**Description**

The digital, connected, sensor rich world is generating extraordinary amounts and variety of data (“Big Data”). CSS is an exciting new scientific perspective that incorporates new methods and models for studying human behavior from the level of neurons to collective behavior. This change in approach has already made breakthroughs possible in understanding human creativity, scientific performance, the sharing economy, human conflict, and consumer behavior.

This seminar will teach computational analysis skills. These skills include null model design and programming, and data mining for structured and unstructured data (topic models, bag of words, etc.). Students will leave the course with the technologies and intuitions needed for sophisticated independent research.

**Prerequisites**

Students must possess an understanding of how to program in Python before the course begins. Students **must have completed** **one of two options** before the course starts.

1. Have taken NICO-101 Introduction to programming for big data (P/NP or A/B/C are both allowed).
2. Register for [Datacamp](https://www.datacamp.com/) on-line and pass a set of courses at your own pace. Kellogg has negotiated access:  
   i. Go to [www.datacamp.com](http://www.datacamp.com) and create a fee account  
   ii. Go the NUIT Datacamp Access form. <https://app.smartsheet.com/b/form?EQBCT=b0e9f1016ecc4c47856f63dd1f52f62a>  
   iii. When filling out the form you will see a “Prerequisite” section that says you must first have taken one of two free introductory courses. Kellogg students may ignore that requirement and simply write “I need this for a Kellogg course” in the box labelled “Request Exception.”  
     
   This option is intended for those that learn best on their own or already know the basics of programming (in Python or another language). The following courses must be passed before the start of CSSMA:
   1. Intro to Python for Data Science
   2. Intermediate Python for Data Science
   3. Python Data Science Toolbox (Part 1)
   4. Python Data Science Toolbox (Part 2)
   5. pandas Foundations
   6. Manipulating DataFrames with pandas
   7. Importing Data in Python (Part 1)
   8. Importing Data in Python (Part 2)

**Course Materials**

The predominant course materials that we will use are ones that I have created and are freely available. The other course materials are primary academic literature.

**Honor Code**

It is expected that code submitted for all assignments and projects will be original and independently written. The copying of code directly from online resources (e.g. stackoverflow) is explicitly prohibited. However, **students are allowed, and encouraged**, to help each other understand programming concepts, errors, and how to approach problems.

**Grading**

Since this is a graduate-level course, the primary evaluation is based on an independent research project that leverages course methods and concepts. There are also three homework assignments at the beginning of class to refresh core programming abilities and one presentation on a primary academic article during the quarter.

|  |  |
| --- | --- |
| ***Assignment*** | ***Percentage*** |
| *Homework 1* | 10 |
| *Homework 2* | 10 |
| *Homework 4* | 10 |
| *Paper Presentation* | 10 |
| *Project Proposal* | 10 |
| *Project Write-up* | 50 |