

Universal Provider Award
Data & Computational Science Series (DCS²)
Summary Report for 2019

Prepared for
The University of Cincinnati Office of the Provost

December 2019

Universal Provider Award
Data & Computational Science Series (DCS²)
Summary Report for 2019

Report prepared by

Amy Latessa, PhD, Research Coordinator, IT@UC Research Computing Services, latessak@uc.edu

Amy Koshoffer, Assistant Director for the Research & Data Services, UC Libraries, koshofae@uc.edu

Suggested citation:

Koshoffer, A. & Latessa, A. (2019 December). *Data & Computational Science Series (DCS²) Summary Report for 2019*. Cincinnati, Ohio: UC Information Technologies and UC Libraries, University of Cincinnati.

Universal Provider Award
Data & Computational Science Series (DCS²)
Summary Report for 2019

Table of Contents

I.	Proposal Background	5
a.	Team	5
b.	Goals	5
c.	Acknowledgements	5
II.	Events & Workshops	6
	Table 1. DCS ² 2019 Workshops & Events	6
a.	XSEDE Monthly HPC Workshops	6
b.	Text Mining Hathi Trust Resources Using Python with Eleanor Dickson Koehl	7
c.	National Center for Supercomputing Application (NCSA)	8
d.	Ohio Supercomputer Center: OnDemand Hands-On Workshop and Presentation	9
e.	Reproducible Research with R Studio	9
f.	Engagement and Performance Operations Center (EPOC) UC Research Case Study Deep Dive	10
g.	DCSS Social at Kingsgate Bearcat Lounge	11
h.	Quantifying and Charactering Reuse of Biomedical Research	11
i.	Persian Twitter: Evolution of Social Media Landscape	13
j.	Research Computing: A View from the Trenches	14
k.	GIS Day	14
III.	Results	15
IV.	Marketing & promotional campaign & post-workshop feedback	15
V.	Recommendations for DCSS 3.0	16
VI.	DCSS 2020 Proposal	16
VII.	Conclusion	17
	Appendix A: University of Cincinnati Libraries Blog	18
	Student Research Consultant's experience at the XSEDE Big Data Workshop	18
	The Data & Computational Science Series Presents: Research Computing: A View from the Trenches	19
	The Data & Computational Science Series presents – GIS Day 2019 Nov 13th	20
	Appendix B: Post-workshop Survey Data	21

XSEDE Big Data Feb 2019 Post-workshop survey.....	21
XSEDE GPU Programming March 5, 2019_ Post-workshop survey.....	22
XSEDE HPC Boot Camp June 3-6, 2019_ Post-workshop survey.....	24
Appendix C: Sample Flyer.....	26

I. Proposal Background

The University of Cincinnati Libraries (UCL) and IT@UC Research and Development (R&D) received \$24,966, from the Office of the Provost's Universal Provider Award to host the *Data and Computational Science Series 2.0*, enriching faculty knowledge and scholarship with multi-disciplinary advanced computing and data scholarship workshops and tools. In brief, the *Data and Computational Science Series (DCS²) 2019* brought to the University of Cincinnati research and education community innovative workshops and distinguished speakers on advanced research data topics such as high performance computing, cloud computing, data visualization and analytics, research story-mapping, spatial analysis, artificial intelligence and machine learning.

a. Team

The organizational team was comprised of Amy Koshoffer (UCL), Don Jason (UCL), Rebecca Olson (UCL), Richard Johansen (UCL), Mark Chalmers (UCL), Amy Latessa (IT@UC R&D) and Jane Combs (IT@UC R&D). We collaborate on various initiatives including UC's Data Day and our job spheres often overlap in the fields of data management, data repositories, cyberinfrastructure and computational resources. We are the representatives at the UC new faculty orientation, host consultation hours in the Faculty Enrichment Center (FEC) and feel that our partnership makes the UC research community more fortified.

b. Goals

- Increase faculty awareness and adoption of enhanced research techniques and tools, software and cloud tools,
- Create a foundation for a community of researchers and educators to share best practices and techniques to enhance research,
- Connect disparate faculty across college and departmental lines to enhance potential research and education collaborations, and,
- Coordinate these data-centric efforts & resources amongst UC departments under a single platform.

c. Acknowledgements

We would like to thank many other UC faculty and staff for their support to make the series a success. We could not have hosted all the events without the help of people for room reservations, catering assistance and promotional materials and distribution. Thank you to the Faculty Enrichment Center for their new support infrastructure, and special thanks to the Office of the Provost Universal Funding Award, that without, these workshops and collaborations were but a vision.

II. Events & Workshops

Table 1. DCS² 2019 Workshops & Events

Date	Workshop/Event Title
February 12-13, 2019	XSEDE Monthly HPC Workshop: Big Data
February 26-27, 2019	Text Mining HathiTrust Resources Using Python
March 5, 2019	XSEDE Monthly HPC Workshop: GPU Programming Using OpenACC
March 12, 2019	Rob Sisneros, National Center for Supercomputing Application (NCSA)
April 3, 2019	Ohio Supercomputer Center: OnDemand Hands-On Workshop and Presentation
April 16, 2019	Mine Cetinkaya-Rundel, R Studio
April 26, 2019	Engagement and Performance Operations Center (EPOC) UC Research Case Study Deep Dive
June 3-6, 2019	XSEDE Monthly HPC Workshop: HPC Boot Camp
July 11, 2019	DCSS Social at Kingsgate Bearcat Lounge
September 17, 2019	Lisa Federer, PhD, Data Science and Open Science Librarian at the National Library of Medicine
October 1-2, 2019	XSEDE Monthly HPC Workshop: Big Data
October 10, 2019	Emad Khazraee, PhD, Data Scientist at Indeed, Fellow at Berkman-Klein Center for Internet and Society, Harvard University
November 5, 2019	XSEDE Monthly HPC Workshop: OpenMP
November 7-8, 2019	George Turner, Indiana University Chief Systems Architect, University Information Technology Services (UITs) Research Technologies
November 13, 2019	GIS Day

a. XSEDE Monthly HPC Workshops

The University of Cincinnati is one of the select group of universities to serve as a satellite-site for these monthly high-performance computing workshops hosted by XSEDE¹ and the Pittsburgh Supercomputing Center. These monthly workshops included:

1. GPU programming
2. 4-day HPC boot camp from beginner to advanced level with topics on MPI, OpenMP, OpenACC and accelerators
3. OpenMP to give C and Fortran programmers a hands-on introduction to MPI programming
4. Big Data with topics such as Hadoop and Spark

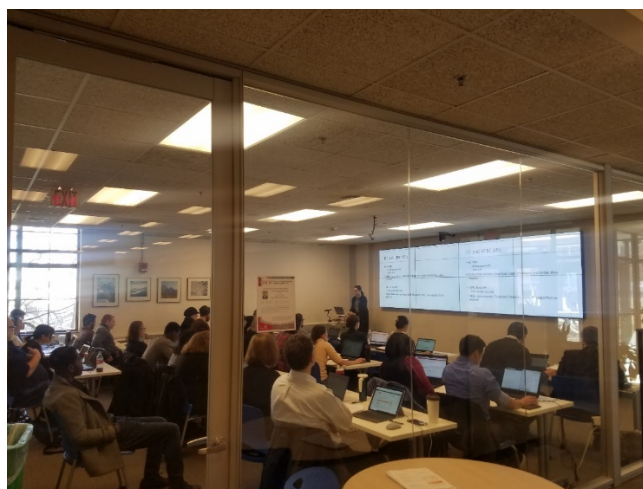
¹ The Extreme Science and Engineering Discovery Environment (XSEDE) is a single virtual system that scientists can use to interactively share computing resources, data and expertise. People around the world use these resources and services — things like supercomputers, collections of data and new tools — to improve our planet.



XSEDE Big Data Workshop on October 1-2, 2019

Instructor: John Urbanic, Parallel Computing Scientist, Pittsburgh Supercomputing Center					
Date	Workshop (Format: Wide-area classroom WAC) training platform	Attendees			
		Faculty	Graduate Students/ Postdocs	Staff	Total
2/12-13	Big Data	3	12	3	19
3/5	GPU Programming Using OpenACC		4	1	5
6/3-6	HPC Boot Camp	2	6		8
10/1-2	Big Data	5	5	3	13
11/5	OpenMP		2	1	3

b. Text Mining Hathi Trust Resources Using Python with Eleanor Dickson Koehl



February 26 presentation on Hathi Trust Resources

This workshop, presented by Eleanor Dickson Koehl, the Digital Scholarship Librarian at HathiTrust, introduced attendees to the HathiTrust Research Center's tools and services for utilizing the massive HathiTrust Digital Library in computational text analysis. The HTRC leverages the scope and scale of HathiTrust Digital Library's holdings to allow researchers the opportunity to perform text data mining. Workshop attendees developed skills that will allow them to conduct text analysis research using HathiTrust data.

Topics covered include:

- How the HTRC makes HathiTrust volumes available for text mining.
- How to identify relevant volumes and build worksets (datasets) of content for analysis.
- How to access HathiTrust data and metadata via provided APIs, request procedures, and open datasets.
- How to acquire and analyze Extracted Features data from the HathiTrust Research Center using Python.

Instructor: Eleanor Dickson Koehl, the Digital Scholarship Librarian at HathiTrust					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
2/26	Presentation: The Hathi Trust Research Center	8	7	7	28

c. National Center for Supercomputing Application (NCSA)

The presentation and workshop focused on scientific visualization using VisIT, presented by Robert Sisneros, a Senior Research Scientist at the National Center for Supercomputing Applications (NCSA) and director of the NCSA's Data Analysis and Visualization Group. NCSA is tasked with supporting science teams utilizing NSF HPC resources as well as furthering the state of scientific visualization through cutting edge research. The focus of the presentation included: in situ visualization, data models and representations, parallel analysis algorithms, I/O parameter optimization, and "big data" analytics.

About the NCSA:

The [National Center for Supercomputing Applications \(NCSA\)](#) is a hub of [transdisciplinary research and digital scholarship](#) where University of Illinois faculty, staff, and students, and collaborators from around the globe, unite to address research grand challenges for the benefit of science and society. Current research focus areas are [Bioinformatics and Health Sciences](#), [Computing and Data Sciences](#), [Culture and Society](#), [Earth and Environment](#), [Materials and Manufacturing](#), and [Physics and Astronomy](#).

Instructor: Rob Sisneros, Senior Research Scientist at the National Center for Supercomputing Applications (NCSA)					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
3/12	Presentation & Workshop: Scientific Visualization Using VisIT	5	2	7	14

d. Ohio Supercomputer Center: OnDemand Hands-On Workshop and Presentation

The OSC @ UC Getting Started: at OSC hands-on workshop helps new users set up accounts and included activities such as:

- Learn high performance computing concepts
- Navigate OSC storage and file systems
- Learn how to create an account and request a project in the client portal
- Access OSC clusters using the web portal OnDemand
- Submit jobs to the batch system and learn to review output

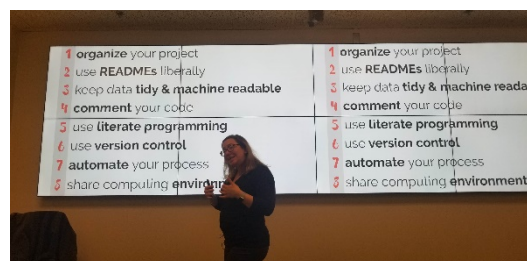
Instructor: Kate Cahill					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
4/3	Hands-on workshop	1	2	1	4

e. Reproducible Research with R Studio



Dr. Mine Cetinkaya-Rundal

For a data analysis to be reproducible, the data and code should be assembled in a way such that results (e.g. tables and figures) can be re-created. While the scientific community is by and large in agreement that reproducibility is a minimal standard by which data analyses should be evaluated, and a myriad of software tools for reproducible computing exist, it is still not trivial to reproduce someone's (sometimes your own!) results without fiddling with unavailable analysis data, external dependencies, missing packages, out of date software, etc. In this talk we present good, better, and best workflows for reproducibility that touch on everything from data storage, cleaning, analysis, to communicating final results.



Workshop: Data Visualization in R with ggplot2 and ganimate

ggplot2 is a system for declaratively creating graphics in R, based on The Grammar of Graphics. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details. In this workshop we will introduce various features of ggplot2 as well as extensions to this package for customizing your plots. Additionally, we'll touch on gganimate, which extends ggplot2 by adding new grammar classes that customizes how a plot should change with time. This two-hour workshop assumes no prior R experience, but a willingness to dive right into hands-on exercises.

Instructor: Dr. Mine Cetinkaya-Rundal, Associate Professor at Duke University					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
4/16	Presentation: Reproducible Research with R Studio	12	16	2	30
4/16	Workshop: Data Visualization in R with ggplot2 and gganimate	12	16	2	30

f. Engagement and Performance Operations Center (EPOC) UC Research Case Study Deep Dive

The National Science Foundation's Engagement Performance and Operations Center (EPOC) team has recently been funded to 'support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process'. One of their services is to help document scientific use cases and needs for the advanced research tools available at universities.



UC research faculty representation included: Criminal Justice Research Institute (CECH), Physics (A&S), Math (A&S), Aerospace Engineering (CEAS), Genomics (CCHMC), Pediatrics (CCHMC), Biomedicine (COM)

EPOC Support Activity for UC: Case Study Deep Dive:

The Application Deep Dives worked closely with the application communities to understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues.

- Science Background
- Collaborators
- Instruments and Facilities
- Process of Science
- Remote Science Activities
- Software Infrastructure
- Network and Data Architecture
- Cloud Services
- Resource Constraints
- Parent Organization(s)
- Outstanding Issues

EPOC Final Report: <https://escholarship.org/uc/item/6t58p052>

Instructor: Jason Zurawski, Science Engagement Engineer, Energy Sciences Network and Hans Addleman, Principal Network Systems Engineer, Indiana University					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
4/26	1-day deep dive: faculty presentations	9	1	15	25

g. DCSS Social at Kingsgate Bearcat Lounge

Beyond speakers and hands-on trainings, we organized a DCS² Social on July 11, 2019. Feedback from UC faculty attendees indicated that they seek a space for conversation and informal networking with other research-oriented faculty. We provided soda, water and appetizers and there was a broad pool of attendees. Bringing people together in these ways encourages interdisciplinary thinking and results and reaches beyond simple attendance numbers. We believe that the community-building component of our collective efforts is essential component for long-term faculty retention.

Location: Kingsgate Hotel Bearcat Lounge					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
7/11	Open Social	7	8	10	24

h. Quantifying and Characterizing Reuse of Biomedical Research

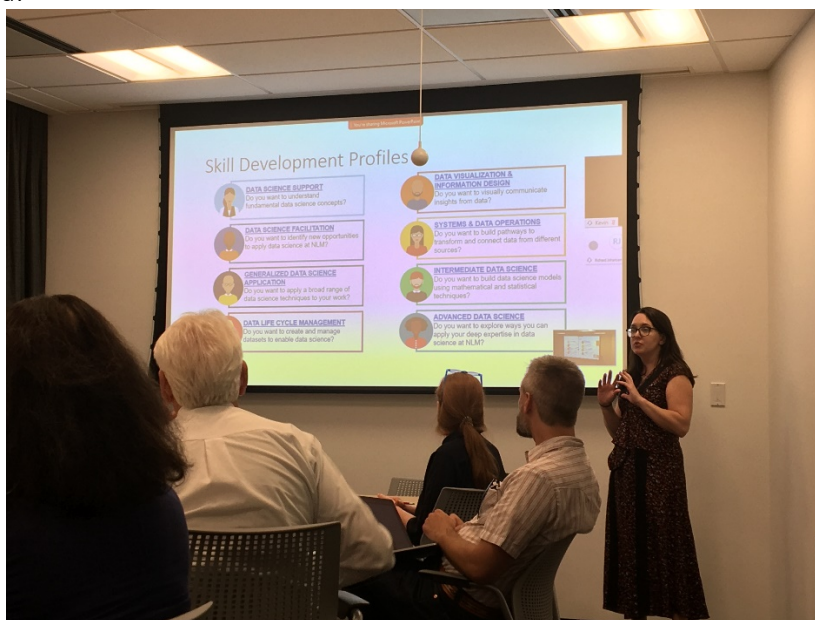
Lecture and Luncheon: *If you Share it, Will They Come? Quantifying and Characterizing Reuse of Biomedical Research Data*. Since the mid-2000s, new data sharing mandates have led to an increase in the amount of research data available for reuse. Reuse of data benefits the scientific community and the public by potentially speeding scientific discovery and increasing the return on investment of publicly funded research. However, despite the potential benefits of reuse and the increasing availability of data, research on the impact of data reuse is so far sparse. This talk provided a deeper understanding of the impacts of shared biomedical research data by answering the questions “what happens with datasets once they are shared?”

Specifically, this talk will demonstrate that data are often reused in very different contexts than for which they were originally collected, as well as explore how patterns of reuse differ between dataset types. This talk also considers patterns of data reuse over time and the topics of the most highly reused datasets to determine whether it is possible to predict which datasets will go on to be highly reused over time. Finally, career stage and geographic location of data reusers provide an understanding of who benefits from shared research data. These findings have implications for several stakeholders, including researchers who share data and those who reuse it, funders and institutions developing policies to reward and incentivize data sharing, and repositories and data curators who must make choices about which datasets to curate and preserve.

Workshop: Endless Forms Most Beautiful: Creating Customized Data Visualization with ggplot2 in R

It's said that a picture is worth a thousand words, and a picture can also be worth a thousand (or more) data points. Visualizing data can be an important way of understanding and communicating the meaning and patterns that lie hidden within. Whether you're creating a quick exploratory chart to identify patterns in a dataset or designing a complex visualization to communicate your findings to an external audience, data scientists need the skills to visually demonstrate their data.

This workshop focused on creating visualizations with the R package ggplot2. This package builds upon a design theory called the Grammar of Graphics and allows users to create visualizations that are customizable to a nearly endless degree. The workshop will also use RColorBrewer, a package that makes it easy to create custom color palettes, including color-blind friendly palettes. Attendees learned how to use these tools to create visualizations that incorporate the elements of design to effectively communicate data.



Dr. Lisa Federer, Data Science and Open Science Librarian at the National Library of Medicine

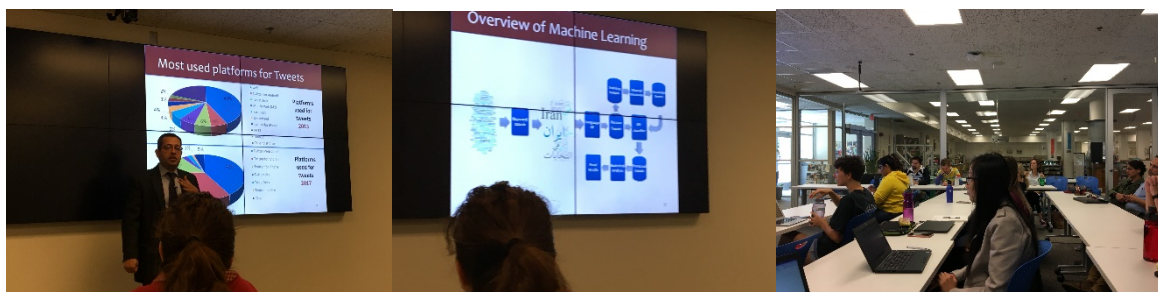
Instructor: Lisa Federer, PhD, Data Science and Open Science Librarian at the National Library of Medicine					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total

9/17	Breakfast with UC Libraries Research & Data Services Team	5		5	10
9/17	Lecture: Quantifying & Charactering Reuse of Biomedical Research	11	4	12	27
9/17	Workshop: Creating Customized Data Visualization with Ggplot2 in R	12	6	7	25
9/18	Breakfast with Data, Informatics and IT Leaders	4		4	8
9/18	Lunch with UC Institute for Interdisciplinary Data Science	7		1	8

i. Persian Twitter: Evolution of Social Media Landscape

“The free Twitter lecture and workshop easily beat the whole data science event at the business school.”²
– Assistant Professor in EECS

Social media have transformed our societies and contributed to the creation of online public spaces. In the past few years, we witnessed how social media were central to any debate of socio-political movements around the world. Social media were cited as the new catalysts of social change in these contexts. However, still, controversies exist about the role they played in these movements. Studying these online spaces becomes a challenge considering the pressure of repressive cultural environments. In such environments, accessing users freely is not possible in most cases. Moreover, the scale and complexity of data require employing multiple methods to achieve a more nuanced understanding of online publics. To overcome these challenges and to gain a better understanding of the dynamics of the online public environments in Iran, Khazraee started a project for the Cartography of Iran’s online publics. The goal of the project was collecting empirical evidence that helps us to achieve a high-resolution image of public online environments in Iran. As part of this project, he conducted research on Twitter use among Iranians during two presidential elections in Iran in 2013 and 2017. This study aimed to map the political landscape of Persian Twitter between these two elections and to investigate how it has been transformed during this period. The study, also, revealed the transformation of the power structure in Persian Twitter between the two elections, as well as the role of various political communities and their influence on the larger communication network.



Workshop: Analyzing Twitter Data with R

² Referring to the Data Science Symposium at the Lindner College of Business

Social media provided an unprecedented opportunity to investigate social interactions at a large scale with high resolution. However, social media also generate massive volumes of data which requires the use of computational methods for analysis. Computational methods enable us to conduct a distant reading of social phenomena and identify areas of interest for close readings. This workshop explored how to use R for analyzing Twitter data. In the first section, the workflow for social media data analysis was discussed; then, in the following part, the class worked on a real-world data set for hands-on practice.

Instructor: Emad Khazraee, PhD, Data Scientist at Indeed, Fellow at Berkman-Klein Center for Internet and Society, Harvard University					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
10/10	Lunch & Learn	4	13	6	23

j. Research Computing: A View from the Trenches

Chief Systems Architect for Research Technologies at the [Pervasive Technologies Institute](#), Indiana University gave an insider's presentation on research computing in higher education. George has helped build one of the most abundant and successful research centers in the United States and has been assisting the University of Cincinnati to set up its first ever [Advanced Research Computing](#) Cluster, now available to UC faculty and students. His talk will address challenges in research computing and HPC and share the benefits that a successful University research computing center can have upon a research institution.

[Linux 101 Workshop](#): This free workshop provided a brief introduction on the layout of the Linux operating system. It provided an overview Linux concepts with a few exercises and will help people to understand the paradigms and why this is important in the bigger scheme of things. Like most of our presentations, this was video recorded, and we hope to incorporate the Linux 101 workshop into our ARC curriculum in the 2020 series.

Instructor: George Turner					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
11/7	Presentation: George Turner	5	13	4	22
11/7	Workshop Linux 101: Himakar Ganti	2	4	2	8

k. GIS Day

This community building event brings to together researchers and practitioners in the field of Geographic Information Systems (GIS) for a celebration of the application and power of GIS. In addition to the Provost Office contribution through the UP grant, the event is sponsored by the Geography Graduate Student Association, Graduate Student Governance Association, Department of Geography & GIS, and Joint Center for GIS and Spatial Analysis.



GIS Day 2019

Instructor:					
Date	Format	Attendees			
		Faculty	Graduate Students	Staff	Total
11/13	GIS Day 2019 - Machine Learning with Arc GIS	10	54	1	65

III. Results

The DCS² resulted in many positive outcomes, including:

- Organized, hosted and planned 8 speakers/seminars and 18 workshops and hands-on training sessions.
- Trained over 369 people, including 112 faculty (30.4%). Faculty attendees came from 8 UC colleges, representing 14 departments/disciplines: A&S (math, chemistry, biology), Lindner College of Business (OBAIS), College of Medicine (environmental health, biomedical informatics), CEAS (computer science, mechanical & materials engineering, engineering education), CECH (education, criminal justice), College of Nursing, College of Allied Health Sciences, and University of Cincinnati Libraries.
- Connected faculty with external experts, resources and tools. Introduced UC faculty to free regional and national research and computing resources, technical and disciplinary experts, and enhanced research practices to increase quality of research outputs. Numerous faculty have established accounts to utilize these new resources, including (1) the University of Pittsburgh's supercomputing resource [Bridges](#), (2) 42 new UC users to UC's new [Advanced Research Computing](#) (ARC) cluster and, (3) the National Science Foundation's Extreme Science and Engineering Discovery Environment ([XSEDE](#)), a single virtual system that scientists can use to interactively share computing resources, data and expertise. Faculty across disciplines have consulted with technical experts and instructors, including Eleanor Dickson-Koehl (Digital Scholarship Librarian at HathiTrust), Dr. Mine Cetinkaya-Rundal (R Studio, Associate Professor at Duke University), and Rob Sisneros (Senior Research Scientist at the National Center for Supercomputing Applications (NCSA)).

IV. Marketing & promotional campaign & post-workshop feedback

In addition to the flyer template designed by Melissa Cox-Norris, Directory of Library Communication, University of Cincinnati Libraries (UCL) for the 2018 series, we also set up a permanent website.

<https://libraries.uc.edu/research-teaching-support/research-data-services/data-computational-science-series.html>

Most of the events' registration were through the [Faculty Development OneStop](#),³ which allowed us to set workshop size, share completion of workshop certificates with attendees, and have a repository of our events. There was no way to limit registration to faculty only, which we were initially concerned about, but as our registration numbers show, it turned out to not be a problem, as we had mixed faculty, student and staff at each event. Each event was posted on the UC Liblog (Appendix A) and the UCIT R&D website.

After each workshop we shared a 6 questions Qualtrics survey to assess attendee satisfaction. Although the complete rate of the survey is low, we did gather some insights. See results Appendix B.

- Q1 - How did you hear about the workshop?*
- Q2 - How well did the workshop meet the description?*
- Q3 - How well did this class meet your personal objectives?*
- Q4 - Overall, how effective was the training session?*
- Q5 - What did you hope to learn in this workshop?*
- Q6 - Please list suggestions for other courses, questions and/or comments*

V. Recommendations for DCSS 3.0

- a. Further streamline our efforts:
 - i. Marketing and communications
 - ii. Incorporate pedagogy – how to use these resources in the classroom/curriculum
 - iii. The team will meet in mid-December to sketch the series; we were awarded the requested amount for our 2020 proposal, so the proposed speakers and workshops are already outlined.
- b. Richard Johansen, UC's Data Visualization specialist will join our team to promote and utilized the new data visualization wall in the Geology-Math-Physics library
- c. Gather more pre- and post- faculty input
- d. Work with the Faculty Enrichment Center to integrate with and promote their mission and objectives.
- e. Provide parking passes to faculty so we can host the events on both East and West campus and encourage participation from satellite campuses.

VI. DCSS 2020 Proposal

This year's proposed activities for the *Data and Computational Science Series 3.0* includes an extension of the 2.0 events along with new content, encompassing the XSEDE monthly HPC workshops, GIS Day, an Advanced Research Computing workshop series, a manuscript symposium vis-à-vis open access publishing and data sharing policies, a Cyberinfrastructure (CI) Day, a research faculty social and the flagship Data Day 2020 event. We plan to host most of these events in the Faculty Enrichment Center and request parking passes to encourage all campus participation (West, East, Blue Ash & Clermont). Planned areas of focus:

³ The XSEDE monthly HPC workshops maintained their own registration since they needed to provide users with free log-on accounts.

1. Advanced Research Computing (ARC) Initiative
2. XSEDE Monthly HPC Workshops
3. GIS Day
4. Manuscript Accepted!
5. Cyberinfrastructure (CI) Day
6. DCS² Social
7. Data Day 2020

We were awarded the full requested amount of \$24,589 and we look forward to a bigger and better 2020 series.

VII. Conclusion

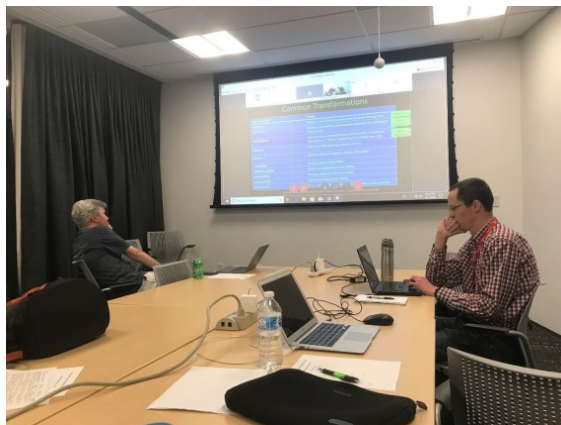
The team had some learning curves; organizing faculty-wide events is not for the faint of heart. However, our mixed levels of expertise and knowledge about how to coordinate and operate visiting guests and workshops at UC proved a good mix. The UCL and UCIT R&D representatives are in direct contact with UC research faculty, supporting their data and research infrastructure needs, thus this extra funding only enhances the services and resources that we can introduce and provide to increase the success of UC faculty research, teaching, funding, publications, community-building and retention.

Appendix A: University of Cincinnati Libraries Blog

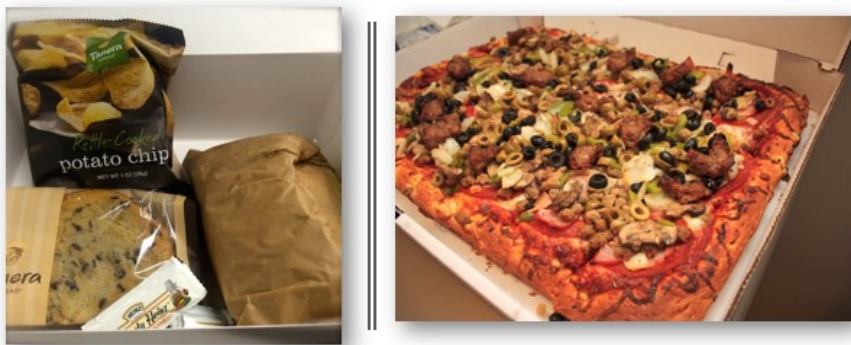
Student Research Consultant's experience at the XSEDE Big Data Workshop

Zhiyuan Yao is a Geography PhD candidate and a student research consultant in the Research Labs@GMP, in the Geology-Math-Physics Library. She is a GIS expert in the UC Libraries Research & Data Services Team and can help you with your GIS and other research data questions. Below she describes her recent experience attending the XSEDE workshop on Big Data.

The Extreme Science and Engineering Discovery Environment (XSEDE) is an NSF-funded virtual organization that integrates and coordinates the sharing of supercomputers with researchers nationally to support science. It has five computer resource infrastructures located across the nation, and Pittsburgh Supercomputing Center (PSC) is a part of it. PSC offers workshops from time to time throughout the year and aims to promote advanced data analysis among a variety of research fields using its supercomputer, Bridges. Thanks to [IT@UC Research & Development, Innovation and Partnerships](#), who cooperates with PSC to offer the free workshop to UC community. The workshop is one of many events in the [Data & Computational Science Series](#) funded by the Provost Office. I was glad to attend the Big Data workshop and give you a glimpse of the informative workshop.



The Big Data workshop was held on Oct 1st to Oct 2nd, from 11:00 AM to 5:30 PM. The workshop was well organized and followed a working flow: introducing background information, providing examples, and doing exercises. I really appreciated the background information introduced in this workshop, since it provided a holistic view of this workshop which was friendly to beginners. Besides, this workshop talked about several programs for big data analysis, such as Hadoop and Spark, and Tensorflow. The Big data workshop is one of PSC [HPC workshop series](#), such as MPI, OpenMP, OpenACC, et al. If you are interested in one of these workshops, you can check if IT@UC offers this workshop at [Workshops & Trainings in Faculty OneStop](#).



Besides the free knowledge, you also have free lunch! [IT@UC Research & Development, Innovation and Partnerships](#) offered the yummy lunch through the [Data & Computational Science Series](#) funded by the Provost Office. If you are interested in learning things about high performance computing, please come and take advantage of the free resources and the yummy lunch!

The next XSEDE event is coming up soon:

[XSEDE Monthly HPC Workshop: OpenMP](#)

Tuesday November 5, 2019 – 11:00am to 5:00pm
Langsam Library Room 475

[October 9, 2019 Amy Koshoffer](#)

The Data & Computational Science Series Presents: Research Computing: A View from the Trenches

Posted on behalf of Amy Latessa, Research Coordinator for IT@UC Research & Development

Coming up on Thursday November 7

The Data & Computational Science Series Presents: *Research Computing: A View from the Trenches*

Visualization Lab (GMP Library, 240 Braunstein Hall)

11am – 1pm: Lunch & Learn with George Turner, Chief Systems Architect for Research Technologies at the [Pervasive Technologies Institute](#), Indiana University will give an insider's presentation on research computing in higher education. George has helped build one of the most abundant and successful research centers in the United States and has been assisting the University of Cincinnati to set up its first ever [Advanced Research Computing](#) Cluster, now available to UC faculty and students. His talk will address challenges in research computing and HPC and share the benefits that a successful University research computing center can have upon a research institution.

1:30pm – 3:30pm: Linux 101 Workshop: This free workshop will provide a brief introduction on the layout of the Linux operating system. It will overview Linux concepts with a few exercises and will help people to understand the paradigms and why this is important in the bigger scheme of things. No prerequisites. No scripting. Please Bring a Laptop.

All events are free. [Registration is required.](#)

Flyer: [Turner DCSS Nov7](#)

Contact Amy Latessa at latessak@ucmail.uc.edu for more information

The Data & Computational Science Series presents – GIS Day 2019 Nov 13th

This year's UC's Celebration of National GIS Day is coming up on Nov 13th, 2019 and is organized by UC Libraries, UC Department of Geography and GIS, UC Joint Center of GIS & Spatial Analysis, UC Environmental Studies Program and the UC College of Arts and Sciences.

We invite you to come and celebrate with us. Lunch will be provided and this event is free and open to all.

Please register at <https://webapps2.uc.edu/ce/FacDev/Workshops/Details/13645>.

Dr Hao Hu, Product Engineer with ESRI is the featured speaker and will present on the intersection of Machine Learning and Geographic Information Sciences (GIS) and how ESRI is addressing these opportunities in spatial applications with ArcGIS. Dr Hao Hu is a product engineer in the Geoprocessing/Raster Analysis team at Esri, and works on large raster data processing that leverages distributed computing and storage technology. Before joining Esri, Dr. Hao completed his Doctoral degree in Geography from the University of Illinois at Urbana-Champaign (UIUC) researching on GIS and cyberinfrastructure, spatial decision support system, geospatial data uncertainty, and spatial optimization.

This talk is co-funded by the Provost Funded Data & Computational Science Series (awarded to UC Libraries and IT@UC Research and Development), the Graduate Student Association of UC and the Department of Geography – College of Arts and Science of UC.

The event will be held in 400 A-B TUC from 10 am to 2 pm

Hour	Talk
------	------

10:00am	Main talk (colloquium) -Machine Learning with ArcGIS
---------	--

10:45am Coffee break	
-----------------------------	--

11:00am	Practical session (Showcase) – Machine Learning in ArcGIS Platforms
---------	---

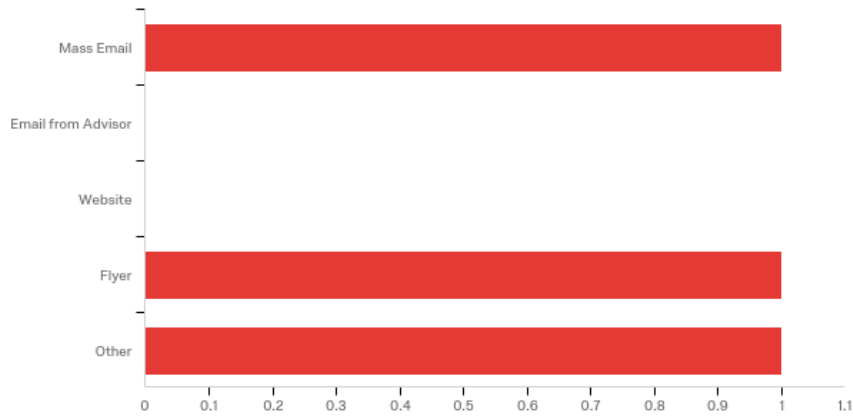
1:00pm Lunch	
---------------------	--

1:30pm	Link-GIS talk
--------	---------------

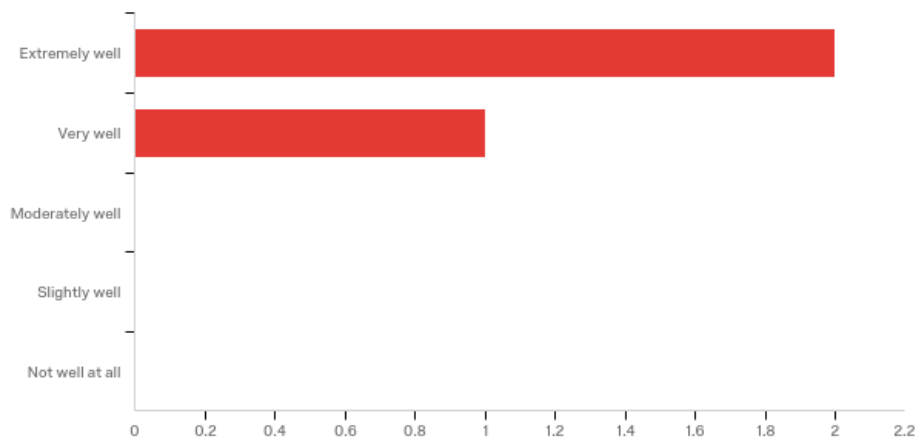
Appendix B: Post-workshop Survey Data

[XSEDE Big Data Feb 2019 Post-workshop survey](#)

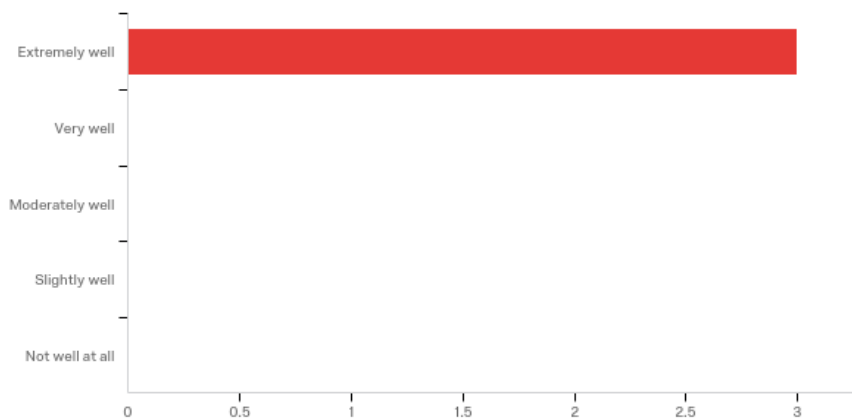
Q1 - How did you hear about the workshop?



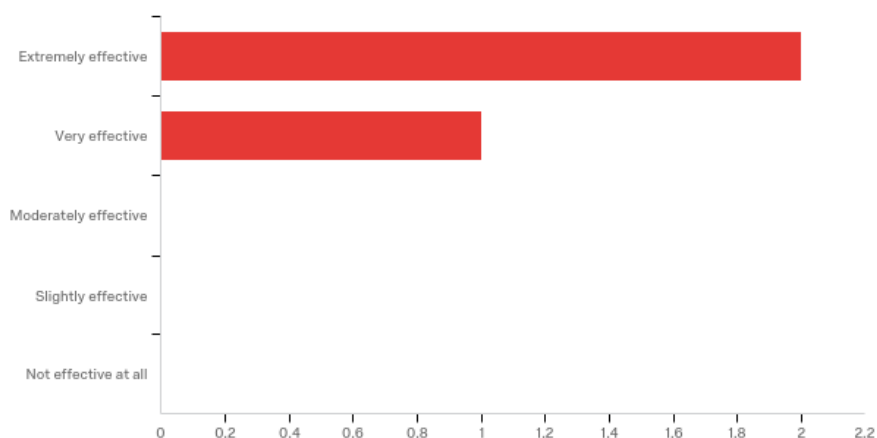
Q2 - How well did the workshop meet the description?



Q3 - How well did this class meet your personal objectives?



Q4 - Overall, how effective was the training session?



Q5 - What did you hope to learn in this workshop?

Introduction to Spark and Tensorflow applications.

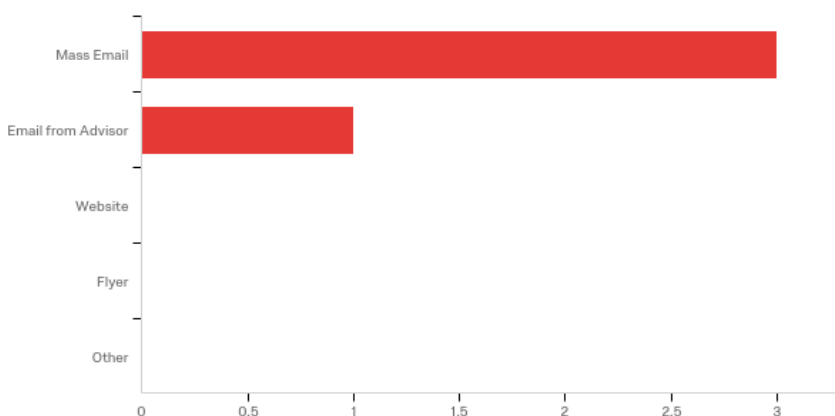
Understand the behind the scene of Machine Learning and Neural Networks.

Q6 - Please list suggestions for other courses, questions and/or comments:

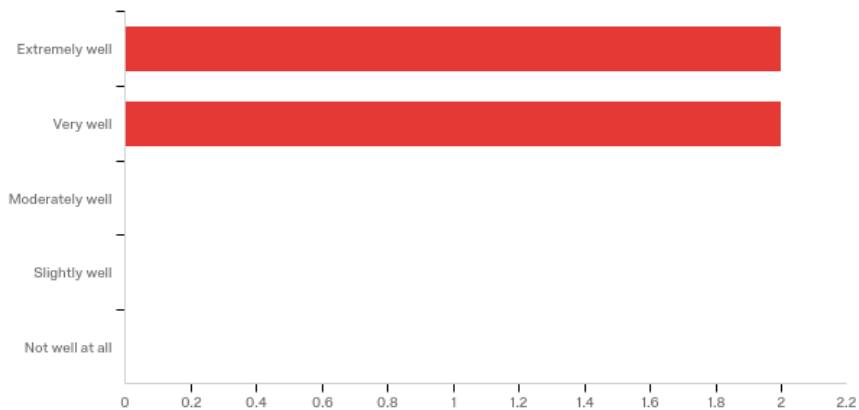
Bioinformatics

[XSEDE GPU Programming March 5, 2019_ Post-workshop survey](#)

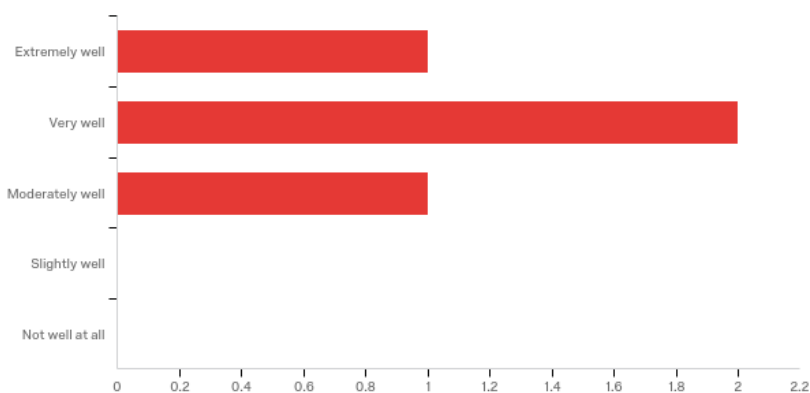
Q1 - How did you hear about the workshop?



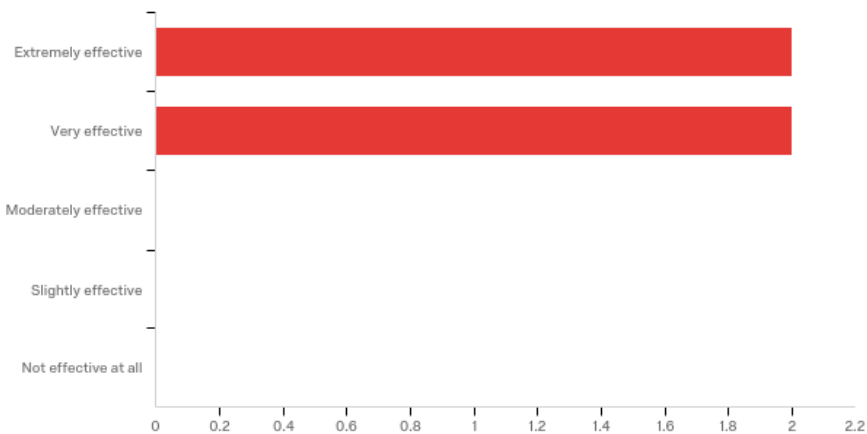
Q2 - How well did the workshop meet the description?



Q3 - How well did this class meet your personal objectives?



Q4 - Overall, how effective was the training session?



Q5 - What did you hope to learn in this workshop?

Just to refresh knowledge that I only half-assimilated from having taking a parallel computing course (using CUDA, on OSC's cluster) in college.

I hoped to learn how to use the GPUs in my simulation and also got experience using Bridges HPC cluster.

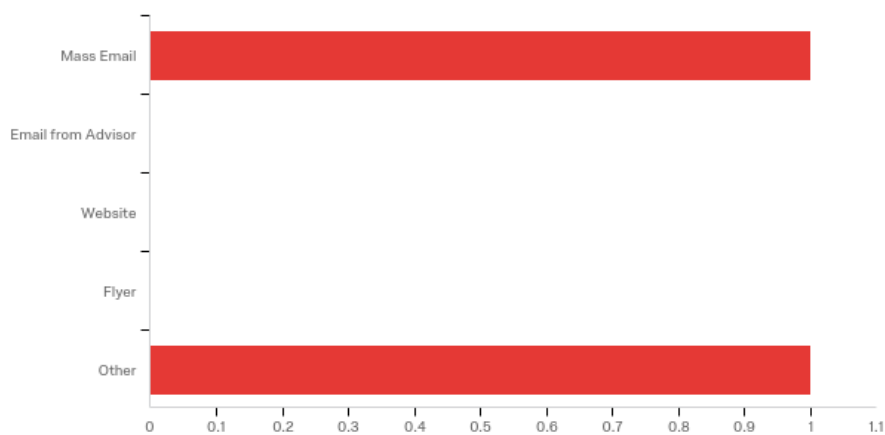
Q6 - Please list suggestions for other courses, questions and/or comments:

None that I can think of. Everything went pretty well. :)

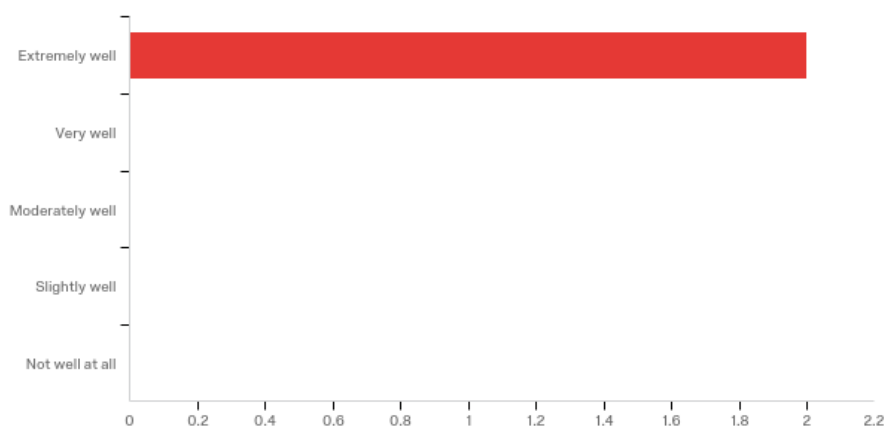
Hands on workshop always helps to teach you the skills and debug during that time.

[XSEDE HPC Boot Camp June 3-6, 2019_ Post-workshop survey](#)

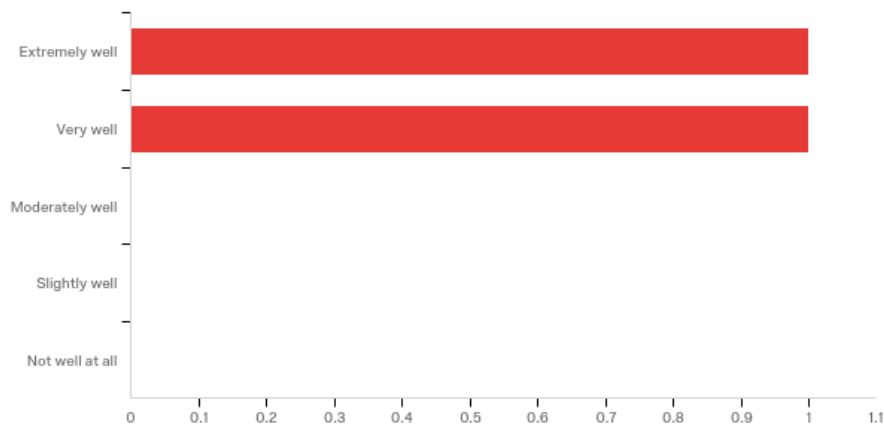
Q1 - How did you hear about the workshop?



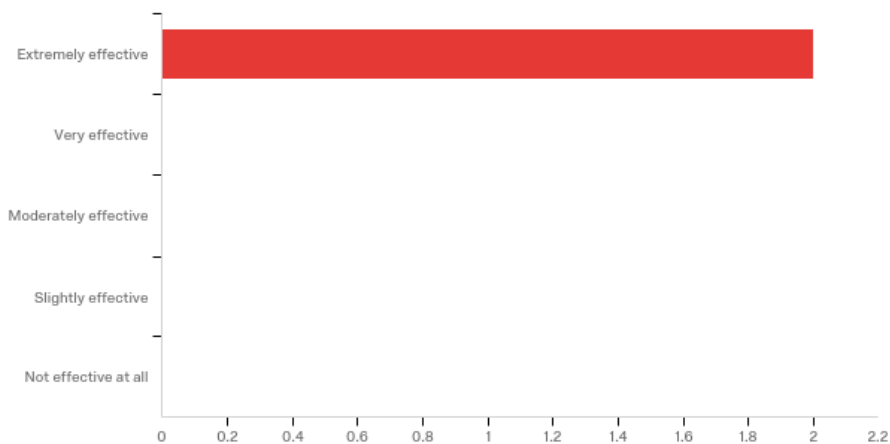
Q2 - How well did the workshop meet the description?



Q3 - How well did this class meet your personal objectives?



Q4 - Overall, how effective was the training session?



Q5 - What did you hope to learn in this workshop?


I hope to learn what parallel computing is and how can I take advantage of supercomputers through different kinds of methods.

Q6 - Please list suggestions for other courses, questions and/or comments:

The workshop introduced how we can get access to super computers by C and Fortran codes. Personally, I use Python and R programming in my research. I am curious if the workshop can talk a little bit of how we can get access to super computers by Python or R codes.


Machine Learning

Appendix C: Sample Flyer



Data & Computational Science Series *2019*

IT@UC and UC Libraries Presents:



George Turner
Chief Systems Architect for Research Technologies
Pervasive Technologies Institute, Indiana University


Thursday November 7, 2019

11:00AM to 12:00PM	Lecture	<i>Research Computing: A View from the Trenches</i>
12:00PM to 1:00PM	Luncheon	<i>Meet and Greet with George Turner</i>
1:30PM to 3:30PM	Workshop	<i>Linux 101 presented by Himakar Ganti</i>

Lecture, Luncheon and Workshop will be held in the Visualization Lab inside the Geology-Math-Physics (GMP) Library, 240 Braunstein Hall

*These events are free and open to all. **Registration is required.***

Register for Lecture & Luncheon [HERE](#). Register for Workshop [HERE](#).
(UC Faculty Development OneStop)



The DCS² series is presented by UC Libraries and UCit Research & Development and is supported by an Universal Provider award from UC's Office of the Provost for Faculty Development.

Questions?
Amy Latessa, IT@UC research coordinator | latessak@uc.edu