

Visualización de Datos Electorales Presidenciales: Herramientas para la Visualización de Grandes Conjuntos de Datos

Roberto CAMANA FIALLOS

Facultad de Ingeniería en Sistemas
Universidad Tecnológica Indoamérica
Bolívar 20-35 y Quito, Ambato, Ecuador
robertocamana@yahoo.es
robertocamana@uti.edu.ec

RESUMEN

El objetivo de esta investigación fue aplicar herramientas para la visualización espacial de datos en forma dinámica, es decir animada e interactiva. El proceso se aplicó sobre un conjunto de datos de las elecciones presidenciales de Ecuador, primera vuelta, años 2006, 2009 y 2013. Se obtuvieron visualizaciones que permitieron realizar tres tipos de análisis. En primer lugar, la votación histórica por partidos políticos (2006–2013), utilizando gráficos de burbujas; este análisis tenía la finalidad de conocer la evolución de la votación alcanzada por partidos políticos recientemente creados, frente al oficialismo. La segunda, resultados por cantones, divididos por género (hombre y mujer), concerniente a las elecciones presidenciales 2013, utilizando mapas geoespaciales, con el fin de determinar cuánto influye la decisión de género en las elecciones. La última fue visualizar la cantidad total de votos nulos por cantones, con el propósito de establecer estrategias para campañas políticas, utilizando diagrama de árboles. Para tales efectos se utilizaron la aplicación gratuita en línea Many Eyes y el software comercial TreeMap. Para la visualización de datos geoespaciales, se manejó el software libre Quantum GIS. Estas estrategias permitieron lograr visualizaciones comprensibles y dinámicas para la toma de decisiones para candidatos, consultores, encuestadores y partidos políticos, en cara a futuras elecciones.

PALABRAS CLAVE

Datos electorales, diagrama de árboles, partidos políticos, resultados, visualización de datos.

ABSTRACT

The objective of this research was to apply tools for spatial visualization of data, in a way that is dynamic, lively, and interactive. The process was applied on a dataset drawn from presidential elections in Ecuador, first round, 2006, 2009, and 2013. Different visualizations allowed performing three kinds of analyses. First, representing the historic vote by political parties (2006–2013), using bubble graphs; this analysis seeks to understand the evolution of results obtained by newly created political parties against the ruling party. The second visualization allowed representing results by cantones, divided by gender (female/male), concerning the 2013 presidential elections; the aim was to determine how gender influences voting, which was shown in geospatial maps. The last analysis allowed visualizing the total number of invalid votes by cantones, in order to establish strategies for future political campaigns; these results were displayed in tree diagrams. Visualization was possible thanks to the use of the free online application Many Eyes, the commercial software TreeMap, and the free software Quantum GIS for geospatial data visualization. These analyses allowed achieving dynamic visualizations, very useful for the decision-making process of candidates, consultants, pollsters, and political parties, facing future elections.

KEYWORDS

Electoral data, tree diagram, results, political parties, data visualization.

Introducción

Con el avance de las Tecnologías de la Información y la Comunicación (TICS), la computadora comenzó a jugar un papel importante en la vida cotidiana. La evolución imprescindible de esta herramienta electrónica ha logrado también transformar la forma de presentar la información de los grandes volúmenes de datos.

Con la era informática, en los años 90, surgieron nuevos procesos netamente innovadores, como son las visualizaciones de datos dinámicos. Es entonces cuando se realizan importantes integraciones: interfaces gráficas de usuario, bases de datos, navegación y simulación. Esto dio nacimiento a una nueva tendencia de visualización de conjuntos de datos sobre potentes sistemas computacionales, capaces de crear visualizaciones dinámicas, es decir animadas e interactivas. Actualmente, la sociedad de la información está inundada por grandes volúmenes de datos de difícil consumo e interpretación [1]. Por lo tanto, la visualización de la información es la mejor opción para mostrar los datos de forma ordenada, creativa y visual.

Uno de los campos poco explotados es la visualización dinámica de resultados provenientes de elecciones, cuya representación gráfica se limita a barras y diagramas de torta provistos en programas informáticos. Sin embargo, ya existen estrategias para presentar cientos de datos, de forma sencilla y fácil de interpretar. Estas nuevas gráficas son animadas e interactivas, permitiendo introducir datos en tiempo real y en el proceso interno computacional. Este proceso es de gran potencia, debido a la variedad de métodos y análisis matemáticos. Además, permite obtener visualizaciones rápidas, novedosas y, sobre todo, que faciliten la interpretación de la representación por parte de los usuarios [2]. Para dicho efecto juega un papel importante el conocimiento y uso de herramientas adecuadas para realizar diversas visualizaciones de datos.

En Ecuador, el tema de la visualización de datos electorales, está en fase de evolución y son pocos los estudios efectuados al respecto. Sin embargo, ya existe un primer acercamiento y tratamiento y análisis de la información electoral en el país [3].

El objetivo de esta investigación fue utilizar herramientas de visualización dinámica que permitan analizar la votación histórica por partidos políticos en el periodo 2006–2013. Para ello se aplicaron herramientas de visualización dinámica de datos, utilizando el *software* libre Many Eyes para la aplicación *on-line*, el *software* libre Quantum GIS para la visualización de datos geoespaciales y el *software* comercial TreeMap.

Metodología

Se siguió la metodología de Fry, para la construcción de visualizaciones interactivas [4]. Los pasos de esta metodología se describen a continuación:

Adquisición y conversión de datos

La totalidad de los datos fue obtenida desde el portal web del Consejo Nacional Electoral (CNE)[5]. El esfuerzo se concentró en la obtención de resultados de las elecciones para Presidente y Vicepresidente, de las elecciones desarrolladas en febrero de 2013. Los datos incluyeron la votación a nivel nacional y de los residentes en el exterior. En el contexto nacional, se obtuvieron resultados desgregados de las 24 provincias, 221 cantones y 1.228 parroquias urbanas y rurales. En el entorno exterior, correspondiente a emigrantes ecuatorianos, se obtuvieron resultados desgregados por países y agrupados en zonas: América Latina el Caribe y África (20 países); EEUU y Canadá; y Europa, Asia y Oceanía (15 países).

Así mismo, se utilizaron resultados de las elecciones presidenciales de las primeras vueltas, realizadas en los años 2006 y 2009. También se analizaron datos históricos generales (por provincia) de las elecciones presidenciales desde 1980 hasta 2004, proporcionados por el CNE. La finalidad de incluir los datos de posteriores elecciones presidenciales era poder analizar, por ejemplo, cuanto aumentó o disminuyó la votación por un partido con respecto a los datos históricos de la última elección presidencial.

La información se encuentra en formato HyperText Markup Language (HTML); en texto puro que muestra los valores numéricos. La primera parte consta de información estadística, desgregada por provincia, por cantón y por parroquia (por ejemplo Tungurahua > Ambato > Ambatillo). La segunda

parte se refiere al listado de organizaciones políticas con sus candidatos, su votación por hombres y mujeres, con sus respectivos porcentajes.

Con las dos secciones en conjunto, la información estadística y el listado de partidos políticos, se procedió a realizar una estructura de datos, que se describirá a continuación. Se seleccionaron 20.200 Juntas Receptoras del Voto de hombres y 20.251 Juntas Receptoras del Voto de mujeres, mostrados (Tabla 1). Las variables en la Tabla 1, se sumaron a las variables creadas en la siguiente sección.

Análisis y procesamiento de la información

En este paso, los resultados electorales fueron convertidos en una sola estructura (en Excel), con el objetivo de facilitar la manipulación y comprensión. Se usó la opción "Obtener Datos Externos de Texto", para facilitar el trabajo, ya que se necesitaba la extracción de datos desde la página del CNE. Así, se obtuvo un total de 30 campos, distribuidos en 3 grupos: Provincia, Cantón y Parroquia; información estadística por género; y votación por partido político por género.

Tabla 1. Listado de atributos obtenidos del portal web. Fuente: CNE y elaboración propia.

N°	Nombre	Descripción
1	Electores	Número total de electores y juntas, según la provincia, el cantón y la parroquia.
2	Juntas	
3	Total de firmas y huellas dactilares que constan en el padrón electoral	Número de electores que sufragaron.
4	Blancos	Son consideradas papeletas que no tienen ninguna marca o raya a favor de determinado candidato.
5	Nulos	Son aquellas papeletas que tienen más de un candidato con alguna marca o raya.
6	PSP	Lucio Gutiérrez

7	PRIAN	Alvaro Noboa
8	PRE	Nelson Zavala
9	PACHAKUTIK/MPD	Alberto Acosta
10	CREO	Guillermo Lasso
11	SUMA	Mauricio Rodas
12	RUPTURA	Norman Wray
13	PAIS	Rafael Correa Delgado
14	Votos totales	Suma de votos por candidatos con sus respectivos porcentajes.
15	Hombres	Votación disgregada por género con sus respectivos porcentajes.
16	Mujeres	

Adicionalmente a los datos mencionados, se crearon nuevas variables, incluyendo ausentismo, votos válidos (por género) y total de votos. En total, el número registros (filas) fue 1.227 y total de columnas fue 45. En vista de los diferentes usos que se le dio al conjunto de datos, se utilizaron dos formatos: el tipo de formato abierto (CSV) y el formato estándar de Excel (XLS), ambos compatibles con programas de visualización de datos.

En esta etapa también se realizó un filtrado que consiste en quitar los datos sin interés. Los datos obtenidos del CNE, se disgregaron de la siguiente forma: Provincia => Cantón => Parroquia => Zonas electorales. Sin embargo, esta última fue descartada por cuanto es un espacio geográfico delimitado que pertenece a una parroquia urbana o rural, en las que se establecen uno o más recintos electorales [5]. Con este antecedente, se decidió, trabajar en tres niveles (provincial, cantonal y parroquial).

Representación de los datos

Con el fin de presentar la información y permitir la interacción del usuario, se desarrolló una aplicación web, dado que de esta manera el usuario que quisiera emplearla sólo requeriría tener una conexión a internet y un navegador web para poder ingresar. El desafío fue construir aplicaciones de web gratuita que se asemejaran lo más posible a una aplicación de escritorio, tanto en su aspecto visual, como en su posibilidad de interacción.

Las herramientas que se utilizaron para el desarrollo de la aplicación fueron, las siguientes:

Many Eyes. Es una herramienta de visualización de la compañía IBM, completamente gratuita y de acceso libre. Está disponible en el sitio web (<http://www-958.ibm.com>). Su principal característica es la posibilidad que tiene el usuario para subir grandes conjuntos de datos sin importar el formato que estos tengan. Además, se pueden seleccionar algunas representaciones gráficas para analizar texto, para comparar conjuntos de datos, para ver relación entre puntos de datos y para ver los datos parcialmente, entre otras visualizaciones.

Una de las ventajas es que las aplicaciones desarrolladas en la web, es que pueden ser compartidas y visualizadas por los demás usuarios [6]. La finalidad de usar esta herramienta, es mostrar la potencialidad que tiene para manejar grandes volúmenes de datos y presentarlos en gráficos dinámicos e interesantes.

Quantum GIS (QGIS). Es un Sistema de Información Geográfica (SIG), libre y de descarga gratuita (<http://www.qgis.org/es/site/>). Trabaja sin problemas en varias plataformas Windows, Linux, Mac. La fácil conexión a la base de datos ha hecho que QGIS utilice toda su potencia de análisis, con un entorno de trabajo agradable [7]. Este software libre, se utilizó para poder representar la totalidad de datos geoespaciales de las elecciones presidenciales del Ecuador, cuyos datos se presentan por capas (provincial, cantonal y parroquial).

TreeMap. Es un *software* de tipo comercial (<http://www.treemap.com>), para el cual se debe pagar una licencia de uso indefinido; sin embargo, se puede usar gratuitamente por 30 días pudiendo ser renovada la licencia mensualmente. La principal característica de TreeMap es la forma elegante de presentar sus visualizaciones, mediante diagramas de árboles, listos para analizar [8].

La integración de los datos alojados en una base de datos es el requerimiento de todo usuario. La base de datos es capaz de realizar conexiones con MySQL y SQL Server, así como la importación de conjunto de datos en formato de Excel. El objeto de usar este *software* es conectar una base de datos en MySQL, con la aplicación TreeMap.

Enriquecimiento de los datos

Los colores que se emplean en los gráficos producidos por estas herramientas no mantienen una correspondencia con los colores que identifican a cada uno de los partidos. Sumado a esto, los nombres de los partidos en la mayoría de los casos tienen una longitud que imposibilita que sean presentados de manera prolija y legible en la visualización.

Para mejorar estos aspectos se decidió volver a la primera etapa de adquisición de datos, en este caso con el objetivo de enriquecer la información existente. Se buscó, en caso de ser conocido, cual es el color característico de cada partido y se generó un nombre abreviado para ser utilizado en el gráfico.

Sin embargo, en Many Eyes no existe la posibilidad de personalizar los colores correspondientes a cada partido político; en este software la asignación de colores se realiza de forma aleatoria de acuerdo al número de partidos. Los colores distintivos de los partidos políticos en la Tabla 2, se utilizaron para distinguir a las diferentes fuerzas políticas por cantón, por medio de QGIS.

Tabla 2. Colores distintivos que utilizan los partidos políticos

COLOR	NOMBRE ORGANIZACIÓN POLÍTICA	SIGLAS
	Partido Sociedad Patriótica "21 de Enero". Lista 3	PSP
	Partido Renovador Institucional Acción Nacional. Lista 7	PRIAN
	Partido Roldosista Ecuatoriano. Lista 10	PRE
	Unidad Prurinacional de las Izquierdas. Listas 18, 15, 29	PK-MPD-RED
	Creando Oportunidades. Lista 21	CREO
	Movimiento Sociedad Unidad MAS. Lista 23	SUMA
	Ruptura. Lista 25	R25
	Movimiento Patria Altiva i Soberana. Lista 35	MIPAIS

Interacción con la Representación de los Datos

Como se mencionó previamente, el objetivo principal de esta investigación es mostrar gráficos dinámicos interactivos de calidad. Las vistas iniciales son similares a las provistas por el sitio del CNE (<http://www.resultados/cne.gob.ec>), con la diferencia que en este sitio web solo se permite seleccionar dos tipos simples de gráficos: de barra y circular.

Resultados y Discusión

La aplicación de herramientas para la visualización de conjuntos de datos electorales es un reto de la computación gráfica, sobre todo si se trata de mostrar millones de datos en una forma que sea atractiva y dinámica. Esta tarea se hizo menos complicada con la ayuda de herramientas apropiadas. A continuación, se presentan los resultados de la visualización en las tres herramientas aplicadas en entorno web.

Visualización histórica de la votación total por partidos políticos (2006–2013), utilizando Many Eyes

Para elaborar la visualización (Fig. 1), se tomaron datos de las 24 provincias del Ecuador, utilizando dos variables, partido/organización política y votación (2006–2013). Esta última variable fue la que dividió cada círculo en pequeñas porciones, es decir, tomando en cuenta el porcentaje que cada organización política obtuvo en las elecciones presidenciales de los años 2006, 2009 y 2013.

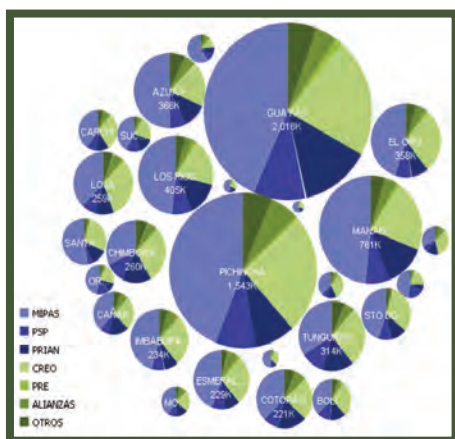


Figura 1. Visualización promedio elecciones presidenciales (2006–2013) por partido y por provincia.

El objetivo de realizar esta visualización fue mostrar en un solo vistazo el porcentaje histórico alcanzado por partidos políticos en las tres elecciones del periodo 2006–2013. A primera vista, se nota la supremacía de MIPAS con el 43,59%. Sin embargo, no se disgregan resultados por cada elección individual. Por ejemplo, en el 2013 el debutante partido CREO alcanzó 24,28% y se proyecta con mayores aspiraciones para las elecciones futuras. En estas mismas elecciones, decreció la votación de partidos tradicionales como PRIAN con 11,72% y PSP 9,17%; también decrece el PRE 0,90% (aunque en menor grado), superado por ALIANZAS entre partidos 5,03% y OTROS PARTIDOS 5,32%.

También se tomó una muestra de las provincias que obtuvieron más de 100.000 votos, con la finalidad de que se conozca dónde se obtuvieron los resultados (Tabla 3).

Tabla 3. Datos de ausentismo por provincia, cantón y parroquia.

PROVINCIA/PAÍSES	PARTIDO/ORGANIZACIÓN POLÍTICA	PROMEDIO DE VOTOS (2006-2013)
GUAYAS	MIPAS	876,69
PICHINCHA	MIPAS	677,14
GUAYAS	CREO	447,25
PICHINCHA	CREO	387,66
MANABI	MIPAS	367,51
GUAYAS	PRIAN	269,52
GUAYAS	PSP	202,49
LOS RIOS	MIPAS	195,70
AZUAY	MIPAS	182,94
MANABI	CREO	165,59
EL ORO	MIPAS	161,25
PICHINCHA	PSP	138,12
PICHINCHA	PRIAN	135,78
GUAYAS	OTROS	110,66
IMBABURA	MIPAS	106,95
TUNGURAHUA	MIPAS	106,31
EL ORO	CREO	103,31
ESMERALDAS	MIPAS	101,52

Lo importante de este tipo de visualizaciones no es solamente sus gráficos, sino la interesante y novedosa información que se puede obtener. Esta tecnología será un instrumento para candidatos, consultantes, encuestadores, partidos políticos, agencias de gobierno y profesionales en relaciones públicas, de cara a las próximas elecciones.

Resultados por cantón y por género en las elecciones presidenciales 2013, utilizando QSIG

Se seleccionaron 221 cantones de las 24 provincias del país, divididos en dos tablas (una de hombres y otra de mujeres) con sus respectivos partidos políticos (PSP, PRIAN, PRE, PK-MPD-RED, CREO, SUMA, R25, MIP AIS).

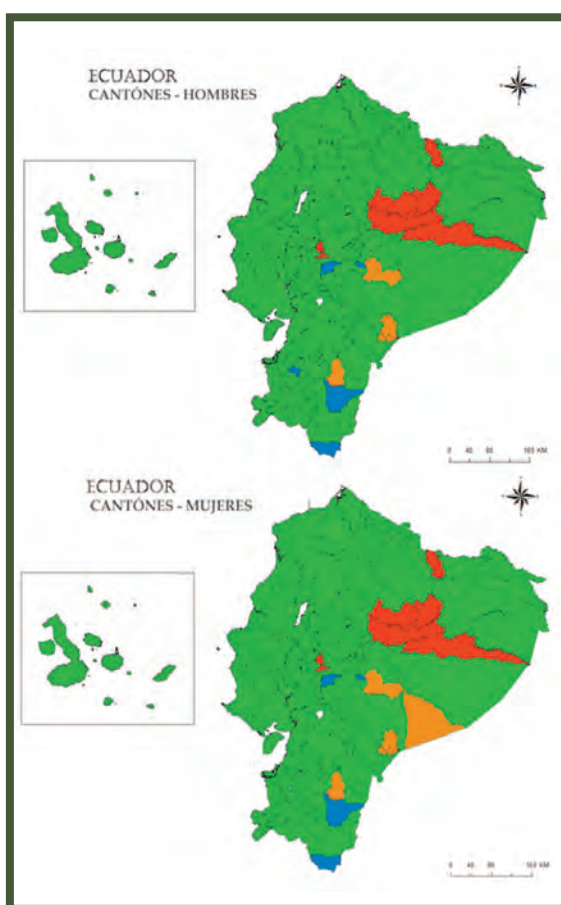


Figura 2. Votación por partido disgregada por cantón. Arriba hombres y abajo mujeres.

En las dos vistas obtenidas (Fig. 9), se asignó un color distintivo a cada partido político, sea (verde, azul, rojo, etc.). Cada color es visualizado en el cantón, de acuerdo a la mayor votación obtenida. Por ejemplo, en el cantón Ambato, provincia de Tungurahua, MIP AIS (color verde), alcanzó la mayor votación entre hombres (44.372 votos) y mujeres (48.434 votos). También se observa que en un mismo cantón dos partidos políticos alcanzan la mayor votación, tanto entre hombres como entre mujeres. Por ejemplo en el cantón Atahualpa, provincia de El Oro, CREO (color azul claro) tiene mayor votación entre hombres, con 1.056 votos, frente a mujeres, con 1.035 votos.

Así mismo, se observó que en los cantones de color rojo (PSP), gana un partido político tradicional que viene participando desde las elecciones de 2006, manteniendo su supremacía electoral en las provincias de la región oriente, norte (Sucumbíos, Orellana, y Napo).

Tabla 4. Partidos políticos en cantones

Cantones	Números Cantones	Diferencia
MIP AIS (Hombres)	193	Ninguna
MIP AIS (Mujeres)	193	
PSP (Hombres)	5	
PSP (Mujeres)	5	
CREO (Hombres)	4	Atahualpa (El Oro)
CREO (Mujeres)	5	
PK-MPD-RED (Hombres)	4	Twintza (Morona Santiago)
PK-MPD-RED (Mujeres)	5	

Como se puede observar en la Figura 9 y la Tabla 4, MIP AIS tiene supremacía en los dos géneros, entre los cantones del país.

Visualización de votos nulos por cantones, utilizando TreeMap

Para elaborar esta visualización se trabajó con datos de los 221 cantones a nivel nacional. Se seleccionaron las variables provincia, cantón, total_nulos, nulos_hombres y nulos_mujer. Como se esperaba, las provincias de Manabí, Pichincha, Guayas y Los Ríos, al tener mayor número de votantes, son las mismas que tienen mayor número votación nula.



Figura 3. Disgregación del voto del color cantones por provincias.

Lo novedoso de esta visualización (Fig. 3) es que cada cuadrado representa un cantón y el conjunto de cuadrado forma una provincia. El degradado del color tiene un valor máximo y un valor mínimo. La variable total_nulos (variable categórica) fue utilizada por TreeMap para crear rangos automáticos, con la desventaja que concentraba en rangos superiores la mayoría de votos_nulos. Para ser más equitativos, se probó con una nueva variable rango, en la cual se crearon categorías: 1, 2, 3,... n. Esto permitió una mejor disgregación del color, diferenciando desde el mayor (color gris con blanco) al menor (color azul marino) (Fig. 3).

El objetivo de esta visualización fue utilizar un diagrama de árboles (TreeMap) que pretende seleccionar los cantones con el mayor número de votos nulos. Esta votación que representa indecisión, se convierte en un nicho para ganar los votos de los electores y convertirlos en votos válidos para un candidato de interés. Además de las capitales provinciales como Quito, Guayaquil, Portoviejo, entre otras capitales, existen también cantones representativos por su número de votantes, que demuestran mayor número de votación nula (Tabla 5).

Tabla 5. Cantones con mayor votación nula

PROVINCIA	CANTÓN	% NIVEL NACIONAL	VOTOS NULOS
PICHINCHA	SAN MIGUEL DE LOS BANCOS	1.59%	12683
MANABÍ	PEDERNALES	1.44%	11484
LOS RÍOS	VINCES	1.17%	9342
LOJA	ZAPOTILLO	0.92%	7353
AZUAY	NABON	0.88%	7071
IMBABURA	URCUQUI	0.61%	4905
ESMERALDAS	QUININDE	0.53%	4258
GUAYAS	YAGUACHI	0.50%	3987
COTOPAXI	PUJILI	0.46%	3711
CHIMBORAZO	COLTA	0.31%	2500
STO DOMINGO TSACHILLAS	SANTO DOMINGO	0.29%	2339
TUNGURAHUA	PELILEO	0.27%	2171
MORONA SANTIAGO	GUALAQUIZA	0.23%	1878
TUNGURAHUA	PILLARO	0.22%	1757

Conclusiones

Sin duda, estas nuevas formas de representar la información, son más versátiles que la tradicional forma de presentar datos (barras, círculos, etc.) y que solo permiten trabajar con pequeñas cantidades de datos. El escrutinio dinámico de estos datos permite que sean utilizados para la toma de decisiones de partidos políticos, candidatos,

consultoras, encuestadores, partidos políticos y profesionales en general, de cara a futuras elecciones.

Agradecimientos

Se agradece el apoyo de la Universidad Tecnológica Indoamérica, Facultad de Ingeniería en Sistemas, así como el aporte valioso de investigadores (análisis y diseño), estudiantes (limpieza de datos e integración a la base de datos) e instituciones del estado como el Consejo Nacional Electoral (CNE) y el Instituto Ecuatoriano de Estadísticas y Censos (INEC).

Referencias

[1] Ajuntament de Barcelona. Cibernarium. [en línea]. [Fecha de consulta: noviembre de 2013]. Disponible en: http://w144.bcn.cat/cibernarium/es/activitats/editActivitat/herramientas-para-la-visualizacion-de-datos-convierte-informacion-en-conocimiento.do?codildioma=2&id=651471&id_activitat_mestre=627528&dia=14&mes=11&any=2013

[2] Freeman, L. C. 2000. Visualizing Social Networks, *Journal of Social Structure* 1. [en línea]. [Fecha de consulta: abril de 2013] <http://www.cmu.edu/joss/content/articles/volume1/Freeman.html>

[3] Camana, R. 2012. Aplicación de Técnicas de Minería de Datos para la Indagación y Estudio de Resultados Electorales. *CienciAmerica* 1: 85–94.

[4] Fry, B. 2008. *Visualizing Data: Exploring and Explaining Data with the Processing Environment*. O'Reilly Media. 1 edition.

[5] Consejo Nacional Electoral. [en línea]. [Fecha de consulta: noviembre de 2013] <http://www.cne.gob.ec/index.php/boletines-de-prensa/articulos/recorridos-por-nueva-zona-electoral-pata-de-mula.html>

[6] Tutorial básico para Many Eyes. [en línea]. [Fecha de consulta: abril de 2013] <http://es.slideshare.net/contrapunto/tutorial-bsico-para-manyeyes>

[7] Visualización de datos, Big Data: Graficar datos en Javascript [Fecha de consulta: abril del 2013]. <http://www.visualthinking.es/tag/visualizacion-de-datos/>

[8] TreeMap 4.0 Documentación en línea. [en línea] [Fecha de consulta: abril de 2013]. <http://www.cs.umd.edu/hcil/treemap/doc4.0/toc.html>