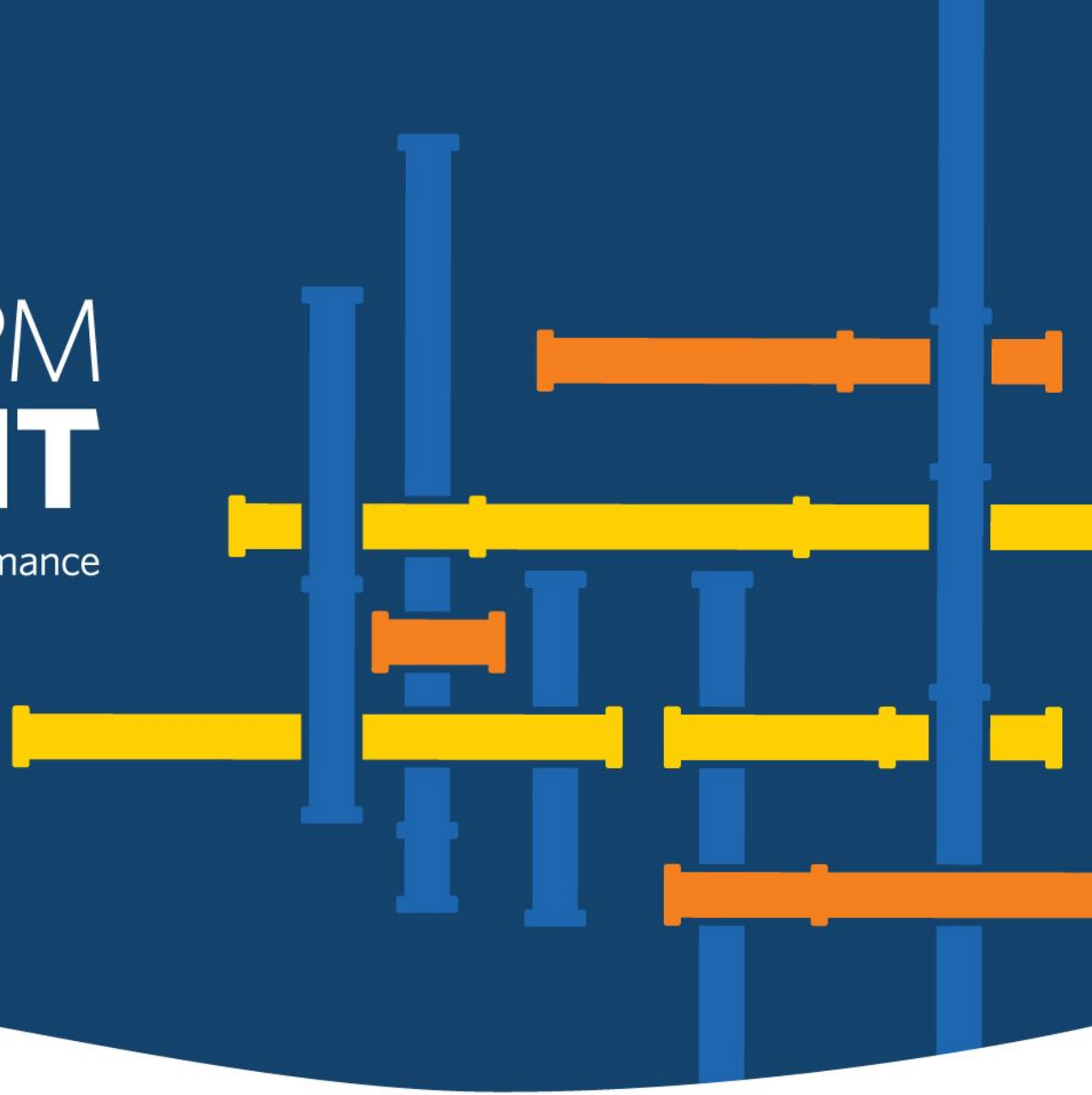
2020 AFPM SUMMIT

Excellence in Plant Performance



AFPM

Storytelling Using Data Science

Summit - 2020

galvanize

- Computational Biologist
- Industry, Academia, National Labs
- Healthcare, Genomics
- Built AI Platforms
- Educator, Data Scientist

Adam Richards, PhD
Head of Enterprise Data Science
Galvanize, Inc
adam.richards@galvanize.com

Currently leading a diverse team of data scientists and data engineers. We help build and enable the leading industry data talent.

I consider one of the most important duties of any scientist the teaching of science to students and to the general public. ---Isaac Asimov *The Tragedy of the Moon* (1973), p. 224

2020 AFPM SUMMIT
Excellence in Plant Performance

Agenda

1 ANALYTICS, AI AND THE WHY

Data science has come so far recently, but why?

2 DEMYSTIFYING MACHINE LEARNING

It is in the examples that we see the business value.

DEMO: Predicting equipment failure

3 DATA-DRIVEN DECISION MAKING AND TECHNICAL COMMUNICATION

The key is being able to decision making is to visualize the stories

DEMO: Change point analysis

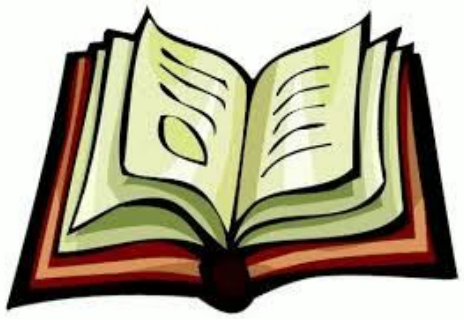
Analytics, AI and the Why

Elements of a good story

1. **Characters** → end-users, customers, stakeholders, data professionals
2. **Settings** → backstory, business scenario, data
3. **Plot** → central theme, business opportunity

How can a data product address a given business opportunity?

4. **Conflict** → What are the obstacles?
5. **Resolution** → What is the plan? What benefit or ROI can we expect?



A data analyst in 2020

See Kaggles 2019 DS Survey

20 years ago..

- Excel, VBA
- User interfaces
- Proprietary tools
- SQL
- SAS, Stata, Matlab
- Installable Applications

Today

- Test-Driven Development (TDD)
- Version control (Git)
- Open-source tools
- SQL and noSQL (MongoDB)
- Python, R
- Docker Containers

A modern analyst should be driven by the ***scientific method*** **and** ***reproducibility.*** They should also possess ***strong communication skills*** that help align their work with ***business needs.***

Now called a **data translator**

Basic programming has been democratized...

Data Science by the numbers

This McKinsey report estimates that AI will have a 1.2-2 Trillion dollar impact on the Supply chain management and manufacturing sector. In order of importance….

1. Predictive Maintenance
2. Yield, Energy, Throughput
3. Procurement, Spend Analytics
4. Inventory and Parts Optimization
5. Logistics Network and Warehouse Optimization
6. Sales and Demand Forecasting

Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI will transform in the next several years. --Andrew Ng

Demystifying Machine Learning

The essential terminology

Artificial Intelligence (AI) - Machines able to perform tasks that normally require human intelligence

Machine Learning (ML) - Subset of AI where machines **learn patterns given inputs**.

- **Supervised learning** - learn a mapping from inputs x to outputs y
- **Unsupervised learning** - given only x, learn interesting patterns in x

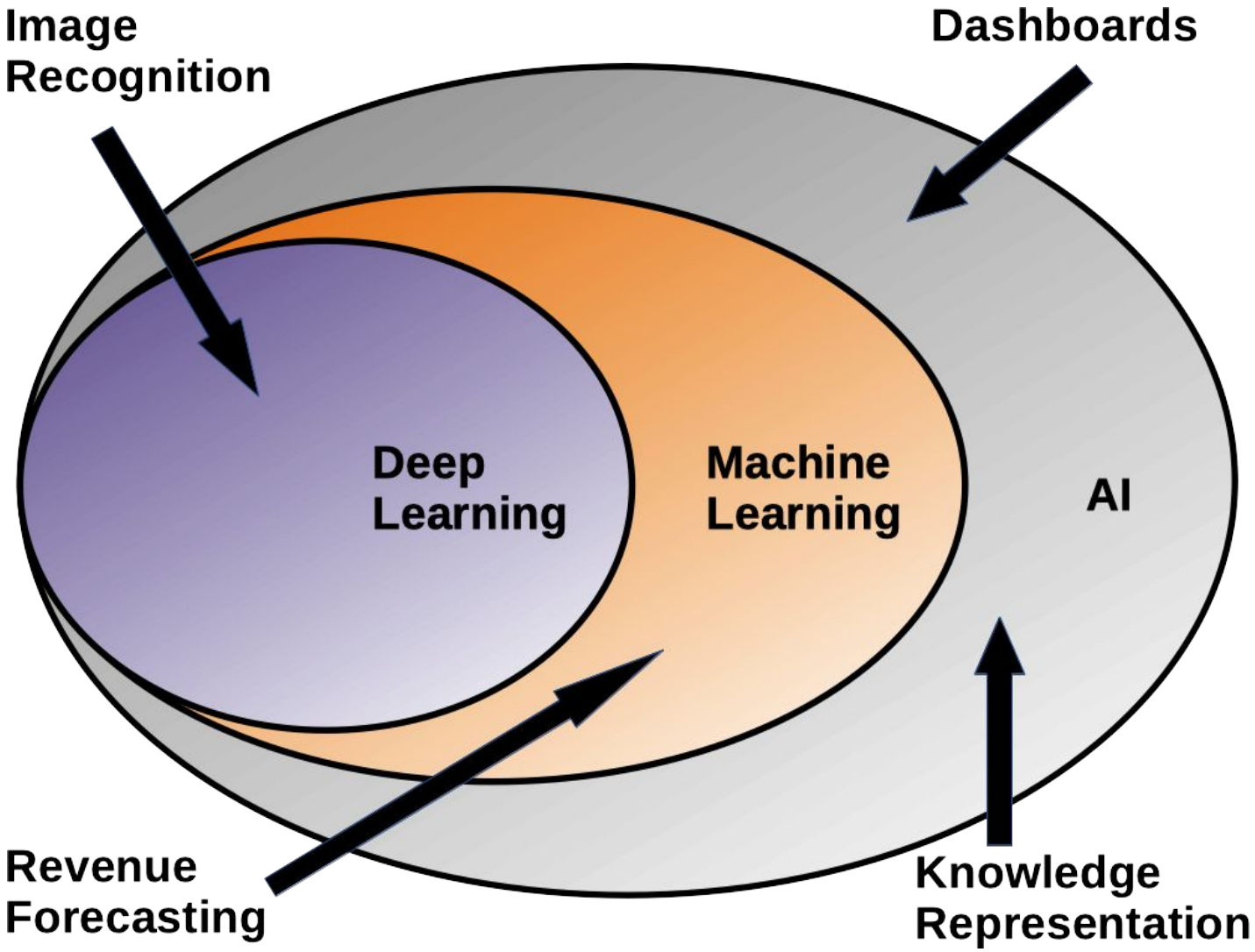
Neural network - A sub-field of ML that, originally was inspired by neurons in the brain, that leverages graphs or networks.

- **Deep learning** - Networks like neural networks and Boltzmann Machines that have at least 3 hidden layers. Image recognition, speech recognition, autonomous driving

Big data - Any data set with enough observations or features that a single machine struggles

The essential terminology (in a picture)

- Deep-learning is mostly the use of neural networks
- Machine learning is mostly comprised of supervised-learning and unsupervised-learning
- AI is used to describe any of these systems



The plot thickens

Business Opportunity

New

Improved

Automated

1. How can I do more with less?
2. If it takes less than one second for a human to come to a decision then good candidate for a data product..

- Tools to help make decisions
  - Prioritizing safety reports for human review
  - Staffing optimization algorithms
- Marketing
  - Churn detection system with automated reporting
  - Identifying ideal customer profile
- A/B testing
  - Refining product recommendations
  - Comparing product strategies that affect revenue/yield

Examples

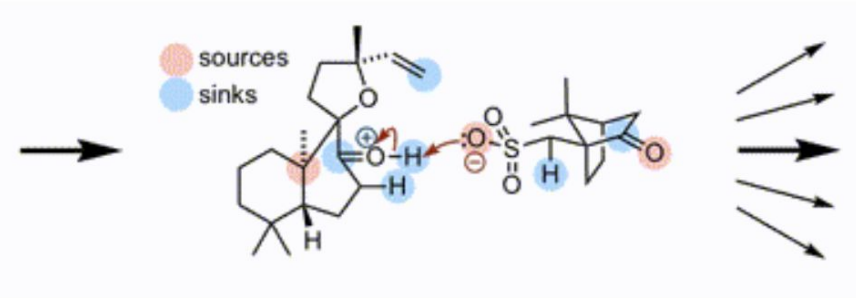
- Sensational examples of data science are only part of how it is applied…
- There is a lot of untapped potential for most organizations through process improvement and automation
- … AI can optimize routing of delivery traffic, thereby improving fuel efficiency

Paper

Deep learning for chemical reaction prediction

David Fooshee, Aaron Mood, Eugene Gutman, Mohammadamin Tavakoli, Gregor Urban, Frances Liu, Nancy Huynh, David Van Vranken and Pierre Baldi

We describe a deep learning-based system for predicting chemical reactions and identifying experimentally-observed masses.



From the themed collection: MSDE most-read Q1 2019

The article was first published on 01 Dec 2017
Mol. Syst. Des. Eng., 2018, **3**, 442-452
https://doi.org/10.1039/C7ME00107J

Paper collection: Machine Learning and Data Science in Materials Design

Measuring value from data projects

- Data Analyst
- Data Translator
- Data Scientist
- Data Engineer
- ML Engineer
- ML Researcher

- Jupyter Notebooks
- Flask/Shiny Apps
- Reports, Dashboards
- Docker
- Packages
- APIs

Business Opportunities

1. **New Opportunities**
2. **Improved or Optimized**
3. **Automated**

Measuring value from data projects

Data Maturity

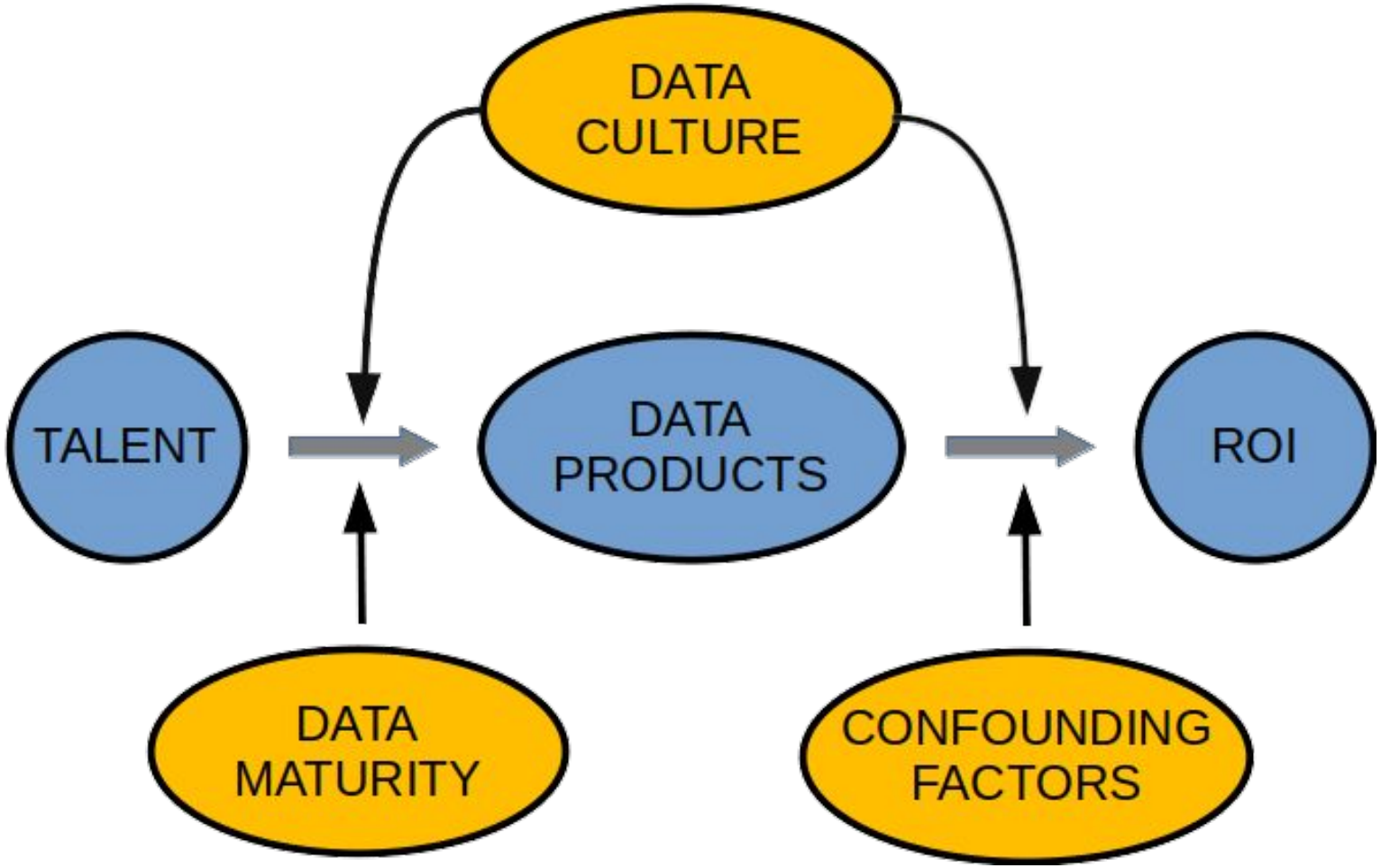
- Centralized Data Lake
- API's, Data Streams

Data Culture

- Infrastructure, Team Org.
- Business Responsiveness

Confounding Factors

- Software releases, bugs
- Marketing, Seasonal trends



Demo - Predicting Equipment Failure

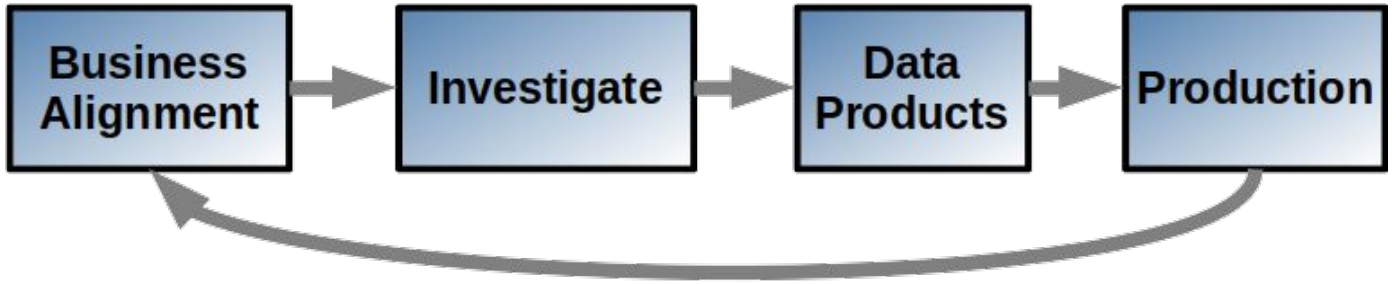
Predictive Maintenance

- **Related concepts**
  - Time to failure
  - Remaining useful life (RUL)
- **Using classical *survival analysis* techniques**
  - Distribution of the time to an event
  - Cox's proportional hazards model
  - Interpretability, Insight into the models
- **Using Bayesian treatment of survival methods**
  - All the benefits of classical methods
  - Highly customizable
- Using **machine learning**
  - Superior to other methods in terms of predictive performance
  - Not always easy to glean insights from the models themselves

Data-driven Decision Making and Technical Communication

Data-Driven Organizations - Process

1. They use a **formalized process** and have a strong data culture



Data Translator

Data Scientist

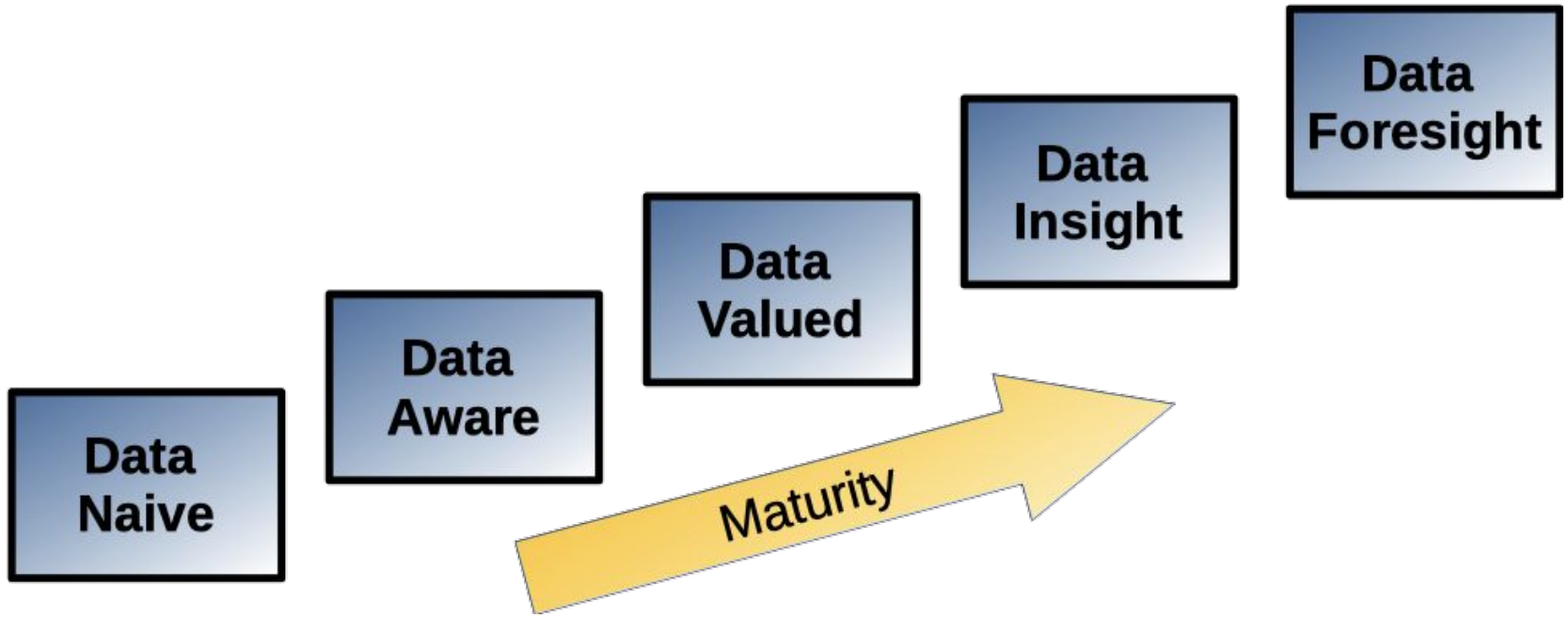
Data Engineer

Examples include:

- CRISP-DM
- OSEMN
- Design Thinking

Data-Driven Organizations - Strategy

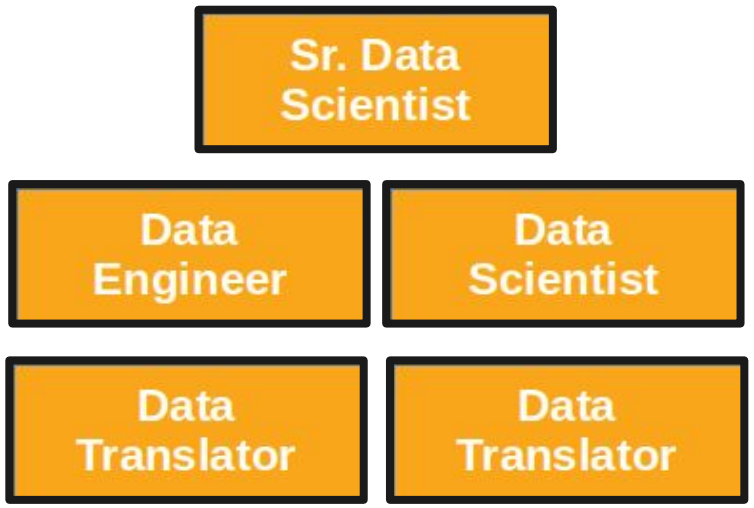
2. They are past the point of a **data-valued** organization



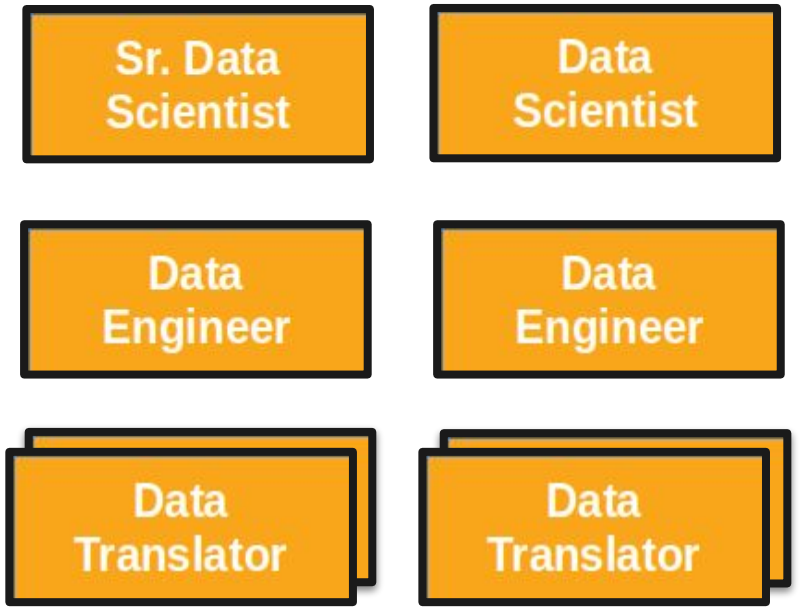
Data-Driven Organizations - Talent

3. They have the right talent

Small Teams



Large Teams



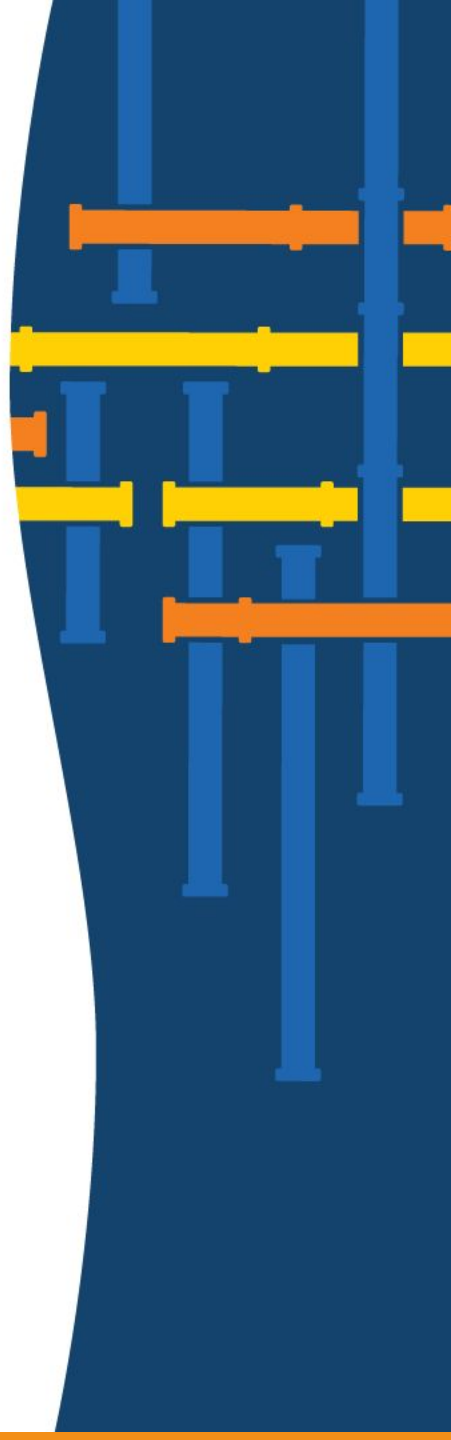
Storytelling and data visualization

- Very few people remember statistics

> According to the McKinsey Global Institute data-driven organizations are 23 times more likely to acquire customers, 6 times as likely to retain customers, and 19 times as likely to be profitable.

- But it is likely that the ability to tell and consume stories has been selected for positively during human evolution.

> How many details can you recall from *Goldilocks and the Three Bears*, The Hobbit, or your favorite movie?



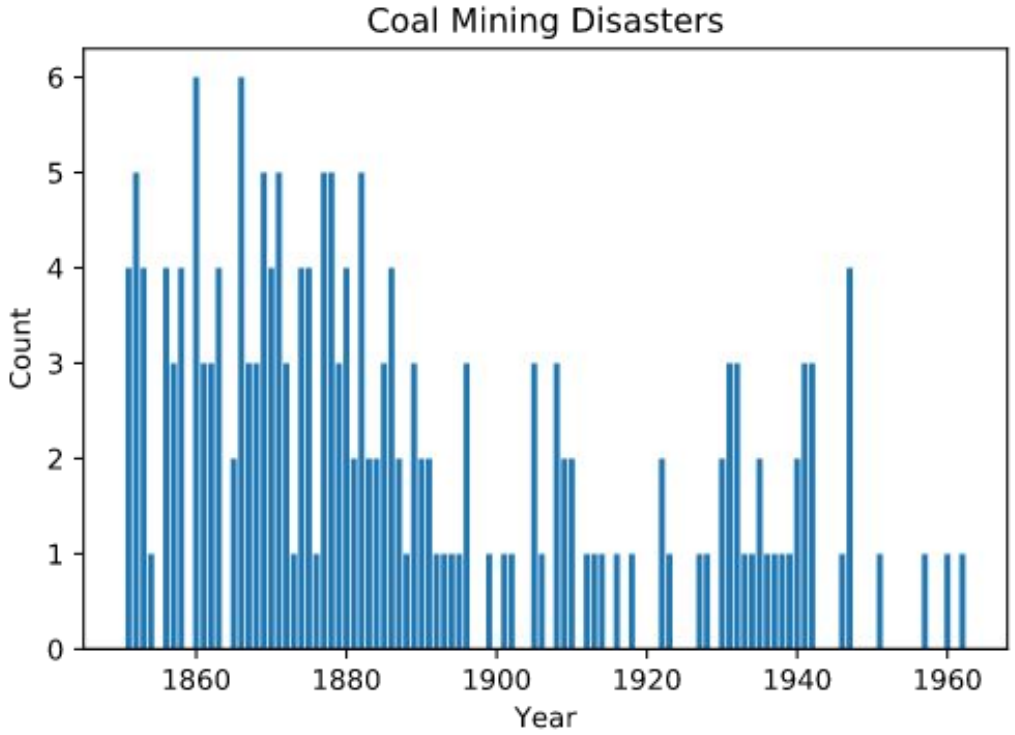
Storytelling and data visualization

- The business opportunity is the and the central hypotheses guide the stories.
- The days of Homer and Brothers Grimm (even slideshows) are long past.
- A good story **interacts with the audience** (interactive code).
- Data visualization is the single most effective prop to help us spin our tales

The best stories are told by professionals with BOTH domain expertise and modern flexible tools.

Demo - Change Point Analysis

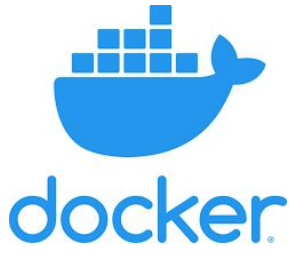
- Changepoint or switchpoint analysis --- Detecting changes in data.
- The coal mining disasters data set is the classical data set



- Did the restructuring of business unit X have an effect on revenue?
- Did the training initiative result in more sales?
- Did the investment in re-skilling of our workforce translate into more data products?

Where do data products come from?

- **Build vs Buy** --- The best data products will be created by folks with BOTH domain expertise and modern flexible tools.
- **Open-source** --- The reality today is that the best tools are written using open-source technologies
- **Talent goes beyond technical skills**: pair-programming, Agile, microservices architecture, CI/CD, technical communication, team-player.
- **A consistent ecosystem** for data teams has a synergistic effect on ROI when combined with modern software engineering practices like version control and pair-programming

Recap

1. Data science has come so far recently, but why?
2. Demystifying Machine learning: It is in the examples that we see the business value.
3. The key is being able to decision making is to visualize the stories

DEMO: Predicting equipment failure
DEMO: Change point analysis