Data Cleaning with OpenRefine

Learning Objectives

- Describe OpenRefine's uses and applications.
- Differentiate data cleaning from data organization.
- Experiment with OpenRefine's user interface.
- Locate helpful resources to learn more about OpenRefine.

Exercise

- Using faceting, find out how many years are represented in the census.

- Is the column formatted as Number, Date, or Text? How does changing the format change the faceting display?

- Which years have the most and least observations?

Exercise

- Try to change the name of the second new column to "species". How can you correct the problem you encounter?

Exercise

- What scientific names (genus and species) are selected by this procedure?

- How would you restrict this to one of the species selected?

Exercise

You might like to look for trends in your data by month of collection across years.

- How do you sort your data by month?

- How would you do this differently if you were instead trying to see all of your entries in chronological order?