Hands-On OpenRefine Exercise

1. Download and install openrefine from http://openrefine.org/download.html
2. Download the sample data file to your computer
3. Create a project in Openrefine. Make sure the data is properly formatted
4. Check the data set and clean up any errors you notice. Check, e.g., for spelling errors, for inconsistencies, for non-numerical values in the count column. Delete the rows with non-numerical values in this column. [hint: the newest version of openrefine seems not to properly recognize numerical values here…]
5. You should now have a dataset with 10 states, 42 stations, around 163 species (depending on your knowledge about aphids that number may differ…), 20518 rows, 20019 of those containing observations of female and 499 containing observations of male animals.