

**Chapter 3 – Relationships Between Categorical Variables – Contingency Tables****Section 3.1****1. College value?**

	Poor	Only Fair	Good	Excellent	DK/NA	Total
US Adults	321	900	750	107	64	2142
Presidents	32	222	622	179	0	1055
Total	353	1122	1372	286	64	3197

- a) The percent of college presidents who think that higher education provides a poor value is  $32/1055 \approx 3\%$ .
- b)  $(750+107)/2142 = 857/2142 \approx 40\%$  of U.S. adults think that the value provided by the higher education system is either good or excellent.  $(622+179)/1055 = 801/1055 \approx 75.9\%$  of college presidents think that the value provided by the higher education system is either good or excellent.
- c) 15% of U.S. adults view college as a poor value, but only 3% of college presidents do. Similarly, U.S. adults are twice as likely to view college as an only fair value compared to the presidents (42% to 21%). Presidents are much more likely (76%) to rate colleges as a good or excellent value compared to U.S. adults (only 40%). So, in short, college presidents have a much higher opinion of the value of college than U.S. adults do.
- d) In this random sample of U.S. adults,  $107/2142 \approx 5.00\%$  responded that college provides an excellent value. This is only an estimate, however, of the percentage of all U.S. adults who feel the same way. The percentage of all U.S. adults is probably close to, but not exactly 5%.

**2. Cyber comparison shopping.**

	<\$30K	\$30K – \$49.9K	\$50K – \$74.9K	>\$75K	Total
Yes	207	115	134	204	660
No	625	406	260	417	1708
Total	832	521	394	621	2368

- a)  $207/832 \approx 24.9\%$  of those earning less than \$30K cyber compare shop.  $204/621 \approx 32.9\%$  of those earning more than \$75K cyber compare shop.
- b)  $207/660 \approx 31.4\%$  of those who cyber compare shop earn less than \$30K.
- c)  $660/2368 \approx 28\%$  of those in this random sample of people who owned cell phones use their phones to cyber compare shop. First of all, not all shoppers have cell phones, so we wouldn't expect the percentage of all shoppers to be similar to the percentage found in this sample. Even if we were only concerned with all shoppers with phones, we would only expect the percentage of all shoppers to be reasonably close to, but not exactly, 28%.

**Section 3.2****3. College value again.**

- a) The conditional distribution of college presidents' opinions about the value of a college education is 3% poor; 21% only fair; 59% good; 17% excellent
- b) Omitting the 64 DK/NAs, the conditional distribution of the opinions U.S. adults about the value of a college education is 15% negative; 43% middle; 41% positive.

**4. Cyber comparison shopping again.**

- a) The conditional distribution of income for those who do not cyber compare prices is 37% <\$30K; 24% \$30K–\$49.9K; 15% \$50K–\$74.9K; 24% >\$75K.
- b) The conditional distribution of income for those who do cyber compare prices is 31% <\$30K; 17% \$30K–\$49.9K; 20% \$50K–\$74.9K; 31% >\$75K.

## **22 Part I Exploring and Understanding Data**

### **Section 3.3**

#### **5. Diet and politics.**

- a) The distribution of political alignment among carnivores is about 25% conservative, 40% moderate, and 35% liberal. Omnivores were generally less conservative and more liberal, with about 10% conservative, 35% moderate, and 55% liberal. Vegetarians were even more liberal than the other groups, with 0% conservative, 30% moderate, and 70% liberal.
- b) The differences are quite large. There appears to be a strong association between diet preference and political alignment.

#### **6. Diet and politics revisited.**

- a) The distribution of diet preference among liberals is approximately 10% carnivore, 75% omnivore, and 15% vegetarian. Moderates tend to be slightly more carnivorous, with approximately 18% carnivores, 73% omnivores, and 9% vegetarians. Conservatives tend to be even more carnivorous, with 35% carnivores, 65% omnivores, and 0% vegetarians.
- b) Yes, the differences are real. It does appear there is a strong association between diet preference and political alignment.

#### **7. Fish and prostate cancer revisited**

- a) Looking at the horizontal axis only, approximately 1% of the men had prostate cancer.
- b) There are more men who didn't have cancer and never or seldom ate fish. The rectangles are approximately the same height, but the bar for "no cancer" is much wider.
- c) The percentage of men who never/seldom ate fish is lower in the group with no cancer than in the group with cancer. Disregard the width, and look only at the height to compare the conditional distribution of fish consumption within each cancer group. The bar for "never/seldom" is slightly shorter within the "no cancer" group.

#### **8. College value? revisited.**

- a) Approximately 33% (or one-third) of the respondents were college presidents. Look only at the horizontal axis to estimate this percentage.
- b) No, the excellence regions appear to have approximately the same areas, but because the U.S. adults region is so much wider, the regions are difficult to compare.
- c) Yes, it is clear that a greater percentage of college presidents said that college provides an excellent value, when compared to the percentage of U.S. adults who said the same thing. The bar for college presidents is much taller.
- d) The differences seem quite large. It does appear there is a strong association between the responses and whether the respondent was a college president. College presidents were much more likely to say that college provides an excellent or good value, and less likely to say that college provides a poor value than U.S. adults in general.

### **Section 3.4**

#### **9. Diet and politics III.**

**Men**

	Carnivore	Omnivore	Vegetarian	Total
Liberal	9	74	5	88
Moderate	12	54	1	67
Conservative	9	14	0	23
Total	30	142	6	178

## 9. (continued)

**Women**

	Carnivore	Omnivore	Vegetarian	Total
Liberal	4	53	12	69
Moderate	4	27	6	37
Conservative	1	4	0	5
Total	9	84	18	111

- a) Men are more likely to be conservative carnivores.  $9/178 \approx 5.1\%$  of the men are conservative carnivores, while only  $1/111 \approx 0.9\%$  of the women are conservative carnivores.
- b) Liberal vegetarians are more likely to be women. Of the 17 liberal vegetarians, 12 of them are women.  $12/17 \approx 70\%$  of liberal vegetarians are women.

10. **Being successful.**

- a) The table contains column percentages, so we can add percentages within each column.  $18\% + 48\% = 66\%$  of young women consider it very important or one of the most important things for them to be personally successful.
- b) The percentage is lower for men.  $11\% + 47\% = 58\%$  of young men consider it very important or one of the most important things for them to be personally successful.
- c) We cannot determine what percent of all women (from both age groups) feel this way, since we don't have the sample sizes from each age group, only the percentages.
- d) Young women are slightly more likely than young men (66% compared to 58%) to consider financial success important. Young men are slightly more likely to view this as somewhat important or not important.

**Chapter Exercises**11. **Movie genres and ratings.**

- a)  $452/1529 \approx 29.6\%$  of the films were rated R.
- b)  $124/1529 \approx 8.1\%$  of the films were R-rated comedies.
- c)  $124/452 \approx 27.4\%$  of the R-rated films were comedies.
- d)  $124/312 \approx 39.7\%$  of the comedies were rated R.

12. **Not the labor force.**

- a)  $2207/12872 \approx 17.1\%$  of the unemployed were available to work now.
- b)  $1048/12872 \approx 8.1\%$  of the unemployed were available to work now and aged 25 to 54 years.
- c)  $208/4158 \approx 5.0\%$  of the unemployed 16- to 24-year-olds were in school or training.
- d)  $4158/12872 \approx 32.3\%$  of the unemployed were aged 16 to 24 years.

13. **Tables in the news.**

Answers will vary.

14. **Graphs in the news.**

Answers will vary.

15. **Poverty and region 2012.**

The differences in poverty are not huge, but they may be real. The Northeast and Midwest have the lowest percentages of people living below the poverty level: 12.7% and 13.7%, respectively. In the West, 15.4% live below the poverty level, and the South has the highest rate at 16.8%.

## **24 Part I Exploring and Understanding Data**

### **16. Moviegoers and ethnicity.**

- a) Caucasians are 66% of the population but only 56% of ticket buyers. Hispanics are only 16% of the population, but they are 26% of ticket buyers. African Americans are more consistent: 12% of the population and 11% of the moviegoers.
- b) There are ethnic differences among moviegoers, but they are not all that pronounced.

### **17. Death from the sky.**

- a)  $100 - 60 - 30 - 1 - 0.2 - 0.17 = 8.63\%$  of estimated deaths are attributed to causes not listed here.
- b) Regardless of the type of display chosen, it is difficult to display causes of death with percentages as low as 0.2% and 0.17%. A display of the two largest values would not show the two smallest ones.

### **18. Cartoons.**

- a) Without a fun item: 61.4%, 59.1%, 61.4%, 17.0%, 22.7%  
With fun item: 48.8%, 52.5%, 55.0%, 20.0%, 27.5%

Using relative frequencies makes these easier to compare, since the number of students in each group is different.

- b) There is no evidence that cartoons help with understanding drawing a sample, the meaning of “random”, or mean versus median, since the percentage of students answering the item correctly was higher in the group without the fun item. The percentage of students correctly answering the question about categorical versus quantitative variables was higher for the group with the fun item, but not significantly higher. There appears to be some evidence that students learn better using a fun item when studying the inadequacy of mean without standard deviation. The percentage of students answering the question correct in the fun item group was 4.8 percentage points higher than in the group with no fun item, 27.5% to 22.5%. However, with only 27.5% accuracy, it can't be said that they mastered the topic.

### **19. Smoking.**

- a) The smoking rate for 18-24-year-old men was 42.1% in 1974.
- b) From 1974 to 2014, the smoking rate for 18-24-year-old men dropped from 42.1% to 18.5%
- c) Men who were 18-24 years old in 1974 are in the 35-44 age group in 1994, the 45-54 age group in 2004, and the 2014. The smoking rate for this cohort has been decreasing through the years, from 42.1% to 31.4% to 26.5% to 19.9%. Although we don't have data on deaths in this table, it may very well be that the smokers have a higher death rate than the non-smokers, so this decrease doesn't necessarily mean that men in this cohort are quitting smoking.

### **20. Smoking women.**

- a) In each year, the women had a lower smoking rate than the men. Women had a 18-24-year-old smoking rate of 34.0% in 1974, compared to the 42.1% rate for men of the same age. The 18-24-year-old female smoking rates decline slowly over the years, and by 2014, the smoking rate for women is 16.5%, only 2 percentage points lower than the men.
- b) Smokers are prone to lung cancer and other diseases. The smokers are less likely to live into old age and those left are less likely to be smokers.

### **21. Mothers and fathers 1965-2011.**

- a) Fathers spend the more of their time on paid work, while mothers spend more time on child care and house work.
- b) The time fathers spend on paid work has decreased, and the time they spend on child care and housework has increased. For mothers, the number of hours spent on paid work has significantly increased, and they have also increased their time spent on child care while reducing housework time.
- c) Parents are spending more time on child care and paid work (13 hours to 21 hours and 50 hours to 58 hours). The time spent on housework has decreased from 36 hours to 28 hours.

**21.** (continued)

- d)** Overall, parents in 2011 reported spending more time total on these tasks, a total of 107 hours in 2011 compared to 99 hours in 1965. Mothers increased their total working time by 3 hours, from 50 hours to 53 hours, while fathers increased their total working time by 5 hours, from 49 hours to 54 hours.

**22. Mothers' and fathers' aspirations.**

- a)** The actual percentages for working full-time differ the most, with 84% of fathers and 48% of mothers working full-time. This is a difference of 36 percentage points.
- b)** 48% of mothers work full-time, while only 27% desire to work full-time.
- c)** Only 7% of fathers stay at home, compared to 30% of mothers.

**23. Teen smokers.**

According to the Monitoring the Future study, teen smoking brand preferences differ somewhat by region. Although Marlboro is the most popular brand in each region, with about 58% of teen smokers preferring this brand in each region, teen smokers from the South prefer Newports at a higher percentage than teen smokers from the West, 22.5% to approximately 10%, respectively. Camels are more popular in the West, with 9.5% of teen smokers preferring this brand, compared to only 3.3% in the South. Teen smokers in the West are also more likely to have no particular brand than teen smokers in the South. 12.9% of teen smokers in the West have no particular brand, compared to only 6.7% in the South. Both regions have about 9% of teen smokers that prefer one of over 20 other brands.

**24. Being successful revisited.**

- a)** The four groups are approximately the same size, since the columns are approximately the same width.
- b)** *Age* is the more important factor. The male/female distributions are more similar to each another within each age group than the male/male and female/female distributions are across age groups. The older group is more likely to respond Not important; the younger more likely to respond Very important but not the most.
- c)** Women aged 18–34 are most likely to respond that being successful is the most important.

**25. Diet and politics IV.**

- a)** There are more men in the survey. The male columns are generally wider than the female columns.
- b)** We can't compare the genders within each category of political ideology since the sample sizes differ. We can, however, note that the male bars are narrowest in the Liberal category, and widest in the Conservative category, indicating that a small share of liberals and a large share of conservatives among the men. The women show the opposite association. The widest female bar is Liberal, while the narrowest is Conservative. In other words, there is an association between gender and political ideology. Males tend to be more conservative and females tend to be more liberal.
- c)** There is an association between politics and diet. Conservatives are more likely to be carnivores, while liberals are more likely to be vegetarians.
- d)** The association between politics and diet seems to differ between men and women. Differences in vegetarianism across political ideology is more pronounced in females than in males. Differences in carnivorous eating habits across political ideology is more pronounced in males than females.

**26. Handguns.**

76.4% of handguns involved in Milwaukee buyback programs are small caliber, while only 20.3% of homicides are committed with small caliber handguns. Along the same lines, only 19.3% of buyback handguns are of medium caliber, while 54.7% of homicides involve medium caliber handguns. A similar disparity is seen in large caliber handguns. Only 2.1% of buyback handguns are large caliber, but this caliber is used in 10.8% of homicides. Finally, 2.2% of buyback handguns are of other calibers, while 14.2% of homicides are committed with handguns of other calibers. Generally, the handguns that are involved in buyback programs are not the same caliber as handguns used in homicides in Milwaukee.

## **26 Part I Exploring and Understanding Data**

### **27. Job satisfaction.**

- a) This is a table of column percents. The columns add up to 100%, while the rows do not.
- b) i) This can't be found from the table. We don't know what the percent of respondents who are very satisfied.
- ii) This can't be found from the table. We don't know what the percent of respondents who are dissatisfied.
- iii) 39% of respondents who are dissatisfied with their current job are actually better off than their parents were at the same age.
- iv) This can't be found from the table. We don't know what the percent of respondents who are very satisfied.

### **28. A sense of identity.**

- a) These are row percents. The rows add up to nearly 100%, while the columns do not. The differences from 100% are too large to be attributed to rounding error, so it is likely that some respondents replied something else other than the provided responses.
- b) i) This can't be found from the table. We don't know what the percent of respondents who are self-employed.
- ii) 65% of respondents who are employed by non-profits get a sense of identity from their job.
- iii) This can't be found from the table. We don't know what the percent of respondents who think of their job as just something to do for a living.
- iv) This can't be found from the table. We don't know what the percent of respondents who get a sense of identity from their job.

### **29. Seniors.**

- a) A table with marginal totals is shown below. There are 268 White graduates and 325 total graduates.  $268/325 \approx 82.5\%$  of the graduates are white.

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>

- b) There are 42 graduates planning to attend 2-year colleges.  $42/325 \approx 12.9\%$
- 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 white graduates.  $36/268 \approx 13.4\%$
- e) There are 42 graduates planning to attend 2-year colleges, and 36 of them are white.  $36/42 \approx 85.7\%$

### **30. Politics.**

- a) There are 192 students taking Intro Stats. Of those, 115, or about 59.9%, are male.
- b) There are 192 students taking Intro Stats. Of those, 27, or about 14.1%, consider themselves to be "Conservative".
- c) There are 115 males taking Intro Stats. Of those, 21, or about 18.3%, consider themselves to be "Conservative".

**30.** (continued)

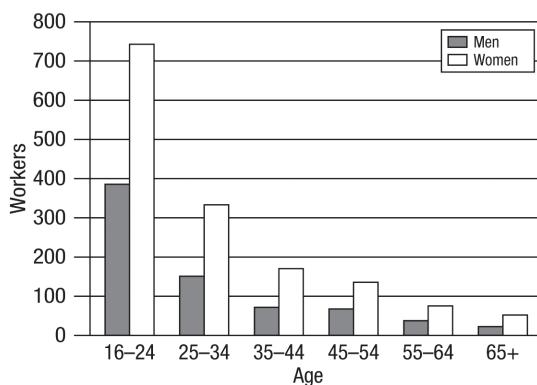
- d) There are 192 students taking Intro Stats. Of those, 21, or about 10.9%, are males who consider themselves to be “Conservative”.

**31. Movies 06-15.**

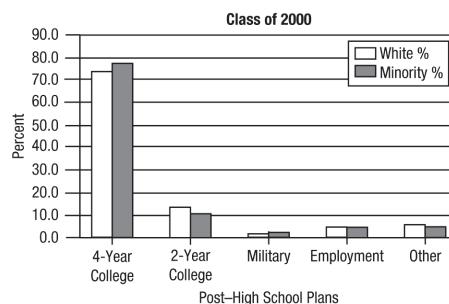
- a) This is a table of column percents. The columns add up to 100%, while the rows do not.
- b) Movies rated G and PG have become slightly less common, while movies rated PG-13 and R have become slightly more common.
- c) For Dramas, the percentages of PG and R have decreased while the percentage of PG-13 dramas has increased significantly. For Comedies, there has been a large increase in the percentage of R-rated films.

**32. Minimum wage workers.**

- a)  $7701/37972 \approx 20.3\%$  of the women are aged 16-24.
- b) The number of workers at or below minimum wage is highest for each sex at 16-24 years, with 384,000 male workers and 738,000 female workers. The number of workers at or below minimum wage decreases as the age increases for each sex, with only 22,000 male and 50,000 female workers. At each age category, the number of male workers at or below minimum wage is roughly half the number of female workers.


**33. More about seniors.**

- a) For white students, 73.9% plan to attend a 4-year college, 13.4% plan to attend a 2-year college, 1.5% plan on the military, 5.2% plan to be employed, and 6.0% have other plans.
- b) For minority students, 77.2% plan to attend a 4-year college, 10.5% plan to attend a 2-year college, 1.8% plan on the military, 5.3% plan to be employed, and 5.3% have other plans.
- c) A side-by-side bar chart is a good display of these data.



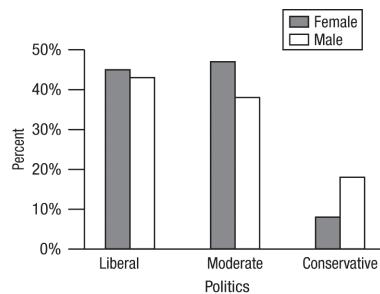
## 28 Part I Exploring and Understanding Data

### 33. (continued)

- d) 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.

### 34. Politics revisited.

- a) The females in this course were 45.5% Liberal, 46.8% Moderate, and 7.8% Conservative.  
b) The males in this course were 43.5% Liberal, 38.3% Moderate, and 18.3% Conservative.  
c) A side-by-side bar chart comparing the distributions is shown below.



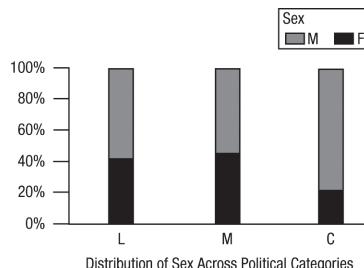
- d) Politics and sex do not appear to be independent in this course. Although about the same fraction of males and females classified themselves as politically Liberal, females were slightly more likely to be Moderate. A much higher percentage of males (18%) considered themselves to be Conservative than did the females (8%).

### 35. Magnet schools revisited.

- a) There were 1755 qualified applicants to the Houston Independent School District's magnet schools program. Of those, 292, or about 16.6% were Asian.  
b) There were 931 students accepted to the magnet schools program. Of those, 110, or about 11.8% were Asian.  
c) There were 292 Asian applicants. Of those, 110, or about 37.7%, were accepted.  
d) There were 1755 total applicants. Of those, 931, or about 53%, were accepted.

### 36. More politics.

a)



- b) The overall percentage of females in the class is about 40%. That's true among Liberals and Moderates, but the percentage of females among Conservatives is only about half that. This suggests that sex and politics are not independent.

**37. Back to school.**

There were 1755 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 25.3% of those turned away. It appears that the admissions decisions were not independent of the applicant's ethnicity.

**38. Parking lots.**

- 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	Percentages
American	212	59.05%
European	45	12.53%
Asian	102	28.41%
<b>Total</b>	<b>359</b>	<b>99.99%</b>

- e) The conditional distributions are shown below.

Student *Origin* distribution

Origin	Driver	
	Student	Percentage
American	107	54.87%
European	33	16.92%
Asian	55	28.21%
<b>Total</b>	<b>195</b>	<b>100.00%</b>

Staff *Origin* distribution

Origin	Driver	
	Staff	Percentage
American	105	64.02%
European	12	7.32%
Asian	47	28.66%
<b>Total</b>	<b>164</b>	<b>100.00%</b>

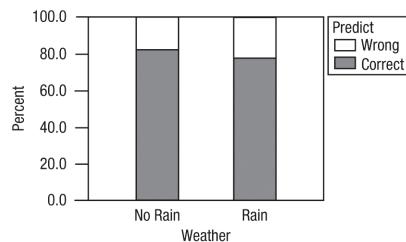
- f) No, the marginal distributions look slightly different. A higher percentage of staff drive American cars, 64% compared to about 55% among students. Bar charts or pie charts could be used to compare.

## 30 Part I Exploring and Understanding Data

### 39. Weather forecasts.

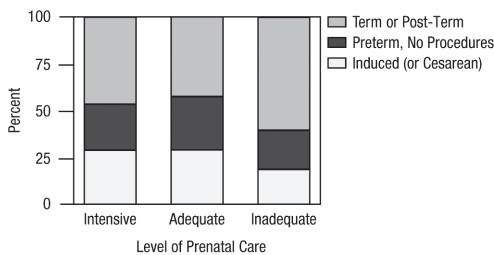
Forecast	Actual Weather		Total
	Rain	No Rain	
Rain	27	63	90
No Rain	7	268	275
Total	34	331	365

- a) The table above 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, or  $90/365 \approx 24.7\%$  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) On rainy days, rain had been predicted 27 out of 34 times (79.4%). On days when it did not rain, forecasters were correct in their predictions 268 out of 331 times (81.0%). These two percentages are very close. There is no evidence of an association between the type of weather and the ability of the forecasters to make an accurate prediction.



### 40. Twin births.

- a) Of the 278,000 mothers who had twins in 1995–1997, 63,000 had inadequate health care during their pregnancies, or  $63,000/278,000 = 22.7\%$ .
- b) There were 76,000 induced or Caesarean births and 71,000 preterm births without these procedures, or  $(76,000 + 71,000)/278,000 = 52.9\%$ .
- 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, or  $(12,000 + 13,000)/63,000 = 39.7\%$ .
- d) A segmented bar chart is shown below.



**40.** (continued)

- 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.

**41. Blood pressure.**

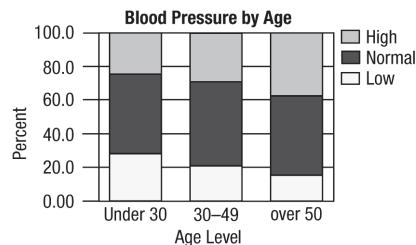
- a) The marginal distribution of blood pressure for the employees of the company is the total column of the table, converted to percentages. 20% low, 49% normal and 31% high blood pressure.

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 shown below.

Blood pressure	under 30	30–49	over 50
low	27.6%	20.7%	15.7%
normal	49.0%	50.8%	47.2%
high	23.5%	28.5%	37.1%

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



- 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.

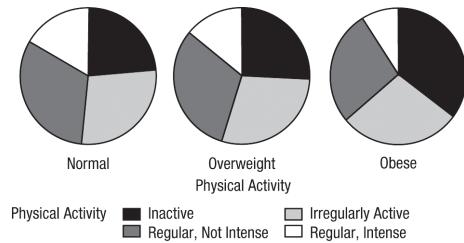
**42. Obesity and exercise.**

- 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.

## 32 Part I Exploring and Understanding Data

### 42. (continued)

- b) Pie charts of the conditional distributions of level of physical activity by Body Mass Index category are shown below.



- 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 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 physically activity.

### 43. Anorexia.

These data provide no evidence that Prozac might be helpful in treating anorexia. Of the patients who took Prozac, 71.4% were diagnosed as “Healthy” (or 28.6% relapsed), while about 72.7% of the patients who took a placebo were diagnosed as “Healthy” (or 27.3% relapsed). Even though the percentage was higher for the placebo patients, this does not mean that Prozac is hurting patients. The difference between 71% and 73% is not likely to be statistically significant.

### 44. Antidepressants and bone fractures.

These data provide evidence that taking a certain class of antidepressants (SSRI) might be associated with a greater risk of bone fractures. Approximately 10% of the patients taking this class of antidepressants experience bone fractures. This is compared to only approximately 5% in the group that were not taking the antidepressants.

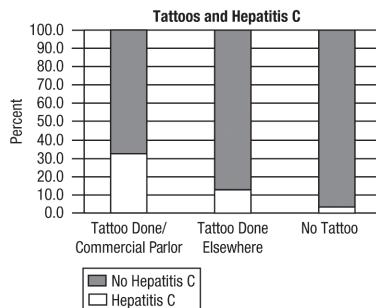
### 45. Driver's licenses 2014.

- There are 8.5 million drivers under 20 and a total of 214.1 million drivers in the U.S. That's about 4.0% of U.S. drivers under 20.
- There are 105.9 million males out of 214.1 million total U.S. drivers, or about 49.5%.
- Each age category appears to have about 50% male and 50% female drivers. At younger ages, males form the slight majority of drivers. The percentage of male drivers shrinks until female drivers hold a slight majority, among older drivers.
- 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.

### 46. Tattoos.

The study by the University of Texas Southwestern Medical Center provides evidence of a strong 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.

46. (continued)



47. Diet and politics shuffled.

- a) The upper right bar chart shows the original data.
- b) It's clear that shuffled data did not produce associations as obvious as the original data. The randomly scrambled data look different from the original data, and this supports the belief that there is an association between the variables.

48. College values shuffled.

- a) The bottom left mosaic plot shows the original data.
- b) It's clear that none of the shuffled data produced associations as strong as the original data. All the mosaic plots from the randomly scrambled data look different from the original data, and this supports the belief that there is an association between the variables.

49. Hospitals.

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

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

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

- b) Yes. 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.

## 34 Part I Exploring and Understanding Data

### 49. (continued)

- f) The small hospital performs a higher percentage of minor surgeries than major surgeries. 250 of 300 surgeries at the small hospital were minor (83.3%). 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.

### 50. 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.
- 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.

### 51. Graduate admissions.

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

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 below 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%	7.0%

- 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.

**52. Be a Simpson.**

Answers will vary. The three-way table below shows one possibility. The number of local hires out of new hires is shown in each cell.

	<b>Company A</b>	<b>Company B</b>
Full-time New Employees	40 of 100 = 40%	90 of 200 = 45%
Part-time New Employees	170 of 200 = 85%	90 of 100 = 90%
Total	210 of 300 = 70%	180 of 300 = 60%

