

0123456789

Study Guide

THE CHICAGO GUIDE

TO WRITING ABOUT

NUMBERS

SECOND EDITION

JANE E. MILLER

THE UNIVERSITY OF CHICAGO PRESS

Chicago and London

Study Guide to accompany *The Chicago Guide to Writing about Numbers*,
second edition, copyright © 2004, 2015 by The University of Chicago.

Introduction v

- 1 Why Write about Numbers?** 1
Suggested Course Extensions 1
- 2 Seven Basic Principles** 2
Problem Set 2
Suggested Course Extensions 5
Solutions 7
- 3 Causality, Statistical Significance, and Substantive Significance** 8
Problem Set 8
Suggested Course Extensions 11
Solutions 13
- 4 Five More Technical Principles** 15
Problem Set 15
Suggested Course Extensions 21
Solutions 23
- 5 Basic Types of Quantitative Comparisons** 27
Problem Set 27
Suggested Course Extensions 31
Solutions 33
- 6 Creating Effective Tables** 35
Problem Set 35
Suggested Course Extensions 38
Solutions 40
- 7 Creating Effective Charts** 42
Problem Set 42
Suggested Course Extensions 46
Solutions 49
- 8 Choosing Effective Examples and Analogies** 53
Problem Set 53
Suggested Course Extensions 55
Solutions 57
- 9 Writing about Distributions and Associations** 58
Problem Set 58
Suggested Course Extensions 61
Solutions 63

CONTENTS

- 10 Writing about Data and Methods 65**
 - Problem Set 65
 - Suggested Course Extensions 68
 - Solutions 71

- 11 Writing Scientific Papers and Reports 73**
 - Problem Set 73
 - Suggested Course Extensions 74
 - Solutions 76

- 12 Speaking about Numbers 78**
 - Problem Set 78
 - Suggested Course Extensions 79
 - Solutions 81

- 13 Writing for Applied Audiences:
Issue Briefs, Chartbooks, Posters, and General-Interest Articles 88**
 - Problem Set 88
 - Suggested Course Extensions 90
 - Solutions 93

INTRODUCTION

This study guide was designed to provide practice applying the principles and tools introduced in *The Chicago Guide to Writing about Numbers*, 2nd Edition, with a problem set and a series of suggested course extensions for each chapter. A series of podcasts and other supplemental learning materials are available on the website, where you can also find a podcast that provides a brief overview of the online materials (PC0.1).

The problem sets reinforce the concepts and skills from each chapter. Some require simple calculations, others involve creating or critiquing tables, charts, or sentences. They can be used as homework assignments for a research methods, statistics, or writing course in which the book is being used, or by readers working independently. Solutions for the odd-numbered problems can be downloaded separately. See also podcasts PC0.2, PC0.3, and PC0.4 and appendix 0.1 on teaching how to write about numbers.

The suggested extensions apply the skills and concepts from *The Chicago Guide to Writing about Numbers*, 2nd Edition, to the actual writing process. They involve reviewing existing work, applying statistics, writing, and revising—using either your own work in progress or published materials (books, articles, reports, or web pages) in your field or that of your intended audience. Many of the suggested exercises for writing or revision entail peer-editing and are most effective if done with one or more others, whether as part of a course in which class time is devoted to these exercises or working with a peer. These exercises often involve writing or revising work to meet the instructions for authors for a leading journal in your field. Identify one or two such journals before you begin those exercises, allowing you to generate a coherent finished product for submission to that journal.

The “applying statistics” questions require access to a computerized database that includes several nominal, ordinal, and interval or ratio variables (see chapter 4 of *The Chicago Guide to Writing about Numbers*, 2nd Edition) for at least several hundred cases. Ideally these variables should be related to a research question involving application of statistical analysis that you can use for the exercises throughout the study guide, yielding a comprehensive analysis for a complete research paper. These exercises also require access to the accompanying documentation describing the study design, data collection, coding, use of sampling weights, and related methodological issues for the data set from which your variables are taken. If you do not have a data set and documentation that fits these criteria, you can download data sets from the supplemental online materials provided on the website that accompanies this study guide. Alternatively, you can often find suitable data sets on CD-ROMs that accompany research methods or statistics textbooks, or you can download data from sites such as the Inter-University Consortium for Political and Social Research (ICPSR).

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. Find a journal article from your field about an application of a quantitative analysis. Identify the audience for that journal in terms of
 - a. their discipline(s).
 - b. their expected level of familiarity with the kinds of quantitative methods used in the article.
 - c. their expected use of the results (e.g., research, policy, education).
2. In that article:
 - a. Circle one numeric fact or comparison each in the introduction, results section, and concluding section. For each:
 - i. Identify its purpose. Does the author explicitly or implicitly convey the purpose, or is it left unclear?
 - ii. Evaluate the ease of understanding the fact or comparison. Does the author convey its meaning and interpretation?
 - b. Are there other places in the article where a number or numeric comparison would be helpful? Identify the purpose of the number for each such situation.
 - c. What tools (tables, charts, prose) are used to present numbers? Do they suit the objective and audience for the article?
3. Find an article in the popular press that refers to an application of a quantitative analysis. (The science and health sections of newspapers, magazines, and websites are good resources.)
 - a. Who is the intended audience (e.g., what is their expected reading level and amount of statistical training)?
 - b. What is the objective of the article?
 - c. Is the article written with appropriate vocabulary and examples for that audience?
 - d. What tools are used to present numbers in the article? Do they suit the objective and audience?



WHY WRITE ABOUT NUMBERS?

2

SEVEN BASIC PRINCIPLES

PROBLEM SET

1. Use complete sentences to describe the relative sizes of the cities shown in table 2A.

Table 2A. Population of the five largest cities worldwide, 2012

City	Population (millions)
Tokyo, Japan	37.1
Jakarta, Indonesia	26.1
Seoul, South Korea	22.5
Delhi, India	22.2
Shanghai, China	20.9

Source: Worldatlas, 2014. "Largest Cities of the World." Available online at <http://www.worldatlas.com/citypops.htm>

2. What is missing from each of the following descriptions of table 2B? Note: Each statement is correct for some portion of the table (row or column).

Table 2B. Final medal standings of the top four countries, 2014 Olympic winter games

Country	Gold	Silver	Bronze	Total
Russia	13	11	9	33
United States	9	7	12	28
Norway	11	5	10	26
Canada	9	10	5	24

- a. "Russia did the best at the 2014 Winter Olympic Games, with 33 medals, compared to 28 for the United States, 26 for Norway, and 24 for Canada."
 - b. "The most medals were bronze (12), followed by gold (9), then silver (7)."
 - c. "The United States won more medals than all other countries, followed by Norway, Russia, and Canada."
3. For each of the following situations, specify whether you would use prose, a table, or a chart.
 - a. Statistics on five types of air pollutants in the 10 largest US cities for a government report.
 - b. Trends in the value of three stock market indices over a one-year period for a web page.
 - c. Notification to other employees in your corporation of a change in shipping fees.
 - d. Distribution of voter preferences for grade-level composition of a new middle school (grades 5–8, grades 6–8, or grades 6–9) for a presentation at a local school board meeting.

- e. National estimates of the number of uninsured among part-time and full-time workers for an introductory section of an article analyzing effects of employment on insurance coverage in New York City.
4. For each of the situations in the previous question, state whether you would use and define technical terms or avoid jargon.
5. Identify terms in the following sentences that need to be defined or restated for a nontechnical audience.
- “The Williams family’s income of \$25,000 falls below 185% of the Federal Poverty Threshold for a family of four, qualifying them for food stamps.”
 - “A population that is increasing at 2% per year has a doubling time of 35 years.”
6. Rewrite the sentences in the previous question for an audience with a fifth-grade education. Convey the main point, not the calculation or the jargon.
7. Read the sentences below. What additional information would someone need in order to answer the associated question?
- “Brand X costs twice as much as Brand Q. Can I afford Brand X?”
 - “My uncle is 6'6" tall? Will he fit in my new car?”
 - “New Diet Limelite has 25% fewer calories than Diet Fizzjuice. How much faster will I lose weight on Diet Limelite?”
 - “It has been above 25 degrees every day. We’re really having a warm month, aren’t we?”
8. Rewrite each of these sentences to specify the direction and magnitude of the association.
- “In the United States, race is correlated with income.” See table 2C.

Table 2c. Median income by race and Hispanic origin of householder, United States, 2011

Race/Hispanic origin	Median income
White	\$52,214
Black	\$32,229
Asian/Pacific Islander	\$65,129
Hispanic (can be of any race)	\$38,624

Source: US Bureau of the Census, 2012. Income, Poverty and Health Insurance Coverage in the United States: 2011, Table A. Median Household Income. Available online at http://www.census.gov/newsroom/releases/archives/income_wealth/cb12-172.html#tablea

- “There is an association between average speed and distance traveled.” (Pick two speeds to compare.)
- Write a hypothesis about the relationship between amount of exercise and weight gain.

9. Use the “generalization, example, exception” (GEE) approach on pp. 33–36 of *Writing about Numbers*, 2nd Edition, to describe the patterns in figure 2A. Include an introductory sentence about the purpose of the chart before summarizing the patterns.

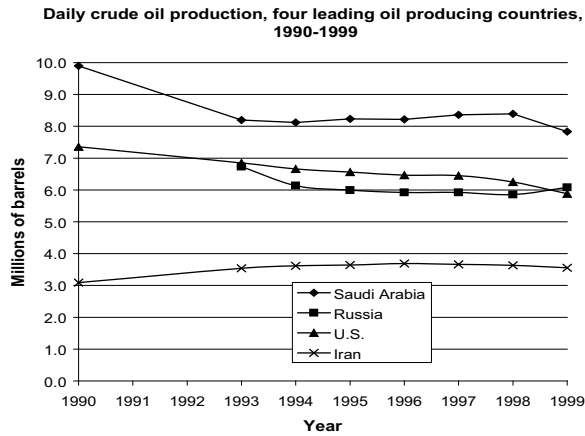


Figure 2A

SUGGESTED COURSE EXTENSIONS

A. Reviewing

Find a journal article in your field about a quantitative analysis. Use it to answer the following questions.

1. Is the context (Ws) of the study specified? If not, which Ws are missing or poorly defined?
2. Identify the intended audience for that journal and their expected level of familiarity with the statistical methods in the article. Evaluate the technical language.
 - a. Are definitions provided for all technical and statistical terms unfamiliar to the audience?
 - b. Are all acronyms used in the paper spelled out and defined?
 - c. Are methods or concepts named using terms familiar to that audience?
3. List the major tools (prose, tables, charts) used to present numbers in the article.
 - a. For one example of each type of tool, identify its intended purpose in that context (e.g., presenting detailed numeric values, conveying a general pattern).
 - b. Use the criteria in chapter 2 of *The Chicago Guide to Writing about Numbers*, 2nd Edition, to evaluate whether it is an appropriate choice of tool for that task.
 - c. If so, explain why. If not, suggest a more effective tool for that context.
4. Find a numeric fact or comparison in the introduction or conclusion to the article.
 - a. Is it clear what question those numbers are intended to answer?
 - b. Are the raw data reported in the text, a table, or a chart?
 - c. Are the values interpreted in the text?
 - d. Revise the paragraph to address any shortcomings you identified in parts a through c.
5. Find a description of an association between two variables. Are the direction and magnitude of the association specified? If not, rewrite the description.
6. Find a description of a pattern involving more than three values, subgroups, or sets of model results shown in a table or chart.
 - a. Is the purpose of the chart or table explained?
 - b. Is the pattern generalized or is it described piecemeal?
 - c. Are representative values reported to illustrate the pattern?
 - d. Are exceptions to the general pattern identified?

2

SEVEN BASIC PRINCIPLES

- e. Rewrite the description of the table or chart using the GEE approach to address any shortcomings you identified in parts a through d.

B. Writing papers

1. Describe a bivariate association among variables in your data, including the Ws, units, direction, and magnitude.
2. Graph the distribution of a continuous variable; describe it using an analogy.
3. Use the GEE approach to describe a three-way association among variables in your data.

C. Revising papers

1. Repeat questions A.1 through A.6 for a paper you have previously written.
2. For the introduction and discussion sections of a paper you have previously written:
 - a. Evaluate your use of technical language, following the guidelines in chapter 2 of *The Chicago Guide to Writing about Numbers*, 2nd Edition, and keeping in mind the intended audience for your paper.
 - b. Revise the wording to fix any problems you identified in part b.
3. Have someone who is unfamiliar with your research question peer-edit the answers to questions C.1 and C.2, using the checklist from chapter 2. Revise your drafts according to the feedback you received.

SOLUTIONS

1. Use complete sentences to describe the relative sizes of the cities shown in table 2A.

“In 2012, the world’s largest city, Tokyo, had a metropolitan-area population of 37 million people. With roughly 26 million residents, Jakarta, Indonesia—the next largest city—was only about 70% as large as Tokyo. Seoul and Delhi followed in third and fourth places, with about 22 million persons apiece, while fifth-place Shanghai weighed in at nearly 21 million.”

3. For each of the following situations, specify whether you would use prose, a table, or a chart.
 - a. Table to show detailed figures and organize the 50 numbers.
 - b. Chart to illustrate approximate pattern.
 - c. Prose (memo).
 - d. Pie chart.
 - e. Prose (few sentences).
5. Identify terms in the following sentences that need to be defined or restated for a nontechnical audience. Answer shown in bold.
 - a. “The Williams family’s income of \$25,000 falls below **185% of the Federal Poverty Threshold** for a family of four, qualifying them for food stamps.”
 - b. “A population that is increasing at 2% per year has a **doubling time** of 35 years.”
7. Read the sentences below. What additional information would someone need to be able to answer the associated question?
 - a. How much does Brand Q (or Brand X) cost? How much money do you have?
 - b. How big is the door opening to your car? The headroom and legroom?
 - c. How many calories does Diet Fizzjuice (or Diet Limelite) have?
 - d. Where are you located? What month is it? Is temperature being measured in degrees Fahrenheit or degrees Celsius?
9. “Figure 2A shows trends in daily crude oil production in the world’s four leading oil-producing countries during the 1990s. Over the course of that decade, Saudi Arabia consistently had the highest crude oil production, followed by Russia, the United States, and Iran. However, downward trends in production in the top three oil producing countries, coupled with steady production in Iran, led to a narrowing of the gap between those countries between 1990 and 1999. In 1990, Saudi Arabia produced 30% more oil than the United States and more than three times as much as Iran (10 million, 7 million, and 3 million barrels per day, respectively). By 1999, Saudi Arabia’s advantage had decreased to 25% more than the United States or Russia, and about twice as much as Iran.”

2

SEVEN BASIC PRINCIPLES

3

CAUSALITY, STATISTICAL SIGNIFICANCE, AND SUBSTANTIVE SIGNIFICANCE

PROBLEM SET

1. Evaluate whether each of these statements correctly conveys statistical significance. If not, rewrite the sentence so that the verbal description matches the numbers; leave the numeric values unchanged.
 - a. There was a statistically significant increase in average salaries over the past three years ($p = .04$).
 - b. The p -value for the t -test for difference in mean ozone levels equals 0.95, so we can be 95% certain that the observed difference is not due to chance.
 - c. The difference in voter participation between men and women was not statistically significant ($p = 0.35$).
 - d. The p -value for the t -test for difference in mean ozone levels equals 0.95. This test shows we can be 95% certain that the difference in ozone levels can be explained by random chance, hence the difference is not statistically significant.
 - e. The price of gas increased by \$0.05 over the past three months, meaning that the p -value = 0.05.
 - f. The p -value comparing trends in gas prices = 0.05, hence the price of gas increased by \$0.05.
 - g. Voter participation was 20% higher among Democrats than among Republicans in the recent local election. Statistical tests show $p < .01$, so we can be 99% certain that the observed difference is not due to chance.
 - h. The average processor speed was slightly higher for Brand A than for Brand B; however, $p = .09$, so the effect was not statistically significant. If the sample size were increased from 40 to 400, the difference in processor speeds between the two brands would increase, so it would become statistically significant.
 - i. The average processor speed was slightly higher for Brand A than for Brand B; however, $p = .09$, so the effect was not statistically significant. If the sample size were increased from 40 to 400, the standard error would decrease, so the difference might become statistically significant.
2. For each of the following findings, identify background facts that could help decide whether the effect is big enough to matter. Look up your suggested facts for one of the research questions. What do you conclude about the substantive significance of the finding?
 - a. Jo's IQ score increased 2 points in one year.
 - b. The average response on a political opinion poll for two adjacent counties differed by 2 points. The question was scaled "agree strongly," "agree," "neither agree nor disagree," "disagree," and "disagree strongly."

- c. The Dow Jones Industrial Index dropped 2 points since this morning.
- d. Bed rest is expected to prolong Mrs. Peterson's pregnancy from 34 to 36 weeks gestation.
3. Discuss whether each of the following research questions involves a causal relationship. If the relationship is causal, describe one or more plausible mechanisms by which one variable could cause the other. If the relationship is not causal, give alternative explanations or mechanisms for the association.
- April showers bring May flowers.
 - People with blue eyes are more likely to have blond hair.
 - Pollen allergies increase rapidly with longer daylight hours.
 - Eating spicy foods is associated with less risk of heartburn.
 - Prices and sales volumes are inversely related, so high sales volumes cause prices to drop.
 - Fair-skinned people sunburn faster than those with dark skin do.
 - Average reading ability increases dramatically with height between 4' and 5'.
4. For each of the studies summarized in table 3A
- explain how you would describe the findings in the results section of a scientific paper;
 - identify the criteria you used to decide how to discuss the findings for that study.

Table 3A. Hypothetical study results

Topic I: Effect of new math curriculum on test scores*	Effect size	Statistical	Sample size
		significance (<i>p</i> -value)	
Study 1	+1/2 point	$p < .01$	1 million
Study 2	+1/2 point	$p = .45$	1 million
Study 3	+5 points	$p < .01$	1 million
Study 4	+5 points	$p = .07$	1 hundred
Study 5	+5 points	$p = .45$	1 million
Topic II: Effect of white hair on mortality**			
Study A	+ 5%	$p < .01$	1 million
Study B	+ 5%	$p = .45$	1 million
Study C	+50%	$p < .01$	1 million
Study D	+50%	$p = .07$	1 hundred
Study E	+50%	$p = .45$	1 million

* Effect size for math curriculum studies = scores under new curriculum – scores under old curriculum

** Effect size for hair color studies = death rate for white-haired people – death rate for people with other hair colors

5. For each of the topics in table 3A, indicate whether you would recommend a policy or intervention based on the results, and explain the logic behind your decision.
6. For each of the topics in table 3A, write one or two sentences to describe the conclusions of each study to a lay audience.

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. In your statistics textbook or a journal article in your field, find an example of a highly correlated association.
 - a. Is that association causal? Why or why not?
 - b. List facts or comparisons that could be used to evaluate the substantive meaning of the association.
 - i. What facts do the authors report and interpret in the article?
 - ii. What are some other facts or comparisons that could be used to improve the explanation in the article?
2. In your statistics textbook or a journal article in your field, find an example of an association with a low correlation or nonstatistically significant association.
 - a. Is that association causal? Why or why not?
 - b. List facts or comparisons that could be used to evaluate whether the association is substantively meaningful.
 - i. What facts do the authors report and interpret in the article?
 - ii. What are some other facts or comparisons that could be used to improve the explanation in the article?
3. In a newspaper or magazine, find a scientific or policy problem and solution that are currently being touted for implementation.
 - a. Evaluate how the article addresses each of these aspects of “importance.” Does the article
 - i. specify a cause-and-effect type of relationship?
 - ii. provide a plausible argument for a causal association?
 - iii. discuss bias, confounding, or reverse causation?
 - iv. report results of statistical tests for that association?
 - v. assess whether the expected benefits of the proposed solution are big enough to outweigh costs or otherwise matter in a larger social context?
 - b. Given your answers to part a, write a short description of the appropriateness of the proposed solution.

B. Writing and Revising

1. Identify an aspect of your main research question that involves the association between two variables. Do you hypothesize that that association is causal?
 - a. If so, describe the mechanisms through which the hypothesized causal variable affects the hypothesized outcome variable.
 - b. If not, explain how those variables could be correlated. Identify possible bias, confounding factors, or reverse causation.

3

CAUSALITY, STATISTICAL SIGNIFICANCE, AND SUBSTANTIVE SIGNIFICANCE

- c. Rewrite your research question as a hypothesis, making it clear whether the association you are studying is believed to be causal.
 - d. What background facts could you find to help assess the substantive meaning of the association? Look them up and make the assessment.
 - e. Write a description of the substantive importance of the association for a discussion section of a scientific paper.
 - f. Write a statement for a lay audience, explaining the nature of the association between the variables.
2. For one or two key statistical results pertaining to the main research question in your paper, identify ways to quantify the broad social or scientific impact of that finding.
 - a. Locate statistics on the prevalence of your issue.
 - b. Find information on the consequences of the issue. E.g., what will it cost in terms of money, time, and other resources? What are its benefits? What does it translate into in terms of reduced side effects, improved skills, or other dimensions suited to your topic?
 - c. Use the information from parts a and b in conjunction with measures of effect size and statistical significance from your analysis to make a compelling case for or against the importance of the topic.
 3. Repeat question B.2 for a paper you have previously written about an application of a quantitative analysis.

SOLUTIONS

1. Evaluate whether each of these statements correctly conveys statistical significance. If not, rewrite the sentence so that the verbal description matches the numbers; leave the numeric values unchanged.
 - a. Correct.
 - b. Incorrect. A p -value of 0.95 corresponds to only a 5% probability that the observed difference is *not* due to chance (e.g., a 95% probability that the observed difference *is* due to chance.) “The p -value for the t -test for difference in mean ozone levels equals 0.95, so we can be 95% certain that the observed difference is due to chance.”
 - c. Correct.
 - d. Correct.
 - e. Incorrect. This sentence doesn’t reveal anything about statistical significance of that change. The most we can say from the information given is “The price of gas increased by \$0.05 over the past three months.”
 - f. Incorrect. Test-statistics and p -values are indicators of statistical significance. They do not measure the size of the association—in this case, the difference between two values, which cannot be calculated from the information given. The most we can say is “The p -value comparing trends in gas prices = 0.05.”
 - g. Correct.
 - h. Incorrect. Sample size does not affect size of a difference between values—in this case, difference in average processor speeds. See part i of this question for correct wording.
 - i. Correct.
3. Discuss whether each of the following research questions involves a causal relationship.
 - a. Causal (partly). The flowers would likely bloom in May whether or not it rains in April, but they will bloom more nicely if it rains.
 - b. Non-causal association. In many populations, blue eyes and blond hair co-occur, but neither causes the other.
 - c. Spurious. Positive correlations between both pollen allergies and daylight with more flowers blooming causes a spurious association between allergies and daylight. In other words, if you could have more daylight without more blooming plants, there wouldn’t be an association of daylight hours with pollen allergies.
 - d. Could be causal or reverse causal. For example, people with heartburn might stop eating spicy foods if they think those foods irritate their heartburn.
 - e. Reverse causal. Low prices probably induced greater sales. Could be causal in the long run if greater sales allow economies of scale in production, which in turn could lower prices.

3

CAUSALITY, STATISTICAL SIGNIFICANCE, AND SUBSTANTIVE SIGNIFICANCE

- f. Causal. Lack of protective pigment in fair-skinned people allows them to sunburn faster.
 - g. Spurious. Both reading ability and height increase dramatically with age, which is the real causal factor for both. Comparing kids with the same age but different heights would likely show much less difference in reading abilities than if age isn't taken into account.
5. For both topics in table 3A, the findings of studies 1 and 3 are statistically significant, studies 2 and 5 are not, and study 4 is borderline because the p -value is slightly above 0.05 and the sample size is small. However, the white hair/mortality association in topic II is spurious, so substantive and statistical significance are irrelevant. For topic I (curriculum change and test scores), where there is a plausible causal explanation, only the findings of study 3 are likely to be of substantive interest because the effect size in study 1 is so small.

PROBLEM SET

- For each of the following topics, indicate whether the variable or variables used to measure it are continuous or categorical, and single or multiple response.
 - Respondent's current marital status.
 - Respondent's number of siblings.
 - Siblings' heights.
 - Current marital status of siblings.
 - Temperature at 9 A.M. today.
 - Form of today's precipitation.
- A new school is being considered in your hometown. Several possible grade configurations are being considered (Plan A: grades K–3, 4–5, 6–8, 9–12; Plan B: grades K, 1–4, 5–7, 8–12). The current configuration is K–5, 6–9, and 10–12. Design a question to collect information from school principals on the age distribution of students, making sure the data collection format provides the detail and flexibility needed to compare the different scenarios for the district now and in five years.
- Your stopwatch is accurate to the nearest tenth of a second. In nine trials, the average time for a mouse named Squeeky to solve a maze was 10.44444444 seconds. Write a sentence to report that average.
- In a microbiology lab exercise, the size of viral cells being compared ranged from 0.00000018 meters (m) in diameter for Parovirus to 0.000001 m in length for Filoviridae (American Society for Microbiology, 1999). What scale would you use to report those data in a table? In the text?
- Write one or two sentences to compare the four specimens in table 4A. Which specimen is the heaviest? The lightest? By how much do they differ? What steps do you need to take before you can make the comparison?

Table 4A. Mass of four specimens

Specimen	Mass
1	1.2 lbs.
2	500 grams
3	0.7 kilograms
4	12 ounces

4

FIVE MORE TECHNICAL PRINCIPLES

6. For each of the figures 4.3a through 4.3e (*Writing about Numbers*, 2nd Edition, 80–81), choose
- a typical value;
 - an atypical value;
 - a plausible contrast (two values to compare).
- Explain your choices, with reference to range, central tendency, variation, and skewness.
7. Identify pertinent standards or cutoffs and other information needed to answer each of the following questions.
- Does Mr. Jones deserve a speeding ticket?
 - Is the new alloy strong enough to be used for the library renovations?
 - How tall is five-year-old Susie expected to be next year?
 - Can Leah go on the Ferris wheel at the amusement park?
 - Is this year's projected tuition increase at Public U unexpected?
 - Should we issue an ozone warning today?
8. Indicate whether each of the following sentences correctly reflects table 4B. If not, rewrite the sentence so that it is correct. Check both correctness and completeness of the data.
- Between 1964 and 1996, there was a steady decline in voter participation, from 95.8% in 1964 to 63.4% in 1996.
 - Voter turnout was better in 1996 (63.4%) than in 1964 (61.9%).
 - Almost all registered voters participated in the 1964 US presidential election.
 - The best year for voter turnout was 1992, with 104,600 people voting.
 - Less than half of the voting-age population voted in the 1996 presidential election.
 - A higher percentage of the voting-age population was registered to vote in 1996 than in 1964.

Table 4B. Voter turnout, United States presidential elections, 1964 through 1996

Year	Total Vote (1000s)	Registered		Voting Age	
		Voters (RV) (1000s)	Vote/RV (%)	Pop. (VAP) (1000s)	Vote/VAP (%)
1964	70,645	73,716	95.8	114,090	61.9
1968	73,212	81,658	89.7	120,328	60.8
1972	77,719	97,329	79.9	140,776	55.2
1976	81,556	105,038	77.6	152,309	53.5
1980	86,515	113,044	76.5	164,597	52.6
1984	92,653	124,151	74.6	174,466	53.1
1988	91,595	126,380	72.5	182,778	50.1
1992	104,600	133,821	78.2	189,529	55.2
1996	92,713	146,212	63.4	196,511	47.2

Source: Institute for Democracy and Electoral Assistance, 1999.

9. A billboard reads: “1 in 250 Americans is HIV positive. 1 in 500 of them knows it.”
- According to the two statements above, what share of Americans are HIV positive and know it? Does that seem realistic?
 - Rewrite the second statement to clarify the intended meaning
 - as a fraction of HIV-positive Americans.
 - as a fraction of all Americans.
10. An advertisement for a health education program included figure 4A to show the prevalence of two common health behavior problems among teenage girls. What is wrong with the graph?

Prevalence of smoking and teen pregnancy (%)

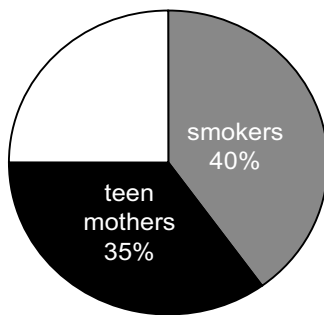


Figure 4A

11. You are involved in a research team that is conducting a study of commuting. One of the team members submits the following question to be included on the questionnaire:
- “How do you commute to work?
- Car _____
- Public transportation _____
- Train _____
- Carpool _____
- Walk _____”
- Critique the wording of the question using the guidelines in chapter 4 of *Writing about Numbers*, 2nd Edition.
 - Revise the question to correct the problems you identified in part a.
12. For each of figures 4.3a through 4.3e (*Writing about Numbers*, 2nd Edition, 80–81), come up with an example variable (topic) and context for which that distribution might be plausible.
13. Each of these statements contains an error. Identify the problem and rewrite the statement to correct the error. If additional information would be needed to make the correction, indicate what kind of information is needed.
- The proportionate increase in income during the 1990s was 20%.
 - Male infants outnumbered females (sex ratio at birth = 0.95).

- c. A majority of respondents (0.67) agreed that there should be a waiting period before buying a gun.
- d. Cancer accounted for two out of every ten deaths, equivalent to a death rate of 20%.

14. What is wrong with the following fictitious set of instructions for authors from a scientific journal that frequently publishes results of inferential statistical tests?

“In the interest of saving space, round all numeric results to the nearest single decimal place.”

15. Fill in table 4C to indicate whether the following values make sense for the concepts and units mentioned. If not, explain why not and suggest a plausible range of values for that variable.

Table 4c

Concept	Units	Value	Plausible (Y/N?)	If no, specify a plausible range of values
a. Height of an adult female in the US, 2014	Inches	65		
b. Height of an adult female in the US, 2014	Inches	120		
c. Height of a 6-year-old child in the US, 2014	Inches	65		
d. Height of a 6-year-old child in the US, 2014	Centimeters	120		
e. Observed daily low temperature, New York City, January	Degrees Fahrenheit	30		
f. Observed daily low temperature, New York City, January	Degrees Fahrenheit	-10		
g. Observed daily low temperature, New York City, January	Degrees Celsius	30		
h. Observed daily low temperature, New York City, January	Degrees Kelvin	-10		
i. Mean hourly wage, fast food employee, Los Angeles, CA, 2010	Dollars	8		
j. Mean annual wage, fast food employee, Los Angeles, CA, 2010	Dollars	8		

k. Share of population that is poor	Percentage	25		
l. Share of population that is poor	Percentage	-5		
m. Share of population that is poor	Proportion	25		
n. Share of population that is poor	Proportion	0.05		
o. Change in share of population that is poor	Percentage change	0.05		
p. Change in share of population that is poor	Percentage change	-5		

16. A weather scale based on daily observations is to be used to quantify weather in cities worldwide in all seasons. Dr. Blowhard suggests that the scale should be constructed for each observation (date, location) by adding together values of the following variables:

- observed high temperature, °F
- rainfall, inches
- snowfall, inches
- an indicator of whether the wind speed exceeded 20 MPH, coded 1 = yes, 2 = no
- a variable coded 1 = sunny all day, 2 = sunny part of the day, 3 = cloudy all day
 - a. Critique the scale construction method using the criteria discussed in chapter 4 of *Writing about Numbers*, 2nd Edition.
 - b. If you think a valid scale could be constructed using the listed variables, provide instructions on how to do so.
 - c. If you do not think a valid scale could be constructed using the listed variables, describe how you could use the variables (e.g., in a study of how weather affects immigration).

Table 4D. Coding of items about abortion attitudes, 2000 US General Social Survey

<i>"Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion . . ."</i>	Don't		
	Yes	No	know
If there is a strong chance of serious defect in the baby?	1	2	8
If she is married and does not want any more children?	1	2	8
If the woman's own health is seriously endangered by the pregnancy?	1	2	8
If she is not married and does not want to marry the man?	1	2	8
If she became pregnant as a result of rape?	1	2	8
If the woman wants it for any reason?	1	2	8

Source: National Opinion Research Center, 2000.

17. Answer the following questions about construction of an abortion attitudes scale using the six items shown in table 4D.
- a. How should “don’t know” answers for each of the items be treated when computing the scale?
 - b. If each respondent’s scores on the individual items are *summed* to compute the value of the scale for them,
 - i. what is the minimum possible valid value of the scale? The maximum?
 - ii. what would be the substantive interpretation of a scale score of 12 for an individual respondent?
 - c. If each respondent’s scores on the individual items are *averaged* to compute the scale, what is the minimum possible valid value of the scale? The maximum?
 - d. If the value of the scale for each respondent consists of a *tally* of the number of items with which that respondent agreed, what is the minimum possible valid value of the scale? The maximum?
 - e. If only valid (“yes” or “no”) answers to each abortion question were included in the calculation of the scale, what can you say about the set of respondents who would be included in the final analytic sample?

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. In a statistics textbook, find a discussion of an association between two or three variables. For each of those variables, identify
 - a. the type of variable (nominal, ordinal, interval, or ratio);
 - b. whether it is single- or multiple-response.
 - c. For continuous variables, identify
 - i. the system of measurement (e.g., British, metric, or other);
 - ii. the unit of analysis;
 - iii. the scale of measurement;
 - iv. the appropriate number of digits and decimal places for reporting the mean value in the text and a table.
 - d. For categorical variables, identify the categories for each variable, and units if pertinent.
 - e. If the items requested in c and d aren't described in your book, list plausible versions of that information. For example, if you are studying family income in the United States, you would expect the system of measurement to be United States dollars, the unit of analysis to be the family, and the scale of measurement to be either dollars or thousands of dollars.
2. Read the textbook's description of the variables you listed in question A.1. Does it provide the recommended information about the distribution of that type of variable? If not, what additional information is needed?
3. Read the literature in your field to determine whether standard cutoffs or standard patterns are used to assess one of the variables in the association you listed in question A.1. Find a reference source that explains its application and interpretation.
4. Repeat questions A.1–A.3 using variables described in a journal article in your field of study.
5. In a journal article that uses a scale or index variable, read the description in the methods section about how that scale or index was constructed.
 - a. Evaluate whether the items used to construct that scale
 - i. are a consistent level of measurement (type of variable);
 - ii. have similar levels and ranges of values to one another.
 - b. Consider whether the author(s) evaluated or described the distribution of values of the scale or index.
 - c. Note whether the author(s) included citation(s) about the items and method used to construct the scale.
 - d. If you found problems in parts a or b, make suggestions of additional information needed or a revised approach to analyzing the items.

4

FIVE MORE TECHNICAL PRINCIPLES

B. Applying Statistics

1. Repeat question A.1 for the key predictor and outcome variables for a research question that can be analyzed using variables available in your database.
2. Using the same data,
 - a. calculate the frequency distribution for each variable;
 - b. create a simple chart of the distribution;
 - c. select and calculate the appropriate measure of central tendency for that type of variable;
 - d. determine whether central tendency typifies the overall distribution. Why or why not? If not, what is a more typical value?
 - e. For continuous variables, identify the minimum and maximum values and the first and third quartiles of the distribution.
3. Complete the “Getting to Know Your Variables” exercise, available on the website of supplemental materials for *The Chicago Guide to Writing about Numbers*, 2nd Edition. Pay special attention to the following issues for your key predictor and outcome variables in your data set:
 - a. issues of missing by design
 - b. other missing values (e.g., item non-response)
 - c. definitional limits on the values of your variables
 - d. how substantive issues related to the concepts under study affect the plausible range of values
 - e. how the units of measurement affect plausible range of values
 - f. how the context of your study affects the plausible range of values
4. For a variable in your data set for which a standard or cutoff is used in the literature,
 - a. repeat question A.3.
 - b. use the standard or cutoff to classify what percentage of cases fall below the cutoff in your data.
 - c. evaluate whether your distribution follows the expected pattern for that variable based on information in the literature. If not, identify possible reasons for those differences (e.g., different context).
5. Repeat questions A.1–A.3 for an index or scale in your data set.
6. Compare the eligibility thresholds for your state’s State Children’s Health Insurance (S-CHIP) for families for the most recent year available against the Federal Poverty Thresholds (see websites for your state’s S-CHIP program and the “Poverty” page on the US Census website). What is the highest income that would qualify for free S-CHIP benefits for a family of one adult and one child? A family of one adult and two children? A family of two adults and two children?

SOLUTIONS

- For each of the following topics, indicate whether the variable or variables used to measure it are continuous or categorical, and single or multiple response.
 - Categorical, single-response.
 - Continuous, single-response.
 - Continuous, multiple-response.
 - Categorical, multiple-response.
 - Continuous, single-response.
 - Categorical, multiple-response.
- “Squeaky the mouse solved the maze in an average of 10.4 seconds over nine trials.”
- All measurements must be converted into consistent units (scale and system of measurement). I chose to convert all measurements to kilograms, using the conversion factor 2.2 pounds/kilogram (see revised table 4A). “Of the four specimens compared here, specimen 3 is the heaviest (0.70 kilograms [kg.]). It is about twice as heavy as the lightest (specimen 4, which weighed 0.34 kg.). The other two specimens were about 70% as heavy as specimen 3.”

Table 4A. Mass of four specimens

Specimen	Weight (original units)	Weight (kg.)
1	1.2 lbs.	0.54
2	500 grams	0.50
3	0.7 kilograms	0.70
4	12 ounces	0.34

- Identify pertinent standards or cutoffs for each of the following questions.
 - Speed limit where he was driving and his actual speed.
 - The weight-bearing capacity of the alloy (in weight per unit area) and the expected weight load (again, in weight per unit area) in the library.
 - Her current height and a growth chart (height for age) for girls.
 - Leah’s height and the minimum height requirement for the Ferris wheel.
 - The rate of inflation, current tuition, and rates of tuition increase at Public U over the past few years.
 - Today’s ozone measurement and the cutoff for an ozone warning.

4

FIVE MORE TECHNICAL PRINCIPLES

9. Regarding the questions about HIV-positive Americans:
- Taken together, the two statements imply that 1 in 125,000 Americans are HIV positive and know it, clearly a misstatement of the facts.
 - Rewrite the second statement to clarify the intended meaning.
 - “Half of HIV-positive Americans know they are infected.”
 - “1 in 500 Americans is HIV positive and knows it.”
11. Critique the commuting questionnaire question.
- First, the responses are not mutually exclusive. For example, “car” and “carpool” overlap, as do “public transportation” and “train.” Second, the responses aren’t exhaustive, excluding bus and bicycle, among other possibilities, and omitting an “other (specify)” response. Third, they don’t provide a way for people to record more than one mode of transportation. Fourth, there is no appropriate response for people who don’t work or for those who work at home. And finally, there are no instructions given about how many responses are allowed.
 - “How do you commute to work? (Mark all that apply.)”

Car	___
Train	___
Bus	___
Bicycle	___
Walk	___
Other (specify)	_____
I work at home	___
I do not work	___”
13. Identify the errors and rewrite.
- Proportion and percentage are not consistent units. Write “The proportionate increase in income during the 1990s was 0.20,” or “Income increased by 20% during the 1990s.”
 - The reported sex ratio indicates a lower number in the numerator than the denominator. Either write “Male infants outnumbered females (sex ratio at birth = 1.05 males per female)” (flipping over the ratio to be consistent with the wording, and reporting units as males per female), or “There were slightly fewer male than female infants (sex ratio at birth = 0.95 males per female)” (revising the wording to be consistent with the numeric value, and reporting units as males per female).
 - The value 0.67 does not indicate a majority unless labeled as a proportion. Better to express the value as a percentage. Write “A majority of respondents (67%) agreed that there should be a waiting period before buying a gun.”
 - A death rate is expressed relative to the population (e.g., number of living people), not as a percentage of deaths (e.g., relative to the total

number of deaths). Unless the total population and number of deaths are known, the first half of the sentence doesn't include enough information to calculate the death rate. Write "Cancer accounted for two out of every ten deaths."

15. Table 4C should be completed to indicate whether the following values make sense for the concepts and units mentioned, and if not, to indicate a plausible range of values for that variable.

Table 4c

Concept	Units	Value	Plausible (Y/N?)	If no, specify a plausible range of values
a. Height of an adult female in the US, 2014	Inches	65	Yes	
b. Height of an adult female in the US, 2014	Inches	120	No, 10' is much too tall	58" to 78"
c. Height of a 6-year-old child in the US, 2014	Inches	65	No, too high to be biologically plausible	40" to 47"
d. Height of a 6-year-old child in the US, 2014	Centimeters	120	Yes	
e. Observed daily low temperature, New York City, January	Degrees Fahrenheit	30	Yes	
f. Observed daily low temperature, New York City, January	Degrees Fahrenheit	-10	Yes	
g. Observed daily low temperature, New York City, January	Degrees Celsius	30	No, 30°C corresponds to 86°F, which is too high for NYC in January	
h. Observed daily low temperature, New York City, January	Degrees Kelvin	-10	No, 0 is the lowest possible value for temperature on the Kelvin scale	250 to 280 K
i. Mean hourly wage, fast food employee, Los Angeles, CA, 2010	Dollars	8	Yes	
j. Mean annual wage, fast food employee, Los Angeles, CA, 2010	Dollars	8	No, much too low for annual wages in the specified time, place, and units	\$2,000 (part-time summer) to \$15,000 (full-time year round)

(continued)

Table 4c (continued)

Concept	Units	Value	Plausible (Y/N?)	If no, specify a plausible range of values
k. Share of population that is poor	Percentage	25	Yes	
l. Share of population that is poor	Percentage	-5	No, a percentage share cannot be negative	0% to 100% are the upper and lower definitional limits for a percentage share
m. Share of population that is poor	Proportion	25	No, a proportion cannot exceed 1.0	0.0 to 1.0 are the upper and lower definitional limits for a proportion
n. Share of population that is poor	Proportion	0.05	Yes	
o. Change in share of population that is poor	Percentage change	0.05	Yes	
p. Change in share of population that is poor	Percentage change	-5	Yes	

17. Answers to questions about construction of an abortion attitude scale using the items shown in table 4D.

- a. The “don’t know” answers should be treated as missing values, meaning that the numeric code of 8 will not be treated numerically in the computation of the scale.
- b. If the scores on the individual items are *summed* to compute the scale,
 - i. the minimum valid value = 6 and the maximum = 12.
 - ii. a scale score of 12 for an individual respondent would mean that that person said “no” to all six items.
- c. If the scores on the individual items are *averaged* to compute the scale, the minimum valid value of the scale = 1 and the maximum = 2.
- d. If the scale consists of a *tally* of the number of items with which a respondent agreed, the minimum valid value of the scale = 0 (didn’t agree with any items), and the maximum = 6 (agreed with all items).
- e. If only valid (“yes” or “no”) answers to each abortion question were included in the calculation of the scale, any respondent who said “don’t know” to one or more of the items would be excluded from the final analytic sample because missing values on any item excludes them from the calculation.

PROBLEM SET

1. Identify the type of quantitative comparison used in each of the following statements.
 - a. "Yesterday, New York City received 5.5 inches of snow."
 - b. "Ian Thorpe's margin of victory in the 400-meter freestyle was 0.74 seconds."
 - c. "A panel of independent tasters preferred new Wheat Whistles 3 to 1 over their regular snack."
 - d. "The Dow Jones Industrial Average dropped 0.6% since this morning's opening."
 - e. "On sale, the scanner cost \$10 less than the suggested list price."
 - f. "Cornstarch has twice the thickening power of flour; for each teaspoon of flour called for in a recipe, substitute on half teaspoon of cornstarch."
 - g. "Median income for the metro region was \$31,750."
 - h. "At 6'3", Joe is two standard deviations taller than the average adult man."
 - i. "Sixty-eight percent of registered voters turned out for the primary election."
 - j. "State U was seeded first in the tournament."
2. In the 2000 presidential election, Al Gore received 50,996,116 votes while George W. Bush received 50,456,169 votes.
 - a. Write a sentence to describe the ranks of the two candidates.
 - b. Calculate the difference between the number of votes each candidate received. What impression does that information alone convey?
 - c. Calculate the percentage difference between the number of votes each candidate received. What impression does that information give?
3. Indicate whether each of the following statements is correct. If not, rewrite the second part of the sentence to agree with the first.
 - a. "Brand X lasts longer than Brand T, with an average lifetime 60% as long as Brand T's."
 - b. "Mean attendance at Root4 U increased 25% since last year, from 4,000 to 5,000 fans per game."
 - c. "The ratio of flour to butter in shortbread is 2:1; it uses twice as much butter as flour."
 - d. "At this time of year, reservoirs are usually 90% full. Currently, with reservoirs at 49% of capacity, water levels are only about 54% of normal."
 - e. "Nadia's test score was higher than 68% of students nationwide ($Z = 1.0$).

5

BASIC TYPES OF QUANTITATIVE COMPARISONS

- f. “A panel of 200 consumers rated ISP A four to one over ISP B. In other words, four more panelists preferred Company A as their Internet service provider.”
 - g. “Matt is in the 91st percentile for height. He is among the tallest 10% of boys his age.”
 - h. “Valueland is advertising 15% off everything in the store. That \$200 camera will cost only \$170.”
 - i. “The value of mutual fund ABCD tripled since last year, going from 100 to 33.”
4. In the 1999 Diallo case in New York City, 41 bullets hit the victim. Write down the criteria that you would intuitively use to interpret that number. Against what are you comparing the number of bullets?
5. Each of the following statements correctly describes part of table 5A, but each description is incomplete. Fill in the missing information.

Table 5A. Median income by race and Hispanic origin of householder, United States, 2011

Race/Hispanic origin	Median income
White	\$52,214
Black	\$32,229
Asian/Pacific Islander	\$65,129
Hispanic (can be of any race)	\$38,624

Source: US Bureau of the Census, 2012. Income, Poverty and Health Insurance Coverage in the United States: 2011, Table A. Median Household Income. Available online at http://www.census.gov/newsroom/releases/archives/income_wealth/cb12-172.html#tablea

- a. “In the US in 2011, median income for Asian households was about twice as much.”
 - b. “Median income for Hispanic households was \$6,395 more.”
 - c. “White households rank second.”
 - d. “The percentage difference for Asian households was 20%.”
6. Use table 5B to perform the tasks listed below.

Table 5B. Price per gallon for regular unleaded gasoline at selected gas stations, June 2012 and June 2013

Gas station	June 2012	June 2013
AAA	\$3.45	\$3.71
Bosco	\$3.37	\$3.75
Cargo	\$3.48	\$3.68
Dart	\$3.30	\$3.66
Essow	\$3.46	\$3.74

- a. Rank the stations from highest to lowest gas price for each of the two dates.

- b. Write a description of the distribution of prices in each year. Use difference and ratio in your description to convey the differences between the two distributions.
- c. Describe how you might use rank in conjunction with difference or ratio in deciding where to buy gas.
7. For each of the phrases listed below
- identify other phrases on the list that have the same meaning;
 - write the equivalent dollar value, assuming comparison against a price of \$200;
 - write the corresponding ratio. For statement i, for example, the ratio would be 0.25.
 - “25% of the original price”
 - “costs 25% less than . . .”
 - “costs 25% more than . . .”
 - “priced 25% off”
 - “125% of the original price”
 - “marked down 75%”
 - “75% of the original price”
 - “costs 75% as much as . . .”
8. The homicide rate in the US dropped from 5.6 homicides per 100,000 persons in 2002 to 4.2 per 100,000 in 2012. Calculate and write sentences to describe
- the differences between the homicide rates in the two periods;
 - the ratio of the homicide rates in the two periods;
 - the percentage change between the two periods using
 - the 2002 rate as the denominator;
 - the average of the two rates as the denominator.
9. In table 5C, fill in the z-score for height for each boy in the sample.

Table 5C. Heights of a sample of six-year-old boys

Name	Height (cm)	Z-score
David	117.51	
Jamal	113.90	
Ryan	124.81	
Luis	115.45	
JC	112.73	

Note: Standard population: mean = 115.12 cm; SD = 4.78 cm.

- Describe how Ryan’s, Luis’s, and JC’s heights compare to the national norms for boys their age based on their z-scores. (See table 5.3 in *Writing about Numbers*, 2nd Edition, 108, for ways to avoid using “z-scores” as you write).

- b. Two boys have heights that are about equidistant from the mean— one above and one below average. Who are they and about how far are their heights from those of average six-year-old boys?
- c. A new boy, Mike, joins the class. He is one standard deviation taller than the average boy his age. How tall is Mike?
10. One thousand people lived in Peopleland in 2010 and the population was growing at an annual rate (r) of 2.0% per year.

Table 5D. Population of Peopleland, 2010–2020

Year	Population	Increase from previous year	Cumulative increase since 2010	Percentage change since 2010
2010	1,000			
2011				
2012				
2013				
2014				
2015				
2016				
2017				
2018				
2019				
2020				

- a. Use the formula $P_t = P_0 \times e^{rt}$ to fill the population for each year into table 5D. The year 2010 is year 0, t is the number of years since 2010, r (the annual growth rate, expressed as a proportion) is 0.02, and e is the base of the natural logarithms.
- b. Calculate the increase in population from the preceding year for each year in the table. Write a sentence explaining the pattern of annual population increase across the 10-year period.
- c. The cumulative increase is the total number of people added to the population since 2010. How many more people live in Peopleland in 2020 than in 2010?
- d. Calculate the percentage change relative to 2010 for each year. Write a sentence to describe the percentage change in population between 2010 and 2020.
- e. What is the ratio of the population size for 2020 compared to 2010? How does that ratio relate to the percentage change over that 10-year period?
- f. How do the annual rate of growth and the percentage change between 2010 and 2020 relate?

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. Find a report about recent patterns in mortality, fertility (National Center for Health Statistics website), or unemployment (Bureau of Labor Statistics website).
 - a. Identify an example in that report of each of the following: rank, difference, ratio, and percentage difference or change.
 - b. For each example, identify the reference value. Does it come from within their data or some other source (e.g., a historic value or a reference population)?
 - c. Read the explanation of each of your chosen examples. Is each one clear? If not, use the criteria outlined in chapter 5 to improve the explanation.
 - d. Identify at least one instance where a different (or additional) comparison would be useful. Perform the calculations and describe the results, not the mathematical steps involved in the comparison.

B. Applying Statistics

1. Find a problem from a statistics or research methods textbook that involves comparison of two or more numbers.
 - a. Identify a pertinent comparison value for the topic and units under study.
 - b. Choose two ways to compare the numbers (e.g., rank, difference, ratio, or percentage difference).
 - c. Explain your choice of types of quantitative comparisons, with reference to the level of measurement for that variable and to a related research question or issue.
 - d. Calculate the pertinent comparisons.
 - e. Write a paragraph that interprets the results of the comparisons from part d.
 - f. Use the checklist at the end of chapter 5 in *Writing about Numbers*, 2nd Edition, to evaluate completeness and clarity of your writing.
2. Use a spreadsheet to complete the population growth question from this chapter's problem set.
 - a. Repeat the growth projections using the formula for annual compounding in place of the formula for continuous compounding (see "Annual Growth Rates" in *Writing about Numbers*, 2nd Edition, 111–12). Which formula predicts faster growth?
 - b. Repeat the population projection using a negative growth rate ($r = -0.02$). How much population is lost over the 10-year period?

5

BASIC TYPES OF QUANTITATIVE COMPARISONS

C. Writing and Revising

1. Identify a quantitative background fact for the introductory section of your paper to compare with information for other time periods or cases.
 - a. Select two pertinent types of quantitative comparisons for that fact.
 - b. Explain your choice of types of quantitative comparisons, with reference to the level of measurement for that variable and to the topic of your paper.
 - c. Look up the pertinent comparison data and calculate the comparisons.
 - d. Write a paragraph that interprets the results of the comparisons from part c.
 - e. Use the checklist at the end of chapter 5 to evaluate the completeness and clarity of your description.
2. Repeat question B.1 with data from the results section of your paper.

SOLUTIONS

1. Identify the type of quantitative comparison used in each of the following statements.
 - a. Value
 - b. Difference
 - c. Ratio
 - d. Percentage change
 - e. Difference
 - f. Ratio
 - g. Rank (median is the 50th percentile)
 - h. Z-score
 - i. Value (in this case, the units of measurement are percentage points)
 - j. Rank

3. Indicate whether each of the following statements is correct. If not, rewrite the second part of the sentence to agree with the first; changes shown in **bold**.
 - a. “Brand X lasts longer than Brand T, with an average lifetime 60% **longer than** Brand T’s.”
 - b. Correct as written.
 - c. “The ratio of flour to butter in shortbread is 2:1; it uses twice as much **flour as butter.**”
 - d. Correct as written.
 - e. “Nadia’s test score was higher than **84%** of students nationwide ($Z = 1.0$). (Sixty-six percent are within 1 standard deviation of the mean [e.g., ± 1 standard deviation], but you must also include those below $z = -1.0$ to answer this question correctly.)
 - f. “A panel of 200 consumers rated ISP A four to one over ISP B. In other words, four **times as many** panelists preferred Company A as their Internet service provider.”
 - g. Correct as written.
 - h. Correct as written.
 - i. “The value of mutual fund ABCD tripled since last year, going **from 33 to 100.**”

5. Additional information shown in **bold**.
 - a. “In the US in 2011, median income for Asian households was about twice **that for black households.**”
 - b. “Median income for Hispanic households was \$6,395 higher **than that for black households.**”
 - c. “White households ranked second **in terms of median income, below only Asians and Pacific Islanders.**”
 - d. “**Median income for** Asian households was 20% higher **than that for white households.**”

5

BASIC TYPES OF QUANTITATIVE COMPARISONS

7. With a comparison value of \$200:
- “25% of the original price” (i) and “marked down 75%” (vi) have the same meaning. Both correspond to a price of \$50, equivalent to a ratio of 0.25.
 - “costs 25% less than . . .” (ii), “priced 25% off” (iv), “75% of the original price” (vii), and “costs 75% as much as . . .” (viii) correspond to a price of \$150, equivalent to a ratio of 0.75.
 - “costs 25% more than . . .” (iii) and “125% of the original price” (v) correspond to a price of \$250 and a ratio of 1.25.
9. In table 5C, fill in the z-score for height for each boy in the sample.

Table 5c. Heights of a sample of six-year-old boys

Name	Height (cm)	Z-score
David	117.51	0.50
Jamal	113.90	-0.26
Ryan	124.81	2.03
Luis	115.45	0.07
JC	112.73	-0.50

Note: Standard population: mean = 115.12 cm; SD = 4.78 cm.

- Ryan is approximately two standard deviations above the average height for a six-year-old boy, while Luis is just about average and JC is about half a standard deviation below average for his age.
- David and JC are half a standard deviation taller and shorter than the average six-year-old boy, respectively.
- Mike stands 119.90 cm tall.

PROBLEM SET

1. Write a title for table 6A.

Table 6A.

Year	Median age (years)
1970	
1980	
1990	
2000	
2010	

Source: US Census of Population, National data, various dates.

2. Answer the following questions for tables 6.2 through 6.8 in *Writing about Numbers*, 2nd Edition.
 - a. Who is described by the data?
 - b. To what date or dates do the data pertain?
 - c. Where were the data collected?
 - d. What are the units of measurement? Are they the same for all cells in the table?
 - e. Where in the table are the units of measurement defined?
 - f. Does the table use footnotes? If so, why? If not, are any needed?
 - g. Are panels used within the table? If so, why? If not, would the addition of panels improve the clarity of the table?
3. Table 6B needs several footnotes to be complete. What information would those footnotes provide?

Table 6B. EPA estimates of carbon dioxide concentration and mean annual global temperature, 1960–2010

	CO ₂ concentration (ppmv.)	Mean temperature (°F)
1960		
1970		
1980		
1990		
2000		
2010		

6

CREATING EFFECTIVE TABLES

4. What is missing from table 6C?

Table 6c. Math and English test scores by homeroom teacher

Ms. Michaelson	65.7	69.0
Mr. Cifuentes	72.1	70.8
Mr. Williams	70.1	62.1
Ms. Andrews	76.2	78.2
Ms. Smith-Henderson	65.2	72.5

5. Design a table for each of the following topics. Provide complete labeling and notes, show column spanner and panels if pertinent, and indicate what principle(s) you would use to organize items within the rows and/or columns.
- Average commuting costs per month, by mode of transportation (bicycle, bus, car, train, walk, other), US, 2010. (One number per type of transportation.)
 - Age (years), gender, race, and educational attainment composition of a study sample, Houston, 2012.
 - Annual number of people receiving college degrees by gender, from 2000 to 2010.
 - Measures of association between height (cm), weight (kg), percentage body fat, systolic blood pressure (mm Hg), and resting pulse (beats per minute) from a sample of 150 elderly adults, Norway, 2008.
 - Estimates of dates using two different dating techniques for each of 15 artifacts taken from archaeological sites in three North African and two European countries.
 - Type of contraceptive (condom, diaphragm, implant/injectable, oral contraceptive [the Pill], surgical sterilization, other, none) by 10-year age groups of women aged 15 to 45, US, 2010. (Some women use more than one method.)

6. A journal for which you are writing an article allows no more than two tables, but your current draft has three. Combine tables 6D and 6E below into one table of 18 or fewer rows.

Table 6D. Number of wildfires by month, United States, 2010–2012

Month	2010	2011	2012	30-year average ^a
January				
February				
March				
April				
May				
June				
July				
August				
September				
October				
November				
December				
Total				

^a1980–2010.

Table 6E. Number of acres consumed by wildfire, by month, United States, 2010–2012

Month	2010	2011	2012	30-year average ^a
January				
February				
March				
April				
May				
June				
July				
August				
September				
October				
November				
December				
Total				

^a1980–2010.

6

CREATING EFFECTIVE TABLES

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. Find a simple table in a newspaper or magazine article. Evaluate whether tables in that article can stand alone without the text, using the guidelines in chapter 6 of *Writing about Numbers*, 2nd Edition. Suggest ways to improve labeling and layout.
2. In a professional journal in your field, find a table that includes column spanners or panels.
 - a. Evaluate whether you can interpret all the numbers in the table without reference to the text, using the guidelines in chapter 6. Suggest ways to improve labeling and layout.
 - b. Using information in the article, revise the table to correct those errors.
 - c. Consider whether a different layout would work more effectively.
 - d. Assess whether other tables are needed in the paper. Can two (or more) tables from the paper be combined into one?
 - e. Take one of the tables in the article and simplify it into one or more tables for a lay audience.
 - f. Pick a chart from the article. Draw a table to present the same information. Show what would go into the rows and columns, whether the table would have spanners or panels, labels, notes, etc.
3. In a journal article from your field, find a table that presents the relationship between a nominal predictor variable with more than two categories, and an outcome variable.
 - a. Identify the principle used to organize the categories of the nominal variable in the rows or columns of the table, referring to the criteria in chapters 6 and 7 of *Writing about Numbers*, 2nd Edition.
 - b. Critique whether that organization coordinates with the associated narrative.
 - c. Sketch a revised version of the table that addresses any shortcomings you identified in part b.

B. Applying Statistics

1. Create a table to display univariate frequency distributions for several variables in your data.
2. Create separate tables to show each of the following types of bivariate associations between variables in your data set, using the guidelines in chapter 6 of *Writing about Numbers*, 2nd Edition, about table structure for those types of statistics.
 - a. correlations between pairs of continuous variables
 - b. cross-tabulations between pairs of categorical variables

- c. differences in means for a continuous outcome variable according to values of categorical independent variables
3. Make a list of two or three simple tables to show two-way or three-way associations that pertain to your research question. Write individualized titles for each table.
4. Obtain a copy of the instructions for authors for a leading journal in your field. Revise the tables you created in questions B.1 through B.3 to satisfy their criteria.

C. Writing and Revising

Hint: Save the tables you create for use in the suggested course extensions for chapters 9 and 11.

1. Design a table to report results of a bivariate analysis involving a nominal predictor variable with more than two categories. Specify which organizing principle(s) you would use to display the categories of the predictor variable in the rows, referring to the criteria in chapters 6 and 7 of *Writing about Numbers*, 2nd Edition. Justify your choice, with reference to the specific objectives of your analysis.
2. Evaluate a table of bivariate statistics that you created previously for a paper, using the checklist in chapter 6, the criteria for organizing data in charts (chapter 7), and the instructions for authors for a leading journal in your field. Revise that table to incorporate any needed changes.
3. Exchange drafts of bivariate tables from questions C.1 and C.2 with a peer. Evaluate them using the checklist in chapter 6 and the instructions for authors for their selected journal. Revise according to the feedback you receive.
4. Read through the results section of your paper and identify topics or statistics for which to create additional tables related to some aspect of your research question. Draft them with pencil and paper, including table structure and complete title, labels, and footnotes.

6

CREATING EFFECTIVE TABLES

SOLUTIONS

1. Write a title for table 6A: “Median age of the US population, 1970 to 2010.”
3. Provide footnotes to table 6B.
 - Spell out EPA (Environmental Protection Agency).
 - Spell out ppmv (parts per million volume).
 - Spell out CO₂ (carbon dioxide).
 - Cite the data sources.
5. Design a table for each of the following topics.
 - a. Design elements:
 - Title: “Average monthly commuting costs (\$) by mode of transportation, US, 2010”
 - Mode of transportation in the rows
 - A column for monthly commuting costs
 - Items in the rows organized in ascending or descending order of cost
 - Note citing data source
 - b. Design elements:
 - Title: “Demographic composition of study sample, Houston, 2012”
 - One row for each demographic characteristic, with the categories for each demographic variable indented in rows below the label for that variable
 - Columns for number and percentage of cases, with units and weighting labeled
 - Note citing data source
 - c. Design elements:
 - Title: “Number of people (thousands) receiving college degrees by gender, 2000 to 2010, place”
 - One row for each year between 2000 and 2010 (chronological order)
 - One column for each gender
 - Note citing data source
 - d. Design elements:
 - Title: “Pearson correlation coefficients between height, weight, percentage body fat, systolic blood pressure, and resting pulse, elderly adults, Norway, 2008”
 - One row and one column for each variable, with label indicating units or footnote callout for abbreviated units (see *Writing about Numbers*, 2nd Edition, table 6.7, for an example)
 - Correlations reported in the below-diagonal cells
 - Symbols in the table cells to identify $p < 0.05$, with a note to explain the meaning of the symbol
 - Other notes to define unit abbreviations and specify data source

e. Design elements:

- Title: “Comparison of estimated dates for various artifacts, selected sites in North Africa and Europe”
- Artifacts arranged in the rows, grouped first by site, with artifacts indented under pertinent location; within site, artifacts arranged in rows in ascending or descending order of age
- Column naming each dating technique
- Footnotes citing data sources and references about dating methods

f. Design elements:

- Title: “Contraceptive use (%) by age group of women aged 15 to 45, US, 2009”
- One row for each type of contraceptive, arranged in conceptual groups (e.g., hormonal contraception) and then in descending order of prevalence within those groups
- One column for each 10-year age group, with a column spanner specifying units as “percentage of age group using specified method”
- Notes citing data source and explaining that each woman could report more than one contraceptive type, so percentages within each age group can sum to more than 100%

7

CREATING EFFECTIVE CHARTS

PROBLEM SET

1. List what is missing from the charts in figures 7A and 7B.

Age distribution of the elderly population
United States, 2000

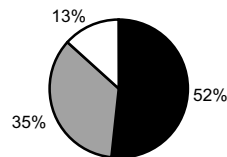


Figure 7A

Median sales price of new one-family homes,
by region, United States, 1980–2000

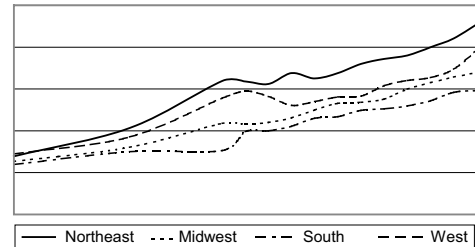


Figure 7B

2. Answer the following questions for figures 7.3, 7.5a, and 7.9 in *Writing about Numbers*, 2nd Edition.
 - a. Who is described by the data?
 - b. To what date or dates do the data pertain?
 - c. Where were the data collected?
 - d. What criteria were used to organize the values of the variables on chart axes? (*Hint*: Consider type of variable.)
 - e. What are the units of measurement? Are they the same for all numbers shown in the chart?
 - f. Are there footnotes to the chart? If so, why? If not, are any needed?
3. State which type of chart would be most appropriate for each of the following topics. Specify which variable goes on which axis, in the legend, as needed.
 - a. Annual number of people receiving college degrees by gender, from 1990 to 2010.
 - b. Average commuting costs per month, by mode of transportation (bicycle, bus, car, train, walk, other), US, 2010. (One number per type of transportation.)
 - c. Current market share for Coca-Cola, Pepsi, and other cola brands.
 - d. Distribution of SAT mathematics scores in 2013. (Range = 200 to 800 in increments of 10 points.)
 - e. Type of contraceptive (condom, diaphragm, implant/injectable, oral contraceptive [the Pill], surgical sterilization, other, none) by 10-year age groups of women aged 15 to 45 in the United States in 2010. (Some women use more than one method.)
 - f. Estimates of dates for each of 15 archeological artifacts, with margin of error for each estimate.

- g. Relationship between systolic blood pressure (mm Hg) and percentage body fat for a sample of 150 elderly people in Norway, 2008.
 - h. Educational attainment distribution (< HS, = HS, > HS) for native-born United States residents and for immigrants from each of the following locations: other North American countries, Africa, Asia, Australia & New Zealand, Europe, and Latin America, 2010.
 - i. Trends in mean annual global temperature (°F) and carbon dioxide (CO₂) concentration (ppmv.) from 1960 to 2010.
4. Draw complete rough drafts of charts for two of the topics from the preceding question, complete with appropriate title, labels, and other features. Make up data to illustrate the interior part of the chart (e.g., trend lines, bars, slices, etc.).
 5. For each of the topics in question 3, identify
 - a. the number of variables to be presented;
 - b. the types of variables to be presented (e.g., nominal, ordinal, interval, ratio);
 - c. (for XY charts only) which principle you would use to decide what order to display values on the x-axis. See chapters 6 and 7 of *Writing about Numbers*, 2nd Edition, for a list of organizing principles.
 6. Joe Schmoe is a Republican candidate for congress in his district. His campaign manager contracted a local survey firm to study opinions of major issues among his constituents. Results of the survey are shown in figure 7C.

Public opinion on pre-election issues, Anytown, 2012

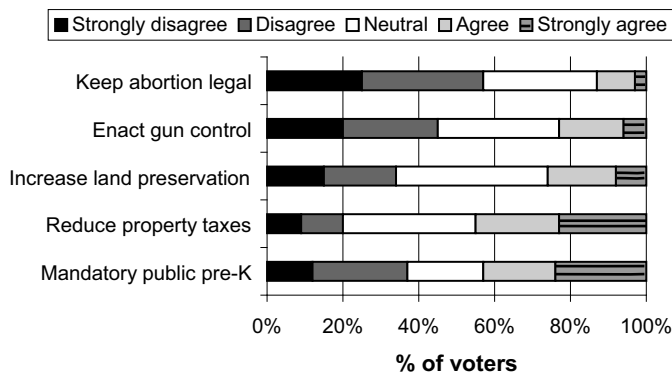


Figure 7C

- a. Which issues should he feature in his platform to maximize his chances in the primary election? Explain.
- b. Mr. Schmoe personally favors land preservation. Is that attitude likely to help or hurt him in the election?
- c. Would you characterize his district as liberal or conservative?

Use the information in table 7A to answer questions 7 through 10.

Table 7A. Total student aid and funds used to finance the postsecondary education expenses of undergraduates (in billions of 2012 \$), academic years 1990–91 to 2010–11

	Total Federal Loans	Total Federal Grants	Institutional Grants	All other
1990–91	14.7	11.4	8.6	6.9
1994–95	24.8	12.1	11.6	8.1
1998–99	30.6	13.7	14.5	17.1
2002–03	37.2	19.5	17.2	27.6
2006–07	44.9	20.6	22.9	42.8
2010–11	74.7	50.8	31.8	43.2

Note: “All other” includes education tax benefits, private and employer grants, state grant programs, nonfederal loans, and federal work-study.

Source: College Board Advocacy and Policy Center, 2014. Trends in Student Aid 2013. Trends in Higher Education Series. Table 1A. Available online at <https://trends.collegeboard.org/sites/default/files/student-aid-2013-full-report-140108.pdf>

7. Create a chart to show the distribution of aid and funds to finance undergraduates’ education in the academic year 2010–11, maintaining the units of measurement reported in table 7A.
 - a. What type of chart fits this task? Explain why.
 - b. Will your chart include data labels? If so, what type of information will they report?
 - c. Write the following information for the chart.
 - i. Title
 - ii. Legend, if needed. If not needed, explain why.
 - iii. Axis or other labels, if needed. If not needed, explain why.
8. Revise your chart from the previous question to portray the percentage distribution of values for 2010–11 instead of the original monetary units.
 - a. Will your chart include data labels? If so, how are they different from those used in the chart you created for the preceding question?
 - b. How, if at all, will the following elements of the chart change from the version you created for the preceding question?
 - i. Title
 - ii. Legend, if needed. If not needed, explain why.
 - iii. Axis or other labels, if needed. If not needed, explain why.
9. Create a stacked bar chart to present the data on undergraduate education funding sources for each of the academic years 1990–91 through 2010–11. To help you plan your chart, answer the following questions, then draw an approximate stacked bar chart, allowing the level to vary by date.
 - a. Which variable goes on the x-axis?
 - b. Which variable goes in the slices (and legend)?
 - c. Which variable goes on the y-axis, and in what units is it measured?

10. Revise your chart from the previous question to illustrate the relative importance (share) of different sources of funding in each year.
 - a. What aspects of the chart remain the same as in the previous question? What aspects change?
 - b. What are the advantages and disadvantages of the two versions of the chart in terms of portraying various aspects of trends in funding sources for undergraduate education over the period shown?

7

CREATING EFFECTIVE CHARTS

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. In a newspaper or journal article in your field, find a pie chart. Critique it using the criteria in chapter 7 of *Writing about Numbers*, 2nd Edition.
 - a. Use table 7.1 in *Writing about Numbers*, 2nd Edition, to assess whether that type of chart is appropriate for the types of variables involved.
 - b. Evaluate whether you can understand the meaning of the numbers in the chart based only on the information in the chart. Suggest ways to improve labeling and layout.
 - c. Using information in the article, revise the table to correct those errors.
 - d. Consider whether a different chart format would be more effective.
 - e. Pick a table from the article. Draft a chart to present the same information, including complete title, axis labels, legend, and notes.
2. Repeat parts a through d of question A.1 for a simple bar chart.
3. Repeat parts a through d of question A.1 for a single-line chart.
4. Repeat parts a through d of question A.1 for a multiple-line chart.
5. Repeat parts a through d of question A.1 for a clustered bar chart.
6. In a journal article in your field, find a table that presents data on a relationship between two variables. Draft a chart to present the same information, including complete title, axis labels, legend, and notes.
7. In a published article or website such as the Bureau of Labor Statistics (www.bls.gov) or US Census (www.census.gov), find data on time trends in two related concepts (e.g., trends in unemployment rates for men and for women, or values of two different stocks over a few weeks). Create a chart to depict those patterns, complete with a title, axis labels, legend, and notes.
8. In a journal article in your field, find a chart that presents the relationship between a nominal predictor variable with more than two categories, and an outcome variable.
 - a. Identify the principle the authors used to organize the categories of the nominal variable on the axis of the chart, with reference to the criteria in chapter 7.
 - b. Critique whether that organization coordinates with the associated narrative.
 - c. Sketch a revised version of the chart that addresses any shortcomings you identified in part b.

B. Applying Statistics

Hint: Save the charts you create for use in the suggested course extensions for chapters 9 and 11.

1. Estimate a frequency distribution on a nominal variable in your data set. Create a chart to present that information, complete with title, axis labels, footnotes, and legend (if pertinent). See table 7.1 in *Writing about Numbers*, 2nd Edition, to decide on the best format of chart for the type of variable.
2. Repeat question B.1 for an ordinal variable in your data set.
3. Repeat question B.1 for a continuous variable in your data set.
4. Run a cross-tabulation of two categorical variables in your data set, one of which has only two possible values. Create a chart to present the results, complete with title, axis labels, footnotes, and legend (if pertinent).
5. Estimate a difference in means for a continuous outcome variable according to values of a categorical predictor. Create a chart to present the results, complete with title, axis labels, footnotes, and legend (if pertinent). Explain the criteria you used to organize the values of the categorical predictor on the x -axis.
6. Obtain a copy of the instructions for authors for a leading journal in your field. Revise the charts you created in questions B.1 through B.5 to satisfy their criteria.

C. Writing and Revising

1. Design a chart to portray the frequencies or mean values of a series of related items (e.g., several symptoms, sources of income, or attitudinal measures) in your data set. Specify which of the principle(s) in chapter 7 you would use to organize those items on the x axis, and explain your choice
 - a. for a description in the results section of an academic paper;
 - b. for a chart to be used as a source of secondary data for other users.
2. Evaluate a chart you created previously to portray the univariate distribution of a nominal variable with more than two categories, using the checklist in chapter 7 of *Writing about Numbers*, 2nd Edition, and the instructions for authors for a journal in your field. Explain your choice of organizing principle, with reference to the specific objectives of your analysis.
3. Repeat question C.2 for an ordinal variable.
4. Repeat question C.2 for a continuous variable.

5. Evaluate a chart you created previously to portray results of a bivariate analysis involving a nominal predictor variable with more than two categories. Explain your choice of organizing principle, with reference to the specific objectives of your analysis and the criteria in chapter 7.
6. Read through the results section of your paper to identify topics or statistics for which to create additional charts that pertain to your research question. Draft them using pencil and paper, including complete title, labels, legend, and notes. If they include nominal variables or a series of related variables, identify the principle(s) you will use to organize those items on the axis and in the legend of the chart.
7. Identify a table in your paper that would be more effective as a chart. Draft and create that chart, including complete title, labels, legend, and notes. If it includes nominal variables or a series of related variables, identify the principle(s) used to organize those items on the axis and in the legend of the chart.
8. Exchange drafts of your charts from questions C.1 through C.5 with another student. Peer-edit their charts using the checklist at the end of chapter 7 and the instructions for authors for their chosen journal. Revise your charts to reflect the feedback you receive.

SOLUTIONS

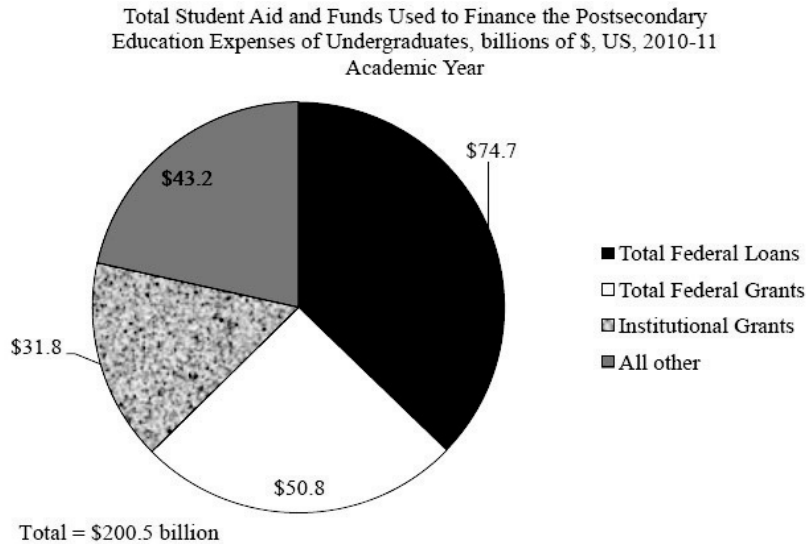
1. Figure 7A is missing a legend; 7B is missing axis titles, axis labels, and units of measurement.
3. Information on which type of chart would be most appropriate for each of the following topics, and which variables will go on each axis, or into the legend.
 - a. Multiple line chart with one line for each gender, year on the x -axis, number of persons receiving a college degree on the y -axis.
 - b. Simple bar chart, with one bar for each mode of transportation.
 - c. Pie chart with one slice for each cola brand.
 - d. Line chart (preferred with this many values) or histogram, with one data point for each 10-point score range (arranged on the x -axis), percentage of sample with each score on the y -axis.
 - e. Clustered bar chart with one cluster for each contraceptive type, a different bar color for each age group (in the legend), and the percentage of the age group using that contraceptive on the y -axis.
 - f. High/low/close chart (high and low show either standard error or 95% CI), or a bar chart with error bars added. Artifacts (named) on x -axis, age of artifact on y -axis.
 - g. Scatter chart with percentage body fat on the x -axis, systolic BP on the y -axis, and one point for the combination of values for each person in the sample.
 - h. Stacked bar chart, with one bar for each continent of origin and one slice for each educational attainment level, with percentage of the origin group on the y -axis. Each bar totals 100% of respondents of that nativity group to control for different numbers of respondents from the different continents.
 - i. Line chart with two different Y -scales, with year on the x -axis, one line for temperature (labeled on the lefthand y -axis), another line for carbon dioxide concentration (labeled on the righthand y -axis).
5. Information on number of variables, types of variables, and organizing principle for charts listed in question 3.
 - a. Annual number of people receiving college degrees by gender, from 1990 to 2010.
 - i. Three variables: number of people receiving degrees; gender; year.
 - ii. number of people receiving degrees is a continuous (ratio) variable; gender is a nominal variable; year is an ordinal variable.
 - iii. Maintain chronological order of years on the x -axis.

7

CREATING EFFECTIVE CHARTS

- b. Monthly commuting cost by mode of transportation, US, 2010.
 - i. Two variables: mode of transportation; monthly commuting cost.
 - ii. Mode of transportation is a nominal variable; monthly commuting cost is a ratio-level variable.
 - iii. Arrange mode of transportation on the x -axis in descending order of commuting cost.
- c. Current market share for Coca-Cola, Pepsi, and other cola brands.
 - i. One variable: cola brands.
 - ii. Cola brand is a nominal variable.
 - iii. Not applicable.
- d. Distribution of mathematics SAT score.
 - i. One variable: mathematics SAT score.
 - ii. Mathematics SAT score is a ratio-level variable.
 - iii. Maintain ascending numeric order of SAT scores on x -axis.
- e. Type of contraceptive (condom, diaphragm, implant/injectable, oral contraceptive [the Pill], surgical sterilization, other, none) by 10-year age groups of women aged 15 to 45 in the United States in 2010.
 - i. Two variables: type of contraceptive; 10-year age group.
 - ii. Type of contraceptive is a nominal variable; 10-year age group is an ordinal variable.
 - iii. Arrange contraceptive methods on the x -axis in conceptual groups (e.g., hormonal methods) and then descending order of prevalence within those groups. Maintain sequence of age groups in the legend.
- f. Ages of archaeological artifacts, with error bars.
 - i. Two variables: type (name) of artifact; age of artifact.
 - ii. Name of artifact is a nominal variable; age of artifact is a ratio-level variable.
 - iii. Arrange artifacts on x -axis in ascending or descending age.
- g. Relationship between systolic blood pressure (mm Hg) and percentage body fat for a sample of 150 elderly people.
 - i. Two variables: systolic blood pressure; percentage body fat.
 - ii. Both variables are continuous (ratio-level).
 - iii. Maintain ascending numeric order of percentage body fat on x -axis.
- h. Educational attainment by nativity.
 - i. Two variables: continent of origin (nativity); level of educational attainment.
 - ii. Continent of origin is a nominal variable; grouped level of educational attainment is an ordinal variable.
 - iii. Arrange continents on x -axis in ascending order of percentage with less than a high school education. Maintain ordinal sequence of educational attainment in the legend.

- i. Trends in mean annual global temperature (°F) and carbon dioxide (CO₂) concentration (ppmv.) from 1960 to 2010.
 - i. Three variables: temperature; CO₂ concentration; year.
 - ii. All three variables are continuous.
 - iii. Maintain chronological order of years on the x-axis.



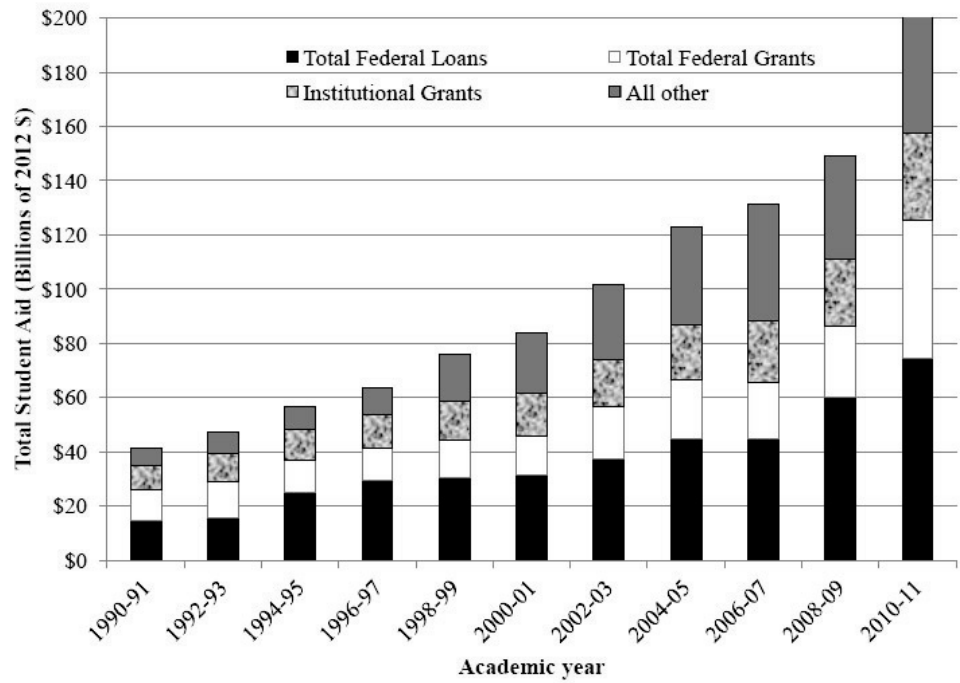
Note: "All other" includes education tax benefits, private and employer grants, state grant programs, nonfederal loans, and federal work-study.

Source: College Board Advocacy and Policy Center, 2014. Trends in Student Aid 2013. Trends in Higher Education Series. Table 1A. Available online at <https://trends.collegeboard.org/sites/default/files/student-aid-2013-full-report-140108.pdf>

Figure 7D

7. Figure 7D shows the distribution of values in 2010–11 in original units.
 - a. Figure 7D is a pie chart because the task is to present the distribution (composition) of a single nominal variable (category of funding) in one year.
 - b. Yes, data labels are included to report dollar amount from the pertinent category of undergraduate funding because that information does not appear elsewhere in the chart.
 - c. The title of the chart is drawn from the title of the table, changing the date to refer to a single academic year. The legend is needed to identify the funding sources. No axes in a pie chart.

Total Student Aid and Funds Used to Finance the Postsecondary Education Expenses of Undergraduates Academic Years 1990-91 to 2010-11



Note: "All other" includes education tax benefits, private and employer grants, state grant programs, nonfederal loans, and federal work-study.

Source: College Board Advocacy and Policy Center, 2014. Trends in Student Aid 2013. Trends in Higher Education Series. Table 1A. Available online at <https://trends.collegeboard.org/sites/default/files/student-aid-2013-full-report-140108.pdf>

Figure 7E

9. Figure 7E is a stacked bar chart to present the data shown in table 7A. Title same as source table. Axis titles, labels, and legends as shown.

PROBLEM SET

1. For each of the following topics, give an analogy to suit a general audience.
 - a. A 12-inch snowfall.
 - b. Two numbers at opposite ends of a distribution.
 - c. An erratic pattern of change.
 - d. Something moving rapidly.
 - e. A few things.
 - f. Something very heavy.
 - g. Prices that are rising rapidly.
 - h. Something that has been level for a long time and then declines suddenly and substantially.
 - i. A repetitive pattern.
2. Repeat the previous question but for a scientific audience in your field.
3. Devise short phrases to convey the concept of small size to the people listed below.
 - a. A cooking aficionado
 - b. A gardening nut
 - c. An artist
 - d. A sports fanatic
4. Each of the following analogies would work better for some audiences than others. Name a suitable audience, an unsuitable audience, and an improved analogy for the latter group.
 - a. "The size of a Blackberry."
 - b. "The gasoline shortage of the early 1970s."
5. For each of the following topics, state whether information from Illinois in 2010 would be useful as a numeric example. If so, give an example of a type of contrast in which that information could be used.
 - a. Chicago in 2010.
 - b. Illinois in 2000.
 - c. Illinois schoolchildren in 2010.
 - d. Iowa voters in 2004.
6. Your state is considering three alternative income tax scenarios: a stable tax rate (at 5%), an increase of 0.5 percentage points, and an increase of 1.0 percentage points. Your local representative wants to know how each scenario would affect low-, moderate-, and high-income residents.
 - a. What criteria could you use to define "low," "moderate," and "high" income?
 - b. What kinds of numeric contrasts would you use to compare the different scenarios?

8

CHOOSING EFFECTIVE EXAMPLES AND ANALOGIES

- c. Create a table to present the effects of those changes to the government budget agency.
 - d. Create a chart to illustrate the effects of those changes to citizens of the state.
7. State whether a \$1.00 increase would be a useful contrast for each of the following topics. If not, suggest a more reasonable increment. Assume that all examples are for the year 2010 in the United States.
 - a. Weekly allowance for an eight-year-old child
 - b. Weekly wages for a supermarket checkout clerk
 - c. Hourly wages for a supermarket checkout clerk
 - d. Sales tax on a package of cigarettes
 - e. Sales tax on a new SUV

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. Complete the following for an article in the popular press or a journal in your field.
 - a. Circle all analogies or metaphors used to illustrate quantitative patterns or relationships.
 - i. Does the author explicitly or implicitly convey the purpose in each example, or is it left unclear?
 - ii. Is it easy to understand the analogy and the pattern or relationship it is intended to illustrate?
 - b. Choose one analogy or metaphor that is unclear in that paper. Revise it using the principles in chapter 8 of *Writing about Numbers*, 2nd Edition.
 - c. Are there other places in the article where an analogy or metaphor would be helpful? Identify the purpose of the analogy or metaphor for each such situation.
 - d. Design an analogy or metaphor to suit one instance where you have suggested adding one (from part c), using the principles in chapter 8.
 - e. Identify the intended audience for the article. Choose a different audience (e.g., more quantitatively sophisticated; younger) and rewrite one analogy to suit them.
2. In the same article, circle all numeric examples where a single number is reported (e.g., not a comparison of two or more numbers). For each, indicate whether the author conveys the purpose of the example (e.g., whether it is a typical or unusual value).
3. In the same article, circle all numeric contrasts.
 - a. Indicate whether in each instance the author provides enough information for you to assess whether it is a realistic difference or change for the research question.
 - b. Evaluate whether different or additional size contrasts would be useful for the intended audience, considering
 - i. plausibility;
 - ii. real-world application;
 - iii. measurement issues.
 - c. Identify an audience that would be interested in different applications than the audience for whom the article is currently written. Describe how you would select numeric contrasts to meet their interests.

8

CHOOSING EFFECTIVE EXAMPLES AND ANALOGIES

B. Applying Statistics

1. Choose an audience to whom you want to explain results of a statistical analysis.
 - a. Devise an analogy to describe one of the main numeric patterns or relationships in your results, keeping in mind the principles described in chapter 8 of *Writing about Numbers*, 2nd Edition.
 - b. Select numeric contrasts that fit your topic, data, and context using the principles in chapter 8. Write a short description of how you chose those contrasts, including references to pertinent citations.
 - c. Review the literature in your field to determine whether standards or cutoffs are commonly used. If so, calculate and describe a contrast between your data and that standard.

C. Writing and Revising

1. Critique a paper you have written previously, using the criteria in questions A.1 through A.3 and B.1 and the checklist at the end of chapter 8. Revise your paper to correct any shortcomings you identify.
2. Exchange your answers to question B.1 with someone studying writing about a different topic, data, or audience. Peer-edit each other's work using the criteria in chapter 8 for effective analogies and examples. Revise your drafts according to the feedback you receive.

SOLUTIONS

1. For each of the following topics, give an analogy to suit a general audience.
 - a. “Knee deep.”
 - b. “Polar opposites.”
 - c. “All over the map.”
 - d. “Faster than a speeding bullet.”
 - e. “A handful.”
 - f. “As heavy as an elephant.”
 - g. “Going through the roof.”
 - h. “Like it fell off a cliff.”
 - i. “Like a broken record.”

3. Analogies for small size.
 - a. Pea-sized
 - b. Like a grain of sand or a seed
 - c. Like a speck of paint
 - d. Like a drop of water in an Olympic-sized pool

5. Use of information from Illinois in 2010.
 - a. Comparison of the state and its largest city in the same year.
 - b. Analysis of trends over time in the entire state.
 - c. Comparison of one age group to the total population.
 - d. A poor choice, as too many dimensions differ (time, place, and age).

7. Using a \$1.00 increase as a contrast.
 - a. Reasonable.
 - b. Too small. Look at \$50 difference.
 - c. Reasonable.
 - d. Okay or possibly too large. Also consider smaller contrasts.
 - e. Way too small. Look at difference of several hundred dollars.

8

CHOOSING EFFECTIVE EXAMPLES AND ANALOGIES

9

WRITING ABOUT DISTRIBUTIONS AND ASSOCIATIONS

PROBLEM SET

1. Write a description of the distribution shown in table 9A, following the guidelines for writing about univariate distributions in chapter 9 of *Writing about Numbers*, 2nd Edition.

Table 9A. Total amount borrowed by 2009 by US students beginning postsecondary education in 2003–04

Amount borrowed	Percentage of students	Percentage of borrowers
Did not borrow	43%	NA
\$1 to \$10,000	25%	44%
\$10,001 to \$20,000	16%	28%
\$20,001 to 30,000	8%	14%
\$30,001 to \$50,000	5%	9%
More than \$50,000	2%	4%

Source: College Board Advocacy and Policy Center, 2014. Trends in Student Aid 2013. Trends in Higher Education Series. Figure 11C. Available online at <https://trends.collegeboard.org/sites/default/files/student-aid-2013-full-report-140108.pdf>

2. Write a description of the race/household type patterns in table 6.1 (*Writing about Numbers*, 2nd Edition, 124), using the “generalization, example, exception” (GEE) approach explained on pp. 220–24 and 349–58 of *Writing about Numbers*, 2nd Edition. *Hint*: To compare across racial/ethnic groups, report percentage distribution of household type within each race. Why are percentages preferred to counts in this case?
3. Write a description of the age pattern of mortality shown in figure 7.17b (*Writing about Numbers*, 2nd Edition, 183), following the guidelines in chapter 9 for writing about bivariate associations. Use descriptive phrases to convey the shape of the pattern, then document with appropriate numeric evidence. Incorporate selected quantitative comparisons to illustrate size of differences in the chart.
4. What kind of test would you perform to assess the statistical significance of differences in household type by race? (Do not perform the calculation, just name the statistic!)

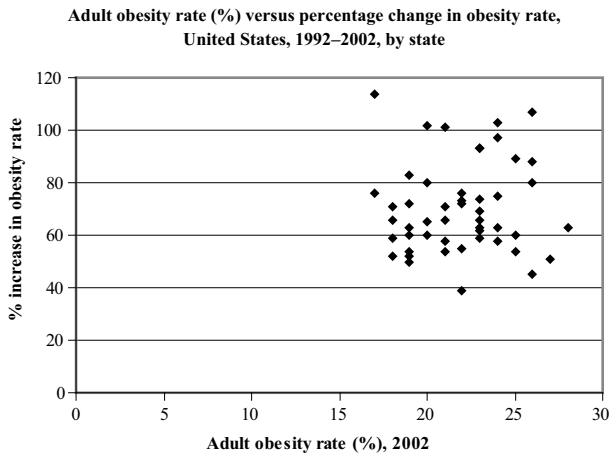


Figure 9A

5. Answer the following questions about figure 9A.
 - a. What types of variables are shown?
 - b. What kind of chart is used to portray the association? Explain why this is an appropriate choice.
 - c. Write a description of the pattern in figure 9A, reporting the level and range of each of the two variables and portraying the shape and extent of association between them.
 - d. What statistic would you use to measure the extent of association between those variables?

6. In figure A.2 (*Writing about Numbers*, 2nd Edition, 355), the pattern of emergency room use varies by race and family income. What is such a pattern called in statistical terms? In GEE lingo? Write paragraphs to describe that pattern to
 - a. a group of first-year high school students;
 - b. a group of senior undergraduate statistics majors.

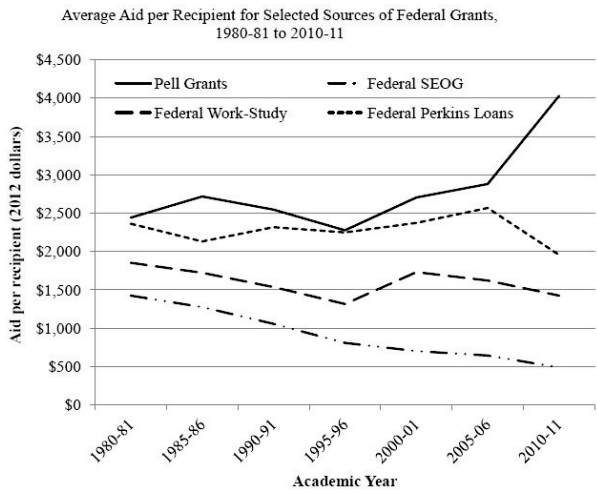


Figure 9B

7. Write a GEE description of the pattern in figure 9B.
 - a. Write topic sentences for
 - i. a paragraph about how levels of aid change across time;
 - ii. a paragraph about how levels of aid differ across sources at one point in time.
 - b. Write evidentiary sentences to illustrate the general pattern in each of those paragraphs.
 - c. Write sentences to describe and document any exceptions to those patterns, then summarize the implications of those differences for the overall shape of the association among the variables shown in the chart.
8. Write a description of the pattern in the chart you created in question 9 of the problem set for chapter 7.

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. In a journal article in your field, find a description of univariate distributions for each of the following types of variables. Critique them, using the criteria described in chapter 9 of *Writing about Numbers*, 2nd Edition.
 - a. A nominal variable.
 - b. An ordinal variable.
 - c. An interval or ratio variable with many possible values.
 - d. Rewrite the descriptions to correct any shortcomings you find in parts a through c.
2. Refer to each of the descriptions in question A.1.
 - a. Identify the criteria the author is using to choose which value to highlight. Does that value match the research question and introductory material in the article?
 - b. If all values are described with equal emphasis, assess whether one or more values should be featured and why.
3. In a journal article in your field, find a description of a table or chart portraying a bivariate association between two categorical variables.
 - a. Critique it, using the criteria in chapter 9 for writing about that type of association.
 - b. Rewrite the description to correct any shortcomings you find.
4. Repeat question A.3 for a description of a bivariate association between a categorical predictor and a continuous outcome variable.
5. Repeat question A.3 for a description of one or more bivariate correlations among continuous variables.
6. In a journal article in your field, find a description of a table or chart portraying time trends in two or more related concepts (e.g., values of several different stocks over a few weeks, or trends in unemployment rates for two or more states or regions).
 - a. Critique the description using the criteria in chapter 9 for writing about three-way associations.
 - b. Rewrite it using the “generalization, example, exceptions” (GEE) approach explained on pp. 33–36, 220–24, and 349–58 of *Writing about Numbers*, 2nd Edition.
7. In a journal article in your field, find a chart depicting a three-way association such as a clustered bar chart or simple bar chart with two or more panels.
 - a. Critique the description using the criteria in chapter 9 for writing about three-way associations.
 - b. Rewrite it using the GEE approach.

9

WRITING ABOUT DISTRIBUTIONS AND ASSOCIATIONS

B. Applying Statistics and Writing

Hint: Use the tables or charts of univariate distributions and bivariate associations that you created based on the suggested course extensions for chapters 6 and 7.

1. Using frequency distributions on one nominal, one ordinal, and one interval or ratio variable in your data set:
 - a. Write a brief description of each distribution, emphasizing the modal value using the criteria in chapter 9 of *Writing about Numbers*, 2nd Edition. Summarize the overall shape of the distribution, then report key indicators of central tendency.
 - b. Write a second description of each distribution, this time highlighting a value of interest other than the mean or mode, such as a minority group, unusual value, or most recent value.
2. Write a brief description of each of the following types of bivariate associations, using the criteria in chapter 9.
 - a. An association between two categorical variables.
 - b. An association between a categorical predictor variable and a continuous outcome variable.
 - c. Bivariate correlations among a series of continuous variables.
3. Using variables from your data set, run a three-way association among two categorical predictor variables and a continuous outcome. Write a description of that association using the GEE approach.

C. Revising

1. Evaluate descriptions of univariate statistics (distributions, central tendency) from a paper you have written previously, using the criteria in chapter 9 of *Writing about Numbers*, 2nd Edition. Revise them to rectify any shortcomings you find.
2. Evaluate descriptions of bivariate statistics (cross-tabulations, differences in means, or correlations) from the same paper. Revise them to rectify any shortcomings you find.
3. Evaluate a description of a three-way association from a results section you have written previously, using the GEE approach. Revise it to rectify any shortcomings you find.
4. Exchange drafts with a peer who is working with a different topic and data. Peer-edit each other's work and revise according to the feedback you receive.

SOLUTIONS

1. As of 2009, more than half (57%) of US students who began their postsecondary education during the 2003–4 academic year had borrowed money to pay for their education. Of those, just over 4 out of 10 (25% out of 57%) had accumulated a total educational debt of less than \$10,000 during the five years after starting their postsecondary education, while the remaining borrowers were approximately equally split between those with debt between \$10,000 and \$20,000, and those with debt of \$20,000 or more (28% and 26% of all borrowers, respectively).
3. “In 1996, age-specific death rates in the United States traced the familiar J-shape, with a substantial decline between infancy and early childhood, slowly increasing rates until middle age, and then an accelerating increase into old age. The lowest death rate was observed among children aged 5–9 years (19.4 deaths per 100,000 persons)—a 50-fold decrease from the death rate among infants (755 per 100,000). From age 50 onward, the death rate increased about 50% for each successive 5-year age group. Approximately one out of every six persons aged 85 or older died in 1996—the highest death rate of any age group.”
5. Referring to figure 9A.
 - a. Both the percentage of adults who are obese in 2002 and the percentage change in obesity rates between 1992 and 2002 are continuous variables. Percentage obese is a ratio variable; percentage change is an interval variable (no absolute zero point).
 - b. A scatter chart is used in this case because many X values (obesity rates) have more than one Y value (change in obesity rate). Also, using a scatter chart conveys the almost complete lack of an association between obesity rate and percentage change in obesity rate.
 - c. “In 2002, at least 17%, and as many as 28%, of adults in each of the 50 United States were obese. Obesity rates increased by 40% to 114% between 1992 and 2002. Somewhat surprisingly, there was virtually no correlation between the obesity rate in 2002 and the percentage change in obesity over the preceding decade. Large percentage increases were observed across the full range from low- to high-obesity-rate states.”
 - d. The Pearson correlation coefficient is the usual measure of association between two continuous variables.
7. Referring to figure 9B.
 - a. Topic sentences for two paragraphs about patterns in figure 9B.
 - i. “Over the 30-year period between academic years 1980–81 and 2010–11, Pell grants remained the single largest federal

9

WRITING ABOUT DISTRIBUTIONS AND ASSOCIATIONS

source of educational aid, followed by Perkins loans. Federal Work-Study and Federal Supplemental Educational Opportunity Grants (SEOG) were consistently the third and fourth sources, respectively.”

- ii. “Throughout the same period, aid per recipient from three of the four major US federal education aid sources remained flat or declined slightly when corrected for inflation.”
- b. Evidentiary sentences to follow the topic sentences in part a.
- i. “In the 1980–81 academic year, for instance, Pell grants and Perkins loans each averaged about \$2,400 per aid recipient, while Federal Work-Study and SEOG averaged 76% and 58% of that amount, respectively.”
 - ii. “For example, Federal Perkins loans declined in real value (constant 2012 dollars) from an average aid amount of \$2,362 in 1980–81, to \$1,955 in 2010–11.”
- c. Exceptions to the generalizations in part a, and implications for the overall pattern. To follow the description in 7.b.ii:
- “In contrast, Pell grants increased nearly 40% between the 2005–6 and 2010–11 academic years, from \$2,880 to \$4,028 per aid recipient in constant 2012 dollars. That increase, coupled with level or declining average aid from each of the three other federal sources, yielded a substantial widening in the gap between Pell and each of those sources. By 2010–11, average Perkins loan aid per recipient declined to 49%, Work-Study to 35%, and SEOG to only 12% of the average level of Pell aid in that year.”

PROBLEM SET

1. For each of the following scenarios, list what information you would report in a data section for a scientific paper. *Hint*: What additional information would you want to know?
 - a. A three-year study of a six-month drug rehabilitation program that recruited 200 subjects to examine cure and relapse rates.
 - b. A study of calcium intake among 50 pregnant women, based on their recall over a two-week period.
2. Dr. Dollar is conducting a study of 2009 poverty patterns in the United States based on annual income data from the 2010 census. She defines a categorical measure of income group comparing family income (calculated from income of individual family members, alimony, and four types of social benefits) against the federal poverty thresholds. Classifications are defined in terms of multiples of the threshold: $<.50$, $.50-.99$, $1.00-1.84$, $1.85-2.99$, and 3.00 or greater. Search for “poverty” on the US Census website (www.census.gov) for more detail. State how you would describe the poverty measure in
 - a. a one-page summary of the study for a local newspaper;
 - b. documentation of a new data set that has collected data on each of the income components as part of a written questionnaire;
 - c. a journal article on poverty patterns, written for people who are familiar with poverty thresholds.
3. Making use of newly available data from a three-year panel study of a sample of 10,000 people drawn from the 2010 census, Dr. Dollar describes movement in and out of poverty and duration of poverty (in months) over the study period. Poverty was defined as family income below the threshold (<1.0). Data were collected annually, with retrospective recall of income in each of the previous 12 months. What information would you want to add to item 2.c to describe these data for this research question?
4. A researcher in a meteorology lab accidentally erased a file containing information from two years’ worth of climatic data. Embarrassed, he went ahead and analyzed data for the other 28 years in the study. What assumptions did he implicitly make about the missing data?

10

WRITING ABOUT DATA AND METHODS

Table 10A. Coding of items about abortion attitudes, 2000 US General Social Survey

<i>“Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion . . .”</i>	Yes	No	Don't know
If there is a strong chance of serious defect in the baby?	1	2	8
If she is married and does not want any more children?	1	2	8
If the woman's own health is seriously endangered by the pregnancy?	1	2	8
If she is not married and does not want to marry the man?	1	2	8
If she became pregnant as a result of rape?	1	2	8
If the woman wants it for any reason?	1	2	8

Source: National Opinion Research Center, 2000.

5. Write a paragraph for a methods section describing how the abortion attitudes scale would be constructed by tallying the number of items listed in table 10A with which each respondent agreed. Be sure to cover the following elements:
 - a. The number and coding of the original items;
 - b. The computational method used to construct the scale;
 - c. How missing values were handled in the construction of the scale;
 - d. The valid range of values for the scale.

6. A study seeks to analyze factors affecting citizens' decisions about whether to take early retirement (at age 62) as part of a survey of the noninstitutionalized US adult population.
 - a. What subset(s) of respondents should be included in an analysis of this research question?
 - b. Design filter question(s) to identify which respondents should be asked a hypothetical series of contingent questions about reasons for early retirement.
 - c. Working from your answers to parts a and b, write a short description for the methods section about who will be included in the analysis of this topic, and how they were identified in the hypothetical survey.
 - d. Suppose that some respondents who are eligible to answer the contingent questions do not respond. What additional issues pertaining to the final analytic sample should be discussed in the methods section of a paper about that topic and data?

7. For each of the following data, methods, and objectives, write a short discussion of strengths and limitations for the concluding section of a general interest newspaper article.
 - a. Study: twenty subjects were interviewed at the Snooty Golf Club at noon on a Friday in early April regarding their preferred color and fit of jeans. Objective: a marketing study by the Abercrombie and Fitch clothing store.

- b. Study: two classes of second graders in different schools from the same town were given a math test in September. One class was then taught with a new math curriculum, the other with the standard curriculum. The classes were tested again in June. Objective: an evaluation of the new math curriculum.
- c. Study: data on hair color and age were collected for everyone aged 25–85 in a city of 200,000 people. Deaths over a two-year period were ascertained from death certificates. Two models were estimated: one with hair color as predictor and mortality as the outcome; the second with age as another predictor. Objective: understand the potential benefit of hair dye in improving survival.

10

WRITING ABOUT DATA AND METHODS

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. In a one- or two-page article in the health or science section of a newspaper or magazine, find the information on data and methods.
 - a. Critique the presentation of that information, using the guidelines for content of data and methods for general interest articles for a lay audience in chapter 10 of *Writing about Numbers*, 2nd Edition.
 - b. Assess whether additional information would be helpful for people seeking information to compare with data from another study.
 - c. Evaluate the discussion of how the data and methods affect interpretation of the findings.
2. Find a short article about a quantitative analysis in an academic journal from your field.
 - a. Critique the data and methods section using the guidelines for content of data and methods for scientific articles and the checklist in chapter 10.
 - b. List additional information needed by researchers seeking to replicate the data collection approach.
 - c. List additional information needed by researchers seeking to replicate the statistical analysis.
 - d. Assess how well the article discusses ways in which the data and methods affect interpretation of the findings.
 - e. Indicate whether the authors suggest directions for future research.
 - f. Revise the description of data and methods in the discussion section to address the problems you found in parts d and e.
3. Go to a data website such as the US Census Bureau, National Center for Health Statistics, or the Bureau of Labor Statistics and identify a topic of interest involving two or three variables. Evaluate the website in terms of how easy it is to find information about the following aspects of the data source for those variables:
 - a. the type of study design (e.g., cross-sectional sample survey, retrospective, prospective);
 - b. the data sources (e.g., vital registration forms, questionnaires, administrative records);
 - c. the overall study response rate;
 - d. the wording of questions used to collect the variables of interest to you;
 - e. the units or coding of those variables;
 - f. the item non-response rates for those variables;
 - g. design issues that affect that number of cases for which those variables are available (e.g., skip patterns, split-sample module, subgroup size);
 - h. sampling weights, if applicable.

4. In a journal article in your field, read the methods section to evaluate the author's description of the following items:
 - a. whether any of the key predictor or outcome variables for their analysis were missing by design, and if so, in what way(s) (e.g., skip patterns, split-sample modules);
 - b. item non-response rates for their key predictor and outcome variables;
 - c. how these issues affected who was included in their final analytic sample;
 - d. how many and what percentage of cases were dropped from the analytic sample due to each of the exclusion criteria;
 - e. representativeness of the final analytic sample.

B. Writing

1. Outline the data section for your analysis for a scientific paper, using the checklist in chapter 10 of *Writing about Numbers*, 2nd Edition.
2. Write a one-paragraph description of the data and methods for the same analysis for a lay audience.
3. Write the portion of your methods section related to exclusion criteria used to arrive at your final analytic sample. Be sure to cover the following elements:
 - a. whether any of your key predictor or outcome variables were missing by design, and if so, in what way(s) (e.g., skip patterns, split-sample modules, small subgroup sample size);
 - b. item non-response rates for your key predictor and outcome variables;
 - c. how these issues affected who was included in your final analytic sample;
 - d. how many and what percentage of cases you dropped due to each of the exclusion criteria;
 - e. representativeness of your final analytic sample, relative to the population it is intended to reflect.
4. For a scale or index used in your research question, write a paragraph for a methods section describing how the scale was constructed. Be sure to cover the following elements:
 - a. the number, wording, and coding of the original items;
 - b. the computational method you used to construct the scale;
 - c. how you handled missing values in the construction of the scale;
 - d. the valid range of values for the scale.
5. Complete the exercise on "Planning How to Create the Variables You Need from the Variables You Have," which is available on the website of supplemental materials for *The Chicago Guide to Writing about Numbers*,

2nd Edition. Write a description of how you created those variables, to be included in the data and methods section.

6. Write a discussion of the strengths and limitations of your data and methods for a scientific audience.
7. Exchange your answers to questions B.1 through B.6 with someone studying a different topic or data. Peer-edit each other's work.

C. Revising

1. Repeat question A.3 for a paper you have written previously. Revise your data and methods to rectify any shortcomings you identify.
2. Use the criteria in question B.3 to review a description you have previously written about how you arrived at your final analytic sample. Revise your description to correct any shortcomings.
3. Use the criteria in question B.4 to review a description you have previously written about how you constructed an index or scale. Revise your description to correct any shortcomings.
4. Review your discussion of strengths or limitations of the data and methods in a paper you have written previously. Identify additional issues that should be discussed for a scientific audience, using the criteria in chapter 10. Revise your discussion section to correct those weaknesses.
5. Exchange your answers to questions C.1 through C.4 with someone studying a different topic or data. Peer-edit each other's work and revise your drafts according to the feedback you receive.

SOLUTIONS

1. List what information you would report in a data section for a scientific paper.
 - a. What were the demographic characteristics (when, where, who) of those in the study? How were they recruited? What was the baseline response rate among recruits? What percentage of the initial sample was lost to follow-up and how? How did the sample compare demographically to all clients at that rehab center? How were “cure” and “relapse” defined and measured? By whom were these assessments made?
 - b. Again, the Ws. How were the study subjects recruited, what was the response rate, and how did the sample compare to all pregnant women? Were they asked specifically about calcium intake or to list foods? Open- or closed-ended questions about food?
3. Loss-to-follow-up (in addition to baseline response rate), how income data were collected (total or by components; in what ranges? continuous or categorical?), and item-non-response to income questions at each round of the survey.
5. “A scale was constructed to summarize respondents’ attitudes about abortion by tallying the number of circumstances in which the respondent thought it should be legal for a woman to obtain an abortion. Wording of the six component items is shown in table 10A. Possible responses to each item included “yes” (e.g., respondent thought abortion should be legal under the stated circumstance), “no,” and “don’t know.” When constructing the scale, “don’t know” and missing responses to individual items were treated as not agreeing with the item. The resulting scale ranged from 0 (respondent did not agree with any item) to 6 (respondent agreed with all items). Thus, higher values on the scale reflected a larger number of circumstances under which the respondent thought it should be legal for a woman to obtain an abortion.”
7. Write a short discussion of strengths and limitations for the concluding section of a general interest newspaper article.
 - a. “The findings from this study are probably of little use for the average Abercrombie and Fitch store. Because the data were collected during a weekday at an exclusive golf club, the opinions likely represent those of relatively affluent, nonworking adult women—a fairly small share of the Abercrombie and Fitch market. Future studies should sample younger persons of both genders from a range of income and employment groups, as well as students—the groups that comprise the more usual target audience.”

10

WRITING ABOUT DATA AND METHODS

- b. “Strengths of this study include its longitudinal nature, with testing both before and after adoption of the new curriculum in schools with each type of curriculum. However, it isn’t clear whether random assignment was used to decide which class followed which curriculum, or whether schools made their own choices of curriculum. In addition, possible differences in socioeconomic, demographic, and educational characteristics that might explain observed differences across schools or changes across time were not included.”
- c. “This study demonstrates that the association between hair color and mortality is spurious, being completely explained by their mutual association with age. Strengths of the study include the large sample size and the wide range of ages and hair colors represented, allowing their association to be observed.”

PROBLEM SET

Answer questions 1 through 5 for a scientific paper about AIDS knowledge for different language groups in the United States (results shown in table 6.2 on p. 130 of *Writing about Numbers*, 2nd Edition). Assume you are writing for a social science journal with a 5,000-word limit for research articles (e.g., several double-spaced pages apiece for the introduction, literature review, and conclusion).

1. For that scientific paper,
 - a. write an outline of the introduction, including complete topic sentences for each major paragraph.
 - b. list the kinds of numeric background information you would incorporate, and suggest types of quantitative comparisons to highlight why the topic is interesting or important.
2. Write an outline of the literature review, including headings for the different topics for which you would summarize published literature.
3. Write the results section for the scientific paper. Summarize direction, magnitude and statistical significance of differences
 - a. across language groups;
 - b. across knowledge topic areas.
4. Write an outline of the concluding section, including notes on the following issues.
 - a. How would you summarize the main numeric results?
 - b. How would the statistical significance of findings influence the way you discuss the results?
 - c. List the types of numeric background information you would use to show how findings of that study might be applied to health education programs in the United States.
5. Write a title, structured 150-word abstract, and keywords for the paper.



WRITING SCIENTIFIC PAPERS AND REPORTS



**WRITING
SCIENTIFIC
PAPERS AND
REPORTS**

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. Find a journal article about a quantitative study in your field. Evaluate the following aspects of the article, using the principles in chapter 11 of *Writing about Numbers*, 2nd Edition.
 - a. Clear introduction of the broad issues or questions to be investigated
 - b. Review of the previous literature to identify pertinent theories and existing evidence
 - c. Topic sentences that introduce the purpose of each table, chart, or quantitative comparison
 - d. Descriptions of direction, magnitude, and statistical significance of associations
 - e. Good explanations of how specific numeric findings address the questions under study
 - f. A discussion and conclusions section that summarizes numeric findings and relates them back to the research question and to previous studies
 - g. Consideration of causality and substantive significance of findings in the conclusions
 - h. Clear story line that ties together all sections of the paper
2. Critique the abstract, keywords, and title to the article used in the previous question, using the guidelines in chapter 11 and the instructions for authors for that journal. Revise them to correct any problems you identify.

B. Writing

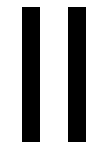
Note: If you are writing a paper on a new quantitative analysis, complete these questions. If you have already written a draft of your paper, complete section C instead.

1. Write an introductory section for your paper following the guidelines in chapter 11 and the instructions for authors for a journal in your field.
2. Write the results section of your paper.
 - a. Organize the quantitative analysis into manageable chunks, each presented in a separate table or chart.
 - b. Use paragraphs and subheadings within the results section to organize written descriptions of each table or chart, working from the versions you created for the suggested course extensions for chapters 6 and 7.
 - c. Write topic sentences that introduce the purpose of each table, chart, or quantitative comparison.
 - d. Include generalizations about direction, magnitude, and statistical significance where relevant.

- e. Write transition sentences to orient readers about how the different tables or charts relate to one another and to the overall research questions.
3. Write the discussion and conclusions section of your paper, including:
 - a. summaries of major numeric findings rather than repetition of detailed numbers from the results;
 - b. a discussion of causality, statistical significance, and substantive significance of findings (see also chapter 3 of *Writing about Numbers*, 2nd Edition);
 - c. an explanation of how your findings relate to initial hypotheses and to findings of other studies.
 4. Write an abstract for your paper following the guidelines in chapter 11 and the instructions for authors for a journal in your field.
 5. Investigate which online databases list the leading journals in your field. Write keywords to satisfy the criteria for that database, following the guidelines in chapter 11 and the instructions for that database.
 6. Following the guidelines in chapter 11 and the instructions for your selected journal, write two different titles for your paper that are
 - a. worded as a statement.
 - b. worded as a rhetorical question.

C. Revising

1. Critique the introductory section of a paper about a quantitative analysis that you have written previously, using the criteria in question A.1. Rewrite it to rectify any problems you have identified.
2. Critique the results section of that paper using the criteria listed in question B.2. Rewrite it to rectify any problems you have identified.
3. Critique the discussion and conclusions section of that paper using the criteria listed in question B.3. Rewrite it to rectify any problems you have identified.
4. Repeat question A.2 for your paper.
5. Exchange initial and revised drafts of the materials in questions C.1 through C.4 with someone writing about a different topic or data. Peer-edit each other's work and revise according to the feedback you receive.



WRITING SCIENTIFIC PAPERS AND REPORTS

SOLUTIONS

1. For a scientific paper about AIDS knowledge:
 - a. Outline of introduction to study of AIDS knowledge by language in the United States
 - i. (Paragraph on why AIDS is of concern) Topic sentence: “AIDS (Acquired Immunodeficiency Syndrome) is a leading cause of death in the United States.”
 - ii. (Paragraph on why it is important to assess AIDS knowledge) Introductory sentences: “In the absence of a vaccine against AIDS, prevention must rely on individual behavior to avoid spread of the disease. It is unlikely that appropriate behavioral change will occur without knowledge about AIDS and how it is transmitted; hence, it is important to assess levels of AIDS knowledge in the general population.”
 - iii. (Paragraph on why language is an important possible mechanism related to AIDS knowledge) Topic sentence: “Language can affect AIDS knowledge either through linguistic barriers or cultural differences.”
 - b. Kinds of numeric information to incorporate and suggested quantitative comparisons for an introduction to the AIDS knowledge paper

For paragraph i, incorporate statistics on levels and trends in AIDS prevalence and mortality in the United States for the period under study, using percentage change to quantify trends over time in AIDS prevalence and mortality rates, and rank as a cause of death to indicate overall importance. If data are more than 5 years old, consider adding current statistics on these outcomes.

For paragraph ii, incorporate evidence on how knowledge about AIDS or other similar diseases such as STDs translates into changes in preventive behaviors.

For paragraph iii, incorporate statistics on how AIDS prevalence and mortality vary by language ethnic group, with supplementary evidence by race or socioeconomic status if statistics by language are not available. Use ratios or percentage difference to contrast rates across groups. Include information on the number of persons, percentage share, and trends in number and share of major language groups in the United States around the time of the study period (and more recently if statistics are dated).
3. “An analysis of survey data from the late 1990s in New Jersey showed that respondents were more likely to understand the ways AIDS can be transmitted than ways AIDS cannot be transmitted (table 6.2). On average, English speakers answered 91.7% of the “likely” transmission questions

correctly, compared to an average of 59.8% of “unlikely” questions correctly. Similar patterns were observed among Spanish speakers.

“Although more than 88% of respondents from each of the three language groups answered the question about transmission via sexual intercourse correctly, for the other three questions about likely transmission modes, incorrect answers were more common among Spanish speakers, especially those who completed the questionnaire in Spanish. For example, only two-thirds of that group knew that the AIDS virus is very likely to be spread via shared IV drug needles, compared to 92% of English speakers and 91% of Spanish respondents who completed the questionnaire in English.

“Responses to questions about transmission via casual contact were more troubling (bottom panel of table 6.2), with fewer than two-thirds of respondents knowing the correct answers to those questions. Most striking was the widespread misconception about transmission from a medical provider: only one-third knew that such spread is unlikely. Differences across language groups were sizeable and statistically significant for all but the medical provider question.”

5. Title: “Differences in AIDS Knowledge among Spanish and English Speakers in New Jersey by Ability to Speak English”

Abstract:

Objectives: Previous studies have shown that Hispanic persons know less about AIDS than non-Hispanics in the United States, but few studies have examined the role of language.

Methods: We use data from a 1998 population-based sample survey of New Jersey adults to compare levels of AIDS knowledge among English and Spanish speakers, taking into account ability to speak English.

Results: Spanish-speakers, especially those who filled out the questionnaire in Spanish, know less than English speakers about AIDS transmission. Respondents of all language groups were more likely to understand ways that AIDS is likely to be transmitted than ways in which transmission is unlikely.

Conclusions: Differences between Hispanics who speak primarily English and those who speak primarily Spanish underscore the idea that linguistic and cultural barriers vary within a cultural group and should be taken into account in the design of AIDS education programs.

Keywords: Acquired Immunodeficiency Syndrome; Ethnic Groups; Hispanic Americans; Knowledge, Attitudes and Practices; Language.

12

SPEAKING ABOUT NUMBERS

PROBLEM SET

1. Adapt the material in text box 13.3 and figure 13.5 (*Writing about Numbers*, 2nd Edition, 343–45) into slides for a 10-minute presentation. Include a good title for each slide and write comments that explain how the material illustrates the principles of how to write about numbers from chapter 2.
2. Write the speaker's notes to accompany the slides you created for the previous question.
3. Create one or more slides to present the following material to a scientific audience. "The Center for Epidemiological Studies-Depression Scale (CES-D) is a 20-item scale for epidemiological research that was developed by the National Institute of Mental Health. Respondents are asked to choose from four possible responses in a Likert format, where '0' is 'rarely or none of the time (less than one day per week),' and '3' is 'almost or all of the time (five to seven days per week).' The theoretical range is from 0 to 60, with higher scores reflecting greater levels of depressive symptoms. The CES-D has four separate factors: depressive affect, somatic symptoms, positive affect, and interpersonal relations. The CES-D has very good internal consistency with alphas of .85 for the general population and .90 for a psychiatric population (Radloff 1977)."
4. Write the speaker's notes to accompany the slides you created for the previous question.
5. Write "Vanna White" notes to introduce the tables or charts listed below.
 - a. Table 4.1c ("Poverty rates by age group," *Writing about Numbers*, 2nd Edition, 74)
 - b. Table 6.1 ("Households by type, race, and Hispanic origin," *ibid.*, 124)
 - c. Figure 7.2b ("Federal outlays by function," *ibid.*, 154)
 - d. Figure 7.5a ("Odds ratios of emergency room visits for asthma, by race and income," *ibid.*, 159)
 - e. Figure 7.17b ("Death rates by age," *ibid.*, 183)
6. Practice presenting (out loud) one table and one chart from question 5, using the Vanna White notes you wrote for that exhibit. Evaluate each of those mini-presentations using the checklist in chapter 12. Revise the oral presentation of each slide to fit within two minutes.
7. Adapt the following tables into simpler tables or charts for use on slides for a speech. Aim for one concept or series of closely related concepts per chart.
 - a. Table 6.1 ("Households by type, race, and Hispanic origin," *Writing about Numbers*, 2nd Edition, 124)
 - b. Table 6.8 ("Drug use by arrestees in selected major United States cities by type of drug and sex, 1999," *ibid.*, 138)

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. Attend a professional research presentation related to your interests.
After you return, write a critique evaluating the following aspects of the talk, using the guidelines for effective speeches to academic audiences in chapter 12 of *Writing about Numbers*, 2nd Edition.
 - a. Title of the presentation
 - b. Contents of slide titles
 - c. Choice of slide types (e.g., text, table, chart, or image) for each component of the talk (e.g., presentation of numeric facts, theoretical relationships among variables, context of the data, etc.)
 - d. Introductory slides for conveying importance of the topic
 - e. Data and methods slides
 - f. Results slides
 - g. Explanations of patterns in tables and charts (purpose, layout, patterns)
 - h. Tailoring of material to suit the specific audience
 - i. Overall story line of the talk (maintaining an orientation to the overall purpose of the study; order of topics; appropriate balance of “big picture” and little details)
 - j. Clarity and engagement of the spoken presentation
 - k. Layout, type size, use of color, and other formatting on tabular and chart slides
 - l. Layout, type size, use of color, and other formatting on text slides
 - m. Whether the speech fit within the allotted time, including allocation of appropriate amounts of time to each section of the material

B. Writing

1. Create slides for a 20-minute presentation about a quantitative analysis to a scientific audience in your field, following the guidelines in chapter 12. Include slides for each major section of the paper, including introduction, literature review, data and methods, results (several charts or tables; see next question), and conclusions.
2. Adapt charts or tables from your paper to be used on the slides, using the guidelines in chapter 12.
3. Write speaker’s notes for the presentation, including “Vanna White” directions for the slides created in the preceding question, following the guidelines on pp. 306–310 of *Writing about Numbers*, 2nd Edition.
4. Exchange draft slides and speaker’s notes with a peer who is working on a different topic and data. Evaluate each other’s work, using the checklist at the end of chapter 12. Revise your slides and speaker’s notes according to the feedback you receive.

12

SPEAKING ABOUT NUMBERS

5. Ask a test audience to evaluate a live presentation of your talk for your specified audience and allotted time, using the criteria under “Dress Rehearsal” on pp. 311–12 of *Writing about Numbers*, 2nd Edition.
6. Make revisions to slides and speaker’s notes based on what you learned in your rehearsal.

C. Revising

1. Evaluate the slides you have previously created for a 15–20 minute speech to a scientific audience, using the criteria in chapter 12 of *Writing about Numbers*, 2nd Edition, and in question A.1. Revise the slides to rectify any shortcomings you identify.
2. Pick one large table from the results section of your paper. Revise it into several simpler table slides or chart slides using the guidelines in chapter 12, and incorporate those slides into the presentation used in question C.1.
3. Write “Vanna White” notes to introduce and explain one table and one chart from your revised presentation, using the guidelines in chapter 12. Incorporate them into your speaker’s notes.
4. Evaluate the full set of speaker’s notes for the same speech. Revise them to rectify any shortcomings you identify.
5. Exchange your revised work from questions C.1 through C.4 with someone working on a different topic and data. Peer-edit each other’s work.
6. Repeat questions B.5 and B.6 for a speech you have written previously.

SOLUTIONS

1. Figures 12A, 12B, 12C, 12D, 12E, 12F, and 12G are slides for a presentation about the physical impact of the planes on the Twin Towers in box 13.3 and figure 13.5.

12

SPEAKING ABOUT NUMBERS

Annotated example of good writing

- Article from front section of New York Times:
 - “First Tower to Fall Was Hit at Higher Speed, Study Finds”
 - E. Lipton and J. Glanz (2/23/02).
- Tailoring to the audience and objectives:
 - An educated lay audience.
 - Two page article.

Figure 12A

Airplane speed

- “The FBI said the government's analysis put the speeds at 586 m.p.h for the United flight and 494 m.p.h. for the American one.”
 - *Basic principle: report numbers.*
- “In both cases, the planes were flying much faster than they should have been at that altitude. The aviation agency's limit below 10,000 feet is 287 m.p.h.”
 - *Basic principle: compare against a standard to help interpret number.*

Figure 12B

Energy and impact of planes

- “The energy of motion carried by any object, called the kinetic energy, varies as the square of its velocity, so even modest differences in speed can translate into large variations in what the building had to absorb.”
 - *Basic principle: define concepts using simple wording.*
- “That means that while the United jet was traveling only about a quarter faster than the American jet, it would have released about 50 percent more energy on impact.”
 - *Tool: relative difference and % difference calculations.*

Figure 12C

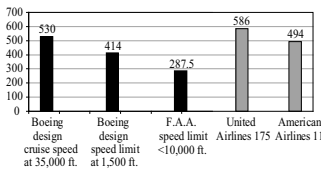
Just how much energy is that?

- “Even at a speed of only about 500 m.p.h., a partly loaded Boeing 767 weighing 132 tons would have created about three billion joules of energy at impact, the equivalent of three-quarters of a ton of T.N.T.”
 - *Basic principle: interpret numbers and relate them to familiar quantities.*

Figure 12D

How did speeds compare to design limits?

Impact speed of 9/11 flights* and comparison speeds



- Uses a bar chart to illustrate speed of planes relative to important benchmarks.
 - *Basic principle: choose the right tools.*
- Describe patterns in chart by pointing out that planes’ speeds exceeded design limits.
 - *Basic principle: compare against meaningful cutoffs.*

Figure 12E

Why do design limits matter?

- Such speeds threatened the structural integrity of the planes even before they struck the buildings, because the lower the plane goes, the thicker the air becomes, so the slower the plane must travel to avoid excessive stress.”
 - *Basic principle: explain complex concepts in simple terms.*
 - In this case, explaining principles of physics.

Figure 12F

Authors’ use of tools and principles

- Explained complex ideas without (much) jargon.
 - Energy on impact.
 - Effect of altitude on stress.
- Compared against
 - Useful benchmarks
 - FAA speed limit.
 - Design speed limit.
 - Familiar examples
 - TNT.
- Used appropriate tools.
 - Chart to show relative speed.
 - Prose to:
 - Report a few numbers.
 - Explain patterns.
 - Define terms.
 - Types of quantitative comparisons:
 - Absolute difference.
 - Relative difference.
 - Percentage difference.

Figure 12G

3. Figures 12H and 12I are slides about data and methods regarding the CES-D scale for a scientific audience.

CESD scale

- Center for Epidemiological Studies Depression (CESD) Scale
 - Developed by National Institute of Mental Health (NIMH)
- 20 items on frequency of symptoms in past week
 - Each scaled from 0 (“rarely or none of the time”) to 3 (“almost or all of the time”).
- Very good internal consistency:
 - $\alpha = .85$ for the general population.
 - $\alpha = .90$ for a psychiatric population

Source: Radloff, 1977.

Figure 12H

Factors within the CESD scale

- Four separate factors:
 - Depressive affect.
 - Somatic symptoms.
 - Positive affect.
 - Interpersonal relations.

Figure 12I

5. “Vanna White” notes to tables and charts. Bracketed material describes accompanying gestures and is *not* intended to be spoken.
- a. “Table 4.1c presents poverty rates by age group in the United States in 2012. There is one row for each age group, and a row for all age groups combined [point to each row in turn]. Columns report the number of poor persons [point] and population [point], each reported in thousands of persons, and the poverty rate as a percentage of persons in the pertinent age group [point]. In 2012, the poverty rate for children was 21.8% of children, compared to 13.7% of people aged 18 to 64, and 9.1% of those aged 65 or older.”
 - b. “Table 6.1 shows the distribution of households by type of household, race, and ethnic origin in the United States in 1997. All numbers are reported in units of thousands, meaning that there were 102 million when all races and household types were combined [point to bottom row in first column of numbers]. Households are divided into family and nonfamily households, shown in the middle and right-hand sections of the table [point]. The racial distribution of all households is in the leftmost column of numbers [wave vertically down that column]. Distributions of household types are also broken out separately by race and Hispanic origin—shown in the rows [gesture down the column of row labels].” [Then name the modal household type and how it compares to other household types of interest, overall, and by race.]

- c. “The distribution of federal outlays by major function in the United States in 2000 is shown in figure 7.2b. Human resources (the black wedge [point]) comprised by far the largest single category of federal outlays (61% of the \$1.8 trillion spent that year). The second largest category—national defense (dotted fill)—accounted for only about one quarter as much as human resources (16% of the total). Net interest, physical resources, and other functions together amounted for the remaining 23% of total outlays [point to each wedge as you mention its category].”
 - d. “Figure 7.5a shows how the relative chances of emergency room use among asthmatic children varied by family income and race in the United States in 1991. Income groups (poor, near poor, and non-poor) are arranged from left to right along the horizontal axis [wave along the axis]. There is a different color bar for each racial group—black for black children and gray for non-black children [point at legend or one cluster of bars]. The heights of the bars show the relative chances of emergency room use for asthma.” [Then describe the patterns by race and by income as in appendix A.]
 - e. “The age pattern of death rates in the United States in 1996 is shown in figure 7.17b. Age is grouped into five-year categories across the axis. Death rates are shown on a logarithmic scale on the *y*-axis, measured in number of deaths per 100,000 persons in the pertinent age group. The logarithmic scale is used to allow a wide range of death rates to be shown on a single graph without obscuring differences at the low end of the range. [For a lay audience, add: That means that the distance between successive marks on the *y*-axis corresponds to a 10-fold increase in death rates.] For example, death rates increase by nearly a multiple of 10 between the age category of 20–24 and that of 50–59 (from 101.3 deaths per 100,000 persons to 851.3 per 100,000). Death rates increase by another factor of 10 between the age category of 50–59 and that of 80–84 years.” [Then go on to describe the J-shaped pattern, as in the answer to question 3 from chapter 9.]
7. Create simplified tables or charts for use on slides.
- a. From table 6.1 in *Writing about Numbers*, 2nd Edition.
 - i. Text slide to present distribution by major family type, all races combined. Extracted from cells in bottom row of table 6.1. Text bullet: “Sixty-nine percent of the 102,528 US households in 1997 were family households.”
 - ii. Pie charts (figures 12J and 12K) to present racial distribution and Hispanic origin distribution, based on information in the first numeric column of table 6.1. Titles modified from the title to table 6.1 to reflect the specific variable in each chart.

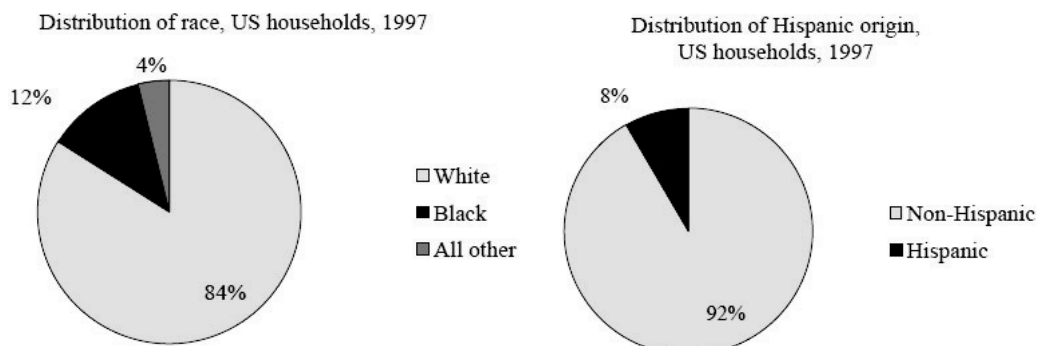


Figure 12J

Figure 12K

iii. Two simple tables to present distribution of family types by race (table 12A) and by Hispanic origin (table 12B), created by splitting the table into separate tables for race (top rows of table 6.1) and Hispanic origin (bottom rows of table 6.1). Tables 12A and 12B put household type in the rows and race or Hispanic origin in the columns so the material will fit better on a slide with an adequate (minimum 16 point) type for all labels and numbers. Table titles reflect the specific comparisons in each table.

Table 12A Family households (thousands) by type and race, United States, 1997

	White	Black	All other	Total
All Households	86,106	12,474	3,948	102,528
Family households	59,511	8,408	2,961	70,880
Married couple	48,066	3,921	2,330	54,317
Other families				
Female householder	8,308	3,926	418	12,652
Male householder	3,137	562	212	3,911
Non-family households	26,596	4,066	986	31,648
Female householder	14,871	2,190	455	17,516
Male householder	11,725	10,901	532	14,133

Table 12B Family households (thousands) by type and Hispanic origin, United States, 1997

	Non-Hispanic	Hispanic	Total
All Households	93,938	8,590	102,528
Family households	63,919	6,961	70,880
Married couple	49,513	4,804	54,317
Other families			
Female householder	11,040	1,612	12,652
Male householder	3,366	545	3,911
Non-family households	30,018	1,630	31,648
Female householder	16,762	754	17,516
Male householder	13,258	875	14,133

b. Figures 12L (males) and 12M (females) are clustered bar charts, each showing the rates of positive drug tests by city (x-axis, sorted in descending order of use for all drugs) and type of drug (in legend, sorted in descending order of frequency, L to R within each cluster). Data for males is from the lefthand set of columns in table 6.8; females from the righthand set of columns. Titles to figures 12L and 12M are modified from the title to table 6.8 to reflect the gender subgroup in each chart.

Percentage of arrestees testing positive for drugs, by city and type of drug, US males, 1999

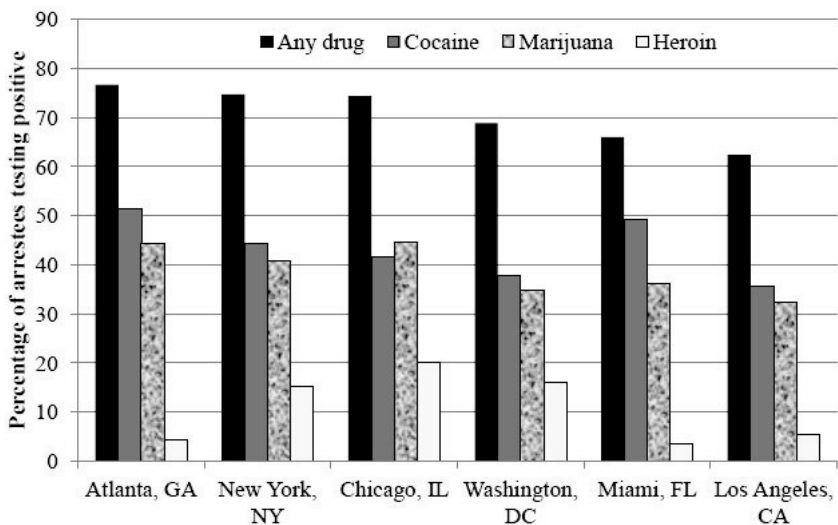
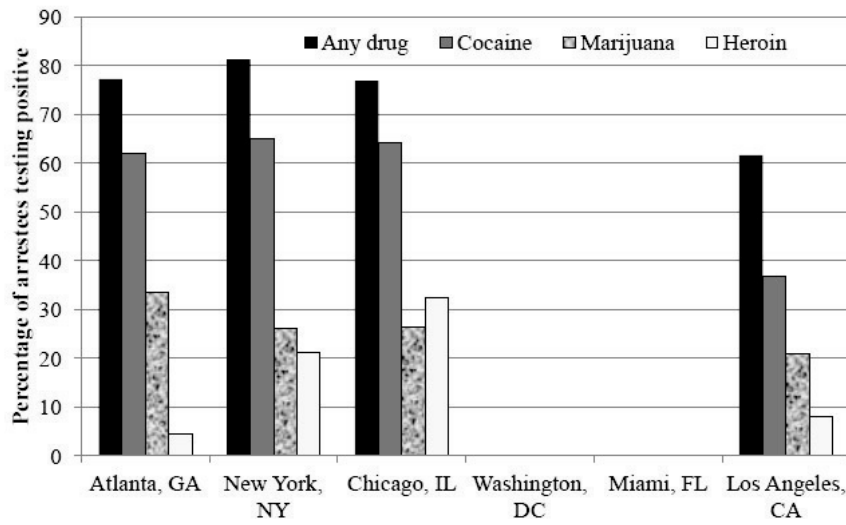


Figure 12L

Percentage of arrestees testing positive for drugs, by city and type of drug, US females, 1999



Data for female arrestees not available for Washington DC or Miami, FL.

Figure 12M

13

WRITING FOR APPLIED AUDIENCES

ISSUE BRIEFS, CHARTBOOKS, POSTERS, AND GENERAL- INTEREST ARTICLES

PROBLEM SET

Answer questions 1 through 4 using the information in text boxes 10.1, 10.2, 10.3, 11.1, 11.2b, and 11.3 in *Writing about Numbers*, 2nd Edition.

1. Write the following components of a two-page issue or policy brief about the birth weight study, following the guidelines on pp. 333–38 of *Writing about Numbers*, 2nd Edition.
 - a. A title
 - b. A statement of why the topic is important
 - c. One or two simplified tables or charts to summarize the key results
 - d. Short descriptions of each table or chart from part c of this question
 - e. Paragraphs explaining how the findings apply to at least two sets of stakeholders
 - f. A sidebar describing the study methods
2. Revise the material in box 10.2 for a general interest newspaper article, following the guidelines on p. 341 of *Writing about Numbers*, 2nd Edition.
3. Design a research poster about the birth weight study for an applied audience, following the guidelines on pp. 328–32 of *Writing about Numbers*, 2nd Edition. Sketch the poster layout and provide notes about the contents of each page, adapting them from the tables, charts, and text boxes in chapters 10, 11, and 12 of *Writing about Numbers*, 2nd Edition.
4. Outline a descriptive report about the birth weight study for a lay audience, following the guidelines on pp. 340–41 of *Writing about Numbers*, 2nd Edition.
 - a. Write the section headings—one for each major question or topic covered in that study.
 - b. Adapt tables 6.4, 6.5, 9.1, and 11.1 into simplified tables or charts, each of which focuses on one finding or set of related findings. Write the titles for the charts or tables that would go in each section of the report.

Answer questions 5 through 8 using the tables, charts, and prose you wrote as solutions to questions 7, 9, and 10 in the problem set for chapter 7, and questions 1, 7, and 8 in the problem set for chapter 9. It may be helpful to download the College Board report from which these data were drawn:

College Board Advocacy and Policy Center, 2014. Trends in Student Aid 2013. Trends in Higher Education Series. Available online at <https://trends.collegeboard.org/sites/default/files/student-aid-2013-full-report-140108.pdf>

5. Using the guidelines on pp. 338–40 of *Writing about Numbers*, 2nd Edition, design chartbook pages to present information on patterns of student aid in the United States to an applied audience.

6. Using the guidelines on p. 341 of *Writing about Numbers*, 2nd Edition, write an introductory paragraph for a general interest article about trends in the levels and sources of student aid in the United States. List the kinds of background information you would incorporate and suggest useful types of quantitative comparisons to highlight why the topic is interesting or important.
7. Using the guidelines on pp. 340–42 of *Writing about Numbers*, 2nd Edition, write an executive summary for the report on student aid in the United States.
8. Identify three different stakeholder groups with interests in student aid for postsecondary education in the United States. Write a short paragraph for each group highlighting key conclusions of the analysis and how they might respond.

13

WRITING FOR APPLIED AUDIENCES

ISSUE BRIEFS, CHARTBOOKS, POSTERS, AND GENERAL- INTEREST ARTICLES

SUGGESTED COURSE EXTENSIONS

A. Reviewing

1. At a departmental research presentation day or professional conference in your field, find a poster related to your interests. Discuss the research project with the poster's author. After you return, evaluate the following, using the criteria under "Posters" on pp. 328–32 of *Writing about Numbers*, 2nd Edition.
 - a. Title of the poster
 - b. Ease of understanding of data and methods description for (i) researchers in your field; (ii) nonstatisticians
 - c. Accessibility of research findings to (i) researchers in your field; (ii) nonstatisticians
 - d. Relevance of conclusions for an applied audience
 - e. Clarity of the overall story line on the poster
 - f. Poster layout
 - g. Type size and other formatting
 - h. Availability and quality of handouts
 - i. Length and clarity of the presenter's oral description of the poster contents
2. Find an issue brief or policy brief related to a research topic in your field or at a website such as the Urban Institute (<http://www.urban.org>). Evaluate the following elements of the brief, using the guidelines under "Issue and Policy Briefs" on pp. 333–38.
 - a. Ease of understanding for nonstatisticians
 - b. Simplicity of tables and charts
 - c. Appropriateness of vocabulary for the intended audience
 - d. Layout
3. Find a chartbook about a research topic in your field or at a website such as the US Social Security Administration (<http://www.ssa.gov/policy/docs/chartbooks/>) or Healthy People 2020 (<http://www.healthypeople.gov/>). Evaluate it using the criteria on pp. 338–39.
4. Find a descriptive report about a topic in your field or at a website such as the Office of Human Services Policy (<http://aspe.hhs.gov/topics0.cfm/>). Evaluate it using the criteria on p. 340.
5. Find a newspaper or magazine article that summarizes findings of a quantitative study. Critique the following elements of that article, using the principles in chapter 13 of *Writing about Numbers*, 2nd Edition.
 - a. Introduction
 - b. Explanation of study data and methods
 - c. Tables and charts
 - d. Sentences to introduce each table, chart, or quantitative comparison

- e. Interpretation of the numeric evidence
- f. Summary of the findings

B. Writing

1. Adapt charts or tables from your paper to be used on the poster, including specific titles and bulleted text annotations.
2. Write “Vanna White” notes for each table or chart on your poster, following the guidelines on pp. 306–10 of *Writing about Numbers*, 2nd Edition.
3. Create a 4' by 8' poster about a research project you are conducting, to be presented at a departmental or professional conference, following the guidelines on pp. 328–32 of *Writing about Numbers*, 2nd Edition.
 - a. Design poster components for each major section of the paper, including an introduction, literature review, data and methods, results (using the charts and tables you created for question B.1), and conclusions.
 - b. Draft the layout of the poster, including a title, abstract, and the components you designed in part a of this question.
4. Write a narrative to accompany your poster. Include short modules for each of the following.
 - a. An introduction to your topic and project
 - b. A description of the data used in your analysis
 - c. A summary of important theories on your topic
 - d. A summary of key previous studies on your topic
 - e. An explanation of your methods for someone familiar with inferential statistics
 - f. An explanation of your methods for nonstatisticians
 - g. The key findings of your study
 - h. The policy or program implications of your work
 - i. The research implications of your study (strengths, weaknesses, directions for future work)
5. Create handouts for the poster you created in questions B.1 and B.3 for
 - a. a statistical audience.
 - b. an applied audience.
6. Evaluate and revise the poster, narrative, and handouts you created in questions B.1 through B.5.
 - a. Ask a peer to evaluate your poster and associated narrative and handouts, given your specified audience and using the criteria under “Posters” on pp. 328–33 of *Writing about Numbers*, 2nd Edition.
 - b. Revise the poster, narrative, and handouts based on the feedback you received from your rehearsal of the poster presentation.

7. Write a two-page issue brief about a quantitative analysis, following the guidelines on pp. 333–38 of *Writing about Numbers*, 2nd Edition.
8. Write a two- or three-page general-interest article about the purpose, findings, and implications of your study, following the guidelines on p. 341 of *Writing about Numbers*, 2nd Edition.
9. Write a chartbook about your study, following the guidelines on pp. 338–40 of *Writing about Numbers*, 2nd Edition.
10. Repeat questions B.1 through B.5 from the suggested course extensions to chapter 12, writing a ten-minute oral presentation to a lay audience about the same study you used for those questions.

C. Revising

1. Repeat question A.1 for a poster you have created previously about an application of a quantitative analysis.
2. Revise the poster, narrative, and handouts to rectify any problems you identified in the previous question.
3. Evaluate a report you have written previously for a lay audience about an application of a quantitative analysis, using the criteria on pp. 317–26 of *Writing about Numbers*, 2nd Edition. Revise it to rectify any problems you identified.
4. Ask a peer who is familiar with the statistical and substantive knowledge level of your intended audience to critique the revised draft of the report you used in the preceding question, using the criteria on p. 317–26 of *Writing about Numbers*, 2nd Edition. Revise it based on the feedback you received.
5. Evaluate a speech you have written previously for an applied audience about an application of a quantitative analysis, using the criteria in chapters 12 and 13. Revise it to rectify any problems you identified.
6. Ask a peer who is familiar with the statistical and substantive knowledge level of your intended audience to listen to the revised speech you used in the preceding question. Have them critique it, using the criteria in chapters 12 and 13. Revise it based on the feedback you received.

SOLUTIONS

1. Components of two-page policy brief about the birth weight study.
 - a. Title: “Does socioeconomic status explain racial/ethnic differences in birth weight in the United States?”
 - b. Statement of importance of the research question, edited from box 11.1:

“Low birth weight is a widely recognized risk factor for infant mortality and a variety of health and developmental problems that can persist into childhood and even adulthood. In 1999, US infants born weighing less than 2,500 grams (5.5 pounds) were 24 times as likely as normal birth weight infants to die before their first birthday (60.5 deaths per 1,000 live births and 2.5 deaths per 1,000, respectively; Mathews, MacDorman, and Menacker 2002). Although they comprised about 7.5% of all births, low birth weight infants accounted for more than 75% of infant deaths (Paneth 1995).

“Costs associated with low birth weight are substantial: in 1995, Lewit and colleagues estimated that \$4 billion—more than one-third of all expenditures on health care for infants—was spent on the incremental costs of medical care for low birth weight infants. Higher risks of special education, grade repetition, hospitalization, and other medical costs added more than \$85,000 (in 1995 dollars) per low birth weight child to costs incurred by normal birth weight children through age 15 (Lewit et al. 1995).

“Despite considerable efforts to reduce the incidence of low birth weight, the problem remains fairly intractable: between 1981 and 2000, the percentage of low birth weight infants rose from 6.8% to 7.6% of all infants, in part reflecting the increase in multiple births (Martin et al. 2002). Rates of low birth weight among black infants remained approximately twice those among white infants over the same period (13.0% and 6.5% in 2000, respectively). This study analyzes the extent to which that pattern is due to the lower socioeconomic status of black children compared to children of other races in the United States.”

13

WRITING FOR APPLIED AUDIENCES

ISSUE BRIEFS, CHARTBOOKS, POSTERS, AND GENERAL- INTEREST ARTICLES

c. Tables and charts:

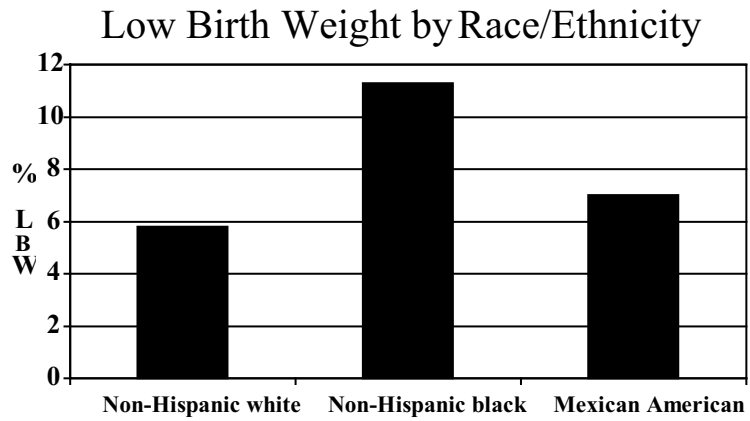


Figure 13A

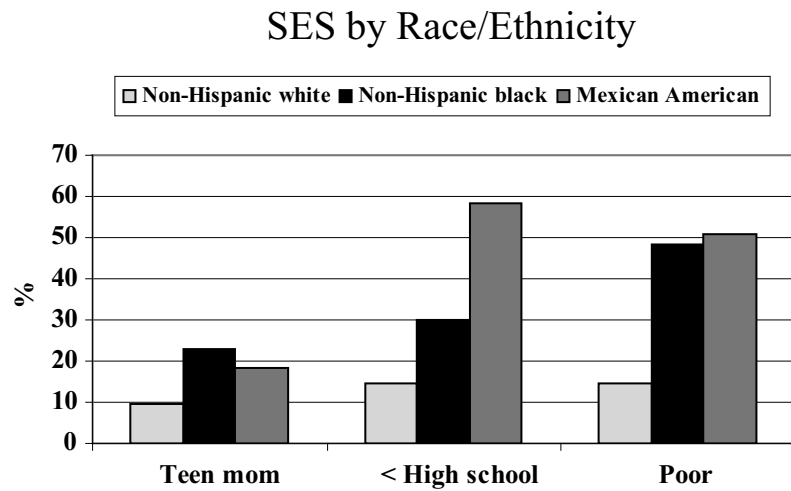


Figure 13B

Maternal Smoking by Race/Ethnicity	
	<i>Smoked cigarettes (%)</i>
Non-Hispanic white	26.8
Non-Hispanic black	22.9
Mexican American	10.1

Figure 13C

- d. Descriptions of the tables or charts, adapted from material in box 11.2b:
- i. Figure 13A: “Non-Hispanic black infants were nearly twice as likely as non-Hispanic white infants to be low birth weight (11.3% versus 5.8%). Mexican-American infants were only about 20% more likely than whites to be low birth weight.”
 - ii. Figure 13B: “In every dimension of socioeconomic status studied here, non-Hispanic black and Mexican American mothers were substantially disadvantaged relative to their non-Hispanic white counterparts. They were twice as likely as non-Hispanic white mothers to be teenagers at the time they gave birth, and two to three times as likely to be high school dropouts. Mean income/needs ratios for non-Hispanic black and Mexican American families were roughly half that of non-Hispanic white families.”
 - iii. Figure 13C: “In contrast to the socioeconomic patterns, maternal smoking—an important behavioral risk factor for low birth weight—was more common among non-Hispanic white women (27%) than non-Hispanic black (23%) or Mexican American women (10%).”
- e. Paragraphs describing how findings apply to two sets of stakeholders.
- i. “Obstetricians should be aware that non-Hispanic black infants at all socioeconomic levels are more likely than non-Hispanic

white infants to be low birth weight. They should provide outreach to increase enrollment rates of non-Hispanic black women into prenatal care, and provide appropriate screening for risk factors for preterm birth to reduce the prevalence of low birth weight.”

ii. “Researchers should continue to investigate mechanisms that have been proposed for higher risks of low birth weight among non-Hispanic black infants, including less access to health care, higher rates of poor health behaviors, greater social stress, and intergenerational transmission of risk factors for low birth weight.”

f. Sidebar of study methods:

“Data are from a national survey conducted in the United States in 1988–1994. All information was collected at an interview with an adult in the child’s household. Low birth weight (LBW) is defined as less than 2,500 grams or 5.5 lbs (World Health Organization).”

3. Figure 13D is a diagram of the layout for a poster about the birth weight study, referring to figures or boxes in *Writing about Numbers, 2nd Edition*, from which the poster components are to be drawn.

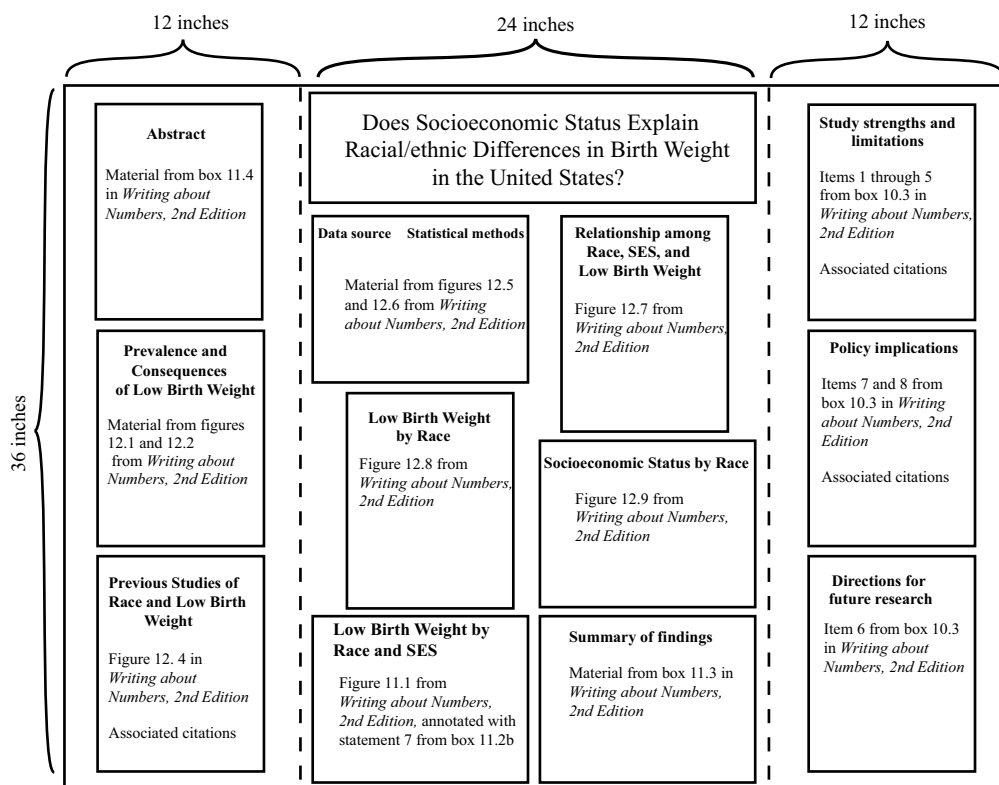


Figure 13D

5. Chartbook pages to present information on patterns of student aid in the United States to an applied audience.

Total Student Aid and Funds Used to Finance the Postsecondary Education Expenses of Undergraduates, billions of \$, US, 2010-11 Academic Year

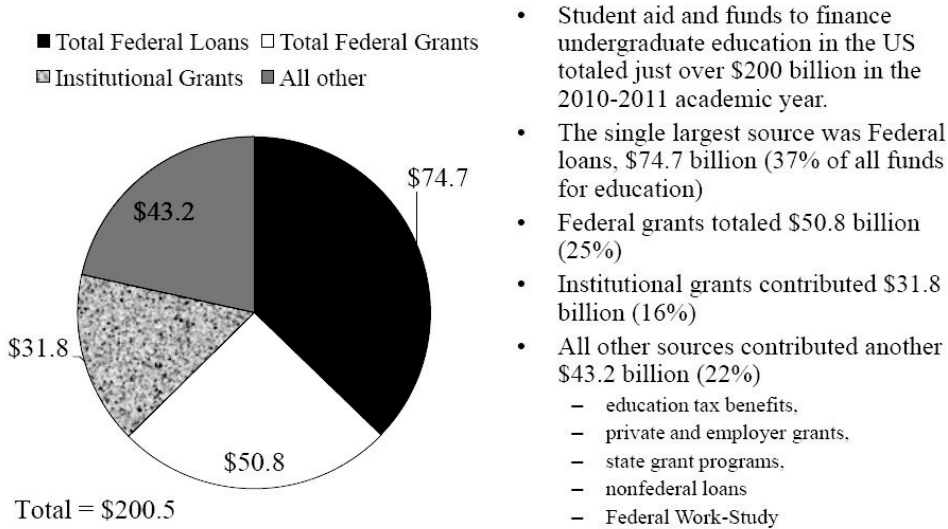


Figure 13E

Total Student Aid and Funds Used to Finance the Postsecondary Education Expenses of Undergraduates Academic Years 1990-91 to 2010-11

- Between 1990-91 and 2010-11, total student aid and funds to finance undergraduate education expenses quintupled, from \$42 billion to \$201 billion adjusted for inflation.

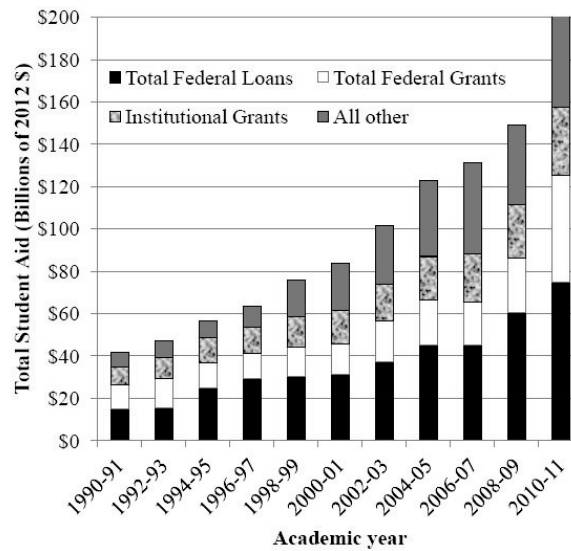
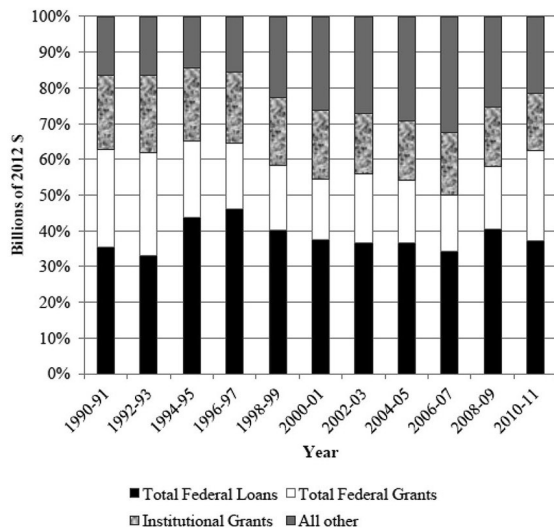


Figure 13F

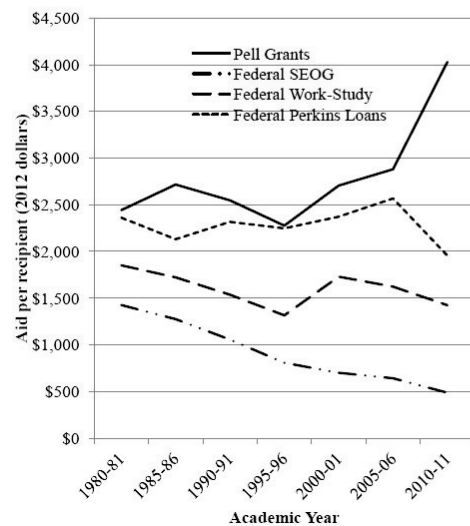
Shares of Total Student Aid and Funds Used to Finance the Postsecondary Education Expenses of Undergraduates by Major Source, Academic Years 1990-91 to 2010-11



- Throughout that period, Federal loans remained the single largest source of financing
 - 33% to 46% of all funds
- Between 1990 and 2006, Federal grants declined from 29% to 16% of all funds before rising to 25% in 2010-11.
- The share from all other sources doubled between 1990 (17%) and 2006 (33%) before declining to 22% in 2010-11.
- Institutional grants contributed roughly one-fifth of all education funds throughout the period.

Figure 13G

Average Aid per Recipient for Selected Sources of Federal Grants, 1980-81 to 2010-11



- Between 1980-81 and 2010-11, Pell grants were the single largest Federal source of educational aid, followed by Perkins loans, Federal Work-Study, and Federal Supplemental Educational Opportunity Grants (SEOG).
 - E.g., in 1980-81, Pell grants and Perkins loans each averaged about \$2,400 per aid recipient, while Federal Work-Study and SEOG averaged 76% and 58% of that amount, respectively.
- Pell grants increased nearly 40% between the 2005-6 and 2010-11 academic years, from \$2,880 to \$4,028 per aid recipient in constant 2012 dollars.
- The gap between Pell grants and each of the three other Federal sources widened considerably. As a percentage of Pell Grant aid, by 2010-11 average aid from these sources declined to:
 - Average Perkins loan aid per recipient: 49%
 - Work-Study: 35%
 - Supplemental Educational Opportunity Grants: 12%

Figure 13H

Total Amount Borrowed by 2009 by Students Beginning Postsecondary Education in 2003-04, US

- Of US students who began their postsecondary education during the 2003-4 academic year, within five years, 57% borrowed money to pay for their education.
- Of those who borrowed,
 - About 4 out of 10 had accumulated a total educational debt of less than \$10,000
 - 28% had debt between \$10,000 and \$20,000
 - 26% had a debt of \$20,000 or more.

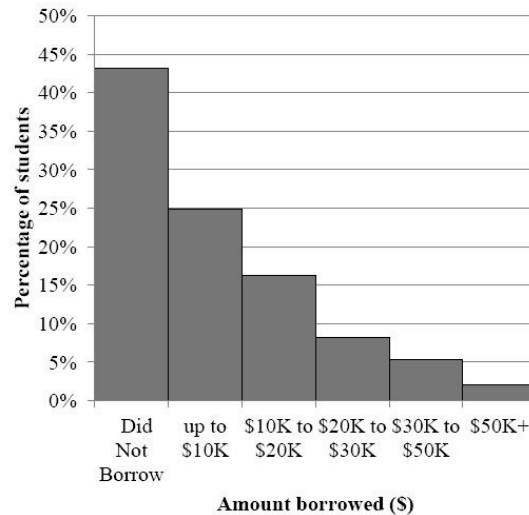


Figure 13I

7. An executive summary for the report on student aid.
 - This report documents grant aid from US federal and state governments, colleges and universities, employers, and other private sources, as well as loans, tax benefits, and Federal Work-Study assistance. In addition to reporting 2012 funding levels and types, financial aid per recipient, and debt incurred by students as they pursue higher education, the report examines trends in those measures over the past 30 years.
 - Data are from the National Center on Educational Statistics.
 - Key findings:

Types of student aid

- Student aid and funds to finance undergraduate education in the US totaled just over \$200 billion in the 2010–11 academic year (Figure 13E).
- The single largest source was Federal loans: \$74.7 billion (37% of all funds for education).
- Federal grants constituted 25%.
- Institutional grants contributed 16%.
- All other sources, including education tax benefits, private and employer grants, state grant programs, nonfederal loans, and Federal Work-Study, contributed another 22%.

Trends in student aid

- Between the 1990–91 and 2010–11 academic years, total student aid and funds to finance undergraduate education expenses quintupled, from \$42 billion to \$201 billion in constant 2012 dollars (Figure 13F).
- Throughout that period, Federal loans remained the single largest source of financing, contributing 33% to 46% of all aid funds (Figure 13G).
- Between 1990 and 2006, Federal grants declined from 29% to 16% of aid funds before rising to 25% in 2010–11.
- The share from all other sources doubled between 1990 (17%) and 2006 (33%) before declining to 22% in 2010–11.
- Institutional grants contributed roughly one-fifth of all education aid funds throughout the period.

Average financial aid per recipient, by source

- Between 1980–81 and 2010–11, Pell grants were the single largest Federal source of educational aid, followed by Perkins loans, Federal Work-Study, and Federal Supplemental Educational Opportunity Grants (SEOG) (Figure 13H).
- Pell grants increased nearly 40% between the 2005–6 and 2010–11 academic years, from \$2,880 to \$4,028 per aid recipient in constant 2012 dollars.
- The gap between Pell grants and each of the three other Federal sources widened considerably. By 2010–11, average Perkins loan aid, Work-Study, and SEOG aid per recipient were 49%, 35%, and 12%, respectively, of Pell Grant levels.

Student debt

- Of US students who began their postsecondary education during the 2003–4 academic year, within five years, 57% had borrowed money to pay for their education (Figure 13I).
- Of those who borrowed,
- about 4 out of 10 had accumulated a total educational debt of less than \$10,000;
- 28% had debt between \$10,000 and \$20,000;
- 26% had a debt of \$20,000 or more.