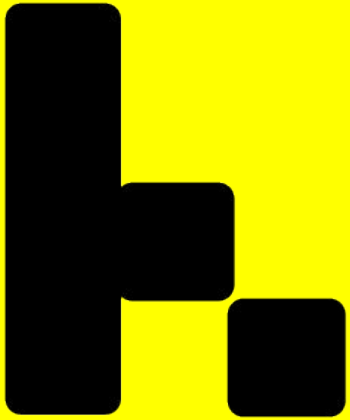


half stack data science

(Three) lessons from the deep end of data science

David Asboth & Shaun McGirr

Welcome

Team behind Half Stack Data Science podcast

Enable smarter decisions, faster at Cox Automotive UK

Here tonight thanks to General Assembly London!

The plan

But please interrupt any time!

1. What is data science?
2. Why "half stack" data science?
3. Lessons to apply tomorrow
    a. Find the right stakeholders
    b. Build what nobody asked for
    c. Avoid analysis paralysis
4. How to learn more
5. Q&A

WOE is DS

Anyone game?

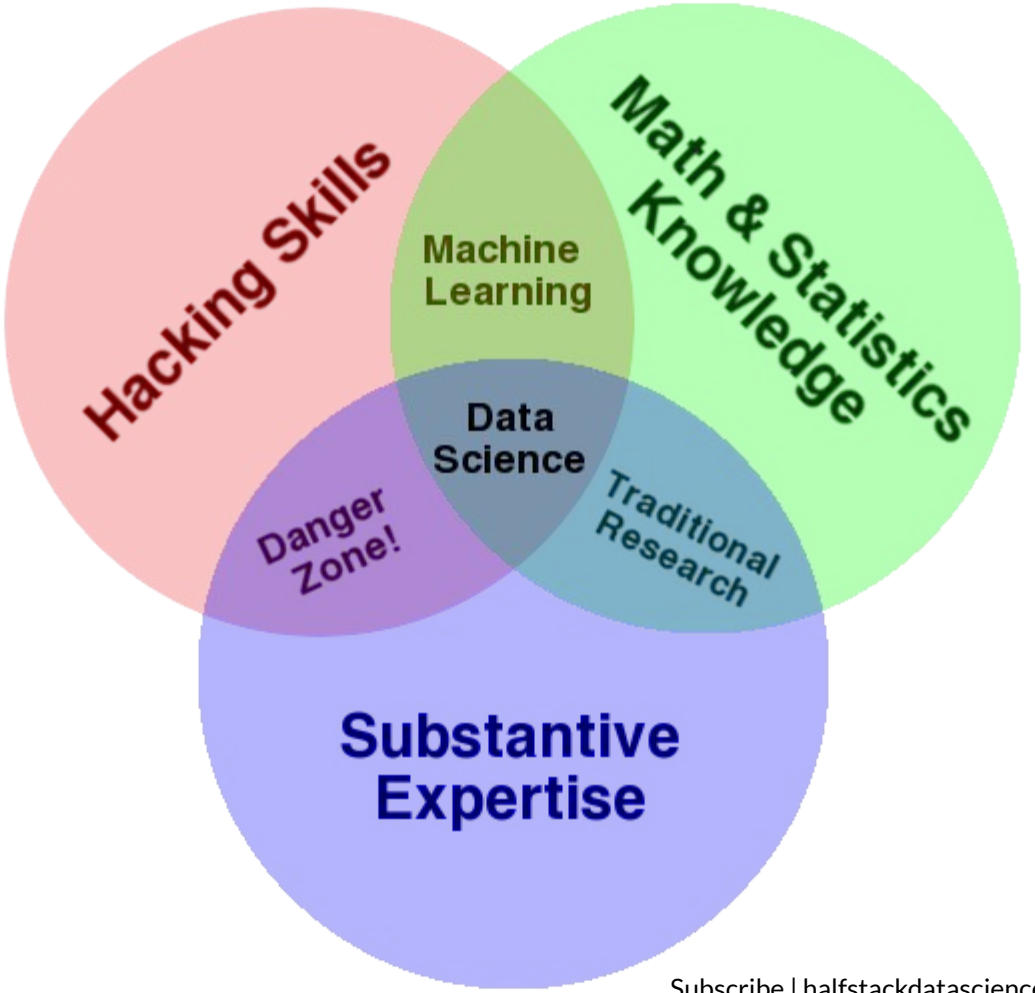
Time for *that* venn diagram

Drew Conway gets a lot of hot takes

But this is actually pretty good i.f.f. you think of data science as a **team sport**

There are no unicorns, at least not in the places where 90% of data science is happening right now

WE CANNOT ALL BE FACEBOOK / GOOGLE / STITCHFIX / AIRBNB



Why can't we all be Facebook?

Because in half-stack world:

- Data often collected by accident
- Colleagues usually not techy
- Questions are ambiguous
- Success criteria undefined
- Interpretability matters

Takeaways:

- Data scientists need to be pragmatic problem solvers, not PhD statisticians
- This is the majority of data science (despite what you may read online)

Data science process (the 3 Ps)

Proof of concept

Prototype

Production

Data science hats

Statistician - for avoiding unwarranted conclusions

Scientist - for thinking about hypotheses to test, not going in blind

Customer - to ensure you solve the right problem in the right way

Developer - to stay D.R.Y. and facilitate move from PoC to production

Lesson 1: find the right stakeholders

Why this matters

Stakeholder is anyone who **wants** you to do work

You may **want** to do that work

Question: how to work out what organisation **needs** most?

In half stack world YOU must answer this question yourself

What you can do tomorrow

Very simple way to play matchmaker

Don't necessarily write this on a public whiteboard!

	Stakeholder wants	**Stakeholder doesn't want**
We want	Actively pursue	Cautiously nurture
We don't want	Enthusiastically divert	Politely ignore

Concrete example

Sales drive a lot at Cox Automotive

One sales team focus specifically on vehicle buyers

Most sophisticated tools we've delivered were for them but...

Politely ignored -> Divert (to BI) -> Nurture curiosity -> LOVE

Takeaway: do initial cut but continue reevaluating "matches"

Lesson 2: build things no one asked for

Why this matters

Innovation requires iteration AND failure

Your job to identify high value work… nobody in the business knows what you do or how to ask for what they really need

Did you get in to data science to satisfy curiosity?

Endlessly analysing A/B test results may not cut that mustard

What you can do tomorrow

Hackathons/dedicated R&D time

Keep a catalogue of PoC to show (even months later)

Attend stakeholder meetings to listen for underlying problems

Ruthlessly show off your work!

Concrete example

Classic half stack problem: who could be our customers?

- Lead generation can be very time consuming, manual
- Ideally you've already joined customer data across silos

Our solution:

- Hackathon built a generic fuzzy matching algorithm
- Used open data to find all the potential customers

Takeaway: building what nobody asked for can pay off

Lesson 3: avoid “analysis paralysis”

The dreaded email...

From: CEO

Hi Shaun,

I've always wondered whether we sell more ice cream on days following a hot day or days following a cold day.

I'm presenting at the Ice Cream Producers of GB annual conference tomorrow and it would be great to have some fresh insights from our plentiful data!

Best,
Your CEO

Why this matters

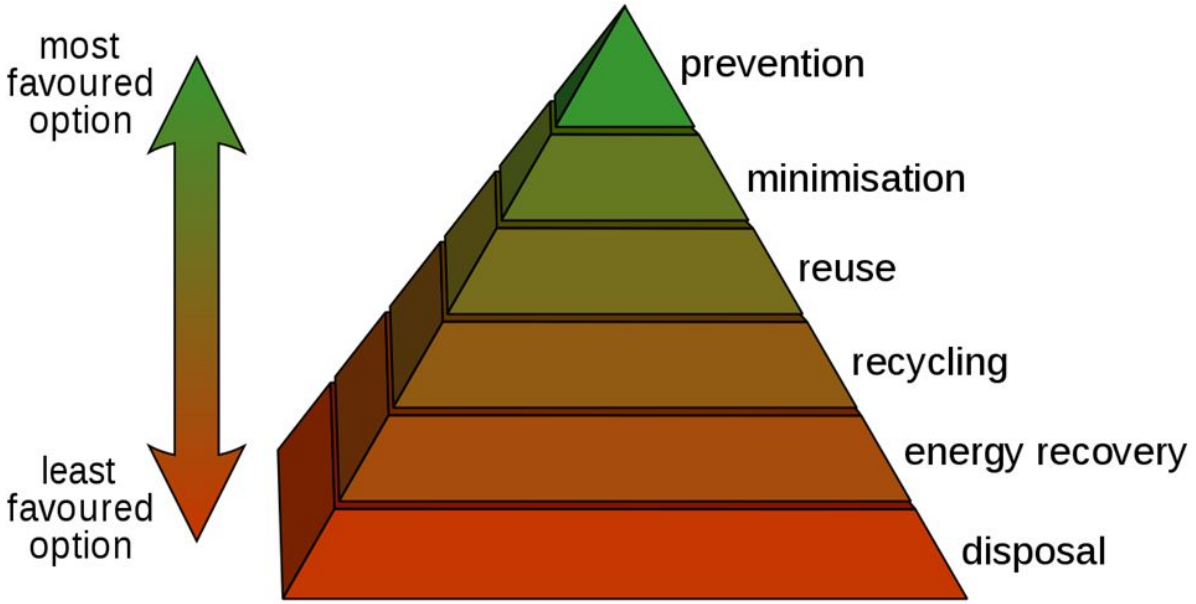
In half stack world you will experience "drive bys"

Senior stakeholders asking questions that sound important

Highest risk of **analysis paralysis** happens here because not at all clear whether any action would (or could!) follow

Can't say no...

Drive bys: waste management



Concrete example

Asked to compare vehicles sold in Dec with those sold in Jan

- then AM vs PM
- then Location A vs Location B
- then Channel 1 vs Channel 2

Given who asked the question, prevention not an option

Invested in better waste management with each ask

Takeaway: proactively manage down risk of analysis paralysis

Lesson 3+1: half stack data science is more than technical skills

What are these other skills you need to succeed?

According to Prof Roger Peng's *The Tentpoles of Data Science*

"Data Science is

1. the application of **design thinking** to data problems;
2. the creation and management of **workflows** for transforming and processing data;
3. the negotiation of **human relationships** to identify context, allocate resources, and characterize audiences for data analysis products;
4. the application of **statistical methods** to quantify evidence; and
5. the transformation of data analytic information into coherent **narratives and stories**"

We agree all of these are needed for success, but most courses only have time to teach 2 & 4

Where can you learn more?

Most data science courses don't have time for these lessons

Some believe you must learn these lessons on the job

We already did, so you can get a head start...

Join our journey

We're only getting started!

Podcast

halfstackdatascience.com

More events, more interactive

Sign up to list in survey

Online

@davidasboth

@shaunmcgirr

Q&A

half stack
data science