

**Chatham University**  
**Course Syllabus**

---

**GENERAL INFORMATION**

**Course Title:** Introduction to Data Science  
**Course Number:** DSA 150  
**Semester/Year:** Spring 2018  
**Classroom:** Falk 117  
**Class Time:** M W F 10:00-10:50 AM  
**Instructor:** Dr. Stephanie Rosenthal  
**Contact Information:** [s.rosenthal@chatham.edu](mailto:s.rosenthal@chatham.edu)  
**Office Hours:** Falk 116C by appointment

---

**COURSE INFORMATION**

**Course Description:**

Data Science is the study of the tools and processes used to extract knowledge from data. This course introduces students to this important, interdisciplinary field with applications in business, communication, healthcare, etc. Students learn the basics of data collection, data organization, packaging, and delivery. Simple algorithms and data mining techniques are introduced.

**Student Learning Outcomes:**

Program Level Student Learning Outcomes

Graduates of bachelor's-level programs in data science will be able to demonstrate that they possess academic skills, professional attributes and broad-based and in-depth knowledge of data science concepts and functions. This course fulfills the following Program Level Student Learning Outcomes for undergraduate applied data science majors:

- Create effective solutions to computing challenges in analytical projects.
- Critically analyze problems and identify analytical solutions.
- Critically evaluate ethical, privacy and security challenges in data analytics
- Communicate analytics problems, methods, and findings effectively orally, visually, and in writing

Specific learning outcomes for this course include:

- Write algorithms and computer code to collect, organize, analyze and visualize data
- Create algorithms and corresponding computer code given desired data specifications
- Assess and predict the outcome of running code segments on provided data
- Analyze tradeoffs between different algorithms or code segments especially for large datasets

**Required Texts and Materials:**

Doing Data Science. Kathy O'Neill and Rachel Schutt. O'Reilly Media. 2013.

**Optional Texts:**

Python for Data Analysis. Wes McKinney. O'Reilly Media. 2012.

Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications. Laura Igual and Santi Seguí. 2017.

**Course Requirements:**

This course includes 4 data science assignments of increasing difficulty, one data science project, a midterm, and a final exam.

## Data Science Assignments

All programs must be the students' own work (See Course Policies), and must be submitted on time with the student's email and assignment number as the file name. For example, if Stephanie Rosenthal's email is **s.rosenthal@chatham.edu**, her assignments will be named s.rosenthal-1.py, s.rosenthal-2.py, etc. If multiple files are required, they should be zipped (archived) as s.rosenthal-1.zip. Any deviation from this naming scheme will result in a 0 for the assignment. All code must also be commented and include readable variable names for easy evaluation by the instructor. Because we will be going over the answers to the assignments in class, all assignments should be submitted on their **due date by 11:59PM**. A 0 will be given for any assignment not submitted by this deadline.

During the semester, we will focus on creating an analytical solution to a research question of your choice. MORE INFORMATION IS AVAILABLE ONLINE!

*Assignment 1 – Data Collection:* Choose from one of three data sets to analyze this semester. Create a BASH or Python script code to download the information you choose onto your computer. The script must automatically download data at least once a day. Submit the code and the data.

*Assignment 2 – Exploratory Data Analysis and Cleaning:* Describe the data types that exist in your data and whether they need to be converted to other types. Then, write up a description of the existing data including the means and standard deviations of the data, ranges, categories, any data you substituted and/or changed, and explanations of each field.

*Assignment 3 – Visualization:* Data is never perfect when it is downloaded. You will need to transform your data into a format that is easier for analysis. Create code to produce 5 new features of your data. Write up a readme file describing what each new field represents. Make sure that at least one of your features combines two or more of the original fields and at least one of your features analyzes a textual field in the original file. These features should represent the independent variables we will use in your next assignment to perform your data analysis. Visualize the new features.

*Assignment 4 – Data Analysis:* Using the features from Assignment 3, perform either regression or classification on the data. You must compare and contrast three different algorithms to do the analytics. Then, write up your results. What do the algorithms compute? How are they different? What features are your independent variables and what is your dependent variable? Do you get different results from the different algorithms? Why or Why not? Discuss the implications of your results to the dataset and application you collected.

## Data Science Project

Students can choose a data science project of their choice or have one assigned to them. These projects are expected to be between 100-200 lines of code. Students are required to write a proposal describing the goals of their program and algorithms they will include. Proposals should be grammar and spelling checked before submission and may require a meeting with the instructor to clarify, modify, or discuss the project. A mid-point write-up indicating progress, challenges, and next steps as well as a code submission is required three weeks later as well as a meeting with the instructor. The last week of the semester, the final code is due along with a presentation describing final algorithms, security and privacy concerns, and coding challenges. Please use the course website for instructions on how to submit these milestones. As with the assignments, all project milestones should be submitted on their **due date by 11:59PM**. A 0 will be given for any milestone not submitted by this deadline.

The grades will be based on the assignment and project as follows:

- Assignment 1: 5%
- Assignment 2: 10%
- Assignment 3: 10%
- Assignment 4: 10%
- Midterm Exam: 10%
- Project Proposal: 10%
- Project Mid-Point: 10%
- Final Project code: 10%
- Final Project Presentation: 10%
- Final Exam: 10%
- Class Participation: 5%

## Course Policies:

### Attendance and/or participation:

Every student enrolled at Chatham accepts the responsibility to attend all required class meetings. To obtain the fullest benefit from their courses, students must participate fully. This implies attending regularly, engaging in course activity, completing work on time, and making up work missed because of an emergency absence. **It is the student's responsibility to let the course instructor know within the drop-add period if he or she will have to miss class for religious reasons, athletics, or other.**

Attendance will be recorded on a sign-in sheet available at the room entrance. You are not permitted to sign-in for other students. Leaving early constitutes an absence unless I am advised of this necessity at least 24 hours in advance. If you miss a class, it is up to you to get the notes for that day from a classmate. **Six unexcused absences will result in a failure in the course.**

### Code Help, and Collaboration Requirements

**Unless otherwise specified, all code must be completed individually.** In other words, it is okay (and highly recommended) to collaborate in studying the course material, but the "writing on the page" or the "code in the file", as examples, should be your own thought product. **We strongly discourage using the Internet to look for exact answers, but you can use the internet for help. Similarly, help from classmates is encouraged but the helper should not show their code or answers when helping someone else (only the help requester should show their code to their classmate(s)).** Paper and pencil and/or whiteboards are the recommended collaborative study and help tools to prevent any possibility of copying working code.

**At the end of the class, the class participation score will be computed by your classmates. Each classmate will rate your participation on homework assignments and projects.**

If portions of your individual assignments have been significantly influenced by someone else, you should prominently give them credit for their contribution. Proper attribution is critically important -- and is an absolute defense against charges of Academic Dishonesty, Cheating, or Plagiarism.

### Grades will be assigned as follows:

94-100% = A	90-93% = A-	
87-89% = B+	83-86% = B	80-82% = B-
77-79% = C+	73-76% = C	70-72% = C-
67-69% = D+	63-66% = D	60-62% = D-
Below 60% = F		

\* See Chatham catalog for Grade Point Values

### Midterm Grades

Mid-term grades will be posted by noon on Friday February 23rd, 2018. The last date to withdraw from the course is Tuesday March 20<sup>th</sup>.

### Missed Exams/Assignments

**Only absences that are excused by an appropriate authority qualify as excused.** If you must miss a class, please obtain the notes for that day from a classmate and check the class website for updates. Exams must be taken on the scheduled day. There is no provision for makeup tests other than for extreme emergency situations that must be documented by an appropriate authority.

### Laptops and cell/smart phones

Use of electronics for note taking or other functions directly related to class activities is permitted when appropriate. **Note: some days, laptops will be required for in-class programming assignments.** Unrelated online/electronic activity that interferes with your and/or others' ability to pay attention in class

is not permitted. Failure to abide by this policy will result in dismissal from the class and an unexcused absence.

### Behavior

Common courtesy is expected. Every student is expected to come to class prepared to learn and participate in a meaningful way. Under no circumstances is a student's behavior to detract from the learning environment of others in the class. (Disruptive behavior may result in the student's dismissal from the course.)

### **Course Calendar/Schedule:**

The instructor reserves the right to change the daily topics and/or assignment deadlines. Those changes will be made available on the course website.

Week 1: What is Data Science?

- Data science process
- Applications of data science

Week 2: Introduction to Statistical Thinking

- Review of Probability
- Review of Research Methods

Week 3: Data Collection

- Downloading data
- BASH scripting
- Daemons for continual collection

**Assignment 1 out**

Week 4: Exploratory Data Analysis

- Means, Standard Deviations
- Counts
- Data summarization

**Assignment 1 due, Assignment 2 out**

Week 5: Data Cleansing

- Parsing
- Converting data types
- Combining features

**Assignment 2 due, Assignment 3 out**

Week 6: Visualization

- Principles of Visualization
- Counts and Histograms
- Tables, charts, and graphs

Week 7: Introduction to Bayesian Modeling

- Bayes Rule for statistics
- Implications for analytics
- Python SciKit-Learn

**Assignment 3 due, Assignment 4 out**

Week 8 and 9: Machine Learning Part 1

- Classification
- Naïve Bayes
- Logistic Regression
- SVMs
- Decision Trees
- Regression
- Linear Regression
- SVR

**Assignment 4 due, Project out**

Week 10-12: Other Algorithms

- Neural Networks

Graphs  
Text Analysis  
Artificial Intelligence

**Project Proposal due**

Week 13: Privacy and Data Security

Research Ethics  
Research on Privacy and Security  
Current Events

**Project Midpoints**

Week 14: Applications of Data Science

Guest speakers invited

Week 15: **Project Presentations and Discussion**

**Important Dates:**

- Add/Drop Ends Wednesday January 10<sup>th</sup>
- Martin Luther King Day Monday January 15<sup>th</sup>
- Last Day to Withdraw Tuesday March 20<sup>th</sup>

**Plan to attend the Monthly B&E Department Career-Themed Networking Mixers:**

- Thursday Jan. 18 Meet New B&E Dept. Program Chair/Director, Dr. Bonnie Richley  
<http://www.chatham.edu/events/details.cfm?EventID=20500>
- Tuesday Feb. 13 Marketing <http://www.chatham.edu/events/details.cfm?EventID=20499>
- Wednesday March 21 Student- Advisory Board  
Reception <http://www.chatham.edu/events/details.cfm?EventID=20498>
- Thursday April 5 Sustainability  
<http://www.chatham.edu/events/details.cfm?EventID=20501>

## **POLICY STATEMENTS**

### **Chatham University Honor Code:**

*Chatham University students pledge to maintain the Honor Code, which states in part: "Honor is that principle by which we at Chatham form our code of living, working, and studying together. The standards of honor at Chatham require that all students act with intellectual independence, personal integrity, honesty in all relationships, and consideration for the rights and well being of others."*

Information about the Honor Code is available in the [Student Handbook](#).

### **Cheating and Plagiarism:**

Cheating is defined as the attempt, successful or not, to give or obtain aid and/or information by illicit means in meeting any academic requirements, including examinations. Plagiarism is defined as the use, without proper acknowledgement, of the ideas, phrases, sentences, or larger units of discourse from another writer or speaker.

### **Turnitin.com and FERPA**

In all classes, faculty must notify students if the Turnitin service may be used. Student papers are protected by the Family Educational Rights and Privacy Act as they are educational records that contain personally identifiable information. If faculty submits a paper or an excerpt from a paper on behalf of a student for evaluation by Turnitin, an alias must be used instead of the student's name and faculty will ensure that any identifiable personal information is removed before submission.

### **Disability Statement:**

Chatham University is committed to providing an environment that ensures that no individual is discriminated against on the basis of her/his disability. Students with disabilities, as defined under the Americans with Disabilities Act of 1990 (ADA) and who need special academic accommodations, should notify the assistant dean of the PACE Center as soon as possible. The PACE Center will work with students and the course instructor to coordinate and monitor the provision of reasonable academic accommodations.

### **Non-Registered Students Policy:**

In accordance with University policy, only officially registered students may attend this class and all other classes offered at the University after the drop/add period. Please confer with your academic advisor if you need assistance with the registration process or you need additional information.

### **Minimum Grade Requirements:**

Graduate students must earn a grade of B- or above in all courses. Undergraduates must earn a grade of C- or above in all courses completed after spring 2011 used to fulfill major or minor requirements. Please refer to the University catalog or individual program manuals for additional information.

## MINIMUM TECHNOLOGY REQUIREMENTS:

Internet Access	Broadband cable or DSL with a minimum connection speed of 768kbit is recommended; slower connections may not provide optimal course experience and performance
Operating System	Microsoft Windows 7 or higher (PC) Mac OS X 10.6 or higher (MAC) Current students may purchase Operating System upgrades from the Chatham Helpdesk
Processor Type	2.0 GHz or higher
System Memory	4GB RAM or higher
Monitor	1024x768 or higher screen resolution
Software	Microsoft Office 2013 or higher (PC) Microsoft Office 2011 or higher (MAC) All students will be provided with Microsoft Office 365 Current students may purchase Microsoft Office from the Chatham Helpdesk
Web Browser	Mozilla Firefox (Recommended for Moodle), or Google Chrome Incognito (Recommended for myPortal); other browsers such as Internet Explorer, Opera and Apple Safari are not recommended
Storage	500GB of hard drive or greater
Audio	Computer speakers and headphones
Visual	Web Camera
E-mail	Chatham University e-mail account (Microsoft Office 365)
Web Conferencing	Courses using web conferencing for online meetings require the following: <ul style="list-style-type: none"> <li>• For audio: headphones and microphone</li> <li>• For video: web camera</li> </ul>
Plug-ins	Course content may include file types that require special plug-in software, which are typically available as free downloads (ex: Real Player, Java, QuickTime, Silverlight, Adobe Reader and Adobe Flash)
Mobile Devices	Some resources are available via smartphones and tablets. Please note: Mobile devices will not be able to complete all course requirements. Students will still need regular access to a computer.
On Campus Resources	Current students have access to the following resources: 24 Hour Computer Lab – JKM Library 106 Computer Lab – JKM Library 101 Computer Lab – Buhl 236 (no printer) Computer Lab – Coolidge 42 Computer/CAD Lab – Eastside 209 Chatham IT Helpdesk – Woodland 100, Eastside 219, Eden Hall Lodge Library
Off Campus Resources	Current students have access to the following resources: Atomic Learning ( <a href="http://www.atomiclearning.com/">http://www.atomiclearning.com/</a> ) Chatham IT ( <a href="http://www.chatham.edu/its">http://www.chatham.edu/its</a> ) Chatham IT Helpdesk ( <a href="http://services.chatham.edu">http://services.chatham.edu</a> )
Current Technologies	For the most up-to-date technology, please visit Chatham IT ( <a href="http://www.chatham.edu/its">http://www.chatham.edu/its</a> )