Why Two Sigma Contributes to Open Source

Julia Meinwald
Two Sigma Investments
Disclaimer

This document is being distributed for informational and educational purposes only and is not an offer to sell or the solicitation of an offer to buy any securities or other instruments. The information contained herein is not intended to provide, and should not be relied upon for, investment advice. The views expressed herein are not necessarily the views of Two Sigma Investments, LP or any of its affiliates (collectively, “Two Sigma”). Such views reflect the assumptions of the author(s) of the document and are subject to change without notice. The document may employ data derived from third-party sources. No representation is made by Two Sigma as to the accuracy of such information and the use of such information in no way implies an endorsement of the source of such information or its validity.

The copyrights and/or trademarks in some of the images, logos or other material used herein may be owned by entities other than Two Sigma. If so, such copyrights and/or trademarks are most likely owned by the entity that created the material and are used purely for identification and comment as fair use under international copyright and/or trademark laws. Use of such image, copyright or trademark does not imply any association with such organization (or endorsement of such organization) by Two Sigma, nor vice versa.
Oh, hello!

Hello
my name is

Julia
How I got here
Can we really do that?

secret mysteries
What do the next 20 minutes hold?

- Why is OSS scary?
- Why is OSS worth it?
- Lessons learned
- How Two Sigma is trying to help
Two Sigma and Open Source: A Brief History

- **2012**: Preliminary OSS meetings
- **2013**: Open Source Committee Established
- **2014**: Beaker open sourced
- **2015**: Satellite open sourced
- **2016**: Cook open sourced, Flint open sourced, BeakerX pivot
- **2017**: Formal OSS approval process launched, IP Steering Committee formed, Phillip Cloud joins TS, Collaboration With Jupyter team, Waiter open sourced, Major commits to Apache Spark
- **2018**: Legal entity TSOS established, IP Steering Committee formed, Jeff Reback joins TS, Marbles open sourced, Afshin Darian joins TS
Why Contribute?
Why Contribute?

Reward

Risk

BAD code.
Why Contribute?

#1: Leveraging other people’s work
#2: Shaping the ecosystem
#3: Avoiding isolation
#4: Being cool
#5: Building your legacy
#6: Making the world a better place
I’m standing on the shoulders of giants!

Cool
#2 - Shape the Products/Ecosystem
#3 - Avoid Isolation
#4 - Be Cool
#5 - Build Your Legacy
#6 - Make the World Better
Lessons Learned
Lessons Learned: Beaker Notebook

Batteries included, so you can get your job done faster.
Lessons Learned: Beaker Notebook vs. Project Jupyter

Wow, you’re popular

Thanx
Lessons Learned: Beaker → BeakerX
Lessons Learned: BeakerX

DISCOVERABLE DATA

DATA ANALYSIS & MODELING

SCALABLE & DISTRIBUTED COMPUTE

PUBLICATION & COLLABORATION
Lessons Learned: Building Community is Hard

Can I come hang?

Sure!
Lessons Learned: The Urge to Merge
Lessons Learned: Satellite

Satellite
Lessons Learned: Spark
Lessons Learned: Flint

Flint
OSS, I need you!
A Day Without Open Source

pandas

$y_t = \beta x_t + \mu + \epsilon_t$

NumPy

jupyter

Apache Spark

git
We’re Not The Only Ones

Companies Using Linux

Public Digital Infrastructure is not Free, Nor Invulnerable

• Widely Used ≠ Widely Supported
• Infrastructure isn’t free!
• Need time and money to grow products
• New challenge
Bystander Effect
We Can Affect The Future
How Two Sigma Got Involved
Ways Two Sigma Is Helping

• **Hiring** OSS Engineers and consultants
• **Open Sourcing** tools we develop in-house
• **Contributing** to OSS projects we rely on
• **Supporting** OSS Startups
• **Building** community
Ways Two Sigma Is Helping

• **Hiring** open source engineers

ABOUT ME

- Li Jin (jcelloess)
- Software Engineer @ Two Sigma Investments
- Analytics Tools Smith
- Apache Arrow Committer
- Other Open Source Projects:
  - Flint: A Time Series Library on Spark
Ways Two Sigma Is Helping

- **Open Sourcing** tools we develop in-house

BeakerX  
Cook  
Git-meta  
Flint  
marbles  
Waiter
Ways Two Sigma Is Helping

• **Contributing** to OSS projects we rely on
Ways Two Sigma Is Helping

• **Supporting** Open Source Startups

URSA LABS
Innovation Lab for Data Science Tools
Ways Two Sigma Is Helping

• Building community

NUMFOCUS
OPEN CODE = BETTER SCIENCE
Thank You

julia.meinwald@twosigma.com

www.juliameinwald.com

@JuliaEve

twosigma.io

As users of open source software, we’re passionate about giving back to the community.

Flint
- 77 • 400
Flint is our project that provides time-series analysis in Spark
Learn More

git-meta
- 21 • 75
Our project that lets you effectively manage many repositories with git submodules.
Learn More

BeakerX
- 246 • 1817
BeakerX is a collection of kernels and extensions to the Jupyter interactive computing environment. It provides JVM support, interactive plots, tables, forms, publishing, and more.
Learn More

Marbles
- 3 • 25
Marbles is a Python unittest extension that allows test authors to write richer tests that expose more information on test failure.
Learn More