

This assignment addresses some tasks that are often required to prepare datasets for further analysis. It follows up on class discussions and uses data from the repository at `ftp://sidads.colorado.edu/pub/DATASETS/AGDC/nsidc0591_fudge/`. The tasks outlined below may be performed using varying levels of automation. In general, you should favor greater automation but should ensure you complete the tasks regardless of method.

1. Download a copy of the `00README.txt` file from the repository to your own computer.
 - (a) Briefly describe the method you used to create this copy.
 - (b) Compute and report the MD5 checksum of the file.
 - (c) Comment on the significance, if any, of the `00` prefix in the file name.
2. Download copies of all files in the repository to your own computer. [Hint: There are methods that do not require repetitious tasks.]
 - (a) Briefly describe the method you used to create these copies.
 - (b) Compute the MD5 checksum of each file and report the results in a suitable tabular format.
3. Consider the dataset `e3331.e50`.
 - (a) Compute and report the mean, minimum, maximum, and standard deviation of the conductance values in the dataset.
 - (b) Briefly describe the method you used for the above.
 - (c) Repeat Qs. 3a and 3b for every column of every dataset in the repository, reporting the results in an appropriate tabular format.
4. Generate a derived dataset that combines data from all the datasets in the repository. The resulting dataset should have a single *depth* column and as many additional columns as needed to capture all the data in the repository.
 - (a) Describe the structure and semantics of your derived dataset, highlighting its relation to the underlying datasets.
 - (b) Compute the mean, minimum, maximum, and standard deviation for each column of the derived dataset. Report your results in a suitable tabular format.
 - (c) Briefly describe the method you used for the above tasks.
 - (d) Outline how much additional or repeated work will be needed to regenerate the derived dataset when one or more underlying datasets change.
5. Create a *gzipped tar* file that packages the following, and submit the file using the procedure outlined in the syllabus.
 - (a) A PDF file with your typed answers to the questions above.
 - (b) The derived dataset of Q. 4.
 - (c) A README file with appropriate contents.