Before the tutorial:

- Before the in-person tutorial, we encourage you to go through the following Python basics online tutorial: [https://github.com/haoopeng/INFO-1590-On-ramp/blob/master/python_tutorial.ipynb](https://github.com/haoopeng/INFO-1590-On-ramp/blob/master/python_tutorial.ipynb)

- Please make an account on Microsoft Azure so that you can use Azure Jupyter Notebook (especially if you plan to use the SSRC’s computers): [https://notebooks.azure.com/](https://notebooks.azure.com/)

- For debugging, if you are bringing your own laptop, please install the following software:
  1. Anaconda 4.0.0: [https://www.continuum.io/downloads](https://www.continuum.io/downloads)
  2. PyCharm IDE (the Community version is free): [https://www.jetbrains.com/pycharm/download/](https://www.jetbrains.com/pycharm/download/)
  3. Scikit-learn: [http://scikit-learn.org/stable/install.html](http://scikit-learn.org/stable/install.html) (note: the same package is also available on Azure Jupyter, so you need not install it if you are using online Python notebook)

- Please save the following materials to your own Dropbox or OneDrive account so that you can use them directly in Azure Jupyter Notebook (if you do not have either a Dropbox or OneDrive account, please sign up for one):
  1. [https://www.dropbox.com/sh/q2kdf4f7chygvo0c/AAACu3EaV5DlyMDTQCwrYhDVEa?dl=0](https://www.dropbox.com/sh/q2kdf4f7chygvo0c/AAACu3EaV5DlyMDTQCwrYhDVEa?dl=0)
  2. [https://www.dropbox.com/sh/8yewenvj3v6sp1/AAAAIOv8AuFyK7C1Jt0mTYk7Ya?dl=0](https://www.dropbox.com/sh/8yewenvj3v6sp1/AAAAIOv8AuFyK7C1Jt0mTYk7Ya?dl=0)
**Tutorial schedule:**

1. **Basics of Python** (20 minutes)
   - We encourage you to go through the following Python basics before the tutorial: [https://github.com/haoopeng/INFO-I590-On-ramp/blob/master/python_tutorial.ipynb](https://github.com/haoopeng/INFO-I590-On-ramp/blob/master/python_tutorial.ipynb)
   - Analysis of the Donald Trump NYC speech on stakes of the election using Regular expression, word cloud, etc.

2. **Introduction to debugging using PyCharm IDE** (10 minutes)

3. **Data Analysis and Visualization: Introduction to NumPY and Pandas** (40 minutes)
   - Data analysis on sampled data from Microsoft Academic Graph Project using Pandas, Numpy, Matplotlib.
   - A simple regression model with significance test using Statsmodels.
   - Visualization of high dimensional data using t-SNE package.

4. **Machine Learning using ScikitLearn** (50 minutes)
   - Introduction to Machine Learning
   - Various tasks which Machine Learning Algorithm can perform
   - Curve fitting and Error functions
   - Logistic Regression Using Scikit-Learn
   - Classification of messages into spam and non-spam using NaiveBayes