

Summer 2017 - W205 – Storing and Retrieving Data
Week 1 Live Class Session Agenda
Kevin R. Crook

- Introductions
 - Instructor
 - Students – Round Robin
 - Name
 - Where you are from originally? Now?
 - Where you are working? In Data Science?
 - Academic and Professional Background
 - How far along are you in the program?

- KevinCrook.com

- Grading
 - Labs – easy compared to the exercises – students are usually able to complete them and get high marks
 - Exercises 1 & 2 and Project
 - grading rubrics are more rigorous than the labs
 - 90 % - if you meet all of the requirements
 - > 90 % - requires excellence above and beyond

- Late Work
 - Please see late penalties on schedule
 - Illness
 - 1 waiving of late penalties, except the project (will have to take an incomplete for the course as the project cannot be extended)
 - > 1 need to discuss an ADA accommodation
 - Business trips, busy at work, busy with other classes, vacations, etc.
 - Not valid excuses

- Connectivity Issues
 - Frequent problems? – consider getting a US cloud based desktop and connect to it from your computer using remote desktop

- Slack.com
 - Students will sign up and join the channels

- GitHub.com
 - Students need to sign up for an academic account that allows private repositories before the next class session. Next session we will go over creating a repository (“repo”) for this course and directories for the exercises and project

- Amazon Web Services (AWS)
 - Students need to sign up for AWS.
 - It will be needed for Lab 1 which is due next week.
 - For this course you will need to use the Berkeley AMI’s in AWS to complete the assignments. Everything is required to run against these and not your own installation. If you want to install software locally, it’s a great learning experience, but you must install and configure it on your own, we cannot support it. I cannot spend class time nor office hour time debugging your configuration for you.

- Tableau
 - Students will need an academic version of Tableau in a few weeks. It would be best to get it now.
 - Don’t get the 14 day free trial – get the academic version.

- Windows users
 - PuTTY – you will need to download this for use to connect to AWS at the Linux command line

- Mac users
 - Mac users had a lot of issues last semester – many more issues than Windows or Linux users. Even students who had purchased support from Apple were told that the employees at the Apple store did not know how to support Mac command line. I opened a ticket with the faculty Apple helpdesk and was told Mac command line in Apple is not guaranteed to work, they don’t support it, and it’s at your own risk.

- Berkeley Library
 - O’Reilly books and other books are available at the Berkeley library free in electronic format.

(next page)

- Things you will need to come up to speed quickly for success in this class. All are required for Exercise 1, so please be sure you come up to speed in time to complete Exercise 1.
 - Python Programming (should have had the Python class or equivalent)
 - Amazon Web Services
 - Linux Command Line
 - Bash Shell Programming
 - Data Modeling using Entity-Relationship Diagrams (ERDs) in Third Normal Form (3NF)
 - SQL

- Asynchronous
 - Readings
 - some of these are out of date in ISVC, please go by the syllabus instead of ISVC
 - some papers are highly theoretical
 - some are dated at this point, but included because they are landmark papers frequently referred to
 - Videos – please watch these before class
 - Quizzes – not so important – don’t count - a lot of them don’t work

- Live Class Sessions – what will we do?
 - Asynchronous
 - Asynchronous is way too much material to rehash everything – most units cover at a high level what would be a third to a half a semester’s course
 - Instructor will usually spend a few minutes going over at a high level what the instructor feels is important in the asynchronous papers or videos
 - If the asynchronous is extremely theoretical, the instructor may present a practical application to help explain how it applies to the real world
 - Won’t spend a lot of time on the asynchronous material
 - Labs (10)
 - Instructor will spend a few minutes highlight the next lab
 - Videos are provided on KevinCrook.com for the labs
 - Exercises (2)
 - Instructor will spend a few minutes highlighting exercises (there are 2)
 - Videos are provided on KevinCrook.com for the exercises
 - Project
 - Once the project has been assigned, last semester it seemed to be very helpful to go around round robin each week and get a quick update / issues / concerns related to projects.
 - There are scheduled presentations at several points

- Break Out Sessions
 - Students work on a problem related to a practical application of the asynchronous, or to work on something related to a lab or exercise.
- Students working in Data Science or Analytics
 - After a few weeks, we will start having any students who are working in Data Science or Analytics to talk each week for a few minutes about their job and projects they work on

- Lab 1

- Break Out Exercise
 - An non-profit area where Data Science can make the world a better place
 - Give an example of an old school company, can be Fortune 500, that has been around for at least 30 years, that has become a Data-Driven Organization