

Data Science for Undergraduates

Nicholas Horton

2015-07-23T21:13:14-05:00

Data Science for Undergraduates

As the first entry in this blog, we thought it would be appropriate to begin with the 2018 consensus report “Data Science for Undergraduates: Opportunities and Options”. Nick was a co-author of this National Academies Report and it provides an accessible overview of undergraduate data science courses and programs. Co-chairs of the committee were Laura Haas (University of Massachusetts/Amherst, <https://www.cics.umass.edu/faculty/directory/haas-laura>) and Al Hero (University of Michigan, <https://hero.engin.umich.edu>). The report can be downloaded for free from <https://nas.edu/envisioningds>

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

CONSENSUS STUDY REPORT

DATA SCIENCE FOR UNDERGRADUATES OPPORTUNITIES AND OPTIONS



From the blurb: As our economy, society, and daily life become increasingly dependent on data, work across nearly all fields is becoming more data driven, affecting both the jobs that are available and the skills that are required. At the request of the National Science Foundation, the National Academies of Sciences, Engineering, and Medicine were asked to set forth a vision for the emerging discipline of data science at the undergraduate level. The study committee considered the core principles and skills undergraduates should learn and discussed the pedagogical issues that must be addressed to build effective data science education programs. *Data Science for Undergraduates: Opportunities and Options* underscores the importance of preparing undergraduates for a data-enabled world and recommends that academic institutions and other stakeholders take steps to meet the evolving data science needs of students.

What did the report find? What implications are there for educators working to help their students extract meaning from data? What key concepts and capacities are needed at the undergraduate level?

While the authors (wisely) didn't define data science, chapter two of the report defined *data acumen*:

Finding 2.3: A critical task in the education of future data scientists is to instill **data acumen**. This requires exposure to *key concepts in data science, real-world data and problems* that can reinforce the limitations of tools, and *ethical considerations* that permeate many applications.

Key concepts involved in developing data acumen include the following:

- 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, and (last but not least...)
- Ethical problem solving.

Each of these areas has an associated list of concepts and topics. As an example, *Data Management and Curation* includes the following:

- Data provenance;
- Data preparation, especially data cleansing and data transformation;
- Data management (of a variety of data types);
- Record retention policies;
- Data subject privacy;
- Missing and conflicting data; and
- Modern databases.

A number of programs and departments have been using the report as they consider changes to their statistics and data science curricula. More such work is needed.

Many other resources are available in addition to the report at <https://nas.edu/envisioningds>, including webinars (slides + recording) on a number of a key topics:

- overview of the study
- building data acumen
- incorporating real-world applications
- faculty training and curriculum development
- communication skills and teamwork
- inter-departmental collaboration and institutional organization
- ethics
- assessment and evaluation
- diversity, inclusion, and increasing participation
- two year colleges and institutional partnerships

While reading the full report or skimming these resources will take longer than our usual 20-30 minutes we think that it will be worthwhile to do so as a way of developing a baseline familiarity with the state of the art for undergraduate data science circa 2018.

Learn more

- <https://nas.edu/envisioningds>
- <http://nas.edu/dsert>
- <https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-060116-053930>