

**CISC 5420 - Applied Statistics & Probability**  
**Department of Computer and Information Science**  
**Luigi Patruno, Fall 2017**

**Class Times:** Tuesday, 5:30 - 7:45 PM in LL 404

**Instructor:** Luigi Patruno

**Email:** [lpatruno@fordham.edu](mailto:lpatruno@fordham.edu)

**Office Hours:** Tuesday 7:45 - 8:45 pm, 3rd Floor Lounge. Please email me to schedule additional office hours!

**Text:** Think Stats 2 by Allen B. Downey. This book is available free of charge in PDF form (<http://greenteapress.com/thinkstats2/>). A paper version of the text can be purchased from Amazon, if desired. We will also be using OpenIntro Statistics 3rd edition. This book is available free of charge in PDF form (<https://www.openintro.org/stat/textbook.php>).

**Course Website:** <https://lpatruno.github.io/stats-fall17/>

**Course Description:** This course is an introduction to exploratory data analysis with an emphasis on practical tools and methods for applying statistical methods to real world data. Rather than emphasizing mathematical theory, the course takes a computational approach to analysis in order to equip the student with the skills necessary to ask and answer questions about real world phenomena by analyzing data. Students will be required to write and execute code in the Python programming language in order to: import and clean data, perform single, pairwise, and multivariate analyses, perform hypothesis testing, and visualize their results. Each student will select and complete a midterm project and final project.

**Objectives:**

The objective of this course is to develop the ability to answer questions about the real world by performing statistical analyses on real data sets. The student will gain experience with performing all of the steps, including searching for and cleaning data, in order to perform these analyses.

**Attendance and Class Participation:** It is important to attend and be prepared for every class. For this course, being prepared means completing the weekly reading assignment, running through and understanding (as best you can) the sample code, and completing the weekly online quiz. The classes will be collaborative and discussion based in order facilitate and foster ideas. If you are going to miss a class, please let me know beforehand.

**Academic Honesty:** All work produced in this course should be your own unless it is specifically stated that you may work with others. You may discuss the homework problems with other students generally, but may not provide complete solutions to one another; copying of homework solutions is always unacceptable and will be considered a violation of Fordham's academic integrity policy. Violations of this policy will be handled in accordance with university policy which can include automatic failure of the assignment and/or failure of the course.

**Grading:** The percentages given below are guidelines for both the student and the instructor and minor changes may be made during the course. Students will be informed promptly of any such changes.

Homework & Quizzes	30%
Midterm Project	30%
Final Project	30%
Participation	10%

**Projects:** The course will include a midterm and final project where you will be asked to apply the techniques you've learned to a real world data set. I encourage each of you to work on projects that relate to your interests. For instance, search for data sets pertaining to the music industry if you're passionate about music, or work on data sets related to athletics, financial information, or social media. We can also discuss techniques for retrieving data sets that are not as easily accessible (for example, scraping craigslist.com to get a data set of all items for sale in New York City). I will ask you to submit proposals for these projects.