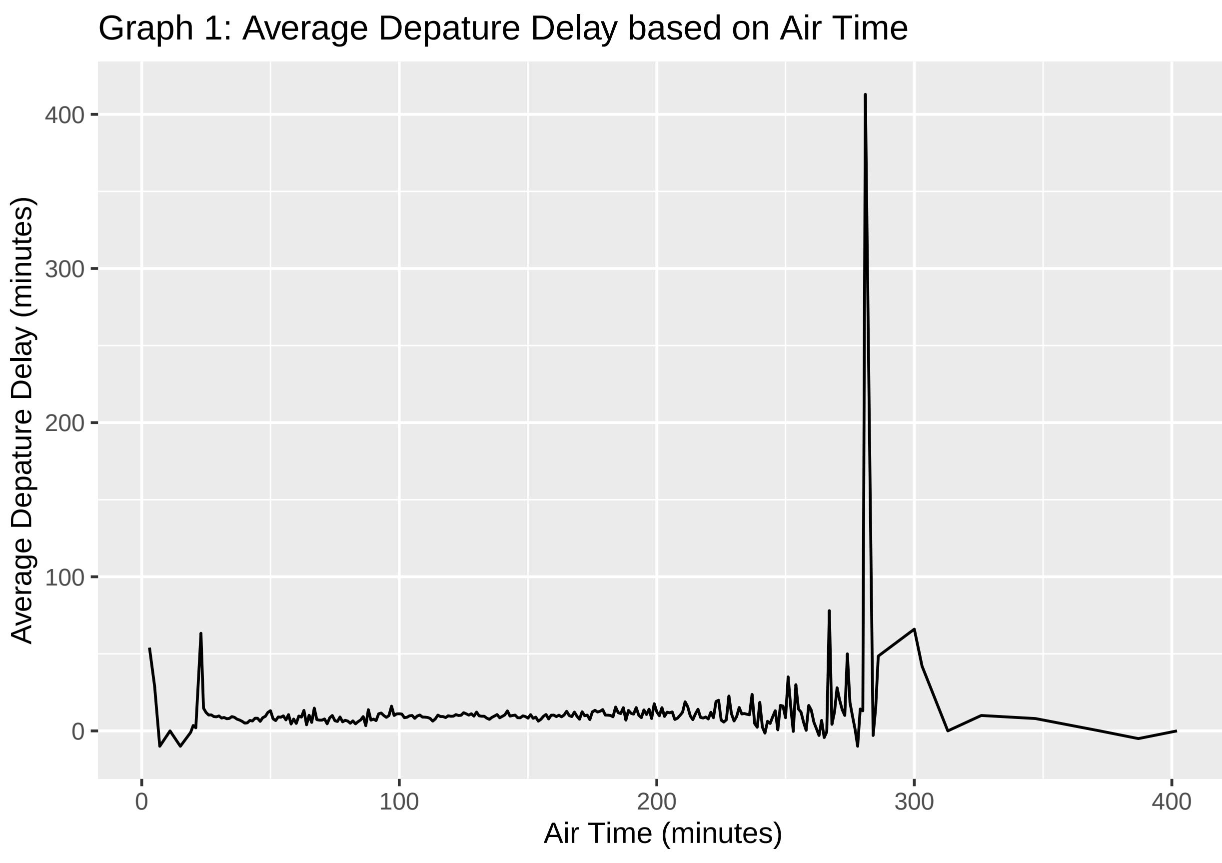
ECO 395M: Exercises 1

Haokun Zhang, Jyun-Yu Cheng and Lu Zhang

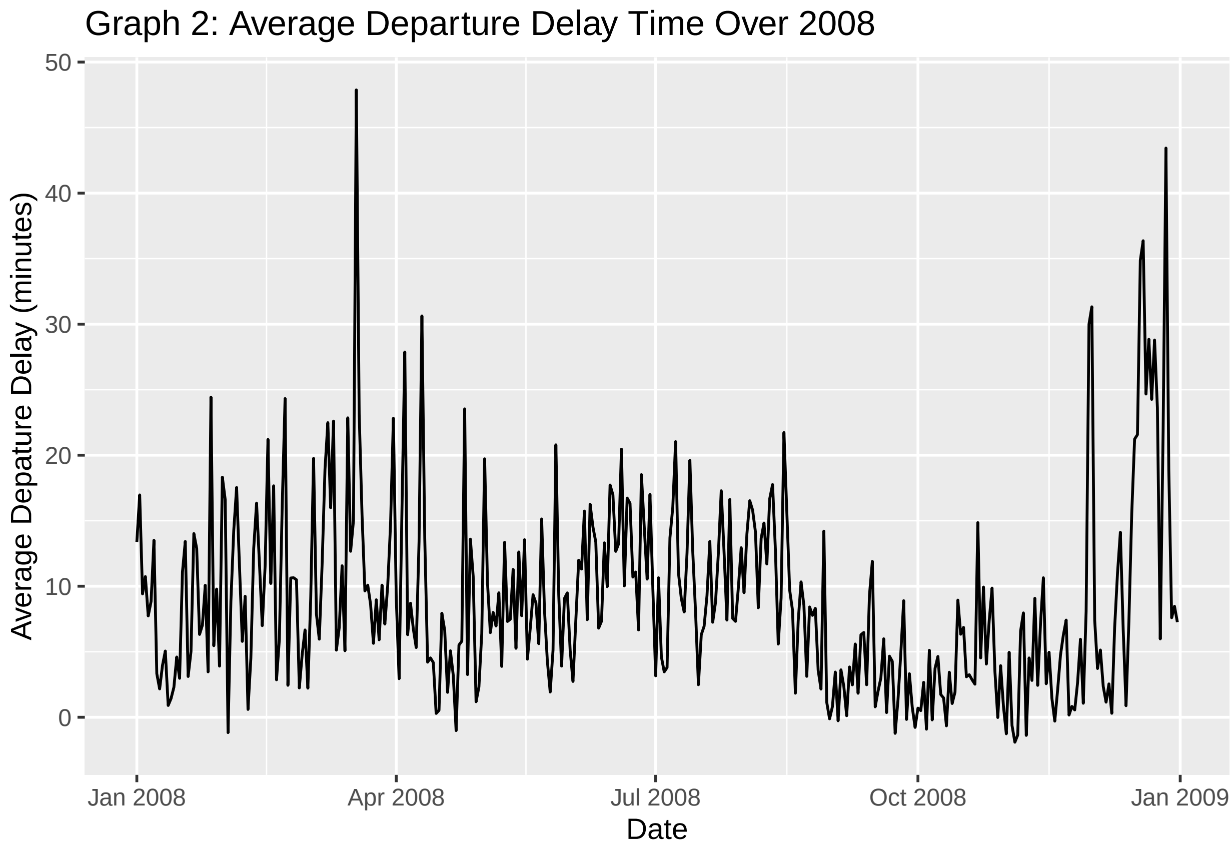
2023-01-25

1) Data visualization: flights at ABIA

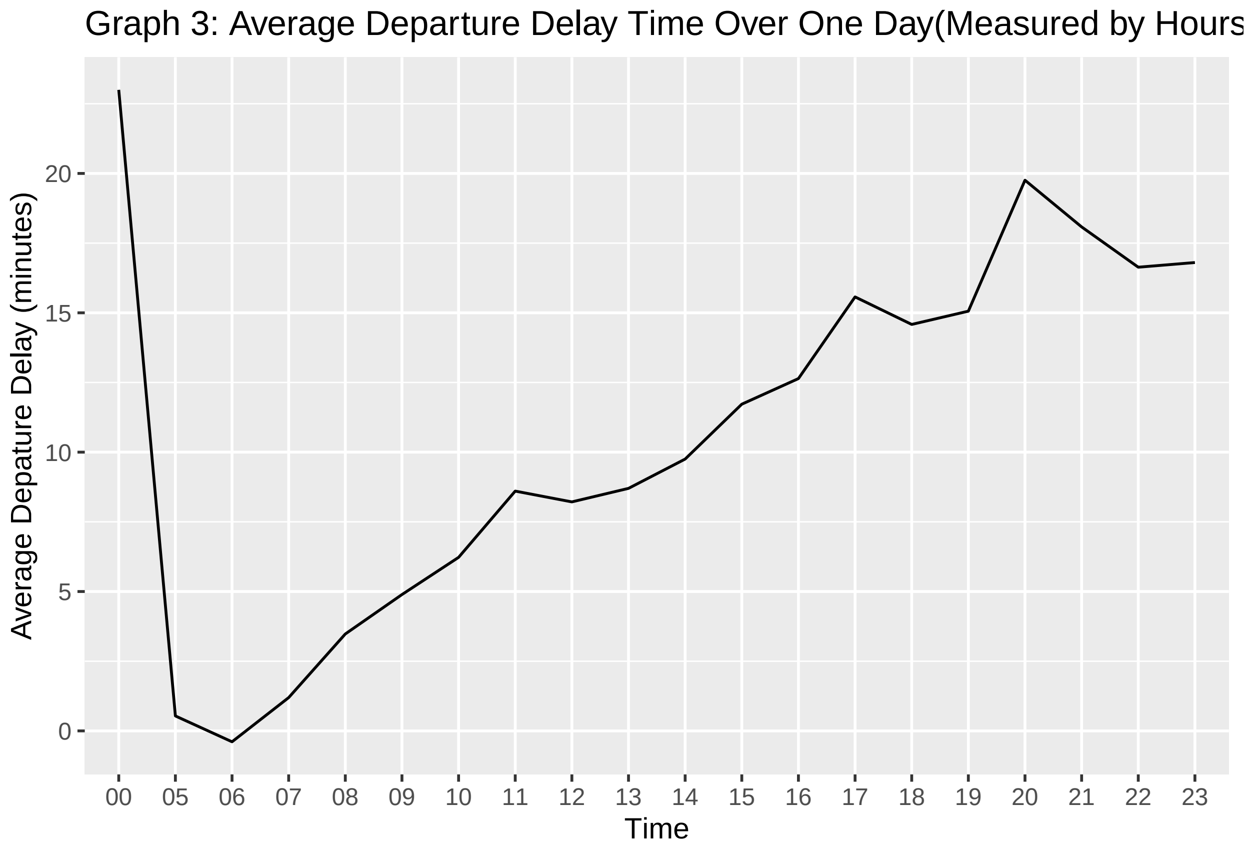
Your task is to create a figure, or set of related figures, that tell an interesting story about flights into and out of Austin.



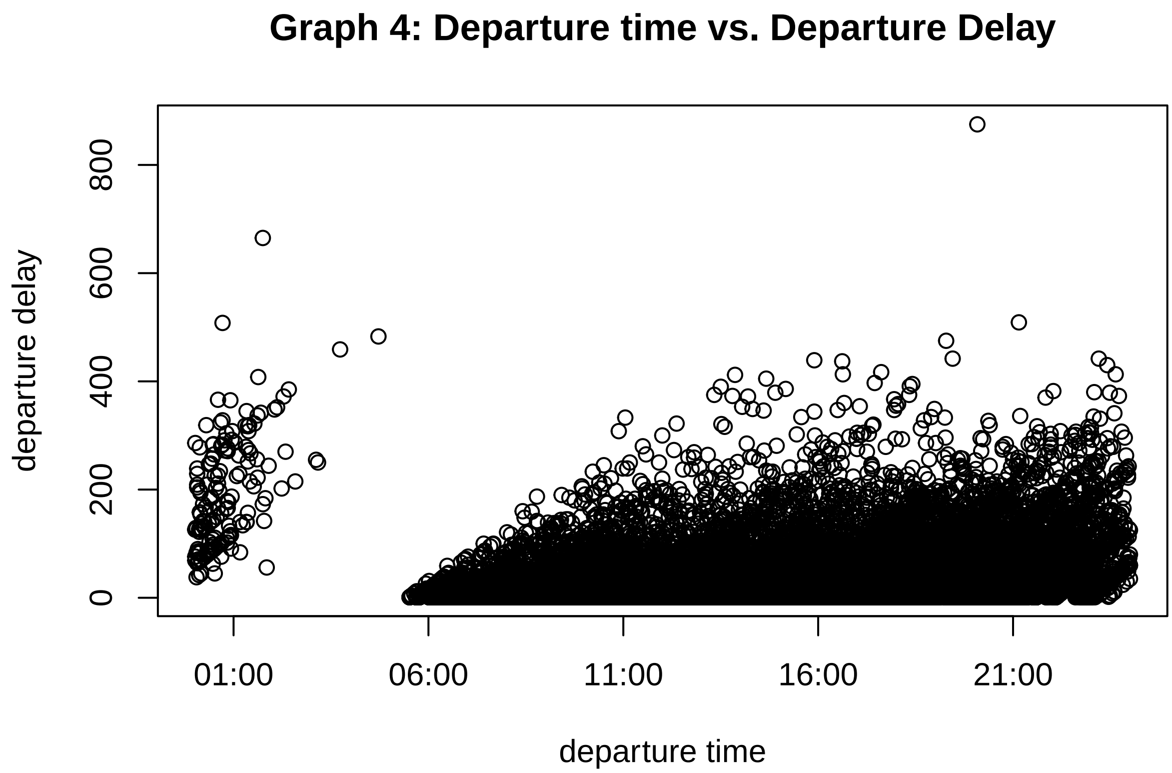
The graph shows that average time of departure delays are similar to each other and slightly above 0 when the time of flight is between 50 to 200 minutes. When it exceeds 200 minutes, the average length of delay became larger and more volatile. When the air time is 281 minutes, the average departure delay is 413 minutes.



We can see the result that in most time, the delay time fluctulate between 0 and 20, and around Oct, the delay time is relatively small.

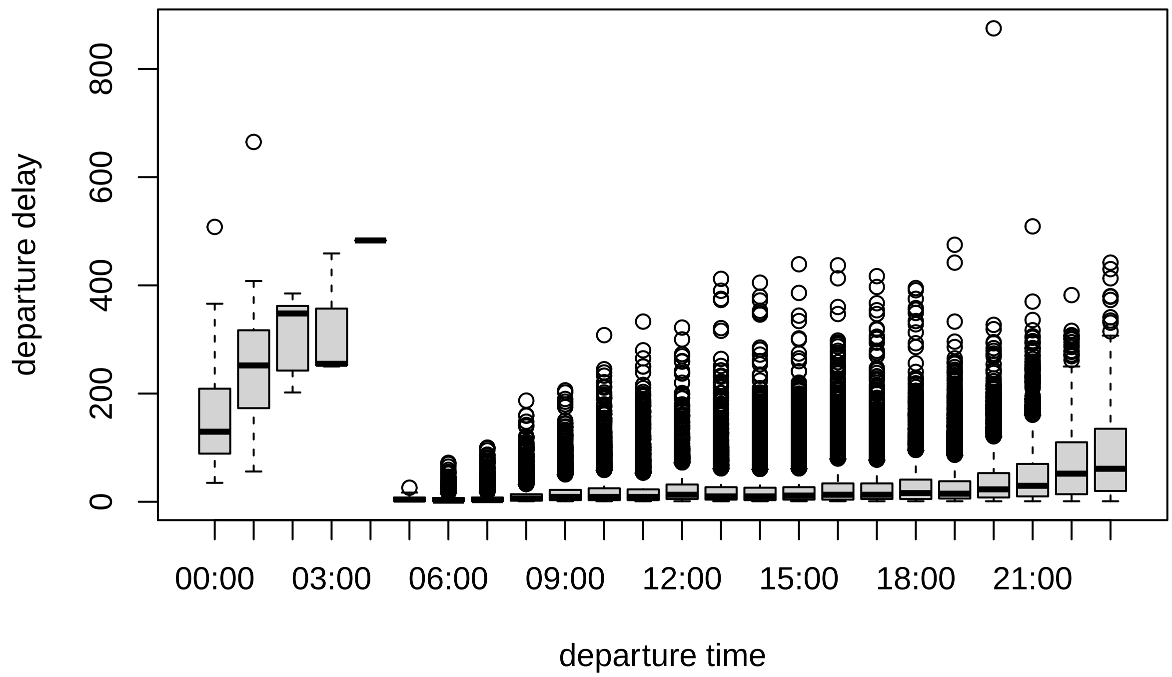


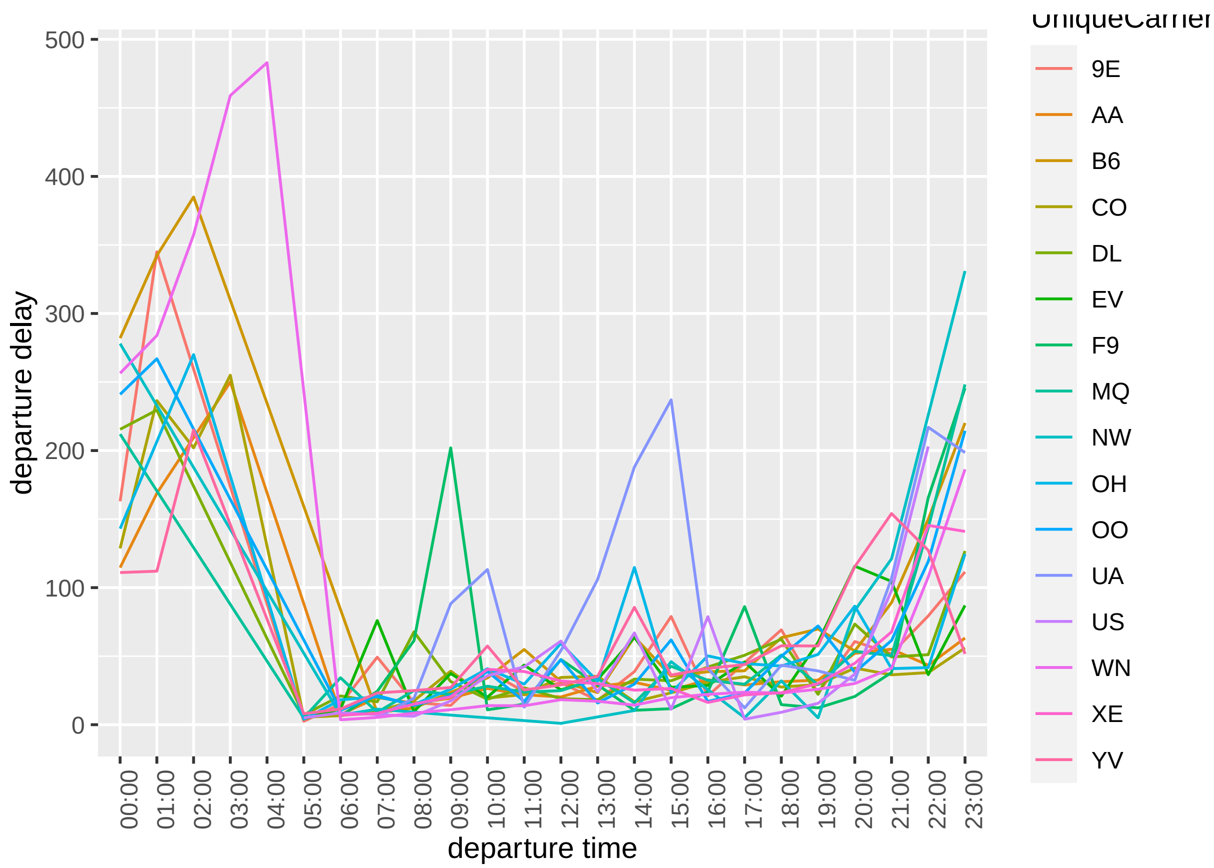
We can find from 6am, the average delay time is increased in general, and due to schedule arrangement, there is no flight delarture between 00 and 06. So we give a suggestion: try to catch earlier flight rather than later flight.



From 0AM-5AM,there are fewer flights and longer delays during this time period. The best time of day to fly to minimize the delays will be 5-8 AM, and the average delay time is less than 10 minutes. For 8-10 AM, the average delay time is still less than 20 minutes.

Graph 5: Departure Time on Hourly Base





The above trends may differ depends on airlines. In 7-8AM, the EV airline has an average delay departure time more than one hour. In other case, for example, F9 may has extreme delay departure time, more than 200 minutes, for 9-10 AM.

2) Wrangling the Olympics

A) What is the 95th percentile of heights for female competitors across all Athletics events (i.e., track and field)? Note that `sport` is the broad sport (e.g. Athletics) whereas `event` is the specific event (e.g. 100 meter sprint).

Table 1: 95% heights for female competitors

	quantile	height
95%	95%	183

The 95th percentile of heights for female competitors across all Athletics events is 183cm.

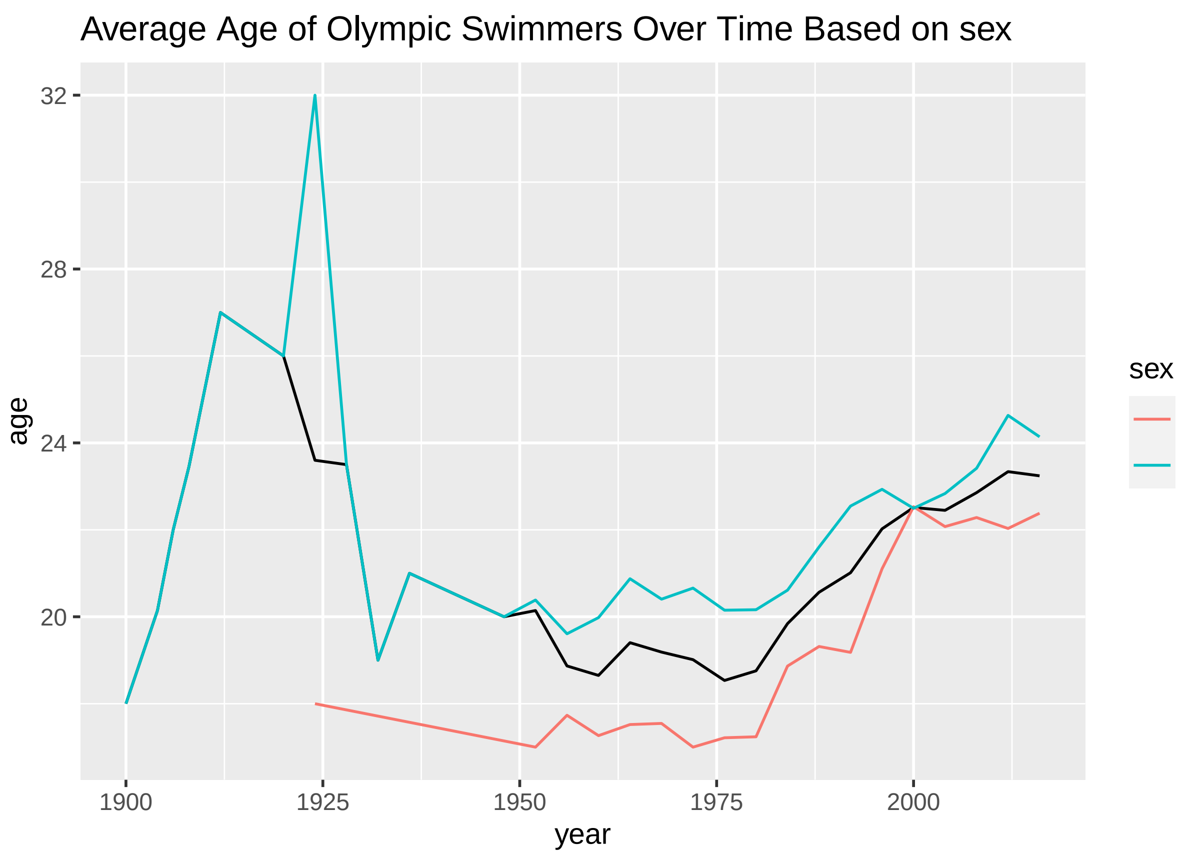
B) Which single women's `event` had the greatest variability in competitor's heights across the entire history of the Olympics, as measured by the standard deviation?

Table 2: Top 5 Female Events that Have the Greatest Variability in Height

	event	height_std_dev
Rowing Women's Coxed Fours	Rowing Women's Coxed Fours	10.865490
Basketball Women's Basketball	Basketball Women's Basketball	9.700255
Rowing Women's Coxed Quadruple Sculls	Rowing Women's Coxed Quadruple Sculls	9.246396
Rowing Women's Coxed Eights	Rowing Women's Coxed Eights	8.741931
Swimming Women's 100 metres Butterfly	Swimming Women's 100 metres Butterfly	8.134398
Volleyball Women's Volleyball	Volleyball Women's Volleyball	8.101521

The Rowing Women's Coxed Fours had the greatest variability in competitor's heights across the entire history of the Olympics, as measured by the standard deviation of 10.865490.

C) How has the average age of Olympic swimmers changed over time? Does the trend look different for male swimmers relative to female swimmers?



The average age of Olympic swimmers decreased after 1924, when female swimmers joined Olympic for the first time. Before 1924, there was only male Olympic swimmers. From 1950 to 2008, the average age shows an increasing trend, with a gap between 2-3 years old. The age of male swimmers is higher than female, except for 2000, when female swimmer's average age spiked and overlapped with male swimmer's average age.

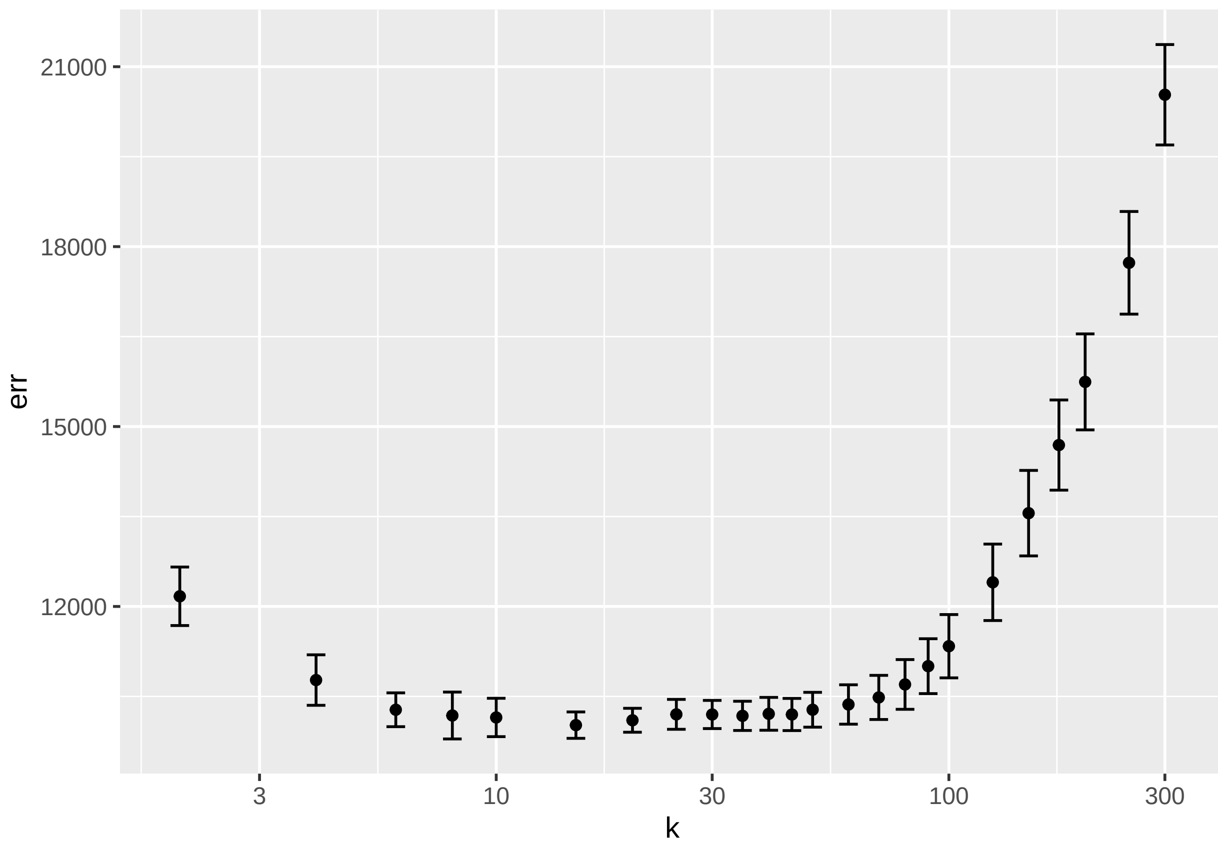
3) K-nearest neighbors: cars

The data in sclass.csv contains data on over 29,000 Mercedes S Class vehicles—essentially every such car in this class that was advertised on the secondary automobile market during 2014. For websites like Cars.com or Truecar that aim to provide market-based pricing information to consumers, the Mercedes S class is a notoriously difficult case. There is a huge range of sub-models that are all labeled "S Class,"" from large luxury sedans to high-performance sports cars; one sub-category of S class has even served as the safety car in Formula 1 Races. Moreover, individual submodels involve cars with many different features. This extreme diversity—unusual for a single model of car—makes it difficult to provide accurate pricing predictions to consumers.

Use K-nearest neighbors to build a predictive model for price, given mileage, separately for each of two trim levels: 350 and 65 AMG. That is, Treating the 350's and the 65 AMG's as two separate data sets.

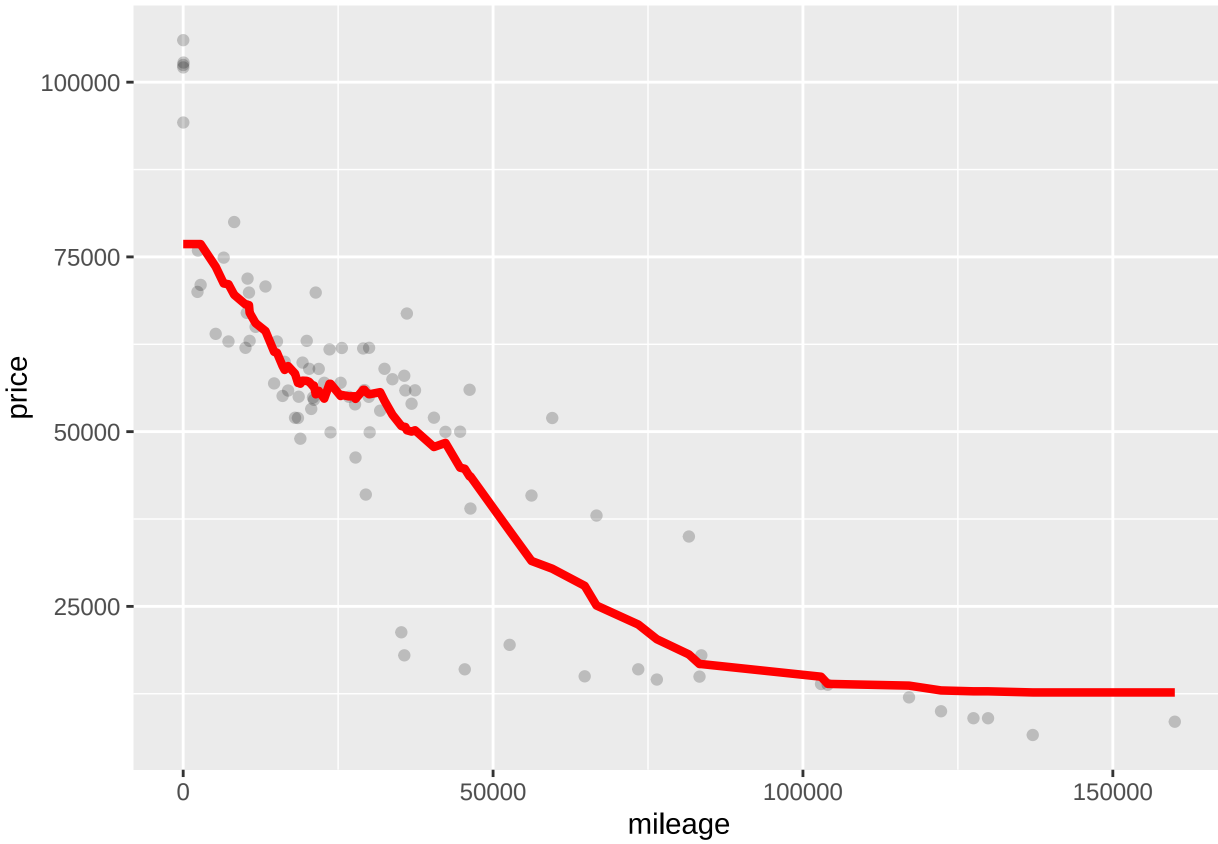
Trim 350

KNN Test Plot



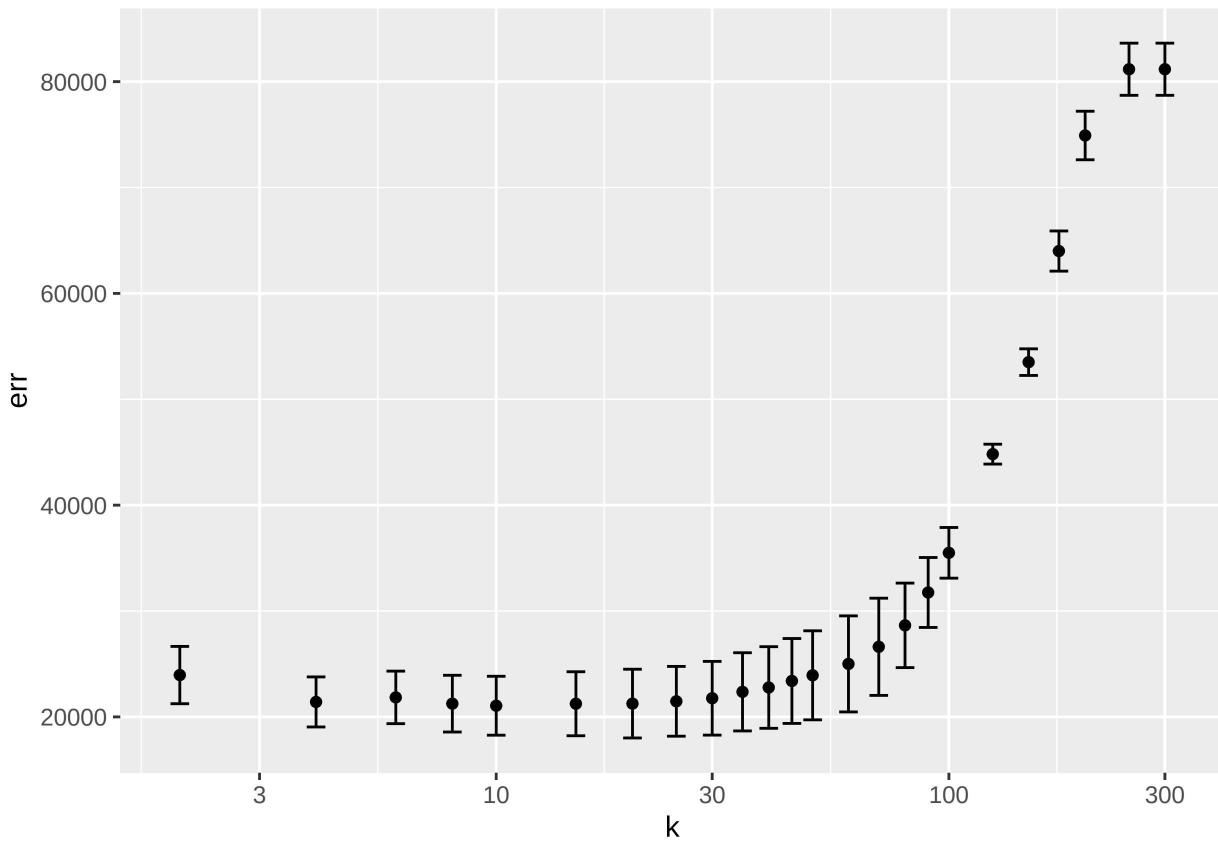
For Trim 350, K=32 is a ideal selection.

Trim 350 Prediction Plot



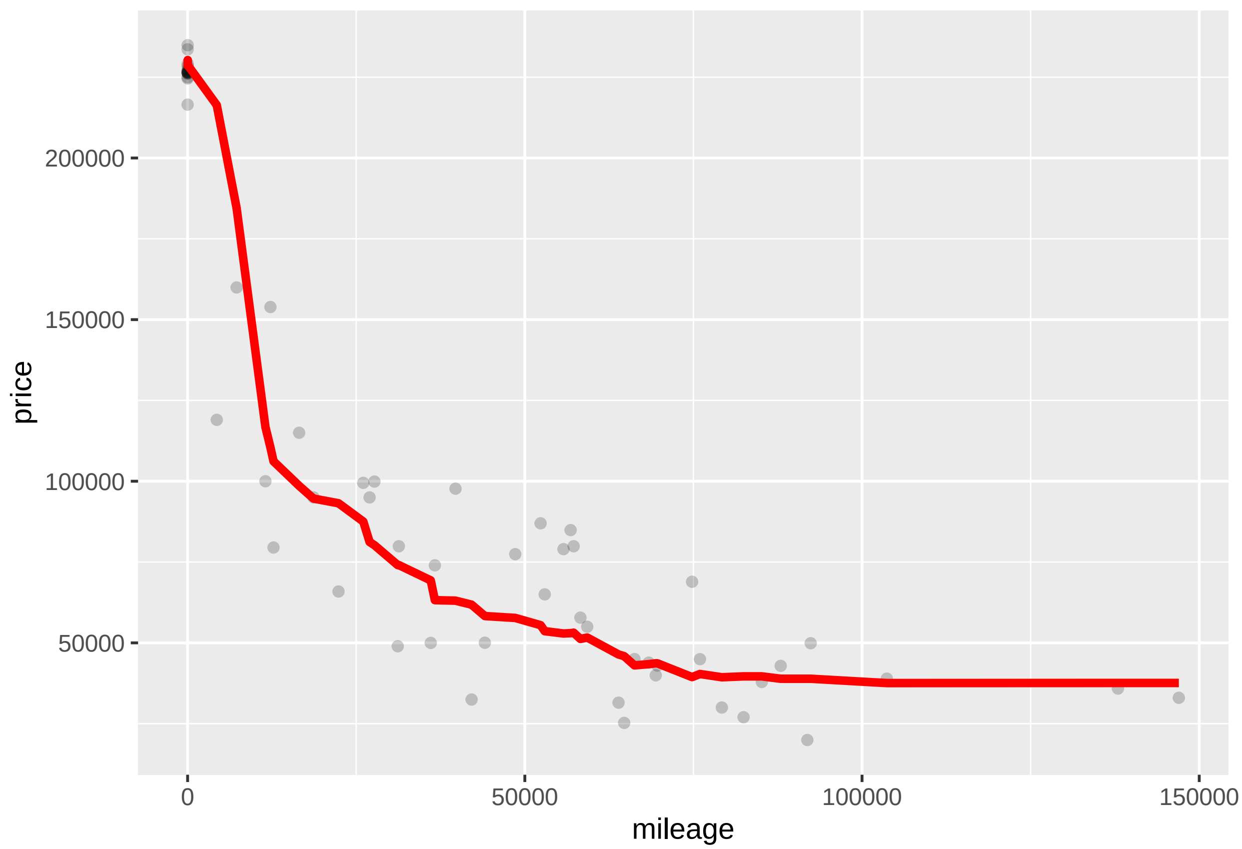
Trim 65 AMG

Trim 65 AMG KNN Test Plot



For Trim 65 AMG, k=30 is an ideal selection with the lowest RMSE.

Trim 65 AMG Prediction Plot



There are 292 samples of Trim 65 AMG and 416 samples of Trim 350. By eyeballing, Trim 350 has the most optimal trade off between bias and variance at k=32. Trim 65 AMG has the most optimal trade off at k=30.