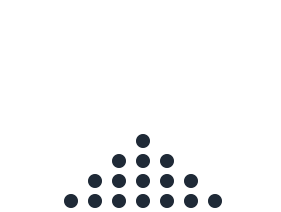


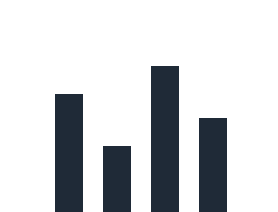
PROBLEM PACKET

Unit 1A

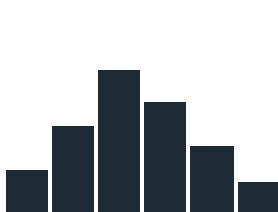
Exploring One-Variable Data



DOTPLOT



BAR CHART



HISTOGRAM

1 | 2 5 8
2 | 0 3 4 7
3 | 1 6 6 9
4 | 2 5
5 | 1

STEM-AND-LEAF

READ THIS FIRST

How this packet works

This packet is problems only. Your notes and worked examples are in the Guided Notes. The eight modules (A–H) match the order we cover things in class, so work them in order. Each module starts easy and builds to exam-level questions, with a mix of short answer, multiple choice, and free response. Show your work, and always answer in context. The stars tell you how hard a problem is:

- | | | | |
|---------|---|---------|--|
| ★ ★ ★ ★ | WARM-UP: recall & one step | ★ ★ ★ ★ | PROFICIENT: standard multi-step skill |
| ★ ★ ★ ★ | ADVANCED: true exam level, interpret & justify | ★ ★ ★ ★ | CHALLENGE: hardest exam reasoning |

Knowing Your Data

The data-analysis process; observational units & variables; categorical vs. quantitative; population, sample, parameter & statistic

1 True or False. For any you mark false, give a one-line reason.

★★★★

a) Statistics is the science and art of collecting, analyzing, and drawing conclusions from data.

b) For a data set to be *bivariate*, every variable in it must be quantitative.

2 For each of 80 students who used a new study app for one week, a researcher records the student's score on a vocabulary quiz. Identify the observational units and classify the quiz-score variable as categorical or quantitative.

★★★★

3 A school district wants to know whether students who eat breakfast tend to score higher on a standardized math test. From the 2400 students in the district, researchers randomly select 200. For each they record whether the student ate breakfast that morning (yes/no) and the student's score on the math test (0–100 points).

★★★★

a) Identify the population and the sample.

b) Identify the observational units and the two variables, classifying each.

c) Suppose 64% of the 200 sampled students ate breakfast. Explain why this 64% is a *statistic*, not a parameter, and describe what parameter it estimates.

- 4** **Error Analysis.** A classmate writes: *"In Problem 3, the population is the 200 students who were selected, because those are the students we actually have data on. So the 64% is a parameter."* Identify **both** mistakes in this reasoning and write a corrected sentence.

- 5** A principal claims: "Students at our school who participate in clubs have a higher GPA, so joining a club raises your GPA." Using only the ideas developed so far in this unit:

a) Identify the observational units and the two variables involved (classify each).

b) The data come from observing students as they already are. Explain, in words only, why any conclusion must stop at "club students tend to have higher GPAs" and cannot reach "joining a club raises GPA."

- 6** A snack company selects 60 of the 1500 protein bars produced in a factory shift and measures the sugar content (grams) in each. Identify the population and the sample. State n and N .

- 7** A phone factory produces 5000 screens in a day. An inspector tests 80 randomly chosen screens and finds 3 defective. The proportion $3/80 = 0.0375$ is best described as



- A) a parameter, because it describes the day's production.
B) a statistic, because it describes the sample of 80 screens.
C) the population, because it is the group of interest.
D) a census, because every screen was tested.
E) an observational unit.

8

★★★★



Which statement correctly distinguishes a parameter from a statistic?

- A) A parameter describes a sample; a statistic describes a population.
- B) A parameter is always larger than a statistic.
- C) A parameter describes a population; a statistic describes a sample.
- D) A statistic can never be used to estimate a parameter.
- E) A parameter is computed from data; a statistic is unknown.

9

★★★★

A polling firm surveys 1200 randomly chosen teens and finds 47% plan to watch the upcoming season of a popular show. Identify the population, the sample, the parameter, and the statistic.

10

★★★★

A school district owns 1000 Chromebooks. To estimate the proportion of all its Chromebooks that need repair, a tech aide randomly selects 50 and finds 8 need repair.

- a) Identify the population, sample, parameter, and statistic.

- b) State the values of n and N .

- c) Compute the value of the statistic.

- d) Explain in one sentence why the statistic might differ from the true parameter.

11

★★★★

Always / Sometimes / Never. For each statement write **A**, **S**, or **N** and give a one-line reason.

- a) A census makes the value of the statistic equal to the parameter.

- b) A sample that is smaller than the whole population can still give a statistic exactly equal to the parameter.

- c) For a fixed population, a parameter is a single fixed number (even if we don't know it).

- d) Two different random samples from the same population give the same value of a statistic.

12

★★★★

An esports club keeps a data table on the 60 players on its roster. A coach randomly selects 15 players and, for each, records: gamertag, main role (duelist, controller, sentinel, ...), average reaction time (milliseconds), whether the player has competed in a tournament, and number of ranked matches played last week.

- a) Identify the population and the sample. State n and N .

- b) Identify the observational units. Explain why "gamertag" is *not* a variable.

- c) List the variables; classify each as categorical or quantitative, and each quantitative one as discrete or continuous.

- d) Of the 15 sampled players, 9 have competed in a tournament. Identify $9/15 = 60\%$ as a parameter or a statistic, and explain.

13

★★★★

Classify each variable as categorical (C), discrete quantitative (DQ), or continuous quantitative (CQ).

Recall: categorical is also called *qualitative*; quantitative is also called *numerical*.

- | | |
|---|---|
| a) ethnicity | f) number of days absent from school |
| _____ | _____ |
| b) number of times directions (North, etc.) were used | g) room temperature |
| _____ | _____ |
| c) gender | h) the number of tests taken |
| _____ | _____ |
| d) number of social events attended in a week | i) grade level (freshman, sophomore, ...) |
| _____ | _____ |
| e) the time it takes to complete a test | |
| _____ | |

14 **Error Analysis.** A student classified two variables as quantitative “because their values are numbers”: a phone’s *area code* and a player’s *jersey number*. Explain why each is actually categorical, and state the one-sentence test that decides whether a number-valued variable is quantitative.

15 Which of the following is quantitative **and** continuous?

★★★★



- A) the number of free throws a player makes in a game
- B) a person's blood type (A, B, AB, O)
- C) the time (minutes) a runner takes to finish a 5K race
- D) a student's grade level (9, 10, 11, 12)
- E) the number of apps installed on a phone

16 A data table has one row per mobile game and these columns: app-store ID (a 9-digit number), average revenue per user (dollars), 5-digit developer studio code, number of in-app purchases offered, and a “fun score” from 1–100 produced by averaging players’ 1–5 ratings. How many columns are **quantitative** variables? Choose the answer with the best reasoning.

★★★★



- A) 4, since every column but the game name is a number.
- B) 3, since revenue, in-app purchases, and fun score are quantitative; the two ID codes are labels.
- C) 2, since only revenue and in-app purchases count; the fun score comes from ratings.
- D) 5, since all five columns are numbers.
- E) 3, since revenue, studio code, and app-store ID are quantitative.

17 A streaming service stores a data table in which each row is one subscriber. The columns are: account number, subscription plan (Basic/Standard/Premium), age (years), number of devices registered, total hours watched last month, country, and whether the account is shared.

★★★★

- a) Identify the observational units and the variables.

- b) Classify each variable as categorical or quantitative.

- c) For each quantitative variable, state discrete or continuous.

18

★★★★

Two researchers disagree about how to classify three variables collected on a sample of concertgoers: (i) seat number (e.g., 214, 307), (ii) an enjoyment level reported on a 0–10 scale, and (iii) ticket tier (GA, lower bowl, upper bowl, VIP). Researcher 1 says “anything recorded as a number is quantitative.” Researcher 2 disagrees.

- a) Classify each variable (and discrete/continuous where quantitative). For each, give a one-sentence justification appealing to labels-vs-amounts and whether averaging is meaningful.

- b) The 0–10 enjoyment scale is the genuinely debatable one. Give the strongest argument for treating it as quantitative *and* for treating it as categorical, then state which side you take and why.

- c) Construct one clear counterexample to Researcher 1's rule from this data set, and state the correct test for whether a numeric variable is quantitative.

19

★★★

Matching. Write the letter of the example that best fits each term. Each is used once.

- | | |
|-----------------------|--|
| 1. observational unit | A. the 75 cafeteria lunches inspected today |
| 2. variable | B. all lunches the cafeteria will serve this year |
| 3. population | C. one inspected lunch tray |
| 4. sample | D. number of calories on a tray |
| 5. parameter | E. the true mean calories of all lunches this year |
| 6. statistic | F. the mean calories of the 75 inspected trays (612) |

20

★★★★

A student council wants to study the school's WiFi reliability. Plan how you would collect the data. Your answer must:

- a) define the observational units explicitly (careful: they are *not* the routers);

- b) name at least two variables, classify each (with discrete/continuous where relevant), and phrase each name so it implies different values;

- c) distinguish the population from a realistic sample, and state which number would be a statistic and which unknown number would be the parameter it estimates.

Picturing One Categorical Variable

Frequency & relative-frequency tables; bar charts; pie charts; comparing the same categorical variable across data sets

USE THIS DATA FOR PROBLEMS 21–22

A class recorded the primary mode of transportation to school for all 25 students:

Bus Car Bus Walk Bus Car Bike Bus Car Walk
Bus Car Bus Walk Bus Car Bus Bike Car Bus
Walk Car Bus Car Bus

21 Make a **frequency table** for the distribution of primary mode of transportation. Include a Total row, and verify the frequencies add to 25.

★★★★

MODE OF TRANSPORTATION	FREQUENCY
TOTAL	

22 Make a **relative frequency table** for the same data. Round each relative frequency to three decimal places, and state which mode of transportation is most common.

★★★★

MODE OF TRANSPORTATION	RELATIVE FREQUENCY
TOTAL	

23
★★★★

In a survey of 480 students about their favorite Chipotle order, 156 chose a burrito. Compute the relative frequency of "burrito" (as a decimal to three places and as a percentage).

24
★★★★

A class of 32 students reports favorite fast-food spot: 18 Chick-fil-A, 8 Chipotle, 6 other. The correct relative frequency for "Chipotle" is



- A) 0.18
- B) 0.20
- C) 0.25
- D) 0.56
- E) 8

25
★★★★

Error Analysis. A student turns in this relative-frequency table for favorite lunch ($n = 200$): Pizza 0.30, Sushi 0.25, Salad 0.20, Tacos 0.15, Wraps 0.15. Without any other information, explain how you know the table contains an error, and describe the two most likely causes.

26
★★★★

A relative frequency table of preferred study music for 200 students lists: Lo-fi 0.355, Pop 0.250, Hip-hop 0.210, Classical 0.085, and "No music" for the rest. **(a)** What proportion chose "No music"? **(b)** How many students is that?

27
★★★★

A relative frequency table for favorite music genre lists Pop 0.33, Hip-hop 0.25, Rock 0.17, Country 0.13, and Other. The relative frequency for "Other" must be



- A) 0.00
- B) 0.12
- C) 0.17
- D) 0.88
- E) cannot be determined

28

★★★★

The frequency table shows the favorite music genre of 300 students.

GENRE	STUDENTS
Pop	96
Hip-hop	84
Rock	54
Country	42
Other	24
TOTAL	300

(a) Make a relative frequency table (three decimals). (b) Which genre is most common? (c) What percentage chose Rock?

GENRE	RELATIVE FREQUENCY
Pop	
Hip-hop	
Rock	
Country	
Other	
TOTAL	

29

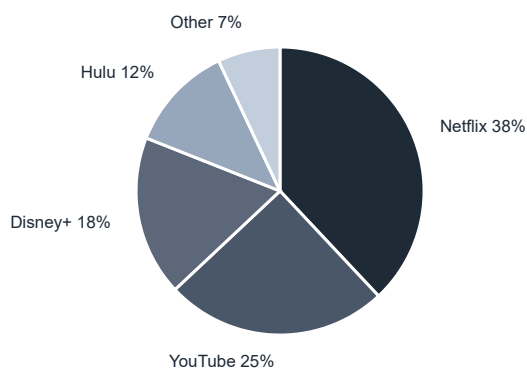
★★★★

A pie chart shows favorite season for a class: Summer 40%, Fall 25%, Winter 20%, Spring 15%. (a) About what percentage chose Summer or Fall combined? (b) Which season was least preferred?

30

★★★★

A sample of students named their most-used streaming service. The pie chart shows the distribution.

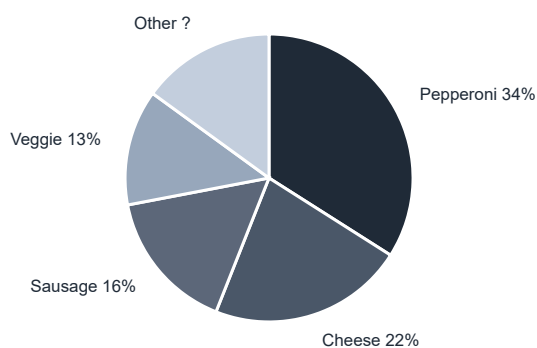


(a) Which service was most used, and which was least? (b) What percentage used Netflix or YouTube combined? (c) If 400 students were surveyed, about how many chose Disney+?

31

★★★★

A pizzeria charts customers' favorite topping, but the "Other" slice label is missing.



(a) Find the percentage for the "Other" slice. (b) Explain why this set of categories makes a pie chart appropriate. (c) Which single topping makes up about one-third of choices?

32

★★★★

True or False. For any you mark false, give a one-line reason.

a) A pie chart is an appropriate display for a single categorical variable.

b) The relative frequencies in a one-variable table add to 1 (up to rounding).

c) A category's slice of a pie chart and its bar on a bar chart of the same data represent the same relative frequency.

d) If one category is left out of the table, the relative frequencies that remain will still add to 1.

33

★★★★

True or False. For any you mark false, give a one-line reason.

a) A bar graph should always use bars of equal width.

b) Pie charts and bar graphs work best for numerical datasets.

c) Bar graphs should only be used with categorical data.

d) The number of degrees for a slice in a pie chart is the relative frequency for the category multiplied by 100.

USE THIS FREQUENCY TABLE FOR PROBLEMS 34–36

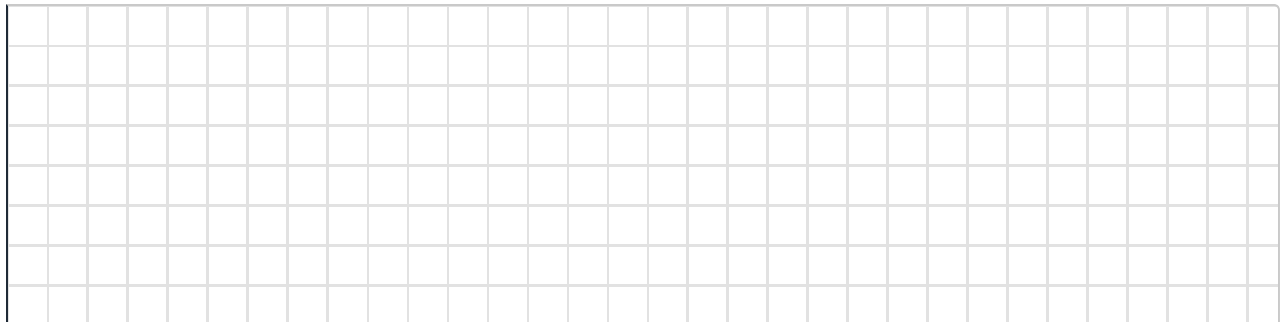
A high school esports club recorded the main console or platform every member plays on.

PLATFORM	MEMBERS
PS5	64
Xbox	52
Switch	16
PC	12
Mobile	6
TOTAL	150

34

★★★

Make a **frequency bar chart** for these platform data (labeled vertical axis starting at 0, equal-width bars). Then describe what you see in one sentence.



PLATFORM

35

★★★

Convert the platform table to relative frequencies (round to three decimals). Would it be appropriate to display these data with a pie chart? Explain.

36

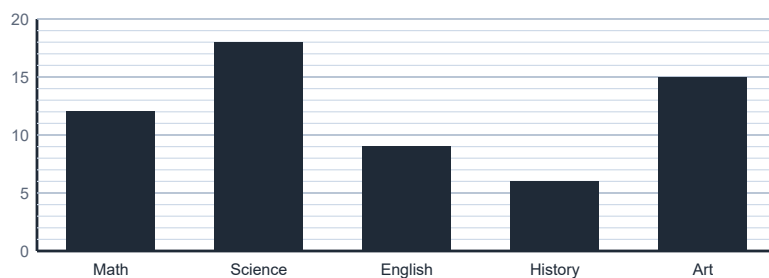
★★★★

Suppose you build a pie chart of the platform data. The slice for a category has a central angle equal to its relative frequency times 360° . **(a)** Compute the central angle (degrees) for the PS5 slice and for the Switch slice. **(b)** Explain why all five central angles must sum to 360° .

37

★★★★

A class of 60 students reported their favorite school subject.



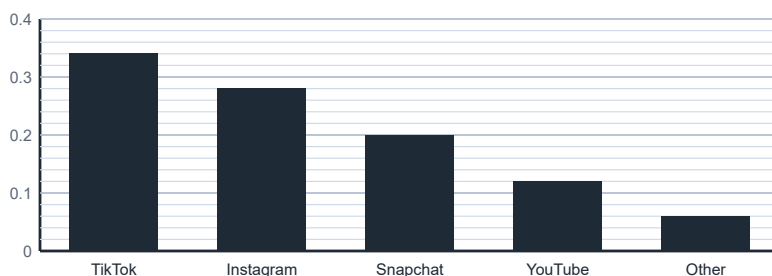
FAVORITE SUBJECT · FREQUENCY

(a) Which subject was most popular? Least? **(b)** Compute the relative frequency of "Science." **(c)** Verify the five frequencies total 60.

38

★★★★

A relative-frequency bar chart shows the social media app each surveyed teen uses most.



APP USED MOST · RELATIVE FREQUENCY

(a) Read the relative frequency for "Other." (b) If 250 teens were surveyed, how many chose Instagram? (c) Explain why the five bar heights must add to 1.

39

★★★★



Which statement about a pie chart of a single categorical variable is TRUE?

- A) The slices may add to more than 100%.
- B) A pie chart can show the exact count for each category without a legend.
- C) The area of each slice is proportional to that category's relative frequency.
- D) A pie chart is appropriate even when an "Other" category is omitted.
- E) A pie chart works for a quantitative variable as well as a categorical one.

40

★★★★

The relative-frequency table summarizes the colors of vehicles sold worldwide in 2020.

COLOR	% OF VEHICLES	COLOR	% OF VEHICLES
Black	19	Red	5
Blue	7	Silver	9
Brown	3	White	36
Gray	15	Yellow	2
Green	1	Other	??

(a) What percentage of vehicles fall in the “Other” category? (b) Would it be appropriate to make a pie chart of these data? Why or why not?

41

★★★★



For which of the following would it be INAPPROPRIATE to display the data with a single pie chart?

- A) The distribution of car colors for vehicles purchased in the last month
- B) The distribution of unemployment percentages for each of the 50 states
- C) The distribution of favorite sport for a sample of 30 middle-school students
- D) The distribution of shoe type worn by shoppers at a local mall
- E) The distribution of presidential-candidate preference for voters in a state

42

★★★★



A class wants to compare the favorite-lunch distribution of its 18 students with the favorite-lunch distribution of the whole school (1200 students), using the *same* categories. Which is the best reason to compare with relative frequencies rather than raw counts?

- A) Relative frequencies are always whole numbers.
- B) Counts cannot be displayed in a bar chart.
- C) The two groups are very different sizes, so proportions allow a fair comparison.
- D) Relative frequencies do not need to add to 1.
- E) Pie charts require counts, not proportions.

USE THIS INFORMATION FOR PROBLEM 43

Two boba shops each surveyed their own customers about their most-ordered drink: the *same* categorical variable in two separate data sets. Shop A surveyed 80 customers; Shop B surveyed 240 customers.

DRINK	SHOP A (COUNT)	SHOP B (COUNT)
Milk tea	32	60
Classic boba	16	108
Fruit tea	24	36
Matcha	8	36
TOTAL	80	240

43

★★★★

Compare the two shops' drink distributions. (a) A manager says, "Shop B sells way more classic boba, so classic boba is more popular at Shop B." Compute the relative frequency of classic boba *within each shop* and evaluate this claim. (b) For which shop is fruit tea a larger *share* of orders? Show your work. (c) Explain why comparing the raw counts directly would be misleading here.

44

★★★★

A relative-frequency bar chart of "primary way I listen to music" is shown for one group, but two values are smudged. Known: **Spotify 0.32**, **YouTube 0.40**, Apple Music is exactly half of SoundCloud, these are the only remaining categories, and there is no "Other."

a) Find the relative frequencies for SoundCloud and Apple Music.

b) If the group had 200 people, reconstruct the count in every category and verify they sum to 200.

c) Explain why, if even one category were left out of the chart, the relative frequencies could no longer be trusted to sum to 1.

45

★★★

A school store recorded the type of every item bought during one week (all 600 purchases).

ITEM	PURCHASES
Snacks	288
Drinks	150
School supplies	108
Spirit gear	42
Stickers	12
TOTAL	600

- a) Make a relative frequency table (round to three decimals).
- b) Make a relative frequency bar chart; label and scale both axes.
- c) The manager claims, "A majority of our purchases are snacks." Use your table to justify whether the data support this.

ITEM	RELATIVE FREQUENCY
Snacks	
Drinks	
School supplies	
Spirit gear	
Stickers	
TOTAL	

46

★★★★

A survey asks, "Which of these streaming services do you use? (Select ALL that apply.)" Of 500 respondents: Netflix 310, Disney+ 240, Hulu 180, Max 150. A designer makes a single pie chart of these four numbers, computing "shares" as $310/880$, $240/880$, etc.

- a) Explain why a pie chart is invalid here, referring to $310+240+180+150 = 880$ versus 500 respondents.

- b) What property must the categories of a single categorical variable satisfy for a pie chart to be appropriate?

- c) Propose a correct display, and state what each bar's height would represent.

Picturing Numbers

Building and reading dotplots, stem-and-leaf plots, and histograms for one quantitative variable

47 True or False. For any you mark false, give a one-line reason.

★★★★

a) Dotplots are often used to summarize large categorical datasets.

b) If a data set has 200 observations, then the recommended number of intervals (bins) in a histogram is 14–15.

c) A pie chart and a histogram are most useful for numeric data.

48 Goals scored per game by a soccer team (24 games):

★★★★

1 2 0 1 3 2 4 1 2 5 1 0 2 6 1 3 2 1 4 0 2 5 1 9

(a) Make a dotplot; label and scale the axis. (b) In what proportion of games did the team score 4 or more goals?

NUMBER OF GOALS

49

★★★★

Daily phone pickups (times unlocked) for 18 students:

27 31 28 33 39 26 29 41 27 29 31 32 31 35 31 32 33 49

(a) Make a dotplot. (b) What percentage of these students picked up their phone more than 35 times in a day?

DAILY PHONE PICKUPS

50

★★★

A dotplot shows quiz scores (out of 10) for 20 students.



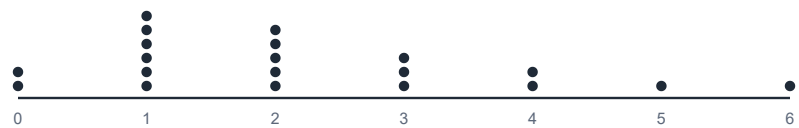
QUIZ SCORE

(a) What proportion of students scored 9 or higher? (b) Describe the shape. (c) Which score(s) occurred most often?

51

★★★

A dotplot shows the number of siblings for 20 students.



NUMBER OF SIBLINGS

(a) What is the most common number of siblings? (b) What proportion of students have 3 or more siblings? (c) Describe the shape of the distribution.

52

★★★★

MC

Which is the clearest difference between a histogram and a bar chart?

- A) A histogram always has more bars than a bar chart.
- B) A histogram displays a quantitative variable (numeric intervals); a bar chart displays a categorical variable.
- C) A bar chart can never have gaps between bars.
- D) Only histograms can use relative frequencies.
- E) A histogram's bars must all be the same height.

53

★★★★

Error Analysis. A student displays favorite gaming platform (PS5, Xbox, Switch, PC) by drawing four bars *with no gaps between them* and calls it a histogram. Give the two things wrong with this, and state the correct graph and how its bars should look.

54

★★★★

Weights (grams) of 16 mini chocolate bars:

21.2 22.5 21.8 22.0 21.5 22.7 21.9 22.3 20.9 21.1 21.6 22.1 21.4 24.8 22.2 21.7

(a) Make a stem-and-leaf plot (first two digits as the stem, last digit as the leaf). Include a key. (b) What interesting feature does the graph reveal?

Hours of video gaming last week for 20 friends:

4.0	0.3	16.3	0.6	2.2	15.4	7.1	3.4	0.2	1.5
11.9	7.8	8.4	0.5	1.9	2.3	9.4	5.6	5.0	0.7

Make a frequency histogram using intervals of width 2 starting at 0 (0 to <2, 2 to <4, ...). Label and scale both axes.



HOURS OF GAMING LAST WEEK

Number of text messages sent yesterday by 25 students:

8 12 15 4 22 31 6 18 25 9 14 11 27 5 19
33 7 16 21 10 13 28 17 6 24

(a) Make a frequency histogram using intervals of width 5 starting at 0. (b) Make a relative-frequency histogram of the same data. How do the two graphs compare in shape?



TEXT MESSAGES SENT · FREQUENCY

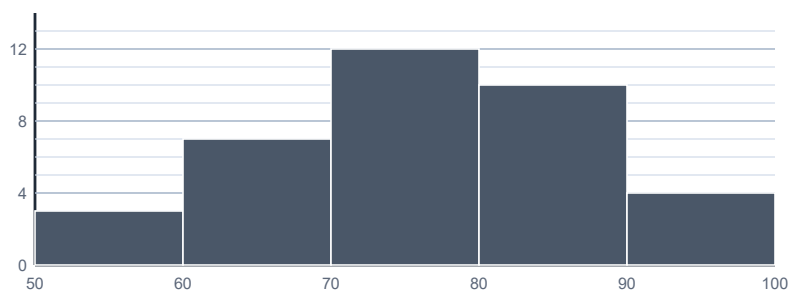


TEXT MESSAGES SENT · RELATIVE FREQUENCY

57

★★★★

A frequency histogram shows test scores for a class.



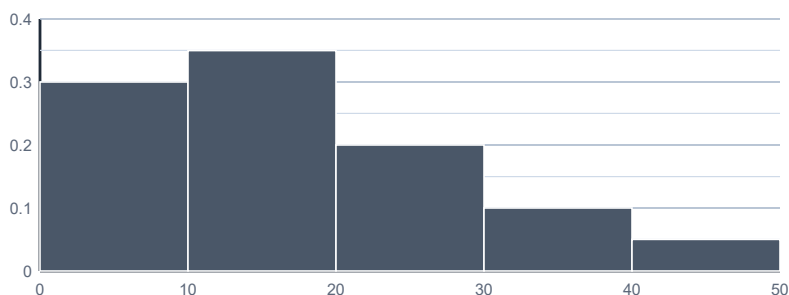
TEST SCORE · FREQUENCY

(a) Which interval contains the most students? (b) What proportion of students scored at least 80? (c) How many students took the test?

58

★★★★

A relative-frequency histogram shows commute times (minutes) for 200 workers.



COMMUTE TIME (MIN) · RELATIVE FREQUENCY

(a) What proportion of workers commute less than 20 minutes? (b) About how many workers are in the 10-to-<20 interval? (c) Verify the bar heights sum to 1.

59

★★★★

Daily Spotify listening time (minutes) for one student over 22 days:

```

7 | 6 8 9
8 | 0 1 1 3 4 4 6 7 8 9
9 | 0 1 2 3 5 7 8
10 | 1 4
Key: 8|4 = 84 minutes

```

(a) On how many days did the student listen at least 90 minutes? (b) Find the median listening time. (c) Describe the shape and identify any potential outliers.

60

★★★★

A stem-and-leaf plot shows the heights (cm) of 16 students.

```

15 | 2 5 8
16 | 0 1 3 3 6 8 9
17 | 0 2 4 5 7
18 | 1
Key: 15|2 = 152 cm

```

(a) How many students are at least 170 cm tall? (b) Find the median height. (c) What is the range?

61

★★★★

Which graph BEST displays the distribution of one quantitative variable while still showing every individual value for a small data set ($n = 20$)?

MC

- A) A stem-and-leaf plot
- B) A bar chart
- C) A pie chart
- D) A relative-frequency histogram with very wide bins
- E) A scatterplot

62

★★★★★

MC

In a histogram, the convention used in this course for a value at an interval boundary is to

- A) include the right endpoint and exclude the left
- B) include the left endpoint and exclude the right
- C) include both endpoints in both bars
- D) split the value evenly between the two bars
- E) always put boundary values in the taller bar

USE THIS STEM-AND-LEAF PLOT FOR PROBLEMS 63–64

Length (seconds) of 16 songs:

```

12 | 0 4 8
13 | 1 1 5 6 9
14 | 0 2 2 7
15 | 3 8
16 | 1
17 |
18 | 9
Key: 14|2 = 142 seconds

```

63

★★★★★

MC

The number of songs that are at least 150 seconds long is

- A) 2
- B) 3
- C) 4
- D) 5
- E) 6

64

★★★★★

MC

Which statement best describes the distribution above?

- A) Skewed to the left with no outliers
- B) Roughly symmetric and bimodal
- C) Approximately uniform
- D) Skewed to the right, with a possible outlier at 189 seconds
- E) Skewed to the right, with a possible outlier at 120 seconds

65

★★★★★

Investigation. Number of stickers on each student's laptop, for 20 students in an AP Statistics class:

6 4 9 5 8 7 6 5 11 8 7 9 6 7 5 33 8 6 31 10

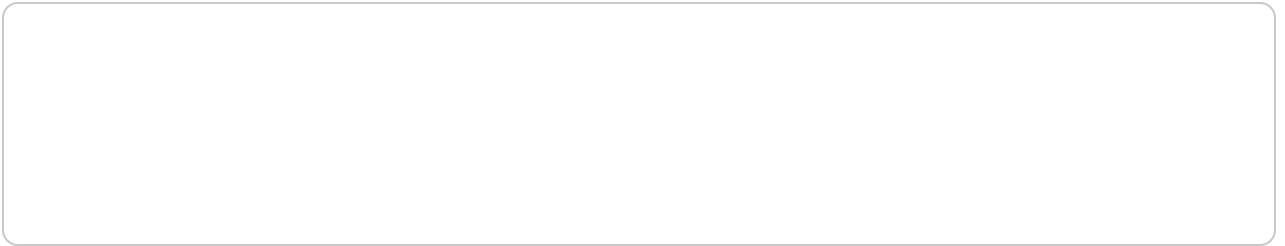
- a) Make a stem-and-leaf plot **without** splitting the stems.

- b) Make a second stem-and-leaf plot **with** split stems (0–4 and 5–9 on separate rows).

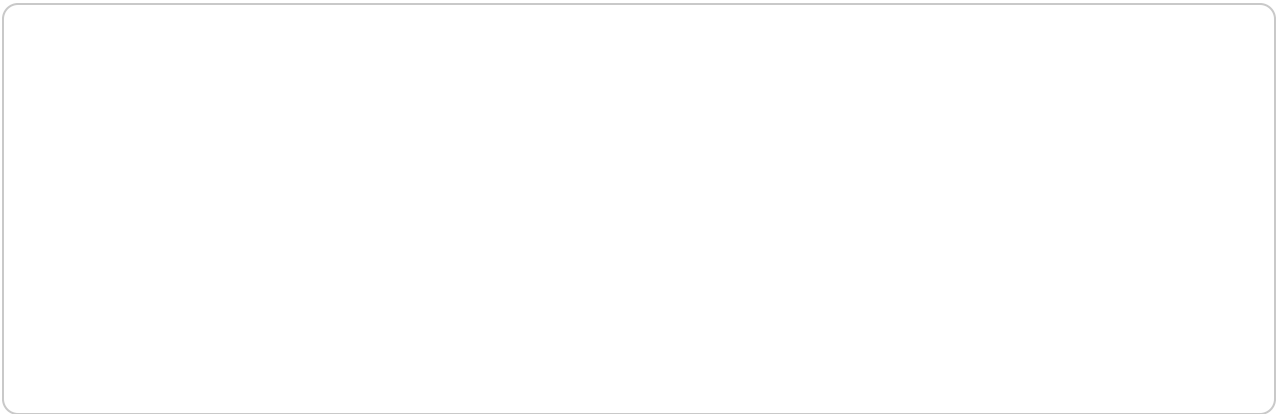


- c) Which of the two gives a better picture of this distribution, and why? Refer to a specific feature one reveals that the other hides.

- (a) Without splitting stems



- (b) With split stems (0–4 and 5–9)



66

★★★★

Test results for two sections of AP Stats. Make a back-to-back stem-and-leaf plot to compare them, then say which section did better and how you can tell.

1st Period: 98 52 92 92 60 66 90 86 86 70 72 84 82 82 82 74 74 76 80 80

2nd Period: 96 95 92 54 57 58 86 85 82 82 82 60 66 66 66 68 80 77 76 76 75 74 74 74 73 72
72 71 70

← 1st Period leaves · stem (center) · 2nd Period leaves →

67

★★★★

Which is the best reason for choosing a stem-and-leaf plot rather than a histogram to display a quantitative variable?

MC

- A) Stem-and-leaf plots allow you to split stems; histograms don't.
- B) Stem-and-leaf plots display every individual data value.
- C) Stem-and-leaf plots are better for very large sets of data.
- D) Stem-and-leaf plots never require rounding of values.
- E) Stem-and-leaf plots make it easier to determine the shape of a distribution.

68

★★★★

Grams of sugar per serving in 24 boba drinks:

7.2 8.4 8.0 7.8 9.1 8.6 7.5 8.9 9.3 8.2 7.9 8.7 9.0 8.3 7.6 9.2 8.5 8.1 7.7 8.8 9.4 7.3
12.6 8.0

(a) Make a stem-and-leaf plot with a key. (b) Identify one feature that stands out and explain what it means in context.

69

★★★★

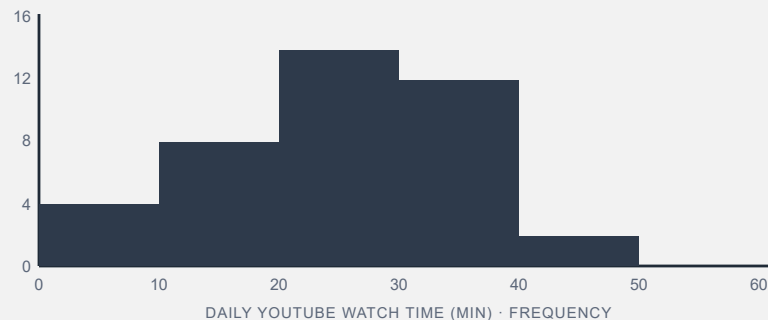


A histogram is redrawn using bins HALF as wide as before. The most accurate statement is

- A) The shape can never change when bin width changes.
- B) Narrower bins always make the distribution look perfectly symmetric.
- C) The center of the data physically moves to the right.
- D) Narrower bins always reduce the number of peaks shown.
- E) Narrower bins show more detail and may reveal additional peaks, possibly changing the apparent shape.

USE THIS HISTOGRAM FOR PROBLEM 70

Daily YouTube watch time (min) for 40 students:



70

★★★★★



The proportion of these 40 students with a daily watch time of at least 30 minutes is closest to

- A) 0.05
- B) 0.30
- C) 0.35
- D) 0.55
- E) 0.65

71

★★★★★

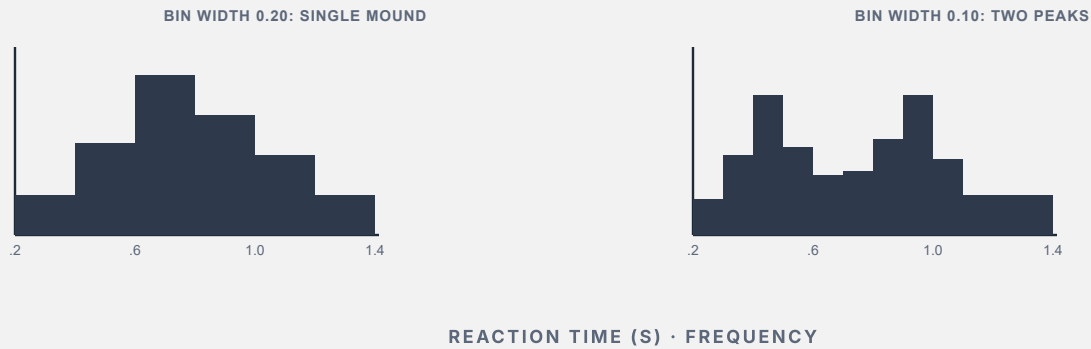


When comparing the distribution of a quantitative variable for two groups, it is best to use relative-frequency histograms rather than frequency histograms when

- A) the distributions have different shapes.
- B) the distributions have different amounts of variability.
- C) the distributions have different centers.
- D) the distributions have different numbers of data values.
- E) at least one of the distributions has outliers.

USE THESE TWO HISTOGRAMS FOR PROBLEM 72

The SAME 40 gaming reaction-time measurements (seconds), graphed two ways. Only the bin width differs.



72 Investigation.

★★★★

- a) Describe how the apparent SHAPE differs between the two histograms (number of clear peaks).

- b) Explain HOW changing bin width can hide or create the appearance of multimodality (refer to what happens when neighboring bins merge).

- c) A researcher believes the times come from two groups (distracted vs. focused). Which histogram better supports investigating that, and what would you do NEXT to decide whether the second peak is real rather than an artifact of the binning?

Telling a Distribution's Story

Shape; outliers; center; variability; describing and comparing distributions of one quantitative variable in context

- 73** ★★★ A distribution whose graph has a long tail stretching toward the SMALLER values, with most data piled up on the right, is best described as

MC

- A) skewed to the left
- B) roughly symmetric
- C) skewed to the right
- D) approximately uniform
- E) bimodal

- 74** ★★★ Name the shape of each dotplot using the best term(s): roughly symmetric, skewed left, skewed right, approximately uniform, unimodal, or bimodal.



- 75** ★★★ Which of the following distributions are more likely to be skewed to the LEFT than to the right? Explain each.

- a) the number of right-handed people in repeated samples of 10 people

- b) the number of hours spent studying for this course over the summer

- c) the number of credit cards an adult has in their wallet

- 76** ★★★ When asked to “describe the distribution” of a quantitative variable on the AP exam, a complete answer must address

MC

- A) only the shape
- B) only the center and variability
- C) shape and a correlation coefficient
- D) the mean and standard deviation only
- E) shape, outliers, center, and variability, all in context

77

★★★★

Error Analysis. A student describes a distribution of house prices: *"It's skewed left because a few very expensive houses pull a long tail out to the right, and that right tail drags the median above the mean."* This one sentence contains two errors. Identify both and write the corrected sentence.

78

★★★★

Always / Sometimes / Never. Write **A**, **S**, or **N** with a one-line reason.

a) In a right-skewed distribution, the mean is greater than the median.

b) A distribution with an outlier is skewed.

c) The median is affected when a single extreme value is made even more extreme.

d) A roughly symmetric distribution has its mean approximately equal to its median.

Teaspoons of sugar a barista adds to each of 24 iced coffees at a café:

2.0 0.0 3.5 1.0 2.5 4.0 0.0 3.0 1.5 2.0 5.0 3.0 0.0 2.5 1.0 4.0 3.0 1.0 2.5 3.5 4.5 2.0 5.0
4.0

- (a) Make a dotplot. (b) Write a complete description of shape, outliers, center, and variability in context. (c) The AHA suggests a daily max of 6 tsp added sugar for women. Can the dotplot justify that a *majority* of these drinks contain less than HALF that maximum? Justify with a calculation.



80

★★★★★

Price (dollars) of each boba drink on a shop's menu (22 drinks):

1.50 1.75 1.50 3.00 1.50 1.50 1.50 1.75 1.25 2.00 1.75
2.50 1.25 1.75 1.50 1.50 1.25 1.50 2.50 1.25 1.75 1.75

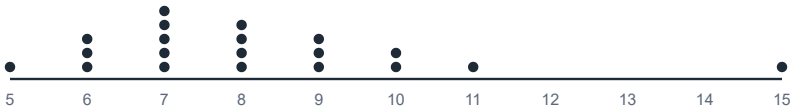
(a) Make a dotplot of the prices. (b) Describe the distribution (shape, outliers, center, variability) in context. (c) A flyer claims four friends with a \$100 budget can each buy at least 12 *different* drinks. Does your graph support this? Justify (cost for four people each buying the 12 cheapest drinks).

DRINK PRICE (\$)

81

★★★★★

A dotplot shows the daily step count (in thousands) for 20 people on one day.



DAILY STEPS (THOUSANDS)

(a) Write a complete description (shape, outliers, center, variability) in context. (b) A fitness app claims "most users hit 10,000 steps." Does this dotplot support that? Justify with a calculation.

82

★★★★★

A survey of 1640 teens who built their own gaming PC setup (amounts in hundreds of dollars):

INTERVAL (\$100S)	PEOPLE	INTERVAL (\$100S)	PEOPLE
0 to <2	360	8 to <10	210
2 to <4	410	10 to <12	150
4 to <6	255	12 to <14	25
6 to <8	215	14 to <16	15

a) Describe the shape in context.

b) About what percentage spent at least \$200? Show your method.

c) A campaign claims teens "typically spend at least \$500." Do the data support that? Identify the interval containing the median and explain.

83

★★★★

You collect two data sets: (i) the number of TikTok followers of all 30 students in one class, where a few have gone viral with far more followers; and (ii) the heights (cm) of 200 randomly chosen high-school students of the same age and sex. For EACH set: (a) predict the shape and explain your reasoning in context; (b) choose the best graph (dotplot, stem-and-leaf plot, or histogram) and justify based on the set's size and what you want to show.

84

★★★★

A teacher asked each of 24 students how many hours they spent gaming last week:

31 17 13 11 9 24 17 12 11 8
 20 15 12 11 8 19 14 12 10 8 18 13 11 9

- a) Make a frequency histogram using intervals of width 5 starting at 0. Label and scale both axes.
- b) Write a complete description (shape, outliers, center, variability) in context.

- c) A student claims, "Most of us gamed fewer than a dozen hours." Use your histogram to decide whether the data support this. Show your calculation.

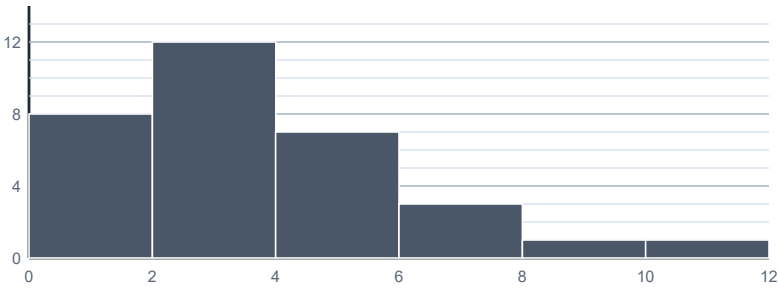


HOURS SPENT GAMING LAST WEEK

85

★★★★★

A frequency histogram shows daily social-media time (hours) for 32 students.



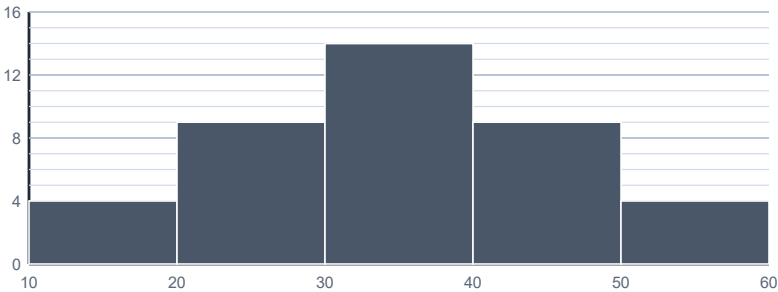
HOURS ON SOCIAL MEDIA · FREQUENCY

(a) Write a complete description (shape, outliers, center, variability) in context. (b) Without computing them, predict whether the MEAN is greater than, less than, or about equal to the median. Justify from the shape.

86

★★★★★

A frequency histogram shows the ages of 40 visitors to a museum exhibit.



AGE (YEARS) · FREQUENCY

(a) Describe the shape. (b) Which interval contains the median? Explain how you located it.

87

★★★★★

A coach recorded daily phone screen-time (minutes) for student athletes and non-athletes:

Non-athletes	stem	Athletes
9 8 8 6	5	2 4 6 8
6 4 4 2 1 0 0	6	0 1 3 5 7 9
8 6 4 2	7	0 2 4
5 0	8	1

Key: stem 6 with leaf 0 = 60 minutes. Non-athlete leaves grow LEFT, athletes grow RIGHT.

- a) Compare the SHAPES of the two distributions in context.

- b) Compare the CENTERS using the median of each group; report both values.

- c) Compare the VARIABILITY by computing and comparing the range (max – min) of each group.

- d) Which group tends to have lower screen-time? Explain why this comparison makes sense, and note one advantage of a back-to-back stem-and-leaf plot.

88

★★★★

The dotplots show goals per game for two soccer teams over 15 games each, on the same scale.

Team A



Team B



GOALS PER GAME

Compare the two distributions in context. Address (a) shape, (b) center (which team tends to score more), and (c) variability (which team is more consistent).

89

★★★★

A car-shopping app reports that the asking prices of 60 used cars teens browse have a median of \$11,800, a minimum of \$4,200, and a maximum of \$46,500, and that “most cars cluster in the low-to-mid teens of thousands, but a handful of specialty vehicles are priced far above the rest.” No graph is provided.

- a) From the description alone, sketch the likely dotplot of asking price and name the shape. Justify.
- b) Do you expect the MEAN to be greater than, less than, or about equal to the median of \$11,800? Justify from the shape you inferred in part (a).
- c) The app advertises “Average price under \$12,000!” Explain why this could mislead about a typical car a shopper encounters. Reference center AND variability.

ASKING PRICE (\$)

90

★★★★

Construct & Prove. Build data sets (whole-number quiz scores 0–10, repeats allowed).

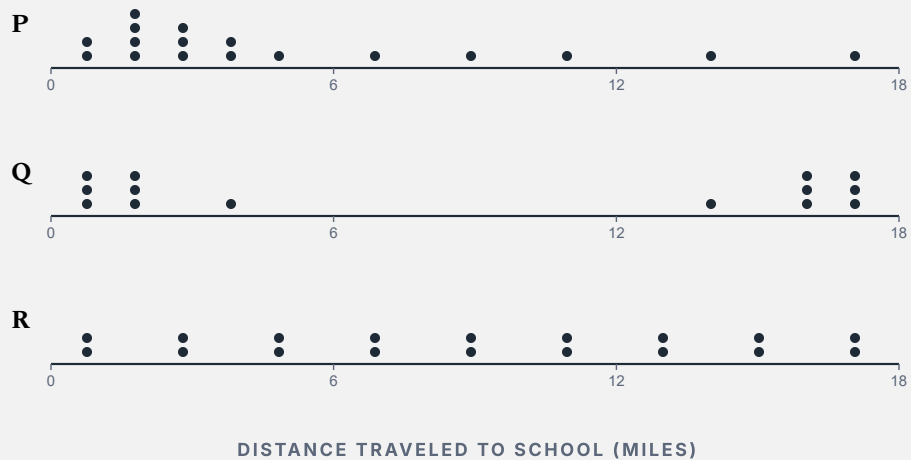
- a) Build 9 scores that are skewed LEFT with a median of 8. List them in order and explain how you know.

- b) Build a DIFFERENT set of 9 scores that is roughly symmetric, has median 8, minimum 4, and maximum 10.

- c) Explain why it is IMPOSSIBLE to build 9 whole-number scores (0–10) that are strongly skewed RIGHT with a median of 10. Justify from the definitions of median and right skew.

USE THESE DISPLAYS FOR PROBLEM 91

Three dotplots (P, Q, R) of distance students travel to school (miles), plus three verbal clues.



Clue 1: Travel distance is fairly even across the whole range, no real pile-up. **Clue 2:** Almost everyone lives close; a few travel much farther, stretching one long tail. **Clue 3:** Two kinds of students, a near group and a far group, with a clear gap between them.

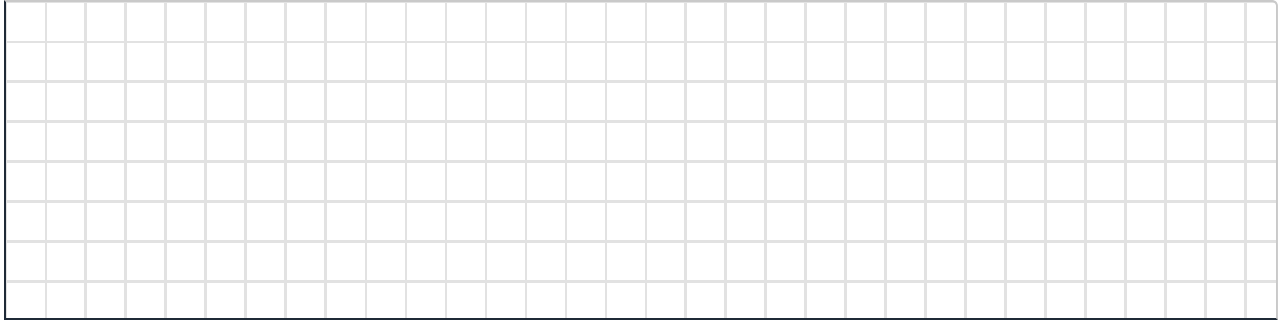
91

★★★★

Matching + Reasoning.

- a) Match each dotplot (P, Q, R) to a clue and name its shape (skewed right, bimodal, or approximately uniform).
- b) For the dotplot you identified as skewed right, is the MEAN likely larger or smaller than the median? Why?

- c) Sketch the rough shape of a histogram (bin width 4) for the bimodal set, and explain whether a histogram or a dotplot makes the gap easier to see for this small data set.



DISTANCE TO SCHOOL (MILES)

92

★★★★

Finishing times (minutes) for 30 runners in a school charity 5K:

```

1 | 8 9
2 | 0 2 2 3 4 4 5 6 6 7 8 9
3 | 0 1 1 2 2 3 3 4 5 7 8
4 | 2 5 9
5 | 1
6 |
7 | 4

```

Key: 2|3 = 23 minutes

- a) Write a complete description (shape, outliers, center, variability) in context. Address clusters/peaks and gaps in SHAPE, and say what the modes or outliers might represent (serious racers vs. casual walkers).

- b) The director claims "a typical runner finishes in under half an hour." Use the median (state its position and value) to evaluate this.

- c) Without computing it, argue whether the MEAN is greater than, less than, or about equal to the median, and justify from shape and extreme values.

Summarizing One Variable with Numbers

Mean, median & position (quartiles, percentiles); range, IQR & standard deviation; outliers ($1.5 \times \text{IQR}$ and 2-SD rules); resistance & choosing the right summary

93 Which of the following is a measure of *center* (not variability) of a quantitative variable?

★★★★

MC

- A) Range
- B) Interquartile range
- C) Standard deviation
- D) Median
- E) Variance

94 Hours of sleep last night for 7 students:

★★★★

6 7 5 8 7 6 9

a) Find the mean (round to two decimals).

b) Find the median.

c) Is the mean larger or smaller than the median here? Briefly say why.

95 Texts sent in the last hour by 8 students:

★★★★

12 9 15 22 8 14 19 11

a) Order the values and find the median.

b) Find the minimum, maximum, and range.

96 Daily steps (in thousands) recorded on 11 days:

★★★★

6 8 5 12 9 7 10 8 11 4 9

- a) Find the median, then the first quartile Q_1 and third quartile Q_3 .

- b) Find the IQR and interpret it in context (one sentence).

97

★★★

Prices paid (\\$) by 7 students for their sneakers, already ordered:

60 65 70 72 75 80 240

- a) Find the range and the IQR.

- b) Which one better describes the *typical* spread of these prices, and why?

98

★★★

Five quiz scores (out of 10): **4, 6, 8, 7, 5**. Find the standard deviation *by hand*.

- a) Find the mean.
b) Make a table of each deviation (value – mean) and its square; add the squares.
c) Divide by $n - 1$ to get the variance, then take the square root for the standard deviation s .

VALUE	DEVIATION (VALUE – MEAN)	SQUARED DEVIATION
4		
6		
8		
7		
5		
SUM		

99

★★★

In an ordered list of daily screen-time values for 20 students, your value is the 18th from the bottom (18 of the 20 values are at or below yours).

a) What percentile is your screen time?

b) Interpret that percentile in context (one sentence).

c) The median of a data set is always which percentile?

100

★★★☆☆



A pediatrician says a baby's length is at the **30th percentile** for their age. This means:

- A) About 30% of babies this age are this length or shorter.
- B) The baby is 30% as long as an average baby.
- C) About 30% of babies this age are longer.
- D) The baby is 30 units below average.
- E) Exactly 30 babies are shorter than this one.

101

★★★★☆

Messages posted in a class group chat by 9 students:

18 22 14 12 20 16 24 10 60

a) Give the five-number summary (min, Q_1 , median, Q_3 , max).

b) Find the IQR and the lower and upper fences (the **$1.5 \times \text{IQR}$** rule).

c) Identify any outliers.

102

★★★★☆

In a track club, 100-meter times have mean $\mu = 14.0$ s and standard deviation $\sigma = 1.2$ s. Use the **2-standard-deviation rule** (a value more than 2 SD from the mean is a potential outlier).

a) Find the lower and upper cutoffs.

b) Is a time of 17.0 s a potential outlier? Justify.

c) Is a time of 11.5 s a potential outlier? Justify.

103

★★★

A set of quiz scores (out of 20) has **mean 16** and **standard deviation 2**.

a) The teacher rescales to a 100-point quiz by *multiplying every score by 5*. What are the new mean and new standard deviation?

b) Instead, suppose she *adds 4 points* to every original score as a curve. What happens to the mean? To the standard deviation?

104

★★★

Five test scores: **78, 82, 85, 88, 90**. A sixth student's score of **30** is then added.

a) Find the mean and median of the original five, then recompute both after adding the 30.

b) Which measure changed more?

c) Explain "resistant" vs. "non-resistant," and name a resistant measure of center and of variability.

105

★★★

For each variable, would you report the **mean & standard deviation** or the **median & IQR**? Justify each choice in a few words.

a) Yearly household incomes in a town (strongly skewed right).

- b) Heights of students in a class (roughly symmetric, no outliers).

- c) Sale prices of homes in a neighborhood that includes a few mansions.

106

★★★★

Battery life (hours) for two phone brands, 5 phones each:

Brand A:	10	11	12	12	15
Brand B:	8	11	12	13	16

- a) Find the mean and range for each brand.

- b) The two brands have the same center. Which is more *consistent*, and how do you know?

107

★★★★

Error analysis. A student finds the standard deviation of **2, 4, 6, 8** like this: “The mean is 5. The deviations are -3, -1, 1, 3. They add to 0, so the standard deviation is 0.”

- a) Identify the mistake.

- b) Find the correct standard deviation (round to two decimals).

108

★★★★

Always, Sometimes, or Never? For any you mark *Sometimes*, give a one-line example or condition.

- a) A standard deviation is negative.

- b) The mean of a data set is greater than its median.

- c) Adding the same constant to every value increases the standard deviation.

- d) The range is at least as large as the IQR.

109

★★★★★

Commute times (minutes) for 10 students:

5 8 10 10 12 15 15 18 20 45

- a) Find the mean and the five-number summary.

- b) Use the $1.5 \times \text{IQR}$ rule to check for outliers.

- c) Is the mean or the median a better “typical” commute here? Justify from shape/outliers.

- d) Which measure of variability should be reported alongside it, and why?

Picturing the Summary: Boxplots

The five-number summary; building and reading boxplots; modified boxplots that flag outliers; reading shape and mean-vs-median from a boxplot

110

★★★★

A data set, already ordered:

3 5 6 8 9 10 12

a) Write the five-number summary.

b) Find the IQR.

111

★★★★

In a boxplot, the *box* (from Q_1 to Q_3) represents

- A) the middle 50% of the data.
- B) all of the data values.
- C) the mean plus or minus one standard deviation.
- D) the lowest 25% of the data.
- E) the range of the data.

MC

112

★★★★

A data set has five-number summary **min = 2, Q_1 = 5, median = 8, Q_3 = 12, max = 20** (no outliers). Draw and scale a number line from 0 to 20 and construct the boxplot.

SCALE YOUR NUMBER LINE EVENLY FROM 0 TO 20

113

★★★★

Minutes spent on homework by 11 students:

6 8 9 10 11 12 13 14 15 16 38

- a) Give the five-number summary and the IQR.
- b) Use the $1.5 \times \text{IQR}$ rule to identify any outlier(s).

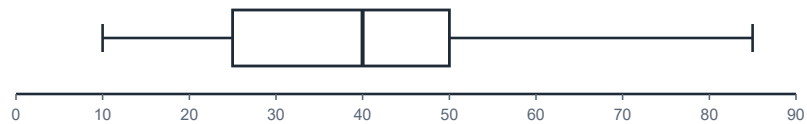
- c) Draw a **modified boxplot**: whiskers reach only to the most extreme values that are *not* outliers, and outliers are marked separately.

MINUTES ON HOMEWORK

114

★★★

The boxplot shows daily active minutes on a fitness app for one class.



DAILY ACTIVE MINUTES

- a) Read off the five-number summary.

- b) Find the IQR and the range.

- c) About what percent of the class is above 50 active minutes?
-

115

★★★



On a boxplot, the left whisker and the left half of the box are short, while the right half of the box and the right whisker are long. The distribution is most likely

- A) skewed to the right.
- B) skewed to the left.
- C) roughly symmetric.
- D) bimodal.
- E) uniform.

116

★★★★

A distribution of home prices is strongly skewed to the right. Which statement about its mean and median is correct?

MC

- A) The mean is greater than the median.
- B) The mean is less than the median.
- C) The mean equals the median.
- D) The mean and median are both 0.
- E) You cannot compare them without the standard deviation.

117

★★★★

Matching. Match each distribution shape to the boxplot feature you would expect. Write the letter for each.

1. Roughly symmetric
 2. Skewed right
 3. Skewed left
- A. Median sits left of center; the right whisker is much longer.
 - B. Median near the middle of the box; whiskers about equal.
 - C. Median sits right of center; the left whisker is much longer.

118

★★★★

Error analysis. A data set has $Q_3 = 30$, $IQR = 8$, and a maximum value of 55. The $1.5 \times IQR$ upper fence is 42, so 55 is an outlier. A student drew the right whisker all the way out to 55 anyway.

- a) Explain what is wrong with extending the whisker to 55.

- b) Describe how the modified boxplot should be drawn instead.

119

★★★★

Eight values: **4, 7, 9, 10, 12, 13, 15, 20**. Which is the correct five-number summary?

MC

- A) 4, 8, 11, 14, 20
- B) 4, 9, 11, 14, 20
- C) 4, 7, 11, 15, 20
- D) 4, 8, 11, 15, 20
- E) 4, 9, 12, 13, 20

120

★★★★

Which feature of a distribution can a boxplot *not* reveal?

MC

- A) The median
- B) The interquartile range
- C) The maximum value
- D) Potential outliers
- E) The presence of two distinct peaks (bimodality)

121

Hours of screen time yesterday for 10 students:

★★★

1 2 2 3 4 4 4 5 6 9

- a) Find the five-number summary and check for outliers with the $1.5 \times \text{IQR}$ rule.
- b) Draw and scale a number line and construct the boxplot.

HOURS OF SCREEN TIME

122

★★★

Dollars spent on a class trip by 11 students:

12 15 18 20 21 22 24 25 26 28 60

- a) Find the five-number summary and IQR.
- b) Use the $1.5 \times \text{IQR}$ rule to identify any outlier(s) and state where each whisker ends.

- c) Draw the modified boxplot and describe the shape in context.

DOLLARS SPENT

Comparing Distributions & Relative Position

Comparing two or more distributions with dotplots, stem-and-leaf plots & parallel boxplots; comparing with summary statistics; z-scores & relative position

123 When you compare two distributions of a quantitative variable, a complete answer should compare

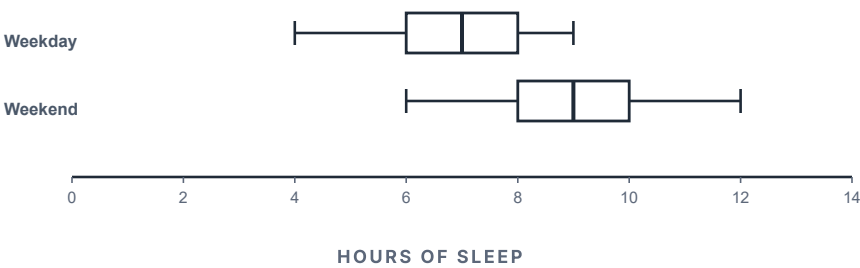
★★★★



- A) only the centers.
- B) only which data set has more values.
- C) only the means, never the medians.
- D) only the maximum values.
- E) shape, center, variability, and outliers, all in context.

124 Parallel boxplots show hours of sleep for the same students on a weekday vs. a weekend night.

★★★★



a) Compare the centers (medians) in context.

b) Compare the variability (IQR and range).

c) Write a one-sentence conclusion about weekday vs. weekend sleep.

125 A back-to-back stem-and-leaf plot shows resting heart rate (bpm) for student athletes and non-athletes.

★★★★

Athletes		stem		Non-athletes
8 6		5		
9 7 4 2		6		2 5
5 3		7		1 3 4 6 8
0		8		0 2 5

Key: 2 | 6 | 5 means an athlete 62 bpm and a non-athlete 65 bpm

- a) Compare the centers of the two groups in context.

- b) Compare shape and variability.

- c) What does the plot suggest about athletes' resting heart rates?

126

★★★☆☆

Test scores for two classes of 5 students each:

Class P: 70 75 80 85 90
Class Q: 60 75 80 85 100

- a) Find the median and range of each class.

- b) The medians are equal. Which class is more variable, and what does that mean here?

127

★★★★☆

Summary statistics for the battery life (hours) of two headphone brands:

BRAND	MEAN	SD	MEDIAN	IQR
A	22	2	22	3
B	25	5	24	8

- a) Which brand lasts longer on average, and by about how much?

- b) Which brand is more consistent? Cite a statistic.

c) If you wanted the most predictable battery life, which would you choose and why?

128

★★★

Error analysis. Asked to compare two distributions of daily steps, a student writes only: "Group A is bigger."

a) List what is missing from this comparison.

b) Rewrite it as a proper comparison, using comparison words and naming center, variability, and context (invent reasonable values).

129

★★★

On a math test, scores have mean $\mu = 70$ and standard deviation $\sigma = 10$. A student scores **85**.

a) Compute the student's z-score.

b) Interpret it in one sentence.

130

★★★

A runner's time has a z-score of **-1.2**. Which statement is correct?

- A) The time is 1.2 seconds below the mean.
- B) 1.2% of times are faster.
- C) The time is 1.2 standard deviations above the mean.
- D) The time equals the mean.
- E) The time is 1.2 standard deviations below the mean.

MC

131

★★★

Heights in a class have mean $\mu = 60$ in and standard deviation $\sigma = 8$ in. A student's height has z-score **-0.5**. What is the student's height?

132

★★★

Two seniors compare college-admissions scores from different tests.

- a) Amara scored **1300** on the SAT (mean 1000, SD 200). Find her z-score.

- b) Ben scored **28** on the ACT (mean 21, SD 5). Find his z-score.

- c) Relative to other test-takers, who performed better? Explain why z-scores let you compare across the two tests.

133

★★★

Amir scored **87** on a test with mean 75 and SD 8. In a different class, Bella scored **89** on a test with mean 82 and SD 4.

- a) Find each student's z-score.

- b) Who did better *relative to their own class*? Justify with the z-scores.

134

★★★

Always, Sometimes, or Never? For any you mark *Sometimes*, give a one-line condition.

- a) A z-score is negative.

- b) A value equal to the mean has a z-score of 0.

- c) Within the *same* distribution, a larger data value has a larger z-score.

- d) A z-score is measured in the same units as the original data.

135
★★★★

Maya's 5K time places her at the **15th percentile** of her team, and her time has a z-score of **-1.04**. (Remember: for race times, *smaller is faster*.)

a) Explain what the 15th percentile says about Maya's standing on the team.

b) Explain why a fast runner has both a *low* percentile and a *negative* z-score here.

136
★★★★

Two coding clubs report the minutes students took to finish a challenge.

CLUB	N	MEDIAN	IQR	MEAN	SD
Morning	24	18	6	19	5
After-school	24	25	10	27	9

a) Compare center and variability of the two clubs in context.

b) For each club, is the distribution more likely symmetric or skewed right? Use mean vs. median to argue.

c) A Morning student finished in 9 minutes; an After-school student finished in 18 minutes. Using z-scores, who was faster *relative to their own club*?

137
★★★★

Two PE classes recorded the number of push-ups students could do.

Class 1: 10 12 14 15 16 18 20 21 22 40
Class 2: 18 19 20 22 23 24 25 26 27 28

a) Find the median and IQR for each class, and check Class 1 for outliers with the $1.5 \times \text{IQR}$ rule.

- b) Compare the two distributions (shape, center, variability, outliers) in context.

- c) Jordan did 22 push-ups in Class 1; Riley did 26 in Class 2. Class 1 has mean 18.8 and SD 8.2; Class 2 has mean 23.2 and SD 3.3. Find each z-score and decide who ranks higher *within their own class*.

- d) Which class performed better overall? Justify using both center and the effect of the outlier.

When Graphs Lie

Truncated axes; pictographs; pooled bins; how a display can distort an honest one-variable data set

138 A bar chart's vertical axis begins at 40 instead of 0. The most likely effect is that the graph

★★★★



- A) makes all categories look equal.
- B) converts the counts into proportions.
- C) understates the differences between the bars.
- D) has no effect on how the bars are perceived.
- E) exaggerates the differences in height between the bars.

139 A bar graph shows the distribution of favorite subject for a sample of 1000 students, but its vertical axis begins at 100. What is the most serious problem?

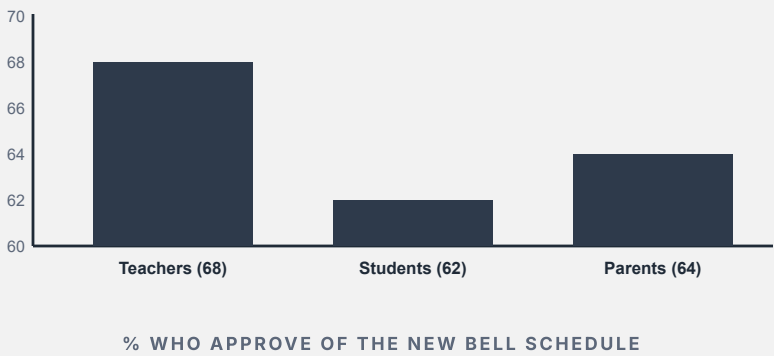
★★★★



- A) The subjects are not listed in the correct order.
- B) This distribution should be displayed with a pie chart.
- C) The vertical axis should show the percentage of students.
- D) The vertical axis should start at 0 rather than 100.
- E) The foreign-language bar should be broken up by language.

USE THIS GRAPH FOR PROBLEM 140

The school news account posted a bar chart titled “Percent who approve of the new bell schedule.” **The vertical axis runs from 60 to 70** (note the truncated scale).



140

★★★★

Error Analysis. (a) Identify the specific feature that makes this graph misleading. (b) Explain the false impression a casual reader would get. (c) Describe how you would fix the graph, and what the corrected bars would look like.

141

★★★★

An ad displays a pictograph comparing monthly users of two apps. App X has 20 million users (a small phone icon); App Y has 40 million users (a phone icon twice as tall *and* twice as wide). (a) App Y has how many times as many users as App X? (b) Explain why the picture makes the difference look much larger than it really is. (c) State the general rule this illustrates.

142

★★★★



A pictograph compares followers gained by two creators by drawing phone images scaled in both height and width. Creator B gained twice as many as Creator A and is drawn twice as tall and twice as wide. B's image has about how many times the *area* of A's image?

- A) 1 time
- B) 2 times
- C) 3 times
- D) 4 times
- E) 8 times

143

★★★

Always / Sometimes / Never. Write **A**, **S**, or **N** with a one-line reason.

- a) Starting a bar chart's axis above 0 exaggerates differences between bars.

- b) Scaling a picture symbol in both height and width keeps the comparison fair.

- c) To compare how a categorical variable is distributed across two groups of very different sizes, a bar chart of raw counts gives a fair comparison.

- d) Two distributions shown side by side should be drawn on the same axis scales so the comparison is fair.

144

★★★★



A histogram of daily screen-time (minutes) looks visually symmetric EXCEPT that its single rightmost bin, "300 to <400," secretly pools every student using 300 minutes or more, including a few over 1,000. Treating that pooled bin honestly, the distribution is best described as

- A) roughly symmetric, because the bars look balanced as drawn.
- B) skewed to the LEFT, because the tallest bars are on the right.
- C) approximately uniform, because most bins have similar height.
- D) skewed to the RIGHT, because the pooled bin hides a long tail of very large values.
- E) bimodal, because there are two regions of tall bars.

145

★★★★



A post compares average daily screen time for two age groups. The vertical axis runs from 3.0 to 4.0 hours. Group X's bar tops at 3.2 and Group Y's at 3.8. A reader says "Group Y uses about three times as much screen time." The reader's "3 times" impression is wrong because the true ratio is closest to

- A) 3.0 times, so the reader is correct.
- B) 6.0 times, since the bars differ by 0.6 and the axis spans 1.0.
- C) 0.84 times, because Group X is actually larger.
- D) 1.19 times ($3.8/3.2$); the truncated axis (starting at 3.0) inflated a 0.6-hour gap into an apparent 3-to-1 difference.
- E) cannot be determined without the sample sizes.

146

★★★★

Capstone critique. A school store's "snack sales" infographic for one week shows the number of bags sold of each snack with three design choices: (i) the bar chart's vertical axis starts at 200, not 0; (ii) the top-selling snack is drawn as a chip-bag icon scaled to be 3× as tall *and* 3× as wide as the others; and (iii) a companion pie chart's slice labels read 30%, 30%, 25%, 20%, and 15%.

- a) For each of the three choices, name the specific distortion and the false impression it creates.

b) The pie chart cannot be a valid display of these data. Show why with a one-line calculation.

c) Describe an honest redo of the bar chart, stating what changes and what each axis should show.
