

## Lesson: Misuses of Data

**Eighth Grade Objective: 4.03 Identify misuses of statistical and numerical data.**

### Lesson:

It is impossible to go through almost any day without coming into contact with some type of data. Frequently, that data is presented to us by a person or group who is trying to influence our opinions about something. It is important for us to be educated about misleading graphs, data, and statistics so that we can make informed decisions about everything from what we eat for lunch to how many different sizes of each jersey a clothing store should order.

First, we need to recall some definitions: Measures of central tendency include the following:

The **arithmetic mean**, usually just called the mean, is the sum of all the elements in a set, divided by the number of elements in the set. Ex: The mean of the set {2, 4, 3, 1} is found by  $(2 + 4 + 3 + 1)/4$ .

The **median** is found by putting all the numbers in a set of data in ascending order and locating the middle number. If there are two middle numbers, the median is the mean of the two numbers. Ex. The median of the set {2, 4, 3, 1, 5} is found by putting the numbers in ascending order: 1, 2, 3, 4, 5 and locating the middle number, 3. Another example: The median of the set {2, 4, 3, 1} is found by putting the numbers in ascending order: 1, 2, 3, 4, locating the middle numbers (2 and 3) and taking their mean  $((2+3)/2 = 2.5)$ .

The **mode** is the number or element in a set of data that occurs the most.

All three of these are frequently just referred to as the “average”. It is important to know which average is being referred to, since sometimes data can be misrepresented by using an inappropriate measure of central tendency.

Take, for example, a clothing store that sells jerseys. Last week, the store sold the following sizes of jerseys: XL, L, L, M, L, XL, S, M, L, XL, M, L, L, S. The manager is placing an order for new jerseys today. Which measure of central tendency should she use to determine which size to order the most of?

**The manager should use the mode. She is looking for the size that is bought most often, which is the element in the data set that occurs the most. There were 2 small jerseys sold, 3 medium, 6 large, and 3 extra-large. Most of the jerseys she should order should be size large.**

A gas station is trying to attract customers by advertising “Fuel! Averaging only \$2.99 per gallon!” The actual fuel prices are:

Type of Fuel	Cost
Non-Highway Grade Diesel	\$0.84
Diesel	\$3.52
87 Octane Unleaded Gasoline	\$3.48
89 Octane Unleaded Gasoline	\$3.53
93 Octane Unleaded Gasoline	\$3.58

The customers are very angry when they pay for their gasoline. They argue that they were lied to. Were they lied to?

No, the customers were not lied to, they were misled. The arithmetic mean of the fuel costs is \$2.99. There is, however, an outlier that pulls the mean down significantly. A better representation of the actual gas prices is the median, \$3.52.

The gas station in the problem above is located off of a major highway. Jeremy is asking the drivers that are fueling their vehicles at the gasoline pumps how many miles per gallon their vehicle gets. Jeremy is pleased to find that the average vehicle gets 27 miles per gallon. Why might this be misleading?

This might be misleading for at least two reasons:

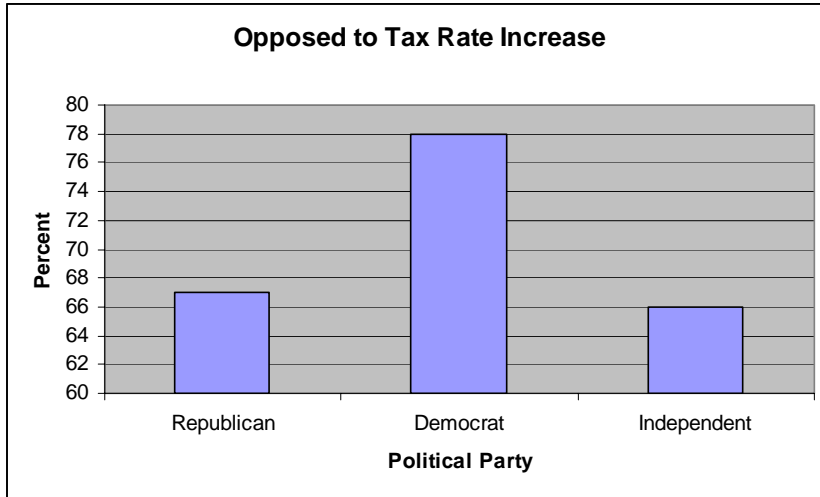
1. Jeremy is at the gasoline pump. Large trucks typically use diesel and get fewer miles to the gallon. Jeremy’s survey does not include those vehicles.
2. Jeremy only surveyed people at one gas station located off a major highway. Cars running at consistent speeds get better gas mileage than they do if they are in stop-and-go traffic. The average Jeremy found might be high because most of the drivers are recording highway miles.

Kayleigh is surveying local residents on where they shop for electronics. She surveys every third person who enters the nearby Best Buy store. Will the results of her survey be representative of the population?

No, her results will not be representative, they will be biased. While surveying every third person appears to be random (which is good!), all the people are potential customers of Best Buy – an electronics store! If they are going into Best Buy, it is likely they will answer her survey by saying they shop for electronics at Best Buy.

Graphs can also be misleading:

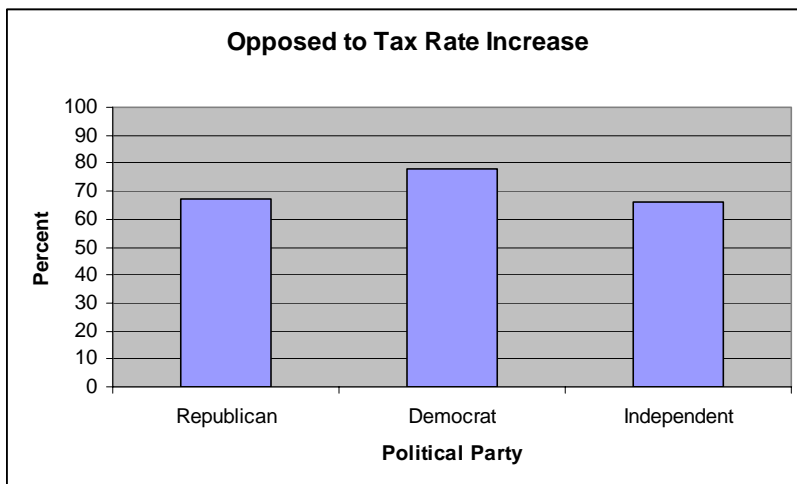
The graph below shows the percent of people, broken down by political party, who are opposed to a tax rate increase:



What is misleading about the graph?

According to the image, three times as many Democrats are opposed to the tax rate increase as Independents. This is not true, since 66% of Independents are opposed and 78% of Democrats are opposed. While the values are only 12% different, the way the graph is drawn makes it look as though there is a more significant difference.

The person who created this graph was able to manipulate the scale on the y-axis to make the difference look larger. The scale on the y-axis should start at zero.



It becomes clear, when the axis is adjusted to begin at zero, that the difference between the parties is not as great as the first graph hoped to convey.



Graphs are frequently manipulated in the following ways:

- unclear or missing titles and labels
- scales that do not begin with zero
- uneven intervals (not counting by the same number on one or both axes)
- breaks in the graph (although not wrong, they can mislead)

Try these on your own:

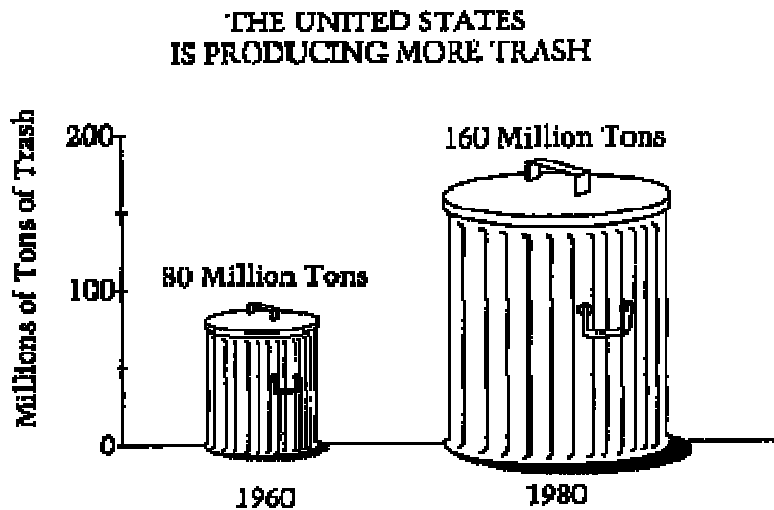
1. The following is the distribution of students with different grades in math class:

Grade on Algebra Test	Frequency of Students
95	3
90	7
85	6
80	2

Marcia says the average grade on the Algebra test was an 87.5. Is she correct? Why or why not?

2. Jerel is surveying people in his community about whether they voted in the most recent election. Jerel calls every second person on the registered voters list. He determines that 72% of the people in his community voted in the most recent election. Are his results valid?

3. Why could the graph below be considered misleading?



(Source: NAEP 1992 Released Items)

Check your answers:

1. No, she is not correct. She found the arithmetic mean of the numbers in the left column only. She neglected to consider that 3 people have 95's, 7 people have 90's, and so on. The actual arithmetic mean would be computed by:

$$(95+95+95+90+90+90+90+90+90+90+90+85+85+85+85+85+85+80+80)/18 = 88.1$$

The median is 90, as is the mode.

2. No, his results are not valid. He only surveyed registered voters. Seventy-two percent of registered voters in his community may have voted, but most likely, his entire community is not registered. It would have been more accurate for Jerel to have chosen every 2<sup>nd</sup> person from the phone book or every 3<sup>rd</sup> person leaving the area's shopping mall.

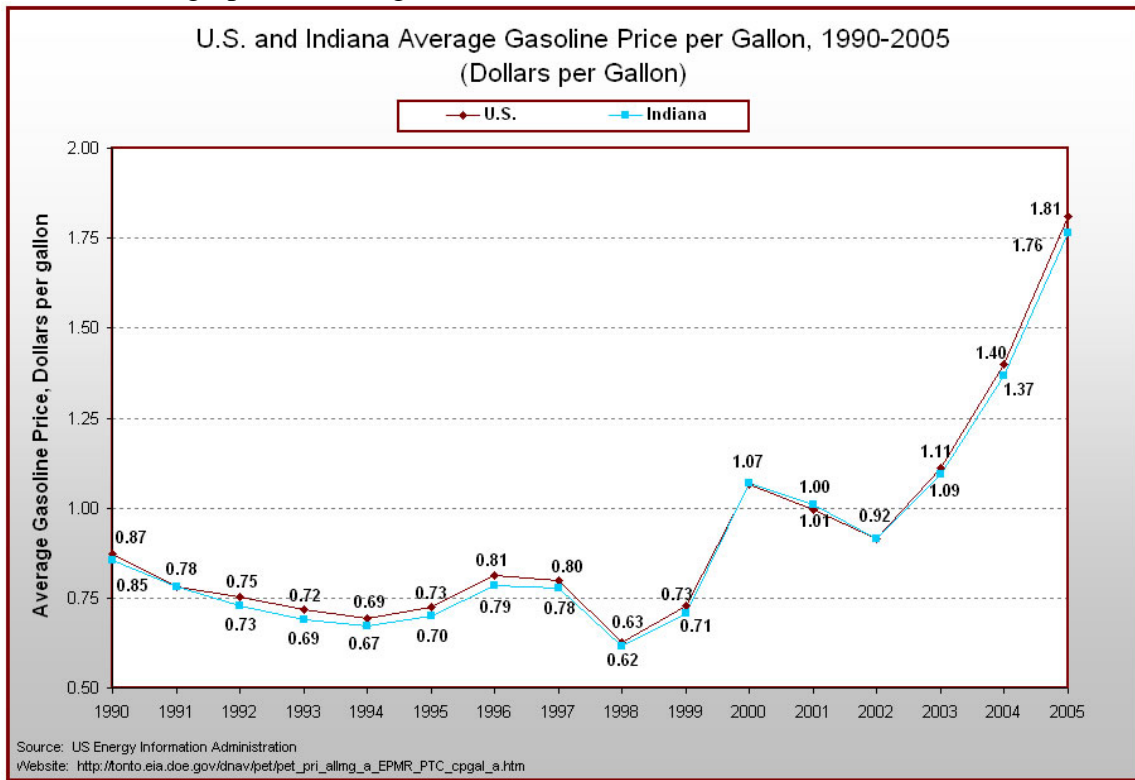
3. The amount of trash doubled. The height of the second trash can doubled, but so did the radius, so the volume of the trashcan actually increased by a factor of 8. The picture shows trash increasing 8 fold, while it actually doubled.

Quiz Yourself:

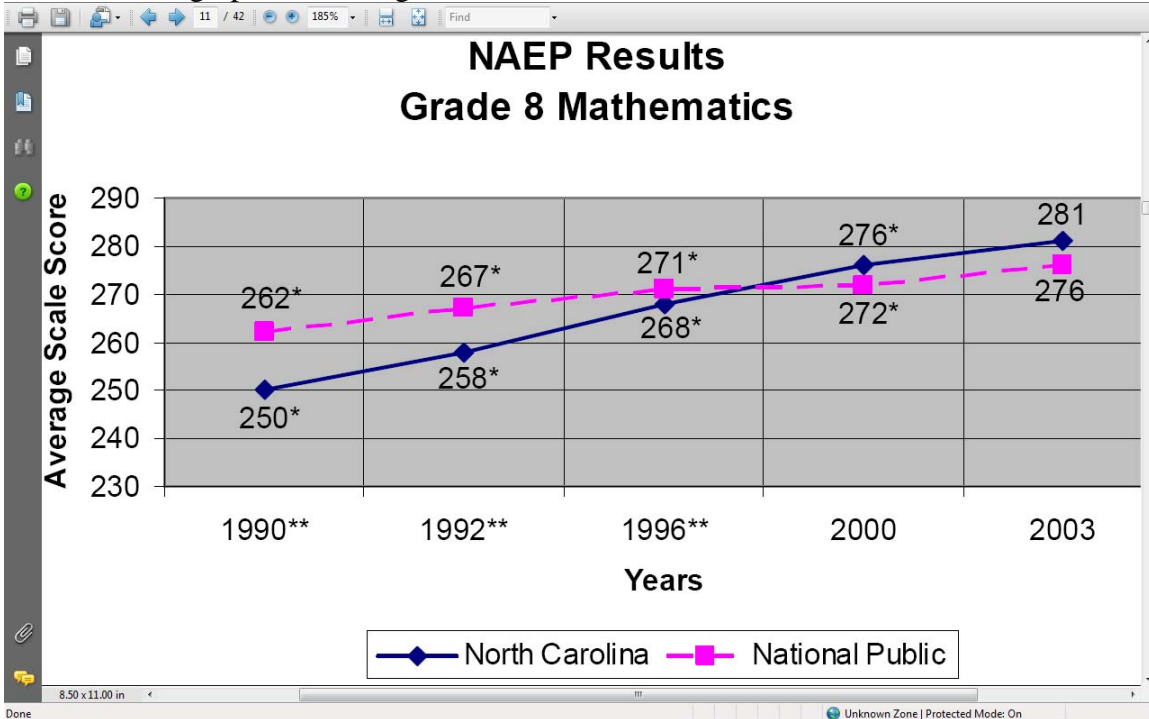
1. The average cost of hamburgers at local fast food chains is reported in the newspaper as \$2.12. The costs at the four restaurants surveyed is: A: \$1.29, B: \$1.35, C: \$1.49, and D: \$4.35. Why is the statistic misleading? Would any of the restaurants be pleased with the misleading report?

2. A survey is to be taken in a city to determine the most popular sport. Would sampling opinions at a baseball game be a good way to collect this data? Explain your answer. (Source: NAEP 2003 Released Items)

3. How is this graph misleading?



4. How is this graph misleading?



(Source:

<http://www.ncpublicschools.org/docs/accountability/2003NAEPMathematicsReport.pdf>)

Check your answers:

1. The statistic is misleading because the outlier raises the mean significantly. Restaurants A, B and C all benefit from this because they are below the “average” cost.
2. No, a baseball game would not be a good place to conduct the survey. Since the people are at a baseball game, it is likely that a higher than average percent of people would say their favorite sport is baseball.
3. The graph is misleading because the y-axis does not begin at zero. The x-axis should also be labeled “years”.
4. The interval on the x-axis is inappropriate, it begins by counting by 2, then 4, then 4, then 3, all with the same spacing in between. The y-axis does not begin with zero. The data points most likely should not be connected because the data is not continuous.