

# Análisis Exploratorio de Datos Geográficos Voluntarios: estudio de caso en la administración zonal Manuela Sáenz de Quito

## *Exploratory Data Analysis of Volunteered Geographic Information: case study in the Manuela Sáenz administrative zone of Quito*

R. Castro, M. Luna

**Resumen**—El desarrollo tecnológico de la primera década de este siglo provocó una explosión de contenidos generados por usuarios y compartidos en la web. En geografía, esta tendencia se denomina Información Geográfica Voluntaria; OpenStreetMap (OSM) es su proyecto más reconocido. Si bien se ha investigado a OSM extensivamente —en especial para conocer su calidad— tales investigaciones no se han realizado en Ecuador, a pesar de que sí se ha mapeado en este país. Este artículo propone un Análisis Exploratorio de Datos como primer acercamiento a conocer las características de OSM. El área de estudio es la Administración Zonal Manuela Sáenz de Quito. Se utilizó software libre SIG y de programación, aplicando métodos no-gráficos y gráficos. Los resultados se presentan en varios gráficos y se discuten los fenómenos relacionados a número de nodos y de keys, condición de área y los keys: building, highway y source. Se concluye indicando aspectos de OSM que requieren futura investigación.

**Palabras clave**—análisis exploratorio de datos, lenguaje R, OpenStreetMap, VGI

**Abstract**—Technological development during the first decade of this century provoked a boom of user-generated, web-shared media. In geography, such trend is called Volunteered Geographic Information; OpenStreetMap (OSM) is its most famed project. While several research has been made about OSM -particularly, concerning its quality-, none of such has been made in Ecuador in despite that, indeed, there has been mapping in this country. This paper proposes an Exploratory Data Analysis as first approach to understanding the features of OSM. Study area is the Manuela Sáenz administrative zone of Quito. Free GIS and programming software were used, while applying non-graphic and graphic methods. Results are displayed in several plots and phenomena are discussed, regarding number of keys and nodes, area condition, and building, highway and source keys. The paper concludes identifying OSM aspects for further research.

**Index Terms**—exploratory data analysis, OpenStreetMap, R language, VGI

### I. INTRODUCCIÓN

DURANTE la primera década del siglo corriente, el desarrollo de un conjunto de tecnologías -particularmente la denominada Web 2.0- permitió dar el salto hacia una internet más colaborativa; fruto de este salto fue la explosión de contenidos generados por usuarios [1] [2] [3]. El ámbito geográfico no fue una excepción: la generación colaborativa de datos geoespaciales ha recibido considerables atención y análisis; de esta manera, se han desarrollado conceptos como la “Información Geográfica Voluntaria” (VGI, por sus siglas en inglés) [1] y la “Wikificación del SIG” [2].

Entre las iniciativas que se han creado en el marco VGI, OpenStreetMap (OSM) es posiblemente la más reconocida. Desde 2004 es “un esfuerzo internacional por crear una fuente de datos libre, disponible en un mapa, a través del esfuerzo voluntario” [1]. Así como otros proyectos VGI, la naturaleza OSM es comunitaria, de roles fluidos, de propiedad común, y “palimpsestica” (un producto incompleto pero que crece continuamente) [3]; de todas maneras, es una alternativa económica e inmediata en varios lugares del mundo donde un mapa digital no existe o no se encuentra liberado.

Aunque la comunidad científica, las instituciones públicas y privadas, las industrias y la gente en general han reconocido el potencial de OSM, aún existe escepticismo respecto a la calidad de estos datos geoespaciales; escepticismo justificado por las motivaciones y las maneras por medio de las cuales los usuarios los aportan [3]. Esto ha motivado, en la última década, una serie de investigaciones alrededor de OSM que se refieren -en general- a cuatro aspectos: Administración y calidad de datos; Contexto social; Modelado de redes y rutas; y Administración

Article history:

Received 07 September 2018

Accepted 24 September 2018

R. Castro y M. Luna pertenecen al Departamento de Ciencias de la Tierra y de la Construcción de la Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador. (emails: rrcastro1@espe.edu.ec, mpluna@espe.edu.ec)

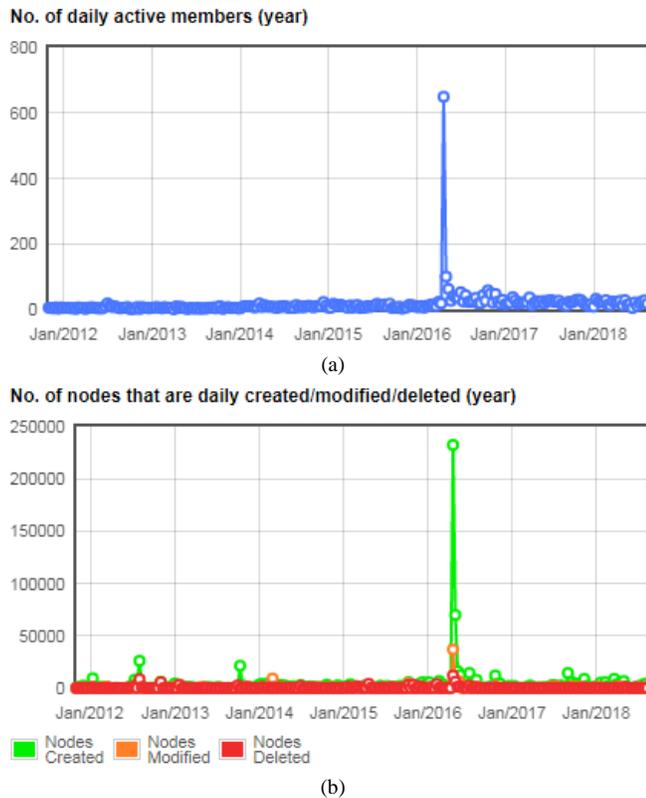


Fig. 1. Miembros activos (a) y nodos creados/modificados/eliminados (b) diariamente en Ecuador, en el periodo enero 2012 – agosto 2018. Observar el pico en abril de 2016. Fuente: <http://osmstats.neis-one.org>.

territorial y urbana [4]. De ellos, el primer aspecto es el que ha sido más investigado. Dado que un proyecto como OSM posee “su propia geografía a través del tiempo y del espacio” (varía considerablemente entre épocas y países), y que “algunos países están emergiendo [...] y aún hay grandes vacíos” [4], es necesario continuar su investigación.

En Ecuador existe una comunidad OSM pequeña en comparación a otros países más desarrollados; no obstante, se han realizado considerables contribuciones a este proyecto. Particularmente, de acuerdo al sitio web <http://osmstats.neis-one.org>, en abril de 2016 se alcanzó un pico de miembros activos y nodos creados -649 y 232462 respectivamente- en un solo día (Fig. 1); este nivel de participación no ha vuelto a suceder [5]. Allende esta comunidad que participativamente ha mapeado en Ecuador, hasta la fecha no se han desarrollado estudios en ninguno de los aspectos mencionados.

En este contexto, el presente artículo propone realizar un Análisis Exploratorio de Datos (EDA, por sus siglas en inglés) en un conjunto de vías OSM, con el propósito de comenzar a caracterizar esta información en el caso particular de una urbe en Ecuador. El EDA es un proceso “fundamental después de la recolección y pre-procesado de datos, donde los datos son simplemente visualizados, ploteados y manipulados sin ninguna suposición, con el propósito de ayudar a evaluar la calidad de los datos y construir modelos” [6]. Típicamente, un EDA incluye métodos de dos tipos: no-gráficos (tabulación, parámetros estadísticos...) y gráficos (histograma, boxplot...); métodos de ambos tipos fueron aplicados en este trabajo.

El EDA es un proceso empírico, descriptivo y útil pues,

```
<way id="31171329">
  <nd ref="346913487"/>
  <nd ref="1488977244"/>
  <nd ref="346913488"/>
  <nd ref="346913490"/>
  <nd ref="346913492"/>
  <nd ref="346913487"/>
  <tag k="addr:city" v="Quito"/>
  <tag k="addr:street" v="Avenida Río Amazonas"/>
  <tag k="building" v="apartments"/>
  <tag k="building:levels" v="13"/>
  <tag k="name" v="Centro Comercial Unicornio II"/>
  <tag k="source" v="Reconocimiento cartográfico por KG 2016"/>
  <tag k="website" v="http://www.ecuadortur.com"/>
</way>
```

Fig. 2. Estructura XML por etiquetas (tags) de una entidad (way) dentro de un archivo .osm. El primer nodo (naranja) y el último (amarillo) son el mismo, indicando que se trata de un área.

realizado adecuadamente, conduce al investigador a descubrir nuevos fenómenos [7]. La aplicación de este y otros métodos propios de la ciencia de datos es coherente en el marco de la VGI, puesto que es posible considerar a la misma como *Big Data*. Este concepto es complicado de definir; sin embargo, tradicionalmente existen tres características que permiten identificar lo que es *Big Data*: volumen, velocidad y variedad [8]. En conjunto, las tres identifican a cantidades gigantescas de datos, generados rápidamente, y por fuentes diversas; las tres son características de OSM y, en consecuencia, justifican los métodos utilizados en el presente trabajo.

## II. BASE TEÓRICA

OpenStreetMap (abreviado OSM) es el “proyecto que crea y distribuye datos geográficos gratuitos para el mundo” [9]. Se puede considerar el proyecto VGI más importante y que mayor acogida ha tenido en los Sistemas de Información Geográfica (GIS, por sus siglas en inglés) y el público en general, alrededor de todo el mundo [4]. A la fecha existen más de 250 aplicaciones [9] que dependen directamente de los datos OSM; lógicamente, los beneficios a obtenerse de dichas aplicaciones serán en función de la calidad de esos datos. Los problemas de calidad (error posicional, coherencia lógica...) pueden dirigir al usuario a realizar malas decisiones [10].

En el contexto de OSM, los puntos (tiendas, restaurantes...) se denominan *nodes*; las polilíneas (calles, puentes...), *ways*; los polígonos (edificios, parques...), *areas*; y los conjuntos de elementos, *relations*. Cada entidad posee una serie de atributos almacenados en la forma de etiquetas, compuestas de una clave y un valor:  $tag = key + value$  (Fig. 2). Las entidades pueden accederse, editarse y descargarse gratuitamente en la forma de un archivo con extensión .osm que posee estructura XML (Extensible Markup Language) [11].

En esta estructura los *ways* son simplemente series de *nodes*, mientras que las *areas* son *ways* donde el primer y el último *node* son el mismo; la Fig. 2 presenta un ejemplo de *area*. Es importante considerar que OSM no es un producto estático, sino que prácticamente en todo momento está siendo aumentado, modificado y revisado. Actualmente existen varios métodos para agregar nuevos objetos: importación masiva de bases de datos, subida de *tracks* GPS, calco de imágenes aéreas suministradas por Bing, entre otros [11].



Fig. 3. Ubicación del área de estudio (Manuela Sáenz) con respecto a las demás Administraciones Zonales del Distrito Metropolitano de Quito. Fuente de la cobertura shapefile: <http://gobiernoabierto.quito.gob.ec>.

### III. MATERIALES Y MÉTODOS

El área de estudio seleccionada corresponde a la Administración Zonal Manuela Sáenz, una de ocho divisiones del Distrito Metropolitano de Quito, ubicada al centro del mismo (Fig. 3). Se escogió con la premisa de que, siendo una zona céntrica y urbana, se encuentra bien mapeada en OSM. Para la realización del EDA se utilizó dos clases de software libre: por un lado, un GIS para la manipulación del aspecto espacial de los datos; y por otro lado un ambiente de programación para el aspecto estadístico de los datos. El software usado es QGIS 2.18 y RStudio 1.1 respectivamente.

QGIS fue utilizado para descargar los datos (.osm) correspondientes al área de estudio y transformarlos a una base de datos SpatialLite (.db); se trata de una base de datos relacional SQLite adaptada al ámbito geográfico [12].

RStudio fue utilizado para extraer las tablas *ways\_nodes* y *ways\_tags* de la base de datos, y arreglarlas de manera que los datos se presenten ordenadamente en *tibbles*; en el contexto de R, un *tibble* es una tabla de datos con funcionalidad y representación mejorada [13]. Los paquetes empleados fueron: *RSQLite*, para conectar al .db; *dplyr* y *tidyr*, para manipular y organizar los *tibbles*; y *ggplot2*, para graficar [14].

Las entidades de interés corresponden solamente a *ways*, distinguiendo los que representan áreas y los que no. Se exploró los datos en función de los *keys* presentes en los *tags*, priorizando aquellos que se encuentren en mayor cantidad y aporten información más relevante; el significado de cada *key* se puede hallar en la web de OSM. Se generaron gráficos acordes a la estructura de los datos y la finalidad del EDA.

### IV. RESULTADOS Y DISCUSIÓN

El primer resultado corresponde a contabilizar *keys* y *ways* distintos, hallando en total 267 y 28146 respectivamente; los cuatro *keys* más presentes (Fig. 4a) son: *building*, que identifica la categoría de un edificio; *source*, la fuente de datos; *name*, el nombre de la entidad; y *highway*, la categoría de una vía. Los cuatro *keys* son escogidos para mayor análisis. Es importante notar que ciertos *keys* se excluyen mutuamente; *highway* y

*building* son un ejemplo, puesto que una vía no es un edificio y viceversa. Los edificios se contabilizan entre los *ways* dado que los polígonos son simplemente líneas cerradas. Por otra parte, la mayoría de *keys* no son excluyentes, razón por la cual las columnas no suman 28146 en este gráfico. La escala de color utilizada se interpreta de la siguiente manera: la condición de un *way* como *area* se computó en una variable lógica (*is\_area* = TRUE identifica un *area*); computar la media de esta variable (TRUE = 1, FALSE = 0) equivale a calcular el porcentaje de *ways* que sí son *areas*. En la Fig. 4a, la escala de color indica un hecho lógico: que los edificios son polígonos y las vías no (salvo ciertas excepciones como los redondeles); pero además indica otros fenómenos, como que las etiquetas de fuente y de nombre son más frecuentes en las vías que en los edificios.

La Fig. 4b exhibe la frecuencia por número de *keys*; es decir cuántos *ways* tienen un *key*, cuántos dos, etc. Se empleó un gráfico de barras y no un histograma dado que el valor máximo apenas es 20; la mediana es tres *keys*. Obsérvese la escala logarítmica en el eje Y; debido a esta transformación no se encuentran graficadas las frecuencias de uno (en 16, 18 y 20 *keys*). La escala de color indica que predominan *ways* en valores medios de *keys*, y *areas* en valores bajos y altos; ello es importante en el sentido de que, a mayor número de *keys*, se conoce mayor información de una entidad y se mejora su exactitud temática. El gráfico parece indicar que unos pocos edificios son los que poseen más información.

La Fig. 5a exhibe el resultado de ejecutar *summary* (función que resume por variable el contenido de una tabla) sobre todos los *ways*, con la información relevante escogida (incluyendo los

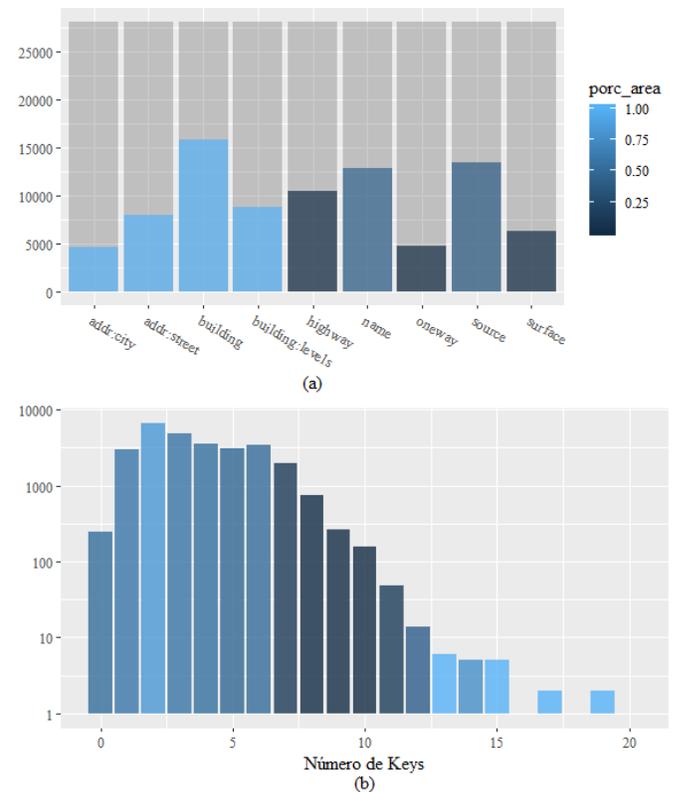


Fig. 4. Gráficos correspondientes a *keys* más presentes (a) respecto al total de *ways* (en gris), y frecuencia por número de *keys* (b). En ambos casos, la escala de color indica el porcentaje de *ways* que son *areas* (1.00 = 100% *areas*).

```

> summary(Tf)
  id          n_nodes      i_node      f_node
Min. : 24159622 Min. : 2.000 Min. :2.617e+08 Min. :2.617e+08
1st Qu.:331558663 1st Qu.: 4.000 1st Qu.:1.625e+09 1st Qu.:1.630e+09
Median :426174164 Median : 5.000 Median :4.245e+09 Median :4.245e+09
Mean :362453554 Mean : 6.361 Mean :3.238e+09 Mean :3.249e+09
3rd Qu.:445113442 3rd Qu.: 7.000 3rd Qu.:4.367e+09 3rd Qu.:4.367e+09
Max. :620418132 Max. :1193.000 Max. :5.862e+09 Max. :5.862e+09

  is_area      n_keys      building      highway
Mode :logical Min. : 1.000 yes : 5046 residential: 4783
FALSE:10671 1st Qu.: 2.000 apartments: 4931 primary : 955
TRUE :17475 Median : 3.000 commercial : 2290 service : 885
Mean : 3.807 house : 1919 secondary : 874
3rd Qu.: 5.000 residential: 1249 footway : 690
Max. :20.000 (Other) : 348 (Other) : 2227
NA's :245 NA's :12363 NA's :17732

  source      name
Yahoo! Aerial Imagery : 6659 Length:28146
Reconocimiento cartográfico de campo 2016 por KG. : 3156 Class :character
Reconocimiento cartográfico por KG 2016. : 998 Mode :character
KG Ground Survey 2016 : 887
Reconocimiento cartográfico de campo 2016 por KG : 724
(Other) : 993
NA's :14729

```

(a)

```

> summary(U)
  is_area      building      highway
Mode :logical apartments : 4931 residential: 4783
FALSE:10671 commercial : 2290 OTROS : 1723
TRUE :17475 house : 1919 primary : 955
OTROS : 5394 service : 885
residential: 1249 secondary : 874
NA's :12363 (Other) : 1194
NA's :17732

  source      n_keys      n_nodes
Bing Imagery : 55 Min. : 0.000 Min. : 2.000
GPS : 87 1st Qu.: 2.000 1st Qu.: 4.000
KG Ground Survey 2016: 6482 Median : 3.000 Median : 5.000
OTROS : 133 Mean : 3.773 Mean : 6.361
Yahoo! Aerial Imagery: 6660 3rd Qu.: 5.000 3rd Qu.: 7.000
NA's :14729 Max. :20.000 Max. :1193.000

```

(b)

Fig. 5. Resumen de variables de interés para todas las entidades; versiones original (a) y corregida (b). Las variables descritas son: id, nodos inicial y final (*i\_node*, *f\_node*), número de nodos y keys (*n\_nodes*, *n\_keys*), condición de área (*is\_area*) y keys de estudio (*building*, *highway*, *source*, *name*).

*keys building*, *source*, *name* y *highway*). Los nodos inicial (*i\_node*) y final (*f\_node*) permiten obtener la condición de *area* (*is\_area*) de un *way*. El valor “Not Available” (NA) en el número de *keys* (*n\_keys*) descubre la presencia de entidades con ninguna etiqueta -el valor correcto sería 0-; por tanto, existen 245 *ways* que se pueden graficar pero carecen información, ni siquiera sus nombres. Se halló 32, 44 y 26 valores diferentes de las variables *building*, *source* y *highway* respectivamente; en muchos casos un valor aparece una única vez. Entonces, se optó por seleccionar en cada variable los valores más abundantes y agrupar los restantes en “OTROS”. Además, se corrigió *source* puesto que algunos valores son en realidad el mismo, escrito de varias maneras: por ejemplo, “Reconocimiento cartográfico por KG 2016.” y “KG Ground Survey 2016”. Con estas consideraciones se generó un *summary* corregido (Fig. 5b); observar que el número medio de *keys* se corrige de 3.80 a 3.77.

El número de nodos (*n\_nodes*) no presenta NAs -puesto que un *way* sin *nodes* simplemente no existe- pero sí valores atípicos, los cuales fueron identificados en un gráfico de caja (boxplot). La Fig. 6 presenta el histograma (gris) de nodos por entidad, recortado hasta un máximo de 100 nodos para evitar dichos valores atípicos; las dos rectas añadidas permiten discernir la condición de *area* (la recta azul representa áreas o polígonos) y la suma de ambas corresponde al histograma (observar que el eje Y es logarítmico). Lo que el gráfico sugiere

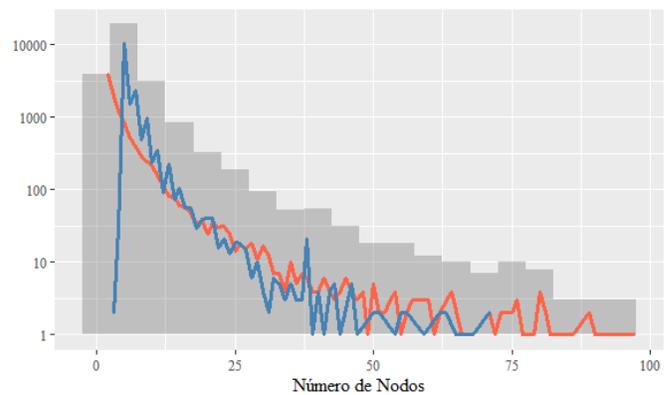


Fig. 6. Histograma (gris) y rectas correspondientes a frecuencia por número de nodos en polilíneas (naranja) y en polígonos (azul).

es que la cantidad de entidades decrece paulatinamente a mayor número de nodos; esto es cierto para las polilíneas, mientras que para los polígonos la frecuencia se dispara desde tres -no hay polígonos con dos nodos- hasta cuatro y cinco nodos, números típicos de vértices en una construcción.

Gráficos de caja fueron implementos para permitir la comparación del número de nodos con las variables *building* y *highway* (Fig. 7). De acuerdo a *building* no hay diferencias significativas en la mediana, excepto para el valor NA -que es difícil de interpretar pues no identifica únicamente a las entidades que no son edificios- que posee menor mediana, pero al mismo tiempo valores atípicos más elevados. En cambio, de acuerdo a *highway*, sí se diferencia la categoría *primary* -vías principales-, con menor mediana. El número de nodos es importante porque está relacionado con qué tan bien se dibujó una entidad: en general, a mayor número de nodos, mejores representación y calidad. También se puede decir que, a breves rasgos, a mayor número de nodos predominan las edificaciones sobre las vías.

Otro aspecto importante del gráfico de caja son los valores atípicos, que aparecen como puntos; tanto en los gráficos como en los estadísticos resalta un valor extremadamente alto de 1193 nodos. A través de su id (561073754), se identificó que se trata un río (<tag k="waterway" v="river"/>) sin información respecto a su nombre o la fuente de datos. No fue posible identificarlo en QGIS debido a que -como se verificó con varios ways-, al descargar datos OSM en este software, se realiza un recorte que deja afuera varios nodos necesarios. Basta que falte un *node* para que no se pueda dibujar un way. En esta situación, el camino alternativo es verificar en el propio mapa OSM en la web; se identificó, de esta manera, que se trata del río Machángara.

El análisis utilizando gráficos de caja se repitió para el número de *keys*, incluyendo la variable *source* (Fig. 8). De acuerdo a *building* se encontró medianas semejantes entre casas, edificios comerciales, residenciales y de apartamentos. Mayor mediana y mayor cantidad de valores atípicamente altos fueron encontrados en la categoría OTROS, donde se agrupó a

los demás valores; tiene sentido pensar que se tratan de establecimientos especializados (educativos, públicos...) y por ende presentan más *keys*. De acuerdo a *highway* poseen mayor mediana las categorías primaria, secundaria y terciaria. Cabe resaltar que en estos gráficos el ancho de la caja es proporcional al número de entidades en esa categoría; una consecuencia interesante de ello es que las tres categorías ya mencionadas contienen menos entidades -sus cajas son más estrechas- pero mayor cantidad de información, comparadas a las vías peatonales, residenciales, de servicio y otras. La misma conclusión es palpable en el análisis de acuerdo a *source*: mientras el calco de imágenes de Yahoo! y el reconocimiento de campo son fuentes de datos muy frecuentes -abarcan el 23.66% y el 23.02% de las entidades-, el calco de imágenes de Bing y la subida de *tracks* GPS son fuentes menos frecuentes pero relacionadas con mayor aporte de información por parte de los usuarios. La categoría OTROS también presenta mayor número mediano de *keys*, y abarca fuentes diversas como: mapa de la ciudad, conocimiento local, visita al campo, e incluso un usuario que escribió "viviendo aquí pues!". Los valores de NA en *source* están relacionados con menos cantidad de *keys*, pero hay que considerar que el propio *source* es un *key*; esto aplica a todos los análisis.

Si bien es posible aducir que "KG Ground Survey 2016" -el segundo *source* más frecuente- es la fuente que explica el pico abrupto de generación de nodos en abril de 2016 (Fig. 1), la manera correcta de comprobar esta hipótesis es analizar una etiqueta denominada *timestamp*. Desafortunadamente se comprobó que, al descargar los datos OSM a través de QGIS, no se obtiene esta y otra información de las entidades; además, aun obteniendo los datos completos -directamente desde la web

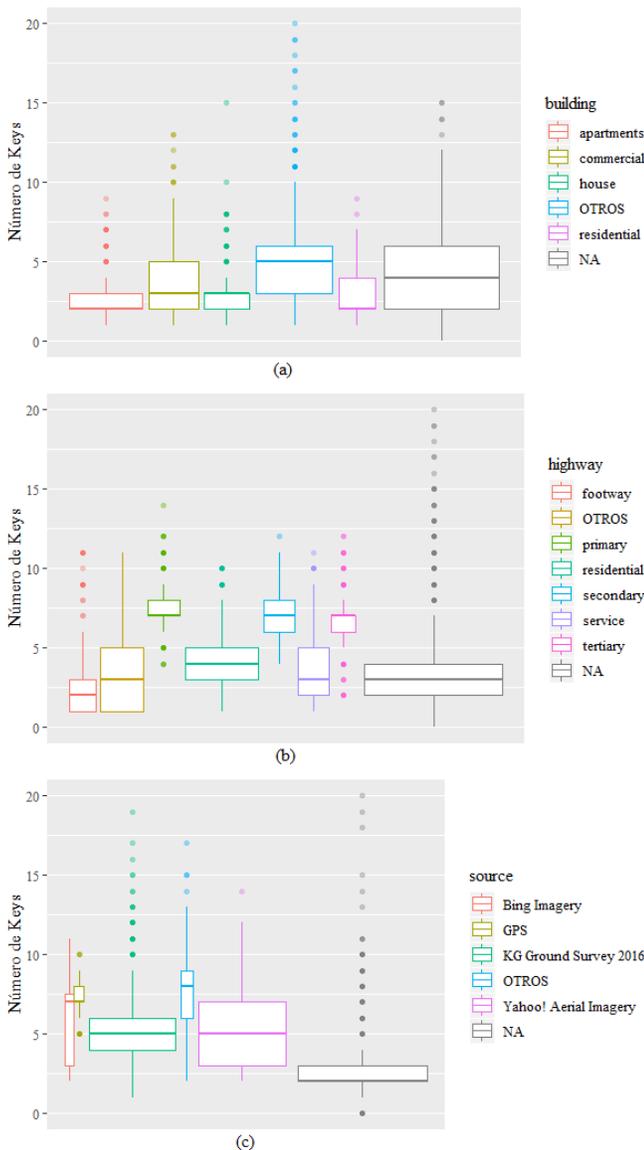


Fig. 8. Gráficos de caja correspondientes a número de keys, según las variables building (a), highway (b) y source (c).

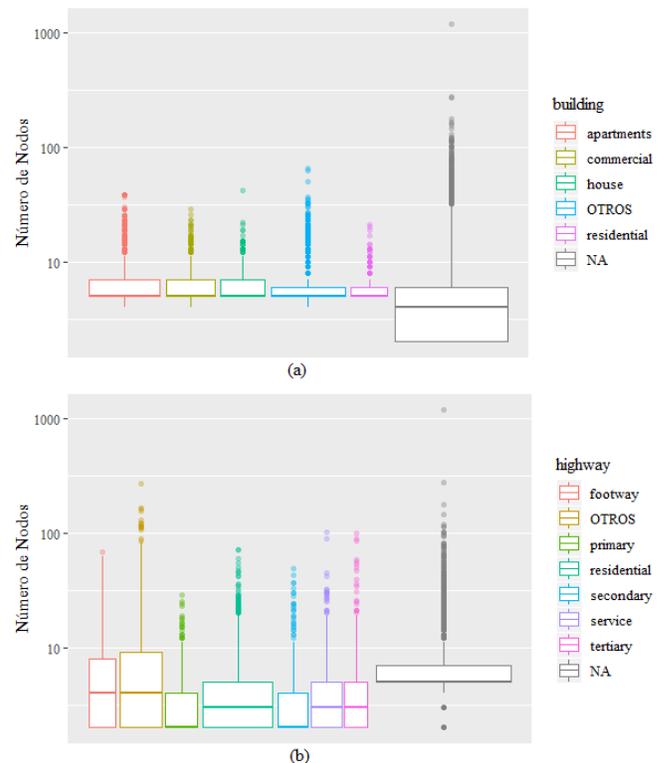


Fig. 7. Gráficos de caja correspondientes a número de nodos, según las variables building (a) y highway (b).

de OSM-, las herramientas de QGIS simplemente no tienen capacidad para obtener esas etiquetas. La comprobación de cuándo fueron aportados estas entidades, así como otros análisis relacionados con el aspecto temporal de las mismas, requerirán de la implementación de herramientas avanzadas de exploración de datos.

## V. CONCLUSIONES

Se demostró la utilidad del EDA y el poder de un ambiente de programación estadístico, como lo es R, en el caso práctico de dar un primer vistazo a un conjunto valioso pero poco explorado -en el área de estudio escogida- de datos libres. La información VGI presenta tanto oportunidades como defectos que requerirán mayor investigación en el futuro, a fin de respaldar su utilización. El presente estudio posee la ventaja de ser aplicable a una escala mayor con pocas modificaciones; una ventaja oportuna, dada las características de Big Data que presentan este y otros conjuntos de datos espaciales.

Posibles maneras de continuar esta investigación incluyen analizar los *keys* que no son extraídos automáticamente por la herramienta de QGIS, pero se encuentran inmersos en la estructura XML de los datos OSM; tales *keys* (*user*, *version*, *timestamp*) pueden aportar información valiosa al EDA y demás análisis que deberían ser realizados posteriormente, en estos datos libres. También es factible integrar aún más el GIS con el ambiente de programación, analizando cómo los aspectos espaciales (longitud de las entidades, exactitud posicional) se relacionan con los aspectos aquí discutidos.

## REFERENCIAS

- [1] M. Goodchild, "Citizens as sensors: the world of volunteered geography," *GeoJournal*, vol. 69, pp. 211–221, Nov. 2007.
- [2] D. Sui, "The Wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS," *Computers, Environment and Urban Systems*, vol. 32, no. 1, pp. 1–5, Ene. 2008.
- [3] D. Coleman, Y. Georgiadou, J. Labonte, "Volunteered Geographic Information: The Nature and Motivation of Producers", *International Journal of Spatial Data Infrastructures Research*, vol. 4, 2009.
- [4] J. Jokar, A. Zipf, P. Mooney, M. Helbich, "An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications," en *OpenStreetMap in GIScience. Experiences, Research, and Applications*, Basilea, Suiza: Springer International Publishing, 2015, ch. 1, pp. 1–15.
- [5] OSMStats (Ago. 2018). Statistics of the free wiki world map (OpenStreetMap.org) created in Central European Time (CET) [Online]. Disponible en <http://osmstats.neis-one.org/?item=countries>
- [6] M. Komorowski, D. Marshall, J. Saliccioli, Y. Crutain, "Exploratory Data Analysis," en *Secondary Analysis of Electronic Health Records*, New York, Estados Unidos: Springer International Publishing, 2016, ch. 15, pp. 185–203.
- [7] B. Haig, "Commentary: Exploratory data analysis," *Frontiers in Psychology*, vol. 6, Ago. 2015.
- [8] M. Goodchild, "The quality of big (geo)data," *Dialogues in Human Geography*, vol. 3, no. 3, pp. 280–284, 2013.
- [9] OpenStreetMap (Ago. 2018). List of OSM-based services [Online]. Disponible en [https://wiki.openstreetmap.org/wiki/List\\_of\\_OSM-based\\_services](https://wiki.openstreetmap.org/wiki/List_of_OSM-based_services)
- [10] P. Hashemi, R. Abbaspour, "Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept", en *OpenStreetMap in GIScience. Experiences, Research, and Applications*, Basilea, Suiza: Springer International Publishing, 2015, ch. 2, pp. 19–36.
- [11] M. Sax-Barnett, "An Introduction to OpenStreetMap," presentado en *State of the Map US*, Washington D.C., Estados Unidos, Abr. 2014.
- [12] OSGeoLive (Ago. 2018). SpatiaLite [Online]. Disponible en [http://live.osgeo.org/es/overview/spatialite\\_overview.html](http://live.osgeo.org/es/overview/spatialite_overview.html)

- [13] Comprehensive R Archive Network (Ago. 2018). Tibbles [Online]. Disponible en <https://cran.r-project.org/web/packages/tibble>
- [14] R Core Team (Ago. 2018). R: A language and environment for statistical computing [Online]. Disponible en <https://www.R-project.org/>



**Roger Castro** se encuentra actualmente se encuentra obteniendo el título de Ingeniero Geógrafo y del Medio Ambiente de la Universidad de las Fuerzas Armadas ESPE, Sangolquí Ecuador. Su investigación se concentra en el Análisis Espacial de Datos y en los Datos Libres.



**Marco Luna** es desde febrero de 1995 profesor en la Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador. Obtuvo el título de Ingeniero Geógrafo de la Escuela Politécnica del Ejército, Quito, Ecuador, en 1994. Recibió los títulos de Máster en Energía y Medio Ambiente de la Escuela Politécnica del Ejército, en 2008, y de Máster en Estadística Aplicada de la Escuela Politécnica Nacional, Quito, Ecuador, en 2012. Obtuvo el título de Doctor en Ingeniería Geográfica de la Universidad Politécnica de Madrid, Madrid, España, en 2017. Sus investigaciones se concentran en las Ciencias Geodésicas y el Análisis de Series Temporales.