

From Open Refine to RAW: Visualizing Change Over Time

RAW is a free and fairly simple web tool that creates colorful, vector-based visualizations from spreadsheet data. This tutorial will show you the basics of how to upload and visualize data in RAW, focusing on graphics that display change over time.

We're going to work with a dataset that needs modification in order to suit the parameters set by RAW's software. In order to prepare the data, we'll first take it into Open Refine, make a quick alteration, then upload the data into RAW.

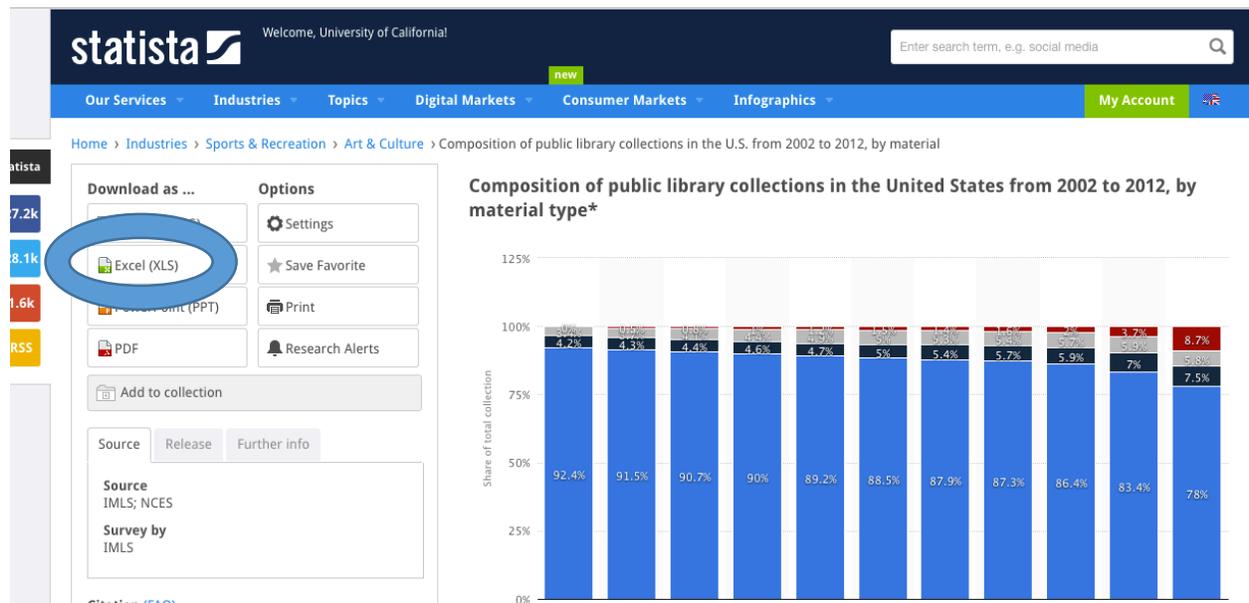
For details on how to install the Open Refine application, see "Open Refine Tutorial 1: Cleaning Messy Data"

Getting Started

We're going to work with statistics on the types of materials found in U.S. libraries from 2002-2012, according to the Institute of Museum and Library Services' Public Library Survey. A chart will make it easier to visualize how collection trends in the U.S. have changed over a decade's time.

To access the data, you'll need to go to the website Statista, which has converted the report's data into a downloadable excel file. Access to this service is free, but you'll need to be logged into a UCLA server to use it (if you're at home, you'll need access to a proxy server)

<https://www.statista.com/statistics/233950/materials-at-us-public-libraries-by-type/>



Find the Excel icon and download the XLS(S) file to your computer. Once downloaded, open the file – you can see that the survey has collected information on the number of print materials, audio materials, video materials, and e-books that have been collected by libraries over the

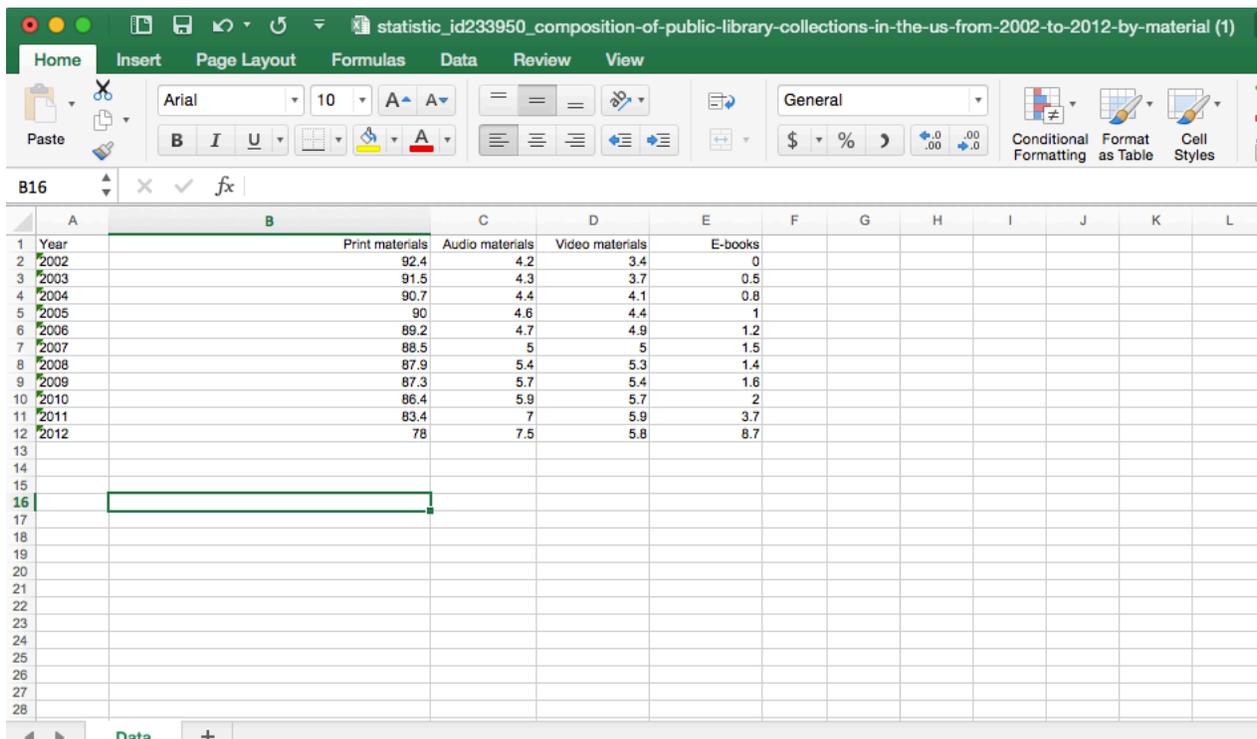
course of a decade. A note on the overview tab provides credits for the dataset and explains that the data appears as a percentage of the total collection materials per 1,000 people.

Cleaning Your Data

First you'll need to format the data so that it is easy to upload in Open Refine. In the Excel document, go to 'File' and click 'Save As' to create a new version of the document. Name the file and save it to the same folder as the original file.

In your new file, you'll first want to delete the 'Overview' tab. Control click over the tab and select 'Delete' then choose 'Delete' again. You should have only the 'Data' tab.

First add the column heading 'Year' to the year column. You'll now want to get rid of all information in the sheet except for the column headings and data. Make sure to delete the empty column A as well. Your spreadsheet should look like this:



Year	Print materials	Audio materials	Video materials	E-books
2002	92.4	4.2	3.4	0
2003	91.5	4.3	3.7	0.5
2004	90.7	4.4	4.1	0.8
2005	90	4.6	4.4	1
2006	89.2	4.7	4.9	1.2
2007	88.5	5	5	1.5
2008	87.9	5.4	5.3	1.4
2009	87.3	5.7	5.4	1.6
2010	86.4	5.9	5.7	2
2011	83.4	7	5.9	3.7
2012	78	7.5	5.8	8.7

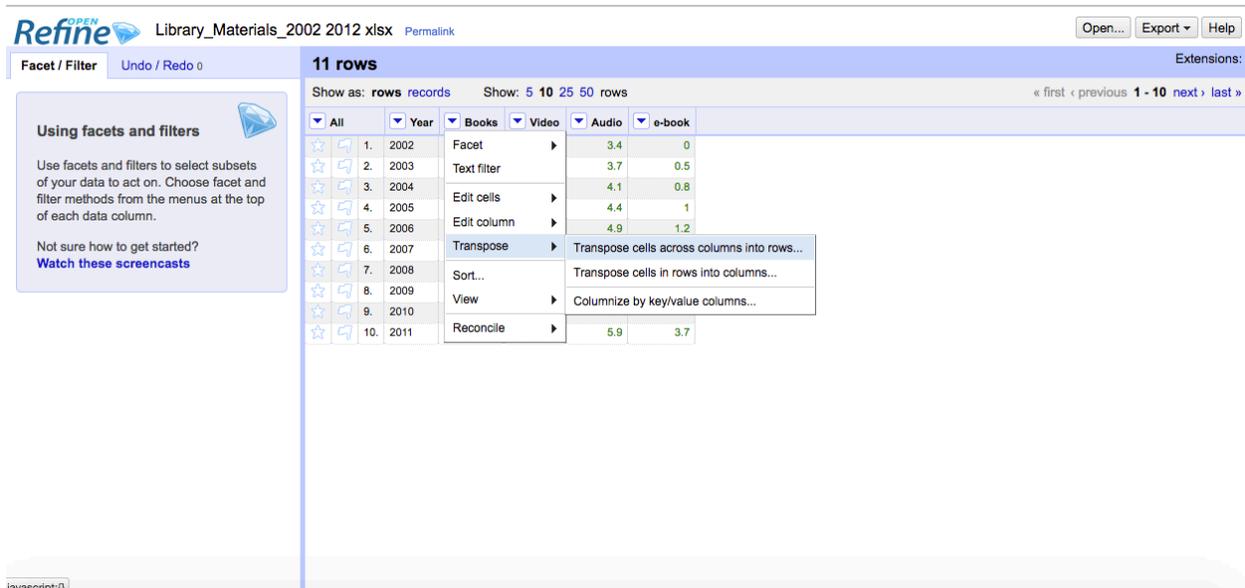
Transposing in Open Refine

As is, the data is spread across five columns, one for year and the other four for the types of materials used in libraries. RAW, however, prefers that data be sorted into three columns: one for the year, one for the group – in this case, the type of material – and one for the value associated with the group, in this case the percentage of each type in relation to the collection total.

Open Refine allows you to create this adjustment through its Transpose function, in just a few clicks. Transposing allows you to flip your data 90 degrees, turning cells into columns or columns into cells.

To see how this function works, launch the Open Refine application so that it pops open in your default web browser. Click the ‘Browse’ button and select your cleaned spreadsheet. Click ‘Next’. On the following page, click ‘Create Project.’

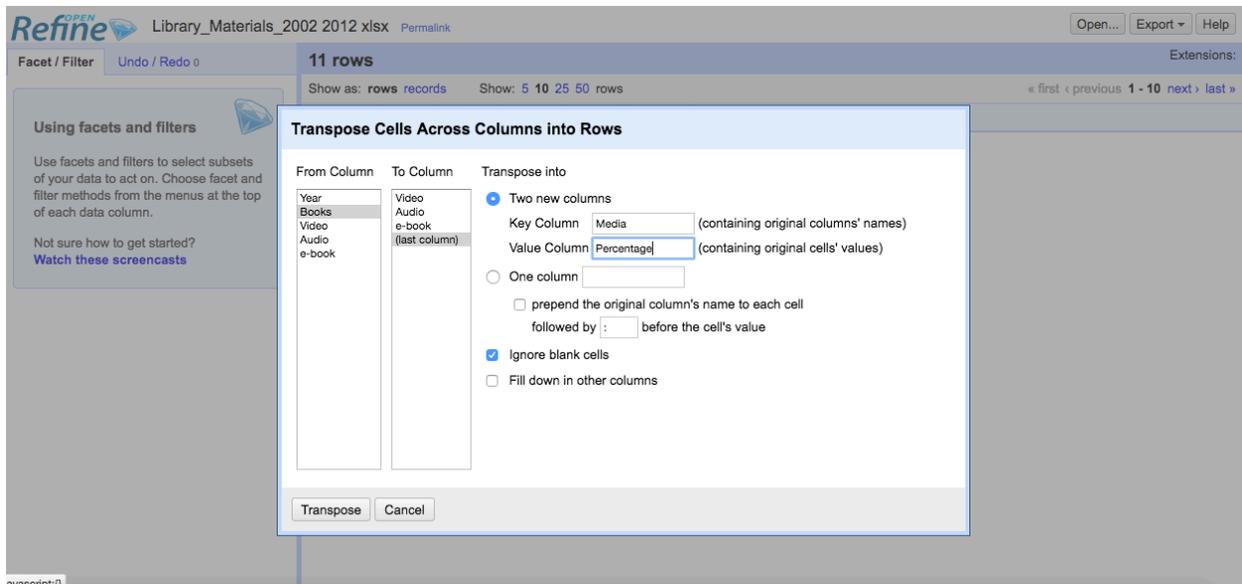
Now that you have your spreadsheet in Open Refine, click on the drop down arrow in the ‘Books’ column. Select ‘Transpose’ → ‘Transpose cells across columns into rows.’



The screenshot shows the Open Refine web interface. At the top, the title bar reads "Refine OPEN Library_Materials_2002 2012.xlsx Permalink" with buttons for "Open...", "Export", and "Help". Below the title bar, there are tabs for "Facet / Filter" and "Undo / Redo 0". The main area displays "11 rows" and "Show as: rows records" with a "Show: 5 10 25 50 rows" dropdown. A table of data is shown with columns for "All", "Year", "Books", "Video", "Audio", and "e-book". The "Books" column has a dropdown menu open, listing actions like "Facet", "Text filter", "Edit cells", "Edit column", "Transpose", "Sort...", "View", and "Reconcile". The "Transpose" option is highlighted, and its sub-menu is open, showing "Transpose cells across columns into rows..." as the selected option. A sidebar on the left contains a "Using facets and filters" section with instructions and a link to "Watch these screencasts".

	All	Year	Books	Video	Audio	e-book
1.	2002	Facet	3.4	0		
2.	2003	Text filter	3.7	0.5		
3.	2004	Edit cells	4.1	0.8		
4.	2005	Edit column	4.4	1		
5.	2006	Transpose	4.9	1.2		
6.	2007	Sort...				
7.	2008	View				
8.	2009	Reconcile	5.9	3.7		
9.	2010					
10.	2011					

In the next window, select that you want to create two new columns and give them each a name. In the example, the Key column is titled ‘Media’ and the ‘Value’ column is titled ‘Percentage.’ Click ‘Transpose’.



You now have three columns, not five, showing the same data but in a different format.

Before you export, go to the drop down arrow in the drop down arrow in the 'Year' column heading and click 'Edit cells' → 'Fill down'. Now all each cell is filled with data.

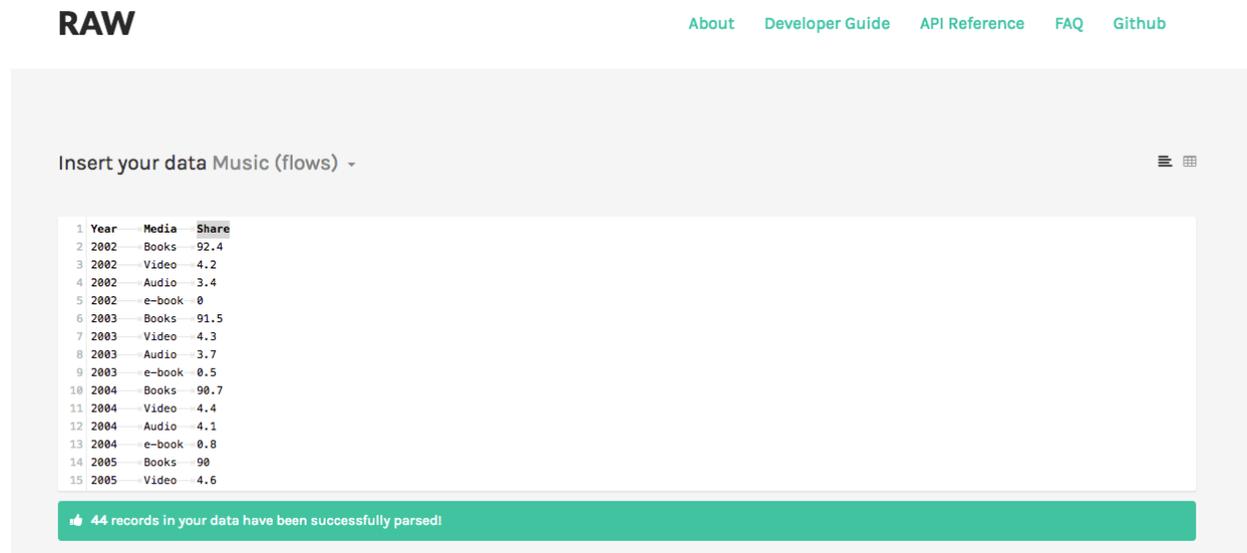
	Year	Media	Percentage
1.	2002	Books	92.4
2.	2002	Video	4.2
3.	2002	Audio	3.4
4.	2002	e-book	0
5.	2003	Books	91.5
6.	2003	Video	4.3
7.	2003	Audio	3.7
8.	2003	e-book	0.5
9.	2004	Books	90.7
10.	2004	Video	4.4
11.	2004	Audio	4.1
12.	2004	e-book	0.8
13.	2005	Books	90
14.	2005	Video	4.6
15.	2005	Audio	4.4
16.	2005	e-book	1
17.	2006	Books	89.2
18.	2006	Video	4.7
19.	2006	Audio	4.9
20.	2006	e-book	1.2
21.	2007	Books	88.5
22.	2007	Video	5

You're ready to export your new chart. In the right hand corner click 'Export' → 'Excel'. Save to your project folder.

Visualizing in RAW

Open the RAW website: <http://app.raw.densitydesign.org/>

To load your data into RAW, simply cut and paste the three columns of data from your new spreadsheet. You should see the three columns and 44 records total.



The screenshot shows the RAW interface. At the top left is the 'RAW' logo. To the right are navigation links: 'About', 'Developer Guide', 'API Reference', 'FAQ', and 'Github'. Below the navigation is a header area with the text 'Insert your data Music (flows) -' and a menu icon. The main content area displays a table with 15 rows of data. Below the table is a green notification bar that says '44 records in your data have been successfully parsed!'.

	Year	Media	Share
1	2002	Books	92.4
2	2002	Video	4.2
3	2002	Audio	3.4
4	2002	e-book	0
5	2003	Books	91.5
6	2003	Video	4.3
7	2003	Audio	3.7
8	2003	e-book	0.5
9	2004	Books	90.7
10	2004	Video	4.4
11	2004	Audio	4.1
12	2004	e-book	0.8
13	2005	Books	90
14	2005	Video	4.6
15			

Scroll down and select a chart. From here, you can explore the various chart options in RAW, seeing which create meaningful visualizations with your data. There are a few, for instance, that especially suit data showing changes over the course of time.

Try out the 'Bump Chart'. After selecting this option, scroll down to 'Map your dimensions.' You'll find that each of your column headings fits within the dimensions listed for this particular kind of chart: 'Media' in Group, 'Year' in Date, and 'Percentage' in Size.

Now scroll down and check out the visualization – it's that easy!

Play around and see if you can find other chart options that make sense with your three data columns.

Exporting from RAW

RAW provides you embed code to display your visualization on a website. You can also download image, vector graphic, and json files – just choose a file type, name it, click download, and you're set!

Credits:

Written by Morgan Currie, September 2016.