

In this lab you will be using Open Refine to clean a data set.

You will be using the eq2015 data set which data about earthquakes of magnitude 3 or more during the first 6 months of 2015.

data attribute glossary: <http://earthquake.usgs.gov/earthquakes/feed/v1.0/glossary.php#net>  
<http://davidhuynh.net/spaces/nicar2011/tutorial.pdf>  
<http://schoolofdata.org/handbook/recipes/cleaning-data-with-refine/>

Google refine cheat sheet <http://arcadiafalcone.net/GoogleRefineCheatSheets.pdf>

Perform the following

nst: fill in missing values.

extract country/state in to a new column called location.

extract the Date without time into a new column called eventdate

Run cluster en edit on location column. Run nearest neighbor and levenshtein distance.

- Does it make sense to merge detected values?
- Why or why not?

Try to do nearest neighbor clustering on "place" column.

- What happens?
- Why?

Tests med chicago crime