Autumn 2023, Phys 576B
Modern Data Analysis Techniques
Miguel Morales & Bryna Hazelton

Welcome to Modern Data Analysis Techniques
Team taught by Miguel Morales (Physics) and Bryna Hazelton (eScience), the goal of this class is to introduce current techniques and best practices in the statistically rigorous analysis of large data sets. The class is organized around four themes: practical statistics, advanced data visualization, collaborative analysis code, and advanced data analysis practices.

Grading
The course is designed to help you with your research. Consequently most of the homework assignments will involve applying concepts from class to your research. Further, this class is designed to scale depending on your interests and time. During the first week you will detail what your goals are, and your grade will be based on how well you achieve your goals. There will be no exams, with the homework and final project forming the basis of your grade.

Syllabus
Themes: Practical Statistics; Data Visualization; Collaborative Analysis; Advanced Data Analysis Practices

Week 1
Th: Welcome; course overview; what does sigma mean? (intro to non-Gaussian statistics)

Homework: Intro quiz

Week 2
T: Introduction to git & GitHub; collaborative analysis
Th: Statistical building blocks (convolution, analytic & non-analytic background distributions)
Homework: Homework #1 (git game)

Week 3
T: Data visualization pt. 1 (data density, perception, how to make plots for personal use)
Th: No class
Homework: Homework #2 (intro to stats)

Stats reference:
Statistics cheat sheet
Online reference chart Links to an external site.
Song paper
Wikipedia entries can be useful, look under 'related distributions'

Week 4
T: Data visualization pt. 2, workshopping plots; analysis plans, worry lists;
Th: Trials factors; parameter distributions
Homework: Homework #3

Week 5
T: Parameters cont.; Fisher matrix; triangle plots; variable backgrounds
Th: Statistically valid plots; jackknife tests

Week 6
T: Developing an analysis plan
Th: Confidence intervals

Week 7
T: Metadata, Provenance & Test Thickets
Th: Stats mini-review; the blob, analysis dragons

Week 8
T: Deconvolution/forward modeling; ML overview
Th: Machine Learning; plots as a language

Week 9
T: Blind & semi-blind analyses; data rampages
Th: Thanksgiving

Week 10
T: Presentations:
Th: Presentations:

Week 11
T: Presentations:
Th: Presentations: