

Course Outline

Course Code : DOS4811
Course Title : Data Visualisation
Class Date : From 15/1/2024 To 19/4/2024
Semester : Semester 2, Academic Year 2023/2024
Faculty : Jeremy Chen, PhD
Department : Analytics & Operations
Email : bizcwj@nus.edu.sg

Overview

Visualisation is an invaluable tool that supports decision making in business. Students will: (i) manipulate relational data sets, aggregate the data and generate visual representations; (ii) build an understanding of data aggregation processes; (iii) learn to use interactivity to support data exploration and counterfactual/what-if analysis; (iv) learn how to tackle realistic data-driven problems and increase confidence of correctness; and (v) learn aspects of communicating facts and messages with data. To succeed in this course, one should actively experiment in a hands-on manner (“create”), participate by sharing what one has created (“show”), and then engage with the feedback (“discuss”).

Before registering, please review text in red in more complete outline appended at the end of this document. More details may be found there.

Course Objectives

The objectives of the course align with its assessment components (below). Respectively, and beginning from the mid-term:

- Being able to clearly communicate the visual encoding of data, as well as the aggregation (summarisation) of data; and conversely being able to act correctly on clear descriptions of the aforementioned
- Being able to, in quantitative problems involving a large amount of data and computations, build confidence in the correctness of one’s solution (because in a professional environment, mistakes can make one liable for negligent misrepresentation)
- Being able to articulate processes for going from data to decision. In particular, going beyond the one-time review of data and making a single set of recommendations (“standard project practice”) to codifying decision making processes and facilitating them using a “data product”, thus supporting decision making even when the “consultants leave and the data evolves”

Details available in more complete outline appended at the end of this document.

Assessment

Class Participation	20%
Mid-term	30%
Individual Assignment	20%
Group Project	30%

Details available in more complete outline appended at the end of this document.

Schedule and Outline

Details available in more complete outline appended at the end of this document.

General Guide & Reading

See more complete outline appended at the end of this document.

Academic Honesty & Plagiarism

Academic integrity and honesty is essential for the pursuit and acquisition of knowledge. The University and School expect every student to uphold academic integrity & honesty at all times. Academic dishonesty is any misrepresentation with the intent to deceive, or failure to acknowledge the source, or falsification of information, or inaccuracy of statements, or cheating at examinations/tests, or inappropriate use of resources.

Plagiarism is ‘the practice of taking someone else's work or ideas and passing them off as one's own' (The New Oxford Dictionary of English). The University and School will not condone plagiarism. Students should adopt this rule - You have the obligation to make clear to the assessor which is your own work, and which is the work of others. Otherwise, your assessor is entitled to assume that everything being presented for assessment is being presented as entirely your own work. This is a minimum standard. In case of any doubts, you should consult your instructor.

Additional guidance is available at:

- [Administrative Policies](#)
- <http://www.nus.edu.sg/registrar/administrative-policies-procedures/acceptance-record#NUSCodeofStudentConduct>
- <http://nus.edu.sg/osa/resources/code-of-student-conduct>

NATIONAL UNIVERSITY OF SINGAPORE
School of Business
Department of Analytics and Operations

DOS4811 Data Visualisation

Instructor: Jeremy Chen
Office: BIZ 2, #02-21
e-mail: bizcwj@nus.edu.sg
Session: Academic Year 2023/2024 Semester 2

(Correct as at: 1 Dec 2023)

Course Description

Visualisation is an invaluable tool that supports decision making in business. Students will: (i) manipulate relational data sets, aggregate the data and generate visual representations; (ii) build an understanding of data aggregation processes; (iii) learn to use interactivity to support data exploration and counterfactual/what-if analysis; (iv) learn how to tackle realistic data-driven problems and increase confidence of correctness; and (v) learn aspects of communicating facts and messages with data. To succeed in this course, one should actively experiment in a hands-on manner (“create”), participate by sharing what one has created (“show”), and then engage with the feedback (“discuss”).

Course Requirements

Students should have familiarity with a spreadsheet application such as Excel at the level of “pivot tables” (“group-by” and “aggregation” operations). In particular, students should know how to generate one from tabular data by manual computation.

Please Read the Outline

Read all text in red before registering. In particular, the information on mid-term scheduling. In addition, please make an effort to consult the Outline prior to making queries on course matters. Review the content on class participation before classes begin.

Sections and Scheduling

Sectional Workshops: **Tue (Venue TBC) & Thu (Venue TBC), 6.30pm – 9.30pm**

Office Hours / Consultation Slots *may* be held from Week 4 virtually at **6pm (45 min slots) on Mon, Wed, and/or Fri** (on a rotating basis). See Canvas for details.

Be sure to read the announcements for updates.

Software and Technology

Bring laptops to class. While there is a small theoretical component, this course is dominantly hands-on. We will use Tableau, a *cross-platform* Business Intelligence software tool, to explore data, create interactive visualisations and develop data-driven presentations. In addition, we will touch on using Tableau Prep to prepare data. You may download them here:

- <https://www.tableau.com/products/desktop/download>
- <https://www.tableau.com/products/prep/download>

The instructional team will strive to keep the content on Tableau Desktop up to date, incorporating relevant and compelling features from the latest releases.

Arrangements will be made with Tableau to obtain licences in bulk. Students may make use of the default 14-day trial in the interim. Outside of the course, students may obtain 1 year student licences via the website. Student status letters that indicate the duration of enrolment will be required as supporting material. To obtain such a document, follow the instructions at <http://nus.edu.sg/osa/services/student-status-letter.html>.

Canvas Discussions will be used for class participation (e.g.: sharing screenshots for discussion), information dissemination for items that do not rise to the level of a “course announcement”, team formation/coordination, and general Q&A. Please post all questions on administration in the General Administration discussion, but be sure to check the outline first. (As a matter of general policy, when posed a question of general interest non-publicly, such as over e-mail or direct message, course staff will direct the requester to “post it in Canvas”.)

An additional persistence layer will be added through shared Google Documents that all may edit. This will be introduced later. If you ask a question in-class and it’s answered, you may be asked to record your question and your understanding of the resolution. This supports retention.

Assessment Breakdown

Class Participation	20%
Mid-term	30%
Individual Assignment	20%
Group Project	30%

What follows is sufficiently detailed guidance for the assessment components. This is a course for senior students and professionals. To some extent, attendees will be expected to interpret the requirements.

Class Participation (20%)

This component of assessment will be partly based on the instructor’s subjective evaluation of the “quality and quantity” of in-class contributions and also on “out-of-class contributions” to the learning of others.

For something you post to be eligible for class participation credit, please include, within your post, your name and the last 4 digits and letter of your *matriculation number* in parentheses. (e.g.: “Ken Chan (1234X)”) This is crucial. See examples in **Introductions** below. *It is your responsibility to ensure this is entered correctly. (Participation scores will arise from a course-staff maintained document, which feeds a script that computes participation scores.)*

Introductions. Each student should, by the end of the second week of the semester:

- For example:
 - Hi, I'm John Li (1234X). I spent the last 4 years in maritime shipping...
 - Greetings everyone, my name is... and that is how I discovered a lifelong love for mathematical optimization and its applications in logistics. (1234X)
- Post a selfie with an introduction in your section channel. Touch on your background, your interests, what you want to get out of the course, which section you are in, and whether you “already have a group”
- Copy this introduction to in Canvas Discussions (Introductions), and the shared documents for Introductions (links will be provided in Canvas Discussions)

Content Sharing. Students should keep a lookout for interesting examples of visualisation in their work or in media, post examples in the respective topical discussion in Canvas and briefly share your thoughts (in a few sentences) in the context of what was covered in the course. Subsequently, to be credited, starting from week 3 (i.e.: 8 to 9 weeks), there will be time for 5 to 8 students to talk about an item that was shared (for a minute or so).

Q&A. Students should ask questions in the relevant Content discussion in Canvas Discussions and participation credit *may* be given to those who help their classmates (and indicate who they are via the relevant parts of their matriculation numbers as above).

As a matter of policy and for purposes of capacity management, questions like “have I been credited” will not be entertained. Neither will questions of the precise breakdown of participation. Please expect course staff to act with integrity, be fair, and act in the general interest of everyone. (Typically, a monotone concave function will be applied to the raw participation scores to generate the final participation scores. This provides a mild moderating effect. But that cannot overcome the impact of non-participation.)

Mid-term (30%)

It will be a closed-book written exam on Visual Encoding of Data and Data Aggregation. Students may use a 2-sided A4 (handwritten/printed) “personal summary sheet”.

It will be held on 20 Feb 2024 (Tue; 6.30pm – 8.00pm) in Week 6 to avoid the midterm rush. The venue(s) will be announced once confirmed. Please make arrangements to be available in this time slot. Do not expect alternative arrangements.

Individual Assignment (20%)

There will be a single “open-everything” electronic assignment executed via ExamSoft. All the required information to answer the questions are provided. As such, most clarification queries will be responded to with: “You should read and interpret the question accordingly.”

For all quantitative problems in the assignment, a core part of the learning experience is checking your work. This reflects necessary practice when working with data professionally. In all such problems, you should validate your methodology.

To validate your methodology, it often helps to start with a simplified data set where the answer to the same question is clear. In addition, “parallel approaches” (e.g.: using Python/Excel) have been effective validation tools in practice. While “analytics” is not yet a profession, which comes with licensing and legal liability for mistakes, the purpose of this is to give students a taste of the essence of working with data like a professional.

Additional Guidance:

- Part of the point of the exercise is using the above recommendations to **do due diligence** and build confidence in your answer.
- You are advised to record your answers if you do the assignment in multiple sittings. ExamSoft may reset your work if you reload it.
- For most numerical response questions, a range will be accepted.
- For “fill-in-the-blank” questions, there may be a set of acceptable answers.
 - For example, if “time and money” is the answer, variations like “money & time” will be accepted.
 - After automatic grading, non-accepted answers may be reviewed and appropriate responses will be added to the set of acceptable answers.
- There are no “long answers” for “fill-in-the-blank” questions. Do answer concisely.
- When answering with formulas, check for correctness, and remove excess spaces.

Due: Week 13 (last week of term) on Monday at 11.30pm with no extensions
(It is *very strongly* recommended that you complete this by Week 11.)

Group Project (30%)

Project groups will be responsible to seek out a narrowly defined issue with an accompanying data set, apply business analytics solution methodologies (as appropriate), present the situation and findings.

All team members should be in the same section, and groups may have up to 5 members. (Special dispensation may be given on a case-by-case basis.)

The data set need not be “big”, but visualisation **must** play a crucial role. While real problems from businesses/non-profits/charities/etc. are more desirable, students are welcome to create simulated problems with “synthetic data sets”.

Note that your projects **must** be oriented around an organisation's priorities and the **decisions to be made**. Your visuals should facilitate making those decisions such as by providing clarifying information (“insights”), *or* even recommending one outright. Credit will be given for appropriate use of counterfactual analysis and analytics tools.

A guiding tenet for the Group Project is that “**groups should make themselves useful**”. This means engaging with a stakeholder (e.g.: a faculty member, a department within a company/non-profit, an established student organisation) and supporting some of their decision making work, doing so using visualisation (and perhaps some analytics). ***While engagement with a stakeholder on a real problem is optional, it is strongly encouraged.*** To incentivise this, a small bonus will be awarded to groups who “make themselves useful” in that way. (This will be on top of how solving real/realistic problems has historically been positively correlated with higher project scores. If your team is not engaging with a real problem consider reviewing case studies to help ground your project.)

Groups should engage their stakeholders early, and identify specific decisions that visuals will be created to support, and share that concurrence with the instructor. The instructor reserves the right to require groups to rework the scope of their project.

If issues of data confidentiality arise, groups are reminded that they may create “synthetic data sets” with the same data field headings and of a similar (statistical) nature to the real data set for presentation and submission.

Your solutions should be **reusable tools**. They should not be oriented around one specific data set, but should be usable across multiple scenarios. (For example, the same decisions might have to be made every week, but the situation and the data would have changed, and hence the specific recommendations might differ. More concretely, consider the decision whether to bring an umbrella when leaving the house, the data to be presented in a “Weather App” varies from day to day, possibly leading to different decisions.)

Revisiting past projects is not objectionable as long as the dominant type of work done did not relate to data visualisation. **If past project work is being revisited, prior approval from the instructor should be sought.**

Grading: Instead of an in-class presentation with many groups rushing and “successively depleting grader energy”, groups will record their presentations as a video. The grader will watch the presentation, and occasionally pause to refer to the report and visuals. Given the time limitations, you should focus on how the relevant decisions are supported using your visuals. (For instance, mundane details on data preparation, details on the data set’s contents, and how the visuals were created are less relevant.) Groups may consider “engineering” “optional pauses” where the grader pauses the video to look at your report/visuals by making reference to the report (section/page number/paragraph) in their presentation. For matters that are “not viewed as optional” (e.g.: how to decide what to do using the visuals), include those in the presentation.

Important: *The philosophy behind the requirements of the Final Project are as follows. In university studies, most will go through the following two levels of demonstrating value: (i) describing the situation, and (ii) studying a situation and giving a recommendation. Here, you are asked to take another incremental step in demonstrating value. In particular, the requirement is to systematise an approach to identifying the right decision, and creating a “data product” from that. That is to say, the core requirement is to develop and implement a process to do (ii) for a range of situations, going beyond a one-off recommendation to something repeatable. Loosely speaking: “the consultants may leave, but their work product should continue to be useful”.*

Final submissions are to be uploaded to Canvas (to the Group Project folder) should generally be a single ZIP file named in the following format [Day]-G[Group Number].zip (where the day should be of the form Mon/Tue/Wed/...) for example Wed-G05.zip. Multi-part zip files will be allowed, in which case the names of the files submitted should take the form [Day]-G[Group Number].[ext]. (A trial folder will be created for experimentation.)

Submissions should contain:

- A report with length no more than 16 pages (12pt font; single spaced; no annexes; the limit excludes the cover page and contents pages) including images for the important use cases (PDF format) (Note: Shorter is better.)
- The data set (for a large public data set, provide instructions for reliably obtaining it)
- A presentation deck (if used in video)
- Dashboards & Other Visuals (twb× format; they should also be featured in the report/presentation; check this; a penalty will be applied if this is not adhered to; single file preferred, but if you choose multiple files, give them sensible names)
- A video presentation of up to 15 minutes (mp4 or mov format; at least 960×480 resolution; all students should speak and each should be introduced in speech or text captions; being on video is not compulsory; it is recommended to name your visuals and number your report sections, then refer to them in your video)

- Overview of video presentation (a list of: time stamps, topic, current presenter(s), page number and paragraph(s) in report; PDF format; it is recommended to create this as a sketch for your video, and update timestamps after recording your video)
- If your submission file is too large for Canvas, split it up, and make multiple submissions with appropriately named files like [Day]-G[Group Number]-[Part Number]-of-[Number of Parts].zip.

Following instructions is crucial to enable efficient grading. The course staff reserves the right to apply penalties for violating the submission guidelines. For example, when a ZIP file is requested, submission of RAR files may be penalised. It is your responsibility to verify compliance with submission guidelines.

Expectations: (i) narrate clearly and pay attention to your body language; (ii) provide but do not dwell on background information (e.g.: organisational priorities and their association with decision making); (iii) state the decisions to be made and relevant decision making criteria; (iv) show how data visualisation supports decision making (potentially over multiple steps; e.g.: user wants to know X, which is important because Y, these are the steps to take); (v) use visual encodings that are self-evident and explain those that aren't; (vi) anticipate and address questions that might arise; and (vii) make good use of filters/parameters (e.g.: insights on groups, and/or counterfactual analysis). Projects demonstrating domain insight and/or technique will be more highly regarded. Where useful, bring in external data and use data analytics methodologies. *Returning to (iv), to be clear, one time recommendations are insufficient, and the expectation is to provide **processes** which each map **data to a decision** (an action selected from among possible alternatives, which might include “no action” as an alternative). Consider illustrating cases with “if you see X then do Y because Z”.*

Not required: (a) extended discussion on mundane data preparation; (b) extended discussion on fields in the dataset; (c) extended discussion on how to create a visual in Tableau.

Again, grading will be done by watching the video as if it were an in-person presentation, with pauses to clarify matters using the report. The Overview document will be used like an index for cross referencing.

As senior students, substantial independence/resourcefulness and some level of production quality is to be expected. Recordings from mobile phone cameras are acceptable as long as content is clear, large enough and stable. It is recommended that you consider the use of screen-recording or video editing software, many free options are available and you should independently identify suitable tools. Submit something that you would be proud of.

In the event that team disputes arise, such as “a non-contributing member”, highlight them as follows: (i) start with an email to the instructor copying all group members (including the “non-contributing member”), (ii) all group members with grievances should state their points of view (even if it is a short affirmation of the claims), and (iii) the “accused” will have the opportunity to provide a defence. Following that, the instructor will take appropriate action. Ideally, groups should air issues openly before conflict of that level arises.

Submissions are due on Week 13 (last week of term) on Monday at 11.30pm. Expect extensions not to be granted. This is due to NUS grade reporting requirements for modules without a final exam.

Useful References

Nathan Yau (2011), Visualize This: The FlowingData Guide to Design, Visualization and Statistics, Wiley.

Cole Nussbaumer Knaflic (2015), Storytelling with Data: A Data Visualization Guide for Business Professionals, Wiley.

Allen B. Downey (2015), Think Stats: Exploratory Data Analysis, 2nd Edition, O'Reilly.

Steve Wexler and Jeffrey Shaffer (2017), The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios, Wiley.

Nancy Duarte (2010), Resonate: Present Visual Stories that Transform Audiences, Wiley.

Tentative Schedule

<i>Week</i>	<i>Visual Analytics</i>	<i>Tableau Desktop / Prep</i>
1 15/1 (M)	Introduction to Data Visualisation Review: Pivot Tables and Data Aggregation	Introduction to Tableau Fast Run: Sheets, Dashboards, and Stories
2 22/1 (M)	Visual Encoding of Data	The Zen of Tableau Data to Visuals: The Charts of Tableau
3 29/1 (M)		
4 5/2 (M)	Data Aggregation <i>(Potential past mid-term question review)</i>	Level of Detail (LOD) Expressions Table Calculations Data Preparation with Tableau Prep
5 12/2 (M)		
6 19/2 (M)	Midterm (Tue; 20 Feb 2024; 6.30pm to 8.00pm)	
-	Recess Week	
7 4/3 (M)	Exploring Data, Descriptive Analytics, and Counterfactual Analysis <i>Testing for Correctness in Analytics (Practical Problem Solving)</i>	Descriptive Statistics via the Analytics Pane Order of Operations Filters and Parameters More visuals
8 11/3 (M)		
9 18/3 (M)	Dashboard Development <i>Engaging Stakeholders (Discovery & Presenting Solutions)</i> Group Project Idea Presentation/Sharing	Dashboards and Stories Cross-Sheet Interactions Organizing and Supplementing Data Sources
10 25/3 (M)		
11 1/4 (M)	Visual Perception and Communication Group Project Visuals Presentation/Sharing	Outstanding Tableau Content Tableau Exercises
12 8/4 (M)		
13 15/4 (M)	– No Classes – <i>Due: Individual Assignment & Group Project (Monday, 11.30pm)</i>	
-	Reading Week	

See Canvas for virtual office hours/consultation time slots.