

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

DSC4215/DOS4811 Data Visualisation

Instructor: Jeremy Chen
Office: BIZ 2, #02-21
e-mail: bizewj@nus.edu.sg
Session: Academic Year 2021/2022 Semester 2

(Correct as at: 16 Nov 2021)

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; and (iv) learn how to communicate facts and ideas effectively with data. This course includes a substantial hands-on-learning component, and supports the development of highly marketable skills in visualisation.

Course Requirements

Students should have familiarity with a spreadsheet application such as Excel at the level of “pivot tables” (“group-by” and “aggregate”). 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.

Sections and Scheduling

Sectional Workshops:

- A1 on Tue 6.30pm – 9.30pm (Venue: BIZ2-B104 with Parallel Zoom Session)
- A2 on Wed 6.30pm – 9.30pm (Venue: BIZ2-B104 with Parallel Zoom Session)

Office Hours / Consultation Slots will be held virtually at 6.30pm – 7pm on one or more days from Week 4 on Mon, Thu, and/or Fri (on a rotating basis). See LumiNUS 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 use 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 licenses 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 licenses via the website. Student status letters that indicate the duration of enrollment 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>.

We also will make use of the NUS-provided messaging tool (Teams). The main “team” for the class will be `datavis-2022`. (Visit LumiNUS Announcements for the join code.)

Teams 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 of general interest in the `General` channel, 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 in the `General` channel”.)

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% |
| Individual Assignment | 20% |
| Mid-term | 30% |
| 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 in Teams to be eligible for class participation credit, please include, within your post, the last 4 digits and letter of your *matriculation number* in parentheses. 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:

- Join the class in Teams using the code provided in LumiNUS announcements
- 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 the `Introductions` 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 the shared document for Introductions and Teams for [Tuesday](#) and [Wednesday](#) (follow the links)

Content Sharing. Students should keep a lookout for interesting examples of visualisation in their work or in media, post examples in `Interesting Examples` and briefly (in a few sentences) discuss the visual in the context of what was covered in the course. Each student should post at least two examples with scoring done on a binary basis ("points" or "no points"). Indicate your full name (per LumiNUS) and the "last 4 digits plus letter of your matriculation number in parentheses" (as per **Introductions** above) to facilitate scoring (having both makes things easier, but the latter is crucial).

Zoom Classes. When joining a virtual class (over Zoom), please choose your names to include (i) group number (when available), (ii) preferred name, and (iii) "last 4 digits plus letter" of your matriculation number. For example, "G05 John Li (1234X)". This facilitates administering the session and crediting you for participation. (If this is no longer possible due to configuration changes, you will have to type them out where applicable.)

Q&A. Students should ask questions in the relevant Content channel in Teams and participation credit may be given to those who help their classmates. Once your answer has been acknowledged as "solving your classmate's problem", organize the question and answer in this [shared document](#) (being sure to identify yourself to receive credit).

As a matter of policy, 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.

Mid-term (30%)

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

It will be held on 15 Feb 2022 (Tue; 7.00pm – 8.30pm) in Week 6 to avoid the midterm rush. 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 in the professional world. In all such problems, you should verify that you are calculating what you intend to calculate. To validate your methodology, it often helps to start with a simplified data set where the answer is clear. In addition, “parallel approaches” (e.g.: using a Python script) have been observed to be effective validation tools in practice. While it may be possible to exclude Tableau from the solution process, please do not, as it defeats the purpose of your taking this course.

Additional Guidance:

- 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, these might also be accepted: “time & money”, “money and time”, “money & time”.
 - After automatic grading, non-accepted answers will be reviewed and appropriate responses will be added to the set of acceptable answers, such as “money, time” and “time, money” for the case above.
- Note that there are no “long answers” for “fill-in-the-blank” questions. So, answer as concisely as possible. Most answers may be found in the Tableau interface. When answering with formulas, remove all 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 teams 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. 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 “synthetic data sets”.

All team members should be in the same section, and groups may have up to 5 members.

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.** It is advisable to communicate with the instructor through a private channel in Teams (create one for your team, and another including the instructor).

Final submissions are to be uploaded to LumiNUS (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]

Submissions should contain:

- A report with length no more than 16 pages (12pt font; single spaced; no annexes) including images for the important use cases (PDF format) (Note: Shorter is better.)
- The data set
- A presentation deck (if used in video)
- Dashboards (twbx format; they should also be featured in the report/presentation; check this; a penalty will be applied if this is not adhered to)
- 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)
- 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; this facilitates grading)

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; (iii) state the decisions to be made and the criteria for decision making (i.e.: priorities); (iv) show how data visualisation supports decision making

or reveals interesting/useful insights with practical use cases being ideal (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 naturally arise; (vii) make good use of filters/parameters; and (viii) all team members should be involved. Projects demonstrating domain insight and/or technique will be more highly regarded. Where useful, bring in external data and use data analytics methodologies.

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 are responsible to identify suitable tools. Submit something that you would be proud of.

In the event that disputes arise, such as a “non-contributing member”. You may highlight this to the instructor in the following way: (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 a previous set of claims), and (iii) the “accused” will have the opportunity to defend himself/herself. Following that, the instructor will take appropriate action. Ideally, groups should air issues openly before conflict of that level arises.

Submissions are Due: 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 Knafllic (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 10/1 (M) | Introduction to Data Visualisation Review: Pivot Tables and Data Aggregation | Introduction to Tableau Fast Run: Sheets, Dashboards, and Stories |
| 2 17/1 (M) | Visual Encoding of Data | The Zen of Tableau Data to Visuals: The Charts of Tableau |
| 3 24/1 (M) | Data Aggregation | Level of Detail (LOD) Expressions Table Calculations |
| 4 31/1 (M) | No Classes (Chinese New Year) | |
| 5 7/2 (M) | Data Aggregation | Data Preparation with Tableau Prep Tableau Online |
| 6 14/2 (M) | Midterm (Tuesday; 15 Feb 2022; 7.00pm to 8.30pm; Venue: TBA) | |
| - | Recess Week | |
| 7 28/2 (M) | Exploring Data, Descriptive Analytics, and Counterfactual Analysis | Descriptive Statistics via the Analytics Pane Order of Operations Filters and Parameters More visuals |
| 8 7/3 (M) | | |
| 9 14/3 (M) | Dashboard Development | Dashboards and Stories Cross-Sheet Interactions Organizing and Supplementing Data Sources |
| 10 21/3 (M) | | |
| 11 28/3 (M) | Visual Design, Communication and Storytelling | Outstanding Tableau Content Tableau Exercises |
| 12 4/4 (M) | Guest Lectures on Visual Analytics in the Enterprise | |
| 13 11/4 (M) | Guest Lectures on Visual Analytics in the Enterprise <i>Due: Individual Assignment & Group Project (Monday, 11.30pm)</i> | |
| - | Reading Week | |

See LumiNUS for virtual office hours/consultation time slots.