

# EECS6414 Data Analytics and Visualization

Winter 2019

## Course Website

[www.eecs.yorku.ca/~papaggel/courses/eecs6414/](http://www.eecs.yorku.ca/~papaggel/courses/eecs6414/)

## Course Description

Data analytics and visualization is an emerging discipline of immense importance to any data-driven organization. This is a project-focused course that provides students with knowledge on tools for data mining and visualization and practical experience working with data mining and machine learning algorithms for analysis of very large amounts of data. It also focuses on methods and models for efficient communication of data results through data visualization.

## Topics

Topics include:

- finding similar items
- frequent item sets
- mining data streams
- clustering
- dimensionality reduction
- link analysis
- mining graphs
- recommendation systems
- value of visualization
- exploratory data analysis
- visualization of multidimensional data
- visualization of networks
- tools/systems for data analytics and visualization

## Instructor

Manos Papagelis

Email: [papaggel@eecs.yorku.ca](mailto:papaggel@eecs.yorku.ca), [papaggel@gmail.com](mailto:papaggel@gmail.com)

Website: <http://www.eecs.yorku.ca/~papaggel>

## Class Hours

Lectures: *Mon 16:00pm-19:00pm at VH 2005 (Vari Hall)*

Office Hours: *Drop by my office or by appointment (LAS3050, Lassonde building)*

## Class Attendance

Attendance of lectures is expected but not required.

## Prerequisite Courses

There are no course prerequisites for this course; students are expected to be competent in:

- Algorithms (basic data structures, dynamic programming, ...)
- Basic probability & linear algebra (moments, typical distributions, MLE, ...)
- Programming (your choice, but Python/C++/Java will be very useful)

If you don't satisfy these, you need to talk with the instructor in the first week of classes to see whether you may remain in the course.

## Textbooks

There is no single text for this course. The course will rely mainly on the following suggested textbooks:

- **BOOK TITLE:** Mining of Massive Datasets, 2nd Edition  
**AUTHORS:** Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman  
**DATE PUBLISHED:** December 2014  
**FORMAT:** Hardback  
**PUBLISHER:** Cambridge University Press  
**ISBN:** 9781107077232  
**Note:** The aforementioned book is freely available online.
- **BOOK TITLE:** Networks, Crowds, and Markets: Reasoning About a Highly Connected World  
**AUTHORS:** David Easley, Jon Kleinberg  
**DATE PUBLISHED:** July 2010  
**FORMAT:** Hardback  
**PUBLISHER:** Cambridge University Press  
**ISBN:** 9780521195331  
**Note:** The aforementioned book is freely available online.
- **BOOK TITLE:** Introduction to Data Mining, Second Edition  
**AUTHORS:** Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar  
**DATE PUBLISHED:** This title has not yet been published. It is scheduled to be available on Jan 2018  
**FORMAT:** Hardback  
**PUBLISHER:** Addison-Wesley, 2017  
**ISBN:** 9780133128901
- **BOOK TITLE:** Social Media Mining: An Introduction  
**AUTHORS:** Reza Zafarani, Mohammad Ali Abbasi, Huan Liu  
**DATE PUBLISHED:** April 2014  
**FORMAT:** Hardback  
**PUBLISHER:** Cambridge University Press  
**ISBN:** 9781107018853  
**Note:** The aforementioned book is freely available online.
- **BOOK TITLE:** The Visual Display of Quantitative Information (2nd Edition)  
**AUTHORS:** Edward R. Tufte  
**DATE PUBLISHED:** 2001  
**FORMAT:** Hardback  
**PUBLISHER:** Graphics Press  
**ISBN:** 9781930824133
- **BOOK TITLE:** Envisioning Information  
**AUTHORS:** Edward R. Tufte  
**DATE PUBLISHED:** 1990

FORMAT: Hardback  
PUBLISHER: Graphics Press  
ISBN: 9781930824140

- **BOOK TITLE:** Interactive Data Visualization for the Web, 2nd Edition  
**AUTHORS:** Scott Murray  
**DATE PUBLISHED:** 2017  
**FORMAT:** Hardback  
**PUBLISHER:** O'Reilly Media  
**ISBN:** 9781491921289

In addition, a number of recent research papers in the area of data analytics and visualization will be distributed in every iteration of the class.

### Communication

The main communication tools for the class will be the *course website* and *Piazza*.

- *Course Website:* Most class materials are available on the course web site; be sure to check regularly. The page has also a link to a discussion board. We are using Piazza.
- *Piazza:* Instead of a discussion board, we are using Piazza, a free Q&A platform. Piazza can get you fast, accurate response to your questions – but it only works if everyone participates! We will also use Piazza to post announcements and updates, so both the website and Piazza is required reading. See below for Piazza signup and class links:

*Signup link:* <http://www.piazza.com/yorku.ca/fall2017/eecs6414/home>

*Class link:* <http://www.piazza.com/yorku.ca/fall2017/eecs6414/home>

**Note:** You will need to sign up with your school email, ending in *yorku.ca*. If you do not have a school email address, please contact your instructor and request to be enrolled with your personal email.

- *Email:* Please use email for personal issues and the discussion board to ask general course-related questions. Include “eecs6414” in all email subject lines to ensure your message is correctly filtered and filed. An informative subject line like “eecs6414: Question related to X” really helps. I try to respond to email frequently. However, due to volume, it may take longer, especially on weekends and near due dates.

### Grading Policy

The evaluation of the course will be based on projects done by teams (of up to three)

Milestone	Weight
Project proposal	10%
Project midterm report	20%
Project midterm in-class presentation	10%
Project final report	40%
Project final in-class presentation	20%

## **Final Examination**

There is no final examination.

## **Working with a Partner**

You have the option of partnering with other (currently enrolled) students for your project, and we encourage you to do so. The ability to work effectively in a team will be very important in your career, and that involves many skills beyond the purely technical aspect of creating working code. You may choose your own partner(s). If you do have a partner, submit only a single copy of your work. Jointly submitted assignments will be graded in the usual way and all partners will receive the same mark. Working with a partner has the potential to lighten your workload or to increase it, depending on how well you work together. Be aware that simply dividing the work and assembling your separate pieces at the end is a poor strategy for completing successful assignments. And of course, you are responsible for learning the course material underlying all parts of the assignments. You will have the most success if you truly work together.

## **Assignment Policies**

Here assignment refers to project milestones. You must make sure that all your assignments are running and are sufficiently documented. Code that doesn't compile, fails to run or lacks documentation, will be marked as not working.

## **Late Work Policy**

Here assignment refers to project milestones. The late policy is strict. All assignments will be submitted electronically. Late assignments will be handled based on a system of "*grace days*", as follows: Each student begins the term with 3 *grace days*. One *grace day* is 24 hours. If an assignment is due at 10:00 p.m. on a Friday then an assignment handed in by 10:00 p.m. on Saturday uses one *grace day*. The *grace days* are intended for use in emergencies (e.g., system failure or illness). Do not use all of them to buy an extension because of a busy week or you will be out of luck in a true emergency. Assignments submitted after the due date when all *grace days* have been used will receive a grade of 0.

If you are at risk of missing a deadline due to a busy week, rather than using your *grace days* you should hand in a working (and tested) version of a simpler program. In the event of an illness or other catastrophe, get proper documentation (e.g., medical certificate), and contact me (by email or in person) as soon as possible. Do not wait until the due date has passed. It is always easier to make alternate arrangements before the due date or test day.

Assignments are submitted electronically and will often be tested using an automated testing program; you must follow the submission instructions exactly. If you do not, you will most likely lose substantial marks on the assignment. If you find you have submitted the wrong file or omitted a file, please notify your instructor as soon as possible.

## **Remarking**

Here assignment refers to project milestones. If you feel an error was made in marking an assignment or test please submit a remark request. Requests for remarking must be submitted using a university remarking request form explaining what your concern is **no later than a week after** the assignment (or test) has been returned back.

**Academic Offenses**

All of the work you submit must be done by you and your work must not be submitted by someone else. Plagiarism is academic fraud and is taken very seriously. The department uses software that compares programs for evidence of similar code. Please read the Rules and Regulations from the [York University's Academic Integrity](#) and the [York University's Senate Policy on Academic Honesty](#) documents.

**Accessibility Needs**

York University is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom or course materials, please contact [York University's Counselling & Disability Services](#).