Succeeding as a Statistician in a Liberal Arts College

Nicholas Horton, nhorton@amherst.edu
April 23, 2019

Same story as Murali and Karen's advice

- Career planning
- Mentoring
- Publication process
- People skills and communication
- Professional service and leadership
- Ensuring emotional research
- Boundaries and balance
- Have fun!

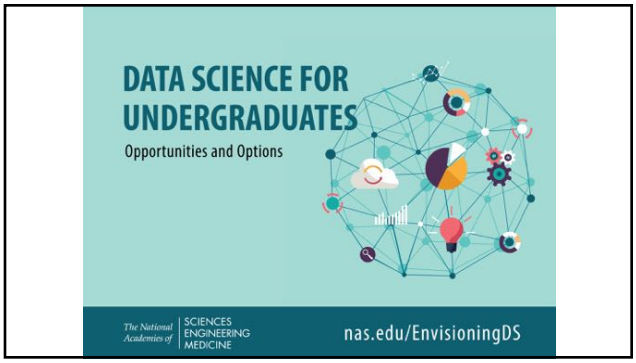
Amherst College Statistics

SEEKING TRUTH... THROUGH DATA...

SINCE 1821

Amherst College in a Nutshell

- n=1800 students, 200 faculty
- Talented and diverse students
- 45% of students are domestic students of color
- 20% Pell Grant eligible
- 11% first generation students
- 1 faculty member in probability or statistics in 2004
- 6 faculty members (5 tenure track plus one permanent lecturer) now
- Thriving statistics major



Finding 2.3 A critical task in the education of future data scientists is to instill data acumen. Key concepts include:

- Mathematical foundations
- Computational foundations
- Statistical foundations
- Data management and curation
- Data description and visualization
- Data modeling and assessment
- Workflow and reproducibility
- Communication and teamwork
- Domain-specific considerations
- Ethical problem solving.

Finding 2.3 A critical task in the education of future data scientists is to instill data acumen. Key concepts include:

- Mathematical foundations
- Computational foundations
- Statistical foundations
- Data management and curation
- Data description and visualization
- Data modeling and assessment
- Workflow and reproducibility
- Communication and teamwork
- Domain-specific considerations
- Ethical problem solving.

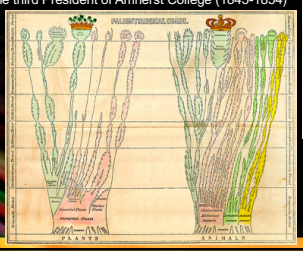
Data science: a liberal art with a long history

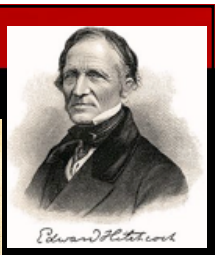
Learning from the past: founding members of the ASA

Hon. Stephen C. Phillips	Salem
Abel L. Pierson, M.D.	Salem
Benjamin Merrill, Esq.	Salem
Rev. Charles W. Upham	Salem
Asahel Huntington, Esq.	Salem
Elisha Bartlett, M.D.	Lowell
Luther V. Bell, M.D.	Charlestown
Hon. Caleb Cushing	Newburyport
Prof. Edward Hitchcock	Amherst
Prof. Joseph Alden, D.D.	Williamstown
Josiah Noyes, M.D.	Needham
Emory Washburn, Esq.	Worcester
Hon. William D. Williamson	Bangor, Me.

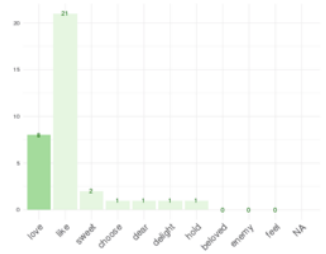
Amherst College's Edward Hitchcock

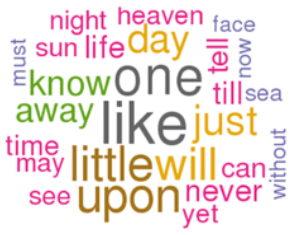
- Geology professor (who liked to count!)
- Later became third President of Amherst College (1845-1854)





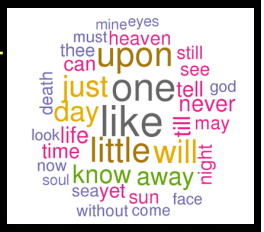
STAT231 (Data Science) text analytics





Emily Dickinson

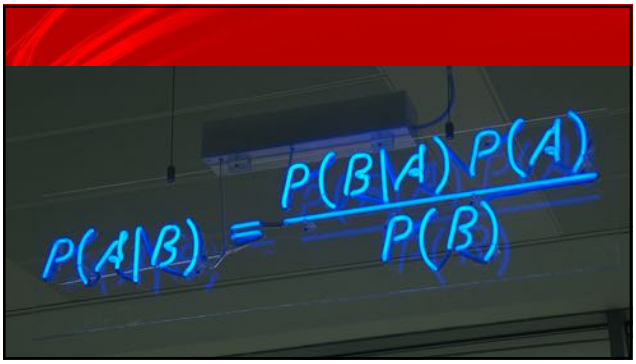
- https://r.amherst.edu/apps/nhorton/Dickinson1/
- https://r.amherst.edu/apps/nhorton/shiny-dickinson/



Hope is the Thing With Numbers: Poems and Stats

What is happening at Amherst? Statistics major

- First four statistics majors graduated in 2015, four in 2016, 10 in 2017, 21 in 2018, 21 in 2019
- New capstone course (Advanced Data Analysis)
- New 200-level Data Science course
- Increased focus on data wrangling, ethics, and communication in all courses
- Computing and workflow integrated into year-long probability/theoretical statistics sequence
- Comprehensive project to assess completion of the learning outcomes of the major



Curriculum (as of 2019)

- Intro Stats with Modeling (STAT135)
- Intermediate Statistics (STAT230)
- Data Science (STAT231)
- Probability (STAT360)
- Theoretical Statistics (STAT370)
- Advanced Data Analysis (STAT495)
- Electives: Multivariate (STAT240), Statistical Communication (STAT320), Epidemiology and Causal Inference (STAT340), COSC 247 (Data Mining), COSC 255 (Databases)

Statistics and Data Science Fellows program

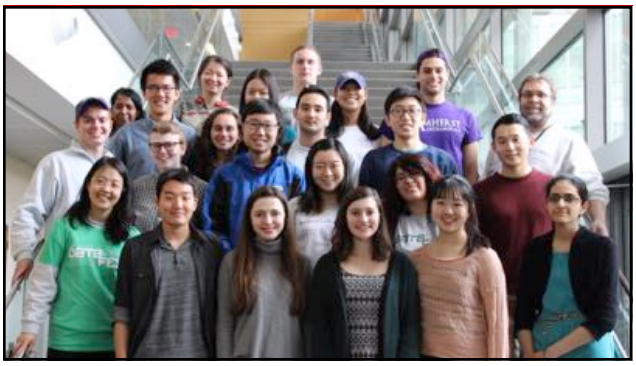
- Instituted in 2013 to create a cohort of students that could address data challenges across the campus
- Now expanded to ten students (rising sophomores, juniors, and seniors) with competitive selection process
  - Provide drop-in hours for intro stats and data wrangling questions
  - Organize the Festival of R and R Markdown each semester
  - Prepare and present workshops to lead up to DataFest
  - Conduct data-related projects across campus

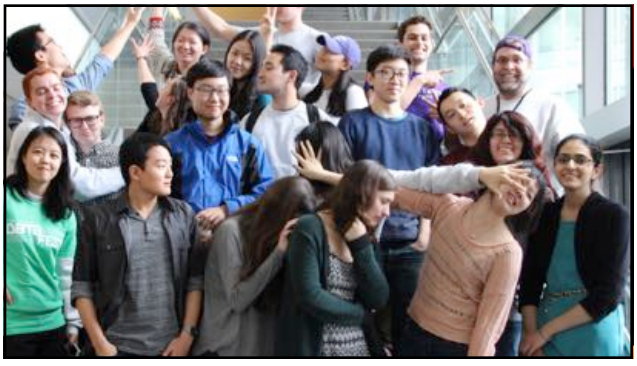
  All under the direction of the Statistics Faculty

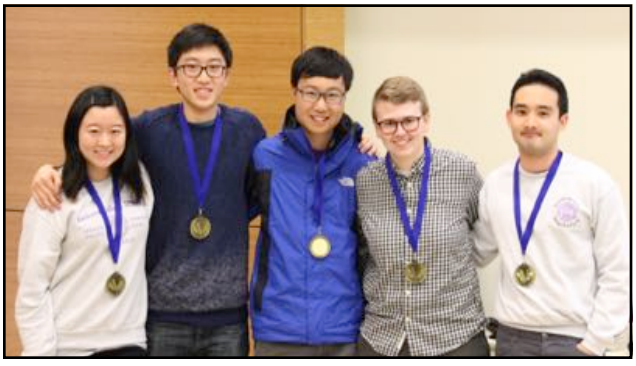
Moss Quantitative Center module development

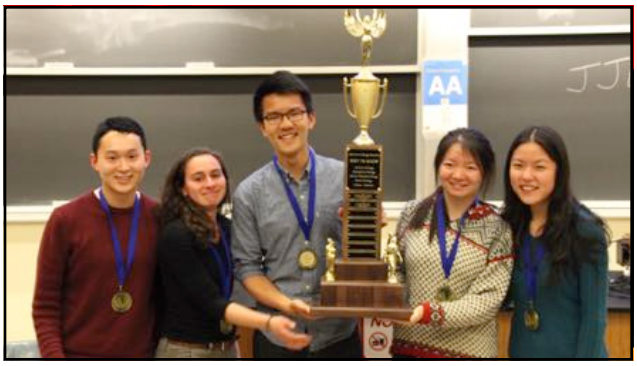
- Pairing up faculty to develop quantitative modules for social science and humanities courses
- Austin Sarat (LJST) “America’s Death Penalty” (COLQ234): logistic regression models to disentangle predictors of receiving a death penalty sentence
- Kerry Ratigan (Government) “Politics of Protest” (POSC330): focus on survey methods, multiple regression, and data visualization
- Amy Coddington (English) “Digital Humanities”

Data Fest: weekend-long data challenge









"Research" College Model

- 2-2 course load, 20-30 students per course, intensive engagement
- Hard-money (9 month) contract + 2 months possible summer salary
- Resources to support projects with students, conference travel, related research expenses
- Many opportunities for collaboration on campus (and beyond)
- Modern computation available (grid and cluster computing)
- Regular sabbaticals (one semester every three years)
- Research expectations (ongoing scholarly engagement)

Questions?

Succeeding as a Statistician in a Liberal Arts College

Nicholas Horton, nhorton@amherst.edu