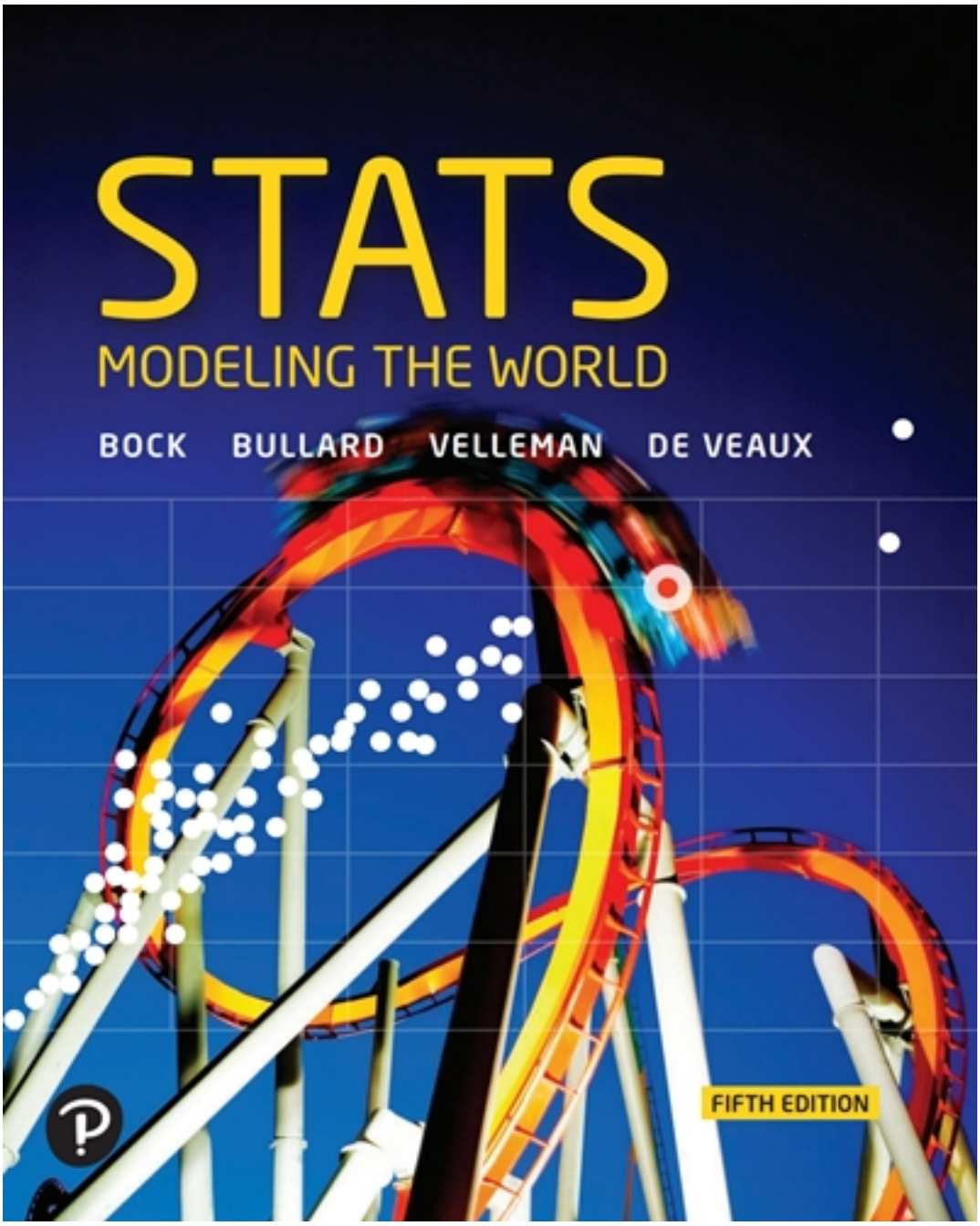
Solutions for Stats Modeling the World 5th Edition by Bock

CLICK HERE TO ACCESS COMPLETE Solutions



Solutions

Chapter 2 – Displaying and Describing 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. **Movie genres.**
   - **a)** A pie chart seems appropriate from the movie genre data. Each movie has only one genre, and the 891 movies constitute a "whole".
   - **b)** "Other" is the least common genre. It has the smallest region in the chart.
6. **Movie ratings.**
   - **a)** A pie chart seems appropriate for the movie rating data. Each movie has only one rating, and the 891 movies constitute a "whole".
   - **b)** The most common rating is R. It has the largest region on the chart.
7. **Movie genres again.**
   - **a)** Thriller/Suspense films were more common than Adventure films. The bar for Thriller/Suspense is taller than the bar for Adventure.
   - **b)** This is easier to see on the bar chart. The percentages are so close that the difference is nearly indistinguishable in the pie chart. Also, the bar chart is organized by height while the pie chart is not, making it difficult to compare genres with areas similar in proportion.
8. **Movie ratings again.**
   - **a)** The least common rating was NC-17. It has the shortest bar.
   - **b)** While it is easy in both the pie chart and the bar chart, it may be easier in the pie chart. In the pie chart, ratings are ordered clockwise by increasing area while in the bar chart, G and NC-17 are inconsistent with an increasing order.
9. **Movie ratings.**

   i. C (This chart has 4 ratings and the proportion of G ratings is smallest.)

   ii. A (This chart has 4 ratings and a slightly higher proportion of G ratings.)

   iii. D (This chart has 3 ratings with PG more common than R.)

   iv. B (This chart has 3 ratings with PG and R roughly equally as common.)
10. **Marriage in decline.**

    i. D (This bar and pie chart are the only ones for which Bad Thing and No Difference are not roughly equally common)

ii. A (Bad Thing and No Difference are roughly equal, while Don't Know/No Response is noticeably the least frequent response)

iii. B or C (These charts are indistinguishable.)

iv. B or C (These charts are indistinguishable.)

11. Magnet schools.

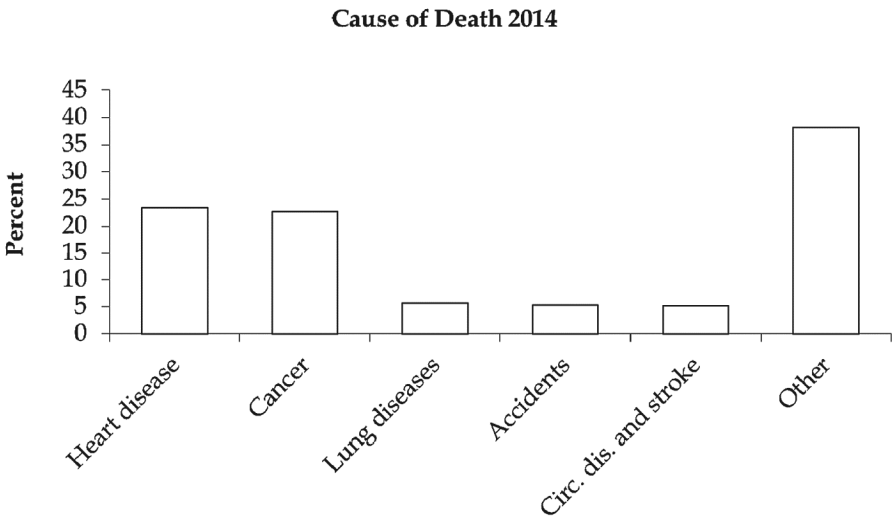
There were 1,755 qualified applicants for the Houston Independent School District's magnet schools program. Approximately 53% were accepted, 17% were wait-listed, and the other 30% were turned away for lack of space.

12. Magnet schools again.

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

13. Causes of death 2014.

a) Yes, it is reasonable to assume that heart or lung diseases caused approximately 29% of U.S. deaths in 2014, since there is no possibility for overlap. Each person could only have one cause of death.



b) Since the percentages listed add up to 61.9%, other causes must account for 38.1% of US deaths.

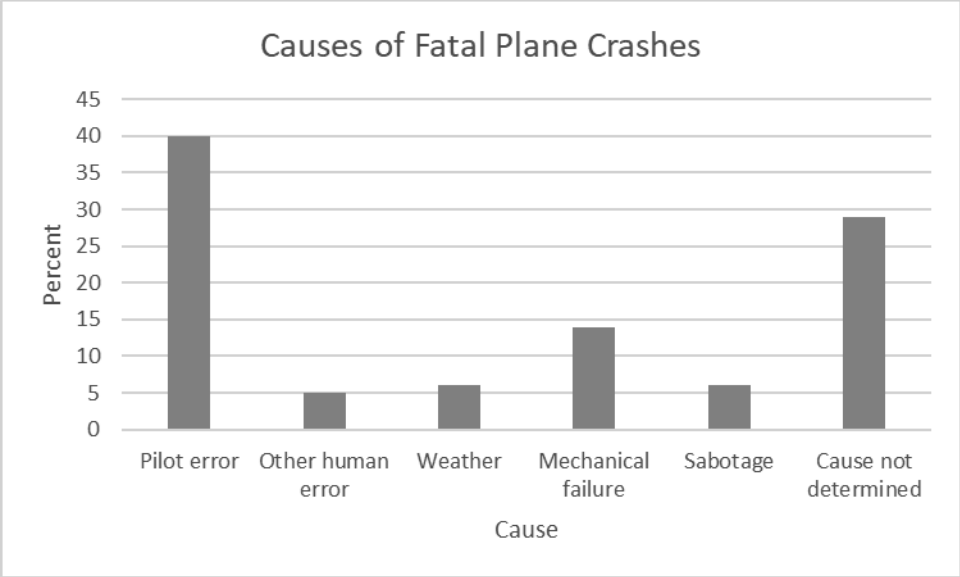
c) A bar chart is a good choice (with the inclusion of the "Other" category). Since causes of US deaths represent parts of a whole, a pie chart would also be a good display.

14. Plane crashes.

a) As long as each plane crash had only one cause, it would be reasonable to assume that weather or mechanical failures were the causes of about 20% of recent plane crashes.

b) Since the percentages listed add up to 71%, other causes (not determined) must account for 29% of recent plane crashes.

c) A relative frequency bar chart is a good choice. A pie chart would also be a good display, as long as each plane crash has only one cause.



15. Oil spills as of 2016.

a) Grounding, accounting for 150 spills, is the most frequent cause of oil spillage for these 460 spills. A substantial number of spills, 136, were caused by collision. Less prevalent causes of oil spillage in descending order of frequency were hull failures, other/unknown causes, fire/explosions, and equipment failure.

b) If being able to differentiate between these close counts is required, use the bar chart. Since each spill only has one cause, 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 fire/explosions or hull failure. If you want to showcase the causes of oil spills as a fraction of all 460 spills, use the pie chart.

16. Winter Olympics 2016.

a) There are too many categories to construct an appropriate display. In a bar chart, there are too many bars. In a pie chart, there are too many slices. In each case, we run into difficulty trying to display those countries that didn't win many medals.

b) Perhaps we are primarily interested in countries that won many medals. We might choose to 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 use a bar chart. A bar chart is also better for comparisons.

17. Global warming.

Perhaps the most obvious error is that the percentages in the pie chart add up to 141%, when they should, of course, add up to 100%. This means that each individual region and any resulting sums will occupy less area of the display than their percentages imply. Furthermore, the three-dimensional perspective view distorts the regions in the graph, violating the area principle. The regions corresponding to "Global warming isn't happening" and "Can't reduce global warming even if it is happening" should be the same size, at 25% of respondents. However, the "Global warming isn't happening" region looks bigger. Always use simple, two-dimensional graphs.

18. Modalities.

a) The bars have false depth, which can be misleading. This is a bar chart, so the bars should have space between them.

b) Since each trainer was asked to list 3 modalities, the expected sum should be 300% rather than 100%.

19. 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 to 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 9% of teen smokers that prefer one of over 20 other brands.

20. Handguns.

76% 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.

21. Movie genres and ratings.

a) 452 of these films were rated R. 452/1,529 ≈ 29.56%

b) 124 of these films were R-rated comedies. 124/1,529 ≈ 8.1%

c) 124 of the 452 R-rated films were comedies. 124/452 ≈ 27.43%

d) 124 of the 312 comedies were R-rated. 124/312 ≈ 39.74%

22. Not the labor force.

a) 2207 of the unemployed population were available to work now. 2207/12,872 ≈ 17.1%

b) 1,048 of the unemployed population were available to work now and aged 25 to 54 years. 1,048/12,872 ≈ 8.14%

c) 208 of 4,158 unemployed 16-24 year olds were in school or training. 208/4,158 ≈ 5%

d) 4,158 of the unemployed population were aged 16-24 years. 4,158/12,872 ≈ 32.3%

23. Seniors.

A table with marginal totals is to the right.

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
TOTAL	268	57	325

a) 268 seniors were white. 268/325 ≈ 82.5%

b) 42 seniors are planning to attend a 2-year college. 42/325 ≈ 13%

c) 36 seniors are white and planning to attend 2-year colleges. 36/325 ≈ 11.1%

d) 36 of the 268 white seniors are planning to attend 2-year colleges. 36/268 ≈ 13.4%

e) There are 42 graduates planning to attend 2-year colleges. 36 are white. 36/42 ≈ 85.7%

24. Politics.

a) There are 192 students taking Intro Stats. Of those, 115 are male. 115/192 ≈ 59.9%.

b) 27 students in the course consider themselves to be "Conservative". 27/192 ≈ 14%.

c) There are 115 males taking Intro Stats. Of those, 21 consider themselves to be "Conservative". 21/115 ≈ 18.26%.

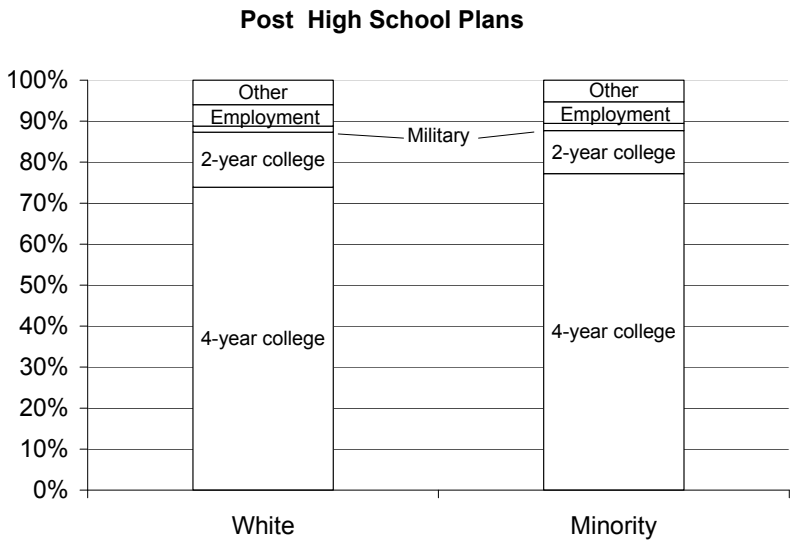
d) 21 of the students in the course are males who consider themselves to be "Conservative". 21/192 ≈ 10.94%

25. 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 segmented bar chart is a good display of these data:



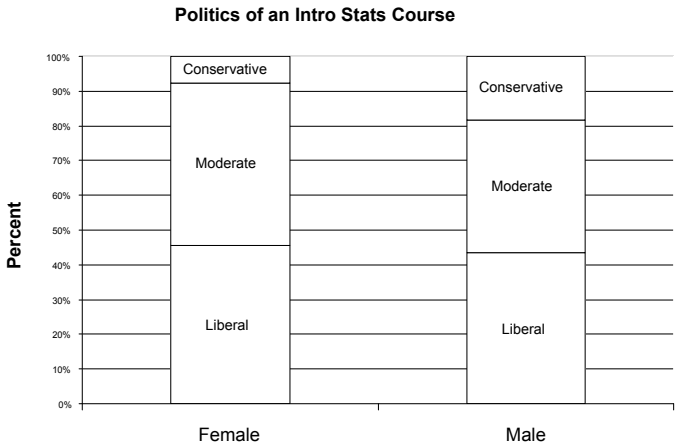
d) The conditional distributions of plans for Whites and Minorities are similar:
White – 74% 4-year college, 13% 2-year college, 2% military, 5% employment, 6% other.
Minority – 77% 4-year college, 11% 2-year college, 2% military, 5% employment, 5% other.
Caution should be used with the percentages for 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.

26. 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 segmented bar chart comparing the distributions is at the right.



d) Politics and sex do not appear to be independent in this course. Although the percentage of liberals was roughly the same for each sex, females had a greater percentage of moderates and a lower percentage of conservatives than males.

27. Magnet schools revisited.

a) There were 1,755 qualified applicants to the Houston Independent School District's magnet schools program. Of those, 292 were Asian. $292/1{,}755 \approx 16.6\%$

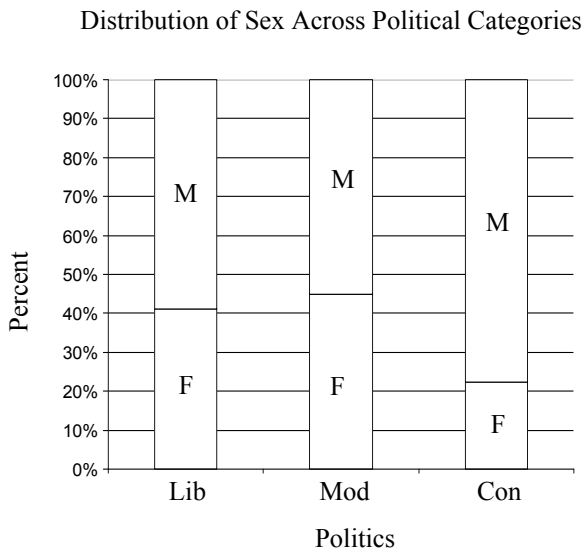
b) There were 931 students accepted to the magnet schools program. Of those, 110, were Asian. $110/931 \approx 11.8\%$.

c) There were 292 Asian applicants. Of those, 110 were accepted. $110/292 \approx 37.7\%$.

d) There were 1,755 total applicants. Of those, 931 were accepted. $931/1{,}755 \approx 53\%$.

28. More politics.

a)



b) 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 Conservatives who are female. This suggests that *sex* and *politics* are not independent.

29. 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 25.3% of those turned away. It appears that the admissions decisions were not independent of the applicant's ethnicity.

30. 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 ≈ 40.95%.

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

b) There are 212 American cars of which 107 or 107/212 ≈ 50.47% were owned by students.

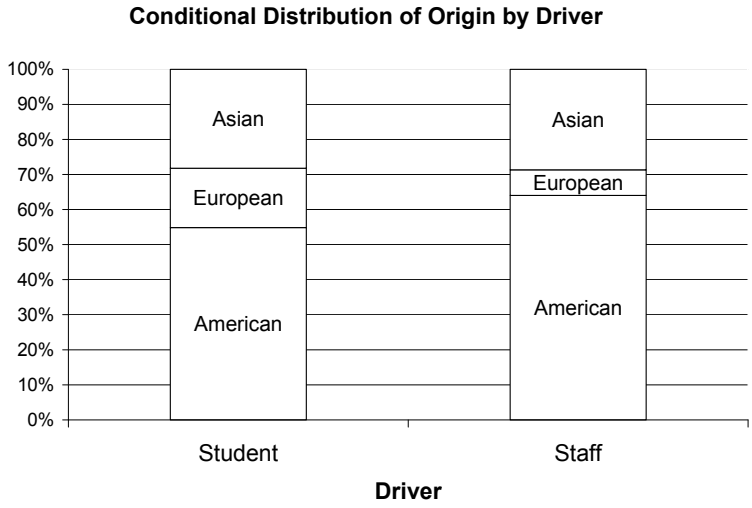
c) There are 195 students of whom 107 or 107/195 ≈ 54.87% owned American cars.

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%)
Total	359

e) The conditional distribution of Origin for Students is: 54.8% (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 of the car.

31. Super students.

a) Some possible observations: The most common choice of super power among students was the ability to Fly at just over 30%. The least common choice was Super Strength at less than 10%. Freeze Time and Telepathy were both chosen by just over 20% of students, the second most frequent choices.

b) It is likely that the ability to Fly is characteristic of the whole population because it was chosen both most frequently and just over 30% in each of the samples. Super Strength is likely to be the least popular choice in the population as it was the least selected characteristic in two of the three other samples. In the third sample, it was only slightly more popular than Invisibility but still by only 10% of students. It is also likely that Freeze Time and Telepathy are equally desirable and just above 20%. Telepathy was consistently chosen by slightly more than 20%. Freeze Time was chosen by about 26% of students in one sample, but was consistently at 20% in the others.

32. Super students II.

a) Some possible observations: Males were about twice as likely to select Fly for their Super Power compared to females. Males were also more likely to choose Freeze Time or Super Strength than females. Females, on the other hand, were far more likely to select Telepathy than males, about 40% to 8% respectively.

b) It appears that observation that males were twice as likely to choose Fly compared to females was a quirk of the first sample. Subsequent samples suggest the females are equally as likely to choose Fly as males. The three new samples support the claim that males are more likely to choose Freeze Time and Super Strength compared to females. While females in the subsequent surveys were more likely to choose Telepathy compared to males, consistent with the first sample, the contrast may not be as severe as in two of the samples females were only twice as likely to choose Telepathy.

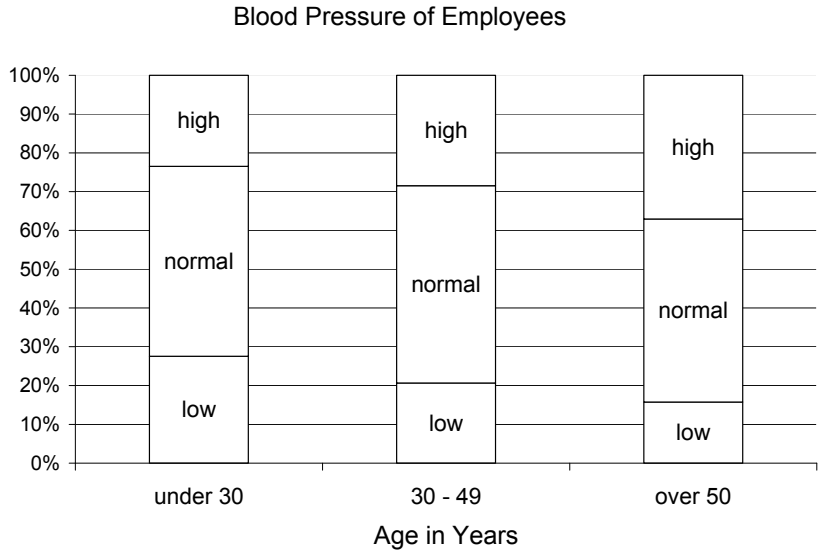
33. Blood pressure.

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
Total	98	179	197	474

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 below.



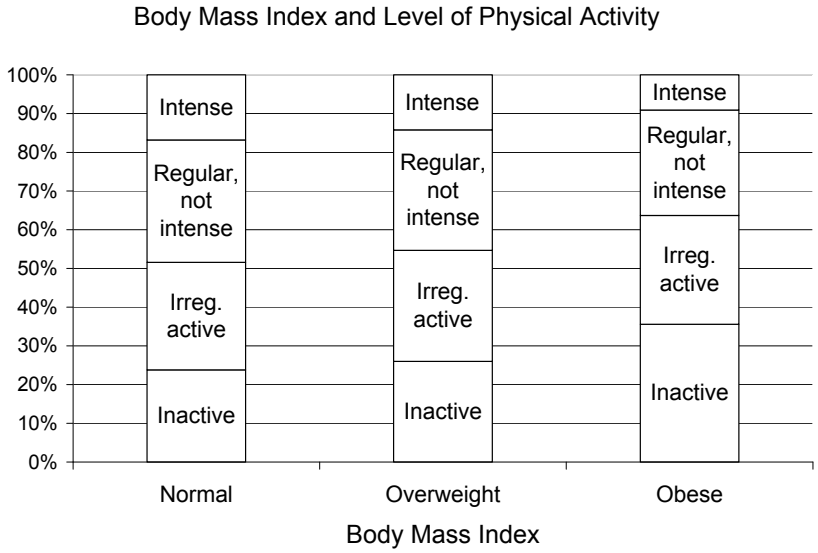
d) In this company, as age increases, the percentage of employees with low blood pressure decreases, and the percentage of employees with high blood pressure increases.

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.

34. 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 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.

35. Anorexia.

These data provide no evidence that Prozac might be helpful in treating anorexia. About 71% of the patients who took Prozac were diagnosed as "Healthy", while about 73% of the patients who took a placebo were diagnosed as "Healthy". 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.

36. 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.

37. Driver's licenses 2014.

a) There are 8.5 million drivers under 20 and a total of 214.1 million drivers in the U.S. $8.5/214.1 \approx 3.97\%$

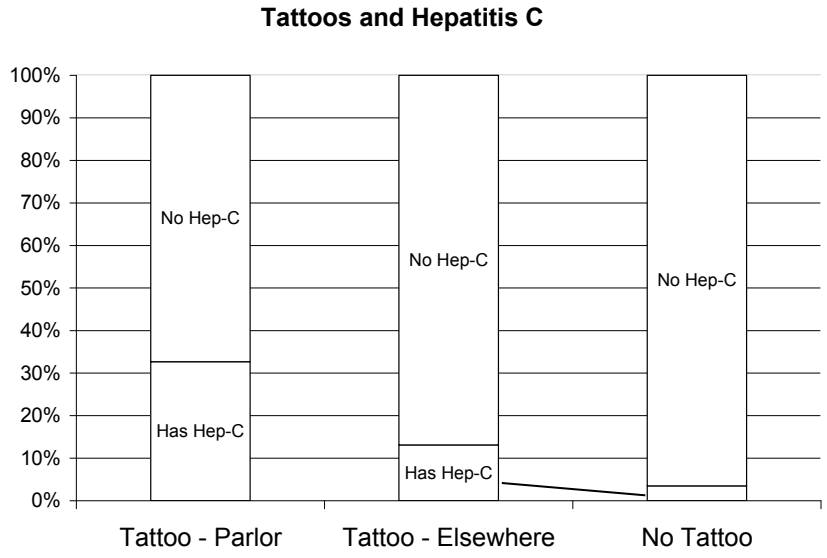
b) There are 105.9 million males out of 214.1 million U.S. drivers. $105.9/214.1 \approx 49.46\%$

c) Each age category appears to have about 50% male and 50% female drivers. At the youngest ages, males form the slight majority of drivers. This percentage shrinks until the percentages are 50% male and 50% female for middle aged drivers. The percentage of male drivers continues to shrink until, at around age 45, female drivers hold a slight majority. This continues into the 85 and over category.

d) There is 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.

38. 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.



39. 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) 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.

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.

40. 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.

41. 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
Total	**1022 of 2165**	**262 of 849**	**1284 of 3014**

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.

42. 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.

	Company A	Company 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%