

**SOLUTIONS MANUAL**

Intro **Stats**



De Veaux Velleman Bock

Includes  by Paul Velleman

### Chapter 3 - Displaying Categorical Data

- 1. **Graphs in the news.** Answers will vary.
- 2. **Graphs in the news II.** Answers will vary.
- 3. **Tables in the news.** Answers will vary.
- 4. **Tables in the news II.** Answers will vary.

5. **Magnet Schools.**

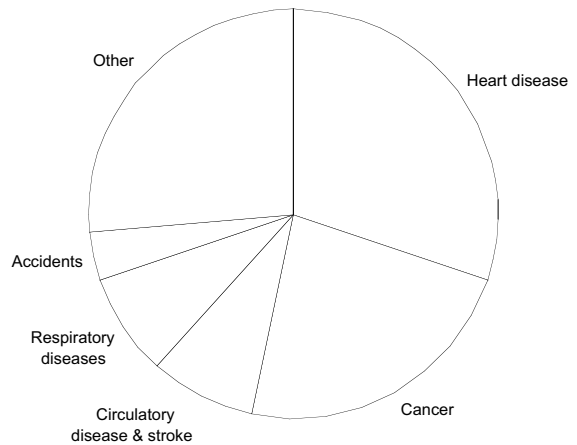
There were 1,755 qualified applicants for the Houston Independent School District’s magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away for lack of space.

6. **Magnet schools, again.**

There were 1,755 qualified applicants for the Houston Independent School District’s magnet schools program. 29.5% were Black or Hispanic, 16.6% were Asian, and 53.9% were white.

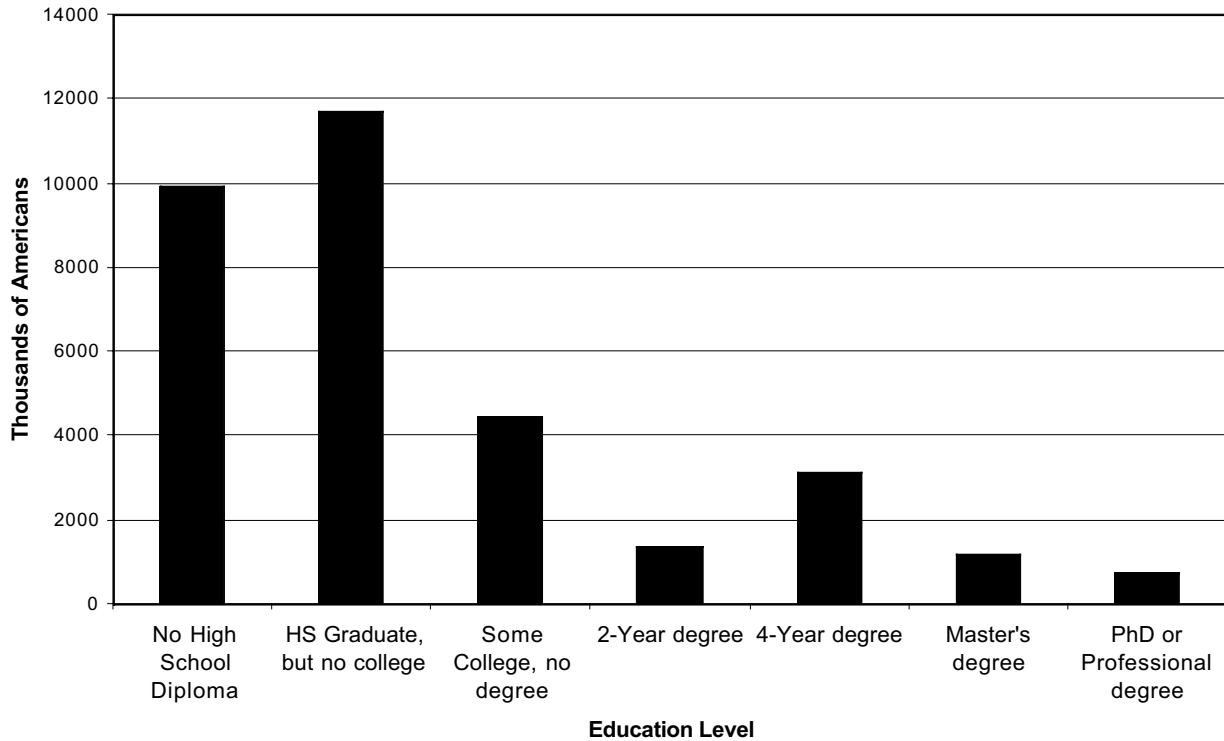
7. **Causes of death.**

- a) Yes, it is reasonable to assume that heart and respiratory disease caused approximately 38% of U.S. deaths in 1999, since there is no possibility for overlap. Each person could only have one cause of death.
- b) Since the percentages listed add up to 73.7%, other causes must account for 26.3% of US deaths.
- c) A pie chart is a good choice (with the inclusion of the “Other” category), since causes of US deaths represent parts of a whole. A bar chart would also be a good display.



8. Education.

Education Level of Americans over 65



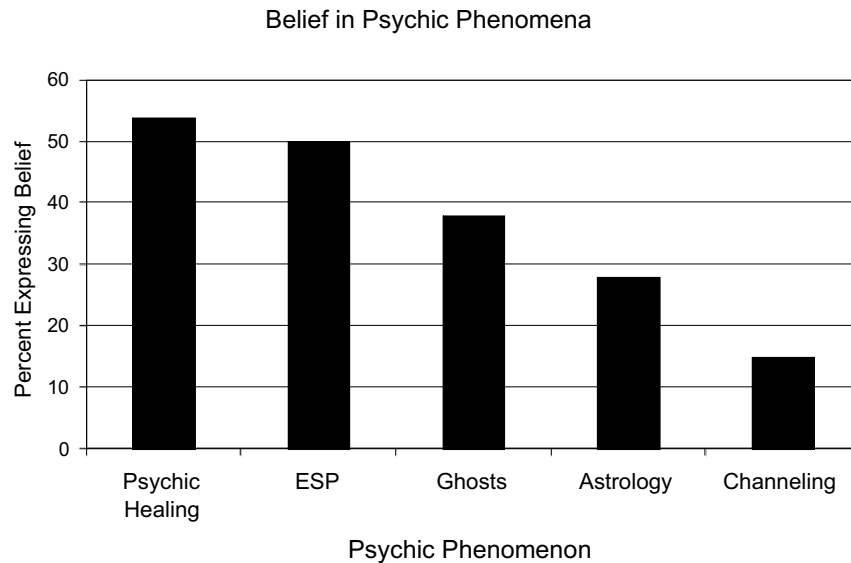
Information obtained by the Census Bureau in December 2000 reports the level of educational attainment of Americans over 65. Most older Americans have, at most, a high school diploma. Only about 15% of older Americans obtained a 4-year or higher degree.

9. Ghosts.

- a) It is NOT reasonable to assume the 66% of those polled expressed a belief in either ghosts or astrology. The percentages in the table add up to 185%! This tells us that we are not dealing with parts of a whole and that some respondents believe in more than one of the psychic phenomena listed. In other words, belief in ghosts and belief in astrology are not mutually exclusive. There is no way to know what percent of respondents believe in ghosts or astrology.
- b) As in 7a, since the percentages are not meant to add up to 100%, there is no way to determine what percentage of respondents did not believe in any of these phenomena. With data collected in this fashion, the only way to determine the percentage of people who did not believe in any of the psychic phenomena would be to add another category to the survey when collecting the data.

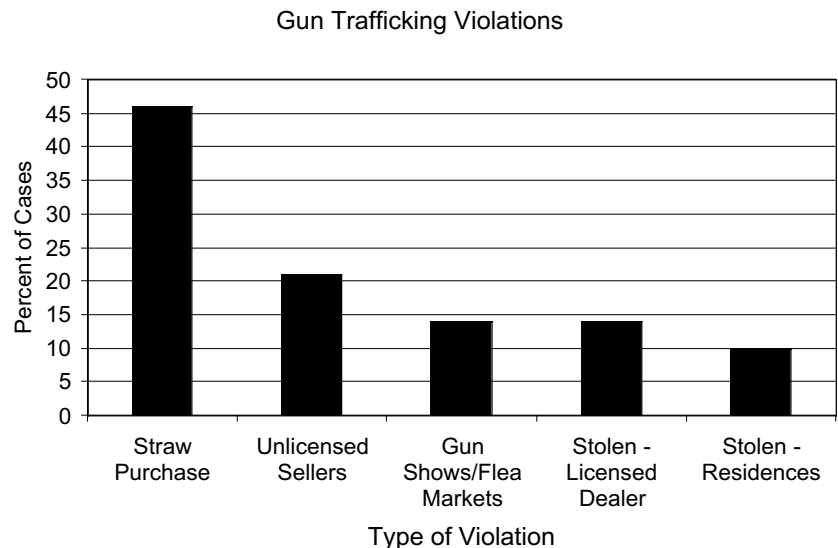
## 8 Part I Exploring and Understanding Data

- c) Since the percentages were not intended to add up to 100%, a pie chart is not appropriate. A bar chart nicely displays the percentages as relative heights of bars. Also, we are probably interested in the most popular phenomena, it makes sense to make a Pareto Chart, with the most common belief occurring first in the chart.



## 10. Illegal guns.

- a) *Who* – 1530 investigations into illegal gun trafficking. *What* – Percentage of cases that were the result of each of the five gun trafficking violations (straw purchase, unlicensed sellers, gun shows and flea markets, stolen from federally licensed dealers, stolen from residences). *When* – July 1996 – December 1998. *Where* – United States. *Why* – These data are being used to track the type of cases involved when gun trafficking violations occur. *How* – The BATF keeps track of this information on each case. The 1530 investigations tracked here probably represent a sample of all gun trafficking cases.
- b) The percentages listed total more than 100%. Either one of the numbers is incorrect, or some cases involved more than one type of gun trafficking violation.
- c) The bar chart at the right displays the information:



- d) A survey of 1530 gun trafficking investigations conducted by the Bureau of Alcohol, Tobacco, and Firearms between July 1996 and December 1998 indicates that the greatest percentage of investigations involve straw purchases, where a legal gun buyer acts on behalf of an illegal gun buyer. Straw purchases are involved in 46% of the investigations. Unlicensed sellers are involved in 21% of investigations. Gun shows and flea markets, thefts from licensed dealers, and thefts from residences are each involved in 10-15% of investigations. It should be noted that these percentages add up to more than 100%. This indicates that either some investigations involve more than one type of violation, or that one or more of the percentages may be incorrect.
- e) Corrupt licensed dealers were involved in only 9% of investigations, but were linked to almost half of the illegal firearms trafficked. One explanation for this is that each of these investigations involved large numbers of illegal firearms. Likewise, each straw purchase may have only involved a relatively small number of firearms. Analyzing these data by percentage of investigations only was obscuring crucial information. In light of this new information, corrupt licensed dealers appear to be a bigger problem.

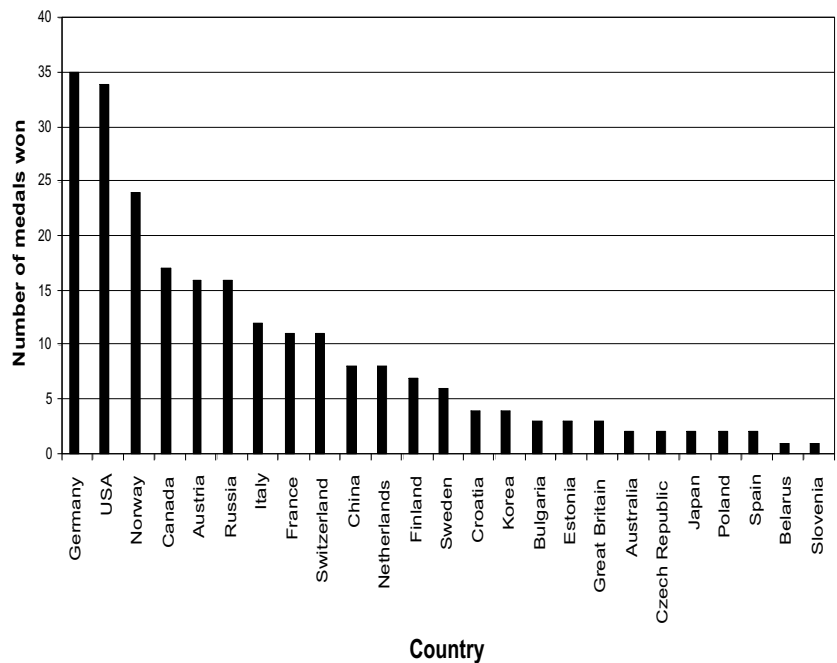
11. Oil spills.

The bar chart shows that grounding is the most frequent cause of oil spillage for these 50 spills, and allows the reader to rank the other types as well. If being able to differentiate between these close counts is required, use the bar chart. The pie chart is also acceptable as a display, but it's difficult to tell whether, for example, there is a greater percentage of spills caused by grounding or hull failure. If you want to showcase the causes of oil spills as a fraction of all 50 spills, use the pie chart.

12. Winter Olympics 2002.

- a) Here are two displays of the data:

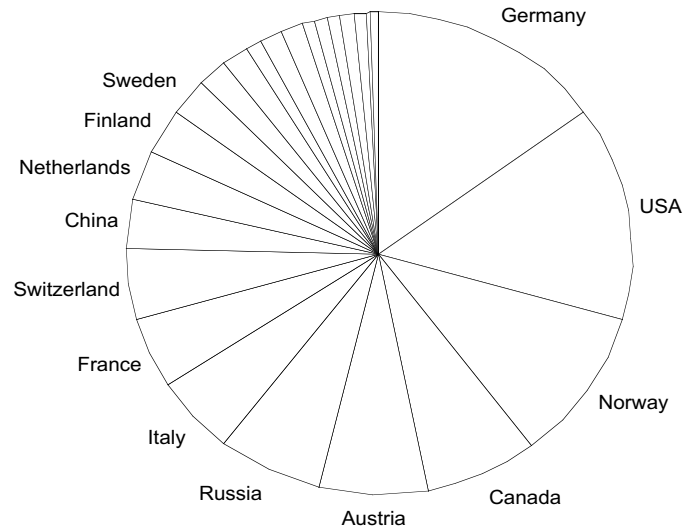
2002 Winter Olympics - Medals Won



This bar chart is confusing. There are simply too many categories!

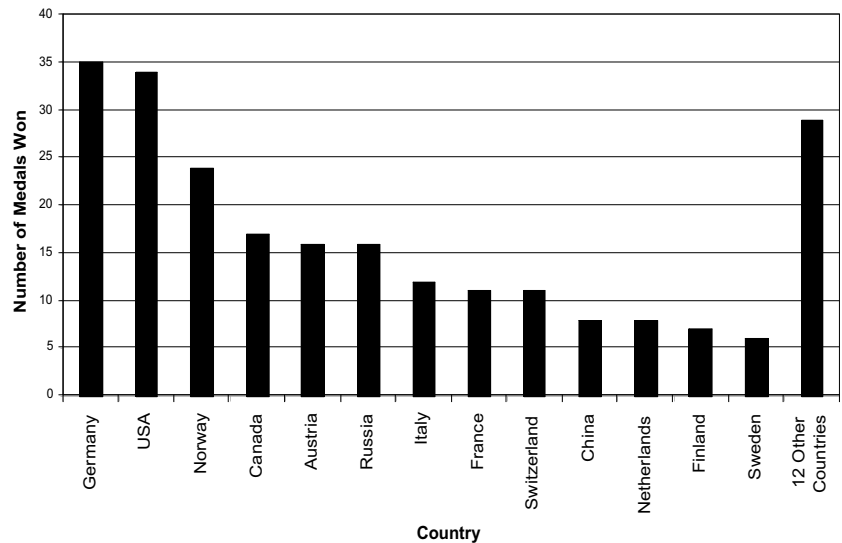
2002 Winter Olympics - Percentage of Medals Won

A pie chart of the percentage of medals won by each country is even more confusing! The sections of the chart representing countries that won fewer than 5 medals are too small to even label properly.



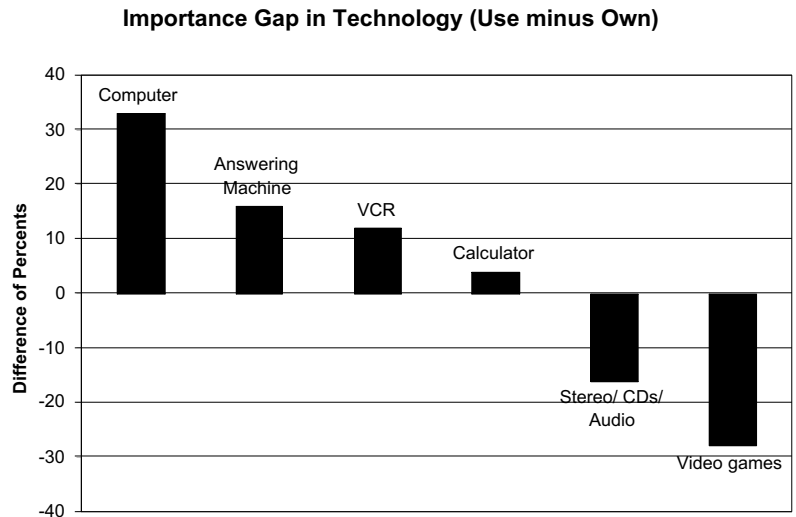
b) Perhaps we are primarily interested in countries that won many medals. Let's combine all countries that won fewer than 6 medals into a single category. This will make our chart easier to read. We are probably interested in number of medals won, rather than percentage of total medals won, so we'll stick with the bar chart. A bar chart is also better for comparisons.

2002 Winter Olympics - Medals Won



13. Teens and technology.

- a) 67% of teens use a calculator and 46% of teens use an answering machine, for a difference of 21%
- b) The display makes it appear that more than twice as many teens use calculators as use answering machines.
- c) The vertical scale should display percentages starting at 0. This isn't simply an improvement, but rather a necessity. The display provided is misleading.
- d) The display at the right is the Importance Gap in teens' opinions about technology. It displays the differences in percentages of teens who stated that the piece of technology was critical to use and the percentage of teens who actually own the piece of technology.



- e) Teens tend to think that some pieces of technology are critically important to own although they don't use them daily. The Importance Gap display highlights the fact that computers and video games fall on the two extremes of this difference. 77% of teens feel computers are important to own, yet fewer than half use them daily.

14. Teens and technology II.

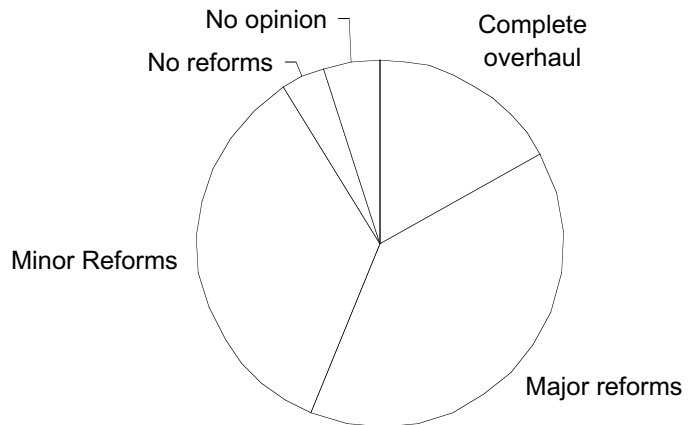
The main error is that this is simply the wrong type of display. The percentages of students who use each piece of technology aren't meant to represent parts of a whole, but rather overlapping categories. A bar chart would be the appropriate display. Even if a pie chart were acceptable, this one has some other problems. The three-dimensional display distorts the percentages. When using technology, always choose the two-dimensional displays. The percentages don't correspond to the angles of the pie pieces. For example, 85% looks smaller than 67% on the display. To top it all off, the display is unlabeled, giving us no context at all. Appropriate graphs are always self-explanatory.

## 12 Part I Exploring and Understanding Data

### 15. Auditing reform.

- a) The pie chart at the right displays the same information as the bar chart.
- b) The bar chart makes it easier to compare the percentages. It also preserves the order in which the questions were originally asked.
- c) Most respondents in a Gallup poll conducted February 8-10, 2002 favored some sort of reform in the way that corporations are audited. 39% of the respondents advocated major reforms, 35% said minor reforms were needed, and 17% favored a complete overhaul of the corporate auditing system. Only 4% were in favor of no reforms to the current system.

### What Kind of Changes Are Needed in the Way Corporations Are Audited?



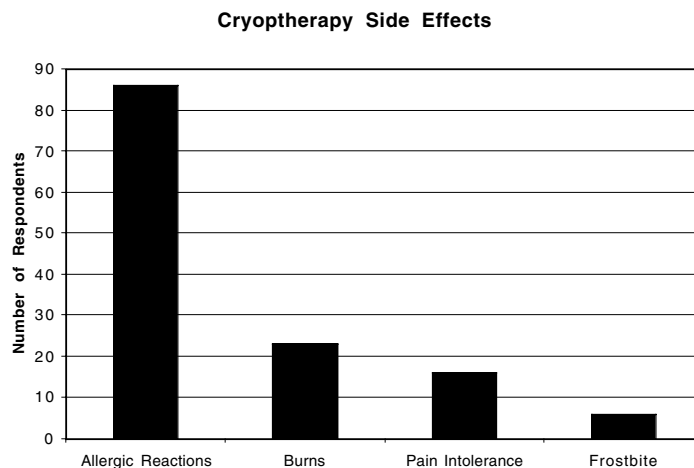
### 16. Modalities.

- a) The bars have false depth, which can be misleading. This is a bar chart, so the bars should have space between them. Running the labels on the bars from top to bottom and the vertical axis labels from bottom to top is confusing.
- b) The percentages sum to 100%. Normally, we would take this as a sign that all of the observations had been correctly accounted for. But in this case, it is extremely unlikely. Each of the respondents was asked to list *three* modalities. For example, it would be possible for 80% of respondents to say they use ice to treat an injury, and 75% to use electric stimulation. The fact that the percentages total greater than 100% is not odd. In fact, in this case, it seems wrong that the percentages add up to 100%, rather than correct.



## 17. Complications.

- A bar chart is the proper display for these data. A pie chart is not appropriate, since these are counts, not fractions of a whole.
- The *Who* for these data is athletic trainers, which should be a cause for concern. A trainer who treated many patients with cryotherapy would be more likely to have seen complications than one who used cryotherapy rarely. We would prefer a study in which the *Who* referred to patients, so we could assess the risks of each complication.



## 18. Politics.

- There are 192 students taking Intro Stats. Of those, 115, or about 59.9%, are male.
- There are 192 students taking Intro Stats. Of those, 27, or about 14.1%, consider themselves to be “Conservative”.
- There are 115 males taking Intro Stats. Of those, 21, or about 18.3%, consider themselves to be “Conservative”.
- There are 192 students taking Intro Stats. Of those, 21, or about 10.9%, are males who consider themselves to be “Conservative”.
- There are 77 females taking Intro Stats. Of those, 35, or about 45.5%, consider themselves to be “Liberal”.
- There are 115 males taking Intro Stats. Of those, 50, or about 43.5%, consider themselves to be “Liberal”.
- The conditional distribution of *politics* for males is 43.5% Liberal, 38.3% Moderate, and 18.3% Conservative. For females, the conditional distribution of *politics* is 45.5% Liberal, 46.8% Moderate, and 7.8% Conservative. The marginal distribution of *politics* is 44.2% Liberal, 41.7% Moderate, and 14.1% Conservative. These distributions are sufficiently different to suggest that *politics* and *sex* are NOT independent. The females in the class tend to identify themselves less as Conservative and more as Moderate than their male classmates.

## 19. Magnet schools revisited.

- There were 1755 qualified applicants to the Houston Independent School District’s magnet schools program. Of those, 292, or about 16.6% were Asian.
- There were 931 students accepted to the magnet schools program. Of those, 110, or about 11.8% were Asian.
- There were 292 Asian applicants. Of those, 110, or about 37.7%, were accepted.

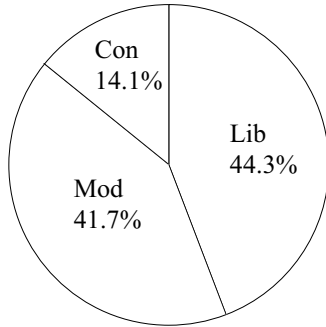
14 **Part I Exploring and Understanding Data**

d) There were 1755 total applicants. Of those, 931, or about 53%, were accepted.

20. **More politics.**

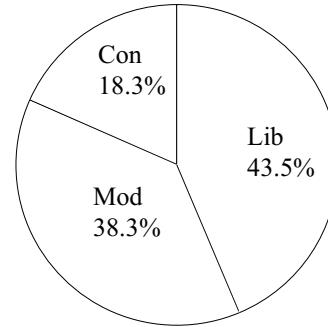
a)

Marginal Distribution of Politics



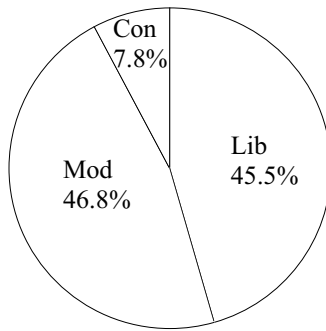
b)

Conditional Distribution of Politics Among Males



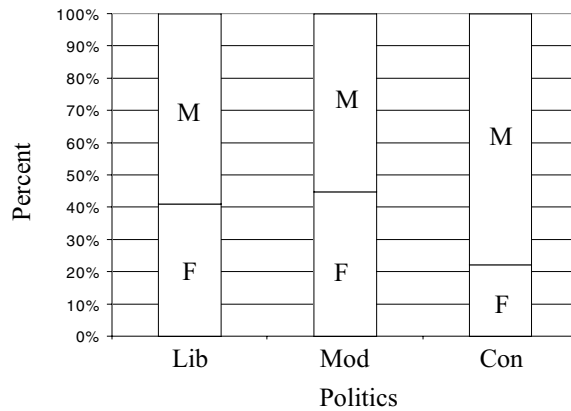
c)

Conditional Distribution of Politics Among Females



d)

Distribution of Sex Across Political Categories



e) The percentage of males and females varies across political categories. The percentage of self-identified Liberals and Moderates who are female is about twice the percentage of self-identified Conservatives who are female. This would suggest that *sex* and *politics* are not independent.

21. **Back to school.**

There were 1,755 qualified applicants for admission to the magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away. While the overall acceptance rate was 53%, 93.8% of Blacks and Hispanics were accepted, compared to only 37.7% of Asians, and 35.5% of whites. Overall, 29.5% of applicants were Black or Hispanics, but only 6% of those turned away were Black or Hispanic. Asians accounted for 16.6% of applicants, but 24.9% of those turned away. It appears that the admissions decisions were not independent of the applicant's ethnicity.

22. Cars.

- a) In order to get percentages, first we need totals. Here is the same table, with row and column totals. Foreign cars are defined as non-American. There are  $45+102=147$  non-American cars or  $147/359 \approx 40.95\%$ .

Origin	Driver		Total
	Student	Staff	
American	107	105	212
European	33	12	45
Asian	55	47	102
<b>Total</b>	<b>195</b>	<b>164</b>	<b>359</b>

- b) There are 212 American cars of which 107 or  $107/212 \approx 50.47\%$  were owned by students.  
 c) There are 195 students of whom 107 or  $107/195 \approx 54.87\%$  owned American cars.

- d) The marginal distribution of Origin is displayed in the third column of the table at the right: 59% American, 13% European, and 28% Asian.

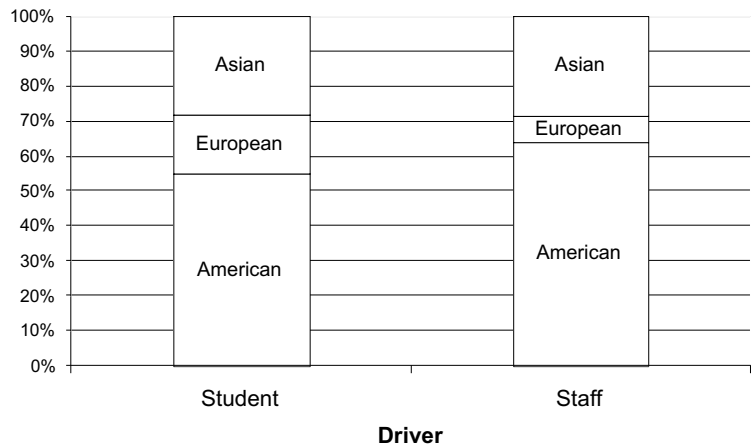
Origin	Totals
American	212 (59%)
European	45 (13%)
Asian	102 (28%)
<b>Total</b>	<b>359</b>

- e) The conditional distribution of Origin for Students is: 55% (107 of 195) American, 17% (33 of 195) European, and 28% (55 of 195) Asian. The conditional distribution of Origin for Staff is: 64% (105 of 164) American, 7% (12 of 164) European, and 29% (47 of 164) Asian.

- f) The percentages in the conditional distributions of Origin by Driver (students and staff) seem slightly different. Let's look at a segmented bar chart of Origin by Driver, to compare the conditional distributions graphically.

The conditional distributions of Origin by Driver have similarities and differences. Although students appear to own a higher percentage of European cars and a smaller percentage of American cars than the staff, the two groups own nearly the same percentage of Asian cars. However, because of the differences, there is evidence of an association between Driver and Origin.

Conditional Distribution of Origin by Driver



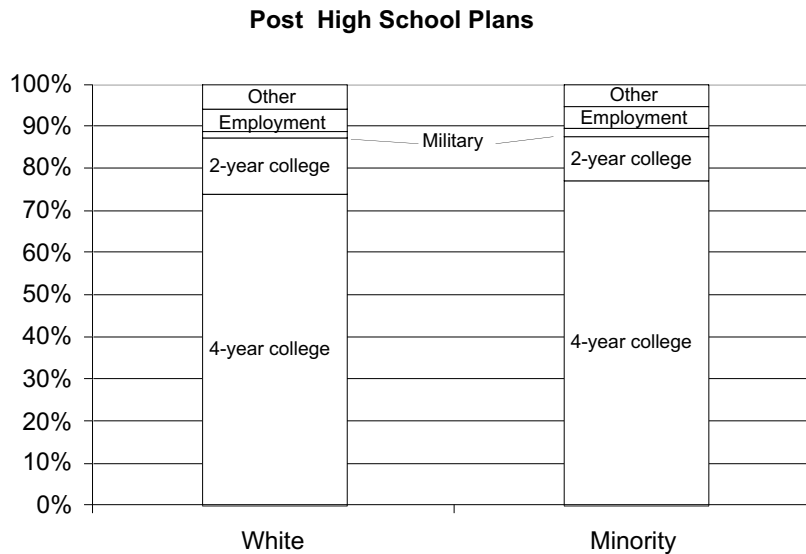
23. Class of 2000.

- a) A table with marginal totals is to the right. There are 268 White graduates and 325 total graduates.  $268/325 \approx 82.5\%$  of the graduates are White.  
 b) There are 42 graduates planning to attend 2-year colleges.  $42/325 \approx 12.9\%$

Plans	White	Minority	TOTAL
4-year college	198	44	242
2-year college	36	6	42
Military	4	1	5
Employment	14	3	17
Other	16	3	19
<b>TOTAL</b>	<b>268</b>	<b>57</b>	<b>325</b>

**16 Part I Exploring and Understanding Data**

- c) 36 white graduates are planning to attend 2-year colleges.  $36/325 \approx 11.1\%$
- d) 36 white graduates are planning to attend 2-year colleges and there are 268 whites graduates.  $36/268 \approx 13.4\%$
- e) There are 42 graduates planning to attend 2-year colleges.  $36/42 \approx 85.7\%$
- f) A segmented bar chart is a good display of these data:



- g) The conditional distributions of plans for Whites and Minorities are similar:  
 White - 74% 4-year college, 13% 2-year college, 2% military, 5% employment, 6% other.  
 Minority - 77% 4-year college, 11% 2-year college, 2% military, 5% employment, 5% other.  
 Caution should be used with the percentages for Minority graduates, because the total is so small. Each graduate is almost 2%. Still, the conditional distributions of plans are essentially the same for the two groups. There is little evidence of an association between race and plans for after graduation.

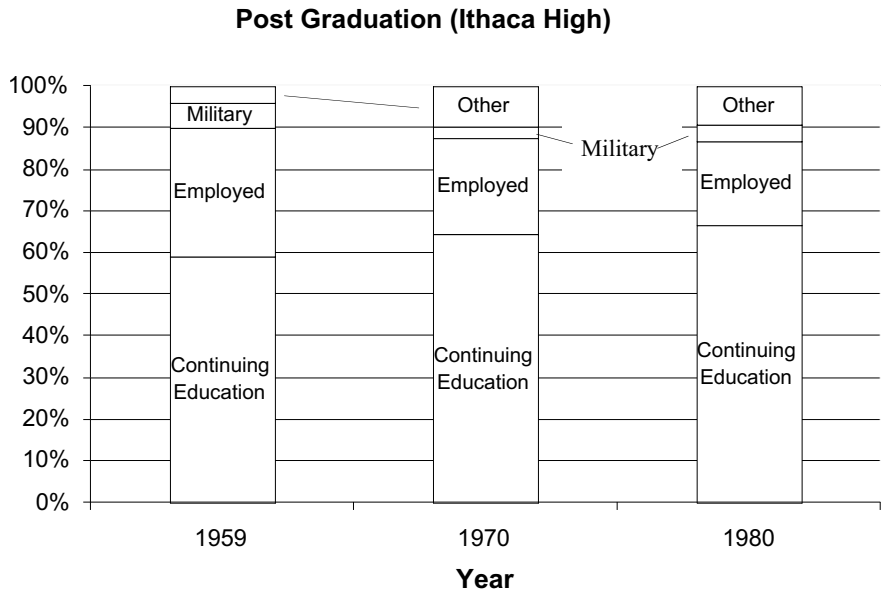
**24. After high school.**

- a) 56 graduates joined the military, out of 1415 total graduates.  $56/1415 \approx 4.0\%$

What graduates did	1959	1970	1980	TOTAL
<b>Continuing education</b>	197	388	320	905
<b>Employed</b>	103	137	98	338
<b>In the military</b>	20	18	18	56
<b>Other</b>	13	58	45	116
<b>Total</b>	333	601	481	1415

- b) In 1970, 601 students graduated.  $601/1415 \approx 42.5\%$
- c) In 1970, 18 graduates joined the military.  $18/601 \approx 3.0\%$
- d) 18 of the 56 students who joined the military were 1970 graduates.  $18/56 \approx 32.1\%$
- e) Convert the total column to percentages: 64.0% continued education, 23.9% employed, 4.0% in the military, and 8.2% other.
- f) Convert the 1959 column to percentages: 59.2% continued education, 30.9% employed, 6.0% in the military, and 3.9% other.

g) There is evidence that the percentage of Ithaca High School graduates continuing their education has increased from 1959 to 1970 to 1980. Likewise, the percentage of students who enter the work force directly after high school appears to have decreased over these years. For example, the percentage of graduates who entered the workforce decreased from 30.9% to 22.8% to 20.4% for the years 1959, 1970, and 1980, respectively.



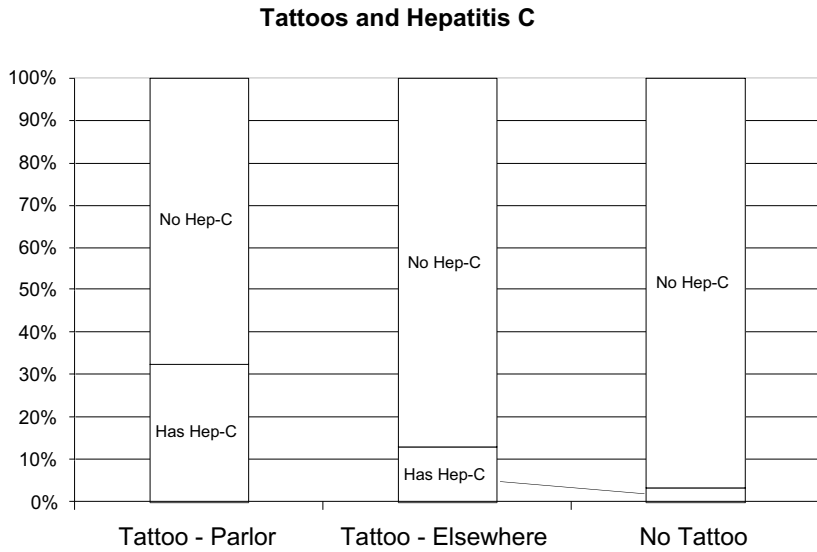
**25. Canadian languages.**

- a) 19,134,000 Canadians speak English only.  $19,134,000 / 28,529,000$  total Canadians  $\approx 67.1\%$
- b) 4,078,000 Canadians speak French only and 4,843,000 speak both French and English, for a total of 8,921,000 French speakers.  $8,921,000 / 28,529,000$  total Canadians  $\approx 31.3\%$
- c) 3,952,000 French and 2,661,000 French and English speakers yield a total of 6,613,000 French speakers in Quebec.  $6,613,000 / 7,045,000$  Quebec residents  $\approx 93.9\%$
- d) 6,613,000 Quebec residents speak French and 8,921,000 Canadians speak French. The percentage of French-speaking Canadians who live in Quebec is  $6,613,000 / 8,921,000 \approx 74.1\%$
- e) If language knowledge were independent of Province, we would expect the percentage of French-speaking residents of Quebec to be the same as the overall percentage of Canadians who speak French. Since 31.3% of all Canadians speak French while 93.9% of residents of Quebec speak French, there is evidence of an association between language knowledge and Province.

18 *Part I Exploring and Understanding Data*

26. Tattoos.

The study by the University of Texas Southwestern Medical Center provides evidence of an association between having a tattoo and contracting hepatitis C. Around 33% of the subjects who were tattooed in a commercial parlor had hepatitis C, compared with 13% of those tattooed elsewhere, and only 3.5% of those with no tattoo. If having a tattoo and having hepatitis C were independent, we would have expected these percentages to be roughly the same.



27. Weather forecasts.

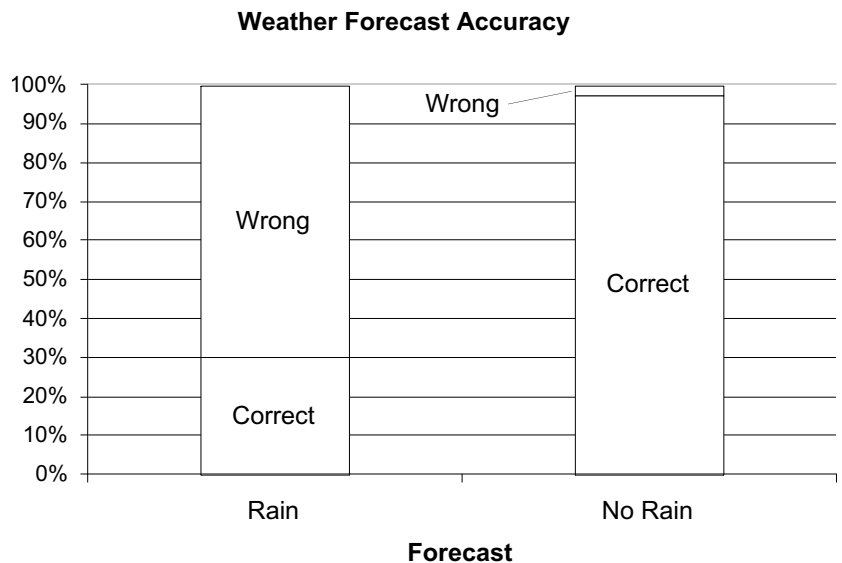
a) The table shows the marginal totals. It rained on 34 of 365 days, or 9.3% of the days.

b) Rain was predicted on 90 of 365 days.  $90/365 \approx 24.7\%$  of the days.

c) The forecast of rain was correct on 27 of the days it actually rained and the forecast of No Rain was correct on 268 of the days it didn't rain. So, the forecast was correct a total of 295 times.  $295/365 \approx 80.8\%$  of the days.

d) There is evidence of an association between the type of weather and the ability of the forecasters to make an accurate prediction. When Rain was forecast, it actually rained only 27 out of 90 times (30%). When the forecasters predicted No Rain, they were correct an amazing 268 out of 275 times (97.5%). The weather forecasters' predictions of Rain are much less reliable than their predictions of No Rain.

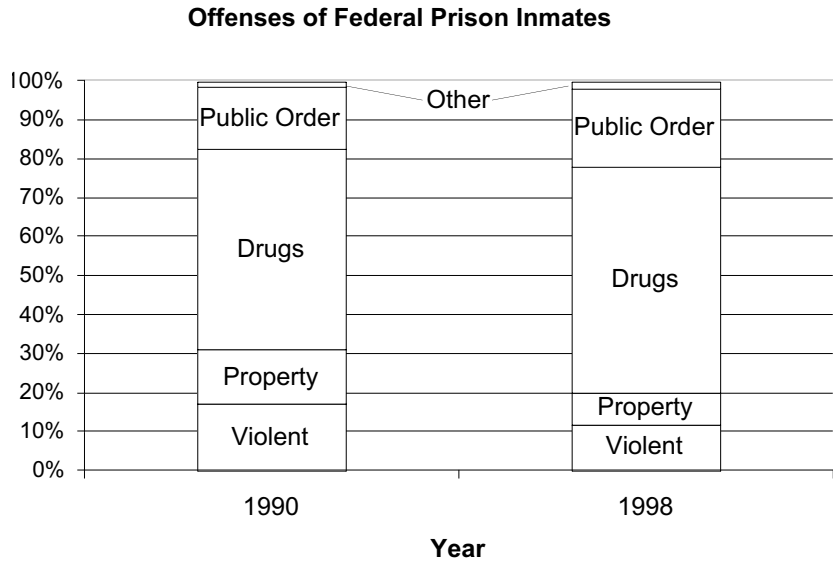
		Actual Weather		Total
		Rain	No Rain	
Forecast	Rain	27	63	90
	No Rain	7	268	275
	Total	34	331	365



However, this is hardly a ringing endorsement of the weather forecasters abilities to predict No Rain. Since the actual weather is No Rain on 90.7% of the days, these predictions are a pretty safe bet.

28. Federal prisons.

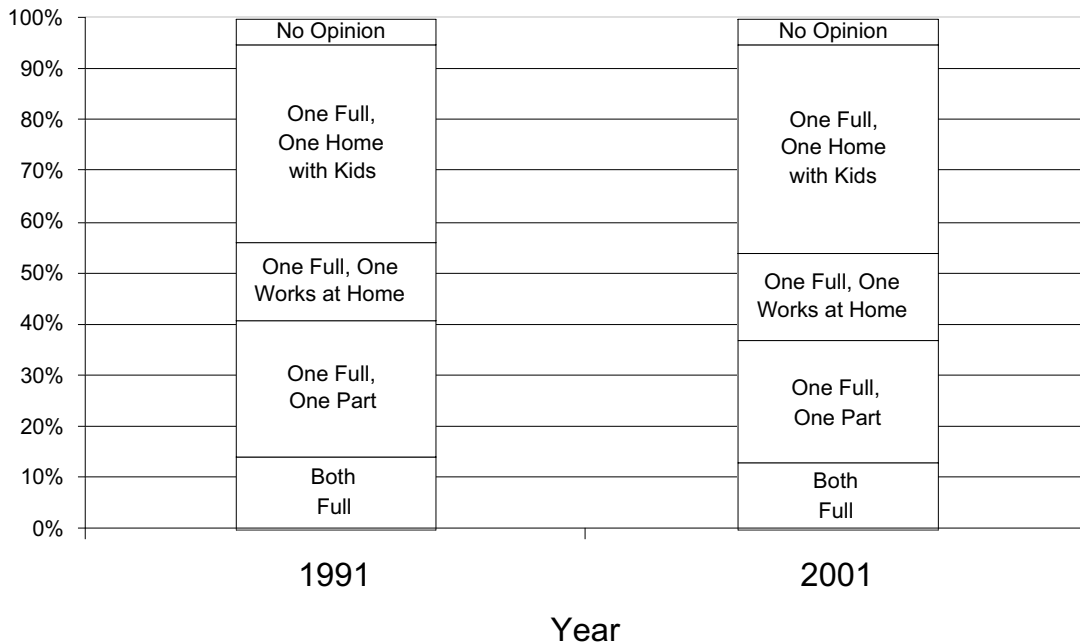
a) From 1990 to 1998, there was an increase in the percentage of the prison population that was incarcerated for drug and public order offenses. Also during this time period, there was an decrease in the percentage of the prison population incarcerated for property and violent offenses. For example, in 1990, about 52% of the prison population was incarcerated for drug offenses. By 1998, that percentage had increased to about 58%.



b) These data do *not* show that there has been an increase in drug *use*, merely an increase in the percentage of the prison population incarcerated for drug offenses. One more point of interest here is that the federal prison population in 1998 was almost double the 1990 population.

29. Working Parents.

Which do you see as the ideal family?



**20 Part I Exploring and Understanding Data**

The Gallup poll doesn't provide strong evidence of a change in people's opinions regarding the ideal family in today's society between the years of 1991 and 2001. The conditional distributions of opinion by year appear roughly the same. For example, the percentage of respondents in 1991 who thought the ideal family had two parents that worked fulltime was 14%, and in 2001, the percentage was 13%.

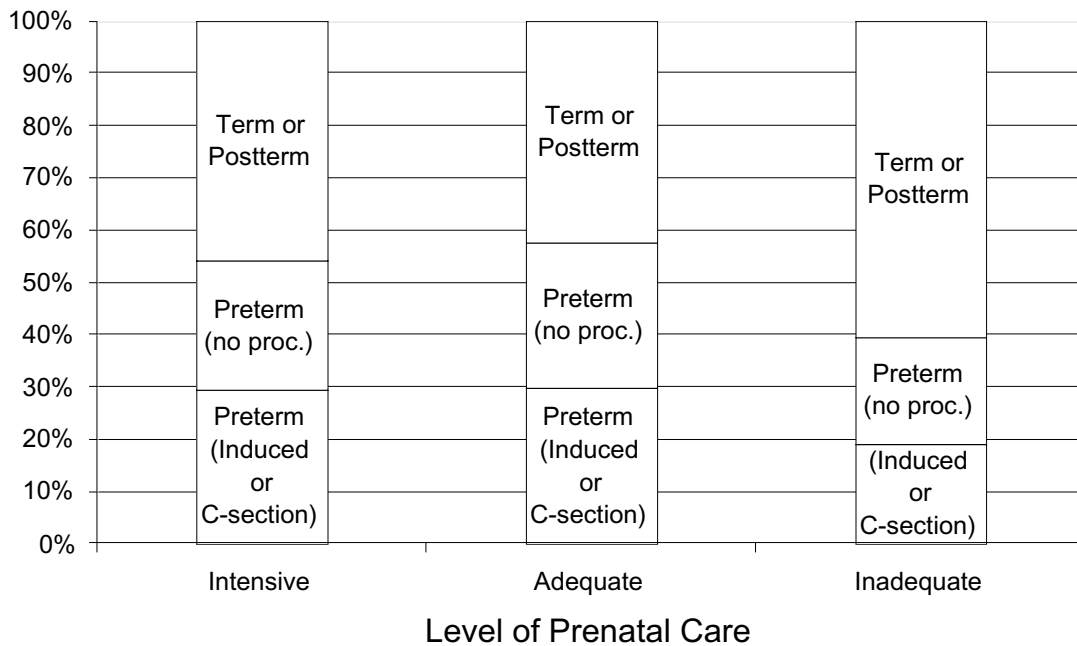
**30. Twins.**

- a) Of the 278,000 mothers who had twins in 1995-1997, 63,000 had inadequate health care during their pregnancies.  $63,000/278,000 = 22.7\%$

Twin Births 1995-97 (in thousands)				
Level of Prenatal Care	Preterm (Induced or Caesarean)	Preterm (without procedures)	Term or Postterm	Total
Intensive	18	15	28	61
Adequate	46	43	65	154
Inadequate	12	13	38	63
<b>Total</b>	<b>76</b>	<b>71</b>	<b>131</b>	<b>278</b>

- b) There were 76,000 induced or Caesarean births and 71,000 preterm births without these procedures.  $(76,000 + 71,000)/278,000 = 52.9\%$
- c) Among the mothers who did not receive adequate medical care, there were 12,000 induced or Caesarean births and 13,000 preterm births without these procedures. 63,000 mothers of twins did not receive adequate medical care.  $(12,000 + 13,000)/63,000 = 39.7\%$
- d)

Twin Birth Outcome 1995-1997





- e) 52.9% of all twin births were preterm, while only 39.7% of births in which inadequate medical care was received were preterm. This is evidence of an association between level of prenatal care and twin birth outcome. If these variables were independent, we would expect the percentages to be roughly the same. Generally, those mothers who received adequate medical care were more likely to have preterm births than mothers who received intensive medical care, who were in turn more likely to have preterm births than mothers who received inadequate health care. This does *not* imply that mothers should receive inadequate health care do decrease their chances of having a preterm birth, since it is likely that women that have some complication *during* their pregnancy (that might lead to a preterm birth), would seek intensive or adequate prenatal care.

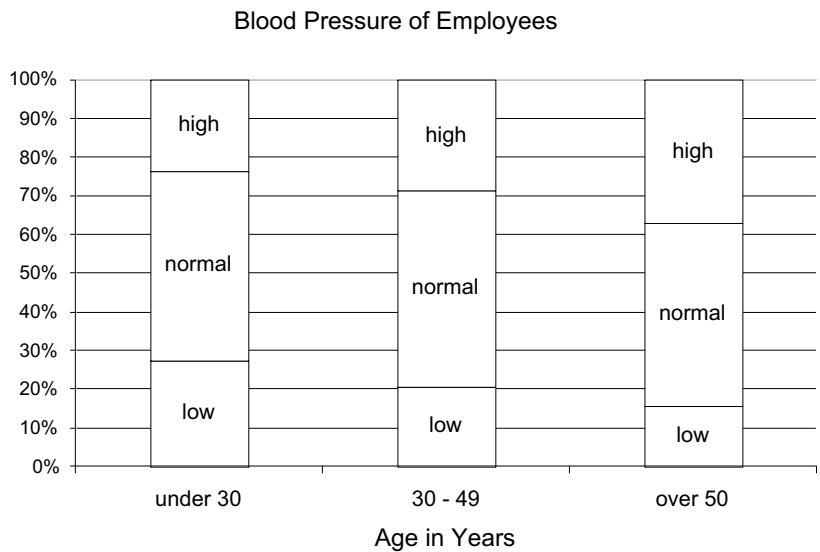
**31. Blood pressure.**

- a) The marginal distribution of blood pressure for the employees of the company is the total column of the table, converted to percentages. 20% low, 49% normal and 31% high blood pressure.

Blood pressure	under 30	30 - 49	over 50	Total
low	27	37	31	95
normal	48	91	93	232
high	23	51	73	147
<b>Total</b>	<b>98</b>	<b>179</b>	<b>197</b>	<b>474</b>

- b) The conditional distribution of blood pressure within each age category is:  
 Under 30 : 28% low, 49% normal, 23% high  
 30 - 49 : 21% low, 51% normal, 28% high  
 Over 50 : 16% low, 47% normal, 37% high

- c) A segmented bar chart of the conditional distributions of blood pressure by age category is at the right.



- d) In this company, as age increases, the percentage of employees with low blood pressure decreases, and the percentage of employees with high blood pressure increases.

- e) No, this does not prove that people’s blood pressure increases as they age.

Generally, an association between two variables does not imply a cause-and-effect relationship. Specifically, these data come from only one company and cannot be applied to all people. Furthermore, there may be some other variable that is linked to both age and blood pressure. Only a controlled experiment can isolate the relationship between age and blood pressure.

## 22 Part I Exploring and Understanding Data

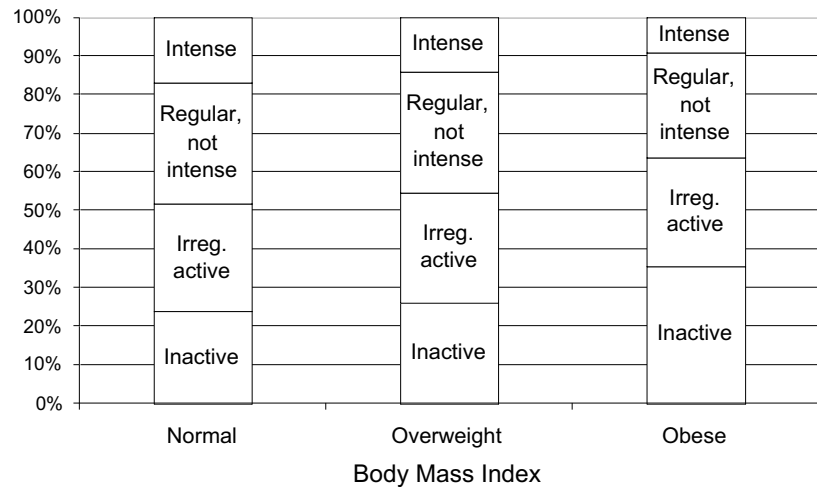
### 32. Obesity and exercise.

a) Participants were categorized as Normal, Overweight or Obese, according to their Body Mass Index. Within each classification of BMI (column), participants self reported exercise levels. Therefore, these are column percentages. The percentages sum to 100% in each column, *not* across each row.

b) A segmented bar chart of the conditional distributions of level of physical activity by Body Mass Index category is at the right.

c) No, even though the graphical displays provide strong evidence that lack of exercise and BMI are not independent. All three BMI categories have nearly the same percentage of subjects who report “Regular, not intense” or “Irregularly active”, but as we move

Body Mass Index and Level of Physical Activity



from Normal to Overweight to Obese we see a decrease in the percentage of subjects who report “Regular, intense” physical activity (16.8% to 14.2% to 9.1%), while the percentage of subjects who report themselves as “Inactive” increases. While it may seem logical that lack of exercise causes obesity, association between variables does not imply a cause-and-effect relationship. A lurking variable (for example, overall health) might influence both BMI and level of physical activity, or perhaps lack of exercise is *caused by* obesity. Only a controlled experiment could isolate the relationship between BMI and level of physical activity.

### 33. Family planning.

The percentage of unplanned pregnancies decreased as education level increased. However, this does *not* mean that more schooling taught young women better family planning. Association between variables is not evidence of a cause-and-effect relationship. In this case, other socioeconomic variables might be related to both unplanned pregnancy and education level. Perhaps some women even had to leave school *because* of an unplanned pregnancy.

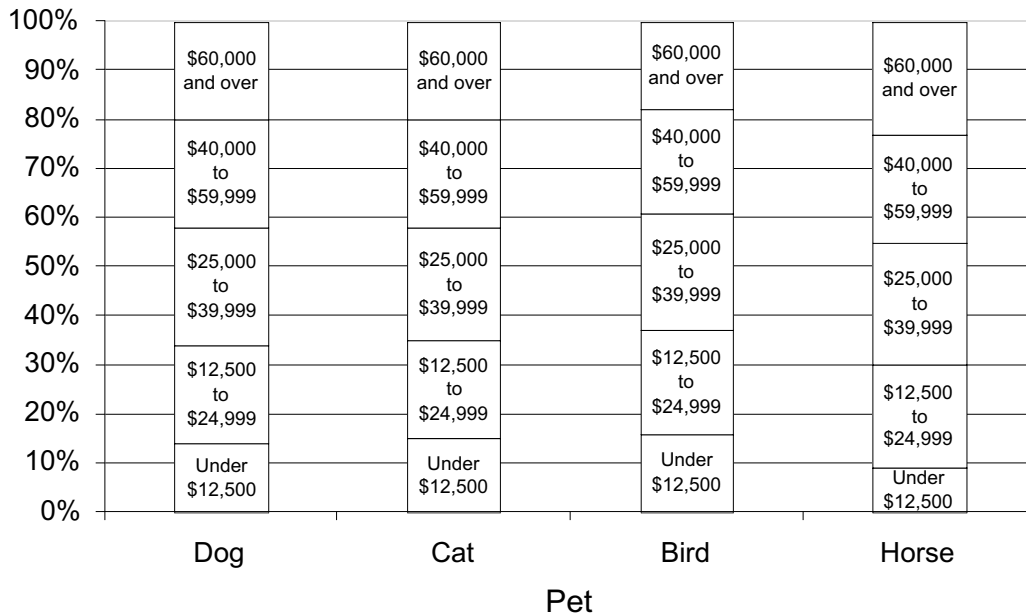
### 34. Pet ownership.

a) No, the income distributions of households by pet ownership wouldn't be expected to be the same. Caring for a horse is much more expensive, generally, than caring for a dog, cat, or bird. Households with horses as pets would be expected to be more common in the higher income categories.

b) These are column percentages, since each column totals 100%. Each pet was classified as belonging to a family in one of the income level categories.

c)

Income Distribution of Households Owning Pets



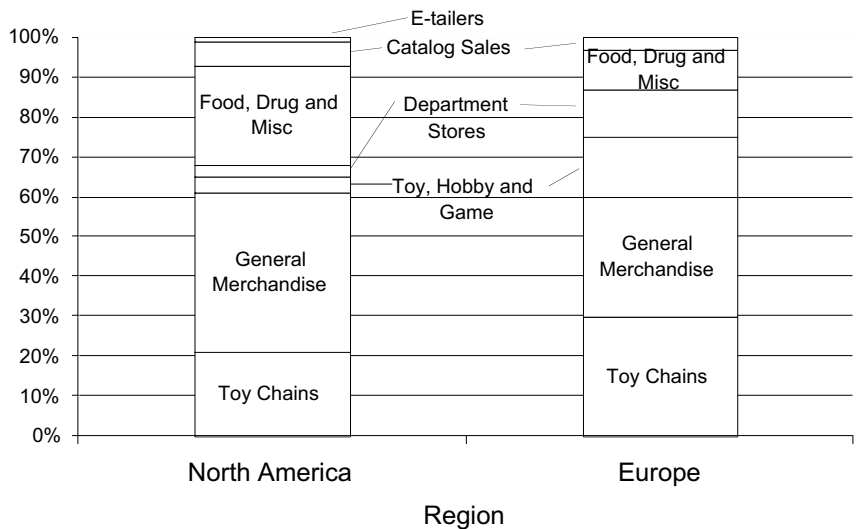
The data support the initial guess to a certain extent. The percentage of horses whose owners have income less than \$12,500 is only 9%, compared to percentages in the 20s for other income levels, while the income levels of owners of other pets were distributed in roughly the same percentages. However, with the exception of those earning less than \$12,500, the percentages in each income level among horse owners weren't much different.

35. Worldwide toy sales.

- a) 2042.16 million dollars worth of toys are sold in European department stores, and 5987.6 million dollars worth of toys are sold in all department stores.  $2042.16/5987.6 \approx 34.1\%$
- b) 2843.8 million dollars worth of toys are sold through catalogs. The "World" total of the table is 74395 million dollars worth of toys sold. The percentage of all toys sold through catalogs is  $2843.8/74035 \approx 3.8\%$ .

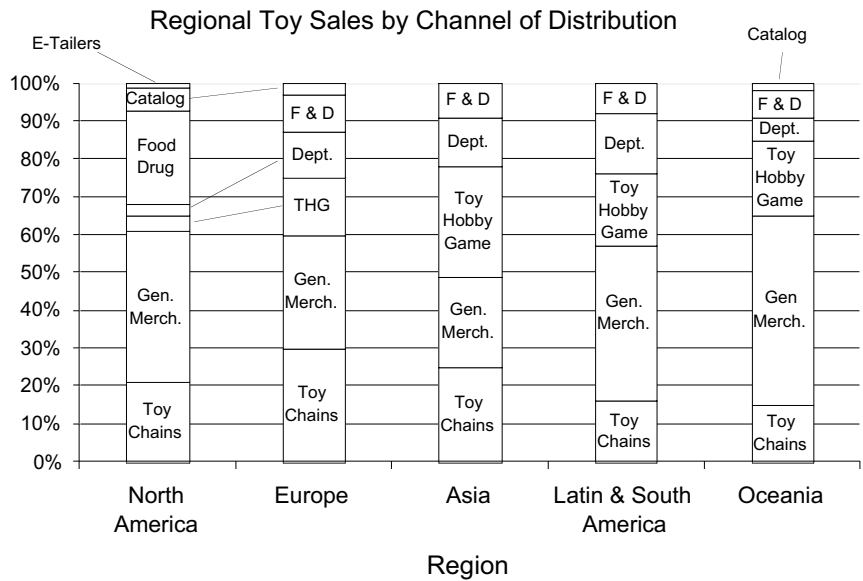
Regional Toy Sales by Channel of Distribution

- c) Toy Chains are more popular sellers of toys in Europe than in North America, as are Toy, Hobby and Game retailers and Department Stores. In North America, General Merchandise and Food, Drug and Misc. Outlets are more popular than they are in Europe.



24 Part I Exploring and Understanding Data

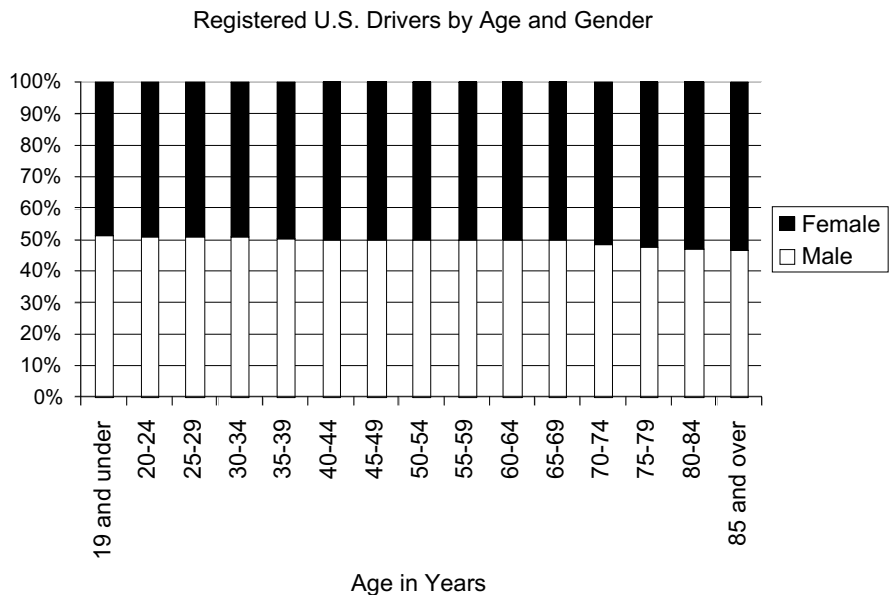
d) The distribution channel varies by region, although General Merchandise stores are either the most popular or a close second in all regions. Toy Chains are especially important in Europe and Toy, Hobby and Game Shops are important in Asia. Only North America has a significant percentage of toys sold in Food and Drug Stores. In the year 1999, E-tailing of toys was of little significance.



36. Driver's licenses.

- a) A table with marginal totals is provided at the right. There 9,743,519 drivers under 20 and a total of 190,625,024 drivers in the U.S. That's about 5% of U.S. drivers under 20.
- b) There are 95,796,069 males out of 190,625,024 total U.S. drivers, or about 50%.
- c) Each age category appears to have about 50% male and 50% female drivers.

The segmented bar chart shows a pattern in the deviations from 50%. At younger ages, males form the slight majority of drivers. This percentage shrinks until the percentages are 50% male and 50% for middle aged drivers. The percentage of male drivers continues to shrink until, at around age 65, female drivers hold a slight majority. This continues into the 85 and over category. It should be noted that this relationship is *very* slight, and may just be a coincidence.



- d) There appears to be a slight association between age and gender of U.S. drivers. Younger drivers are slightly more likely to be male, and older drivers are slightly more likely to be female.

## 37. Hospitals.

- a) The marginal totals have been added to the table:

		Discharge delayed		
		Large Hospital	Small Hospital	Total
Procedure	Major surgery	120 of 800	10 of 50	130 of 850
	Minor surgery	10 of 200	20 of 250	30 of 450
	Total	130 of 1000	30 of 300	160 of 1300

160 of 1300, or about 12.3% of the patients had a delayed discharge.

- b) Major surgery patients were delayed 130 of 850 times, or about 15.3% of the time. Minor Surgery patients were delayed 30 of 450 times, or about 6.7% of the time.
- c) Large Hospital had a delay rate of 130 of 1000, or 13%. Small Hospital had a delay rate of 30 of 300, or 10%. The small hospital has the lower overall rate of delayed discharge.
- d) Large Hospital: Major Surgery 15% delayed and Minor Surgery 5% delayed. Small Hospital: Major Surgery 20% delayed and Minor Surgery 8% delayed. Even though small hospital had the lower overall rate of delayed discharge, the large hospital had a lower rate of delayed discharge for each type of surgery.
- e) No. While the overall rate of delayed discharge is lower for the small hospital, the large hospital did better with *both* major surgery and minor surgery.
- f) The small hospital performs a higher percentage of minor surgeries than major surgeries. 250 of 300 surgeries at the small hospital were minor (83%). Only 200 of the large hospital's 1000 surgeries were minor (20%). Minor surgery had a lower delay rate than major surgery (6.7% to 15.3%), so the small hospital's overall rate was artificially inflated. Simply put, it is a mistake to look at the overall percentages. The real truth is found by looking at the rates after the information is broken down by type of surgery, since the delay rates for each type of surgery are so different. The larger hospital is the better hospital when comparing discharge delay rates.

## 38. Delivery service.

- a) Pack Rats has delivered a total of 28 late packages (12 Regular + 16 Overnight), out of a total of 500 deliveries (400 Regular + 100 Overnight).  $28/500 = 5.6\%$  of the packages are late. Boxes R Us has delivered a total of 30 late packages (2 Regular + 28 Overnight) out of a total of 500 deliveries (100 Regular + 400 Overnight).  $30/500 = 6\%$  of the packages are late.
- b) The company should have hired Boxes R Us instead of Pack Rats. Boxes R Us only delivers 2% (2 out of 100) of its Regular packages late, compared to Pack Rats, who deliver 3% (12 out of 400) of its Regular packages late. Additionally, Boxes R Us only delivers 7% (28 out of 400) of its Overnight packages late, compared to Pack Rats, who delivers 16% of its Overnight packages late. Boxes R Us is better at delivering Regular and Overnight packages.

**26 Part I Exploring and Understanding Data**

- c) This is an instance of Simpson’s Paradox, because the overall late delivery rates are unfair averages. Boxes R Us delivers a greater percentage of its packages Overnight, where it is comparatively harder to deliver on time. Pack Rats delivers many Regular packages, where it is easier to make an on-time delivery.

**39. Graduate admissions.**

- a) 1284 applicants were admitted out of a total of 3014 applicants.  
 $1284/3014 = 42.6\%$

Program	Males Accepted (of applicants)	Females Accepted (of applicants)	Total
1	511 of 825	89 of 108	600 of 933
2	352 of 560	17 of 25	369 of 585
3	137 of 407	132 of 375	269 of 782
4	22 of 373	24 of 341	46 of 714
<b>Total</b>	<b>1022 of 2165</b>	<b>262 of 849</b>	<b>1284 of 3014</b>

- b) 1022 of 2165 (47.2%) of males were admitted.  
 262 of 849 (30.9%) of females were admitted.

- c) Since there are four comparisons to make, the table at the right organizes the percentages of males and females accepted in each program. Females are accepted at a higher rate in every program.

Program	Males	Females
1	61.9%	82.4%
2	62.9%	68.0%
3	33.7%	35.2%
4	5.9%	7%

- d) The comparison of acceptance rate within each program is most valid. The overall percentage is an unfair average. It fails to take the different numbers of applicants and different acceptance rates of each program. Women tended to apply to the programs in which gaining acceptance was difficult for everyone. This is an example of Simpson’s Paradox.