

The Awesomeness of Scientific Python and Effective Computation in the Physical Sciences

Anthony Scopatz

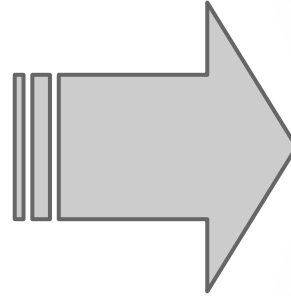
UT Austin Astronomy Graduate Student Postdoc Seminar

2015-05-01



THE UNIVERSITY OF
CHICAGO





UNIVERSITY OF
SOUTH CAROLINA

THE UNIVERSITY OF
CHICAGO



<http://physics.codes/>

O'REILLY®

Effective Computation in Physics

FIELD GUIDE TO RESEARCH
WITH PYTHON



Anthony Scopatz &
Kathryn D. Huff

<http://physics.codes/>

O'REILLY®

Effective Computation in Physics

FIELD GUIDE TO RESEARCH
WITH PYTHON



Early Release
RAW & UNEDITED

Anthony Scopatz &
Kathryn D. Huff

Get 40% off of the
print book and 50%
off of the ebook
version of this book
by entering discount
code **AUTHD**.

<http://physics.codes/>

O'REILLY

Effective Computation in Physics

FIELD GUIDE TO RESEARCH
WITH PYTHON

Early Release
RAW & UNEDITED

Anthony Scopatz &
Kathryn D. Huff

Get 40% off of the
print book and 50%
off of the ebook
by entering discount
code **AUTHD**.

<http://katyhuff.github.io/>

[@katyhuff](https://twitter.com/katyhuff)





Disclaimer: opinions mine.

Why?

“If there's a book that you want to read, but it hasn't been written yet, then you must write it.”

— Toni Morrison

“If there's a book that you want to read, but it hasn't been written yet, then you must write it.”

— Toni Morrison



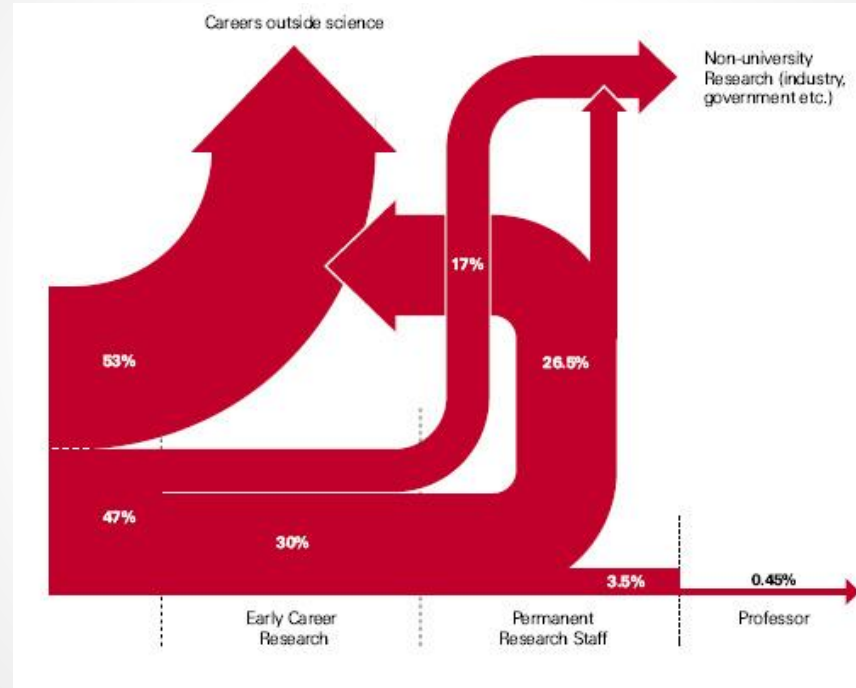
Staffing Problem - Too Few?

Hits different domains differently

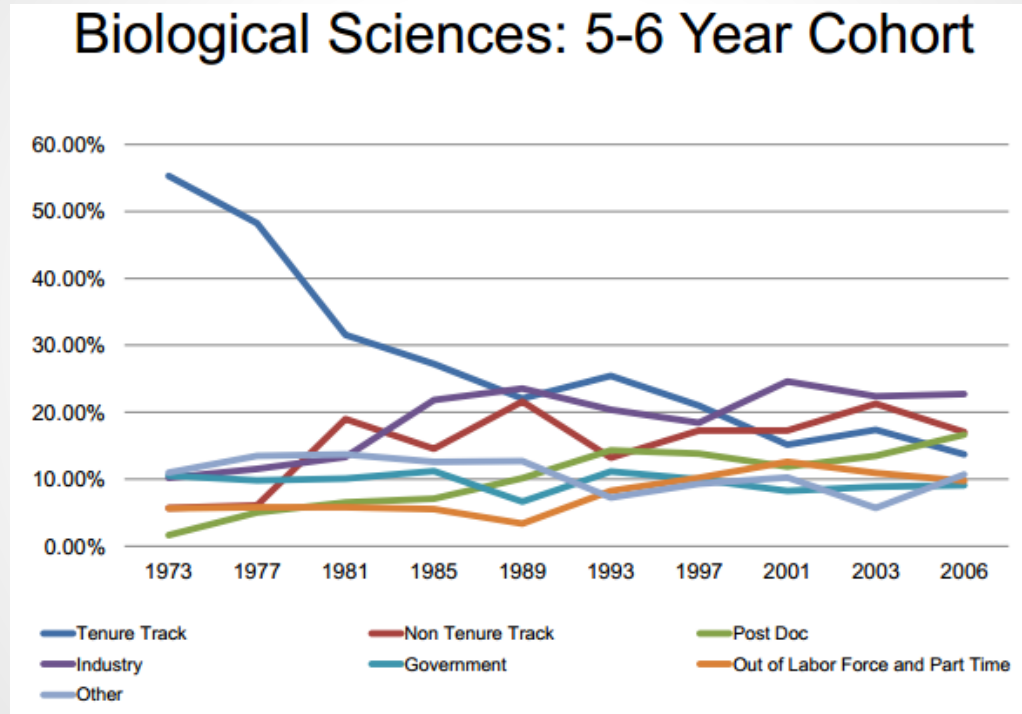
Number H-1B visas remain debate topic

But really, new graduates don't know how to use a computer (as an experimental device...)

Staffing Problem - Too Many?



Staffing Problem - Too Many?



<http://www.theatlantic.com/business/archive/2013/02/the-phd-bust-americas-awful-market-for-young-scientists-in-7-charts/273339/>

Proposition

scientist + software dev \neq scientific software

**Does anyone think
computation has not
helped physics?**

What's Awesome?

Dark Sky Simulations

Sam Skillman et al.

<http://arxiv.org/abs/1407.2600>

http://darksky.slac.stanford.edu/halo_world.html

Mainstream languages are finally becoming usable by mortals.



Open Source is Winning

And if it isn't open it isn't science...

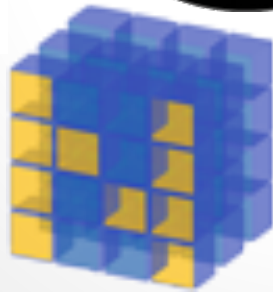
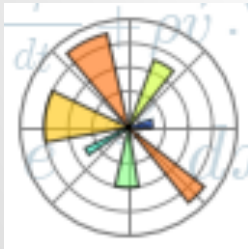
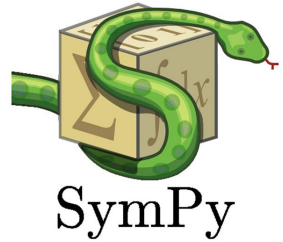
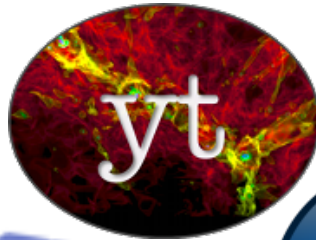
Open Source is Winning

And if it isn't open it isn't science...

...it is probably technology, though.

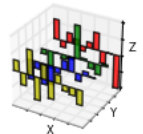
The only general purpose language that has a robust scientific ecosystem and engaged community is Python.

IP[y]:



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



ATX is a hotspot for scientific computing.

Awesome but Unleveraged

Exascale

Awesome but Unleveraged

Exascale

Haskell et al.

Awesome but Unleveraged

Exascale

Haskell et al.

Containers

What's Less Awesome?

What are we dealing with?

- Diversity
 - open source participation
 - code of conduct
 - identity

What are we dealing with?

- Diversity
 - open source participation
 - code of conduct
 - identity
- Legacy & Maintenance Codes

What are we dealing with?

- Diversity
 - open source participation
 - code of conduct
 - identity
- Legacy & Maintenance Codes
- Data Science is the new Quant. Finance
 - But with less money and occasional morals

Diversity

What does diversity mean?

In STEM we tend to only talk about gender diversity.

What does diversity mean?

In STEM we tend to only talk about gender diversity.

This is bad.

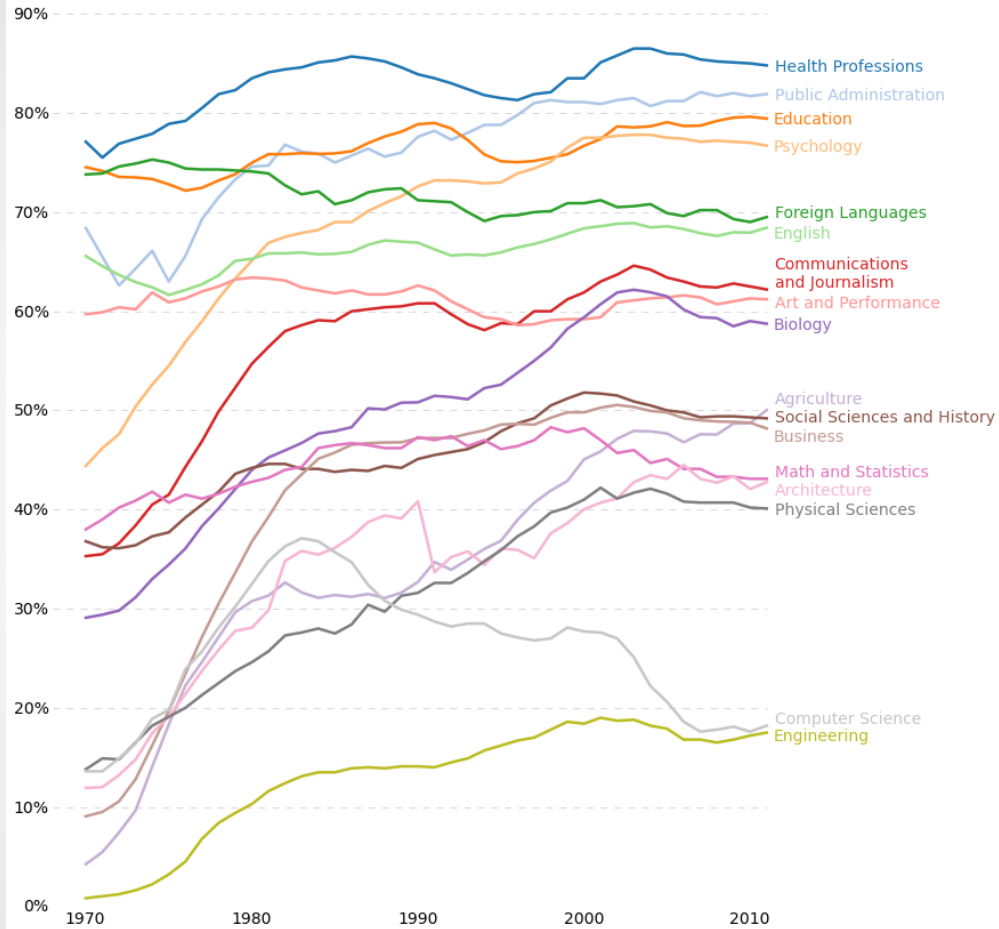
What does diversity mean?

In STEM we tend to only talk about gender diversity.

This is bad.

After decades of effort, things are getting better, except...

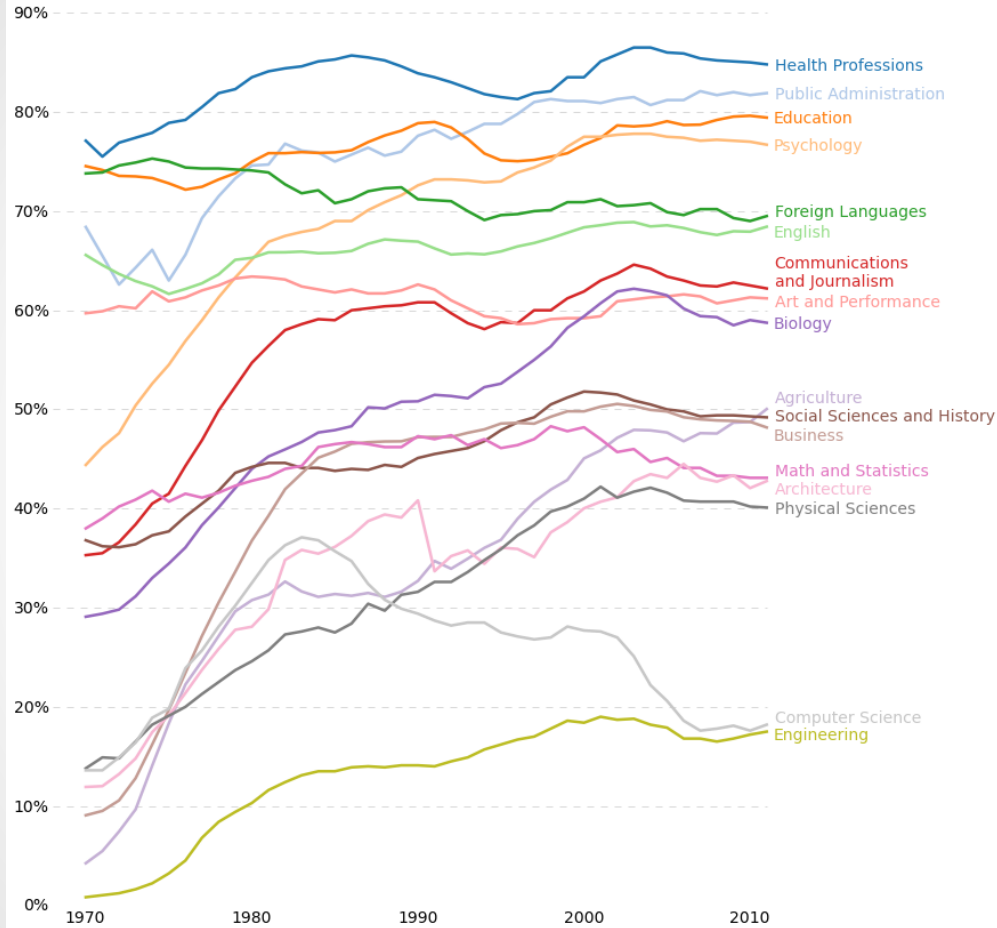
Percentage of Bachelor's degrees conferred to women in the U.S.A., by major (1970-2012)



Data source: nces.ed.gov/programs/digest/2013/menu_tables.asp
 Author: Randy Olson (randalolson.com / [@randal_olson](https://twitter.com/randal_olson))

Note: Some majors are missing because the historical data is not available for them

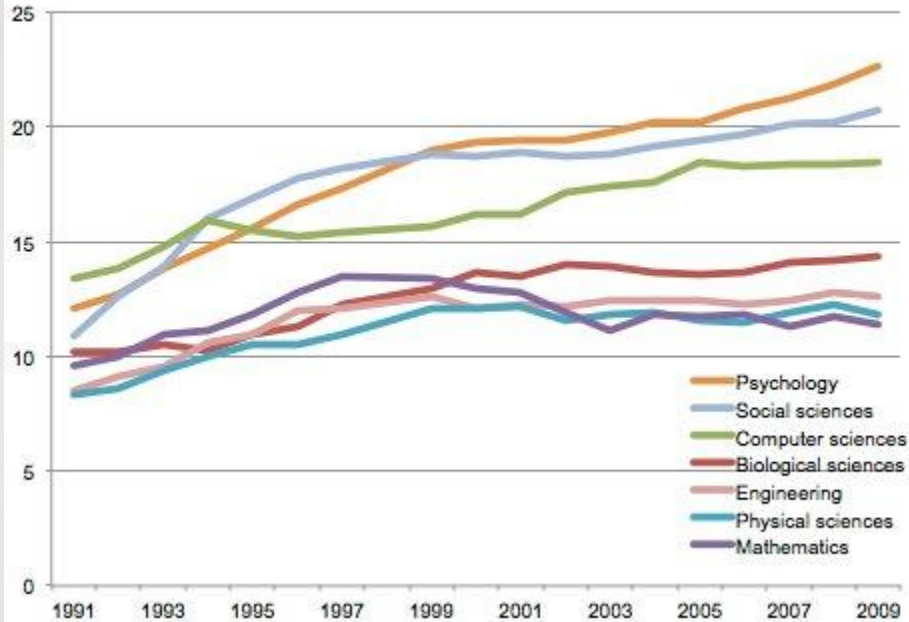
Percentage of Bachelor's degrees conferred to women in the U.S.A., by major (1970-2012)



Data source: nces.ed.gov/programs/digest/2013/menu_tables.asp
Author: Randy Olson (randalolson.com / [@randal_olson](https://twitter.com/randal_olson))
Note: Some majors are missing because the historical data is not available for them

Computational Science looks more like the Computer Science curve than it looks like the Physical Sciences curve.

Percent of B.A.'s in a Field Received by Historically Underrepresented Racial/Ethnic Minorities



“For whatever reason, there have been greater gains by underrepresented minorities in computer science and engineering than there have been for women, although in both cases blacks, Hispanics and Native Americans still receive a small share of total bachelor’s degrees awarded.”

[Source:](#) "Women, Minorities, and Persons with Disabilities in Science and Engineering," National Science Foundation. Underrepresented minorities refer to blacks, Hispanics and Native Americans.

Non-privileged Participaction

Engagement with open source by underrepresented classes approaches zero for core development.

Non-privileged Participaction

Engagement with open source by underrepresented classes approaches zero for core development.

Delicate issue:

- People aren't badges.
- Open source is a waste of time?

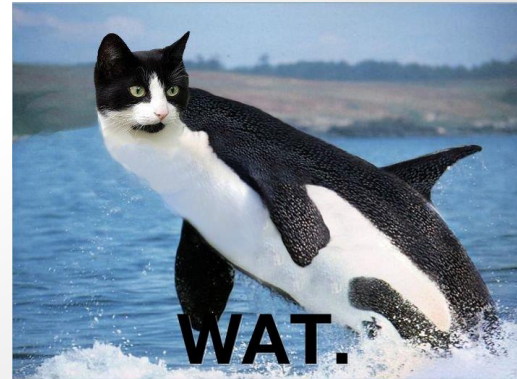
Identity

Codes of conduct can help drive conference attendance (e.g. PyCon, SciPy). Some people oppose the legalistic structure on principle.

Identity

Codes of conduct can help drive conference attendance (e.g. PyCon, SciPy). Some people oppose the legalistic structure on principle.

Not everyone who develops software thinks of themselves as a programmer.



Legacy & Maintenance Software

Technical Debt

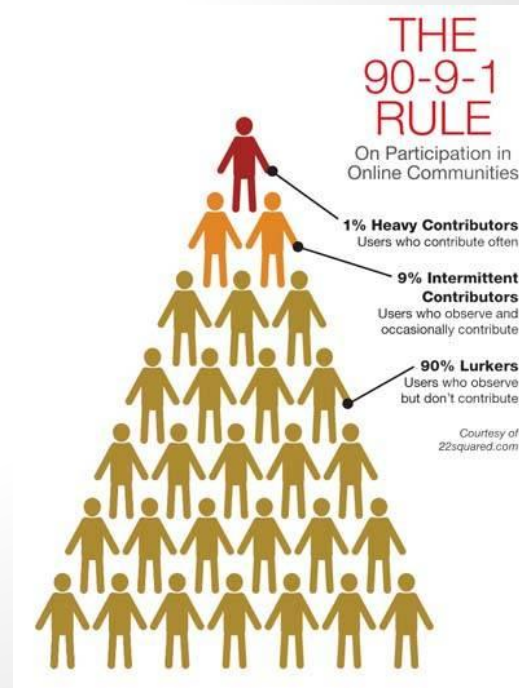
Computational Physics is paying off technical debt from the 1950s.

Technical Debt

Computational Physics is paying off technical debt from the 1950s.

Current core projects lack large support bases.

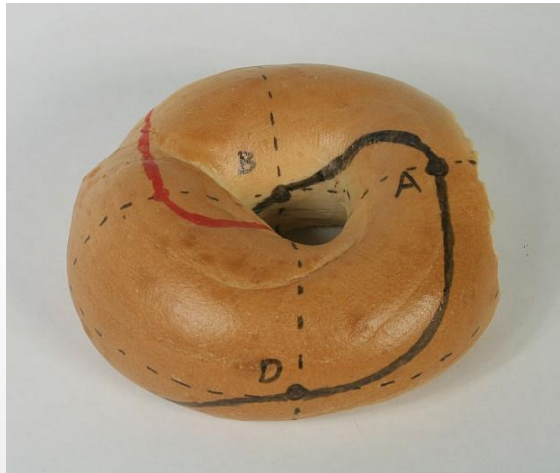
90-9-1 rule or worse.



Data?
Science?

Challenge Problem 1: Bagelocation

Statement: Find the point on Earth's surface that is furthest from any bagel.



Challenge Problem 1: Bagelocation

Statement: Find the point on Earth's surface that is furthest from any bagel.

Lemma 1: This problem is solvable.

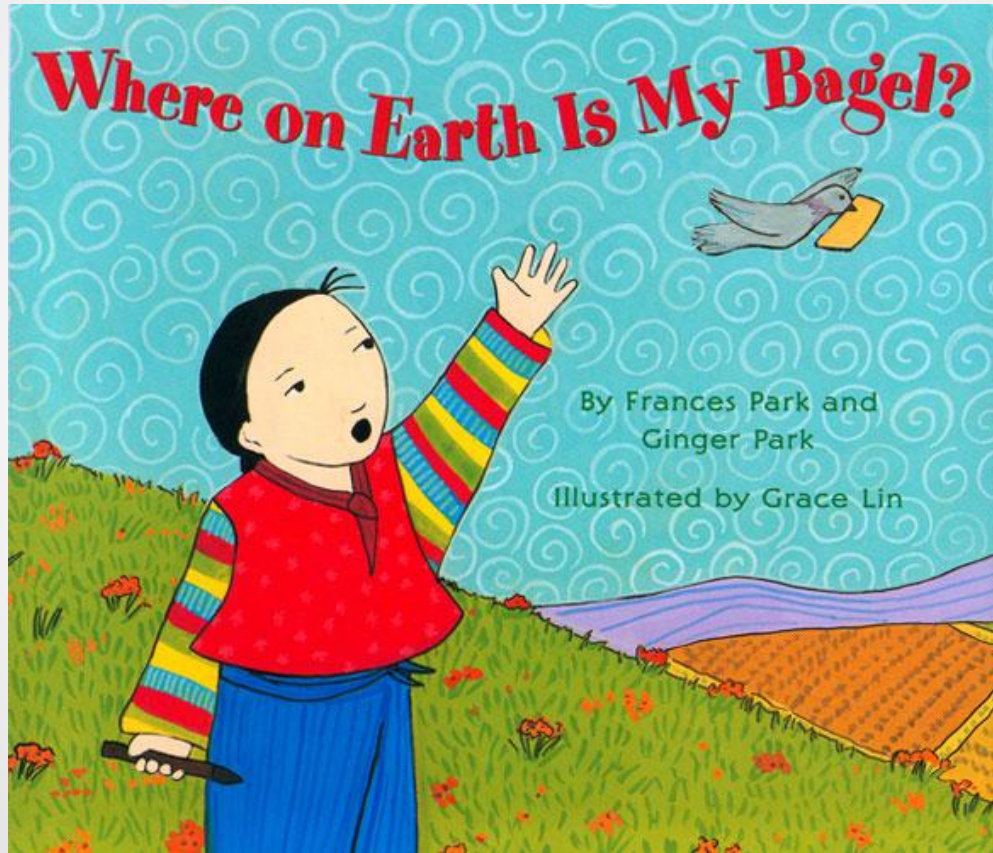
Challenge Problem 1: Bagelocation

Statement: Find the point on Earth's surface that is furthest from any bagel.

Lemma 1: This problem is solvable.

Lemma 2: We lack needed data.

Challenge Problem 1: Bagelocation



Challenge Problem 2: Methodology

Suppose Data Science is a science in the traditional sense.

What are the hypotheses that are being tested, affirmed, disproven, and reproduced?

Challenge Problem 3: Truth

How does data science derive truth?

S - reproducible observations

T - craft has desirable aesthetics

E - better match objectives given constraints

M - propositional logic

**Organizations To Be
Excited About**



<http://numfocus.org/>



<http://software-carpentry.org/>



<http://datacarpentry.org/>



**AFRICAN INSTITUTE
FOR
MATHEMATICAL SCIENCES**

<http://physics.codes/>

O'REILLY®

Effective Computation in Physics

FIELD GUIDE TO RESEARCH
WITH PYTHON

Early Release
RAW & UNEDITED

Anthony Scopatz &
Kathryn D. Huff

Get 40% off of the
print book and 50%
off of the ebook
version of this book
by entering discount
code **AUTHD**.