Chapter 1 Study Guide: Exploring Data
WITH TI84 PLUS

ACTIVITY 1.2

Includes TPS 1.53+ 1.105

Individuals: The objects described by a set of data. (*ex: people, animals, things…*)

Variable: Any characteristic of an individual.

Categorical Variable: Places an individual into groups or categories. (*ex: colors*)

Quantitative Variable: Variables that are numbers. (*ex: height, weight*)
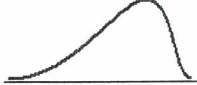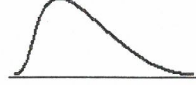
Describe Distribution:
Shape: symmetric or skewed
Outliers: value that is outside pattern
Center: mean/median
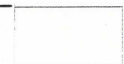Spread(Variability): range
} + CONTEXT

Compare Distributions:
Use comparative words
(similar, greater than, less than)

SHAPE of a distribution: Use –ly words (slightly, moderately, strongly)

Symmetric	Skewed Left	Skewed Right
Mean = Median	Mean < Median	Mean > Median

Other words to describe shape:

Unimodal	Bimodal	Uniform

CENTER of a distribution:
~**Mean:** ⟶ Use with symmetric data
~**Median:** middle point of a distribution $\left(location = \frac{n+1}{2}\right)$ ⟶ Use with skewed data

SPREAD/VARIABILITY of a distribution:
~**Range** = max − min
~ **Standard Deviation** = $\sqrt{variance}$ ⟶ The average distance from the mean
⟹ The __(context)__ typically varies by __(SD)__ from the mean by __($\bar{x}$)__ .
~**IQR** = IQR = $Q_3 - Q_1$

Resistant: A measure that is unaffected by extreme values (ex-median)
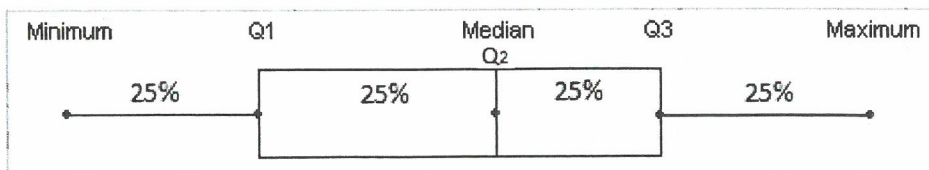~Skewed data or outliers → Use median and IQR
~Symmetric data → Use mean and standard deviation

OUTLIERS:
LFence: $Q_1 - 1.5(IQR)$
UFence: $Q_3 + 1.5(IQR)$

5 Number Summary: MIN Q_1 MED Q_3 MAX ⟶ Use to make boxplot
Outliers: way too small < $Q_1 - 1.5 IQR$ and way too big > $Q_3 + 1.5 IQR$
Boxplots:

53. **Traveling to work** How long do people travel each day to get to work? The following table gives the average travel times to work (in minutes) for workers in each state and the District of Columbia who are at least 16 years old and don't work at home.[30]

AL	23.6	LA	25.1	OH	22.1
AK	17.7	ME	22.3	OK	20.0
AZ	25.0	MD	30.6	OR	21.8
AR	20.7	MA	26.6	PA	25.0
CA	26.8	MI	23.4	RI	22.3
CO	23.9	MN	22.0	SC	22.9
CT	24.1	MS	24.0	SD	15.9
DE	23.6	MO	22.9	TN	23.5
FL	25.9	MT	17.6	TX	24.6
GA	27.3	NE	17.7	UT	20.8
HI	25.5	NV	24.2	VT	21.2
ID	20.1	NH	24.6	VA	26.9
IL	27.9	NJ	29.1	WA	25.2
IN	22.3	NM	20.9	WV	25.6
IA	18.2	NY	30.9	WI	20.8
KS	18.5	NC	23.4	WY	17.9
KY	22.4	ND	15.5	DC	29.2

(a) Make a histogram of the travel times using classes of width 2 minutes, starting at 14 minutes. That is, the first class is 14 to 16 minutes, the second is 16 to 18 minutes, and so on.

(b) The shape of the distribution is a bit irregular. Is it closer to symmetric or skewed? About where is the center of the data? What is the spread in terms of the smallest and largest values? Are there any outliers?

(Handwritten notes)

STUDY TIP FOR CH 1 TEST

TPS 1.53 – See my website for notes on CALC COMMANDS

Goal: become familiar with Calculator.

STEP 1: FIND THE MEAN, STANDARD DEVIATION AND SAMPLE SIZE AND USE THE CORRECT VARIABLE NOTATIONS

STEP 2: FIND THE 5 NUMBER SUMMARY

STEP 3: CALC. IF THEIR ANY OUTLIERS

STEP 4: BASED ON THE MEAN AND MEDIAN WHAT DO YOU THINK THE SHAPE WILL BE?

PART A – Complete and add a boxplot
PART B – Complete and answer w/ CUSS and BS

ANSWER KEY TO GIVEN QUESTIONS

53. **Traveling to work** How long do people travel each day to get to work? The following table gives the average travel times to work (in minutes) for workers in each state and the District of Columbia who are at least 16 years old and don't work at home.[30]

AL	23.6	LA	25.1	OH	22.1
AK	17.7	ME	22.3	OK	20.0
AZ	25.0	MD	30.6	OR	21.8
AR	20.7	MA	26.6	PA	25.0
CA	26.8	MI	23.4	RI	22.3
CO	23.9	MN	22.0	SC	22.9
CT	24.1	MS	24.0	SD	15.9
DE	23.6	MO	22.9	TN	23.5
FL	25.9	MT	17.6	TX	24.6
GA	27.3	NE	17.7	UT	20.8
HI	25.5	NV	24.2	VT	21.2
ID	20.1	NH	24.6	VA	26.9
IL	27.9	NJ	29.1	WA	25.2
IN	22.3	NM	20.9	WV	25.6
IA	18.2	NY	30.9	WI	20.8
KS	18.5	NC	23.4	WY	17.9
KY	22.4	ND	15.5	DC	29.2

(a) Make a histogram of the travel times using classes of width 2 minutes, starting at 14 minutes. That is, the first class is 14 to 16 minutes, the second is 16 to 18 minutes, and so on.

(b) The shape of the distribution is a bit irregular. Is it closer to symmetric or skewed? About where is the center of the data? What is the spread in terms of the smallest and largest values? Are there any outliers?

CH 1 STUDY TIP TPS 1.53

KEY X = STATE AVG. TRAVEL TIME TO WORK (MIN)

STEP 1: Sample Mean = $\bar{X}$ = 23.1

Sample S.D = S_X = 3.6

Sample size = n = 51

STEP 2:
MIN = 15.5
Q1 = 20.8
MED = 23.4
Q3 = 25.2
MAX = 30.9

IQR = 4.4
R = 15.4

STEP 3:
IQR = 25.2 − 20.8 = 4.4

UB: 1.5 * 4.4 + 25.2 = 31.8+

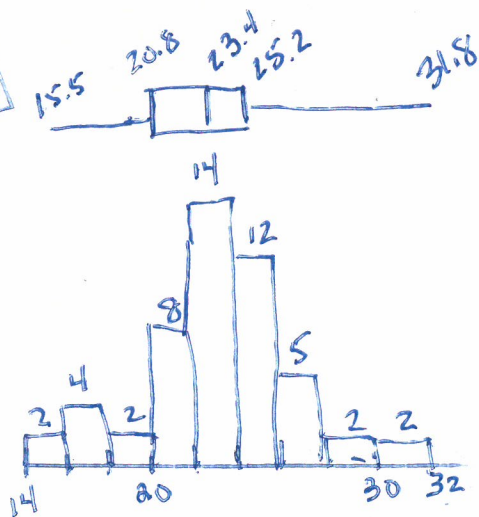LB: 20.8 − 1.5 * 4.4 = ∠14.2

NO OUTLIERS   MAX = 30.9 < 31.8
MIN = 15.5 > 14.2

STEP 4: Since the mean (23.1) and median (23.4) are close, I am expecting a somewhat symmetric shape.

Part A

15.5   20.8   23.4   25.2   31.8

PART B

FOR THE DISTRIBUTION OF TIME TRAVELED TO WORK, THE STATE DATA IS ROUGHLY SYMMETRIC. THE CENTER TRAVEL TIME IS ABOUT 23 minutes TO WORK WITH A SPREAD OF ABOUT 3.6 MINUTES. THE APPEARS TO BE NO OUTLIERS

FREQUENCY

14   12   8   5   4   2   2   2   2

14   20   30   32

TRAVEL TIME (MIN)

Need to label + include UNITS ON ALL GRAPHS

A

53

NOTE: GRAPHS DO NOT
NEED TO BE ARTIST
MASTER PIECES. IN STATS,
WE GRAPH TO GET A
VISUAL UNDERSTANDING
OF THE DATA. LABELS
ARE IMPORTANT.

Frequency

must
label

Leave space

Label
classes
at end pts

TRAVEL TIME (minutes) ←
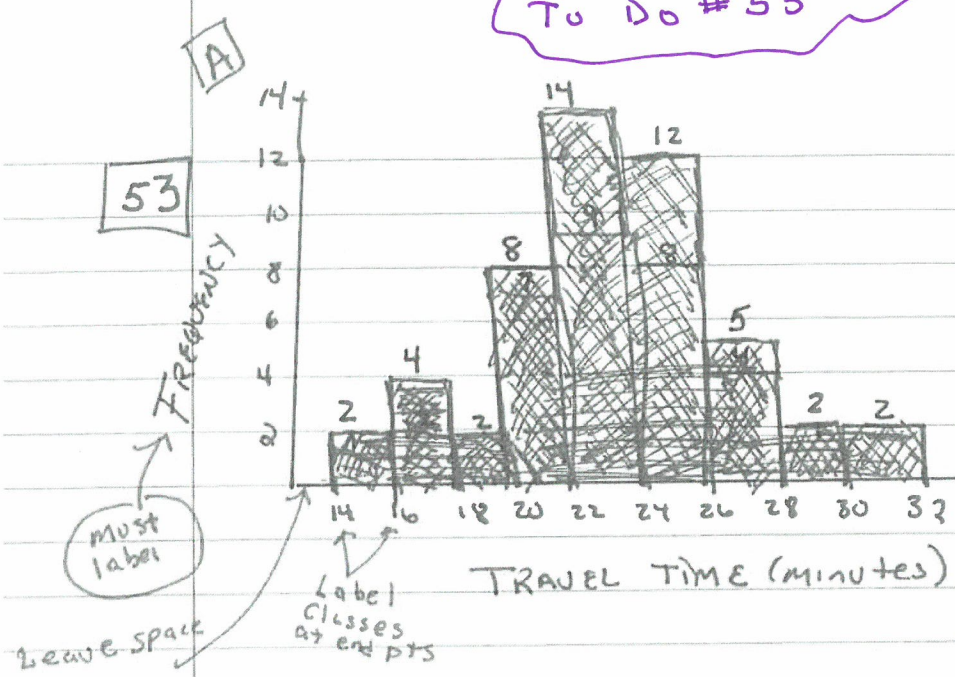
Must label +
Give units

Make sure to get to know your calculator. You
can use it on all tests. It's a GREAT tool.

STEPS: ① Put the data in a LIST (STAT) EDIT L1

② (ZOOM) (STAT:9) to see the graph

③ Set the (Window) to easily
graph $\begin{cases} Xmin = 14 \\ Xmax = 32 \\ Xscl = 2 \end{cases}$

the problem
ask you
to set up like this

$Ymin = 0$
$Ymax = 14$
$Yscl = 2$

these
may vary

③ USE (TRACE) TO GET FREQUENCY COUNTS
FOR EACH BAR (YOU DO NOT NEED TO INCLUDE
THESE ON YOUR GRAPH, BUT THEY CAN BE
HELPFUL INTERPRETING YOUR DATA.

④ GET STATISTICS (STAT) (CALC) 1: 1-VAR STATS (L1)

$\bar{X} = 23.1$ $S_x = 3.57$ $n = 51$ MIN $= 15.5$ MED $= 23.4$ MAX $= 30.9$

B   ALTHOUGH THE GRAPH IS SOMEWHAT IRREGULAR,
THE DISTRIBUTION IS ROUGHLY SYMMETRIC
FROM THE GRAPH THE CENTER IS ABOUT 23 MIN
THE SPREAD MEASURED BY THE RANGE IS ABOUT 15 MIN
THERE DO NOT APPEAR TO BE ANY OUTLIERS.

105. SSHA scores Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

pg 66

STEP 4

| 154 | 109 | 137 | 115 | 152 | 140 | 154 | 178 | 101 |
| 103 | 126 | 126 | 137 | 165 | 165 | 129 | 200 | 148 |

and for 20 first-year college men:

108	140	114	91	180	115	126
92	169	146	109	132	75	88
113	151	70	115	187	104	

Do these data support the belief that women have better study habits and attitudes toward learning than men? (Note that high scores indicate good study habits and attitudes toward learning.) Follow the four-step process. *

THE 4 STEP PROCESS IS A GUIDE TO HELP YOU AND YOU DO NOT NEED TO DO THE STEPS EXACTLY.

STEPS:

① (STATE) read question clearly and understand what you ARE ASKED TO ANALYZE. (mental step). What are you asked to ANALYZE? _____

_____

_____

② (PLAN/DO) What graphs + statistics do you need to SUPPORT YOUR CLAIM?

Ⓐ STATISTICS: To Provide include samplesize (n), mean ($\bar{x}$), Std dev (S_x) and the 5number summary

Ⓑ HISTOGRAMS: To see shape of distributions

Ⓒ BOX PLOTS: To identify outliers   or   do the calc's

$$LF = Q1 - 1.5\ IQR$$
$$UF = Q3 + 1.5\ IQR$$

Show work on BACK →

③ (CONCLUDE) ANSWER QUESTION IN CONTEXT.

① (STATE) The question is "Does the data for first year college students indicate women have better study habits and attitudes towards learning"

② (PLAN/DO) Ⓐ Sample statistics

→ [L1] women SSHA
→ [L2] men SSHA

(CALC—since sample sizes are different USE: 1-VAR STATS)

Group	n	$\bar{x}$	S_x	MIN	Q_1	Med	Q_3	MAX
Women	18	114	26	101	126	139	154	200
men	20	121	33	70	98	115	143	187

Ⓑ HISTOGRAMS (STACK THEM TO SEE DIFFERENCES AND USE SAME SCALE) AND SHAPE OF DISTRIBUTIONS)

Ⓒ BOX PLOTS (stack them + make sure to label and show scale, ALSO identify outliers

Graphs do NOT need to be but percise visually show (differences)

SSHA SCORE

Females

Females histogram: 4, 7, 5 over SSHA SCORE 75 100 125 150 175 200 225

MALES

Males histogram: 1, 4, 7, 4, 2, 2, 1 over SSHA SCORE 50 75 ... 225

Females box plot with * outlier
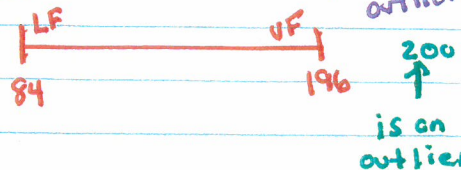MALES box plot
70 90 110 130 150 180 210   SSHA SCORE
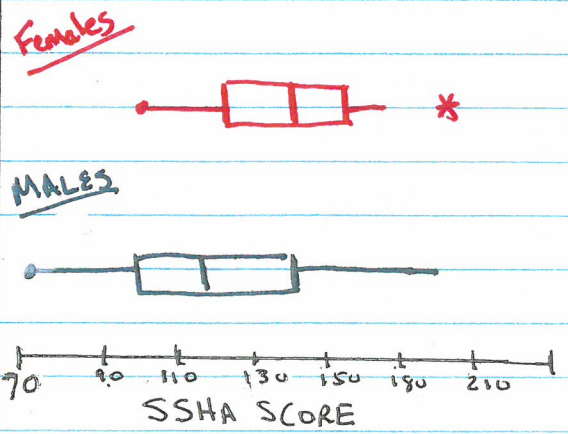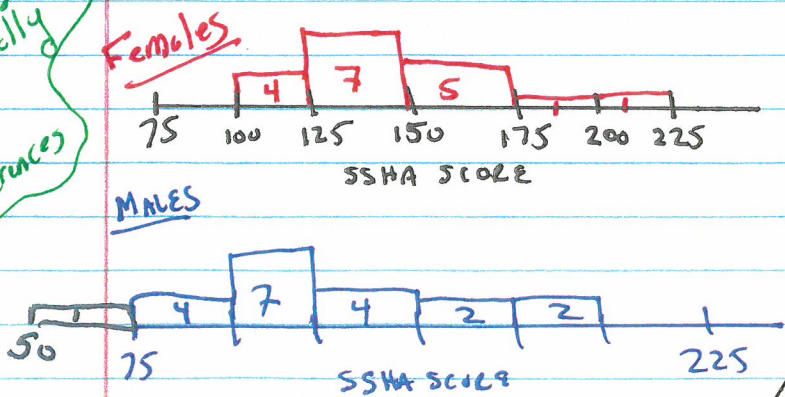
NOTE: TI 84 BOX PLOTS INDICATE OUTLIERS WITH A * OTHERWISE YOU NEED TO CALC:

Females   IQR = 154 - 126 = 28
LF: 126 - 1.5(28) = 84 (NO OUTLIER)
U.F: 154 + 1.5(28) = 196 (OUTLIER - 200)

Visually   ANY DATA OUTSIDE THE FENCES ARE outliers
LF ———————— UF
84   196
200 ↑ is an outlier

③ CONCLUDE:

GUIDELINES FOR COMPARISONS:
    You must compare Males and Females
    Shape, Center, Spread, outliers
    IN CONTEXT.

From the statistics and graphs, you can see 1st year female college students have higher SSHA (STUDY HABITS AND ATTITUDE) score THAN MALES.

① The ^Female center is higher than males for SSHA (based on mean, median, box plots)

② The Female SSHA spread is smaller than for males (based on SD, histograms, boxplots) indicating their is less variability among female SSHA scores.

③ Both Females and Males appear to have symmetric shape (based on histograms)