<u>CS661A: Big Data Visual Analytics</u>
<u>Sample questions</u>

1. Explain the difference between Overview and Detail and Focus + Context based visualization design.
2. Can you think of a scenario when data visualization is not necessary and another scenario when data visualization can be extremely helpful?
3. What is Visual Analytics? What are the components of a visual analytics framework? lec1-52
4. What are visual variables? Which one of them is the strongest? lec2-10
5. Discuss three techniques that are used to denoise data sets after it is acquired.
6. Why do we need data reduction techniques? What are some standard data reduction techniques?
7. What is the basic difference between image processing, computer graphics, and visualization?
8. Discuss the fundamental steps in the Visualization pipeline.
9. In the scientific visualization domain, given a data set that has cube cells, what technique can be used to estimate data value at a given location when the location is not a grid point? Can you write a pseudocode of a function that will return the value at a given location inside a cube cell if data values are at the eight corner points of the cube cell?
10. Given 3D scalar data with cube cells, what algorithm is used to extract isocontours from it? Is there a limitation to this algorithm? If yes, then how is this limitation addressed in practice?
11. Why do we need transfer functions for volume rendering? How can gradient information in a transfer function design be helpful?
12. What is hybrid parallelism? Can you describe step by step how volume rendering can be done using hybrid parallelism?
13. Are there some risks that are involved when we use information visualization techniques to visualize large data?
14. What is a Box and Whisker plot? What information about the data is shown using it?
15. Discuss the limitations of PCA.
16. When interpreting the structure of high dimensional data from a t-SNE projection, what are the key points that we should keep in mind? What would be the best way to apply t-SNE to real-life high-dimensional data, and what precautions should be taken? When would you prefer UMAP over t-SNE?