

# Introduction

# Course Logistics

# The (tentative) Schedule

Lecture Number	Topic(s)
1	Introduction Bayesian Inference
2	Linear Regression and Categories
3	Multiple Regression and Confounds
4	Curves, Prediction, and Overfitting
5	Markov chain Monte Carlo & Generalized Linear Models: Logistic and Binomial
6	Sensitivity Proxies Generalized Linear Models: Poisson
7	Ordered Categories
8	Bayesian Hierarchical Models
9	Bayesian Hierarchical Models (cont)
10	Network Analysis
11	Gaussian Processes
12	Missing Data
13	Clustering and Mixture Models
14	Variational Inference

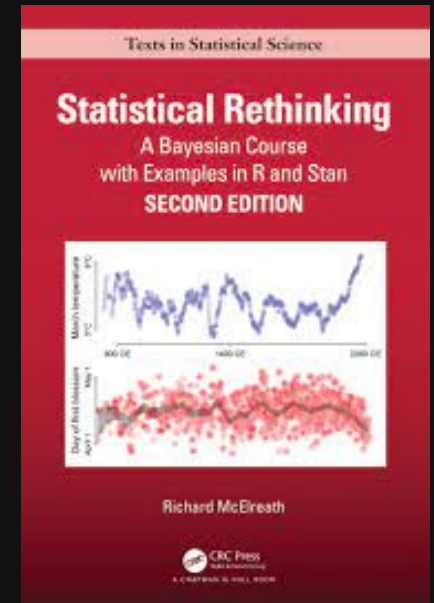
- Available under Content on Brightspace
- Likely to change
- Includes:
  - Topics
  - Assigned Reading
  - Homework Due Dates

# Evaluation

- Homework Assignments (60%)
  - 8 in total
  - Lowest score dropped
  - Late submission allowed (10% penalty)
    - additional 10% penalty per 24 hours
- Course Project (40%)

# Pre-Lecture Reading

- McElreath, R. (2020). Statistical Rethinking: A Bayesian Course with Examples in R and STAN. United States: CRC Press.
- Reading assigned for each lecture
- Corresponding Python code on JupyterHub



# Office Hours

(start Thursday, September 5th)

- Professor Reeves
  - Wednesday: 11 AM - 1 PM
- Course Assistant Office Hours
  - Available on Brightspace

# Our Computational Tools

# Python

- Prior Python programming experience required
- Version 3.9
- Docs: <https://docs.python.org/3.9/>



# NumPy and SciPy

- Core parts of Python's scientific computing ecosystem
- Numerical Python (NumPy) for general math/linear algebra
- Scientific Python (SciPy)
  - probability distributions
  - optimization routines
  - sparse matrix support
  - etc
- NumPy Docs: <https://numpy.org/doc/1.24/>
- SciPy Docs: <https://docs.scipy.org/doc/scipy/>

# PyMC

- Open-source, probabilistic programming language (PPL)
- Bayesian data analysis
- Contains some overlapping functionality with NumPy/SciPy
- Docs: <https://www.pymc.io/projects/docs/en/stable/api.html>

# Pandas

- Python data analysis/manipulation library
- Lots of functionality
- Docs: <https://pandas.pydata.org/docs/>

# Matplotlib

- Low-level visualization/plotting library
- Modeled on MATLAB's visualization framework
- Huge amount of functionality
- Docs: <https://matplotlib.org/stable/index.html>

# Seaborn

- Data visualization library
- Built on top of pandas and matplotlib
- Simplifies creating beautiful, clear visualizations
- Docs: <https://seaborn.pydata.org/api.html>

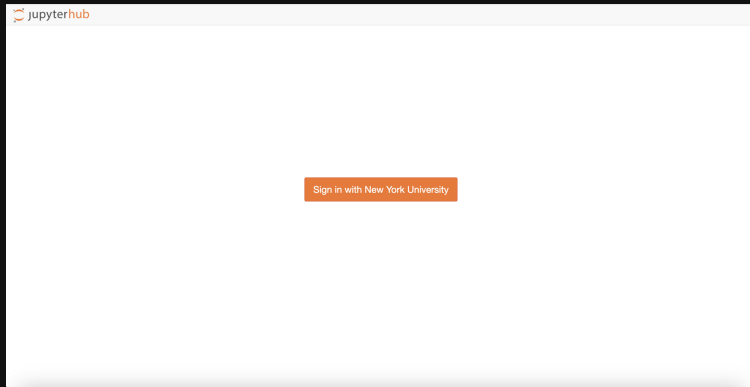
# ArviZ

- Exploratory analysis of Bayesian models
- Diagnostics and model comparison
- Docs: <https://python.arviz.org/en/latest/api/index.html>

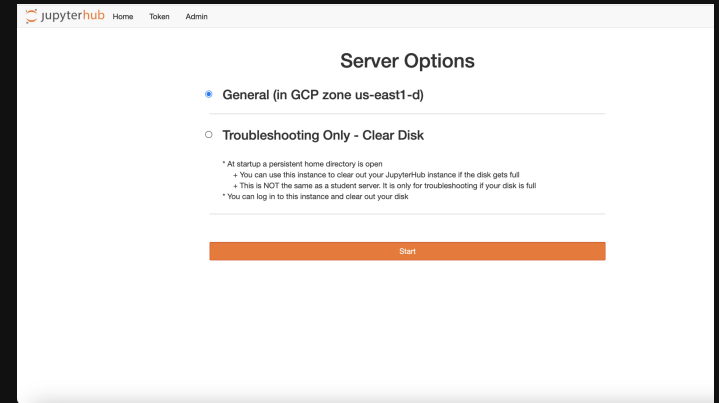
# JupyterHub

- Provided by NYU High Performance Computing
- No need to install any software
  - Jupyter notebooks allow Python execution in browser
- Textbook code (Python)
- Homework assignments
- Course site: <https://csgy-6053-fall.rcnyu.org/>

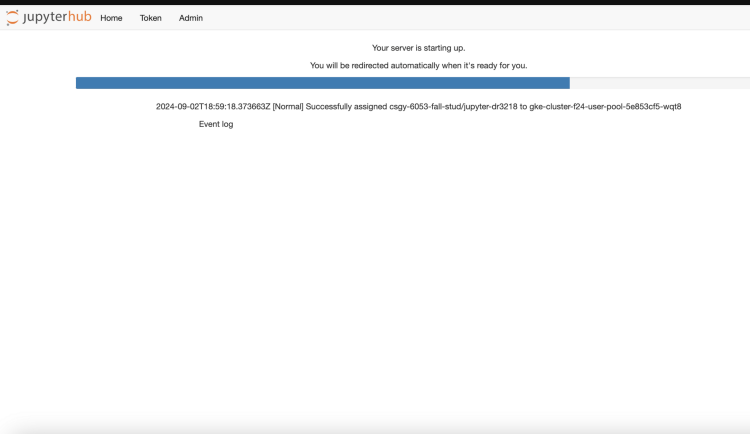
# JupyterHub



login →



start server ↙



access  
course →  
materials

