CS125 Overview

From MSCS

Contents

What is CS125?

CS125: Computer Science for Scientists and Mathematicians is a course offered at St. Olaf College offering a statistics-infused introduction to computer science. This course was developed as part of a successful collaboration between the Statistics and Computer Science Programs to meet the demand for a growing population of "data-centric" students who need to understand and use computational tools for data handling in a variety of fields [1]. This course was first taught in spring 2013, and has developed since its first offering to emphasize the interdisciplinary fields of sustainability and computing.

> **Course Description:** `This course focuses on handling data: visualization, finding patterns, and communicating with data. The primary tools are Python (for transforming data) and R (for visualization and classification). Students work individually and in teams to apply basic principles and explore real-world datasets with a sustainability theme. Prerequisite: calculus or consent of the instructor. Counts toward statistics and mathematical biology concentrations.` - *St. Olaf Course catalog*

It is a recommended elective for the statistics concentration at St. Olaf, while also serving as a prerequisite for courses in the computer science major as well as a way for students to satisfy their general education requirement in abstract and quantitative reasoning.

Summary of Course Content

Link to complete list of homeworks (http://www.cs.stolaf.edu/wiki/index.php/CS125_S15_homework)

This course teaches basic programming skills in two languages: Python and R. The course begins by introducing basic Python, then shifts R, and returns to introduce more advanced Python. The structure of this course was specifically designed to promote student confidence and familiarity with both languages, allowing students to assess their advantages and disadvantages for different applications in data handling, mining, and visualization.

A typical class day involves half the period spent answering questions from previous material and introducing new concepts. The remainder of the period is often spent on environment diagrams to promote student understanding of the underlying structure and evaluation processes of their programs.

- Python data structures and control features are introduced, followed by more complex topics in Python including web scraping using BeautifulSoup. Objects are only briefly discussed. The textbook for this section of the course is taught using Charles Severance's *Python for Informatics: Exploring Information* [2]. Emphasis is placed on using Python to reformat data that does not immediately follow traditional rows-as-observational units and columns-as-variables formatting (e.g. unstructured text).

- The R section of the course begins with students translating their basic programming skills form Python to R. Norm Matloff's text is then used to delve into R's basic data structures including vectors, matrices, lists, data frames, factors, and tables [3]. Students also use R for data visualization and for the majority of team projects.

Team Projects:

Team projects conducted groups of 3 or 4 students are integral to the CS125, taking the place of a scheduled lab time during the week in exchange for an approximately 3-5 hour weekend commitment. The team projects serve to integrate an emphasis on sustainability, a topic that is rarely integrated into an introductory computing class. Through this interdiscplinary approach, students have the opportunity to:

- Explore and visualize monthly global temperatures worldwide from January 1881 to December 2012

- Overlay data from weather satellite images taken of North America at night with a map of US population to identify natural gas flares as a source of light pollution in low-population areas

- Create an animation (http://www.stolaf.edu/people/olaf/cs125/animationProject.gif) in R visualizing seasonal variation in Antarctic sea ice levels over a recent three year period.

- Use BeautifulSoup in Python to scrape a messy U.S. Endangered and Threatened Wildlife list web page and identify at-risk Minnesota wildlife

- Visualize St. Olaf campus energy and electricity use data and then apply those techniques to visualize energy usage on a global scale. Students use machine-learning with a random forest machine learning algorithm to impute missing data and interpret results at both campus and the global levels.

- Utilize Edward Tufte's guidelines to create effective graphics of St. Olaf "farm-to-fork" local food purchases and of Minneapolis - St. Paul temperature data

- All projects are written and submitted in Markdown: Link to Team Projects (http://www.cs.stolaf.edu/wiki/index.php/CS125_S15_team_projects)

Final Project:

In addition to completing daily homeworks and weekly team projects, students are expected to produce a final project applying the skills they have learned throughout the course to a dataset of their choosing.

- Guidelines for the Final Project (http://www.cs.stolaf.edu/wiki/index.php/CS125_Final_Project_Description)

Future Directions

- Link to Biology Collaboration Home Page (http://www.cs.stolaf.edu/wiki/index.php/Team_Project_BIO_231_Collaboration)

Contact Information

- Dr. Olaf Hall-Holt: `olaf@stolaf.edu`
- Dr. Paul Roback: `roback@stolaf.edu`
- Dr. Kevin Sanft

Resources

1. D.G. Sullivan. A data-centric introduction to computer science for non-majors. In *Proceedings of the 44th SIGCSE technical Symposium on Computer Science Education (SIGCSE '13)* pages 71-76, 2013.
2. C. Severance. Python for informatics: Exploring information v. 0.08, 2013. Retrieved from http://www.py4inf.com/book.php
3. N. Matloff. *The Art of R Programming: A Tour of Statistical Software Design* No Starch Press, 2011.