

# Computing for Medicine: Phase 3, Seminar 1 Project

Jennifer Campbell

Associate Professor, Teaching Stream

[campbell@cs.toronto.edu](mailto:campbell@cs.toronto.edu)



# Seminar 1 Project

- The project handout, starter code, and data are posted:
  - <http://c4m.cdf.toronto.edu/cohort1/phase3/>
- Two options for doing your work:
  - Use the Computer Science Teaching Labs computing network from your computer (internet connection required).
  - Use only your computer.

**WORKING ON CS TEACHING  
LABS NETWORK**

# CS Teaching Labs Account

- Information you will need:
  - **Username:** your UTORid
  - **Password:** See email “C4M: Phase 3 account information” for your temporary password.
  - **Key:** See email “C4M: Phase 3 computing labs key” for the `cdf_nx_key.key` file

# Connecting to CS Teaching Labs

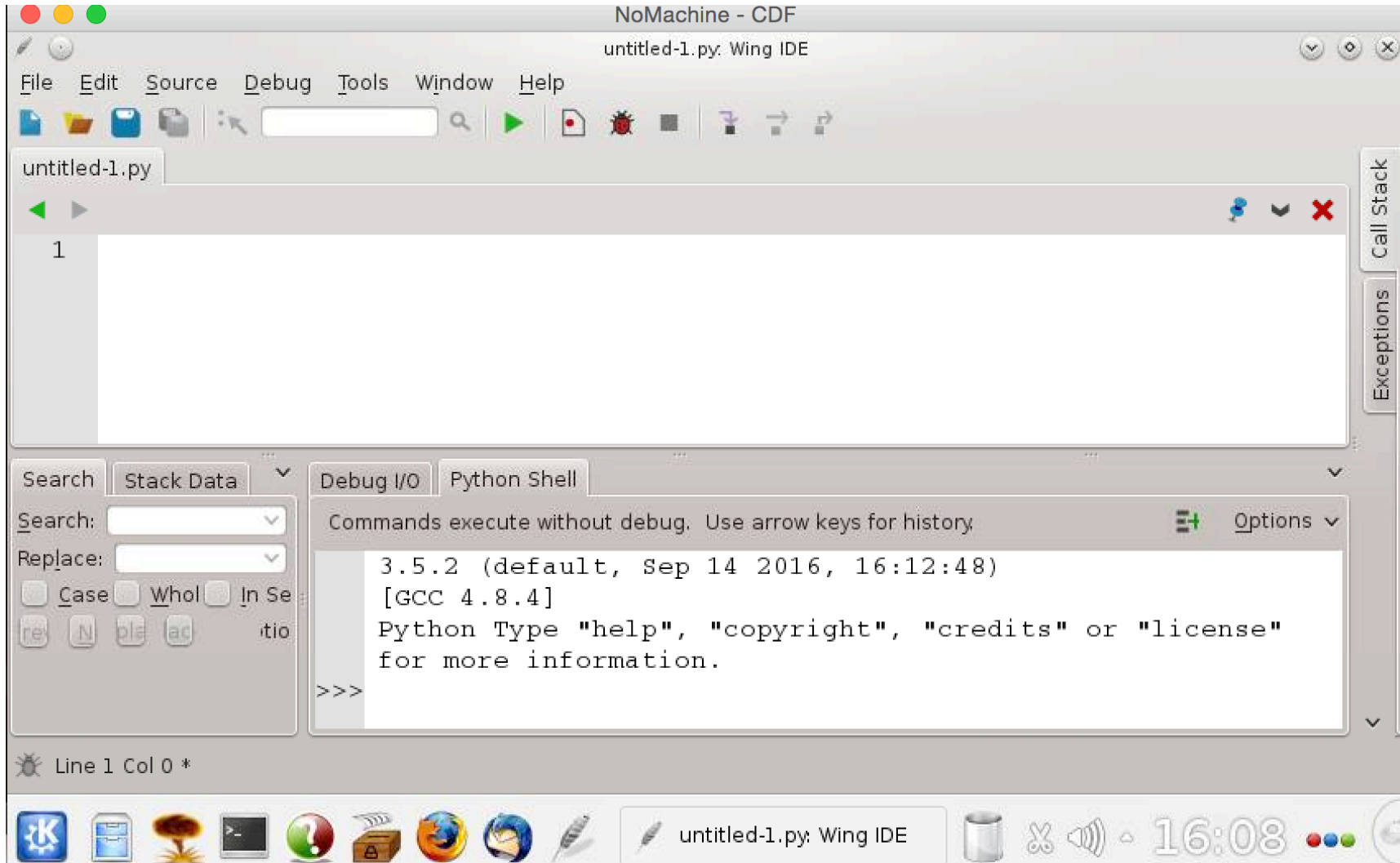
- To connect to the CS Teaching Labs from your computer, please follow the instructions here:
  - <http://computing.help.inf.ed.ac.uk/nomachine>
  - Name: CS Teaching Labs
  - Host: nxsrv1.cdf.toronto.edu
  - Protocol: SSH
  - Port: 22
  - Key: use the file sent to you by email

# Changing your password

- Once you've logged in to the CS Teaching Labs, please change your password.
- At the bottom of the screen, to change your password use:
  - Application Launch Menu -> First Year CDF -> CDF Reset Password



# Wing 101 IDE (instead of Pyzo)



**WORKING ON YOUR  
OWN LAPTOP**



# Software installation

- To work on your own laptop, you will need to install the following software:
  - <http://www.nltk.org/install.html>
  - <http://scikit-learn.org/stable/install.html>

For Windows users, follow the instructions on the websites above.

For Mac OS X users, open a Terminal window and type these commands (you'll need Python 3.4 or higher installed first):

```
sudo pip3 install -U numpy
sudo pip3 install -U nltk
sudo pip3 install -U scipy
sudo pip3 install -U scikit-learn
```

# Download data (and tagger)

- Download the ZIP file that contains the project starter code and data from the C4M website:  
<http://c4m.cdf.toronto.edu/cohort1/phase3/>)
- (Optional – not needed unless you want to run the preprocessing code yourself)
  - Download the [Stanford Log-linear Part-Of-Speech Tagger](http://nlp.stanford.edu/software/stanford-postagger-2015-12-09.zip) from <http://nlp.stanford.edu/software/stanford-postagger-2015-12-09.zip> and save it in the same directory as the starter code and data.
  - Unzip the file stanford-postagger-2015-12-09.zip

# MODULES & PACKAGES

# Modules and Packages used in project

- `nltk` (Natural Language Toolkit)
  - <http://www.nltk.org/api/nltk.html>
- `csv` (Comma Separated Values files)
  - <https://docs.python.org/3/library/csv.html>
- `math` (Mathematical functions)
  - <https://docs.python.org/3/library/math.html>
- `numpy` (Numerical Python)
  - <https://docs.scipy.org/doc/numpy/reference/routines.html>

Click on the links above for documentation for each module/package. You can also use `dir()` and `help()`.

# NumPy

- A scientific computing package for Python.
- For this project, you'll use NumPy's N-dimensional array.
  - NumPy's 2D array vs Python's nested lists
    - NumPy's array may contain only elements of the same type, whereas Python's lists may contain different types.
    - NumPy's arrays are more efficient and take less space.
    - NumPy supports a variety of array operations.

# NumPy 2D array demo

```
>>> import numpy as np
>>> my_array = np.array([[1,
2, 3], [4, 5, 6]])
>>> my_array.shape
(2, 3)
>>> my_array.size
6
>>> my_array.sum()
21
>>> my_array.min()
1
>>> my_array.max()
6
>>> my_array.mean()
3.5
>>> my_array.var()
2.9166666666666665
>>> my_array.std()
1.707825127659933
```

# UPCOMING SEMINARS

# Seminar 2: Dr. Fanny Chevalier

- Tuesday November 1 2016 6-8pm
- Location: DCS Innovation Lab
- Topic: Data Visualization
- <http://fannychevalier.net/>

Please RSVP by completing this Google form:

[https://docs.google.com/forms/d/e/1FAIpQLSeh3L9kLd\\_8H8LnSF14VVg-ZoVNPoBUqKR99YgK6LJSa2ywxg/viewform](https://docs.google.com/forms/d/e/1FAIpQLSeh3L9kLd_8H8LnSF14VVg-ZoVNPoBUqKR99YgK6LJSa2ywxg/viewform)



# SEMINAR FEEDBACK

# Phase 3, Seminar 1 Survey

- You should have received an email with the subject “C4M: Phase 3, Seminar 1 Feedback Survey”.
- Please complete that survey now:
  - <https://www.surveymonkey.com/r/C4MSeminar1>