

Chapter 1: Introduction to Data

Section 1.2: Classifying and Storing Data

- 1.1 There are nine variables: “Male”, “Age”, “Eye Color”, “Shoe Size”, “Height, Weight”, “Number of Siblings”, “College Units This Term”, and “Handedness”.
- 1.2 There are eleven observations.
- 1.3 a. Handedness is categorical.
b. Age is numerical.
- 1.4 a. Shoe size is numerical.
b. Eye color is categorical.
- 1.5 Answers will vary but could include such things as number of friends on Facebook or foot length. *Don't copy these answers.*
- 1.6 Answers will vary but could include such things as class standing (“Freshman”, “Sophomore”, “Junior”, or “Senior”) or favorite color. *Don't copy these answers.*
- 1.7 The label would be “Brown Eyes” and there would be eight 1's and three 0's.
- 1.8 There would be nine 1's and two 0's.
- 1.9 Male is categorical with two categories. The 1's represent males, and the 0's represent females. If you added the numbers, you would get the number of males, so it makes sense here.
- 1.10

Units	Full
16.0	1
13.0	1
5.0	0
15.0	1
19.5	1
11.5	0
9.5	0
8.0	0
13.5	1
12.0	1
14.0	1

- 1.12 a. The data is unstacked.
b. Labels for columns will vary.

Age	5 p.m.
31	1
34	1
46	1
47	1
50	1
24	0
18	0
21	0
20	0
20	0

- 1.11 a. The data is stacked.
b. 1 means male and 0 means female.
c.

Female	Male
9.5	9.4
9.5	9.5
9.9	9.5
	9.7

1.13 a. Stacked and coded

Calories	Sweet
90	1
310	1
500	1
500	1
600	1
90	1
150	0
600	0
500	0
550	0

The second column could be labeled “Salty” with the 1’s being 0’s and the 0’s being 1’s.

b. Unstacked

Sweet	Salty
90	150
310	600
500	500
500	550
600	
90	

1.14 a. Stacked and coded

Cost	Male
10	1
15	1
15	1
25	1
12	1
8	0
30	0
15	0
15	0

The second column could be labeled “Female” with the 1’s being 0’s and the 0’s being 1’s.

b. Unstacked

Male	Female
10	8
15	30
15	15
25	15
12	

Section 1.3: Organizing Categorical Data

1.15 a.

	Men	Women	Total
Yes, Older S	12	55	$12 + 55 = 67$
No, Older S	11	39	50
Total	23	$55 + 39 = 94$	117

b. $12/23 = 52.2\%$

c. $11/23 = 47.8\%$

d. $55/94 = 58.5\%$

e. $67/117 = 57.3\%$

f. $55/67 = 82.1\%$

g. $0.585(600) = 351$

1.16 a.

	Men	Women	Total
Work	15	65	$15 + 65 = 80$
Not Work	23	28	51
Total	38	$65 + 28 = 93$	131

b. $15/38 = 39.5\%$

c. $23/38 = 60.5\%$

d. $65/93 = 69.9\%$

e. $80/131 = 61.1\%$

f. $65/80 = 81.25\%$

g. $15/80 = 18.75\%$

h. $65/93 \times 800 = 0.698925(800) = 559$

- 1.17 a. $15/38$, or 39.5%, of the class were male.
 b. $0.641(234) = 149.99$, or about 150, men in the class
 c. $0.40x = 20$

$$x = \frac{20}{0.40}$$

$$= 50 \text{ people in the class}$$

- 1.18 a. $0.35(346) = 121$ male nurses
 b. $66/178 = 37.1\%$ female engineers
 c. $0.65x = 169$

$$x = \frac{169}{0.65}$$

$$= 260 \text{ lawyers}$$

1.19 The frequency of women is 7, the proportion is $7/11$, and the percentage is 63.6%.

1.20 The frequency of righties is 9, the proportion is $9/11$, and the percentage is 81.8%.

1.21 The answers follow the steps given in the Guided Exercises.

a. and b.

	Men	Women	Total
Right	4	5	9
Left	0	2	2
Total	4	7	11

- c. $5/7 = 71.4\%$ e. $9/11 = 81.8\%$
 d. $5/9 = 55.6\%$ f. $0.714(70) = 50$

1.22 a. and b.

	Men	Women	Total
Brown	3	5	8
Blue	1	1	2
Hazel	0	1	1
Total	4	7	11

- c. $5/7 = 71.4\%$ e. $8/11 = 72.7\%$
 d. $5/8 = 62.5\%$ f. $0.714(60) = 42.84$ or about 43

1.23 $0.202x = 88,547,000$

$$x = \frac{88,547,000}{0.202}$$

$$x = 438,351,485 \text{ (final value could be rounded differently)}$$

1.24 $0.055x = 12,608,000$

$$x = \frac{12,608,000}{0.055}$$

$$x = 229,236,364 \text{ (final value could be rounded differently)}$$

1.25 The answers follow the steps given in the Guided Exercises.

1–3:

State	AIDS/HIV	Rank Cases	Population	Population (thousands)	AIDS/HIV per 1000	Rank Rate
New York	192,753	1	19,421,005	19,421	$\frac{192,753}{19,421} = 9.92$	2
California	160,293	2	37,341,989	37,342	$\frac{160,293}{37,342} = 4.29$	5
Florida	117,612	3	18,900,773	18,901	$\frac{117,612}{18,901} = 6.22$	3
Texas	77,070	4	25,258,418	25,258	$\frac{77,070}{25,258} = 3.05$	6
New Jersey	54,557	5	8,807,501	8,808	$\frac{54,557}{8808} = 6.19$	4
District of Columbia	9257	6	601,723	602	$\frac{9257}{602} = 15.38$	1

4: No, the ranks are not the same. The District of Columbia had the highest rate and had the lowest number of cases. (Also, the rate for Florida puts its rank above California, and the rate for New Jersey puts it above Texas in ranking.)

5: The District of Columbia is the place (among these six regions) where you would be most likely to meet a person diagnosed with AIDS/HIV, and Texas is the place (among these six regions) where you would be least likely to do so.

1.26 a.

State	Population	Area (square miles)	Population Density	Rank
Pennsylvania	12,448,279	44,817	$\frac{12,448,279}{44,817} = 277.76$	3
Illinois	12,901,563	55,584	$\frac{12,901,563}{55,584} = 232.11$	5
Florida	18,328,340	53,927	$\frac{18,328,340}{53,927} = 339.87$	2
New York	19,490,297	47,214	$\frac{19,490,297}{47,214} = 412.81$	1
Texas	24,326,974	261,797	$\frac{24,326,974}{261,797} = 92.92$	6
California	36,756,666	155,959	$\frac{36,756,666}{155,959} = 235.68$	4

b. Texas has the lowest population density.

c. New York has the highest population density.

1.27

Year	Percentage
1990	$\frac{112.6}{191.8} = 58.7\%$
1997	$\frac{116.8}{207.2} = 56.4\%$
2000	$\frac{120.2}{213.8} = 56.2\%$
2007	$\frac{129.9}{235.8} = 55.1\%$

The percentage of married people is decreasing over time (at least with these dates).

1.28

Year	Percentage
2006	$\frac{2426}{4266} = 56.9\%$
2007	$\frac{2424}{4316} = 56.2\%$
2008	$\frac{2473}{4248} = 58.2\%$
2009	$\frac{2437}{4131} = 59.0\%$
2010	$\frac{2452}{4007} = 61.2\%$

The rate of death as a percentage of the rate of birth tends to go up over this time period. This is primarily due to the birth rate decreasing.

1.29 We don't know the percentage of female students in the two classes. The larger number of women at 8 a.m. may just result from a larger number of students at 8 a.m., which may be because the class can accommodate more students because perhaps it is in a large lecture hall.

1.30 We don't know the rate of fatalities—that is, the number of fatalities per pedestrian. There may be fewer pedestrians in Hillsborough County, and that may be the source of the difference.

Section 1.4: Collecting Data to Understand Causality

1.31 Observational study

1.35 Controlled experiment

1.32 Observational study

1.36 Observational study

1.33 Controlled experiment

1.37 Observational study

1.34 Controlled experiment

1.38 Controlled experiment

1.39 This was an observational study, and from it you cannot conclude that the tutoring raises the grades.

Possible confounders (answers may vary): 1. It may be the more highly motivated who attend the tutoring, and this motivation is what causes the grades to go up. 2. It could be that those with more time attend the tutoring, and it is the increased time studying that causes the grades to go up.

1.40 a. If the doctor decides on the treatment, you could have bias.

b. To remove this bias, randomly assign the patients to the different treatments.

c. If the doctor knows which treatment a patient had, that might influence his opinion about the effectiveness of the treatment.

d. To remove that bias, make the experiment double-blind. The talk-therapy only patients should get a placebo, and no one should know whether they have a placebo or antidepressant.

1.41 a. It was a controlled experiment, as you can tell by the random assignment. This tells us that the researchers determined who received which treatment.

b. We can conclude that the early surgery caused the better outcomes, because it was a randomized controlled experiment.

- 1.42 This is an observational study, because researchers did not determine who received PCV7 and who did not. You cannot conclude causation from an observational study. We must assume that it is possible that there were confounding variables (such as other advances in medicine) that had a good effect on the rate of pneumonia.
- 1.43 Answers will vary. However, they should all mention randomly dividing the 100 people into two groups and giving one group the copper bracelets. The other group could be given (as a placebo) bracelets that look like copper but are made of some other material. Then the pain levels after treatment could be compared.
- 1.44 a. Heavier people might be more likely to choose to eat meat. Also, people who are not prepared to change their diet very much (such as by excluding meat) might also not change other variables that affect weight, such as how much exercise they get.
b. It would be better to randomly assign some of the subjects to eat meat and some of the subjects to consume a vegetarian diet.
- 1.45 No. This was an observational study, because researchers could not have deliberately exposed people to weed killers. There was no random assignment, and no one would randomly assign a person to be exposed to pesticides. From an observational study, you cannot conclude causation. This is why the report was careful to use the phrase *associated with* rather than the word *caused*.
- 1.46 a. The survival rate for TAC (473/539, or 87.8%) was higher than the survival rate for FAC (426/521, or 81.8%).
b. Controlled experiment: Yes, we can conclude cause and effect, because this was a controlled experiment with random assignment. The random assignment balances out other variables, so the only difference is the treatment, which must be causing the effect.
- 1.47 Ask whether the patients were randomly assigned the full or the half dose. Without randomization there could be bias, and we cannot infer causation. With randomization we can infer causation.
- 1.48 Ask whether there was random assignment to groups. Without random assignment there could be bias, and we cannot infer causation.
- 1.49 This was an observational study: vitamin C and breast milk. We cannot conclude cause and effect from observational studies.
- 1.50 This is likely to be from observational studies. It would not be ethical to assign people to overeat. We cannot conclude causation from observational studies because of the possibility of confounding variables.
- 1.51 a. LD: $\frac{4}{4+46} = \frac{2}{25} = 8\%$ tumors; LL: $\frac{14}{14+36} = \frac{7}{25} = 28\%$ tumors
b. A controlled experiment; you can tell by the random assignment.
c. Yes, we can conclude cause and effect because it was a controlled experiment, and random assignment will balance out potential confounding variables.
- 1.52 a. $\frac{43}{43+10} = \frac{43}{53}$, or 81.1%, of the males who were assigned to Scared Straight were rearrested.
 $\frac{37}{37+18} = \frac{37}{55}$, or 67.3%, of those receiving no treatment were rearrested. So the group from Scared Straight had a higher arrest rate.
b. No, Scared Straight does not cause a lower arrest rate, because the arrest rate was higher.

Chapter Review Exercises

- 1.53 a. Dating: 81/440, or 18.4%
b. Cohabiting: 103/429, or 24.0%
c. Married: 147/424, or 34.7%
d. No, this was an observational study. Confounding variables may vary. Perhaps married people are likely to be older, and older people are more likely to be obese.
- 1.54 No, this was an observational study. There is no mention of random assignment. We cannot conclude causation from observational studies because of the possibility of confounding factors.

1.55 a.

	Boy	Girl	Total
Violent	10	11	21
Nonviolent	19	4	23
Total	29	15	44

- b. For the boys, 10/29, or 34.5%, were on probation for violent crime. For the girls, 11/15, or 73.3%, were on probation for violent crime.
- c. The girls were more likely to be on probation for violent crime.

1.56 For those getting the antivenom, 87.5% got better. For those given the placebo, only 14.3% got better.

	Antivenom	Placebo	Total
Better	7	1	8
Not Better	1	6	7
Total	8	7	15

- 1.57 Answers will vary. *Students should not copy the words they see in these answers.* Randomly divide the group in half, using a coin flip for each woman: Heads she gets the vitamin D, and tails she gets the placebo (or vice versa). Make sure that neither the women themselves nor any of the people who come in contact with them know whether they got the treatment or the placebo (“double-blind”). Over a given length of time (such as three years), note which women had broken bones and which did not. Compare the percentage of women with broken bones in the vitamin D group with the percentage of women with broken bones in the placebo group.
- 1.58 Answers will vary. *Students should not copy the words they see here.* Randomly divide the group in half, using a coin flip for each person: Heads they get Coumadin, and tails they get aspirin (or vice versa). Make sure that neither the subjects nor any of the people who come in contact with them know which treatment they received (“double-blind”). Over a given length of time (such as three years), note which people had second strokes and which did not. Compare the percentage of people with second strokes in the Coumadin group with the percentage of people with second strokes in the aspirin group. There is no need for a placebo, because we are comparing two treatments. However, it would be acceptable to have three groups, one of which received a placebo.
- 1.59 a. The treatment variable was Medicaid expansion or not and the response variables were the death rate and the rate of people who reported their health as excellent or very good.
- b. This was observational. Researchers did not assign people either to receive or not to receive Medicaid.
- c. No, this was an observational study. From an observational study, you cannot conclude causation. It is possible that other variables that differed between the states caused the change.
- 1.60 a. The treatment variable is whether the person has both forms of HIV infection (HIV-1 and HIV-2) or only one form (HIV-1). The response variable is the time to the development of AIDS.
- b. This was an observational study. No one would assign a person to a form of HIV.
- c. The median time to development of AIDS was longer for those with both infections.
- d. No, you cannot infer causation from an observational study.
- 1.61 No, we cannot conclude causation. There was no control group for comparison, and the sample size was very small.
- 1.62 No, it does not show that the exercise works. There is no control group. (Also, the sample size is very small.)

