JSM 2016

Learning Communities: An Emerging Platform for Research in Statistics

August 2, 2016

Mark Daniel Ward
Purdue University

2014-15 STAT-LLC students
2015-16 STAT-LLC students

Introduction

- MCTP: Sophomore Transitions: Bridges into a Statistics Major and Big Data Research Experiences via Learning Communities

Purdue Statistics

NSF Living-Learning Community 2015-16
Purdue Statistics Living-Learning Community (STAT-LLC)

- Sponsored by National Science Foundation, Division of Mathematical Sciences
- $1.5$ million grant
- Fall 2014 through spring 2019
- Vision: Complete revamp of how we teach data analysis at the undergraduate level

Living Learning Community

- 20 sophomore undergraduates per year
- 100 students (altogether) by 2018
- 12-month experience for each sophomore (August to August)
- 12 Co-PI’s and Senior Personnel from Purdue Statistics plus 56 letters of collaboration
Why did we choose sophomores?

- At Purdue: 55 first-year learning communities with 78 total sections.....
- but only 3 sophomore-year LC’s!
- Only 13% of universities offer sophomore/junior seminars, while 90% offer first-year and senior seminars. (Berrett, 2012)
- “In 2002-03, only 8 percent of NSF-funded undergraduate researchers were sophomores” —Wilson and Crowe, 2010

Sophomore Slump

- Sophomores are referred to in literature as:
  - “Abandoned”
  - “Forgotten”
  - “Invisible to the institution”
- Sophomore year is a time of transition:
  - Finish(ing) calculus
  - Diving into most foundational courses in major
  - Identifying or refining long-term job goals
Attrition

- The sophomore slump leads to attrition of
  - Females
  - Minorities
  - Students with disabilities
  - First-generation students

- Especially in the mathematical sciences...

- Sophomore year is a time of transition!

Prerequisites (or lack thereof!)

- Only 1 prerequisite: Calculus III (MA 26100) before the start of their sophomore year.

- No computational or statistics requirements

- US Citizens/Nationals/Permanent Residents

- Students from any major can apply

- For 2016-17, ≈ 60 apply, but only 20 stipends
Blended Experience

- Sophomore year experiences that blend these 4 are unknown, nationwide.
  - Academic (Curricular)
  - Research
  - Residential Life
  - Professional Development

Research with Faculty

- *Students choose* from 50+ research projects, each with a dedicated faculty mentor.
- They all have relevance to Big Data analysis.
- Not a large spreadsheet, but (often) gigabytes or terabytes of data.
Research Thrusts

- Atmospheric Science
- Biostatistics
- Healthcare Engineering
- Probability/Theoretical Statistics
- Human Development and Family Studies
- Consulting
- Coastal Margin Observation and Prediction
- Saving Nature with Statistics (Forestry and Natural Resources)

Other Department Collaborators

- Biology, and Biomedical Engineering
- Chemical Engineering, and Chemistry,
- Coastal Margin Observation and Prediction,
- Computer Science,
- Economics,
- Electrical and Computer Engineering,
- Forestry and Natural Resources,
- Health and Human Sciences,
- Mathematics, and Physics,
- Regenstrief Center for Healthcare Engineering
Research Groups

- Students are encouraged to participate in
  - Research literature groups
  - Computational projects
  - Large research group / laboratory meetings
  - Conference calls
  - Meetings with research visitors
  - Discussions with clinicians/practitioners

- Basically, students should learn all aspects of research and be valued team members.

Big Data research; supported by curriculum

- Probability:
  - fundamental; random variables

- Statistical Theory:
  - also fundamental; builds on Probability course

- Introduction to Big Data Analysis (new!)
  - directly impacts the research effort;
  - allows students to build comfort with comp. stat.
"Flipped" Courses

- Big Data Analysis and Probability are both taught in a "flipped" style
- 237 videos for probability theory
- Each video has full page of notes
- Big Data Analysis is project-oriented and students work in teams all semester.
Become very comfortable with computing

Proceeding cores, for data analysis needs.

384 GB of RAM, 50 TB of disk, and 24

Rack-mounted (remote access) server with

The data will not fit in laptop / desktop

Students have a paradigm shift

Introduction to Big Data Analysis

Final presentations from students (2 weeks)

XML (1 week)

SQL (2 weeks)

awk/regular expressions (1 week)

UNIX (1 week)

visualizing data (2 weeks)

R (6 weeks)

Platforms/tools taught in nuts and bolts style:

Introduction to Big Data Analysis
Residential Life

- Faculty as mentors beyond class/research.
- The life of a faculty can include family.
Faculty Fellow: Impacting students’ lives outside of the classroom

- Student dinners
  - Hillenbrand every Thursday
  - Occasionally at Ward family home.
- Makes a scientist’s non-academic life be more tangible
  (grasp full picture of being a scientist)

Wearing Many Different Hats
Wearing Many Different Hats
Big Data Workforce

- Students are prepared for data-driven workforce
  - Data Scientists
  - Graduate school (Stat, Math, CS, many applied areas)
  - Consulting, R&D, analytics, databases, etc...

- Also learn from networking with alumni

- Big picture of what matters in the workforce:
  - how to partner with teammates
  - communicate their results
  - problem solving (e.g., how to get unstuck)

Overnight Visit to Math Conf. at Rose-Hulman Inst. of Tech. in April
Overnight Visit to Math Conf. at Rose-Hulman Inst. of Tech. in April

New ASA DataFest competition

1st Annual Purdue Statistics branch of ASA DataFest
Long Term Future

- The Department of Statistics would love to keep the STAT-LLC operating, long after the NSF grant has expired.

- We have approximately 450 students pursuing a Statistics major.

- Purdue Statistics is one of the largest UG Stat programs... and many students are interested in pursuing data science!

Statistics Living Learning Community

Questions: Mark Daniel Ward  mdw@purdue.edu