

LTEC 6514: Foundation of Data Science & Learning Analytics

Youngjin Lee

August 28, 2019

E-mail: Youngjin.Lee@unt.edu

Office Hours: By appointment

Office: Discovery Park (G155)

Web: youngjin-lee.rbind.io

Class Hours: Wednesday 6:00 - 8:00 PM (Synchronos)

Class Room: D209A @ Discovery Park (Synchronous)

Course Description

This is an applied course for doctoral students with little-to-no programming experience who are interested in data science and learning analytics. The focus of the course is on building foundational skills required for conducting data science and learning analytics research. Major emphasis is placed on a pragmatic understanding of core principles of programming and packaged implementations of methods. Students will leave the course with basic computational skills implemented through various computational methods and approaches to data science and learning analytics. Students will be able to apply, adapt and expand these skills for new questions and problems they will encounter.

Course Objectives

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

1. Construct and execute basic programs in R and tidyverse packages
2. Transform raw data into a tidy format
3. Conduct an exploratory data analysis (EDA) on tidy data
4. Visualize information using appropriate graphical techniques
5. Create interactive Web apps using Shiny

Textbooks

1. R Programming for Data Science (**RP4DS**, <https://bookdown.org/rdpeng/rprogdatascience/>)
 2. R for Data Science (**R4DS**, <https://r4ds.had.co.nz/>)
- These books are freely available at the above Websites.
 - If you prefer paper copies, you can purchase them from Amazon.

Software

Students must install the following software applications on their computer. All software applications are free and available for macOS and Windows operating systems. Detailed installation instructions will be provided on the first synchronous meeting. Students attending the synchronous meetings in person must **bring their own laptop** with these software and R packages installed.

- RStudio (<https://www.rstudio.com/products/rstudio/download/#download>)
- R (<https://cloud.r-project.org>)
- R packages:
 - knitr
 - tidyverse
 - nycflights13
 - forcats
 - maps
 - gapminder
 - Lahman
 - rcfss
 - cluster
 - factoextra
 - kohonen
 - shiny
 - tibble
 - stringr
 - devtools
 - region5air

Topics & Class Schedule

Although some course activities take place asynchronously, we will have several synchronous meetings during the semester. Synchronous sessions will be held in NTDP D209A on Wednesday evenings between 6 - 8:00 pm Central Time for students who wish to attend in person. Distant students can participate in the synchronous meetings through Zoom. All synchronous sessions will be recorded and made available to students for their review. Weeks not marked as synchronous are asynchronous sessions during which students are expected to complete the assigned reading, and execute all R codes in the reading. Before executing the R code, try to **predict** the result, and **compare** your predictions to the observed outcome. If you did not get what you predicted, think about why you got a different result, and read the textbook again. Repeat these processes until you fully understand the R code you are executing, and can correctly predict the outcome of your R code.

Week 1 (August 28): Introduction (Synchronous session #1)

- Review of syllabus
- Install R, Rstudio, and R packages
- Reproducible research with R Markdown

Week 2 (September 4): R Basics I

- Reading assignments:
 - [R Markdown](#)
 - [R Nuts and Bolts](#)
 - [Tibbles](#)
 - [Subsetting R Objects](#)
 - [Vectorized Operations](#)
 - [Functions](#)
- * Sections 14.4, 14.5 and 14.6 cover advanced concepts. Skim through these sections for now and come back to them later when you need them.

Week 3 (September 11): R Basics II (Synchronous session #2)

- Review of important concepts and skills introduced in Week 2
- Hand-on computing activities

Week 4 (September 18): Visualization and Grammar of Graphics I

- Reading assignments:
 - [Introduction](#)
 - [Explore: Introduction](#)
 - [Data Visualization](#)
 - [Workflow: Basics](#)
- Assignment 1 (Exploratory Data Analysis & Data Visualization) out

Week 5 (September 25): Visualization and Grammar of Graphics II (Synchronous session #3)

- Review of important concepts and skills introduced in Week 4
- Hand-on computing activities

Week 6 (October 2): Data Transformation I

- Reading assignment:
 - [Data Transformation](#)
 - [Workflow: Scripts](#)

Week 7 (October 9): Data Transformation II (Synchronous session #4)

- Review of important concepts and skills introduced in Week 6
- Hand-on computing activities

Week 8 (October 16): Data wrangling I

- Reading assignment:
 - [Wrangle: Introduction](#)
 - [Data Import](#)
 - [Tidy Data](#)
- Assignment 2 (Data Transform, Tidying and Munging) out

Week 9 (October 23): Data wrangling II (Synchronous session #5)

- Review important concepts and skills introduced in Week 8
- Hand-on computing activities

Week 10 (October 30): Relational Data & Factor I

- Reading assignment:
 - [Relational Data](#)
 - [Factors](#)
- Assignment 3 (Relational Data & Factor) out

Week 11 (November 6): Relational Data & Factor II (Synchronous session #6)

- Review of important concepts and skills introduced in Week 10
- Hand-on computing activities

Week 12 (November 13): Clustering & Self-Organizing Map I

- Reading assignment
 - [K-means Cluster Analysis](#)
 - [Hierarchical Cluster Analysis](#)
 - [Modeling Self Organising Maps in R](#)
 - * Note that R code is using an old version of kohonen package. I will explain how to use the current version of kohonen package in the next synchronous session.
 - [Self-Organizing Maps](#)(skim through it)
- Assignment 4 (Clustering & Self-Organizing Map) out
- Final project out

Week 13 (November 20): Clustering & Self-Organizing Map II (Synchronous session #7)

- Review of important concepts and skills introduced in Week 12
- Hand-on computing activities
 - Creating a Self-Organizing Map using R kohonen package version 3.x

Week 14 (November 27): Building Shiny App I

- Reading assignment:
 - [RStudio's Shiny Tutorial #1](#)
 - [RStudio's Shiny Tutorial #2](#)
 - [RStudio's Shiny Tutorial #3](#)
 - [RStudio's Shiny Tutorial #4](#)
 - [RStudio's Shiny Tutorial #5](#)
 - [RStudio's Shiny Tutorial #6](#)
 - [RStudio's Shiny Tutorial #7](#)

Week 15 (December 4): Building Shiny App II & Course De-brief (Synchronous session #8)

- Review of important concepts and skills introduced in Week 14
- Hand-on computing activities

Course Policies

Correspondence

Detailed descriptions of learning activities will be posted in Canvas. Students must check Canvas regularly and submit their work to the designated place in Canvas.

To reach me, email me at youngjin.lee@unt.edu any time, and expect a response within 48 hours during the work week (please include [LTEC 6514] in the subject line).

Platform and Basic Technical Requirements

This course uses [Canvas](#) as its main platform for file sharing and communication, and will integrate a significant selfdirected study.

Also, a university-assigned student email account is the official means of communication with all students. Students are responsible for all information sent to them via this universityassigned email account. If a student chooses to forward their university email account or to use some other email account, he or she is still responsible for all information sent to the official university email account. Be sure to read messages sent with Canvas as well.

Grading

Final grades will be based on the percentage of points earned toward the total possible points (points earned divided by points possible times 100 = percentage).

Grade	Percentage
A	90% - 100% of total allocated points
B	80% - 89%
C	70% - 79%

Grade	Percentage
D	60% - 69%
F	59% and below

- 15% of your grade will be determined by your synchronous participation (e.g., exercise).
- 55% of your grade will be determined by 4 assignments:
 - Exploratory Data Analysis (EDA) & Visualization (15%)
 - Data Transform and Munging (15%)
 - Relational Data and Factor (10%)
 - Clustering and Self-Organizing Map (SOM) (15%).
- 30% of your grade will be determined by your final project:
 - Applying data science and learning analytics techniques and skills to your own or provided data set
 1. Come up with three or more interesting questions about data set
 2. Conduct an EDA, including appropriate visualizations, to answer your questions
 3. Conduct a clustering and/or SOM analysis
 4. (Optional) Share the analysis result as a Shiny app
 5. Write a written report (1) describing the data set, questions, EDA results, and clustering/SOM analysis results (optionally Shiny app if you chose to do the optional Shiny part); and (2) proposing a possible follow-up study

Late Assignments

All assignments are to be submitted by the scheduled due date. Late assignments will be penalized affecting both project and course letter grade. Works may not be accepted if they are submitted one week after the due. Exceptions to this policy will only be made in extraordinary circumstances. As a general rule, you should not request an exception to this late policy.

Incompletes

The department complies with university policy regarding the assignment of an Incomplete Grade in any course. Please see <http://essc.unt.edu/registrar/academicrecordincomplete.html> for information. Per UNT policy, a grade of Incomplete can only be awarded to a student who is 1) passing the course at the time of the request, and 2) has a justifiable and documented reason, beyond the control of the student, for not completing the course work on schedule. Notification and submission of documentation must be provided to the instructor at the time of the emergency.

Cheating and Plagiarism

Cheating is the actual or attempted practice of fraudulent or deceptive acts for the purpose of improving one's grade or obtaining course credit; such acts also include assisting another student to do so.

Under all circumstances, you are bound by the UNT policies on academic dishonesty and cheating. Any materials you have used or adapted must be fully credited the original author. Any verified act of plagiarism, no matter how seemingly small or inconsequential, will result in an F

in the course (subject to appeal based on a finding of plagiarism) and possible additional sanctions by the University. The UNT Dean of Students requires official reporting of all instances of plagiarism.

All works submitted for credit must be original works created by the student uniquely for the class. Any exceptions to this (e.g., collaborative work or a variation of something submitted to another course) must be documented and approved by the instructor in advance. It is considered inappropriate and unethical, particularly at the graduate level, to make duplicate submissions of a single work for credit in multiple classes, unless specifically requested by the instructor. Work submitted at the graduate level is expected to demonstrate higherorder thinking skills and be of significantly higher quality than work produced at the undergraduate level. You are encouraged to become familiar with the university's Student Standards of Academic Integrity policy: <https://policy.unt.edu/policy/06-003>

Students with Disabilities

University of North Texas recognizes its responsibility for creating an institutional climate in which students with disabilities can thrive. In accordance with university policy, if you have a documented disability and require accommodations to obtain equal access in this course, please contact the instructor at the beginning of the semester or when given an assignment for which an accommodation is required. Students with disabilities must verify their eligibility through the Office of Student Disability Services.

Medical Emergencies

If you have a medical emergency that will prevent you from completing the course as scheduled, you **MUST** notify your instructor at the time of your medical problem. Requests for a grade of incomplete or to drop the course should come at the time of the medical emergency and must be supported by a letter from your physician indicating that your medical situation will **NOT** permit you to work at your computer in order to submit your assignments as scheduled.

Requests based on a medical emergency cannot be granted after the medical emergency except in the rarest cases when prior communication was not possible. All requests for a grade of incomplete must be initiated by the student and meet standard university policy as outlined in the Incomplete Grade Information section below.