OIDD 105: Developing Tools for Data Access and Analysis (Fall 2016)
Lorin M. Hitt (lhitt@wharton.upenn.edu)

This course provides an introduction to 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 "big data" tools such as Hadoop/Hive. The 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 and necessary skills.

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 OPIM/OIDD concentration and the Information Systems and Business Analytics tracks.

Course Format: The course will meet in the JMHH375 Computer Lab. Most sessions will include at least half the time on in-class computer exercises. While this will likely lighten the out of class

---

[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. Given enrollment, space for auditors will be limited and will require a personal commitment to attend class and complete all assignments.

preparation the tradeoff is that attendance will be mandatory this semester and some class exercises will be graded evaluations.

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. The Mac version (either 2016 or 2011) is not acceptable and cannot be used.

SQL Azure ("Azure"). I have arranged for free accounts for all students. Details will be provided at the beginning of the semester.

Microsoft SQL Server Management Studio ("SSMS") (free, preferred) or *Navicat Essentials for SQL Server* (*"Navicat"*) ($80, for Mac users only). These are database access tools. I also encourage you to install the full version of SQL Server on your local machine if you have a PC or are running Windows on Bootcamp although it is not necessary given you will have an Azure account.

Tableau 9 ("Tableau") (free) for PC or Mac. I have arranged for free versions for all students.

(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.

…But can I use the (physical) labs?

Mostly yes. Excel and Access run just fine in the labs. I am told that SSMS will be available in JMHH375 (lab). With SSMS and SQL Azure you should be fine. However… you will be spending a LOT of time in the labs if you pursue this route and I do not recommend it.

… but what about Macs?

Using a Mac is becoming easier but is still not as simple as it could be.   Here are some suggestions.

1) Get a PC. Decent laptop PCs can be had for less than $500 from Acer, HP, Dell, Lenovo and Toshiba. Decent desktops can be had for less than $300 if you already have a screen, keyboard and mouse. This is the minimum nuisance option.

2) Install Bootcamp/Windows 7 on your Mac.[2] We can provide Windows 7 for use for the class which you can install on your Mac using Bootcamp. Provided your Mac can handle it, this is a good option. You will need an external mouse if you do this.

3) Virtual Labs. This is a good solution for Excel/VBA. You will not have a database client, however.   To fill in the gap, you can buy Navicat Essentials for SQL Server ($80, online) which will provide the minimum required capability to access the database. Alternatively, you can do all the database work in JMHH375 (again, not recommended).

Grading and Evaluation. The final grade will be based on performance in three areas: Assignments, Quizzes, and Class Participation.

Assignments (40%): The assignments in this course are critical, and probably where you will do the most learning.   There will be something to turn in almost every session as well as a series of at-home exercises ("short assignments"), although most will be graded for "being there". There will also be 3-4 larger assignments, and a "build something" final project that has the same weight as a larger assignment and will be graded more carefully. In-class assignments can usually be done collaboratively. Larger exercises can be done in small groups (no more than 3).   You will be permitted to miss one in-class exercise and one short assignment.

---

[2] In the past I have recommended using VMWare or Parallels to run Windows as a virtual machine on your Mac. I no longer recommend this option due to technical and performance limitations and it is no longer supported.

Quizzes (50%): There will be two quizzes that are equally weighted, one on Programming/VBA and the other on Databases/SQL. They are scheduled roughly at the 1/3 and 2/3 points in the semester although they may get pushed back depending on the schedule. There is no midterm or final exam. These exams will happen when we finish the material.

Class Participation (10%). Students are expected to prepare the readings, 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. This will have a neutral effect on grades for most students.

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 personal conflict with any of the assignment due dates, please arrange to complete the assignment early. For exercises graded for "being there" just turn in what you have by the deadline (it doesn't have to work or be complete).

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 in the lab 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 to copy code for the assignments graded for "being there". All collaborators 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.

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 (please limit your use of e-mail to personal communications with the instructors – this is always welcome but should not be the primary source of information). Any question requiring more than a two or three line reply, or more than a minute or two of examining code should be done during office hours. Open ended questions ("can you help me, I don't know what is wrong…") should also be handled in person. Please do not post sample solution code to Piazza – if you need code evaluation, send your spreadsheet or SQL code by e-mail or bring it in person. Please also note that while you can keep your Piazza comments anonymous to other students, instructors and TAs can see your real name… so be nice.

PRELIMINARY: SUBJECT TO SUBSTANTIAL CHANGE.

EX=Graded Exercise (larger assignment); SA = Short Assignment (graded for completion). In-class assignments are not listed as they will be determined as we go.

Date	Day	Session	Outside of Class Assignments
8/31/2016	Wed	Course Introduction	
9/5/2016	Mon	No Class (labor day)	
9/7/2016	Wed	Excel for Analysts	SA0: Tools Screen Shot Due (Friday)*
9/12/2016	Mon	Basic Programming and Functions	
9/14/2016	Wed	Programming and Functions (II)	SA1: Function
9/19/2016	Mon	Algorithms and Complexity	EX1: Excel Data Analysis
9/21/2016	Wed	Subroutines (I)	SA2: Subroutine
9/26/2016	Mon	Subroutines and Error Handling	
9/28/2016	Wed	User Interfaces	EX2: Functions (Due Friday)
10/3/2016	Mon	Object Orientation	SA3: UI
10/5/2016	Wed	More Objects	
10/10/2016	Mon	Subroutine Exercise Discussion	EX3: Subroutines
10/12/2016	Wed	XML and Web Services	
10/17/2016	Mon	Testing	
10/19/2016	Wed	Review Session	
10/24/2016	Mon	Quiz I (in class)	Quiz I (in class)
10/26/2016	Wed	Single Table Queries	
10/31/2016	Mon	Complex Queries	SA4: Query Practice I
11/2/2016	Wed	Relational Database Intro	
11/7/2016	Mon	Relational Joins	
11/9/2016	Wed	Complex Joins/Subqueries	SA5: Query Practice II
11/14/2016	Mon	DDL and Scripts	
11/16/2016	Wed	VBA-SQL Integration	
11/21/2016	Mon	Regular Expressions	SA6: Scripting/Integration
11/23/2016	Wed	No class (schedule shift)	
11/28/2016	Mon	Database Exercise Discussion	EX4: Database
11/30/2016	Wed	Review Session	
12/5/2016	Mon	Quiz II (in class)	Quiz II
12/7/2016	Wed	Big Data I: Data Mining in SQL	
12/12/2016	Mon	Big Data II: Hadoop	SA7: TBD/optional
12/16/2015	Wed	Scheduled Exam	Quiz II