

1.1 Analyzing Categorical Data

Read 2–4

What’s the difference between categorical and quantitative variables?

Do we ever use numbers to describe the values of a categorical variable? Do we ever divide the distribution of a quantitative variable into categories?

What is a distribution?

Alternate Example: US Census Data

Here is information about 10 randomly selected US residents from the 2000 census.

State	Number of Family Members	Age	Gender	Marital Status	Total Income	Travel time to work
Kentucky	2	61	Female	Married	21000	20
Florida	6	27	Female	Married	21300	20
Wisconsin	2	27	Male	Married	30000	5
California	4	33	Female	Married	26000	10
Michigan	3	49	Female	Married	15100	25
Virginia	3	26	Female	Married	25000	15
Pennsylvania	4	44	Male	Married	43000	10
Virginia	4	22	Male	Never married/ single	3000	0
California	1	30	Male	Never married/ single	40000	15
New York	4	34	Female	Separated	30000	40

- (a) Who are the individuals in this data set?
- (b) What variables are measured? Identify each as categorical or quantitative. In what units were the quantitative variables measured?
- (c) Describe the individual in the first row.

Read 8–12

What is the difference between a data table, a frequency table, and a relative frequency table? When is it better to use relative frequency?

What is the most important thing to remember when making pie charts and bar graphs? Why do statisticians prefer bar graphs?

When is it inappropriate to use a pie chart?

What are some common ways to make a misleading graph?

What is wrong with the following graph?



HW: page 7 (1, 3, 5, 7, 8), page 22 (11, 17, 18)

1.1 Analyzing Categorical Data

Read 12–19

What is a two-way table? What is a marginal distribution?

What is a conditional distribution? How do we know which variable to condition on?

What is a segmented bar graph? Why are they good to use?

What does it mean for two variables to have an association? How can you tell by looking at a graph?

Alternate Example: Super Powers

A sample of 200 children from the United Kingdom ages 9–17 was selected from the CensusAtSchool website. The gender of each student was recorded along with which super power they would most like to have: invisibility, super strength, telepathy (ability to read minds), ability to fly, or ability to freeze time.

(a) Explain what it would mean if there was no association between gender and superpower preference.

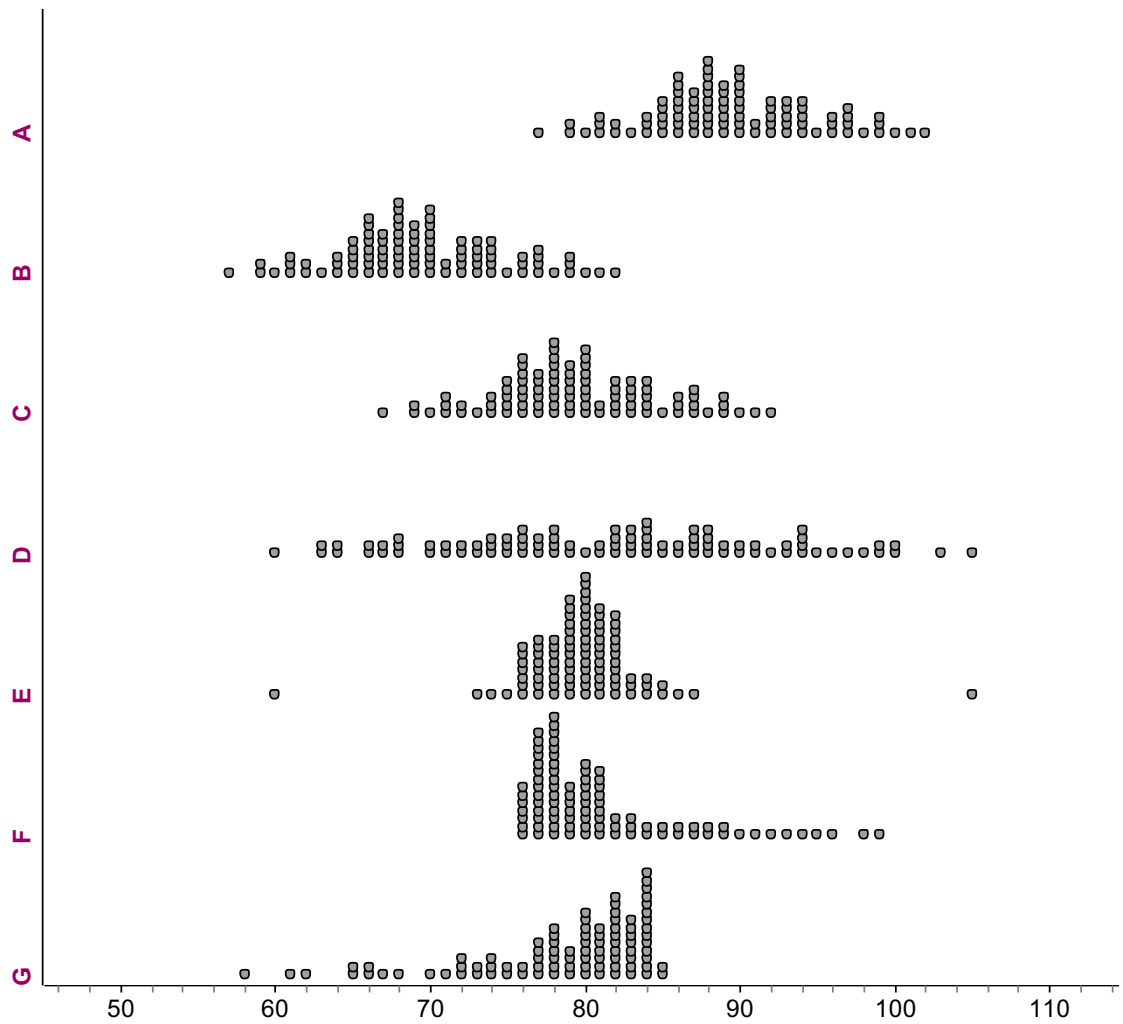
(b) Based on this data, can we conclude there is an association between gender and super power preference? Justify.

	Female	Male	Total
Invisibility	17	13	30
Super Strength	3	17	20
Telepathy	39	5	44
Fly	36	18	54
Freeze Time	20	32	52
Total	115	85	200

HW: page 24 (20, 22, 23, 25, 27–32)

1.2 Displaying Quantitative Data with Graphs

Brian and Jessica have decided to move and are considering seven different cities. The dotplots below show the daily high temperatures in June, July, and August for each of these cities. Help them pick a city by answering the questions below.



1. What is the most important difference between cities A, B, and C?
2. What is the most important difference between cities C and D?
3. What are two important differences between cities D and E?
4. What is the most important difference between cities C, F, and G?

Read 27–29

When describing the distribution of a quantitative variable, what characteristics should be addressed?

Read 29–31

Briefly describe/illustrate the following distribution shapes:

Symmetric

Skewed right

Skewed left

Unimodal

Bimodal

Uniform

Alternate Example: Frozen Pizza

Here are the number of calories per serving for 16 brands of frozen cheese pizza, along with a dotplot of the data.

340 340 310 320 310 360 350 330
260 380 340 320 360 290 320 330

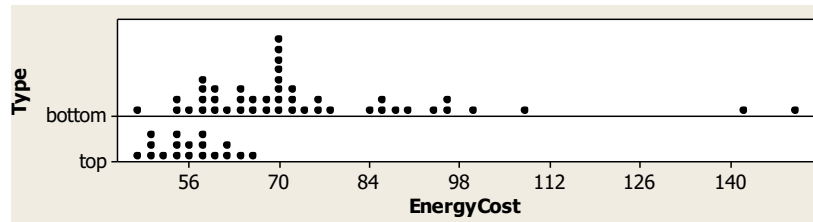
Describe the shape, center, and spread of the distribution. Are there any outliers?

Read 31–32

What is the most important thing to remember when you are asked to compare two distributions?

Alternate Example: Energy Cost: Top vs. Bottom Freezers

How do the annual energy costs (in dollars) compare for refrigerators with top freezers and refrigerators with bottom freezers? The data below is from the May 2010 issue of *Consumer Reports*.



Read 33–34 (word for word)

What is the most important thing to remember when making a stemplot?

Alternate Example: Which gender is taller, males or females? A sample of 14-year-olds from the United Kingdom was randomly selected using the CensusAtSchool website. Here are the heights of the students (in cm). Make a back-to-back stemplot and compare the distributions.

Male: 154, 157, 187, 163, 167, 159, 169, 162, 176, 177, 151, 175, 174, 165, 165, 183, 180

Female: 160, 169, 152, 167, 164, 163, 160, 163, 169, 157, 158, 153, 161, 165, 165, 159, 168, 153, 166, 158, 158, 166

HW: page 42 (39, 43, 44, 45)

Wednesday, September 4: 1.2 Histograms

The following table presents the average points scored per game (PPG) for the 30 NBA teams in the 2012–2013 regular season. Make a dotplot to display the distribution of points per game. Then, use your dotplot to make a histogram of the distribution.

Team	PPG	Team	PPG	Team	PPG
Atlanta Hawks	98.0	Houston Rockets	106.0	Oklahoma City Thunder	105.7
Boston Celtics	96.5	Indiana Pacers	94.7	Orlando Magic	94.1
Brooklyn Nets	96.9	Los Angeles Clippers	101.1	Philadelphia 76ers	93.2
Charlotte Bobcats	93.4	Los Angeles Lakers	102.2	Phoenix Suns	95.2
Chicago Bulls	93.2	Memphis Grizzlies	93.4	Portland Trail Blazers	97.5
Cleveland Cavaliers	96.5	Miami Heat	102.9	Sacramento Kings	100.2
Dallas Mavericks	101.1	Milwaukee Bucks	98.9	San Antonio Spurs	103.0
Denver Nuggets	106.1	Minnesota Timberwolves	95.7	Toronto Raptors	97.2
Detroit Pistons	94.9	New Orleans Hornets	94.1	Utah Jazz	98.0
Golden State Warriors	101.2	New York Knicks	100.0	Washington Wizards	93.2

Read 35–39

How do you make a histogram?

Why would we prefer a *relative* frequency histogram to a frequency histogram?

Read 39–41 (*skip #2*)

What will cause you to lose points on tests and projects (and make the rest of Sever's hair fall out)?

HW: page 45 (49, 51, 55, 59, 60, 67, 68)

1.3 Describing Quantitative Data with Numbers

Read 50–52

What is the difference between \bar{x} and μ ?

What is a resistant measure? Is the mean a resistant measure of center?

How can you estimate the mean of a histogram or dotplot?

Read 53–55

Is the median a resistant measure of center? Explain.

How does the shape of a distribution affect the relationship between the mean and the median?

Read 55–57

What is the range? Is it a resistant measure of spread? Explain.

What are quartiles? How do you find them?

What is the interquartile range (*IQR*)? Is the *IQR* a resistant measure of spread?

Alternate Example: McDonald's Fish and Chicken Sandwiches

Here are data on the amount of fat (in grams) in 9 different McDonald's fish and chicken sandwiches. Calculate the median and the *IQR*.

Sandwich	Fat (g)
Filet-O-Fish [®]	19
McChicken [®]	16
Premium Crispy Chicken Classic Sandwich	22
Premium Crispy Chicken Club Sandwich	33
Premium Crispy Chicken Ranch Sandwich	27
Premium Grilled Chicken Classic Sandwich	9
Premium Grilled Chicken Club Sandwich	20
Premium Grilled Chicken Ranch Sandwich	14
Southern Style Crispy Chicken Sandwich	19

Read 57–58

What is an outlier? How do you identify them? Are there outliers in the chicken/fish sandwich distribution?

Here is data for the amount of fat (in grams) for McDonald's beef sandwiches. Are there any outliers in this distribution?

Sandwich	Fat
Big Mac [®]	29
Cheeseburger	12
Daily Double	24
Double Cheeseburger	23
Double Quarter Pounder [®] with cheese	43
Hamburger	9
McDouble	19
McRib [®]	26
Quarter Pounder [®] Bacon and Cheese	29
Quarter Pounder [®] Bacon Habanero Ranch	31
Quarter Pounder [®] Deluxe	27
Quarter Pounder [®] with Cheese	26

Read 58–60

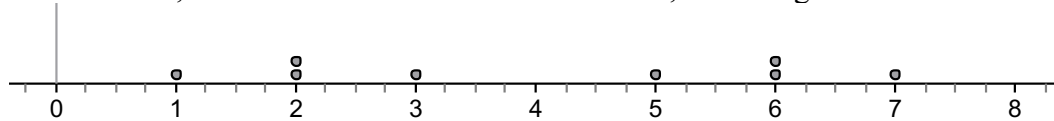
What is the five-number summary? How is it displayed?

Draw parallel boxplots for the beef and chicken/fish sandwich data. Compare these distributions.

HW: page 48 (69–74), page 70 (83, 85, 88, 91, 93, 94a)

1.3 Standard Deviation

In the distribution below, how far are the values from the mean, on average?



What does the standard deviation measure?

What are some similarities and differences between the range, *IQR*, and standard deviation?

Read 62–64 *Do the by-hand calculation for dotplot before doing the reading!*

How is the standard deviation calculated? What is the variance?

What are some properties of the standard deviation?

Alternate Example: A random sample of 5 students was asked how many minutes they spent doing HW the previous night. Here are their responses (in minutes): 0, 25, 30, 60, 90. Calculate and interpret the standard deviation.

Read 65–67

What factors should you consider when choosing summary statistics?

Don't need to do four-step process

HW: page 72 (99, 102, 103, 105, 107-111)