INSTRUCTOR'S SOLUTIONS MANUAL

ZHWEI ZHU

University of Louisiana at Lafayette

STATISTICS FOR BUSINESS: DECISION MAKING AND ANALYSIS

THIRD EDITION

Robert Stine

Wharton School of the University of Pennsylvania

Dean Foster

Emeritus, Wharton School of the University of Pennsylvania

Pearson

330 Hudson Street, NY NY 10013

Reproduced by Pearson from electronic files supplied by the author.

Publishing as Pearson, 330 Hudson Street, NY NY 10013

1 16

ISBN-13: 978-0-13-449735-8
ISBN-10: 0-13-449735-X

Contents

Chapter 2: Data

Mix and Match

1. Variable Name: brand of car; Type: nominal; Cases: drivers

2. Variable Name: household income; Type: numerical; Cases: households

3. Variable Name: color preference; Type: nominal; Cases: consumers in focus group

4. Variable Name: customer counts; Type: numerical; Cases: outlets of retail chain

5. Variable Name: item size; Type: ordinal; Cases: unknown (could be stocks in stores or purchase amounts)

6. Variable Name: shipping cost; Type: numerical; Cases: unknown (could be a time series or could be the costs for various items or destinations)

7. Variable Name: stock price; Type: numerical; Cases: companies (though the question is vague)

8. Variable Name: number absent; Type: numerical; Time Series Frequency: days

9. Variable Name: Sex; Type: nominal; Cases: respondents in survey

10. Variable Name: Education; Type: ordinal; Cases: customers

True/False

11. False. Zip codes are numbers, but these numbers are used only for identification and would not have any numerical meaning.

12. True.

13. False. Cases is another name for the rows in a data table.

14. True.

15. True.

16. False. A row holds an observation.

17. False. A Likert scale is used for ordinal data.

18. True.

19. False. Aggregation collapses a table into one with fewer rows.

20. True.

Think About It

21. (a) The data are cross sectional.
(b) The variables are whether the employee opened an IRA (nominal) and the Amount saved (numerical with dollars as the units).
(c) Did employees respond honestly, particularly when it came to the amount they reported to have saved?

22. (a) The data are cross sectional.
(b) The variables are Reaction to increase (categorical, or perhaps ordinal if asked to rate the chance of moving to another bank), Current balance and other aspects of the customer that would be useful additions to the data. (Bank may not care if it loses unprofitable customers.)
(c) How many customers responded to the questionnaire? Were their responses about leaving the bank sincere?

23. (a) The data are cross sectional.
(b) The variable is the Service rating (ordinal most likely, using a Likert scale).
(c) With only 450 replying, are the respondents representative of the other guests?

24. (a) The data are cross sectional.
(b) The variables are whether a coupon was used (nominal) and Purchase amount (numerical with dollars as the units).

(c) How were these homes chosen? Was there a time limit on redemption?

25. (a) The data are a time series.
(b) The variable is the Exchange rate of the US dollar to the Canadian dollar (numerical ratio of currencies).
(c) Are the fluctuations in 2016 typical of other years?

26. (a) The data are a time series.
(b) The variable is the Average time spent on the lot for ten car models (numerical for each model).
(c) Did dealers accurately report this information? Were all dealers surveyed, or just some of them? If it's a survey, did it concentrate more in some regions than others?

27. (a) The data are cross-sectional.
(b) The variables are the Quality of the graphics (ordinal from bad to good) and the Degree of violence (ordinal from none to too much).
(c) Did some of the participants influence the opinions of others?

28. (a) The data are cross sectional.
(b) The variables are Income (numerical with units in dollars), Sex (nominal), Location (nominal), Number of cards (numerical count) and Profit (numerical with dollars as the units, derived from other data).
(c) Why were these accounts sampled and not all of them?

29. (a) The data are cross sectional (though they could be converted to a time series).
(b) The variables are Name (nominal), Zip code (nominal), Region (nominal), Date of purchase (nominal or numerical, depending on the context; the company could compute the average length of time since the last purchase), Amount of purchase (numerical with dollars as the units) and Item purchased (nominal).
(c) Presumably the region was recorded from the zip code.

30. (a) The data are a time series.
(b) The variable is Vehicle type (nominal or ordinal as compact, regular, large and SUV)
(c) The mix of cars on the weekend may not be the same as on a weekday. Do employees get an accurate count since they have other things to do as well?

4M Economic Time Series

(a) Answers will vary, but should resemble the following.
By merging the data, we can see how sales of Best Buy move along with the health of the general economy. If sales at Best Buy rise and fall with disposable income, we might question the health of this company if the government predicts a drop in the amount of disposable income.

(b) A row in the data from FRED2 describes the level of disposable income in a month whereas a row in the company-specific data is quarterly, summarizing a quarter (3 months).

(c) The columns are both numbers of dollars, but with different multipliers. The national disposable income is in billions (so the value for January 2010 means that consumers have $11.041 trillion annually to spend). The quarterly sales are in millions (so Best Buy's net sales in the first quarter of 2010 were $3.036 billion).

(d) We can aggregate the monthly numbers into a quarterly number such as by taking an average (FRED2 will do this for you if you want to return to the web site). Alternatively, we could take the quarterly number and spread it over the months. That's a bit hard to do, so the first path is more common.

(e) Name the columns Net Sales ($ billion) and Disp Income ($ trillion) and scale as shown previously. That avoids lots of extraneous zeros if you were, for example, to label them all as dollars. The dates might best be recorded in a single column as, say, 2010:1, 2010:2, and so forth, or in the style shown in the following table.

(f) Here's the merged data table for 2010:

Quarter	Net Sales ($ billion)	Disp Income ($ trillion)
Jan-2010	$3.036	$33.124
Apr-2010	$3.156	$33.593
Jul-2010	$3.233	$33.860
Oct-2010	$4.214	$34.277

(g) Sales at Best Buy rocket up in the fourth quarter (30% higher during the holiday season), but consumers don't have that much more money to spend. Looks like some people spend a lot more during the holidays, no surprise there!

4M Textbooks

(a) Various sources report that books cost about $100 per class. In 2003, U.S. Senator Charles E. Schumer of New York released a study showing that the average New York freshman or sophomore pays $922 for textbooks in a year. So reducing the cost 5% would save $46.10 a year and by 10% would save $92.20 a year.

(b) Your table should have headings like these. You should use the names of the stores you shopped at if different from these. The first two columns are nominal, with the first identifying the book and the second giving the label. The two columns of prices are both numerical.

Book Title	Type	Price at Amazon	Price at B&N

(c) These will vary. Presumably, you've got five textbooks from your current classes. Hopefully, you've also got some other personal books. For popular books, you might consider books on one of the best-seller lists or those at the top of the lists offered on-line.

(d) You may have to change the list of books, particularly for textbooks. Some on line sites have a limited selection of these.

(e) You should include all of the relevant costs. Some Internet retailers add high shipping costs.

(f) Again, answers will vary depending on the choice of books and the choice of stores. The key to notice is the value of comparison. Because you've got two prices for the same books, you can compare apples to apples and see whether one retailer is systematically cheaper than the other

Chapter 3: Describing Categorical Data

Mix and Match

In each case, unless noted, bar charts are better to emphasize counts whereas pie charts are better to communicate the relative share of the total amount.

1. Proportion of autos: pie chart is the most common; a bar chart or Pareto chart can also be used.

2. Types of defects: Pareto chart (a bar chart with the categories sorted in order of the most common defect)

3. Coupons: bar chart or Pareto chart (these are counts) or perhaps a table (only three values)

4. Type of automobile: bar chart or Pareto chart (counts) or pie chart (shares)

5. Destination: bar chart or Pareto chart (counts) or pie chart (shares)

6. Hanging up: Pareto chart (counts)

7. Excuses: Pareto chart (counts)

8. Brand of computer: bar chart (counts) or pie chart (shares)

9. Software: pie chart (shares) or perhaps a table (only three values)

10. Camera: bar chart (counts), pie chart (shares), or a table (only three values)

11. Ratings: Bar chart or table (only four values). Because the values are ordinal, avoid a pie chart.

12. Loans: Bar chart or table (only three values). Because the data is ordinal, it should not be put into a pie chart – even though the plot shows shares.

True/False

13. True, but only in general. For variables with few categories, a frequency table is often better, particularly when the analysis requires knowing the detailed frequencies.

14. False. The measure of variation has to be within one category.

15. False. The frequency is the count of the items.

16. False. A relative frequency is a proportion.

17. True. It would be false if the variable were ordinal; you should not put the shares of an ordinal variable into a pie chart.

18. False. The proportion must match the relative frequency.

19. True.

20. False. It has fewer bars.

21. True.

22. False. The median only applies to ordinal variables and identifies the category of the middle value.

Think About It

23. The message is that customers tend to stick with manufacturers from the same region. Someone trading in a domestic car tends to get another domestic car whereas someone who trades in an Asian car tends to buy another Asian car. There's not a lot of switching of loyalties. The more subtle message, one that is disturbing to domestic car makers, is that those who own Asian cars are more loyal (78% buy another Asian car compared to 69% who stick with a domestic car). That makes it hard for domestic manufacturers to win back customers, even if they improve the quality of their cars.

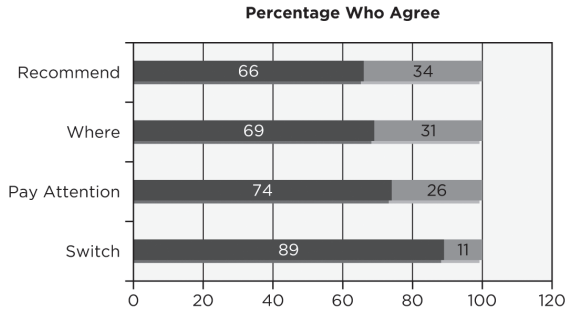
24. The answer is yes. Since lighting makes up 37% of the use of electricity, reducing the demand for electricity by using more efficient bulbs can have a substantial impact. Compact fluorescent bulbs produce the same amount of light with much less, say one-quarter, of the electricity used by an incandescent bulbs. Less energy also implies less heat and lower cooling costs. That said, the benefit of these savings for utilities is less pronounced because these savings happen mostly at night, not during the times of peak load that occur during the daytime.

25. This is a bar chart if you think about the underlying data as labeling the dollars held in these countries. The intent of the plot is to show the relative sizes of these counts, comparing the shares of U.S. debt held in these countries.

26. No, this is not a bar chart in the sense of this chapter. The chart uses bars to show a very short time series with five data points, the annual revenue in 2011-2015. Hence, it is a time plot that uses bars to show the data.

27. (a) No, these categories are not mutually exclusive. These percentages summarize four dichotomous variables, not one variable.

(b) Divided bars such as these might work well. This style is commonly used in reporting opinion poll results in the news. Sorting the values so that the percentages are in order also makes for a cleaner presentation.



28. (a) No. Each customer could report several of these items, so the categories are not mutually exclusive.
(b) A figure such as the divided bars used in Exercise 27 would be useful to illustrate the varying shares.

29. No. These percentages only list the percent of executives that report each problem. The categories are not mutually exclusive; some of the executives listed several issues.

30. The percentages do not add to 100; we need another category (which has a 9% share of the market).

31. A bar chart would not be appropriate for this situation since the variable, the amount spent by the last 200 customers, is a continuous variable.

32. This grouping data into categories is one step in constructing a frequency table. A histogram should be used since it reports a continuous variable, not a bar chart (a bar chart should be used when you deal with category variable).

33. The bar chart would have one very long bar (height 900) and five shorter bars of height 20 each. The plot would not be very useful, other than to show the predominance of one category.

34. A pie chart would devote 90% of its area to the main category and divide the remaining area into five small slices, each with equal area of 2%.

35. The bar chart would have five bars, each of the same height.

36. The bar chart. It would be hard to tell in the pie chart that the slices were of the same size (however, if the slices were labeled with the percentages it would be the same).

37. A bar chart is preferred because the categorical nature of the variable. A pie chart also can be used to present this data. A frequency table is not appropriate in this situation.

38. With so many categories (the 51 states, including Washington D.C.), some aggregation by region might be useful. Alternatively, it might be good to highlight the most common states, and combine the rest together into a separate, other category. A bar chart or pie chart could be used, and a frequency table would be fine if there were only a few states represented.

39. The mode is Public. There's no median for this chart since this is nominal data.

40. The East is the modal location. To find the median size, notice there are 50 sizes given, so the median is the size in position 25 or 26. Both lie in the category 10,000 to 19,999. The percentages of enrollment categories, not counts, should be shown in a pie chart.

41. The manufacturers want to know the modal preference because it identifies the most common color preference. Color preferences cannot be ordered, the median color preference can't be defined.

42. A median rating of Excellent implies that at least one-half rated the service as Excellent. A modal rating of Excellent implies that this is the most common rating, but far fewer than half might have picked this rating.

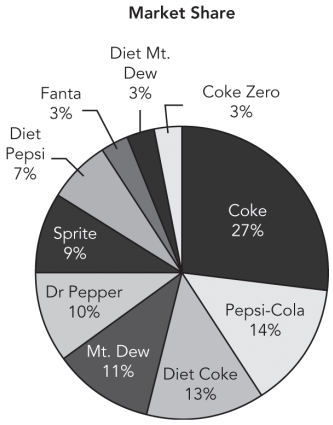
43. The radius of the circle for consumer electronics, for example, would have to be sqrt (10.5/9.1) times the radius of the circle for cell phones, sqrt(10.5/6.3) times the radius for chips, and sqrt(10.5/5.7) times the radius for LCD panels.

44. For instance, render as dollar bills with area determined by the amounts.

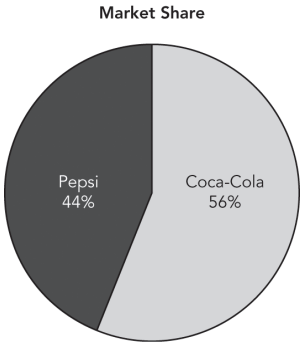
You Do It

45. (a) It probably accumulates case sales by brand over some period, such as daily or weekly. It is unlikely that every case is represented by a row.

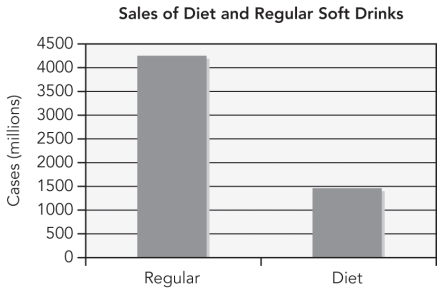
(b) A pie chart emphasizes shares.



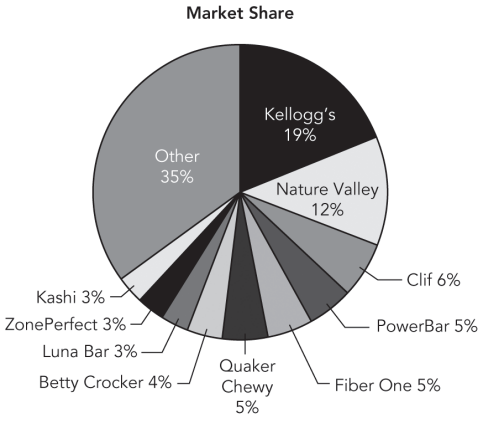
(c)



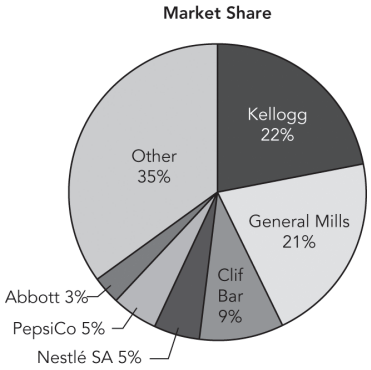
(d)



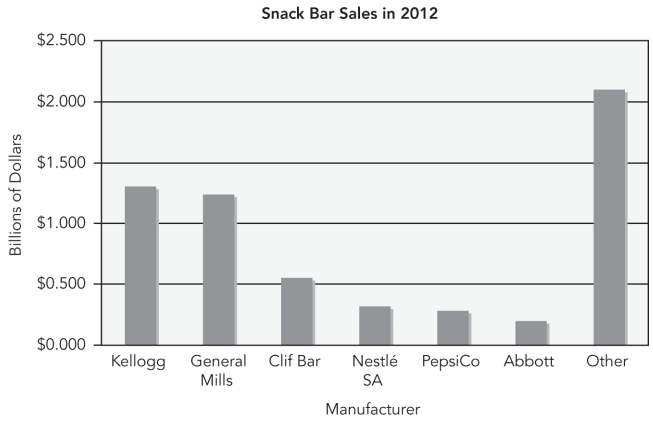
46. (a)



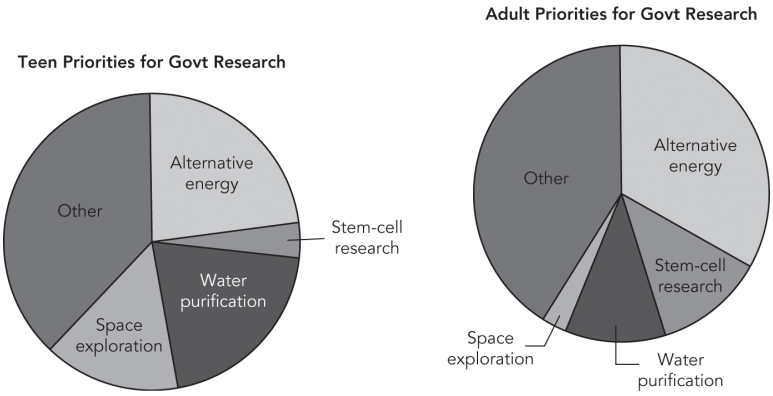
(b)



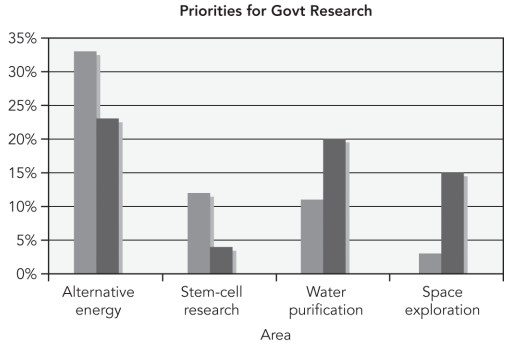
(c)



47. (a) The Other category forms an additional row in the tables so that each column adds up to 100%. The addition of this extra row makes up a big part of both pie charts.



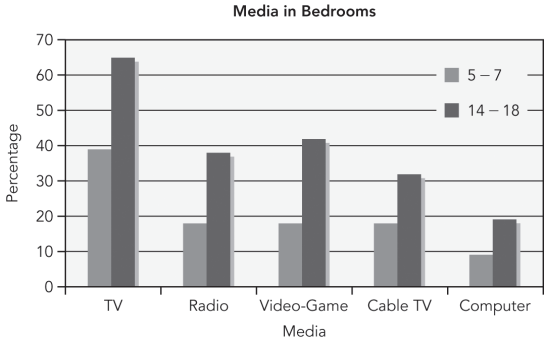
(b) The side-by-side bar chart works well for this. Notice that we no longer need the Other category that dominates the pie charts.



(c) No, because the categories would no longer partition the cases into distinct, non-overlapping subsets. A pie chart should only be used to summarize mutually exclusive groups.

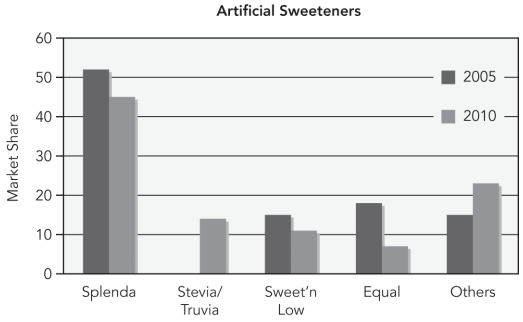
48. (a) The totals within a row do not sum to 100%. The columns give the proportion with different types of media, and these can sum to more than 100% as well. The represented categories do not divide the homes into different groups; a bedroom could have all of these media.

(b) The side-by-side bar chart shows more of every type of media in the rooms of older children.

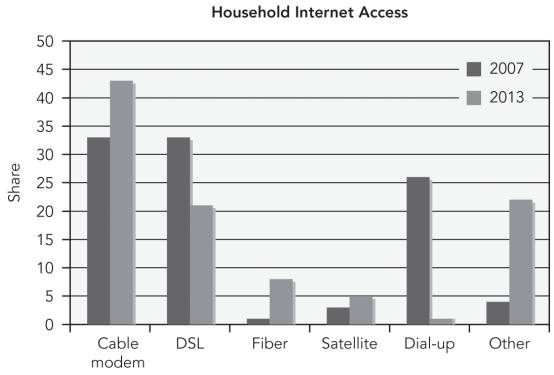


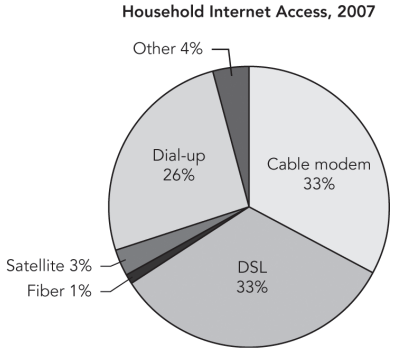
(c) The big adoption of games appears to happen in the 8-13 age range.

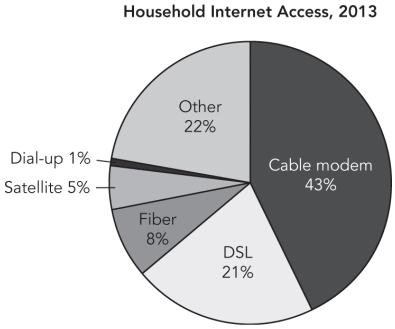
49.



50. (a) The pie charts are favored slightly over the side-by-side bar charts.

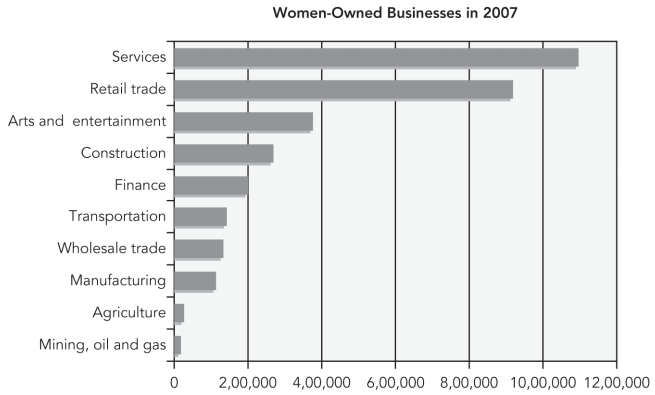


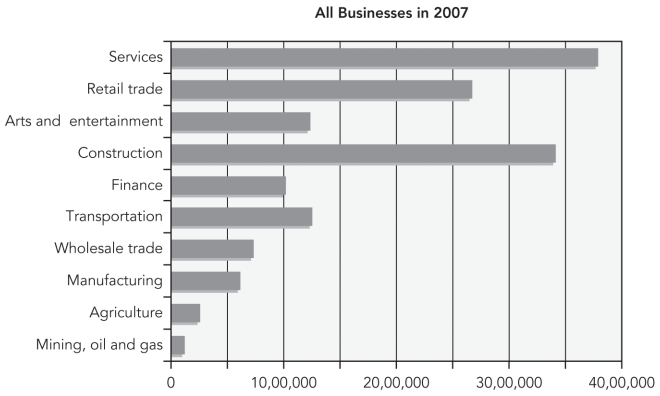




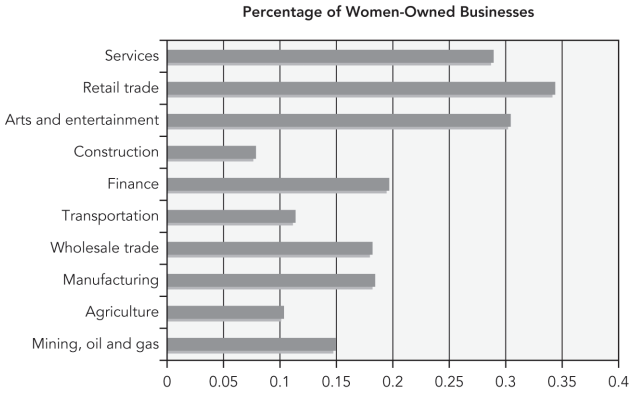
(b) No; to compare, count, not percentages, are needed.

51. (a)





(b)

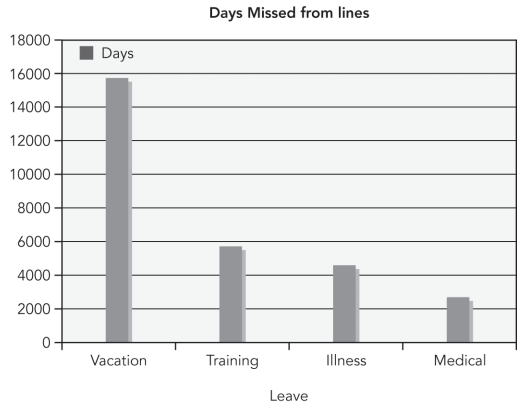


52. (a) No, the percentages do not add; some respondents gave more than one reason.
(b) No, unless you think there is a natural way to prioritize the reasons.
(c) A bar chart, with the length given by the percentage.

53. (a) Use a table with the two rows and the percentages (or proportions)

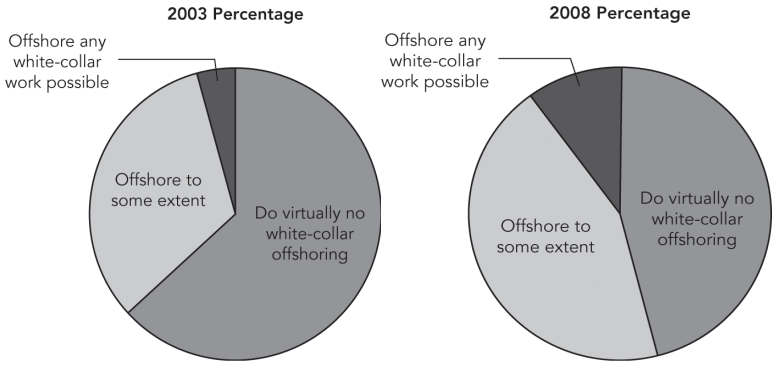
Unexpected illness	4,463	15.8%
Planned leave	23,735	84.2%

(b) A Pareto chart shows the categories in order of size.

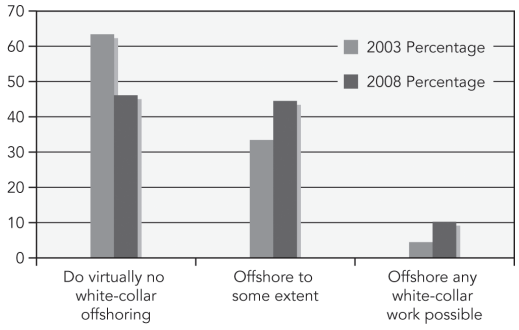


54. (a) Kraft plus Cadbury (15.2%) becomes the mode.
(b) Yes, if the Hershey items are much less expensive than those of other brands.

55. (a) Yes, pie charts are fine because the responses are mutually exclusive and sum to 100%.



(b) Various answers are possible. The following layout is reasonable.

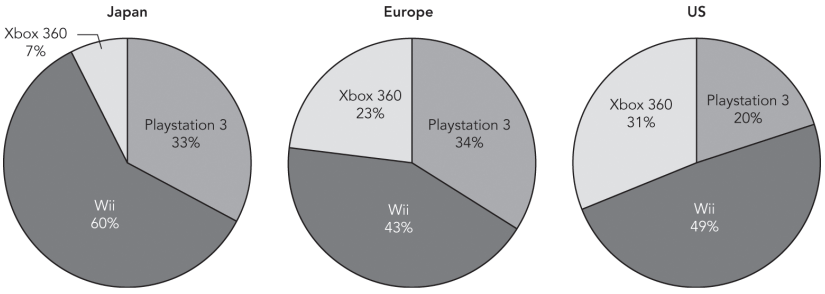


(c) The bar chart facilitates comparison. The pie chart makes the relative shares more apparent. For example, the 2003 pie shows a predominant share for taking no action, the only choice anticipated to fall in 2008.

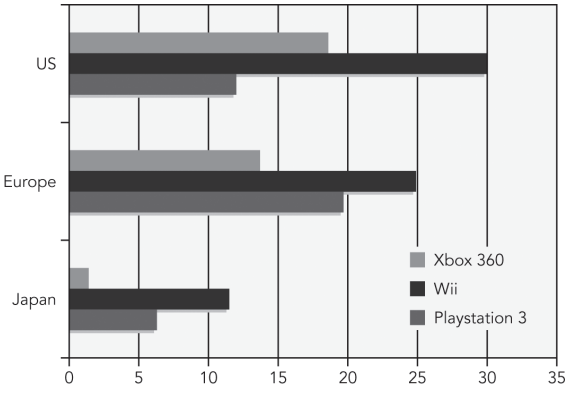
(d) The mode and median agree (virtually none) in 2003, but differ in 2008 as responses shift from the more consistent response to a tendency to do more off shoring.

56. (a) The radius for each region should be proportional to the total sales in that region, as suggested in this table:

	Japan	Europe	US
Playstation 3	6.3	19.7	12
Wii	11.5	24.9	30
Xbox 360	1.4	13.7	18.6
	19.2	58.3	60.6
radius	1	1.742543639	1.7765838
diamenter	2	3.485087278	3.553167601



(b)



57. (a)

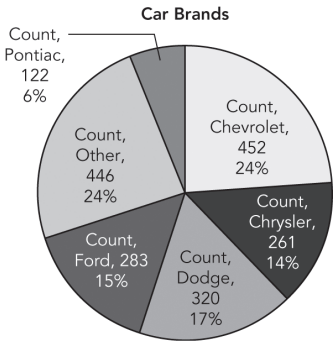
BUICK	Frequency 16	Relative Freq 0.00849
CHEVROLET	452	0.23992
CHRYSLER	261	0.13854
DODGE	320	0.16985
FORD	283	0.15021

(b) Chevrolet.

(c) The pie chart has too many slices.

(d)

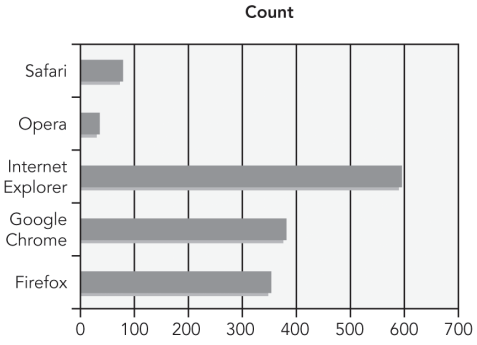
Brand	Count
Chevrolet	452
Chrysler	261
Dodge	320
Ford	283
Other	446
Pontiac	122

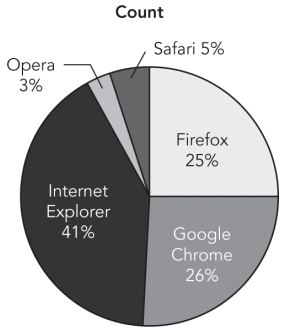


58. (a) Frequency table

Browser	Count
Firefox	354
Google Chrome	382
Internet Explorer	596
Opera	36
Safari	79

(b) The pie chart is easiest since it shows the percentages.





(c) Internet Explorer is losing share and Google Chrome has grown.

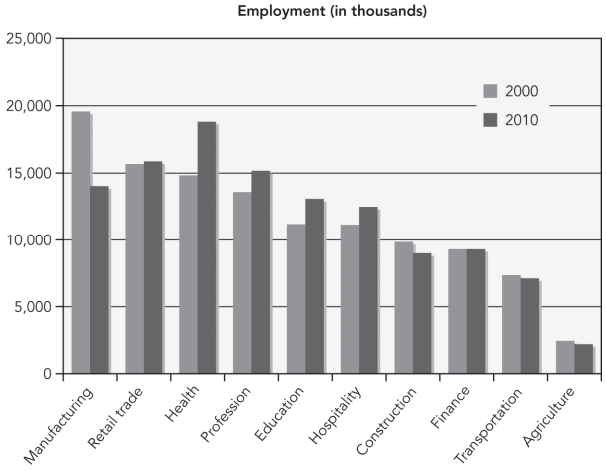
59. 4M Growth Industries

(a) It could use trends suggested by the table to indicate how to shift its sales force from declining industries to those that appear stronger and growing.

(b) A bar chart would show the counts and make it simpler to compare the counts within a year. A pie chart would emphasize the shares among industries in the two years.

(c) By looking at the changes from 2000 to 2010, you can see which industries are growing and which are shrinking. You could also use a side-by-side chart, but a chart of the differences does the subtraction for us.

(d) Grouped bar chart, order by size in 2010 to emphasize the change in manufacturing.



(e) Changes are shown in the differences in adjacent bar heights.

(f) One might choose to show just those that change, but also useful to see those that remained steady. Ten works fine.

(g) One can see the increase in health care and services and the big loss in manufacturing and in construction (mortgage debt crisis).

(h) Percentage shares are not evident, and the plot hides other types of employment.

60. This table shows the distribution of the hosts.

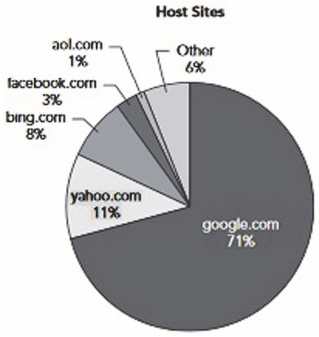
Host	Count	Proportion
google.com	3561	0.7122
yahoo.com	535	0.107
bing.com	384	0.0768
facebook.com	156	0.0312
aol.com	78	0.0156
imdb.com	43	0.0086
ask.com	41	0.0082
comcast.net	38	0.0076
humblebundle.com	33	0.0066
buyholidaygift.com	27	0.0054
reddit.com	23	0.0046
gevella.com	22	0.0044
yidio.com	20	0.004
youtube.com	20	0.004
redbubble.com	19	0.0038

Answers vary for this question. A sample answer is:

Motivation Amazon would benefit by forming a relationship with newly popular hosts that send shoppers to its Web sites.

Method Use bar or pie charts to compare the distribution of hosts in 2014 to that seen in the text example.

Mechanics A pie chart such as the following works well here to show the domination of Google. It is useful to combine the smaller hosts into an "Other" category. (Compare to Figure 3.6 in the text.)



Message Google dominates the search market and, not surprisingly, is the major host that sends shoppers to Amazon. Google was the third most popular host in 2004, but in 2014 it dominates, sending more than 70% of the visitors who use a host. An important caveat notes that these are shoppers who use a host, omitting those who come to Amazon directly.