



MANGO  
SOLUTIONS

# Looking over the Fence

Richard Pugh

Chief Data Scientist

✉ [rich@mango-solutions.com](mailto:rich@mango-solutions.com)

🐦 [@richatmango](https://twitter.com/richatmango)



# Agenda



Introductions



History of Data Science



Data Science Today



Learnings



Discussion





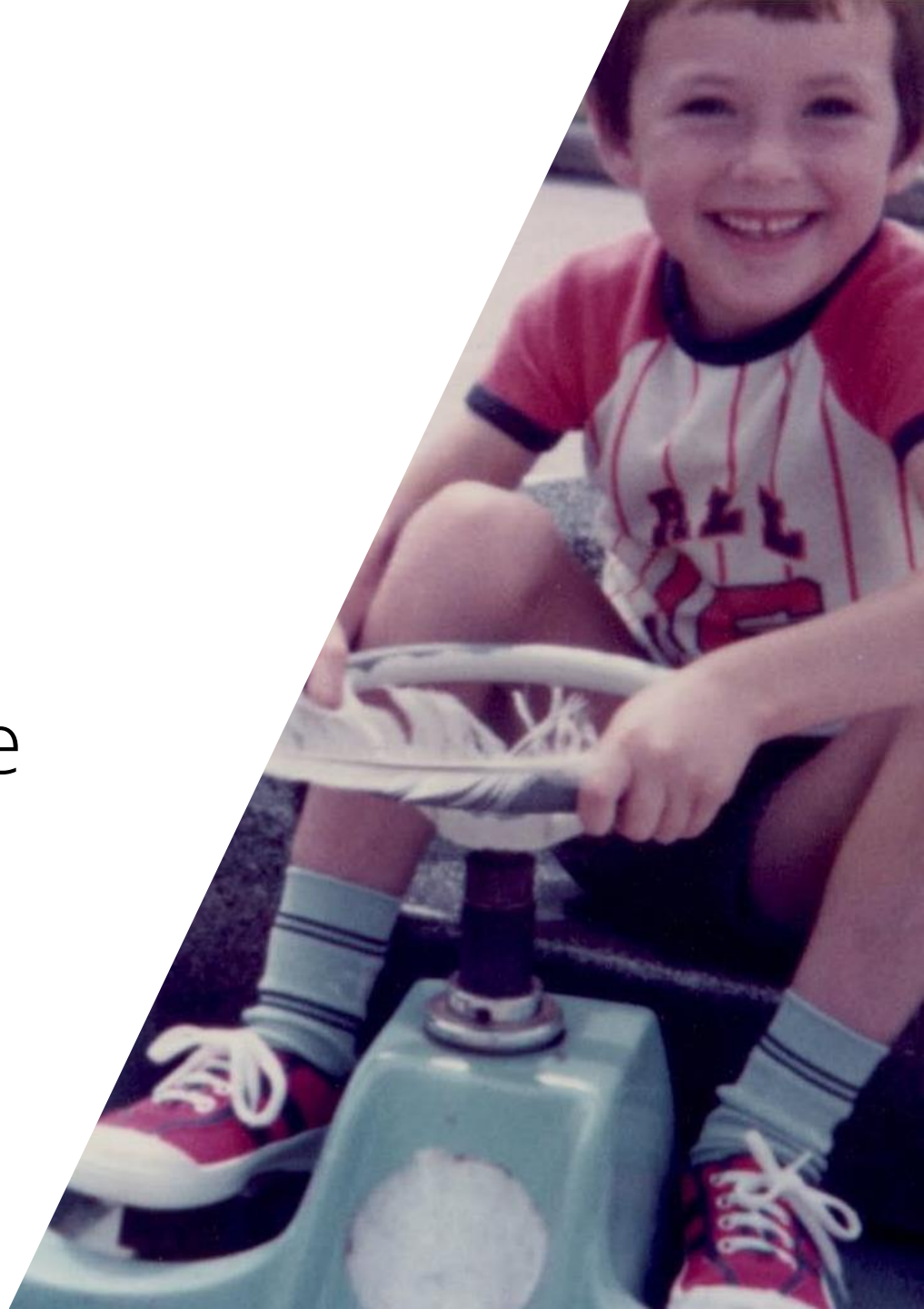
# Introductions





# About Me

- Maths & Stats
- SAS > S+ > R
- RConsortium
- RSS Data Science
- Strategic Advice



# About Mango

- What we do
  - Data Science
  - Data Engineering
  - Strategic Advice
  - Education
- R, Python, Spark
- ~70 People
- +15 Years



# My Perspective





# A (Brief) History of Big Data & Data Science



# History of Statistics

- Basic forms of statistics used throughout history
- Statistical theory during 18<sup>th</sup> Century
- Modern Statistics early 20<sup>th</sup> Century
- GLM in 1972 (Nelder & Wedderburn)
- Advanced of Computer Science





# Professor Jeff Wu



Doug  
Laney

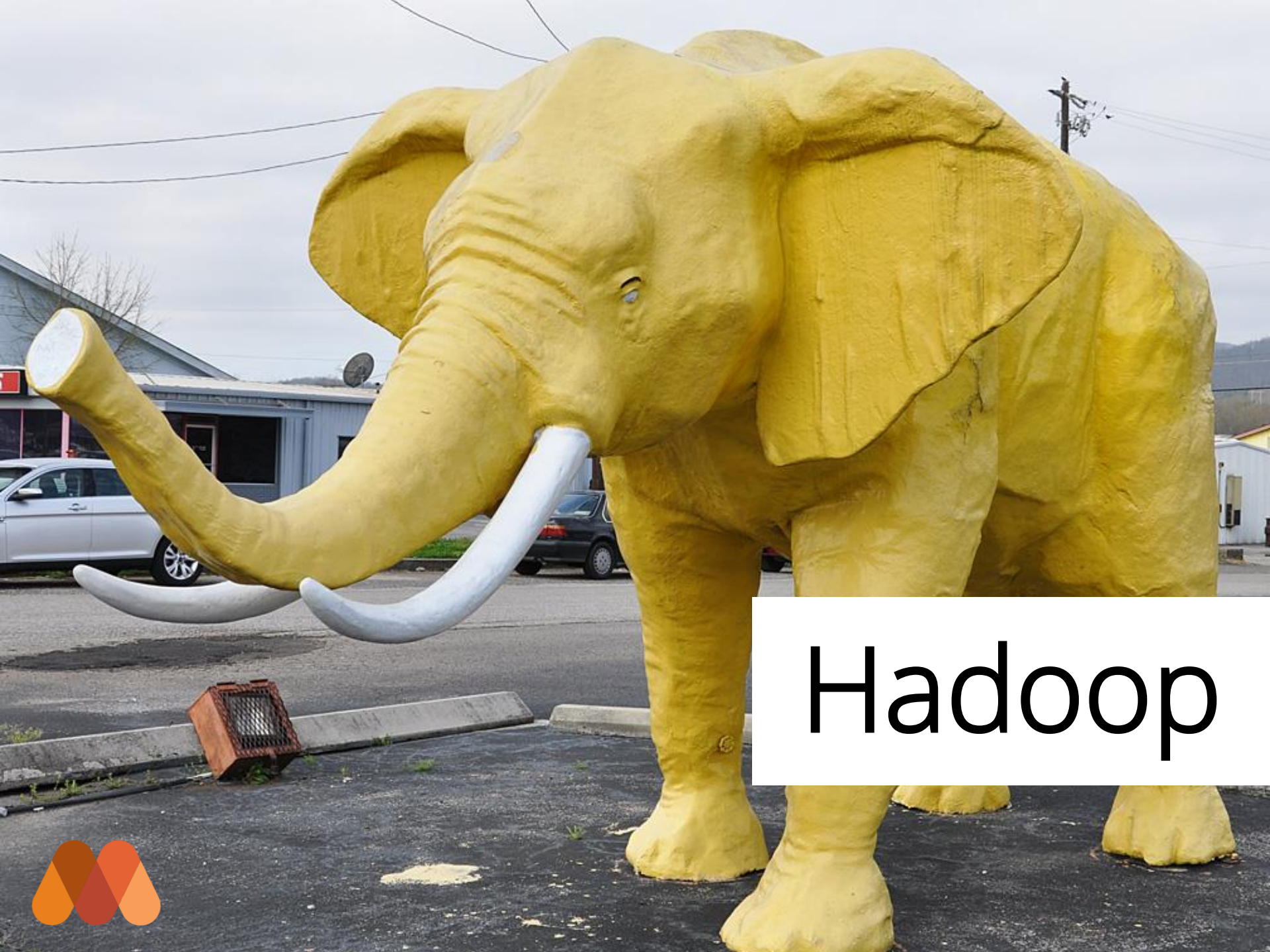




# The 3 Vs of Data

- Volume – size of data
- Velocity – speed at which new data created
- Variety – types of data (Unstructured, Structured)
- *Veracity* – *quality of data*





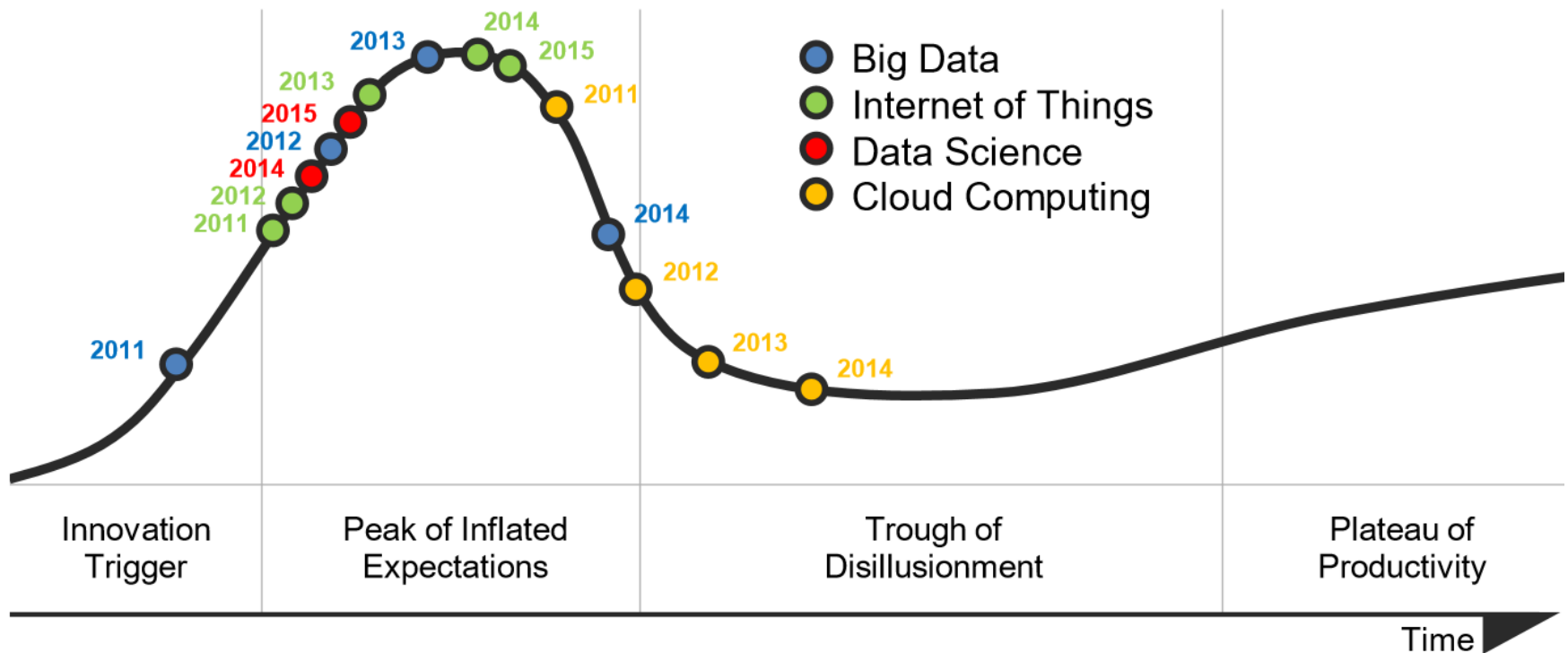
# Hadoop



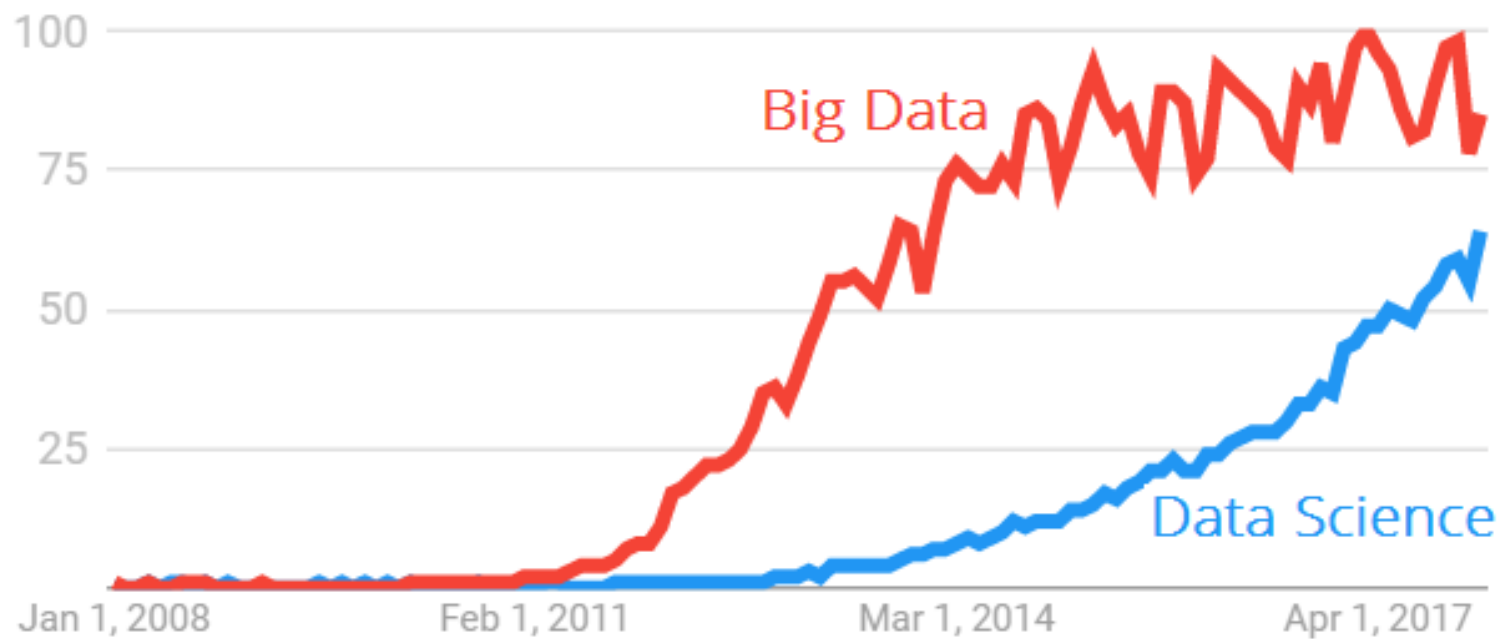


# The Marketing Hype

## The Gartner Hype Curve: 2011-2016



# Growth of Data Science





# Data Science Today





A Data Scientist is a statistician who lives in San Francisco



**Likelihood T. Prior**  
@CjBayesian

A Data Scientist is a device for turning coffee and data into better decisions



**Big Data Borat**  
@BigDataBorat

Data Science is Statistics on a Mac.





**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.



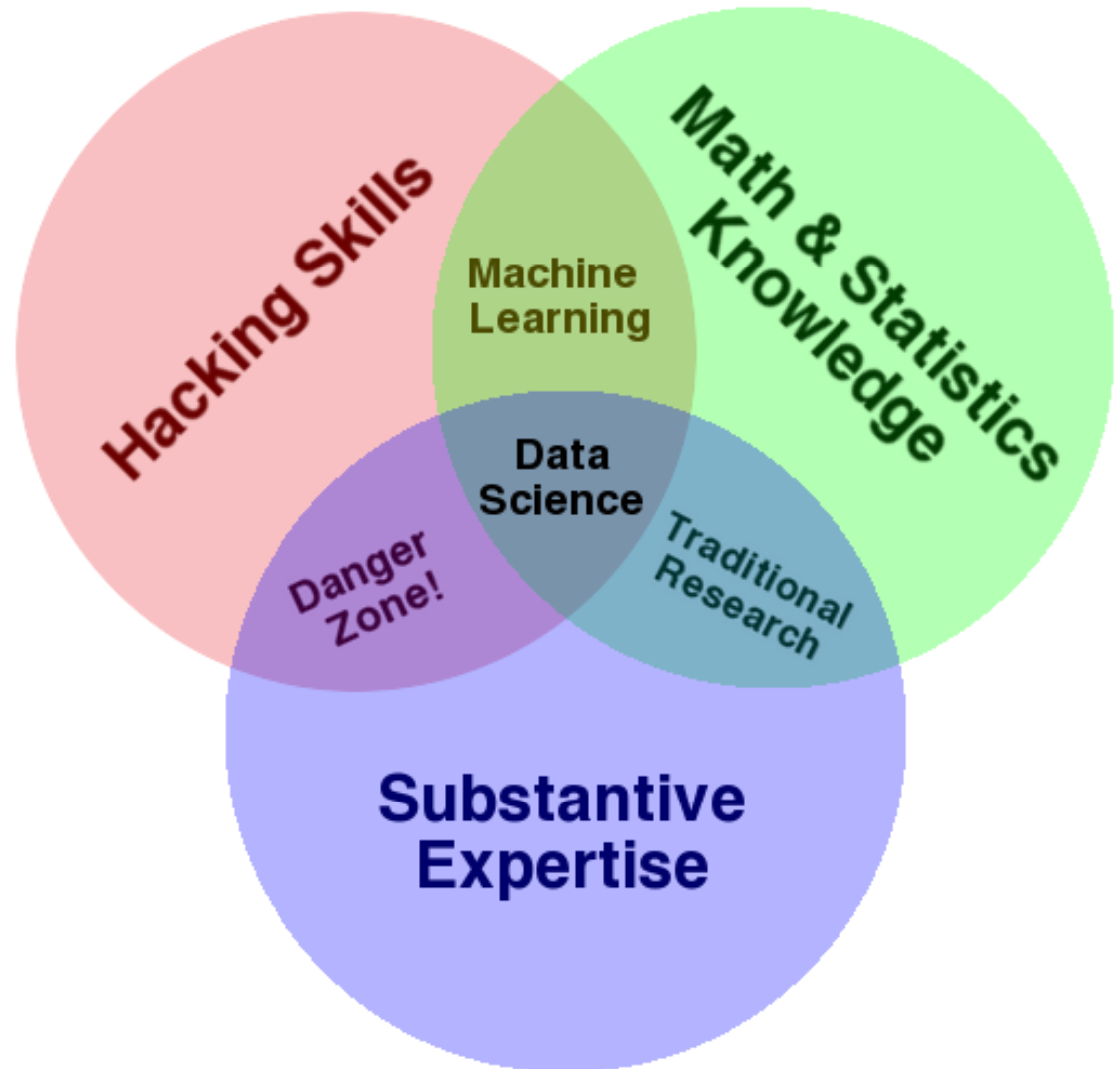


@drewconway



**Drew Conway**

@drewconway



Data Science is ... the  
Proactive use of Data and  
Advanced Analytics to Drive  
Better Decision Making



Data Science is ... the  
Proactive use of **Data** and  
Advanced Analytics to Drive  
Better Decision Making





Data Science is ... the  
Proactive use of Data and  
Advanced Analytics to Drive  
Better Decision Making



Data Science is ... the  
**Proactive** use of Data and  
Advanced Analytics to Drive  
Better Decision Making



Data Science is ... the  
Proactive use of Data and  
Advanced Analytics to Drive  
Better Decision Making





Data Science is ... the  
Proactive use of Data and  
Advanced Analytics to Drive  
Better Decision Making



Engage with Business

Work with Data

Data Science is ... the  
Proactive use of Data and  
Advanced Analytics to Drive  
Better Decision Making

Analytic Breadth

Industrialise > Code



# Learnings from Other Industries





# Where are we trying to get to?

- Becoming “Data-Driven” Companies
  - Data valued as major asset
  - Insight used to drive better decisions
- Benefits
  - Gain/Maintain competitive advantage
  - Stay ahead of “disruptors”



# Many Different Organisations

Analytically  
Mature

Pharmaceutical  
Insurance

Banking  
Retail

Government  
Energy

Analytically  
Immature

Recruitment  
Sports

Legal  
HR

Media  
SMEs

Disruptors

airbnb  
Facebook

uber  
alibaba

SocietyOne  
Netflix



# Many Different Organisations

Analytically  
Mature

Pharmaceutical  
Insurance

Banking  
Retail

Government  
Energy

Analytically  
Immature

Recruitment  
Sports

Legal  
HR

Media  
SMEs

Disruptors

airbnb  
Facebook

uber  
alibaba

SocietyOne  
Netflix



# #1 Analytic Remit

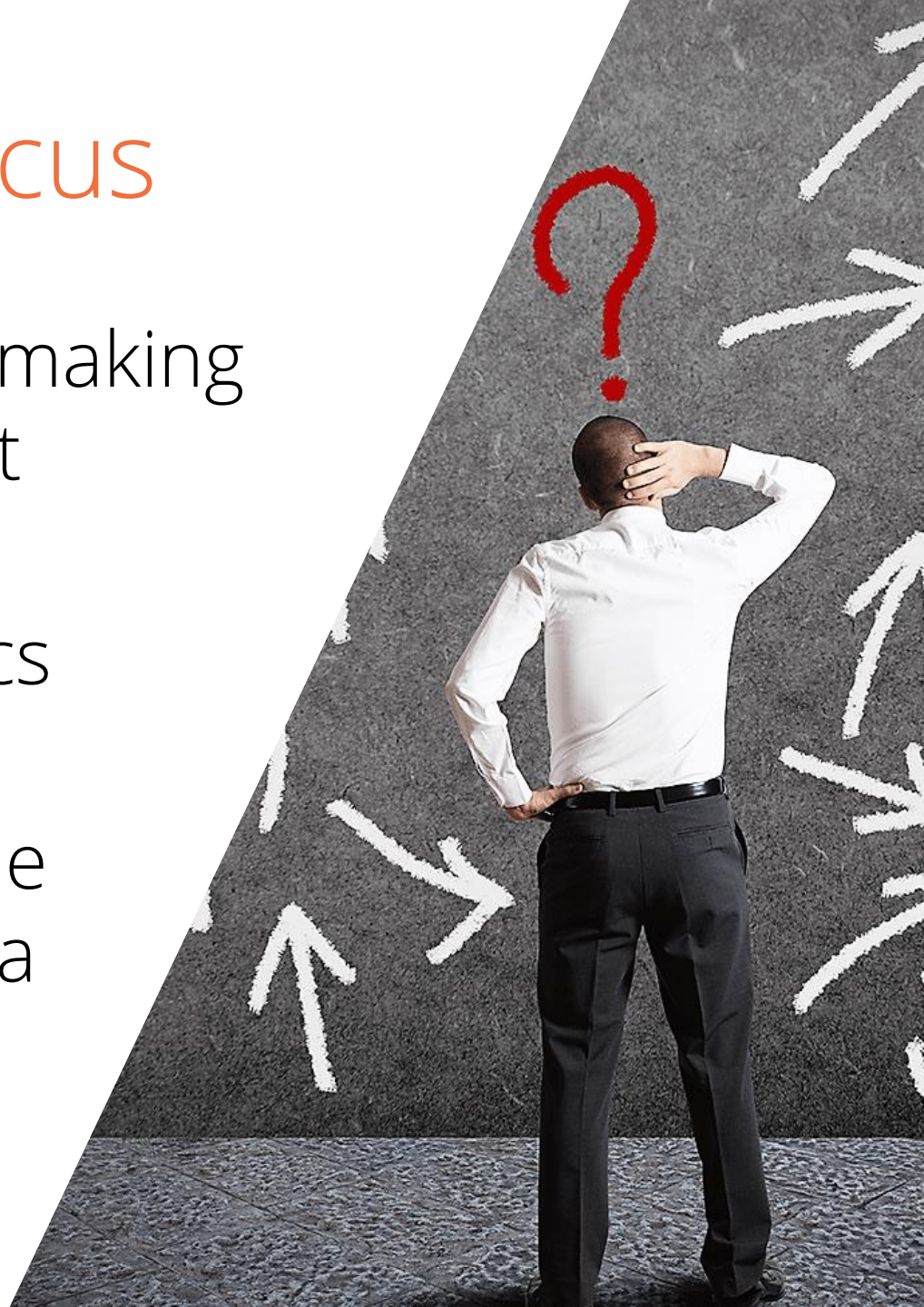
- Analytics now a more Strategic Initiative
- Asked to solve wider range of challenges
- Great time to be a Data Scientist!





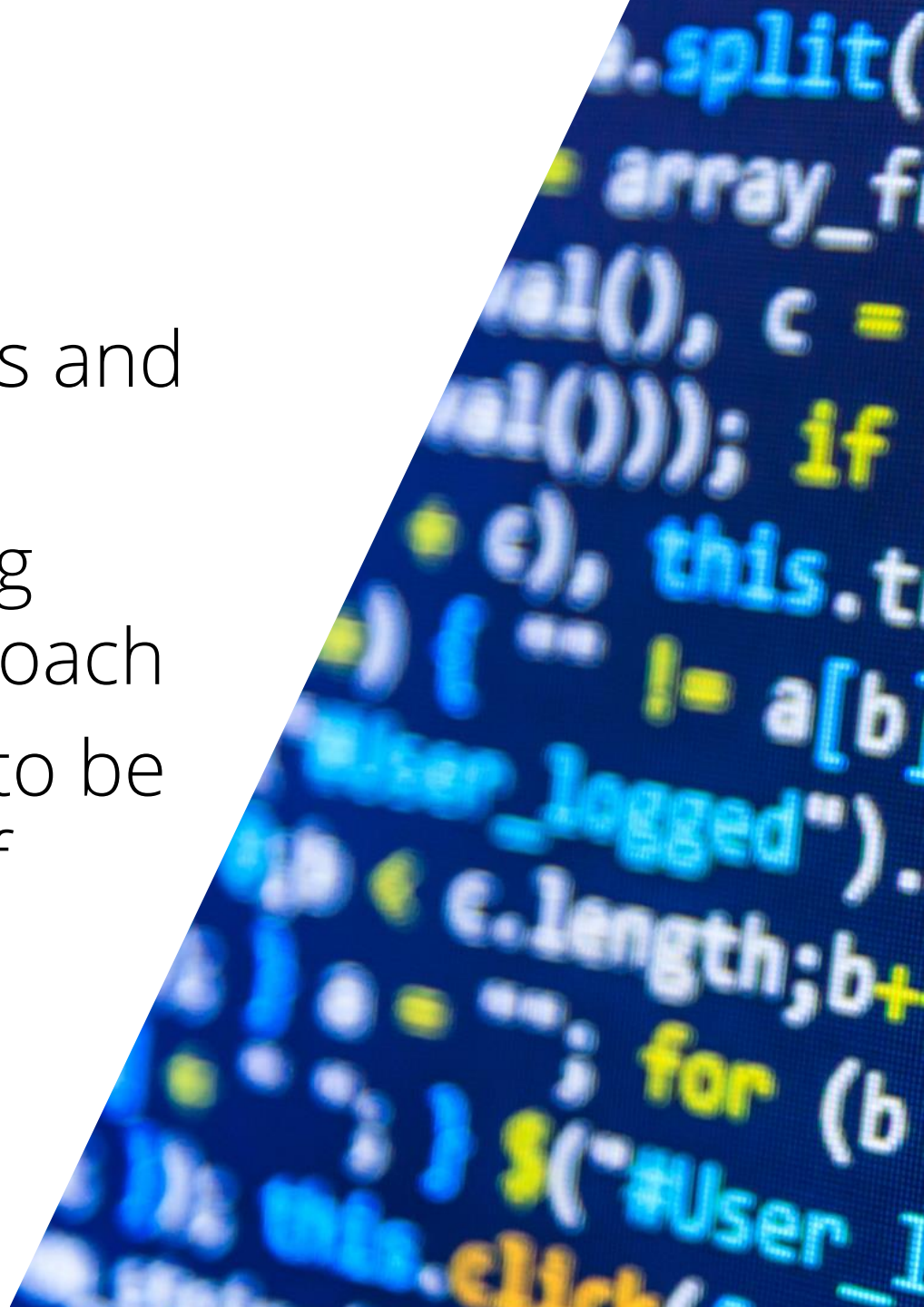
## #2 Decision Focus

- Focus on decision making to drive investment
- Great device for connecting analytics with business
- Establishes baseline and success criteria



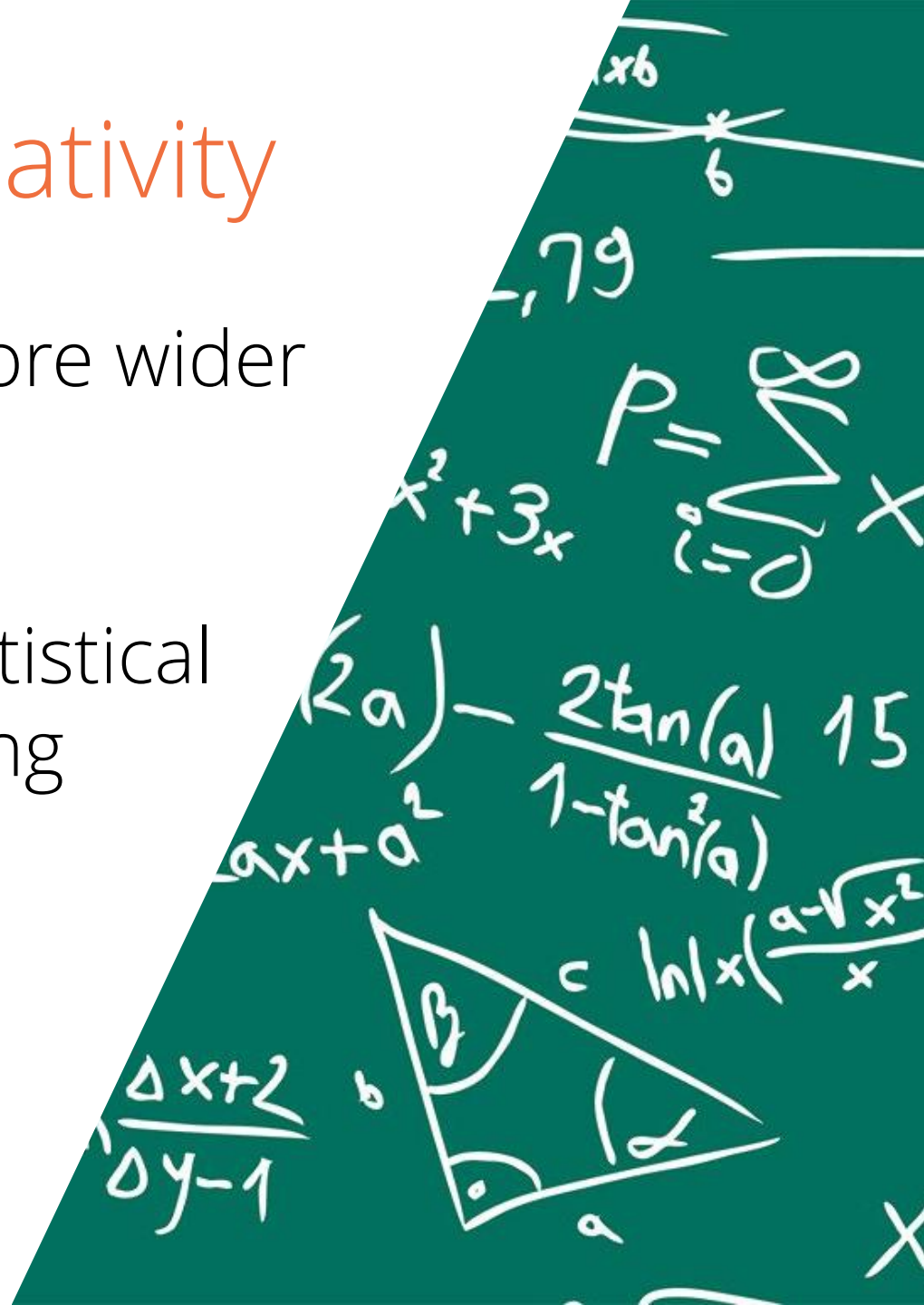
# #3 DataOps

- The mix of analytics and DevOps
- More formal coding and technical approach
- Enabling analytics to be increasingly part of systematic analytic approach



# #4 Analytic Creativity

- Great time to explore wider range of analytics approaches
- Comparison of Statistical vs Machine Learning





# #5 Innovation

- Willingness to invest in innovation
- Great time to look at new data technologies or data sources





# Summary



# Summary

- The Big Data and Data Science fields have been driven largely by “hype”
- They still contain great ideas and technologies that should be explored
- They represent a great opportunity to promote the use of Analytics and the creation of Data-Driven companies



# Questions?

Richard Pugh

✉ [rich@mango-solutions.com](mailto:rich@mango-solutions.com)

🐦 [@richatmango](https://twitter.com/richatmango)

