

Welcome! We will begin shortly



Beyond the Numbers

Discovering the fascinating history of data science

Learning Outcomes

- 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

Guidelines



Listen only mode



Ask questions at the interest of the larger audience



Questions in the Q&A Box

Thank you

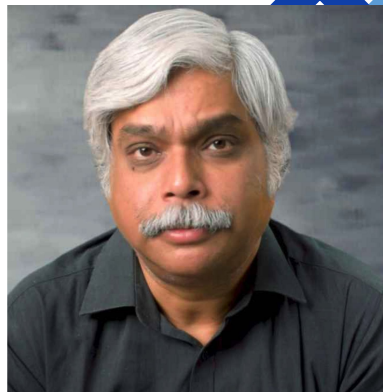
Kindly utilize the chat box for **subject-relevant questions only** to maximize your learnings from the session.

Your **questions** are being managed by the academic team, and they will be **answered**.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Meet Your Speaker



Dr. Abhinanda Sarkar
Academic Director at Great Learning

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

This file is meant for personal use by jacesca@gmail.com only.

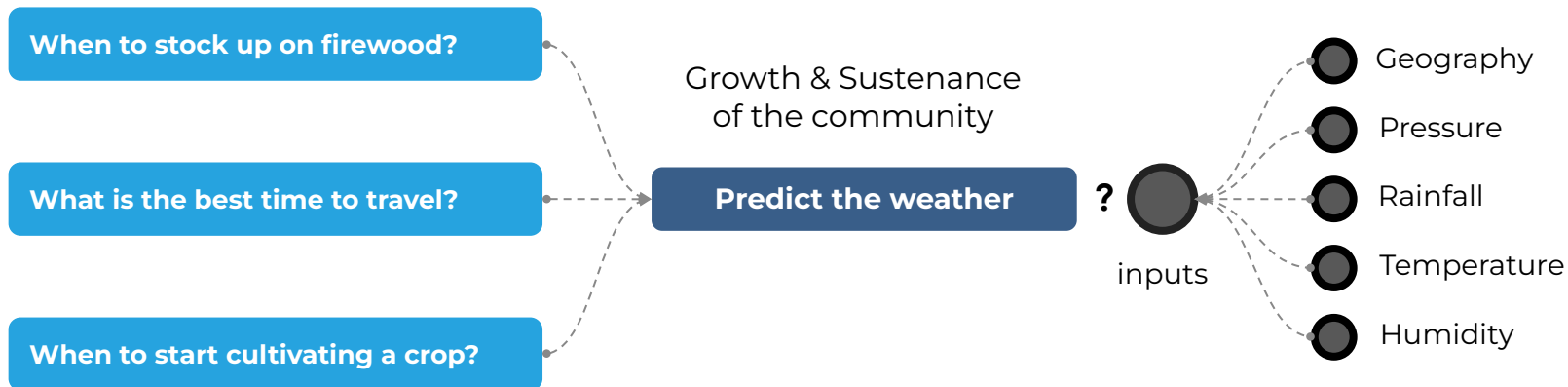
Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Origin of Decisions

Decisions were always **data-driven**

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



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

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Paradigms in data science

Inferential

Make predictions on population based on sample data

Use statistical methods to draw conclusions / infer from data

Representativeness of data

1. Effectiveness of a new medication through randomized trial
2. Impact of a new policy on citizens

Computational

Leverage computational methods and technology to scale insight generation

Implement algorithms and computational methods to analyse data


Complexity of algorithms and cost of training large models

1. Weather forecasting based on historical and weather patterns
2. Optimize routing of vehicles to minimize costs

 Focus

 Methods

 Limitations

 Examples

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (Early Stats & Computing)

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science

Inferential

Central Tendency

Expected Values

Probability Theory



Telegraph

Difference Engine

Audio Tapes

Computational

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (1940s to 1950s)

Bayesian 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

Evolution of Data Science (1940s to 1950s)

Digital Computers

Electronic device which can do math and logical calculations

- Electronic Numerical Integrator & Computer (ENIAC)
- Developed in 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 behavior of neutrons in nuclear reactor
- Now - Predict weather patterns, financial markets

Programming Languages

To have a computer understand instructions & execute them

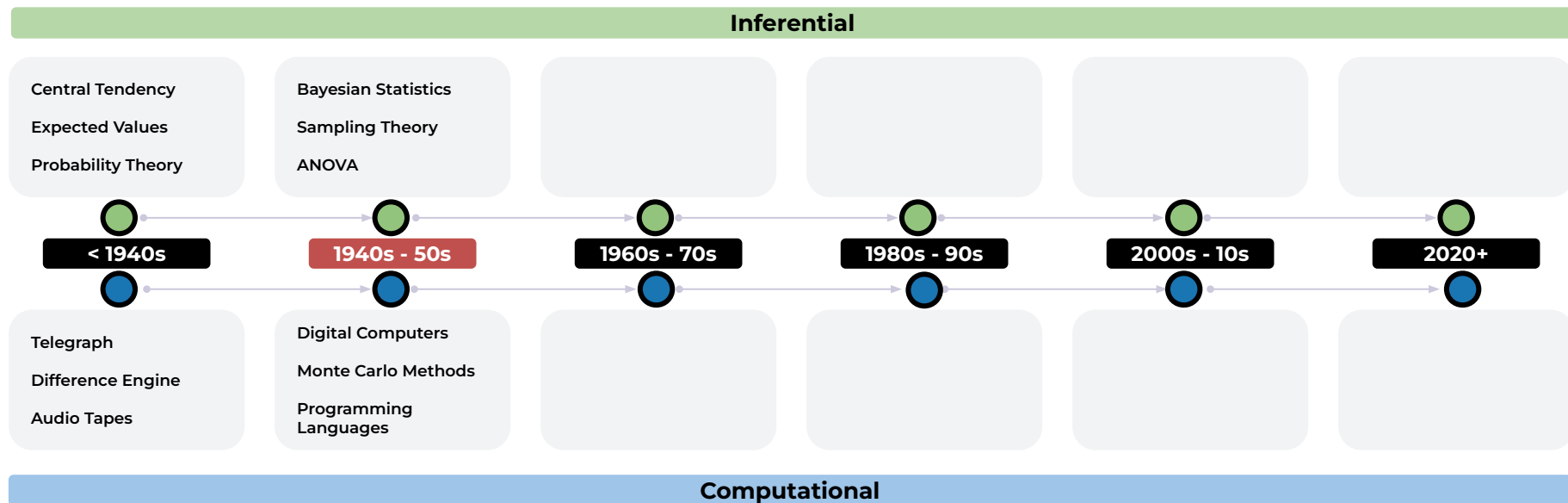
- Fortran - the 1st programming language created
- Made for scientific and engineering calculations
- Led to the development of World Wide Web
- Revolutionized communication & learning

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (1960s to 1970s)

Non-Parametric Methods

Rely on ranking / ordering of data rather than the distribution

- Frank Wilcoxon - Wilcoxon Rank-sum test
- Effect of store ambience on customer behaviour
- 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

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (1960s to 1970s)

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, 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
- **Common Business Oriented Language** - COBOL
- Laid the roots for the creation of MySQL in 1995

Time Sharing Systems

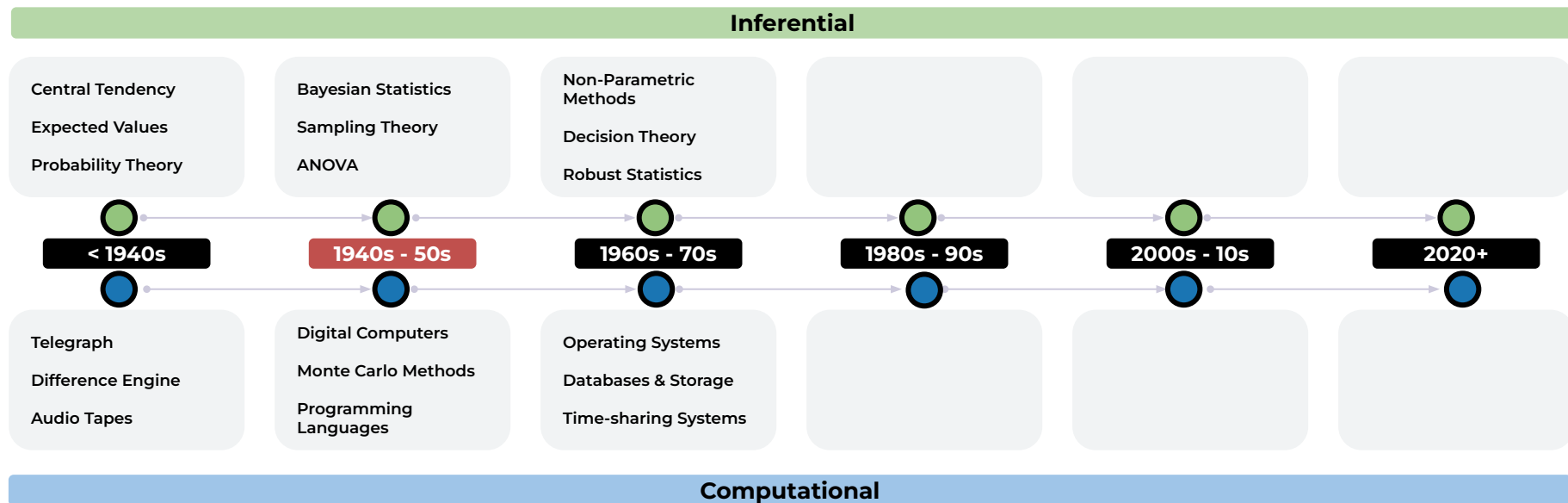
Many users can access a computer at the same time!

- **Compatible Time-Sharing System** (CTSS)
- Created by MIT to access one IBM computer
- The seed thought for modern networking systems
- Today's cloud exists because of the experiment!

This file is meant for personal use by jacesca@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (1980s to 1990s)

Resampling Methods

Simulate multiple datasets from original data for analysis

- Bradley Efron- Bootstrap technique
- Market Research survey - calculate 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

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (1980s to 1990s)

Personal Computers

Small, lightweight, affordable - used by a single person

- Altair 8880 - 1st PC - 1975 - 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

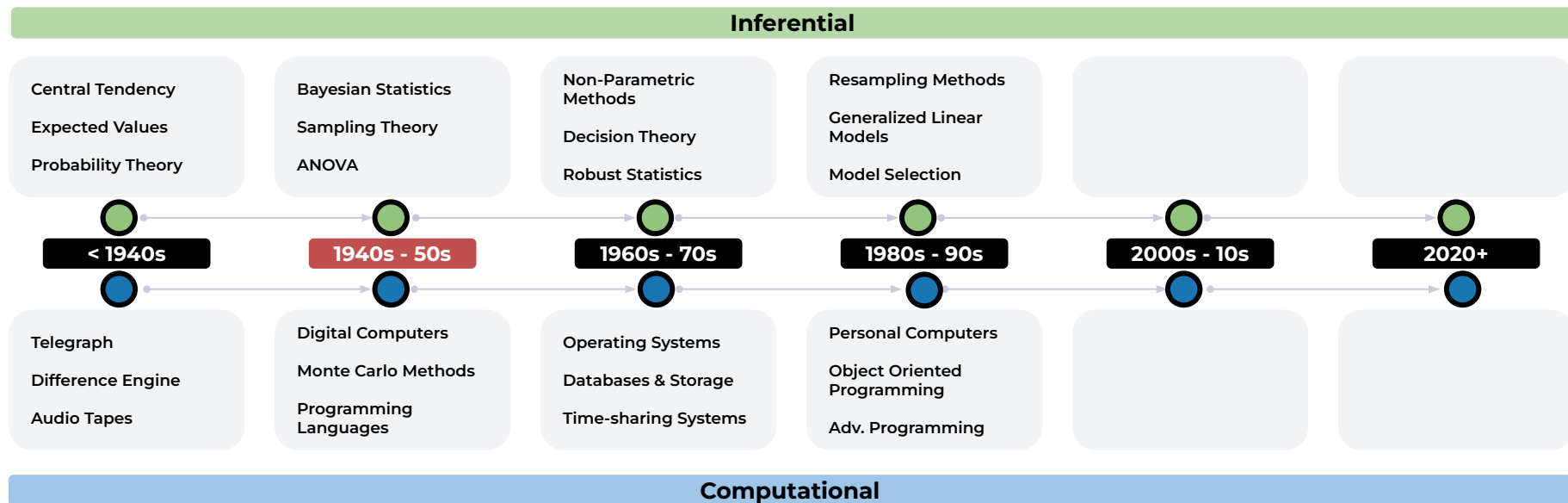
- Earliest was Fortran - created in the mid 50s
- Laid the foundation for Java, Ruby, and Python
- Able to create complex applications - at scale
- Democratization of programming - Anybody can Code!

This slide is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (2000s to 2010s)

Bayesian Networks & Graphical Models

Relationship b/w variables in a dataset using graphs

- Judea Pearl - Directed Acyclic Graphs - Turing award
- Healthcare - chances of patient having a disease
- Easily interpret complex relationship 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

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (2000s to 2010s)

Personal Computers

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

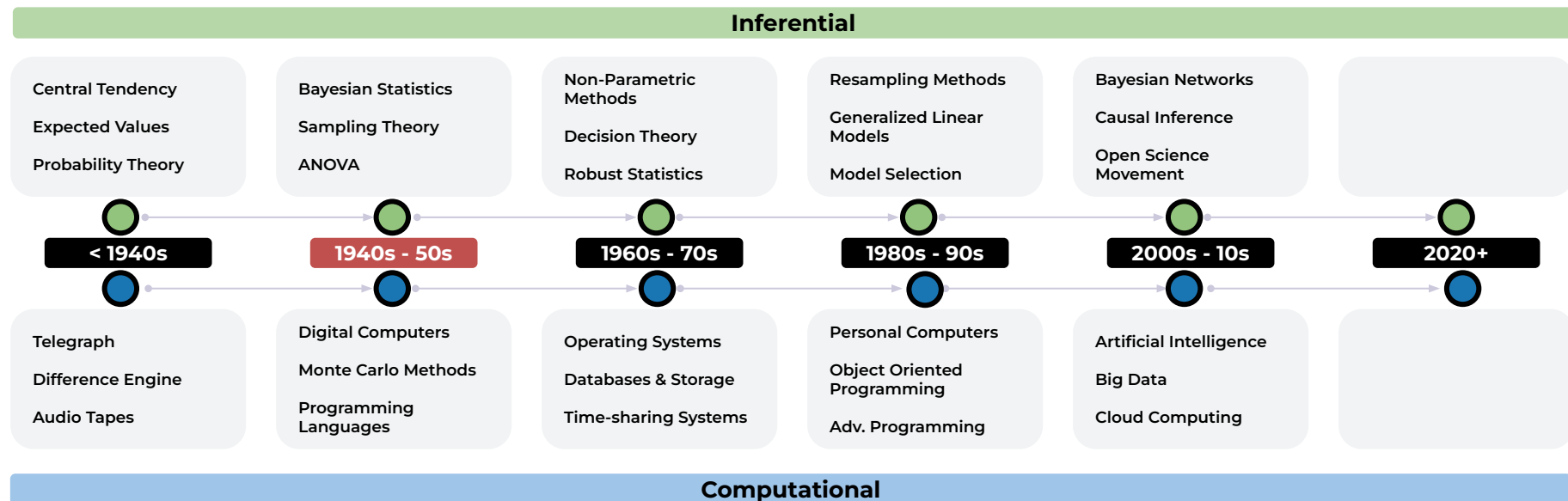
- Coined by Eric Schmitt - 2006, Google CEO
- 2006: **A**ma**z**on **W**eb **S**ervices, the 1st cloud provider
- Followed by Microsoft Azure and Google Cloud
- Grew due to the reduction in cost of computing

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (2020s+)

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 occurring w/o manipulating factors

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

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (2020s+)

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)

Edge Computing

Compute directly at the source of data, instead of remote

- Took shape in the early 2000s - Internet of Things (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

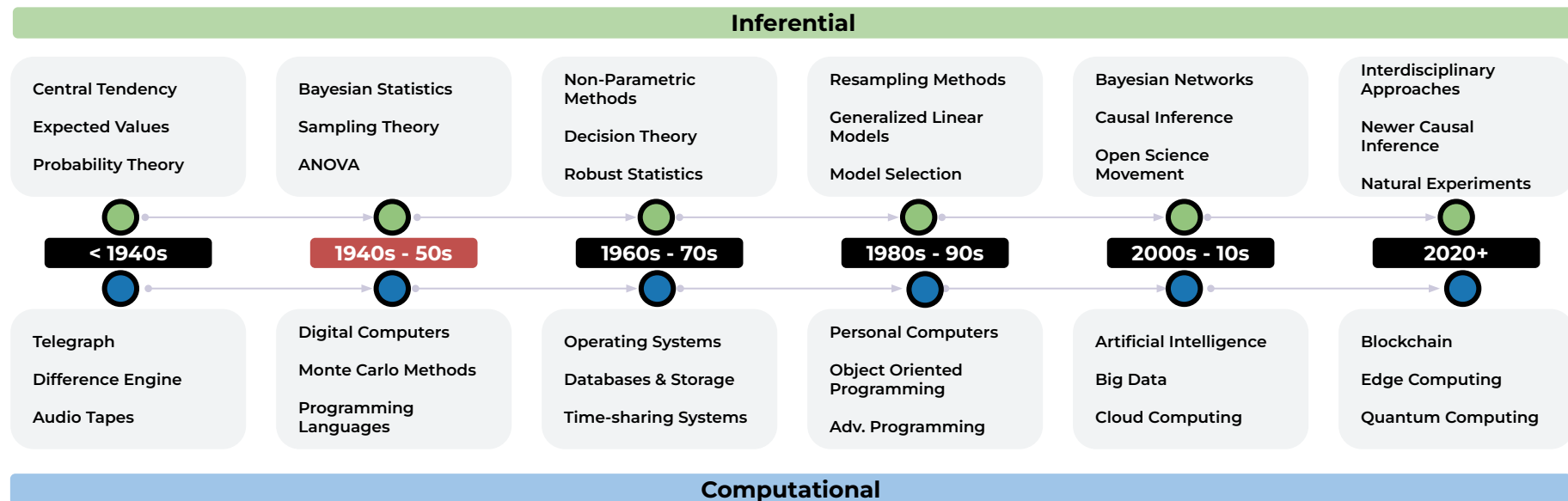
- 1st built in 1998 - Los Alamos Laboratory New Mexico
- Impact areas: Cryptography, Chemistry & Optimization
- In early stages, a lot of opportunities are still theoretical and under experimentation

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

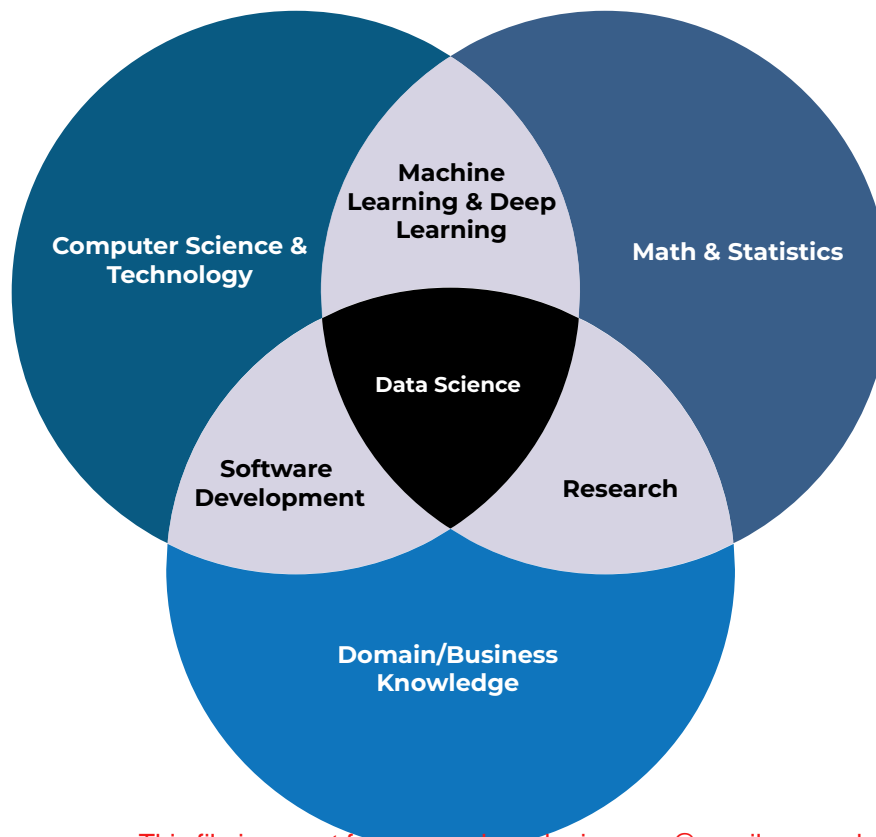
Questions?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Let's conclude by defining data science...



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



Happy Learning !



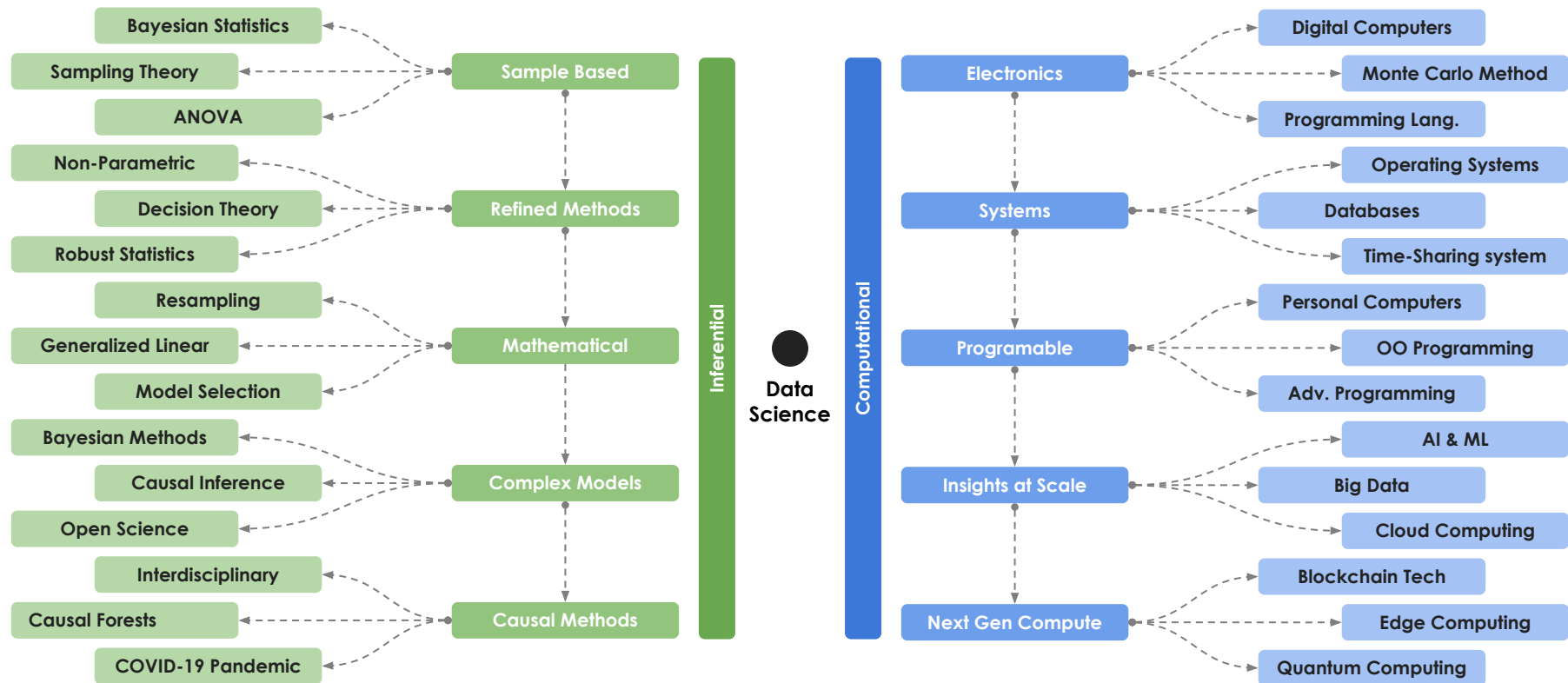
Appendix

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Summary - mind map of the history

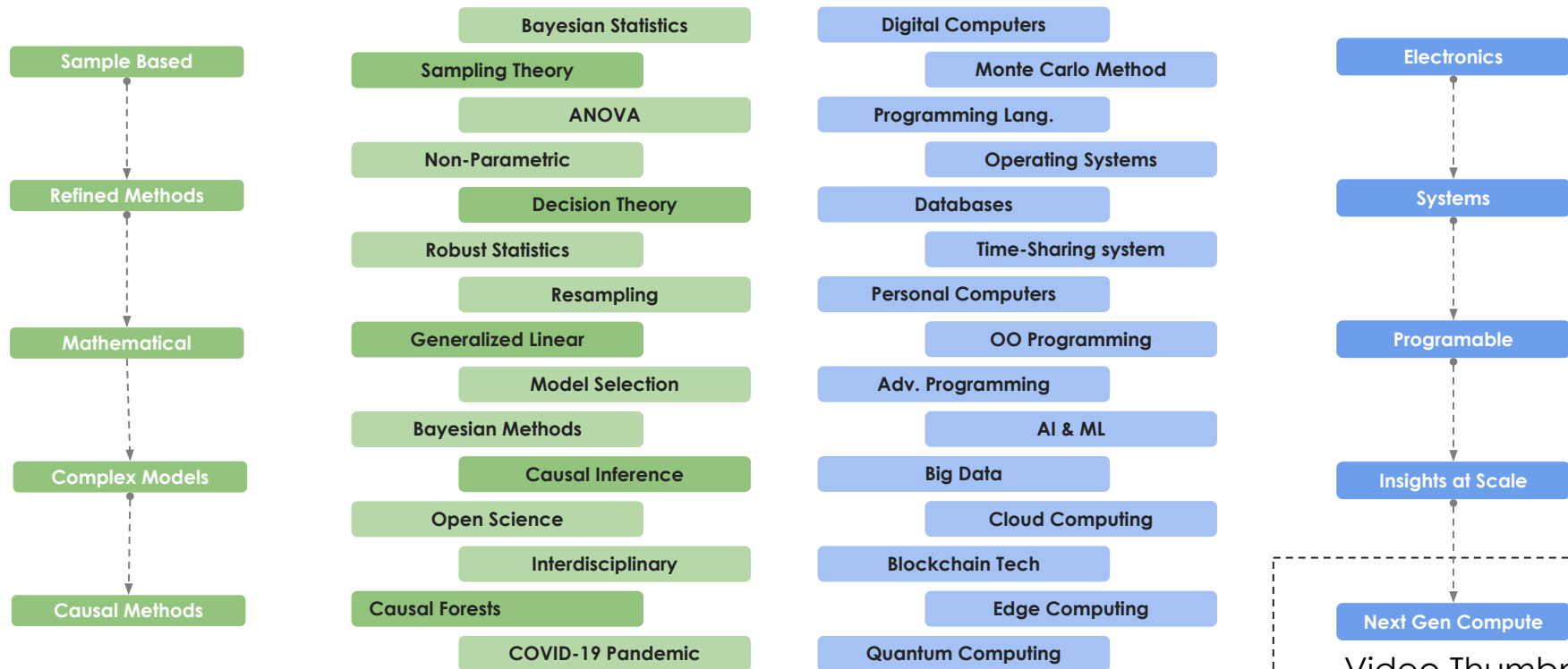


This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Summary - mind map of the history



Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (1940s-1950s)

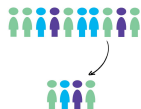
Inferential

Bayesian Statistics



- Analysis of data and the inference of probabilities
- **How?** - Using Bayes' theorem

Sampling Theory



- Challenges will large populations
- **How?** - Drawing inferences from small subsets

ANOVA



- Test for significant differences between groups
- **How?** - variability within and between the groups

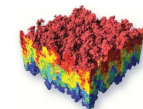
Computational

Digital Computers



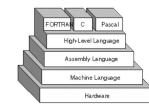
- First electronic digital computers
- **How ?** - Developments in electronic computing technology

Monte Carlo Methods



- To simulate the behavior of the process
- **How?** - Model is created using probability distributions

Programming Languages



- Automate complex calculations and tasks
- **How?** - Fortran & COBOL were discovered

This file is meant for personal use by jacesca@gmail.com only.

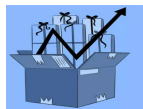
Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (1960s-1970s)

Inferential

Non-parametric methods



- When data do not meet the standard assumptions
- How?** - by using techniques like ranking

Decision Theory



- Framework for making decisions under uncertainty
- How?** - evaluating and choosing best alternatives

Robust Statistics



- Methods less sensitive to violations of assumptions
- How?** - resisting to outliers & other non-normality

Computational

Operating systems



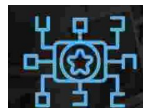
- Manages hardware and software resources
- How?** - using device drivers, system libraries, and system utilities

Databases



- Tool for managing/analyzing large data.
- How?** - using tables, queries, indexes

Time-sharing systems



- System to access computer simultaneously
- How?** - by dividing the computing resources

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

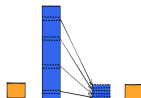
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (1980s-1990s)

Inferential

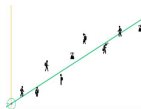
● Resampling Methods

- To evaluate the performance of a statistical model
- **How?** - Using Bootstrap/Cross validation resampling



● Generalized Linear Models

- To handle variety of variables & error distributions
- **How?** - connecting predictor function to the expected value of the response variable



● Model Selection Techniques

- To select the best statistical model
- **How?** - using techniques like AIC and BIC



Computational

● Personal computers

- Computer designed for use by an individual
- **How?** - using small, programmable computing devices



● Object-Oriented Programming

- "objects" to represent data and functionality
- **How?** - using blueprints of behaviors of the objects



● Advanced Programming Languages

- Address the limitations of earlier languages
- **How?** - C++, Python, and Java



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (2000s-2010s)

Inferential

Bayesian Networks & graphical Models

- Probabilistic representation of complex relationships
- **How?** - Using graphs to represent the relationships



Causal Inference

- Analyzing how one event/action leads to another
- **How?** - Using techniques like DID & IV



Open Science Movement

- emphasizing transparency, and community-driven innovation
- **How?** - by making things freely available



Computational

Artificial Intelligence and Machine Learning

- Enables to make predictions/decisions
- **How?** - Utilizing large amounts of data



Big Data

- Large and complex data sets
- **How?** - generated by digital systems & applications



Cloud Computing

- Delivery of on-demand computing services
- **How?** - through service based model



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Computational

[illegible]

- ## ● Newer Causal Inference Methods



- ## Natural Experiments



- ## Blockchain technology



- Secure & transparent storage/transfer of data
- How?** - creating a digital ledger that records transactions

● Edge Computing



- Computing directly at the source of data
- **How?** - Deploy resources within the product

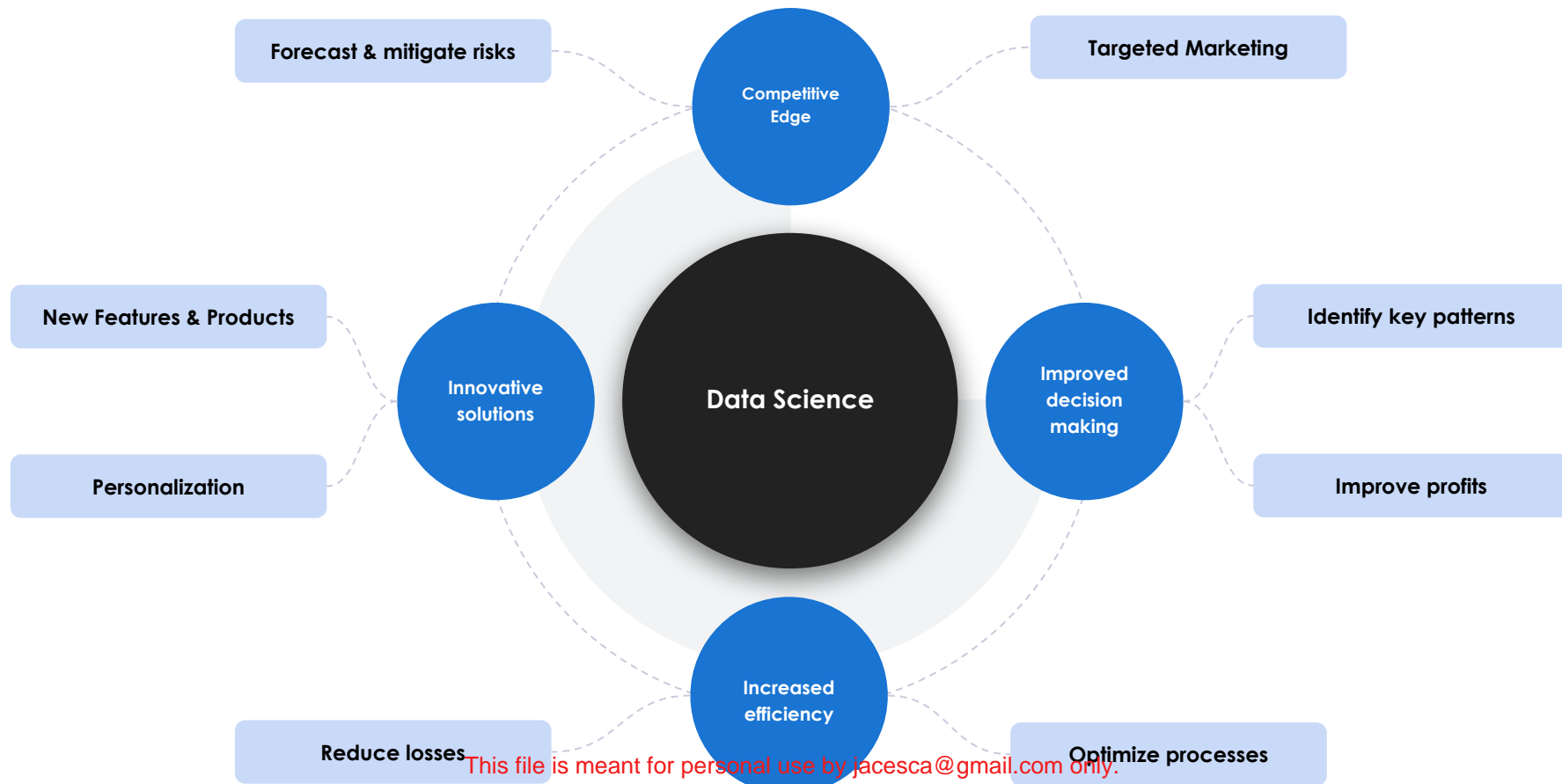
Quantum Computing



- Providing more processing power
- **How?** - use quantum bits or qubits

Sharing or publishing the contents in part or full is liable for legal action.
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Why data science?



Paradigms in data science



Focus

Inferential

Make predictions on population based on sample data

Computational

Leverage computational methods and technology to scale insight generation



Methods

Use statistical methods to draw conclusions / infer 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

This file is meant for personal use by jacesca@gmail.com only.

Evolution of Data Science

Inferential

Central Tendency
Expected Values
Probability Theory



< 1940s



1940s - 50s



1960s - 70s



1980s - 90s



2000s - 10s



2020+



Telegraph
Difference Engine
Audio Tapes

Computational

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Video Thumbnail
DO NOT REMOVE

Evolution of Data Science (1960s-1970s)

Non-Parametric Methods

Rely on ranking / ordering of data rather than the distribution

- Frank Wilcoxon - Wilcoxon Rank-sum test
- Effect of store ambience on customer behaviour
- 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

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, 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
- **Common Business Oriented Language** - COBOL
- Laid the roots for the creation of MySQL in 1995

Time Sharing Systems

Many users can access a computer at the same time!

- **Compatible Time-Sharing System** (CTSS)
- Created by MIT to access one IBM computer
- The seed thought for modern networking systems
- Today's cloud exists because of the experiment!

Evolution of Data Science (1980s-1990s)

Resampling Methods

Simulate multiple datasets from original data for analysis

- Bradley Efron- Bootstrap technique
- Market Research survey - calculate 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

Personal Computers

Small, lightweight, affordable - used by a single person

- Altair 8880 - 1st PC - 1975 - 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!

Evolution of Data Science

Inferential

Central Tendency
Expected Values
Probability Theory

Bayesian Statistics
Sampling Theory
ANOVA

Non-Parametric
Methods
Decision Theory
Robust Statistics



< 1940s

1940s - 50s

1960s - 70s

1980s - 90s

2000s - 10s

2020+

Telegraph
Difference Engine
Audio Tapes

Digital Computers
Monte Carlo
Methods
Programming
Languages

Operating Systems
Databases &
Storage
Time-sharing
Systems

Computational

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science

Inferential

Central Tendency
Expected Values
Probability Theory

Bayesian Statistics
Sampling Theory
ANOVA

Non-Parametric
Methods
Decision Theory
Robust Statistics

Resampling
Methods
Generalized Linear
Models
Model Selection



< 1940s

1940s - 50s

1960s - 70s

1980s - 90s

2000s - 10s

2020+

Telegraph
Difference Engine
Audio Tapes

Digital Computers
Monte Carlo
Methods
Programming
Languages

Operating Systems
Databases &
Storage
Time-sharing
Systems

Personal
Computers
Object Oriented
Programming
Adv. Programming

Computational

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (2000s-2010s)

Bayesian Networks & Graphical Models

Relationship b/w variables in a dataset using graphs

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

Causal Inference

Is change in one variable changing the other?

- 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

Artificial Intelligence & Machine Learning

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: Amazon Web Services, the 1st cloud provider
- Followed by Microsoft Azure and Google Cloud
- Grew due to the reduction to cost of computing

Evolution of Data Science

Inferential

Central Tendency
Expected Values
Probability Theory

Bayesian Statistics
Sampling Theory
ANOVA

Non-Parametric
Methods
Decision Theory
Robust Statistics

Resampling
Methods
Generalized Linear
Models
Model Selection

Bayesian Networks
Causal Inference
Open Science
Movement



< 1940s

1940s - 50s

1960s - 70s

1980s - 90s

2000s - 10s

2020+



Telegraph
Difference Engine
Audio Tapes

Digital Computers
Monte Carlo
Methods
Programming
Languages

Operating Systems
Databases &
Storage
Time-sharing
Systems

Personal
Computers
Object Oriented
Programming
Adv. Programming

Artificial
Intelligence
Big Data
Cloud Computing

Computational

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Evolution of Data Science (2020s+)

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 occurring w/o manipulating factors

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

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)

Edge Computing

Compute directly at the source of data, instead of remote

- Took shape in the early 2000s - **Internet of Things (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

Evolution of Data Science

Inferential

Central Tendency
Expected Values
Probability Theory

Bayesian Statistics
Sampling Theory
ANOVA

Non-Parametric
Methods
Decision Theory
Robust Statistics

Resampling
Methods
Generalized Linear
Models
Model Selection

Bayesian Networks
Causal Inference
Open Science
Movement

Interdisciplinary
Approaches
Newer Causal
Inference
Natural
Experiments



< 1940s



1940s - 50s



1960s - 70s



1980s - 90s



2000s - 10s



2020+



Telegraph
Difference Engine
Audio Tapes

Digital Computers
Monte Carlo
Methods
Programming
Languages

Operating Systems
Databases &
Storage
Time-sharing
Systems

Personal
Computers
Object Oriented
Programming
Adv. Programming

Artificial
Intelligence
Big Data
Cloud Computing

Blockchain
Edge Computing
Quantum
Computing



Computational

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.