

Complete Solutions Manual to Accompany

Introduction to Statistics & Data Analysis

FIFTH EDITION

Roxy Peck

California Polytechnic State University,
San Luis Obispo, CA

Chris Olsen

Grinnell College
Grinnell, IA

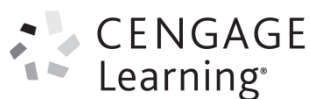
Jay Devore

California Polytechnic State University,
San Luis Obispo, CA

Prepared by

Michael Allwood

Brunswick School, Greenwich, CT



© 2016 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: 978-130526581-3
ISBN-10: 1-305-26581-5

Cengage Learning
20 Channel Center Street
Fourth Floor
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,
visit **www.cengage.com**.

Purchase any of our products at your local college
store or at our preferred online store
www.cengagebrain.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, posted electronically or distributed, except that portions of the Supplement may be provided to your students IN PRINT FORM ONLY 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. 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.

Table of 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.....	29
Chapter 4	Numerical Methods for Describing Data.....	79
Chapter 5	Summarizing Bivariate Data.....	98
Chapter 6	Probability.....	148
Chapter 7	Random Variables and Probability Distributions.....	179
Chapter 8	Sampling Variability and Sampling Distributions.....	229
Chapter 9	Estimation Using a Single Sample.....	241
Chapter 10	Hypothesis Testing Using a Single Sample.....	261
Chapter 11	Comparing Two Populations or Treatments.....	296
Chapter 12	The Analysis of Categorical Data and Goodness-of-Fit Tests.....	353
Chapter 13	Simple Linear Regression and Correlation: Inferential Methods.....	376
Chapter 14	Multiple Regression Analysis.....	424
Chapter 15	Analysis of Variance.....	463
Chapter 16	Nonparametric (Distribution-Free) Statistical Methods.....	496

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 2014 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

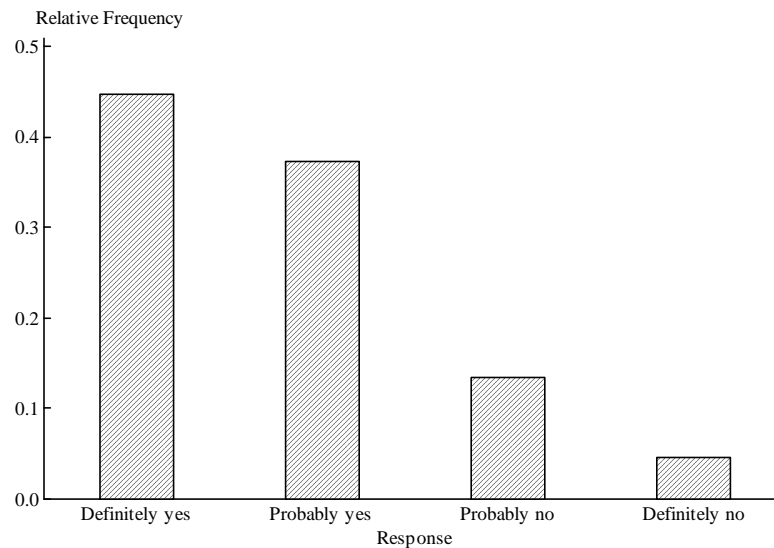
1.17 a Gender of purchaser, brand of motorcycle, telephone area code

b Number of previous motorcycles

c Bar chart

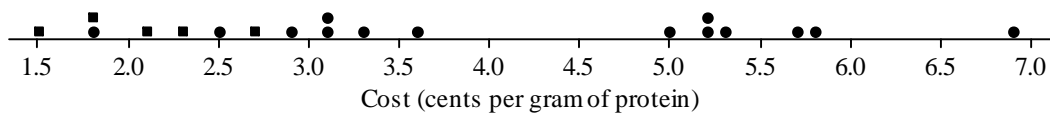
d Dotplot

1.18 a



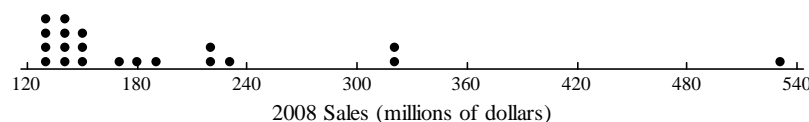
b “Large Majority of Seniors Say They’d Choose the Same College Again”

1.19 a

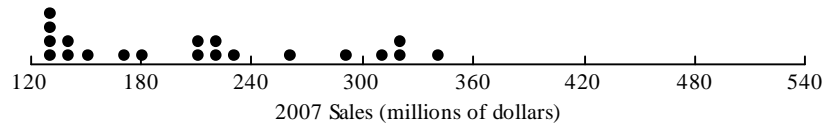


b The costs per gram of protein for the meat and poultry items are represented by squares in the dotplot above. With every one of the meat and poultry items included in the lowest seven cost per gram values, meat and poultry items appear to be relatively low cost sources of protein.

1.20 a

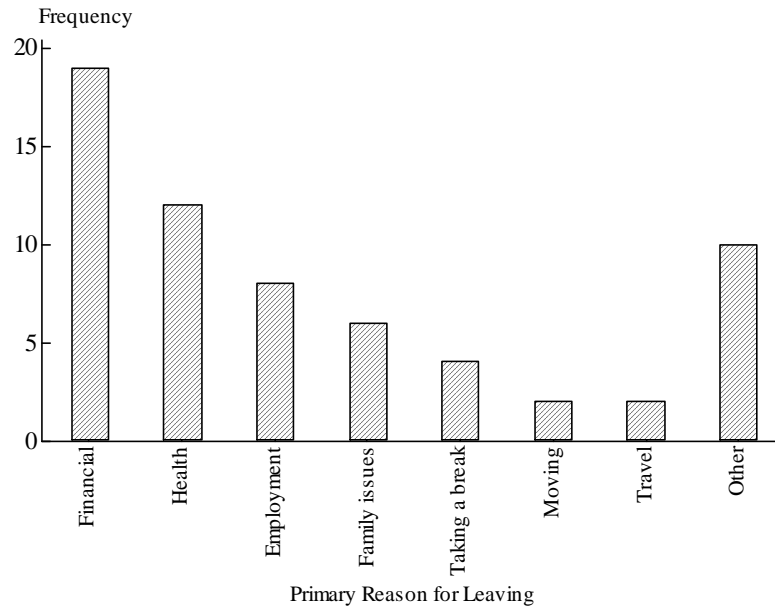


A typical sales figure for 2008 was around 150 million dollars. There is one extreme result at the upper end of the distribution. If this point is disregarded then the values range from 127.5 to 318.4. The greatest density of points is at the lower end of the distribution.

b

A typical sales figure for 2007 was around 210 million dollars, with sales figures ranging from around 128 to around 337 million dollars. The greatest density of points was at the lower end of the distribution. There were no extreme results in 2007.

- c** Sales figures were generally speaking higher in 2007 than in 2008. There was one extreme result in 2008, and no extreme result in 2007. If the extreme sales figure is taken into account, the variation in the sales figures (among the top 20 movies) was far greater in 2008 than in 2007. However, if the extreme result is disregarded, the variation was greater in 2007. 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 reason was financial, this accounting for 30.2% of students who left for non-academic reasons. The next two most common reasons were health and other personal reasons, these accounting for 19.0% and 15.9%, respectively, of the students who left for non-academic reasons.

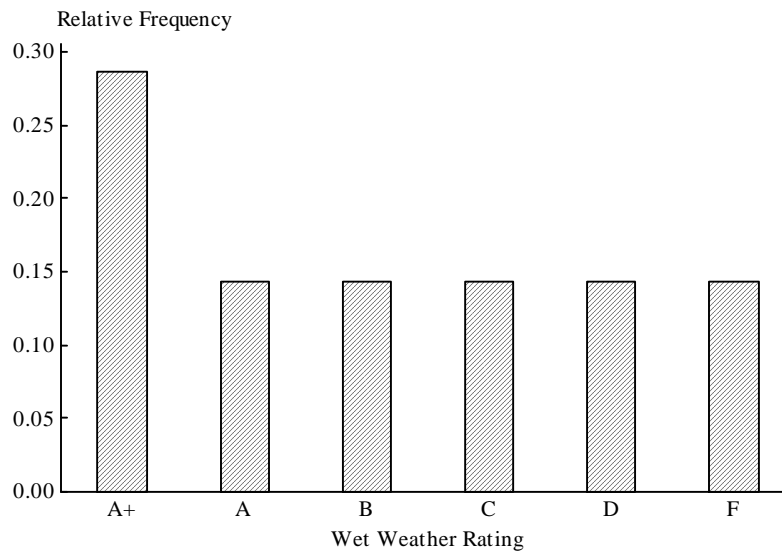
1.22 a Categorical

- 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 two sites that received far greater numbers of visits than the remaining 23 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. This is the case even when only the 23 less popular sites are considered.
- b** Again, 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 23 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 23 less popular sites are considered.
- c** The statistic “visits per unique visitor” tells us how heavily the individuals are using the sites. Although the table tells us that the most popular site (Facebook) in terms of the other two statistics also has the highest value of this statistic, the dotplot of visits per unique visitor shows that no one or two individual sites are far ahead of the rest in this respect.
- 1.24 a** It would not be appropriate to use a dotplot because rating is a categorical variable.

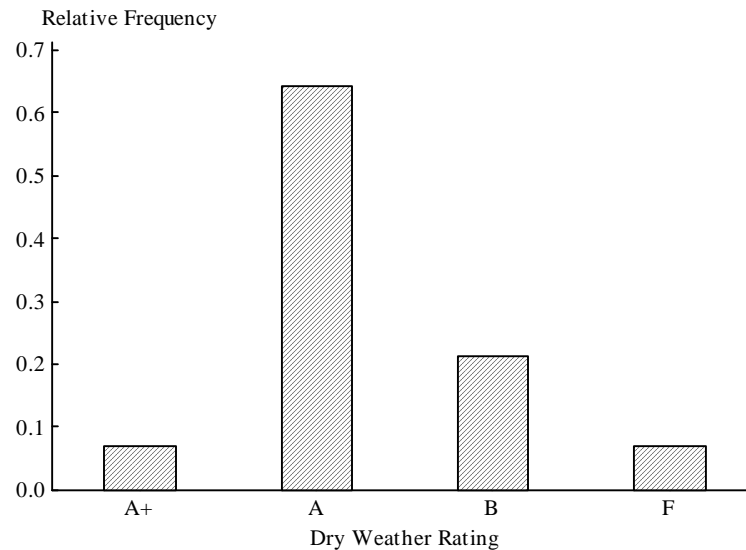
b

Wet Weather Rating	Frequency	Relative Frequency
A+	4	0.286
A	2	0.143
B	2	0.143
C	2	0.143
D	2	0.143
F	2	0.143



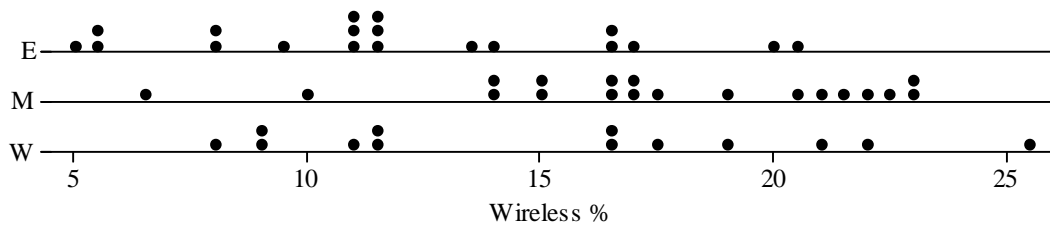
c

Dry Weather Rating	Frequency	Relative Frequency
A+	1	0.071
A	9	0.643
B	3	0.214
F	1	0.071



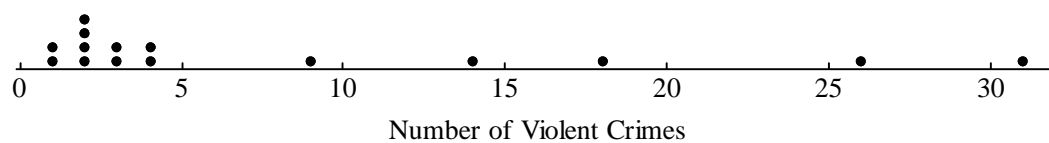
d Yes. Apart the greater proportion of “A+” ratings for wet weather than for dry weather, the beaches on the whole receive higher ratings in dry weather than in wet weather, with only 28.6% of beaches receiving below an A in dry weather, compared to 57.1% in wet weather.

1.25 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.

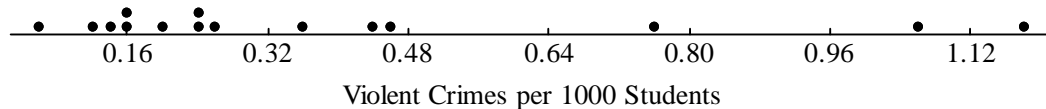
1.26 a



Five schools seem to stand out from the rest, these being, in increasing order of number of crimes, Florida International, Florida A&M, University of Florida, University of Central Florida, and Florida State University.

b

University/College	Violent Crime Rate Per 1000 Students
Edison State College	0.234
Florida A&M University	1.060
Florida Atlantic University	0.158
Florida Gulf Coast University	0.233
Florida International University	0.202
Florida State University	0.755
New College of Florida	1.183
Pensacola State College	0.260
Santa Fe College	0.065
Tallahassee Community College	0.133
University of Central Florida	0.445
University of Florida	0.363
University of North Florida	0.123
University of South Florida	0.464
University of West Florida	0.167



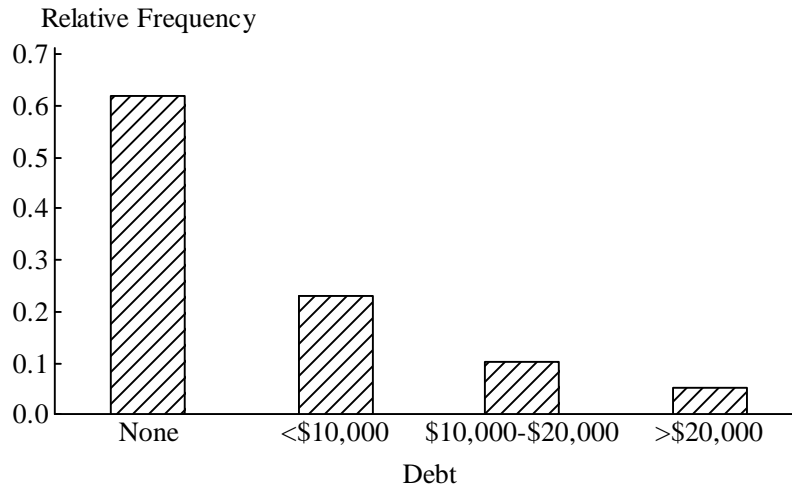
The colleges that stand out in violent crimes per 1000 students are, in increasing order of crime rate, Florida State University, Florida A&M University, and New College of Florida. Only Florida A&M stands out in both boxplots.

- c** For the number of violent crimes, there are five schools that stand out by having high numbers of crimes, with the majority of the schools having similar, and low, numbers of crimes. There seems to be greater consistency for crime rate (per 1000 students) among the 15 schools than there is for number of crimes, with just three schools standing out as having high crime rates, and no schools with crime rates that stand out as being low.

- 1.27 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

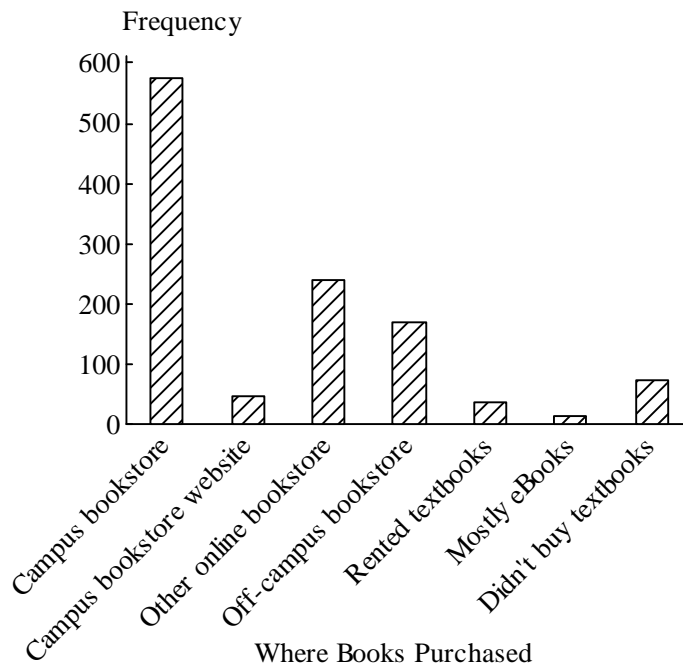
10,000 flights. All the other airlines have rates below 1.6, with the best rating being for Southwest, with a rate of only 0.1 delayed flights per 10,000.

1.28 a



b Most public community college graduates have no debt at all, and a debt of \$10,000 or less accounts for 85% of the graduates. Among the small minority (15%) of the graduates who have a debt of more than \$10,000, only one third (5% of all graduates) have a debt of more than \$20,000.

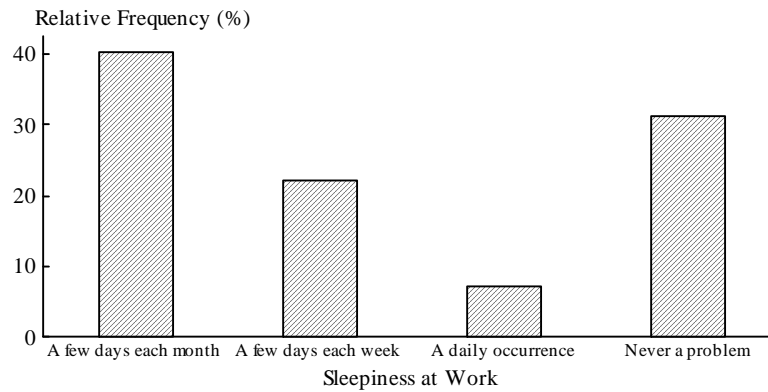
1.29 a



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.

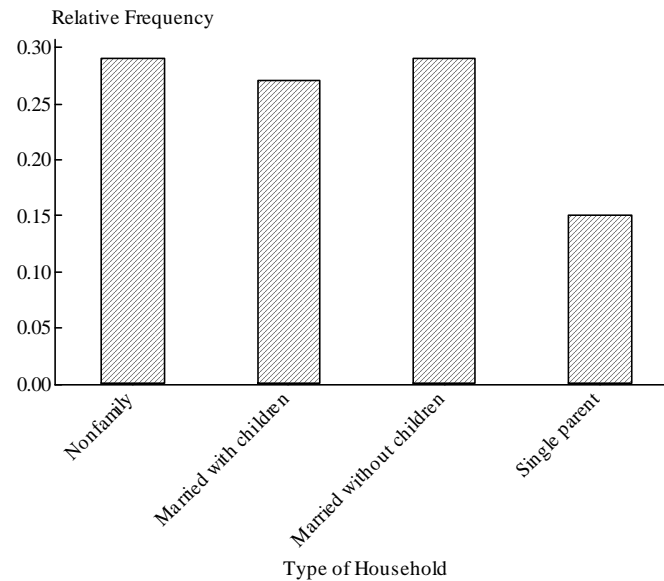
1.30

Sleepiness at Work	Relative Frequency (%)
A few days each month	40
A few days each week	22
A daily occurrence	7
Never a problem	31

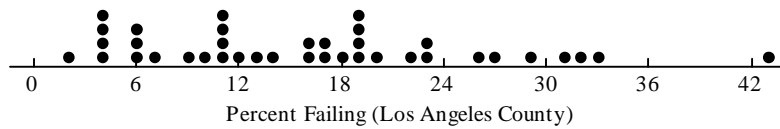


1.31

Type of Household	Relative Frequency
Nonfamily	0.29
Married with children	0.27
Married without children	0.29
Single parent	0.15

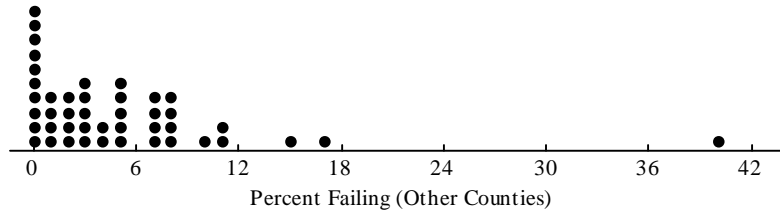


1.32 a The dotplot for Los Angeles County is shown below.



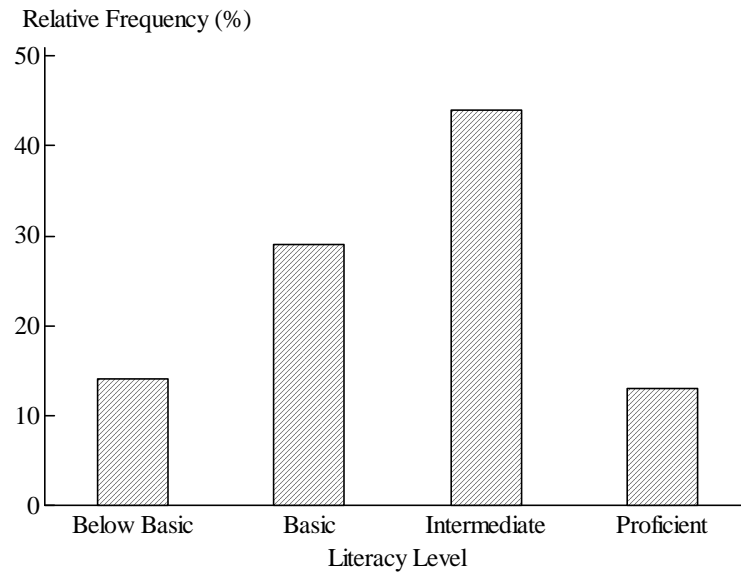
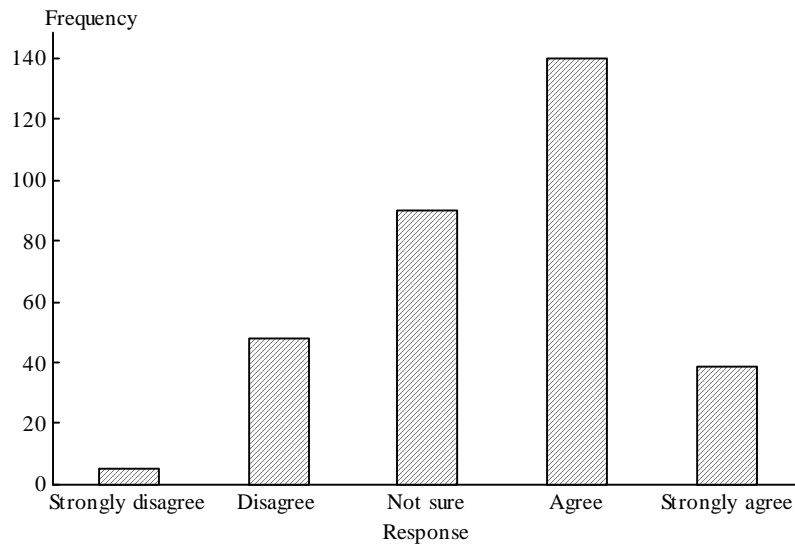
A typical percent of tests failing for Los Angeles County is around 16. There is one value that is unusually high (43), with the other values ranging from 2 to 33. There is a greater density of points toward the lower end of the distribution than toward the upper end.

- b** The dotplot for the other counties is shown below.

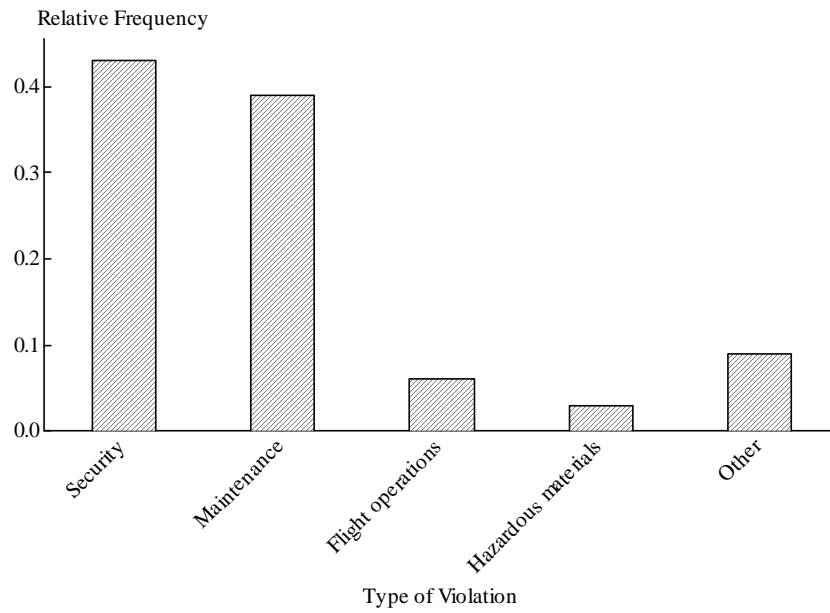


A typical percent of tests failing for the other counties is around 3. There is one extreme result at the upper end of the distribution (40); the other values range from 0 to 17. The density of points is highest at the left hand end of the distribution and decreases as the percent failing values increase.

- 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.

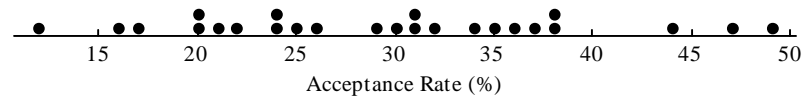
1.33 a Categorical**b****c** No, since dotplots are used for numerical data.**1.34**

1.35 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%).

1.36 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.

1.37

