Unit 2: Stemplots

PREREQUISITES

Stemplots require familiarity with place value in the number system. There are no statistical prerequisites.

ADDITIONAL TOPIC COVERAGE

Additional coverage of stemplots can be found in *The Basic Practice of Statistics*, Chapter 1, Picturing Distributions with Graphs.

ACTIVITY DESCRIPTION

Students are often more invested if they have an opportunity to analyze data that they have collected. Use the questions suggested in the Unit 2 Activity survey (these questions are listed below for your reference) or create alternative questions of your own.

MATERIALS

Survey questionnaire, one copy per student.

Prior to handing out the survey, ask students to wait a moment while you get things ready. Take your time – so that students have to wait a few moments. (This wait time is related to question 1.) Then hand out the survey.

Once students have answered the survey, they will need to turn in their responses. Combine student responses to each question into a single table (See Table T2.1 in the activity solutions for an example.) Make sure that the same units are used by all students – for example, height in inches. These data will be revisited in Unit 5's activity (boxplots), so save these data. (If you decide not to collect data from your class, use the sample data from Table T2.1 instead.)

As students make stemplots of the data on each variable, encourage them to experiment with using different stems. Sometimes it is helpful to expand the stems and other times it is helpful to truncate the data values and collapse the stems. The idea is to get a stemplot that reveals

information about the data. Below is a copy of the suggested survey questionnaire. Feel free to adapt or revise the questions.

Survey Questionnaire

1. How long (in seconds) did you wait while your instructor was getting ready for this activity?

2. How much money in coins are you carrying with you right now?

3. To the nearest inch, how tall are you?

4. How long (in minutes) do you study, on average, for an exam?

5. On a typical day, how many minutes do you exercise?

6. Circle your gender: Male Female

Return your answers to your instructor.

THE VIDEO SOLUTIONS

1. Sample answer: head circumference, upper arm circumference, foot length, foot width, height.

2. The foot length data was fairly symmetric with a single peak. The center was around 26.8 inches.

3. City miles per gallon (mpg).

4. There were outliers at the upper end of fuel efficiency. A few cars got great gas mileage.

5. The data for the 2012 models exhibited more spread. There were vehicles that were more fuel efficient (for example, the Prius) in 2012 compared to 1984, but there were vehicles that were less fuel efficient (for example, SUVs) in 2012 compared to 1984.

UNIT ACTIVITY SOLUTIONS

Sample solutions are based on the data in Table T2.1.

Question 1 Wait (sec)	Question 2 Coins (cents)	Question 3 Height (in)	Question 4 Study (min)	Question 5 Exercise (min)	Question 6 Gender
40	77	68	30	75	Male
40	62	67	60	20	Female
75	175	73	30	0	Male
40	189	72	20	45	Male
50	120	71	15	90	Male
40	54	68	45	75	Female
45	26	66	60	30	Female
45	145	75	30	30	Male
40	0	69	120	0	Female
35	0	71	45	60	Male
45	35	72	30	30	Male
45	47	64	15	45	Female
45	125	72	20	90	Male
45	55	71	30	0	Male
40	35	69	60	45	Male
45	78	63	45	90	Female
55	157	65	45	20	Male
40	225	62	75	40	Female
40	92	64	30	60	Female
50	85	62	60	0	Female
35	35	64	45	30	Female
45	59	66	45	90	Female
50	145	60	30	45	Female
30	137	70	30	30	Male
50	142	69	20	45	Male
45	62	69	30	60	Male

Table T2.1. Sample data from unit activity survey.

1. Sample answers based on sample data in Table T2.1.

Time (sec)

3	0
3	55
4	00000000
4	555555555
5	0000
5	5
6	
6	
7	
7	5

The stemplot for Time is single-peaked and fairly symmetric. The middle is somewhere in the 40s. There is one outlier at 75.

Money (cents) Leaf unit = 10

0	00
0	2333
0	4555
0	6677
0	89
1	
1	223
1	4445
1	7
1	8
2	
2	2

For the stemplot above of the money (cents), we truncated the pennies place. These data are not symmetric. There are two clumps of data and one single high value in the 220s.

Height (in)

6	0
6	223
6	4445
6	667
6	889999
7	0111
7	2223
7	5

The height data are somewhat mound-shaped and roughly symmetric. The middle is around 67. There are no outliers.

Study Time (min)

1	55
2	000
3	000000000
4	555555
5	
6	0000
7	5
8	
9	
10	
11	
12	0

These data do not appear symmetric. The study times are mostly under 45 minutes. There is at least one outlier, the largest of which is 120 minutes.

Exercise (min)

0	0000
1	
2	00
3	00000
4	055555
5	
6	000
7	55
8	
9	0000

There are some gaps in the data. There are some people who don't exercise and the same number who exercise, on average, 90 minutes per day. The middle appears around 45.

2. The time estimates from male students were more spread out than for female students. A middle value for the female students looks to be at around 40 seconds and for male students at around 45 seconds. One male student was an outlier, at 75 seconds.

Time

Female		Male
	3	0
5	3	5
00000	4	000
5555	4	55555
00	5	00
	5	5
	6	
	6	
	7	
	7	5

3. The change in male students' pockets split into two groups. One group was at the low end of the change carried by female students. However, the other group tended to have more change than the female students. One female was an outlier. She carried $2.25 in change, more than anyone else in the class.

Money

Female		Male
0	0	0
32	0	33
554	0	5
76	0	67
98	0	
4	1	
	1	223
	1	445
	1	7
	1	8
	2	
2	2	

EXERCISE SOLUTIONS

1. Sample answer: time to run one mile; time to complete an obstacle course; number of pull-ups completed without stopping, number of sit-ups completed without stopping; resting pulse rate (a low rate means more fit).

2. a.

2	25
3	45
4	1166679
5	449
6	0

b. It is roughly symmetric with the center at stem 4. It is unimodal.

c. The center is around 46. (There are 15 observations – the 7th, 8th, and 9th value of the ordered data are all 46.)

d. No. Although 60 is the largest data value it is not outside the pattern that includes two 54s and a 59. There is no gap between 60 and the other data values.

3. a.

46	99
47	99
48	25579
49	33345599
50	9
51	234557
52	033
53	99
54	268
55	5
56	348
57	01256
58	03456
59	02369

b. Sample answer: The distribution appears to have three peaks, one around the 490's, another around the 510's and a third around the 580's (or around the 570's – 590's). The distribution doesn't look very symmetric. At the lower end, there are few scores in the 460's and 470's and then the number of scores increases for the 480's and 490's. On the higher end, the pattern is reverse. There are a larger number of scores in the 590's, 580's and 570's and then the number of scores decreases in the 560's and 550's. The middle number is 523, which can serve as the center of the distribution. (Or students might suggest the middle is in the 520's because 24 scores are below this stem and 24 scores are above this stem.)

c.

Writing		Math
93	45	7
954	46	9
96654310	47	
97	48	79
9997751	49	00369
985	50	01112289
7631	51	13568
92	52	13568
6	53	79
765	54	1135
6431	55	09
973321	56	589
97553	57	023
	58	
1	59	1113
	60	2468
	61	27

d. Sample answer: The average Math SATs are more spread out than the average Writing SATs. The lowest Math SAT was 457 and the highest was 617, for a spread of 160 points. The lowest Writing SAT was 453 and the highest was 591, for a spread of 138 points. The center of the Writing SATs is in the 510's and the center of the Math SATs is in the 520's. (The actual middle number of the ordered data is 511 for the Writing SATs and 527 for the Math SATs.) There are gaps in the Math SAT data in the 470's and 580's; the average scores of 457 and 469 might be considered outliers. There is a gap in the 580's for the Writing SATs and 591 is a potential outlier. The Writing SATs appear multimodal, with a peak around the 470's and another around the 560's. Neither of the distributions appears to be roughly symmetric.

4. The distribution is not symmetric. The data is concentrated toward the lower numbers and then trails off as the numbers get larger. In other words, this show attracts a mostly young audience. The center is around 19 (the 22nd observation in the ordered data of 44 values). There is an outlier at 120, which is considerably larger than the second largest data value of only 65. Most likely this is a typo – maybe the person was 12 or 20 but certainly not 120. The other possibility is that someone was being funny and responded that he/she was 120.

0	566899
1	0012223445667789
2	001233467
3	0135
4	28
5	0025
6	05
7	
8	
9	
10	
11	
12	0

REVIEW QUESTIONS SOLUTIONS

1. a.

0	3344
0	55555556667788
1	02
1	79
2	01
2	68
3	
3	
4	
4	7
5	23
5	678
6	44
6	7788
7	01344
7	789
8	0
8	799
9	3

b. The states divide into two clusters, with students from one group of states participating in SATs at a very low rate and students in the second group participating at a much higher rate. The lower cluster varies from 3% to 28% and the upper cluster from 47% to 93%. The lower cluster is concentrated around the single digits and then trails in the teens and twenties. The upper cluster is unimodal and roughly symmetric with its center at around 70%. (In some states, most college-bound students take the SATs. In other states, the rival American College Testing, or ACT, exams dominate and only students applying to selective colleges take the SATs. This explains, in part, the two clusters.)

c. Similar to the 2010/2011 data, the 1990 data breaks into two clusters. The lower cluster of the 1990 percentages has a similar spread to the 2010/2011 percentages and has roughly the same number of data values in the lower cluster. The gap between lower and upper clusters begins at the same stem for both years, but is slightly wider for the 2010/2011 data than for

the 1990 data. The upper cluster is reasonably symmetric in both years. However, the spread of the upper cluster is wider for the 2010/2011 data (47% to 93% for a difference of 46%) than for the 1990 data (42% to 74% for a difference of 32%). The center of the upper cluster for the 2010/2011 data is around 70%; the center for the 1990 upper cluster is only around 58%.

1990		2010-11
4	0	3344
99866555	0	555555566
4322221000	1	02
765	1	79
420	2	01
85	2	68
	3	
	3	
4422	4	
95	4	7
442	5	23
98875	5	678
4220	6	44
987	6	7788
420	7	01344
	7	789
	8	0
	8	799
	9	3

2. Sample answer: The Army should stock boots that fit foot widths from 90 millimeters to 113 millimeters. There was one outlier, a soldier with a foot width of 119 millimeters – 6 millimeters larger than the second largest foot width. The boot for that soldier should be specially ordered

(See stemplot on next page...).

9	011
9	2223
9	455
9	77
9	888
10	0001
10	222333
10	445
10	77
10	88
11	000011
11	3
11	
11	
11	9

3. a.

Leaf unit = 0.1

Girls		Boys
4	13	257
97744210	14	0458
9988531	15	45577
53	16	03
	17	356
	18	
	19	
	20	6
	21	
	22	8
	23	
	24	5
2	25	
9	26	

b. Sample answer: The overall pattern for boys' BMI is a flat mount shape that is roughly symmetric. There is a gap in the 18s and 19s and then three possible outliers: 20.6, 22.8, and 24.5.

c. Sample answer: The overall pattern for girls' BMI is mound-shaped and roughly symmetric. However, there appear to be two outliers: 25.2 and 26.9.

d. Sample answer: Ignoring the outliers identified in (b) and (c), the girls' data is less spread out than the boys' data. Just by eyeballing the data, the girls' data is centered around 15.1 and the boys' data is centered a little higher at about 15.5. The outliers in the girls' data are more extreme than the potential outliers in the boys' data.