4 - Critiquing data

Analyze datasets by critiquing/noting the shape, center, spread, outliers.

Example 1. Critique the data at right.

Shape:
Both distributions are skewed right: the
distance from the minimum to the median is much
smaller than the distance from the median to
the maximum.

Center:
Travel times to work are generally longer in NY
than in NC. The median, both quartiles, and the
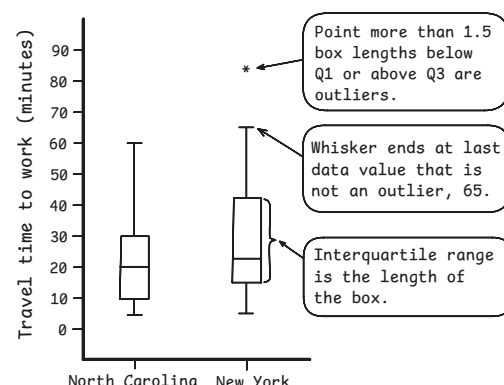maximum are all larger in NY.

Spread:
Travel times are also more variable in NY, as shown by the lengths
of the boxes (the IQR) and the range.

Outliers:
The maximum travel time of 85 minutes is an outlier for the NY data.
There are no outliers in the NC sample.

Figure. Boxplots comparing the travel times to work of samples of workers in North Carolina and New York.

Point more than 1.5 box lengths below Q1 or above Q3 are outliers.

Whisker ends at last data value that is not an outlier, 65.

Interquartile range is the length of the box.

Example 2. For their final project, some statistics students compared texting
habits of males and females at their school. They asked a random sample of
students to record the number of text messages sent and received over a two-day
period. Here are their data:

Males: 127 44 28 83 0 6 78 6 5 213 73 20 214 28 11
Females: 112 203 102 54 379 305 179 24 127 65 41 27 298 6 130 0
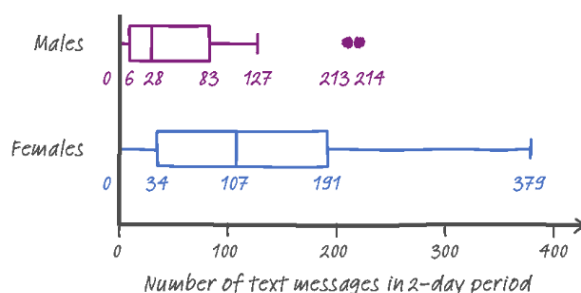
What conclusions should the students draw?

PLAN: Make boxplots. Then calculate summary
parameters. Then compare shape, center,
spread, outliers of the two distributions.

DO: Use R commands in Google Colab:

```
male <- c(127,44,28,83,0,6,78,6,5,213,73,20,214,28,11)
female <- c(112,203,102,54,379,305,179,24,127,65,41,27,298,6,130,0)
boxplot(male, female, names=c('Male', 'Female'))
summary(male)
summary(female)
```

Due to the strong skewness and outliers, we
use the median and IQR instead of the mean
and standard deviation when comparing
center and spread.

	median	mean	pop.sd
male	28	62.4	68.95
female	107	128.2	112.27

Number of text messages in 2-day period

Shape: Both distributions are strongly right-skewed.

Center: Females typically text more than males. The median number of texts for females (107) is about four times that for males (28). In fact, the median for the females is above the third quartile for the males. So over 75% of the males texted less than the "typical" (median) female.

Spread: There is much more variation in texting among the females than the males. The IQR for females (157) is about twice the IQR for males (77).

Outliers: There are two outliers in the male distribution: students who reported 213 and 214 texts in two days. The female distribution has no outliers.

CONCLUDE: This survey gives strong evidence that male and female texting habits differ considerably at the school. A typical female sends and receives about 79 more text messages in a two-day period than a typical male. The males as a group are also much more consistent in their texting frequency than the females.

First Homework:

Section 3 (Tues-Thurs class) due Thurs Feb 5.
Section 4 (Mon-Wed class) due Wed Feb 11.

Problem 1. (No R needed.)
In a 2008 article titled "Letting Our Fingers Do the Talking," the New York Times reported that Americans now send more text messages than they make phone calls. According to a study by Nielsen Mobile, "Teenagers ages 13 to 17 are by far the most prolific texters, sending or receiving 1742 messages a month." Mr. Williams, a high school statistics teacher, was skeptical about the claims in the article. So he collected data from his first-period statistics class on the number of text messages and calls they had sent or received in the past 24 hours. Here are the texting data:

0 7 1 29 25 8 5 1 25 98 9 0 26 8 118 72 0 92 52 14 3 3 44 5 42

(a) Make a boxplot of these data by hand. Be sure to check for outliers.
(b) Explain how these data seem to contradict the claim in the article.

Problem 2. (Use R in Google Colab.)
Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154 109 137 115 152 140 154 178 101 103 126 126 137 165 165 129 200 148

and for 20 first-year college men:

108 140 114 91 180 115 126 92 169 146 109 132 75 88 113 151 70 115 187 104

Do these data support the belief that men and women differ in their study habits and attitudes toward learning? (Note that high scores indicate good study habits and attitudes toward learning.) Follow the four-step process.