

DAS732: Data Visualization -- Course Evaluation Guide

Jaya Sreevalsan Nair
International Institute of Information Technology Bangalore
Oct 25, 2024¹

Grading scheme, as announced on August 1, 2024:

- 20% of the final grade for each of 3 assignments (A1, A2, A3)
 - Each of the assignments has a demo for 5%, which may be submitted as a video as per instructions
- 5% for reading-writing assignment (RWA) (or a take-home exam, depending on the course logistics)
- 15% for mid-term
- 15% for end-term
- 5% for class attendance (specifics to be announced during the course)

Submission instructions:

- Programming assignments and RWA: 65% of the final grade
 - It is compulsory to attempt all 3 assignments and the RWA.
 - The assessment of the programming assignment is based on deliverables, i.e. code+report+folder of images+demo for each assignment.
 - The report is where one can elaborate on the data used, hypothesis, analytical methodology, inferences, and responses to any questions posed in the assignment.
 - The report should have images of the visualizations generated, with a clear explanation of the choice of marks and channels, and your interpretation. The image captions should be sufficiently descriptive.
 - The code submissions must be source code in text format, say Python code in .py text file format, and not as Jupyter notebooks.
 - The video must demonstrate your rationale behind the choice of visualizations and its marks/channels, interactions, and other inferences.
 - For visualizations of time series datasets or the same data with structured changes in parameters, gif images may be generated from the series of frames. It is important to provide metadata in the plot on the time instance, variable type, etc.

¹ Earlier versions were on Aug 22 and Sep 25, 2024

- It is important to submit a folder of images as it is difficult to get a full gist of the images from the video or the report.
 - A README is required that gives instructions on how to run the code, the annotation for images in the folder, and other details related to the execution of your code.
- All submissions must be done on LMS.
 - If your entire submission is larger than the permissible size for LMS submissions, upload your submission on Outlook OneDrive; and submit a document containing the URL to the submission.
 - It is the onus of the student to ensure the correct access permissions are provided in the repository if the submission files reside outside of LMS so that there is no difficulty in accessing the files for assessment. In such cases, these repositories have to be accessible until the course grades are announced.
- The scheduling of assignments provides **2-3** weeks to complete each of the programming assignments [A1, A2, A3], and 2-4 weeks for RWA.
 - All assignment submissions must be done by Monday midnight IST, as per schedule.
 - All programming assignments are group projects with groups of 3 members. The TAs will help with the logistics of team creation, dataset selection for the team, etc.
 - Only one member of the team needs to submit on behalf of the team.
 - All team members are welcome to submit their individual contributions over and above the group submission.
 - Please mention in the comments section for each submission as to if it is for the group or individual contribution.
 - Assignment announcement date and submission deadline:
 - **A1: Aug 22, Sep 16;**
 - **Dataset announcement: Aug 23**
 - **Datataset finalization: Aug 25**
 - **A2: ~~Oct 14, Nov 04~~; Oct 24, Nov 11;**
 - **Dataset announcement: ~~Oct 16~~ Oct 19**
 - **Dataset finalization: ~~Oct 18~~ Oct 20**
 - **A3: ~~Nov 04, Dec 25~~; Oct 24, Dec 09;**
 - **Dataset additions finalization: Nov 08**
 - **RWA: Sep 16, Oct 21;**
 - **Paper announcement: Sep 16**
 - **Paper finalization: Sep 19**

- 2 written exams - 30% of the final grade
 - Proctored exams in person during mid-term and end-term weeks.
 - Open notes, only hand-written notes.
 - 1 report-writing - 5% of the final grade
 - This can be based on a research paper or a theme. It will be allocated on a first-come-first-serve basis.
 - The topics will be published on Sep 16, 2024.
 - The choice of topic/paper by the student must be communicated by Sep 18, 2024. If not received by the deadline, a randomly picked topic/paper by the instructor by Sep 19, 2024.
 - The report is due on Oct 21, 2024.
 - If it is based on a research paper, the report must say why the method is important, and its impact on the research community (using the papers that have cited the paper, a state-of-the-art paper that explains the value of the paper), etc.
 - The technical report is to be written in the IEEE conference paper format.
 - This also includes references to papers/articles/etc. by citing them appropriately in-place in the report. There will be negative points for not doing citation references within the article and bibliography properly.
 - A more detailed description of how to prepare the report will be posted at the time of the announcement of the papers.
-

A1: Visual Exploration Using Visual Analytics Tools

Date of announcement: August 22, 2024 (Thursday)

Date of submission: 11:59 pm IST, September 16, 2024 (Monday)

Summary: A group project on the visual exploration of a sufficiently complex dataset

Dataset: As chosen by the group from the list provided by the instructor and TAs

A1 tasks and requirements:

1. Read the selected dataset, process the same, and explore the data using visualization and optionally, other simple data analysis techniques, such as statistics.
2. Since it is visual exploration, multiple visualizations should collectively lead to knowledge discovery from the dataset.
 - a. A data story must preferably evolve from multiple visualizations, instead of presenting several visualizations.
3. The assignment implementation must start with a question that can be answered by a list of visualization tasks implemented on the dataset. Visualization tasks include exploration and summary, along with descriptive verbs such as "overview", "trends", "search", etc. (If you are interested more in understanding tasks, please refer to Schulz et al. [1]). Your solution should solve n sets of tasks for an n -member team.
 1. The visualizations may be generated on independent tools or an integrated tool, e.g., Tableau, PowerBI, etc. We recommend the use of Tableau.
 2. You are advised to not split/fragment a task across different team members, as much as possible, and also ensure that all team members have tasks that will enable all in the team to make equal contributions.
 3. You may create tasks based on cohesive subsets of variables in the dataset.
4. The report shall contain a detailed description of the tasks you have tried, and eventually implemented, who did what in the team, etc. Specifically, mention the contributions of each member.

A1 assessment:

1. Assessment is based on the report, images folder, and a video demo of the data stories.
 - a. The report should contain details on dataset description, tasks, visualizations, and inferences/conclusions. The report should have a section on Author Contributions where the contributions of the team members are explicitly mentioned.
 - b. Appropriate Python code or Tableau workbook shall be submitted. For the latter, appropriate access shall be managed for assessment.
 - c. The video demo should be crisp and at most 5 minutes long, where each member explains their task, visualization solutions, and inferences. The first minute can be used by the team leader or the data processing contributor to mention the preprocessing of the data.

- d. The visualizations are expected to be simple plots that are understandable by laymen.
- e. The folder of images must be organized properly, and all images in the folder must be featured in the report. The images in the folder may be named as Fig<number>.jpg/png/<image file extension> where the number is the index of the image in the report. You may add additional images that are not contained in the report, in which case, include a readme file to indicate the significance of such images.

Rubrics for assessment:

- Choice of visualizations - 5 points
- Inferences - 5 points
- Video demo (including use of slides, clarity of presentation, Tableau demo) - 7.5 points
- Report (including good writing and presentation, the inclusion of images, captions, rationale on the choice of visualizations, inferences) - 7.5 points
- Completeness of submission (report, code, images, README) - 5 points

References:

[1] Schulz, H. J., Nocke, T., Heitzler, M., & Schumann, H. (2013). A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2366-2375.

RWA: Reading-Writing Assignment

Date of announcement (of paper allocation): 11:59 pm IST, September 25, 2024 (Wednesday)

Deadline for paper selection: 11:59 pm IST, September 18, 2024 (Wednesday)

Paper finalization: September 19, 2024 (Thursday)

Date of submission: 11:59 pm IST, October 21, 2024 (Monday)

Summary: The assignment is reading a visualization research paper, understanding, and writing a report on the same.

Paper: As selected from a list published by the instructor/TA.

Assignment:

- Read the allocated paper thoroughly.
- Write a technical report using LaTeX (use overleaf to help your case) and IEEE conference proceeding 2-column format, with appropriate bibliography, citing references in-place and as required.
 - The report must have three parts: (i) why the method is important (what gaps did it fill), (ii) the proposed method, and (iii) its impact on the research community (using the papers that have cited the paper, a state-of-the-art paper that explains the value of the paper), etc.

Guidelines:

- Follow appropriate BibTeX to format the bibliography.
 - Use the citations given in BibTeX format in Google Scholar to have consistent referencing.
 - Use the standard sectioning used in Computer Science papers: abstract, introduction, related work including the gaps addressed by the chosen paper, methodology in the paper, and the impact of the paper (use Google Scholar to find around 2-3 influential papers citing the chosen paper) including state-of-the-art, and conclusions.
 - Refrain from using first-person narratives in the paper, refer to the authors of the paper. If there are up to two authors, refer to them by their last names, e.g., “Doe and Smith [21] have discussed ... “. In case there are 3+ authors, use the last name of the first author followed by the phrase “et al.”, e.g., “Doe et al. [42] have proposed ... “
 - In the case of paper review, the title of the paper must be A Critical Review of “<topic-title>”. In the case of the exposition of a topic, the title of the paper must be The State-of-the-Art in “<topic-title>”. You could come up with variants of these titles.

- Your report may contain figures and tables borrowed from other papers or created on your own. If borrowed, mention the source as “Image courtesy: XYZ et al. [32]” in the figure caption, and likewise for tables. For Tables, Equations, etc., regenerate the same in your article and avoid including screenshots of the same from the original papers.
- The figure caption goes below the figure, and the table caption goes above the table. All figures and tables must be referred to in your text content. e.g. “Figure 2 shows how” or “Table 2 gives a comprehensive analysis of ...”
- The technical report writing exercise is purely a reading-writing exercise, and no coding is expected. That said, if you would like to run small experiments to improve your paper, you are welcome.
- Run a spelling and grammar-checking tool(s) on your document and edit your report appropriately before submission.
- Do not have more than 8 contiguous words verbatim from any source you are citing, otherwise, it would be flagged as plagiarized material.

Rubrics for assessment:

- Editorial correctness (including grammar, citations, references, and spelling) - 4 points
- Completeness (including past and later paper references) - 4 points
- Propriety (no plagiarism, original content, etc.) - 4 points
- Content (intellectual contribution, effective summarization) - 4 points

Note on rubrics:

"original content" implies it should be the original work of the student without copy-paste from the paper -- look out for tables/figures/etc. -- Ideally, the captions of the figures must be rewritten in the context needed in the report by the student.

The screenshot of the image shouldn't include the caption and figure indexing given in the paper. The caption should clearly say that it is from the paper being reviewed -- should be written as "Image courtesy: [1]" or similar. Hence, the paper being reviewed should be included in the bibliography, and the report should say it is "a review of the work by ABC et al. [1]."

The propriety should also be that the review doesn't speak of the work in first person, e.g. "We propose ..." should be "ABC et al. have proposed ...", etc. Based on the degree of these errors, the points are to be given on a scale of 1 to 4.

A2: Visualization Methods for Data Types

Date of announcement: October 24, 2024 (Thursday)

Date of submission: 11:59 pm IST, November 11, 2024 (Monday)

Summary: A group project on scientific visualization (SciVis) and information visualization (InfoVis) method implementation on selected datasets

Dataset: As assigned by the instructor and TAs, and A1 datasets

A2 tasks and requirements:

1. Use Python to implement SciVis methods, namely, color mapping, contour mapping, and quiver plots (optionally, experiment with streamlines), and generate animation gifs using 5-10 time instances (i.e. at least 5 and at most 10 temporal samples) of sea surface data.
2. Use Gephi or Networkx to implement a node-link diagram of a sufficiently complex network and compare at least 3 graph layout algorithms.
3. Use Javascript to implement Infovis methods as browser visualizations, namely, the parallel coordinates plot with user interactions and treemap visualizations with at least 3 different experiments.

Description of SciVis dataset:

Data source: <https://www.climatologylab.org/gridmet.html>

The gridMET dataset provides daily meteorological data for a period of over 40 years. Each team will work on a contiguous three-month period (e.g., January-March 2020, or November 2019-January 2020) that the team has already selected. The time instances of the dataset can be sampled either uniformly or randomly to select 5-10 samples. The rationale behind choosing the time period and the time instances must be explained in the assignment report.

Steps for Downloading the gridMET dataset:

1. On the gridMET page (given as the data source above) on a browser, find the "Download" tab in the navigation menu at the top and click on it.
2. Scroll down the page, and under the Direct Downloads section, find the link labeled "Direct download of NetCDF files".
3. Click on this link to be redirected to the repository that contains the NetCDF files.
4. You should now see a list of .nc files. The files are organized by variable and year in the format: <variable abbreviation>_<year>.nc.
5. Use vs (Wind speed at 10 m) and th (Wind direction at 10 m) for the Quiver plot. Use other variables for contour mapping and heatmap or color mapping.

Guidelines and assessment:

1. Assessment is based on the report, images folder, and codebase, as applicable.
 - a. The report shall have images with appropriate captions, and the inferences in the text shall be written about these images.
 - b. The images in the folder may be titled using their figure indices in the report so that the high-resolution images are available for assessment. It is sometimes hard to see the relatively lower-resolution images included in the report.
 - c. The screenshots from Gephi and other GUIs shall be complete and of high resolution.
 - d. The codebase shall include a README for implementation.
 - e. A video demo is not required for this assignment. Hence, it is important to include sufficient inferences in the report.
 - f. The report shall be professionally prepared as done for RWA. Inclusion of appropriate references, as possible, is encouraged.
2. There are 3 tasks for each of the visualization domains, i.e., SciVis and InfoVis. Each student shall implement at least one SciVis and InfoVis visualization tasks. The combined outcome of the team can be proportional to the team's strength.
 - a. The report shall clearly state the contributions of each team member.
 - b. Apart from data cleaning and processing (as required), there cannot be joint ownership of the methods. If a team violates this rule, marks will be deducted accordingly.
3. The contour mapping shall be done using the Marching Squares algorithm and/or contour fill algorithm. Explain the rationale behind the choice in the report.
4. The color mapping shall include experimentation with different colormaps, e.g. continuous, discrete, logarithmic, etc., and different color palettes. As appropriate, you could comment about the use of sequential and diverging colormaps.
5. The color mapping shall comment about parametric mapping using global vs local maxima and minima, across different timesteps.
6. The quiver plot can experiment with one of the following:
 - a. Grid sampling experiments along with a comparison of magnitude-proportional and same-sized vectors.
 - b. Compare and contrast between the streamline method and quiver plots, and consider experiments with different seeding strategies for the streamline implementation.
7. Node-link diagram experiments shall include different graph layouts, and optionally, any desired node filtering implementation.
8. Parallel coordinate plot implementation is complete only with user interactions. Hence, either axes reordering or brushing or both shall be implemented.
9. Treemap experimentations shall include different remodels of tabular data to the tree data structure. Optionally, different spatial partitioning strategies can be experimented with.

Rubrics for assessment:

- Completeness - 4 points
 - All experiments as mentioned in the assignment description must be attempted.
 - The report, code, images, README, etc. must be complete.
 - Correctness - 4 points
 - Inferences must be correct with respect to the context of the assigned dataset, its inherent nature (e.g., space and time), etc.
 - Quality of Inferences - 8 points
 - The complexity of arriving at the inferences, effective use of visualizations, etc. will be considered.
 - Data Handling - 4 points
 - Preprocessing steps, any other data-specific transformations, etc. will be checked.
-

A3: Visual Analytics Workflow

Date of announcement: October 24, 2024 (Thursday)

Date of submission: 11:59 pm IST, December 09, 2024 (Monday)

Summary: A group project on visual analytics

Dataset: A1 dataset, supplemented by relevant datasets as would have been used in A2
(Additional datasets must be ratified by the TAs)

A3 tasks and requirements:

1. Identity requirement of data analysis on the dataset(s) using statistical analysis, classical data mining methods (unsupervised machine learning such as clustering, classification, etc.), supervised machine learning (clustering, classification, forecasting, etc.).
2. Generate a visual analytic workflow that: (a) involves visualization and data analysis, and (b) demonstrates a clear case of feedback loop.
 - a. There should be an inference made on the dataset leading to its transformation and a second iteration of visual analytics workflow (the diamond workflow by Keim et al.)
3. The implementation can use a mix of tools and programs for different processes in the workflow. Optionally, methods can be chosen in such a way that a program script could be written to implement the entire workflow.

A3 assessment:

1. Assessment is based on the report, images folder, and a video demo of the data stories, similar to the deliverables of A1.
 - a. The report should ideally refer to the visual exploration of the core dataset in A1. Hence, add the A1 report as an appendix for the sake of completion of A1.
 - b. The appendix will not be assessed as a part of A3.
 - c. The visual analytics workflow shall have minimal self-loops, and it is important to include its drawing in the report. The feedback loop can be alternatively represented by “unrolling” the workflow diagram.
2. This time, there will not be a live demo. All assessments will be made using the deliverables.
3. The deadline for submission is a hard one, and there will be no scope for late submissions. This is the administrative imposition, given the grading deadline. Hence, consider a few days before the deadline as a soft one and make submissions.

Rubrics for assessment:

30 points - to be scaled down to 20%

- Completeness (4 points) –
 - should have a visual analytics workflow description, rationale behind visualizations, and choice of supplementary data, if any
 - Mining/analysis methods, image folder, and video demo to be considered
- Correctness of inferences (4 points) –

- Inferences are checked for correctness with respect to the inherent nature of the data (e.g., related to time and space, etc.).
 - Data Handling – in addition to that done in A1 (4 points) –
 - The appropriateness of datasets used, and implementation of expected data preprocessing are considered.
 - Implementation of appropriate data analytics/statistical modeling is considered.
 - Quality of inferences from report and video demo (8 points) –
 - The complexity of arriving at the inferences and the effective use of visualizations are considered.
 - Overall quality of report (2 points) -
 - The effort in writing the report, formatting, layout of images, the correctness of references, language, and grammar, etc. are all considered.
 - Overall quality of the video (3 points) -
 - The effort in making the demo, e.g., the use of PowerPoint to aid the viewing of the video must be appreciated.
 - Visual analytics workflow (5 points) -
 - The diagram or drawing of the workflow, and the presence of visualization-analytics-feedback are considered.
 - The feedback loop-transformation in the workflow, choice of methodologies for both visualization and analytics, etc. are considered.
-