HOW TO GO FROM DATA TO FUNCTION

In order for students to truly understand and appreciate the subject of mathematics, it is imperative that they have opportunities to make connections between mathematics and daily life problems. Specifically pertaining to the concept of function, an effective way to help students make the connection is through the process of representing real world data with functions. One approach for accomplishing this process, as suggested by the Texas Educational Agency (TEA), involves instructing students how to examine characteristics of various data sets and construct their relations and functions. Teachers first need to have students closely investigate and make predictions of different data structures. Second, teachers will have students construct the patterns and the relational models. Before constructing models for different types of numerical data sets, it is important for students to distinguish the differences between discrete (or discontinuous) and continuous data.

Discrete vs. Continuous Data

One of the persistent misconceptions about data is the perception that all the data can be represented by a continuous function (Yerushalmy & Gafni, 1992, and Hitt, 1994). That perception is not true for all real life data collection; for instance, the data that represents the change of a country's population over a number of years is an example of a discrete data structure. This kind of misconception can be avoided if the teacher provides students with the fundamental concept of data with definitions and examples of discrete versus continuous data. The following table could help students to make a clear distinction between continuous and discrete data structures:
### Data and Its Mathematical Relationship

Both discrete and continuous data can be represented by a chart, table, or graph. A possible transition to help students distinguish a discrete data set from a continuous data set is by graphing. Before approaching the graphs, students should be allowed to develop a sense of the

<table>
<thead>
<tr>
<th>Discrete data</th>
<th>Continuous data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Discrete data is an attributed data that can be categorized into a classification. Discrete data is based on counts. Only a finite number of values is possible, and the values cannot be subdivided meaningfully.</td>
<td>• Continuous data is information that can be measured on a continuum or scale. Continuous data can have almost any numeric value and can be meaningfully subdivided into finer and finer increments, depending upon the precision of the measurement system.</td>
</tr>
<tr>
<td>• Discrete data can't be broken down into a smaller unit or add additional meaning. It is typically things counted in whole numbers. Population data is attributed because you are generally counting people and putting them into various categories. There is no such thing as 'half a defect.' It doesn't really add additional 'meaning' to the description.</td>
<td>• Continuous data is data that can be measured and broken down into smaller parts and still have meaning. Money, temperature, time, volume, and size are continuous data. As opposed to discrete data like good or bad, off or on, etc., continuous data can be recorded at many different points (length, size, width, time, temperature, cost, etc.).</td>
</tr>
</tbody>
</table>

(see: [http://www.isixsigma.com/dictionary/Continuous Dat-96.htm](http://www.isixsigma.com/dictionary/Continuous Dat-96.htm))
data so that they can identify the input from the output or the dependent from the independent variable. Many teachers have found students have difficulty in identifying the dependent from the independent variable. Therefore, Konold and Higgins suggest the use of a heuristic rule, such as independent variable often varies over time (time series), or the independent variable can be chosen or controlled (Konold & Higgins, 2003). Those may help students overcome the difficulty in identifying variables for various data sets. In general, if a data set is given in the form of a table, the students often assume either the first row or the first column will represent the independent variable. That assumption is generally true for most cases, but not all.

Returning to the discrete and continuous data, the following three tables would help students notice the difference between the graphs on which they should not “connect the dots” and the graphs on which they should “connect the dots”:

Table (a) below represents the number of participants each year in the Turkey Trot 5k:

<table>
<thead>
<tr>
<th>Years</th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>121</td>
<td>151</td>
<td>170</td>
<td>209</td>
<td>250</td>
<td>270</td>
<td>299</td>
</tr>
</tbody>
</table>

Table (b) below represents the length of the Rabbit Run and enclosed area:

<table>
<thead>
<tr>
<th>Length of run in meters</th>
<th>0</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area in square meters</td>
<td>0</td>
<td>10</td>
<td>18</td>
<td>24</td>
<td>30</td>
<td>30</td>
<td>28</td>
<td>24</td>
<td>18</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Table (c) below shows the data from a class survey on each student’s time spent watching television and working on homework:

<table>
<thead>
<tr>
<th>Hours watching TV</th>
<th>2</th>
<th>3</th>
<th>3</th>
<th>4</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>6</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours working on HW</td>
<td>19</td>
<td>25</td>
<td>10</td>
<td>16</td>
<td>19</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>14</td>
<td>9</td>
<td>13</td>
</tr>
</tbody>
</table>
Note: Table (a) represents discrete functional data, table (b) represents continuous functional data, and table (c) represents discrete non-functional data (or a discrete relation.)

Once the concepts of discrete and continuous have been mastered, and once students can model many situations with graphs, the relationships between two variables should be explored. In fact, after graphing the data, the students should recognize that:

A *mathematical relationship* is a connection between two numerical quantities. Changes in the value of one numerical quantity are associated with changes in the numerical value of the other numerical quantity.

At this point, the teachers should have students engaged in the process of examining the graphs that they have constructed. They should also have articulated some predictions or generated trends of the data. By doing so, students are preparing to make a transition from data to function.

**Relations and Functions**

Before leading students to modeling data, it is important to notify students that not all the relations between the two numerical variables form a function. Specifically, the teacher should make clear that ‘a function is a relation’ but ‘a relation is not necessarily a function.’ Only a relationship in which *each input value* is connected with *one and only one output value* is called a *function*. Moreover, a function is the best kind of relationship, which can be used to make predictions. Each input that is given a function will return one and only one unambiguous prediction for the output value. If the relationship is not a function, then students could get several predicted output values for each input. For example, from the Table (c), students can have multiple predictions on hours spent on working on homework for any students who spent nine hours watching television. Therefore, Table (c) only represents a relationship between the two numerical variables but does not represent a function.
Consequently, a given data set can be either a function or a relation. The following concept maps would help students recognize the characteristics of a data set.

![Concept maps of data to relation and function](image)

**Figure 1: Concept maps of data to relation and function**

**Transition from Data to Functions**

The successive step of making a prediction is forming a functional rule. No doubt, students generally have a difficult time transitioning data to functions. Seeing and analyzing the data set are considered fun things to do, but generalizing a pattern with a function, especially non-linear functions, is considered a tough job by many students (Eisenberg, 1982, & Yerushalmy, 1991). Therefore, it is suggested the teacher should first present the students with a data set that fits the pattern of a linear function. For example, using the data from table (a), the students quickly can translate the data to the following graph by labeling the year 1997 as the first year, 1998 as the second year, and so on:

![Graph of Turkey Trot 5k Participants](image)
By observing the graph, the students can tell that the data represents an approximate linear model. By picking any two pairs of points, for example, the pairs (2, 151) and (6, 270), and applying the point-slope equation, \( f(x) = mx + b \), the students are able to come up with a function \( f(x) = 30x + 90 \) which can be used to approximate the number of participants for any given year. Since real life data will not perfectly fit any particular pattern, students should be instructed on how to perform regression on graphing calculators to find the equation of the function that best represents the data set. In order for the students to make a real sense of the function from table (a), the teacher should pose questions to help students cope with the characteristics of the function and its prediction, such as:

1. From the data set, can you make a prediction of how many participants will participate in the race in 2005?
2. Do you really think that in the year of 2022 there will be 840 participants?
3. Does it make any sense to include negative numbers in the above function domain?
   Write the domain set for this function.
4. Can you tell whether the data set represents a discrete or a continuous function?
5. Should we connect all the scattering dots in this graph?
6. Does it make sense to evaluate \( f\left(\frac{3}{4}\right) \) or \( f\left(\frac{25}{6}\right) \)?

This particular example is discrete and linear in nature; but the question here is whether or not students recognize the discrete feature. If the teacher does not make a mark by posing questions like 4 to 6 above, most students routinely will assume that the function is continuous.

Determining a linear function of the form \( y = f(x) \) to model the data is done fairly easy by hand, and is a good exercise. But, if the data does not fit a linear model, then using a graphing calculator and other technology is more helpful for modeling data. Specifically, once
the students have learned the process and the significance of determining functions for data, then
the graphing calculator could be used to enhance the complexity of the examples. In brief,
forming a function from a data structure, students will go through the process of constructing
Table→Graph→Function.

Summary

Using real life data to approach functions has been considered a didactic modeling that
fosters the meaningfulness and usefulness of the so-developed mathematical knowledge. Going
from data to function constitutes an organizing activity that enhances students’ learning
experiences through investigation, interpretation, making predictions, generating patterns, and
finally, constructing a symbolic relation. In addition, the key to effective teaching is ensuring
that students are interested in the subject matter they are studying. Therefore, teachers are
encouraged to bring in applicable examples and have students work on manipulating data and
generating patterns and relations in approaching the concept of function.

References

Dubunsky (Eds.). *The Concept of Function: Aspect of Epistemology and
Pedagogy* (pp. 153-174). Mathematical Association of America, 25. Washington DC:
Mathematical Association of America.

Hitt, F. (1994). Teachers' difficulties with the construction of continuous and
