# Creative Coding

# Lab #3 (Disease.xls Demo) PART 1

Say we want to create some graphs based on the data in Disease.xls. There are many options (6 columns) so we just use the relevant columns for the type of graph we want to create. For example we will not be creating a map at this stage so we don’t need the Longitude and Latitude columns.

In this case the data is clean i.e. there are no gaps or incorrect formatting, but for the ‘Date’ column, we don’t need the full date if we only want to show disease breakdown by year – we only need the year. See option 1 below.

**Option 1:**

For example if we want to create a graph showing disease breakdown by year, we will need to create the relevant table – see Table 1 and Figure 1 below.

Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Disease Breakdown by Year** | | |  |  |  |
| **Year** | **2012** | **2013** | **2014** | **2015** | **2016** |
| Cryptosporidiosis | 28 | 24 | 17 | 39 | 51 |
| Giardiasis | 12 | 10 | 18 | 36 | 40 |
| Verotoxigenic Escherichia coli infection | 41 | 75 | 74 | 72 | 104 |

Figure

**Steps:**

1. First create a table with labels only – see Table 2. We have to populate this table with values from the given data.

Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Disease Breakdown by Year** | | |  |  |  |
| **Year** | **2012** | **2013** | **2014** | **2015** | **2016** |
| Cryptosporidiosis |  |  |  |  |  |
| Giardiasis |  |  |  |  |  |
| Verotoxigenic Escherichia coli infection |  |  |  |  |  |

1. Next we can change the date format from the data:

Select all the cells in **Date** column and click ***format*** -> ***format cells*** from ***cells*** panel from ***Home*** from the main menu***.***

Select ***custom*** from menu and type in yyyy instead of dd/mm/yyyy.

**OR:** we can use the function YEAR() to return the year from a given date. This is the preferred option as the content of the cells change as opposed to just the format.

1. Note that in this case, if the years are not already sorted, we sort them using the Sort function. To do this, select all the data, select ***Data*** from the main menu. Then select ***Sort*** from the ***Sort and Filter*** panel and sort by ‘***Date’***, ***Sort on ‘Cell Values’,*** and ***Order ‘Oldest to Newest’.***
2. When sorted, in the empty cell below **2012** (in the table) count all cells in disease column for 2012 using countif(‘select all the cells in the disease column for 2012 – fix using dollar signs (F4)’, ‘select the cell with Cryptosporidiosis on table 2’).
3. Next use autofill (down the column) to fill disease incidence for Giardiasis and E Coli.
4. Follow the same steps (3-4) for 2013 to 2016.
5. Now create a bar chart based on this table – see Figure 1 above.

**More charts:**

Note we can also create a line chart which is good for showing changes over time (Figure 2) and a bar chart where the series are the years instead of the diseases (Figure 3).

Figure

Figure

**Option 2:**

Say if we want to show gender breakdown by disease. See Figure 4 below.

**Steps:**

1. First create a table with labels only. We have to populate this table with values from the given data – see Table 3.
2. First sort the Disease column - to sort the **Disease** column, we select all the data, select ***Data*** from the main menu. Then select ***Sort*** from the ***Sort and Filter*** panel and sort by ‘***Disease’***, ***Sort on ‘Cell Values’,*** and ***Order ‘A – Z’.***
3. Fill in Table 3 using countif(), autofill etc. similar to option 1.

Table 3

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gender Breakdown by Disease From 2012 - 2016**   |  |  |  |  | | --- | --- | --- | --- | | **Disease** | **Cryptosporidiosis** | **Giardiasis** | **Verotoxigenic Escherichia coli infection** | | **F** |  |  |  | | **M** |  |  |  | | **U** |  |  |  | |
|  |
|  |
| Figure |
|  |
|  |
| **Option 3:**  Create Figure 5 below from disease.xls raw data. Note that you will have to divide the ages into appropriate intervals.  **Steps:**  **Note:** You will only need the disease + age columns.   * Sort the disease column. * Get the max and min age. * Calculate the number of age intervals + interval width using formulae from lab 2. * Now work out the frequency for Crypto.   **See below:**   |  |  |  | | --- | --- | --- | | **ages** |  | **frequency (Crypto)** | | 0 to 8 | 8 | 68 | | 9 to 17 | 17 | 27 | | 18 to 26 | 26 | 23 | | 27 to 35 | 35 | 22 | | 36 to 44 | 44 | 12 | | 45 to 53 | 53 | 2 | | 54 to 62 | 62 | 3 | | 63 to 71 | 71 | 1 | | 72 to 80 | 80 | 0 | | 81 to 89 | 89 | 1 | |  |  | 159 |  * Complete the frequency distribution table for Giardiasis and E Coli. You will now have three tables – one for Crypto (above), one for Giardiasis. and one for E Coli. * Now create the clustered column chart below based on the three tables. |
| Figure |

**Questions:**

1. Attempt options 1, 2 and 3 using the methods described above.
2. Attempt options 1 and 2 using countifs().

**Useful functions:**

=countif()

=countifs()

=frequency()

=left()/right()/mid()

=year()/month()/day()

=text(address, “mmm”)

=max()

=min()

=quartile()

=median()

=average()

Sort (from Data tab)

Format -> Format Cells (from Home tab)

count()

countA()

Filter (from Data tab)

Text to columns (from Data tab)

Etc.