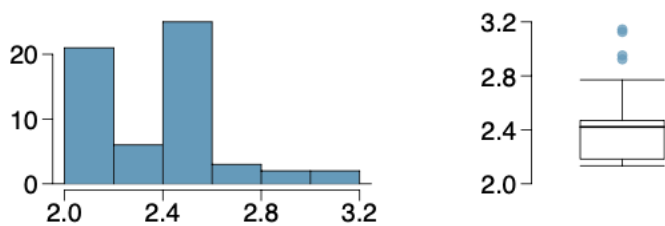


Data 606 Chapter 2 Presentation

Leticia Salazar

9/10/2021

2.34 Marathon winner. The histogram and box plots below show the distribution of finishing times for male and female winners of the New York Marathon between 1970 and 1999.



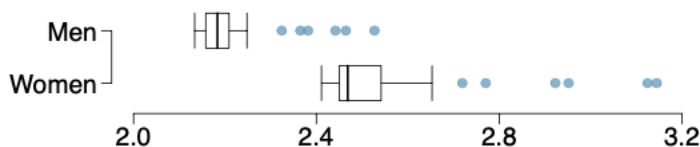
(a) What features of the distribution are apparent in the histogram and not in the box plot? What features are apparent in the box plot but not in the histogram?

- *In a histogram you get a good sense of the distribution of the data. It shows the shape of the data and you can get an overall scope of the data set. For instance you can see that there's two modes in the histogram above and that the data set is right skewed. With a box plot you get to see a summary of distribution of individual points unlike a histogram where you see grouped frequencies of the data. In the box plot above you see the median and outliers are more visible.*

(b) What may be the reason for the bimodal distribution? Explain

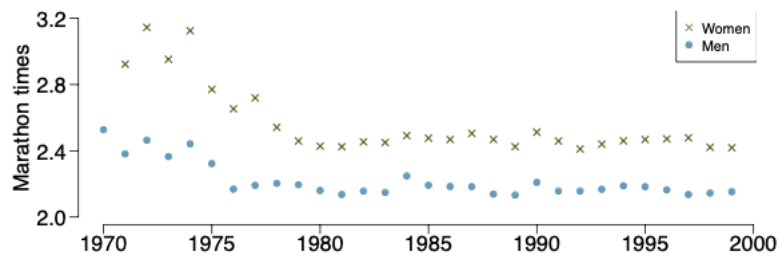
- *Since the data set is displaying the finishing times for male and female winners, the bimodal distribution is showing both of these outcomes.*

(c) Compare the distribution of marathon times for men and women based on the box plot shown below.



- *The men's finishing times are much lower than the women. The few outliers that men have end around 2.5, where the women's outliers are closer to 3.2.*

(d) The time series plot shown below is another way to look at these data. Describe what is visible in this plot but not in the others.



- *The time series plot actually shows the time from men and women and how they are dispersed throughout time. Overall, women's marathon times are much higher than the men. Yet, they both start to decrease from 1975-2000.*