Data Analytics

Yong Zheng

Illinois Institute of Technology
Chicago, IL, 60616, USA

School of Applied Technology
ILLINOIS INSTITUTE OF TECHNOLOGY

Intro

- Who am I
- Data Management and Data Analytics
- Topics in This Class
- Syllabus, Blackboard system and Policy

Intro

- Who am I
- Data Management and Data Analytics
- Topics in This Class
- Syllabus, Blackboard system and Policy

Who am I

Yong Zheng, PhD
2018 – Present, Assistant Professor at IIT, USA
2016 – 2018, Senior Lecturer at IIT, USA
2017 – Present, Consultant at NPAW, Barcelona, Spain
2016 – 2017, Adjunct Lecturer at DePaul University
2016, PhD in CIS, DePaul University, Chicago, USA
2015, Data Scientist at Pandora, Inc., Oakland, USA
Research: Data Science for Web Intelligence
Particular: Recommender Systems (RecSys)
Website: http://yongzheng.me/

Research Projects

Research Projects

- Recommender Systems and Personalizations
- Data Science for Technology-Enhanced Learning
- Data Science in CyberSecurity/MobileSecurity

Development Projects

- Java-Based Web information system for educations

Research Assistants

Qualifications

- Must: Good in math and data analytics
- Must: Good in programming for algorithm implementations
- Must complete ITMD 527 and ITMD 525 classes; or, you already have practical knowledge and skills in data mining and machine learning

How to be RA

- We do not have funded RA positions in ITM department
- You can take ITMD/ITMT 597 – Independent Study, and work with me on research projects (no stipend, 1-3 credits)
- Or, you can work when you are available anytime

Self-Introduction

1st-year or 2nd year student?

Any experience in statistics or data analytics?

Can you program in Java?

How about R or Python?

Career goals? Software engineer, DBA, Data analyst?

What are you expected to learn?

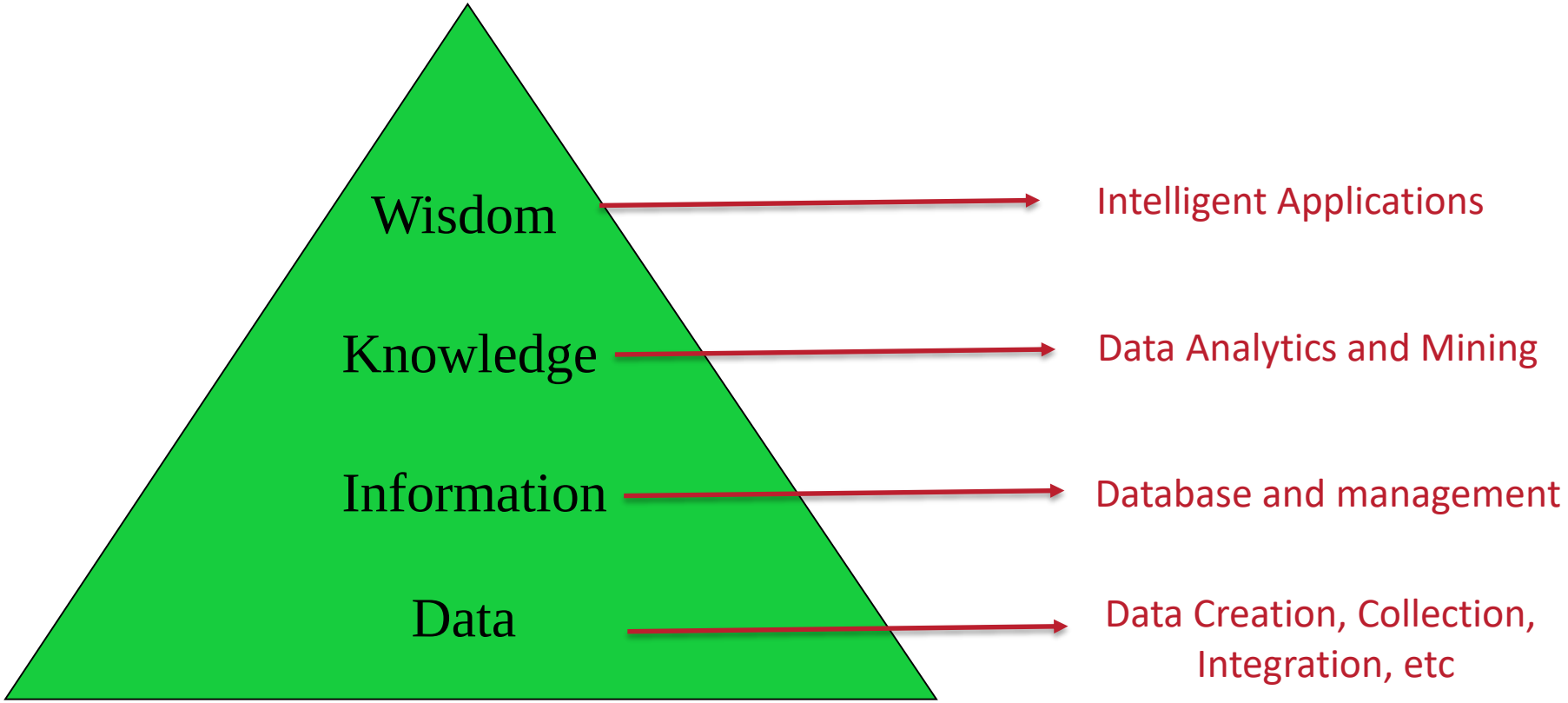
Self-Introduction

- You do not need to have background in statistics
- Most of the students do not have experience in data analytics or statistics. Note: data warehousing is not counted as data analytics
- We will teach you from the very beginning

Intro

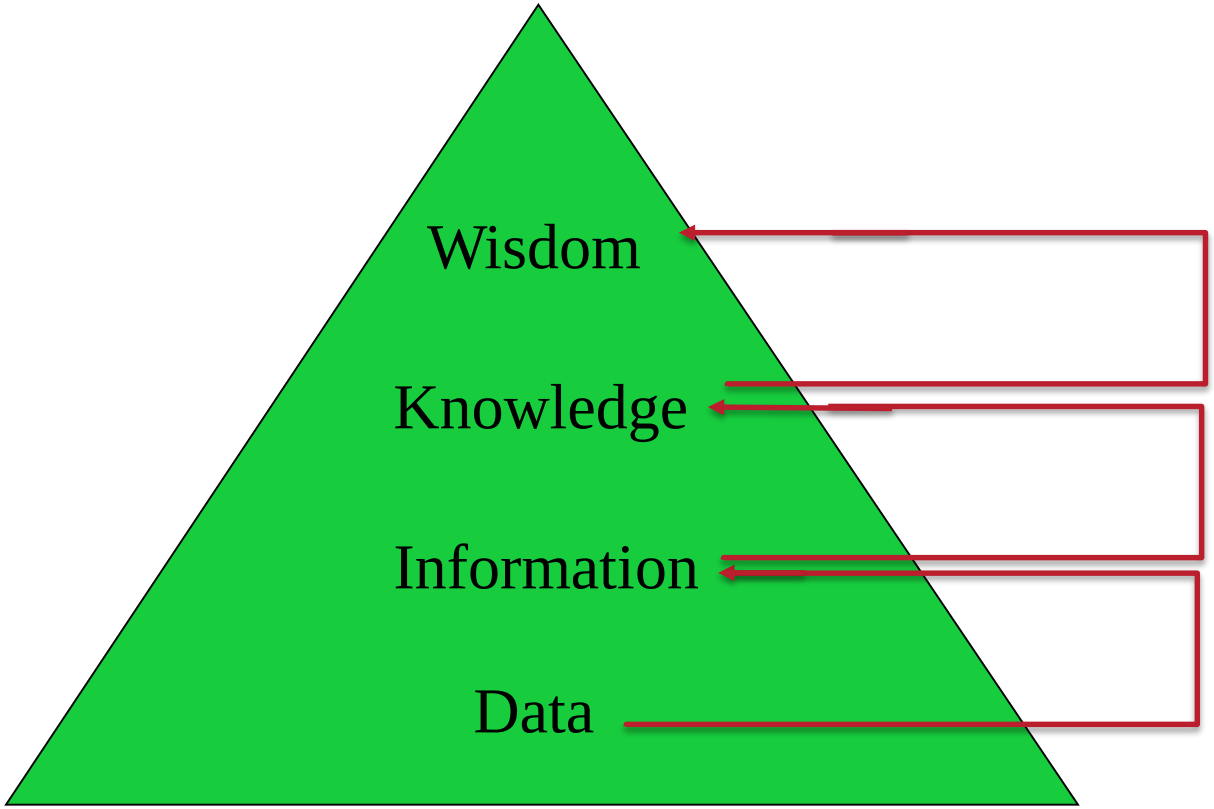
- Who am I
- Data Management and Data Analytics
- Topics in This Class
- Syllabus, Blackboard system and Policy

Intro: Data Management



The Information Hierarchy

Intro: Data Management



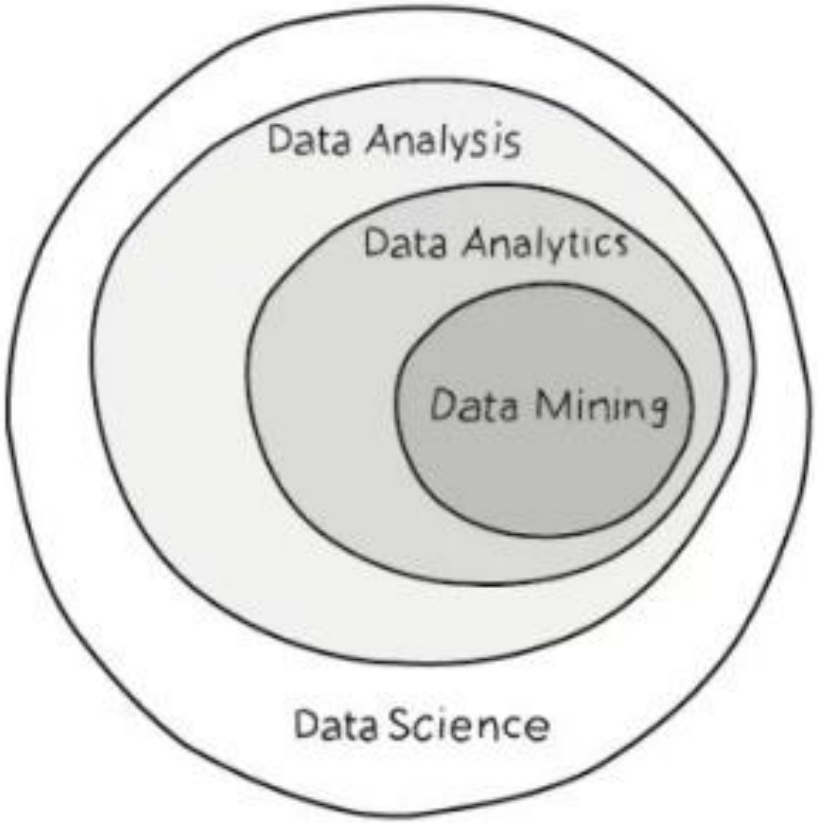
The Information Hierarchy

Business Intelligence

Data Analytics and Mining
ITMD 525, 526, 527, 529

Data Preprocessing & Storage
ITMD 525, 523

What is Data Analytics



What is Data Analytics

- Generally, Data Analytics is a general term. It is involved with broader scope of topics than data mining.
- Data mining is related to explorations, knowledge discovery and predictive models.
- Practically, data analytics focus more on data statistics, distributions and visualizations, hypothesis testing, statistical inference and predictive analytics
  ➔ statistical and predictive analysis

Job Market

Traditional Positions in IT/CS/CIS

- Software Engineer
- Database Manager
- Network Architect
- UX Designer
- etc

Popular Data-related Job Positions

- Data Analyst
- Data Engineer
- Data Scientist
- etc

Job Market

Traditional Positions in IT/CS/CIS

- Software Engineer, $50K-$100K
- Senior Software Engineer, $70K-$160K
- DBA, $50K-$80K
- Senior DBA, $70K-$ 120K

Popular Data-related Job Positions

- Data Analyst, $50K- $70K
- Senior Data Analyst, $65K- $110K
- Data Engineer, $70K- $115K
- Senior Data Engineer, $100K- $165K
- Data Scientist, $85K- $170K
- Senior Data Scientist, $120K- Secret

Data Science

- Data Science = Statistics + DM + ML + App/Domain

Popular Jobs in Data Science

Data Analyst

- Statistics, Data processing, analytics, visualizations
- Tools: R, SAS, SPSS, Matlab, Excel, etc

Data Engineer

- Database, data management, big data
- Tools: Hadoop, Hive, Pig, Spark, Oracle, SAP, etc

Data Scientist

- Statistics, data mining, machine learning, AI, etc
- Tools: R, Python, Hadoop, Spark, Weka, etc

Intro

- Who am I
- Data Management and Data Analytics
- Topics in This Class
- Syllabus, Blackboard system and Policy

Why Data Analytics

- Take jobs and salaries for example. Assume you got multiple job offers
  - Chicago, Data Analyst, $70K
  - San Francisco, Soft Engineer, $92K
  - Los Angeles, Data Analyst, $80K
  - And so forth
- So, which job you should take? Let's ignore other factors, and only consider the salary level

Why Data Analytics

- You may have several questions. For example
  - $70K as a data analyst in Chicago is good?
  - But I can get $80K in Los Angeles for the same title
  - How about soft engineer? It seems the salary is better
  - It seems that it is easier to compare the position of data analyst between Chicago and Los Angeles, since they are the offers with the same title but different salary. However, is it comparable for the offer of soft engineer in SF vs the data analyst in Chicago?

ITMD 527: Data Analytics

- There are three major topics covered in this class
  - Statistic Basics
    - Data and data types
    - Describe your data numerically
    - Describe your data by visualizations
  - Hypothesis Testing
    - Validate your assumptions by statistical methods
  - Predictive Models
    - Make predictions on numerical or categorical variables

ITMD 527: Data Analytics

Basics in Data Analytics

- Descriptive Statistics
- Hypothesis Testing
- Linear Regressions by R
- ANOVA

data statistics, distributions and visualizations, hypothesis testing, statistical inference, predictive models

---------- Midterm Exam ----------

More Topics in Data Analytics

- Linear Regressions by SAS
- Classification Models
- Advanced Topics
  - Data mining vs machine learning
  - Other regression models, etc

More predictive models

Learning Objectives

- Understand and be familiar with statistical basics
- Be able to analyze real-world data, and use statistical basics (including hypothesis testing) to perform basic analytical tasks. For example, the questions given in the example of job offers
- Understand and be able to build predictive models
- Be able to use tools (R or SAS) for real-world practice
- Foster analytical mind and skills ➔ understand techniques, know how to use them, can identify problems in the models, and fix these issues, finally, be able to interpret the outputs/results, and derive meaningful findings or conclusions

Misunderstandings, Mistakes or Problems

- This is a math class
  We do learn statistics, but they are easier than math if you can understand the concepts well
- This is a coding class
  - We do use R or SAS for analytical tasks, but they are not complicated as programming languages
  - Several students focus too much on codings, and ignore the underlying theories, analytical minds and skills
- This is a theory class
  - We learn concepts which are not complicated
  - We actually use real-world data for analysis
  - We have many practice in class and assignments

Teaching Style

- We use case-study learning style
  - We introduce a case/data before the discussions or the learnings of a specific technique
  - Most of the data are real-world data sets
  - We use the case as example to analyze step by step, and use other cases/data as practice
- We need in-class communications
  - You need to think together with the class
  - You need to actively interact with the class
  - You need to answer my questions in the class

How to Learn Data Science/Analytics

- First of all, you must focus on understandings
  Ask the following questions frequently
  - Do you know and understand this method works? Try to remember how it works by closing notes/books
  - Do you know a method/technique should be used in which situations? And its pros & cons?
  - Do you know the potential issues and corresponding solutions?
- Moreover, practical skills are also important
  - Use tools (R, SAS, Python) for analytics
  - Be able to analyze, debug, explain or interpret the outputs

Intro

- Who am I
- Data Management and Data Analytics
- Topics in This Class
- Syllabus, Blackboard system and Policy

ITMD 527 (Tuesdays & Thursdays)

The following policies and introductions are applied to the sections below:

- IT-D 827-01 and ITMD 527-01, Live Section
- IT-D 827-02 and ITMD 527-02, Online Section
- ITMD 527-03, Remote Students from India

Students in Live Section are expected to come to the classroom. We will randomly issue roster sheets

Ultimate Goals

- You can identify research problems
- You can figure out appropriate solutions to solve specific research problems
- You can use data analytics tools to apply the approaches on real-world data set, and evaluate how good your solutions are
- You can analyze and interpret the experimental results
- You can use SAS and R to perform analytical tasks

ITMD 527 (Tuesdays & Thursdays)

- Time and Place

  Time: Tuesdays & Thursdays, 08:35 AM to 09:50 AM
  Location: Stuart Building (SB), IIT Main Campus

- Office Hours

  Tuesdays 1 PM – 4 PM
  Location: Room 221, Perlstein Hall

  It's better to reserve for office hours by sending emails;

  You can only stand by if you did not reserve;

Syllabus and Blackboard

- Advising Rules
  - Questions on lectures and assignments
  - Questions on class projects
  - Questions on research projects & assistants
  - Questions on jobs and careers
  - Usually, I do not help you on debugging…
  - You can also seek for helps on Discussion Forums
- TA
  - TA will grade your assignments and answer your questions
  - TA also have office hours for QAs

Syllabus and Blackboard

- Important Dates

Jan 14	Spring Classes Begin
Jan 21	Martin Luther King, Jr. Day - No Classes
Feb 6	Spring Final Exam Schedule Published Online
Mar 18-23	Spring Break Week, No Classes
Mar 25	Fall Class Schedule Published Online
Apr 1	Last Day to Withdraw from Full Semester Classes
Apr 8	Fall Registration Begins
May 6-11	Final Exam Week
May 15	Final Grades Due

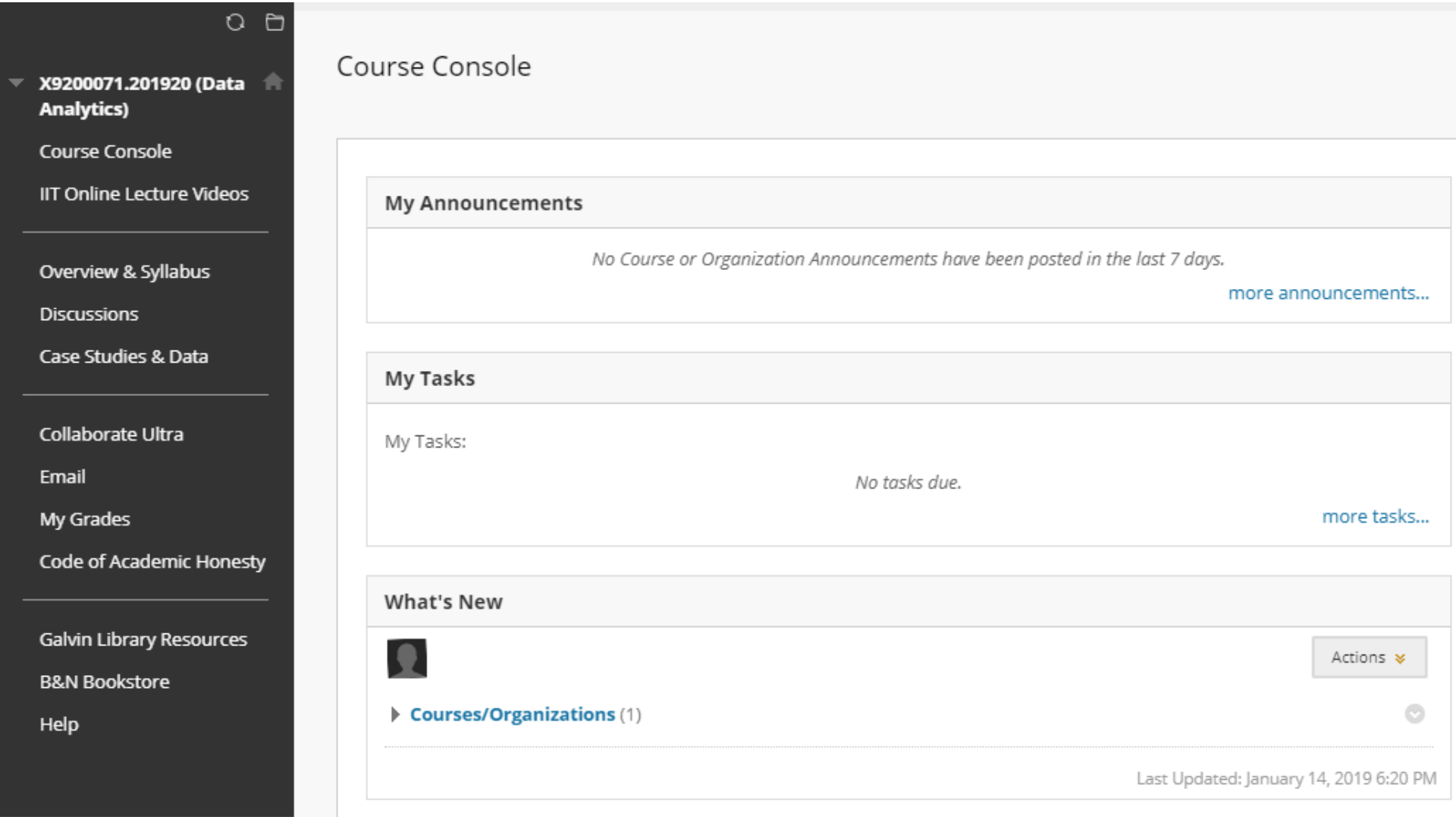
- Jan 26, Last Day to Add/Drop for Full Semester Classes with No Tuition Charges

Syllabus and Blackboard

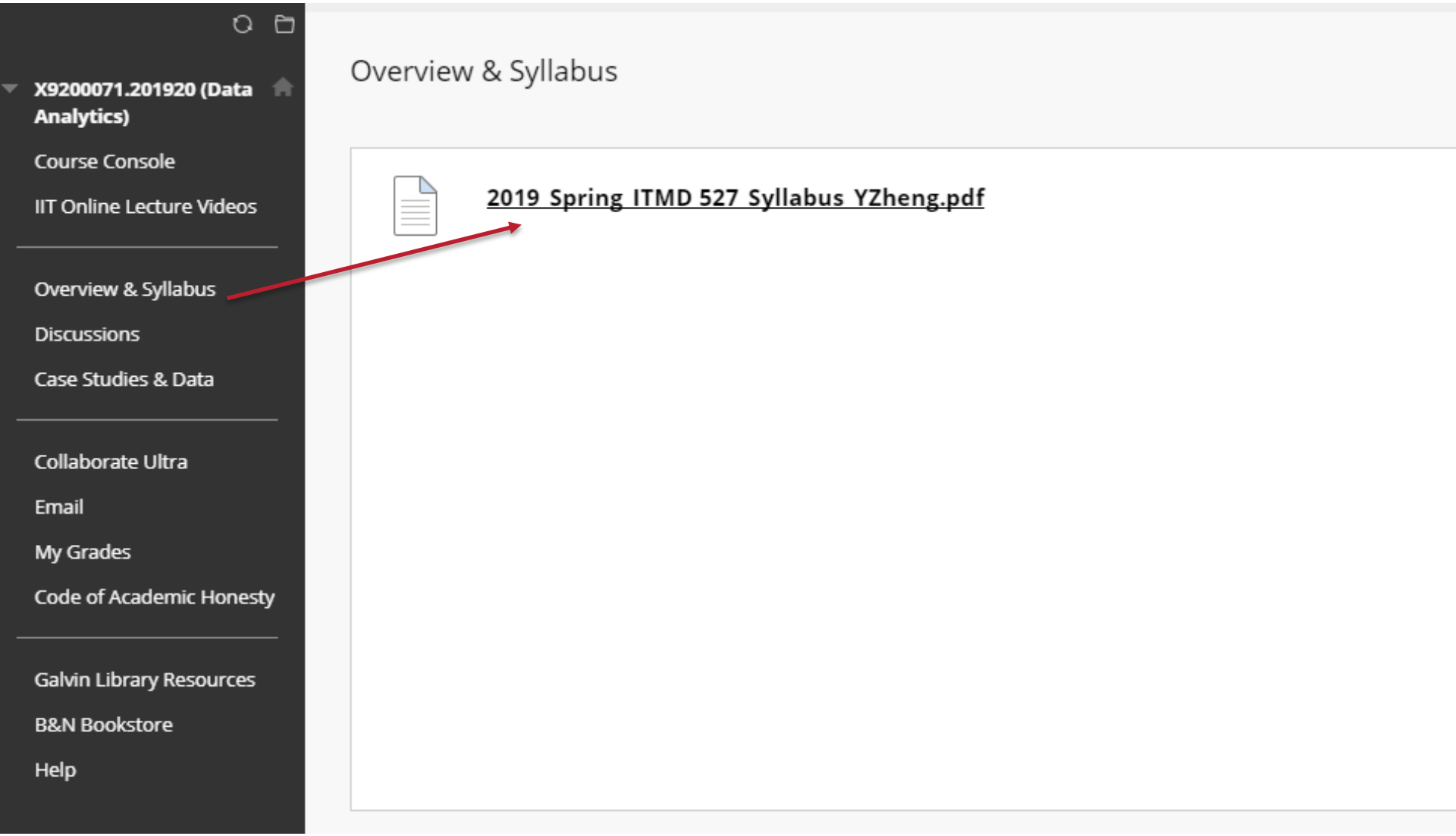
Laptops

- From Fall, 2016, ITM requires students to bring their own laptops to attend the classes.
- We only one Lab in ITM and the lab is usually reserved for undergraduate classes or limited graduate classes
- We may have in-class lab practice, and you should bring your laptop to the class

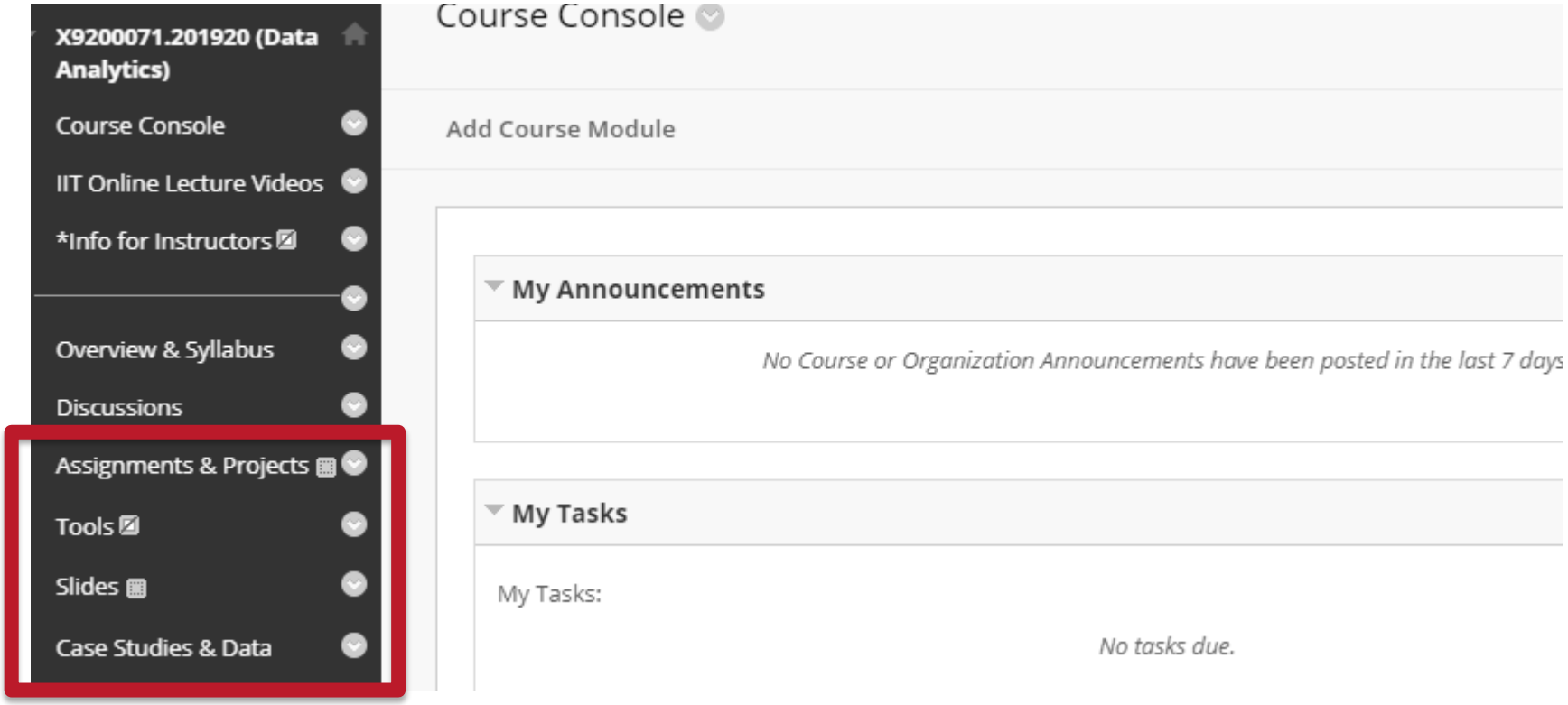
Blackboard


X9200071.201920 (Data Analytics)
Course Console
IIT Online Lecture Videos
Overview & Syllabus
Discussions
Case Studies & Data
Collaborate Ultra
Email
My Grades
Code of Academic Honesty
Galvin Library Resources
B&N Bookstore
Help
Course Console
My Announcements
No Course or Organization Announcements have been posted in the last 7 days.
more announcements...
My Tasks
My Tasks:
No tasks due.
more tasks...
What's New
Actions
Courses/Organizations (1)
Last Updated: January 14, 2019 6:20 PM

Blackboard

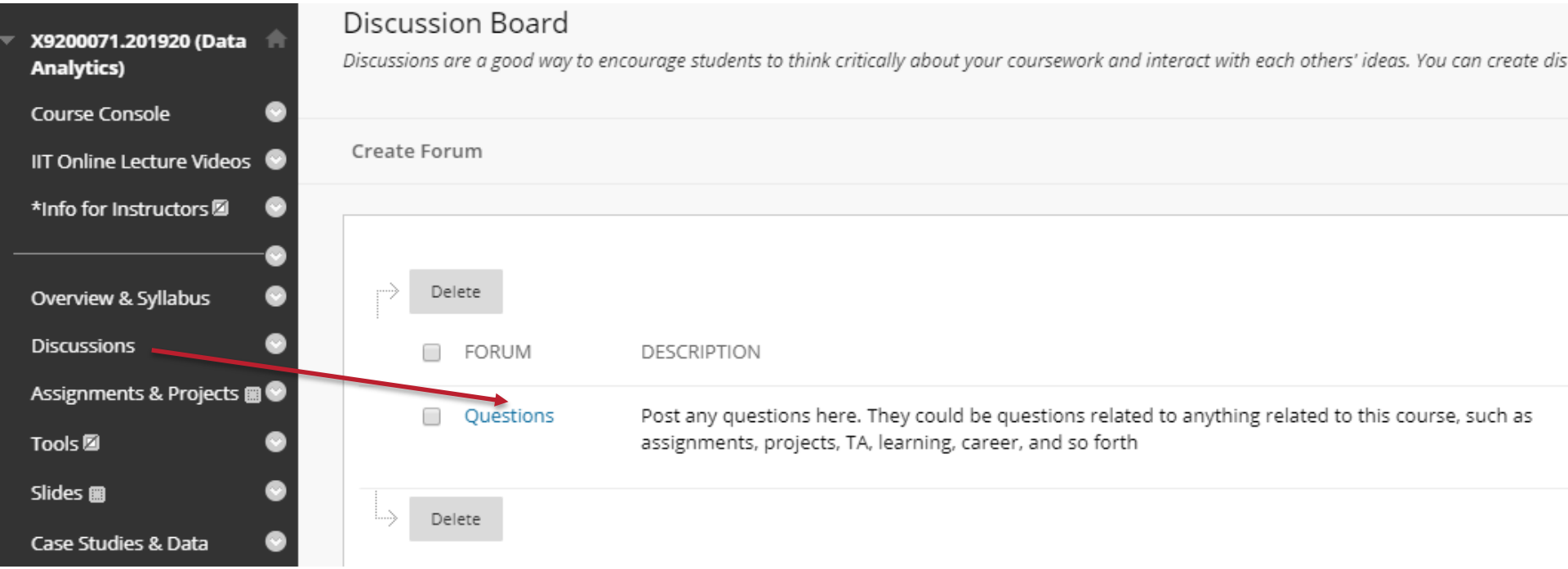

X9200071.201920 (Data Analytics)
Course Console
IIT Online Lecture Videos
Overview & Syllabus
Discussions
Case Studies & Data
Collaborate Ultra
Email
My Grades
Code of Academic Honesty
Galvin Library Resources
B&N Bookstore
Help
Overview & Syllabus
2019_Spring_ITMD 527_Syllabus_YZheng.pdf

Blackboard


X9200071.201920 (Data Analytics)
Course Console
IIT Online Lecture Videos
*Info for Instructors
Overview & Syllabus
Discussions
Assignments & Projects
Tools
Slides
Case Studies & Data
Course Console
Add Course Module
My Announcements
No Course or Organization Announcements have been posted in the last 7 days
My Tasks
My Tasks:
No tasks due.

Blackboard

- Discussion Board



Cannot share assignment/exam results on board

Syllabus and Blackboard

- Email Rules

1) Clear title

   ITMD527 – I want to …

2) Clear descriptions

   tell me which class are you in (MW or Tu/Th)

   clearly describe what questions you have.

   For office hour reservations, you should estimate how many minutes you need for our talk

Syllabus and Blackboard

- Example of Bad Emails


Meeting with you Inbox x

[redacted]

Hello Professor,
Are you available today after 3pm? I would like to meet you on certain clarifications. Do let me know you convenient time.
Thank you so much. Have a nice day.

Regards,

- Which course/class?
- What are your questions?
- Which pieces of work (lectures or assignments) you need clarifications?

Syllabus and Blackboard

- Example of Bad Emails

Fwd: Regarding Late Submission Inbox x

10/10/16

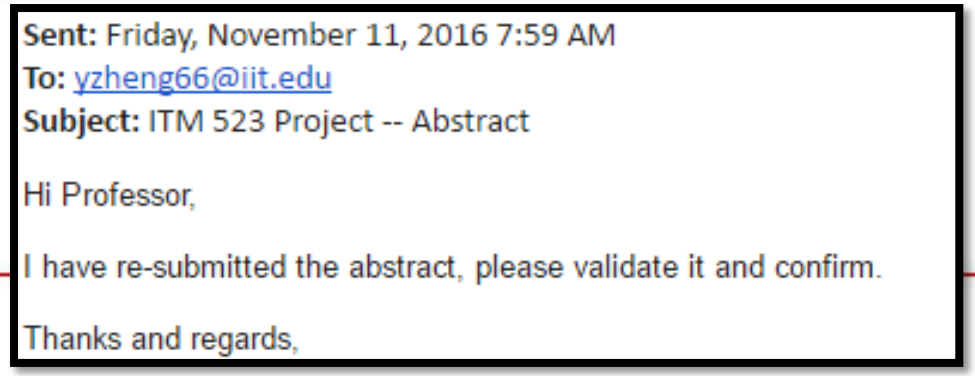
Hello Sir,
Thanks for your time,

My Name is [redacted] and I could not submit my assignment(HW5) on time.I submitted on OCT 5,1:31A.M as it was a very time taking and lengthy.Please forgive me this time ,from the next time I will take care to not submit it late.

Please consider my request for the first and last time.I will be very thankful to you.

Regards,

- Which course/class are you in?



Sent: Friday, November 11, 2016 7:59 AM
To: yzheng66@iit.edu
Subject: ITM 523 Project -- Abstract

Hi Professor,

I have re-submitted the abstract, please validate it and confirm.

Thanks and regards,

Syllabus and Blackboard

Textbooks

- William Mendenhall, Terry Sincich. “A Second Course In Statistics: Regression Analysis (7th Edition)”. Pearson Education, ISBN: 9780321691699; Note: You can also use the 6th Edition
- Alvin C. Rencher and G. Bruce Schaalje. "Linear Models in Statistics (2nd Edition)". John Wiley & Sons, Inc., 2008; ISBN 978-0-471-75498-5;
- Geoff Der and Brian S. Everitt. "A Handbook of Statistical Analyses using SAS (2nd Edition)". Chapman & Hall/CRC, 2002; You may find an eBook online.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. "An Introduction to Statistical Learning with Applications in R"; Springer; You may find an eBook online.

Syllabus and Blackboard

Supplemental Textbooks

- W. N. Venables, D. M. Smith and the R Core Team. "An Introduction to R";
- SAS Team. "Step-by-Step Programming with Base SAS 9.4";
- John W. Foreman. "Data Smart: Using Data Science to Transform Information into Insight". John Wiley & Sons, Inc., 2013; ISBN: 978-1-118-66146-8; [This is related to Excel]

Syllabus and Blackboard

Online Resources

- KDnuggets, Data Mining Community's Top Resource for Data Mining and Analytics info. http://www.kdnuggets.com/
- Statistics Concepts, http://www.statsoft.com/Textbook/Elementary-Statistics-Concepts
- R tutorial, https://www.tutorialspoint.com/r/
- SAS tutorial, http://support.sas.com/training/tutorial/
- Excel, http://www.excel-easy.com/data-analysis.html

Syllabus and Blackboard

Assignments

- Written Assignments
  Examine your understanding and skills in analytics
  May ask you to process real-world data sets
  May ask you to use tools to perform data analytics
- Final Project: could be team project; more details will be given after midterm exam

Syllabus and Blackboard

About Final Project

- Write a project proposal [after midterm exam]
  Introduce your research problems, which data you will use, what are the solutions and evaluations, what are the expected outcomes
- Work on the experiments
- Present your work and submit final project reports
  For remote student from India, you can record a video of your presentation and send it to me

Syllabus and Blackboard

Examinations

- Exam-Regression: closed books/notes/laptops/ mobile devices, and written exam
- Exam-Classification: closed books/notes/laptops/ mobile devices, and written exam
- Final Project and presentations: could be team project; you should present your project and submit project reports eventually. More details will be given after midterm exam

Note: you can bring a calculator (without communication functions) to exams

Syllabus and Blackboard

Rules in Assignments and Examinations

- Assignments

  Usually, no late submission is allowed

  15% penalty will be applied for late submission (in 1 week). Late submission will be ignored if later than 1 week

  e.g., due date is March 1st 11:59 PM

  No penalty if submitted before due

  15% penalty if between due time and Mar 8th 11:59 PM. A zero score if submitted later than Mar 8th 11:59 PM

Syllabus and Blackboard

Rules in Assignments and Examinations

- Grading criteria

Regular assignments	27%
Exam1	25%
Exam2	20%
Final Projects	25%
Attendance	3%

Note: the attendance for live students will be examined by signature on the roster sheets. The roster sheets will be distributed randomly.

The attendance for online/remote students will be evaluated by filling a google form. I will ask TA to prepare a google form every week, you have the chance to sign it by Friday. After Friday, the form will be closed, and you cannot sign it again.

Syllabus and Blackboard

Scales for your final grades

Grading: Grading criteria for this course will be as follows:

A	*Outstanding work reflecting substantial effort*	90-100%
B	*Adequate work fully meeting that expected of a graduate student*	80-89.99%
C	*Satisfactory work meeting minimum expectations*	60-79.99%
F	*Unsatisfactory work*	0-59.99%

The final grade for the class will be calculated as follows:

Regular Assignments	**27%**
Exam 1	**25%**
Exam 2	**20%**
Final Project and Presentations	**25%**
Class Participation	**3%**

Syllabus and Blackboard

Academic Honesty

- Plagiarism is a very serious problem, and it is forbidden in all submissions, including assignments, paper reviews, midterm exams and final projects, as well codings/reports in these submissions.
- New rules from Fall, 2018 ➔ you cannot share assignments/exams online

Syllabus and Blackboard

So, what is plagiarism?

- Cheating in the exams or final projects
- Cheating or copying answers from other students or other resources (such as online materials) in assignments, exams or final projects
- Cheating or copying texts from other resources without references in paper writing or reviews

Blackboard system has a special function to indicate percentage of similar texts from online resources

Syllabus and Blackboard

Example of plagiarism in assignments

- For any concept questions in the assignments, such as “what is statistics?. You can learn by searching answers from Internet, but you cannot simply copy the original texts online in your assignments. You should use your own language/texts as answers based on your understandings.

Syllabus and Blackboard

Policies of plagiarism

- 1st Time, you will get a zero score and warning
- 2nd Time, you will get a zero score, and a fail (E) in this class, and an Academic Honesty Violation Report (AHVR) will be filed
- You will be expelled from the school if you get two AHVR

Note: only the final project could be a team project. Other assignments are individual home work

Examples of Unreasonable Requests

Reasonable requests can be accepted in some situations. Unreasonable requests will be ignored.

- Can you give me a second chance on the exam or the assignments?
- Can you give me extra practice or assignments so that I can improve my grade?
- Can you accept my super-late submissions (more than 1 week), because I have a medical issue this semester?

Important Notes

Syllabus and Blackboard

- It is your responsibility to read syllabus online and be familiar with the Blackboard system

Medical Issues and Disabilities

- You must introduce your medical situation as early as possible. You cannot use it as an excuse at the end of the semester
- Your medical issues or disabilities must be verified by the center for disability resources (CDR); telephone 312.567.5744 or disabilities@iit.edu

Next Class

Data and Statistic Basics

Case Study 1 has been uploaded to Blackboard, you can take a look at it and start thinking about relevant questions or concerns