

Syllabus

Last updated: August 18, 2021

Course coordinator. Assoc. Prof. Nelson Uhan ✉ uhan@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
3	Basic data transformations: binning and counting, aggregation 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	Review Exam 1
7	Interactive visualization: tooltips, pan and zoom, dynamic queries
8	Data sources in Altair Cartographic visualization: point maps, symbol maps, choropleth maps, lookup transforms
Data wrangling with Pandas	
9	The Series and DataFrame objects Basic arithmetic operations on Series and DataFrame objects, broadcasting Filtering observations
10	Selecting and dropping data Creating new variables

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
- 3 Basic data transformations: binning and counting, aggregation
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 Review
Exam 1
- 7 Interactive visualization: tooltips, pan and zoom, dynamic queries
- 8 Data sources in Altair
Cartographic visualization: point maps, symbol maps, choropleth maps, lookup transforms

Data wrangling with Pandas

- 9 The Series and DataFrame objects
Basic arithmetic operations on Series and DataFrame objects, broadcasting
Filtering observations
- 10 Selecting and dropping data
Creating new variables

Week Topics

- 11 Review
Exam 2
- 12 Grouped operations: split-apply-combine
- 13 Working with missing values
Tidy data: long vs. wide data, pivoting, separating, uniting
- 14 Combining data: concatenate, merge
- 15 Working with strings
Working with dates and times

Additional topics

- Web scraping
- Passing data between R and Python
- 16 Review or additional topics