City University of Hong Kong
Course Syllabus

offered by School of Data Science
with effect from Semester A 2024/25

Part I Course Overview

Course Title:	Exploratory Data Analysis and Visualization
Course Code:	SDSC5002
Course Duration:	One Semester
Credit Units:	3
Level:	P5
Medium of Instruction:	English
Medium of Assessment:	English
Prerequisites: *(Course Code and Title)*	Nil
Precursors: *(Course Code and Title)*	Nil
Equivalent Courses: *(Course Code and Title)*	Nil
Exclusive Courses: *(Course Code and Title)*	Nil

Part II Course Details

1. Abstract

The goal of this course is to introduce students to the essential exploratory techniques for summarizing data and associated visual methods. Exploratory data analysis is typically applied before formal modelling commences. It can help formulate hypotheses and inform the development of more complex statistical models. The course begins with an introduction of basic graphical techniques used in exploratory data analysis, continues with typical statistical methods for exploratory analysis including clustering and dimension reduction techniques that allow you to make graphical displays of high dimensional data, then focuses on visualization techniques and methods for a broad range of data types. Principles from perception will be introduced to design effective data visualizations. Students will work through a series of case studies and hands-on projects to learn the skills for working with real-world data.

2. Course Intended Learning Outcomes (CILOs)

(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

No.	CILOs	Weighting (if applicable)	Discovery-enriched curriculum related learning outcomes (please tick where appropriate)		
			A1	*A2*	*A3*
1.	**Describe** basic concept of exploratory data analysis and its relationship to exploratory data analysis, data mining, and the applications in network analysis, spatiotemporal analysis and machine learning.	20%	✓		
2.	**Describe and apply** basic visualization techniques used in exploratory data analysis	20%	✓	✓	
3.	**Describe and apply** statistical methods for exploratory analysis in high-dimensional data, network analysis and spatiotemporal analysis.	20%	✓	✓	
4.	**Discuss** the principle of perception and be able to select suitable visualization techniques and methods for diverse types of datasets in physical and social sciences.	20%	✓	✓	
5.	**Explain** how exploratory data analysis and visualization can be applied to real life problems in various applications with data-driven insights.	20%	✓	✓	✓
		100%			

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: *Accomplishments*
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

3. Learning and Teaching Activities (LTAs)

(LTAs designed to facilitate students' achievement of the CILOs.)

LTA	Brief Description	CILO No.					Hours/week (if applicable)
		1	2	3	4	5	
Lectures	Students will participate in formal lectures to acquire a comprehensive understanding of essential exploratory techniques for summarizing data and associated visual methods. These lectures will serve as a valuable platform to study visualization tools, gain software package usage training, and practice real-world data analysis with state-of-the-art visualization techniques.	✓	✓	✓	✓	✓	39 hours/Semester
Tutorial	Students will participate in tutorial activities aimed at expanding their proficiency in Python programming. These tutorials will provide hands-on experience and practical applications of Python specifically for data visualization purposes. By actively engaging in these activities, students will develop a deeper understanding of the language and its application in the field of data visualization.	✓	✓	✓	✓	✓	6 hours/Semester (included in the lecture hours)
Group project	Students will work collaboratively in groups to enhance their understanding and application of exploratory techniques and data visualizaion methods learned in the class. Through producing a research report and a presentation, they will have the opportunity to apply their knowledge to real-world scenarios. This hands-on approach will help them consolidate their learning and gain practical experience.	✓	✓	✓	✓	✓	6 hours/Semester (included in the lecture hours)

4. Assessment Tasks/Activities (ATs)

(ATs are designed to assess how well the students achieve the CILOs.)

Assessment Tasks/Activities	CILO No.					Weighting	Remarks
	1	2	3	4	5		
Continuous Assessment: 100 %							
Group Project	✓	✓	✓	✓	✓	40%	The students form groups to work on real-world datasets. They are required to use visualization and exploratory data analysis tools to draw meaning insights by analyzing and visualizing the data.
Individual Coursework	✓	✓	✓	✓	✓	25%	Individual coursework involves the analysis and visualization of real-world datasets.
Test		✓	✓	✓	✓	35%	The midterm test evaluates the conceptual capability and skills of students in exploratory data analysis and data visualization.
Examination: 0 % (duration: , if applicable)							
						100%	

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

Applicable to students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Fair (C+, C, C-)	Marginal (D)	Failure (F)
1. Group Project	Based on **oral presentation and submitted written report** to evaluate understanding of subject matter, evidence of knowledge base, capacity to analyse and synthesize, and evidence of original and critical thinking	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Individual Coursework	Based on **submitted written work and lab attendance** to evaluate understanding of subject matter, evidence of knowledge base, capacity to analyse and synthesize, and evidence of original and critical thinking.	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Test	Based on **submitted written work** to evaluate understanding of subject matter, evidence of knowledge base, capacity to analyse and synthesize, and evidence of original and critical thinking	High	Significant	Moderate	Basic	Not even reaching marginal levels

Applicable to students admitted from Semester A 2022/23 to Summer Term 2024

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B)	Marginal (B-, C+, C)	Failure (F)
1. Group Project	Based on **oral presentation and submitted written report** to evaluate understanding of subject matter, evidence of knowledge base, capacity to analyse and synthesize, and evidence of original and critical thinking	High	Significant	Basic	Not even reaching marginal levels
2. Individual Coursework	Based on **submitted written work and lab attendance** to evaluate understanding of subject matter, evidence of knowledge base, capacity to analyse and synthesize, and evidence of original and critical thinking.	High	Significant	Basic	Not even reaching marginal levels
3. Test	Based on **submitted written work** to evaluate understanding of subject matter, evidence of knowledge base, capacity to analyse and synthesize, and evidence of original and critical thinking	High	Significant	Basic	Not even reaching marginal levels

Part III Other Information (more details can be provided separately in the teaching plan)

1. Keyword Syllabus

(An indication of the key topics of the course.)

- Introduction of visualization, e.g., classical visualizations, historical remarks, examples of scientific visualizations and information visualizations.
- Principles of visual design, perception and color theory.
- Basic graphical techniques used in exploratory data analysis, multivariate data visualization
- Visualization of high-dimensional data, clustering and dimension reduction techniques.
- Visualization of scalar fields and vector fields.
- Applications and case studies of exploratory data analysis and visualization in different topics, e.g., network science, spatiotemporal analysis and machine learning.
- Introduction of the programming language (e.g., Python) and the software (e.g., Tableau) for data visualization.

2. Reading List

2.1 Compulsory Readings

(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

1.	Lecture note

2.2 Additional Readings

(Additional references for students to learn to expand their knowledge about the subject.)

1.	Interactive Data Visualization for the Web By Scott Murray, O'Reilly Media, 2012.
2.	Information Visualization: Perception for Design By Colin Ware, 2012
3.	Visual Thinking: for Design By Colin Ware, 2008
4.	The Visual Display of Quantitative Information By Edward R. Tufte, 2001
5.	Visualizing Data: Exploring and Explaining Data with the Processing Environment By Ben Fry, O'Reilly Media, 2007.