

Análisis Visual de encuesta de Satisfacción Estudiantil de la Universidad de los Andes

Stevenson Contreras, Felipe González, Juan Camilo Ortiz, Jessica Jiménez, Jose Tiberio Hernández & John Alexis Guerra

Departamento de Ingeniería de Sistemas y Computación

Universidad de los Andes
Bogotá, Colombia

{js.contreras2187, jc.ortiz939, jf.gonzalez695, jes-jime, ja.guerrag, jhernand}@uniandes.edu.co

Resumen—Este proyecto busca presentar una herramienta de análisis visual sobre datos de una encuesta de satisfacción de una población estudiantil, para realizar análisis comparativos, presentar un resumen del resultado de la encuesta y encontrar correlaciones existentes entre los niveles de satisfacción y las características de los públicos encuestados. Este análisis construye sobre los principios de analítica visual e interacción en visualización, teniendo en cuenta la necesidad de la navegación jerárquica para el contexto del problema. El proyecto considera especialmente la diversidad de perfiles de usuario que usarán la herramienta y, como tal, apela al uso de gestos intuitivos y de visualizaciones de fácil comprensión y navegación. Esto se asegura a través de validaciones con expertos en visualización, diseño y un experto del dominio, así como con el uso de usuarios sobre iteraciones de prototipos construidos a lo largo del proyecto.

Palabras Clave— *Analítica Visual; interacción web; navegación jerárquica.*

I. INTRODUCCIÓN

La Encuesta de Satisfacción Estudiantil ESAT es una encuesta realizada anualmente por el Departamento de Planeación a los estudiantes de la Universidad de los Andes, con el fin de medir la percepción de satisfacción en la población estudiantil. La encuesta está compuesta por 43 preguntas con respuestas en escala Likert de seis niveles de satisfacción posibles. La encuesta es realizada en promedio a 15000 estudiantes anualmente, tiene un

porcentaje de respuesta del 30% aproximadamente, por lo que se cuenta con cerca de 5000 registros por año de realización de la encuesta.

Actualmente el Departamento de Planeación cuenta con los datos de respuestas de las encuestas realizadas desde el año 2011 hasta el 2015. Con estos datos se generan reportes estadísticos con herramientas de computación tradicionales, cuya creación representa un proceso laborioso. El Departamento pretende realizar análisis comparativos del nivel de satisfacción entre diferentes perfiles poblacionales, acción que no le es posible realizar con las herramientas estadísticas con las que generan su reporte.

Este artículo presenta el proceso de diseño, desarrollo y evaluación de una herramienta de análisis visual para el estudio de los datos de la encuesta ESAT. La herramienta es construida para tecnologías web (HTML, Javascript y CSS) utilizando el framework de visualización de datos D3, con una base de datos MySQL sobre la que se realizan consultas de la información de las encuestas.

Este proyecto busca extraer valor de los datos, y potencializarlo a través de técnicas de interacción. Luego de un estudio del estado del arte y la caracterización de los datos y del contexto de diseño, se presenta una propuesta de solución en donde la interacción entre componentes gráficos es de especial importancia para el usuario de la herramienta.

II. ESTADO DEL ARTE

Se realiza una investigación de bibliografía seleccionada tanto de soluciones a situaciones

relacionadas con el dominio del problema, como de proyectos cuyos objetivos fueran similares a lo que el Departamento de Planeación necesita.

A. Sobre el dominio del problema

Survey Data Visualization Application [1] es una aplicación Web desarrollada por J.Zhang utilizando D3 y jQuery. Esta aplicación le permite al usuario seleccionar la escala Likert que se ajusta a sus datos y distribuye las respuestas de un archivo de valores separados por comas en dos gráficos complementarios: un diagrama de barras y un diagrama de ejes paralelos. Los datos entre los dos diagramas están coordinados y relacionados, por lo que una acción sobre uno de los gráficos afecta ambas visualizaciones, ofreciendo una mayor retroalimentación al usuario. La herramienta también permite el reordenamiento de las variables y el análisis semántico de respuestas de texto.

Evergreen Data [2] presenta una aplicación de información de encuestas de satisfacción en gráficos de barras apiladas. Esta aproximación permite establecer un nivel de proporción y categorización eficaz de las diferentes respuestas obtenidas en la encuesta. Incluye una categorización por tono que acentúa la diferencia categórica, pero se mantiene en un nivel de detalle mínimo, por la misma naturaleza de los datos que visualiza esta herramienta.

La herramienta *SurveyVisualizer*, desarrollada por Macrofocus GmbH [3] es una herramienta construida para analizar grandes y cuantiosos resultados de encuestas de formato complejo. Permite la agregación de variables en diferentes niveles, selección y zoom sobre segmentos de datos y descripción de líneas de tendencia. A diferencia de los dos referentes anteriores, *SurveyVisualizer* contempla una noción temporal de los datos, es decir, utiliza tipos de datos de mediciones hechas en algún momento del tiempo.

B. Fuera del Dominio del Problema

También se analizan referentes de aplicaciones fuera del dominio del problema pero que comparten objetivos o datos categorizados de forma similar a los presentados en este proyecto.

JobMarketTracker, desarrollado por A.Van Dam y R.Lightner [4], es una aplicación web que relaciona la cantidad de empleos ganados o perdidos por sector económico y el cada uno de los meses del año desde el 2009 a la actualidad. A la pérdida o ganancia de cada sector asigna un tono en una escala de color. Con esta información presenta un gráfico de dispersión donde cada punto es un sector económico, su ubicación en el eje ordenado es el momento de la medición y la posición en el eje de abscisas es la cantidad de pérdida o ganancia en puntos porcentuales. Este gran gráfico es complementado con una tabla relacionada que toma el momento puntual de tiempo y expande la información numérica. *JobMarketTracker* también incluye un gráfico de línea con el porcentaje de desempleo e incluye filtros de género, edad, educación y raza. La aplicación presenta un buen manejo de datos individuales en la gráfica de dispersión y como tal permite establecer tendencias en los datos, aún así la escala vertical introduce errores de concordancia con el eje de medición y es posible que en la misma posición en el eje ordenado se encuentren dos mediciones diferentes.

También se consulta *Urban Governance* [5], una iniciativa de la comisión Habitat de las Naciones Unidas que busca construir una base de datos global de los modelos de gobierno de las ciudades para responder a la pregunta “¿cómo se gobiernan las ciudades?”. Esta iniciativa consigna en un sitio web una encuesta de gobernanza urbana realizada a 78 ciudades en seis áreas temáticas. La encuesta se viene realizando desde el 2014. Se analiza este caso al tener un modelo de realización de encuesta similar al del caso de aplicación. En el portal web se presentan los resultados de la encuesta con las respuestas de los ciudadanos agrupadas por ciudades. Es interesante la representación individual de cada ciudad en la construcción de un tipo de diagramas de barras que representan las respuestas a cada una de las preguntas de la encuesta. También es destacable la utilización de las transiciones para la construcción de un *storytelling* sencillo que le da al usuario un entendimiento de los datos que está explorando.

III. CONTEXTO DEL PROBLEMA

A. Definición del contexto de usuario

Si bien el Departamento de Planeación de la Universidad de los Andes realiza el encargo de este proyecto, los usuarios de la herramienta son también entidades y personas de otras dependencias de la institución, entre las que se encuentran dependencias académicas como coordinadores y decanos de facultades. El interés principal de estos usuarios incluye realizar comparaciones con otros programas que pueden ser de otra facultad o de diferente nivel de estudios.

También se contempla la utilización de la herramienta por parte de dependencias no académicas como Gerencia del Campus, vicerrectorías y la Decanatura de Estudiantes, cuya labora se aleja de los aspectos académicos y se enfoca en los procesos y decisiones relacionados con la administración de los recursos y la promoción del bienestar estudiantil. Para estas dependencias es de especial interés la capacidad de evaluar la percepción de satisfacción en el tiempo, de acuerdo a las políticas o acciones implementadas en el campus en general.

En ese orden de ideas, la herramienta construida pretende presentar el resumen del resultado de las encuestas de satisfacción, realizar comparaciones del nivel de satisfacción de los estudiantes discriminando por los atributos que tienen en los datos y encontrar la correlación existente entre los niveles de satisfacción y las características de los públicos encuestados.

B. Caracterización de los datos

La encuesta está estructurada en cuarenta y tres (43) preguntas categorizadas en cuatro (4) ejes temáticos. Las categorías y su explicación se presentan a continuación.

- Satisfacción General: una sola pregunta encaminada a captar la percepción general de los estudiantes con la universidad.
- Imagen Institucional: percepción de las acciones de la universidad en relación a la educación de calidad y su impacto en el desarrollo del país.

- Procesos Académicos: percepción sobre las acciones y procesos que tienen que ver con el ejercicio de la docencia y pedagogía en la institución.
- Servicios Prestados por la Universidad: percepción sobre los servicios complementarios de la universidad. En esta última categoría se distinguen seis subtemas que caracterizan aún más los servicios involucrados. Estos son apoyo al aprendizaje, bienestar, efectividad institucional, personal de apoyo, recursos y vida en la universidad.

La respuesta para cada una de estas preguntas está incluida en una escala Likert de seis opciones: muy satisfecho, satisfecho, ni satisfecho ni insatisfecho, insatisfecho, muy insatisfecho y NS/NC (no sabe, no conoce).

Cada registro de los datos obtenidos de esta encuesta representa la respuesta de un estudiante, de un nivel de estudios, de un género, que pertenece a un programa, departamento y facultad en un año determinado, a una pregunta que hace parte de un tema, y a su vez puede o no hacer parte de un subtema.

Estos registros están agregados de manera tal que cada fila de la tabla de datos agrupa la cantidad de respuestas de la escala likert que fueron obtenidas para una pregunta de acuerdo al perfil del estudiante. El perfil es definido como la combinación de las categorías año de medición, nivel de estudio, facultad, departamento y programa.

Por lo tanto, se puede establecer que existen 10 atributos categóricos de los datos, incluyendo el año de la medición que es categórico ordenado. El número de respuestas en la agregación mencionada es el único atributo de tipo ordenado.

Con los datos obtenidos de un archivo Excel se construye una base de datos SQL sobre la que la herramienta hará consultas de acuerdo a los modos de interacción y a las acciones realizadas por el usuario. Las consultas obtendrán respuestas en formato JSON tanto del modelo de preguntas como de los datos de las encuestas en sí. La definición del modelo de

preguntas permite escalar la herramienta y extender sus funcionalidades acordes a las modificaciones que, desde el negocio, se deseen realizar en el futuro.

IV. MÉTODO PROPUESTO

A. Estrategia de vistas y representación de los datos

Inicialmente, por la naturaleza de los datos se plantea una solución en tres vistas diferentes. Esta propuesta se apoya fuertemente sobre el principio de visualización de B.Shneiderman [6] “*overview first, zoom and filter, then details on demand*” – visión general primero, enfoque y filtrado, luego detalles a petición-. Como tal se definieron tres vistas con tres tipos de gráfico diferentes que respondían a los intereses definidos anteriormente.

La primera vista planteada (Figura 1) es una vista de coordenadas paralelas en donde cada uno de los ejes representaba una de las 43 preguntas de la encuesta. La vista permitía hacer un filtro de las preguntas que eran mostradas, desplazarse horizontalmente por la gráfica y reordenar los ejes a voluntad del usuario. Al seleccionar un segmento de un eje se mostraba un resumen estadístico de la selección en un cuadro inferior.

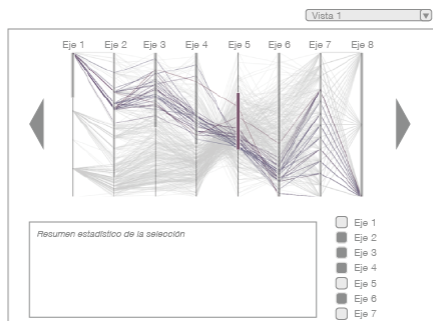


Figura 1: Mockup de la primera vista, ejes paralelos.

La segunda vista (Figura 2) pretendía sobreponer dos poblaciones definidas por el cruce de los atributos de los datos que fueron descritos en la sección ‘Caracterización de los datos’. Cada población estaría representada por un elemento, una geometría que se ubicaría sobre unas coordenadas polares. Cada eje de este sistema de coordenadas es una de las opciones de respuesta de la escala Likert. La gráfica resultante es

un polígono radial. Para comparar los resultados, las geometrías son superpuestas una a otra.

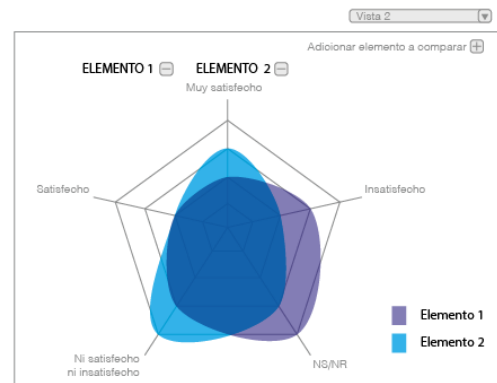


Figura 2. Mockup de la segunda vista, geometrías superpuestas.

El uso de geometrías superpuestas fue descartado por introducir interpretaciones erróneas sobre la intersección de las figuras. Este aspecto es discutido con detalle en la sección ‘Resultados y Evaluación’.

La tercera y última vista (Figura 3) pretendía dar cuenta de la evolución en el tiempo de la satisfacción frente a una categoría o grupo de preguntas de la encuesta. Esta vista se componía de un diagrama de líneas en donde se relacionara el nivel de satisfacción y el año de medición para cada una de las preguntas. Cada pregunta tenía una línea de un tono diferente al resto y de igual manera se podía filtrar las preguntas que eran visualizadas en la gráfica.

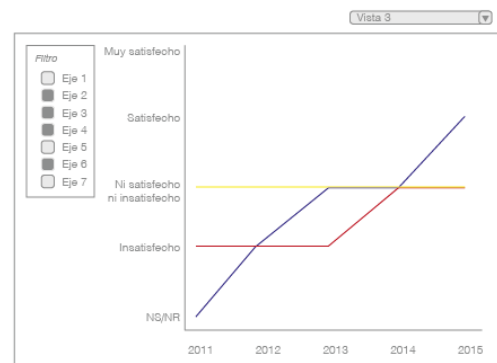


Figura 3. Mockup de la tercera vista, gráfico de líneas

Esta vista también fue descartada por introducir interpretaciones erróneas en la continuidad de la línea, pues se debe tener en cuenta que el año es una variable categórica ordenada y como tal no posee continuidad.

B. Definición de la Navegación Jerárquica

El dominio del problema requiere de un manejo jerárquica de las categorías de preguntas de la encuesta y de los programas de estudios de la universidad. Por lo tanto, se experimentó con dos representaciones gráficas de información jerárquica: árbol n-ario (Figura 4) y árbol-vp (Figura 5).

El modelo de datos de las encuestas contempla que un programa hace parte de un departamento que a su vez hace parte de una facultad. Adicionalmente, se pueden clasificar los programas por el nivel de estudio. Por lo tanto se toma el nivel de estudio como primer nivel de jerarquía, bajo el cual se encuentra el nivel de facultad, luego el nivel de departamentos y por último los programas.

Para representar esta jerarquía se pensó en utilizar un árbol n-ario que fuera ubicado en el lado izquierdo de la gráfica de ejes paralelos. El despliegue y selección de las hojas de el árbol, que corresponden a programas de estudio de la universidad, permitiría la agregación de la serie de datos que corresponda a la población especificada por la ruta de navegación en el árbol (Nivel de estudio > Facultad > Departamento > Programa) a la gráfica de ejes paralelos.

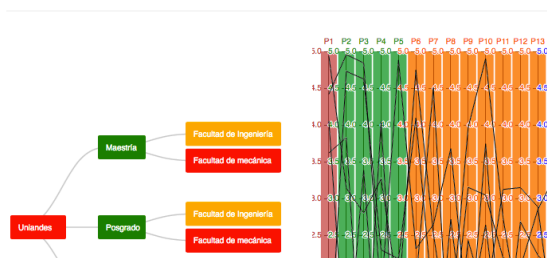


Figura 4. Ejemplo de navegación con un árbol n-ario lateral

Al escalar la cantidad de programas de estudio esta opción deja de ser viable, por lo que se decide implementar un árbol-vp. Esta estructura gráfica asigna una región definida para el universo de muestra y a través de acciones de enfoque se puede explorar los hijos de cada uno de los nodos en el espacio.

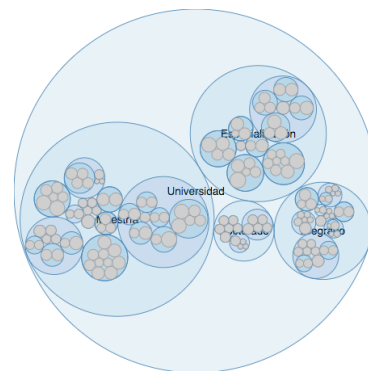


Figura 5. Ejemplo de navegación con un árbol-vp

C. Técnicas de interacción y definición final

Para mantener la estrategia de vistas, por la complejidad de las dimensiones de los datos se introduce una relación entre el árbol-vp y el gráfico de ejes paralelos. Se desea mantener la exploración del universo de muestra que provee el árbol-vp con la definición de tendencias que provee la representación de los resultados de la encuesta en los ejes paralelos. Por lo tanto, se construye un elemento adicional para el gráfico de ejes paralelos que permite arrastrar y soltar nodos del árbol-vp y así definir las series de datos que serán visualizados. La unión y correspondencia entre estos construye una unidad en la interfaz (Figura 6).

Al soltar el nodo sobre la barra a la izquierda del gráfico de ejes paralelos se agrega la serie de datos correspondiente al gráfico. Por lo tanto el usuario pasa por un proceso de exploración en el árbol-vp, selección y arrastre del nivel deseado a la barra lateral. Al hacer esto puede realizar un análisis de tendencia por los datos mostrados en los ejes paralelos. Estos ejes pueden ser ordenados a criterio del usuario, pues se conoce que cada perfil de usuario tiene intereses específicos sobre diferentes preguntas. Esto le permite reorganizar las preguntas de su interés y desplazar las que no desea ver. Al posicionar el mouse sobre los ejes el usuario puede ver a la pregunta a la que hace referencia el eje, con su tema y subtema si aplica.

Así mismo, se incluye un diagrama de barras a manera de resumen. Esta diagrama presenta un resumen promedio de satisfacción del nodo en el que se encuentra el usuario en su proceso de exploración por año, de esta forma se evidencia el cambio en el

tiempo en la percepción de satisfacción en la población estudiantil del nodo actual.



Figura 6. Interfaz de la herramienta con los elementos de interacción

V. RESULTADOS Y EVALUACIÓN

Durante la construcción de la herramienta se realizan diferentes evaluaciones durante el proceso con expertos en diseño, visualización y un experto del dominio. Estas evaluaciones son de carácter cualitativo y buscan analizar la eficacia y comprensión de la herramienta construida.

La evaluación inicial se realiza sobre la estrategia de tres vistas. Esta evaluación se realiza con expertos en visualización y en diseño utilizando prototipos en papel de las vistas ya definidas. Se pretende revisar la comprensión de los elementos de interfaz y de la intención de cada una de las vistas.

La evaluación fue realizada con tres usuarios expertos en visualización y dos en diseño. En la evaluación con los expertos de visualización hay un observador y el experto. En la evaluación con expertos en diseño hay dos observadores y los expertos. La evaluación consiste en poner sobre la mesa los prototipos en papel y cuestionar al experto sobre el objetivo de cada una de las vistas para evaluar la comprensión la vista en general, luego de haberlo contextualizado brevemente sobre el proyecto y el dominio del problema. Luego de que el experto ha dado su explicación de la vista se le solicita evaluar los elementos de la interfaz y su utilidad y relevancia. Finalmente, el experto emite una revisión final con sus consideraciones y recomendaciones.

Esta primera evaluación confirmó la utilidad de los filtros por la complejidad del modelo de datos que tiene el proyecto, que debió ser explicado a detalle a los expertos. En esta evaluación se descartaron la segunda vista de geometrías sobrepuestas y la tercera, el gráfico de líneas para los años de medición.

En la gráfica de geometrías sobrepuestas (ver Figura 2), si bien en las pruebas se entendía la utilidad de las geometrías individuales, al momento de sobreponerlas los expertos asociaban el área de intersección entre las geometrías como una nueva relación que podía crearse y que daba cuenta de algún fenómeno o interacción entre los datos que podía surgir, aunque este no es el caso. Por lo tanto la vista fue descartada pues la distracción que esta intersección introducía era mayor a la utilidad que proveía la vista en general.

De igual forma, la gráfica de línea continua (ver Figura 3) connotaba una continuación en la medición por la continuidad de la línea, lo que implica la existencia de un crecimiento o decrecimiento gradual en la medida de satisfacción. Sin embargo, la encuesta son mediciones puntuales, por lo que se decide abordar esta situación con un método diferente, a manera de filtro.

La siguiente evaluación se realizó sobre la navegación jerárquica de árboles de la estructura de los programas. El objetivo de la evaluación fue conocer la capacidad de un usuario de comprender la selección del programa en la interfaz para visualizar los datos. Esta evaluación se realizó con usuarios comunes con poco destreza en el manejo de herramientas tecnológicas, pues un perfil de usuario de interés para el cliente son los coordinadores académicos de áreas ajenas a la ingeniería y a la tecnología que no suelen usar herramientas computacionales en su ejercicio profesional.

La evaluación consistía en un prototipo funcional de la navegación sobre la que se evaluaba la facilidad de interacción con la estructura y la conveniencia de los gestos de selección. Esta evaluación permitió verificar la eficacia de la estructura en la comprensión del dominio de los datos, pero presentó inconvenientes de visualización al momento de extender sus funcionalidades pues la extensión del árbol sobrepasaba los límites del espacio. Se

aplicaron técnicas de enfoque como ojo de pescado y *scroll*, pero estas no fueron efectivas en la solución del problema (Figura 7), por lo que se migró la navegación a un árbol-vp.

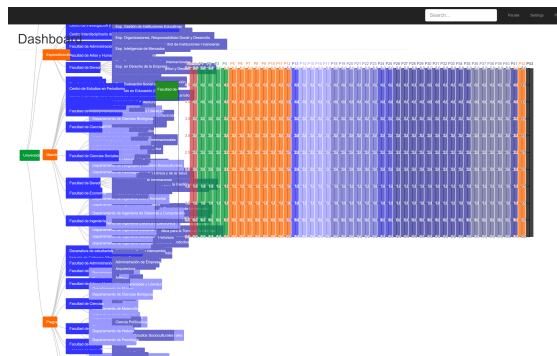


Figura 7. La extensión total del árbol de navegación causó inconvenientes de visualización e interacción en la interfaz.

VI. DISCUSIÓN Y CONCLUSIÓN

Se construyó una herramienta de análisis visual de encuestas de satisfacción. Se analizó diferentes estrategias de diseño y desarrollo de otras soluciones tanto para el dominio del problema como para casos con modelos de datos similares. Se verificó la conveniencia de una interfaz amigable y gestos de interacción intuitivos para los perfiles de usuario de la herramienta que permite realizar análisis comparativos de tendencias de satisfacción y obtener el resumen por año de medición.

Pedagógicamente se utilizó la metodología de T. Munzner[7] para la definición de tareas y concepción del proyecto de analítica visual. Se evidencia la relevancia de la analítica visual y del pensamiento de diseño en el ejercicio de la ingeniería en su relación con otras disciplinas. Es clara la necesidad de establecer objetivos y tareas específicas para este tipo de proyectos, de tal forma que estas tareas definan las acciones, transformaciones y representaciones de los datos. La evaluación y validación con expertos del dominio y de visualización y diseño arrojó hallazgos relevantes

sobre el proceso de construcción de la herramientas y sobre la lógica de negocio a la que deben responder los datos y la visualización propuesta. La iteración sobre prototipos y la evaluación sobre la marcha también resultaron ser procesos de diseño de suma importancia para agregar valor al proceso de construcción y, consecuentemente, entregar valor al usuario final.

Se espera para la siguiente fase del proyecto mejorar la funcionalidad para algunos casos de movimiento de elementos por la interfaz, para dar continuidad al gesto y no afectar la concentración del usuario en su tarea al utilizar la herramienta.

REFERENCIAS

- [1] J.Zhang (2015, Nov. 24). *Survey Analytics Web Application* [Video] Disponible: https://www.youtube.com/watch?v=at0gjhv_jOQ
- [2] S. Evergreen (2014, Nov.12) *Visualizing likert-type Data: Try Aggregated Stacked Bars* [En línea] Disponible: <http://stephanieevergreen.com/aggregated-stacked-bars/>
- [3] D.Brodbeck y L.Girardin,(2003, Oct.) “*Visualization of Large-Sacel Customer Satisfaction Surveys Using Parallel Coordinate Tree*”, Macrofocus GmbH, Seattle, InfoVis 2003 Symp. On Inform. Visualization.
- [4] A. Van Dam y R.Lightner,(2016, Dec. 2) “Track National Unemployment, Job Gains and Job Losses” [En línea] Disponible: <http://graphics.wsj.com/job-market-tracker/>
- [5] LSECities et al.(2014) “*¿Cómo se gobiernan las ciudades? Construyendo una base de datos global de los actuales modelos de gobernanza urbana*” [En línea] Disponible: <https://urbangovernance.net/es/>
- [6] B. Shneiderman (1996) “*The Eyes Have It: A Taks by Type Taxonomy for Information Visualization*” Washington, Proceedings of the IEEE Symposium on Visual Languages, pg. 336-343.
- [7] T.Munzner (2014, Dic.) “*Visualization Analysis and Design*” K Peters/CRC Press.