Introduction to Data Analysis
Session 3, Nov. 15
Today's Overview

• Morning lecture
  - New tools
  - Data visualization
  - Harry Potter on Wikipedia
  - Working on data sets in Google docs

• Lunch

• Afternoon sessions
  - Continuation of this morning's lecture
  - matplotlib session
  - Self-directed

• Closing remarks
while loops

```python
>>> while True:
    ...
    ...   ((do something))
    ...
    ...   if ((condition)):
    ...       break
```
string.join

>>> list = ['a', 'b', 'c']
>>> list
['a', 'b', 'c']

>>> ','.join(list)
'a,b,c'
File Reading Review

```python
>>> file = open('myfile', 'r')

• 'r' stands for “read” mode

>>> line = file.readline()
>>> line
Some text from the file
```
Read the file in other ways

```python
>>> for line in file:
...     ((do something))
```
csv Module

- It's common to store data in “comma-separated value” format (csv)
- Data looks like
- We want to be able to process this to aggregate and do analysis
- Built-in csv module allows us to do so
>>> # Read in file as line list
>>> import csv
>>> with open('hp_wiki.csv', 'r') as csvfile:
...     reader = csv.reader(csvfile,
...               delimiter=',' )
...     for row in reader:
...         ((do something))
>>> # Read in file as list of dictionaries
>>> import csv
>>> reader = csv.DictReader(open('file', 'r'))
>>> for line in reader:
...   ((do something))
Live Data Processing Demos!