1. **[7.71/7.71 Points]** DETAILS PREVIOUS ANSWERS **ASWSBE14 2.E.025.ALT.**

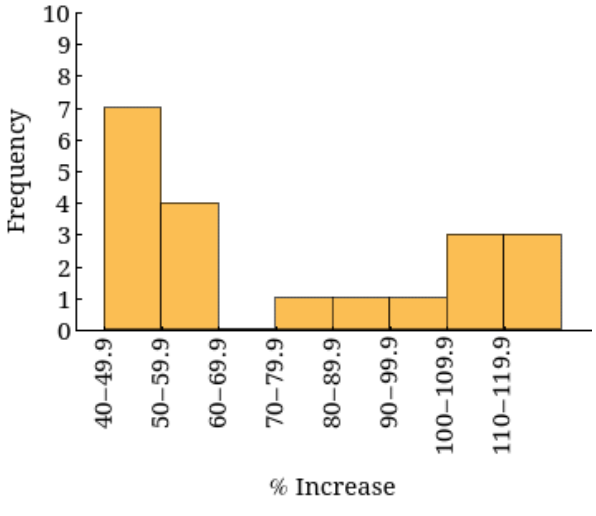
MY NOTES ASK YOUR TEACHER

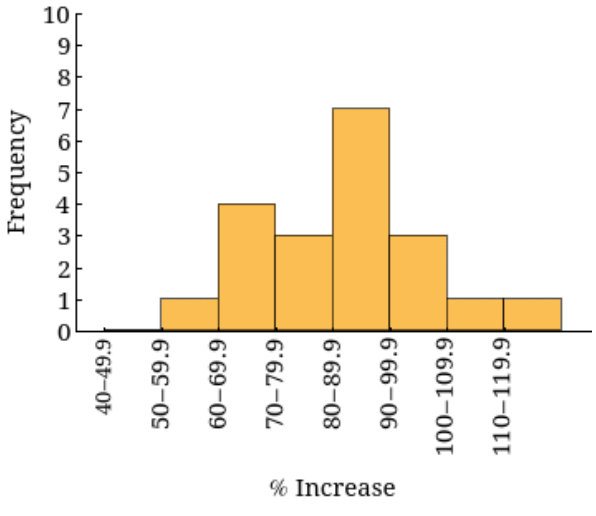
DATAfile: BestPayingDegrees

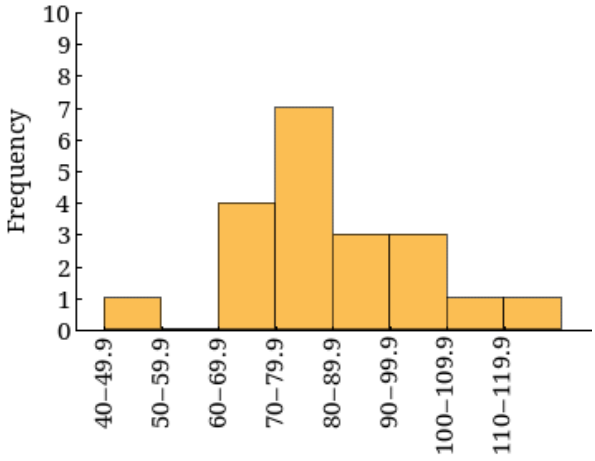
Each year America.EDU ranks the best paying college degrees in America. The following data show the median starting salary, the mid-career salary, and the percentage increase from starting salary to mid-career salary for the 20 college degrees with the highest mid-career salary.†

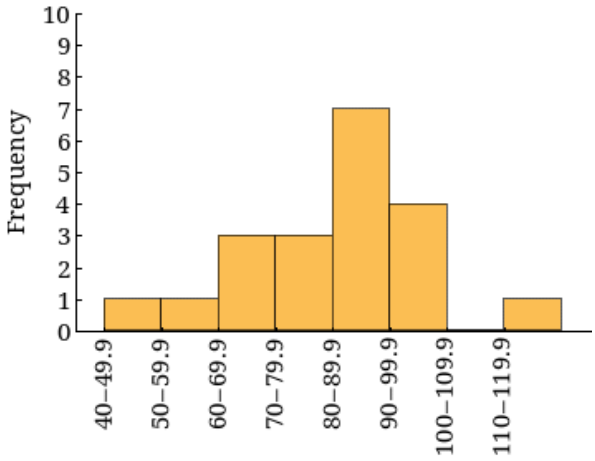
Degree	Starting Salary	Mid-Career Salary	% Increase
Aerospace engineering	59,400	108,000	82
Applied mathematics	56,400	101,000	79
Biomedical engineering	54,800	101,000	84
Chemical engineering	64,800	108,000	67
Civil engineering	53,500	93,400	75
Computer engineering	61,200	87,700	43
Computer science	56,200	97,700	74
Construction management	50,400	87,000	73
Economics	48,800	97,800	100
Electrical engineering	60,800	104,000	71
Finance	47,500	91,500	93
Government	41,500	88,300	113
Information systems	49,300	87,100	77
Management info. systems	50,900	90,300	77
Mathematics	46,400	88,300	90
Nuclear engineering	63,900	104,000	63
Petroleum engineering	93,000	157,000	69
Physics	50,700	99,600	96
Software engineering	56,700	91,300	61
Statistics	50,000	93,400	87

(a) Using a class width of 10, construct a histogram for the percentage increase in the starting salary.


Frequency
0
1
2
3
4
5
6
7
8
9
10
40–49.9
50–59.9
60–69.9
70–79.9
80–89.9
90–99.9
100–109.9
110–119.9
% Increase


Frequency
0
1
2
3
4
5
6
7
8
9
10
40–49.9
50–59.9
60–69.9
70–79.9
80–89.9
90–99.9
100–109.9
110–119.9
% Increase





(b) Comment on the shape of the distribution.

The histogram is skewed to the right .

(c) Develop a stem-and-leaf display for the percentage increase in the starting salary. (Enter numbers from smallest to largest separated by spaces. Enter NONE for stems with no values.)

Leaf Unit = 1

4	3
5	NONE
6	1 3 7 9
7	1 3 4 5 7 7 9
8	2 4 7
9	0 3 6
10	0
11	3

(d) What are the primary advantages of the stem-and-leaf display as compared to the histogram? (Select all that apply.)

- [ ] stem-and-leaf display is hard to read
- [x] the stem-and-leaf display is easier to construct by hand
- [ ] stem and leaf display provides exactly the same information as a histogram
- [x] stem-and-leaf display shows the actual data
- [ ] stem and leaf display provides less information than a histogram

Need Help? Read It

2. [7.69/7.69 Points] DETAILS PREVIOUS ANSWERS ASWSBE14 2.E.028.MI.SA.

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

This question has several parts that must be completed sequentially. If you skip a part of the question, you will not receive any points for the skipped part, and you will not be able to come back to the skipped part.

Tutorial Exercise

The following observations are for two quantitative variables, *x* and *y*.

Observation	*x*	*y*	Observation	*x*	*y*
1	57	35	11	18	72
2	88	21	12	76	24
3	57	56	13	35	68
4	78	22	14	43	71
5	55	44	15	66	53
6	25	94	16	37	62
7	77	23	17	12	83
8	12	73	18	38	22
9	21	93	19	37	66
10	83	34	20	72	34

(a) Develop a crosstabulation for the data, with *x* as the row variable and *y* as the column variable. For *x* use classes of 10–29, 30–49, and so on; for *y* use classes of 20–39, 40–59, and so on.

(b) Compute the row percentages.

(c) Compute the column percentages.

(d) What is the relationship, if any, between *x* and *y*?

Step 1

(a) Develop a crosstabulation for the data, with *x* as the row variable and *y* as the column variable. For *x* use classes of 10–29, 30–49, and so on; for *y* use classes of 20–39, 40–59, and so on.

A crosstabulation displays the frequency of data as it occurs in particular categories. The types of data displayed can be either categorical or quantitative. If the data is quantitative, ranges will often be used instead of each numerical value from the given data. Be careful when determining the ranges for use within a crosstabulation. If the ranges are too large or too small, then this display will not be meaningful.

Here we are given quantitative data for variables *x* and *y*. We are also given specific classes to use. The crosstabulation format is given below. The rows correspond to *x* values in the specified classes 10–29, 30–49, etc., and the columns correspond to *y*

values in the specified classes 20–39, 40–59, etc. The Grand Total column will be the sums of the previous columns, and the Grand Total Row will be the sums of the above rows.

		y				Grand Total
		20–39	40–59	60–79	80–99	
x	10–29					
	30–49					
	50–69					
	70–90					
Grand Total						

Here are the data.

Observation	*x*	*y*
1	57	35
2	88	21
3	57	56
4	78	22
5	55	44
6	25	94
7	77	23
8	12	73
9	21	93
10	83	34

Observation	*x*	*y*
11	18	72
12	76	24
13	35	68
14	43	71
15	66	53
16	37	62
17	12	83
18	38	22
19	37	66
20	72	34

To fill in the table, first determine how many observations are in each *x* class. Place this value in the corresponding row of the Grand Total column. Of these, determine how many are in the corresponding classes for *y*.

Observations 6, 8, 9, 11, and 17 have *x* values between 10 and 29, so a 5 will be placed in the first row of the Grand Total column. Of these 5 observations, there are none with *y* values between 20 and 39, or 40 and 59. There are 2 observations with *y*-values between 60 and 79. There are 3 observations with *y*-values between 80 and 99. Therefore 2 and 3 will be placed in the first row of the 60–79 and 80–99 columns, respectively.

		y				Grand Total
		20–39	40–59	60–79	80–99	
x	10–29			2 ✔	3 ✔	5 ✔
	30–49					

	50–69					
	70–90					
Grand Total						

Step 2

Use the given data to complete the table.

Observation	*x*	*y*	Observation	*x*	*y*
1	57	35	11	18	72
2	88	21	12	76	24
3	57	56	13	35	68
4	78	22	14	43	71
5	55	44	15	66	53
6	25	94	16	37	62
7	77	23	17	12	83
8	12	73	18	38	22
9	21	93	19	37	66
10	83	34	20	72	34

		y				Grand Total
		20–39	**40–59**	**60–79**	**80–99**	
x	**10–29**			2	3	5
	30–49	1 ✔		4		5
	50–69	1	3 ✔			4 ✔
	70–90	6				6
Grand Total		8 ✔	3	6 ✔	3	20 ✔

Step 3

(b) Compute the row percentages.

The frequencies from the crosstabulation will be used to compute the row percentages. Like the crosstabulation, there will be a Grand Total column. The percentages for each row should sum to 100, but there may be discrepancies due to rounding.

The crosstabulation from part (a) is given below.

		y				Grand Total
		20–39	40–59	60–79	80–99	
x	10–29			2	3	5
	30–49	1		4		5
	50–69	1	3			4
	70–90	6				6
Grand Total		8	3	6	3	20

For the first row with x values between 10 and 29, note that there are 5 observations from the data. Two of these 5 have y values between 60 and 79. Therefore, $\frac{2}{5}$, or written as a percentage $\boxed{40}$ ✔ %, of the data in the first row has an x value between 10 and 29 and a y value between 60 and 79. The remaining $\boxed{60}$ ✔ % of the data in the first row has a y value between 80 and 99.

		y				Grand Total
		20–39	40–59	60–79	80–99	
x	10–29			$\text{percentage} = \frac{\text{cell value}}{\text{row total}} \cdot 100\% = \frac{2}{5} \cdot 100\% = \boxed{40}$ ✔	$\text{percentage} = \frac{\text{cell value}}{\text{row total}} \cdot 100\% = \frac{3}{5} \cdot 100\% = \boxed{60}$ ✔	100
	30–49					
	50–69					
	70–90					

Step 4

Recall the crosstabulation.

		y				Grand Total
		20–39	40–59	60–79	80–99	
x	10–29			2	3	5
	30–49	1		4		5
	50–69	1	3			4
	70–90	6				6
Grand Total		8	3	6	3	20

Complete the table of row percentages.

	y			Grand

		20–39	40–59	60–79	80–99	Total
x	10–29			40.0	60.0	100
	30–49	20.0		80 ✔		100
	50–69	25 ✔	75.0			100 ✔
	70–90	100 ✔				100

Step 5

(c) Compute the column percentages.

Column percentages will be computed in a similar manner as row percentages. Instead of a final column for the Grand Total of the row percentages, a final row will be used to display the Grand Total.

The crosstabulation from part (a) is given below.

		y				Grand Total
		20–39	40–59	60–79	80–99	
x	10–29			2	3	5
	30–49	1		4		5
	50–69	1	3			4
	70–90	6				6
Grand Total		8	3	6	3	20

In the first column with y values between 20 and 39, note that there are 8 total observations. One of these 8 has an x value between 30 and 49. Therefore, $\frac{1}{8}$, or 12.5 ✔ %, of the data in the first column has a y value between 20 and 39 and an x value between 30 and 49.

Moving to the next row, there is 1 value that has a y value between 20 and 39 and an x value between 50 and 69. This represents 12.5% of the data.

The final row has 6 values with y between 20 and 39 and x between 70 and 90. This represents 75 ✔ % of the data in this column.

		y			
		20–39	40–59	60–79	80–99
x	10–29				
	30–49	$\text{percentage} = \frac{\text{cell value}}{\text{column total}} \cdot 100\% = \frac{1}{8} \cdot 100\% = 12.5$ ✔			
	50–69	12.5			

	70–90	$\text{percentage} = \frac{\text{cell value}}{\text{column total}} \cdot 100\% = \frac{\boxed{6}}{8} \cdot 100\% = \boxed{75}$ ✓			
Grand Total		100 ✓			

Step 6

Recall the crosstabulation.

		y				**Grand Total**
		20–39	**40–59**	**60–79**	**80–99**	
x	**10–29**			2	3	5
	30–49	1		4		5
	50–69	1	3			4
	70–90	6				6
Grand Total		8	3	6	3	20

Complete the table of column percentages. If necessary, round your answers to one decimal place.

		y			
		20–39	**40–59**	**60–79**	**80–99**
x	**10–29**			33.3 ✓	100 ✓
	30–49	12.5		66.7	
	50–69	12.5	100.0		
	70–90	75.0			
Grand Total		100	100 ✓	100 ✓	100

Step 7

(d) What is the relationship, if any, between *x* and *y*?

The crosstabulation, table of row percentages, or table of column percentages can be used to determine if there is a relationship between *x* and *y*. Since we are investigating any type of relationship between *x* and *y*, we will use the raw data in the crosstabulation.

		y				**Grand Total**
		20–39	**40–59**	**60–79**	**80–99**	
x	**10–29**			2	3	5

	30–49	1		4		5
	50–69	1	3			4
	70–90	6				6
Grand Total		8	3	6	3	20

First observe that in the top two rows (lower values of x), that is for x-values between 10 and 49, 9 of the 10 corresponding y-values are between 60 and 99 (higher values of y). This tells us that lower values of x correspond to higher ✔ values of y.

Next, observe that in the bottom two rows, that is for x-values between 50 and 90 (higher values of x), all of the corresponding y-values are between 20 and 59 (lower values of y). This tell us that higher values of x correspond to lower ✔ values of y.

You have now completed the Master It.

Need Help? Read It

3. **[7.57/7.68000000000001 Points]** DETAILS PREVIOUS ANSWERS **ASWSBE14 2.E.035.**

MY NOTES ASK YOUR TEACHER

DATAfile: FuelData2018

The U.S. Department of Energy's Fuel Economy Guide provides fuel efficiency data for cars and trucks.† A portion of the data from 2018 for 341 compact, midsize, and large cars is shown in the following table.

Car	Size	Displacement	Cylinders	Drive	Fuel Type	City MPG	Hwy MPG
1	Compact	1.4	4	F	R	27	40
2	Compact	1.4	4	F	R	27	35
3	Compact	1.4	4	F	R	28	38
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
190	Compact	2.5	4	F	R	27	36
191	Large	2.5	4	F	R	22	30
192	Midsize	2.5	4	F	R	21	32
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
339	Large	6.0	12	R	P	13	21
340	Large	6.0	12	R	P	13	22
341	Large	6.0	12	R	P	13	20

The data set contains the following variables.

- Size: Compact, Midsize, and Large
- Displacement: Engine size in liters
- Cylinders: Number of cylinders in the engine
- Drive: All wheel (A), front wheel (F), and rear wheel (R)
- Fuel Type: Premium (P) or regular (R) fuel
- City MPG: fuel efficiency rating for city driving in terms of miles per gallon
- Hwy MPG: fuel efficiency rating for highway driving in terms of miles per gallon

The complete data set is contained in the file named FuelData2018.

(a) Prepare a crosstabulation of the data on Size (rows) and Hwy MPG (columns). Use classes of 20–24, 25–29, 30–34, 35–39, and 40–44 for Hwy MPG.

Size	Hwy MPG					Total
	20–24	25–29	30–34	35–39	40–44	
Compact	13 ✔	25 ✔	49 ✔	29 ✔	6 ✔	122 ✔
Large	10 ✔	31 ✔	19 ✔	11 ✔	1 ✔	72 ✔
Midsize	15 ✔	35 ✔	61 ✔	29 ✔	7 ✔	147 ✔

Total	38 ✓	91 ✓	129 ✓	69 ✓	14 ✓	341 ✓

(b) Comment on the relationship between Size and Hwy MPG.

The least efficient size seems to be Large ✓ .

(c) Prepare a crosstabulation of the data on Drive (rows) and City MPG (columns). Use classes of 10–14, 15–19, 20–24, 25–29, and 30–34 for City MPG.

Drive	City MPG					Total
	10–14	15–19	20–24	25–29	30–34	
A	3 ✓	43 ✓	57 ✓	5 ✓		108 ✓
F		8 ✓	48 ✓	82 ✓	16 ✓	157 ✗
R	10 ✓	33 ✓	32 ✓	4 ✓		79 ✓
Total	13 ✓	84 ✓	137 ✓	91 ✓	16 ✓	341 ✓

(d) Comment on the relationship between Drive and City MPG.

Higher fuel efficiencies are associated with front ✓ wheel drive cars.

(e) Prepare a crosstabulation of the data on Fuel Type (rows) and City MPG (columns). Use classes of 10–14, 15–19, 20–24, 25–29, and 30–34 for City MPG.

Fuel Type	City MPG					Total
	10–14	15–19	20–24	25–29	30–34	
P	13 ✓	58 ✓	94 ✓	16 ✓	1 ✓	182 ✓
R		26 ✓	43 ✓	75 ✓	15 ✓	159 ✓
Total	13 ✓	84 ✓	137 ✓	91 ✓	16 ✓	341 ✓

(f) Comment on the relationship between Fuel Type and City MPG.

Higher fuel efficiencies are associated with cars that use regular ✓ gas.

Need Help? Read It

4. **[7.69/7.69 Points]** DETAILS PREVIOUS ANSWERS **ASWSBE14 2.E.045.**

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

Suppose the table below contains the median household income for a family with two earners for each of the fifty states.

State	Median Income (000's)
Alabama	76.2
Alaska	98.4
Arizona	79.7
Arkansas	70.9
California	91.2
Colorado	89.3
Connecticut	107.5
Delaware	89.9
Florida	75.5
Georgia	79.7
Hawaii	89.7
Idaho	67.1
Illinois	89.7
Indiana	76.7
Iowa	81.3
Kansas	80
Kentucky	76.4

State	Median Income (000's)
Louisiana	82.6
Maine	77.8
Maryland	108.5
Massachusetts	106.8
Michigan	81.0
Minnesota	90.1
Mississippi	70.9
Missouri	77.0
Montana	73.6
Nebraska	78.3
Nevada	74.9
New Hampshire	93.9
New Jersey	110.7
New Mexico	77.6
New York	95.2
North Carolina	76.5
North Dakota	87.0

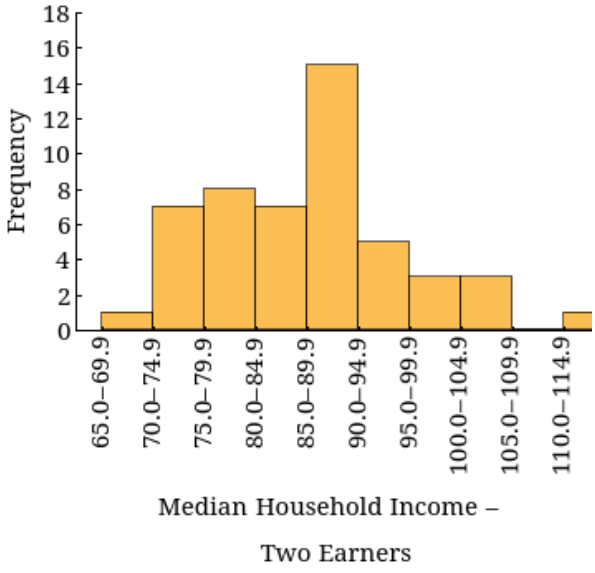
State	Median Income (000's)
Ohio	80.9
Oklahoma	74.5
Oregon	78.7
Pennsylvania	86.8
Rhode Island	94.9
South Carolina	77.1
South Dakota	72.0

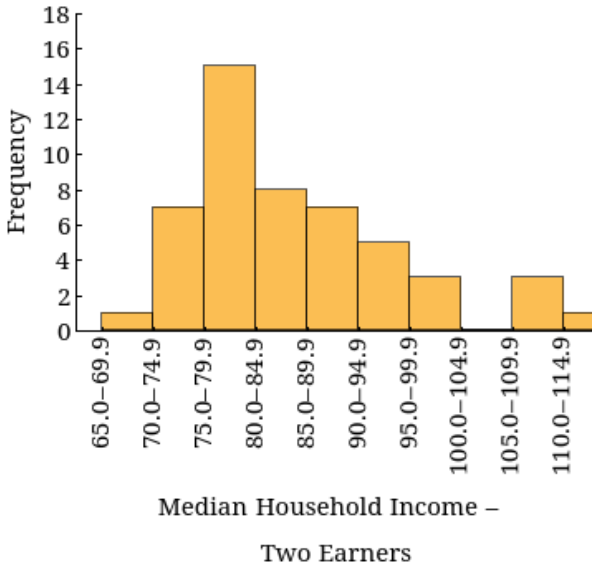
Tennessee	73.4
Texas	82.0
Utah	75.0
Vermont	83.1
Virginia	97.2
Washington	91.6
West Virginia	76.8
Wisconsin	82.3
Wyoming	87.9

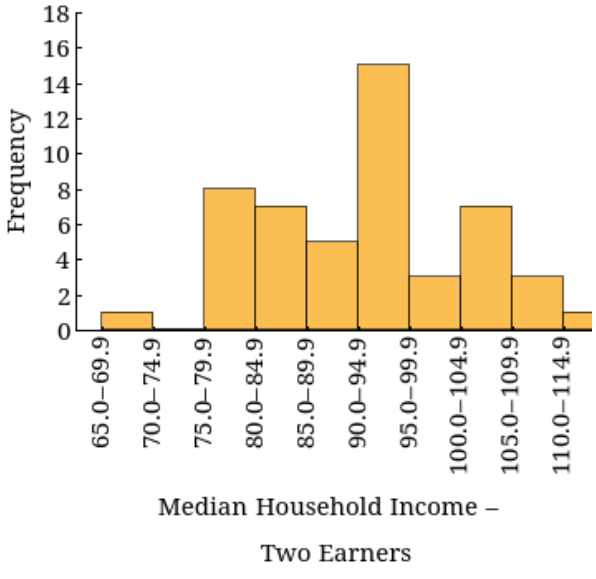
(a) Construct a frequency and a percent frequency distribution of median household income. Begin the first class at 65.0 and use a class width of 5.

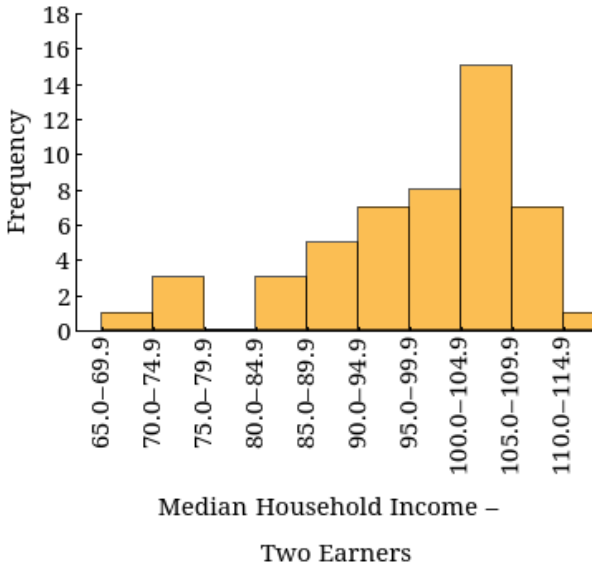
Median Household Income	Frequency	Percent Frequency
65.0–69.9	1 ✔	2 ✔ %
70.0–74.9	7 ✔	14 ✔ %
75.0–79.9	15 ✔	30 ✔ %
80.0–84.9	8 ✔	16 ✔ %
85.0–89.9	7 ✔	14 ✔ %
90.0–94.9	5 ✔	10 ✔ %
95.0–99.9	3 ✔	6 ✔ %
100.0–104.9	0 ✔	0 ✔ %
105.0–109.9	3 ✔	6 ✔ %
110.0–114.9	1 ✔	2 ✔ %
Total	50 ✔	100 ✔ %

(b) Construct a histogram.


Frequency
18
16
14
12
10
8
6
4
2
0
65.0–69.9
70.0–74.9
75.0–79.9
80.0–84.9
85.0–89.9
90.0–94.9
95.0–99.9
100.0–104.9
105.0–109.9
110.0–114.9
Median Household Income – Two Earners


Frequency
18
16
14
12
10
8
6
4
2
0
65.0–69.9
70.0–74.9
75.0–79.9
80.0–84.9
85.0–89.9
90.0–94.9
95.0–99.9
100.0–104.9
105.0–109.9
110.0–114.9
Median Household Income – Two Earners





(c) Comment on the shape of the distribution.

The distribution is [skewed to the right] ✓ . There is a gap in the [$100.0–$104.9] ✓ thousand range. The most frequent range for the median household income is [$75.0–$79.9] ✓ thousand.

(d) Which state has the highest median income for two-earner households?

- (●) New Jersey
- ( ) Maryland
- ( ) Arkansas
- ( ) Idaho

✓

(e) Which state has the lowest median income for two-earner households?

- ○ New Jersey
- ◉ Idaho
- ○ Maryland
- ○ Arkansas

Need Help? Read It

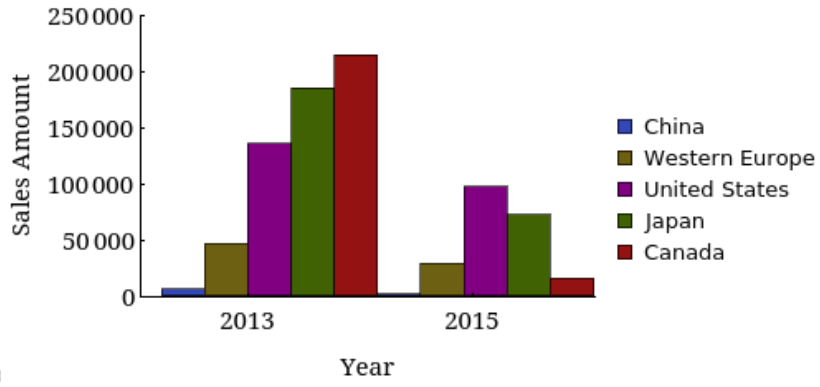
5. [7.69/7.69 Points] DETAILS PREVIOUS ANSWERS ASWSBE14 2.E.057.

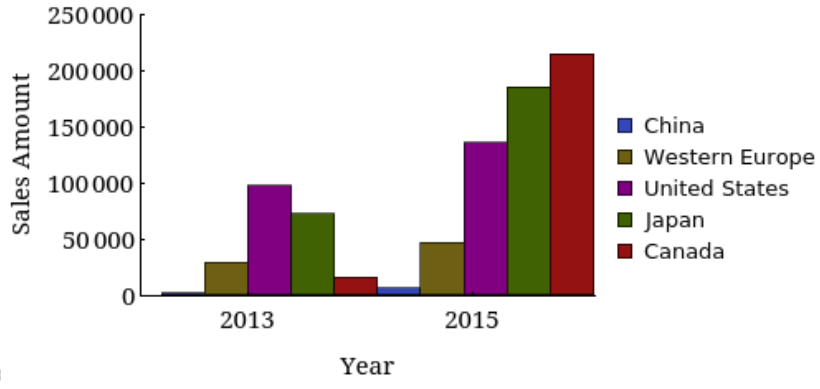
MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

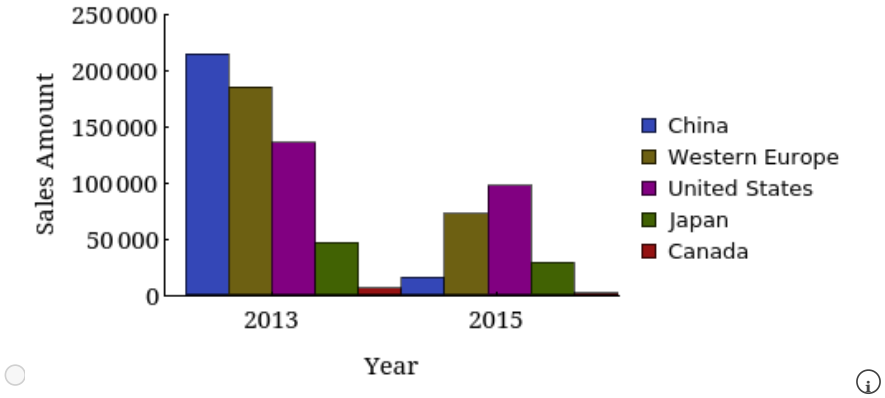
Electric plug-in vehicle sales have been increasing worldwide. The table below displays hypothetical data on electric plug-in vehicle sales in the world's top markets in 2013 and 2015.

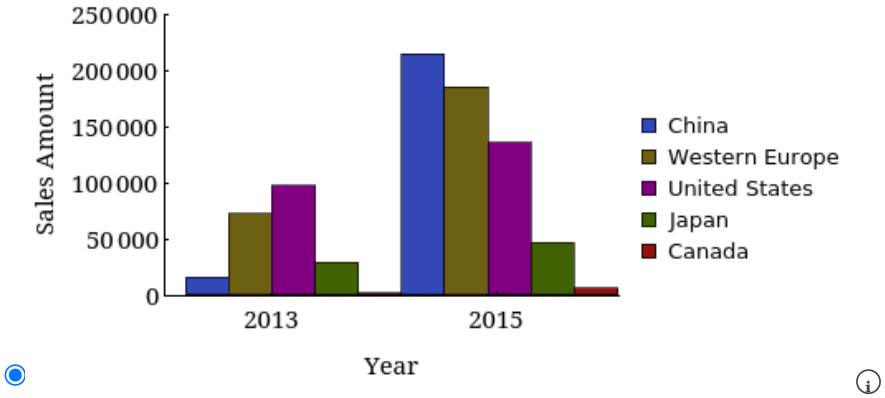
Region	2013	2015
China	15,004	214,283
Western Europe	72,233	184,500
United States	97,102	135,262
Japan	28,716	46,339
Canada	1,031	5,484

(a) Construct a side-by-side bar chart with year as the variable on the horizontal axis.





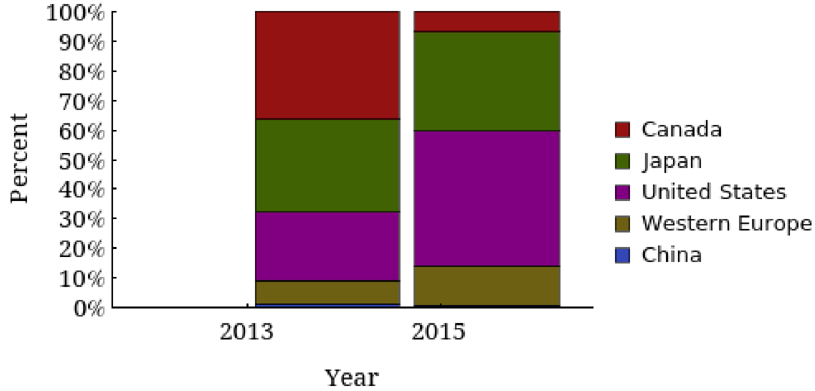


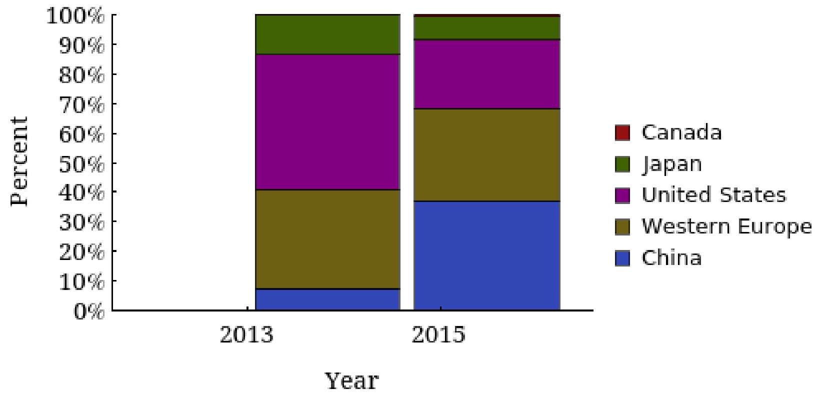


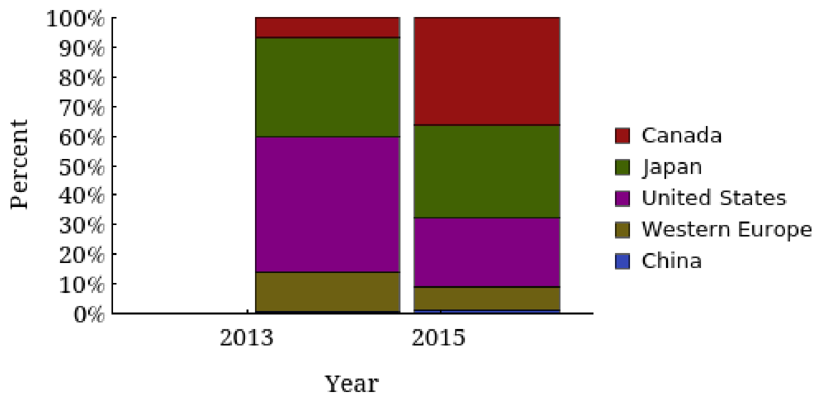
(b) Convert the above table to percentage allocation for each year. (Round your answers to one decimal place.)

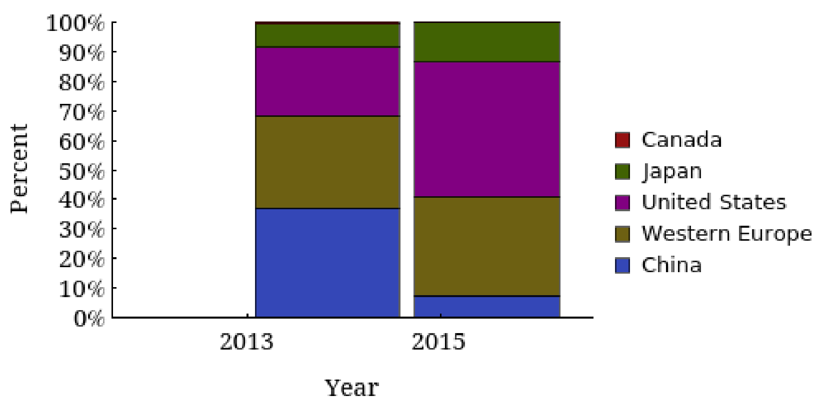
Region	2013	2015
China	7 %	36.6 %
Western Europe	33.7 %	31.5 %
United States	45.4 %	23.1 %
Japan	13.4 %	7.9 %
Canada	0.5 %	0.9 %
Total:	100 %	100 %

Construct a stacked bar chart with year as the variable on the horizontal axis.









(c) Is the display in part (a) or part (b) more insightful? Explain.

The way data is presented in part (b) is more insightful than in part (a). This is because of the way the data is stacked upon

each other for a particular categorical value. Displaying data in the side-by-side bar chart can become cumbersome if there are more attributes associated with the data

Score: 0.55 out of 0.55

Comment:

Need Help? Read It

6. [7.68/7.68000000000001 Points] DETAILS PREVIOUS ANSWERS ASWSBE14 3.E.031.MI.SA.

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

This question has several parts that must be completed sequentially. If you skip a part of the question, you will not receive any points for the skipped part, and you will not be able to come back to the skipped part.

Tutorial Exercise

According to the 2016 Consumer Expenditure Survey, Americans spend an average of $1,124 on cellular phone service annually.† Suppose that we wish to determine if there are differences in cellular phone expenditures by age group. Therefore, samples of 10 consumers were selected for three age groups (18–34, 35–44, 45 and older). The annual expenditure (in dollars) for each person in the sample is provided in the table below.

18–34	35–44	45 and Older
1,505	1,019	1,140
165	584	961
1,506	1,842	415
2,095	1,550	1,379
1,671	1,327	1,254
1,144	1,106	840
1,987	2,072	773
1,250	1,400	1,202
1,717	1,636	1,310
1,440	1,465	1,525

(a) Compute the mean, variance, and standard deviation for each of the three samples.

(b) What observations can be made based on these data?

Step 1

(a) Compute the mean, variance, and standard deviation for each of the three samples.

We are asked to compute the mean dollar amount spent on cellular phone service in the last year for each age groups. Recall the formula to calculate the mean where x_i is the ith observation and n is the sample size.

$$\bar{x} = \frac{\sum x_i}{n}$$

There are n = 10 observations for each age group. Use the given data to compute each mean.

18–34	35–44	45 and Older
1,505	1,019	1,140

165	584	961
1,506	1,842	415
2,095	1,550	1,379
1,671	1,327	1,254
1,144	1,106	840
1,987	2,072	773
1,250	1,400	1,202
1,717	1,636	1,310
1,440	1,465	1,525

Compute the mean for the 18–34 age group.

$$\bar{x} = \frac{\sum x_i}{n}$$

$$= \frac{1{,}505 + 165 + 1{,}506 + 2{,}095 + 1{,}671 + 1{,}144 + 1{,}987 + 1{,}250 + \boxed{1717} + 1{,}440}{10}$$

$$= \$\boxed{1448}$$

Compute the mean for the 35–44 age group.

$$\bar{x} = \frac{\sum x_i}{n}$$

$$= \frac{1{,}019 + 584 + \boxed{1842} + 1{,}550 + 1{,}327 + 1{,}106 + 2{,}072 + 1{,}400 + 1{,}636 + 1{,}465}{10}$$

$$= \$\boxed{1400.1}$$

Compute the mean for the 45 and older age group.

$$\bar{x} = \frac{\sum x_i}{n}$$

$$= \frac{1{,}140 + 961 + 415 + 1{,}379 + 1{,}254 + 840 + \boxed{773} + 1{,}202 + 1{,}310 + 1{,}525}{10}$$

$$= \$\boxed{1079.9}$$

Step 2

The mean for the 18–34 age group was computed to be 1,448. Since the data are from a sample, the formula for a sample variance should be used.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

A table like the one below can be used to organize the observations, sample mean, deviations about the mean, and squared deviations about the mean. Fill in the missing values.

Observation x_i	Sample Mean $\bar{x}$	Deviation About the Mean $x_i - \bar{x}$	Squared Deviation about the Mean $(x_i - \bar{x})^2$
1,505	1,448	57	3,249

165	1,448	−1,283	1,646,089
1,506	1,448	58	3364 ✔
2,095	1,448	647 ✔	418,609
1,671	1,448	223	49729 ✔
1,144	1,448	−304	92,416
1,987	1,448	539	290,521
1,250	1,448	-198 ✔	39,204
1,717	1,448	269	72361 ✔
1,440	1,448	−8	64
			$\sum (x_i - \bar{x})^2 =$ 2,615,606 ✔

Step 3

Use the found sum $\sum (x_i - \bar{x})^2 = 2{,}615{,}606$ and sample size of $n = 10$ to find the sample variance of the 18–34 age group, rounding the result to two decimal places.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{2{,}615{,}606}{9} = 290622.89$$ ✔

Step 4

The mean for the 35–44 age group was found to be 1,400.1. Complete the table below, rounding values to two decimal places if they are not exact.

Observation x_i	Sample Mean $\bar{x}$	Deviation About the Mean $x_i - \bar{x}$	Squared Deviation about the Mean $(x_i - \bar{x})^2$
1,019	1,400.1	−381.1	145,237.21
584	1,400.1	−816.1	666,019.21
1,842	1,400.1	441.9 ✔	195,275.61
1,550	1,400.1	149.9	22470.01 ✔
1,327	1,400.1	−73.1	5,343.61
1,106	1,400.1	-294.1 ✔	86,494.81
2,072	1,400.1	671.9	451,449.61
1,400	1,400.1	−0.1	0.01
1,636	1,400.1	235.9	55648.81 ✔

1,465	1,400.1	64.9	4,212.01
			$\sum (x_i - \bar{x})^2$ = 1,632,150.90 ✔

Step 5

Use the found sum $\sum (x_i - \bar{x})^2$ = 1,632,150.9 and sample size of n = 10 to find the sample variance of the 35–44 age group, rounding the result to two decimal places.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{1{,}632{,}150.90}{9} = 181350.1$$ ✔

Step 6

The mean for the 45 and older age group was found to be 1,079.9. Complete the table below, rounding the values to two decimal places if they are not exact.

Observation x_i	Sample Mean $\bar{x}$	Deviation about the Mean $x_i - \bar{x}$	Squared Deviation about the Mean $(x_i - \bar{x})^2$
1,140	1,079.9	60.1	3,612.01
961	1,079.9	-118.90 ✔	14,137.21
415	1,079.9	−664.9	442,092.01
1,379	1,079.9	299.1	89,460.81
1,254	1,079.9	174.1	30310.81 ✔
840	1,079.9	−239.9	57,552.01
773	1,079.9	−306.9	94,187.61
1,202	1,079.9	122.1 ✔	14,908.41
1,310	1,079.9	230.1	52946.01 ✔
1,525	1,079.9	445.1	198,114.01
			$\sum (x_i - \bar{x})^2$ = 997,320.90 ✔

Step 7

Use the found sum $\sum (x_i - \bar{x})^2$ = 997,320.9 and sample size of n = 10 to find the sample variance of the 45 and older age group, rounding the result to two decimal places.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{997{,}320.90}{9} = 110813.4333$$ ✔

Step 8

Recall that the sample standard deviation is the positive square root of the sample variance. The units of the standard deviation are the same units as the given data.

The variance for the 18–34 age group was found to be 290,622.89. Find the standard deviation for this age group, rounding the result to two decimal places.

$s = \sqrt{s^2}$

$= \sqrt{290,622.89}$

$= \$539.09$ ✔

The variance for the 35–44 age group was found to be 181,350.10. Find the standard deviation for this age group, rounding the result to two decimal places.

$s = \sqrt{s^2}$

$= \sqrt{181,350.10}$

$= \$425.85$ ✔

The variance for the 45 and older age group was found to be 110,813.43. Find the standard deviation for this age group, rounding the result to two decimal places.

$s = \sqrt{s^2}$

$= \sqrt{110,813.43}$

$= \$332.89$ ✔

Step 9

(b) What observations can be made based on these data?

In part (a), the following values were found.

	18–34	35–44	45+
mean	$1,448	$1,400.1	$1,079.9
variance	290,622.89	181,350.10	110,813.43
standard deviation	$539.09	$425.85	$332.89

The age group that spends the least amount of money per year on cellular phone service is the 45 and older ✔ age group.
The age group that spends the most amount of money per year on cellular phone service is the 18-34 ✔ age group.

You have now completed the Master It.

Need Help? Read It

7. **[7.69/7.69 Points]** DETAILS PREVIOUS ANSWERS **ASWSBE14 3.E.042.**

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

Many families in California are using backyard structures for home offices, art studios, and hobby areas as well as for additional storage. Suppose that the mean price for a customized wooden, shingled backyard structure is $3,200. Assume that the standard deviation is $1,400.

(a) What is the *z*-score for a backyard structure costing $2,400? (Round your answer to two decimal places.)

z = -0.57 ✓

(b) What is the *z*-score for a backyard structure costing $6,700?

z = 2.5 ✓

(c) Interpret the *z*-scores in parts (a) and (b). Comment on whether either should be considered an outlier.

The cost $2,400 is 0.57 standard deviations below ✓ the mean, therefore it is not an outlier ✓ . The cost $6,700 is 2.50 standard deviations above ✓ the mean, therefore it is not an outlier ✓ .

(d) The cost for a backyard shed-office combination built in Albany, California, is $11,950. Compute the *z*-score for this structure.

z = 6.25 ✓

Should this structure be considered an outlier? Explain.

This structure is an outlier ✓ because the *z*-score is greater than 3 ✓ .

Need Help? Read It

8. [7.69/7.69 Points] DETAILS PREVIOUS ANSWERS ASWSBE14 3.E.045.

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

A newspaper article reported that a computer company has unveiled a new tablet computer marketed specifically to school districts for use by students. The new tablets will have faster processors and a cheaper price point in an effort to take market share away from a competing company in public school districts. Suppose that the following data represent the percentages of students currently using the company's tablets for a sample of 18 U.S. public school districts. (Round your answers to two decimal places.)

14	21	11	20	25	17	41	28	63	19	14	21	17	23	26
23	25	18												

(a) Compute the mean and median percentage of students currently using the company's tablets.

mean 23.67 ✓ %
median 21 ✓ %

(b) Compute the first and third quartiles (as percentages) for these data.

Q_1 17 ✓ %
Q_3 25.25 ✓ %

(c) Compute the range and interquartile range (as percentages) for these data.

range 52 ✓ %
interquartile range 8.25 ✓ %

(d) Compute the variance and standard deviation (as a percentage) for these data.

variance 140.82 ✓
standard deviation 11.87 ✓ %

(e) Are there any outliers in these data?

There are 0 values ✓ below the lower limit and are 2 values ✓ above the upper limit.

(f) Based on your calculated values, what can we say about the percentage of students using the company's tablets in public school districts?

- ○ Use of the tablets is very low for all school districts.
- ○ Use of the tablets is very high for all school districts.
- ○ Relative to the mean, there are some school districts where much fewer students are using the tablets.
- ○ Relative to the mean, use of the tablets is similar for all school districts.
- ◉ Relative to the mean, there are some school districts where many more students are using the tablets. ✓

Need Help? Read It

9. **[7.69/7.69 Points]** DETAILS PREVIOUS ANSWERS **ASWSBE14 3.E.053.**

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

A magazine's list of the world's most admired companies for 2014 is provided in the data below. The data in the column labeled "Return" shows the one-year total return (%) for the top ranked 50 companies. For the same time period the S&P average return was 18.4%.

Rank	Company Name	Return (%)
1	Company A	15.8
2	Company B	28.8
3	Company C	50.3
4	Company D	12.7
5	Company E	28.1
6	Company F	1.4
7	Company G	32.8
8	Company H	24.4
9	Company I	83.5
10	Company J	13.3
11	Company K	39.6
12	Company L	9.1
13	Company M	32.8
14	Company N	15.1
15	Company O	3.1
16	Company P	−13.4
17	Company Q	6.8
18	Company R	−14.9
19	Company S	20.5
20	Company T	13
21	Company U	−12.9
22	Company V	1.8
23	Company W	28.3
24	Company X	33.2
25	Company Y	31.3

Rank	Company Name	Return (%)
26	Company Z	64.2
27	Company AA	2.5
28	Company BB	5.5
29	Company CC	−10.6
30	Company DD	15.3
31	Company EE	4.7
32	Company FF	20.3
33	Company GG	−4
34	Company HH	8
35	Company II	29.6
36	Company JJ	5.1
37	Company KK	−7.3
38	Company LL	118.1
39	Company MM	20.7
40	Company NN	15.7
41	Company OO	25.2
42	Company PP	12.7
43	Company QQ	33.3
44	Company RR	−6.1
45	Company SS	8.3
46	Company TT	−8.1
47	Company UU	14.3
48	Company VV	115.6
49	Company WW	6.4
50	Company XX	−3.5

(a) Compute the median return (in %) for the top-ranked 50 companies.

13.8 ✔ %

(b) What percentage of the top-ranked 50 companies had a one-year return greater than the S&P average return?

40 ✔ %

(c) Develop the five-number summary (in %) for the data.

minimum	-14.9 ✔ %
first quartile	2.95 ✔ %
median	13.8 ✔ %
third quartile	29 ✔ %
maximum	118.1 ✔ %

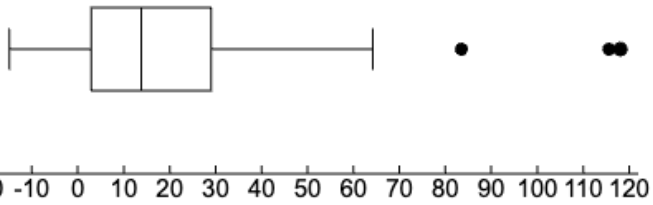
(d) Compute the lower and upper limits (in %).

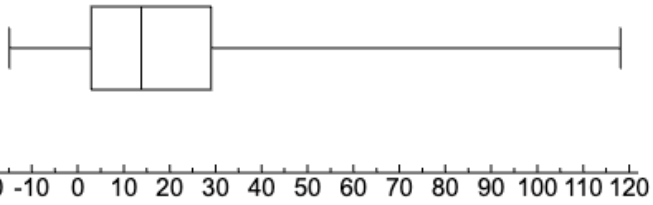
lower limit	-36.125 ✔ %
upper limit	68.075 ✔ %

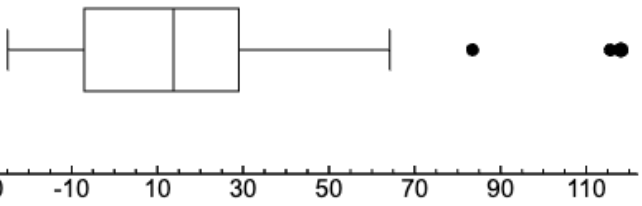
Are there any outliers?

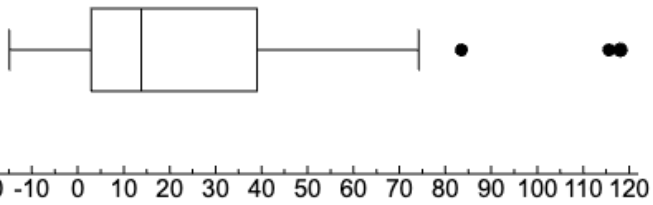
Returns less than the lower ✔ bound or greater than the upper ✔ bound are considered outliers. Thus, there are three ✔ outliers.

(e) Develop a boxplot for the one-year total return.









✔

Need Help? Read It Watch It

10. [7.69/7.69 Points] DETAILS PREVIOUS ANSWERS ASWSBE14 2.E.058.MI.SA.

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

This question has several parts that must be completed sequentially. If you skip a part of the question, you will not receive any points for the skipped part, and you will not be able to come back to the skipped part.

Tutorial Exercise

A zoo has categorized its visitors into three categories: member, school, and general. The member category refers to visitors who pay an annual fee to support the zoo. Members receive certain benefits such as discounts on merchandise and trips planned by the zoo. The school category includes faculty and students from day care and elementary and secondary schools; these visitors generally receive a discounted rate. The general category includes all other visitors. The zoo has been concerned about a recent drop in attendance. To help better understand attendance and membership, a zoo staff member has collected the following data.

Visitor Category	Attendance			
	2011	2012	2013	2014
General	151,213	156,204	160,933	166,606
Member	118,023	107,295	100,937	83,717
School	85,385	82,376	84,470	83,790
Total	354,621	345,875	346,340	334,113

(a) Construct a bar chart of total attendance over time. Comment on any trend in the data.

(b) Construct a side-by-side bar chart showing attendance by visitor category with year as the variable on the horizontal axis.

(c) Comment on what is happening to zoo attendance based on the charts from parts (a) and (b).

Step 1

(a) Construct a bar chart of total attendance over time. Comment on any trend in the data.

Recall that a bar chart is used to display categorical data summarized as frequencies, relative frequencies, or percent frequencies. Data for each category will be displayed in a separate bar. Here, the data are given as frequencies.

Note that the data are given in terms of visitor category per year. Since the zoo is concerned about a drop in attendance, the total attendance for the given years will be compared. Each bar will represent the attendance for one year.

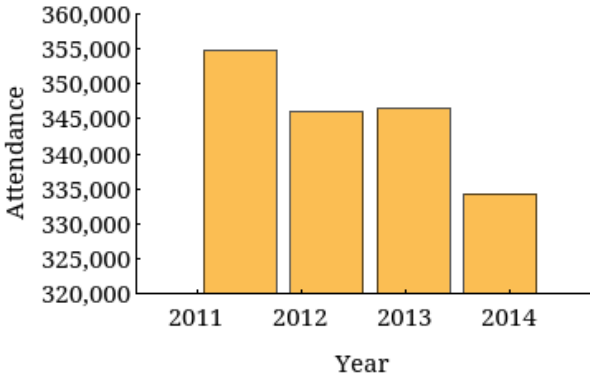
The height of each bar will be the sum of the attendance for the general, member, and school categories, given in the Total row. The data are given below.

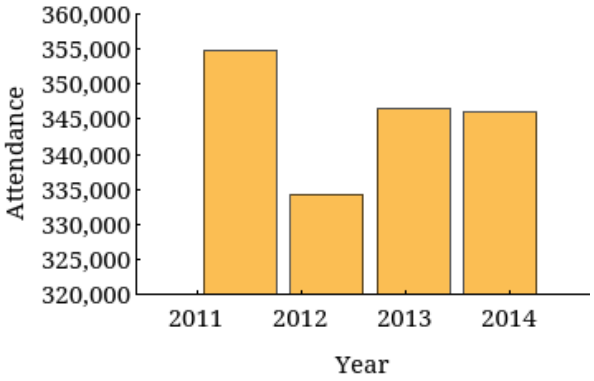
Visitor Category	Attendance			
	2011	2012	2013	2014
General	151,213	156,204	160,933	166,606
Member	118,023	107,295	100,937	83,717
School	85,385	82,376	84,470	83,790

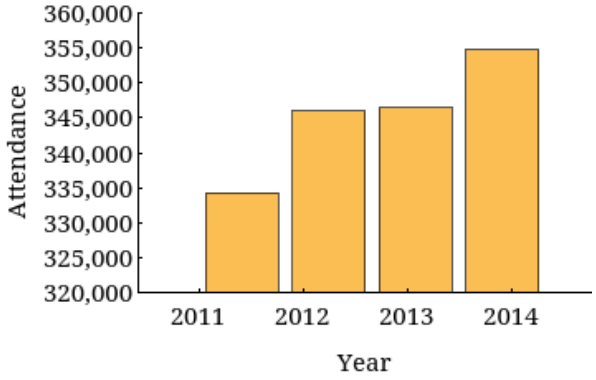
Total	354,621	345,875	346,340	334,113

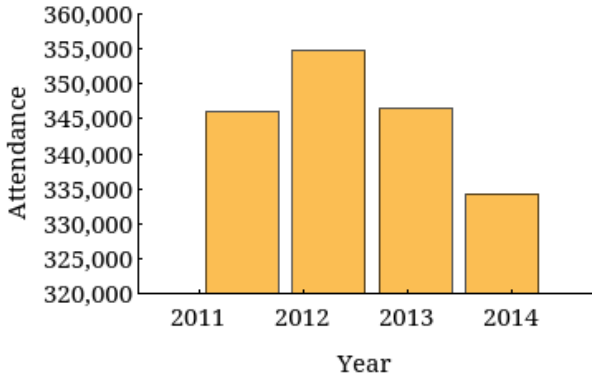
Thus, the height of the bar for 2011 will be 354,621. The height of the bar for 2012 will be 345,875 ✔ . The height of the bar for 2013 will be 346,340, and the height of the bar for 2014 will be 334,113 ✔ .

Construct a bar chart that reflects these values.



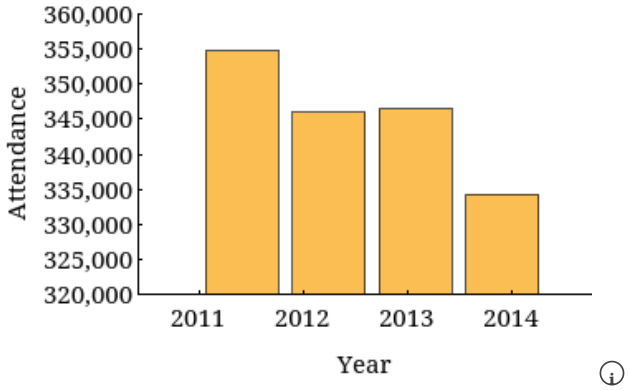






Step 2

The bar chart showing the attendance is given below.



The heights of the bars are close for the years 2012 and 2013 , indicating a similar attendance for these years. The highest attendance occurred in the year 2011 . The lowest attendance occurred in the year 2014 . This indicates that the total attendance for the zoo is decreasing .

Step 3

(b) Construct a side-by-side bar chart showing attendance by visitor category with year as the variable on the horizontal axis.

A side-by-side bar chart is useful for displaying data covering multiple categories. It is important to display each category in the same order for the different variables along the horizontal axis. Often the same color will be used for categories for different variables along the horizontal axis. This allows for conclusions to be drawn among the various categories as they relate to the variable on the horizontal axis.

Year will be the variable along the horizontal axis. For each year, there will be a bar corresponding to each of the visitor categories General, Member, and School. Therefore, for this data, there will be 3 bars for each value of year. Each of these will use a different color.

The years for which data was collected are 2011, 2012, 2013, and 2014. Thus, there will be 4 groupings of the visitor categories.

Step 4

The side-by-side bar chart will have 4 groupings, one for each year, of 3 bars, one for each type of visitor. The data are given below.

Visitor Category	Attendance			
	2011	2012	2013	2014
General	151,213	156,204	160,933	166,606
Member	118,023	107,295	100,937	83,717
School	85,385	82,376	84,470	83,790
Total	354,621	345,875	346,340	334,113

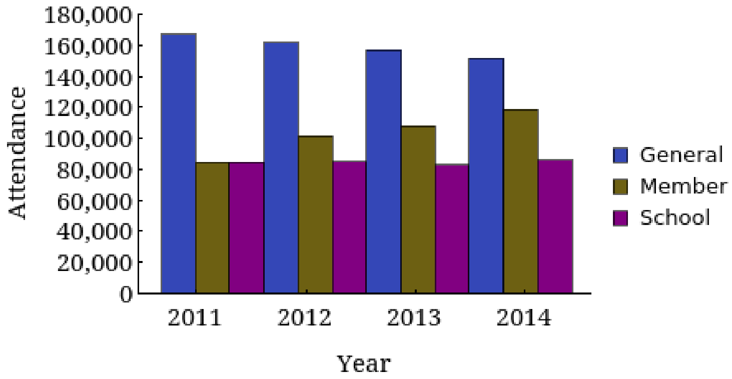
For the year 2011, the height of the bar for the General visitor category will be 151,213, the height corresponding to the Member visitor category will be 118,023, and the height for School will be 85,385.

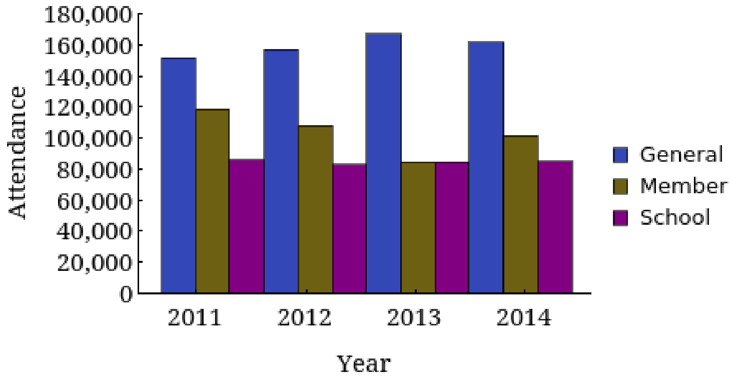
For the year 2012, the height of the bar for the General visitor category will be 156,204 , the height corresponding to the Member visitor category will be 107,295, and the height for School will be 82,376 .

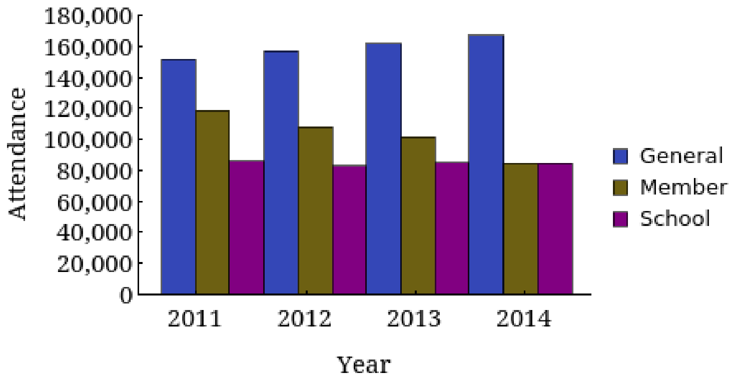
For the year 2013, the height of the bar for the General visitor category will be 160,933, the height corresponding to the Member visitor category will be 100,937 , and the height for School will be 84,470.

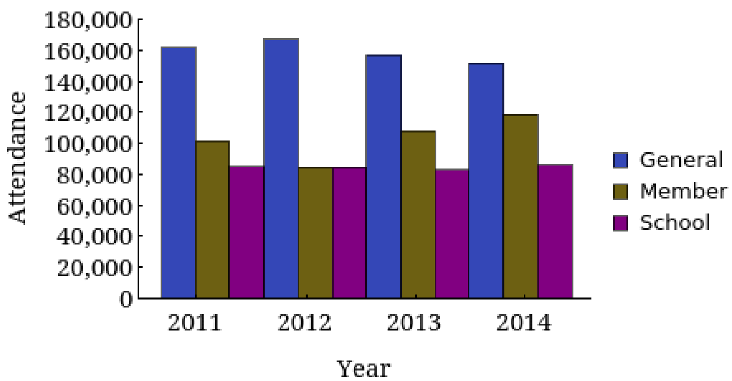
For the year 2014, the height of the bar for the General visitor category will be 166,606, the height corresponding to the Member visitor category will be 83,717, and the height for School will be 83,790 .

Construct the side-by-side bar chart.





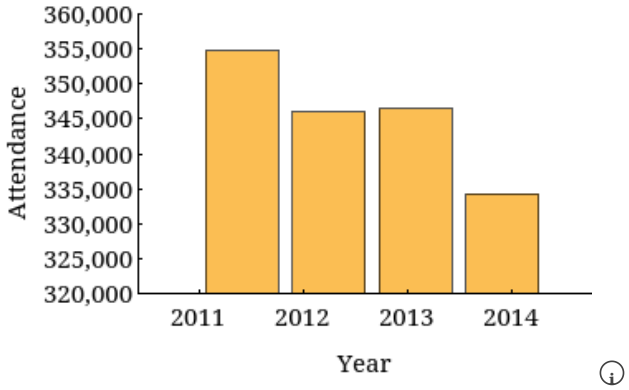




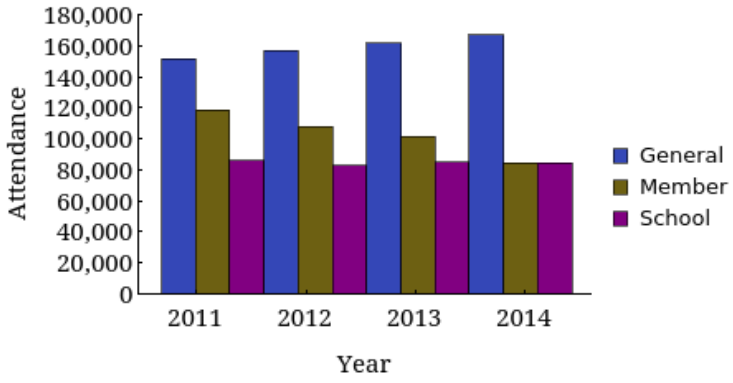
Step 5

(c) Comment on what is happening to zoo attendance based on the charts from parts (a) and (b).

The bar chart for the total attendance is given below. This representation of the data indicates that overall attendance is decreasing over time.



The side-by-side bar chart is given below.



Even though the general attendance is increasing ✔ , member attendance is decreasing ✔ . This difference is causing overall attendance to decrease from year to year. School attendance is roughly staying the same.

You have now completed the Master It.

Need Help? Read It

11. **[7.69/7.69 Points]** DETAILS PREVIOUS ANSWERS **ASWSBE14 3.E.059.**

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

Over the past 40 years, the percentage of homes in the United States with smoke detectors has risen steadily and has plateaued at about 96% as of 2015.† With this increase in the use of home smoke detectors, what has happened to the death rate from home fires? Suppose the following contains 17 years of data on the estimated percentage of homes with smoke detectors and the estimated home fire deaths per million of population.

Percentage of Homes with Smoke Detectors	Home Fire Deaths per Million of Population
0.51	22.8
0.68	20.7
0.75	17.2
0.77	20.4
0.78	19.3
0.83	18.8
0.82	17.5
0.86	16.1
0.87	13.5
0.89	14.3
0.91	14.2
0.93	12.9
0.95	10.8
0.96	10.8
0.97	10.1
0.96	8.3
0.96	8.0

(a) Do you expect a positive or negative relationship between smoke detector use and deaths from home fires? Why or why not?

We would expect there would be a [negative] ✔ relationship between smoke detector use and deaths from home fires. As more households have smoke detectors, warning of a fire would help the inhabitants of the home escape and lead to [fewer] ✔ deaths from home fires.

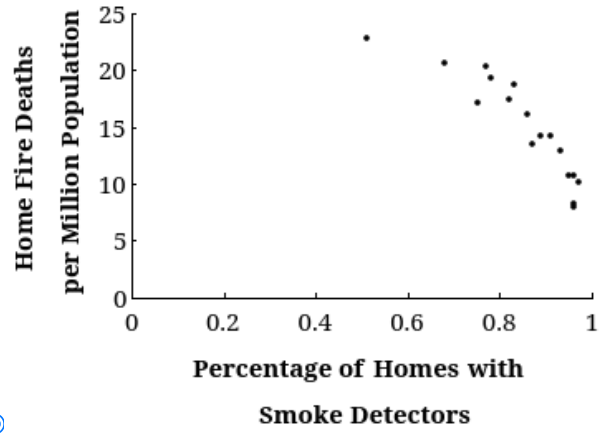
(b) Compute and report the correlation coefficient. (Round your answer to four decimal places.)

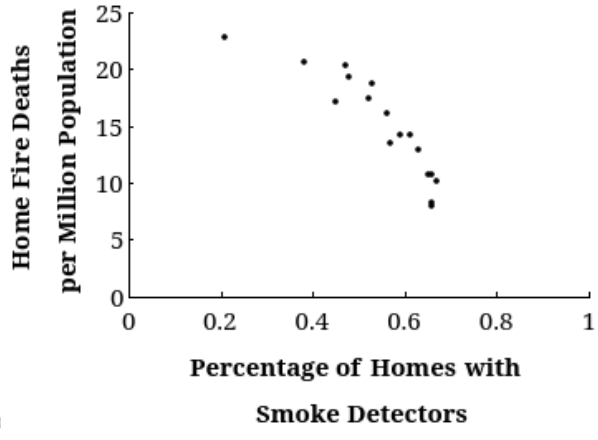
[-0.8971] ✔

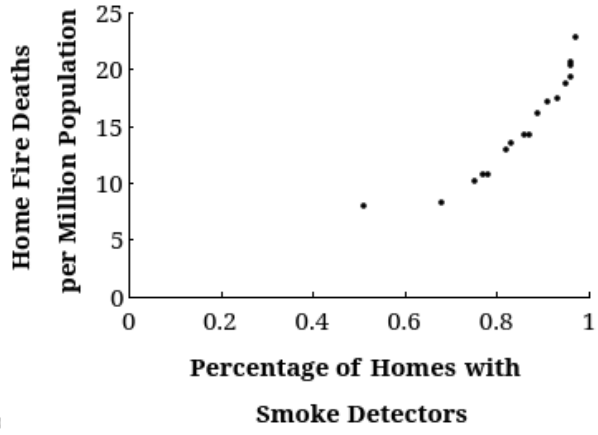
Is there a positive or negative correlation between smoke detector use and deaths from home fires? Comment.

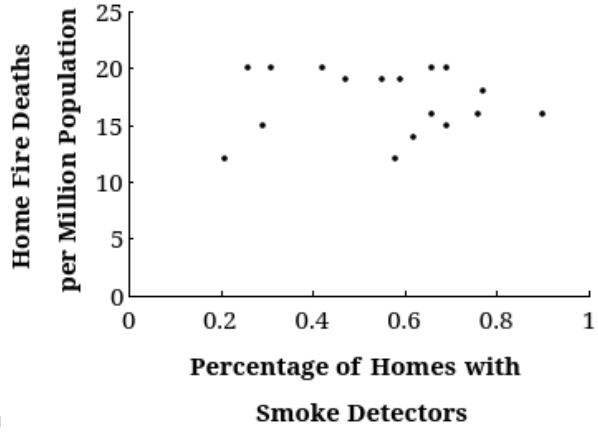
There is a strong negative ✔ correlation between smoke detecter use and death from home fires.

(c) Show a scatter plot of the death rate per million of population and the percentage of homes with smoke detectors.









Need Help? Read It

12. [7.69/7.69 Points] DETAILS PREVIOUS ANSWERS ASWSBE14 3.E.064.

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

DATAfile: WaitTracking

The average waiting time for a patient at an El Paso physician's office is just over 29 minutes, well above the national average of 21 minutes. In order to address the issue of long patient wait times, some physician's offices are using wait tracking systems to notify patients of expected wait times. Patients can adjust their arrival times based on this information and spend less time in waiting rooms. The following data show wait times (minutes) for a sample of patients at offices that do not have an office tracking system and wait times for a sample of patients at offices with an office tracking system.

Without Wait Tracking System	With Wait Tracking System
24	31
67	11
17	14
20	18
31	12
44	37
12	9
23	13
16	12
37	15

(a) What are the mean and median patient wait times (in min) for offices with a wait tracking system?

mean 17.2 ✓ min
median 13.5 ✓ min

What are the mean and median patient wait times (in min) for offices without a wait tracking system?

mean 29.1 ✓ min
median 23.5 ✓ min

(b) What are the variance and standard deviation (in min) of patient wait times for offices with a wait tracking system? (Round your answers to one decimal place.)

variance 86.2 ✓
standard deviation 9.3 ✓ min

What are the variance and standard deviation (in min) of patient wait times for visits to offices without a wait tracking system? (Round your answers to one decimal place.)

variance 275.7 ✓
standard deviation 16.6 ✓ min

(c) Do offices with a wait tracking system have shorter patient wait times than offices without a wait tracking system? Explain.

Offices with a wait tracking system have shorter ✔ patient wait times than offices without a wait tracking system.

(d) Considering only offices without a wait tracking system, what is the *z*-score for the seventh patient in the sample? (Round your answer to two decimal places.)

-1.03 ✔

(e) Considering only offices with a wait tracking system, what is the *z*-score for the fifth patient in the sample? (Round your answer to two decimal places.)

-0.56 ✔

How does this *z*-score compare with the *z*-score you calculated for part (d)?

As indicated by the negative ✔ *z*-scores, both patients had wait times that are lower than ✔ the means of their respective samples. Even though the patients had the same wait time, the *z*-score for the fifth patient in the sample who visited an office with a wait tracking system is much larger ✔ because that patient is part of a sample with a smaller ✔ mean and a smaller ✔ standard deviation.

(f) Based on *z*-scores, do the data for offices without a wait tracking system contain any outliers?

The data for offices without a wait tracking system contains no ✔ outliers.

Based on *z*-scores, do the data for offices with a wait tracking system contain any outliers?

The data for offices with a wait tracking system contains no ✔ outliers.

Need Help? Read It

13. [7.72/7.72 Points] DETAILS PREVIOUS ANSWERS ASWSBE14 3.E.070.

MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

A magazine provides an annual list of the 500 best hotels in the world. The magazine provides a rating for each hotel along with a brief description that includes the size of the hotel, amenities, and the cost per night (in dollars) for a double room. A sample of 12 of the top-rated hotels in the United States follows.

Hotel	Location	Rooms	Cost/Night
Hotel A	Phoenix, AZ	217	499
Hotel B	Orlando, FL	721	340
Hotel C	Los Angeles, CA	285	585
Hotel D	Boston, MA	267	495
Hotel E	Washington, DC	151	495
Hotel F	Asheville, NC	219	279
Hotel G	Phoenix, AZ	404	279
Hotel H	Island of Hawaii	340	455
Hotel I	Laguna Beach, CA	256	595
Hotel J	Chicago, IL	408	367
Hotel K	Dana Point, CA	406	675
Hotel L	Colorado Springs, CO	694	420

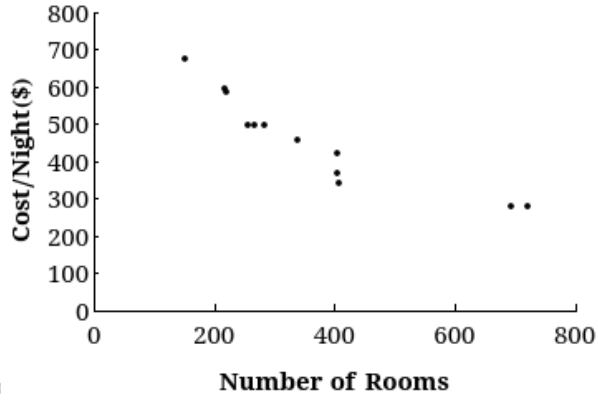
(a) What is the mean number of rooms?

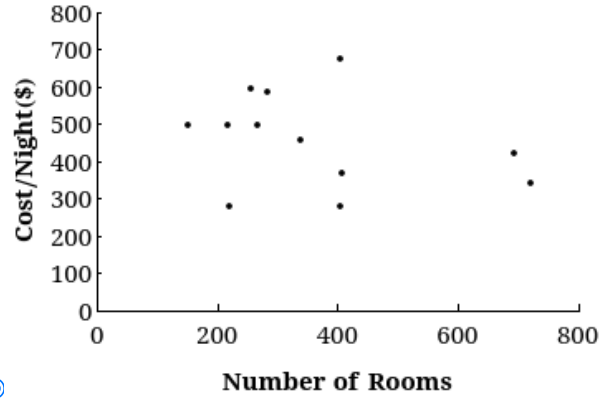
364 ✓ rooms

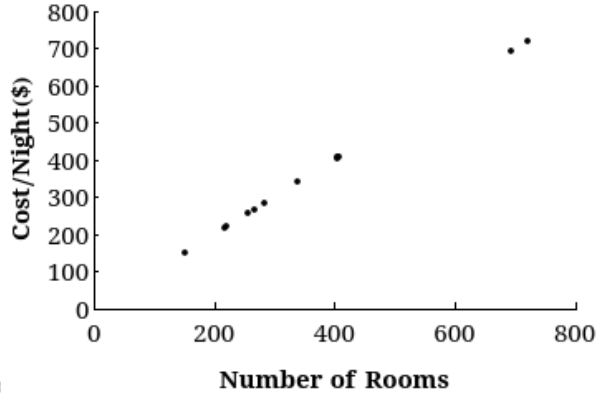
(b) What is the mean cost per night (in dollars) for a double room?

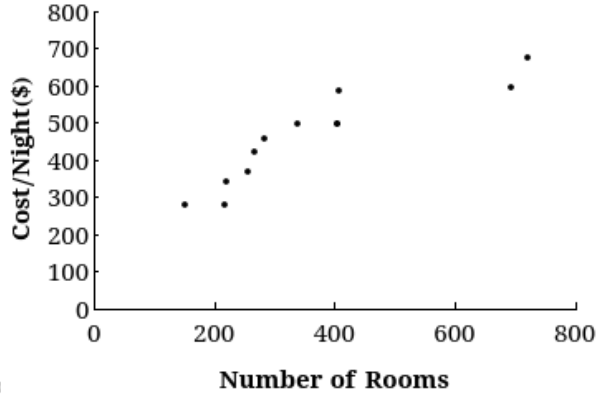
$457 ✓

(c) Develop a scatter diagram with the number of rooms on the horizontal axis and the cost per night on the vertical axis.









Does there appear to be a relationship between the number of rooms and the cost per night? Discuss.

There appears to be no ✓ relationship between the two variables. When the number of rooms becomes larger, there is no ✓ indication that the cost per night changes.

(d) What is the sample correlation coefficient? (Round your answer to three decimal places.)

-0.293 ✓

What does it tell you about the relationship between the number of rooms and the cost per night for a double room?

The sample correlation coefficient indicates that there is a slightly negative ✓ linear relationship between the number of rooms and the cost per night for a double room. This means that higher room rates tend to be associated with smaller ✓ hotels.

Does this appear reasonable? Discuss.

- (●) Yes, economies of scale allow larger hotels to charge less per night and still achieve a nice profit.
- ( ) Yes, economies of scale allow smaller hotels to charge less per night and still achieve a nice profit.
- ( ) No, economies of scale allow larger hotels to charge less per night and still achieve a nice profit.
- ( ) No, economies of scale allow smaller hotels to charge less per night and still achieve a nice profit.

✓

Need Help? Read It