Human Centered Data Science

DATA 512 — Jonathan T. Morgan & Os Keyes

Ethics & Privacy | Week 2 | October 04, 2018

Introduction

Overview of the day

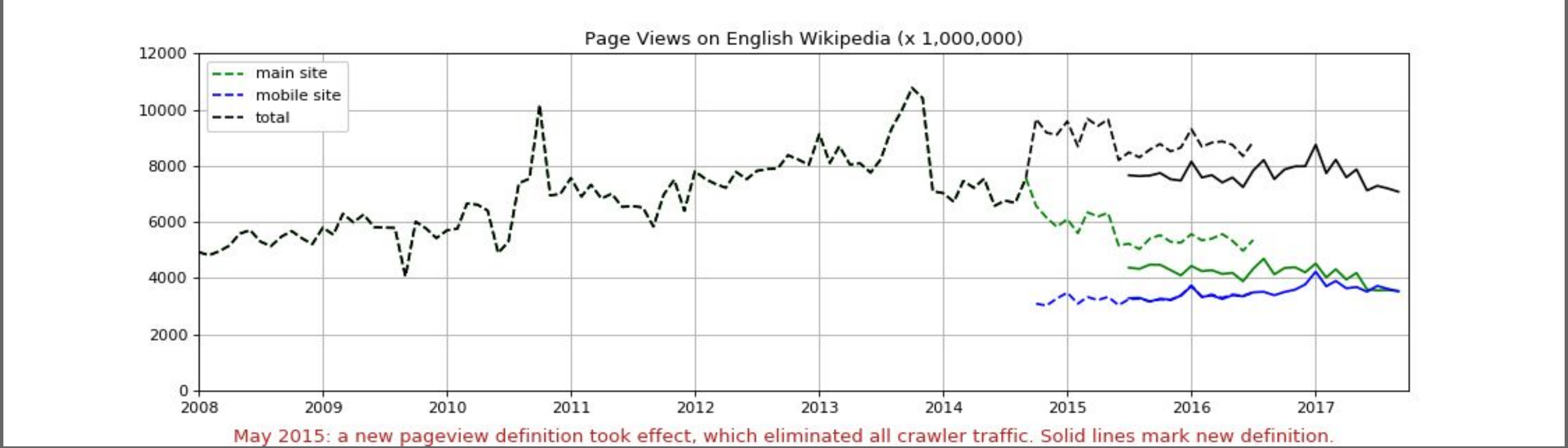
- Assignment 1 Intro (Jonathan)
- A short history of research ethics & regulation in the United States (Jonathan)
- Guest lecture 1 (Javier & Mark)
- Guest lecture 2 (Javier)
- Contextual Integrity in Data Science (Os)

First assignment assigned!

A1: Data curation

5 points, due October 18

https://wiki.communitydata.cc/Human_Centered_Data_Science_(Fall_2018)/
Assignments#A1:_Data_curation



Goal: make a graph like this one, using the Wikimedia REST API as a data source, and document your process and outcomes according to best practices for open, reproducible research.

We will go through this assignment in detail next week, but you're encouraged to get started (more at end of class)

Graph: Rex Thompson, 2017. MIT License. Used with permission.

A short history of research, ethics & regulations

(in the twentieth century, in the United States, abridged)

Why is this relevant?

A lot of the most notable regulatory advances in the 20th century addressed issues that are just now being raised in a data science context, such as...

- **Consumer protections**
- **Human subjects protections**
- **Labor practices**
- **Individual liberty**
- **Right to privacy**
- **Social equity**

Why is this relevant?

A lot of the most notable regulatory advances in the 20th century addressed issues that are just now being raised in a data science context, such as...

- **Consumer protections:** GDPR
- **Human subjects protections:** Online experiments in social media
- **Labor practices:** Algorithmic work allocation, gig economy, human moderators
- **Individual liberty:** Automated content moderation, platform speech policies
- **Right to privacy:** online surveillance, corporate data sharing practices
- **Social equity:** algorithmic bias

Where regulations come from

Sometimes, regulations are direct responses to *perceived serious violations* of widely held *social norms*—shared expectations about how people should act in particular situations (including ethical codes).

Sometimes, regulations are ‘inspired’ by existing formal policies or practices:

- Non-binding statements of principles (Nuremburg Code)
- Policies from professional organizations (AMA Code of Ethics)
- Regulations in other jurisdictions (GDPR)

What regulations provide (ideally)

- *Legal accountability*
    - **Compliance** frameworks
    - **Penalties** for non-compliance
    - **Recourse** for those harmed
- *Social accountability*
    - **Transparency** into organizational and government processes
    - **Affirmation** of widely-held beliefs and ethical values
    - **Precedent** for future social action (and potentially, further regulation)

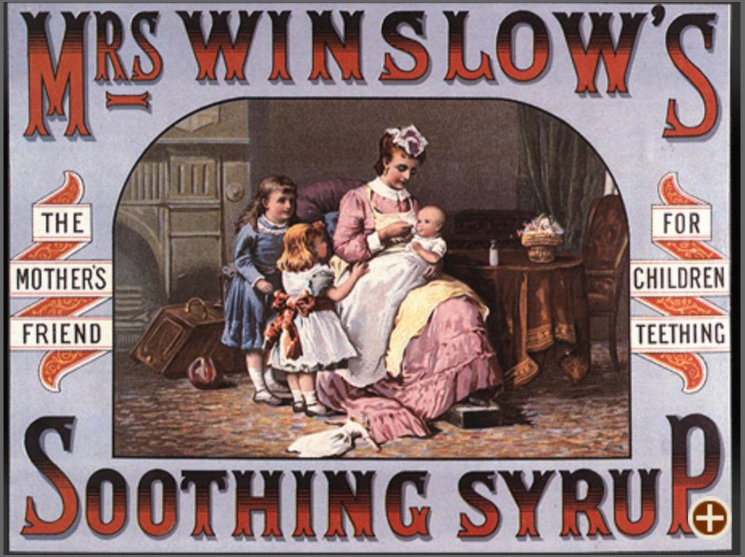
How regulations fall short

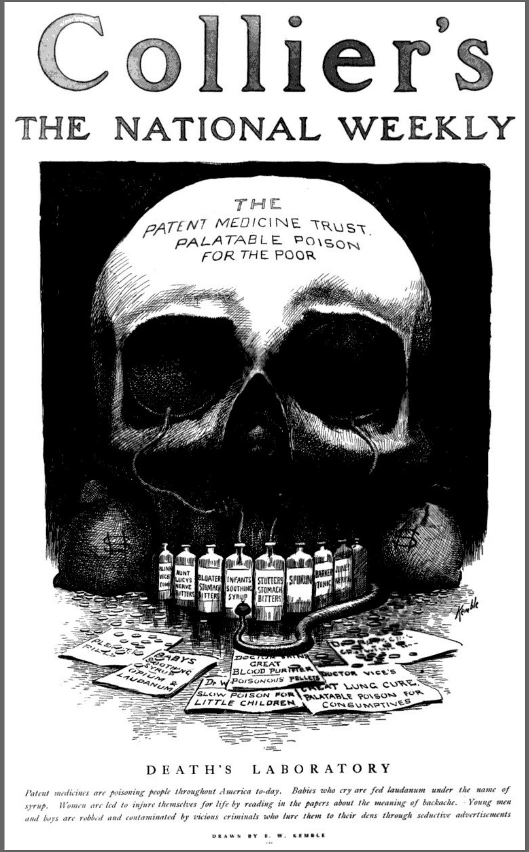
- **Scoping:** too limited, too broad
- **Design:** undue influence, under/over specification, subject matter expertise
- **Implementation:** limited jurisdiction, enforcement, and evolutionary flexibility

Case studies

Research ethics and consumer protection regulations in health and medicine

Pure Food & Drug Act (1906)





Collier's

THE NATIONAL WEEKLY

DEATH'S LABORATORY

Patent medicines are poisoning people throughout America to-day. Babies who cry are fed laudanum under the name of syrup. Women are led to injure themselves for life by reading in the papers about the meaning of backache. Young men and boys are robbed and contaminated by vicious criminals who lure them to their dens through seductive advertisements

DRAWN BY E. W. KEMBLE

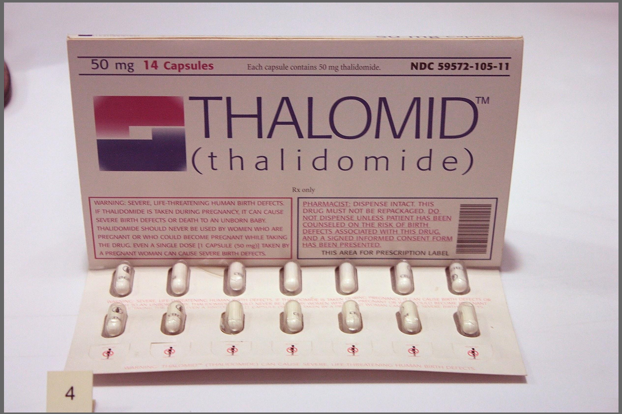
Pure Food & Drug Act (1906)

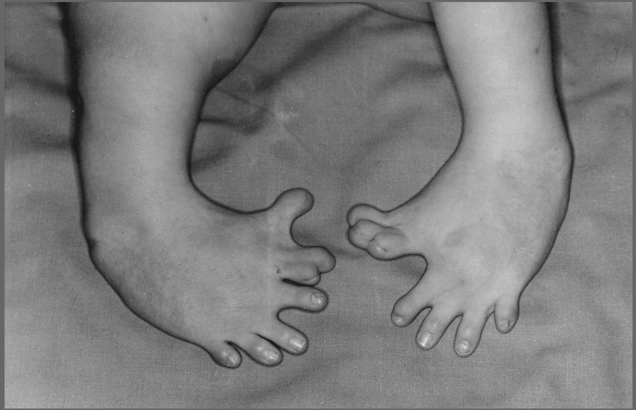
- Targeted at the patent medicine industry
- Required accurate labeling of presence and dosage of 10 “addictive” or “dangerous” ingredients in consumer products
- Forbid manufacturers making fraudulent claims about effectiveness (later struck down by Supreme Court)
- Created inspector officers and inspection process

Food, Drug and Cosmetic Act (1938)

- Spurred by mass poisoning due to drug additives in sulfidamide medication
- Expanded and clearly defined the FDA’s scope
- Required premarket safety testing for *side effects*
- Required subject matter expertise in reviewing applications for new food, drugs, ingredients, and cosmetics

Kefauver Harris Amendment (1962)


50 mg 14 Capsules
Each capsule contains 50 mg thalidomide.
NDC 59572-105-11
THALOMID™
(thalidomide)
Rx only
WARNING: SEVERE, LIFE-THREATENING HUMAN BIRTH DEFECTS. IF THALIDOMIDE IS TAKEN DURING PREGNANCY, IT CAN CAUSE SEVERE BIRTH DEFECTS OR DEATH TO AN UNBORN BABY. THALIDOMIDE SHOULD NEVER BE USED BY WOMEN WHO ARE PREGNANT OR WHO COULD BECOME PREGNANT WHILE TAKING THE DRUG. EVEN A SINGLE DOSE [1 CAPSULE (50 mg)] TAKEN BY A PREGNANT WOMAN CAN CAUSE SEVERE BIRTH DEFECTS.
PHARMACIST: DISPENSE INTACT. THIS DRUG MUST NOT BE REPACKAGED. DO NOT DISPENSE UNLESS PATIENT HAS BEEN COUNSELED ON THE RISK OF BIRTH DEFECTS ASSOCIATED WITH THIS DRUG, AND A SIGNED INFORMED CONSENT FORM HAS BEEN PRESENTED.
THIS AREA FOR PRESCRIPTION LABEL
4



Kefauver Harris Amendment (1962)

- Manufacturers must prove their drug is both safe **and effective** to gain FDA approval
- Clinical trials must include informed consent
- Side effects and efficacy information must be provided to consumers
- Limitations on claims for rebranded drugs

Tuskegee Syphilis Study (1932-72)

- Run by US Public Health Service
- 400 African-American men who had syphilis, disease progression monitored over 40 years
- Subjects not told what disease they had; not treated, even after Penicillin
- Resulted in 128 deaths, 40 sex-related transmissions of the disease, 19 children born with congenital syphilis

Problems

- Informed consent
- Medical malpractice
- Social justice
- The reliance on *abstraction*:
    - The scientists acted in ignorance of (or aware of) their own racism
    - The scientists fell into treating their research subjects as “just data”

The Belmont Report and the National Research Act (1974)

- Full title: *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, Report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*
- Spurred by outrage over Tuskegee
- Set new federal standards for clinical trials and Institutional Review Boards
- Acknowledged that research ethics were a matter of national importance

The Belmont Report: Principles

- **Respect for persons:** people should be treated as autonomous; those with diminished autonomy should be provided extra protection
- **Beneficence:** potential risks and benefits must be assessed and the risk/benefit ratio (for subjects & society) must be justified
- **Justice:** fair distribution of risk and benefit among individuals and groups

The Belmont Report

- **Informed consent**
  - *Information:* subjects must be given all relevant information
  - *Comprehension:* subjects must understand procedures, risks, benefits
  - *Voluntariness:* Subjects must be free to consent
- **Assessment of risk and benefit**
  - Nature and scope of potential risks and intended benefits
  - Systematic assessment of potential R&B
- **Selection of subjects**
  - Individual justice
  - Social justice

Limitations

- Designed primarily around the needs of medical research
- Challenges with differentiating research from practice
- Focused primarily on protecting subjects, not society
- No mechanism for prioritizing justice, beneficence, or respect for persons
- Scoped primarily to clinical trials, drug applications, and govt-funded research
- Almost 40 years old

Guest Lectures

Javier Salido and Mark van Hollebeke (Microsoft)

A Practitioner’s View of Privacy & Data Protection

Differential Privacy

In-class activity

- Pose **at least 1 question** for one or both of this week's guest lecturers
- List **at least three important/useful things you learned** from these lectures (and *why* they're important).
- The question, and each of the "important things", should be **at least 1 full sentence each.**

You are encouraged, but NOT required, to ask you question during the lecture Q&A.

This is an individual activity, due by 11:59pm tomorrow. It is worth 2 points.

See: https://wiki.communitydata.cc/HCDS_(Fall_2018)/Assignments#Weekly_in-class_activities

Reading Reflections

(a) reading reflection format

- The paper is about X
- This aligns with Y principle of human centered data science
- Summary of what I found particularly interesting
- A question?

“However, as the authors recognised, contextual informational norms are generally not fixed but may evolve over time. What happens when certain practices, such as re-identifying and linking different pseudonyms to one entity, becomes commonplace and thus becomes the contextual norm? ...Does this mean some uses of data once considered intrusive may no longer need to be disclosed when it becomes commonplace?”

- Edmund

“Are there any ways data scientists can think through these issues methodically and design systems that are more just and uphold the moral obligations of data collectors?”

- Ankit

“Is an individual's consent to be taken as an absolute green signal for a deep dive into the data? Must companies take legal consent with a grain of salt and allow human intervention to filter inferences made so that steps to ensure an individual's privacy is maintained, even after they have given consent to use their information?”

- Nimisha

“What is our legal system doing, if anything, to govern the far-reaching uses of big data (not just from a PII and privacy perspective)? Clearly, anonymity and consent can go only so far to 'protect' society from the evil data scientist and their voodoo mathemagics.”

- Luke

Consent, Privacy and Integrity

Problems with consent

- How consensual is consent?
- Not unique to “big data”, just magnified by it
- Nissenbaum’s concerns
- Leads to problems with privacy

Contextual integrity

- Follow the norms of the space you’re in
- *Appropriateness* of collection
- *Distribution*
- subject, sender, recipient, info type, and transmission principle

Problems with contextual integrity

- What about when things become “commonplace”?
- “Presumption in favour of the status quo”
- Who determines “norms”?

Questions?

Week 3 preview

Homework due next week

- Read & reflect
    - danah boyd and Kate Crawford, *Six Provocations for Big Data (2011)*

- Start Assignment 1: Data Curation (due in 2 weeks)
    - **First step:** read Read Chapter 2 "Assessing Reproducibility" and Chapter 3 "The Basic Reproducible Workflow Template" from *The Practice of Reproducible Research*. *(no reading reflection for this; part of the assignment)*
    - Here's a sample notebook (view, download) that you can use to get started with the API queries.

https://wiki.communitydata.cc/Human_Centered_Data_Science_(Fall_2018)#Week_3:_October_11