

Date	Topic	Reading (before class)
Introduction		
9/4	Course overview, slack, and jupyter	
9/9	Python Programming Assessment (ungraded) & Python Review	
Python Toolkits for Data Manipulation		
9/11	Regular Expressions	Python docs
9/16	Pandas I: Numpy, Series and DataFrames	Chapters 3, 4, 5
9/18	Pandas II: Loading data & Data Cleaning	Chapters 6, 7
9/23	Pandas III: Joining & Combining	Chapter 8
9/25	Pandas IV: Aggregation & Grouping	Chapter 10
9/30	Pandas V: In-class problem set	
	Other structured data formats: HTML, XML, JSON	
10/1	Assignment 1 due: Pandas (10%)	
Scalable Structured Data		
	Databases I: Introduction	
	Databases II: DDL	

	Databases III: DML 1	
	Databases IV: DML2	
	Databases V: ETL and ORM	
	Assignment 2 due: Databases (10%)	
10/28	Midterm (Cumulative up to and including Databases IV)	
10/11	Project Pitches (2 min lightning + 800 words)	
Big Data		
	Big Data I:	
	Big Data II:	
	Big Data III:	
	Big Data IV:	
	Assignment 3 due: Big Data	
Other Forms of Data		
	Time Series, Natural Language	
	Networks, Images	
Extended Topics		
	Project Presentations I	
	Project Presentations II	
	Project Due	
	Advanced python/pandas topics: performance, cython	
	Advanced databases topics: triggers, ufuncs, data modelling	
	Advanced scalability topics: service oriented computing with aws	