

Complete Solutions Manual
to Accompany

**Introduction to Statistics
and Data Analysis**

6th Edition

Roxy Peck

California Polytechnic State
University, San Luis Obispo

Tom Short

West Chester University of
Pennsylvania

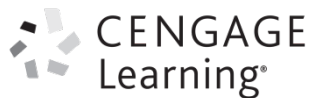
Chris Olsen

Grinnell College

Prepared by

Stephen Miller

Winchester Thurston
School, Pittsburgh, PA



© 2020 Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher except as may be permitted by the license terms below.

For product information and technology assistance, contact us at
Cengage Learning Customer & Sales Support,
1-800-354-9706.

For permission to use material from this text or product, submit
all requests online at **www.cengage.com/permissions**
Further permissions questions can be emailed to
permissionrequest@cengage.com.

ISBN-13: 978-1-337-79417-6
ISBN-10: 1-337-79417-1

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at: **www.cengage.com/global**.

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions or to purchase any of our products at our preferred online store, visit **www.cengage.com**.

NOTE: UNDER NO CIRCUMSTANCES MAY THIS MATERIAL OR ANY PORTION THEREOF BE SOLD, LICENSED, AUCTIONED, OR OTHERWISE REDISTRIBUTED EXCEPT AS MAY BE PERMITTED BY THE LICENSE TERMS HEREIN.

READ IMPORTANT LICENSE INFORMATION

Dear Professor or Other Supplement Recipient:

Cengage Learning has provided you with this product (the "Supplement") for your review and, to the extent that you adopt the associated textbook for use in connection with your course (the "Course"), you and your students who purchase the textbook may use the Supplement as described below. Cengage Learning has established these use limitations in response to concerns raised by authors, professors, and other users regarding the pedagogical problems stemming from unlimited distribution of Supplements.

Cengage Learning hereby grants you a nontransferable license to use the Supplement in connection with the Course, subject to the following conditions. The Supplement is for your personal, noncommercial use only and may not be reproduced, or distributed, except that portions of the Supplement may be provided to your students in connection with your instruction of the Course, so long as such students are advised that they may not copy or distribute any portion of the Supplement to any third party. Test banks, and other testing materials may be made available in the classroom and collected at the end of each class session, or posted electronically as described herein. Any

material posted electronically must be through a password-protected site, with all copy and download functionality disabled, and accessible solely by your students who have purchased the associated textbook for the Course. You may not sell, license, auction, or otherwise redistribute the Supplement in any form. We ask that you take reasonable steps to protect the Supplement from unauthorized use, reproduction, or distribution. Your use of the Supplement indicates your acceptance of the conditions set forth in this Agreement. If you do not accept these conditions, you must return the Supplement unused within 30 days of receipt.

All rights (including without limitation, copyrights, patents, and trade secrets) in the Supplement are and will remain the sole and exclusive property of Cengage Learning and/or its licensors. The Supplement is furnished by Cengage Learning on an "as is" basis without any warranties, express or implied. This Agreement will be governed by and construed pursuant to the laws of the State of New York, without regard to such State's conflict of law rules.

Thank you for your assistance in helping to safeguard the integrity of the content contained in this Supplement. We trust you find the Supplement a useful teaching tool.

Contents

Chapter 1: The Role of Statistics and the Data Analysis Process	1
Chapter 2: Collecting Data Sensibly	14
Chapter 3: Graphical Methods for Describing Data	28
Chapter 4: Numerical Methods for Describing Data	70
Chapter 5: Summarizing Bivariate Data	88
Chapter 6: Probability	136
Chapter 7: Random Variables and Probability Distributions.....	167
Chapter 8: Sampling Variability and Sampling Distributions	215
Chapter 9: Estimation Using a Single Sample	226
Chapter 10: Hypothesis Testing Using a Single Sample	248
Chapter 11: Comparing Two Populations or Treatments	287
Chapter 12: The Analysis of Categorical Data and Goodness-of-Fit Tests	343
Chapter 13: Simple Linear Regression and Correlation: Inferential Methods.....	361
Chapter 14: Multiple Regression Analysis	410
Chapter 15: Analysis of Variance	454
Chapter 16: Nonparametric (Distribution-Free) Statistical Methods.....	488

Chapter 1

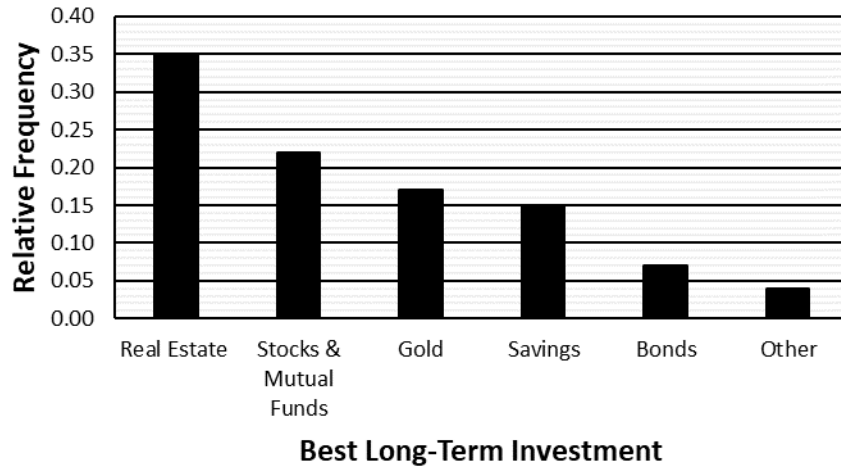
The Role of Statistics and the Data Analysis Process

- 1.1** *Descriptive statistics* is the branch of statistics that involves the organization and summary of the values in a data set. *Inferential statistics* is the branch of statistics concerned with reaching conclusions about a population based on the information provided by a sample.
- 1.2** The *population* is the entire collection of individuals or objects about which information is required. A *sample* is a subset of the population selected for study in some prescribed manner.
- 1.3** The proportions are stated as population values (although they were very likely calculated from sample results).
- 1.4** The sample is the set of 2121 children used in the study. The population is the set of all children between the ages of one and four.
- 1.5**
- a** The population of interest is the set of all 15,000 students at the university.
 - b** The sample is the 200 students who are interviewed.
- 1.6** The estimates given were computed using data from a sample.
- 1.7** The population is the set of all 7000 property owners. The sample is the 500 owners included in the survey.
- 1.8** The population is the set of all 2019 Toyota Camrys. The sample is the set of six cars that are tested.
- 1.9** The population is the set of 5000 used bricks. The sample is the set of 100 bricks she checks.
- 1.10**
- a** The researchers wanted to know whether the new surgical approach would improve memory functioning in Alzheimer's patients. They hoped that the negative effects of the disease could be reduced by toxins being drained from the fluid filled space that cushions the brain.
 - b** First, it is not stated that the patients were randomly assigned to the treatments (new approach and standard care); this would be necessary in a well designed study. Second, it would help if the experiment could have been designed so that the patients did not know whether they were receiving the new approach or the standard care; otherwise, it is possible that the patients' knowledge that they were receiving a new treatment might in itself have brought about an improvement in memory. Third, as stated in the investigators' conclusion, it would have been useful if the experiment had been conducted on a sufficient number of patients so that any difference observed between the two treatments could not have been attributed to chance.
- 1.11**
- a** The researchers wanted to find out whether taking a garlic supplement reduces the likelihood that you will get a cold. They wanted to know whether a significantly lower proportion of people who took a garlic supplement would get a cold than those who did not take a garlic supplement.

- b** It is necessary that the participants were *randomly* assigned to the treatment groups. If this was the case, it seems that the study was conducted in a reasonable way.
- 1.12**
- a** Numerical (discrete)
 - b** Categorical
 - c** Numerical (continuous)
 - d** Numerical (continuous)
 - e** Categorical
- 1.13**
- a** Categorical
 - b** Categorical
 - c** Numerical (discrete)
 - d** Numerical (continuous)
 - e** Categorical
 - f** Numerical (continuous)
- 1.14**
- a** Discrete
 - b** Continuous
 - c** Discrete
 - d** Discrete
- 1.15**
- a** Continuous
 - b** Continuous
 - c** Continuous
 - d** Discrete
- 1.16** For example:
- a** Ford, Toyota, Ford, General Motors, Chevrolet, Chevrolet, Honda, BMW, Subaru, Nissan.
 - b** 3.23, 2.92, 4.0, 2.8, 2.1, 3.88, 3.33, 3.9, 2.3, 3.56, 3.32, 2.4, 2.8, 3.9, 3.12.
 - c** 4, 2, 0, 6, 3, 3, 2, 4, 5, 0, 8, 2, 5, 3, 4, 7, 3, 2, 0, 1
 - d** 50.27, 50.67, 48.98, 50.58, 50.95, 50.95, 50.21, 49.70, 50.33, 49.14, 50.83, 49.89
 - e** In minutes: 10, 10, 18, 0, 17, 17, 0, 17, 12, 19, 12, 13, 15, 15, 15

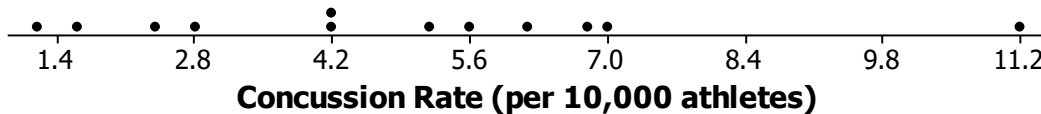
- 1.17** a Gender of purchaser, brand of motorcycle, telephone area code
 b Number of previous motorcycles
 c Bar chart
 d Dotplot

1.18 a

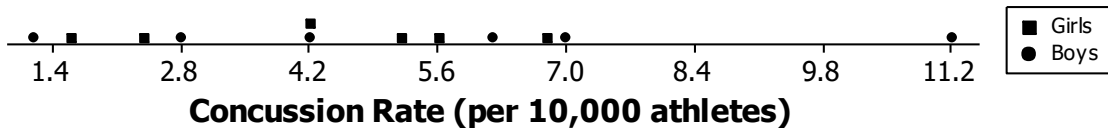


b Over half of the responses (57%) were from people who indicated that the best long-term investments were real estate (35%) and stocks & mutual funds (22%). The remaining 43% of respondents indicated that gold (17%), savings (15%), bonds (7%), and other (4%) were the best long-term investments.

1.19 a The dotplot below shows the concussion rate (concussions per 10,000 athletes).

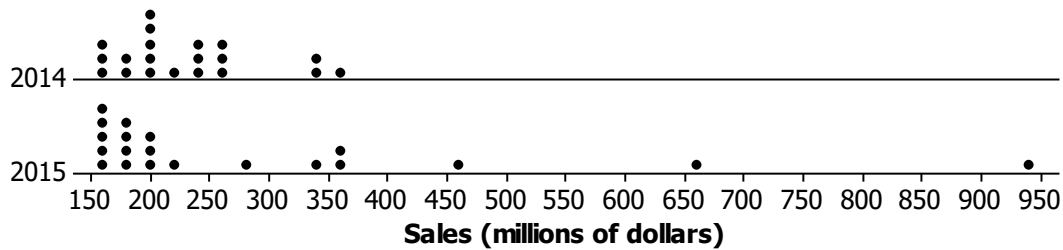


b The dotplot below shows the concussion rate (concussions per 10,000 athletes), with different symbols for boys and girls.



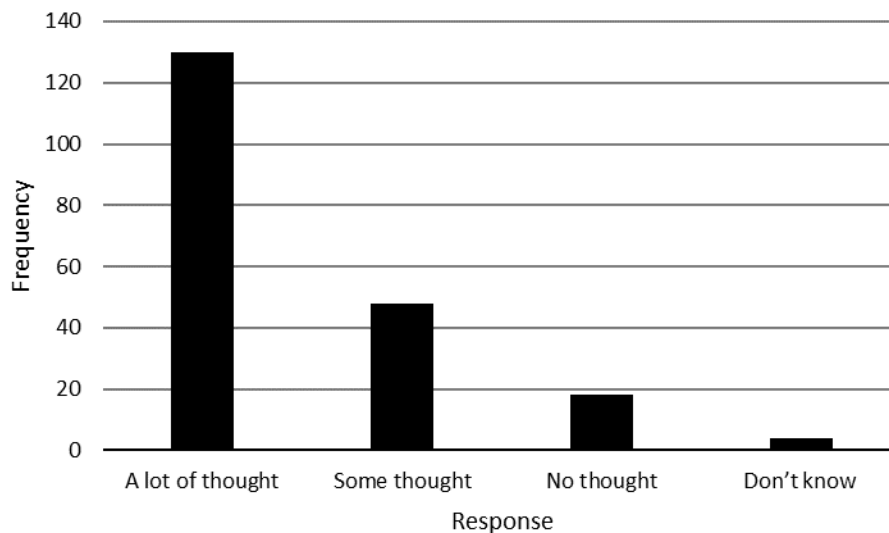
The sport with an unusually high (compared to all the other sports) concussion rate is football. Without considering football, the concussion rates for girls' sports is essentially the same as the concussion rate for boys' sports. However, if we consider football, the concussion rate for girls' sports tends to be lower than that for boys' sports.

- 1.20** Dotplots (for parts **a** and **b**) drawn on the same scale for the 2014 and 2015 sales data are shown below.



- a** A typical sales figure for 2014 was around 215 million dollars, with sales figures ranging from around 150 to around 350 million dollars. The greatest density of points was at the lower end of the distribution. There were no extreme observations in 2014.
- b** A typical sales figure for 2015 was around 200 million dollars, with sales figures ranging from around 155 to around 937 million dollars. The greatest density of points was at the lower end of the distribution. There are two extreme observations in 2015, namely, 936.7 million dollars and 652.3 million dollars.
- c** Sales figures were generally speaking higher in 2015 than in 2014. There were two extreme observations in 2015, and no extreme observations in 2014. If the extreme sales figures are taken into account, the variation in the sales figures (among the top 20 movies) was far greater in 2015 than in 2014. If the extreme sales figures are disregarded, the variation was still greater in 2015 than in 2014, but by not nearly as much. The distributions are similar in shape, with the greatest density of points being at the lower end of the distribution in both cases.

- 1.21 a**

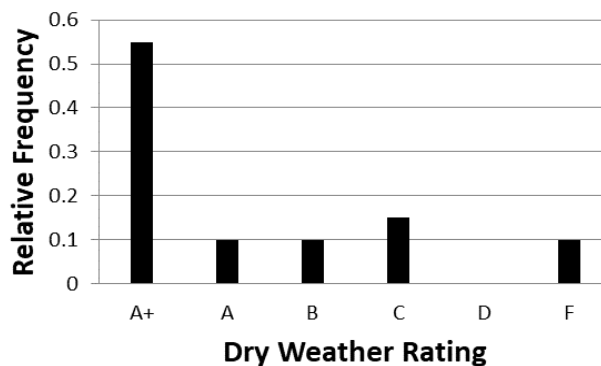


- b** The most common response was “A lot of thought”, accounting for 130 (or 65%) of the students who started college but did not complete a degree. The next two most common responses were “Some thought” and “No thought”, accounting for 48 (or 24%) and 18 (or 9%), respectively, of the students who started college but did not complete a degree. Finally,

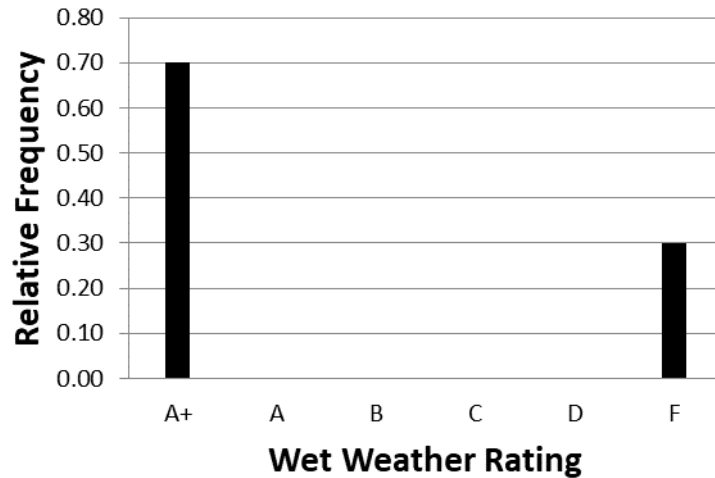
- 4 of the 200 respondents (2%) indicated that don't know how much thought they have given to going back to school.
- 1.22**
- a** Categorical
 - b** Since the variable being graphed is categorical, a dotplot would not be suitable.
 - c** If you add up the relative frequencies you get 107%. This total should be 100%, so a mistake has clearly been made.
- 1.23**
- a** The dotplot shows that there were three sites that received far greater numbers of visits than the remaining 6 sites. Also, it shows that the distribution of the number of visits has the greatest density of points for the smaller numbers of visits, with the density decreasing as the number of visits increases.
 - b** It is clear from the dotplot that there were two sites that were used by far greater numbers of individuals (unique visitors) than the remaining 7 sites. However, these two sites are less far above the others in terms of the number of unique visitors than they are in terms of the total number of visits. As with the distribution of the total number of visits, the distribution of the number of unique visitors has the greatest density of points for the smaller numbers of visitors, with the density decreasing as the number of unique visitors increases. This is the case even when only the 7 less popular sites are considered.
 - c** The statistic “visits per unique visitor” tells us how heavily the individuals are using the sites. The table tells us that the most popular sites (Facebook and YouTube) in terms of total visits and unique visitors do not have the highest value of this statistic. The dotplot of visits per unique visitor shows that there are two individual sites are far ahead of the rest in this respect (Pinterest and Twitter).

1.24

Rating	Relative Frequency
A+	$11/20 = 0.55$
A	$2/20 = 0.10$
B	$2/20 = 0.10$
C	$3/20 = 0.15$
D	$0/20 = 0.00$
F	$2/20 = 0.10$

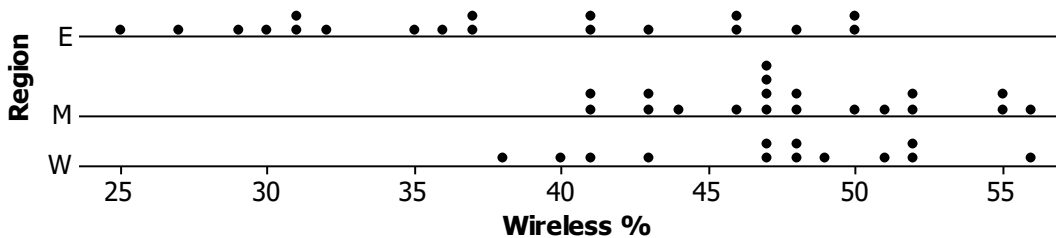


1.25 a



b Seventy-five percent (75%) of the dry weather ratings are B or higher, and 70% of wet weather ratings are B or higher, indicating that dry weather ratings are higher than wet weather ratings. Note that the wet weather ratings are only A+ or F, so the wet weather ratings are more extreme than dry weather ratings. If we only consider A+ ratings, then the wet weather ratings tend to be better than dry weather ratings because only 55% of dry weather ratings are A+, compared with 70% of wet weather ratings being A+.

1.26 a



b Looking at the dotplot we can see that Eastern states have, on average, lower wireless percents than states in the other two regions. The West and Middle states regions have, on average, roughly equal wireless percents.

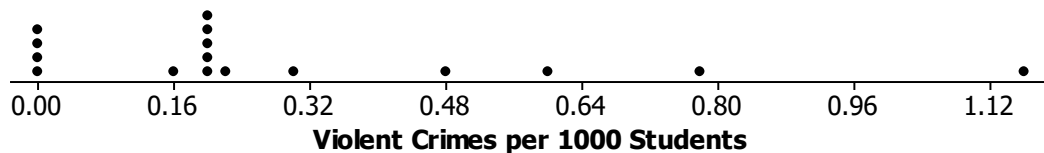
1.27 a



Three schools seem to stand out from the rest, these being, in increasing order of number of crimes, University of Central Florida (14 crimes reported), Florida International University (15 crimes reported), and Florida State University (20 crimes reported).

b

University/College	Violent Crime Rate Per 1000 Students
Florida A&M University	0.60435
Florida Atlantic University	0.19750
Florida Gulf Coast University	0.20225
Florida International University	0.30131
Florida South Western State College	0.00000
Florida State University, Tallahassee	0.48984
New College of Florida	1.16144
Pensacola State College	0.00000
Santa Fe College	0.00000
Tallahassee Community College	0.16071
University of Central Florida	0.22239
University of Florida	0.19745
University of North Florida	0.19139
University of South Florida, St. Petersburg	0.00000
University of South Florida, Tampa	0.19017
University of West Florida	0.78351



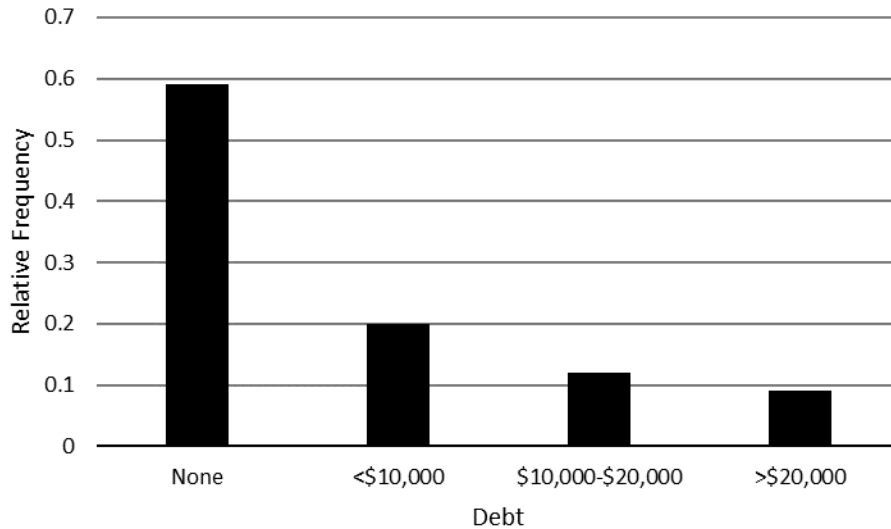
The colleges that stand out in violent crimes per 1000 students are, in increasing order of crime rate, Florida State University, Florida A&M University, University of West Florida, and New College of Florida. Only Florida State University stands out in both dotplots.

- c For the number of violent crimes, there are three schools that stand out by having high numbers of crimes, with the majority of the schools having similar, and low (10 or fewer), numbers of crimes. There seems to be greater consistency for crime rate (per 1000 students) among the 16 schools than there is for number of crimes, with four schools standing out as having high crime rates, and four schools with crime rates that stand out as being noticeably low.

- 1.28 a When ranking the airlines according to delayed flights, one airline would be ranked above another if the probability of a randomly chosen flight being delayed is smaller for the first airline than it is for the second airline. These probabilities are estimated using the *rate per 10,000 flights* values, and so these are the data that should be used for this ranking. (Note that the *total number of flights* values are not suitable for this ranking. Suppose that one airline had a larger number of delayed flights than another airline. It is possible that this could be accounted for merely through the first airline having more flights than the second.)
- b There are two airlines, ExpressJet and Continental, which, with 4.9 and 4.1 of every 10,000 flights delayed, stand out as the worst airlines in this regard. There are two further airlines that stand out above the rest: Delta and Comair, with rates of 2.8 and 2.7 delayed flights per

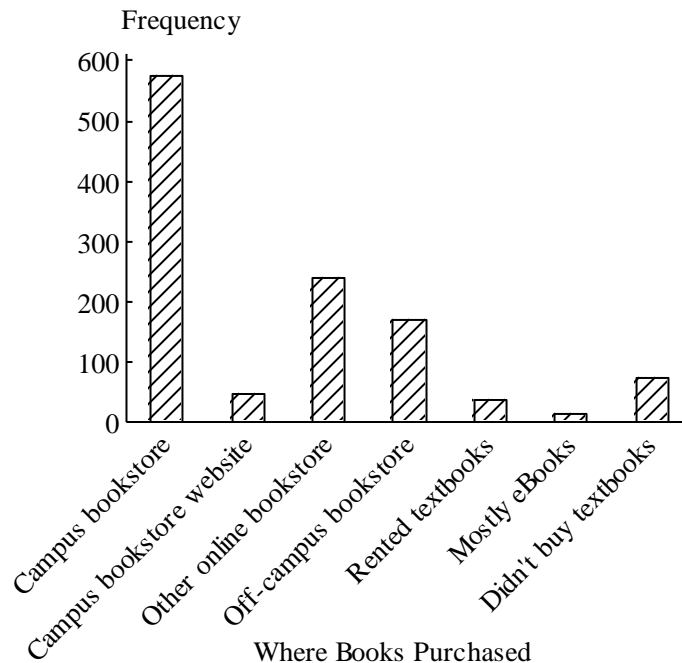
10,000 flights. All the other airlines have rates below 1.6, with the best rating being for Southwest, with a rate of only 0.1 delayed flights per 10,000.

1.29 a



b Most public community college graduates have no debt at all, and a debt of \$10,000 or less accounts for 79% of the graduates. Among the 21% of the graduates who have a debt of more than \$10,000, nearly 43% (9% of all graduates) have a debt of more than \$20,000.

1.30 a



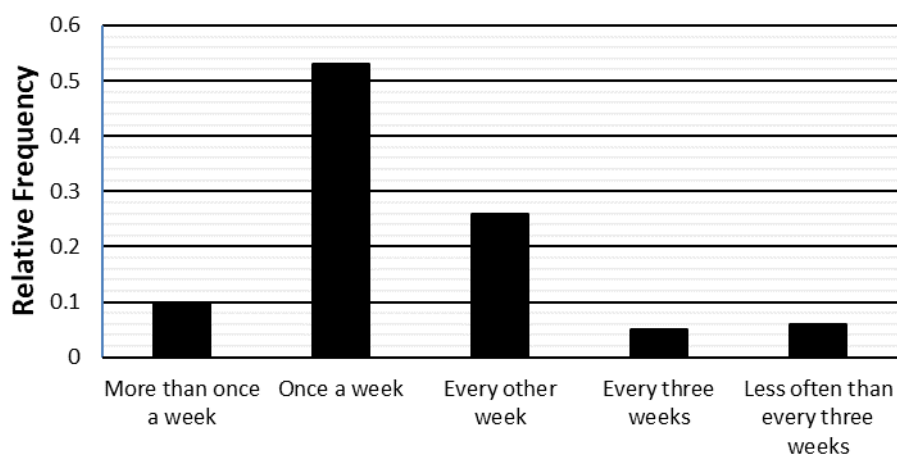
b By far the most popular place to buy books is the campus bookstore, with half of the students in the sample buying their books from that source. The next most popular sources are online bookstores other than the online version of the campus bookstore and off-campus bookstores,

with these two sources accounting for around 35% of students. Purchasing mostly eBooks was the least common response.

1.31 a

How Often	Relative Frequency
More than once a week	0.10
Once a week	0.53
Every other week	0.26
Every three weeks	0.05
Less often than every three weeks	0.06

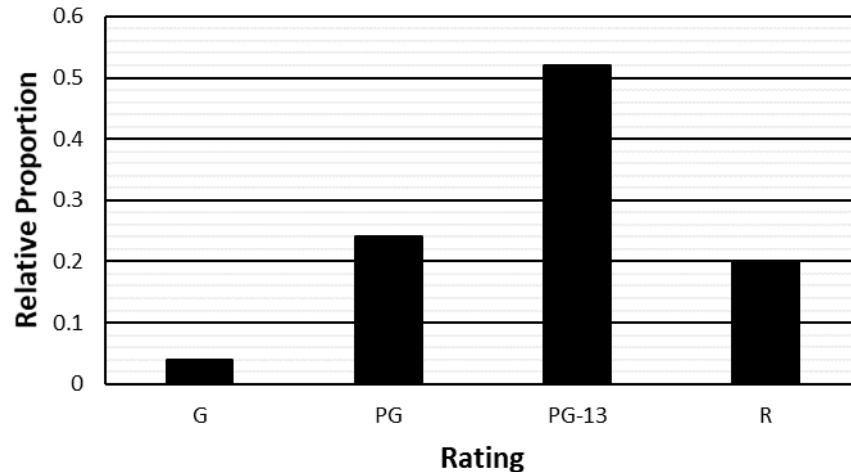
b



How often do you change the sheets?

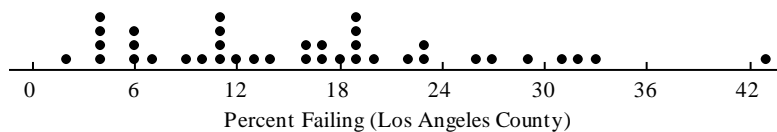
1.32 The relative frequency distribution is:

Rating	Relative Frequency
G	$1/25 = 0.04$
PG	$6/25 = 0.24$
PG-13	$13/25 = 0.52$
R	$5/25 = 0.20$



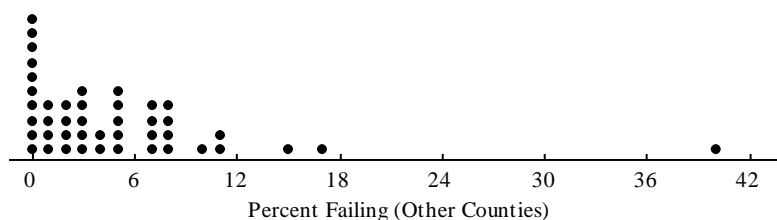
PG-13 is the rating with the highest relative proportion (0.52), followed by PG (0.24), R (0.20), and G (0.04). Seventy-two percent (72%) of the top 25 movies of 2015 are PG-13 or R, and the remaining 28% are rated G or PG.

1.33 a The dotplot for Los Angeles County is shown below.



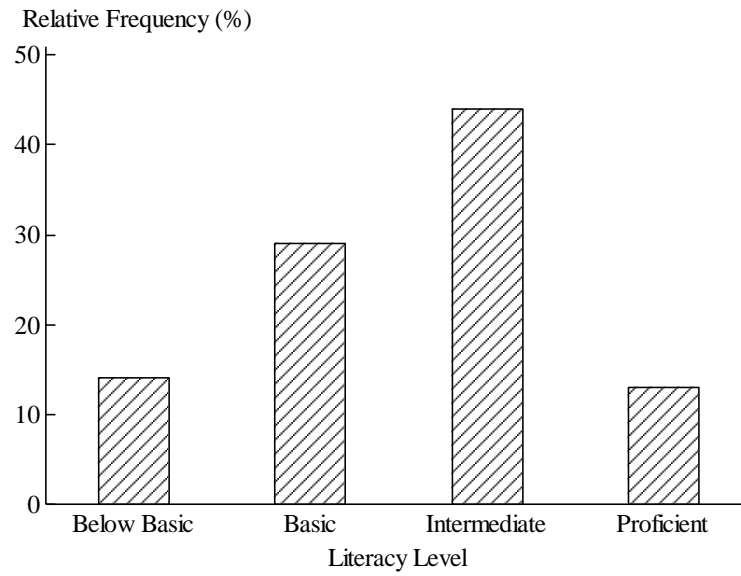
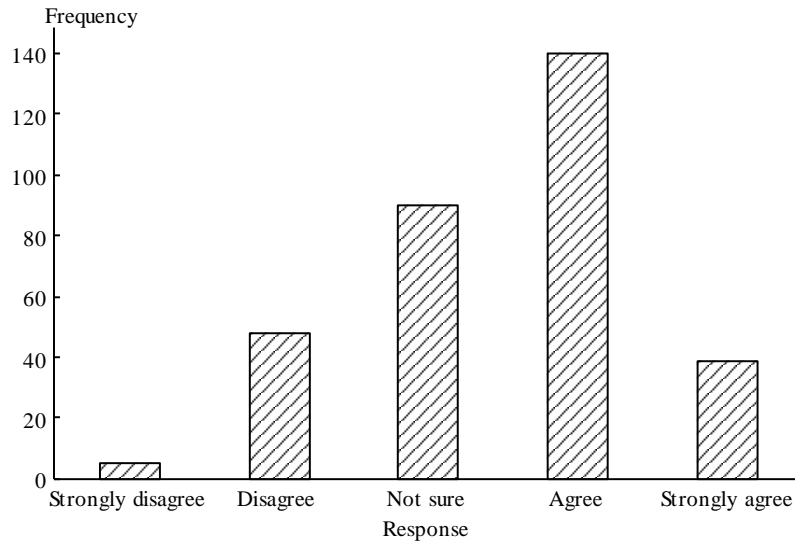
A typical percent of tests failing for Los Angeles County is around 16. There is one value that is unusually high (43), with the other values ranging from 2 to 33. There is a greater density of points toward the lower end of the distribution than toward the upper end.

b The dotplot for the other counties is shown below.

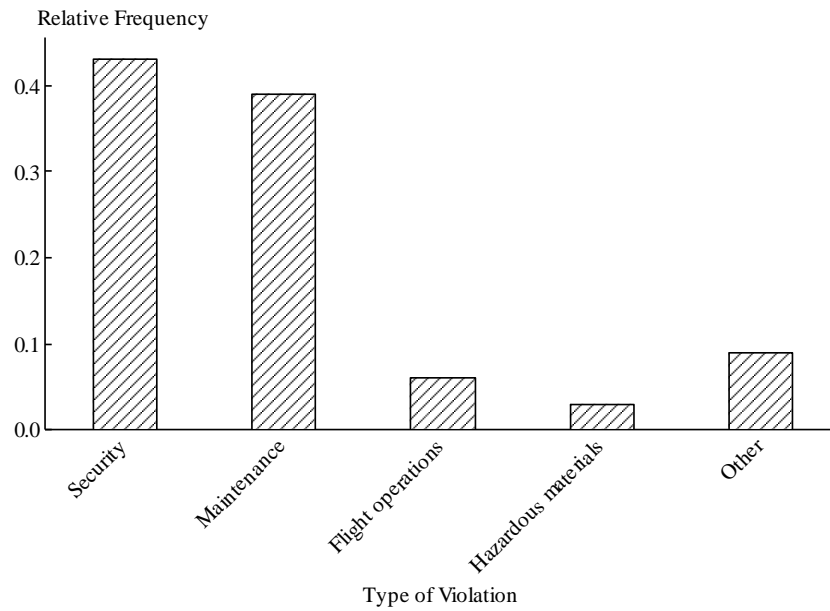


A typical percent of tests failing for the other counties is around 3. There is one extreme result at the upper end of the distribution (40); the other values range from 0 to 17. The density of points is highest at the left hand end of the distribution and decreases as the percent failing values increase.

c The typical value for Los Angeles County (around 16) is greater than for the other counties (around 3) and, disregarding the one extreme value in each case, there is a greater variability in the values for Los Angeles County than for the other counties. In the distribution for Los Angeles County the points are closer to being uniformly distributed than in the distribution for the other counties, where there is a clear tail-off of density of points as you move to the right of the distribution.

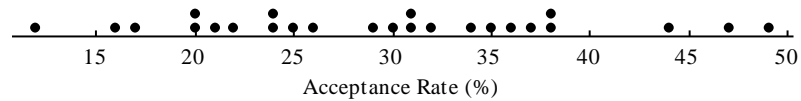
1.34 a Categorical**b****c** No, since dotplots are used for numerical data.**1.35**

1.36 a



- b** By far the most frequently occurring violation categories were security (43%) and maintenance (39%). The least frequently occurring violation categories were flight operations (6%) and hazardous materials (3%).

1.37 a



- b** A typical acceptance rate for these top 25 schools is around 30, with the great majority of acceptance rates being between 19 and 39. There are no particularly extreme values. The pattern of the points is roughly symmetrical.

1.38

