OIDD 105: Developing Tools for Data Access and Analysis (Fall 2017)
Lorin M. Hitt (lhitt@wharton.upenn.edu)

This course introduces the construction and use of data analysis tools that are commonly used for business analysis, especially in consulting and finance. The course builds on the spreadsheet and analytical skills developed in OIDD101, providing a much more extensive treatment of spreadsheet application development (using *Visual Basic for Applications*) and database management (using SQL as implemented in the SQL Server database system – *Transact-SQL*). In addition, the course will cover related skills needed to effectively develop computer applications including specification development, interface design, and testing, as well as more specialized topics directly related to the use of databases such as data structures, database design, and data security. Time permitting, we will also explore analytics tools such as Hadoop or some of the data mining capabilities of SQL server. This course is not a substitute for a computer science course in programming, algorithm design or databases, but will draw on computer science concepts as needed to provide a working knowledge of the necessary theory.

The course is intended for students without prior experience in programming[1], but students must have familiarity with computer-based tools as covered in OIDD101 or equivalent, or through personal experience.   The course is definitely INTRODUCTORY in that it does not require prior knowledge of the material. That does not mean it will be easy since computers can be unforgiving when you make a programming mistake and some concepts, like object orientation and set-based reasoning, are intellectually challenging. We expect the course to be especially useful for students seeking entry-level analyst positions in data-intensive firms, or those generally seeking to broaden their knowledge and skills in the construction and use of computer-based analytical tools. The course counts toward the general OIDD concentration and the Information Systems and Business Analytics tracks.

Course Format: The course will have regular meetings where we discuss the material. In addition, there will be a series of programming or analysis tasks which are expected to be completed on your own prior to attending class (these are to be submitted but will not be graded for anything other than "being there").   In some cases, I will hold optional lab sessions to assist in these class preparation exercises. There will also be a series of graded exercises (1 required, and 4 others of which you must complete 3) and two in-class Quizzes.

---

[1] This course is not intended for students with an extensive computer science background. Students in this category are strongly encouraged to take an alternative class such as OPIM311 (Business Computer Languages) instead of OPIM105.   Please contact me immediately if you think you fall into this category and we can figure out the appropriate course of action.

Course Materials. There are two **required** texts:

(PPVBA) Alexander and Kusleika (2016). *Excel 2016 Power Programming with VBA*. (ISBN: 1119067723)

 (SQLD) Syyverson and Murach (2016). *Murach's SQL Server 2016 for Developers* (ISBN: 1890774960)

These are trade books and available from a wide variety of sources, including Amazon and other discounters, in both paper and digital form. Note that the digital version of Murach uses a non-standard copy protection scheme ("Locklizard") and is therefore not recommended. Neither of these books are free (they are also not very expensive) so please do not use illegal copies.

Mandatory Computer Resources

While you can use the labs, you will find yourself at a disadvantage in the course if you rely strictly on them for computing resources (disadvantage = measurably lower grade!). If you have a PC with Windows 7 or better you are in good shape. If you have a relatively new Mac (excluding the Macbook Air), you can make it work with a little effort.

We will be using:

Office 365 for Windows ("Excel" and "Access"). Available for free through your Penn O365 account or through the Wharton virtual labs. The native Mac version is not acceptable and cannot be used.

SQL Azure ("Azure"). A basic free account is available to all students. If you need more for your semester project, a bigger "free trial" account is available. Details will be provided at the beginning of the semester.

Microsoft SQL Server Management Studio ("SSMS") (free, preferred, Windows-only). Windows users can install it directly. Mac users can access SSMS through a server set up for this class or you can use a native SQL client (I recommend *Navicat Essentials for SQL Server,* available directly from Navicat for $80).

We may be using additional software for the analytics sessions (TBD).

 (optional but recommended). If you are using a laptop, get an external mouse. This will increase your programming productivity significantly (best $5 you will ever spend!). I also highly recommend that programmers use large screens (24" or better). Studies have shown this increases developer productivity.

<u>…But can I use the (physical) labs?</u>

Yes..but…. Excel and Access run just fine in both the physical and virtual labs. We will also have SSMS installed in at least one lab. However, if you rely entirely on the labs you will be spending lots of time in the labs. Since I have found that this can be a problem, as part of Exercise 1 I am requiring everyone to demonstrate the ability to run VBA and SQL on their own machine.

<u>… but what about Macs?</u>

We have many options for using Macs that work ok. However, this is one situation where things really are better on the PC since neither VBA nor SSMS run natively on the Mac. You should be fine with some combination of: virtual labs (for Excel), remote desktop (for SSMS), a native SQL client for the Mac (Navicat replaces SSMS), or working in the labs. If you already have Windows running on your Mac (either through a virtual machine or bootcamp) you can also try to use the native applications. However, the cost-benefit ratio of this option is not enough for us to recommend setting up Windows on your Mac if you don't already have it working.

<u>Grading and Evaluation</u>.

Graded Exercises (30% of final grade). There is one required activity (Exercise 1) and four additional Exercise opportunities (I will take the best 3 of 4, so you can skip one or do them all to try to get a better score). Exercises may be done in groups of no more than 3 (except Exercise 1 which is individual).

Pre-Class Preparation (10%). Most class sessions will have some type of preparatory work. I will request submission of some of these prior to class and they will be graded lightly (full credit for being there on time). You will be able to skip at least two for full credit.

Class Project (10%). There is a class project which is either a) build something interesting, or b) do an interesting data analysis. There are two deliverables: a proposal due just after mid semester the final product which includes a 5-page writeup of what you did and supporting code/data/analysis files. Ideal group size is 3. I will consider larger groups on an individual basis for larger/more ambitious projects.

Quizzes (40%). There are two in-class quizzes, equally weighted and non-cumulative. The first is on VBA and the second is on Database. There is no final exam and we will NOT be using the scheduled final exam time.

Class Participation (10%). Students are expected to prepare, attend class, actively participate, and make good use of course resources (including the support staff and the instructors out of class

time). The class participation grade will reflect our subjective evaluation on these dimensions as well as objective observation of class attendance (see below).

Grade Distribution. There is no pre-specified grade distribution. Historically, we gave approximately 40% A's and 60% B's. Most of the variance in grades is driven by quiz scores (homework scores tend to have modest variation other than missed/late assignments). Grades lower than a "B" are unlikely if you complete all the assigned work and otherwise follow course guidelines.

Other Course Policies

Regrades. Any requests for regrades should be submitted in writing to your assignment submission folder before the next assignment is due. The request must be labeled clearly and explain why you believe your answer is correct.   Please note that we do not consider regrade requests regarding partial credit awarded to incorrect answers (in other words, if your answer is not correct, it is not eligible for regrade consideration).

Deadlines. Assignment deadlines are firm because we often review the assignments in class immediately following the deadline. If for some reason you are not able to complete an assignment (e.g., you can't get your program to work…) submit what you have by the deadline.   If you have a conflict on a deadline date, skip the assignment or submit it early. Note that you are permitted to skip at least one exercise and at least two of the class preparatory activities.

Collaboration. You are free to discuss any and all course material with your fellow students and the course staff, including approaches to the assignments. You can also work together on most assignments in small groups. However, you are not allowed to share code or answers on any graded assignments outside your small work team or copy code for the assignments graded for "being there". You are also not permitted to use materials from prior iterations of OPIM105 or OIDD105 in preparing your written work or to copy code directly from Internet sources. All collaborators or should be identified by name in the submitted documents (distinguishing between your work team and anyone you spoke with in preparation of the assignment). If you worked in a group for the assignments, you should submit a common paper for the assignment. Doing an assignment in a group and then creating a private version of the group work violates the "no sharing of code" guideline and is not allowed.   You are not required to work in a group.

Regardless…We strongly discourage "divide and conquer" strategies on assignments where questions are divided among group members or "you drive, I watch" programming where one student writes all the code and the other watches, gets coffee, etc. You cannot learn these skills without actual personal experience. Programmers write code, and you can't write and test code without touching the computer.

Attendance. You are expected to come to class and to be prepared. From time to time, something may happen in class that requires your physical presence like a cold call, returned assignment, or an unannounced in-class exercise. You are permitted to miss two of these over the course of the semester before it affects your grade (this is in addition to any University-approved absences such as religious observances).   You do not need to tell me why you are missing class or get permission. If you need to miss class due to a religious holiday, I am happy to go over the material by appointment or during office hours or to record a session of the class by request.

Support. There will be office hours by both the instructors as well as undergraduate and graduate teaching assistants. We will be using Piazza, and online discussion tool, for online course questions. A few guidelines about the use of Piazza which will make everyone happier:

- If you have a general question or something about the course material, use Piazza. If you have a personal question, e-mail the instructor.
- Please do post code to Piazza as an open message. If you need a quick evaluation of your code, post it is a private message to instructors. If you have a more complicated question ("why doesn't this work?") that is probably best done in person or by e-mail.
- Please do not spam questions on Piazza. If you have lots of questions, come see me or someone on the course staff.
- You can make your questions anonymous to other students but the instructors and TAs can see your real name… so be nice.
- You too can answer questions on Piazza. This is appreciated by the course staff.

Electronics. Unless you are taking notes electronically, computers and tablets are not permitted in class. If you do wish to use your computer to take notes, please come see me for permission. You are not permitted to make audio or video recordings of class sessions. Cell phones should be turned off or put in airplane mode. If you must take a call or respond to a message, please leave the room. Anyone using an electronic device in class without permission will not receive credit for the class (no warnings).

Preliminary Schedule (Subject to substantial change)

Date	Day	Session	Assignments
8/30/2017	Wed	Course Introduction	
9/4/2017	Mon	No Class (labor day)	
9/6/2017	Wed	Excel Review/Introduction to VBA	
9/8/2017	Fri		Exercise 1: Get the Tools
9/11/2017	Mon	Programming and Functions (I)	
9/13/2017	Wed	Programming and Functions (II)	
9/18/2017	Mon	Algorithms and Complexity	
9/20/2017	Wed	Subroutines (I)	Exercise 2: Function
9/25/2017	Mon	Subroutines and Error Handling	
9/27/2017	Wed	User Interfaces	
10/2/2017	Mon	Introduction to Object Orientation	
10/4/2017	Wed	XML and Web Services	Exercise 3: Subroutine
10/9/2017	Mon	Built-In Objects	
10/11/2017	Wed	Review Session	
10/16/2017	Mon	Quiz I (in class)	Quiz I
10/18/2017	Wed	SQL: Single Table Queries	
10/23/2017	Mon	SQL: Advanced Queries	
10/25/2017	Wed	Relational Database Concepts	Project Proposal (Due Friday)
10/30/2017	Mon	SQL: Relational Joins	
11/1/2017	Wed	SQL: Complex Joins/Subqueries	
11/6/2017	Mon	DDL and Scripting	
11/8/2017	Wed	Complex Joins/Subqueries	Exercise 4: Analysis with SQL
11/13/2017	Mon	DDL and Scripts	
11/15/2017	Wed	VBA-SQL Integration	
11/20/2017	Mon	Regular Expressions	
11/22/2017	Wed	No class (schedule shift)	
11/27/2017	Mon	Review Session	
11/29/2017	Wed	Quiz II (in class)	Quiz II
12/4/2017	Mon	Analytics Session I (TBD)	
12/6/2017	Wed	Analytics Session I (TBD)	
12/11/2017	Mon	Analytics Session I (TBD)	Exercise 5: Analytics/Projects "due"*

* No late penalty for project submissions up until noon on 12/20

There is no final exam and we will not be using the scheduled exam times for anything