Syllabus

Course coordinator. Assoc. Prof. Daphne Skipper   ✉ skipper@usna.edu

Course objectives. By the end of this course, students will be able to:

(i) Wrangle (i.e. clean and manipulate) large, messy data sets into forms suitable for modeling and analysis (in particular, optimization, simulation, and statistical models).

(ii) Create sophisticated visualizations of large data sets that provide useful insights for decision-making as well as further modeling and analysis.

Textbook. There is no required textbook for this course. You will be provided with lesson notes and links to supplementary readings throughout the semester.

Schedule. Here is a tentative schedule.

Week	Topics
Warm up	
1	A survival course in Python and Jupyter A very brief introduction to Pandas Method chaining
Data visualization with Altair	
2	Altair basics: data types, encoding channels, graphical marks Basic data transformations: binning and counting, aggregation
3	Advanced data transformations: calculate, filter, aggregate, window
4	Encoding channel configuration: scales, axes, colors, sort order Top-level chart configuration
5	Multi-view composition: layers, concatenation, facets, repeat
6	Interactive visualization: tooltips, pan and zoom, dynamic queries Data sources in Altair
7	Cartographic visualization: point maps, symbol maps, chloropleth maps, lookup transforms **Exam 1**
Data wrangling with Pandas	
8	The Series and DataFrame objects Basic arithmetic operations on Series and DataFrame objects, broadcasting Filtering observations
9	Selecting and dropping data Creating new variables
10	Grouped operations: split-apply-combine

Week	Topics
11	Working with missing values
12	Tidy data: long vs. wide data, pivoting, separating, uniting **Exam 2**
13	Combining data: concatenate, merge Working with strings, dates, and times

Additional topics

14	Regular expressions
15	Web scraping
16	Review