CCLA: Data Science, 2023–2024

Instructors:
HaiYing Wang, Associate Professor, Department of Statistics, University of Connecticut
Jun Yan, Professor, Department of Statistics, University of Connecticut

Lectures:
Location: TBA
Sunday: 11:00–11:50 ET

Course Description: Data science is a fast developing science of extracting meaningful information from massive data for better decision making. It is interdisciplinary by nature, involving statistics, computing, and domain knowledge. Important principles of data science will be elaborated through interactive simulations, games and examples.

Prerequisites: Some exposure to programming and data science would be a plus, but is not required.

Topics:

1. Introduction to Julia, random number generation, data summaries
2. Exploratory data analysis, visualization
3. Paradoxes explained
4. At the infinity horizon (as sample size increases)
5. Hide-and-seek: Estimating the unknowns
6. Sampling design, data collection
7. Correlation and regression
8. Hypothesis testing
9. A real-world data science project

Computing: Please have Julia preinstalled: `https://julialang.org`

References:

- Joshi, A. (2016). Julia for Data Science. Packt Publishing Ltd.
- Storopoli, J., Huijzer, R., & Alonso, L. (2022). Julia Data Science. Editeur indépendant. `https://github.com/JuliaDataScience/JuliaDataScience`
- Nazarathy, Y., & Klok, H. (2021). Statistics with Julia: Fundamentals for data science, machine learning and artificial intelligence. Springer Nature. `https://github.com/h-Klok/StatsWithJuliaBook`