

Developing HPC Skills Across the University Community

Brian Gregor, Ph.D.
Research Computing Services
Boston University



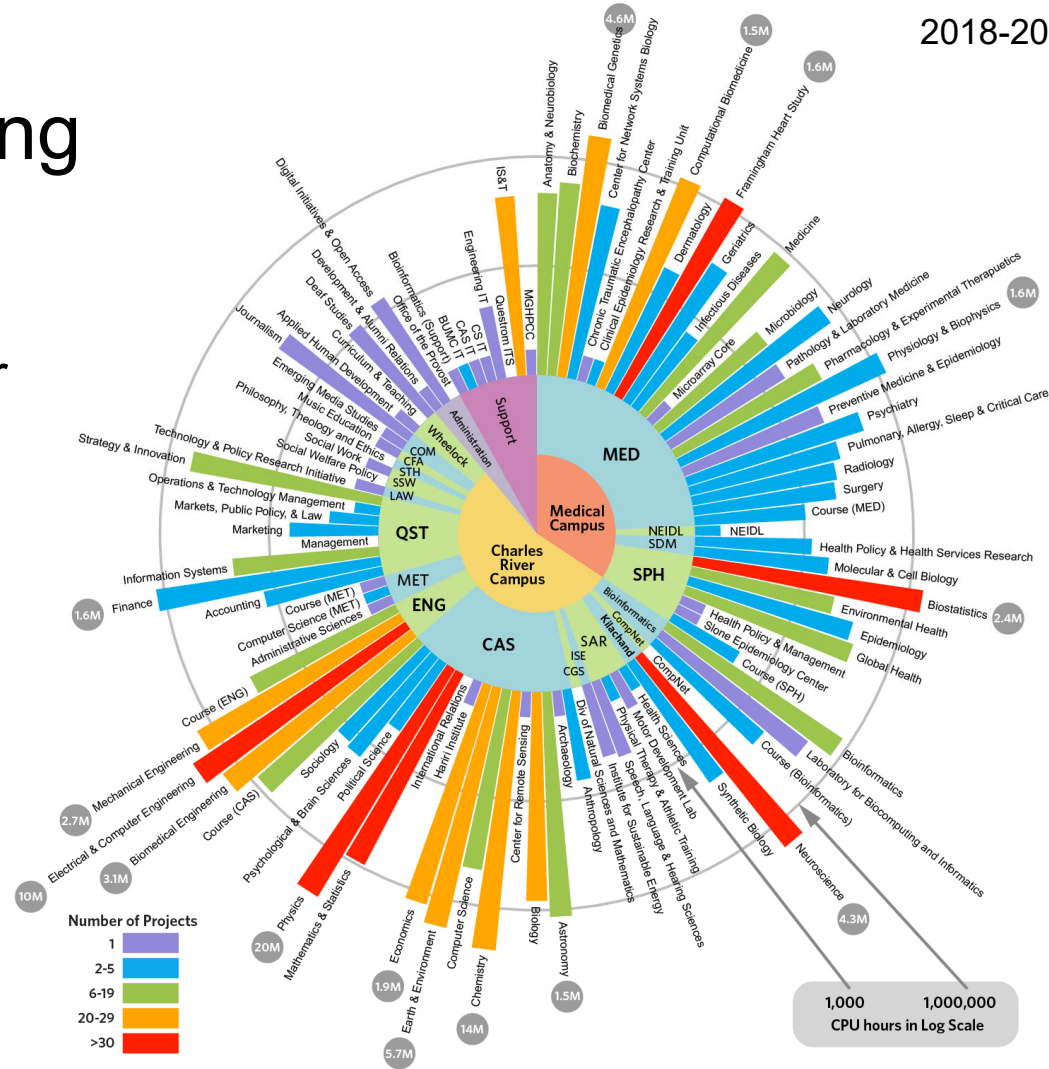
BU Research Computing Services (RCS)

- 16 staff members
 - 8 are in Applications Support working directly with the BU research community
- Shared Computing Cluster
 - Housed at the MGHPCC facility in Holyoke, MA
 - ~19,000 CPU cores
 - >6 PB of storage
 - 281 Nvidia GPUs
 - Grid Engine queuing system



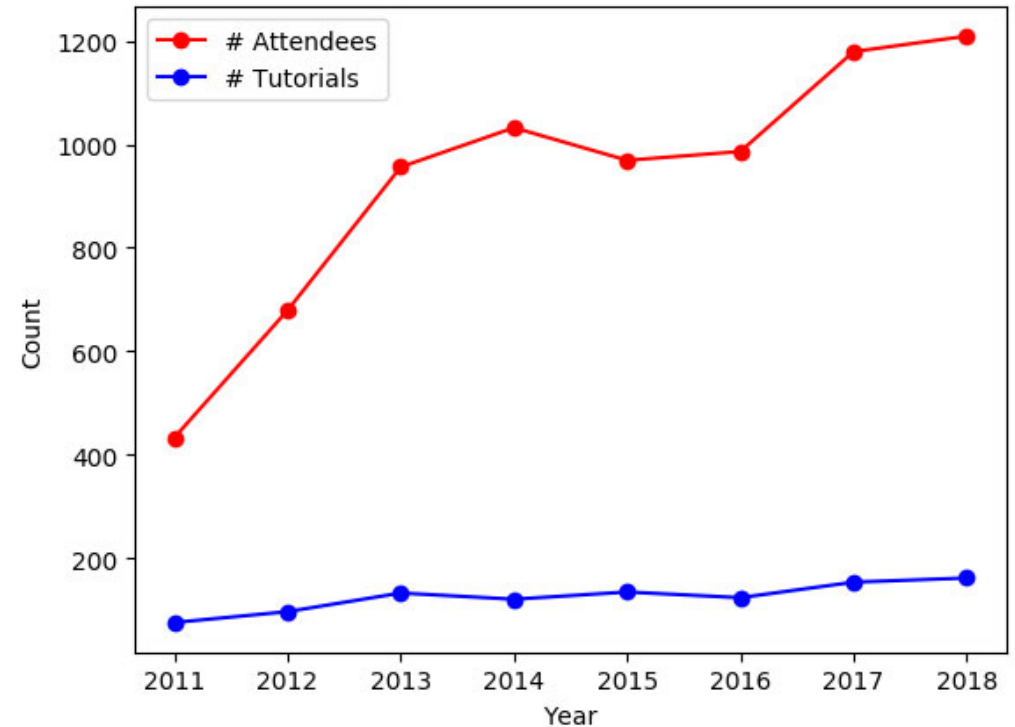
Interest in HPC is increasing

- Interest in data science continues to grow at BU leading more people to our HPC resources.
- >2000 cluster users across our Charles River and Medical campuses.
- RCS works with every college and almost every academic department at BU.



RCS Tutorials

- Offered in the Spring, Summer 1, and Fall semesters
 - Most are 1 or 2 sessions, 2 hours each.
- Topics:
 - Linux & cluster basics
 - Programming Languages
 - Data analysis
 - HPC / parallel programming tools
 - Visualization tools
 - Specific domains (e.g. neuroscience tools)



RCS Tutorials



■ Python:

- Intro (2 parts)
- Numpy & Scipy
- Data Analysis
- Visualization
- Scikit-learn
- Optimization (Jan 2020)

■ R:

- Intro
- Data Wrangling
- Graphics
- Graphics with ggplot2
- Programming techniques
- Large datasets with Data.table
- Optimization

Fall 2019 Python Tutorial Demand

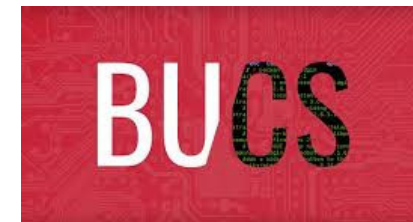
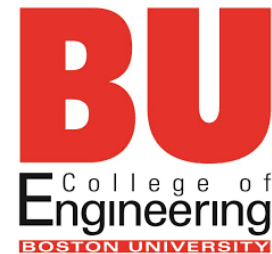


- Fall 2019: 3 complete sequences of Python tutorials scheduled
 - 6 tutorial sessions per sequence
 - Max of 25 attendees per tutorial
- All sessions were fully registered in early September.
- We received >40 emails asking to register.
- A fourth sequence was added for October. All sessions are fully registered.



Academic Class Support

- A growing area of involvement over the past 4 years
- Training on SCC usage, HPC software, general programming topics
- Engineering & Comp. Sci. deep learning courses
- Machine learning for Computational Biomedicine
- Computational neuroscience
- Biostatistics (SPH)
- Biophysics



- A data science degree program from BU's Mathematics Dept.
- Students come from:
 - Computer science, mathematics, economics, finance, physical sciences, engineering
 - Some are making a mid-career switch
- MSSP teaches R and Python alongside statistics and related math topics.



- RCS is teaching:
 - SCC usage
 - R Optimization
 - Python and “big data” tools
 - ex. Spark and PySpark
 - SQL and database usage

Computer-Assisted Text Analysis



- Course focus is on using large text resources
 - Social media, gov. databases, etc.
- “Data science” is an unfamiliar phrase to the students.
- Journalism / Communications graduate students have little or no background in programming.
- RCS is teaching:
 - Practical use of R for textual analysis
 - Best practices for R code



Future Efforts



- Implementation of OnDemand for academic access to the BU cluster
- How-To videos for introductory tutorials
- Expanded tutorial offerings
- Internship program for graduate students in research facilitation
- Continued participation in the Northeast Cyberteam program