Week by Week Schedule

Week 1 Setting up a data science working environment

- Key tools:
  - the anaconda environment
  - vscode
  - Rstudio
- Projects
  - projects, files, directories
  - the terminal
  - self documenting code
    - jupyter notebooks, markdown, and R markdown
  - essentials of REPL python in jupyter
    - numeric variables, arithmetic, scientific functions;
    - lists and arrays; indexing and slicing; vectorization
  - basics of R
    - datatypes, arrays, arithmetic, slicing and indexing, vectorization
  - a minimal introduction to plotting in R and Python

Week 2 Probability and Statistics: a first look at the normal distribution

- Working with normally distributed populations
  - Events and Outcomes
  - The Normal Distribution
    - probability mass functions and area under the curve
  - Mean and Variance
  - Hypothesis testing
    - sampling from a univariate normal distribution
    - Null and alternative hypotheses
    - p-values, statistical significance and confidence intervals (for univariate normally distributed populations)
  - illustrated with examples in R and Python

Week 3 More on programming with data in R and Python

- Key tools
  - R dataframes and the tidyverse
  - Numpy and Pandas
  - functions in R and Python
- working with files and I/O in R and Python
- pandas Series and dataframe basics (reading files, indices, selecting, summarizing data)
- R factors and dataframes (reading files, indices, selecting, summarizing)

Week 4 Linear Algebra

- Geometry of n-dimensional space, vectors, addition and scalar multiplication of vectors, the dot product, orthogonality
- Matrices, matrix multiplication, column space of a matrix
- Ordinary Least Squares as an illustration(?) of geometry and linear algebra
- computational examples in both R and Python (numpy)

Week 5 Partial derivatives and the gradient

- discussion of functions of several variables:
  - graphs of functions
  - contour graphs and level surfaces
- review of the derivative in one dimension; rates of change
- partial derivatives
- directional derivatives and the gradient
- contour plotting

Week 6 Slicing and dicing data in R and Python

- One day on pandas grouping, summarizing, selecting data
- One day on R grouping, summarizing, selecting data (tidyverse)

Week 7-8 Statistical models and hypothesis testing

- What is a statistical model?
- Likelihood and model parameters
- Maximum likelihood
- Another look at the normal distribution; the multivariate normal
- covariance, correlation
- significance and confidence intervals
- null and alternative hypotheses
- p-values

Week 9 Advanced topics in programming

- Data structures; object oriented concepts
- Essential notions from data structures:
  - stacks, lists, hashing
- Python classes
  - data attributes and methods

Week 10 A deeper dive into visualization

- more on ggplot and its capabilities, dashboards
- python plotting packages and dashboards

Week 11 Version Control

- Git as a tool (command line and through R studio)
- commits, branches
- remotes and github
- using github to host a web page for a project
- collaboration using Git – pull requests; contributing to open source projects

Week 12 Databases

- What is a relational database? tables, keys, indices, joins
- Basic SQL for getting data from a database

Additional topics as time permits

Discrete Probability and Bayes Theorem

- Discrete probability;
  - events and outcomes;
  - mean and variance
  - independent events;
  - conditional probability and Bayes theorem in the discrete case
  - bernoulli and binomial distributions
  - false positives, false negatives, versions of the base rate fallacy and simpson's paradox
  - illustrated with R and Python examples

More advanced topics in Linear Algebra

- Eigenvalues and Eigenvectors
- Orthogonality and orthogonal projection
- The spectral theorem for real symmetric matrices