UNIVERSITY of NICOSIA

Course Syllabus

Course Code	Course Title	ECTS Credits
COMP-543DL	Managing and Visualizing Data	10
Prerequisites	**Department**	**Semester**
COMP-542DL	Computer Science	Spring
Type of Course	**Field**	**Language of Instruction**
Required	Data Science	English
Level of Course	**Lecturer(s)**	**Year of Study**
2nd Cycle	Dr. D. Trihinas	1st
Mode of Delivery	**Work Placement**	**Corequisites**
Distance Learning	N/A	None

Course Objectives:

The main objectives of the course are to:

- Introduce the importance of data management and visualization for data science projects.
- Present fundamental concepts and stages comprising data analysis pipelines.
- Provide a comprehensive overview of the relational data model and relational databases.
- Demonstrate how to devise queries to efficiently store and extract data from relational databases.
- Present algorithmic techniques to measure and validate data quality in a data science project.
- Introduce a problem-solving framework for “thinking with data” that targets the scientific treatment for data curation (e.g., data preparation, cleaning and management).
- Present intuitive data types and effective data structuring techniques to speed data curation.
- Present basic techniques for measuring the performance of data intensive applications and showcase methods to improve the efficiency of data manipulation.
- Provide a comprehensive overview of the practice of data visualization.
- Discuss the persuasive power of data visualization and the benefits of storytelling through data to visually communicate analysis results.
- Introduce techniques and aesthetic rules for creating effective visualizations based on principles and perspectives derived from the visual arts, perceptual psychology, journalism and cognitive science.
- Introduce a number of common data domains and corresponding visualization methods

and algorithms for multivariate data, graphs, cartography and timestamped data.
- Demonstrate methods and algorithms to model, curate and visually depict distributions, uncertainty and error-bounds to aid the exploration of hidden relationships and correlations among visualized data.
- Present algorithms and techniques to efficiently and effectively model, curate and visually depict large datasets.

Learning Outcomes:

After completion of the course students are expected to be able to:

- Understand the basic concepts involved in computational thinking towards solving data-oriented problems.
- Understand the fundamental concepts of relational algebra and calculus.
- Demonstrate the ability to create appropriate database schemas based on given problem descriptions.
- Be proficient in writing (near-) optimal queries for certain (complex) problem descriptions.
- Be familiar with the methods for measuring and optimizing the performance of relational databases and data curation tasks.
- Apply in practice the principles for pre-processing datasets and for indexing data to improve the efficiency of data trawling and extraction.
- Apply in practice the principles for combining, aggregating and summarizing datasets.
- Understand the algorithmic process for cleaning impure datasets and acknowledge the scientific treatments for dealing with missing data.
- Develop an understanding of how the different big data dimensions impact data analysis.
- Develop an understanding of the fundamentals of storytelling through data and the alignment around basic concepts that are required to effectively present and communicate analysis results.
- Distinguish between informative and deceitful data visualizations.
- Create relatively high-quality visualizations to aid statistical and exploratory data analysis by applying suitable visualization techniques and aesthetic principles.
- Apply appropriate principles to depict data distributions, uncertainty and error bounds to extract and explore hidden relationships and correlations among visualized data.
- Leverage incremental and approximation techniques to balance the complexity and clutter when visualizing large datasets.
- Demonstrate the ability to curate and visually depict spatio-temporal datasets.
- Understand the basic principles and concepts implied to visually apprehend and depict high-dimensional data.

Course Content:

1. Introduction to Data Management and Visualization
    a. Impact of Data Management and Visualization
    b. Data Analysis Pipeline
2. The Relational Data Model
    a. Relational Database Management Systems
    b. Relational Algebra and Calculus
    c. Data Schemas and Indexes
    d. SQL
3. Query Processing and Optimization
    a. Algorithms for SELECT-JOIN-SORT-PROJECT queries
    b. Query Pipelines
    c. Query Optimization
    d. Query Performance
4. Data Quality and Cleaning
    a. The Data Quality Continuum
    b. Measuring Data Quality
    c. “Dirty”, Missing and Impure Data
    d. Data Imputation
5. Data Transformation and Aggregation for Relational Data Warehouses
    a. Combining and Reshaping Data
    b. OLAP Techniques
    c. Measuring the Impact of Lookup and Range Queries
    d. Data Cubes, Data Spheres and Pivot Tables
6. The Big Data Era: Models and Architectures
    a. The Big Data Dimensions (3V’s, 5V’s)
    b. Big Data and the Data Science Process
    c. Models and Architectures
7. Storytelling Through Data
    a. The Art of Visual Communication
    b. Exploratory vs Explanatory Visualizations
    c. Separation of Content and Design
    d. Aesthetics (e.g., Coordinate Systems, Color Scales, Axes)
    e. Visualizing Univariate Data (e.g., Bar, Pie, Scatter Plots)
8. Plotting Data Distributions and Uncertainty
    a. Histograms and Density Plots
    b. Cumulative Distribution Functions
    c. Correlograms
    d. Uncertainty in Empirical Data
    e. Error-Bounds, Box Plots and Q-Q plots
9. Timeseries Exploration and Textual Data
    a. Timestamped Data

b. Sliding Windows
c. Categorical Data
d. Word Clouds

10. Data Visualization at Scale
    a. Interactive Latency and Big Data Navigation
    b. Incremental Models and Approximation Techniques
    c. Sampling, Filtering, Change Detection
11. Visualizing Spatio-Temporal Data
    a. Point and Distance Data
    b. Graphs and Networks
    c. Plotting Maps
12. Visualizing High-Dimensional Data
    a. Multivariate Data
    b. Dimensionality Reduction (e.g., PCA, t-SNE, Isomaps)
    c. Paired Diagrams, 3D plots, Glyphs
    d. Layered and Overlapping Data

Learning Activities and Teaching Methods:

Lectures, Exercises, Software Tool Tutorials, Case-Study Presentations, Discussions.

Assessment Methods:

Final Exam, Homework, Semester Project.

Required Textbooks / Readings:

Title	Author(s)	Publisher	Year	ISBN
Exploratory Data Mining and Data Cleaning (2nd edition)	T. Dasu and T. Johnson	Wiley Interscience	2006	978-0-471-26851-2
Fundamentals of Data Visualization*	Claus O. Wilke	O'Reilly Media	2018	978-1-492-03108-6

Recommended Textbooks / Readings:

Title	Author(s)	Publisher	Year	ISBN
Fundamentals of Database Systems (7th edition)	Ramez Elmasri and Shamkant B. Navathe	Pearson	2015	978-0-133-97077-7
Visualization Analysis and Design	Tamara Munzner	CRC Press	2014	978-1-466-50891-0
Python for Data Analysis (2nd edition)	Wes McKinney	O'Reilly Media	2017	978-1-491-95766-0

* Made freely available online by the author: https://serialmentor.com/dataviz/