

# Learning Communities for Data Analysis and Statistics Technology

**eCOTS: May 17, 2016**

Mark Daniel Ward (Purdue University)

This material is based upon work supported by the NSF under Grant No. DMS-1246818.

# 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 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

# 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
- In 2016-17, have  $\approx$  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 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.

# 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)

# 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.

# Introduction to Big Data Analysis

- Platforms/tools taught in nuts and bolts style:
  - R (6 weeks)
  - visualizing data (2 weeks)
  - UNIX (1 week)
  - awk/regular expressions (1 week)
  - SQL (2 weeks)
  - XML (1 week)
  - Final presentations from students (2 weeks)

# Introduction to Big Data Analysis

- Students have a paradigm shift
- The data will not fit in laptop / desktop
- Rack-mounted (*remote access*) server with 384 GB of RAM, 50 TB of disk, and 24 processing cores, for data analysis needs.
- Become very *comfortable* with computing

# Introduction to Big Data Analysis



# 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)



# Hillenbrand Residence Hall



# Big Data Workforce

- Students are prepared for data-driven workforce
  - Data Scientists: New kind of job—very “in demand”!
  - 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)

# Questions?

- Please feel welcome to contact me:

Mark Daniel Ward, Ph.D.

Associate Professor and Undergraduate Chair

Department of Statistics

Purdue University

[mdw@purdue.edu](mailto:mdw@purdue.edu)

phone: (765) 496-9563

- Our website contains much more information:

<http://llc.stat.purdue.edu>