Class Sessions

Mondays and Wednesdays, 12pm – 1:30pm
425 E 61st Street, 3rd Floor Conference Room
January 7, 2019 – April 3, 2019

Instructor

Debby D'Angelo, MS
ded2011@med.cornell.edu
402 East 67th Street, LA-0004 (Floor C2)
Office Hours: by appointment only

Teaching Assistant

Kaylee Ho, MS
kah4001@med.cornell.edu
Office Hours: Wednesdays 11:00am-12:00pm, 425 E 61st St, Room 310

Course Description

In this course, you will learn to design and build relational databases, to write queries using the SQL programming language, and to pull data from the Web using APIs. Application of skills learned in this course will be geared toward research and data science settings in the healthcare field; however, these skills are transferrable to many industries and application areas.

You will begin the course examining the pitfalls of using Excel spreadsheets as a data storage tool and then learn how to build properly-designed relational databases to eliminate the issues related to spreadsheets and maintain data integrity when storing and modifying data. You will then learn two aspects of the SQL programming language: 1) the data manipulation language (DML), which allows you to retrieve data from and populate data into database tables (e.g., SELECT, INSERT INTO, DELETE, UPDATE, etc.), and 2) the data definition language (DDL), which allows you to create and modify tables in a database (e.g., CREATE, ALTER, DROP, etc.). Finally, you will learn how to pull data from the web using APIs.

This course is formatted as a flipped classroom, which means that lectures are watched on video outside of class, and "homework" is done during the class sessions (see "Course Structure" section for more details). Emphasis is placed on skill-building through practical programming exercises in Microsoft Excel, Microsoft Access, MySQL, R, and SAS.

Course Learning Objectives

Students who successfully complete this course will be able to:

- Describe the limitations of using Excel spreadsheets for data storage
- Design and build fully-functioning, normalized, relational databases using Microsoft Access and MySQL
- Formulate SQL queries to 1) create and modify database tables, and 2) populate and retrieve data from database tables using Microsoft Access, MySQL, R, and SAS
- Extract data from websites using APIs

Prerequisites

Students should have taken an introductory biostatistics course and have some familiarity working with data. No prior database development experience or SQL programming experience is required.

Course Structure

This course is formatted as a "flipped classroom," which means that lectures are watched on video prior to class, and assignments are done during class meetings under the guidance of the teaching team. This format allows you to learn material at your own pace by re-watching lectures as needed, and makes your time spent on exercises more efficient by allowing you to get clarification from instructors in real-time as you work.

At the beginning of each week (i.e., Mondays, in weeks without holidays), each class meeting will start with a brief review of the current week's lecture material. We will then review the previous week's assignments, and spend the remainder of the week's class sessions working on programming assignments. If you have a question as you work, please raise your hand and either the TA or the instructor will come over to assist you.

All course materials can be accessed on the course Canvas site:
https://medcornell.instructure.com/courses/1624515

Textbooks

There are no required textbooks for this course; however, the books below are excellent, comprehensive resources on building databases in Access and MySQL. The MySQL book also provides a clear and thorough overview, with examples, on SQL programming syntax.

- Access 2013: The Missing Manual by Matthew MacDonald. ISBN: 9781449357412
- Murach's MySQL, 2nd Edition by Joel Murach. ISBN-13: 978-1890774820

Software

We will use **Microsoft Excel**, **Microsoft Access**, **MySQL**, **R,** and **SAS** software in this class. Since Microsoft Access is only available for Windows, **you must have a laptop that can run Windows, even if it is a Mac**. Please install both Microsoft Access and MySQL on Windows to allow these programs to communicate when we connect them later in the semester. Excel, R, and SAS software can be installed on either platform.

Weill Cornell Medicine students can download Microsoft Office software (which includes Excel and Access) for free by following these instructions:
https://its.weill.cornell.edu/sites/default/files/o365_for_home_computers.pdf

MySQL software is also free and can be downloaded from https://dev.mysql.com/downloads/. Installation instructions will be posted to Canvas toward the middle of the semester.

Assessment and Grading Policy

Student grades will be based on:

In-Class Assignments 30%
Quizzes 20%
Final Project 30%
Final Exam 20%

In-Class Assignments: (30% of final grade) We will cover 1 to 2 programming assignments in class per week (17 in total), which will be graded out of 100%. The in-class assignments are designed to give practice implementing the concepts learned in lectures, and to gain experience with troubleshooting issues that may arise with various techniques. You will work in pairs for each assignment, which are randomly selected for each week of the semester (see "People" section on Canvas for team assignments). Each member of the group should submit an attestation in the comments of the assignment submission stating **"I have contributed significantly to this assignment"** in order to receive a grade for that assignment.

You are highly encouraged to take advantage of the flipped classroom format and seek feedback from peers and the teaching team as you work on assignments, but you each team is expected to submit a unique solution for grading.

Grading rubrics will be posted simultaneously with the assignment so that you may keep the grading scheme in mind as you work. In addition to a completed grading rubric, you will receive detailed feedback to help you improve your skills and clarify concepts that you may have missed in the assignment.

Assignments have been designed to be completed during the duration of the weekly class meetings, but the official deadline for all assignments will be at 11:59pm following class on Wednesdays. Assignments submitted past the deadline without prior approval will have 10% out of the 100% total deducted per day, and if an assignment is not submitted prior to the next class meeting, no credit will be given.

Your final in-class assignment grade will be calculated as follows:

In-Class Assignment Average Score (out of 100%) * 30%

Quizzes: (20% of final grade) In lieu of a midterm exam, you will take weekly quizzes in class every Wednesday at the end of class, to assess your mastery of course objectives. Lecture videos for the following week will be unlocked immediately after class on Wednesdays. Each video contains demonstrations that are designed to help build your skills on various topics. You should follow along with each demo, using the files provided, to prepare for the following week's in-class assignments and quiz.

Quizzes will be relatively short (~10 questions) and will consist of a mix of true/false, multiple choice, fill in the blank, matching, and short answer questions. **You may <u>not</u> refer to course materials or to the internet while taking quizzes.** There will be a 20-minute time limit for each quiz, and the grades will be scored out of 100%. **No make-up quizzes will be given.**

Your lowest quiz grade will be dropped prior to calculating your final quiz grade, which will be calculated as follows:

Average Quiz Score (out of 100%) * 20%

Final Project: (30% of final grade) For your final project, you will be asked to build a small, normalized, relational database in Microsoft Access using all of the skills you learned during the semester. The database should have a healthcare-related focus, but you may choose another topic with permission. The final project will be "scaffolded," with portions of it due at 6 intervals throughout the semester, instead of the entire project being due at once. An overview of the

project and its components and deadlines is available on Canvas. Specific requirements for individual components of the project will be posted separately, as they are assigned. Each portion of the final project will have its own grading rubric, which will be provided to you in advance. Each rubric will be returned to you with feedback within approximately 1-2 weeks of submission. For your last submission, you will be allowed to make corrections to recoup up to half credit on points lost in the initial submissions. This will also allow you to have a work sample for your portfolio that is as error-free as possible. The last submission of the final project will be due on the last day of class, **Wednesday, April 3rd, by 11:59pm.**

As with the in-class assignments, you will work on the final project in pairs. While in-class assignment pairs are randomly selected, you will be allowed to choose your own pairs for the final project. Instructions for team sign-ups will be given a few weeks into the semester, once the first portion of the project has been assigned.

Your final project grade will be calculated as follows:

Total Score from Final Submission (out of 100%) * 30%

Final Exam: (20% of final grade) In addition to your final project, you will also have a comprehensive final exam to assess your overall knowledge from the course, especially on material not included in the final project. The exam will be given during **the week of April 8th** (exact date and time TBA) and will be comprehensive, covering material from weeks 1-13. The exam will be given on paper and will be in a similar format to weekly quizzes.

Your final exam grade will be calculated as follows:

Final Exam Score (out of 100%) * 20%

Letter Grades: Letter grades will be assigned as follows, based on your overall final score:

A+	98-100%	4.3
A	93-97%	4.0
A-	90-92%	3.7
B+	88-89%	3.3
B	83-87%	3.0
B-	80-82%	2.7
C+	78-79%	2.3
C	73-77%	2.0
C-	70-72%	1.7
D	60-69%	1.3
F	Below 60%	0

Academic Integrity

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity: https://cuinfo.cornell.edu/aic.cfm. Any work submitted by a student in this course for academic credit will be the student's own work. **For this course, collaboration is allowed only for in-class assignments and the final project.**

You are encouraged to study together and to discuss information and concepts covered in lecture and the sections with other students. You can give "consulting" help to or receive "consulting" help from such students. However, this permissible cooperation should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an e-mail, an e-mail attachment file, a diskette, or any other digital or hard copy. **Additionally, you may never submit any part of a class assignment solution file as your own work.**

Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

During examinations, you must do your own work. Talking or discussion is not permitted during the examinations, nor may you compare papers, copy from others, or collaborate in any way. Any collaborative behavior during the examinations will result in failure of the exam, and may lead to failure of the course and University disciplinary action.

Accommodations for Students with Disabilities

In compliance with the Cornell University policy and equal access laws, I am available to discuss appropriate academic accommodations that may be required for student with disabilities. Requests for academic accommodations are to be made during the first three weeks of the semester, except for unusual circumstances, so arrangements can be made. Students are encouraged to register with Student Disability Services to verify their eligibility for appropriate accommodations.

Inclusivity Statement

We understand that our members represent a rich variety of backgrounds and perspectives. The Biostatistics and Data Science Program is committed to providing an atmosphere for learning that respects diversity. While working together to build this community we ask all members to:

- Share their unique experiences, values and beliefs.
- Be open to the views of others.
- Honor the uniqueness of their colleagues.
- Appreciate the opportunity that we have to learn from each other in this community
- Value each other's opinions and communicate in a respectful manner.
- Keep confidential discussions that the community has of a personal (or professional) nature.
- Use this opportunity together to discuss ways in which we can create an inclusive environment in this course and across the Cornell community.

Course Schedule

The following is a tentative course schedule for the Spring 2019 semester. The schedule may change to accommodate the needs of the class.

All course-related documents can be found under the **Files** section of Canvas.
Lecture videos can be accessed under the **Modules** section of Canvas.
Assignments and quizzes should be submitted through the **Assignments** section of Canvas.

Week 1 – Course Introduction, Limitations of Excel for Data Storage

Learning Objectives:

- Define course objectives and logistics
- Describe limitations of using Excel spreadsheets for data storage

1/7/19 Lectures (In Class):

- Course/Syllabus Overview
- Limitations of Excel for Data Storage

Quiz: Week 1 Quiz: Introduce Yourself!

Assignments:

- In-Class Assignment 1 (Excel) – Due Wednesday, 1/9 at 11:59pm
- In-Class Assignment 2 (Excel) – Due Wednesday, 1/9 at 11:59pm

1/9/19 Assignments (continued):

- In-Class Assignment 1 (Excel) – Due Wednesday, 1/9 at 11:59pm
- In-Class Assignment 2 (Excel) – Due Wednesday, 1/9 at 11:59pm

Continued on next page

Week 2 – Relational Databases, Table Design

Learning Objectives:

- Define database, relational database
- Build database tables in Access containing:
  - variable names that follow good naming conventions
  - correct data types assigned to each variable
  - correct identifiers selected as primary key (composite and autonumber)
  - indexes on variables, where appropriate
  - validation rules applied to variables, where appropriate
  - a combo box linked to a lookup table
  - no HIPAA identifiers
- Connect tables to create relationships that:
  - are joined on the correct fields
  - have referential integrity enforced
  - show the correct relationship type (e.g., 1-to-1, 1-to-many)
  - have cascade update/delete selected appropriately

1/14/19 Lectures (Due Today): Lecture Videos for Week 2

Assignments:

- In-Class Assignment 3 (Access) – Due Wednesday, 1/16 at 11:59pm
- Final Project Part 1: Proposal – Due Wednesday, 1/23 at 11:59pm

1/16/19 Assignments (Continued):

- In-Class Assignment 3 (Access) – Due Wednesday, 1/16 at 11:59pm
- Final Project Part 1: Proposal – Due Wednesday, 1/23 at 11:59pm

Quiz: Week 2 Quiz

Continued on next page

Week 3 – Data Anomalies, Normalization Rules

Learning Objectives:

- Describe insertion, deletion, and update anomalies
- Build database tables normalized up to 3rd Normal Form
- Compare normalization rules to "Tidy Data" rules

1/21/19 NO CLASS – MARTIN LUTHER KING, JR. HOLIDAY

1/23/19 Lectures (Due Today): Lecture Videos for Week 3

Assignments:

- In-Class Assignment 4 (Access) – **Due Friday, 1/25 at 11:59pm** →
- Final Project Part 1: Proposal – Due Wednesday, 1/23 at 11:59pm

Quiz: Week 3 Quiz

Week 4 – SQL Data Manipulation Language (DML)

Learning Objectives:

- Describe the types of operations performed in the SQL DML
- Write SQL queries containing the following clauses: SELECT, FROM, WHERE, GROUP BY, HAVING, ORDER BY, INSERT INTO, UPDATE, DELETE
- Differentiate query results produced with outer vs. inner joins
- Write SQL queries containing outer and inner joins
- Write SQL queries containing a subquery
- Add aggregate functions and functions that calculate across columns to SELECT queries
- Compare the use of joins vs. subqueries when multiple tables are involved in one query
- Apply SQL Style Guide conventions to SQL query syntax

1/28/19 Lectures (Due Today): Lecture Videos for Week 4

Assignments:

- In-Class Assignment 5 (Access) – Due Wednesday, 1/30 at 11:59pm
- In-Class Assignment 6 (Access) – Due Wednesday, 1/30 at 11:59pm
- Final Project Part 2: Tables – Due Wednesday, 2/6 at 11:59pm

1/30/19 Assignments (Continued):

- In-Class Assignment 5 (Access) – Due Wednesday, 1/30 at 11:59pm
- In-Class Assignment 6 (Access) – Due Wednesday, 1/30 at 11:59pm
- Final Project Part 2: Tables – Due Wednesday, 2/6 at 11:59pm

Quiz: Week 4 Quiz

Week 5 – SQL Data Definition Language (DDL)

Learning Objectives:

- Describe the types of operations performed in the SQL DDL
- Write SQL queries to create and alter database tables

2/4/19 Lectures (Due Today): Lecture Videos for Week 5

Assignments:

- In-Class Assignment 7 (Access) – Due Wednesday, 2/6 at 11:59pm
- Final Project Part 2: Tables – Due Wednesday, 2/6 at 11:59pm

2/6/19 Assignments (Continued):

- In-Class Assignment 7 (Access) – Due Wednesday, 2/6 at 11:59pm
- Final Project Part 2: Tables – Due Wednesday, 2/6 at 11:59pm

Quiz: Week 5 Quiz

Week 6 – SQL in R and SAS

Learning Objectives:

- Write SQL queries using the 'sqldf' package in R
- Write SQL queries using proc sql in SAS

2/11/19 Lectures (Due Today): Lecture Videos for Week 6

Assignments:

- In-Class Assignment 8 (R) – Due Wednesday, 2/13 at 11:59pm
- In-Class Assignment 9 (SAS) – Due Wednesday, 2/13 at 11:59pm
- Final Project Part 3: Queries – Due Wednesday, 2/20 at 11:59pm

2/13/19 Assignments (Continued):

- In-Class Assignment 8 (R) – Due Wednesday, 2/13 at 11:59pm
- In-Class Assignment 9 (SAS) – Due Wednesday, 2/13 at 11:59pm
- Final Project Part 3: Queries – Due Wednesday, 2/20 at 11:59pm

Quiz: Week 6 Quiz

Week 7 – Data Entry Forms

Learning Objectives:

- Create new data entry forms
- Add objects to forms
- Modify form and object properties
- Add subforms and ensure their correct connection to main form
- Add buttons to forms to open/close forms and navigate between records
- Write VBA code to alter form and object properties, and to perform data validation between different variables
- Describe the purpose of Data Entry Mode
- Apply good design practices in form layout and formatting (e.g., labels, colors, object spacing and alignment)

2/18/19 NO CLASS – PRESIDENTS' DAY HOLIDAY

2/20/19 Lectures (Due Today): Lecture Videos for Week 7

Assignments:

- In-Class Assignment 10 (Access) – **Due Friday, 2/22 at 11:59pm**
- Final Project Part 3: Queries – Due Wednesday, 2/20 at 11:59pm

Quiz: Week 7 Quiz

Week 8 – Reports

Learning Objectives:

- Design report to show summary of records without grouping
- Design report to show summary of records with grouping
- Apply good design practices in report layout and formatting (e.g., labels, colors, object spacing and alignment)

2/25/19 Lectures (Due Today): Lecture Videos for Week 8

Assignments:

- In-Class Assignment 11 (Access) – Due Wednesday, 2/27 at 11:59pm
- Final Project Part 4: Forms – Due Wednesday, 3/6 at 11:59pm

2/27/19 Assignments (Continued):

- In-Class Assignment 11 (Access) – Due Wednesday, 2/27 at 11:59pm
- Final Project Part 4: Forms – Due Wednesday, 3/6 at 11:59pm

Quiz: Week 8 Quiz

Week 9 – Introduction to MySQL

Learning Objectives:

- Locate relevant parts of MySQL Workbench interface
- Open and save new SQL scripts
- Run queries using SQL DML

3/4/19 Lectures (Due Today): Lecture Videos for Week 9

Assignments:

- In-Class Assignment 12 (MySQL) – Due Wednesday, 3/6 at 11:59pm
- Final Project Part 4: Forms – Due Wednesday, 3/6 at 11:59pm

3/6/19 Assignments (Continued):

- In-Class Assignment 12 (MySQL) – Due Wednesday, 3/6 at 11:59pm
- Final Project Part 4: Forms – Due Wednesday, 3/6 at 11:59pm

Quiz: Week 9 Quiz

Week 10 – SQL DDL and SQL DML in MySQL

Learning Objectives:

- Create a new schema in MySQL
- Add and modify tables within a schema
- Set primary keys, foreign keys, and indexes within tables
- Produce EER diagram to show table relationships
- Populate and retrieve data using SQL DML

3/11/19 Lectures (Due Today): Lecture Videos for Week 10

Assignments:

- In-Class Assignment 13 (MySQL) – Due Wednesday, 3/13 at 11:59pm
- Final Project Part 5: Reports – Due Wednesday, 3/20 at 11:59pm

3/13/19 Assignments (Continued):

- In-Class Assignment 13 (MySQL) – Due Wednesday, 3/13 at 11:59pm
- Final Project Part 5: Reports – Due Wednesday, 3/20 at 11:59pm

Quiz: Week 10 Quiz

Week 11 – SQL Misc: Views, Case Statements, Triggers, Special Joins, USING

Learning Objectives:

- Define view and describe its uses
- Use a case statement within a SELECT query to conditionally recode a variable
- Apply triggers to validate data
- Write SQL queries that apply self-joins, cross joins, and full joins appropriately
- Write joins with USING as an alternative to ON

3/18/19 Lectures (Due Today): Lecture Videos for Week 11

Assignments:

- In-Class Assignment 14 (MySQL) – Due Wednesday, 3/20 at 11:59pm
- In-Class Assignment 15 (MySQL) – Due Wednesday, 3/20 at 11:59pm
- Final Project Part 5: Reports – Due Wednesday, 3/20 at 11:59pm

3/20/19 Assignments (Continued):

- In-Class Assignment 14 (MySQL) – Due Wednesday, 3/20 at 11:59pm
- In-Class Assignment 15 (MySQL) – Due Wednesday, 3/20 at 11:59pm
- Final Project Part 5: Reports – Due Wednesday, 3/20 at 11:59pm

Quiz: Week 11 Quiz

Week 12 – Connecting an Access Front End to a MySQL Back End

Learning Objectives:

- Develop a “front end” interface in Access for MySQL tables
- Apply a password to the Access “front end”
- Establish ODBC connection between Access and MySQL

3/25/19 Lectures (Due Today): Lecture Videos for Week 12

Assignments:

- In-Class Assignment 16 (MySQL and Access) – Due Wednesday, 3/27 at 11:59pm
- Final Project Part 6: Corrections – Due Wednesday 4/3 at 11:59pm

3/27/19 Assignments (Continued):

- In-Class Assignment 16 (MySQL and Access) – Due Wednesday, 3/27 at 11:59pm
- Final Project Part 6: Corrections – Due Wednesday 4/3 at 11:59pm

Quiz: Week 12 Quiz

Week 13 – Pulling Web Data Using APIs

Learning Objectives:

- Define API
- Extract data from the Web via API using the httr package in R

4/1/19 Lectures (Due Today): Lecture Videos for Week 13

Assignments:

- In-Class Assignment 17 (R) – Due Wednesday, 4/3 at 11:59pm
- Final Project Part 6: Corrections – Due Wednesday 4/3 at 11:59pm

4/3/19 Assignments:

- In-Class Assignment 17 (R) – Due Wednesday, 3/20 at 11:59pm
- Final Project Part 6: Corrections – Due Wednesday 4/3 at 11:59pm

Quiz: Week 13 Quiz

Final Exam

Week of 4/8/19 Final Exam – Time and location TBA