Introduction to Astrostatistics and Data-Intensive Astronomy

ASTR 598A

mjuric@astro.washington.edu

ajc@astro.washington.edu

This Week

- Why statistics and ML in astronomy?
- Administrivia
- Getting set up (Zoom, Slack, JupyterHub, GitHub,…)

About Us: Mario Juric

Artist's impression: 👇

- Mario Juric (mar-ee-oh you-rich)
  - Astronomy Prof & eScience Institute Fellow
  - Office: C320 (Zoom/Slack/mjuric@astro.washington.edu)
- What I do:
  - Rubin Observatory Legacy Survey of Space and Time (LSST)
  - Astronomical algorithms and software research
  - Science derived from large surveys: Galactic structure, properties of the Solar System

Learns by trial and error: 👈

About Us: Andrew Connolly

- Andy Connolly
  - Astronomy Prof, Director of eScience
  - Office: C355 (Zoom/Slack/ajc@astro.washington.edu)
- What I do:
  - Lead LINCC Frameworks (software for astronomy)
  - Machine learning and statistical approaches for astronomy
  - Rubin Observatory Legacy Survey of Space and Time (LSST)
  - Cosmology, Large Scale Structure, Solar System, Galaxy Spectral Classification

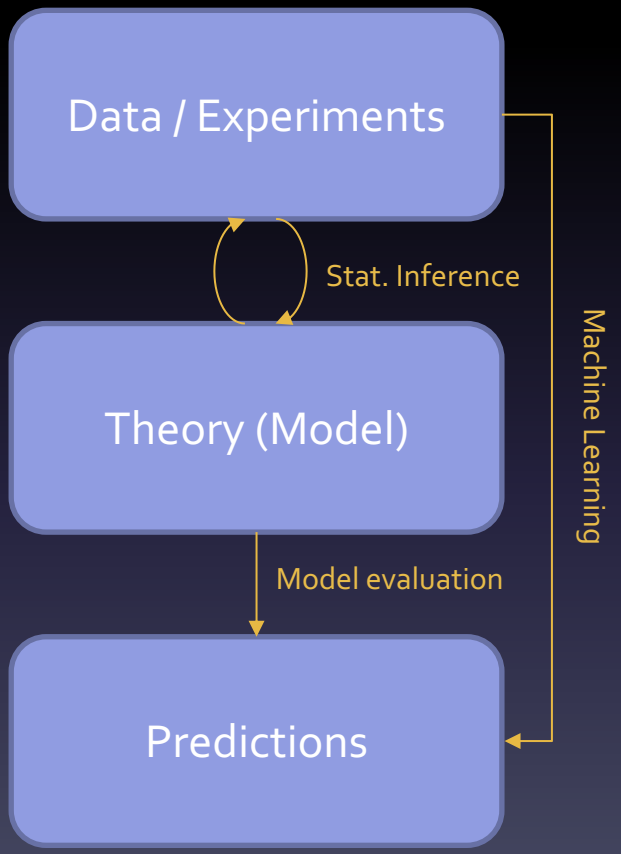
There are a lot of us… be afraid

Why Statistics and ML?

- Our goal as scientists is to *understand* the laws governing the world around us (theories) based on *observations and experiments* (data), and make *predictions* about yet unseen observations and experiments.

- Statistics (statistical inference) gives us the mathematics to correctly interpret observations and their impact on theories.

- Machine learning allows us to sometime (in part) skip the "understanding" part, jumping straight from data to predictions. Interpreting ML models (understanding the *"why?"*) is one of the most interesting areas of ML today.



A new topic every week

1. Getting started with technology
2. Introduction to probability and statistics I
3. Introduction to statistics II
4. Maximum likelihood and applications in astronomy
5. Bayesian inference and model selection
6. Introduction to MCMC and model parameter estimation
7. Dimensionality reduction
8. Project Proposals and Discussions
9. Time series analysis
10. Basics of Machine Learning
11. Basics of Deep Learning

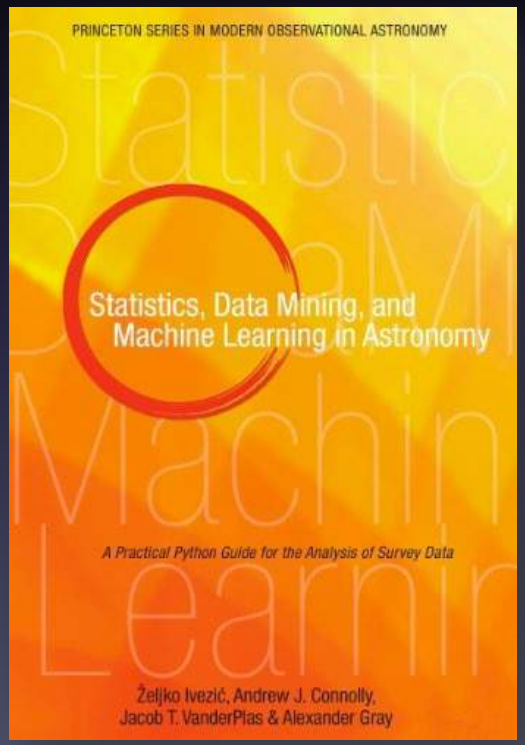
https://dirac.us/git598

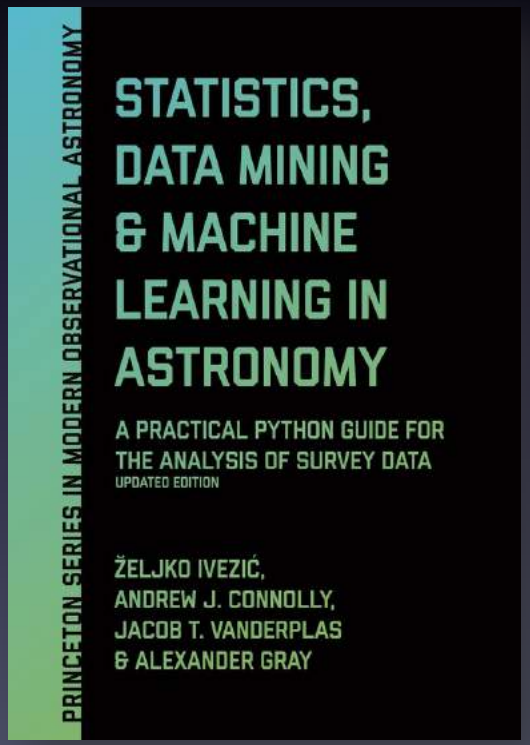
Learning Goals

- At the end of this course, you should know:
  - How to think probabilistically, and correctly interpret probability
  - How to correctly summarize measurements
  - How to estimate model parameters given observations, and when to reject poor models. Understand the theory behind why this works.
  - How to think of probabilities of parameter values, and how to derive those using Markov Chain Monte Carlo techniques.
  - How to measure and interpret time series
  - Apply all this to an research project

Textbooks & Reference Material

- We will be using the "Statistics, Data Mining, and Machine Learning in Astronomy" textbook by Ivezic, Connolly, VanderPlas and Gray





+ many (many) online writeups / notebooks / blog posts that I will point you to over the next few weeks.

Class Meetings

- When: TBD
- Were & How:
  - Lectures delivered via YouTube: https://dirac.us/videos598
    - Generally Jupyter notebooks; best to follow along as you watch.
    - Supplement by readings from the textbook.
  - In-person discussions

Flipped Classroom

- "Flipped classroom":
  1. Lecture videos and homework assignments will be posted by Friday morning (YouTube and GitHub)
  2. Monday afternoon, you will be asked to fill out:
     - A short quiz covering the material (Canvas)
     - Anonymous survey about anything that was unclear in the lectures (Canvas)
  3. MW: We'll spend our in-person time discussing and doing homeworks
     - M: Discussion (groups of 2-3), followed by a joint discussion (~40 minutes)
     - W: Work on the homework (groups of 2-3). Ideally, you can finish your homework in class!

Communication: Slack

- In this class, we won't be using a mailing list but an instant messaging (-like) tool called Slack (http://slack.com). Slack is heavily used today by many research & technology companies and projects.
- Signing up for Slack:
  a) https://join.slack.com/t/uw-astronomy/signup
  b) Join the #astr-598-astrostats channel

- What to use it for:
  - Asking questions, discussing the class, exchanging snippets of code, discussing homeworks.
  - Please prefer asking questions via Slack to sending me e-mails. Two reasons:
    - Everyone can benefit from the question and answer.
    - Your colleagues may be able to help!

Action: Let's make sure everyone's on Slack!

Course Materials

- I'll be adding most of what we need to the following organization on GitHub:

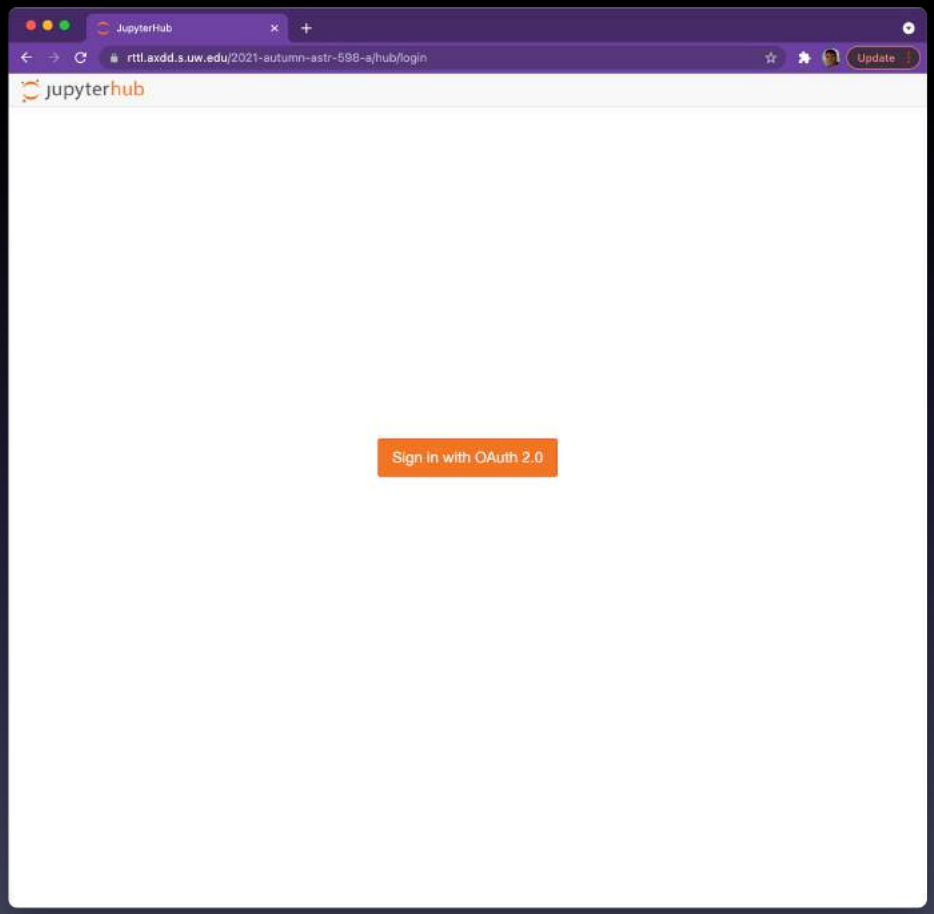
`https://github.com/uw-astrostats`

- You'll turn in homeworks via JupyterHub (more next week).

Action: Let's make sure everyone can access this repository.

JupyerHub

- JupyterHub: running Jupyter notebooks remotely.
- Go to: https://dirac.us/hub598
- This environment has all you need for the class. No need to install anything on your laptops!
- Caveats
  - To conserve resources, it will close your notebooks after 1hr of inactivity.



Action: Let's try this out!

Homeworks and Grades

Homeworks (50% of the grade):

- Jupyter notebooks. Designed to exercise what we've learned in any given week. Roughly ~one per week.
- All homeworks will be turned in via JupyterHub, two weeks after being assigned.
- Late homework policy:
  - -20% for being up to 1wk late
  - -50% for more than 1wk late.

- You will propose a project (a piece of software, or an analysis) to build or improve using the techniques and libraries we'll learn about in the course. Ideally, this is something that helps with your research. Keep your eyes open for ideas!
- **LINCC Frameworks infrastructure and science with Rubin (e.g. ZTF, SDSS, etc)**

Quizes (10% of the grade):

- Multiple-choice questions every due every Monday morning (12noon).

Final project (40% of the grade):

Details are in the Syllabus in the class git repository

Next steps

- Relax. This is a graduate-level course. We're here to teach & learn, not to fail you.
- Expect lectures for next week tomorrow. I'll post a note to Slack.
- Don't forget to do the quiz by noon Monday.