

### Chapter 3 Data Visualization

---

#### Solutions:

1.

<b>Salesperson</b>	<b>Total Sales (\$)</b>	<b>Average Performance Bonus Previous Years (\$)</b>	<b>Customer Accounts</b>	<b>Years with Company</b>
Smith, Michael	325,001	12,499	124	14
Yu, Joe	13,678	240	9	7
Reeves, Bill	452,359	21,987	175	21
Hamilton, Joshua	87,424	7,643	28	3
Harper, Derek	87,654	1,250	21	4
Quinn, Dorothy	234,091	14,568	48	9
Graves, Lorrie	379,402	27,981	121	12
Sun, Yi	31,734	673	7	1
Thompson, Nicole	127,845	13,323	17	3

Some of the changes made in this table include:

- Deleting unnecessary gridlines in the table

- Removing bolded font except for column titles

- Left-align text column and right-align numerical columns

- Adding commas to dollar values to ease readability and removing unnecessary digits to right of decimal place

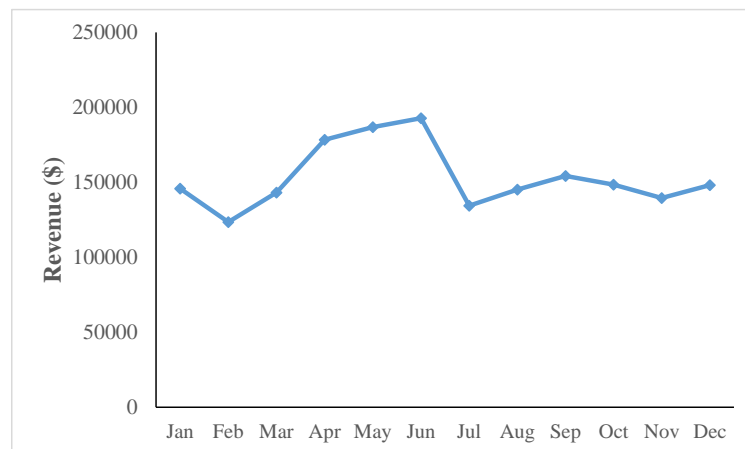
2. a. The readability of the table could be improved by removing unnecessary gridlines, left-aligning the first column because it contains text entries, right-aligning all other columns because they contain numerical values, reducing the number of digits displayed by displaying numerical values in millions or billions of dollars, and using shading to differentiate the columns. Note that we could also sort these values by GDP in 2010 if this was the value of most interest.

b.

Country	Gross Domestic Product (in Billions of US Dollars, \$)					
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Albania	7.4	8.1	9.7	11.6	10.8	10.6
Argentina	169.7	198.0	241.0	301.3	285.1	339.6
Australia	704.5	758.3	916.9	983.0	934.2	1178.8
Austria	272.9	290.7	336.8	375.8	344.5	341.4
Belgium	335.6	355.4	408.5	451.7	421.4	416.5
Brazil	756.8	935.5	1175.0	1407.0	1370.8	1782.4
Canada	1056.8	1193.9	1332.1	1404.8	1245.2	1469.9
Costa Rica	18.9	21.2	24.7	28.0	28.0	33.9
Czech Republic	111.7	128.7	156.6	194.6	170.9	172.6
Finland	169.9	180.2	214.8	236.6	206.7	207.9
France	1915.0	2015.0	2312.8	2541.6	2367.9	2301.8
Germany	2516.9	2632.8	2984.7	3258.9	2973.7	2966.1
Greece	217.2	234.3	273.8	307.1	292.6	269.1
Ireland	177.7	195.3	229.6	233.3	199.5	186.2
Israel	122.4	133.2	153.4	185.7	178.7	199.8
Italy	1597.3	1661.9	1892.7	2063.9	1900.0	1836.9
Mexico	823.3	928.5	1011.9	1085.6	858.3	1010.3
Netherlands	567.3	600.9	694.8	775.4	708.0	700.8
New Zealand	108.9	105.1	129.0	125.7	112.9	136.2
Peru	72.3	84.4	98.5	117.8	120.4	144.3
Poland	267.8	300.1	371.9	463.4	382.6	413.4
Portugal	165.3	172.9	200.1	218.7	206.0	200.5
Saudi Arabia	317.4	358.4	386.7	477.3	374.4	436.2
Singapore	119.7	139.1	167.0	179.3	173.8	209.7
South Africa	220.3	232.2	254.4	247.5	256.9	328.8
Spain	1012.0	1100.0	1293.2	1459.0	1361.0	1287.9
Switzerland	350.6	368.2	409.2	474.7	464.9	498.2
Turkey	425.5	467.9	579.0	656.6	557.7	655.8
United Kingdom	2030.3	2178.0	2504.6	2381.9	1959.2	2005.6
United States	12579.7	13336.3	14010.9	14369.5	14113.3	14601.6

3. a. The chart contains unnecessary gridlines, the y-axis label values are spaced much too close together, and the shading of the chart does not add value.

b.



Note that here we have also added small markers on the line chart at each data point to indicate that the data are not continuous.

4.

Row Labels	Count of Major	Average of Monthly Salary
Accounting	28	\$4,020
Finance	21	\$3,695
Info Systems	16	\$4,000
Management	24	\$3,180
Marketing	22	\$3,345
<b>Grand Total</b>	<b>111</b>	<b>\$3,640</b>

- The PivotTable shows that accounting major has the greatest number of graduates with 28 students.
- Accounting has the highest average starting monthly salary at \$4,020.
- By changing the Value Field Settings for Monthly Salary from Average to Max, we see that an accounting student has the highest starting salary at \$5,650.  
By changing the Value Field Settings for Monthly Salary to Min, we see that a management student has the lowest starting salary at \$2,240.

5.

Column Labels	0-9999	10000-19999	20000-29999	30000-39999	Grand Total
Count of # U.S. Locations	13	3	1	3	20

- Thirteen franchises have between 0 and 9999 locations.
- Three franchises have more than 30,000 locations.

6. a.

Row Labels	10-20	20-30	30-40	50-60	Grand Total
DE	1	25	1		27
FI	9	1			10
IE	2	3	2	1	8
<b>Grand Total</b>	<b>10</b>	<b>28</b>	<b>4</b>	<b>1</b>	<b>45</b>

**PivotTable Fields**

Choose fields to add to report:

- Fund Name
- Fund Type
- Net Asset Value (\$)
- 5 Year Average Return (%)
- Expense Ratio (%)
- Morningstar Risk (Star)

MORE TABLES...

Drag fields between areas below:

**FILTERS:**

**COLUMNS:** 5 Year Average Return L...

**ROWS:** Fund Type

**VALUES:** Count of 5 Year Averag...

b. Fixed Income (FI) funds have had lower average returns than domestic equity (DE) or international equity (IE) funds; no FI funds has an average return greater than 19.99% and 9 of the 10 have average returns less than 9.99%. IE funds have had greater returns with no IE fund having an average return of less than 10%. IE funds are the only funds to have surpassed a 30% average return.

7. a.

Row Labels	Sum of Adjusted Gross Income [in Thousands]
ND	\$14,923,737
WY	\$15,216,842
VT	\$15,246,153
AK	\$17,312,637
SD	\$17,825,579
DC	\$18,177,370
MT	\$20,045,506
DE	\$22,983,203
RI	\$26,532,233
ME	\$28,934,364
ID	\$30,292,719
HI	\$30,592,981
WV	\$32,243,698
NM	\$38,144,031
NH	\$38,175,000
NE	\$41,565,443
MS	\$47,387,966
AR	\$49,783,295
UT	\$55,426,178
KS	\$65,216,514
NV	\$65,272,642
IA	\$68,946,841
OK	\$70,394,493

**PivotTable Fields**

Choose fields to add to report:

- State Abbreviation
- County Name
- Total Number of Tax Returns
- Adjusted Gross Income [in Thousands]
- Wages and Salaries Income [in Thousands]

MORE TABLES...

Drag fields between areas below:

**FILTERS:**

**COLUMNS:**

**ROWS:** State Abbreviation

**VALUES:** Sum of Adjusted Gross L...

North Dakota (ND) has the smallest sum of adjust gross income with \$14,923,737,000.

b.

Row Labels	Sum of Adjusted Gross Income (In Thousands)
TX	\$504,386,600
Loving County	\$788
King County	\$4,477
Kenedy County	\$4,651
Borden County	\$13,634
Travis County	\$25,231,356
Bexar County	\$31,208,385
Tarrant County	\$40,950,778
Dallas County	\$50,561,800
Harris County	\$89,933,750
<b>Grand Total</b>	<b>\$504,386,600</b>

Loving County had the smallest sum of adjusted gross income in Texas; Harris County had the largest sum of adjusted gross income.

c. Harris County has the highest percentage of adjusted gross income in Texas at 17.83%.

d.

State Abbreviation	Sum of Adjusted Gross Income (In Thousands)
CA	12.56%
NY	7.06%
TX	6.94%
FL	5.91%
IL	4.66%
PA	4.31%
NJ	3.89%
OH	3.56%
MI	3.12%
VA	2.99%
MA	2.78%
GA	2.75%
NC	2.69%
WA	2.42%
MD	2.32%
MN	1.97%
IN	1.93%
WI	1.93%
AZ	1.85%
MO	1.80%

New York provides 7.06% of the total adjusted gross income in the United States.

8. a.

Count of Bank Name	Column Labels	2000	2001	2002	2003	2004	2007	2008	2009	2010	2011	2012	Grand Total
GA		1					1	5	25	21	23	4	80
Acworth										1			1
Alpharetta							1	2	1				4
Atlanta		1							9	1	1		9
Bamesville										1			1
Brunswick											1		1
Carrollton										2			2
Cartersville										1	1		2
Clarkesville											1		1
Clayton											1		1
Commerce									1				1
Cornelia										1			1
Cumming											1		1
Dallas											1		1
Dawsonville										1			1
Decatur											1		1
Doraville												1	1
Douglasville											1		1
Duluth							1				1		2

Georgia (GA) had the greatest number of bank closures between 2000 and 2012.

b. Nevada experienced 4 bank closures in 2010. These occurred in Carson City, Las Vegas, and Reno.

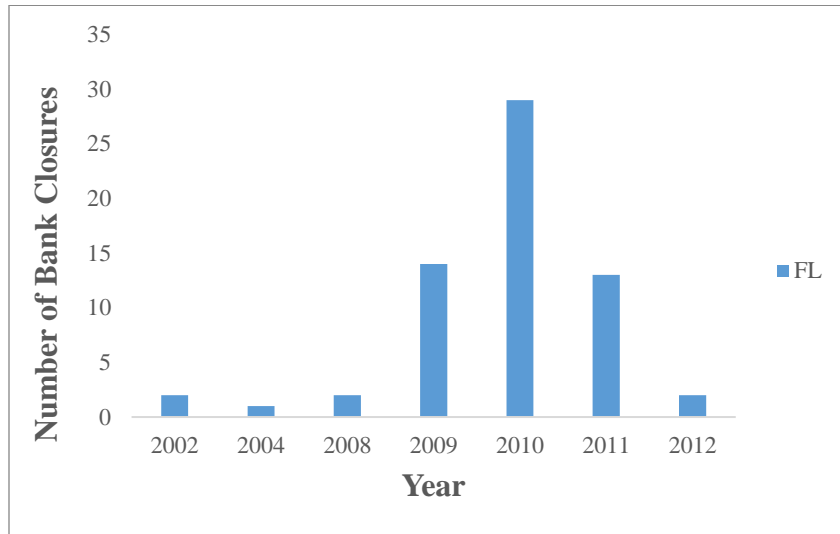
c.

Count of Bank Name	Column Labels	2009	2010	2011	2012	Grand Total
FL		14	29	13	2	58
CA		17	12	4		33
TX		5	1	1		7
NY		1	3			4
Grand Total		37	45	18	2	102

There were 102 bank closures between 2009 and 2012 in the states of California, Florida, Texas, and New York.

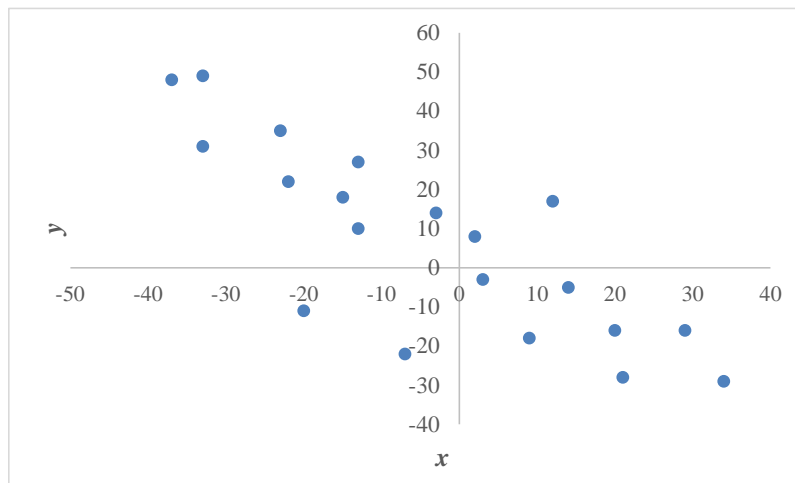
d. Naples had the greatest number of bank closures in Florida between 2009 and 2012 with 4 bank closures.

e.

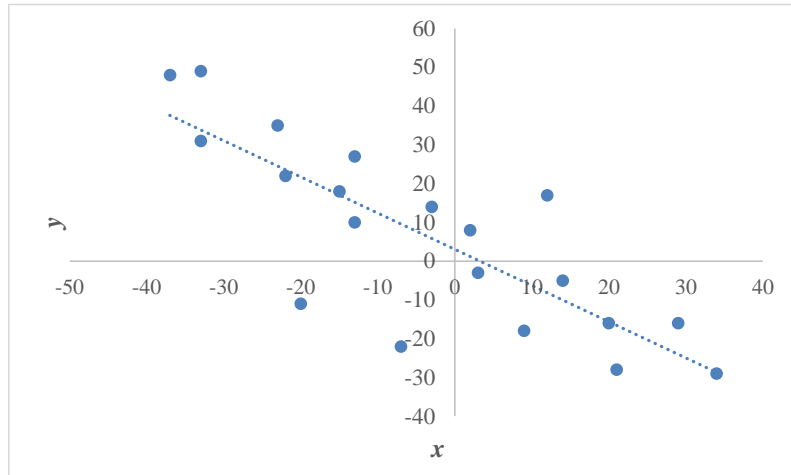


Bank closures peaked in 2010 in Florida and have decreased since then.

9. a.

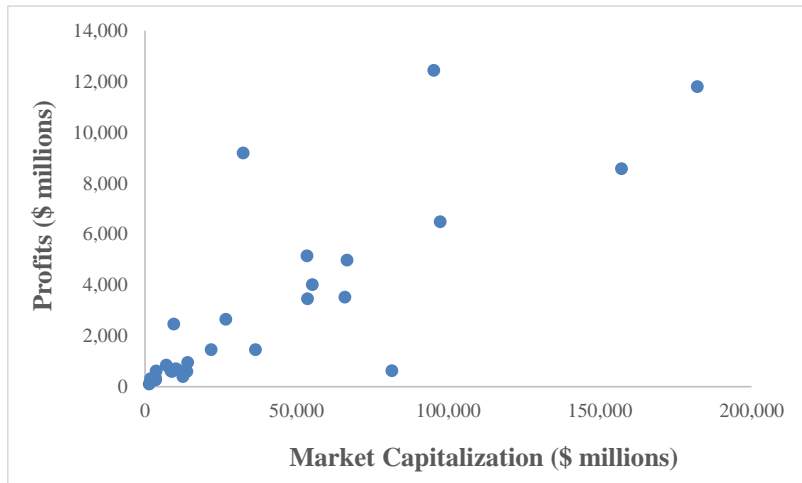


b.



There appears to be a negative linear relationship between the  $x$  and  $y$  variables.

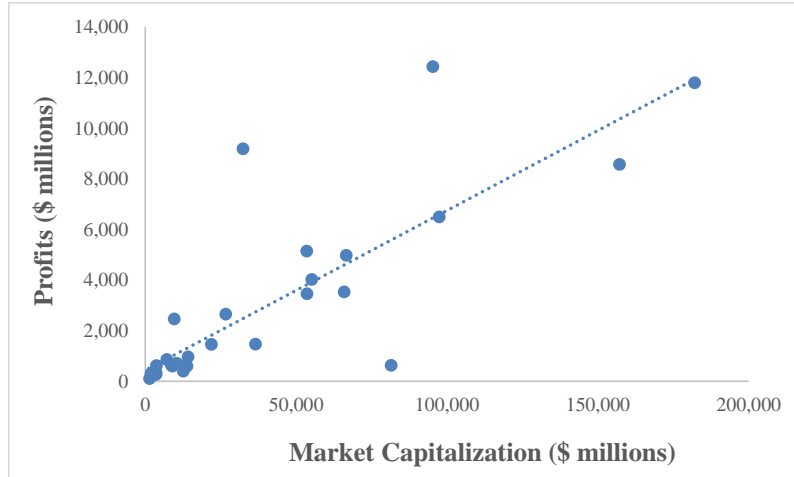
10. a.



There appears to be a positive linear relationship between profits and market capitalization. As profit increases, market capitalization also increases.

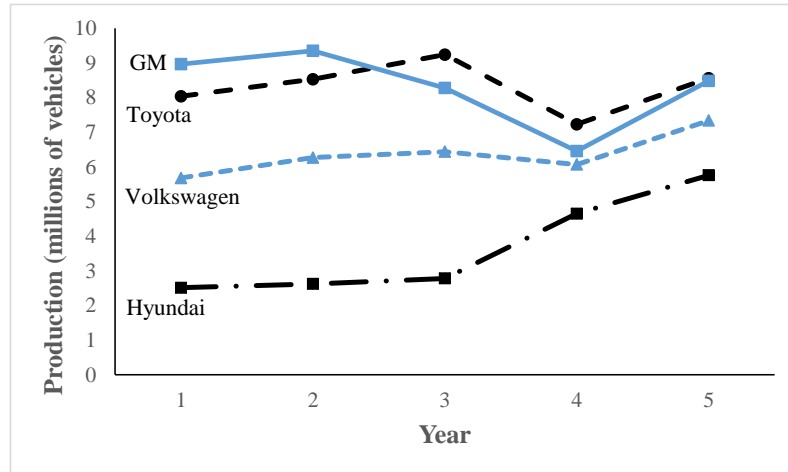
b.





The trendline confirms that there is a positive linear trend between profits and market capitalization.

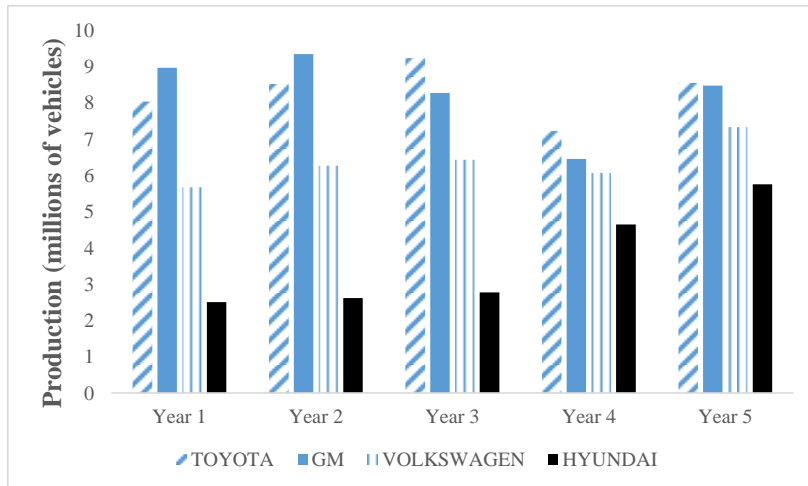
11. a.



Note that we have made several changes to the line chart here to improve readability. We have added axes labels, we have differentiated the lines using different line styles and we have placed labels in the chart for each line.

b. GM produced the greatest number of vehicles in Years 1 and 2, but was passed in year 3 by Toyota. Both GM and Toyota produced fewer vehicles in Year 4. By Year 5, Toyota and GM were producing approximately the same number of vehicles. Hyundai has seen significant increases in vehicle production in Years 4 and 5.

c.



Leading Manufacturers

Year 1: GM

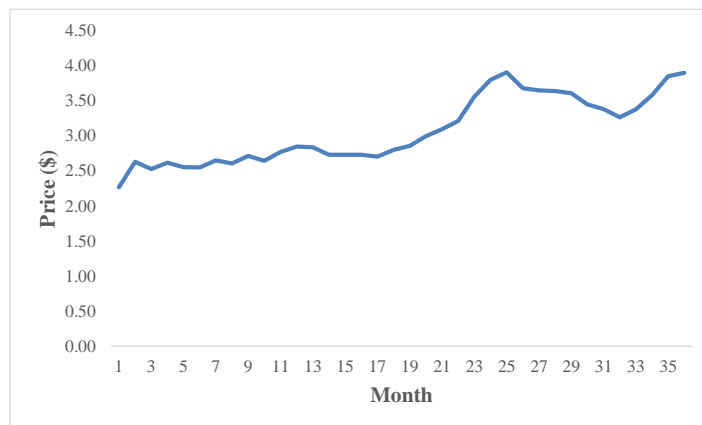
Year 2: GM

Year 3: Toyota

Year 4: Toyota

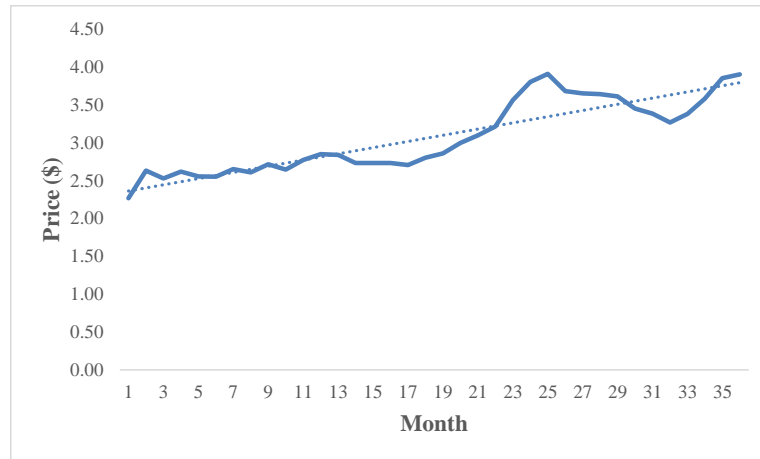
Year 5: Toyota (although GM and Toyota are producing about the same number of vehicles in Year 5)

12. a.



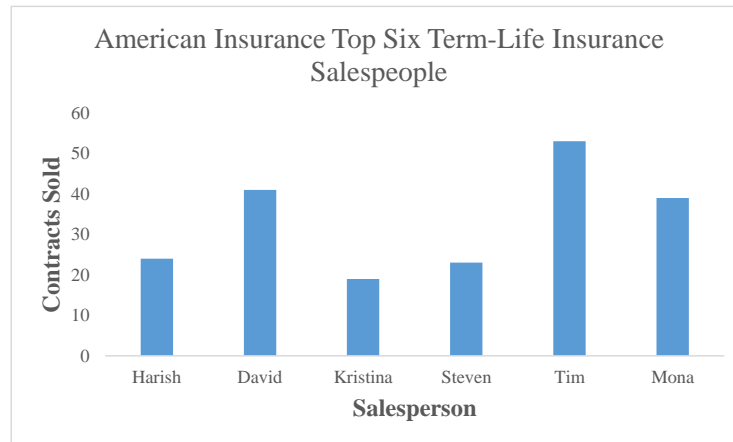
Gasoline prices were relatively steady for about the first 16 to 18 months and then increased rapidly through about month 25 before falling before rising in the last few months. Overall the price of gasoline appears to be increasing over the 36 months, but it is not a constant increase.

b.

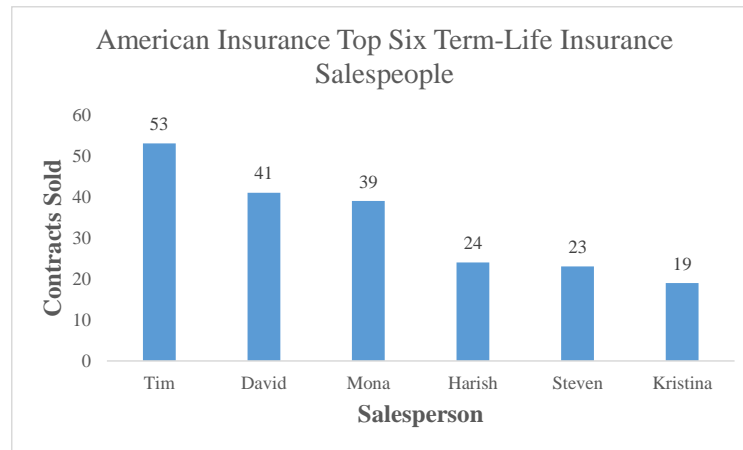


The trendline confirms that there is an overall linear increase in the price of gasoline over the 36 months.

13. a.



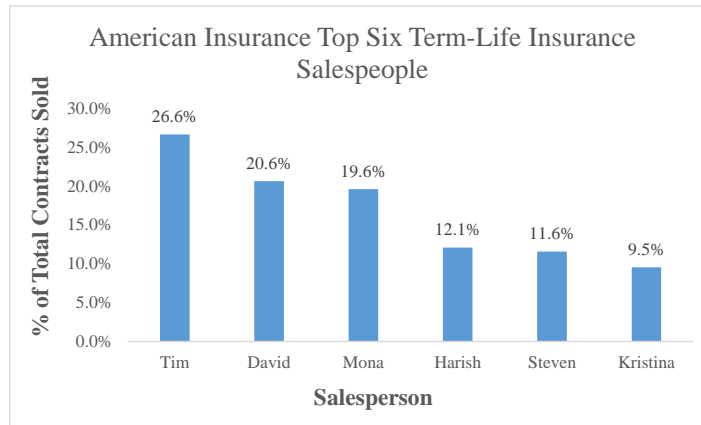
b. and c.



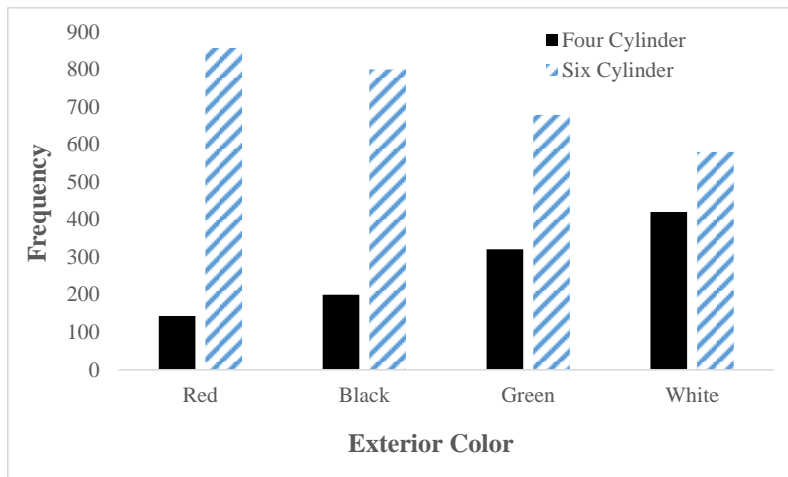
Sorting can be done by selecting the data in Excel and then using the **Sort** function in the **Sort & Filter** group under the **DATA** tab. Data labels can be added by right-clicking on one of the columns in the chart and selecting **Add Data Labels**.

14. a. It is difficult to distinguish the relative sizes of the different pieces of the pie chart. It can also be difficult to distinguish the different colors in the pie chart. Finally, it takes a lot of work for the reader to match the salesperson names to the different pieces of the pie chart.
- b. A sorted column or bar chart would be preferable to display these data.

c.

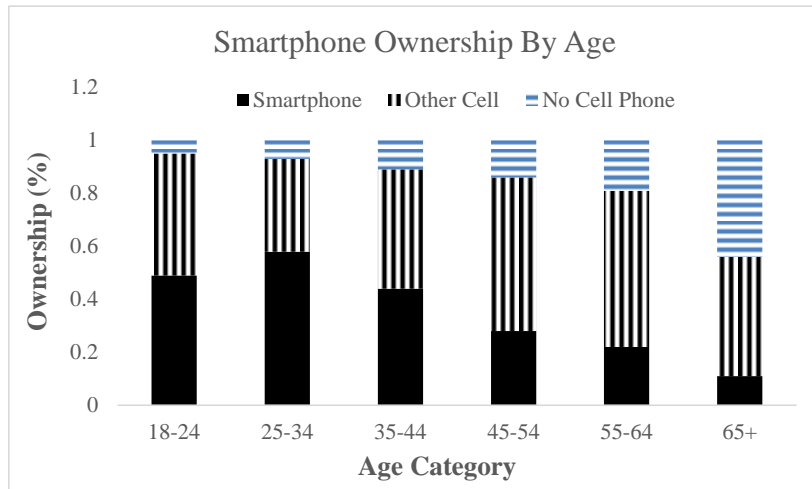


15. a.

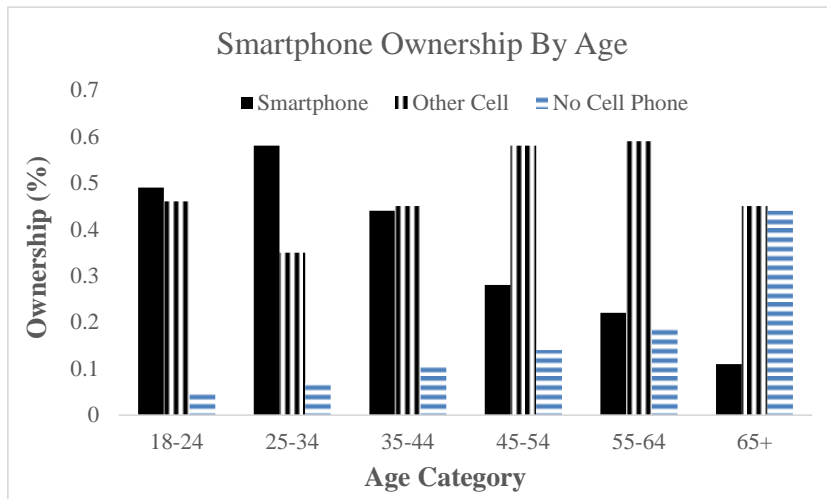


- b. All customers preferred the six-cylinder engine type, regardless of which exterior color they preferred. However, few customers who preferred red or black exterior colors preferred the four-cylinder engine, but customers who preferred the white exterior color were more evenly split between the four- and six-cylinder engine. If this is a representative sample of customers, it appears that the company will sell very few red models with four-cylinder engines, but may sell more white models with four-cylinder engines.

16. a.



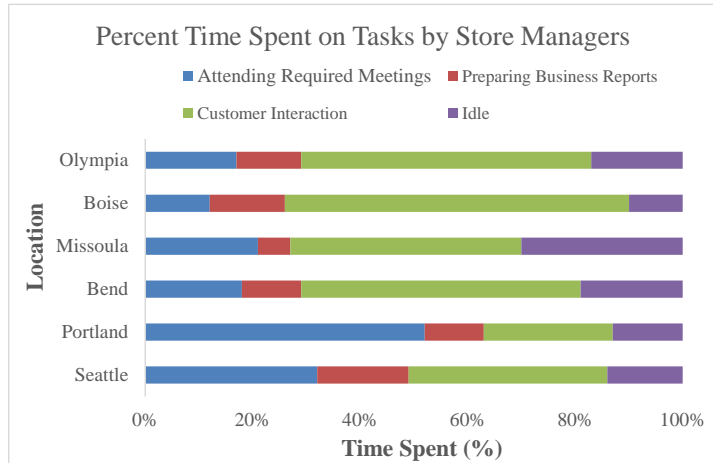
b.



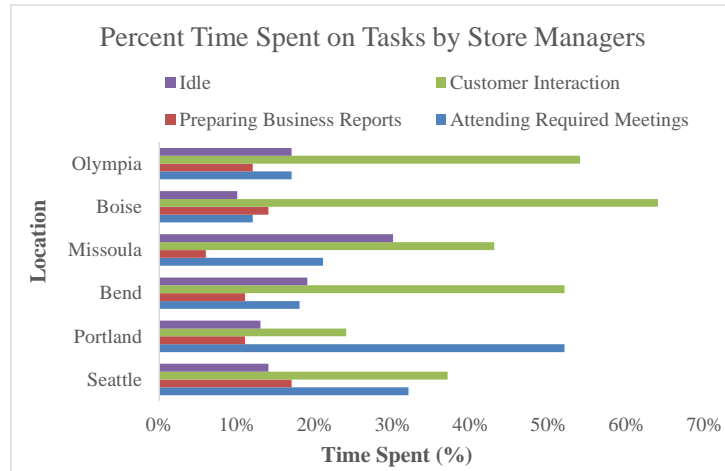
c. Younger respondents are more likely to own smartphones; older respondents are more likely to have other cell phones or to own no cell phone, particularly those in the 65+ age category.

The clustered column chart makes it easier to compare the relative percent ownership values within an age category. It is much easier to interpret from the clustered bar chart that a greater percentage of respondents age 18–24 own smartphones than other cell phones than it is to interpret this from the stacked bar chart.

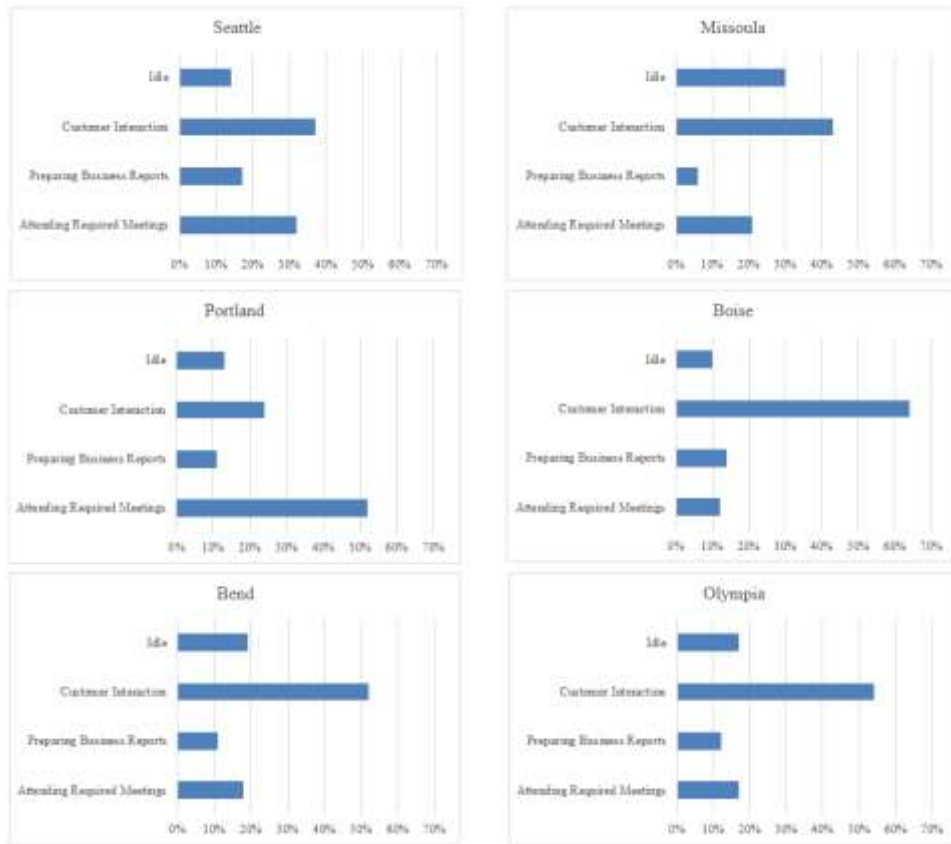
17. a.



b.



c.



Note that here we have taken care to make sure that the horizontal axes in each bar chart have the same range (0–70%) for easy comparisons among locations. We have also retained the vertical gridlines here to ease comparisons, but this is mainly a matter of stylistic preference.

- d. Both the stacked and clustered bar charts become very busy for these data, so many readers will prefer the individual bar charts. However, some readers may prefer the clustered bar chart which may make comparisons between locations easier.
- e. The managers in Boise, Bend, and Olympia spend more time (relatively) in customer interactions than managers in Seattle, Missoula, and Portland. The Portland manager in particular appears to spend an excessive time attending required meetings while the manager in Missoula appears to have more idle time.



18. a.



b. Projects 6, 3, 2 and 5 appear to be on the efficient frontier.

19. a. The screen shot below shows the first 15 respondents of the survey results with the blank cells highlighted.

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Respondent Number										
3	1	5	1	2	4	2	4	3	1	1	1
4	2	5	1	1	1	4	3	1	2	2	3
5	3	4	1		1	1	5	5	1	1	4
6	4		2		5	2	4	4	4	3	4
7	5	1	5	1	5	5		3	3	3	3
8	6	2	3	5		1	5	5	1	1	1
9	7	1	5	1	2	3	3	2	2	3	4
10	8	1		3	2	2	3	1	2	5	3
11	9	2	1	4	3	4	4	5	5	2	5
12	10	2	3	3	1	4	3	1	1	2	1
13	11	2	1	2	3	2	2	4	1	1	1
14	12	5	2	5	2	4	4	3	2	5	2
15	13	1	1	1	1	1	5	1	4	1	3
16	14	2	3	2	3	4	3	4	4	4	3
17	15	3	5	3	1	5	4	3	3	3	

b. Question 1: Respondents 4, 18  
 Question 2: Respondents 8, 74, 87, 100  
 Question 3: Respondents 3, 4, 78, 101  
 Question 4: Respondents 6, 15, 23, 76, 82, 106  
 Question 5: None  
 Question 6: Respondents 5, 70, 86  
 Question 7: Respondents 34, 81  
 Question 8: Respondents 82, 96  
 Question 9: Respondents 23, 50  
 Question 10: Respondents 15, 37, 45  
 Question 4 has the highest nonresponse rate with six respondents not providing answers.

20. a.

	A	B	C	D	E	F	G	H
1		<b>Revenue (\$)</b>						
2	<b>Company</b>	<b>Jan</b>	<b>Feb</b>	<b>Mar</b>	<b>Apr</b>	<b>May</b>	<b>Jun</b>	
3	Blue Sky Media	8995	9285	11555	9530	11230	13600	
4	Innovate Technologies	18250	16870	19580	17260	18290	16250	
5	Timmler Company	8480	7650	7023	6540	5700	4930	
6	Accelerate, Inc.	28325	27580	23450	22500	20800	19800	
7	Allen and Davis, LLC	4580	6420	6780	7520	8370	10100	
8	Smith Ventures	17500	16850	20185	18950	17520	18580	

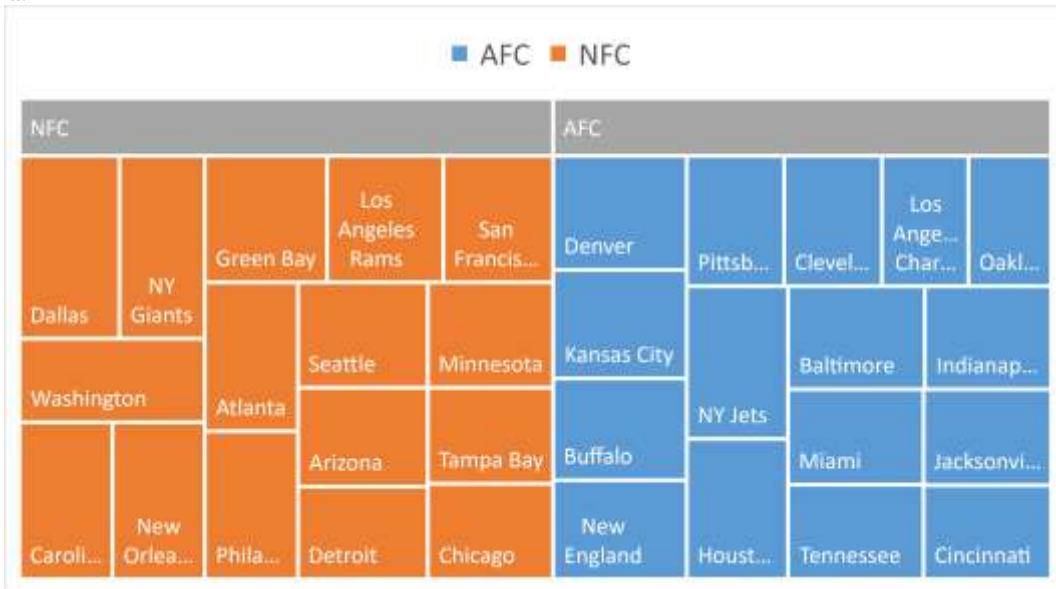
b. Timmler Company and Accelerate, Inc. appear to have generally decreasing revenues over these six months. Allen and Davis, LLC appears to have had the most consistent growth over these six months. Blue Sky Media, Innovate Technologies, and Smith Ventures have revenues that have both increased and decreased over the six months.

c.

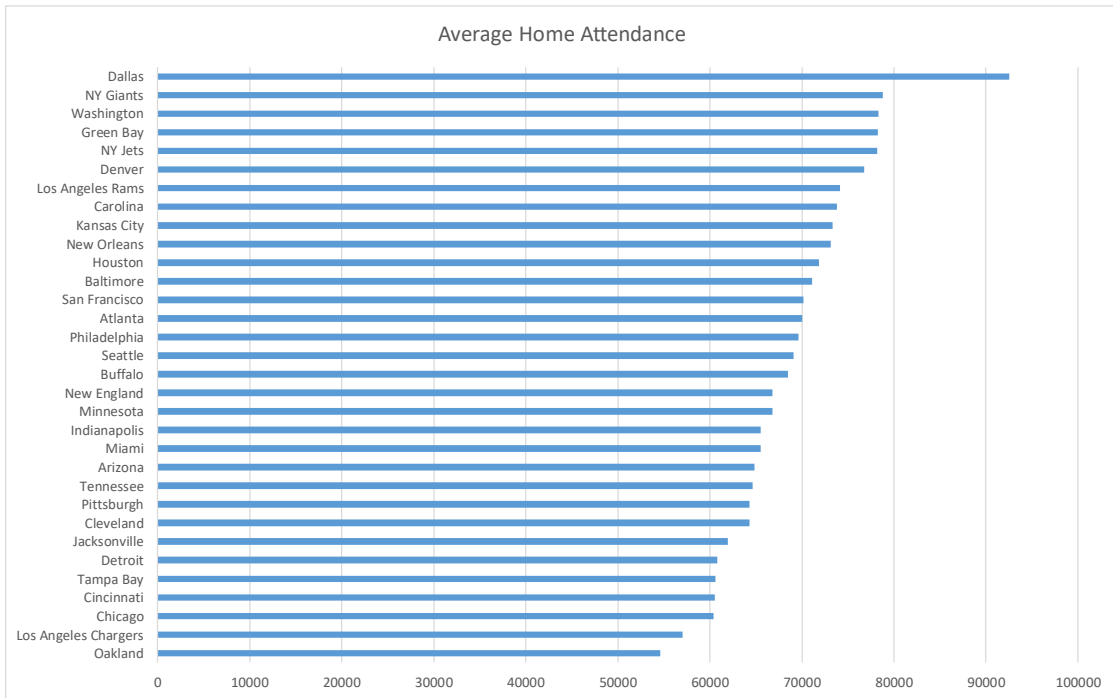
	A	B	C	D	E	F	G
1		<b>Revenue (\$)</b>					
2	<b>Company</b>	<b>Jan</b>	<b>Feb</b>	<b>Mar</b>	<b>Apr</b>	<b>May</b>	<b>Jun</b>
3	Blue Sky Media	8995	9285	11555	9530	11230	13600
4	Innovate Technologies	18250	16870	19580	17260	18290	16250
5	Timmler Company	8480	7650	7023	6540	5700	4930
6	Accelerate, Inc.	28325	27580	23450	22500	20800	19800
7	Allen and Davis, LLC	4580	6420	6780	7520	8370	10100
8	Smith Ventures	17500	16850	20185	18950	17520	18580

It is difficult to create a heat map that effectively conveys the overall trend of revenues during the six months for each company. The heat map shows the relative magnitude of the revenues which is absent from the sparklines, but the trend for each company is less apparent.

21. a.

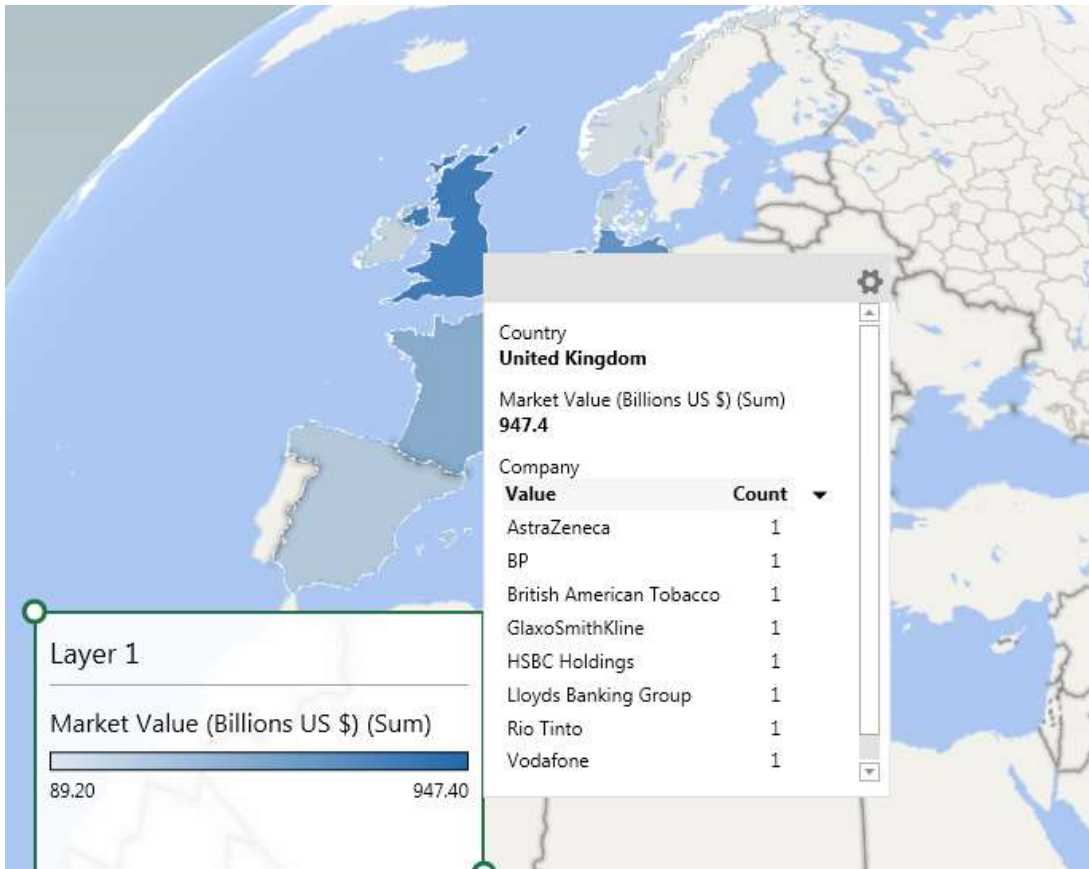


b.



c. The treemap chart makes it very easy to identify the conference affiliation for each team. However, the sorted bar chart makes it much easier to compare home attendance values for each team. Which chart is best really depends on what information is trying to be conveyed to the reader. If it's more important to know conference affiliations for each team, the treemap is likely better; if it is more important to compare attendance numbers for each team, the sorted bar chart is preferred.

22. a.



The United Kingdom has the highest total market value for Global 100 companies. The Global 100 companies located in the United Kingdom are AstraZeneca, BP, British American Tobacco, GlaxoSmithKline, HSBC Holdings, Lloyds Banking Group, Rio Tinto, and Vodafone. These companies have a total market value of 947.4 billion US \$.

b.



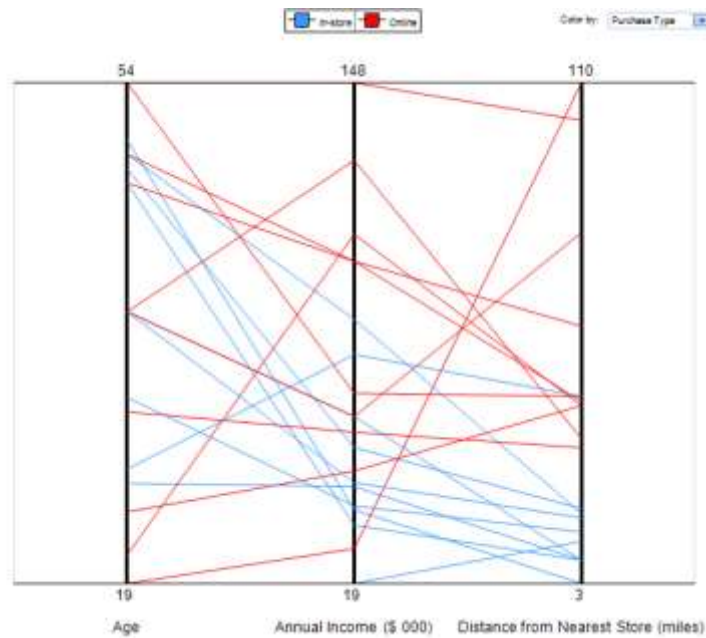
There is no longer any differentiation among the countries in Europe. This is because including North America adds the United States to the heat map. The market value of Global 100 companies in

the United States is much larger than any other country: over 8 trillion US \$. Therefore, on the heat map, the United States is shaded dark blue and all other countries have the same light gray shading. This is a common occurrence in heat maps that include an outlier as is the case here.



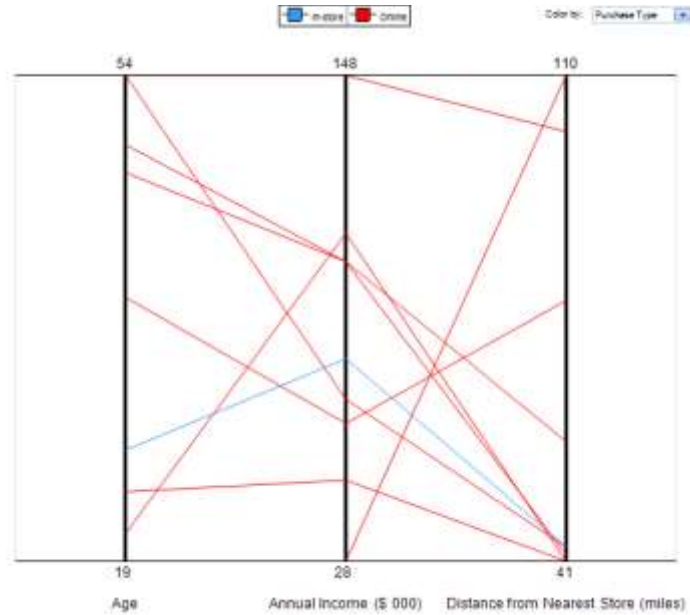
23. Online customers appear to be generally younger, have higher annual income, and live further distance away from a store.

24. a.



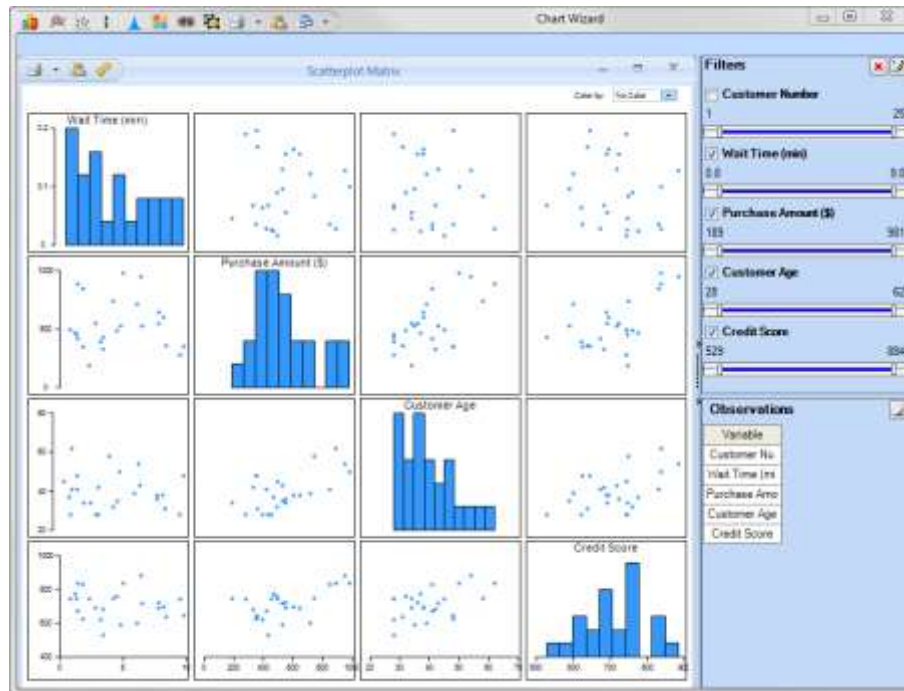
- b. There does not appear to be a clear differentiation between the ages of online versus in-store customers for electronics. However, online customers for electronics still appear to have higher annual incomes and to live farther from a store.

c.



All but one customer who lives more than 40 miles away made their purchase online.

25. a. Some possible key performance indicators include patient wait times, and current and projected utilization of the resources available (technicians, physicians, nurses, equipment, etc.). Other information that would be helpful includes number of patients currently waiting to be seen, number of patients expected to arrive today (by procedure type, etc.), number of technicians, physicians, nurses available, etc.
- b. The CEO would probably be more interested in key performance indicators such as revenue generated by each clinic, costs at each clinic separated by cost type (salaries for physicians, nurses, operating expenses, etc.). Patient satisfaction scores and/or patient wait times for each clinic could also be very helpful. It would probably be important for the CEO to also see these data plotted over recent history so that she could infer trends and plan for the future.
26. a.



- b. Purchase amount appears to have a positive relationship with customer age and credit score. Older customers and customers with higher credit scores appear to place larger purchase amounts. There are no obvious relationships between wait time and the other variables. Customer age and credit score also appear to have a positive relationship. Older customers appear to have higher credit scores.

# Chapter 3

## Data Visualization

---

### Case Problem 1: Pelican Stores

1. There were 70 promotional customers and 30 regular customers. Because there are 100 observations in the sample, the frequency and percent frequency distribution are the same. Percent frequency distributions for many of the variables are given.

#### Number of Items

Row Labels	Count of Customer
1	29.00%
2	27.00%
3	10.00%
4	10.00%
5	9.00%
6	7.00%
7	1.00%
8	1.00%
9	3.00%
10	1.00%
13	1.00%
17	1.00%
<b>Grand Total</b>	<b>100.00%</b>

#### Net Sales

Row Labels	Count of Customer
0-25	9.00%
25-50	30.00%
50-75	25.00%
75-100	10.00%
100-125	12.00%
125-150	4.00%
150-175	3.00%
175-200	3.00%
>200	4.00%
<b>Grand Total</b>	<b>100.00%</b>



### Method of Payment

Row Labels	Count of Customer
American Express	2.00%
Discover	4.00%
MasterCard	14.00%
Proprietary Card	70.00%
Visa	10.00%
<b>Grand Total</b>	<b>100.00%</b>

### Gender

Row Labels	Count of Customer
Female	93.00%
Male	7.00%
<b>Grand Total</b>	<b>100.00%</b>

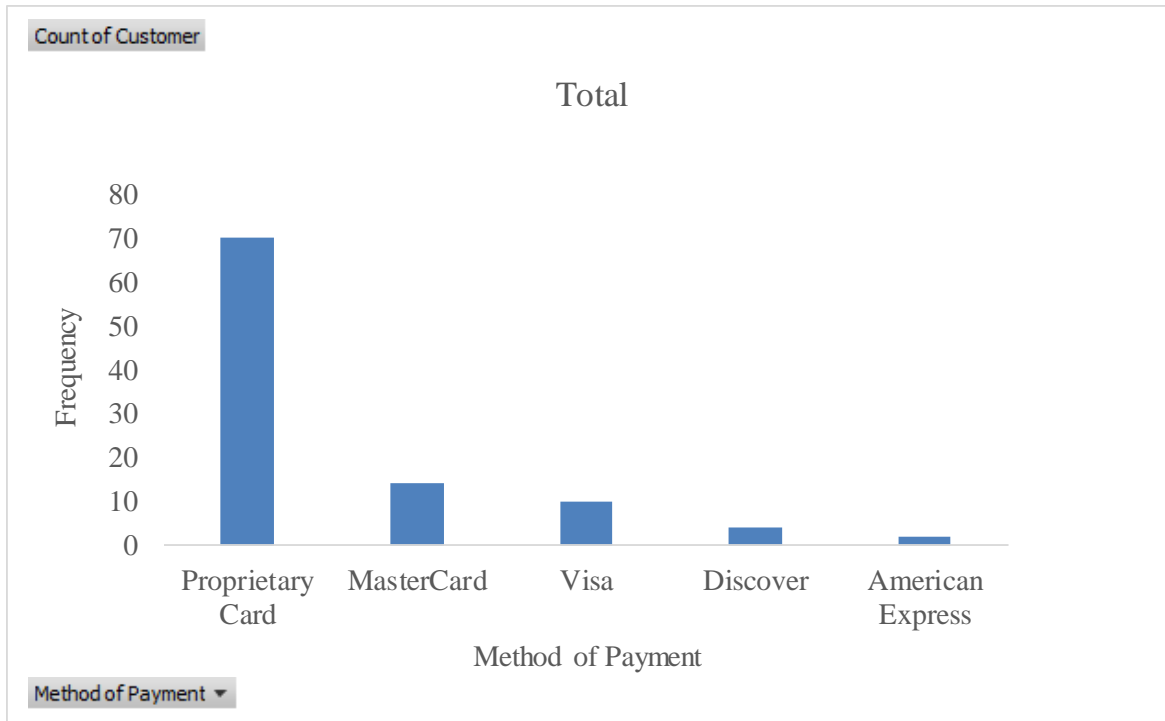
### Age

Row Labels	Count of Customer
20-29	10.00%
30-39	30.00%
40-49	33.00%
50-59	16.00%
60-69	7.00%
70-80	4.00%
<b>Grand Total</b>	<b>100.00%</b>

These percent frequency distributions provide a profile of Pelican's customers. Many observations are possible, including:

- A large majority of the customers use National Clothing's proprietary credit card.
- More than half of the customers purchase one or two items, but a few make numerous purchases.
- The percent frequency distribution of net sales shows that 61% of the customers spent \$50 or more.
- Customers are distributed across all adult age groups.
- The overwhelming majority of customers are female.
- Most of the customers are married.

2.

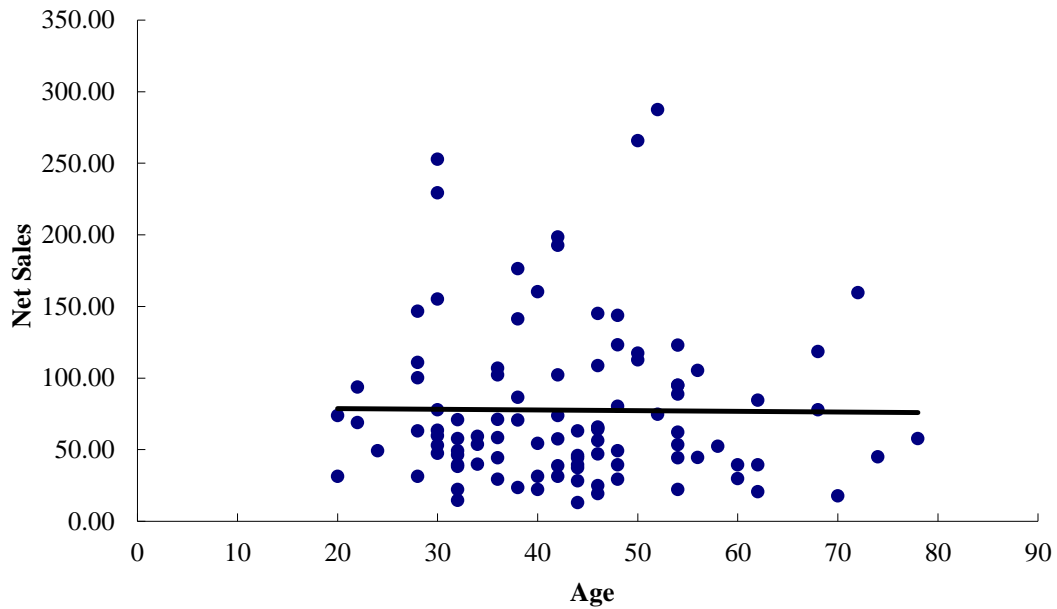


3. A crosstabulation of type of customer versus net sales is shown.

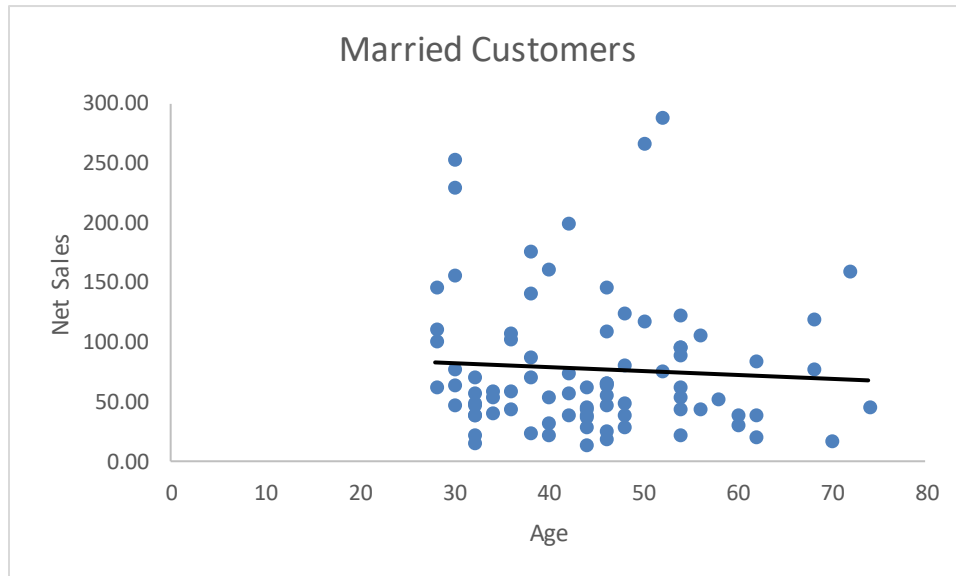
Count of Customer	Column Labels	0-25	25-50	50-75	75-100	100-125	125-150	150-175	175-200	>200	Grand Total
Promotional		7	17	17	8	9	3	2	3	4	70
Regular		2	13	8	2	3	1	1			30
<b>Grand Total</b>		<b>9</b>	<b>30</b>	<b>25</b>	<b>10</b>	<b>12</b>	<b>4</b>	<b>3</b>	<b>3</b>	<b>4</b>	<b>100</b>

From the crosstabulation it appears that net sales are larger for promotional customers.

4. A scatter diagram of Net Sales versus Age is shown as follows. A trendline has been fitted to the data. From this, it appears that there is no relationship between net sales and age.

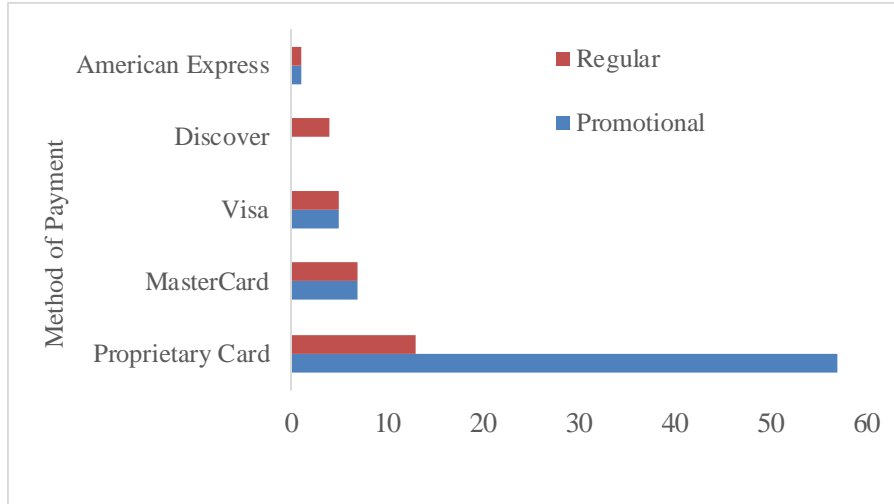


5. There are several ways we can examine this relationship using charts. One way is to sort the observations by marital status and then create scatter charts for net sales and age for married customers separately from single customers. The two scatter charts appear below. We have fitted a trendline to the data in each chart.



By putting these on separate scatter charts we see that there may be a difference in the relationship depending on whether the customer is married or single. There appears to be a slight decrease in net sales with higher age for married customers, but net sales appear to increase with age for single customers.

6. The side-by-side bar chart appears below.



We can see the number of regular customers and promotional customers who use methods of payments other than the National Clothing proprietary card are nearly identical. But many more promotional customers use the National Clothing proprietary card than regular customers. This is likely because the promotion was sent to customers of other National Clothing stores, and one way to identify such customers is to target customers who have a National Clothing credit card but who have not made previous purchases from Pelican Stores. So, the targeted customers are more likely to use their National Clothing proprietary card.

## Case Problem 2: Movie Theater Releases

This case provides the student with the opportunity to use tabular and graphical presentations to analyze data from the movie industry. Developing and interpreting frequency distributions, percent frequency distributions, and scatter diagrams are emphasized. The interpretations and insights can be quite varied. We illustrate some possible insights below.

### Frequency Distributions

We can view tabular summaries of these variables as frequency distributions by using the PivotTable tool in Excel. Because there are 100 observations in the sample, the frequency and percent frequency distribution are the same. The choice of the classes for frequency distributions can be expected to vary. (Note that the Excel PivotTable tool omits bins where no observations are present when using the PivotTable Grouping function by default. To display empty bins, right-click on any item in the **Row Labels** column, click **Field Settings...**, click the **Layout & Print** tab and select the check box for **Show items with no data.**) The frequency distributions we developed are as follows:

Opening Gross Sales (\$ millions)

Row Labels	Count of Movie Title
<0	
0-10	14
10-20	34
20-30	22
30-40	10
40-50	5
50-60	3
60-70	1
70-80	2
80-90	1
90-100	
100-110	2
110-120	
120-130	
130-140	3
140-150	
150-160	1
160-170	1
170-180	1
>180	
<b>Grand Total</b>	<b>100</b>

Total Gross Sales (Millions)

Row Labels	Count of Movie Title
<0	
0-50	34
50-100	36
100-150	11
150-200	6
200-250	3
250-300	1
300-350	3
350-400	3
400-450	1
450-500	1
500-550	1
>550	
<b>Grand Total</b>	<b>100</b>

Number of Theaters

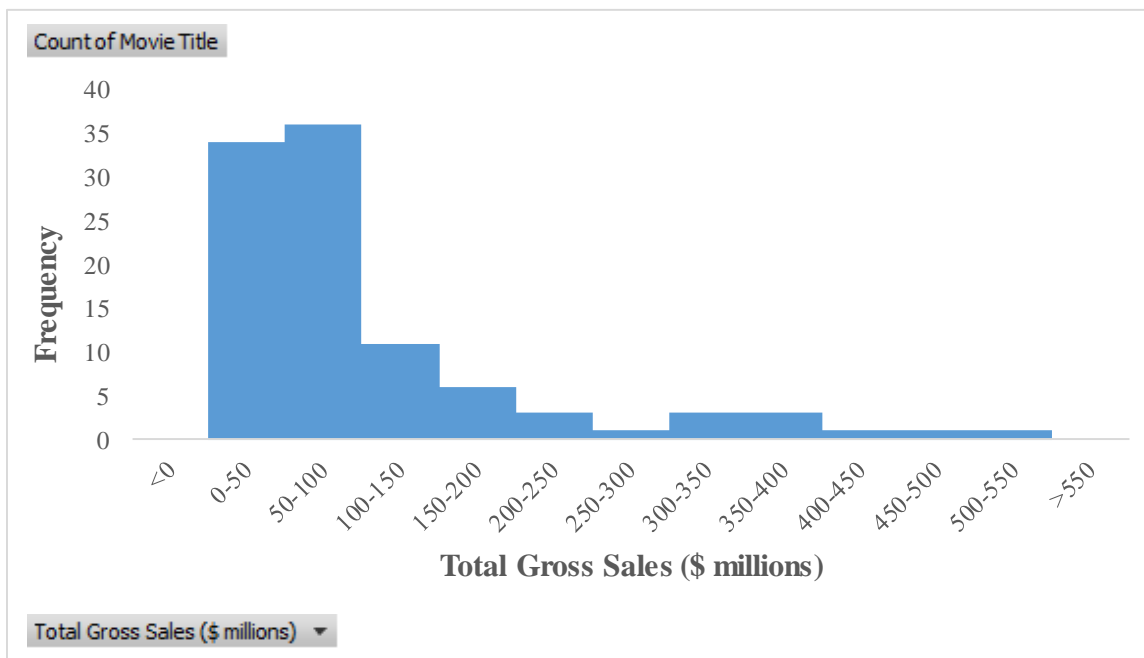
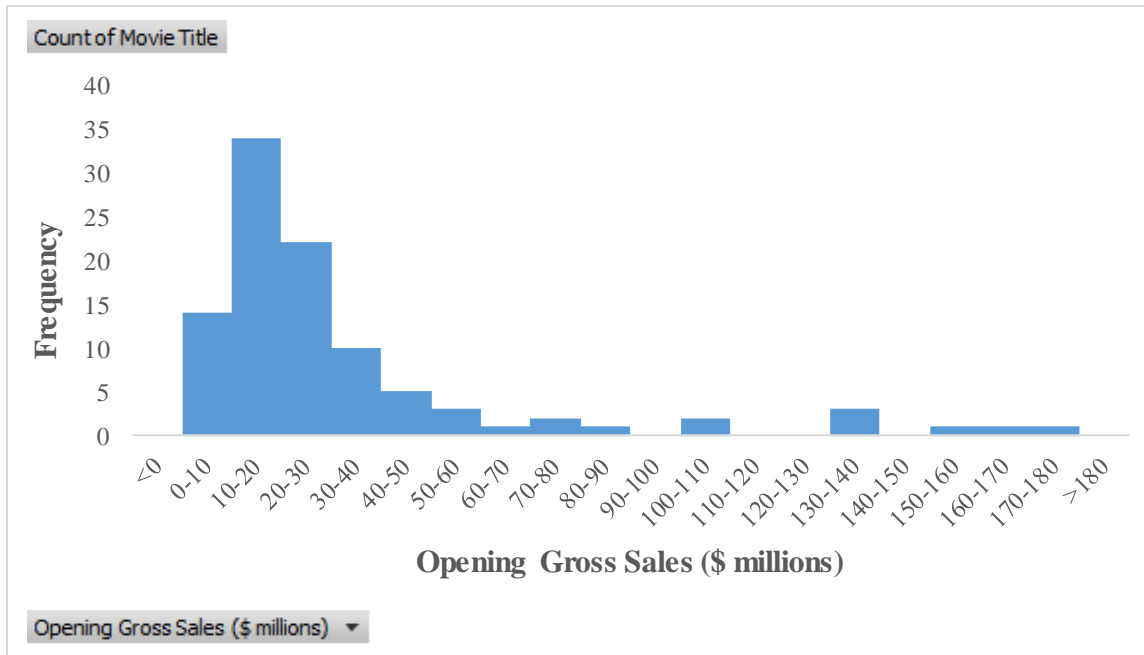
Row Labels	Count of Movie Title
<1000	
1000-1499	1
1500-1999	4
2000-2499	6
2500-2999	17
3000-3499	37
3500-3999	21
4000-4500	14
>4500	
<b>Grand Total</b>	<b>100</b>

Weeks in Release

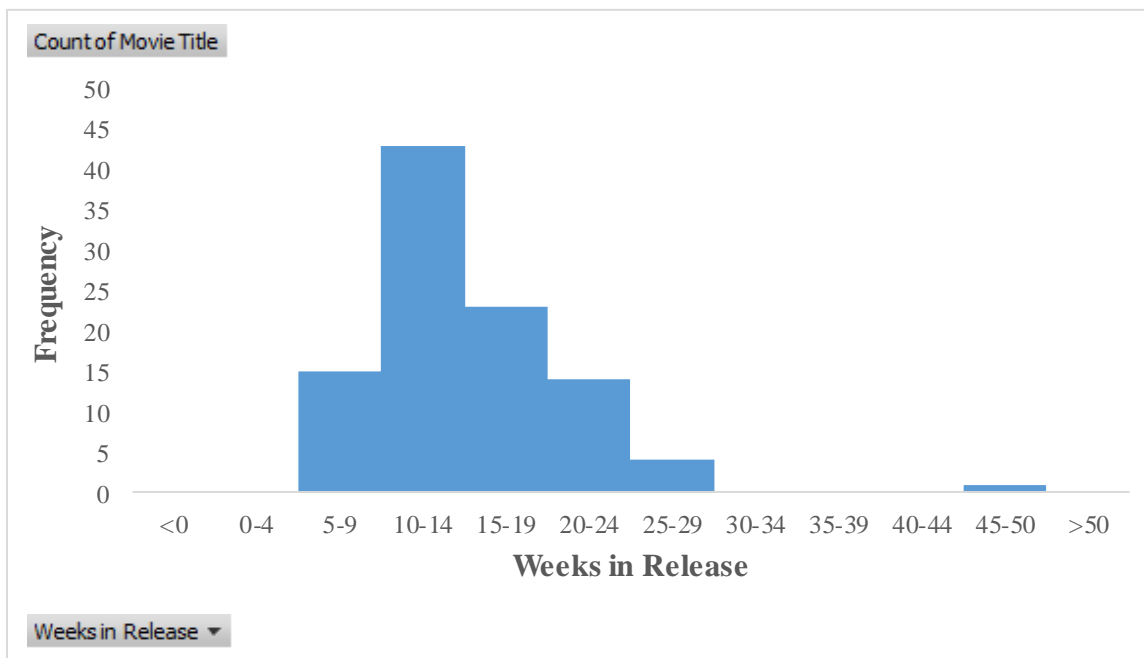
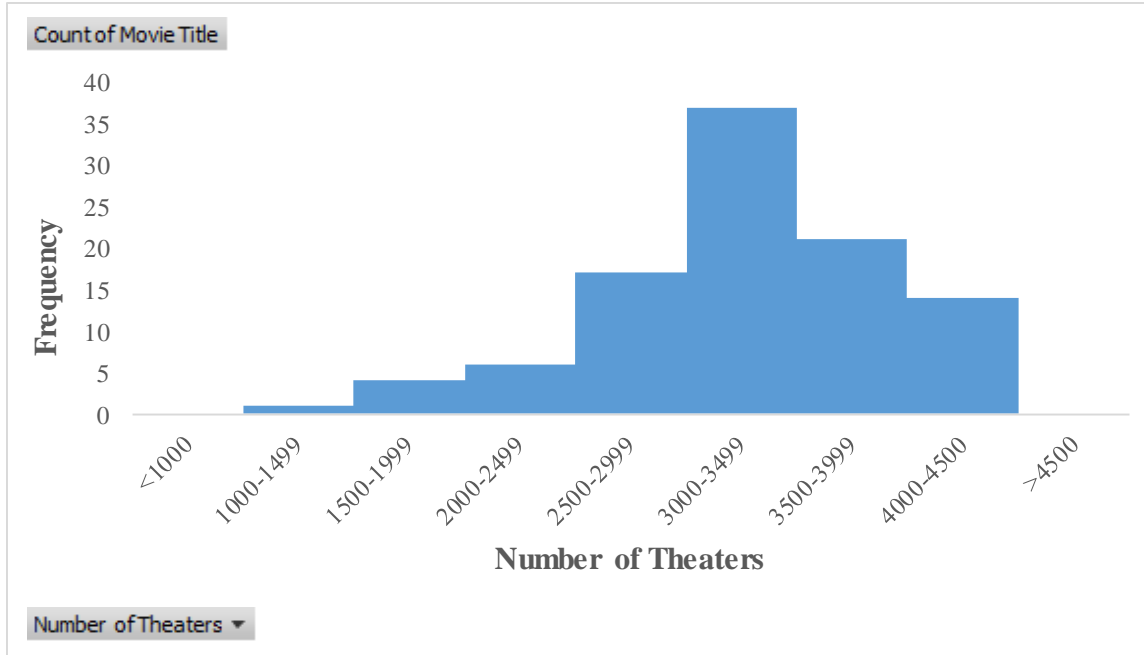
Row Labels	Count of Movie Title
<0	
0-4	
5-9	15
10-14	43
15-19	23
20-24	14
25-29	4
30-34	
35-39	
40-44	
45-50	1
>50	
<b>Grand Total</b>	<b>100</b>

**Histograms**

We can examine the variables graphically by creating histograms. We can create histograms in Excel for these variables by creating bar charts of the PivotTable outputs from above.







*Interpretation*

**Opening Weekend Gross Sales** Numerous movies have somewhat low opening weekend gross sales, while a relatively few (8%) have an opening weekend gross sales of \$100 million or more. Only 3% had opening weekend gross sales of \$150 million or more. Eighty percent of the movies had opening weekend gross sales less than \$40 million, and 92% of the movies had opening weekend gross sales less than \$100 million. Using the terminology from Chapter 2, we say that this distribution is skewed to the right.

**Total Gross Sales** Again, the majority of the movies have relatively low total gross sales with 70% of movies having gross sales less than \$100 million and 91% less than \$300 million. Highly successful blockbuster

movies are rare. Total gross sales of more than \$400 million occurred only 3% of the time, and gross sales of more than \$500 million occurred only 1% of the time. Unless there is something unusually attractive about the movie, a total gross sales less than \$100 million appears typical. Using the terminology from Chapter 2, we say that this distribution is also skewed to the right.

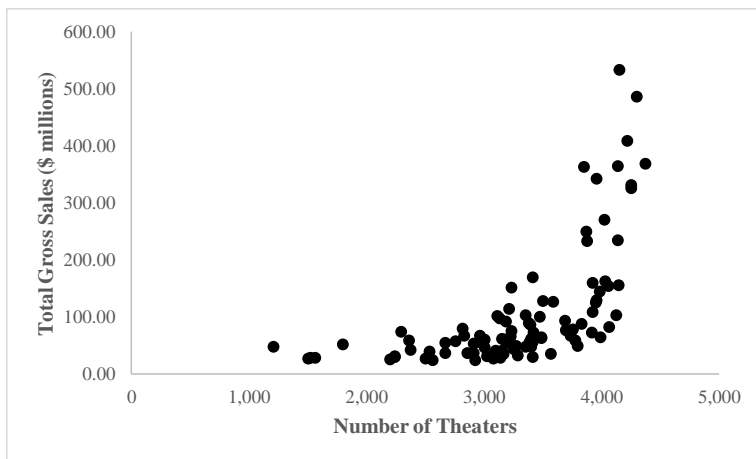
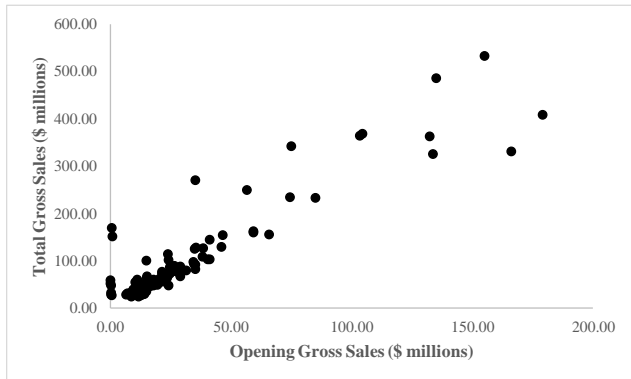
**Number of Theaters** The number of theaters range from slightly more than 1,000 to almost 4,500. Eighty-nine percent of the movies had large market exposure, playing in 2,500 or more theaters. No movies were in fewer than 1,000 theaters, and only 11% were in fewer than 2,500 theaters. Most top movies in 2016 appeared to receive large market exposure in 2,500 or more theaters. Using the terminology from Chapter 2, we say that this distribution is skewed to the left.

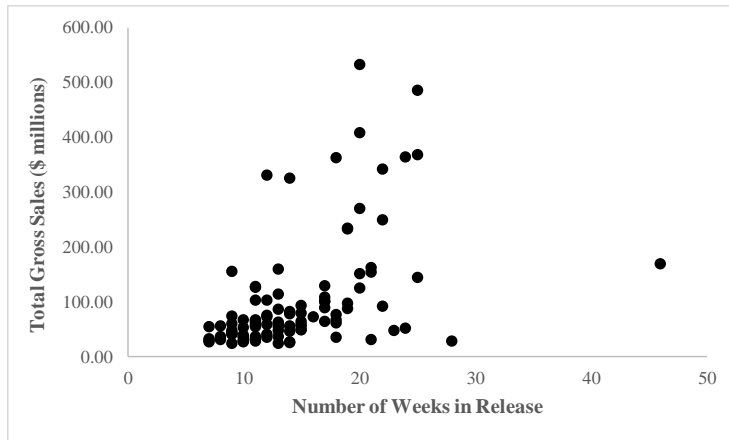
**Number of Weeks in Release** Almost all movies in 2016 spent at least 10 weeks in release. Only 15% of movies in 2016 spent fewer than 10 weeks in release. One movie (*Hidden Figures*) spent much longer in release than any other movie at 46 weeks. Using the terminology from Chapter 2, we say that this distribution is skewed to the right, but less skewed than the distributions on sales.

**General Observations** The data show there are relatively few high-end, highly successful movies. The financial rewards are there for the pictures that make the blockbuster level. But the majority of movies will have relatively low opening weekend gross sales and low total gross sales. Movies being shown in more than 2,500 theaters and movies that spend at least 10 weeks in release are common.

Scatter Diagrams

Three scatter diagrams are suggested to show how Total Gross Sales is related to each of the other three variables.





### Interpretation

**Opening Weekend Gross Sales** The scatter plot of total gross sales and opening weekend gross sales shows a strong positive relationship. Movies with the highest total gross sales were those with the highest opening gross sales. How a movie does during its opening weekend should be a strong predictor of how the movie will do in terms of total gross sales. Note in the scatter diagram that the majority of the movies show a low opening weekend gross sales and a low total gross sales.

**Number of Theaters** The scatter plot of the total gross sales and number of theaters also shows a positive relationship. For movies playing in fewer than 3,500 theaters, the total gross sales were significantly less than those movies playing in more than 3,500 theaters. If the movie is shown in more theaters, higher total gross sales are anticipated. For movies playing in more than 3,500 theaters, the positive relationship is especially strong. This scatter chart also appears to show a nonlinear relationship because movies playing in the most theaters increase in total gross sales rapidly compared to those playing in fewer theaters. This scatter chart also shows that the relationship may be nonlinear.

**Number of Weeks in Release** The scatter plot of the total gross sales and number of weeks in release shows a positive relationship, but this relationship appears to be the weakest of the three relationships studied. Generally, the more successful movies with higher gross sales are in release for more weeks. However, this is not always the case. The longest released movie (*Hidden Figures*) had less in total gross sales than many movies that had shorter release times. And many movies that were in release for more than 20 weeks had less total gross revenue than those with fewer than 20 weeks in release. This suggests that in some cases blockbuster movies with high gross sales may run their course quickly and not have an excessively long run in release. At the same time, perhaps quality movies with a limited audience may not generate the high total gross sales but may still show a run of 20 or more weeks. The number of weeks in release does not appear to be the best predictor of total gross sales.