Summer Institute in Statistics for Big Data (SISBID)

PIs: Ali Shojaie & Daniela Witten
University of Washington

sisbid.uw.edu
UW Biostat’s Summer Institutes

• Existing:
  • Statistical Genetics
  • Modeling Infectious Disease
• BD2K R25: Statistics for Big Data

http://www.biostat.washington.edu/suminst
Format of SISBID

• In-person instruction
• Five courses, or “modules”
• Each module has 2 world-class instructors
• Each module is 2.5 days long:
  – Monday 8 AM – Wednesday 12 PM
  – Wednesday 1:30 PM – Friday 5:30 PM
• July 2015, 2016, 2017
What is “Statistics for Big Data”?

- Data Wrangling
- Data Analysis
- Data Visualization
- Reproducible Research
What is “Statistics for Big Data”?

Module 1

Module 2

Modules 3 & 4

Module 5
Module 1: Big Data Wrangling

Jeffrey Leek (Johns Hopkins) & Andrew Jaffe (Johns Hopkins)
Module 2: Data Visualization

Dianne Cook (Monash University) & Heike Hofmann (Iowa State)
Module 3: Supervised Learning

Noah Simon & Ali Shojaie or Daniela Witten (Univ. Washington)
Module 4: Unsupervised Learning

Genevera Allen (Rice University) & Yufeng Liu (UNC Chapel Hill)
Module 5: Reproducible Research

Keith Baggerly (MD Anderson) and Roger Peng (Johns Hopkins)
Module Themes

• Applications of key concepts:
  – All modules involve R coding.
  – Lectures & hands-on labs.
  – Topics motivated by case studies.

• Accessible to the broader community:
  – Course materials are available on Github.
  – Recorded lectures from 2015 are free online.
Module Pre-Reqs

• Basic knowledge of statistics and probability
• Prior experience programming

*Can fulfill all pre-reqs by taking courses in Johns Hopkins Data Science MOOC!*
Participants

• All levels:
  – Students
  – Post-docs
  – Faculty
  – Research staff

• All backgrounds:
  – Academia
  – Industry
  – Government
  – Non-profits

• Biologists and (classically trained) statisticians

• Average 100 participants/module in 2015 and 2016
Has the way you think about statistics changed as a result of your attending this Summer Institute?

- No (0)
- Yes (101)
Per-Module Student Evaluations

The course content was:
- Excellent (23)
- Very Good (22)
- Good (13)
- Poor (1)
- Very Poor (0)

The course content was:
- Excellent (17)
- Very Good (27)
- Good (12)
- Poor (1)
- Very Poor (0)

The course content was:
- Excellent (36)
- Very Good (29)
- Good (7)
- Poor (1)
- Very Poor (0)

The course content was:
- Excellent (26)
- Very Good (27)
- Good (5)
- Poor (0)
- Very Poor (0)

The course content was:
- Excellent (17)
- Very Good (13)
- Good (11)
- Poor (2)
- Very Poor (0)
R25 Funding

• Instructors: expenses and stipend
• Participants: travel & registration scholarships
  – In 2016, 50% of participants received full registration scholarships
Lessons Learned

• In-person instruction
  – Students like it!
  – But it’s a lot of work.

• 2015 lecture recordings free online
  – Students like it!
  – But it was a lot of work.

• Critical to piggy-back on UW Biostat’s existing summer institute program.
Thank You!!

sisbid.uw.edu