# Dynamic-Interactive Graphics for Statistics (26 Years Later)

### Gráficas dinámicas interactivas para estadística (26 años después)

Pedro Valero-Mora[1,a], Rubén Ledesma[2,3,b]

[1]Departamento de Metodología de las Ciencias del Comportamiento, Facultad de Psicología, Universitat de València, València, España

[2]Consejo Nacional de Investigaciones Científicas y Técnicas, Mar del Plata, Argentina

[3]Universidad Nacional de Mar del Plata, Mar del Plata, Argentina

### Abstract

This paper briefly reviews the history of dynamic-interactive graphics for statistics, introduces an example of such graphics, and provides a few glimpses as to the current state of things and the future trends we envision. The general conclusion is that dynamic-interactive graphics for statistics are thriving more than ever as they shift from the desktop to the internet. Thus, dynamic-interactive graphics are becoming increasingly important as they: 1) provide non-experts in statistics with the means to carry out analyses on their own; and 2) teach the basic concepts of statistics to students and practitioners with low to moderate mathematics skills. Their increasing popularity makes the lessons learned over the past 26 years of research on the subject more important than ever.

***Key words***: Data Visualization, Dynamic Graphics, Statistical Graphics.

### Resumen

Este artículo revisa brevemente la historia de las gráficas dinámicas interactivas en estadística, introduce un ejemplo y muestra del estudio actual del arte y las tendencias hacia el futuro. La conclusión general es que las gráficas dinámicas interactivas para la estadística son más populares que nunca y se han desplazado del computador a la Internet. Su importancia en ascenso se debe a que: 1) brindan a los no expertos en estadística los medios para llevar a cabo análisis; 2) enseñan los caminos básicos a estudiantes y practicantes con habilidades bajas o moderadas en matemáticas. Su popularidad hace que las lecciones aprendidas sobre los últimos 26 años sean más importantes que nunca.

***Palabras clave***: gráficas dinámicas, gráficas estadísticas, visualización.

[a]Professor. E-mail: valerop@uv.es

[b]Professor. E-mail: rdledesma@conicet.gov.ar

# 1. Introduction

This paper provides an overview of the history of dynamic-interactive graphics and offers some insights on their current state and a glimpse into their future. Dynamic-interactive graphics are graphics that can move smoothly and change in response to a data analyst's actions, with the changes being computed and presented in real time. We use a compound word because as its two components capture slightly different aspects of the graphics we are referring to. Dynamic was the term originally used in the literature to denote the property to change as opposed to traditional graphics, which are static. However, dynamic could be understood as driven by itself, as in a movie or an animation, or driven by the user's interests or intentions, as in computer-managed interactivity. We used this term in Young, Valero-Mora & Friendly (2006) and although its use is not universally accepted, others have also used it as well (Cook & Swayne 2007).

Twenty six years ago, in 1988, Cleveland & McGill (1988) edited a book titled *Dynamic Graphics for Statistics*. This book marked a turning point in the subject at hand, as it summarized many systems that were implemented on special hardware and also described systems based on personal computers which, at the time, were beginning to dominate the scene. Indeed, prior to the publication of this book, experimenting with dynamic-interactive graphics was restricted to a few fortunate researchers with access to highly expensive facilities. However, at the beginning of the 1990s, hardware capable of supporting these graphics had become sufficiently affordable for their use to become more widespread. The most significant event in the history of dynamic-interactive graphics has been their move to Internet, which have made them accessible to all of users not only those with formal training in statistics. In particular, we will discuss the impact that these graphics may have on the modern teaching of statistics and in providing public access to statistical databases. Before doing so, however, we will describe an example of a dynamic-interactive graphic so as to give the reader a sense of the concept: this example consists of applying the Box-Cox transformation (Box & Cox 1964) using a slider in ViSta (Young et al. 2006).

# 2. An Example of Dynamic-Interactive Graphics: Interactive Box-Cox

In this section we will introduce an example to illustrate the characteristics and appeal of dynamic-interactive graphics for data analysis. The data shows how these graphics can be applied to obtain visual transformations of variables, using in this case the Box-Cox (Power) family of transformations.

The Box-Cox (or Power) family of transformations is very useful for normalizing univariate distributions, often helping to make the variance of bivariate distributions homogeneous. A set of transformations capable of dealing with this kind of problem is the Box-Cox family of transformations (Hoaglin, Mosteller &

Tukey 1983). The formula for the Box-Cox family of transformations is:

$$f(y) = \left\{ \begin{array}{ll} (y^p - 1)/p & for\ p \neq 0 \\ log(y) & for\ p = 0 \end{array} \right. \tag{1}$$

Values of $p$ in (1) are usually considered in the range of $-1$ to 2. In this range they produce approximations to many well-known transformations. For example, a value of $p = 2$ gives the square transformation, $p = 0.5$ is the square root, $p = 0$ is the log transformation, and $p = -1$ is the inverse transformation. Values of $p$ below 1, are known to improve the distribution of positively asymmetric distributed variables. In turn, values of $p$ above 1, will benefit negatively asymmetric distributed variables. Generally, the analyst will try successive $p$-values, looking for plots of transformed data that conform to the intended shape; specifically, seeking symmetric histograms for individual variables and closely linear and homoscedastic scatterplots for pairs of variables. Univariate changes are not only useful per se, but also because of their effects on the multivariate distribution of the data. These effects are: 1) making non-linear relationships more linear; 2) promoting constant variability across groups; and 3) improving additivity of a response in relation to two or more factors (Hoaglin et al. 1983). For a complete view of the results of the transformation process, all these consequences in the data must be examined simultaneously. Trying different transformations to implement statistical graphics in a conventional way can be a cumbersome process. However, using a program with dynamic-interactive graphics makes very easy to display the effect of different transformations successively until the most satisfactory one is found. This is possible because of certain features, such as the linking of different graphic representations.

We will use the ViSta program (Young et al. 2006) to demonstrate dynamic-interactive transformations with data taken from the *World Almanac and Book of Facts* (1993), which Rossman (Rossman 1994) presented as a good example for teaching correlations and transformations. In ViSta, this dataset is included in a file named **tele.lsp** and contains the following variables on the forty largest countries in the world (according to 1990 population figures):

TABLE 1: Variables in the file tele.lsp.

| | |
|---|---|
| LifeExpec | Life expectancy |
| LifeExpecMal | Life expectancy for males |
| LifeExpecFem | Life expectancy for females |
| PeopleTV | Number of people per television set |
| PeoplePhy | Number of people per physician |

A preliminary examination of this data reveals a strong positive asymmetry in the variables **PeopleTV** and **PeoplePhy**. We will search for a Box-Cox transformation that reduces this asymmetry using a visualization in ViSta.

Figure 1 shows ViSta's multi-plot tool for performing and visualizing transformations in an interactive way. It includes a scatterplot matrix of the variables, a quantile plot of the variable currently selected (by clicking on the diagonal of the scatterplot matrix), a slider that controls the transformation parameter, a line

plot of the original data versus the transformed data, and a list of the labels of observations. Initially, the transformations are all linear and the first variable in the dataset is the selected variable.

The key to this visualization is the slider, that allows changing the $p$ parameter in (1). In fact, this parameter is linked to a slider so that manipulating the slider modifies the argument and determines which function will be used to transform the variables.

We call the combination of plots shown here a *spreadplot* (Young, Valero-Mora, Faldowski & Bann 2003) but others have used the term dashboard for a similar layout (Few 2004). In this spreadplot, the slider is linked to the currently selected variable. Hence, pointing and clicking the mouse on the arrows of this slider will apply the Box-Cox formula to this variable using the $p$-value shown on the slider. Remember that transformations with $p > 1$ make left-skewed distributions more symmetric and transformations with $p < 1$ have the same effect on right-skewed distributions. $p = 1$ means no transformation at all. The quantile plot in the Box-Cox spreadplot allows assessment of this effect since it is connected to the transformation of the variable currently selected. A data object for the transformed data is created when the Box-Cox menu item is used, and the data object is updated automatically. Thus, it is possible to check the transformed values at any time. The scatterplot matrix, providing us with information on the linearity and homoscedasticity of the bivariate relationships among variables, varies as changes are made in the slider. Finally, the line-plot named "power of the transformation" gives us information on the relationship between original and transformed data, and changes as the slider is moved.

The scatterplot matrix reveals several non-linear and non-homoscedastic relationships. For example, the relationship between **PeopleTV** and **LifeExpec** seems clearly non-linear. Not surprisingly, the relationship between **LifeExpec** and its subsets, **LifeExpecMal** and **LifeExpecFem**, is clearly linear. Box-Cox transformations with $p < 1$ will probably reduce this asymmetry and will bring the data closer to the assumptions of univariate and multivariate normality.

Successive values of $p$ can be tested until more satisfactory plot shapes are identified. The visualization in Figure 2 shows the set of transformations that the data analyst liked best. Values of $p$ were $-0.30$ for **PeopleTV** and $-0.35$ for **PeoplePhy**. The other variables were left untouched (hence $p = 1$). An examination of the matrix of scatterplots in Figure 2 reveals linear relationships between the transformed variables and the other variables.

In summary, this spreadplot allows the user to view the effects of a transformation on several plots simultaneously, thereby allowing to find the right transformation more efficiently.
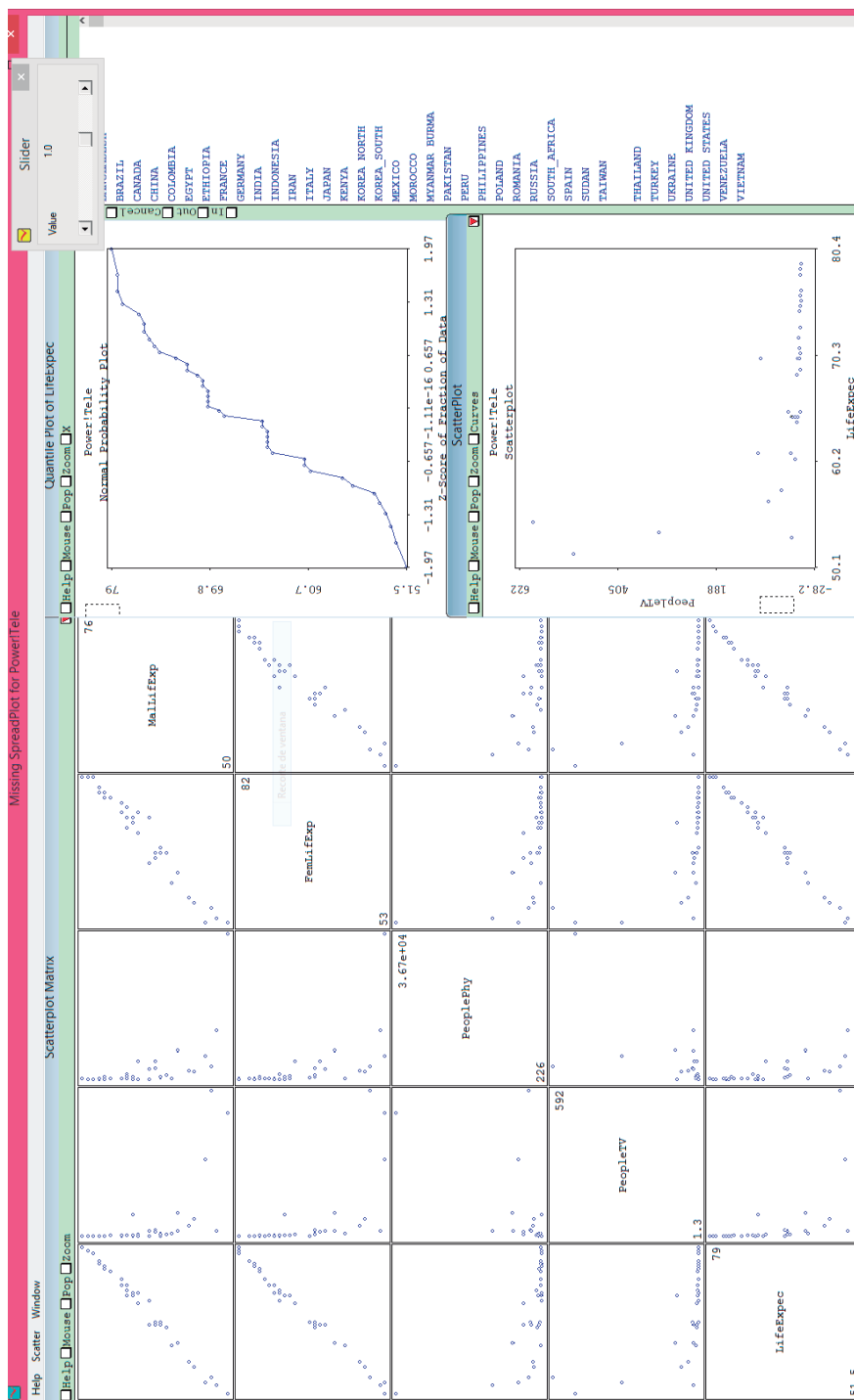
FIGURE 1: ViSta's Box-Cox visualization for the original (untransformed) variables.
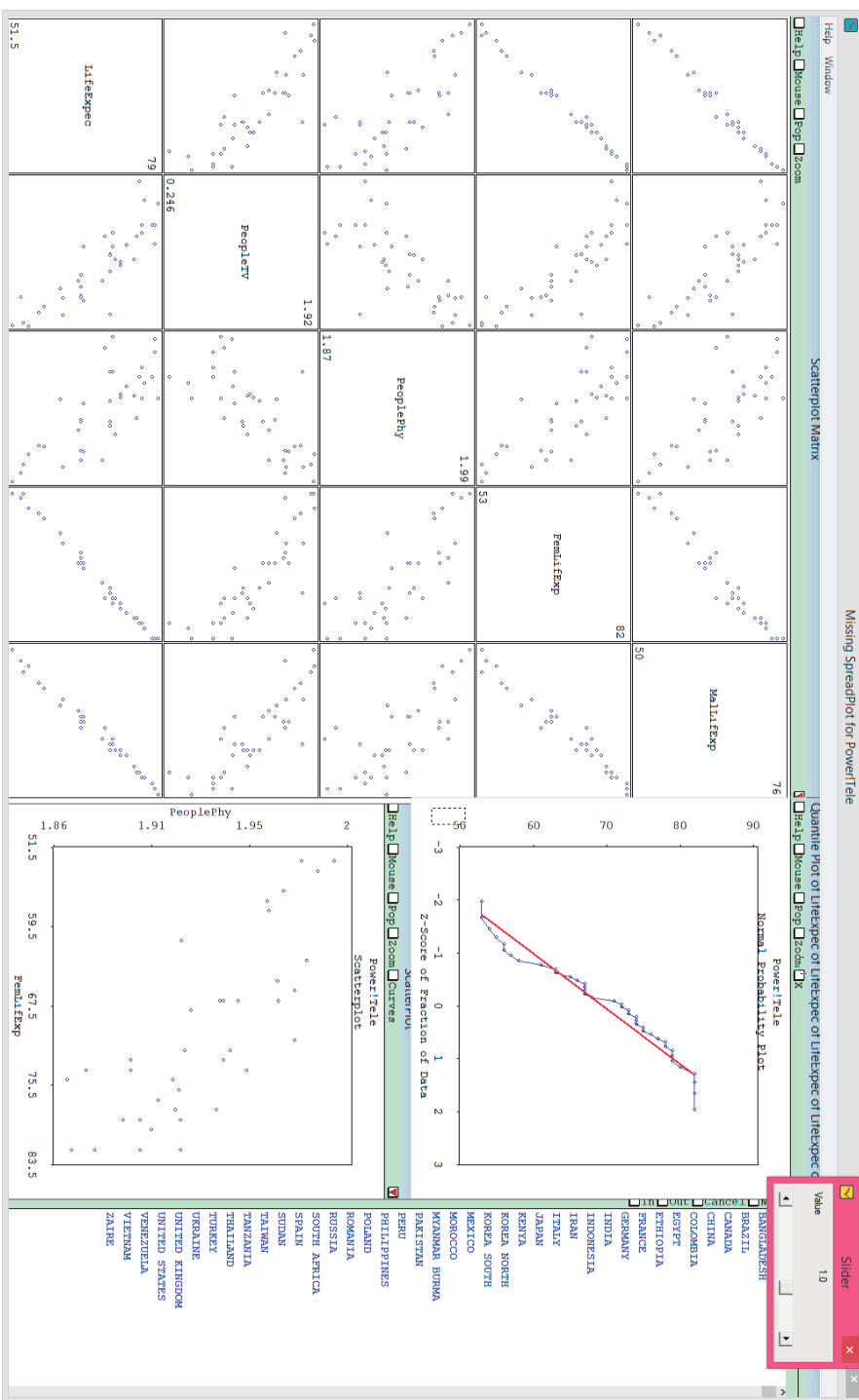
FIGURE 2: Box-Cox transformation spreadplot after applying the transformations to the variables. Note the linearity of the scatterplots.

# 3. A Brief Review of the History of Dynamic-Interactive Graphics

As mentioned in the introduction, the book *Dynamic Graphics for Statistics*, published in 1988, was the first collection on a series of different experiments on what was very innovative hardware and software, at the time. To a great extent, its chapters inherited Fisherkeller, Friedman & Tukey's (1975) work in Prim9 and covered topics such as real-time graphics for analyzing multivariate data, how to control graphics, interacting with scatterplots, handling different plots, rotation, etc. The book also included the names of programs that, several years later, would form part of the toolbox of those interested in dynamic-interactive graphics: `MacSpin`, `Explor4` and `VisuaALS`.

The years that followed witnessed the emergence of several commercial and non-commercial packages implementing concepts from the book and facilitating their use to everyone. Whereas some previous programs required the use of hardware which, in many cases, was priced in the millions, widespread accessibility to personal computers, mainly those with a graphical user interface, made it possible to run the software on computers that students could afford (as one of the coauthors of this paper can attest to). This GUI revolution in computing led to the development of two commercial programs that included many of the techniques described in the Cleveland and McGill book, as well as other new techniques. These commercial programs are `DataDesk` and `JMP`.

- `DataDesk`, developed by Velleman & Velleman (1985), included almost all the features required in a dynamic-interactive graphics program. Plots were linked by features, color, activation, state and so forth. Analyses were linked to plots, in a manner that removing some dots from a plot might result in excluding them from a dataset and, consequently, recomputing parameters and statistics. The list of features is very long and this is not the place to describe all; however, interested readers may download a demo version of this program covering all its features at `www.datadesk.com`. The program still performs very well and, used on modern hardware, is faster than it used to be. In fact, for those familiar with this program, it seems puzzling that it has not become much more popular and developed to the point of competing with the big names in statistical software.

- `JMP`, on the other hand, supported by `SAS`, one of the big names in statistical software, has undergone continuous development to this day. The very capable `JMP` package is at version 11, and includes a large array of statistical procedures mixed with statistical graphics, much in the same way of `DataDesk`. A demo version of the current version is also available.

A few months after the publication of the aforementioned book, a new language for statistics and graphic statistics was released into the public domain. This language was `Lisp-Stat` and it included a barebones Lisp programming environment developed by Tierney (1990) called `XLisp-Stat` (Valero-Mora & Udina 2005). This

system facilitated the task of exploring dynamic-interactive graphics significantly and had a strong impact on the development of statistical visualization systems. This should not seem odd, taking into account the three primary motivations that Tierney gave for developing this language: 1) To provide a vehicle for experimenting with dynamic graphics and for using dynamic graphics in education; 2) To explore the use of object-oriented programming ideas for building and analyzing statistical models; and 3) To experiment with an environment supporting functional data.

`XLisp-Stat` gave statisticians the opportunity to implement ideas related to dynamic graphics in much the same way the prior statistical language `S` had already provided a general statistical programming language (Becker, Chambers & Wilks 1988). In fact, some `XLisp-Stat` features were inspired by similar features or functions in `S`. One of the strong points of `XLisp-Stat`, as Tierney and others have mentioned, was the fact that it was based on `Lisp`, a general purpose high (and low) level programming language that was (and is) well known and mature. This guaranteed a solid technical foundation and a strong set of basic programming tools. Based on this language, two major projects, `ViSta` and `Arc`, implemented a number of techniques previously discussed in the literature or existing in other programs, and at the same time were very active in the development of new ideas and in promoting new advances.

However, these two programs were not the only ones experimenting with interactive graphics. Although a complete listing of such programs is beyond the scope of this paper, two deserve mentioning: `Manet` and `XGoby`.

- `Manet` is one of several statistical visualization programs developed by the computer-oriented statistics and data analysis group of the University of Augsburg (Unwin, Hawkins, Hofman & Siegl 1996). Manet originally focused on the visual estimation of missing values but later incorporated many other visualization techniques. It is particularly outstanding for visualizing categorical data.

- `XGoby` (Swayne, Cook & Buja 1998) is a data visualization system designed to explore high-dimensional data with graphical views that can be brushed and linked. These views include dynamic-interactive scatterplots, parallel coordinate plots and scatterplot matrices.

Many of these non-commercial programs have been merged with `R` and there is now a version of `XGoby` called `RGoby`, the research group that developed Manet is now behind `iPlots`, a package for the `R` environment incorporating many characteristics found in Manet but within a programmable environment. However, this activity on the non-commercial side did not receive an echo of similar success on the commercial side. Thus, at the turn of the millennium, what we believed was the next logical step: inclusion of interactive graphics as part of large statistical packages `did` not occur. For example, `SPSS` included a module of interactive graphics that was discontinued after a few versions. The other big statistical package, `SAS`, did not include this type of graphics, relying instead on JMP as a means to offer them. `DataDesk` developers, on the other hand, maintained their solid

product from the '90s without major changes, but did not manage to gain a large user base.

Apparently, dynamic- interactive graphics were losing traction. However, a new change in technology opened additional opportunities for interactive graphics. This change was the launching of the Internet as the platform for distributing statistical data and computing statistical analysis. New opportunities arising from this change are discussed below.

# 4. The Present and Future of Interactive Graphics

In recent years, the trend in computing has been to shift the functionality of desktop software to the cloud, and statistics software is no exception. So, in what it is an important advance, several new Internet-based systems have been designed for data to be delivered publicly, bundled with a set of tools allowing analysis and visualization of data. The availability of the data and the tools means that users can pursue their own interests, given that the system allows selecting the data, graphics, statistical summaries or models. It is worth noting that many of these potential users are not experts in statistics, and therefore systems must be carefully designed to avoid common mistakes they might make.

Another important development is the growing availability of software that can be used to teach statistics and allow students to interact with demonstrations and carry out certain analysis. We will first discuss the use of interactive graphics on the Internet and then continue with the topic of self-analysis.

## 4.1. Teaching Statistics

Teaching statistics to students and practitioners of areas other than statistics itself is an endeavor of giant proportions. Around the world, thousands and thousands of students without much formal training in mathematics are exposed to the rudiments of hypothesis testing, ANOVA, regression and other techniques considered basic by experts in statistics yet exceedingly difficult for novices. These students demand an education enabling them to understand applications, main concepts and usefulness of different statistical techniques, but without the burden of having to learn what many of them perceive as the useless intricacies of arcane methods. In his autobiography, Mosteller admits he was reluctant at first to accept the idea of teaching general concepts without some thorough understanding of theory, but after years of contact with experts in other, he changed his mind (Mosteller, Fienberg, Hoaglin & Tanur 2010).

Visual demonstrations of techniques, such as those in Valero-Mora & Ledesma (2011), are quite possibly a good way of achieving educational goals. In that paper, we showed how cluster analysis and factor analysis can be carried out by "hand" using simple actions such as rotating or selecting different views of datasets. The software described in the above paper gives students an understanding of what certain multivariate techniques can produce and how results can be interpreted

without going into details. This overview of the general usefulness of the techniques can be a first step in their education.

The idea that interactive visual displays are useful in teaching statistics is widely embraced, as evidenced by the fact that even a cursory internet search will turn up a number of demos and short demonstrations on many statistics topics. Additionally, the two leading introductory statistics handbooks by De Veaux, Velleman & Bock (2012) and by Moore (2011) incorporate demonstrations of this kind. Further, Xie (2013) provides a package for `R` to create animations that can be used to demonstrate statistics concepts. There are countless other examples, a complete list beyond the scope of this paper. There is broad consensus as to the possible benefits of this approach, although research on its effectiveness from an educational standpoint is not as abundant as the available software. With respect to interactive graphics, they are in many cases very close to techniques that were developed for actual analysis, and not only for purposes of teaching the analysis. Although we cannot other than applaud the efforts carried out in this direction, it is also true that this separation might limit the potential impact of interactive graphics in statistics. Moreover, students must learn two different software systems for statistics: one used for teaching the concept and another required for performing the analysis. Developers of statistics teaching software might claim that the software's learning curve is very low and the time required to master the program is negligible. Nonetheless, the literature on human-computer interaction often shows how apparently very-easy-to-operate devices often pose significant challenges to users (Norman 2013). Additionally, the material to be covered in statistics courses is already quite extensive in relation to available time, and it is therefore quite common more attention be devoted to software that will be used for real analysis whilst little or none is dedicated to interactive demonstrations in the classroom.

In our opinion, interactive graphics should not be left out of statistical packages so that its learning ought to be part of the basic tool kit that students are required to go through in introductory courses. In the previous section, we examined a Box-Cox transformation as an example. If this capability were included as part of any scatterplot in a statistical package, students might learn the importance of transformations with the same tool that can later be used to transform the data in practice. `DataDesk` and `ViSta`, for example, are statistical packages that include these two roles, and therefore may be used both for teaching purposes and for performing out the actual task.

In conclusion, teaching statistics is one area of application for interactive graphics with enormous potential, but we believe that for interactive graphics to yield the impact they deserve, they have to be where the real action is: in statistical packages and systems that are actually used in practice.

## 4.2. Self-Analysis by Non-Experts

In 2007, the OECD organized a workshop on dynamic graphics for presenting statistical indicators. Amongst the presentations, Hans Rosling's Gapminder

presentations (`http://www.gapminder.org/`) stood out because of their high impact media. As anyone who has played with Gapminder knows, its intuitive approach to public data analysis is both fascinating and instructive, and may even be used by non-experts to gain knowledge about complex issues which would otherwise need to be communicated in written reports lacking the joy of self-discovery. Gapminder is fun and has shown how statistical data can be more democratically distributed to all. Presently, supported by software such as Tableau, Spotfire, Qlikview, Quadrigram, and Google's data explorer, to mention a few, this approach is growing exponentially and will perhaps be more influential than the authors of the Cleveland and McGill book ever dreamt. New software allowing the creation of graphical user interfaces in `R` (Valero-Mora & Udina 2005) can have an even greater impact; one example is the shiny package (`http://shiny.rstudio.com`), which facilitates the task of building dynamic-interactive graphics and inserting them in webpages for programmers already familiar with the `R` language.

# 5. Discussion and Conclusion

We wish to make a few comments on this topic not only from the perspective of researchers who have been working on dynamic-interactive graphics for 15 years, but also from the standpoint of statistics teachers and, lastly, of people interested in the service that statistics may provide to the world. Let us address these views one by one.

As researchers on dynamic-interactive graphics, we observe that many developments currently applied actually ignore the base of knowledge that has been built over the past two decades. It seems that much of the software is reinventing ideas that have been explored previously and failing to give others demonstrating good potential the attention deserved. Of course, this is not unusual. It is well known that information visualization and statistical graphics communities work in parallel without a exchange of ideas between what look to be very close fields of research. Therefore, we do not take offense in not being considered as source of insight. In other words, to paraphrase Newton, being the dwarves on whose shoulders they stand on. However, it is not much fun at all to witness the reinventing of the wheel. We cannot help but be wary of developments that do not include sufficient expertise from seasoned statisticians. Exciting as they may be, systems that are developed by non-statisticians may lack the necessary expertise and be flawed in critical aspects. Examples abound and the comments made by Wilkinson (2001) about a software company are as relevant now as they were then:

> [said software company] implemented boxplots by placing the center at the mean, the edges of the box at one sample standard deviation on either side of the mean, and the ends of the whiskers at two standard deviations on either side of the mean. To this company, it appears, symmetry is the highest form of art.

As teachers of statistics, we regard these new systems as an opportunity. We can use these interactive systems in the classroom and our students will learn

with real data. This might make our lessons much more memorable although the pedagogical challenges are huge. However, we also feel that we should guide our students to be part of this effort. In the future, they might well be in charge of putting a front-end for the analysis of a database. We do not believe our students should have to learn the technology in our classes-this should come after the statistics-but perhaps the rudiments of human-computer interaction would be useful, as we have learned from our own experiences designing a good interactive graphic often involves decisions that fall into this domain. With this knowledge, a data analyst not skilled in computer programming may clearly convey to a computer scientist how things should be implemented. It is very likely that the end result of this collaboration would be much fruitful satisfactory if the statistician has an idea on how the user will interact with the computer.

Lastly, as statistics users in our own applied fields of research, we feel that dynamic-interactive graphics (and graphics in general) in actual practice do not receive the recognition deserved. Although the literature on statistical graphics has taken big steps in demonstrating their value, and highly influential books such as Cleveland's (1994) *Visualizing Data*, Tufte's (1983) *The Visual Display of Quantitative Information*, and Wilkinson's (2005) *The Grammar of Graphics* enjoy enormous popularity, their impact, in our opinion, falls far short of what it should be. Despite being limited, we are starting to see static graphics finding their way into academic journals and Wilkinson's (2001) impression that not even a boxplot could be found in the journals he had browsed is no longer the case. But dynamic-interactive graphics are a completely different story. Of course, the problem here is that a dynamic-interactive graphic tells a story that cannot be displayed easily using static media such as paper (or its electronic counterpart as we now very often read from a screen); a different form of expression is needed. Videos or movies are maybe the answer to the problem and in some cases they can illustrate the process of discovery that dynamic-interactive graphics are so clever in supporting. However, it is possible that other forms of expression may be better suited for the purpose of communicating something that has continuity but at the same time demanding significant pausing in order to appreciate specific aspects, easily overlooked in a constantly running movie. Mccloud's (1994) reflections on the formal aspects of sequential art provide a framework for thinking about media that is somewhere between the written word, static images, and animations. Also, new technologies now support document viewers that include all these different types of media seamlessly. Actually, some academic journals accept these new media types as part of their standard content. Thus, if the trend continues, it will probably not be far-fetched to believe that in the near future dynamic-interactive graphics will play a more important role in these new journals than today.

# References

Becker, R. A., Chambers, J. M. & Wilks, A. R. (1988), *The New S Language:A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole computer science series, Chapman & Hall, California, USA.

Box, G. E. P. & Cox, D. R. (1964), 'An analysis of transformations', *Journal of the Royal Statistical Society. Series B* **26**, 296–311.

Cleveland, W. C. & McGill, M. E. (1988), *Dynamic Graphics for Statistics*, Wadsworth & Brooks/Cole Statistics/Probability Series, Chapman and Hall/CRC, Boca Raton, FL, USA.

Cleveland, W. S. (1994), *Visualizing Data*, Hobart Press, USA.

Cook, D. & Swayne, D. F. (2007), *Interactive and Dynamic Graphics for Data Analysis*, Springer, New York.

De Veaux, R., Velleman, P. F. & Bock, D. E. (2012), *Stats: Data and Models*, 3 edn, Pearson, Boston.

Few, S. (2004), *Show me the Numbers*, Analytic Press, Oakland, CA.

Fisherkeller, M. A., Friedman, J. H. & Tukey, J. W. (1975), PRIM-9: A data display and analysis system, Pacific Regional Conference of the Association for Computing Machinery, San Francisco, CA.

Hoaglin, D., Mosteller, F. & Tukey, J., eds (1983), *Transforming Data*, John Wiley, chapter Emerson, J. D. and Stoto, M. A., pp. 97–128.

Mccloud, S. (1994), *Understanding Comics*, Harper Collins, USA.

Moore, D. S. (2011), *The Basic Practice of Statistics*, 5 edn, W.H.Freeman & Co Ltd.

Mosteller, F., Fienberg, S. E., Hoaglin, D. C. & Tanur, J. M. (2010), *The Pleasures of Statistics: The Autobiography of Frederick Mosteller*, Springer, New York.

Norman, D. A. (2013), *The Design of Everyday Things*, Basic Books, United States. Revised and Expanded Edition.

Rossman, A. J. (1994), 'Televisions, physicians, and life expectancy', *Journal of Statistics Education* **2**(2), 54–63.

Swayne, D., Cook, D. & Buja, A. (1998), 'XGobi: Interactive dynamic data visualization in the X window system', *Journal of Computational and Graphical Statistics* **7**, 113.

Tierney, L. (1990), *LISP-STAT: an Object Oriented Environment for Statistical Computing and Dynamic Graphics*, Wiley-Interscience, New York.

Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire.

Unwin, A. R., Hawkins, G., Hofman, H. & Siegl, B. (1996), 'Interactive graphics for data sets with missing values - MANET', *Journal of Computational and Graphical Statistics* **5**(2), 113–122.

Valero-Mora, P. M. & Ledesma, R. D. (2011), 'Using interactive graphics to teach multivariate data analysis to psychology students', *Journal of Statistics Education* **19**(1), 1–19.

Valero-Mora, P. M. & Udina, F. (2005), 'The health of Lisp-Stat', *Journal of Statistical Software* **13**(10), 1–5.

Velleman, P. F. & Velleman, A. Y. (1985), *Data Desk*, Data Description Inc.

Wilkinson, L. (2001), EDA and the Internet, Technical report, SPSS Inc.
\*http://www.spss.com/research/wilkinson/Publications/tukey.pdf

Wilkinson, L. (2005), *The Grammar of Graphics*, Springer, New York.

Xie, Y. (2013), 'Animation: An R package for creating animations and demonstrating statistical methods', *Journal of Statistical Software* **53**(1), 1–27.

Young, F., Valero-Mora, P., Faldowski, R. & Bann, C. M. (2003), 'Gossip: The architecture of spreadplots', *Journal of Computational and Graphical Statistics* **12**(1), 80–100.

Young, F. W., Valero-Mora, P. M. & Friendly, M. (2006), *Visual Statistics: Seeing Data with Dynamic Interactive Graphics*, 1 edn, Wiley-Interscience, New York.