Stat 5014

Bob Settlage

August 2017

This course is an introduction to computing for statistics. Recently, open source platforms have proliferated and are becoming the *defacto* standard for Data Scientists. R and Python are arguably the two most important languages in a Data Scientists toolbox. In this class, we introduce both languages with a focus on R. We touch on notebook style analysis for enabling and performing Reproducibile Research. Throughout the course, we will use LaTeX via Markdown for type setting. In many collaborative environments, version control and collaborative development is a crucial technology and concept. Here, we will use Git as the backbone of the course for homework submission, collaborative learning via code sharing, and crowd sourced design. Finally, we end the class with methods for large (TB) datasets and discuss how these methods could be used to handle other large (long) computes.

Course learning objectives:

- Good programming practices
- Reproducible research concepts
- Data cleaning and munging
- R programming
- Git fundamentals
- Markdown
- Python basics