

Visual Analytics aplicado a la Deforestación en Colombia

Andrés F. Tenorio, José F. Moreano, and Javier A. Ortiz

Abstract—Este trabajo tiene como objetivo principal presentar los resultados del proyecto de aplicación de visual analytics al informe de Deforestación en Colombia de la revista Semana. El marco teórico sobre el cual se sustenta la aplicación de las técnicas es el propuesto por la profesora Tamara Munzner, sin menoscabo de una revisión complementaria del estado del arte, que se espera sea de utilidad para el lector. A partir de la identificación de las tareas principales del informe de deforestación, validadas de manera iterativa con el equipo investigador de la revista Semana, y de la revisión de los datos del estudio, se realiza una propuesta de diseño para la visualización que es el resultado de interpretar adecuadamente el dominio del problema y abstraer de manera acertada las tareas, datos y usuarios de las visualizaciones.

Index Terms—Narrative visualization, data journalism, visual analytics, storytelling

INTRODUCCIÓN

A partir de los trabajos seminales de medios digitales como The New York Times, The Guardian o The Washington Post, en los cuales se realizaron producciones periodísticas dirigidas por datos, se abrió una vertiente de investigación interdisciplinar en la que se encuentran el diseño, la analítica, el periodismo, y las ciencias de la computación, con el propósito de aportar a lo que se ha denominado el periodismo de datos.

En Colombia los grandes medios digitales han incursionado también en esta vertiente, y han realizado algunos especiales multimedia enmarcados de manera primaria en el periodismo de datos.

De manera específica la revista Semana, uno de los medios impresos y digitales más importantes del país, desde hace algunos años se ha preocupado por generar especiales multimedia que presenten las problemáticas más relevantes de la nación de una manera interactiva y de fácil interpretación por parte de los usuarios.

Una de estas líneas de investigación de la revista Semana es la de sostenibilidad, dentro de la cual se enmarca el especial de Deforestación en Colombia, el cual tiene como objetivo principal la presentación del avance de la tragedia ambiental en el país, identificando las principales causas, y haciendo énfasis en regiones muy bien determinadas sobre las cuales se realiza una investigación periodística de campo.

En el marco de este especial sobre deforestación se decide realizar un ejercicio de aplicación de *visual analytics* que contribuya a la generación de *insights* con base en los datos dispuestos para el informe y que permitan la interacción del usuario final con los resultados de la investigación periodística.

De esta manera, el valor agregado del proyecto se encuentra en la aplicación del marco conceptual y teórico que permita definir el diseño óptimo de visualizaciones para determinar los *insight* específicos, y su posterior despliegue en un desarrollo de licenciamiento abierto que sea de uso de la revista Semana para su informe.

1 ESTADO DEL ARTE

En el marco del proyecto es importante destacar la relevancia que el periodismo de datos (*data journalism*) ha venido adquiriendo en los últimos años y la manera en la se ha desplegado un núcleo de investigación completo sobre problemáticas asociadas al mismo.

En [2] se presentan algunas de las implicaciones más importantes que la creación de *data stories* generan para el periodismo, y se realiza una propuesta respecto a los flujos de trabajo novedosos que deben propiciarse a raíz de la aparición de estos nuevos formatos. De esta manera, se está significando la relevancia que tienen esta nueva forma de producir investigaciones periodísticas, y cómo se hace necesario adaptar los procesos de producción del periodismo a estos flujos interdisciplinarios.

Esta propuesta de *workflows* es complementaria a la presentada por [8] en donde se presentan las fases del proceso de *storytelling* a partir de tres pasos principales: la exploración de los datos, la concepción de la historia y la manera en la que se cuenta la misma.

En esta misma línea en [3] se proponen las siguientes fases de desarrollo del proceso de *visual storytelling*, por lo que se discute sobre el primer paso: encontrar *insights*, seguido de la conversión de estos en una narración (hacer la historia) y finalmente, comunicar la historia.

De esta manera se constata que aún cuando en la actualidad se hace uso de marcos conceptuales y técnicas novedosas en el periodismo de datos, todavía se persigue el mismo objetivo, el cual es la efectividad en el mensaje, de ahí su relación con el *storytelling*.

Desde esta perspectiva se han iniciado líneas de investigación que están más estrechamente relacionadas con el efecto que pueden llegar a tener las visualizaciones sobre el usuario y las mejores prácticas para aumentar la efectividad de las mismas.

En [4] se propone un *framework* analítico sobre el diseño de técnicas de visualización que pueden afectar significativamente la interpretación del usuario final.

Así mismo, en [5] se presenta una perspectiva crítica en torno a la pregunta de si realizar representaciones gráficas de los datos permite transmitir el mensaje con mayor persuasión. Con base en la realización de experimentos sobre 720 participantes encuentran que el poder de persuasión de las visualizaciones depende de la actitud inicial del usuario.

En [6] se presenta la importancia de lograr con las *narrative visualizations* transmitir los *insights* hallados en datos complejos. Desde esta perspectiva se parte de caracterizar los *insights*, destacando el hecho que '*el propósito de la visualización es el insight no el gráfico*'.

Otro aspecto relevante que se evalúa en [7] es la comparación entre el entendimiento de la visualización y el grado de recordación de la misma por parte del usuario. Los autores realizan una revisión sobre las visualizaciones con mayor recordación en diferentes categorías: publicaciones científicas, medios de comunicación, organizaciones gubernamentales e infografías. Encuentran que el tipo de visualización en cada sector es diferente, y que las visualizaciones más recordadas no necesariamente son las que más entienden los usuarios, sino aquellas con los cuales adquieren una mayor vinculación por su aspecto estético en un primer momento.

-
- Andrés F. Tenorio. E-mail: af.tenorio@uniandes.edu.co
 - José F. Moreano. E-mail: jf.moreano@uniandes.edu.co
 - Javier A. Ortiz. E-mail: Ja.ortiz905@uniandes.edu.co

Este aspecto estético es complementario a uno de los factores que ha contribuido de manera significativa a la expansión del uso de visualización: la interactividad. En [8] se indica que una de las principales causas por las cuales las visualizaciones son tan útiles, es porque tienen la habilidad de representar grandes cantidades de datos complejos en una historia, y que su uso en los medios digitales se debe a la facilidad de interacción que se le ofrece al usuario.

La importancia que tiene la interactividad de las visualizaciones ha generado una serie de investigaciones que ahondan en la mejor manera de realizar las transiciones y animaciones entre las diferentes gráficas estadísticas. En [10] se propone una taxonomía de transición de las visualizaciones que incluye:

- Transformación de la vista: la cual implica un cambio en el punto de vista, simulado mediante el movimiento de cámara en un espacio virtual. Dentro de esta categoría se encuentra el efecto de zoom.
- Transformación en el entorno: esto afecta por ejemplo los ejes de la gráfica, como puede ser una transformación logarítmica sobre los ejes.
- Transformación por Filtro: lo que implica la visualización de un subconjunto de elementos.
- Transformación por ordenamiento: mediante lo cual se realiza un arreglo especial de los datos que permitan por su naturaleza darles un orden.
- Transiciones de paso de tiempo: lo que implica aplicar cambios a partir de la dimensión tiempo a los valores de las gráficas.
- Cambios de visualizaciones: lo cual implica afectar todo el nivel sintáctico de la gráfica, esto es, las marcas como formas, color y tamaño.
- Cambios en el esquema: realizado mediante una agregación de nuevas dimensiones columnares de la tabla de datos al gráfico.

Esta taxonomía permite identificar las principales transformaciones que pueden aplicarse a una secuencia interactiva de visualizaciones. Además, citando a Kosslyn, identifican tres niveles de análisis para los gráficos basados en datos:

Sintáctico: el cual se enfoca en el tipo de marcas que se utilizan.

Semántico: en el que el análisis se enfoca en los canales que se utilizan con base en el tipo de datos.

En Segel [11] se sugieren estrategias de diseño para la visualización narrativa periodística a partir de la revisión de visualizaciones propuestas The New York Times, The Guardian, Financial Times, The Washington Post. Las principales estrategias identificadas son:

- Anotación: lo que implica generar marcas sobre las visualizaciones con notas.
- Matching sobre el contenido: esto es mantener el mismo esquema entre las diferentes visualizaciones de tal manera que se facilite la comparación entre las mismas por parte del lector.
- Manejo de la luz: lo cual facilita el resaltado de contenido específico mediante el control de la luminosidad o del color.
- Barra de progreso: que permite al lector seguir una ruta narrativa.

- Detalles por demanda: lo cual da la oportunidad al usuario de generar solicitudes específicas a las visualizaciones.

Estas estrategias de diseño complementan la taxonomía de Heer [10] en el sentido que permiten generar un equilibrio entre la propuesta narrativa del autor y el descubrimiento de la historia por parte del usuario. De esta manera Segel [11] propone los siguientes mecanismos no visuales que facilitan la conducción narrativa:

- Ordenamiento: la ruta que puede seguir el usuario a través de la visualización
- Interactividad: las distintas maneras en las que el usuario puede manipular las visualizaciones
- Mensajería: las distintas maneras en las cuales la visualización se transmite al espectador.

Con base en estas tres tácticas de estructuras narrativas proponen tres patrones de diseño para historias interactivas:

- Copa de Martini: en el cual la narración visual es principalmente dirigida por el autor.
- Diapositivas interactivas: en el que se hace uso de una presentación interactiva que permite al usuario explorar determinados tópicos
- Desglose de historias: que es el enfoque más próximo a dar al usuario complete autonomía sobre el desarrollo narrativo de la visualización.

En un trabajo posterior, Stolper [12] realiza una revisión de las técnicas de narración basadas en datos dentro de las cuales incluyen las de Heer. Las técnicas categorizadas por Stolper son:

Narración y explicación de datos: que consiste en el uso de una gran cantidad de texto para transmitir los puntos más relevantes de la investigación, intercalando visualizaciones y su detalle textual.

Vinculación de elementos: que permite una relación más interactiva entre el texto y las visualizaciones. Dentro de esta categoría se encuentran las técnicas que engloba el brushing & linking, lo que permite la vinculación de elementos a través del color, la animación o la interacción.

Mejora de la estructura y la navegación: lo que facilita el seguimiento de la estructura de la narración mediante el orden: se incluyen técnicas como scrolling, el breadcrumbs, los botones de encabezado de sección y el timeline.

Proporcionar exploración controlada: aquí se incluyen técnicas como consultas dinámicas, exploración visual embebida y la exploración visual por separado, pero con la motivación principal de restringir el espacio de interacción del usuario de tal manera que no se pierda en los datos.

2 CARACTERIZACIÓN

La propuesta inicial está basada en las tareas principales presentadas por el cliente, la identificación conjunta de los usuarios y la revisión preliminar de los datos puestos a disposición del proyecto.

2.1 What

Existen varias causas de deforestación en Colombia. El Ideam estableció que las principales causas de la deforestación son seis: el acaparamiento de tierras, los cultivos ilícitos, la construcción de infraestructura, la ganadería extensiva, los incendios forestales y la minería, en ese orden de importancia.

Para este proyecto y de acuerdo con el requerimiento de la revista Semana se realizará énfasis en la causa de cultivos ilícitos y

cuál es el impacto en la deforestación en Colombia en los últimos 10 años.

En este orden de ideas se cuenta con un dataset tipo tabla cuyos elementos principales son los municipios de Colombia y los siguientes atributos:

Tabla 1. Resumen del dataset dispuesto por el cliente

Elemento / Atributo	Tipo	Periodo de tiempo
Municipio	Catégorico	
Hectáreas deforestadas	Ordenado secuencial	1990-2016
Hectáreas de cultivos de coca	Ordenado secuencial	1999-2016
Hectáreas de cultivos de amapola	Ordenado secuencial	1999-2016
Hectáreas superficies Bosques Naturales	Ordenado secuencial	1990-2016

2.2

2.3 Why

Las actividades principales establecidas a partir de la sesión de trabajo con el cliente, están asociadas a dos usuarios principales:

Investigadores y reporteros de Semana
Público en general / Lectores

En ambos casos la tarea de más alto nivel es la de consumir información pero se espera enfocar el descubrimiento de hallazgos interesantes para los investigadores y reporteros del cliente, cuidando que el público en general tenga un disfrute y fácil entendimiento de la información presentada en la visualización.

Adicionalmente, se plantean las siguientes tareas específicas:

- Identificar las correlaciones entre la deforestación y los cultivos de coca en el periodo de tiempo analizado.
 - Identificar la tendencia de la deforestación y sus principales causas a nivel nacional y municipal.
 - Identificar los municipios con mayor número de hectáreas deforestadas por año.
 - Localizar características específicas. Será posible seleccionar un municipio y ver cómo se comporta la deforestación y sus causas asociadas.
- ☐ Comparar similitudes en el comportamiento de la deforestación en el tiempo por departamento.

2.4 How

Teniendo en cuenta que el dataset es una tabla, y las tareas principales, se propone el uso los siguientes modismos principales:

- ☐ Barchart
- ☐ Piechart
- ☐ Radial chart
- ☐ Linechart
- ☐ Radial Treemap

3 REPLANTEAMIENTO DE LA SOLUCIÓN.

Inicialmente la propuesta está basada en las tareas principales presentadas por el cliente, la identificación conjunta de los usuarios y la revisión preliminar de los datos puestos a disposición del proyecto Durante la revisión realizada con el cliente del planteamiento inicial, se identifica que los datos obtenidos presentan un contenido temporal importante que puede ser mejor visualizado a partir de componentes dinámicos. Este análisis fue realizado siguiendo las fases descritas anteriormente para el proceso de story telling, ya que es la metodología que se considera apropiada para que el usuario pueda explorar los datos de manera divertida.

A partir de la aplicación de revisiones iterativas se replantearon dos modismos principales que hacían parte de la solución inicial. Estos dos modismos fueron: Scatterplot y Parallel Coordinates, los cuales son expresamente retirados de la solución por solicitud del cliente, teniendo en cuenta la dificultad para su interpretación por parte de los usuarios de Semana.

Adicionalmente, se cuenta con una limitación sobre la totalidad del uso de los datos, dado que no todos son públicos, y por lo tanto, se debe generar una solución que considere solo una de las casus principales de deforestación, en particular, los cultivos de coca.

LISTA ORGANIZADA DE LOS DEPARTAMENTOS CON MAYOR DEFORESTACION EN EL PERIODO SELECCIONADO

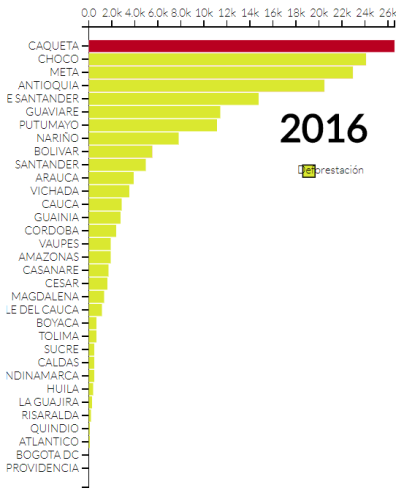


Fig. 1. Uso de barchart en la solución

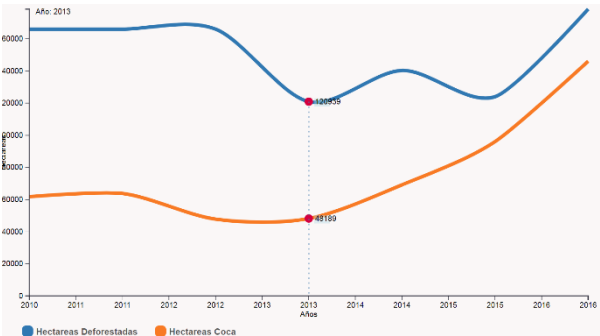


Fig. 2. Uso de linechart en la solución

Tabla 2. Descripción modismos para la visualización

Modismo	Tareas
Linechart	Encontrar las correlaciones entre la deforestación y los cultivos de coca.
Barchart	Localizar los departamentos con mayor número de hectáreas deforestadas por año.
Parallel coordinates	Encontrar las correlaciones entre la deforestación y los cultivos de coca
Radial Treemap	Resumir los municipios con mayor número de hectáreas deforestadas.

Así mismo, se realiza la vinculación de tres modismos mediante la técnica linking & brush, de tal manera que se

4 EVALUACIÓN

Las pruebas de uso sobre la solución se generaron en dos escenarios diferentes. En primer lugar, se realizó una presentación a expertos en el uso de técnicas de visual analytics del curso de la Universidad de los Andes. De esta sesión, el principal resultado estuvo relacionado con la escalabilidad de la solución teniendo en cuenta la cantidad de municipios de análisis, razón por la cual se optó por trabajar a nivel departamental.

Posteriormente se realizó presentación de prototipos iterativos con el equipo de diseñadores del cliente, lo cual permitió tener un refinamiento preciso sobre lo deseado en la solución. Los principales resultados de estas sesiones se relacionan con el uso de modismo de fácil entendimiento para los lectores de la revista Semana, descartando el uso de interacciones y, en general, modismos, que requieran de un entrenamiento previo por parte del usuario.

5 CONCLUSIONES

Con base en los datos obtenidos y las reuniones realizadas con el cliente se identifica una amplia oportunidad de mejora en la visualización del problema de la deforestación en Colombia. Aplicando los conceptos actuales sobre visualización de datos y con el apoyo de la metodología expuesta por Tamara es posible brindar a los usuarios una herramienta que permita explorar mejor los datos garantizando siempre cumplir con los principios de efectividad y expresividad.

Teniendo en cuenta las características de los datos, y la especificación de los usuarios finales, se plantea una solución completa que hace uso de modismos de fácil entendimiento y que permiten presentar los hallazgos principales que son de utilidad para el cliente: La tendencia constante en tres departamentos principales en los cuales la deforestación ha alcanzado los más altos índices desde 1990 (Caquetá, Meta y Antioquia); y la tendencia común que existe entre la deforestación y los cultivos de coca en el país.

AGRADECIMIENTOS

Los autores agradecen al equipo de periodistas e investigadores de la revista Semana.

REFERENCIAS

- [1] T. Munzner. Visualization Analysis and Design. AK Peters 2014.
- [2] W. Weber., H. Rall. Data visualization in online journalism and its implications for the production process. 2012
- [3] B. Lee., N. Henry., P. Isenberg., S. Carpendal. More than telling a story: Transforming data into visually shared stories.
- [4] J. Hullman., N. Diakopoulos. Visualization Rhetoric: Framing effects in narrative visualization.
- [5] A. Vikram., A. Manivannan., O. Nov., M. Satterthwaite., E. Bertini. The Persuasive Power of Data Visualization.
- [6] G. Dove., S. Jones. Narrative visualization: Sharing insights into complex data. 2012.
- [7] M. Borkin., A. Vo., Zo. Bylinski., P. Isola. What makes a visualization memorable?. 2013.
- [8] K. Gupta, S. Sampat, M. Sharma, V. Rajamanickam. 2016. Visualization of election data: Using interaction design and visual discovery for communicating complex insights.
- [9] J. Hullman, S. Drucker, N. Riche, B. Lee, D. Fisher, and E. Adar. A Deeper Understanding of Sequence in Narrative Visualization. IEEE Transactions on Visualization and Computer Graphics, 19(12):2406–2415, Dec. 2013
- [10] J. Heer, G. Robertson. 2007. Animated transitions in statistical data graphics.
- [11] E. Segel, J. Heer. 2010. Narrative visualization: Telling stories with data. IEEE Transactions.
- [12] C. Stolper, B. Lee, N. Henry, J. Stasko. 2016. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories.
- [13] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, Sheelagh Carpendale. More than Telling a Story: A Closer Look at the Process of Transforming Data into Visually Shared Stories. IEEE Computer Graphics and Applications, Institute of Electrical and Electronics Engineers, 2015, 35 (5), pp.84–90.
- [14] M. Bostock, V. Ogievetsky., J. Heer. D3: Data Driven Documents. 2011
- [15] Miqing Li, Liangli Zhen, Xin Yao. How to Read Many-Objective Solution Sets in Parallel Coordinates. 2017
- [16] Bahador Saket, Arjun Srinivasan, Eric D. Ragan, Alex Endert. Evaluating Interactive Graphical Encodings for Data Visualization. 2015
- [17] Derya Birant, Alp Kut, An algorithm for clustering spatial-temporal data. 2006.
- [18] Mihael Ankerst, David H. Jones, Anne Kao, Changzhou Wang DataJewel: Tightly Integrating Visualization with Temporal Data Mining. 2008.