Computational & Data Sciences

The Division of Computational & Data Sciences (DCDS) at Washington University in St. Louis trains students interested in problems from across a range of disciplines that share a common reliance on data and computing.

The introduction of now-standard tools from statistical analysis and hypothesis testing transformed the practice of natural and social science in the mid-20th century. Emerging tools from computational and data science have the potential to bring about an even larger transformation of scientific practice, especially in the social sciences. The questions raised by data generated by and about human behavior are engaging and profound. However, many if not most of these questions can only be tackled using a multidisciplinary approach that combines a deep knowledge of the capabilities and operation of data science techniques with the domain expertise needed to apply them effectively to the problems under consideration.

Doctoral students in Computational & Data Sciences receive strong methodological training in modern computational and statistical methods, and they also acquire expertise in a particular social science application area.

The program is inherently interdisciplinary and brings together leading experts from across the university who are using data to solve some of the greatest challenges that our world faces today. Faculty include both data and computing experts as well as domain experts from different application areas.

Faculty

Deanna Barch (https://psych.wustl.edu/people/deanna-barch)  
Professor and Chair, Psychological & Brain Sciences  
PhD, University of Illinois

Michael Bechtel (https://artsci.wustl.edu/faculty-staff/michael-m-bechtel)  
Associate Professor, Political Science  
PhD, University of Konstanz

Ryan Bogdan (https://psych.wustl.edu/people/ryan-bogdan)  
Associate Professor, Psychological & Brain Sciences  
PhD, Harvard University

Todd Braver (https://psych.wustl.edu/people/todd-braver)  
Professor, Psychological & Brain Sciences, Radiology, and Neuroscience  
PhD, Carnegie Mellon University

Derek Brown (https://brownschool.wustl.edu/Faculty-and-Research/Pages/Derek-Brown.aspx)  
Associate Professor, Brown School  
PhD, Duke University

Sanmay Das (https://engineering.wustl.edu/Profiles/Pages/Sanmay-Das.aspx)  
Track Chair, Computational Methodologies  
Associate Professor, Computer Science & Engineering  
PhD, Massachusetts Institute of Technology

Brett Drake (https://brownschool.wustl.edu/Faculty-and-Research/Pages/Brett-Drake.aspx)  
Professor, Brown School  
PhD, UCLA

Christine Ekenga (https://brownschool.wustl.edu/Faculty-and-Research/Pages/Christine-Ekenga.aspx)  
Assistant Professor, Brown School  
PhD, New York University

Patrick Fowler (https://brownschool.wustl.edu/Faculty-and-Research/Pages/Patrick-Fowler.aspx)  
Track Chair, Social Work & Public Health  
Associate Professor, Brown School  
PhD, Wayne State University

Roman Garnett (https://engineering.wustl.edu/Profiles/Pages/Roman-Garnett.aspx)  
Assistant Professor, Computer Science & Engineering  
PhD, University of Oxford

Chris Gill (https://engineering.wustl.edu/Profiles/Pages/Christopher-Gill.aspx)  
Professor, Computer Science & Engineering  
DSc, Washington University in St. Louis

Roch Guérin (https://engineering.wustl.edu/Profiles/Pages/Roch-Gu%C3%A9rin.aspx)  
Professor and Chair, Computer Science & Engineering  
PhD, California Institute of Technology

Shenyang Guo (https://brownschool.wustl.edu/Faculty-and-Research/Pages/Shenyang-Guo.aspx)  
Professor, Brown School  
PhD, University of Michigan - Ann Arbor

Ross Hammond (https://brownschool.wustl.edu/Faculty-and-Research/Pages/Ross-Hammond.aspx)  
Associate Professor, Brown School  
PhD, University of Michigan

Jenine Harris (https://brownschool.wustl.edu/Faculty-and-Research/Pages/Jenine-Harris.aspx)  
Associate Professor, Brown School  
PhD, Saint Louis University

CJ Ho (https://cse.wustl.edu/faculty/Pages/faculty.aspx?bio=116)  
Assistant Professor, Computer Science & Engineering  
PhD, University of California, Los Angeles
Upon joining the PhD program, each student is assigned an initial adviser from the DCDS faculty. This adviser meets with the student to assess their background and to advise them on course selection. Immediately prior to each fall semester (starting in 2019), DCDS faculty conduct a “boot camp” in
mathematics, statistics and programming to help bring incoming students up to the level needed to succeed in the initial course work and the program.

All students complete a common core curriculum as well as a domain depth requirement in a social science area. The focus of the first year is on acquiring a common set of tools and an understanding of the ranges and types of problems students may work on as they progress through the program. The entire incoming cohort takes a unique two-semester seminar sequence solely for DCDS students, which includes both general topics and a series of data-driven dives into the types of research questions that may be encountered in each of the domain areas.

In addition, students will be exposed to research in different areas through “rotations” that start in November of their first year. By the end of the summer following their first year, each student will put together an advisory committee of at least two DCDS faculty members (preferably from different tracks) and identify the specific track in which they plan to do research and pursue their degree.

**Curriculum**

**Required Core Courses (24 credit hours)**

- **CSE 502 (3 credits):** This is an existing fundamental course in algorithms and data structures, including significant implementation in an object-oriented programming language (currently Java). We expect that many students will already have this background; the course is intended as a pathway for students with very little computational training.

- **Quantitative Methods (QM) I and II (6 credits):** This two-semester sequence covers essential probability and statistics, including hypothesis testing, inference and experimental methodology using a modern statistical computing language like R. The introductory courses offered by the departments of Psychological & Brain Sciences (PBS 5066) and Political Science (PS 581) will be cross-listed and count for QM I credit. QM II is a course that includes maximum-likelihood methods, Bayesian and nonparametric models, generalized linear models and sampling techniques. The course is currently taught as Political Science 582 and will be cross-listed across participating departments.

- **CSE 5XX: “Data Wrangling”:** We are in a new era in terms of the volume and modalities of data generated by efforts to measure human behavior. This will be a new cross-listed course that introduces students to the tools and techniques used to collect, maintain and process large-scale data sets of the kind generated in the course of studying people and social systems.

- **Machine Learning I and II: CSE 417T and 517A (3 credits):** This is a two-semester sequence in machine learning. Together, the two courses cover the fundamental principles of supervised learning, including generalization, overfitting, regularization, cross-validation, model selection, core ML techniques and algorithms, including linear models like logistic regression, gradient descent, tree-based and ensemble methods, kernel methods, deep neural networks and topics in unsupervised learning.

- **Computational and Data Sciences (CDS) Seminar I and II (6 credits):** This two-semester seminar sequence is cross-listed across participating departments and team-taught by participating faculty.

  - **CDS Seminar I will be structured around topics and ideas that do not need detailed specific-content background. The topics covered will include ethics, the nature of research, robustness and reproducibility of research, and presentations from across the different areas of interest to give students an understanding of research in human and social data analytics across the university.**

  - **CDS Seminar II will be structured as a series of deep dives into data-driven approaches in each of the domain areas, including a module on computational methodologies. In each of these modules, the students will either be given a specific data set to investigate or a specific hands-on task to complete (e.g., developing a visualization, assessing how easy a computational tool is for social scientists to use). Students will work in teams on these projects.**

**Domain Depth Tracks**

Students will choose one of four focus tracks: Political Science, Psychological & Brain Sciences, Social Work & Public Health, or Computational Methodologies. Depending on the track, students must complete the following domain depth requirements:

1. **Political Science track:** Students must complete three substantive classes in one subfield (e.g., American politics, comparative politics, international relations) from a specified list for each subfield as well as a research design course (PS 540).

2. **Psychological & Brain Sciences track:** Students must complete three substantive classes in one subfield (e.g., brain, behavior and cognition, clinical science, social/personality, development and aging). With permission, students may substitute the Psychological & Brain Sciences Research Methods Course (PBS 5011) for one of the substantive classes, depending on their background in psychological science.

3. **Social Work & Public Health track:** Students must complete a three-course core doctoral seminar series, including conceptual foundations of social science, advanced research methods, and a theory seminar, either in public health or social work. Students will also be required to take an advanced substantive course from an approved list in their area of interest.
4. **Computational Methodologies track**: Students must take CSE 541T Advanced Algorithms and either CSE 511A Introduction to Artificial Intelligence or CSE 515T Bayesian Methods in Machine Learning. In addition, students must take two substantive classes in their area of interest (i.e., political science, psychological & brain sciences, or social work & public health) from among the classes acceptable for students in that track as noted above.

**Sample Curriculum**

A typical progression of classes is described below, with separate examples for students who enter with and without more extensive computational backgrounds.

**Students Without Much Computer Science Background**

<table>
<thead>
<tr>
<th>Units</th>
<th>Fall</th>
<th>Spring</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Structures and Algorithms (CSE 502N)</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Quantitative Methods I</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CDS Seminar I</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Introduction to Machine Learning (CSE 417T)</td>
<td>—</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Data Wrangling</td>
<td>—</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>CDS Seminar II</td>
<td>—</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Second Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative Methods II</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Domain course</td>
<td>3</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Domain course or elective</td>
<td>3</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Machine Learning (CSE 517A)</td>
<td>—</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

**Students With More Computer Science Background**

<table>
<thead>
<tr>
<th>Units</th>
<th>Fall</th>
<th>Spring</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduction to Machine Learning (CSE 417T) or domain course</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Quantitative Methods I</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CDS Seminar I</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Introduction to Machine Learning (CSE 417T) or domain course</td>
<td>—</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Data Wrangling</td>
<td>—</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>CDS Seminar II</td>
<td>—</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

**Further Requirements**

Additional requirements for this program are as follows:

- A minimum of 72 credit hours beyond the bachelor’s level, with a minimum of 37 being course credits (including the core curriculum)
- A minimum of 24 credit hours of doctoral dissertation research
- Students must maintain a cumulative average grade of B (3.0 grade-point average) for all 72 credit hours.
- Required courses must be completed with no more than one grade below a B-.
- Up to 24 graduate credit hours may be transferred with the approval of the Graduate Studies Committee, chaired by the director of graduate studies.

In addition to fulfilling the course and research credit requirements, students must do the following:

- Complete at least two three-month-long research rotations.
- Pass a qualifying exam.
- Successfully defend a thesis proposal.
- Present and successfully defend a dissertation.
- Complete a teaching requirement consisting of two semesters of mentored teaching experience.