**OPIM 105:  Developing Tools for Data Access and Analysis (Fall 2013)**
**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 OPIM101, 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.  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 OPIM101 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 concentration and the Information Systems and Business Analytics tracks.

Course Materials.  There are two required texts:

(PPVBA) Walkenbach, J. (2013).  *Excel 2013 Power Programming with VBA*. (ISBN:  1118490398)

 (SQLD) Syverson, B and Murach, J. *Murach's SQL Server 2012 for Developers* (ISBN: 1890774693)

These are trade books and available from a wide variety of sources, including Amazon and other discounters, in both paper and digital form.  PPVBA provides a free PDF copy if you buy the book in paper; you can buy a digital copy of SQLD but it uses a nonstandard copy protection scheme

---

[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.

that you have to install ("Locklizard").  These books are not free, so please do not download pirated 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!).

You will need access to a computer that can run Windows (XP SP3 or greater, Windows 7 or better recommended) with the following software tools installed:

Office for Windows that includes Excel and Access.  Versions prior to Office 2010 are not recommended.  There are many ways to get Office inexpensively (less than $100) including Microsoft student programs and the Computer Connection.  Use of Office versions prior to 2010 is not supported or recommended and will cause problems.  Access will only be used for one session, so a trial version would be acceptable.

Microsoft SQL Server Express 2012 with Tools.  This is a free download from Microsoft (about 660MB).[2]

SQL Azure.  I will provide access keys for SQL Azure which is a "cloud" database server.  You will be required to use SQL Azure for some of the homework assignments.  If you are using a Mac, you may find SQL Azure better for doing the class preparation material as well (see below).

(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!).

…But can I use the labs?

Yes and No.  Excel and Access run just fine in the labs.  However, SQL Server can only be installed on a virtual machine on certain lab computers (currently, the first two rows of F75).  This copy of SQL server works fine for class preparation, but it is too slow to be used for the assignments.  Regardless, unless you plan to spend some serious time in the labs, you are better off having your own machine for this class.

---

[2] The SQL Server Express 2012 page is:  http://www.microsoft.com/en-us/download/details.aspx?id=29062. Select the appropriate edition (32-bit or 64-bit depending on your operating system configuration).  You will need the database engine and Express Tools (the minimal install package is ENU\x64\SQLEXPRWT_x64_ENU.exe;  the package with advanced services is also fine but about twice as large and I do not expect to need the extra tools in this package).

**… but what about Macs?**

Macs older than about 3 years, and low performance Macs (e.g. MacBook Air) cannot run the requisite software.  Newer, higher performance Macs, can be fine as long as you do the following:

1) Install a virtualization engine like VMWare Fusion or Parallels.

2) Install a legal copy of Windows 7 into the virtual machine.  Verify that your Mac can access the Internet from Windows (this can be a problem with some older MacBooks and renders the machine unusable for the class).

3) Install the required course software (SQL Server, Excel for Windows – trial or full version)

We can provide free academic copies of VMWare Fusion and Windows 7 for use with the course.  For Windows, you will have to come to the OPIM department and install it from disk there (we cannot lets these disks leave the department offices).  This should be done in the first two weeks of class because you will need it to do the screen shot in Short Assignment 1.  There may be some variations on the best procedure when the semester rolls around and I'll update this process then.

NOTE: After 3 years of trying various variations on this policy, this is the only successful strategy for using your Mac with the class.

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

*Assignments (40%)*:  Since the course is focused on building and using tools, the majority of the class grade is determined by the course assignments.  Expect there to be something "due" every week although most of these will be short preparation exercises  - the short assignments are should take a modest amount of time and are graded for "being there". There will be 3-4 larger problem sets which are graded more carefully and two quizzes (the second of which will be on the Final Exam day).  Preparation exercises and quizzes must be done alone; problem sets can be done in small groups (no more than 3). All submissions are electronic through Canvas.

*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, but there may be a larger last assignment due at the end of class.

*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.

*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).

<u>Other Course Policies</u>

*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.

*Collaboration*.  You are free to discuss any and all course material with your fellow students and the course staff.  However, any Assignments must be prepared solely by you, or, for group assignments, members of your group.  You are still free to discuss your approach with fellow students, but you are not permitted to share code or analysis results (in any form).  You can't groupsource answers when you are working in a consulting firm or investment bank, so you may as well get used to it now.  We also 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.

All collaborators (other than course staff) should be identified in your assignment write-up.  Any submitted work may be examined by automated analysis tools to detect potential violations of course policy.

*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.

| Date | Day | Session | Assignments |
|------|-----|---------|-------------|
| 8/28/2013 | Wed | Course Introduction | |
| 9/2/2013 | Mon | No Class - Labor Day | |
| 9/4/2013 | Wed | Excel for Analysts | SA0: Tools Screen Shot Due (Friday)* |
| 9/9/2013 | Mon | Basic Programming and Functions | |
| 9/11/2013 | Wed | Programming and Functions (II) | SA1: Function |
| 9/16/2013 | Mon | Algorithms and Complexity | Ex1:  Excel Data Analysis |
| 9/18/2013 | Wed | Subroutines (I) | SA2: Subroutine |
| 9/23/2013 | Mon | Subroutines and Error Handling | |
| 9/25/2013 | Wed | User Interfaces | Ex2:  Functions (Due Friday) |
| 9/30/2013 | Mon | Object Orientation | SA3:  UI |
| 10/2/2013 | Wed | More Objects | |
| 10/7/2013 | Mon | XML and Web Services | |
| 10/9/2013 | Wed | Subroutine Exercise Discussion | Ex3:  Subroutines |
| 10/14/2013 | Mon | Testing | |
| 10/16/2013 | Wed | Review Session | |
| 10/21/2013 | Mon | Quiz I (in class) | Quiz I (in class) |
| 10/23/2013 | Wed | Single Table Queries | |
| 10/28/2013 | Mon | Complex Queries | SA4: Query Practice I |
| 10/30/2013 | Wed | Relational Database Intro | |
| 11/4/2013 | Mon | Relational Joins | |
| 11/6/2013 | Wed | Complex Joins/Subqueries | SA5:  Query Practice II |
| 11/11/2013 | Mon | DDL and Scripts | |
| 11/13/2013 | Wed | VBA-SQL Integration | |
| 11/18/2013 | Mon | Catch-Up Day | SA6:  Scripting/Integration |
| 11/20/2013 | Wed | Regular Expressions | |
| 11/25/2013 | Mon | Database Exercise Discussion | Ex4: Database |
| 11/27/2013 | Wed | No Class (Thanksgiving) | |
| 12/2/2013 | Mon | Big Data Tools | SA7:  TBD |
| 12/4/2013 | Wed | Course Conclusion | |
| 12/9/2013 | Mon | Review Session | |
| 12/16/2013 | Monday | Scheduled Exam | Quiz II |
| | | | |
| * - Obtaining tools/screen shot cannot be skipped; otherwise, may skip 1 SA | | | |