Applied Data Analytics
Fall 2022

Exploratory Data Analysis

Introductions	Data science and friends, RStudio, R Markdown, Github	*wk 1*
Raw Data	Relational and non-relational databases, Data types, Base R	*wk 2*
Data Wrangling	Grammar of data manipulation, Summary statistics, `dplyr`	*wk 3*
Data Visualization	*The Visual Display of Quantitative Information*, Base R graphics, `ggplot`	*wk 4*
Data in the World	Missing data and data imputation, Data privacy	*wk 5*
Data Scraping	Scraping legality, etiquette, and technique	*wk 6*
Text Data	String processing, Bags of words	*wk 7*
Data Ethics	Transparency, Interpretability, Reproducibility	*bonus*

Data Modelling

Probability	Random variables, Distributions, Monte Carlo sampling	*wk 8*
Hypothesis Testing	Common hypothesis tests, p-values, Confidence intervals	*wk 9*
Multiple Testing	p-hacking, Publication bias, Crisis of reproducibility	*wk 10*
Linear Regression	Correlation, Linear models, Prediction	*wk 11*
Multiple Regression	Diagnostics, Model fit, Outliers, Causality	*wk 12*
High-Dimensional Data	Collinearity, Feature selection	*wk 13*
Logistic Regression	Modelling binary and count data, Data transformations	*wk 14*

Presentations

Project Presentations	*wk 15*