Reading and writing CSV files

Presenter: Steve Baskauf
steve.baskauf@vanderbilt.edu

Jean & Alexander Heard
LIBRARIES

CodeGraf landing page

- vanderbi.lt/codegraf

Format of CSV files

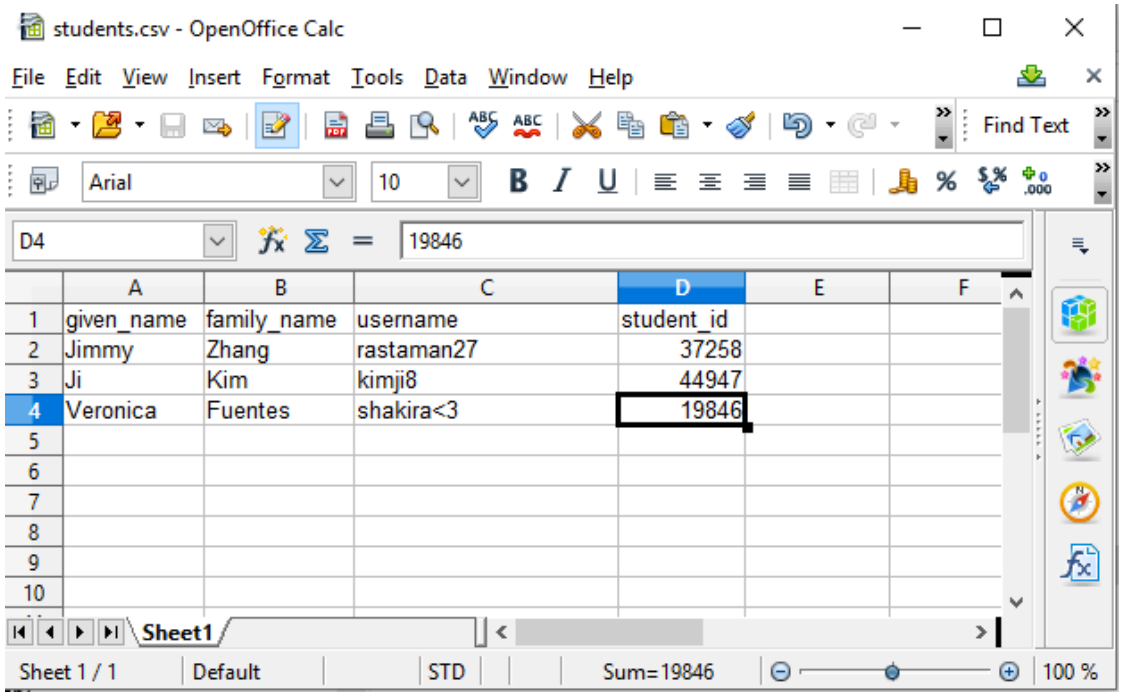
Jean & Alexander Heard
LIBRARIES

Example of a CSV file

raw text

```
given_name,family_name,username,student_id
Jimmy,Zhang,rastaman27,37258
Ji,Kim,kimji8,44947
Veronica,Fuentes,shakira<3,19846
```

rendered as a table by a spreadsheet program



	A	B	C	D	E	F
1	given_name	family_name	username	student_id		
2	Jimmy	Zhang	rastaman27	37258		
3	Ji	Kim	kimji8	44947		
4	Veronica	Fuentes	shakira<3	19846		
5						
6						
7						
8						
9						
10						

CSV details and cautions

- Saving a CSV for the first time is critical for determining the delimiter and character encoding
- Delimiters are sometimes tabs (TSV) or pipes (|)
- The CSV file scheme has special tricks for handling strings that contain the field delimiter (comma) or the text delimiter (double quotes).
- Best not to manage parsing or writing CSVs the hard way – use an editor or Python library functions.
- Excel will always read in CSV text like "1-26" as dates like "January 26". There is no way to turn this off! **So don't use Excel with CSV files!!!**
- Libre Office is probably the best editor for working with CSV files.

Reading and writing CSV file data

Jean & Alexander Heard
LIBRARIES

The Python **`csv`** module

- **`csv.reader()`** and **`csv.DictReader()`** objects are iterable objects created by reading from CSV files.
- Each line (ended by a newline character) is turned into one of the iterable items in the object.
- Each **reader** iterated item is a list with items corresponding to the columns of that row.
- Each **DictReader** iterated item is a dictionary with a key for the column header and a value from the column of that row.
- To reuse the row items, append them to a list for later reference.

Writing data to CSV files

- To use the **`csv.writer()`** object, each line must be in the form of a list.
- The lists are written to the writer object one at a time.
- When the writer object is closed, the file is written as a CSV.

- Using the **`csv.DictWriter()`** object is similar, with each row in the form of a dictionary.
- The header row must be specified and written in a separate operation from the rows.

Reading CSV data from a URL

- CSV files acquired through a URL using the **`requests`** module can't be iterated directly.
- However, the response text can be turned into an iterable list using the **`.splitlines()`** method.
- The iterable list can be passed into one of the reader functions and used to generate lists of lists or dictionaries.