

# DATA210N : Data Wrangling with R and Python

Data wrangling is the process of working with large, messy and diverse datasets to prepare them for data analysis. Students will learn techniques of uploading and merging source data, diagnosing data issues that require intervention before analysis (missing data, misformatted data, and other unclean data issues such as outliers and corrupt data), fixing data issues (cleaning) and preparing and exporting usable analytical data files. Students will also learn methods of joining fields (such as concatenating text fields), sub setting and filtering data, and large-scale dimensionality reduction techniques (principal components and factor analysis). The posit integrated development environment using elements of R, Python and SQL languages in synchrony will be used.

**Class Hours** 2

**Lab Hours** 2 **Credits** 3

**Prerequisite Courses**

DATA101N