Instructions on Adding Zeros to the Comtrade Data

Required: An excel spreadshheet with the commodity codes for all products you want included. In this exercise we will want all 4-digit SITC Revision 2 commodity codes. These are included in the file ***4DigitSITC2Codes.xls***

Step 1 – (Download Data from Comtrade)

> 1.a) Go to http://comtrade.un.org/db/dqQuickQuery.aspx it should be accessible in all browsers
>
> 1.b) Click the radio button for **SITC Rev. 2** under *Select Classification*
>
> 1.c) Under *Enter Commodity Codes/Text* enter **????** (this tells it to include all 4 digit codes)
>
> 1.d) Under *Enter Reporters Codes / Text* enter the code for your reporting country. E.g. **842** for the United States. You can click *Lookup* to add reporters by name instead of code.
>
> 1.e) Under *Enter Partners Codes / Text* enter the code for your partner countries, separated by commas if there are more than one. E.g. **124, 484** for Canada and Mexico, respectively. You can click *Lookup* to add partners by name instead of code.
>
> 1.f) Under *Enter Years* enter all years that you want data for, separated by commas. E.g. **1990,1991,1992,1993,1994,1995** for 1990 to 1995. There is no way give it just a start and end point to have it include a range; each year you want included must be listed explicitly.
>
> 1.g) Under *Select Trade Flow* click the check boxes next to the trade flows you want, e.g. **Imports** and/or **Exports.**
>
> 1.h) Click *Submit Query*
>
> 1.i) When the *Basic Query Results* page loads click *Direct Download*. Click OK then *Save as* and download the data to your computer. ******If your query returns over 50000 records you will be unable to download it. If this is the case then you must go back and include fewer years or fewer partners** If you get a server error and it won't download, just go back and retry until it works.
>
> *Alternative:* Comtrade data can also be downloaded by creating an account at WITS. https://wits.worldbank.org/WITS/WITS/Restricted/Login.aspx although instructions will not be provided for this site.

Step 2 – Use a Pivot Table to Organize Data by Years

2.a) Open your downloaded data from Step 1 in Excel.

    For Excel 1997-2003:

2.b) First click on a cell containing some data and then click *Data* on the top and then *PivotTable and PivotChart Report…* from the dropdown menu

2.c) Leave *Microsoft Excel list or database* and *PivotTable* selected and click *Next*.

2.d) It will ask you what data you want to use and should automatically guess and be filled in with the correct range if you first clicked a cell in your dataset in step 2.b. Click *Next*

2.e) When it asks *Where do you want to put the PivotTable report* click the radio button next to **New Worksheet**. Click *Finish.*

2.f) In the new sheet with the empty pivot table drag **Years** to *Drop Column Fields Here*, drag **Commodity Code** to the *Drop Row Fields Here* and drag **Value** to *Drop Data Items Here.* It should list it as *Sum of Value*, which is what you want (average, min, and max are also fine since there is only one data point).

    For Excel 2007-2010

2.b) First click on a cell containing some data and then click *Insert* on the top and then *PivotTable* in the below ribbon.

2.c) It will ask you what data you want to use and should automatically guess and be filled in with the correct range if you first clicked a cell in your dataset in step 2.b. Leave the radio buttons next to **Select a table or range** and **New Worksheet** selected. Click *Next*

2.d) In the new sheet with the empty pivot table drag **Years** to *Column Labels*, drag **Commodity Code** to the *Row Labels* and drag **Value** to *Values.* It should list it as *Sum of Value*, which is what you want (average, min, and max are also fine since there is only one data point).

Step 3 – Get Values for all Commodity Codes

3.a) We're going to want to use this pivot table data elsewhere so we need to copy it to a new spreadsheet.

- To select all data in the rectangle given by the first commodity code in the upper left corner and the value for the last year and last commodity code for the lower right corner. You can do this by clicking and dragging with the mouse or by holding shift and using the arrow keys (you can press the end key and an arrow key to

move immediately to the last cell with data in it in that direction). Don't include the totals in the last column or last row of the pivot table in your selection.
- After the Data is selected press **Ctrl+C** or right click in your selection and click *Copy*
- Create a new worksheet (**Shift+F11**) and in this worksheet select cell A1 and paste the data with **Ctrl+V** or by right clicking and selecting *Paste*
- You should end up with just the inside of the pivot table, with no headers and commodity codes in the first column, the values for the first year of your period in the second column and so on to the values for the last year in your period in the last column. (The **Vlookup** function doesn't work inside active pivot tables so that's why we need the new sheet)

- Double click the worksheet name at the bottom and name it **pivotdata**

3.b) We now need the list of commodity codes and the data we just pasted in the same worksheet. Open the *4DigitSITC2Codes.xls* excel file and copy the worksheet to your excel file with the comtrade data.
- To do this right click on the worksheet name *Codes* and select **Move or Copy**.
- In the *To book:* select the name of the excel file containing your comtrade data
- Click *OK*

3.c) We are now going to use the **Vlookup** function to match up the data with the full list of codes. Click cell **C1** in the *Codes* worksheet and press **=** to start typing a formula.

- **Vlookup** takes the following arguments in the following order
  - *lookup_value :* this is the value we want to lookup. It will be the cell of the corresponding row in the first column. For the formula in C1 we want **A1**
  - *table_array :* this the table you are searching. We want it to be all the data in the *pivotdata* worksheet. We want to put **pivotdata!$A1:??** where **??** is the bottom right cell containing data in *pivotdata* with **$** signs before both the letter and number. E.g. **pivotdata!$A$1:$G$742** (the $ signs are very important, they denote absolute versus relative references. Note no $ infront of the 1 in $A1, because we want the 1 to change as we copy the formula to other cells)
  - *col_index_num* : This is the column that you want to return a value from when you find the row you're looking for. A value of 1 returns the first column (the code) while a value of 2 returns the year for your first period and so on. We'll want this to be **2** for C1. *[range_lookup]* :This is asking whether you want approximate matches. Always set this to **False**.
- Cell C1 should have something like **=VLOOKUP($A2,copypivot!$A$1:$G$742,2,FALSE)** in it

3.d) We now want to copy the formula to all cells.

- First copy it horizontally along the same column by selecting *C1* and copying the cell (**Ctrl+C**) and then selecting the cells to the right of it equal to the number of periods you have and then pasting the formula into them by pressing **Ctrl+V** or Enter.
- IMPORTANT: We need to change the *col_index_num* in the Vlookup function for each of the columns. So in cell *D1* we want it equal to **3**, in cell *E1* we want it equal to **4** and so on. This means that column D will contain the data for the second year in your period, E for the third year, etc. You can use the function **row**() to update the column as you copy the formula for other years.
- We now want to copy the formula to all the cells below the first row down to the last row (789 in this case). To do that select all the cells with the Vlookup function in the first row and copy them (**Ctrl+C**). Then highlight all the cells below them down to the last row and paste the formula with **Ctrl+V** or Enter.

Step 4 – Replace #N/A's with Zeros

4.a) When **Vlookup** fails to find what it's looking for (in this case indicating a dropped zero) it returns #N/A; but we want those cells to be zeros

4.b) Copy all the data in the worksheet with the **Vlookup** formulas by pressing **Ctrl+A** and then **Ctrl+C**

4.c) Create a new worksheet (**Shift+F11**) and select the place where you want to paste your data (e.g. **A1**) and right click in that cell. Select **Paste Special** from the Right click menu and then select **Values** and click okay to paste as values. Note a normal paste will not work here.

4.d) To now replace the #N/A's with zeros press **Ctrl+H** to bring up the find and replace option. Under *Find what:* put **#N/A** and under *Replace with:* put **0** then click *OK*.

4.e) We now have added the dropped zeros back into the comtrade data and we can use this data to complete the least traded exercise.

- It may be useful to rename this worksheet with the final data to a descriptive name or save it to its own spreadsheet as well as to label it with years and such. At this point you can also delete the first column of the commodity codes without descriptions if desired.
- Alternatively, step 4 can be done by using the IF function combined with the **IsNumber** function in step 3, to automatically return a 0 if #N/A is reported.

Step 5 – Sorting the Data

Instructions from here are checked for Excel 2007 and later only. May be slightly different for Excel 2003.

5.a) Create a new sheet and name it something, for example **Sort and Bin**. Copy over all the data by using **Ctrl+A** then **Ctrl+C** to copy all the data from step 4. Then go the new page and **paste as values** (see step 4.c)**.**

5.b) I deleted all columns except the first 3 years and the last year, but that isn't necessary. In the first empty column after your data, put the name **First3Years**, and then use the **Average** function (or **Sum** function) to combine the value over the first 3 years so you can sort it. Copy the formula to the whole column.

5.c) Right click on a value in your **First3Years** column other than the title, and go to *sort* and then *sort smallest to largest.*

5.d) In the next empty column, create a column named **Fraction** which contains the fraction of trade each product accounted for in 1989. I use the formula =**C2/SUM($C$2:$C$790)** in the example to calculate it in row 2 and then copy the formula down, where column C has the value of the product in 1989.

5.e) Now create one more column named **Cumulative Fraction** which contains the cumulative fraction of trade by each product, according to the first 3 year sort order. I use the formula =**SUM($C$2:C2)/SUM($C$2:$C$790)** in the example to calculate it in row 2 and then copy the formula down, where column C has the value of the product in 1989.

Step 6 – Binning the Data

6.a) We now want to assign each product to a bin. Skip an extra empty column, and then make 10 columns for the 10 bins. Name them 1 through 10.

6.b) I am going to use the formula =**IF($I2<L$1/10,1-SUM($K2:K2),IF($I1<L$1/10,($H2-($I2-L$1/10))/$H2,0))** to assign products to bins. I'll break down the formula and what it does.

IF($I2<L$1/10, : **If** *the cumulative fraction in row 2 is smaller than the Bin in column L divided by 10 (recall I named the bins 1 through 10)*

Then *1-SUM($K2:K2), : put 1, since it's in the bin. Except if the product has already been placed in a previous bin, then subtract what's already been put into another bin (this is for when a product is split between two bins). Note K2 is an empty column.*

Else If *IF($I1<L$1/10, : the previous product was in this bin (indicating this product is a transition product split between two bins). Note: you could also just have an if statement for if the above value =1*

Then *($H2-($I2-L$1/10))/$H2, : put in the fraction of this product required to make the cumulative sum exactly meet the bin cutoff. The formula is the Fraction minus how much the Cumulative Fraction exceeds the cutoff, divided by the Fraction again.*

Else *0)) : put 0 since the product is not in the bin.*

6.c) Now we want to compute the value each product contributes to each bin in the end year. So create 10 more columns with the bin names, this times for values, and then multiply the end year value times the number we computed in 6.b) for each number. For example I have **=$F2*L2** to give me the value of the first product in the first bin, and then I copy that formula to all bins and products.

6.d) Now we want to calculate the number of products in each bin and the new value of trade accounted for by each bin. To do this go to the bottom of the worksheet, and under each bin just use the **sum** function to sum up all the values. Summing under the values from 6.b) gives the number of products, and summing under the values from 6.c) gives the value of trade in the end period.

Then compute the fraction of trade accounted for by each bin, by taking it's value of trade divided by the total value of trade for all bins.

Step 7: Plotting the Data

7.a) Go to **Insert** at the ribbon at the top, and then **Insert Chart** and choose **Bar Graph (2-d).** You may just have to click a little picture of a bar graph. Right click the empty graph, and click *Move Chart* and then *New Sheet*. Name the new sheet, e.g. **BinChart**.

7.b) Right click the chart and click *Select Data*. Choose the *series name* to be the bin names (with the title of my axis, I want the bins to be named .1 and .2 instead of 1 and 2 though; so either change the axis label or create a row of labels .1, .2,.3, …). For the *series value* choose the new fractions of trade accounted for by each bin that we just computed at the end of step 6.d.

7.c) **Format the graph to look "pretty."** Among other things, this means no legends, have large font everywhere, name the graph/axis if needed, choose colors that people can see (it's worth keeping in mind some economists are color blind), no unnecessary gaps or spaces, etc.. An easy way to do this is make one graph that you like, copy it, and then click your new unformatted graph and do *Paste Special* and then *Formats.* You can paste special by clicking the arrow under paste on the home tab, or by doing *Ctrl+Alt+V*.

7.d) Label each of the bars with the number of products in each bin. I don't know a fast way to do it all at once without a visual basic macro, so you can either do it completely manually by copying the number of products and pasting it **or** you can click each label, and then in the function area put the cell corresponding to the value, e.g. **='Sort and Bin'!$L$792**. Then at least if you change the value of cell L792 the label will update itself.