Introduction

Data scientists, including industry analysts, scientific researchers and data-driven policy makers, often want to analyze data that contains sensitive personal information that must remain private. However, common techniques for data sharing that attempt to preserve privacy either bring great privacy risks or great loss of information. Moreover, the increasing ability of big data, ubiquitous sensors, and social media to record lives in detail brings new ethical responsibilities to safeguard privacy.

The traditional approach to protecting privacy when sharing data is to remove "personally identifying information," but it is now known that this approach does not work, because seemingly innocuous information is often sufficient to uniquely identify individuals. A long literature has shown that anonymization techniques for data releases are generally open to reidentification attacks. Indeed, there have been many high-profile examples in which individuals in supposedly anonymized datasets were re-identified by linking the remaining fields with other, publicly available datasets. Aggregated information can reduce but not prevent this risk, while also reducing the utility of the data to researchers.

This class will provide an overview of the risks of private data leakage in data science applications and a firm foundation in how to measure and protect against these risks using the framework of differential privacy, together with a hands-on examination of how to build
algorithms and software to preserve privacy, including a review of the deployed solutions in industry and government.

Differential privacy, deriving from roots in cryptography, is a formal, mathematical conception of privacy preservation. It guarantees that any released statistical result does not reveal information about any single individual. That is, the distribution of answers one would get with differentially private algorithms from a dataset that does not include myself must be indistinguishable from the distribution of answers where I have added my own information.

Using differential privacy enables us to provide wide access to statistical information from a privacy sensitive dataset without worries of individual-level information being leaked inadvertently or due to an adversarial attack. There is now both a rich theoretical literature on differential privacy and numerous efforts to bring differential privacy closer to practice, including large-scale deployments by Google, Apple, Microsoft and the US Census Bureau. This course will set out a foundation in the underlying theory of differential privacy, and then consider the practical elements of implementing and deploying privacy-preserving techniques for data analysis.

Format and Goals

The class will typically alternate between lecture/discussion meetings, which will focus on learning the fundamentals and discussion of important issues, and problem-solving sessions where there will be some lecture, some demonstration code, and some hands-on computer work. Homeworks will typically involve some analytical/mathematical work to learn techniques, and increasingly as the term progresses, hands-on data immersive coding tasks to test and experiment with approaches to privacy preservation within the context of real datasets and data science questions.

The main components of the course are as follows:

- **Reading and commenting:** For every class meeting, we will assign reading for you to do in advance. You will be expected to read and comment on this material prior by midnight before lecture, using the online forum NB.

- **Class participation:** Attendance is mandatory, as our meetings will be highly interactive.

- **Problem Sets:** There will be problems sets approximately every week or every second week. These will be progressive, and require reuse of previous solutions, so it is important both to keep up on the problems, review feedback to submissions, and organize and document previous submitted code so that it can be reused.

- **Final Project:** You will do a final project on a topic of your choosing. Projects can be done individually or in pairs, with groups of three allowed for ambitious projects. You can do a project that is experimental, or involves system-building, or is theoretical. The project should provide good opportunities to connect the course material to your other interests and get some exposure to the frontier of research in differential privacy. The project will involve submitting topic ideas for feedback (in mid-February, and revised in mid-March), a
detailed project proposal (due mid-April), a written paper (draft due in reading period, final version in exam period), and a project presentation (in exam period). We will post more details about the final project, including some directions to look for topics, early in the course.

By the end of the course, we hope that you will all be able to:

- Identify and demonstrate risks to privacy in data science settings,
- Correctly match differential privacy technology with an application,
- Safely implement privacy solutions, and experimentally validate the performance and utility of algorithms.
- Understand differential privacy at a level sufficient to engage in research about best practices in implementation, apply the material in practice, and/or connect it to other areas,
- Formulate and carry out an interesting, short-term independent research project, and present the work in both written and oral form.

**Prerequisites**

Basic probability, algorithms, and programming at the level of CS109/AC209. STAT110 and CS124 should also be sufficient preparation.
Outline of Topics

- Attacks
  - Reidentification attacks: see exercise in CMU course on implementing Narayanan-Shmatikov
  - Reconstruction attacks
  - Differencing attacks
  - Membership attacks (Homer et al., Fingerprinting, ML/Deep Learning)
- Foundations of differential privacy
  - Epsilon indistinguishability definition
  - Composition
  - Core differentially private mechanisms (Laplace, Gaussian, exponential, geometric, randomized response)
  - Survey of known algorithms and experimental validation
- Centralized DP
  - Curators, interaction, and budgeting
- Local DP
  - Changes in trust model, utility
  - Randomized response, histograms
- Existing deployments and challenges in them
  - Census
  - Google - RAPPOR, ESA/Prochlo
  - Apple
  - Microsoft
- Existing DP software/implementations/tools
  - PSI
  - PinQ (& wPinQ)
  - Ektelo
  - GUPT
  - Uber library (Song et al.)
- Dangers in DP implementations
  - Mironov attack
  - DP under fire (timing)
  - Sources of randomness
- Policy considerations
  - Setting epsilon, utility vs. harm considerations
  - Relation to privacy law
  - Combining DP with other controls, tiered access models

Grading
For grading, we will place approximately equally weight on each of the following 3 categories:

- Reading and commenting on the reading and class participation
- Problem sets
- Final project

Problem set solutions must be typed and submitted electronically. The deadline will typically be set at 5pm on a Friday. You are allowed 6 late days for the semester, of which at most 4 can be used on any individual problem set. (1 late day = 24 hours exactly). The problem sets may require a lot of thought, so be sure to start them early. You are encouraged to discuss the course material and the homework problems with each other in small groups (2-3 people). Discussion of homework problems may include brainstorming and verbally walking through possible solutions, but should not include one person telling the others how to solve the problem. In addition, each person must write up their solutions independently, and these write-ups should not be checked against each other or passed around.