

Project Title: iChart - Interactive Exploration of Data Charts

eSTEEeM Final Report

Submission date: 19 July 2013

Project Leader: Michel Wermelinger

Key staff associated with the project: Paul Piwek

Contact: michel.wermelinger@open.ac.uk

Executive Summary

Data charts are used in many magazines, newspapers and official reports. Data journalism is on the rise. Big data is a familiar buzz word. And yet the public in general often does not have an easy way to explore the data and statistics that are thrown at them to justify a particular argument.

The iChart (Interactive Exploration of Data Charts) eSTEEeM project aims to contribute to the development of tools that allow users (and students in particular) to explore data sets in an interactive and simple way, without the learning curve associated to the professional statistical software packages.

In particular, the project has delivered iScatter, a prototype software that can be used across a range of disciplines and contexts because it produces a widely used chart type (points in a two dimensional space) for a widely used platform (a web browser).

In particular, iScatter reads multivariate data (i.e. where each data point is characterised by two or more attributes) exported from an spreadsheet, and displays it as an interactive scatterplot or bubble chart (a scatterplot that additionally maps the area of circles to a third attribute) within a web page. iScatter can show subsets of the data in different colours and can compute common statistics (median, maximum, upper quartile, standard deviation, etc.) on each subset, to make comparisons possible. Most of the statistics can be shown visually through a reference line. To avoid cluttering, information is only displayed on demand, when hovering the mouse over the subset, data point, or reference line.

Aims and scope of the project

The main aim was to improve numeracy skills by delivering a tool - an interactive visualisation tool with textual summarisation, to improve accessibility of data charts - and then try it out in some modules.

Details about the visualisation part are already given in the executive summary.

The aim of the second part, led by Paul Piwek, was to explore the addition of automatic graph summarisation capabilities to the iChart prototype. For this purpose, students were asked to investigate algorithms for automatically mapping numerical data to English summaries of those data.

The algorithms should use chart data compliant with the iChart specification.

Two students took up this project topic in the 2012-2013 cohort of M801 (Computing MSc final project). Both projects were completed with the following results:

One project explored specification of an algorithm for domain-independent summarisation of time series data. In particular, it investigated how to summarise graphs with two lines (in contrast with previous research on single line charts). It resulted in the specification and partial implementation of an algorithm which systematically maps segments of a two-line graph to sentences that verbalise the underlying trends. The approach was compared with human-

generated summaries and a number of recommendations were formulated for follow up research.

A second project developed a fully implemented prototype for summarising web analytics data. The prototype is compatible with the iChart data specification, but limited to the domain of web analytics. The domain specificity of the prototype makes its reuse in the context of the iChart project less attractive.

In summary, both projects worked with iChart compliant data, but only the first project yielded results that could in the long run be of benefit to iChart. Further work is however required to fully implement the algorithm specification from the first project. For this reason, the topic of "Graph Summarisation" is currently offered to the new cohort of MSc students, with the idea that a student that takes up this topic will build on the insights of the aforementioned project.

Activities

Unfortunately, the technical hurdles took all the available time.

There is no expertise in the OU for developing interactive graphs with modern technologies (SVG, Javascript) and therefore the initial LTS developer associated to the project was of little help and was himself overburdened with his own work. The interactive tool was hence basically developed by myself. A second developer, hired to make it look better and iron out some problems, also had a steep learning curve with the particular graphics library used and is currently still solving some problems.

Having become head of department also didn't help towards a quicker progression of this project.

There was an initial workshop to gather some requirements, and various stages of the tool were presented at eSTeeM events, the 2013 Learn About Fair, and an eSTeeM newsletter. This was very useful and feedback has been positive. The OpenScience Laboratory might be interested in using it. IET would also like to use it if it only worked in Internet Explorer 8.

Impact

There is none to report as the tool itself is not yet complete. There will be however follow-up actions in presenting the final version to the OpenScience Lab and module teams.

List of deliverables

iScatter will be made available. The interim version (not the one the second developer is currently finalising) is available at:

<https://dl.dropboxusercontent.com/u/5760429/iScatter/doc/demo.html>

Produced a tutorial which is available at:
<http://dl.dropboxusercontent.com/u/5760429/iScatter/doc/tutorial.html>