

4

CHAPTER

Describing and Summarizing Data from One Variable

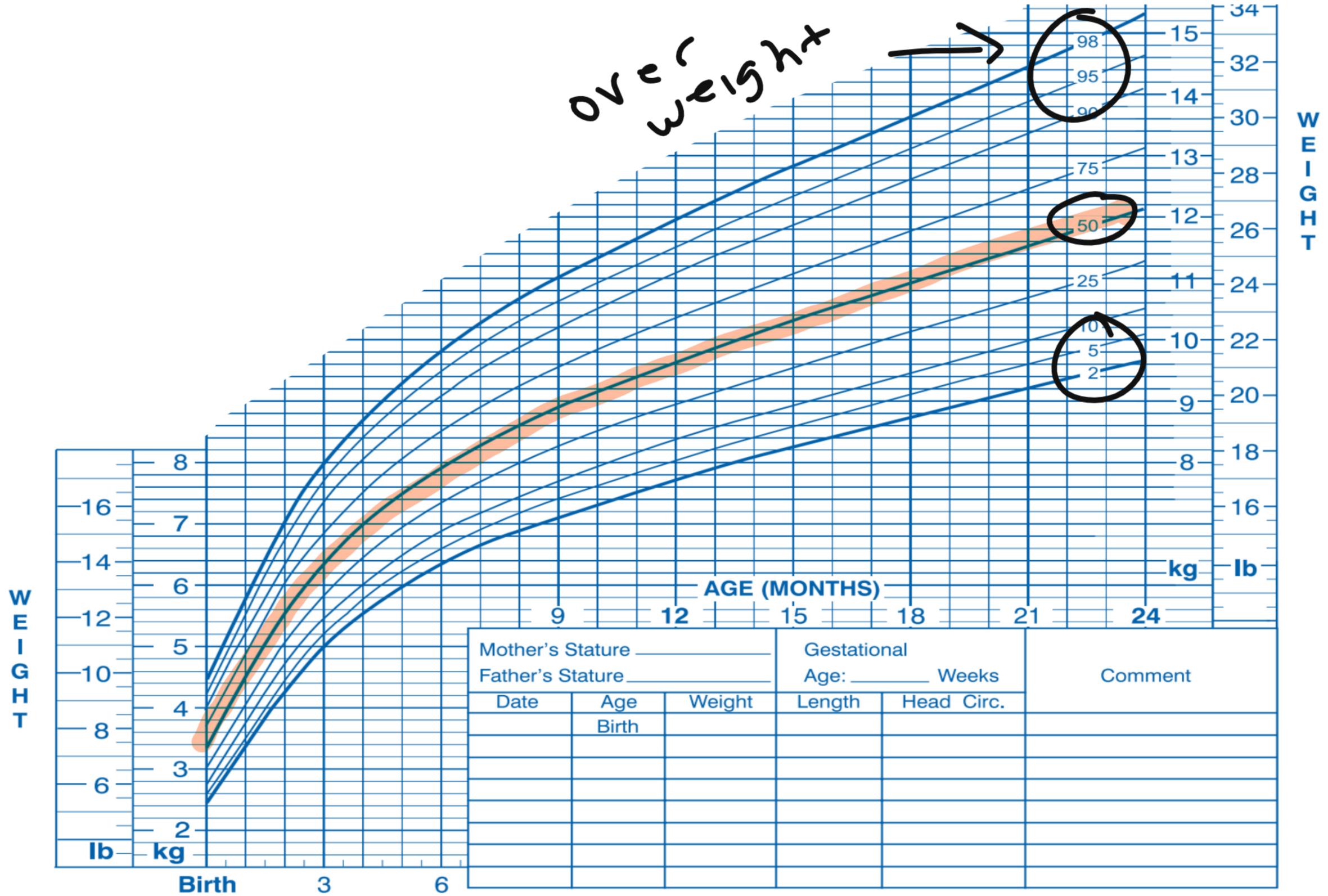
4.3 Measures of Relative Position,
Box Plots, and Outliers

4.6 Proportions and Percentages

P^{th} Percentile

Given a set of data x_1, x_2, \dots, x_n , the P^{th} **percentile** is a value, say x , such that approximately P percent of the data is less than or equal to x and approximately $(100 - P)$ percent of the data is greater than or equal to x .

DEFINITION



Published by the Centers for Disease Control and Prevention, November 1, 2009



Finding the P^{th} Percentile

Arrange the data in ascending order

To find the location of the P^{th} percentile in the ordered data, calculate,

$$l = n \left(\frac{P}{100} \right)$$

where n is the number of observations in the ordered data.

No

Is l an integer?

Yes

Average the

52%

.52

Find the 50th percentile for the following data on the number of spelling errors found on 7 pages of a web site.

3, 5, 0, 1, 9, 2, 7

0, 1, 2, 3, 5, 7, 9

$$50^{\text{th}} \rightarrow 7 \times .50 = 3.5 \rightarrow 4$$

3

$$80^{\text{th}} \rightarrow 7 \times .80 = 5.6 \rightarrow 6$$

7

Suppose that 40 members of your company are given a screening test for a new position. These scores are reported in the table on the left below. To inform potential employees of their screening test performance you may wish to report various percentiles for the test scores. Find the 10th and 88th percentiles for the test.

$$10^{th} \rightarrow 40 \times .10 = \boxed{4} \quad \textcircled{28}$$

$$88^{th} \rightarrow 40 \times .88 = 35.2$$

$$\rightarrow 36$$

$$\textcircled{73}$$

Ordered Test Scores			
18	43	54	66
21	44	55	67
21	45	55	69
$\boxed{27}$	45	56	70
$\boxed{29}$	46	57	71
31	47	58	$\textcircled{73}$
32	48	61	77
33	49	62	80
34	52	63	81
41	54	64	82

Percentile

The **percentile** of some data value x is given by

$$\text{percentile of } x = \frac{\text{number of data values less than or equal to } x}{\text{total number of data values}} \cdot 100.$$

FORMULA

Suppose that 40 members of your company are given a screening test for a new position. These scores are reported in the table on the left below. To inform potential employees of their screening test performance you may wish to report various percentiles for the test scores. Find the 10th and 88th percentiles for the test.

$$\frac{7}{40} = 0.175$$

18%

Ordered Test Scores			
18	43	54	66
21	44	55	67
21	45	55	69
27	45	56	70
29	46	57	71
31	47	58	73
32	48	61	77
33	49	62	80
34	52	63	81
41	54	64	82

Quartiles

The 25th, 50th, and 75th percentiles are known as **quartiles** and are denoted as Q_1 , Q_2 , and Q_3 respectively.

DEFINITION

25th → Q_1

50th → Q_2
Median

75th → Q_3

Interquartile Range

The **interquartile range** is a measure of dispersion which describes the range of the middle fifty percent of the data. It is calculated as follows.

$$\text{IQR} = Q_3 - Q_1$$

FORMULA

Suppose that 40 members of your company are given a screening test for a new position. These scores are reported in the table on the left below. To inform potential employees of their screening test performance you may wish to report various percentiles for the test scores. Find the 10th and 88th percentiles for the test.

$$25^{th} \rightarrow Q_1 \quad 40 \times .25 = 10$$

$$50^{th} \rightarrow \text{Median} \quad 40 \times .5 = 20$$

$$75^{th} \rightarrow Q_3 \quad 40 \times .75 = 30$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 65 - 42 = 23 \end{aligned}$$

Ordered Test Scores			
18	<u>43</u>	54	66
21	44	55	67
21	45	55	69
27	45	56	70
29	46	57	71
31	47	58	73
32	48	61	77
33	49	62	80
34	52	63	81
<u>41</u>	54	64	82

$$\text{Min} = 18$$

$$Q_1 = 42$$

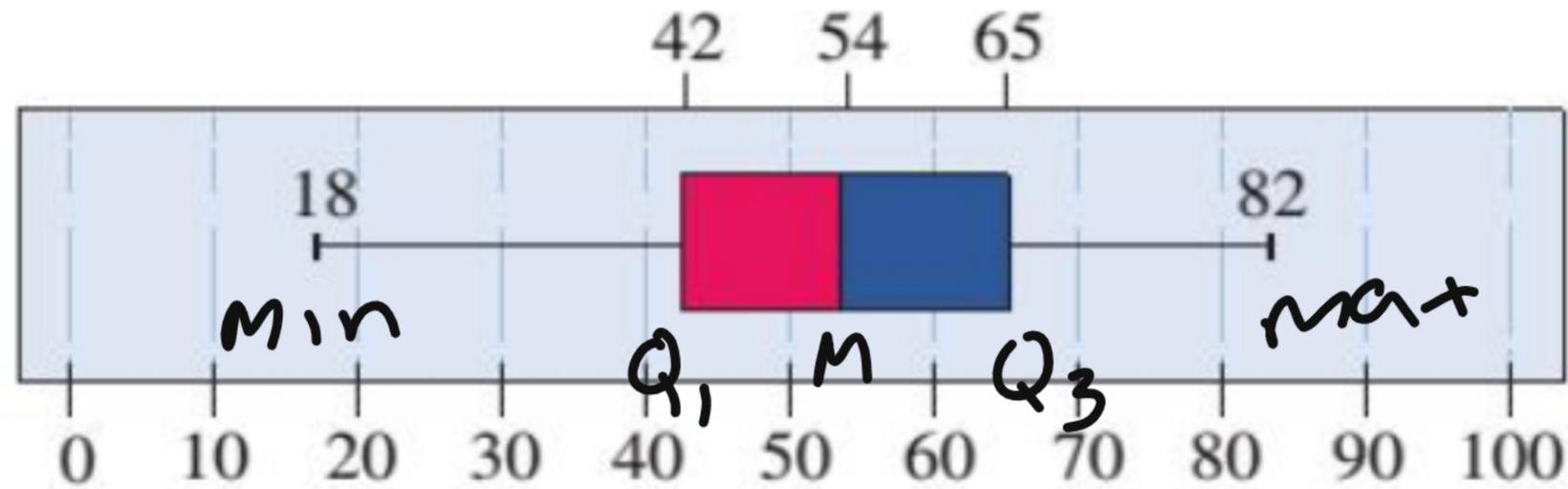
$$\text{Median} = 54$$

$$Q_3 = 65$$

$$\text{Max} = 82$$

$$\begin{aligned} \text{Range} \\ 82 - 18 = 64 \end{aligned}$$

Box Plot of Screening Test Scores



5-Number Summary

For a set of data, the 5-number summary consists of the following five values:

1. Minimum
2. First quartile, Q_1
3. Second quartile, Q_2 , or the median
4. Third quartile, Q_3
5. Maximum

DEFINITION

Outlier

A data point is considered an **outlier** if it is 1.5 times the interquartile range above the 75th percentile or 1.5 times the interquartile range below the 25th percentile.

DEFINITION

$$\begin{array}{l}
 \min = 18 \quad Q_1 = 42 \quad M = 54 \quad Q_3 = 65 \quad \max = 82 \\
 IQR = 23 \quad \text{Outlier} \quad 1.5 \times 23 = 34.5 \rightarrow 35 \\
 Q_1 - 35 = 7 \quad Q_3 + 35 = 100 \\
 \leftarrow 7 \quad \text{Outliers} \quad 100 \rightarrow
 \end{array}$$

- e. Explain the meaning of these percentiles in the context of the on-base percentages.
- f. Calculate the interquartile range.
- g. Construct a box plot for the on-base percentages. Are there any outliers?

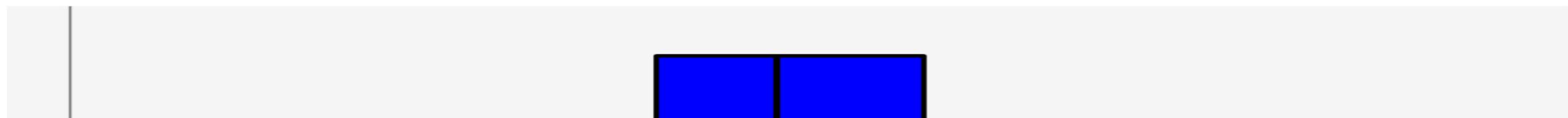
Min	Q1	Median	Q3	Max
0.277	0.317	0.326	0.337	0.373

$$IQR = 0.337 - 0.317 = 0.020$$

$$0.020 \times 1.5 = 0.030$$

$$0.317 - 0.030 = 0.287$$

$$0.337 + 0.030 = 0.367$$



g. Construct a box plot for the on-base percentages. Are there any outliers?

Min	Q1	Median	Q3	Max
0.277	0.317	0.326	0.337	0.373

IQR =

$$IQR = 0.337 - 0.317 = 0.020$$

$$\text{Outlier } 0.020 \times 1.5 = 0.030$$

$$Q_1 - 0.030 = 0.287$$

$$Q_3 + 0.030 = 0.367$$

US Violent Crime Rate (per 100,000 Residents) 2010-2014

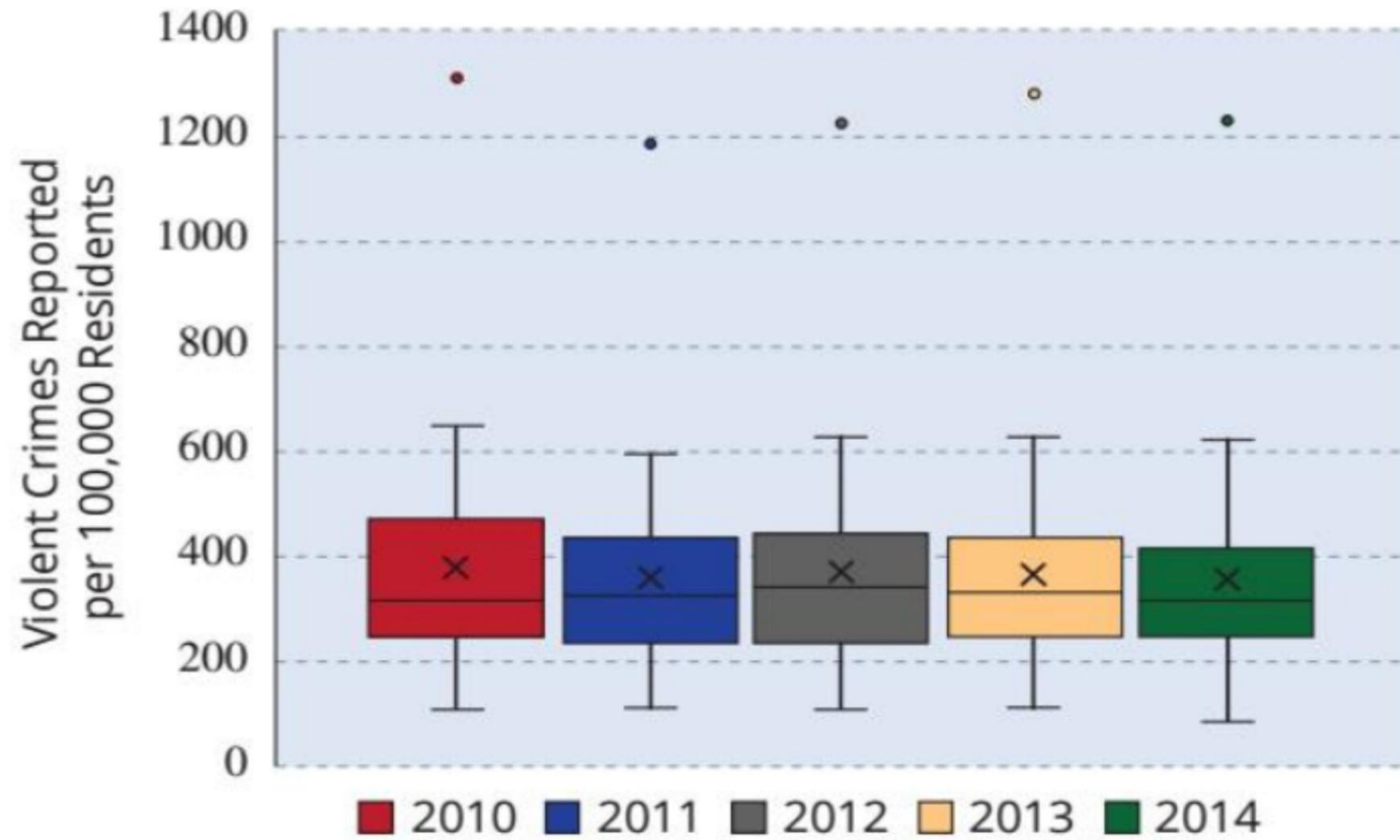
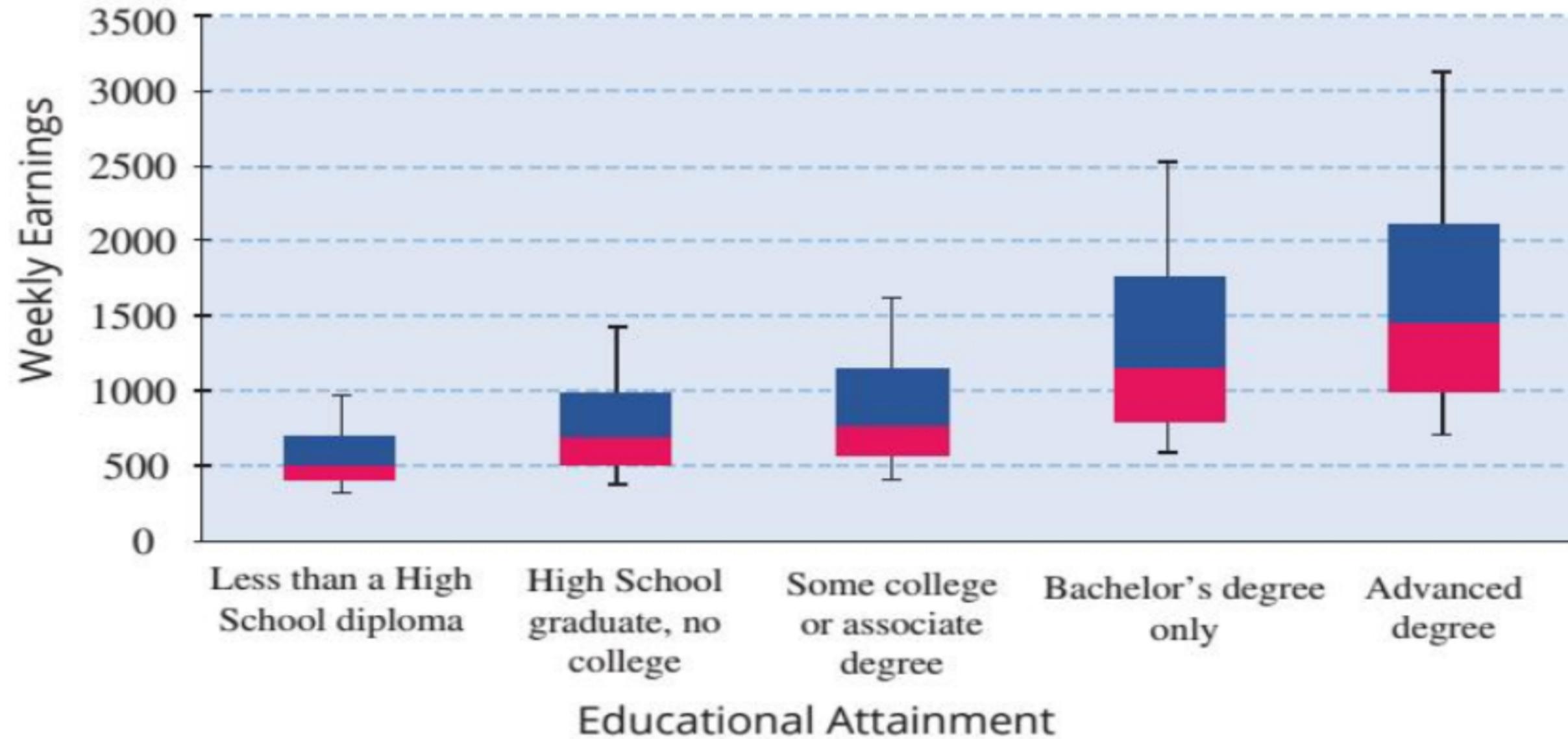


Figure 4.3.5

Education and Earnings 2016



Proportion

A **proportion** measures the fraction of a group that possesses some characteristic.

DEFINITION

Suppose your statistics class is composed of 48 students of which 4 are left-handed.

What proportion of the class is left-handed?

$$\frac{4}{48} = 0.0833$$
$$8.3\%$$

According to a study administered by the National Bureau of Economic Research, half of Americans would struggle to come up with \$2000 in the event of a financial emergency. The majority of the 1900 Americans surveyed said they would rely on more than one method to come up with emergency funds if required. In the survey, 532 people said that they “certainly” would not be able to cope with an unexpected \$2000 bill if they had to come up with the money in 30 days, and 418 people said they “probably” would not be able to cope.²

- a.** What percentage of Americans “certainly” would not be able to produce \$2000 in the event of an emergency according to the study?
- b.** What percentage of Americans would “probably” not be able to pay a \$2000 bill in 30 days if required?
- c.** What does this say about the savings habits of Americans?

