Shared Research Computing Policy Advisory Committee (SRCPAC)

Spring 2017 Meeting
Friday, March 3, 2017

Welcome & Introductions

Chris Marianetti, Chair of SRCPAC

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Habanero Update

Kyle Mandli, Chair of the Habanero Operating Committee

Rob Lane, Manager of Research Computing, CUIT

Habanero HPC Cluster

Cluster Nodes

Type	Nodes	Cores	K80 GPU
Standard	176	4,224	
High Memory	32	768	
GPU	14	336	28
Total	**222**	**5,328**	**28**

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Standard Node

CPU (2 per node)	E5-2650v4
Clock Speed	2.2 GHz
Cores	2 x12
Memory	128 GB

High Memory Node

Memory	512 GB

GPU Node

GPU (2 per node)	Nvidia K80
GPU Cores	2 x 4,992

Faculty RFP Committee

- Formed to advise CUIT on requirements, and participate in vendor selection.

- 7 Faculty Participants:
  - Ryan Abernathey (Earth & Environmental Sciences)
  - David Kipping (Astronomy)
  - Kyle Mandli (Applied Physics & Applied Mathematics)
  - Chris Marianetti (Applied Physics & Applied Mathematics)
  - Bob Mawhinney (Physics)
  - Lorenzo Sironi (Astronomy)
  - Brent Stockwell (Biological Sciences/Chemistry)

- Met four times from January – March 2016.

- Narrowed 16 bids from 11 vendors down to 2 winners (compute and storage).

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Participation

- 32 research groups purchased nodes
- Education tier established (A&S- and SEAS-funded)

- One-year rentals – coming soon!
- Free tier – coming soon!

Business Rules

- Business rules set by Habanero Operating Committee.

- Habanero launched with rules similar to those used on Yeti.

- Operating committee reviewed rules at January meeting; decided to keep current rules and review in the Fall.

Support Office Hours

HPC support staff are available to answer your Habanero questions in-person on the first Monday of every month.

Where: Science & Engineering Library, Northwest Corner Building
When: 3-5 pm first Monday of the month

Next session: 3-5 pm Monday, March 6

or

hpc-support@columbia.edu

Group Information Sessions

- HPC support staff can speak to your group at anytime.

- Topics can be general and introductory *or* tailored to your group.

- Contact hpc-support@columbia.edu to schedule a session.

Workshops

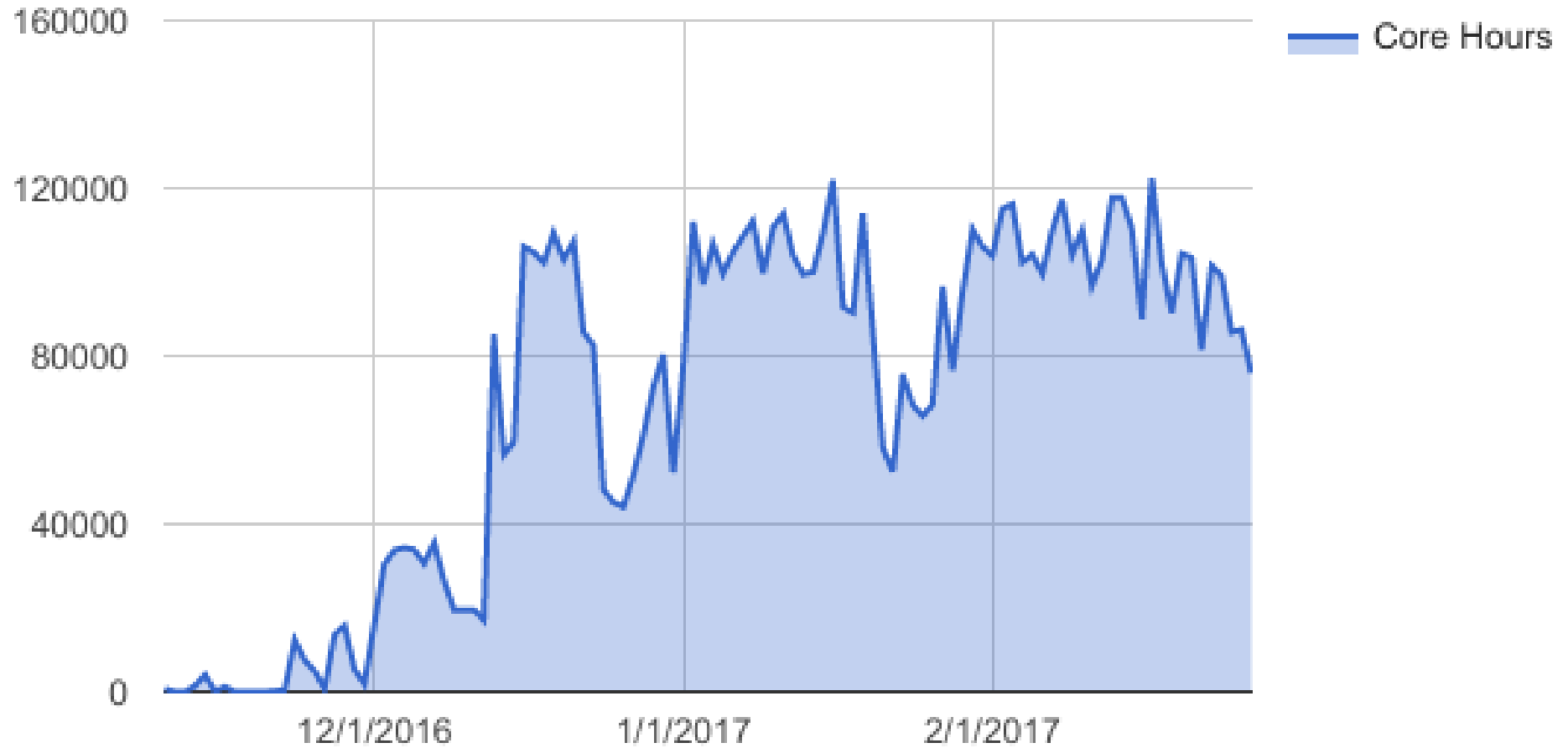Introductory Workshops run by CUIT & Libraries.

Next Week:

Tuesday, March 7, 2017
1:00 PM–2:30 PM
Science & Engineering Library, Northwest Corner Building
Registration: http://bit.ly/CUSELworkshop

Habanero Daily Usage

HPC Expansion
& Annual Purchase Cycle

Chris Marianetti, Chair of SRCPAC

Habanero Expansion Timeline 2017

What?	When?
Notification of Purchase Options	Early April
Order Period	May 15 – June 15
Equipment Delivery	September
Go-Live!	November

Overview of Training Services in FY17

Mark Newton, Director, Center for Digital Research and Scholarship
Ryan Abernathey, Assistant Professor, Earth & Environmental Sciences
Victoria Hamilton, Executive Director of Research Initiatives

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Digital/Computational Literacy *Informal* Efforts (Libraries, CUIT, EVP for Research, CTL, GSAS)

- GSAS Digital Orientation
- Programs and Workshops
- Open Labs
- Online Resources
- Technologies

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Software Carpentry Workshop

Software Carpentry

- Not-for-profit started in 1998 to teach STEM researchers statistical programming methodologies.

- Volunteer instructors run hundreds of events annually; lesson materials are reusable in the Creative Commons.
    - Six instructors sent to Columbia.

Administration

- Led by Ryan Abernathey, Assistant Professor, Earth & Environmental Sciences.

- Logistics support from the Office of the Executive Vice President for Research.

- Software Carpentry administrative fee, plus instructor travel/hotel and catering = **$18.8k** (or $142pp)
    - 50% covered by EVP for Research; 50% covered by participating departments.
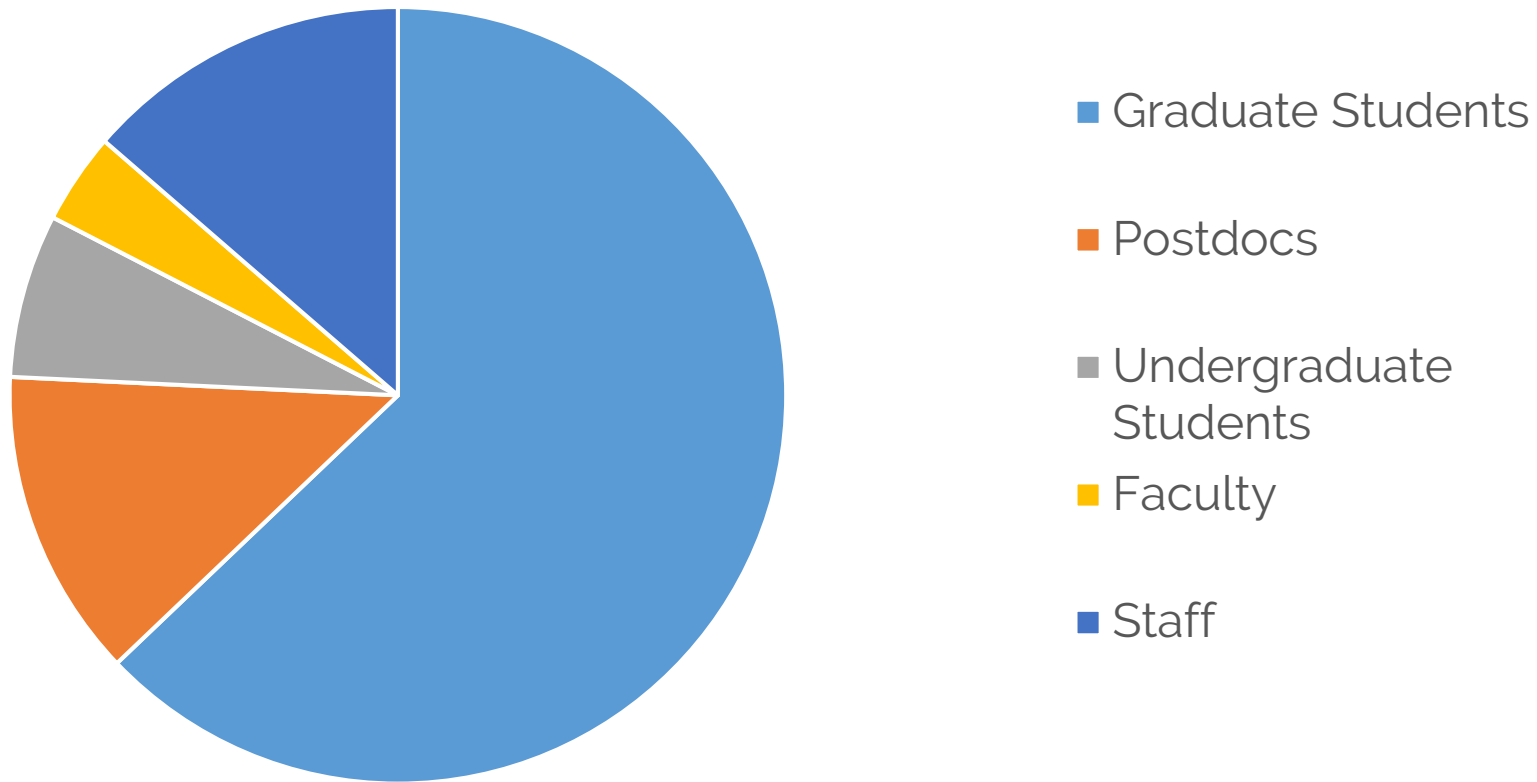
Institutional Partnership

- Columbia may join as a Software Carpentry *Member*. An annual fee of $7,500 includes:
    - 4 Coordinated Workshops, and 33% Discount on Additional Coordinated Workshops.
    - No Charge for Self-Organized Workshops at Partner Organizations.
    - 6 Instructors (Trained Online), Which Would *Eliminate Travel and Hotel Costs.*
    - Seat on the Software Carpentry Advisory Board.

Software Carpentry Workshop

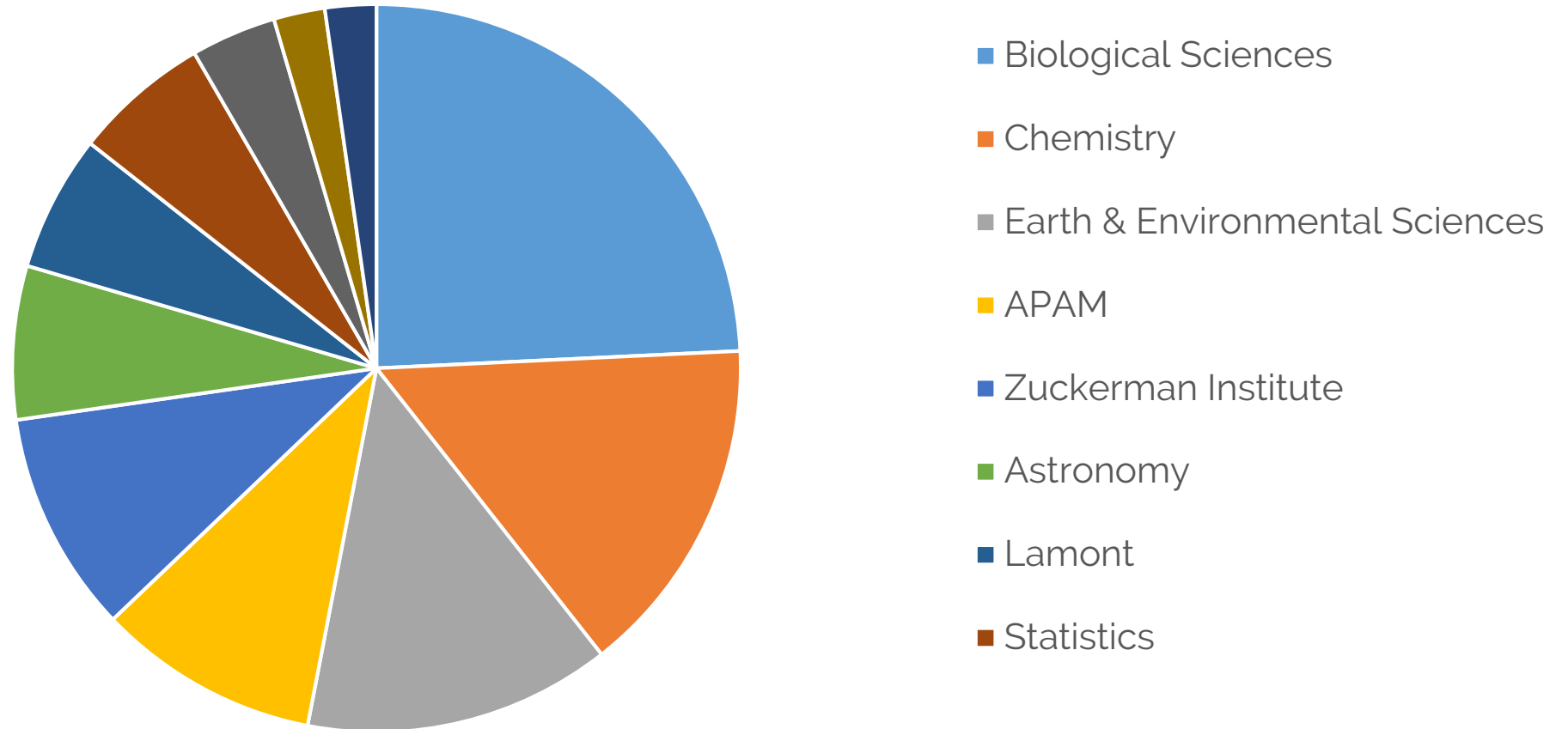- Held on September 1st-2nd in Mudd Building
- 210 registrations for 132 seats

Software Carpentry Workshop

Grad Students (83), Postdocs (17), Undergrads (9), Faculty (5), and Staff (18)

- Graduate Students
- Postdocs
- Undergraduate Students
- Faculty
- Staff

COLUMBIA|RESEARCH
Shared Research Computing
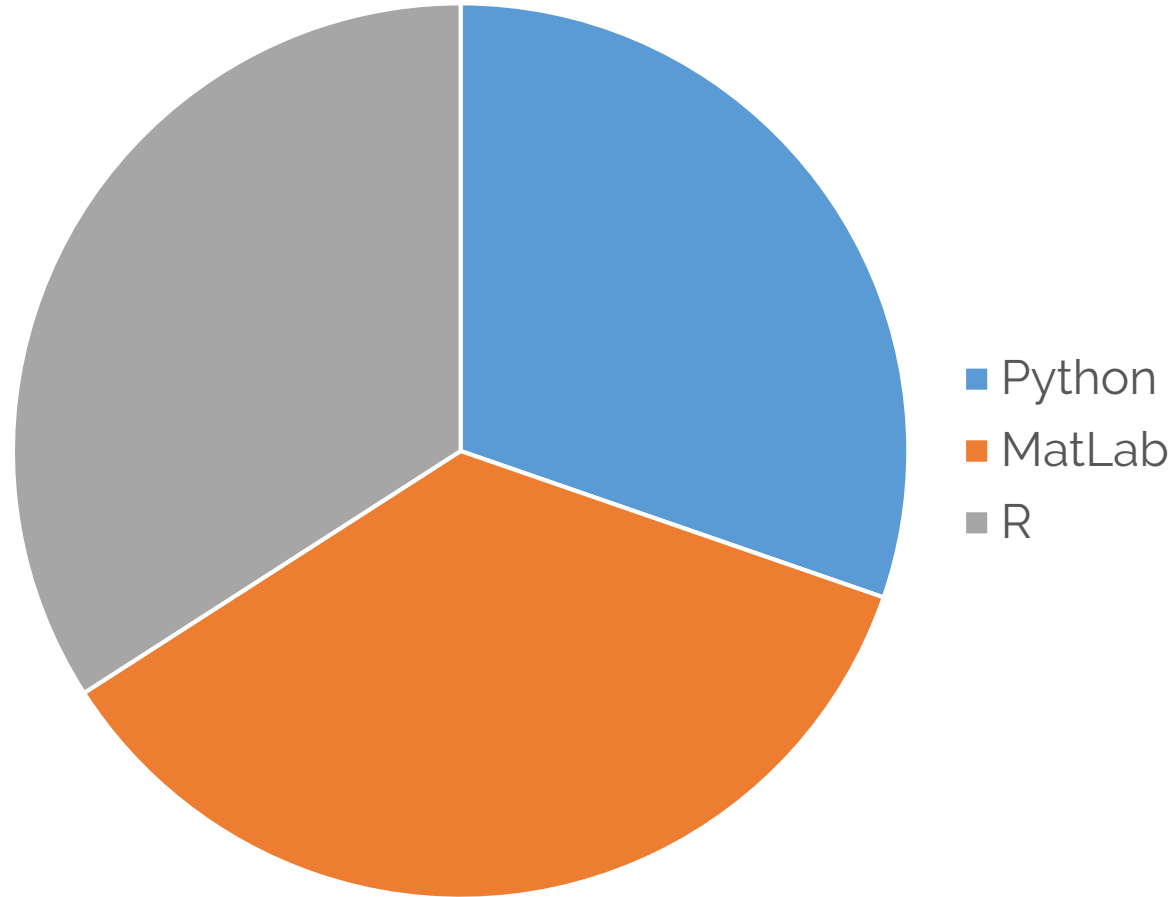Policy Advisory Committee

Software Carpentry Workshop

Representation from Biological Sciences (32), Chemistry (20), Earth & Environmental Sciences (18), APAM (13), Zuckerman Institute (13), Astronomy (9), Lamont (8), Statistics (8), Libraries (5), Teachers College (3), and Earth & Environmental Engineering (3)
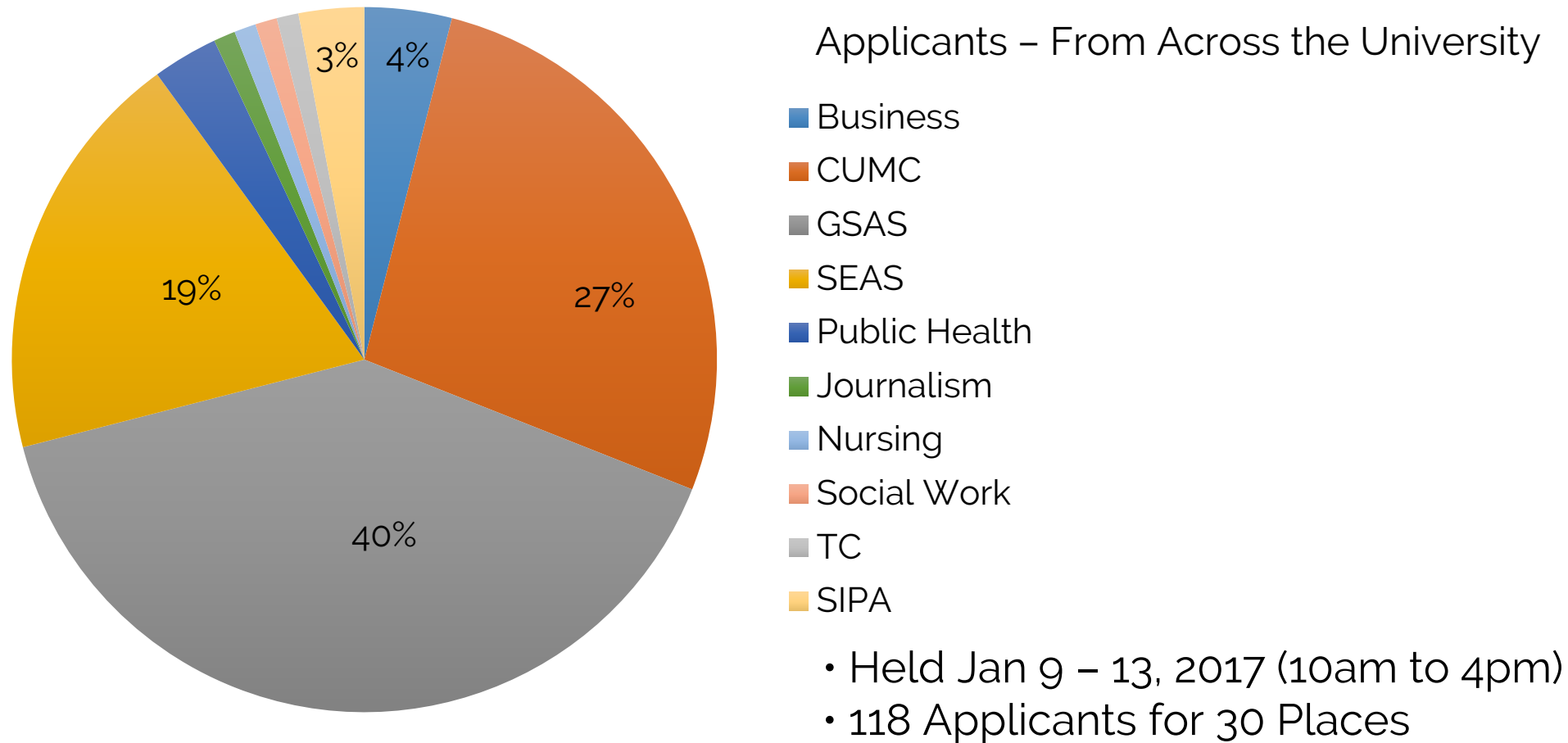

- Biological Sciences
- Chemistry
- Earth & Environmental Sciences
- APAM
- Zuckerman Institute
- Astronomy
- Lamont
- Statistics

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Software Carpentry Workshop

- Three courses:
  - Python (40)
  - MatLab (47)
  - R (45)

■ Python
■ MatLab
■ R

COLUMBIA|RESEARCH
Shared Research Computing
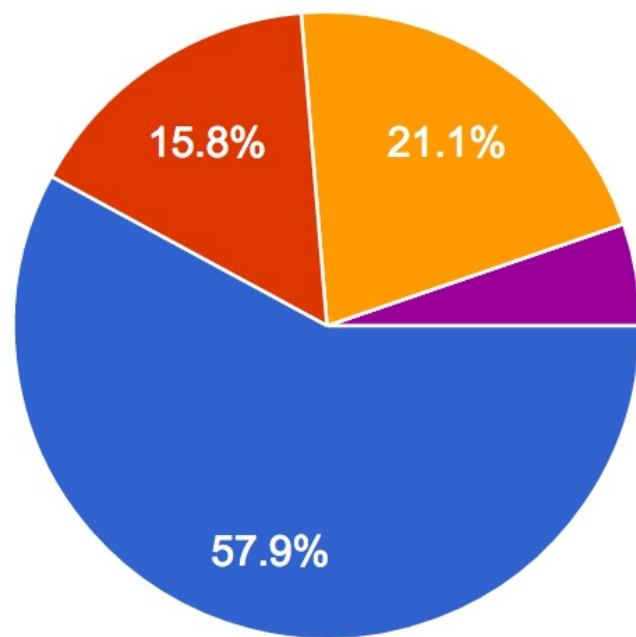Policy Advisory Committee

Software Carpentry Workshop

- Post-Event Survey Results (80% response rate):
  - Attendees want more training courses;
  - Attendees were not sufficiently aware of internal offerings;
  - Attendees want a combination of formal and informal training opportunities, including courses, workshops, trainings, and online tutorials;
  - Attendees believe their research will be aided by knowledge gained in the courses.

Collaboratory – Data Science Bootcamp

Applicants – From Across the University
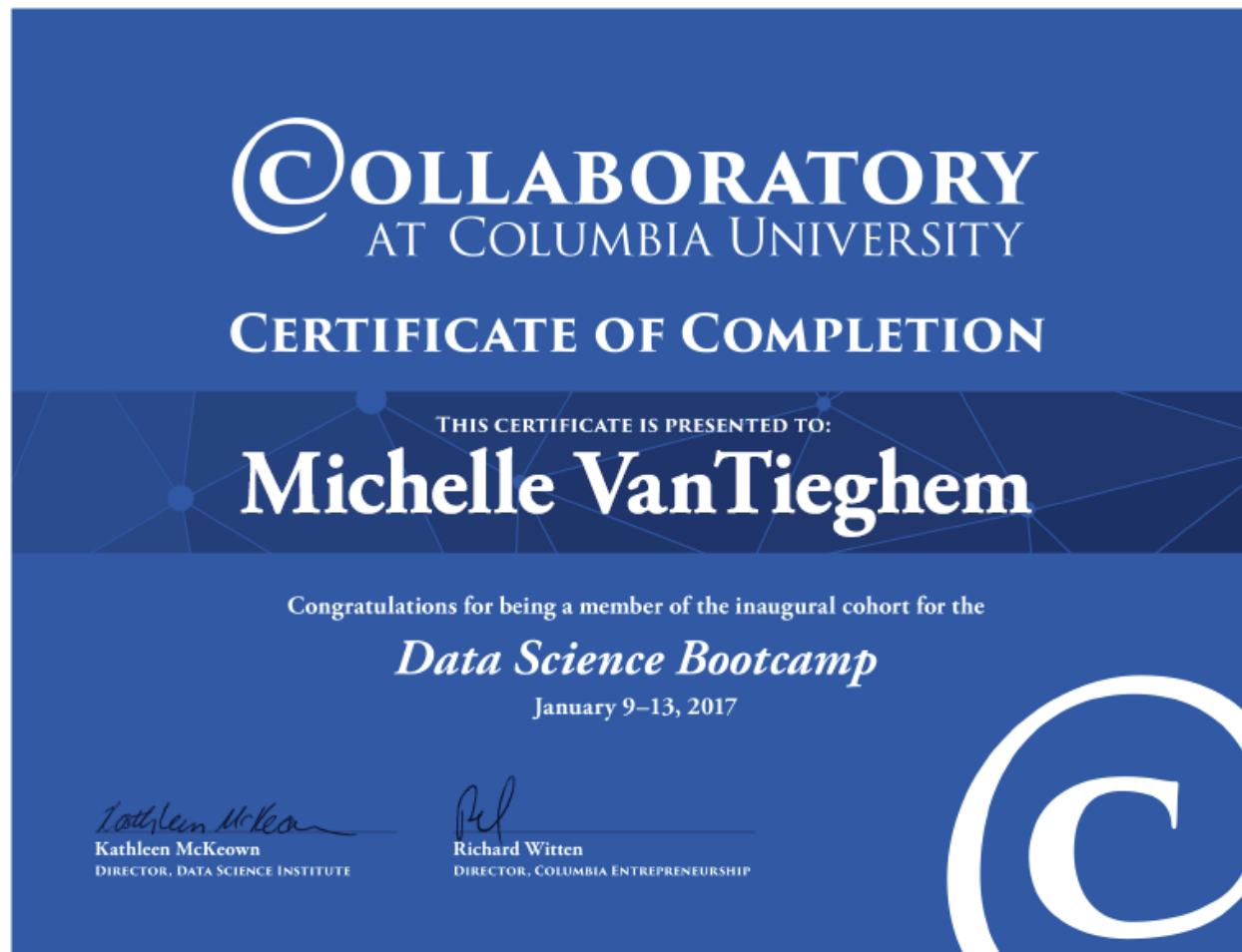
- Business
- CUMC
- GSAS
- SEAS
- Public Health
- Journalism
- Nursing
- Social Work
- TC
- SIPA

- Held Jan 9 – 13, 2017 (10am to 4pm)
- 118 Applicants for 30 Places

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Collaboratory – Data Science Bootcamp

Legend:
- PhD candidate
- Postdoc
- Research Scientist
- Assistant Professor
- Software Engineer
- Other

Pie chart values: 57.9%, 15.8%, 21.1%

Participants
Researchers in biomedical informatics, cancer biology, chemistry, communications, computational mechanics, earth science, neuroscience, operations research, physics, and more!

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Bootcamp Certificate

Data & Society Task Force

Chris Marianetti, Chair of SRCPAC

Victoria Hamilton, Executive Director of Research Initiatives

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Data & Society Task Force

- Created in 2015 by President Bollinger.
- Venue for relevant Deans and faculty to:
  - Assess Columbia's strengths and opportunities in data sciences
  - Gather consensus on recommendations to the President and Trustees.
- Chairs: Deans Mary Boyce (SEAS) and David Madigan (A&S).
- Very broad, University-wide membership.

Data & Society Task Force: Cyberinfrastructure Subcommittee

- Goal: Identify high-impact ideas for computational infrastructure.
- Co-Chairs: Mike Purdy (EVP for Research) and Gaspare LoDuca (CIO).
- Could prove to be compelling campaign objectives.
- Faculty Advisory Group:
  - Chris Marianetti, APAM (and SRCPAC)
  - Ryan Abernathey, Earth & Environmental Sciences
  - Louisa Gilbert, Social Work
- Two Focus Groups totaling 24 faculty participants.

Data & Society Task Force: Recommendations

- Four ideas proposed at the Task Force's February meeting:
  1. Managed public/private cloud infrastructure, including for highly-regulated data.
  2. Professional programming/software support staff to assist faculty in performing various sorts of computational research (very broadly speaking). This might include assistance in preparation of relevant grant proposals.
  3. Guidance/assistance regarding storage, from working to archival.
  4. Recruiting a cadre of intellectual leaders in the field of software tool development.

Yeti & Habanero Publications Reporting

Chris Marianetti, Chair of SRCPAC

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Publications Reporting

- CUIT and EVP for Research annually request data on publications emerging out of Habanero- and Yeti-hosted research.
- Data is critical for:
  - Reassuring the NIH that funds are being used optimally;
  - Encouraging Columbia leadership to continue their support;
  - Promoting research accomplishments within and outside of Columbia.
- Email shortly forthcoming from SRCPAC; updates can always be sent to srcpac@columbia.edu.

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Closing Remarks

Chris Marianetti, Chair of SRCPAC

COLUMBIA|RESEARCH
Shared Research Computing
Policy Advisory Committee

Appendix 1:
Habanero Expansion Timeline 2017 – Detailed

What?	When?
Vendor Assessment and Decision	March
Notification of Purchase Options	Early April
Vendor Pricing	April – May
Ordering Period	May 15 – June 15
Finalize Vendor Specifications	June 15 – July 15
Purchase Order Process	July 15 – August 15
Equipment Delivery	September
Go-Live!	November