

# Workshop Statistics: Discovery with Data, Second Edition

## Topic 3: Displaying and Describing Distributions

### Activity 3-6: Cars' Fuel Efficiency

- (a) The centers of distribution of these cars' MPG increases as the cars' size decreases (i.e. large < family < small).
- (b) There is a wide spread for sports, a medium spread for small, and a narrow spread for upscale.
- (c) There are probably 2 categories of cars represented in this dotplot. The centers of their clusters are at about 2500 and 4100, respectively.
- (d) There is a low outlier just about 14. The rest of the distribution does show a bimodal shape.

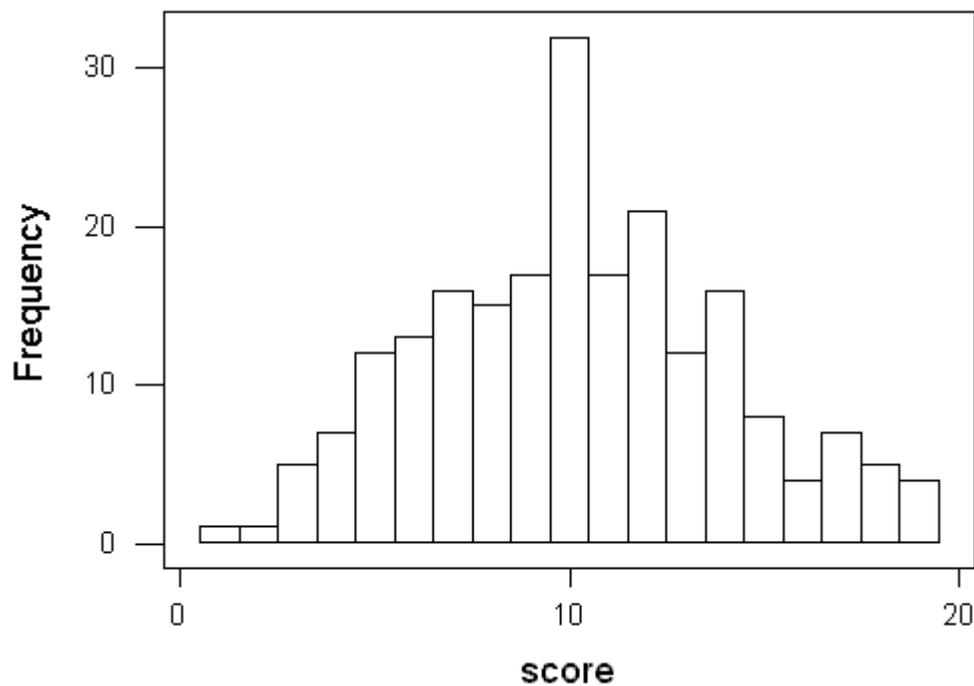
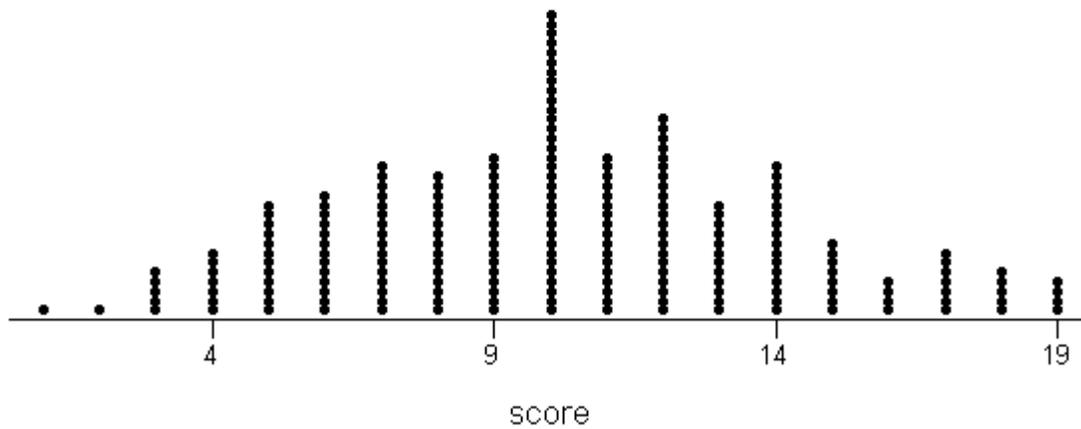
### Activity 3-7: Features of Distributions (*cont.*)

The histogram with the horizontal axis labeled "years" is skewed to the right. The histogram with the horizontal axis labeled "% urban" is roughly symmetric (maybe a little skewed to the left). The histogram with the horizontal axis labeled "price" is skewed to the left.

### Activity 3-8: Placement Exam Scores

(a)

Dotplot for score



(b) The shape is roughly symmetric, with the center of the distribution at a score of 10. The spread is from a score of 1 to 19.

(c) The number of people who scored a 10 is much higher than the next highest

subinterval (12). The placement scores are also all integer values (granularity).

### Activity 3-9: Hypothetical Manufacturing Processes

The center of distribution of process A is at about 11.5cm. There is not much variability in rod diameter in process A. The center of distribution of process B is at about 12.0cm. There is a lot of variability in rod diameter in process B. The center of distribution of process C is at about 11.7cm. There is a moderate amount of variability in rod diameter in process C. The center of distribution of process D is at about 12.0cm. There is only slightly more variability in rod diameter in process D than process A. Process D is best as is because its center of distribution is right on target, and it appears that the other points are within specifications. Process A is most stable because it has the least variability in rod diameter. Process B is least stable because it has the most variability in rod diameter. Process A produces rods whose diameters are generally farthest from the target value.

### Activity 3-10: College Tuitions

- (a) 33 (the third bar)
- (b) 21,  $21/117$  which is .179
- (c) No, \$20,000 is a midpoint, so you can't tell how many of the 8 colleges with this midpoint are actually above \$20,000.
- (d) There are 2 peaks in the distribution. One is at roughly \$5,000, and the other is at roughly \$13,750. The first peak is probably state schools, while the second is probably private schools.
- (e) While using 50 subintervals shows the distribution in more detail, it may then be harder to see the overall shape. Ten subintervals shows the general trends with moderate detail, while labeling enough to make it readable.

### Activity 3-11: Parents' Ages (*cont.*)

- (a) 4
- (b)  
1| 4  
1| 66677888999  
2| 000001112334444  
2| 6788  
3| 123  
3|  
4| 0

(c) This distribution is skewed to the right. The center of distribution is at about 21. There is one outlier, namely 40 years of age.

### Activity 3-12: Parents' Ages (*cont.*)

Answers will vary from class to class.

### Activity 3-13: Marriage Ages

#### husbands' ages

(a)

1| 9

2| 3355556699

3| 0111458

4|

5| 144

6| 02

7| 1

(b) The distribution is skewed to the right for the husbands. The center of distribution is at about 31.

#### wives' ages

(a)

1| 6

2| 22334456778

3| 023669

4| 457

5| 0

6| 0

7| 3

(b) The distribution is skewed to the right for the wives. The center of distribution is at about 30.

### Activity 3-14: Hitchcock Films

(a)

8| 1

9|

10| 1335888

11| 136679

12| 00068

13| 026

(b) The distribution is skewed to the right. The center of distribution is at about 116. There is one outlier at 81 minutes.

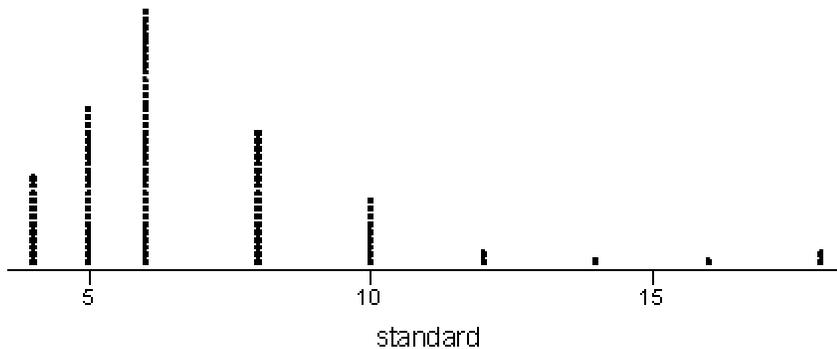
(c) Since there is one outlier, it can be argued that this film is significantly different from the others, based solely on running time. Therefore, an intelligent choice would be *Rope*.

### Activity 3-15: Jurassic Park Dinosaur Heights

- (a) Heights1 is roughly symmetrical with one peak. Heights2 has a lot more variability in heights and has 3 clusters.
- (b) The natural population is represented by heights1, while the controlled population is represented by heights2 because of its 3 clusters.
- (c) There are no outliers and it is a very smooth curve. The actual distribution occurring in nature would probably not be quite so "regular" or perfectly shaped.

### Activity 3-16: Tennis Simulations

Dotplot for standard

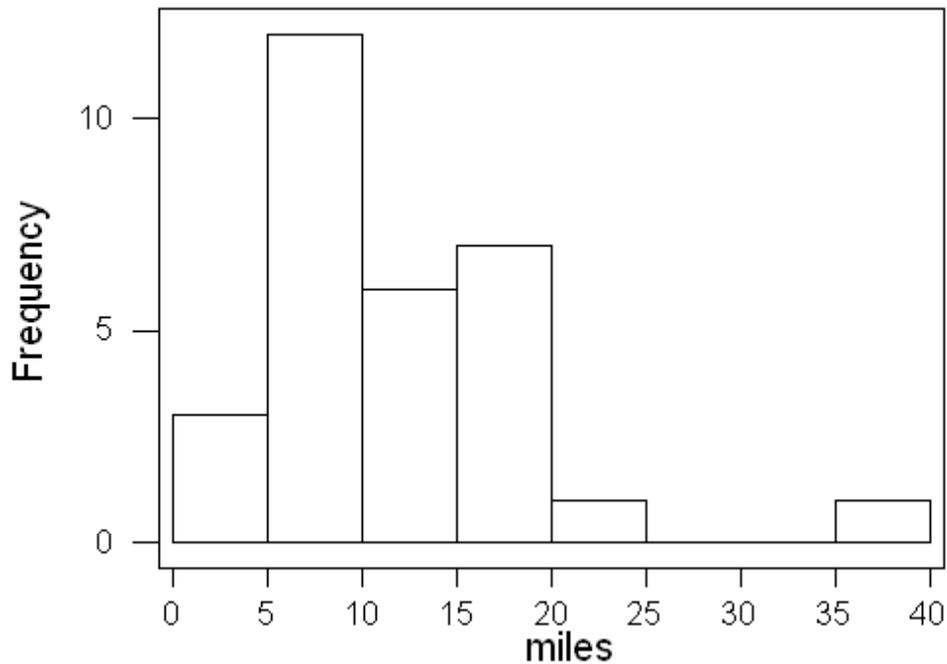


- (a) The distribution is skewed to the right. The center of distribution is 6 points.
- (b) There is an unusual granularity to this distribution. All games, except those 21 that were over in 5 points, were over in an even number of points. This could be explained by the fact that most of these games were a result of a tie after 4 points. Since one must win by 2 points, an even number of total points must usually be reached before the end of the game.

### Activity 3-17: Turnpike Distances

(a)

miles	0.1-5.0	5.1-10.0	10.1-15.0	15.1-20.0	20.1-25.0	25.1-30.0	30.1-35.0
tally (count)	3	12	6	7	1	0	0



(b)

(c) The peak and the center of distribution are in the 5.1-10.0 range. The spread is from 0.1 to 40.0. There is one outlier in the 35.1-40.0 range. There does not seem to be a recognizable shape to this distribution, though might be skewed to the right.

(d) 10.05 miles is a value such that half of the exits are more than 10.05 miles apart, and half are less than 10.05 miles apart. This value is not unique because any value greater between 9.8 miles and 10.7 miles would fit this criteria.

(e) She is very likely to make it. 28 out of 30 exits are 20 miles or fewer apart. This is a proportion of about .93.

(f) Half of all the exits are 10 miles or fewer apart. This is a proportion of .5, so this an equal chance that she will make it as that she will not.

### Activity 3-18: ATM Withdrawals

### Dotplot for withdrawals

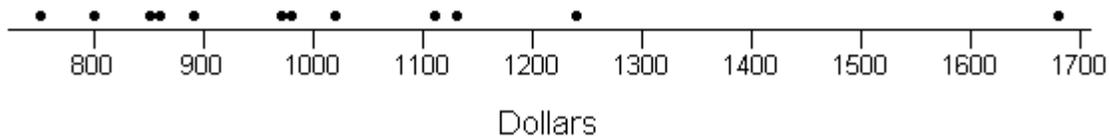


(a)

There is a cluster about 9 withdrawals with a couple of high outliers, and one low outlier.

(b)

### Dotplot for Dollars



There is a fairly even distribution from \$750 to \$1240, and one high outlier.

(c) December had both the most withdrawals and the most cash withdrawn. Christmas is in December, and many people spend a lot of money on items for Christmas.

(d) The trips were in July. The ATM is used the fewest times in this month, and the least total cash is withdrawn. It can be difficult to get to an ATM when away from home. People usually get traveler's checks for trips, so this individual would not need the ATM as much.

### Activity 3-19: Students' Salary Expectations

Answers will vary from class to class.

### Activity 3-20: Fan Cost Index (*cont.*)

(a) The program distribution is roughly symmetrical. The center of distribution is at the peak of about \$4.00. The spread is from about \$2.00 to about \$6.00. There is some variability in price. The cap distribution has no nice shape. There are two peaks, at about \$10.00 and about \$12.00. The spread is from about \$7.75 to about \$15.10. There is some variability in price.

(b) In almost all cases in both distributions, program and cap prices were on the dollar, that is with no fractional dollar amount. There are a few observations that violate this granularity in both distributions.

(c) Five cities violate this granularity. Of them, three are American. Of these three, two violate the granularity by adding \$0.25 or \$0.50 to the dollar amount, which Americans do not consider unusual. The third of these cities is Tampa Bay. Only Tampa Bay's cap price violates the granularity. The two cities that are not American are Montreal and Toronto, both Canadian. Both cities violate the granularity in both cap and program prices. This is probably due to the conversion rate between American and Canadian dollars.