

# Boxes and Cats

## Statistics From the Beginning

One-Page Overview

By Robert B. Brown, The Ohio State University

Topics:

Data Analysis, Mathematics as Communication, Statistics

Levels:

Grades 5 – 8

Problem:

Q. Why is a box like a cat?      A. Both have whiskers.

Getting Started:

Ask your students to answer these questions:

- a. What is your pulse rate?
- b. Estimate the time that it takes you to get from your home to school.
- c. Estimate how much time has gone by in this class period already.
- d. How many times have you seen the movie, "The Wizard of Oz"?
- e. On a scale of 1 – 5, rate "The Wizard of Oz." (1-low, 5-high, x-not seen)
- f. What is the probability, 0 to 1, that you will see "The Wizard of Oz?" If you have seen it, record 1.
- g. Record the musical instruments that you play and the vocal groups that you participate in.
- h. What is your favorite vacation spot?

**Ohio Academic Content Standards, 2002**

5-7	8-10	11-12
1. Number, Number Sense and Operations	1. Number, Number Sense and Operations	1. Number, Number Sense and Operations
2. Measurement	2. Measurement	2. Measurement
3. Geometry and Spatial Sense	3. Geometry and Spatial Sense	3. Geometry and Spatial Sense
4. Patterns, Functions and Algebra	4. Patterns, Functions and Algebra	4. Patterns, Functions and Algebra
5. Data Analysis and Probability	5. Data Analysis and Probability	5. Data Analysis and Probability
<b>Mathematical Processes</b> Connections Communication	<b>Mathematical Processes</b> Connections Communication	<b>Mathematical Processes</b>

**NCTM Principles and Standards, 2000**

6-8	9-12
1. Number and Operations	1. Number and Operations
2. Algebra	2. Algebra
3. Geometry	3. Geometry
4. Measurement	4. Measurement
5. Data Analysis and Probability	5. Data Analysis and Probability
6. Problem Solving	6. Problem Solving
7. Reasoning and Proof	7. Reasoning and Proof
8. Communication	8. Communication
9. Connections	9. Connections
10. Representation	10. Representation

Note: Capital X denotes major emphasis; lower case x denotes minor emphasis.

# Boxes and Cats

## Statistics From the Beginning

By Robert B. Brown, The Ohio State University

<u>Topics:</u> Data analysis, Mathematics as communication, Statistics	
<u>Levels:</u> Grades 5 - 8	<u>Timing:</u> One hour
<u>Materials:</u> Classroom clock with a second hand	<u>Prerequisites:</u> None

Problem:

- Q. Why is a box like a cat?  
A. Both have whiskers.

Goals:

- Introduce the statistical notions of range, median, quartile
- Create box-and-whiskers plots
- Create stem-and-leaf plots
- Ask questions about data sets
- Practice finding differences between data sets

Procedure:

1. Give each student a sheet of paper to fold into eight rectangles. Label the rectangles A through H, and fill in the answers to these questions:
  - A. What is your pulse rate?
  - B. Estimate the time that it takes you to get from your home to school.
  - C. Estimate how much time has gone by in this class period already.
  - D. How many times have you seen the movie, "The Wizard of Oz" either in a theater or on TV?
  - E. On a scale of 1 – 5, rate "The Wizard of Oz." (1-low, 5-high, x-not seen)
  - F. What is the probability, 0 to 1, that you will see "The Wizard of Oz?" If you have seen it, record 1.
  - G. Record the musical instruments that you play and the vocal groups that you participate in.
  - H. What is your favorite vacation spot?
2. Record students' pulse rates on the board. Have the class as a whole discuss what interesting questions there might be about pulse rates. Then have them discuss how these data might be organized to enable one to quickly answer the questions.
3. A suggestion that the students will probably make is to write the pulse rates in order from lowest to highest (or the other way around).
4. Do steps 2 and 3 for B and for C.
5. Are there any differences in the data sets for A, B, and C?
6. What questions might the students ask now about the data sets? Are the data sets organized to their satisfaction to try to answer the questions?
7. Ask the students to discuss which of the other data sets would be interesting to analyze? Would they want to ask the same or different questions about each one?
8. Now you are in a good position to introduce some of the simplest of the standard statistics that are used to describe data sets: median, upper quartile, lower quartile, and range. When these statistics have been determined, a box-and-whiskers plot can be made for each data set. This lets anyone compare the data at a glance. (Definitions of all of these terms with examples are in the Statistics section.)
9. Break the students into groups and have them make stem-and-leaf plots for the same data sets. Stem-and-leaf plots are described in the Statistics section, too.

Closure:

Displaying data sets and interpreting differences between them is partly science and partly an art. It is often very difficult to tell if the differences that you see reflect differences in how the data were generated or are just chance fluctuations. Sometimes further investigation is needed, such as generating more data by the same processes that gave you the original data sets.

Extension:

Separate the pulse rate data by male and female. Draw a box-and-whiskers plot of each separately. Are there any detectable differences between the two sets of data?

Statistics:

Here is an example of a data set taken from a table of random numbers:

00, 43, 63, 61, 99, 31, 36, 19, 70, 85, 71, 12, 98, 23, 10, 14, 98, 90, 94, 33, 85, 32, 07, 55, 87, 75, 34, 28, 59, 91, 55, 99, 54, 11, 23, 53, 94, 24, 87, 14, 18, 53, 32, 56, 69, 79, 88, 21, 57, 05

Here is the same data set arranged in order from smallest to largest.

00, 05, 07, 10, 11, 12, 14, 14, 18, 19, 21, 23, 23, 24, 28, 31, 32, 32, 33, 34, 36, 43, 53, 53, 54, 55, 55, 56, 57, 59, 61, 63, 69, 70, 71, 75, 79, 85, 85, 87, 87, 88, 90, 91, 94, 94, 98, 98, 99, 99

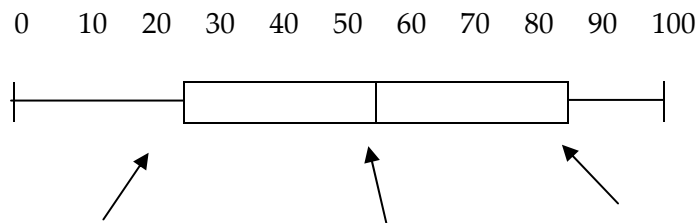
A possible discussion could be to determine a quick way to check whether there have been any mistakes made in rearranging the numbers in the data set.

The *range* of a data set is the difference between the highest number in the set and the lowest. In this set the highest is 99 and the lowest is 0, so the range is  $99 - 0 = 99$ .

The *median* of the data set is the middle number moving from lowest to highest (or from highest to lowest). Because there are fifty numbers in this data set, there is no exact middle, and the median would be the 25th or the 26th number. The 25th number counting up is 54 and the 26th is 55. So the median would be 54 or 55. You could have the class discuss which would be the best choice, or would something else be a better number to use as the central number? Because there is very little difference in choosing one over the other, the class could simply agree on which definition they are going to use.

The *upper quartile* is the median of just the upper half of the data set, and the *lower quartile* is the median of the lower half. In our data set each half has 25 numbers, an odd number, so there are unambiguous choices for these quartiles. The upper quartile is 85, and the lower quartile is 23.

Create a box-and-whiskers plot by marking off the median and the upper and lower quartiles on a number line, as shown below.



Lower quartile: 23

Median: 55

Upper quartile: 85

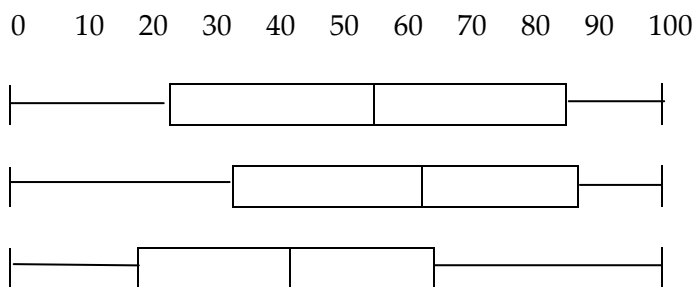
Here is a second set of data. These fifty numbers are between 0 and 99 and are formed by the 4th through 103rd decimal digits of the decimal expansion of the number pi.

59 26 53 58 97 93 23 84 62 64 33 83 27 95 02 88 41 97 16 93 99 37 51 05 82  
 09 74 94 45 92 30 78 16 40 62 86 20 89 98 62 80 34 82 53 42 11 70 67 98 21

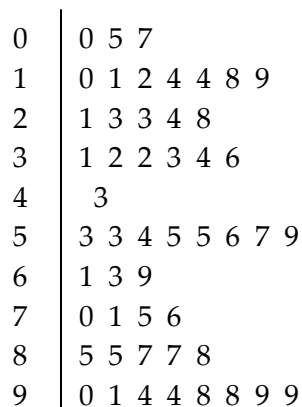
Here is a third set of data. These fifty numbers are between 0 and 96 and are formed by the 104th through 203rd decimal digits of the decimal expansion of the number pi.

48 08 65 13 28 23 06 64 70 93 84 46 09 55 05 82 23 17 25 35 94 08 12 84 81  
 11 74 50 28 41 02 70 19 38 52 11 05 55 96 44 62 29 48 95 49 30 38 19 64 42

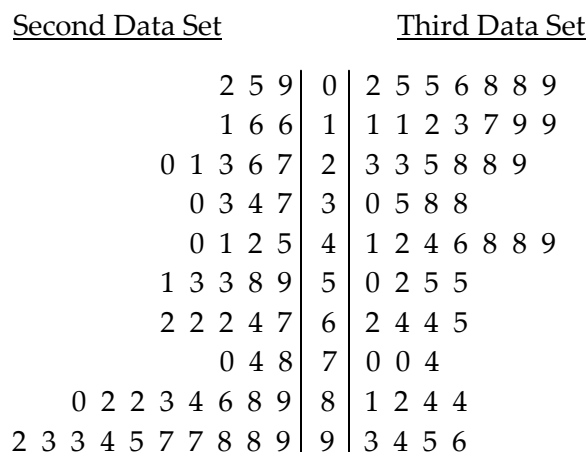
Here are box-and-whiskers plots of all three data sets. The different shapes of the plots can tell you about differences in the data.



Stem-and-leaf plots are another simple way to exhibit data. For two-digit data all possible ten's digits are written vertically followed by a vertical line. Then the unit's digits for each piece of data is put in the row with its correct ten's digit, and finally the unit's digits in each row are put into numerical order. Here are the results for the first data set. Its stem-and-leaf plot is:



A two-sided stem-and-leaf plot can be used to compare two data sets. Here is a two-sided plot for the second and third data sets from above. The unit's digits for the second data set are on the left; the unit's digits for the third set are on the right.



Does this stem-and-leaf plot show any differences between the two data sets? Do you think the differences that you see, if any, are important? Do you have any reasons to expect differences? What if you make a two-sided stem-and-leaf plot of male pulse rates vs. female pulse rates?

Relationships to the Ohio Academic Content Standards, 2002:

Grades 5-7:

Data Analysis and Probability Standard

The student will be able to...

- Read, create and use line graphs, histograms, circle graphs, box-and-whisker plots, stem-and-leaf plots, and other representations when appropriate.
- Collect, organize, display and interpret data for a specific purpose or need.
- Determine and use the range, mean, median and mode to analyze and compare data, and explain what each indicates about the data.
- Find all possible outcomes of simple experiments or problem situations, using methods such as lists, arrays and tree diagrams

Mathematical Processes Standard

The student will be able to...

- Relate mathematical ideas to one another and to other content areas; e.g., use area models for adding fractions, interpret graphs in reading, science and social studies.
- Communicate mathematical thinking to others and analyze the mathematical thinking and strategies of others.

Grades 8-10:

Data Analysis and Probability Standard

The student will be able to...

- Create, interpret and use graphical displays and statistical measures to describe data; e.g., box-and-whisker plots, histograms, scatterplots, measures of center and variability.
- Find, use and interpret measures of center and spread, such as mean and quartiles, and use those measures to compare and draw conclusions about sets of data.

Mathematical Processes Standard

The student will be able to...

- Apply mathematical knowledge and skills routinely in other content areas and practical situations.
- Locate and interpret mathematical information accurately, and communicate ideas, processes and solutions in a complete and easily understood manner.

Relationships to the NCTM Principles and Standards, 2000:

Grades 6-8:

Data Analysis and Probability Standard

Instructional programs from pre-kindergarten through grade 12 should enable all students to...

- Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.
- Select and use appropriate statistical methods to analyze data.

Communication Standard

Instructional programs from pre-kindergarten through grade 12 should enable all students to...

- Organize and consolidate their mathematical thinking through communication.
- Communicate their mathematical thinking coherently and clearly to peers, teachers, and others.

Connections Standard

Instructional programs from pre-kindergarten through grade 12 should enable all students to...

- Recognize and apply mathematics in contexts outside of mathematics.