

**ECO541: Health Care Data Analytics**  
**Online Class (Monday 5:45-7:15 PM )**

updated: 2/17/2022

---

**INSTRUCTOR:** Jin Man Lee

Office: Loop Campus (1 E Jackson DPC 6230)

Phone: (312)362-5970

Email: jlee141@depaul.edu

**COURSE DESCRIPTION**

Health Care Data Analytics introduces health care data and analytical methods for students who have had little previous data analytics coursework or experience. Topics to be covered in this course include creating datasets, the structure of datasets, an introduction to data warehousing and working with large databases, an introduction to public health and health-care datasets, methods for descriptive analytics, and an introduction to predictive analytics. Students will gain skills in data manipulation for program evaluation and analysis. Some statistical programming will be required, and students are expected to have some facility with a statistical package using R.

**COURSE LEARNING OBJECTIVES**

- Practice the issues related to the collection, sharing, and structuring of healthcare data using R with relevant packages
- Build skills to know how to find, download/extract, clean, and do descriptive analyses on a variety of "Big Data" analytics
- Technical analysis to understand concepts related to healthcare innovation using descriptive analytics and predictive analytics. Examples from
  - National Health Interview Survey (NHIS)
  - American Community Survey (ACS)
  - Healthcare Cost and Utilization Project(HCUP)
  - Behavioral Risk Factor Surveillance System (BRFSS)
  - Pandemic and Epidemic Data (COVID19)
- Apply critical thinking using machine learning (ML) and artificial intelligence (AI) technical skills to inform business and policy decisions
- Understand current barriers in healthcare management and how data analytics can provide potential solutions to improve quality, lower cost, and advance the outcome

**COMMUNICATION**

The best way to reach me is to send an email. Please use ECO541 as a prefix on the subject line to get my attention. If you don't receive my reply within 24 hours, please remind me

again. Due to some email filters, your email might be lost. Before sending the email, read the discussions on D2L. You will find the most frequent questions and answers there. It is required to use the DePaul email address to communicate with the instructor for security.

### **ZOOM CLASS and OPEN OFFICE HOURS**

- **ZOOM CLASS:** Every Monday 5:45-7:15 PM. This is a required class to attend. Attendance will be automatically recorded by ZOOM for the attendance grade. Make sure to use your full name as a screen name for attendance credit.
- **ZOOM OPEN OFFICE HOUR:** Every Monday 4:00-5:00 Thursday 4:00-5:00 PM. This is an individual office hour and first come, first served. You will be in the waiting room if someone is in the room. Any suggested topics are welcome via email before the meeting if you come as a group.

### **PREREQUISITE**

Applied Quantitative Analysis or passed the equivalent tests at the graduate school level

### **TEXTBOOK**

- Healthcare Data Analytics, Edited by Chandan K. Reddy and Charu C. Aggarwal (2015), CRC Preess, ISBN: 978-1-482-23212-7
- R for Health Data Science, Ewen Harrison and Riinu Pius (2021), CRC Preess, ISBN: 978-0-367-85542-0, available from [https://argoshare.is.ed.ac.uk/healthyr\\_book/](https://argoshare.is.ed.ac.uk/healthyr_book/)
- Healthcare Analytics for Quality and Performance Improvement, Trevo L. Strome (2013), John Wiley & Sons Inc. ISBN: 978-1-118-51969-1
- More Reading List will be available on D2L every week

### **SOFTWARE USED in CLASS (required for Assignments)**

- R for any platform. Current version or Above to 3.6.0. This is a free public domain statistical program available to install. We will use R studio server from <http://bigblue.depaul.edu:8787>

### **GRADE**

- Weekly Assignments (40%), Attendance (5%), Midterm Project (25%), Final Project Presentation and PPT(5%), and Final Project (25%)
- Scale of grade: A: 93 or above, A-: 88-92.9, B+: 85-87.9, B: 80-84.9, B-: 77-79.9, C+: 75-76.9, C: 70-74.9, C-: 68-69.9, D+: 65-67.9, D: 60-64.9, F: Below 60

### **PROJECTS**

- Midterm Project: Guided Health Data Analytics using R (5/9/2021 to 5/15/2021 at 10:00 PM)
- Final Project Proposal via ZOOM presentation (PPT required) on 5/30 (each group has about 10 minutes of presentation and 5 minutes of discussion)
- Final Project Due on 6/11 (Saturday) at 10:00 PM, upload all files to D2L

## **ASSIGNMENTS**

Weekly Assignment related to review of computational work using example data.

- All assignments are to be prepared individually. You risk an academic integrity violation if you submit the same work and answers with others. Group study is encouraged but not the submission of Assignments.
- All statistical code needs to be submitted as a txt file, For example, FileName.R.txt.
- All submitted codes should be tested without any error. If there are any errors, the assignment will get zero credit.
- All weekly assignments should be uploaded to D2L by Sunday at 10:00 PM
- Late submission will get a steep penalty without prior approval by the instructor.

## **DISCUSSIONS in D2L**

We will have weekly discussions in D2L. You can ask any questions related to the material we covered each week, including homework. This is an excellent online-community space, so you are welcome to give answers or explanations to the questions. I will review the discussion board and leave comments if needed.

## **ACADEMIC HONESTY**

Work done for this course must adhere to the University Academic Integrity Policy. Violations include but are not limited to the following categories: cheating; plagiarism; fabrication and academic misconduct.

- Cheating: any action that violates University norms or an instructor's guidelines for the preparation and submission of assignments. Such actions may include using or providing unauthorized assistance or materials on course assignments, or possessing unauthorized materials during an examination.
- Plagiarism: the representation of others' work as your own. You are to prepare your own assignments. Violations may result in the failure of the assignment, failure of the course, and/or additional disciplinary actions.
- Misconduct: This includes but is not limited to attempts to bribe an instructor for academic advantage; persistent hostile treatment of, or any act or threat of violence against, an instructor, advisor or other students. Violations may result in additional disciplinary actions by other university officials and possible civil or criminal prosecution.

You may review the Academic Integrity Policy in the Student Handbook or by visiting Academic Integrity at DePaul University (<http://academicintegrity.depaul.edu>)

**STUDENT WITH DISABILITY**

Students with Disability may register The Productive Learning Strategies (PLuS) Program. You may request your exam schedule arrangement by requesting through the PLuS program. For more information on the PLuS program, you may visit <http://studentaffairs.depaul.edu/plus/> or call: 312-362-8000.

---

## SUMMARY OF WEEKLY SCHEDULE

1. Sunday 10:00 AM : Weekly Material will be posted in D2L including Lecture Note and Homework
  2. Monday 4:00-5:00 PM: ZOOM Open Office Hour
  3. Monday 5:45-7:15 PM: ZOOM Online Class (Attendance Required)
  4. Thursday 4:00-5:00 PM: ZOOM Open Office Hour
  5. Sunday 10:00 PM : Weekly Assignment Due
- 

## TENTATIVE SCHEDULE OF TOPICS

(The instructor may change the order or contents by needs, any special material needs for the class will be available on D2L)

- WEEK 1-2. Introduction to Health Care Data in R
  - TOPIC 1: Introduction to health analytics and Public Data
    - \* Introduction to Public Health Data
    - \* Administrative Claims Data
    - \* Research design and methods for Health Data Analysis
  - TOPIC 2: Introduction to R and Useful Packages
    - \* Introduction to R
    - \* Understanding data types
    - \* dplyr and ggplot packages in R
  - TOPIC 3: Public Health Care Data Analysis Using R
    - \* Survey data analysis in R
    - \* National Health Interview Survey (NHIS) Data
    - \* Health Insurance Data from the American Community Survey (ACS)
- WEEK 3-4 Health Care Costs and HCUP Inpatient Data Analysis
  - TOPIC 4: Health Care Costs Data Analysis
    - \* Basic definitions: costs, expenditures, prices, charges
    - \* Payment design underlying claims data
    - \* Measuring costs, reporting spending and prices
    - \* Healthcare Cost and Utilization Project (HCUP) Inpatient Sample Data (ISD)
  - TOPIC 5: Database and Descriptive Analytics on HCUP
    - \* Descriptive Analytics using R
    - \* Correlation Analysis

- \* Tabulation and Group Data Analysis
  - \* Visualization using ggplot
  - \* Database and Structural Query Language (SQL)
- Week 5-6: Analysis on Measuring Health and Quality
  - TOPIC 6: Health related quality measures
    - \* Behavioral Risk Factor Surveillance System (BRFSS) data
    - \* Health outcome measures
    - \* Measuring incidence and prevalence
    - \* Health care quality measures
    - \* Measures of illness, severity, co-morbidity
  - TOPIC 7: Predictive Analytics on Behavioral Risk Factor Surveillance System (BRFSS) data
    - \* Classification using Clustering and Principal Component Analysis
    - \* Regression Analysis to evaluate outcomes and quality measurement
  - MidTerm Project on Health Data Descriptive Analytics
- Week 7-8: Uses of health analytics on Infectious Diseases and Pandemics
  - TOPIC 8: Pandemic and Epidemic Analysis and Health Policy Evaluation
    - \* Epidemic Analysis and Trend
    - \* Infection Rate on Demography
    - \* Health policy evaluation and design
  - TOPIC 9: Classification and on Pandemic and Epidemic Data
    - \* Regression Analysis of Geographic and Demographic Information on Pandemics
    - \* Time Trend Analysis on Pandemic Data
    - \* Binary and Multinomial Classification Analysis on Outcome of Infection
    - \* Survival Analysis in Disease Control
- Week 9: Health analytics in practice using Machine Learning (ML) and Deep Learning (DL)
  - TOPIC 10: Machine Learning on health care outcomes
    - \* Random Forest Model Development on Health Care Outcome
    - \* Artificial Neural Network Model
    - \* Evaluations of Machine Learning Models and Further Topics on DL
- Week 10: Research Project on Health Care and Final Project
 

Each person will work on the topics of interest. The proposal PowerPoint includes introduction the data and descriptive analytics. The project will start after the midterm exam, and the related techniques will be developed throughout the class. If allowed, students can bring their own data or any public data they are interested in. The maximum number of members are three in a group.
- FINAL PROJECT DUE : June 11 (Saturday) 10:00 PM