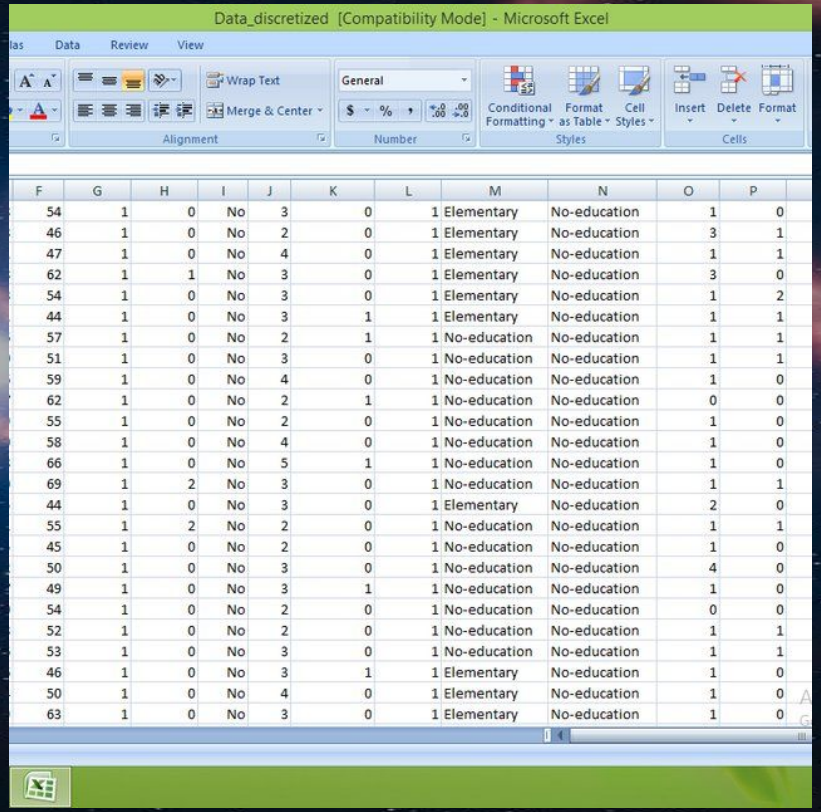
Lecture 6

Working with Data

Workshop 2 Answer Review

Importing External Data

- Most of the data you work with will not be created in python
- In these cases the data will need to be imported into python
- Python does this very easily (another reason it's used so much)

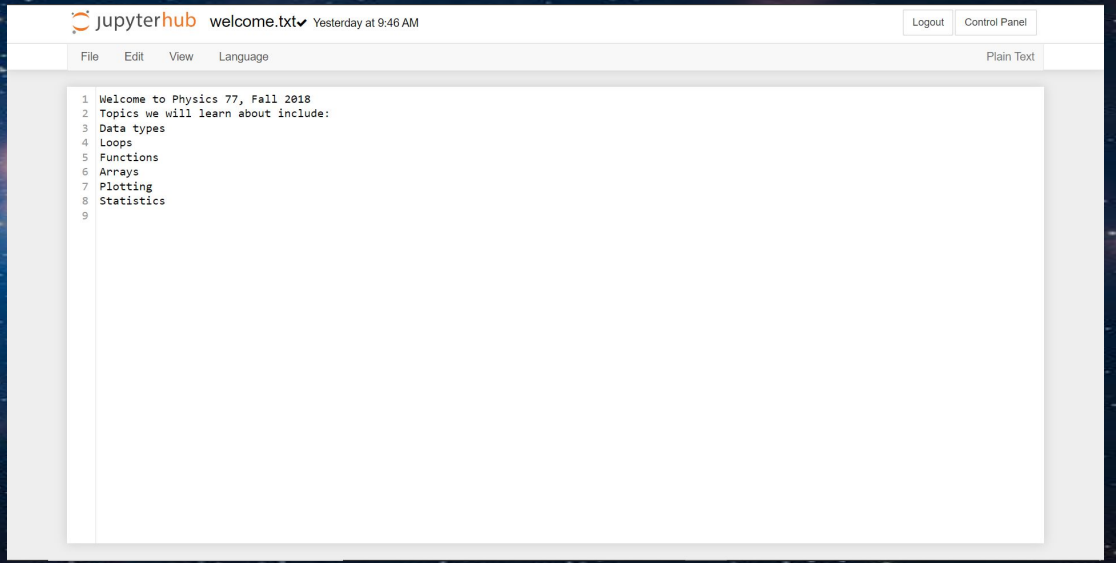


F	G	H	I	J	K	L	M	N	O	P
54	1	0	No	3	0	1	Elementary	No-education	1	0
46	1	0	No	2	0	1	Elementary	No-education	3	1
47	1	0	No	4	0	1	Elementary	No-education	1	1
62	1	1	No	3	0	1	Elementary	No-education	3	0
54	1	0	No	3	0	1	Elementary	No-education	1	2
44	1	0	No	3	1	1	Elementary	No-education	1	1
57	1	0	No	2	1	1	No-education	No-education	1	1
51	1	0	No	3	0	1	No-education	No-education	1	1
59	1	0	No	4	0	1	No-education	No-education	1	0
62	1	0	No	2	1	1	No-education	No-education	0	0
55	1	0	No	2	0	1	No-education	No-education	1	0
58	1	0	No	4	0	1	No-education	No-education	1	0
66	1	0	No	5	1	1	No-education	No-education	1	0
69	1	2	No	3	0	1	No-education	No-education	1	1
44	1	0	No	3	0	1	Elementary	No-education	2	0
55	1	2	No	2	0	1	No-education	No-education	1	1
45	1	0	No	2	0	1	No-education	No-education	1	0
50	1	0	No	3	0	1	No-education	No-education	4	0
49	1	0	No	3	1	1	No-education	No-education	1	0
54	1	0	No	2	0	1	No-education	No-education	0	0
52	1	0	No	2	0	1	No-education	No-education	1	1
53	1	0	No	3	0	1	No-education	No-education	1	1
46	1	0	No	3	1	1	Elementary	No-education	1	0
50	1	0	No	4	0	1	Elementary	No-education	1	0
63	1	0	No	3	0	1	Elementary	No-education	1	0

Review: Uploading files to Datahub

What can we do with external file types?

- First let's look at txt files
- These files are easy to manipulate both within and outside of python
- Really useful for storing information you want to retrieve easily



Numpy: np.loadtxt

Some of the most useful optional arguments are:

dtype: data type of the resulting array

comments: the character that indicates the start of a comment (e.g. '#')

lines following these characters will be ignored, and not read into the array

delimiter: the character used to separate values. Often, it's whitespace,

but it could also be ',', '|', or others

skip_header: how many lines to skip at the beginning of the file

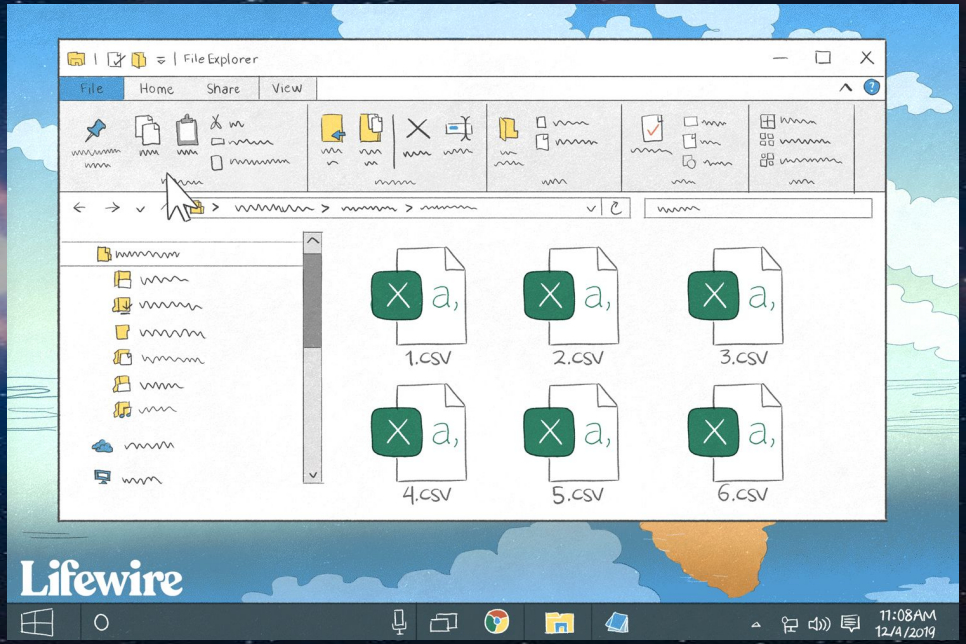
skip_footer: how many lines to skip at the end of the file

use_cols: which columns to load and which to ignore

unpack: If True (the default is False), the array is transposed

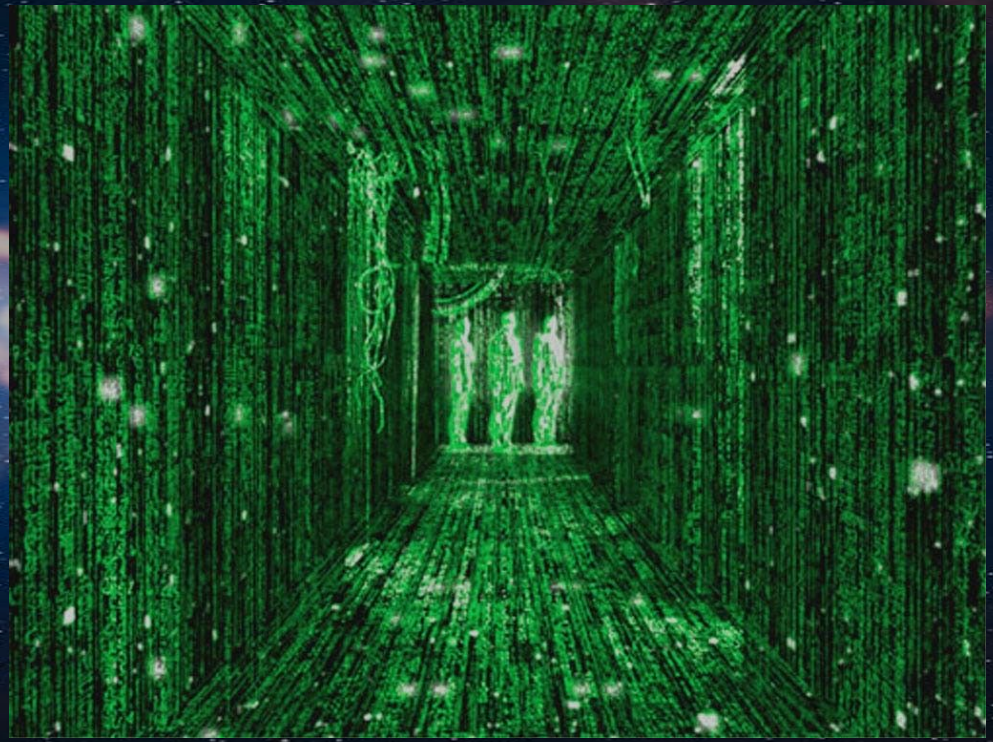
CSV Files

- Similar to txt files
  - Can be written both within and outside of python
  - Used to store data and information
- More complex than txt files
- Can create data for csv files in excel and move them to python



Binary Files

- Unlike txt and csv files that are written in such a way we can understand
- Binary files are written in such a way that python can read the file’s contents, but we cannot access them directly
- Commonly Used in python as an .npy file



Coding/CSV file creation Demo