NUMB3RS Activity: I Never Metadata I Didn't Like
Episode: "Democracy"

Topic: Pattern analysis and recognition **Grade Level:** 9 - 12
Objective: To associate sets of numbers with contexts for their possible meaning.
Materials: paper, pencil, and (time permitting) access to reference materials.
Time: 15 - 20 minutes (plus research time, if possible)

Introduction

In "Democracy," Charlie investigates the death of his friend Rachel Lawton. He finds columns of numbers on her computer. Because they have no labels, he does not know what the numbers mean. Charlie explains, "Numbers tend to come in groups…it's called *metadata*—information about data. For instance the numbers 1, 2, 3, 4, 5 are data. They could refer to anything. But if you know they are a ZIP code, they mean something very specific—in this case Schenectady, New York."

In this activity, students will analyze sets of data and try to identify a reasonable meaning for each set. While students can work alone on this activity, it is designed for a class (or at least small groups) to work on cooperatively. Ideally, different students will recognize different pieces of data and will appreciate the contributions of each person, based on their individual backgrounds. If possible, patterns that students do not recognize can be put aside for later consideration. Students may learn that letting a problem "rest" for a while can result in a new perspective and enable a fresh approach to solving it.

Students may also benefit from developing data sets for their classmates to analyze, like data unique to their school or community. The student who creates the set must also offer an explanation of what clues are present in the data. Time permitting, students could have a contest to see who can make the most interesting, most difficult, etc.

Discuss with Students

Point out to students that the data sets in this activity do not necessarily have a single correct answer. The purpose is to examine the data and come up with a plausible possibility of what they might represent. The conclusion is meaningless without the justification. The answers on the following page represent only one possible interpretation. If a student comes up with another conclusion and can justify it, then that answer should be accepted, too. On *NUMB3R*S, while Charlie's conclusions are logical, they are not always the only possible interpretation, and he remains open to other explanations.

Bro. Patrick Carney, De Paul Catholic High School, Wayne, NJ

Student Page Answers:

1a. *Blood pressures. Note the fractional form and the range of values.* **1b.** *Eye test results. While also fractions, these all have a numerator of 20.* **1c.** *Body temperatures. They are all near the "normal" 98.6. It is probably a medical office.* **2a.** *Letter grades at a school that uses an A, B, C, D, F grading system.* **2b.** *SAT scores (math, of course). They are all 3 digit numbers ending in 0 in the range between 200 and 800.* **2c.** *Grades expressed as a GPA on a 4.0 scale, because of the range and the inclusion of the lone 0 after the decimal, as well as differing number of places after the decimal. It could be a guidance office or the office of a teacher.* **3.** *Musical notes. Only the letters from A to G are used. (These notes are the notes in "Yankee Doodle.")*

4a. *Gasoline prices, because of the 2-digit decimal followed by the.* $\frac{9}{10}$ **4b.** *Baseball batting averages. They are written as 3-digit decimals even if they end in zeroes. In addition, the numbers are within the range common to batting averages.* **4c.** *Dewey decimal numbers for library books. All have a three-digit integer part that includes a leading zero if necessary, and the number of places after the decimal point varies.* **4d.** *Dates as they are sometimes expressed in shorthand. The first two digits are all between 01 and 12, and the next two are from 01 to 31. The last two digits are birth years. The leading zero is printed (in some states like Delaware, these could be license plates.)* **4e.** *Time on a clock. To the left of the colon is a number from 1 to 12 and to the right are two digits from 00 to 59.* **4f.** *Odds for events. The ratios are reduced to lowest terms and not like the time in the previous problem.* **4g.** *Dollar amounts (like prices). All values have two decimal places, even when the final one is zero. Many of the decimals are 0.95 to 0.99, which are common for prices.* **4h.** *Lumber dimensions. The "×" is read "by" and these are standard sizes used in construction (the 5/4 is for deck flooring).* **4i.** *The numbers of the players on the starting offense for two football teams. There are two groups of 11 numbers, and the numbers correspond to the numbers allocated to the various positions by the NFL.* **4j.** *United States FM radio frequencies. There is a single odd digit to the right of the decimal and the numbers to the left of the decimal are within the range for FM broadcast band frequencies in the US (88 – 108 MHz).*

Extensions Answers:

The language is Hawaiian, which has only the vowels a, e, i, o, and u, and the consonants h, k, l, m, n, p, and w.

The numbers are in octal (base 8). A leading question might be to ask that with this many digits, what is the probability that there would be no 8s or 9s if the numbers were in base 10?

Bro. Patrick Carney, De Paul Catholic High School, Wayne, NJ

Name: _____ Date: _____

NUMB3RS Activity: I Never Metadata I Didn’t Like

In “Democracy,” Charlie investigates the death of his friend Rachel Lawton. He finds columns of numbers on her computer. Because they have no labels, he does not know what the numbers mean. Charlie explains, “Numbers tend to come in groups…it’s called *metadata*—information about data. For instance the numbers 1, 2, 3, 4, 5 are data. They could refer to anything. But if you know they are a ZIP code, they mean something very specific—in this case Schenectady, New York.”

For example, in the United States, if a piece of data has the form XXX-XX-XXXX, it is likely to be a Social Security number. The slightly different form XXX-XXX-XXXX is probably a telephone number.

For this activity, examine the sets of data below and use clues in the data to tell what sort of information they could represent. Just as Charlie sometimes has a logical reason for a conclusion that turns out to be incorrect, there is not necessarily one correct explanation for each data set in this activity. The answer is not as important as the reasoning. If there is a logical explanation for the conclusion, then it is a plausible answer. For each data set, identify the type of data and your reasoning.

1. In a certain office, the following sets of data are discovered. What could each set of data represent, and in what kind of office?

a. 180/90, 140/80, 120/50, 130/60, 140/70, 190/100, 110/70, 160/90, 200/120, 130/80

b. 20/40, 20/140, 20/100, 20/10, 20/15, 20/200, 20/70, 20/20, 20/80, 20/60

c. 98.7, 97.8, 98.6, 98.8, 98.2, 98.5, 99.2, 98.6, 100.2, 99.0

2. In another office, a page is found with the following lists of data. What could each of these be, and in what kind of office?

a. A, B, B, D, C, A, C, A, F, C

b. 780, 430, 600, 540, 800, 610, 430, 200, 390, 590

c. 3.91, 4.0, 2.82, 1.3, 3.72, 2.0, 1.92, 3.1, 2.8, 3.5

3. What could this list represent?

C, C, D, E, C, E, D, C, C, D, E, C, B, C, C, D, E, F, E, D, C, B, G, A, B, C, C

Bro. Patrick Carney, De Paul Catholic High School, Wayne, NJ

4. Give a possible meaning for each of the following data sets. Justify your reasoning.

a. $1.97\frac{9}{10}$, $2.06\frac{9}{10}$, $2.17\frac{9}{10}$, $2.08\frac{9}{10}$, $2.12\frac{9}{10}$, $2.34\frac{9}{10}$, $2.45\frac{9}{10}$, $1.99\frac{9}{10}$

b. .293, .313, .275, .300, .281, .333, .198, .310, .288

c. 083.12, 510.751, 618.6, 519.3, 863.9215, 421.6, 187.0, 282.4

d. 120364, 021841, 010491, 100352, 082789, 112077, 030207, 061163

e. 12:31, 1:15, 11:20, 4:29, 3:37, 10:04, 7:31, 6:00, 1:01

f. 5:2, 13:1, 8:1, 3:2, 3:1, 2:1, 10:1, 6:5, 100:1

g. 12.95, 78.98, 143.00, 7632.00, 1720.00, 47.99, 32.97, 1.25

h. 2 × 3, 4 × 4, 5/4 × 6, 1 × 8, 2 × 6, 1 × 4

i. 11, 24, 28, 35, 54, 62, 75, 78, 86, 81 and 17, 21, 32, 45, 55, 61, 68, 71, 80, 88

j. 101.5, 94.7, 88.3, 92.9, 104.5, 91.7, 102.5, 107.5, 90.3

Bro. Patrick Carney, De Paul Catholic High School, Wayne, NJ

The goal of this activity is to give your students a short and simple snapshot into a very extensive mathematical topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extensions

Introduction

The prefix “meta” is used to express a level that is beyond the ordinary level of understanding. In philosophy, metaphysics deals with a higher level of abstraction of physics. Other “meta” fields include *metacognition* (thinking about how people think) and *metalanguage* (how a language works). Similarly, although mathematics itself is a type of abstraction, *metamathematics* refers to how mathematics works and is studied. Hence, metadata refers not to the data itself but goes beyond the data to study its properties. Examples in this activity are simple cases used in many modern applications. Charlie uses more complicated methods to solve crime.

For the Student

Sometimes, it is not the type of data that can be determined, but a different property. For example, suppose the following code is intercepted:
WAKUL LIMKK PAOOL NUWEL HLLNO PNOLK LWIOU ULKIO
Assuming that people tend to write code in their own language, research what language the author may have used.

Suppose the data set is: 307, 1025, 2654, 1000, 7273, 767, 23401, 10036, 201, 777, 2164
Research the concept of the “base” of a number system. What is the most likely base for these numbers?

Construct data sets (perhaps unique to your school or town) and share them with classmates. Have them try to discover a plausible meaning for each data set. As in this activity, mind that the justification is just as important as the result.

Additional Resources

Metadata is much more complex than this introductory exercise. To see a much more detailed explanation see:
http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

Professor Joe Gallian from the University of Minnesota at Duluth has used techniques similar to Charlie’s to decode drivers’ license numbers for many states. For example, if a sequence has a three-digit sequence that falls between 001 and 366, it could be a birth date using the day of the year. A sequence with twice the range could encode the birth date and gender with three digits. Research whether he has done your state and read the reasoning he used to determine what the various numbers and letters represent. For an explanation of how Gallian decoded Minnesota license codes (although they have since changed the system) see:
http://www.sciencenews.org/pages/sn_arc98/10_17_98/mathland.htm

Bro. Patrick Carney, De Paul Catholic High School, Wayne, NJ