Assignment 2 Brief

DATA5002 2025T3
UNSW Sydney

Table of Contents

Version History

This document will be updated from time to time, particularly in response to student queries. Text added since the last version will be underlined.

- **16 October:** Initial version
- **29 October:** Clarification that marking standard will be the same regardless of group size
- **4 November:** Clarification for how “the story you are planning to tell” should be interpreted for dashboards

Summary of Information

Assessment title: Data Visualisation Project

Weighting: 30% of course mark

Type: A visualisation and 2 reports

Group work: Up to 5, or individual

Length: See below

Feedback Details: Feedback will be given in the form of detailed rubric marks. Additional feedback will be provided upon request.

Aligned CLOs: 1–5

Due Date:

- **21 November (Friday of Week 10):** Final submission, 11:59 PM

Suggested Timeline:

- **End of Week 6:** Join a group for the project; group selection will be open until the middle of Week 7.
- **Weeks 6–7:** Work on the proposals.

- **Friday of Week 7:** Try to submit your first draft proposal by this date.
    - Note that this is not a hard deadline for the proposal stage, but the longer you take to get yours approved, the less time you will have to work on the visualisation.
- **Weeks 7–8:** Incomplete or insufficient proposals sent back and iterated.
- **Friday of Week 8:** Your proposal should be approved.
- **Weeks 8–10:** Work on the visualisation.

Rationale

In this assignment, you will apply the principles and the techniques introduced in this course to create a sophisticated interactive dashboard or narrative infographic.

Task Description

In this task, you will create a sophisticated data storytelling narrative and/or an interactive dashboard.

You will select a real-world dataset and develop either a narrative infographic using data storytelling principles in the form of a web page or similar medium; an interactive dashboard which enables the user to explore and understand the data; or anything in between.

Some suggested data sources are provided below.

The task will comprise two stages:

1. the proposal, in which you will select a suitable project and receive approval; and
2. the final submission.

Proposal

The proposal stage comprises the following steps:

1. Select your theme and dataset(s). Some data sources are provided below.
2. Explore them sufficiently to ascertain that the task is feasible. (Don’t underestimate the time it takes to clean and understand data!)
3. Submit a short (less than 1,000 words, plus wireframe) proposal describing
    - the dataset (very briefly),
    - what you are planning to do,
    - the story you are planning to tell (for infographics),
    - how the user might use the interface, both in the sense of how and when they will interact with it and what they might learn from it (for dashboards),
    - your target audience, including 2 complete personas,
    - a list of similar visualisations of same or similar data that already exist, if any (doesn’t count towards the word limit),

        - **Note:** It is your responsibility to make sure that this list is as thorough as possible. If it is later found that a visualisation similar to yours can be easily found online, particularly if the source code for it is available, this may be considered to be academic misconduct.
    - the medium in which you plan to develop it (e.g., a web page, a Shiny app, etc.),
    - a wireframe diagram for your visualisation or dashboard,
    - acknowledgements of assistance received, including help and advice, programming or graphic design, and an overview of how, if at all, you used generative AI. (Reference information does not count towards the word limit.)

We will do our best to provide feedback to your proposal within 72 hours (3 days) of your submission. The feedback will comprise the following:

- a Yes/No mark; proposals marked "No" will automatically reopen for resubmission, and there is no limit on the number of attempts; and
- if the mark is "No", a brief explanation for why the proposal has been rejected; common reasons include:
    - the project is too simple;
    - the project is too similar to something that already exists, particularly with published code;
    - the proposal is missing required components, such as a wireframe diagram or personas.

Note that a proposal being accepted means only that it meets the minimal requirements. It does not mean that all components will receive full marks, nor does it mean that executing it will receive full marks, merely that the level is minimally acceptable for passing.

After your proposal is approved, you will be asked to upload a final version for marking alongside the final submission.

Final submission

The medium must be appropriate for the project, and further instructions will be provided as to the form of the final submission. At minimum, it should include:

- A URL for the infographic or the dashboard itself. It does not have to be public, but it needs to be accessible to the marker. Techniques for hosting interactive visualisations will be discussed in class.
- Runnable R code and data to reproduce the data visualisations used and/or to create the dashboard.
- Any additional instructions needed to do so.
- A short report (500 words at most) with the following components:
    - a brief reflection on major challenges encountered in implementing the proposal, how what was accomplished deviates from the proposal, and why;
    - a brief reflection on what could be improved given more time and resources;

    - acknowledgements of assistance received, including help and advice, programming or graphic design, and an overview of how, if at all, you used generative AI. (Reference information does not count towards the word limit.)

Additional submission instructions will be provided closer to the due date.

Marking information

The proposal, the visualisation, and the final report will be assessed together.

All submissions will be marked to the same (high) standard, regardless of whether one is by an individual or a group and regardless of the size of the group.

Assessment criteria

- Proposal:
    - Quality and details of the wireframe.
        - Note that the final product does not have to match the wireframe exactly.
    - Quality and suitability of the personas.
    - Reflection on the proposal.
- Analysis:
    - Level of understanding of the data and its context.
    - Accuracy of data representation (i.e., not being misleading).
    - Methodological sophistication.
    - Reflection on the implementation.
- Storytelling (for infographics):
    - Application of storytelling principles and narrative.
    - Suitability for the target audience.
- Interactivity (for dashboards):
    - Sophistication and power of the interface.
    - Usability by the target audience.
- Design and Polish:
    - Application of design principles.
    - Simplicity and usability.
    - Aesthetic choices.
    - Professionalism (e.g., spelling, grammar, and capitalisation, robustness to unexpected user input).
    - Layout and accessibility (e.g., relative sizes of panels, fonts, etc.; colour schemes).
    - Acknowledgement, attribution, and referencing.

Teamwork

- Each team member will be required to assess all team members' teamwork skill and contribution at the time of submitting the project. The teamwork scores will be used by Moodle to determine team members individual marks from the team's common project report mark; see https://webpaproject.lboro.ac.uk/WorkedExample for the details of the individual mark calculation algorithm.
- Failure to complete and submit your own evaluation of your group members' individual contributions to your group works by the required time will unnecessarily disadvantage yourself.

Academic Integrity

Permitted resources

You may work on this assignment individually or in groups of up to 5. Note that the number of students permitted to submit individually may be limited.

Outside of the following exceptions, you may not receive help on this assessment from anyone except for the instructor, and you must not communicate this assessment to anyone.

- You may solicit and receive, in person or electronically, general help such as "How do I align legends of two ggplot figures so that their titles are at the same height?" or "I get this error when I run this code. How do I fix it?" If so, provide only minimal working examples (MWE) rather than project drafts. Any such help must be acknowledged in your submission.
- You are permitted, and encouraged, to solicit and share interesting data sources.

If in doubt, ask the instructor first.

Generative AI

For this assessment, the *baseline* Generative AI policy is the **Assistance with Attribution** policy template:

> This assessment requires you to write/create a first iteration of your submission yourself. You are then permitted to use generative AI tools, software or services to improve your submission in the ways set out below.
>
> Any output of generative AI tools, software or services that is used within your assessment must be attributed with full referencing.
>
> If outputs of generative AI tools, software or services form part of your submission and are not appropriately attributed, your Convenor will determine whether the omission is significant. If so, you may be asked to explain your submission. If you are unable to satisfactorily demonstrate your understanding of your submission you may be referred to UNSW Conduct and Integrity Office for investigation for academic misconduct and possible penalties.

Over and above that, the following recommendations apply to the two components of the project:

Proposal and Final Report: You are encouraged to limit your use of generative AI to a search engine (research) and an editing aid for the initial proposal and the final report. You may use

generative AI for generating graphical assets, including the picture of the personas, but **not** the wireframe diagram.

Implementation: There is no restriction on the use of generative AI for implementing your visualisation, whether in data management, asset generation, or writing software code.

To the extent possible, preserve your generative AI interaction logs, as you may be asked to provide an overview of how you used generative AI and, if any questions arise, details.

If in doubt, ask the instructor first.

Declaration

The University regards plagiarism as a form of academic misconduct, and has very strict rules regarding plagiarism. See UNSW policies, penalties, and information to help you avoid plagiarism, as well as the guidelines in the online ELISE tutorials for all new UNSW students. Note, in particular, the policies on Contract Cheating: sharing your assignment, and, in general, publicly or privately soliciting or obtaining help with it in ways not expressly authorised by the instructor is not permitted and will be investigated and punished. The information can be found at the following links:

- https://subjectguides.library.unsw.edu.au/elise/check
- https://www.student.unsw.edu.au/plagiarism
- https://www.unsw.edu.au/assurance-integrity/conduct-integrity/conduct-unsw/cheating

By submitting this assessment item, you declare that this assessment item is your own work, except where acknowledged, and acknowledge that the assessor of this item may, for the purpose of assessing this item:

- Reproduce this assessment item and provide a copy to another member of the University; and/or
- Communicate a copy of this assessment item to a plagiarism checking service (which may then retain a copy of the assessment item on its database for the purpose of future plagiarism checking).

You also certify that you have read and understood the University Rules with respect to Student Conduct.

Suggested data sources

Note that you can use other sources as well.

Global and Government Data

- **Australian Government Open Data:** https://data.gov.au
  Datasets from Australian government agencies covering health, environment, education, and infrastructure.
- **NSW Government Open Data Portal:** https://data.nsw.gov.au
  Access to datasets from NSW government departments including local council and public services.
- **NSW Transport Open Data:** https://opendata.transport.nsw.gov.au
  Real-time and historical data on public transport and traffic in NSW.

- **United Nations Statistics Division:** https://unstats.un.org/home
  Global development indicators including population, gender, and sustainability.
- **World Bank Open Data:** https://data.worldbank.org
  Economic and development data from countries worldwide.
- **WHO Global Health Observatory:** https://www.who.int/data/gho
  Health statistics including disease burden and healthcare access.
- **Global Forest Watch:** https://www.globalforestwatch.org
  Interactive maps and datasets on deforestation and conservation.
- **Zenodo Research Repository:** https://zenodo.org
  Scientific research data and publications across disciplines.
- **OpenStreetMap Data Extracts:** https://download.geofabrik.de
  Geospatial data for mapping and urban analysis.
- **Data.World Community Datasets:** https://data.world/datasets/world
  Collaborative platform for sharing datasets across domains.

Education and Learning

- **NCSU Teaching and Learning Datasets:** https://www.lib.ncsu.edu/formats/teaching-and-learning-datasets
  Curated datasets for instructional use in data science and statistics.
- **University of Rochester Data and Stats Guide:** https://libguides.lib.rochester.edu/data-stats
  Guide to finding datasets for academic research and teaching.
- **Open Syllabus Project:** https://opensyllabus.org
  Aggregates university syllabi to analyse trends in course content.
- **World Bank EdStats:** https://datatopics.worldbank.org/education
  Global education statistics including enrollment and literacy.

Science, Health, and Environment

- **NSW Health Statistics:** https://www.healthstats.nsw.gov.au
  Health indicators for NSW including disease prevalence and hospital use.
- **Ecological Data Sources:** https://www.rforecology.com/post/top-five-ish-sources-of-ecological-data/
  Datasets for biodiversity, climate, and conservation research.
- **PubChem Chemical Database:** https://pubchem.ncbi.nlm.nih.gov
  Information on chemical compounds and bioactivity.
- **DataHub Collections:** https://datahub.io/collections
  Aggregated datasets across domains including finance and climate.
- **FiveThirtyEight Datasets:** https://data.fivethirtyeight.com
  Datasets used in articles on politics, sports, and economics.
- **KDnuggets Big Data Sources:** https://www.kdnuggets.com/2017/12/big-data-free-sources.html
  Free big data sources for analytics and machine learning.

Business, Finance, and Economics

- **GlobalData Market Intelligence:** https://www.globaldata.com
  Commercial datasets on industries and market trends.
- **Quandl Financial and Economic Data:** https://www.quandl.com
  Financial and alternative datasets for investment and research.
- **OpenCorporates Company Registry:** https://opencorporates.com
  Open database of companies including registration and ownership.
- **Crunchbase Open Data:** https://data.crunchbase.com
  Startup and tech company data including funding and growth.

Culture, Media, and Entertainment

- **BFI Film Industry Insights:** https://www.bfi.org.uk/industry-data-insights
  UK film industry data including box office and production.

- **Movie Datasets for ML Projects:** https://imerit.net/blog/13-best-movie-data-sets-for-machine-learning-projects-all-pbm/
  Movie-related datasets for machine learning and recommendation systems.
- **Reddit Datasets Community:** https://www.reddit.com/r/datasets/
  Subreddit for sharing and requesting datasets across domains.
- **SteamSpy Game Analytics:** https://steamspy.com
  Data on video game popularity and player engagement.
- **IMDb Datasets Interface:** https://www.imdb.com/interfaces/
  Structured data including movies, ratings, and cast information.

Technology and Research

- **UCI Machine Learning Repository:** https://archive.ics.uci.edu/ml/datasets
  Repository of datasets for machine learning and modelling.
- **StatSci Dataset Archive:** http://www.statsci.org/datasets.html
  Datasets for statistical analysis and teaching.
- **Tableau Public Data Sets:** https://www.tableau.com/learn/articles/free-public-data-sets
  Curated datasets for practising data visualisation.
- **GDELT Global News Database:** https://www.gdeltproject.org
  Tracks global news events and sentiment analysis.
- **Media Cloud Analysis Platform:** https://mediacloud.org
  Platform for studying media coverage and framing.