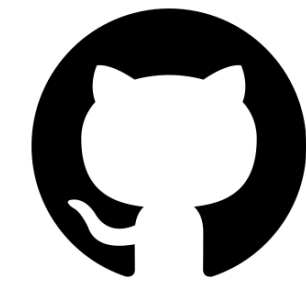


Building a student-driven graduate data science group

Sena Agezo, Tamara Dahl, Nael McCarty, Elyse Morin, David Nicholson, Stephanie Prince, Kaitlyn Roman, Lance Waller. Emory University.

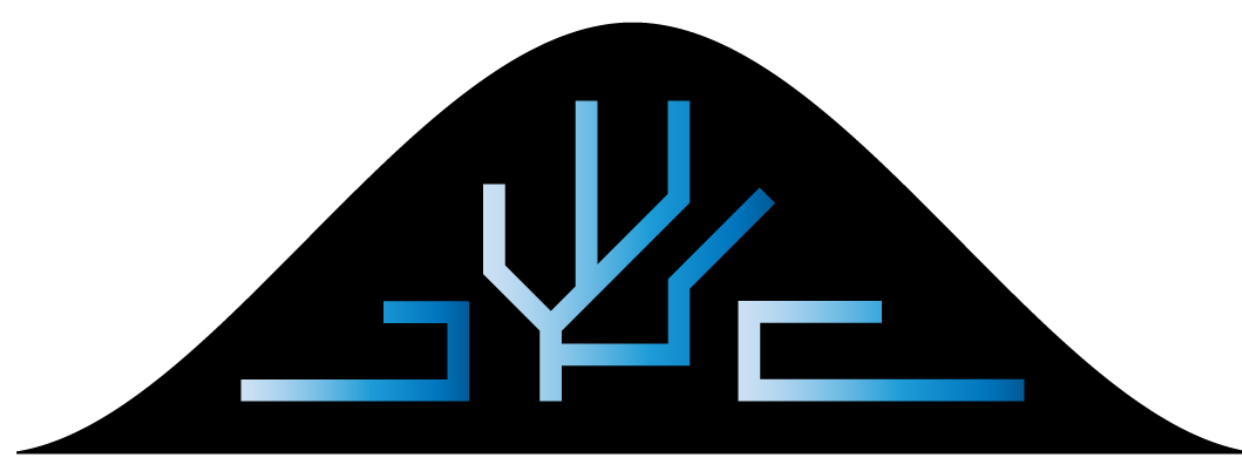


@datasci4sciATL



<https://github.com/Data-Science-for-Scientists-ATL>

Vision



DATA SCIENCE
for SCIENTISTS ATL

- A definition of **data science**:
 - Statistics and modeling, especially of large, noisy data sets
 - **and** reproducible analysis pipelines based on best practices adopted from open-source software
- Hence, graduate students need:
 - fluency in statistics and modeling
 - computational skills
- These skills also translate directly to careers outside academia
- To help graduate students acquire these skills, our group:
 - holds workshops, bootcamps, and short courses targeted to graduate students
 - meets regularly so members can
 - share knowledge
 - collaborate on projects
 - connect with invited guests from academia and industry



Activities

- Software Carpentry workshops
 - funded by Laney Graduate school and QTM
- General body meetings
 - academic speakers
 - Melinda Higgins, reproducible research with R and RStudio
 - Christian McDaniel and Andrew Durden from Quinn lab at UGA, SciPy 2018 presentations on biomedical data science
 - industry speakers
 - Jessica Owensby and Jay Lee from AT&T
 - Jeff Kolve, eHire, "Transitioning from academia to industry"
 - blog post from genetics graduate student Crystal Grant
- co-working sessions
 - based on Mozilla Study Group Handbook



Upcoming Events

- April General Body Meeting: Amber Willett, Chief Data Scientist, Terbium
- Workshop on version control with git and GitHub
- 1st annual datathon
 - modeled on ASA DataFest
 - two datasets: public health, demographic
 - sponsors include:
 - Emory graduate chapter of Association for Women in Science
 - Rollins School of Public Health Biostatistics and Bioinformatics



Some participants, helpers, and instructors from our second Software Carpentry workshop

Future Directions

- outreach
 - data science for public good projects
 - computational literacy workshops for other groups, e.g. high-schoolers, under-represented minorities, undocumented students
- short courses
 - e.g. based on <https://merely-useful.github.io>



Acknowledgements

Initial group development supported by the BEST grant (<http://best.emory.edu/index.html>) and the Laney Graduate Division of Biomedical and Biological Sciences (<http://www.biomed.emory.edu/index.html>). Funding for Data Science for Scientists ATL comes from Laney Graduate Student Council, with additional support from Rollins School of Public Health Biostatistics and Bioinformatics, Nell Hodgson Woodruff School of Nursing Center for Data Science, and the Institute for Quantitative Theory and Methods (QTM).