UMBC

DATA 605
Ethical & Legal Issues in Data Science

SPRING 2022

SUNELA THOMAS

MARCH 10, 2022

AGENDA

- Questions?
- Paper and Discussion Graded
- AI with Ethical Impacts
- Privacy and Ethics
- Breakout Discussion

AI with Ethical Impacts

Driverless cars https://theconversation.com/the-everyday-ethical-challenges-of-self-driving-cars-92710

Lethal autonomous weapons systems https://www.hrw.org/report/2018/08/21/heed-call/moral-and-legal-imperative-ban-killer-robots

Information Technology Erodes Privacy

Computers, databases, and Internet enable ever-improving information

- collection
- exchange
- combination
- distribution

Easier than ever to get information about others, including total strangers

Scott McNealy: “You have zero privacy anyway. Get over it.”

Is privacy important? If so, can we protect it?

Defining Privacy

Privacy related to notion of access

Access

- Physical proximity to a person
- Knowledge about a person

Privacy is a “zone of inaccessibility”

Privacy violations are an affront to human dignity

Too much individual privacy can harm society

Where to draw the line?

Harms of Privacy

Can be a cover for illegal or immoral activities

Can be a burden on the nuclear family

Can hide dysfunctional families

People on society’s fringes can be ignored

Benefits of Privacy

Necessary for each individual's growth as a unique person

Signals that individuals are responsible for themselves

Recognizes everyone's true freedom

Let's people be themselves

Allows people to shut out world so they can focus, be creative, and grow intellectually and spiritually

Fosters the development of loving, trusting, caring, intimate relationships

Is There a Natural Right to Privacy? (1 of 3)

Argument in favor

Right to privacy may have grown out of property rights

- Europeans have historically viewed the home as a sanctuary
- English common law tradition: “A man’s home is his castle”
- Coercive Acts (1773) led to 3rd Amendment to US Constitution: “**No soldier shall, in time of peace be quartered in any house, without the consent of the Owner, nor in time of war, but in a manner to be prescribed by law.”**

Warren and Brandeis

- Warren shocked at newspaper coverage of daughter’s wedding
- “The Right to Privacy” published in 1890
- Defined privacy as “the right to be let alone”
- Right to privacy now recognized in courts across America

“Right to Be Let Alone” (Warren and Brandeis)

Warren and Brandeis argued that the legal system should protect people’s “right to be let alone.” (PhamousFotos/Splash News/Newscom)

Is There a Natural Right to Privacy? (2 of 3)

Nobody seems to know what privacy is

Problems with defining privacy as “the right to be let alone”

- On the one hand, definition is too narrow – doesn’t include covert spying
- On the other hand, definition is too broad – does include assault

Whenever a right to privacy is violated, another right is violated as well

Therefore, no need to define privacy or privacy rights precisely

Privacy is not a **natural** right, but it is a **prudential** right

Rational people agree to recognize some privacy rights because granting these rights benefits society

Privacy and Trust

Perhaps modern life is actually more private than life centuries ago

- Most people don’t live with extended families
- Automobile allows us to travel alone
- Television v. public entertainment

Challenge: we now live among strangers

Remedy: establishing reputations

- Ordeal, such as lie detector test or drug test
- Credential, such as driver’s license, key, ID card, college degree

Establishing reputation is done at the cost of reducing privacy

Case Study: New Parents

Sullivans have a baby girl

Both work; they are concerned about performance of full-time nanny

Purchase program that allows monitoring through laptop’s camera placed in family room

They do not inform nanny she is being monitored

Rule Utilitarian Evaluation

If everyone monitored nannies, it would not remain a secret for long

Consequences

- Nannies would be on best behavior in front of camera
- Might reduce child abuse and parents' peace of mind
- Would also increase stress and reduce job satisfaction of childcare providers
- Might result in higher turnover rate and less experienced pool of nannies, who would provide lower-quality care

Harms appear greater than benefits, so we conclude action was wrong

Social Contract Theory Evaluation

It is reasonable for society to give people privacy in their own homes

Nanny has a reasonable expectation that her interactions with baby inside home are private

Sullivan’s decision to secretly monitor the nanny is wrong because it violates her privacy

Kantian Evaluation

Imagine rule, “An employer may secretly monitor the work of an employee who works with vulnerable people”

If universalized, there would be no expectation of privacy by employees, so secret monitoring would be impossible

Proposed rule is self-defeating, so it is wrong for Sullivans to act according to the rule

Virtue Ethics Evaluation

Sullivans are responsible for well-being of their daughter

Chose nanny through concern for baby: characteristic of good parents

Daughter is truly defenseless, unable to communicate with them

Decision to monitor can be viewed as characteristic of good parents

Would also expect them to cease monitoring once assured nanny is doing well

Public Records

Public record: information about an incident or action reported to a government agency for purpose of informing the public

Examples: birth certificates, marriage licenses, motor vehicle records, criminal records, deeds to property

Computerized databases and Internet have made public records much easier to access

Information Held by Private Organizations

Credit card purchases

Purchases made with loyalty cards

Voluntary disclosures

Posts to social network sites

Data Gathering and Privacy Implications

Facebook tags

Enhanced 911 services

Rewards or loyalty programs

Body scanners

RFID tags

Implanted chips

Mobile apps

Facebook Login

OnStar

Automobile “black boxes”

Medical records

Digital video recorders

Cookies

Facebook Tags

Tag: Label identifying a person in a photo

Facebook allows users to tag people who are on their list of friends

About 100 million tags added per day in Facebook

Facebook uses facial recognition to suggest name of friend appearing in photo

Does this feature increase risk of improper tagging?

Enhanced 911 Services

Cell phone providers in United States required to track locations of active cell phones to within 100 meters

Allows emergency response teams to reach people in distress

What if this information is sold or shared?

Rewards or Loyalty Programs

Shoppers who belong to store’s rewards program can save money on many of their purchases

Computers use information about buying habits to provide personalized service

- ShopRite computerized shopping carts with pop-up ads

Do card users pay less, or do non-users get overcharged?

Body Scanners

Some department stores have 3-D body scanners

Computer can use this information to recommend clothes

Scans can also be used to produce custom-made clothing

RFID Tags

RFID: Radio frequency identification

An RFID tag is a tiny wireless transmitter

Manufacturers are replacing bar codes with RFID tags

- Contain more information
- Can be scanned more easily

If tag cannot be removed or disabled, it becomes a tracking device

Implanted Chips

Taiwan: Every domesticated dog must have an implanted microchip

- Size of a grain of rice; implanted into ear
- Chip contains name, address of owner
- Allows lost dogs to be returned to owners

RFID tags approved for use in humans

- Can be used to store medical information
- Can be used as a “debit card”

Mobile Apps

Many apps on Android smartphones and iPhones collect location information and sell it to advertisers and data brokers

- Angry Birds
- Brightest Flashlight

Flurry: a company specializing in analyzing data collected from mobile apps

- Has access to data from > 500,000 apps

Facebook Login

Allows people to login to Web sites or apps using their Facebook credentials

App’s developer has permission to access information from person’s Facebook profile: name, location, email address, and friends list

OnStar

OnStar manufactures communication system incorporated into rear-view mirror

Emergency, security, navigation, and diagnostics services provided subscribers

Two-way communication and GPS

Automatic communication when airbags deploy

Service center can even disable gas pedal

Automobile “Black Boxes”

Modern automobiles come equipped with a “black box”

Maintains data for five seconds:

- Speed of car
- Amount of pressure being put on brake pedal
- Seat belt status

After an accident, investigators can retrieve and gather information from “black box”

Medical Records

Advantages of changing from paper-based to electronic medical records

Quicker and cheaper for information to be shared among caregivers

- Lower medical costs
- Improve quality of medical care

Once information in a database, more difficult to control how it is disseminated

Digital Video Recorders

TiVo service allows subscribers to record programs and watch them later

TiVo collects detailed information about viewing habits of its subscribers

Data collected second by second, making it valuable to advertisers and others interested in knowing viewing habits

Cookies

Cookie: File placed on computer’s hard drive by a Web server

Contains information about visits to a Web site

Allows Web sites to provide personalized services

Put on hard drive without user’s permission

You can set Web browser to alert you to new cookies or to block cookies entirely

General Data Protection Regulation

General Data Protection Regulation (GDPR): set of rules governing collection of information from citizens of European Union

Requires companies to…

- Disclose information they are seeking to collect
- Disclose why they are collecting it
- Get permission before collecting it

Responding to GDPR, most large American companies are adopting new privacy guidelines

- Web-site banners informing users, asking for consent

Data Mining Defined

Searching records in one or more databases, looking for patterns or relationships

Can be used to create profiles of individuals

Allows companies to build more personal relationships with customers

Google's Personalized Search

Secondary use: Information collected for one purpose use for another purpose

Google keeps track of your search queries and Web pages you have visited

- It uses this information to infer your interests and determine which pages to return
- Example: "bass" could refer to fishing or music

Also used by retailers for direct marketing

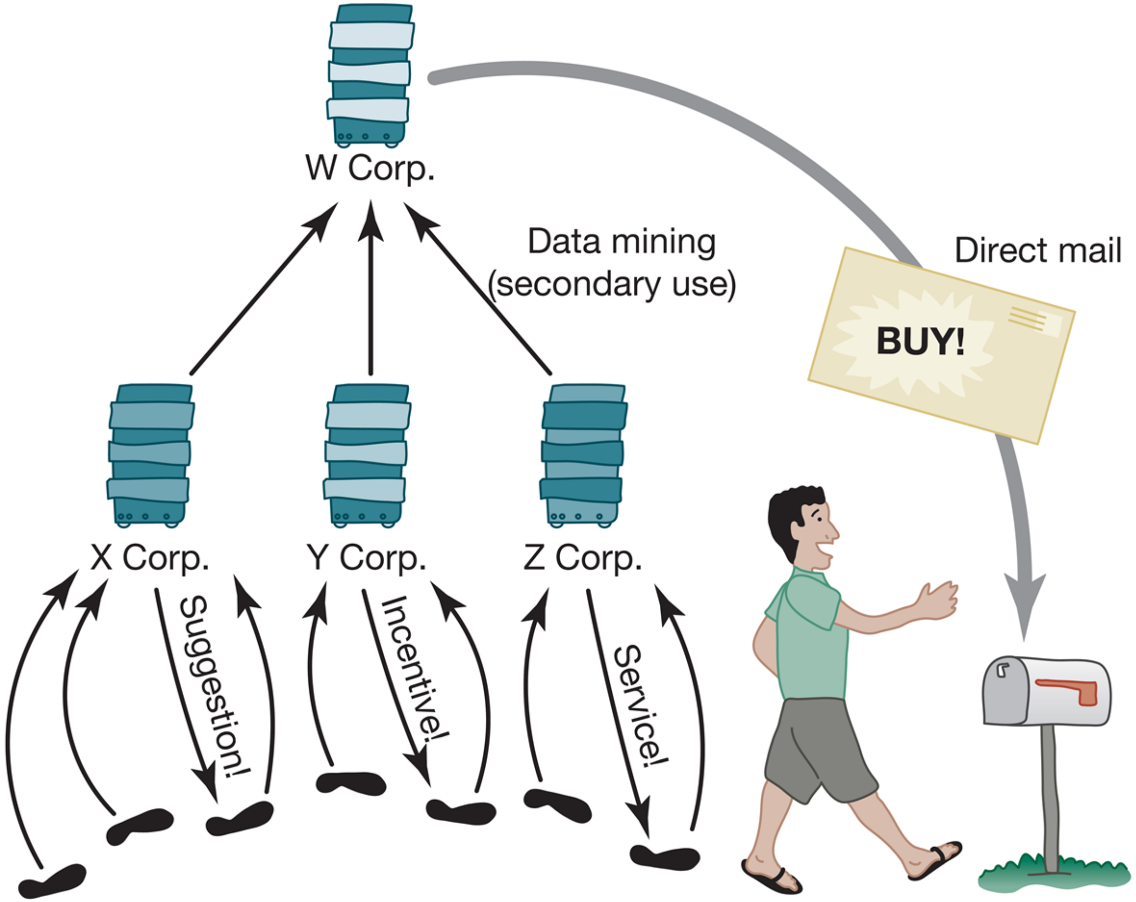
Limiting Information Google Saves

You can limit amount of information Google saves about your activities

Privacy Checkup lets you pause collection of personal information

- Search queries and other Google activity
- Location information collected from signed-in devices
  - Where you have gone
  - How often you have gone there
  - How long you have stayed
  - Customary routes of travel
- Contact and calendar information
- Recordings of your voice and accompanying audio
- YouTube search queries
- YouTube videos you have watched

Secondary Uses of Information



Collaborative Filtering

Form of data mining

Analyze information about preferences of large number of people to predict what one person may prefer

- Explicit method: ask people to rank preferences
- Implicit method: keep track of purchases

Used by online retailers and movie sites

Ownership of Transaction Information

Who controls transaction information?

- Buyer?
- Seller?
- Both?

Opt-in: Consumer must explicitly give permission before the organization can share info

Opt-out: Organization can share info until consumer explicitly forbid it

Opt-in is a barrier for new businesses, so direct marketing organizations prefer opt-out

“Target”-ing Pregnant Women

Most people keep shopping at the same stores, but new parents have malleable shopping habits

Targeting pregnant women, a good way to attract new customers

Target did data mining to predict customers in second trimester of pregnancy

- Large amounts of unscented lotion, extra-large bags of cotton balls, nutritional supplements

Mailings included offers for unrelated items with offers for diapers, baby clothes, etc.

Credit Reports

Example of how information about customers can itself become a commodity

Credit bureaus

- Keep track of an individual’s assets, debts, and history of paying bills and repaying loans
- Sell credit reports to banks, credit card companies, and other potential lenders

System gives you more choices in where to borrow money

Poor credit can hurt employment prospects

Targeted Direct Mail

Businesses mail advertisements only to those most likely to purchase products

Data brokers provide customized mailing lists created for information gathered online and offline

Example of making inferences for targeted direct mail

- Shopping for clothes online + frequent fast-food dining + subscribing to premium cable TV channels ⟶ more likely to be obese

Two shoppers visiting same site may pay different prices based on inferences about their relative affluence

Microtargeting

Political campaigns determine voters most likely to support particular candidates

- Voter registration
- Voting frequency
- Consumer data
- GIS data

Target direct mailings, emails, text messages, home visits to most likely supporters

Social Network Analysis

Collect information from social networks to inform decisions

Companies offers special promotions to “influencers”

Police use Facebook and Twitter posts to deploy officers on big party nights

Banks combine social network data with credit reports to determine creditworthiness

Controlling Your Facebook Info (1 of 2)

You can change your Facebook settings to minimize who can see what you’re doing

Privacy settings

- Who can see your friends list?
- Who can see your future posts?
- Who can look you up using your email address?
- Who can look you up using your phone number?
- Do you want search engines to link to your profile?
- Limit audience for posts you’ve shared?

Controlling Your Facebook Info (2 of 2)

Timeline and Tagging

- Who sees tag suggestions when photos look like you?
- Review posts you're tagged in?
- Review tags people add to your posts?

Location History

Ads – Based on

- Relationship status
- Employer
- Job title
- Education
- Data from partner
- Activity on Facebook Company Products
- Social actions

Netflix Prize

Netflix offered $1 million prize to any group that could come up with a significantly better algorithm for predicting user ratings (2006)

Released more than 100 million movie ratings from a half million customers

- Stripped ratings of private information

Researchers demonstrated that ratings not truly anonymous if a little more information from individuals was available

U.S. Federal Trade Commission complaint and lawsuit

Netflix canceled sequel to Netflix Prize (2010)

AOL Search Dataset

AOL researcher Dr. Chowdhury posted three months’ worth of user queries from 650,000 users (2006)

No names used; random integers used to label all queries from particular users

Researchers identified some users from queries; e.g., many people performed searches on their own names

New York Times investigation led to public outcry

AOL took down dataset, but already copied and reposted

AOL fired Dr. Chowdhury and his supervisor

Marketplace: Households

Lotus Development Corporation developed CD with information on 120 million Americans

Planned to sell CD to small businesses that wanted to create mailing lists based on various criteria, such as household income

More than 30,000 consumers complained to Lotus about invasion of privacy

Lotus dropped plans to sell CD

Facebook Beacon

2007: Facebook announced Beacon, a targeted advertising device

- Facebook user makes purchase
- Facebook broadcasts purchase to user’s friends
- Based on opt-out policy: users enrolled unless explicitly asked to be excluded

A significant source of advertising revenue for Facebook

MoveOn.org led online campaign lobbying Facebook to switch to an opt-in policy

Mark Zuckerberg apologized, and Facebook switched to an opt-in policy

Malls Track Shoppers' Cell Phones

In 2011 two malls recorded movement of shopper by tracking locations of cell phones

- How much time people spend in each store?
- Do people who shop at X also shop at Y?
- Are there unpopular areas of mall?

Small signs informed shoppers of study

After protest, mall quickly halted study

iPhone Apps Upload Address Books

In 2012 a programmer discovered Path was uploading iPhone address books without permission

Internet community pointed out this practice violated Apple’s guidelines

CEO of Path apologized; app rewritten

Twitter, Foursquare, and Instagram also implicated for same practice

Instagram's Proposed Change to Terms of Service

Late 2012: Instagram announced changes

- Privacy policy
- Terms of service

Legal experts: Instagram and Facebook would have right to use photos in ads without permission

Instagram CEO: New policy misunderstood

Changed advertising section of terms of service agreement back to original version

Behavioral Big Data (BBD)

Large, rich, multi-dimensional datasets on human behaviors, actions, and interactions.

Available to companies, governments, and researchers; such data come with opportunities and drawbacks.

Special because they capture individuals’ actions and interactions, self-reported opinions, thoughts, feelings.

It involve issues like intention, deception, emotion, reciprocation, herding, and other social and human aspects.

What do you think?

Question: if people know that data is being collected, do they change their behavior?

BBD are different from medical data. While both involve human subjects, medical data focuses on physical measurements. Participants in clinical trials know they are participating and have a vested interest. However, subjects of behavioral experiments with BBD may not know and typically receive no direct value.

Additional Resources on BBD

- Analyzing Behavioral Big Data: Methodological, Practical, Ethical, and Moral Issues
- Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide
- What Amazon Echo and Google Home Do With Your Voice Data
- https://www.sporttechie.com/toshiba-using-facial-images-fans-staples-center/

Data Privacy

- "The relationship between the collection and dissemination of data, technology, the public expectation of privacy, legal, and political issues surrounding them
- "Prevention of anything that can be learned about a respondent from the statistical database without access to the database
- An analysis of data preserves data privacy if: (1) you learn something useful from the analysis, and (2) the analysis does not violate the privacy of any individual.

Additional Resources on Privacy

https://www.linkedin.com/pulse/entering-new-age-data-privacy-us-learning-from-gdpr-daniel-solove/

https://pubsonline.informs.org/do/10.1287/orms.2020.02.37p/full/

Data Privacy vs. Data Security

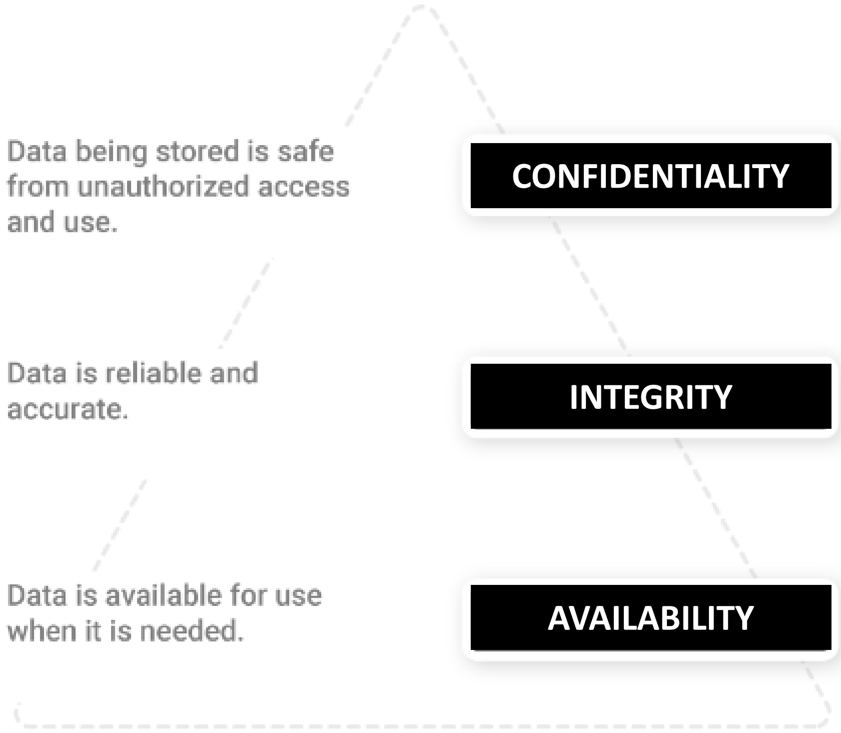
Data security is concerned with who can touch the data in one of two ways:

- (1) ensuring only appropriate people can *view* the data (confidentiality), and
- (2) ensuring that only appropriate people can *modify* the data (integrity).

Data privacy is concerned with what can be learned from the data, i.e., its *information content*.

"You can't have privacy without security, but you can have security without privacy."

Data Security



Data being stored is safe from unauthorized access and use.

CONFIDENTIALITY

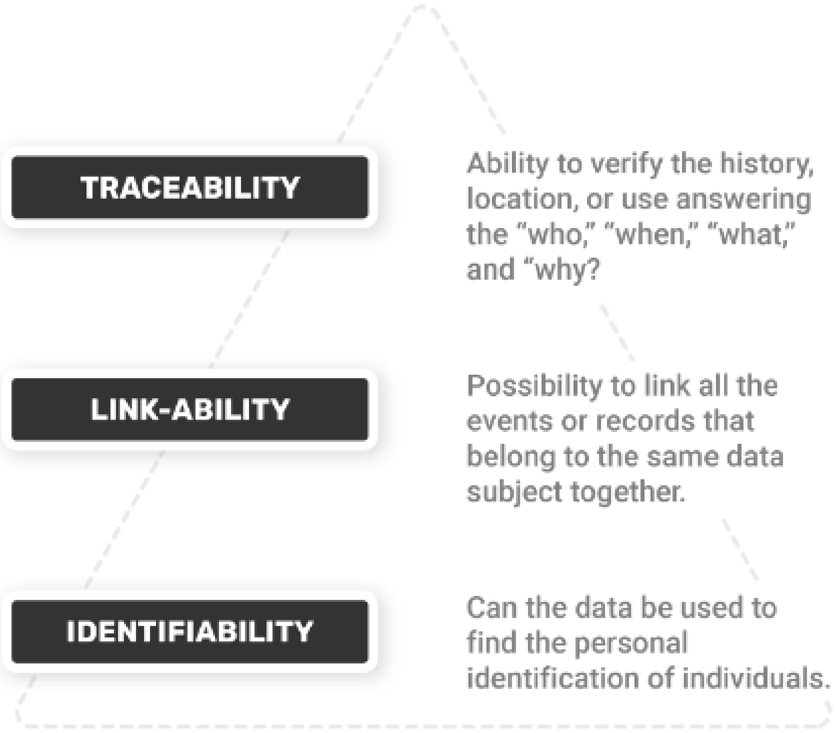
Data is reliable and accurate.

INTEGRITY

Data is available for use when it is needed.

AVAILABILITY

Data Privacy



TRACEABILITY

Ability to verify the history, location, or use answering the “who,” “when,” “what,” and “why?

LINK-ABILITY

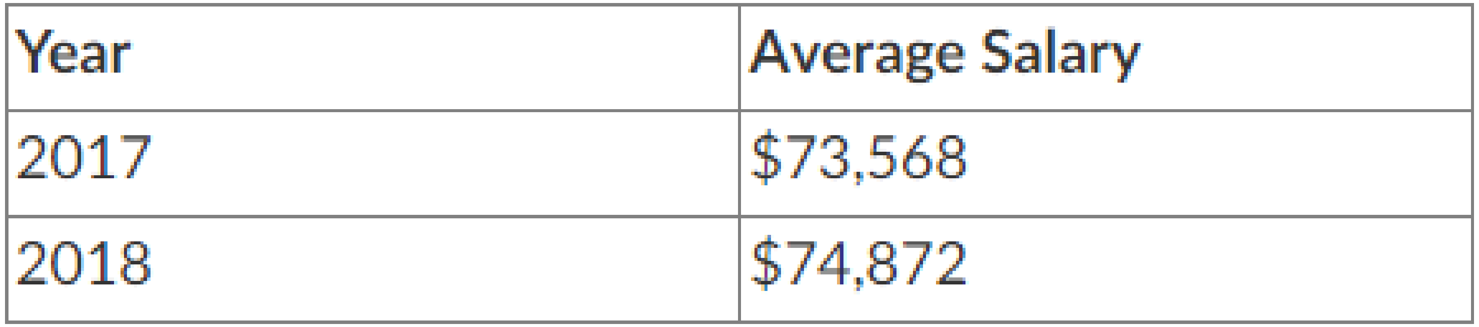
Possibility to link all the events or records that belong to the same data subject together.

IDENTIFIABILITY

Can the data be used to find the personal identification of individuals.

Methods of Protecting Privacy

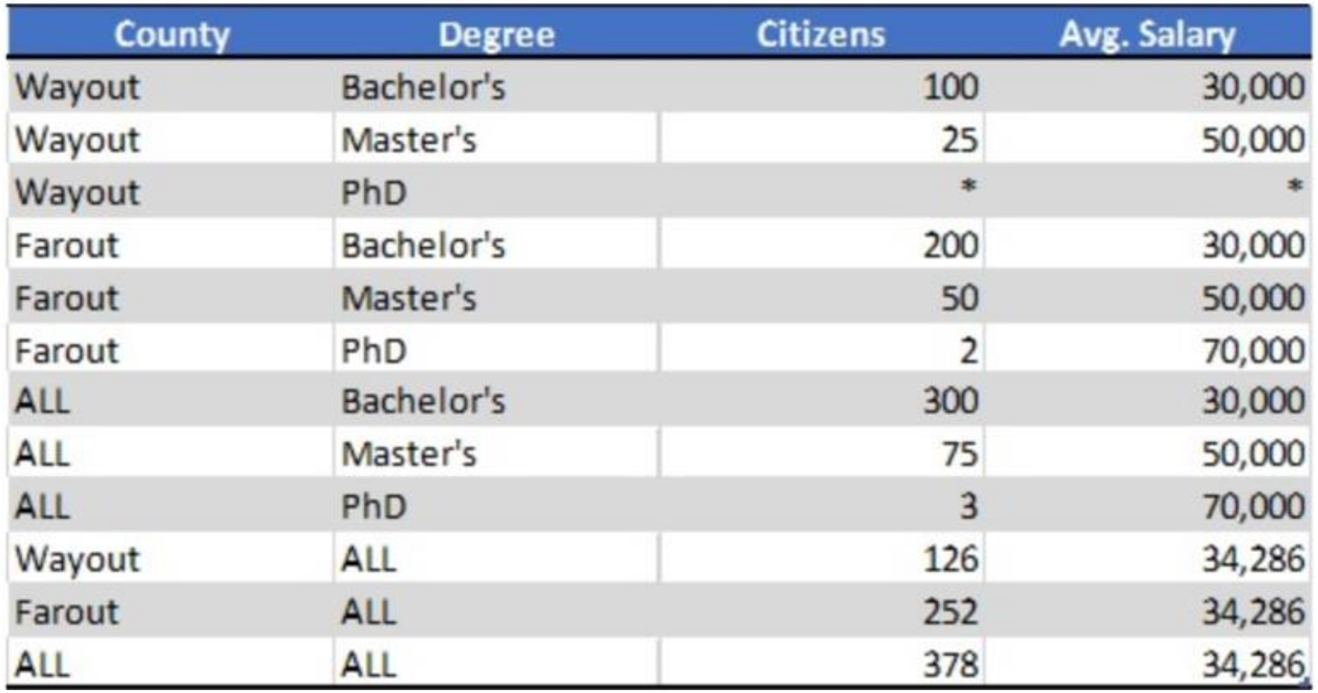
Aggregation



Year	Average Salary
2017	$73,568
2018	$74,872

Methods of Protecting Privacy

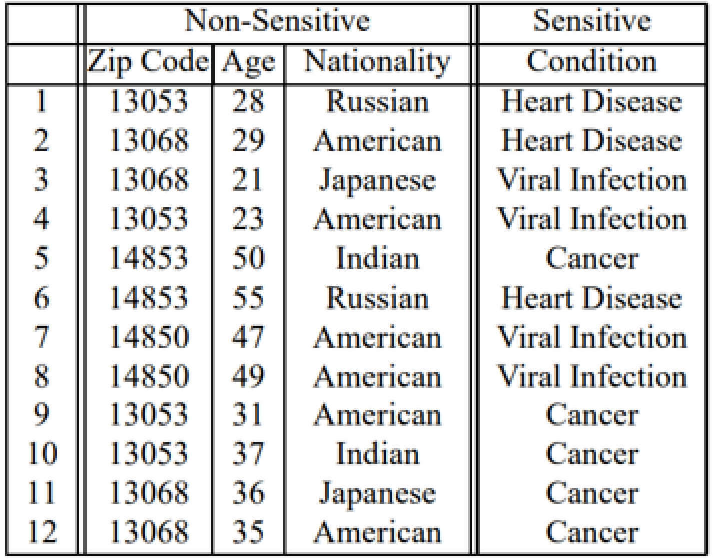
Cell Suppression



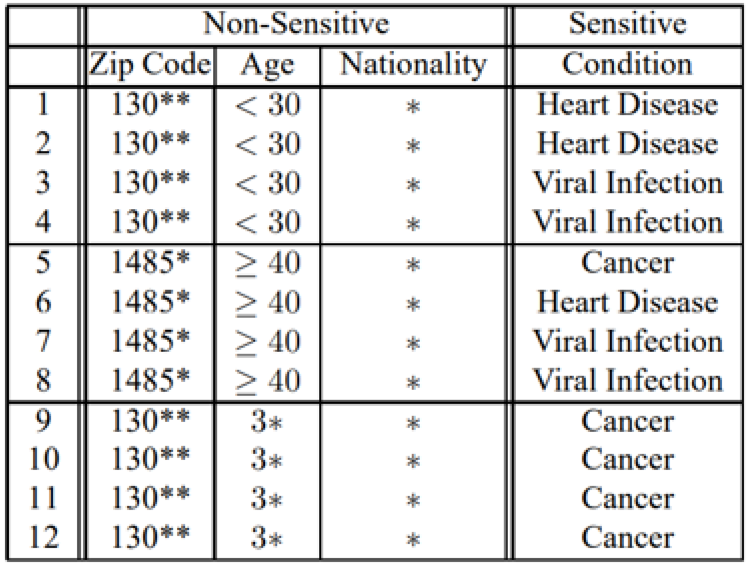
County	Degree	Citizens	Avg. Salary
Wayout	Bachelor's	100	30,000
Wayout	Master's	25	50,000
Wayout	PhD	*	*
Farout	Bachelor's	200	30,000
Farout	Master's	50	50,000
Farout	PhD	2	70,000
ALL	Bachelor's	300	30,000
ALL	Master's	75	50,000
ALL	PhD	3	70,000
Wayout	ALL	126	34,286
Farout	ALL	252	34,286
ALL	ALL	378	34,286

Methods of Protecting Privacy

k-Anonymity and l-Diversity



	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer



	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Differential Privacy

Differential privacy is a rigorous mathematical definition of privacy.

In the simplest setting, consider an algorithm that analyzes a dataset and computes statistics about it (such as the data's mean, variance, median, mode, etc.).

Such an algorithm is said to be differentially private if by looking at the output, one cannot tell whether any individual's data was included in the original dataset or not.

In other words, the guarantee of a differentially private algorithm is that its behavior hardly changes when a single individual joins or leaves the dataset.

Random noise (epsilon)

Additional Resource on Differential Privacy

https://www.wired.com/story/apple-differential-privacy-shortcomings/

https://www.oreilly.com/radar/podcast/how-privacy-preserving-techniques-can-lead-to-more-robust-machine-learning-models/