# Proposal for a Psychology graduate student “Data Science option”

## Summary Description

The proposed Data Science Option (DSO) is designed to meet a critical educational gap at the intersection of Psychology (PSYCH) and Data Science. Psychology graduate students will receive credentialed training in the analysis of large datasets. The Data Science option provides students an introduction to the world of data science, giving them the skills to understand a variety of techniques and tools. The goal of this option is to educate all students in the foundations of data science, so they may apply those methods and techniques in current research. The Psych DSO is designed for students with little or no background in data science, computer science or coding.

The requirements for the Psych DSO are as follows:

1. Courses from two out of three of the following areas:
2. Software development for data science
3. Statistics and machine learning
4. Data management and data visualization
5. 2 quarters of the eScience Community Seminar (<http://escience.washington.edu/get-involved/escience-community-seminar/>)
6. Fulfillment of the Psychology Department Statistics and General Methodology requirements. These are currently the following:

All students must achieve a grade of at least 2.7 or place out of (and replace with a higher level course) in the following courses:

**Psych 522 Laboratory in Statistical Computation I**

**Psych 523 Laboratory in Statistical Computation II**

**Psych 524 Introduction to Statistics and Data Analysis**

**Psych 525 Linear Models and Data Analysis**

These classes are taught annually and are part of the current 1st year requirements for Psychology graduate students.

In addition, ALL students are required to take at least one additional course in quantitative methods (statistics, mathematics, or computational science) OR in a methodology directly relevant to their area of focus. Such courses are to be specified in each student’s Individualized Training Plan. Some classes in the Data Science Option will meet this course requirement, including (though not limited to STAT 416, STAT 527, 535, 509, 512, 513)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

## 1. Software development for data science

HIGHLY RECOMMENDED:

1. Software Development for Data Scientists (CSE 583)
2. Software Engineering for Molecular Data Scientists (ChemE 546)

## 2. Statistics and machine learning

HIGHLY RECOMMENDED:

1. Introduction to Machine learning: CSE416/STAT416

ALTERNATIVES:

1. Nonparametric regression and classification (STAT 527)

ADVANCED OPTION:

1. Machine Learning: (CSE 546 or STAT 535) also serves for the “Advanced Data Science Option”
2. Introduction to Mathematical Statistics (STAT 509) and Statistical Inference (STAT 512-513) also serves for the “Advanced Data Science Option”

## 3. Data management and data visualization

HIGHLY RECOMMENDED:

1. Introduction to Database Systems (CSE 414)
2. Information for Visualization (HCDE 411/511)

ADVANCED OPTION:

1. Principles of DBMS (CSE 544) also serves for the “Advanced Data Science Option”
2. Data Visualization (CSE 512)  also serves for the “Advanced Data Science Option”

## eScience Community Seminar

Two quarters of the eScience Community Seminar

## Rationale for Adding a Data Science Option

The path to deep scientific discoveries is changing rapidly. Most disciplines, from physical to life sciences, have entered an era where discovery is no longer limited by the collection and processing of data, but by the management, analysis, and visualization of this information. Novel developments in instrumentation have led to a tremendous increase in the volume of this data, forcing scientists to perform analyses on data that is too big for standard desktop computing tools. Thus, rising scholars need the skills to process big data.

To harness the opportunities that big data brings, the next generation of scientists requires education both in a domain science and in methods for data management, analysis, and visualization. Thus, many graduate students in Psychology need an education that focuses on building the next generation of data science tools and knowledge in the application of these tools in a discipline-specific manner.

In addition, given the small number of tenure-track positions available to students and postdocs, experience in data science opens up additional career paths for graduate students.

## Administrative Location

The Psychology department will administer the DSO and the ADSO for its own students.

## Related Programs

See description under overall framework above.

## Timeline for Implementation

Our faculty voted to approve the Data Science Option beginning in Fall 2018.

## Relationship to Institutional Role, Mission, and Academic Unit Priorities:

Many departments have already approved the “Advanced Data Science Option” and the “Data Science Option”.

These two options will give graduate students a choice of how deeply they wish to be trained in the data science field.

The courses that form the curriculum underlying both options are all existing courses. The Psychology graduate program advisor and coordinator (GPC/GPA) will carry out student tracking and determine whether classes are eligible to meet requirements. The Psychology member of the eScience education committee member will act as an advisor to GPC/GPA.

## Documentation of Need for Program

A graduate “option” is the only transcriptable method to offer a credentialed recognition of our students’ efforts and to serve a recruiting tool for the units involved. We already have around 12 enrolled graduate students who have expressed a strong interest in taking the data science option – so we expect roughly 10-15% of our graduate student cohort to choose to participate in the DSO.

## Program Oversight

The Psychology graduate program director (GPC) and the eScience Education Working Group chair (currently Bing Brunton) will provide formal oversight to the program. The Psychology department eScience education committee member (currently Ione Fine) will play an advisory role.

To apply for the Psychology Data Science option, a student must be a full-time graduate student in Psychology. The student must send an email to the graduate advisor in their department and declare interest in pursuing the Data Science track. The research advisor of the graduate student must approve the application.

## Administration

Because the oversight required for students to participate in the option is minimal, we will not require any administrative resources nor other support services beyond the existing graduate advising staff in the participating departments.

## Students

Eligible students for the DSO include all full time students in the Psychology graduate program (PhD) in good standing who have completed the first year statistics requirements. There are no other requirements to apply.

In cases where a graduate student does not graduate but instead leaves only after completing the master’s part of the graduate program, the student will retain the “Data Science option” recognition on their transcript if the student finishes all the requirements of the option before leaving the graduate program. This point is important because the DSO opens the door to a larger number of employment opportunities in industry. We will advertise the availability of the option by posting information about the option on the department website.

## Diversity

Only students from participating departments will be able to register for the Data Science option. As a result, the option will directly leverage the departments’ efforts to increase diversity. Psychology has historically been very successful in its ability to recruit URM students into STEM.

## Program Assessment

The department will collect feedback from the students and their advisors at least once per year to get their input on the program. This will be done through an online survey and in-person interviews. Once students start graduating from the DSO program we will also collect data from students who have recently left the program.

The Psychology ASDO/DSO steering committee may also recommend changes to the program based on the feedback from the students and faculty from Psychology and other departments.