Working with messy data sets? Two useful and free tools

Description

Tweet

I have just come across two useful *apps* (aka *software packages* (aka *tools*)) for when you are working with someone elseâ??s data sets and/or data sets from multiple sources and times. Or, just your own data that was in a less than perfect state when you last left it :-)

- OpenRefine: Initially developed by Google and now open source with its own support and development community. You can explore the characteristics of a data set, clean it in quick and comprehensive moves, transform its layout and formats, as well as reconcile and match multiple data sets. There is documentation and videos to show you how to do all this. There is also a book, which you can purchase.The wikipedia entry provides a good overview.
- Tabula: This package allows you to extract tables of data from pdfs, a task which otherwise can be very tiresome, messy and error prone

And some other packages I have yet to explore

- Google Fusion Tables: A web service that provides means for visualizing data with pie charts, bar charts, lineplots, scatterplots, timelines, and geographical maps. Data is exported in a comma-separated values file format.
- And a review of 22 free tools for data visualization and analysis (including those above) that looks worth exploring

Category

1. Uncategorized

Date
16/11/2026
Date Created
25/04/2014
Author
admin