

Example of cut and paste data gathering:

Incidence of colorblindness by country (html table).

http://www.wrongdiagnosis.com/c/color_blindness/stats-country.htm

1. First, cut and paste into Word as unformatted text
2. Then, use search and replace to remove all of the gratuitous warnings about extrapolated data. For example, replace WARNING (Details) with nothing
3. Use search to check for double tabs and double spaces
4. The final wrinkle is that the superscript footnote numbers at the end of each population estimate are now appended to the population numbers (eg. An extra 1 or 2). Fortunately, they put this on each and every population value. So, you can use Word's advanced replace feature to remove any number preceding a paragraph marker. Use the special symbols menu so you can remove any number followed by an endline. (^#\^p with ^p)
5. Now you have a tab separated set of data for Country, population that are colorblind, and total population. However, there are extra rows that specify the world region for the country that must be reshaped into another column labeling the following rows. This must be done by hand and is a bit tedious but straightforward. Finally, a couple of global replacements are needed to remove the extra words around the region name.
6. Now you can import the data into Excel, which makes it readable by Tableau
7. Unfortunately, using Tableau to graph this data vs. population demonstrates that the data was calculated from a single colorblindness rate in the first place – so in fact, there is no real data here. Better to realize this early than to base your project upon it.

Example of data reshaping:

Our survey data has several groups of columns that contain answers to multi-answer questions (eg, courses taken, programming languages known, etc). An Excel plugin, together with some simple cutting and pasting, can be used to reshape this data for use by Tableau. As an example, the sheet labeled Q36 in the Lab 3 survey data INFO424-SurveyF07 v3.xls was created by manual selection from the full Survey sheet. The ResponseID for each row acts as a unique identifier. Using the “Reshape Data” command in the menu bar, the plugin created the reshaped Q36-Tableau sheet, which can be used to create distribution graphs for types of pets.

Plugin Link: <http://courses.washington.edu/info424/survey/Setup-Tableau-Add-In-3.1.zip>

Example of downloadable data:

Summary climate data from the National Climate Data Center

<http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv>

Two separate text files provide 1) data, keyed by weather station index, and 2) Information about each weather station in a format easy to import into Excel.

1. Use the Data>Import function, or simply cut and paste the climate data file into an Excel spreadsheet.
2. Similarly, import the weather station information into a separate sheet in the same Excel file.
3. Connect to this file from Tableau.
4. Tableau can Join the two sheets using the weather station index field.

Many similar sets are listed at: <http://www.ncdc.noaa.gov/oa/mpp/freedata.html> The set above is from the Free Data J table.