

NSF GRFP Fellowship Report

16-17 Richard Clark Fitzgerald

Please write a 2-3 paragraph SUMMARY of your fellowship activities and major accomplishments within the last year. This should be written for the public, and should address both the Intellectual Merit and the Broader Impact of your work.

In the fall I partnered with a fellow graduate student, Irene Kim, to initiate and organize a student led statistics department seminar. This broadened our collective awareness of active areas of statistics research. It also resulted in a greater sense of community among the graduate students. Additionally, I met with several groups of community college students interested in transferring and joining our department.

Last summer I wrote a software package called `rddlist` integrating R with Apache Spark through the use of distributed data structures. This was in the context of an internship with the R Consortium working on the `ddR` (distributed data in R) package. These projects are efforts towards the broader goal of building statistical software that scales up to larger amounts of data. All development was done in an open source environment, so the code is freely available for public use. I've also continued contributing to more established open source projects through patches and code review.

Throughout the year I presented several talks on parallel R programming techniques and concepts. This motivates and relates to my continued efforts researching code analysis and parallel programming. The first talk in October 2016 was through the NSF funded Research Training Group, which drew in a broad audience of mostly graduate students from different departments. Another talk was for the undergraduate organized `iidata` data science convention in February 2017. I went deeper into an example of GPU programming to show how to improve code run time by orders of magnitude. Professionals as well as undergraduate and graduate students from all across campus attended the talk.