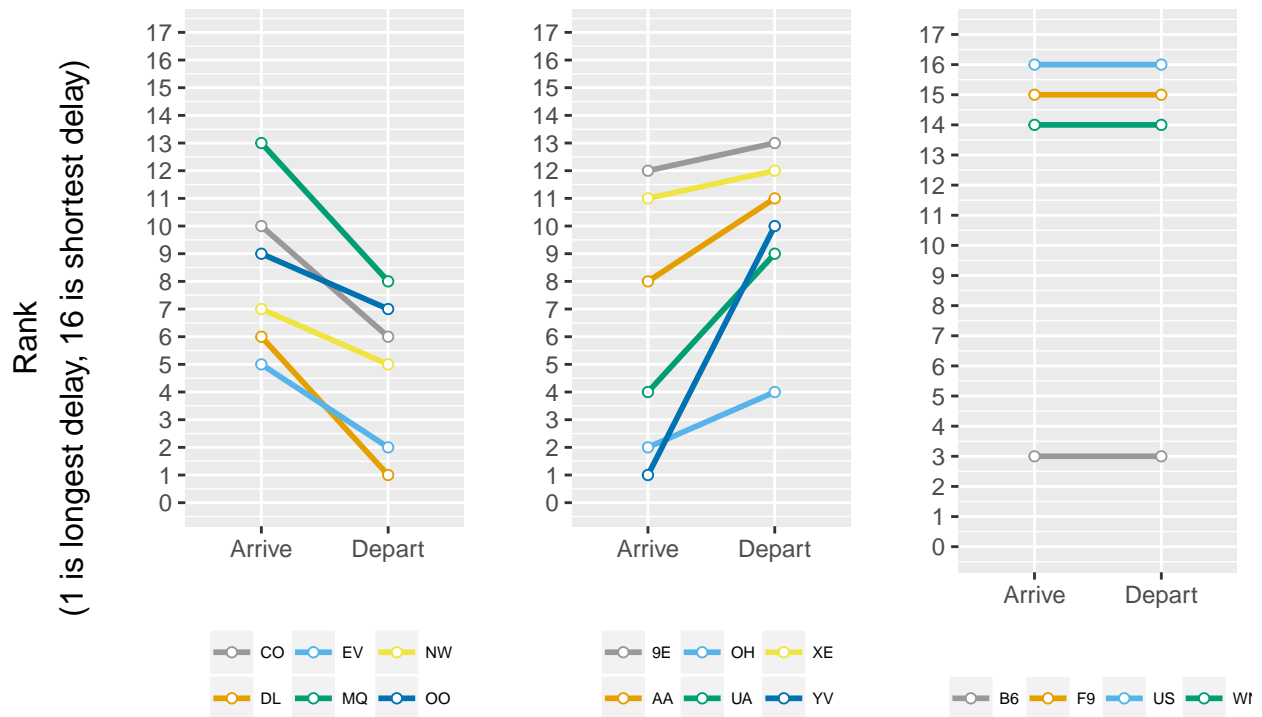


The plots above show the annual average delays per flight carrier, separated by inbound and outbound flights related to Austin, Texas. Each bar shows the average total delay in minutes (including all delay categories), and there is a standard error bar applied to each carrier.

For both graphs, the right side of the graph (carriers with lower average annual delays) seems to have a tighter error range on average, while those carriers clocking in at higher delays also have larger error ranges.

Some carriers are in different orders from arrivals to departures. How do carriers rank against one another in arrivals vs. departures?

Carrier Ranking in Average Delay Times For Arrivals and Departures

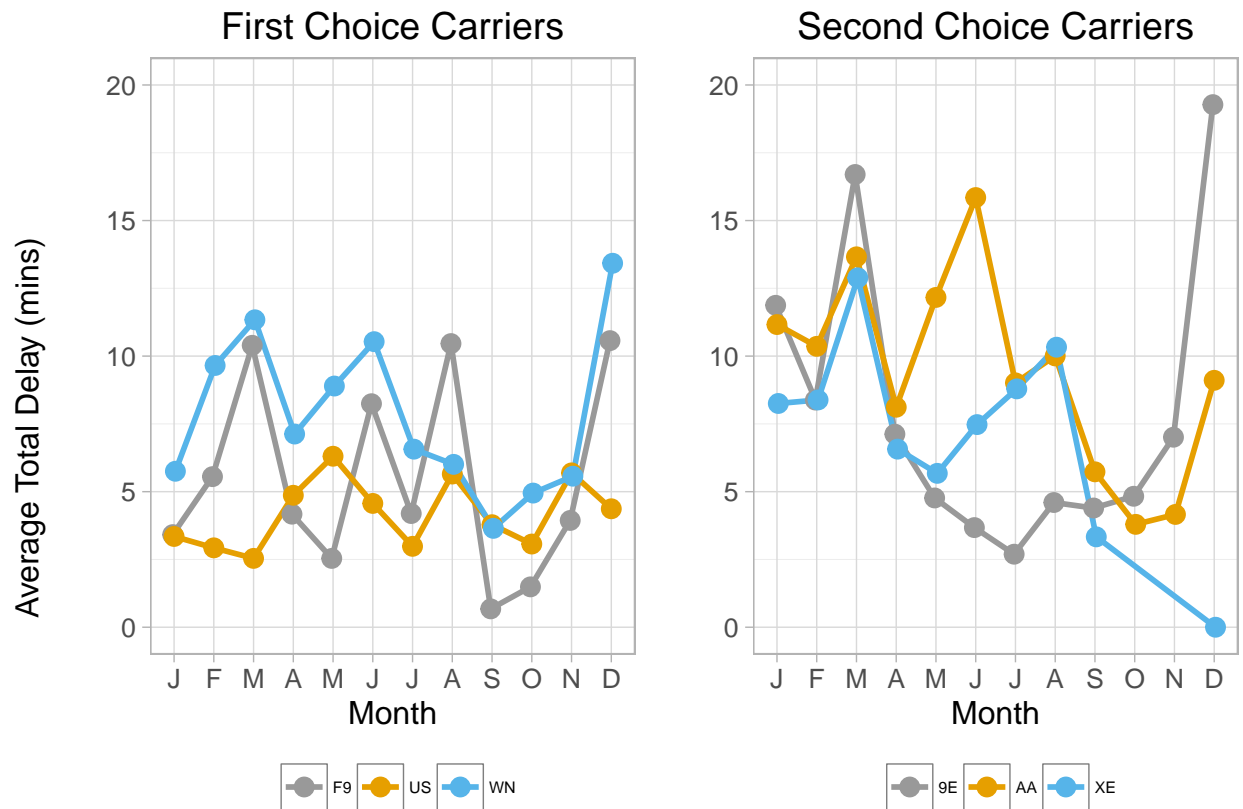


Carriers F9, US, and WN consistently have the shortest delays for both arrivals and departures. Depending on the priority, different airlines seem to fair better keeping to the schedule arriving in Austin or departing from Austin.

If a person is leaving Austin to go to an event (conference, wedding, family reunion) in another city, he or she will probably be more concerned about arriving to that destination on time, rather than the time s/he returns home in a few days.

In that case, the safest options would be the carriers US, F9, WN as a first choice. Second choice options would be 9E, XE, or AA. These three airlines will see a longer delay, on average, returning to Austin, but will be relatively good at minimizing delays on the outbound flights.

Which airline and time of the year is the safest bet to reduce delays leaving Austin? In this case, money is no object as long as a person can guarantee to the family that s/he will not be late to the reunion this time.



For consistently low average delays throughout the year, US is the clear winner. Even during the peak delay months for other airlines (March, June, and December), US manages to keep its outbound delays fairly low and stable.

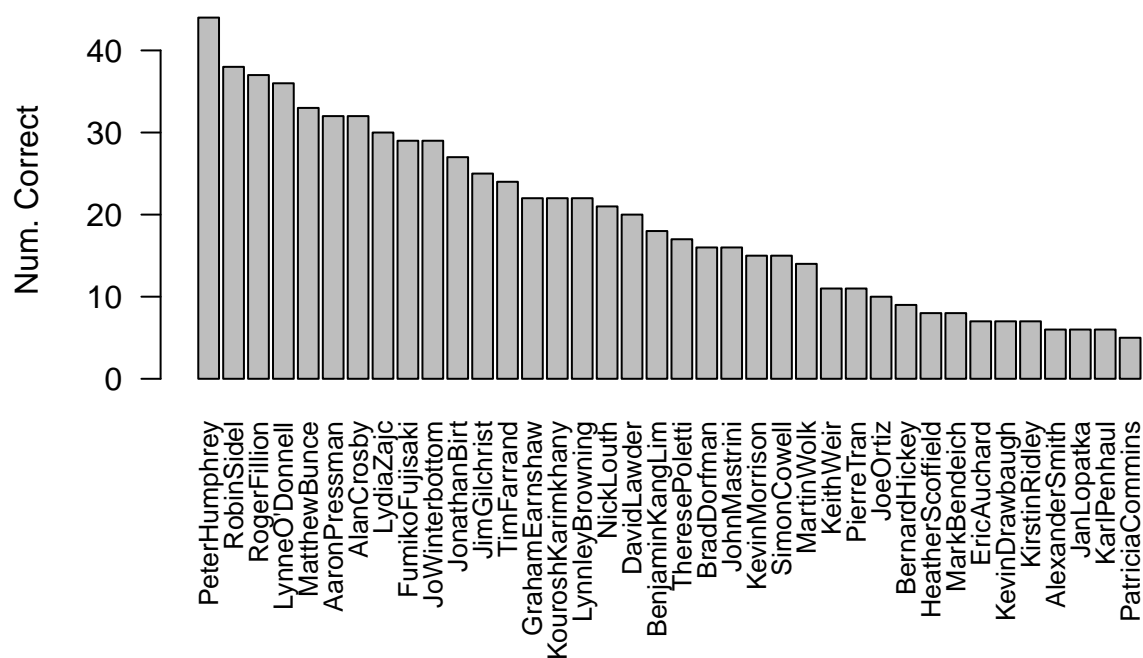
If the reunion is in September, October, or November, it may be worthwhile to consider F9. Also, although there is no data for October and November for XE, in 2008 it had 0 delays over 17 flights in December. This may be a trend worth investigating for someone flying in December, especially as XE may be looked over by other delay-wary passengers, who would likely choose F9 or US.

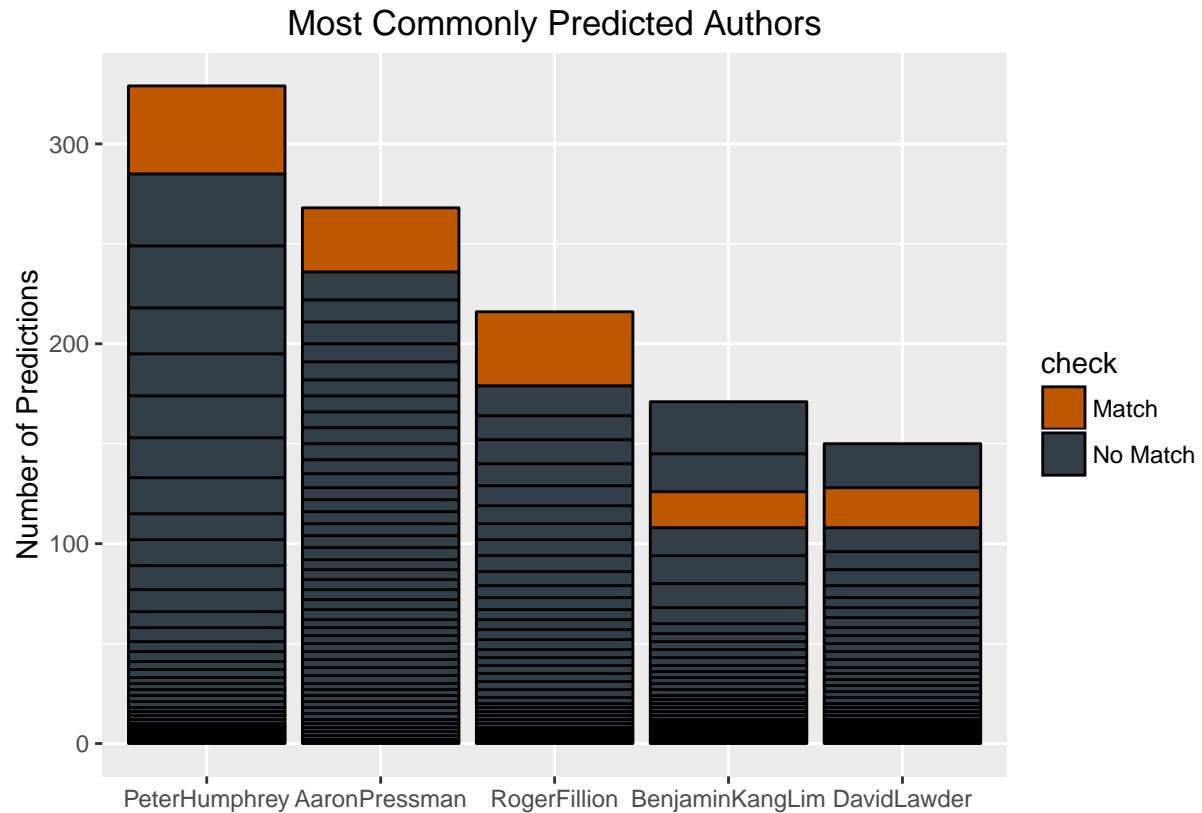
Question 2: Author Attribution

Naive Bayes model, sparsity at 90%. Able to correctly predict 30.24% of articles' authors correctly.

The model predicted the following authors correctly more than 10% of their test set articles.

Correctly Predicted Authors > 10%



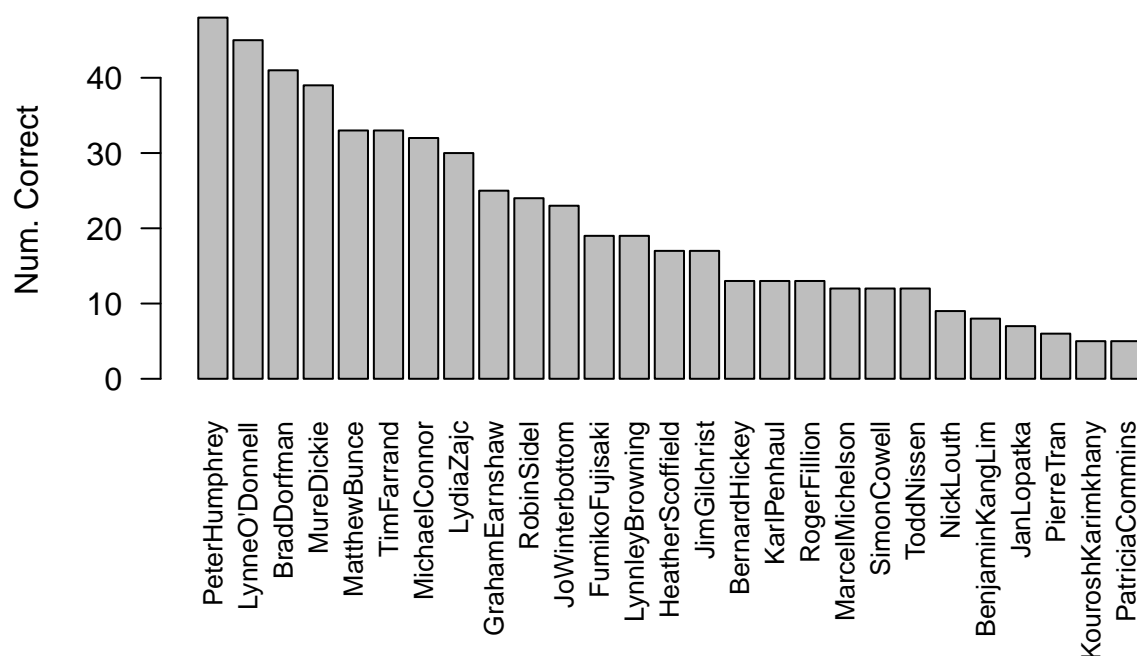


For Naive Bayes, Peter Humphrey was predicted the most often, for many different authors. The most times he was predicted was for his own articles, as is the case for Aaron Pressman and Roger Fillion/ Benjamin Kang Lim and David Lawder, however, were actually predicted more frequently for different authors than for themselves.

Multinomial logistic regression was able to correctly predict 23.56% of articles' authors.

The model predicted the following authors correctly more than 10% of their test set articles.

Correctly Predicted Authors > 10%



Question 3: Groceries

What are the more popular items?

Rule	Support	Confidence	Lift
{other vegetables,yogurt} => {whole milk}	0.0223	0.5129	2.0072
{tropical fruit,yogurt} => {whole milk}	0.0151	0.5174	2.0248
{other vegetables,whipped/sour cream} => {whole milk}	0.0146	0.507	1.9844

The more “popular” baskets of customers are all very practical; these are shoppers on a mission. They are stocking up on essentials like fruits, veggies, yogurt, and milk. With slightly more than 50% of customers who have the first items in their baskets also purchasing whole milk, it is a good idea to leverage this trend in groceries stores. This analysis supports the store layout methodology to keep milk at the back of the store. When the “fruits and veggies” customers are more than likely to seek out milk as well, it makes sense to have them walk past as many other options as possible on the way.

What are some interesting trends?

Rule	Support	Confidence	Lift
{Instant food products,soda} => {hamburger meat}	0.0012	0.6316	18.9957
{popcorn,soda} => {salty snack}	0.0012	0.6316	16.6978
{baking powder,flour} => {sugar}	0.001	0.5556	16.4081
{ham,processed cheese} => {white bread}	0.0019	0.6333	15.0455
{Instant food products,whole milk} => {hamburger meat}	0.0015	0.5	15.0382
{domestic eggs,processed cheese} => {white bread}	0.0011	0.5238	12.4436

To find surprising basket rules, allow the “support” to go to 0.1 % of the customers. They are not frequent among customers, but when they buy the first two items, these customers are much more likely to purchase the third. These customers are either shopping for a specific purpose, or their “stock up” trips look different from an average family shopper. For example, customers shopping for popcorn and soda are 16.7 times more likely to need a salty snack to complete their movie nights, as compared to an average shopper.