**GCD Assignments week 3**

The assignments of week 3 deal with cleaning a dataset with Microsoft Excel or Libre Office and then translating the operations to Python.

**Preparation**

No preparation needed before this lesson.

**Activities**

*Activity 1 [2.5 hours max]*

Follow tutorial <http://schoolofdata.org/handbook/recipes/cleaning-data-with-spreadsheets/> (Solve Problem 1-6).

Problem 1 can be omitted, since this exercise deals with layout only.

*Activity 2*

Open iPython Notebook:

In Terminal/Command window type: ipython notebook

**Tip**: To open a Terminal/Command window at a certain Folder

Mac OS X

* In System Preferences > Keyboard > Shortcuts > Services:

Enable New Terminal at Folder

* In Finder: Right-mouse click on the preferred folder

Windows 7

* In Documents: Hold Shift and right-click the preferred folder
* Select ‘Open command window here’

Rewrite the statements used in the Excel Data Cleaning tutorial to Python statements.

The Excel operations used in the tutorial:

1. Autofilter – quickly find out what’s in a column of data, and show which bits you want
2. Data Validation – choose what values are allowed in a cell
3. Defined data ranges – create lists of data that can be used to make comparisons
4. Find and Replace – look for one thing, and replace it with another
5. Paste Special – remove unwanted formatting
6. Pivot tables – summarise your data and see it in completely new ways
7. Regular Expressions – clever way of searching for data
8. Text-to-Columns – split up cells of data that have more than one entry in them
9. Formulae that use some different spreadsheet functions, including CLEAN, CONCATENATE, COUNTA, IFBLANK, LEFT, SEARCH, SUBSTITUTE, and TRIM.