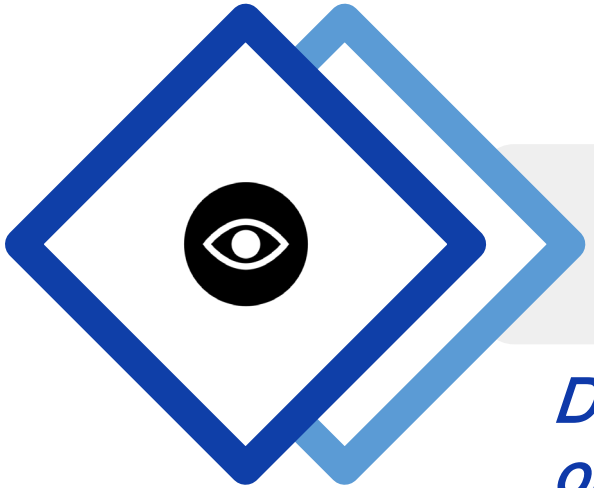
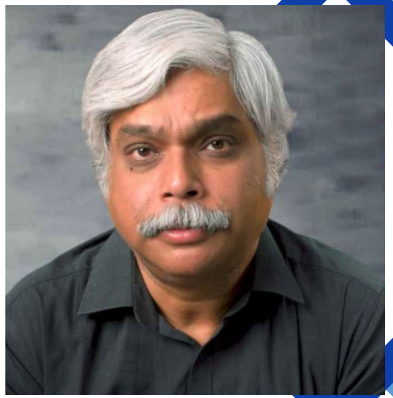


Beyond the Numbers

Discovering the fascinating history of Data Science

Meet Your Speaker



Dr. Abhinanda Sarkar
Academic Director at Great Learning

- Alumnus - Indian Statistical Institute, Stanford University
- Taught at - MIT, Indian Institute of Management, Indian Institute of Science
- Experienced in applying probabilistic models, statistical analysis and machine learning to diverse areas
- Certified as Master Black Belt in Lean Six Sigma and Design for Six Sigma in GE

Video Thumbnail
DO NOT REMOVE

Learning Objectives

Upon completion of this module, you will be able to:

- Describe the historical context of decision-making.
- Illustrate and differentiate between inferential and computational paradigms of data science.
- Explore the evolution of inferential and computational paradigms over the years through key innovations and their drivers.
- Interpret the impact of these innovations on further advancements and developments.

Video Thumbnail
DO NOT REMOVE

Learning Outcomes

You should now be able to:

- Comprehend what data science and analytics means.
- Understand the idea of data-driven decision making.
- Illustrate how data science has evolved over the past century.
- Identify the difference approaches within data science.

Video Thumbnail
DO NOT REMOVE

Agenda

In this session, we will discuss:

- Origin of Decisions
- Paradigms in Data Science
- Evolution of Data Science (From the 1940s till now)

Video Thumbnail
DO NOT REMOVE

Origin of Decisions

Decisions were always **data-driven.**

Let's consider a few situations that early civilizations might have faced:

Decisions are made today by businesses in the same way - but the methods have become more **accurate** and **faster** owing to the evolution of **statistical techniques** & **computing capabilities.**

Video Thumbnail
DO NOT REMOVE

Paradigms in Data Science

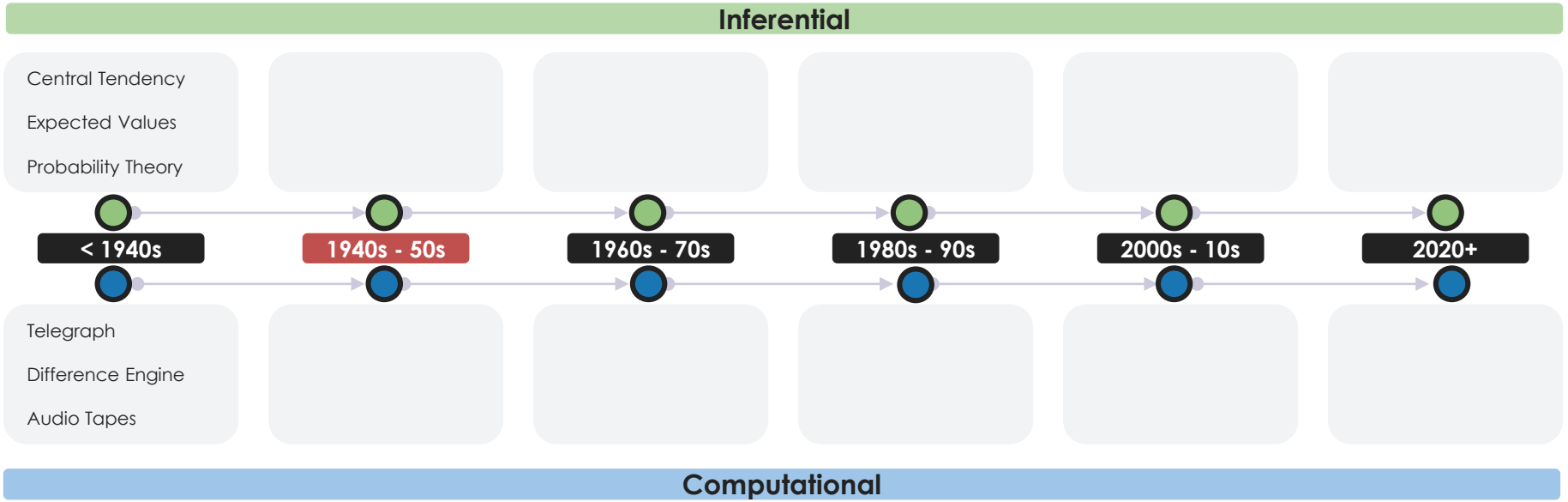
	Inferential	Computational
Focus	Make predictions on population based on sample data	Leverage computational methods and technology to scale insight generation
Methods	Use statistical methods to draw conclusions / inferences from data	Implement algorithms and computational methods to analyse data
Limitations	Representativeness of data	Complexity of algorithms and cost of training large models
Examples	1. Effectiveness of a new medication through randomized trial 2. Impact of a new policy on citizens	1. Weather forecasting based on historical and weather patterns 2. Optimize routing of vehicles to minimize costs

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (Early Stat istics & Computing)

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science



Video Thumbnail
DO NOT REMOVE

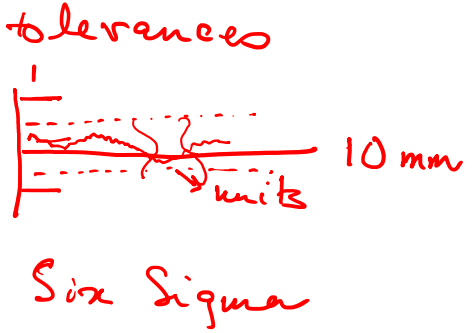
Evolution of Data Science (1940s to 1950s)

Inferential

Industrial Statistics

Use prior knowledge to predict future uncertainties

- Control Chart Theory - Dr. Shewhart @ Bell Labs
- Telephone manufacturing - Quality Control Process
- Led to economic benefits - High quality with efficiency
- Lower wastage + Higher product quality



Sampling Theory

Make inferences about a population, using a sample

- Frederick Taylor - Father of Scientific Management
- Manpower Productivity Assessments
- To improve manufacturing processes + Efficiency
- Now - Advertising by Google/Meta - Target audiences

ANOVA

Compare the means of 3+ groups - Evidence of difference

- Ronald A Fisher - Statistician & geneticist
- Analyze experiments in Agriculture
- Effect of different fertilizers >> Differences in yields
- Now - Market research, finance, quality control

Video Thumbnail
DO NOT REMOVE

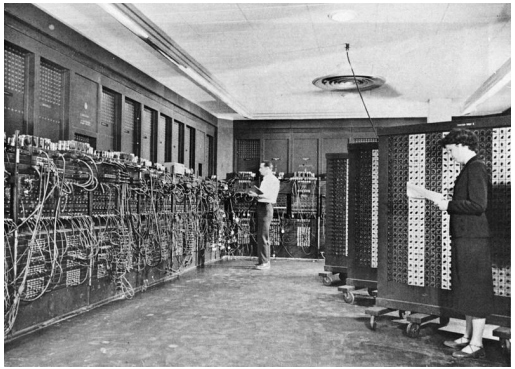
Evolution of Data Science (1940s to 1950s)

Computational

Digital Computers

Electronic device which can do math and logical calculations

- **Electronic Numerical Integrator & Computer (ENIAC)**
- Developed in the US - World War II - Artillery Firing Tables
- Finding trajectories for different types of guns
- Type of ammo + external conditions (temp & wind)



Monte Carlo Methods

Run simulations with random inputs to arrive at conclusions

- Origin - Monte Carlo Casino – Chance-based games
- Developed during the Manhattan Project - 1940s
- Simulate the behaviour of neutrons in a nuclear reactor
- Now - Predict weather patterns, financial markets

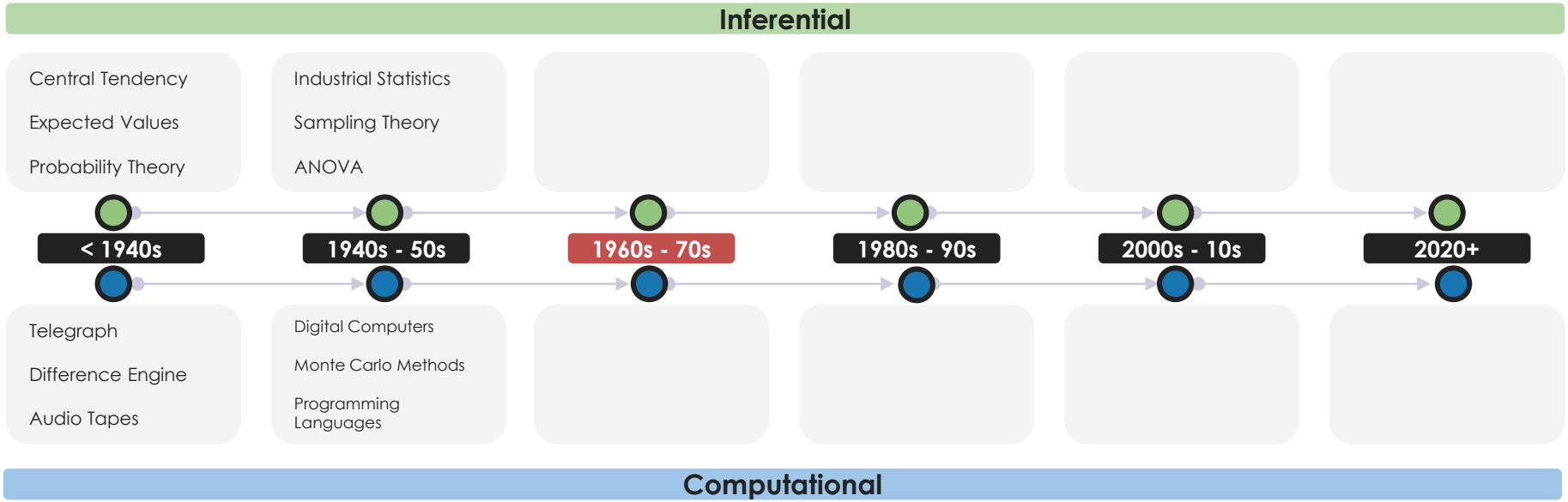
Programming Languages

To have a computer understand instructions & execute them

- Fortran – The 1st programming language created
- Made for scientific and engineering calculations
- Led to the development of the World Wide Web
- Revolutionized communication & learning

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science



Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (1960s to 1970s)

Inferential

Non-Parametric Methods

Rely on ranking/ordering of data rather than the distribution

- Frank Wilcoxon - Wilcoxon Rank-sum test
- Effect of store ambiance on customer behavior
- Sales difference b/w 2 groups in 2 different stores
- More effective marketing and pricing strategies

Decision Theory

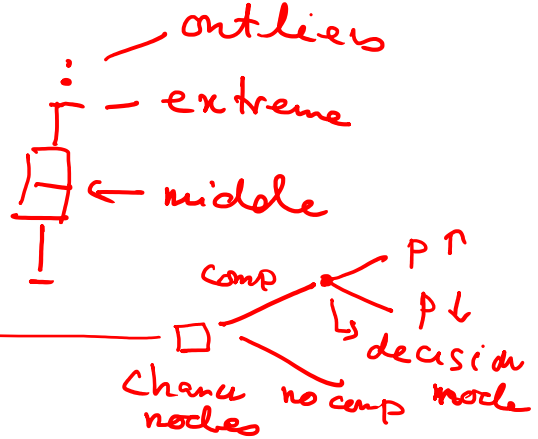
Assign probabilities to different outcomes to make a decision

- Howard Raiffa - Economist - Negotiation Processes
- Government agencies - achieve favorable outcomes
- Identify optimal strategies using decision trees
- Improved negotiation - labor, international trade

Robust Statistics

Provide accurate results despite outliers/extreme values

- John W Tukey - Statistician - Contributed to EDA
- Improve QC process - Identify and remove outliers
- Box plots - identify outliers and variation
- Manufacturing better products, lesser defects



Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (1960s to 1970s)

Computational

Operating Systems

A software that manages resources & apps in a computer

- General Motors are responsible for creating the 1st OS
- GM-NAA I/O - designed for their **IBM 704 mainframe**
- To manage hardware and use them efficiently
- Today - Windows, MAC, Android, and Linux are everywhere

Databases & Storage

Store, Organize & Query large amounts of data quickly

- Charles Bachman - IBM - 1st DBMS ever created
- Businesses wanted databases to be standardized
- **Co**mmon **B**usiness **O**riented **L**anguage - COBOL
- Laid the roots for the creation of MySQL in 1995

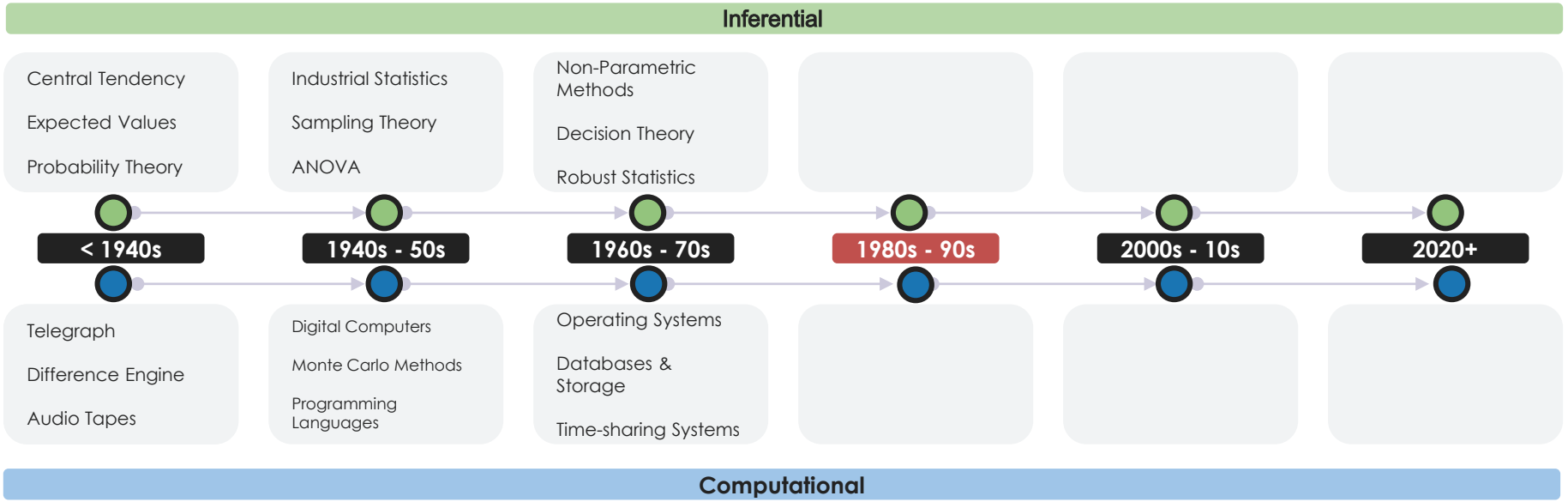
Time Sharing Systems

Many users can access a computer at the same time!

- **C**ompatible **T**ime-**S**haring **S**ystem (CTSS)
- Created by MIT to access one IBM computer
- The seed thought for modern networking systems
- Today's cloud existences based on these ideas

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science



Video Thumbnail
DO NOT REMOVE

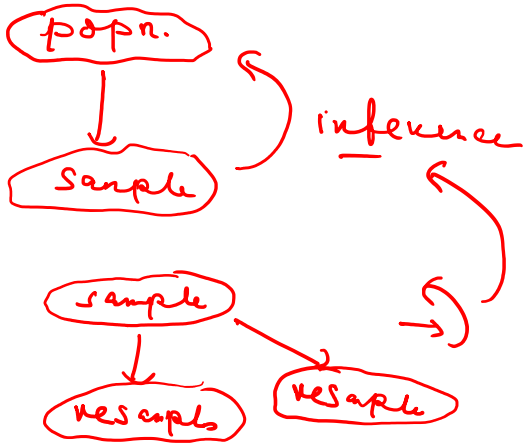
Evolution of Data Science (1980s to 1990s)

Inferential

Resampling Methods

Simulate multiple datasets from original data for analysis

- Bradley Efron- Bootstrap technique
- Market Research survey – Calculate the uncertainty of data
- Estimate sampling distribution + hypothesis testing
- More accurate estimates + better decision making



Generalized Linear Models

Analyze data where outcome is not normally distributed

- John Nelder + Robert Wedderburn
- Insurance claim modeling - Identify risk exposure
- Model different types of response variables
- More accurate + flexible modeling of several data types

Model Selection Techniques

Selecting best mathematical model for a process

- Akaike - Akaike Information Criterion (AIC)
- Demand forecasting - Accurate predictions
- AIC - Evaluate model - Fit, explainability, accuracy
- Better model selection, improved prediction accuracy

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (1980s to 1990s)

Computational

Personal Computers

Small, lightweight, affordable - used by a single person

- Altair 8880 - 1st PC - 1975 – A company called MITS
- This was primitive - **Apple II in 1977** made PCs popular
- Mostly used by hobbyists & technicians
- You could work & collaborate from anywhere

Object-Oriented Programming

An abstract entity with its own set of properties & functions

- Popularized by C++ and Java
- Revolutionized software development
- Flexible, Modular, Reusable & Easy to maintain codes
- Browsers, apps, games - Impossible w/o OOP

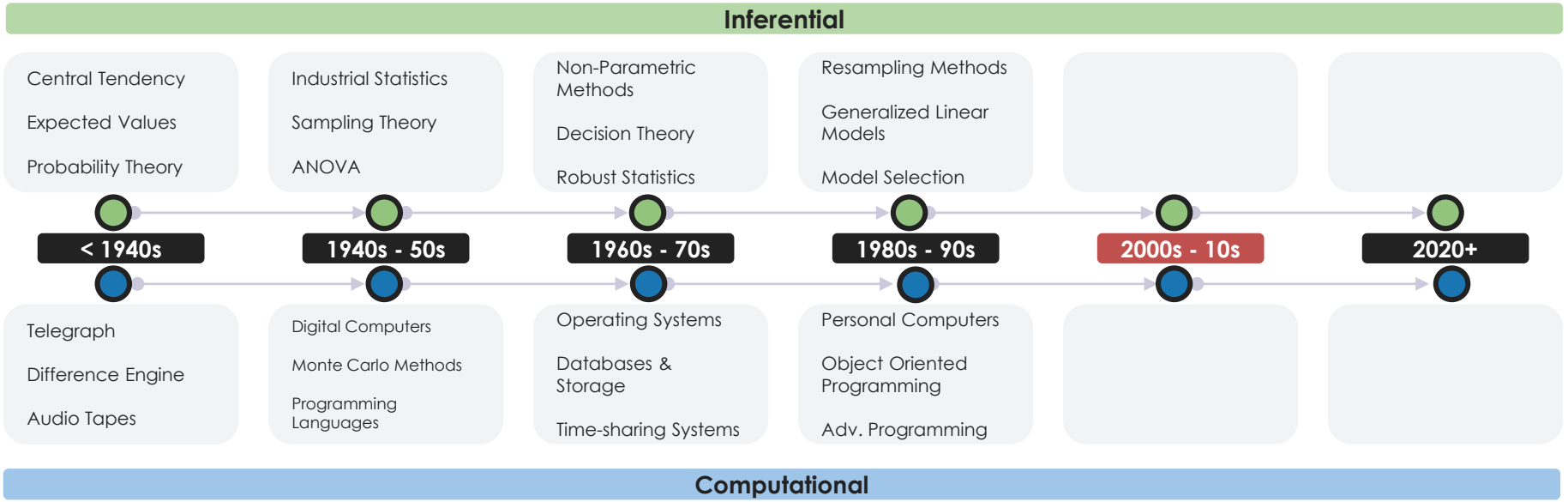
Advanced Programming Languages

High level - data structures - loops - objects - conditions

- Earliest was Fortran - Created in the mid-50s
- Laid the foundation for Java, Ruby, and Python
- Able to create complex applications - At scale
- Democratization of programming - Anybody can Code!

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science



Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (2000s to 2010s)

Inferential

Bayesian Networks & Graphical Models

Relationship b/w variables in a dataset using graphs

- Judea Pearl - Directed Acyclic Graphs - Turing award
- Healthcare - Chances of the patient having a disease
- Easily interpret complex relationships b/w variables
- Accurate diagnosis + improved outcomes, decisions

Causal Inference

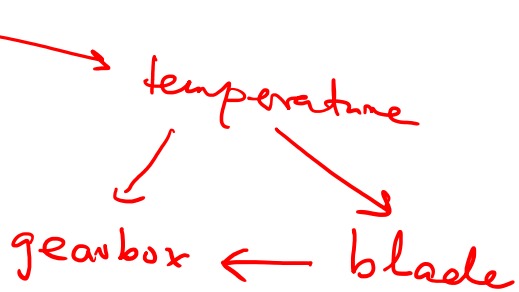
Is change in one variable changing the other?

- Donald Rubin - Rubin Causal Model
- Education - Summer program on student outcomes
- Using outcomes to represent causal relationship
- Better understanding of variables + decision making

Open Science Movement

Making research accessible, collaborative, transparent

- Eli Lilly - Open Innovation Drug Discovery (OIDD)
- Pharma - Develop new drugs and treatments
- Greater collaboration, transparency, reproducibility
- Faster drug development + therapies



Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (2000s to 2010s)

Computational

Artificial Intelligence

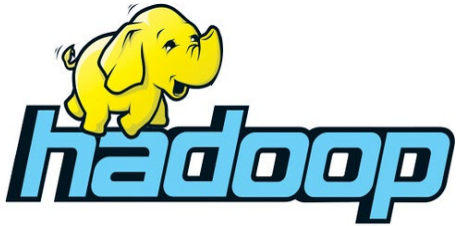
Machines responding/doing tasks at human-level intelligence

- Alan Turing - Machine Intelligence - Imitation Game
- Frank Rosenblatt - "Built the Perceptron" - late 50s
- 2012: Geoffrey Hinton - "Deep Neural Networks"
- 2016: AlphaGo defeats Human Go Champion

Big Data

Massive digital information generated every second

- Nutch Search Engine - Optimize speed of search
- Doug Cutting, 2005: **Hadoop** (son's toy elephant)
- Paved the way for in-memory computing: Spark
- Big Data Analytics - Started with Google & Facebook



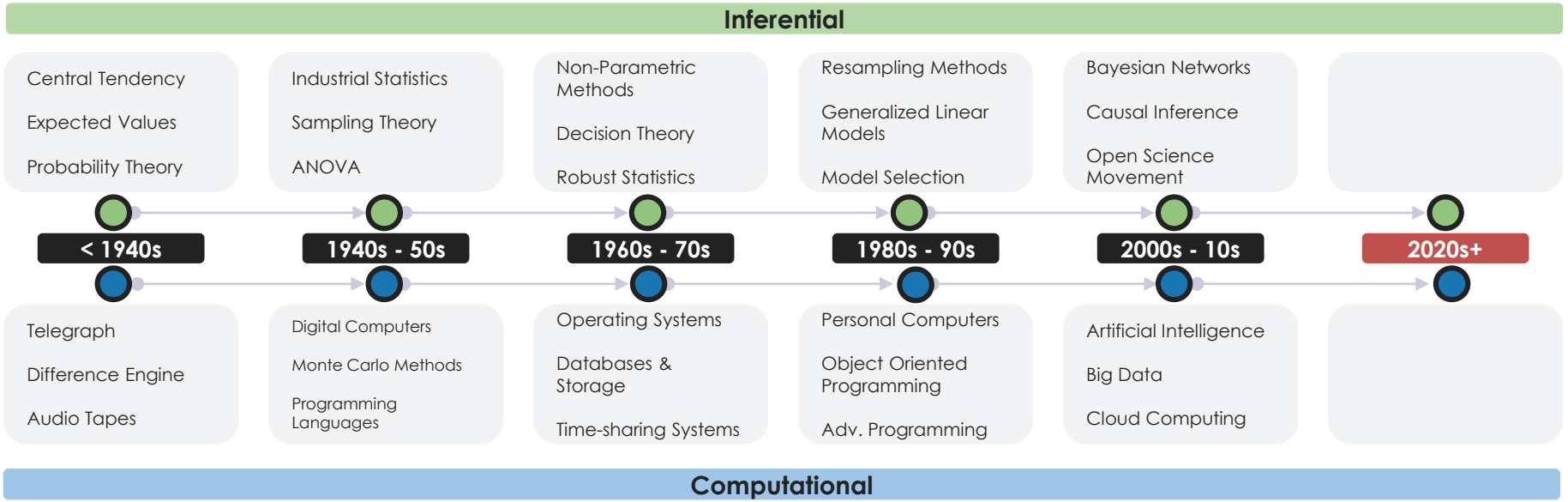
Cloud Computing

Computing power & resources for everyone, on-demand

- Coined by Eric Schmitt - 2006, Google CEO
- 2006: **A**mazon **W**eb **S**ervices, the 1st cloud provider
- Followed by Microsoft Azure and Google Cloud
- Grew due to the reduction in the cost of computing

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science



Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (2020s+)

Inferential

Interdisciplinary Approaches

Knowledge from multiple disciplines for problem solving

- Tesla - Advances in battery + electric motor tech
- Model S - Range of 400 km in a single charge
- Accelerated transition from fossil fuels
- Innovative solutions to complex problems

Newer Causal Inference methods

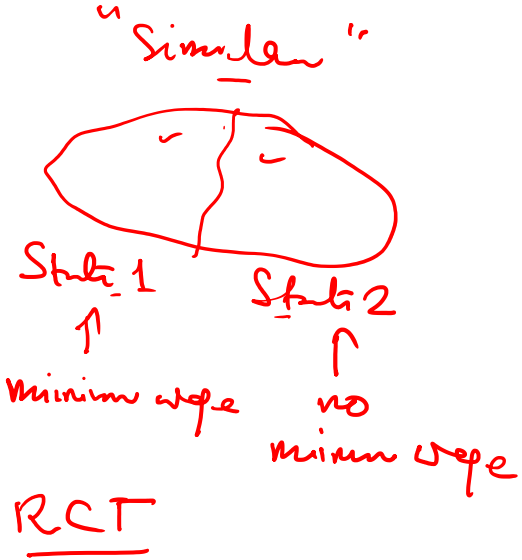
Making causal inference accurate and reliable

- Amazon - Personalized Marketing Campaigns
- Recommend products likely to be purchased
- Causal Inference methods - analyse user behavior
- Increased sales, customer satisfaction

Natural Experiments

Observe events naturally occuring w/o manipulating factors

- Journal of Public Economics - Study of policy impact
- Effectiveness of Public health interventions
- Impact of business closure due to the pandemic on jobs
- Investigate complex phenomena + precise conclusions



Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (2020s+)

Computational

Blockchain

Share information - secure, transparent, & tamper-proof

- Created for the proposal of a Virtual Currency System
- 2008: **BitCoin** - Underlying tech was Blockchain
- The concept is a threat for Traditional Banking Systems
- Extreme Security + Low Fees (No Central Authority)

Price of Bitcoin in USD

Edge Computing

Compute directly at the source of data, instead of remote

- Took shape in the early 2000s - **I**nternet **o**f **T**hings (IoT)
- Tesla's advancements with Autonomous Vehicles
- Opportunities: Real time monitoring & analysis (Medical Devices, Defence, Smart Homes)

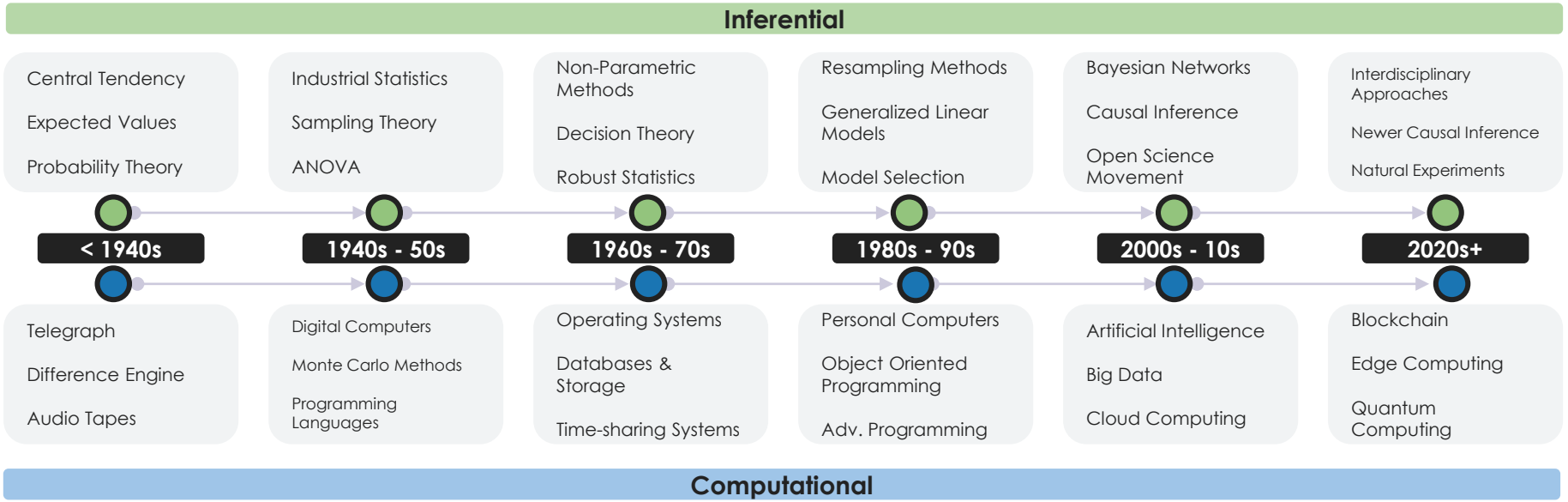
Quantum Computing

Use the principles of quantum physics to compute

- 1st built in 1998 - Los Alamos Laboratory New Mexico
- Impact areas: Cryptography, Chemistry & Optimization
- In early stages, a lot of opportunities are still theoretical and under experimentation

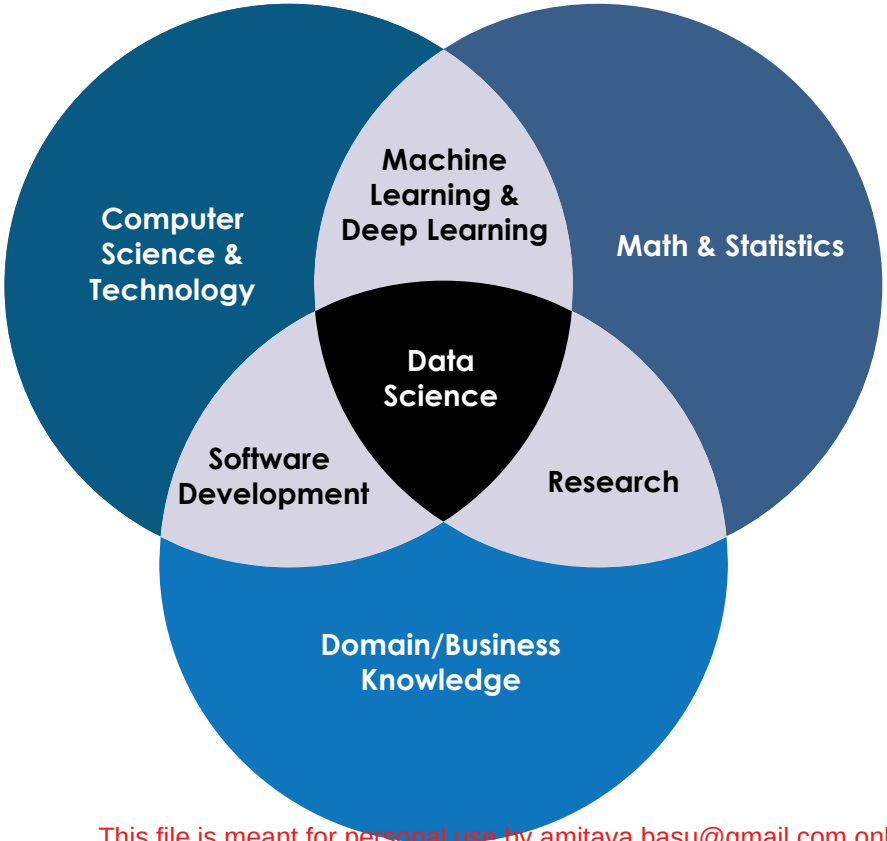
Video Thumbnail
DO NOT REMOVE

Evolution of Data Science



Video Thumbnail
DO NOT REMOVE

Let's conclude by defining data science



Video Thumbnail
DO NOT REMOVE

Summary

Upon completion of this module, you will be able to:

- Early civilizations made use of decisions to solve critical problems that impacted the growth and sustenance of the community.
- Data Science involves the amalgamation of two paradigms - the inferential and computational paradigms.
- The Inferential paradigm focuses on the statistical methods for analysis, while the computational paradigm focuses on computational methods and algorithms.
- Data Science has evolved from simple probabilistic models and primary computers in the early 1950s to highly advanced inference methods and computing in the 2020s.
- The developments in inferential and computational paradigms act as catalysts that propel further advancements in these categories as the evolution continues.

Video Thumbnail
DO NOT REMOVE

Great Learning

Happy Learning !