Introduction

Jingchen (Monika) Hu

Vassar College

Data Confidentiality

Outline

1 Course orientation

Outline

1 Course orientation

General info

Instructor: Jingchen (Monika) Hu - jihu@vassar.edu
                      RH 403

Lecture: Tuesdays 3:10-5:10pm
                      RH 101

Lab: Some lectures will be used as labs.

Office hours: Wednesday 10:00am-12:00pm & Thursday 11:30am-12:30pm, or by appointment.

Required materials

Prerequisite: MATH 347 Bayesian Statistics, and interests in data confidentiality

Readings: A collection of selected journal articles and manuscripts

Software: We will use the software R/RStudio for labs and project. Download R from *http://www.r-project.org/* and RStudio, from *https://www.rstudio.com/*

Webpage: The course GitHub page *https://github.com/monika76five/Data-Confidentiality*. Tentative schedule, to-do lists, datasets, R scripts...

Course topics

- <u>Disclosure risks</u> in microdata
- <u>Protection methods</u> for microdata
  - ▶ Bayesian data synthesis
    - ★ for continuous variables
    - ★ for binary variables
    - ★ for categorical variables
- <u>Evaluation</u> of protection methods
  - ▶ data utility
  - ▶ disclosure risks
- Other protection methods
  - ▶ differential privacy
  - ▶ differentially private data synthesis

Course components

- The three parts of material: <u>disclosure risks, protection, evaluation</u>

Course components

- The three parts of material: <u>disclosure risks, protection, evaluation</u>

- Each part is a combination of a selection of the following: readings, lectures, labs, homework, discussions (in-class and online), and project

Course components

- The three parts of material: <u>disclosure risks, protection, evaluation</u>

- Each part is a combination of a selection of the following: readings, lectures, labs, homework, discussions (in-class and online), and project

- The course project (individual or in pair; cross-campus collaboration is highly encouraged!) is a final product, using Bayesian data synthesis to provide privacy protection

Course project

- Find a dataset and demonstrate disclosure risks
- Design and implement protection methods
- Evaluate protection methods
- Write a paper and present the results

Course project

- Find a dataset and demonstrate disclosure risks
- Design and implement protection methods
- Evaluate protection methods
- Write a paper and present the results

Questions? Ideas to discuss and share?