<u>The visualisation of data in a digital context</u>

David A. Ellis[1]

Hannah L. Merdian[2]

[1]Department of Psychology, Lancaster University, Lancaster, UK

[2]School of Psychology, University of Lincoln, Lincoln, UK

David A. Ellis

Lancaster

Lancashire

LA1 4YF

email: d.a.ellis@lancaster.ac.uk

telephone: +44 (0) 1524 593207

1. Introduction

Science is often driven by data sets and a curiosity to explore them. While these investigations are typically guided by hypotheses, acting as a lens during exploration, good study design also allows researchers to create a story with their data (McCandless, 2009). The advent of digital research methods has led to the emergence of larger and multifaceted data sets, be it through the appearance of new platforms for established research methods (e.g., online social media) or the generation of mobile and cloud-based systems for data capture (e.g., smartphones); thus, transitioning from raw data to a story that can be shared with others has become increasingly complex, often requiring the simultaneous development of computational models alongside larger and constantly evolving data sets (McGrath, & Scanaill, 2013).

In psychology and the social sciences more broadly, two main issues have emerged as a consequence of increased digitalisation; (1) the ethics of social research, with regards to maintaining ethical research practices alongside secure data ownership, and (2) the approach to the collection, management and analysis of very large data sets. While the former remains a challenging (and complex) topic in itself, we have dedicated this chapter to the exploration of the latter, which can also be termed "scientific storytelling". We have argued previously that the digitalisation of psychological research requires a "digitalisation" of our approach to data analysis and data visualisation, which could provide new benefits for psychological research, including applied domains within clinical and forensic settings (Ellis & Merdian, 2015). This also provided examples of interactive data visualisation as a way to enhance the communication and teaching of basic and complex statistical functions (Lowe, 2003). Therefore, the aim of this chapter is to introduce interactive visual graphics as one response to the digitalisation of psychological research. We will consider current challenges and the development of new visualisations before discussing the future impact on psychology as a whole.

2. Data visualisation in psychology

While psychological scientists have traditionally designed and conducted research to test predictions, visualise data, and provide statistical output, the introduction of digital data has changed the research environment dramatically. For example, it has now become standard practice to run experiments and surveys online. Psychologists are now able to access pre-existing administrative data sets or access increased samples through the use of mobile technology (Ellis & Jenkins, 2012; Woods, Velasco, Levitan, Wan, & Spence, 2015). Field research also has the capacity to produce very large data sets (e.g., Andrews et al., 2015), both in terms of the absolute sample size as well as the amount of information collected about each participant (Keuleers & Balota, 2015). Similar to theoretical paradigm development, research methods evolve gradually with accumulative knowledge and in response to social processes (Kuhn, 2012) and psychological research has indeed made considerable progress when it comes to exploratory, non-linear research designs (Columb & Sagadai, 2006; Huberty & Morris, 1989; Rothman, 1990; Sainani, 2009; Saville, 1990). However, developments concerning other statistical practices, data management, and visualisation have not kept pace with these developments[1].

Historically, psychological research has seen a number of paradigms with regards to what defined the generation of “new knowledge”. In 1987, Gigerenzer and Murray identified the *inference revolution*, the adoption of significance testing, as the essential paradigm for conducting empirical research and an essential component if research was to become published. But a shift in focus from significance testing to the exploration of effect sizes and confidence intervals (Sainani, 2009; Schmidt, 1996; Wilcox, 2006) suggests a move away from the significant *vs.* non-significant dichotomy that has dominated the empirical literature in social sciences. This correlates with a gradual rise in innovative methods for visualising and interpreting data (Friendly, 2008), allowing a more flexible approach to data management. However, despite the importance of visualisation and exploration of data for psychology as a discipline, it has failed to become integrated into standard statistical practice (Gelman, 2004; Zuur & Leno, 2016). What is often presented within or as a supplement to existing publications does not reflect recent advances in data collection

[1] Many developments concerning data visualisation have taken place in other disciplines (e.g., computer science). However, these have never been universally adopted across the social sciences.

and analysis. The communication of most research continues to rely on numerical or visual representations that are centered around the concept of static, paper-based journals.

In reality, almost all of our standard visual representations of data remain grounded in developments of almost 100 years ago. Friendly (2009) defined the second half of the 19th century as *The Golden Age of Statistical Graphics*, due to

> *the unrestrained enthusiasm not only of statisticians, but of government and municipal authorities, by the eagerness with which the possibilities and problems of graphic representation were debated and by the graphic displays which became an important adjunct of almost every kind of scientific gathering* (p. 505).

According to Friendly, this rapid development was due to a combination of the increasing quantification of social data, advances in statistical thinking, and advances in technology that allowed images to be created more quickly. In contrast, in the first half of the 20th century, Friendly and Denis (2000) argued that visual data representations largely disappeared from many journals, particularly within social and personality psychology, returning to representations in the form of numbers and tables. This move was partly driven by the development of statistical theory, which led to the standardised implementation of correlation and regression (Stanton, 2001). By 1908, Gosset had developed the t-test and between 1918-1925 followed by Fisher, who extended early ideas underpinning the ANOVA and experimental design further (Lovie, 1981). Through the enhancement of statistical theory in psychology, numbers came to be viewed as more precise than graphs, with visual representations of data used predominantly to support rather than confirm the statistical narrative [Figure 1].

[Figure 1: Typical static data visualisations used in psychology (a) scatter plot, (b) line graph, (c) bar graph and the less common (d) box plot]

These static graphical illustrations may remain perfectly adequate in some instances; however, their static nature can lead to a lack of transparency and information loss. In many cases, they can be fundamentally misleading (Allen, Erhardt & Callhoun, 2012). For example, Ansccombe’s Quartet [Figure 2] reveals how very different distributions of data can be associated with numerically equal summary statistics; these different distributions would be readily disguised within traditional data portrayals, such as bar graphs. In addition, the benefits of illustrating data and its associated distributions that were previously emphasised in many publications and introductory statistical textbooks are frequently ignored within academic publications (Zuur & Ieno, 2016). For example, data visualisations can typically provide insights concerning fundamental assumptions required by traditional statistical tests (e.g. normality). The lack of appropriate data visualisation in psychological journals has resulted in a lack of concrete examples that directly relate to psychological data, leading to new developments being frequently overlooked even within statistics itself (Gelman, Pasarica & Dodhia, 2002).

[Figure 2: Ansccombe's Quartet effectively demonstrates how very different distributions of data can return identical summary statistics]

We have previously summarised some of the limitations of static data visualisation, pointing to the restricted depth of these representations that are not adequate as larger and/or more complex data sets emerge (Ellis & Merdian, 2015; Heer & Kandel, 2012; Weissgerber, Milic, Winham, & Garovic, 2015). Indeed, the use of static graphs, such as bar charts or box-plots, as a standard way of reporting data can still lead to information loss that can be alleviated even by a small addition of interactivity. For example, interactive visualisations allow users to display the raw data points on top of any plot, thus instantly revealing the consistency of any underlying effect; these data points can then be hidden so the data is displayed in a summary format. Similarly, a trend line that is hidden or revealed can allow for the additional exploration of any linear (or non-linear) trend. Moreover, recent research has demonstrated that static data representations such as bar graphs are routinely misinterpreted by students (Newman & Scholl, 2012), with Moreau (2015) suggesting that visual and dynamic visualisations may be more appropriate when teaching complex statistical concepts. To exemplify this, in a previous paper, we presented an interactive regression as a teaching example where the model could be recalculated in real-time according to the research question, for example to investigate differences between male and female participants separately [Figure 3].[2] In addition, these visualisations in isolation raise additional questions about the data itself or may encourage an alternative analysis; here, in contrast to static figures, dynamic representations can provide a limitless supply of additional information (Ellis & Merdian, 2015).

[Figure 3: An interactive dashboard that provides a range of visualisaion and analysis options (Ellis & Merdian, 2015)]

In response to the digitalisation of psychological research methods, data figures have and will continue to evolve in order to maximise their expressive power, both with regards to conveying the content and structure of the data as well as informing the analysis process (Campitelli & Macbeth, 2014; Marmolejo-Ramos, 2014). So far, this has included computational developments, such as the expansion of boxplots to include information about both distribution and density of the data (Marmolejo-Ramos & Matsunaga, 2009; Marmolejo-Ramos & Tian, 2010) or explorations of different data visualisations for particularly skewed data sets (Ospina et al., 2014). Many research groups have also started to develop new ways of visualising moderation analyses (e.g. Tay, Parringon, Huang & LeBreton, 2016). Of course, while visualisation can provide tools for driving domains of interest, its adoption and adaptation is ultimately determined by the ability to build accurate and relevant models that produce useful answers. Surprisingly, very few papers actually explain how to build accurate data representations, and available resources appear to neglect the possibility of dynamic, software-based visualisations (Fry, 2007).

3. Developing dynamic data visualisations

To exemplify the ease of application when creating interactive data visualisations from scratch, we have developed both a website and related materials that will help anyone build simple visualisations and statistical applications that can quickly be deployed online[3]. These use *Shiny* (https://shiny.rstudio.com/), a web application framework for *R* (http://www.r-project.org). *Shiny* allows for the rapid development of data visualisations that enable researchers, practitioners, and members

[2] https://psychology.shinyapps.io/example3/
[3] https://sites.google.com/site/psychvisualizations/

of the public to interact with the data in real time and generate custom graphs and tables as required. The development of such dynamic visualisations previously required specialist knowledge of HTML, CSS or Java, which would often turn many researchers away at the outset. However, it is now possible to quickly produce interactive data visualisations that can supplement and support a variety of research findings. Digital researchers are therefore in a prime position to make meaningful contributions when it comes to developing innovative methods for visualising and interpreting findings.

At a universal level, the development of any interactive visualisation can be split into several smaller steps (Ellis & Merdian, 2015). This includes the preparation of data; here, it is crucial to ensure that the narrative the researcher is aiming to convey with the data is easy to understand by a reader, without the need to conduct any additional analysis. At this stage, the researcher may conclude that static visualisation is adequate. However, dynamic visualisations will often add additional value and transparency to the dataset, particularly when it comes to enabling reviewers and other interested parties to explore the data in a fashion that is not reported within the original publication (Tay, Parringon, Huang, & LeBreton, 2016).

Following our earlier discussion concerning the limitations of bar graphs, the working example below demonstrates the value of interactive visualisation even within a relatively small data set, and provides a further introduction into the development of interactive graphics.[4]

[4] Instructions and computer code required to build the example outlined in this chapter can also be found at the above address.

<u>Case Study: Predicting smartphone operating system from personality and individual differences</u>

This research study aimed to investigate if smartphones could provide clues about the individual behind the screen (Shaw, Ellis, Kendrick, Ziegler & Wiseman, 2016). Extended Self-Theory suggests that the greater power and control a person has over an object, the more it becomes part of his or her self-identity (Belk 2013). Aspects of a smartphone, including the type of operating system, can therefore be considered within this context and one can make sensible predictions grounded in the context of brand personality, which is quantifiably different between Android and iPhone smartphones.

This study therefore employed a straightforward between-subjects design where the researchers wanted to compare differences between two groups of smartphone users based on their operating system of choice (*iPhone* or *Android*). These as these devices account for over 90 percent of all smartphones sold worldwide (IDC 2015) and the final sample in this study consisted of 310 iPhone owners and 219 Android users. Participants completed a standard personality assessment, which measured 6 factors of personality (Ashton & Lee, 2009). They were also required to answer a number of additional questions about themselves and their attitudes towards smartphones.

The data can be reported in several ways, depending on the original research questions; for example, comparing means alongside a series of t-tests will reveal differences between the two groups. However, a more advanced analysis may subsequently aim to develop several regression models that could then predict future smartphone ownership based on the differences observed between the two groups.

Quantifying and assessing group differences remains a key part of psychological science and traditionally, if one wanted to test for a significant effect between two independent groups, they would compare the differences between means using several t-tests. However, these, like ANOVAs, are not always robust to outliers, particularly when only a small sample is available (Wilcox & Keselan, 2003; Wilcox, 2012). Similarly, measures of effect size including Cohen's *d* are also not robust in this respect (Wilcox, 2006). Graphically, even with a seemingly straightforward data set like this one, static data visualisations (a single table or graph) could lead to information loss and incorrect conclusions. For example, a bar graph displaying the mean difference between two groups or conditions does not inform the reader about variance or outliers. Here, even a small addition of interactivity to the visualisation and statistical analysis can provide considerable insight. Displaying the original data points that make up each result is highly revealing when it comes to any outliers that might be driving significant results. Some argue that displaying raw data points leads to a rather messy and cluttered visual presentation, but dynamic presentation actually provides the best of both worlds in this instance - transparency of analysis alongside some detailed and informative visualisations.

To exemplify, we have built an interactive visualisation plot of the outlined study above using *Shiny* (https://psychology.shinyapps.io/smartphonepersonality/). The controls on the left allow you to explore the data in more detail, and compare the impact of any combination of different variables.

This data explorer allows anyone with a basic knowledge of statistics to become familiar with the data set. To the left, a direct link to the location of the data is visible. Copying the URL link into any web browser will allow anyone to download the original data set. The x-axis variable will default to *Smartphone*, which in this case is a categorical variable. Changing the y-axis will update the data display accordingly and allow a visitor to explore the differences between Android and iPhone users. Individual data points in the scatter graph can also be coloured based on a specific variable.

All graphics on the right will update in real-time based on the selections made on the left. This includes scatter plots and box-plots. It is also possible to zoom into specific areas of these graphs by dragging the mouse over a specific area and double-clicking the mouse in the selected zone. It is possible to reset the viewpoint by double-clicking anywhere on the graph. Finally, the table will update with data points that are within the area of interest.

This simple visualisation can be adapted to suit other data sets where two groups or conditions are being compared. Provided the raw-data set is in a similar format, the data will simply transfer through the application within little or no modification to the existing code.

4. Into the cloud: Generating impact with digital dashboards

Static visualisations typically become exponentially more difficult to understand as the complexity of the content they aim to display increases (Teknomo & Estuar, 2014). However, with careful management, dynamic data visualisations can provide a data format that is more accessible for a variety of user groups including other researchers, practitioners, interested members of the public, and students.

Benefits to the research community

Given the sheer number of scientific publications that appear every day, the challenge for researchers not only concerns getting papers published, but ensuring that others notice and engage with their work. This is before any attempt is made to reach out to other interested parties and stakeholders affected by the research findings. Providing access to an interactive storyboard, which can be shared with the public and journalists, is likely to lead to increased numbers of people engaging with the research; in the long-run, this may lead to more citations (Piwowar & Vision, 2013)[5].

Perhaps most importantly, the addition of dynamic data visualisation tools may also encourage active data sharing and transparency within the analysis. It is almost impossible for a reviewer to keep their statistical expertise current given the rate of progress concerning analysis (Siebert, Machesky, & Insall, 2015). Therefore, ensuring that the process of analysis and visualisation is clear will help share these developments, while simultaneously illuminating results. Computer code that sits alongside visualisations further enhance this usability (Gorgolewski & Poldrack, 2016).

An online, interactive dashboard also encourages good data management particularly if the associated data is also available from an online repository or directly from within the visualisation. Open data is almost certainly going to become standard practice for psychology in the next few years, with the 2015 RCUK Concordat on Open Research Data (Grand, Davies, Holliman, & Adams, 2015) stating that good data management is fundamental to all stages of the research process

[5] Other advocates of dynamic visualisation have suggested that academic papers as a whole need to be re-designed for computer screens that are not space orientated like traditional physical journals http://worrydream.com/ - !/ScientificCommunicationAsSequentialArt

and should be established at the outset; all data must be curated so that it is accessible, discoverable, and useable.

The benefits of data dashboards for psychology are not simply limited to improved understanding and dissemination of research (and the related data), but also feed into issues surrounding transparency and replication. Data sets can be now be issued and cited with DOI numbers, and these might be taken into account when it comes to future grant applications or replication research. But this data will only become valuable if it is visible and easily accessible. This would allow researchers to revisit a paper's analysis and re-analyse the original data with improved methods in the future.

While cloud-based data repositories can improve research practices by ensuring that past research is integrated into current work, they still require a user-friendly interface that allows for rapid re-analysis and visualisation if they are to be successful. Given the fast nature of these developments and the associated ethical implications, it remains imperative that the 21st century digital researcher remains vigilant and pro-active throughout. For example, the ability to compare multiple or pairs of replications side by side is now possible by providing suitable user interfaces. Tsjui and colleagues (2014), for example, have recently developed the concept of Community-Augmented Meta-Analysis (CAMA), which involves a combination of meta-analysis and an open repository (e.g., PsychFileDrawer.org; Spellman, 2012). Developments such as these can improve research practices by ensuring that past research is integrated into current work, contributing to accumulative knowledge in psychology and beyond.

Benefits to society (bridging the practitioner researcher gap)

We have argued previously, that dynamic data visualisation and cloud-based data sharing has particular benefits for applied psychology, where research sits at the interface between science and practice (Ellis & Merdian, 2015). In forensic and clinical psychology, research is often based on single-case studies (e.g., clinical intervention research) or small and diverse sample sizes (e.g., clinical subtypes, offending populations). Frequently, a large amount of data is collected for each case, resulting in limitations when it comes to the type of analysis that can be conducted as the assumptions for linear testing are not fulfilled, for example with regards to offender profiling (e.g., Canter & Heritage, 1990; Goodwill, Lehmann, Beauregard, & Andrei, 2014) or offender classification and risk assessment (e.g., Merdian, Moghaddam, Boer, Wilson, Thakker, Curtis, & Dawson, 2016). Dynamic data visualisations are ideally placed to aid in the data display, exploration, outlier identification, and analysis of applied data sets, especially when it comes to more complex visual representation as required for Multidimensional Scaling. Using dynamic data plots, a researcher could easily display different types of Cluster Analysis (e.g., using different distance measures) and compare these alongside each other; or, by adapting examples outlined previously, develop an application that plots the progress of individual clients over several years, providing information on treatment change, outliers, and group trends over time. Thus, dynamic data visualisations hold specific appeal for applied psychology, and have the potential to become a shared platform for both scientists and practitioners.

Benefits for teaching

Such data-rich representations are likely to be helpful when teaching statistical concepts as data (and the relationships between variables) can be explored in real time, and from different angles, allowing the student to match their engagement with the data based on their skills and knowledge. Many undergraduate psychology students are now being taught how to use R, Shiny, and other related visualisation tools (Barr, 2016). This represents a sensible step forward for the next generation of academics and practitioners. However, little research exists on its effectiveness within an educational context (Valero-Mora & Ledesma, 2014) and until recently, very little was known cognitively about how interacting with dynamic information can facilitate learning (Scaife & Rogers, 1996). Many questions remain unresolved in terms of how visualisation can assist with understanding. Memorability, for example, may be a factor as, interestingly; less memorable visualisations tend to be very traditional in their presentation (e.g., line and bar graphs). People appear to be better at remembering visualisations that are colourful and include a human recognisable object, but that does not confirm that viewers fully understand the content that has been presented (Borkin, Vo, Bylinskii, Isola, Sunkavalli, Oliva, & Pfister, 2013). In addition, while an expert user may believe they have created something practical and aesthetically pleasing, much of the literature surrounding human-computer interaction repeatedly demonstrates how a seemingly straightforward system that an expert considers ‘easy’ to operate often poses significant challenges to new users (Norman, 2013). Future research is required in order to fully understand the effect interactive visualisations could have on, for example, a student’s understanding of complex statistical concepts.

5. Future challenges to dynamic data visualisation

The digitalisation of psychological research has sparked the development of new ways to collect, store, analyse, and visualise data; however, if we are to move into the productive use of digital data management, several challenges remain, most predominantly, how to deal with “big data”, how to respond to the lack of adequate software, and yet continue to encourage the psychological profession to adjust their own practice surrounding data management and visualisation within existing disciplinary constraints.

Dealing with “Big Data”

As technology has transitioned from lab-centred, computer-based systems towards Internet and remote collection, researchers are able to amass larger participant samples and thousands of data points from participants. For example, recent research has effectively demonstrated how we can better understand individuals with millions of data points derived from smartphones alongside a variety of other wearable and Internet connected devices (Piwek, Ellis, Andrews, & Joinson, 2016). As outlined above, the expansion of these data sets brings with it many benefits but also new methodological concerns. Debates surrounding issues of replication alongside data that is arguably no longer suitable for traditional inferential statistics continue to dominate the research landscape (Keuleers & Balota, 2015).

Classic statistical analysis was based on the great insight that not every member of a population needs to be examined to make inferential statements about the whole population, based on the likelihoods of certain events occurring. Unfortunately, if not managed correctly, the focus on big data, and the need to collect, and process, more and more information, can actually be a step backwards with regards to knowledge generation within psychology. Large numbers of variables can also mean that key effects are missed or misinterpreted. However, we would argue that dynamic visualisation could help overcome these issues. It allows the researcher to focus on key variables based on existing empirical and theoretical knowledge of interest in the first instance and build in complexity as required. Even if the work is purely exploratory in nature, large data sets allow researchers to consider multiple questions simultaneously, or different questions at different times, leading to the development of new hypothesises. Returning to our worked example, the dynamic data exploration provided an open-access reference point for other researchers, the popular press, and interested members of the public. This also helped stimulate a number of additional research projects. In summary, this dynamic data visualisation did not replace or repeat any statistical analysis within the original publication, but instead acted as a counterweight to help communicate and develop the research further (Shaw, Ellis, Kendrick, Ziegler, & Wiseman 2016).

Into the Cloud: The Development and Uptake of Responsive Visualisation Software

In the social sciences, little is publicised concerning new ways to visualise and organise large data sets, with new developments often confided to computer science (Ellis & Merdian, 2015). Software to develop interactive visualisations has existed long before the Internet, but today there remains a lack of commercially available software with a simple graphical user interface that can help psychologists develop interactive, data-driven graphics, which can be easily uploaded online. When it comes to analysing and visualising data, almost every psychologist trained in the UK will be aware of *SPSS*. However, visualisation functions within *SPSS* are limited, as the software is no longer as actively developed with many standard analysis and visualisation functions unavailable (JASP Team, 2016). Similarly, the user interface requires refinement and has become overly complex, which makes statistical errors more difficult to spot (Smith & Mosier, 1986). Excellent free alternatives to mainstream statistical packages include *GNU PSPP* (https://www.gnu.org/software/pspp/) and *JASP* (https://jasp-stats.org/) and for interactive visualisations *R* and *Shiny* portray accessible alternatives.

A handful of alternative, cloud-based solutions also now exist online that remove the need for any programme code. For example, *Plotly* (https://plot.ly/feed/) allows for the development of interactive dashboards without a single line of code. However, going beyond basic functionality will require additional time and funding. The most powerful solution at present involves the use of *D3.js* (https://d3js.org/), which is a JavaScript library for producing interactive data visualisations. However, this requires extensive knowledge from the user concerning other aspects of JavaScript. Commercial cloud services are easier to use[6], but lack the ability to customise specific visualisation requirements that are very specific to each research design and data set. *Shiny* on the other hand sits somewhere in the middle between usability and flexibility. A basic knowledge of *R* is essential, but many free online courses are avalibale and such training is frequently being built into postgraduate and even some undergraduate psychology courses (Barr, 2016). We would argue that recent technological advances will require social science courses to include basic coding modules as part of their research methods training.

Our examples demonstrate that the tools are relatively easy to access, and we hope this motivates the reader to consider their use in future work. The reality for many researchers is that in the digital domain, a small amount of programming knowledge can go a long way - dynamic visualisations can make a technically proficient user more productive, while also empowering students and those with limited programming abilities (Ellis & Merdian, 2015). At this stage, we would favour the flexibility of producing code to fine-tune each visualisation rather than to rely on commercial or point-and-click interfaces. Making these available as part of any publication is essential, for both the peer review process, future replication and the sharing of methodological developments.

Data Visualisation within Psychological Science

While these developments are exciting and promise many benefits for psychology as a whole, they have yet to become commonplace for for a number of

[6] *Tableau* has been used by online mainstream media outlets for many years, but, at the time of writing, lacks any statistical functionality without additional development http://www.tableau.com

reasons. As outlined, psychologists are currently ill-equipped to develop these visualisations and even with those skills, it takes longer to develop interactive visualisations in comparison to static graphs. This is further compounded when there is no agreed software platform that will help psychologists develop these new tools.

However, it is our view that open data and the additional transparency offered by dynamic visualisations is going to become standard practice for a number reasons. Firstly, as the issues surrounding research impact are becoming unavoidable, more psychologists will start to use additional methods, including dynamic visualisation, to make these data sets usable and accessible to a larger audience. Secondly, large funders in the UK and abroad increasingly require data to be made usable and available to other researchers. Finally, there is a growing consensus that regardless of funder requirements, data analyzed as part of any publication should be freely available and useable when required by other researchers. Until this is built into the publication process directly, psychology will continue to deal with the associated fallout (Van der Zee, Anaya, & Brown, 2017).

Psychological journals are playing catch-up at this stage, but it remains likely that the same will apply when journals start to integrate interactive dashboards and visualisations as part of their publication output. While publishers may conclude that space remains an issue, this does not apply in a digital context. Graphics have also repeatedly been shown to actually take up less space than their table-based counterparts (Gelman, Pasarica, & Dodhia, 2004). Researchers themselves will govern the speed of this development and if there is a clear demand, old and new journals may start to support this additional interactivity within their publications (Weissgerber, Milic, Winham & Garovic, 2015). In doing so, psychology itself will lead the way ensuring that old and new data sets can, for the first time, become an essential resource in their own right.

Theoretical Frameworks to Guide Future Development

Finally, the broader introduction of interactive data dashboards calls for the consideration of a theoretical and structural framework underlying these developments. Significant attention has been paid to the graphics used to present data in a variety of contexts, particularly following influential work by Tufte (1983); however, this knowledge has not been integrated into theoretical developments. Often in science the theoretical background for a subject is considerable with little applied literature implementing it, but the literature on data visualisation is at the opposite end of the spectrum and often contradictory. Examples are plentiful in many journals concerned with quantitative analysis; however, few articles within psychology are dedicated to the theory of specific graphical forms. A new generation of statistical graphics as a whole, dynamic or static, may encourage the development of a theoretical rationale concerning visualisations in the digital domain for both quantitative and qualitative data. New evaluative methodologies could be built into future visualisations for example, to better understand the links between data visualisation and cognition. Scientists could then efficiently use these new tools to their advantage, particularly when attempting to get the attention of those outside of academia (Chen 2005; Chen, Haerdle, & Unwin, 2007).

Beyond theoretical contributions, dynamic data dashboards, like visualisations of the past will continue to depend on similar advances in technology, data collection and statistical theory (Friendly, 2008). However, researchers also have to reconsider some ethical aspects of their work, particularly when it comes to issues surrounding data ownership, and the long-term security of participants' data. For example, many digital data collection tools and devices collect identifiable information from participants, such as their IP address, and in some countries, even store the data on the collection platform (e.g., Qualtrics). The establishment of a ethical frameworks for digital data management and data sharing remains a pressing issue, but existing ethical guidelines for internet-mediated research may act as a useful starting point (British Psychological Society, 2013).

6. Conclusion

The current chapter was written in response to the increased digitalisation of research, especially within psychology and across social science. We argue that digital research methods have the potential to revolutionise data management in our respective disciplines but that it has so far failed to be recognised and/or implemented by the research community. We further argued that dynamic data visualisation provide great mechanisms to promote this change, and have exemplified their development with a worked example. We have described the benefits of dynamic data visualisations, and considered some of the challenges resulting from the digitalisation of psychological research as a whole. We hope this chapter will highlight the way towards an improved research toolset for the psychological community and help improve access to future research for both psychological peers and other interested parties.

Of course, we recognise that, at times, interactive visualisations will be unwarranted or less useful than static visualisations or tables, and distinctions need to be made between those that benefit experts and end-users with a more general interest. Perhaps more importantly, any form of data visualisation can only provide answers and insights into well-constructed questions. Dynamic data visualisation will never become a magic wand that can make sense from data chaos; however, careful execution and guidance combined with appropriate research questions will often ensure that simple and complex data sets are correctly interpreted and communicated in the digital domain.

References

Ashton, M. and Lee, K. (2009). The HEXACO-60: a short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345.

Allen, E. A., Erhardt, E.B. and Callhoun, V. D. (2012). Data visualization in the neurosciences: Overcoming the curse of dimensionality. *Neuron, 74,* 603-608.

Andrews, S., Ellis, D. A., Shaw, H. and Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-world smartphone use. *PLOS ONE. 10(10),* e0139004.

Barr, D. J. (2016). No more excuses: R is better than SPSS for psychology undergrads and students agree [Web blog post]. Retrieved from https://datahowler.wordpress.com/2016/09/10/no-more-excuses-r-is-better-than-spss-for-psychology-undergrads-and-students-agree/

Belk, R. W. (2013). Extended self in a digital world. *Journal of Consumer Research, 40*, 477-500.

British Psychological Society (2013). Ethics Guidelines for Internet-mediated Research. URL http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf.

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A. and Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics, 19(12),* 2306-2315.

Campitelli, G. and Macbeth, G. (2014). Hierarchical graphical Bayesian models in psychology. *Rev. Colomb. Estadística 37*, 319–339.

Canter, D. and Heritage, R. (1990). A multivariate model of sexual offence behaviour: Developments in 'offender profiling'. *The Journal of Forensic Psychiatry, 1(2),* 185-212.

Chen, C. (2005). Top 10 unsolved information visualization problems. *IEEE computer graphics and applications, 25(4),* 12-16.

Chen, C. H., Härdle, W. K. and Unwin, A. (Eds.). (2007). *Handbook of data visualization*. Springer Science & Business Media.

Columb, M. O. and Sagadai, S. (2006). Multiple comparisons. *Current Anaesthesia & Critical Care, 17(3),* 233-236.

Ellis, D. A. and Jenkins, R. (2012). Weekday affects attendance rate for medical appointments: Large-scale data analysis and implications. *PLOS ONE, 7(12)*:e51365

Ellis, D. A. and Merdian, H. L. (2015). Thinking outside the box: Developing dynamic data visualizations for psychology with shiny. *Frontiers in Psychology. 6*:1782.

Friendly, M. (2008). A brief history of data visualization. In *Handbook of data visualization* (pp. 15-56). Springer Berlin Heidelberg.

Friendly, M. (2009). The golden age of statistical graphics. *Statistical Science, 23(4), 502-535*.

Friendly, M. and Denis, D. (2000). Discussion and comments. Approche graphique en analyse des données. The roots and branches of modern statistical graphics. *Journal de la société française de statistique, 141(4),* 51-60.

Fry, B. (2007). *Visualizing Data*. O'Reilly: Canada

Gelman, A., Pasarica, C. and Dodhia, R. (2002). Let's practice what we preach: turning tables into graphs. *The American Statistician, 56(2),* 121-130.

Gelman, A. (2012). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics.14 (4),* 755-779.

Grand, A., Davies, G., Holliman, R. and Adams, A. (2015). Mapping public engagement with research in a UK university. *PLOS ONE, 10(4),* e0121874.

Goodwill, A. M., Lehmann, R. J., Beauregard, E. and Andrei, A. (2014). An action phase approach to offender profiling. *Legal and Criminological Psychology*. doi:10.1111/lcp.12069

Gorgolewski, K. J. and Poldrack, R. A. (2016). A practical guide for improving transparency and reproducibility in neuroimaging research. *PLOS BIOLOGY,14(7),* e1002506.

JASP Team (2016). *JASP (Version 0.7.5.5)* [Computer Software]

Heer, J. and Kandel, S. (2012). Interactive analysis of big data. *XRDS: Crossroads, The ACM Magazine for Students, 19(1),* 50-54.

Huberty, C. J. and Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological bulletin, 105(2),* 302.

IDC. (2016). Smartphone OS market share, Q2. Retrieved from www.idc .com/prodserv/smartphone-os-market-share.jsp

Keuleers, E., and Balota, D. A. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *The Quarterly Journal of Experimental Psychology 68(8),* 1457-1468.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago press.

Lovie, A. D. (1981). On the early history of ANOVA in the analysis of repeated measure designs in psychology. *British Journal of Mathematical and Statistical Psychology, 34(1),* 1-15.

Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *External and Internal Representations in Multimedia Learning, 13(2)*,157-176.

Marmolejo-Ramos, F. and Matsunaga, M. (2009). Getting the most from your curves: Exploring and reporting data using informative graphical techniques. *Quant. Methods Psychol, 5,* 40-50.

Marmolejo-Ramos, F. and Tian, T. S. (2010). The shifting boxplot. A boxplot based on essential summary statistics around the mean. *Int. J. Psychol. Res. 3,* 37–45.

McCandless, D. (2009). *Information is Beautiful*. Collins: London.

McGrath, M. J. and Scanaill, C. N. (2013). *Sensor Technologies: Healthcare, Wellness and Environmental Applications*. Apress: New York

Merdian, H. L., Moghaddam, N., Boer, D. P., Wilson, N., Thakker, J., Curtis, C. and Dawson, D. (Epub ahead of print). Fantasy-Driven Versus Contact-Driven Users of Child Sexual Exploitation Material Offender Classification and Implications for Their Risk Assessment. *Sexual abuse: a journal of research and treatment*.

Newman, G. E. and Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, *19*(4), 601-607.

Norman, D. A. (2013). *The Design of Everyday Things, Revised and Expanded Edition.* New York, NY: Basic Books.

Ospina, R., Larangeiras, A. M. and Frery, A. C. (2014). Visualization of skewed data: a tool in R. *Rev. Colomb. Estadística 37*, 399–417.

Piwek, L., Ellis, D. A., Andrews, S. and Joinson, A. (2016). The rise of consumer health wearables: Promises and barriers. *PLOS MEDICINE. 13(2),* e:1001953.

Piwowar, H. A. and Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ, 1,* e175.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology, 1(1),* 43-46.

Sainani, K. L. (2009). The problem of multiple testing. *PM&R, 1(12)*, 1098-1103.
Saville, D. J. (1990). Multiple comparison procedures: the practical solution. *The American Statistician, 44*(2), 174-180.

Scaife, M. and Rogers, Y. (1996). External cognition: how do graphical representations work? *International journal of human-computer studies, 45(2)*, 185-213.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1(2),* 115.

Shaw, H., Ellis, D. A., Kendrick, L. R., Ziegler, F. and Wiseman, R. (2016). Predicting smartphone operating system from personality and individual differences. *Cyberpsychology, Behavior, and Social Networking, 19(12)*, 727-732.

Siebert, S., Machesky, L. M. and Insall, R. H. (2015). Overflow in science and its implications for trust. *ELife, 4,* e10825.

Smith, S. L. and Mosier, J. N. (1986). *Guidelines for designing user interface software*. Bedford, MA: Mitre Corporation.

Spellman, B. A. (2012). Introduction to the special section on research practices. *Perspectives in Psychological Science. 7,* 655–656.

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education, 9(3).*

Tay, L., Parrigon, S., Huang, Q. and LeBreton, J. M. (2016). Graphical descriptives: A way to improve data transparency and methodological rigor in psychology. *Perspectives on Psychological Science. 11(5)*, 692-701.

Teknomo, K. and Estuar, M. R. (2014). Visualizing gait patterns of able bodied individuals and transitional amputees with the use of accelerometry in smart phones. *Rev. Colomb. Estadística 37,* 471–488.

Tsuji, S., Bergmann, C. and Cristia, A. (2014). Community-augmented meta-analyses toward cumulative data assessment. *Perspectives in Psychological Science, 9,* 661–665.

Tufte, E. R. (2001) [1983]. *The Visual Display of Quantitative Information* (2nd Edition), Cheshire, CT: Graphics Press.

Valero-Mora, P. and Ledesma, R. (2014). Dynamic-interactive graphics for statistics (26 years later). *Revista Colombiana de Estadística, 37(2)*, 247-260.

van der Zee, T., Anaya, J., & Brown, N. J. (2017). Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab. *PeerJ Preprints, 5*, e2748v1.

Weissgerber, T. L., Milic, N. M., Winham, S. J. and Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLOS BIOLOGY, 13*, e1002128.

Wilcox, R. R. and Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods, 8*, 254-274.

Wilcox, R. R. (2006). Graphical methods for assessing effect size: Some alternatives to Cohen's d. *Journal of Experimental Education, 74,* 353-367.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing.* Academic Press.

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X. and Spence, C. (2015). Conducting preception research over the internet: A tutorial review. *PeerJ*. e1058.

Zuur, A. F. and Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution, 7(6),* 636-645.