```
library(dplyr)

r-ladies_global %>%
  filter(city = 'London')
```

@alice_data
@AnalyticsPanda

Demystifying Data Science

London Business Analytics Group - Feb ‘17

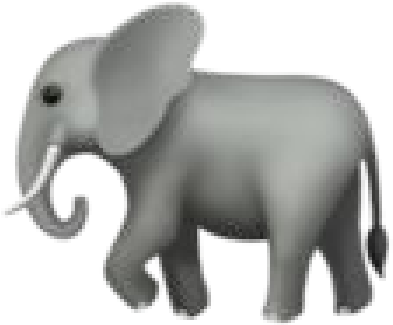
Who are we?

Alice Daish

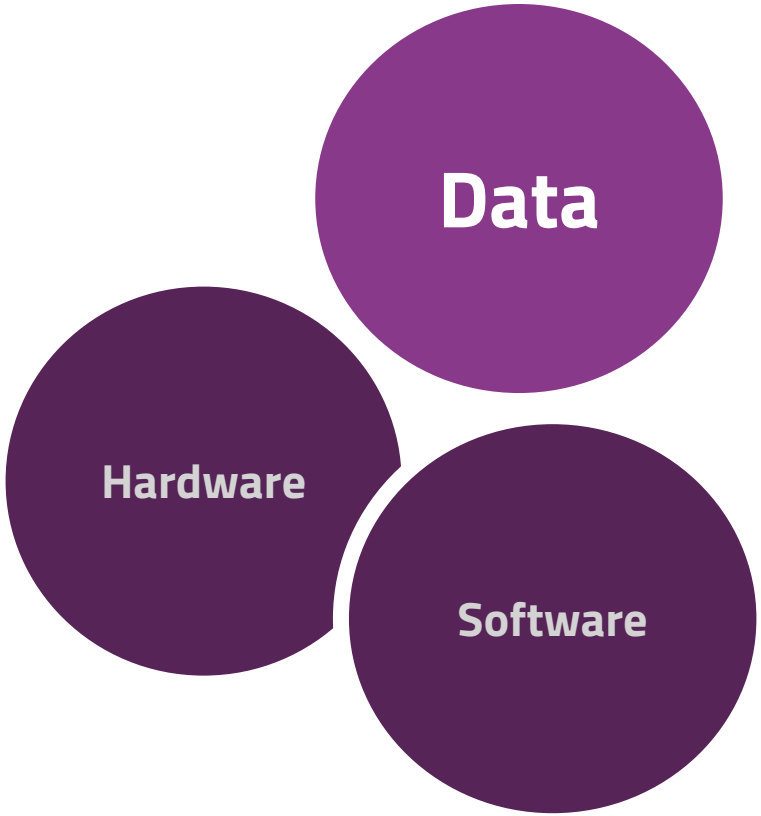
Chiin-Rui Tan

Co-Founder of R-Ladies Global

Welcome



Tech Industry - not all Apps, Robots & PhDs

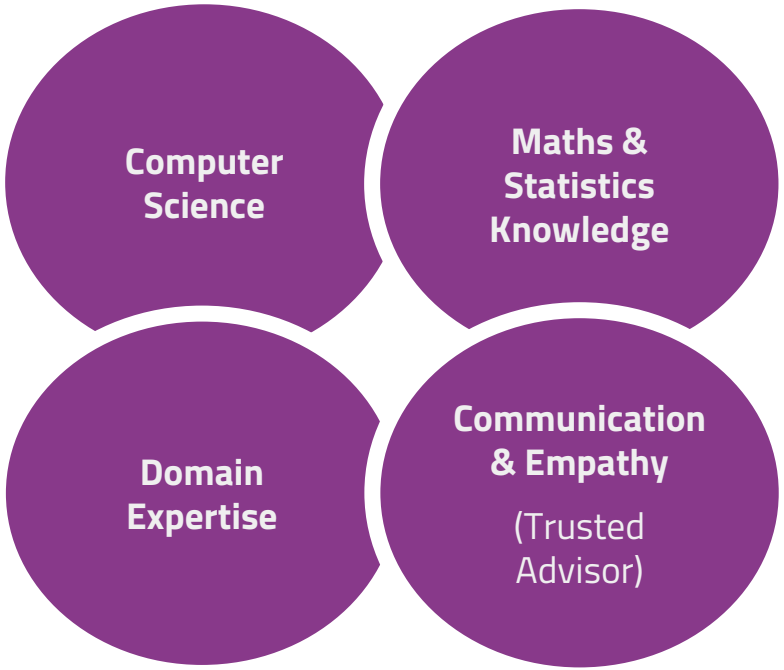


P.S you do **not** have to have a PhD in Computer Science to work in Tech!!

"69% of all (coders) tell us they are at least partly self-taught.

13% of respondents across the globe tell us they are ***only*** self-taught."

—Stack Overflow
Developer Survey Results 2016
56,033 coders in 173 countries



What is Data Science?

Our suggestion:

“The application of scientific methods to Data which enable new discoveries to be made & utilised”

- General consensus = There is no consensus!
- DJ Patil & Jeff Hammerbacher purport to have first coined the term “Data Scientist” as a job title in 2008
- Hybrid & Interdisciplinary

- P.S Don’t assume an Organisation recruiting for a Data Scientist actually knows what a Data Scientist is…..!

Data Science - Misconceptions

Data Science != Hadoop

Data Science != Big Data

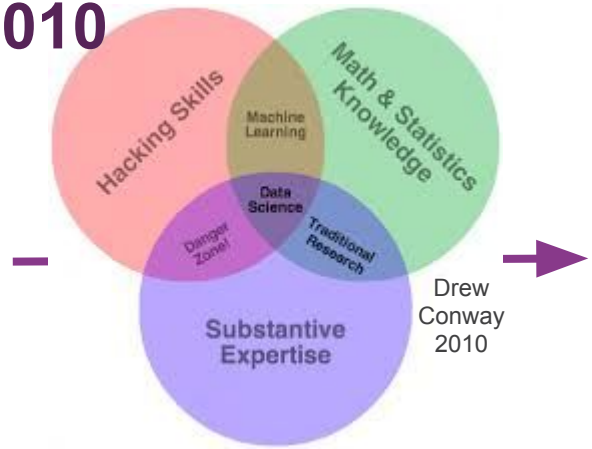
Data Science != Machine Learning

Data Science != Just Programming

Data Science != Successful Data Science

Skills

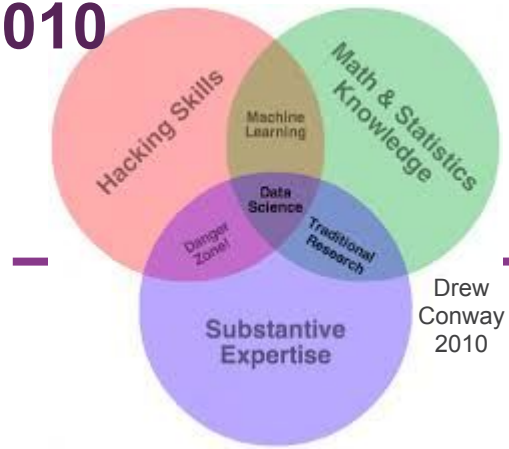
2010



Venn Timeline

David Taylor - http://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

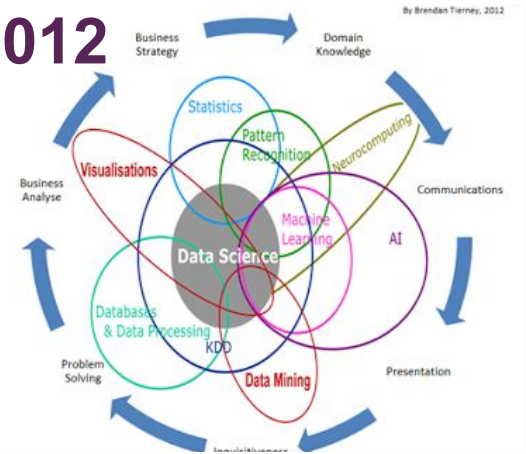
2010



–

Drew Conway 2010

2012

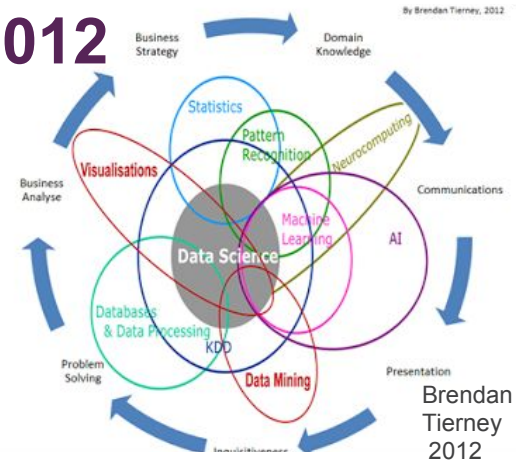


Venn Timeline

2010

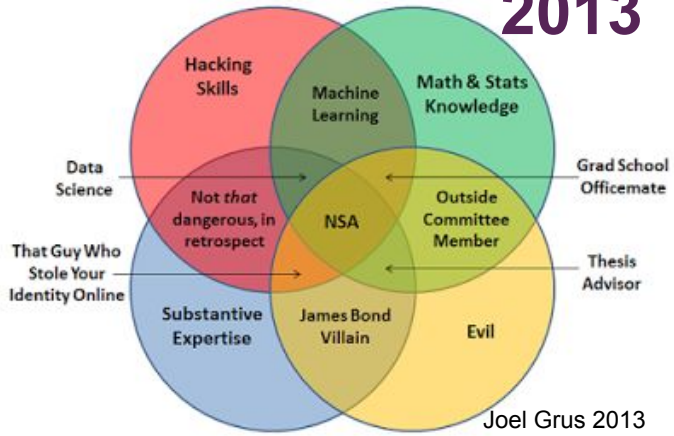
Drew Conway 2010

2012



Brendan Tierney 2012

2013



Joel Grus 2013

Venn Timeline

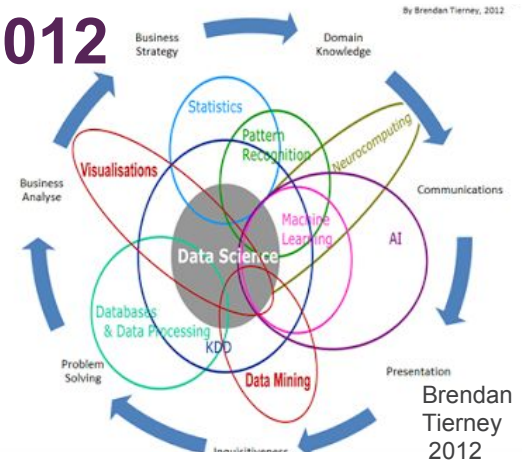
David Taylor - http://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

Venn Timeline

2010

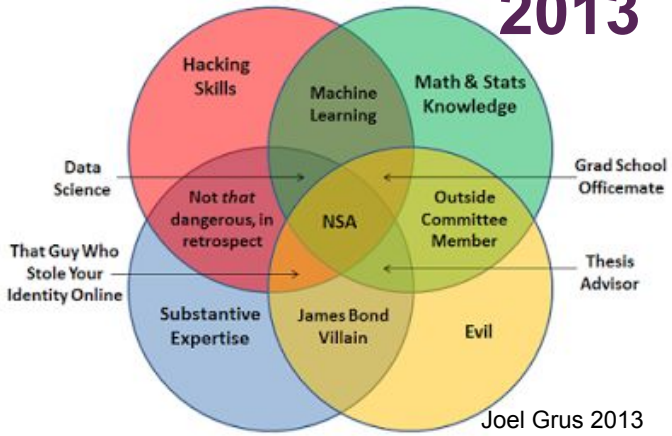
Drew Conway 2010

2012

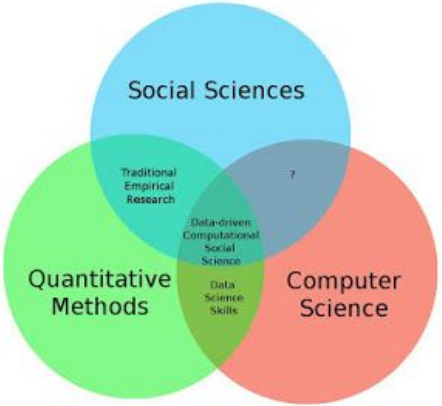


Brendan Tierney 2012

2013



Joel Grus 2013



Ulrich Matter 2013

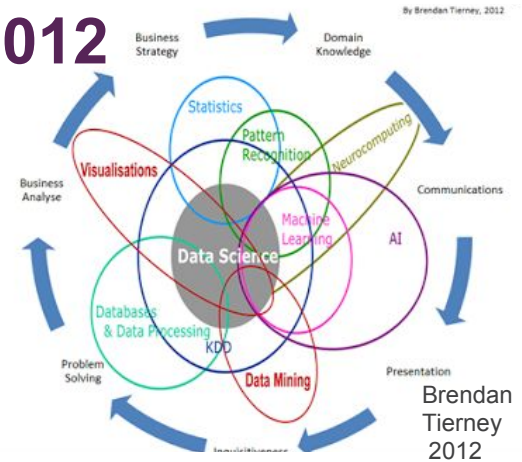
David Taylor - http://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

Venn Timeline

2010

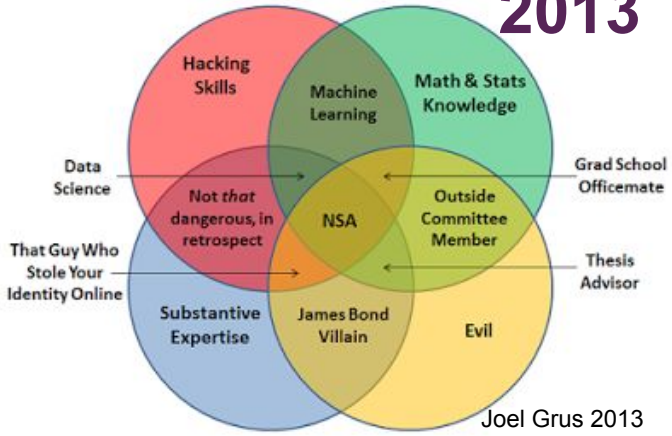
Drew Conway 2010

2012

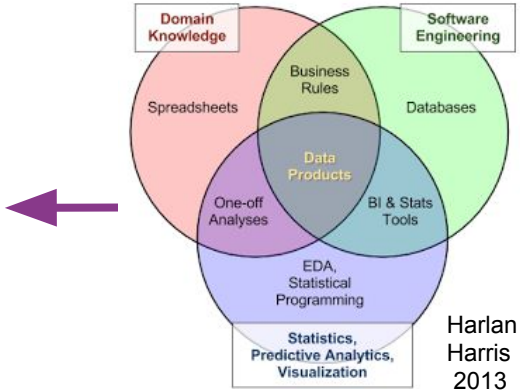


Brendan Tierney 2012

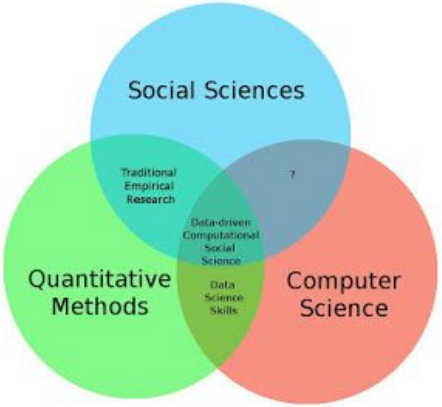
2013



Joel Grus 2013



Harlan Harris 2013



Ulrich Matter 2013

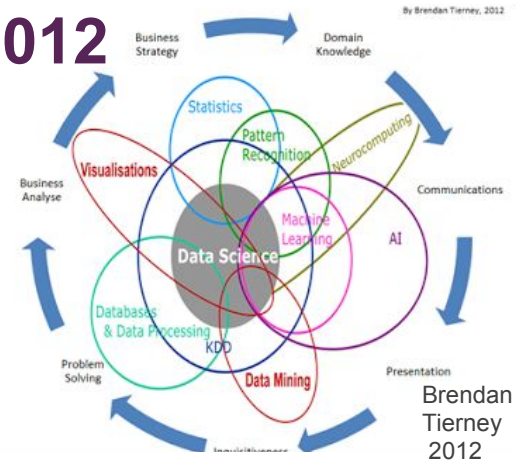
David Taylor - http://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

Venn Timeline

2010

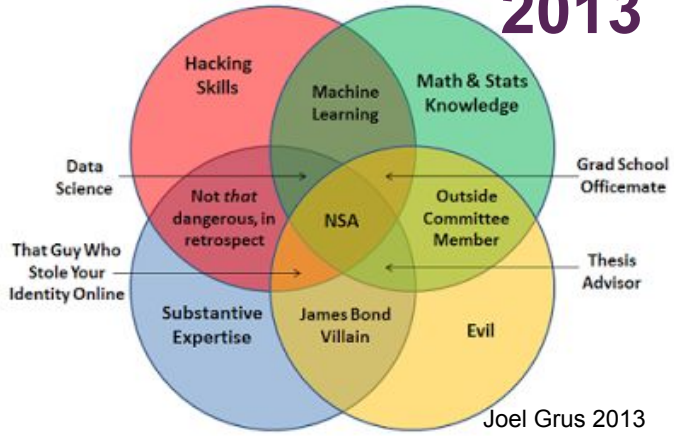
Drew Conway 2010

2012



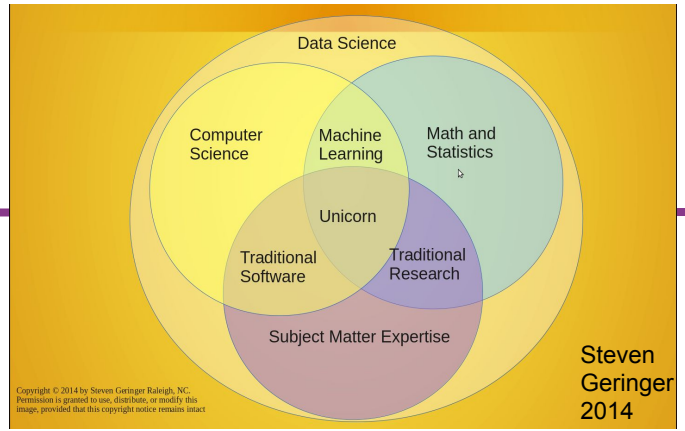
Brendan Tierney 2012

2013

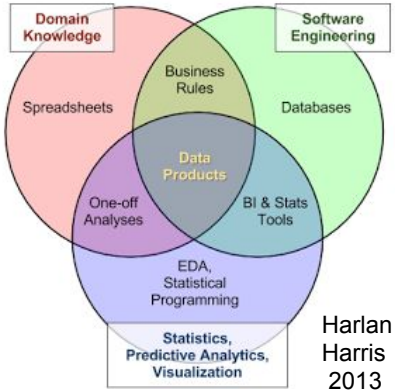


Joel Grus 2013

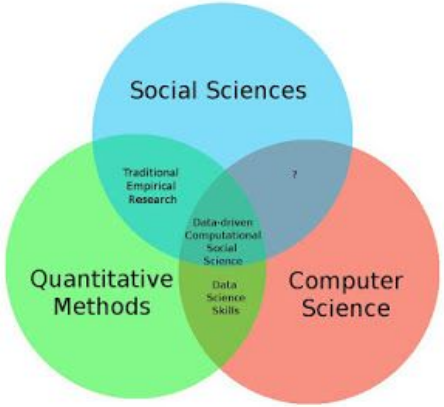
2014



Steven Geringer 2014

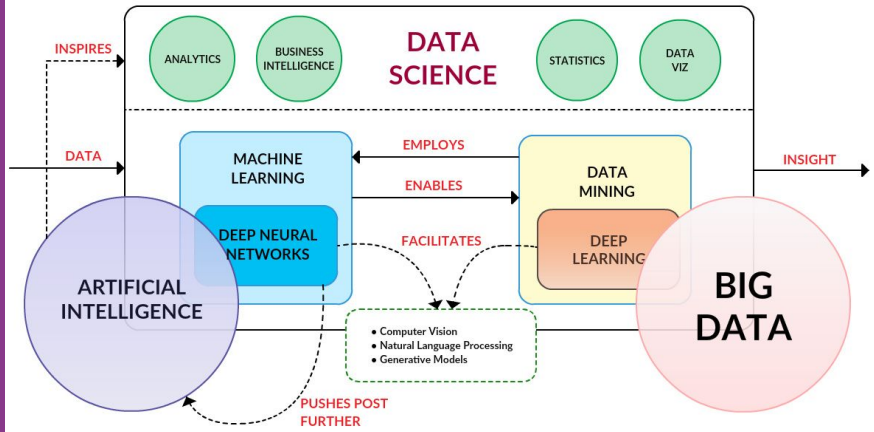


Harlan Harris 2013

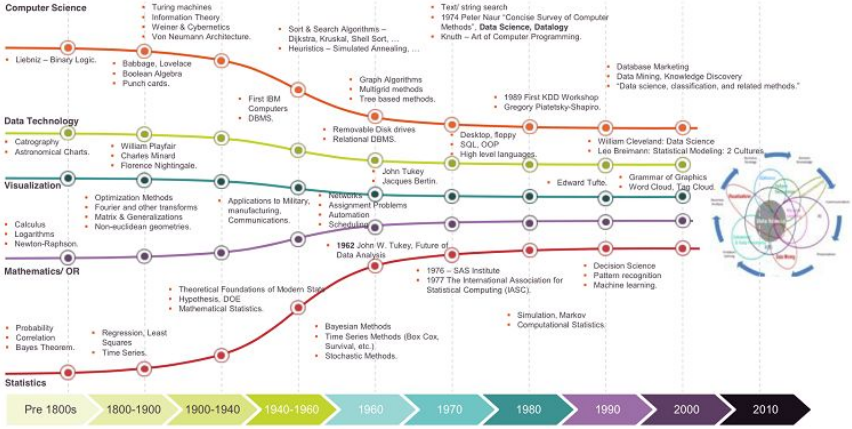


Ulrich Matter 2013

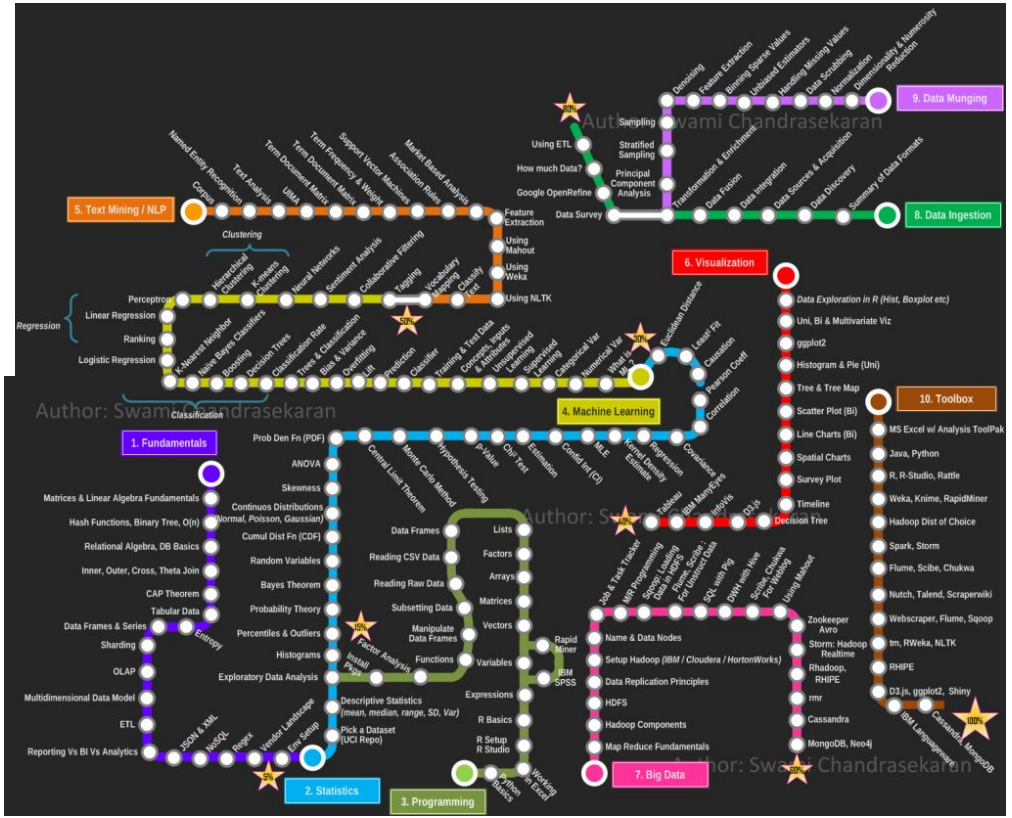
David Taylor - http://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html



http://www.kdnuggets.com/2016/10/deep-learning-key-terms-explained.html

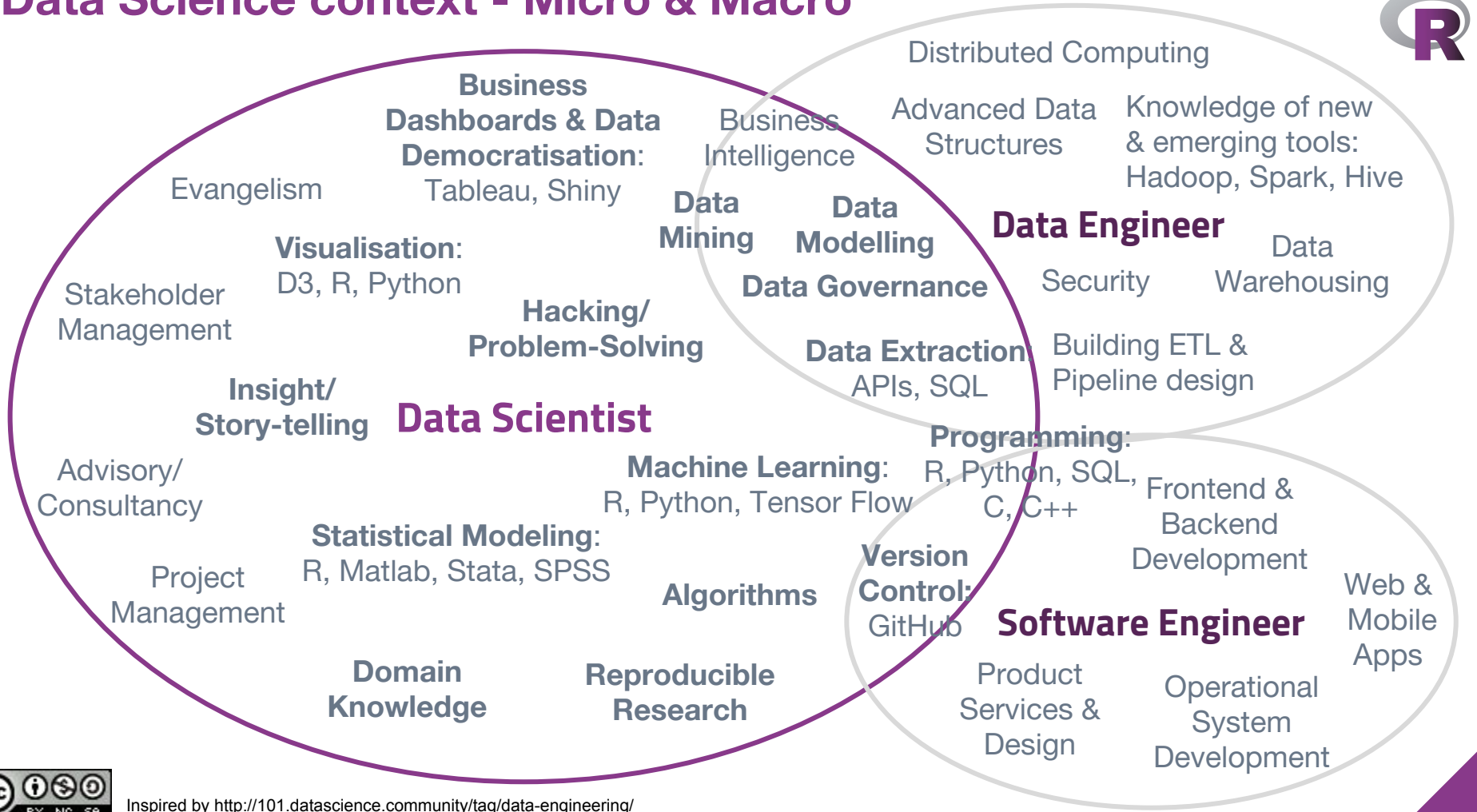


http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/



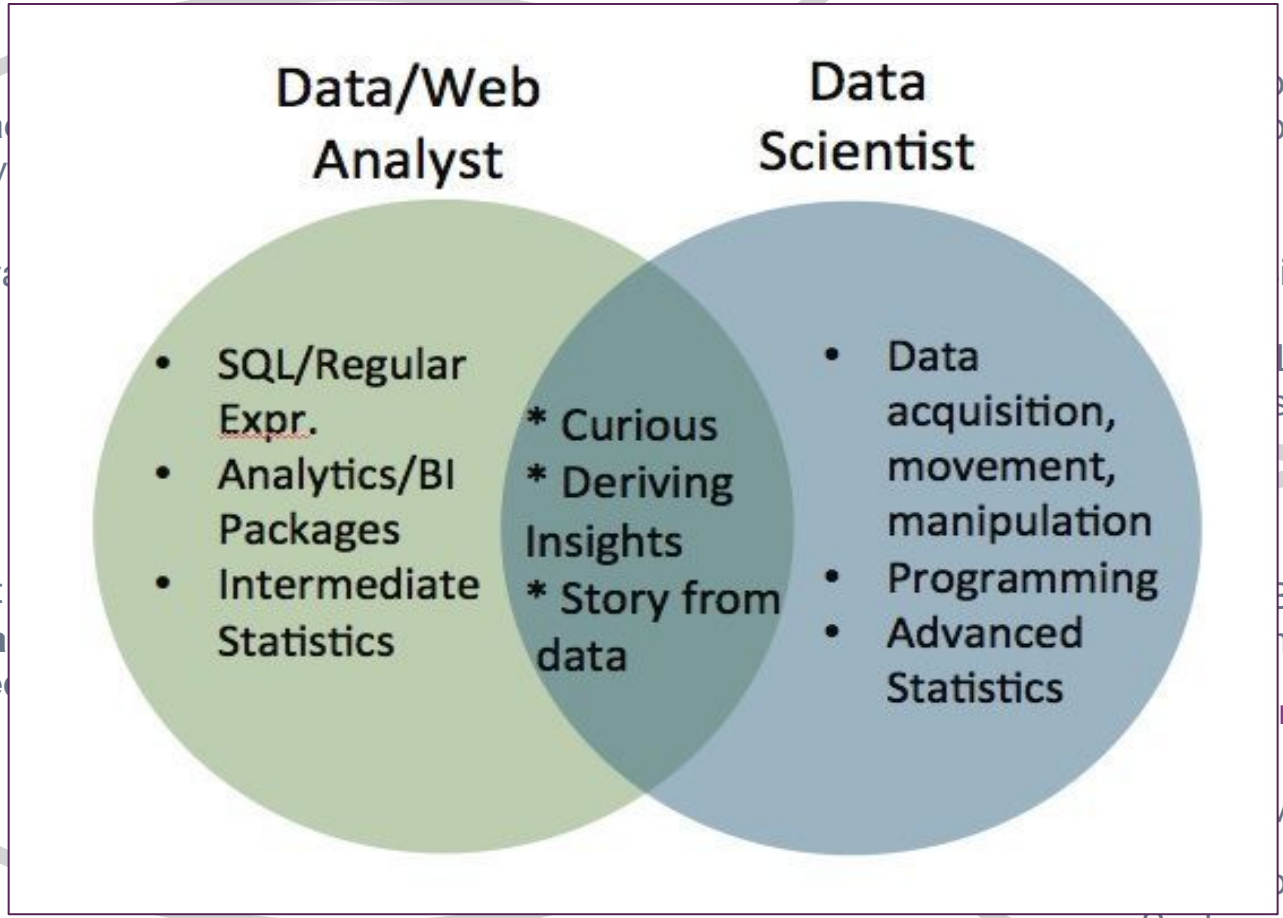
http://nirvacana.com/thoughts/becoming-a-data-scientist/

Data Science context - Micro & Macro



Inspired by http://101.datascience.community/tag/data-engineering/

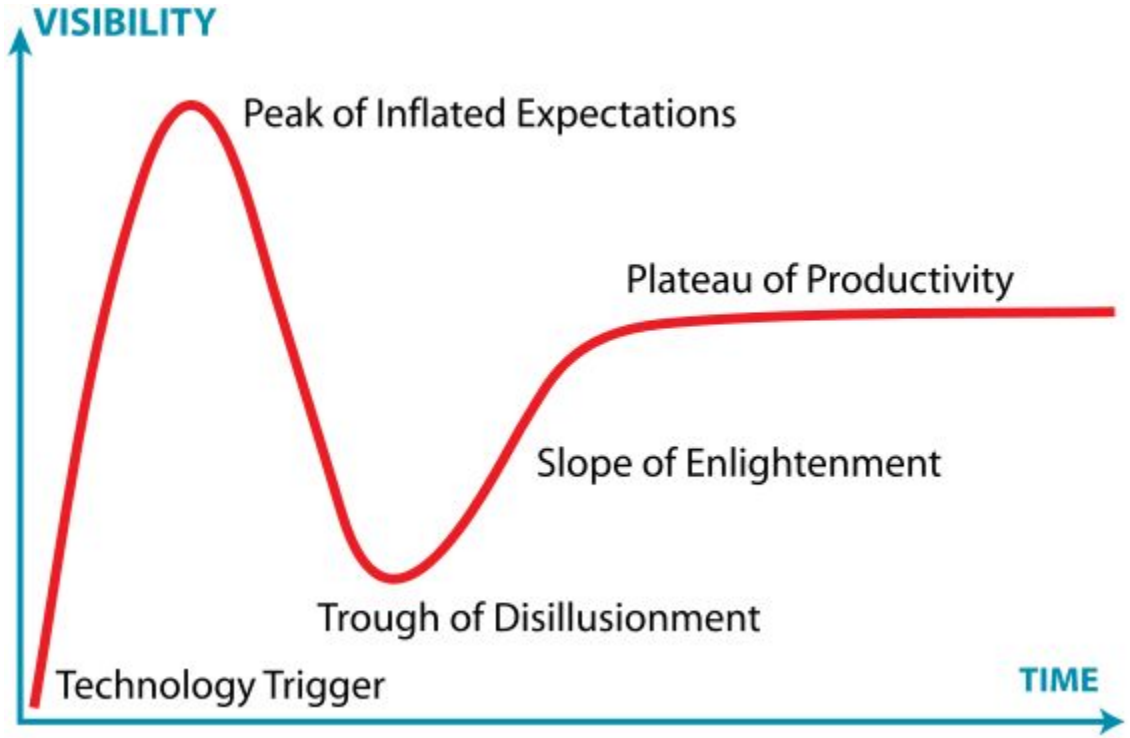
Data Science in context - Macro & Micro



Inspired by http://101.datascience.community/tag/data-engineering/

Hype

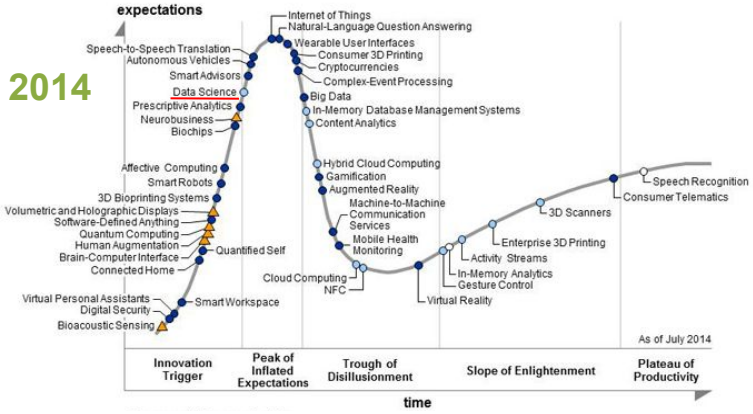
Hype Cycle



https://en.wikipedia.org/wiki/Hype_cycle

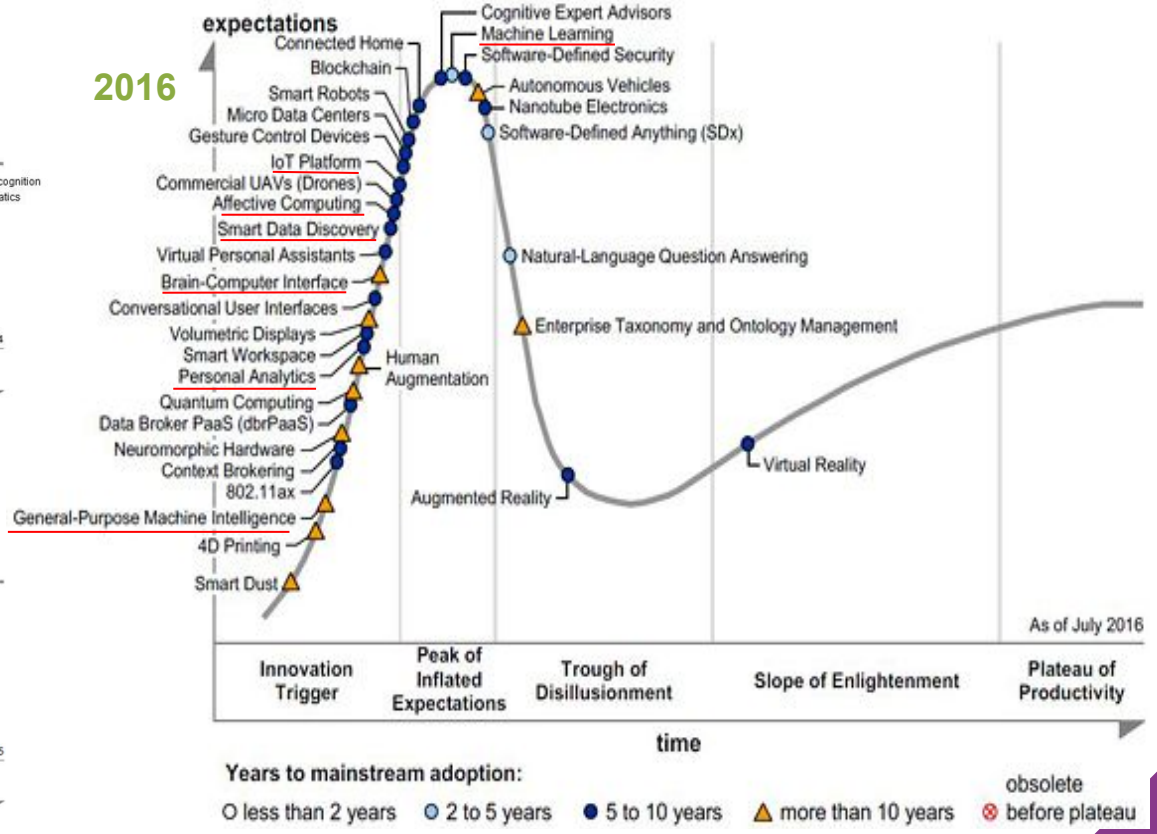
Hype Cycle - Silver Bullets?

2014



2015

2016



http://www.gartner.com/

Hype Cycles - Has Data Science survived?

Data Science does not start by analysing big data but instead by asking key business questions.

What do we know?
What do we not know?
What have not been answered yet?
What are the known unknowns?
What are the known known?

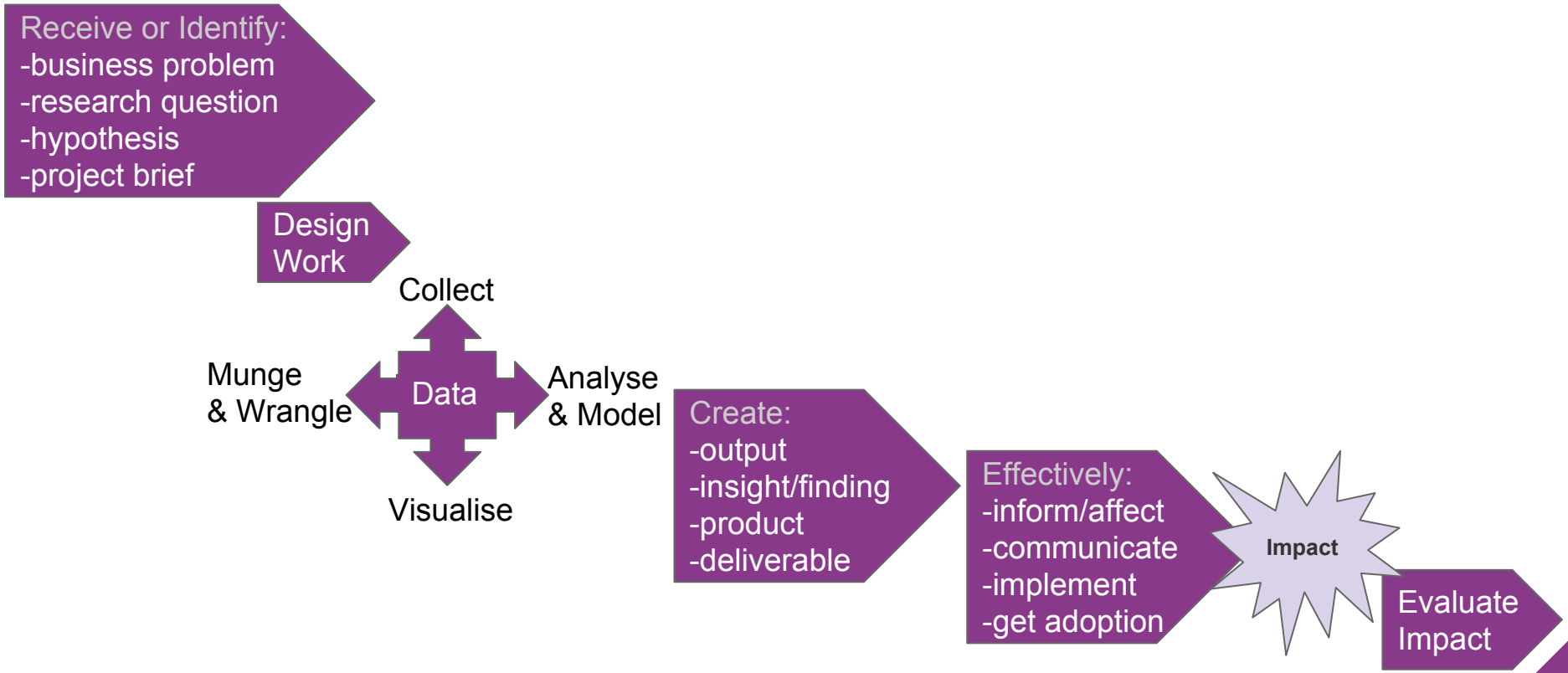
YES

"Data Science is about discovering and showing people things they didn't know using data."

http://datasciencelondon.org/data-science-london/

Workflow

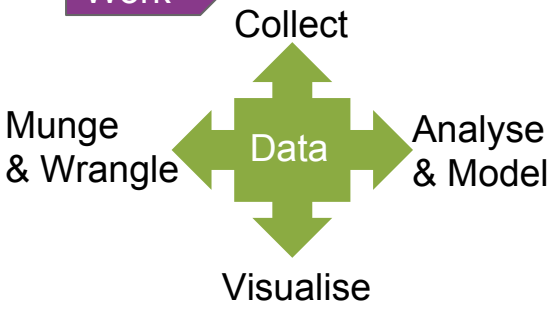
Data Science - Workflow



Data Science - Workflow

Receive or Identify:
-business problem
-research question
-hypothesis
-project brief

Design Work



Create:
-output
-insight/finding
-product
-deliverable

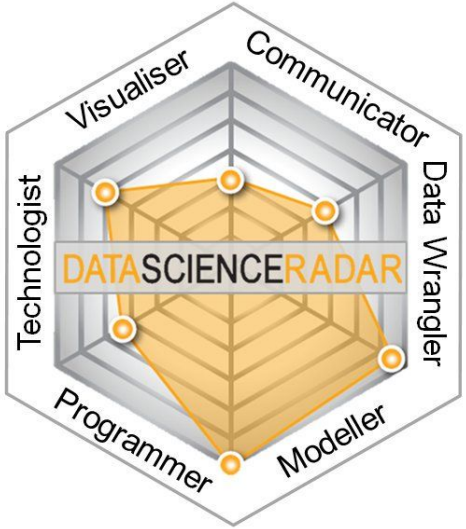
Effectively:
-inform/affect
-communicate
-implement
-get adoption

Impact

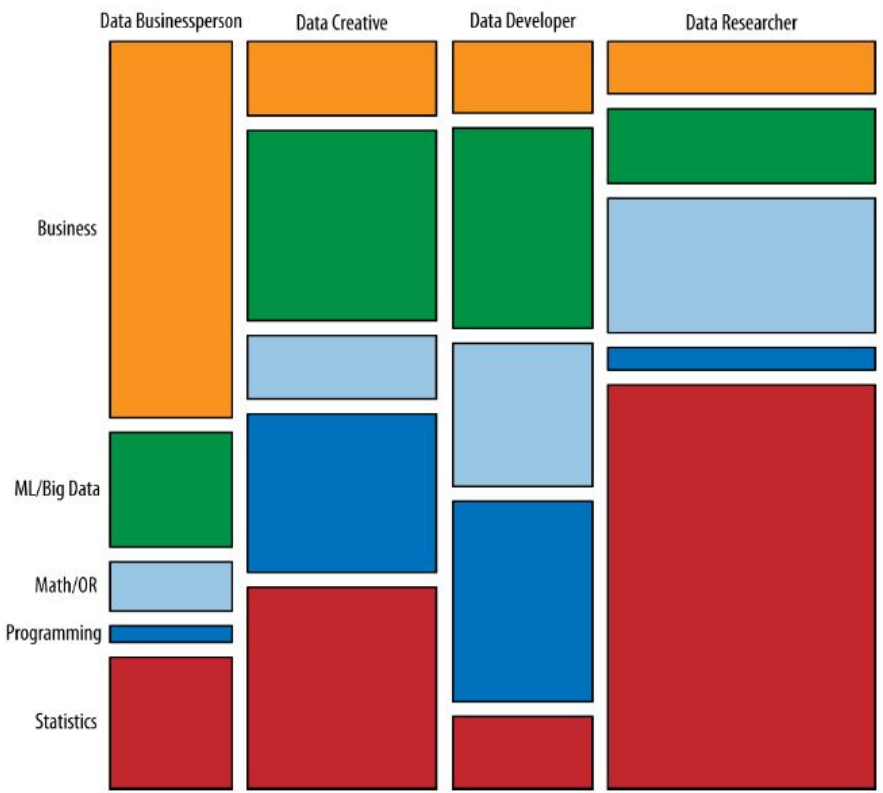
Evaluate Impact

Computational data skills are Necessary but **not** Sufficient!!

Data Science - The Unicorn = Group of Horses



https://www.mango-solutions.com/radar/



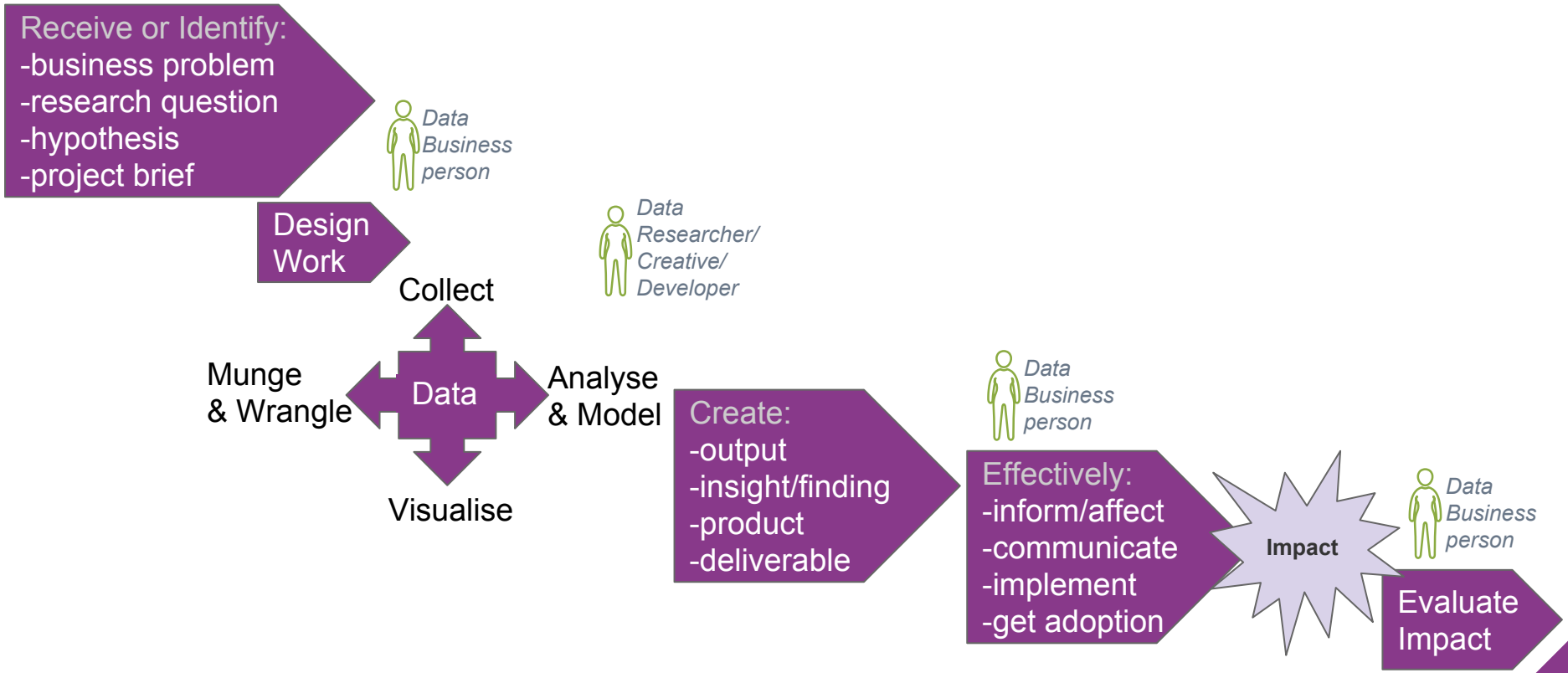
http://radar.oreilly.com/2013/06/theres-more-than-one-kind-of-data-scientist.html

Data Science - The Unicorn = Group of Horses

- **Data Businesspeople** are the product and profit-focused data scientists. They're leaders, managers, and entrepreneurs, but with a technical bent. A common educational path is an engineering degree paired with an MBA.
- **Data Creatives** are eclectic jacks-of-all-trades, able to work with a broad range of data and tools. They may think of themselves as artists or hackers, and excel at visualization and open source technologies.
- **Data Developers** are focused on writing software to do analytic, statistical, and machine learning tasks, often in production environments. They often have computer science degrees, and often work with so-called "big data".
- **Data Researchers** apply their scientific training, and the tools and techniques they learned in academia, to organizational data. They may have PhDs, and their creative applications of mathematical tools yields valuable insights and products.

http://radar.oreilly.com/2013/06/theres-more-than-one-kind-of-data-scientist.html

Data Science Workflow



Pitfalls

Data Science Pitfalls - Noise

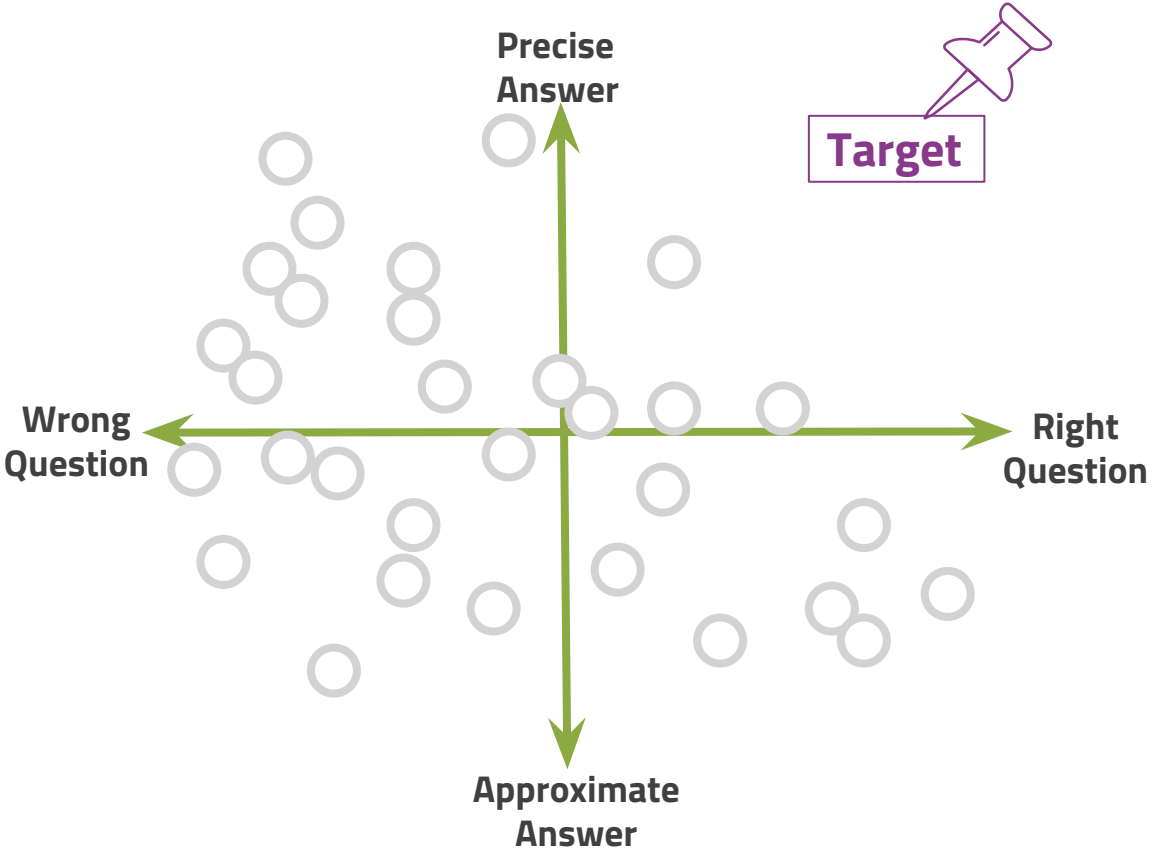
Data Science Pitfalls - Beware of the HiPPO!

Data Science Pitfalls - Wrong Question

"Far better an **approximate answer to the right question**, which is often vague, than an exact answer to the wrong question, which can always be made precise."

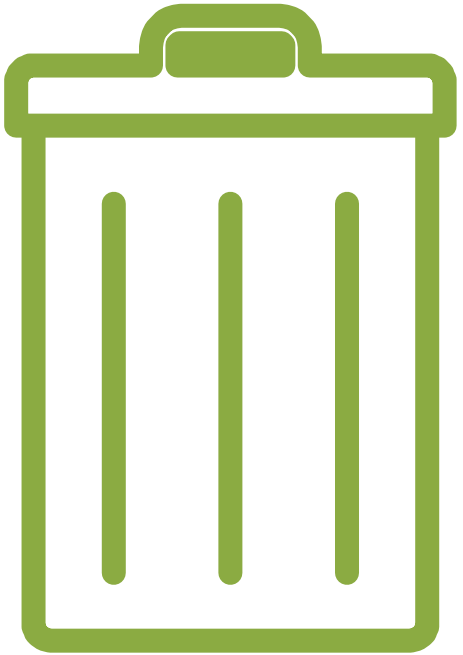
—John W. Tukey

The Future of Data Analysis (July 1961)

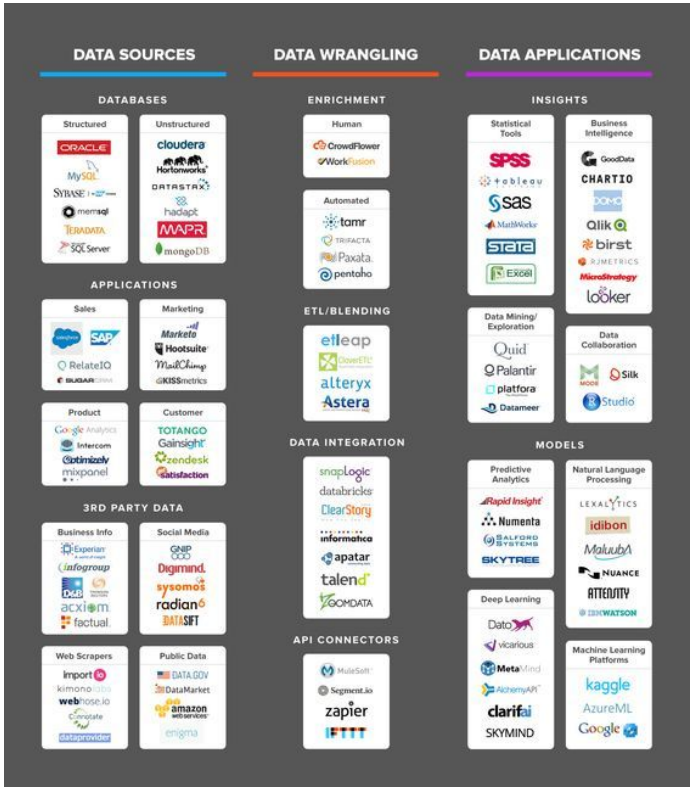


Data Science - Pitfalls

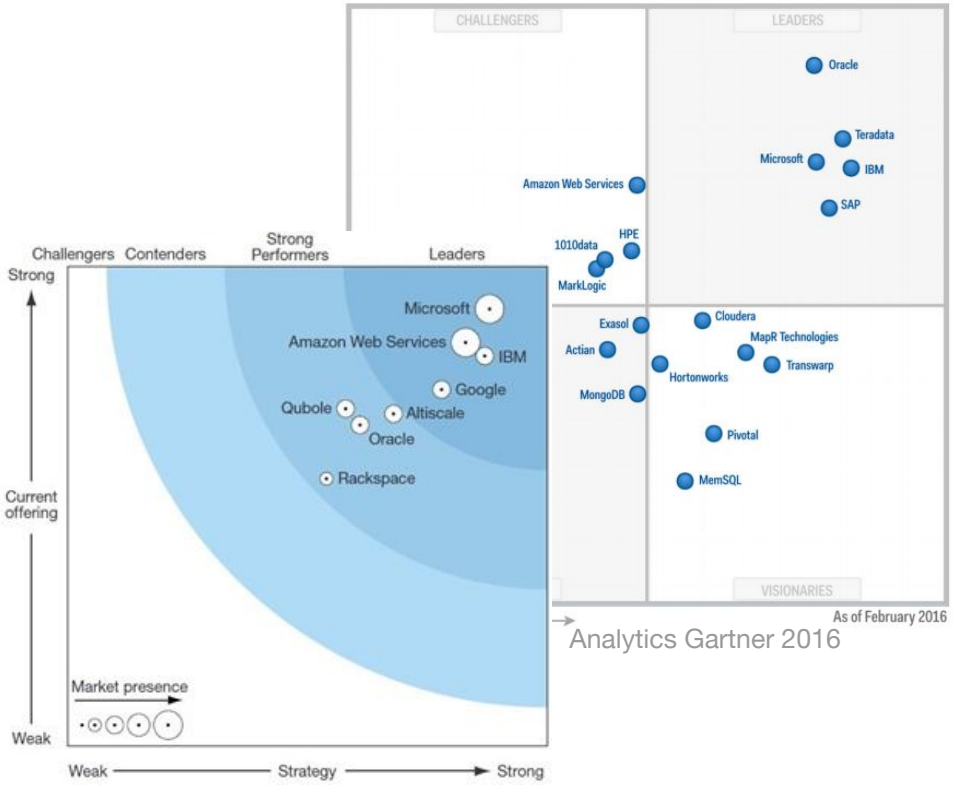
“80%
of Data Science projects never go to production”



NO SOLUTION IS A SILVER BULLET



https://www.crowdflower.com/the-data-science-ecosystem/



Analytics Gartner 2016

Big Data Forrester 2016

Data Science Pitfalls - Blind Leading the Blind?

Data Scientist

The Co-op

Manchester, England

£60,000 to £80,000 p.a

We're looking to recruit an experienced Data Scientist to join our team, and unearth information from our data assets that'll help Co-op make smarter decisions for our members. You'll work as an internal data consultant with people from right across the Co-op family to find simple solutions to complex problems. And if you can bring us the combination of data expertise & great people skills we need, we can offer you a unique opportunity to help build our data capabilities from the ground up. To be successful as a Data Scientist at Co-op, you'll need to be an established data expert. Statistical models will be your speciality, but in addition to tried and tested technical ability, it's important you've got the skills to put these techniques to good use in our business.

What you'll do:

- Understand Co-op objectives, and create algorithms that deliver positive impacts for members
- Be capable and confident in tackling the most complex data-driven and analytical problems
- Work across several simultaneous projects with partners from various Co-op businesses
- Take on different roles on different projects (expert adviser/statistical guru/lead developer)
- Write code when you need to in order to tackle and troubleshoot the most complex tasks
- Make sure new problem solving approaches comply with existing Co-op rules and values
- Plan projects over a 12 month horizon, and manage budgets using Agile principles
- Contribute to the coaching & mentoring of a pool of less experienced colleagues
- Design solutions that create outputs which are accessible to non-technical audiences
- Be a great advocate for Data Science, promoting our services to Co-op colleagues

Data Science Pitfalls - Blind Leading the Blind?

Data Scientist

Telegraph Media Group

London, GB

Requirements:

- Robust experience in SQL (Hive QL or any other flavour) with databases spanning tens of millions of records
- Solid fundamental understanding of statistics, and experience with statistical analysis using R or Python from university degree or previous role
- Be knowledgeable about the digital ecosystem
- You are a communicative person that values building strong relationships with colleagues and stakeholders and have the ability to explain complex topics in simple terms
- A self-starter who's comfortable working autonomously
- Solid commercial skills and business awareness

To be successful in this role, you will need:

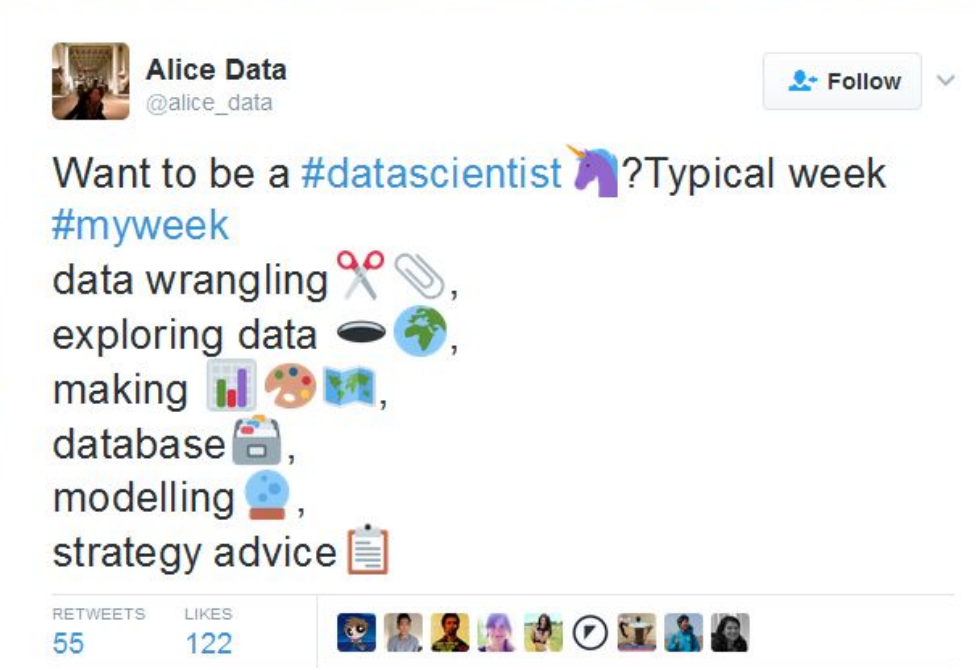
- Outstanding logical thinking, a very strong background in a quantitative discipline such as mathematics,physics, engineering, operations research, data science or similar, and a scientific approach to problem solving
- Demonstrate ability to manage stakeholder, show diplomacy and be an effective influencer
- Good technical skills (at least some practical experience with statistical/pattern recognition/machine learning methods)
- A sharp focus on getting results and insights using the most practical and fastest approach (even if it's not the most challenging or interesting for you)
- Ability to perform in a high pressure, fast paced environment and be comfortable with a high level of ambiguity
- Ability to communicate complex concepts concisely and clearly, including to senior level executives
- High energy and a can do, no-excuses attitude
- Curiosity and passion for making an impact
- Good interpersonal skills

Behavioural Competencies:

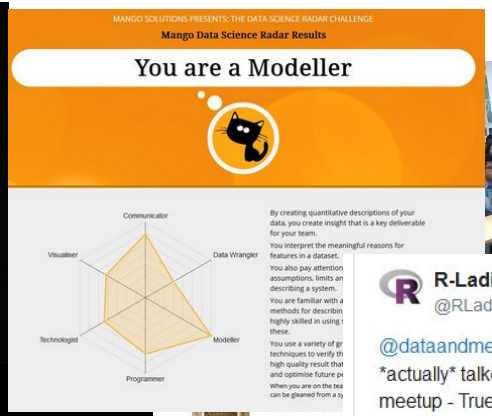
- Bright, driven, love challenges and want to change the editorial world

Data Science is awesome!!!

Data Science is awesome!!!



Data Science is awesome!!!



Alice Data @alice_data · Jan 9

❤️ documentation, schemas and beautiful code - what #datascientistlove ! 💻🦄

R-Ladies London @RLadiesLondon — Follow

@dataandme @lawscholar (Thanks!!) Exactly! @alice_data *actually* talked about #Rstats for over TWO hours straight last meetup - True R ❤️ !

Alice Data @alice_data — Follow

Want to be a #datascientist🦄?Typical week #myweek
data wrangling ✂️📎,
exploring data 🕳🌍,
making 📊🎨🗺,
database 🗃,
modelling 🔮,
strategy advice 📋

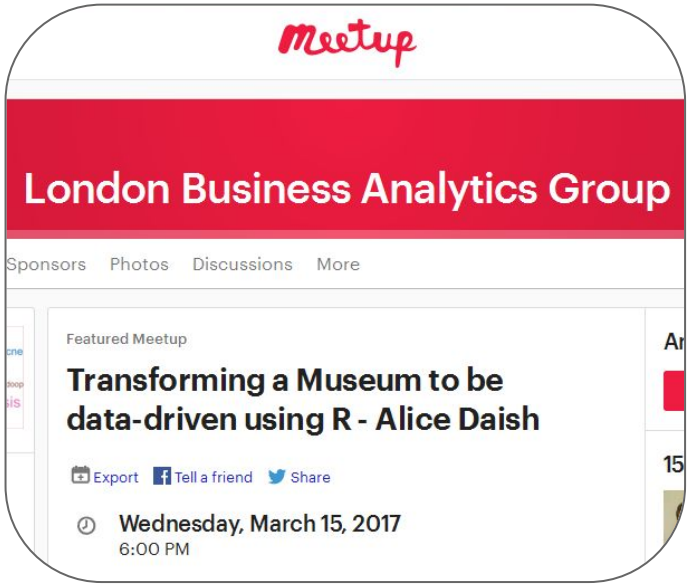
RETWEETS 55 LIKES 122

@britishmuseum - I had a super fun day learning #d3js & updating our data sources list #datascience

Alice Data @alice_data — Follow

Presented at @turinginst Both NEW #datascience initiatives! hope we can work & grow together! #DataScienceLeadership

Transforming a museum to be data-driven



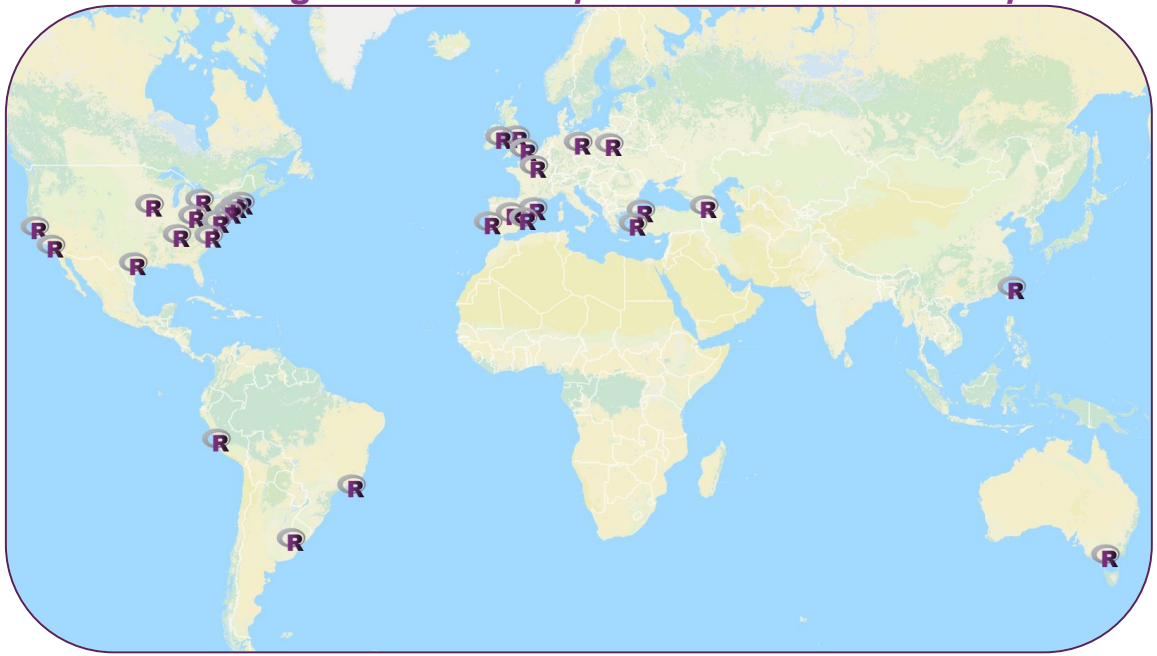
R-Ladies

rladies.org/

@RLadiesGlobal

@RLadiesLondon

Promoting Gender Diversity in the Global R community



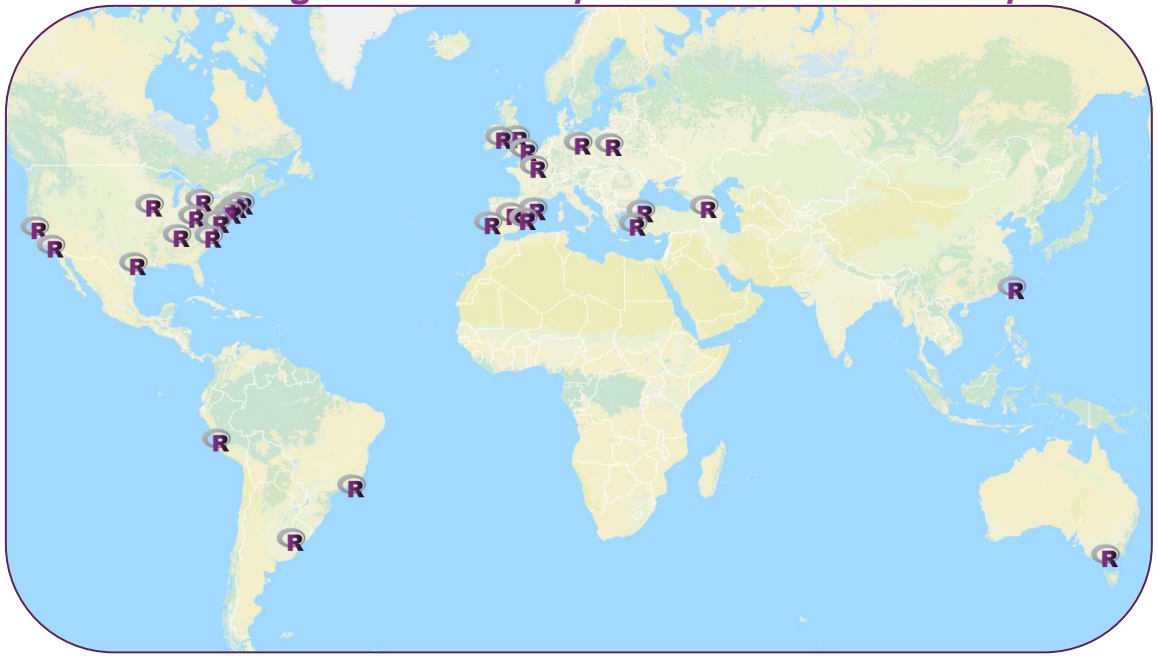
R-Ladies

rladies.org/

@RLadiesGlobal

@RLadiesLondon

Promoting Gender Diversity in the Global R community



@RLadiesLondon

@alice_data
@analyticspanda

#RLadies

Thank you!

london@rladies.org

Abstract

Demystifying Data Science

We all know the description of the Data Scientist being “The Sexiest Job of the 21st Century”, but do you have a clear understanding of what Data Science actually is?

In this talk, Data Science practitioners Alice and Chiin will present experience-based insights that enable clarity about this undoubtedly red hot but nebulous trend, by sharing coherent definitions, frameworks, and facts that help clear up some of the key areas of confusion. They will outline their view of Data Science best-practices, common misconceptions, pitfalls, and practical tips for those looking to make a career transition. This talk will ultimately deliver awareness about the reality of the Data Science profession, and help demystify the hype.

Alice Daish

Alice Daish is the Data Scientist at The British Museum focusing on making the museum data-driven by 2018. Co-Founder of R- Ladies Global and mentor at R-Ladies London. Registered Scientist previously trained in ecology and quantitative biology. Interests include R and data science, datafication, data analysis, predictive modelling, data visualisation, data communication, gender diversity in STEM.

Chiin-Rui Tan

Chiin is the Head of Data Science at the Foreign & Commonwealth Office, leading the establishment of a new data science capability. She has an Economics degree from the University of Cambridge, and interest in applied microeconomics topics. Founder & Co-Founder of R-Ladies London & Global, Chiin is a keen R evangelist and active promoter of computing to underrepresented minorities. Professional specialisms include data-centric leadership, econometrics, digital analytics, visualisation, data democratisation, and strategic insight.

END