

## Visualización Estadística: Narrar Historias con Software de Acceso Libre

MariCarmen González-Videgaray<sup>1</sup>, Rubén Romero-Ruiz<sup>2</sup> y Nora del Consuelo Goris-Mayans<sup>3</sup>

### Resumen:

---

La visualización, gracias a la existencia de grandes bases de datos que se actualizan y enriquecen automáticamente, al entorno web dinámico y a las herramientas de programación avanzadas, permite hoy en día narrar historias a través de estadísticas. La estadística se ha convertido, a partir de formas visuales claras e interactivas, en una forma de hacer conciencia sobre problemas importantes o de advertir relaciones antes insospechadas. Para cristalizar la visualización se requieren datos que pueden existir en muchos sitios, pero también aplicaciones de software con este propósito. Algunas de ellas son costosas y difíciles de aprender a utilizar. Sin embargo, internet ofrece un conjunto interesante de herramientas de libre acceso con esta finalidad. Por ello, el propósito de nuestra investigación es brindar un panorama de las aplicaciones libres para desarrollar visualización estadística y comparar tanto sus ventajas como sus desventajas. Para ello se realizó un perfil de búsqueda en internet, a partir del cual se extrajeron tanto aplicaciones que se usan a través de la web como otras que se instalan en la computadora del usuario. Se analizaron sus beneficios potenciales y su valor tanto didáctico como para crear conciencia social acerca de problemas reales. Se ofrece al lector una muestra de datos procesados con ellas y se invita a continuar indagando acerca de la visualización estadística, este año 2013: Año Internacional de la Estadística.

**Palabras clave:** datos, Processing, R, JavaScript, HTML5, programación.

### Abstract:

---

Visualization, thanks to the existence of large databases that are updated and automatically enriched, the dynamic web environment, and advanced programming tools, can now tell stories through Statistics. Statistics has become a way to raise awareness about important issues or to notice previously unsuspected relationships. To crystallize visualization, required data may exist in many places, but also there should be software applications for this purpose. Some of them are expensive and difficult to learn. However, the Internet offers an interesting set of freely available tools for this purpose. Therefore, the purpose of our research is to provide an overview of free applications to develop statistical visualization and compare both their advantages and disadvantages. For this we conducted a web search profile, from which we extracted both applications used across the web and others that are installed on the user's computer. We analyzed the potential benefits and value both teaching potential and capability to create social awareness of real problems. The study offers the reader a sample processed data and invites to continue inquiring about viewing statistics, this year 2013: International Year of Statistics.

**Keywords:** data, Processing, R, JavaScript, HTML5, programming.

**Modalidad:** Ponencia.

---

<sup>1</sup> Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Acatlán, México, mcgv@unam.mx.

<sup>2</sup> Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Acatlán, México, rubenr@unam.mx.

<sup>3</sup> Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Acatlán, México, goris.nora@gmail.com.

## 1. Introducción

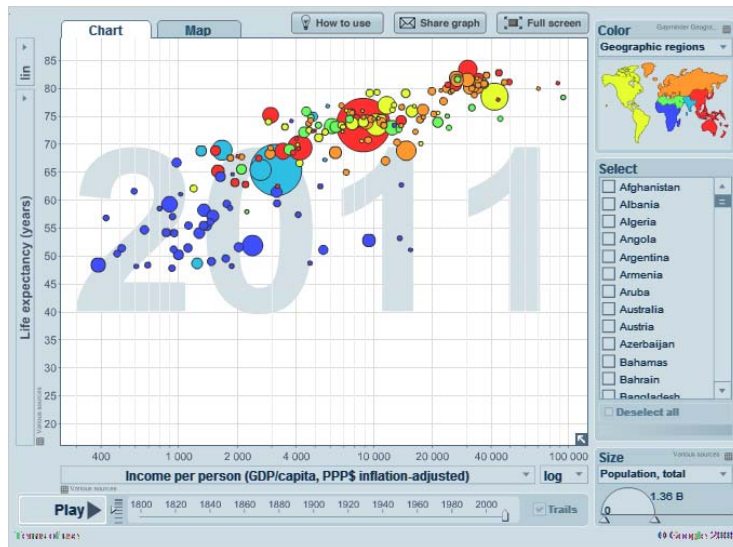
### 1.1 El arte de contar historias a partir de datos

El Año Internacional de la Estadística, 2013, ha sido testigo de avances notables en la visualización estadística. Este recurso se ha vuelto tan cercano y familiar que hoy en día, cualquier usuario de Facebook puede visualizar las estadísticas de su actividad personal en esta red social, de manera interactiva, con un solo clic. ¿Cómo? Bien, vayamos poco a poco.

En primera instancia, ¿qué es la visualización estadística? Autores como NathanYau[1] explican que se trata del arte de contar historias a través de los datos. En sus propias palabras: “Los datos son representaciones de la vida real. No se trata sólo de cubetadas de números. Existen historias en cada cubeta. Existe significado, verdad y belleza. Y tal como en la vida real, a veces las historias son simples y directas; otras veces son complejas y llenas de rodeos.” Sin embargo, en muchas ocasiones resulta difícil advertir estas historias si los datos se presentan como cuadros o tablas llenos de números. Este tipo de presentación, inclusive, puede ocasionar cierta repulsión en quienes la observan. Una forma de organizar los datos y presentarlos de manera más clara es recurrir a gráficos, es decir, a elementos visuales que pueden ser más fáciles de entender y atractivos.

Por supuesto, la estadística descriptiva ofreció desde sus inicios una gran variedad de elementos visuales[2], tales como histogramas; polígonos de frecuencia; gráficos de pastel, barras o columnas; diagramas de dispersión o series de tiempo, entre muchos otros. ¿Cuál es entonces la novedad de la hoy llamada visualización estadística? Veamos un par de ejemplos notables.

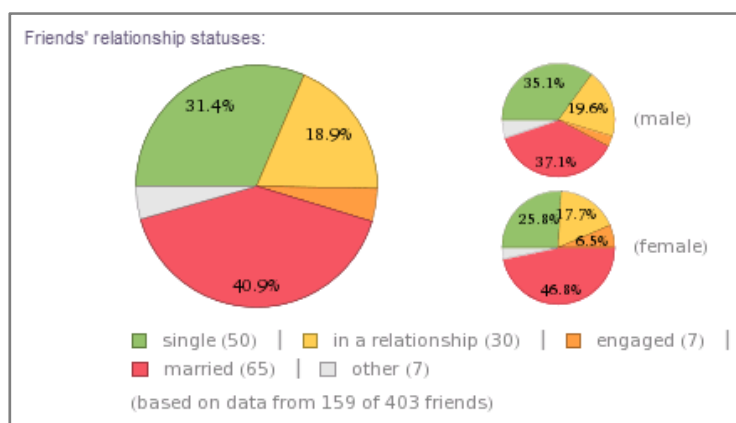
Hans Rosling[3] ha creado uno de los mejores sitios para visualizar estadísticas llamado GapMinder y la herramienta para ello, Trendalyzer, adquirida por Google en 2007 [4]. Una de las visualizaciones de este sitio es la que presentamos en la figura 1. Esta figura tiene una gran riqueza de contenido. Los colores representan los continentes. El tamaño de los círculos es proporcional al tamaño de la población. Además, la figura es dinámica: al colocar el apuntador sobre un círculo despliega sus datos; pueden marcarse países específicos y seguir su desarrollo; es posible ver de manera dinámica los datos correspondientes a los años 1800 – 2011, utilizando el botón *Play*. Esta visualización *cuenta una historia*, la historia del desarrollo histórico en cuanto a salud y economía de los países.



**Figura 1. Visualización estadística de la esperanza de vida vs. el ingreso per cápita. Fuente: (Rosling, 2013).**

Ahora, si se acude al sitio del motor de conocimiento WolframAlpha[5], puede efectuarse el registro (*Sign in*) a través de la red social Facebook (Elegir la opción: *Sign in with Facebook*). De esta forma se conectará WolframAlpha con nuestra información personal en la red social. Hecho esto, puede describirse en la ventana que brinda WolframAlpha para hacer cálculos, la frase: “*Facebook report*”, para luego oprimir el signo igual (=).

Con esto se obtendrá un conjunto muy interesante de resultados, de los cuales se muestran sólo dos en las figuras 2 y 3. En este caso particular, además de la presentación atractiva y clara, se agrega el hecho de que los datos son de gran interés personal: ¡son los datos de la actividad propia! Cuentan, además, una historia que uno no percibe en el mar de textos e imágenes colocados en la red. Es sólo hasta que los datos son organizados estadísticamente cuando es posible advertir tendencias y patrones [1].



**Figura 2. Visualización estadística del estatus de relaciones de amigos en Facebook. Fuente: (WolframAlpha, 2013).**

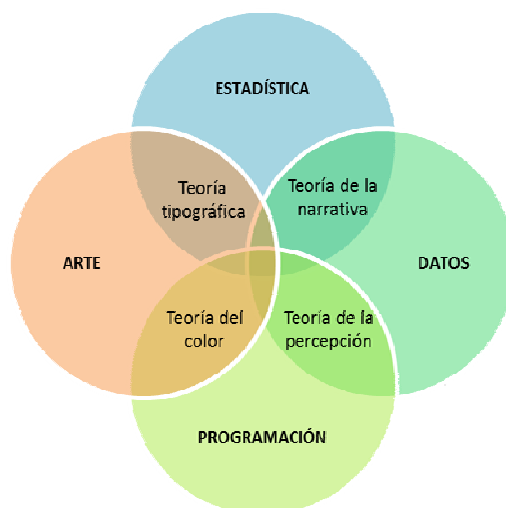


**Figura 3. Visualización estadística de las palabras más utilizadas en Facebook. La dimensión del texto es proporcional al número de veces que se ha mencionado. Fuente: (WolframAlpha, 2013).**

Como se advierte en estos ejemplos, la visualización estadística acerca la posibilidad de interpretar con facilidad grandes cantidades de datos, lo cual puede ser muy útil para los investigadores en general. Además, se convierte en un elemento con gran potencial para promover el aprendizaje de la estadística.

### *1.2 Qué se requiere para hacer visualización estadística*

Como se puede notar con los ejemplos anteriores, la visualización estadística es una técnica que requiere de la convergencia armónica de varias disciplinas (figura 4).



**Figura 4. Convergencia de disciplinas en la visualización estadística. Fuente: Elaboración propia.**

En primer lugar, para crear visualizaciones es indispensable contar con datos. Éstos pueden provenir de fuentes primarias, es decir, pueden ser recabados directamente por el investigador, o pueden surgir a partir de fuentes secundarias, tales como los sitios del Instituto Nacional de Estadística y Geografía, el Banco de México, el Banco Mundial, la

Organización para la Cooperación y el Desarrollo Económico, la UNESCO, o GapMinder, entre muchos otros. Todos estos sitios ofrecen, de manera gratuita y accesible, bases de datos acerca de muchos temas nacionales e internacionales. De hecho, la “avalancha” [6] de datos que hoy en día se generan representa, al mismo tiempo, retos y oportunidades para la visualización.

En segundo lugar, se requiere conocer posibles tratamientos estadísticos que revelen las historias contenidas en los datos. Para ello es necesario conocer la mayor parte de las estadísticas descriptivas. Sin embargo, con ingenio es posible crear visualizaciones de formas muy simples. Por ejemplo, en una hoja de cálculo de Excel donde aparezca un conjunto de datos, puede indicarse al programa que colorea las celdas de los números con diferente intensidad según el valor numérico. Esto permitirá descubrir rápidamente patrones o relaciones. La ventaja de la visualización estadística es que, si se tiene imaginación, será posible idear formas atractivas y claras de presentar información y no sólo datos.

Por otro lado y en tercer término, las grandes bases de datos sólo pueden explotarse de forma adecuada a través de lenguajes de programación que puedan manejar los valores y darles una representación visual lógica. Es decir, para convertir los datos en información estadística se requiere programar o usar herramientas que programen e indiquen a la computadora qué resultado se desea obtener. En este aspecto se basa precisamente este trabajo: pretendemos indagar acerca de las aplicaciones de software de acceso libre que permiten hacer visualización estadística.

Ahora, en cuarto lugar, es necesario hacer buenas representaciones visuales, es decir, buenas gráficas interactivas y dinámicas. Existen textos de gran utilidad para ello, como el de Cheny col.[2], pero hoy en día se reconoce que también inciden en este factor elementos artísticos[7], narrativos [8], creativos y psicológicos.

## **2. Metodología**

El objetivo del presente trabajo fue identificar las principales aplicaciones para hacer visualización estadística, que pueden ser de utilidad para manejo de datos o para la enseñanza de esta asignatura. Se pretende ofrecer una guía para quien desee aplicar estas herramientas a la investigación, al campo profesional o a la enseñanza.

Con el fin de localizar aplicaciones para hacer visualización estadística, se realizó una búsqueda en Google con el siguiente perfil:

(data OR statistic OR statistical) AND (visualization) AND (tool OR software OR site)

El perfil se refinó para detectar cuáles de las aplicaciones localizadas constituyen software de acceso libre. Dentro de esta búsqueda se localizó un artículo de Brian Suda [9] y una selección de [10] que resultaron excelentes guías para el trabajo. A continuación se clasificaron y se analizaron las características de cada herramienta.

La metodología fue de corte cualitativo, ya que se hizo un análisis de categorías de los recursos localizados en cuanto a su tipo: biblioteca de JavaScript, software de acceso libre, plataforma, software comercial, y entorno y lenguaje de programación. Se hizo una breve descripción de cada elemento.

### 3. Resultados y análisis

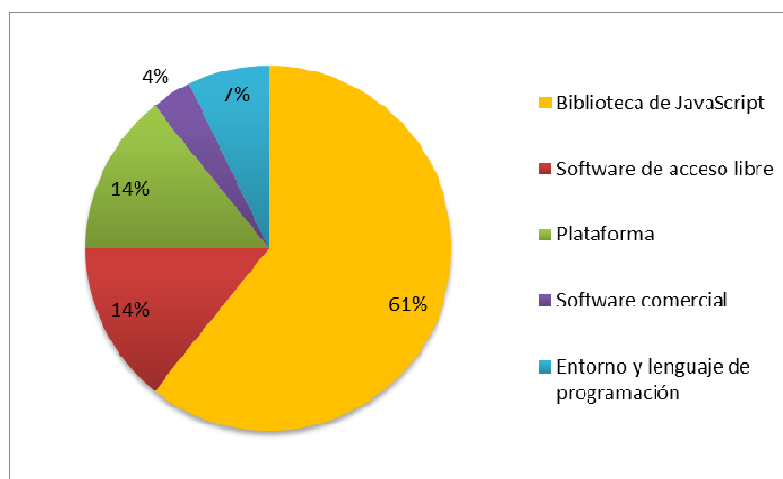
La tabla 1 resume las características del software localizado. Las aplicaciones se presentan en orden alfabético.

**Tabla 1. Comparativo de software de acceso libre para hacer visualización estadística.**

<b>Software</b>	<b>Sitio web</b>	<b>Descripción</b>
Arbor.js	<a href="http://arborjs.org/">http://arborjs.org/</a>	Biblioteca de JavaScript para árboles interactivos.
CartoDB	<a href="http://cartodb.com/">http://cartodb.com/</a>	Biblioteca de JavaScript con mapas interactivos.
Circos	<a href="http://circos.ca/">http://circos.ca/</a>	Software para visualizaciones circulares.
Cubism	<a href="http://square.github.com/cubism/">http://square.github.com/cubism/</a>	Biblioteca de JavaScript para visualizar series de tiempo.
D3.js	<a href="http://d3js.org/">http://d3js.org/</a>	Biblioteca de JavaScript para gráficos dinámicos.
Envision.js	<a href="http://www.humblesoftware.com/envision">http://www.humblesoftware.com/envision</a>	Biblioteca de JavaScript para gráficos dinámicos.
Excel	-	No es un software libre, pero es accesible y permite hacer visualizaciones sencillas.
Flot	<a href="http://www.flotcharts.org/">http://www.flotcharts.org/</a>	Biblioteca de JavaScript para gráficos dinámicos.
Gephi	<a href="http://gephi.org/">http://gephi.org/</a>	Plataforma de visualización interactiva.
Google Chart API	<a href="https://developers.google.com/chart/">https://developers.google.com/chart/</a>	Permite visualizaciones sencillas dinámicas.
Highcharts JS	<a href="http://www.highcharts.com/">http://www.highcharts.com/</a>	Biblioteca de JavaScript para gráficos dinámicos.
JavaScript InfovisToolkit	<a href="http://philobg.github.com/jit/">http://philobg.github.com/jit/</a>	Biblioteca de JavaScript para gráficos dinámicos.
Kartograph	<a href="http://kartograph.org/">http://kartograph.org/</a>	Biblioteca de JavaScript con mapas interactivos.
Leaflet	<a href="http://leafletjs.com/">http://leafletjs.com/</a>	Biblioteca de JavaScript con mapas interactivos.
ManyEyes	<a href="http://www-958.ibm.com/software/data/cognos/manyeyes/">http://www-958.ibm.com/software/data/cognos/manyeyes/</a>	Sitio para hacer visualizaciones con Java.
Miso Dataset	<a href="http://misoproject.com/dataset/">http://misoproject.com/dataset/</a>	Biblioteca de JavaScript para gráficos dinámicos.
NodeBox	<a href="http://nodebox.net/node/">http://nodebox.net/node/</a>	Software para visualizar grandes conjuntos de datos.
OpenLayers	<a href="http://openlayers.org/">http://openlayers.org/</a>	Biblioteca de JavaScript con mapas interactivos.
PolyMaps	<a href="http://polymaps.org/">http://polymaps.org/</a>	Biblioteca de JavaScript con mapas interactivos.
Processing	<a href="http://processing.org/">http://processing.org/</a>	Lenguaje de programación.
Protovis	<a href="http://mbostock.github.com/protovis/">http://mbostock.github.com/protovis/</a>	Biblioteca de JavaScript para gráficos dinámicos.
R	<a href="http://www.r-project.org/">http://www.r-project.org/</a>	Entorno de programación y paquete estadístico.
Raphaël	<a href="http://raphaeljs.com/">http://raphaeljs.com/</a>	Biblioteca de JavaScript para gráficos dinámicos.

Software	Sitio web	Descripción
Rickshaw	<a href="http://code.shutterstock.com/rickshaw/">http://code.shutterstock.com/rickshaw/</a>	Biblioteca de JavaScript para visualizar series de tiempo.
TableauPublic	<a href="http://www.tableausoftware.com/public/">http://www.tableausoftware.com/public/</a>	Software para visualizar datos.
Tangle	<a href="http://worrydream.com/Tangle/">http://worrydream.com/Tangle/</a>	Biblioteca de JavaScript para documentos reactivos.
Visual.ly	<a href="http://visual.ly/">http://visual.ly/</a>	Más enfocado a infografías que a visualización estadística. Se puede conectar con Facebook y Twitter.
Weka	<a href="http://www.cs.waikato.ac.nz/ml/weka/">http://www.cs.waikato.ac.nz/ml/weka/</a>	Software para minería de datos en java.

A partir de esta tabla se obtuvo la clasificación que se muestra en la figura 5, donde destaca que la mayoría son bibliotecas de JavaScript.



**Figura 5. Clasificación de las herramientas para visualización estadística.**  
Fuente: Elaboración propia.

#### 4. Conclusiones

Como puede verse, existe un buen número de herramientas para hacer visualización estadística hoy en día. En este trabajo hemos registrado 28, de las cuales la mayoría son bibliotecas de JavaScript enfocadas a elaborar gráficos dinámicos con datos propios. Algunas son sitios donde pueden colocarse los datos, como ManyEyes, Highcharts JS y Google Chart API.

Dentro de los entornos y lenguajes de programación formales, destacan R, Processing, Weka y JavaScript. En particular, Processing permite también programar en JavaScript con la herramienta Processing.js. Tanto R como Processing son funcionales en plataformas Windows, Mac y Linux. R y Weka han sido desarrollados por universidades de Nueva Zelanda, mientras que Processing fue originalmente una iniciativa del Massachusetts Institute for Technology.

Consideramos que es útil contar con la información anterior de manera concentrada, para:(a) integrar las diversas aplicaciones a los planes de estudio de las licenciaturas relacionadas con la estadística y (b) darlas a conocer a los investigadores que pueden hacer uso de ellos en sus trabajos, tanto para el análisis de la literatura como para organizar los datos recabados.

Coincidimos con Wang y col.[11] en cuanto a que la interactividad de las computadoras y el software actuales son un elemento clave que potencia el aprendizaje de la estadística porque disminuye la carga cognitiva, capta la atención y mejora la comprensión de los conceptos teóricos. Sin embargo, consideramos que la visualización estadística es aún más poderosa, por el hecho de que permite descubrir patrones y relaciones en grandes conjuntos de datos, lo cual hace patente el valor de los conceptos teóricos para aplicaciones muy interesantes. Además, permite que los estudiantes jueguen con datos propios, lo cual agrega motivación.

En cuanto a los investigadores, la visualización permite crear verdaderas narraciones a partir de los datos, de manera que aún la investigación de corte cualitativo puede descansar en ella, puesto que facilita el examen de textos extensos, audios y videos. La visualización, además, permite ciertas estructuras narrativas [8] que pueden ser de gran interés.

Por último, cabe mencionar el uso de la visualización estadística para las revisiones de literatura, con un par de ejemplos notables: Eigenfactor[12] y Microsoft AcademicSearch[13].

Resultará esencial continuar investigando acerca de los recursos disponibles para hacer visualización estadística y las mejores formas de usarlos. En particular, consideramos que deberá discutirse su inclusión en los planes de estudio que abarcan la enseñanza de la estadística y en la formación para la investigación.

## **Agradecimientos**

Los autores agradecemos el apoyo recibido al proyecto PAPIME PE 300713 titulado “Vientos de cambio: Estrategias y buenas prácticas para el uso de ambientes virtuales en la educación superior”, de DGAPA UNAM FES Acatlán.

## **Referencias**

- [1] YAU, N. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*: Wiley 2011.
- [2] CHEN, C. H., HÄRDLE, W. & UNWIN, A. *Handbook of Data Visualization*: Springer-Verlag 2008.
- [3] ROSLING, H. *Gapminder*. 2013. [Fecha de consulta: 2013/01/13.] Disponible en: <http://www.gapminder.org/>
- [4] MAYER, M. *A world in motion*. Google Official Blog 2007. [Fecha de consulta: 2013/01/03.] Disponible en: <http://googleblog.blogspot.mx/2007/03/world-in-motion.html>



- [5] WOLPHRAM ALPHA. *Wolfram Alpha: Computational Knowledge Engine*. 2013. [Fecha de consulta: 2013/01/03.] Disponible en: <http://www.wolframalpha.com/>
- [6] RAO, C. R. *Handbook of Statistics: Data Mining and Data Visualization*: Elsevier Science 2005.
- [7] BYRON, L. & WATTENBERG, M. *Stacked graphs—geometry & aesthetics*. Visualization and Computer Graphics, IEEE Transactions on. 2008; 14(6):1245-1252.
- [8] SEGEL, E. & HEER, J. *Narrative visualization: Telling stories with data*. Visualization and Computer Graphics, IEEE Transactions on. 2010; 16(6):1139-1148.
- [9] SUDA, B. *The top 20 data visualisation tools*. 2012. [Fecha de consulta: 2013/01/03.] Disponible en: <http://www.netmagazine.com/features/top-20-data-visualisation-tools>
- [10] DATAVISUALIZATION.CH. *Selected Tools*. 2013. [Fecha de consulta: 2013/01/03.] Disponible en: <http://selection.datavisualization.ch/>
- [11] WANG, P.-Y., VAUGHN, B. K. & LIU, M. *The impact of animation interactivity on novices' learning of introductory statistics*. Computers & Education. 2011; 56(1):300-311.
- [12] UNIVERSITY OF WASHINGTON. *Eigenfactor*. 2012. [Fecha de consulta: 2013/06/03.] Disponible en: <http://www.eigenfactor.org/>
- [13] MICROSOFT RESEARCH. *Academic Search*. 2013. [Fecha de consulta: 2013/06/03.] Disponible en: <http://academic.research.microsoft.com/>