Problem Sheet DTC — Practical Assignment
"Parallel Coordinates Plots"

Recommended Hours: (**level 1:** 2 hrs, **level 2**: 4-6 hrs, **level 3**: 12-14 hrs, **level X**: 10-14 hrs)
Initial Assessment: a 3-hour Practical session in Week 5
Submission Deadline: 12:00 noon, Wednesday, 26 February 2014
Guest Lecturers: Eamonn Maguire and Alfie Abdul Rahman
Lab. supervisors and Assessors: Alfie Abdul Rahman and Eamonn Maguire

Summary

You are required to implement one or more parallel coordinates plotting tools to meet the requirements at one of the three levels specified below. Each level is associated with **40** marks, and the marking will depend on the quality of your software as well as that of your paper-based report. In other words, attempting a Level does not guarantee to gain 40 marks, while attempting a level unsuccessfully may still gain some marks. If one does not attempt a particular level, of course, no mark can be given to that level. **Only a maximum of 100 marks will be counted towards the total mark of this module**.

Level X is only for students who have limited programming background, and have serious difficulties in completing Level 2 and Level 3 part of this assignment.

You are allowed to build your tools based on existing examples in the D3.js repository, though you must acknowledge the relevant examples explicitly in your source code and paper-based report. The grading scale has already taken into account the availability of such examples, while recognising the effort needed to adapt them for this assignment. However, no student should make use of others (e.g., online coding service) to complete their assignment, or take programs from another student.

The recommended programming environment is D3.js, for which a 2-hour lecture is scheduled in Week 5. In addition, there is a 3-hour laboratory session in the same week. If you wish to use a different programming environment instead of D3.js, please let the course lecturer know in advance.

Please submit your work to Dr. Alfie Abdul Rahman. Each piece of work will be marked by either Dr. Alfie Abdul Rahman or Eamonn Maguire. Please submit your paper-based report on Wednesday 26 February. The paper-based report should include the followings:
- A summary (no more than 1 A4 page), which outlines the level that you have attempted, the existing examples in the D3.js repository that you have adapted, the functionality that you have successfully completed, and the estimated number of hours that you spent on this assignment (for help us balancing the workload in future years).
- One or more screenshots (no more than 2 A4 pages) to demonstrate the visualization results of your work.
- Source code with essential inline documentation.

LEVEL 1: Simple Adaptation (40 marks)

The software tool at this level should be able to display the tabular data in the spreadsheet of **WorldEconomy2012.xlsx** using a parallel coordinates plot. The table includes seven columns with 19 measured values in each column, that is, seven 19-value data series for the seven variables. The variables are:

- **Region:** Codes for different regions
- **H-Exp:** Household consumption expenditure
- **G-Exp:** General government final consumption expenditure
- **GPF:** Gross capital formation (GPF)
- **Exports:** Exports of goods and services
- **Imports:** Imports of goods and services
- **GDP:** Gross Domestic Product (GDP)

Note that the region codes are not the standard UN codes. The numbers are chosen to provide appropriate grouping of sub-regions in each continent.

Your software must have the following essential functionality:

- Display a line for each of the 19 regions intersecting with the 7 axes at correct places.
- Display labels for each axis.
- Provide a brushing utility.

LEVEL 2: Competent Tool Development (40 marks)

You must complete LEVEL 1 before attempting LEVEL 2.

The software tool at this level should be able to display the tabular data in three **Population** spreadsheets using a parallel coordinates plot. You are required to use **PopulationAmeN.xlsx** (i.e., your Dataset 1) plus two other spreadsheets based on the first letter of your first name and the first letter of your surname. For example, the two letters for John Smith is J and S. The allocation is given in the following table:

1st letter of first name	Dataset 2	1st letter of surname	Dataset 3
A, J, S	PopulationAfrE.xlsx	A, J, S	PopulationAmeC.xlsx
B, K, T	PopulationAfrM.xlsx	B, K, T	PopulationAmeCa.xlsx
C, L, U	PopulationAfrN.xlsx	C, L, U	PopulationAmeS.xlsx
D, M, V	PopulationAfrS.xlsx	D, M, V	PopulationOcean.xlsx
E, N, W	PopulationAfrW.xlsx	E, N, W	PopulationAsiW.xlsx
F, O, X	PopulationEurE.xlsx	F, O, X	PopulationAsiC.xlsx
G, P, Y	PopulationEurN.xlsx	G, P, Y	PopulationAsiE.xlsx
H, Q, Z	PopulationEurS.xlsx	H, Q, Z	PopulationAsiS.xlsx
I, R	PopulationEurW.xlsx	I, R	PopulationAsiSE.xlsx

The datasets are formulated based on the 2010 UN population statistics, with additional Human Development Index (HDI) data for 2012. Each dataset has 12 columns:

- **Region:** The abbreviation for a region in the world
- **Country or SR:** The name of a country or a special region
- **P-Both:** Population of both genders (in thousands)
- **P-Male:** Male population (in thousands)

- **P-Female:** Female population (in thousands)
- **Rate:** Population growth rate (between 2005 and 2010)
- **Mdn-Age:** Median age of the population
- **Density:** the number of people per square km
- **Child-D**: Child dependency ratio (ratio of population aged 0-19 per 100 population 20-64)
- **Old-D**: Old-age dependency ratio (ratio of population aged 65+ per 100 population 20-64)
- **HDI:** Human development index (HDI), where a 0 indicates a missing datum
- **HDI-Qt:** Human development index quartiles, *very high*, *high*, *medium* and *low*. There is a fifth nominal value *unknown* for any missing HDI value.

Your software must have the following essential functionality:
- Display a line for each of the countries (or special regions) in your three datasets, intersecting with the 12 axes at correct places.
- Colour-code lines based on their regions (i.e., one colour per dataset).
- Display labels for each axis.
- Provide a brushing utility, and an axis-swapping utility.

LEVEL 3: Problem Solving with a Personal Design (40 marks)

You must complete LEVEL 2 before attempting LEVEL 3.

The software tool at this level should be able to display the tabular data in the spreadsheet of **WorldEconomy1990-2010.xlsx** using a specially-designed parallel coordinates plot. This is a time varying dataset with five tables. Each table consists of a dataset similar to the one at LEVEL 1, except that they are for 1990, 1995, 2000, 2005, and 2010 respectively.

There are no specific requirements at this level, except that your software should try to help users visualize the temporal relationships among data values within each column (i.e., the same concept in different years), as well as the relationships among different columns (i.e., different concepts in the same year). For example, you may choose to show 5 parallel coordinates plots simultaneously, or associate each axis with a time series plot or bar chart. Please beware that programming slightly complex interactions may take some time to get them right.

In order to make data at different years comparable, it is recommended to set the value range for each axis based on its global MIN and MAX values (during 1990 ~ 2010).

LEVEL X: Problem Solving with a Personal Design (60 marks)

You must complete LEVEL 1 before attempting LEVEL X.

Write a short essay on a topic of visualization literacy (1000-1500 words, excluding references). You may aid your essay with illustrations. You should choose your own title and decide on the structure of your essay. You are also expected to support your essay with appropriate references. Below are some optional topics:

(a) In many applications of visualization, many ordinary users (including some analysts and most decision makers) have found that learning parallel coordinates plots is more difficult than learning scatter plots. Discuss the possible reasons and suggest ways to teach parallel coordinates plots to analysts and decision makers.

(b) Most familiar visual representations are at least 300 years old. This seems to suggest that learning new visual representations is very difficult. Do you agree with this hypothesis? Explain your answer in detail with the aid of appropriate examples.

(c) There have been some heated debates about the merits and demerits of visual embellishments. Describe the central issue of this debate and present your view point on this topic.

(d) How do people acquire the skills of map reading, or spatial navigation? Is there a parallel between data visualization and map reading (or spatial navigation).