



# Workshop: **Introduction to Python**



## **Statistics**

Christian C. Luhmann  
Stony Brook University

# Relevant Packages

- scipy
  - Distributions
  - Simple stats (e.g.,  $t$ ,  $\chi^2$ ,  $z$ ,  $r$ , 1-way ANOVA)
- statsmodels
- pymc3
- bambi

# Relevant Packages

- scipy
  - Distributions
  - Simple stats (e.g.,  $t$ ,  $\chi^2$ ,  $z$ ,  $r$ , 1-way ANOVA)
- statsmodels
  - More complex stats (e.g., GLM, mixed models, survival analysis)
- pymc3
- bambi

# Relevant Packages

- scipy
  - Distributions
  - Simple stats (e.g.,  $t$ ,  $\chi^2$ ,  $z$ ,  $r$ , 1-way ANOVA)
- statsmodels
  - More complex stats (e.g., GLM, mixed models, survival analysis)
- pymc3
  - Full-featured Bayesian modeling (think Stan)
- bambi

# Relevant Packages

- `scipy`
  - Distributions
  - Simple stats (e.g.,  $t$ ,  $\chi^2$ ,  $z$ ,  $r$ , 1-way ANOVA)
- `statsmodels`
  - More complex stats (e.g., GLM, mixed models, survival analysis)
- `pymc3`
  - Full-featured Bayesian modeling (think Stan)
- `bambi`
  - Streamlined, Bayesian GLMs built on top of `pymc3` (think `brms`)

# Statistics

Let's go do some stats!

# scikit-learn

- Machine learning
  - Supervised
    - Classification (e.g., GLM, LDA, SVM, random forests)
    - Regression (e.g., ridge, lasso)
  - Unsupervised
    - Clustering (k-means)
    - Dimension reduction (e.g., PCA)
- All the extras needed to fit, evaluate, and use these tools

# Take-homes

- You have now seen some stats done in Python
- Seen some of the functionality that relevant packages provide
  - pandas
  - jupyter (notebook)
  - matplotlib
- What data exploration looks like and the flexibility these tools provide



# Outline

1. Overview
2. Ways of using Python
3. Python basics
4. Data set overview
5. Data wrangling
6. Statistics
7. Plotting
8. Experiment creation