Schedule (subject to change)

All readings should be completed by the following week.
All assignments are due on the date listed, prior to the start of class, at 7pm.

Date	Topic / Guest / Readings	Assignments due
2015-09-01	Introductions; Jupyter and command line basics; VM setup. **Guest**: Shmuel Ben-Gad, Gelman Library Readings Required: Software Carpentry Lesson: The Unix Shell, http://software-carpentry.org/lessons.html Required: JHU Reproducible Research on Coursera, week one videos, https://www.coursera.org/course/repdata/ (about one hour) Recommended: Data Science at the Command Line, chapters 1-5	None
2015-09-08	The command line; input, output, and pipelines; csvkit; data types. Book review project. Readings Required: Wickham, "Tidy Data." http://vita.had.co.nz/papers/tidy-data.pdf Required: Software Carpentry Lesson: Using Databases and SQL, Topics 1-5, http://software-carpentry.org/lessons.html Recommended: Data Science at the Command Line, chapters 6-8	#1
2015-09-15	Command line filters; parallel processing. Readings Required: Software Carpentry Lesson: Using Databases and SQL, Topics 6-10, http://software-carpentry.org/lessons.html Required: Database System Concepts, chapters 1-3 (slides at http://codex.cs.yale.edu/avi/db-book/; text recommended) Optional: Learning SQL, chapters 1-4	#2
2015-09-22	**No class**	None
	RDBMS: schema, keys, basic SQL operations, aggregate	#3, and book

2015-09-29	functions, subqueries. Readings Required: Database System Concepts, chapters 4, 5, 7, 8 (slides at http://codex.cs.yale.edu/avi/db-book/; text recommended) Optional: Learning SQL, chapters 5, 6, 7, 9, 10 Optional: Write Great Code, Volume I: Understanding the Machine (online through GW Libraries at http://findit.library.gwu.edu/item/5966168), chapters 2-5 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/)	reviews start
2015-10-06	RDBMS: joins, integrity, transactions, functions, triggers, schema design and E-R models, normal forms. Group project. Readings Required: Database System Concepts, chapters 11-13 (slides at http://codex.cs.yale.edu/avi/db-book/; text recommended) Optional: Learning SQL, chapters 12, 13, 14	
2015-10-13	**No class**	#4
2015-10-20	RDBMS: indexes, query processing and optimization Readings Required: Star Schema, chapters 1-3	
2015-10-27	Warehouses: facts and dimensions, architectures, schemas **Guest**: Luis Novoa Readings Required: Star Schema, chapters 4-5	#5 (project 1)
2015-11-03	Warehouses: dimension design Readings Required: Star Schema, chapters 6-7	
2015-11-10	Warehouses: fact table design **Guest**: Jackie Kazil	

	Readings Required: Star Schema, chapter 11 Optional: Star Schema, chapters 16-18	
2015-11-17	**Midterm exam (online, due Sunday, 11/22)** Warehouses: design and implementation 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 Optional: Redis project. http://redis.io/ and Try Redis http://try.redis.io/ 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	
2015-11-24	Final notes on data warehouses, then noSQL and beyond: Map/Reduce, Hadoop, Redis, Spark **Tutorial:** Spark (Mokeli and Nisha) Readings Required: CAP theorem. https://en.wikipedia.org/wiki/CAP_theorem Required: Apache Spark. https://spark.apache.org/ Required: Kudu. http://getkudu.io/ Required: AWS Redshift. https://aws.amazon.com/redshift/ Required: AWS Kinesis. https://aws.amazon.com/kinesis/	book reviews end
2015-12-01	Spark, PySpark Group Project presentations (1)	#7 (project #2)
2015-12-15	Group Project presentations (2)	