CSE 5520 – Data Visualization and Communication

Fall 2021, Wed 5:00 – 7:30 pm

Instructors: **Seung-Hyun Hong**
ITEM Room 213, x3654
sshong@engr.uconn.edu

Dong-Guk Shin
ITEB Room 263, x2783
shin@engr.uconn.edu

Office Hours: **Mon 11:00 am - 12:30 pm (DGS)**
Tue 11:00 am - 12:30 pm (SHH)
or by appointment

Build the big picture and develop intuition!

CSE 5520 – Data Visualization and Communication

Data visualization is a form of communication utilized by a wide spectrum of scientists and practitioners to convey information to both novice or expert audiences. This course covers basic and advanced skill sets which help one learn how to transform massive amount of intricate data sets into patterns, graphs, plots and other visual metaphors to present trends, biases and changes. The course will use multiple real-world example data sets, starting with simple ones and will gradually move to complex ones. It will introduce two to three popular data visualization packages and teach how to use the packages to engage in hands-on data visualization case studies.

CSE 5520 – Data Visualization and Communication

No textbook

Internet Resources will be heavily used.

- Blogs
- YouTube
- Published articles

Data Sets

- Only publicly available datasets will be used.
- Directly from publicized packages.
- Uploaded at HuskyCT/CSE5520

Teaching modality

- Distance learning
- Exclusive use of HuskyCT
- QUIZ – Lockdown browser
- Office Hrs – HuskyCT/WebEx

Lecture Recording

- HuskyCT lecture will be recorded.

Student Presentation

- WebEx will be used without recording.

Reference book:
Visualizing Graph Data, Manning Publication, 2016

CSE 5520 – Data Visualization and Communication

Importance of Hands-On practices in learning Data Visualization methods

Phase I

Manual calculation using toy example.

Code the toy example.

Extend the toy example data set and visually inspect the process to develop good understanding of methods.

Phase II

Test/extend your code using a dataset from the real-world and develop intuition.

Repeat the process using multiple datasets from the real-world.

Phase III

Combine different visualization techniques to solve/predict trends for a dataset from a real-world.

Tools & Environment: Python (Jupyter lab or PyCharm)
R (RStudio)

Topics*

1. Basic statistical graphs

Mean, standard deviation, histogram, PDF, CDF

Various types of distributions

2. Test statistics & p-value; FDR

t-test, p-values, FDR and ROC

3. Random Sampling, various probability distribution and Monte Carlo method

Monte Carlo method visualization

4. Linear regression (slope) vs. Pearson correlation

Comparing scatter plots of two or more variables

5. Network graph/Network diagram

Showing relationships between points; Chord diagram, Arc diagram, Sankey diagram, etc.

Topics* (continued)

6. Markov Chain method

Computing stationary probabilities

7. Bayesian MCMC

Bayesian prediction and Bayesian statistical inference

8. Kernel Density Estimate/Gaussian Mixed Model

Converting discrete distribution to continuous distribution (2D and 3D)

KDE vs. Gaussian Mixed Models; Analyzing a cohort of samples

9. Bayesian Belief Network

Let them learn computational tools

10. Other visualization topics of data mining

Dimension reduction methods (PCA, t-SNE, UMAP)

Clustering and classification visualization methods

Visualizing natural language processing outcomes

*** Choice of topics will be adjusted as needed.**

CSE 5520 – Data Visualization and Communication

Specific goals for the course:

1. Students can work **as part of a team (two members)** on a significant project.
2. Students will **generate and articulate functional requirements** for a term project.
3. Students will be able to **contrast and critique efficacy of various data visualization methods they practiced in the term project through verbal and computer demo presentations.**
4. Each student will submit **a written report summarizing the project outcomes in a term paper.**

Micro Hands-On Lab

Weekly meeting will include hands-on real-time data visualization practices.

For this reason, each student should have a “data visualization ready” computing environment in every class.

→ Participation credit can be acquired.

→ WebEx can be used for face-to-face discussion.

CSE 5520 – Data Visualization and Communication

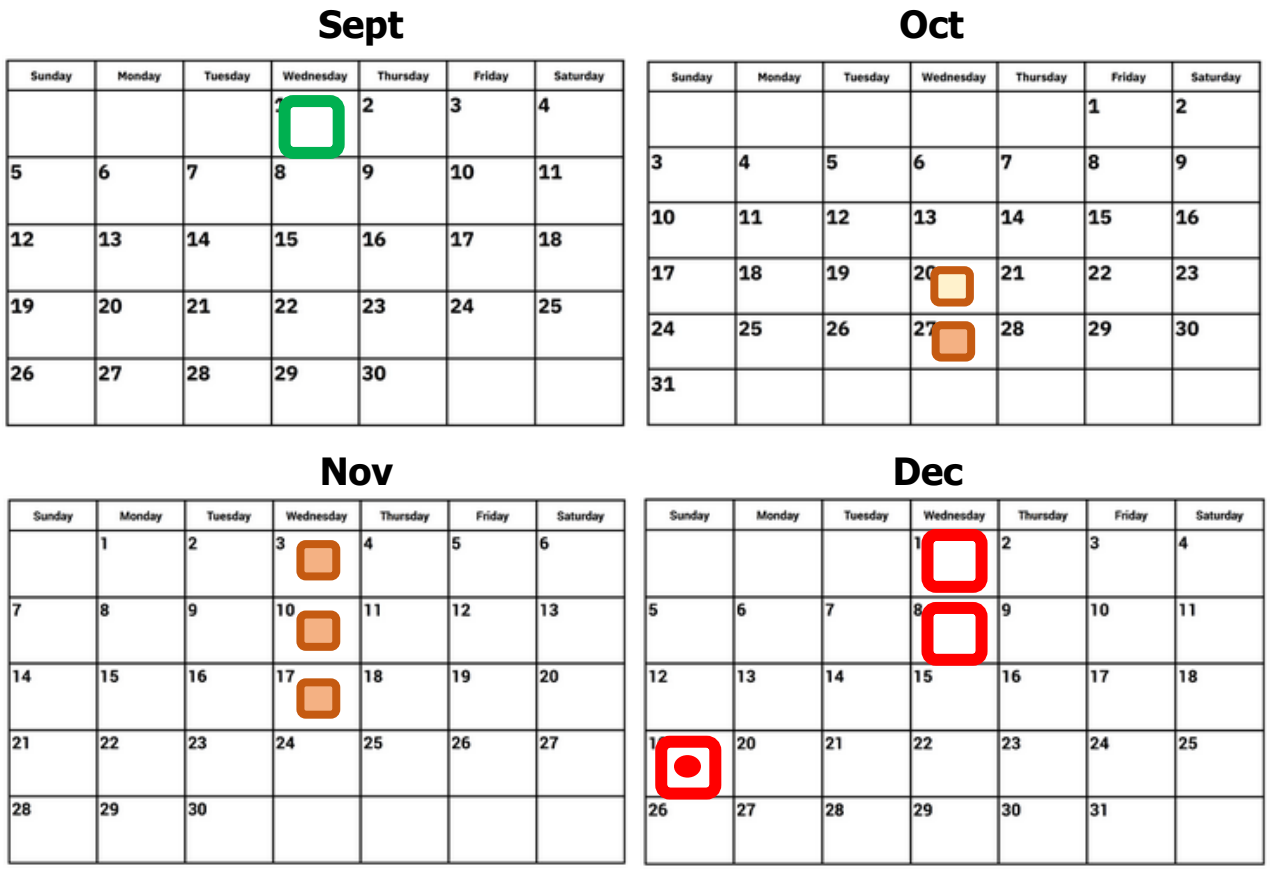
Grading Policy

Grading will be based on student's performance in five areas: participation (5%), homework (25%), quiz (15%), project (15%), in-class presentation (20%) and term paper (20%).

Schedule (14 weeks)

9/1 (Wed)	Fall semester begins.
10/20 (Wed)	Project proposal abstract due.
10/27 (Wed)	Project proposal begins*.
11/21-11/27	Thanksgiving Recess
12/1 (Wed)	Student Presentation
12/8 (Wed)	Student Presentation; Class ends.
12/19 (Sun)	Final exam ends; Term paper due.

* Specific date and time may change and the student presentation ordering will be determined FCFS. Team size should be 2. Term paper should be solo.



CSE 5520 – Data Visualization and Communication

HWs and Projects: 5% penalty for one day late submission.

Classroom NOs: cell phone usage, gossiping, food (drinks OK)

Academic Misconduct: Refer to Responsibilities of Community Life: The Student Code

21. “Student Conduct Officer”, “investigating Student Conduct officer”, or “investigator” means a University staff member who is authorized to investigate and determine the appropriate resolution of an alleged violation of The Student Code. Subject to the provisions in this code, this individual is vested with the authority to, among other duties, investigate an alleged violation of The Student Code; decline to pursue a complaint; refer identified disputants to mediation or other appropriate resources; establish The Student Code alleged violations regarding a respondent; approve a case resolution form; and impose sanctions or affect other remedies as appropriate.

CSE 5520 – Data Visualization and Communication

UConn Student Authentication Plan for Distance Learning

Faculty Member(s) Name(s): Sean Hong, Dong-Guk Shin
Course Name and Number: CSE5520
Session and year this plan first implemented: Fall, 2021
Today's Date: 9/1/2021

Student Authentication Methods: UConn is required to verify the identity of students who register in online or remote courses and to establish that students who register in these courses are the same students who participate in and complete the course activities and assessments and receive academic credit. Though students in transitioned courses may not have originally registered in an online or remote course, following the steps shown serves to provide reassurance that student verification and authentication has occurred.

CSE 5520 – Data Visualization and Communication

For this course, the following methods will be employed:

I. **Secure NetID and Log In (University Wide)**
HuskyCT, the learning management system in which all online courses are hosted authenticates student user accounts and passwords through NetID access. This process is already in place for all students enrolled in online courses at UConn and provides the first point of student authentication.

II. **Additional Authentication Method (Faculty Determined)**
In order to authenticate that the student completing the course activities/assessments and the student receiving credit for the course is the same student enrolled in the course, the following will be implemented (enter text below):

- Faculty member will check the StudentAdmin roster photographs at the beginning of the online teaching portion of the semester. Through monitoring of video chat, threaded discussions, and ongoing monitoring of multiple student submissions, faculty will be able to associate student work with individual students ensuring they are the same students receiving credit.
- Students will put a video on for all the synchronous classes and exams.