

Shared Research Computing Policy Advisory Committee

Spring 2019 Meeting
Thursday, April 25, 2019
10:00 a.m. – 11:30 a.m.

 COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee





Today's Agenda

Introductions

HPC Update

Foundations for Research Computing Update

RCEC Plans

Introductions

Everyone!

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee

HPC Update

Kyle Mandli, Chair of HPC Operating Committee

George Garrett, Manager of Research Computing, CUIT

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee

Topics

- Governance
- Support
- Yeti
- Habanero
- Terremoto
- Singularity



HPC Governance

- Shared HPC is governed by the faculty-led HPC Operating Committee, chaired by **Kyle Mandli**.
- The committee reviews business and usage rules in open, semiannual meetings.
- The last meeting was held on March 11, 2019. Next meeting will be in Fall 2019.
- **All HPC Users (Terremoto, Habanero) are invited.**



HPC Support Services

- Email
 - hpc-support@columbia.edu
- Office Hours
 - In-person support from 3pm – 5pm on 1st Monday of month
 - RSVP required (Science & Engineering Library, NWC Building)
- Group Information Sessions
 - HPC support staff present with your group



Cloud Computing Consulting

- Overview of features of cloud service providers (AWS, Google, Azure)
- **Cost estimates** and **planning workflow** for efficiency and/or price
- **Creation and initial configuration of images**, including software installation



Yeti Cluster – Retired

Publication Outcomes

- Research conducted on Yeti has led to over **60 peer-reviewed publications** in top-tier research journals.

Retirement

- Yeti Round 1 retired November 2017
- Yeti Round 2 retired March 2019

Habanero

Specifications

- 302 compute nodes (7,248 cores)
- 740 TB storage (DDN GS7K GPFS)
- 397 TFLOPS of processing power

Lifespan

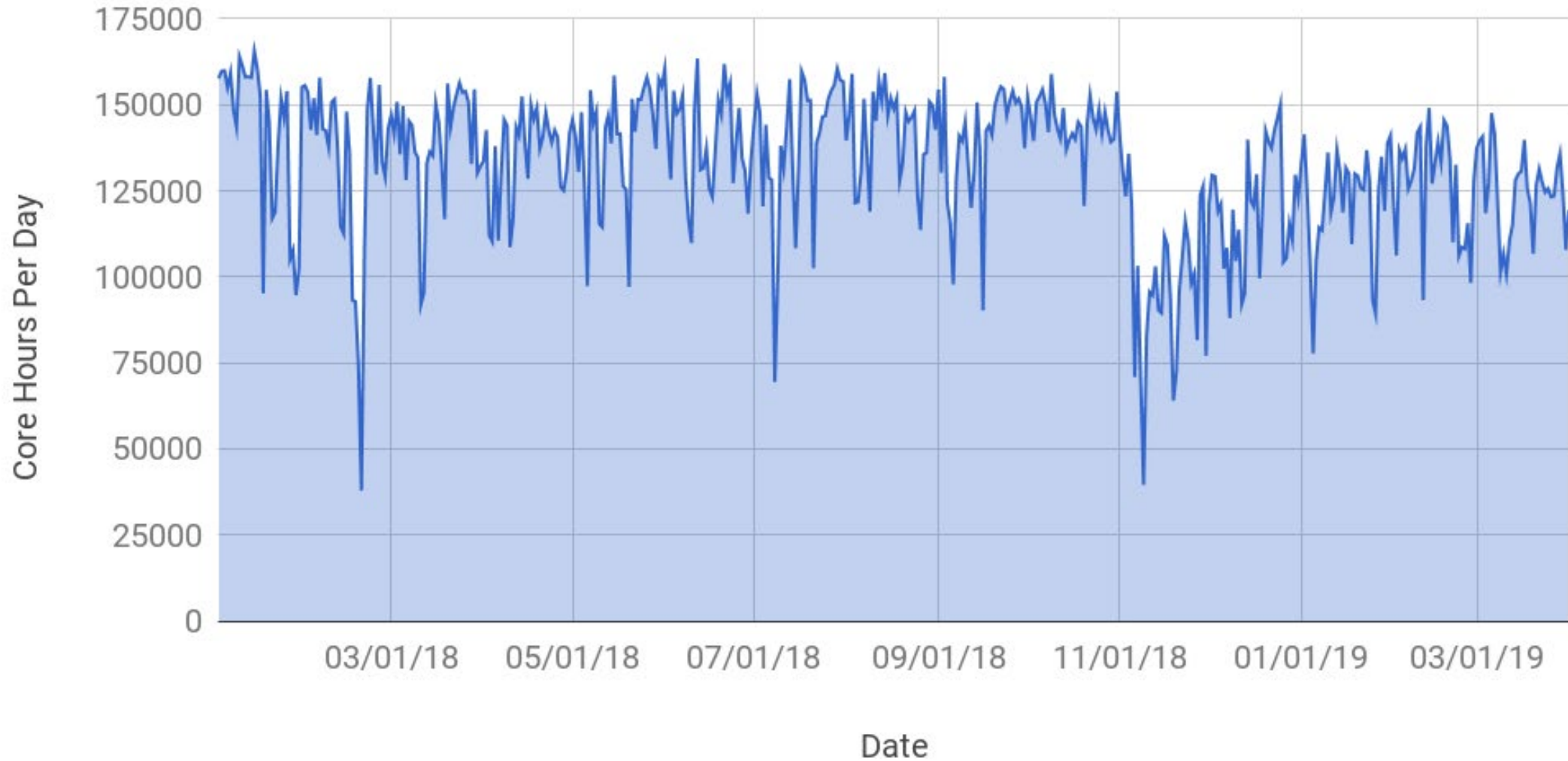
- 222 nodes expire 2020
- 80 nodes expire 2021



Habanero – Participation and Usage

- 44 groups
- 1,550 users
- 9 renters
- 160 free tier users
- Education tier
 - 15 courses since launch

Habanero – Cluster Usage in Core Hours



Terremoto

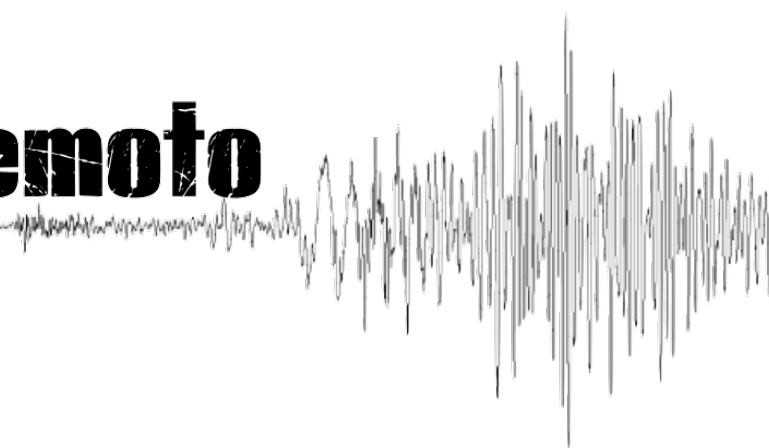
Launched in December 2018.

- 24 research groups
- 5 year lifetime





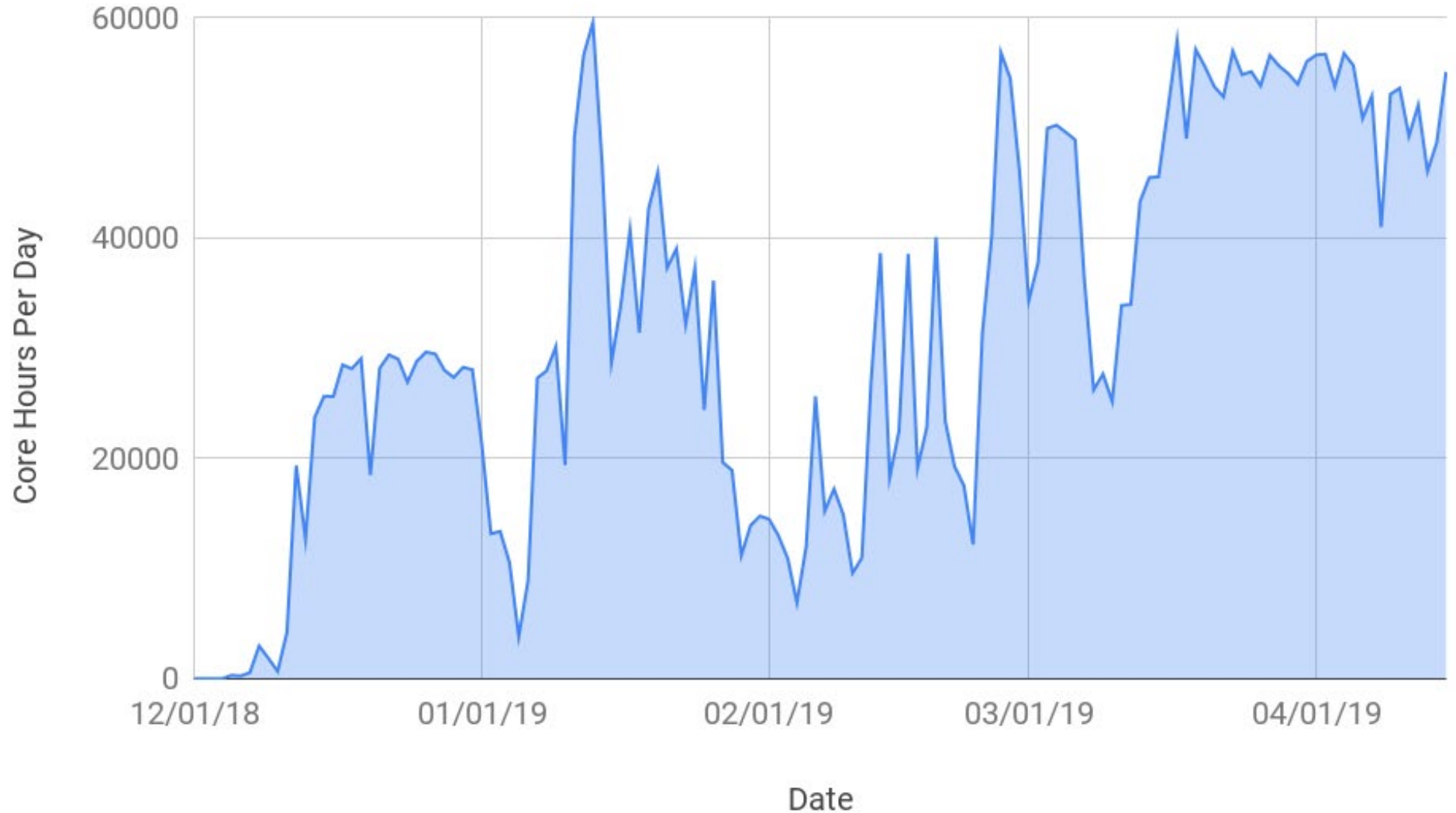
Terremoto



Specifications

- 110 Compute Nodes (2640 cores)
 - 92 Standard nodes (192 GB)
 - 10 High Memory nodes (768 GB)
 - 8 GPU nodes with 2 x NVIDIA V100 GPUs
- 430 TB storage (Data Direct Networks GPFS GS7K)
- 255 TFLOPS of processing power
- Dell Hardware, Dual Skylake Gold 6126 cpus, 2.6 Ghz, AVX-512
- 100 Gb/s EDR Infiniband, 480 GB SSD drives

Terremoto – Cluster Usage in Core Hours





Terremoto 2019 HPC Expansion Round

- **No RFP.** Same CPUs and GPUs as Terremoto 1st round.
- Purchase round to commence in **May 2019.**
- Go-live in late **Fall 2019.**

If you are aware of potential demand, including new faculty recruits who may be interested, please contact us at rsc@columbia.edu.

Singularity



- Easy to use, secure containers for HPC.
- Enables running different Operating Systems (Ubuntu, etc.)
- Brings reproducibility to HPC.
- Instant deployment of complex software stacks (Genomics, OpenFOAM).
- Rapidly deploy the newest versions of software (Tensorflow).
- Bring your own container (use on Laptop, HPC, Cloud).
- **Available now on Terremoto and Habanero!**

Consumer GPU Cluster Experience

Sander Antoniadou, Senior Research Systems Administrator, Zuckerman Institute

Jochen Weber, Scientific Computing Specialist, Zuckerman Institute

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee



Use of Consumer Grade GPU Cards in Research

Nvidia, the dominant GPU vendor has multiple offerings, in research computing there are two major categories.

Enterprise (Tesla, Kepler)

- Custom built for GPU computer servers.
- Supported by major server vendors (such as HP and Dell)
- Offered as part of CUIT HPC clusters since Yeti.
- Expensive.

Consumer (GeForce)

- No error correcting memory
- Against Nvidia's terms of service, as such isn't supported by many vendors.
- No support advanced features such as large memory and nvlinc connections.
- Can be as much as 1/10 the price, and can fit in regular workstations.

The GPU Cluster Pilot

- Researcher need for GPUs was increasing, and many researchers were buying workstations with multiple consumer grade GPUs inside them to do machine learning.
- One researcher estimated he was going to need 100 GPUs for an upcoming project and working in the cloud or traditional HPC clusters was going to be too expensive.
- A PI was willing to fund a pilot to see if it would be feasible to build a dedicated GPU cluster, primarily for the neurotheory group
- The initial order was for three servers from the vendor Advanced HPC, containing 24 GeForce 1080ti GPUs which were delivered and set up last June.
- Work in conjunction with RCS a scheduler was set up, however the servers have largely been used directly by individual researchers.
- Some success, but need for GPU resources is still evolving.



Observations

- GPU computing isn't as flexible as traditional server solutions.
- Specifying hardware for GPU workloads is complicated.
- GPU lifecycles and performance increases are still changing very fast.
- Lack of support for vendor consumer GPUs is a major hurdle.
- The cost benefit of using consumer GPUs at the moment is too great to ignore.

CUIT Updates

George Garrett, Manager of Research Computing, CUIT

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee

Globus Update

- Provides secure, unified interface to research data.
- “***Fire and Forget***” high-performance data transfers between systems within and across organizations.
- Share data with collaborators.
- Columbia has procured an **enterprise license**.
- **Columbia Globus World Tour** workshop held on April 24, sponsored by CUIT and ZI.
- Contact **RCS** to get started with **Globus**.



Foundations for Research *Computing Update*

Marc Spiegelman, Chair of *Foundations* Advisory Committee

Patrick Smyth, *Foundations* Program Coordinator

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee



Foundations Goals

1. Address demand for informal training in computational research
2. Serve novice, intermediate, and advanced users with targeted programming
3. Foster a Columbia-wide community around research computing
4. Leverage existing University-wide investments in research computing infrastructure



Tiered Training Structure

Novice

- Software Carpentry Bootcamps
- Introductory Workshops
- Python User Group

Intermediate

- Distinguished Lectures in Computational Innovation
- Workshop series (modeled on HPC collaboration)
- Domain-specific intensives

Advanced

- Coordination with departmental curriculum

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee



Demand for Informal Instruction

First Bootcamp, August 2018

462 registrations for 90 seats

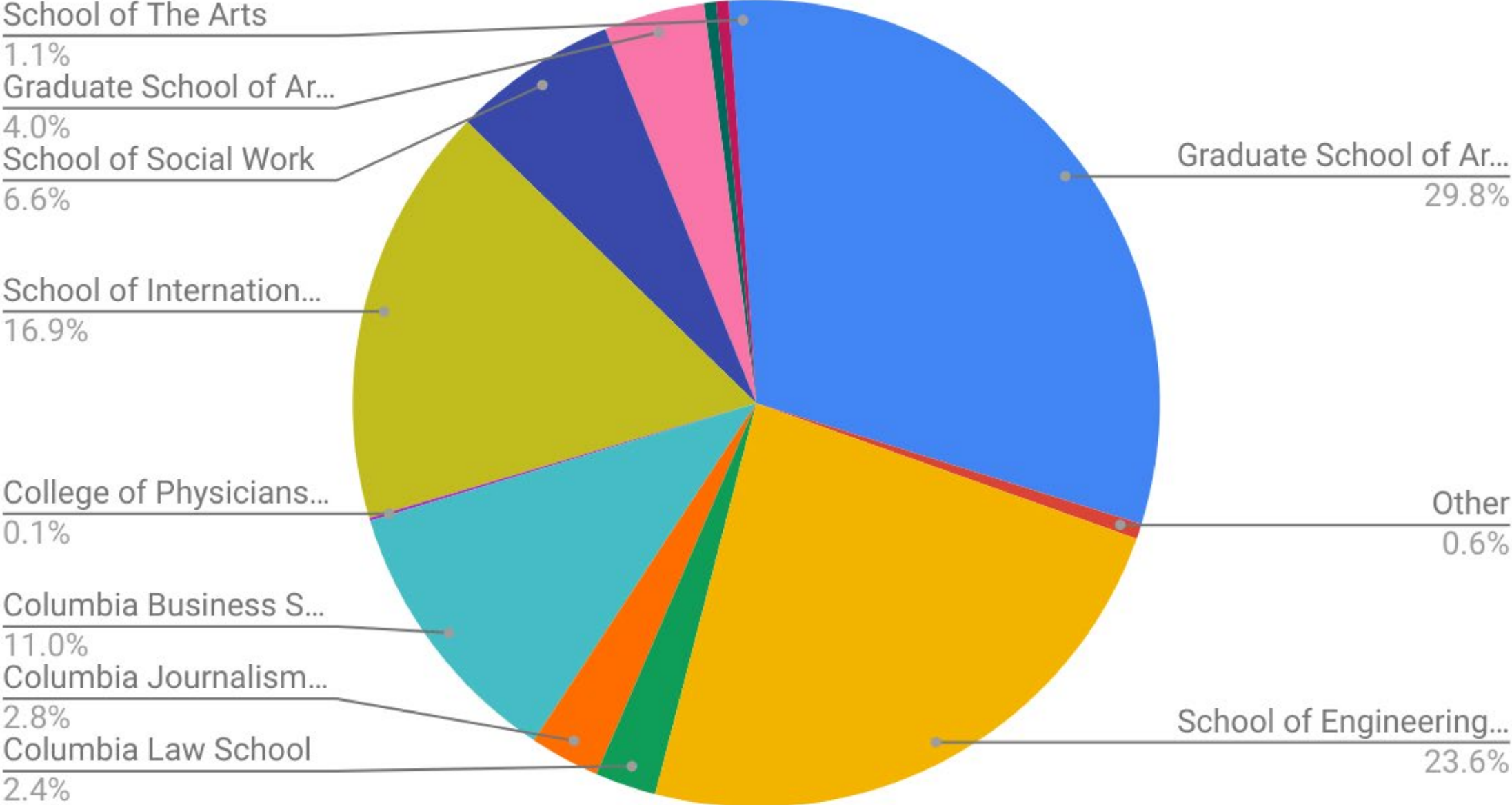
Second Bootcamp, January 2019

850 registrations for 120 seats

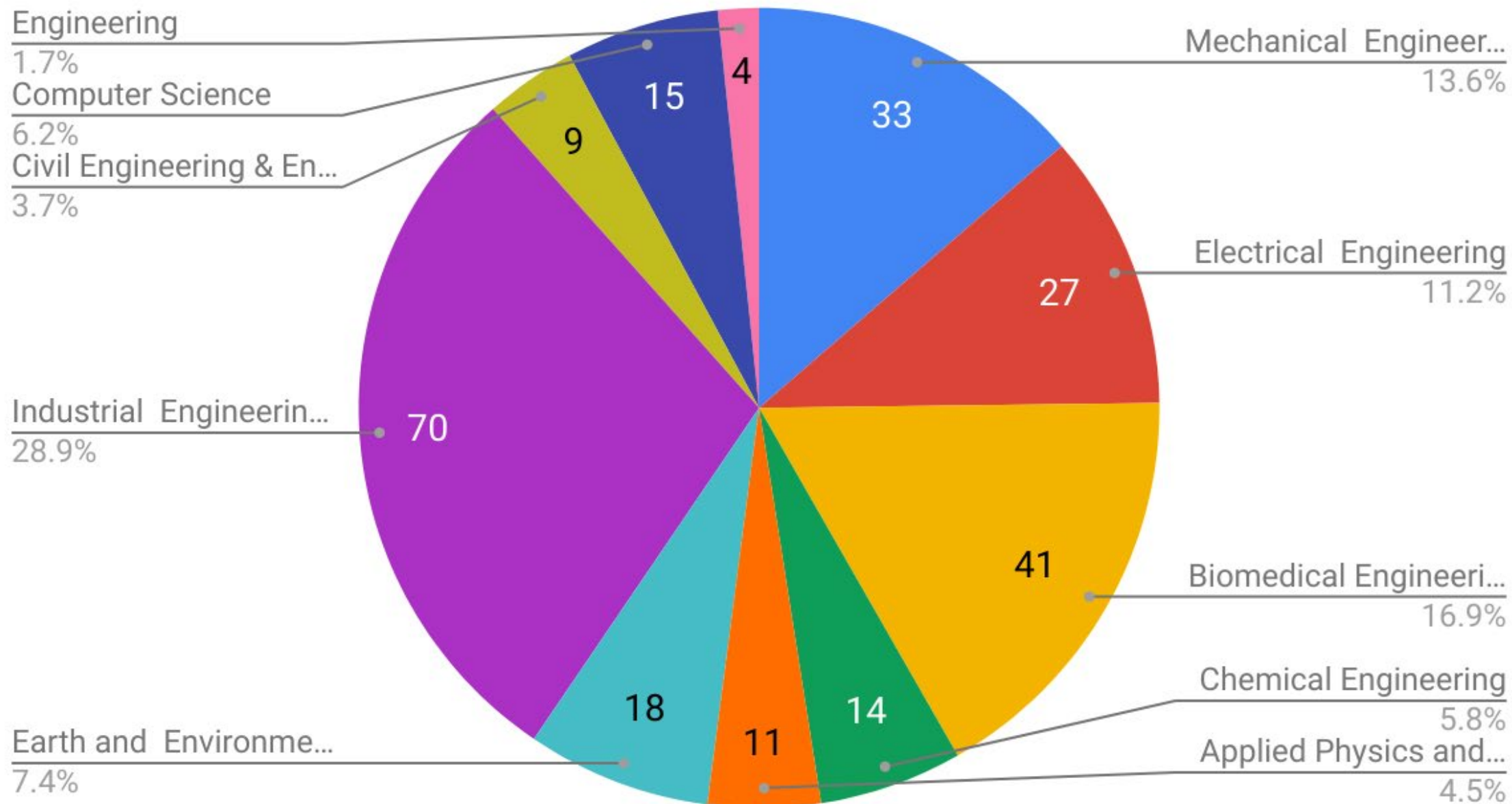
Spring break bootcamp for waitlisted students

Drew from waitlist, **45 students served**

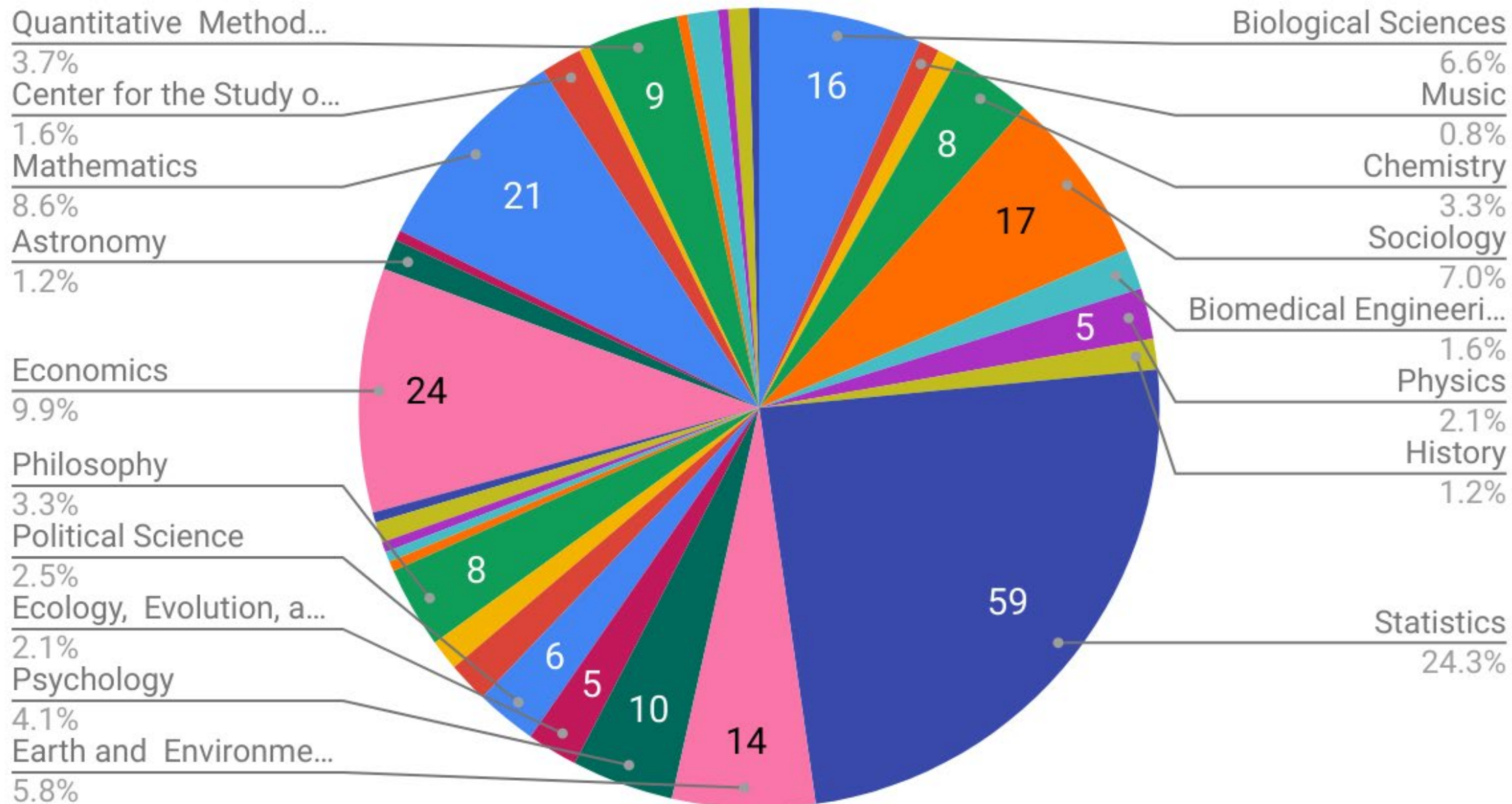
School Distribution, January 2019 (819 applicants)



Applicants by Department, SEAS only (Jan 2019)




Applicants by Department, Arts and Sciences only (Jan 2019)



Foundations Engagement

- **700+** total in-person engagements
- **235** served at two-day bootcamps
- **340+** attending direct instruction (bootcamps + workshops)
- **380+** attendees at 6 Distinguished Lectures
- **40+** attendees at Python User Group
- **14** instructors trained, 6+ in next training
- **1950+** contacts on mailing list



The Carpentries

- Software Carpentry, Data Carpentry, Library Carpentry
- Non-profit organization with a train-the-trainer model
- SC curriculum includes UNIX, Git, Python, and R, emphasizes applications



Columbia Instructors

- Silver membership, exploring increase in Columbia participation
- 14 instructors trained, 21 by end of year
- Instructors from CUIT, Libraries, CS, CUIMC, Business, Psychology, SPS

A vertical image on the left side of the slide shows a person from the back, wearing a light blue shirt, looking at a computer monitor. The monitor displays a complex software interface with various panels and data. The person is in a professional or academic setting.

Collaborations

Partner-led Collaborations

- DSI and Brown on *Distinguished Lecture* series
- RCS on cluster computing training
- CUIMC internal training collaboration in June

Instructor-led Collaborations

- R training at CUIMC
- Python workshops at Business School
- Text mining workshop at Center for Population Research
- Early stage of collaboration with Psychology Department



Scaling Up

Year Two

Focus on instruction, intermediate programming

Doubling direct instruction time

More instructors, increased community support

Expanded community programming

Exploring new partnerships, models



Intermediate Instruction

Expanded Programming

Targeted programming based upon student feedback

Third pre-semester bootcamp targets intermediate

Four workshop series (12 workshops total)

Half-day intensives in domain applications

After-hours Python User Group (outside speakers)

Foundations Questions?

Marc Spiegelman, Chair of *Foundations* Advisory Committee

Patrick Smyth, *Foundations* Program Coordinator

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee



2019 Research Computing Executive Committee

1. **HPC Update**

Publications Reporting

2. *Foundations for Research Computing* Annual Review

Thank You!

 COLUMBIA|RESEARCH

Shared Research Computing
Policy Advisory Committee