Data and Society
COVID-19 models – Lecture 3

2/1/21

Announcements

- Personal Essay due 2/7:
  - 450-525 words / 11 point font / 10 points
  - Send .docx to bermaf@rpi.edu before/by **Sunday, February 7 at midnight.**
  - Rubric and instructions in Lecture 2 on the web

Readings for February 4

The New York Times

Opinion

The Coup We Are Not Talking About

We can have democracy, or we can have a surveillance society, but we cannot have both.

By Shoshana Zuboff

Dr. Zuboff, a professor emeritus at Harvard Business School, is the author of "The Age of Surveillance Capitalism."

Jan. 29, 2021

920

What Privacy Is For

Harvard Law Review, Vol. 126, 2013

24 Pages • Posted: 15 Nov 2012 • Last revised: 23 Nov 2012

Julie E. Cohen

Georgetown University Law Center

Date Written: November 5, 2012

Abstract

Privacy has an image problem. Over and over again, regardless of the forum in which it is debated, it is cast as old-fashioned at best and downright harmful at worst — anti-progressive, overly costly, and inimical to the welfare of the body politic. Yet the perception of privacy as antiquated and socially retrograde is wrong. It is the result of a conceptual inversion that relates to the way in which the purpose of privacy has been conceived. Like the broader tradition of liberal political theory within which it is situated, legal scholarship has conceptualized privacy as a form of protection for the liberal self. Its function is principally a defensive one; it offers shelter from the pressures of societal and technological change. So characterized, however, privacy is reactive and ultimately inessential.

In fact, the liberal self who is the subject of privacy theory and privacy policymaking does not exist. The self who is the real subject of privacy law- and policy-making is socially constructed, emerging gradually from a preexisting cultural and relational substrate. For this self, the purpose of privacy is quite different. Privacy shelters dynamic, emergent subjectivity from the efforts of commercial and government actors to render individuals and communities fixed, transparent, and predictable. It protects the situated practices of boundary management through which self-definition and the capacity for self-reflection develop.

So described, privacy is anything but old-fashioned, and trading it away creates two kinds of large systemic risk. First, privacy is an indispensable structural feature of liberal democratic political systems. Freedom from surveillance, whether public or private, is foundational to the capacity for critical self-reflection and informed citizenship. A society that permits the unchecked ascendancy of surveillance infrastructures cannot hope to

Read this first: Shoshana Zuboff, "**The Coup we are not talking about**", NYTimes
https://www.nytimes.com/2021/01/29/opinion/sunday/facebook-surveillance-society-technology.html

- Julie Cohen, "**What Privacy is For**", Harvard Law Review, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2175406

Date	Topic	Speaker	Date	Topic	Speaker
1-25	Introduction	Fran	1-28	The Data-driven World	Fran
2-1	Data and COVID-19	Fran	2-4	Data and Privacy -- Intro	Fran
2-8	Data and Privacy – Differential Privacy	Fran	2-11	Data and Privacy – Anonymity	Fran
2-15	NO CLASS / PRESIDENT'S DAY		2-18	Data and Privacy – Law	Ben Wizner
2-22	Digital rights in the EU and China	Fran	2-25	Data and Discrimination 1	Fran
3-1	Data and Discrimination 2	Fran	3-4	Data and Elections 1	Fran
3-8	Data and Elections 2	Fran	3-11	NO CLASS / WRITING DAY	
3-15	Data and Astronomy	Alyssa Goodman	3-18	Data Science	Fran
3-22	Digital Humanities	Brett Bobley	3-25	Data Stewardship and Preservation	Fran
3-29	Data and the IoT	Fran	4-1	Data and Smart Farms	Rich Wolski
4-5	Data and Self-Driving Cars	Fran	4-8	Data and Ethics 1	Fran
4-12	Data and Ethics 2	Fran	4-15	Cybersecurity	Fran
4-19	Data and Dating	Fran	4-22	Data and Social Media	Fran
4-26	Tech in the News	Fran	4-29	Wrap-up / Discussion	Fran
5-3	NO CLASS				

Today (2/1/21)

- Lecture 3 / Discussion
- Student Presentations

Lecture 3 – Data and COVID-19 -- models

Why good modeling is hard

Epidemiology and pandemic modeling

Vaccine distribution modeling

Modeling is critical but hard

- Scientific modeling used to **understand and predict what may happen.**
- Requires identifying **relevant aspects of a real world situation** to inform the model.
- **Accuracy of a model matters**: Does it predict things that will really happen, at the same frequency, in the same environment?
- Many parts of the model to get right:
  - Process / simulation
  - Parameters
  - Interpretation of results, etc.

Interpreting the Data: Correlation and Causation

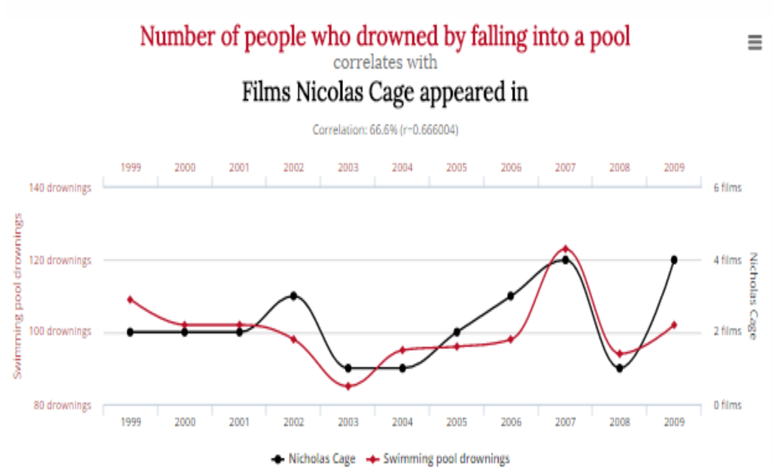
- **Correlation** is a statistical measure that indicates the extent to which two or more variables fluctuate together. [http://whatis.techtarget.com/definition/correlation]

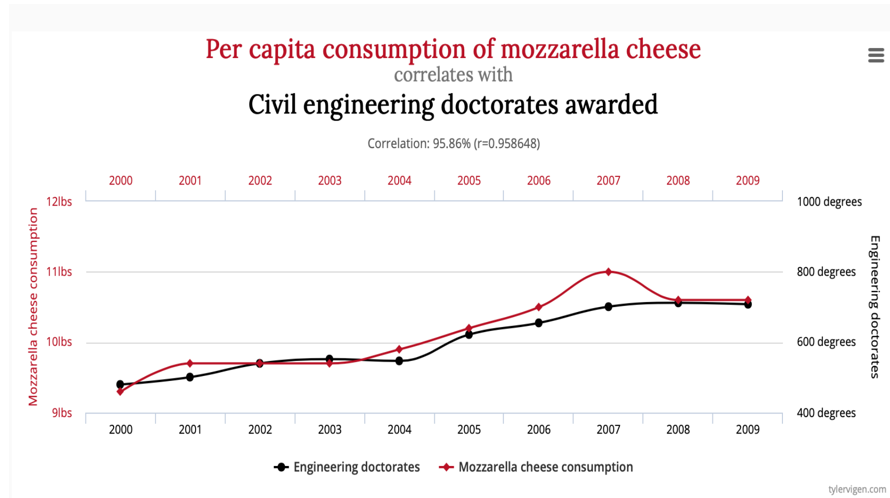
- **Causation**, or causality, is the capacity of one variable to influence another. The first variable may bring the second into existence or may cause the incidence of the second variable to fluctuate.

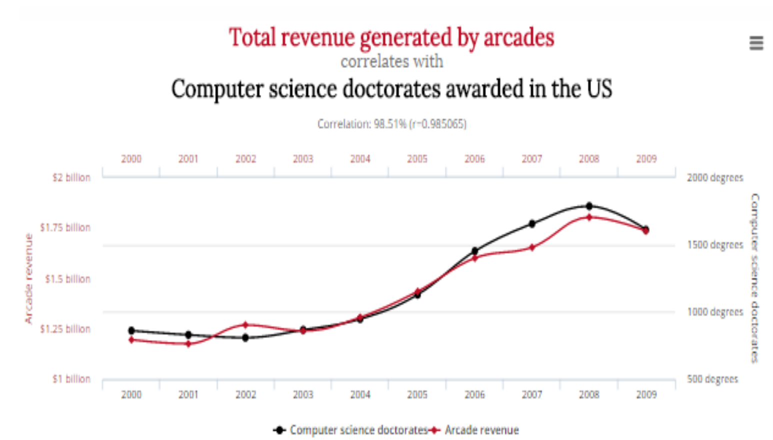
- **Causation is often confused with correlation.** Causation implies correlation but **correlation by itself does not imply causation**. There may be a third factor, for example, that is responsible for the fluctuations in both variables. [http://whatis.techtarget.com/definition/causation]

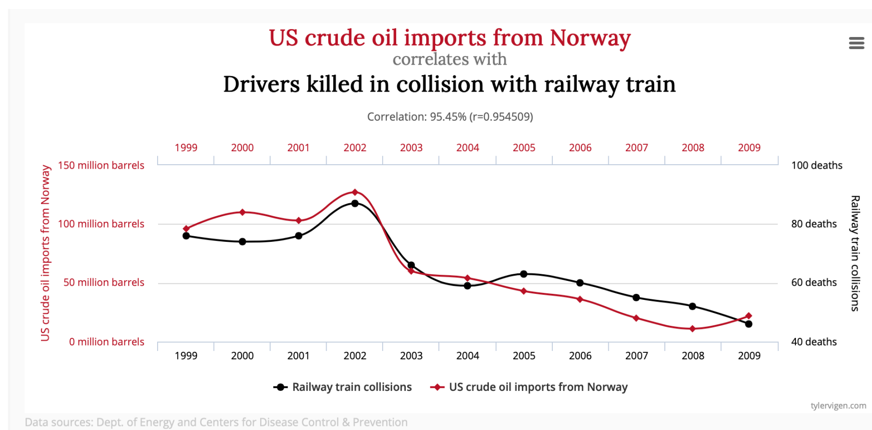
Correlated but unlikely to be causal ☺

http://tylervigen.com/spurious-correlations









Epidemiology and Pandemic Modeling

- 1860's-1870's, London, Cholera epidemic → early use of data used to track, investigate, contain disease
  - **John Snow (doctor)** – theorized (correctly) that cholera was spread through drinking water
    - Source of outbreak: improper decontamination from water from River Lea
  - **William Farr (doctor and statistician)** – tracked correlations of disease and water sources and companies in London; his work helped invent epidemiology
- Farr's epidemiological approach to curbing cholera focused **not on cure and treatment but on containment and information**
  - Data provided information on where the disease was spreading and where it is most prevalent; valuable input used to contain last big cholera outbreak in London

Street disinfecting in London, 1877

https://www.nytimes.com/interactive/2020/06/10/magazine/covid-data.html

Birth of epidemiological approach: data, models, visualization, interpretation

- **William Farr:** "*Facts, however numerous, do not constitute a science.* Like innumerable grains of sand on the seashore, single facts appear isolated, useless, shapeless; it is only when compared, when arranged in their natural relations, when crystalized by the intellect, that they constitute the eternal truths of science."

- **Farr's Law** (the curve you flatten when you "flatten the curve"): Epidemic events rise and fall in a roughly symmetrical pattern. The time-evolution behavior can be captured by a single mathematical formula that could be approximated by a bell-shaped curve.

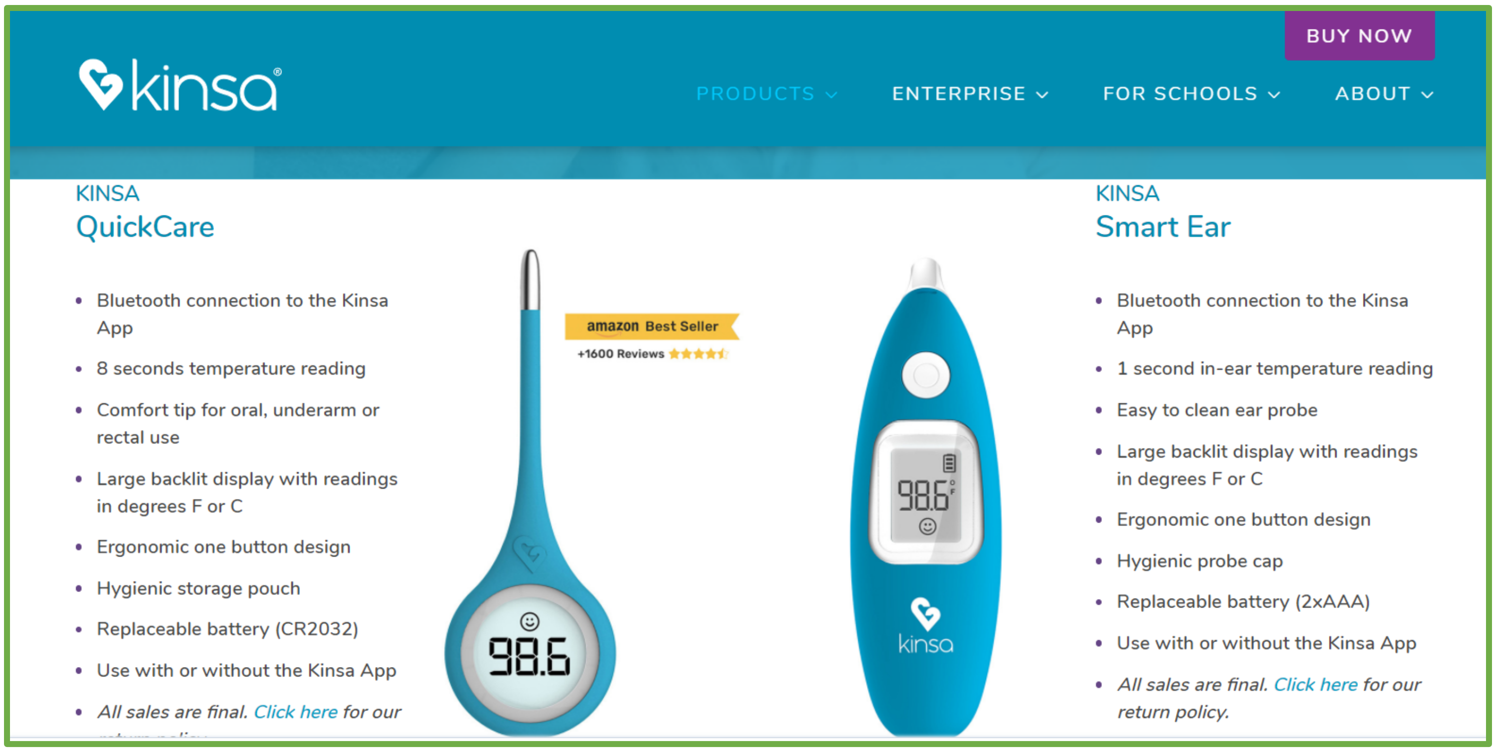
Data is at the heart of epidemiology

- “Farr was among the first to think **systematically** about how **data** on outbreaks, their **distribution in space and over time**, could be used to curb them as they unfolded – and to minimize future ones.” (NY Times, see ref.)
  - Farr focused on gathering data about deaths, births, age, occupation, location, etc. and developed patterns to answer questions about the populations
- **Coronavirus epidemic surprisingly similar to cholera epidemic**: still studying / developing vaccines, no cures, many questions about spread and who is likely to fare the worst. Epidemiological approaches (data collection, containment, modeling) are critical approaches to protect uninfected.

Opportunities and Challenges for Epidemiological Data Gathering for COVID-19

- **Much data available** at fine grain: health status, co-morbidities, genetic strains of the virus, antibody test results, etc.
- In early days of COVID-19, **no single data repository existed**. “the vast majority of public-health data during epidemics are still largely organized on pen, paper, Excel and PDFs.” [NY Times]
- Information on COVID-19 generally **captured too late** – end stages of the disease (hospitalization or death). Multi-day initial period between contraction and symptoms hampers good forecasting.
- **Data-driven approaches can supplement “late stage” information**
  - Early-stage testing, especially in vulnerable populations (“**sentinel surveillance**”)
  - Symptom tracking (“**syndromic surveillance**”) supplements official data from patients in the health care system with data tracking of the appearance of disease symptoms (e.g. temperature, Google searches on key words, etc.)

Syndromic surveillance via smart thermometers (part of the Internet of Things!)



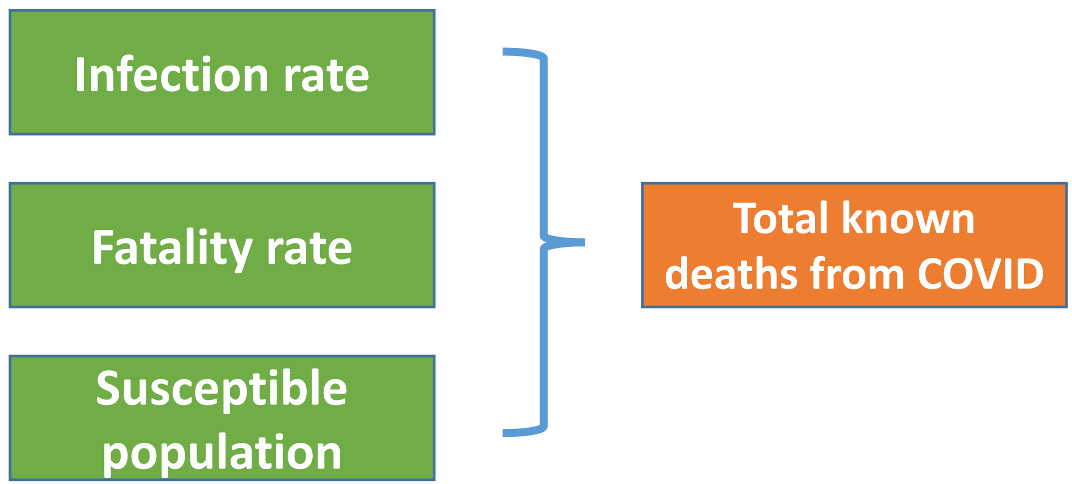
https://www.kinsahealth.co/products/

Other approaches to epidemiological modeling: Expand data collection beyond humans …

- Test **environmental sources**: water, air, etc. to detect changes in pathogen levels.
- Monitor **animal populations**:
  - Some diseases (COVID, SARS, MERS, swine flu, bird flu, Ebola, HIV, Zika, etc.) were at one point animal diseases
  - Humans can surveil animal diseases to try to stop them before they make the jump to humans.
    - Animal surveillance in Cambodia: Hotline that tells the government when a number of chickens die. Government will replace chickens and clean up environment.
    - Surveillance of bats, pigs, birds could go similarly

Creating accurate and useful COVID-19 models is incredibly hard

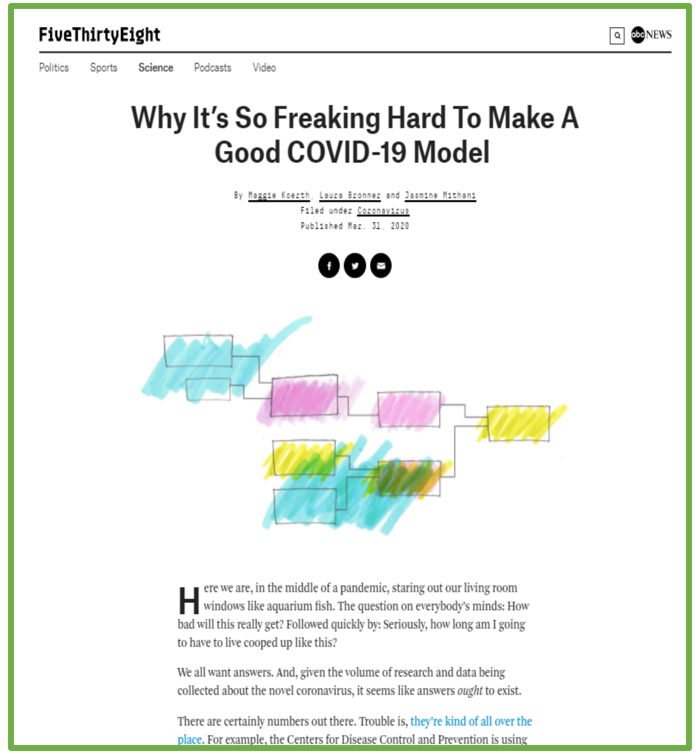
- **Basic approach** [538 reading]:



- *"Think of it like making a pie. If you have a normal recipe, you can do it pretty easily and expect a predictable result that makes sense. But if the recipe contains instructions like "add three to 15 chopped apples, or steaks, or brussels sprouts, depending on what you have on hand" … well, that's going to affect how tasty this pie is, isn't it."* [538 reading]

Reading for Today's Discussion

- **"Why It's So Freaking Hard to Make a Good COVID-19 Model"**, 538
- https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a-good-covid-19-model/, on class website



FiveThirtyEight

Politics Sports Science Podcasts Video

Why It's So Freaking Hard To Make A Good COVID-19 Model

By Maggie Koerth, Laura Bronner and Jasmine Mithani
Filed under Coronavirus
Published Mar. 31, 2020

Here we are, in the middle of a pandemic, staring out our living room windows like aquarium fish. The question on everybody's minds: How bad will this really get? Followed quickly by: Seriously, how long am I going to have to live cooped up like this?

We all want answers. And, given the volume of research and data being collected about the novel coronavirus, it seems like answers *ought* to exist.

There are certainly numbers out there. Trouble is, they're kind of all over the place. For example, the Centers for Disease Control and Prevention is using

Current COVID-19 models

- **"SEIR model"** – estimates how people move among 3-4 states:
  - **S:** individuals **susceptible** to the virus
  - (**E:** Individuals who are **exposed** but not yet infectious)
  - **I:** individuals **infected** by the virus
  - **R:** individuals immune to the virus (**recovered or dead**)
- Motion between different states varies for different demographics (age, sex, health status, employment, number of contacts, etc.)
- Approach used in the Imperial College Model (March, 2020) that projected >2M deaths if the U.S. did nothing.

Many parameters used to estimate COVID impact

- **Population** size and density
- **Hospital** capacity
- $\boldsymbol{R_0}$ -- number of people to whom one infected person will pass the virus (reproduction number)
- **Number of days** the virus can be passed before symptoms appear
- **How infectious people are** before, during, after symptoms
- Whether people are **naturally immune**
- Whether (and how long) people are **immune after having the disease**
- Probability that an infected person will **spread the virus** to a susceptible person
- **Movement model:** equation-based (social subset for population group), agent-based (individual movement), etc.

Questions to ask about COVID-19 models
(from New England Journal of Medicine)

https://www.nejm.org/doi/full/10.1056/NEJMp2016822

1. What is the **purpose and time frame** of this model? For example, is it a purely statistical model intended to provide short-term forecasts or a mechanistic model investigating future scenarios? These two types of models have different limitations.

2. What are the **basic model assumptions**? What is being assumed about **immunity** and **asymptomatic transmission**, for example? How are **contact parameters** included?

3. How is **uncertainty** being displayed? For statistical models, how are confidence intervals calculated and displayed? Uncertainty should increase as we move into the future. For mechanistic models, what parameters are being varied? Reliable modeling descriptions will usually include a table of parameter ranges — check to see whether those ranges make sense.

4. If the model is fitted to data, **which data are used**? Models fitted to confirmed Covid-19 cases are unlikely to be reliable. Models fitted to hospitalization or death data may be more reliable, but their reliability will depend on the setting.

5. Is the **model general, or does it reflect a particular context**? If the latter, is the spatial scale — national, regional, or local — appropriate for the modeling questions being asked and are the assumptions relevant for the setting? Population density will play an important role in determining model appropriateness, for example, and contact-rate parameters are likely to be context-specific.

Different assumptions, different results

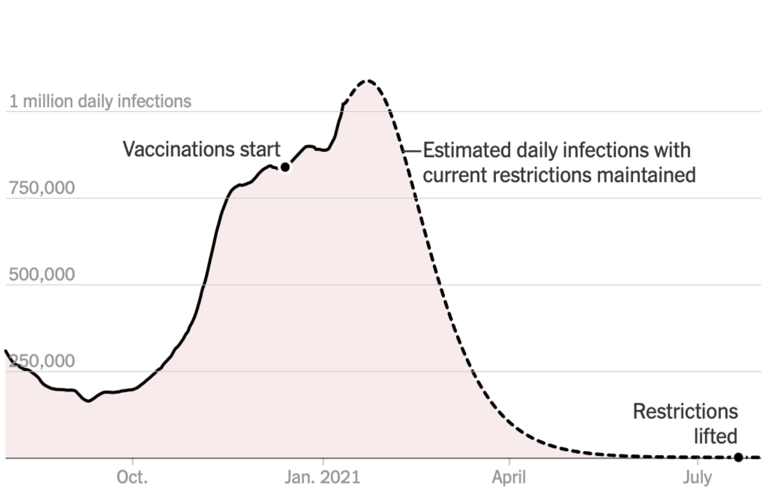
- https://projects.fivethirtyeight.com/covid-forecasts/

Modeling vaccine uptake

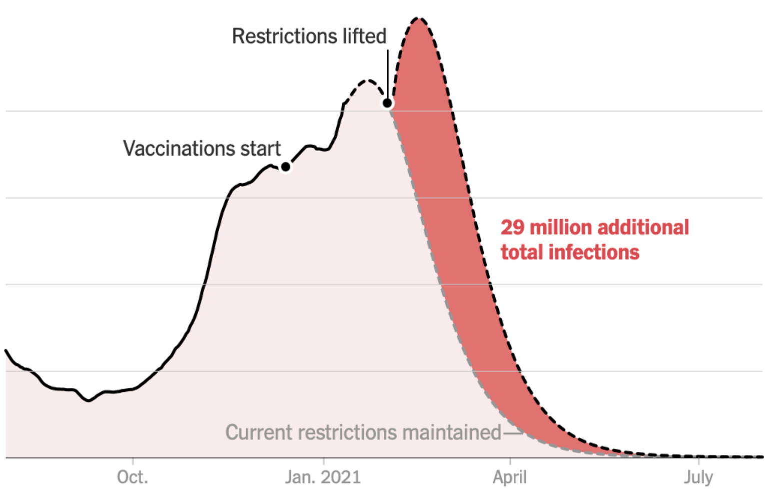
(https://www.nytimes.com/interactive/2021/01/24/us/covid-vaccine-rollout.html, New York Times)

- “current restrictions”: working remotely, limiting travel, wearing masks, etc.
- Goal is “herd immunity” so that the probability of getting the disease is low enough that we can resume “normal life”.
- Herd immunity will happen when enough people have had the disease and/or have been vaccinated against the disease

If current restrictions remain in place until late July



If restrictions are lifted in February



Methodology and parameters

(https://www.nytimes.com/interactive/2021/01/24/us/covid-vaccine-rollout.html, Galanti and Shaman/Columbia and Pfizer)

Methodology

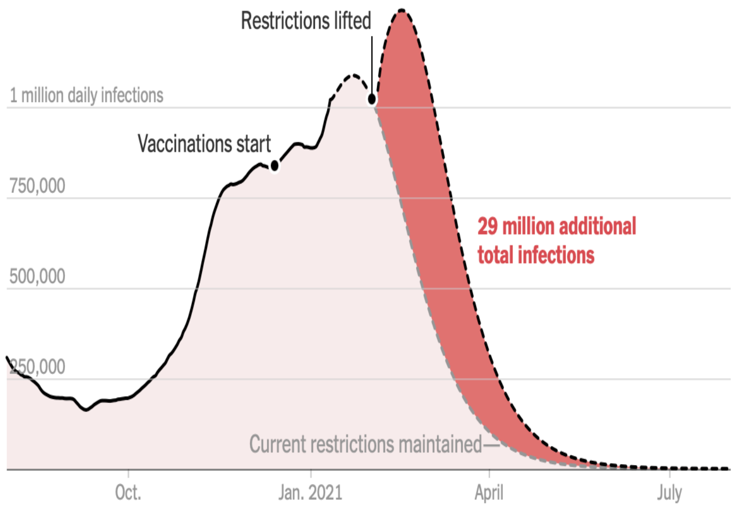
- The model assumes a **vaccine efficacy of 90 percent 10 days after the first dose and 95 percent one week after the second dose** — and that those vaccinated will ultimately receive both doses. The model also assumes a **vaccination schedule following the C.D.C. guidelines** for different groups, followed by a distribution ramping up to five million doses per week of other adults and children.

- **Historical case data is used to estimate the number of past infections** and effective reproduction rate in each state as of Jan. 10, 2021. Infection projections are based on the **virus having a basic reproduction number of 2.4** without any nonpharmaceutical interventions, including social distancing restrictions.

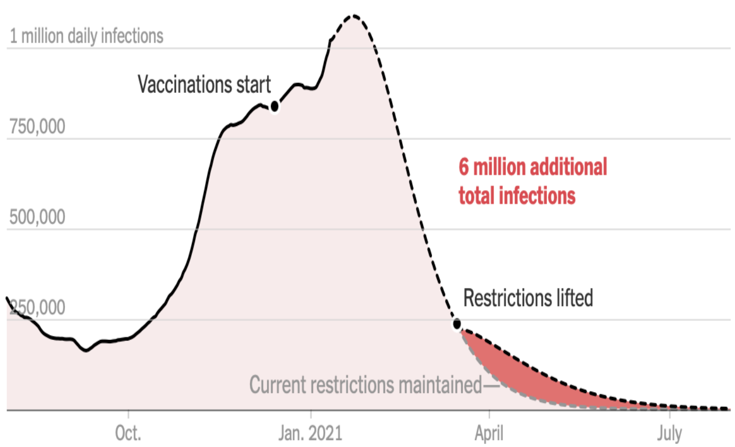
- In each scenario, **restrictions are fully relaxed at the date indicated**, except those in which restrictions are strengthened. In these scenarios, restrictions are gradually relaxed over the following months. **For scenarios in which restrictions are strengthened, the basic reproduction number is lowered to 1.5**.

Results

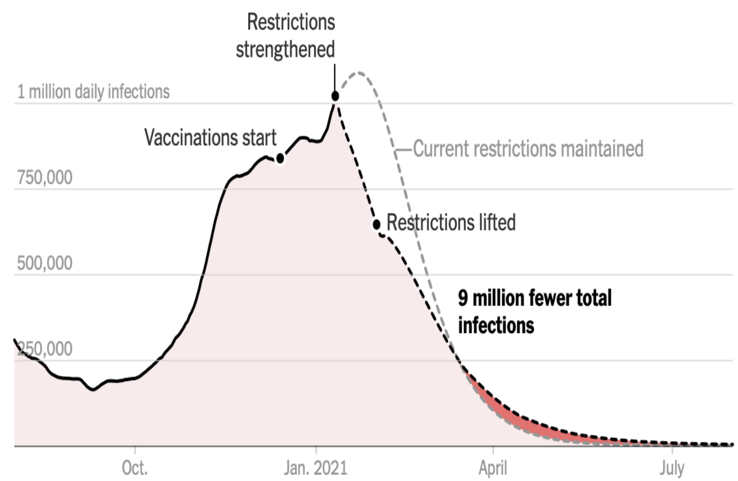
If restrictions are lifted in February



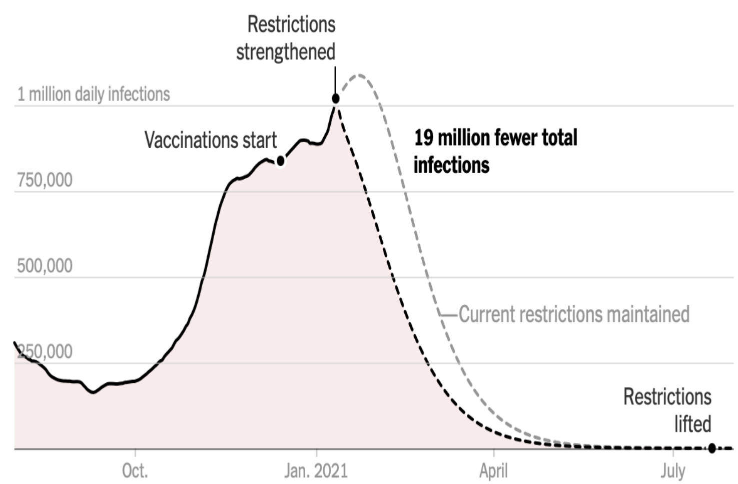
If restrictions are lifted in mid-March



If restrictions are strengthened until February



If restrictions are strengthened until late July



Lecture 3 Resources

- "**How data became one of the most powerful tools to fight an epidemic**", New York Times Magazine, https://www.nytimes.com/interactive/2020/06/10/magazine/covid-data.html. *Historical perspective about the birth of epidemiology and its use for epidemics.*
- "**Where the Latest COVID-19 Models Think We're Headed and Why They Disagree**," 538, https://projects.fivethirtyeight.com/covid-forecasts/. *Current COVID-19 models and their differentiators.*
- **COVID-19 Forecast Hub**, https://covid19forecasthub.org/. *Rich and authoritative source of data and models.*
- *"***Special report: The simulations driving the world's response to COVID-19",** Nature, https://www.nature.com/articles/d41586-020-01003-6. *Detail on modeling and the Imperial College COVID-19 Model.*
- "**Modeling the Pandemic: Attuning Models to their Contexts**", BMJ Global Health, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7307539/. *How models for COVID-19 differ from other pandemic models.*
- *"***Why Vaccines Alone will not end the Pandemic***", New York Times, https://www.nytimes.com/interactive/2021/01/24/us/covid-vaccine-rollout.html?referringSource=articleShare*

Presentations

Upcoming Presentations

- **Presentations for February 4**
  - '**Are vaccine providers selling your health data? There's not much stopping them**.", Vox, https://www.vox.com/recode/22251118/vaccine-health-data-privacy-laws-philadelphia (Hannah L.)
  - "**WHO plans privacy, security rules for COVID-19 vaccine certificates**," Wall Street Journal, https://www.wsj.com/articles/who-plans-privacy-security-rules-for-covid-19-vaccine-certificates-11610706601 (Justin O.)
- **Presentations for February 8**
  - "**Changes to the census could make small towns disappear**," New York Times, https://www.nytimes.com/interactive/2020/02/06/opinion/census-algorithm-privacy.html (Liam M.)
  - "**Can a set of equations keep U.S. census data private?**," Science, https://www.sciencemag.org/news/2019/01/can-set-equations-keep-us-census-data-private (Davis E.)
- **Presentations for February 11**
  - "**We're banning facial recognition. We're missing the point**." New York Times, https://www.nytimes.com/2020/01/20/opinion/facial-recognition-ban-privacy.html (Josh M.)
  - "**This site published every face from Parler's Capitol riot videos**", Wired, https://www.wired.com/story/faces-of-the-riot-capitol-insurrection-facial-recognition/ (Nate S.)

Today’s Presentations

- “**The swiss cheese model of pandemic defense**”, New York Times, https://www.nytimes.com/2020/12/05/health/coronavirus-swiss-cheese-infection-mackay.html (Isaac L.)

- “**The architecture of mass vaccine distribution**,” Bloomberg CityLab, https://www.bloomberg.com/news/features/2021-01-22/the-architecture-of-covid-vaccine-distribution
(Adam M.)