

Workshop Statistics: Discovery with Data, Second Edition

Topic 2: Data, Variables, and Technology

Activity 2-5: Scrabble Names (*cont.*)

Answers will vary from student to student.

Activity 2-6: Fan Cost Index (*cont.*)

(a) highest:

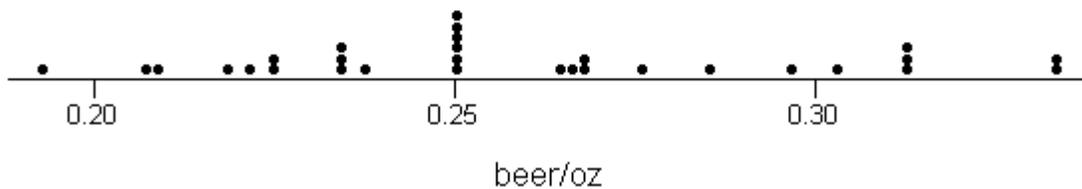
- 1. N.Y. Mets, \$5.50
- 2. Anaheim, \$5.00
- 3. Atlanta, \$4.75

lowest:

- 30. Toronto, \$2.48
- 29. Florida, \$2.50
- 28. Montreal, \$2.81

(b)

Dotplot for beer/oz



The peak is at about \$0.25 per oz. Most points lie between \$0.20 per oz. and \$0.32 per oz.

(c) highest:

- 1. Oakland, \$0.33 per oz.
- 2. Boston, \$0.33 per oz.
- 3. Anaheim, \$0.31 per oz.

lowest:

- 30. Houston, \$0.19 per oz.
- 29. Toronto, \$0.21 per oz.
- 28. Florida, \$0.21 per oz.

These are not the exact same ballpark from (a). However, we still see Anaheim in the top 3, and Toronto and Florida in the bottom 3.

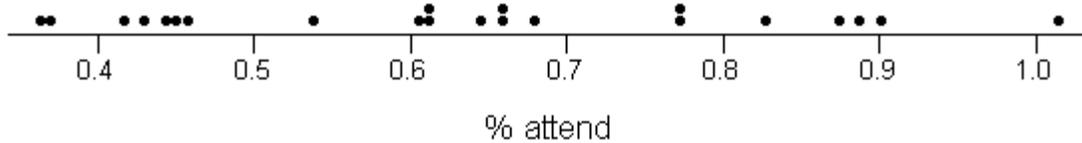
Activity 2-7: Hazardousness of Sports

- (a) Based on number of injuries, bike riding is more hazardous than football;
soccer > ice hockey; swimming > skateboarding
- (b) To calculate the rate per thousand we multiply by 1000 instead of 1000 (let $c_4 = c_3/c_2 * 1000$)
- (c) Based on rate of injury, football (16.6) > bike riding (12.1); ice hockey (40.8) > soccer (10.9) and skateboarding (7.6) > swimming (1.4)
- (d) They all reversed. Compare to answer to Preliminary 8
- (e) least hazardous: billiards, bowling, archery
most hazardous: ice hockey, basketball, football
- (f) Many possible answers, e.g. seriousness of injury

Activity 2-8: Broadway Shows

- (a) The observational units are the different shows.
- (b) Type: categorical - binary; Receipts: quantitative; Attendance: quantitative; Capacity: quantitative
- (c)

Dotplot for % attend



highest: *The Lion King*, 101%; lowest: *Smokey Joe's Cafe*, 37%

(d) There is a fairly even distribution of percentage of capacity. The middle of the distribution seems to be between 60% and 70% of capacity. There are two outliers below 40% (*Smokey Joe's Cafe* and *Kat and the King*), and one outlier above 100% (*The Lion King*).

(e) attendance: 11; receipts: 9; % of capacity: 3

(f) The show consistently filled the theater close to capacity, but the theater is smaller than others (16 theaters are larger) so that the attendance and overall revenue is lower in comparison. Other shows could have also charged more, increasing their receipts compared to *Death of a Salesman*.

Activity 2-9: Box Office Blockbusters

(a) *Star Wars: The Phantom Menace*, \$64.8110 million

(b)

Dotplot for weekend1



About half of the points lie between \$20 million and \$40 million. *Tarzan's* \$34,221,968 is the middle value. *Star Wars: The Phantom Menace* (\$64,810,968) and *The Blair Witch Project* (\$1,512,054) were the two main outliers.

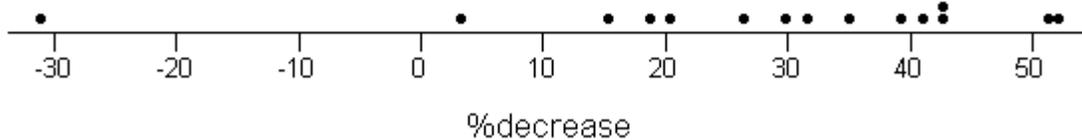
Answers will vary for two additional films, e.g., *The World is Not Enough* and *Runaway Bride* both made about the same amount, just above *Tarzan*.

(c) highest: *Big Daddy* (51.8%); lowest: *The Blair Witch Project* (-30.8%). The negative value indicates an increase in revenue.

(d) No, *Star Wars: The Phantom Menace* had a drop of over 20%, while *The Blair Witch Project* actually had a negative drop, meaning that second weekend revenue was higher than first weekend revenue.

(e)

Dotplot for %decrease



Most movies had at least a 15% drop. There are two outliers: *The Blair Witch Project* (about -31%) and *The Sixth Sense* (about 3%).

(f) *The Blair Witch Project* opened on very few screens initially, but created a sensation and word spread quickly. As it gained in popularity, word spread and it opened on more screens. Because few people saw it the first weekend, there were many people who had not yet seen it, some of which did so on the second weekend.

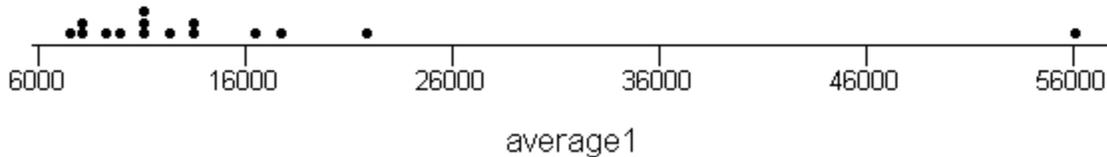
(g) There was about a 1,376% increase.

Activity 2-10: Box Office Blockbusters (*cont.*)

(b) highest: *The Blair Witch Project*, \$56,002 per screen; lowest: *Analyze This*, \$7,300.80 per screen; middle: *Tarzan*, \$11,388.30 per screen

(c)

Dotplot for average1

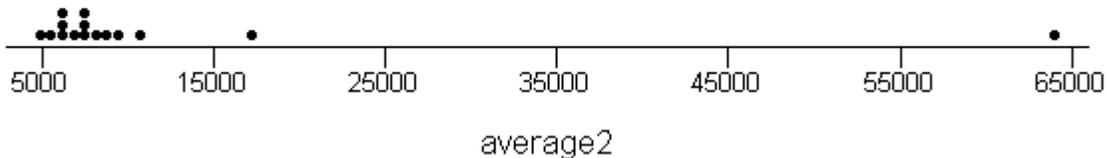


Most of the points lie between \$16,000 and \$22,000 per screen, with one outlier at about \$56,000 per screen (*Blair Witch Project*).

(d,b) highest: *The Blair Witch Project*, \$63,812 per screen; lowest: *Wild, Wild West*, \$5,037.10 per screen; middle: *The Mummy*, \$7,705 per screen

(d,c)

Dotplot for average2



Most of the points lie between \$5,000 and \$11,000 per screen. There is one outlier at about \$17,000 per screen (*Star Wars*), and another at about \$64,000 per screen (*The Blair Witch Project*).

(e) In general for these data, the films with the highest revenue per screen in the first weekend tend to be the same films with the highest revenue per screen in the second weekend, probably because people tend to see good films more than once, as well as new viewers. The films with the lowest revenue per screen tend to shift around near the bottom, but it's the same 3 or 4 films at the bottom. Word gets out that it's not a very good movie, so not many people go see it the second weekend.

Activity 2-11: Uninsured Americans

There is a typographical error in the text at this point. The table fails to include the states of Missouri and Wyoming. Therefore, this Activity can be done two ways: with Missouri and Wyoming, and without Missouri and Wyoming. First we will do the activity as it appears in the text, without Missouri and Wyoming. You can use the files `uninsured48.mtw` and `uninsured48.dat` for this activity.

(a)

- starting with highest: California, Texas, New York
- starting with lowest: Vermont, South Dakota, Hawaii

(b)

- starting with highest: Arizona (25%), Texas (25%), Arkansas (24%)
- starting with lowest: Hawaii (7.5%), Wisconsin (8%), Minnesota (9.2%)

(c) The states with the highest uninsured rates are not necessarily the states with the most uninsured residents. The states with the lowest uninsured rates do not necessarily also have the fewest uninsured residents. In fact Vermont, the state with the lowest number of uninsured residents, has the fourth highest uninsured rate. This is due in part to the fact that Vermont has such a low population.

(d) If one looks only at the raw numbers of uninsured residents, one does not take into account the total populations of the states. By looking at the rates of uninsured residents, one can see comparable percentages that take into account not only how many residents are uninsured, but also how many total residents are in each state. This makes for a much better state to state comparison.

We will now do Activity 2-11: Uninsured Americans again, this time including Missouri and Wyoming as they appear in the files `uninsured.mtw` and `uninsured.dat`.

(a)

- starting with highest: California, Texas, New York
- starting with lowest: Vermont, Wyoming, South Dakota

(b)

- starting with highest: Arizona (25%), Texas (25%), Arkansas (24%)
- starting with lowest: Hawaii (7%), Wisconsin (8%), Minnesota (9.2%)

(c) Texas is second highest on both lists, but it is not true that all the states with the highest uninsured rates have the most uninsured residents. The states with the lowest uninsured rates do not necessarily have the fewest uninsured residents. Rates take population into consideration, whereas the number of uninsured residents does not.

(d) If one looks only at the raw numbers of uninsured residents, one does not take into account the total populations of the states. By looking at the rates of uninsured residents, one can see comparable percentages that take into account not only how many residents are uninsured, but also how many total residents are in each state. This makes for a much

better state to state comparison.

Activity 2-12: Signature Measurements (*cont.*)

Answers will vary from class to class.

Activity 2-13: Driver Safety

No, this does not establish that younger drivers are better and safer than older drivers. First of all, we do not know how many people were between the ages of 16 and 20, and over 55 in 1997. In particular, we suspect (see next paragraph) there are many more drivers in the over 55 age group than in the 16-20 age group, and therefore we would expect a higher *count* of accidents in the older group because there are more of them on the road. A more fair comparison of the accident rate of the two groups would be to take into account the number of drivers in each age category.

For instance, some states, like New Jersey, do not issue driver's licenses until the driver has reached the age of 17. This lessens the total number of drivers who fall between the 16-20 category. No state in the union voids a driver's license just because a driver gets too old, so the over 55 category does not have this problem.

Activity 2-14: Personal Comparison

Answers will vary from student to student.