

# Data Science

What? Why? How?

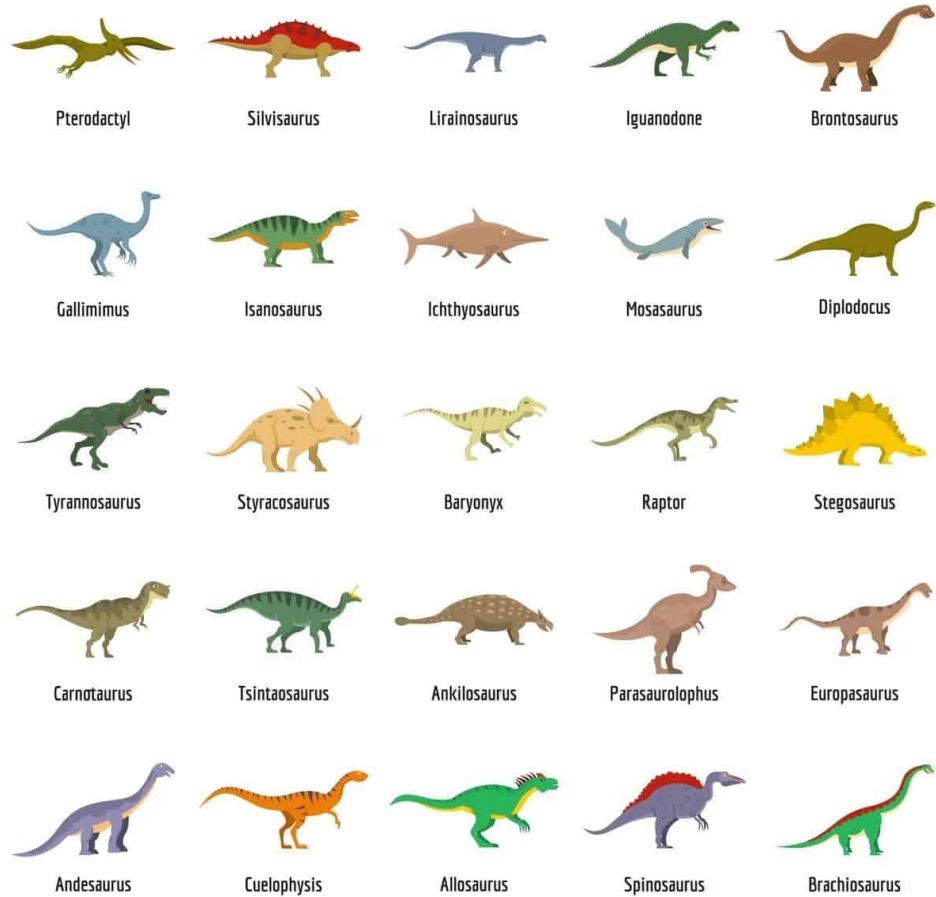
**Dr. Panchatcharam M**

Associate Professor & Head, Department of Mathematics and Statistics, IIT Tirupati



# Biryani





# Time Table



|    | A                  | B     | C     | D     | E     | F     | G     | H     | I     | J     | K     | L        | M                                                                                                                                                                                                                                                                                                                 | N                                                          |
|----|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| 1  | Institute          | 08-09 | 09-10 | 10-11 | 11-12 | 12-13 | 13-14 | 14-15 | 15-16 | 16-17 | 17-18 | After 18 | Source                                                                                                                                                                                                                                                                                                            | Slot System                                                |
| 2  | IIT (BHU) Varanasi | Y     | Y     | Y     | Y     | N     | P     | P     | P     | P     | P     | N        | <a href="https://www.iitbhu.ac.in/contents/institute/academics/fresher/doc/time-table_2024-25_uq.pdf">https://www.iitbhu.ac.in/contents/institute/academics/fresher/doc/time-table_2024-25_uq.pdf</a>                                                                                                             | Departmental wise                                          |
| 3  | IIT Bhilai         | N     | N     | Y     | Y     | Y     | N     | P     | Y     | Y     | P     | N        | <a href="https://www.iitbhu.ac.in/index.php?pid=TimeTablefor1stYearStudents2020">https://www.iitbhu.ac.in/index.php?pid=TimeTablefor1stYearStudents2020</a>                                                                                                                                                       | No Slot System                                             |
| 4  | IIT Bhubaneswar    | P     | Y     | Y     | Y     | P     | N     | Y     | Y     | N     | Y     | N        | BHUBANESWAR                                                                                                                                                                                                                                                                                                       | No Slot System                                             |
| 5  | IIT Bombay         | N     | Y     | Y     | Y     | Y     | N     | Y     | Y     | Y     | Y     | Y        | <a href="https://acad.iitb.ac.in/sites/default/files/Autumn_TTC_26-27_09.07.2026.pdf">https://acad.iitb.ac.in/sites/default/files/Autumn_TTC_26-27_09.07.2026.pdf</a>                                                                                                                                             | 1A, 2B Etc                                                 |
| 6  | IIT Delhi          | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | P     | P        | <a href="https://timetable.iitd.ac.in/public/storage/uploads/schedule/Slotting_Pattern_L_Abolished_1728266587.pdf">https://timetable.iitd.ac.in/public/storage/uploads/schedule/Slotting_Pattern_L_Abolished_1728266587.pdf</a>                                                                                   |                                                            |
| 7  | IIT Dharwad        | P     | Y     | Y     | P     | P     | N     | Y     | Y     | N     | N     | N        |                                                                                                                                                                                                                                                                                                                   |                                                            |
| 8  | IIT Goa            | Y     | Y     | Y     | Y     | Y     | N     | Y     | Y     | Y     | P     | N        | GOA                                                                                                                                                                                                                                                                                                               | A-G AND P-S                                                |
| 9  | IIT Indore         | P     | P     | P     | P     | N     | P     | P     | P     | P     | P     | P        | IIT Indore                                                                                                                                                                                                                                                                                                        | A to H, M to P (practical), 8.30 AM to 6.25 PM With 55 min |
| 10 | IIT (ISM) Dhanbad  | N     | Y     | Y     | N     | Y     | N     | Y     | Y     | Y     | Y     | Y        | DHANBAD                                                                                                                                                                                                                                                                                                           |                                                            |
| 11 | IIT Jammu          | Y     | Y     | Y     | Y     | P     | N     | N     | N     | N     | N     | N        | <a href="https://www.iitjammu.ac.in/pdf/program/timetable/2026/BT_first_year.pdf">https://www.iitjammu.ac.in/pdf/program/timetable/2026/BT_first_year.pdf</a>                                                                                                                                                     |                                                            |
| 12 | IIT Kanpur         | Y     | Y     | Y     | N     | Y     | Y     | Y     | Y     | Y     | Y     | N        | <a href="https://www.iitk.ac.in/doi/data/Coreschedule2026-27-1.pdf">https://www.iitk.ac.in/doi/data/Coreschedule2026-27-1.pdf</a>                                                                                                                                                                                 |                                                            |
| 13 | IIT Kharagpur      | Y     | Y     | Y     | Y     | Y     | N     | Y     | Y     | Y     | Y     | N        | <a href="https://erp.iitkgp.ac.in/Acad/Aut_Time_Table_26-27.pdf">https://erp.iitkgp.ac.in/Acad/Aut_Time_Table_26-27.pdf</a>                                                                                                                                                                                       |                                                            |
| 14 | IIT Mandi          | Y     | Y     | Y     | Y     | Y     | N     | N     | N     | N     | Y     | N        | IIT Mandi                                                                                                                                                                                                                                                                                                         | A to H, L1 to L5 (laboratory), 50 min timeslots            |
| 15 | IIT Palakkad       | Y     | Y     | Y     | Y     | Y     | N     | Y     | N     | Y     | Y     | N        | IIT Palakkad                                                                                                                                                                                                                                                                                                      |                                                            |
| 16 | IIT Patna          | Y     | Y     | Y     | Y     | Y     | N     | N     | Y     | Y     | Y     | N        | <a href="https://academics.iitp.ac.in/images/tttt-aut-24/First_Year_04_08_24.pdf">https://academics.iitp.ac.in/images/tttt-aut-24/First_Year_04_08_24.pdf</a>                                                                                                                                                     |                                                            |
| 17 | IIT Roorkee        | Y     | Y     | Y     | Y     | Y     | N     | Y     | Y     | Y     | Y     | Y        | <a href="https://iitr.ac.in/Academics/static/TimeTable/Aut%20Sem%202023-24/Notification_Slot_Wise_time_Table_for_Autumn_Semester_2023-24_AutoRecovered.pdf">https://iitr.ac.in/Academics/static/TimeTable/Aut%20Sem%202023-24/Notification_Slot_Wise_time_Table_for_Autumn_Semester_2023-24_AutoRecovered.pdf</a> |                                                            |
| 18 | IIT Ropar          | Y     | Y     | Y     | Y     | Y     | N     | Y     | Y     | Y     | Y     | Y        | <a href="https://www.iitrpr.ac.in/math/2page_id-2119">https://www.iitrpr.ac.in/math/2page_id-2119</a>                                                                                                                                                                                                             | Department wise, no slot system                            |
| 19 | IIT Guwahati       | Y     | Y     | Y     | Y     | Y     | N     | Y     | Y     | Y     | Y     | N        | Guwahati                                                                                                                                                                                                                                                                                                          | NO                                                         |
| 20 | IIT Jodhpur        | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | N        | Jodhpur                                                                                                                                                                                                                                                                                                           | A1-A6, B1-B7                                               |
| 21 | IIT Gandhinagar    | P     | Y     | Y     | Y     | Y     | N     | Y     | Y     | Y     | Y     | N        | Gandhinagar                                                                                                                                                                                                                                                                                                       | NO                                                         |
| 22 | IIT Hyderabad      | N     | Y     | Y     | Y     | P     | N     | Y     | Y     | Y     | P     | Y        | Hyderabad                                                                                                                                                                                                                                                                                                         | A-G AND Q-S                                                |
| 23 | IIT Madras         | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | Y     | N        | Madras                                                                                                                                                                                                                                                                                                            | A-M & A1-M1 & P1-S1 & P2-S2                                |
| 24 | IIT Tirupati       | Y     | Y     | Y     | Y     | Y     | N     | Y     | N     | N     | N     | N        | Tirupati                                                                                                                                                                                                                                                                                                          | A - H & P1                                                 |
| 25 |                    |       |       |       |       |       |       |       |       |       |       |          |                                                                                                                                                                                                                                                                                                                   |                                                            |
| 26 |                    |       |       |       |       |       |       |       |       |       |       |          |                                                                                                                                                                                                                                                                                                                   |                                                            |
| 27 |                    |       |       |       |       |       |       |       |       |       |       |          |                                                                                                                                                                                                                                                                                                                   |                                                            |

# References



- ▶ Kenneth S B, Bruce B A, **Research Design and Methods, A Process Approach**, 8<sup>th</sup> Edition, McGraw Hill, 2011
- ▶ Michael M, **Head First Data Analysis**, O'Reilly, 2009
- ▶ Lyman R O, Michael L, An Introduction to Statistical Methods and Data Analysis, 6th Edition, Cengage, 2010
- ▶ <https://www.questionpro.com/blog/data-analysis-in-research/>
- ▶ <https://www.geckoboard.com/blog/how-to-analyze-data/>
- ▶ <https://www.analyticsfordecisions.com/importance-of-data-analysis-in-research/>
- ▶ [Qualitative Data Analysis Techniques](#)
- ▶ [Collecting and Analyzing Data](#)

# Data

What is Data?

1

# Data



# Data is Everywhere



- ▶ Everyone deals with mountains of data – Are they Data Analysts?
- ▶ Drives real-world action: How to translate raw numbers into intelligence
- ▶ How to break down?
- ▶ How to structure the complex problems and data sets to get right at the heart of the problems in their research?

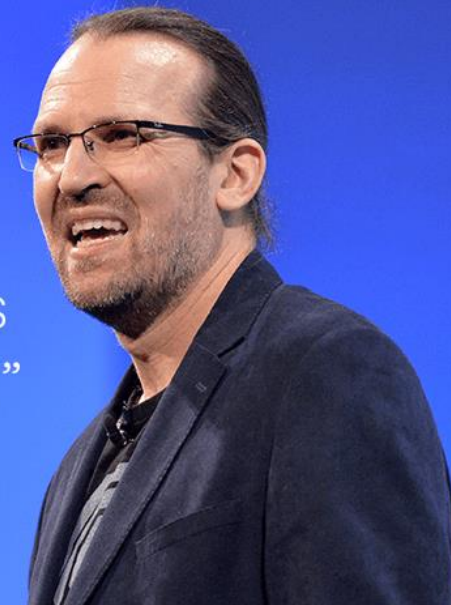
# Data is the Fuel



XEROCON 2016

“Data is the fuel of our future. When you have lots of data, it changes things.”

DAVE COPLIN, MICROSOFT



# Data Collection

1a

# Data Collection



- ▶ Primary Data
- ▶ Secondary Data
- ▶ Methods of Data Collection
- ▶ Important Data Available for Scientists in Web
- ▶ Public Domain Database
- ▶ How to use Public Domain Database in Research

# Data Analysts

Who is a Data Analyst?

1b

# Data Analysts



- ▶ Dataset – 1
  - ▶ Questions and Answers
- ▶ Dataset – 2
  - ▶ Questions and Answers
- ▶ Dataset – 3
  - ▶ Questions and Answers

# Data Analysts



- ▶ Knowing how to use Excel qualifies you as a Data Analyst?
- ▶ Good Data Analysts always want to see data

# Data Analysis

Let's explore what is Data Analysis

2



# What does Data Analysis mean?

- ▶ Refer to one method or many?
- ▶ Collection of different procedures?
- ▶ Without background of Mathematics and Statistics, can you learn to identify and use data analysis in your work?



# What does Data Analysis mean?

- ▶ Careful thinking about evidence



- ▶ Define: Know your problem
- ▶ Disassemble: Break problems and data into smaller pieces
- ▶ Evaluate: Draw your conclusions
- ▶ Decide: Make or recommend a decision

# Data Science: What?

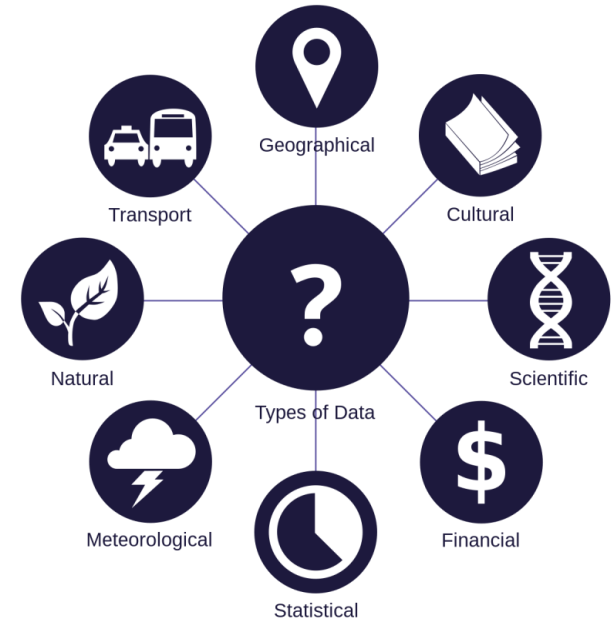
What is Data Science?

3

# Data Science: What?



- ▶ Also known as data-driven science
- ▶ An interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, either structured or unstructured



<sup>1</sup><https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf>



# Data Science: What?

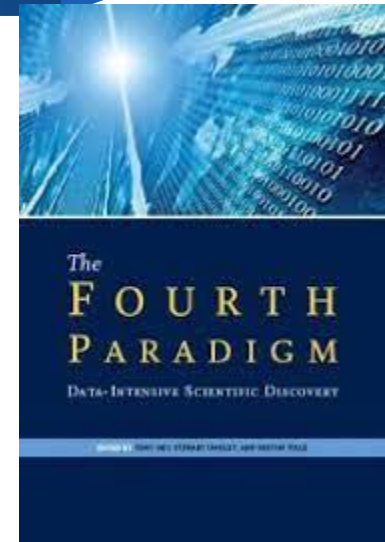
- ▶ Intends to analyze and understand actual phenomena with data
- ▶ To reveal the features or the hidden structure of complicated natural, human and social phenomena with data from a different point of view from the established or traditional theory and method.

<sup>1</sup><https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf>



# Data Science: What?

- ▶ Fourth Paradigm
- ▶ Changes of all sciences moving from observational to theoretical to computational and now to the 4<sup>th</sup> paradigm – Data-Intensive Scientific discovery
- ▶ Observational:
  - ▶ Think of Newton, Kepler, Archimedes
- ▶ Theoretical:
  - ▶ Riemann, Galois, Weierstrass, Einstein
- ▶ Computational:
  - ▶ Last 50 years of the research in the world



<sup>1</sup><https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf>

# Data Science and Big Data

Difference

3

# Data Science and Big Data



Are they the same thing? No

## Big Data

- ☐ Crude Oil
- ☐ Extracting Crude oil
- ☐ Transporting it in mega tankers
- ☐ Siphoning it through pipelines
- ☐ Storing it in Massive Silos

## Data Science:

- ✓ Refining the Crude Oil

Carlos Samohano

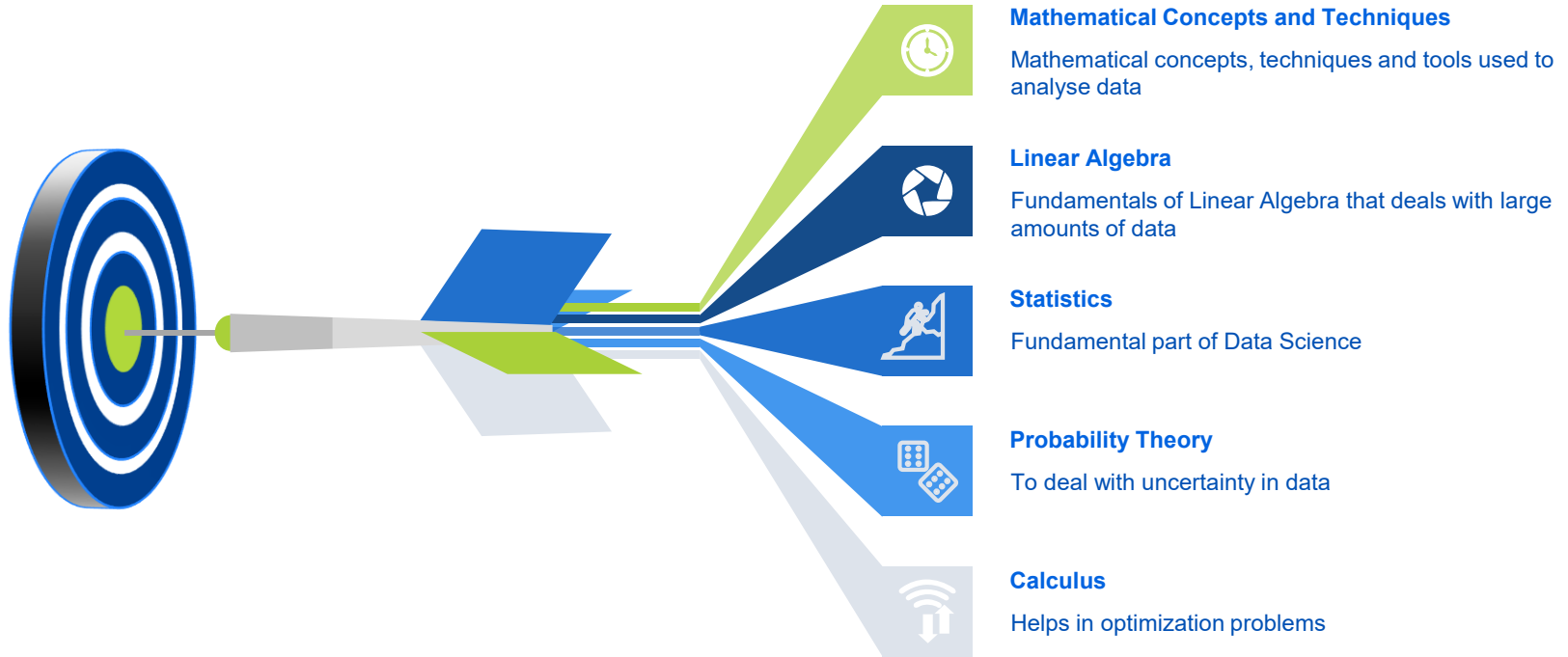
Founder, Data Science London

# Data Science: What?

What is Data Science?

3

# Data Science



# Data Analysis in Research: When?

How often? When?

5



# Data Analysis in Research: When?

01

Data Collection

- ✓ Should start no later than you begin your work
- ✓ Before you begin, in order to establish a baseline

02

When you collected all of it

- ✓ Best to analyse your information
- ✓ You can look at it as a whole
- ✓ Summative Evaluation

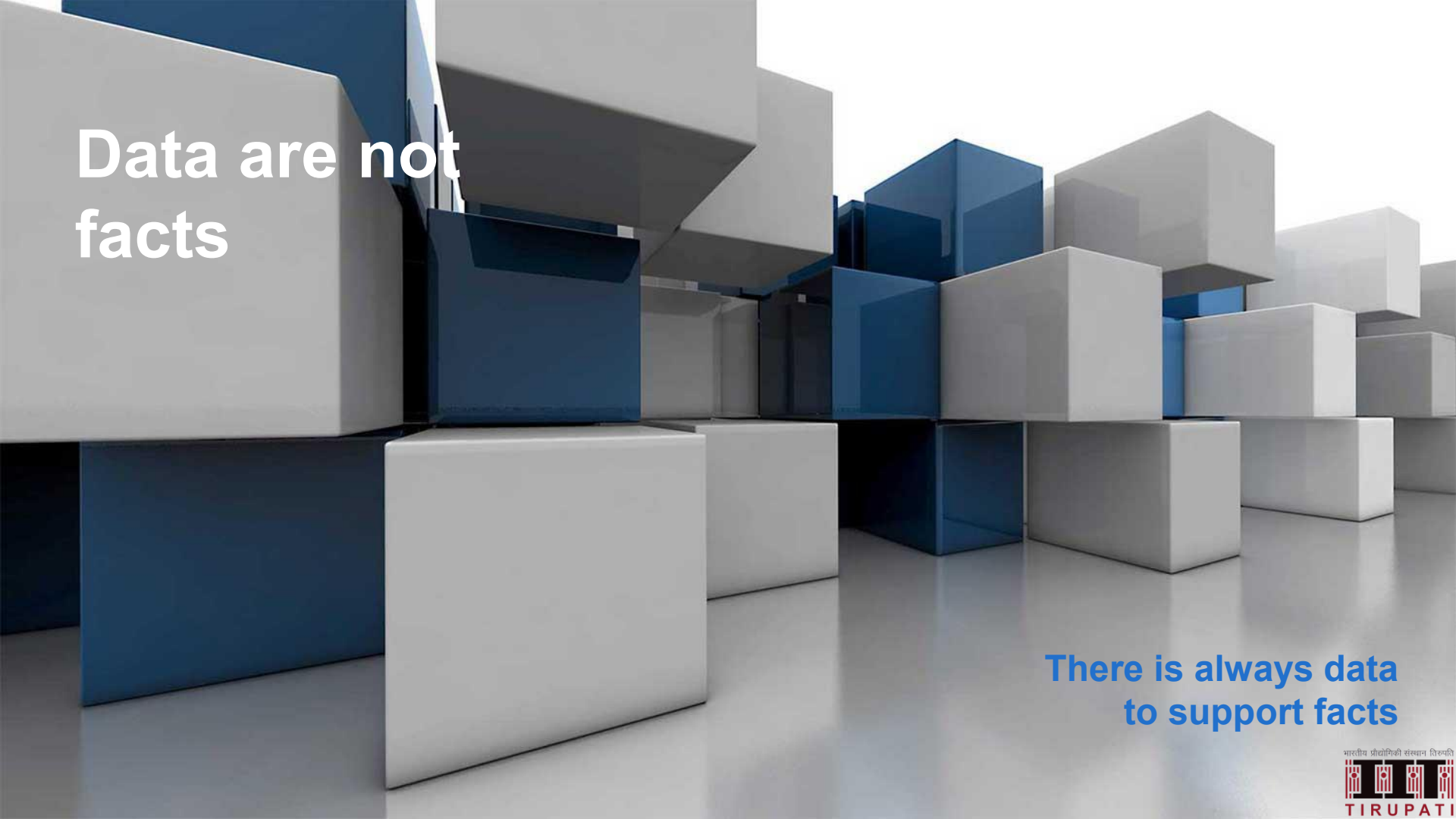
03

As you go along

- ✓ Adjust your thinking about what information you actually need
- ✓ Adjust your program to respond to the information you are getting
- ✓ Formative Evaluation

# Data Science: Why?

4

The background of the slide is an abstract 3D rendering of numerous cubes in white and blue. The cubes are arranged in a complex, overlapping pattern that recedes into the distance. The surface they sit on is a light gray with a subtle reflection, giving the scene a clean, modern feel. The lighting is soft, creating gentle shadows and highlights on the faces of the cubes.

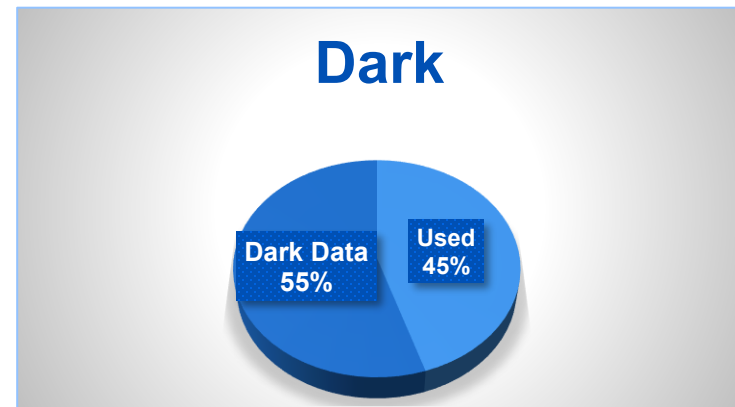
**Data are not  
facts**

**There is always data  
to support facts**



# Data Analysis: Why?

- ▶ Critical for all employees
- ▶ Many researchers/companies struggle with data organization and analysis
- ▶ Global Survey Splunk<sup>1</sup>: 55% of all data collected – dark data
- ▶ Dark Data: Data collected but never used
- ▶ Data collected, but do not know how to analyse it



<sup>1</sup><https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf>



# Data Analysis: Why?

- ▶ 76% executives<sup>1</sup> believe training current employees in data science will help to understand dark data
- ▶ If you understand how to analyse different types of data, you can make better output



<sup>1</sup><https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf>

# Big concept





# Why analyse data in Research?

## Researchers/Data Analysts:

- ☐ Have Heavy data
- ☐ Have Story to tell
- ☐ Have problems to solve

## Research/Analysis:

- ✓ Starts with a question
- ✓ **Data:** Answer to the question
- ✓ No question to ask
- ✓ **Data:** Explore – Data Mining

# Data Science: Why?



- ▶ Mathematics is crucial in data-related fields for several reasons
- ▶ **Statistics:** Statistics, a branch of mathematics, helps in drawing meaningful conclusions from data, estimating parameters, and making predictions.
- ▶ **Machine Learning:** Algorithms used in machine learning are often based on mathematical concepts, such as linear algebra, calculus, and probability theory.



# Data Science: Why?



- ▶ **Optimization:** Mathematical optimization techniques are used to fine-tune models and algorithms.
- ▶ **Data Visualization:** Geometry and statistics play a role in creating effective data visualizations.



# Data Science: Why?

## Detailed View



- ▶ **Data Analysis:** Mathematics provides the tools to analyze and interpret data effectively. Techniques like statistical tests, regression analysis, and hypothesis testing are all based on mathematical principles



# Data Science: Why?

## Detailed View



- ▶ **Quantification:** Mathematics allows us to quantify and measure various aspects of data, such as central tendencies (mean, median), dispersion (variance, standard deviation), and correlations.
- ▶ **Probability and Statistics:** Probability theory and statistics are the backbone of Data Science. They help in making sense of uncertain data, drawing meaningful conclusions, and making predictions.



# Data Science: Why?

## Detailed View



- ▶ **Machine Learning:** Machine learning algorithms, a core component of Data Science, often rely on linear algebra, calculus, and optimization techniques, all of which are mathematical in nature.
- ▶ **Data Visualization:** Creating effective data visualizations requires an understanding of geometric principles and mathematical concepts to accurately represent data.



# Data Science: Why?

## Detailed View



- ▶ **Predictive Modeling:** Mathematical modeling is used to build predictive models that can forecast future trends or outcomes based on historical data.
- ▶ **Data Preprocessing:** In data cleaning and preprocessing, mathematical techniques are used to handle missing values, outliers, and scale data appropriately.





## Detailed View

- 
- Individuals - PCA**
- Dim2 (8%)
- Dim1 (14.2%)
- Groups
- KO (blue circle)
  - OE (yellow triangle)
  - WT (red square)
- The PCA plot shows three distinct clusters of individuals. The KO group (blue circles) is clustered in the upper left quadrant, the OE group (yellow triangles) is in the upper right quadrant, and the WT group (red squares) is in the lower right quadrant. The axes are labeled Dim1 (14.2%) and Dim2 (8%).

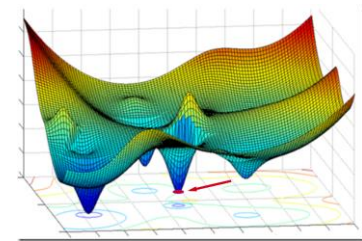


# Data Science: Why?

## Summary



- ▶ In summary, mathematics provides the foundation for understanding, processing, and extracting insights from data.
- ▶ It offers the language and tools needed to tackle the challenges of working with data, making it a fundamental aspect of Data Science.



**Data Science: How?**

**6**

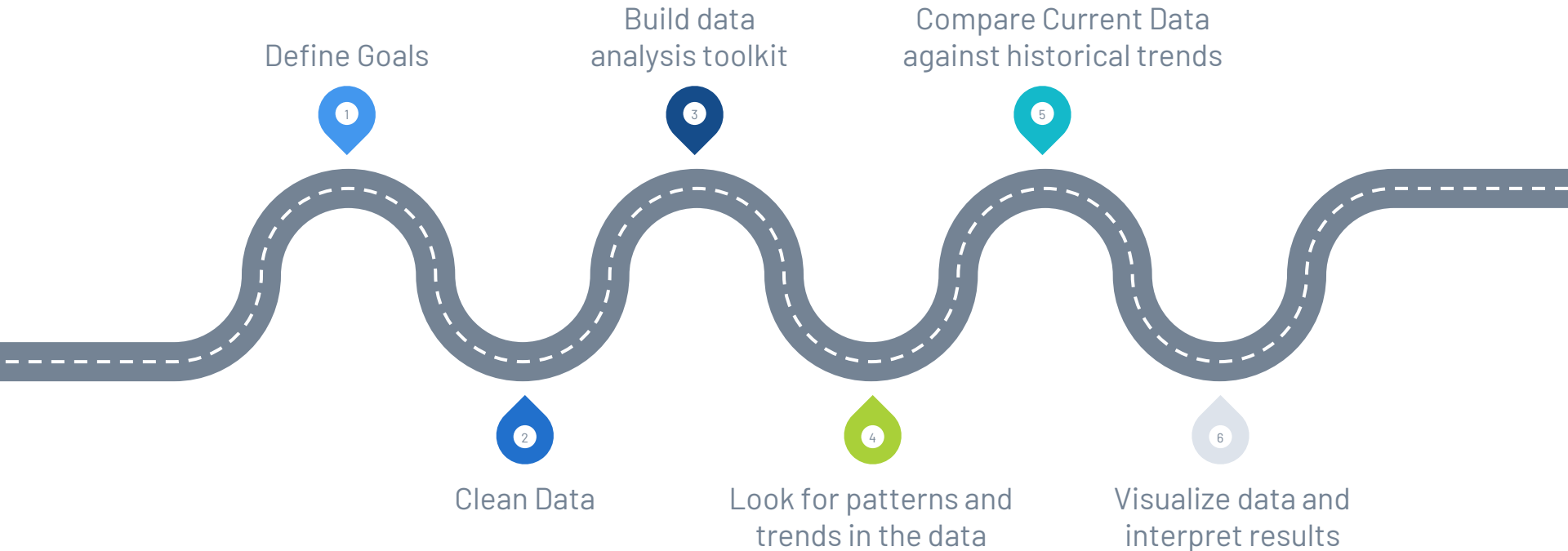
# Data Science: How?



- ▶ Data Analysis: A skill you can learn
  - ▶ By doing advanced statistics course
  - ▶ In-depth training module to understand how to analyse data



# Data Science: How?



# Data Science: How?

## Define Goals



- ▶ Before starting to analyse data
  - ▶ Set some clear objectives
- ▶ If no idea, you will stare at spreadsheet or scrolling through dataset
- ▶ Goal depends on the research, data collected and your role in research

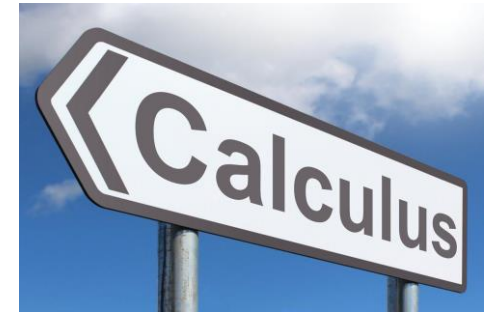
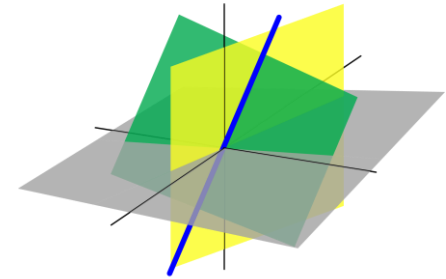


# Data Science: How?

## Linear Algebra and Calculus



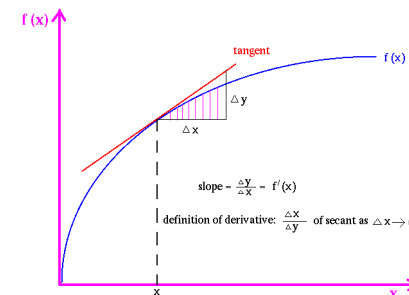
- ▶ **Linear Algebra:** Used in dimensionality reduction, eigenvalue decomposition, and linear regression.
- ▶ **Calculus:** Helps in optimization problems, like gradient descent in machine learning.





# Probability and Statistics, DEs

- ▶ **Probability and Statistics:** Used for hypothesis testing, probability distributions, and sampling.
  - ▶ Descriptive, Inferential, Regression, Time series analysis
- ▶ **Differential Equations:** Applied in modeling dynamic systems, like in epidemiology or physics-based simulations.

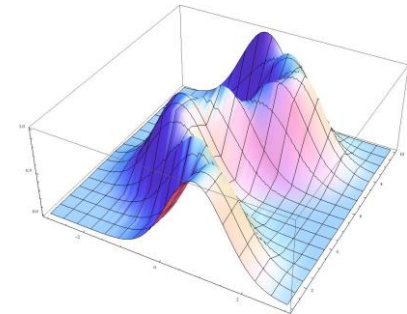
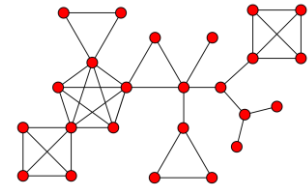


# Data Science: How?

## Graph Theory and Numerical Analysis



- ▶ **Graph Theory:** Useful for analysing networks and relationships in data.
- ▶ **Numerical Analysis:** Helps in solving complex mathematical problems efficiently.

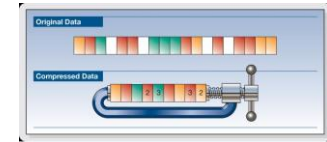


# Data Science: How?

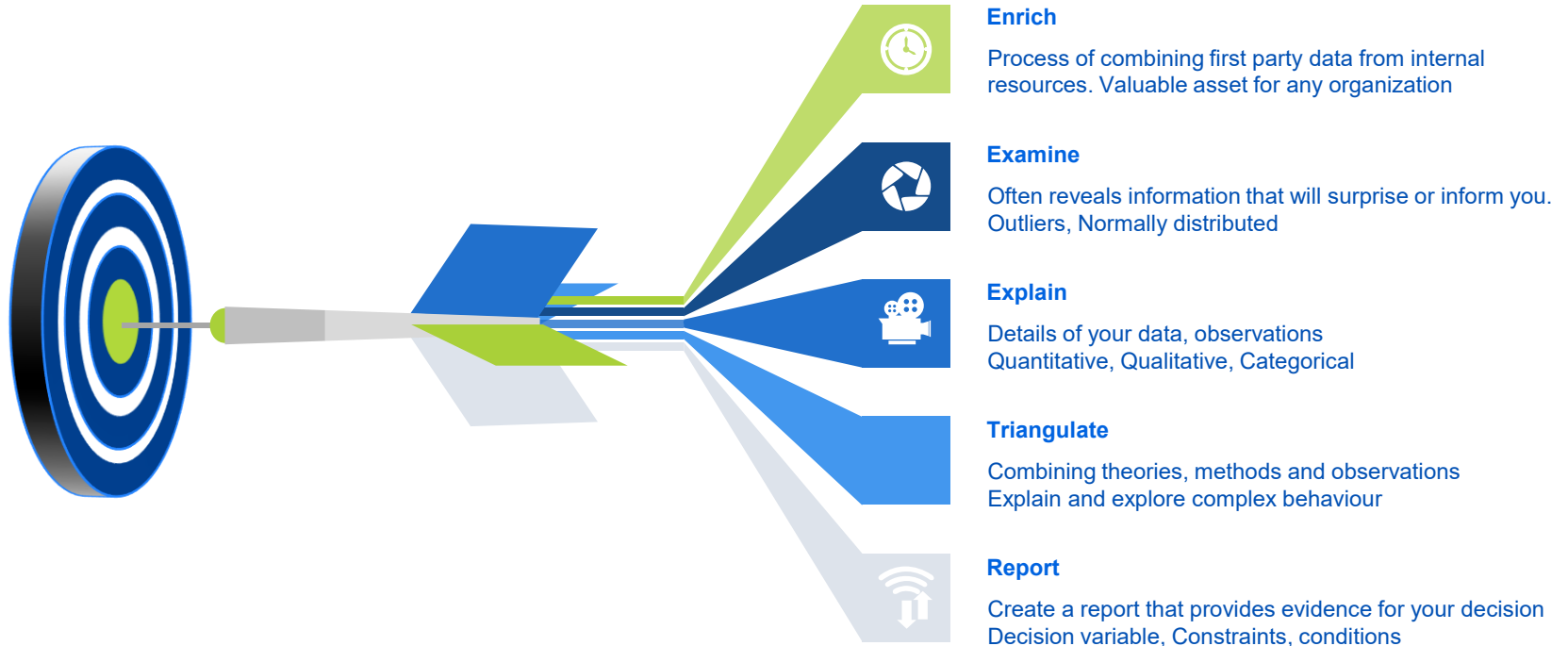
## Graph Theory and Numerical Analysis



- ▶ **Data Compression:** Mathematical techniques like the Huffman coding algorithm are used for data compression, reducing the storage space required.
- ▶ **Optimization Problems:** Mathematics is used to solve optimization problems, such as finding the best route for delivery trucks or the optimal allocation of resources.



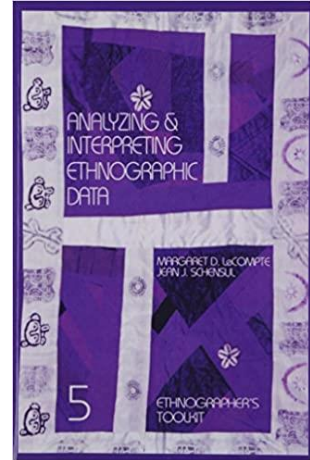
# Data Analysis in Research





# What is Data Analysis in Research?

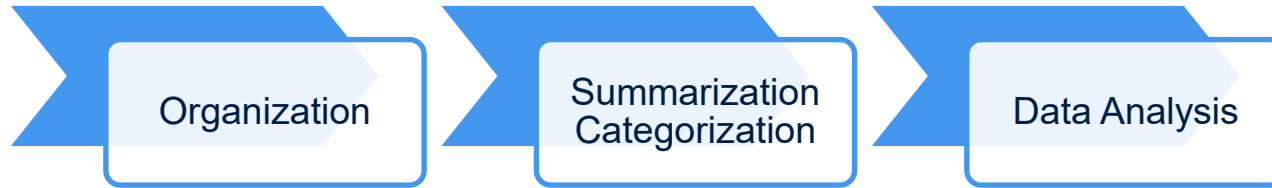
- ▶ The process a researcher uses to reduce data to a story and its interpretation
- ▶ Reducing large amounts of collected data into smaller fragments to make sense of them





# Three Essential Things

- ▶ Three essential things during data analysis

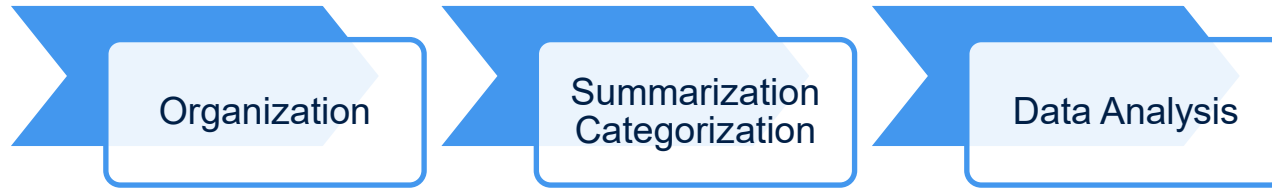


- ▶ Summarization and Categorization → Data Reduction
  - ▶ Finding patterns and themes in data for easy identification and linking



# Three Essential Things

- ▶ Three essential things during data analysis



- ▶ Data Analysis:
  - ▶ Field Analysis – Inscription, Description, Transcription
  - ▶ Top down and bottom up approach

## Why analyse data in research?

A complex vision idea can be conveyed with just a single still image, namely making it possible to absorb large amounts of data quickly.





# Why analyse data in Research?

- ▶ Irrelevant to types of data
  - ▶ Researchers explore their mission
  - ▶ Researchers find the patterns to shape the story
- ▶ Stay open and remain unbiased towards unexpected patterns, expressions and results
- ▶ Tells most unforeseen stories that were not expected
- ▶ Exploratory Research



# Why analyse data in Research?

- ▶ It makes studying data a lot simpler and more accurate
- ▶ Helps to straightforwardly interpret the data
- ▶ Derive insights from data
- ▶ Researcher cannot produce a substantial report without analysing data
- ▶ Helps to find validity
- ▶ Helps to reproduce



# Why analyse data in Research?

- ▶ Data can show whether there was any significant change in variables you are interested in
- ▶ To uncover factors that may be associated with changes in dependent variables
- ▶ To show connections between various factors that have an effect on the results of research evaluation



# Why analyse data in Research?

- ▶ Can shed light on reasons that your work was effective or less effective
- ▶ Provide credible evidence to show
  - ▶ Stakeholders that your research is successful
  - ▶ You have uncovered a truth
  - ▶ Limitations on your research
- ▶ To show you are serious about your evaluation and improving work

# Why analyse data in Research?



- ▶ Show the field what you are learning
- ▶ It can pave a new path for others to implement or do research
- ▶ Helps to improve community efforts and quality of life for people

# Data Analysis in Research: How?

## Clean Data



- ▶ Data can be patchy, inaccurate, inconsistent
  - ▶ Insights will be incomplete or misleading
- ▶ Clean the data for consistency
  - ▶ Don't include duplicate information
- ▶ Small dataset
  - ▶ Easy to clean, manually



# Data Analysis in Research: How?

## Clean Data: Simple Steps



- ▶ Add title rows
  - ▶ Easy to understand what information you got
- ▶ Remove duplicate rows or columns
  - ▶ If you have multiple copies of same records
- ▶ Delete rows or columns
  - ▶ If you are not going to use
- ▶ Standardize your data
  - ▶ Numerical values: numbers, dates or currency



# Data Analysis in Research: How?

## Clean Data



- ▶ Extensive Dataset
  - ▶ Harder to clean data manually
- ▶ Use Data cleaning tools
  - ▶ OpenRefine, Talend
- ▶ Dedicated Data cleaning tools
  - ▶ Cleans up messy, inconsistent information
  - ▶ Ready to use
- ▶ Data Governance strategy
  - ▶ Set clear guidelines

### OpenRefine

OpenRefine is a powerful free, open source tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

Our goal is to empower everyone to meaningfully engage with data by providing an accessible open source tool and nurturing a diverse, supportive community.

[Download](#)

[Participate in the 2026 user survey](#)



#### Main features



##### Faceting

Drill through large datasets using facets and apply operations on filtered views of your dataset.



##### Clustering

Fix inconsistencies by merging similar values thanks to powerful heuristics.



##### Reconciliation

Match your dataset to external databases via reconciliation services.



##### Infinite undo/redo

Revert to any previous state of your dataset and replay your operation history on a new version of it.



##### Privacy

Your data is cleaned on your machine, not in some dubious data laundering cloud.



##### Wikibase

Contribute to Wikidata, the free knowledge base anyone can edit, and other Wikibase instances.

And much more to discover in our documentation.

[Qlik](#) | [Why Qlik?](#) | [Products](#) | [Solutions](#) | [Pricing](#) | [Learn](#)

[Try for Free](#) [Contact Us](#)

#### Qlik AND TALEND TOGETHER

### Qlik Talend: Deliver Trusted, AI-Ready Data

Talend is now part of Qlik. Seamlessly integrate, transform, and govern your data with the power of Qlik Talend!

[Learn More](#)

[Try for Free](#)



# Data Analysis in Research: How?

## Data Governance: Best Practices



- ▶ Create a standard process for when and how to collect data
- ▶ Adopt standardized naming conventions
  - ▶ To reduce inconsistencies in your data
- ▶ Automated Data Collection
  - ▶ Watch out for any error messages
  - ▶ Incorrect data
  - ▶ Diagnose for errors
- ▶ Edit and update data collected in the past
  - ▶ To meet new quality standards



# Cleaning and Standardizing Data

- ▶ Essential step for analysing data.
- ▶ You are less likely to draw incorrect conclusions based on inconsistent data.
- ▶ More likely to get usable inputs



# Data Analysis in Research: How?

## Building Data Analysis Toolkit



- ▶ Many researchers depend on Excel or Google Sheets to store and analyse data
- ▶ There are many other platforms available to analyse data
- ▶ Data Analysis tool depends on the following



# Data Analysis in Research: How?

## Building Data Analysis Toolkit



### Types of Data

The type of data you are analysing

#### Quantitative/ Qualitative Data

Quantitative Data: Ideal for presenting spreadsheets and visualization tools

Qualitative Data: Answers to Questionnaires, Survey Responses, Support Tickets, Social Media messages. Hard to draw out usable insights in a spreadsheet

Categorize or structure qualitative data to analyse

### Amount of Data

The amount of data you are analysing

#### Small/Big Data

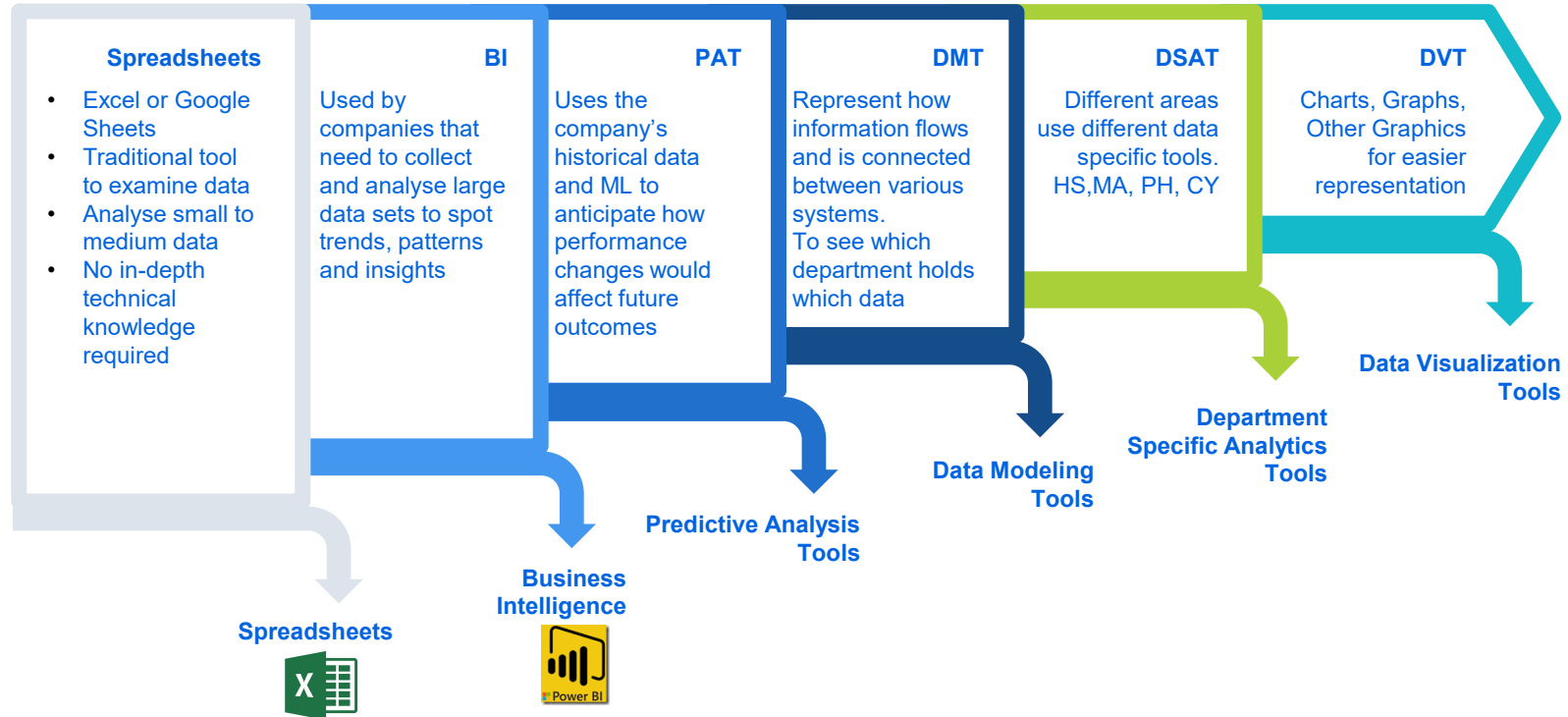
Small Data: For each week/month, analyse information manually

Big Data: Need to invest in tools that automate the data collection and analysis process. Reduce the likelihood of human error, speed up the process



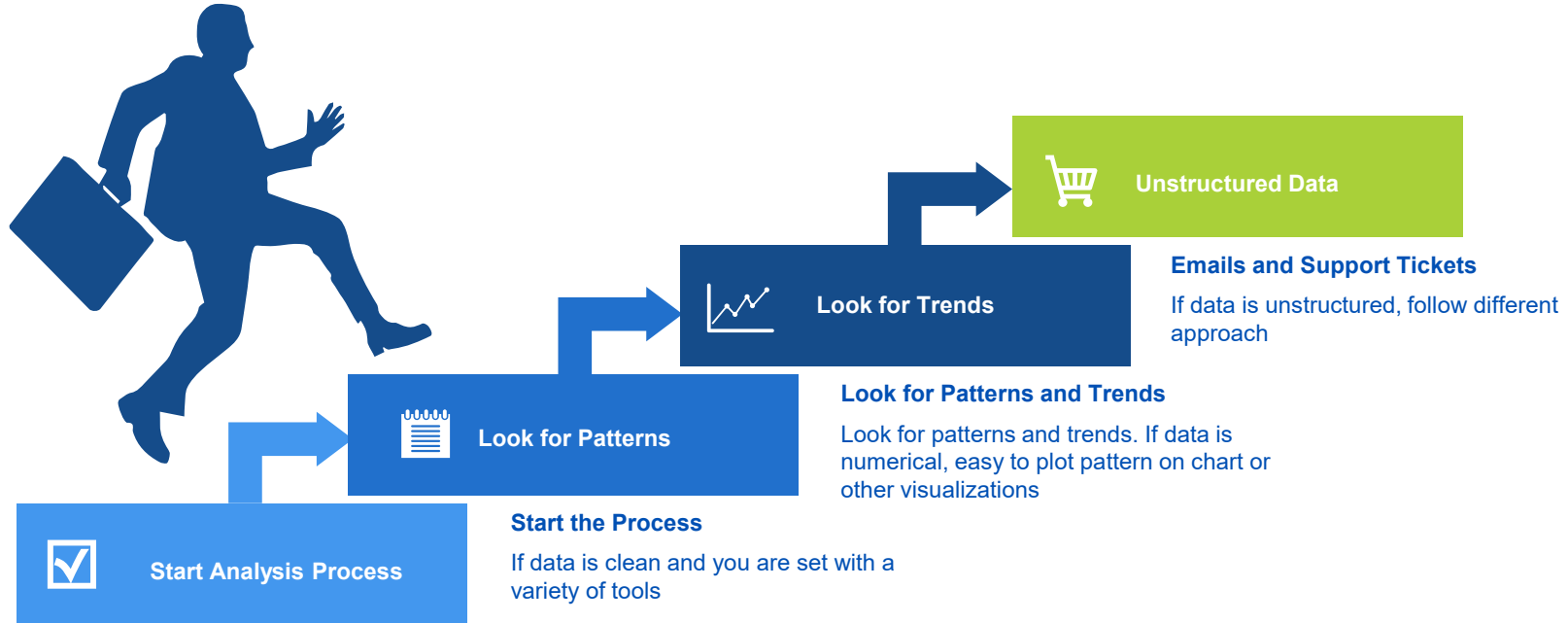
# Data Analysis in Research: How?

## Building Data Analysis Toolkit



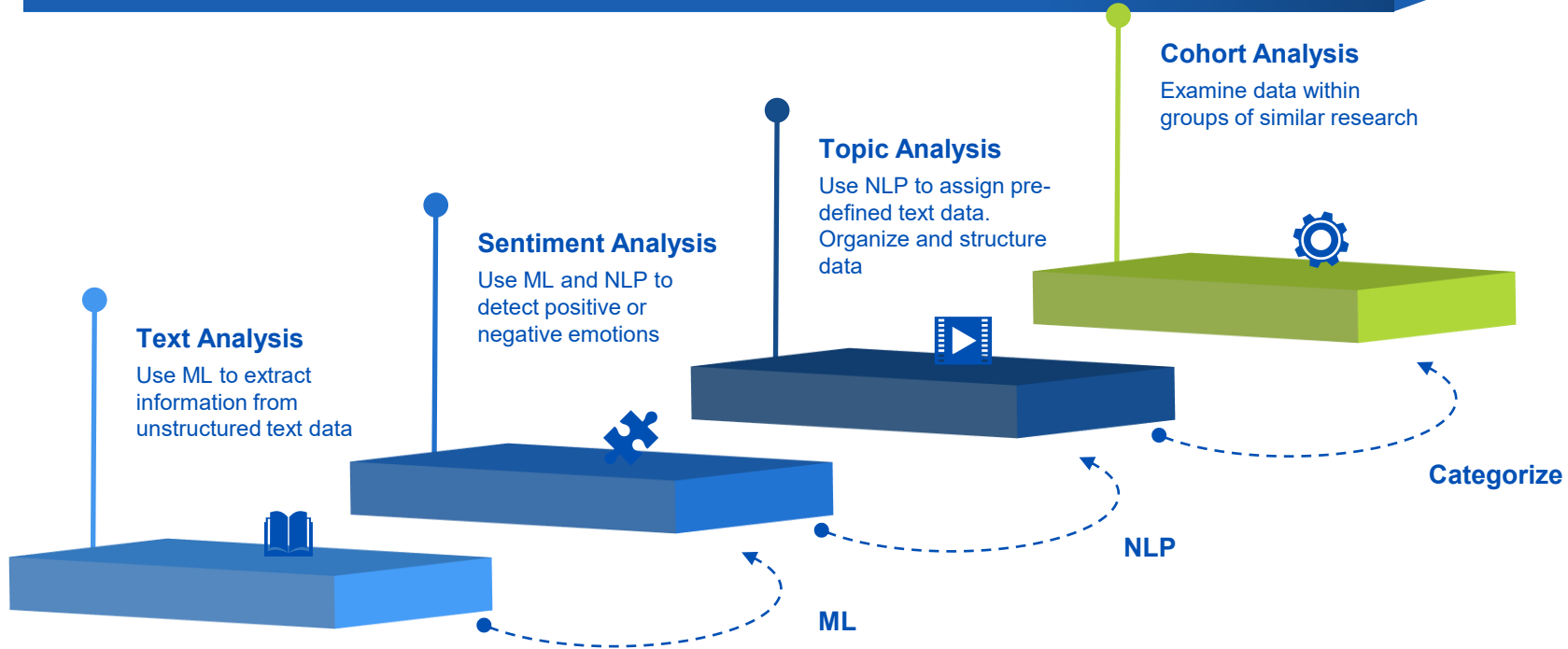
# Data Analysis in Research: How?

Look for patterns and trends in the data



# Data Analysis in Research: How?

Look for patterns and trends in the data



# Data Analysis in Research: How?

Look for patterns and trends in the data: Common Mistakes



- ▶ Don't assume correlation means causation
  - ▶ Assuming that correlation between two things means that one caused the other is called false causality
- ☑ Take time to gather enough evidence to make sure your insights are correct



# Data Analysis in Research: How?

## Compare Current Data Against Historical Trends



- ▶ Struggle to find patterns in your data
  - ▶ Maybe you are seeing your data in a single snapshot
  - ▶ See: How your current data compares to previous time frames
- ▶ Compare against past findings
- ▶ Check for industry benchmark
  - ▶ Caution: Use them as a reference point
- ▶ Completely new: No benchmarks, make a benchmark



# Data Analysis in Research: How?

Look for data that goes against your expectations



- ▶ Set Clear goals and expectations for what you wanted
  - ▶ Confirmation bias
  - ▶ More likely to notice trends that support your existing assumptions
- ▶ Keep an open mind by looking for trends or data points that go against your expectations
  - ▶ Look for outliers
  - ▶ Avoid Cherry-picking findings that support your beliefs



# Data Analysis in Research: How?

Look for data that goes against your expectations



- ▶ Find anomalies in your data
  - ▶ Investigate further: Find explanation
- ▶ How much do outliers in your data skew your results
- ▶ Significant outliers can easily skew averages in your data
  - ▶ Track the median rather than mean
  - ▶ Less skewed by outliers
  - ▶ Alternatively discount these outliers from analysis



# Data Analysis in Research: How?

## Visualization of Data and Interpretation of Results



- ▶ Easy to understand and interpret data when it is presented visually instead of in a spreadsheet
  - ▶ Google Data Studio
  - ▶ Tableau
- ▶ Represent your data
  - ▶ Charts, Graphs or other Graphics
  - ▶ Clearly explain your results
- ▶ Large Data Set
  - ▶ Don't try to communicate too much information at once in your visualizations



# Data Analysis in Research: How?

## Visualization of Data and Interpretation of Results



- ▶ Simple Charts
  - ▶ Easy to view and understand your messages and findings from your data
  - ▶ Series of data visualization tips
- ▶ Dashboards



# Data Analysis in Research: How?

Deeper Details

7

# Data Analysis: How?

## Implement your measurement system



- ▶ Clearly define and describe what measurements are needed
  - ▶ To enable observers to agree on what they're observing and reliably record data in the same way.
- ▶ Select and train observers.
  - ▶ Need training to know what to record;
  - ▶ to recognize key behaviors, events, and conditions;
  - ▶ to reach an acceptable level of inter-rater reliability



# Data Analysis: How?

## Implement your measurement system



- ▶ Conduct observations
  - ▶ At the appropriate times for the appropriate period of time.
  - ▶ Reviewing archival material;
  - ▶ conducting interviews, surveys, or focus groups;
  - ▶ engaging in direct observation; etc.
- ▶ Record data in the agreed-upon ways
  - ▶ Pencil and paper, computer in the field, entering numbers into a program, audio or video, journals, etc



# Data Analysis: How?

Organize the data you have collected



- ▶ Enter necessary data into the system
  - ▶ Comments, description etc
- ▶ Transcribe any audio or video
  - ▶ For clarity and understand the context
- ▶ Record Score tests
- ▶ Sort your information in an appropriate way
  - ▶ Category of information, date, event, place, group etc
- ▶ Transform qualitative into quantitative data
  - ▶ Number of times specific issues were mentioned



# Data Analysis: How?

Data Graphing, Visual Inspection, Statistical Analysis and Other Operations



- ▶ Simple Counting
- ▶ Graphing
- ▶ Visual Inspection of Frequency
- ▶ Rate of behaviour, events over time
- ▶ Using visual inspection of patterns over time to identify discontinuities in the measures over time
- ▶ Calculating the mean, median, mode of a series of measurements
- ▶ Using qualitative interviews, conversations, and participant observations



# Data Analysis: How?

## Significant or Interesting Results



- ▶ Difference within people or groups
- ▶ Difference between or among two or more groups
- ▶ Results that show significant changes
- ▶ Correlations
- ▶ Patterns
- ▶ Obvious important findings



# Data Analysis: How?

## Interpret Results



- ▶ After data organized
  - ▶ Run through statistical or other analysis
- ▶ Most common question
  - ▶ Whether it works or it made a difference
- ▶ In Research
  - ▶ What are the effects of the independent variable on the dependent variable



# Data Analysis: How?

## Interpret Results



- ▶ Your research had exactly the effects on the dependent variables you expected
  - ▶ Clear positive effect at a high level of significance
- ▶ Your research had no effect
  - ▶ No significant results on the dependent variable
- ▶ Your research had negative effect
  - ▶ Unintended consequences
- ▶ Your research had positive effect and side effects
- ▶ Your research had neutral or multiple or mixed effects

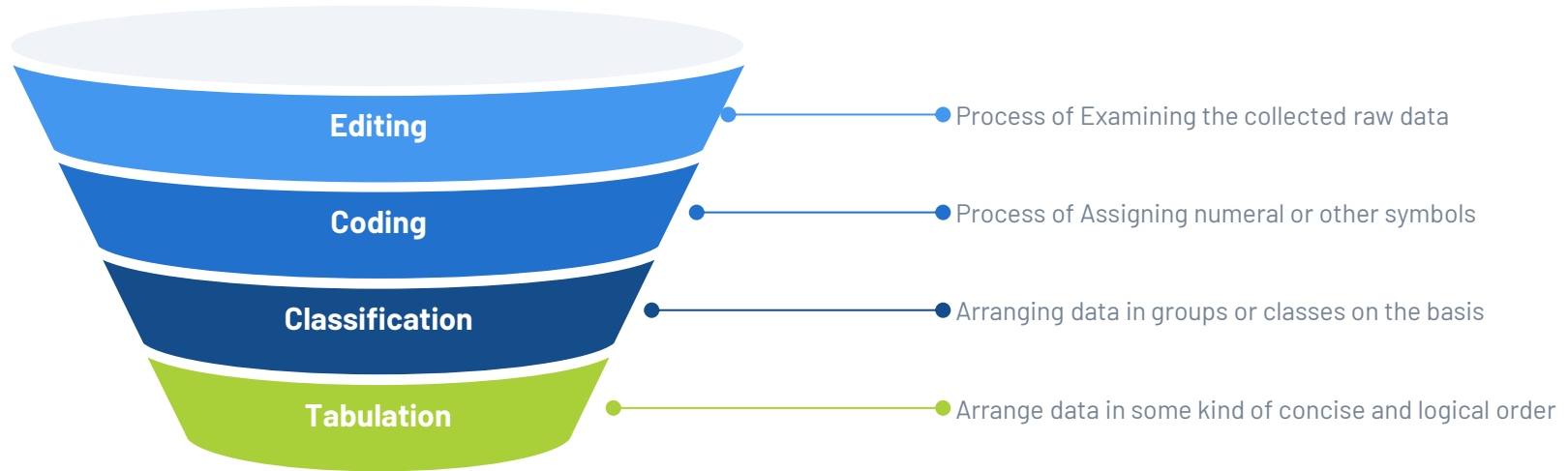


# Processing of Data

Essential for scientific study

8

# Processing Operations



# Processing of Data

## Editing

8a

# Processing Operations

## Editing



- ▶ Process of examining the collected raw data to detect errors and omissions and correct errors
- ▶ Careful scrutiny of the complete questionnaires, surveys
- ▶ Assure that
  - ▶ Data accurate, consistent with facts gathered, uniformly entered

# Processing Operations

## Editing



### Field Editing

- ▶ Review of the reporting forms by investigator for completing
- ▶ At the time of recording
- ▶ Necessary as individual writing style may differ and difficult to decode
- ▶ Do it immediately after field visit
- ▶ Don't fill the gap as you wish

### Central Editing

- ▶ Take place after the field visit to office
- ▶ All editing by a single editor or by a team of editors
- ▶ Correct obvious errors
  - ▶ Entry in wrong place, units, time ..
- ▶ Find proper answer in missing data
  - ▶ Strike out if proper answer is unknown
- ▶ Drop off wrong data

# Processing Operations

Editing: Key points to remember for editors



- ▶ Should be familiar with instructions given to the interviewers and coder
- ▶ Should be familiar with editing instructions supplied and its purpose
- ▶ When crossing out an original entry for a valid reason, should draw a single line
- ▶ Must make entries on the form using distinctive colour and standardised form
- ▶ Keep your initials on all answers you change or supply
- ▶ Keep initials and date of editing on each completed form

# Processing of Data

Coding

8b

# Processing Operations

## Coding



- ▶ Process of assigning numerals or other symbols
  - ▶ To answer the responses that can be put into a limited number of classes
- ▶ Must possess the exhaustive characteristics
  - ▶ Mutually exclusive so that only one cell in a given category set
- ▶ Define every class in one concept
- ▶ It is necessary for efficient analysis
  - ▶ Several replies can be reduced to a small number of classes
- ▶ Coding decision
  - ▶ Decide while designing the questionnaire

# Processing Operations

## Coding



- ▶ Code in the margin with coloured pencil
- ▶ Transcribe the data from the questionnaire to a coding sheet
- ▶ Avoid coding errors
- ▶ Try to minimize coding errors

# Processing of Data

## Classification

8c

# Processing Operations

## Classification



- ▶ Large volume of data
  - ▶ Reduce into homogeneous groups to get meaningful relationships
- ▶ Arranging data in groups or classes on the basis of common characteristics
- ▶ Based on Attributes
  - ▶ Descriptive (Literacy, Honesty)
    - Qualitative phenomena: Presence or absence in the class
  - ▶ Numerical (Height, Weight, Marks)
  - ▶ Statistics attributes

# Processing Operations

## Classification



- ▶ Based on Attributes
  - ▶ Simple classification
    - Only one attribute: Divide into two classes (1) Possess attribute and (2) Does not possess attribute
    - Example: Covid vaccinated
  - ▶ Manifold Classification
    - Two or more attributes simultaneously
    - Divide into number of classes  $2^n$
  - ▶ Should have least possibility of doubt or ambiguity

# Processing Operations

## Classification



- ▶ Based on Class Intervals
  - ▶ Numerical Characteristics
  - ▶ Quantitative phenomena
    - Measured through statistical units
      - Income, Age, Weight
    - Statistics of Variables
  - ▶ Divide classes based on intervals called class-intervals
    - Upper limit, lower limit – class limits, class magnitude
    - Frequency of the given class
  - ▶ Group frequency or simple frequency distribution

# Processing Operations

Classification: Based on Class Intervals – Major Problems



- ▶ How many classes should be there?
  - ▶ No specific answer, but provide a meaningful classification
  - ▶ 5 to 15 are recommended
- ▶ What should be their magnitudes?
  - ▶ Equal magnitude, some cases with unequal magnitude
  - ▶  $i = \frac{R}{1+3.3 \log N}$ ,  $i = \text{size of class interval}$ ,  $R = \text{Range}$ ,  $N = \text{Total}$

# Processing Operations

Classification: Based on Class Intervals – Major Problems



- ▶ How to choose class limits?
  - ▶ Choose the interval such that mid-point of the interval and the actual average of the items of that class interval remain as close as possible
  - ▶ Exclusive: End points or upper limit should be excluded
    - Eg: 10-20
    - Discrete cases
  - ▶ Inclusive: Including end points
    - Continuous



# Processing Operations

Classification: Based on Class Intervals – Major Problems

- ▶ How to decide the frequency of each class?
  - ▶ Tally sheets or mechanical aids

| Income        | Frequency       | Number of frequencies |
|---------------|-----------------|-----------------------|
| < 1 Lakh      | III III III     | 14                    |
| 1 Lakh-2 Lakh | III III III II  | 17                    |
| 2 Lakh-4 Lakh | III III III III | 18                    |
| >4 Lakh       | III II          | 7                     |



# Processing Operations

Classification: Based on Class Intervals – Major Problems

- ▶ How to decide the frequency of each class?
  - ▶ Tally sheets or mechanical aids

| Income        | Frequency       | Number of frequencies |
|---------------|-----------------|-----------------------|
| < 1 Lakh      | III III III     | 14                    |
| 1 Lakh-2 Lakh | III III III II  | 17                    |
| 2 Lakh-4 Lakh | III III III III | 18                    |
| >4 Lakh       | III II          | 7                     |

# Processing of Data

Tabulation

8d

# Processing Operations

## Tabulation



## Tabulation

- ▶ When a mass of data has been assembled, it is necessary for a researcher to arrange the same in some kind of concise and logical order
- ▶ A process of summarising raw data and displaying the compact form
- ▶ Orderly arrangement of data in rows and columns



## Why Tabulation

- ▶ Conserves space and reduces explanatory and descriptive statement to a minimum
- ▶ Facilitates the process of comparison
- ▶ Facilitates summation of items and the detection of errors and omissions
- ▶ Provides a basis for various statistical computations

# Processing Operations

## Tabulation



- ▶ It can be done by hand or mechanical or electronic devices
- ▶ It depends on the size, type of study, cost considerations, time pressures and tabulating computers
- ▶ Hand tabulation for small inquiries and small questionnaires
  - ▶ Direct tally
- ▶ Computer tabulation for large inquiries
- ▶ Tally sheet – similar to frequency table

# Processing Operations

## Tabulation



- ▶ Simple Tabulation
  - ▶ Gives information about one or more groups of independent questions
  - ▶ One way tables to supply answers to questions about one characteristic of data
- ▶ Complex Tabulation
  - ▶ Shows the division of data in two or more categories
  - ▶ Designed to give more information concerning one or more sets of inter-related questions
  - ▶ Two-way tables, three-way table or still higher order or manifold tables
  - ▶ Provides information about several interrelated characteristics of data

# Processing Operations

## Tabulation: General Principles



1. Table should have clear, concise and adequate title
  - a. Should be self-explanatory
  - b. Title should be above the table
2. Every table should be given a distinct number for easy reference
3. Column heading and row headings should be clear
4. Units of measurement under each heading or sub-heading must be indicated
5. Explanatory footnotes, if any, should be placed beneath the table with symbols, if any

# Processing Operations

## Tabulation: General Principles



6. Sources from where data in table obtained
7. Separate columns by lines, make it readable and attractive
8. Keep lines at the top and bottom of the table and column heading (captions)
9. Keep thick lines to separate the data under one class from the data under another class
10. Keep thin lines to separate sub-divisions of the classes
11. Number the columns to facilitate reference

# Processing Operations

## Tabulation: General Principles



12. Columns whose data are to be compared should be kept side by side
13. Keep percentages, averages close to the data
14. Better to approximate figures before tabulation
15. Emphasise relative significance of certain categories, different kinds of type
16. Use appropriate spacing and indentations
17. All column figures should be properly aligned
18. Keep decimal points, + and – signs in perfect alignment

# Processing Operations

## Tabulation: General Principles



19. Avoid abbreviations as much as possible
20. Don't use ditto marks
21. Miscellaneous and exception items should be kept at the last row of the table
22. Make logical, clear, accurate and simple table
23. If data is very large, don't provide crowded single table, it would be inconvenient
24. Total of rows should be placed in the extreme right column
25. Total of columns should be placed at the bottom of the table

# Processing Operations

## Tabulation: General Principles



26. Arrangement of categories in a table may be

- a. Chronological
- b. Geographical
- c. Alphabetical
- d. According to magnitude

27. Table must suit the needs and requirements of an investigation

# Processing of Data

Problems in Processing

8e

# Processing Operations

## Problems in Processing



- ▶ Don't Know Responses or DK responses
  - ▶ Some responses are difficult to handle
  - ▶ If DK group is simple, it is of little significance
  - ▶ If DK group is big, it is a major concern
- ▶ Is the question which elicited DK response useless
  - ▶ Researcher actually may not know the answer
    - Fine. Legitimate DK response
  - ▶ Researcher may fail in obtaining the appropriate information
    - More likely to be a failure of the questioning process

# Processing Operations

## Problems in Processing



- ▶ DK Responses
  - ▶ Best way is to design a better type of question
  - ▶ Eliminate allocation of DK answers
  - ▶ Keep DK responses in a separate category in tabulation
- ▶ Use of Percentages:
  - ▶ Percentages are used in data presentation because they simplify numbers
  - ▶ Reduces to the range 0 to 100
  - ▶ It is a kind of normalization for relative comparisons

# Processing Operations

## Problems in Processing



- ▶ Use of Percentages:
  - ▶ Two or more percentages must not be averaged unless each is weighted by the group size from which it has been derived
  - ▶ Avoid use of too large percentages – difficult to understand, confusing, defeats the purpose
  - ▶ It can hide the base from which they have been computed
  - ▶ Real differences may not be correctly stated
  - ▶ Percentage decrease can never exceed 100% for calculating percentage of decrease
  - ▶ Higher should invariably be taken as the base

# Processing Operations

## Problems in Processing



- ▶ Use of Percentages:
  - ▶ Should generally be worked out in the direction of the causal factor in case of two-dimensional tables
  - ▶ Select the most significant factor out of the two given factors as the causal factor.

# Types of Data

Qualitative, Quantitative, Categorical

9



# Types of data

## ▶ Qualitative Data

- ▶ Words and descriptions
- ▶ Subjective
- ▶ Harder to analyse
- ▶ Taste, experience, opinion
- ▶ Focus groups, interviews

## ▶ Quantitative Data

- ▶ Data expressed in numbers
- ▶ Numerical figures
- ▶ Distinguished into categories, grouped, measured, calculated or ranked
- ▶ Age, rank, cost, length

## ▶ Categorical Data

- ▶ Data presented in groups
- ▶ An item included in one categorical data cannot belong to more than one group
- ▶ Marital status, Drinking habit

# Data Analysis in Qualitative Research

10

# Data Analysis in Qualitative Research

## Finding Patterns in the Qualitative Research



- ▶ It is different from numerical data because the quality is made up of words, descriptions, images, objects, and symbols
- ▶ Getting insight from complicated information is a complicated process
- ▶ Used for exploratory research
- ▶ Several ways to find patterns
  - ▶ Textual information: a word-based method
    - Global technique, finds repetitive and commonly used words

# Data Analysis in Qualitative Research

## Finding Patterns in the Qualitative Research



- ▶ Several ways to find patterns
  - ▶ Keyword context: Try to understand the concept by analysing the context
  - ▶ Scrutiny-based: One of the highly recommended text analysis
    - To identify a quality data pattern
  - ▶ Compare and contrast
    - To differentiate how a specific text is similar or different from each other
  - ▶ Metaphors: To reduce the data pile and find patterns in it, easier to connect data with theory
  - ▶ Variable partitioning: Used to split variables so that researchers can find more coherent descriptions

# Data Analysis in Qualitative Research

## Methods used



- ▶ **Content Analysis:** Widely accepted and most frequently employed technique
  - ▶ To analyse the documented information from text, images
  - ▶ Depends on research questions to predict when and where to use this method
- ▶ **Narrative Analysis:** Used to analyse content gathered from various sources such as personal interviews, field observation, surveys
  - ▶ Stories, opinions shared by people
  - ▶ Find answers to research questions

# Data Analysis in Qualitative Research

## Methods used



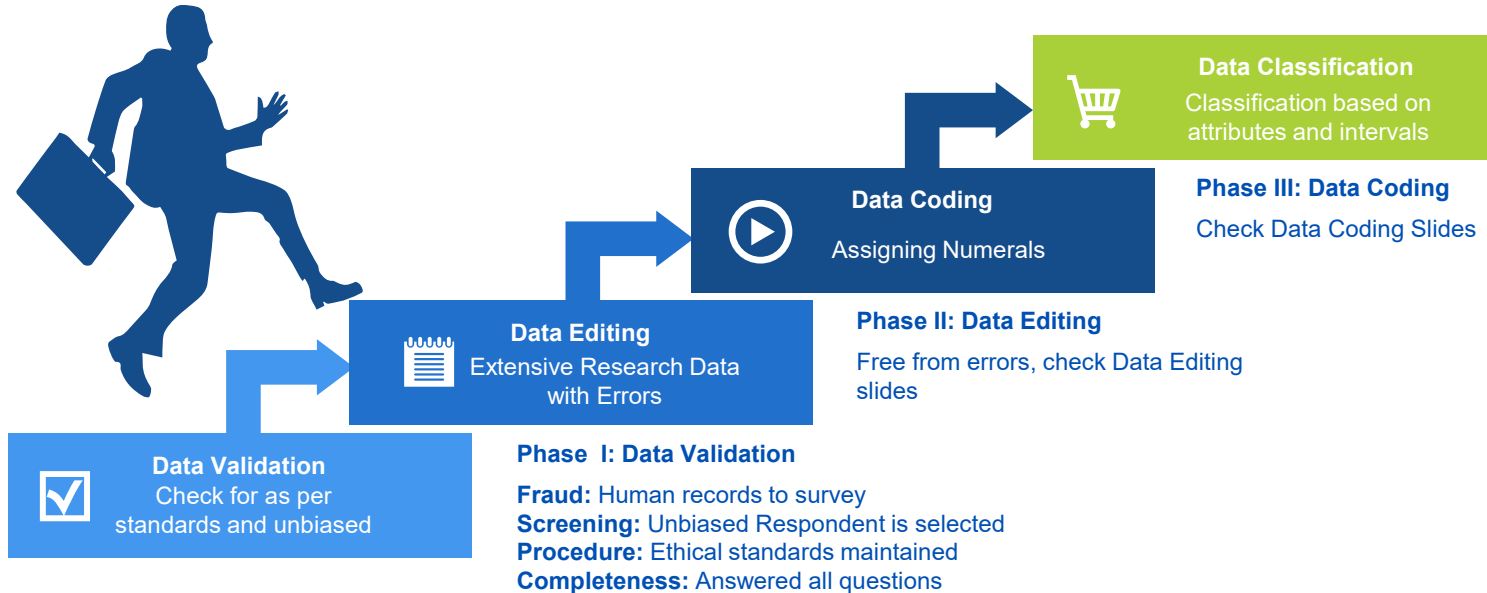
- ▶ **Discourse Analysis:** Similar to narrative analysis, to analyse the interactions with people.
  - ▶ Considers the social context under which or within which communication between the researcher and respondent takes place
  - ▶ Focuses on lifestyle and day-to-day environment
- ▶ **Grounded Theory:** When you try to explain why a particular phenomenon happened
  - ▶ Applied to study data about the host of similar cases occurring in different settings
  - ▶ Researcher may alter explanations or produce new ones until they arrive at some conclusion

# Data Analysis in Quantitative Research

11

# Data Analysis in Quantitative Research

## Preparing Data for Analysis



# Data Analysis in Quantitative Research

## Main Characteristics



- ▶ The data collected are precise and structured
- ▶ Results rely mostly on large sample sizes
- ▶ Due to its high reliability, researchers can replicate and repeat the studies
- ▶ The research questions are clearly defined with an intention to seek precise answers related to the objectives
- ▶ All collected data comprises numbers and statistics
- ▶ The analysts rearranged the received data into tables, figures, charts and non-textual forms

# Data Analysis in Quantitative Research

## Methods used



- ▶ **Descriptive Statistics/Univariate Analysis:** Describes the basic features of various types of data
  - ▶ Presents a meaningful pattern
  - ▶ Does not go beyond conclusions
  - ▶ Conclusions are based on hypotheses formulated by the researcher
- ▶ **Inferential Statistics/Multivariate Analysis:** Describes how to make predictions about a large population from a collected sample. E.g. Movie review from audience
  - ▶ Estimating parameters: Statistics from sample research data, demonstrates population parameters
  - ▶ Hypothesis Test: Sampling Data to answer research questions

# Data Analysis in Quantitative Research

Methods used: Descriptive Statistics



- ▶ Measures of Frequency:
  - ▶ Count, percent, frequency
  - ▶ To denote particular event occurrence
  - ▶ To showcase the response
- ▶ Measures of Central Tendency:
  - ▶ Mean, Median, Mode
  - ▶ To demonstrate distribution by various points
  - ▶ To showcase averagely or commonly indicated response

# Data Analysis in Quantitative Research

Methods used: Descriptive Statistics



- ▶ **Measures of Dispersion or Variation:**
  - ▶ Range, Variance, Standard Deviation
  - ▶ High or low points
  - ▶ Difference between observed score and mean
  - ▶ Identify the spread of scores
  - ▶ Showcase spread out
- ▶ **Measures of Position:**
  - ▶ Percentile ranks, Quartile Ranks
  - ▶ Standardized scores to identify the relationship between different scores
  - ▶ To compare scores with the average count

# Data Analysis in Quantitative Research

Methods used: Inferential Statistics



- ▶ **Correlation:**
  - ▶ Interested to find the relationship between two or more variables
- ▶ **Cross-Tabulation**
  - ▶ Contingency Tables
  - ▶ Used to analyse relationship between multiple variables
  - ▶ E.g: Gender and Age.
    - Two-dimensional cross-tabulation could help to show number of males in each age category

# Data Analysis in Quantitative Research

Methods used: Inferential Statistics



## ▶ Regression Analysis:

- ▶ To understand the strong relationship between two variables
- ▶ Predictive Analysis Method
- ▶ It has an essential factor called dependent variable
- ▶ It has multiple independent variables
- ▶ Both independent and Dependent Variables are error-free random numbers

## ▶ Frequency Tables:

- ▶ For Testing Degree to which two or more vary or differ in an experiment
- ▶ ANOVA Table
- ▶ Analysis of Variance

# Data Analysis in Quantitative Research

## Tips to follow while Reporting



- ▶ Must explain all the collected data in detail along with all the statistical treatment utilized while investigating the research problems
  - ▶ No need to explain his/her interpretation in this section
- ▶ Must report all the limitations and unanticipated events that happened while collecting the data
- ▶ If any, explain how the actual analysis differed from the planned analysis in the current study
- ▶ Must also explain how missing data was handled
- ▶ Justify the missing data does not undermine the credibility of the analysis

# Data Analysis in Quantitative Research

## Tips to follow while Reporting



- ▶ Justify cleaning the data set
  - ▶ Mention the techniques they used to clean their data set
- ▶ Do not use complex and critical statistical methods unless it's mandatory
- ▶ Justify reasons for using the analytical process
  - ▶ Mention the computer program and references used while preparing the report
- ▶ All assumptions taken by analysts must be discussed systematically
  - ▶ Suggest various steps used to conduct the report without violating any rules

# Data Analysis in Quantitative Research

## Tips to follow while Reporting



- ▶ If inferential statistics is used, mention the descriptive statistics, sample sizes of each variable, confidence intervals.
- ▶ To report the actual p-Value, must provide the value of the test statistics, the direction of the test statistics, the degree of freedom and the level of significance
- ▶ To represent exact values, it's crucial to use tables and figures
- ▶ While mentioning confidence intervals, better to include graphic representations
- ▶ Give a few details about your tables and figures, making it easier for readers to know what the tables & figures represent

# Final Remarks

12



# Final Remarks

- ▶ Researchers/Analysts must have the necessary skills to analyse the data
- ▶ Getting trained to demonstrate a high standard of research practice
- ▶ Researchers/ Analysts must possess more than a basic understanding of the rationale for selecting one statistical method over another to obtain better data insights
- ▶ Data analysis methods in research differ by scientific discipline
  - ▶ Getting statistical advice at the beginning of analysis helps design a survey questionnaire
  - ▶ Select data collection methods, and choose samples appropriately



# Final Remarks

- ▶ The primary aim: To derive ultimate insights that are unbiased
- ▶ A biased mind while collecting data, selecting a method, or choosing an audience sample can lead to a biased inference
- ▶ Irrelevant to the sophistication used in research, data analysis is not enough to rectify the poorly defined objective outcome measurements
- ▶ Irrespective of faulty design or unclear intentions, lack of clarity might mislead readers, so avoid the practice



# Final Remarks

- ▶ Motive behind data analysis in research: To present accurate and reliable data
- ▶ Avoid statistical errors
- ▶ Find a way to deal with outliers, missing data, data altering, data mining
- ▶ Develop graphical representation.

# Thanks!

## Any questions?

You can find me at:

- ▶ [panch.m@iittp.ac.in](mailto:panch.m@iittp.ac.in)

