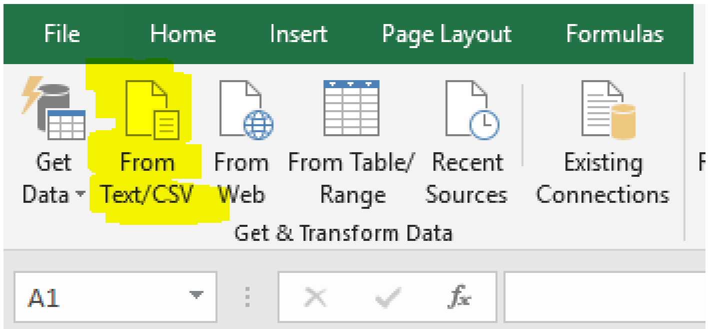
How to import text files to Microsoft Excel 2016:

You would use these directions if you get a delimited text file from a government agency (or some other source). This might be tab-delimited, comma-delimited or uses some other symbol as a delimiter. A comma-delimited file (or .csv extension) will open automatically in Excel if you double-click on it from File Explorer. However, if it doesn't bring all the values in correctly, you might need to use this import process. For example, if you have schools data with a field that says “17-18” for the schoolyear, Excel might interpret that as a date. Or if you have a unique identifier with leading zeros, Excel might treat that as a number and drop the leading zeros. Using this import process allows you to tell Excel exactly how to treat those things. And you will definitely need it for tab-delimited or files delimited with other symbols.

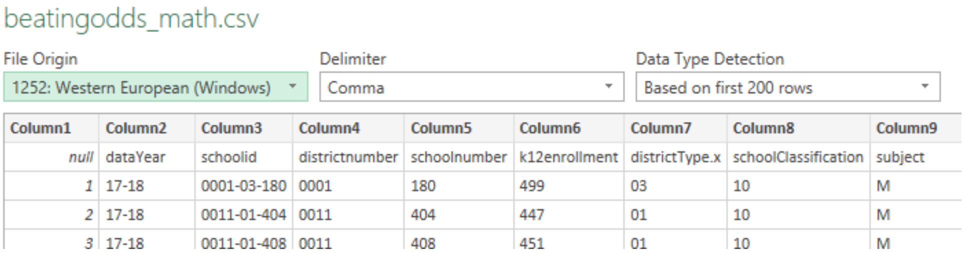
For this we will use this csv file:
http://mjwebster.github.io/DataJ/spreadsheets/importing_sample_data.csv

1. Launch Excel and open a blank workbook
2. Go to the Data menu and on the left side of the menu bar you will see “Get & Transform data” options.



You can EITHER click on “From Text/CSV” or go through the Get Data menu and choose From File….From Text/CSV. It will ask you to find you file on your computer and click Open.

3. A preview of your data will appear and at the top it will tell you that Excel has guessed what the delimiter is and that is detecting the data types for each field based on the first 200 rows. If those things are incorrect, you can change that here.   At the bottom of the preview window, click on the EDIT button. You can see on this preview that it's not recognizing my column names (it merely says “Column1”, “Column2” and then my column names look like they are the first row of data). We can fix this on the next step.



beatingodds_math.csv

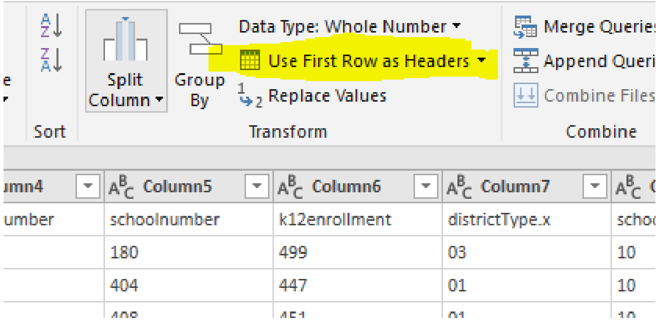
File Origin: 1252: Western European (Windows)
Delimiter: Comma
Data Type Detection: Based on first 200 rows

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9
null	dataYear	schoolid	districtnumber	schoolnumber	k12enrollment	districtType.x	schoolClassification	subject
1	17-18	0001-03-180	0001	180	499	03	10	M
2	17-18	0011-01-404	0011	404	447	01	10	M
3	17-18	0011-01-408	0011	408	451	01	10	M

4. It will look like your data is now in Excel, but it is not! In this example, it hasn't recognized our first row as the column headers. Usually it will. The problem here is that our first column doesn't have a name/header.

5. Let's check out the data types that Excel has assigned to each column. Because Excel still thinks our header row is a row of data, it is guessing that all of our columns are text (ABC). Especially look at districtnumber and schoolnumber. Notice that these columns have leading zeros.

   To fix this the header row problem, go up in the Transform section of the menu bar and choose "Use First Row as Headers"



umn4	Column5	Column6	Column7	C
umber	schoolnumber	k12enrollment	districtType.x	schoo
	180	499	03	10
	404	447	01	10

6. Now look at what happened to districtnumber and schoolnumber. Notice that the leading zeros are gone? And if you scroll to the right you'll see it has set some columns as numeric (123) or decimal (1.2). We don't have any date fields in here, but if we did you would see a calendar image next to the name.

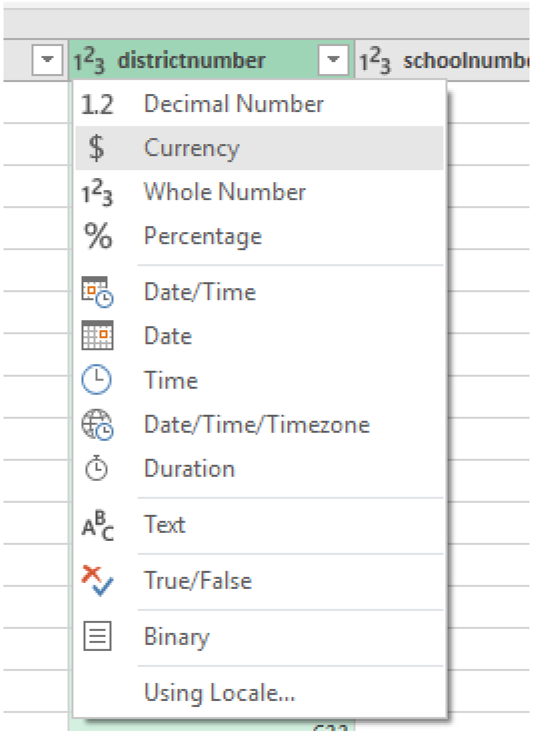
   What you need to do with any data import (regardless of software you're using) is CAREFULLY look through all the columns and make sure nothing is going wrong. Some common problems: dates not coming in correctly; unique identifiers that lost their leading zeros; zip codes losing leading zeros (several East Coast states have leading zeros on zip codes).

   Some basic rules of thumb: any numeric columns that are actually codes and won't be used for math – such as zip codes or unique identifiers – should be set as text; Make sure date columns are set as dates otherwise they won't sort properly; Make sure numeric columns that you want to do math on are set as numeric.

   In this example, it is trying to convert districtnumber and schoolnumber to numeric. They are actually unique identifier codes (which we won't use for math) and they originally had leading zeros.

For most data type changes, you can click on the data type symbol on that field and it will give you a list of things you can change it to.

However, with this leading zero problem, doing this will simply give you the values you see now without the leading zeros.



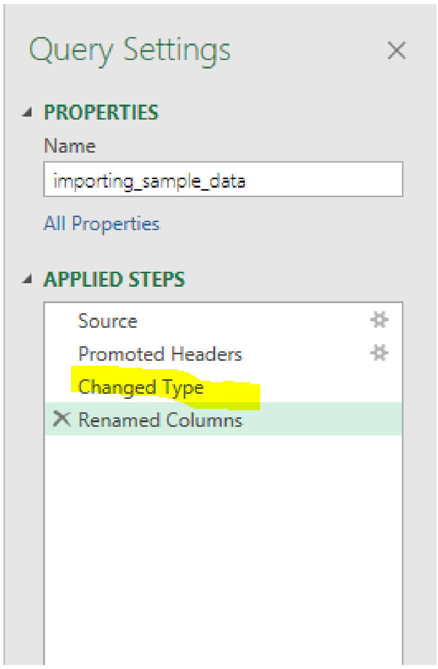
7. We can get the leading zeros back by going to the APPLIED STEPS window (on the right side of your screen). These are the various things we've done to our data so far. It allows you to "undo" past steps, in case you made a mistake.

   In this case you'll see that it "Changed type"

   Click on "Changed Type" and you'll see an X to the left of it. It will ask – are you sure you want to delete this step? Say "Delete"

   You'll see that ALL of the columns have reverted back to text (ABC).

   Unfortunately, now you need to go find the ones you want to set as numeric or decimal and manually convert each of those to what you want. (Using the directions in step 6)



Query Settings

PROPERTIES

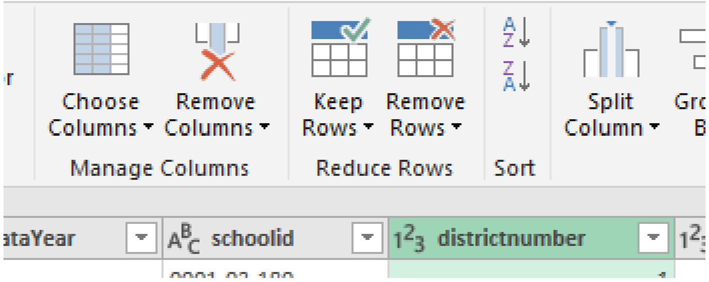
Name: importing_sample_data

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Renamed Columns

8. Excel also gives you some other options for editing your data as part of the import. In the menu bar are options for choosing which columns to keep and which to exclude; another lets you keep or exclude certain rows. For example, let's say you're bringing in a national dataset and you only want the records pertaining to Minnesota you could use the Keep Rows option to tell it to only keep rows where a particular column has the value of "MN" (Or whatever the value is that indicates it's Minnesota).

   In this example, we've got that first column that doesn't have a name. This is actually a column of row numbers that came with the file when I exported the data from R. It's an extraneous column we don't need. (I generally caution AGAINST excluding columns unless you are 100% certain you truly don't need it or that it's truly junk that will be useless to you. This is only something you can discern from the people you got the data from).

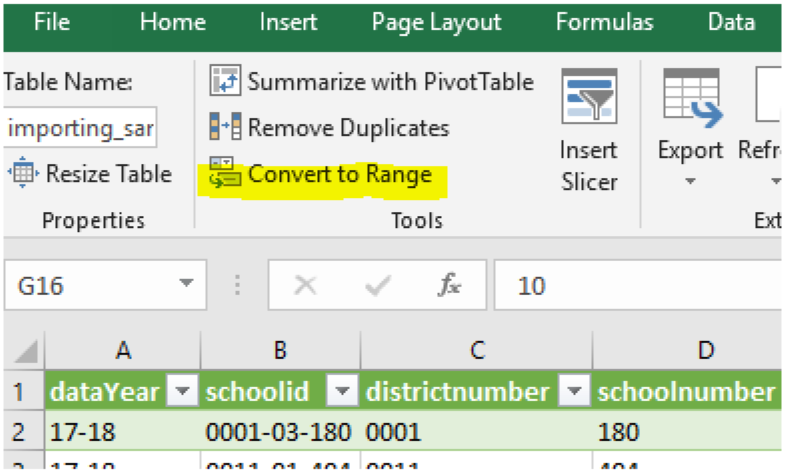


   Click on Keep Rows and you can uncheck the first column (meaning exclude it) and click OK. You'll see that column disappear from the preview.

9. Once you're happy with the settings, click the Close & Load button in the upper left corner. It will load it to the open, blank worksheet. Alternatively you can use that pull-down menu to "Close and Load to…" and specify where you want it to go.

   Excel will load it into a "Table" which has different properties and features than what you'd find in your typical Excel file. You'll see that it has colored the rows in alternating shades. The filters are automatically turned on. And there are lots of things going on in the background.

10. If you want to get it back to a more traditional Excel structure, go to the Menu bar and choose "Convert to Range" (it's under the Design tab and in the Tools section)



	A	B	C	D
1	dataYear	schoolid	districtnumber	schoolnumber
2	17-18	0001-03-180	0001	180

   You'll notice that the filters disappear, but the colors remain. If you really don't want the colors on there, highlight the whole datasheet by clicking on the triangle between the 1 and A…………………………..

And then click the "Normal" button in the menu bar (It's under the Home tab in the Styles section)

11. Now to go to the File menu and choose "Save As…" Find a place to save it, give it a name that is meaningful to you, and set the file type to "Excel Workbook (*.xlsx)"

12. Spend a little time perusing your data and make sure everything looks the way it should. One thing to look for is to make sure the data fell into the correct columns. For example, are the values in the address field really an address? Try turning on the filters and see what values are there. For example, look at the filter for Subject – are all the values M and R?

    If you are seeing values in the list that don't seem like they belong, it MIGHT be an import problem. Or you might just have dirty data. Regardless, now is the time to deal with it.

    It's especially important to check the columns on the far right because if there's an import problem it will affect all the columns to the right of whichever column the problem occurred in. So if you have 20 columns and something went wrong in column 15, then it will only be those last 6 columns that have problems.