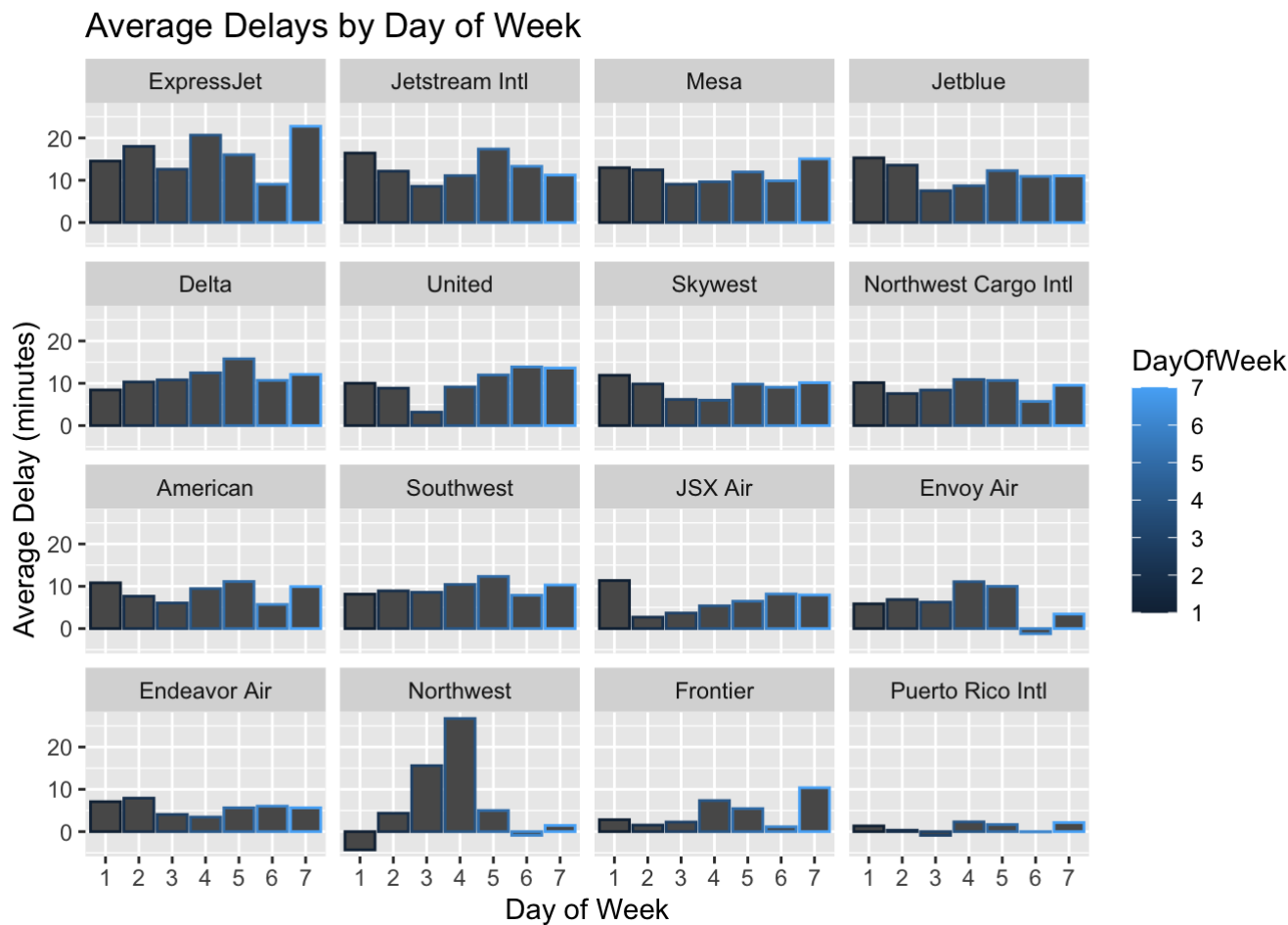# Exercises 1

Abby Johnson

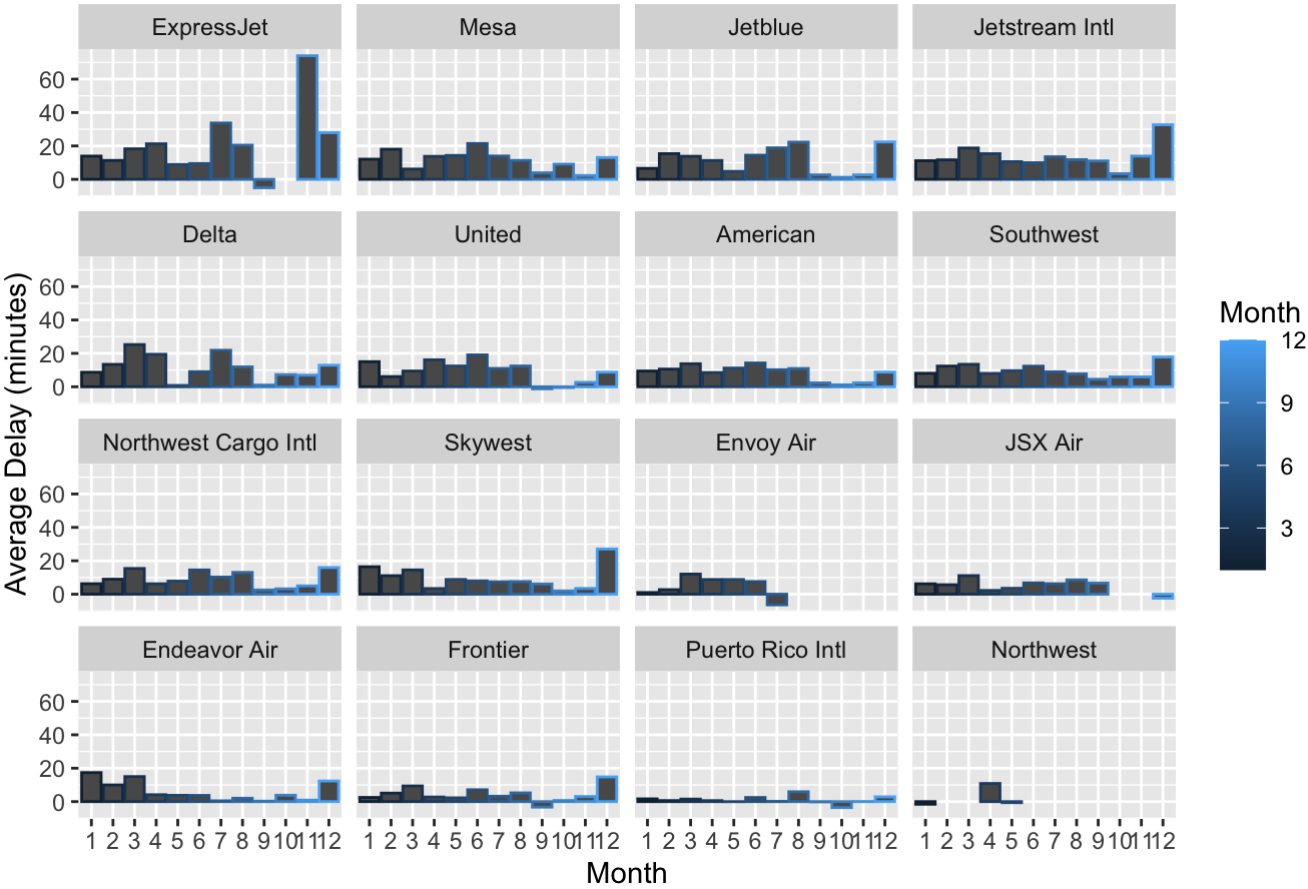2/11/2022

1) Data visualization: flights at ABIA

Does average airline departure delay vary by day of the week? Or maybe by time of the year?

After averaging departure delays by day of week and month, we can understand how airline departure delays may vary.

ExpressJet, Jetstream Intl, Mesa, and Jetblue are the four airline carriers with the longest average departure delays, while Endeavor Air, Northwest, Frontier, and Puerto Rico Intl have the shortest delays. Moreover, most carriers experience their longest delays at the beginning of the week or the weekend. Meanwhile, the day with shortest delays typically is Tuesday or Wednesday.

Average Delays by Month of Year

Compared to the average daily delays, the same four airlines have the longest delays and the same four airlines have the shortest delays throughout the year. However, across most airlines we see the lowest delays around the fall. This makes sense when we consider weather and travel patterns throughout the year.

From these results, we can conclude that traveling early-mid week in the fall on Endeavor Air, Frontier, Puerto Rico Intl, or Northwest airlines will most likely result in the shortest departure delays out of Austin.

2) Wrangling the Billboard Top 100

The Billboard Top 100 helps us understand what artists and songs are popular throughout musical history. From Billboard Top 100 data, we can better understand top songs, musical diversity, and long-run hits.
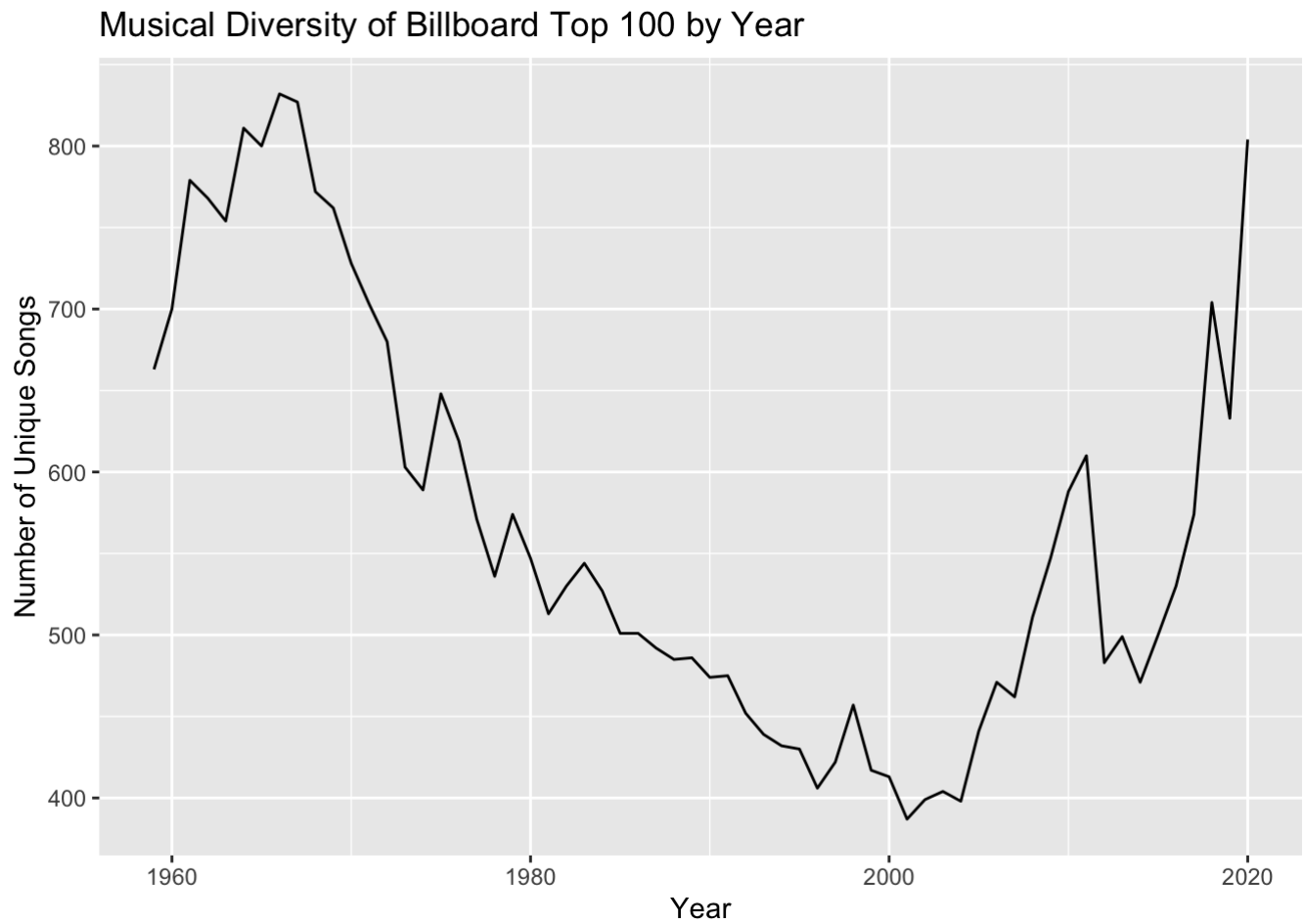
Part A

Top 10 Songs on Billboard Top 100 by Total Number of Weeks

Performer	Song	Count
Imagine Dragons	Radioactive	87
AWOLNATION	Sail	79

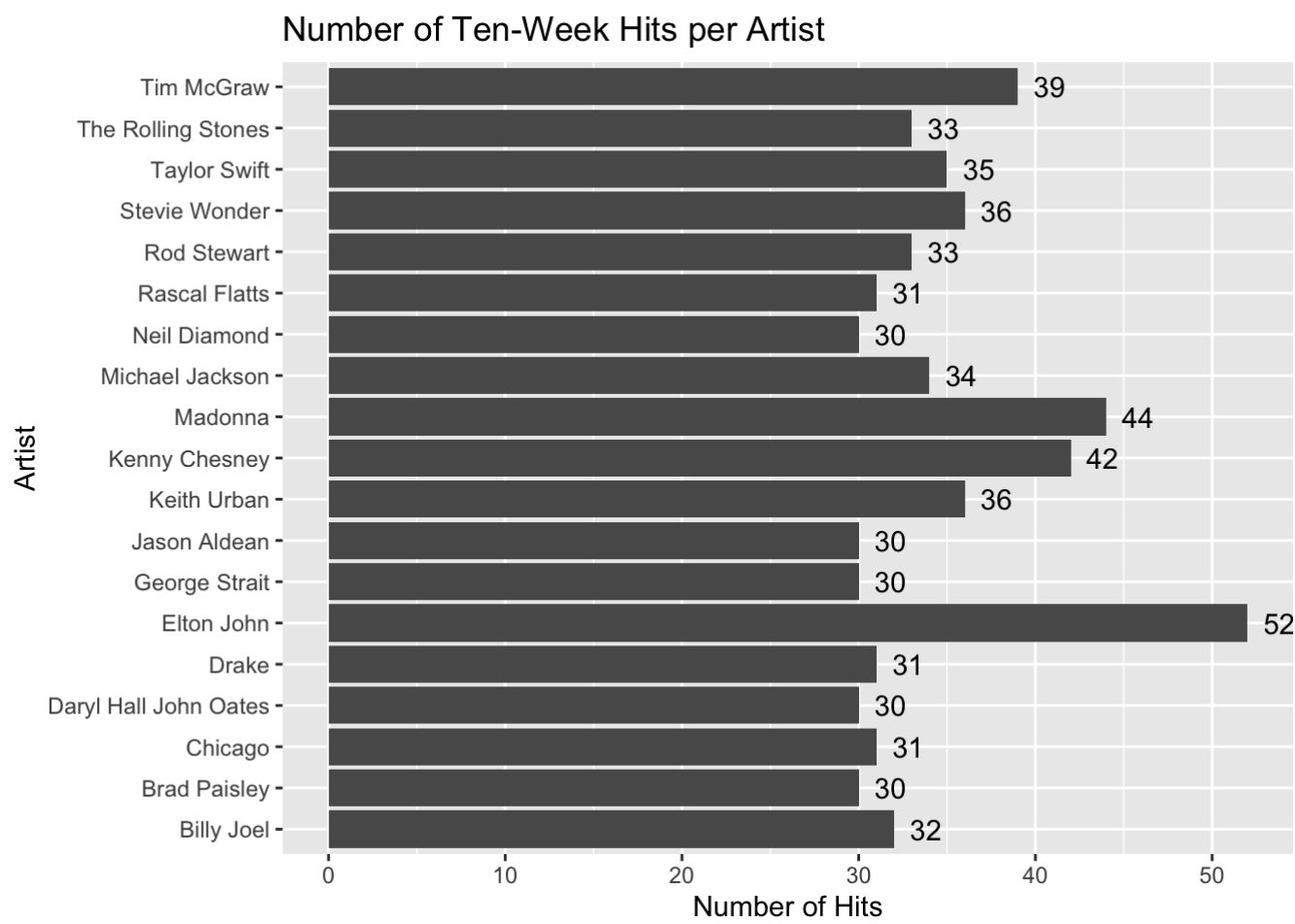| Performer	Song	Count
Jason Mraz	Im Yours	76
The Weeknd	Blinding Lights	76
LeAnn Rimes	How Do I Live	69
LMFAO Featuring Lauren Bennett & GoonRock	Party Rock Anthem	68
OneRepublic	Counting Stars	68
Adele	Rolling In The Deep	65
Jewel	Foolish Games/You Were Meant For Me	65
Carrie Underwood	Before He Cheats	64

The number one song is Radioactive by Imagine Dragons with 87 total weeks spent on the Billboard Top 100. Coming in at number ten is Before He Cheats by Carrie Underwood, with 64 weeks on the Billbaord Top 100.

Part B

Musical Diversity of Billboard Top 100 by Year

The musical diversity (or number of unique songs) of the Billboard Top 100 was highest in the 1960s and around 2020. From about 1980 to 2010, there was less musical diversity on the Billboard Top 100, which suggests that listeners preferred a less diverse collection of msuic during this time.

Part C

Number of Ten-Week Hits per Artist

Artist	Number of Hits
Tim McGraw	39
The Rolling Stones	33
Taylor Swift	35
Stevie Wonder	36
Rod Stewart	33
Rascal Flatts	31
Neil Diamond	30
Michael Jackson	34
Madonna	44
Kenny Chesney	42
Keith Urban	36
Jason Aldean	30
George Strait	30
Elton John	52
Drake	31
Daryl Hall John Oates	30
Chicago	31
Brad Paisley	30
Billy Joel	32

Only 19 aritsts had at least 30 songs on the Billboard Top 100 for at 10 weeks or longer. Of these 19 artists, we can see that the top three artists are: Elton John with 52 hits, Madonna with 44 hits, and Kenny Chesney with 42 hits.

3) Wrangling the Olympics

Part A

The 95th percentile of heights for female competitors across all Athletics events

Event	95th Percentile
Athletics Women's 1,500 metres	172.00

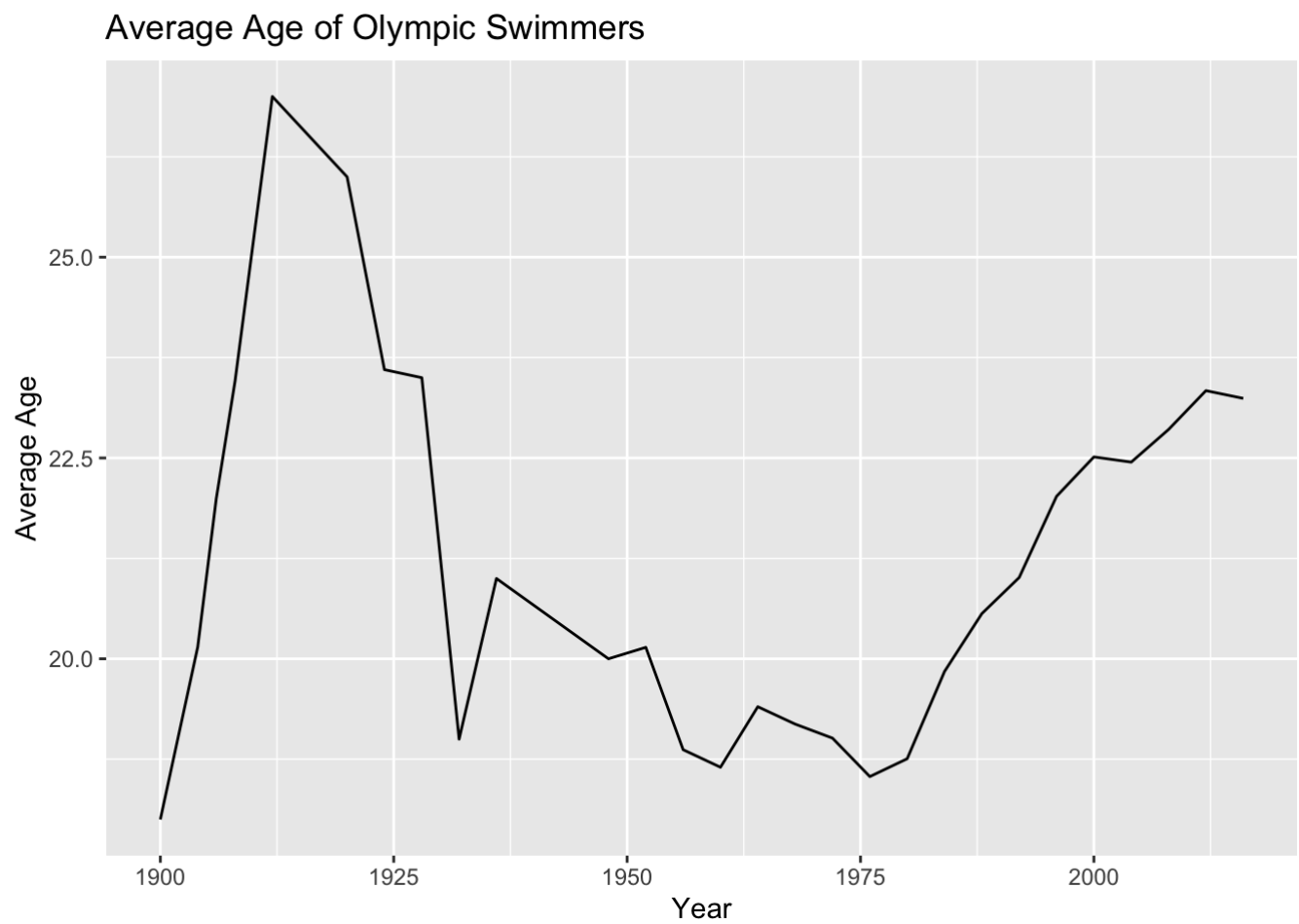| Event	95th Percentile
Athletics Women's 10 kilometres Walk	170.00
Athletics Women's 10,000 metres	167.80
Athletics Women's 100 metres	179.60
Athletics Women's 100 metres Hurdles	176.00
Athletics Women's 20 kilometres Walk	173.00
Athletics Women's 200 metres	180.00
Athletics Women's 3,000 metres	170.00
Athletics Women's 3,000 metres Steeplechase	176.80
Athletics Women's 4 x 100 metres Relay	176.00
Athletics Women's 4 x 400 metres Relay	178.00
Athletics Women's 400 metres	180.00
Athletics Women's 400 metres Hurdles	181.40
Athletics Women's 5,000 metres	170.75
Athletics Women's 80 metres Hurdles	174.80
Athletics Women's 800 metres	175.75
Athletics Women's Discus Throw	188.40
Athletics Women's Hammer Throw	183.90
Athletics Women's Heptathlon	188.70
Athletics Women's High Jump	188.00
Athletics Women's Javelin Throw	186.20
Athletics Women's Long Jump	181.35
Athletics Women's Marathon	171.40
Athletics Women's Pentathlon	179.00
Athletics Women's Pole Vault	183.00
Athletics Women's Shot Put	191.50
Athletics Women's Triple Jump	186.90

Part 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?

Event	Standard Deviation
Rowing Women's Coxed Fours	10.86549

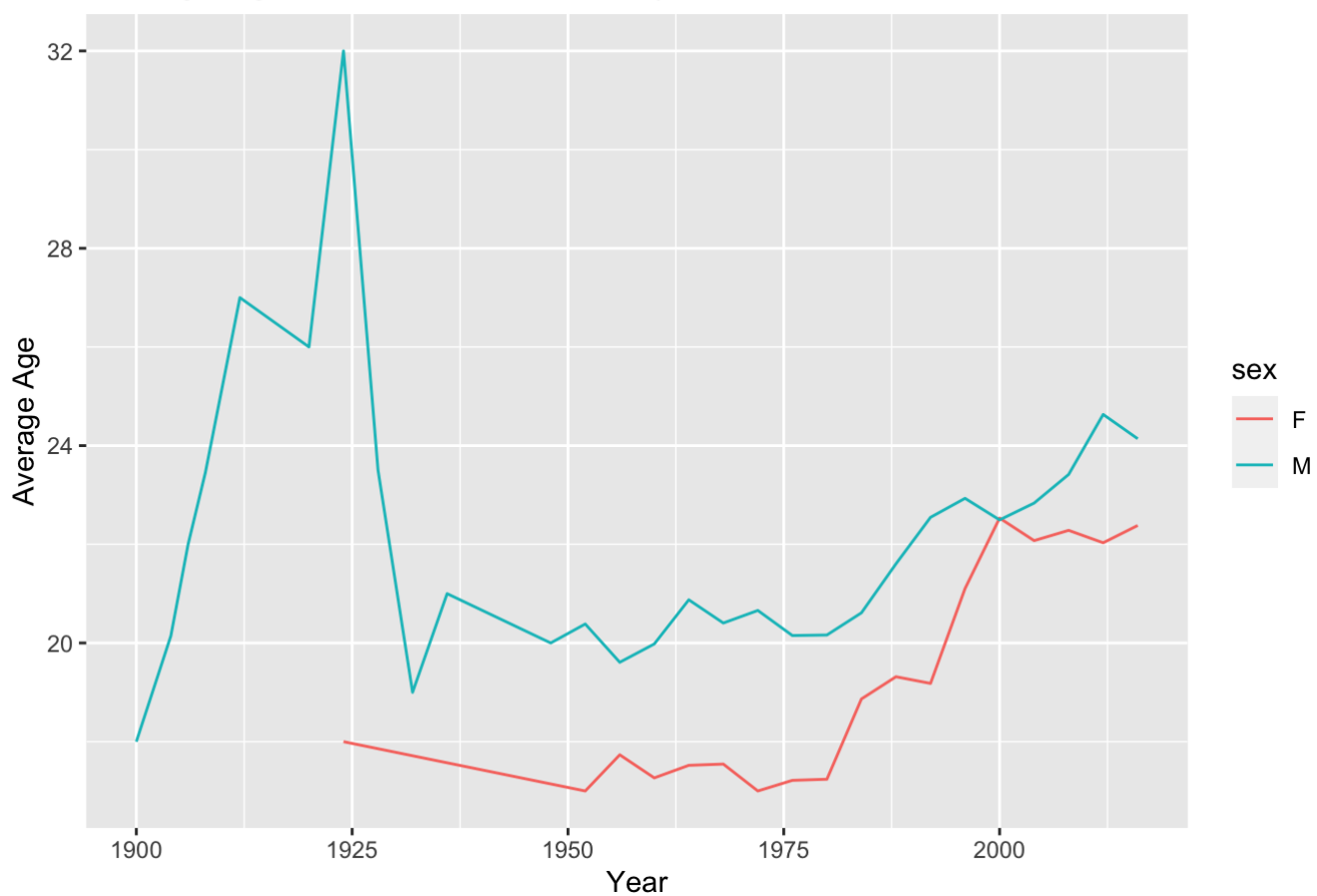The Women's Rowing Coxed Fours has the highest variablity in height across women's single Olympic events.

Part C

How has the average age of Olympic swimmers changed over time?

Average Age of Olympic Swimmers

The average age of all Olympic swimmers in the early 1900s is around 27 years old. However, the average age dips to around 20 during the early and mid 1900s. By the 2010s, however, the average age is back up to around 23.

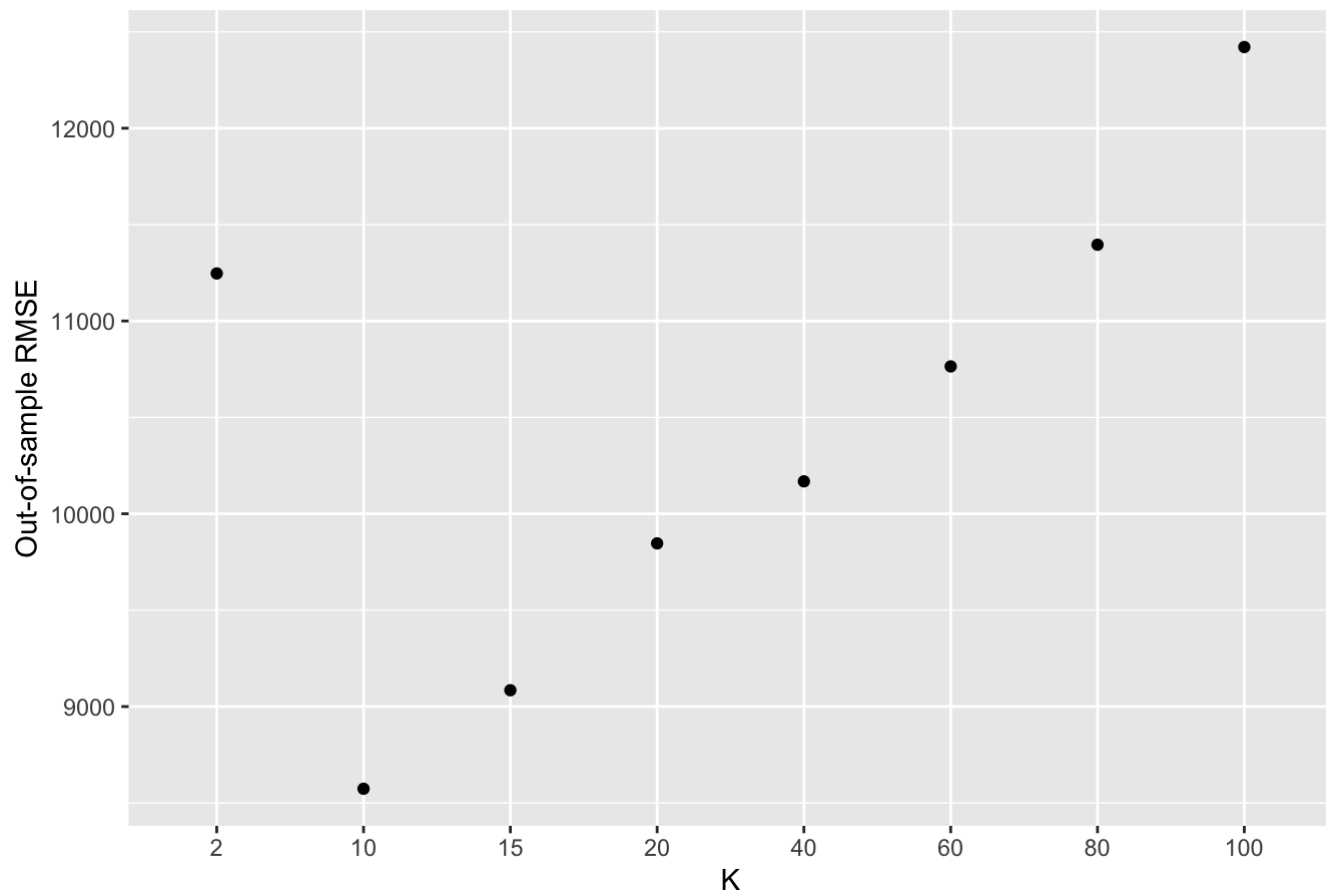## Average Age of Male and Female Olympic Swimmers

The average age for male Olympic swimmers starts around 16 years old in 1900, but shoots up to about 32 in 1925 before dropping back to about 20. By the 2010s, the average male age is about 24. Meanwhile, for female Olympic swimmers, the average age is starts around 16 in 1925 and stays on an upward trend to around 22 in the 2010s.
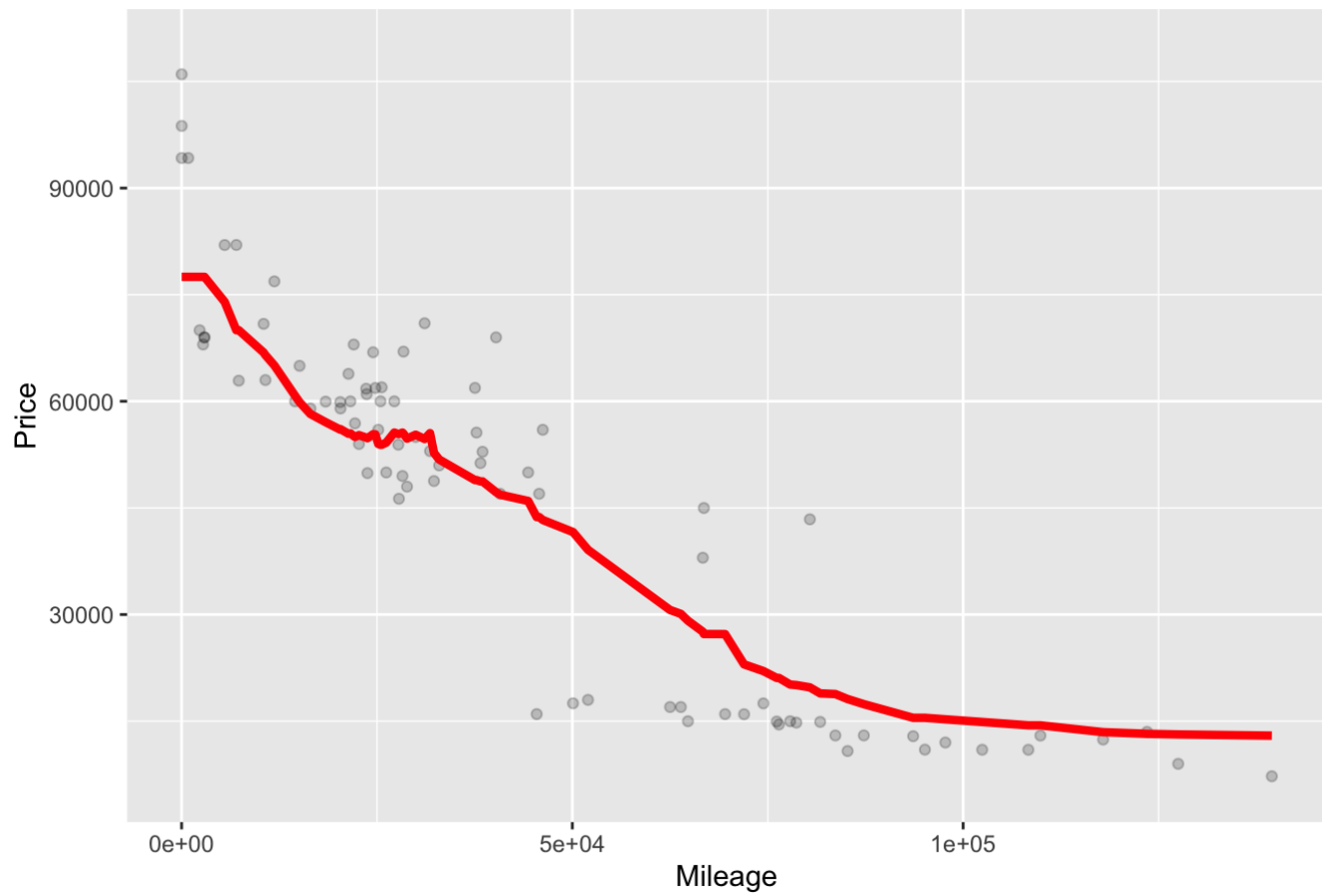
4) K-nearest Neighbors

S CLass 350

S Class 350: Out-of-sample RMSE for Values of K

After plotting the RMSE for different values of K, it is clear the the RMSE bottoms out around K=40 for the S Class 350.
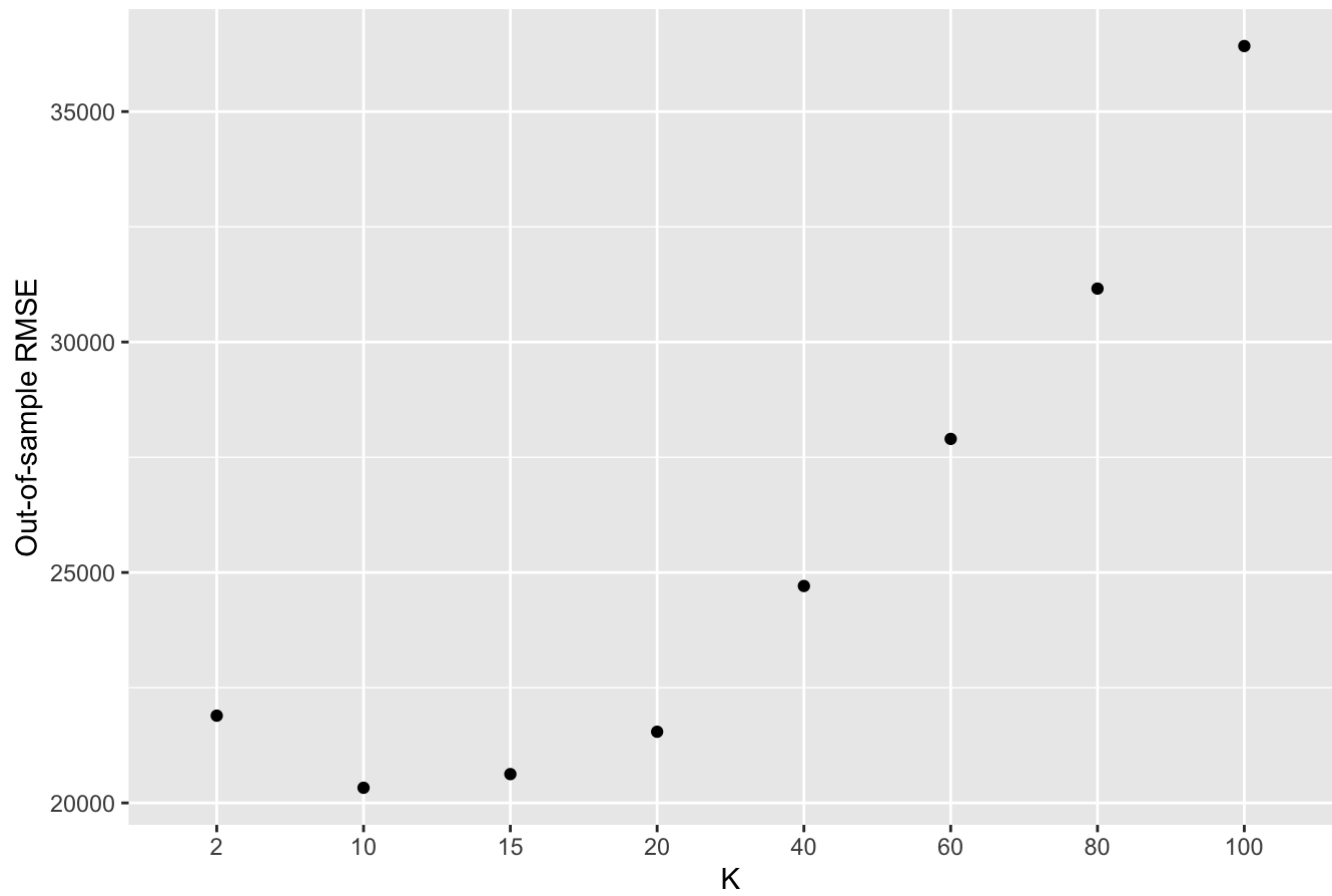
S Class 350: Price & Mileage for K=40

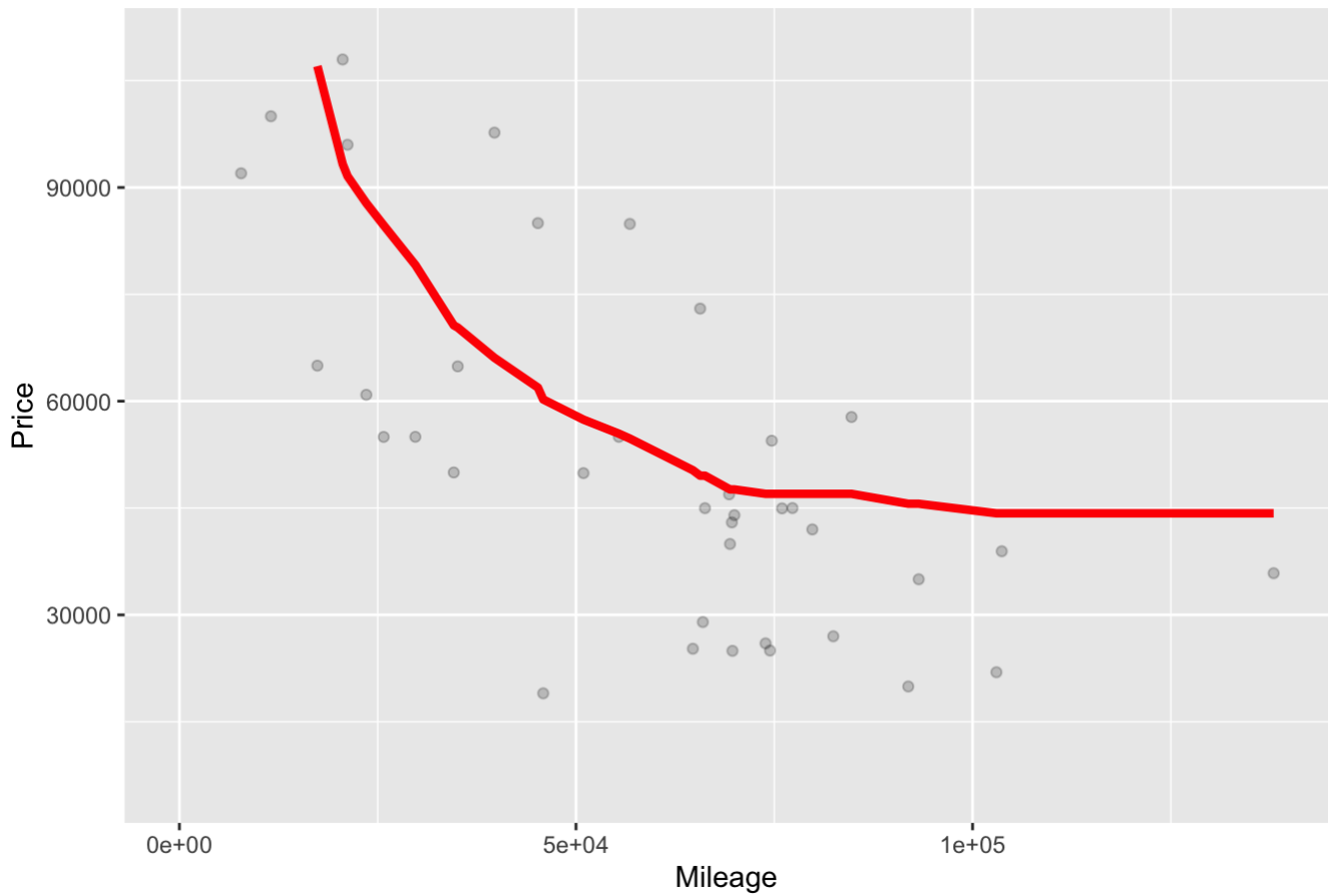Using the optimal value of K=40 for the S Class 350, we can plot the fitted model of price against mileage.

S Class 65 AMG

S Class 65 AMG: Out-of-sample RMSE for Values of K

After plotting the RMSE for different values of K, it is clear the the RMSE bottoms out around K=60 for the S Class 65 AMG.

S Class 65 AMG: Price & Mileage for K=60

Using the optimal value of K=60 for the S Class 65 AMG, we can plot the fitted model of price against mileage.

Relative to the S Class 350, the S Class 65 AMG has a higher optimal value of K. The S Class 65 AMG may have a larger optimal value of K, because the predicted values are not as closely grouped together. Therefore, a larger K is necessary to get a more accurate estimate.