Announcements

CSE 6242 / CX 4242 Data & Visual Analytics
Fall 2016

Polo Chau
Georgia Tech
Thu, 8/25
How to address Polo?

**Grammatically correct**

Prof. Chau

Dr. Chau

**Grammatically incorrect, but popular**

Prof. Polo

Dr. Polo
Course Homepage

Google “cse6242”

poloclub.gatech.edu/cse6242/
Join Piazza ASAP

goo.gl/LBlsBw

Announcements and Discussion on Piazza

We use Piazza for discussion and all announcements.

Post your questions there. Our teaching staff and your fellow classmates will help answer them; we also use Pizza to find project teammates.

T-square will only be used for submission of assignments and projects.

While we welcome everyone to share their experiences in tackling issues and helping each other, we ask that you not post your answers, as that may affect the learning experience of your fellow classmates.
It’s Important to Join Piazza, because…

**The fastest way** to get help with homework assignments is to post your questions on Piazza. If you prefer that your question addresses to only our TAs and the instructor, you can use the *private post* feature (i.e., check the "Individual Students(s) / Instructors(s)" radio box).
It’s Important to Join Piazza, because…

• Polo will announces events related to this class and data science in general

• Distinguished lectures

• Seminars

• Hackathons (free food, prizes)

• Company recruitment events (free food, swag)
While collaboration is allowed for homework assignments, each student **must** write up their own answers. All GT students must observe the honor code. Any suspected plagiarism and academic misconduct will be reported and directly handled by the **Office of Student Integrity (OSI)**.
Late Submissions Policy

- **Homework:** each student has **4 slip days** total. No questions asked.
- **Project:** each team has **3 slip days** total. No questions asked. Slip days may not be used on in-class activities (e.g., proposal presentation, poster presentation, etc.).
- To use slip days, **specify the number of days you have used in the textbox on T-Square** (when you submit your work).
- **Each slip day equals 24 hours.** E.g., if a submission is late for 30 hours, that counts as 2 slip days.
- After all slip days are used up, **5% deduction for every 24 hours of delay.** (e.g., 5 points for a 100-point homework)
- We **will not consider late submission of any missing parts** of an homework assignment or project deliverable. To make sure you have submitted everything, download your submitted files to double check.
- **No penalties for medical reasons or emergencies.** You **must** submit a doctor's note or an official letter explaining the emergency.
Distance Learning Sections (Q & Q3)

A standard 3-day lag applies to all homework and project deliverables. For project presentation, a group that has DL student member can choose to:

1. Present in class without 3-day lag; or
2. Submit a video presentation with 3-day lag (e.g., screen capture)
Prerequisites & Expectation
For both CSE 6242 (grad) and CX 4242 (undergrad)

Students are expected to complete significant programming assignments (homework, project) that may involve higher-level languages or scripting (e.g., Java, R, Matlab, Python, C++, etc.). Some assignments may involve web programming and D3 (e.g., Javascript, CSS).

You are expected to quickly learn many new things. For example, an assignment on Hadoop programming may require you to learn some basic Java and Scala quickly, which should not be too challenging if you already know another high-level language like Python or C++. Please make sure you are comfortable with this.

Please take a look at the assignments (homework and project) of the previous offerings of this course, which will give you some idea about the difficulty level of the assignments.

Basic linear algebra, probability knowledge is expected.

http://poloclub.gatech.edu/cse6242/2016spring/
Getting to know you better

• Thursday after class, Polo invites 5 students to coffee (randomly chosen) for a casual chat
Tue, 8/30
Getting to know you better
(Casual chat. You can ask me anything)

• Last Thursday’s coffee after-session session was great! 5 students: Chris, Sam, Jack, Anmol, Nitin

• Expand to Tuesday after class?
Course Project

• You’re welcome to look for teammates now. Deadline: **Friday, Sept 23**

• **4-6 people** in one team

• Polo suggests teaming with students in the **same section**, but OK to mix if really needed. See “Teaming” section on project description: [http://poloclub.gatech.edu/cse6242/2016fall/project.html](http://poloclub.gatech.edu/cse6242/2016fall/project.html)
Course Project

- Two example projects presented on **Thu, Sept 8**
  - **Firebird**: Predicting Fire Risk in Atlanta
  - **PASSAGE**: Safe Path Recommendation
  - Q&A after presentations
Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta

Michael Madaio
Carnegie Mellon University
Pittsburgh, PA, USA
mmadaio@cs.cmu.edu

Shang-Tse Chen
Georgia Tech
Atlanta, GA, USA
schen351@gatech.edu

Oliver L. Haimson
University of California, Irvine
Irvine, CA, USA
ohaimson@uci.edu

Wenwen Zhang
Georgia Tech
Atlanta, GA, USA
wzhang300@gatech.edu

Matthew Hinds-Aldrich
Atlanta Fire Rescue Dept.
Atlanta, GA, USA
mhinds-aldrich@atlantaga.gov

Xiang Cheng
Emory University
Atlanta, GA, USA
xcheng7@emory.edu

Bistra Dilkina
Georgia Tech
Atlanta, GA, USA
bdilkina@cc.gatech.edu

Duen Horng Chau
Georgia Tech
Atlanta, GA, USA
polo@gatech.edu

ABSTRACT
The Atlanta Fire Rescue Department (AFRD), like many municipal fire departments, actively works to reduce fire risk by inspecting commercial properties for potential hazards and fire code violations. However, AFRD’s fire inspection practices relied on tradition and intuition, with no existing data-driven process for prioritizing fire inspections or identifying new properties requiring inspection. In collaboration with AFRD, we developed the Firebird framework to help municipal fire departments identify and prioritize commercial property fire inspections, using machine learning, geocoding, and information visualization. Firebird computes fire risk scores for over 5,000 buildings in the city,
PASSAGE: A Travel Safety Assistant With Safe Path Recommendations For Pedestrians

Matthew Garvey  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
mgarvey5@gatech.edu

Meghna Natraj  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
mnatraj@gatech.edu

Nilaksh Das  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
nilakshdas@gatech.edu

Bhanu Verma  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
bhanuverma@gatech.edu

Jiaxing Su  
College of Engineering  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
jiaxingsu@gatech.edu

Abstract

Atlanta has consistently ranked as one of the most dangerous cities in America with over 2.5 million crime events recorded within the past six years. People who commute by walking are highly susceptible to crime here. To address this problem, our research group has developed a mobile application, PASSAGE, that uses real-time crime data to find "safe paths" for pedestrians in Atlanta.

Figure 1: Paths recommended by PASSAGE.
HW1

- Out this Friday
- At least 2 weeks to complete
Tue, 9/6
HW1 is out!
Start now.
Yahoo Tech Talk on Tue, 9/20

• Industry guest speaker + info session during class time (4:30-6pm)

• Learn about what’s REALLY happening at tech companies

• Free food and swag!

• Doesn’t hurt to have your CV handy!

• Some companies start making full-time offers on-site
Form Teams by **Fri, 9/23** for Course Project

- **4-6 people** in one team

- Polo suggests teaming with students in the **same section**, but OK to mix if really needed. See “Teaming” section on project description: [http://poloclub.gatech.edu/cse6242/2016fall/project.html](http://poloclub.gatech.edu/cse6242/2016fall/project.html)
Coffee After Class

• Two changes
  • Will sample with replacement (i.e., students could be repicked)
  • 5 randomly chosen + 2 volunteers
Today: 2 Example Projects

- **Firebird**: Predicting Fire Risk in Atlanta
- **PASSAGE**: Safe Path Recommendation
- Q&A after each presentation
Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta

Michael Madaio  
Carnegie Mellon University  
Pittsburgh, PA, USA  
mmadaio@cs.cmu.edu

Shang-Tse Chen  
Georgia Tech  
Atlanta, GA, USA  
schen351@gatech.edu

Oliver L. Haimson  
University of California, Irvine  
Irvine, CA, USA  
ohaimson@uci.edu

Wenwen Zhang  
Georgia Tech  
Atlanta, GA, USA  
wzhang300@gatech.edu

Xiang Cheng  
Emory University  
Atlanta, GA, USA  
xcheng7@emory.edu

Matthew Hinds-Aldrich  
Atlanta Fire Rescue Dept.  
Atlanta, GA, USA  
mhinds-aldrich@atlantaga.gov

Duen Horng Chau  
Georgia Tech  
Atlanta, GA, USA  
polo@gatech.edu

Bistra Dilkina  
Georgia Tech  
Atlanta, GA, USA  
bdilkina@cc.gatech.edu

ABSTRACT
The Atlanta Fire Rescue Department (AFRD), like many municipal fire departments, actively works to reduce fire risk by inspecting commercial properties for potential hazards and fire code violations. However, AFRD’s fire inspection practices relied on tradition and intuition, with no existing data-driven process for prioritizing fire inspections or identifying new properties requiring inspection. In collaboration with AFRD, we developed the Firebird framework to help municipal fire departments identify and prioritize commercial property fire inspections, using machine learning, geocoding, and information visualization. Firebird computes fire risk scores for over 5,000 buildings in the city,
PASSAGE: A Travel Safety Assistant With Safe Path Recommendations For Pedestrians

Matthew Garvey  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
mgarvey6@gatech.edu

Meghna Natraj  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
mnatraj@gatech.edu

Nilaksh Das  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
nilakshdas@gatech.edu

Bhanu Verma  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
bhanuverma@gatech.edu

Jiaxing Su  
College of Engineering  
Georgia Institute of Technology  
Atlanta, GA 30332, USA  
jiaxingsu@gatech.edu

Abstract

Atlanta has consistently ranked as one of the most dangerous cities in America with over 2.5 million crime events recorded within the past six years. People who commute by walking are highly susceptible to crime here. To address this problem, our group has developed a mobile application, PASSAGE, which uses crowd-sourced and crime data to find "safe paths" for users in Atlanta.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Tue, 9/13
Google on Campus

Many Googlers (GT alumnus) and HR representatives are visiting Georgia Tech campus this week, starting today. Below is the schedules for Google event during this week. If you are interested in full-time positions (assuming graduation by next Summer) or internship, please don't miss this great opportunity to meet Googlers. Mark on your calendar!

Google culture + Intern lightning talk: 9/13 Tue (6pm - 7:30pm, KACB 1116)

Women@CC Lunch: 9/14 Wed (11:30am - 1pm, CCB 345)

Interview prep workshop: 9/14 Wed (6pm - 8pm, Klaus Atrium)

Navigating the hiring process: 9/15 Thr (6pm - 8pm, KACB 1443)

Lunch with Googlers: 9/16 Fri (11:30am - 1pm, CCB 345)

CoC Career Fair: 9/20 Tue (9:30am - 3:30pm, Klaus Atrium) -- Registration required to CoC Career fair.
Thu, 9/15
In-class Pop Quiz

Pop Quiz #1 at ~5pm today
This first quiz is a “dry run”
Does NOT contribute to any bonus points to your course grade

Future pop quizzes WILL take place on random days
In-class Pop Quizzes

• Everyone should have received an email from classmarker.com last night.

• **Pop Quiz #1** around 5pm today
  
  • This first quiz is a “dry run”
  
  • Does NOT contribute bonus points to your course grade
  
  • 0 (zeroes) may be part of the code you received

• Future pop quizzes will take place on random days and **WILL** contribute bonus course grade points
Thu, 9/22
Form Teams by Tomorrow (Friday, 9/23)

• https://t-square.gatech.edu/access/content/group/gtc-7467-3646-5b71-8e4c-f64b39a4370c/https%3A__docs.google20160921184356.URL
If you want feedback on your project ideas before proposal presentation...

• Come talk to me during my office hour

• Join our after-class coffee

• Send a private Piazza post to our TAs and me (You’ll get a faster response. Don’t email me directly — I get hundreds of emails each day!)
Thu, 9/29
Everyone should have joined a project team by now

• If not,

  • Please stay behind today

  • Check out the team spreadsheet. There are teams still looking for more members.

  • Come talk to me. The later you come talk to me, the less help I can offer.

• If you have joined a team, that means you have made a commitment.
Tips for Working Successfully in a Group

1. Meet people properly
2. Find things you have in common
3. Make meeting conditions good
4. Let everyone talk
5. Check your egos at the door
6. Praise each other
7. Put it in writing
8. Be open and honest
9. Avoid conflict at all costs
10. Phrase alternatives as questions

https://www.cs.cmu.edu/~pausch/Randy/tipoForGroups.html
HW2

- Q3 part ii and iii revised. Thanks Alyse Keim for spotting the issues.

- **Q:** Is it OK to start with code found on the web?  
  **A:** No  
  **Q:** Why?  
  **A:** Here’s why…
Do not plagiarize!

- Using code as reference does not mean copying and pasting that code. Nor does that mean copying in a block of code and then modifying parts of it.

- If you want to use some code for reference, you should go over it, understand what it is doing, and then try to accomplish what it is trying to do **using your own code**. And it’s a good practice to cite the sources (e.g., as part of your code comments).

- The analogy is like how you would write an essay or a speech. You can get inspirations from others, but you should **use your own words**, otherwise it will be considered plagiarism. As I mentioned in class, and in the beginning of every homework, plagiarism can lead to heavy consequences.

After-class coffee today…
with TAs
Thu, 10/6
Everyone now has 6 slips days

- As announced on Piazza
- You can use your remaining slip days on HW2, 3 and 4
Project

- Everyone should have joined a team by now!

- Consider using LaTeX for your project
  - e.g., ACM
    https://www.acm.org/publications/proceedings-template
  - Especially helpful if your team plans to turn the project into publication (just swap out LaTeX style file)
  - Easily cite and organize papers your team has reviewed

- Overleaf for collaborative LaTeX editing
Thu, 10/13
Project

- Proposal and slides due on Monday 10/17
- In-class project presentations on Tue and Thu. Mandatory attendance.
Please back up your work!

- use private Github
- Computers died
- Computers got stolen!
Tue, 10/23
Project

- Proposal presentation grades + feedback to be released tomorrow (Wed, 10/26)
Quiz

1. There may or may not be a quiz this Thursday
Thu, 10/27
Project

- Proposal presentation grades + feedback to be released tomorrow (Thu, 10/27)
Quiz

- There will be a quiz today, around 30min from now.
- **10min** to finish.
- Count towards **1 bonus course** grade point
- Quizzes are meant to refresh your memory
  - and NOT to stress you out
- **Lowest scoring quiz will be dropped**
  - e.g., if we give 3 quizzes total this semester, only the top 2 scores count
  - **Reduces, but not remove** the chance of missing extra credit
Today’s random 5 students (+2 volunteers) for after-class coffee ☕

No coffee today.

Polo will hold office hour 6-7pm in my office Klaus 1324
Tue, 11/1
HW3: AWS and Azure

• Turn off EVERYTHING when you’re done (don’t leave them on unnecessarily).

• If you get charged by AWS
  • First, don’t panic
  • Second: call them (yes, use your phone) and they will be able to waive the charges
HW3: academic misconduct

- Someone posted a question on Stackoverflow and asked others to solve it.
- Clearly not allowed!!
Quiz

• There may or may not be a quiz next week…

• Will go through last quiz’s answers this Thu
Thu, 11/3
Quiz

• There may or may not be a quiz next week…

• Going through last quiz’s answers today
What’s your experience using AWS and Azure?

- Azure clunky; Azure seems to be quicker to set up
  - errors on AWS (verification, limit of instance)
- Azure seems to be using credit quite quickly
- timing harder to predict
- need to ssh into azure
- cost
Good Read about AI: White House Report

Preparing for The Future of Artificial Intelligence

https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
Tue, 11/8
Project

- Progress reports due Sunday
- Project proposal documents to be released by Wed
Quiz

- No quiz today
- Possibly a quiz on Thursday
Today’s random 5 students (+2 volunteers) for after-class coffee ☕

1. Will ask for volunteers at the end of class
Thu, 11/10
Project

- Progress reports due Sunday
- Project proposal documents to be released today o T-square, to contact persons, by 7:30pm. Sorry for the delay!
Quiz

• Quiz later today