

So you want to be a data science consultant (or hire one)?

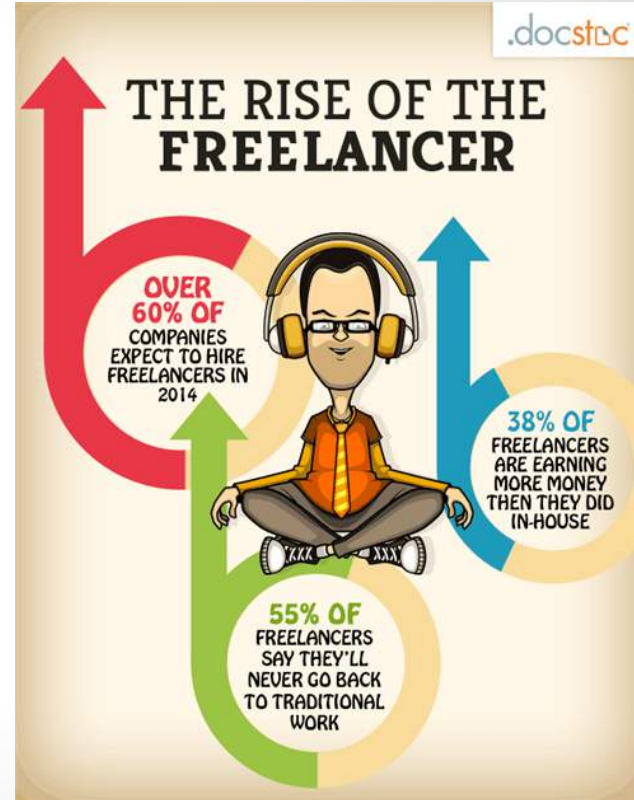
Radim Řehůřek, Ph.D.
rare-technologies.com



Who am I?

- PhD in CS (scalability of semantic text analysis)
- open source (mostly Python world)
- worked for a search engine company as research lead
- freelancing 2009+
- founded RaRe Technologies 2014

Why freelancing?



#1 Beware of hidden costs of consulting



- expect to bill ~40% hours: not 1:1 salary!
 - rest: keeping up with tech, research
 - prospecting, accounting, proposals, marketing
 - internal trainings, code reviews
 - +open source “free support”...
- **charge for business analysis!**
 - R&D vs. trainings, architecture reviews and design
 - project-based utopia



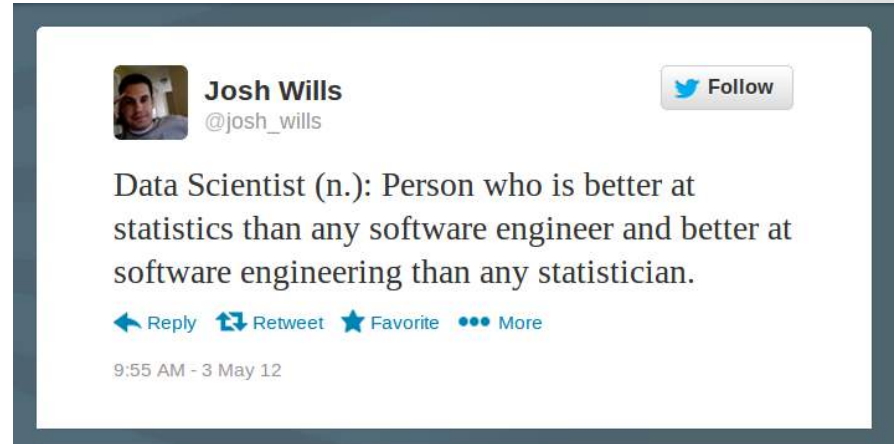
RaRe Technologies Ltd.





What is “data science”?

- engineering
- math & statistics
- visualization
- business insights



#2 Nobody cares about expert lingo or academic ivory towers



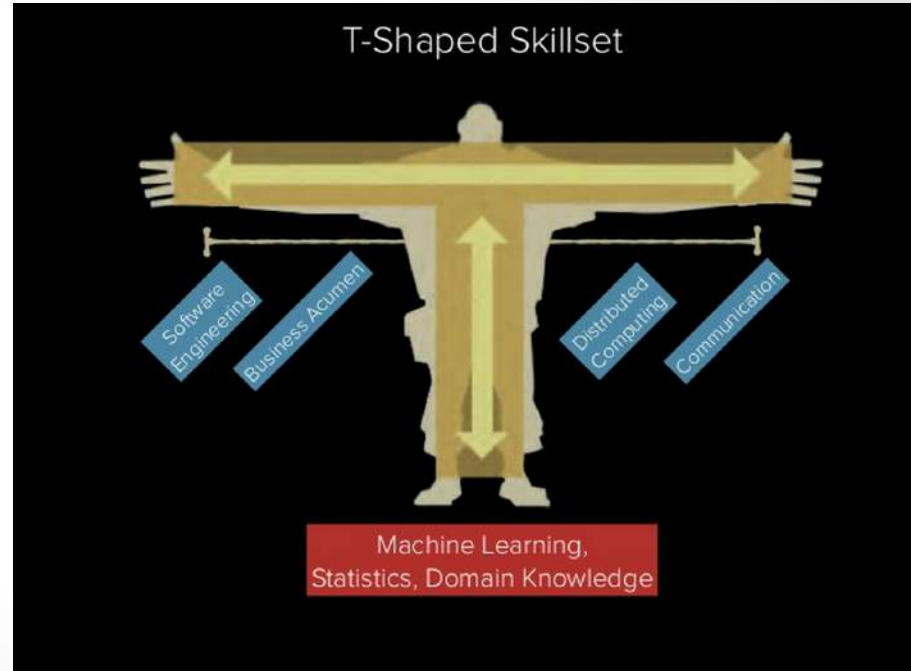
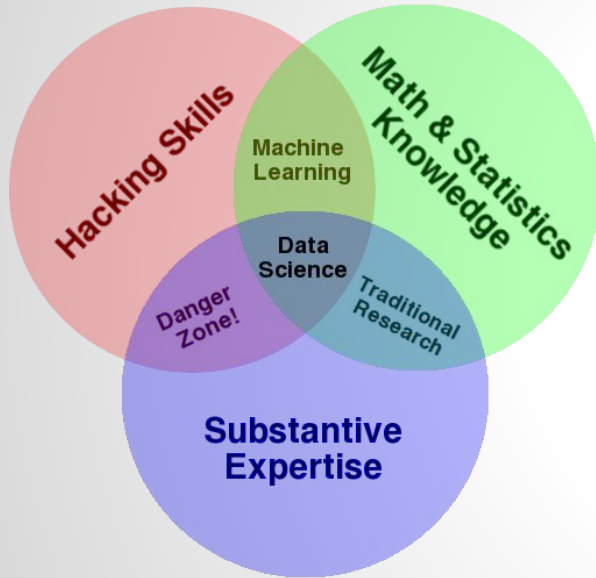
“The purpose of computation is insight, not numbers.”

Richard Hamming

- clear processes
- always eye toward business impact
 - vs. technologist mindset and tinkering for fun



Lots of existing resources



<http://www.slideshare.net/ryanorban/how-to-become-a-data-scientist>

#3 Prepare for difficult, poorly scoped “challenges”



Good

- get experience w/ verticals, overall “market feel”
- learn how to scope, price, analyze
- develop ancillary skills: accounting, taxes, IP law...

Bad

- uncertainty not for everyone
- transfer of domain expertise difficult
- expectations to master all new tech instantly, constantly

Cambrian explosion





IP[y]: IPython
Interactive Computing



#4 Take “progress” easy



mat kelcey @mat_kelcey · May 20

pretty much every paper i've ever read....

method	score
previous approach	good
our approach	almost as good
our approach + last minute hack	slightly better than good!



[View photo](#)



#5 Master your tools

“Do one thing and do it well.”

Doug McIlroy

"Every program attempts to expand until it can read mail. Those programs which cannot so expand are replaced by ones which can."

Zawinski's law of software development



#6 Expect non-glamorous work

1. problem discovery, scoping, APIs
 2. data collection, cleaning
 3. evaluation, reporting, visualizations
 4. coding, integration, automation
- vs. what novices expect: models, math, algorithms
- vs. suits & armchair data philosophers

#7 Avoid wasting time on “unrealistic” clients



Good

- small/medium companies with established market fit
- reporting to CEO / project lead
- ideally longer term collaboration: sense of ownership

Bad

- enterprise “demonstrate innovation to execs” projects
- start-ups “golden goose... on \$5k budget!”
- veiled attempts at hiring

How?

- Price yourself out of “I have an idea” range
- Manage expectations clearly, terminate dead ends quickly

#8 Be wary of working for equity

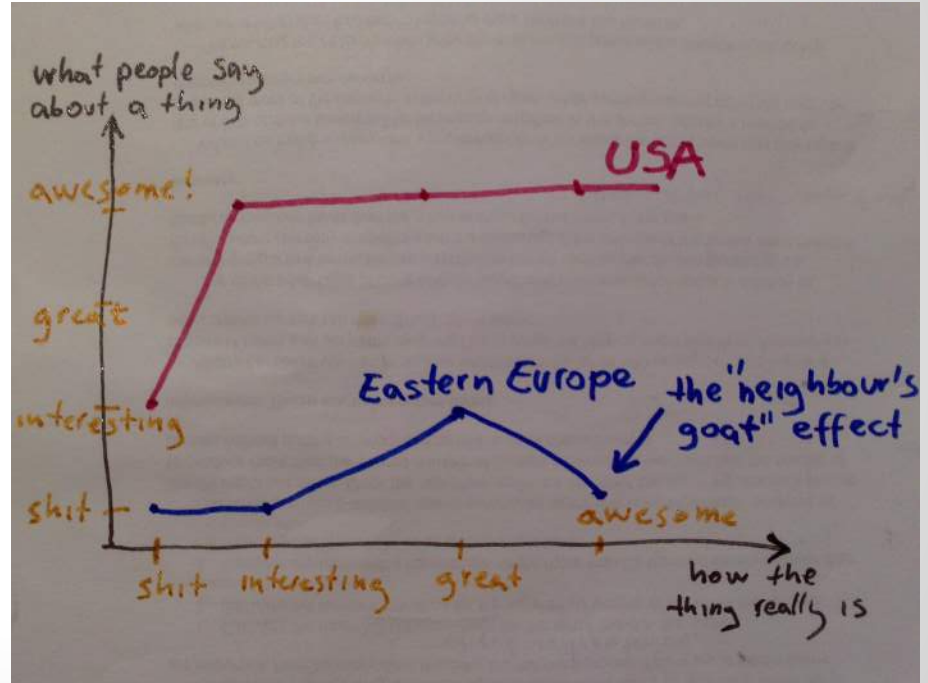


- too risky
 - do you really want to be an early **investor**?
- core product tech designed and developed by an external consultant = bad idea all around
 - esp. remote consulting
- passion mismatch



#9 Learn to navigate hype

- It's **your** responsibility to manage expectations and advise on tech stack
- Diplomacy needed: "big data" & co
- Cultural differences



People



RaRe Braintrust





#10 Scaling consulting sucks

Good

- people help scaling up
- take on larger projects with larger teams

Bad

- people don't help scaling up
- fresh grads & Coursera instant experts

...Teach yourself data science in ten years

RaRe Incubator & Internships



Academic Incubator

We know good data scientist are hard to find – so we grow them.

Students

Want to gain industry experience?

We offer practical thesis topics in data mining. For talented students worldwide we have paid internships.

[JOIN INCUBATOR](#)

Universities

Want to partner up?

We mentor and review students' theses. We also partake in official industry partnerships for large consortial projects.

[HAVE A PROJECT?](#)





Bonus tip #11: take time to appreciate your luck

You have the luxury of choice, live in the middle of a revolution, amazing times!

Radim Řehůřek

RaRe Technologies Ltd.

<http://rare-technologies.com>

(previously radimrehurek.com)

@radimrehurek

Blog (mostly tech, machine learning):

<http://radimrehurek.com/blog/>