Workshop Statistics: Discovery with Data, Second Edition

Topic 5: Measures of Spread

Activity 5-8: City Temperatures (*cont.*)

(a) Answers will vary from student to student.
(b) Minneapolis (22.16) > Chicago (18.87) > Pittsburgh (16.71) > Phoenix (14.87) > Atlanta (13.72) > San Antonio (13.05) > San Diego (5.64) > Honolulu (3.194)
(c) Answers will vary from student to student.
(d)

	range	IQR
Atlanta	38	28.25
Chicago	52	38.25
Honolulu	8	6.75
Minneapolis	62	45
Phoenix	40	28.5
Pittsburgh	46	33.25
San Antonio	36	25.75
San Diego	16	11

The ordering of the cities by the magnitude of their ranges, as well as the ordering of the cities by the magnitude of their interquartile ranges corresponds with the ordering by the magnitude of their standard deviations.
(e) Answers will vary from student to student.

Activity 5-9: Hypothetical Manufacturing Processes (*cont.*)

Process A has the smallest mean, followed by process C. Processes B and D have larger means which are probably pretty close to each other. The big difference between B and D is variability. Process D should have much less variability (so smaller SD). Thus A = process 3 (smallest mean), C = process 4 (next smallest mean), B = process 1, and D = process 2 (smaller SD).

Activity 5-10: Climatic Conditions

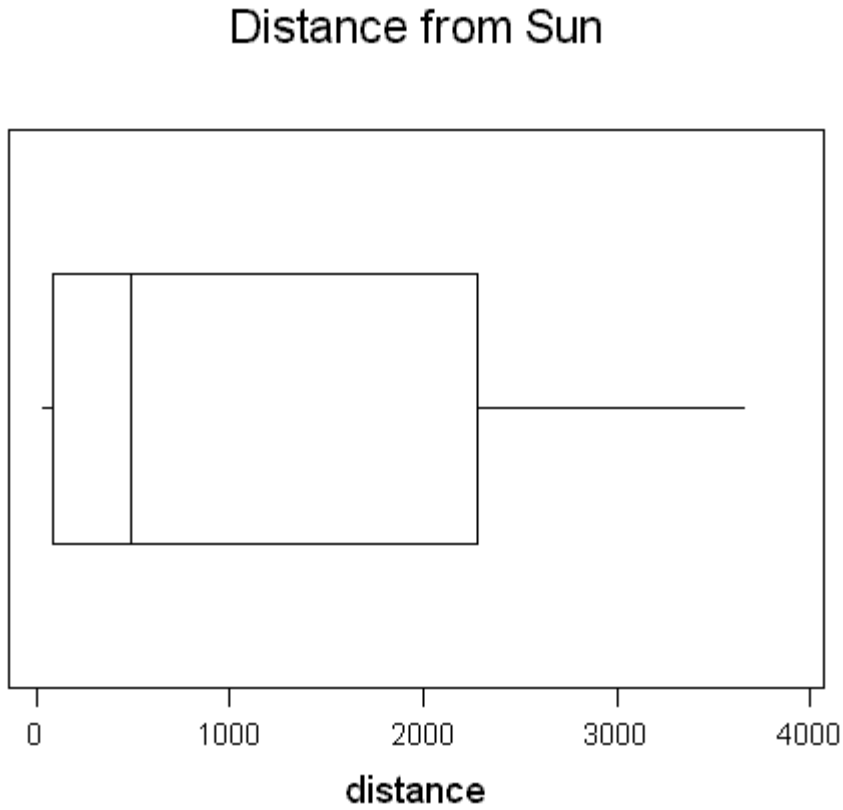
(a) January high IQR: 23; July high IQR: 6.9
(b) January high SD: 14.08; July high SD: 7.28

(c) The variable "January high" has greater variability in its distribution.
(d) July high temperatures are generally higher than January high temperatures.
(e) No, variability has to do with the spread of the values, not the relative size of the values themselves.

Activity 5-11: Planetary Measurements (*cont.*)

(a)

min	Q1	median	Q3	max
36	80	484	2278	3654



(b)
(c) skewed right

Activity 5-12: Word Lengths (*cont.*)

Answers will vary from class to class.

Activity 5-13: Tennis Simulations (*cont.*)

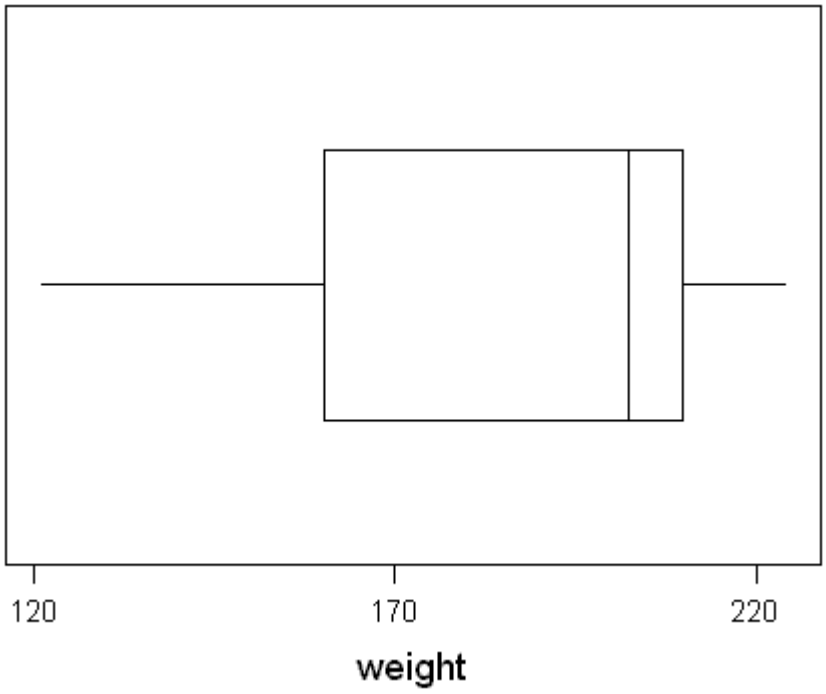
conventional scoring SD: 2.740
no-ad scoring SD: 1.022

This confirms what we saw earlier, that there is a larger spread with conventional scoring than with no-ad scoring.

Activity 5-14: Rowers' Weights (*cont.*)

(a)

min	Q1	median	Q3	max
121	160	202.5	210	224



(b)

(c) Boxplot doesn't show any of the clustering or individual values like the dotplot did.

Activity 5-15: Sudents' Data (*cont.*)

Answers will vary from student to student.

Activity 5-16: Mothers' Ages (*cont.*)

(a) The shape of this distribution is skewed right, not symmetric, so the empirical rule does not apply.
(b) mean: 22.257; SD: 5.622
(c) .71 This is 71%, while the empirical rule would predict 68%.
(d) .97 This is 97%, while the empirical rule would predict 95%.
(e, a) The distribution is still skewed to the right.
(e, b) mean: 21.735; SD: 4.769

(e, c) .705 This is about 71%, while the empirical rule would predict 68%.
(e, d) .94 This is 94%, while the empirical rule would predict 95%.

The numbers for one SD didn't change enough to make a difference, but two SDs dropped 3 percentage points. Without the high outlier to increase the SD, less data can fit within a given number of SDs.

Activity 5-17: Ideal Temperatures

Answers will vary from class to class.

Activity 5-18: GPAs

(a) Ted: z=(3.25-2.8)/.33=1.36 using the information for Washington High
Frank: z=(3.17-2.7)/.37 = 1.27 using the information from Jefferson High
Ted, with his higher z-score, did better relative to his peers than Frank did.
(b) For Ted to have a z-score of 1.127, need (x-2.8)/.33 = 1.27
Solving for x, x=.33(1.27)+2.8 = 3.22
(c) Torsten happens to be exactly one standard deviation above the mean. So we know 68% of the data are in the interval from one standard deviation below the mean to one standard deviation above the mean from the empirical rule. Thus, 32% of the data are outside that interval, and by symmetry, half is above. So 16% of GPAs are above Torsten's.

Activity 5-19: Heights of Volleyball Players

(a) 95% of heights should be within 2 standard deviations, so with 2(2.1)=4.2 inches, of the mean. So in between 69-4.2=64.8 and 69+4.2 = 73.2 inches.
(b) .997 of the players should be between 62.7 inches and 75.3 inches.
(c) Most observations are within 3 standard deviations, so we wouldn't expect a height taller than 69+3(2.1) = 75.3 inches

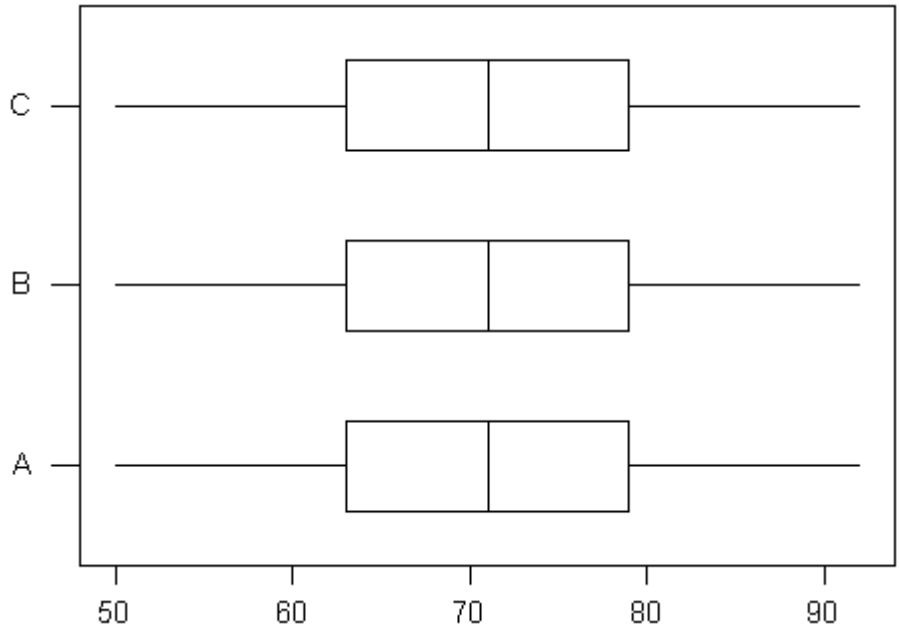
Activity 5-20: Guessing Standard Deviations

Answers will vary from student to student.

Activity 5-21: Limitations of Boxplots

(a) These dotplots reveal differences in the distributions of the three classes' scores. For instance, the scores from class B are fairly evenly distributed, while the other two classes have obvious clusters.
(b)

	min	Q1	median	Q3	max
class A	50	63	71	79	92
class B	50	63	71	79	92
class C	50	63	71	79	92



(c)
(d) If we had been shown only the boxplots and not the data, we would not be able to tell the differences in the three distributions.

Activity 5-22: Creating Examples (*cont.*)

Answers will vary from student to student. Some hints:

(a) 1 1 2 4 5 6 7 9 9 10
(b) Lower and upper quartiles are the same, but not all the values are the same
(c) 4 4 4 4 4 6 6 6 6 6
(d) Scores evenly split between 0 and 100
(e) Lower quartile is 0 and upper quartile is 100 but are some values in-between 0 and 100

Activity 5-23: More Measures of Center

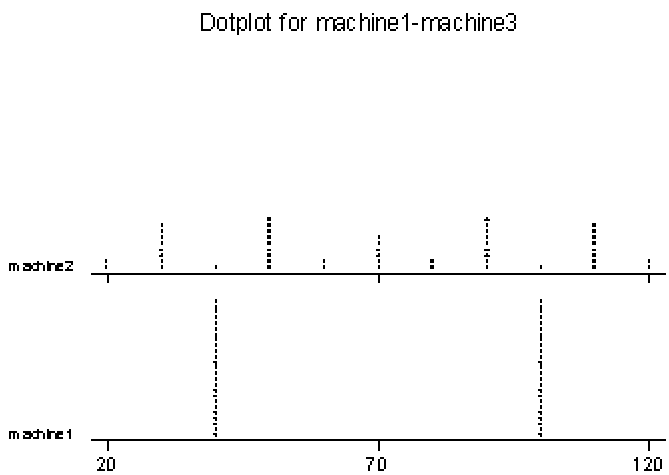
(a) Trimmed mean is resistant to outliers because it trims them off the data set before calculating the mean. Midhinge is also resistant to outliers because it does not include anything less than Q1 or greater than Q3 to affect its calculation. Midrange is not resistant to outliers because it is calculated based solely on the minimum and maximum values, which could be outliers.

(b) midrange: 16; midhinge: 13

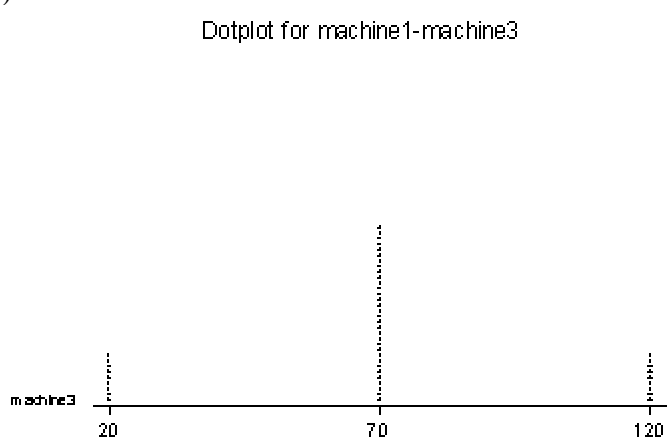
The mean was 13.44, and the median was 11. Midhinge is quite close to the mean, while midrange is way off from either the mean or the median.

(c) Answers will vary from class to class.

Activity 5-24: Hypothetical ATM Withdrawals



(a)



Each distribution is perfectly symmetric.

(b)

	mean	SD
machine 1	70	30.30
machine 2	70	30.30
machine 3	70	30.30

Mean and standard deviation are identical for each machine.
(c) The distributions are quite different. The first consists of just two values, fairly spread apart, the second is more bimodal with a wider range of values,
the third has a high peak in the middle with some extreme observations on both sides.
Thus the mean and standard deviation do not tell us everything about a
distribution.

Activity 5-25: Five on Five

(a)

min	Q1	median	Q3	max
1	1.5	3	4.5	5

(b) no

Activity 5-26: Properties of Measures of Spread (*cont.*)

(a) 32, 29, 23, 18, 16, 14, 13, 11, 10
(b) range: 22; IQR: 14; SD: 7.89
(c) The measures of spread for the new data are identical to those for the original data.
(d) range: 90; IQR: 30; SD: 19.29
original values: 18; 6, 3.859
(e) The new statistics are 5 times larger than the old statistics.
(f) The measures of spread will be 5/9 times the original measures.