LESSON QUESTIONS: Introducing the sum function

FOCUS QUESTION: How can I transform the data to give more meaningful results?

Contents

- EXAMPLE 1: Load the NYC contagious disease data set (load .mat files)
- EXAMPLE 2: Calculate totals by year and by month (sum)
- EXAMPLE 3: Plot yearly total of measles by year (plot)
- EXAMPLE 4: Compare monthly totals of measles and mumps (multiple plots)
- EXAMPLE 5: Attempt to plot all the measles data as a single time series (linear representation)
- EXAMPLE 6: Correctly plot measles data as a single time series (transpose)
- EXAMPLE 7: Define the x-axis scale for a time series (subintervals a:inc:b)
- EXAMPLE 8: Plot measles as a time series, setting x-axis scale (computed scale)
- EXAMPLE 9: Plot a pie chart, comparing yearly measles counts for the first decade

EXAMPLE 1: Load the NYC contagious disease data set (load .mat files)

```matlab
load NYCDiseases.mat;
```

EXAMPLE 2: Calculate totals by year and by month (sum)

```matlab
measlesByMonth = sum(measles, 1);
measlesByYear = sum(measles, 2);
measlesTotal = sum(measlesByMonth);
```

<table>
<thead>
<tr>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>What does sum represent?</td>
<td>This identifier represents a MATLAB function whose name is sum. The items enclosed in parenthesis are the arguments to the function. MATLAB sends the arguments to the function as inputs when it executes (calls) the function.</td>
</tr>
<tr>
<td>What does sum(measles, 1) do?</td>
<td>This command sums along dimension 1, that is it adds up each column of measles.</td>
</tr>
<tr>
<td>How big is sum(measles, 1)?</td>
<td>Since measles is an array with 41 rows and 12 columns, the result will be a single row of 12 entries. In other words measlesByMonth is a 1 x 12 array.</td>
</tr>
<tr>
<td>Questions</td>
<td>Answers</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Since <code>measlesByMonth</code> has one row of 12 columns, why doesn't <code>plot(measlesByMonth)</code> show 12 different lines, each consisting of one point?</td>
<td>The MATLAB <code>plot</code> command normally plots each column of an array as a separate line. However, if the array only has one row, MATLAB plots all of its values as a single line.</td>
</tr>
<tr>
<td>How does <code>sum(measles)</code> differ from <code>sum(measles, 1)</code>?</td>
<td>In this case they are the same. If you omit the dimension, MATLAB sums along the first non-singleton dimension (the first dimension that has more than 1 element). In this case the first non-singleton dimension is 1, the row dimension.</td>
</tr>
<tr>
<td>What does <code>sum(measles, 2)</code> do?</td>
<td>This function computes the sum of <code>measles</code> along dimension 2 (i.e., it collapses the columns). The call returns a single column consisting of the row sums of <code>measles</code>.</td>
</tr>
<tr>
<td>How big is <code>sum(measles, 2)</code>?</td>
<td>Since <code>measles</code> is an array with 41 rows and 12 columns, the result will be a single column of 41 entries.</td>
</tr>
<tr>
<td>What happens if I omit the 2?</td>
<td>MATLAB uses the default value, which is the first non-singleton dimension (1 in this case).</td>
</tr>
<tr>
<td>How big is <code>measlesByYear</code>?</td>
<td>The <code>measlesByYear</code> variable holds a single column of 41 values corresponding to the 41 row sums of <code>measles</code>.</td>
</tr>
</tbody>
</table>

**EXAMPLE 3: Plot yearly total of measles by year (plot)**

```matlab
figure

plot(years, measlesByYear./1000)

xlabel('Year')

ylabel('Total cases (in thousands)')

title('NYC measles cases')
```
What does \( A./B \) represent in MATLAB?

\( A./B \) designates element-by-element division of the array \( A \) by the array \( B \). MATLAB creates a new array whose the elements are the elements of \( A \) divided by the corresponding elements of \( B \). If \( B \) is just a single number, then all of the elements in \( A \) are divided by the value of \( B \).

What did dividing \( \text{measlesByYear} \) by 1000 accomplish?

In EXAMPLE 3, the y-axis would run from 0 to 80,000 without scaling. MATLAB shortens the labels and displays an exponent of \( 10^4 \) above the y-axis to make the display more readable. However, the exponent is easy to miss, and the \( 10^4 \) is hard to comprehend. By converting to thousands and adjusting the y-axis label accordingly, you will make the sizes easier for your viewer to understand.

EXAMPLE 4: Compare monthly totals of measles and mumps (multiple plots)

```matlab
figure
hold on
plot(measlesByMonth./1000, '-rs')
plot(sum(mumps)./1000, '-ko')
hold off
```
Why bother with plot markers when you have colors?  Unless you pick colors carefully the lines may not be distinguishable when printed on a black and white printer. Also, a significant number people experience color blindness of various sorts.

**EXAMPLE 5: Attempt to plot all the measles data as a single time series (linear representation)**

```matlab
figure
plot(measles(:));
```
Questions | Answers
---|---
**What does `measles(:)` mean?** | `measles(:)` is the linear representation of the `measles` array. MATLAB forms the linear representation by placing the columns of `measles` end-to-end to make a single vertical column.

**Does the linear representation of an array contain the same values as the original array?** | Yes, the values don't change, only the arrangement in rows and columns.

**What is order of elements in the linear representation of an array?** | All of the elements from the first column (in order) are followed by all of the elements of the second column (in order), etc.

---

**EXAMPLE 6: Correctly plot measles data as a single time series (transpose)**

```matlab
measlesFlip = measles';

figure

plot(measlesFlip(:))
```
What does the single quote mean here?
The prime (') is the transpose operator. Note: that if the word measles were enclosed in single quotes it would be a string, but here it designates an array operator.

What does the transpose operator do?
The transpose operation flips an array on its main diagonal, making rows into columns and columns into rows. Thus, measles' is a new array with the same values as measles, but the rows become the columns.

**EXAMPLE 7: Define the x-axis scale for a time series (subintervals a:inc:b)**

```plaintext
yearStart = 1931;
yearInc = 1/12;
yearEnd = 1972 - yearInc;
yearScale = yearStart:yearInc:yearEnd;
```
<table>
<thead>
<tr>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>What does the notation ( a:b ) mean?</td>
<td>( a:b ) creates a row of values that are one apart, starting with the value of ( a ). None of the values can be bigger than ( b ). Thus, 1931:1971 is a single row consisting of the values 1931, 1932,...,1971. This is how we created the ( \text{years} ) variable.</td>
</tr>
<tr>
<td>Does ( a:b ) still work if ( a ) and ( b ) are not integers?</td>
<td>Yes, the list still starts with the value of ( a ), and the values are still one apart. However, the list may end with a value less than ( b ), if ( a ) and ( b ) are not separated by an integral value.</td>
</tr>
<tr>
<td>How would I produce a column containing the values 1931, 1932, ..., 1971?</td>
<td>Use the transpose operator to turn a row into a column. Thus ((1931:1971)') is a column containing the values 1931, 1932, ... 1971.</td>
</tr>
<tr>
<td>What does ( a:\text{inc}:b ) do?</td>
<td>This expression creates a row vector of the values that are ( \text{inc} ) apart: ( a, a + \text{inc}, a + 2*\text{inc}, ..., b ). If ( b-a ) isn't evenly divisible by ( \text{inc} ), the list stops before reaching ( b ).</td>
</tr>
<tr>
<td>What if ( \text{inc} ) doesn't evenly divide ( b-a )?</td>
<td>The row vector will include the largest value of ( a + n*\text{inc} ) that is less than ( b ).</td>
</tr>
<tr>
<td>Why is ( \text{yearInc} ) equal to 1/12?</td>
<td>The points are spaced one month apart, but the units are in years. One month is 1/12 of a year.</td>
</tr>
<tr>
<td>Why is the ending point not 1972?</td>
<td>The last point in the data set represents December of 1971, not January of 1972.</td>
</tr>
</tbody>
</table>

**EXAMPLE 8: Plot measles as a time series, setting x-axis scale (computed scale)**

```matlab
figure

plot(yearScale, measlesFlip(:)./1000);

xlabel('Year')

ylabel('Cases (in thousands)')

title('NYC measles cases')
```
EXAMPLE 9: Plot a pie chart, comparing yearly measles counts for the first decade

Create a new cell in which you type and execute:

```matlab
yearLabel={'1931','1932','1933','1934','1935','1936','1937','1938','1939','1940'};

figure

pie(measlesByYear(1:10), yearLabel)

title('Annual count of NYC measles cases (1930-1941)')
```

<table>
<thead>
<tr>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why would I want to use a pie chart?</td>
<td>Pie charts are a useful way to communicate the percentage that each element in a vector or matrix contributes to the sum of all elements</td>
</tr>
<tr>
<td>Why do I need to type all those values in for a label?</td>
<td>MATLAB's <code>pie</code> command only uses text (not numbers) to label the individual pie pieces, so these must be entered separately, and enclosed in curly braces.</td>
</tr>
</tbody>
</table>
These questions were written by Kay A. Robbins of the University of Texas at San Antonio and last modified by Dawn Roberson on 4-Jan-2014. Please contact krobbins@cs.utsa.edu with comments or suggestions.