

Missouri University of Science & Technology

Information Science and Technology 3420

INTRODUCTION TO DATA SCIENCE AND MANAGEMENT

Spring Semester 2015 – Syllabus

Class Information

Class: Monday, Wednesday, and Friday 1:00-1:50

Class Room: Computer Science (CS) 206

Contact Information

Instructor: Prof. Michael G. Hilgers Ph.D. CQF

Email: hilgers@mst.edu

Phone: 341-6484(office)

Office: Fulton Hall 106D

Office Hours: Monday, Wednesday, & Friday (10:00-12:00)

(Or by appointment; please email me to set a time: hilgers@mst.edu)

Course Information:

"By 2018, the United States alone will face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as a shortfall of 1.5million data-savvy managers with the know how to analyze big data to make effective decisions."

Catalog Description

Introduces students to increasing business success through analysis of large-scale data collections. Topics include: import/export of data, summary statistics, cross-tabulation, data transformations (sub setting, merging, sorting and aggregation), modeling methods, and visualization. Significant programming in R is expected.

Extended Description

This course works with two intertwined ideas: "data management" and "data science". In referring to the management of data in this setting we are concerned with collecting and organizing information which is often a mixture of sources and formats.

Data science seeks to answer business questions by identifying trends, creating predictive models for forecasting, and optimizing business processes for enhanced performance. Three main categories of analytics are:

1. **Descriptive**—the use of data to find out what happened in the past;
2. **Predictive**—the use of data to find out what could happen in the future; and
3. **Prescriptive**—the use of data to prescribe the best course of action for the future

Course Prerequisites:

Programming – Two semesters of prior programming experience such as IST 1552 or CS 1510

Software:

We will program in R. As can be seen from the comparison to the left, R is at the top of the pay scale. Be careful in the interpretation of this. Companies are hiring people who understand analytics and data science.

R skills attract the highest salaries

Two recent salary surveys have shown that [R language](#) skills attract median salaries in excess of \$110,000 in the United States.

In the [2014 Dice Tech Salary Survey](#) of over 17,000 technology professionals, the highest-paid IT skill was R programming. While [big-data skills in general](#) featured [strongly](#) in the top tier, having R at the top of the list reflects the strong demand for skills to make sense of, and extract value from big data.

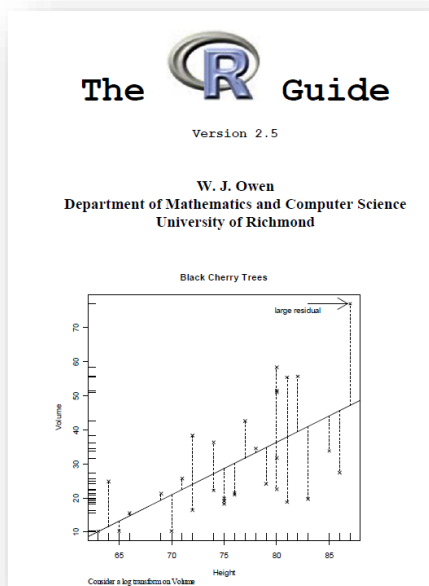
AVERAGE SALARY FOR High Paying Skills and Experience		
SKILL	2013	YR/YR CHANGE
R	\$ 115,531	n/a
NoSQL	\$ 114,796	1.6%
MapReduce	\$ 114,396	n/a
PMBok	\$ 112,382	1.3%
Cassandra	\$ 112,382	n/a
Omnigraffle	\$ 111,039	0.3%
Pig	\$ 109,561	n/a
SOA (Service Oriented Architecture)	\$ 108,997	-0.5%
Hadoop	\$ 108,669	-5.6%
Mongo DB	\$ 107,825	-0.4%



Free sources of R

In this course we will use RStudio (<http://www.rstudio.com>). I have used it for several semesters and the students like it best among various choices. It is free and can be downloaded for your personal machine and is available in the computer lab.

Textbooks:



There are numerous free online resources. These include textbooks, research papers, and help manuals. We will take advantage of these.

A very valuable source of information and documentation for R is

<https://cran.r-project.org/>

R Reference Card 2.0

Public domain, v2.0 2012-12-24.
V 2 by Matt Baggott, matt@baggott.net
V 1 by Tom Short, tshort@ieee.org
Material from *R for Beginners* by permission of Emmanuel Paradis.

Getting help and info

`help(topic)` documentation on topic
`?topic` same as above; special chars need quotes: for example `"&&"`
`help.search("topic")` search the help system; same as `??topic`
`apropos("topic")` the names of all objects in the search list matching the regular expression "topic"

Operators

`<-` Left assignment, binary
`>-` Right assignment, binary
`=` Left assignment, but not recommended
`<<-` Left assignment in outer lexical scope; n for beginners
`$` List subset, binary
`-` Minus, can be unary or binary
`+` Plus, can be unary or binary
`~` Tilde, used for model formulae
`:` Sequence, binary (in model formulae: interaction)
`::` Refer to function in a package, i.e. `pkg::function`; usually not needed
`*` Multiplication, binary
`/` Division, binary
`^` Exponentiation, binary

Course Elements

Instructional Methods:

To achieve an understanding of the material, several techniques and methods will be used:

- I will lecture over foundational material. Typically, some form of notes will be provided, though I am not a strong supporter of PowerPoint in the classroom.
- Various sources of reading matter will be given including textbook, web, and personally developed notes. Please do the reading
- I like to work examples. Expect to spend some time studying data with software tools such as R.
- Analysis is not a spectator sport. Expect a homework problem or two most every night.

Assignments:

Homework/programming assignments will be made frequently. Please observe the following:

- Work is to be done individually unless otherwise specified. If you submit the work of another person as your own, you will receive a zero for the assignment and your name given to the Vice Provost of Undergraduate/Graduate studies.
- Homework is to be completed on the specified date and time.
- Late homework will NOT be accepted and will result in a zero for that assignment unless prior arrangement is made.

Course Content

“Big Data is an emerging phenomenon. Computing systems today are generating 15 petabytes of new information every day—eight times more than the combined information in all the libraries in the U.S.; about 80% of the data generated every day is textual and unstructured data.”

The following should give you an idea of how we will proceed in the course. It is not ridged in that I will slow down rather than lose everyone or speedup if I am boring everyone.

R and Data

We will learn basic data operations using numeric and character data then extending the idea to vectors and a critical structure known as a data frame.

Data Science

We will move step-by-step through a series of data examples of increasing complexity to learn the basic types of analysis and interpretation of data.

Foundations of Modeling

The next important step is to understand mathematical modeling. We will examine how to select a model, estimate the model accuracy, tradeoffs between accuracy and form and how we can predict outcomes.

Linear Models

Linear models are foundational. We will consider topics like sum of squares error, R^2 , P values, correlation, covariance, anova tables, predicting outcomes, prescribing actions (linear optimization)

Classification

If time permits, we can peek at logistic modeling, estimating coefficients, lift curves, multivalent, prediction

Course Activity:

Given what is described about, the breakdown of activity is as follows:

Activity	Course Percentage
Programming Exercises	20%
Tests	80%

This is very general. This course is being created. I might need to change things once I see how the class is responding. I might make a "test" a larger "program". And "programming exercises" might be short problem sets worked during class. Others might be more substantial.

Grading Breakdown:

Grades will be based on total points, as defined below. There may be bonus points from time to time, which would be added to whatever category the bonus applies to. Boundaries for grades may be adjusted downward slightly, if deemed needed.

Grades:

- A: 100% - 90%
- B: 89% - 80%
- C: 79% - 70%
- D: 69% - 60%
- F: Below 59%

Learning Objectives

	Communication Skills	Critical Thinking	Information Technology	Teamwork and Leadership
Develop basic skills needed to understand and manipulate the mathematical models forming the foundations of data analytics		X		
Learn how to use R to visualize large multidimensional data sets and explain	X	X	X	
Identify proper mathematical model for a given data set				
Perform Linear regression on a multi-dimension data set		X	X	
Be able to use nonparametric techniques		X	X	
Be able to use logistic models to analyze data sets		X	X	
Be able to use classification techniques to see patterns in data		X	X	
Be able to use clustering techniques to see patterns in data		X	X	
Be able to explain the role of business analytics in corporate environments	X			

COURSE POLICIES

Attendance:

Attendance is required, particularly as the assignments will be based on the important definitions and concepts presented in the lectures. You will likely want to ask questions. The class moves quickly and it is easy to fall behind and not get caught up. The more you miss class, the more material that will be foreign to you. If a student has missed an extended or excessive amount of classes or has failed to turn in multiple assignments, the instructor will send that student an Academic Alert. The alert will be emailed to the student and student's advisor. The student must meet with the instructor within three days or the instructor will send out another alert. If the student has not met with the instructor after the second alert, the instructor reserves the right to drop the student. If emergency circumstances arise, please contact the instructor soon to avoid penalties, and to try to catch up to the rest of the class.

Academic Integrity Statement

(<http://registrar.mst.edu/academicregs/>):

Violations of the University's academic code include, but are not limited to, possession of or use of unauthorized materials during quizzes or tests; providing unauthorized information to another student; or copying the work of another person. Violations may result in academic penalties in addition to receiving an "F" on the assignment in question. (See page 30 of S&T's "Student Academic Regulations" handbook for further details about student standards of conduct relative to the system's Collected Rules and Regulations section 200.010.)

The most common attempt at dishonesty is submitting the program of another person with only some changes to deceive me. These are easy to recognize and not be tolerated.

Academic Alert System

(<http://academicalert.mst.edu/>):

S&T is committed to the success of its students by providing an environment conducive to teaching and learning. To ensure that every student takes full advantage of the educational opportunities and support programs on campus, the University has implemented an Academic Alert System, a web-based application. The purpose of the System is to improve the overall academic success of students by:

- Improving communication between students, instructors, and advisors;
- Reducing the time required for students to be informed of their academic status;
- Informing students of actions they need to perform in order to meet the academic requirements in the courses they are taking.

To assist you, I will initiate an academic alert for students who are not meeting academic course requirements through poor performance on assignments or poor attendance. When an alert is initiated, an email is immediately sent to the instructor, student, and advisor. You are encouraged to respond quickly to all academic alerts. If you fail to open the alert within one week, email notification is sent to your advisor.

Title IX

(<http://titleix.mst.edu>)

Missouri University of Science and Technology is committed to the safety and well-being of all members of its community. US Federal Law Title IX states that no member of the university community shall, on the basis of sex, be excluded from participation in, or be denied benefits of, or be subjected to discrimination under any education program or activity. Furthermore, in accordance with Title IX guidelines from the US Office of Civil Rights, Missouri S&T requires that all faculty and staff members report, to the Missouri S&T Title IX Coordinator, any notice of sexual harassment, abuse, and/or violence (including personal relational abuse, relational/domestic violence, and stalking) disclosed through communication including but not limited to direct conversation, email, social media, classroom papers and homework exercises.

Missouri S&T's Title IX Coordinator is Vice Chancellor Shenethia Manuel. Contact her directly (manuels@mst.edu; (573) 341-4920; 113 Centennial Hall) to report Title IX violations. To learn more about Title IX resources and reporting options (confidential and non-confidential) available to Missouri S&T students, staff, and faculty, please visit <http://titleix.mst.edu>.

Disability Support Services

(<http://counsel.mst.edu/>):

If you have a documented disability and anticipate needing accommodations in this course, you are strongly encouraged to meet with me early in the semester. You will need to request that the Disability Services staff send a letter to me verifying your disability and specifying the accommodation you will need before I can arrange your accommodation. If you have a disability that might require academic accommodations, please visit Disability Support Services in 204 Norwood Hall (341-4211; dss@mst.edu) very early in the semester.

Classroom Egress Maps

(<http://registrar.mst.edu/links/egress/>):

Please familiarize yourself with the classroom egress maps posted on line so you will know where emergency exits are located.

LEAD Learning Assistance

(<http://lead.mst.edu/>):

The Learning Enhancement across Disciplines Program (LEAD) sponsors free learning assistance in a wide range of courses for students who wish to increase their understanding, improve their skills, and validate their mastery of concepts and content in order to achieve their full potential. LEAD assistance starts no later than the third week of classes. Check out the online schedule at <http://lead.mst.edu/assist>, using zoom buttons to enlarge the view. Look to see what courses you are taking have collaborative LEAD learning centers (bottom half of schedule) and/or Individualized LEAD tutoring (top half of the schedule). For more information, contact the LEAD office at 341-7276 or email lead@mst.edu.

Cell Phones

Students are asked to set phones and pagers to vibrate or silent while in the laboratory. Cell phone use is not permitted. This includes texting. Please remember that voices carry and that speakers using cell phones tend to speak more loudly than normal conversational tones. In general, talking and texting are disrespectful.

With that said, I have had students ask to take a picture of something I had written in class. Provided you ask my permission, I am comfortable with this. Do not take pictures of other student's work.