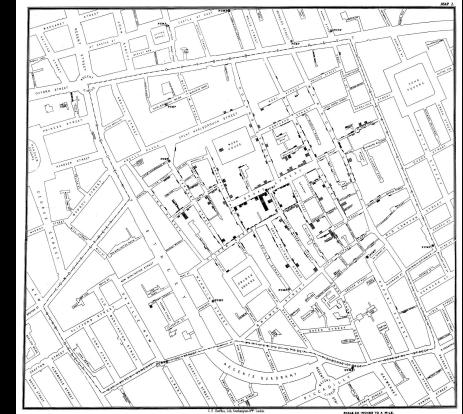
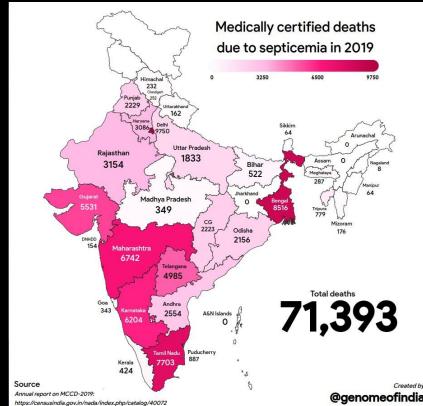
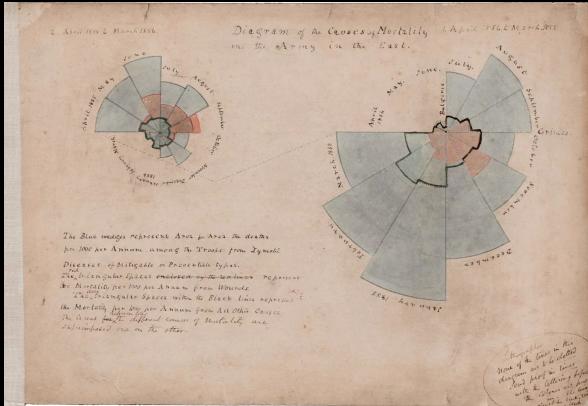


Introduction to the course



Saket Choudhary
saketc@iitb.ac.in

Introduction to Public Health Informatics
DH 302

Lecture 01 || Wednesday, 08th January 2025

Welcome to DH302!

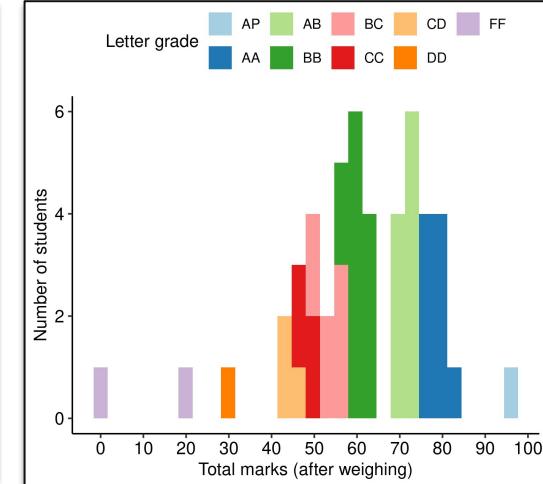
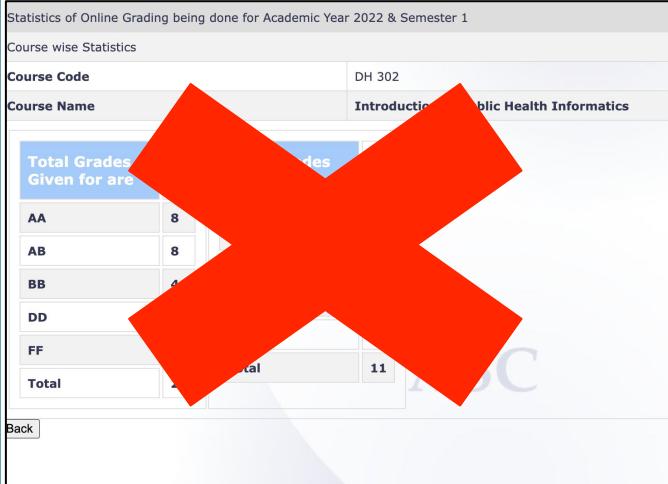
*“Somewhere, something incredible is
waiting to be known”*



Carl Sagan

American astronomer and planetary scientist

If you are here because of this, rethink...



Logistics upto Midsems (50%)

- **Assignments: 10% (Best 2/3 out of 3/4)**
 - Due + graded online
 - Best 2 (or 3) will be considered for grading out of 3 (or 4)
 - Late submission policy: 10% penalty per day
 - One submission per student (Attribute if you discussed with someone)
- **Mid-sem: 30%**
 - Closed book and offline (no collaboration)
- **Surprise quizzes: 10%**
 - Based on content taught in the past
 - Participation points if you interact (and not just come to the class)
- **Class participation:** Relative bonus points (Scaled to the maximum total)
- Possibility of absolute bonus points

Final grades: RG (Relative grading)

Logistics - Office hour(s)

- Lecture: Wednesdays and Fridays, 11:05am – 12:30pm || LH 101
- Instructor Office: G-22, KCDH, KReSIT Basement
- Instructor Office Hours: **Wednesdays, 4:00 - 5:00pm** or by appointment
- For appointments outside office hours: <https://cal.com/saketkc/>
- Contact: saketc@iitb.ac.in | Ext: 3785 (+91 22 2159 3785)

Use email preferably only for personal requests - if you have a question, someone else might also have a similar one.

Logistics - Reading material

- Material is based on a mix of topics
- No one textbook
- Slides will be text + (digitally) handwritten
- Lecture will contain sufficient references + key papers

Logistics - Programming/Coding

- I will cover some preliminary coding exercise in the hands-on class
- Please bring your **laptops** for hands-on sessions (will be announced)
- But if this is your first exposure to programming, please use the [programming resources](#) to familiarize yourself
- R installation and a brief tutorial is available on website

!!! Logistics - Programming/Coding - !!!

- First half of the course will **require programming in R (and only R)**
- If you are not comfortable with R, we will provide templates to learn, but you are expected to pick up R programming on your own
- Resources and tutorials are available [here](#) - they are lengthy but worth it if you are new to the language
- **The following list of non-exhaustive excuses will not be entertained:**
 - I do not know R
 - I do not like R
 - I prefer Python/C/xyz.



!!! Collaboration policy and Academic Integrity !!!

- You are expected to work on your own for most part of the course.
- For assignment problems, If you get stuck, you are welcome to discuss it with other students (in-person or online). However, the **solutions must be your work**. If you discussed with someone, **please mention their name and what you received help with** in your submission. If you do not attribute and we find similarities in the final submissions - **this will automatically count as plagiarism!**
- **Mid-semester exam (closed book). No collaboration is allowed.**
- **Write/speak what you understand.** If you write something, it is assumed you understand it - and hence are open to being quizzed by it
- Simply: **DTRT - Do the right thing**

"I declare that I will adhere to all principles of academic honesty and integrity throughout my stay in the Institute. I will not seek or give unauthorized assistance in tests, quizzes, examinations or assignments. I will not misrepresent, fabricate or falsify any idea/data/fact/source in my project submissions. I understand that any violation of the above will be cause for disciplinary action as per the rules and regulations of the Institute."

[See Policy](#)

TAs and office hours



Anisha Karmakar
23D1622@iitb.ac.in
Friday, 3-4 pm, BSBE
(Lab 605)

Chetan Patil
20b030012@iitb.ac.in
Wednesday, 2-3 PM,
KCDH Lab

Devendra Singh
devendrasb@iitb.ac.in
Friday, 5-6 PM KCDH
Lab

Kriti A
210100083@iitb.ac.in
Tuesdays, 5-6 PM, ME
Department

Shobhit Aggarwal
20d100026@iitb.ac.in
Wednesdays, 4-5PM,
KCDH Lab

Sunny Gupta
sunnygupta@iitb.ac.in
Friday, 4-5 pm, Medal
EE

Code of conduct

1. Be on time (classes/assignments/projects)
2. Be respectful on Moodle (and all other platforms).
3. Maintain a convivial & collegial atmosphere

Questions and clarifications are always welcome during/after the class or during office hours.

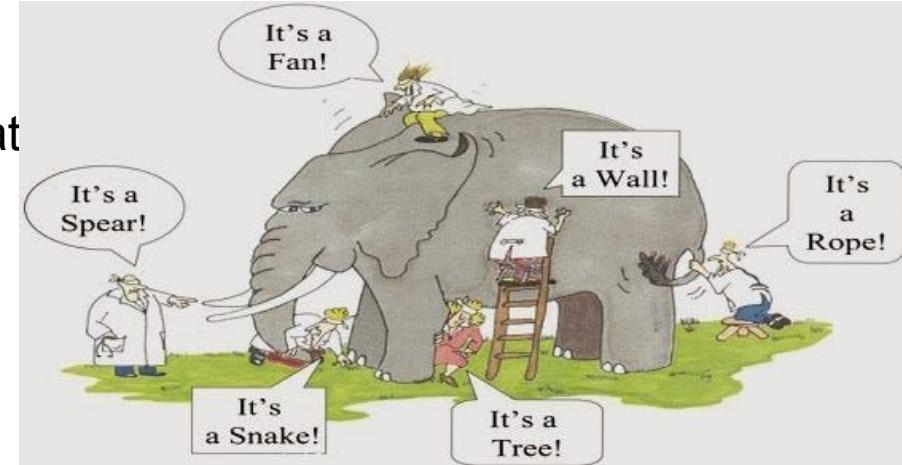
What is the first part of the course about?

- What is health, public health and health informatics
- Statistical techniques underlying healthcare informatics
- How to analyze healthcare data: survey data, birth and mortality data, disease surveillance data
- What is genome sequencing and its role in healthcare?

What is the first part of the course about?

Goal 1: Give you a flavour of **science**

- **Science:** Essence of science is “inquiry”: concrete descriptions of what we observe; theories about what drives those observations
- **Engineering:** “Design”: expands the scope of human plans results



[Source](#)

Goal 2: Equip you with **fundamental analytical framework** to answer your own questions

What is this course not about?

- I am looking for an easy course
- I want a guaranteed good grade
- I want to learn <R/Python/C++>
- I want to learn AI/ML
- I want to learn image processing, time series, ..., <specialized topic>

Tentative topics

Week	Date	Topic	Slides	Assignment	Resources
01	01-08 (Wed)	1. Introduction to the course and history of public health informatics			
01	01-10 (Fri)	2. Statistical models for health data			
02	01-15 (Wed)	3. Exploratory data analysis			
02	01-17 (Fri)	4. Dimensionality reduction for healthcare data		• Assignment 01 released	
03	01-22 (Wed)	5. Basics of ML for health data - Regression			
03	01-24 (Fri)	6. Multivariate analysis of health data		• Assignment 01 due • Assignment 02 released	
04	01-29 (Wed)	7. Basics of ML for health data - Classification			
04	01-31 (Fri)	8. Introduction to DNA sequencing		• Assignment 02 due • Assignment 03 released	
05	02-05 (Wed)	9. Modeling high-throughput data			
05	02-07 (Fri)	10. Models for predicting disease outbreaks		• Assignment 03 due • Assignment 04 released	
06	02-12 (Wed)	11. Analysis of image data			
06	02-14 (Fri)	12. Invited Lecture		• Assignment 04 due	
07	02-19 (Wed)	13. Buffer			
07	02-21 (Fri)	14. Buffer			
08	02-26 (Wed)	Midsem week			
08	02-28 (Fri)	Midsem week			

****subject
to change**

Why you might want to take this course:

- Hands-on activities and real-world case studies related to public health problems
- First principle approach to solving problems in public health and healthcare
- Health Informatics and digital health is **ubiquitous** abroad:
 - Academic labs
 - Industry: [Epic systems](#), Pharma companies, [Siemens](#), [Abbott](#), [Google \(health\)](#), [Apple \(health\)](#), ..
- Health Informatics is becoming ubiquitous in **India**
 - Academic labs
 - Industry: Pharma companies, Genomics ([Strand genomics](#), [MedGenome](#), [MapMyGenome](#), [Precision health](#)), [GoQii](#), [Samsung](#), ...
- Learn how to *analyse* (“do **hands-on data science**”) large-scale datasets

Nature of assignments and mid-sems

- Some theory questions on the analytical techniques that we discuss in class
- Some analytical questions based on data interpretation
- MidSem: Closed book, no collaboration

Assessment:

- If you collaborate on your assignments with a friend, mention their name
- If we detect identical or very similar answers we are bound to award you zero and report it to the academic affairs office

Expectations

- Some prior exposure to mathematics - linear algebra and probability preferably
- Put more effort and time if the material is extremely new to you
- Ask questions
- Do not speak/write/use things (particularly tools) you do not understand
- Attend class and take notes (“I cannot understand anything from the slides on my own”)
- You are expected to KNOW and could be asked to explain what you write (assignment/quiz/exams) - So THINK before <ChatGPT>ing
- Do not use brute-force to make things work somehow; but do explore (more on this later)

Use LLMs wisely...

What technology will you use to profile and what is the underlying principle behind the technology that you are leveraging - will it work both for human and viral transcriptomes? If not, what modification would you need to make it work (you can also mention caveats)

To profile the transcriptome in the gut across different cell types, I would utilize single-cell RNA sequencing (scRNA-seq). This technology would be applied to dissect gene expression at the resolution of individual cells needed for understanding the complexity of a large heterogenic tissue such as the gut.

Underlying Principle scRNA-seq works by:

1. Cell Isolation: Using techniques such as microfluidics (for example, 10x Genomics Chromium), droplet encapsulation, or plate-based sorting, individual cells are caught.
2. mRNA Capture: mRNA is reverse-transcribed into cDNA using oligo(dT) primers.
3. Amplification: The cDNA is amplified and tagged with UMIs for accurate quantification.
4. Sequencing: High-throughput sequencing is used to generate reads corresponding to the transcriptome of each cell.
5. Data Analysis: The output data is analyzed to identify differentially expressed genes, cell types, and pathways.

Use LLMs wisely - Please DON't do this

Dear Sir,

I hope this email finds you well. I am writing to express my strong interest in enrolling in the DH302 course Introduction to Public Health Informatics, titled Introduction to Public Health Informatics. The comprehensive and interdisciplinary nature of the course content, particularly its emphasis on public health promotion, disease prevention, biomedical informatics, and healthcare ethics, aligns deeply with my academic and career aspirations.

Despite my earnest effort to secure a spot through pre-registration, I was unable to enroll due to the course cap of 120 students. Given the high demand for this course and its significant relevance to contemporary global health challenges, I humbly request your kind consideration to either accommodate my enrollment or increase the course cap to include additional interested students.

I assure you of my utmost commitment and enthusiasm for actively participating in and contributing to the course.

If required, I would be happy to provide any further details or follow any specific procedure.

Thank you for considering my request. I understand the constraints on resources and teaching capacity, and I genuinely appreciate the efforts you and KCDH department invests in delivering such an impactful course.

Looking forward to your response.

Questions?

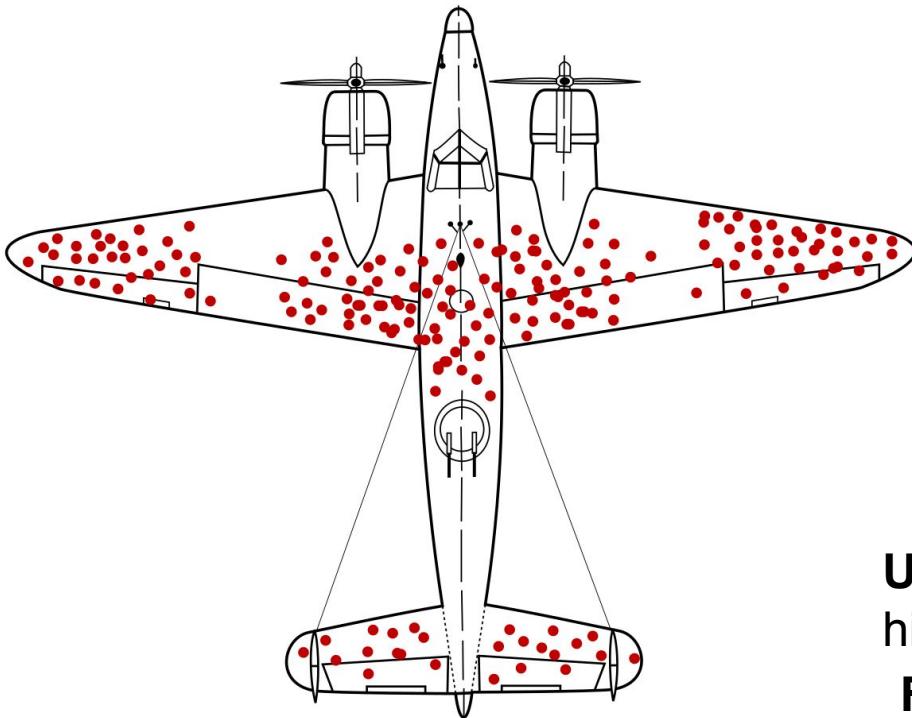


Goals for today

1. A short(est) introduction to health, healthcare and public health
2. What is health informatics?
3. 20,000 feet picture of the course

A simple problem from World War II

Where should you put the armour?



Section of plane	Bullet holes per square foot
Engine	1.11
Fuselage	1.73
Fuel system	1.85
Rest of the plane	1.50

Usual answer: Put armour where the bullet holes are maximum (Fuel system)

First principles thinking: Armour goes on the engine!

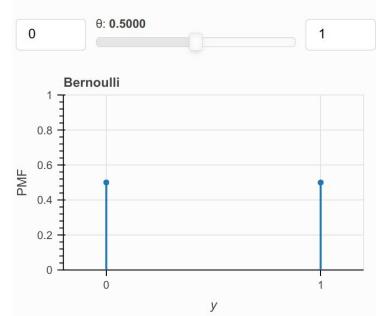
Course vignettes

Probability models for health data

- How should the number of Covid-19 cases be modeled?
- What is the correct statistical model for representing deaths as a function of time?
- What is the distribution of height of males in a village? What about children in village? What about children in a village known to be suffering from stunting?

How to think about distributions? The most important ones..

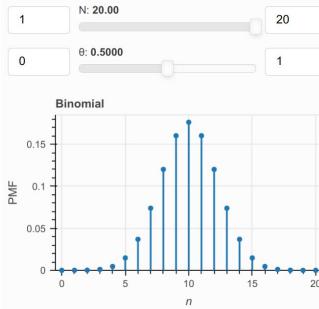
Discrete



$$f(y; \theta) = \begin{cases} 1 - \theta & y = 0 \\ \theta & y = 1. \end{cases}$$

Mean: θ

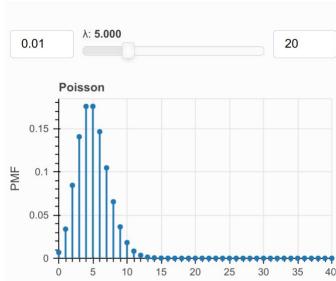
Variance: $\theta(1 - \theta)$



$$f(n; N, \theta) = \binom{N}{n} \theta^n (1 - \theta)^{N-n}.$$

Mean: $N\theta$

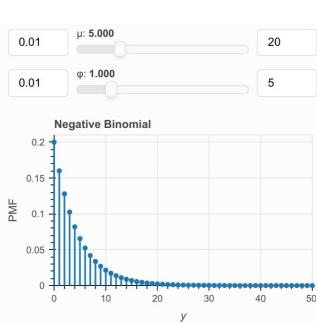
Variance: $N\theta(1 - \theta)$



$$f(n; \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

Mean: λ

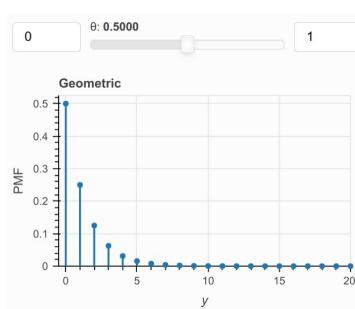
Variance: λ



$$f(y; \mu, \phi) = \frac{\Gamma(y + \phi)}{\Gamma(\phi) y!} \left(\frac{\phi}{\mu + \phi} \right)^{\phi} \left(\frac{\mu}{\mu + \phi} \right)^y.$$

Mean: μ

$$\text{Variance: } \mu \left(1 + \frac{\mu}{\phi} \right).$$



$$f(y; \theta) = (1 - \theta)^y \theta.$$

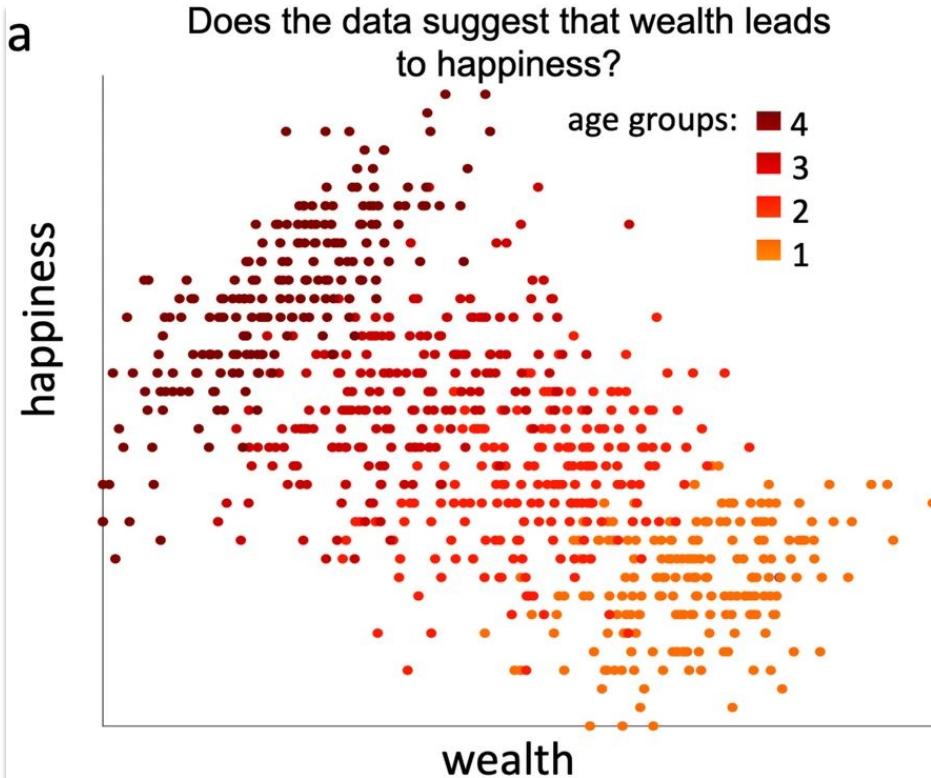
Mean: $\frac{1 - \theta}{\theta}$

Variance: $\frac{1 - \theta}{\theta^2}$



Simeon Poisson,

Exploratory data analysis

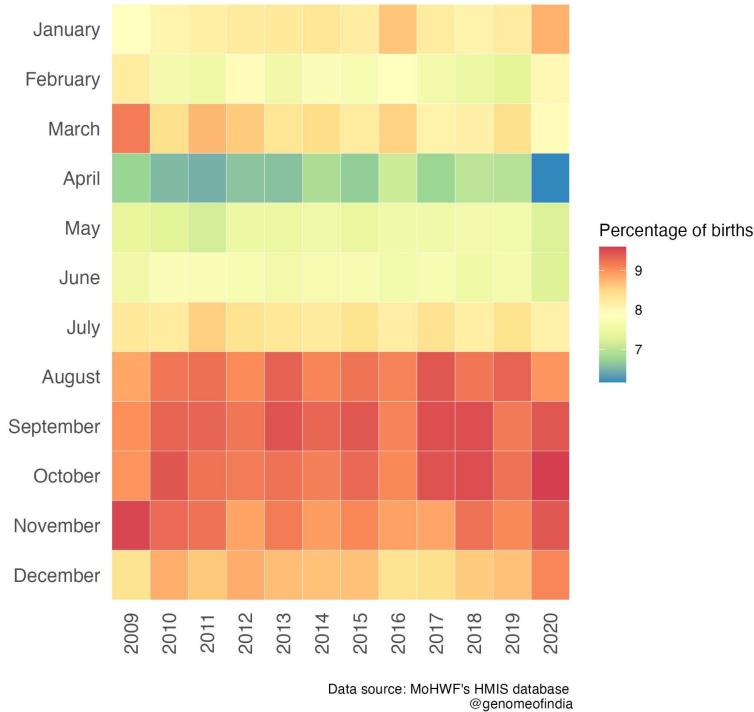


Exploratory data analysis

When is your birthday? 1st half or 2nd half?

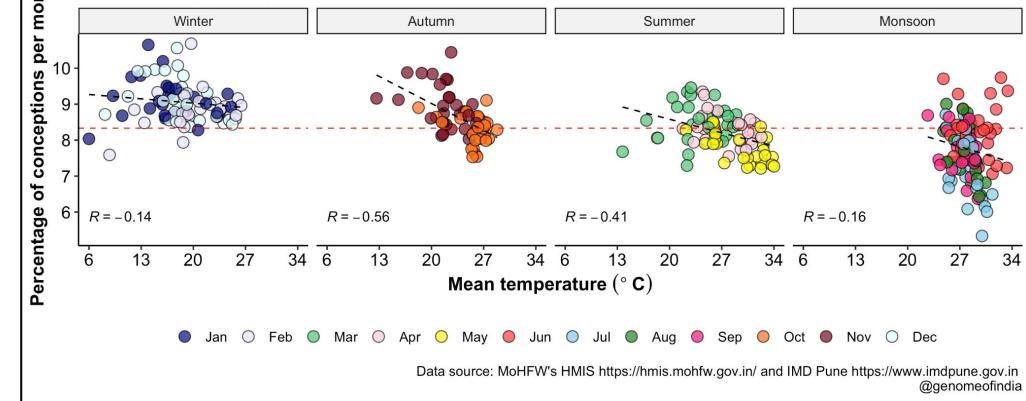
Seasonality of births in India

Percentage of live births per month in a year in India



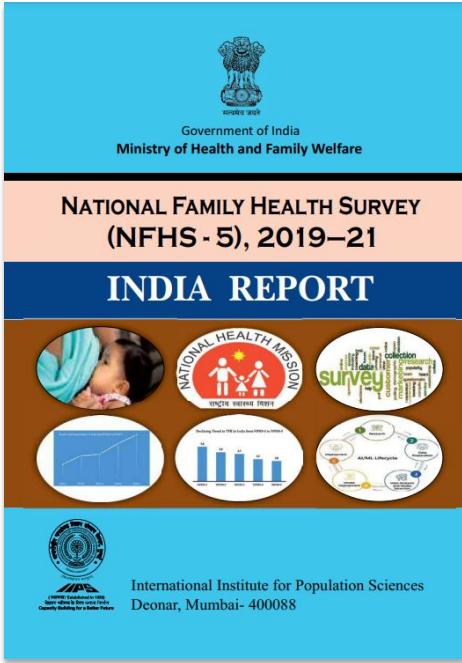
Seasonality of conceptions in India

Correlation b/w percentage of estimated conceptions and temperature across seasons

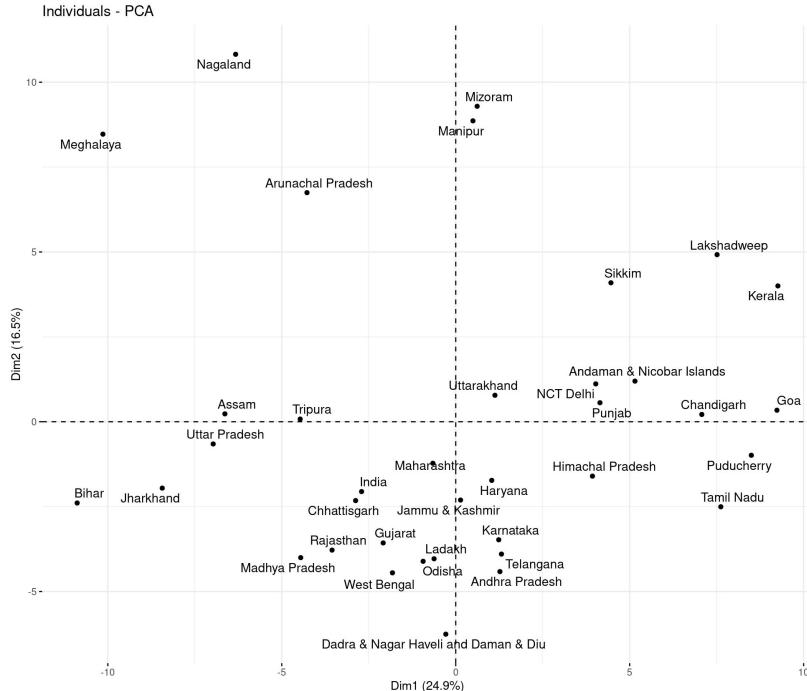


<https://genomeofindia.substack.com/p/wake-me-up-when-august-ends>

Dimensionality reduction for health care data



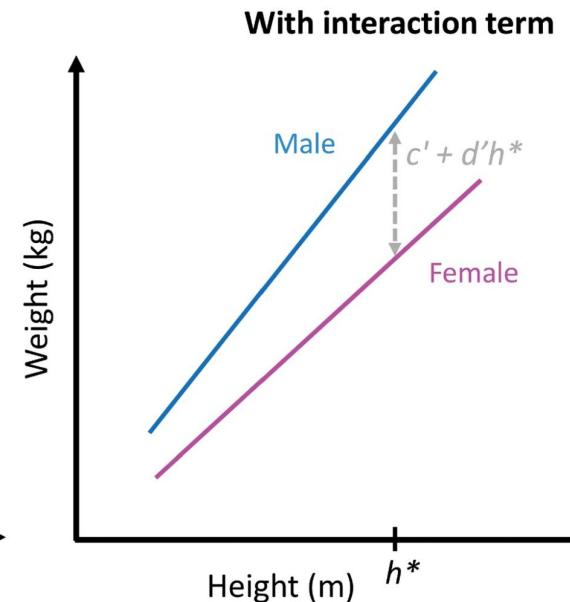
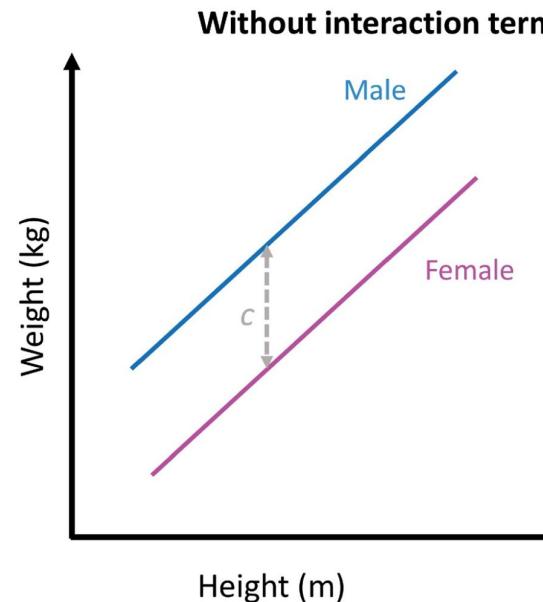
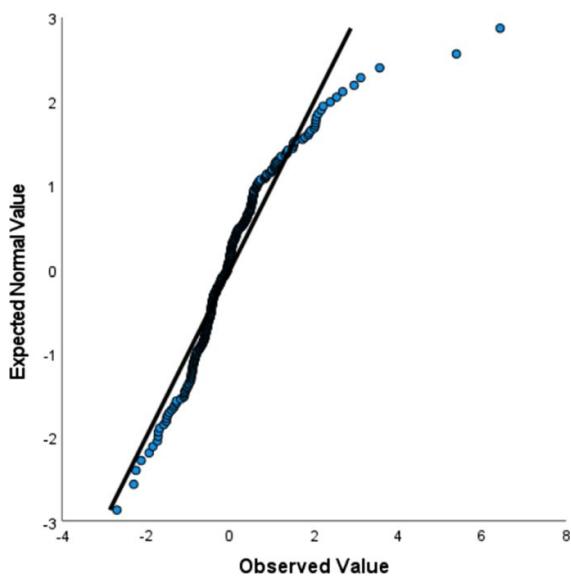
Problem: Are different states performing similarly on multiple health metrics?



Solution: Dimensionality reduction, clustering

Multivariate regression for health-related data

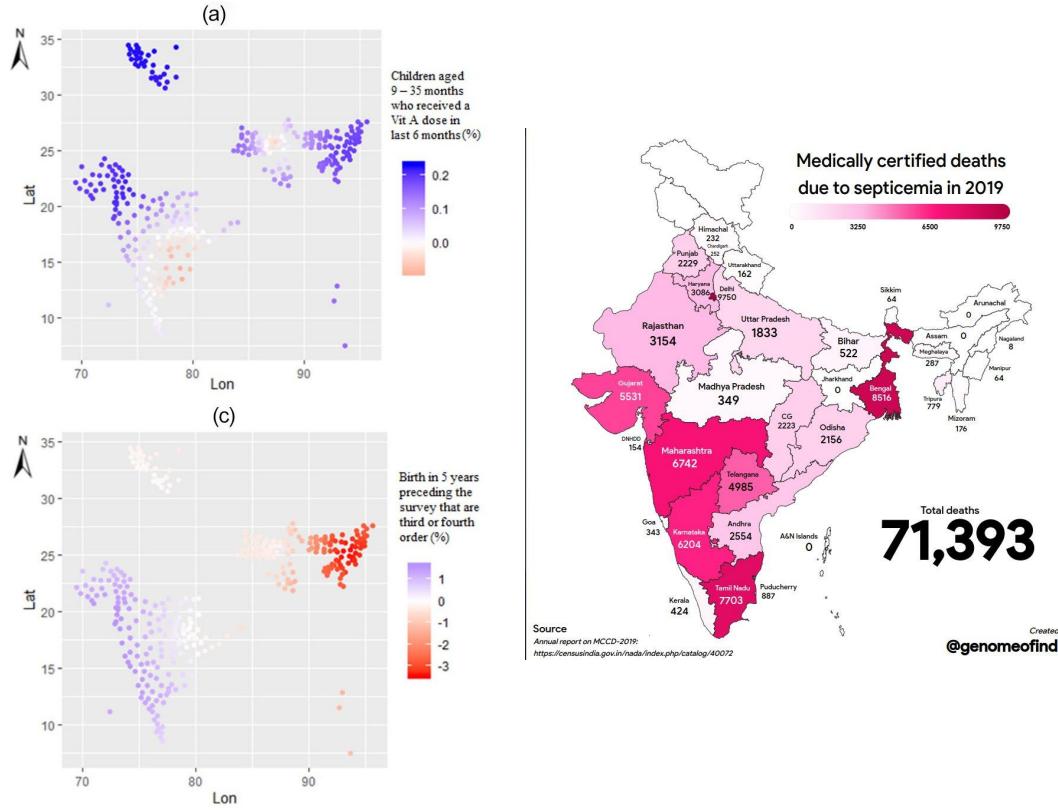
Problem: Is the relationship between height and weight similar across males and females? Is it linear?



[Source](#)

[Source](#)

Multivariate regression for health-related data



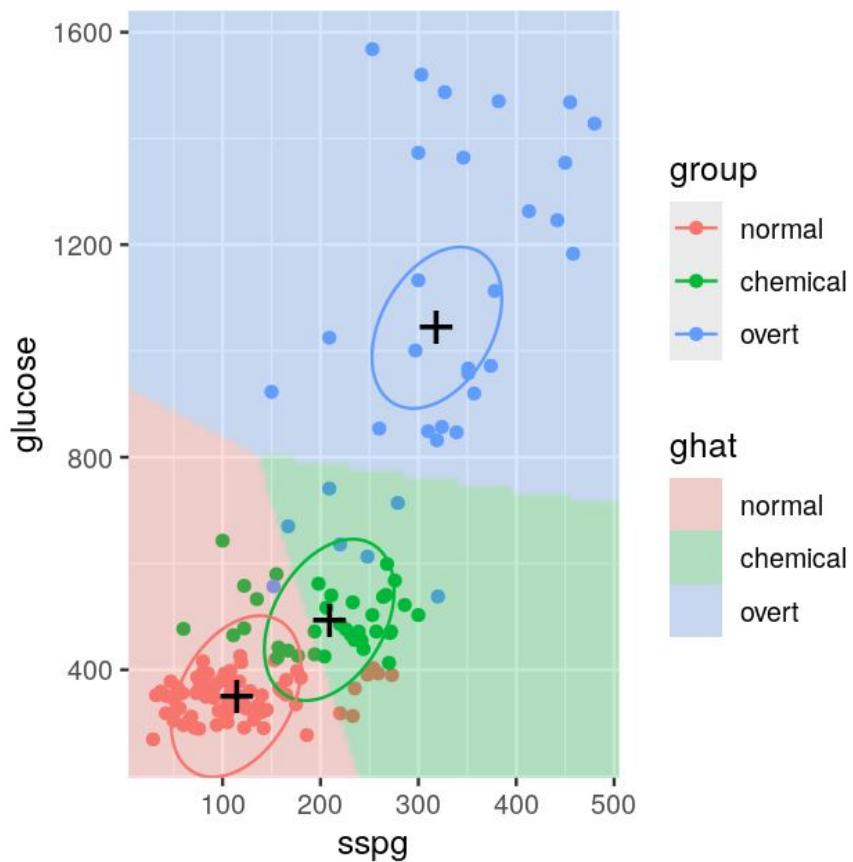
Problem: What are factors associated with disease X in India?

Solution: Geographically aware regression to identify factors and spatial clusters

[Source](#)

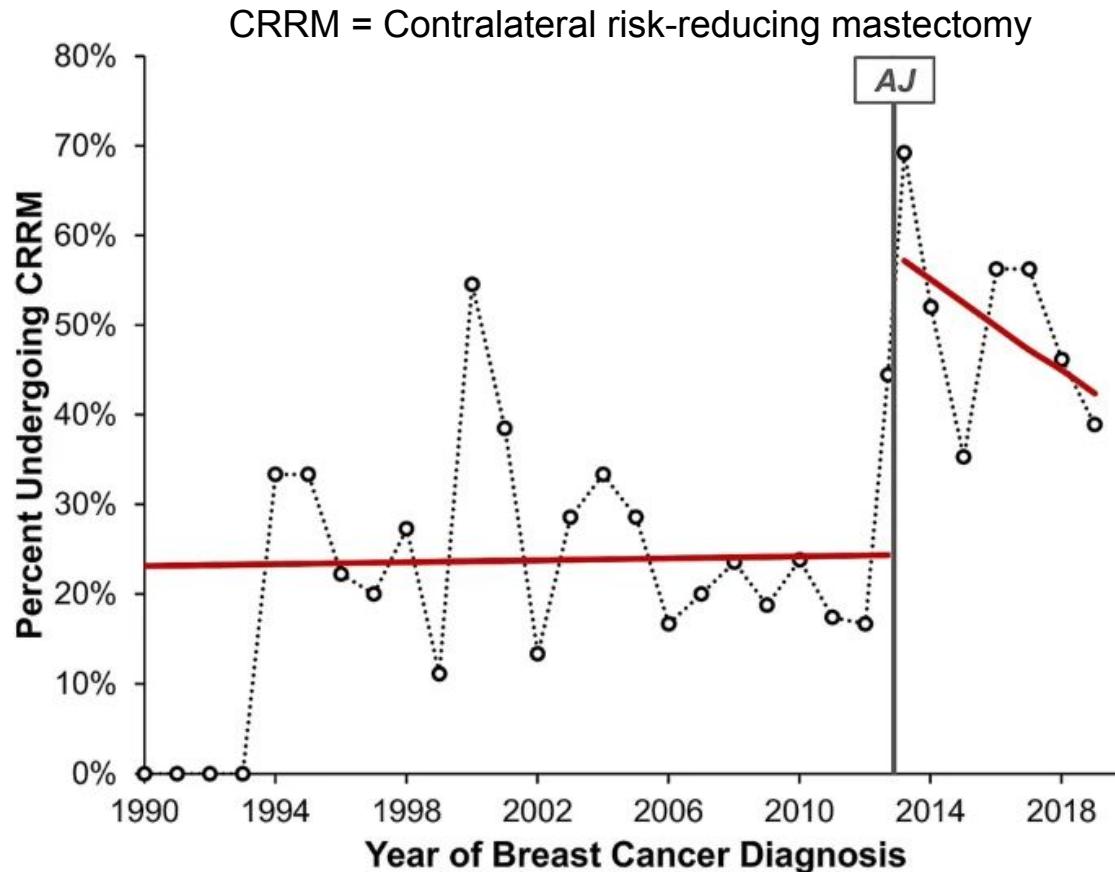
[Source](#)

Classification problems in healthcare



Problem: Predicting the severity of disease (or no-disease)

The Angelina Jolie Effect

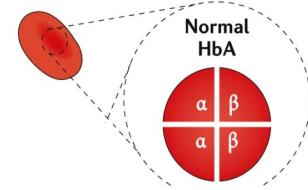


Genomics for health

Person with
HBB/HBB genotype

<i>HBB</i>	CAC	CTG	GAC	TGA	GGA	CTC	CTC
	[purple]	[blue]	[yellow]	[purple]	[orange]	[yellow]	[purple]
	GUG	GAC	CUG	ACU	CCU	GAG	GAG
	[yellow]	[blue]	[purple]	[blue]	[purple]	[yellow]	[yellow]
	Val	His	Leu	Thr	Pro	Glu	Glu

<i>HBB</i>	CAC	CTG	GAC	TGA	GGA	CTC	CTC
	[purple]	[blue]	[yellow]	[purple]	[orange]	[yellow]	[purple]
	GUG	GAC	CUG	ACU	CCU	GAG	GAG
	[yellow]	[blue]	[purple]	[blue]	[purple]	[yellow]	[yellow]
	Val	His	Leu	Thr	Pro	Glu	Glu



Problem: How are DNA mutations identified?

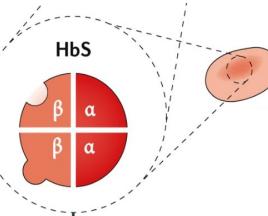
Person
with SCA

β^s allele

	CAC	CTG	GAC	TGA	GGA	CAC	CTC
	[purple]	[blue]	[yellow]	[purple]	[orange]	[purple]	[purple]
	GUG	GAC	CUG	ACU	CCU	GUG	GAG
	[yellow]	[blue]	[purple]	[blue]	[purple]	[yellow]	[yellow]
	Val	His	Leu	Thr	Pro	Val	Glu

β^s allele

	CAC	CTG	GAC	TGA	GGA	CAC	CTC
	[purple]	[blue]	[yellow]	[purple]	[orange]	[purple]	[purple]
	GUG	GAC	CUG	ACU	CCU	GUG	GAG
	[yellow]	[blue]	[purple]	[blue]	[purple]	[yellow]	[yellow]
	Val	His	Leu	Thr	Pro	Val	Glu



Genomics for health

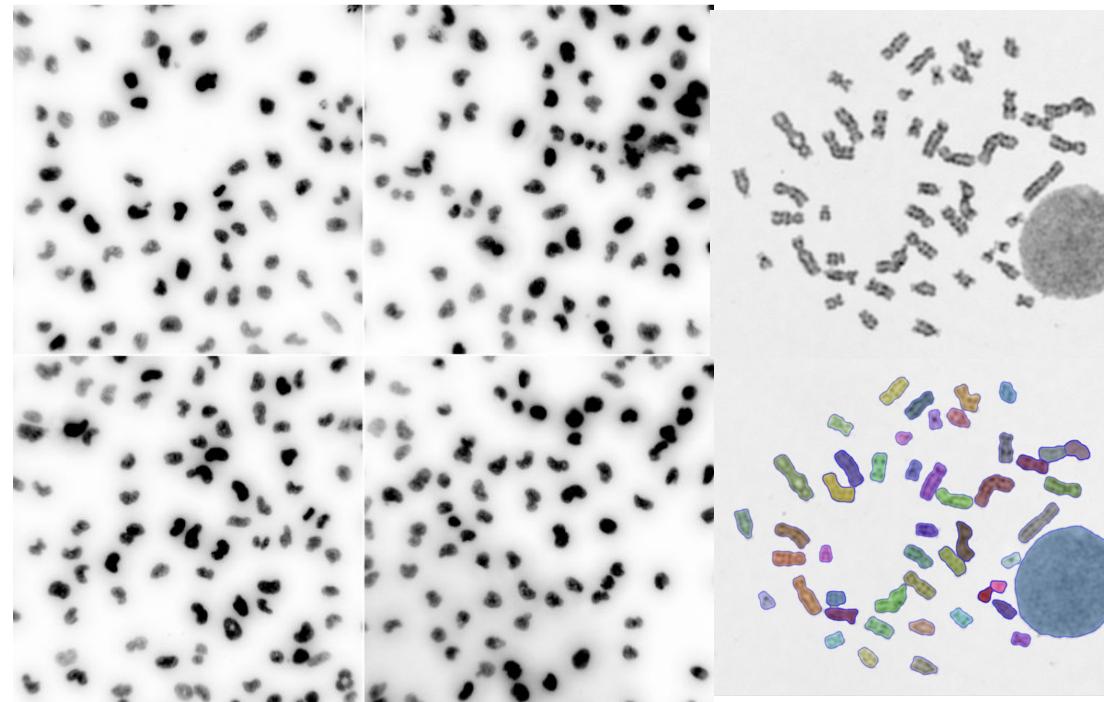
Problem: Which variants are (were) driving the uptick in Covid-19 cases?

Estimated cases (weekly average) in New York City by variant

Estimation based on a multinomial fit to weekly genomic surveillance data deposited to GISAID



Analysis of image data for healthcare



Problem: Extracting features from medical images; Classifying images into grades of diseases

Short(est) introduction to public health informatics

What is public health?

SCIENCE

FRIDAY, JANUARY 9, 1920

CONTENTS

<i>The American Association for the Advancement of Science:</i> —	
<i>The Untilled Fields of Public Health: PROFESSOR C-E. A. WINSLOW</i>	23
<i>The Organization of Research: PROFESSOR H. P. ARMSBY</i>	33
<i>Scientific Events:—</i> <i>Conference of British Research Associations;</i>	



CEA Winslow

THE UNTILLED FIELDS OF PUBLIC HEALTH¹

A SHORT time ago two Yale undergraduates came to my laboratory to consult me in regard to the choice of a career. One of them was a son of a public health administrator of the highest eminence; and they particularly wanted to know something about the field of public health, what it included, what was the nature of the work involved, what were the qualifications required, and what the financial rewards and the more intangible emoluments to be expected by those who

“Public health is the science and the art of preventing disease, prolonging life, and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, the organization of medical and nursing service for the early diagnosis and preventive treatment of disease, and the development of the social machinery which will ensure to every individual in the community a standard of modest duties which their task entails. living adequate for the maintenance of health”

— CEA Winslow, Science (1920)

The ten essential public health services

Assessment

1. Monitor health status to identify community health problems
2. Diagnose and investigate health problems and health hazards in the community

Policy Development

3. Inform, educate, and empower people about health issues
4. Mobilize community partnerships to identify and solve health problems
5. Develop policies and plans that support individual and community health efforts

Assurance

6. Enforce laws and regulations that protect health and ensure safety
7. Link people to needed personal health services and assure the provision of health care when otherwise unavailable
8. Assure a competent public health and personal healthcare workforce
9. Evaluate effectiveness, accessibility, and quality of personal and population-based health services

Serving All Functions

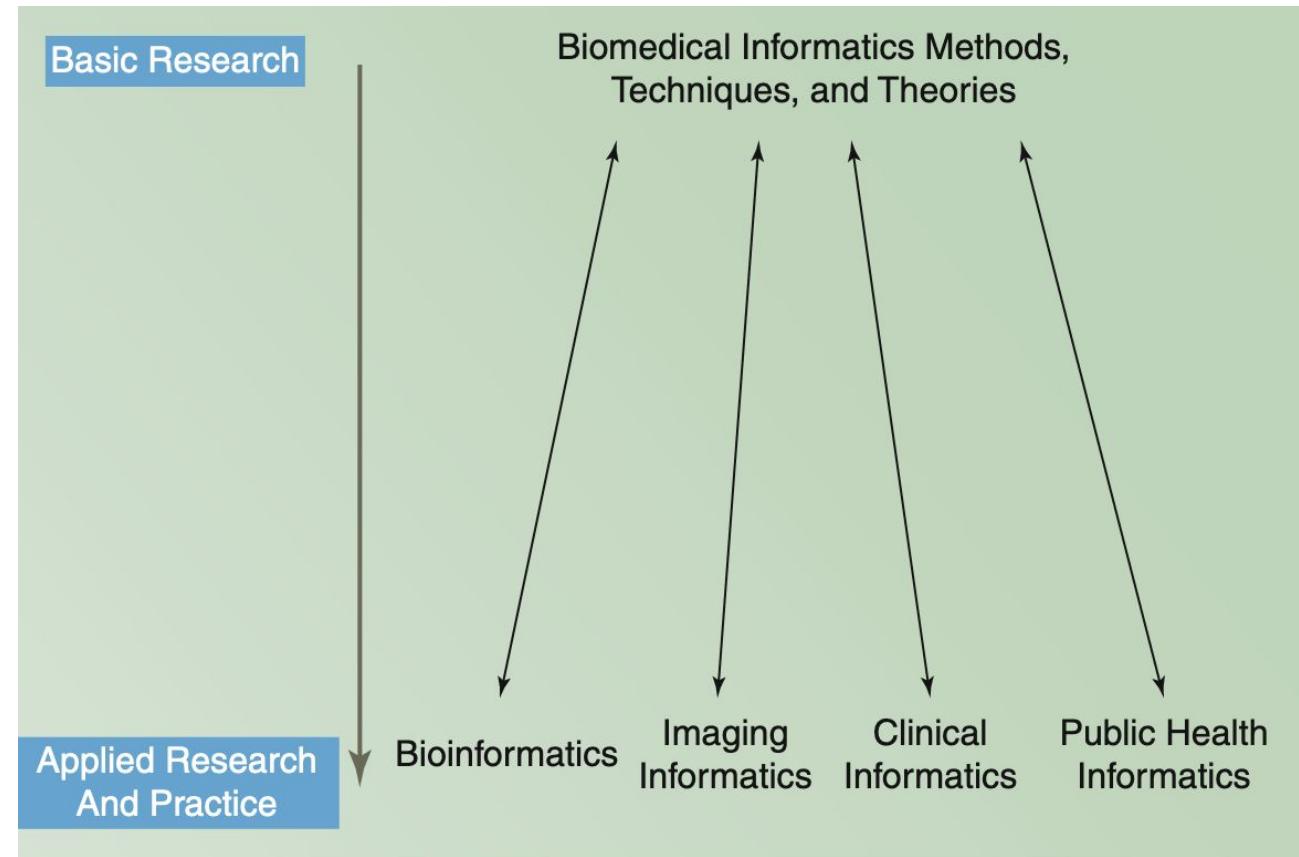
10. Research for new insights and innovative solutions to health problems

[Source](#)

What is health informatics?

Public health informatics =
Systematic application of
data, technology, and
information systems to
public health practice and
research

Goal: Use data to improve
efficiency, accuracy and
diagnosis of health care
needs



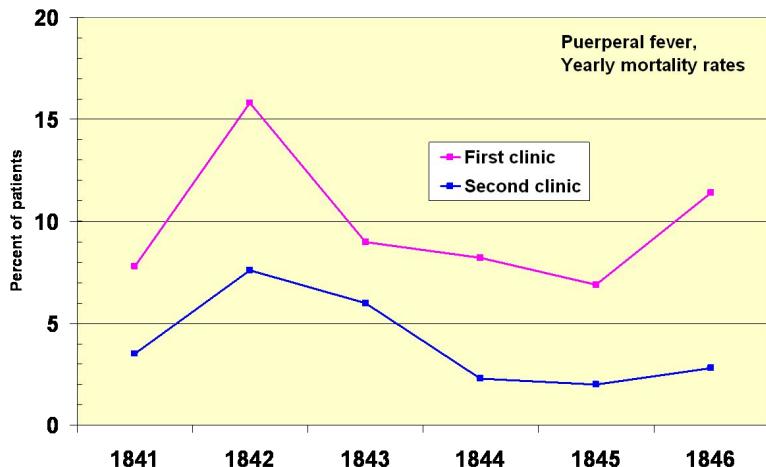
Story time

Blast from the past

Ignaz Semmelweis: Saviour of Mothers



- In 19th century, postpartum infection also called puerperal fever was common cause of death with its reason largely remaining unknown (germ theory had not been proven yet!)
- Ignaz was appointed as an assistant physician at Vienna General Hospital → came across pregnant women “begging” to not be admitted to the “First” clinic
- First clinic mortality rate: 10%, Second clinic (just down the road, more crowded): 4%
- Mortality rate of women giving birth on the road (to avoid the “First”): << 10%



Why is the mortality different across the two clinics?

First clinic had doctors training who would conduct everything; Second clinic had mostly midwives assisting in birth

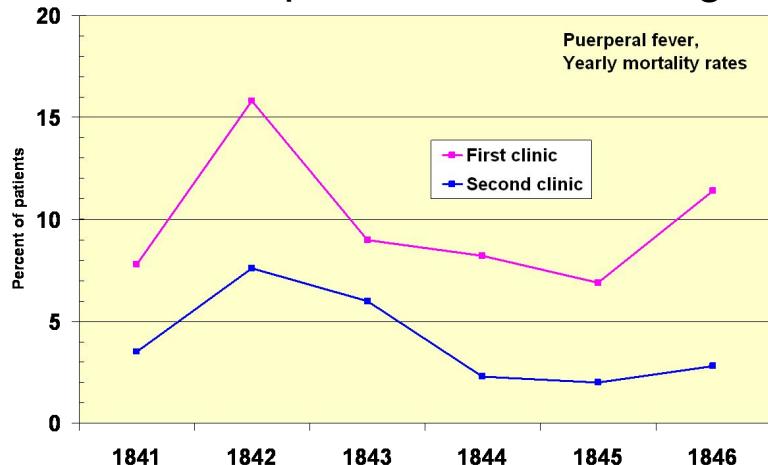
[Source](#)

[Source](#)

Ignaz Semmelweis: Saviour of Mothers

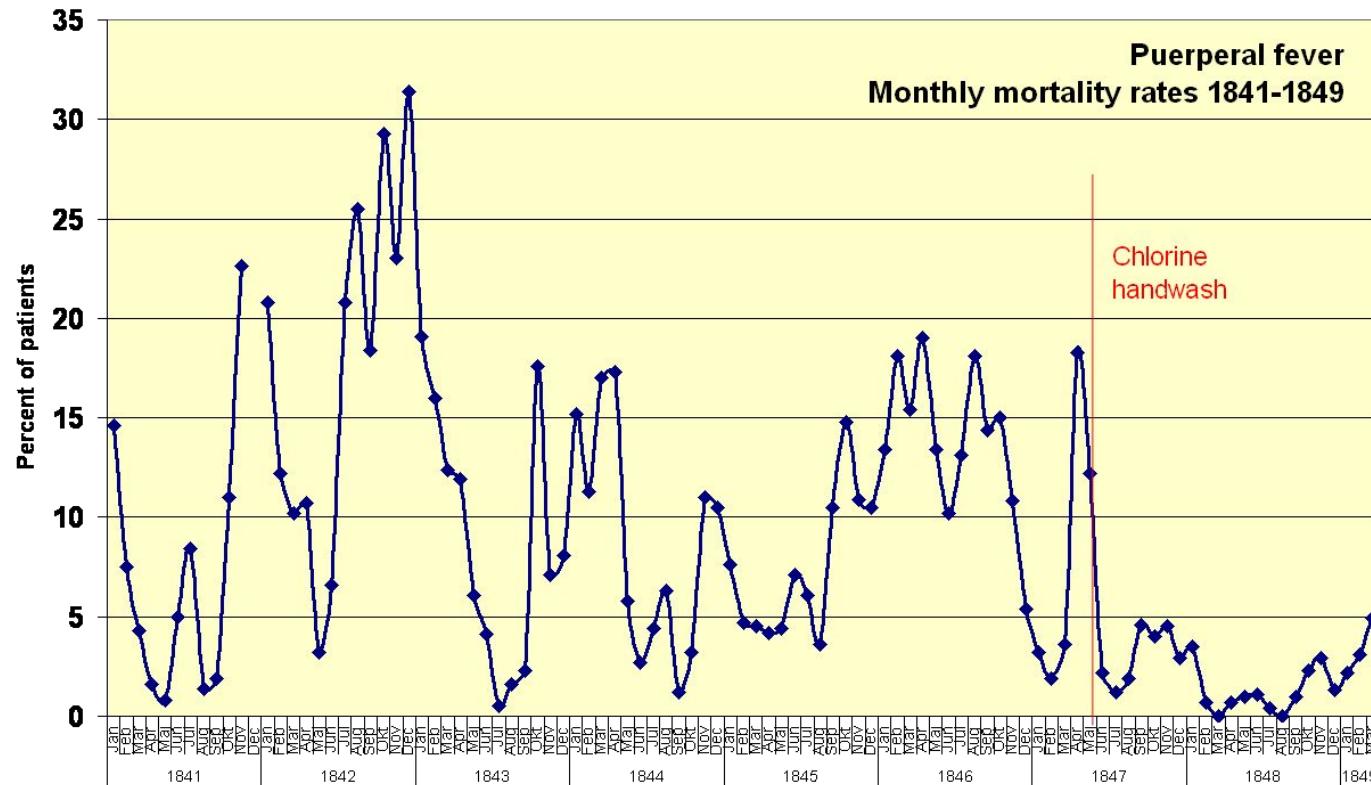


- Semmelweis' own friend died in 1847 after being accidentally poked with a scalpel while conducting an autopsy → his autopsy had similar pathophysiology of that of women dying in “First” clinic
- Doctors in the first clinic would conduct autopsies in the morning and then go and help in deliveries during the afternoon



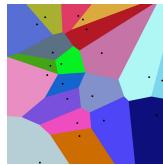
- Hand-washing did NOT exist!
- Semmelweis suggested washing hands with chlorine after any operation/autopsy
- Backlash; driven to insanity → asylum → beaten → died

Ignaz Semmelweis: Saviour of Mothers

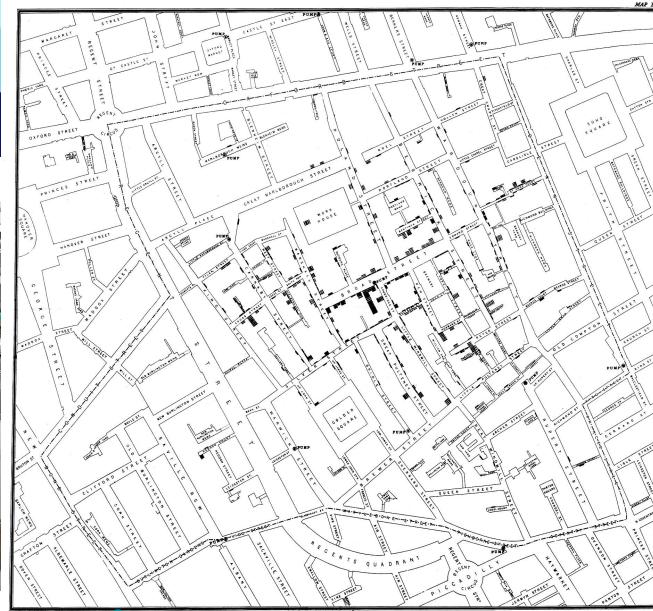


The Cholera outbreak of 1854 on Broad street, London

- 616 people died because of cholera in 1854
- Attended by Florence Nightingale
- John Snow identified the source of outbreak to be a handpump on Broad street
- Chemical and microscopic identification did not prove anything
- But a dot map was convincing → the handle of the pump was disabled → cases declined (but possibly were already declining before his argument)
- The pump was dug next to a cesspit!



John Snow used a dot map to illustrate how cases occurred around this pump → one of the earliest uses of voronoi diagram



Source

The pioneer of data visualization in health

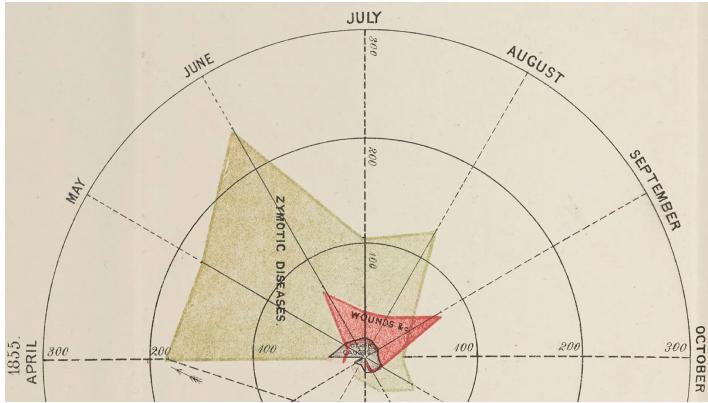
Florence
Nightingale



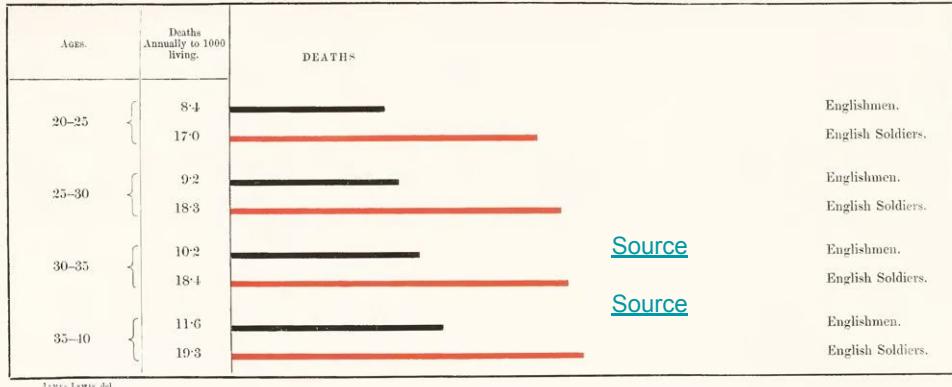
Lady with the lamp.

[Source](#)

Mortality in British East Asia army

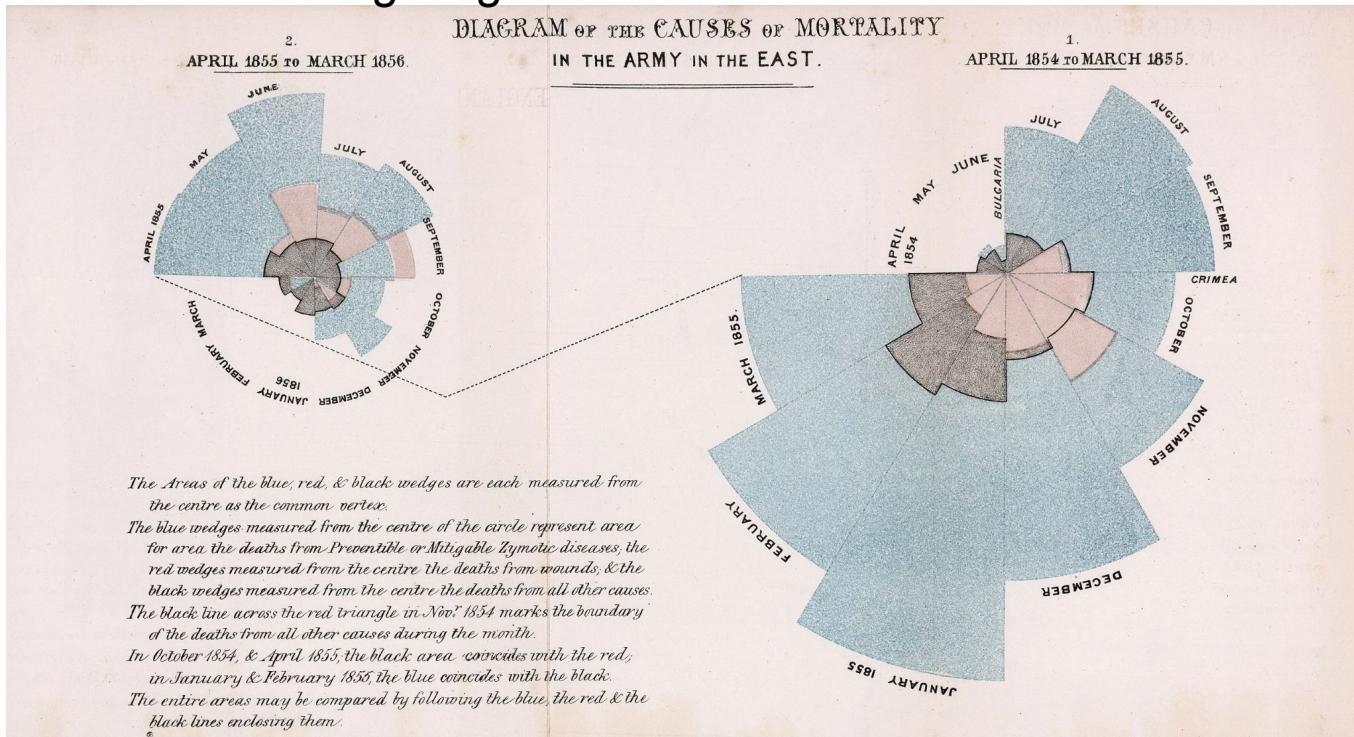


Representing the Relative Mortality of the Army at Home and of the English Male Population at corresponding Ages.



The pioneer of data visualization in health

“Printed tables and all-in double columns, I do not think anyone will read. None but scientific men ever look in the Appendix of a Report. And this is for the vulgar* public.” – Florence Nightingale



Questions?

