Statistics 159 & 259 — Fall 2015 Project
Reproducible and Collaborative Statistical Data Science

Form teams	Sept. 17
Project proposal	Oct. 1
Progress presentation	Nov. 3 & 5
Draft report	Nov. 12
Project presentation	Dec. 1 & 3
Final report	Dec. 14

Learning objectives: Investigate a published fMRI study; collaborate on a software project; work with complex and large datasets; convolution (hemodynamic modeling, smoothing); interpolation (slice time correction, image resampling); optimization (registration, advanced statistics); basic linear algebra (statistics).

Project overview: The semester project involves "investigating" a published result using the analysis of functional Magnetic Resonance Imaging (MRI).[1] Functional MRI (fMRI) allows scientists to localize which parts of the brain are associated with specific cognitive tasks. There is no expectation that you will have a background in neuroscience. You will learn everything you need to know about the method in the course. The intention of focusing on fMRI data is merely to provide a concrete problem domain that exemplifies the types of programming and statistical challenges present in many modern statistical applications. Additionally, the weekly labs will prepare you—through a series of guided exercises—with the basic skills and background you need for the group project.

While you will learn the basic methods and tools during lecture and lab, you will be expected to do additional research and reading in the course of working on the group project. For instance, you may need to use a specialized analysis method or a Python package not covered in the lectures or labs.

As part of your final project grade, you will be required to work on your project using GitHub's pull request and code review mechanism. During lecture, I will cover the exact workflow you will be expected to follow. Please note that as this course is explicitly about reproducibility and collaboration your project grade will not be entirely based on your final report, but will also reflect how well your group work is reproducible and how effectively you collaborated using the techniques taught in the course (e.g., pull requests, code review, testing, etc.)

Forming teams: Teams will be decided by **Thursday, September 17th**. Each team should consist of 4 or 5 students. You should ensure that your team is composed of individuals with different strengths. For example, each team will need people with strong computational as well as

[1] Exactly what each team means by "investigate" will need to be defined by each team. For example, for one team this might mean a very careful reanalysis of the data using the original methods as closely as possible. However, this shouldn't mean merely running existing scripts used in the original analysis; rather, the team would need to reimplement the analysis scripts. Another team might decide to conduct a different analysis than used in the original study. For instance, the published work might use a parametric approach and the team project might attempt to use a nonparametric technique such as permutation testing. A third team might focus on a careful validation of the modeling assumptions made by the original analysis.

statistical skills.

Project proposal: Each team will submit a project proposal on **Thursday, October 1st**. The proposal should be written in LaTeXand submitted to your team's GitHub repository. Use this template `http://www.jarrodmillman.com/stat159-fall2015/project/proposal.tex`.

Progress presentation: Each team will present a 5 minute progress report in class on **Tuesday, November 3rd** or **Thursday, November 5th**.

Draft report: Each team will submit a written draft report by 21:00 **Thursday, November 12th**. For more information see the template `http://www.jarrodmillman.com/stat159-fall2015/project/report.tex`.

Project presentation: Each team will present a 5 minute progress report in class on **Tuesday, December 1st** or **Thursday, December 3rd**.

Final report: Each team will submit a final written report by 21:00 on **Monday, December 14**. For more information see the template `http://www.jarrodmillman.com/stat159-fall2015/project/report.tex`.

Additional resources:

- `https://www.coursera.org/course/fmri`