



Unfortunately, data isn't always in the format that Tableau likes it. That's okay – we can fix that. The following is a step by step guide to reshaping your data so you can make the most of it in Tableau.

First, make sure there are no blank rows or columns within your data set. That's usually pretty easy to fix manually but if there's a systematic problem (like every fifth row is empty) you might prefer to use this nifty tool called [Data Wrangler](#) (more on this later) developed by the good people at the Stanford Visualization Group.

The data also needs to be in a data table format where each row contains only one piece of data. What on earth does that mean you ask? Well, it basically means that cross tabulations are bad.

**Normalized Data (you want your data shaped this way):**

General Electric	2010	11644
General Electric	2009	11025
General Electric	2008	17410
Siemens AG	2010	5554
Siemens AG	2009	3650
Siemens AG	2008	8504
Koninklijke Philips Electronic	2010	1948
Koninklijke Philips Electronic	2009	608
Koninklijke Philips Electronic	2008	-262

### Crosstab Data (you don't want it shaped like this):

Net Income (millions of dollars)			
Company	2010	2009	2008
Citi Group	10602	-1606	-27684
General Electric	11644	11025	17410
Siemens AG	5554	3650	8504
Koninklijke Philips Electronic	1948	608	-262

While crosstabs make sense to people, it makes our software's head hurt because each row contains three pieces of data -- in this case the company's net income in 2010, 2009 and 2008. Tableau feels much more comfortable when data is in a normalized format, where each row contains only one net income figure.

It can be a huge pain to do all the copy and pasting necessary to change data into this form. But the Data Wrangler can help with this too. Here's a step by step guide using the cross-tab data from above:

Example Data:

Company	2,010	2,009	2,008		
Citi Group	10,602		(1,606)		(27,684)
General Electric		11,644		11,025	17,410
Siemens AG	5,554	3,650	8,504		
Koninklijke Philips Electronics	1,948	608	(262)		

After pasting your data click "wrangle."

Wrangle

Don't worry if the columns don't line up

Now promote the top row of data to be the column headers by highlighting it and selecting "promote" from the rows drop down.

rows

- Split data repeatedly on 'tab'
- Delete row 1

Select the first row.

2	Citi Group	10602
3	General Electric	11644
4	Siemens AG	5554
5	Koninklijke Philips Electronics	1948

Text Columns Rows Table Clear

Delete  
Fill  
Promote

Delete row 1

Select "promote" to turn the first row into column headers.

Remember to click the "plus" button to confirm the change.

Text Columns Rows Table Clear

Promote row 1 to header

Click the "plus" sign to execute.

Split data repeatedly on newline into rows

Split data repeatedly on 'tab'

Delete row 1

Promote row 1 to header

Text Columns Rows Table Clear

Fold \_2010, \_2009, \_2008 using header as a key

Fold \_2010, \_2009, \_2008 using 1 as a key

Fold \_2010, \_2009, \_2008 using 1, 2 as keys

Company	_2010	_2009	_2008
1 Citi Group	10602	-1606	-27684
2 General Electric	11644	11025	17410
3 Siemens AG	5554	3650	8504
4 Koninklijke Philips Electronics	1948	608	-262

Select the columns (ctrl+shift) for each year of net income.

Then select the "fold" option that reshapes your data so that there is only one net income value per row.

Company	fold	value
1 Citi Group	_2010	10602
2 Citi Group	_2009	-1606
3 Citi Group	_2008	-27684
4 General Electric	_2010	11644

Delete the underscore that Data Wrangler added to your data. We don't want that. Select the text in one cell and then select "cut from fold."

Split **data repeatedly** on **tab**

Delete **row 1**

Promote **row 1** to header

Fold **\_2010, \_2009, \_2008** using **header** as a key

3	Unit Group	2008	2008
4	General Electric	2010	2010
5	General Electric	2009	2009
6	General Electric	2008	2008
7	Siemens AG	2010	2010
8	Siemens AG	2009	2009
9	Siemens AG	2008	2008
10	Koninklijke Philips Electronics	2010	2010
11	Koninklijke Philips Electronics	2009	2009
12	Koninklijke Philips Electronics	2008	2008

Text Columns Rows Table Clear

Split **fold** between positions **0, 1**

Split **fold** on **'\_'**

Split **fold** before **'2010'**

Split **fold** before **'any number'**

Extract from **fold** between positions **0, 1**

Extract from **fold** on **'\_'**

Extract from **fold** before **'2010'**

Extract from **fold** before **'any number'**

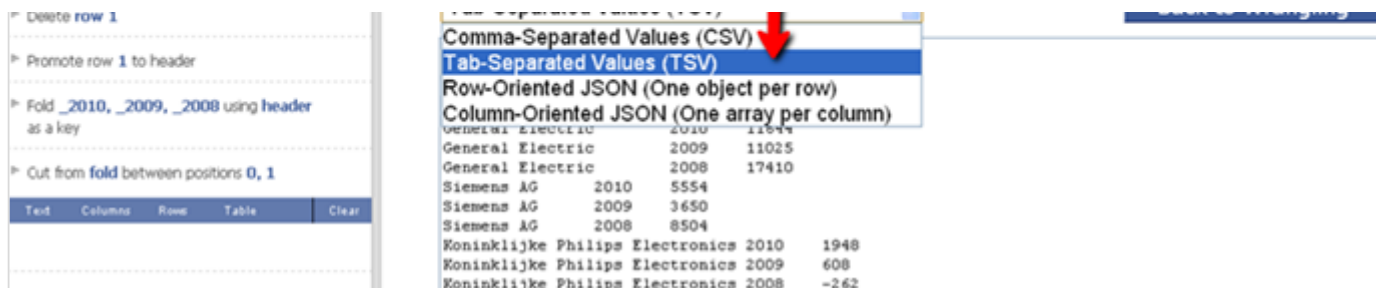
Cut from **fold** between positions **0, 1**

Blog

Highlight the undesired text.

Select the "cut from fold" option.

All that remains is to export the data back to Excel and connect to Tableau:


[Previous Post](#)
[Next Post](#)

## Comments

Do you guys like Data

Submitted by Tim on Fri, 2011-04-08 12:46 | [Permalink](#)

Do you guys like Data Wrangler better than the Tableau-Excel Addon?

Props to the Wrangler team!

Submitted by Joe Hellerstein on Fri, 2011-04-08 13:42 | [Permalink](#)

Props to the Wrangler team! For the record, DataWrangler is a Stanford/Berkeley collaboration, part of the [d^p](#) project.

Here's to collaboration!



## I need some excel help when

Submitted by Randy on Mon, 2011-08-01 12:27 | [Permalink](#)

I need some [excel help](#) when submitting my excel data into Tableau. I'm not sure if this data is in the format that Tableau likes. My excel spreadsheet only has one piece of data in each row and each cell. It still isn't uploading correctly. Is it better to save my excel in a csv format?

## Please read about an

Submitted by Petros on Wed, 2012-05-02 18:10 | [Permalink](#)

Please read about an automated Excel unpivot utility, similar to data wrangler

<http://www.spreadsheet1.com/unpivot-data.html>

Advantages:

The conversion utility can accommodate any number of fixed columns to the left the data and a header row that is not immediately above the data.

Blank or zero cells could be excluded from the output database.

Column A in the database lists the row number of the input table. This information is useful when a pivot table is created from the new database table and rows should be sorted according the original sort order, for example when converting financial statements e.g. a multi-year balance sheet.

Arrays are used in order to minimize computing time by 'hitting' the Excel grid only twice.

The input parameters can be saved as a named template for future use on identical spreadsheets.

## They're different. My first

Submitted by victor blaer on Mon, 2013-01-28 16:26 | [Permalink](#)

They're different. My first plan of attack is the add on, if the data is too crazy I bring in the Wrangler.





**Products**

Tableau Desktop

Tableau Server

Tableau Online

Tableau Public *Free*

Tableau Reader *Free*

**Legal**

Privacy Policy

Data Policy

Terms of Service

Patents & IP

**About**

Current Status

FAQ

Careers

Contact Us



English