Guidelines for CS661 course project

Project group size: 8 members
Deadline to finalize and submit your initial project proposal: March 7th 11:59pm

For the final course project, you and your teammates must pick up a topic of interest (a domain) on which you will work on. Once you pick up a topic of interest, you must ensure availability and accessibility of a dataset from that topic so that you can work on it. You must read relevant materials about this domain so that you have some idea about the domain. The dataset does not necessarily have to be extremely large, but it must be complex enough and containing enough information that can be analyzed and visualized.

- **Then you must identify a set of tasks (at least 5) relevant to that dataset** that you want to perform and visualize the data and extracted information. Often, just visualizing the raw data may not be sufficient, hence you may first analyze and process the data, derive new information and then use visual analytics to show the information to the users.

You are expected to build a visual analytics system (web-based/standalone) that consists of several (5-6) charts and plots with some form of interactive capabilities that will allow data selection, filtering, querying, etc. At the end, you should be telling a story about the dataset and the application domain and explain what problems you are solving using your visual analytics system.

- You can use any visualization library to build this system. Software such as PowerBI, Tableau are not allowed to use. You must write the code yourself with help of some libraries such Bokeh/Plotly/Matplotlib/Dash to build your analysis system from scratch. If you are comfortable with JavaScript-based libraries such as D3, you are welcome to use it. **If you are unsure if you can use some library, please talk to me.**

- If you are interested in working with data sets from scientific visualization domain, such as medical data sets, or simulation data sets, then you can consider building your system using VTK for rendering and using QT library for building user interaction interface. Both VTK and QT can be combined into a single interface in Python.

Kaggle is a great source of various types of data sets. So, you can go over the available data sets and pick a sufficiently complex one to use in your project. Other open resources for getting your data are also acceptable. For example, here are a few that I found interesting:

1. https://www.kaggle.com/datasets/manishabhatt22/tweets-onchatgpt-chatgpt
2. https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-india
3. https://www.kaggle.com/datasets/rajkumarpandey02/countries-by-carbon-dioxide-emissions
4. https://www.kaggle.com/datasets/luxcem/new-york-taxi-2022

At the end of the project, your team must submit a project report, describing the details of every step of your project, adding screenshots from your visualization system, and explaining the results that you have obtained. I will later share a LaTeX template for project report.

You should put all your codes into a GitHub repository, provide the link of that repository in your report, add a detailed instruction about running your code in a README.txt file along with all the libraries used.

You will submit the project report via HelloIITK and I will schedule a time (sometime during the final examination week) when you and your teammates will give a live demonstration of your visualization system and present your project outcomes for your final evaluation. **The time limit will be ~20 mins for each team.**

Please submit a ~2 page write up of your project proposal in pdf format by March 7th 11:59pm in HelloIITK.
You should mention project title, what application domain you have picked, what is the source of your data, what specific tasks you want to perform using your visual analytics system, how will the overall solution look like (a web-based system or a standalone visualization system, etc.), and given your team has n members, who will do what for this project. I am attaching a sample proposal from last year as a reference.

Note that individual members in a team will be evaluated and graded during the final presentation, so each member should contribute equally to the project and mention clearly what they have done in the report. Project marking will be assigned individually.

Some Inspirations/examples:

https://sciviscontest2018.org/
https://www.uni-kl.de/sciviscontest/#tasks
https://vast-challenge.github.io/2021/description.html

Some topics from previous groups who took this course:

1. Visualizing Rainfall Data and its correlation with temperature
2. Stories from the FIFA World Cups
3. Visualization of COVID-19 India Data
4. Seeing Beyond the Concrete: A Visual Journey into Rainfall and Air Pollution
5. Visualization and Analysis of Air Quality Data of India
6. Crop Production Data Visualization