Data Analytics in Health Care

Madhavan Mukund

https://www.cmi.ac.in/~madhavan

FDP on Data Analytics in Healthcare

SSN College of Engineering

17 January, 2022

A long history

- Major outbreak of cholera in London in 1854

A COURT FOR KING CHOLERA.

A long history

- Major outbreak of cholera in London in 1854
- Prevailing wisdom was that cholera was due to "bad air"—miasma theory

A COURT FOR KING CHOLERA.

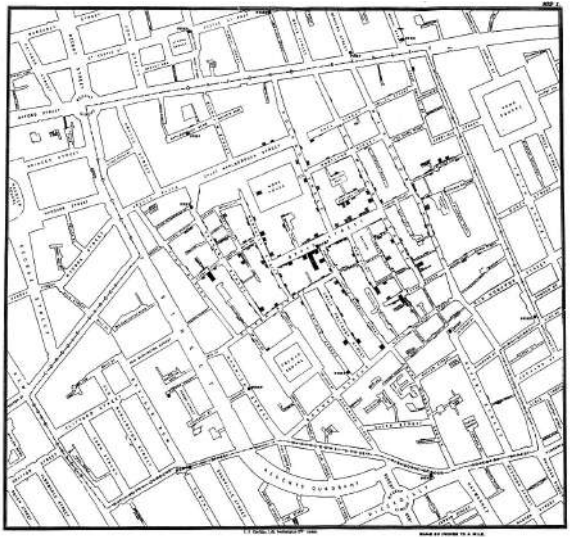
A long history

- Major outbreak of cholera in London in 1854
- Prevailing wisdom was that cholera was due to "bad air"—miasma theory
- Physician John Snow believed it was due to contaminated water
  - Germ theory due to Louis Pasteur still to come, 1861

A COURT FOR KING CHOLERA.

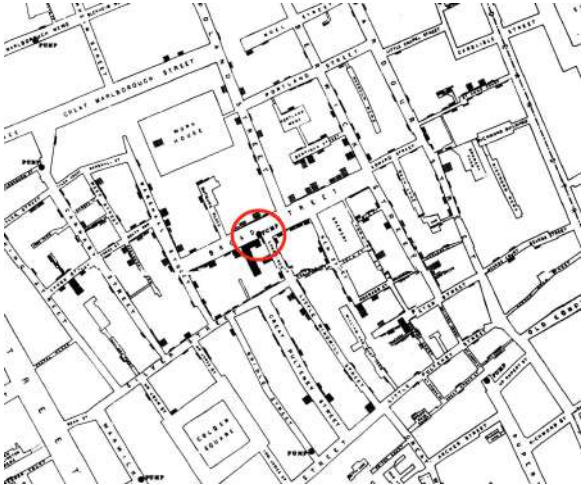
A long history

- Major outbreak of cholera in London in 1854
- Prevailing wisdom was that cholera was due to "bad air"—miasma theory
- Physician John Snow believed it was due to contaminated water
  - Germ theory due to Louis Pasteur still to come, 1861
- Snow plotted the locations where cholera was reported



A long history

- Major outbreak of cholera in London in 1854
- Prevailing wisdom was that cholera was due to "bad air"—miasma theory
- Physician John Snow believed it was due to contaminated water
  - Germ theory due to Louis Pasteur still to come, 1861
- Snow plotted the locations where cholera was reported
- Cluster of points pointed to contaminated water pump on Broad Street



The many facets of Data Analytics and Health Care

The many facets of Data Analytics and Health Care

Data Analytics

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning
- Searching for "patterns"
  - Unsupervised learning

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning
- Searching for "patterns"
  - Unsupervised learning

Health Care

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning
- Searching for "patterns"
  - Unsupervised learning

Health Care

- Understanding disease
  - Risk factors, genetics

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning
- Searching for "patterns"
  - Unsupervised learning

Health Care

- Understanding disease
  - Risk factors, genetics
- Diagnosis
  - Interpreting tests

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning
- Searching for "patterns"
  - Unsupervised learning

Health Care

- Understanding disease
  - Risk factors, genetics
- Diagnosis
  - Interpreting tests
- Treatment
  - Targetted drug delivery

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning
- Searching for "patterns"
  - Unsupervised learning

Health Care

- Understanding disease
  - Risk factors, genetics
- Diagnosis
  - Interpreting tests
- Treatment
  - Targetted drug delivery
- Infrastructure management
  - Resource allocation in hospitals

The many facets of Data Analytics and Health Care

Data Analytics

- Descriptive statistics
- Sampling
- Data visualization, interpretation
- Predictive analytics
  - Mathematical models
  - Supervised learning
- Searching for "patterns"
  - Unsupervised learning

Health Care

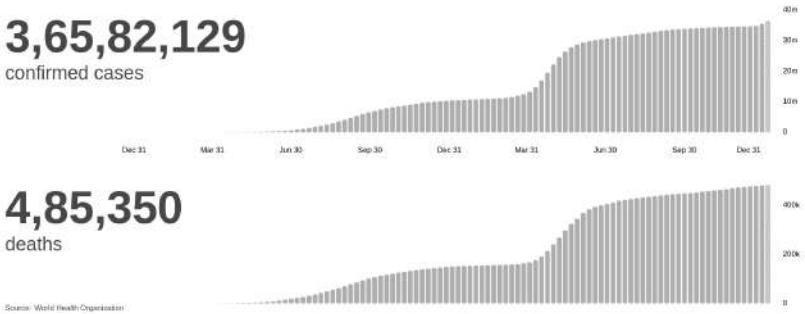
- Understanding disease
  - Risk factors, genetics
- Diagnosis
  - Interpreting tests
- Treatment
  - Targetted drug delivery
- Infrastructure management
  - Resource allocation in hospitals
- Public health
  - Epidemiology, interventions

Data visualization and interpretation

- Covid case data

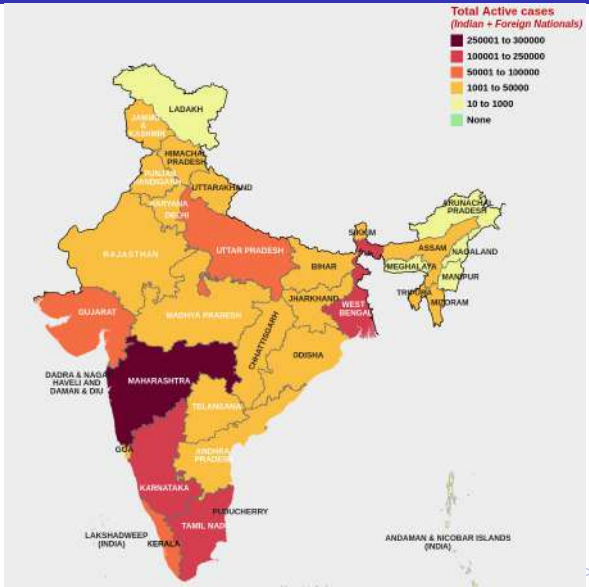
Data visualization and interpretation

- Covid case data
- Progress in time



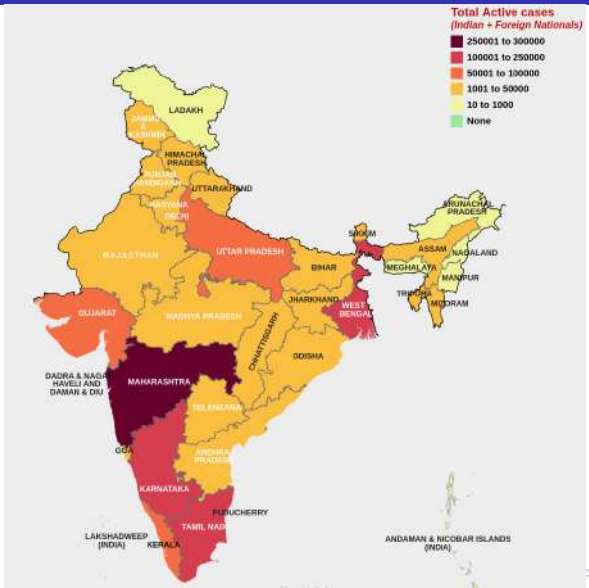
Data visualization and interpretation

- Covid case data
- Progress in time
- . . . and space



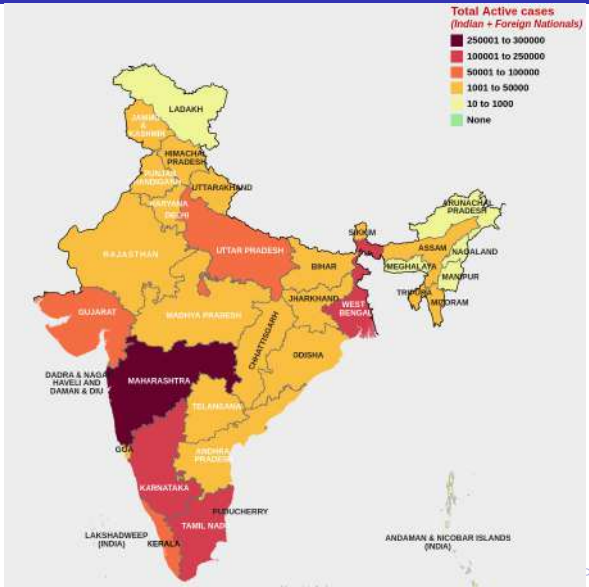
Data visualization and interpretation

- Covid case data
- Progress in time
- . . . and space
- Plan allocation of resources



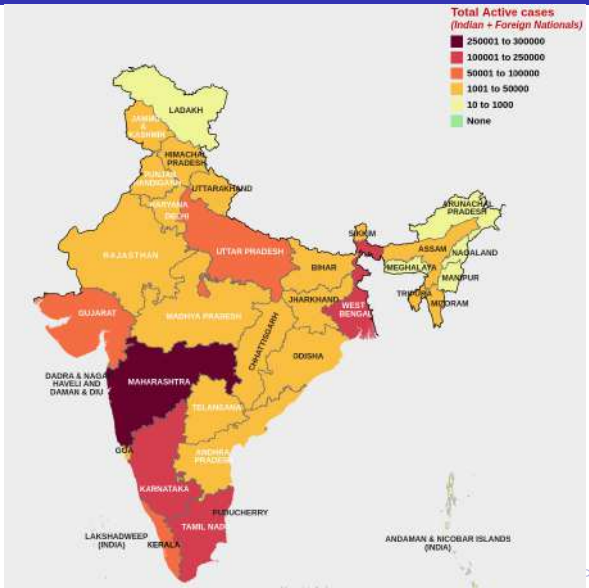
Data visualization and interpretation

- Covid case data
- Progress in time
- . . . and space
- Plan allocation of resources
- More effective triage



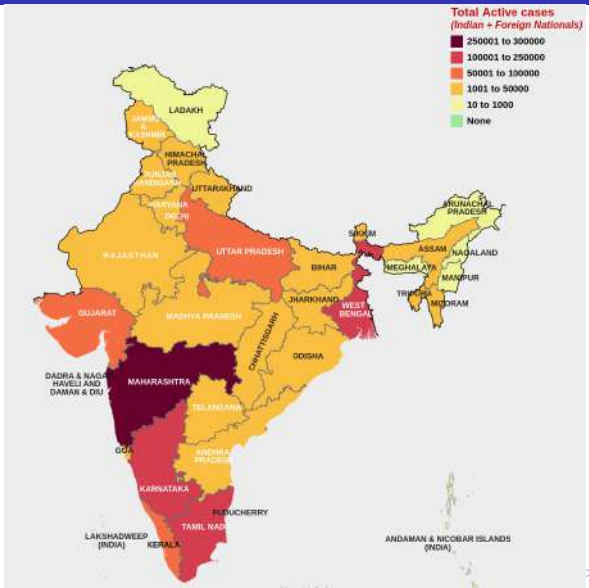
Data visualization and interpretation

- Covid case data
- Progress in time
- . . . and space
- Plan allocation of resources
- More effective triage
- Provisioning specialized equipment
  - ICU beds, oxygen concentrators



Data visualization and interpretation

- Covid case data
- Progress in time
- . . . and space
- Plan allocation of resources
- More effective triage
- Provisioning specialized equipment
  - ICU beds, oxygen concentrators
- Designing testing protocols

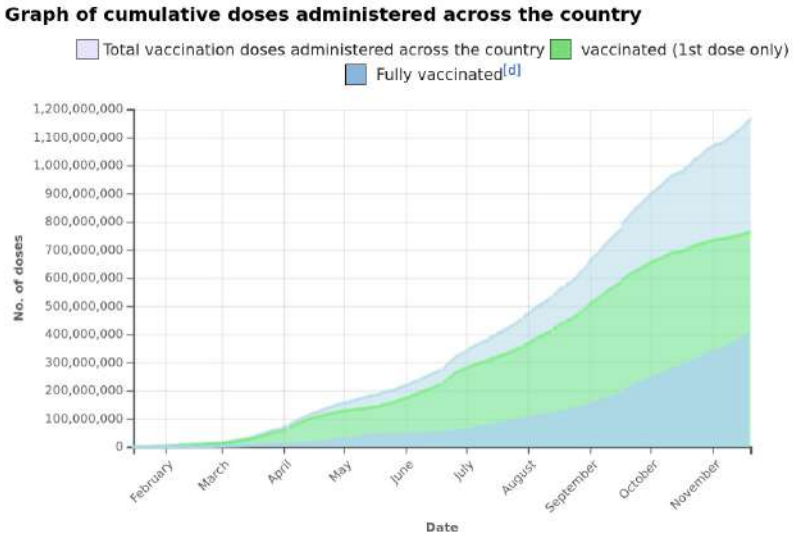


Data visualization and interpretation

- Vaccination data

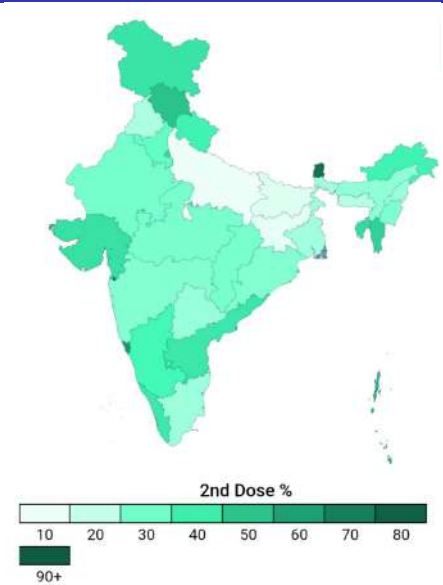
Data visualization and interpretation

- Vaccination data
- Progress in time



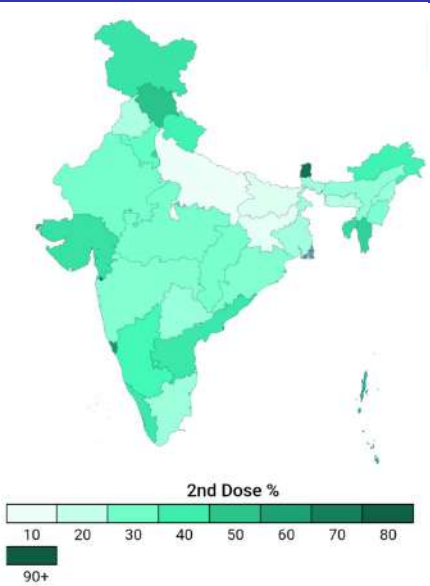
Data visualization and interpretation

- Vaccination data
- Progress in time
- . . . and space



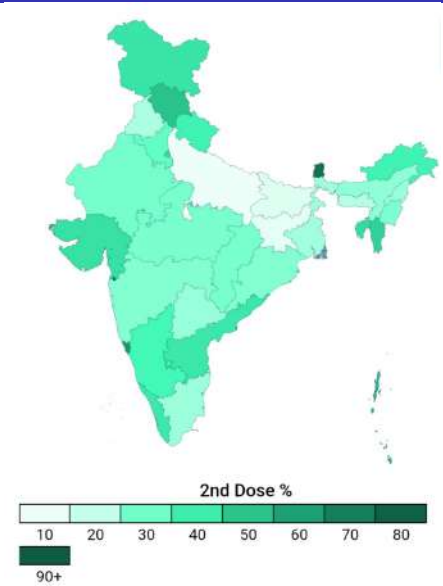
Data visualization and interpretation

- Vaccination data
- Progress in time
- . . . and space
- Slice and dice along different axes
  - Urban vs rural
  - Gender
  - Socio-economic status



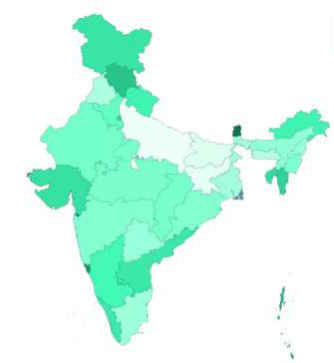
Data visualization and interpretation

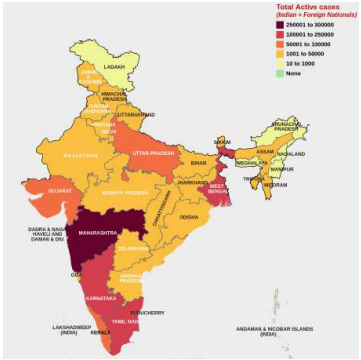
- Vaccination data
- Progress in time
- . . . and space
- Slice and dice along different axes
  - Urban vs rural
  - Gender
  - Socio-economic status
- Data-driven policy
  - Targetted camps
  - Awareness campaigns



Data visualization and interpretation

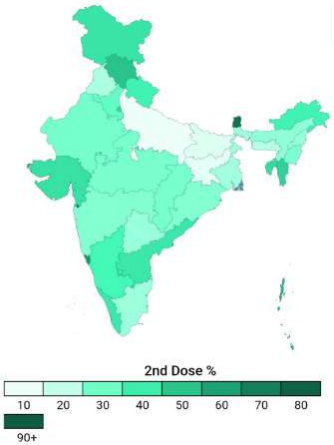
- Vaccine data vs Covid case data

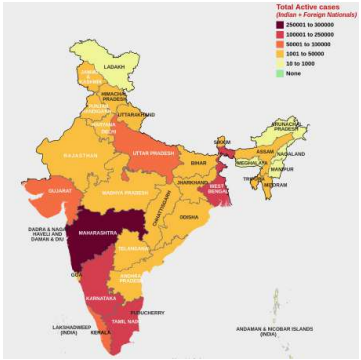




Data visualization and interpretation

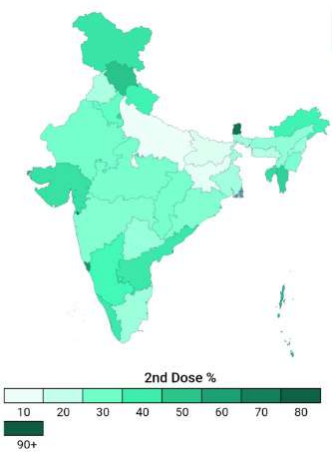
- Vaccine data vs Covid case data
- What's the difference?

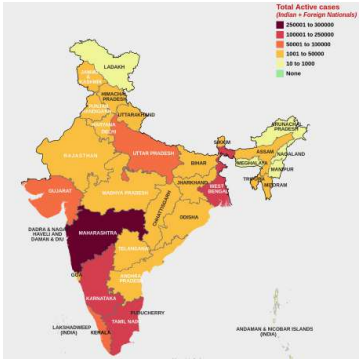




Data visualization and interpretation

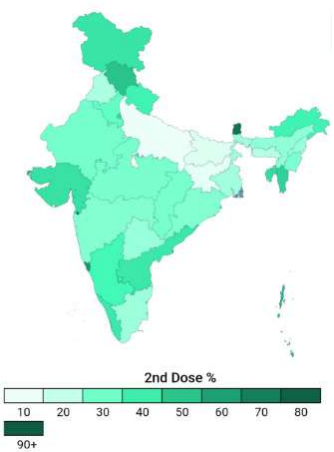
- Vaccine data vs Covid case data
- What's the difference?
- How reliable are the data sources?

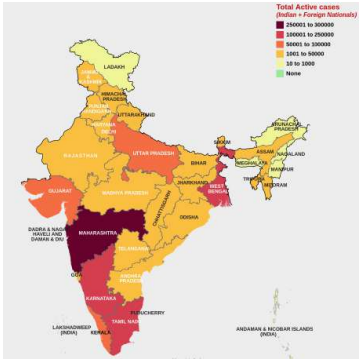




Data visualization and interpretation

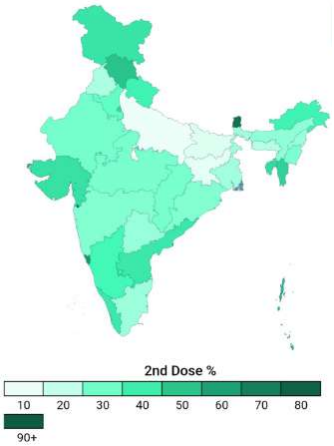
- Vaccine data vs Covid case data
- What's the difference?
- How reliable are the data sources?
- Vaccinations
  - Administered in controlled centres

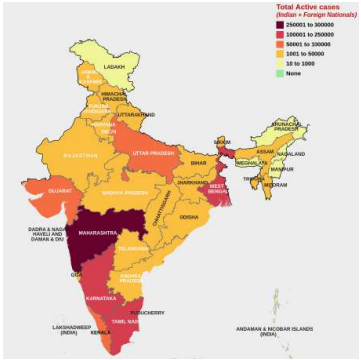




Data visualization and interpretation

- Vaccine data vs Covid case data
- What's the difference?
- How reliable are the data sources?
- Vaccinations
  - Administered in controlled centres
- Covid cases
  - Unreliable self-reporting
  - Uncertain attribution of deaths





Predictive analytics

- Epidemiology

Predictive analytics

- Epidemiology
- Compartmental models

Predictive analytics

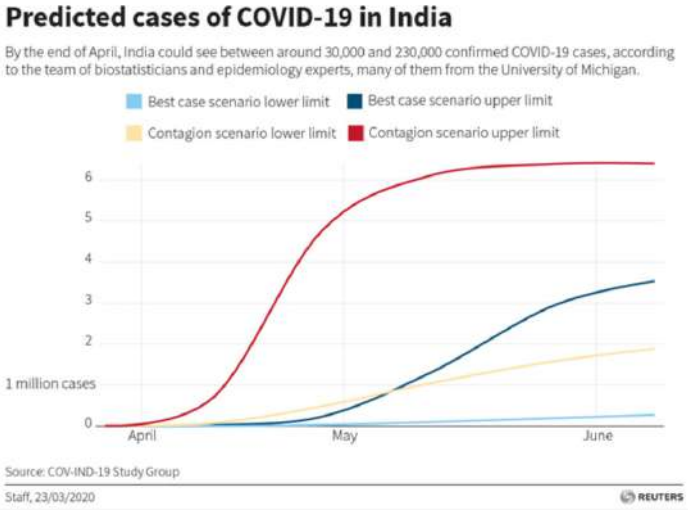
- Epidemiology
- Compartmental models
- Simple assumptions about movement of people, contact, spread of infection

Predictive analytics

- Epidemiology
- Compartmental models
- Simple assumptions about movement of people, contact, spread of infection
- What are the effects of interventions?
  - Lockdowns, travel restrictions
  - Mask mandates

Predictive analytics

- Epidemiology
- Compartmental models
- Simple assumptions about movement of people, contact, spread of infection
- What are the effects of interventions?
  - Lockdowns, travel restrictions
  - Mask mandates
- "Data driven" predictions
  - Speculative, unfalsifiable



Machine learning

- Medical imaging

Machine learning

- Medical imaging
- ML based systems outperform trained radiologists

Machine learning

- Medical imaging
- ML based systems outperform trained radiologists
- Supervised learning
  - Large volume of manually labelled training data
  - Train a classification model — typically a deep neural network

Machine learning

- Medical imaging
- ML based systems outperform trained radiologists
- Supervised learning
  - Large volume of manually labelled training data
  - Train a classification model — typically a deep neural network
- Can apply, in principle to ECGs, EEGs, voice patterns, ...

Machine learning

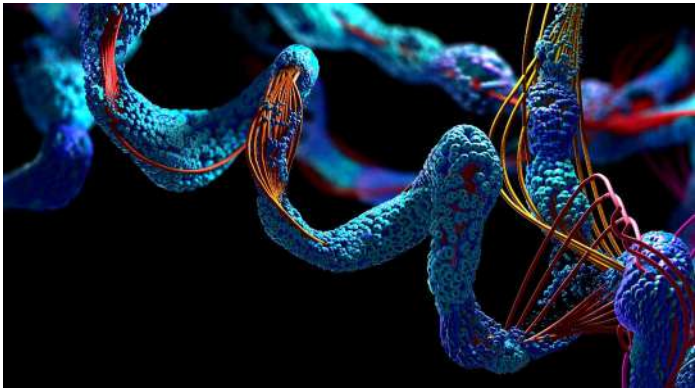
- Medical imaging
- ML based systems outperform trained radiologists
- Supervised learning
  - Large volume of manually labelled training data
  - Train a classification model — typically a deep neural network
- Can apply, in principle to ECGs, EEGs, voice patterns, ...
- How much data is available? Necessary?

Machine learning

- Protein folding

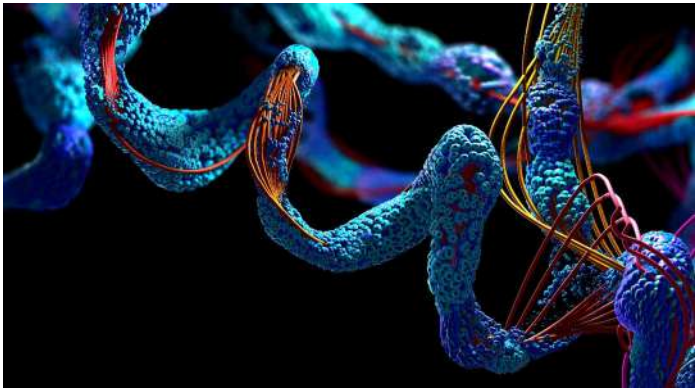
Machine learning

- Protein folding
- Shape of a protein determines how it acts



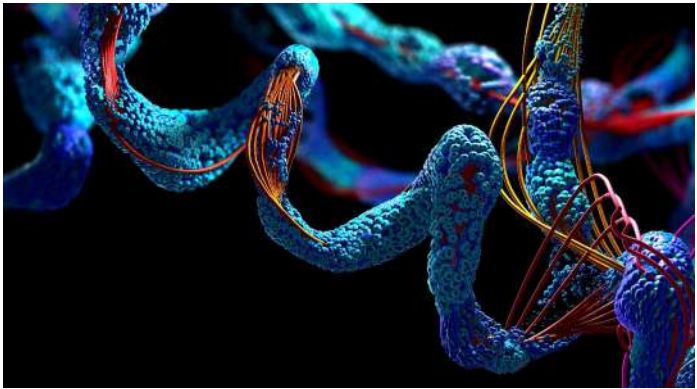
Machine learning

- Protein folding
- Shape of a protein determines how it acts
- Predict the shape from the chemical formula



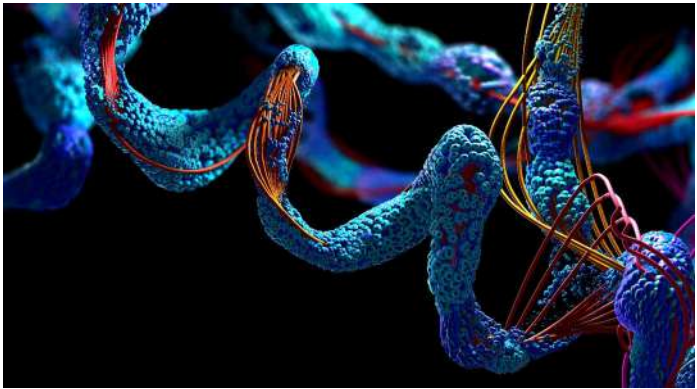
Machine learning

- Protein folding
- Shape of a protein determines how it acts
- Predict the shape from the chemical formula
- AlphaFold 2 achieves record-breaking accuracy (Nov 2020)



Machine learning

- Protein folding
- Shape of a protein determines how it acts
- Predict the shape from the chemical formula
- AlphaFold 2 achieves record-breaking accuracy (Nov 2020)
- But why does a protein fold the way it does?

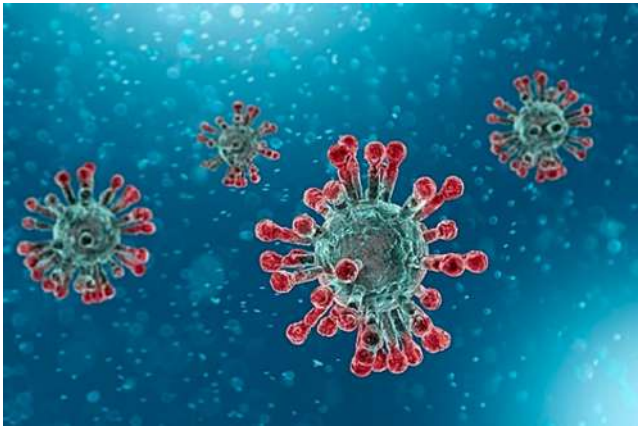


Bio-informatics

- The shape of SARS-CoV-2

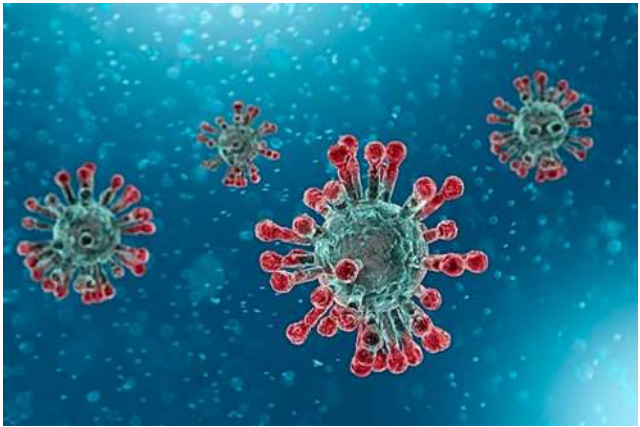
Bio-informatics

- The shape of SARS-CoV-2
- The now-familiar spiky ball



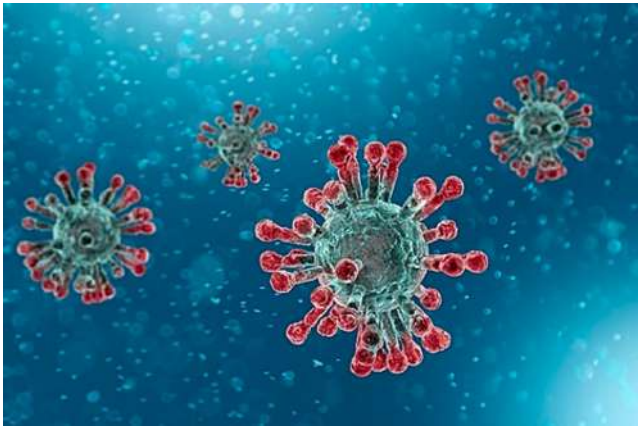
Bio-informatics

- The shape of SARS-CoV-2
- The now-familiar spiky ball
- Genetic structure unravelled very quickly
  - Accurately track mutations



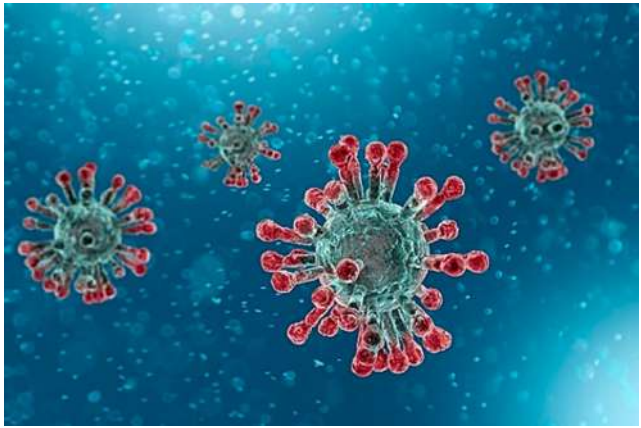
Bio-informatics

- The shape of SARS-CoV-2
- The now-familiar spiky ball
- Genetic structure unravelled very quickly
  - Accurately track mutations
- A new generation of vaccines
  - Rigorous clinical trials in record time
  - mRNA vaccines — adapt better to virus mutations



Bio-informatics

- The shape of SARS-CoV-2
- The now-familiar spiky ball
- Genetic structure unravelled very quickly
  - Accurately track mutations
- A new generation of vaccines
  - Rigorous clinical trials in record time
  - mRNA vaccines — adapt better to virus mutations
- Side effects — why, who, how many?



Smart diagnostics and treatment

- Wearable devices to monitor health parameters

Smart diagnostics and treatment

- Wearable devices to monitor health parameters
- Targetted treatment
  - Insulin pumps

Smart diagnostics and treatment

- Wearable devices to monitor health parameters
- Targetted treatment
  - Insulin pumps
- IOT allows extensive data collection, remote monitoring

Smart diagnostics and treatment

- Wearable devices to monitor health parameters
- Targetted treatment
  - Insulin pumps
- IOT allows extensive data collection, remote monitoring
- Data privacy?

Smart diagnostics and treatment

- Wearable devices to monitor health parameters
- Targetted treatment
  - Insulin pumps
- IOT allows extensive data collection, remote monitoring
- Data privacy?
- Security?

Challenges

- Robust data collection
  - Controversy about authenticity of Covid data
  - Incentives for truthful self-reporting, both individual and governmental
  - Effective sampling, dealing with bias

Challenges

- Robust data collection
  - Controversy about authenticity of Covid data
  - Incentives for truthful self-reporting, both individual and governmental
  - Effective sampling, dealing with bias
- Interpreting the data
  - Many confounding factors
  - How to determine cause and effect?

Challenges

- Robust data collection
  - Controversy about authenticity of Covid data
  - Incentives for truthful self-reporting, both individual and governmental
  - Effective sampling, dealing with bias
- Interpreting the data
  - Many confounding factors
  - How to determine cause and effect?
- Data privacy
  - "Triangulate" medical data to identify individuals
  - Balancing individual privacy and public good
  - Example — contact tracing

Challenges . . .

- ML for diagnosis
  - Clinical data is often sparse
  - Mismatch across data sources
  - How reliable are the conclusions?

Challenges . . .

- ML for diagnosis
  - Clinical data is often sparse
  - Mismatch across data sources
  - How reliable are the conclusions?
- Interpretability and explainability
  - Machine learning models are notoriously opaque
  - Why does this scan indicate a tumour?
  - Tied to trustworthiness, reliability

Challenges . . .

- ML for diagnosis
  - Clinical data is often sparse
  - Mismatch across data sources
  - How reliable are the conclusions?
- Interpretability and explainability
  - Machine learning models are notoriously opaque
  - Why does this scan indicate a tumour?
  - Tied to trustworthiness, reliability
- Security of devices
  - Wearable devices connect to the internet
  - How to prevent malicious interference

Summary

- Covid has brought health care analytics to the forefront

Summary

- Covid has brought health care analytics to the forefront
- Spectacular successes — vaccinations

Summary

- Covid has brought health care analytics to the forefront
- Spectacular successes — vaccinations
- And failures — socio-economic impact of lockdowns

Summary

- Covid has brought health care analytics to the forefront
- Spectacular successes — vaccinations
- And failures — socio-economic impact of lockdowns
- The need for reliable data has never been greater

Summary

- Covid has brought health care analytics to the forefront
- Spectacular successes — vaccinations
- And failures — socio-economic impact of lockdowns
- The need for reliable data has never been greater
- Increasing reliance on data driven diagnostics, decision-making

Summary

- Covid has brought health care analytics to the forefront
- Spectacular successes — vaccinations
- And failures — socio-economic impact of lockdowns
- The need for reliable data has never been greater
- Increasing reliance on data driven diagnostics, decision-making
- Need to build up trust

Summary

- Covid has brought health care analytics to the forefront
- Spectacular successes — vaccinations
- And failures — socio-economic impact of lockdowns
- The need for reliable data has never been greater
- Increasing reliance on data driven diagnostics, decision-making
- Need to build up trust
- There is no shortage of irrational skeptics!