ACM Data Science Task Force Course Example

DSC 321 – Data Visualization
Northern Kentucky University, Kentucky
Seth Akonor Adjei
adjeis1@nku.edu

Knowledge Areas that contain competencies (knowledge, skills, and dispositions) covered in the course

Knowledge Area	Total Number of Contact Hours
Analysis and Presentation (AP)	38
Data Acquisition and Governance (DG)	6
Programming (P)	4

Where does the course fit in your undergraduate Data Science curriculum?
This course is part of the Data Science major. Students typically take this course in the third or fourth year of the curriculum. This course has a prerequisite but not a required post-requisite. The only time I taught this class, 8 students took the course.

Is this course from or used in other curricula/majors?
Yes, the course was designed explicitly for the data science major. About 50% of the topics have generally not changed. However, the remaining 50% which relates to the use of data visualization tools evolves over time. But these changes are not as frequent as the non-tool related content.

What is covered in the course?
Design principles and techniques for visualizing data; visualization techniques for spatial and geospatial data; visualization techniques for multivariate data; networks visualization; interaction concepts and techniques; designing effective visualizations.

What is the format of the course?
This is a face-to-face course that has 3 hours of contact time per week for a 16-week semester. For each week, the first hour is presented as a lecture, the second hour is a class discussion and the third hour is for working on labs and project assignments. This course is offered once every year in the Fall.

How are students assessed?
A variety of assessment types are used for this course: midterm exams, written paper assignment, in-class paper and chapter presentations by students, and project work. Students are expected to spend 9 hours each week on assignments (which include lab exercises,

projects and homework assignments). The following table lists the number of each of the assessment types:

Assessment Type	Count
Weekly Assignments	10
Presentations	2
Two-Part Project	1
Final Project Presentation	1

Course tools and materials

A combination of materials is used for this course. The course uses one required textbook and a recommended textbook, each of which is available for free online. The programming languages used for this course include Python, R, JavaScript and HTML. For each of these programming languages, no specific programming environment is required. However, I use RStudio, PyCharm and SublimeText, which are all free for students. There is no specific dataset that is required for this course. Students are encouraged to identify datasets that meet a set of predefined criteria which I provided for each assignment. The students are then provided with list of portals from where data can be freely sourced. Examples of such open data sources include the City of Cincinnati Open Data Portal (https://data.cincinnati-oh.gov/) and UCI Machine Learning Data Repository.

Why do you teach the course this way?

The goal of this course is to ensure that by the end of the course, students will be able to:

1. Explain the data visualization design principles
2. Explain the data visualization techniques applicable to spatial, temporal and other types of data of interest.
3. Apply the visualization techniques to answer data analytics related questions from clients
4. Use available data visualization tools, environments and APIs to create effective visualizations
5. Apply characteristics of an effective data visualization to evaluate the effectiveness of visualizations

Generally, students consider the course as interesting and challenging due to the number of tools they are required to use. We are therefore considering a revamp of the course to limit the amount of material covered in the course. Since this is the first time I have taught this course, I do not have the history requested by this question.

Body of Knowledge coverage

KA	**Sub-domain**	**Competencies Covered**	**Hours**
PDA	Programming	*Use standard libraries for a given programming* language	4

DG	Data Acquisition and Governance	Data acquirement, information extraction, data integration, data sampling, data reduction, compression, transformation and cleaning algorithms	6
AP	Analysis and Presentation	Explaining data and inferences made from data in oral, written and graphical formats	6
		Use standard APIs and tools to create visual displays of data, including graphs charts, tables and histograms	12
		Apply a variety of visualization techniques to different types of data	12
		Propose a suitable visualization design for a particular combination of data characteristics and application tasks	3
		Analyse the effectiveness of a given visualization for a particular task	3
		Be aware that the client for a presentation is often not a data scientist	1
		For an identified client, undertake and document an analysis of their needs.	1