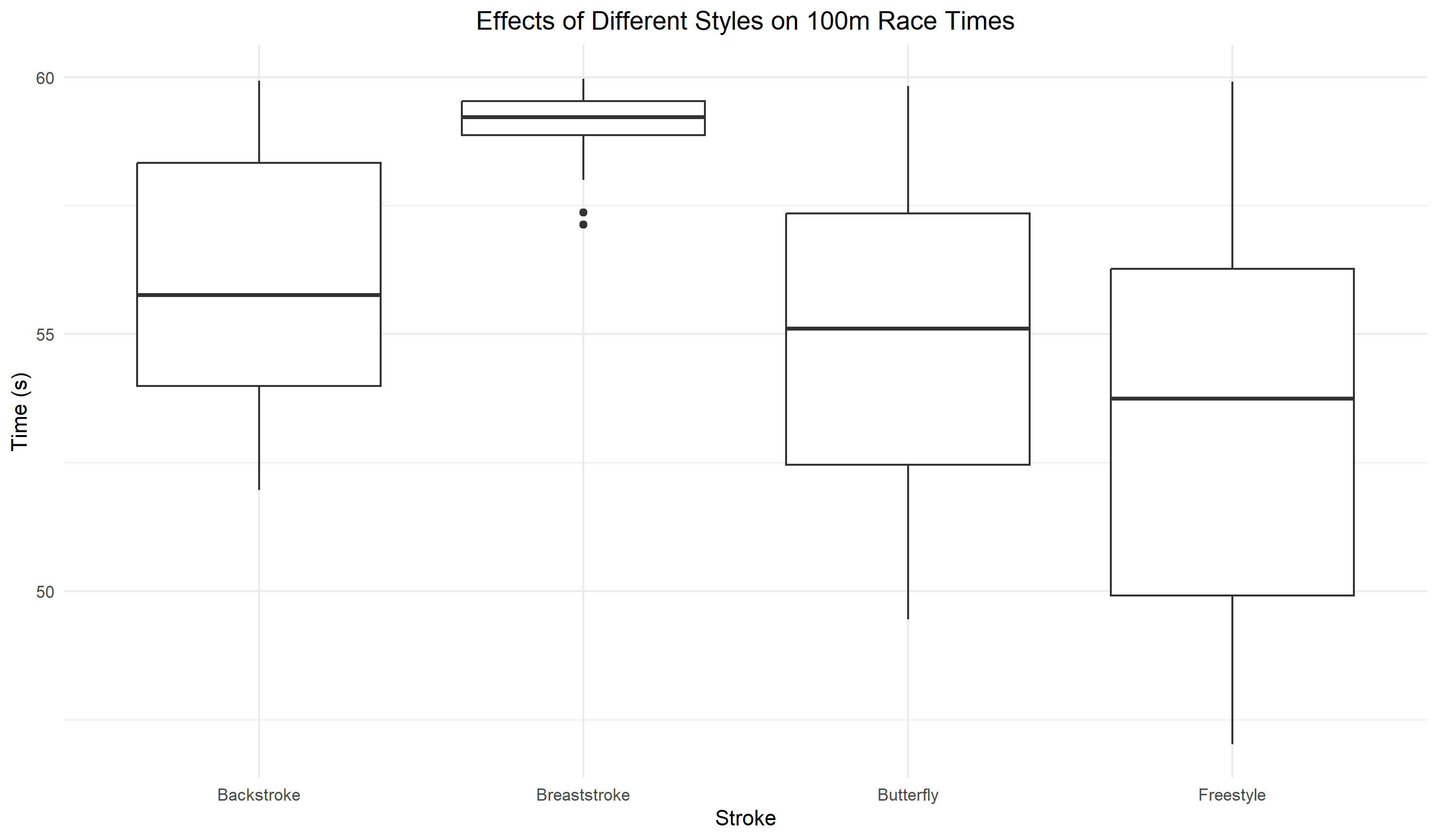
**Analyzing Olympic Swim Times with Boxplots and Hypothesis Tests**

The Summer Olympics are traditionally held every four years, with an extensive history that goes back to 1896. Since then, swimming has been an integral part of the events. The sport has attracted thousands of athletes from over 150 countries around the world. The Olympics allows nations to put aside their differences and come together to provide a sense of unity to millions of fans.

We will be investigating the results of the 50 and 100 meter men and women races from 1924 to 2020. The dataset includes 750 swimmers and 9 variables.

Our main goals are to find which styles of swimming are the best for the 100m race and whether race times have changed significantly for the 100 and 50 meter races over the past century or so. Each style of swimming has its own advantages and disadvantages. The backstroke makes breathing easier but makes it difficult for the swimmer to know where he is going. The breaststroke allows the swimmer’s head to be out of the water for the longest amount of time, however it requires the most strength and endurance of the four. The butterfly stroke, similar to the breaststroke, makes breathing easier because the swimmer’s head spends significant time out of the water, however it requires lots of strength and timing. The freestyle stroke uses your full body and is easiest to learn, however the swimmer keeps their head in the water most of the time so it can be difficult to breathe.

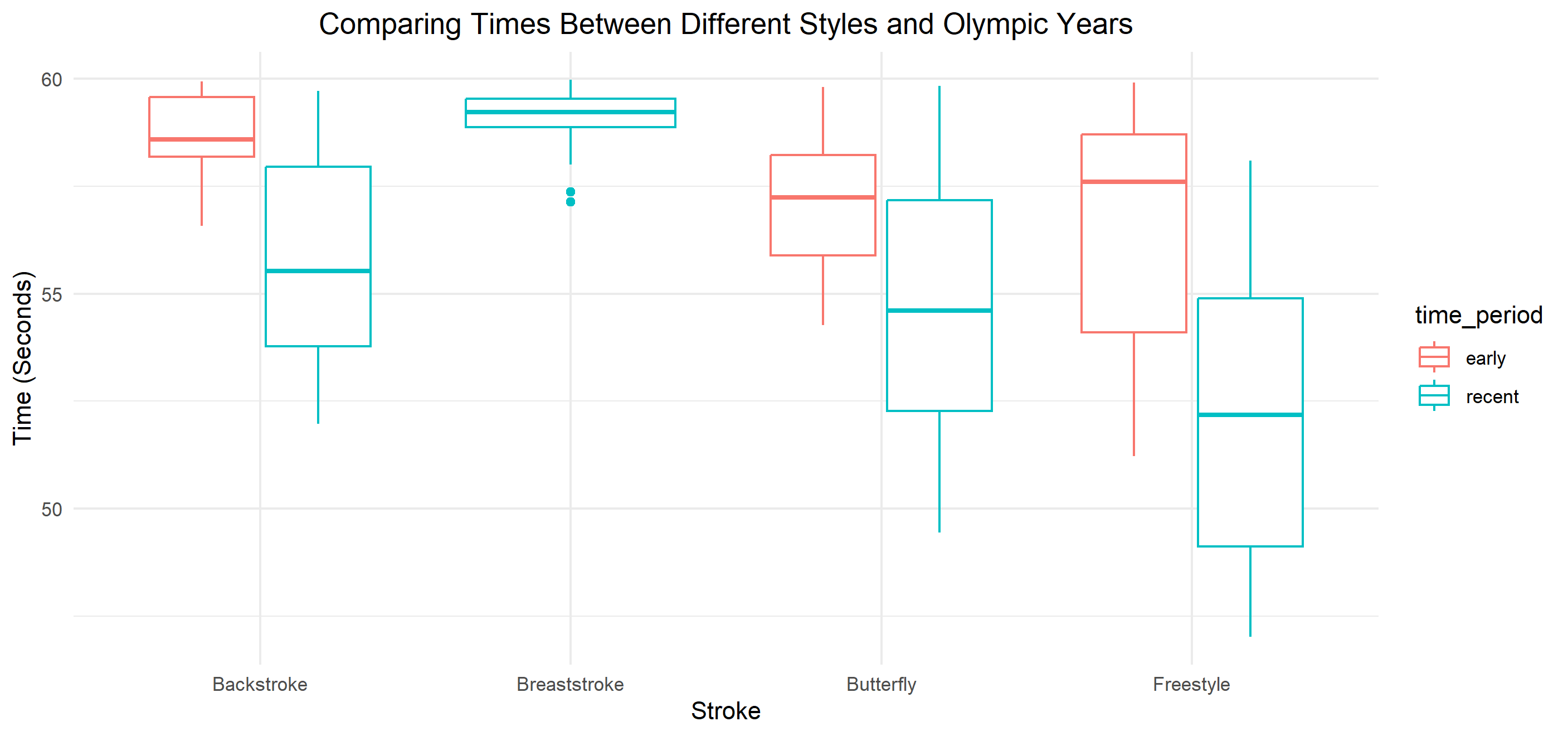
The side-by-side boxplots below display race times for different strokes in the 100-meter event. Use it to answer the questions below.



1. Interpret the side-by-side boxplots above. Be sure to compare medians and note any skew outliers, and other interesting features.

The backstroke, breaststroke, and freestyle boxplots have approximately no skew, while the butterfly boxplot has a slight positive skew towards lower event times. There are two upper (lower times) outliers under the breaststroke category. The freestyle has the lowest median time of the four styles, with butterfly, backstroke, and breaststroke following. Freestyle has the highest range and IQR, while breaststroke has the lowest range and IQR. The third quartile of the breaststroke is lower than any of the first quartiles for the other techniques.

Use the boxplot below to answer the following 2 questions.

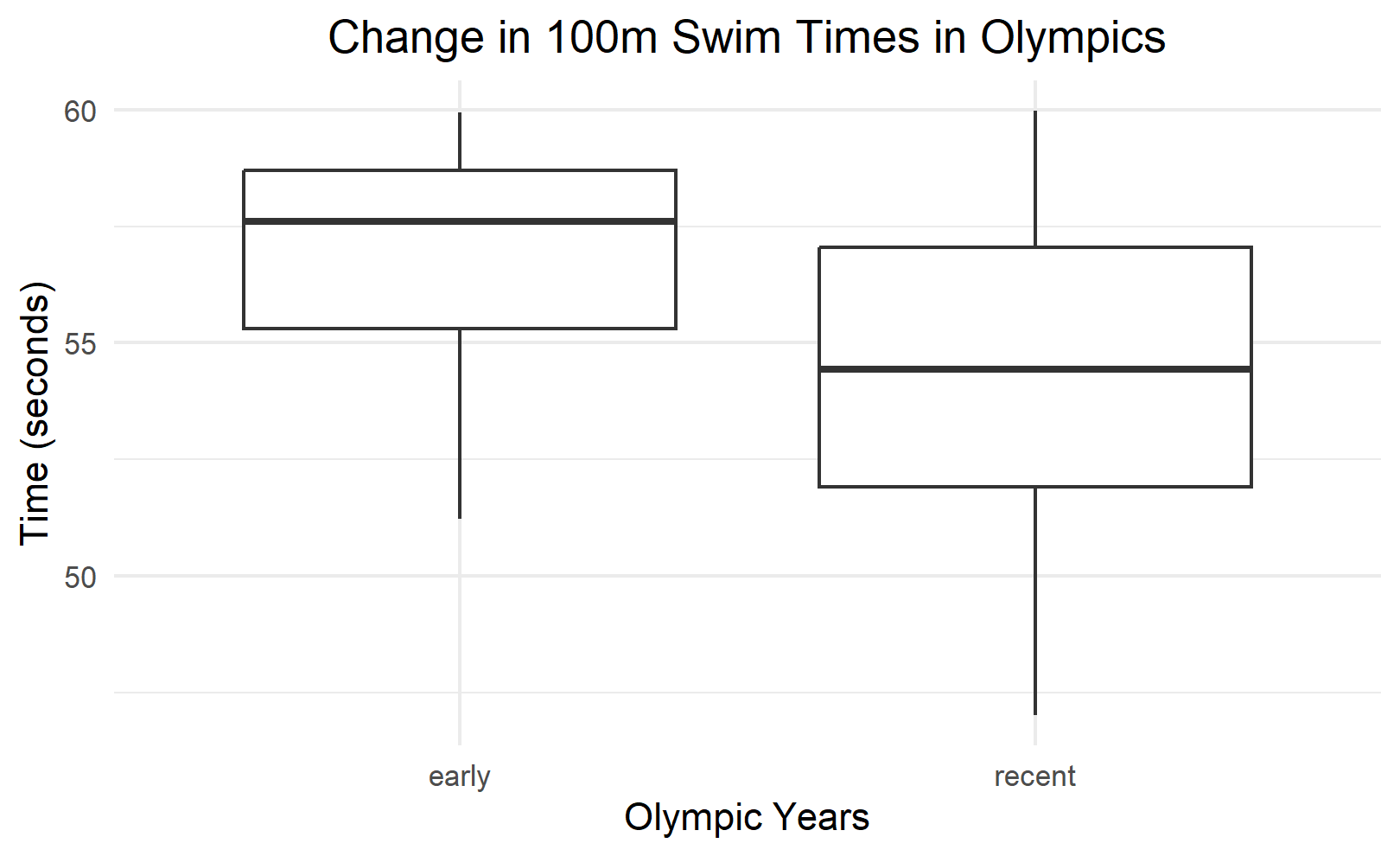


1. How do the older results compare to the more recent results from the 100m event? What changed and what remained similar?

One of the main differences is the median butterfly time is lower than the median freestyle time during the event from 1924 to 1972. Also, the introduction of the breaststroke during the more recent years may have affected how the other styles were used and practiced by the Olympians. The butterfly and freestyle remained quicker than the backstroke during both the older and more recent years. Overall, the freestyle, butterfly, and backstroke times all got quicker as the years went on.

1. How do these results compare to the overall results on the first page?

The more recent data is very similar to the overall results with the breaststroke having the same distribution and the freestyle being the fastest style followed by butterfly and backstroke. For the older Olympics, besides not having the breaststroke, the freestyle has a strong positive skew towards quicker times and the backstroke has more uneven tails in the older Olympics with a longer tail towards the lower times and a shorter tail towards the higher times.



1. Look at the boxplot above. Brainstorm some ideas on whether or not the 100m swim times got significantly faster from the early years of the event (1924-1972) to the recent years (1976-2020).

Responses will vary. Maybe mention the medians and quartiles and how recent times are quicker.

1. Let’s use a hypothesis test to examine these ideas. The mean finish time for recent Olympics is 54.27 seconds, standard deviation 3.4 and 447 recorded times. For earlier Olympics the mean finish time is 56.86 seconds with standard deviation 2.39 with 159 recorded times. Conduct a difference in means t-test to determine if there is enough evidence to conclude that Olympic swimmers have gotten faster in the 100m race. You can assume all conditions are met.

Population: n = 606, mean = 54.71, sd = 3.39

x̄1 = 54.27, S1 = 3.4, n1 = 447,x̄2 = 56.86, S2 = 2.39, n2 = 159

Null Hypothesis: x̄1 = x̄2

Alternative Hypothesis: x̄1 < x̄2

**t-statistic:**