Luis Amaral amaral@northwestern.edu

Adam Pah [a-pah@kellogg.northwestern.edu](mailto:a-pah@kellogg.northwestern.edu)

**Description**

Our digital, connected, sensor rich world is generating extraordinary amounts of data (“Big Data”) that are being used to purposes as diverse as teaching a computer to win at Jeopardy or offering taxi alternatives. The skills needed to go from data to knowledge and application, which go under the name of Data Science, are in big demand in industry, government, and academia. This course provides an introduction to the foundational skills needed by data scientists.

**Software**

In this course we will be using the Anaconda Python 3.6 distribution ([download here](https://www.continuum.io/downloads)), **you must have the software installed on your laptop before the first class starts.** We have videos detailing the installation here --- [OS X](https://www.youtube.com/watch?v=UQhOyZXHkxI) and [Windows](https://www.youtube.com/watch?v=w16iUU6IA5E) (ignore any mentions of ‘bootcamp’ and make sure you download **version 3.6**).

**Course Materials**

We have created the materials for this course, there are no additional texts to purchase.

You can download the materials [here](http://bit.ly/nico101) and watch a video detailing how to use the materials [here](https://www.youtube.com/watch?v=Pul4cqoPLbE).

**Important Note**

Since this class is not on the regular timing for fall quarter, it **does not** follow the administrative timing of other courses. If you do no intend to finish the class for some reason, you must **withdraw on Caesar by Friday, September 7th.**

**Grading**

*Assignments* (90%) A series of daily individual assignments will give you experience applying the tools from class. Each assignment is topically related to the content from that day in class and will be made available at the end of each day on [Vocareum](http://vocareum.com/) (we will detail how to use Vocareum in-class on the first day).

There are 8 assignments (one for each day) and each assignment is worth 11.25% of your overall grade. Each of these assignments must be completed individually.

Assignments 1-4 will be due on Monday, September 10th by 9am.

Assignments 5-8 will be due on Tuesday, September 18th by 9am.

*Participation* (10%) We are utilizing a Canvas app called YellowDig that allows for discussion amongst classmates. The app is structured similarly to the ‘Facebook News Feed’, except you receive points for each post and comment that you make. Participation on Yellowdig (both posting and answering questions) will comprise your participation grade.

**Honor Code**

It is expected that code submitted for all assignments will be original and independently written. The grading platform has plagiarism checks built-in to detect between student copying of code and usage of code from on-line resources (i.e. stackoverflow), which will be used.

Students are allowed, **and encouraged**, to help each other understand programming concepts, errors, and how to approach problems.

Summary of Assignments

*All assignments are due by 9am on the day indicated.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Title | Distributed | Due | Percentage |
| 1 | *Individual* | *Basic data types* | Sept. 3 | Sept. 12 | 7.5% |
| 2 | *Individual* | *Collection data types* | Sept. 4 | Sept. 12 | 7.5% |
| 3 | *Individual* | *File IO* | Sept. 5 | Sept. 12 | 7.5% |
| 4 | *Individual* | *Dictionaries* | Sept. 6 | Sept. 12 | 7.5% |
| 5 | *Individual* | *Text analysis and Regular expressions* | Sept. 9 | Sept. 19 | 7.5% |
| 6 | *Individual* | *Web scraping* | Sept. 10 | Sept. 19 | 7.5% |
| 7 | *Individual* | *Structured analysis* | Sept. 11 | Sept. 19 | 7.5% |
| 8 | *Individual* | *Image analysis* | Sept. 12 | Sept. 19 | 7.5% |

Warning: This schedule is subject to change

Week 1 — Python Basics

*Day 1. Basics*

Morning Session:

* Course Overview
* Jupyter Notebook Introduction
* Basic Data Types

Afternoon Session:

* Flow Control
* Errors

*Day 2. Collections and Files*

Morning Session:

* Lists, Tuples, and Sets
* File I/O

Afternoon Session:

* Section Review

*Day 3. Imports, Plots, and Functions*

Morning Session:

* Python Standard Library
* Data Visualization

Afternoon Session:

* Functions
* Review

*Day 4. Dictionaries and Skill Review*

Morning Session:

* Dictionaries
* Review

Afternoon Session:

* Mini-Project: Understanding the impact of college major on employment

Week 2 — Data Analysis

*Day 5. Textual Analysis*

Morning Session:

* Text Analysis

Afternoon Session:

* Regular Expressions
* Sentiment Analysis

*Day 6. Working with the Web*

Morning Session:

* Introduction to APIs
* Reading and Posting with APIs

Afternoon Session:

* Web Scraping

*Day 7. Statistics and Structured Data*

Morning Session:

* Statistical Analysis with Python
* Bootstrapping Monte Carlo Chains
* Model Fitting

Afternoon Session:

* Structured data analysis Pt. 1
* Structured data analysis Pt. 2

*Day 8. Image Analysis*

Morning Session:

* Image Manipulation
* Image Analysis

Afternoon Session:

* Mini-Project: Identifying Cells