Course Title: Big Data Visual Analytics
Course No: CS661
Instructor's Name: Soumya Dutta
Email: soumyad@cse.iitk.ac.in
Credits: 3-0-0-0-9
Prerequisites: Basics of Linear Algebra, Statistics, and knowledge in Programming (C/C++, Python)
Who can take the course: PhD, Masters, 3rd and 4th year UG Students
Departments that may be interested: CSE, EE, SDS, ME
Class Timing: MW 2:00pm – 3:15pm
Venue: RM-101
Office Hours: By email appointment

Course Objective: The necessity of visual analytics capabilities for big data is becoming omnipresent due to its significant demand in the current age of data science and analytics. Interactive data visualization techniques enable us to comprehend and explore diverse types of complex data sets efficiently so that patterns and features from the data can be readily identified and studied in detail. As the data grows larger and become intricate, it poses significant challenges to manage, curate, and explore such large data sets. These data sets can come from various scientific simulations as well as from social media, IoT, various sensors, and many other industry and application domains. In this course, we will cover a comprehensive view of data visualization techniques with a focus on the techniques that are suitable for big data. We will discuss the theory and foundations of visualization techniques and have hands-on exercises on visualizing different types of data sets using available visualization software and libraries. We will study scientific and information visualization techniques with afocus on data compression, statistical and information theory techniques, and selected high-dimensional visualization algorithms. Next, we will discuss how modern machine learning and deep learning techniques are adapted for big data visual analytics. Finally, we will learn about Exascale computing and state-of-the-art in situ analysis techniques and conclude by discussing the future paradigms of the big data visual analytics domain. The contents for this course will be based on research papers from top-tier journals and conferences such as IEEE TVCG, CGF, ACM CHI, IEEE/ACM Supercomputing, IEEE Visualization, EuroVis and EuroGraphics, IEEE Pacific Visualization, IEEE LDAV, EGPGV etc.

Course Contents:

Index	Module	Topics Covered
1	Fundamentals of Data Visualization	Introduction to Visualization and Visual Analytics
		Foundations of Data Visualization, Visual Abstractions, Visual Variables, Various types of Data
2	Scientific Visualization (SciVis)	Big Data Characteristics, Data Reduction, Various Data Models; Visualization Pipeline
		Scientific Visualization Software such as VTK, ParaView, etc.
		Linear Interpolation; Isosurface Algorithm; Volume Rendering Algorithm
3	Information Visualization (InfoVis)	Fundamentals of Information Visualization, Software for Information Visualization
		High Dimensional Data Visualization Techniques
4		Big Data Analytics, Statistical Modeling

	Big Data Analysis and Visual Computing Techniques	Information Theory Techniques for Visualization
		Time-varying Data, Ensemble Data, and Uncertainty Visualization
5	Machine Learning for Visual Computing of Large Data	Machine Learning for Visual Computing and Visual Analytics
		Applications of Machine Learning to Big Data Visual Computing
		Visual Analytics and Explainability of Machine Learning Models
6	Advanced Topics	Extreme-scale Data Analytics, Parallel and High-Performance Visualization,
		Exascale Computing, Future Paradigms

Policies:

- Please be on time for the lectures.
- **Attendance will be taken in class from time to time and the marks will be awarded based on it.**
- You are expected to submit your assignments on time.
    - There is a 10% penalty each day after the submission deadline for up to 20% (2 late days). After that, you get zero. This policy will be strictly followed.
- **Students caught cheating or plagiarizing will be dealt with heavy punishment and could automatically fail the course and will be reported to the institute.**
    - **Please cite your sources properly in your work.**
    - **Your assignments should be your own original work.**
- If you are unwell, please follow the standard IITK procedure.

Noteworthy points:

1. We might add new, drop existing, or reorder topics depending on the progress and class feedback. Things may be changed by mutual consent after discussion in class.
2. Grading will be relative.
3. The recommended reading material is part of the course material.
4. If required, extra classes will also be conducted in weekends.

Evaluation:

Category	Split
Attendance	5%
Quiz	10%
Assignments	30%
Mid Semester Exam	25%
Final Semester Project	30%

Books, Reference, and Resources:

1. Data Visualization: Principles and Practice by Alexandru C. Telea, CRC Press.
2. Visualization Analysis and Design by Tamara Munzner, A K Peters Visualization Series, CRC Press.
3. The Visualization Handbook edited by Charles D. Hansen and Chris R. Johnson.
4. In Situ Visualization for Computational Science, Editors: Hank Childs, Janine C. Bennett, Christoph Garth, Springer publication.
5. Research papers provided during the class to cover selected topics.