Ministry of Statistics and Programme Implementation

Monitoring Social Sustainable Development Goals in India Use of Big Data

Hyderabad

27th September, 2019

Sustainable Development Goals (SDG)

- A transformative plan of action based on 17 Sustainable Development Goals
- To address urgent global challenges over the next 15 years.
- A road map for people and the planet
  - Ensure sustainable social and economic progress worldwide
  - Seeks to eradicate extreme poverty
  - And integrate and balance the three dimensions of sustainable development
    - Economic, social and environmental
    - in a **comprehensive global vision**.

SDG Vision

The new agenda is a promise by leaders to all people everywhere.

It is a universal, integrated and transformative vision for a better world.

It is an agenda for people, to end poverty in all its forms.

An agenda for the planet, our common home.

An agenda for shared prosperity, peace and partnership.

It conveys the urgency of climate action.

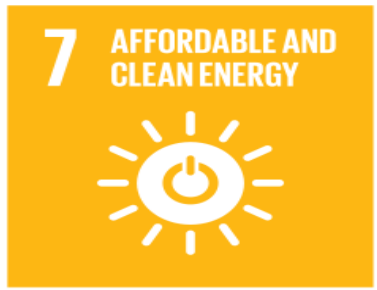
It is rooted in gender equality and respect for the rights of all.

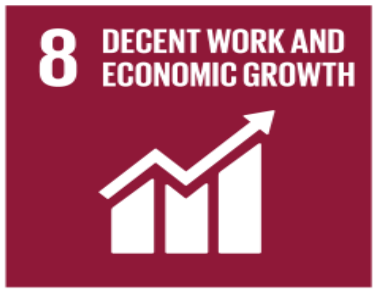
Above all, it pledges to **Leave No One Behind**.

BAN Ki-Moon

Ex-Secretary-General, United Nations

SUSTAINABLE DEVELOPMENT GOALS





Sustainable Development implies

- Promotion of sustainable, inclusive and equitable economic growth
- Creating greater opportunities for all
- Reducing inequalities
- Raising basic standards of living
- Fostering equitable social development and inclusion
- Promoting integrated and sustainable management of natural resources and ecosystems

Transition from MDGs to Global SDGs

MDGs (2000-2015)		SDGs (2015-2030)
Developing country focused	→	Universal
Social	→	Social, Economic, and Environmental
Foreign Aid	→	Domestic Investment, Private Flows, and Aid
Official Statistics and Administrative Data	→	Big Data, Citizen Generated Data, Geospatial and Earth Observation Data, Open Data, and more

SDG Goals (2015-2030)			MDG Goals (2000-2015)	
1	No Poverty	→	MDG 1	End hunger and poverty
2	Zero Hunger			
3	Good Health & Well Being		MDG 4	Reduce Child Mortality
		→	MDG 5	Improve Maternal Health
			MDG 6	Combat HIV/AIDS, Malaria and TB
4	Quality Education	→	MDG 2	Achieve Universal Primary Education
5	Gender Equality	→	MDG 3	Promote Gender Equality and Empower Women
6	Clean Water and Sanitation		**New Indicators**	
7	Affordable and Clean Energy			
8	Decent Work and Economic Growth			
9	Industry, Innovation and Infrastructure			
10	Reduced Inequalities			
11	Sustainable Cities and Communities			
12	Responsible Consumption and Production		MDG 7	Ensure Environmental Sustainability
13	Climate Action	→		
14	Life below water			
15	Life on Land			
16	Peace, Justice and Strong Institutions		**New Indicator**	
17	Partnerships for the Goals	→	MDG 8	Develop Global Partnership for Development

Seven New Goals under SDG

New Goals	
Goal 6	Ensure availability and sustainable management of **water and sanitation for all**
Goal 7	Ensure access to affordable, reliable, sustainable and modern **energy for all**
Goal 8	Promote sustained, inclusive and **sustainable economic growth**, full and productive **employment** and decent work for all
Goal 9	Build resilient **infrastructure**, promote inclusive and sustainable industrialization and foster innovation
Goal 10	Reduce **inequality** within and among countries
Goal 11	Make **cities and human settlements** inclusive, safe, resilient and sustainable
Goal 16	Promote **peaceful** and **inclusive societies** for sustainable development, provide access to **justice for all** and build effective, accountable and inclusive institutions at all levels

SDG Key Features

- SDG not legally binding
  - But will become de facto international obligation/commitment
  - Reorient domestic spending priorities of countries
- Member Countries
  - Take ownership of the SDGs
  - Establish a national framework for achieving these goals.
  - Align/design sustainable development policies and programmes.
  - Responsible for follow up and review at the national, regional and global level
- Global interconnect

Relevance of Indicators

- Two-fold relevance:
  - Assist Governments to stay on course
  - Accountability to hold and connect all stakeholders to the SDGs
  - For management purposes, the indicators need to be accurate and frequent.
- Regular review of Indicators necessary
  - Introduction of new ones
  - Replacement/removal of earlier ones that are no longer relevant

SDG Roadmap - India

- To achieve SDG goals, India needs to mainstream SDGs into ongoing national policies and closely monitor progress.
- India has been pursuing the path of ***Sabka Saath Sabka Vikaas***
  - Consistent with the pledge of ***'no one will be left behind'***.
  - Programmes and other developmental schemes address SDG targets .
- Success in SDG attainment dependent on
  - Coordinated efforts of all stakeholders (GOI, States, Civil Society etc)
  - Robust monitoring mechanism.
- Close coordination amongst policy formulators and statistical establishments.
- MoSPI playing a lead role in statistical coordination

Indicator Development Process

- MoSPI developed the National Indicator Framework for monitoring progress on SDGs and associated targets.
  - Scientifically design/evolve indicators for various targets
  - Periodically measurable
  - Indicator should meet the criterion of fitment for purpose.

- Government is to ensure that
  - Measurement of progress of goals and targets under SDGs is undertaken as per the **statistically robust methodology** and
  - Adopt only robust indicators which can withstand statistical scrutiny now or in future.

Challenges in National Indicator Framework

- Definitional/Measurement issues
- Disaggregation
  - Geographic/spatially
  - Temporal/
  - Gender
  - Social groups
- Periodicity/Timeliness
- Baseline Data to be made available
- Capacity Development of statistical personnel

Measurement Challenges - Global Indicators

- Food Insecurity Experience Scale (FIES)
- Proportion of population living in households with access to basic services
- Prevalence of undernourishment
- Proportion of total adult population with secure tenure rights to land
  - With legally recognized documentation **and**
  - Who perceive their rights to land as secure, by sex and by type of tenure
- Average income of small-scale food producers, by sex and indigenous status
- Proportion of agricultural area under productive and sustainable agriculture
- Global food loss index

Role of States

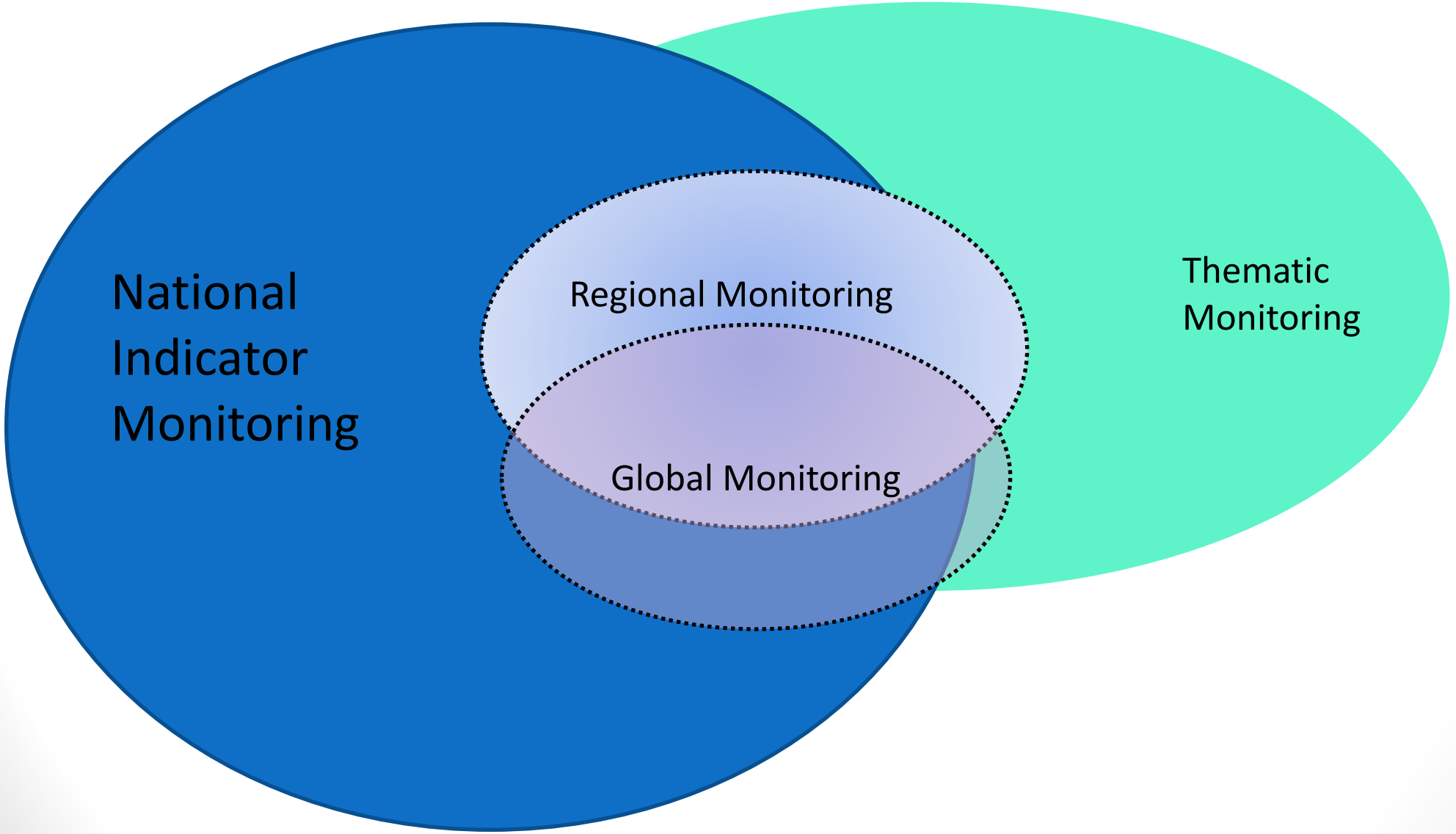
- MoSPI to bring out annual national report on SDGs based on inputs from
  - State Government
  - Central Ministries / Departments
- States Governments role
  - State SDG frameworks to be developed in sync with the National Indicator Framework
  - Align state statistical systems for effective monitoring and implementation of SDGs.
  - Using technology to ensure regular data flow system from States to Centre
  - State statistical system to also regularly undertake NSS surveys with state sample for disaggregated data

SUSTAINABLE DEVELOPMENT GOALS: DATA ROADMAPS

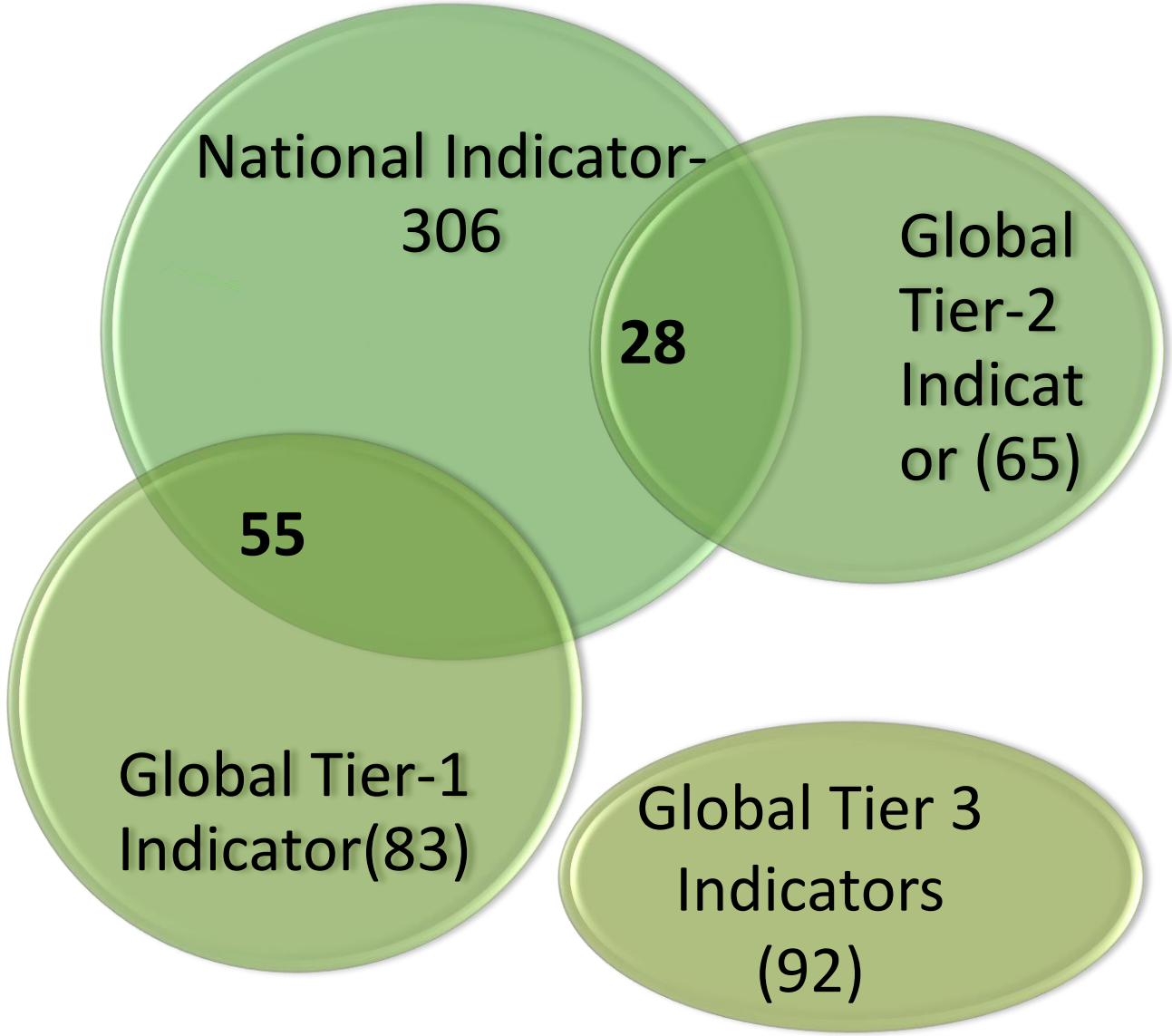
Comprehensive & Holistic National Framework

- 17 Goals
- 169 Targets
- >300 Indicators
- 15 Nodal Ministries
- 56 Data Source Agencies
- Highly interconnected
- Launching a data revolution for the SDGs

National vs Global comparisons



National Indicators vis-à-vis Global Indicators



Data/Measurement Challenges and Interpretation

THE CHALLENGES:

- Data on entire groups and key issues are unavailable.
- Data are not dynamic or disaggregated.
- Data quality is poor and major gaps remain.
- Data that exist are often not useable.
- Data that are useable are not accessible or open.
- Data that are accessible are often not used effectively.

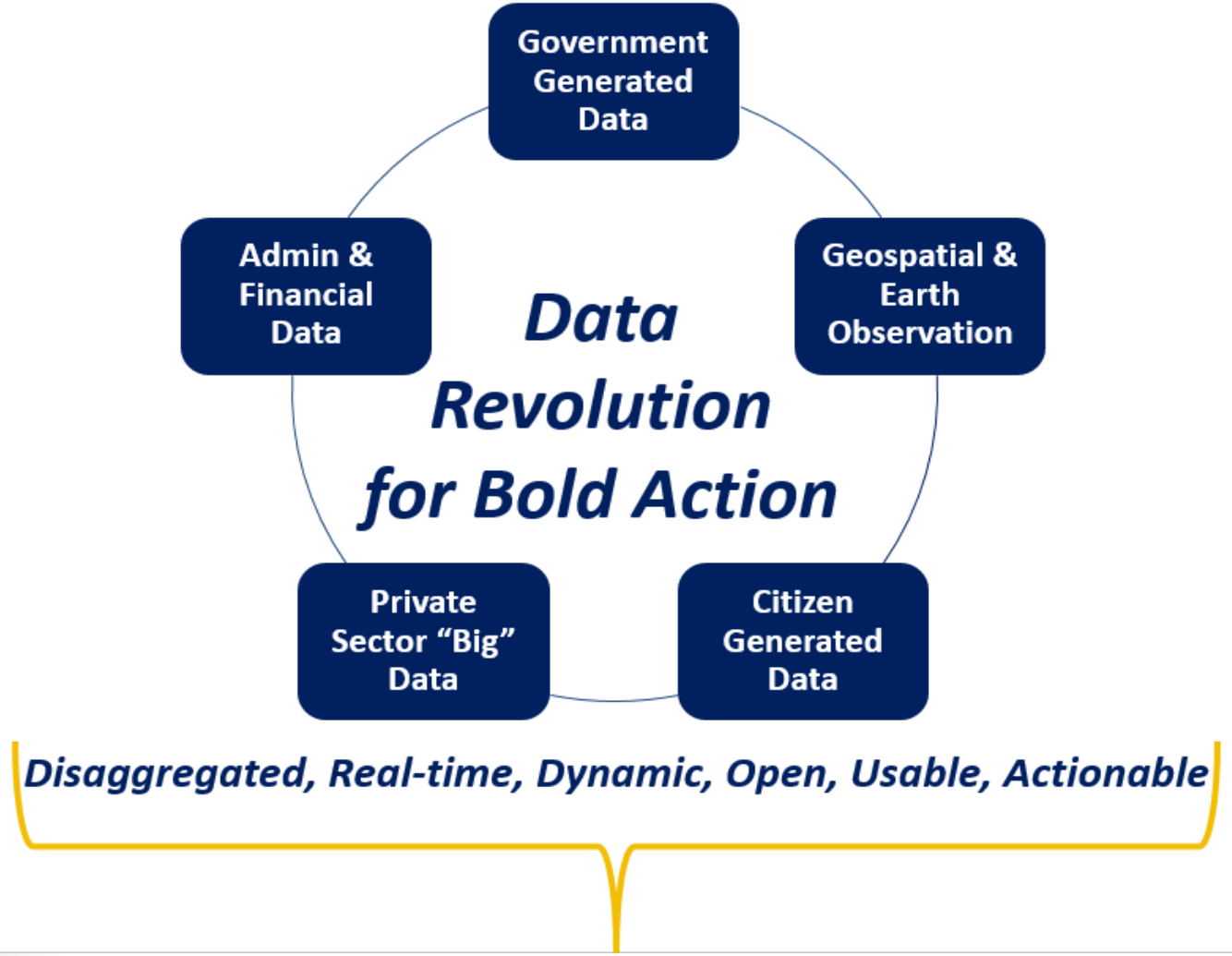
DATA CHALLENGES ….. LEAVE TOO MANY BEHIND

Extensive use of Information Technology

- **Migration to online MIS**
- **Increase openness and leverage existing data.**
  - By bringing together high-value data sets across sectors for achieving the SDGs.
- **Improve the effective use of data.**
  - By providing robust APIs to empower data to be combined in new ways for further insights.
- **Enable data for action and decision making**
  - By empowering the developer community to create rich applications and visualizations.

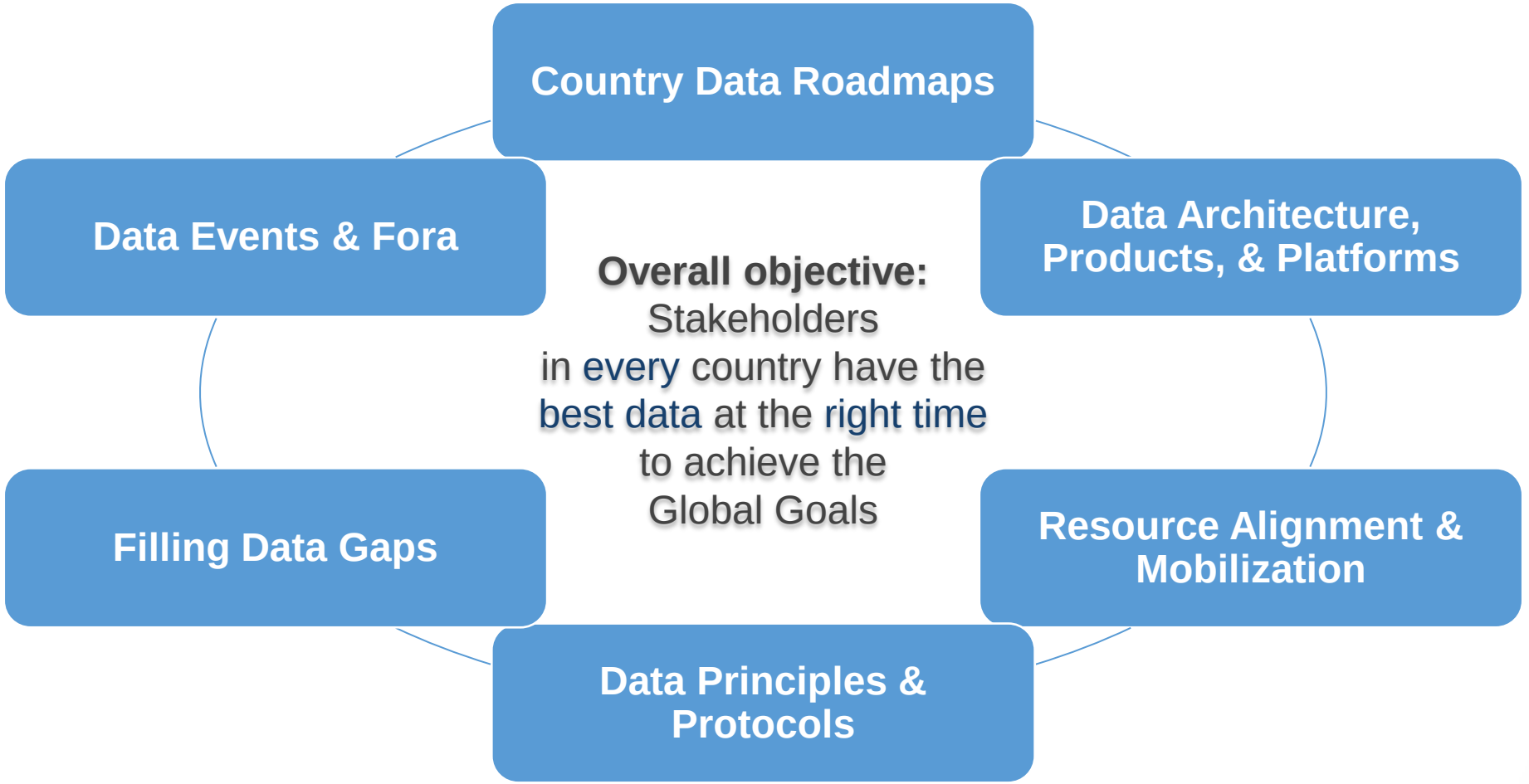
Harnessing the Data Revolution

"Data is the Oil of the 21st Century"



- Fill data gaps more efficiently, frequently and cost effectively
- Real-time, dynamic, disaggregated data
- Official and non-official data
- Use innovative approaches and range of stakeholder to solve problems

Convening, Connecting, Catalyzing



Data Roadmaps to harness Data Revolution

- It's an **action plan** with short and long-term **goals** for addressing specific **data needs and priorities** in regards to **SDG implementation**.
- A data roadmap is ideally developed by governments at local, subnational or national levels according to local context and priorities:
  - Multi-stakeholder involvement
  - Situation assessment
  - Priority Mapping
  - Data/technology gaps assessment
  - Requirements analysis
  - Commitments to Action
- A data roadmap is part of an **iterative and adaptive** planning process.

Country Led Approaches

- The Data Roadmaps for Sustainable Development approach is iterative, based on experiences and implementation models from partner countries

DATA FOR WHAT?

Improved Decision-Making and Policy

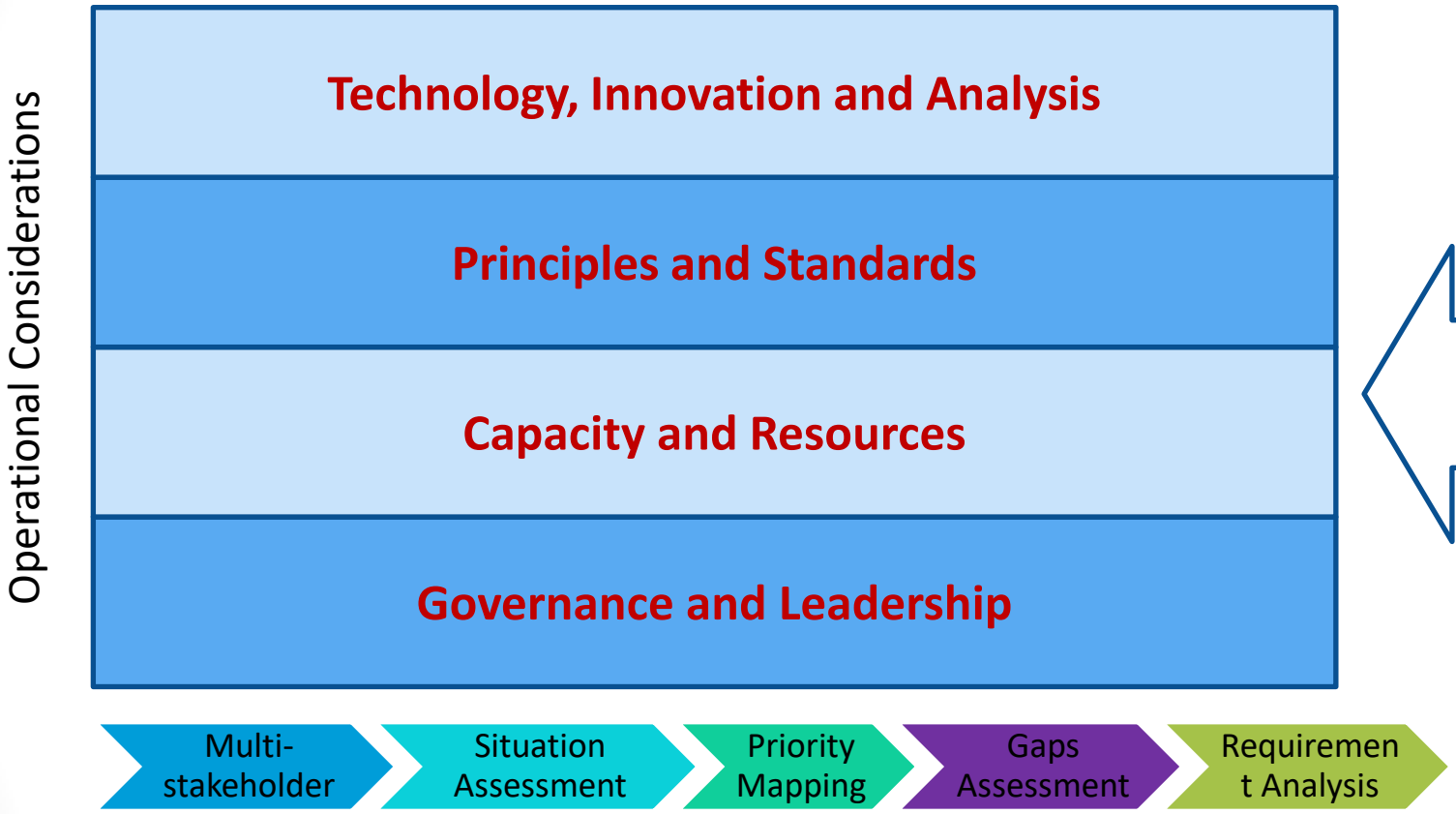
Increased Citizen Empowerment

Increased Innovation and Entrepreneurship

DATA FOR ACTION!!!

To Achieve and Monitor Sustainable Development

Data Support for National and Subnational Stakeholders



- Many processes are already in place.
- Tools and methods have been developed or are being developed.
- A data roadmap process acknowledges good practice, lessons learned, and tools applied from peers to inform its development.

MOSPI's Data Strategy Ecosystem

- Increase role of statistics in policy, development and social well-being
- **Enhanced role of IT** in data life cycle (collection to dissemination)
  - Reduction in respondent burden
  - Improved quality
  - New products
  - New processes
  - National Data Warehouse – highly interconnected and integrated system
- **Leveraging partnership** with leading institutions
  - Technical knowhow
  - Capacity building on contemporary methodologies
- **Data Sharing Policy** to address privacy and confidentiality issues
  - Sharing protocols
  - Privacy of information
  - Ownership of Data

Creating a Data Ecosystem

- Highly integrated and interconnected data system
- API Highways and Digital Marketplace
- Data Sharing Policy to address privacy and confidentiality issues
- Ownership of Data

"In God we trust, all others must bring data." **W Edwards Deming**

"If you torture the data long enough, it will confess." **Ronald Coase**

"An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question" **John Tukey**

"If we have data, let's look at data. If all we have are opinions, let's go with mine." **Jim Barksdale, former CEO of Netscape**

"Information is the oil of the 21st century, and analytics is the combustion engine." **Peter Sondergaard, Sr. Vice President, Gartner Research**

Future Steps

- About Data Analytics
- Data Visualization
- Big Data: Data Mining and Artificial Intelligence

Data Analytics

- Provides a better understanding of a phenomenon through the data generated.

- Tools, techniques and processes for continuous exploration and investigation of past data
  - Gain insights
  - Assist decision making and problem solving – data driven.

- An integration between
  - Phenomenon/problem context
  - Technology

FRAMEWORK- DATA-DRIVEN DECISION MAKING

Problem or Opportunity Identification

Domain knowledge is important
Organisational capacity to leverage capabilities of analytics.

Collection of relevant data

“Relevant data" need not be known in advance.
Interactive and Iterative process

Data Pre-processing

Data preparation and data processing – most critical part.
Data imputation, Missing Values, Creation of additional variables.

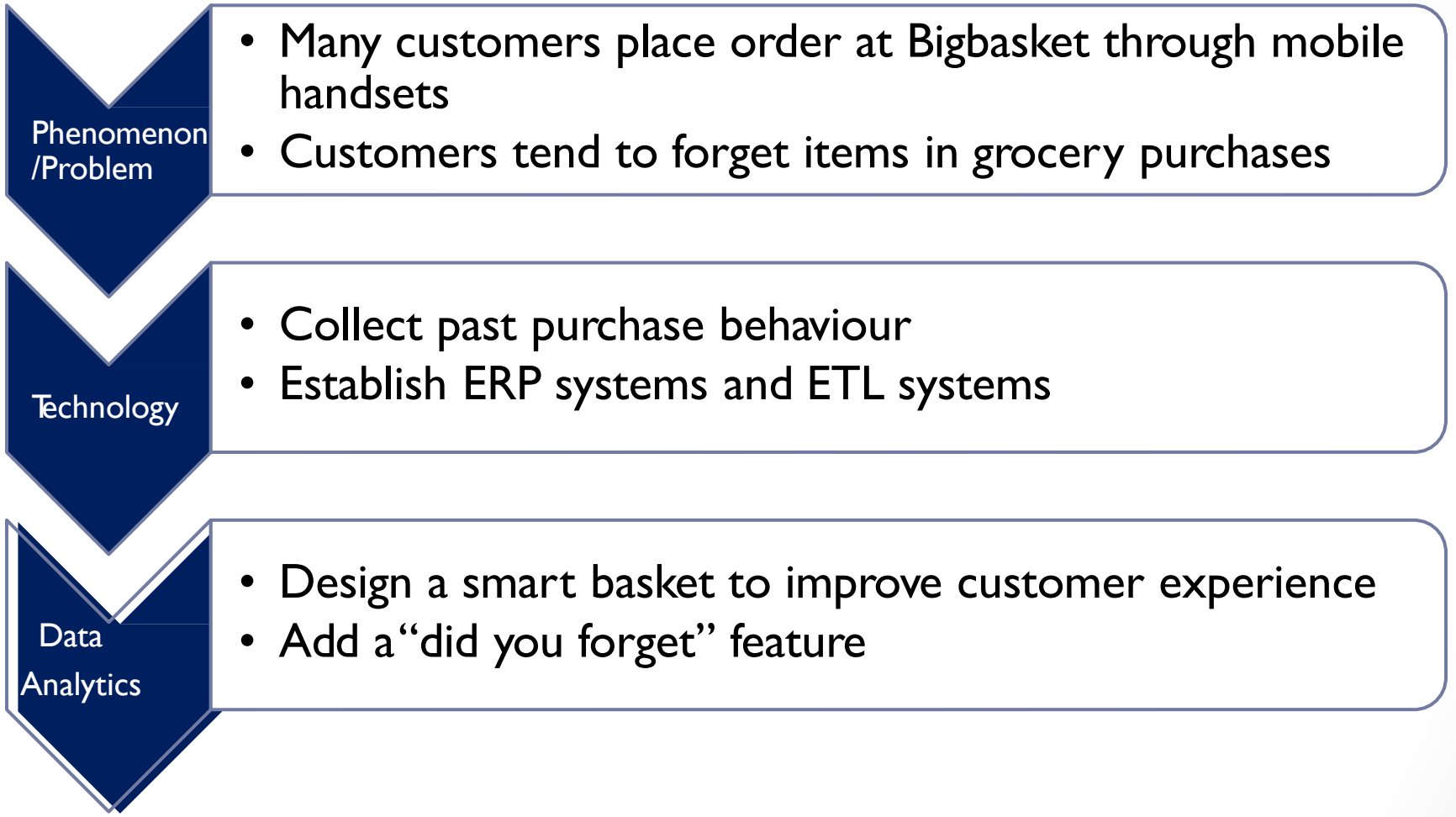
Model Building

Analytics model building is an iterative process.
Several analytical tools and solution procedures will be used to find the best model.

Communication of the data analysis

Communication strategy to stakeholders.
Innovative data visualization techniques to enthuse users.

Analytics in E-Commerce (Big Basket)



Big
Bas

Search for more than 10,000 products

Search

Vour ..ket
Oitems

CHtO:OUT

SHOP

OFFERS

NEW ARRIVALS

SHOPBYUST

HOME > SMART BASKET

Smart Basket (125)

llection of products that you on most or buy often

stL£CTID PRODUCTS (O)
SeL"Ctal!

FOR SELECTED PRODUCTS:

ADD TO BASKET

COPYTOUST U

.- Fnuts & Vegetables (28)

Select all

0

On1on • MrdJum
11;;
{1700
AOOw

Potato
1
t2SOO

Pom rar.atc Kcs.r
1
{18100

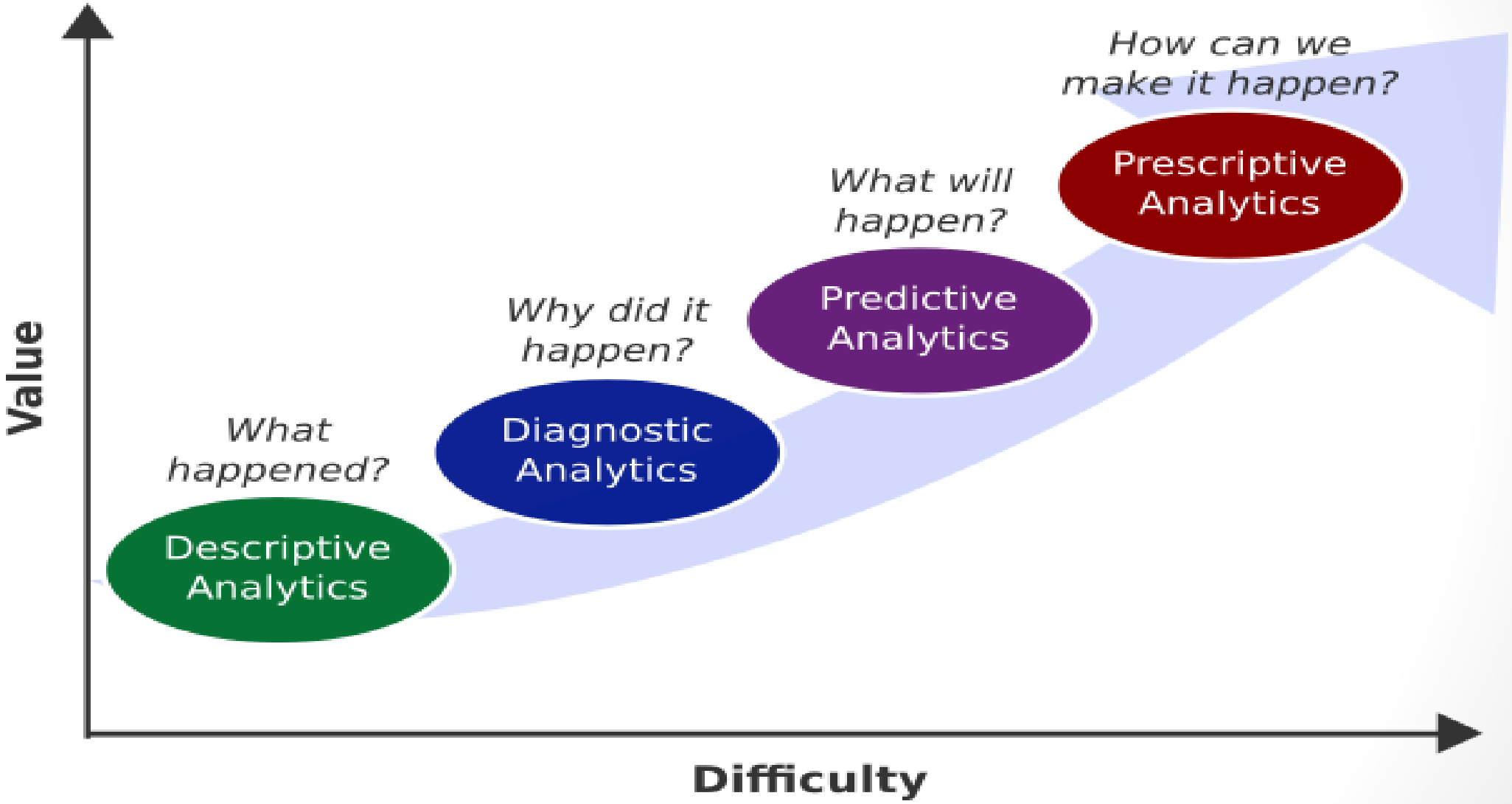
Cauliflower (Mrdium) •
Grade A
SOO
t Ɒ00

B&nce • Robu=uScm1
Ri;>c Gr t A Super) (7
t32.00
ADD,.,

Why Data Analytics?

- Provides ability to make better decisions.
- Better way to solve problems.
- Removes inefficiency in the system/organization.
- Competitive advantage.

Data Analytics Classification



Descriptive Analytics Applications

- Most shoppers turn towards right when they enter a retail store.
- Conversion rate of women shoppers is higher than male shoppers among electronic gadgets purchasers.
- Women car buyers prefer women sales person

Predictive Analytics Problems

- Which product is the customer likely to buy in his next purchase (recommender system).

- Which customer is likely to default in his/her loan payment.

- Who is likely to cancel the product that was ordered through e-commerce portal.

Prescriptive Analytics Problem

- What is the optimal route for a delivery truck.
- Whether a company should introduce a new product?
- What is the optimal product mix?
- How to manage the fleet of vehicles owned by a company for employee drop and pick up?

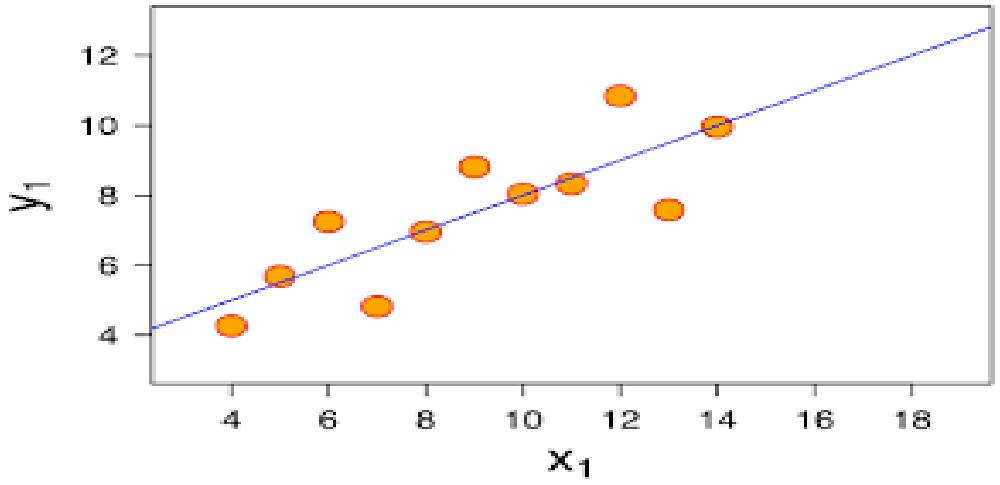
Anscombe's Quartet

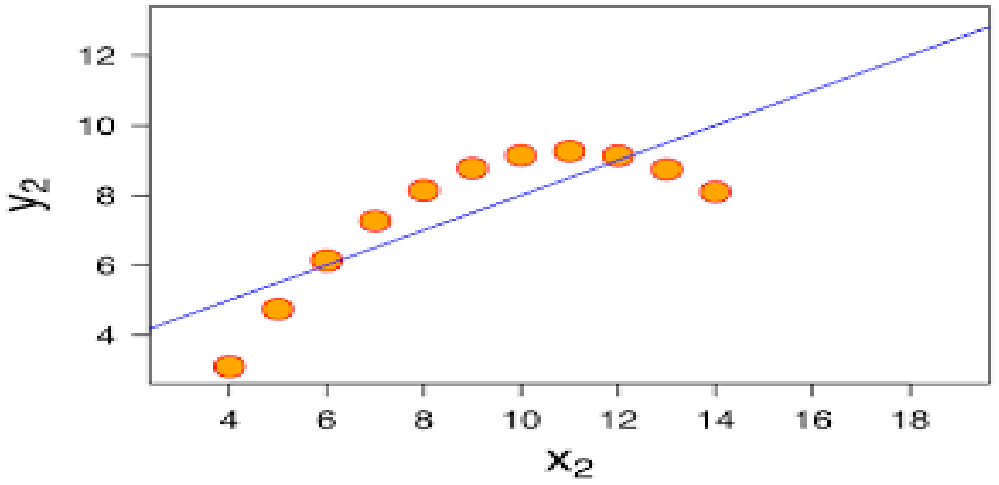
- Four datasets that have nearly identical simple statistical properties, yet appear very different when graphed.
- Each dataset consists of eleven (x,y) points.
- Constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers on statistical properties.

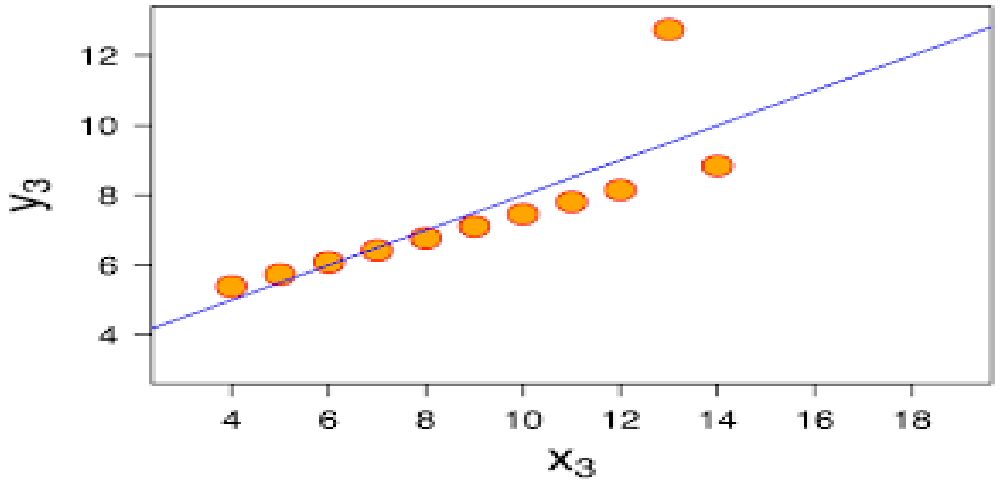
Anscombe's Quartet

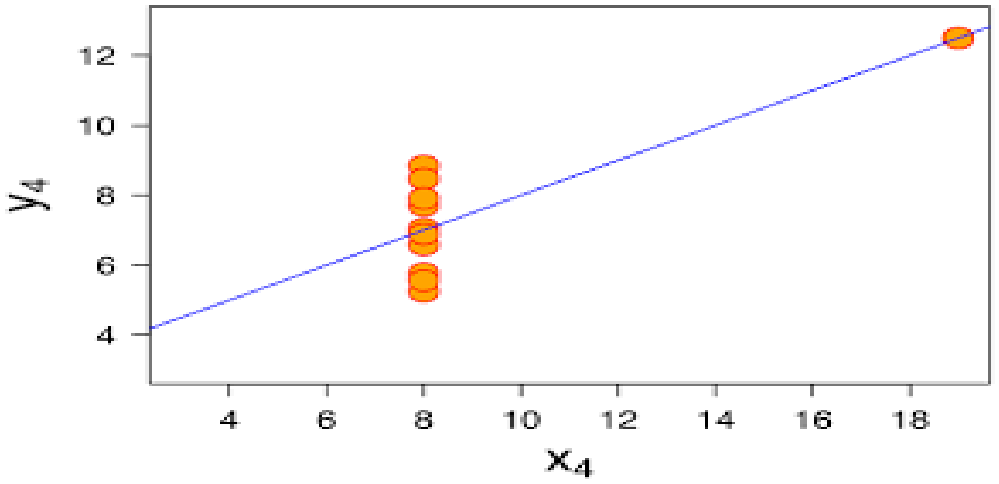
	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

Anscombe's Quartet









Exploratory Data Analysis (EDA)

- An approach to analyzing data sets - **often with visual methods**.
- Leverages our brains' built-in "software" to identify patterns and communicate relationships
- Identify trends and outliers, discover or search for insights
- Seeing what data can tell beyond formal modeling
- Visual analysis aids analytical reasoning
- Inspire new questions & further exploration
- Leverages the incredible capabilities and bandwidth of the visual system

Data Visualization Types/Purpose

Visualisation: Moves information from point A to point B

1. Exploratory Visualization

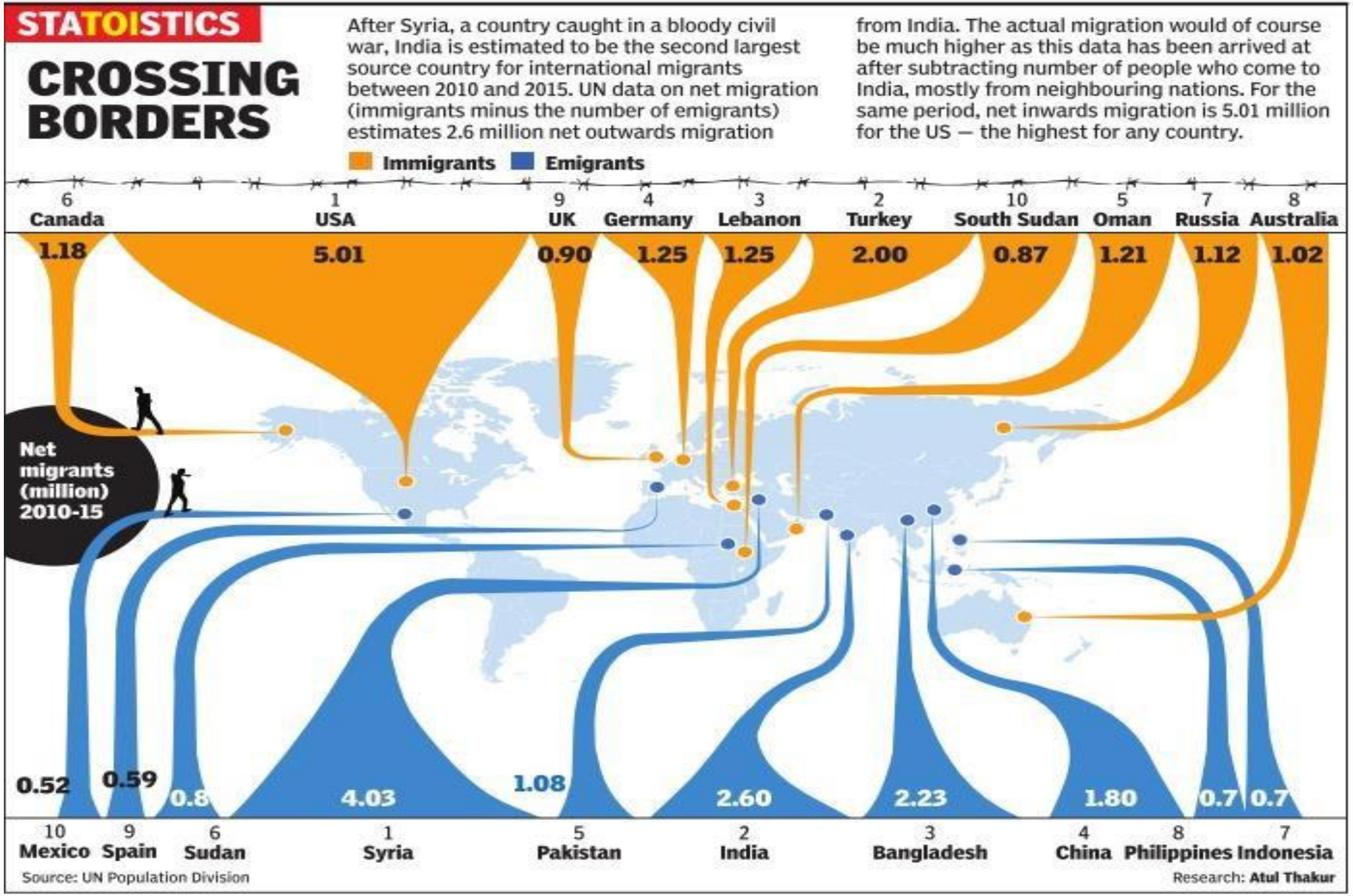
   A: Dataset

   B: Designer's mind

2. Explanatory visualization

   A: Designer's mind

   B: Reader's mind


STATOISTICS
CROSSING BORDERS
After Syria, a country caught in a bloody civil war, India is estimated to be the second largest source country for international migrants between 2010 and 2015. UN data on net migration (immigrants minus the number of emigrants) estimates 2.6 million net outwards migration from India. The actual migration would of course be much higher as this data has been arrived at after subtracting number of people who come to India, mostly from neighbouring nations. For the same period, net inwards migration is 5.01 million for the US — the highest for any country.
Immigrants
Emigrants
6 Canada
1 USA
9 UK
4 Germany
3 Lebanon
2 Turkey
10 South Sudan
5 Oman
7 Russia
8 Australia
1.18
5.01
0.90
1.25
1.25
2.00
0.87
1.21
1.12
1.02
Net migrants (million) 2010-15
0.52
0.59
0.8
4.03
1.08
2.60
2.23
1.80
0.7
0.7
10 Mexico
9 Spain
6 Sudan
1 Syria
5 Pakistan
2 India
3 Bangladesh
4 China
8 Philippines
7 Indonesia
Source: UN Population Division
Research: Atul Thakur

What is Big Data Problem?

Big Data Problems have two major constraints:

1. Technology that can handle the volume of data that is generated at high velocity.
2. *Many problems do not have straight forward algorithms.*

Big Data Problem is "collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications"

Data Mining

- Data mining, in simple terms, is finding useful patterns in the data.
- The art of data mining combines the knowledge of statistics, subject matter expertise, database technologies, and machine learning techniques to extract meaningful and useful information from the data

- **WHAT DATA MINING IS NOT**
  - **Descriptive statistics**
  - **Exploratory visualization**
  - **Dimensional slicing**
  - **Hypothesis testing**
  - **Queries**

What is Artificial Intelligence

Artificial intelligence (**AI**) or **Machine intelligence** (**MI**) is **the capability of a machine to perform tasks that normally require human intellect,** such as visual perception, speech recognition and decision-making.

An **automated system** is one in which a computer reasons by a clear if–then–else, **rule-based structure**, and does so deterministically, meaning that for each input the system output will always be the same (unless if something fails).

An **autonomous system** is one **that reasons probabilistically given a set of inputs**, meaning that it makes **guesses about best possible courses of action given sensor data input,** it also means given different ecosystems the autonomous system will give different results.

AI in Governance

Local law enforcement can use pattern detection to detect anomalous behaviour in individual actors, or to predict dangerous crowd behaviour. Intelligent perception systems can protect critical infrastructure, such as airports and power plant.

AI can play an integral part in shaping the way we generate, transmit and use energy through demand response aggregation that synchronizes energy supply with real-time demand.

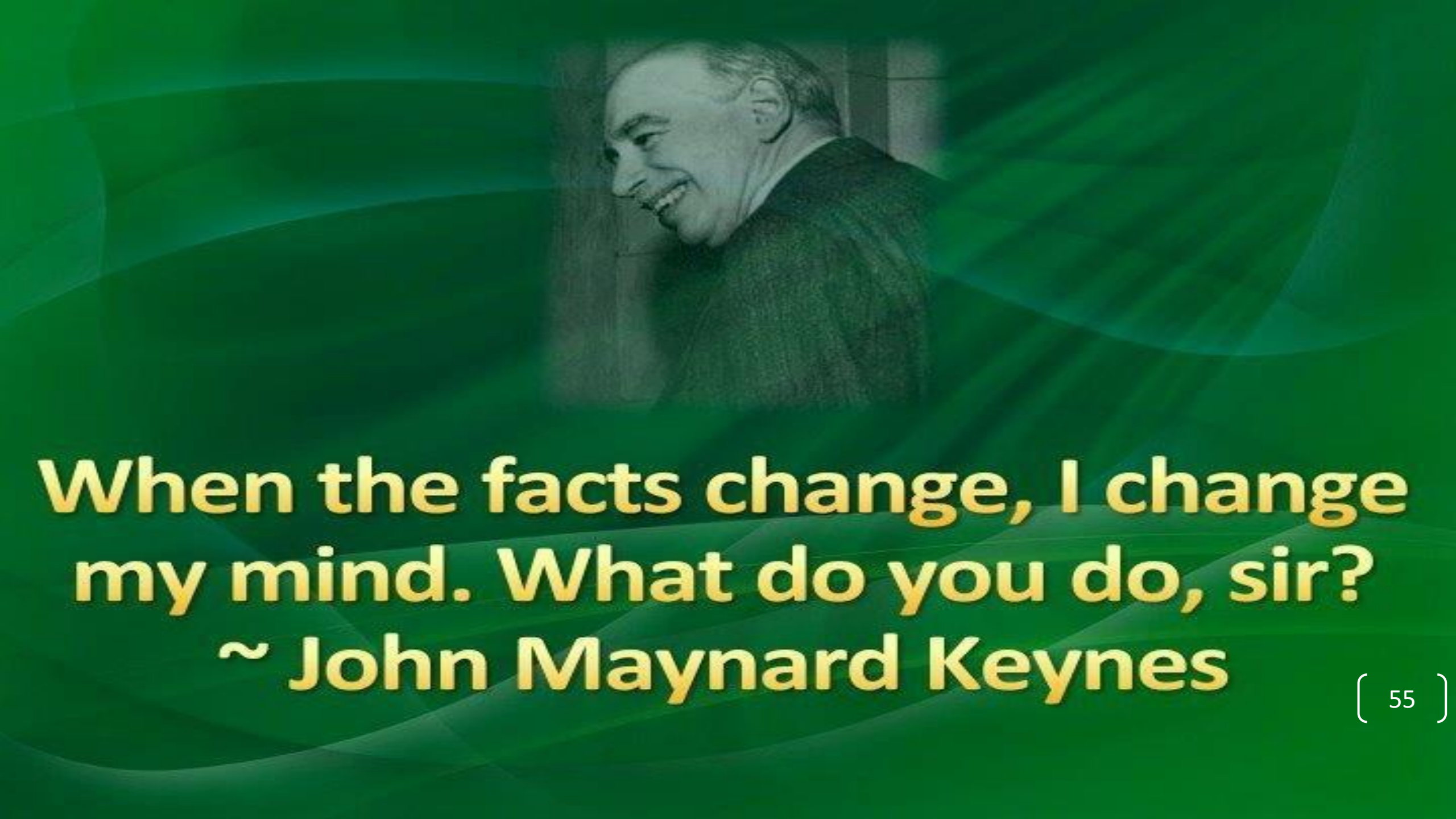
AI can be leveraged in Banking and Financial Services Industry by capturing real time as well as historical time series data for deep learning and predictive analytics to improve stakeholder relationship and mitigate risk.

Key Governance Sectors for AI

Leveraging AI in healthcare: population health insight, to analyze where populations might need the most attention; augmented intelligence to make a function stronger; and precision engagement in which personalization is taken to the next level.

AI in agriculture: (i) agricultural robotics, (ii) soil and crop monitoring, and (iii) predictive analytics using sensors and soil sampling

Primary Education: improve children's knowledge of concepts rather than rote learning. Augmented Reality involves integrating digital information and technology with the user's physical environment to incentivise the learning progress


When the facts change, I change my mind. What do you do, sir?
~ John Maynard Keynes

Thank you

Pravin Srivastava

Ministry of Statistics & Programme Implementation

Email: secretary@mospi.gov.in

http://mospi.gov.in/sites/default/files/film.mp4