Chapter 60
Teaching Visualisation in the Age of Big Data:
Adopting Old Approaches to Address New Challenges

Belinda A. Chiera
University of South Australia, Australia

Malgorzata W. Korolkiewicz
University of South Australia, Australia

ABSTRACT

Technological advances have led to increasingly more data becoming available, a phenomenon known as Big Data. The volume of Big Data is to the order of zettabytes, offering the promise of valuable insights with visualisation the key to unlocking these insights, however the size and variety of Big Data poses significant challenges. The fundamental principles behind tried-and-tested methods for visualising data are still as relevant as ever, although the emphasis necessarily shifts to why visualisation is being attempted. This chapter outlines the use of graph semiotics to build data visualisations for exploration and decision-making and the formulation of elementary, intermediate- and overall-level analytical questions. The public scanner database Dominick's Finer Foods, consisting of approximately 98 million observations, is used as a demonstrative case study. Common Big Data analytic tools (SAS, R and Python) are used to produce visualisations and exemplars of student work are presented, based on the outlined visualisation approach.

INTRODUCTION

Recent technological advances have led to data collection at a rate never seen before, from sources such as climate sensors, transaction records, social media and videos, to name a few. With the advent of *Big Data* come insights at a previously unseen depth and breadth of detail. Operational decisions are increasingly based on data rather than experience or intuition (McAfee & Brynjolfsson, 2012) and more

broadly, a shift in perspective is under way on the relationship between data and knowledge generation (Ekbia et al, 2015).

Big Data is typically defined in terms of its *Variety*, *Velocity* and *Volume*. Variety refers to expanding the concept of data to include unstructured sources such as text, audio, video or click streams. Velocity is the speed at which data arrive and how frequently data change. Volume is the size of the data, which for Big Data runs to the order of petabytes (1000^5) through to zettabytes (1000^7).

Visualisation is a potentially valuable way to make sense of Big Data, to uncover features, trends or patterns to produce actionable analysis and provide deeper insight (SAS Institute Inc., 2014). There is an increased focus on visualisation over formal data analysis, partly due to the proliferation of easy-to-use web-based visualisation tools (e.g., Tableau Online), and partly due to an added emphasis on the power of well-designed visualisations and the demand for a new type of Big Data 'dataviz' analyst (McCosker, & Wilken, 2014). However, the opaque character of large data sets makes it difficult to describe in a systematic way how to effectively translate Big Data into visual or other kinds of knowledge. Furthermore, a 'black box' approach to generating data visualisations puts data analysts at risk of producing ornate and visually pleasing graphics that are otherwise useless.

The use of visualisation as a tool for data exploration and/or decision-making is not a new phenomenon. Data visualisation has long been an important component of data analysis, whether the intent is that of data exploration or as part of a model building exercise. However, the challenges underlying the visualisation of Big Data are still relatively new; often the choice to visualise is between simple graphics using a simple palette of colours to distinguish information or to present overly complicated but aesthetically pleasing graphics, which may obfuscate and distort key relationships between variables. Previous work in the literature suggests a tendency to approach Big Data by repeating analytical behaviours typically reserved for smaller, purpose-built data sets (e.g., Gelper, Wilms, & Croux 2015; Toro-Gonz´alez McCluskey, & Mittelhammer 2014; Huang, Fildes, & Soopramanien 2014). There appears, however, to be less emphasis on the exploration of Big Data itself to formulate questions that drive analysis.

In this chapter, a portable framework is proposed to arm aspiring visual analysts with tools to decide what is/is not useful when visualising Big Data. The target audience is working professionals seeking to expand or formalise their data analytics skills through postgraduate university study. Based on the authors' teaching experience in this context, introducing a framework for visualising Big Data through real-life case studies is advocated, instead of presenting students with an encyclopaedia of methods and graphical displays.

Despite the new challenges introduced by the emergence of Big Data, visualisation techniques themselves need not be new – *lessons from the past* in the form of tried-and-tested graphs can still be effective – however with a new way to perceive the data. This chapter adopts the seminal work of Bertin (1967) and extensions in Tufte (1983) in the field of *graph semiotics,* to explore and harness the scope of Big Data to formulate meaningful questions and drive analysis. To illustrate graph semiotics, the Dominick's Finer Foods (DFF) scanner database is used. Dominick's Finer Foods (DFF) is a Chicago supermarket chain and its scanner database is publicly available at https://research.chicagobooth.edu/kilts/marketing-databases/dominicks courtesy of James M. Kilts Center, University of Chicago Booth School of Business. The use of graph semiotics is further demonstrated to formulate three levels of questions from the supermarket database: elementary-, intermediate- and overall-level, before presenting conclusions.

There are many visualisation tools available; some come from established analytics software companies (e.g., Tableau, SAS or IBM), with others emerging as open source applications. For the purposes

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/teaching-visualisation-in-the-age-of-big-data/217881

Related Content

An Efficient MapReduce Computing Model for Imprecise Applications
Changjian Wang, Yuxing Peng, Mingxing Tang, Dongsheng Li, Shanshan Liand Pengfei You (2016). *International Journal of Web Services Research (pp. 46-63).*
www.irma-international.org/article/an-efficient-mapreduce-computing-model-for-imprecise-applications/161802

An Efficient Service Discovery Method and its Application
Shuiguang Deng, Zhaohui Wuand Jian Wu (2012). *Innovations, Standards and Practices of Web Services: Emerging Research Topics (pp. 382-404).*
www.irma-international.org/chapter/efficient-service-discovery-method-its/59932

Web Service Enabled Online Laboratory
Yuhong Yan, Yong Liang, Abhijeet Royand Xinge Du (2012). *Innovations, Standards and Practices of Web Services: Emerging Research Topics (pp. 363-381).*
www.irma-international.org/chapter/web-service-enabled-online-laboratory/59931

Using Web Service Enhancements to Establish Trust Relationships with Privacy Protection (Extended and Invited from ICWS 2006 with id 47)
Alfred C. Weaverand Zhengping Wu (2009). *International Journal of Web Services Research (pp. 49-68).*
www.irma-international.org/article/using-web-service-enhancements-establish/3133

On Feature Interactions Among Web Services
Michael Weissand Babak Esfandiari (2005). *International Journal of Web Services Research (pp. 22-47).*
www.irma-international.org/article/feature-interactions-among-web-services/3068