# ASTR 598: Astro-statistics and Machine Learning

Andy Connolly & Željko Ivezić

University of Washington, Winter Quarter 2018

**Location and Time:** Tuesdays and Thursdays 11:00-12:20, **Room: B305**

**Office Hours:** Any time when our office doors are open;
After class, or Tue and Thu mornings by appt. are the best.

**Class web site:** https://github.com/dirac-institute/uw-astr598-w18

**Reference textbook:**
Ivezić, Connolly, VanderPlas & Gray: *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data* (Princeton University Press, 2014)
See http://press.princeton.edu/titles/10159.html

**Learning Goals:** This course will introduce graduate students to most common statistical and computer science methods used in astronomy and other physical sciences. It will combine theoretical background with examples of data analysis based on modern astronomical datasets. Practical data analysis will be done using python tools, such as astroML module (see www.astroML.org). While focused on astronomy, this course should be useful to all graduate students interested in data analysis in physical sciences and engineering. The lectures will be aimed at graduate students and the main discussion topics will be based on Chapters 3-5, and selected topics from Chapters 6-10, from the reference textbook.

By taking this course, students will develop working knowledge of topics such as robust statistics, hypothesis testing, maximum likelihood analysis, Bayesian statistics, model parameter estimation, the goodness of fit and model selection, density estimation and clustering, unsupervised and supervised classification, dimensionality reduction, regression and time series analysis. Most of these topics will be applied in class homeworks and projects to analysis of astronomical data.

**Prerequisites:** The students taking this class are required to have taken basic calculus and statistics, to have knowledge of scientific measurements at the level of a senior lab (in science or engineering), as well as to have basic python skills.

**Lecture format:** After a few introductory lectures, new material will be introduced on a schedule that will enable independent student work on a publication-quality class project.

**Class Schedule (tentative!):**

- WEEK 0 (first class on Jan 4, Željko): Introduction to class (syllabus, literature, astroML, class project, DataLab)
- WEEK 1 (starting Jan 8, Željko): Introduction to statistics (probability, distributions, robust statistics, Central Limit Theorem, hypothesis testing).
- WEEK 2 (starting Jan 15): Work on the Class Project (step 1).
- WEEK 3 (starting Jan 22): Maximum likelihood and applications in astronomy.
- WEEK 4 (starting Jan 29): Bayesian statistics and introduction to MCMC; Class Project step 2 report.
- WEEK 5 (starting Feb 5): Model parameter estimation and model selection.
- WEEK 6 (starting Feb 12): Regression and Time series analysis; Class Project step 3 report.
- WEEK 7 (starting Feb 19): Dimensionality reduction; Class Project step 4 report.
- WEEK 8 (starting Feb 26): Density estimation and clustering; Class Project step 5 report.
- WEEK 9 (starting Mar 5): Supervised Classification; writing assignments for Class Project paper
- FINAL EXAM: Mar 13 (Tue, 11:00-12:20, B305): cake and closed book final exam.

**Class Project**

A quarter-long survey data analysis project is described separately (see the class website). We will use GitHub and Jupyter notebooks for progress tracking.