

UNIVERSITY OF WISCONSIN-MILWAUKEE
School of Information Studies

INFOST 582 – Introduction to Data Science
Section 201 and 202 - Online
Spring 2019

SYLLABUS

Instructor: Margaret Kipp
E-mail: kipp@uwm.edu (best contact method)
Office: NWQB 2574
Office Hours: TBD
SOIS Office: 414-229-4707
SOIS Fax: 414-229-6699

CATALOG DESCRIPTION

Introduces basic concepts, background, theoretical, practical and technological aspects of data science. 3 credits

GENERAL DESCRIPTION

This course provides an introduction to data science. Data science has developed as a set of methods for analysing massive data sets to extract useful knowledge. A data scientist is a person who has the skills and knowledge to perform these analyses. This course will cover topics necessary to develop data-science solutions to problems including data collection, data cleaning and integration, data analysis, and data presentation.

PREREQUISITES

- Junior Standing. For 500 and 600 level courses it is recommended that an undergraduate student first consult with the appropriate instructor and/or advisor concerning the applicability of this specific course.
- Basic computer facility and technology literacy as listed in the SOIS policy are required: <http://uwm.edu/informationstudies/academics/graduate/mlis/?target=curriculum/#computer-literacy>
- Optional: Basic statistical knowledge (measures of central tendency) would be an asset for later in the course, but is not required. Basic Excel experience would also be an asset.

OBJECTIVES/OUTCOMES

Upon completion of the course, students will be able to:

1. effectively develop researchable questions; (**Paper or project**)
2. identify data sources, collect, clean and merge data; (**Selecting a data set, Cleaning data**)
3. manipulate structured or unstructured data sources; (**Querying a Database, Creating Metadata**)
4. identify and apply appropriate statistical methods for analysing data; (**Analysing Data, Project**)
5. critically evaluate tools for working with data; (**Project, Cleaning Data**)

6. address multilingual and multicultural issues in data creation and analysis; **(Creating Metadata, Querying a Database, Readings and Discussions)**
7. identify emerging trends and stay current with issues in data science. **(Readings and Discussions)**

ALA COMPETENCIES (for MLIS students)

1. The systems of cataloguing, metadata, indexing, and classification standards and methods used to organize recorded knowledge and information.
2. Information, communication, assistive, and related technologies as they affect the resources, service delivery, and uses of libraries and other information agencies.
3. The application of information, communication, assistive, and related technology and tools consistent with professional ethics and prevailing service norms and applications.
4. The principles and techniques necessary to identify and analyse emerging technologies and innovations in order to recognize and implement relevant technological improvements.

METHOD

Lecture/Discussion/Readings/Examples/Exercises – to achieve a satisfactory understanding of the course material and to fulfil requirements of the assignments, students are expected to attend the lectures, read and comment on the readings, participate in discussions and in-class exercises, and explore examples and tutorials.

TIME COMMITMENT

This course requires a weekly time commitment. General university guidelines indicate that a 3 credit course requires a minimum 144 hour time commitment over the course of a term. This time commitment represents a minimum of 9-10 hours of work per week per course. For an on-site class 3 of these hours represent on-site instruction in a classroom; in an online class this time would be spent on independent reading, discussions and in-class exercises.

Each week you may be required to read notes and readings from the reading list associated with that class, participate in discussions, write summaries of readings, complete in-class exercises, explore examples, or complete assignments and projects. It is your responsibility to plan your time in order to complete all activities based on the schedule outlined in this syllabus.

ACCOMMODATIONS

If you need accommodations due to illness, disabilities, scheduling conflicts with religious observances, or other life events (e.g. military service) contact the instructor as soon as possible, preferably by the third week of class as per university policy. Official documentation may be required depending on the nature of the considerations requested per university policy (http://www4.uwm.edu/secu/docs/faculty/1895R5_Uniform_abus_Policy.pdf).

TEXTBOOK AND READINGS

Shron, Max. 2014. *Thinking with Data: How to Turn Information into Insight*. O'Reilly Media. ISBN: 978-1449362935 (Available in Paperback, Kindle, EPUB, etc.) [Required]

Watt. 2014. *Database Design 2nd Edition*. BCcampus <https://opentextbc.ca/dbdesign01/> [Open Textbook]

Lane. n.d. *Online Statistics Education: An Interactive Multimedia Course of Study*.

<http://onlinestatbook.com/2/index.html> [Open Textbook]

Foreman, John W. 2015. *Data Smart*. Wiley. ISBN: 978-1118661468 (UWM Library Full Text, Also available in Paperback, Kindle, EPUB, etc.) [Optional, but will be used in 691: Data Analysis for Data Science]

Readings are listed in the course outline for each class. Readings should be completed before the class. Other course materials, including this syllabus, are available through D2L (<http://d2l.uwm.edu/>).

COURSE OUTLINE

Class	Date	Topics	Readings (complete before class)
1	Jan 25	Introduction to Data Science and Big Data	<ul style="list-style-type: none"> Shron. 2014. <i>Thinking with Data</i>. Preface, Chapter 1 (18p); Loukides. 2010. What is Data Science? (12p) (D2L); Zhu & Xiong. 2015. Defining Data Science. (8p) http://arxiv.org/abs/1501.05039 [cs.DB]; Miller. 2013. Data Science: The Numbers of Our Lives, APRIL 11, 2013, New York Times (4p) http://nyti.ms/10QarGu; Chandrasekaran. 8 July 2013. Becoming a Data Scientist. http://nirvacana.com/thoughts/becoming-a-data-scientist/;
2	Feb 1	Developing Data Based Questions	<ul style="list-style-type: none"> Shron. 2014. <i>Thinking with Data</i>. Chapters 2-4 (50p);
3	Feb 8	Choosing Datasets and Collecting Data	<ul style="list-style-type: none"> Shron. 2014. <i>Thinking with Data</i>. Chapters 5-6 (25p); Mattmann. 2013. Computing: A vision for data science. <i>Nature</i> 493, p.473–475. (UWM Library Full Text); Marx. 2013. Biology: The big challenges of big data. <i>Nature</i> 498, p.255–260. (UWM Library Full Text); Wallis, et al. 2013. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. <i>PLOS One</i>. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067332;
4	Feb 15	Privacy, Ethics and Data	<ul style="list-style-type: none"> O'Leary. 2015. Big Data and Privacy: Emerging Issues. <i>Intelligent Systems, IEEE</i> 30(6): 92-96. (D2L); Perera, et al. 2015. Big Data Privacy in the Internet of Things Era. <i>IT Professional</i> 17(3): 32-39 (D2L); Daries, et al. 2014. Privacy, Anonymity, and Big Data in the Social Sciences. <i>Communications Of The ACM</i> 57(9): 56-63. (D2L); Shilton. 2012. Participatory personal data: An emerging research challenge for the information sciences. <i>Journal of the American Society for Information Science and Technology</i> 63(10): 1905-1915. (UWM Library Full Text); Kugler. 2016. What Happens When Big Data Blunders? <i>Communications Of The ACM</i> 59(6): 15-16. (D2L);
5	Feb	Metadata and	<ul style="list-style-type: none"> Gilliland. 2016. "Setting the Stage" In <i>Introduction to</i>

	22	the Semantic Web	<p><i>Metadata Version 3.</i> http://www.getty.edu/publications/intrometadata/setting-the-stage/;</p> <ul style="list-style-type: none"> • Guha, Brickley and Macbeth. 2016. Schema.org: Evolution of Structured Data on the Web. <i>Communications of the ACM</i> 59(2): 44-51. (D2L); • Nykyrl and Niininen. 2015. The future of metadata: Open, linked and multilingual. <i>Scandinavian Library Quarterly</i> 48(1-2). (UWM Library Full Text); • Bernstein et al. 2016. A New Look at the Semantic Web. <i>Communications of the ACM</i> 59(9): 35-37. (D2L); • Introducing Linked Data and the Semantic Web. LinkedDataTools.com. http://www.linkeddatatools.com/semantic-web-basics/;
6	Mar 1	Databases and Data Stores	<ul style="list-style-type: none"> • Watt. 2014. <i>Database Design 2nd Edition</i>. Chapters 1-4, 7-8. BCcampus https://opentextbc.ca/dbdesign01/; • Pokorny. 2015. Database technologies in the world of big data. In <i>Proceedings of the 16th International Conference on Computer Systems and Technologies 15</i>, Rachev and Smrikarov (Eds.). ACM, New York, NY, USA, 1-12. (D2L); • McKenzie. 2010. Falsehoods Programmers Believe About Names. http://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/; • MSDN. 2016. Locale and Culture Awareness (section). https://docs.microsoft.com/en-gb/globalization/locale/locale-and-culture;
7	Mar 8	Working with Structured and Semi-structured Data	<ul style="list-style-type: none"> • Watt. 2014. <i>Database Design 2nd Edition</i>. Chapters 15-16 (SQL). BCcampus https://opentextbc.ca/dbdesign01/; • Florescu. 2005. Managing Semi-Structured Data. <i>ACM Queue</i> 2005 18-24. (D2L); • Jackson et al. 2008. Dublin Core Metadata Harvested Through OAI-PMH. <i>Journal of Library Metadata</i> 8:1 p. 5-21. http://hdl.handle.net/2142/9091;
8	Mar 15	Unstructured Data	<ul style="list-style-type: none"> • Polnaszek, et al. 2016. Overcoming the Challenges of Unstructured Data in Multisite, Electronic Medical Record-based Abstraction. <i>Medical Care</i> 54(10): 65-72. (UWM Library Full Text); • Shim, Koh, Fister and Seo. 2016. Phonetic Analytics Technology and Big Data: Real-World Cases. <i>Communications of the ACM</i> 59(2): 84-90. (D2L); • Veeranjaneyulu, et al. 2014. Approaches for Managing and Analyzing Unstructured Data. <i>International Journal on Computer Science and Engineering (IJCSSE)</i> 6(1). http://www.enggjournals.com/ijcse/doc/IJCSE14-06-01-020.pdf; • Log File Analysis: The Ultimate Guide • http://builtvisible.com/log-file-analysis/;
9	Mar	Spring Break -	<ul style="list-style-type: none"> • No Readings

	22	No Class	
10	Mar 29	Cleaning and Merging Data	<ul style="list-style-type: none"> • Rahm. (n.d.) Data Cleaning: Problems and Current Approaches, University of Leipzig, Germany. http://lips.informatik.uni-leipzig.de/files/2000-45.pdf; • Top Ten Ways to Clean Your Data. Microsoft.com. https://support.office.com/en-us/article/Top-ten-ways-to-clean-your-data-2844b620-677c-47a7-ac3e-c2e157d1db19; • Using a spreadsheet to clean up a dataset. 2013. • http://schoolofdata.org/handbook/recipes/cleaning-data-with-spreadsheets/; • Loshin. 2015. Integrating Data from Multiple Sources • http://community.embarcadero.com/index.php/blogs/entry/integrating-data-from-multiple-sources-by-david-loshin; • Data Journalism Handbook http://datajournalismhandbook.org/1.0/en/understanding_data_2.html;
11	April 5	Describing Data	<ul style="list-style-type: none"> • Lane. (n.d.). <i>Online Statistics Education: An Interactive Multimedia Course of Study</i>. Chapters 1-3, 17. http://onlinestatbook.com/2/index.html;
12	April 12	Analysing Data	<ul style="list-style-type: none"> • Lane. (n.d.). <i>Online Statistics Education: An Interactive Multimedia Course of Study</i>. Chapters 4, 6-7, 11-12, 14. http://onlinestatbook.com/2/index.html;
13	April 19	Machine Learning	<ul style="list-style-type: none"> • Alpaydin. 2014. <i>Introduction to Machine Learning Third Edition</i>. MIT Press. Chapter 1 (UWM Full Text); • Lane. (n.d.). <i>Online Statistics Education: An Interactive Multimedia Course of Study</i>. Chapter 5 (probability). http://onlinestatbook.com/2/index.html; • Optional: Foreman. 2015. <i>Data Smart</i>. Wiley. Chapters 1-2 (UWM Library Full Text);
14	April 26	Analysing Textual Data	<ul style="list-style-type: none"> • A Companion to Digital Humanities, ed. Schreibman, et al. Oxford: Blackwell, 2004. Part 1 http://www.digitalhumanities.org/companion/; • Reagan et al. 2016. The emotional arcs of stories are dominated by six basic shapes. arXiv.org. (10p + supplementary data) https://arxiv.org/abs/1606.07772; • Optional: Foreman. 2015. <i>Data Smart</i>. Wiley. Chapter 3 (UWM Library Full Text);
15	May 3	Visualizing Data	<ul style="list-style-type: none"> • Verdinelli and Scagnoli. 2013. Data Display in Qualitative Research. <i>International Journal of Qualitative Methods</i> http://journals.sagepub.com/doi/pdf/10.1177/160940691301200117; • NIST/SEMATECH e-Handbook of Statistical Methods, Chapter 1: Exploratory Data Analysis http://www.itl.nist.gov/div898/handbook/; • Kandel, et al. Research directions in data wrangling: Visualizations and transformations for usable and credible data. <i>Information Visualization</i> 0(0) 1–18. http://vis.stanford.edu/files/2011-DataWrangling-IVJ.pdf;

			<ul style="list-style-type: none"> Google Fusion Tables https://support.google.com/fusiontables/answer/2571232?hl=en;
16	May 10	Data Science as a Career	<ul style="list-style-type: none"> Invasion of the data scientists: Hot job of 2016 expands beyond tech http://www.denverpost.com/business/ci_29451303/invasion-data-scientists-hot-job-2016-expands-beyond; Data Scientist: The Sexiest Job No One Has http://www.informationweek.com/big-data/big-data-analytics/data-scientist-the-sexiest-job-no-one-has/d/d-id/1112832; Interviews with Data Scientists (one interview) http://www.datascienceweekly.org/pdf/DataScienceWeekly-DataScientistInterviews-Vol1-April2014.pdf (long);

ASSIGNMENTS

Assignment	Graduate	Undergrad [1]	Associated Classes	Deadline
Selecting a Dataset [2] Identify a question that interests you. Identify a data set on this topic that could be used to answer the question. Explain the kinds of information available in the dataset and how the data is structured. (400 words)	5	5	1-3	Class 4 [3]
Metadata Select 2 objects and create a metadata record for each using a metadata schema of your choice. Your records should contain enough information to thoroughly describe the object. Use Dublin Core or Schema.org encoded in XML, RDF or JSON.	5	5	5	Class 6
Database Create a simple database in Access or MySQL with three or more joined tables. Populate the tables with enough data to provide useful results for your queries. Create two SQL queries that extract useful data from the database.	5	5	6-7	Class 8
Short Paper Write a short paper on a data science related topic. (800-1000 words)	20	n/a	All	Proposal: Class 3 Paper: Class 10
Cleaning Data Identify problems in a data set which might interfere with analysis of the data (e.g. typos, structure problems, poor adherence to standards). Describe the	5	5	7-10	Class 11

problems and write policy statements or suggested solutions for solving these problems. (400 words)				
Analysing Data Develop a basic analysis of a data set using the tools discussed in class. Describe your findings in report format. Use graphs, charts or other tools to present your findings as required. (500U/1000G words) Note: It is recommended that you submit a rough draft/early draft of your project, though you may choose to do a separate data analysis.	10	10	1-13	Class 13
Project Select a topic and gather a small test set of data, clean it, analyse it and present the results in a report. Your report may be written or multimedia based. (800U/1500G words or equivalent, charts, tables, etc. do not count towards the word limit) Note: Graduate students are expected to provide additional critical analysis and reflection of the data including potentially locating and citing appropriate supporting materials from published sources.	30	50	All	Proposal: Class 9 Project: Last class
Participation (see below)	20	20	All	Last class
Total	100	100		

[1] Different requirements for graduate and undergraduate levels will be specified in the directions for each assignment where appropriate.

[2] Full assignment descriptions are in D2L.

[3] Class numbers are listed in the Course Outline Table. Each class has an associated Class Number (#), Date, Topic, Readings and may have In-class Exercises, Discussions or Tutorials. The assignment table is keyed to the course outline's class numbers. To determine the exact date an assignment is due, go to the appropriate class number in the course outline table or use the D2L calendar.

Participation

Students are expected to participate in discussion and in-class exercises as a demonstration of their ability to articulate key concepts. Participation is mandatory and constitutes almost one quarter of the points available for this class. Participation will consist of all of the following: individual summaries of readings, participation in discussions, contributed articles, and responses to others.

Participation will consist of all of the following:

- **Completion of the Syllabus Quiz**

- The syllabus quiz must be completed in the first 2 weeks of class. Points will automatically be entered in D2L.
- **Individual Summaries of Readings**
 - Post 3 summaries of the weekly readings to the appropriate weekly discussion group based on the class associated with each reading.
 - You must post 3 summaries in total, but you may choose the classes for which you wish to contribute the summaries.
 - Sign up for 3 sets of readings on the signup sheet posted in the news section of D2L.
 - Responses need not exceed 300 words.
 - Summaries posted before the date of the class earn a half bonus point each. Be sure to mark this on your course completion checklist to ensure you receive the bonus.
- **Contributed Article**
 - Contribution of a new article, video, cartoon, etc. relevant to the class and a short summary (approximately 100 words) explaining its relevance to class. This should be posted to the appropriate weekly discussion group based on the topic. You may choose which week you wish to contribute this item.
 - A signup sheet will be posted in the news section of D2L.
- **Individual Summaries of In-Class Exercises**
 - Participation in the in-class exercises included in most weeks. Post individual summaries to the appropriate weekly discussion group.
 - You must post 6 summaries of in-class exercises in total, but you may choose the classes for which you wish to contribute the summaries.
 - Responses need not exceed 300 words.
- **Participation in Weekly Discussions**
 - Participation in weekly discussions including responding to discussion prompts and reading and/or responding to weekly reading summaries and other information posted to the weekly discussion groups by classmates. Points will be allocated based on your reading level (i.e. many, few, nothing read) and/or your responses to others (i.e. many, few, no responses).
 - Generally frequent participation requires that you participate at least once a week in most weeks.
- **Submission of the Course Checklist to the participation dropbox**
 - The completed checklist with all required course elements listed submitted to the dropbox before the last class. You should complete as much as possible of the checklist. Use the checklist throughout the term to ensure you are on track to complete all course requirements.

Working with Classmates

All assignments except the short paper and participation may be completed in pairs or trios. Assignments completed in pairs/trios must identify all work partners by full name at the top of the assignment. You must each submit the same assignment to the dropbox. If you simply assisted each other but did not do the whole assignment together, you must also note this at the top of the assignment. Unacknowledged borrowing is seen as academic misconduct, so be sure to document your teamwork to avoid this.

Formatting Guidelines for Assignments

Assignments should be written using Arial or another Sans-Serif style font. Do not use red text or highlights for emphasis or to highlight your answers to questions. Remove all extraneous information before submission (e.g. assignment instructions or tips).

Use whatever citation format you prefer. If you are not using a common format such as MLA or APA you should include information about which style guide you are using in the assignment.

Paper submissions will not be accepted. All assignments must be typed on a computer and submitted electronically. Handwritten submissions will not be accepted, even if scanned and submitted electronically.

Assignments may not be submitted in Pages, Microsoft Works, or Microsoft Project as I cannot open these formats. You should save these as a PDF instead. Other common file formats should be acceptable including Open/Libre Office formats. If you are using an unusual format you can always check with me first before submission to ensure I can open it.

Due Dates and Assignment Submission

All assignments and projects should be submitted through D2L to the appropriate dropbox before midnight (Central Time) on the due date. Points for late assignments will be reduced 10% per day late after the due date. The dropbox will remain open for the submission of late assignments until the late penalty reaches 100%.

Participation items, including in-class exercises, should be submitted to the appropriate discussion group (see the participation section below) before the discussion group closes. Discussion groups will be open for 1 week before and 1 week after the date of the associated class.

Emailed submissions will only be accepted as a backup to a D2L submission (or in case of D2L errors).

Everything must be submitted by the Last Class (this includes all assignments, papers, projects, and participation). All project and assignment deadlines are in the syllabus. For discussion deadlines check the discussion groups or the D2L calendar. The D2L calendar also contains all project and assignment deadlines. It is your responsibility to keep track of deadlines using the tools provided or by creating your own list of deadlines.

Items submitted early will not be evaluated until their due date. Students are encouraged to complete all Associated Classes listed under Assignments before submitting the assignments since the material in these classes constitutes preparation for the assignments. Submission well before the due date is not encouraged.

Extensions

Students must contact the instructor before each due date for any extensions. Extension requests made prior to the due date do not require any documentation or explanation as long as they are not longer than a week. Simply provide a date/time by which you will submit the assignment. After the deadline the penalties listed under Due Dates will be enforced. Material

submitted late after an extension will also be subject to these penalties. Plan your time accordingly.

Technical Issues

You are responsible for accessing tools used in this class in a timely manner in order to complete in-class exercises and assignments. This course assumes you have the required basic computer facility and technology literacy skills as listed in the SOIS policy. Technical issues do not absolve you from the requirement to complete material. If you are having technical issues, you should switch to Firefox as a first step. You may also find the tools do not work from your work place, in which case you should try them from home or on the school machines. I will attempt to provide technical assistance with common problems, but you can also contact soistech@uwm.edu for assistance. If you are a Safari user you should be aware that Safari is the least capable browser for technical work and you will encounter difficulties with the tools in this class.

Extra Credit or Other Special Considerations

Per university policies (see <http://www4.uwm.edu/secu/docs/other/S29%2Ehtm>) extra credit assignments and other special consideration are not possible. Students should make use of the extensions policy outlined above or provide appropriate documentation of special circumstances as outlined elsewhere in the syllabus.

Code of Conduct/Expectations for this Class

This is a professional programme and professional, courteous behaviour is expected of all participants. It is expected that class members will show consideration for all other members of the class and contribute in a constructive manner which is conducive to a good learning environment. Class members should consider the relevance and appropriateness of their contributions to the class before contributing to the class. Violations of these expectations will result in reduced participation points or other sanctions depending on severity.

Plagiarism and Referencing

Plagiarism is the unacknowledged borrowing of ideas or material from someone else's work. It is considered an academic offence and can be considered grounds for failure in a course or expulsion from the programme. Cite all references and provide credit for all other materials. This applies to all material including images, sounds or videos. A citation (in the format of your choice) with a functioning URL (if relevant) is the minimum required for a reference. (<http://guides.library.uwm.edu/content.php?pid=235714&sid=1949820#6509804>)

You may not resubmit assignments already submitted in other courses or in a previous instance of this course, nor may you submit other people's work as your own. Plagiarism will be dealt with on a case by case basis but will result in a lowered mark on the assignment, failure on the assignment or failure in the course depending on severity and the number of plagiarized items submitted. Points lost through plagiarism may not be replaced by bonus points on other assignments.

It is expected that you will consult and cite the research and professional literature where merited and not rely solely on encyclopaedias, newspapers or unpublished, online sources. Papers where the majority of sources are blogs and Wikipedia (or similar sites) will not be accepted.

Use a common style manual for citations (e.g. APA, MLA, Chicago). Ideally you would choose a citation style guide you have used before, or one you are using in another class.

GRADING SCALE

96-100	A	Superior work		74-76	C	Work is below standard
91-95	A-			70-73	C-	
87-90	B+			67-69	D+	
84-86	B	Satisfactory, but undistinguished work		64-66	D	Unsatisfactory work
80-83	B-			60-63	D-	
77-79	C+			Below 60	F	

GRADE REQUIREMENT FOR A CORE COURSE

If you are pursuing an MSIST degree or an MLIS IO or IT Concentration, you need to earn at least a B (does not include B-) in this course.

UWM AND SOIS ACADEMIC POLICIES

The following link will take you to UWM pages/links which contain university policies affecting all UWM students. <http://uwm.edu/secu/wp-content/uploads/sites/122/2016/12/Syllabus-Links.pdf>

The following link will take you to pages/links which contain SOIS policies affecting all SOIS students. <http://www4.uwm.edu/sois/resources/formpol/policies.cfm>

Undergraduates may also find the *Panther Planner and Undergraduate Student Handbook* useful. <http://uwm.edu/studenthandbook/>

For graduate students, there are additional guidelines from the Graduate School.

<http://uwm.edu/graduateschool/>



This document is licensed under a Creative Commons Non-commercial Alike 3.0 United States Licence except where other rights exist.