Overview of D2K Learning Lab
For Rice & External Researchers

What is the D2K Learning Lab?
In the D2K Learning Lab, researchers at Rice University or the Texas Medical Center can sponsor teams of students that work on a semester-long data science project defined by the sponsor.

- Sponsoring a D2K Learning Lab team is a great way to jump start a collaborative data science research project!
- Teams of 4-6 students work with the sponsor’s data and on the data science analysis or research problem. Teams consist of students from diverse disciplines and experience levels (e.g. undergraduates and graduate students) that are matched to projects based on interest and relevant skills.
- Teams produce a data science research or analysis report and deliver all software and scripts used in the report. Teams exclusively use open-source software such as R and Python.
- There is no sponsorship fee for researchers.
- Sponsors must agree to mentor the student team throughout the semester along with a Rice data science faculty co-mentor.
- Research sponsors can come from any discipline, but must be collaborating with or form a new collaboration with a Rice data science faculty member.

What types of projects are appropriate?

- Projects should be real-world, impactful research projects that would be meaningful to the sponsor and motivating for the students.
- Projects should be challenging, but with well-defined objectives. It is recommended that the objectives be tiered in nature with some that are easily achievable within a semester and with others that are a stretch to motivate and challenge the students.
- While project objectives can focus on any aspect of the data analysis pipeline, the most effective projects will encompass aspects of each of the four stages of data science:
  1. **Data Wrangling** is the process of cleaning and unifying messy and complex data sets for easy access and data analysis.
  2. **Exploratory Data Analysis** (or Data Mining) visually explores data and finds interesting features or patterns in large data sets.
  3. **Data Science Modeling** uses statistical or machine learning models to address specific objectives.
  4. **Communication & Validation of Results** is critical for confirming data-driven discoveries and assessing the accuracy of data science models.
The best projects are organized in such a way that student teams can get started on day one.

- While student teams will need to learn discipline-specific terms and discipline data complexities, the more these can be minimized or abstracted away, the more time student teams will be able to spend on the data science aspects.
- In cases where data-wrangling is a significant portion of the project, it is helpful to provide students with a subset of “cleaned” data that they can use as a model to begin thinking about how to address the project’s objectives.

What data is appropriate for the D2K Learning Lab?

- We encourage sponsors NOT to provide confidential data or divulge any confidential information to the student teams.
  - Confidential data requires additional protections and training for students and faculty that may delay work on the project.
  - For 2018-2019 academic year projects, we will not be able to support projects utilizing confidential or protected data.
- We encourage sponsors to share data that they own but that is not proprietary or confidential. The data should not be personally identifiable and should not be data protected by HIPPA or FIRPA (such as health, medical, or educational data).

What are the expectations for D2K Learning Lab sponsors & mentors?

- Research sponsors must apply to have their projects considered for the D2K Learning Lab. See application process and timeline below.
- Sponsors should fully prepare their project and data for the team. D2K Lab faculty may provide brief feedback on the projects after applications are due to make sure they are most effectively conceived for the D2K Learning Lab.
- Researchers from established or new collaborations are welcome to apply. Each research project for the D2K Learning Lab must consist of at least one faculty researcher from any discipline with data challenges and at least one Rice data science faculty.
- During the semester, both sponsor co-mentors are expected to meet with the student team at least twice a month, but preferably once a week.
Timeline for Spring 2019 D2K Learning Lab Projects

Before semester starts:
- **November 15, 2018** - Project applications due to d2k@rice.edu.
- **December 1, 2018** - Applicants notified of project acceptances.
- **January 1, 2019** - Project descriptions due. Data transferred to D2K Lab.

During Spring 2019 semester:
- **January 9, 2019 4pm - 5:30pm** – Orientation to the D2K Learning Lab and sponsors meet student teams.
- **April TBD, 2019** – Spring D2K Learning Lab Showcase.
- **May 1, 2019** – Final report and code due.
Application to Sponsor a D2K Learning Lab Project Spring 2019

For Rice & External Researchers

Please send the following application materials to d2k@rice.edu by November 15, 2018:

1. Biosketches of all project sponsors. These can be NSF or NIH-style biosketches.

2. Project description. Please provide a short (2-4 page) description of the project that includes:
   a. Project Title
   b. Project Pitch
      • Please discuss the background and significance of project in a manner that would be accessible to all students in the D2K Learning Lab.
   c. Project Description
      • Please discuss the project scope of work.
   d. Project Objectives
      • Please provide clear objectives for the D2K Learning Lab team.
   e. Data Description
   f. Personnel & Collaboration
      • Is this a new or established collaboration? If established, is this support by a grant? If new, are there plans to apply for funding?
   g. References