CS732/DS732: Data Visualization | Assignment Booklet

Jaya Sreevalsan Nair jnair@iiitb.ac.in
IIIT Bangalore

Term I (2019-20) | Document updated on October 21, 2019[1]

Logistics for Assignments:

Assignment/Assessment	**Submission Deadline**	**Grading**
A1: 2D scalar/vector field visualization	11:59:59 pm, IST, Aug 31, '19 (extended to Sep 03, '19)	10%
A2: 3D scalar field visualization/2D vector field visualization	11:59:59 pm, IST, Sep 18, '19	20%
Midterm	Sep 25 - Oct 01, '19	10%
RWPA: Technical report	11: 59:59 pm, IST, Nov 02, '19	-
A3: Information visualization	11:59:59 pm, IST, Nov 09, '19	20%
A4: Visual analytics	11:59:59 pm, IST, Nov 30, '19	20%
RWPA: Presentation+technical report	11:59:59 pm, IST, Nov 30, '19	10%
Finals	Dec 02 - Dec 08, '19	10%

1. Demo for A1, A2, A3, A4 will be 5% each, and remaining grade will be for report/code/data analysis/etc.
2. For RWPA (Reading-writing-presentation assignment), 5% is for report and 5% for presentation.
3. Midterm/Finals are 3-hour long closed-book written examinations.

[1] This document was previously updated on August/08/2019, August/13/2019, August/14/2019, September/03/2019, September/26/019.

Submission Instructions:

A1,A2,A3,A4: A single .tar.gz file named as <RollNumber>_A*.tar.gz (n= 1, . . . , 4) which it decompresses to a folder named <RollNumber>_A*.

- Submission for a programming assignment would be a tarred-gzipped folder comprising of: a subfolder containing the source files (named "source"),
- a subfolder containing header files (named "include"),
- README (in .txt format),
- a summary file or a report (in .pdf format),
- a subfolder containing screenshots (named "images"),
- a Makefile, and
- an empty subfolder (named "bin").

Important points:

1. The Makefile shall create the executables in the bin folder.
2. All or some of the datasets used in the assignment may be optionally included in a subfolder named "data". However, care must be taken that the tarred-gzipped folder is within the data limit specified for submission through LMS. During the demonstration, the executables can be run with datasets not submitted in the "data" folder.
3. There shall be penalty proportional to non-compliance of instructions, e.g. for submissions containing intermediate files (e.g., *.o, *.C , etc.) or submissions without some of the aforementioned files (e.g., report) and folders (e.g. screenshots), uncropped screenshots (unless it is a screenshot of the Graphical User Interface itself).
4. README.txt shall contain information on sources referred to for help on the assignment, instructions on how to compile and run the application, expected input-outputs, and any notable defects/effects when running the application.
5. The summary file or report, in .pdf format, shall be the student's original document, consisting of the problem statement, the approach taken for the solution, experiments (if any), discussions, and conclusions.
6. Students are encouraged to walk through their report to the instructor during the assignment demonstration.
7. The report shall be prepared using LaTeX. This document shall also contain the answers to the questions posed for each assignment.
8. Assignment demonstrations shall be shown for the last submitted version of the source code of the programming assignment.

Reading-Writing-Presentation Assignment: A single file in .pdf format: <RollNumber>_RWPA.pdf, prepared in IEEE conference format using LaTeX style files, to be e-mailed to instructor. After the presentation, the report, as well as presentation, must be submitted together in a tarred-zipped folder, as <RollNumber>_RWPA.tar.gz

A1: 2D scalar/vector field visualization

Dataset: Structured 3D dataset at http://sciviscontest-staging.ieeevis.org/2004/data.html

Description: Consider scalar fields of cloud moisture mixing ratio, pressure, and temperature; and 2D vector field of (u,v), i.e. X and Y wind speeds. The volumetric data is in x, y, z coordinates, in 500x500x100 sized grids. We are interested in only studying the x-y planes (corresponding to longitude-latitude data). Hence, the first task is to aggregate data about the z-axis. Uniformly aggregate "n" layers in the z-axis, such that n is a manageable number for running an interactive application. Experiment values of n to be 5, 10, and 20. The aggregation operation can be averaging. Visualize the scalar fields and vector field for each (aggregated) x-y layer, looping through values of z. The scalar fields can be visualized using colour mapping, contour mapping, and height/elevation mapping, and vector field using the quiver/hedgehog plot.

Use color and geometry visual encodings judiciously. For color mapping, visualize using 1-colour, 2-colour, 3-colour, and rainbow colour spectrum.

Write in your report:

1. What are your inferences on the value of n? Did large values of n lead to loss of information? If yes, which variable(s) was/were affected?
2. What inferences can you make on the choice of your colour spectra? Did you find any artifacts introduced by the rainbow colour spectrum?
3. Comment on the visual encodings you have used for each of your visualizations. Would you have liked to try other alternative encodings as well?

Optional: try out other variables given in the dataset. Do you see any patterns? You could also combine multiple visual encodings with respect to multiple variables/fields in a single visualization, as discussed in class.

<u>A2: 3D scalar field/2D vector field visualization</u>

Dataset: The same dataset in A1.

Description: Demonstrate streamlines of 2D wind speed field (in u,v directions) in sea surface changing over 5 different time instances. Demonstrate slicing and isosurface extraction in the volumetric data for 3 different scalar fields, and show their changes over time using the same 5 different time instances.

Choice of time-instances: Generate a box-plot of pressure variable across time instances using pre-computed statistical descriptors[2] for each time instance (or uniformly subsampled time instances -- should have at least 12 time instances). Use this box plot to determine time instances for which you would generate visualizations.

Use color and geometry visual encodings judiciously. For color mapping, visualize using 1-colour, 2-colour, 3-colour, and rainbow colour spectrum.

Write in your report:

1. Other than using the boxplot, which other ways will you identify the time instances for further exploration? What other scalar fields could you have used for a similar study?
2. What is your conclusion on the evolution or timeline of Isabel Hurricane based on your visualization results? Is tracking the eye of the hurricane sufficient to manage disasters owing to the hurricanes?
3. How would you build a real-time hurricane tracking visualization application? What visualizations will you include and why?

Relevant reading material related to Hurricane Isabel and visualizations:

1. http://vis.computer.org/vis2004contest/vrvis/IsabelWithSimVis.pdf
2. https://earthobservatory.nasa.gov/images/12152/hurricane-isabel

[2] https://stackoverflow.com/questions/23655798/matplotlib-boxplot-using-precalculated-summary-statistics

A3: Information visualization

Datasets:

Multivariate: AutoMPG http://mlr.cs.umass.edu/ml/datasets/Auto+MPG
Network: Jazz Musicians http://konect.uni-koblenz.de/networks/arenas-jazz
Hierarchical: https://www.kaggle.com/fernandol/countries-of-the-world/download

Description: Use D3.js or Processing to implement treemap, sunburst, node-link diagram, matrix visualization, and parallel coordinates plot; along with appropriate user interactions. Use Gephi as a network visualization tool to analyze network data.

1. Visualize all the three datasets in its format, using visualizations specified above and routinely used for the dataset type.
2. Remodel one of the datasets to a different dataset type, and visualize the same dataset using a method apt for the new dataset type.

Write in your report:

1. Compare the visualizations of the raw and remodeled data. What semantics of the data did you exploit for the visualization?
2. How did you use Gephi for network visualization and analysis?
3. What are the user-interactions implemented for each visualization? Which, in your opinion, are user-interactions which enhance your visualization?

A4: Visual analytics

Dataset: Of your choice, preferably from publicly available dataset.
Description: Get your team, dataset, and analytical tasks on the data approved by the instructor.
Use all data analytical techniques, including visualization, as a requirement.

Reading-Writing-Presentation Assignment:

Choose and read a paper from the list of papers published on LMS. These papers are seminal in the area of visualization.

The technical report and presentation should articulate the background of the problem the paper solves, related work (if any), the methodology used, its impact on the computer graphics community, and the areas that the method affected.

The technical report shall be prepared using LaTeX.

The technical report shall be prepared using the template of an IEEE conference paper. The report shall include the following sections: Abstract, Introduction, Related work, Background of the assigned paper, Discussions, and Conclusions, and References. The technical report shall be original work, whose content includes the findings on the allocated paper, its influence on the visualization community, and its consequences in terms of its succeeding techniques, algorithms, and/or applications.

A class presentation of the report, if applicable, shall be submitted in .pdf format, in which case the submission for the assignment will be a tarred-gzipped folder containing the presentation and the report. If not, the submission will be the report alone, which shall be named <RollNumber>_RWPA.pdf.

Depending on the strength of the class, the instructor may perform a closed viva-voce for a larger class as opposed to the class presentation (for a smaller class). Appropriate instructions regarding the submission and viva-voce shall be given in the class.

The technical report is expected to be submitted <u>before</u> the scheduled class presentation or viva voce.