Schedule (2016-08-30)

(Subject to change)
All required readings should be completed by the following week.
All exercises are due on the Friday morning following class, at 12pm.
All reviews are due prior to the start of class, at 7pm.

Date	Topic / Readings	Deadlines
2016-08-30	Introductions; Jupyter notebook and command line shell basics; Git and GitHub basics; computing setup. Readings for next week: Required: Software Carpentry Lesson: The Unix Shell, http://swcarpentry.github.io/shell-novice/ Required: Roger Peng on Reproducible Research (three videos): http://tinyurl.com/jhu-reproducible-research Optional: Software Carpentry Lesson: Version Control with Git, http://swcarpentry.github.io/git-novice/	Exercise #1, Friday, 9/2, 12pm
2016-09-06	The command line shell: input, output, and pipelines; csvkit; data types. Readings Required: Wickham, "Tidy Data." http://vita.had.co.nz/papers/tidy-data.pdf Optional: Data Science at the Command Line, chapters 1-5	Exercise #2, Friday, 9/9, 12pm
2016-09-13	Command line filters in the shell and Python; parallel processing in the shell. Readings Required: Software Carpentry Lesson: Using Databases and SQL, Topics 1-5, http://software-carpentry.org/lessons.html Optional: Data Science at the Command Line, chapters 6-8	Project #1, Friday, 9/23, 12pm
2016-09-20	RDBMS: schema, keys, basic SQL operations, aggregate functions, subqueries. Readings Required: Software Carpentry Lesson: Using Databases and SQL, Topics 6-10, http://software-carpentry.org/lessons.html Optional: Learning SQL, chapters 1-4	Exercise #3, Friday 9/23, 12pm Review #1, Tuesday, 9/27, 7pm
2016-09-27	RDBMS: joins, integrity, transactions, functions, triggers, schema design and E-R models, normal forms.	Exercise #4, Friday 9/30,

	Readings Optional: Learning SQL, chapters 5, 6, 7, 9, 10 Optional: A Gentle Introduction to Algorithm Complexity Analysis (online at http://discrete.gr/complexity/) Optional: Visualizing Algorithms (online at http://bost.ocks.org/mike/algorithms/)	12pm
2016-10-04	RDBMS: advanced SQL, indexes, query processing, analysis, and optimization. **Note:** no office hours on Tuesday, October 4. Readings Required: Star Schema, chapters 1-5 Optional: Learning SQL, chapters 12, 13, 14	Project #2, Friday 10/15, 12pm
2016-10-11	**No class**	
2016-10-18	Warehouses: facts and dimensions, architectures, schemas Readings Required: Star Schema, chapters 4-7	Exercise #5, Friday, 10/21, 12pm Review #2, Tuesday, 11/1, 7pm
2016-10-25	**No class (fall break)**	
2016-11-01	Warehouses: dimension design Readings Required: Star Schema, chapter 11 Required: AWS Redshift. https://aws.amazon.com/redshift/	Exercise #6, Friday, 11/4, 12pm
2016-11-08	Warehouses: fact table design Readings Required: Dean and Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters." http://research.google.com/archive/mapreduce.html Required: Drake, "Command-line tools can be 235x faster than your Hadoop cluster." http://aadrake.com/command-line-tools-can-be-235x-faster-than-your-hadoop-cluster.html	Project #3, Friday, 11/18, 12pm

	Optional: Chang et al. "Bigtable: A Distributed Storage System for Structured Data." http://research.google.com/archive/bigtable.html Optional: DeCandia et al. "Dynamo: Amazon’s Highly Available Key-value Store", http://www.read.seas.harvard.edu/~kohler/class/cs239-w08/decandia07dynamo.pdf	
2016-11-15	Contemporary data management tools: Hadoop, map/reduce, Dynamo, Trifacta Readings Required: Apache Spark. https://spark.apache.org/ Required: Lambda Architecture. http://lambda-architecture.net/	Exercise #7, Friday, 11/18, 12pm Review #3, Tuesday, 11/22, 7pm
2016-11-22	Contemporary data management tools: Spark introduction Readings Required: CAP theorem. https://en.wikipedia.org/wiki/CAP_theorem Required: Kudu. http://getkudu.io/ Required: AWS Kinesis. https://aws.amazon.com/kinesis/	Exercise #8, Tuesday 11/29, 7pm
2016-11-29	Contemporary data management tools: Spark SQL, DataFrames, MLib, Streaming	Final Project, Friday 12/9, 12pm
2016-12-06	Final Project presentations, course wrap-up	