**BAIS:3250 - Data Wrangling:**

“Data wrangling” encompasses collecting, cleaning, transforming, integrating, and describing data. It is estimated that data scientists spend up to [80% of their time on data wrangling tasks](https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists). In addition to being the most time-consuming step of most analytics workflows, wrangling is critical, as it impacts the outcome of all subsequent experiments/analyses. This course introduces essential R programming skills for data wrangling.

**Students will learn:**

* R and R Studio
* File Handling
* Data Cleaning and Transformation
* Web Scraping
* APIs
* Data Integration
* Descriptive Analytics and Visualization

**Students will be able to:**

* Build foundational programming skills, including variables, data types, data structures, functions, and subsetting
* Read and write plain text, delimited, and nested file formats (XML, JSON), including file encoding
* Learn methods to identify and correct common types of errors, handle missing values, and create derived features
* Access data via APIs, Web scraping, and external database connections
* Perform vertical and horizontal data integration
* Use statistical methods (summary tables, hypothesis testing, regression) and visualizations to describe data
* Apply techniques for time-series analysis and text analytics, including regular expressions and term frequency representation