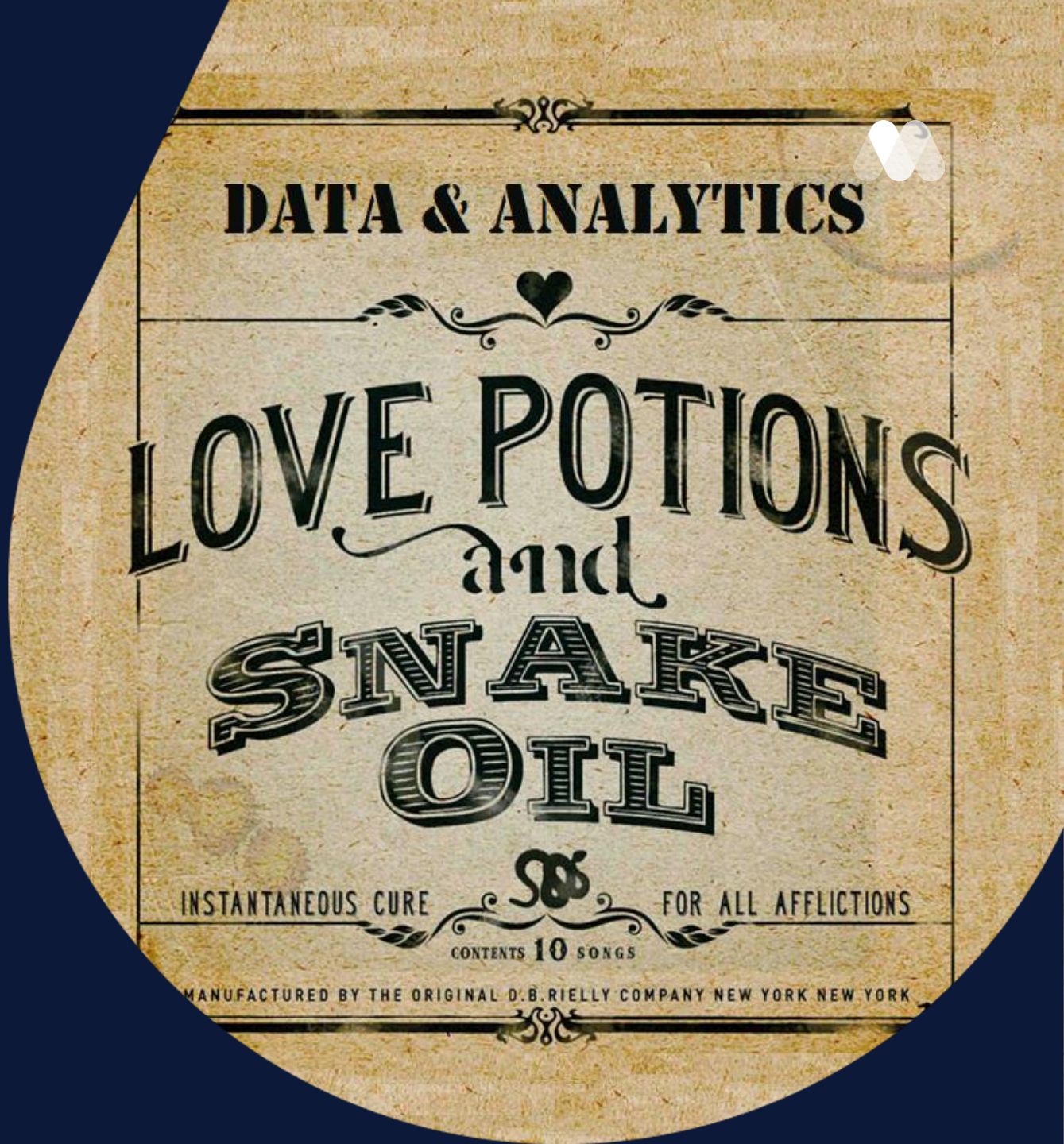
MANGO SOLUTIONS

PSI Webinar

What does the Data-Driven Hype mean for Pharma Analytics?

Richard Pugh
Chief Data Scientist
rich@mango-solutions.com
@richatmango



Agenda

Introductions

Data, hype & snake oil

Never mind the buzzwords

What this means for pharma

Summary

PSI Webinar

Introductions

My background

- Statistician working in pharmaceutical industry
- SAS > S+ > R
- Founded Mango in 2002
- Work across sectors
- Advice on data & analytics

Mango helps companies to deliver data-driven value, create a data-first culture and to build a lasting data science capability

ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

Data Science Section

- Established in 2017
- Representatives from business, industry, government and academia
- Formed to address emerging topics that will impact the long term success of data science as a profession
- Our remit is to be a professional body that represents data scientists in the UK

PSI Webinar

Data, hype & Snake Oil

The visibility and remit of analytics has changed over the last 20 years

Organisations believe that **future success** will depend largely on an **ability to use data** to **make optimal business decisions** and drive efficiencies

A **Data Driven Company** is one that generates value from data by integrating it into the DNA of its decision making processes

Dominic Cummings's Blog

Home | About | An Index of blogs, articles, papers | Education reform 2010-14 | My essay on an 'Ody

JANUARY 2, 2020 BY DOMINICCUMMINGS

'Two hands are a lot' — we're hiring data scientists, project managers, policy experts, assorted weirdos...

'This is possibly the single largest design flaw contributing to the bad Nash equilibrium in which . many governments are stuck. Every individual high-functioning competent person knows they ca make much difference by being one more face in that crowd.' Eliezer Yudkowsky, AI expert, LessWrong etc.

'[M]uch of our intellectual elite who think they have "the solutions" have actually cut themselves c from understanding the basis for much of the most important human progress.' Michael Nielsen physicist and one of the handful of most interesting people I've ever talked to.

'People, ideas, machines — in that order.' Colonel Boyd.

We want to hire an unusual set of people with different skills and backgrounds to work in Downing Street with the best officials, some as spads and perhaps some as officials. If you are already an official and you read this blog and think you fit one of these categories, get in touch.

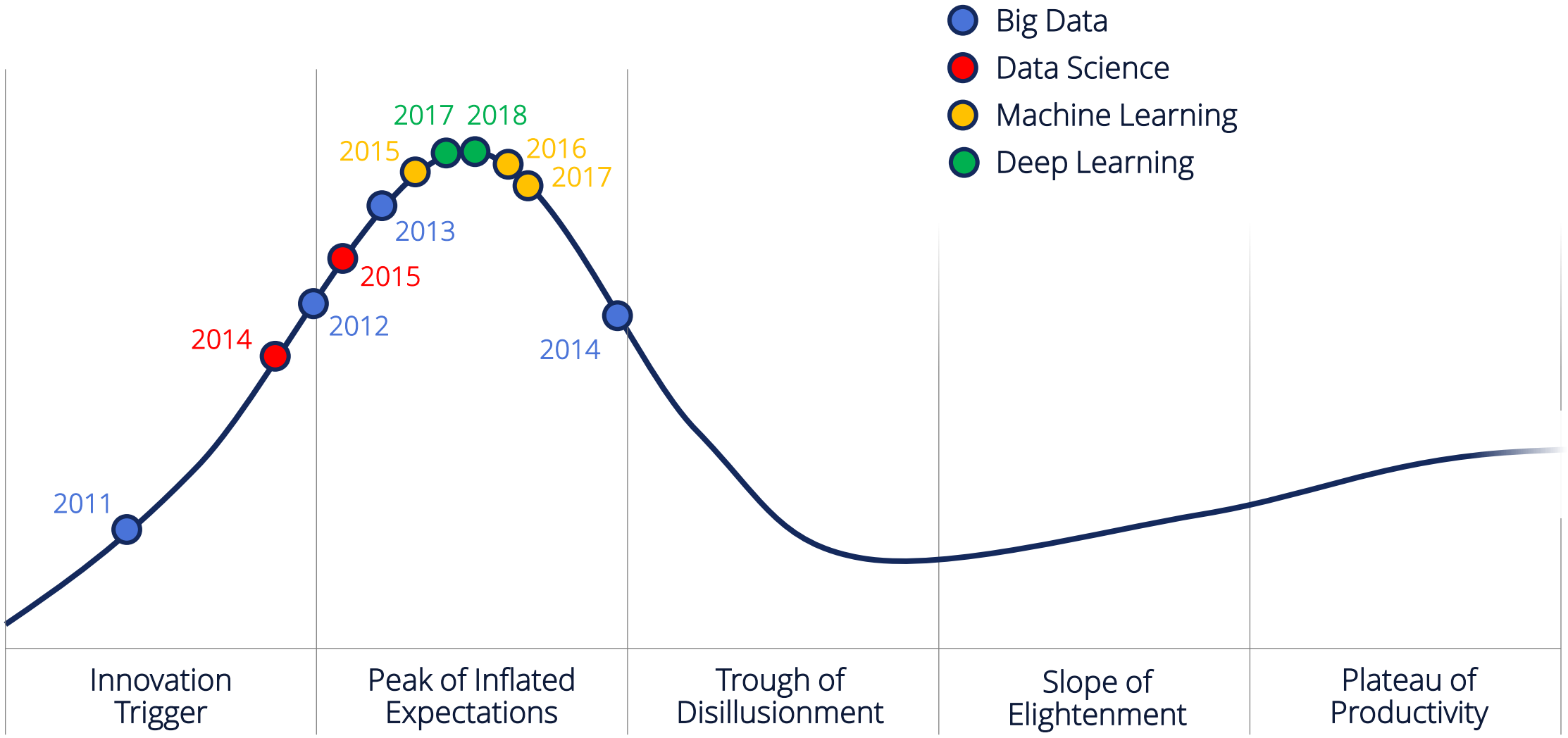
The categories are roughly:

- Data scientists and software developers
- Economists
- Policy experts
- Project managers
- Communication experts
- Junior researchers one of whom will also be my personal assistant
- Weirdos and misfits with odd skills

We want to improve performance and make me much less important — and within a year largely redundant. At the moment I have to make decisions well outside what Charlie Munger calls my 'circle of competence' and we do not have the sort of expertise supporting the PM and ministers that is needed. This must change fast so we can properly serve the public.

So what changed?

The Gartner Hype Curve



Who is driving the hype and why?

Large technology vendors looking to sell highly-priced technical solution

Large consulting firms looking to sell analytic consulting services

Startups looking for funding

The story so far …

- Over the last 20 years, the hype has raised the profile of data & analytics
- Analytics is now a strategic topic for leadership
- The hype focused on a set of “buzzwords” which are now in common use
- What do these “buzzwords” mean, and how does it impact pharmaceutical analysis?

PSI Webinar

Never mind the buzzwords

$$2 > -3 \quad \infty \quad \pm \quad \times \quad 0.999\ldots = 1 \quad \pi \approx 3.14 \quad \div \quad \sqrt{2} \quad 5^2 \quad 1 + 2 \cdot 3 \quad (1 - 2) + 3 \quad 5(2 + 2) \quad 101_2 = 5_{10}$$

AI

Big Data

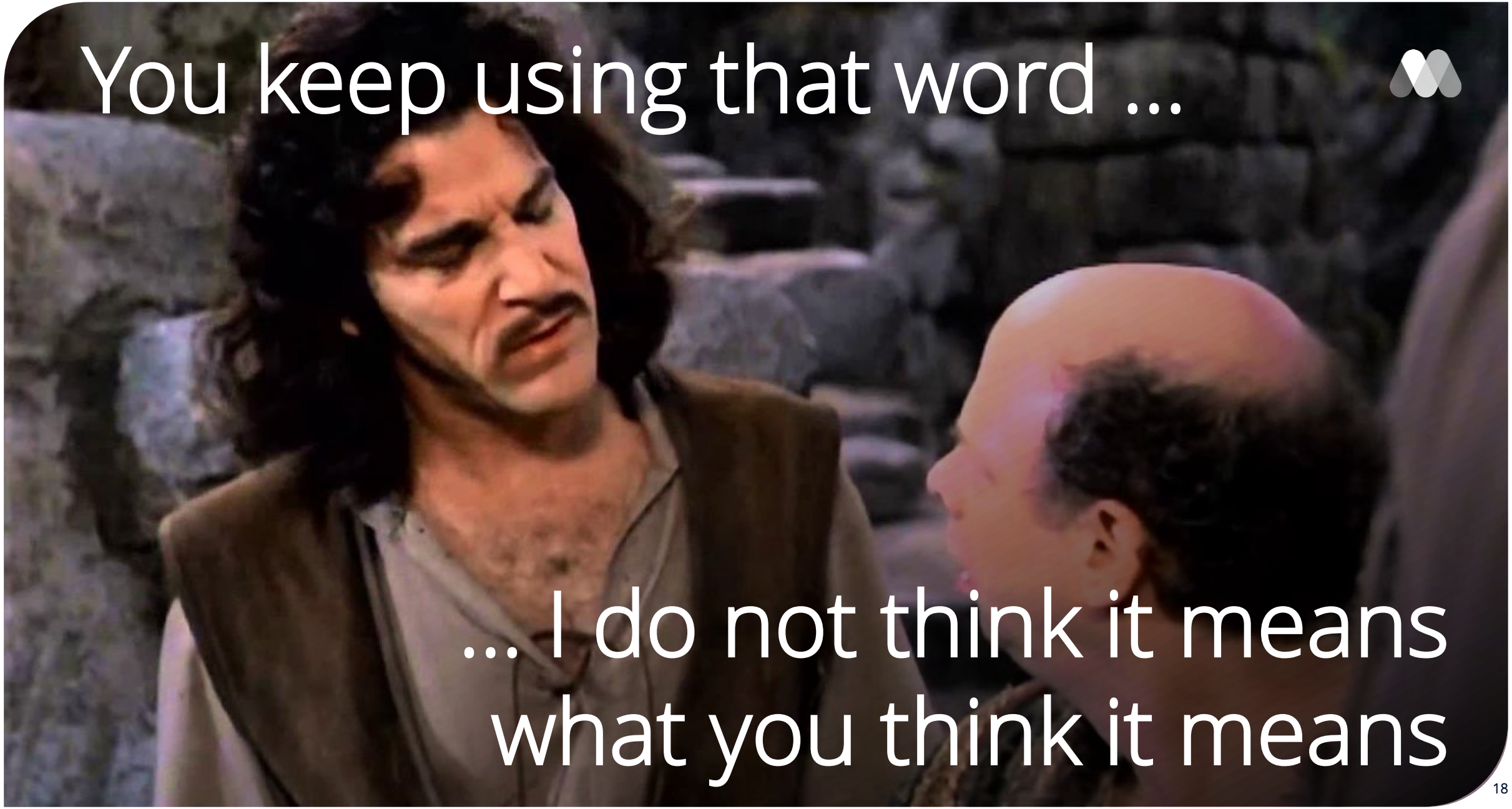
Analytics

Machine Learning

Data Science

Statistics

You keep using that word …



… I do not think it means what you think it means

Data & Analytic Terminology

Impact on the business

Influence decision making or ensure visibility on performance

Analytics

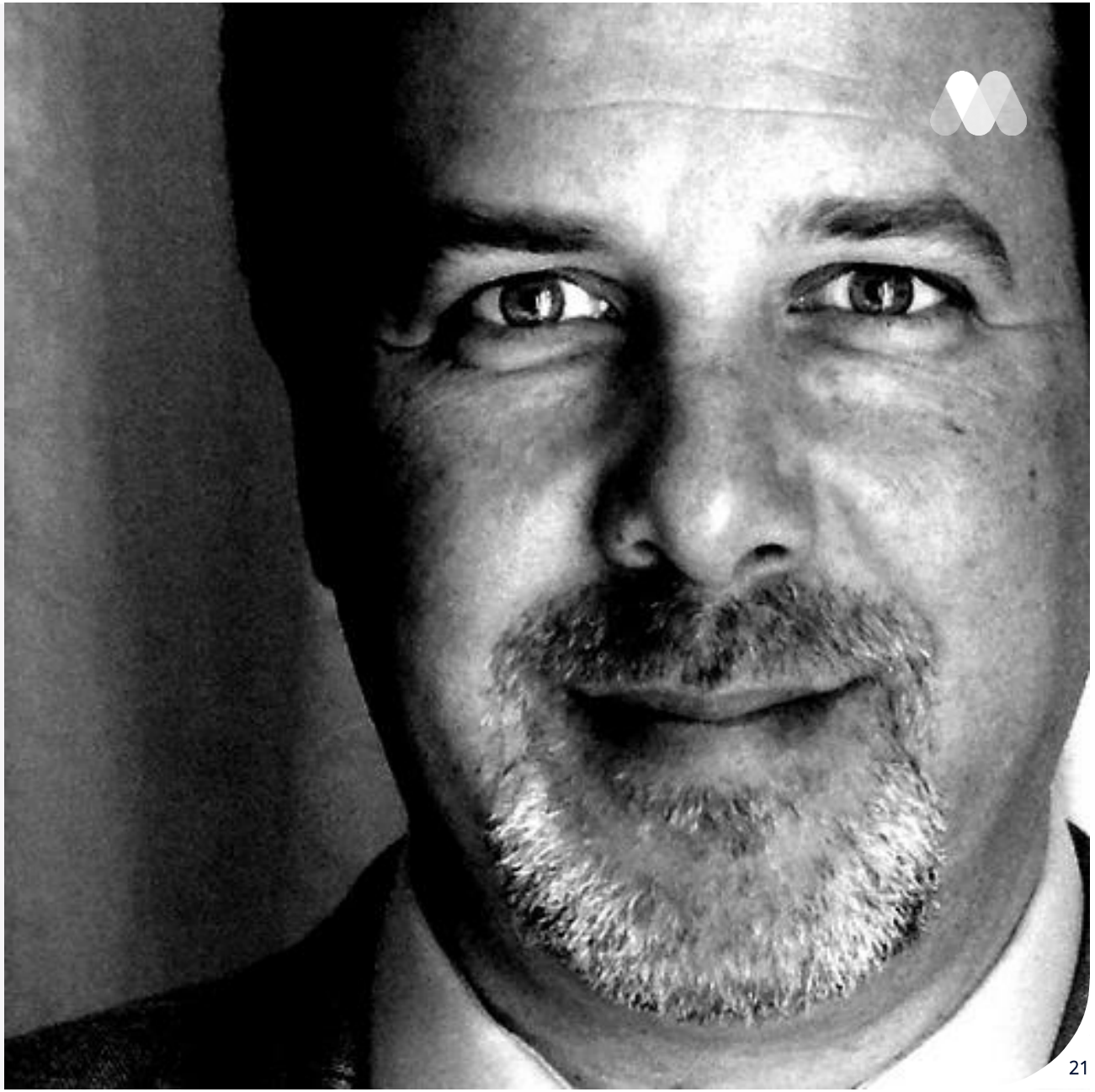
Turn data into knowledge to be communicated to the business

Data

Information captured from internal or external sources, stored and managed on technical platforms

Data Dimensions

- Categorised as the "3 (or 4) Vs":
  - Volume – Data Size
  - Velocity – Real time decisions
  - Variety – Structure, Unstructured
  - *Veracity – Data Quality*



Hadoop

- Technology solution to scaling data storage
- Considered “obsolete” on the 2017 Hype cycle
- Spawned set of technologies to manage data across the “V” dimensions

Data & Analytic Terminology

Impact on the business

Influence decision making or ensure visibility on performance

Analytics

Turn data into knowledge to be communicated to the business

Data

Information captured from internal or external sources, stored and managed on technical platforms

Data & Analytic Terminology

Impact on the business
Influence decision making or ensure visibility on performance

Analytics
Turn data into knowledge to be communicated to the business

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

Data & Analytic Terminology

Impact on the business

Influence decision making or ensure visibility on performance

Descriptive Analytics

The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Diagnostic Analytics

The use of modelling to understand the factors that influence a particular outcome, based on historical data

Predictive Analytics

The use predictive modelling approaches to understand likely outcomes from a process under new circumstances

Prescriptive Analytics

The use of predictive modelling and optimisation used to understand how to optimise future outcomes

Big Data

A set of techniques that enable access to structured and unstructured data of varying size

Structured Data

Structured data in rectangular formats (tables, databases)

Unstructured Data

Non-standard data forms (image, text, video, sound, streams)

Data & Analytic Terminology

Business Intelligence
Delivering management information via reports and dashboards

Impact on the business
Influence decision making

Advanced Analytics
The application of data and math to model real-world processes and likely future outcomes

Descriptive Analytics
The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Diagnostic Analytics
The use of modelling to understand the factors that influence a particular outcome, based on historical data

Predictive Analytics
The use predictive modelling approaches to understand likely outcomes from a process under new circumstances

Prescriptive Analytics
The use of predictive modelling and optimisation used to understand how to optimise future outcomes

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

Data & Analytic Terminology

Business Intelligence
Delivering management information via reports and dashboards

Impact on the business
Influence decision making

Advanced Analytics
The application of data and math to model real-world processes and likely future outcomes

Descriptive Analytics
The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

Data & Analytic Terminology

Business Intelligence
Delivering management information via reports and dashboards

Impact on the business
Influence decision making

Advanced Analytics
The application of data and math to model real-world processes and likely future outcomes

Descriptive Analytics
The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Statistical Modelling
Human-centric process of building models through data understanding using iterative mathematical approaches

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

Data & Analytic Terminology

Business Intelligence
Delivering management information via reports and dashboards

Impact on the business
Influence decision making

Advanced Analytics
The application of data and math to model real-world processes and likely future outcomes

Descriptive Analytics
The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Statistical Modelling
Human-centric process of building models through data understanding using iterative mathematical approaches

Machine Learning
Compute-centric creation of mathematical models by learning about patterns in data and how they influence outcomes

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

Data & Analytic Terminology

Business Intelligence
Delivering management information via reports and dashboards

Impact on the business
Influence decision making

Advanced Analytics
The application of data and math to model real-world processes and likely future outcomes

Descriptive Analytics
The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Statistical Modelling
Human-centric process of building models through data understanding using iterative mathematical approaches

Machine Learning
Compute-centric creation of mathematical models by learning about patterns in data and how they influence outcomes

Artificial Intelligence
A narrow set of machine learning algorithms that appear to have "human-like" qualities, typically applied to unstructured data

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

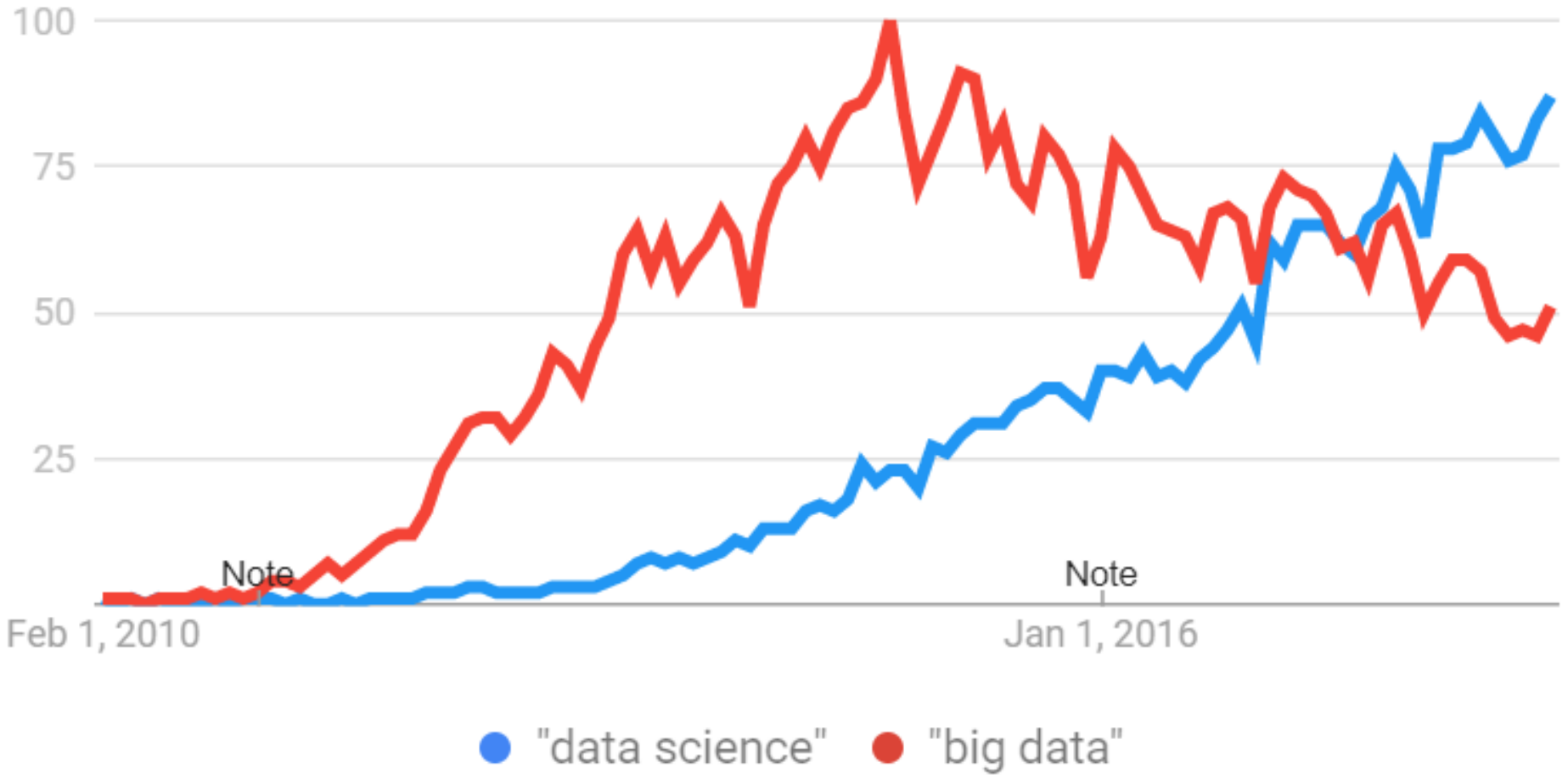
Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

Data Science

- Term coined in 1997
- Professor Jeff Wu suggested it as an alternative to “statistics”

The growth of data science

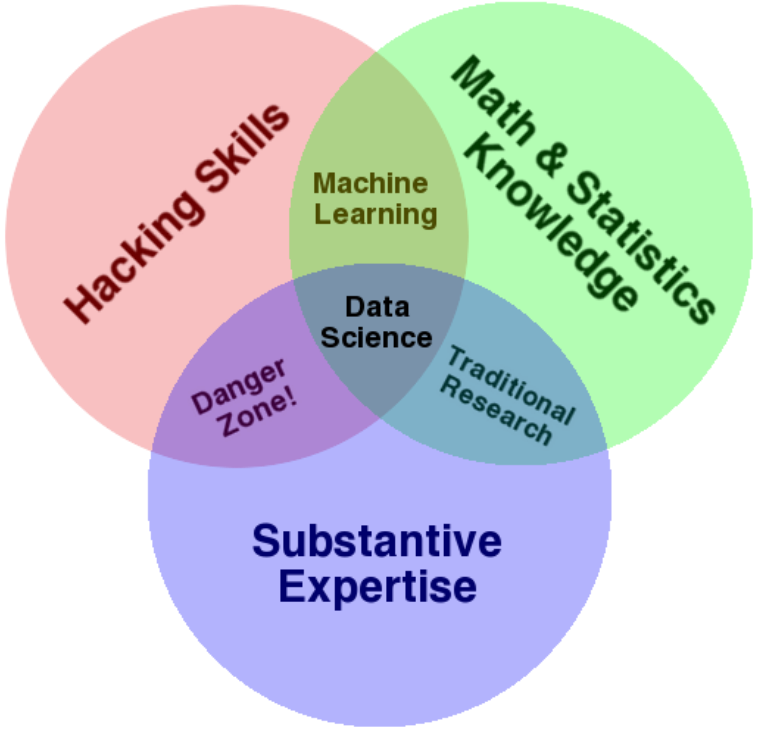


Definitions of data science

Josh Wills
@josh_wills

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

Drew Conway
@drewconway



What is Data Science?

The **proactive** use of **data** and **advanced analytics** to drive better **decision making**

Data & Analytic Terminology

Business Intelligence
Delivering management information via reports and dashboards

Data Science
The proactive application of **data** and **advanced analytics** to drive better decision making

Advanced Analytics
The application of data and math to model real-world processes and likely future outcomes

Descriptive Analytics
The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Statistical Modelling
Human-centric process of building models through data understanding using iterative mathematical approaches

Machine Learning
Compute-centric creation of mathematical models by learning about patterns in data and how they influence outcomes

Artificial Intelligence
A narrow set of machine learning algorithms that appear to have "human-like" qualities, typically applied to unstructured data

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

Warning: These definitions are badly misused and can vary greatly!

PSI Webinar

What this means for pharma?

The pharma industry is, to some extent, already **data-driven**

However, the data science “movement” is broadening the **remit** of analytics, which is impacting the role of **analytics** and its **practitioners**

There are 5 ways in which Data Science is impacting pharma statisticians

REMIT

The interest in data science and AI is broadening the remit of analytic teams, opening up a wider range of challenges and moving the insight closer to the decision

ROLE

The changing remit of analytics can impact the role of practitioners as we're increasingly asked to engage with the business and explore new opportunities

METHODS

A new focus on analytics enables practitioners to look at a broader range of analytic techniques to solve problems, including those based on machine learning approaches

DATA

Big data technologies are allowing a wider range of data sources to be collected and analysed, resulting in the use of new analytic approaches

TECH

All of this is having an impact on both the technology we employ, and the way in which we use technology in a more "DevOps" style to create repeatable insight generation

The broadening of the analytic remit is impacting our role and approach

REMIT

- Increased visibility of data and analytics to support decision making
- Broader range of challenges to which analytics can be applied
- Increasing use of prescriptive analysis to drive decisions
- More openness to try new approaches and innovate

Novartis is a data company, and I think data science is going to allow us unlock even more insights across every element of our business

I'm incredible excited about the power of these technologies to enable us to do even more as a company

Vas Narasimhan, CEO

NOVARTIS

The broadening of the analytic remit is impacting our role and approach

REMIT

- Quantitative Decision Making (QDM) framework built into development process
- Framework describes the quantitative characteristics of the proposed study design
- QDM helps understand probability of trial success
- QDM rolled out to all clinicians and clinical statisticians
- Data science tools built to facilitate and standardise process for analysts

Across our R&D, we are using AI to help us decipher a wealth of information with the aim of gaining a better understanding of the diseases we want to treat; identifying new targets for novel medicines; recruiting for and designing better clinical trials; driving personalised medicine strategies and speeding up the way we design, develop and make new drugs

Jim Weatherall
VP, Data Science & AI, R&D

The nature of our roles is evolving as the remit of analytics changes

ROLE

- More emphasis on soft skills needed to engage with broader business
- Focus on modern techniques and technologies
- Curiosity to explore new approaches and answer new questions

The nature of our roles is evolving as the remit of analytics changes

ROLE

The Data Scientist is **accountable for driving projects** and improvement activities through **technical consulting** and **value delivery leadership** …

Advanced communications skills are critical for this role, with particular strengths required in distilling and **communicating complex concepts** …

… new and emerging data science technologies such as **Machine/Deep Learning** and **Artificial Intelligence** …

… will have a **passion for discovering solutions** hidden in large data sets and working with stakeholders to **improve business outcomes** …

New methodologies can be applied, initially in non-clinical space

METHODS

- Pharma analysis is primarily focused on (largely frequentist) statistical methods
- The use of machine learning and other techniques is becoming widespread, mostly in non-clinical areas
- The use of Bayesian methods in Clinical Statistics gives a good indication of the pace of change we could expect

New methodologies can be applied, initially in non-clinical space

METHODS

ML methods used for subgroup analysis in clinical trials in respiratory

AstraZeneca using DS and AI to help them recruit for, and design, better clinical trials

Supervised learning techniques used in cheminformatics, high throughput screening

Many examples of deep learning applied to images to create new endpoints for study

Transfer learning to analyze molecular and imaging libraries as well as patient datasets to uncover complex biomarker patterns

NLP and Deep Learning used to automate the classification and extraction of information from medical papers

https://blog.benchsci.com/pharma-companies-using-artificial-intelligence-in-drug-discovery

DATA

Big data technologies enable analyse of wider variety of data formats

- Big data tech allows us to collect, store and manage new data sources
- Unstructured data sources such as text, image and video, as well as more traditional (but large) data sources such as genomics data
- Also supports analysis of data streams from devices (e.g. wearables, or devices that trigger alerts)

Example Uses of Wearables

- Asthma devices with IoT to monitor and analyse correct dosing
- Sensors used to supplement, or replace, pain endpoints
- Devices to monitor patient health
- Devices to trigger alerts on patient falls within healthcare facilities
- IoT used to trigger alerts for patients in the home

Demands on analysis changing our use of technology

TECH

As technology evolves, and the remit of analysis broadens, there is more demand to leverage innovations improve drug development

Push towards tech that offers modern capabilities, flexibility and scale (R, Py, Spark)

Innovation around access to intelligence - advanced reporting, lightweight apps, interactivity

This changes the technology we use and relationship with it

Use of DevOps approaches becoming more common place where code is a primary output

Collaborative working in teams around version control for efficient creation of common IP

Demands on analysis changing our use of technology

TECH

Creating applications to allow rich review of results

Many initiatives looking at developing apps in technologies such as “Shiny” to provide richer experience for data review

This includes initiatives to build applications to support FDA review of Tables, Figures and Listings

Reproducible research impacting reporting flows

Programming approaches managing to separate components of reports to speed turnaround time to results from database lock

Approaches based on technologies such as R and Markdown, with concepts based on reproducible research (e.g. parameterised publication of standard reports)

PSI Webinar

Summary

Over the last 20 years, the hype has raised the profile of, and expectations on, data & analytics

The hype focused on a set of “buzzwords” which are now in common use

Data & Analytic Terminology

Business Intelligence
Delivering management information via reports and dashboards

Data Science
The proactive application of **data** and **advanced analytics** to drive better decision making

Advanced Analytics
The application of data and math to model real-world processes and likely future outcomes

Descriptive Analytics
The use of reporting and basic analysis to summarise historical data using charts, tables and statistics, with results often displayed in dashboards and reports

Statistical Modelling
Human-centric process of building models through data understanding using iterative mathematical approaches

Machine Learning
Compute-centric creation of mathematical models by learning about patterns in data and how they influence outcomes

Artificial Intelligence
A narrow set of machine learning algorithms that appear to have "human-like" qualities, typically applied to unstructured data

Big Data
A set of techniques that enable access to structured and unstructured data of varying size

Structured Data
Structured data in rectangular formats (tables, databases)

Unstructured Data
Non-standard data forms (image, text, video, sound, streams)

There are 5 ways in which Data Science is impacting pharma statisticians

REMIT

The interest in data science and AI is broadening the remit of analytic teams, opening up a wider range of challenges and moving the insight closer to the decision

ROLE

The changing remit of analytics can impact the role of practitioners as we're increasingly asked to engage with the business and explore new opportunities

METHODS

A new focus on analytics enables practitioners to look at a broader range of analytic techniques to solve problems, including those based on machine learning approaches

DATA

Big data technologies are allowing a wider range of data sources to be collected and analysed, resulting in the use of new analytic approaches

TECH

All of this is having an impact on both the technology we employ, and the way in which we use technology in a more "DevOps" style to create repeatable insight generation

Enabling your data-driven journey
Helping you to thrive on data science

Richard Pugh
Chief Data Scientist
rich@mango-solutions.com
@richatmango
linkedin.com/in/richatmango/

RSS Data Science Section
groups.io/g/datasciencesection
linkedin.com/company/rssdatascience