

Week 1 Setting up a data science working environment (for Python)

- Key tools:
 - the anaconda environment
 - vscode
- Working with vscode and anaconda
 - projects, files, directories
 - the terminal
 - packages and the package manager
 - markdown and jupyter notebooks; self-documenting code
 - essentials of REPL python in jupyter
 - * numeric variables, arithmetic, scientific functions;
 - * lists and arrays; indexing and slicing; vectorization
 - * booleans
 - * list comprehensions
 - pandas basics, reading files, series, indices, selecting

Week 2 Setting up a data science working environment (for R)

- Key tools:
 - Rstudio
- Working with Rstudio
 - comparison of files, projects, directories with the python situation
 - packages, libraries management within Rstudio
 - R Markdown and self-documenting code
 - basics of the R language
 - * datatypes, arrays, factors, functions, arithmetic
 - * slicing and indexing, vectorization, booleans
 - R dataframes basics, reading files, indices, selecting

Week 3 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 4 Plotting in R and Python

- One day on plotting in R using ggplot and base R
- One day on plotting in python using seaborn (?)

Week 5 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 6-7 Multivariate calculus

- 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
- relationship of the gradient to the level curves
- Use of calculus to solve the OLS problem done by linear algebra in week 5 (exact solution)
- Use of gradient descent to solve OLS problem
- includes programming examples for gradient descent in the least squares case

Week 8 Probability (Discrete case)

- 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
 - discussion of the Naive Bayes spam filter (?)
 - illustrated with R and Python examples

Week 9 Probability Distributions

- Continuous random variables
- probability mass distributions and cumulative distributions
- area under the curve and probability
- mean and variance in the continuous case
- the normal distribution and its mean and variance
- illustrated with examples in R and Python

Week 10 Statistical models

- What is a statistical model?
- Likelihood and model parameters
- Maximum likelihood estimation and gradient descent

Week 11-12 Statistics

- Basics of experimental design
- Estimation of parameters
- confidence intervals
- significance
- null and alternative hypothesis
- p-values

Week 13 Version Control

- Git as a tool
- 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 14 Databases

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