Ethics & Privacy

Plan for Today

- Institutional Review Board (IRB)
- Readings for this Week
  - "Adaptive Privacy-Preserving Visualization Using Parallel Coordinates"
  - "Agile Ethics for Massified Research and Visualization"
- Tangentially Related Graphics/Vision Topics
- Assignment 7: Final Project Ideas
- Reading for Friday
• World War II: German physicians conducted medical experiments on concentration camp prisoners without their consent. Tested blood clotting (shooting them), vaccines (infesting them), effectiveness of poison bullets, and effects of high altitude and low oxygen.

• In the 1950’s, thalidomide given to pregnant women to help with sleep and nausea, but they did not know it was experimental nor did they given consent.

• Tuskegee, Alabama (1940s-1970s): Low-income African-American males with high incidence of syphilis infection were given free medical examinations, but not told about their disease, and researchers intervened to prevent treatment.

• 1961, Milgram obedience study (the shock machine): lack of proper attention to debriefing, didn’t reveal the purpose of the study, didn’t comfort subjects ethical qualms about having inflicted pain on a fellow human, didn’t offer his participants an opportunity to opt out of the study.

• Zimbardo's prison experiment (Stanford): The study did meet the criteria of his IRB in 1973!

• In the 70’s various federal regulations established IRB at all research institutions.

---

**Institutional Review Board (IRB)**

- Privacy, Confidentiality, Anonymity, and Informed Consent
- Reduce risk (physical/mental/privacy) to the participants engaged in research
Institutional Review Board (IRB)

Rensselaer Polytechnic Institute is committed to protecting the rights and welfare of human subjects of research conducted on the campus or sponsored by the Institute.

Rensselaer subscribes to the basic ethical principles that underlie the conduct of biomedical and behavioral research involving human subjects as set forth in the Belmont Report, and in accordance with Title 45, Code of Federal Regulations, Part 46.

The Institutional Review Board (IRB) has the responsibility and authority to review, approve, disapprove, or require changes in research activities involving human subjects. This policy applies to all faculty, staff, and student projects, regardless of whether the project is funded externally, internally, or receives no funding support.

Researchers should refer to Rensselaer’s Guidelines for Human Subjects Research to determine whether or not their research is indeed human subjects research, and/or if their research satisfies the requirements for expedited review by the IRB.

IRB Training Requirements

As federally mandated and required by the Rensselaer IRB, all investigators must complete a self-study course in human subject protection via the CITI Training Program. Each investigator on a research project involving human subjects is required to provide a signed CITI training certificate before proceeding in the project.

https://oasis.cs.rpi.edu/
Spa:ally Augmented Reality (SAR) Projection

- Camera detects design geometry
- 6 projectors augment design
- Design sketched with foam-core walls

Participation is voluntary. We anticipate no risk or discomfort beyond routine use of a computer and the Internet.

Construction of a model averages 5-10 minutes, depending on the complexity and depth of analysis. Your models and written feedback will be collected for use in future publications and the improvement of our tool.

No personal information is collected during the registration process. If you choose to provide an email address, researchers may contact you with optional follow-up questions. We will not share this email with anyone.

There is no remuneration offered for participation in this study. You retain ownership of the architectural models designed in our system.

For questions or concerns please contact:

Barbara Cutler, cutler@cs.rpi.edu,
Phone: 518-276-2274
Rensselaer Polytechnic Institute
Max Espinosa, espinosa@epi.rpi.edu,
Rensselaer Polytechnic Institute

Chair, Institutional Review Board
Rensselaer Polytechnic Institute
CII 5015110 8th Street
Troy, NY 12180
(518) 276-4873
Plan for Today

- Institutional Review Board (IRB)
- Readings for this Week
  - "Adaptive Privacy-Preserving Visualization Using Parallel Coordinates"
  - "Agile Ethics for Massified Research and Visualization"
- Tangentially Related Graphics/Vision Topics
- Assignment 7: Final Project Ideas
- Reading for Friday
Different purposes for “Visualization”:
- Convey information or art? Is misleading necessarily bad when visualizing for art purposes?
  Context matters.
- Find the appropriate balance between art (eye catching & memorable) and scientific accurate & honest data presentation.

Some visualization techniques are already quite lossy (e.g., pie chart) or just confusing (e.g., parallel coordinates?) and obfuscation may not be necessary
- Making false connections
- Limit visualization to 500 pixels to thwart attacks(?)
- Rounding pixel coordinates

Impractical in the real world?
So much effort to make a visualization less useful.
(Why make the visualization at all? Who is the target audience of this visualization?)

Need to investigate reliability of data before reporting it
K-anonymity & l-diversity
(Trusted) server does clustering & only sends clusters to (untrusted) client
Writing
- Nice to read about the process and ideas for techniques that ultimately didn’t work
- Would have been better written/easier to read if they had stated upfront what they wanted to accomplish
- Diagrams & charts not well explained

Could be further secured:
- Don’t place unclustered data on the public facing machine
- Don’t ever re-cluster the data (prevent clustering attacks)

Client-server model: Computation & networking costs?
How well does this work for small datasets?
Hadn’t considered visualization as a means of data breach, reassuring that people are researching the problem.
How much accuracy in analysis do we lose?
- Obsuring & blurring data (opposite of our usual focus on clarity & accuracy!)

What about more sophisticated attacks? ...
- Maybe this isn’t 100% secure, but it’s important to do something!

Would like a user study comparing their Privacy-Preserving Parallel Coordinates to original full data Parallel Coordinates.
Quasi-Identifiers (is an anonymous form really anonymous?)
Privacy policies that declare how they can and will share your personal data...

- Privacy, confidentiality, anonymity, and informed consent
  - Minimize risk to participants
  - Observational study does not require consent
  - (current) IRB process does not/cannot work at massive scale
  - A single person can more easily do what used to take a team of researchers much more time.
- Agile ethics: too high level to be enforceable?
- Even though individuals made this data available, it is researchers responsibility to not put them in danger
- Not showing individuals, only general trends
- If the research team’s information is similarly public/vulnerable, its ok?
- Geometry for twitter
- Public-private greyscale
- “Now that you’ve read this... Just be more careful, ok?”
- Writing
  - Confusing paper organization
  - Low resolution images
  - Unconventional acronyms
  - Footnotes at end of paper (prefer at end of each page)

- Jargon-y
- Unnecessarily lengthy?
- Websurfing is dangerous
- Twitter is scary (amount of personal data available surprising)
  - Users can (now?) disable location tracking
  - Are Twitter “protected” accounts new?
  - Read the fine print before sharing your data!
- TimeRose visualization is new to me
- Scattered topics
- Fitness tracking applications offer to post your morning run on facebook are dangerous
- Proposed solutions are too idealistic?
Society of Professional Journalists' Code of Ethics

• Is this actually used? Impractical (won’t have a long career). Seems like many media outlets do the opposite
  – Examples?

• First amendment, expose negative qualities/actions of people in power

• Create emotions that can influence viewers – Tufte says to avoid garnishes (for this reason?)

http://visual.ly/about/ethics

As an organization that both practices and recognizes quality data-journalism, Visual.ly subscribes to the code of ethics of the Society of Professional Journalists and agrees to abide by all of its principles.
We also agree to the following principles to support data analysis and visualization:
Data will be accurate and verifiable - Visual.ly will not "lie with statistics."
Proper Sourcing & Attribution - Visual.ly will always give credit where due and will do its own reporting.
Best Practices in Visual Representation - Visual.ly will not exploit idiosyncrasies of the human visual system to exaggerate or misrepresent data.
Most succinctly stated, Visual.ly's policy is one that encompasses accuracy, honesty, and transparency.
While Visual.ly will do our best to promote these standards, the policy applies only to the visualizations we create ourselves and those we feature as staff picks, not to those uploaded by members of the community.
Visual.ly's Code of Ethics for Data Visualization Professionals

- Data analysis is important
- Too narrow focus or omission can lead to bias
- Incorrect analysis must be avoided
- Be open to criticism, learn from past work

Interesting Tidbits

- Internet is an ocean of data
- Research results poured back into ocean of data
- Surveillance: Shopping malls are private spaces, but made to feel like public spaces
- File/log planned data collections in advance (pre-planning required, data more precious)
- Researchers should make themselves equally public
Informed Consent

• Do you carefully read every document you sign? Every “agree to terms” button you click?
• Data can be taken out of context, used in ways other than intended
• Previously: required a team of researchers to gather data
• Now: a single person can do it alone - lost informal peer consultation of ethics concerns

• Does an individual student’s grade rise over time (repeated submissions)?
• Do students who start submitting earlier in the week have a higher final grade for that homework?
• Do students who submit more times get higher grades?
What is sufficient to anonymize data?

• Remove explicit identifiers
• Small datasets cause problems
• Quasi identifiers
• What are the sensitive attributes
• Sanitize data on the fly, constraining the interaction
• Assume data holder is aware of data sensitivity (and appropriately concerned)

Privacy & Visualization

• Most visualization computation assumes unrestricted access to data
• How do we do this computation with partial information?
• How do you design hardware/software system to ensure data security?
• Who would potentially benefit from access to this data? (Why is this a grey area?)
  – Scientific discovery
  – Improve healthcare

• What data has privacy concerns?
  – Corporate secrets
  – Health records
  – Personal finances
  – Personal location

Risks to users/participants?

• Quasi-identifiers & Doxing/doxxing (document tracing):
  “Internet-based practice of researching and publishing personally identifiable information about an individual. The methods employed in pursuit of this information range from searching publicly available databases and social media websites like Facebook, to hacking, and social engineering. It is closely related to cyber-vigilantism, hacktivism and cyber-bullying.” (definition from Wikipedia)

• If you’re not interesting (now or ever in the future), you probably have privacy?
Facebook generation of oversharers?

• It’s your choice to share or not use the service at all (Is this true?)
• Generation that believes privacy doesn’t/can’t exist for anyone. Is there now or will there be regret for what has been shared?
• Do we have an obligation to educate young internet users on (lack of) online privacy? On how easy it is to connect the dots even without usernames or obvious identifiers?

When are MIT students asleep?
Leon Lin and Aaron Scheinberg

Health Insurance Portability and Accountability Act (HIPAA)

- Long Title: “An Act To amend the Internal Revenue Code of 1986 to improve portability and continuity of health insurance coverage in the group and individual markets, to combat waste, fraud, and abuse in health insurance and health care delivery, to promote the use of medical savings accounts, to improve access to long-term care services and coverage, to simplify the administration of health insurance, and for other purposes.”
- Unintended negative outcomes
  - Reduced retrospective chart-based research (responses dropped from 96% to 34% in one study on heart-attack followup surveys)
  - Legalistic details on privacy preservation techniques has made informed consent forms even longer and less user-friendly
  - Stiff penalties for violations, lead doctors to withhold information (even sometimes from people who have rights to see it!)
  - Expensive to implement
  - Requires training healthcare providers

Plan for Today

- Institutional Review Board (IRB)
- Readings for this Week
  - "Adaptive Privacy-Preserving Visualization Using Parallel Coordinates”
  - "Agile Ethics for Massified Research and Visualization”
- Tangentially Related Graphics/Vision Topics
- Assignment 7: Final Project Ideas
- Readings for Friday
Assignment 7: Final Project Ideas

• ½ points of a regular 1 week assignment
• Invent 2 different Final Project Ideas
  – “Who” (audience), "why" (research question), "what" (the finished visualization)
  – One technical challenge for the project. What makes it difficult? What is a potential “risk” for completion? For example:
    • acquiring the data,
    • working with very large data,
    • implementing a new visualization design,
    • implementing a novel interaction scheme, or
    • revising the visualization design to validate your hypothesis.
  – Do you already have a partner?
• Make LMS post by Thursday
• Reply to 3 other students on LMS by Monday

Plan for Today

• Institutional Review Board (IRB)
• Readings for this Week
  – "Adapтивe Privacy-Preserving Visualization Using Parallel Coordinates”
  – "Agile Ethics for Massified Research and Visualization”
• Tangentially Related Graphics/Vision Topics
• Assignment 7: Final Project Ideas
• Readings for Friday
Reading for Friday: *(choose one)*

- "LabelMe: online image annotation and applications"  
  Torralba, Russell, & Yuen, IEEE, 2010

- "QSplat: A Multiresolution Point Rendering System for Large Meshes",  
  Rusinkiewicz & Levoy,  
  SIGGRAPH 2000