The Government of Russian Federation

The Federal State Autonomous Institution of Higher Education
"National Research University – Higher School of Economics"

Faculty of Business and Management

School of Business Informatics

Department of Innovation and Business in Information Technologies

38.04.05 «Business Informatics»

Master's programme "Big Data Systems"

Data Visualization – Spring 2017 (Module 3 & 4)

Tentative Course Syllabus

Author: P.B. Panfilov, Professor, Department of Innovation and Business in IT, ppanfilov@hse.ru

Recommended: UMS Section on Business Informatics, Chair Y.V.Taratukhina

Approved: Meeting of the Department of Innovation and Business in IT, Head of Department S.V.Maltseva

Approved: Academic Council of Faculty of Business and Management, Secretary N.M.Lyubakova

Moscow 2016

Rationale:
Data and information visualization is the graphical communication of data and information for the purposes of presentation, confirmation, exploration, and analysis. Images can be used to convey numbers, concepts, and relationships using techniques such as maps, icons, graphs, and other visual forms. In the past decade, visualization has evolved into a discipline, drawing from such fields as computer graphics, human-computer interaction, perceptual psychology, and art.

Course Objectives
To introduce students to the fundamental problems, concepts, and approaches in the design and analysis of data visualization systems. To familiarize students with the stages of the visualization pipeline, including data modeling, mapping data attributes to graphical attributes, perceptual issues, existing visualization paradigms, techniques, and tools, and evaluating the effectiveness of visualizations for specific data, task, and user types.
On completion of the course, the student should:

- understand the evolution of the data visualization from its early beginnings as text only, to text and 2D images, to the emerging rich digital media environment of text, 2D images, 3D graphics, animations, interactive visualizations and virtual environments;
- understand the existing visualization paradigms and perceptual issues;
- understand application areas of data visualization technology, such as e-commerce, scientific visualization, data mining and virtual worlds;
- be familiar with existing data visualization techniques, authoring and validation tools;
- have built a data visualization application using available technologies (e.g., OpenGL, Java2D, Java3D, Processing, or X) and/or visualization tools; understand evaluation techniques for successful design and development of efficient and effective data visualization applications.

The emphasis of the course will be on exposing students to the current research issues and on identifying potential research topics in data visualization as it applies to large-scale big data systems.

Prerequisites: 2D and 3D Computer Graphics; Data Mining

Approach
Formal lectures, home assignments, term project, research study/presentation.

Assignments: The monthly assignment will consist of several components:

- Readings: Every week we will discuss a visualization topics based on a different chapters of the book.
- Exercises: Every week students will be offered a set of exercises included with the specific chapters of the book.
- Visualization Project: Each student, either working alone or in pairs, will be responsible for selecting data that is of interest to him (them) and designing and developing a visualization tool to display and explore that type of data. Students can use sample programs included with the textbook as starting points for their programming projects. Alternatively, student(s) can identify an existing tool for data visualization and create a web page to describe the tool and show examples of its use on data selected. For the

end of the term each student should create a report (5-10 pages) and a poster about his/her/their project.

- Visualization Study: Each student (individual) is responsible for researching a topic in visualization, reviewing 2 or more papers on the topic, and creating a presentation and 15 minute talk on the topic. In the ideal situation, this topic would be related to the visualization project (either on the same kind of data or using a technique that you are planning to implement). Alternatively, students can prepare presentations on different visualization topics selected on the basis of additional readings from reports in Chapter 16. The talks will be presented at the end of first half of the term (Module 3).

Visualization Project: The steps of the visualization project are as follows:

1. Select some socially relevant data set or information source as a focus for visual analysis. Confirm your topic with instructor.
2. Locate 1 to 3 papers that present methods for visualizing this kind of data. Summarize and include references to them in your project report.
3. Design or extend a visualization to allow exploration of your data/information. You are not allowed to just use Excel! There should be some programming involved.
4. Explore your dataset and identify a modest number of "interesting" features in the data.
5. Write a short (between 5 and 10 pages, single spaced) paper describing the data, the papers you read related to visualizing this type of data, the process you followed in developing your visualization, the methods used for exploration, and the things you discovered. Include screen shots and relevant references.
6. Create a poster or web-page describing your data and how you visualized it. Show more than one view and, if possible, more than one data set.

Software Resources: OpenGL, Java2D, Java3D, Processing, or X can be used for software development. In most cases, students can get by with 2-dimensional graphics, though for some types of visualization, 3-D is essential.

This project is due by the start of our last class. We will hold a special in-class session for all to see what everyone has been doing. Students can demonstrate their visualization systems in action if projects involve some programming.

Visualization Study: The topics for visualization study – within research reports published in recent years as it is presented in Chapter 16, including (numbers refer to the Bibliography in textbook, pp.513-548):

- Chris Johnson's list of top problems in scientific visualization [208];
- Chaomei Chen's list of top unsolved problems in information visualization [70];
- MacEachren and Kraak's report on research challenges in geovisualization [279];
- the NVAC research and development agenda for visual analytics [84];
- the NIH/NSF visualization research challenges report [207];
- the Grand Challenge in Information Visualization panel [160].

Students are directed to the original reports for in-depth descriptions, as well as identification of research areas specific to particular branches of the ever-enlarging field of visualization.

Assessable work
1 Visualization Project will constitute 30% of the student's assessment.
1 Visualization Study/Paper presentation will constitute 30% of the student's assessment.
1 Formal examination(s), 30% of student's assessment.
The final exam will cover the entire course. The written exam will consist of 10 problems, giving 10 points each, total 100 points for the exam.
Class attendance and participation will make remaining 10% of student's assessment.

Grading formula
Cumulative grade for the subject = 0,5*Project grade + 0,4*Study Grade + 0,1*Attendance
Final grade for the subject = 0,7*Cumulative grade + 0,3* Grade for the exam

Grading table

ECTS Grade	Ten-point scale	Five-point scale
A+	10 (Excellent)	5 (Excellent)
A	9 (Very Good)	
A-	8 (Very Good)	
B+	7 (Good)	4 (Good)
B	6 (Good)	
C+	5 (Satisfactory)	3 (Satisfactory)
C	4 (Satisfactory)	
F	1, 2, 3 (Fail)	1, 2 (Fail)

Textbook:
M. Ward, G. Grinstein, and D. Keim. Interactive Data Visualization: Foundations, Techniques, and Applications, N.-Y.: A K Peters/CRC Press, 2015. - 558p. ISBN-13: 9781482257373
Textbook web-site: http://www.idvbook.com/

The textbook covers mathematical and analytical aspects from its foundations to human visual perception, and from coded algorithms for different types of data, information and tasks to the design and evaluation of new visualization techniques. It includes exercises, programming projects, and related readings for each chapter, and provides sample programs as starting points for building one's own visualization tools.

The book concludes with an examination of several existing visualization systems and projections on the future of the field.

Additional readings/Resources:

M. Ward, G. Grinstein, and D. Keim. Interactive Data Visualization: Foundations, Techniques, and Applications, AK Peters Ltd, 2010. ISBN 978-1-56881-473-5.
U. Fayyad, G. Grinstein, and A. Wierse. Information visualization in data mining and knowledge discovery. Morgan Kaufmann 2001. ISBN-10: 1558606890
A. Kirk. Data Visualization: a Successful Design Process.

Course outline (Chapter numbers refer to the textbook)

Part I: Introduction to the field of data visualization

Chapter 1. Introduction *(1 week)*
- What is visualization?
- History of Visualization
- Relationship Between Visualization and Other Fields
- The Visualization Process

Chapter 2. Data Foundations *(1 week)*
- Types of Data
- Structure within and between Records
- Data Preprocessing

Chapter 4. Visualization Foundations *(1-2 weeks)*
- The Visualization Process in Detail
- Semiology of Graphical Symbols
- The Eight Visual Variables
- Historical Perspective
- Taxonomies

Part II: Data visualization techniques.

Chapter 5. Visualization Techniques for Spatial Data *(1 week)*
- One-, Two-, and Three-Dimensional Data
- Dynamic Data
- Combining Techniques

Chapter 6. Visualization Techniques for Geospatial Data *(1 week)*
- Visualizing Spatial Data
- Visualization of Point Data, Line Data, and Area Data
- Other Issues in Geospatial Data Visualization

Chapter 7. Visualization Techniques for Time-Oriented Data
- Definitions: Characterizing Time-Oriented Data
- Visualization Time-Oriented Data
- TimeBench: A Data Model and Software Library for Visual Analytics of Time-Oriented Data

Chapter 8. Visualization Techniques for Multivariate Data *(1 week)*
- Point-based, Line-Based, and Region-Based Techniques
- Combinations of Techniques

Chapter 9. Visualization Techniques for Trees, Graphs, and Networks *(1 week)*
- Displaying Hierarchical Structures
- Displaying Arbitrary Graphs/Networks
- Other Issues

Chapter 10. Text and Document Visualization *(1 week)*
- Levels of Text Representations
- The Vector Space Model
- Single Document, Document Collection, and Extended Text Visualizations

Part III: Interactive visualization

Chapter 3. Human Perception and Information Processing *(1 week)*
- What Is Perception?
- Physiology
- Perceptual Processing
- Perception in Visualization
- Metrics

Chapter 11. Interaction Concepts *(1 week)*
- Interaction Operators
- Interaction Operands and Spaces
- A Unified Framework

Chapter 12. Interaction Techniques *(1 week)*
- Screen Space
- Object-Space (3D Surfaces)
- Data Space (Multivariate Data Values)
- Attribute Space (Properties of Graphical Entities)
- Data Structure Space (Components of Data Organization)
- Visualization Structure Space (Components of the Data Visualization)
- Animating Transformations
- Interaction Control

Part IV: Visualization systems

Chapter 13. Designing Effective Visualizations *(1 week)*
- Steps in Designing Visualizations
- Problems in Designing Effective Visualizations

Chapter 14. Comparing and Evaluating Visualization Techniques *(1 week)*
- User Tasks
- User Characteristics
- Data Characteristics
- Visualization Characteristics
- Structures for Evaluating Visualizations
- Benchmarking Procedures
- An Example of Visualization Benchmarking

Chapter 15. Visualization Systems *(1 week)*
- Systems Based on Data Type
- Systems Based on Analysis Type
- Text Analysis and Visualization
- Modern Integrated Visualization Systems
- Toolkits

Part V: Visualization R&D

Chapter 16. Research Directions in Visualization *(1 week)*