INM433 Visual Analytics (PRD1 A 2020/21)

Individual coursework

This **individual** coursework assesses your ability to **use a Visual Analytics approach** to **investigate some research questions** using a **suitable dataset** of your choice and your ability to **critically reflect** on this. The marks will be based on a written report which must be submitted via Moodle by **Sunday 10th January** at **23:59**.

Data

You can choose whichever datasets you like for your project as long they are complex enough (and that you have any required permissions to use them). As a guide, if you consider your data as one big table, we're looking for something like 150 row/cols x 10 cols/rows as an absolute minimum. This is simply to help increase the chances that there is something worth investigating in there. Combining more than one dataset is not necessary, but may improve the scope of what to investigate.

Tools

You can use **whichever software tools you like** for your project, as long as you use them using a visual analytics approach. This means that your interpretation of visual outputs must inform the computational analysis and computational outputs must inform the way you visualise and interpret the outputs. You may one or multiple tools. You may do it all in Python. You may use tightly-coupled visual analytic approaches (e.g. Orange, V-Analytics) or loose-coupled approaches such as file imports/exports from/to R, Python and Excel. You may even write bits of your own software, though marks will only be for the way the tools are applied.

Format and marking criteria

You must use this template.

The report length **must** be less than 4000 words and up to 10 figures in total, obeying the specific limitations given for each section. Exceeding the limit for any section by more than 5% will **decrease your mark**, because you are supposed to be able to express your ideas and describe your work succinctly. The figures must have appropriate size to be readable.

1. **Problem statement (5%; <250 words).** Describe the analysis problem you aim to solve. Specify the phenomenon you are studying (e.g., bank transactions, city traffic, social media, sport, climate, etc) and the question(s) you strive to answer. Briefly present the content and structure of the data you are going to analyse and explain why these data are suitable for answering your questions.
You'll be marked on the (A) choice of the questions (how interesting and non-trivial they are) and (B) suitability of the data.
2. **State of the art (15%; <500 words).** Find 2-3 papers published in respected journals or conferences that deal with similar data and problems and apply visualisations for solving the problems (see the groupwork description and below for ideas). Briefly present the specifics of the data they analysed, questions they addressed, and the approaches they applied. Discuss to what extent these approaches are applicable to your data and problem. Substantiate your judgements by comparing your problem setting with those from the surveyed papers, paying particular attention to any assumptions taken. Specify what you have learned from these papers that is useful for your study.
You'll be marked on the (A) choice of the papers to survey (relevance to your problem, use of visual analytics approaches), (B) quality of the summarised description of the problems and approaches, and (C) quality of discussing the applicability of the approaches and deriving useful lessons.
3. **Properties of the data (15%; <500 words; <=2 images).** Describe in more detail the dataset(s) you are using. How were the data collected? Describe the structure, i.e., the fields, their meanings, types of the values. In case of two or more datasets used, describe how they are linked, e.g., by references in one dataset to data items in another dataset. Specify the amount of the data and the coverage (territory in space, period in time, number of distinct objects, etc.) In case of spatial and/or temporal data, specify the precision and resolution.
Describe your process of investigating the quality of the data and the problems detected, such as erroneous values, missing values, outliers, peculiar distributions, gaps or non-uniformity in spatial and/or temporal coverage, noisiness, and biases. Include 1-2 figures with visualisations that helped you to explore the data quality. Tell how you are going to account for the data properties and problems in your analysis.
You'll be marked on the quality of the (A) investigation of the data properties and (B) deriving conclusions for conducting the further analysis.
4. **Analysis (45%).** The analysis section includes 3 subsections: approach, process, and results. The mark will be based on their joint content, but each subsection must be present.

specify where in the workflow human reasoning or judgement is required in order to proceed with the analysis. What information needs to be presented to the human analyst for this purpose? What visualisation method is suitable for presenting this information? It is expected that your problem is sufficiently complex and cannot be solved simply by applying some algorithm and interpreting its outcomes. It also should not be solvable simply by considering some visual representation. Essential is involvement of human reasoning in the analysis process, not only in making final interpretation and conclusions. The human reasoning needs to be supported by appropriate visual displays. However, the involvement of human reasoning needs to be justified, so that the human does not do the work that can be done by computers. Make sure that the human is doing what humans do best and the computational methods are doing what computational methods do best.

Apart from visual analysis, the analytical workflow may include data transformations and/or application of computational analysis methods. The use of computational methods is not strictly required. It is essential that results of some of the steps in the workflow are used in the further analysis or affect the further analysis. The possible types of results may be derived data, decisions of what method or transformation to apply next, or choice of appropriate parameter settings.

You'll be marked on the (A) appropriateness of the approach, (B) appropriateness of the activity assigned to the human analyst, and (C) interrelations between human cognition and computer operations. The mark does not depend on the complexity of the chosen methods.

- **Analysis Process (<1500 words, <=7 images).** Describe the process of your analysis, in which you apply the approach presented in the previous subsection. Give particular attention to the human side of the process, including interpretation of information displays, reasoning about what is seen, making decision concerning the ways to proceed with the analysis, and drawing conclusions that answer the questions of your study. The key thing is to show how the repeated use of visual and computational methods and/or data transformations helped refine your answers and informed subsequent analytical steps and how - ultimately - they generated the findings reported in the next section. Show the stopping condition - i.e. how do you know when you've reached a good solution? Justify your use of parameters and variables with your investigations or with insights from other studies. Illustrate the description of your analysis with images of the information displays that were essential for your reasoning, decisions, and conclusions.
  You'll be marked on the (A) correctness of the use of the visual and computational techniques, including justified selection of variables, parameters, and transformations, (B) logics of your reasoning and justification of the steps taken.
- **Analysis Results (<200 words, <=2 images).** Summarise the most important findings and use these to answer your research questions. Illustrate the answers with appropriate visuals. Discuss the implications of the findings for your chosen application/domain. If you cannot answer your research questions fully, say so and expand in the next section.
  You'll be marked on how convincing your findings are and how well you apply them to your research questions.

5. **Critical reflection (20%; <500 words).** Reflect on the chosen approach and the course of the analysis process. Discuss where and how your thinking was important for taking the further steps or for obtaining components of your answer. Discuss the role of the visual representations for your reasoning. If you did not find complete answers to your research questions, discuss the reasons, such as properties of the data or weaknesses of the techniques you used. Irrespectively of the results, try to imagine what could be done differently, e.g., what software functionality could make your analysis more efficient. Discuss the limitations of your approach (e.g., assumptions, particular requirements to data, disregard of some aspects, etc.) and its potential applicability to other data and problems. Present lessons you have learnt and give recommendations to other analysts having to deal with similar data.
  You'll be marked on your ability to (A) think critically of what you have done and draw useful lessons and (B) understand the limitations and generalisability of your data, research questions and methods.

6. **References.** Your report must include full references to the papers you reviewed in Section 2 (State of the art). It may also include references to other relevant materials, such as the source of the data, papers describing the algorithms you used, web sites of the software tools you used, and your own code (e.g., Python notebook). Put the list of references at the end of the report. The references are not included in the word count.

7. **Table of word counts**. At the end of the report, include a table with the word number in each section. To count the words, you may use the functions of MS Word or another text editor. The table content is not included in the word count.

See below for an **indicative** descriptions what we expect at different grade levels, but please use the marking criteria listed above.

distinction	85-100	A	A+	Outstanding	This impressive work is an excellent piece of visual analytics. It is extremely well-motivated, identifying excellent research questions that are suitable for being answered by visual analytics approaches using the datasets chosen. The analytical tasks and approaches are perfectly suited for answering the research questions in which visual and computation techniques fully inform each other. The analytical steps perfectly demonstrate how visual analytics techniques were used to arrive at the answer, including how the parameters used were chosen. The findings clearly follow from the analytical steps. The reflection shows a high degree of independent thought in its construction and a high degree of clarity and creativity. This report is of publishable quality.
	80-84		A	Excellent	
	75-79		A-	Very Good	This is a very good piece of visual analytics. It is well-motivated, identifying very good research questions that are suitable for being answered by visual analytics approaches using the datasets chosen. The analytical tasks and approaches are very well suited for answering the research questions in which visual and computation techniques inform each other. The analytical steps demonstrate how visual analytics techniques were used to arrive at the answer, including how the parameters used were chosen. The findings follow from the analytical steps. The reflection has attention to detail and clarity, but a more creative synthesis could have improved this report.
	70-74				
merit	67-69	B	B+	Good	This is a good piece of visual analytics. It is well-motivated, identifying good research questions that are suitable for being answered by visual analytics approaches using the datasets chosen, though there may be some limitations. The analytical tasks and approaches are able to help answer the research questions in which visual and computation techniques inform each other to some degree, but there may be more
	64-66		B		

	60-63		B-	though not completely. The reflection and discussion may be limited in places and there may be some omissions.
pass	57-59	C	C+	Satisfactory	This is an adequate piece of visual analytics. It identifies research questions that can answered with visual analytics with the datasets chosen, though there will be limitations. The analytical tasks and approaches can help answer the research questions the visual and computational methods may not inform each other very well and may not be entirely appropriate. The analytical steps indicate how the visual analytics techniques helped answer the research questions but some important details will be missing. The findings will follow from the analytical steps to some degree. The reflection and discussion will be limited and lack depth, comprehensiveness and clearness.
	54-56		C	
	50-53		C-	
fail	47-49	D	D+	Poor	This poor piece of visual analytics works to some degree but contains many omissions. The report is poorly structured and papers may not illustrate the approaches well. The report lacks clarity and critical reflection.
	44-46		D	
	40-43		D-	Very Poor	This is a problematic submission in general. Although some knowledge of basic concepts have been demonstrated, there is no clear structure. Or it is not a valid submission.
	20-40	E	E	
	0-20			

Some tips

- Follow the structure provided - it's in your interest to make it as easy for us to mark as possible!
- Don't forget that both positive and negative results are equally valid, as long as your data and research methods are appropriate and you understand the limitation of the methods.
- You're marked on your application of visual analytics to answering research questions. Marks will only be awarded in the context of data analysis and its interpretation. Credit will not be given for sophistication of coding, complexity of techniques, quantity of statistical methods or similar, unless it has direct analytical benefit.
- If you use a method like clustering, make sure you justify its use and the features you use to cluster.
- Please do use graphics, but only if directly referred to in the text. Don't expect the marker to interpret them for you.
- Late submissions will not be marked.

Some datasets

You're free you use any suitable dataset you like, including ones provided in other modules.

There's a useful list of good and free data sources here. There's annoying splash screen - click on "continue to site" in the top right.

In addition, we have provided some datasets that you might want to consider:

- London census
- London Tweet events (original; New Year, 2015)
- London Tweet events (9-16th February, 2015)
- London Tweet events (16-23th February, 2015)
- London Tweet times
- London Tweet residents vs tourists
- London Output Area Classification (OAC)
- Milan traffic events
- Milan traffic times
- U.S. Daily Temperature Anomalies 1964-2013
- UC Irvine Machine Learning Repository
- The first one million London bike hires [more information]
- Chicago bike sharing data, including some customer info and weather during trips
- https://www.bigdatanews.datasciencecentral.com/profiles/blogs/big-data-50-fascinating-and-free-data-sources-for-data
- European social survey
- Global terrorism database

Sources of literature

Your coursework report must include an overview of 2-3 papers presenting visual analytics approaches to analyzing data and answering questions similar to yours. You should choose papers that were published in respected journals or conferences, such as journals and conferences of IEEE and ACM (especially IEEE VIS conference and IEEE Transactions on Visualization and Computer Graphics), Computer Graphics Forum, Information Visualization journal. To find relevant papers, you can look into existing surveys focusing on different research and application areas of visualisation and visual analytics.

Here are some references to potentially useful surveys:
- https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.13212
- https://link.springer.com/article/10.1007%2Fs11390-013-1383-8
- https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.13210
- https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.13092
- https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.12931
- https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.13211
- http://dx.doi.org/10.1109/TITS.2017.2683539

- https://onlinelibrary.wiley.com/doi/full/10.1111/cgi.13217

Last modified: Thursday, 8 October 2020, 7:00 PM

◄ Module outline

Jump to...

Submission area: coursework ►

City, University of London is an independent member institution of the University of London. Established by Royal Charter in 1836, the University of London consists of 18 independent member institutions with outstanding global reputations and several prestigious central academic bodies and activities.

Contact us

+44 (0)20 7040 5060
Make an enquiry

Back to top ^

Legal

Help & Support

Accessibility