# Introduction to Text Analysis

## Workshop Preliminaries

### Bryce J. Dietrich

University of Iowa

For the workshop, please do the following:

1. Install R Studio. To do so, go to this website: https://www.rstudio.com

2. Install miniconda. Miniconda helps us set up a virtual environment. That way, we are all working with the same Python installation. To install minconda, go to this website: http://conda.pydata.org/miniconda.html BE SURE TO select a Python 3.5 installer and 64-bit (unless your machine is super old, then shoot me an email).

3. After you have done this, install Sublime Text 3. Sublime is a universal text editor that works on the Windows and Mac operating systems. I personally use Atom, but it is only works on the Mac. To install Sublime, go to this website: http://www.sublimetext.com/3 BE SURE TO select the 64-bit installation (unless your machine is super old, then shoot me an email).

4. Hopefully, you went to the Python workshop. If not, you can learn the basics of Python using DataCamp. This is a free online tutorial. You can find the tutorial here: https://www.datacamp.com/courses/intro-to-python-for-data-science

For the workshop, please read following:

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3): 267–297.

Once you are done, watch the following video:

https://youtu.be/GVTa-NhR4jk

In the video, you will find me working (struggling?) my way through the New York Times Articles API. To use the API, you need to register for an API key here. The code I produced in the video and the data set can be found here:

http://www.brycejdietrich.com/files/nyt_demo.zip

Don't worry – we are going to walk through the script during the workshop.

In the meantime, try creating a Python virtual environment using miniconda. According to the eresearch cookbook (https://uoa-eresearch.github.io/eresearch-cookbook/recipe/2014/11/20/conda/):

> A virtual environment is a named, isolated, working copy of Python that that maintains its own files, directories, and paths so that you can work with specific versions of libraries or Python itself without affecting other Python projects. Virtual environments make it easy to cleanly separate different projects and avoid problems with different dependencies and version requirements across components. The conda command is the preferred interface for managing installations and virtual environments with the Anaconda Python distribution.

To create a virtual environment, you need to open your terminal. On Mac machines, this is called the "Terminal." On Windows machines, this is called the "Command Prompt." Once you have it open, type the following:

```
1 #create python envrionment
2 conda create −−name python−UI numpy scipy scikit−learn
      matplotlib python=3
3
4 #verify it has been created
5 conda info −−envs
6
7 #activate virtual environment
8 source activate python−UI
```

2

```
 9
10  #deactivate  virtual  environment
11  source  deactivate  python−UI
```

All we are doing here is making sure we are all working with the same Python installation. If you took the Python workshop, feel free to use whatever you are most comfortable using. The most important thing is you are using Python 3 NOT Python 2.

Finally, you can find all the sample code and data here:

http://www.brycejdietrich.com/files/workshop_code.zip