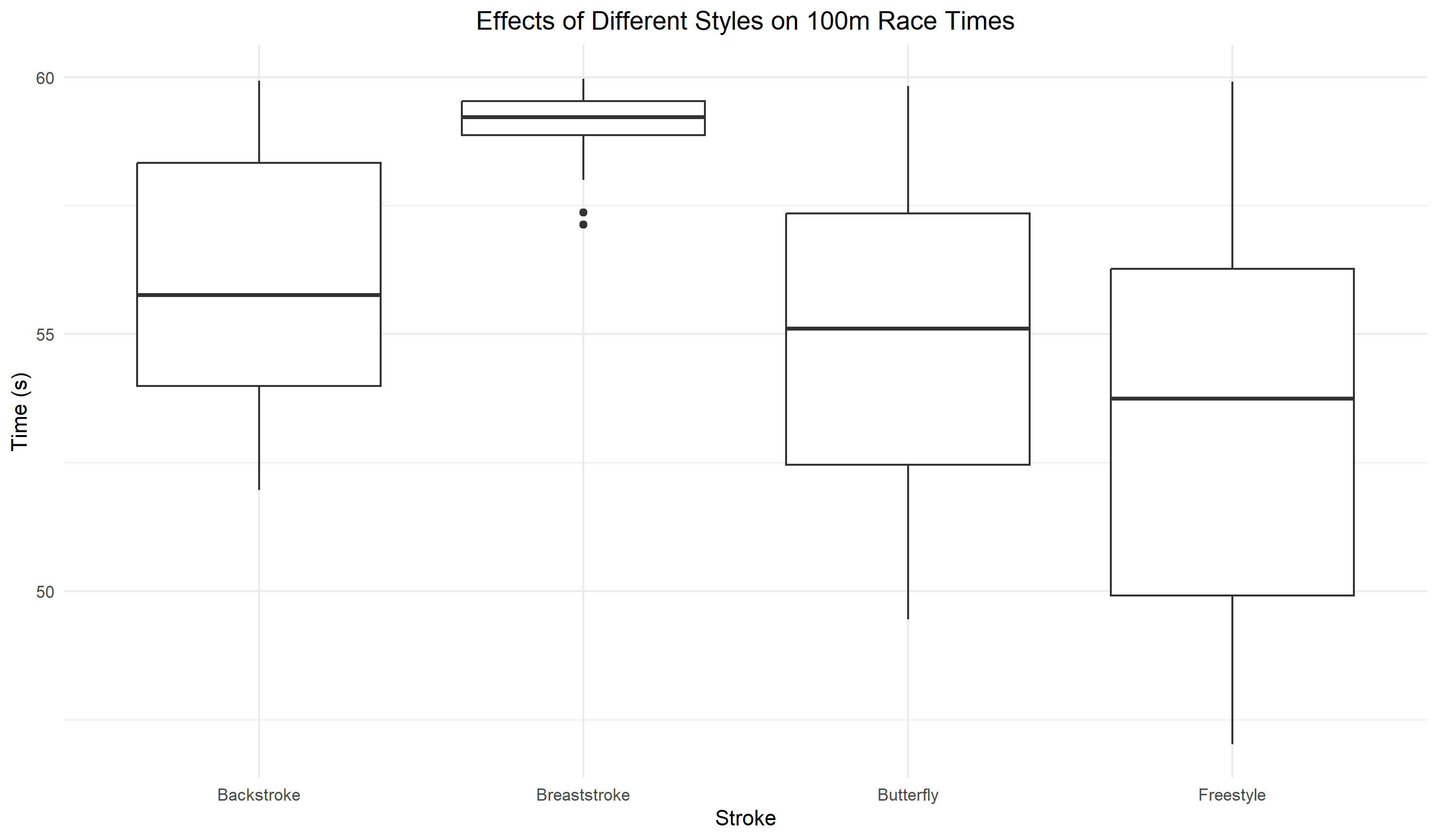
**Analyzing Olympic Swim Times with Boxplots and Hypothesis Tests**

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 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 race 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. What swimming styles seem to be the quickest?

Use the boxplot below to answer the following 2 questions.

A graph with different styles and colors

Description automatically generated with medium confidence

1. How do the older results compare to the more recent results from the 100m event? What changed and what remained similar?
2. How do these results compare to the overall results on the first page?

A graph with black lines

Description automatically generated

1. Look at the boxplot above. Brainstorm some ideas on whether the 100m swim times got significantly faster from the early years of the event (1924-1972) to the recent years (1976-2020).
2. 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. Make sure to write out your null and alternative hypothesis, find your t-stat, write the p-value, and write a conclusion.
3. Now let’s look at the strokes, or styles, of swimming. The two fastest strokes were consistently the butterfly and the freestyle. The mean finishing time for 271 freestyle finishers was 53.41 seconds with standard deviation 3.66. For 173 butterfly swimmers the mean was 55.03 seconds with standard deviation 2.82. Conduct a difference in means t-test to determine if there is enough evidence to conclude that Olympic freestyle swimmers are faster than butterfly swimmers in the 100m race. You can assume all conditions are met. Make sure to write out your null and alternative hypothesis, find your t-stat, write the p-value, and write a conclusion.
4. How do the hypothesis test conclusions compare to what you saw in the boxplots displaying this data?